PQStat Software
Statystyczne Oprogramowanie Obliczeniowe

Podręcznik Użytkownika - PQStat

Barbara Więckowska

Copyright © 2010-2026 U-smile Stats

Wszelkie prawa zastrzeżone

Do wersji 2.0.0

Wydawca: U-smile Stats, Poznań
Data wydania: 8 października 2026
ISBN 978-83-983915-0-4

www.pqstat.pl

Spis treści

1 WYMAGANIA SPRZĘTOWE 8

2 INSTALACJA PQSTAT 8

3 USTAWIENIA 9

4 NARZĘDZIA AI 13

5 AI: INTERPRETATOR WYNIKÓW 13

6 KONTROLA JAKOŚCI DANYCH 17

7 OBSŁUGA PROGRAMU 20

7.1 PRACA Z ARKUSZAMI DANYCH . . . 22

7.1.1 DOŁĄCZANIE, USUWANIE I EKSPORTOWANIE ARKUSZY DANYCH . . . 22

7.1.2 UMIESZCZANIE DANYCH W ARKUSZU . . . 22

7.1.3 OKNO ARKUSZA DANYCH . . . 24

7.1.4 WŁAŚCIWOŚCI ZMIENNEJ . . . 25

7.1.5 EDYCJA DANYCH . . . 27

7.1.6 SORTOWANIE DANYCH . . . 29

7.1.7 ZAMIANA DANYCH SUROWYCH NA TABELĘ KONTYNGENCJI . . . 30

7.1.8 ZAMIANA TABELI KONTYNGENCJI NA DANE SUROWE . . . 30

7.1.9 PRZEKSZTAŁCANIE DANYCH MIĘDZY POSTACIĄ SZEROKĄ A DŁUGĄ . . . 31

7.1.10 FORMUŁY . . . 35

7.1.11 GENEROWANIE DANYCH . . . 38

7.1.12 BRAKI DANYCH . . . 39

7.1.13 TRANSFORMACJE . . . 42

7.1.14 STANDARYZACJA BEZPOŚREDNIA I POŚREDNIA . . . 46

7.1.15 SYMULACJA PRÓBKOWANIA . . . 47

7.1.16 MACIERZ PODOBIEŃSTWA . . . 50

7.2 PRACA Z ARKUSZAMI WYNIKÓW (RAPORTAMI) . . . 60

7.3 MENU I USTAWIENIA JĘZYKOWE PROGRAMU . . . 61

7.3.1 JĘZYK . . . 61

7.3.2 MENU . . . 61

8 ORGANIZACJA PRACY Z PROGRAMEM 63

8.1 FORMA ORGANIZACJI DANYCH . . . 63

8.2 OBSZAR ROBOCZY ARKUSZA . . . 65
8.2.1 Organizowanie zmiennych w zestawy . . . 65
8.2.2 Aktywacja/dezaktywacja przypadków . . . 66
8.2.3 Zaznaczenie obszaru . . . 67
8.2.4 Zapisane zaznaczenie . . . 68
8.2.5 Filtr danych . . . 68
8.3 WIELOKROTNE WYKONYWANIE ANALIZ . . . 71
8.4 ORGANIZOWANIE RAPORTÓW W ZESTAWY . . . 71
8.5 INFORMACJE ZWRACANE DO RAPORTU . . . 72
8.6 OZNACZANIE WYNIKÓW ISTOTNYCH STATYSTYCZNIE . . . 72

9 WYKRESY 73
9.1 GALERIA WYKRESÓW . . . 73
9.1.1 Wykresy kolumnowe . . . 74
9.1.2 Wykresy błędów estymatora . . . 79
9.1.3 Wykresy typu ramka-wąsy . . . 80
9.1.4 Wykresy typu skrzypcowy . . . 82
9.1.5 Wykresy punktowe . . . 83
9.1.6 Wykresy liniowe . . . 85
9.2 LOKALNE TECHNIKI WYGŁADZANIA LINIOWEGO . . . 86
9.2.1 LOWESS . . . 86
9.2.2 Wygładzanie jądrowe . . . 86
9.3 ESTYMACJA JĄDROWA . . . 90
9.3.1 Jednowymiarowy estymator jądrowy . . . 90
9.3.2 Wykres Blanda-Altmana . . . 94
9.4 Macierze korelacji . . . 98

10 MOC TESTU I LICZNOŚĆ PRÓBY 107
10.1 Niezbędna liczność próby . . . 107
10.2 Moc i liczność dla testu . . . 110
10.2.1 Test t-Studenta dla pojedynczej próby . . . 112
10.2.2 Test t-Studenta dla grup zależnych . . . 115
10.2.3 Test t-Studenta dla grup niezależnych . . . 117
10.2.4 Test chi-kwadrat wariancji pojedynczej próby . . . 118
10.2.5 Test chi-kwadrat dwóch wariancji Fishera-Snedecora . . . 120
10.2.6 Test chi-kwadrat zgodności . . . 121
10.2.7 Test chi-kwadrat (RxC) . . . 122

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

10.2.8 Test dwóch niezależnych proporcji, Chi-kwadrat (2x2) . . . 124
10.2.9 ANOVA dla grup niezależnych . . . 125
10.2.10 Test dla jednej proporcji . . . 127

11 ANALIZY OPISOWE 128
11.1 Skale pomiarowe . . . 129
11.2 Tabele . . . 130
11.2.1 Tabele liczności i rozkłady empiryczne danych . . . 130
11.2.2 Raport tabeli . . . 135
11.2.3 Analizy dla tabel kontyngencji . . . 140
11.3 Statystyki opisowe . . . 144
11.3.1 Miary tendencji centralnej . . . 145
11.3.2 Inne miary położenia . . . 146
11.3.3 Miary rozproszenia . . . 147
11.3.4 Inne atrybuty rozkładu . . . 148
11.4 Zestawienia opisowe . . . 150

12 ROZKŁADY PRAWDOPODOBIEŃSTWA 156
12.1 CIĄGŁE ROZKŁADY PRAWDOPODOBIEŃSTWA . . . 158
12.2 KALKULATOR FUNKCJI DYSTRYBUCJI . . . 161

13 TESTOWANIE HIPOTEZ 165

14 TESTY NORMALNOŚCI ROZKŁADU 168
14.0.1 Testy normalności jednowymiarowej . . . 168
14.0.2 Testy normalności wielowymiarowej . . . 177

15 PORÓWNANIE - 1 GRUPA 182
15.1 TESTY PARAMETRYCZNE . . . 183
15.1.1 Test t-Studenta dla pojedynczej próby . . . 183
15.1.2 Test chi-kwadrat wariancji pojedynczej próby . . . 186
15.2 TESTY NIEPARAMETRYCZNE . . . 190
15.2.1 Test Wilcoxona (rangowanych znaków) . . . 190
15.2.2 Test chi-kwadrat zgodności . . . 193
15.2.3 Testy dla jednej proporcji . . . 196

16 PORÓWNANIE - 2 GRUPY 200
16.1 TESTY PARAMETRYCZNE . . . 201
16.1.1 Test Fishera-Snedecora . . . 201

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

16.1.2 Test t-Studenta dla grup niezależnych . . . 203
16.1.3 Test t-Studenta z korektą Cochrana-Coxa . . . 203
16.1.4 Test t-Studenta dla grup zależnych . . . 208
16.2 TESTY NIEPARAMETRYCZNE . . . 211
16.2.1 Test U Manna-Whitneya . . . 211
16.2.2 Test Wilcoxona (kolejności par) . . . 216
16.2.3 Testy chi-kwadrat . . . 220
16.2.4 Test chi-kwadrat dla dużych tabel . . . 221
16.2.5 Test Fishera dla dużych tabel . . . 224
16.2.6 Test chi-kwadrat dla małych tabel . . . 228
16.2.7 Poprawki testu chi-kwadrat dla małych tabel . . . 231
16.2.8 Test chi-kwadrat dla trendu . . . 232
16.2.9 Relatywne Ryzyko i Iloraz Szans . . . 237
16.2.10 Test Z dla dwóch niezależnych proporcji . . . 239
16.2.11 Test McNemara, test wewnętrznej symetrii Bowkera . . . 244
16.2.12 Test Z dla dwóch zależnych proporcji . . . 250

17 PORÓWNANIE - WIĘCEJ NIŻ 2 GRUPY 254

17.1 TESTY PARAMETRYCZNE . . . 255
17.1.1 ANOVA dla grup niezależnych . . . 255
17.1.2 Kontrasty i testy POST-HOC . . . 259
17.1.3 ANOVA dla grup niezależnych z korektą F^* i F'' . . . 265
17.1.4 Test Browna-Forsythe'a i test Levene'a . . . 273
17.1.5 ANOVA powtarzanych pomiarów . . . 275
17.1.6 Sferyczność Mauchly'ego . . . 279
17.1.7 ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA . . . 280
17.2 TESTY NIEPARAMETRYCZNE . . . 288
17.2.1 ANOVA Kruskala-Wallisa . . . 288
17.2.2 Test Jonckheere-Terpstra dla trendu . . . 294
17.2.3 Test wariancji rang Conovera . . . 298
17.2.4 ANOVA Friedmana . . . 302
17.2.5 Test Page dla trendu . . . 309
17.2.6 ANOVA Durbina (brakujących danych) . . . 311
17.2.7 ANOVA Skillings-Mack (brakujących danych) . . . 317
17.2.8 Test chi-kwadrat dla wielowymiarowych tabel kontyngencji . . . 320
17.2.9 ANOVA Q-Cochrana . . . 323

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

18 WIELOKROTNE PORÓWNANIA **328**

19 HOTELLING T-KWADRAT **332**

19.0.1 Test T-kwadrat Hotellinga dla pojedynczej próby . . . 332
19.0.2 Test T-kwadrat Hotellinga dla grup zależnych . . . 336
19.0.3 Test T-kwadrat Hotellinga dla grup niezależnych . . . 340
19.0.4 Test T-kwadrat Hotellinga dla grup niezależnych z korektą dla różnych wariancji . . . 345
19.0.5 Test Box'a równości macierzy kowariancji . . . 347

20 JEDNOCZYNNIKOWA MANOVA **349**

20.0.1 Jednoczynnikowa MANOVA dla grup niezależnych . . . 349

21 ANALIZA DLA WARSTW **356**

21.1 Metoda Mantela-Haenszela dla wielu tabel $2x2$ . . . 356
21.1.1 Iloraz szans Mantela-Haenszela . . . 357
21.1.2 Relatywne ryzyko Mantela-Haenszela . . . 361

22 KORELACJA **364**

22.1 TESTY PARAMETRYCZNE . . . 365
22.1.1 Współczynniki korelacji liniowej . . . 365
22.1.2 Istotność współczynnika korelacji Pearsona . . . 366
22.1.3 Istotność współczynnika nachylenia prostej . . . 370
22.1.4 Predykcja . . . 371
22.1.5 Porównanie współczynników korelacji . . . 371
22.1.6 Porównanie nachylenia prostych regresji . . . 373
22.2 TESTY NIEPARAMETRYCZNE . . . 375
22.2.1 Współczynniki korelacji monotonicznej . . . 375
22.2.2 Test t do sprawdzania istotności współczynnika korelacji rangowej Spearmana . . . 376
22.2.3 Test do sprawdzania istotności współczynnika korelacji $\tilde{\tau}$ Kendalla . . . 379
22.2.4 Współczynniki tabel kontyngencji i ich istotność statystyczna . . . 381

23 ANALIZA ZGODNOŚCI **387**

23.1 TESTY PARAMETRYCZNE . . . 388
23.1.1 Współczynnik korelacji wewnątrzklasowej i test badający jego istotność . . . 388
23.2 TESTY NIEPARAMETRYCZNE . . . 395
23.2.1 Współczynnik zgodności Kendalla i test badający jego istotność . . . 395
23.2.2 Współczynnik Kappa Cohena i test badający jego istotność . . . 398
23.2.3 Współczynnik Kappa Fleissa i test badający jego istotność . . . 404

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

24 TESTY DIAGNOSTYCZNE **409**

24.0.1 Ocena testu diagnostycznego . . . 409
24.0.2 Krzywa ROC . . . 414
24.0.3 Wybór optymalnego punktu odcięcia . . . 417
24.0.4 Porównywanie krzywych ROC . . . 423

25 MODELE WIELOWYMIAROWE **430**

25.0.1 Dopasowanie grup . . . 430
25.0.2 Przygotowanie zmiennych do analizy w modelach wielowymiarowych . . . 442
25.1 LINIOWA REGRESJA WIELORAKA . . . 446
25.1.1 Liniowa regresja wieloraka . . . 446
25.1.2 Predykcja na podstawie modelu i walidacja zbioru testowego . . . 462
25.1.3 Walidacja i predykcja dla regresji wielorakiej . . . 462
25.2 PORÓWNYWANIE MODELI LINIOWEJ REGRESJI WIELORAKIEJ . . . 466
25.2.1 Porównywanie modeli regresji wielorakiej . . . 466
25.3 REGRESJA LOGISTYCZNA . . . 475
25.3.1 Regresja logistyczna . . . 475
25.3.2 Predykcja na podstawie modelu i walidacja zbioru testowego . . . 499
25.3.3 Walidacja i predykcja dla regresji logistycznej . . . 499
25.4 PORÓWNYWANIE MODELI REGRESJI LOGISTYCZNEJ . . . 507
25.4.1 Porównywanie modeli regresji logistycznej . . . 507
25.4.2 ANOVA czynnikowa – GLM . . . 524
25.4.3 ANCOVA . . . 537
25.4.4 Efekt mediacji . . . 545

26 MODELE MIESZANE **549**

26.0.1 Wprowadzenie: dane zależne, efekty stałe i efekty losowe . . . 549
26.1 REGRESJA WIELORAKA MIESZANA . . . 553
26.1.1 Regresja wieloraka mieszana (liniowy model mieszany) . . . 553
26.1.2 Walidacja i predykcja dla regresji wielorakiej mieszanej . . . 568
26.1.3 Porównywanie modeli regresji wielorakiej mieszanej . . . 570
26.2 REGRESJA LOGISTYCZNA MIESZANA . . . 575
26.2.1 Regresja logistyczna mieszana (uogólniony model mieszany) . . . 575
26.2.2 Walidacja i predykcja dla regresji logistycznej mieszanej . . . 583
26.2.3 Porównywanie modeli regresji logistycznej mieszanej . . . 585
26.3 REGRESJA PH COXA Z EFEKTEM LOSOWYM . . . 590
26.3.1 Regresja PH Coxa z efektem losowym (frailty) . . . 590

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

26.3.2 Porównywanie modeli regresji PH Coxa z efektem losowym . . . 597

27 REDUKCJA WYMIARÓW I GRUPOWANIE 601

27.0.1 Analiza składowych głównych . . . 601

27.1 ANALIZA SKUPIEŃ . . . 613

27.1.1 Metody hierarchiczne . . . 613

27.1.2 Metoda k-średnich . . . 620

28 ANALIZA PRZEŻYCIA 626

28.0.1 Tabele przeżycia . . . 627

28.0.2 Krzywe Kaplana-Meiera . . . 631

28.0.3 Porównywanie krzywych przeżycia . . . 634

28.0.4 Regresja proporcjonalnego hazardu Coxa . . . 643

28.0.5 Porównywanie modeli regresji PH Coxa . . . 647

28.0.6 Regresja Coxa ze zmiennymi zależnymi od czasu . . . 657

28.0.7 Analiza ryzyk konkurencyjnych – model Fine'a-Graya . . . 663

29 META-ANALIZA 670

29.0.1 Wprowadzenie do meta-analizy i przygotowanie danych . . . 670

29.0.2 Efekt podsumowujący – efekt stały i efekt zmienny . . . 675

29.0.3 Badanie heterogeniczności . . . 679

29.0.4 Badanie wrażliwości . . . 681

29.0.5 Badanie asymetrii (obciążenie publikacyjne) . . . 683

29.0.6 Meta-analiza skumulowana . . . 687

29.0.7 Porównanie grup . . . 689

29.0.8 Meta-regresja . . . 698

30 ANALIZA RZETELNOŚCI 703

30.0.1 Rzetelność skali . . . 703

31 ZESTAWIENIA TESTÓW 710

32 KREATOR 719

33 INNE UWAGI 720

33.1 FORMATY PLIKÓW . . . 720

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1 WYMAGANIA SPRZĘTOWE

Minimalne wymagania dla komputera, aby można było na nim uruchomić PQStat to:

- Procesor: Intel Pentium II 500MHz
- Pamięć RAM: 256MB
- Karta Graficzna: SVGA 800x600
- Wolne miejsce na dysku: 0.2GB
- W przypadku wersji na nośniku CD: CD-ROM
- Inne : klawiatura, mysz
- Microsoft Windows 10 lub 11 (64-bit)

2 INSTALACJA PQSTAT

Proces instalacji rozpoczyna się od uruchomienia instalatora aplikacji - pliku PQStat-setup_x86-FULL (dla wersji 64-bitowej PQStat-setup_x64-FULL.exe).

Instalator wyświetli najpierw ekran powitalny – aby rozpocząć instalację, klikamy ”Dalej”. Instalacja wymaga zaakceptowania umowy licencyjnej programu: jeśli zgadzasz się z jej postanowieniami, zaznacz opcję ”Akceptuję warunki umowy” i przejdź ”Dalej”; w przeciwnym razie wybierz ”Nie akceptuję warunków umowy” i kliknij Anuluj, aby zakończyć instalację.

Kolejne okno pozwala zmienić docelowe miejsce instalacji i pokazuje ilość dostępnego miejsca na dysku – zalecane jest instalowanie aplikacji w domyślnie podanej lokalizacji.

Dalej można wybrać pełną instalację programu albo instalację bez przykładowych plików danych. Pliki te są wykorzystywane w pomocy programu.

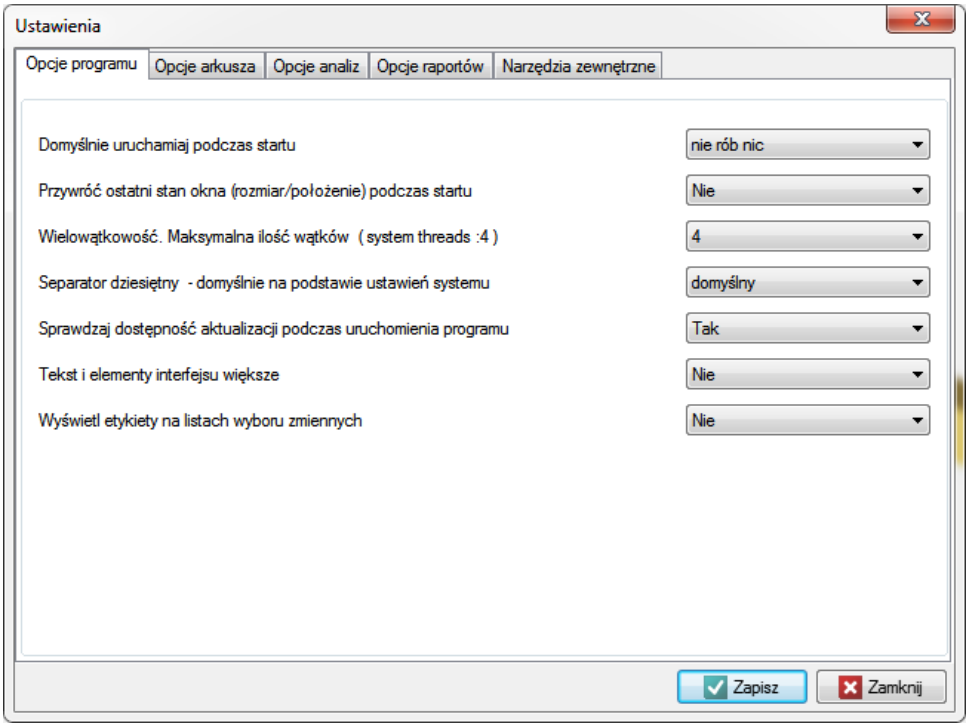
Kolejne okno pozwala zmienić nazwę skrótu, który zostanie utworzony w Menu Start systemu Windows.

Następnie można zdecydować, czy instalator ma utworzyć ikonę programu na pulpicie oraz na pasku szybkiego uruchamiania – zaznacz interesujące Cię opcje i przejdź ”Dalej”.

Ostatnie okno przed rozpoczęciem instalacji pokazuje podsumowanie wybranych opcji. Aby rozpocząć instalację, klikamy przycisk "Instaluj".

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

3 USTAWIENIA



Ogólne ustawienia dotyczące opcji programu, arkusza danych, wykonywania analiz, zwracanych raportów i stosowania narzędzi zewnętrznych możliwe są do zmiany po wybraniu menu Edycja→Ustawienia.

Opcje programu

- Domyślnie uruchamiaj podczas startu - pozwala na wywoływanie podczas startu programu: nowego pustego projektu, projektu który był ostatnio wykonywany, lub nie wywoływania żadnego projektu.
- Przywróć ostatni stan okna (rozmiar/położenie) podczas startu - umożliwia uruchomienie programu w ten sposób, by jego okno znalazło się w miejscu i miało rozmiar okna programu wyświetlanego bezpośrednio przed poprzednim zamknięciem.
- Wielowątkowość. Maksymalna ilość wątków (system threads: 4) - daje możliwość wykonywania jednocześnie zadań w wielu wątkach (maksymalnie w wątkach ośmiu).
- Separator dziesiętny - domyślnie na podstawie ustawień systemu - może być ustawiony jako przecinek lub kropka, bądź też pobierany domyślnie zgodnie z ustawieniami systemu operacyjnego w który program działa.
- Sprawdzaj dostępność aktualizacji podczas uruchamiania programu - daje możliwość sprawdzania i w konsekwencji zainstalowania aktualizacji programu lub rezygnacji z tej usługi.
- Tekst i elementy interfejsu większe - pozwala zwiększyć czcionki i inne elementy interfejsu, co zmniejszy ilość wyświetlanych informacji ale może ułatwić ich czytelność.
- Wyświetl etykiety na listach wyboru zmiennych - Daje możliwość wyświetlania etykiet zmiennych (długich nazw zmiennych) na listach wyboru tych zmiennych lub jako podpowiedź do nazw wyświetlanych w tych listach. Opcja ta przyniesie pożądany skutek, gdy nazwy zmiennych posiadają przypisane etykiety.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Opcje arkusza

- Maksymalna liczba kroków cofnięć w arkuszu - liczba możliwych kroków o jakie można cofnąć ostatnio wykonywane zadania w arkuszu - ma ona wpływ na szybkość pracy programu dlatego zalecana liczba kroków to 10.
- Maksymalna liczba komórek do zapamiętania w jednym kroku - liczba komórek arkusza, których jednoczesna zmiana może być uznana za jeden krok do zapamiętania przez program - ma ona wpływ na szybkość pracy programu dlatego zalecana liczba komórek to 5000.
- Czcionka arkusza danych - umożliwia ustawienie rodzaju czcionki, którą będą wyświetlane dane zawarte w arkuszu danych.
- Działanie podwójnego kliknięcia nagłówka kolumny - pozwala ustawić akcję uzyskaną po podwójnym kliknięciu nagłówka kolumny. Może to być: dopasowanie szerokości kolumny do zawartych w niej danych; wywołanie okna zarządzania zmienną Kody/Etykiety/Format; wywołanie okna Inspekcja zmiennej.

Opcje analiz

- Domyślny poziom istotności testów - daje możliwość zmiany standardowo proponowanego w oknach analiz statystycznych poziomu istotności 0.05.
- MachineEpsilon - informacja o ustawieniu wielkości maszynowego Epsilon czyli precyzji obliczeń (1E-16).
- Metoda dla kurtozy - informacja o ustawieniu metody wyliczania kurtozy (g2, czy b2 Pearson'a), gdzie rozkład normalny charakteryzuje zerowa kurtoza g2 lub b2 bliska wartości trzy.
- Porządek sortowania tabel kontyngencji - daje możliwość posortowania wierszy i kolumn tabeli kontyngencji w sposób rosnący lub malejący. Opcja ta wpływa nie tylko na wyświetlanie wyników w tabelach i na wykresach, ale ma również wpływ na te analizy oparte na tabelach kontyngencji, dla których kolejność występowania kategorii jest ważna.
- Metoda alfabetycznego sortowania wartości tekstowych - umożliwia sortowanie danych oraz wyników zapisanych w postaci tekstowej w sposób naturalny lub alfabetyczny. Naturalny porządek sortowania to poprawiony porządek alfabyteczny, gdzie liczby wielocyfrowe są traktowane niepodzielnie, tj. tak, jakby były pojedynczym znakiem. Na przykład w sortowaniu alfabetycznym „a11" będzie sortowane przed „a2", ponieważ „1" jest sortowane jako mniejsze niż „2", podczas gdy w sortowaniu naturalnym „a2" jest sortowane przed „a11", ponieważ „2" jest sortowane jako mniejsze niż „11"
- Etykiety dla wartości. Podczas analizy wyświetlaj: - daje wybór sposobu wyświetlania w wartości lub etykiet odpowiadających obowiązującym wartościom podczas korzystania z filtru lub w przypadku, gdy analiza wymaga ustawienia poszczególnych wartości wybranej zmiennej.

Opcje raportów

- Precyzja wyświetlanych wartości obliczeń - pozwala ustawić maksymalną liczbę miejsc dziesiętnych podawanych w raportach dla liczb rzeczywistych.
- Precyzja wyświetlanych wartości procentowych - pozwala ustawić maksymalną liczbę miejsc dziesiętnych podawanych w raportach dla liczb podanych w procentach.
- Postać wykładnicza dla wartości p - umożliwia określenie sposobu wyświetlania wartości p testów statystycznych w postaci liczbowej (ze zdefiniowaną liczbą miejsc dziesiętnych) lub w notacji naukowej .

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Kolor wartości p poniżej poziomu istotności - pozwala ustawić kolor, jaki zostanie użyty do oznaczenia wyników istotnych statystycznie na zadanym poziomie.
- Nazwa raportu w drzewie nawigacji - daje wybór rodzaju informacji o wykonanych analizach dokładanej do nazwy analizy w drzewie nawigacji. Możliwe jest podawanie samej nazwy testu lub nazwy wraz z: godziną jego wykonania lub opisem testu, lub wybranym filtrem, lub użytą zmienną grupującą, lub nazwami zmiennych biorącymi udział w analizie.
- Opisuj wartości zmiennych poprzez etykiety - Daje możliwość wyświetlania etykiet zmiennych (długich nazw zmiennych) w raportach analiz. Opcja ta przyniesie pożądany skutek, gdy nazwy zmiennych posiadają przypisane etykiety.
- Czcionka raportu - umożliwia ustawienie rodzaju czcionki, którą będą wyświetlane opisy wyników zawarte w raportach.
- Domyślna wysokość/szerokość wykresów - umożliwia ustawienie obowiązującej domyślnie liczby pikseli dla wysokości i dla szerokości wykresu.
- Domyślne szablony wykresów - daje możliwość wstępnego ustawienia ważnych dla użytkownika opcji wykresu.

Narzędzia zewnętrzne

- Ścieżka do programu gnuplot - Aby możliwe było generowanie wykresu trójwymiarowego dla analizy jądrowego estymatora gęstości 2D należy zainstalować w systemie program gnuplot.
- Ścieżka do modułu IO dla SPSS Statistics - Aby możliwy był odczyt plików sav (format zapisu danych programu IBM SPSS Statistics) wymagany jest moduł producenta IBM IO SPSS, w tym celu należy pobrać plik $IO\_Module\_for\_SPSS\_Statistics\_20001.zip$. Więcej informacji na stronie `http://manuals.pqstat.pl/statpqpl:installpl`

Źródło modelu AI - PQStat może korzystać z lokalnie lub zdalnie uruchomionego modelu językowego (AI). Dostępne są trzy niezależnie zapamiętywane źródła modelu, między którymi można się dowolnie przełączać bez utraty wprowadzonych ustawień pozostałych:

- Lokalny (ten komputer) - korzysta z modelu uruchomionego na tym samym komputerze co PQStat. W polu Silnik lokalny wybieramy program, który udostępnia model:
  - Ollama - wymaga podania pola Adres URL API (domyślnie `http://localhost:11434`) oraz pola Model - nazwy modelu pobranego wcześniej poleceniem `ollama pull`. Przycisk Odśwież listę pobiera z serwera listę aktualnie dostępnych (pobranych) modeli.
  - LM Studio - korzysta z lokalnego serwera LM Studio, który udostępnia model przez API zgodne z OpenAI (bez klucza). LM Studio pobieramy ze strony `https://lmstudio.ai` (gdy PQStat nie wykryje serwera, sam zaproponuje otwarcie tej strony). W LM Studio należy pobrać model, a następnie włączyć serwer: Settings → Local Model API → przełącznik Local API server (ma pokazywać Running); pole Base URL w tym samym oknie pokazuje adres do wpisania w PQStat. Przy włączonej opcji Just-in-time model loading model jest ładowany automatycznie przy pierwszym zapytaniu. W PQStat podajemy Adres URL API (domyślnie `http://localhost:1234/v1`) i wybieramy Model - przyciski Odśwież listę i Testuj połączenie pobierają listę modeli dostępnych w LM Studio. Jeśli wybrany model obsługuje obrazy, zaznaczamy Model obsługuje obrazy (wymuś).

  Ustawienia obu silników są zapamiętywane niezależnie - przełączenie silnika nie kasuje adresu ani modelu drugiego.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

– Zdalny serwer Ollama - korzysta z serwera Ollama uruchomionego na innym komputerze, dostępnego w sieci lokalnej lub przez VPN. Oprócz pól Adres URL API i Model można podać opcjonalny Klucz API, jeśli zdalny serwer wystawiony jest za pośrednikiem (reverse-proxy) wymagającym uwierzytelnienia, oraz zaznaczyć Wymagaj ważnego certyfikatu dla połączeń szyfrowanych (https). Przycisk Testuj połączenie sprawdza dostępność serwera niezależnie od tego, czy model został już wybrany, a Odśwież listę pobiera listę modeli dostępnych na tym serwerze.
– Chmura (API kompatybilne z OpenAI) - korzysta z dowolnej zewnętrznej usługi udostępniającej API zgodne ze standardem OpenAI (np. OpenAI, Google Gemini, LiteLLM API Gateway i inne bramki tego typu, w tym firmowe serwery pośredniczące). Wymaga podania pól Adres URL API (lista rozwijana podpowiada adresy popularnych dostawców) oraz Klucz API (przechowywanego w postaci zaszyfrowanej - checkbox Pokaż pozwala tymczasowo podejrzeć wpisywaną wartość). Po kliknięciu Testuj połączenie program próbuje pobrać z serwera listę dostępnych modeli - jeśli się to powiedzie, pole Model zamienia się w listę rozwijaną do wyboru. Jeśli dana usługa nie udostępnia takiej listy w standardowy sposób (np. natywne API Google wymaga do tego innego uwierzytelniania niż zwykły klucz), nazwę modelu należy wpisać ręcznie, a test połączenia zostanie wtedy wykonany jako rzeczywiste zapytanie do wskazanego modelu. Checkbox Model obsługuje obrazy (wymuś) pozwala ręcznie wymusić traktowanie wybranego modelu jako obsługującego obrazy, gdy program nie potrafi tego rozpoznać automatycznie.

Ustawienia AI dla wszystkich użytkowników komputera. Domyślnie ustawienia AI zapamiętywane są osobno dla każdego konta Windows. Zaznaczenie pola Zapisz jako domyślne dla wszystkich użytkowników tego komputera przed kliknięciem OK zapisuje wybrane źródło modelu (adres URL API, model, zaszyfrowany klucz API) w pliku `C:\ProgramData\PQStat\AI\pqstat_ai.ini`. Z tych ustawień korzysta każdy użytkownik komputera, który nie zmienił ustawień AI u siebie. Ustawienia komputera są tylko domyślne: użytkownik może je zmienić, a jego zmiany obowiązują wyłącznie dla jego konta. Pod polem widać, czy używane są ustawienia komputera, czy własne; przycisk Przywróć ustawienia komputera wraca do ustawień domyślnych. Klucz API zapisany dla komputera może zostać użyty przez każdego jego użytkownika.

Działy IT mogą ustawić te wartości już podczas instalacji, podając instalatorowi parametry: `/AIMODE=local|remot` `/AIENGINE=ollama|lmstudio` (dla `local`), `/AIURL=`*adres*, `/AIMODEL=`*model*, `/AIKEY=`*klucz*, opcjonalnie `/AIVERIFYCERT=0` i `/AIVISION=1`, np.

```
PQStat-setup_x64-FULL.exe /VERYSILENT /AIMODE=remote /AIURL=https://llm.uczelnia.pl
/AIMODEL=gemma2:9b
```

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

4 NARZĘDZIA AI

5 AI: INTERPRETATOR WYNIKÓW

Interpretator AI zamienia tabelę wyników wybranej analizy w opis słowny, napisany w stylu sekcji Wyników publikacji naukowej – gotowy do wklejenia do pracy lub raportu, ewentualnie do dalszej redakcji przez użytkownika. Opis generowany jest przez model językowy (AI) na podstawie treści raportu, rodzaju wykonanego testu oraz (opcjonalnie) własnych wskazówek badacza.

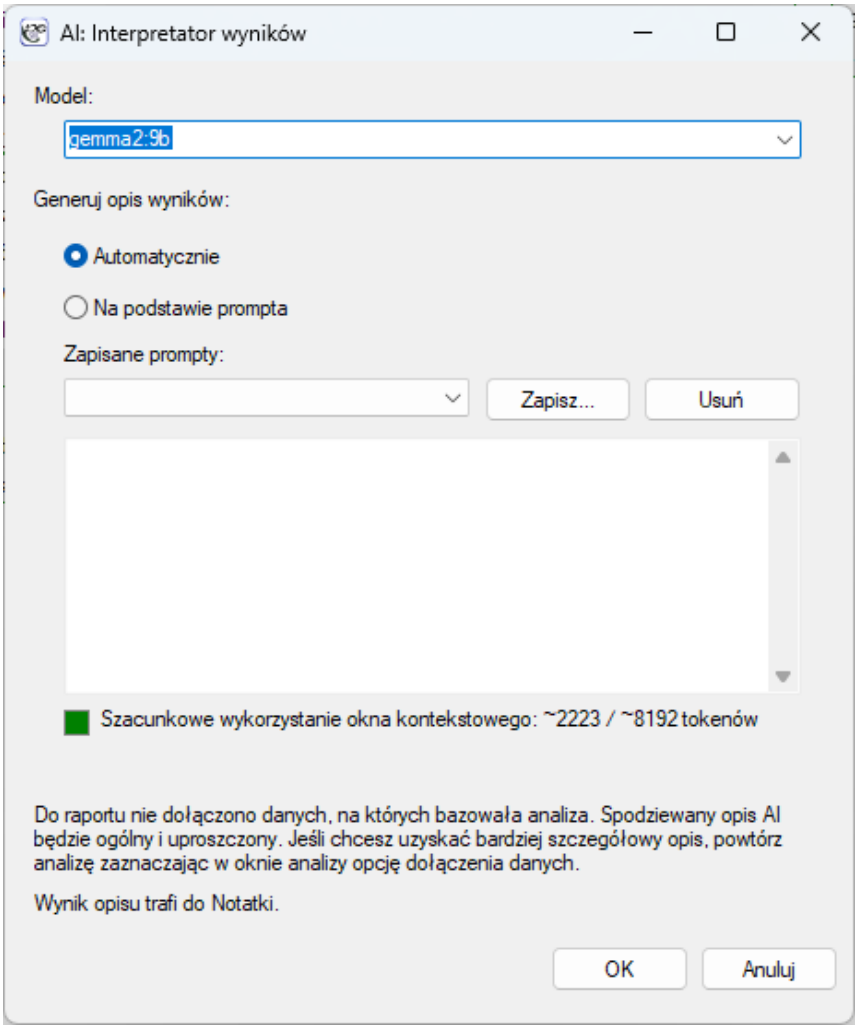
Uwaga! Funkcja wymaga skonfigurowanego źródła modelu AI w Ustawieniach (zakładka Narzędzia zewnętrzne) – lokalnego serwera Ollama, zdalnego serwera Ollama lub usługi zgodnej z API OpenAI. Zobacz rozdział USTAWIENIA po szczegóły konfiguracji poszczególnych źródeł.

JAK URUCHOMIĆ INTERPRETACJĘ

Interpretator wywołujemy dla aktywnego raportu na jeden z trzech równoważnych sposobów:

- menu Pomoc→AI: Interpretator wyników,
- menu kontekstowe raportu (prawy przycisk myszy na raporcie) → AI: Interpretator wyników,
- przycisk AI na pasku narzędzi raportu.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Krok 1. Ustaw aktywny raport, dla którego chcesz uzyskać opis.

Krok 2. Wywołaj interpretator jednym z powyższych sposobów.

Krok 3. W oknie opcji wybierz model AI (domyślnie ten wybrany w Ustawieniach, ale można go dla tej jednej interpretacji zmienić).

Krok 4. Zdecyduj, czy opis ma zostać wygenerowany Automatycznie, czy Na podstawie prompta – czyli z dodatkowymi, własnymi wskazówkami dla modelu (np. *"opisz wynik jednym zdaniem"* albo *"napisz po angielsku"*). Często używany własny prompt można zapisać pod wybraną nazwą przyciskiem Zapisz... i użyć ponownie przy kolejnej analizie.

Krok 5. Jeśli do raportu dołączony jest wykres, a wybrany model obsługuje obrazy, można zaznaczyć Dołącz wykres do zapytania, by model uwzględnił go w opisie.

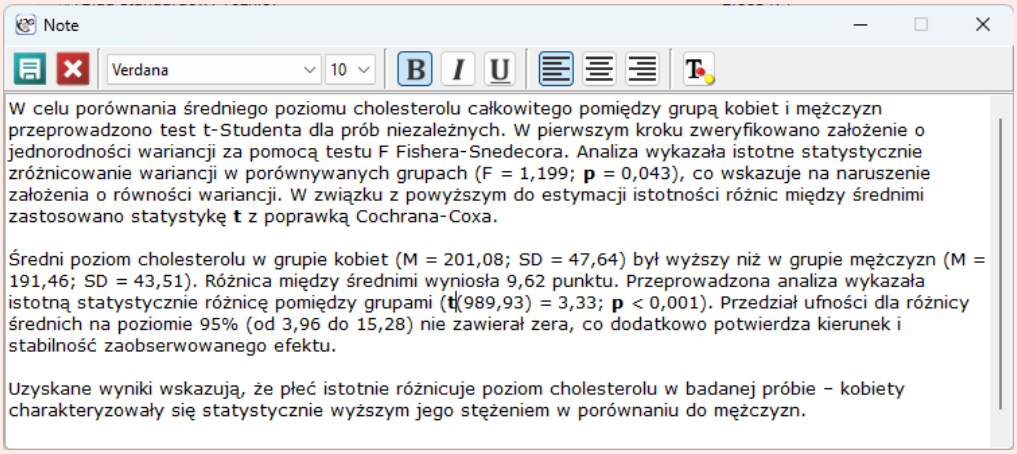
Krok 6. Uruchom generowanie przyciskiem OK.

Uwaga! Pasek pod wyborem modelu pokazuje szacunkowe wykorzystanie okna kontekstowego wybranego modelu (na zielono – z zapasem, na pomarańczowo – blisko granicy, na czerwono – przekroczone, co grozi niepełną lub urwaną odpowiedzią). Bardzo obszerne raporty mogą wymagać modelu z większym oknem kontekstowym.

WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PRZYKŁAD 5.1. Po wykonaniu dowolnego testu statystycznego (np. testu t-Studenta) wywołujemy interpretator z menu kontekstowego raportu i uruchamiamy generowanie w trybie Automatycznie.



Wygenerowany opis trafia do okna Notatki raportu (widocznej na raporcie lub w drzewie projektu, oznaczonej ikoną notatki), gdzie można go swobodnie edytować, uzupełnić lub poprawić przed ostatecznym zapisaniem.

INTERPRETACJA ZBIORCZA WIELU RAPORTÓW

Zaznaczając na liście raportów (widok Zbiorczy) kilka pozycji jednocześnie i wybierając z menu kontekstowego tej listy AI: Interpretator wyników, można wygenerować opisy dla wielu raportów naraz, bez okna opcji dla każdego z nich osobno – interpretacja przebiega wtedy automatycznie, z ustawieniami aktywnego źródła modelu z Ustawień.

ROZMOWA Z MODELEM Z POZIOMU NOTATKI

Wygenerowany opis nie musi być końcem pracy z modelem. W oknie Notatki raportu, pod polem tekstowym notatki, znajduje się panel rozmowy: pole na kolejne pytanie, przycisk Zapytaj (lub skrót Ctrl+Enter) oraz przycisk Nowa rozmowa. Zadane pytanie trafia do modelu razem z raportem i całą dotychczasową rozmową (wcześniejszymi pytaniami i odpowiedziami), więc model ”pamięta” swoją poprzednią odpowiedź – można prosić o wyjaśnienie fragmentu opisu, jego skrócenie, zmianę stylu, odniesienie do konkretnej wartości z tabeli itp. Odpowiedź jest dopisywana na końcu notatki (poprzedzona treścią pytania), gdzie podlega takiej samej edycji jak pierwotny opis. Historia rozmowy zapisywana jest w projekcie razem z raportem, dlatego do rozmowy można wrócić także po ponownym otwarciu pliku; panel pojawia się tylko w notatkach raportów, dla których wygenerowano opis AI.

Ponieważ przy każdym pytaniu do modelu wysyłana jest cała rozmowa, jej długość rośnie z każdą turą. Pod polem pytania wyświetlany jest wskaźnik wykorzystania okna kontekstowego modelu (kolor zielony, pomarańczowy lub czerwony – tak jak w oknie opcji interpretatora), uwzględniający raport, wszystkie dotychczasowe pytania i odpowiedzi oraz pytanie właśnie wpisywane. Gdy rozmowa przestaje mieścić się w oknie kontekstowym, program ostrzega przed wysłaniem, a przycisk Nowa rozmowa pozwala zacząć od początku: model dostaje wtedy ponownie tylko raport i nowe pytanie, natomiast treść notatki pozostaje bez zmian. Kolejne pytania kierowane są do aktywnego źródła modelu z Ustawień; wykres dołączony do pierwszego zapytania nie jest ponownie wysyłany.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Procedura: opis generowany jest na podstawie trzech elementów łączonych w jeden prompt przekazywany modelowi: (1) wspólnych zasad redagowania wyniku (styl sekcji Wyników publikacji naukowej, zakaz zgadywania wartości spoza tabeli, unikanie sformułowań w stylu ”odrzucamy hipotezę zerową”), (2) instrukcji właściwej dla konkretnego rodzaju wykonanego testu, oraz (3) treści samej tabeli wyników wraz z metadanymi zmiennych (etykietami wartości), jeśli są dostępne. Własny prompt badacza – jeśli podany – ma pierwszeństwo przed powyższymi zasadami, również w kwestii języka odpowiedzi.

Jeżeli do raportu dołączono opcję echa danych surowych, w prompcie przekazywana jest tylko próbka (do 10 pierwszych wierszy), by nie przeciążać okna kontekstowego modelu.

Rozmowa z modelem z poziomu notatki korzysta z tych samych interfejsów (Ollama, LM Studio, API zgodne z OpenAI), które są bezstanowe: serwer nie pamięta poprzednich zapytań, więc przy każdym pytaniu program wysyła od nowa pierwszy prompt (z raportem) i wszystkie dotychczasowe pary pytanie–odpowiedź. Model ”widzi” wcześniejszą rozmowę tylko dlatego, że jest ona dołączana do każdego kolejnego zapytania; stąd rosnące zużycie okna kontekstowego i sens przycisku Nowa rozmowa.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

6 KONTROLA JAKOŚCI DANYCH

Zanim przystąpi się do analizy statystycznej, warto sprawdzić, czy wczytane dane nie zawierają błędów, które mogłyby wypaczyć wyniki – np. kolumny liczbowej, która w rzeczywistości jest zakodowaną kategorią, wartości wyraźnie odstających od reszty, niestandardowych kodów braków danych (np. -99) wymieszanych z prawdziwymi wartościami, czy pojedynczych wpisów niepasujących do reszty kolumny (np. wartość ujemna w kolumnie wiek).

Kontrola jakości danych skanuje wszystkie kolumny aktywnego arkusza, oblicza dla każdej z nich zwięzłe podsumowanie statystyczne, a następnie oddaje to podsumowanie modelowi językowemu (AI), który wskazuje, co w danych wygląda podejrzanie i dlaczego. Funkcja nie modyfikuje danych – jedynie zwraca opisową listę spostrzeżeń do przejrzenia przez użytkownika.

Uwaga! Funkcja wymaga skonfigurowanego źródła modelu AI w Ustawieniach (zakładka Narzędzia zewnętrzne) – lokalnego serwera Ollama, zdalnego serwera Ollama lub usługi zgodnej z API OpenAI. Zobacz rozdział USTAWIENIA po szczegóły konfiguracji poszczególnych źródeł.

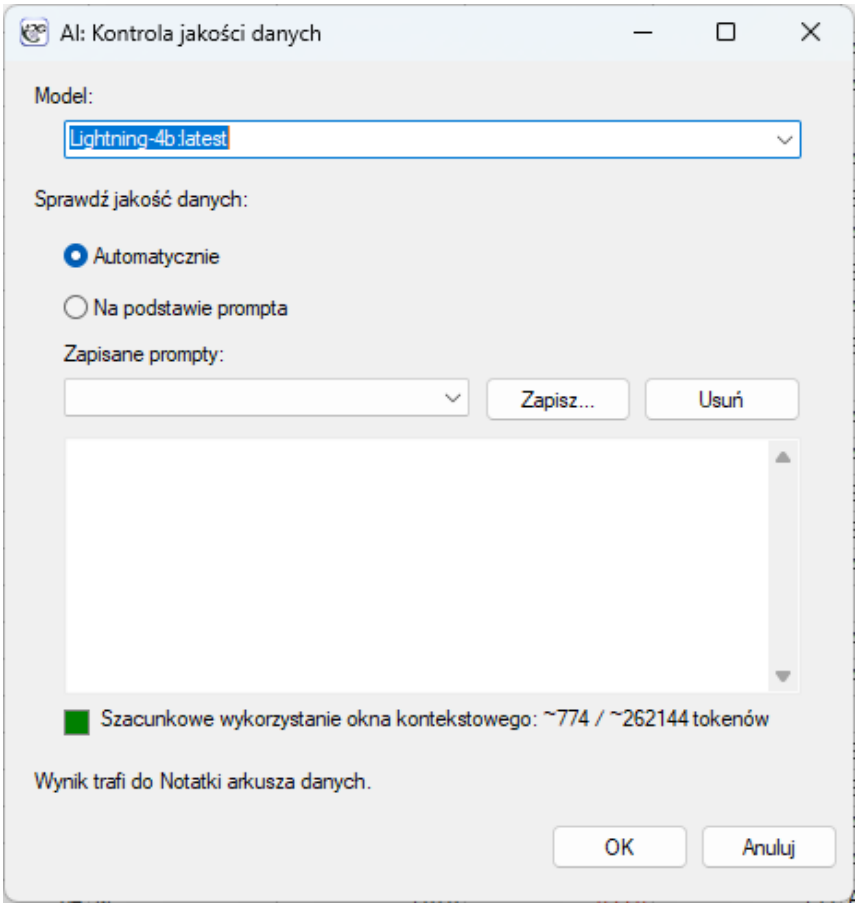
CO JEST SPRAWDZANE

- **Kolumny liczbowe wyglądające na zakodowaną kategorię** – gdy kolumna zawiera wyłącznie liczby całkowite, a liczba różnych wartości jest niewielka względem liczby obserwacji.
- **Wartości odstające** – wykrywane metodą rozstępu międzykwartylowego (IQR): za odstającą uznawana jest wartość leżąca poniżej $Q_1 - 1{,}5 \cdot IQR$ lub powyżej $Q_3 + 1{,}5 \cdot IQR$.
- **Podejrzane, powtarzające się wartości skrajne** – wartość leżąca daleko poza resztą rozkładu, która pojawia się w kolumnie więcej niż raz, co odróżnia ją od pojedynczego, przypadkowego odstępstwa i sugeruje niestandardowy kod braku danych (np. -99, 999).
- **Sensowność wartości względem nazwy kolumny** – np. czy kolumna wiek nie zawiera wartości ujemnych, czy kolumna wyglądająca na pytanie tak/nie nie zawiera innych wartości. Tego rodzaju ocena wymaga zrozumienia znaczenia nazwy kolumny, dlatego wykonuje ją wyłącznie model AI na podstawie przekazanego podsumowania – nie jest to reguła sprawdzana bezpośrednio przez program.

JAK URUCHOMIĆ KONTROLĘ

Okno kontroli jakości danych wywołujemy poprzez menu Dane→Kontrola jakości danych...

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Krok 1. Upewnij się, że w arkuszu aktywny jest zbiór danych, który chcesz sprawdzić.

Krok 2. Wybierz Dane→Kontrola jakości danych...

Krok 3. W oknie opcji wybierz model AI (lub zostaw domyślny) i zdecyduj, czy opis ma zostać wygenerowany automatycznie, czy na podstawie własnego prompta z dodatkowymi wskazówkami.

Krok 4. Uruchom analizę przyciskiem OK.

Uwaga! Funkcja sprawdza **wszystkie** niepuste kolumny aktywnego arkusza jednocześnie – nie ma możliwości ograniczenia sprawdzania do wybranych kolumn.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 6.1. (plik Data_Quality_Check.csv)
Plik zawiera 1000 wierszy (osób badanych) i 10 kolumn opisujących typowe dane medyczne: wiek, płeć, wzrost, wagę, ciśnienie skurczowe, grupę krwi, obecność choroby przewlekłej, liczbę wizyt, wynik testu psychometrycznego oraz tętno spoczynkowe. W każdej kolumnie celowo umieszczono po kilka nietypowych wartości – tabela poniżej pokazuje, jakiego rodzaju problem reprezentuje każda z nich. Po wczytaniu pliku wybieramy Dane→Kontrola jakości danych... i uruchamiamy analizę z modelem domyślnym.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Kolumna	Rodzaj problemu	Przykładowe wartości
wiek	wartości odstające / niemożliwe	$-2, 0, 150, 199$
plec	wartość spoza zbioru kategorii (K/M)	"3", "nieznana"
wzrost_cm	wartości odstające (błąd zapisu)	$17{,}0$ (zam. 170), $1210{,}0$
waga_kg	powtarzający się podejrzany kod	-99 ($5\times$ w różnych wierszach)
cisnienie_skurczowe	wartości odstające	$255, 38$
grupa_krwi	wartość spoza zbioru kategorii (0/A/B/AB)	"5", "Rh+"
choroba_przewlekla	wartość spoza zbioru kategorii (tak/nie)	"kobieta", "2"
liczba_wizyt	wartość niemożliwa dla licznika	$-2, -1$
wynik_testu_psychometrycznego	powtarzający się podejrzany kod	999 ($4\times$ w różnych wierszach)
tetno_spoczynkowe	wartości odstające	$2, 410$

Spodziewany wynik kontroli

Dla każdej z powyższych kolumn kontrola powinna wskazać dokładnie ten rodzaj problemu, jaki został w niej celowo umieszczony – np. dla wiek: ”wartości -2, 0, 150 i 199 są niezgodne z sensem wieku”; dla waga_kg: ”wartość -99 powtarza się kilkukrotnie i może być niestandardowym kodem braku danych”; dla choroba_przewlekla: ”wartość *kobieta* nie pasuje do pozostałych wartości tak/nie”. Dokładne sformułowania zależą od użytego modelu AI – powyższe to jedynie oczekiwany kierunek wskazań, nie dosłowna odpowiedź.

Wynik zostaje zapisany w Notatce arkusza danych (widocznej w drzewie projektu), gdzie można go dowolnie edytować lub uzupełnić własnymi uwagami przed zapisaniem.

DLA ZAINTERESOWANYCH

Procedura: dla każdej kolumny program oblicza niezależnie: liczbę obserwacji, liczbę braków danych, liczbę unikalnych wartości, a następnie – w zależności od charakteru kolumny – albo statystyki liczbowe (minimum, kwartyle, maksimum, wartości odstające wg reguły IQR), albo listę kategorii wraz z ich licznością. To zwięzłe podsumowanie (a nie surowe dane) trafia do modelu AI wraz z instrukcją, by wskazał wyłącznie te kolumny, w których faktycznie coś budzi wątpliwości, i nie zgadywał niczego poza tym, co wynika z podanego podsumowania.

Uwaga o wydajności: ponieważ analizowane jest samo podsumowanie, a nie pełna zawartość arkusza, rozmiar zapytania do modelu pozostaje niewielki nawet dla arkuszy z dużą liczbą wierszy – rośnie on wraz z liczbą **kolumn**, a nie liczbą obserwacji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

7 OBSŁUGA PROGRAMU

Praca w programie PQStat opiera się na projektach - każdy projekt to jeden plik. **Projekt** działa podobnie jak skoroszyt (zeszyt) i mieści w sobie 3 podstawowe elementy:

1. Arkusze danych (w tym arkusze map i macierzy) - w jednym projekcie można ich mieć maksymalnie 1000,
2. Arkusze wyników (raporty) - w jednym arkuszu danych może ich powstać maksymalnie 2000,
3. Zarządca projektu - okno, w którym zmienimy nazwy arkuszy danych i wyników, dodamy własne opisy i notatki oraz wyeksportujemy zawartość projektu.

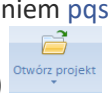
Jednocześnie można mieć otwartych maksymalnie 255 projektów. Pierwszy projekt wraz z pustym arkuszem danych otwiera się automatycznie po uruchomieniu programu, o ile taka opcja jest ustawiona w oknie Ustawienia programu.

Nowy projekt tworzymy przez:

- Plik→Nowy projekt (Ctrl+N) 

- Plik→Nowy arkusz (Ctrl+D) 

Istniejący projekt (plik z rozszerzeniem pqs lub pqx) otwieramy na jeden z kilku sposobów:

- Plik→Otwórz projekt (Ctrl+O) 

- Plik→Otwórz przykłady 

- otwiera przykładowe projekty dołączone do programu,
- przeciągając plik projektu na okno programu,
- klikając dwukrotnie plik projektu.

Projekt można zapisać przez:

- menu Plik→Zapisz (Ctrl+S) 

- Plik→Zapisz jako... 

- przycisk Zapisz w Zarządcy Projektu.

Zapisanie projektu zapisuje jednocześnie wszystkie jego składowe do pliku z rozszerzeniem pqs lub pqx.

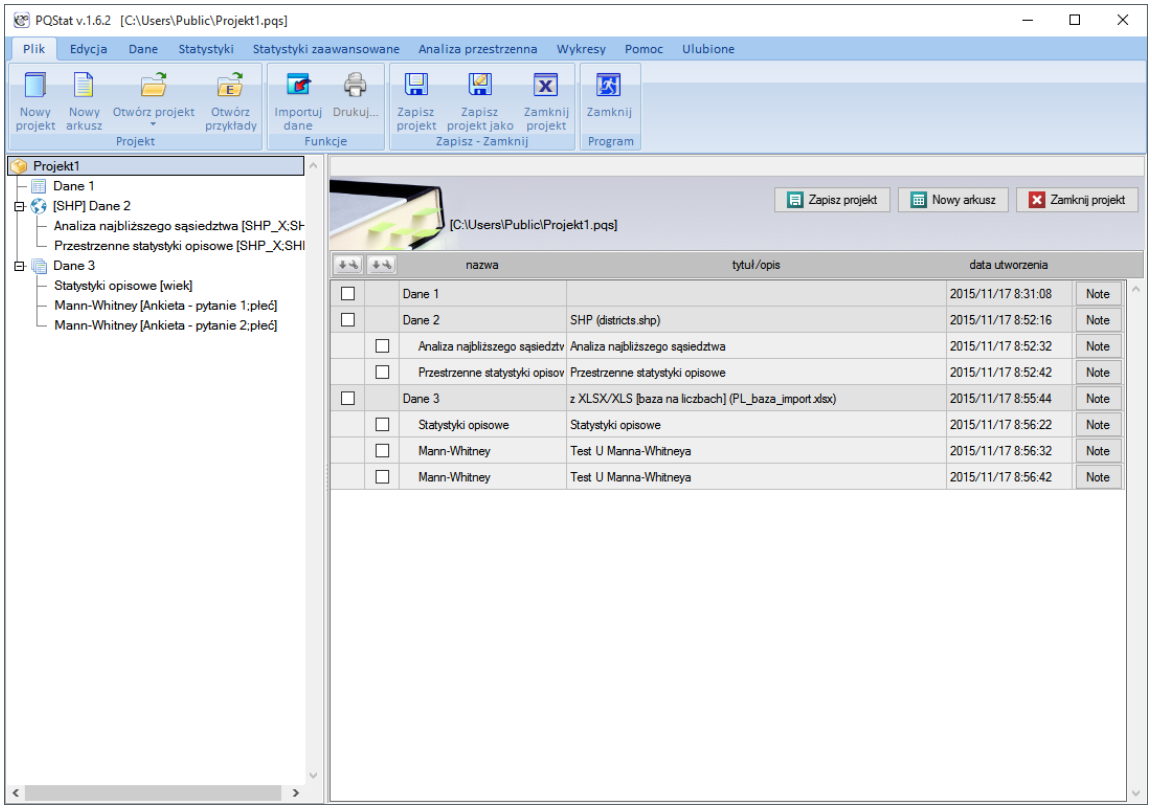
Projekt zamykamy przez:

- menu Plik→Zamknij projekt 

- przycisk Zamknij projekt w Zarządcy Projektu.

Do łatwej nawigacji w projekcie służy Zarządca Projektu, wyświetlany po wybraniu odpowiedniego projektu. W oknie tym można zapisać lub usunąć wybrany projekt, dodać bądź usunąć arkusz danych, usunąć arkusz wyników, a także dodawać notatki. Nazwa projektu jest jednocześnie nazwą jego pliku (pqs/pqx).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

7.1 PRACA Z ARKUSZAMI DANYCH

Najważniejszym elementem każdego projektu jest arkusz danych. Każdy otwarty projekt musi mieć przynajmniej jeden taki arkusz.

DOŁĄCZANIE, USUWANIE I EKSPORTOWANIE ARKUSZY DANYCH

Pierwszy pusty arkusz danych otwiera się automatycznie razem z nowym projektem.

Kolejne arkusze danych dołączamy do projektu przez:

- menu Plik→Nowy arkusz (Ctrl+D) 

- przycisk Nowy arkusz w Zarządcy Projektu.

Arkusz danych usuniemy przez:
- menu kontekstowe Usuń arkusz (Shift+Del) na nazwie arkusza danych w Drzewie nawigacji,
- przycisk  →Usuń w Zarządcy Projektu, gdy zaznaczony jest arkusz/arkusze. **Uwaga!** Jeśli do arkusza danych dołączone są raporty lub mapa, usunięcie arkusza usuwa jednocześnie wszystkie przypisane do niego raporty i mapę. Arkusze możemy opisać w Zarządcy Projektu, nadając im nazwę, tytuł lub dłuższą notatkę.

Arkusze danych utworzone w PQStat można eksportować do formatu csv (txt), dbf i xls. Eksportu dokonuje się w Zarządcy Projektu przez przycisk  →Eksportuj..., gdy zaznaczony jest arkusz/arkusze.

UMIESZCZANIE DANYCH W ARKUSZU

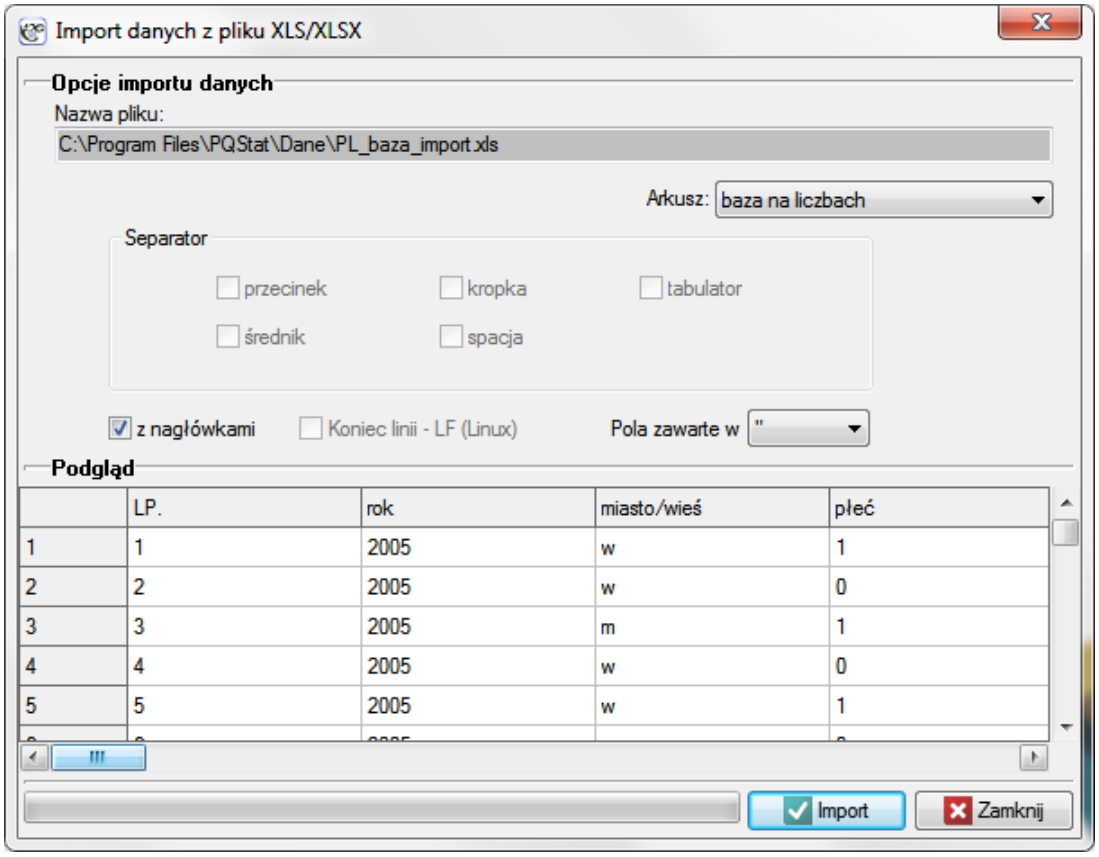
Nowo utworzony arkusz danych jest pusty. Dane można wpisać ręcznie, skopiować z innego arkusza danych albo zaimportować z pliku. Arkusz mieści maksymalnie 4 miliony wierszy i 1000 kolumn, a pojedyncza komórka może zawierać do 40 znaków.

Import danych

Dane można w prosty sposób importować z plików zapisanych w formacie:
- *.xls/*xlsx,
- *.txt/*.csv z wewnętrznym kodowaniem znaków UTF8, Windows-1250,
- *.shp (SHP/SHX/DBF ESRI Shapefile),
- *.dbf (dBase III, dBase IV, dBase VII),
- *.sav (SPSS),
- *.dbf (FoxPro).

Importu dokonujemy wybierając z menu Plik→Importuj dane ... 

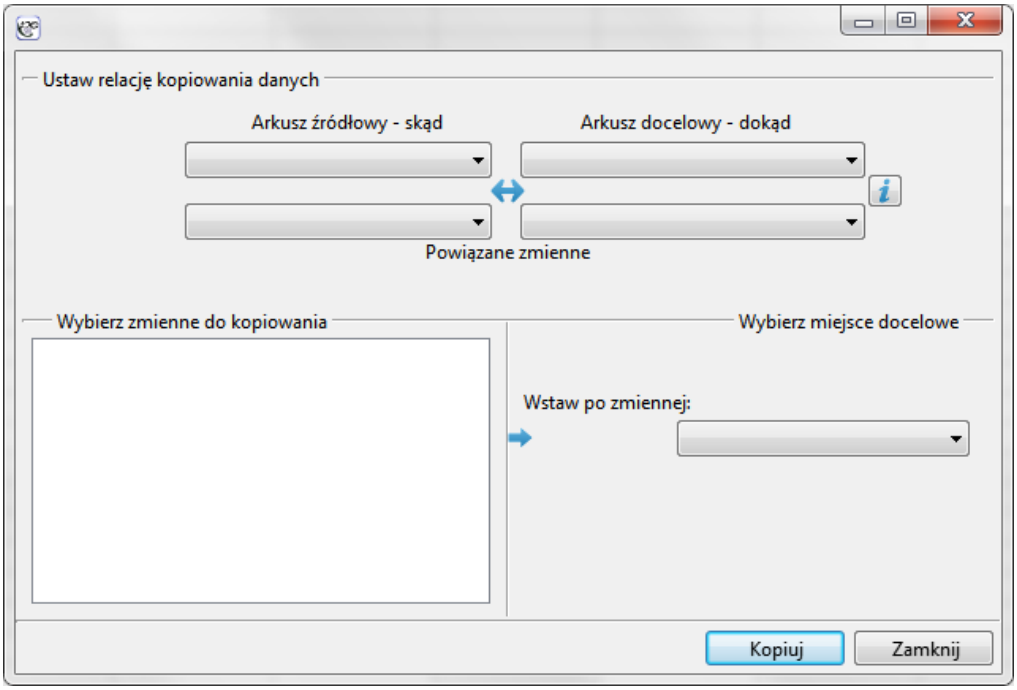
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Okno importu pozwala podejrzeć importowane dane jeszcze przed ich wczytaniem i sprawdzić, jak zostaną zinterpretowane przy wybranych opcjach. Warto zwrócić uwagę, czy polskie znaki wyświetlają się w podglądzie poprawnie - dzięki temu unikniemy błędnej interpretacji kodowania. Dla dużych plików podgląd pokazuje tylko początkową część danych. **Uwaga!** W programach Microsoft Office Excel 2000-2007 domyślnym kodowaniem znaków jest Windows-1250. Import danych z arkuszy Microsoft Excel obejmuje tylko wartości komórek - nie da się zaimportować formatowania ani formuł. **Kopiowanie danych poprzez relacje**

Do wybranego arkusza danych mogą zostać przekopiowane dane z innego arkusza na podstawie relacji. Ten rodzaj kopiowania danych wykonujemy wybierając z menu Dane→Kopiowanie z relacją... 

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

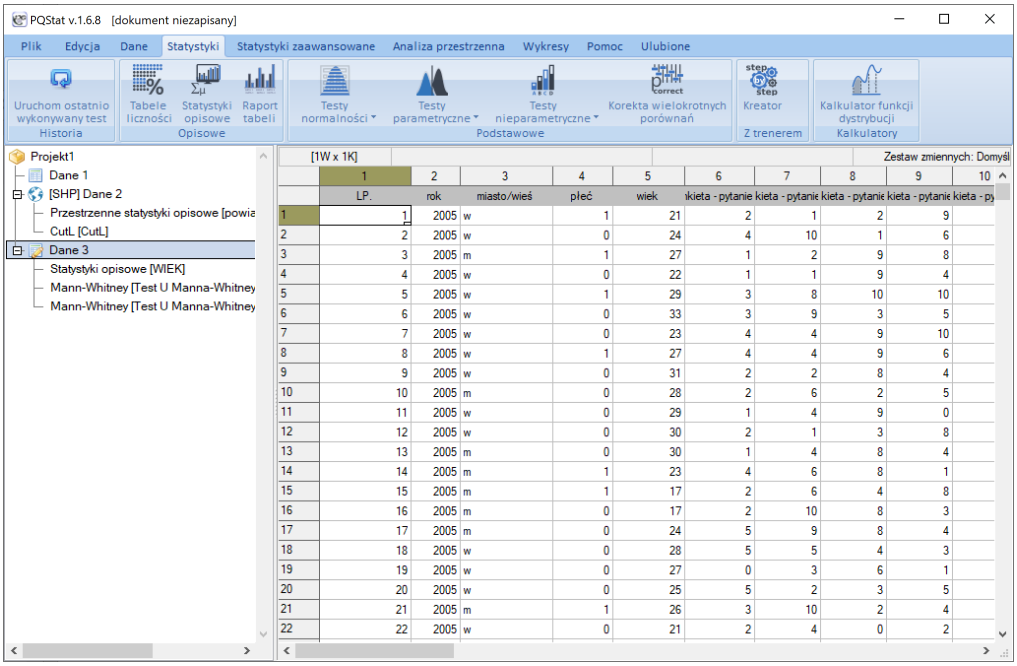


By zbudować relację, wskazujemy arkusz danych, z którego kopiujemy, oraz arkusz danych, do którego trafią skopiowane dane. Oba arkusze muszą mieć ten sam klucz, czyli zmienną, której wartości identyfikują każdy wiersz. Klucz w arkuszu źródłowym musi być unikalny. Powiązanie działa według relacji jeden-do-wielu - jeden wiersz z arkusza źródłowego może odpowiadać wielu wierszom w arkuszu docelowym. Klucze obu arkuszy wskazujemy jako Powiązane zmienne, a następnie wybieramy zmienne do skopiowania i kolumnę, po której mają zostać umieszczone.

OKNO ARKUSZA DANYCH

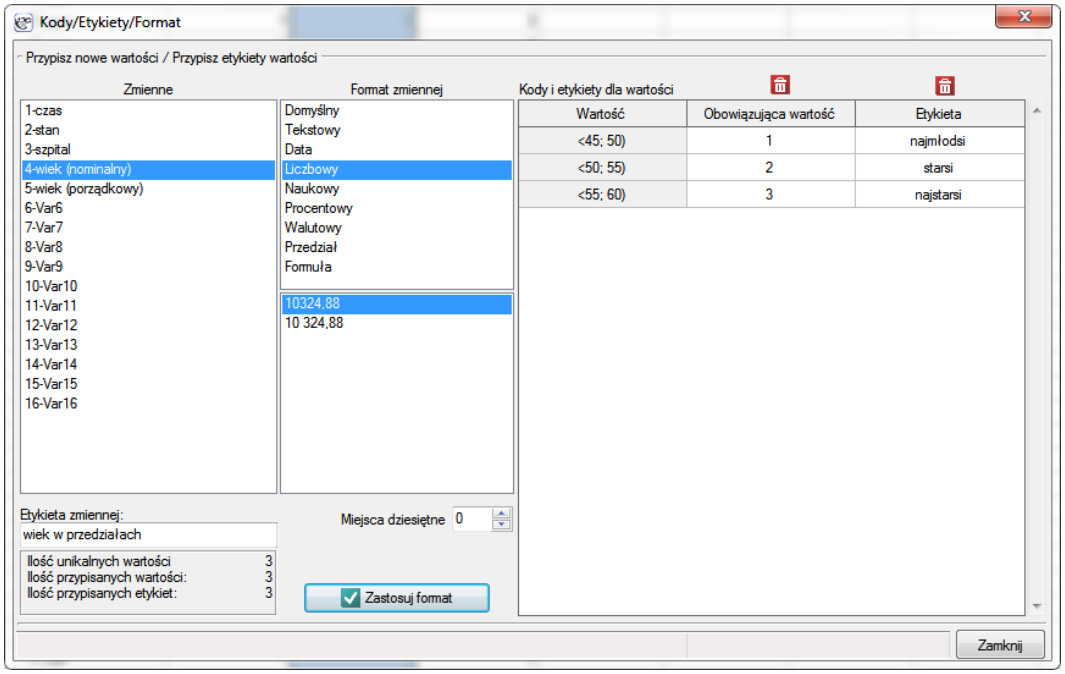
Wiersze i kolumny arkusza danych są ponumerowane kolejnymi liczbami naturalnymi. Każdej kolumnie, w miejscu oznaczonym szarym kolorem, można nadać własny nagłówek. Na górze arkusza znajduje się Pasek komunikatów - tu na bieżąco pojawiają się informacje dla użytkownika. Jego lewa część pokazuje wielkość zaznaczonego obszaru [liczba wierszy, liczba kolumn], środkowa - wartość z zaznaczonej komórki, a prawa jest przeznaczona m.in. na informacje o wykonywanej analizie statystycznej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WŁAŚCIWOŚCI ZMIENNEJ

Dla każdej kolumny arkusza można ustawić jej właściwości, takie jak kody, etykiety i format. Okno właściwości zmiennej otworzymy na dwa sposoby:
- wybierając Właściwości zmiennej → Kody/Etykiety/Format z menu kontekstowego na numerze znajdującym się powyżej nagłówka kolumny,
- klikając dwukrotnie na numerze znajdującym się powyżej nagłówka kolumny - o ile taka akcja podwójnego kliknięcia jest ustawiona w ustawieniach programu.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Kody i etykiety dla wartości - można je przypisać do każdej wartości występującej w danej kolumnie.

> **Obowiązująca wartość** - wypełniając kody, ustalamy, jakie wartości mają obowiązywać w danej kolumnie. Wartości pobierane do obliczeń zostaną w tle zmienione na odpowiednio przypisane kody (obowiązujące wartości).
> **Etykieta** - wartość wpisana w polu Etykieta jest wykorzystywana w raportach i na wykresach jako opis wyniku, ustalony przez użytkownika.

Etykieta zmiennej - jest przypisana do nagłówka danej kolumny. Zwykle jest to krótki opis zawartości zmiennej. Gdy jest ustawiona, zastępuje nagłówek kolumny (nazwę zmiennej) w raportach i na wykresach, dzięki czemu wyniki są czytelniejsze. Jej wykorzystanie jest opcjonalne i zależy od ustawień programu.

Format zmiennej
Każda komórka arkusza (również nagłówek kolumny) może zawierać najwyżej 40 znaków, w tym znaki narodowe. Wprowadzone wartości mogą być sformatowane jako:

- **domyślne** - format domyślny jest formatem, w którym program automatycznie rozpoznaje zawartość komórki w zakresie - dane liczbowe, dane tekstowe;
- **tekstowe** - w formacie tekstowym dane interpretowane są jako tekst (wyrównanie do lewej krawędzi komórki);
- **data** - w formacie daty dane liczbowe interpretowane są jako kolejne wartości daty, i tak wartość 1 oznacza 1899.12.31, wartość 2 oznacza 1900.01.01 itd. W zależności od wybranego formatu daty można również wprowadzać dane tekstowo w wybranym formacie, są to:

  2010.12.31
  31.12.2010
  12.31.2010
  2010/12/31
  31/12/2010
  12/31/2010
  2010-12-31
  31-12-2010
  12-31-2010
  Poniedziałek...
  Styczeń...

  W przypadku formatu Poniedziałek... wartość 1 oznacza poniedziałek, ..., 7 - niedziela, w przypadku formatu Styczeń..., wartość 1 oznacza styczeń, ..., 12 - grudzień.

- **czas** - w formacie czasu dane liczbowe interpretowanie są jako kolejne wartości czasu, część ułamkowa liczby oznacza liczbę milisekund od północy podzielonych przez całkowitą liczbę milisekund dnia (86400000), i tak wartość 0,000694444 oznacza 00:01:00, wartość 0,041666667 oznacza 01:00:00, wartość 0,999988426 oznacza 23:59:59. W zależności od wybranego formatu czasu można również wprowadzać dane tekstowo w wybranym formacie, są to:

  18:31:58
  18:31
  12/31/2010 18:31
  12/31/2010 18:31:58

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **liczbowe** - liczby rzeczywiste w tym formacie występują w postaci rozwinięcia dziesiętnego, przy czym znak oddzielający część całkowitą i ułamkową to przecinek lub kropka (w zależności od ustawień wybranych w oknie Ustawienia w polu Separator dziesiętny), możliwe jest ustawienie liczb miejsc dziesiętnych oraz separatora części tysięcznych;
- **naukowe** - czyli przy użyciu $M \cdot 10^E$, gdzie podstawa, to mantysa M, a wykładnik E jest liczbą całkowitą; tak jak w formacie liczbowym możliwe jest ustawienie liczby miejsc dziesiętnych;
- **procentowe** - zmieniające liczbę na procent poprzez pomnożenie przez 100 i wyświetlenie z symbolem %; tak jak w formacie liczbowym możliwe jest ustawienie liczby miejsc dziesiętnych;
- **walutowe** - wykorzystywane są dla wartości pieniężnych - pozwala to na dodanie symbolu waluty; tak jak w formacie liczbowym możliwe jest ustawienie liczby miejsc dziesiętnych;
- **przedział** - zapisany za pomocą górnej i dolnej granicy; tak jak w formacie liczbowym możliwe jest ustawienie liczby miejsc dziesiętnych;
- **formuła** - wartości wyliczone zgodnie z przypisaną do kolumny formułą; wartość jest automatycznie przeliczana gdy zmieniona zostanie którakolwiek z danych wejściowych.

Po otwarciu nowego arkusza dla każdej komórki standardowo ustawiony jest **format domyślny**.

Cały wiersz nagłówkowy ma na stałe ustawiony format tekstowy. Dla pozostałej części arkusza można ustawić dowolny z powyższych formatów. Formatowaniu podlega zawsze cała kolumna (poza jej nagłówkiem), nie pojedyncza komórka.

Szerokość kolumny można zmienić myszą - wystarczy przeciągnąć linię dzielącą kolumny, zwężając lub rozszerzając kolumnę znajdującą się po jej lewej stronie.

Każdej komórce arkusza można też nadać inny kolor tła (po uprzednim zaznaczeniu obszaru, który chcemy pokolorować). Służy do tego:

- menu Plik→Kolor wypełnienia … 

- polecenie Kolor komórki w menu kontekstowym komórki.

EDYCJA DANYCH

Zaznaczanie spójnego obszaru arkusza można wykonać myszą lub klawiaturą (Klawisze strzałek + Shift). W trakcie zaznaczania pasek komunikatów na bieżąco pokazuje jego rozmiar (liczbę wierszy i kolumn). Cały arkusz zaznaczymy jednym kliknięciem w lewy górny róg arkusza albo z menu Edycja→Zaznacz wszystko (Ctrl+A). Całe wiersze lub kolumny zaznaczamy, klikając ich nagłówki.

Kopiowanie lub przenoszenie komórek wykonujemy poleceniami kopiuj, wytnij i wklej, dostępnymi w kilku miejscach:
- w menu Edycja,
- w menu kontekstowym komórek,
- na pasku narzędziowym  ,
- w menu kontekstowym wierszy i kolumn,
- pod klawiszami skrótów: kopiuj (Ctrl+C), wytnij (Ctrl+X) i wklej (Ctrl+V).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Usuwanie danych z komórek wykonamy przez menu Edycja→Usuń (Del)

Cofanie ostatnio wykonanej operacji wykonamy przez menu Edycja→Cofnij (Ctrl+Z). Domyślnie program pamięta 10 ostatnich operacji, każdą dla maksymalnie 5000 komórek. Te wartości można zmienić w oknie Ustawienia - warto jednak pamiętać, że ich zwiększenie oznacza większe zużycie pamięci komputera.

Wstawianie i usuwanie wierszy i kolumn

Puste wiersze lub kolumny można wstawić powyżej lub po lewej stronie istniejącego wiersza/kolumny - pozostałe komórki przesuną się wtedy w dół lub w prawo. Aby wstawić wiersz (wiersze), zaznaczamy wiersz, przed którym chcemy wstawić nowy, a następnie wybieramy z menu kontekstowego na jego numerze Wstaw wiersz. Kolumny wstawia się analogicznie.

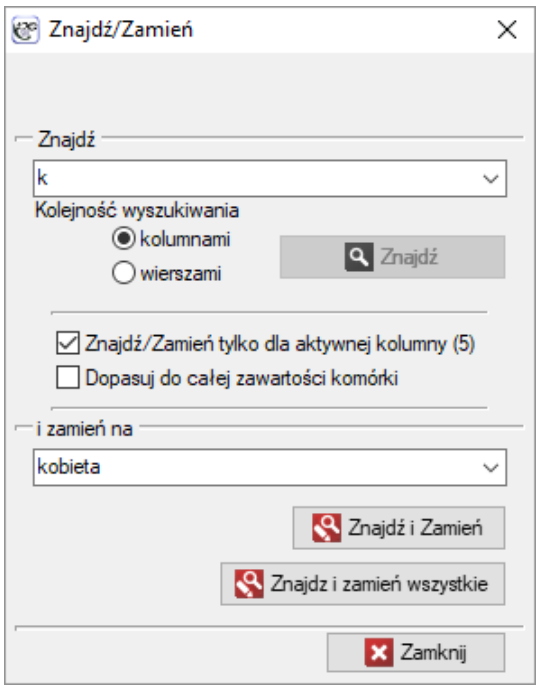
Wiersze i kolumny usuwamy w podobny sposób - zaznaczamy je i wybieramy w menu kontekstowym na ich numerze Usuń wiersz/Usuń kolumnę.

Znajdowanie i zamiana wartości komórki

Do wyszukania lub zastąpienia zawartości komórki inną wartością służy okno Znajdź/Zamień, wywoływane przez menu Edycja→Znajdź/Zamień (Ctrl+F) 

.

Górna część okna służy wyszukiwaniu, dolna - zamianie wartości komórki.



Aby wyszukać dane, w górnej części okna wpisujemy szukany ciąg znaków, ustalamy kierunek wyszukiwania i klikamy przycisk Znajdź.

Aby znaleźć i zastąpić zawartość komórki inną wartością, wypełniamy obie części okna: górną - tak jak przy wyszukiwaniu, dolną - wpisując dane, które mają zastąpić znalezione. Następnie wybieramy Znajdź

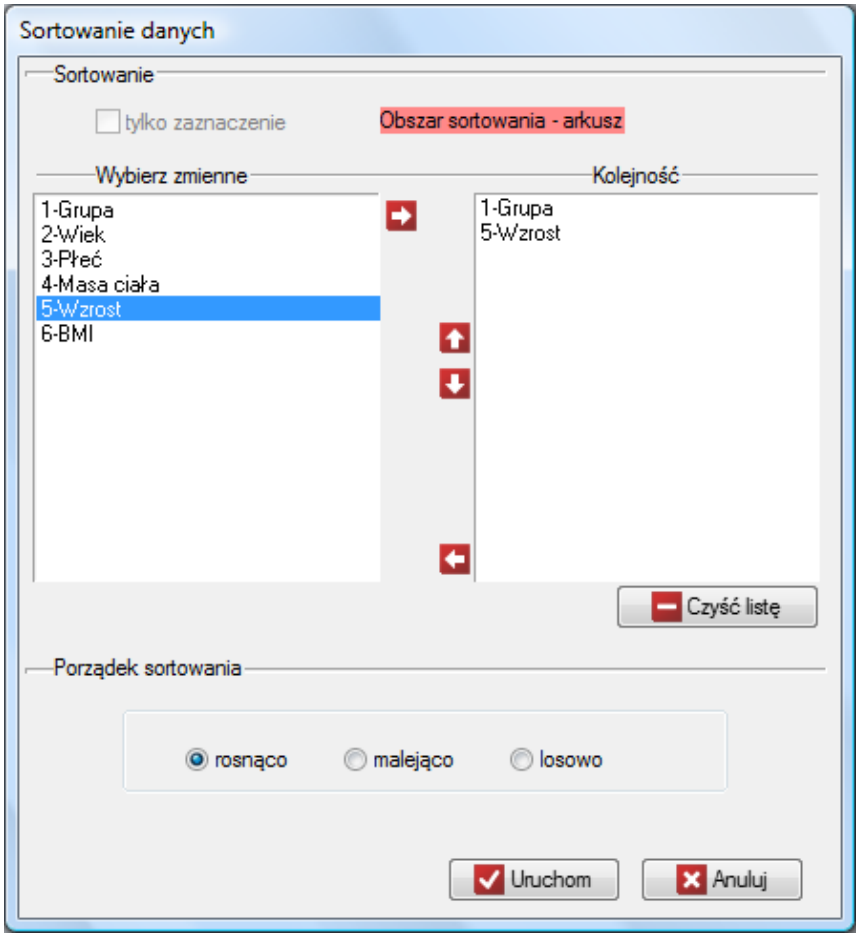
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

i Zamień albo Znajdź i zamień wszystkie, gdy chcemy zastąpić wszystkie wystąpienia naraz. Zarówno wyszukiwanie, jak i zamiana od razu pokazują efekt na arkuszu.

SORTOWANIE DANYCH



Opcje sortowania znajdziemy w menu Dane→Sortuj... albo pod opcją Sortuj... z menu kontekstowego na numerze znajdującym się powyżej nagłówka kolumny. Domyślnie sortowany jest cały arkusz danych - jeśli jednak przed uruchomieniem sortowania zaznaczymy fragment danych, w oknie sortowania będzie można ograniczyć sortowanie tylko do tego zaznaczenia.



W oknie sortowania strzałkami przenosimy z pola Wybierz zmienne do pola Kolejność te zmienne, według których chcemy sortować dane, wybieramy Porządek sortowania i potwierdzamy przyciskiem Uruchom.
Sortować można według co najwyżej 3 kolumn (zmiennych). Przy sortowaniu według kilku zmiennych decyduje kolejność, w jakiej zostały ułożone w polu Kolejność.

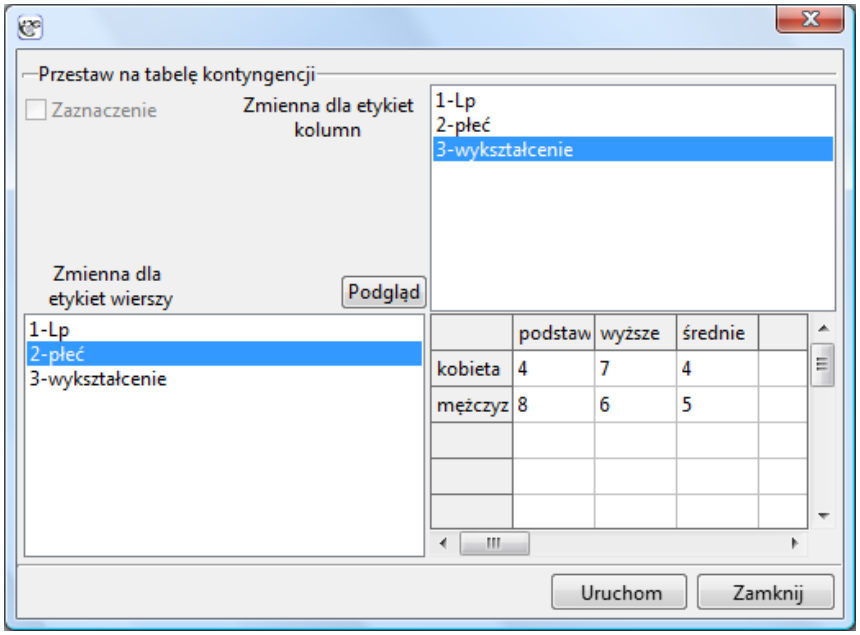
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

ZAMIANA DANYCH SUROWYCH NA TABELĘ KONTYNGENCJI



Operację zamiany danych surowych na tabelę kontyngencji wykonamy z menu Dane→Przestaw na tabelę... Domyślnie brany jest pod uwagę cały arkusz danych - jeśli jednak wcześniej zaznaczymy fragment danych, w oknie przekształcania będzie można ograniczyć się tylko do tego zaznaczenia.

Tabelę kontyngencji projektujemy, wybierając zmienne, które utworzą etykiety wierszy i kolumn. Gdy podgląd tabeli wygląda tak, jak oczekiwaliśmy, potwierdzamy wybór przyciskiem Uruchom. Wynik trafi do nowego arkusza.

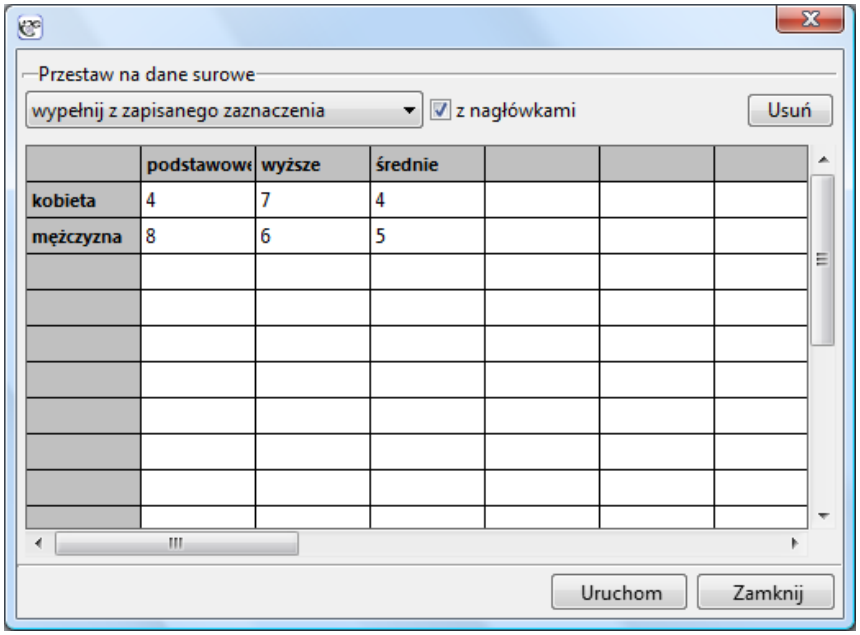


ZAMIANA TABELI KONTYNGENCJI NA DANE SUROWE



Odwrotną operację, czyli zamianę tabeli kontyngencji na dane surowe, wykonamy z menu Dane→Przestaw na dane surowe... W oknie przekształcania wpisujemy liczności oraz nagłówki wierszy i kolumn, a wybór potwierdzamy przyciskiem Uruchom. Wynik trafi do nowego arkusza.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Jeśli przekształcana tabela znajduje się już w arkuszu danych, przed przystąpieniem do zamiany warto ją zaznaczyć (z nagłówkami lub bez) - wtedy w oknie przekształcania pojawi się automatycznie. Można też skorzystać z innych tabel oznaczonych jako zapisane zaznaczenie.

PRZEKSZTAŁCANIE DANYCH MIĘDZY POSTACIĄ SZEROKĄ A DŁUGĄ

Te same dane pomiarowe można ułożyć w arkuszu na dwa sposoby:

- **postać szeroka** – jeden wiersz to jeden obiekt (np. pacjent), a kolejne pomiary tego obiektu (np. pomiar przed leczeniem, po miesiącu, po roku) zapisane są w kolejnych kolumnach; w takim układzie pracują m.in. ANOVA dla grup zależnych i testy dla grup zależnych,
- **postać długa** – jeden wiersz to jeden pomiar; obiekt jest opisany przez kolumnę identyfikatora (powtórzoną tyle razy, ile ma pomiarów), a obok znajduje się kolumna z nazwą lub numerem pomiaru oraz jedna wspólna kolumna z wartością; taki układ wymagają modele mieszane i większość analiz, w których pomiar jest osobną obserwacją.

Przykład tych samych danych w obu postaciach:

postać szeroka			
ID	płeć	przed	po
1	K	120	112
2	M	135	130

postać długa			
ID	płeć	zmienna	wartość
1	K	przed	120
1	K	po	112
2	M	przed	135
2	M	po	130

Przekształcenie w obie strony wykonują polecenia z menu Dane. Wynik zawsze trafia do **nowego arkusza** projektu, a arkusz źródłowy pozostaje bez zmian. Wiersze dezaktywowane oraz kolumny ukryte są pomijane, tak jak w analizach statystycznych.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Przykład danych z pliku cisnienie.pqs, gdzie badamy dwa rodzaje leków (Ai B) oraz ciśnienie na kolejnych wizytach: Postać szeroka:

CS przed	CS w trakcie	CS po	leczenie
166	160	160	Lek A
144	140	120	Lek A
188	170	150	Lek A
170	161	160	Lek A
156	150	130	Lek A
160	150	120	Lek A
148	130	126	Lek A
150	150	150	Lek A
178	160	156	Lek A
180	160	140	Lek A
190	140	138	Lek A
180	170	168	Lek A
176	156	140	Lek A
140	140	120	Lek A
160	150	150	Lek A

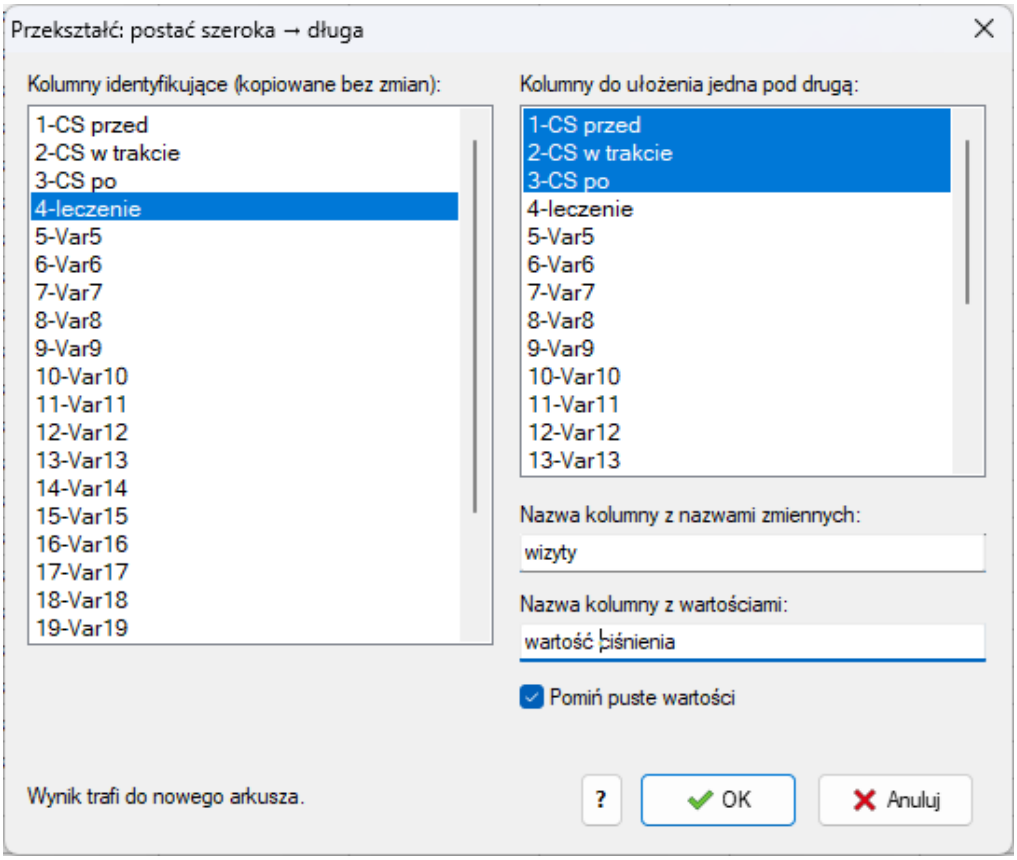
Postać długa:

leczenie	wizyty	wartość ciśnienia
Lek A	CS przed	166
Lek A	CS w trakcie	160
Lek A	CS po	160
Lek A	CS przed	144
Lek A	CS w trakcie	140
Lek A	CS po	120
Lek A	CS przed	188
Lek A	CS w trakcie	170
Lek A	CS po	150
Lek A	CS przed	170
Lek A	CS w trakcie	161
Lek A	CS po	160
Lek A	CS przed	156
Lek A	CS w trakcie	150
Lek A	CS po	130
Lek A	CS przed	160
Lek A	CS w trakcie	150

Postać szeroka → długa

Polecenie Dane→Przekształć: szeroka → długa... układa wybrane kolumny jedna pod drugą (przykład plik cisnienie.pqs).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



W oknie przekształcania wskazujemy:

- **Kolumny identyfikujące** – kolumny opisujące obiekt (np. identyfikator, płeć, grupa), które mają zostać skopiowane bez zmian i powtórzone w każdym wierszu powstałym z tego samego obiektu; ich wybór nie jest obowiązkowy,
- **Kolumny do ułożenia jedna pod drugą** – kolumny z pomiarami; trzeba wybrać co najmniej jedną, a ta sama kolumna nie może być jednocześnie identyfikatorem i kolumną wartości,
- **Nazwę kolumny z nazwami zmiennych** (domyślnie `zmienna`) – w tej kolumnie znajdzie się nazwa kolumny źródłowej, z której pochodzi wartość,
- **Nazwę kolumny z wartościami** (domyślnie `wartość`),
- **Pomiń puste wartości** – gdy opcja jest zaznaczona (ustawienie domyślne), pomiary bez wartości nie tworzą wierszy; po jej odznaczeniu każdy obiekt otrzyma dokładnie tyle wierszy, ile wybrano kolumn, a braki pozostaną pustymi komórkami.

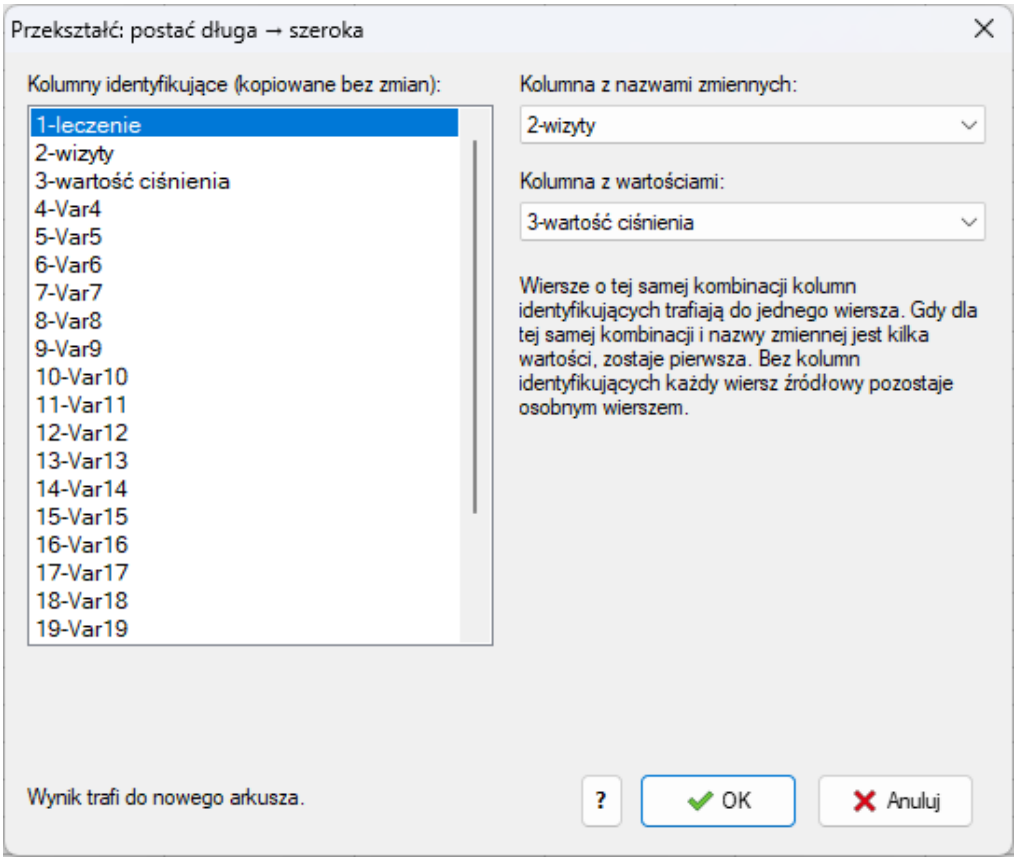
Wynik potwierdzamy przyciskiem OK. Każdy wiersz źródłowy daje tyle wierszy, ile wybrano kolumn do ułożenia, a wiersze są uporządkowane obiekt po obiekcie, w kolejności wybranych kolumn. Nazwy kolumn w nowym arkuszu są automatycznie rozróżniane, jeśli by się powtarzały.

Uwaga! Jeżeli kolumna z nazwami zmiennych ma posłużyć jako czas lub numer pomiaru w modelu mieszanym, warto nadać kolumnom źródłowym nazwy liczbowe (np. `0`, `1`, `12`) albo po przekształceniu przekodować je za pomocą formuł.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Postać długa → szeroka

Polecenie Dane→Przekształć: długa → szeroka... rozkłada jedną kolumnę wartości na kilka kolumn, po jednej dla każdej nazwy zmiennej.



W oknie przekształcania wskazujemy:

- **Kolumny identyfikujące** – kolumny, których wspólna kombinacja wartości wyznacza jeden obiekt (np. identyfikator pacjenta albo identyfikator i płeć); wszystkie wiersze o tej samej kombinacji zostaną połączone w jeden wiersz postaci szerokiej; jeśli nie wybierzemy żadnej kolumny, każdy wiersz źródłowy pozostanie osobnym wierszem,
- **Kolumnę z nazwami zmiennych** – każda unikalna wartość tej kolumny stanie się osobną kolumną nowego arkusza (w kolejności pierwszego wystąpienia w danych); wiersze z pustą nazwą są pomijane,
- **Kolumnę z wartościami** – jej zawartość wypełni nowe kolumny.

Kolumna z nazwami zmiennych i kolumna z wartościami muszą być różne i nie mogą być jednocześnie kolumnami identyfikującymi. Wynik potwierdzamy przyciskiem OK. Nowy arkusz zawiera kolumny identyfikujące (wartości z pierwszego wystąpienia obiektu) oraz kolumny utworzone z nazw zmiennych; kombinacje, dla których w danych nie było pomiaru, pozostają pustymi komórkami.

Uwaga! Jeżeli dla tej samej kombinacji identyfikatorów i tej samej nazwy zmiennej w danych występuje kilka wartości (np. kilka pomiarów tej samej zmiennej u jednego obiektu albo zbyt mało kolumn

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

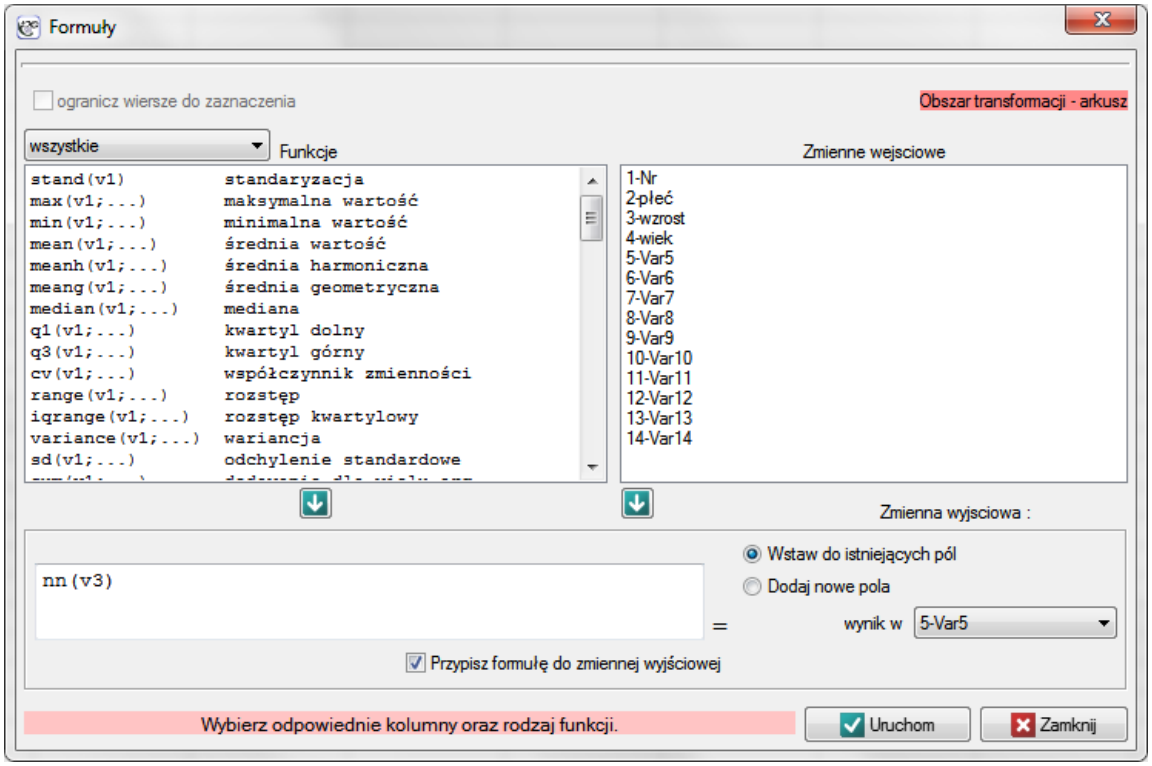
identyfikujących), żadna z nich nie ginie: pierwsza trafia do kolumny o nazwie zmiennej, a kolejne do dodatkowych kolumn `nazwa_2`, `nazwa_3`, …umieszczonych obok niej. Program informuje o liczbie takich powtórzeń. Jeśli kolejne wystąpienia mają rozróżniać się inaczej (np. numerem wizyty), warto dodać odpowiednią kolumnę do identyfikatorów.

FORMUŁY



Formuła to sposób na przeliczenie danych, dzięki któremu otrzymujemy nowe wartości dla zmiennych.

Okno definiowania formuł otwieramy przez Dane→Formuły...



Formuła przypisana do zmiennej jako jej format jest zapamiętywana razem z arkuszem danych. Jej wynik przelicza się automatycznie za każdym razem, gdy zmieni się którakolwiek z danych wejściowych. Formułę przypiszemy w oknie Formuły... albo przez ustawienie Formatu kolumny (Ctrl+F10).

Budowanie formuł

Formuły wprowadzamy w polu edycji:

- zmienne, do których odwołuje się formuła, wskazujemy ich numerami, np. `v1+v2`,
- wartości tekstowe wpisujemy w apostrofach, np. 'dom',
- funkcję wstawiamy dwukrotnym kliknięciem na jej nazwie na liście - wówczas nazwa pojawi się w polu edycji - albo wpisujemy ją samodzielnie (wielkość liter nie ma znaczenia). Argumenty funkcji podajemy w nawiasie, zgodnie ze składnią opisaną przy danej funkcji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wyniki formuł

Wynik formuły pojawi się w wybranej kolumnie.

Jeśli wśród argumentów funkcji znajdzie się wartość, której funkcja nie potrafi zinterpretować, program zapyta, czy pominąć takie dane. Odpowiedź "tak" przelicza formułę z pominięciem niezinterpretowanych danych, odpowiedź "nie" zwraca błąd (NA). Przykładowo dla wartości w kolumnach v1, v2 i v3 odpowiednio: 1, 2, 'ada', funkcja sumowania `sum(v1;v2;v3)` zwróci wynik 3, gdy pominiemy niezinterpretowaną wartość 'ada', albo NA, gdy jej nie pominiemy.

Wartość pusta (brak danych) pojawi się tylko wtedy, gdy wszystkie argumenty formuły są puste.

Liczbę wierszy branych pod uwagę można ograniczyć: wystarczy zaznaczyć w arkuszu odpowiedni zakres wierszy i w oknie formuły wybrać opcję ogranicz wiersze do zaznaczenia.

Operatory

$+$ dodawanie,
$-$ odejmowanie,
$*$ mnożenie,
$/$ dzielenie,
% dzielenie modulo (w wyniku reszta z dzielenia),
$>$ większy,
$<$ mniejszy,
$=$ równy.

Funkcje matematyczne

Funkcje matematyczne wymagają argumentów liczbowych.

ln(v1) - zwraca logarytm naturalny podanej liczby,
log10(v1) - zwraca logarytm przy podstawie 10 dla podanej liczby,
logn(v1) - zwraca logarytm przy podstawie n dla podanej liczby,
sqr(v1) - zwraca kwadrat podanej liczby,
sqrt(v1) - zwraca pierwiastek kwadratowy podanej liczby,
fact(v1) - zwraca silnię podanej liczby,
degrad(v1) - zwraca miarę kąta w radianach (argumentem funkcji są stopnie),
raddeg(v1) - zwraca miarę kąta w stopniach (argumentem funkcji są radiany),
sin(v1) - zwraca sinus podanego kąta, (argumentem funkcji są radiany),
cos(v1) - zwraca cosinus podanego kąta, (argumentem funkcji są radiany),
tan(v1) - zwraca tangens podanego kąta, (argumentem funkcji są radiany),
ctng(v1) - zwraca cotangens podanego kąta, (argumentem funkcji są radiany),
arcsin(v1) - zwraca arcus sinus podanego kąta, (argumentem funkcji są radiany),
arctan(v1) - zwraca arcus tangens podanego kąta, (argumentem funkcji są radiany),
exp(v1) - zwraca wartość liczby e podniesionej do potęgi określonej przez podaną wartość,
frac(v1) - zwraca część ułamkową podanej liczby,
int(v1) - zwraca część całkowitą podanej liczby,
abs(v1) - zwraca wartość bezwzględną podanej liczby,
odd(v1) - zwraca 1, gdy podana liczba jest parzysta, 0 w przeciwnym przypadku,
sum(v1;...) - zwraca wynik dodawania podanych liczb,
multip(v1;...) - zwraca wynik mnożenia podanych liczb,
power(v1;n) - zwraca wynik podniesienia liczby do potęgi n,
norme(v1;...) - zwraca normę euklidesową wektora,
round(v1;n) - zwraca liczbę zaokrągloną do n miejsc po przecinku.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Funkcje statystyczne

Funkcje statystyczne wymagają argumentów liczbowych.

stand(v1) - zwraca wystandaryzowaną wartość podanej zmiennej,
max(v1,...) - zwraca wartość największą,
min(v1,...) - zwraca wartość najmniejszą,
mean(v1,...) - zwraca wartość średniej arytmetycznej,
meanh(v1,...) - zwraca wartość średniej harmonicznej,
meang(v1,...) - zwraca wartość średniej geometrycznej,
median(v1,...) - zwraca wartość mediany,
q1(v1,...) - zwraca wartość kwartyla dolnego,
q3(v1,...) - zwraca wartość kwartyla górnego,
cv(v1,...) - zwraca wartość współczynnika zmienności,
range(v1,...) - zwraca wartość rozstępu,
iqrange(v1,...) - zwraca wartość rozstępu kwartylowego,
variance(v1,...) - zwraca wartość wariancji,
sd(v1,...) - zwraca wartość odchylenia standardowego.

Funkcje tekstowe

Funkcje tekstowe działają na dowolnym ciągu znaków.

upperc(v1) - konwertuje znaki z łańcucha na wielkie,
lowerc(v1) - konwertuje znaki z łańcucha na małe,
clean(v1) - usuwa znaki, które nie mogą być drukowane,
trim(v1) - usuwa przednie i końcowe spacje,
length(v1) - zwraca długość łańcucha znaków,
search('abc';v1) - zwraca pozycję początku szukanego tekstu,
concat(v1;...) - łączy teksty,
compare(v1;...) - porównuje teksty,
copy(v1;i;n) - zwraca część tekstu począwszy od i-tego znaku, gdzie n - to liczba zwracanych znaków,
count(v1;...) - zwraca liczbę komórek, które nie są puste,
counte(v1;...) - zwraca liczbę pustych komórek,
countn(v1;...) - zwraca liczbę komórek, które zawierają liczby.

Funkcje daty i czasu

Funkcje daty i czasu powinny być wykonywane na danych sformatowanych jako data lub jako czas (patrz rozdział 7.1.4). Jeśli tak nie jest, program stara się automatycznie rozpoznać format, a gdy nie jest to możliwe zwraca wartość NA.

year(v1;) - zwraca rok odpowiadający dacie,
month(v1;) - zwraca miesiąc odpowiadający dacie,
day(v1;) - zwraca dzień odpowiadający dacie,
hour(v1;) - zwraca godzinę odopowiadającą podanemu czasowi,
minute(v1;) - zwraca minutę odopowiadającą podanemu czasowi,
second(v1;) - zwraca sekundę odopowiadającą podanemu czasowi,
yeardiff(v1;v2) - zwraca liczbę lat różniącą dwie daty,
monthdiff(v1;v2) - zwraca liczbę miesięcy różniącą dwie daty,
weekdiff(v1;v2) - zwraca liczbę tygodni różniącą dwie daty,
daydiff(v1;v2) - zwraca liczbę dni różniącą dwie daty,
hourdiff(v1;v2) - zwraca liczbę godzin różniącą dwa czasy,
minutediff(v1;v2) - zwraca liczbę minut różniącą dwa czasy,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

seconddiff(v1;v2) - zwraca liczbę sekund różniącą dwa czasy,
compdate(v1;v2) - porównuje daty i zwraca liczbę 1 gdy v1>v2; 0 gdy v1=v2, -1 gdy v1<v2.

Funkcje logiczne

if(pytanie;'tak - odpowiedź';'nie - odpowiedź') - pytanie formułuje się w formie twierdzenia, które może być prawdziwe lub fałszywe; funkcja zwraca jedną wartość jeśli twierdzenie jest prawdziwe a drugą jeśli jest fałszywe,
and - operator koniunkcji - zwraca prawdę (1) gdy wszystkie warunki, które łączy są prawdziwe, fałsz (0) w przeciwnym wypadku,
or - operator alternatywy zwraca prawdę (1) gdy przynajmniej jeden z łączonych przez nią warunków jest prawdziwy, fałsz (0) w przeciwnym wypadku,
xor - operator alternatywy rozłącznej - zwraca prawdę (1) gdy jeden z łączonych przez nią warunków jest prawdziwy, fałsz (0) w przeciwnym wypadku,
not - operator negacji używany w wyrażeniu warunkowym if,
empty(v1) - zwraca prawdę (1) gdy występują puste komórki, fałsz (0) w przeciwnym przypadku,
text(v1) - zwraca prawdę (1) gdy występuje tekst, fałsz (0) w przeciwnym przypadku,
number(v1) - zwraca prawdę (1) gdy występuje liczba, fałsz (0) w przeciwnym przypadku.

GENEROWANIE DANYCH

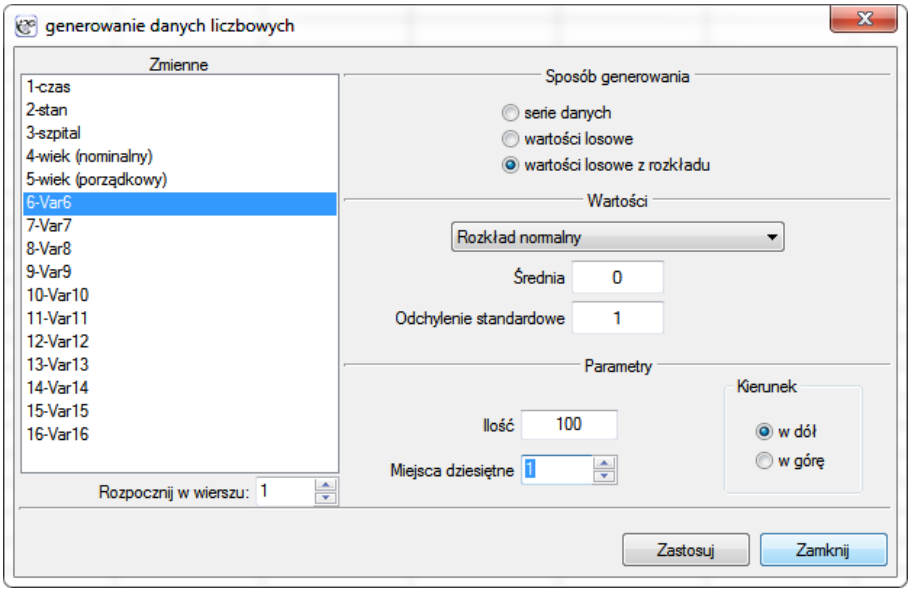


Dane w arkuszu można wygenerować na dwa sposoby:

1. Pierwszy sposób to proste przeciągnięcie zawartości zaznaczonych komórek na komórki sąsiednie za pomocą myszy. Pozwala on powielić te same wartości (tekstowe lub liczbowe) w sąsiednich kolumnach bądź wierszach.
   Zaznaczamy komórkę z odpowiednim wpisem, chwytamy wskaźnikiem myszy (w kształcie znaku +) jej prawy dolny narożnik i przeciągamy przez komórki, które chcemy wypełnić - w górę, w dół, w lewo lub w prawo. Podobnie można przeciągnąć kilka różnych wartości umieszczonych w jednej kolumnie (w lewo lub w prawo) albo w jednym wierszu (w górę lub w dół).

2. Drugi sposób pozwala wygenerować w kolumnie dane liczbowe jako serię, wartości losowe albo wartości losowe z wybranego rozkładu.
   Zaznaczamy komórkę, od której ma się zacząć wypełnianie arkusza, i otwieramy okno generowania danych liczbowych przez menu Dane→Generuj...

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Na początek wybieramy zmienną, w której mają się znaleźć wygenerowane dane.
W środkowej części okna, zależnie od wybranego sposobu generowania, ustawiamy:

- dla serii danych:
  - Wartość początkowa - pierwsza wartość, jaka ma zostać wygenerowana,
  - Krok - o ile mają się różnić kolejne wartości.
- dla wartości losowych:
  - Przedział (od) - początek przedziału, z którego losujemy wartości,
  - Przedział (do) - koniec tego przedziału.
- dla wartości losowych z rozkładu - wybieramy rodzaj rozkładu (normalny albo Chi-kwadrat) i wpisujemy jego parametry.

Liczba wygenerowanych wartości zależy od tego, co wpiszemy w polu ilość, a ich precyzja - od ustawienia pola miejsca dziesiętne. Dane zostaną wpisane od aktywnej komórki w dół lub w górę, zależnie od wybranej opcji. Wybór potwierdzamy przyciskiem Zastosuj.

BRAKI DANYCH

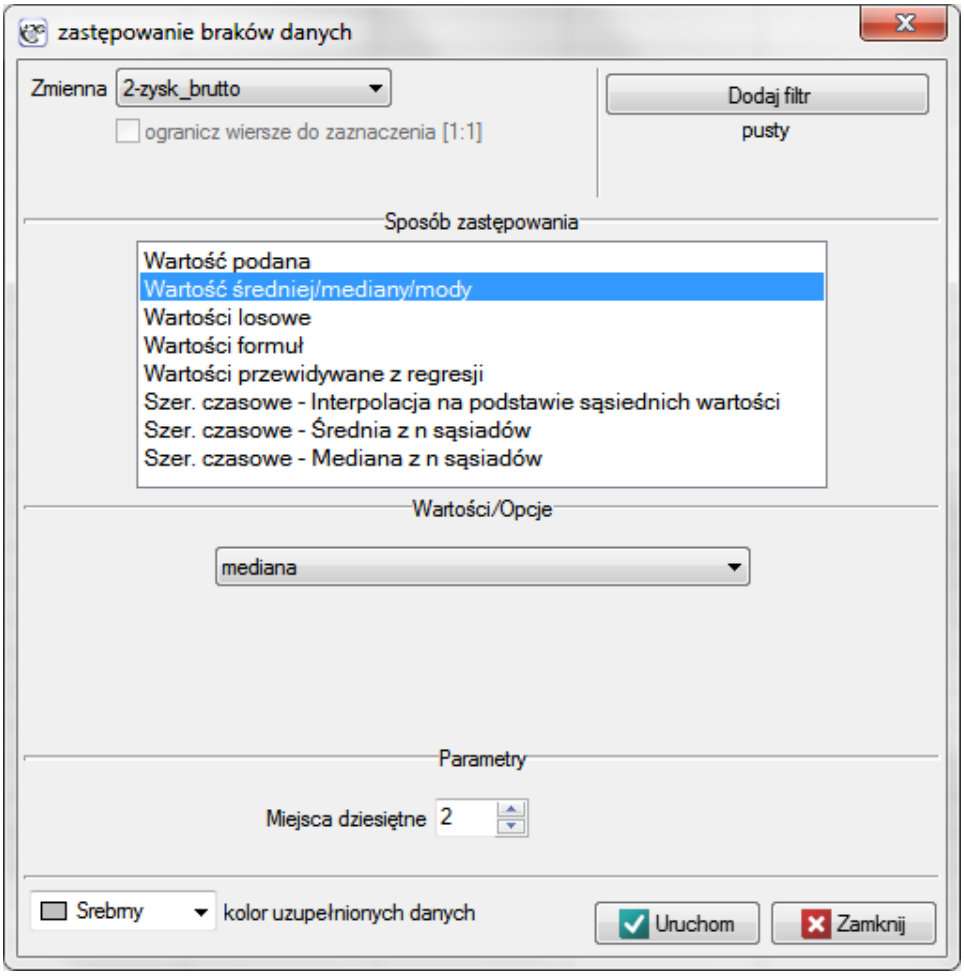


W badaniach bardzo często napotykamy braki danych - to naturalne, zwłaszcza w danych ankietowych. Czasem braki danych same w sobie niosą wartościową informację: liczba braków w odpowiedzi na pytanie o sympatie polityczne mówi coś o liczbie osób niezdecydowanych (albo niechcących przyznać się do swoich sympatii). Niewielka liczba braków nie stanowi problemu w analizach statystycznych, ale ich duża liczba może podważyć wiarygodność badania - dlatego warto już na etapie zbierania danych dbać, by było ich jak najmniej. Najlepiej oczywiście dotrzeć do rzeczywistej wartości, która powinna się znaleźć w miejscu braku, choć nie zawsze jest to możliwe.

Sposób szacowania brakujących danych zależy przede wszystkim od charakteru danych. Program oferuje kilka metod imputacji braków dla poszczególnych zmiennych.

Okno z ustawieniami opcji zastępowania braków danych otwieramy przez menu Dane→Braki danych...

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



1. Wypełnianie jedną wartością

   Wybranie jednej z poniższych opcji spowoduje zastąpienie w wybranej kolumnie wszystkich występujących tam braków tą samą wartością:

   - podaną przez użytkownika,
   - średnią arytmetyczną wyliczoną z danych,
   - średnią geometryczną wyliczoną z danych,
   - średnią harmoniczną wyliczoną z danych,
   - medianą,
   - modą (o ile nie jest wielokrotna).

2. Wypełnianie wieloma wartościami

   Wybranie jednej z poniższych opcji spowoduje zastąpienie braków w wybranej kolumnie wieloma (najczęściej różnymi) wartościami. Wartości te mogą być przewidywane na podstawie kolumny, dla której następuje wypełnienie braków danych ale mogą być również przewidywane na podstawie wartości innych kolumn (zmiennych). Zastąpić brak danych możemy wartościami:

   - losowymi z danych;
   - losowymi z rozkładu normalnego - rozkład normalny definiowany jest na podstawie średniej i odchylenia standardowego z występujących danych;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- losowymi z przedziału podanego przez użytkownika;
- wyliczanymi z funkcji użytkownika - opcja ta pozwala na wykorzystanie danych z innych zmiennych by móc przewidzieć brakującą wartość w wybranej kolumnie;
- wyliczanymi z modelu regresji - opcja ta pozwala na przewidywanie wartości braków danych na podstawie modelu regresji wielorakiej (zasada działania regresji wielorakiej opisana została w dziale 25.1 Liniowa regresja wieloraka);
- interpolacja na podstawie sąsiednich wartości - dotyczy szeregów czasowych - użytkownik musi więc wskazać zmienną czas mówiącą o uporządkowaniu danych; interpolacja polega na wyznaczeniu wartości dla braków danych w ten sposób by graficznie znalazły się na linii prostej łączącej wartości dla danych sąsiadujących z brakami;
- średnia z n sąsiadów - dotyczy szeregów czasowych - użytkownik musi więc wskazać zmienną czas mówiącą o uporządkowaniu danych; interpolacja polega na wyznaczeniu średniej z wartości dla n sąsiadów poprzedzających i n sąsiadów następujących bezpośrednio po brakach danych;
- mediana z n sąsiadów - dotyczy szeregów czasowych - użytkownik musi więc wskazać zmienną czas mówiącą o uporządkowaniu danych; interpolacja polega na wyznaczeniu mediany z wartości dla n sąsiadów poprzedzających i z n sąsiadów następujących bezpośrednio po brakach danych;

Uwaga! By odróżnić dane imputowane od danych rzeczywistych, miejsca zastąpione są oznaczane wybranym kolorem.

PRZYKŁAD 7.1. (plik: brakiDanychWydawca.pqs)
Analiza pliku wydawca.pqs nie zawierającego braków danych została omówiona w dziale Liniowa regresja wieloraka. Tym razem zajmiemy się arkuszem w którym w kolumnie zawierającej zysk brutto ze sprzedaży książek występują braki danych. W przypadku tych braków znane są wartości rzeczywiste (arkusz: ”RZECZYWISTE DANE”), można zatem odnieść wartości wygenerowane w programie dla braków danych do wartości rzeczywistych by porównać wyniki uzyskane różnymi technikami. W przykładzie wykorzystamy 2 sposoby zastępowania braków danych: zastąpienie wartością mediany i wartością wyznaczoną na podstawie modelu regresji. Pozostałe możliwości pozostawiamy do samodzielnej pracy.

Zastąpienie braków danych wartością mediany wykonujemy na arkuszu pierwszym nazwanym ”Wstaw medianę”. Ustawiamy w oknie Braków danych zmienną uzupełnianą jako zysk brutto i wybieramy sposób zastępowania jako wartość mediany. W rezultacie w miejsce braków danych wpisana zostanie wartość 46,85 tysięcy dolarów.

Podejrzewamy, że zyski są większe, gdy dotyczą książek znanych autorów (kodowanych jako 1) a mniejsze, gdy dotyczą tych nieznanych (kodowanych jako 0). Wyliczymy więc medianę zysku brutto oddzielnie dla książek autorów znanych i nieznanych. Imputację wykonujemy na arkuszu danych o nazwie ”Wstaw dwie mediany”. Ustawiamy dwukrotnie filtr dla zmiennej definiującej popularność autora (zmienna 7) - raz podając wartość 1, a raz 0. Uzyskana mediana zysku brutto w grupie książek autorów popularnych to ok. 51 tysięcy dolarów, a wśród tych mniej znanych to ok. 34 tysiące dolarów.

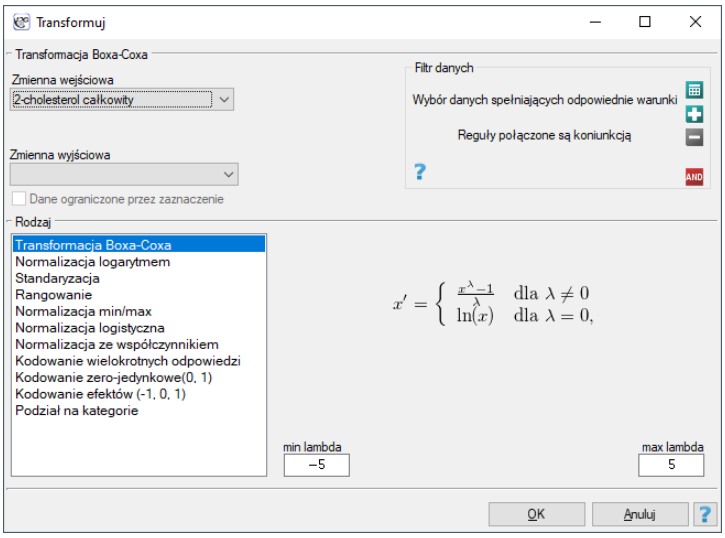
Innym sposobem zastępowania braków jest np. skorzystanie z modelu regresji. Wybieramy arkusz danych ”Wstaw z regresji” i ponownie wybieramy w oknie Braków danych zmienną dotyczącą zysku brutto jako zmienną, którą należy uzupełnić, jako sposób zastępowania wybieramy natomiast Wartości przewidywane z regresji. Zmiennych dzięki którym będziemy przewidywać wartość zysku brutto będzie tym razem więcej - będą to: koszty produkcji (zmienna 3), koszty reklamy (zmienna 4) i popularność autora (zmienna 7). Tym razem wyniki wydają się mniej odbiegać od rzeczywistych wartości, niestety brakuje wyniku dla pozycji o numerze 35. Dla tej książki nie mieliśmy bowiem informacji o koszcie produkcji na którym to między innymi chcieliśmy oprzeć przewidywanie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

TRANSFORMACJE



Okno transformacji otwieramy przez Dane→Transformuj...



Transformacja to przekształcenie danych tak, by spełniały określone kryteria - np. rozkład normalny albo rozciągnięcie w wybranym przedziale.

Transformacja Boxa-Coxa

Transformacja Boxa-Coxa wprowadzona przez Boxa i Coxa w roku 1964[25] sprowadza dane do rozkładu normalnego poprzez przekształcenie oparte na współczynniku λ. Do przeprowadzenia transformacji wymagane są dane dodatnie. Jeśli dane nie są dodatnie zaleca się ich wcześniejsze przekształcenie do liczb dodatnich metodą normalizacji min-max.

Transformacja Boxa-Coxa wyraża się wzorem:

$$x' = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{dla } \lambda \neq 0 \\ \ln(x) & \text{dla } \lambda = 0, \end{cases} \tag{1}$$

gdzie wartość λ wyznaczana jest jako maksymalna wartość logartmu funkcji wiarygodności (LL) w podanym przez badacza przedziale. Domyślny przedział dla poszukiwania wartości λ to przedział [-5, 5], a funkcja LL opisana jest wzorem:

$$LL = -\frac{n}{2}\ln(sd_{pop}^2) + (\lambda - 1)\sum \ln x \tag{2}$$

gdzie:
n - liczność próby,
sd_{pop} - odchylenie standardowe populacji.

Uwaga! Jeśli przed transformacją Boxa-Coxa wykorzystano normalizację min-max, po transformacji Boxa-Coxa można wrócić do poprzedniego przedziału, ponownie stosując tę normalizację.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Transformacja logarytmiczna

Transformacja logarytmiczna może być wykorzystywana do redukcji skośności rozkładu tzn. w sytuacji gdy mamy do czynienia z rozkładem lognormalnym.

$$x' = \ln x \tag{3}$$

Standaryzacja

Standaryzacja, to przekształcenie danych, w wyniku którego zmienna uzyskuje średnią równą 0 a odchylenie standardowe równe 1.

$$x' = \frac{x - \bar{x}}{sd} \tag{4}$$

Rangowanie

Rangi - są to kolejne liczby (zwykle naturalne) przypisane do wartości uporządkowanych pomiarów badanej zmiennej. Często wykorzystywane są w tych testach nieparametrycznych, które bazują wyłącznie na kolejności elementów w próbie. Przypisanie do zmiennej wyliczonych według niej rang zwane jest **rangowaniem**. Rangowanie może odbywać się dla zmiennych sortowanych rosnąco (jest to domyślne ustawienie) lub malejąco.

Powtarzającym się wartościom zmiennej przypisuje się **rangę wiązaną**. Rangą wiązaną może być:
- średnia arytmetyczna wyliczana z proponowanych dla powtarzanych wartości kolejnych liczb naturalnych - jest to domyślne ustawienie;
- dolna granica, czyli najmniejsza z proponowanych dla powtarzanych wartości kolejnych liczb naturalnych;
- górna granica, czyli największa z proponowanych dla powtarzanych wartości kolejnych liczb naturalnych.

Na przykład dla zmiennej o następujących wartościach: 8.6, 5.3, 8.6, 7.1, 9.3, 7.2, 7.3, 7.4, 7.3, 5.2, 7, 9.9, 8.6, 5.7 przypisywane są następujące rangi:

posortowane wartości zmiennej	rangi
5.2	1
5.3	2
5.7	3
7	4
7.1	5
7.2	6
7.3	7.5
7.3	7.5
7.4	9
8.6	11
8.6	11
8.6	11
9.3	13
9.9	14

Przy czym dla zmiennej o wartości 7.3 przypisana jest ranga wiązana wyliczona jako średnia arytmetyczna z liczb: 7 i 8, a dla zmiennej o wartości 8.6 ranga wiązana wyliczona z liczb: 10, 11, 12.

Normalizacja min-max

Normalizacja min-max przy pomocy funkcji liniowej sprowadza dane do wskazanego przez użytkownika przedziału (new_{min}, new_{max}). Powinniśmy przy tym znać zakres jaki mogą osiągnąć dane.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jeśli nie znamy tego zakresu, możemy posłużyć się wartością największą i najmniejszą występującą w analizowanym zbiorze (w oknie Transformacji zaznaczamy wtedy opcję oblicz z próby).

$$x' = \frac{x - \min}{\max - \min} \cdot (new_{\max} - new_{\min}) + new_{\min} \tag{5}$$

Normalizacja logistyczna

Normalizacja przy pomocy funkcji logarytmicznej (S-kształtnej) sprowadza dane poddane standaryzacji do wskazanego przedziału.

$$x' = \frac{e^x}{1 - e^x} \tag{6}$$

Jeśli tak przekształcone dane chcemy rozciągnąć na innym przedziale niż zadany, wówczas w oknie Transformacji należy wprowadzić zakres nowego przedziału.

Funkcja normalizująca ze współczynnikiem

Normalizacja ta sprowadza dane zestandaryzowane do wskazanego przedziału przy pomocy funkcji S-kształtnej o zmieniającym się współczynniku normalizacji α.

$$x' = \frac{x}{\sqrt{x^2 + \alpha}} \tag{7}$$

Zwiększenie wartości współczynnika α tworzy wykres o bardziej łagodnym zboczu.
Jeśli tak przekształcone dane chcemy rozciągnąć na innym przedziale niż zadany, wówczas w oknie Transformacji należy wprowadzić zakres nowego przedziału.

Kodowanie wielokrotnych odpowiedzi

Ten rodzaj kodowania pozwala przygotować odpowiedzi udzielone na pytania wielokrotnego wyboru w taki sposób, by ułatwić ich dalszą obróbkę statystyczną. W efekcie zastosowania tej transformacji wybrana zmienna o k- możliwościach odpowiedziach jest rozbijana na k nowych zmiennych. Przy czym należy podać jaki znak (lub zestwa znaków) jest separatorem poszczególnych kategorii. Np. Zapytano ankietowanych jaki alkohol piją? Dane zapisano w kolumnie Alkohol oddzielając wielokrotne odpowiedzi znakiem średnika. Taki zapis danych nie pozwala nawet na proste podsumowanie. Nie można między innymi szybko zliczyć ile osób pije wino. Po przekodowaniu wielokrotnych odpowiedzi uzyskano trzy nowe kolumny - po jednej dla każdej z możliwych odpowiedzi. Każdą z tych kolumn można teraz poddać analizie statystycznej.

Alkohol	Alkohol(piwo)	Alkohol(wino)	Alkohol(wódka)
piwo;wino	1	1	0
wino	0	1	0
wino	0	1	0
piwo	1	0	0
wódka;wino	0	1	1
wino;wódka	0	1	1
piwo;wódka	1	0	1
piwo;wino;wódka	1	1	1

Kodowanie zero-jedynkowe

Transformacja zmiennej o k kategoriach poprzez kodowanie zero-jedynkowe pozwala na uzyskanie $k-1$ zmiennych fikcyjnych. Taka forma przekształcania wykorzystywana jest przede wszystkim w modelach regresji. Dokładny opis tego typu transformacji można znaleźć w dziale 25.0.2 PRZYGOTOWANIE ZMIENNYCH DO ANALIZY W MODELACH WIELOWYMIAROWYCH.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Kodowanie efektów

Transformacja zmiennej o k kategoriach poprzez kodowanie efektów pozwala na uzyskanie $k-1$ zmiennych fikcyjnych. Taka forma przekształcania wykorzystywana jest przede wszystkim w modelach regresji i ANOVA. Dokładny opis tego typu transformacji można znaleźć w dziale 25.0.2 PRZYGOTOWANIE ZMIENNYCH DO ANALIZY W MODELACH WIELOWYMIAROWYCH.

Podział na kategorie

Taki sposób przygotowania danych pozwala na dowolny podział zmiennych np. cholesterol całkowity możemy podzielić zgodnie z obowiązującymi normami (wówczas wybieramy podział Ręczny, ustawiamy liczbę kategorii i sami wpisujemy ich granice oraz nadajemy odpowiednie etykiety dla każdej kategorii). Jeśli jednak nie mamy gotowego pomysłu na podział naszych danych można skorzystać z zaproponowanych w oknie opcji podziału automatycznego. Możliwe sposoby podziału zmiennej:

- **Naturalny Podział (Jenks)** - metoda polegająca na takim podziale zmiennej na klasy, by zminimalizować wariancję w klasach a zmaksymalizować wariancję pomiędzy klasami.
- **Podział według kwantyli** - metoda polegająca na podziale zmiennej na klasy równej liczności.
- **Odchylenie standardowe** - metoda polegająca na podziale zmiennej na klasy w oparciu o oddalenie od średniej o 1, 2 lub więcej odchyleń standardowych.
- **Błąd standardowy średniej** - metoda polegająca na podziale zmiennej na klasy w oparciu o oddalenie od średniej o 1, 2 lub więcej błędów standardowych średniej.
- **Ręczny** - metoda polegająca na podziale zmiennej na klasy wg dowolnego podziału wprowadzonego ręcznie przez badacza.

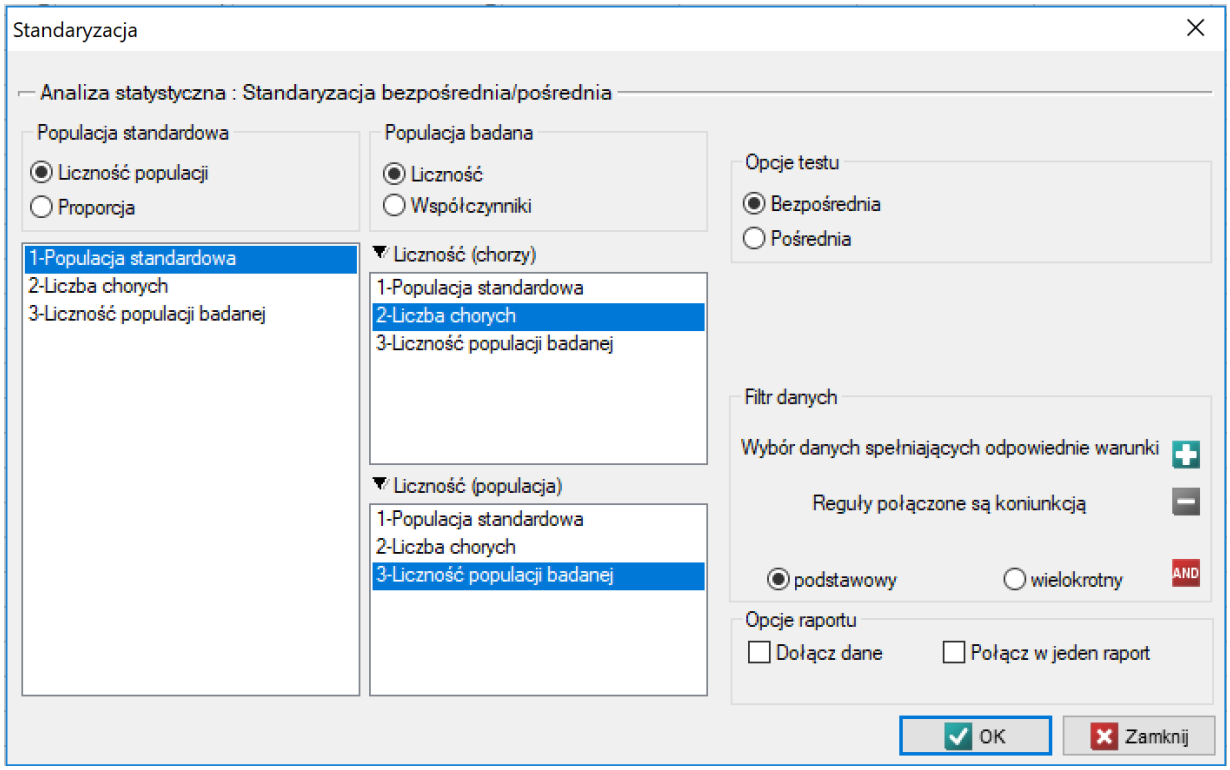
W oknie podziału możliwe jest również wybranie opcji Dodaj schemat kolorów wówczas kolumna, która będzie przechowywała nowe dane zostanie oznaczona kolorami zgodnie ze wskazanym schematem.

PRZYKŁAD 7.2. (plik: normalizacja.pqs)
Dokonaj przekształcenia zmiennych zawartych w pliku:
a) Przekształć wartość trójglicerydów poprzez transformację Boxa-Coxa a następnie sprawdź przy pomocy odpowiedniego testu, czy dane te mają rozkład normalny.
b) Przekształć wartość trójglicerydów poprzez transformację logarytmiczną a następnie sprawdź przy pomocy odpowiedniego testu, czy dane te mają rozkład normalny.
c) Stosując normalizację min-max przekształć wybrane zmienne do przedziału [0,10].
d) Stosując normalizację logistyczną przekształć wybrane zmienne do zadanego przez siebie przedziału.
e) Stosując normalizację ze współczynnikiem przekształć wybrane zmienne do zadanego przez siebie przedziału. Zrób to kilkukrotnie zmieniając wartość współczynnika α.
f) Dokonaj standaryzacji wszystkich danych, które są opisane rozkładem normalnym.
g) Przekształć zmienną obrazującą jak zmieniła się masa ciała podczas stosowania diety tak, by przedstawiała ona rozkład normalny.
h) Pytanie o przebytych chorobach zakaźnych było pytaniem wielokrotnego wyboru. Przygotuj uzyskane odpowiedzi na to pytanie tak, by można było poddać je dalszej obróbce statystycznej tzn. zapisz każdą z wielokrotnych odpowiedzi w innej kolumnie.
i) Przygotuj zmienną wykształcenie tak, by była zapisana przy pomocy zmiennych fikcyjnych o kodowaniu zero-jedynkowym.
j) Przygotuj zmienną cholesterol całkowity dzieląc ją na 3 klasy wg podziału na percentyle (kwartyle). Powstałym klasom nadaj etykiety :”niski”, ”przeciętny”, ”wysoki” i dobierz schemat kolorów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

STANDARYZACJA BEZPOŚREDNIA I POŚREDNIA

Okno bezpośredniej i pośredniej standaryzacji współczynników epidemiologicznych otwieramy przez Dane→Standaryzacja...



Standaryzacji pośredniej i bezpośredniej podlegają współczynniki częstości, np. współczynniki chorobowości. Porównywane populacje zwykle zamieszkują różne rejony geograficzne i różnią się czynnikami mogącymi wpływać na wartość współczynników (np. wiekiem) - dlatego bezpośrednie porównanie surowych współczynników między nimi może maskować występujące różnice. By zniwelować wpływ struktury populacji co do tych cech wikłających, porównania można dokonać na bazie współczynników standaryzowanych. Standaryzacja pozwala "odfiltrować" wpływ znanego czynnika wikłającego (np. wieku) i dzięki temu czyni porównywalnymi współczynniki standaryzowane uzyskane z różnych populacji. Niezbędnym etapem procesu standaryzacji współczynników jest wybór populacji standardowej. Populacją standardową dla populacji zajmującej określony obszar geograficzny w określonym czasie może być populacja obejmująca większy geograficznie obszar, w tym również obszar badany, np. jeśli populacją badaną jest populacja województwa wielkopolskiego, to populacja Polski może być wykorzystana jako populacja standardowa. Można również dokonać wyboru zupełnie innej populacji, odległej geograficznie od populacji badanej. Jednak dobrze jest, aby wybrana populacja stanowiła populację odniesienia nie tylko do przeprowadzanego właśnie badania, ale również dla wielu innych badaczy. Daje to bowiem możliwość porównania wyników badań z wykorzystaniem tej samej populacji standardowej. Wybierając populację, zaleca się zwrócenie uwagi na kilka aspektów dokonywanego wyboru, m.in.:

- jeśli kilka populacji jest porównywanych, wspólna populacja standardowa minimalizuje wariancję (zmienność) uzyskanych współczynników standardowych;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- w analizie trendów polecaną populacją standardową jest populacja przedstawiająca średnią strukturę dla analizowanego odcinka czasu;
- populacja standardowa powinna być jak najbardziej podobna do populacji badanej;
- ta sama populacja standardowa powinna być konsekwentnie wybierana w celu zapewnienia porównywalności badań (wybór innej niż dotychczas używanej powszechnie populacji standardowej powoduje, iż wszystkie historyczne dane należałoby ponownie przeliczyć).

Wiek oraz płeć są najczęściej wykorzystywanymi cechami, względem których dokonuje się standaryzacji, niemniej standaryzacja może być dokonywana również w oparciu o inne, dowolne cechy, które ze względu na ich oczywisty wpływ na dane zjawisko warto „odfiltrować" z przeprowadzanego badania. Cechy takie nazywane są cechami zakłócającymi lub wikłającymi. Przy czym wybierając cechę, względem której chcemy dokonać standaryzacji, należy pamiętać, że standaryzacja będzie możliwa, jeśli mamy wystarczające informacje o rozkładzie tej cechy w populacji badanej i populacji standardowej (Tabela 1). Dodatkowo standaryzacja według wybranej cechy, np. według wieku, kompensuje w pewnym stopniu oddziaływanie innych czynników wikłających związanych z wiekiem, takich jak styl życia, a standaryzacja według płci kompensuje te czynniki, które są związane z płcią, np. wykonywane zawody. Kompensacja innych czynników jest więc ważnym aspektem przy doborze cechy, względem której dokonuje się standaryzacji.

Rodzaje standaryzacji:

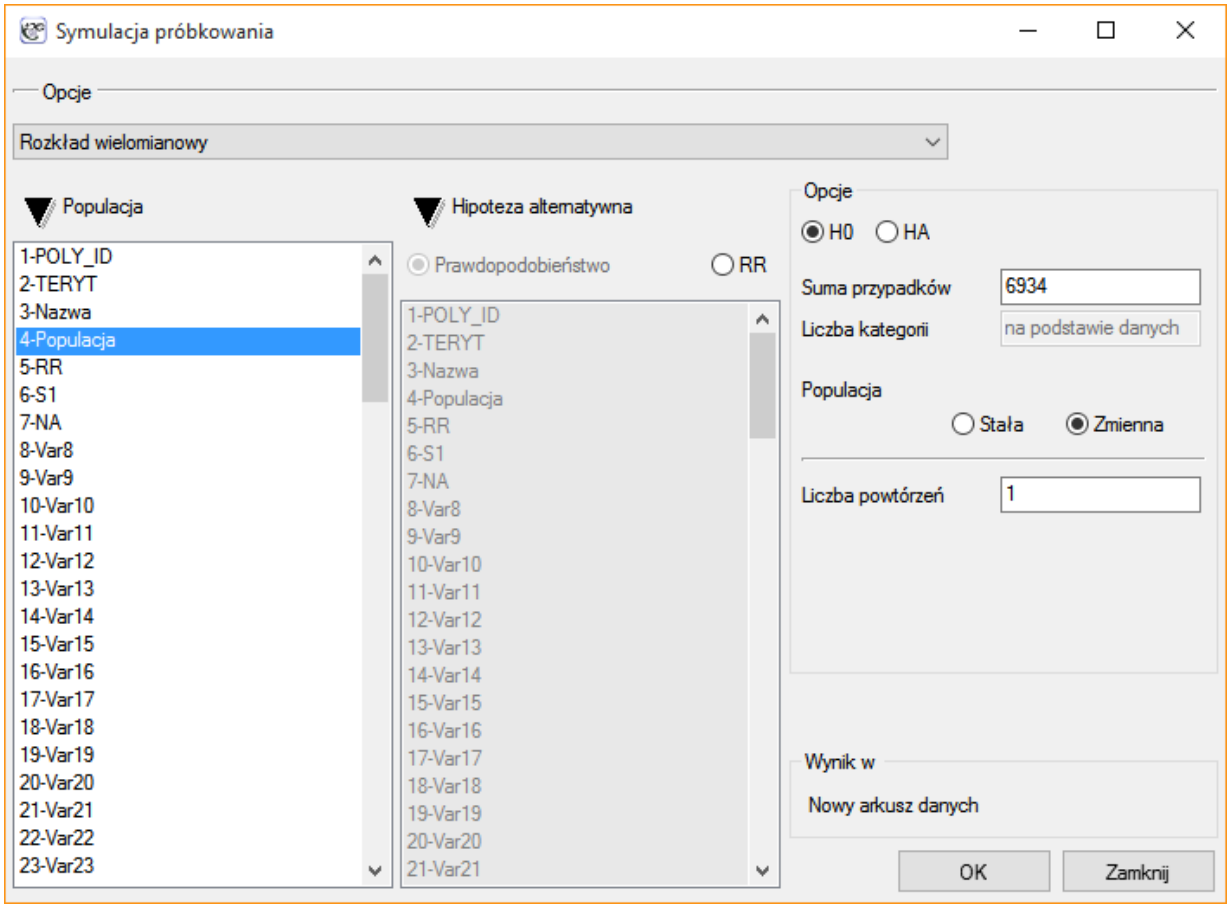
- **standaryzacja bezpośrednia** *(ang. direct standardization)* – uzyskany tą metodą standaryzowany współczynnik częstości podaje, jak wyglądałaby częstość występowania choroby w populacji badanej, gdyby posiadała ona strukturę (np. strukturę wiekową) populacji referencyjnej;
- **standaryzacja pośrednia** *(ang. indirect standardization)* – uzyskany tą metodą standaryzowany współczynnik częstości podaje, jak wyglądałaby częstość występowania choroby w populacji badanej, gdyby występowanie choroby w populacji badanej było takie samo w poszczególnych kategoriach (np. kategoriach wiekowych) jak w populacji referencyjnej.

SYMULACJA PRÓBKOWANIA



Okno próbkowania otwieramy przez Dane→Symulacja próbkowania ...

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Symulacja próbkowania jest jednym ze sposobów generowania danych rozkładu wielomianowego. Polega na przydzieleniu podanej liczby przypadków do poszczególnych kategorii, w sposób zadany przez użytkownika. Wygenerowane dane zwracane są w nowym arkuszu. Generowanie może zostać powtórzone, tak by w arkuszu uzyskać wiele wygenerowanych kolumn w zależności od ustawionej w oknie próbkowania liczby powtórzeń tej operacji.

Opcje:

H0 - hipoteza zerowa zakłada równomierne rozmieszczenie wszystkich przypadków w poszczególnych kategoriach.

HA - hipoteza alternatywna zakłada nierównomierne rozmieszczenie przypadków. Wybranie tej opcji wymaga wskazania kategorii o większym prawdopodobieństwie lub relatywnym ryzyku. Informacja o zdefiniowanym prawdopodobieństwie lub relatywnym ryzyku dla każdej kategorii powinna zostać wprowadzona do wybranej kolumny arkusza przed przeprowadzeniem analizy.

Prawdopodobieństwo należy zdefiniować jako wartość od 0 do 1, przy czym suma prawdopodobieństw podanych dla wszystkich kategorii powinna wynosić 1.

Relatywne ryzyko definiuje ryzyko względem innych kategorii i dla kategorii o zwiększonym ryzyku jest wartością większą niż 1, a ułamkiem mniejszym niż 1 dla kategorii o zmniejszonym ryzyku.

Ustawienie wartości prawdopodobieństwa lub relatywnego ryzyka na tym samym poziomie dla wszystkich kategorii, tożsame jest z rozkładem dla H0.

Populacja stała - zakłada, że użytkownik zainteresowany jest rozmieszczeniem przypadków zgodnie z zaproponowanym rozkładem.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

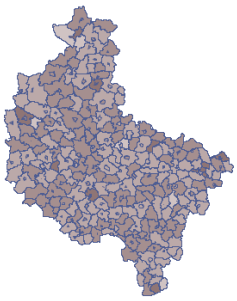
Populacja zmienna - zakłada, że użytkownik zainteresowany jest takim rozmieszczeniem przypadków, aby proporcja przypadków do populacji rozmieszczona została zgodnie z zaproponowanym rozkładem.

Przykład 7.3. (plik symulacje.pqs)
Jako podstawę symulacji użyto populację Wielkopolski w roku 2013, która wynosiła wg GUS N = 3467016 osób. Województwo podzielone jest na 315 gmin. Gminy różnią się znacznie liczbą mieszkańców. Najliczniejsza gmina (stolica województwa) 548028 mieszkańców, najmniej liczna 1454 mieszkańców, mediana i kwartyle to odpowiednio: 6298 (4462; 9621) mieszkańców. Przy założeniu, że w 2013 roku mieszkańców województwa z chorobą X było 6934, należy zasymulować rozlokowanie osób chorych w taki sposób, by uzyskać:

1. Rozkład losowy (na podstawie danych z arkusza ”Losowy”)
2. Czterokrotnie większe częstości występowania choroby we wskazanych gminach niż w pozostałej części województwa (na podstawie danych z arkusza ”Klastery”)

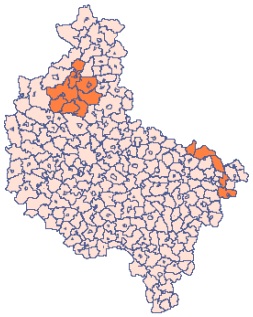
Ad 1. Należy zaznaczyć, że równomierne rozłożenie 6934 chorych w sposób losowy nie oznacza podobnej liczby chorych w każdej gminie. Wiadomo, że gminy o większej liczbie narażonych powinny mieć odpowiednią większą liczbę chorych niż te o mniejszej liczbie mieszkańców. Interesujące jest więc takie rozłożenie chorych, by współczynnik liczby chorych do liczby mieszkańców był względnie stały. Oznacza to przyjęcie hipotezy zerowej H0 i zmiennej populacji. Liczność poszczególnych gmin zapisano w kolumnie o nazwie: populacja.

Wylosowane na podstawie tych założeń dane przedstawiono w pierwszej kolumnie nowego arkusza danych. By móc obserwować losowy rozkład współczynnika chorych w poszczególnych gminach, należy przekopiować uzyskany wynik do arkusza ”Losowo” kolumny ”S1”. Formuła znajdująca się w kolumnie 7 zostanie wówczas ponownie przeliczona (podejrzeć i zmienić formułę można ustawiając Kody/Etykiety/Format we właściwościach kolumny). Na mapie przedstawiamy uzyskany wynik przy pomocy menagera Map  z menu Analiza przestrzenna. Wyrysowana zostaje wówczas proporcja chorych do liczby mieszkańców w poszczególnych gminach. Przykładowy wynik przedstawia poniższa mapa.



Ad 2. W arkuszu ”Klastery” podano, podobnie jak w poprzednim zadaniu, liczność dla populacji badanej. Tym razem oczekujemy wyższej częstości w niektórych gminach (wskazanych na mapie) więc dodatkowo w kolejnej kolumnie arkusza przedstawiono wartość relatywnego ryzyka dla poszczególnych gmin ustawiając ją na 4, dla gmin zwiększonego ryzyka i 1 dla pozostałych gmin.

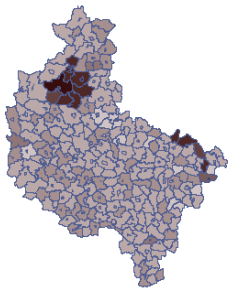
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Odpowiednie próbkowanie wymaga od nas wybrania hipotezy alternatywnej HA (poprzez wybranie kolumny z relatywnym ryzykiem) i zmiennej populacji (poprzez wskazanie kolumny z licznością populacji gmin). Wylosowane na podstawie tych założeń dane przedstawiono w pierwszej kolumnie nowego arkusza danych.

By móc obserwować rozkład współczynnika, przy założeniu większego ryzyka we wskazanych gminach, należy przekopiować uzyskany wynik do arkusza "Klastery" kolumny "S1". Formuła znajdująca się w kolumnie 7 zostanie wówczas ponownie przeliczona. Na mapie przedstawiamy uzyskany wynik przy pomocy menagera Map 

z menu Analiza przestrzenna. Wyrysowana zostaje wówczas proporcja chorych do liczby mieszkańców w poszczególnych gminach. Przykładowy wynik przedstawia poniższa mapa.



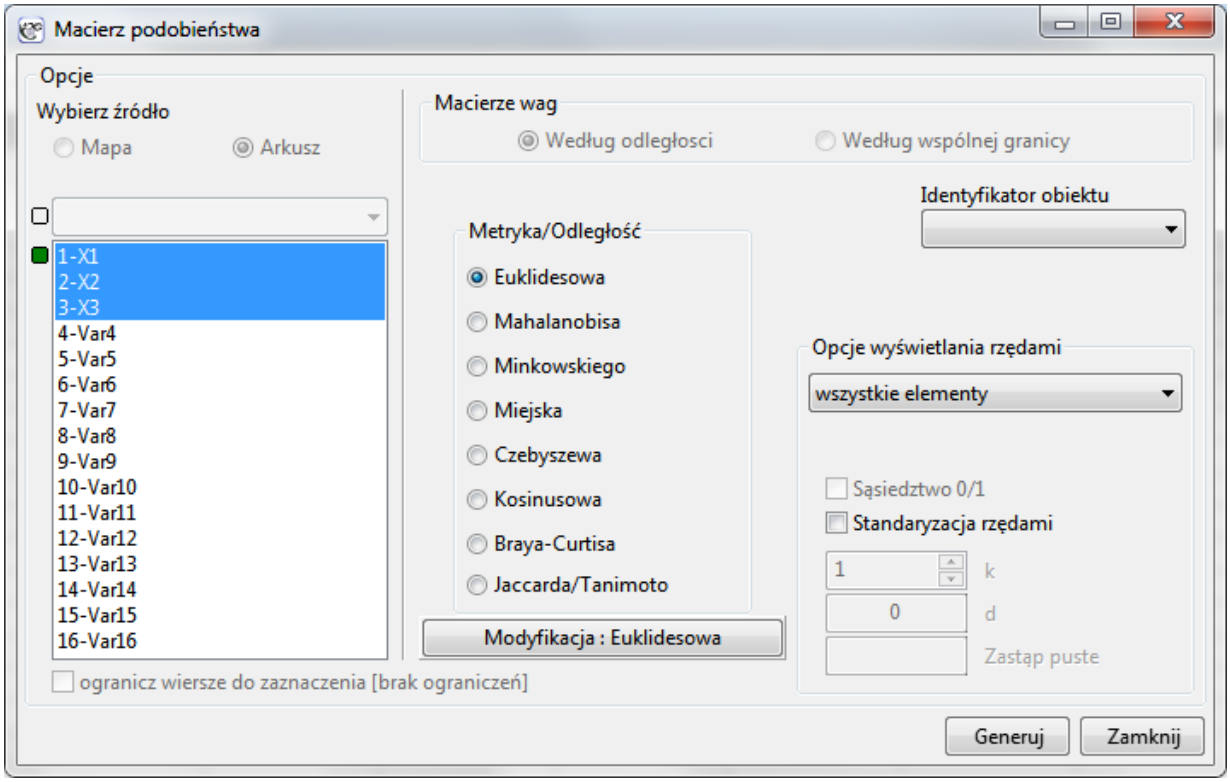
MACIERZ PODOBIEŃSTWA



Relacje między obiektami można wyrazić przez ich odległość albo, bardziej ogólnie, przez niepodobieństwo. Im dalej od siebie znajdują się obiekty, tym bardziej są niepodobne; im bliżej, tym większe jest ich podobieństwo. Odległość obiektów można badać pod względem wielu cech - np. dla porównywanych miast podobieństwo można określać na podstawie: długości drogi je łączącej, gęstości zaludnienia, PKB na mieszkańca, emisji zanieczyszczeń, przeciętnych cen nieruchomości itd. Mając tak wiele różnych cech, trzeba dobrać taką miarę odległości, która najlepiej odda rzeczywiste podobieństwo obiektów.

Okno z ustawieniami opcji macierzy podobieństwa otwieramy przez menu Dane→Macierz podobieństwa…

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Niepodobieństwo/podobieństwo obiektów wyrażamy za pomocą odległości będących najczęściej **metryką**. Nie każda miara odległości jest jednak metryką. Aby odległość mogła być nazwana metryką musi spełniać 4 warunki:

1. odległość pomiędzy obiektami nie może być ujemna: $d(x_1, x_2) \geq 0$,
2. odległość między dwoma obiektami wynosi 0 wtedy i tylko wtedy gdy są one identyczne: $d(x_1, x_2) = 0 \iff x_1 = x_2$,
3. odległość musi być symetryczna, tzn. odległość z obiektu x_1 do x_2 musi być taka sama jak z x_2 do x_1: $d(x, y) = d(y, x)$,
4. odległość musi spełniać warunek trójkąta: $d(x, z) \leq (x, y) + d(y, z)$.

Uwaga! Metryki powinny być wyliczane dla cech o podobnych zakresach wartości. W przeciwnym razie cechy o wyższych zakresach będą miały większy wpływ na wynik podobieństwa niż te o niższych zakresach. Przykładowo, licząc podobieństwo osób na podstawie masy ciała i wieku, masa ciała w kilogramach (zakres 40-150) będzie miała większy wpływ na wynik niż wiek w latach (zakres 18-90). Aby wpływ każdej cechy był zrównoważony, warto każdą z nich znormalizować/wystandaryzować przed analizą. Chcąc samodzielnie zdecydować o sile tego wpływu, po standaryzacji można przy wybranym rodzaju metryki wpisać własne wagi. **Odległość/Metryka:**

Euklidesowa

Kiedy mówimy o odległości nie definiując jej rodzaju zakładamy, że jest to odległość Euklidesowa - najpopularniejszy typ odległości stanowiący naturalny element modeli świata rzeczywistego. Odległość euklidesowa jest metryką i dana jest wzorem:

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^{n} (x_{1k} - x_{2k})^2}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Minkowskiego

Odległość Minkowskiego definiowana jest dla parametrów p i r równych sobie - jest wówczas metryką. Ten rodzaj metryki pozwala sterować procesem wyliczania podobieństwa poprzez podanie wartości p i r ujętymi we wzorze:

$$d(x_1, x_2) = \sqrt[p]{\sum_{k=1}^{n} |x_{1k} - x_{2k}|^r}$$

Zwiększając parametr r zwiększamy wagę przypisaną różnicy pomiędzy obiektami dla każdej cechy, zmieniając p nadajemy większe/mniejsze znaczenie obiektom bliższym/dalszym. Jeśli r i p są równe 2, to odległość Minkowskiego sprowadza się do odległości Euklidesowej, jeśli są równe 1, do odległości Miejskiej, a przy tych parametrach dążących do nieskończoności, do metryki Czebyszewa.

Miejska (inaczej: odległość Manhattan lub odległość taksówkowa)

To odległość pozwalająca na poruszanie się tylko w dwóch prostopadłych do siebie kierunkach. Ten rodzaj odległości przypomina poruszanie się po prostopadle przecinających się ulicach (kwadratowa sieć ulic przypominająca plan Manhattanu). Metryka ta dana jest wzorem:

$$d(x_1, x_2) = \sum_{k=1}^{n} |x_{1k} - x_{2k}|$$

Czebyszewa

Odległość pomiędzy porównywanymi obiektami to największa z uzyskanych odległości dla poszczególnych cech tych obiektów:

$$d(x_1, x_2) = \max_k |x_{1k} - x_{2k}|$$

Mahalanobisa

Odległość Mahalanobisa nazywana jest również odległością statystyczną. Jest to odległość ważona macierzą kowariancji, przez co porównywać można obiekty opisane wzajemnie skorelowanymi cechami. Stosowanie odległości Mahalanobisa daje dwie podstawowe korzyści:

1) Zmienne, dla których obserwowane są większe odchylenia lub większe zakresy wartości nie mają zwiększonego wpływu na wynik odległości Mahalanobisa (stosując macierz kowariancji standaryzujemy bowiem zmienne wykorzystując wariancję znajdującą się na diagonali). W rezultacie przed przystąpieniem do analizy nie ma wymogu standaryzowania/normalizacji zmiennych.

2) Bierze pod uwagę wzajemne skorelowanie cech opisujących porównywane obiekty (stosując macierz kowariancji wykorzystujemy informację o zależności między cechami znajdującą się poza przekątną macierzy).

$$d(x_1, x_2) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})}$$

Wyliczona w ten sposób miara spełnia warunki metryki.

Kosinusowa

Odległość kosinusowa powinna być wyliczana na danych dodatnich ponieważ nie jest ona metryką (nie spełnia pierwszego warunku: $d(x_1, x_2) \geq 0$). Jeśli więc mamy cechy przyjmujące również

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

wartości ujemne powinnyśmy je wcześniej przekształcić stosując np. normalizację do przedziału rozpiętego na liczbach dodatnich. Zaletą tej odległości jest fakt, że (dla dodatnich argumentów) ograniczona jest do przedziału [0, 1]. Podobieństwo dwóch obiektów reprezentuje kąt pomiędzy dwoma wektorami przedstawiającymi cechy tych obiektów.

$$d(x_1, x_2) = 1 - K,$$

gdzie K to współczynnik podobieństwa (kosinus kąta pomiędzy dwoma znormalizowanymi wektorami):

$$K = \frac{\sum_{k=1}^{n} x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^{n} x_{1k}^2} \sqrt{\sum_{k=1}^{n} x_{2k}^2}}$$

Obiekty są podobne, gdy wektory się pokrywają - wtedy kosinus kąta (podobieństwo) wynosi 1 odległość (niepodobieństwo) 0. Obiekty są różne, gdy wektory są prostopadłe - wtedy kosinus kąta (podobieństwo) wynosi 0 odległość (niepodobieństwo) wynosi 1.

Przykład - porównanie tekstów
Tekst 1: na tym przystanku wsiadło kilka osób a na następnym wysiadła jedna
Tekst 2: na przystanku jedna Pani wysiadła a kilka wsiadło
Chcemy wiedzieć, jak podobne są teksty pod względem liczby tych samych słów, nie interesuje nas natomiast kolejność występowania słów.

Tworzymy listę słów z obu tekstów i liczymy jak często wystąpiło każde słowo:

SŁOWA	na	tym	przystanku	wsiadło	kilka	osób	a	następnym	wysiadła	jedna	Pani
tekst 1	2	1	1	1	1	1	1	1	1	1	0
tekst 2	1	0	1	1	1	0	1	0	1	1	1

Kosinus kąta pomiędzy wektorami wynosi 0,784465, a więc odległość miedzy nimi nie jest duża $d(tekst_1, tekst_2) = 1 - 0,784465 = 0,215535$.
Na podobnej zasadzie możemy porównywać dokumenty pod względem występowania słów kluczowych, tak by znajdować te najbardziej odpowiadające zapytaniu.

Braya-Curtisa

Odległość (miara niepodobieństwa) Bray'a-Curtis'a powinna być wyliczana na danych dodatnich ponieważ nie jest ona metryką (nie spełnia pierwszego warunku: $d(x_1, x_2) \geq 0$). Jeśli mamy cechy przyjmujące również wartości ujemne, powinnyśmy je wcześniej przekształcić stosując np. normalizację do przedziału rozpiętego na liczbach dodatnich. Zaletą tej odległości jest fakt, że (dla dodatnich argumentów) ograniczona jest do przedziału [0, 1], gdzie 0 - oznacza, że porównywane obiekty są podobne, 1 - niepodobne.

$$d(x_1, x_2) = \frac{\sum_{k=1}^{n} |x_{1k} - x_{2k}|}{\sum_{k=1}^{n} (x_{1k} + x_{2k})} \tag{8}$$

Wyliczając miarę podobieństwa BC od wartości 1 odejmujemy odległość Bray'a-Curtis'a:

$$BC = 1 - d(x_1, x_2) \tag{9}$$

Jaccarda

Odległość (miara niepodobieństwa) Jaccarda wyliczana jest dla zmiennych binarnych (Jaccard, 1901), gdzie 1 oznacza występowanie danej cech 0- jej brak.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

		objekt 1	
		1	0
objekt 2	1	a	b
	0	c	d

Odległość Jacckarda wyraża się wzorem:

$$d(x_1, x_2) = 1 - J. \tag{10}$$

gdzie:

$J = \frac{a}{a+b+c}$ - współczynnik podobieństwa Jaccarda.

Współczynnik podobieństwa Jaccarda zawiera się w przedziale [0,1], gdzie 1 oznacza najwyższe podobieństwo, 0 - najniższe. Odległość (niepodobieństwo) interpretujemy przeciwnie: 1 - oznacza, że porównywane obiekty są niepodobne, 0 - że bardzo podobne. Sens współczynnika podobieństwa Jaccarda dobrze opisuje sytuacja dotycząca wyboru towaru przez klientów. Przez 1 oznaczymy fakt zakupu danego produktu przez klienta, 0 - klient nie kupił tego artykułu. Wyliczając współczynnik Jaccarda porównamy 2 produkty by dowiedzieć się jaka część klientów kupuje je w tandemie. Nie interesuje nas oczywiście informacja o klientach, którzy nie kupili żadnego z porównywanych artykułów. Jesteśmy natomiast ciekawi jak wiele osób wybierających jeden z porównywanych produktów wybiera jednocześnie ten drugi. Suma $a + b + c$ - to liczba klientów, którzy wybrali któryś z porównywanych artykułów, a - to liczba klientów wybierających oba artykuły jednocześnie. Im wyższy współczynnik podobieństwa Jaccarda, tym bardziej nierozerwalne są artykuły (zakupowi jednego towarzyszy zakup drugiego). Odwrotnie będzie, gdy dostaniemy wysoki współczynnik niepodobieństwa Jaccarda. Będzie on świadczył o dużej konkurencyjności artykułów, tzn. zakup jednego będzie powodował brak zakupu drugiego.

Wzór na współczynnik podobieństwa Jaccarda można również zapisać w ogólnej postaci:

$$J = \frac{\sum_{k=1}^{n} x_{1k}x_{2k}}{\sum_{k=1}^{n} x_{1k}^2 \sum_{k=1}^{n} x_{2k}^2 - \sum_{k=1}^{n} x_{1k}x_{2k}}$$

zaproponowanej przez Tanimoto (1957). Ważną cechą formuły Tanimoto jest fakt, że może być wyliczana także dla cech ciągłych.
W przypadku danych binarnych wzory na niepodobieństwo/podobieństwo Jaccarda i Tanimoto są tożsame i spełniają warunki metryki. Natomiast dla zmiennych ciągłych wzór Tanimoto nie jest metryką (nie spełnia warunku trójkąta).

Przykład - porównanie gatunków
Badamy podobieństwo pod względem genetycznym przedstawicieli trzech różnych gatunków - w sensie liczby genów które są dla nich wspólne. Jeśli gen występuje w organizmie, to dajemy mu wartość 1, 0 - w przeciwny przypadku. Dla prostoty przykładu analizie poddanych jest zaledwie 10 genów.

GENY	gen1	gen2	gen3	gen4	gen5	gen6	gen7	gen8	gen9	gen10
osobnik1	0	1	1	1	1	1	1	0	1	0
osobnik2	0	0	1	1	1	1	1	0	1	0
osobnik3	1	0	1	1	0	0	1	0	0	0

Wyliczona macierz podobieństwa przedstawia się następująco:

OBIEKTY	osobnik1	osobnik2	osobnik3
osobnik1	0	0.857143	0.375
osobnik2	0.857143	0	0.428571
osobnik3	0.375	0.428571	0

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Najbardziej podobni są osobnicy 1 i 2 a najmniej 1 i 3: - podobieństwo Jaccarda osobnika1 i osobnika2 wynosi 0.857143, czyli nieco ponad 85% genów występujących w obu porównywanych gatunkach jest dla nich wspólna.
- podobieństwo Jaccarda osobnika1 i osobnika3 wynosi 0.375, czyli ponad 37% genów występujących w obu porównywanych gatunkach jest dla nich wspólna.
- podobieństwo Jaccarda osobnika2 i osobnika3 wynosi 0.428571, czyli prawie 43% genów występujących w obu porównywanych gatunkach jest dla nich wspólna.

Opcje macierzy podobieństwa wykorzystujemy do wskazania sposobu zwracania elementów w macierzy. Standardowo zwracane są wszystkie elementy macierzy, w takiej formie w jakiej były wyliczone zgodnie z przyjętą metryką. Możemy to zmienić ustawiając:

Elementy macierzy:

- minimum - oznacza, że w każdym wierszu macierzy zostanie wyświetlona tylko wartość minimalna i wartość na głównej przekątnej;
- maksimum - oznacza, że w każdym wierszu macierzy zostanie wyświetlona tylko wartość maksymalna i wartość na głównej przekątnej;
- k minimalnych - oznacza, że w każdym wierszu macierzy zostanie wyświetlonych tyle najmniejszych wartości ile wskaże użytkownik podając wartość k oraz wartość na głównej przekątnej;
- k maksymalnych - oznacza, że w każdym wierszu macierzy zostanie wyświetlonych tyle największych wartości ile wskaże użytkownik podając wartość k oraz wartość na głównej przekątnej;
- elementy poniżej d - oznacza, że w każdym wierszu macierzy zostaną wyświetlone te elementy, których wartość będzie mniejsza niż wskazana przez użytkownika wielkość d oraz wartość na głównej przekątnej;
- elementy powyżej d - oznacza, że w każdym wierszu macierzy zostaną wyświetlone te elementy, których wartość będzie większa niż wskazana przez użytkownika wielkość d oraz wartość na głównej przekątnej;

Sąsiedztwo 0/1
: Wybierając opcję Sąsiedztwo $0/1$ wartości wewnątrz macierzy zastępujemy wartością 1, a miejsca puste wartością 0. W ten sposób oznaczamy na przykład czy obiekty sąsiadują (1) czy nie (0), czyli wyznaczamy macierz sąsiedztwa.

Standaryzacja rzędami
: Standaryzacja rzędami oznacza, że każdy element macierzy dzielony jest przez sumę wiersza macierzy. W rezultacie uzyskane wartości znajdują się w przedziale od 0 do 1.

Zastąp puste
: Opcja Zastąp puste pozwala na wpisanie wartości jaka ma zostać umieszczona w macierzy w miejscu ewentualnych pustych elementów.

Wybrany identyfikator obiektu pozwala nazwać wiersze i kolumny macierzy podobieństwa zgodnie z nazewnictwem przechowywanym we wskazanej zmiennej.

Przykład 7.4. (plik: podobienstwoLokali.pqs)
W procedurach wyceny nieruchomości, zarówno ze względów merytorycznych jak też prawnych, kwestia podobieństwa pełni ważną rolę. Jest na przykład zasadniczą przesłanką umożliwiającą grupowanie

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

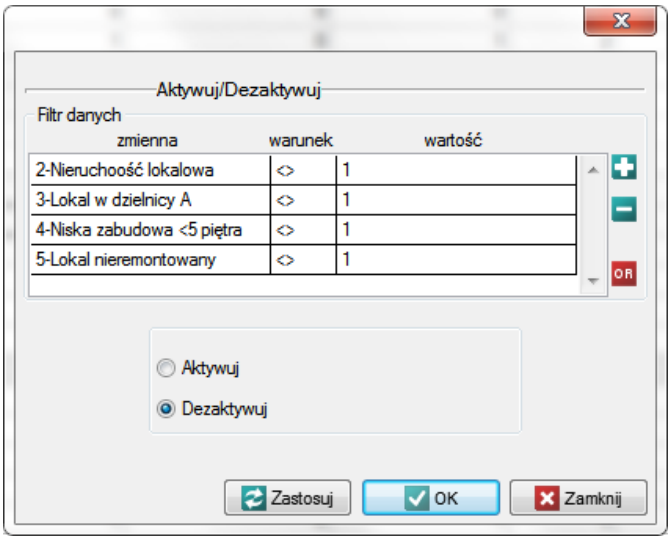
obiektów i przypisywanie do odpowiedniego segmentu.

Załóżmy, że do pośrednika nieruchomości zgłasza się osoba poszukująca mieszkania, która definiuje te cechy, które lokal musi posiadać i te, które mają duży wpływ na decyzję o zakupie ale nie są decydujące. Cechy, które lokal musi posiadać to:

- nieruchomość lokalowa (stanowiąca przedmiot odrębnej własności),
- położona w dzielnicy A,
- położona w niskiej zabudowie wielorodzinnej (do 5-ciu kondygnacji),
- nie remontowana (standard przeciętny lub pogorszony).

Dane dotyczące tych lokali zebrano w tabeli, gdzie 1 oznacza, że lokal spełnia warunki wyszukiwania, 0 że ich nie spełnia.

Te lokale, które nie spełniają warunków wyszukiwania wyłączymy z analizy poprzez dezaktywację odpowiednich wierszy. Poprzez menu Edycja→Aktywuj/Dezaktywuj (filtr)... dezaktywujemy te wiersze, które nie spełniają choćby jednego z postawionych warunków.



Pamiętamy by warunki dezaktywacji połączone były alternatywą (zmieniamy AND na OR).
W rezultacie wyłoniono 11 lokali (lokal 10, 12, 17, 35, 88, 101, 105, 122, 130, 132, 135) pasujących do tego segmentu (spełniających wszystkie 4 warunki).

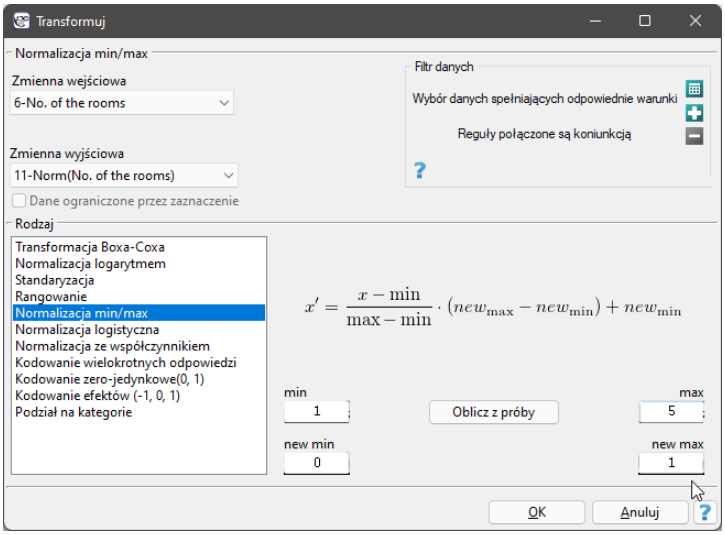
Teraz weźmiemy pod uwagę te cechy, które mają duży wpływ na decyzję klienta, ale nie są decydujące:

- Liczba pokoi = 3;
- Piętro na którym znajduje się lokal = 0;
- Wiek budynku w którym znajduje się lokal = ok. 3 lata;
- Bliskość centrum dzielnicy A (czas jaki zajmuje dotarcie do centrum) = ok. 30 min;
- Bliskość przystanku komunikacji miejskiej = ok. 80 m.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

	Liczba pokoi	Piętro lokalu	Wiek budynku	Dystans do centrum	Odległość przystanku
Poszukiwany	3	0	3	30	80
Lokal 10	2	1	1	0	150
Lokal 12	1	2	1	0	200
Lokal 17	3	1	7	20	500
Lokal 35	2	0	6	5	100
Lokal 88	3	4	6	5	200
Lokal 101	4	2	10	0	10
Lokal 105	2	2	6	0	50
Lokal 122	1	0	6	5	100
Lokal 130	2	0	10	0	20
Lokal 132	3	5	6	30	400
Lokal 135	3	1	6	5	100

Zauważmy, że ostatnia cecha, czyli odległość przystanku komunikacji miejskiej jest wyrażona znacznie większymi liczbami niż pozostałe cechy porównywanych lokali. W rezultacie cecha ta będzie miała znacznie większy wpływ na uzyskany wynik macierzy odległości niż pozostałe cechy. Chcąc temu zapobiec przed analizą normalizujemy wszystkie cechy wybierając dla nich wspólny zakres od 0 do 1 - w tym celu korzystamy z menu Dane→Transformacja.... W oknie normalizacji jako zmienną wejściową ustawiamy "Liczbę pokoi", a jako zmienną wyjściową pustą zmienną nazwaną "Norm(Liczba pokoi)"; rodzaj normalizacji to normalizacja min/max; wartości min i max wyliczamy z próby wybierając przycisk Oblicz z próby - wynik normalizacji zostanie zwrócony do arkusza danych po wybraniu przycisku Wykonaj. Normalizację powtarzamy dla kolejnych zmiennych czyli: "Piętra", "Wieku budynku", "Dystansu do centrum" i "Odległości przystanku".



Znormalizowane dane przedstawia poniższa tabela.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

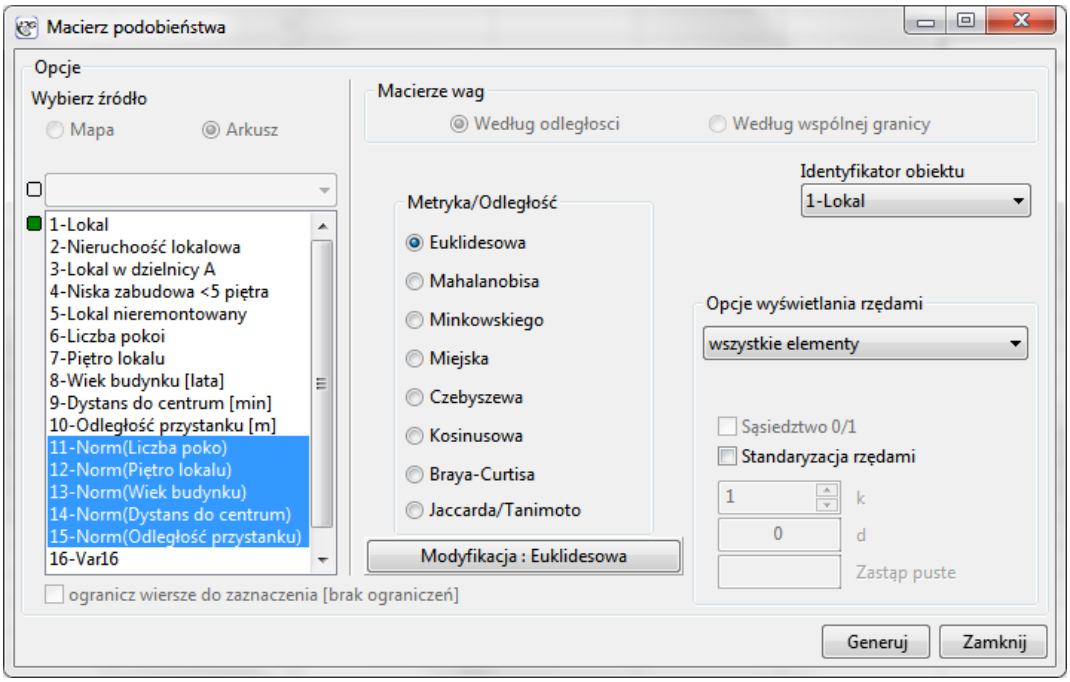
	Norm(Liczba pokoi)	Norm(Piętro lokalu)	Norm(Wiek budynku)	Norm(Dystans do centrum)	Norm(Odległość przystanku)
Poszukiwany	0,666666667	0	0,222222222	1	0,142857143
Lokal 10	0.333333333	0.2	0	0	0.285714286
Lokal 12	0	0.4	0	0	0.387755102
Lokal 17	0.666666667	0.2	0.666666667	0.666666667	1
Lokal 35	0.333333333	0	0.555555556	0.166666667	0.183673469
Lokal 88	0.666666667	0.8	0.555555556	0.166666667	0.387755102
Lokal 101	1	0.4	1	0	0
Lokal 105	0.333333333	0.4	0.555555556	0	0.081632653
Lokal 122	0	0	0.555555556	0.166666667	0.183673469
Lokal 130	0.333333333	0	1	0	0.020408163
Lokal 132	0.666666667	1	0.555555556	1	0.795918367
Lokal 135	0.666666667	0.2	0.555555556	0.166666667	0.183673469

Bazując na danych znormalizowanych wyznaczymy lokale najbardziej dopasowane do zapytania klienta. Do wyliczenia podobieństwa posłużymy się metryką (odległością) euklidesową. Czym mniejszą uzyskamy wartość, tym bardziej podobne będą lokale. Analizę przeprowadzić można zakładając, że każda z pięciu wymienionych przez klienta cech jest tak samo ważna, ale można również wskazać te cechy, które powinny w większym stopniu wpływać na wynik analizy. Zbudujemy dwie macierze odległości euklidesowych:

(1) W pierwszej macierzy znajdą się odległości euklidesowe wyliczone na podstawie równoważnie traktowanych pięciu cech;

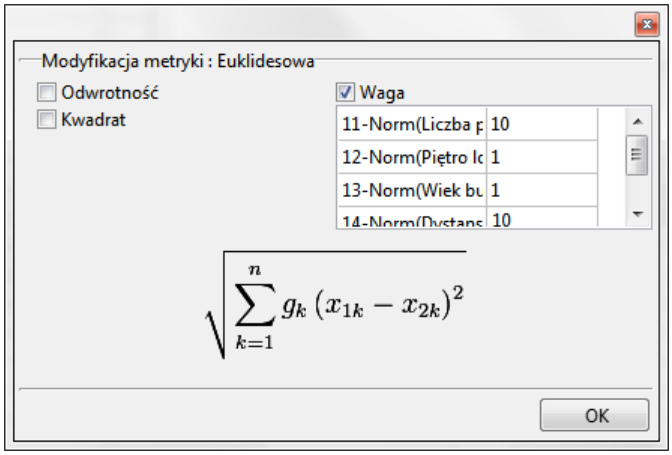
(2) W drugiej macierzy znajdą się odległości euklidesowe, w budowie których największe znaczenie będzie miała liczba pokoi i dystans do centrum dzielnicy.

By zbudować pierwszą macierz, w oknie macierzy podobieństwa wybieramy 5 znormalizowanych zmiennych oznaczonych jako Norm, metrykę Euklidesową i jako Identyfikator obiektu zmienną ”Lokal”.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

By zbudować drugą macierz, w oknie macierzy podobieństwa dokonujemy tych samych ustawień co przy budowie pierwszej macierzy, ale dodatkowo wybieramy przycisk Modyfikacja : Euklidesowa i w oknie modyfikacji wpisujemy większe wagi dla ”Liczby pokoi” i ”Dystansu do centru” np. równe 10, a mniejsze dla pozostałych cech np. równe 1.



W rezultacie uzyskamy dwie macierze. W każdej z nich pierwsza kolumna dotyczy podobieństwa do lokalu szukanego przez klienta:

Euklidesowa	**Poszukiwany**	...
Poszukiwany	0	...
Lokal 10	1.10	...
Lokal 12	1.31	...
Lokal 17	1.04	...
Lokal 35	0.96	...
Lokal 88	1.23	...
Lokal 101	1.38	...
Lokal 105	1.18	...
Lokal 122	1.12	...
Lokal 130	1.32	...
Lokal 132	1.24	...
Lokal 135	0.92	...

Euklidesowa z wagami	**Poszukiwany**	...
Poszukiwany	0	...
Lokal 10	3.35	...
Lokal 12	3.84	...
Lokal 17	1.44	...
Lokal 35	2.86	...
Lokal 88	2.78	...
Lokal 101	3.45	...
Lokal 105	3.37	...
Lokal 122	3.39	...
Lokal 130	3.43	...
Lokal 132	1.24	...
Lokal 135	2.66	...

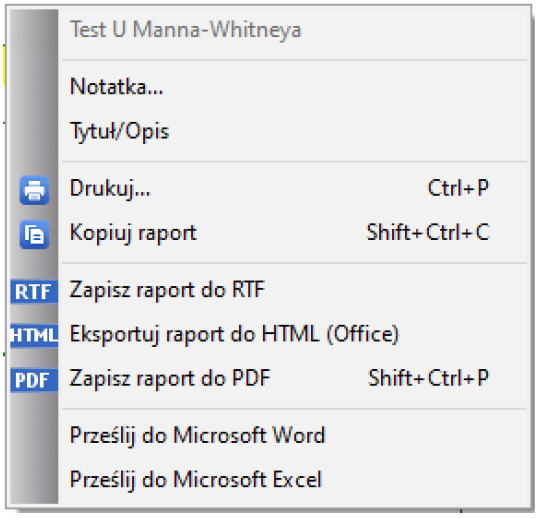
Według niemodyfikowanej odległości euklidesowej warunkom klienta najbardziej odpowiadać powinien lokal 35 i lokal 135. Gdy uwzględnimy wagi, najbardziej zbliżonymi do wymogów klienta będą lokale 17 i 132 - są to lokale, które w pierwszej kolejności są podobne pod względem wymaganej przez klienta liczby pokoi (3) i wskazanej odległości do centrum, mniejszy wpływ na wynik tego podobieństwa mają 3 pozostałe cechy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

7.2 PRACA Z ARKUSZAMI WYNIKÓW (RAPORTAMI)

Raport to element projektu, w którym przechowywane są wyniki wykonanych analiz statystycznych. Program dodaje go do projektu automatycznie, przypisując do aktywnego arkusza danych, gdy tylko procedura statystyczna się zakończy. Raportu nie da się edytować, poza wykresem i tytułem. Wykres edytujemy dwukrotnym kliknięciem myszy albo z menu kontekstowego prawego przycisku myszy, a tytuł - w Zarządcy Projektu, dodając lub zmieniając opis.

Najważniejsze operacje na raporcie dostępne są z menu kontekstowego prawego przycisku myszy w oknie raportu:



- **Drukowanie**
  Opcje drukowania znajdziemy w:
  – menu kontekstowym,
  – menu Plik→Drukuj...

- **Eksportowanie i przesyłanie raportów**
  Raporty z ℙℚStat można eksportować do pliku w formacie *.rtf, *.html i *.pdf, a także przesłać wprost do dokumentów Word lub Excel.
  Eksportując z poziomu Zarządcy Projektu, można umieścić raporty w oddzielnych plikach albo w jednym wspólnym. W tym celu zaznaczamy wybrane raporty, klikamy przycisk i eksportujemy do pliku (lub plików) w wybranym formacie. Pojedynczy raport można też wyeksportować osobno z menu kontekstowego w oknie raportu.

- **Opisywanie raportów**
  Raporty opiszemy w Zarządcy Projektu albo w oknie raportu, dodając tytuł lub notatkę.

- **Edycja wykresów**
  Opcje ogólne i szczegółowe wykresu edytuje się z menu kontekstowego w oknie raportu.

- **Kopiowanie raportów**
  Wyniki analiz można też przenosić do innych aplikacji (np. Word, Excel) za pomocą schowka systemowego.

- **Usuwanie raportów**
  Raport usuniemy przez:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

– menu kontekstowe Usuń raport (Shift+Del) na nazwie raportu w Drzewie nawigacji,
– Zarządcę Projektu.

Uwaga! Jeśli do raportu dołączone są warstwy mapy, usunięcie raportu usuwa jednocześnie wszystkie przypisane do niego warstwy. **Zmiana kolejności raportów** jest możliwa z menu kontekstowego prawego przycisku myszy: Przenieś w górę (Ctrl+Up) lub Przenieś w dół (Ctrl+Down) na nazwie raportu w Drzewie nawigacji.

Do nazwy raportu w Drzewie nawigacji można też **dodać dodatkowe informacje**, takie jak:

- godzina wygenerowania,
- opis,
- filtr,
- nazwa zmiennej grupującej,
- nazwa zmiennej.

Wybieramy je w oknie ustawień programu.

7.3 MENU I USTAWIENIA JĘZYKOWE PROGRAMU

JĘZYK

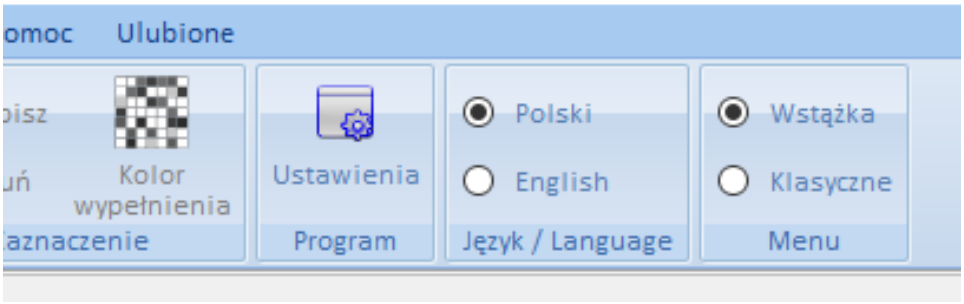
Interfejs programu oraz tworzone raporty mogą być wyświetlane zarówno w języku polskim, jak i w języku angielskim. Język zmienia się poprzez wybranie Language/Język z menu Edycja - nie wymaga to ponownego uruchomienia programu. Raporty otwarte po zmianie języka zostaną automatycznie przetłumaczone, z wyjątkiem nazwy procedury, która stanowi opis podlegający edycji użytkownika.

MENU

Menu programu można wyświetlić w formie **klasycznej** lub jako **menu wstążkowe**. Sposób wyświetlania przełącza się w menu Edycja, wybierając odpowiedni rodzaj menu.

Ulubione
Menu Ulubione umożliwia szybki dostęp do często wykonywanych statystyk i innych operacji. Do zbudowania własnego menu służy okno wyszukiwarki Ulubionych, za pomocą którego można odnaleźć dowolną pozycję menu, a następnie dodać ją do Ulubionych lub z nich usunąć.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

8 ORGANIZACJA PRACY Z PROGRAMEM

Wszystkie procedury analiz statystycznych dostępne są w menu Statystyka.

8.1 FORMA ORGANIZACJI DANYCH

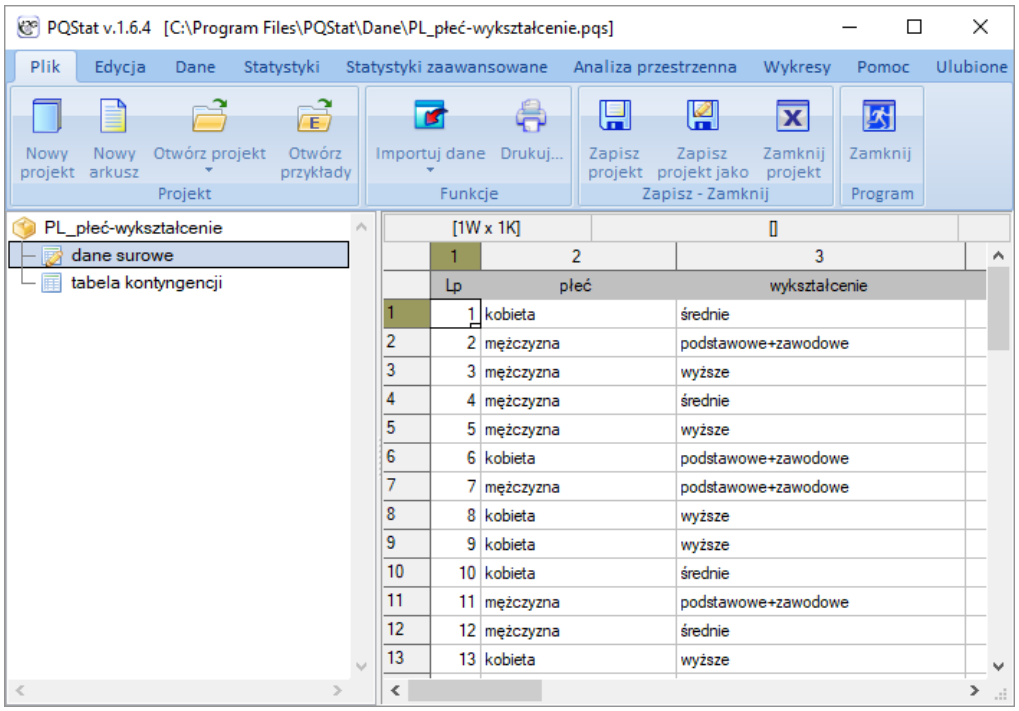
Forma organizacji danych zależy od procedur statystycznych, jakie użytkownik planuje wykonać.

Analiza statystyczna może dotyczyć danych zebranych w postaci tabeli kontyngencji lub w postaci surowej. Dane można zamienić z jednej postaci na drugą:

- z tabeli kontyngencji do postaci surowej – poprzez menu Dane → Przestaw na dane surowe...,
- z postaci surowej do tabeli kontyngencji – poprzez menu Dane → Przestaw na tabelę...

1. Dane w postaci surowych rekordów, to dane zorganizowane w taki sposób, by każdy wiersz zawierał informacje o innym badanym obiekcie (pacjencie, firmie itp.)

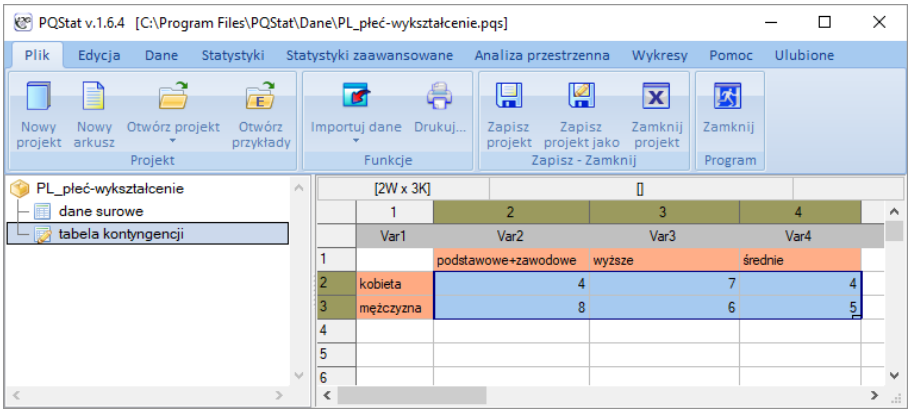
   PRZYKŁAD 8.1. *Dane surowe* (plik płeć-wykształcenie.pqs)



2. Tabela kontyngencji, to tabela przedstawiająca łączny rozkład dwóch zmiennych. Jej wnętrze, to liczności obserwowane (liczby naturalne).

   PRZYKŁAD 8.2. *Tabela kontyngencji* (plik płeć-wykształcenie.pqs)

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

8.2 OBSZAR ROBOCZY ARKUSZA

Zwykle wykonując analizę statystyczną mamy dostęp do całego obszaru roboczego arkusza danych. Użytkownik może jednak swobodnie ograniczać ten obszar wskazując określoną przez siebie część arkusza. Takie wskazanie może dokonać się na kilka sposobów:

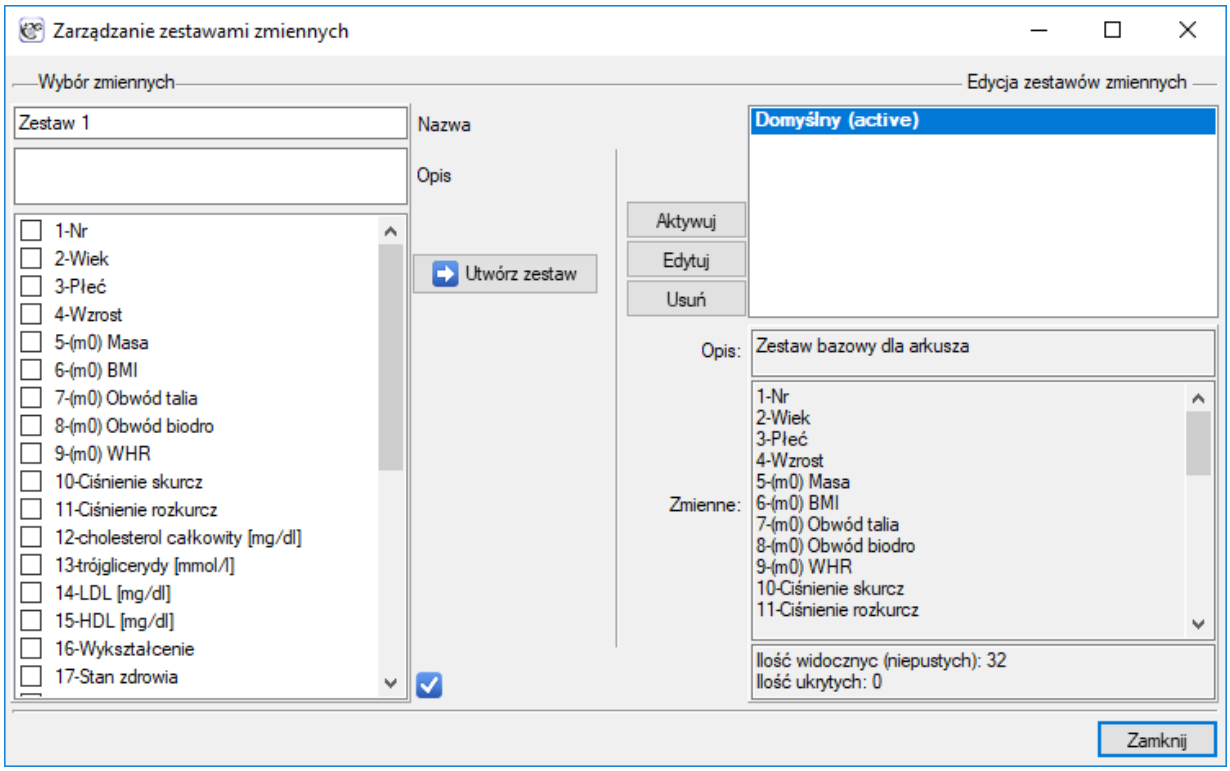
Organizowanie zmiennych w zestawy

Zestawy zmiennych, to podzbiory zmiennych. Definiowanie różnych zestawów ma na celu łatwiejszy wybór zmiennych w oknach analiz poprzez ograniczenie wyświetlanej listy zmiennych. Wybrany zestaw zmiennych jest wyświetlony na listach zmiennych w oknie dialogowym każdej analizy oraz w arkuszu danych.

Standardowo wykorzystywany jest domyślny zestaw zmiennych. Zestaw domyślny jest zestawem złożonym ze wszystkich zmiennych.

DEFINIOWANIE ZESTAWÓW ZMIENNYCH

By zdefiniować własne zestawy zmiennych, w górnej części okna arkusza dwukrotnie klikamy nazwę aktywnego zestawu.



W oknie Zarządzania zestawami zmiennych podajemy nazwę nowego zestawu (ewentualnie jego opis) oraz zaznaczamy na liście zmienne, które mają stanowić zestaw. Potwierdzamy wybór przyciskiem Utwórz zestaw. Zmienna może należeć do wielu zestawów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

EDYCJA ZESTAWÓW ZMIENNYCH

By edytować zestaw, w oknie Zarządzania zestawami zmiennych wybieramy zestaw z listy zestawów i potwierdzamy przyciskiem Edytuj. W oknie wyboru zmiennych pojawią się zaznaczone zmienne wybrane w zestawie oraz zmienne niezaznaczone, nie przynależące do zestawu. Po odpowiednim ponownym zaznaczeniu/odznaczeniu zmiennych potwierdzamy wybór przyciskiem Zapisz zmiany.

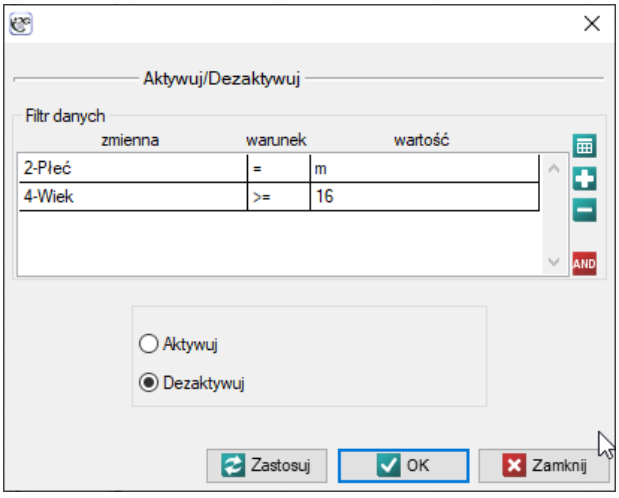
AKTYWOWANIE WYBRANEGO ZESTAWU ZMIENNYCH

Aktywowanie zestawu dokonywane jest w oknie Zarządzania zestawami zmiennych lub poprzez menu kontekstowe w górnej części arkusza danych na nazwie wybranego zestawu.

Aktywacja/dezaktywacja przypadków

Aktywacja/dezaktywacja przypadków jest opcją globalną i nadrzędną w stosunku do innych ograniczeń obszaru dostępnych w programie. Przypadki (wiersze) wskazane jako dezaktywowane są w arkuszu danych zacienione i nie biorą udziału w analizach statystycznych.
By aktywować lub dezaktywować wybrane przypadki należy wybrać jedną z poniższych opcji:

- wskazać wiersze w arkuszu danych i wybierać opcję Aktywuj/Dezaktywuj z menu kontekstowego na ich nazwie;
- wybrać menu Edycja →Aktywuj/Dezaktywuj (filtr)...



Przykład 8.3. (plik filtr.pqs)
Zamierzamy przeprowadzać wiele analiz statystycznych, na danych z pliku filtr.pqs. Analizy te nie mają dotyczyć chłopców, którzy mają 16 lat lub więcej. W tym celu określamy wiersze, których analiza nie będzie dotyczyć: wybieramy przycisk ➕ i ustawiamy regułę dla zmiennej płeć; ponownie wybieramy przycisk ➕ i ustawiamy regułę dla zmiennej wiek. Pamiętamy: aby wykonać to zadanie poprawnie wszystkie warunki filtru powinny być połączone koniunkcją (informuje o tym znak AND). Pozostawiamy wybraną opcję Dezaktywuj i tak ustawione warunki analizy potwierdzamy przyciskiem OK

Przy zawężaniu obszaru roboczego arkusza danych należy pamiętać, że warunki filtru mogą być połączone koniunkcją lub alternatywą. Zmianę alternatywy i koniunkcji dokonujemy przyciskami

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

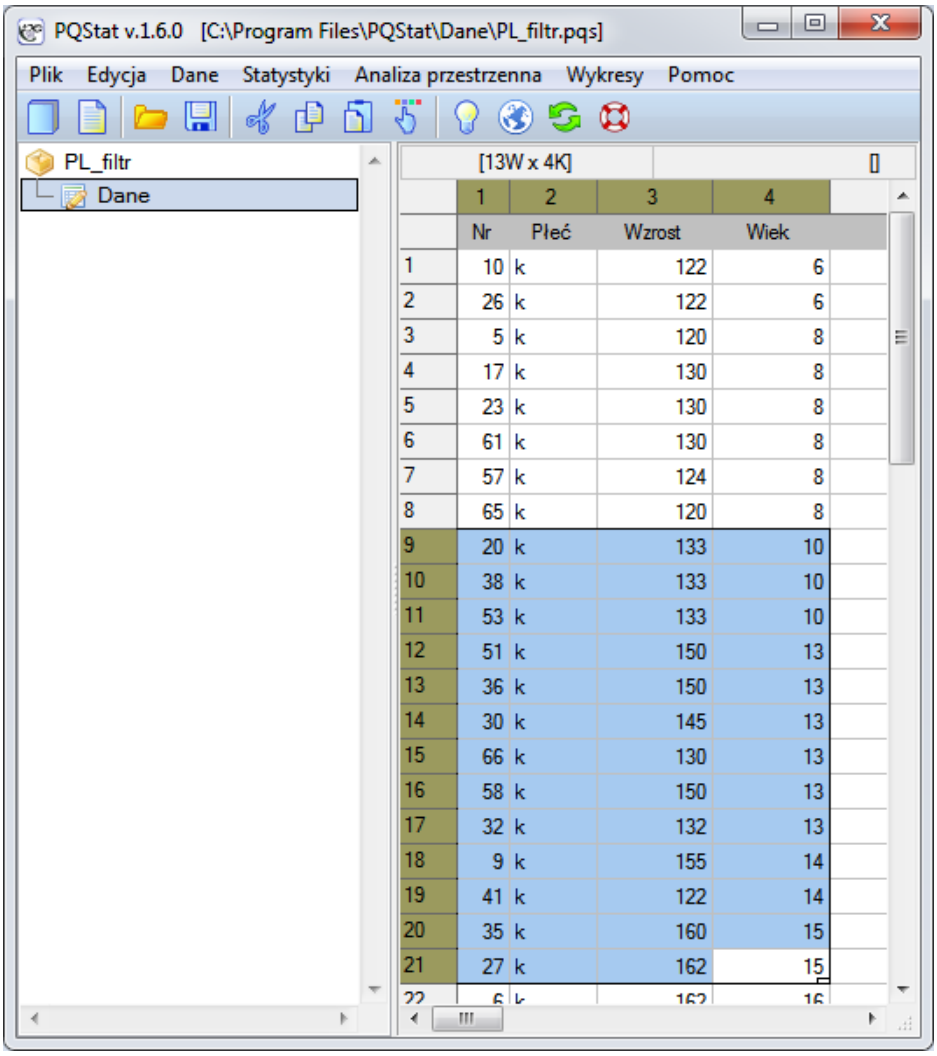
Więcej przykładów wykorzystania filtru znajdziesz tutaj.

By aktywować wszystkie przypadki należy wybrać menu Edycja →Aktywuj wszystkie

Zaznaczenie obszaru

Zaznaczenie spójnego obszaru powoduje, że wybrana przez nas następnie analiza będzie wykonywana tylko na tych wierszach, które znajdują się w zaznaczeniu i będzie miała do dyspozycji tylko te kolumny, które są wewnątrz zaznaczenia i zawierają dane.

PRZYKŁAD 8.4. (plik filtr.pqs)
Chcemy wyznaczyć statystyki opisowe dla wzrostu dziewczynek w wieku pomiędzy 10 a 15-stym rokiem życia. Sortujemy więc dane według kolumny płeć i według kolumny wiek, a następie zaznaczamy spójny obszar zawierający kolumnę ze wzrostem dziewczynek w wieku pomiędzy 10 a 15-stym rokiem życia i wybieramy z menu Statystyka→Analizy opisowe→Statystyki opisowe.



W oknie opcji testu statystyk opisowych wybieramy procedury, jakie chcemy wykonać (np. zaznaczamy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

średnią, odchylenie standardowe, minimum i maksimum) oraz zmienną do analizy (kolumnę z wzrostem) i potwierdzamy wybór przyciskiem OK.

Zawężanie obszaru roboczego arkusza danych poprzez zaznaczenie spójnego fragmentu danych powoduje wyświetlenie w oknie analizy komunikatu:
Dane ograniczone przez zaznaczenie

Zapisane zaznaczenie

Jeśli do arkusza danych przypisane są zaznaczone zakresy, to są one wyróżnione poprzez obramowanie. Można z nich korzystać w przypadku tych analiz, które mają możliwość wprowadzenia danych bezpośrednio w oknie analizy. Wówczas korzystając z przycisku wypełnij z zapisanego zaznaczenia, można wkleić dane znajdujące się w wybranym zakresie.

Przykład 8.5. (plik warstwy.pqs)
Chcemy wyznaczyć statystyki związane z analizą Ilorazu Szans (OR) dla kilku warstw. Skorzystamy z danych zapisanych w postaci 10 tabel – są one zaznaczone (obramowane). Z menu Statystyka wybieramy Analiza dla warstw → Mantel-Haenszel OR/RR. W oknie opcji testu zaznaczamy tabela kontyngencji a następnie ustawiamy liczbę warstw na 10. Każdą powstałą warstwę możemy teraz wypełnić z zapisanego zaznaczenia. Gdy wypełnimy wszystkie tabele, wykonujemy analizę, wybierając przycisk OK.

Uwaga! By przypisać kolejne zaznaczenie do arkusza danych, z menu Edycja wybieramy Zapisz zaznaczenie (Ctrl+T), a by usunąć przypisane zaznaczenia – Usuń zaznaczenia.

Filtr danych

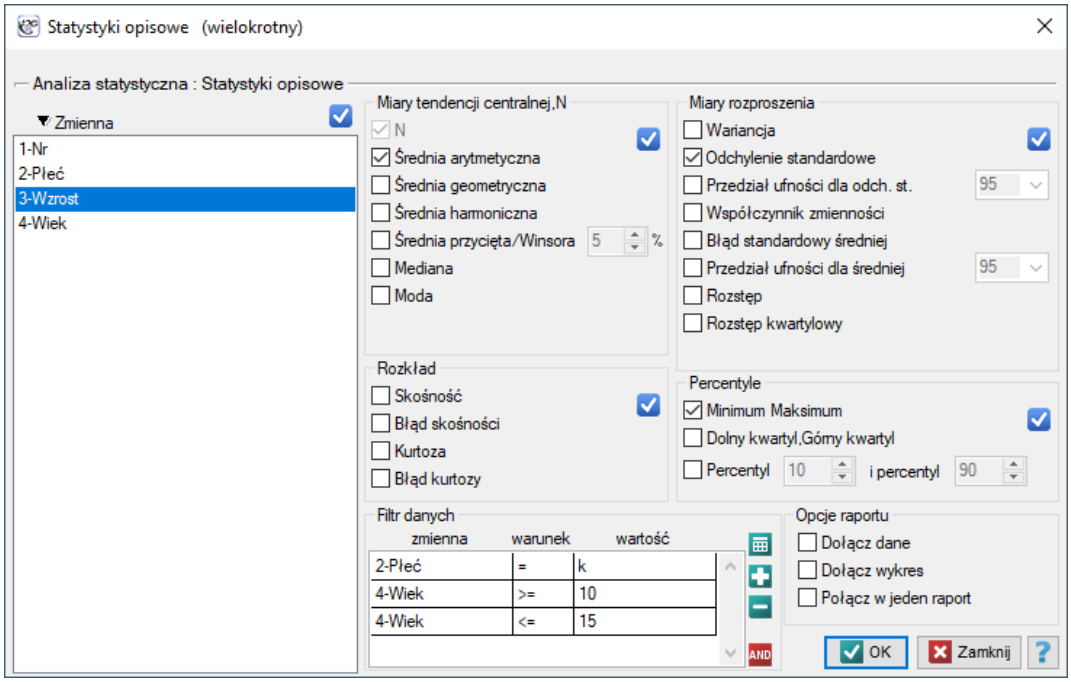
Filtr Danych jest opcją dostępną po wybraniu dowolnej analizy statystycznej. Jego ustawienie powoduje zmniejszenie liczby wierszy biorących udział w tej analizie. Możliwy do ustawienia jest filtr podstawowy lub wielokrotny.

- **Filtr podstawowy** pozwala wybrać jedną, szczególną podgrupę. Wybór może być dokonany na dwa sposoby:
  - **Sposób automatyczny** – w ten sposób możemy nałożyć jeden warunek na zebrane dane używając przycisku ➕, lub kilka warunków wciskając ponownie ➕. Warunki te połączone będą alternatywą lub koniunkcją w zależności od tego czy wybierzemy przycisk AND czy OR. Warunki usuwamy korzystając z przycisku ➖.

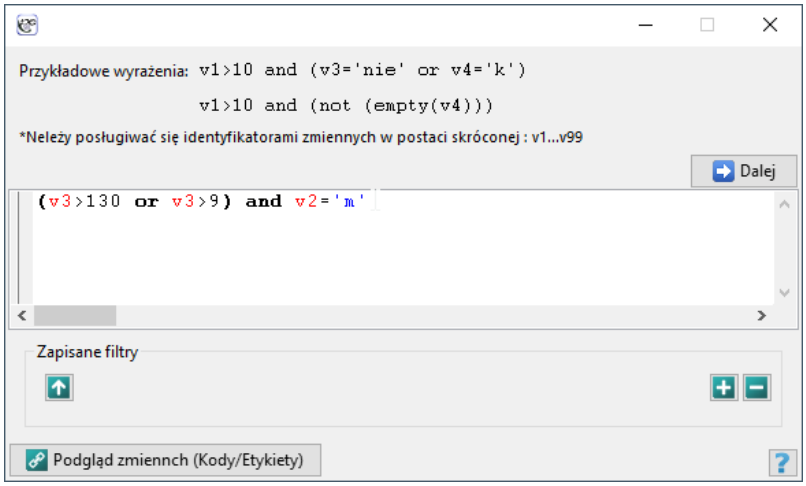
    Przykład 8.6. *Filtr podstawowy automatyczny* (plik filtr.pqs)
    Chcemy wyznaczyć statystyki opisowe dla wzrostu dziewczynek w wieku pomiędzy 10 a 15-stym rokiem życia. Z menu wybieramy Statystyka→Analizy opisowe→Statystyki opisowe. W oknie opcji testu statystyk opisowych wybieramy procedury jakie chcemy wykonać (np. zaznaczamy średnią, odchylenie standardowe, minimum i maksimum) oraz zmienną do analizy (kolumnę ze wzrostem). Filtr ustawiamy dodając reguły przyciskiem ➕. Najpierw regułę ustawiamy dla zmiennej płeć, jako warunek wybieramy znak równości, a jako wartość wybieramy literę k oznaczającą dziewczynki. Następnie dodajemy kolejny warunek i ustawiamy

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

filtr dla zmiennej wiek, jako warunek wybieramy znak >=, a jako wartość wpisujemy 10. W podobny sposób dodajemy warunek wieku <=15. Pamiętamy: aby wykonać to zadanie poprawnie wszystkie warunki filtru powinny być połączone koniunkcją (informuje o tym znak AND). Tak ustawione warunki analizy potwierdzamy przyciskiem OK.



– **Sposób zaawansowany** – w ten sposób możemy zapisać dowolną regułę, tzn. zarówno zupełnie prostą jak też bardziej złożoną. Sposób zaawansowany uruchamiamy przełączając przycisk na a następnie wybierając . W ten sposób uzyskujemy okno wewnątrz którego możemy budować formułę filtru.



Ustawiając filtr zaawansowany posługujemy się numerami zmiennych poprzedzonymi literą *v*. Numery zmiennych i ich zawartość możemy podejrzeć poprzez przycisk Podgląd zmiennych (Kody/Etykiety).

Do wykorzystania mamy następujące funkcje logiczne:
and - operator koniunkcji – sprawdza, czy wszystkie warunki, które łączy, są prawdziwe,
or - operator alternatywy – sprawdza, czy przynajmniej jeden z podanych warunków jest prawdziwy,

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

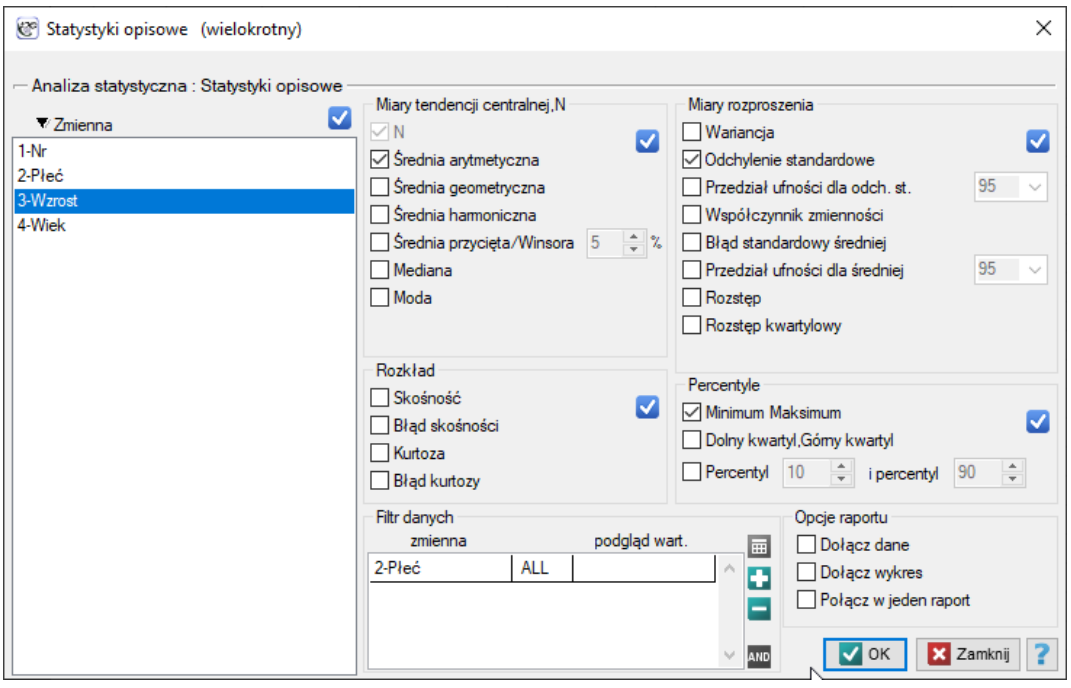
xor - operator alternatywy rozłącznej - sprawdza, czy dokładnie jeden z podanych warunków jest prawdziwy,
not - operator negacji,
empty(v1) – funkcja sprawdzająca czy występują puste komórki,
text(v1) – funkcja sprawdzająca czy w komórkach znajduje się tekst,
numer(v1) – funkcja sprawdzająca czy w komórkach znajduje się liczba.

Gotowy filtr możemy użyć wybierając przycisk Zastosuj, lub zachować na później wybierając przycisk . Zapisane filtry można wykorzystać przesyłając je do okna formuły poprzez przycisk .

PRZYKŁAD 8.7. *Filtr podstawowy zaawansowany* (plik filtr.pqs)
Chcemy wyznaczyć statystyki opisowe dla wzrostu chłopców, którzy są wyżsi niż 130 cm lub mają więcej niż 9 lat. Z menu wybieramy Statystyka→Analizy opisowe→Statystyki opisowe. W oknie opcji testu statystyk opisowych wybieramy procedury jakie chcemy wykonać (np. zaznaczamy średnią, odchylenie standardowe, minimum i maksimum) oraz zmienną do analizy (kolumnę ze wzrostem). Filtr ustawiamy wpisując formułę (v3>130 or v4>9) and v2='m'. Następnie wybieramy Zastosuj i wykonujemy analizę

- **Filtr wielokrotny** — nakłada jeden warunek na zebrane dane dzieląc je na kilka podgrup. Wybrana analiza jest wykonywana wielokrotnie, odrębnie dla każdej podgrupy.

  PRZYKŁAD 8.8. *Filtr wielokrotny* (plik filtr.pqs)
  Chcemy wyznaczyć statystyki opisowe dla wzrostu dziewczynek i oddzielnie dla wzrostu chłopców. Z menu wybieramy Statystyka→Analizy opisowe→Statystyki opisowe. W oknie opcji testu statystyk opisowych wybieramy procedury, jakie chcemy wykonać (np. zaznaczamy średnią, odchylenie standardowe, minimum i maksimum) oraz zmienną do analizy (kolumnę z wzrostem). Filtr wielokrotny ustawiamy dodając regułę przyciskiem i wybierając zmienną Płeć. Potwierdzamy wybrane opcje analizy przyciskiem OK. W rezultacie otrzymamy 2 raporty: oddzielny dla chłopców i oddzielny dla dziewczynek.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

8.3 WIELOKROTNE WYKONYWANIE ANALIZ

Aby usprawnić wielokrotne wykonywanie tych samych analiz można:

1. Wykorzystać możliwość zapamiętania bieżącej analizy i analiz wcześniejszych. Program PQStat automatycznie zapamiętuje ostatnio wykonywaną analizę i ustawione w jej oknie opcje. Możliwość szybkiego powrotu do niej daje przycisk  (menu wstążka -> ) umieszczony na pasku narzędzi.
2. Wykorzystać możliwość zapamiętania analizy na podstawie raportu. Program PQStat wraz z raportem analizy automatycznie zapamiętuje ustawione w oknie analizy opcje. Możliwość szybkiego powrotu do niej daje przycisk  umieszczony w menu kontekstowym prawego przycisku myszy na raporcie w Drzewie nawigacji projektu
3. Wybrać w oknie analizy wiele zmiennych tak, by analiza została przeprowadzona wielokrotnie. Wyniki dla każdej z przeprowadzonych analiz zostaną zwrócone w kolejnych raportach.
4. Skorzystać z filtru wielokrotnego aby analiza została przeprowadzona oddzielnie dla poszczególnych podzbiorów danych. Wyniki dla każdej z przeprowadzonych analiz zostaną zwrócone w kolejnych raportach.

8.4 ORGANIZOWANIE RAPORTÓW W ZESTAWY

Przy większej liczbie raportów wygodnie jest grupować je w zestawy. Przygotowane zestawy ułatwiają zarówno nawigację po wynikach, jak też przyspieszają przenoszenie wyników do innych programów, jak Excel czy Word – przenosimy wówczas całe zestawy raportów, a nie tylko pojedyncze raporty.

TWORZENIE ZESTAWÓW

Poprzez menu kontekstowe Utwórz pusty w Drzewie nawigacji na arkuszu danych lub istniejącym zestawie.

UMIESZCZANIE RAPORTÓW W ZESTAWACH

Raporty mogą być umieszczane w zestawie

- w momencie ich tworzenia – z poziomu analizy – w tym celu w oknie analizy należy wskazać przycisk  i wybrać zestaw, do którego ma trafić tworzony raport analizy, lub przed samą analizą wskazać kursorem w Drzewie nawigacji ten zestaw;
- po wykonanej analizie – poprzez opcję przeciągnij i upuść lub menu kontekstowe na nazwie raportu/raportów w Drzewie nawigacji, albo poprzez Zarządcę Projektu.

EDYCJA ZESTAWÓW

Poprzez menu kontekstowe w Drzewie nawigacji na arkuszu danych lub istniejącym zestawie – Edytuj zestawy.

USUWANIE ZESTAWÓW

Poprzez menu kontekstowe w Drzewie nawigacji na nazwie zestawu – Usuń zestaw (Shift+Del).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

8.5 INFORMACJE ZWRACANE DO RAPORTU

Poza podstawowymi ustawieniami dotyczącymi wykonywanych analiz statystycznych w oknie testu mamy możliwość:

- Dołączenia analizowanych danych do raportu.
  Dane poddane analizie w zależności od wykonywanego testu mogą być zwrócone do raportu:
  - w postaci surowych rekordów,
  - w postaci tabeli kontyngencji.

- Dołączenia wykresu do raportu.
  By odpowiedni wykres znalazł się w raporcie, w oknie wybranej analizy statystycznej zaznaczamy opcję Dołącz wykres.

- Połączenia wyników z wielu analiz w jeden raport.
  Poprzez wybranie wielu zmiennych lub poprzez zastosowanie wielokrotnego filtru danych można jednocześnie wykonać wiele analiz tego samego typu. By ułatwić przeglądanie wyników wszystkie wykonane w ten sposób analizy mogą zostać zwrócone do jednego raportu. W tym celu, w oknie analizy należy zaznaczyć opcję Połącz w jeden raport.

8.6 OZNACZANIE WYNIKÓW ISTOTNYCH STATYSTYCZNIE

W raporcie kolorem czerwonym oznaczona jest wartość p wykonanego testu statystycznego, która jest mniejsza od zadanego przez użytkownika poziomu istotności. Domyślnie dla wszystkich testów wybrany jest poziom istotności $\alpha = 0.05$. Ustawienie to można zmienić na stałe w oknie Ustawienia lub tymczasowo (do czasu zamknięcia aplikacji) w oknie wybranego testu.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

9 WYKRESY

Program PQStat pozwala tworzyć wykresy kolumnowe, wykresy błędów, wykresy typu ramka-wąsy, wykresy punktowe oraz liniowo-punktowe.

Okno z ustawieniami opcji wykresów wywołujemy poprzez menu Wykresy.

Podstawowe parametry wykresu można zmienić bezpośrednio w jego oknie. Poszczególne zakładki tego okna służą do innych celów:

- zakładka Ogólne opcje wykresu pozwala zmienić parametry ogólne, takie jak tytuły, tło, osie, linie siatki czy legendę;
- zakładka Właściwe opcje wykresu pozwala zmienić wygląd samego rysowanego obiektu, np. jego kształt, styl czy kolory;
- zakładka Inne pozwala dorysować na wykresie dodatkowe elementy, np. dodatkową linię.

Wykresy prezentujące wyniki analiz statystycznych dostępne są w oknie wybranej analizy statystycznej, pod opcją Dołącz wykres.

CO MOŻNA ZROBIĆ Z GOTOWYM WYKRESEM

Wykres trafia do raportu, gdzie z poziomu menu kontekstowego (kliknięcie prawym przyciskiem myszy na wykresie) można go:

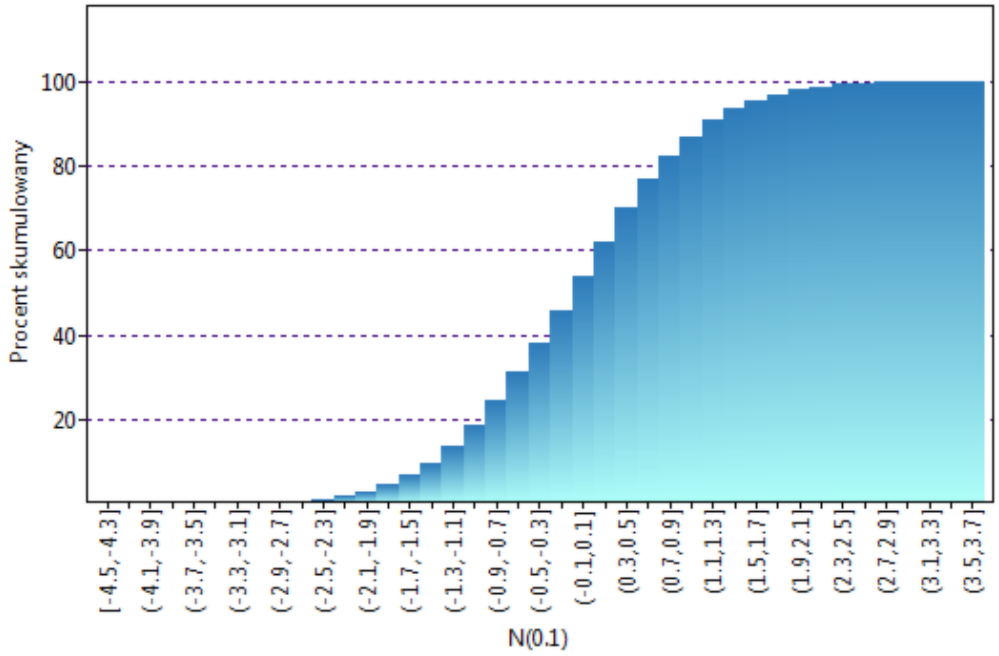
- zapisać – opcja Zapisz wykres jako...;
- wydrukować – opcja Drukuj wykres;
- skopiować – opcja Kopiuj wykres;
- edytować – dotyczy to Ogólnych opcji wykresu i Właściwych opcji wykresu. Aby edytować wykres, wystarczy podwójnie kliknąć na nim myszą albo wybrać z menu kontekstowego opcję Edytuj wykres. W oknie edycji wykresu można też zapisać wykres w wysokiej rozdzielczości.

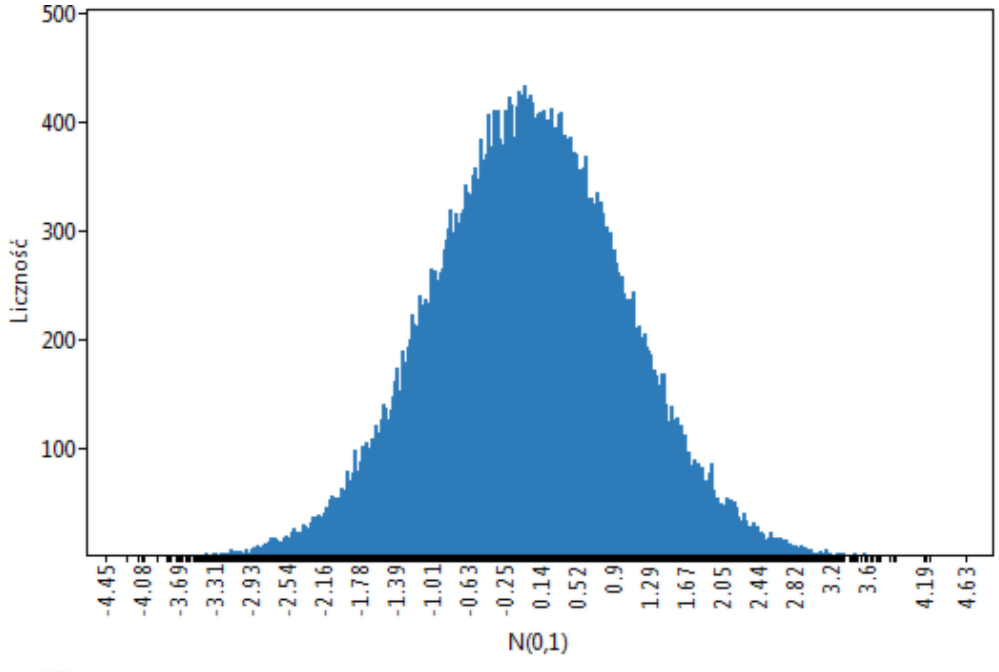
9.1 GALERIA WYKRESÓW

Rodzaj dostępnych wykresów zależy od wykonywanej analizy. Poniżej przedstawiono przykłady poszczególnych typów wykresów.

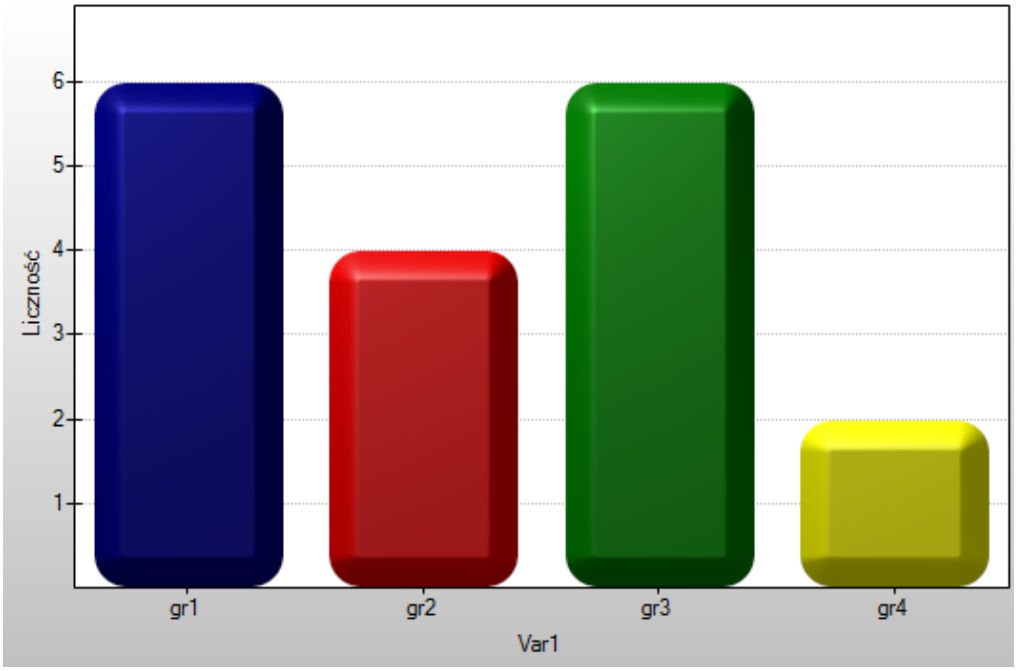
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

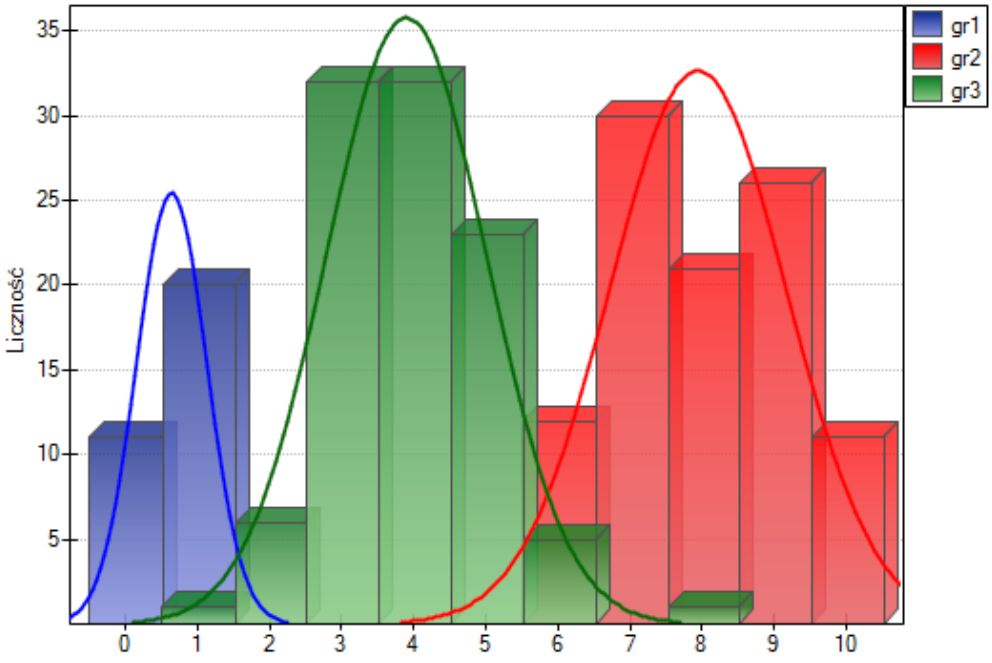
Wykresy kolumnowe



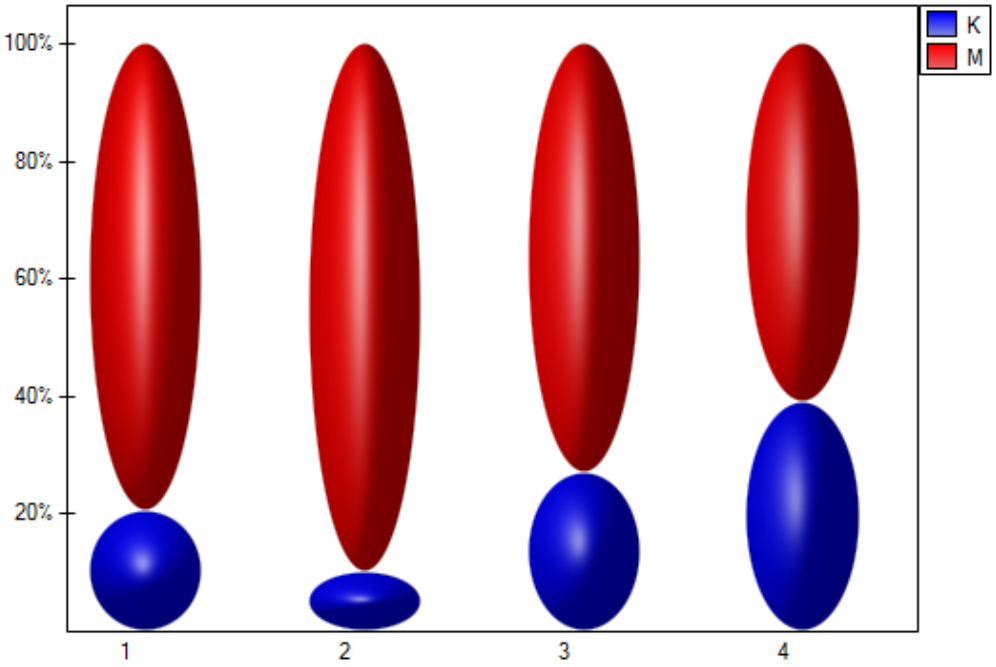


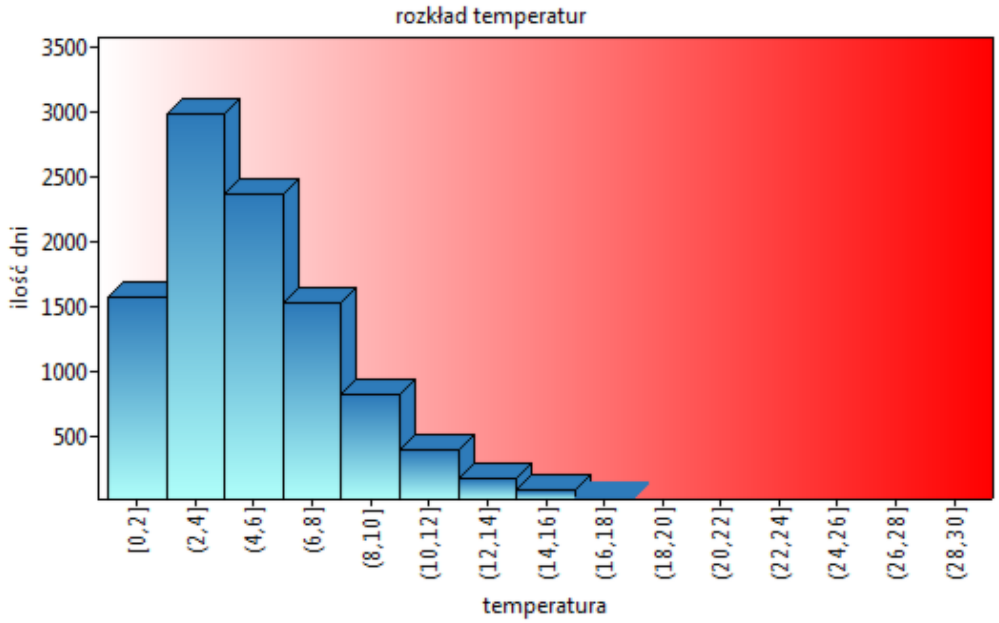
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



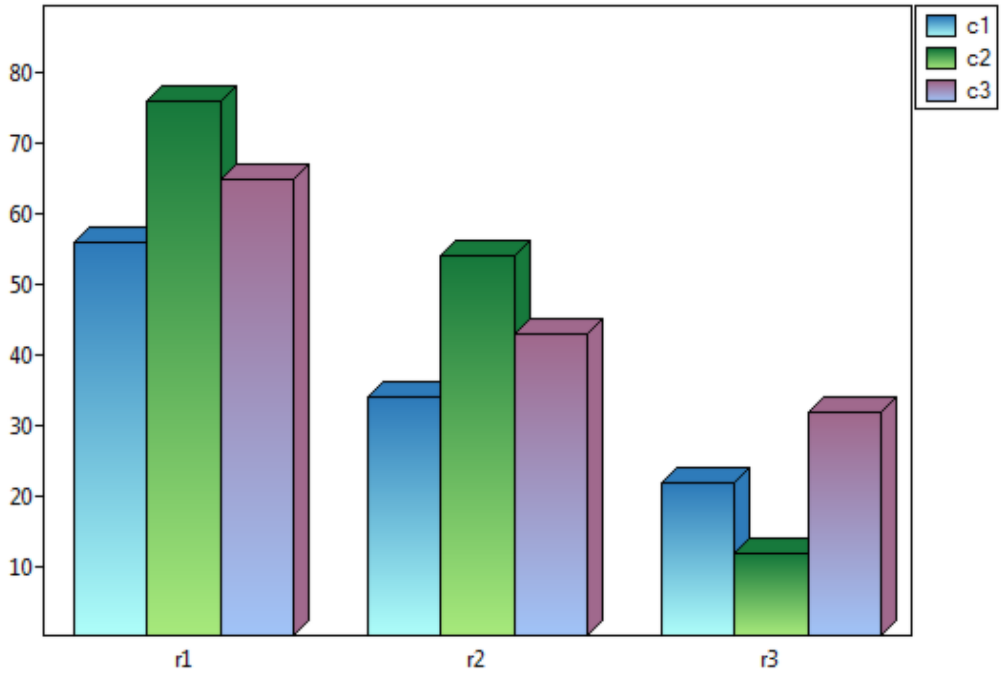


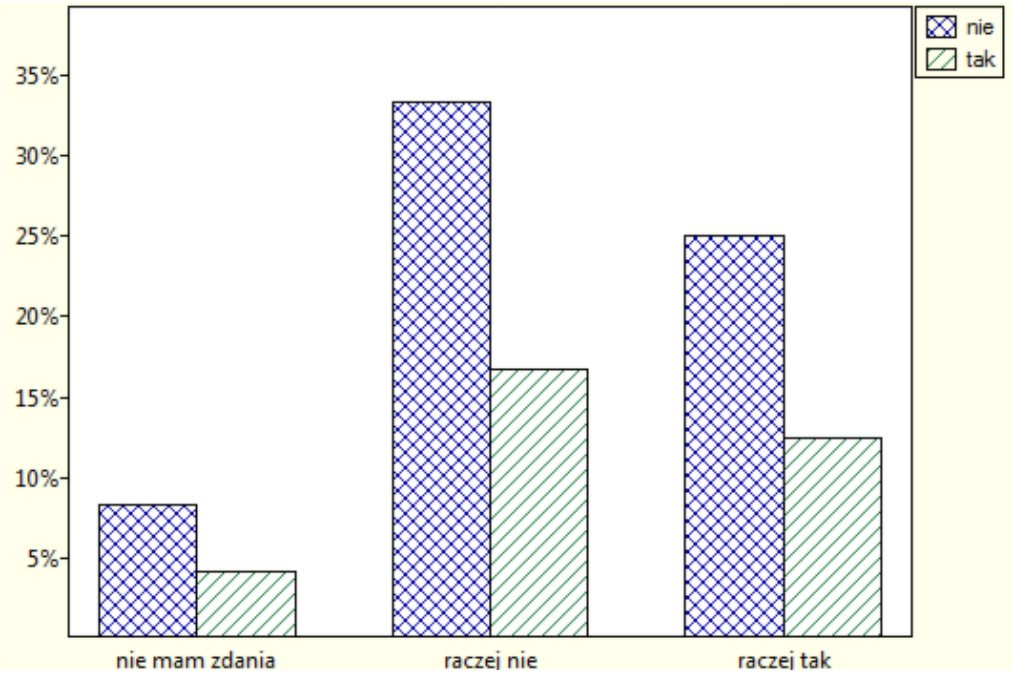
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



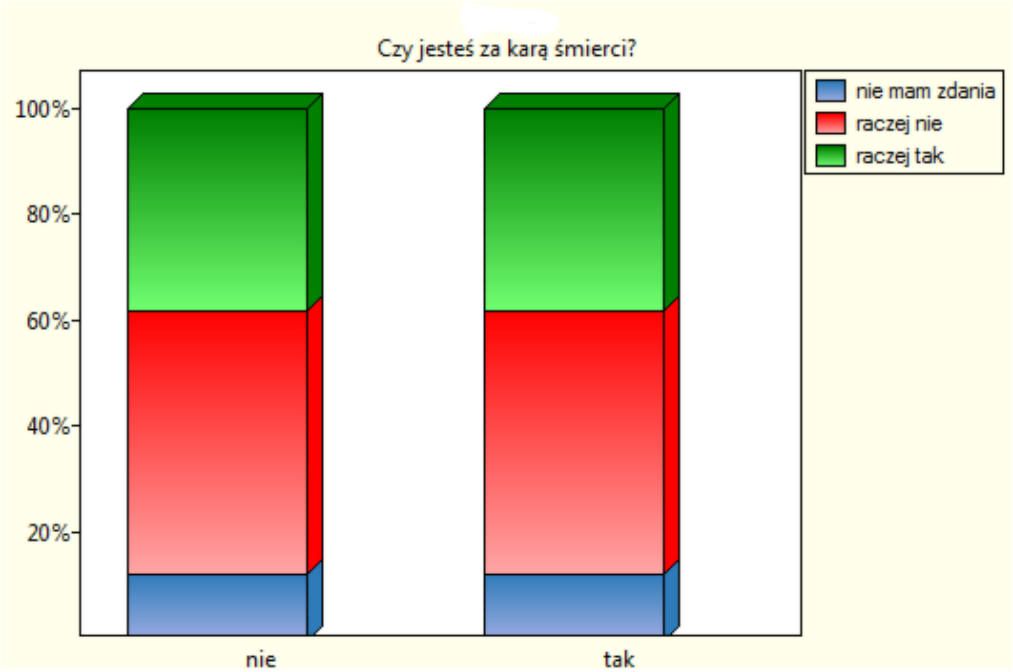


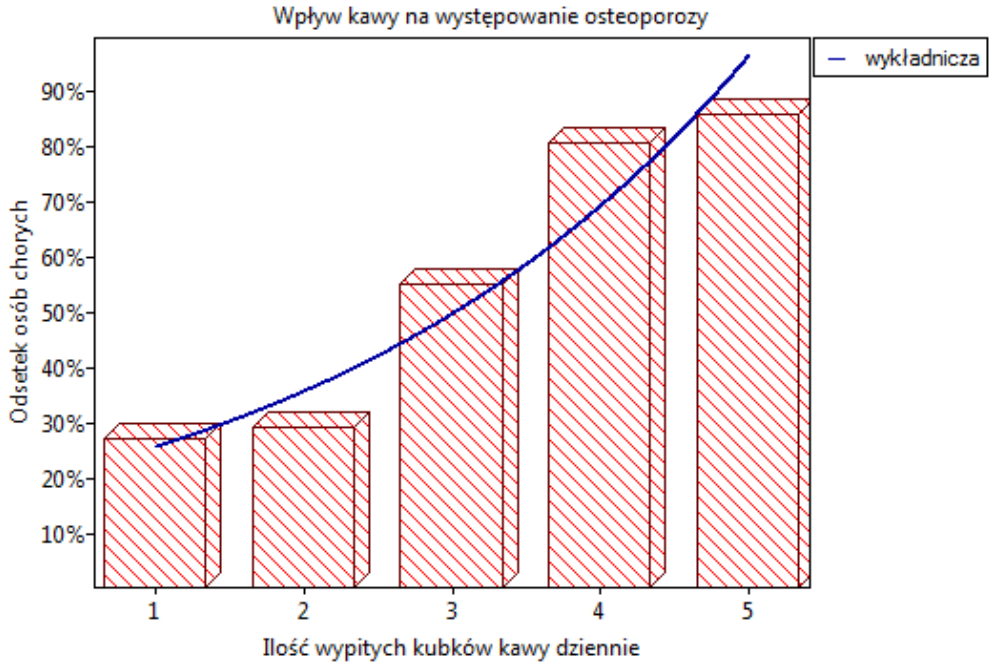
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





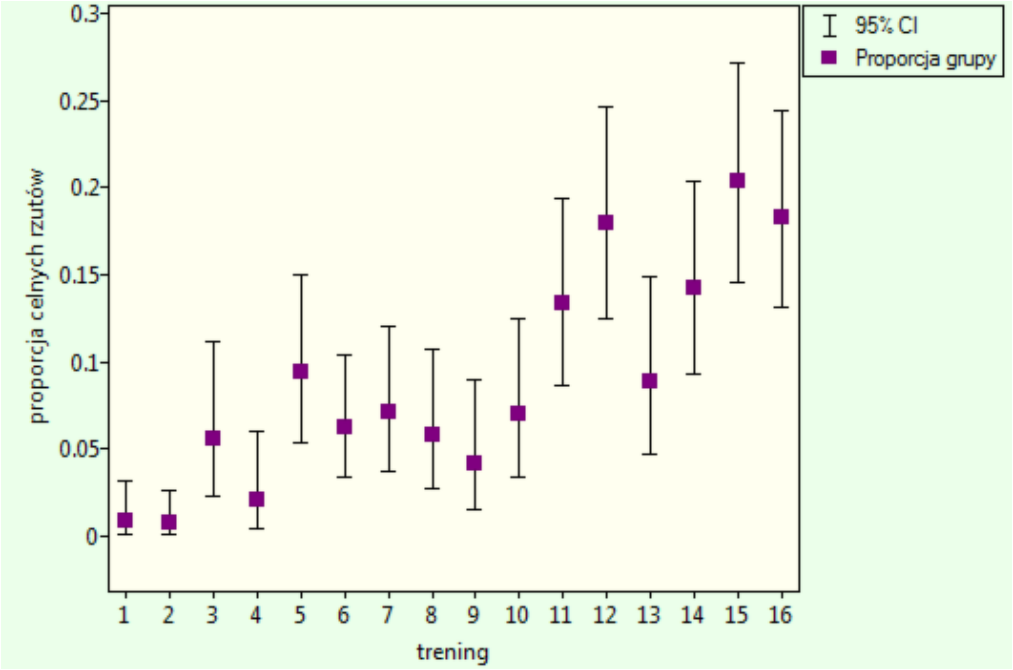
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

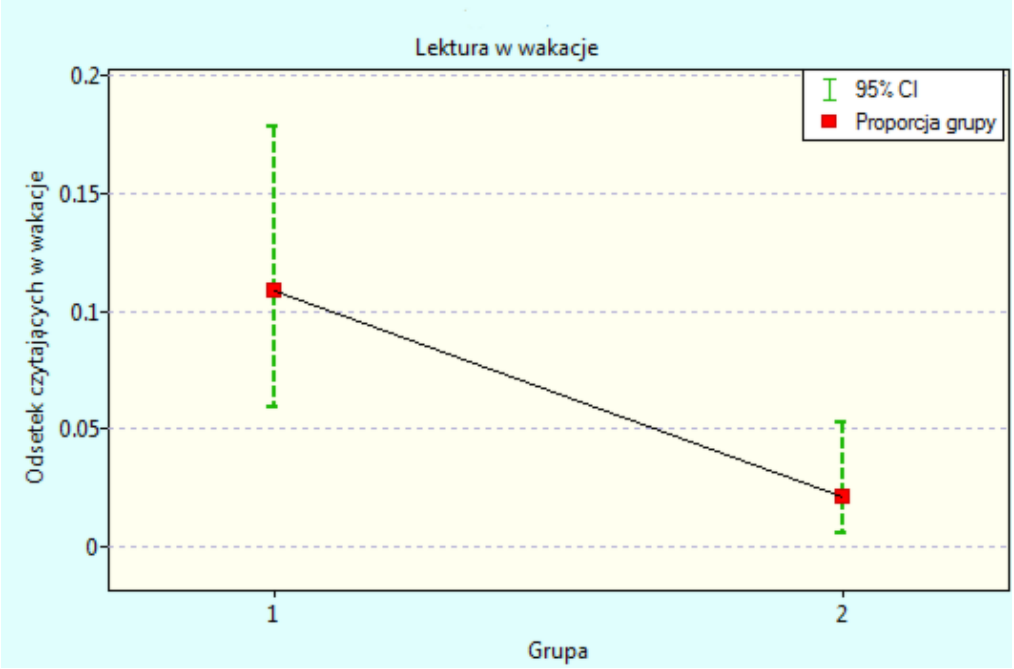




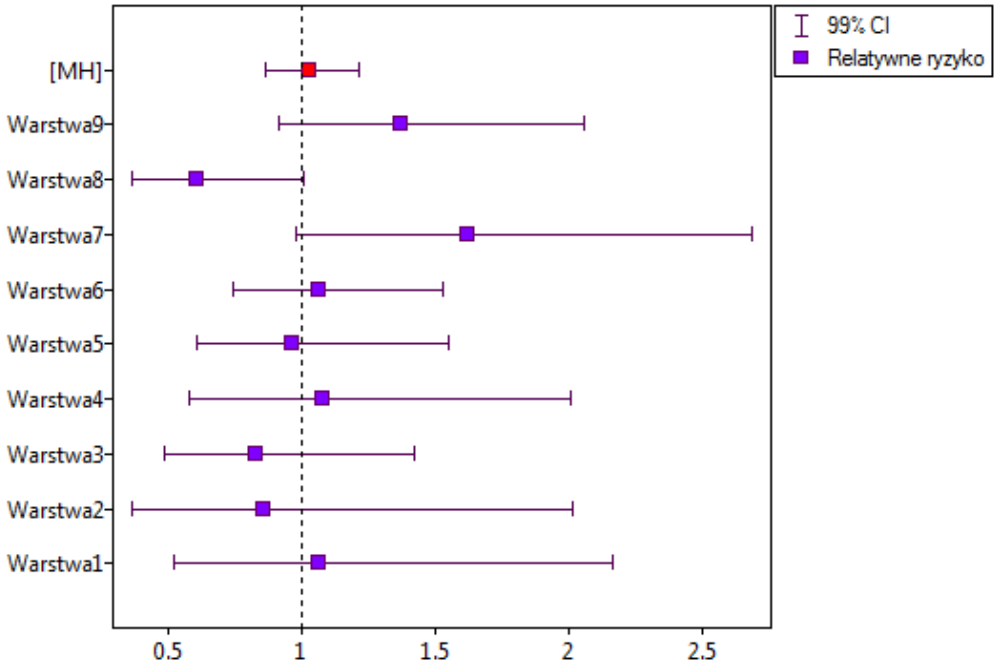
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wykresy błędów estymatora

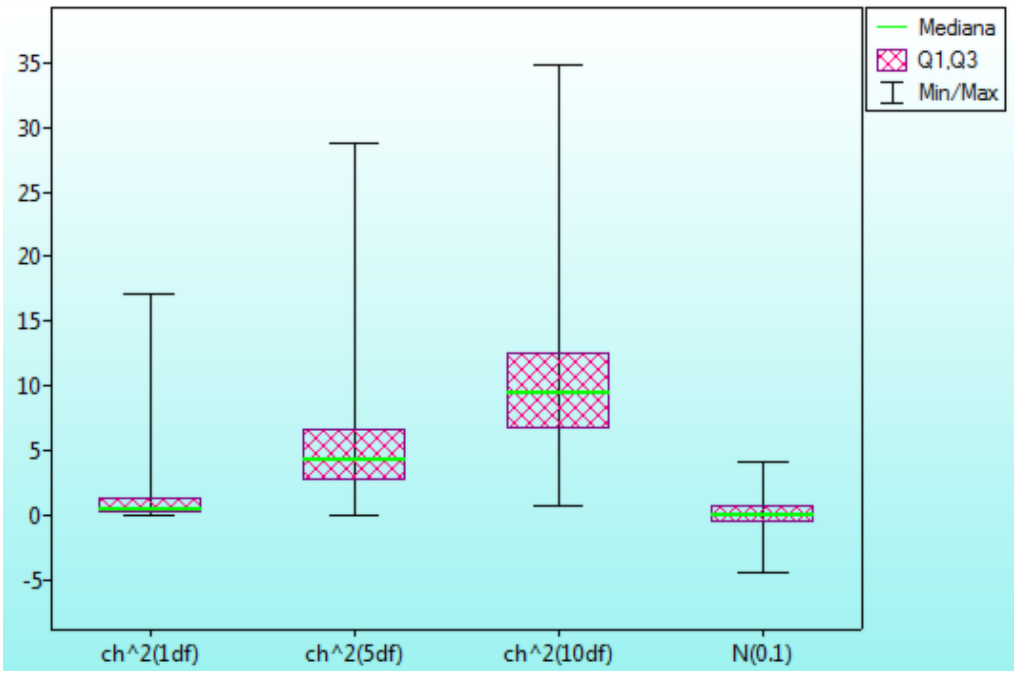




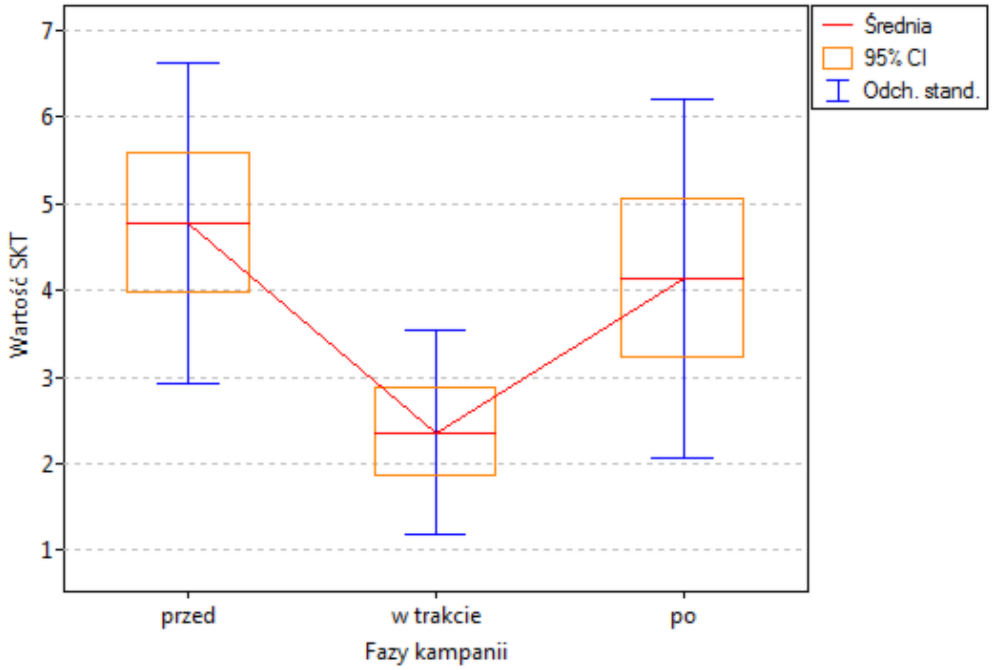
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

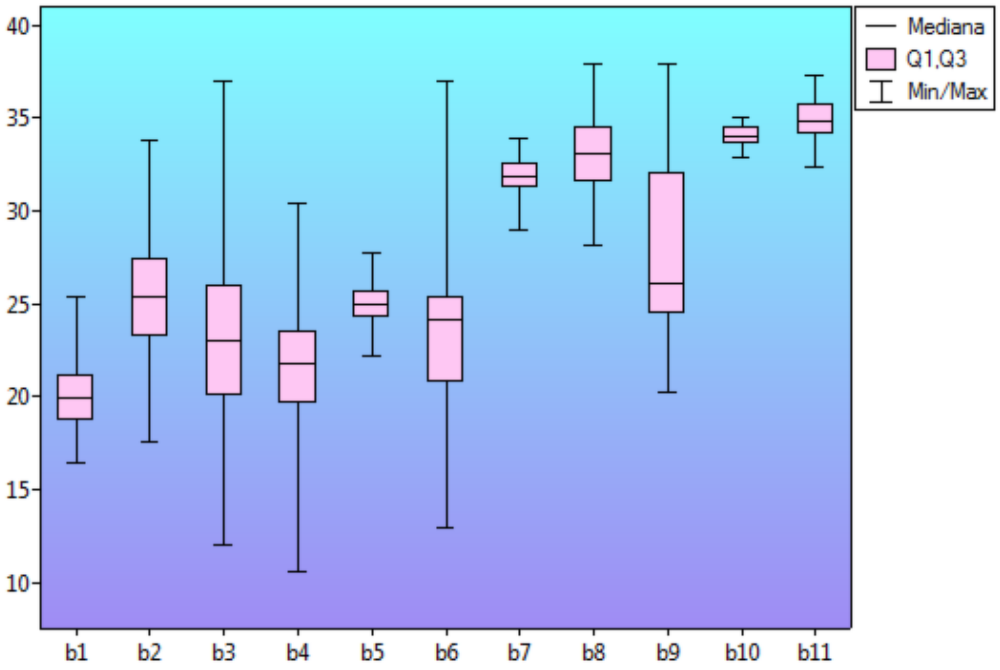


Wykresy typu ramka-wąsy



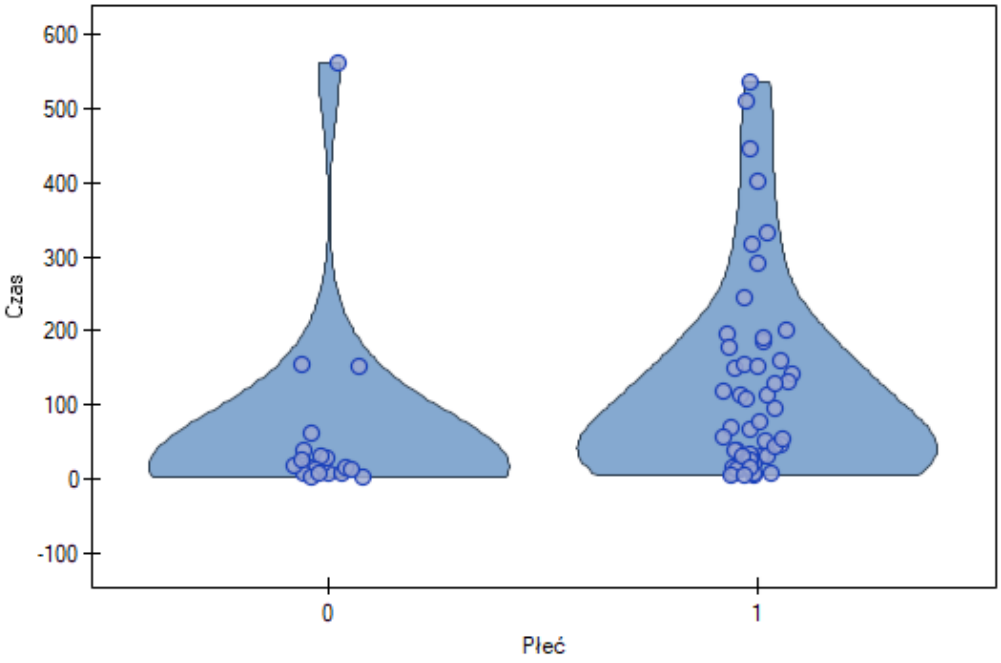
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

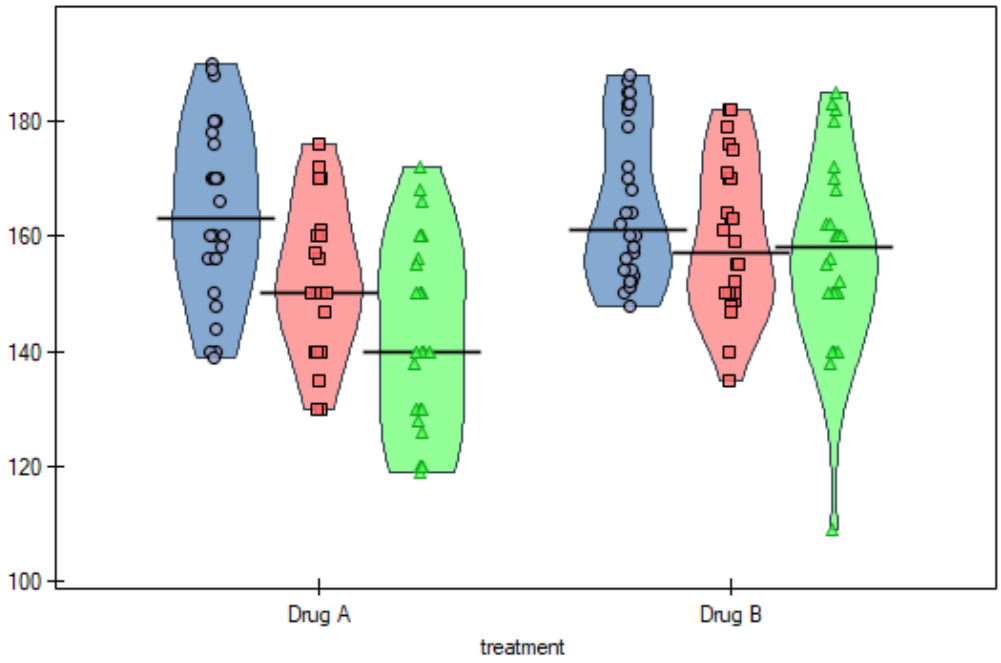




Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

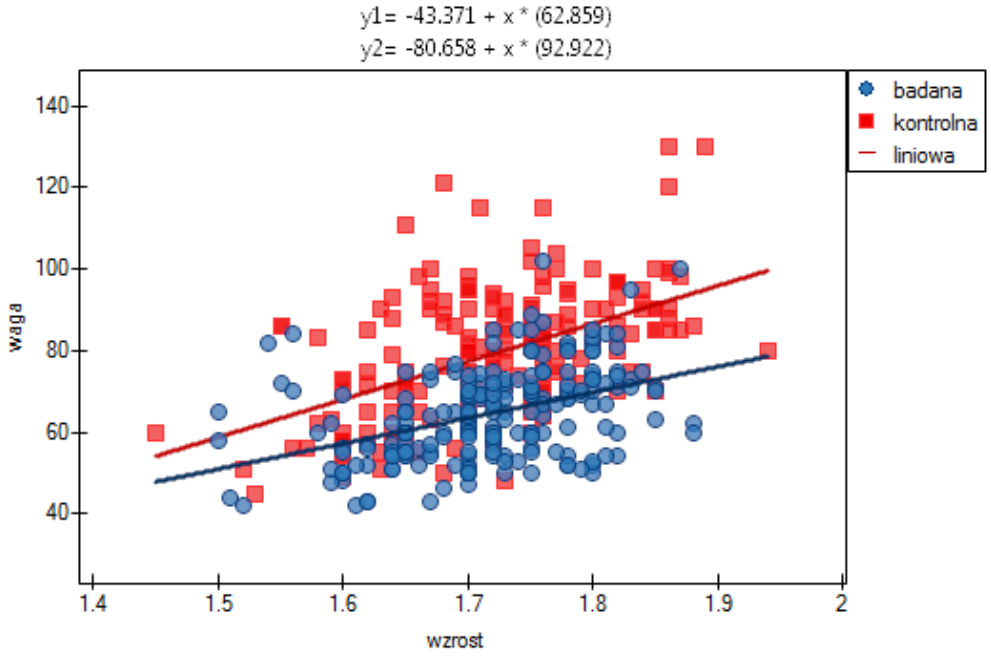
Wykresy typu skrzypcowy

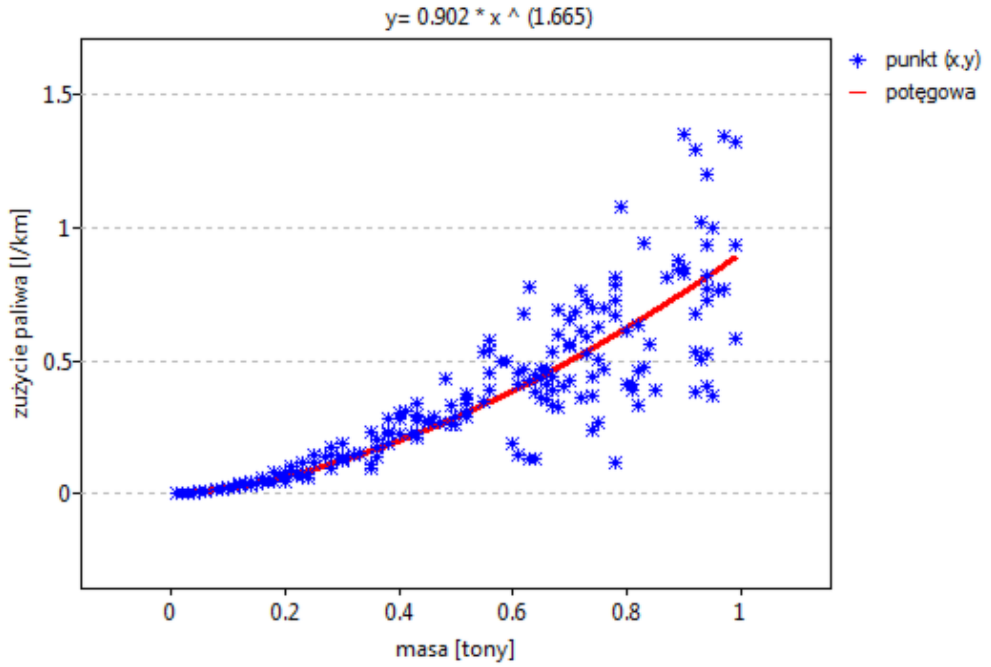




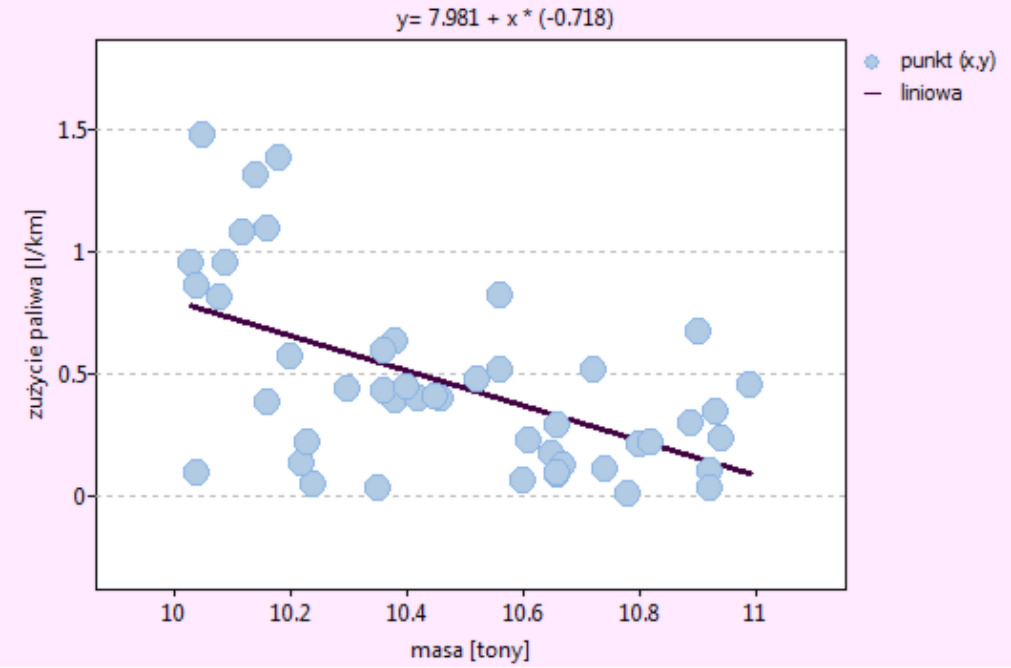
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

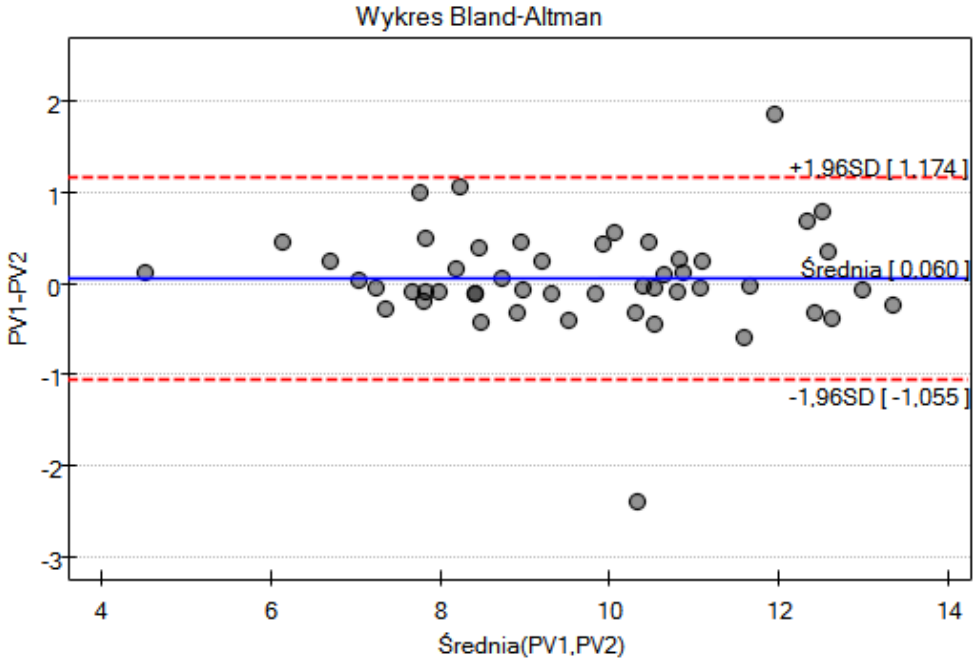
Wykresy punktowe





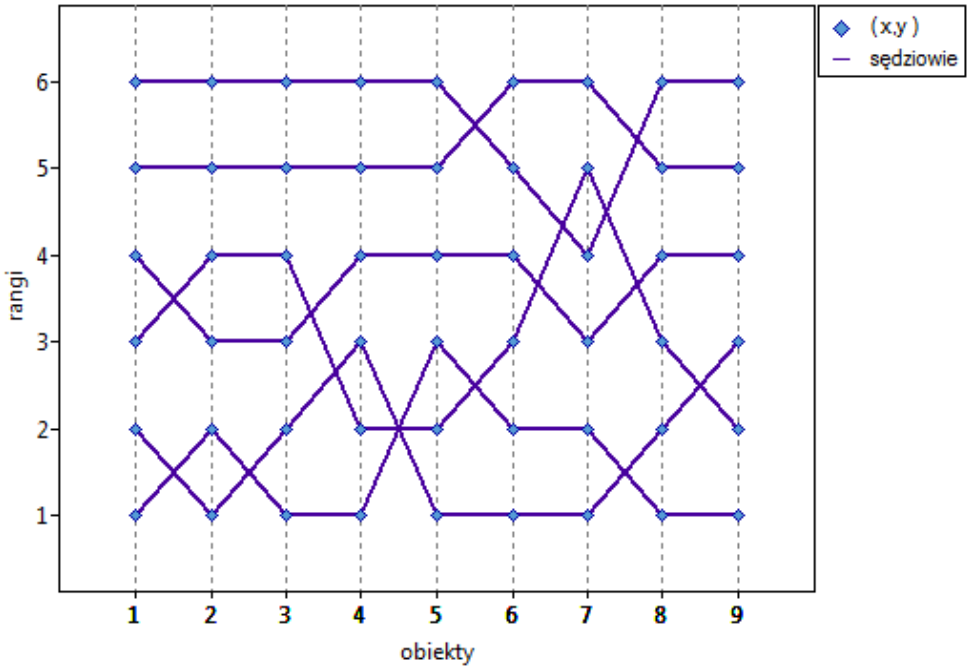
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

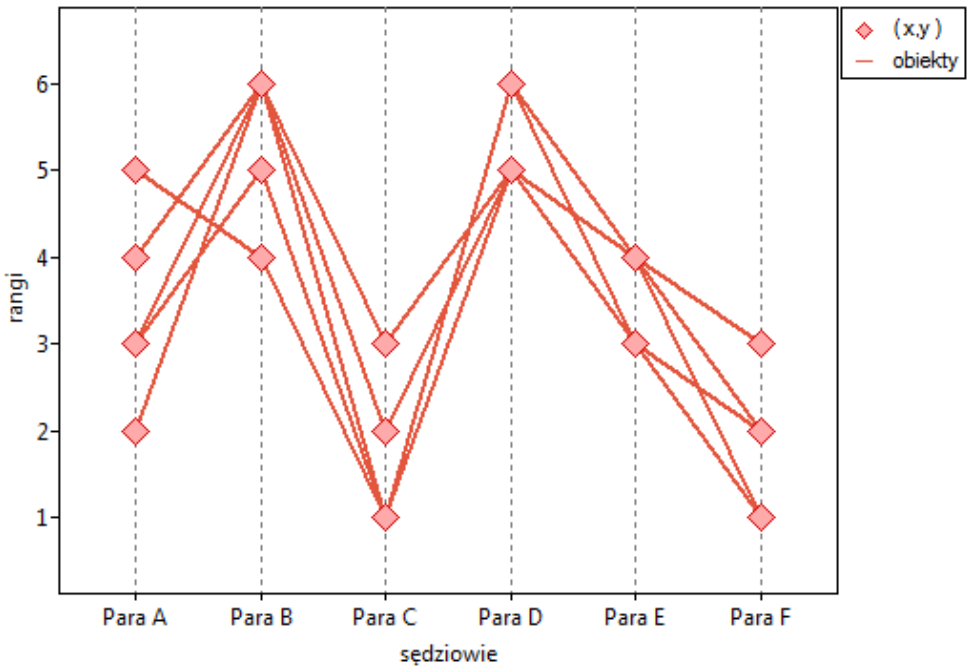




Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wykresy liniowe





Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

9.2 LOKALNE TECHNIKI WYGŁADZANIA LINIOWEGO

LOWESS

KIEDY STOSOWAĆ:

LOWESS *(ang. locally weighted scatterplot smoothing)*, znana również pod nazwą LOESS *(ang. locally estimated scatterplot smoothing)*, pozwala poprowadzić przez chmurę punktów na wykresie punktowym wygładzoną linię trendu, budowaną w sposób nieparametryczny (bez zakładania z góry konkretnej postaci zależności, np. liniowej czy wielomianowej). Jest to jedna z wielu „nowoczesnych" metod modelowania opartych na metodzie najmniejszych kwadratów, łącząca prostotę regresji liniowej z elastycznością regresji nieliniowej. Oszacowanie regresji lokalnej (ang. locally weighted regression) zostało niezależnie wprowadzone na kilku różnych polach pod koniec XIX i na początku XX wieku (Henderson, 1916[80]; Schiaparelli, 1866[150]). W literaturze statystycznej metodę tę wprowadzono niezależnie z różnych punktów widzenia pod koniec lat siedemdziesiątych, min. Cleveland, 1979[38] – i to właśnie na tej pracy oparto metodę wykorzystaną w programie PQstat.

Podstawową zasadą jest to, że funkcja gładka może być dobrze przybliżona przez wielomian niskiego stopnia (w programie zastosowano funkcję liniową, czyli wielomian pierwszego stopnia) w sąsiedztwie dowolnego punktu x. Innymi słowy: zamiast dopasowywać jedną funkcję do wszystkich danych naraz, w każdym punkcie dopasowuje się osobną, lokalną funkcję liniową, budowaną na podstawie sąsiednich obserwacji – im bliżej danego punktu, tym większy wpływ ma dana obserwacja na jego wynik. Szczegółowy algorytm postępowania, wraz ze wzorami, opisano w części „Dla zainteresowanych".

Wygładzanie jądrowe

KIEDY STOSOWAĆ:

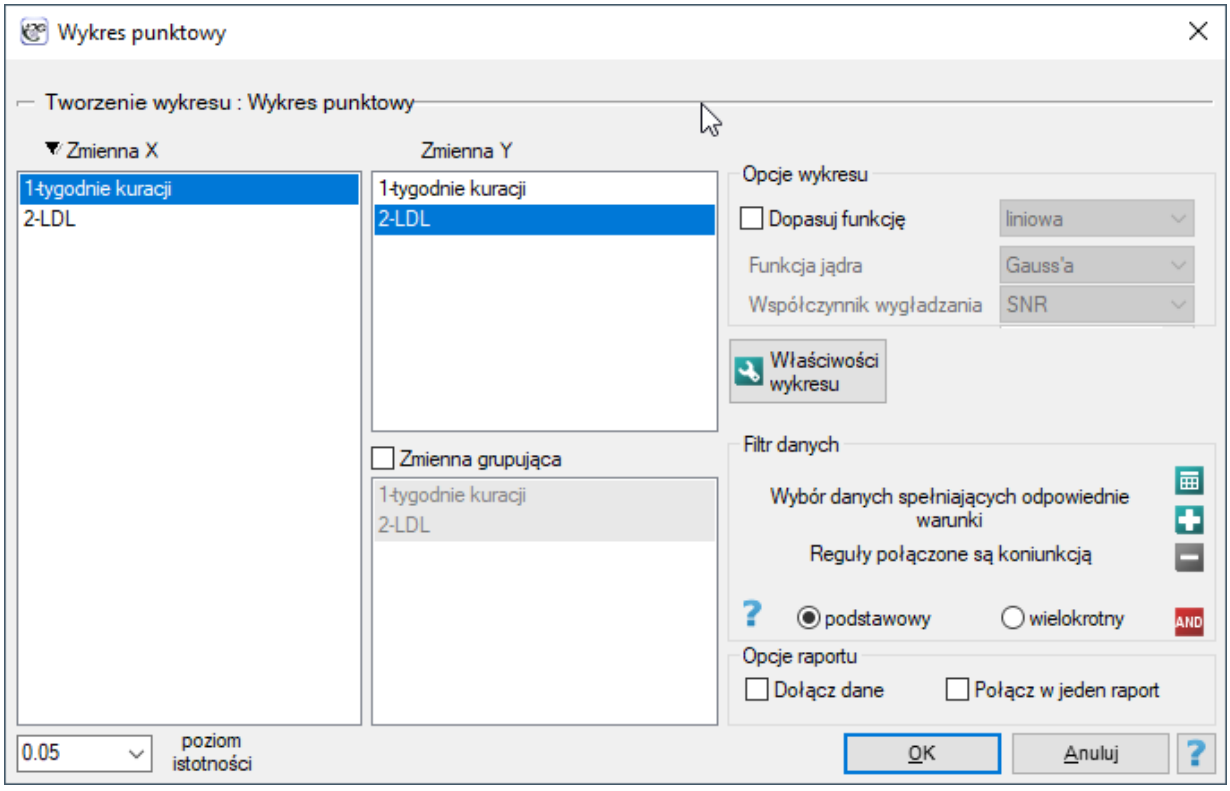
Alternatywnym sposobem wygładzenia wykresu punktowego jest wygładzanie jądrowe, nazywane czasem oszacowaniem Nadaraya-Watson (Nadaraya, 1964[126]; Watson, 1964[171]). Podobnie jak LOWESS, nie zakłada ono z góry konkretnej postaci zależności – w każdym punkcie wyliczana jest średnia ważona sąsiednich obserwacji, przy czym wagi nadawane są zgodnie z funkcją jądra opisaną w rozdziale Estymacja jądrowa. Dokładny wzór podano w części „Dla zainteresowanych".

Na kształt uzyskanej krzywej decydujący wpływ ma **parametr wygładzania** h *(ang. bandwidth)* – im wyższa jego wartość, tym większy stopień wygładzenia. Parametr ten można ustawić dowolnie (wartość użytkownika) albo dobrać automatycznie jedną z metod: SNR, SROT lub OS. Znacznie mniejszy wpływ na wynik ma wybór **funkcji jądra**: Gaussa, jednostajnej (prostokąt), trójkątnej, Epanechnikova lub quartic/biweight (czwartego stopnia). Opis poszczególnych parametrów wygładzania i funkcji jądra można znaleźć we wspomnianym rozdziale Estymacja jądrowa.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji wykresu punktowego z dopasowaną funkcją metodą LOWESS lub poprzez wygładzanie jądrowe znajdziemy w różnych analizach. Wykres ten możemy również wykonać poprzez menu Wykresy→Wykres punktowy.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

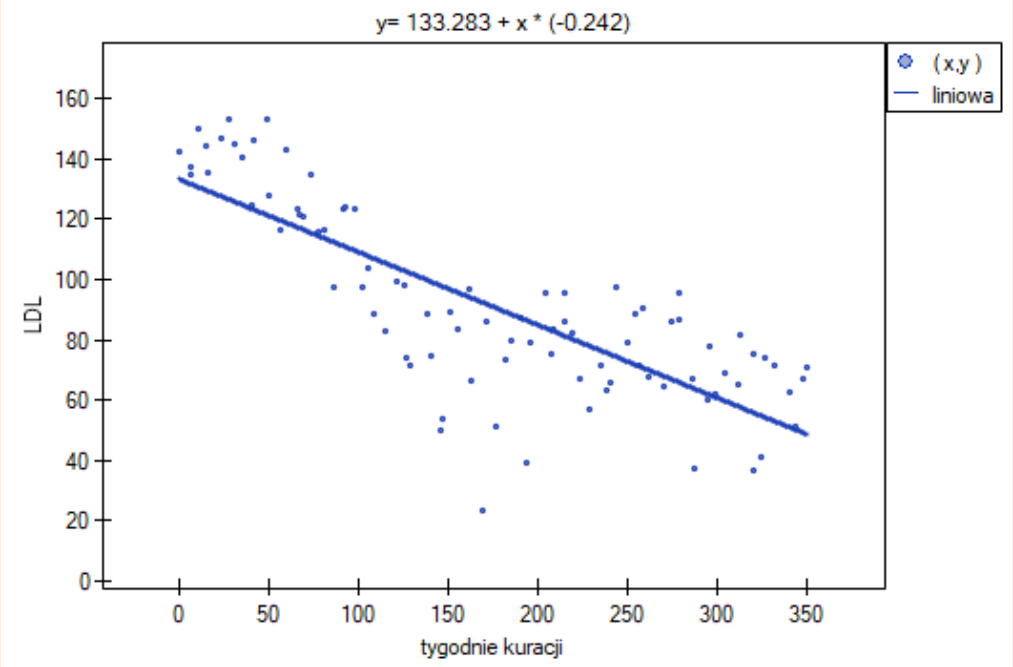


WYKONANIE I INTERPRETACJA

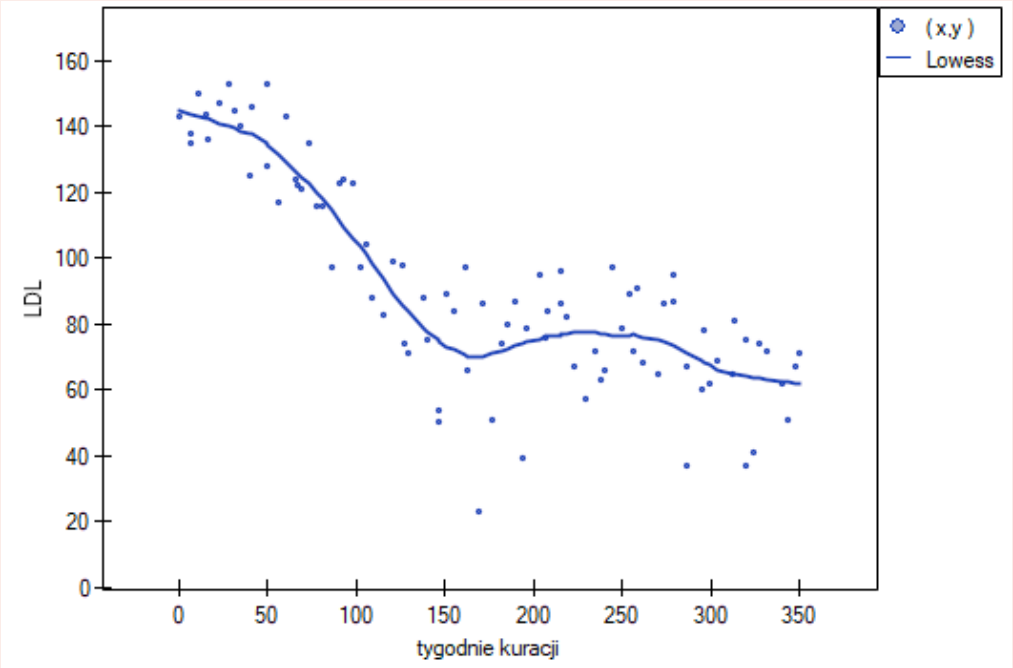
PRZYKŁAD (22.2) c.d. *(plik LDL tygodnie.pqs)*
Badano skuteczność nowej terapii, której celem jest obniżenie poziomu cholesterolu we frakcji LDL. Przebadano 88 osób na różnym etapie kuracji. Sprawdzimy, czy wraz z upływem czasu stosowania kuracji (czas w tygodniach) poziom cholesterolu LDL spada i się stabilizuje.

Wyniki przedstawiono początkowo, dopasowując linię prostą wskazującą kierunek badanej zależności.

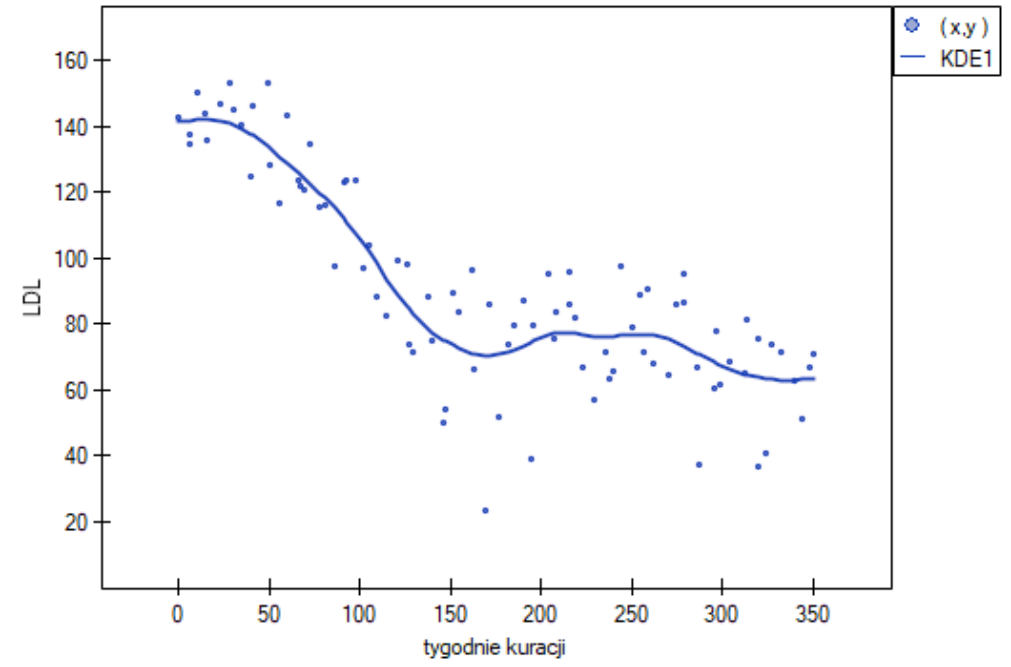
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Jednak ten sposób przedstawienia danych nie oddaje w pełni zachodzących zależności. Z ułożenia punktów można zauważyć, że zależność ta jest początkowo malejąca, a po 150 tygodniach zaczyna się stabilizować. Zależność przedstawiono ponownie, korzystając z metody LOWESS oraz z wygładzania jądrowego Gaussa.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Obydwa sposoby, tzn. zarówno LOWESS, jak i wygładzanie jądrowe, dały zbliżone wyniki i znacznie lepiej opisały dane, wskazując na początkowy spadek LDL, a następnie jego stabilizację na poziomie bliskim 70 mg/dl.

DLA ZAINTERESOWANYCH

Algorytm LOWESS
Procedura:

1. Dla każdego punktu zbioru danych budujemy okno zawierające sąsiednie elementy. To, jak dużo elementów znajdzie się w oknie, wyznacza parametr wygładzania q. Czym wyższa jego wartość, tym bardziej wygładzoną uzyskamy funkcję. Jeśli parametr ten wyniesie np. 0.2, wówczas w oknie znajdzie się około 20% danych i one posłużą do budowania wielomianu (tutaj jednomianu, tzn. funkcji liniowej).
   Ze względu na konieczność utrzymania symetrii okna, tzn. dane w oknie powinny znajdować się powyżej i poniżej punktu, dla którego budujemy model, liczba elementów w oknie powinna być nieparzysta. Dlatego też uzyskana na podstawie parametru q liczba elementów zaokrąglana jest w górę do pierwszej liczby nieparzystej. W ten sposób powstaje okno zawierające element x_0 oraz odpowiednią liczbę $k(x_0)$ elementów znajdujących się przed tym elementem i za tym elementem w uporządkowanym zbiorze danych.

$$k(x_0) = \frac{roundUpNpar(q * n) - 1}{2}$$

   Przykładowo, gdy w oknie ma się znaleźć 7 elementów, to znajdzie się w nim element x_0 oraz trzy wcześniejsze i trzy późniejsze elementy próby. Okno wyznaczone dla pierwszych i ostatnich elementów próby jest tej samej wielkości, ale badany element nie jest w nim umieszczony symetrycznie, tzn. na środku.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

2. W każdym punkcie zbioru danych dopasowany jest wielomian niskiego stopnia (tutaj funkcja liniowa) do podzbioru zlokalizowanych w oknie danych. Dopasowanie dokonywane jest przy użyciu ważonej metody najmniejszych kwadratów, co daje większą wagę punktom w pobliżu punktu, którego odpowiedź jest szacowana, i mniejszą wagę punktom oddalonym. W ten sposób do każdego punktu przypisana jest inna formuła funkcji wielomianu (tutaj funkcji liniowej).
   Stosowane w ważonej metodzie najmniejszych kwadratów wagi mogą być ustawiane dość elastycznie, ale punkty odległe od zadanego muszą mieć mniejszą wagę niż punkty w pobliżu. Tutaj posłużono się wagami zaproponowanymi przez Clevelanda, tzw. tricube

$$w = (1 - d^3)^3,$$

   gdzie odległość d od punktu x_0 wyniosła 0 – dla punktów znajdujących się poza wyznaczonym oknem i była podana jako rzeczywista odległość między punktami, ale znormalizowana do przedziału [0,1] – dla punktów zlokalizowanych w oknie, tak by maksymalne odległości we wszystkich oknach były takie same.

3. W każdym punkcie, na podstawie przypisanej do niego formuły wielomianu (tutaj funkcji liniowej), obliczana jest wartość funkcji $\hat{y}_i$. Na podstawie punktów x_i oraz estymowanych tą metodą punktów $\hat{y}_i$ wyrysowywana jest wygładzona i dopasowana do danych funkcja.

Wygładzanie jądrowe
Wzory:
Szacowanie jądra jest średnią ważoną obserwacji w oknie wygładzania:

$$\hat{y}_i = \frac{\sum_{i=1}^{2} K_h(t_i) y_i}{\sum_{i=1}^{2} K_h(t_i)},$$

gdzie $K_h(t_i)$ jest funkcją jądra opisaną w rozdziale Estymacja jądrowa.

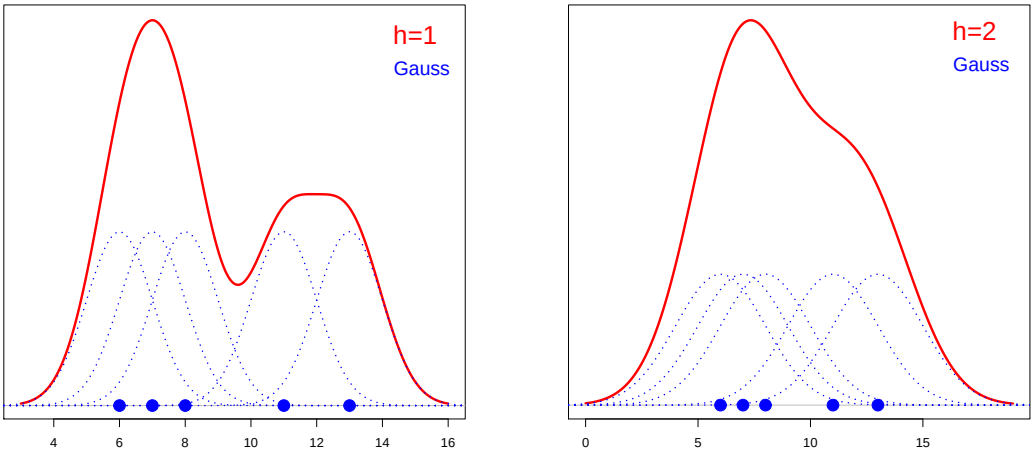
9.3 ESTYMACJA JĄDROWA

Jednowymiarowy estymator jądrowy

KIEDY STOSOWAĆ:

Jednowymiarowy jądrowy estymator gęstości pozwala przybliżyć kształt rozkładu danych za pomocą wygładzonej krzywej, budowanej w sposób nieparametryczny (bez zakładania konkretnej postaci rozkładu). Jest to alternatywa dla tradycyjnego histogramu – tam, gdzie histogram przedstawia rozkład za pomocą schodkowych słupków (co zależy od arbitralnie wybranej szerokości przedziałów), estymator jądrowy daje gładką, ciągłą krzywą, dzięki czemu lepiej oddaje faktyczny kształt rozkładu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Dla każdego punktu x z zakresu danych wyznaczana jest wartość estymatora, czyli lokalna gęstość w tym punkcie. Powstaje ona przez zsumowanie wartości funkcji jąder $K_h(t_i)$ zbudowanych wokół poszczególnych obserwacji:

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(t_i)$$

Jeśli poszczególnym przypadkom nadamy wagi w_i, można zbudować ważony jądrowy estymator gęstości, definiowany wzorem:

$$\hat{f}_K(x) = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i K_h(t_i)$$

DEFINICJE MIAR

Na kształt uzyskanej krzywej wpływają dwa elementy: dobór **współczynnika wygładzania** oraz wybór **funkcji jądra**.

Współczynnik wygładzania (h) *(ang. bandwidth)* decyduje o stopniu wygładzenia krzywej – im wyższa jego wartość, tym mocniej wygładzony (bardziej „spłaszczony") jest uzyskany estymator. Współczynnik ten może być ustalony na kilka sposobów:

Użytkownika – dowolna, dodatnia wartość współczynnika wygładzania wskazana samodzielnie przez użytkownika.

Użytkownika skalowane – wartość ustalana tak, by przy zmianie funkcji jądra zachować podobny poziom wygładzenia, jaki wybrano wcześniej dla jądra Gaussa. Dzięki przeskalowaniu współczynnika (Scott, D. W. 1992[151]) można wygodnie przełączać się między różnymi jądrami bez ponownego dobierania wygładzenia. Przeliczenia oparte są na odchyleniu standardowym:

$$h_2 = \frac{\sigma(K_{h_1})}{\sigma(K_{h_2})} h_1$$

SNR – współczynnik wyznaczany automatycznie metodą Silverman'a (Silverman B. W. 1986[156] str. 45 i 47) z korektą Jones'a (Jones M. C. i inni 1996[91]), wykorzystującą odchylenie standardowe z próby (a nie z populacji), zgodnie z propozycją Silvermana:

$$h_{SNR} = 1.06 sd \cdot n^{1/5}$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Dla jądra innego niż Gaussa współczynnik wygładzania podlega dodatkowo skalowaniu (Scott D. W., 1992[151]).

SROT – współczynnik wyznaczany automatycznie metodą Silverman'a (Silverman B. W. 1986[156] str. 48) z korektą Jones'a (Jones M. C. i inni 1996[91]):

$$h_{SROT} = 0.9 \min\left(sd, \frac{IQR}{1.34}\right) n^{1/5}$$

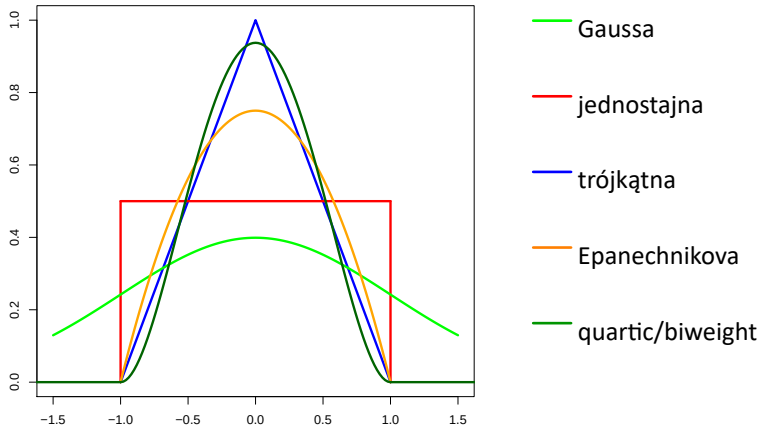
Dla jądra innego niż Gaussa współczynnik wygładzania podlega dodatkowo skalowaniu (Scott D. W., 1992[151]).

OS – współczynnik wyznaczany automatycznie metodą Terrell'a i Scott'a (Terrell G. R. i Scott D. W. 1985[165], Terrell G. R. 1990[164] str. 470):

$$h_{OS} = 1.144 sd \cdot n^{1/5}$$

Dla jądra innego niż Gaussa współczynnik wygładzania podlega dodatkowo skalowaniu (Scott D. W., 1992[151]).

Funkcja jądra ma mniejszy wpływ na wynik niż współczynnik wygładzania, ale również kształtuje uzyskany estymator. Jądro jest funkcją gęstości prawdopodobieństwa budowaną wokół każdego punktu danych x_i – zwykle jest to funkcja symetryczna, osiągająca maksimum w punkcie x_i i malejąca wraz ze wzrostem odległości d_i od tego punktu. Odległość od analizowanego punktu jest modyfikowana przez współczynnik wygładzania h zgodnie ze wzorem $t_i = \frac{d_i}{h}$. W zależności od potrzeb funkcja jądra może przyjmować postać funkcji Gaussa, jednostajnej (prostokąt), trójkątnej, Epanechnikova lub quartic/biweight (czwartego stopnia) – dokładne wzory znajdują się w części „Dla zainteresowanych".



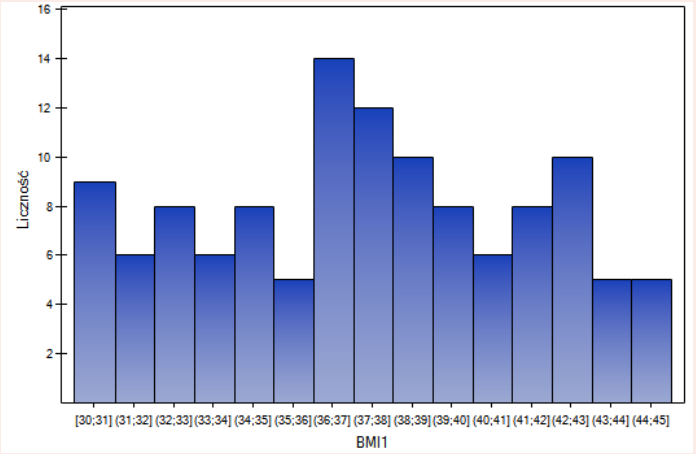
JAK USTAWIĆ ANALIZĘ

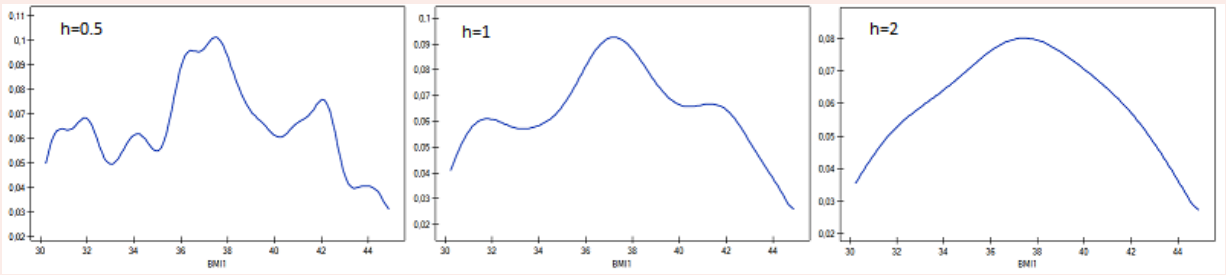
Jądrowy estymator gęstości dostępny jest jako opcja graficzna, wykorzystywana przy okazji wielu różnych analiz opisujących rozkład zmiennej (m.in. przy tworzeniu histogramu). W oknie ustawień takiego wykresu można wybrać funkcję jądra oraz sposób dobrania współczynnika wygładzania (wartość użytkownika, wartość użytkownika skalowana lub jedną z automatycznych metod: SNR, SROT, OS).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

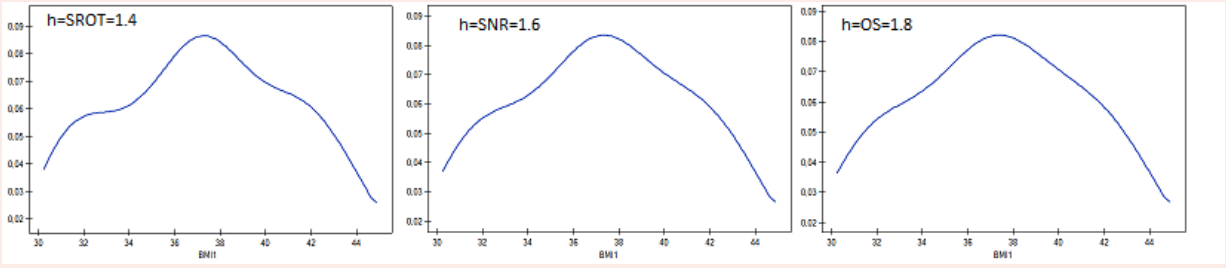
WYKONANIE I INTERPRETACJA

PRZYKŁAD 9.1. (plik BMI.pqs)
Wyliczono wartości współczynnika wagowo-wzrostowego BMI1 dla pewnej grupy osób otyłych. Ich rozkład przedstawiono najpierw za pomocą histogramu z podziałem wartości co 1 jednostkę BMI. Te same dane zobrazowano również za pomocą jądrowego estymatora gęstości, wybierając Gaussowską funkcję jądra i ustawiając współczynniki wygładzania odpowiednio: 0.5, 1, 2.





Widać, że im wyższy współczynnik wygładzania, tym gładsza (mniej „poszarpana”) jest uzyskana krzywa gęstości, co jednak może maskować drobniejsze cechy rozkładu, np. jego wielomodalność. Współczynniki wygładzania sugerowane przez automatyczne metody SROT, SNR i OS osiągają dla tych danych wartości pomiędzy 1.4 a 2.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Funkcja jądra Gaussa

$$K_h(t_i) = \frac{1}{h\sqrt{2\pi}} \exp(-\frac{t_i^2}{2})$$

Funkcja jądra jednostajna (prostokąt)

$$K_h(t_i) = \begin{cases} \frac{0,5}{h} & \text{jeżeli } t_i < 1 \\ 0 & \text{jeżeli } t_i \geq 1 \end{cases}$$

Funkcja jądra trójkątna

$$K_h(t_i) = \begin{cases} \frac{1-t_i}{h} & \text{jeżeli } t_i < 1 \\ 0 & \text{jeżeli } t_i \geq 1 \end{cases}$$

Funkcja jądra Epanechnikova

$$K_h(t_i) = \begin{cases} \frac{3}{4}\frac{1-t_i^2}{h} & \text{jeżeli } t_i < 1 \\ 0 & \text{jeżeli } t_i \geq 1 \end{cases}$$

Funkcja jądra quartic lub biweight (czwartego stopnia)

$$K_h(t_i) = \begin{cases} \frac{15}{16}\frac{(1-t_i^2)^2}{h} & \text{jeżeli } t_i < 1 \\ 0 & \text{jeżeli } t_i \geq 1 \end{cases}$$

Wykres Blanda-Altmana

KIEDY STOSOWAĆ:

Wykres Blanda-Altmana *(ang. Bland-Altman plot)* (Bland i Altman, 1986[22], 1999[8]) stosuje się do graficznej oceny zgodności dwóch metod pomiaru tej samej wielkości, np. dwóch różnych ciśnieniomierzy, dwóch analizatorów laboratoryjnych czy metody nowej i dotychczas stosowanej (tzw. gold standard). W medycynie klinicznej pomiary wykonywane na żywym organizmie ulegają ciągłym zmianom, a ich prawdziwa wartość jest nieznana (np. ciśnienie tętnicze krwi), co wymaga stałego doskonalenia i tworzenia nowych, lepszych narzędzi pomiaru. Zwykle, kiedy powstaje nowa metoda, porównuje się jej wyniki z metodą już uznaną. Nie można przy tym oczekiwać, że nowa metoda da dokładnie taki sam wynik jak metoda dotychczas stosowana, ale warto sprawdzić, na ile uzyskane wyniki się różnią. Aby zastąpić starą metodę nową, różnica pomiędzy wynikami obu metod powinna być na tyle mała, by nie stwarzać problemu w interpretacji klinicznej – np. w pomiarach ciśnienia tętniczego różnica 20mmHg jest już na tyle duża, że może wpłynąć na decyzję o podjęciu leczenia. Metody statystyczne nie odpowiedzą wprost na pytanie, jak duża różnica metod jest dopuszczalna, by uznać je za zgodne, ale odpowiednia ilustracja graficzna uzyskanych różnic i możliwych granic zmienności ułatwia podjęcie takiej decyzji.

DEFINICJE MIAR

Wykres Blanda-Altmana jest wykresem punktowym, na którym:

Oś X – średnia pomiarów dla porównywanych metod;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Oś Y – różnica pomiędzy pomiarami dla porównywanych metod;

Średnia różnic *(ang. Mean difference)* – jeśli wyniki uzyskane nową metodą są stale większe/mniejsze niż metodą starą, występuje przesunięcie zwane obciążeniem *(ang. bias)*, tzn. linia obrazująca średnią różnic nie znajduje się na poziomie 0, ale jest przesunięta znacząco w górę lub w dół od tego poziomu;

95% przedział zgodności *(ang. 95% limits of agreement)* – jeśli różnice mają rozkład normalny, 95% różnic znajdzie się w przedziale (Średnia różnic $\pm$ 1.96SD), gdzie SD to odchylenie standardowe różnic.

Uwaga! Zdefiniowany powyżej przedział zgodności nie jest tym samym co przedział ufności. Nie ma też wymogu, by same dane miały rozkład normalny – wystarczy, by rozkład różnic nie odbiegał znacznie od rozkładu normalnego. Zgodność różnic z rozkładem normalnym można sprawdzić za pomocą testów badających zgodność z rozkładem normalnym lub interpretacji graficznej tego rozkładu.

Precyzja granic przedziału zgodności *(ang. Precision of the limits of agreement)* – przedział ufności dla granic zgodności, czyli przedział dokładności, z jaką wyznaczamy te granice na podstawie reprezentatywnej próby. Im liczniejsza próba i mniejsza wariancja różnic, tym większa uzyskana precyzja.

JAK USTAWIĆ ANALIZĘ

Wykres Blanda-Altmana dostępny jest jako dodatkowa opcja graficzna w oknach analiz służących do porównania dwóch zależnych serii pomiarów tej samej wielkości (np. przy porównaniu wyników dwóch mierników czy dwóch metod oceny). Wystarczy zaznaczyć odpowiednią opcję wykresu w oknie wybranej analizy, by uzyskać wykres zgodności dla porównywanych metod.

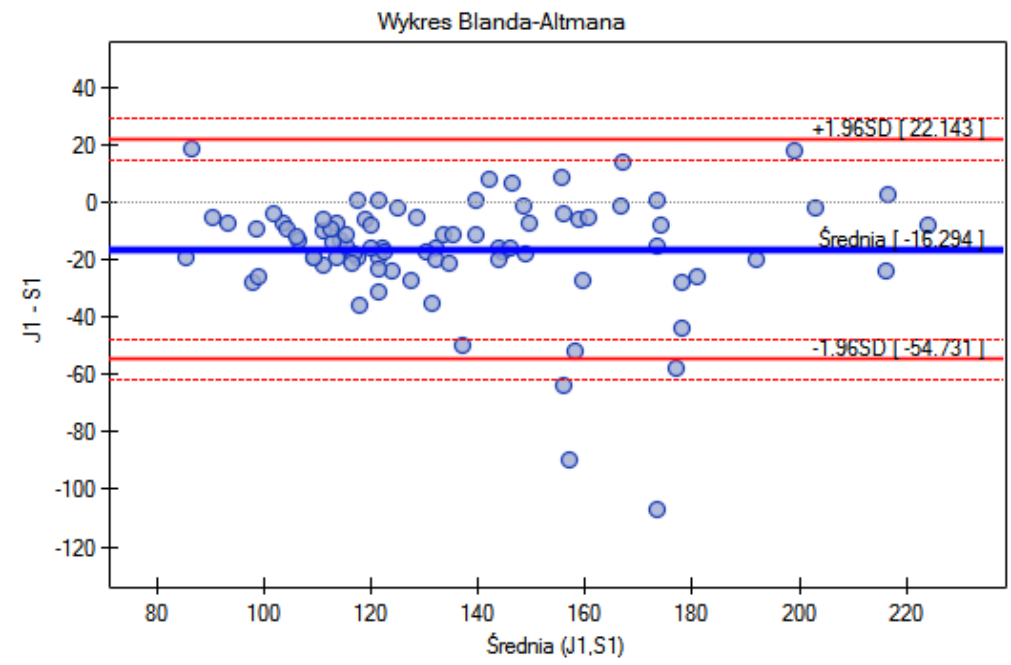
WYKONANIE I INTERPRETACJA

Przykład 9.2. (plik ciśnieniomierz.pqs)
Przykład został zaczerpnięty z pracy Blanda i Altmana (1999[8]). Porównano w niej ciśnieniomierz półautomatyczny (S) z tradycyjnie stosowanym wcześniej klasycznym ciśnieniomierzem (J). W tym celu dokonano pomiaru ciśnienia skurczowego krwi dla 85 pacjentów przy pomocy obu ciśnieniomierzy. Fragment danych przedstawiono poniżej.

J1	S1
100	122
108	121
76	95
108	127
124	140
122	139
116	122
114	130
100	119

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wykres Blanda-Altmana dla zebranych danych wskazuje, że ciśnieniomierz półautomatyczny (S) daje wyższe wyniki niż klasyczny średnio o 16.3mmHG (linia dla średniej różnic znajduje się o 16.3 niżej niż bezwzględna zgodność zobrazowana linią na poziomie 0). Rozpiętość przedziału zgodności wynosi aż 76.9mmHG.



Dla osób z nadciśnieniem (ciśnienie skurczowe $\geq$ 140) zmiany ciśnienia mogą być dość duże, więc badane różnice pomiarów mogą być zaburzone przez rzeczywiste skoki ciśnienia. Dlatego wyodrębniono podgrupę osób z ciśnieniem w normie i nadciśnieniem na podstawie średniej wartości ciśnienia. Dla każdej podgrupy możemy wykonać wykres oddzielnie, ustawiając w oknie testu filtr wielokrotny dla zmiennej *grupa*. Zgodność metod dla osób z ciśnieniem w normie będzie wówczas zdecydowanie lepsza (węższy przedział zgodności).

Wykres Blanda-Altmana dla powtarzanych pomiarów

Powtarzalność pomiarów jest ważnym, lecz często pomijanym aspektem badania zgodności metod. Metoda o większej powtarzalności jest bardziej precyzyjna. Jeśli pomiary jednej z porównywanych metod nie są powtarzalne (tzn. powtórne pomiary wykonane na tych samych obiektach dają raczej inne wyniki), jej zgodność z inną metodą będzie niska. Jeśli powtarzalność obu metod będzie słaba, to ich zgodność będzie jeszcze niższa. W konsekwencji, gdy powtarzalność starej metody będzie słaba, zgodność nowej metody może okazać się niewielka, nawet jeśli nowa metoda charakteryzuje się wysoką powtarzalnością. Dlatego, mimo że w rzeczywistych badaniach wykonuje się zwykle jeden pomiar dla każdego obiektu (pacjenta), w badaniach mających na celu szacowanie zgodności zaleca się wykonywanie pomiarów kilkukrotnie – pozwala to uwzględnić powtarzalność uzyskanych wyników w badaniu zgodności metod.

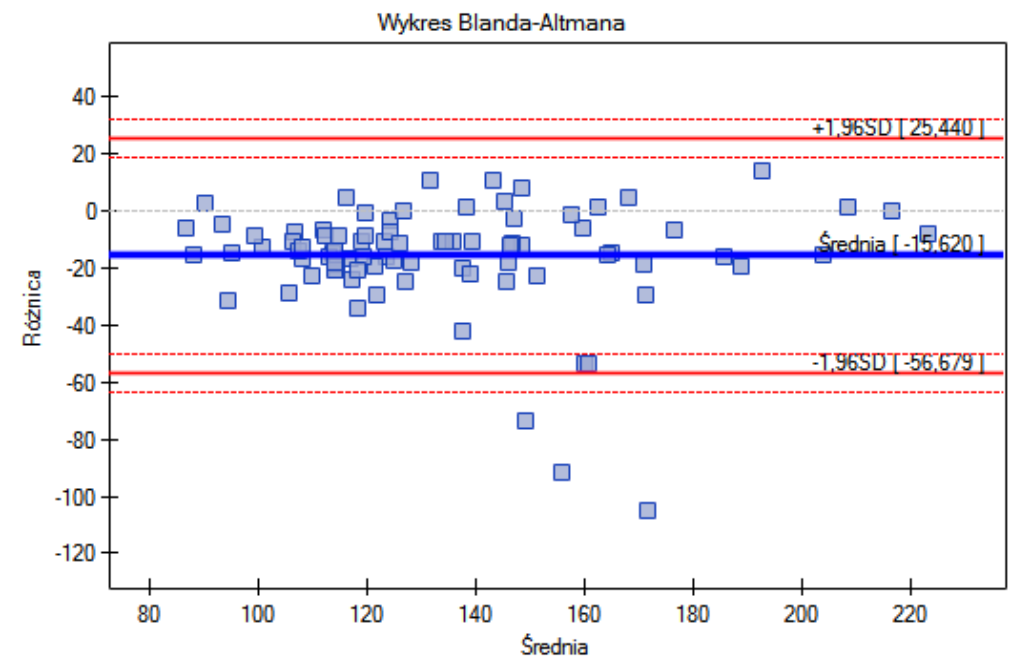
Uwaga! Jako powtarzane pomiary rozumiemy tutaj pomiary wykonywane niezależnie na tych samych obiektach, w takich samych warunkach.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PRZYKŁAD (9.2) c.d. *(plik ciśnieniomierz.pqs)*
W porównaniu zgodności pomiarów dokonanych przy pomocy porównywanych ciśnieniomierzy wzięto pod uwagę również powtarzalność obu metod. Dlatego badanie powtórzono jeszcze dwukrotnie, uzyskując ostatecznie dla każdego pacjenta 3 pomiary wykonane ciśnieniomierzem półautomatycznym i 3 pomiary wykonane ciśnieniomierzem klasycznym. Fragment danych przedstawiono poniżej.

J1	J2	J3	S1	S2	S3
100	106	107	122	128	124
108	110	108	121	127	128
76	84	82	95	94	98
108	104	104	127	127	135
124	112	112	140	131	124
122	140	124	139	142	136
116	108	102	122	112	112
114	110	112	130	129	135
100	108	112	119	122	122

Tym razem granice zgodności są nieco szersze niż przy wykorzystaniu pojedynczego pomiaru dla każdej metody – rozpiętość przedziału zgodności wynosi aż 82.11mmHG. Dzieje się tak dlatego, że uwzględniamy stopień powtarzalności pomiarów. Niestety, uwzględnienie kilku powtórzeń zwiększa szerokość granic, ale przedstawione wyniki lepiej reprezentują rzeczywistość – nieuwzględnienie stopnia powtarzalności oznaczałoby założenie 100-procentowej powtarzalności, co w rzeczywistych warunkach jest niemal niemożliwe.



Podobnie jak poprzednio, zaleca się powtórzyć analizę oddzielnie dla osób z nadciśnieniem i z ciśnieniem w normie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

9.4 Macierze korelacji

KIEDY STOSOWAĆ:

Gdy interesuje nas korelacja pomiędzy wieloma zmiennymi naraz, wygodnym sposobem jej zobrazowania jest zestawienie współczynników korelacji dla wszystkich par zmiennych w postaci jednej tabeli (macierzy). Zamiast wielokrotnie powtarzać tę samą analizę dla kolejnych par zmiennych, otrzymujemy od razu pełny obraz wzajemnych powiązań w badanym zbiorze danych. W zależności od skali, na jakiej zebrano dane, program PQStat pozwala wybrać jeden z następujących współczynników:

- r-Pearsona (skala interwałowa)
- r-Spearmana (skala porządkowa lub silniejsza)
- tau-Kendalla (skala porządkowa lub silniejsza)
- C-Pearsona (skala nominalna lub silniejsza)
- V-Cramera (skala nominalna lub silniejsza)
- Phi (skala nominalna lub silniejsza)
- Q-Yulea (skala nominalna lub silniejsza)

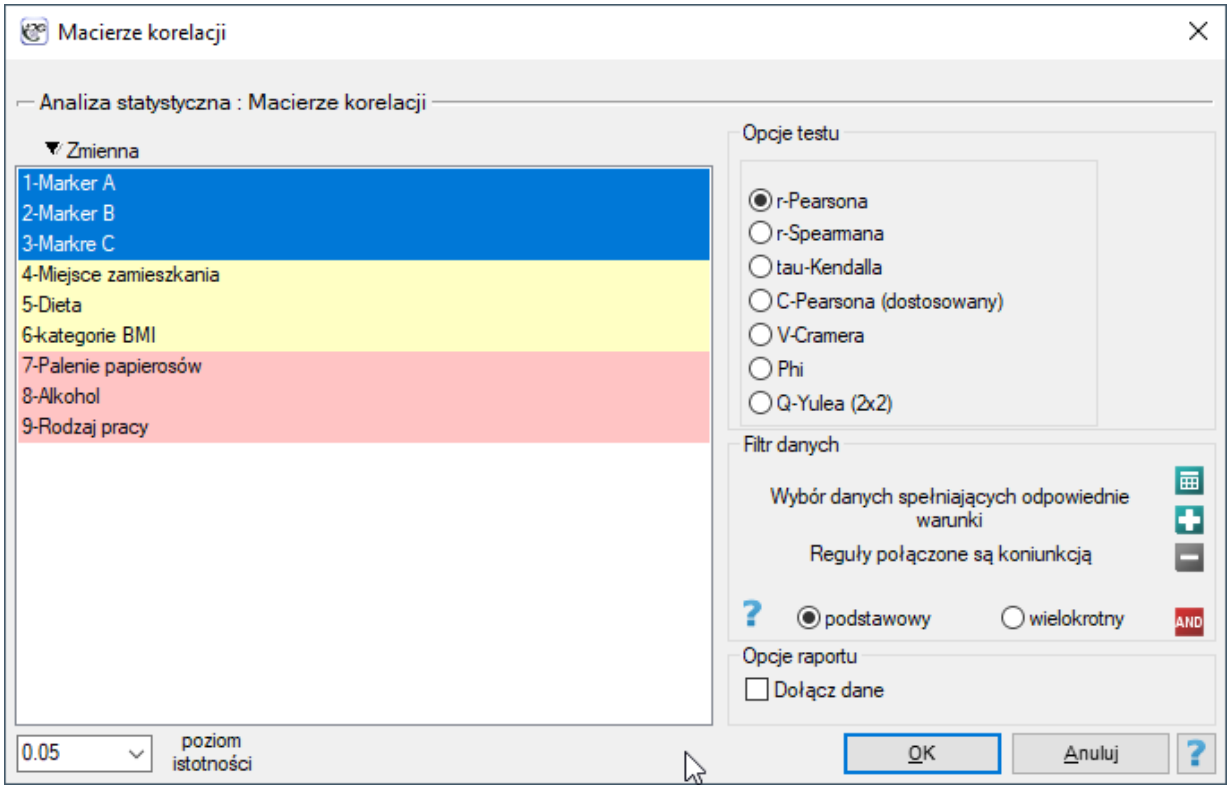
W wyniku analizy powstają dwie macierze: macierz współczynników korelacji oraz macierz wartości p dla testu sprawdzającego istotność statystyczną każdego z tych współczynników (dla r-Pearsona, r-Spearmana i tau-Kendalla są to dedykowane im testy, a dla skali nominalnej – test chi-kwadrat). W miejscu przecięcia się dwóch zmiennych w macierzy współczynników znajduje się wartość ich wzajemnej korelacji, a odpowiadająca jej wartość p znajduje się w tym samym miejscu drugiej macierzy. Kolor komórki w macierzy współczynników jest stopniowany od niebieskiego (współczynniki ujemne) do czerwonego (współczynniki dodatnie), co pozwala od razu wychwycić wzrokiem najsilniejsze powiązania.

Uwaga! Korelację wyznacza się osobno dla każdej pary zmiennych, dlatego braki danych są pomijane parami. Jeśli chcemy, aby z analizy usuwane były całe przypadki mające braki danych również w innych zmiennych (spoza danej pary), powinniśmy skorzystać z filtru zaawansowanego.

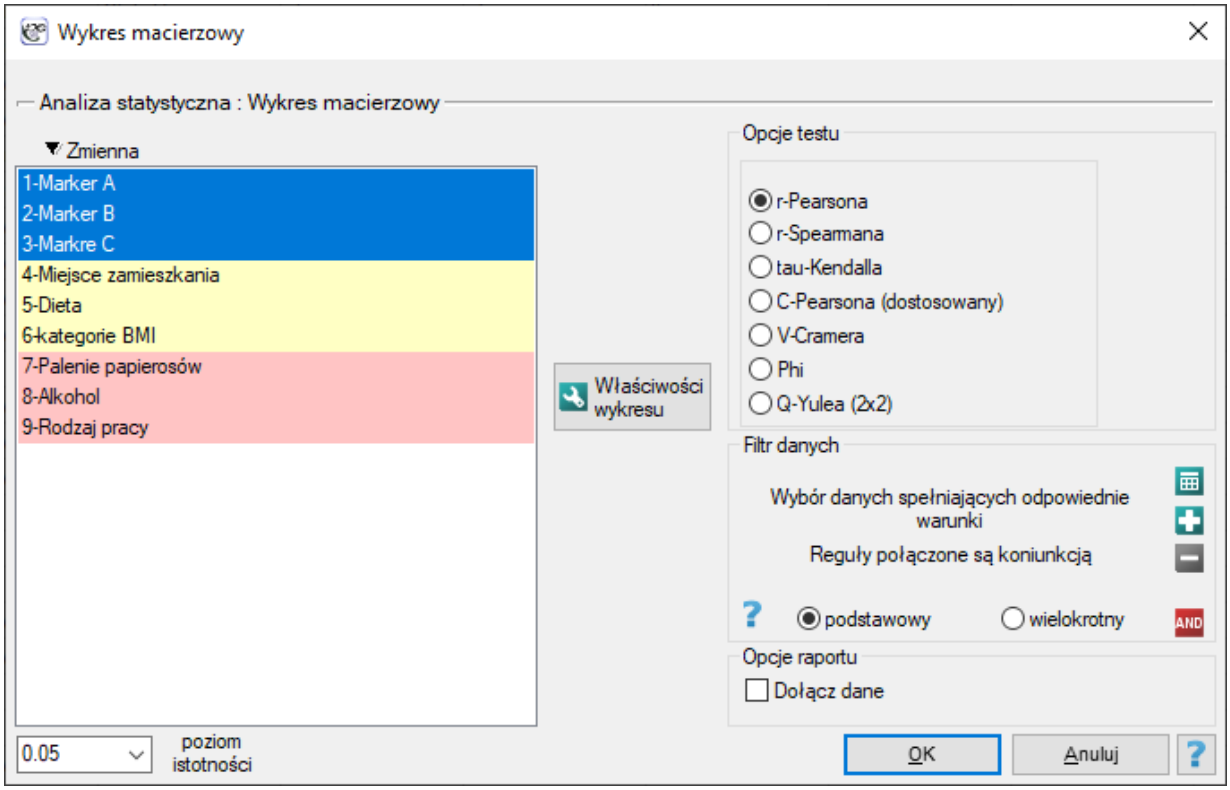
JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji macierzy korelacji wywołujemy poprzez Statystyki→Kalkulatory→Macierze korelacji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Okno z ustawieniami opcji wykresu macierzowego dla macierzy korelacji wywołujemy poprzez Wykresy→Wykres macierzowy.



Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

WYKONANIE I INTERPRETACJA

PRZYKŁAD 9.3. (plik markery inne.pqs)
Przedstawiono fragment większego badania dotyczącego choroby nowotworowej. Zebrane dane dotyczą grupy 100 osób. W badaniu mierzono m.in. wartości markerów nowotworowych (skala interwałowa), wyznaczono kategorie BMI dla pacjentów oraz pytano o opinie na temat możliwego wpływu miejsca zamieszkania oraz diety na zdrowie (skala porządkowa), zanotowano też odpowiedzi pacjentów na temat palenia przez nich papierosów, spożywania alkoholu oraz rodzaju wykonywanej pracy (skala nominalna).
Prowadzenie analiz wielowymiarowych często wymaga wcześniejszego sprawdzenia wzajemnych korelacji zmiennych. Na potrzeby dalszych analiz:
(1) sprawdzimy korelację wewnątrz każdej z tych grup,
(2) sprawdzimy korelację między wszystkimi zmiennymi.

(1) Korelacja wewnątrz poszczególnych grup zmiennych

Dla skali interwałowej, przy założeniu normalności rozkładu, korelację sprawdzimy współczynnikiem korelacji liniowej Pearsona.

Marker A i marker C: $r = 0.8995$, $p < 0.0001$

Najsilniejsza i istotna statystycznie korelacja w tej grupie zmiennych. Najsłabsza, nieistotna statystycznie korelacja dotyczy markera B i markera C ($r = 0.0753$, $p = 0.4567$).

r-Pearsona

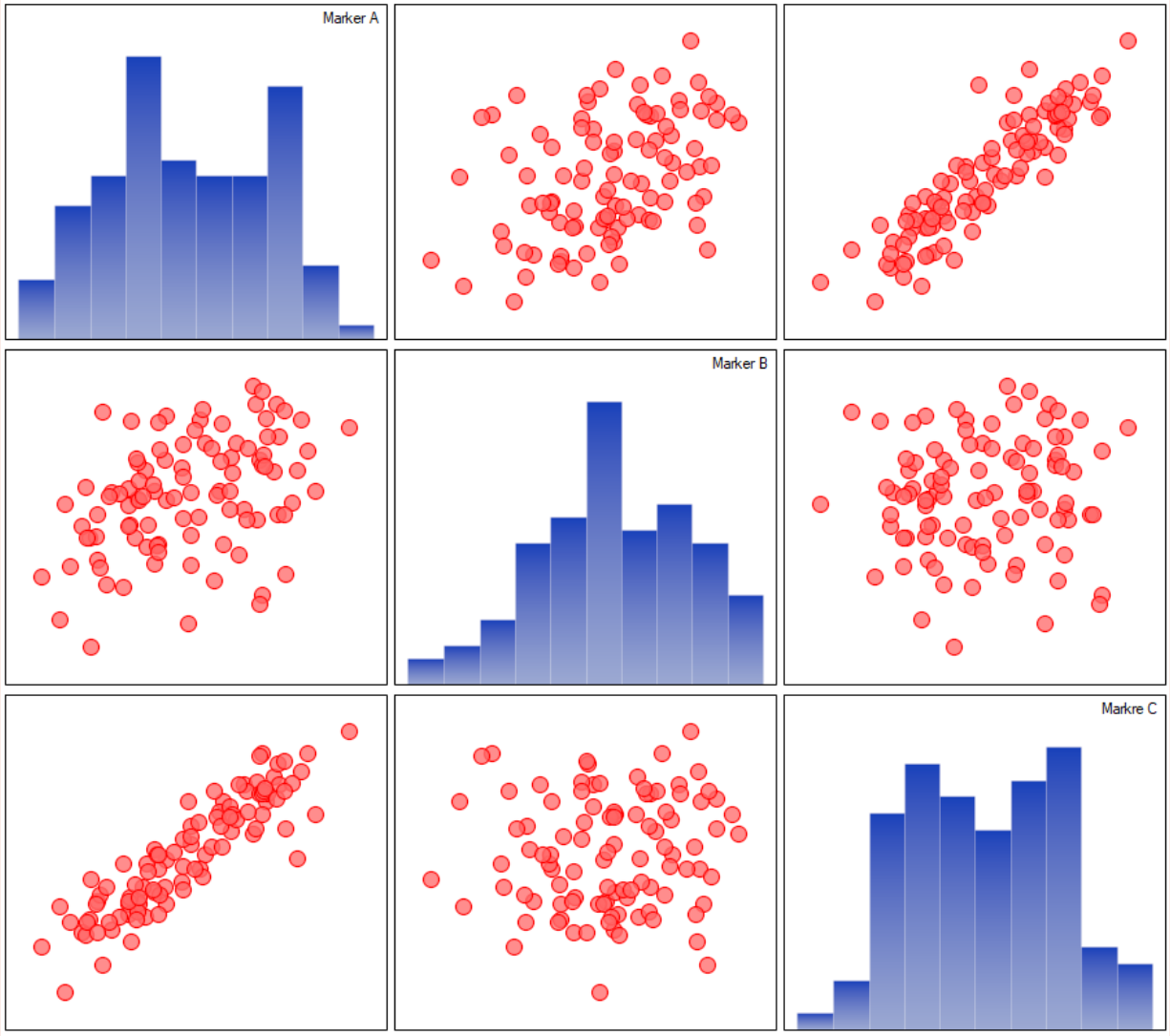
	Marker A	Marker B	Markre C
Marker A		**0.4027**	**0.8995**
Marker B	**0.4027**		0.0753
Markre C	**0.8995**	0.0753	

P-value

	Marker A	Marker B	Markre C
Marker A		<0.0001	<0.0001
Marker B	<0.0001		0.4567
Markre C	<0.0001	0.4567	

Opisane korelacje można obserwować na wykresach rozrzutu (oś X tych wykresów to zmienna opisana w kolumnach, oś Y – w wierszach), a rozkłady poszczególnych zmiennych – na wykresach kolumnowych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Dla skali porządkowej korelację sprawdzimy przy pomocy współczynnika korelacji Spearmana.

Dieta i miejsce zamieszkania: $r = 0.2634$, $p = 0.0081$

Jedyna istotna statystycznie korelacja w tej grupie zmiennych.

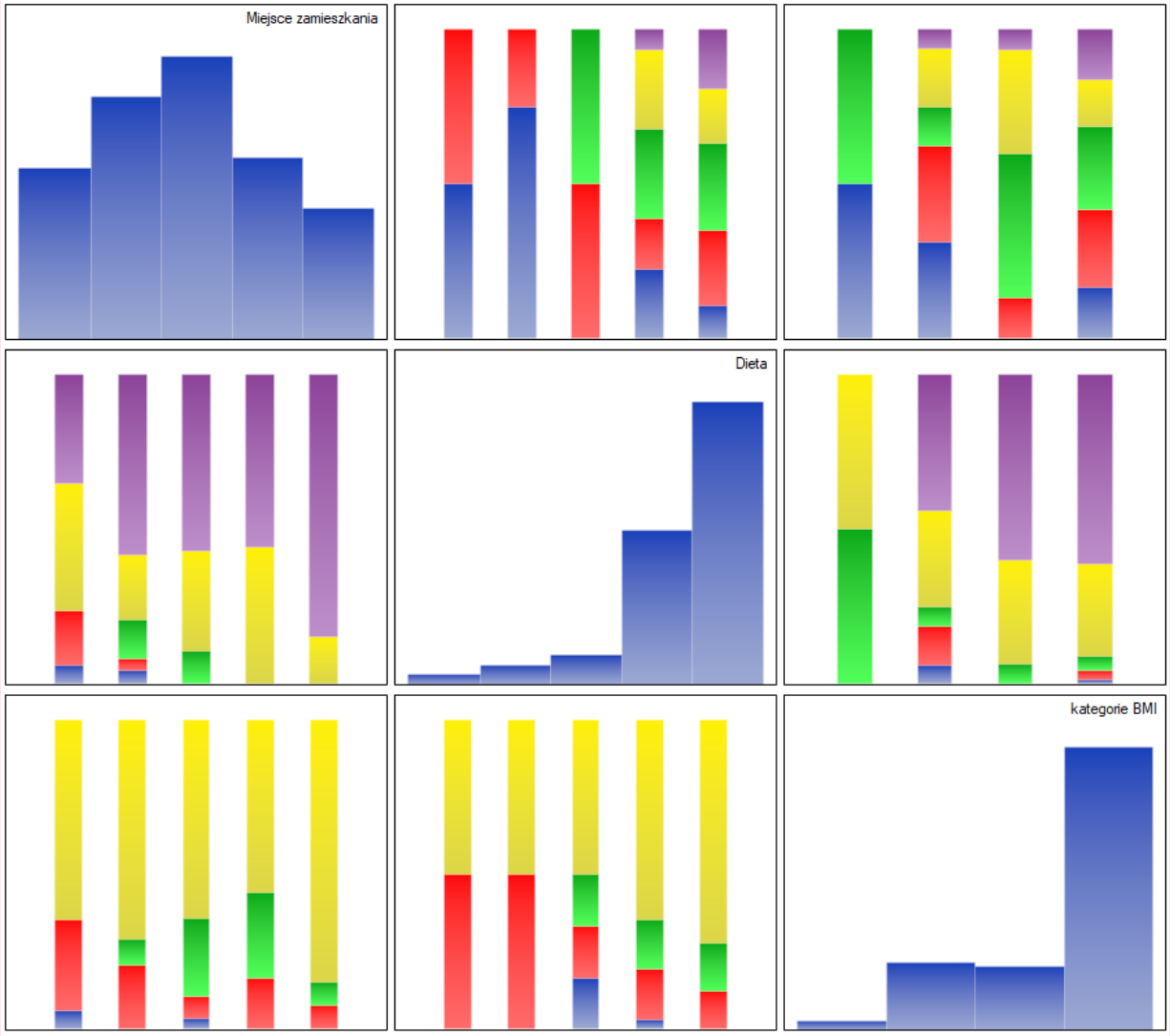
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

r-Spearmana	Miejsce zam	Dieta	kategorie BM
Miejsce zamiesz		**0.2634**	0.0746
Dieta	**0.2634**		0.1735
kategorie BMI	0.0746	0.1735	

P-value	Miejsce zam	Dieta	kategorie BM
Miejsce zamiesz		0.0081	0.4608
Dieta	0.0081		0.0842
kategorie BMI	0.4608	0.0842	

Opisane korelacje można obserwować na wykresach kolumnowych skumulowanych (oś X tych wykresów to zmienna opisana w kolumnach, legenda to zmienna opisana w wierszach), a rozkłady poszczególnych zmiennych – na wykresach kolumnowych znajdujących się na głównej przekątnej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Dla skali nominalnej korelację sprawdzimy przy pomocy współczynnika C-Pearsona skorygowanego na wielkość tabeli. W tej grupie zmiennych nie uzyskaliśmy istotnych statystycznie korelacji.

C-Pearsona (do

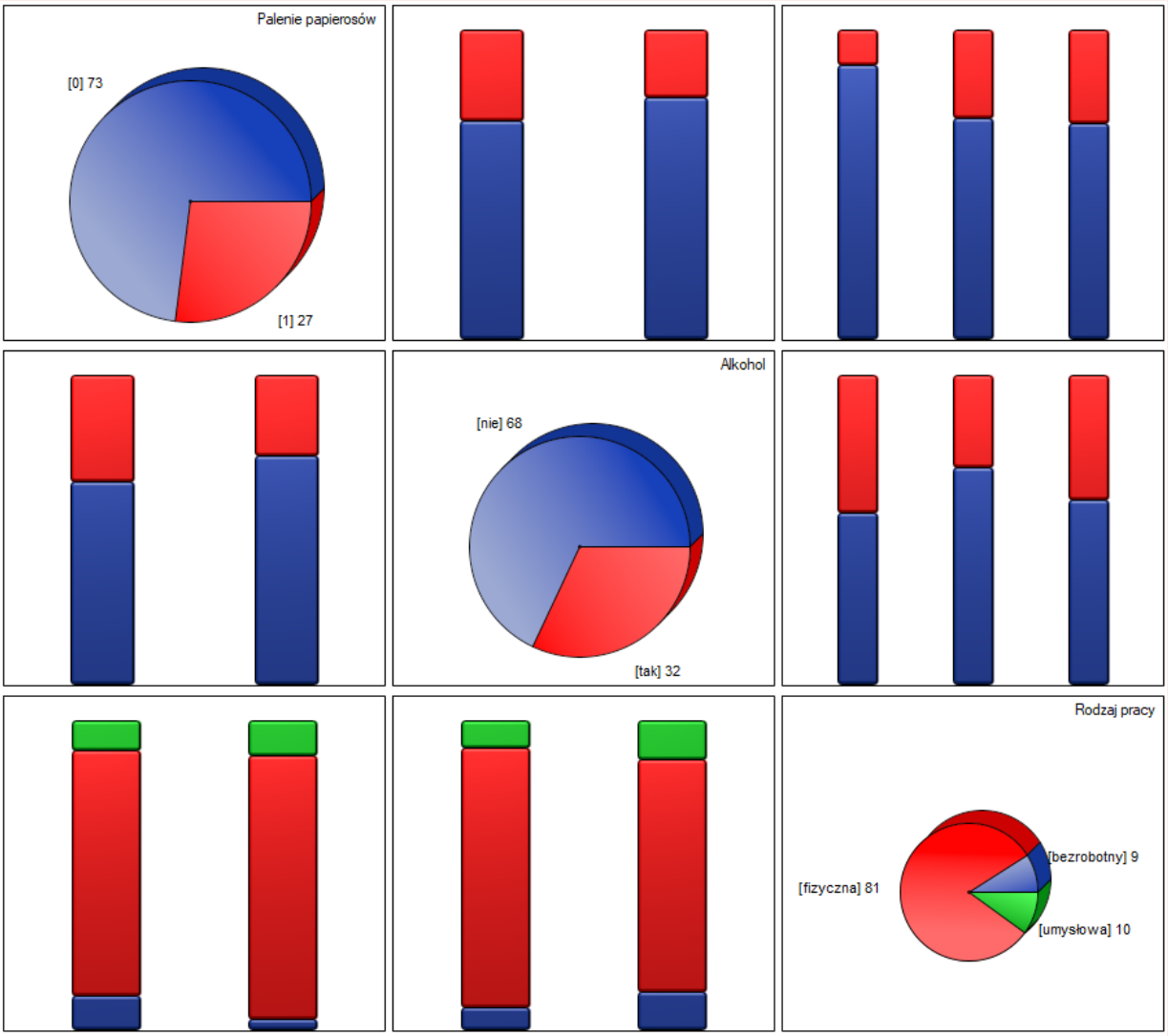
	Palenie papi	Alkohol	Rodzaj pracy
Palenie papieros		0.1116	0.1589
Alkohol	0.1116		0.1504
Rodzaj pracy	0.1589	0.1504	

P-value

	Palenie papi	Alkohol	Rodzaj pracy
Palenie papieros		0.4284	0.5277
Alkohol	0.4284		0.5645
Rodzaj pracy	0.5277	0.5645	

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

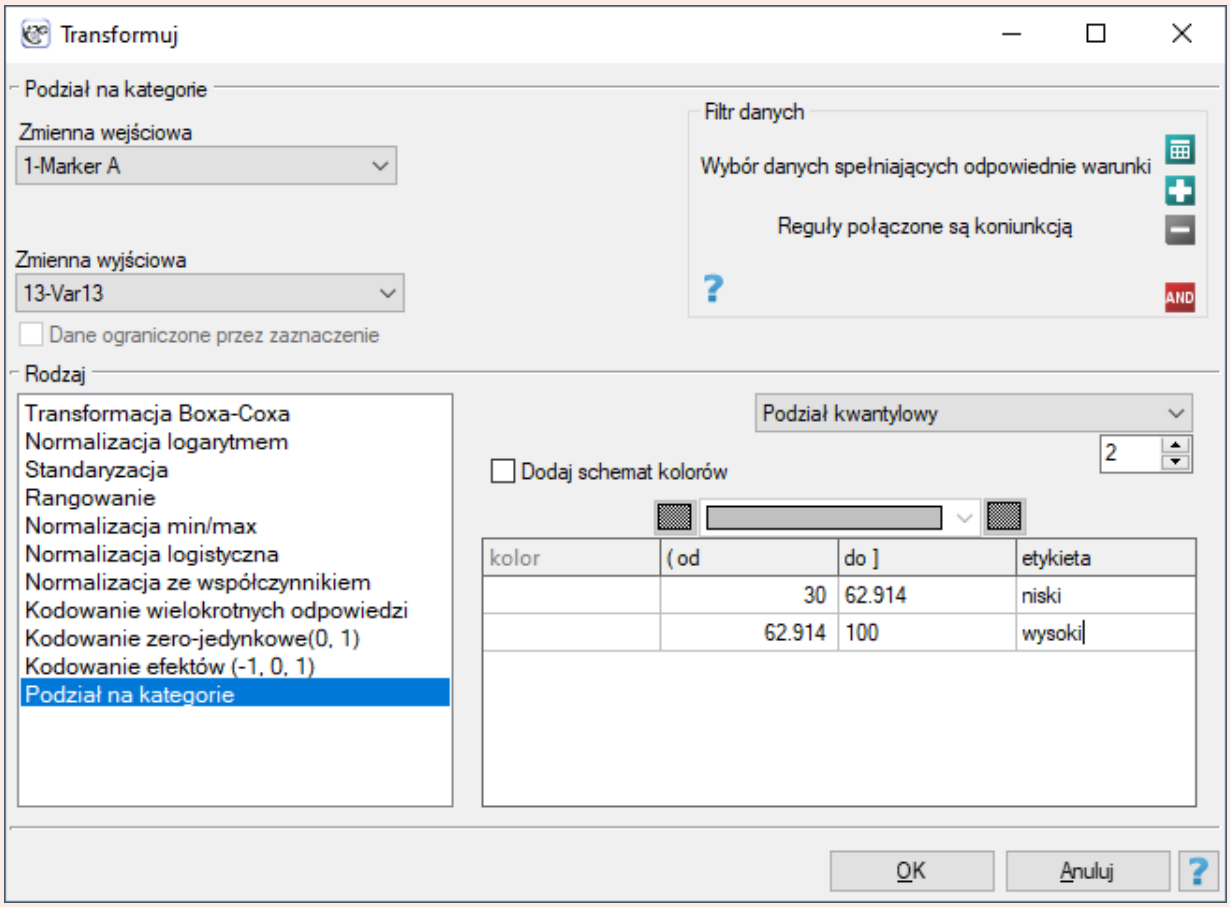
Ewentualne korelacje można by obserwować na wykresach kolumnowych skumulowanych (oś X tych wykresów to zmienna opisana w kolumnach, legenda to zmienna opisana w wierszach), a rozkłady poszczególnych zmiennych – na wykresach kolumnowych znajdujących się na głównej przekątnej.



(2) Korelacja między wszystkimi zmiennymi

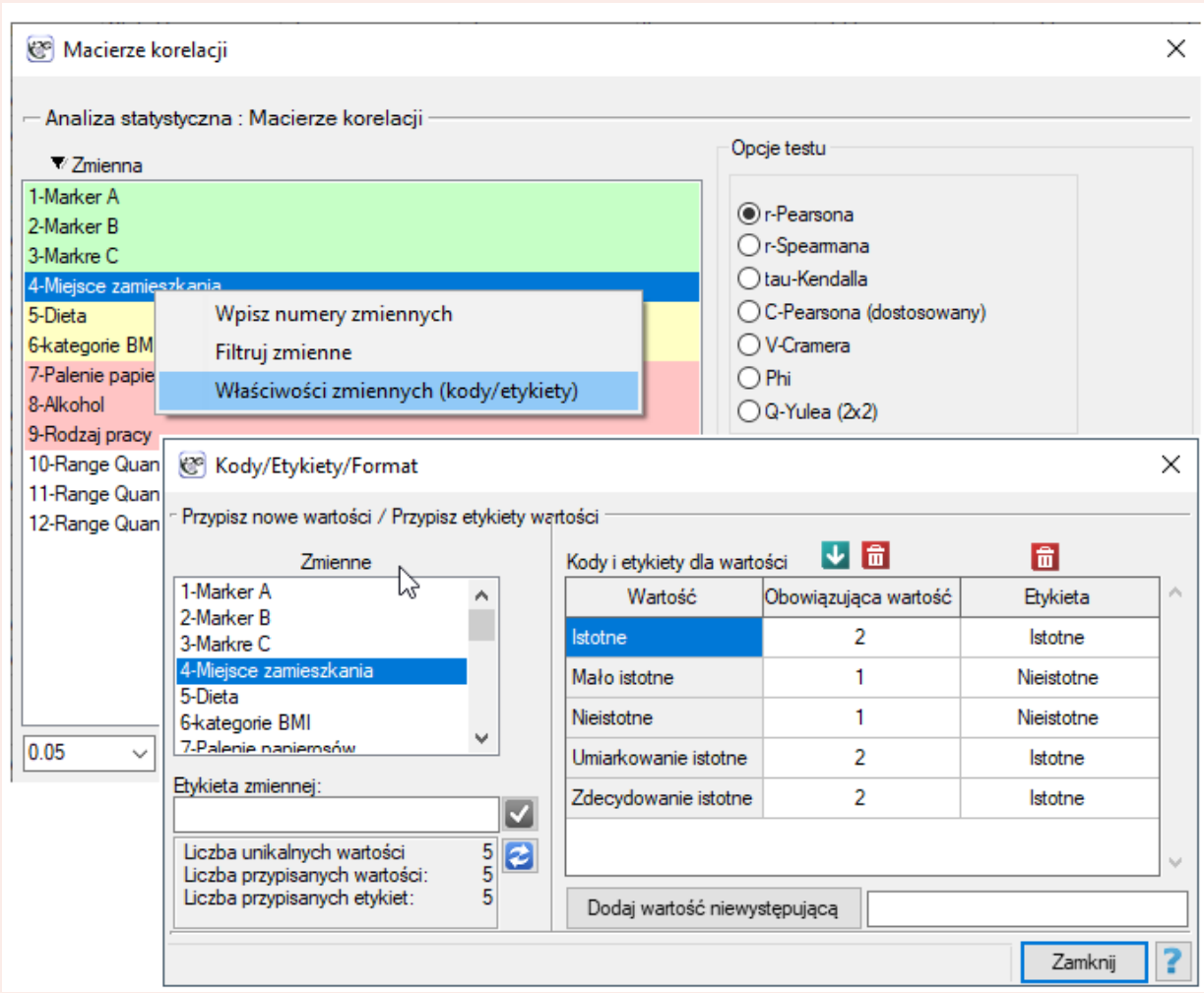
Najłatwiejszym sposobem wyznaczenia korelacji pomiędzy zmiennymi mierzonymi na różnych skalach jest sprowadzenie ich do tej samej skali. W tym celu zapiszemy dane interwałowe, dzieląc je na dwie kategorie – ”niski” i ”wysoki” – np. wg kwantyli. Możemy to zrobić automatycznie w oknie transformacji poprzez menu Dane→Transformuj.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Dane porządkowe również podzielimy na dwie kategorie, ale podziału dokonamy, wybierając Właściwości zmiennych (kody/etykiety) w oknie analizy poprzez menu kontekstowe (prawy klawisz myszy) i wpisując tylko dwie obowiązujące wartości i dwie etykiety.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W wyniku przedstawimy tylko macierz korelacji (bez wykresu), ponieważ wykres dla tak wielu zmiennych nie będzie wystarczająco czytelny.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

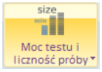
C-Pearsona (do

	Miejsce zam	Dieta	kategorie BM	Palenie papi	Alkohol	Rodzaj pracy	Range Quant	Range Quant	Range Quant
Miejsce zamiesz		**0.4102**	0.2661	0.1335	0.1761	0.107	0.0287	0.143	0.0965
Dieta	**0.4102**		**0.2912**	0.083	0.0102	**0.4159**	<0.0001	0.2349	0.1329
kategorie BMI	0.2661	**0.2912**		0.1533	0.0189	0.24	0.1465	0.0735	**0.3133**
Palenie papieros	0.1335	0.083	0.1533		0.1116	0.1589	0.0953	0.1583	0.192
Alkohol	0.1761	0.0102	0.0189	0.1116		0.1504	<0.0001	0.239	0.0218
Rodzaj pracy	0.107	**0.4159**	0.24	0.1589	0.1504		0.0497	0.1669	0.0353
Range Quantile[	0.0287	<0.0001	0.1465	0.0953	<0.0001	0.0497		**0.3813**	**0.91**
Range Quantile[	0.143	0.2349	0.0735	0.1583	0.239	0.1669	**0.3813**		0.1686
Range Quantile[	0.0965	0.1329	**0.3133**	0.192	0.0218	0.0353	**0.91**	0.1686	

P-value

	Miejsce zam	Dieta	kategorie BM	Palenie papi	Alkohol	Rodzaj pracy	Range Quant	Range Quant	Range Quant
Miejsce zamiesz		0.0024	0.0554	0.3431	0.2094	0.7497	0.8389	0.3093	0.4942
Dieta	0.0024		0.0353	0.5565	0.9424	0.0088	1	0.0921	0.3452
kategorie BMI	0.0554	0.0353		0.2755	0.8935	0.227	0.2978	0.6027	0.0231
Palenie papieros	0.3431	0.5565	0.2755		0.4284	0.5277	0.4992	0.2601	0.1705
Alkohol	0.2094	0.9424	0.8935	0.4284		0.5645	1	0.0863	0.8772
Rodzaj pracy	0.7497	0.0088	0.227	0.5277	0.5645		0.9401	0.4935	0.9693
Range Quantile[	0.8389	1	0.2978	0.4992	1	0.9401		0.0051	<0.0001
Range Quantile[	0.3093	0.0921	0.6027	0.2601	0.0863	0.4935	0.0051		0.2298
Range Quantile[	0.4942	0.3452	0.0231	0.1705	0.8772	0.9693	<0.0001	0.2298	

10 MOC TESTU I LICZNOŚĆ PRÓBY



Do wyznaczania liczności próby możemy podejść na kilka sposobów. Jedną z możliwości jest oszacowanie, jak duża powinna być próba, by w dobry sposób odzwierciedlała populację. Inną możliwością jest oszacowanie liczności próby do sytuacji stosowania konkretnych testów statystycznych – wówczas oprócz niezbędnej liczności próby możemy być zainteresowani również mocą tych testów. Pierwsze, i wydaje się łatwiejsze, podejście zaprezentowane jest w podrozdziale Niezbędna liczność próby, drugie w podrozdziale Moc i liczność dla testu.

10.1 Niezbędna liczność próby

KIEDY STOSOWAĆ:

Badanie całej populacji najczęściej nie jest ani praktyczne, ani możliwe, dlatego wybiera się jej podzbiór – próbę. Próba jest oczywiście mniej liczna niż populacja, ale powinna ją dobrze odzwierciedlać. Jednym z kluczowych aspektów przy planowaniu badania, obok losowości próby, jest ustalenie jej wielkości – tak, by wnioskowanie na temat populacji na podstawie próby było wiarygodne.

Jeśli zainteresowani jesteśmy tym, by proporcja pewnej cechy lub wartość średnia wyliczone dla próby odzwierciedlały proporcję lub wartość średnią w populacji z możliwie małym błędem, możemy jeszcze przed rozpoczęciem badania oszacować, ile osób/obserwacji jest do tego niezbędnych. W programie PQStat możemy w ten sposób wyznaczyć niezbędną liczność próby zarówno dla oszacowania **proporcji**, jak i **wartości średniej** w populacji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji Niezbędnej liczności próby wywołujemy poprzez menu Statystyki zaawansowane→Moc testu i liczność próby→Niezbędna liczność próby. Niezależnie od tego, czy szacujemy niezbędną liczność próby dla proporcji czy dla średniej, musimy podać zakładany margines błędu E, czyli maksymalne odchylenie, jakie jesteśmy skłonni zaakceptować pomiędzy wynikiem uzyskanym z próby a rzeczywistą wartością w populacji.

- Dla **proporcji** podajemy dodatkowo oczekiwaną proporcję cechy w populacji – jeśli jej nie znamy, program przyjmie najbardziej ostrożną wartość $p = 0.5$, dla której niezbędna liczność próby jest największa.
- Dla **średniej** podajemy dodatkowo odchylenie standardowe cechy w populacji, znane np. z wcześniejszych badań.

Jeśli znamy liczność całej populacji (a w szczególności, gdy jest ona relatywnie mała w stosunku do wyliczonej liczności próby), warto zastosować tzw. korektę skończonej populacji (FPC) (Lenth (2001)[103], Armitage i Colton (2009)[12]) – program pozwala na jej uwzględnienie po podaniu wielkości populacji.

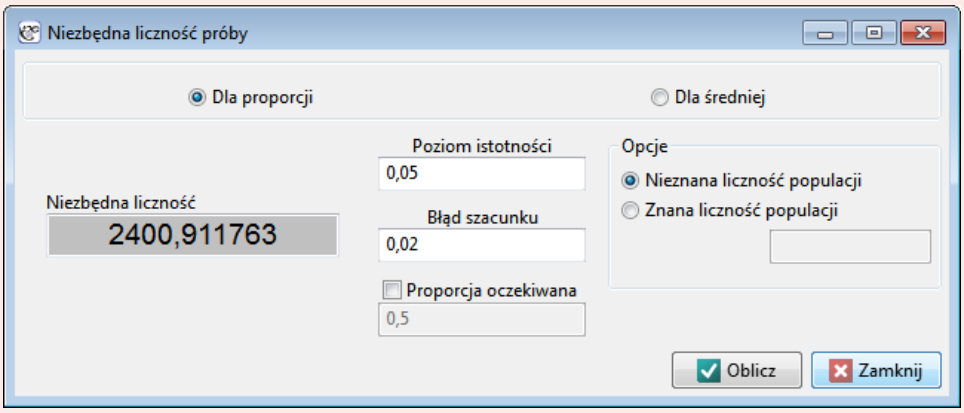
WYKONANIE I INTERPRETACJA

Przykład 10.1. Szacowanie proporcji

Populacja: Uprawnieni do głosowania na prezydenta Polski.

Zainteresowani jesteśmy poparciem dla poszczególnych kandydatów.

Ile osób wybrać, by uzyskany odsetek obarczony był błędem co najwyżej 2%?



Niezbędna liczność = 2401 osób

Przy próbie przynajmniej 2401-elementowej będziemy mieć 95% pewności, że błąd poparcia dla wybranego kandydata na prezydenta nie przekracza 2%.

Oznacza to, że w 95% eksperymentów polegających na pobieraniu losowej, 2401-elementowej próby z populacji, błąd szacunku poparcia dla danego kandydata nie przekroczy 2%, ale w 5% takich eksperymentów może on być większy niż 2%.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

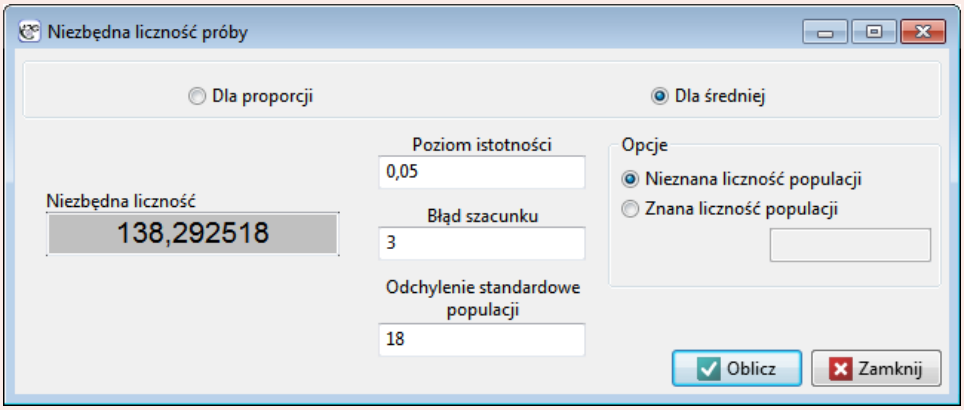
Przy wyborze wielkości dopuszczanego błędu należy zwrócić uwagę, czy w wyborach nie biorą udziału również kandydaci o niewielkim poparciu (na granicy założonego błędu szacunku). Jeśli tak jest, warto zmniejszyć wartość zakładanego błędu – konsekwencją tego będzie wówczas wzrost niezbędnej liczności próby.

PRZYKŁAD 10.2. Szacowanie wartości średniej

Populacja: Osoby chorujące na nadciśnienie w Polsce w latach 2005-2010, w wieku 20-40 lat.

Zainteresowani jesteśmy średnią masą ciała tych ludzi.

Ile osób wybrać, by średnia masa obarczona była błędem co najwyżej $3kg$? Wiemy, że populacyjne odchylenie standardowe masy ciała tych osób wynosi $18kg$.



Niezbędna liczność = 139 osób

By mieć 95% pewności, że w marginesie błędu ($\pm 3kg$) zbudowanym wokół średniej z naszej próby znajduje się średnia populacyjna, potrzebujemy zebrać przynajmniej 139 osób.

DLA ZAINTERESOWANYCH

Liczność próby dla oszacowania proporcji populacji

Wzory: zakładając możliwość popełnienia błędu szacunku wielkości E, wyznaczamy niezbędną liczność próby n_0 – dla nieznanej wielkości populacji, lub n_{FPC} – dla znanej wielkości populacji.

$$n_0 = \frac{Z^2 p(1-p)}{E^2}$$

gdzie:

> p – podana przez badacza oczekiwana proporcja w populacji, przy czym, jeśli wielkość ta nie jest znana, oszacowana niezbędna liczność będzie tak zwiększona, by była wystarczająca dla każdej możliwej proporcji, zatem wykorzystana zostanie wartość $p = 0.5$.

W przypadku, gdy znamy liczność populacji (a w szczególności – gdy liczność jest relatywnie mała w stosunku do n_0, tzn. gdy $n_0/N > 5\%$), powinniśmy stosować tzw. korektę skończonej populacji (FPC)

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

(Lenth (2001)[103], Armitage i Colton (2009)[12]) daną wzorem:

$$n_{FPC} = \frac{n_0 N}{n_0 + (N-1)}$$

Liczność próby dla oszacowania średniej populacji
Wzory: zakładając możliwość popełnienia błędu szacunku wielkości E, wyznaczamy niezbędną liczność próby n_0 – dla nieznanej wielkości populacji, lub n_{FPC} – dla znanej wielkości populacji.

$$n_0 = \frac{Z^2 \sigma^2}{E^2}$$

gdzie:

σ – odchylenie standardowe populacji, znane na podstawie wcześniejszych badań.

W przypadku, gdy znamy liczność populacji (a w szczególności – gdy liczność jest relatywnie mała w stosunku do n_0, tzn. gdy $n_0/N > 5\%$), powinniśmy stosować tzw. korektę skończonej populacji (FPC) daną wzorem:

$$n_{FPC} = \frac{n_0 N}{n_0 + (N-1)}$$

10.2 Moc i liczność dla testu

KIEDY STOSOWAĆ:

Zanim przeprowadzimy badanie (lub gdy już je przeprowadziliśmy), warto wiedzieć, czy zastosowany test statystyczny w ogóle miał szansę wykryć interesującą nas różnicę, związek czy zależność. Do tego właśnie służy **analiza mocy testu** i wyznaczanie **niezbędnej liczności próby**.

Testując hipotezy, możemy popełnić dwa rodzaje pomyłek: uznać, że coś się różni, choć w rzeczywistości nie ma różnicy (**błąd I rodzaju**, oznaczany α), albo przeoczyć realną różnicę, uznając, że jej nie ma (**błąd II rodzaju**, oznaczany β). **Moc testu** to zdolność testu do wykrycia różnicy, związku, zgodności czy innej zależności opisanej w hipotezie alternatywnej – innymi słowy, to prawdopodobieństwo, że test poprawnie „zauważy" efekt, który naprawdę istnieje w populacji. Nie wszystkie testy są tak samo skuteczne – jedne są „czulsze" (mocniejsze), inne „mniej czułe" (słabsze), dlatego mając do wyboru kilka testów pasujących do tego samego problemu, warto sięgnąć po ten o większej mocy.

Sprawdzenie mocy testu można wykonać **a priori**, czyli jeszcze przed zebraniem danych – wtedy planujemy, ile osób/obiektów będziemy potrzebować, by mieć realną szansę wykrycia efektu, który nas interesuje. Można to też zrobić **a posteriori**, czyli już po zebraniu próby – często robią to recenzenci prac lub sami badacze, by ocenić, na ile wiarygodny jest uzyskany (zwłaszcza nieistotny statystycznie) wynik. Niska moc oznacza, że wyniki mogą być niejednoznaczne – inny badacz, powtarzając eksperyment, łatwo mógłby uzyskać odmienny rezultat. Wysoka moc oznacza, że wynik jest solidny – trudno będzie go podważyć. Przykładowo: jeśli przy pomocy testu o mocy 80% wykazujemy, że dwie grupy studentów różnią się istotnie statystycznie liczbą poprawnie rozwiązanych zadań, oznacza to, że powtarzając ten sam eksperyment (ta sama wielkość próby, ten sam test, ten sam poziom istotności) w 80% przypadków inni badacze również potwierdziliby tę różnicę.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

W programie PQStat możemy iść w dwie strony: obliczyć **moc testu**, podając liczność próby, albo obliczyć **niezbędną liczność próby**, podając moc testu, jaką chcemy osiągnąć.

WARUNKI STOSOWANIA

- znajomość zasad stosowania testu, dla którego liczymy moc lub liczność (jego hipotez, założeń, sposobu interpretacji),
- umiejętność określenia wielkości efektu, jaki chcemy być w stanie wykryć – najlepiej na podstawie danych pilotażowych, wcześniejszych badań lub wiedzy merytorycznej o badanym zjawisku.

DEFINICJE MIAR

Moc testu i niezbędna liczność próby są ze sobą ściśle powiązane, a dodatkowo zależą od kilku innych elementów, które musimy ustalić:

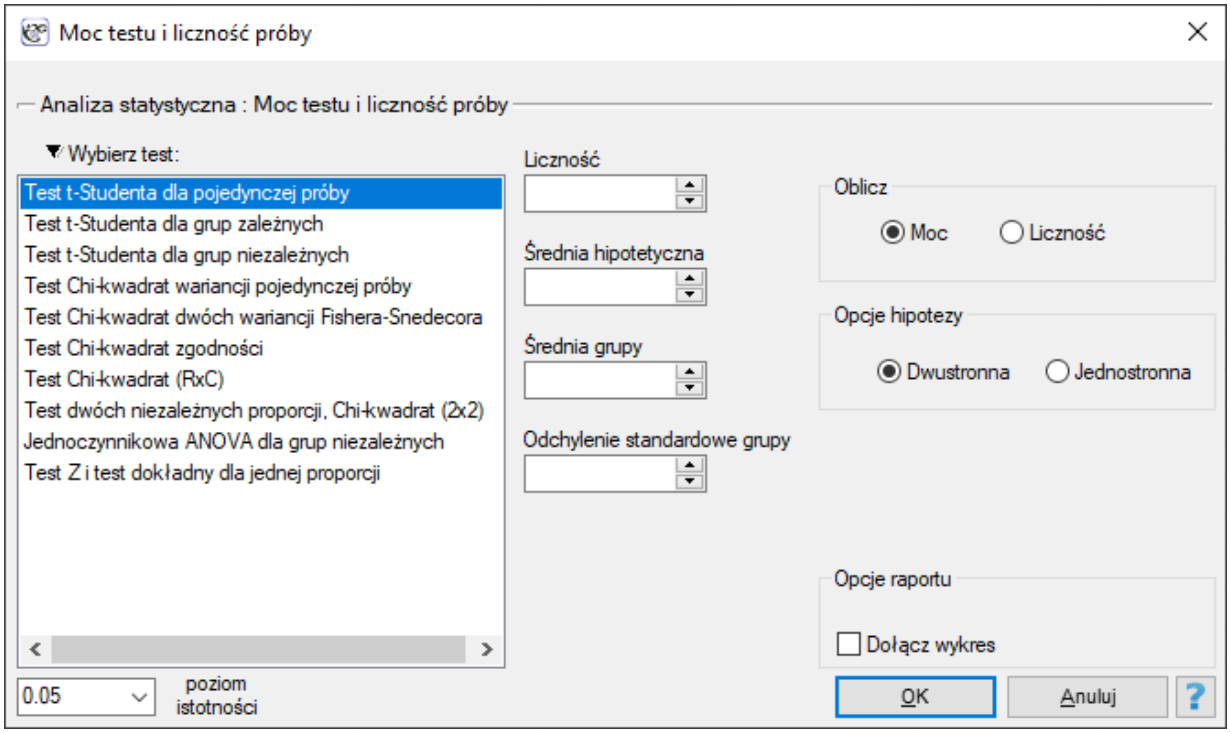
- **Wielkość efektu** ES *(ang. effect size)* – czyli to, jak dużą różnicę, zależność czy zgodność uznajemy za wartą wykrycia. Im większy efekt, tym łatwiej go wykazać – rośnie więc moc testu, a potrzebna liczność próby maleje. Na przykład łatwiej przekonać, że dwie grupy się różnią, gdy różnica wynosi pięć poprawnie rozwiązanych zadań, niż gdy wynosi zaledwie jedno zadanie. Wielkość efektu zwykle trzeba zestandaryzować (np. przy pomocy odchylenia standardowego czy współczynnika korelacji), dlatego do obliczeń często potrzebne są dodatkowe informacje o zmienności danych.
- **Poziom istotności statystycznej** α (błąd I rodzaju) – im większy poziom istotności, tym większa moc testu. To jednak element, na który mamy tylko pozorny wpływ – rzadko zdarza się sytuacja, w której wolno go swobodnie zmieniać, a jeśli już, to zwykle w kierunku jego zmniejszenia (patrz Wielokrotne porównania). Standardowo przyjmuje się $\alpha = 0,05$.
- **Kierunek hipotezy**, czyli hipoteza dwustronna (równość w hipotezie zerowej) lub jednostronna (znak $<$ lub $>$ w hipotezie zerowej). Hipoteza jednostronna daje większą moc, ale rzadziej się ją stosuje – w praktyce rzadko zakładamy z góry, że jedna grupa nie ma żadnych szans „wygrać” z drugą; częściej dajemy obu grupom równe szanse.
- **Moc testu** wyraża się prostym wzorem $Moc = 1 - \beta$, gdzie β (błąd II rodzaju) to prawdopodobieństwo przyjęcia hipotezy zerowej, gdy w rzeczywistości jest ona fałszywa.

Moc testu jest też ściśle związana z **licznością próby** n – im liczniejsza próba, tym większa moc: im więcej osób/obiektów zbierzemy do badania, tym łatwiej będzie przekonać, że wykryte różnice między grupami nie są dziełem przypadku, lecz rzeczywiście występują w populacjach. Dlatego zamiast (lub obok) mocy testu badacz może chcieć wyznaczyć **niezbędną liczność próby**, potrzebną, by przy zadanej wielkości efektu osiągnąć określoną moc testu.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji mocy testu i niezbędnej liczności próby dla wybranego testu wywołujemy poprzez menu Statystyka zaawansowana→Moc testu i liczność próby→Moc i liczność dla testu.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Analiza mocy jest bezpośrednio związana z konkretnymi testami statystycznymi, dlatego w oknie tym wybieramy najpierw interesujący nas test, a następnie – w zależności od potrzeb – podajemy liczność próby (by uzyskać moc testu) albo zadaną moc testu (by uzyskać niezbędną liczność próby), a także pozostałe, właściwe dla danego testu, informacje: wielkość efektu (lub dane, na podstawie których można ją wyliczyć), poziom istotności α oraz kierunek hipotezy. Poniżej opisano szczegółowo, jakich danych wymaga analiza mocy i liczności dla poszczególnych testów dostępnych w programie – wraz z przykładami ich zastosowania.

Test t-Studenta dla pojedynczej próby

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu t-Studenta dla pojedynczej próby wykonujemy wtedy, gdy planujemy (lub już wykonaliśmy) porównanie średniej z próby do zadanej wartości hipotetycznej, i chcemy wiedzieć, jak liczną próbę zebrać (lub jaką mamy moc wykrycia różnicy). Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu to w tym przypadku wielkość różnicy między średnią populacji badanej a hipotetyczną wartością średnią. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o niecentralny rozkład t-Studenta.

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Moc testu	Liczność grupy
Liczność grupy	Moc testu
Średnia hipotetyczna	
Średnia grupy	
Odchylenie standardowe grupy	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

PRZYKŁAD 10.3. Chcemy sprawdzić (na poziomie istotności 0.05) czy czas oczekiwania na dostarczenie przesyłki przez pewną firmę kurierską to przeciętnie 3 dni (tzn. 72 godziny).

Badanie a priori – planujemy przeprowadzić badanie, ale nie zebraliśmy jeszcze danych. Moc testu, jaką chcemy uzyskać, ustalamy na poziomie 80%. Odchylenie standardowe, które podamy, powinno odzwierciedlać różnice w czasie dostarczenia przesyłki, jakich spodziewamy się w planowanym badaniu – zakładamy (na bazie doświadczeń pracowników firmy), że wyniesie ono 1 dzień (24 h).

a) Jaka będzie niezbędna liczność próby, gdy założymy, że wielkość efektu, przy którym chcielibyśmy uzyskać istotność statystyczną, wynosi 12 godzin (0.5 dnia)?
b) Jaka będzie niezbędna liczność próby, gdy założymy, że wielkość efektu, przy którym chcielibyśmy uzyskać istotność statystyczną, wynosi 6 godzin (0.25 dnia)?

Odp a) Wiemy, że zakres od 2.5 do 3.5 dnia mieści się w granicach dopuszczonego błędu. Jako średnią hipotetyczną podajemy więc 3 dni, a jako średnią grupy badanej 2.5 dnia (lub 3.5 dnia).

Moc testu i liczność próby	
Czas analizy	0,01 sek.
Poziom istotności	0,05
Test t-Studenta dla pojedynczej próby	
Hipoteza	dwustronna
Moc - zadana	0,8
Średnia hipotetyczna - zadana	3
Średnia grupy - zadana	2,5
Odchylenie standardowe - zadane	1
Wielkość efektu (różnica średnich)	0,5
Moc przybliżona	0,8078
Liczność	34

Niezbędna liczność = 34 przesyłki

Odp Tyle przesyłek trzeba zebrać, by udowodnić, że efekt przekraczający 12 godzin jest istotny statystycznie.

b) Wiemy, że zakres od 2.75 do 3.25 dnia mieści się w granicach dopuszczonego błędu. Jako średnią hipotetyczną podajemy więc 3 dni, a jako średnią grupy badanej 2.75 dnia (lub 3.25 dnia).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Moc testu i liczność próby	
Czas analizy	0,02 sek.
Poziom istotności	0,05
Test t-Studenta dla pojedynczej próby	
Hipoteza	dwustronna
Moc - zadana	0,8
Średnia hipotetyczna - zadana	3
Średnia grupy - zadana	3,25
Odchylenie standardowe - zadane	1
Wielkość efektu (różnica średnich)	0,25
Moc przybliżona	0,8015
Liczność	128

Niezbędna liczność = 128 przesyłek

Tyle przesyłek trzeba zebrać, by udowodnić, że efekt przekraczający 6 godzin jest istotny statystycznie. **Ba-**

danie a posteriori – zebraliśmy dane do badania – nasza próba zawiera 22 przesyłki (dane w pliku kurier.pqs). Na podstawie zebranych danych wyznaczamy średnią liczbę dni oczekiwania na dostarczenie przesyłki i odchylenie standardowe grupy. W naszym przypadku wynoszą one: średnia=3.727273, odchylenie=1.906925.

a) Jaka jest moc przeprowadzonej analizy?
b) Jak wyglądałaby moc, gdybyśmy zwiększali liczność próby do 100 elementów, pozostawiając niezmienione pozostałe założenia?

Odp a) Moc przeprowadzonej analizy wynosi zaledwie 0.400302.

Moc testu i liczność próby	
Czas analizy	0,00 sek.
Poziom istotności	0,05
Test t-Studenta dla pojedynczej próby	
Hipoteza	dwustronna
Liczność - zadana	22
Średnia hipotetyczna - zadana	3
Średnia grupy - zadana	3,7273
Odchylenie standardowe - zadane	1,9069
Wielkość efektu (różnica średnich)	0,7273
Moc	0,4003

Moc = 0.400302

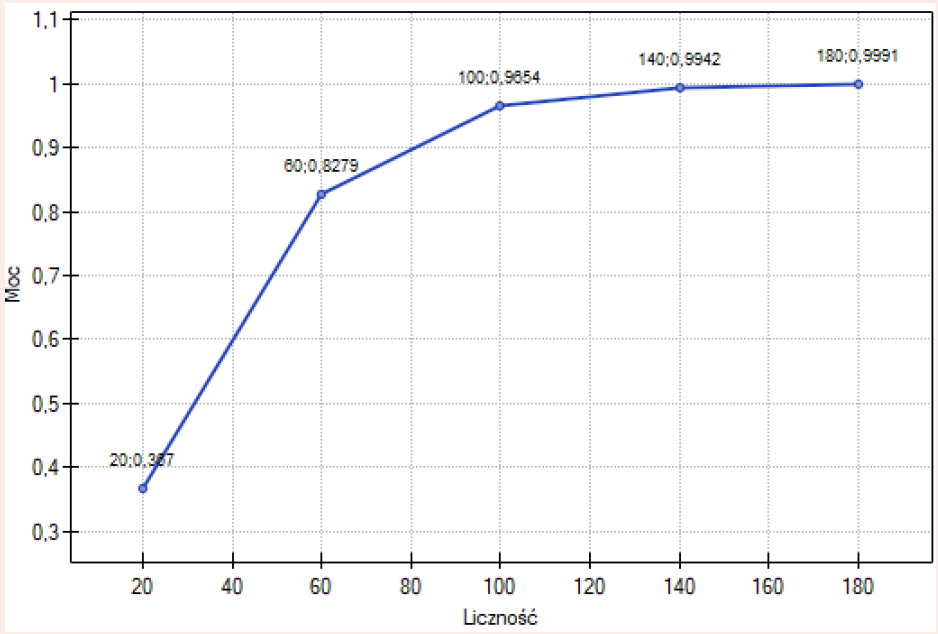
Wiele losowych prób o liczności równej 22 (około 60% takich prób) nie doprowadzi do potwier- Odp dzenia hipotezy alternatywnej.

b) Moc przeprowadzonej przez nas analizy wzrośnie do 0.965364, gdy jej liczność wzrośnie do 100 elementów, a założenia analizy się nie zmienią.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Moc testu i liczność próby	
Czas analizy	0,00 sek.
Poziom istotności	0,05
Test t-Studenta dla pojedynczej próby	
Hipoteza	dwustronna
Liczność - zadana	100
Średnia hipotetyczna - zadana	3
Średnia grupy - zadana	3,7273
Odchylenie standardowe - zadane	1,9069
Wielkość efektu (różnica średnich)	0,7273
Moc	0,9654

Jak zmienia się moc analizy przy zmieniającej się liczności próby i niezmienionych pozostałych założeniach, możemy zobaczyć na wykresie.



Test t-Studenta dla grup zależnych

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu t-Studenta dla grup zależnych wykonujemy, planując lub oceniając badanie, w którym te same osoby/obiekty mierzone są dwukrotnie (np. przed i po interwencji). Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu to w tym przypadku wielkość różnicy między średnimi, jaką spodziewamy się uzyskać w populacji. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o niecentralny rozkład t-Studenta.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Moc testu	Liczność grupy
Liczność grupy	Moc testu
Średnia grupy (badania) 1	
Średnia grupy (badania) 2	
Odchylenie standardowe różnicy (lub odchylenia obu grup i wsp. korelacji Pearsona)	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Przykład 10.4. Chcemy sprawdzić (na poziomie istotności 0.05) czy leczenie zaburzeń odżywiania w pewnej klinice przynosi istotne zmniejszenie masy ciała już po 30 dniach stosowania nowego rodzaju diety. Za istotną zmianę masy ciała uznajemy zmianę wskaźnika BMI o pół jednostki. Jak dużą próbę należy zebrać, by różnica tej wielkości była istotna statystycznie w teście t-Studenta dla grup zależnych?

Ponieważ nie dysponujemy danymi z badania pilotażowego, podstawowe dane do obliczeń podamy na bazie doświadczeń i przypuszczeń pracowników kliniki. Zakładamy, że przeciętna wartość BMI osoby leczonej to 35 – taką wartość wpisujemy w polu dotyczącym pierwszej średniej. Ponieważ zmiana BMI o mniej niż połowę jednostki jest nieistotna klinicznie, dopiero spadek poniżej wartości 34.5 (lub wzrost powyżej wartości 35.5) będzie uznawany za istotny. Jako drugą średnią podajemy więc wartość 34.5 (lub 35.5). Przypuszczamy, że odchylenie standardowe różnicy (BMI przed i BMI po) może być dość duże, gdyż zwykle w grupie znajdą się osoby zdyscyplinowane, przestrzegające diety, i takie, które mimo wszystko pozwalają sobie na dodatkowe przekąski między posiłkami. Dlatego wielkość odchylenia ustalamy na 2.5. Moc analizy, jaką chcemy uzyskać, to 80%.

Moc testu i liczność próby		
Czas analizy	0,02 sek.	0,02 sek.
Poziom istotności	0,05	0,05
Test t-Studenta dla grup zależnych		
Hipoteza	dwustronna	jednostronna
Moc - zadana	0,8	0,8
Średnia grupy 1 - zadana	35	35
Średnia grupy 2 - zadana	34,5	34,5
Odchylenie standardowe różnicy - zadane	2,5	2,5
Wielkość efektu (różnica średnich)	0,5	0,5
Moc przybliżona	0,8017	0,8002
Liczność	199	156

Niezbędna liczność = 199 osób (hipoteza dwustronna) lub 156 osób (hipoteza jednostronna)

Hipoteza dwustronna zakłada, że na skutek diety BMI może spaść lub wzrosnąć; hipoteza jednostronna zakłada jedynie spadek BMI.

Gdybyśmy założyli, że grupa będzie bardziej zdyscyplinowana i odchylenie standardowe różnicy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

wyniesie 1.5 jednostki BMI, wówczas próba będzie mogła być nieco mniejsza.

Moc testu i liczność próby		
Czas analizy	0,01 sek.	0,01 sek.
Poziom istotności	0,05	0,05
Test t-Studenta dla grup zależnych		
Hipoteza	dwustronna	jednostronna
Moc - zadana	0,8	0,8
Średnia grupy 1 - zadana	35	35
Średnia grupy 2 - zadana	34,5	34,5
Odchylenie standardowe różnicy - zadane	1,5	1,5
Wielkość efektu (różnica średnich)	0,5	0,5
Moc przybliżona	0,8023	0,806
Liczność	73	58

Niezbędna liczność = 73 osoby (hipoteza dwustronna) lub 58 osób (hipoteza jednostronna)

Mniejsze zakładane odchylenie standardowe różnicy pozwala uzyskać tę samą moc przy mniejszej próbie.

Test t-Studenta dla grup niezależnych

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu t-Studenta dla grup niezależnych wykonujemy, planując lub oceniając badanie porównujące średnie w dwóch niezależnych grupach (np. chorzy i zdrowi). Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu to w tym przypadku wielkość różnicy między średnimi, jaką spodziewamy się uzyskać pomiędzy populacjami. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o niecentralny rozkład t-Studenta.

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Moc testu	Liczność grupy
Liczność grupy 1 Liczność grupy 2	Moc testu
Średnia grupy 1	
Średnia grupy 2	
Wspólne odchylenie standardowe (lub odchylenia obu grup*)	
Poziom istotności α	

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uwaga! *Gdy znane są liczności obu grup, wspólne odchylenie standardowe wyznaczane jest dokładnym wzorem; gdy liczności nie są znane, stosowany jest uproszczony wzór Cohena (Cohen, J. (1988)[46]) – oba wzory podano w części „Dla zainteresowanych" na końcu rozdziału.

WYKONANIE I INTERPRETACJA

Przykład 10.5. Badamy mężczyzn chorych na chorobę X i zdrowych. Chcemy sprawdzić (na poziomie istotności 0.05) czy chorzy różnią się od zdrowych poziomem cholesterolu HDL. Za istotną klinicznie różnicę uznajemy wielkość 2 mg/dl. Jak dużą próbę należy zebrać, by różnica tej wielkości była istotna statystycznie w teście t-Studenta dla grup niezależnych?

Jako średnią HDL dla mężczyzn chorych podajemy 40 mg/dl, a dla osób zdrowych 42 mg/dl. Stosunek liczności obu grup ustalamy na 1, ponieważ zakładamy grupy równoliczne. Dysponujemy danymi z badania pilotażowego, więc podajemy odchylenie standardowe dla chorych (13 mg/dl) i zdrowych (11 mg/dl), wskazując dodatkowe opcje. Moc analizy, jaką chcemy uzyskać, to 80%.

Moc testu i liczność próby		
Czas analizy	0,06 sek.	0,07 sek.
Poziom istotności	0,05	0,05
Test t-Studenta dla grup niezależnych		
Hipoteza	dwustronna	dwustronna
Moc - zadana	0,8	0,8
Stosunek liczności - zadany	1	2
Średnia grupy 1 - zadana	42	42
Średnia grupy 2 - zadana	40	40
Wspólne odchylenie standardowe - zadane	12,0416	12,0416
Wielkość efektu (różnica średnich)	2	2
Moc przybliżona	0,8007	0,8004
Liczność	571 / 571	855 / 428

Niezbędna liczność = 571 osób (grupy równoliczne, n1/n2=1) lub n1=855, n2=428 (n1/n2=2)

Im mniej równoliczne mają być grupy, tym większa łączna liczba osób jest potrzebna do zachowania tej samej mocy testu.

Test chi-kwadrat wariancji pojedynczej próby

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu chi-kwadrat wariancji pojedynczej próby wykonujemy, planując lub oceniając badanie porównujące zmienność (odchylenie standardowe) w próbie z zadaną wartością hipotetyczną. Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zadana wielkość efektu to w tym przypadku iloraz odchylenia standardowego populacji badanej i odchylenia hipotetycznego. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o rozkład chi-kwadrat.

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Moc testu	Liczność grupy
Liczność grupy	Moc testu
Hipotetyczne odchylenie standardowe	
Odchylenie standardowe grupy	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Przykład 10.6. Przed przystąpieniem do produkcji kolejnej partii pewnego syropu przeciwkaszlowego należy dokonać kontrolnych pomiarów objętości syropu nalewanego do butelek. W butelkach powinno się znaleźć 200 ml syropu. Z dokumentacji technicznej urządzenia dozującego wynika, że dopuszczalny rozrzut objętości syropu mierzony za pomocą odchylenia standardowego wynosi 1 ml. Należy sprawdzić (na poziomie istotności 0.05), czy testowane urządzenie działa prawidłowo. Czy próbka składająca się z 20 butelek będzie wystarczająca, by wykazać nadmierny błąd urządzenia, jeśli taki zaistnieje? Za nadmierny błąd urządzenia uznano odchylenie standardowe przekraczające 1.2 ml.

Ponieważ oczekujemy, że odchylenie standardowe dla urządzenia dozującego będzie takie jak w dokumentacji, wpisujemy wartość 1 ml jako wartość hipotetyczną. Zbyt duży błąd uzyskamy wtedy, gdy odchylenie przekroczy 1.2 ml, i tę wartość wpisujemy w polu dotyczącym odchylenia standardowego grupy butelek, którą będziemy badać.

Jeśli próba będzie liczyła 20 butelek, uzyskana moc przy hipotezie dwustronnej wyniesie zaledwie 0.25, a przy hipotezie jednostronnej 0.34.

Moc testu i liczność próby		
Czas analizy	0,01 sek.	0,01 sek.
Poziom istotności	0,05	0,05
Test Chi-kwadrat wariancji pojedynczej próby		
Hipoteza	dwustronna	jednostronna
Liczność - zadana	20	20
Hipotetyczne odch. standardowe - zadana	1	1
Odchylenie standardowe - zadane	1,2	1,2
Wielkość efektu (iloraz odchyleń)	0,8333	0,8333
Moc	0,2482	0,3405

Moc = 0.25 (dwustronna) / 0.34 (jednostronna) dla n=20

Wartości niskie – mniej niż 35% losowych prób tej wielkości pozwoli wykryć przekroczenie rozrzutu o 0.2 ml. 20 butelek to zbyt mała grupa, by udowodnić zbyt duże odchylenie standardowe urządzenia, jeśli takie rzeczywiście zaistnieje.

Chcielibyśmy uzyskać standardową moc. By uzyskać moc równą 80%, zmieniamy ustawienia programu

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

i wyliczamy niezbędną liczność.

Moc testu i liczność próby		
Czas analizy	1,03 sek.	0,47 sek.
Poziom istotności	0,05	0,05
Test Chi-kwadrat wariancji pojedynczej próby		
Hipoteza	dwustronna	jednostronna
Moc - zadana	0,8	0,8
Hipotetyczne odch. standardowe - zadana	1	1
Odchylenie standardowe - zadane	1,2	1,2
Wielkość efektu (iloraz odchyleń)	0,8333	0,8333
Moc przybliżona	0,8027	0,8032
Liczność	115	92

Niezbędna liczność = 115 (dwustronna) / 92 (jednostronna) przy mocy 80%

Tyle butelek trzeba skontrolować, by przy zachowaniu mocy testu na poziomie 80% wykazać przekroczenie dopuszczalnego rozrzutu.

Test chi-kwadrat dwóch wariancji Fishera-Snedecora

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu Fishera-Snedecora wykonujemy, planując lub oceniając badanie porównujące zmienność (odchylenie standardowe) między dwiema niezależnymi próbami. Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu to w tym przypadku iloraz odchylenia standardowego populacji pierwszej i drugiej. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o rozkład F-Snedecora.

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Moc testu	Liczność grupy
Liczność grupy 1 Liczność grupy 2	Moc testu
Odchylenie standardowe grupy 1	
Odchylenie standardowe grupy 2	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przykład 10.7. Przed przystąpieniem do produkcji kolejnej partii pewnego syropu przeciwkaszlowego należy dokonać kontrolnych pomiarów objętości syropu nalewanego do butelek. W butelkach powinno się znaleźć 300 ml syropu. W rozlewni stosowane są dwa urządzenia dozujące. Chcemy sprawdzić (na poziomie istotności 0.05), czy rozrzut objętości syropu mierzony za pomocą odchylenia standardowego dla obu urządzeń jest taki sam. Przeprowadzono niewielkie badanie pilotażowe i ustalono, że odchylenie standardowe dla pierwszego urządzenia wynosi 1.32, a dla drugiego 1.1. Jeśli różnica będzie niewielka, tzn. iloraz obu odchyleń wyniesie poniżej 1.2 (tak jak w badaniu pilotażowym), oba urządzenia będą wykorzystywane zamiennie, jeśli nie – zostanie wybrane to z mniejszym odchyleniem od średniej. Ile losowo wybranych butelek należy zmierzyć, by móc wykazać, że iloraz wielkości 1.2 jest istotny statystycznie?

Wpisujemy wartości odchyleń standardowych uzyskanych w badaniu pilotażowym i zakładamy 80% mocy testu.

Moc testu i liczność próby		
Czas analizy	0,12 sek.	0,18 sek.
Poziom istotności	0,05	0,05
Test Chi-kwadrat dwóch wariancji Fishera-Snedecora		
Hipoteza	dwustronna	dwustronna
Moc - zadana	0,8	0,8
Stosunek liczności - zadany	1	2
Odch. standardowe grupy 1 - zadane	1,32	1,32
Odch. standardowe grupy 2 - zadane	1,1	1,1
Wielkość efektu (iloraz odchyleń)	0,8333	0,8333
Moc przybliżona	0,8015	0,8006
Liczność	239 / 239	363 / 182

Niezbędna liczność = n1=n2=239 (grupy równoliczne, n1/n2=1) lub n1=363, n2=182 (n1/n2=2)

Tyle butelek dla każdego urządzenia trzeba zmierzyć, by przy 80% mocy testu wykazać istotny statystycznie iloraz odchyleń standardowych równy 1.2.

Test chi-kwadrat zgodności

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu chi-kwadrat zgodności wykonujemy, planując lub oceniając badanie sprawdzające, czy rozkład danych w próbie jest zgodny z rozkładem oczekiwanym (teoretycznym lub hipotetycznym). Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu, czyli ϕ, to w tym przypadku pierwiastek z ilorazu statystyki testu chi-kwadrat i liczności grupy badanej – wzór podano w części „Dla zainteresowanych" na końcu rozdziału. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o niecentralny rozkład chi-kwadrat.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać[46]:

Moc testu	Liczność grupy
Liczność grupy Liczba kategorii	Moc testu
Wielkość efektu ϕ	
Poziom istotności α	

Test chi-kwadrat (RxC)

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu chi-kwadrat dla tabel RxC wykonujemy, planując lub oceniając badanie sprawdzające zależność dwóch cech kategorialnych przedstawionych w tabeli o wielu wierszach i kolumnach. Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu, czyli ϕ, to w tym przypadku pierwiastek z ilorazu statystyki testu chi-kwadrat i liczności grupy badanej – wzór podano w części „Dla zainteresowanych" na końcu rozdziału. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o niecentralny rozkład chi-kwadrat.

JAK USTAWIĆ ANALIZĘ

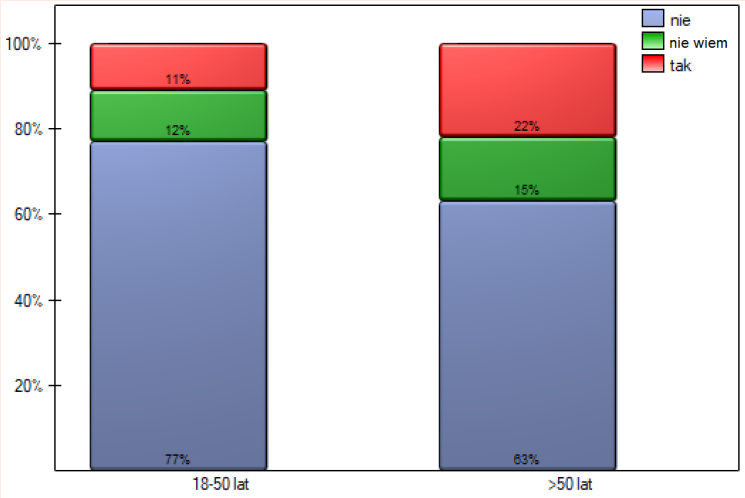
Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać[46][132]:

Moc testu	Liczność grupy
Liczność grupy Liczba kategorii (wierszy) Liczba kategorii (kolumn)	Moc testu
Wielkość efektu ϕ	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Przykład 10.8. Istnieją plany przeprowadzenia szerokiego badania pokazującego stan wiedzy populacji Polski na temat sposobów walki z popularnymi wirusami. W ramach projektu należy ustalić czy działania edukacyjne informujące o nieskuteczności antybiotykoterapii w zakażeniach wirusowych były tak samo skuteczne w starszej grupie wiekowej (tzn. powyżej 50 roku życia) jak u młodszych dorosłych (18-50 lat). Przeprowadzono badanie pilotażowe i w losowo wybranej próbie 200 osób zadano pytanie: „Czy antybiotyki zwalczają wirusy?". Respondenci mieli możliwość wskazania jednej z trzech odpowiedzi: „nie", „nie wiem", „tak". Wyniki badania pilotażowego przygotowano do publikacji. Poniżej zamieszczono fragment opisu umieszczonego w pracy:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



	nie	nie wiem	tak
18-50 lat	77	12	11
>50 lat	63	15	22

Uzyskana wartość p w teście chi-kwadrat była nieistotna statystycznie: p=0.0672. Słuszne zaskoczenie recenzenta pracy budzi informacja, iż dwukrotnie więcej osób powyżej 50 roku życia błędnie wskazało, że antybiotyki zwalczają wirusy (22% w stosunku do 11%), jednak różnica ta nie była istotna statystycznie.

Zgodnie z sugestią recenzenta należy sprawdzić, czy brak istotności statystycznej dla tej różnicy nie wynika ze zbyt niskiej mocy testu oraz podać jak duża powinna być próba, by przy zachowaniu tych samych odsetków uzyskać moc testu chi-kwadrat na poziomie 80%.

Przygotowanie odpowiedzi dla recenzenta

Wyznaczymy współczynnik ϕ dla uzyskanego testu (menu: Chi-kwadrat, Fisher, OR/RR→Współczynniki korelacji...→Phi). Uzyskujemy ϕ=0.1643. Na podstawie 200-elementowej próby, o danych umieszczonych w tabeli o dwóch wierszach i trzech kolumnach, oraz zadanej wielkości współczynnika ϕ wyznaczamy moc testu chi-kwadrat.

Moc testu i liczność próby	
Czas analizy	0,01 sek.
Poziom istotności	0,05
Test Chi-kwadrat CxR	
Hipoteza	dwustronna
Liczność - zadana	200
Liczba kategorii (wiersze) - zadana	2
Liczba kategorii (kolumny) - zadana	3
Wielkość efektu (Phi) - zadane	0,1643
Wartość statystyki testu chi-kwadrat	5,3989
Moc	0,5368

Moc = 0.5368

Jeśli

Wartość niewielka, co potwierdza obawy o niewystarczająco liczną próbę.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

dla próby o innej liczności uzyskamy ten sam rozkład danych, oznacza to, że uzyskamy również ten sam współczynnik ϕ. Do wyznaczenia liczności próby, która dałaby nam 80% moc testu chi-kwadrat, podajemy ponownie współczynnik ϕ=0.1643.

Moc testu i liczność próby	
Czas analizy	0,12 sek.
Poziom istotności	0,05
Test Chi-kwadrat CxR	
Hipoteza	dwustronna
Moc - zadana	0,8
Liczba kategorii (wiersze) - zadana	2
Liczba kategorii (kolumny) - zadana	3
Wielkość efektu (Phi) - zadana	0,1643
Wartość statystyki testu chi-kwadrat	9,637
Moc przybliżona	0,8001
Liczność	357

Niezbędna liczność = 357 osób (przy mocy 80%)

Tylu ankietowanych potrzeba, by przy zachowanym rozkładzie odpowiedzi wykazać istotność statystyczną z mocą testu 80%. Ponieważ jest to dopiero badanie pilotażowe, planujemy zwiększyć grupę badaną do 357 osób w badaniu właściwym. Już w tej chwili zauważamy, że przy pominięciu osób niezdecydowanych (tzn. tych, którzy wybrali odpowiedź „nie wiem") i ponownym przeprowadzeniu analizy, można wykazać istotne różnice (chi-kwadrat, p=0.0251). W grupie osób zdecydowanych odsetki wybierających błędną odpowiedź są ponad dwukrotnie większe na niekorzyść osób powyżej 50 lat (12.5% vs 25.9%).

Test dwóch niezależnych proporcji, Chi-kwadrat (2x2)

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu chi-kwadrat dla tabel 2x2 i testu Z dla dwóch niezależnych proporcji wykonujemy, planując lub oceniając badanie porównujące dwa odsetki (proporcje) w dwóch niezależnych grupach. Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania obu tych testów.

Zadana wielkość efektu to różnica między wyróżnionymi proporcjami. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o rozkład normalny.

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać[36]:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Moc testu	Liczność grupy
Liczność grupy	Moc testu
Wyróżnioną proporcję w pierwszej grupie	
Wyróżnioną proporcję w drugiej grupie	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Przykład 10.9. Rozważmy badanie oceniające skuteczność aspiryny w zmniejszaniu śmiertelności z powodu zawału mięśnia sercowego. Wcześniejsze badania wskazują, że odsetek zgonów z powodu zawału serca wynosi 0.015 dla osób nieużywających i 0.001 dla użytkowników aspiryny. Badacze chcą określić minimalną wielkość próbki wymaganą do wykrycia absolutnej różnicy $|0.001 - 0.015| = 0.014$ przy 80% mocy za pomocą dwustronnego testu o poziomie istotności 5%.

Moc testu i liczność próby	
Czas analizy	0.00 sek.
Poziom istotności	0.05
Test dwóch niezależnych proporcji, Chi-kwadrat (2x2	
Hipoteza	dwustronna
Moc - zadana	0.8
Stosunek liczności - zadany	1
Proporcja grupy 1 - zadane	0.001
Proporcja grupy 2 - zadane	0.015
Wielkość efektu (proporcja 1 - proporcja 2)	-0.014
Moc przybliżona	0.8004
Liczność	635 / 635

Niezbędna liczność = 635 osób w każdej grupie (grupy równoliczne)

Tyle osób trzeba zebrać w każdej z grup, by przy 80% mocy testu wykazać istotność statystyczną różnicy odsetków wynoszącej 0.014.

ANOVA dla grup niezależnych

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla ANOVA dla grup niezależnych wykonujemy, planując lub oceniając badanie porównujące średnie w trzech lub więcej niezależnych grupach. Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tej metody.

Zadana wielkość efektu to w tym przypadku RMSSE, czyli miara standaryzowana stosowana w ANOVA do opisania ogólnego poziomu efektu w populacji. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o niecentralny rozkład F-Snedecora.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Moc testu	Liczność grupy
Liczność grupy	Moc testu
Średnia grupy 1	
Średnia grupy 2	
Średnia grupy 3	
Średnia grupy 4	
Odchylenie standardowe grupy 1	
Odchylenie standardowe grupy 2	
Odchylenie standardowe grupy 3	
Odchylenie standardowe grupy 4	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Przykład 10.10. Badano parametr FVC dla pacjentów z wadami serca (wada serca A, wada serca B i wada serca C). Chcemy się dowiedzieć (na poziomie istotności 0.05), czy pacjenci różnią się wartościami tego parametru. By to sprawdzić, przeprowadzono najpierw badanie pilotażowe. Na podstawie wyników tego badania ustalono przewidywane wielkości efektów, tzn.:
– dla wady serca A: średnia = 3.8, odchylenie standardowe = 1.1,
– dla wady serca B: średnia = 4.5, odchylenie standardowe = 0.6,
– dla wady serca C: średnia = 4.2, odchylenie standardowe = 0.9.
Ile osób należy zebrać, jeśli wielkości te utrzymają się na tym samym poziomie, by udowodnić, że występują istotne statystycznie różnice?

Wpisujemy wartości średnich i odchyleń standardowych.

Moc testu i liczność próby	
Czas analizy	0,03 sek.
Poziom istotności	0,05
Jednoczynnikowa ANOVA dla grup niezależnych	
Hipoteza	dwustronna
Moc - zadana	0,8
Liczba grup - zadana	3
Średnia 1 - zadana	3,8
Średnia 2 - zadana	4,5
Średnia 3 - zadana	4,2
Odchylenie std. 1 - zadana	1,1
Odchylenie std. 2 - zadana	0,6
Odchylenie std. 3 - zadana	0,9
Wielkość efektu (RMSSE)	0,3943
Moc przybliżona	0,8131
Liczność	33

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Niezbędna liczność = 33 osoby w każdej grupie (przy mocy 80%)

Tyle osób trzeba zebrać w każdej z grup badanych, by przy zachowaniu przewidywanych wielkości efektów uzyskać moc testu na poziomie 80%.

Test dla jednej proporcji

KIEDY STOSOWAĆ:

Analizę mocy i niezbędnej liczności próby dla testu dla jednej proporcji wykonujemy, planując lub oceniając badanie porównujące proporcję (odsetek) w próbie z zadaną wartością oczekiwaną. Zanim przystąpimy do obliczeń, warto zapoznać się z zasadami stosowania tego testu.

Zadana wielkość efektu to w tym przypadku wielkość różnicy między zadaną proporcją w populacji badanej a hipotetyczną, oczekiwaną proporcją. Moc testu oraz niezbędna liczność próby wyliczane są w oparciu o rozkład normalny, gdy korzystamy z testu asymptotycznego, lub rozkład dwumianowy, gdy korzystamy z testu dokładnego.

JAK USTAWIĆ ANALIZĘ

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy podać:

Moc testu	Liczność grupy
Liczność grupy	Moc testu
Proporcja oczekiwana	
Proporcja grupy	
Poziom istotności α	

WYKONANIE I INTERPRETACJA

Przykład 10.11. Zapytano 10 wylosowanych rodzin posiadających dzieci w wieku poniżej 10 lat i zamieszkujących w Poznaniu o planowaną przyszłość edukacyjną swoich dzieci. Spośród nich 6 rodzin planowało edukację swoich dzieci na uczelniach wyższych. Jak dużą musielibyśmy zebrać próbę, by na poziomie istotności 0.05 i przy mocy testu równej 0.8 można było stwierdzić, że więcej niż 50% poznańskich rodzin posiadających dzieci w wieku poniżej 10 lat już teraz planuje ich przyszłą edukację na uczelniach wyższych?

Jako oczekiwaną proporcję wpisujemy 0.5, a proporcję grupy 0.6.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Moc testu i liczność próby		
Czas analizy	0,02 sek.	0,01 sek.
Poziom istotności	0,05	0,05
Test Z i test dokładny dla jednej proporcji		
Hipoteza	dwustronna	jednostronna
Moc - zadana	0,8	0,8
Proporcja oczekiwana - zadana	0,5	0,5
Proporcja grupy - zadane	0,6	0,6
Wielkość efektu (różnica proporcji)	0,1	0,1
Moc przybliżona	0,8003	0,8013
Liczność (asymptotycznie)	194	153

Niezbędna liczność = 194 rodziny (hipoteza dwustronna) lub 151 rodzin (hipoteza jednostronna)

Tyle rodzin trzeba zebrać, by przy mocy testu 80% wykazać, że odsetek planujących wyższe wykształcenie dzieci przekracza 50%.

DLA ZAINTERESOWANYCH

Test t-Studenta dla grup niezależnych – wspólne odchylenie standardowe
Wzory: gdy znane są liczności obu grup, wspólne odchylenie standardowe wyraża się wzorem:

$$SD_p = \sqrt{\frac{(n_1-1)sd_1^2 + (n_2-1)sd_2^2}{n_1+n_2-2}},$$

gdy liczności nie są znane, stosowany jest uproszczony wzór Cohena (Cohen, J. (1988)[46]):

$$SD_p = \sqrt{\frac{sd_1^2 + sd_2^2}{2}}.$$

gdzie:
n_1, n_2 – liczności w pierwszej i drugiej próbie,
sd_1, sd_2 – odchylenia standardowe w pierwszej i drugiej próbie.

Test chi-kwadrat zgodności oraz test chi-kwadrat (RxC) – wielkość efektu ϕ
Wzory: zadana wielkość efektu ϕ to pierwiastek z ilorazu statystyki testu chi-kwadrat i liczności grupy badanej:

$$\phi = \sqrt{\frac{\chi^2}{n}}.$$

11 ANALIZY OPISOWE

Dane zebrane przez badacza powinny w pierwszej kolejności zostać opisane. To, jakich miar użyjemy do tego opisu, zależy od sposobu dokonywania pomiarów, czyli od tzw. skali pomiarowej zmiennej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

11.1 Skale pomiarowe

Poprawne określenie rodzaju wykonywanej analizy zależy od skali, na jakiej wyrażone są zebrane dane. Wyróżniamy 3 główne skale pomiarowe:

1. **Skala interwałowa (przedziałowa)** (*ang. interval scale*)

   Zmienna jest wyrażona na skali interwałowej, gdy:

   - można ją uporządkować,
   - można obliczyć, o ile jeden element jest większy od drugiego, a różnica ta ma interpretację w świecie rzeczywistym. Zwykle określona jest jednostka miary.

   Przykład: masa obiektu [kg], powierzchnia obiektu [m], czas [lata], prędkość [km/h] itp.

2. **Skala porządkowa** (*ang. ordinal scale*)

   Zmienna jest wyrażona na skali porządkowej, gdy:

   - można ją uporządkować, czyli ma znaczenie kolejność występowania elementów,
   - nie da się w sensowny sposób określić różnicy ani ilorazu między dwiema wartościami.

   Przykład: wykształcenie, kolejność zawodników na podium itp.

   Uwaga! Jeśli zmienna jest wyrażona na skali porządkowej, to by można było wykonać na niej w sposób prawidłowy obliczenia, powinna być zapisana za pomocą liczb. Liczby te to umowne identyfikatory mówiące o kolejności elementów.

3. **Skala nominalna** (*ang. nominal scale*)

   Zmienna jest wyrażona na skali nominalnej, gdy:

   - nie można jej uporządkować, czyli nie istnieje wynikające z natury danego zjawiska uporządkowanie,
   - nie da się w sensowny sposób określić różnicy ani ilorazu między dwiema wartościami.

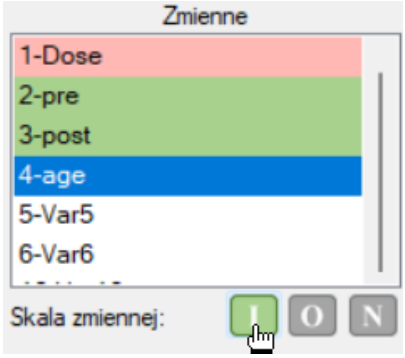
   Przykład: płeć, kraj zamieszkania itp.

   Uwaga! Jeśli zmienna jest wyrażona na skali nominalnej, to może być zapisana za pomocą etykiet tekstowych. Nawet jeśli wartości zmiennej nominalnej są wyrażane liczbowo, to liczby te są tylko umownymi identyfikatorami – nie można więc wykonywać na nich działań arytmetycznych ani ich porównywać.

Przed przystąpieniem do analiz zaleca się przyporządkować skalę pomiarową poszczególnym zmiennym. Takie przypisanie spowoduje, że nagłówki zmiennych zyskają odpowiadający danej skali kolor, tzn. kolor zielony = **skala interwałowa**, kolor żółty = **skala porządkowa**, kolor czerwony = **skala nominalna**. Kolor zmiennych (a więc ich skala) będzie widoczny w arkuszu danych i na liście zmiennych w oknach analiz.

Przypisania skali do wybranej zmiennej możemy dokonać w oknie opcji zmiennych Kody/Etykiety/Format lub w menu kontekstowym na nagłówku wybranej zmiennej.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Dane **ilościowe** to wszelkie informacje, które można określić ilościowo, policzyć lub zmierzyć i nadać im wartość liczbową. Do danych ilościowych zaliczamy skalę interwałową i niekiedy również skalę porządkową.
Dane **jakościowe** mają charakter opisowy, wyrażony raczej w kategoriach językowych niż w wartościach liczbowych. Do danych jakościowych zaliczamy skalę nominalną i niekiedy również skalę porządkową.

Skala porządkowa, która ma wiele zobiektywizowanych kategorii, to dane ilościowe, np. skala jakości życia SF-36 (od 0 do 100 punktów). Jeśli jednak kategorii jest tak niewiele, że można je opisać tekstem, np. wykształcenie (podstawowe, średnie, wyższe), to są to dane jakościowe.

11.2 Tabele

Tabele liczności i rozkłady empiryczne danych

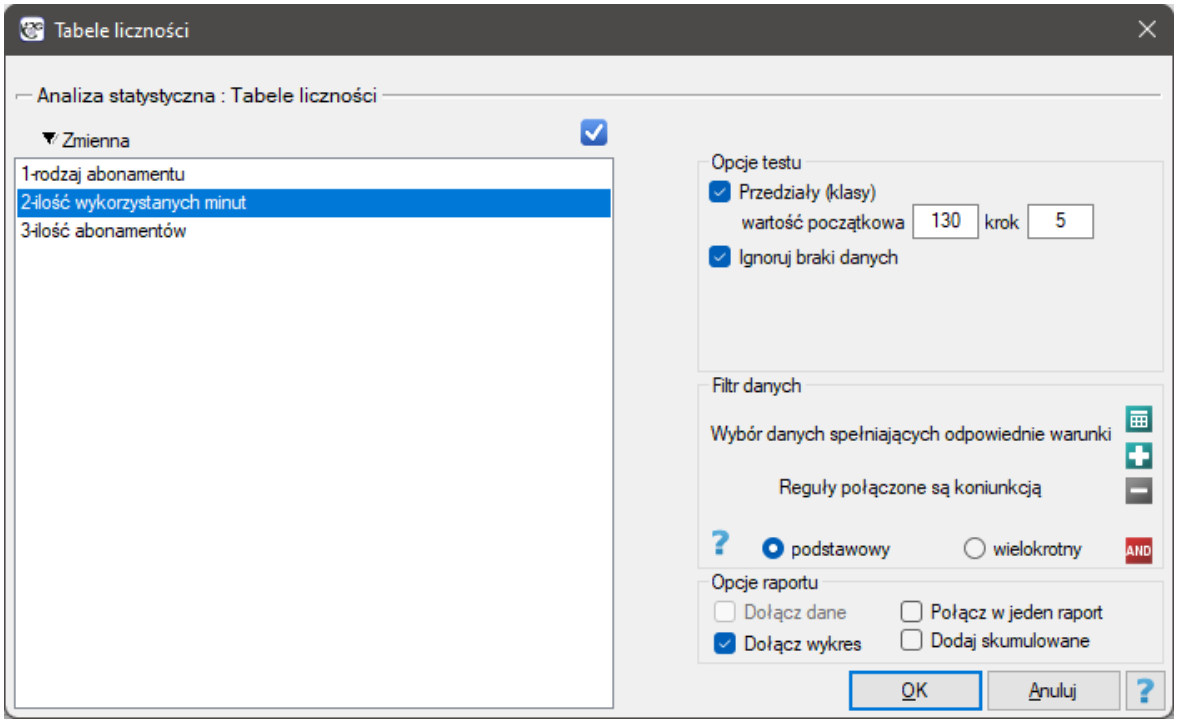
KIEDY STOSOWAĆ:

Podstawą badań statystycznych jest określenie **rozkładu empirycznego**, czyli zaobserwowanego w próbie rozkładu cechy – innymi słowy, sprawdzenie, jak często poszczególne wartości cechy się powtarzają. Rozkład taki przedstawia się w postaci **tabeli liczności** lub wykresu (histogramu). Dla małych zbiorów danych tabele liczności mogą prezentować wszystkie dane – tzw. szeregi rozdzielcze punktowe, a dla większych zbiorów danych tworzy się tzw. szeregi rozdzielcze przedziałowe.

JAK USTAWIĆ ANALIZĘ

Statystyka → Analizy opisowe → Tabele liczności.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W oknie tym wybieramy zmienną do analizy oraz opcje analizy. Wybierając odpowiednie opcje, zwrócony wynik możemy posortować, traktując zmienne jako wartości tekstowe lub jako liczby. Jeśli w analizowanej kolumnie występują puste komórki, to mogą być wliczane do analizy bądź pomijane. Wynik analizy znajdzie się w raporcie dołączonym do arkusza danych, dla których została wykonana.

Jeśli dodatkowo chcemy, by dane zostały zobrazowane wykresem kolumnowym lub histogramem, zaznaczamy w oknie Tabel liczności opcję Dołącz wykres.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 11.1. (plik rozkład.pqs)
Pewien operator telefonii komórkowej przeprowadza szereg badań dotyczących wykorzystania przez klientów liczby przyznanych w abonamencie „darmowych minut". Klienci w każdym miesiącu mogą wykorzystać do 190 takich minut. Badanie przeprowadzono na podstawie losowej próby 200 klientów. Analizowano między innymi informacje o:
– rodzaju wykupionego abonamentu,
– liczbie wykorzystanych darmowych minut,
– liczbie zarejestrowanych na danego klienta abonamentów (nie dotyczy firm).

Chcemy przedstawić rozkład:

1. rodzaju abonamentu,
2. liczby wykorzystanych darmowych minut,
3. liczby zarejestrowanych abonamentów na osoby prywatne.

Uruchamiamy okno Tabele liczności.

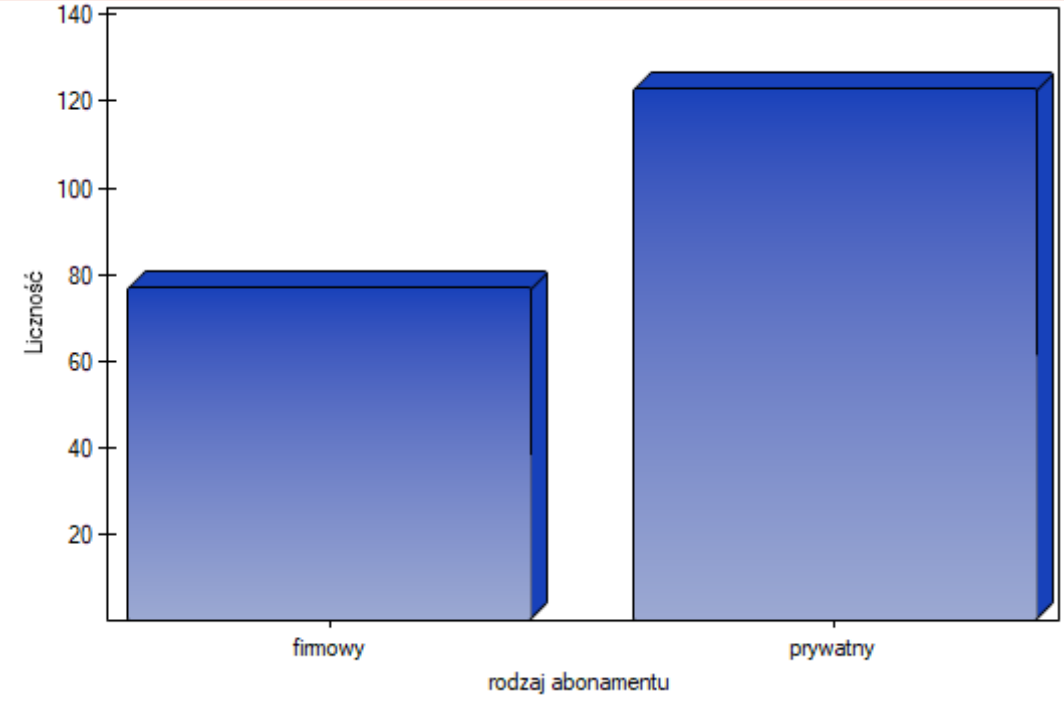
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

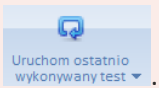
1. Wybieramy Zmienną do analizy: „rodzaj abonamentu” i zaznaczamy opcję Dołącz wykres. Następnie potwierdzamy wybrane ustawienia przyciskiem OK i uzyskujemy wynik w postaci raportu:

Tabele liczności

Analizowane zmienne	rodzaj aboname
Liczność / braki danych	0
Liczność / przypadki dezaktywowane	0

	Liczność	Procent
firmowy	77	38.5%
prywatny	123	61.5%
Razem	200	100%



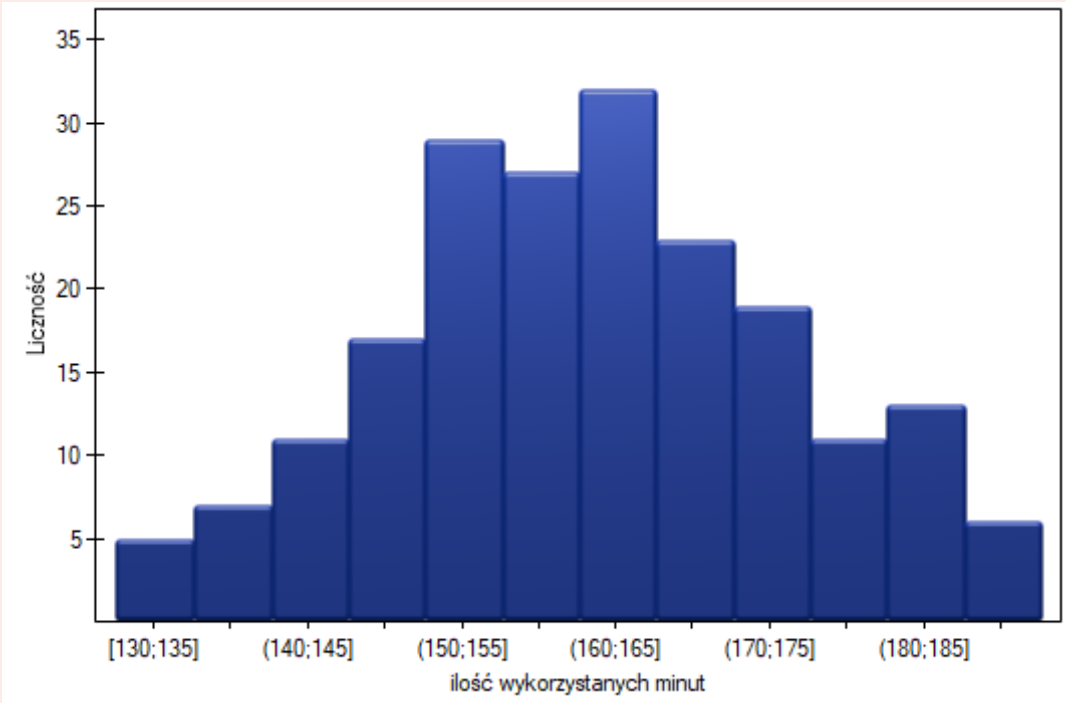
2. Wznawiamy analizę przyciskiem 

. Wybieramy Zmienną do analizy: „ilość wykorzystanych minut” i zaznaczamy opcję Przedziały (klasy), wartość początkową ustawiamy np. na 130, a krok na 5. Możemy również zaznaczyć opcję Dołącz wykres. Następnie wybrane opcje potwierdzamy przyciskiem OK i uzyskujemy wynik w postaci raportu:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Tabele liczności		
Analizowane zmienne	ilość wykorzyst:	
Liczność / braki danych	0	
Liczność / przypadki dezaktywowane	0	

	Liczność	Procent
[130;135]	5	2.5%
(135;140]	7	3.5%
(140;145]	11	5.5%
(145;150]	17	8.5%
(150;155]	29	14.5%
(155;160]	27	13.5%
(160;165]	32	16%
(165;170]	23	11.5%
(170;175]	19	9.5%
(175;180]	11	5.5%
(180;185]	13	6.5%
(185;190]	6	3%
Razem	200	100%



3. Wznawiamy analizę przyciskiem . Ustawiamy filtr tak, by analiza była wykonana wyłącznie dla osób prywatnych. Wybieramy Zmienną do analizy: „Ilość abonamentów”. Ponieważ zmienna ta zawiera również braki danych, uzyskany wynik może uwzględniać te braki w analizie lub nie, w zależności od wybranej opcji dotyczącej ignorowania pustych przypadków:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Tabele liczności		
Analizowane zmienne	ilość abonamen	
Filtr danych:	rodzaj abonai	
Liczność / braki danych	5	
Liczność / przypadki dezaktywowane	0	
	Liczność	Procent
1	94	79.661%
2	12	10.169%
3	11	9.322%
4	1	0.847%
Razem	118	100%

Tabele liczności		
Analizowane zmienne	ilość abonamen	
Filtr danych:	rodzaj abonai	
Liczność / braki danych	5	
Liczność / przypadki dezaktywowane	0	
	Liczność	Procent
1	94	76.423%
2	12	9.756%
3	11	8.943%
4	1	0.813%
puste	5	4.065%
Razem	123	100%

PRZYKŁAD 11.2. (plik nawozy.pqs)
Przeprowadzono doświadczenie, w którym badano stan mikrobiologiczny gleby pod uprawą życicy trwałej zasilanej nawozami biologicznie aktywnymi. Gleby nawożono różnymi rodzajami preparatów mikrobiologicznych i nawozów, a następnie wyliczono ilość mikroorganizmów występujących w 1 gramie suchej masy gleby. Chcemy znać częstość występowania promieniowców na 1 gram suchej masy gleby nawożonej azotem – interesuje nas, jak często w badanej próbie występowało od 0 do 20 promieniowców, od więcej niż 20 do 40 promieniowców, od więcej niż 40 do 60 promieniowców, itd. Zaznaczamy w arkuszu danych tylko 54 pierwsze wiersze, które odpowiadają założeniom analizy (są to promieniowce nawożone azotem), i uruchamiamy okno Tabele liczności.

W oknie opcji wybieramy zmienną do analizy: Ilość mikroorganizmów, a następnie ustawiamy przedziały klasowe w ten sposób, by wartością początkową było 0, a krokiem 20. Na górze okna powinien być widoczny komunikat: Dane ograniczone przez zaznaczenie. Potwierdzamy wybór przyciskiem OK i uzyskujemy wynik w postaci raportu:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Tabele liczności		
Analizowane zmienne	ilość mikroorga	
Liczność / braki danych	0	
Liczność / przypadki dezaktywowane	0	
	Liczność	Procent
[0;20]	1	1.852%
(20;40]	3	5.556%
(40;60]	6	11.111%
(60;80]	21	38.889%
(80;100]	16	29.63%
(100;120]	4	7.407%
(120;140]	3	5.556%
Razem	54	100%

Raport tabeli

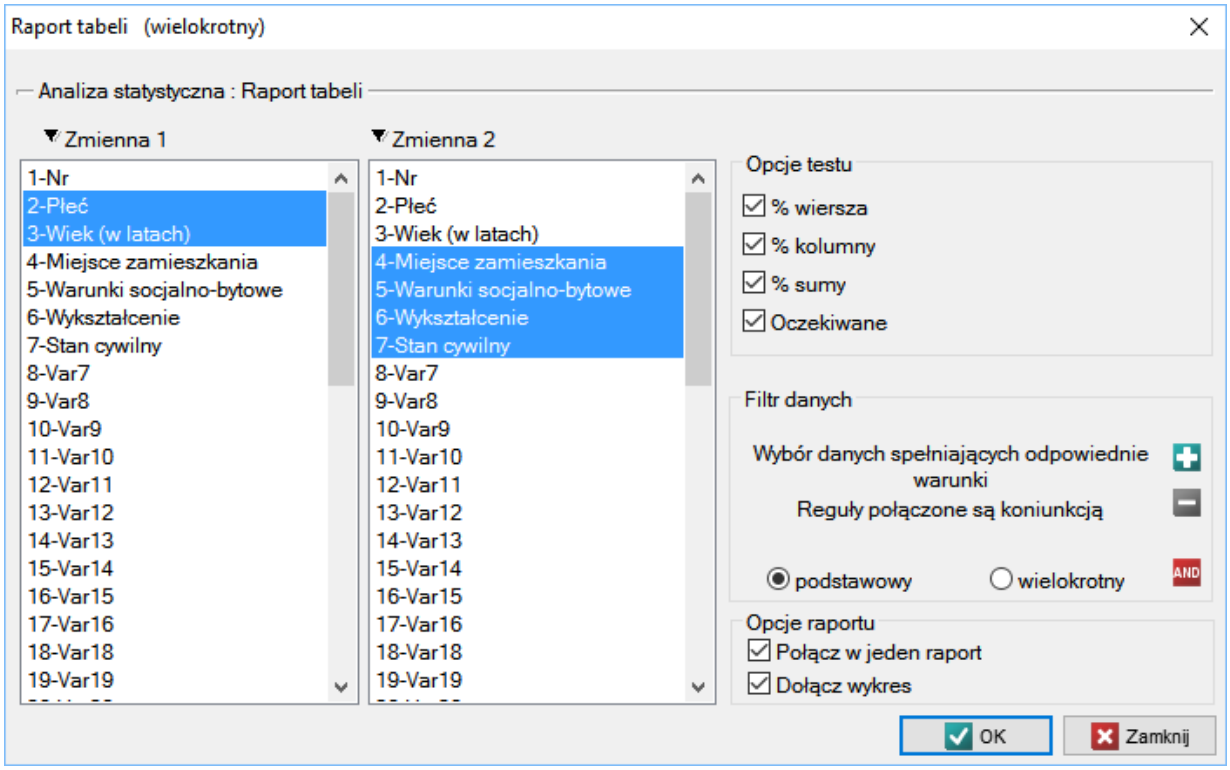
KIEDY STOSOWAĆ:

Raport tabeli pozwala jednocześnie podsumować bardzo wiele danych w postaci tabel dwudzielczych (tabel dwóch cech). Na przykład w ten sposób możemy przedstawić rozkład grup wiekowych według miejsca zamieszkania, wykształcenia itd. Każda tabela jest przedstawiana w postaci liczności w poszczególnych kategoriach oraz dodatkowo można ją podsumować, wyliczając procenty z wiersza, z kolumny lub z sumy całkowitej, a także wyznaczyć tabelę liczności oczekiwanych. Dla takich tabel możliwe jest też automatyczne podsumowanie w postaci wykresu kolumnowego.

JAK USTAWIĆ ANALIZĘ

Statystyka → Analizy opisowe → Raport tabeli.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

PRZYKŁAD 11.3. (plik Tabele.pqs)
W postaci tabel należy podsumować rozkład płci według miejsca zamieszkania, warunków socjalno-bytowych, wykształcenia, stanu cywilnego oraz rozkład grup wiekowych względem tych samych cech. W rezultacie uzyskamy po 4 tabele dla każdej pary cech, czyli 8 tabel dla wszystkich par i odpowiadające im wykresy. Poniżej przedstawiono tylko zestawienie względem płci:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[1] Płeć <> Miejsce zamieszkania

Dane	duże miasto	małe miasto	wieś	Razem
kobieta	51	180	67	298
mężczyzna	24	135	47	206
Razem	75	315	114	504
% wiersza	duże miasto	małe miasto	wieś	
kobieta	17,11%	60,4%	22,48%	
mężczyzna	11,65%	65,53%	22,82%	
% kolumny	duże miasto	małe miasto	wieś	
kobieta	68%	57,14%	58,77%	
mężczyzna	32%	42,86%	41,23%	
% sumy	duże miasto	małe miasto	wieś	
kobieta	10,12%	35,71%	13,29%	
mężczyzna	4,76%	26,79%	9,33%	
Oczekiw.:	duże miasto	małe miasto	wieś	
kobieta	44,35	186,25	67,4	
mężczyzna	30,65	128,75	46,6	

[2] Płeć <> Warunki socjalno-bytowe

Dane	bardzo dobre	bardzo złe	dobre	przeciętn	złe	Razem
kobieta	44	11	148	84	11	298
mężczyzna	33	2	106	59	6	206
Razem	77	13	254	143	17	504
% wiersza	bardzo dobre	bardzo złe	dobre	przeciętn	złe	
kobieta	14,77%	3,69%	49,66%	28,19%	3,69%	
mężczyzna	16,02%	0,97%	51,46%	28,64%	2,91%	
% kolumny	bardzo dobre	bardzo złe	dobre	przeciętn	złe	
kobieta	57,14%	84,62%	58,27%	58,74%	64,71%	
mężczyzna	42,86%	15,38%	41,73%	41,26%	35,29%	
% sumy	bardzo dobre	bardzo złe	dobre	przeciętn	złe	
kobieta	8,73%	2,18%	29,37%	16,67%	2,18%	
mężczyzna	6,55%	0,4%	21,03%	11,71%	1,19%	
Oczekiw.:	bardzo dobre	bardzo złe	dobre	przeciętn	złe	
kobieta	45,53	7,69	150,18	84,55	10,05	
mężczyzna	31,47	5,31	103,82	58,45	6,95	

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

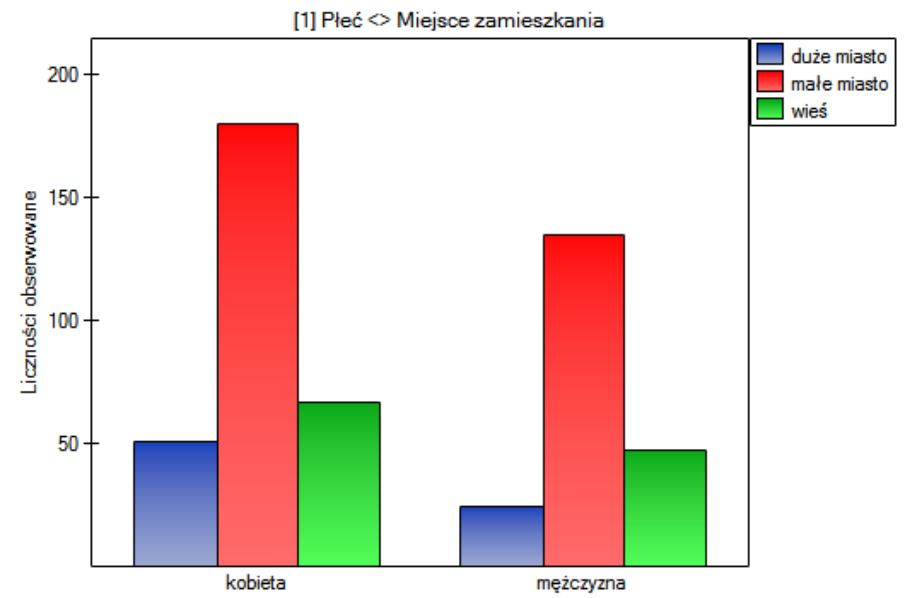
[3] Płeć <> Wykształcenie

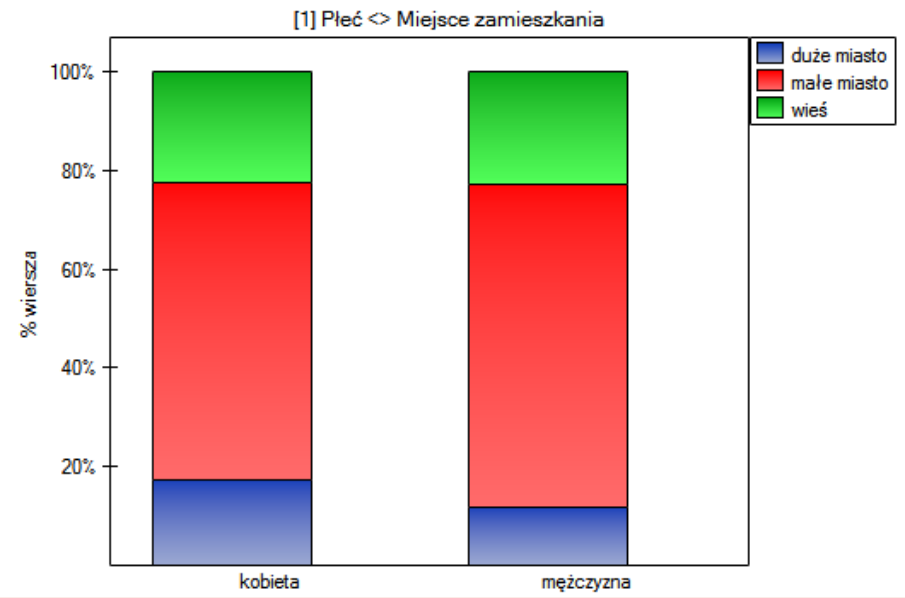
Dane	podstawowe	wyższe	zawodowe	średnie	Razem
kobieta	90	43	59	106	298
mężczyzna	33	39	76	58	206
Razem	123	82	135	164	504
% wiersza	podstawowe	wyższe	zawodowe	średnie	
kobieta	30,2%	14,43%	19,8%	35,57%	
mężczyzna	16,02%	18,93%	36,89%	28,16%	
% kolumny	podstawowe	wyższe	zawodowe	średnie	
kobieta	73,17%	52,44%	43,7%	64,63%	
mężczyzna	26,83%	47,56%	56,3%	35,37%	
% sumy	podstawowe	wyższe	zawodowe	średnie	
kobieta	17,86%	8,53%	11,71%	21,03%	
mężczyzna	6,55%	7,74%	15,08%	11,51%	
Oczekiw.:	podstawowe	wyższe	zawodowe	średnie	
kobieta	72,73	48,48	79,82	96,97	
mężczyzna	50,27	33,52	55,18	67,03	

[4] Płeć <> Stan cywilny

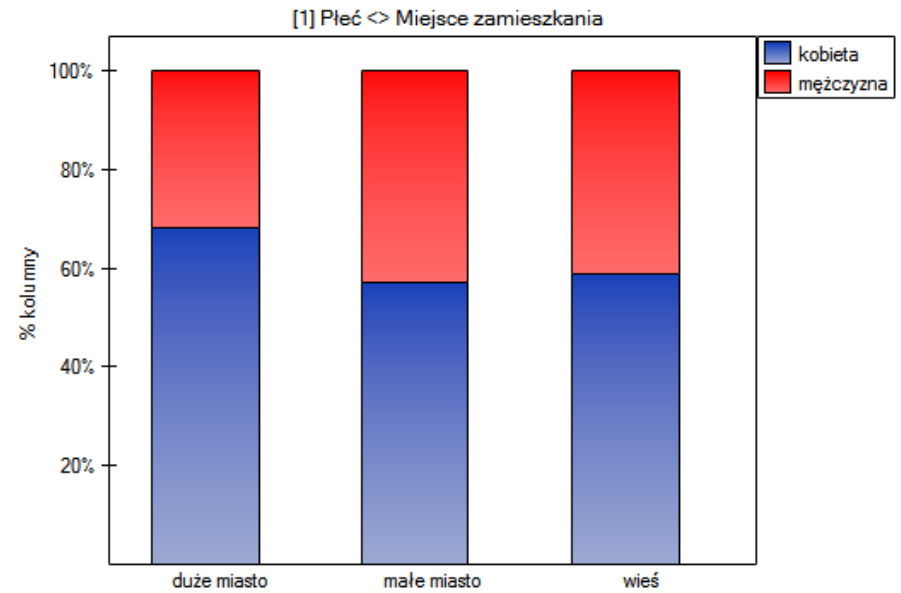
Dane	mężatka/żon	panna/kawal	rozwódka/ro:	wdowa/w	Razem
kobieta	112	24	16	146	298
mężczyzna	131	11	18	46	206
Razem	243	35	34	192	504
% wiersza	mężatka/żon	panna/kawal	rozwódka/ro:	wdowa/w	
kobieta	37,58%	8,05%	5,37%	48,99%	
mężczyzna	63,59%	5,34%	8,74%	22,33%	
% kolumny	mężatka/żon	panna/kawal	rozwódka/ro:	wdowa/w	
kobieta	46,09%	68,57%	47,06%	76,04%	
mężczyzna	53,91%	31,43%	52,94%	23,96%	
% sumy	mężatka/żon	panna/kawal	rozwódka/ro:	wdowa/w	
kobieta	22,22%	4,76%	3,17%	28,97%	
mężczyzna	25,99%	2,18%	3,57%	9,13%	
Oczekiw.:	mężatka/żon	panna/kawal	rozwódka/ro:	wdowa/w	
kobieta	143,68	20,69	20,1	113,52	
mężczyzna	99,32	14,31	13,9	78,48	

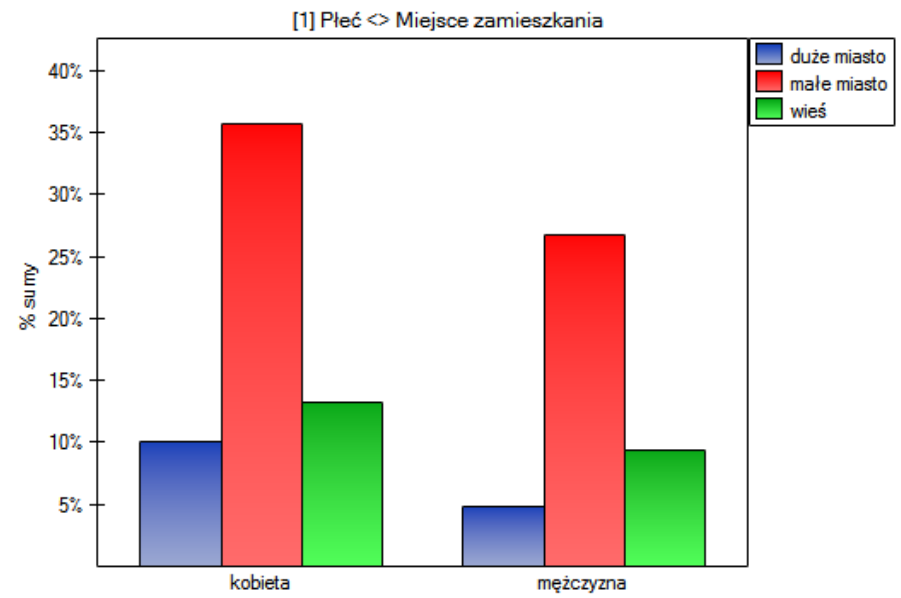
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone





Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





Dla rozkładu względem grup wiekowych utworzono najpierw kategorie wiekowe poprzez kody/etykiety/format.

Analizy dla tabel kontyngencji

KIEDY STOSOWAĆ:

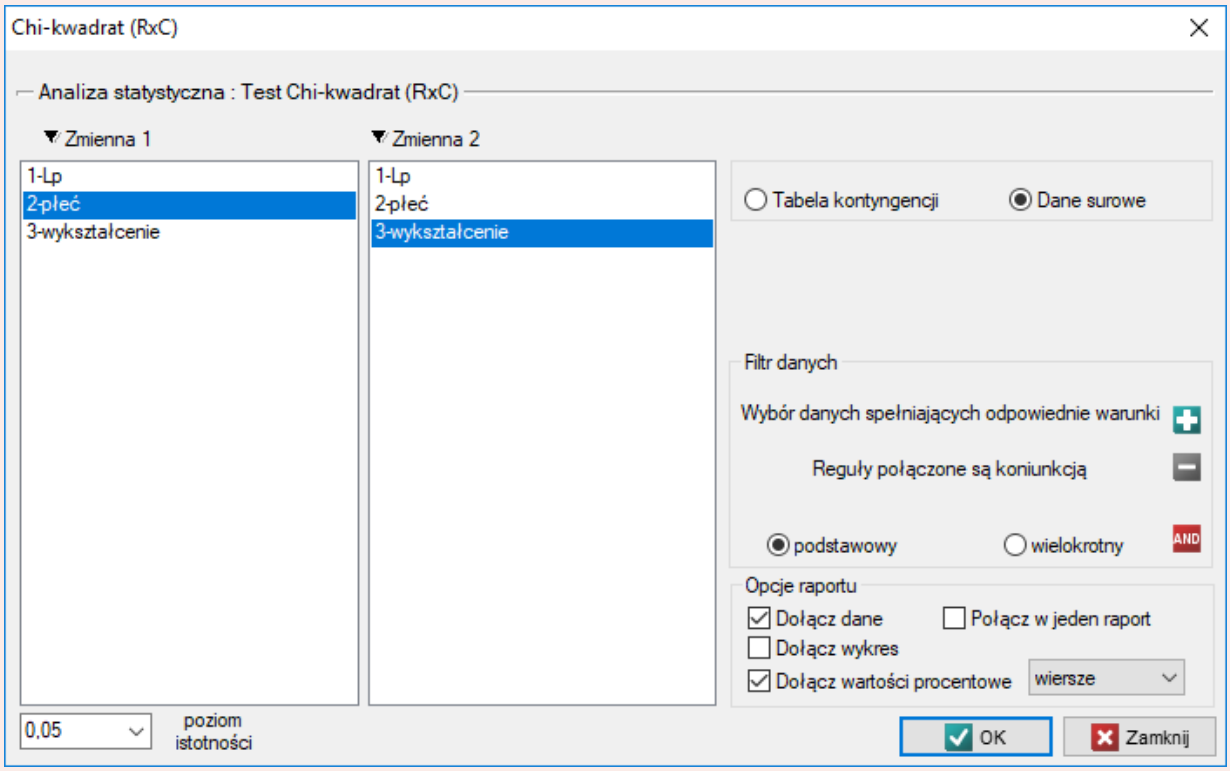
Analizy dla tabel kontyngencji stosuje się, gdy chcemy zbadać zależność między dwiema (lub więcej) cechami jakościowymi – np. czy rodzaj przyjmowanego leku wiąże się z wystąpieniem powikłań. Wyliczane są one na podstawie danych zebranych w tabelę kontyngencji lub bezpośrednio na podstawie

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

danych w postaci surowej. Istnieje przy tym możliwość transformacji danych z tabeli kontyngencji do postaci surowej lub odwrotnie.

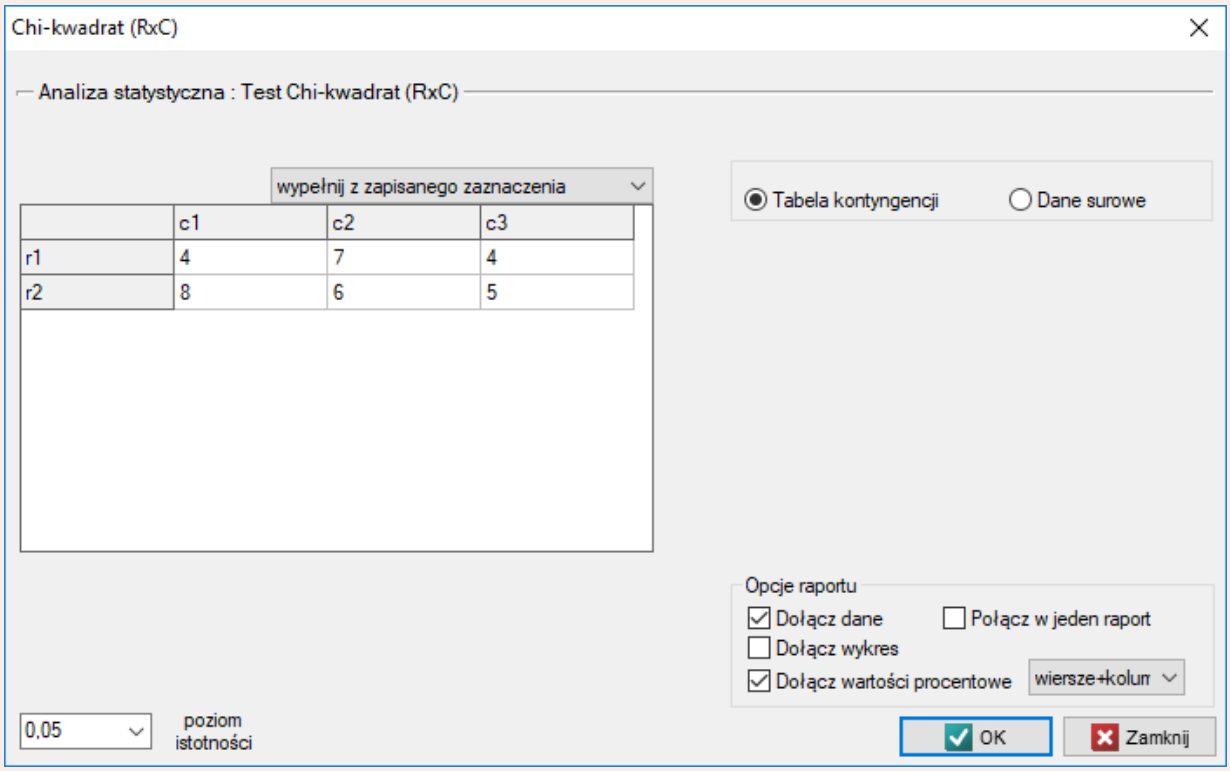
PRZYKŁAD 11.4. (plik płeć-wykształcenie.pqs)
Rozpatrzmy próbę składającą się z 34 osób ($n = 34$). Badamy 2 cechy tych osób (X=płeć, Y=wykształcenie). Płeć występuje w 2 kategoriach (X_1=kobieta, X_2=mężczyzna), wykształcenie w 3 kategoriach (Y_1= podstawowe + zawodowe, Y_2=średnie, Y_3=wyższe).

W przypadku danych surowych, po otwarciu okna opcji testu, np. testu χ^2 dla tabel $C \times R$, zaznaczona będzie automatycznie opcja dane surowe.



W przypadku danych zebranych w tabeli kontyngencji dobrze jest zaznaczyć te dane (wartości liczbowe bez nagłówków) przed uruchomieniem okna testu. Wówczas po otwarciu okna testu zaznaczona będzie automatycznie opcja tabela kontyngencji, a dane z zaznaczenia zostaną wyświetlone.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W oknie testu zawsze możemy zmienić automatycznie wykryte ustawienie dotyczące formy organizacji danych, jak też wpisywać z poziomu okna dane do tabeli kontyngencji.

Uwaga! Warunek Cochrana to podstawowy warunek stosowania wielu testów statystycznych opartych na tabelach kontyngencji, np. testu chi-kwadrat. Warunek ten zakłada duże liczności oczekiwane. Według interpretacji Cochrana 1952[41] żadna z liczności oczekiwanych nie może być < 1 oraz nie więcej niż 20% liczności oczekiwanych może być < 5. Informacja o spełnieniu (bądź niespełnieniu) tego warunku przez dane zebrane w tabeli może być zwrócona do raportu.

Podstawowe testy dla tabel kontyngencji:

- Test chi-kwadrat zgodności
- Test chi-kwadrat RxC (2x2) oraz jego poprawki
- Test chi-kwadrat dla trendu dla tabel Rx2
- Test McNemara, test wewnętrznej symetrii Bowkera
- Test chi-kwadrat dla wielowymiarowych tabel kontyngencji
- ANOVA Q-Cochrana
- Metoda Mantela-Haenszela dla wielu tabel 2x2

Współczynniki dla tabel kontyngencji:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Relatywne Ryzyko i Iloraz Szans
- Współczynniki kontyngencji: Q-Yulea, Phi, V -Cramera, C-Pearsona
- Współczynnik zgodności Kappa Cohena
- Współczynnik Kappa Fleissa
- Czułość i swoistość, PPV, NPV, LR(+), LR(-), wsp. chorobowości, dokładność

DEFINICJE MIAR

W raporcie wynikowym można również umieścić podstawowe podsumowanie tabel:

- **Tabela kontyngencji liczności obserwowanych** – czyli dane w postaci tabeli kontyngencji. Tabela taka przedstawia rozkład obserwacji dla kilku cech (kilku zmiennych). Tabelę dla 2 cech (X, Y), z których pierwsza ma możliwych r, a druga c kategorii, przedstawiono poniżej (tabela(11.1)).

  Tabela 11.1. Tabela kontyngencji $r \times c$ liczności obserwowanych

Liczności obserwowane O_{ij}		Cecha Y				
		Y_1	Y_2	...	Y_c	Suma
Cecha X	X_1	O_{11}	O_{12}	...	O_{1c}	$\sum_{j=1}^{c} O_{1j}$
	X_2	O_{21}	O_{22}	...	O_{2c}	$\sum_{j=1}^{c} O_{2j}$
	...	...	...	...	...	...
	X_r	O_{r1}	O_{r2}	...	O_{rc}	$\sum_{j=1}^{c} O_{rj}$
	Suma	$\sum_{i=1}^{r} O_{i1}$	$\sum_{i=1}^{r} O_{i2}$	...	$\sum_{i=1}^{r} O_{ic}$	$n = \sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij}$

  Liczności obserwowane O_{ij} $(i = 1, 2, \ldots, r; \quad j = 1, 2, \ldots, c)$ przedstawiają częstość występowania poszczególnych kategorii dla obu cech.
  Aby tabela taka była zwrócona przez program, należy w oknie testu wybrać opcję dołącz analizowane dane. Dla danych z przykładu (11.4) tabela kontyngencji liczności obserwowanych przedstawia się następująco:

Dane :			
	podstawov	wyższe	średnie
kobieta	4	7	4
mężczyzn	8	6	5

- **Tabela kontyngencji liczności oczekiwanych** – dla każdej tabeli kontyngencji liczności obserwowanych można utworzyć odpowiadającą jej tabelę **liczności oczekiwanych:** E_{ij} (tabela(11.2)) – czyli liczności, jakich spodziewalibyśmy się w poszczególnych komórkach tabeli, gdyby obie cechy były od siebie niezależne (wzór w *Dla zainteresowanych*).

  Tabela 11.2. Tabela kontyngencji $r \times c$ liczności oczekiwanych

Liczności oczekiwane E_{ij}		Cecha Y			
		Y_1	Y_2	...	Y_c
Cecha X	X_1	E_{11}	E_{12}	...	E_{1c}
	X_2	E_{21}	E_{22}	...	E_{2c}
	...	...	...	...	...
	X_r	E_{r1}	E_{r2}	...	E_{rc}

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Dla danych z przykładu (11.4) tabela kontyngencji liczności oczekiwanych przedstawia się następująco:

Oczekiw.:			
	podstawov	wyższe	średnie
kobieta	5.29	5.74	3.97
mężczyzn	6.71	7.26	5.03

- **Tabela kontyngencji wartości procentowych wyliczanych z sumy kolumn**. Dla danych z przykładu (11.4) tabela ta przedstawia się następująco:

kolumny:			
	podstawov	wyższe	średnie
kobieta	33.33%	53.85%	44.44%
mężczyzn	66.67%	46.15%	55.56%

- **Tabela kontyngencji wartości procentowych wyliczanych z sumy wierszy**. Dla danych z przykładu (11.4) tabela ta przedstawia się następująco:

wiersze:			
	podstawov	wyższe	średnie
kobieta	26.67%	46.67%	26.67%
mężczyzn	42.11%	31.58%	26.32%

- **Tabela kontyngencji wartości procentowych wyliczanych z sumy całkowitej wierszy i kolumn**. Dla danych z przykładu (11.4) tabela ta przedstawia się następująco:

sumy:			
	podstawov	wyższe	średnie
kobieta	11,76%	20,59%	11,76%
mężczyzn	23,53%	17,65%	14,71%

11.3 Statystyki opisowe

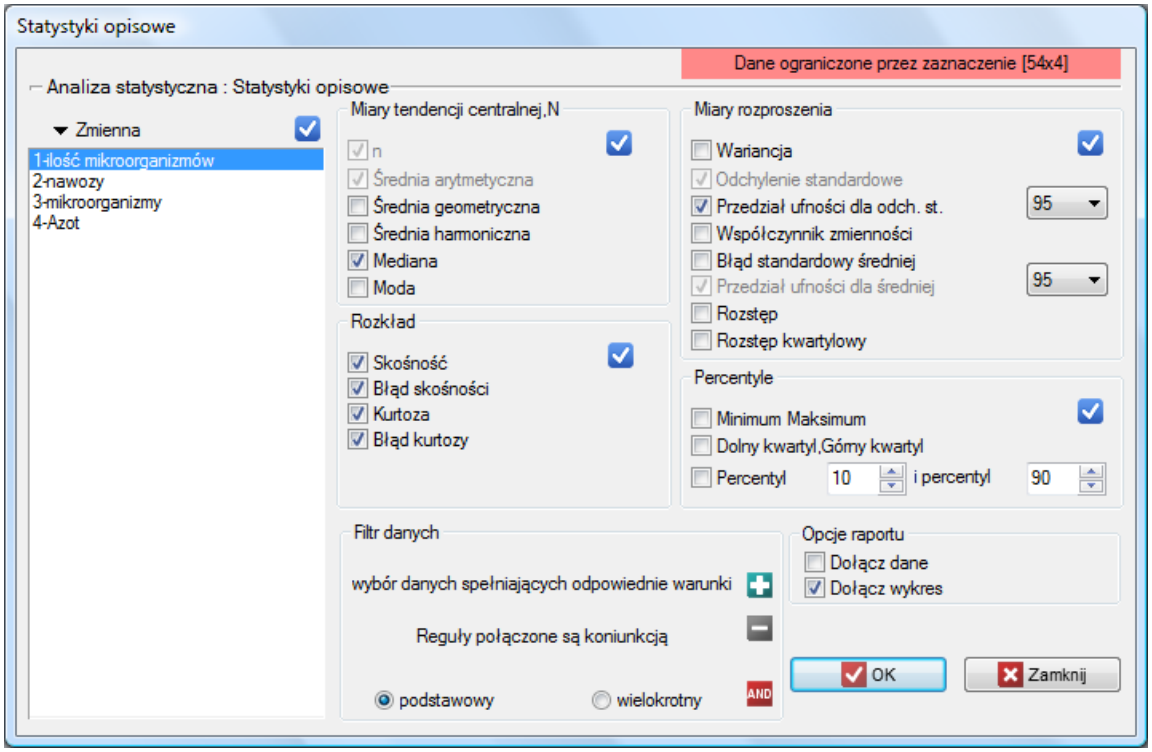
KIEDY STOSOWAĆ:

Celem stosowania metod statystyki opisowej jest podsumowanie zbioru danych za pomocą kilku charakterystyk liczbowych, np. wartości średniej, mediany czy odchylenia standardowego, oraz wyciągnięcie na tej podstawie podstawowych wniosków i uogólnień na temat zbioru – bez tego kroku trudno w ogóle zacząć rozumieć, z jakimi danymi mamy do czynienia.

JAK USTAWIĆ ANALIZĘ

Statystyka → Analizy opisowe → Statystyki opisowe.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



W oknie tym wybieramy zmienną do analizy oraz opcje analizy i zaznaczamy interesujące nas miary statystyk opisowych. Zaznaczać można pojedyncze statystyki lub całe ich grupy, wybierając przycisk ☑. Dokonany wybór potwierdzamy przyciskiem OK. Wynik analizy znajdzie się w raporcie dołączonym do arkusza danych, dla których została wykonana.

Jeśli dodatkowo chcemy, by dane zostały zobrazowane wykresem ramka-wąsy, zaznaczamy w oknie Statystyk opisowych opcję Dołącz wykres.

Miary tendencji centralnej

Miary tendencji centralnej to tzw. miary przeciętne, charakteryzujące średni lub typowy poziom wartości cechy.

DEFINICJE MIAR

Średnia arytmetyczna (*ang. arithmetic mean*) to najbardziej znana z miar przeciętnych – suma wszystkich wartości podzielona przez ich liczbę (wzór w *Dla zainteresowanych*). Jest stosowana dla skali interwałowej. Dla próby przyjmuje się oznaczać ją przez $\overline{x}$, a dla populacji przez μ.

Średnia przycięta (*Trimmed mean*) – wyznaczana jest jako średnia arytmetyczna obliczona po usunięciu z próbki zadanego procentu najmniejszych i największych pomiarów, np. gdy obcinamy 5% pomiarów, to oznacza, że obcinamy 2.5% największych i 2.5% najmniejszych wartości. Gdy uzyskana na podstawie przeliczeń liczba pomiarów przeznaczonych do usunięcia nie jest liczbą całkowitą, jest ona zaokrąglana w dół do najbliższej całkowitej.

Średnia Winsora (*Winsor mean*) – wyznaczana jest jako średnia arytmetyczna obliczona po zastąpieniu odpowiedniego odsetka skrajnych pomiarów wartością najmniejszą i największą, jaka pozostała w zmniejszonym zbiorze wartości. Jeśli zdecydujemy się na obliczanie średniej Winsora, przycinając np.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

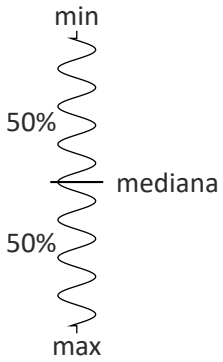
5% pomiarów, to te odrzucone 5% zostanie zastąpione wartością najmniejszą i największą wyznaczoną z pozostałych 95% pomiarów. Podobnie jak dla średniej przycinanej, gdy na podstawie przeliczenia procentu wartości przeznaczonych do zamiany nie uzyskamy liczby całkowitej, zaokrąglamy w dół do najbliższej całkowitej.

Średnia geometryczna (*ang. geometric mean*) to pierwiastek stopnia n z iloczynu wszystkich wartości (wzór w *Dla zainteresowanych*). Jest stosowana dla skali interwałowej, gdy zmienna ma rozkład logarytmiczno-normalny (logarytm zmiennej ma rozkład normalny).

Średnia harmoniczna (*ang. harmonic mean*) to liczba obserwacji podzielona przez sumę odwrotności poszczególnych wartości (wzór w *Dla zainteresowanych*). Jest stosowana dla skali interwałowej.

Mediana (*ang. median*)

W uporządkowanym zbiorze danych mediana jest wartością dzielącą ten zbiór na dwie równe części. Połowa wszystkich obserwacji znajduje się poniżej, a połowa powyżej mediany.



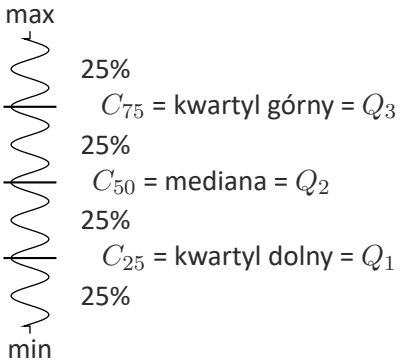
Mediana może być stosowana w skali interwałowej oraz porządkowej.

Moda (*ang. mode*)

Moda — to wartość, która występuje najczęściej wśród uzyskanych pomiarów. Moda może być stosowana w każdej skali pomiarowej.

Inne miary położenia

Kwartyle *(ang. quartiles)*, **decyle** *(ang. deciles)*, **centyle** *(ang. centiles)* to wartości dzielące uporządkowany zbiór danych na równe części – podobnie jak mediana, tylko na więcej niż dwie części.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Kwartyle (Q_1, Q_2, Q_3) dzielą uporządkowany szereg na 4 równe części, decyle ($D_i, i = 1, 2, ..., 9$) na 10 równych części, a centyle (percentyle: $C_i, i = 1, 2, ..., 99$) na 100 równych części. Drugi kwartyl, piąty decyl i pięćdziesiąty centyl są równe medianie. Miary te mogą być stosowane w skali interwałowej oraz porządkowej.

Miary rozproszenia

Znajomość miar tendencji centralnej nie wystarcza do scharakteryzowania struktury zbiorowości statystycznej – badana grupa może charakteryzować się różnym stopniem zmienności w zakresie badanej cechy. Potrzebne są zatem miary, które pozwalają określić rozrzut danych.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej – miary rozproszenia bazują na odległościach między punktami, więc nie liczy się ich dla skali porządkowej ani nominalnej.

DEFINICJE MIAR

Rozstęp *(ang. range)* – różnica między największą a najmniejszą wartością w próbie (wzór w *Dla zainteresowanych*). **Rozstęp kwartylowy (IQR)** to analogiczna różnica pomiędzy górnym a dolnym kwartylem.

Rozstępy dla skali percentylowej (decylowej, centylowej). Rozstępy między percentylami to jedna z miar rozproszenia – określają procent wszystkich obserwacji, których wartość znajduje się pomiędzy wybranymi percentylami.

Wariancja *(ang. variance)* – mierzy stopień rozproszenia pomiarów wokół średniej arytmetycznej. Liczy się ją nieco inaczej dla próby i dla całej populacji (wzory w *Dla zainteresowanych*). Wariancja jest zawsze dodatnia, ale nie jest wyrażona w tych samych jednostkach co wyniki pomiarów.

Odchylenie standardowe *(ang. standard deviation)* – podobnie jak wariancja, mierzy stopień rozproszenia pomiarów wokół średniej arytmetycznej, ale – w odróżnieniu od wariancji – jest wyrażone w tych samych jednostkach co dane (jest pierwiastkiem z wariancji). Im wyższa wartość odchylenia standardowego (lub wariancji), tym bardziej zróżnicowana grupa pod względem badanej cechy.

Uwaga! Odchylenie standardowe z próby jest pewnym przybliżeniem (estymatorem) odchylenia standardowego z populacji. Populacyjna wartość odchylenia standardowego mieści się w pewnym przedziale zawierającym odchylenie standardowe z próby. Przedział ten nazywany jest **przedziałem ufności** *(confidence interval)* dla odchylenia standardowego. **Współczynnik zmienności** *(ang. coefficient of variation)*

Współczynnik zmienności, podobnie jak odchylenie standardowe, pozwala ocenić stopień jednorodności badanej zbiorowości (wzór w *Dla zainteresowanych*). Jest to wielkość niemianowana, co pozwala na porównanie zróżnicowania kilku zbiorowości pod względem tej samej cechy oraz tej samej zbiorowości pod względem kilku różnych cech (wyrażonych w różnych jednostkach). Przyjmuje się, że jeżeli współczynnik V nie przekracza 10%, to cechy wykazują zróżnicowanie statystycznie nieistotne.

Błędy standardowe *(ang. standard errors)* – nie są miarami rozproszenia wyników pomiarowych, lecz

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

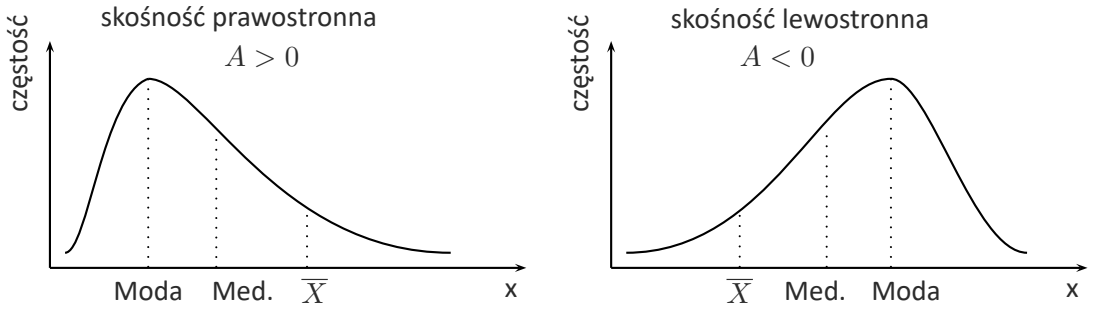
określają stopień dokładności, z jaką możemy określić wartości parametrów populacji na podstawie ich estymatorów wyliczonych z próby. Najczęściej wykorzystywanym błędem standardowym jest **błąd standardowy średniej arytmetycznej** *(ang. standard error of the mean, SEM)* – wzór w *Dla zainteresowanych*.

Uwaga! Na podstawie błędu standardowego estymatora z próby można określić przedział ufności dla parametru populacji.

Inne atrybuty rozkładu

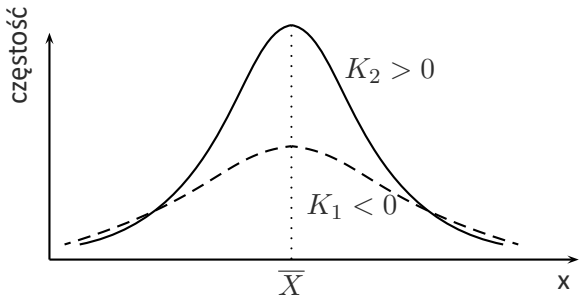
Skośność, inaczej współczynnik asymetrii *(ang. skewness)*

Jest to miara, która mówi o tym, jak bardzo rozkład danych różni się od rozkładu symetrycznego. Im wartość współczynnika asymetrii jest bliższa zeru, tym bardziej symetrycznie wokół średniej rozkładają się dane. Zwykle wartość tego współczynnika zawiera się w przedziale [-1, 1], choć przy szczególnie dużej asymetrii może się poza nim znaleźć. Wartości dodatnie świadczą o skośności prawostronnej (o dłuższym prawym „ogonie"), wartości ujemne zaś o skośności lewostronnej (o dłuższym lewym „ogonie"). Wzór na skośność znajduje się w *Dla zainteresowanych*.



Kurtoza, inaczej współczynnik koncentracji *(ang. kurtosis)*

Jest to miara, która mówi o tym, jak bardzo rozrzut danych wokół średniej jest zbliżony do rozrzutu tych danych w rozkładzie normalnym. Im wartość kurtozy jest większa od zera, tym badany rozkład jest bardziej smukły niż rozkład normalny, a im mniejsza od zera, tym rozkład jest bardziej spłaszczony niż rozkład normalny. Wzór na kurtozę znajduje się w *Dla zainteresowanych*.



WYKONANIE I INTERPRETACJA

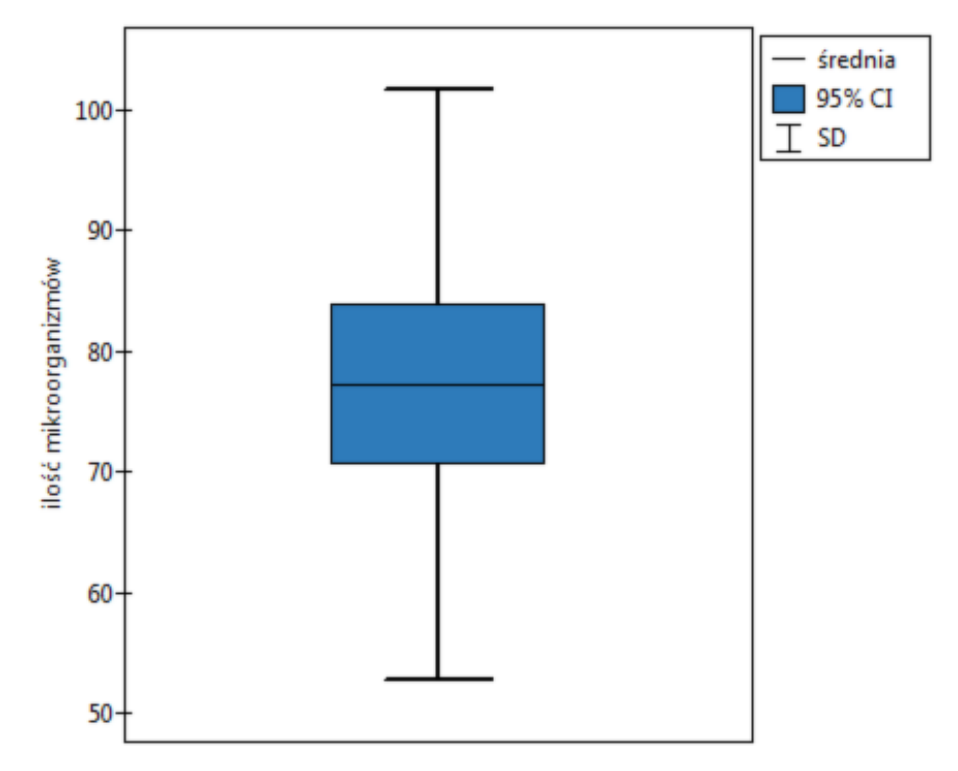
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

PRZYKŁAD 11.5. (plik nawozy.pqs)
W doświadczeniu dotyczącym nawożenia gleby różnymi rodzajami preparatów mikrobiologicznych i nawozów wyliczono liczbę mikroorganizmów występujących w 1 gramie suchej masy gleby. Chcemy wyznaczyć statystyki opisowe liczby promieniowców dla próbki nawożonej azotem i zobrazować uzyskane wyniki za pomocą wykresu ramka-wąsy. Zaznaczamy w arkuszu danych tylko 54 pierwsze wiersze, które odpowiadają założeniom analizy (są to promieniowce nawożone azotem), i uruchamiamy okno Statystyki opisowe.

W oknie opcji testu statystyk opisowych wybieramy zmienną do analizy: Liczby mikroorganizmów, a następnie procedury, jakie chcemy wykonać (np. średnią arytmetyczną wraz z przedziałem ufności, medianę, odchylenie standardowe wraz z przedziałem ufności oraz informacje o skośności i kurtozie rozkładu wraz z błędami). Na górze okna powinien być widoczny komunikat: Dane ograniczone przez zaznaczenie. By w raporcie znalazł się również wykres, zaznaczamy opcję Dołącz wykres i wybieramy interesujący nas rodzaj wykresu ramka-wąsy. Potwierdzamy wybór przyciskiem OK i uzyskujemy wynik w postaci raportu:

Statystyki opisowe	
Czas analizy	0.13 sek.
Analizowane zmienne	ilość mikroorganizmów
Poziom istotności	0.05
Liczność - grupy	54
Średnia arytmetyczna	77.240741
Mediana	78.5
Odchylenie standardowe	24.425424
95% Przedział ufności dla odch. st.	20.532603
	30.153531
95% Przedział ufności dla średniej	70.573883
	83.907598
Skośność	-0.226875
Błąd skośności	0.324556
Kurtoza	0.343163
Błąd kurtozy	0.638893

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



11.4 Zestawienia opisowe

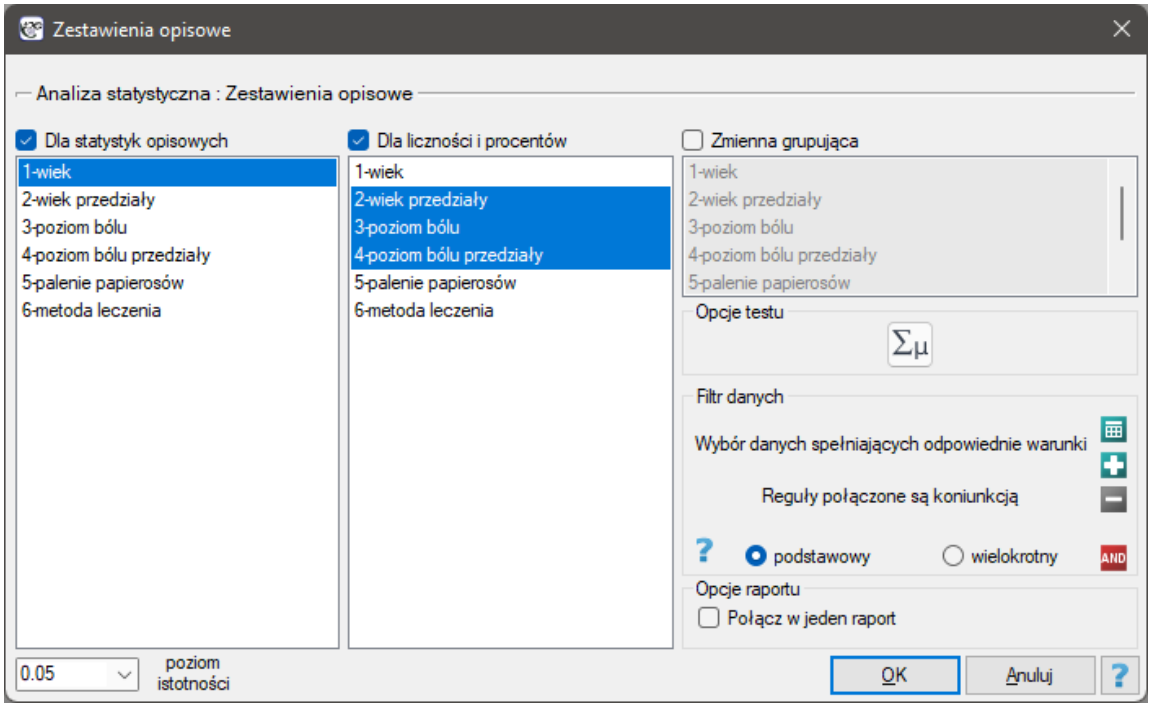
KIEDY STOSOWAĆ:

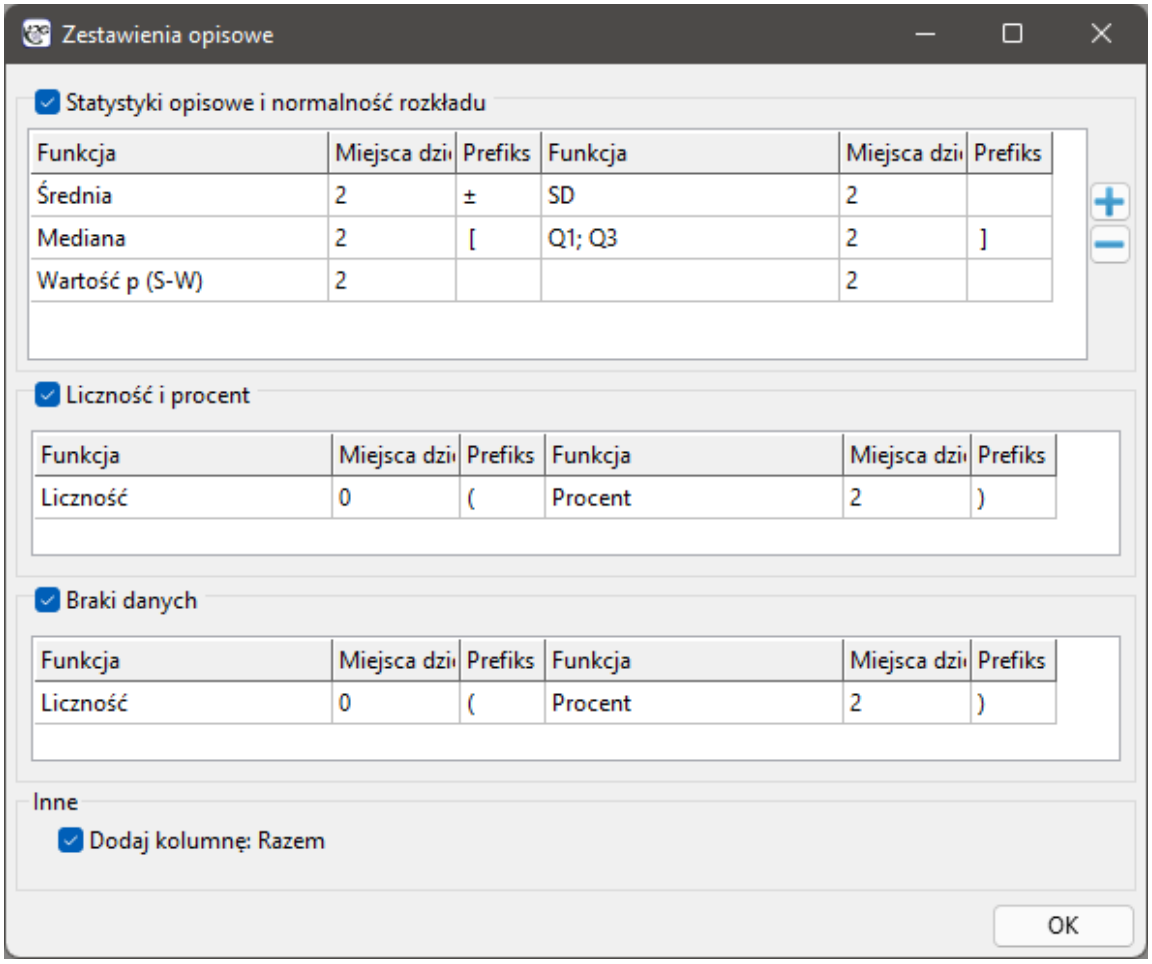
Zestawienia opisowe to szybkie przygotowanie raportu przedstawiającego opis danych, gotowego do umieszczenia bezpośrednio w pracy naukowej. Jest to narzędzie ułatwiające zebranie wszystkich form podstawowego opisu danych w jednym miejscu.

JAK USTAWIĆ ANALIZĘ

Statystyka → Analizy opisowe → Zestawienia opisowe.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





W zależności od sposobu dokonywania pomiarów do opisu zmiennej używamy najczęściej jednej z trzech typów miar:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- średnia $\pm$ odchylenie standardowe, w skrócie $\overline{x} \pm sd$
- mediana [kwartyl dolny; kwartyl górny], w skrócie $Me[Q_1; Q_3]$, lub mediana [min; max], w skrócie $Me[min; max]$
- liczność (procent) poszczególnych kategorii, w skrócie $n(\%)$

W zależności od potrzeb i spełnienia dodatkowych założeń:

- w skali interwałowej dane mogą zostać opisane przy użyciu dowolnej miary – dodatkowo często badaczowi zależy na sprawdzeniu normalności rozkładu takich danych,
- w skali porządkowej do dyspozycji mamy mediany (z kwartylami lub wartością najmniejszą i największą) lub liczności i procenty poszczególnych kategorii,
- w skali nominalnej jedynie liczności i procenty poszczególnych kategorii.

WYKONANIE I INTERPRETACJA

Przykład 11.6. (plik Zestwienia opisowe.pqs)
Przykładowy sposób opisu danych przedstawia uzyskany raport podsumowujący wiek, poziom odczuwanego bólu i status palenia papierosów. Dane te zostały podsumowane w tabeli pierwszej, opisującej wszystkie 100 osób, oraz w tabeli drugiej z podziałem na metodę leczenia.

Zestawienia opisowe		
		(N=100)
wiek		
	Średnia ± SD	66.38±16.31
	Mediana [Q1; Q3]	64.5 [55.75; 76]
wiek przedziały		
	>=70	40(40%)
	60-69	22(22%)
	50-59	23(23%)
	40-49	12(12%)
	<=39	3(3%)
poziom bólu przedziały		
	7-10	49(49%)
	4-6	31(31%)
	0-3	20(20%)
palenie papierosów		
	tak	27(27%)
	nie	73(73%)

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Zestawienia opisowe			
		metoda leczenia A(N=44)	B(N=56)
wiek			
	Średnia ± SD	66.27±17.7	66.46±15.29
	Mediana [Q1; Q3]	64 [56; 74.5]	65.5 [54.75; 76.25]
wiek przedziały			
	>=70	17(38.64%)	23(41.07%)
	60-69	11(25%)	11(19.64%)
	50-59	9(20.45%)	14(25%)
	40-49	5(11.36%)	7(12.5%)
	<=39	2(4.55%)	1(1.79%)
poziom bólu przedziały			
	7-10	16(36.36%)	33(58.93%)
	4-6	15(34.09%)	16(28.57%)
	0-3	13(29.55%)	7(12.5%)
palenie papierosów			
	tak	11(25%)	16(28.57%)
	nie	33(75%)	40(71.43%)

DLA ZAINTERESOWANYCH

Wzory

Tabela kontyngencji liczności oczekiwanych

$$E_{11} = \frac{\sum_{i=1}^{r} O_{i1} \times \sum_{j=1}^{c} O_{1j}}{n}, \quad E_{12} = \frac{\sum_{i=1}^{r} O_{i2} \times \sum_{j=1}^{c} O_{1j}}{n}, \quad E_{1c} = \frac{\sum_{i=1}^{r} O_{ic} \times \sum_{j=1}^{c} O_{1j}}{n}$$

$$E_{21} = \frac{\sum_{i=1}^{r} O_{i1} \times \sum_{j=1}^{c} O_{2j}}{n}, \quad E_{22} = \frac{\sum_{i=1}^{r} O_{i2} \times \sum_{j=1}^{c} O_{2j}}{n}, \quad E_{2c} = \frac{\sum_{i=1}^{r} O_{ic} \times \sum_{j=1}^{c} O_{2j}}{n}$$

$$E_{r1} = \frac{\sum_{i=1}^{r} O_{i1} \times \sum_{j=1}^{c} O_{rj}}{n}, \quad E_{r2} = \frac{\sum_{i=1}^{r} O_{i2} \times \sum_{j=1}^{c} O_{rj}}{n}, \quad E_{rc} = \frac{\sum_{i=1}^{r} O_{ic} \times \sum_{j=1}^{c} O_{rj}}{n}.$$

Średnia arytmetyczna

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n},$$

gdzie x_i to kolejne wartości zmiennej, a n – liczność próby.

Średnia geometryczna

$$\overline{x}_G = \sqrt[n]{x_1 x_2 \ldots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}.$$

Średnia harmoniczna

$$\overline{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}.$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Rozstęp

$$I = \max x_i - \min x_i,$$

gdzie x_i to wartości badanej zmiennej.

$$IQR = \text{rozstęp kwartylowy} = Q_3 - Q_1,$$

gdzie Q_1, Q_3 to dolny i górny kwartyl.

Wariancja *(ang. variance)*

wariancja z próby:

$$sd^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1},$$

gdzie x_i to kolejne wartości zmiennej, $\overline{x}$ to średnia arytmetyczna tych wartości, n – liczność próby.

wariancja z populacji:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N},$$

gdzie x_i to kolejne wartości zmiennej, μ to średnia arytmetyczna tych wartości, N – liczność populacji.

Odchylenie standardowe *(ang. standard deviation)*

odchylenie standardowe z próby:

$$sd = \sqrt{sd^2},$$

odchylenie standardowe z populacji:

$$\sigma = \sqrt{\sigma^2}.$$

Współczynnik zmienności

$$V = \frac{sd}{\overline{x}}100\%,$$

gdzie sd to odchylenie standardowe, $\overline{x}$ to średnia arytmetyczna.

Błąd standardowy średniej *(ang. standard error of the mean)*

$$SEM = \text{błąd standardowy średniej} = \frac{sd}{\sqrt{n}}.$$

Skośność

$$A = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{sd}\right)^3,$$

gdzie:
x_i – kolejne wartości zmiennej,
$\overline{x}$, sd – odpowiednio średnia arytmetyczna i odchylenie standardowe x_i,
n – liczność próby.

Kurtoza

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{sd}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)},$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

gdzie:
$x_i -$ kolejne wartości zmiennej,
$\overline{x}$, $sd -$ odpowiednio średnia arytmetyczna i odchylenie standardowe x_i,
$n -$ liczność próby.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

12 ROZKŁADY PRAWDOPODOBIEŃSTWA

CZYM SĄ ROZKŁADY PRAWDOPODOBIEŃSTWA

Rzeczywisty rozkład danych z próby – **rozkład empiryczny danych** – może być przedstawiony za pomocą Tabel liczności (poprzez wybranie menu Statystyka→Analizy opisowe→Tabele liczności). Na przykład rozkład liczby wykorzystanych darmowych minut przez abonentów pewnego operatora telefonii komórkowej (*przykład (11.1), plik: rozkład.pqs*) przedstawia następująca tabela:

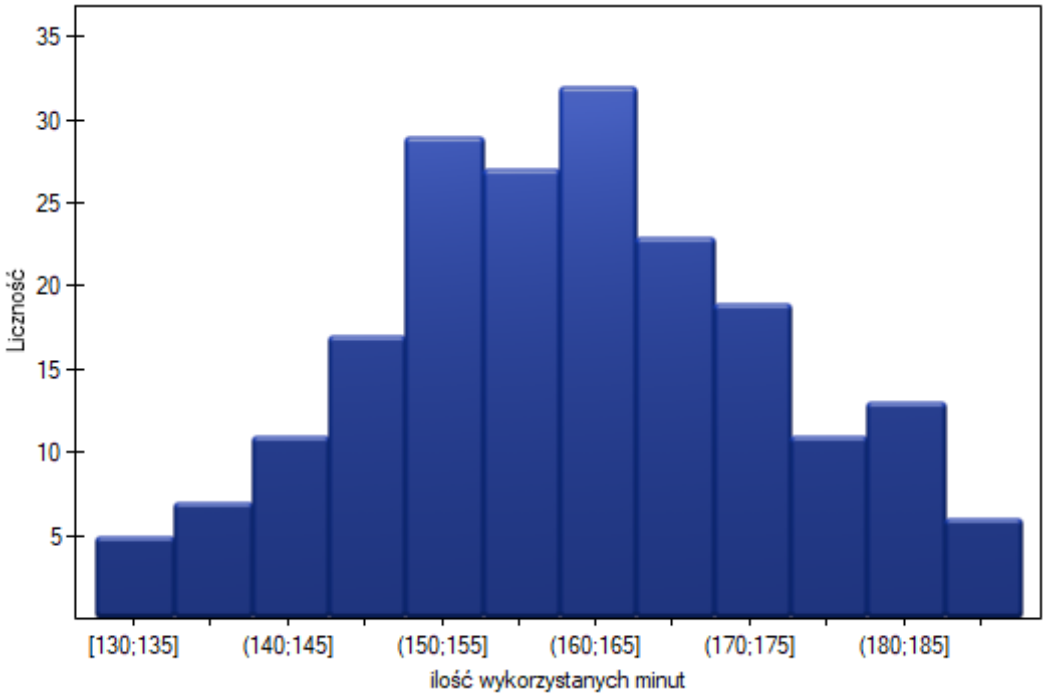
Tabele liczności

Analizowane zmienne	ilość wykorzyst:
Liczność / braki danych	0
Liczność / przypadki dezaktywowane	0

	Liczność	Procent
[130;135]	5	2.5%
(135;140]	7	3.5%
(140;145]	11	5.5%
(145;150]	17	8.5%
(150;155]	29	14.5%
(155;160]	27	13.5%
(160;165]	32	16%
(165;170]	23	11.5%
(170;175]	19	9.5%
(175;180]	11	5.5%
(180;185]	13	6.5%
(185;190]	6	3%
Razem	200	100%

Graficzna prezentacja wyników uzyskanych w raporcie tabeli zwykle dokonywana jest przy pomocy histogramu czy też wykresu kolumnowego.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Wykres taki możemy uzyskać zaznaczając w oknie Tabel liczności opcję Dołącz wykres.

Rozkład teoretyczny danych, zwany również **rozkładem prawdopodobieństwa**, przedstawiany jest zwykle graficznie przy pomocy wykresu liniowego. Taka linia opisana jest funkcją (modelem matematycznym) zwaną **funkcją gęstości rozkładu**. Odpowiednim rozkładem teoretycznym można zastąpić rozkład empiryczny – dzięki temu można opisać dane kilkoma parametrami zamiast wypisywać wszystkie zaobserwowane wartości.

Uwaga! Do zastąpienia rozkładu empirycznego rozkładem teoretycznym nie wystarczy intuicyjne stwierdzenie podobieństwa ich przebiegu. Służą do tego specjalnie skonstruowane testy zgodności.

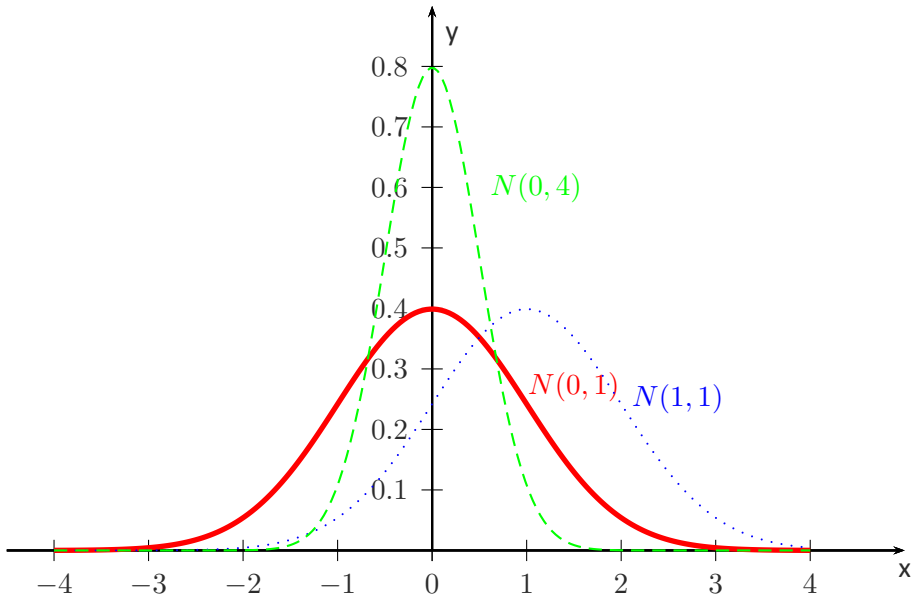
Najczęściej używanym rozkładem prawdopodobieństwa jest rozkład normalny (rozkład Gaussa) – i taki właśnie rozkład, o średniej 161.15 i odchyleniu standardowym równym 13.03, mają dane na temat liczby wykorzystanych darmowych minut (*przykład (11.1), plik: rozkład.pqs*).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

12.1 CIĄGŁE ROZKŁADY PRAWDOPODOBIEŃSTWA

ROZKŁAD NORMALNY (ROZKŁAD GAUSSA)

(ang. normal distribution), nazywany również krzywą dzwonową, jest jednym z najważniejszych rozkładów w statystyce – ma interesujące własności matematyczne i bardzo często opisuje zjawiska występujące w naturze. Oznacza się go zwykle symbolem $N(\mu, \sigma)$, gdzie μ to wartość oczekiwana (średnia w populacji), a σ to odchylenie standardowe – to właśnie te dwie liczby w pełni określają kształt i położenie krzywej.



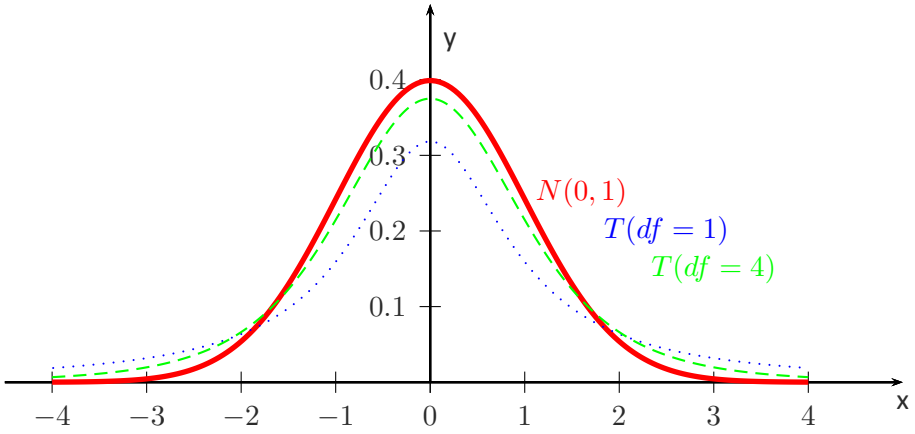
Rozkład normalny jest rozkładem symetrycznym względem prostej prostopadłej do osi odciętych, przechodzącej przez punkt wyznaczający jednocześnie średnią, modę i medianę.

Rozkład normalny o średniej $\mu = 0$ i odchyleniu standardowym $\sigma = 1$ ($N(0, 1)$) to tzw. **rozkład normalny standaryzowany**.

ROZKŁAD T-STUDENTA

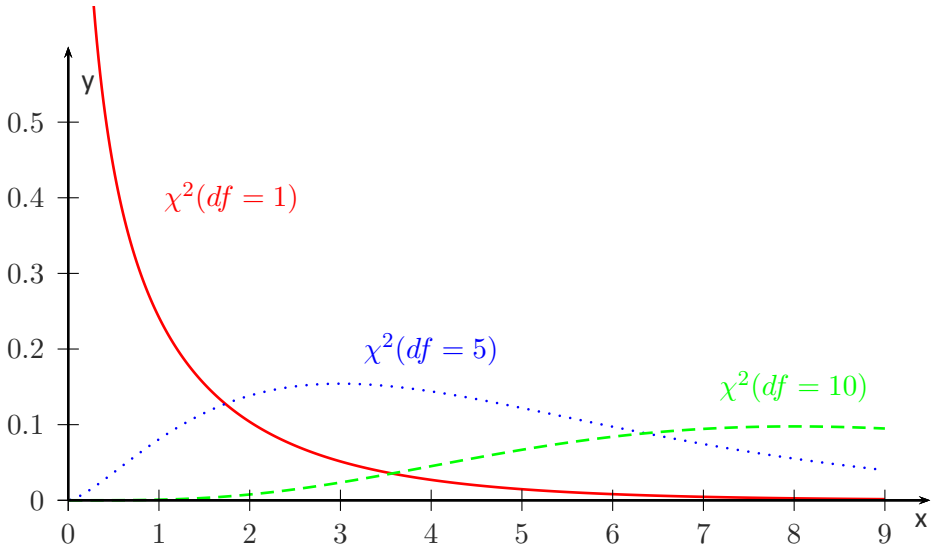
(ang. t-Student distribution) ma kształt podobny do standaryzowanego rozkładu normalnego, ale ma dłuższe "ogony" – czyli wartości skrajne pojawiają się w nim nieco częściej. Jego kształt zależy od liczby stopni swobody df *(ang. degree of freedom)*: im więcej stopni swobody (czyli zwykle im większa próba), tym bardziej rozkład t-Studenta przypomina rozkład normalny.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



ROZKŁAD CHI-KWADRAT (χ^2)

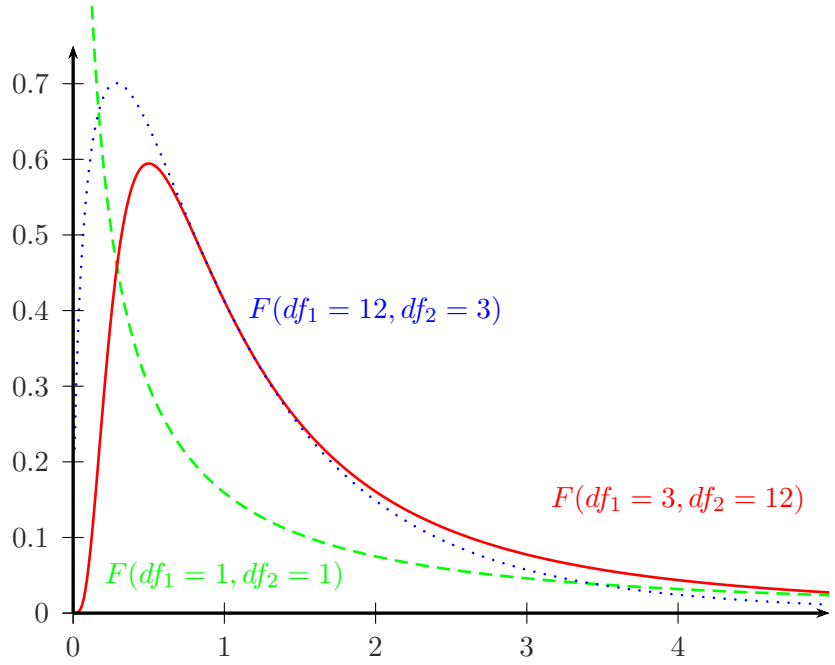
(ang. Chi-square distribution) jest rozkładem prawoskośnym (z dłuższym "ogonem" po prawej stronie), którego kształt zależy od liczby stopni swobody df. Podobnie jak w rozkładzie t-Studenta, wraz ze wzrostem liczby stopni swobody kształt rozkładu χ^2 upodabnia się do kształtu rozkładu normalnego.



ROZKŁAD F SNEDECORA

(ang. Fisher-Snedecor distribution) ma dłuższy prawy "ogon", a jego kształt zależy od dwóch liczb stopni swobody: df_1 i df_2. Rozkład ten powstaje jako iloraz dwóch niezależnych zmiennych o rozkładzie χ^2 (każda podzielona przez swoją liczbę stopni swobody) i pojawia się w testach porównujących wariancje lub w analizie wariancji (ANOVA).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

12.2 KALKULATOR FUNKCJI DYSTRYBUCJI

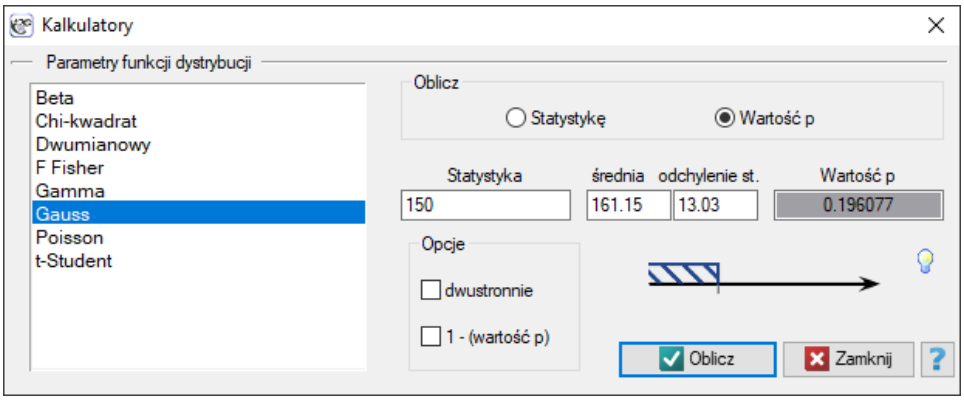
DO CZEGO SŁUŻY



Pole pod krzywą (funkcją gęstości rozkładu) to **prawdopodobieństwo** p wystąpienia wszystkich możliwych wartości badanej zmiennej losowej. Całe pole pod krzywą wynosi $p = 1$. Gdy chcemy zbadać wielkość tylko części tego pola, musimy podać wartość graniczną zwaną **wartością krytyczną** lub Statystyką. Korzystamy w tym celu z okna Kalkulatora funkcji dystrybucji. W oknie tym można wyliczyć wartość pola pod krzywą (Wartość p) zadanego rozkładu na podstawie Statystyki, jak też wyznaczyć wartość Statystyki na podstawie Wartości p.

JAK OTWORZYĆ KALKULATOR

Okno Kalkulatora funkcji dystrybucji uruchamiamy poprzez wybranie menu Statystyka→Kalkulatory→Kalkulator funkcji dystrybucji.



PRZYKŁAD

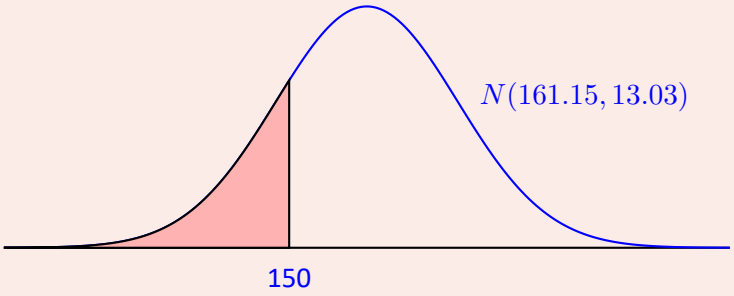
PRZYKŁAD 12.1. Kalkulator funkcji dystrybucji
Pewien operator telefonii komórkowej przeprowadza szereg badań dotyczących wykorzystania przez klientów liczby przyznanych w abonamencie ”darmowych minut”. Na podstawie 200-osobowej próby swoich klientów (w której rozkład wykorzystanych ”darmowych minut” przyjmuje kształt rozkładu normalnego) wyznaczył wartość średnią $\overline{x} = 161.15min.$ i odchylenie standardowe $sd = 13.03min.$ Chcemy wyliczyć prawdopodobieństwo, że wylosowany przez nas klient wykorzystał:

1. 150 minut lub mniej,
2. więcej niż 150 minut,
3. ilość minut z przedziału $[\overline{x} - sd, \overline{x} + sd] = [148.12min., 174.18min.]$,
4. ilość minut spoza przedziału $\overline{x} \pm sd$.

Uruchamiamy okno Kalkulatora funkcji dystrybucji, wybieramy rozkład Gaussa i wpisujemy średnią $\overline{x} = 161.15min.$ i odchylenie st. $sd = 13.03min.$ oraz zaznaczamy, że będziemy wyliczać Wartość p.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

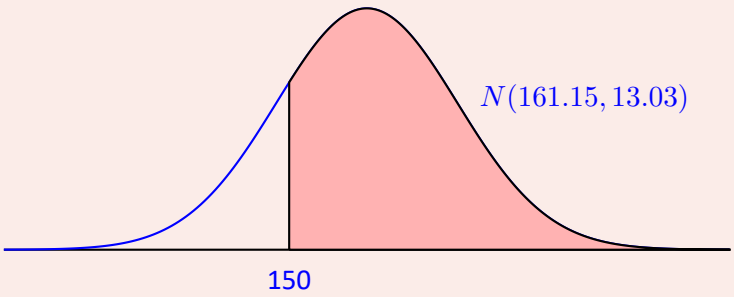
1. By wyliczyć na podstawie rozkładu normalnego (Gaussa), jakie jest prawdopodobieństwo, że klient którego wylosujemy wykorzystał 150 darmowych minut lub mniej, w polu Statystyka wpisujemy wartość 150. Wybrane ustawienia potwierdzamy przyciskiem Oblicz.



Uzyskana Wartość p wynosi 0.193961.

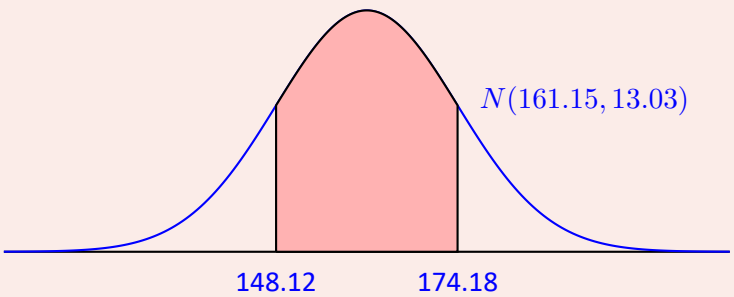
Uwaga! Podobne obliczenia możemy wykonać na podstawie rozkładu empirycznego. Wystarczy wówczas przy pomocy okna Tabele liczności wyznaczyć procent klientów wykorzystujących 150 minut lub mniej *(patrz przykład (11.1), plik: rozkład.pqs)*. W badanej 200-osobowej próbie klientów wykorzystujących 150 minut lub mniej jest 40, co stanowi 20% próby, a zatem szukane prawdopodobieństwo wynosi $P = 0.2$.

2. By wyliczyć na podstawie rozkładu normalnego (Gaussa), jakie jest prawdopodobieństwo, że klient którego wylosujemy wykorzystał więcej niż 150 darmowych minut, w polu Statystyka wpisujemy wartość 150 i zaznaczamy opcję 1- Wartość p. Wybrane ustawienia potwierdzamy przyciskiem Oblicz.



Uzyskana Wartość p wynosi 0.806039.

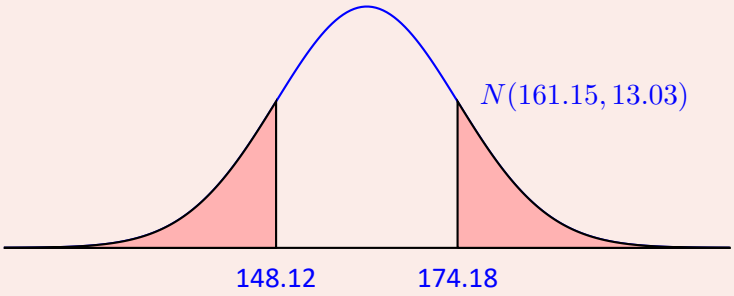
3. By wyliczyć na podstawie rozkładu normalnego (Gaussa), jakie jest prawdopodobieństwo, że klient którego wylosujemy wykorzystał minuty z przedziału $[\overline{x} - sd, \overline{x} + sd] = [148.12min., 174.18min.]$, w polu Statystyka wpisujemy jedną z końcowych wartości przedziału, a następnie zaznaczamy opcję dwustronnie. Wybrane ustawienia potwierdzamy przyciskiem Oblicz.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uzyskana Wartość p wynosi 0.682689.

4. By wyliczyć na podstawie rozkładu normalnego (Gaussa), jakie jest prawdopodobieństwo, że klient którego wylosujemy wykorzystał minuty spoza przedziału $[\overline{x} - sd, \overline{x} + sd] = [148.12min., 174.18min.]$, w polu Statystyka wpisujemy jedną z końcowych wartości przedziału, a następnie zaznaczamy opcje: dwustronnie i 1-wartość p. Wybrane ustawienia potwierdzamy przyciskiem Oblicz.



Uzyskana Wartość p wynosi 0.317311.

DLA ZAINTERESOWANYCH

Rozkład normalny $N(\mu, \sigma)$. Wzór: funkcja gęstości jest zdefiniowana jako:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

gdzie:
$-\infty < x < +\infty$,
μ – wartość oczekiwana populacji (której miarą jest średnia),
σ – odchylenie standardowe.

Rozkład t-Studenta. Wzór: funkcja gęstości jest zdefiniowana jako:

$$f(x, df) = \frac{\Gamma(\frac{df+1}{2})}{\Gamma(\frac{df}{2})\sqrt{df\pi}} \left(1 + \frac{x^2}{df}\right)^{-\frac{df+1}{2}},$$

gdzie:
$-\infty < x < +\infty$,
df – stopnie swobody (liczność próby pomniejszona o liczbę ograniczeń w określonych obliczeniach),
Γ to funkcja Gamma.

Rozkład χ^2. Wzór: funkcja gęstości jest zdefiniowana jako:

$$f(x, df) = \frac{1}{2^{\frac{df}{2}}\Gamma\frac{df}{2}} x^{\frac{df}{2}-1} e^{-\frac{x}{2}},$$

gdzie:
$x > 0$,

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

df – stopnie swobody (liczność próby pomniejszona o liczbę ograniczeń w określonych obliczeniach),
Γ to funkcja Gamma.

Rozkład F Snedecora. Wzór: funkcja gęstości jest zdefiniowana jako:

$$F(x, df_1, df_2) = \frac{\sqrt{\frac{(df_1 x)^{df_1} d_2^{df_2}}{(df_1 x + df_2)^{df_1 + df_2}}}}{x B\left(\frac{df_1}{2}, \frac{df_2}{2}\right)},$$

gdzie:
$x > 0$,
df_1, df_1 – stopnie swobody (przyjmuje się, że jeżeli X i Y są niezależne o rozkładzie χ^2 z odpowiednio df_1 i df_2 stopniami swobody, to $F = \frac{X/df_1}{Y/df_2}$ ma rozkład F Snedecora $F(df_1, df_2)$),
B to funkcja Beta.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

13 TESTOWANIE HIPOTEZ

Uogólnienie wyników otrzymanych dla próby na całą populację dzieli się zasadniczo na 2 części:

- **estymację** – szacowanie wartości parametrów populacji na podstawie próby statystycznej,
- **weryfikację hipotez statystycznych** – sprawdzanie określonych założeń sformułowanych dla parametrów populacji generalnej na podstawie wyników z próby.

ESTYMACJA PUNKTOWA I PRZEDZIAŁOWA

W praktyce najczęściej nie znamy **parametrów** (charakterystyk) całej populacji, a dysponujemy jedynie próbką wylosowaną z tej populacji. **Estymatory punktowe** to charakterystyki wyliczone na podstawie próby losowej, a ich dokładność określa **błąd standardowy**. Prawdziwa wartość parametru populacji znajduje się w okolicy wyznaczonego estymatora punktowego – np. parametr populacyjny średnia arytmetyczna μ znajduje się w okolicy estymatora z próby, jakim jest $\overline{x}$.

Znając teoretyczny rozkład estymatora, możemy z góry określonym **poziomem ufności** $(1-\alpha)$ szacować wartość parametru populacji na podstawie estymatora z próby losowej. Taki zabieg nosi nazwę **estymacji przedziałowej**, a wyznaczony w ten sposób przedział nazywamy **przedziałem ufności** (*ang. confidence interval*); samo α nazywamy zaś **poziomem istotności** (*ang. significance level*).

Najczęściej przyjmuje się poziom istotności równy 0.05, 0.01 lub 0.001.

WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Weryfikacja hipotez statystycznych przebiega w kilku krokach.

Krok 1. Postawienie hipotez. Będą one weryfikowane za pomocą odpowiedniego testu statystycznego. Każdy test podaje ogólną postać hipotezy zerowej – $\mathcal{H}_0$ (*ang. null hypothesis*) i hipotezy alternatywnej – $\mathcal{H}_1$ (*ang. alternative hypothesis*):

$\mathcal{H}_0$: **nie ma** ważnej statystycznie **różnicy** między (średnimi, medianami, proporcjami, rozkładami, itp.) **populacji**,

$\mathcal{H}_1$: **istnieje** ważna statystycznie **różnica** między średnimi, medianami, proporcjami, rozkładami, itp.) **populacji**.

Badacz musi sformułować hipotezę tak, by była zgodna z rzeczywistością i z wymogami wybranego testu statystycznego, np.:

$\mathcal{H}_0$: odsetek kobiet i mężczyzn prowadzących działalność gospodarczą w badanej populacji jest taki sam.

Jeśli nie wiemy, który odsetek (kobiet czy mężczyzn) mógłby być w badanej populacji większy, hipoteza alternatywna powinna być dwustronna, tzn. nie zakładamy kierunku różnicy:

$\mathcal{H}_1$: odsetek kobiet i mężczyzn prowadzących działalność gospodarczą w badanej populacji jest różny.

Tylko w rzadkich przypadkach, gdy mamy pewność co do kierunku spodziewanej różnicy, można zastosować jednostronną hipotezę alternatywną.

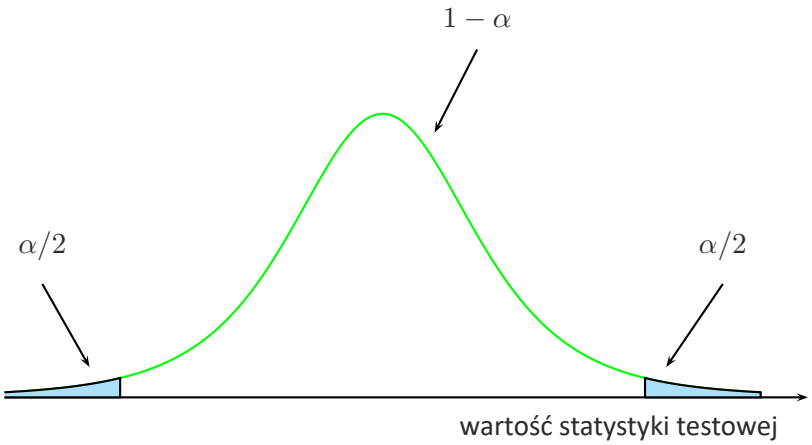
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Krok 2. Wybór testu i wyliczenie statystyki testowej. Sprawdzamy, która z hipotez – $\mathcal{H}_0$ czy $\mathcal{H}_1$ – jest bardziej prawdopodobna. W zależności od rodzaju badania i charakteru zmiennych dobieramy odpowiedni test statystyczny.

Uwaga! Wybór testu statystycznego wiąże się przede wszystkim z odpowiednią skalą pomiarową (interwałową, porządkową, nominalną), jaką reprezentują analizowane dane, oraz z modelem badania (model zależny lub niezależny). Pomiary danej cechy nazywamy **zależnymi (powiązanymi)**, gdy są wykonane kilkukrotnie dla tych samych obiektów. Gdy pomiary danej cechy dotyczą obiektów należących do różnych grup, mówimy o grupach **niezależnych (niepowiązanych)**. Przykłady badań w grupach zależnych: masa ciała pacjentów przed i po terapii odchudzającej, reakcja na bodziec tej samej grupy osób w dwóch różnych warunkach (np. w nocy i w dzień), ocena zdolności kredytowej tych samych klientów dokonana przez 2 różne banki. Przykłady badań w grupach niezależnych: masa ciała w grupie pacjentów zdrowych i grupie pacjentów chorych, efektywność nawożenia kilkoma różnymi rodzajami nawozów, wielkość PKB w różnych krajach.

Uwaga! Zawarty w oknie Kreatora graf ułatwia wybór odpowiedniego testu statystycznego.

Statystyka testowa wybranego testu, wyliczona zgodnie z jej wzorem, podlega odpowiedniemu dla niej rozkładowi teoretycznemu.



Program wylicza wartość statystyki testowej oraz odpowiadającą jej wartość p (czyli tę część pola pod krzywą, która odpowiada wartości statystyki testowej). Wartość p pozwala wybrać spośród hipotezy zerowej i alternatywnej tę bardziej prawdopodobną. Zawsze zakładamy przy tym prawdziwość hipotezy zerowej, a zebrane dane mają dostarczyć wystarczających argumentów przeciwko tej hipotezie:

$$\begin{aligned} &\text{jeżeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ na rzecz } \mathcal{H}_1, \\ &\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Zwykle przyjmuje się poziom istotności $\alpha = 0.05$, godząc się z tym, że w 5% przypadków odrzucimy hipotezę zerową, mimo że jest ona prawdziwa. W szczególnych sytuacjach można wybrać inny poziom istotności, np. 0.01 lub 0.001.

Krok 3. Opis wyników weryfikacji hipotez.

BŁĘDY PIERWSZEGO I DRUGIEGO RODZAJU

Test statystyczny może nie być zgodny z rzeczywistością na dwa sposoby:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

		rzeczywistość	
		$\mathcal{H}_0$: prawdziwa	$\mathcal{H}_0$: fałszywa
wyniki testu	$\mathcal{H}_0$: prawdziwa	OK	β
	$\mathcal{H}_0$: fałszywa	α	OK

Możemy popełnić 2 rodzaje błędów:

α = błąd pierwszego rodzaju – prawdopodobieństwo odrzucenia hipotezy $\mathcal{H}_0$, gdy jest ona prawdziwa,

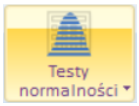
β = błąd drugiego rodzaju – prawdopodobieństwo przyjęcia hipotezy $\mathcal{H}_0$, gdy jest ona fałszywa.

Moc testu to $1 - \beta$.

Wartości α i β są ze sobą powiązane. Przyjętą praktyką jest ustalenie z góry poziomu istotności α i minimalizacja β przez zwiększanie liczności próby.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

14 TESTY NORMALNOŚCI ROZKŁADU



Testy normalności jednowymiarowej

KIEDY STOSOWAĆ:

W badaniu normalności rozkładu zastosowanie mogą mieć różne testy, z których każdy zwraca uwagę na nieco inne aspekty rozkładu Gaussa. Nie można wskazać testu dobrego dla każdego możliwego zestawu danych. Badanie normalności rozkładu można przeprowadzać dla zmiennych lub dla różnic wyznaczonych na podstawie dwóch zmiennych.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej.

DEFINICJE MIAR

Test Kołmogorova-Smirnova *(ang. Kolmogorov-Smirnov test for normality)* – test zaproponowany przez Kołmogorova (1933)[97] jest testem stosunkowo konserwatywnym (trudniej przy jego użyciu udowodnić brak normalności rozkładu). Opiera swoje działanie na wyznaczeniu odległości pomiędzy dystrybuantą empiryczną a teoretyczną rozkładu normalnego. Zaleca się jego stosowanie dla licznych prób, jednak powinien być stosowany w sytuacji, gdy znana jest wartość średnia (μ) i odchylenie standardowe (σ) dla populacji, z której pochodzi próba – wówczas można sprawdzić, czy rozkład jest zgodny z rozkładem zdefiniowanym przez podaną średnią i odchylenie standardowe.

Test Lillieforsa *(ang. Lilliefors test for normality)* – test zaproponowany przez Lillieforsa (1967, 1969, 1973)[106][107][108]. Jest on poprawką testu Kołmogorova-Smirnova, stosowaną, gdy nie jest znana wartość średnia ani odchylenie standardowe dla populacji, z której pochodzi próba. Uznawany jest za nieco mniej konserwatywny od testu Kołmogorova-Smirnova.

Test Shapiro-Wilka *(ang. Shapiro-Wilk test for normality)* – zaproponowany przez Shapiro i Wilka (1965)[152] dla mało licznych grup, a następnie zaadaptowany dla grup liczniejszych (do 5000 obiektów) przez Roystona (1992)[145][146]. Test ten charakteryzuje stosunkowo wysoka moc, co ułatwia dowodzenie braku normalności rozkładu. Ideę działania testu przedstawia wykres Kwantyl-Kwantyl.

Test D'Agostino-Pearsona *(ang. D'Agostino-Pearson test for normality)* – różne typy analiz statystycznych zakładające normalność są w różnym stopniu wrażliwe na różne rodzaje odejścia od tego założenia. Przyjmuje się, że testy odnoszące się w swoich hipotezach do średnich są bardziej wrażliwe na skośność, a testy porównujące wariancje w większym stopniu zależą od kurtozy. Rozkład normalny charakteryzować powinna zerowa skośność i zerowa kurtoza g_2 (lub b_2 bliska wartości trzy). W przypadku braku normalności rozkładu, stwierdzonej przez test D'Agostino (1973)[3], można sprawdzić, czy jest to

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

efektem wysokiej skośności czy kurtozy, poprzez test skośności i test kurtozy. Podobnie jak test Shapiro-Wilka, test D'Agostino charakteryzuje się większą mocą niż test Kołmogorova-Smirnova i test Lillieforsa (D'Agostino, 1990[4]).

Wykres Kwantyl-Kwantyl *(ang. Q-Q plot, Quantile-Quantile plot)* – wykorzystywany jest do przedstawienia zgodności dwóch rozkładów. W przypadku badania zgodności z rozkładem normalnym sprawdza zgodność rozkładu danych (rozkładu empirycznego) z rozkładem teoretycznym Gaussa. Na jego podstawie można wizualnie sprawdzić, jak dobrze krzywa rozkładu normalnego jest dopasowana do danych. Jeśli kwantyle rozkładu teoretycznego i rozkładu empirycznego są zgodne, wówczas punkty rozkładają się wzdłuż linii $y = x$. Oś pozioma przedstawia kwantyle rozkładu normalnego, oś pionowa – kwantyle rozkładu danych.

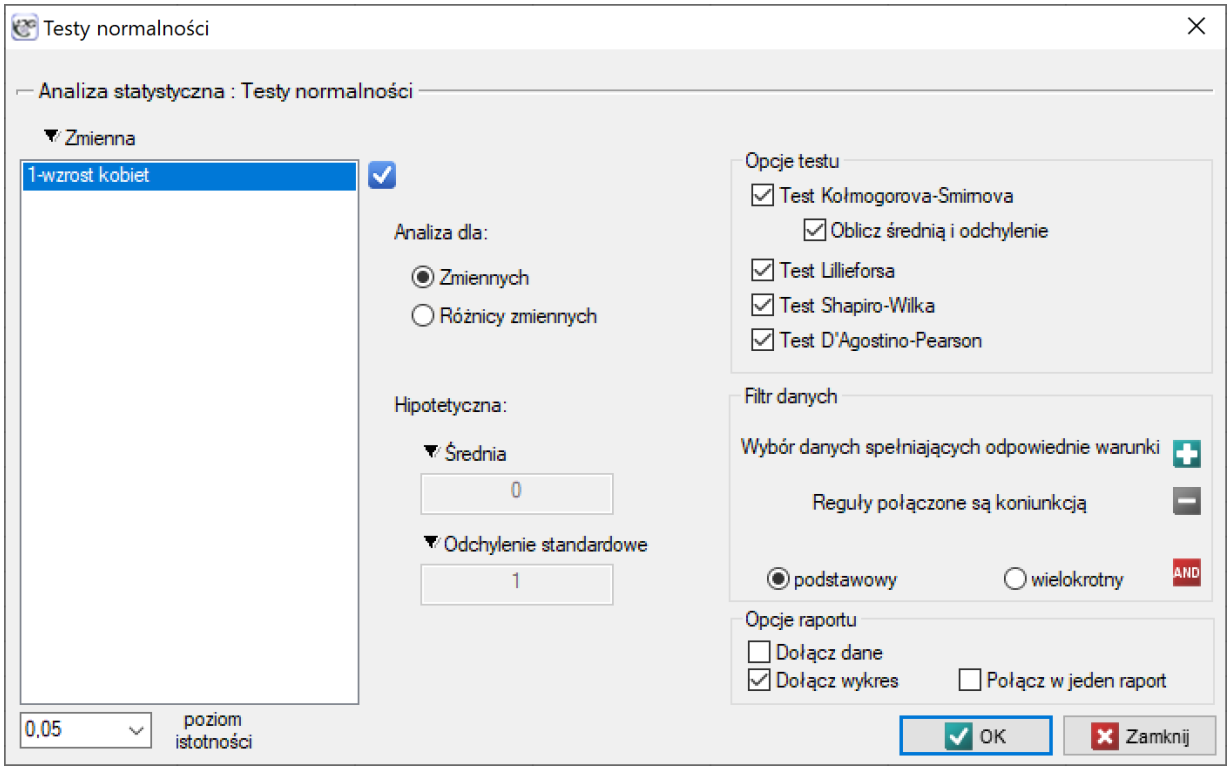
Możliwe są bardzo różne odstępstwa od rozkładu normalnego – interpretację kilku najczęściej występujących opisuje poniższe zestawienie:

- dane rozłożone wzdłuż linii, lecz kilka punktów mocno od niej odbiega – w danych występują wartości odstające;
- punkty po lewej stronie wykresu znajdują się powyżej linii, a po prawej – poniżej – rozkład charakteryzuje większa obecność wartości odległych od średniej niż w rozkładzie normalnym (ujemna kurtoza);
- punkty po lewej stronie wykresu znajdują się poniżej linii, a po prawej – powyżej – rozkład charakteryzuje mniejsza obecność wartości odległych od średniej niż w rozkładzie normalnym (dodatnia kurtoza);
- punkty po lewej i po prawej stronie wykresu znajdują się powyżej linii – rozkład jest prawostronnie skośny (dodatnia skośność);
- punkty po lewej i po prawej stronie wykresu znajdują się poniżej linii – rozkład jest lewostronnie skośny (ujemna skośność).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy normalności → Normalność jednowymiarowa (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

PRZYKŁAD 14.1. (plik Gauss.pqs)

Wzrost kobiet

Załóżmy, że wzrost kobiet jest taką cechą, dla której wartość przeciętna wynosi 168 cm. Większość kobiet spotykanych na co dzień jest wzrostu, który nie odbiega znacznie od tej przeciętnej. Oczywiście zdarzają się kobiety zupełnie niskie, a także bardzo wysokie, ale stosunkowo rzadko. Skoro wartości bardzo niskie i bardzo wysokie występują rzadko, a wartości przeciętne często, możemy się spodziewać, że rozkład wzrostu jest rozkładem normalnym. By się o tym przekonać, zmierzono 300 losowo wybranych kobiet.

Ponieważ nie znamy średniej ani odchylenia standardowego dla wzrostu kobiet, a jedynie mamy przypuszczenia co do tych wielkości, będą one wyznaczane z próby.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

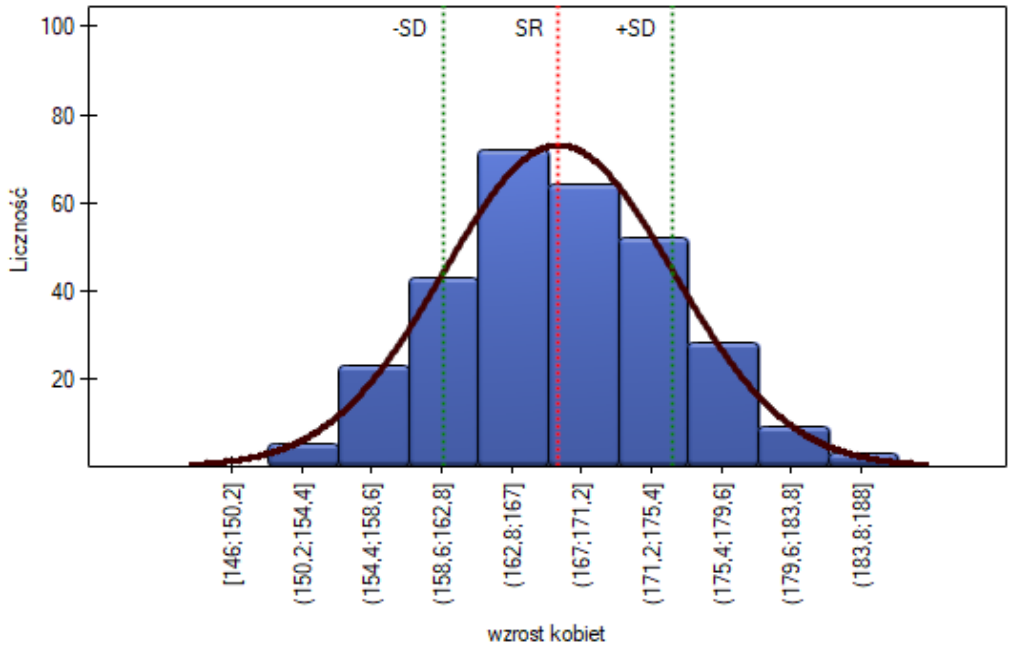
Testy normalności	
Czas analizy	0,15 sek.
Analizowane zmienne	wzrost kobiet
Poziom istotności	0,05
Liczność - grupy	300
Średnia grupy	167,726666667
Odchylenie standardowe grupy	6,886462574
Test Kołmogorova-Smirnova	
Statystyka D	0,042331501
Stopnie swobody	300
Wartość p	0,639671197
Test Lillieforsa	
Statystyka D	0,042331501
Stopnie swobody	300
Wartość p	0,212690323
Test Shapiro-Wilka	
Statystyka W	0,996276564
Statystyka Z	-0,543621967
Wartość p	0,70664918
Test D'Agostino-Pearson	
Statystyka K-kwadrat	0,284183168
Stopnie swobody	2
Wartość p	0,8675418
Test skośności	
Skośność	-0,013223327
Statystyka Z	-0,095228891
Wartość p	0,924133022
Test kurtozy (g2)	
Kurtoza	-0,164654981
Statystyka Z	-0,524513705
Wartość p	0,59992129

wszystkie testy $p > 0.05$ (nieistotne statystycznie, $\alpha = 0.05$)

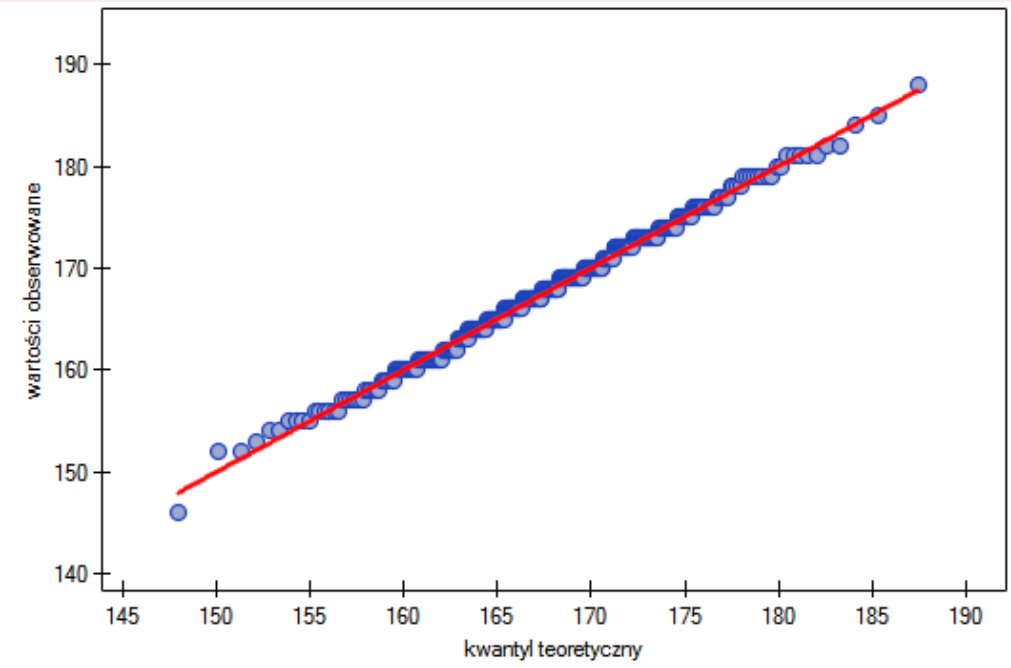
Rozkład wzrostu kobiet w badanej populacji nie odbiega od rozkładu normalnego.

Również test badający skośność i kurtozę nie wykazuje odstępstw. Na wykresie kolumnowym przedstawiliśmy rozkład wzrostu w postaci 10 kolumn. Najliczniejszą grupę stanowią kobiety o wzroście od 167 cm do 171 cm, najmniej liczne są natomiast kobiety niższe niż 150 cm lub wyższe niż 184 cm. Dzwonowa krzywa rozkładu normalnego wydaje się dobrze opisywać ten rozkład.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Na wykresie kwantyl-kwantyl punkty leżą prawie idealnie na linii, co również świadczy o bardzo dobrym dopasowaniu rozkładu normalnego.



Rozkład normalny może zatem zostać uznany za rozkład, jakim charakteryzuje się wzrost kobiet w badanej populacji.

PRZYKŁAD 14.2. (plik Gauss.pqs)
Dochód

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Załóżmy, że badamy dochód osób w pewnym kraju. Oczekujemy, że dochody większości osób będą przeciętne, jednak nie będzie osób zarabiających zupełnie mało (poniżej minimalnej pensji narzuconej przez władze), za to będą osoby zarabiające bardzo dużo (prezesi firm), których jest stosunkowo niewielu. By sprawdzić, czy dochód osób w badanym kraju ma rozkład normalny, zebrano informację o dochodach 264 losowo wybranych osób.

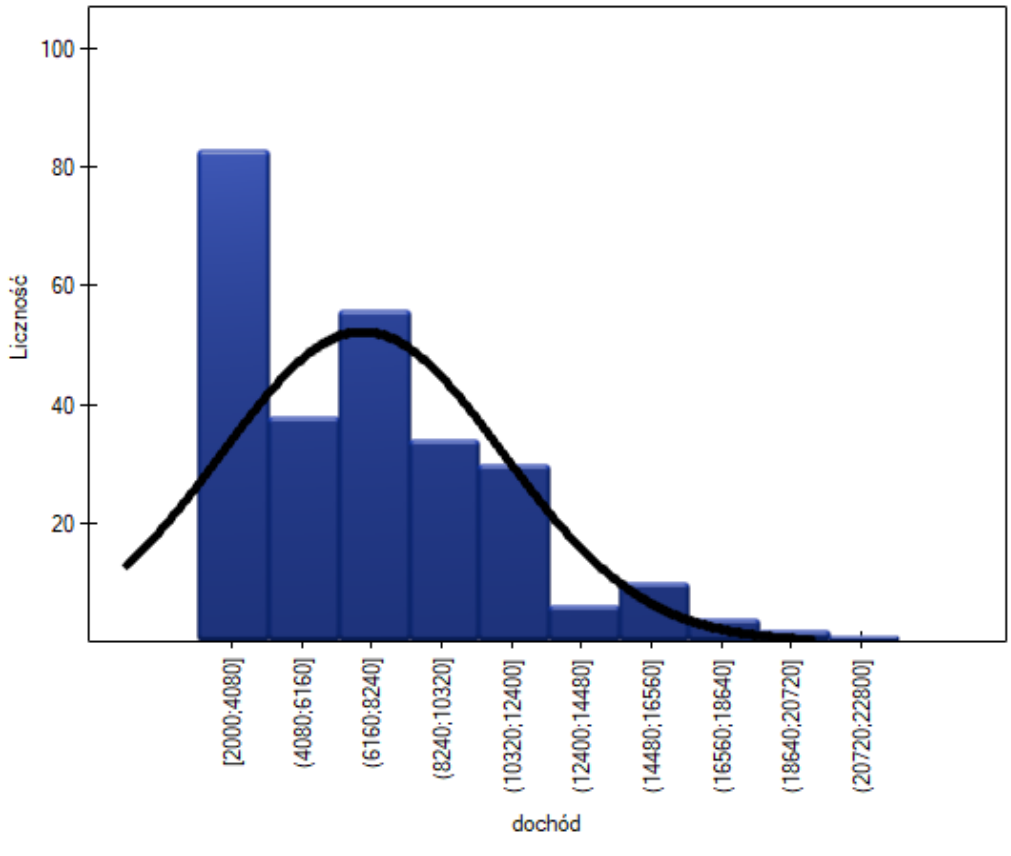
Testy normalności	
Czas analizy	0,05 sek.
Analizowane zmienne	dochód
Poziom istotności	0,05
Liczność - grupy	264
Średnia grupy	6929,924242424
Odchylenie standardowe grupy	4190,689298949
Test Kołmogorova-Smirnova	
Statystyka D	0,119717662
Stopnie swobody	264
Wartość p	0,000937906
Test Lillieforsa	
Statystyka D	0,119717662
Stopnie swobody	264
Wartość p	0,000000001
Test Shapiro-Wilka	
Statystyka W	0,920661679
Statystyka Z	6,332389748
Wartość p	<0.000000001
Test D'Agostino-Pearson	
Statystyka K-kwadrat	28,158080383
Stopnie swobody	2
Wartość p	0,000000768
Test skośności	
Skośność	0,82511525
Statystyka Z	4,959279001
Wartość p	0,000000708
Test kurtozy (g2)	
Kurtoza	0,656828436
Statystyka Z	1,887758505
Wartość p	0,059058378

wszystkie testy $p < \alpha$ (istotne statystycznie, $\alpha = 0.05$)

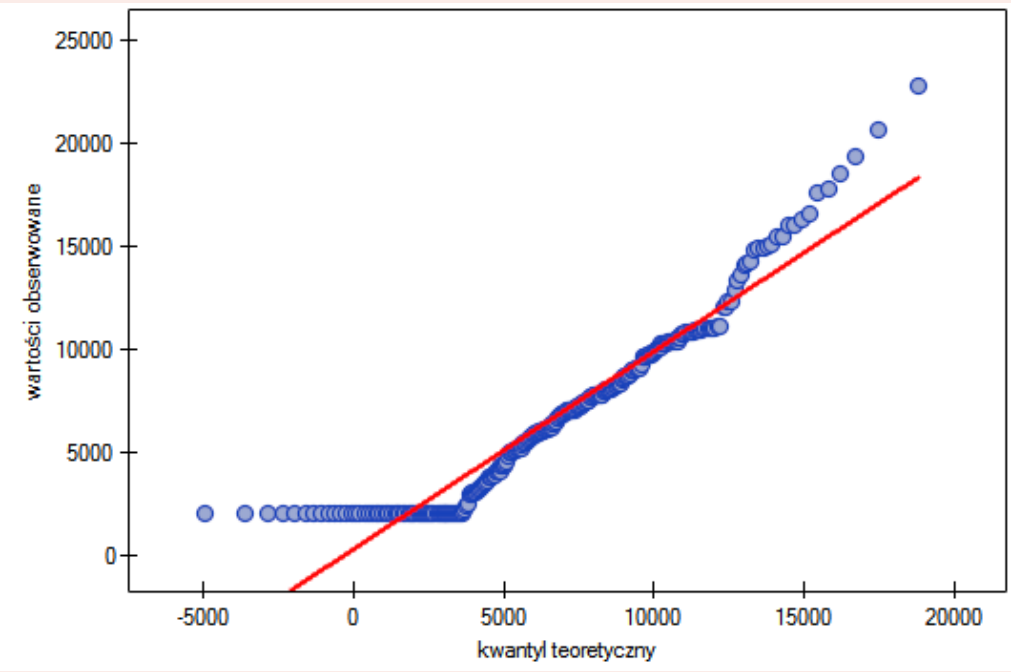
Rozkład dochodów osób w badanej populacji nie jest rozkładem normalnym.

Dodatnia i istotna statystycznie ($p < \alpha$) wartość skośności świadczy o zbyt długim prawym ogonie funkcji. Rozkład jest również bardziej smukły od rozkładu normalnego, ale nie jest to różnica istotna statystycznie (test kurtozy).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Na wykresie kwantyl-kwantyl odstępstwo od rozkładu normalnego jest obrazowane poprzez skośność prawostronną, czyli położenie znacznie powyżej linii początkowych i końcowych punktów wykresu.



Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

W rezultacie zebrane dane nie świadczą o zgodności rozkładu dochodów z rozkładem normalnym.

DLA ZAINTERESOWANYCH

Hipotezy testów normalności rozkładu:

$\mathcal{H}_0$: rozkład badanej cechy w populacji jest rozkładem normalnym,
$\mathcal{H}_1$: rozkład badanej cechy w populacji jest różny od rozkładu normalnego.

Interpretacja: wyznaczoną na podstawie statystyki testowej wybranego testu wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test Kołmogorova-Smirnova
Wzór: w oparciu o dane z próby zebrane w skumulowany rozkład częstości oraz o odpowiednie wartości pola pod teoretyczną krzywą rozkładu normalnego wyznaczamy wartość statystyki testowej D:

$$D = \sup_x |F_n(x) - F(x)|,$$

gdzie:
$F_n(x)$ – empiryczna dystrybuanta rozkładu normalnego wyliczana w poszczególnych punktach rozkładu, dla n-elementowej próby,
$F(x)$ – teoretyczna dystrybuanta rozkładu normalnego.
Statystyka testu podlega rozkładowi Kołmogorova-Smirnova.

Test Lillieforsa
Wyjaśnienie: statystyka testowa D wyznaczana jest na podstawie tej samej formuły, z której korzysta test Kołmogorova-Smirnova, ale podlega rozkładowi Lillieforsa.

Test Shapiro-Wilka
Wzory: statystyka testowa Shapiro-Wilka ma postać:

$$W = \frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

gdzie:
a_i – współczynniki wyznaczane w oparciu o wartości oczekiwane dla statystyk uporządkowanych (*ang. ordered statistics*), przypisanych wag oraz macierzy kowariancji,
$\overline{x}$ – wartość średnia danych z próby.

Statystykę tę przekształca się do statystyki o rozkładzie normalnym:

$$Z = \frac{g(W) - \mu}{\sigma},$$

gdzie $g(W)$, μ i σ zależą od wielkości próby:

- dla prób małych o licznościach $n \in \langle 4; 12)$:

  $g(W) = -\ln(\gamma - \ln(1 - W))$,
  $\gamma = 0.459n - 2.273$,
  $\mu = -0.0006714n^3 + 0.025054n^2 - 0.39978n + 0.5440$,
  $\sigma = \exp(-0.0020322n^3 + 0.062767n^2 - 0.77857n + 1.3822)$;

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- dla prób dużych o licznościach $n \in \langle 12; 5000 \rangle$:

$$g(W) = \ln(1 - W),$$
$$\mu = 0.0038915u^3 - 0.083751u^2 - 0.31082u - 1.5851,$$
$$\sigma = \exp(0.0030302u^2 - 0.082676u - 0.4803),$$
$$u = \ln(n).$$

Test D'Agostino-Pearsona
Wzory: statystyka testowa ma postać:

$$K^2 = Z_A^2 + Z_K^2,$$

gdzie:
Z_A^2 – statystyka testowa testu skośności,
Z_K^2 – statystyka testowa testu kurtozy.
Statystyka ta ma asymptotycznie rozkład χ^2 z dwoma stopniami swobody.

Test skośności D'Agostino
Hipotezy:

$\mathcal{H}_0$: rozkład nie jest skośny (skośność w populacji wynosi zero),
$\mathcal{H}_1$: rozkład jest skośny (skośność w populacji odbiega od zera).

Wzory: statystyka testowa ma postać:

$$Z_A = \delta \ln\left(\frac{Y}{\alpha} + \sqrt{\frac{Y^2}{\alpha^2} + 1}\right),$$

gdzie:
$Y = \sqrt{(b_1)}\sqrt{\frac{(n+1)(n+3)}{6(n-2)}}$,
$\sqrt{(b_1)} = \frac{m_3}{m_2^{(3/2)}}$,
$m_k = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^k}{n}$,
$\beta(\sqrt{(b_1)}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$,
$W^2 = -1 + \sqrt{2(\beta(\sqrt{(b_1)}) - 1}$,
$\delta = \frac{1}{\sqrt{\ln W}}$,
$\alpha = \sqrt{\frac{2}{W^2-1}}$.
Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Test kurtozy D'Agostino
Hipotezy:

$\mathcal{H}_0$: kurtoza w populacji odpowiada kurtozie rozkładu normalnego,
$\mathcal{H}_1$: kurtoza w populacji różni się od kurtozy rozkładu normalnego.

Wzory: statystyka testowa ma postać:

$$Z_K = \frac{(1 - \frac{2}{9H}) - \left(\frac{1-\frac{2}{A}}{1+x\sqrt{\frac{2}{H-4}}}\right)^{1/3}}{\sqrt{\frac{2}{9H}}},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie:
$E(b_2) = \frac{3(n-1)}{n+1}$,
$b_2 = \frac{m_4}{m_2^2}$,
$var(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$,
$x = \frac{b_2 - E(b_2)}{\sqrt{var(b_2)}}$,
$\sqrt{\beta(b_2)} = \frac{6(n^2-5n+2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-5)}}$,
$H = 6 + \frac{8}{\sqrt{\beta(b_2)}}\left(\frac{2}{\sqrt{\beta(b_2)}} + \sqrt{1 + \frac{4}{\beta(b_2)}}\right)$.
Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Testy normalności wielowymiarowej

KIEDY STOSOWAĆ:

Wiele metod analizy wielowymiarowej, w tym MANOVA, testy Hotellinga czy modele regresji, opiera się na założeniu normalności wielowymiarowej. Jeśli zbiór zmiennych charakteryzuje wielowymiarowy rozkład normalny, to każda zmienna posiada rozkład normalny. Jednak gdy wszystkie pojedyncze zmienne charakteryzowane są rozkładem normalnym, ich zestaw nie musi mieć wielowymiarowego rozkładu normalnego – dlatego testowanie jednowymiarowej normalności każdej zmiennej może być przydatne, ale nie można założyć, że jest wystarczające.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej dla wszystkich analizowanych zmiennych.

DEFINICJE MIAR

Różne typy analiz statystycznych zakładające normalność są w różnym stopniu wrażliwe na różne rodzaje odejścia od tego założenia. Przyjmuje się, że testy odnoszące się w swoich hipotezach do średnich są bardziej wrażliwe na skośność, a testy porównujące kowariancje w większym stopniu zależą od kurtozy.

Test Mardia *(ang. Mardia's test for multivariate normality)* – test zaproponowany przez Mardia w 1970 roku[118] i zmodyfikowany w 1974 roku[119] bada normalność rozkładu, analizując oddzielnie rozmiar wielowymiarowej skośności i wielowymiarowej kurtozy.

Test Jarque-Bera *(ang. Jarque-Bera test for multivariate normality)* – Jarque i Bera[90] zaproponowali złączenie dwóch miar Mardia (skośności i kurtozy) w jeden test.

Test Hanusz-Tarasińska *(ang. Hanusz-Tarasinska test for multivariate normality)* – podobny sposób łączenia w jeden test informacji o skośności i kurtozie oferuje metoda Hanusz i Tarasińskiej (2014)[78].

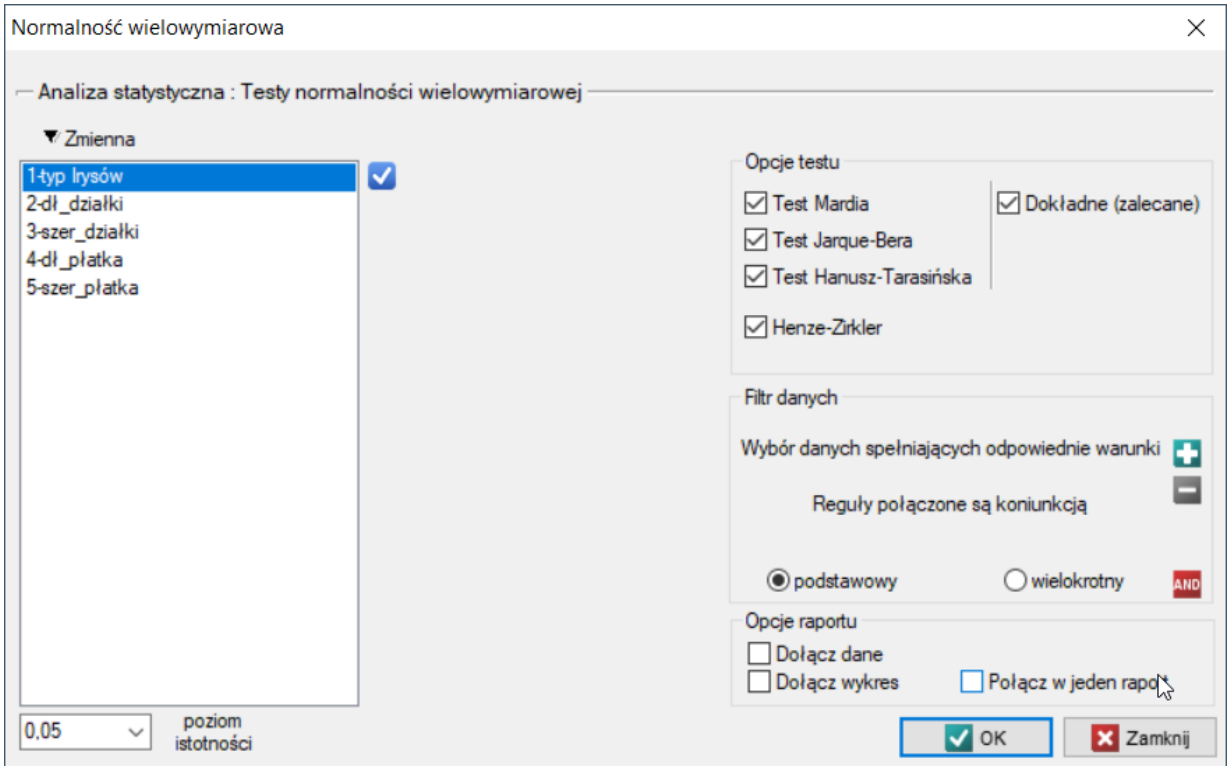
Test Henze-Zirklera *(ang. Henze-Zirkler test for multivariate normality)* – Henze i Zirkler (1990)[81] zaproponowali test badający wielowymiarową normalność rozkładu, rozszerzając pracę Baringhausa

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

i Henzego nad empiryczną funkcją charakterystyczną[55]. W literaturze jest to test uznawany za jeden z najsilniejszych testów poświęconych wielowymiarowemu rozkładowi normalnemu (Thode, 2002)[166].

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy normalności → Normalność wielowymiarowa.



WYKONANIE I INTERPRETACJA

Przykład 14.3. (plik Irysy.pqs)
Badamy normalność rozkładu dla klasycznego zestawu danych R.A. Fishera z 1936 roku[59]. Plik znajduje się w pomocy dołączonej do programu i zawiera pomiary długości i szerokości płatków oraz działek kielicha dla 3 odmian kwiatu irysa. Analiza zostanie przeprowadzona oddzielnie dla każdej odmiany.

W oknie analizy zaznaczamy wszystkie testy oraz wykres, a w celu powtórzenia analizy dla każdej odmiany irysa ustawiamy filtr wielokrotny. Wszystkie wyniki zwrócimy do tego samego arkusza, dlatego zaznaczamy opcję Połącz w jeden raport.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

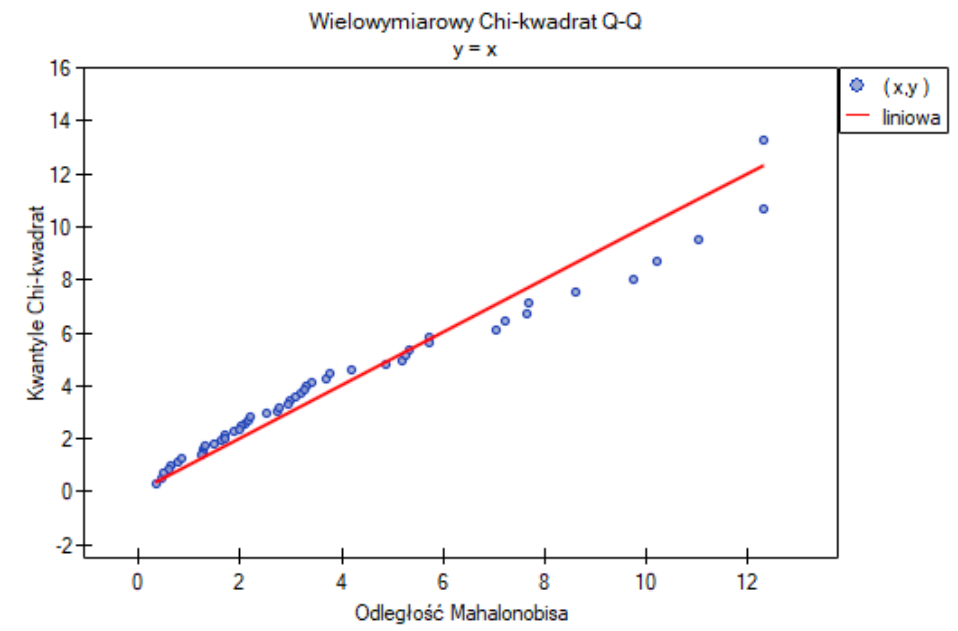
Testy normalności wielowymiarowej	ID1	ID2	ID3
Czas analizy	0,13 sek.	0,13 sek.	0,12 sek.
Analizowane zmienne	dł_działki;szer.	dł_działki;szer.	dł_działki;szer.
Liczność - niezinterpretowane	0	0	0
Liczność - braki danych	0	0	0
Poziom istotności	0,05	0,05	0,05
Filtr danych	typ Irysów=se	typ Irysów=ve	typ Irysów=vir
Liczność	50	50	50
Liczba zmiennych w modelu	4	4	4
Dokładne momenty	Tak	Tak	Tak
Test skośności Mardia			
Skośność	3,0797	3,0222	3,1525
Statystyka Chi-kwadrat	27,8597	27,3394	28,5178
Stopnie swbody	20	20	20
Wartość p	0,1128	0,126	0,0977
Test kurtozy Mardia			
Kurtoza	26,5377	22,8794	24,2991
Statystyka Z	2,1926	-0,1131	0,7817
Wartość p	0,0283	0,9099	0,4344
Test Jarque-Bera			
Statystyka Chi-kwadrat	32,6674	27,3522	29,1289
Stopnie swbody	21	21	21
Wartość p	0,05	0,1595	0,1109
Test Hanusz-Tarasińska			
Statystyka t	1,8578	-0,0967	0,6546
Stopnie swbody	20	20	20
Wartość p	0,078	0,9239	0,5202
Test Henze-Zirkler			
Statystyka	0,9488	0,8388	0,757
Wartość p	0,05	0,2262	0,497

wszystkie testy $p > 0.05$ dla odmian versicolor i virginica; wyniki na pograniczu istotności dla odmiany setosa

Rozkład danych nie odbiega od wielowymiarowego rozkładu normalnego dla odmian versicolor i virginica; dla odmiany setosa wyniki są niejednoznaczne.

Dla odmiany setosa test Mardia dla kurtozy oraz test Henze-Zirkler wskazują na odstępstwa od wielowymiarowego rozkładu normalnego. Możemy obserwować takie odstępstwa również na poniższym wykresie, gdzie wraz z narastaniem odległości Mahalanobisa punkty znajdują się coraz dalej od prostej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



DLA ZAINTERESOWANYCH

Hipotezy (wspólne dla wszystkich testów):

$\mathcal{H}_0$: rozkład danych w populacji = wielowymiarowy rozkład normalny,
$\mathcal{H}_1$: rozkład danych w populacji $\neq$wielowymiarowy rozkład normalny.

Test Mardia

Wzory: Mardia zdefiniował wielowymiarową skośność i kurtozę następująco:

$$skew = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}^3, \quad kurt = \frac{1}{n}\sum_{i=1}^{n} m_{ii}^2$$

gdzie:
$m_{ij} = \left(X_i - \bar{X}\right)^T S^{-1} \left(X_j - \bar{X}\right)$,
$S = \frac{1}{n}\sum_{j=1}^{n} \left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^T$,
$\bar{X}$ – średnia, S – macierz kowariancji.
Dla danych pochodzących z próby, a nie z populacji, wzory na skośność i kurtozę są mnożone odpowiednio: skośność przez $(n/(n-1))^3$, a kurtoza przez $(n/(n-1))^2$.

Test skośności: gdy próba pochodzi z populacji o wielowymiarowym rozkładzie normalnym (hipoteza zerowa), statystyka testowa ma postać (Mardia, 1970):

$$\chi^2(M) = \frac{n}{6} skew$$

lub z poprawką dokładnych momentów dla grup o mniejszych licznościach (< 20) (Mardia, 1974):

$$\chi_c^2(M) = \frac{n}{6}\frac{(n+1)(n+3)(k+1)}{n((n+1)(k+1)-6)} skew$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $df = f = \frac{k(k+1)(k+2)}{6}$ stopniami swobody.

Test kurtozy: gdy próba pochodzi z populacji o wielowymiarowym rozkładzie normalnym (hipoteza zerowa), statystyka testowa ma postać (Mardia, 1974):

$$Z(M) = \frac{kurt - k(k+2)}{\sqrt{\frac{8k(k+2)}{n}}}$$

lub z poprawką (Mardia, 1974):

$$Z_c(M) = \frac{(n+1)kurt - k(k+2)(n-1)}{\sqrt{\frac{8k(k+2)(n-3)(n-k-1)(n-k+1)}{(n+3)(n+5)}}}$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład normalny.

Interpretacja: wyznaczoną na podstawie statystyki testowej wartość p dla obu testów (skośności i kurtozy) porównujemy z poziomem istotności α: gdy $p \leq \alpha$ dla przynajmniej jednego testu, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$ dla obu zastosowanych testów, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test Jarque-Bera
Wzory: test bazuje na statystyce skośności i kurtozy testu Mardia. Statystyka testowa ma postać:

$$\chi^2(JB) = \chi^2(M) + (Z(M))^2$$

lub z poprawką (Mardia, 1974):

$$\chi^2_c(JB) = \chi^2_c(M) + (Z_c(M))^2$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $df = f + 1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test Hanusz-Tarasińska
Wzory: test bazuje na statystyce skośności i kurtozy testu Mardia. Statystyka testowa ma postać:

$$t_c(HT) = \frac{Z_c(M)}{\sqrt{\frac{\chi^2_c(M)}{f}}}$$

Statystyka testowa ma rozkład t-Studenta z $df = f$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test Henze-Zirklera
Wzory: statystyka testowa ma postać:

$$Z(HZ)_\beta = n\left(4I_E + D_{n,\beta} I_{E^c}\right)$$

gdzie:
I_E oraz I_{E^c} – funkcje indykatorowe zależne od osobliwości macierzy kowariancji,
$D_{n,\beta} = \frac{1}{n^2}\sum \exp\left(\frac{-\beta^2||Y_j - Y_k||^2}{2}\right) + (1+2\beta^2)^{-p/2} - 2(1+\beta^2)^{-p/2}\sum \exp\left(\frac{-\beta^2||Y_j||^2}{2(1+\beta^2)}\right)$,

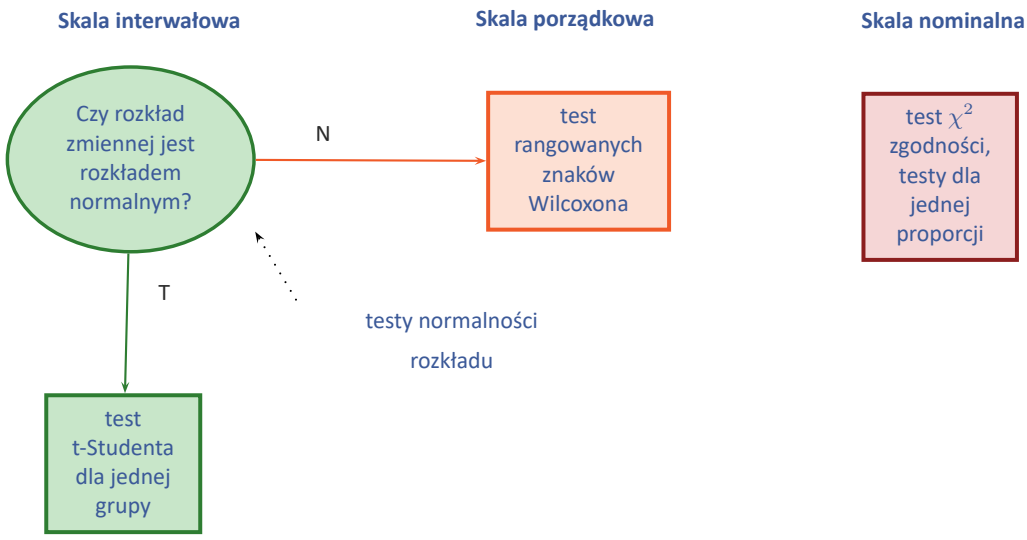
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

$Y_i = S^{1/2}(X_i - \bar{X})$,
$\beta* = 2^{-1/2}\left(\frac{n(2k+1)}{4}\right)^{1/(k+4)}$ – optymalna wartość parametru β.

Statystyka $Z(HZ)_\beta$ ma asymptotycznie (dla dużych liczności) rozkład normalny oparty na średniej i wariancji opisanej przez Henze i Zirklera i odczytywany jednostronnie.

Wyznaczoną na podstawie statystyki testowej jednostronną wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

15 PORÓWNANIE - 1 GRUPA



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

15.1 TESTY PARAMETRYCZNE

Test t-Studenta dla pojedynczej próby

KIEDY STOSOWAĆ TEN TEST:

Test t-Studenta dla pojedynczej próby *(ang. single-sample t test)* służy do weryfikacji hipotezy, że badana próba o średniej $\overline{x}$ pochodzi z populacji, dla której średnia μ to zadana wartość.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanej cechy.

Uwaga! Gdy próba jest duża i znane jest odchylenie standardowe z populacji, statystykę testową można wyznaczyć z wzoru $t = \frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}$, mającego wówczas rozkład normalny. Przy $n \to \infty$ rozkład t-Studenta jest zbieżny do rozkładu normalnego $N(0,1)$ – w praktyce przyjmuje się, że dla $n > 30$ rozkład t-Studenta można aproksymować rozkładem normalnym. Obliczenia mogą bazować na danych w postaci surowych rekordów lub danych uśrednionych, tzn. średniej arytmetycznej, odchyleniu standardowym i liczności próby.

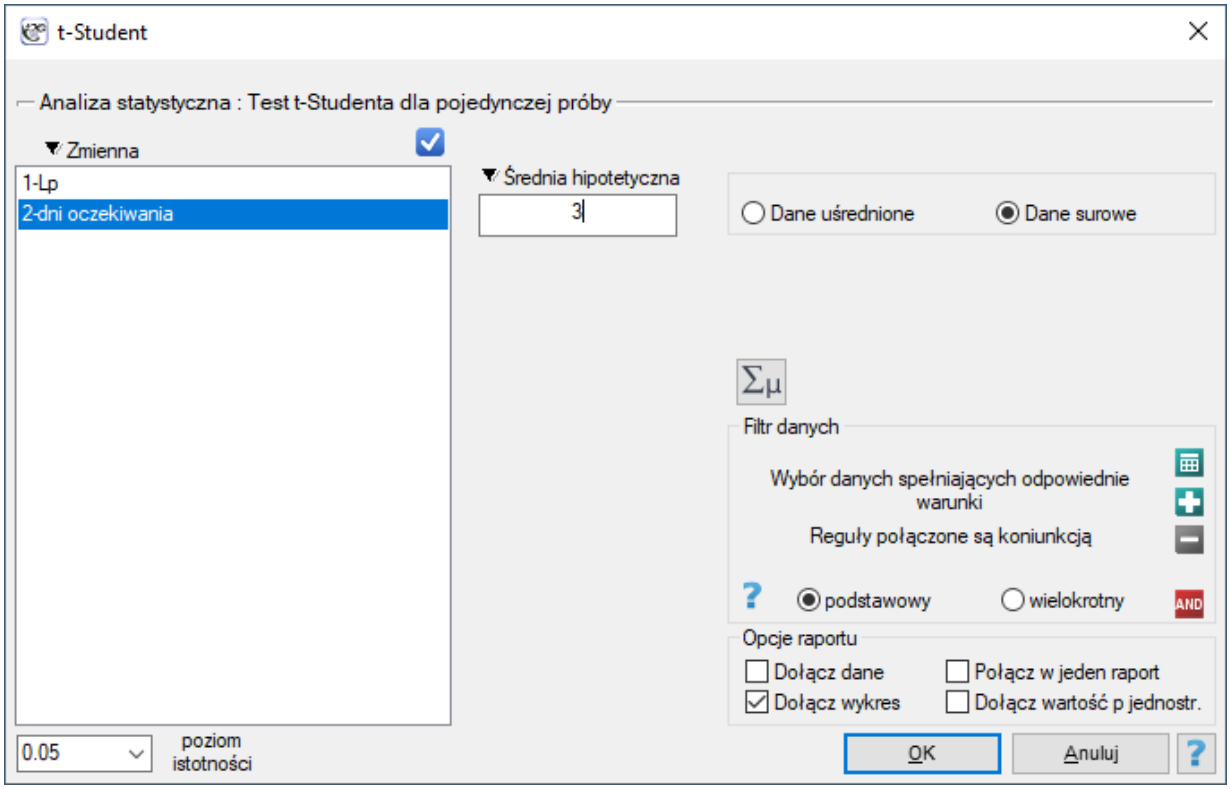
DEFINICJE MIAR

d-Cohena – standaryzowana wielkość efektu, która określa, jak dużą częścią występującej zmienności jest różnica między średnimi. Przy interpretacji efektu badacze często posługują się ogólnymi wskazówkami określonymi przez Cohena[46], definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Testy parametryczne $\rightarrow$ t-Student (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

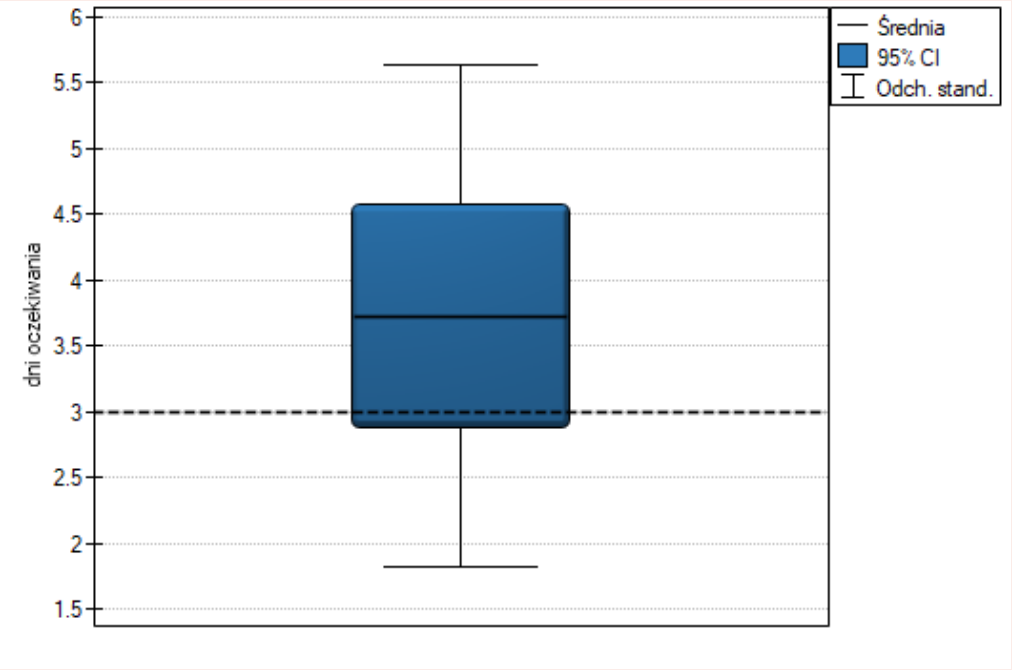
PRZYKŁAD 15.1. (plik kurier.pqs)
Chcemy sprawdzić, czy czas oczekiwania na dostarczenie przesyłki przez pewną firmę kurierską to przeciętnie 3 dni ($\mu_0 = 3$). W tym celu z populacji klientów tej firmy wylosowano próbę liczącą 22 osoby i zapisano informacje o liczbie dni, jakie minęły od dnia nadania przesyłki do jej dostarczenia; były to następujące wielkości: (1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 6, 6, 6, 7, 7).

Liczba dni oczekiwania na przesyłkę w badanej populacji spełnia założenie normalności rozkładu.

Test t-Studenta dla pojedynczej próby	
Czas analizy	0.10 sek.
Analizowane zmienne	dni oczekiwania
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Liczność - grupy	22
Średnia hipotetyczna	3
Różnica średnich	0.7273
-95% CI dla różnicy średnich	-0.1182
+95% CI dla różnicy średnich	1.5728
Statystyka t	1.7889
Stopnie swobody	21
Wartość p dwustronne	0.0881

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Podsumowanie	
Grupa	dni oczekiwa
Liczność	22
Średnia arytmetyczna	3.7273
Błąd standardowy średniej	0.4066
Odchylenie standardowe	1.9069
-95% CI dla średniej grupy	2.8818
+95% CI dla średniej grupy	4.5728



$p = 0.088074$ **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Nie ma podstaw, by odrzucić hipotezę zerową mówiącą, że średnia liczba dni oczekiwania na dostarczenie przesyłki przez analizowaną firmę kurierską wynosi 3.

Dla badanej próby średnia to $\overline{x} = 3.727$, a odchylenie standardowe $sd = 1.907$.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu = \mu_0, \\ \mathcal{H}_1: &\quad \mu \neq \mu_0, \end{aligned}$$

gdzie:
μ – średnia cechy w populacji reprezentowanej przez próbę,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

μ_0 – zadana wartość.

Wzory:

Statystyka testowa ma postać:

$$t = \frac{\overline{x} - \mu_0}{sd}\sqrt{n},$$

gdzie:
sd – odchylenie standardowe z próby,
n – liczność próby.

Statystyka testowa ma rozkład t-Studenta z $n - 1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

d-Cohena

$$d = \left| \frac{\overline{x} - \mu_0}{sd} \right|$$

Test chi-kwadrat wariancji pojedynczej próby

KIEDY STOSOWAĆ TEN TEST:

Test χ^2 wariancji pojedynczej próby *(ang. The Single-Sample Chi-square Test for a Population Variance)* służy do weryfikacji hipotezy, że badana próba pochodzi z populacji, dla której wariancja (lub odchylenie standardowe σ) to zadana wartość. Hipotezy mogą dotyczyć zarówno wariancji, jak i równoważnie odchylenia standardowego.

WARUNKI STOSOWANIA

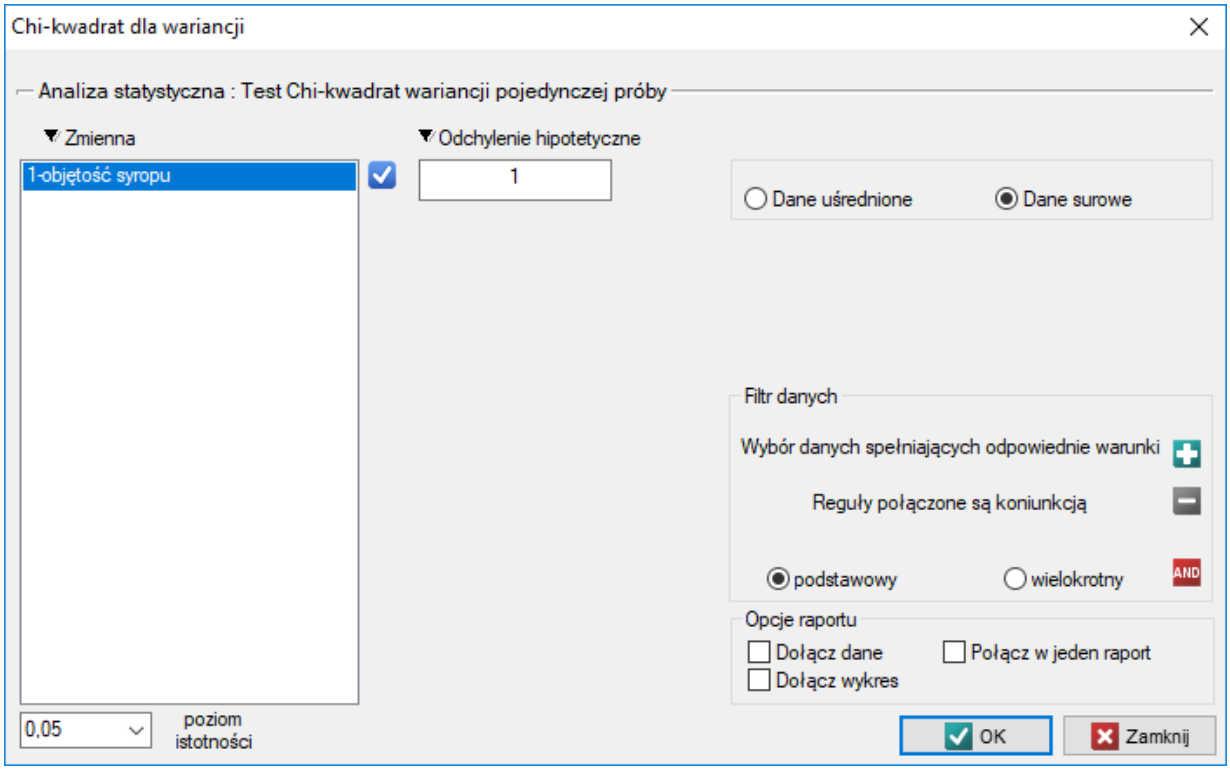
- pomiar na skali interwałowej,
- normalność rozkładu badanej cechy.

Uwaga! Obliczenia mogą bazować na danych w postaci surowych rekordów lub danych uśrednionych, tzn. odchyleniu standardowym i liczności próby.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → Chi-kwadrat dla wariancji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



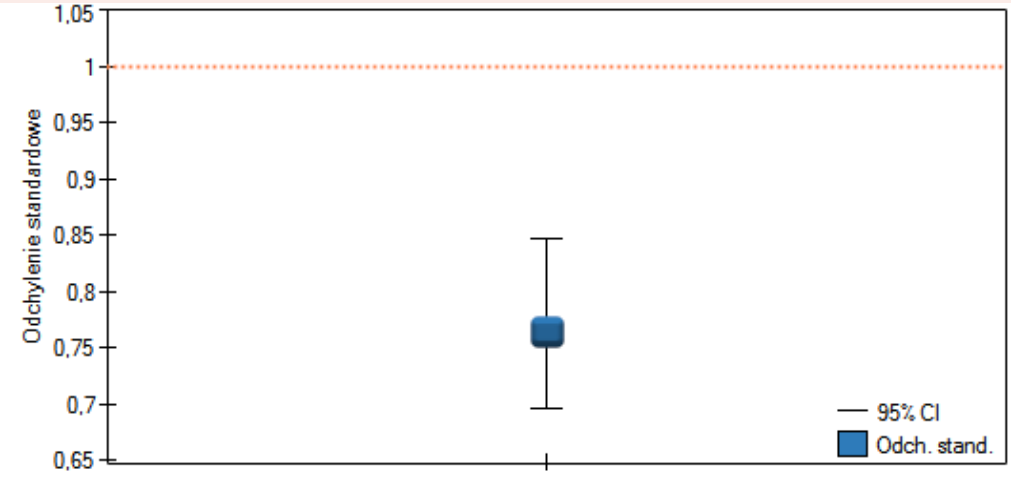
WYKONANIE I INTERPRETACJA

PRZYKŁAD 15.2. (plik dozownik.pqs)
Przed przystąpieniem do produkcji kolejnej partii pewnego syropu przeciwkaszlowego dokonano kontrolnych pomiarów objętości syropu nalewanego do butelek. Z dokumentacji technicznej urządzenia dozującego wynika, że dopuszczalny rozrzut objętości syropu mierzony za pomocą odchylenia standardowego wynosi 1 ml. Należy sprawdzić, czy testowane urządzenie działa prawidłowo.

Sprawdzono (testem Lillieforsa) normalność rozkładu objętości syropu nalewanego do butelek, uzyskując wynik zgodny z tym rozkładem. Analiza dotycząca odchylenia standardowego może być więc przeprowadzona testem chi-kwadrat dla wariancji.

Test Chi-kwadrat wariancji pojedynczej próby	
Czas analizy	0,08 sek.
Analizowane zmienne	objętość syropu
Poziom istotności	0,05
Liczność - grupy	200
Średnia grupy	199,98475
Odchylenie hipotetyczne	1
Odchylenie standardowe grupy	0,764631
-95% CI dla odchylenia standardowego	0,696321
+95% CI dla odchylenia standardowego	0,847919
Statystyka Chi-kwadrat	116,347588
Stopnie swobody	199
Wartość p	0,000001

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p = 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Rozrzut urządzenia dozującego jest inny niż 1 ml.

Działanie urządzenia możemy jednak uznać za prawidłowe, ponieważ odchylenie standardowe w próbie wynosi 0.76 i jest istotnie mniejsze niż dopuszczalna wartość wynikająca z dokumentacji technicznej.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \sigma = \sigma_0, \\ \mathcal{H}_1: &\quad \sigma \neq \sigma_0, \end{aligned}$$

gdzie:
σ – odchylenie cechy w populacji reprezentowanej przez próbę,
σ_0 – zadana wartość.

Wzory: Statystyka testowa ma postać:

$$\chi^2 = \frac{(n-1)sd^2}{\sigma_0^2},$$

gdzie:
sd – odchylenie standardowe z próby,
n – liczność próby.

Interpretacja: Statystyka ta ma rozkład χ^2 z liczbą stopni swobody wyznaczaną według wzoru $df = n - 1$.

Wyznaczoną na podstawie wartości statystyki i rozkładu χ^2 wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Przy czym, jeśli wartość odchylenia standardowego jest mniejsza niż wartość zadana, wartość p wyliczana jest jako podwojona wartość pola pod krzywą rozkładu chi-kwadrat na lewo od odpowiedniej

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

wartości krytycznej, a jeśli jest większa niż wartość zadana – jako podwojona wartość odpowiedniego pola po prawej stronie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

15.2 TESTY NIEPARAMETRYCZNE

Test Wilcoxona (rangowanych znaków)

KIEDY STOSOWAĆ TEN TEST:

Test Wilcoxona rangowanych znaków *(ang. Wilcoxon signed-ranks test)*, znany również pod nazwą testu Wilcoxona dla pojedynczej próby (Wilcoxon, 1945, 1949[175]), służy do weryfikacji hipotezy, że badana próba pochodzi z populacji, dla której mediana (θ) to znana wartość.

WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej.

Uwaga! W zależności od wielkości próby statystyka testowa przyjmuje inną postać – statystykę T dla małej liczności próby lub statystykę Z dla próby o dużej liczności (zob. *Dla zainteresowanych*).

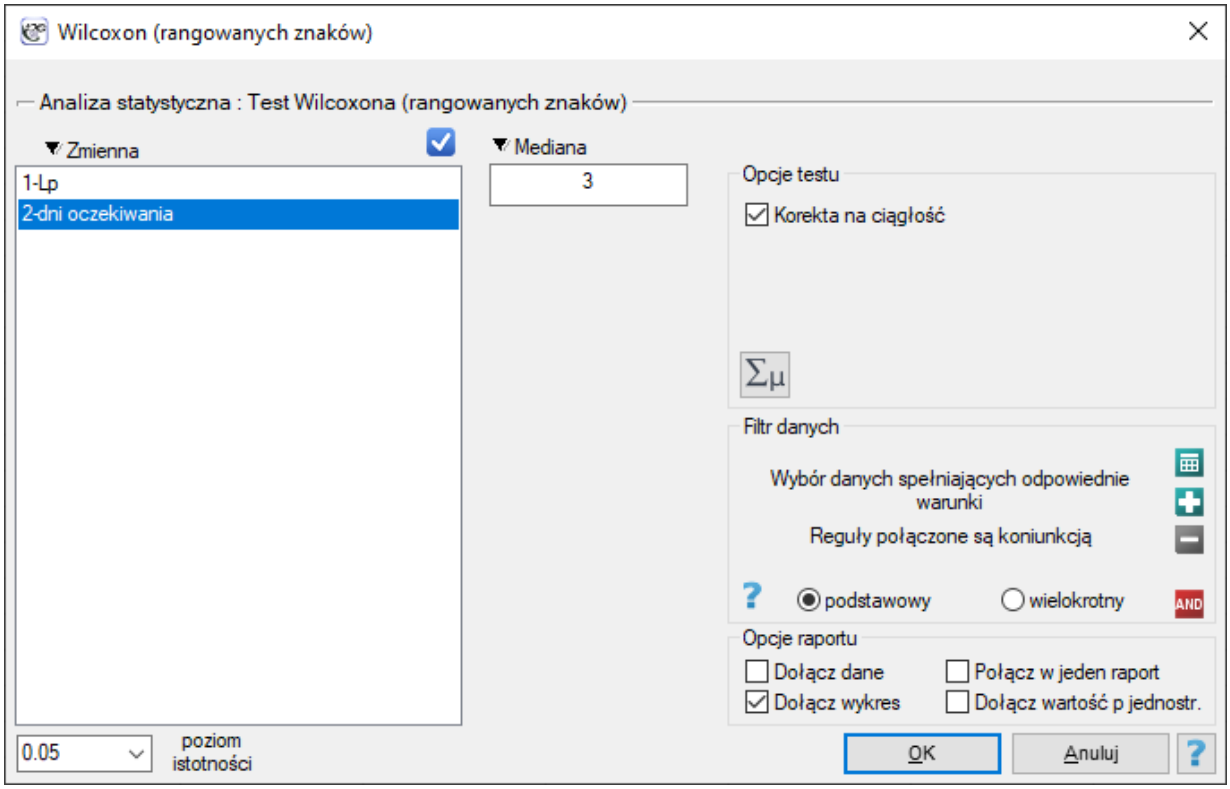
DEFINICJE MIAR

d-Cohena – standaryzowana wielkość efektu. Rozkład statystyki testu Wilcoxona jest aproksymowany przez rozkład normalny, który można przekształcić na wielkość efektu $r = |Z/n|$[67], by następnie uzyskać wartość d-Cohena zgodnie ze standardową konwersją stosowaną przy meta-analizach. Przy interpretacji efektu badacze często posługują się ogólnymi wskazówkami określonymi przez Cohena[46], definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Wilcoxon (rangowanych znaków) (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



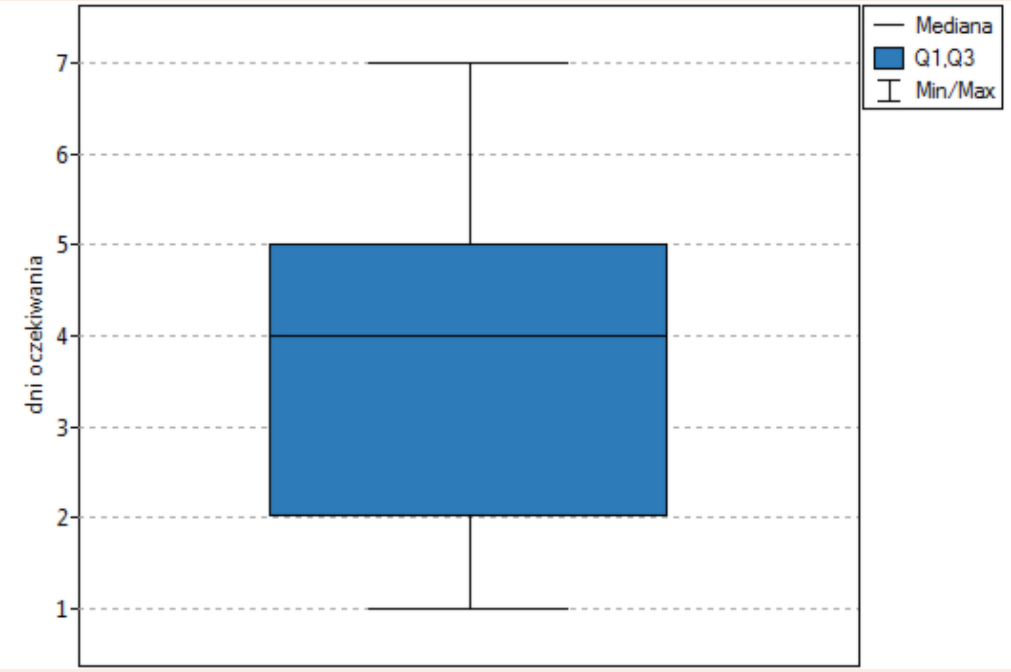
WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 15.1 (plik kurier.pqs): tym razem zamiast średniej sprawdzimy medianę liczby dni oczekiwania na dostarczenie przesyłki.

Test Wilcoxona (rangowanych znaków)	
Czas analizy	0.21 sek.
Analizowane zmienne	dni oczekiwania
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Korekta na ciągłość	Tak
Liczność - grupy	22
Liczność pominiętych wartości (równych medianie)	3
Mediana hipotetyczna	3
Statystyka T	56
Wartość p dwustronne (dokładne)	0.1232
Statystyka Z (z poprawką dla rang wiązanych)	1.5726
Wartość p dwustronne (asymptotyczne)	0.1158

Podsumowanie	
Grupa	dni oczekiwa
Liczność	22
Mediana	4
Dolny kwartyl	2
Górny kwartyl	5

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p = 0.123212$ **dla statystyki** T **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Nie ma podstaw, by odrzucić hipotezę zerową mówiącą, że mediana liczby dni oczekiwania na dostarczenie przesyłki przez analizowaną firmę kurierską wynosi 3.

Taką samą decyzję podjęlibyśmy również na podstawie wartości $p = 0.111161$ lub $p = 0.115817$ testu Wilcoxona opartego o statystykę Z lub Z z poprawką na ciągłość.

DLA ZAINTERESOWANYCH

Hipotezy:
Hipotezy dotyczą równości sumy rang dodatnich i ujemnych lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0 &: \quad \theta = \theta_0, \\ \mathcal{H}_1 &: \quad \theta \neq \theta_0. \end{aligned}$$

gdzie:
θ – mediana badanej cechy w populacji reprezentowanej przez badaną próbę,
θ_0 – zadana wartość.

Wyznaczamy wartość statystyki testowej Z (T – dla małej liczności próby), a na jej podstawie wartość p, którą porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Dla małej liczności próby – wzór:

$$T = \min\left(\sum R_-, \sum R_+\right),$$

gdzie $\sum R_+$ i $\sum R_-$ to odpowiednio suma rang dodatnich i suma rang ujemnych. Statystyka ta podlega rozkładowi Wilcoxona.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dla próby o dużej liczności – wzór:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}},$$

gdzie:
n – liczba rangowanych znaków (liczba rang),
t – liczba przypadków wchodzących w skład rangi wiązanej.
Wzór na statystykę testową Z zawiera poprawkę na rangi wiązane. Poprawka ta powinna być stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych, poprawka ta nie jest wyliczana, gdyż wówczas $\left(\sum t^3 - \sum t\right)/48 = 0$). Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Poprawka na ciągłość testu Wilcoxona (Marascuilo i McSweeney, 1977[117])
Poprawkę na ciągłość stosujemy, by zapewnić możliwość przyjmowania przez statystykę testową wszystkich wartości liczb rzeczywistych, zgodnie z założeniem rozkładu normalnego. Wzór na statystykę testową z poprawką na ciągłość wyraża się wtedy wzorem:

$$Z = \frac{\left|T - \frac{n(n+1)}{4}\right| - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}.$$

d-Cohena – wzór:

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Test chi-kwadrat zgodności

KIEDY STOSOWAĆ TEN TEST:

Test χ^2 zgodności (dobroci dopasowania) *(ang. Chi-square goodness-of-fit test)*, nazywany również testem χ^2 dla pojedynczej próby, przeznaczony jest do testowania zgodności wartości obserwowanych dla r ($r \geq 2$) kategorii $X_1, X_2, ..., X_r$ jednej cechy X z hipotetycznymi wartościami oczekiwanymi dla tej cechy. Wartości wszystkich n pomiarów należy zebrać w postaci tabeli składającej się z r wierszy (kategorii: $X_1, X_2, ..., X_r$). Dla każdej kategorii X_i zapisuje się częstość jej występowania O_i oraz częstość dla niej oczekiwaną E_i lub prawdopodobieństwo jej wystąpienia p_i. Częstość oczekiwana jest wyznaczana jako iloczyn $E_i = np_i$.

WARUNKI STOSOWANIA

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- duże liczności oczekiwane według interpretacji Cochrana (1952)[41],
- suma liczności obserwowanych powinna być taka sama jak suma liczności oczekiwanych, a suma wszystkich prawdopodobieństw p_i powinna wynosić 1.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

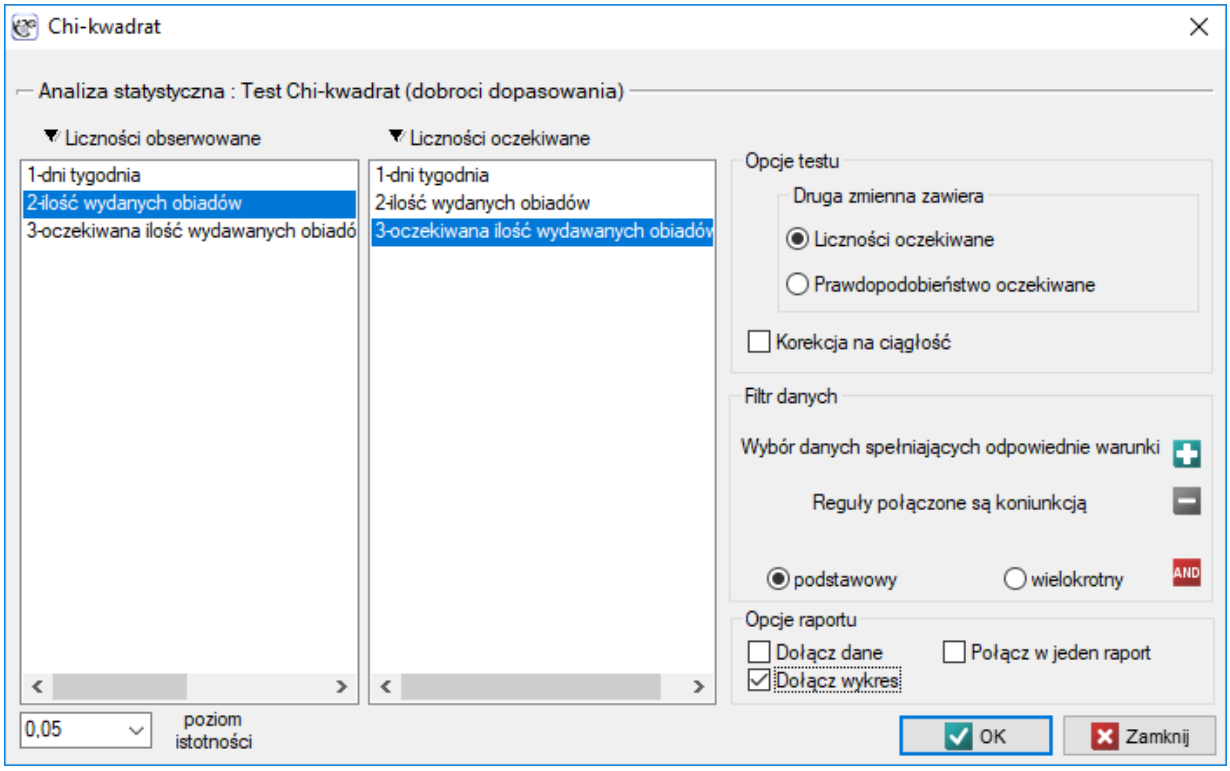
Uwaga! Utworzona tabela danych może przyjąć jedną z poniższych postaci:

Kategorie X_i	O_i	E_i
X_1	O_1	E_i
X_2	O_2	E_2
...	...	...
X_r	O_r	E_r

Kategorie X_i	O_i	p_i
X_1	O_1	p_1
X_2	O_2	p_2
...	...	...
X_r	O_r	p_r

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Chi-kwadrat (lub poprzez Kreator).



WYKONANIE I INTERPRETACJA

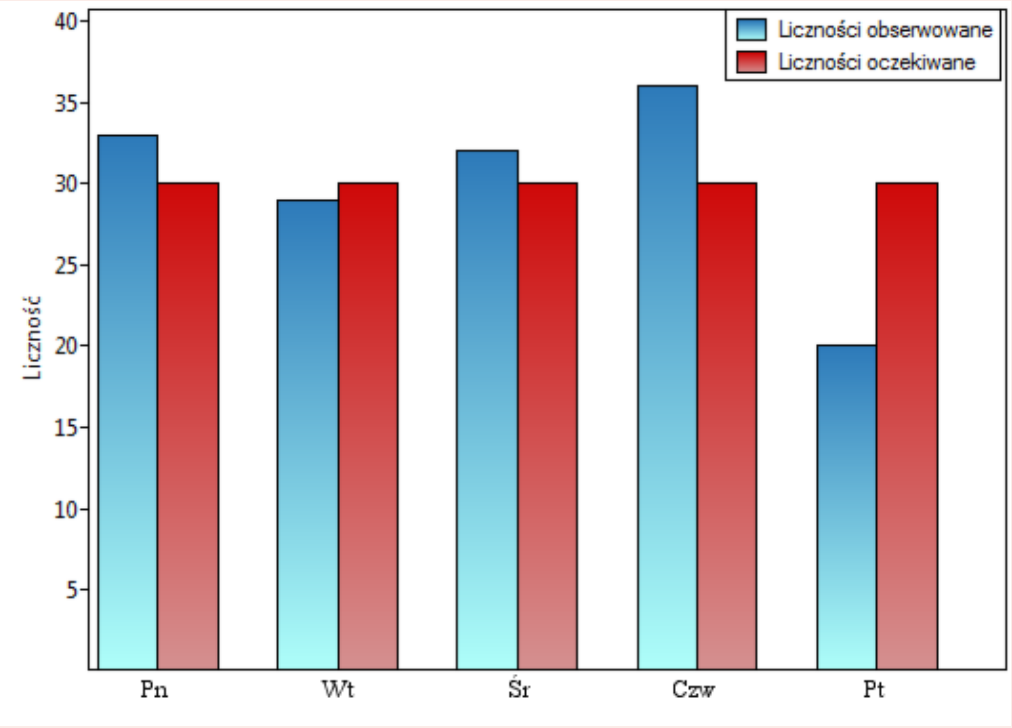
Przykład 15.3. (plik obiady.pqs)
Chcielibyśmy się dowiedzieć, czy liczba wydawanych obiadów w kolejnych dniach tygodnia (od poniedziałku do piątku) w pewnej szkolnej stołówce jest statystycznie taka sama. W tym celu pobrano tygodniową próbę i zapisano dla niej liczbę wydanych obiadów w poszczególnych dniach: poniedziałek – 33, wtorek – 29, środa – 32, czwartek – 36, piątek – 20.

Łącznie przez cały tydzień (5 dni) wydano 150 obiadów. Zakładamy, że w każdy dzień prawdopodobieństwo wydania obiadu jest takie samo, czyli wynosi $\frac{1}{5}$. Oczekiwana liczba wydanych obiadów dla każdego z pięciu dni tygodnia wynosi więc $E_i = 150 \cdot \frac{1}{5} = 30$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

dni tygodnia	ilość wydanych obiadów	oczekiwana ilość wydawanych obiadów
poniedziałek	33	30
wtorek	29	30
środa	32	30
czwartek	36	30
piątek	20	30

Test Chi-kwadrat (dobroci dopasowania)	
Czas analizy	0.03 sek.
Analizowane zmienne	ilość wydanych obiadów,oczekiwana ilość wydawany
Poziom istotności	0.05
Liczność	150
Statystyka Chi-kwadrat	5
Ilość stopni swobody	4
Wartość p	0.287297



$p = 0.287297$ **dla** $df = 4$ **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Nie ma podstaw, aby odrzucić hipotezę zerową mówiącą o zgodności liczby wydawanych obiadów z oczekiwaną liczbą wydawanych obiadów w poszczególnych dniach.

Uwaga! Gdybyśmy chcieli w ramach jednego badania dokonać większej liczby porównań, moglibyśmy zastosować **poprawkę Bonferroniego**[2] lub inną z poprawek opisanych w dziale Wielokrotne porównania. Poprawka ta jest używana, by ograniczyć wielkość popełnianego błędu pierwszego rodzaju, gdy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

porównujemy wartości obserwowane i oczekiwane pomiędzy wybranymi dniami, np.: Pt $\Longleftrightarrow$ Pn, Pt $\Longleftrightarrow$ Wt, Pt $\Longleftrightarrow$ Śr, Pt $\Longleftrightarrow$ Czw – przy założeniu, że porównania wykonujemy niezależnie. Poziom istotności α dla każdego porównania wyznaczamy zgodnie z tą poprawką według wzoru $\alpha = \frac{0.05}{r}$, gdzie r to liczba wykonywanych porównań – dla naszego przykładu wynosi więc $\alpha = \frac{0.05}{4} = 0.0125$. Należy jednak pamiętać, że redukując α dla każdego porównania, zmniejszamy również moc testu.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad O_i = E_i \text{ dla wszystkich kategorii,} \\ \mathcal{H}_1: &\quad O_i \neq E_i \text{ dla przynajmniej jednej kategorii.} \end{aligned}$$

Wzory:

Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z liczbą stopni swobody wyznaczaną według wzoru $df = (r-1)$.

Wyznaczoną na podstawie wartości statystyki i rozkładu χ^2 wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Testy dla jednej proporcji

KIEDY STOSOWAĆ TEN TEST:

Testy dla jednej proporcji stosujemy, gdy mamy do uzyskania dwa możliwe wyniki (jeden z nich to wynik wyróżniony o liczności m) i wiemy, jak często te wyniki pojawiają się w próbie (znamy proporcję p). Testy te służą do weryfikacji hipotezy, że proporcja w populacji, z której pochodzi próba, to zadana wartość. W zależności od wielkości próby n mamy do wyboru **test Z dla jednej proporcji** – dla dużych prób, oraz **test dokładny dwumianowy** – dla prób o małej liczności.

WARUNKI STOSOWANIA

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę.

Uwaga! Dodatkowy warunek dla testu Z dla jednej proporcji: duża liczność (według interpretacji Marascuilo i McSweeney (1977)[117] każda z wartości $np > 5$ i $n(1-p) > 5$). Gdy warunek ten nie jest spełniony, zaleca się skorzystać z testu dokładnego dwumianowego.

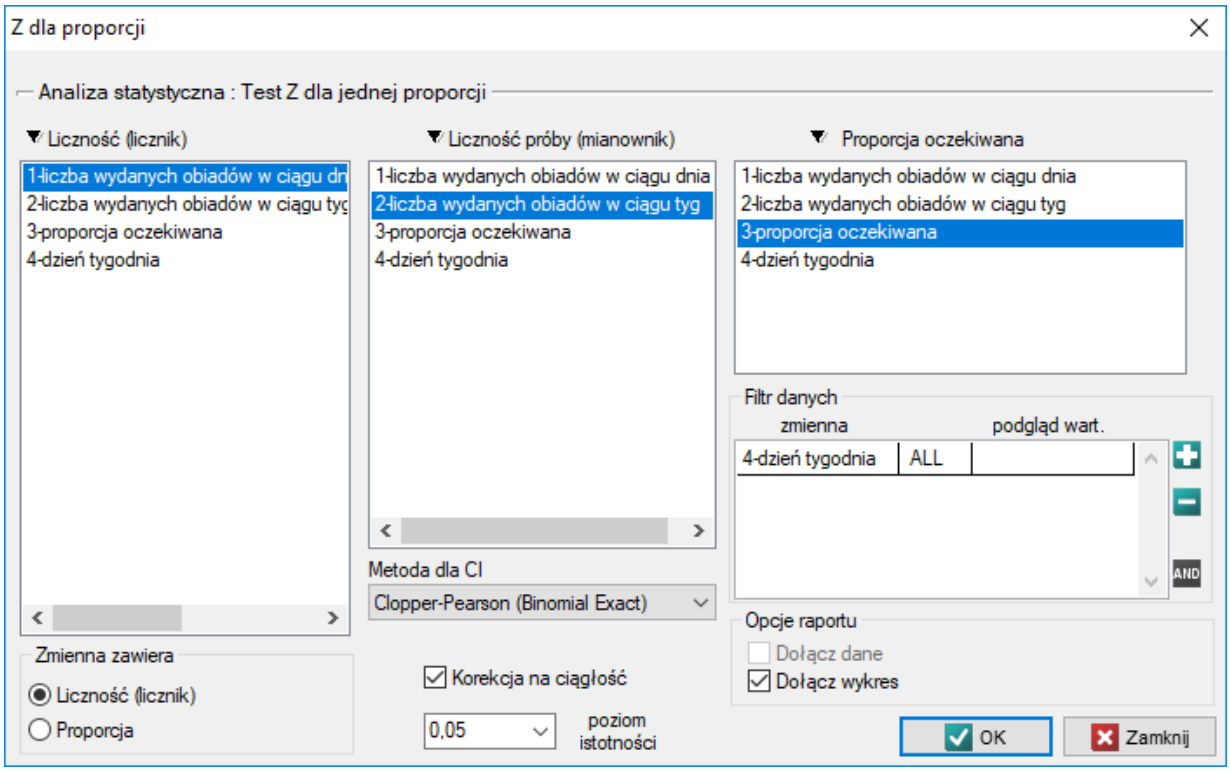
DEFINICJE MIAR

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Przedział ufności dla proporcji – dla estymatora z próby, jakim jest w tym przypadku wartość proporcji p, wyznacza się przedział ufności. Dla prób o dużej liczności można bazować na przedziałach opartych o rozkład normalny – tzw. przedziały Walda. Bardziej uniwersalne są natomiast przedziały zaproponowane przez Wilsona (1927)[178], a także Agresti i Coull (1998)[5]. Przedziały Cloppera i Pearsona (1934)[39] są dokładniejsze dla prób o mniejszej liczności. Porównanie metod budowania przedziałów dla proporcji można znaleźć w pracy Browna L.D. i innych (2001)[30].

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Z dla proporcji.



WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 15.3 (plik obiady.pqs): załóżmy, że chcielibyśmy sprawdzić, czy w piątek wydawana jest $\frac{1}{5}$ spośród wszystkich obiadów wydawanych w szkolnej stołówce w ciągu tygodnia. Dla pobranej próby $m = 20$, $n = 150$.

liczba wydanych obiadów w ciągu dnia	liczba wydanych obiadów w ciągu tyg	proporcja oczekiwana	dzień tygodnia
33	150	0,2	poniedziałek
29	150	0,2	wtorek
32	150	0,2	środa
36	150	0,2	czwartek
20	150	0,2	piątek

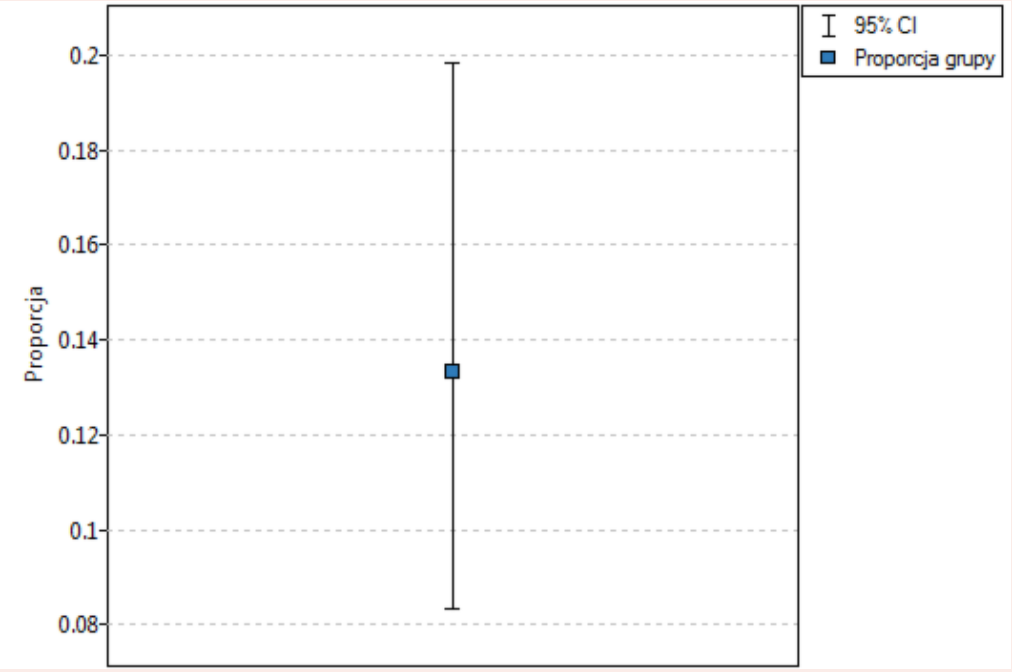
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przy ustawianiu opcji analizy włączamy filtr, wybierając odpowiedni dzień tygodnia – czyli piątek. Brak ustawienia filtru nie generuje błędu, a jedynie skutkuje wyliczeniem kolejnych statystyk dla kolejnych dni tygodnia.

Test Z dla jednej proporcji	
Czas analizy	0,04 sek.
Analizowane zmienne	liczba wydanych obiadów
Poziom istotności	0,05
Korekcja na ciągłość	Nie
Filtr danych	dzień tygodnia=piątek
1	
Proporcja grupy	0,133333
Clopper-Pearson (Binomial Exact)	
-95% CI dla proporcji	0,083384
+95% CI dla proporcji	0,198387
Statystyka Z	-2,041241
Wartość p (asymptotyczne)	0,041227
Wartość p jednostronne (dokładne)	0,022355
Wartość p dwustronne (dokładne)	0,044711

Dane :

v.1	v.2	v.3
20	150	0,2



$p = 0.041227$ dla testu Z bez poprawki na ciągłość, $p = 0.044711$ dla testu dokładnego dwumianowego (istotne statystycznie, $\alpha = 0.05$); $p = 0.052479$ dla testu Z z poprawką na ciągłość (nieistotne statystycznie)

Proporcja obiadów wydawanych w piątek jest statystycznie mniejsza niż $\frac{1}{5}$ obiadów wydawanych przez cały tydzień – z wyjątkiem wyniku testu Z z poprawką na ciągłość.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Proporcja wartości wyróżnionych w próbie to $p = \frac{m}{n} = 0.133$, a 95% przedział ufności Cloppera-Pearsona dla tej frakcji $(0.083, 0.198)$ nie zawiera hipotetycznej wartości 0.2.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad p = p_0, \\ \mathcal{H}_1: &\quad p \neq p_0, \end{aligned}$$

gdzie:
p – prawdopodobieństwo (wyróżniona proporcja) w populacji,
p_0 – prawdopodobieństwo oczekiwane (proporcja oczekiwana).

Test Z dla jednej proporcji *(ang. Z test for one proportion)*
Wzory: Statystyka testowa ma postać:

$$Z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},$$

gdzie:
$p = \frac{m}{n}$ – proporcja dla próby z tej populacji,
m – liczność wartości wyszczególnionych w próbie,
n – liczność próby.
Zmodyfikowana o poprawkę na ciągłość statystyka testowa ma postać:

$$Z = \frac{|p - p_0| - \frac{1}{2n}}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Statystyka Z, zarówno bez korekcji na ciągłość, jak i z tą korekcją, ma asymptotycznie (dla dużych liczności) rozkład normalny.

Test dwumianowy *(ang. Binomial test for one proportion)*
Definicja: Test wykorzystuje w sposób bezpośredni **rozkład dwumianowy**, zwany również rozkładem Bernoulliego, który należy do grupy rozkładów dyskretnych (czyli takich, w których badana zmienna przyjmuje skończoną liczbę wartości). Analizowana zmienna może przyjmować $k = 2$ wartości, pierwszą oznaczaną zwykle mianem sukcesu, a drugą – porażki. Prawdopodobieństwo wystąpienia sukcesu to p_0, a porażki $1 - p_0$.

Wzory: Prawdopodobieństwo dla konkretnego punktu w tym rozkładzie wyliczane jest ze wzoru:

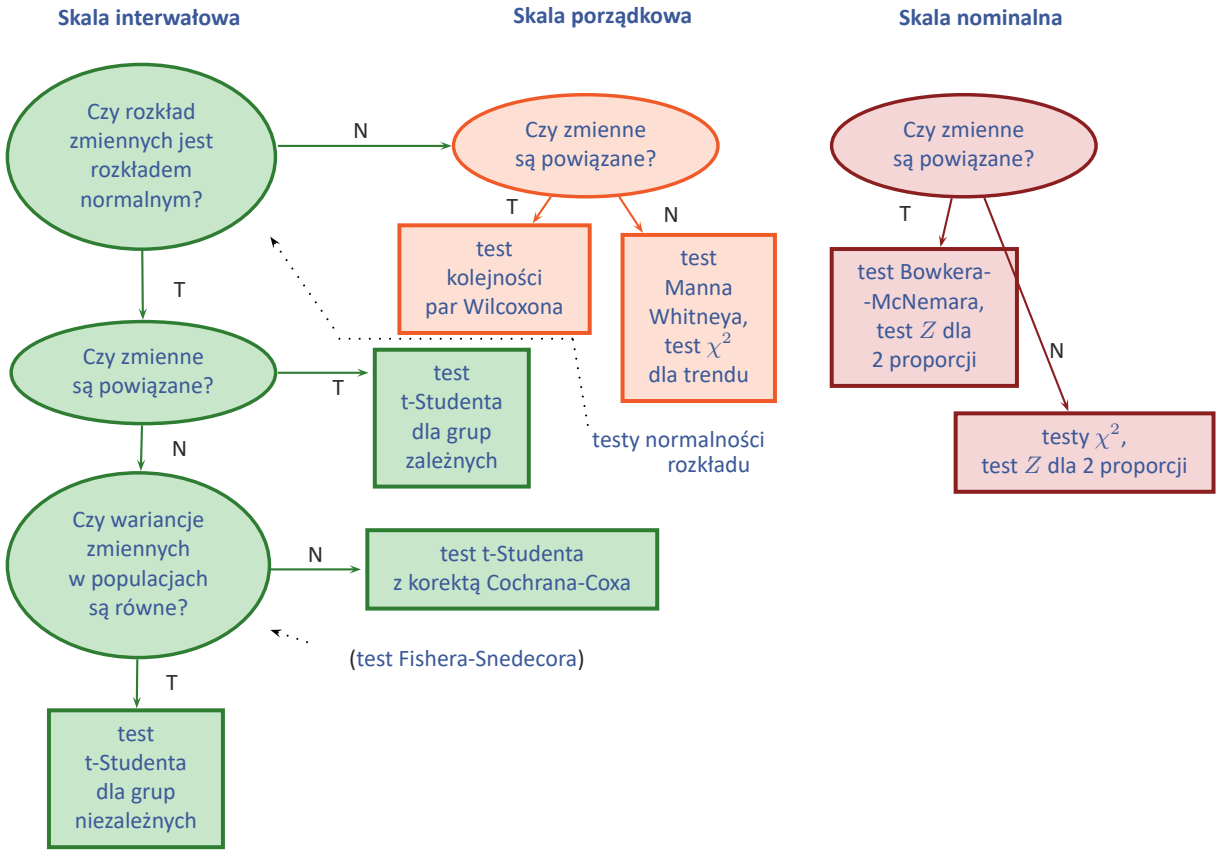
$$P(m) = \binom{n}{m} p_0^m (1 - p_0)^{n-m},$$

gdzie:
$\binom{n}{m} = \frac{n!}{m!(n-m)!}$,
m – liczność wartości wyszczególnionych w próbie,
n – liczność próby.

Interpretacja: Na podstawie sumy odpowiednich prawdopodobieństw P wyznacza się wartość p jednostronną i dwustronną, przy czym dwustronna wartość p jest definiowana jako podwojona wartość mniejszego z jednostronnych prawdopodobieństw. Wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

16 PORÓWNANIE - 2 GRUPY



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

16.1 TESTY PARAMETRYCZNE

Test Fishera-Snedecora

KIEDY STOSOWAĆ TEN TEST:

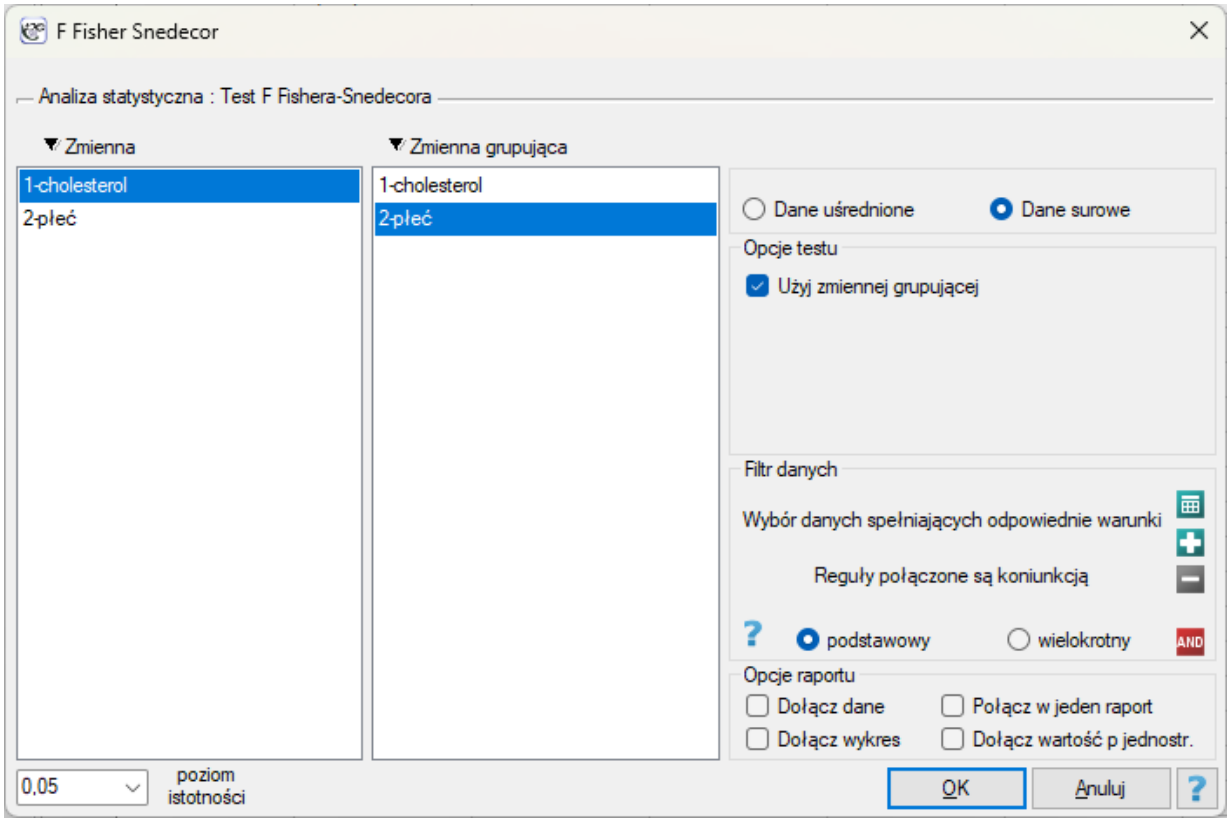
Test Fishera-Snedecora *(ang. F-Snedecor test)* sprawdza, czy dwie porównywane grupy mają podobnie zróżnicowane wyniki (podobną zmienność). Stosuje się go zwykle jako krok wstępny przed testem t-Studenta dla grup niezależnych – jego wynik podpowiada, czy użyć podstawowej wersji testu t, czy jego wariantu z poprawką Cochrana-Coxa.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanej zmiennej w obu populacjach,
- model niezależny.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → F Fisher Snedecor.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krok 1. Wskaż zmienną grupującą (np. płeć) i zmienną wynikową (np. stężenie cholesterolu całkowitego).

Krok 2. Uruchom analizę.

Uwaga! Obliczenia mogą bazować na danych w postaci surowych rekordów lub danych uśrednionych, tzn. odchyleniach standardowych i licznościach prób.

WYKONANIE I INTERPRETACJA

Przykład 16.1. (plik cholesterol.pqs)
Z populacji kobiet i z populacji mężczyzn mieszkających w Polsce, w wieku powyżej 40 roku życia, wylosowano po 500 osób. Badanie dotyczyło oceny ryzyka chorób sercowo-naczyniowych. Wśród badanych parametrów znajduje się stężenie cholesterolu całkowitego. Przed porównaniem średnich wartości tego parametru między kobietami a mężczyznami sprawdzamy najpierw, czy wariancje stężenia cholesterolu całkowitego w obu populacjach są sobie równe.

Rozkład stężenia cholesterolu całkowitego w obu grupach jest rozkładem normalnym (zostało to sprawdzone testem Lillieforsa). Odchylenie standardowe w grupie mężczyzn wyniosło $sd_1 = 47.6$ ($n_1 = 500$), a w grupie kobiet $sd_2 = 43.5$ ($n_2 = 500$).

Test F Fishera-Snedecora	
Analizowane zmienne	cholesterol
	płeć
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0,05
Zmienna grupująca	płeć
Nazwa grupy	kobieta
Liczność / grupy	500
Średnia grupy	201,078
Odchylenie standardowe grupy	47,638711
Nazwa grupy	mężczyzna
Liczność / grupy	500
Średnia grupy	191,456
Odchylenie standardowe grupy	43,514631
Statystyka F	1,198531
Stopnie swobody	499 / 499
Wartość p dwustronne	0,043351

$p = 0.0434$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Wariancje stężenia cholesterolu całkowitego u kobiet i mężczyzn mieszkających w Polsce różnią się istotnie statystycznie, choć różnica ta jest niewielka. W dalszym porównaniu średnich należy więc zastosować test t-Studenta z poprawką Cochrana-Coxa, a nie jego wersję podstawową.

Wartość p wyznaczoną na podstawie statystyki testowej porównujemy zawsze z przyjętym poziomem

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

istotności α: gdy $p \leq \alpha$, różnica wariancji jest istotna statystycznie; gdy $p > \alpha$, nie ma podstaw, by twierdzić, że wariancje się różnią. To rozstrzygnięcie decyduje, którą wersję testu t-Studenta dla grup niezależnych zastosować w kolejnym kroku analizy.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \sigma_1^2 = \sigma_2^2, \\ \mathcal{H}_1: &\quad \sigma_1^2 \neq \sigma_2^2, \end{aligned}$$

gdzie σ_1^2, σ_2^2 – wariancje badanej zmiennej w pierwszej i drugiej populacji.

Wzory: Statystyka testowa ma postać:

$$F = \frac{sd_1^2}{sd_2^2},$$

gdzie sd_1^2, sd_2^2 – wariancje badanej zmiennej w próbach wybranych losowo z pierwszej i drugiej populacji. Statystyka ta podlega rozkładowi F Snedecora z $n_1 - 1$ i $n_2 - 1$ stopniami swobody.

Test t-Studenta dla grup niezależnych

KIEDY STOSOWAĆ TEN TEST:

Test t-Studenta dla grup niezależnych *(ang. t test for independent groups)* służy do weryfikacji hipotezy o równości średnich badanej zmiennej w dwóch populacjach – np. gdy porównujemy wartość jakiejś cechy między dwiema niezależnymi, niepowiązanymi ze sobą grupami (kobiety i mężczyźni, grupa leczona i kontrolna).

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanej zmiennej w obu populacjach,
- model niezależny,

Nie musisz samodzielnie sprawdzać równości wariancji w obu grupach – PQStat robi to automatycznie i dobiera odpowiednią wersję testu (zob. *Jak rozpocząć analizę*).

Test t-Studenta z korektą Cochrana-Coxa

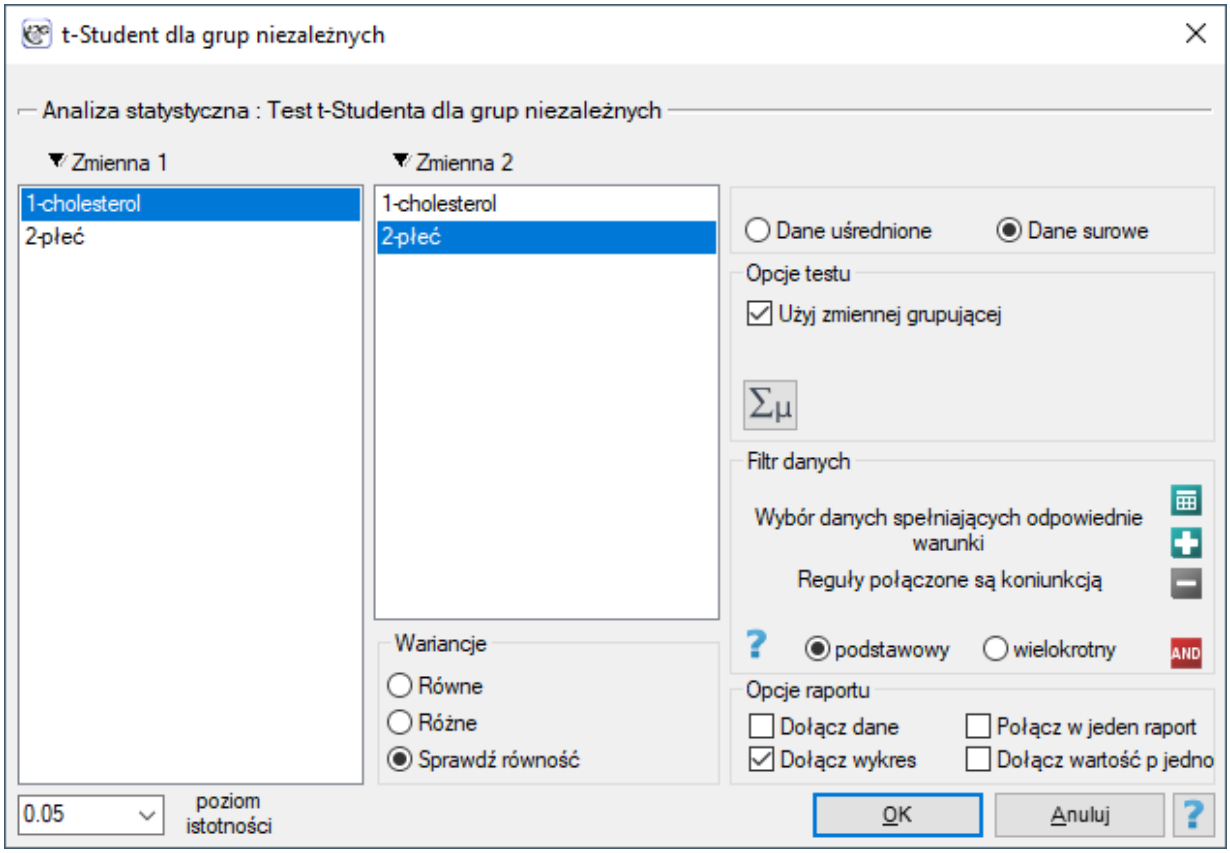
Poprawka Cochrana-Coxa *(ang. Cochran-Cox adjustment)*, (1957)[43] to wariant testu t-Studenta dla grup niezależnych stosowany automatycznie wtedy, gdy wariancje badanej zmiennej w obu populacjach

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

różnią się istotnie. Warunki stosowania i cel testu są takie same jak dla wariantu podstawowego – różni się jedynie sposób wyliczenia statystyki testowej i stopni swobody (zob. *Dla zainteresowanych*).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → t-Student dla grup niezależnych (lub poprzez Kreator).



Krok 1. Wskaż zmienną grupującą (np. płeć) i zmienną wynikową (np. cholesterol całkowity).

Krok 2. W opcji dotyczącej wariancji wybierz jedno z ustawień:

- równe – zostanie wyliczony test t-Studenta dla grup niezależnych,
- różne – zostanie wyliczony test t-Studenta z korektą Cochrana-Coxa,
- sprawdź równość – program sam wyliczy test Fishera-Snedecora i na tej podstawie automatycznie wybierze właściwy wariant testu.

Krok 3. Uruchom analizę.

Uwaga! Obliczenia mogą bazować na danych w postaci surowych rekordów lub danych uśrednionych tzn. średnich arytmetycznych, odchyleniach standardowych i liczności prób.

WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

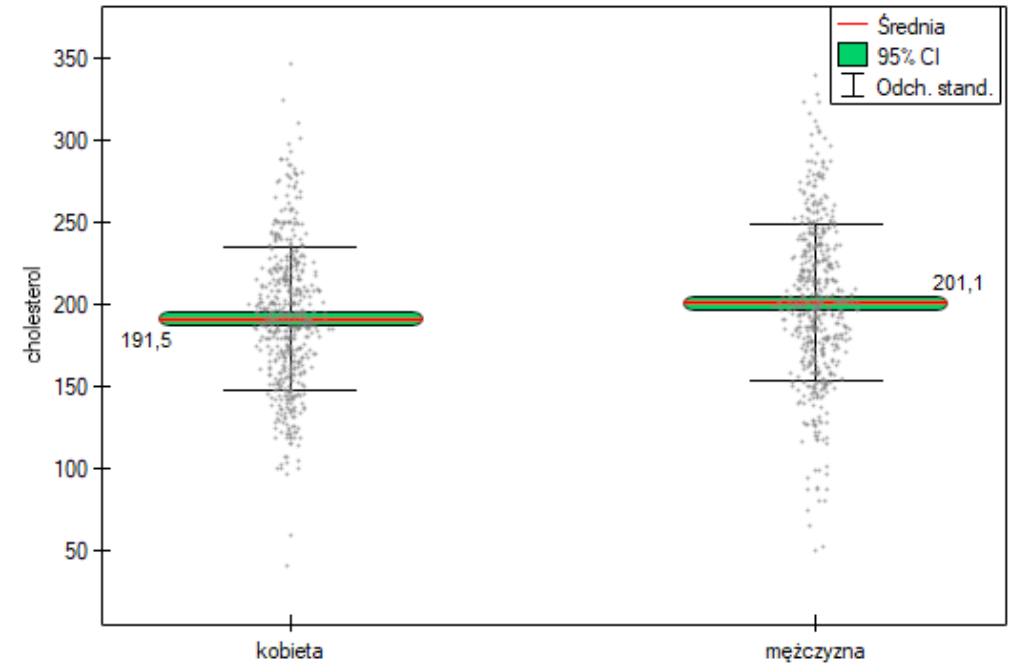
PRZYKŁAD 16.2. (plik cholesterol.pqs)
Z populacji kobiet i z populacji mężczyzn mieszkających w Polsce, w wieku powyżej 40 roku życia, wylosowano po 500 osób. Badanie dotyczyło oceny ryzyka chorób sercowo-naczyniowych. Wśród badanych parametrów znajduje się stężenie cholesterolu całkowitego. Celem tego badania będzie porównanie mężczyzn i kobiet co do tej wartości – chcemy wykazać, że populacje te różnią się już na poziomie stężenia cholesterolu całkowitego, a nie tylko w zakresie poszczególnych frakcji cholesterolu.

Rozkład wieku w obu grupach jest rozkładem normalnym (zostało to sprawdzone testem Lillieforsa). Średnie stężenie cholesterolu całkowitego w grupie mężczyzn wyniosło $\overline{x}_1 = 201.1$, a odchylenie standardowe $sd_1 = 47.6$, w grupie kobiet odpowiednio $\overline{x}_2 = 191.5$ i $sd_2 = 43.5$. Test Fishera-Snedecora wskazuje na niewielkie, lecz istotne statystycznie ($p = 0.0434$) różnice w wariancjach. W analizie zastosowany zostanie test t-Studenta z poprawką Cochrana-Coxa.

Test t-Studenta dla grup niezależnych	
Czas analizy	0.16 sek.
Analizowane zmienne	cholesterol;płeć
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
poprawka dla różnych wariancji	Tak
Zmienna grupująca	płeć
Różnica średnich	9.622
-95% CI dla różnicy średnich	3.9597
+95% CI dla różnicy średnich	15.2843
Błąd standardowy różnicy	2.8855
Wspólne odchylenie standardowe	NA
Statystyka t	3.3346
Stopnie swobody	989.9277
Wartość p dwustronne (Cochran-Cox)	0.0009
Test F Fishera-Snedecora	
iloraz wariancji F	1.1985
Wartość p	0.0434

Podsumowanie		
Grupa	kobieta	mężczyzna
Liczność	500	500
Średnia arytmetyczna	201.078	191.456
Błąd standardowy średniej	2.1305	1.946
Odchylenie standardowe	47.6387	43.5146
-95% CI dla średniej grupy	196.8922	187.6326
+95% CI dla średniej grupy	205.2638	195.2794

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p = 0.0009$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Kobiety i mężczyźni mieszkający w Polsce różnią się istotnie statystycznie stężeniem cholesterolu całkowitego. Przeciętny Polak po 40. roku życia ma wyższe stężenie cholesterolu całkowitego niż przeciętna Polka – różnica wynosi prawie 10 jednostek.

Wartość p wyznaczoną na podstawie statystyki testowej porównujemy zawsze z przyjętym poziomem istotności α: gdy $p \leq \alpha$, różnica średnich jest istotna statystycznie; gdy $p > \alpha$, nie ma podstaw, by twierdzić, że grupy się różnią. Warto dodatkowo zgłosić **wielkość efektu** (d-Cohena) – sama istotność statystyczna nie mówi, jak duża jest różnica; przy dużych próbach nawet niewielkie różnice mogą wyjść istotne statystycznie.

DEFINICJE MIAR

d-Cohena – standaryzowana miara wielkości efektu, która określa, jak dużą częścią występującej w danych zmienności jest zaobserwowana różnica między średnimi. W przeciwieństwie do wartości p, nie zależy od liczności próby, dzięki czemu pozwala ocenić praktyczne znaczenie różnicy, a nie tylko jej istotność statystyczną.

DLA ZAINTERESOWANYCH

Test t-Studenta dla grup niezależnych

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2, \\ \mathcal{H}_1: &\quad \mu_1 \neq \mu_2. \end{aligned}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie μ_1, μ_2 – średnie badanej zmiennej w pierwszej i drugiej populacji.

Wzory:

Statystyka testowa ma postać:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{(n_1-1)sd_1^2 + (n_2-1)sd_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

gdzie $\overline{x}_1, \overline{x}_2$ – średnie w próbach, n_1, n_2 – liczności prób, sd_1^2, sd_2^2 – wariancje w próbach. Statystyka ma rozkład t-Studenta z $df = n_1 + n_2 - 2$ stopniami swobody.

Wspólne odchylenie standardowe:

$$SD_p = \sqrt{\frac{(n_1-1)sd_1^2 + (n_2-1)sd_2^2}{n_1+n_2-2}},$$

Błąd standardowy różnicy średnich:

$$SE_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{(n_1-1)sd_1^2 + (n_2-1)sd_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

Wzór d-Cohena:

d-Cohena wyliczamy ze wzoru:

$$d = \left|\frac{\overline{x}_1 - \overline{x}_2}{SD_p}\right|$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [46] wskazówkami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Test t-Studenta z korektą Cochrana-Coxa

Wzory:

Statystyka testowa ma postać:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}}.$$

Statystyka ma rozkład t-Studenta z liczbą stopni swobody zaproponowaną przez Satterthwaite (1946)[148] i wyliczaną z wzoru:

$$df = \frac{\left(\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}\right)^2}{\left(\frac{sd_1^2}{n_1}\right)^2 \cdot \frac{1}{(n_1-1)} + \left(\frac{sd_2^2}{n_2}\right)^2 \cdot \frac{1}{(n_2-1)}}.$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test t-Studenta dla grup zależnych

KIEDY STOSOWAĆ TEN TEST:

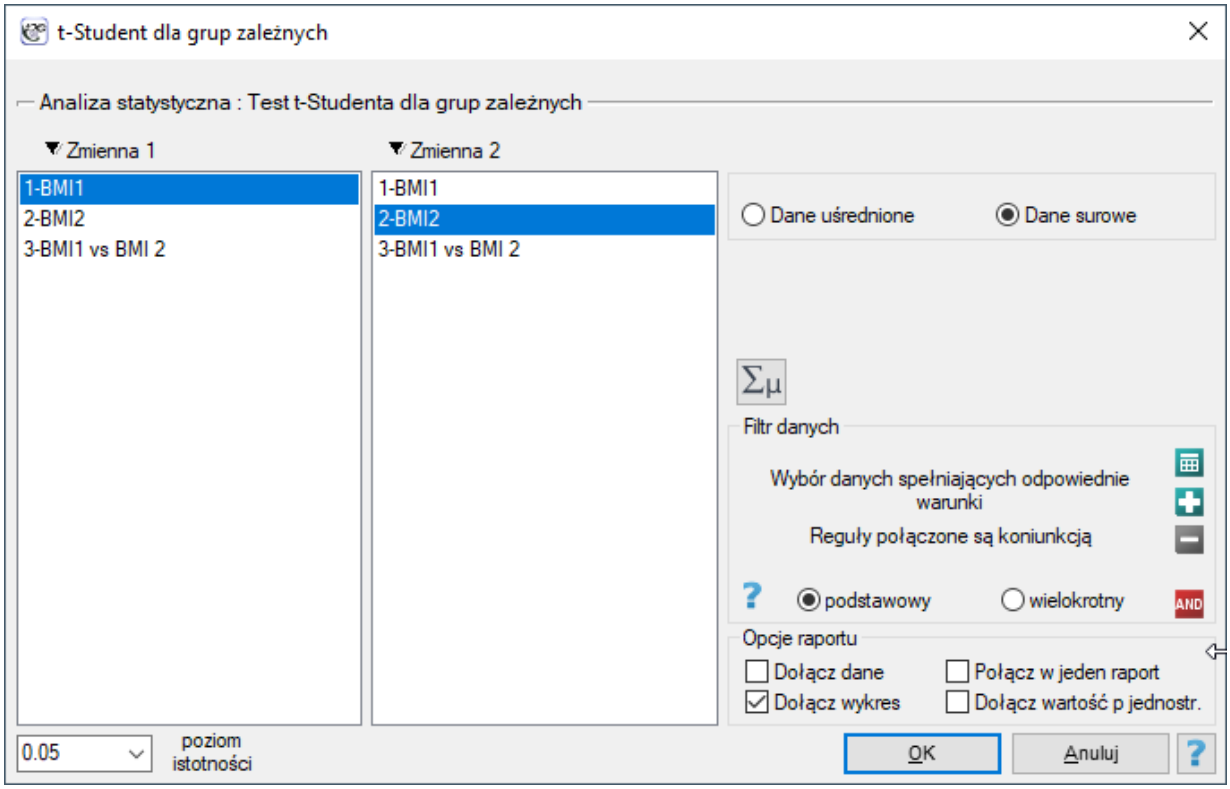
Test t-Studenta dla grup zależnych *(ang. t test for dependent groups)* stosujemy, gdy tę samą cechę mierzymy dwukrotnie u tych samych osób, w różnych warunkach (np. przed i po zastosowaniu diety), i chcemy sprawdzić, czy wynik pomiaru się zmienił.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu różnicy pomiarów d_i,
- model zależny.

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Testy parametryczne $\rightarrow$ t-Student dla grup zależnych (lub poprzez Kreator).



Krok 1. Wskaż dwie zmienne odpowiadające pierwszemu i drugiemu pomiarowi (np. BMI przed dietą i BMI po diecie).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krok 2. Uruchom analizę.

Uwaga! Obliczenia mogą bazować na danych w postaci surowych rekordów lub danych uśrednionych, tzn. średniej różnic, odchyleniu standardowym różnic i liczności próby.

WYKONANIE I INTERPRETACJA

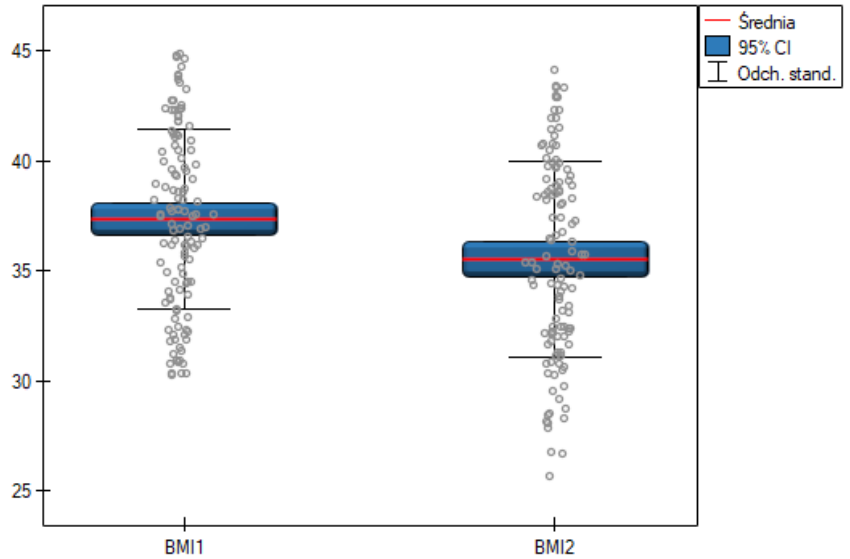
PRZYKŁAD 16.3. (plik BMI.pqs)
W klinice leczącej zaburzenia odżywiania badano wpływ zalecanej „diety A" na zmianę masy ciała. Próbę 120 otyłych chorych poddano diecie. Zbadano dla nich poziom BMI dwukrotnie: przed wprowadzeniem diety i po 180 dniach jej stosowania. Celem badania było porównanie uzyskanych wyników pomiarów BMI, aby ocenić skuteczność diety.

Rozkład różnicy pomiędzy parami pomiarów był rozkładem normalnym (test Lillieforsa, $p = 0.0837$), co pozwoliło zastosować test t-Studenta dla grup zależnych.

Test t-Studenta dla grup zależnych	
Czas analizy	0.13 sek.
Analizowane zmienne	BMI1;BMI2
Liczność - niezinterpretowane	0
Liczność - braki danych (parami)	0
Poziom istotności	0.05
Liczność - ilość par	120
Średnia różnic	1.8219
-95% CI dla średniej różnic	1.4136
+95% CI dla średniej różnic	2.2303
Błąd standardowy średniej różnic	0.2062
Odchylenie standardowe różnicy	2.2592
Statystyka t	8.8342
Stopnie swobody	119
Wartość p dwustronne	<0.0001

Podsumowanie		
Grupa	BMI1	BMI2
Liczność	120	120
Średnia arytmetyczna	37.3532	35.5313
Błąd standardowy średniej	0.3703	0.4083
Odchylenie standardowe	4.0567	4.4727
-95% CI dla średniej grupy	36.6199	34.7228
+95% CI dla średniej grupy	38.0864	36.3397

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p < 0.0001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Średni poziom BMI badanych chorych zmienił się istotnie statystycznie pod wpływem diety – przed jej zastosowaniem był wyższy średnio o niecałe 2 jednostki.

Wartość p wyznaczoną na podstawie statystyki testowej porównujemy zawsze z przyjętym poziomem istotności α: gdy $p \leq \alpha$, różnica jest istotna statystycznie; gdy $p > \alpha$, nie ma podstaw, by twierdzić, że wystąpiła zmiana. Warto dodatkowo zgłosić **wielkość efektu** (d-Cohena) – sama istotność statystyczna nie mówi, jak duża jest zmiana; przy dużych próbach nawet niewielkie różnice mogą wyjść istotne statystycznie.

DEFINICJE MIAR

d-Cohena – standaryzowana miara wielkości efektu, która określa, jak dużą częścią występującej w danych zmienności jest zaobserwowana różnica między średnimi. W przeciwieństwie do wartości p, nie zależy od liczności próby, dzięki czemu pozwala ocenić praktyczne znaczenie różnicy, a nie tylko jej istotność statystyczną.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_0 = 0, \\ \mathcal{H}_1: &\quad \mu_0 \neq 0, \end{aligned}$$

gdzie μ_0 – średnia różnic d_i w populacji.

Wzory:

Statystyka testowa ma postać:

$$t = \frac{\overline{d}}{sd_d}\sqrt{n},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie $\overline{d}$ – średnia różnic d_i w próbie, sd_d – odchylenie standardowe różnic d_i w próbie, n – liczność różnic d_i w próbie. Statystyka testowa ma rozkład t-Studenta z $n-1$ stopniami swobody.

Odchylenie standardowe różnicy:

$$sd_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}},$$

Błąd standardowy średniej różnic:

$$SEM_d = \frac{SD_d}{\sqrt{n}}.$$

Wzór d-Cohena:

d-Cohena wyliczamy ze wzoru:

$$d = \left|\frac{\overline{d}}{sd_d}\right|$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [46] wskazówkami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

16.2 TESTY NIEPARAMETRYCZNE

Test U Manna-Whitneya

KIEDY STOSOWAĆ TEN TEST:

Test U Manna-Whitneya *(ang. Mann-Whitney U test)* stosujemy, gdy porównujemy dwie niezależne grupy pod względem tej samej cechy, a dane nie spełniają założeń testu t-Studenta dla grup niezależnych (np. nie mają rozkładu normalnego). Sprawdzamy nim, czy wyniki w obu grupach się różnią (zakładając przy tym, że rozkłady zmiennej w obu grupach są sobie zbliżone – porównanie wariancji rang można sprawdzić testem dla rang Conovera).

Test znany jest również jako test Wilcoxona Manna-Whitneya *(ang. Wilcoxon Mann-Whitney test)*, Mann i Whitney (1947)[113] oraz Wilcoxon (1949)[177].

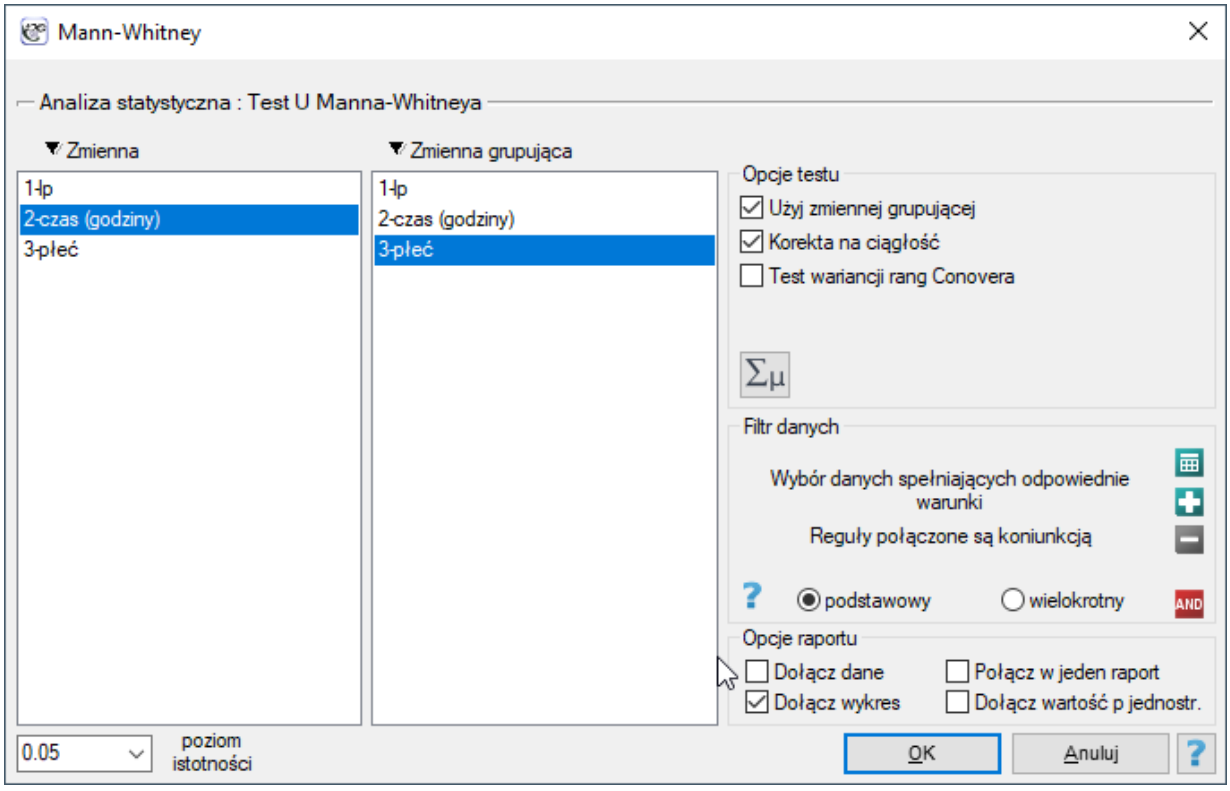
WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej,
- model niezależny.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Mann-Whitney (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienną grupującą (np. płeć) i zmienną wynikową (np. czas spędzany przed komputerem).

Krok 2. Uruchom analizę.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 16.4. (plik komputer.pqs)
Wysunięto hipotezę, że na pewnej uczelni studenci matematyki spędzają statystycznie więcej czasu przed komputerem niż studentki matematyki. W celu weryfikacji tego przypuszczenia z populacji osób studiujących matematykę na tej uczelni wylosowano próbę liczącą 54 osoby (25 kobiet i 29 mężczyzn). Osoby te zapytano o to, jak dużo czasu dziennie spędzają przy komputerze (czas w godzinach) i otrzymano następujące wyniki: (czas, płeć): (2, k) (2, m) (2, m) (3, k) (3, k) (3, k) (3, k) (3, m) (3, m) (4, k) (4, k) (4, k) (4, k) (4, m) (4, m) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, m) (5, m) (5, m) (5, m) (6, k) (6, k) (6, k) (6, k) (6, k) (6, m) (6, m) (6, m) (6, m) (6, m) (6, m) (6, m) (7, k) (7, m) (7, m) (7, m) (7, m) (7, m) (7, m) (7, m) (7, m) (7, m) (8, k) (8, m) (8, m).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test U Manna-Whitneya	
Czas analizy	0.20 sek.
Analizowane zmienne	czas (godziny);płeć
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Korekta na ciągłość	Tak
Zmienna grupująca	płeć
Statystyka U	225.5
Statystyka U'	499.5
Wartość p dwustronne (dokładne)	0.0149
Statystyka Z (z poprawką dla rang wiązanych)	2.413
Wartość p dwustronne (asymptotyczne)	0.0158

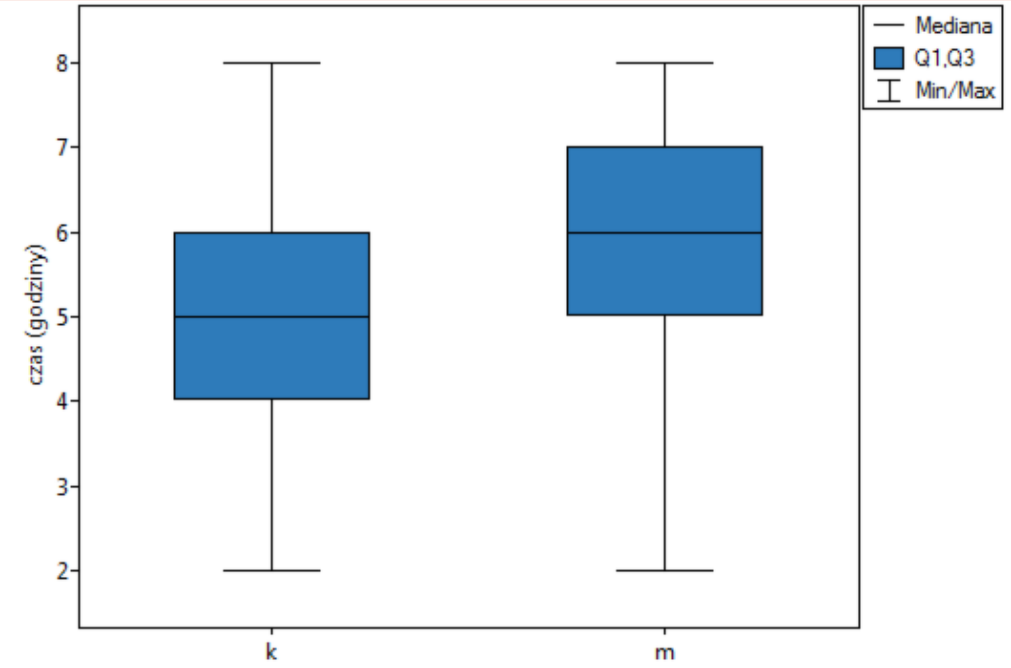
Podsumowanie		
Grupa	k	m
Liczność	25	29
Mediana	5	6
Minimum	2	2
Maksimum	8	8
Dolny kwartyl	4	5
Górny kwartyl	6	7

$p = 0.0149$ **(dokładna statystyka U, istotne statystycznie, $\alpha = 0.05$)**

Istnieją istotne statystycznie różnice pomiędzy studentkami a studentami matematyki w ilości czasu spędzanego przed komputerem – studentki spędzają mniej czasu przed komputerem niż studenci.

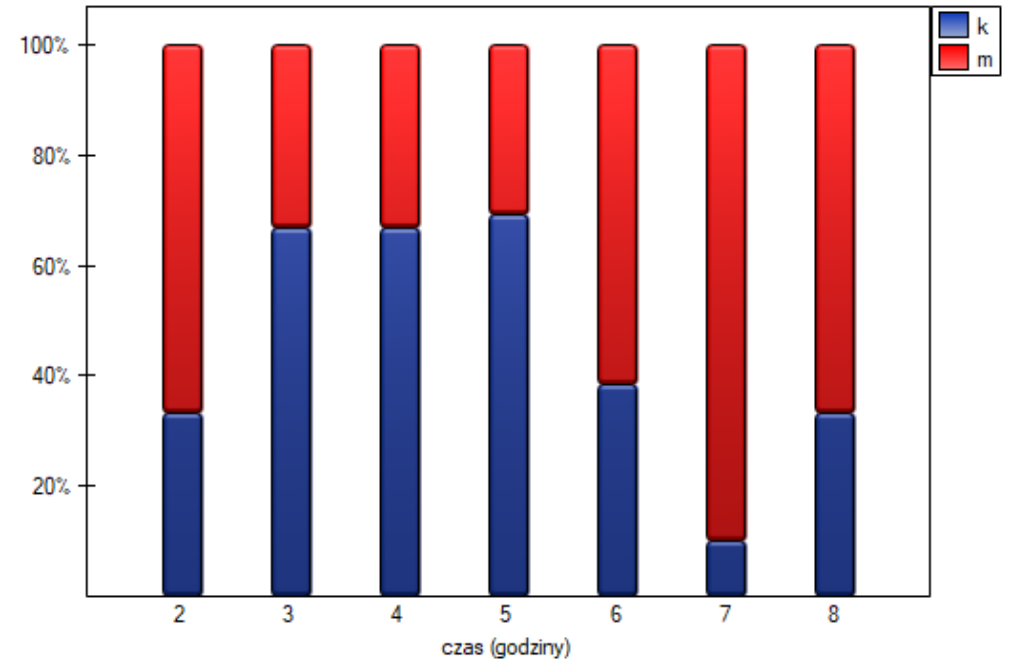
Wartość p wyznaczoną na podstawie statystyki testowej porównujemy zawsze z przyjętym poziomem istotności α: gdy $p \leq \alpha$, różnica jest istotna statystycznie; gdy $p > \alpha$, nie ma podstaw, by twierdzić, że grupy się różnią. Taką samą decyzję podjęlibyśmy na podstawie statystyki Z bez poprawki na ciągłość ($p = 0.0154$) i z poprawką ($p = 0.0158$). Różnice możemy opisać na podstawie mediany, kwartyli oraz wartości największej i najmniejszej, widocznych na wykresie typu ramka-wąsy, a także na podstawie tabeli liczności i procentów (którą uruchamiamy w oknie analizy, ustawiając statystyki opisowe $\Sigma\mu$) lub wykresu kolumnowego. Warto dodatkowo zgłosić **wielkość efektu** (d-Cohena) – sama istotność statystyczna nie mówi, jak duża jest różnica.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Liczność(Procent)		
Grupa	k	m
2	1 (4%)	2 (6.9%)
3	4 (16%)	2 (6.9%)
4	4 (16%)	2 (6.9%)
5	9 (36%)	4 (13.79%)
6	5 (20%)	8 (27.59%)
7	1 (4%)	9 (31.03%)
8	1 (4%)	2 (6.9%)
razem	25	29

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



DEFINICJE MIAR

r (wielkość efektu) – miara siły efektu oparta na przekształceniu statystyki testowej Z; przyjmuje wartości od 0 (brak efektu) do 1 (maksymalny efekt).

d-Cohena – standaryzowana miara wielkości efektu, wyliczana tu na podstawie r zgodnie ze standardową konwersją stosowaną przy meta-analizach. Pozwala ocenić praktyczne znaczenie różnicy niezależnie od liczności próby.

DLA ZAINTERESOWANYCH

Hipotezy:
Hipotezy dotyczą równości średnich rang dla porównywanych populacji lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0: &\quad \phi_1 = \phi_2, \\ \mathcal{H}_1: &\quad \phi_1 \neq \phi_2, \end{aligned}$$

gdzie ϕ_1, ϕ_2 – rozkłady badanej zmiennej w pierwszej i drugiej populacji.

Wzory:
W zależności od wielkości próby statystyka testowa przyjmuje inną postać. **Dla małej liczności próby:**

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1,$$

lub

$$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2,$$

gdzie n_1, n_2 – liczności prób, R_1, R_2 – sumy rang dla prób. Standardowo interpretacji podlega mniejsza z wartości U lub U'. Statystyka ta podlega rozkładowi Manna-Whitneya i nie zawiera poprawki na

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

rangi wiązane. Wartość dokładnego prawdopodobieństwa z rozkładu Manna-Whitneya wyliczana jest z dokładnością do części setnej ułamka. **Dla próby o dużej liczności:**

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n1+n_2+1)}{12} - \frac{n_1 n_2 \sum(t^3-t)}{12(n_1+n_2)(n_1+n_2-1)}}},$$

gdzie U można zastąpić przez U', t – liczba przypadków wchodzących w skład rangi wiązanej. Wzór na statystykę Z zawiera poprawkę na rangi wiązane – stosowaną, gdy rangi wiązane występują (gdy ich nie ma, $\frac{n_1 n_2 \sum(t^3-t)}{12(n_1+n_2)(n_1+n_2-1)} = 0$). Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny. **Poprawka na ciągłość testu Manna-Whitneya** (Marascuilo and McSweeney (1977)[117]) – stosujemy ją, by zapewnić możliwość przyjmowania przez statystykę testową wszystkich wartości liczb rzeczywistych zgodnie z założeniem rozkładu normalnego:

$$Z = \frac{\left|U - \frac{n_1 n_2}{2}\right| - 0.5}{\sqrt{\frac{n_1 n_2 (n1+n_2+1)}{12} - \frac{n_1 n_2 \sum(t^3-t)}{12(n_1+n_2)(n_1+n_2-1)}}}.$$

Wzory (miary efektu):
r wyliczamy ze wzoru $r = |Z/(n_1 + n_2)|$ [67], a d-Cohena:

$$d = \frac{2r}{\sqrt{1-r^2}}$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [46] wskazówkami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Test Wilcoxona (kolejności par)

KIEDY STOSOWAĆ TEN TEST:

Test kolejności par Wilcoxona *(ang. Wilcoxon matched-pairs test)* stosujemy, gdy tę samą cechę mierzymy dwukrotnie u tych samych osób, w różnych warunkach (np. przed i po leczeniu), a dane nie spełniają założeń testu t-Studenta dla grup zależnych (np. nie mają rozkładu normalnego). Sprawdzamy nim, czy wynik pomiaru się zmienił.

Test nazywany jest również testem Wilcoxona dla grup zależnych, Wilcoxon (1945[174], 1949[176]) – jest rozszerzeniem na dwie zależne próby testu rangowanych znaków Wilcoxona (przeznaczonego dla jednej próby).

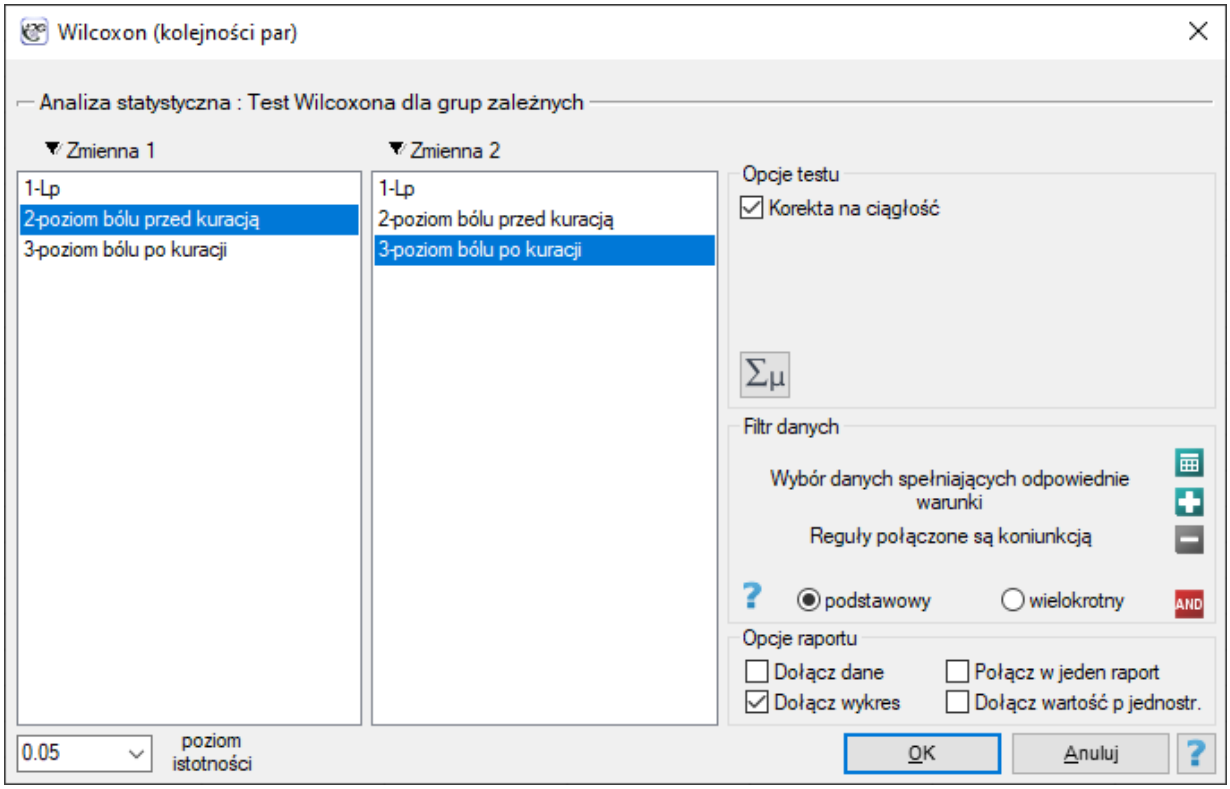
WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej,
- model zależny.

JAK USTAWIĆ ANALIZĘ

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Statystyka → Testy nieparametryczne → Wilcoxon (kolejności par) (lub poprzez Kreator).



Krok 1. Wskaż dwie zmienne odpowiadające pierwszemu i drugiemu pomiarowi (np. poziom bólu przed i po kuracji).

Krok 2. Uruchom analizę.

Uwaga! Od wersji 1.8.0 mediana wyliczana dla kolumny różnica obejmuje wszystkie pary wyników za wyjątkiem tych, których różnica wynosi 0.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 16.5. (plik ból.pqs)
Pobrano próbę 22 pacjentów cierpiących na raka. Badano u nich poziom odczuwanego bólu (na skali od 1 do 10, gdzie 1 to brak bólu a 10 to ból największy). Badanie to powtórzono po miesiącu stosowania kuracji nowym lekiem mającym obniżyć poziom odczuwanego bólu. Otrzymano następujące wyniki: (przed, po): (2, 2) (2, 3) (3, 1) (3,1) (3, 2) (3, 2) (3, 3) (4, 1) (4, 3) (4, 4) (5, 1) (5, 1) (5, 2) (5, 4) (5, 4) (6, 1) (6, 3) (7, 2) (7, 4) (7, 4) (8, 1) (8, 3). Chcemy sprawdzić, czy zastosowana kuracja ma wpływ na poziom odczuwanego bólu w populacji, z której pochodzi próba.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

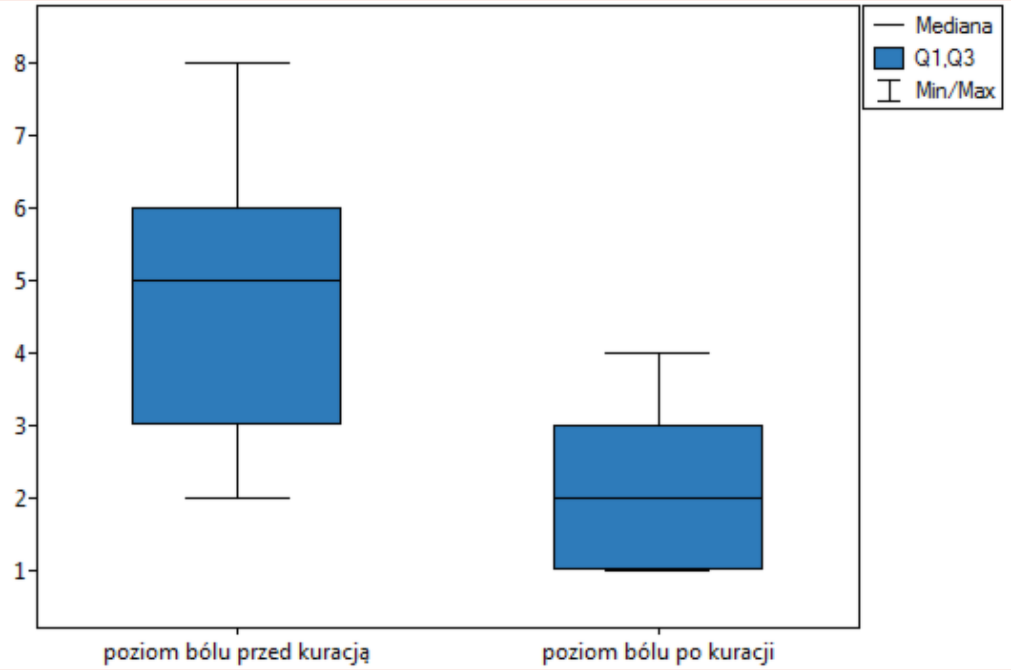
Test Wilcoxona dla grup zależnych	
Czas analizy	0,08 sek.
Analizowane zmienne	poziom bólu przed kuracją;p
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0,05
Korekta na ciągłość	Tak
Liczność - ilość par	22
Liczność - pominięte pary (wartości równe)	3
Mediana różnic	3
Statystyka T	3,5
Wartość p dwustronne (dokładne)	0,0001
Statystyka Z (z poprawką dla rang wiązanych)	3,68486
Wartość p dwustronne (asymptotyczne)	0,000229

Podsumowanie		
Grupa	poziom bólu	poziom bólu
Liczność	22	22
Mediana	5	2
Minimum	2	1
Maksimum	8	4
Dolny kwartyl	3	1
Górny kwartyl	6	3
Suma rang dodat	3,5	
Suma rang ujem	186,5	

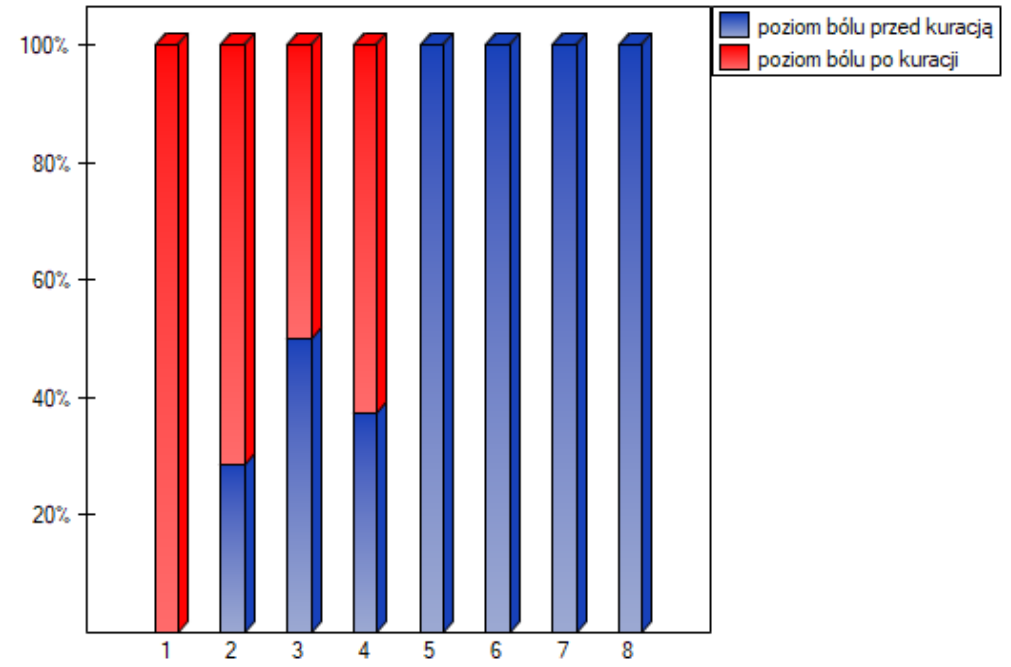
$p = 0.000211$ **(istotne statystycznie,** $\alpha = 0.05$**)**

W badanej populacji poziom odczuwanego bólu istotnie statystycznie zmienił się po miesiącu kuracji – spadł (suma rang ujemnych jest znacznie większa niż suma rang dodatnich).

Wartość p wyznaczoną na podstawie statystyki testowej porównujemy zawsze z przyjętym poziomem istotności α: gdy $p \leq \alpha$, różnica jest istotna statystycznie; gdy $p > \alpha$, nie ma podstaw, by twierdzić, że wystąpiła zmiana. Taką samą decyzję podjęlibyśmy również na podstawie wartości $p = 0.00021$ lub $p = 0.000229$ testu Wilcoxona opartego o statystykę Z (odpowiednio bez i z poprawką na ciągłość). Różnice możemy zaobserwować na wykresie typu ramka-wąsy lub wykresie kolumnowym. Warto dodatkowo zgłosić **wielkość efektu** (d-Cohena) – sama istotność statystyczna nie mówi, jak duża jest zmiana.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



DEFINICJE MIAR

r (wielkość efektu) – miara siły efektu oparta na przekształceniu statystyki testowej Z; przyjmuje wartości od 0 (brak efektu) do 1 (maksymalny efekt).

d-Cohena – standaryzowana miara wielkości efektu, wyliczana tu na podstawie r zgodnie ze standardową konwersją stosowaną przy meta-analizach. Pozwala ocenić praktyczne znaczenie różnicy niezależnie od liczności próby.

DLA ZAINTERESOWANYCH

Hipotezy: dotyczą równości sumy rang dodatnich i ujemnych lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0: &\quad \theta_0 = 0, \\ \mathcal{H}_1: &\quad \theta_0 \neq 0, \end{aligned}$$

gdzie θ_0 – mediana d_i w populacji.

Wzory: w zależności od wielkości próby statystyka testowa przyjmuje inną postać. **Dla małej liczności próby:**

$$T = \min\left(\sum R_-, \sum R_+\right),$$

gdzie $\sum R_+$ – suma rang dodatnich, $\sum R_-$ – suma rang ujemnych. Statystyka ta podlega rozkładowi Wilcoxona i nie zawiera poprawki na rangi wiązane. **Dla próby o dużej liczności:**

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie n – liczba rangowanych znaków (liczba rang), t – liczba przypadków wchodzących w skład rangi wiązanej. Wzór na statystykę Z zawiera poprawkę na rangi wiązane – stosowaną, gdy rangi wiązane występują (gdy ich nie ma, $\frac{\sum t^3 - \sum t}{48} = 0$). Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny. **Poprawka na ciągłość testu Wilcoxona** (Marascuilo and McSweeney (1977)[117]) – stosujemy ją, by zapewnić możliwość przyjmowania przez statystykę testową wszystkich wartości liczb rzeczywistych zgodnie z założeniem rozkładu normalnego:

$$Z = \frac{\left|T - \frac{n(n+1)}{4}\right| - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}.$$

r wyliczamy ze wzoru $r = |Z/n|$ [67], a d-Cohena:

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [46] wskazówkami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Testy chi-kwadrat

KIEDY STOSOWAĆ TE TESTY:

Testy chi-kwadrat sprawdzają, czy istnieje zależność pomiędzy dwiema cechami zebranymi w tabeli kontyngencji (np. czy płeć wiąże się ze zdawalnością egzaminu). W zależności od wielkości tabeli mówimy o teście chi-kwadrat 2×2 (dwie kategorie każdej cechy) lub teście chi-kwadrat $R \times C$ (więcej kategorii):

- test chi-kwadrat dla małych tabel (2×2),
- test chi-kwadrat dla dużych tabel ($R \times C$).

WARUNKI STOSOWANIA

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- model niezależny,
- duże liczności oczekiwane według interpretacji Cochrana (1952)[41].

Uwaga! Wartość p testu porównujemy zawsze z przyjętym poziomem istotności α: gdy $p \leq \alpha$, uznajemy, że badane cechy są ze sobą powiązane; gdy $p > \alpha$, nie ma podstaw, by twierdzić, że taka zależność istnieje.

DODATKOWO

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Gdy warunek Cochrana nie jest spełniony, zamiast testu chi-kwadrat można wyznaczyć: test Fishera dla tabel $R \times C$, test Fishera dla tabel 2×2, test chi-kwadrat z poprawką Yatesa, test mid-p dla tabel 2×2.
- Gdy uzyskamy tabelę $R \times 2$, a kategorie R da się uporządkować, można sprawdzić trend: test chi-kwadrat dla trendu.
- Gdy dla tabeli większej niż 2×2 wykryjemy istotną zależność i chcemy sprawdzić, między którymi kategoriami ona występuje, można wykonać porównania wielokrotne z odpowiednią korektą – w oknie opcji testu zaznaczamy Wielokrotne porównania kolumn (RxC).
- By opisać siłę związku między badanymi cechami, można wyznaczyć: miary zależności.
- By dla tabel 2×2 opisać wielkość wpływu czynnika ryzyka, można wyznaczyć: iloraz szans (OR) i relatywne ryzyko (RR).

DLA ZAINTERESOWANYCH

Hipotezy (w brzmieniu ogólnym):

$$\begin{aligned} \mathcal{H}_0: &\quad O_{ij} = E_{ij} \text{ dla wszystkich kategorii,} \\ \mathcal{H}_1: &\quad O_{ij} \neq E_{ij} \text{ dla przynajmniej jednej kategorii,} \end{aligned}$$

gdzie O_{ij} – liczności obserwowane w tabeli kontyngencji, E_{ij} – liczności oczekiwane w tabeli kontyngencji.

Hipotezy (w brzmieniu testu niezależności):

$\mathcal{H}_0$: nie istnieje zależność pomiędzy badanymi cechami populacji (obie klasyfikacje ze względu na cechę X i na cechę Y są statystycznie niezależne),
$\mathcal{H}_1$: istnieje zależność pomiędzy badanymi cechami populacji.

Test chi-kwadrat dla dużych tabel

KIEDY STOSOWAĆ TEN TEST:

Test χ^2 dla tabel $r \times c$, znany również jako test χ^2 Pearsona *(ang. Pearson's Chi-square test)*, Karl Pearson 1900, stosujemy, gdy chcemy sprawdzić, czy dwie cechy o dowolnej liczbie kategorii (np. kraj zamieszkania i wykształcenie) są ze sobą powiązane. Jest to rozszerzenie testu χ^2 (dobroci dopasowania) na dwie cechy.

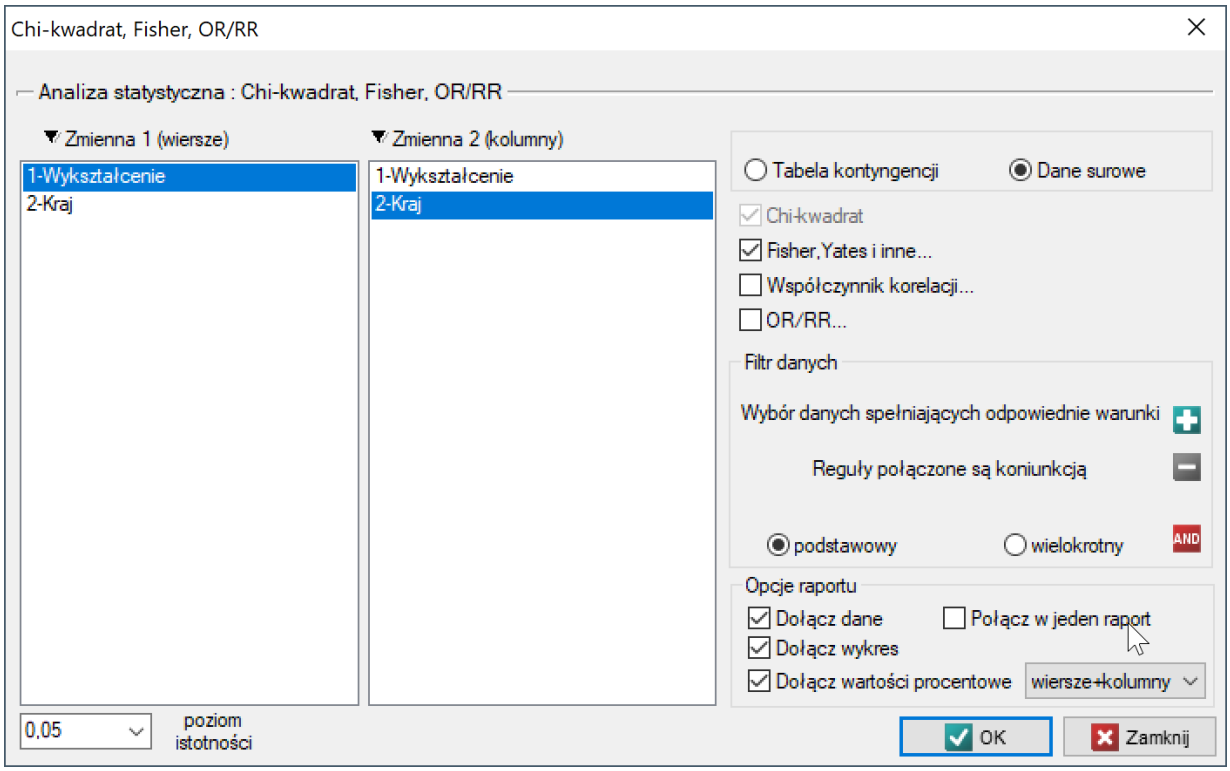
WARUNKI STOSOWANIA

- pomiar na skali nominalnej,
- model niezależny,
- duże liczności oczekiwane według interpretacji Cochrana (1952)[41].

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR (lub poprzez Kreator).



Krok 1. Wskaż dwie zmienne (cechy), których zależność chcesz sprawdzić.

Krok 2. Uruchom analizę.

WYKONANIE I INTERPRETACJA

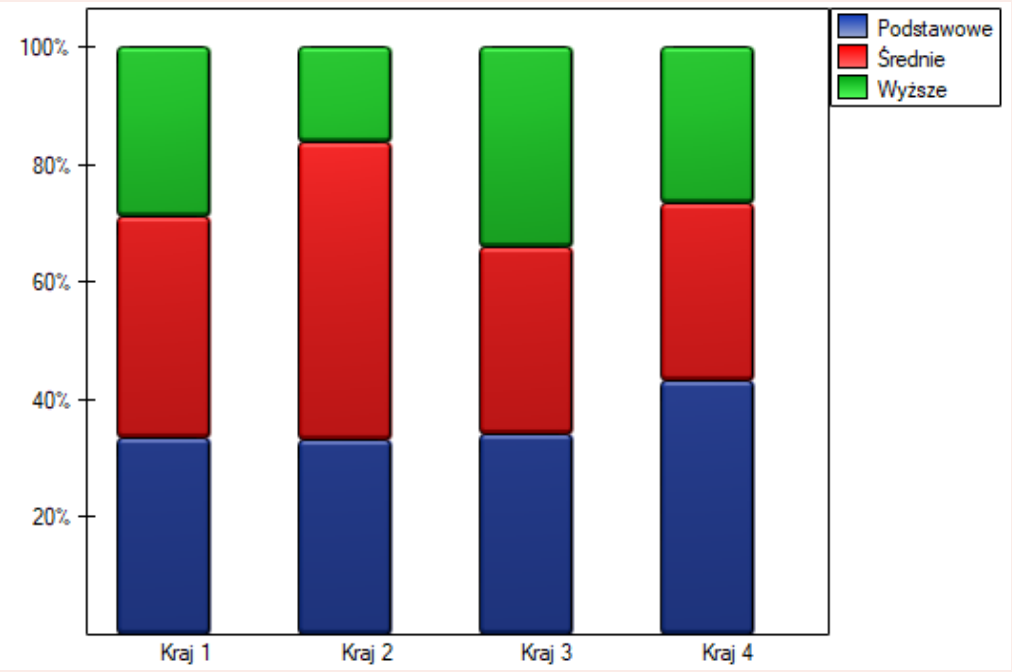
Przykład 16.6. (plik kraj-wykształcenie.pqs)
Rozpatrujemy próbę 605 osób ($n = 605$), dla których badamy 2 cechy (X=kraj zamieszkania, Y=wykształcenie). Pierwsza cecha występuje w 4, a druga w 3 kategoriach (X_1=Kraj 1, X_2=Kraj 2, X_3=Kraj 3, X_4=Kraj 4, Y_1=podstawowe, Y_2=średnie, Y_3=wyższe). Rozkład danych przedstawia tabela kontyngencji:

Dane :				↙Kraj
↓Wykształcenie	Kraj 4	Kraj 3	Kraj 2	Kraj 1
Podstawowe	65	52	51	50
Średnie	45	48	78	56
Wyższe	40	52	25	43

Na podstawie tej próby chcielibyśmy się dowiedzieć, czy w badanej populacji istnieje zależność pomiędzy wykształceniem a krajem zamieszkania.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Chi-kwadrat, Fisher, OR/RR	
Czas analizy	0,21 sek.
Analizowane zmienne	Wykształcenie;Kraj
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0,05
Liczność	605
Warunek Cochrana	spełniony
Statystyka chi-kwadrat Pearsona	23,6104
Stopnie swobody	6
Wartość p	0,0006



$p = 0.0006$ **(istotne statystycznie,** $\alpha = 0.05$**)**

W badanej populacji istnieje zależność pomiędzy krajem zamieszkania a wykształceniem.

Warunek Cochrana jest spełniony. Jeśli interesują nas dokładniejsze informacje na temat wykrytych zależności, uzyskamy je, wyznaczając porównania wielokrotne poprzez opcję Fisher, Yates i inne..., a następnie Wielokrotne porównania kolumn (RxC) i jedną z poprawek, np. Benjamini-Hochberg.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wielokr. por. Benjamini-Hochberg :
Wartość p, Chi^2, *Cochran niesp.

Wykształc	Kraj 1	Kraj 2	Kraj 3	Kraj 4
Kraj 1		0,0314	0,4777	0,2555
Kraj 2	0,0314		0,0015	0,0028
Kraj 3	0,4777	0,0015		0,2555
Kraj 4	0,2555	0,0028	0,2555	

Dokładniejsza analiza pozwala stwierdzić, że jedynie drugi kraj różni się poziomem wykształcenia od pozostałych krajów w sposób istotny statystycznie.

DLA ZAINTERESOWANYCH

Hipotezy:

$\mathcal{H}_0$: nie istnieje zależność pomiędzy badanymi cechami w populacji,
$\mathcal{H}_1$: istnieje zależność pomiędzy badanymi cechami w populacji.

Wzory: Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z liczbą stopni swobody wyznaczaną według wzoru: $df = (r-1)(c-1)$.

Test Fishera dla dużych tabel

KIEDY STOSOWAĆ TEN TEST:

Test Fishera dla tabel $r \times c$, zwany również testem Fishera-Freemana-Haltona *(ang. Fisher-Freeman-Halton test)*, Freeman G.H., Halton J.H. (1951)[64], stosujemy zamiast testu χ^2 dla tabel $r \times c$, gdy liczności oczekiwane w tabeli są zbyt małe (warunek Cochrana nie jest spełniony). Jest to rozszerzenie na tabele $r \times c$ testu dokładnego Fishera, wyznaczające dokładne prawdopodobieństwo wystąpienia konkretnego rozkładu liczb w tabeli przy znanym n i ustalonych sumach brzegowych.

WARUNKI STOSOWANIA

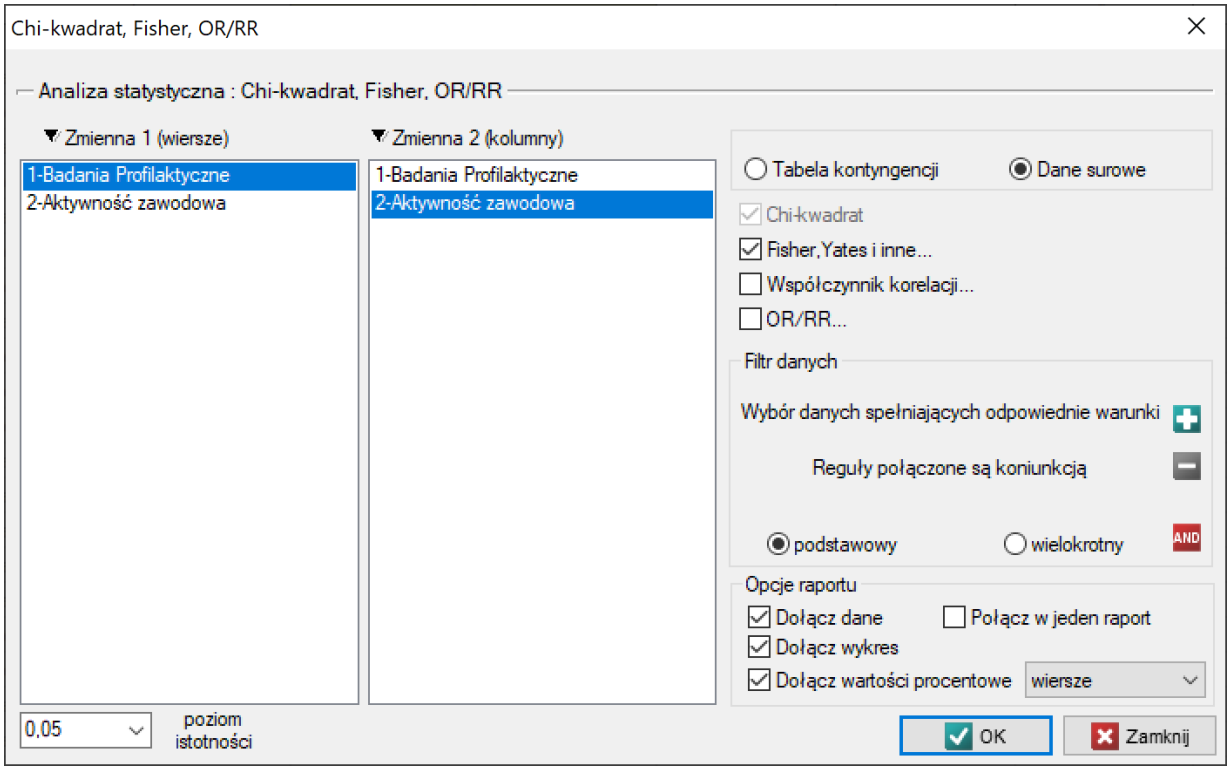
- pomiar na skali nominalnej,
- model niezależny,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– małe liczności oczekiwane (test χ^2 nie jest wskazany).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR (lub poprzez Kreator).



Krok 1. Wskaż dwie zmienne (cechy), których zależność chcesz sprawdzić.

Krok 2. Uruchom analizę.

Uwaga! Procedura obliczania wartości p dla tego testu bazuje na algorytmie opublikowanym w pracy Mehta (1986)[122].

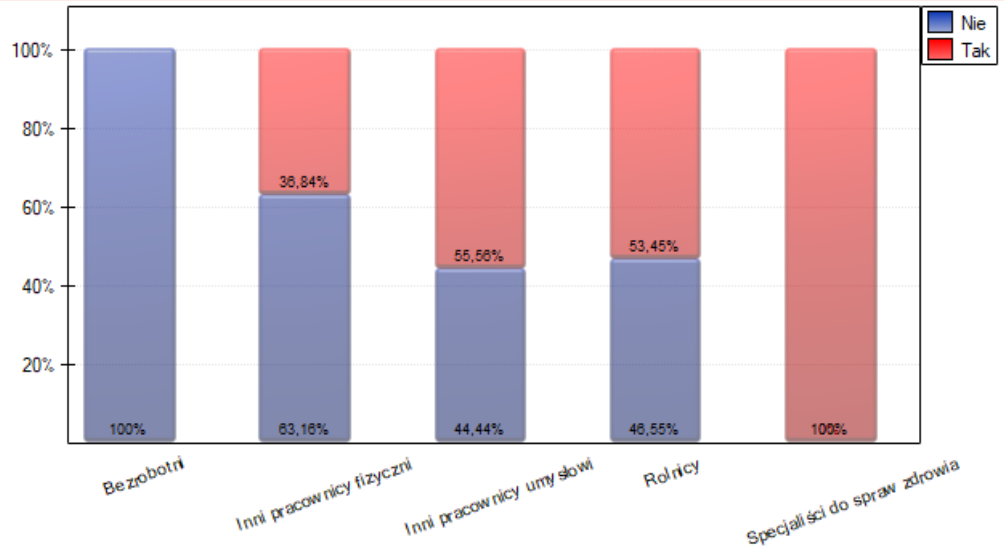
WYKONANIE I INTERPRETACJA

Przykład 16.7. (plik praca-profilaktyka.pqs)
W populacji osób zamieszkujących obszary wiejskie gminy Komorniki badano, czy wykonywanie badań profilaktyki zdrowia jest uzależnione od rodzaju aktywności zawodowej mieszkańców. Zebrano losową próbę 120 osób i zapytano o wykształcenie oraz o to, czy osoby te wykonują badania profilaktyczne. Pełną odpowiedź uzyskano od 113 osób.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Dane :					Aktywność zawodowa
↓Badania profilaktyczne	Bezrobotni	Inni pracownicy fizyczni	Inni pracownicy umysłowi	Rolnicy	Specjaliści do spraw zdrowia
Nie	17	12	4	27	0
Tak	0	7	5	31	10

Chi-kwadrat, Fisher, OR/RR	
Czas analizy	0,11 sek.
Analizowane zmienne	Czy wykonuje bad profilakty
Liczność - niezinterpretowane	0
Liczność - braki danych	7
Poziom istotności	0,05
Liczność	113
Warunek Cochrana	niespełniony
Statystyka chi-kwadrat Pearsona	28,378
Stopnie swobody	4
Wartość p	<0,0001
Test dokładny Fishera	
Wartość p jednostronne	NA
Wartość p dwustronne	<0,0001



$p < 0.0001$ **(test Fishera, istotne statystycznie,** $\alpha = 0.05$**)**

W badanej populacji istnieje zależność pomiędzy wykonywaniem badań profilaktycznych a rodzajem wykonywanej pracy.

Warunek Cochrana nie jest spełniony, dlatego nie powinniśmy stosować testu chi-kwadrat – stąd wybór testu Fishera. Jeśli interesują nas dokładniejsze informacje na temat wykrytych zależności, uzyskamy je, wyznaczając porównania wielokrotne poprzez opcję Fisher, Yates i inne..., a następnie Wielokrotne porównania kolumn (RxC) i jedną z poprawek, np. Benjamini-Hochberg.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wielokr. por. Benjamini-Hochberg : Wartość p, Chi^2/*Fisher(Cochran niesp.)					
↓Badania profilaktyczne	Bezrobotni	Inni pracownicy fizyczni	Inni pracownicy umysłowi	Rolnicy	Specjaliści do spraw zdrowia
Bezrobotni		0,0139*	0,0048*	0,0004	<0,0001*
Inni pracownicy fizyczni	0,0139*		0,4799*	0,2611	0,0039*
Inni pracownicy umysłowi	0,0048*	0,4799*		1*	0,0464*
Rolnicy	0,0004	0,2611	1*		0,0091*
Specjaliści do spraw zdrowia	<0,0001*	0,0039*	0,0464*	0,0091*	

Dokładniejsza analiza pozwala stwierdzić, że specjaliści do spraw zdrowia istotnie częściej niż pozostałe grupy wykonują badania profilaktyczne (100% osób w tej grupie), a bezrobotni istotnie rzadziej (nikt w tej grupie nie wykonał badania). Rolnicy, inni pracownicy fizyczni i inni pracownicy umysłowi w około 50% wykonują badania profilaktyczne, co sprawia, że te trzy grupy nie różnią się od siebie istotnie statystycznie. Wartości p w tabeli oznaczone gwiazdką pochodzą z testu dokładnego Fishera z poprawką Benjaminiego-Hochberga, a wartości nieoznaczone – z testu chi-kwadrat z tą samą poprawką (tam, gdzie warunek Cochrana był spełniony).

DLA ZAINTERESOWANYCH

Hipotezy:

$\mathcal{H}_0$: nie istnieje zależność pomiędzy badanymi cechami w populacji,
$\mathcal{H}_1$: istnieje zależność pomiędzy badanymi cechami w populacji.

Wzory: sumy brzegowe wierszy definiujemy jako:

$$W_i = \sum_{j=1}^{c} O_{ij},$$

gdzie O_{ij} – liczności obserwowane w tabeli kontyngencji, a sumy brzegowe kolumn jako:

$$K_i = \sum_{i=1}^{r} O_{ij}.$$

Przy ustalonych sumach brzegowych, dla różnych układów wartości obserwowanych oznaczonych jako U_{ij}, wyznaczamy prawdopodobieństwa P:

$$P = \frac{D^{-1} \prod_{j=1}^{c} K_j!}{U_{1j}!U_{2j}! \ldots U_{rj}},$$

gdzie

$$D = \frac{(W_1 + W_2 + \ldots + W_r)!}{W_1!W_2! \ldots W_r!}.$$

Interpretacja: dokładny poziom istotności p jest sumą tych prawdopodobieństw P (wyznaczonych dla nowych wartości U_{ij}), które są mniejsze lub równe prawdopodobieństwu P tabeli z wartościami początkowymi O_{ij}.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test chi-kwadrat dla małych tabel

KIEDY STOSOWAĆ TEN TEST:

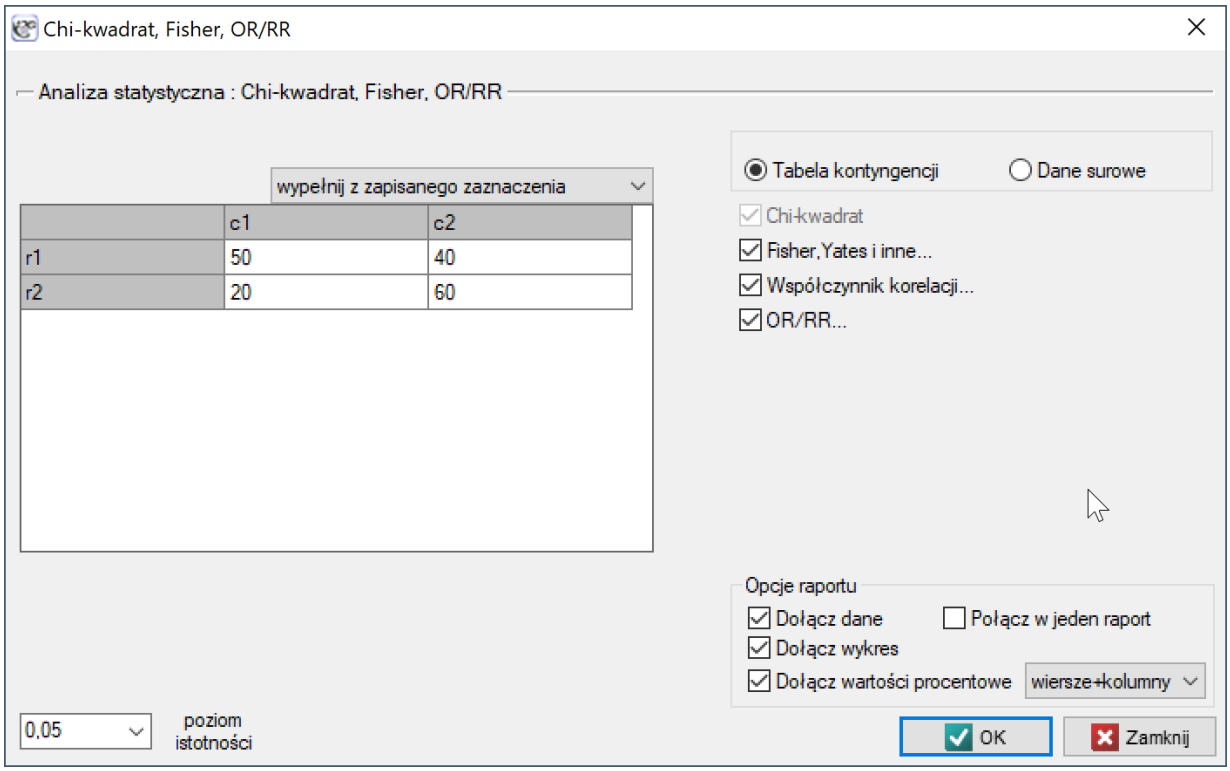
Test χ^2 dla tabel 2×2 *(ang. Pearson's Chi-square test)*, Karl Pearson 1900, stosujemy, gdy chcemy sprawdzić, czy dwie cechy o dwóch kategoriach każda (np. płeć i zdawalność egzaminu) są ze sobą powiązane. Jest to szczególny przypadek testu χ^2 dla tabel $r \times c$ dla najmniejszej możliwej tabeli.

WARUNKI STOSOWANIA

- pomiar na skali nominalnej,
- model niezależny,
- tabela kontyngencji 2×2 (patrz tabela (11.1)).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR (lub poprzez Kreator).



Krok 1. Wskaż dwie zmienne (cechy), których zależność chcesz sprawdzić.

Krok 2. Uruchom analizę.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uwaga! Gdy liczności oczekiwane w tabeli są małe (poniżej 5), zamiast podstawowego testu χ^2 lepiej opierać wniosek na teście z poprawką Yatesa, teście dokładnym Fishera lub teście mid-p – wszystkie są liczone automatycznie razem z testem podstawowym.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 16.8. (plik płeć-egzamin.pqs)
Rozpatrzmy próbę składającą się z 170 osób ($n = 170$), dla których badamy 2 cechy (X=płeć, Y=zdawalność egzaminu). Każda z tych cech występuje w dwóch kategoriach (X_1=k, X_2=m, Y_1=tak, Y_2=nie). Na podstawie tej próby chcielibyśmy się dowiedzieć, czy w badanej populacji istnieje zależność pomiędzy płcią a zdawalnością egzaminu. Rozkład danych przedstawia tabela kontyngencji:

Liczności obserwowane O_{ij}		zdawalność egzaminu		
		tak	nie	suma
płeć	k	50	40	90
	m	20	60	80
	suma	70	100	170

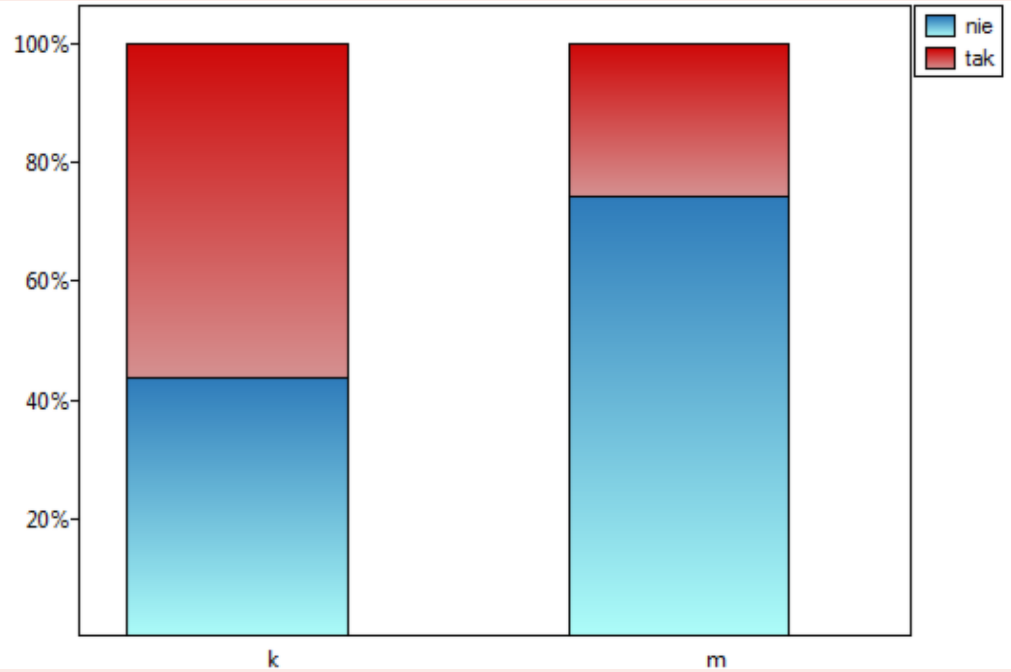
Chi-kwadrat, Fisher, OR/RR (2x2)	
Czas analizy	0,15 sek.
Analizowane zmienne	Tabela kontyngencji
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0,05
Liczność	170
Warunek Cochrana	spełniony
Statystyka Chi-kwadrat	16,325396825
Stopnie swobody	1
Wartość p	0,000053344
Statystyka z pop. Yatesa	15,088258929
Stopnie swobody	1
Wartość p z pop. Yatesa	0,000102599
Test dokładny Fishera	
Wartość p jednostronne	0,000043395
Wartość p dwustronne	0,00008324
Test mid-p	
2 * wartość p jednostronne	0,000054186

Oczekiw.:		✓Tabela ko
ı	tak	nie
k	37,06	52,94
m	32,94	47,06

Dane :		✓Tabela ko
ı	tak	nie
k	50	40
m	20	60

% wiersza		✓Tabela ko
ı	tak	nie
k	55,56%	44,44%
m	25%	75%

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p = 0.000053$ **(test chi-kwadrat, istotne statystycznie,** $\alpha = 0.05$**)**

W badanej populacji istnieje zależność pomiędzy płcią a zdawalnością egzaminu.

Tabela liczności oczekiwanych nie zawiera wartości mniejszych niż 5, więc warunek Cochrana jest spełniony – podstawowy test χ^2 można było zastosować wprost. Do tego samego wniosku prowadzą pozostałe warianty testu: test z poprawką Yatesa ($p = 0.000103$), test dokładny Fishera ($p = 0.000083$) i test mid-p ($p = 0.000054$).

DLA ZAINTERESOWANYCH

Hipotezy:

$\mathcal{H}_0$: nie istnieje zależność pomiędzy badanymi cechami w populacji,
$\mathcal{H}_1$: istnieje zależność pomiędzy badanymi cechami w populacji.

Wzory: Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z jednym stopniem swobody.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Poprawki testu chi-kwadrat dla małych tabel

KIEDY STOSOWAĆ TE WARIANTY:

Gdy w tabeli 2×2 liczności oczekiwane są małe (warunek Cochrana nie jest spełniony), podstawowy test χ^2 może dawać zawyżoną wiarygodność wyniku. W takiej sytuacji lepiej oprzeć wniosek na jednym z poniższych, bardziej odpowiednich dla małych liczności wariantów. Wszystkie są liczone automatycznie razem z testem podstawowym.

WARUNKI STOSOWANIA

- pomiar na skali nominalnej,
- model niezależny,
- tabela kontyngencji 2×2 (patrz tabela (11.1)).

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Testy nieparametryczne $\rightarrow$ Chi-kwadrat, Fisher, OR/RR (lub poprzez Kreator) – wszystkie warianty pojawiają się w jednym raporcie razem z testem podstawowym.

TEST CHI-KWADRAT Z POPRAWKĄ YATESA NA CIĄGŁOŚĆ

Test χ^2 z poprawką Yatesa *(ang. Chi-square test with Yates correction)*, Frank Yates (1934)[180], jest bardziej konserwatywny od testu χ^2 – trudniej niż test χ^2 odrzuca hipotezę zerową. Poprawka na ciągłość ma zapewnić możliwość przyjmowania przez statystykę testową wszystkich wartości liczb rzeczywistych zgodnie z założeniem rozkładu χ^2.

TEST FISHERA DLA TABEL 2×2

Test Fishera dla tabel 2×2, nazywany również testem dokładnym Fishera *(ang. Fisher exact test)*, R. A. Fisher (1934[57], 1935[58]), określa dokładne prawdopodobieństwo wystąpienia konkretnego rozkładu liczb w tabeli przy znanym n i ustalonych sumach brzegowych.

TEST MID-P

Test mid-p jest korektą testu dokładnego Fishera, rekomendowaną przez wielu statystyków (Lancaster 1961[100], Anscombe 1981[9], Pratt i Gibbons 1981[138], Plackett 1984[137], Miettinen 1985[123], Barnard 1989[15], Rothman 2008[143]) jako metoda zmniejszenia konserwatyzmu testu dokładnego Fishera – testem mid-p szybciej odrzucimy hipotezę zerową niż dokładnym testem Fishera. Dla dużych prób wartość p testu χ^2 z poprawką Yatesa i testu Fishera są zbliżone, natomiast wartość p testu χ^2 bez korekcji koresponduje z wartością mid-p.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Test chi-kwadrat z poprawką Yatesa – statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}.$$

Test Fishera – przy znanych sumach brzegowych, dla różnych układów wartości obserwowanych wyznaczamy prawdopodobieństwa P:

$$P = \frac{\binom{O_{11}+O_{21}}{O_{11}}\binom{O_{12}+O_{22}}{O_{12}}}{\binom{O_{11}+O_{12}+O_{21}+O_{22}}{O_{11}+O_{12}}}.$$

Dokładny poziom istotności p jest sumą tych prawdopodobieństw, które są mniejsze lub równe badanemu prawdopodobieństwu.

Test mid-p – jednostronna wartość p wyznaczana jest przez przekształcenie wartości prawdopodobieństwa testu dokładnego Fishera:

$$p_{I(mid-p)} = p_{I(Fisher)} - 0.5 \cdot P_{punktu(tabeli \quad zadanej)},$$

gdzie $p_{I(mid-p)}$ – wartość jednostronna p testu mid-p, $p_{I(Fisher)}$ – wartość jednostronna p testu dokładnego Fishera. Dwustronna wartość p jest zdefiniowana jako podwojona wartość mniejszego z jednostronnych prawdopodobieństw:

$$p_{II(mid-p)} = 2p_{I(mid-p)}.$$

Test chi-kwadrat dla trendu

KIEDY STOSOWAĆ TEN TEST:

Test χ^2 dla trendu, nazywany również testem Cochrana-Armitage *(ang. Cochran-Armitage test for trend)*[42][10], stosujemy, gdy mamy uporządkowane kategorie jednej cechy (np. poziom wykształcenia) i dwie grupy drugiej cechy (np. pali/nie pali), a chcemy sprawdzić, czy proporcje w kolejnych kategoriach rosną lub maleją w sposób systematyczny – czyli czy widoczny jest trend.

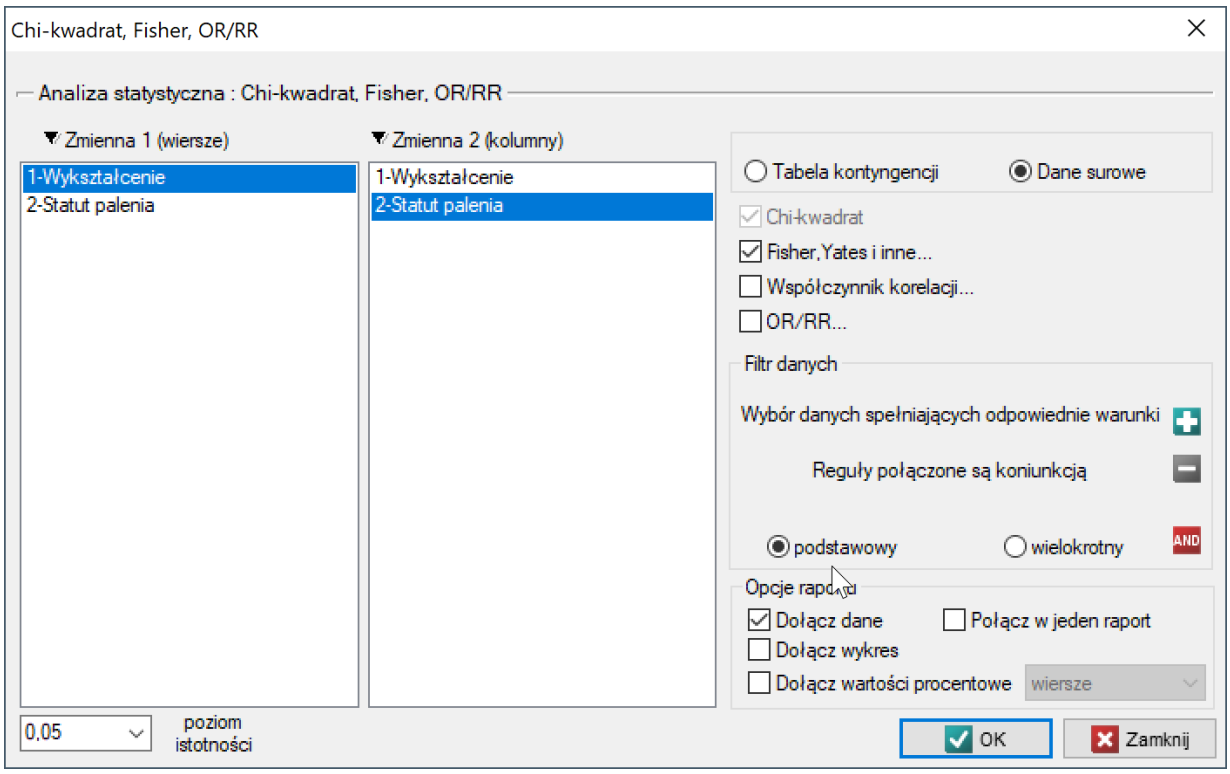
WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej (kategorie muszą dać się uporządkować),
- model niezależny (druga cecha to 2 niezależne grupy).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR, opcja Fisher, Yates i inne... → Chi-kwadrat dla trendu.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Krok 1. Wskaż cechę uporządkowaną (np. wykształcenie) i cechę dwukategorialną (np. palenie).

Krok 2. Uruchom analizę.

Uwaga! Przed analizą trzeba wskazać kolejność, w jakiej mają się pojawiać kategorie cechy uporządkowanej. W tym celu we właściwościach zmiennej wybieramy Kody/Etykiety/Format... i nadajemy kolejność, podając kolejne liczby naturalne, a następnie przypisujemy im etykiety.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 16.9. (plik palenie-wykształcenie.pqs)
Sprawdzamy, czy palenie papierosów związane jest z wykształceniem mieszkańców pewnej wsi. Wylosowano próbę 122 osób. Zakładamy, że zależność może być dwojakiego typu: im bardziej wykształceni ludzie, tym częściej palą, albo im bardziej wykształceni ludzie, tym rzadziej palą – poszukujemy więc rosnącego lub malejącego trendu.

Kody i etykiety dla wartości

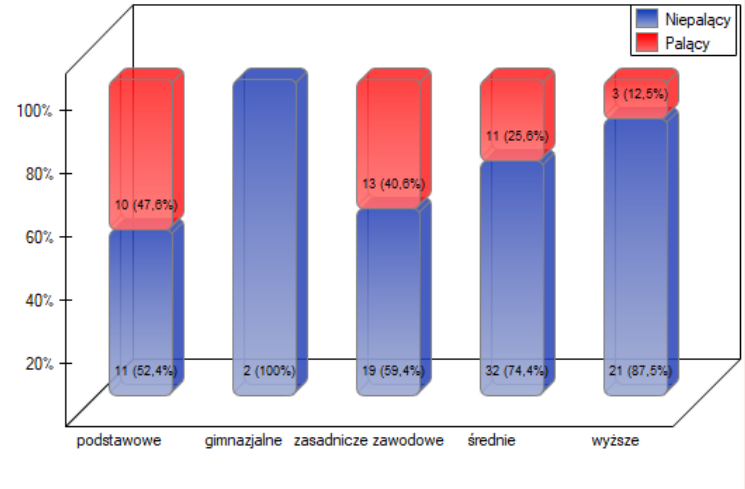
Wartość	Obowiązująca wartość	Etykieta
gimnazjalne	2	gimnazjalne
podstawowe	1	podstawowe
wyższe	5	wyższe
zasadnicze zawodowe	3	zasadnicze zawodowe
średnie	4	średnie

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Statystyka chi-kwadrat dla trendu	6,8098
Stopnie swobody	1
Wartość p	0,0091

$p = 0.0091$ **(istotne statystycznie,** $\alpha = 0.05$**)**

W badanej populacji istnieje trend w częstości palenia wraz ze wzrostem wykształcenia.



Jak wynika z wykresu, im osoby są bardziej wykształcone, tym rzadziej palą. Wynik uzyskany przez osoby o wykształceniu gimnazjalnym odbiega jednak od tego trendu – ponieważ ta kategoria dotyczy tylko dwóch osób, nie miało to dużego wpływu na wynik. Ze względu na bardzo małą liczność tej grupy analizę powtórzono dla połączonych kategorii wykształcenia podstawowego i gimnazjalnego.

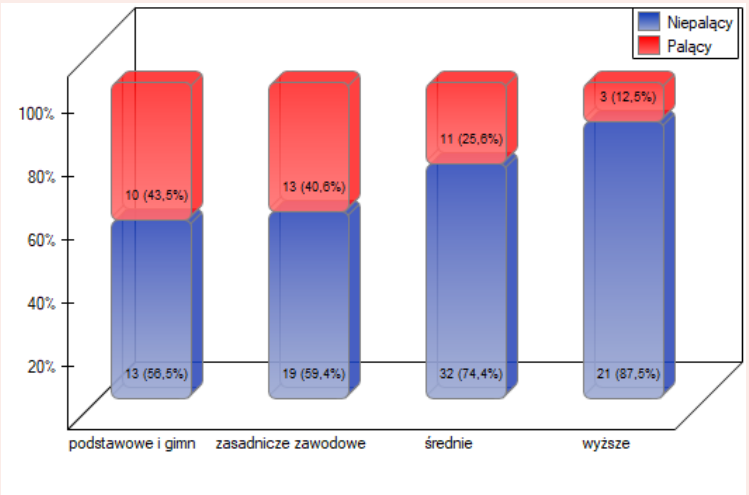
Kody i etykiety dla wartości

Wartość	Obowiązująca wartość	Etykieta
gimnazjalne	0	podstawowe i gimn
podstawowe	0	podstawowe i gimn
wyższe	3	wyższe
zasadnicze zawodowe	1	zasadnicze zawodowe
średnie	2	średnie

$p = 0.0078$ **(po połączeniu kategorii, istotne statystycznie,** $\alpha = 0.05$**)**

Trend w częstości palenia wraz ze wzrostem wykształcenia potwierdza się również po połączeniu najmniej licznych kategorii.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Przykład 16.10. (plik widzowie.pqs)
Z powodu spadku oglądalności pewnego serialu telewizyjnego przeprowadzono badanie opinii widzów. Przepytano 100 osób, które rozpoczęły oglądanie serialu niedawno, i 300, które oglądają go systematycznie od początku. Zapytano ich między innymi o stopień zaabsorbowania losami bohaterów. Wyniki zapisano w tabeli poniżej:

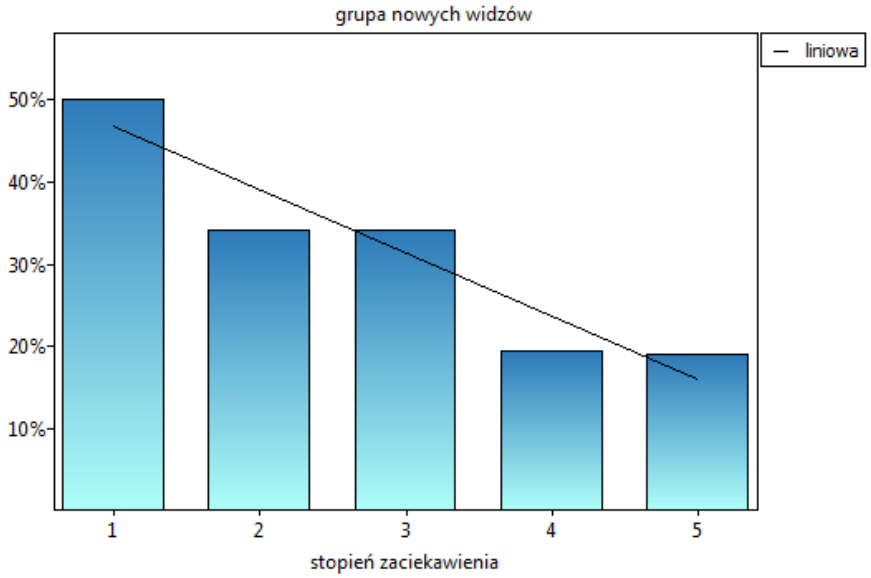
Stopień zaciekawienia	grupa		
	grupa nowych widzów	grupa stałych widzów	suma
raczej niewielki	7	7	14
przeciętny	13	25	38
raczej wysoki	30	58	88
wysoki	24	99	123
bardzo wysoki	26	111	137
suma	100	300	400

Nowi widzowie stanowią 25% badanych. Taka proporcja nie utrzymuje się jednak dla każdej kategorii stopnia zaciekawienia:

Stopień zaciekawienia	grupa		
	grupa nowych widzów	grupa stałych widzów	suma
raczej niewielki	p_1=50.00%	50.00%	100%
przeciętny	p_2=34.21%	65.79%	100%
raczej wysoki	p_3=34.09%	65.91%	100%
wysoki	p_4=19.51%	80.49%	100%
bardzo wysoki	p_5=18.98%	81.02%	100%
suma	**25.00%**	**75.00%**	**100%**

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test Chi-kwadrat dla trendu	
Czas analizy	0.03 sek.
Analizowane zmienne	stopień zaciekawienia;grupa
Poziom istotności	0.05
Liczność - grupy1	100
Liczność - grupy2	300
Statystyka Chi-kwadrat	12.3702523
Ilość stopni swobody	1
Wartość p	0.0004362



Oczekiw.:

	grupa now	grupa stał
1	3.5	10.5
2	9.5	28.5
3	22	66
4	30.75	92.25
5	34.25	102.75

Dane :

	grupa now	grupa stał
1	7	7
2	13	25
3	30	58
4	24	99
5	26	111

wiersze:

	grupa now	grupa stał
1	50%	50%
2	34.21%	65.79%
3	34.09%	65.91%
4	19.51%	80.49%
5	18.98%	81.02%

$p = 0.000436$ **(istotne statystycznie,** $\alpha = 0.05$**)**

W populacji widzów serialu istnieje trend w proporcji nowych widzów: im grupa jest bardziej zainteresowana losami bohaterów, tym mniejszą jej część stanowią nowi widzowie (trend malejący).

DLA ZAINTERESOWANYCH

Oznaczenia: tabela kontyngencji $r \times 2$ liczności obserwowanych:

Liczności obserwowane O_{ij}		Cecha 2 (grupa)		
		G_1	G_2	Suma
Cecha 1 (cecha X)	X_1	O_{11}	O_{12}	$W_1 = O_{11} + O_{12}$
	X_2	O_{21}	O_{22}	$W_2 = O_{21} + O_{22}$
	...	...	...	...
	X_r	O_{r1}	O_{r2}	$W_r = O_{r1} + O_{r2}$
	Suma	$C_1 = \sum_{i=1}^{r} O_{i1}$	$C_2 = \sum_{i=1}^{r} O_{i2}$	$n = C_1 + C_2$

Niech $p_1, p_2, ..., p_r$ oznaczają proporcje $p_1 = \frac{O_{11}}{W_1}$, $p_2 = \frac{O_{21}}{W_2}$,..., $p_r = \frac{O_{r1}}{W_r}$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Hipotezy:

$\mathcal{H}_0$: w badanej populacji nie istnieje trend w proporcjach $p_1, p_2, ..., p_r$,
$\mathcal{H}_1$: w badanej populacji istnieje trend w proporcjach $p_1, p_2, ..., p_r$.

Wzory: Statystyka testowa ma postać:

$$\chi^2 = \frac{\left[\left(\sum_{i=1}^{r} i \cdot O_{i1}\right) - C_1 \left(\sum_{i=1}^{r} \frac{i \cdot W_i}{n}\right)\right]^2}{\frac{C_1}{n}\left(1 - \frac{C_1}{n}\right)\left[\left(\sum_{i=1}^{n} i^2 W_i\right) - n\left(\sum_{i=1}^{n} \frac{i \cdot W_i}{n}\right)^2\right]}.$$

Statystyka ta ma rozkład χ^2 z 1 stopniem swobody.

Relatywne Ryzyko i Iloraz Szans

KIEDY STOSOWAĆ TE MIARY:

Ilorazu Szans i Relatywnego Ryzyka używamy, gdy chcemy sprawdzić, jak narażenie na jakiś czynnik (np. palenie papierosów) wiąże się z pojawieniem się badanego zjawiska (np. choroby) – innymi słowy, o ile większa (lub mniejsza) jest szansa albo ryzyko wystąpienia zjawiska w grupie narażonej na czynnik w porównaniu z grupą nienarażoną. Obie miary wyznacza się na podstawie danych zebranych w tabeli kontyngencji 2×2.

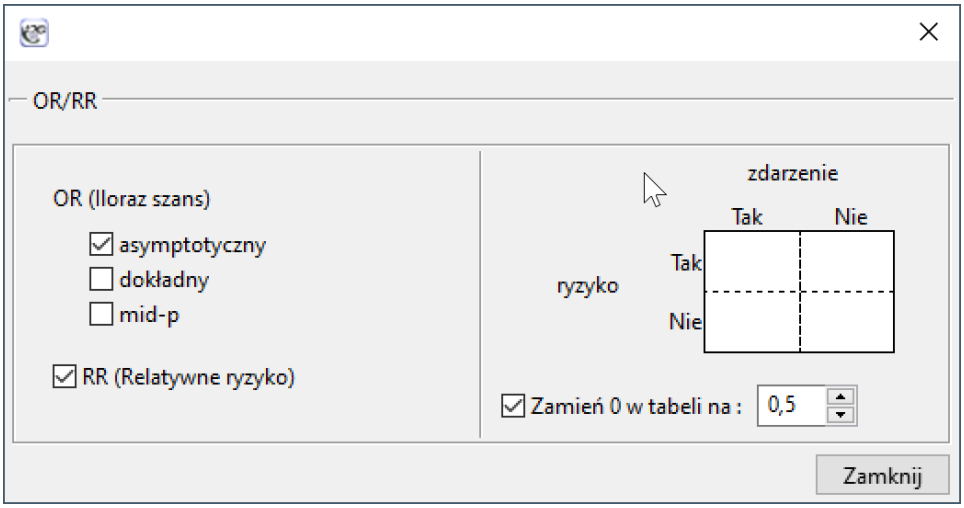
	chorzy	zdrowi
palący	O_{11}	O_{12}
niepalący	O_{21}	O_{22}

Jeśli przeprowadzone badanie jest badaniem typu **przypadek-kontrola** (zwykle retrospektywnym – badacz sam decyduje o wielkości próby, w której badane zjawisko występuje, i próby kontrolnej, wolnej od tego zjawiska), wylicza się **Iloraz Szans**. Jeśli natomiast badanie jest badaniem **kohortowym** (zwykle prospektywnym – badacz dba o to, by struktura występowania zjawiska w próbie była zbliżona do struktury w populacji), wylicza się **Relatywne Ryzyko**.

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Testy nieparametryczne $\rightarrow$ Chi-kwadrat, Fisher, OR/RR, opcja OR/RR... (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Uwaga! Gdy w tabeli danych znajdują się zera, wyliczenie ilorazu szans lub relatywnego ryzyka może być niemożliwe. Aby temu zaradzić, w oknie analizy można zaznaczyć opcję Zamień 0 w tabeli na: (korekta na ciągłość) – powoduje ona dodanie do wszystkich komórek tabeli zadanej wartości.

DEFINICJE MIAR

Iloraz Szans (OR) (*ang. Odds Ratio*) – określa, ile razy szansa wystąpienia badanego zjawiska w grupie narażonej na czynnik ryzyka jest większa (lub mniejsza) niż w grupie nienarażonej. Wyznacza się go dla badań typu przypadek-kontrola.

Relatywne Ryzyko (RR) (*ang. Relative Risk*) – określa, ile razy ryzyko wystąpienia badanego zjawiska w grupie narażonej na czynnik ryzyka jest większe (lub mniejsze) niż w grupie nienarażonej. Wyznacza się go dla badań kohortowych.

Przy interpretacji istotności obu miar często posługujemy się wyznaczonym dla nich przedziałem ufności – badamy wówczas, czy przedział ten zawiera wartość 1 (brak różnicy między grupami). Można też porównać wartość p testu istotności (patrz „Dla zainteresowanych") z przyjętym poziomem istotności α: gdy $p \leq \alpha$, różnica jest istotna statystycznie.

DLA ZAINTERESOWANYCH

Iloraz Szans (tabela 2×2)

Wzory:
Szansę bycia przypadkiem w grupie narażonej na czynnik ryzyka i w grupie nienarażonej wyliczamy ze wzorów:

$$szansa_{ryzyko=tak} = \frac{O_{11}/(O_{11}+O_{12})}{O_{12}/(O_{11}+O_{12})} = \frac{O_{11}}{O_{12}},$$

$$szansa_{ryzyko=nie} = \frac{O_{21}/(O_{21}+O_{22})}{O_{22}/(O_{21}+O_{22})} = \frac{O_{21}}{O_{22}}.$$

Iloraz Szans to:

$$OR = \frac{O_{11}/O_{12}}{O_{21}/O_{22}} = \frac{O_{11}O_{22}}{O_{12}O_{21}}.$$

Test do sprawdzenia istotności OR weryfikuje hipotezę o tym, że szansa na wystąpienie badanego

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

zjawiska jest taka sama w grupie narażonej i nienarażonej na czynnik ryzyka.
Hipotezy:

$$\begin{aligned}\mathcal{H}_0: &\quad OR = 1,\\ \mathcal{H}_1: &\quad OR \neq 1.\end{aligned}$$

Wzory:
Statystyka testowa wyraża się wzorem:

$$z = \frac{\ln(OR)}{SE},$$

gdzie $SE = \sqrt{\frac{1}{O_{11}} + \frac{1}{O_{12}} + \frac{1}{O_{21}} + \frac{1}{O_{22}}}$ – błąd standardowy logarytmu Ilorazu Szans. Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład normalny.

Relatywne Ryzyko (tabela 2×2)

Wzory:
Oszacowane Ryzyko wystąpienia badanego zjawiska wyraża się wzorem $R = \frac{O_{11}+O_{21}}{n}$, natomiast Relatywne Ryzyko:

$$RR = \frac{O_{11}/(O_{11}+O_{12})}{O_{21}/(O_{21}+O_{22})}.$$

Test do sprawdzenia istotności RR weryfikuje hipotezę o tym, że ryzyko wystąpienia badanego zjawiska jest takie samo w grupie narażonej i nienarażonej na czynnik ryzyka.
Hipotezy:

$$\begin{aligned}\mathcal{H}_0: &\quad RR = 1,\\ \mathcal{H}_1: &\quad RR \neq 1.\end{aligned}$$

Wzory:
Statystyka testowa wyraża się wzorem:

$$z = \frac{\ln(RR)}{SE},$$

gdzie $SE = \sqrt{\frac{1}{O_{11}} - \frac{1}{O_{11}+O_{12}} + \frac{1}{O_{21}} - \frac{1}{O_{21}+O_{22}}}$ – błąd standardowy logarytmu Relatywnego Ryzyka. Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład normalny.

Test Z dla dwóch niezależnych proporcji

KIEDY STOSOWAĆ TEN TEST:

Test Z dla dwóch niezależnych proporcji stosujemy w podobnych sytuacjach jak test χ^2 (2×2): gdy mamy 2 niezależne próby, w których badana cecha przyjmuje jeden z dwóch możliwych wyników, a chcemy sprawdzić, czy odsetek (proporcja) wyróżnionego wyniku jest taki sam w obu populacjach, z których pochodzą próby.

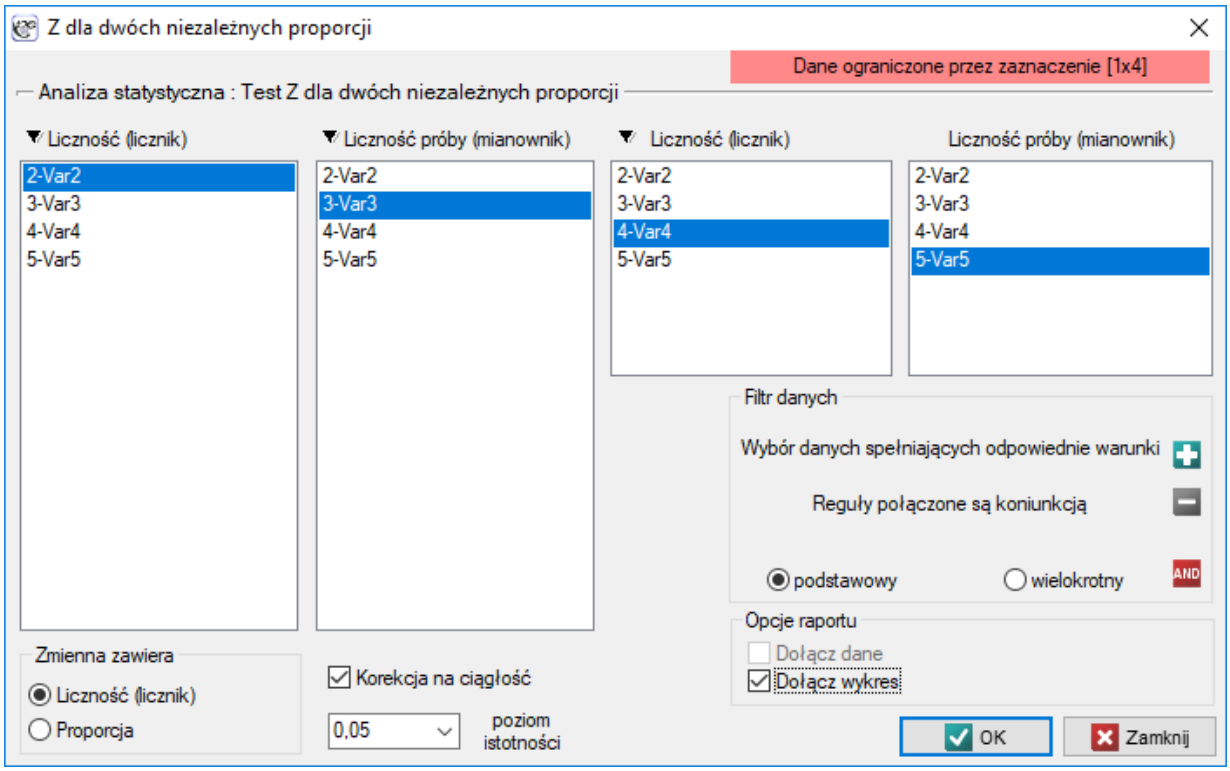
WARUNKI STOSOWANIA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- model niezależny,
- duża liczność.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Z dla dwóch niezależnych proporcji.



Krok 1. Wskaż dane dla obu prób – jako liczbę sukcesów i liczność próby, lub jako proporcję i liczność próby.

Krok 2. Uruchom analizę.

Uwaga! Gdy w arkuszu danych znajduje się więcej informacji, przed rozpoczęciem analizy należy zaznaczyć odpowiedni obszar (dane bez nagłówków). W oknie testu należy dodatkowo wybrać opcję opisującą zawartość zmiennej (liczność (licznik) lub proporcja).

Uwaga! Program oprócz różnicy proporcji wylicza również wskaźniki NNT i/lub NNH. Przedziały ufności dla różnicy dwóch niezależnych proporcji (od wersji PQStat 1.3.0) estymowane są metodą Newcomba-Wilsona[18][127][178], a przedział ufności dla NNT/NNH – jako odwrotność przedziału dla proporcji, metodą Altmana[6].

WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

PRZYKŁAD 16.11. **PRZYKŁAD (16.8) c.d.** *(plik płeć-egzamin.pqs)*
Wiemy, że $\frac{50}{90} = 55.56\%$ kobiet w próbie zdało egzamin pozytywnie i $\frac{20}{80} = 25.00\%$ mężczyzn w próbie zdało egzamin pozytywnie. Chcemy sprawdzić, czy w badanej populacji proporcja zdawalności różni się między płciami.

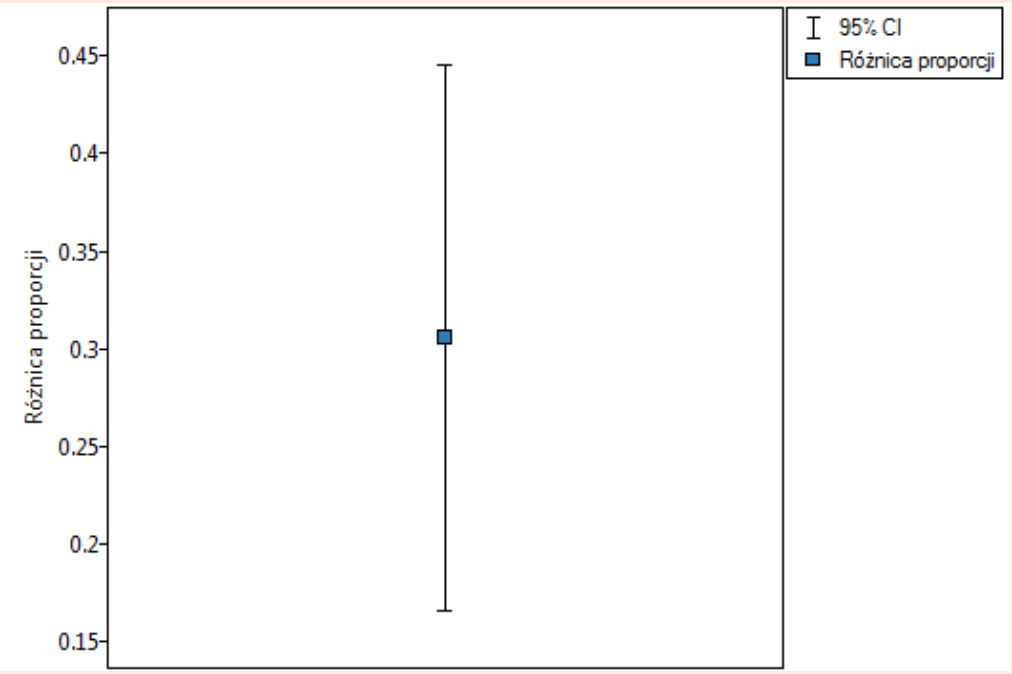
Dane możemy zapisać na dwa sposoby – jako licznik i mianownik dla każdej próby, lub jako proporcja i mianownik dla każdej próby:

dla testu Z (liczność)	licznik - kobiety	mianownik - kobiety	licznik - mężczyźni	mianownik - mężczyźni
	50	90	20	80
dla testu Z (proporcja)	proporcja - kobiety	mianownik - kobiety	proporcja - mężczyźni	mianownik - mężczyźni
dla testu Z (proporcja)	0,555555555556	90	0,25	80

Test Z dla dwóch niezależnych proporcji	
Czas analizy	0,32 sek.
Analizowane zmienne	Var2;Var3;Var4;Var5
Poziom istotności	0,05
Korekcja na ciągłość	Nie
Różnica proporcji	0,305556
-95% CI dla różnicy proporcji	0,158695
+95% CI dla różnicy proporcji	0,433518
NNT	3,272727
-95% CI NNT	2,306711
+95% CI NNT	6,301412
Statystyka Z	4,04047
Wartość p (asymptotyczne)	0,000053

Dane :

v.1	v.2	v.3	v.4
50	90	20	80



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$p = 0.000053$ **(bez poprawki na ciągłość, istotne statystycznie,** $\alpha = 0.05$**)**

Proporcja mężczyzn zdających egzamin pozytywnie różni się od proporcji kobiet zdających egzamin pozytywnie w badanej populacji.

Różnica proporcji wyróżnionych w próbie to 30.56%, a jej 95% przedział ufności $(15.90\%, 43.35\%)$ nie zawiera 0. Do tego samego wniosku prowadzi test Z z poprawką na ciągłość ($p = 0.0001$) – podobnie jak wcześniej wykonane testy: dokładny Fishera, mid-p, chi-kwadrat i chi-kwadrat z poprawką Yatesa. Istotnie częściej ten egzamin zdają kobiety (55.56% z próby) niż mężczyźni (25.00% z próby).

Przykład 16.12. Załóżmy, że choroba ma śmiertelność 100% bez leczenia, a terapia zmniejsza śmiertelność do 50% – jest to wynik 20-letnich badań. Chcemy wiedzieć, ilu osób trzeba leczyć, aby zapobiec w ciągu 20 lat 1 śmierci. W tym celu pobrano dwie 100-osobowe próby z populacji osób chorych: w próbie nieleczonych wszyscy 100 pacjentów umiera, w próbie leczonych przeżywa 50 ze 100 pacjentów.

Chorzy – nie leczeni		Chorzy – leczeni	
liczność licznik	liczność próby (mianownik)	liczność licznik	liczność próby (mianownik)
100	100	50	100

Test Z dla dwóch niezależnych proporcji	
Czas analizy	0,05 sek.
Analizowane zmienne	Var2;Var3;Var4;Var5
Poziom istotności	0,05
Korekcja na ciągłość	Tak
Różnica proporcji	0,5
-95% CI dla różnicy proporcji	0,396962
+95% CI dla różnicy proporcji	0,596168
NNT	2
-95% CI NNT	1,677378
+95% CI NNT	2,519135
Statystyka Z	8,001666
Wartość p (asymptotyczne)	<0.000001

Dane :

v.1	v.2	v.3	v.4
100	100	50	100

$p < 0.000001$, $NNT = 2$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Różnica między proporcjami jest istotna statystycznie, a wskaźnik NNT wynosi 2 – oznacza to, że leczenie 2 chorych przez 20 lat zapobiegnie 1 śmierci.

Wyliczony 95% przedział ufności dla NNT należy zaokrąglić do wartości całkowitych, co daje NNT od 2 do 3 chorych.

Przykład 16.13. Wartość pewnej różnicy proporcji w badaniu porównującym skuteczność leku 1 vs lek 2 wynosiła (95%CI) $= -0.08$ (-0.27 do 0.11). Ta ujemna różnica proporcji sugeruje, że lek 1 był mniej skuteczny niż lek 2, a jego zastosowanie naraziło chorych na ryzyko.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

NNH(95%CI)$= 2.5$ **(**3.7 **do** ∞ **do NNT** 9.1**)**

Wynik nie jest istotny statystycznie – ponieważ przedział ufności dla NNH zawiera nieskończoność i przechodzi z NNH do NNT, nie mamy podstaw, by mówić o istotnym zwiększeniu ryzyka[6].

DEFINICJE MIAR

NNT (*ang. number needed to treat*) – wskaźnik stosowany w medycynie, oznacza liczbę pacjentów, których trzeba poddać leczeniu przez określony czas, aby wyleczyć jedną osobę, która w innych okolicznościach nie wyzdrowiałaby.

NNH (*ang. number needed to harm*) – wskaźnik stosowany w medycynie, oznacza liczbę pacjentów, których narażenie na ryzyko przez określony czas powoduje uszczerbek na zdrowiu u jednej osoby, która w innych okolicznościach nie doznałaby uszczerbku.

DLA ZAINTERESOWANYCH

Definicja: Niech $p_1 = \frac{m_1}{n_1}$ i $p_2 = \frac{m_2}{n_2}$ oznaczają wyróżnione proporcje w próbach o licznościach n_1, n_2, gdzie m_1, m_2 – liczności wyniku wyróżnionego w pierwszej i drugiej próbie.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad P_1 = P_2, \\ \mathcal{H}_1: &\quad P_1 \neq P_2, \end{aligned}$$

gdzie P_1, P_2 – wyróżnione proporcje w pierwszej i drugiej populacji.

Wzory: Statystyka testowa ma postać:

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

gdzie $p = \frac{m_1+m_2}{n_1+n_2}$.

Zmodyfikowana o poprawkę na ciągłość statystyka testowa ma postać:

$$Z = \frac{p_1 - p_2 - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Statystyka bez korekcji na ciągłość, jak i z tą korekcją, ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wzory – NNT/NNH: NNT wyliczane jest ze wzoru:

$$NNT = \frac{1}{|p_1 - p_2|},$$

i jest cytowane wtedy, gdy różnica $p_1 - p_2$ jest dodatnia. NNH wyliczane jest w ten sam sposób co NNT, ale cytowane wtedy, gdy różnica $p_1 - p_2$ jest ujemna.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Interpretacja: Im węższy przedział ufności, tym bardziej precyzyjne oszacowanie. Jeśli przedział ufności dla różnicy proporcji zawiera 0 (a dla NNT/NNH – zawiera ∞), jest to wskazanie, by dany wynik traktować jako nieistotny statystycznie.

Metoda estymacji przedziału ufności – przedziały ufności dla różnicy dwóch niezależnych proporcji, od wersji PQStat 1.3.0, estymowane są metodą Newcomba-Wilsona (Bender 2001[18], Newcombe 1998[127], Wilson 1927[178]); wcześniej stosowano metodę Walda. Metoda Walda jest odpowiednia dla dużych prób i różnicy proporcji dalekiej od 0 lub 1 – dla małych prób i wartości bliskich tym skrajnym może prowadzić do wyników niewiarygodnych (Newcombe 1998[127], Miettinen 1985[124], Beal 1987[17], Wallenstein 1997[169]). Metoda Newcomba-Wilsona jest odpowiednia również dla skrajnych wartości proporcji.

Test McNemara, test wewnętrznej symetrii Bowkera

KIEDY STOSOWAĆ TE TESTY:

Testy stosujemy, gdy tę samą cechę mierzymy dwukrotnie u tych samych osób (np. opinię przed i po jakimś wydarzeniu) i chcemy sprawdzić, czy wyniki się zmieniły. **Test McNemara** *(ang. McNemar test)*, McNemar (1947)[121], stosujemy, gdy badana cecha ma tylko 2 kategorie (oznaczane **(+)** i **(−)**). **Test wewnętrznej symetrii Bowkera** *(ang. Bowker test of internal symmetry)*, Bowker (1948)[24], jest jego rozszerzeniem na cechy o więcej niż dwóch kategoriach ($c > 2$). Oba testy można wyliczać na podstawie danych surowych albo z gotowej tabeli kontyngencji.

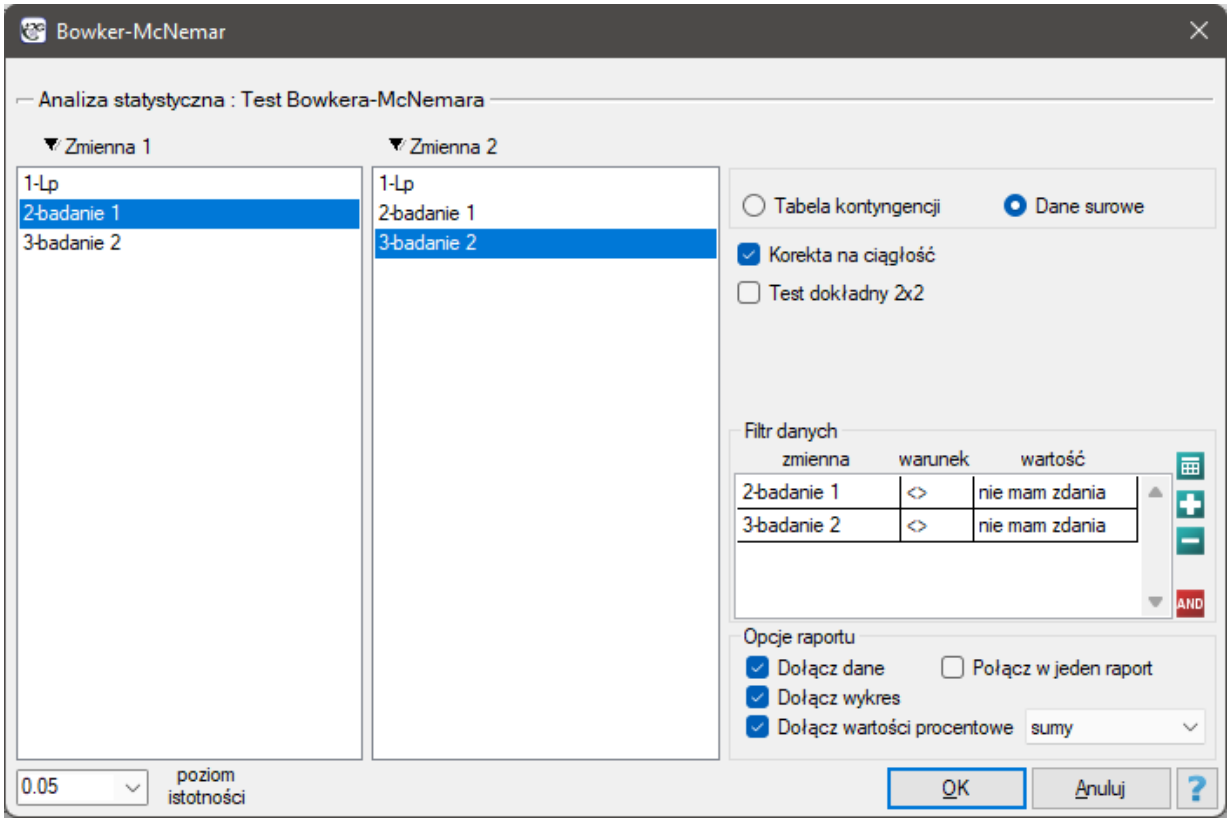
WARUNKI STOSOWANIA

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- model zależny.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Bowker-McNemar (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż dwie zmienne odpowiadające pierwszemu i drugiemu pomiarowi tej samej cechy.

Krok 2. Uruchom analizę.

Uwaga! Powszechną zasadą ważności asymptotycznego testu McNemara jest **warunek Rufibach**: liczba niezgodnych par powinna być większa niż 10 ($O_{12} + O_{21} \geq 10$)[147]. Gdy warunek ten nie jest spełniony, lepiej oprzeć wniosek na dokładnej wartości prawdopodobieństwa testu (opartej na rozkładzie dwumianowym) lub na jej mniej konserwatywnej wersji, teście mid-p McNemara[56] – obie są liczone automatycznie razem z testem podstawowym.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 16.14. (plik opinia.pqs)
Przeprowadzono 2 badania opinii studentów na temat określonego wykładowcy akademickiego. Oba badania pozwalały ocenić wykładowcę negatywnie, pozytywnie, lub wybrać odpowiedź neutralną – nie mam zdania. Oba badania przeprowadzono na tej samej próbie 250 studentów, z tym że pierwsze badanie wykonano dzień przed egzaminem z przedmiotu prowadzonego przez ocenianego wykładowcę, a drugie dzień po egzaminie. Poniżej przedstawiono fragment danych w postaci surowej oraz całość danych w postaci tabeli kontyngencji. Chcemy zbadać, czy obydwa badania dają podobne wyniki.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

badanie 1	badanie 2
pozytywnie	negatywnie
negatywnie	nie mam zdania
poztywnie	negatywnie
negatywnie	negatywnie

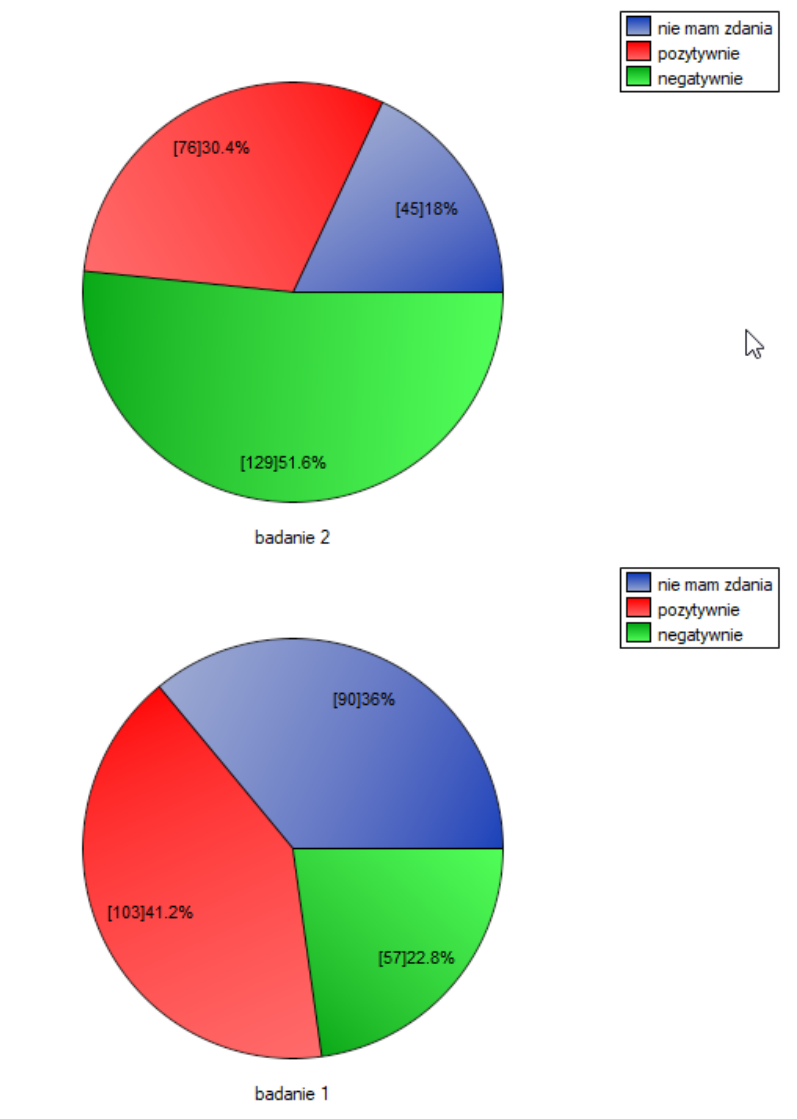
	negatywnie	pozytywnie	nie mam zdania
negatywnie	50	4	3
pozytywnie	44	54	5
nie mam zdania	35	18	37

Test Bowkera-McNemara	
Analizowane zmienne	badanie 1
	badanie 2
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Korekta na ciągłość	Tak
Liczność / ilość par	250
Iloraz szans	NA
-95% CI dla ilorazu szans	NA
+95% CI dla ilorazu szans	NA
Bowker test	
Statystyka Chi-kwadrat	63.2378
Stopnie swobody	3
Wartość p	<0.0001

Dane :			↙badanie 2
↓badanie 1	nie mam zdа	pozytywnie	negatywnie
nie mam zdania	37	18	35
pozytywnie	5	54	44
negatywnie	3	4	50
Razem	45	76	129

% sumy :			↙badanie 2
↓badanie 1	nie mam zdа	pozytywnie	negatywnie
nie mam zdania	14.8%	7.2%	14%
pozytywnie	2%	21.6%	17.6%
negatywnie	1.2%	1.6%	20%

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p < 0.0001$ **(test Bowkera, istotne statystycznie,** $\alpha = 0.05$**)**

Opinie studentów na temat wykładowcy zmieniły się pomiędzy oboma badaniami.

Z tabeli wynika, że istotnie więcej było tych studentów, którzy zmienili swoją opinię na negatywną po egzaminie, niż tych, którzy zmienili ją na pozytywną – wielu studentów oceniających wykładowcę pozytywnie przed egzaminem po egzaminie nie wyrażało już takiego zdania.

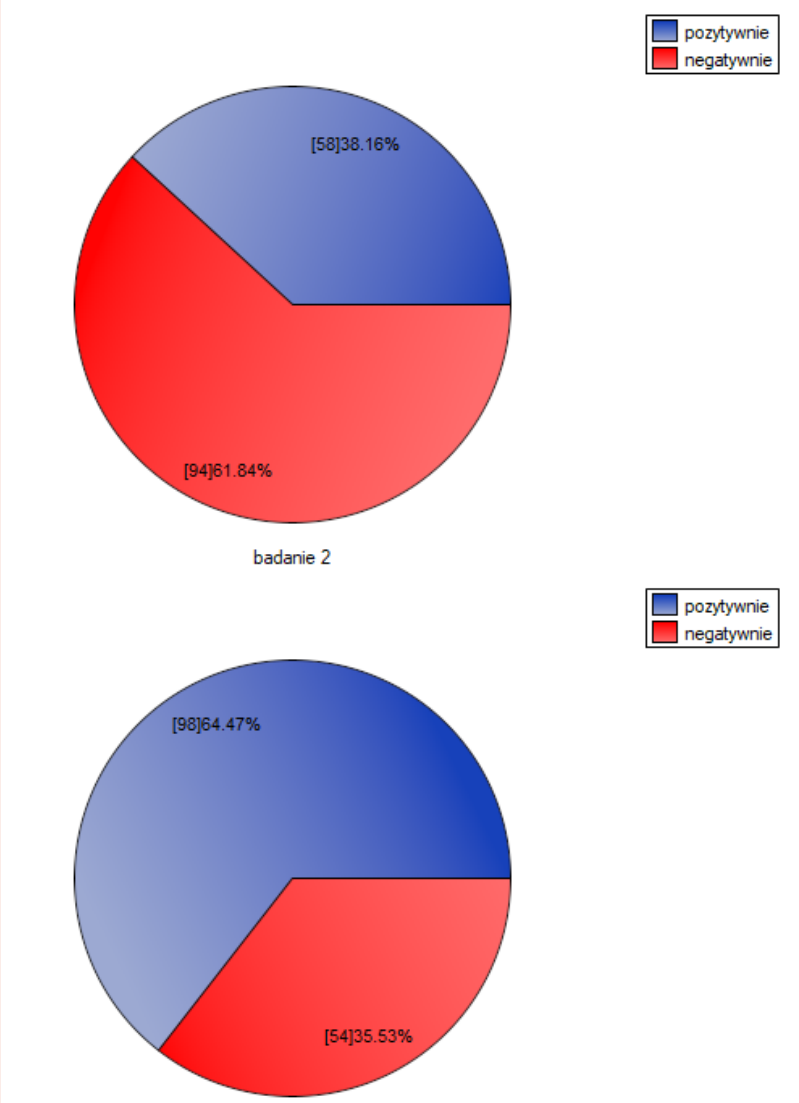
Gdybyśmy ograniczyli badanie do osób mających zdefiniowany pogląd na temat wykładowcy (oceniających go tylko pozytywnie lub negatywnie), moglibyśmy wykorzystać prostszy test McNemara:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test Bowkera-McNemara	
Analizowane zmienne	badanie 1
	badanie 2
Filtr danych	badanie 1<>nie mam zdania
	i badanie 2<>nie mam zdani
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Korekta na ciągłość	Tak
Liczność / ilość par	152
Iloraz szans	11
-95% CI dla ilorazu szans	3.9525
+95% CI dla ilorazu szans	30.6139
McNemar test	
Statystyka Chi-kwadrat	31.6875
Stopnie swobody	1
Wartość p	<0.0001
Warunek Rufibach	spełniony

Dane :		✓badanie 2	
↓badanie 1	pozytywnie	negatywnie	Razem
pozytywnie	54	44	98
negatywnie	4	50	54
Razem	58	94	152

% sumy :		✓badanie 2
↓badanie 1	pozytywnie	negatywnie
pozytywnie	35.526%	28.947%
negatywnie	2.632%	32.895%



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$p < 0.0001$ **(test McNemara, istotne statystycznie, $\alpha = 0.05$)**

Wśród studentów z jednoznaczną opinią, jej rozkład (pozytywna/negatywna) zmienił się pomiędzy oboma badaniami.

Istotnie więcej było tych studentów, którzy zmienili swoją opinię na negatywną po egzaminie, niż tych, którzy zmienili ją na pozytywną. Szansa zmiany opinii z pozytywnej (przed egzaminem) na negatywną (po egzaminie) jest jedenaście razy większa ($\frac{44}{4}$) niż szansa zmiany w przeciwną stronę ($\frac{4}{44}$).

DEFINICJE MIAR

Iloraz szans (OR) (*ang. odds ratio*) – jeśli badanie tej samej cechy na tych samych obiektach przeprowadzono dwukrotnie, iloraz szans opisuje, ile razy szansa zmiany wyniku w jedną stronę (np. z **(+)** na **(−)**) jest większa lub mniejsza od szansy zmiany w stronę przeciwną.

DLA ZAINTERESOWANYCH

Oznaczenia:
Oznaczenia tabeli kontyngencji 2×2 dla McNemara:

Liczności obserwowane O_{ij}		$X^{(2)}$ **(+)**	$X^{(2)}$ **(−)**	$X^{(2)}$ **Suma**
$X^{(1)}$	**(+)**	O_{11}	O_{12}	$O_{11} + O_{12}$
	(−)	O_{21}	O_{22}	$O_{21} + O_{22}$
	Suma	$O_{11} + O_{21}$	$O_{12} + O_{22}$	$n = O_{11} + O_{12} + O_{21} + O_{22}$

Test McNemara – hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad O_{12} = O_{21}, \\ \mathcal{H}_1 &: \quad O_{12} \neq O_{21}. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$\chi^2 = \frac{(O_{12} - O_{21})^2}{O_{12} + O_{21}}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z jednym stopniem swobody.

Poprawka na ciągłość testu McNemara jest testem bardziej konserwatywnym od podstawowego (trudniej odrzuca hipotezę zerową). Statystyka testowa z poprawką ma postać:

$$\chi^2 = \frac{(|O_{12} - O_{21}| - 1)^2}{O_{12} + O_{21}}.$$

Wzory (iloraz szans):
Szansa zmiany wyniku z **(+)** na **(−)** wynosi O_{12}, a z **(−)** na **(+)** wynosi O_{21}. Iloraz szans wyliczamy ze wzoru:

$$OR = \frac{O_{12}}{O_{21}}.$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przedział ufności dla ilorazu szans buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{1}{O_{12}} + \frac{1}{O_{21}}}.$$

Dla prób o niewielkich licznościach można dodatkowo wyznaczyć dokładny zakres przedziału ufności dla ilorazu szans[105].

Test wewnętrznej symetrii Bowkera rozszerza test McNemara na tabele $c \times c$ ($c > 2$ kategorii):

Liczności obserwowane O_{ij}		$X^{(2)}$				
		$X_1^{(2)}$	$X_2^{(2)}$	...	$X_c^{(2)}$	Suma
$X^{(1)}$	$X_1^{(1)}$	O_{11}	O_{12}	...	O_{1c}	$\sum_{j=1}^{c} O_{1j}$
	$X_2^{(1)}$	O_{21}	O_{22}	...	O_{2c}	$\sum_{j=1}^{c} O_{2j}$
	...	...	...	...	...	...
	$X_c^{(1)}$	O_{c1}	O_{c2}	...	O_{cc}	$\sum_{j=1}^{c} O_{cj}$
	Suma	$\sum_{i=1}^{c} O_{i1}$	$\sum_{i=1}^{c} O_{i2}$	...	$\sum_{i=1}^{c} O_{ic}$	$n = \sum_{i=1}^{c}\sum_{j=1}^{c} O_{ij}$

Hipotezy (gdzie $j \neq i$, $j, i \in \{1, 2, ..., c\}$, a O_{ij} i O_{ji} to liczności symetrycznych par):

$$\begin{aligned} \mathcal{H}_0 &: \quad O_{ij} = O_{ji}, \\ \mathcal{H}_1 &: \quad O_{ij} \neq O_{ji} \text{ dla przynajmniej jednej pary } O_{ij}, O_{ji}. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^{c} \sum_{j>i} \frac{(O_{ij} - O_{ji})^2}{O_{ij} + O_{ji}}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody wyliczaną według wzoru $df = \frac{c(c-1)}{2}$.

Test Z dla dwóch zależnych proporcji

KIEDY STOSOWAĆ TEN TEST:

Test Z dla dwóch zależnych proporcji stosujemy, gdy tę samą cechę o dwóch możliwych odpowiedziach (np. tak/nie, zgadzam się/nie zgadzam się) mierzymy dwukrotnie u tych samych osób, i chcemy sprawdzić, czy odsetek jednej z odpowiedzi zmienił się między pomiarami (np. przed i po jakimś wydarzeniu). Jest to podobna sytuacja do testu McNemara.

WARUNKI STOSOWANIA

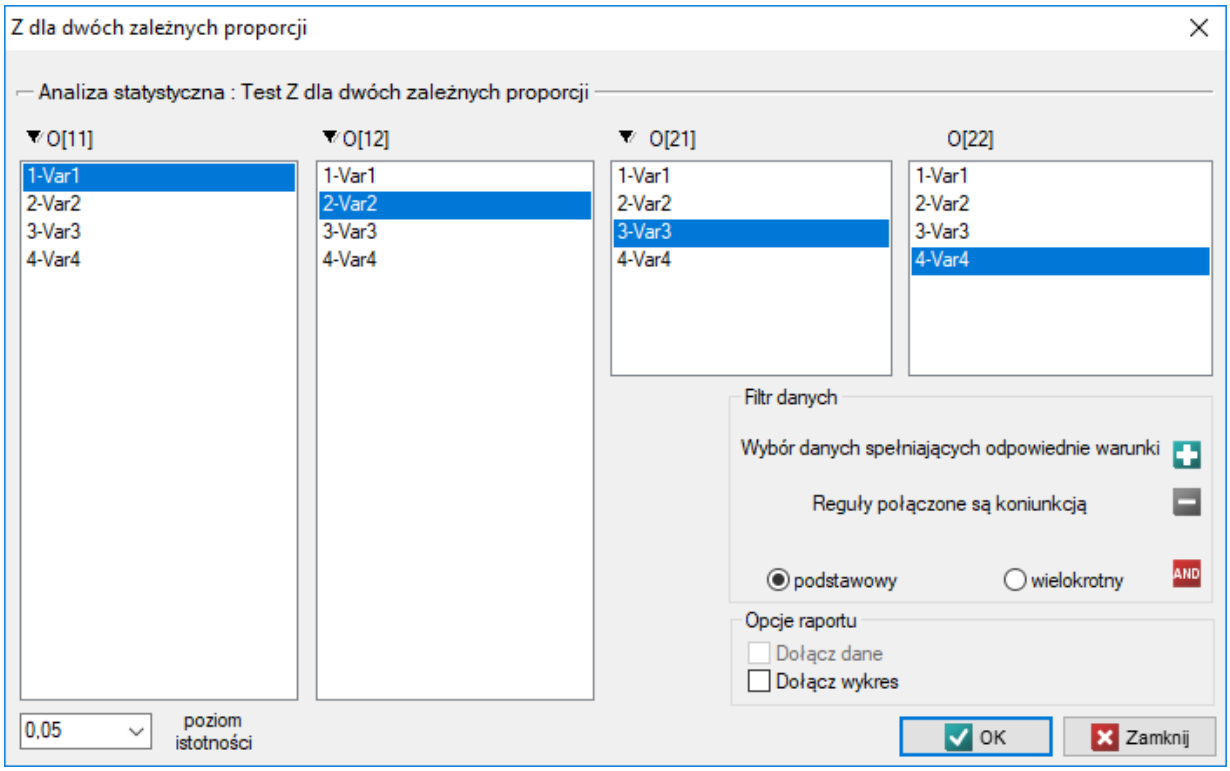
- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- model zależny,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– duża liczność.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Z dla dwóch zależnych proporcji.



Krok 1. Wskaż dwie zmienne odpowiadające pierwszemu i drugiemu pomiarowi tej samej cechy.

Krok 2. Uruchom analizę.

Uwaga! Przedział ufności dla różnicy dwóch zależnych proporcji estymowany jest w oparciu o metodę Newcomba-Wilsona.

WYKONANIE I INTERPRETACJA

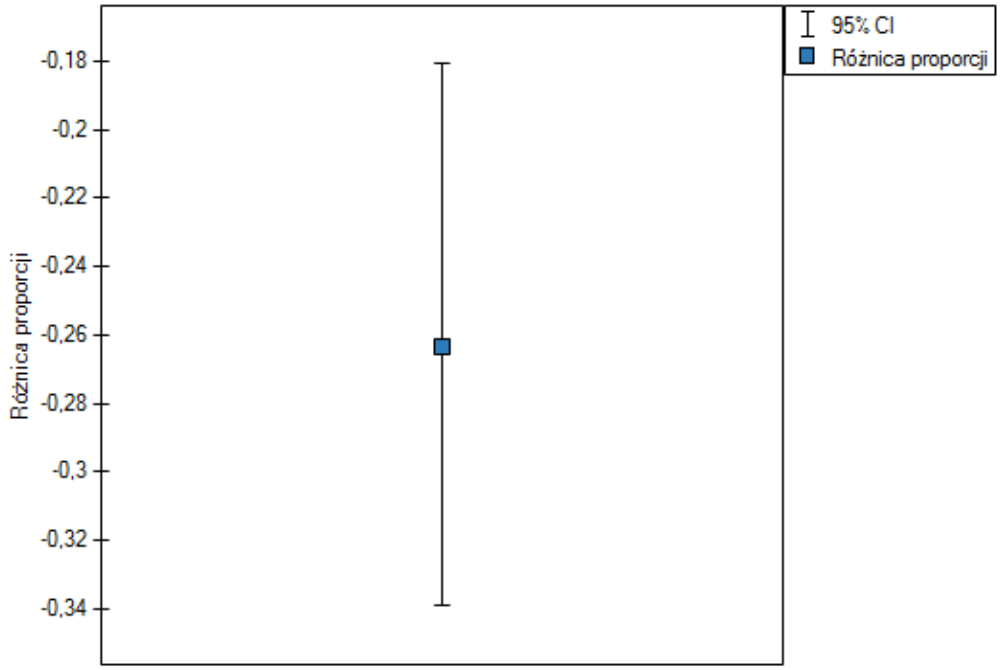
Przykład 16.15. **Przykład (16.14) c.d.** *(plik opinia.pqs)*
Gdy ograniczymy nasze badanie do osób mających zdefiniowany pogląd na temat wykładowcy (oceniających go tylko pozytywnie lub negatywnie), uzyskamy 152 takich studentów. Wśród nich: 50 studentów oceniało wykładowcę pozytywnie zarówno przed, jak i po egzaminie; 4 studentów oceniało go dobrze przed egzaminem, ale źle po egzaminie; 44 studentów oceniało go źle przed egzaminem, a dobrze po egzaminie; 54 studentów oceniało go negatywnie zarówno przed, jak i po egzaminie. Wiemy więc, że $\frac{50+4}{152} = 35.53\%$ studentów przed egzaminem wyrażało negatywną opinię, a po egzaminie odpowiedni procent wynosił $\frac{50+44}{152} = 61.84\%$. Chcemy sprawdzić, czy proporcja negatywnych ocen zmieniła się między pomiarami.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Test Z dla dwóch zależnych proporcji	
Czas analizy	0,11 sek.
Analizowane zmienne	Var1;Var2;Var3;Var4
Poziom istotności	0,05
Korekcja na ciągłość	Nie
O[11]+O[12]	54
O[11]+O[21]	94
Proporcja 1	0,355263
Proporcja 2	0,618421
Różnica proporcji	-0,263158
-95% CI dla różnicy proporcji	-0,338845
+95% CI dla różnicy proporcji	-0,180717
Statystyka Z	-5,773503
Wartość p (asymptotyczne)	<0.000001

Dane :

v.1	v.2	v.3	v.4
50	4	44	54



$p = 0.0001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Proporcja negatywnych ocen wykładowcy przed egzaminem różni się istotnie statystycznie od proporcji negatywnych ocen po egzaminie.

Różnica proporcji wyróżnionych w próbie to 26.32%, a jej 95% przedział ufności $(18.07\%, 33.88\%)$ nie zawiera 0 – do tego samego wniosku prowadzi test McNemara. Po egzaminie wykładowca jest istotnie częściej oceniany negatywnie.

DLA ZAINTERESOWANYCH

Definicja: Oznaczenia tabeli kontyngencji dla dwóch zależnych pomiarów tej samej cechy X (pierwszy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

pomiar $X^{(1)}$, drugi pomiar $X^{(2)}$), każdy o dwóch możliwych odpowiedziach **(+)** i **(−)**:

Liczności obserwowane O_{ij}		$X^{(2)}$		
		(+)	**(−)**	**Suma**
$X^{(1)}$	**(+)**	O_{11}	O_{12}	$O_{11}+O_{12}$
	(−)	O_{21}	O_{22}	$O_{21}+O_{22}$
	Suma	$O_{11}+O_{21}$	$O_{12}+O_{22}$	$n=O_{11}+O_{12}+O_{21}+O_{22}$

gdzie O_{11} – liczba osób z odpowiedzią **(+)** w obu pomiarach, O_{12} – liczba osób, które zmieniły odpowiedź z **(+)** na **(−)**, O_{21} – liczba osób, które zmieniły odpowiedź z **(−)** na **(+)**, O_{22} – liczba osób z odpowiedzią **(−)** w obu pomiarach.

Dla tych danych wyliczamy wyróżnione proporcje $p_1=\frac{O_{11}+O_{12}}{n}$ i $p_2=\frac{O_{11}+O_{21}}{n}$.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0:&\quad P_1-P_2=0,\\ \mathcal{H}_1:&\quad P_1-P_2\neq 0,\end{aligned}$$

gdzie P_1, P_2 – wyróżnione proporcje w populacji, dla pierwszego i drugiego pomiaru.

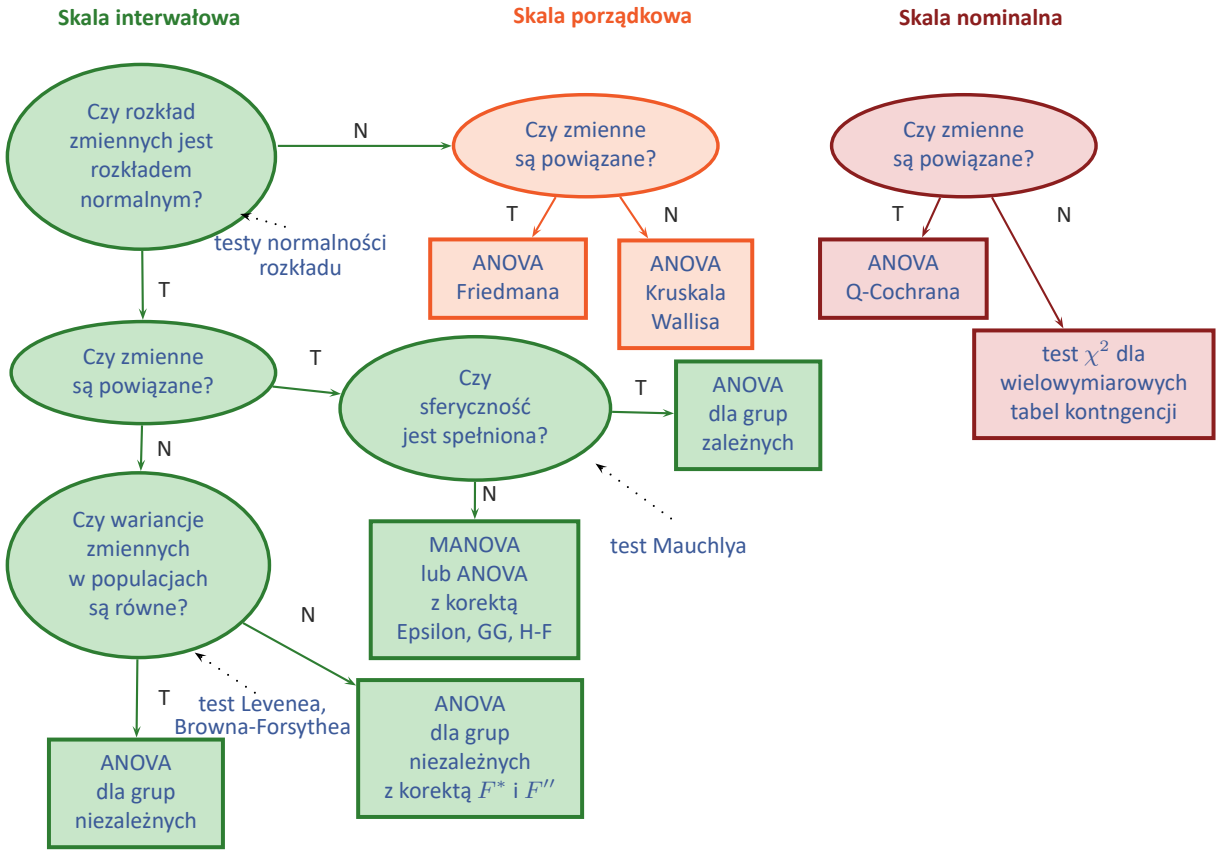
Wzory: Statystyka testowa ma postać:

$$Z=\frac{p_1-p_2}{\sqrt{O_{21}+O_{12}}}\cdot n,$$

i ma asymptotycznie (dla dużych liczności) rozkład normalny.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

17 PORÓWNANIE - WIĘCEJ NIŻ 2 GRUPY



Uwaga!
Zaproponowany schemat wyboru testów do porównania wielu grup nie jest jedynym możliwym schematem i nie zawiera wszystkich zaproponowanych w programie testów przeznaczonych do tego porównania.

Uwaga!
Jednoczesne porównywanie większej liczby grup NIE może być zastąpione wielokrotnym wykonywaniem testów dla porównania dwóch grup. Wynika to z konieczności kontrolowania błędu pierwszego rodzaju α. Wybierając poziom α i stosując k-krotnie wybrany test dla porównania dwóch grup zwiększylibyśmy znacznie założony poziom α. Tego błędu unikniemy, stosując odpowiednie testy ANOVA (*ang. Analysis of Variance*) oraz dedykowane im kontrasty, czy też testy POST-HOC.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

17.1 TESTY PARAMETRYCZNE

ANOVA dla grup niezależnych

KIEDY STOSOWAĆ TEN TEST:

Jednoczynnikowa analiza wariancji (ANOVA) dla grup niezależnych, zaproponowana przez Ronalda Fishera, pozwala sprawdzić, czy średnia wartość badanej cechy jest taka sama w co najmniej trzech niezależnych od siebie populacjach (np. grupach pacjentów, wariantach produktu, firmach). Test wskazuje jedynie, czy takie różnice w ogóle istnieją – nie mówi, między którymi konkretnie grupami występują; do tego służą testy POST-HOC, opisane w kolejnym rozdziale.

WARUNKI STOSOWANIA

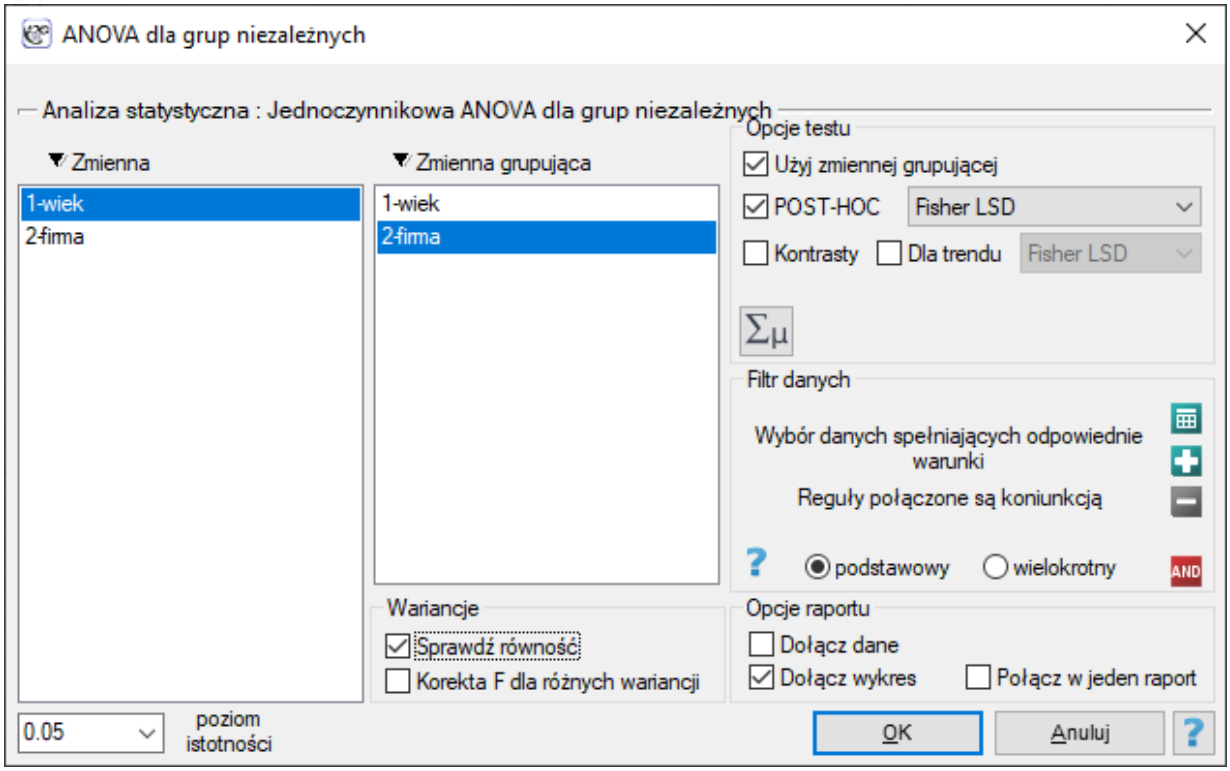
- pomiar na skali interwałowej,
- normalność rozkładu badanej zmiennej w każdej populacji,
- model niezależny,
- równość wariancji badanej zmiennej we wszystkich populacjach.

Gdy założenie równości wariancji nie jest spełnione, stosuje się ANOVA dla grup niezależnych z korektą F^* i F''.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → ANOVA dla grup niezależnych (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienną grupującą (np. nazwę firmy) oraz zmienną wynikową (np. wiek pracownika).

Krok 2. Uruchom analizę.

WYKONANIE I INTERPRETACJA

Przykład 17.1. (plik wiek ANOVA.pqs)
W pewnym doświadczeniu bierze udział 150 osób wybranych w sposób losowy z populacji pracowników 3 różnych firm przewozowych. Z każdej firmy do próby wybrano 50 osób. Przed przystąpieniem do eksperymentu należy sprawdzić czy średni wiek pracowników tych firm jest podobny, od tego bowiem zależeć będzie kolejny etap eksperymentu. Wiek każdego uczestnika eksperymentu zapisano w latach.
Wiek (przewoźnik 1): 27, 33, 25, 32, 34, 38, 31, 34, 20, 30, 30, 27, 34, 32, 33, 25, 40, 35, 29, 20, 18, 28, 26, 22, 24, 24, 25, 28, 32, 32, 33, 32, 34, 27, 34, 27, 35, 28, 35, 34, 28, 29, 38, 26, 36, 31, 25, 35, 41, 37
Wiek (przewoźnik 2): 38, 34, 33, 27, 36, 20, 37, 40, 27, 26, 40, 44, 36, 32, 26, 34, 27, 31, 36, 36, 25, 40, 27, 30, 36, 29, 32, 41, 49, 24, 36, 38, 18, 33, 30, 28, 27, 26, 42, 34, 24, 32, 36, 30, 37, 34, 33, 30, 44, 29
Wiek (przewoźnik 3): 34, 36, 31, 37, 45, 39, 36, 34, 39, 27, 35, 33, 36, 28, 38, 25, 29, 26, 45, 28, 27, 32, 33, 30, 39, 40, 36, 33, 28, 32, 36, 39, 32, 39, 37, 35, 44, 34, 21, 42, 40, 32, 30, 23, 32, 34, 27, 39, 37, 35

Przed przystąpieniem do analizy ANOVA potwierdzono normalność rozkładu danych. W oknie analizy sprawdzono również założenie równości wariancji, uzyskując w obydwu testach $p > 0.05$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednoczynnikowa ANOVA dla grup niezależnych	
Czas analizy	0.10 sek.
Analizowane zmienne	wiek;firma
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Zmienna grupująca	firma
ANOVA dla grup niezależnych	
Eta-kwadrat	0.0692
Całkowita suma kwadratów (SS[T])	5151.8933
Suma kwadratów między grupami (SS[BG])	356.4133
Suma kwadratów wewnątrz grup (SS[WG])	4795.48
Średnia kwadratów między grupami (MS[BG])	178.2067
Średnia kwadratów wewnątrz grup (MS[WG])	32.6223
Stopnie swobody między grupami (df[BG])	2
Stopnie swobody wewnątrz grup (df[WG])	147
Całkowite stopnie swobody (df[T])	149
Statystyka F	5.4627
Wartość p	0.0051
Równość wariancji - Brown-Forsythe	
Statystyka F	0.8484
Wartość p	0.4302
Równość wariancji - Levene	
Statystyka F	0.8671
Wartość p	0.4223

$p = 0.0051$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Średni wiek pracowników trzech badanych firm przewozowych różni się istotnie statystycznie.

Na podstawie samego wyniku ANOVA nie wiemy jednak, które konkretnie firmy różnią się między sobą pod względem wieku pracowników – odpowiedzi na to pytanie udziela analiza POST-HOC, opisana w kolejnym rozdziale.

Dokładny opis danych możemy przedstawić, wybierając w oknie analizy statystyki opisowe 

Podsumowanie			
Grupa	przewoźnik 1	przewoźnik 2	przewoźnik 3
Liczność	50	50	50
Średnia arytmetyczna	30.26	32.68	33.98
Błąd standardowy średniej	0.74	0.8992	0.7754
Odchylenie standardowe	5.2326	6.3582	5.4828
-95% CI dla średniej grupy	28.7729	30.873	32.4218
+95% CI dla średniej grupy	31.7471	34.487	35.5382

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DEFINICJE MIAR

Cząstkowa η^2 – miara wielkości efektu określająca, jaka część zmienności badanej cechy wynika z przynależności do porównywanych grup (czynnika), a jaka pozostaje niewyjaśniona. Im wyższa wartość, tym silniejszy jest związek pomiędzy przynależnością do grupy a wynikiem badanej cechy.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_k, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, ..., k), \end{aligned}$$

gdzie μ_1,μ_2,...,μ_k – średnie badanej zmiennej w populacjach, z których pobrano próby.

Wzory:

Statystyka testowa ma postać:

$$F = \frac{MS_{BG}}{MS_{WG}},$$

gdzie:

$MS_{BG} = \dfrac{SS_{BG}}{df_{BG}}$ – średnia kwadratów między grupami,

$MS_{WG} = \dfrac{SS_{WG}}{df_{WG}}$ – średnia kwadratów wewnątrz grup,

$SS_{BG} = \displaystyle\sum_{j=1}^{k} \frac{\left(\sum_{i=1}^{n_j} x_{ij}\right)^2}{n_j} - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}\right)^2}{N}$ – suma kwadratów między grupami,

$SS_{WG} = SS_T - SS_{BG}$ – suma kwadratów wewnątrz grup,

$SS_T = \left(\displaystyle\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}^2\right) - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}\right)^2}{N}$ – całkowita suma kwadratów,

$df_{BG} = k - 1$ – stopnie swobody (między grupami),

$df_{WG} = df_T - df_{BG}$ – stopnie swobody (wewnątrz grup),

$df_T = N - 1$ – całkowite stopnie swobody,

$N = \sum_{j=1}^{k} n_j$,

n_j – liczności prób dla $(j = 1, 2, ...k)$,

x_{ij} – wartości zmiennej w próbach dla $(i = 1, 2, ...n_j)$, $(j = 1, 2, ...k)$.

Statystyka ta podlega rozkładowi F Snedecora z df_{BG} i df_{WG} stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Cząstkowa η^2

$$\eta^2 = \frac{SS_{BG}}{SS_{BG} + SS_{res}}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Kontrasty i testy POST-HOC

KIEDY STOSOWAĆ TEN TEST:

Wynik ANOVA informuje jedynie o tym, czy pomiędzy badanymi populacjami istnieją istotne statystycznie różnice – nie wskazuje, które konkretnie grupy się różnią. Odpowiedzi na to pytanie dostarczają porównania wielokrotne: **kontrasty** (gdy porównania są zaplanowane z góry, zwykle tylko wybrane, tzw. a'priori) lub **testy POST-HOC** (gdy różnic szukamy dopiero po wykonanej ANOVA, zwykle pomiędzy wszystkimi możliwymi parami grup).

WARUNKI STOSOWANIA

- wcześniej wykonana ANOVA dla grup niezależnych wykazała istotną statystycznie różnicę,
- wybór metody zależy od spełnienia założenia równości wariancji: testy LSD Fishera, Scheffego i Tukeya stosuje się, gdy wariancje nie różnią się istotnie.

Dla grup o istotnie różnych wariancjach stosuje się metody opisane w rozdziale ANOVA dla grup niezależnych z korektą F^* i F''.

JAK USTAWIĆ ANALIZĘ

Testy POST-HOC ustawiamy w tym samym oknie opcji, w którym uruchamiana jest ANOVA dla grup niezależnych.

Krok 1. Wznów analizę przyciskiem  .

Krok 2. W oknie opcji testu wskaż metodę porównań wielokrotnych, np. Tukey HSD.

Krok 3. Zaznacz opcję dołączenia wykresu.

Krok 4. Uruchom analizę.

WYKONANIE I INTERPRETACJA

Grupy jednorodne

Dla każdego testu post-hoc budowane są grupy jednorodne. Każda grupa jednorodna przedstawia zbiór grup, które nie różnią się od siebie w sposób istotny statystycznie. Na przykład, załóżmy, że podzieliliśmy badanych na sześć grup odnośnie statusu palenia: Nonsmokers (NS), Passive smokers (PS), Noninhaling smokers (NI), Light smokers (LS), Moderate smokers (MS), Heavy smokers (HS) i badamy dla nich parametry wydechowe. W przeprowadzonej analizie typu ANOVA uzyskaliśmy istotne statystycznie różnice w parametrach wydechowych pomiędzy badanymi grupami. Chcąc wskazać, które grupy różnią się istotnie, a które nie, wykonujemy testy typu post-hoc. W rezultacie oprócz tabeli z wynikami poszczególnych par porównań i podanej istotności statystycznej w postaci wartości p:

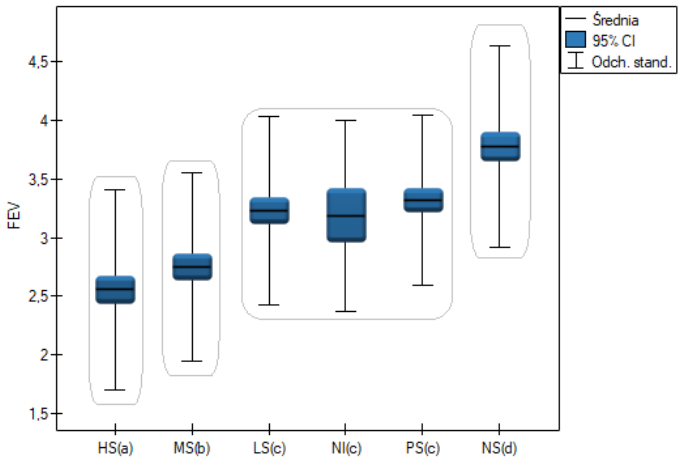
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Statystyk	HS	MS	LS	NI	PS	NS
HS		2,434752	8,276928	4,920204	9,41745	15,066567
MS	2,434752		5,842176	3,380331	6,982697	12,631814
LS	8,276928	5,842176		0,314585	1,140522	6,789639
NI	4,920204	3,380331	0,314585		1,035914	4,60873
PS	9,41745	6,982697	1,140522	1,035914		5,649117
NS	15,066567	12,631814	6,789639	4,60873	5,649117	
Wartość p	HS	MS	LS	NI	PS	NS

uzyskujemy podział na grupy jednorodne:

Jednorodı	HS(a)	MS(b)	LS(c)	NI(c)	PS(c)	NS(d)
A	*					
B		*				
C			*	*	*	
D						*

W tym przypadku uzyskano 4 grupy jednorodne, tzn. A, B, C i D, co wskazuje na możliwość przeprowadzania badania w oparciu o mniejszy podział, tzn. zamiast sześciu grup, które badaliśmy pierwotnie, można prowadzić dalsze analizy w oparciu o cztery wyznaczone tu grupy jednorodne. Kolejność grup ustalona została na podstawie średnich ważonych wyliczonych dla poszczególnych grup jednorodnych, w taki sposób, by litera A przypisana została do grupy o najniższej średniej ważonej, a dalsze litery alfabetu – kolejno do grup o coraz wyższych średnich.



Kontynuacja przykładu 17.1 (plik wiek ANOVA.pqs): ANOVA wykazała istotną statystycznie różnicę średniego wieku pracowników trzech firm przewozowych ($p = 0.0051$). Sprawdzamy teraz, które konkretnie firmy różnią się między sobą wiekiem pracowników, wykonując test POST-HOC Tukeya.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

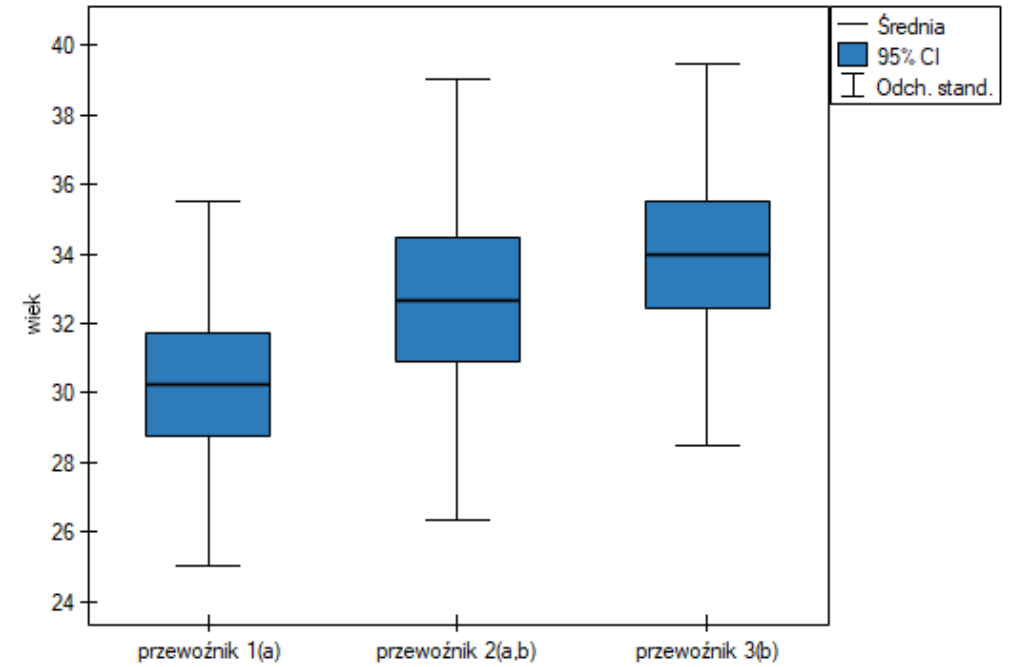
POST-HOC (Tukey HSD)			
	przewoźnik 1	przewoźnik 2	przewoźnik 3
Różnica średn			
przewoźnik 1		2.42	3.72
przewoźnik 2	2.42		1.3
przewoźnik 3	3.72	1.3	
NIR	przewoźnik 1	przewoźnik 2	przewoźnik 3
przewoźnik 1		2.7309	2.7309
przewoźnik 2	2.7309		2.7309
przewoźnik 3	2.7309	2.7309	
Statystyka q	przewoźnik 1	przewoźnik 2	przewoźnik 3
przewoźnik 1		2.996	4.6054
przewoźnik 2	2.996		1.6094
przewoźnik 3	4.6054	1.6094	
Wartość p	przewoźnik 1	przewoźnik 2	przewoźnik 3
przewoźnik 1		0.0897	0.004
przewoźnik 2	0.0897		0.4923
przewoźnik 3	0.004	0.4923	
Jednorodne	przewoźnik 1(	przewoźnik 2	przewoźnik 3
A	*	*	
B		*	*

$p < 0.05$ **(istotne statystycznie tylko dla porównania 1. i 3. firmy)**

Istotna różnica średniego wieku dotyczy wyłącznie pracowników pierwszej i trzeciej firmy przewozowej.

Najmniejsza istotna różnica (NIR) wyznaczona dla każdej pary porównań jest taka sama (ponieważ liczności grup są sobie równe) i wynosi 2.7309. Porównanie wartości NIR z wartością różnicy średnich wskazuje, że istotne różnice występują tylko pomiędzy wartością średnią dla wieku pracowników pierwszej i trzeciej firmy przewozowej (tylko w przypadku porównania tych dwóch grup wartość NIR jest mniejsza od różnicy średnich). Ten sam wniosek wyciągniemy, porównując wartości p testu POST-HOC z poziomem istotności $\alpha = 0.05$. Pracownicy pierwszej firmy są młodsi średnio o nieco ponad 3 lata od pracowników trzeciej firmy. Uzyskano dwie, zazębiające się grupy jednorodne, które zaznaczono również na wykresie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Test dla trendu. Jeśli oprócz samego istnienia różnic między grupami interesuje nas, czy średnie zmieniają się zgodnie z określonym, uporządkowanym trendem (np. rosną lub maleją wraz z kolejnymi grupami), można wykorzystać wariant testów POST-HOC oparty na kontrastach – tzw. test dla trendu. Uruchamiamy go w tym samym oknie opcji co pozostałe testy POST-HOC, wskazując oczekiwaną kolejność grup. Konstrukcję kontrastów trendu oraz wzór statystyki testowej opisano w części „Dla zainteresowanych".

DLA ZAINTERESOWANYCH

Wzory:

Liczba wszystkich możliwych porównań prostych wyliczana jest z wzoru:

$$c = \binom{k}{2} = \frac{k(k-1)}{2}$$

Kontrasty

Porównania w ramach kontrastów i testów POST-HOC zapisuje się w postaci hipotez.

Przykład 1 – **porównania proste** (porównanie pomiędzy sobą 2 wybranych średnich):

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2, \\ \mathcal{H}_1: &\quad \mu_1 \neq \mu_2. \end{aligned}$$

Przykład 2 – **porównania złożone** (porównanie kombinacji wybranych średnich):

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \tfrac{\mu_2+\mu_3}{2}, \\ \mathcal{H}_1: &\quad \mu_1 \neq \tfrac{\mu_2+\mu_3}{2}. \end{aligned}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

By można było zdefiniować wybrane hipotezy, należy dla każdej średniej przypisać wartość kontrastu c_j, $(j = 1, 2, ...k)$. Wartości c_j są tak wybierane, by ich sumy dla porównywanych stron były liczbami przeciwnymi, a ich wartość dla średnich nie biorących udziału w analizie wynosiła 0.

Przykład 1: $c_1 = 1$, $c_2 = -1$, $c_3 = 0, ...c_k = 0$.

Przykład 2: $c_1 = 2$, $c_2 = -1$, $c_3 = -1$, $c_4 = 0$,..., $c_k = 0$.

Wyboru właściwej hipotezy możemy dokonać:

(i) Porównując różnicę wybranych średnich z **najmniejszą istotną różnicą (NIR)** wyznaczoną odpowiednim testem POST-HOC: gdy różnica średnich $\geq NIR$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy różnica średnich $< NIR$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

(ii) Porównując wyznaczoną na podstawie statystyki testowej odpowiedniego testu POST-HOC wartość p z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test LSD Fishera
Dla porównań prostych i złożonych, zarówno równolicznych jak i różnolicznych grup, gdy wariancje nie różnią się istotnie.

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,1,df_{WG}}} \cdot \sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n_j}\right) MS_{WG}},$$

gdzie $F_{\alpha,1,df_{WG}}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz dla stopni swobody odpowiednio: 1 i df_{WG}.

(ii) Statystyka testowa ma postać:

$$t = \frac{\sum_{j=1}^{k} c_j \overline{x}_j}{\sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n_j}\right) MS_{WG}}}.$$

Statystyka ta podlega rozkładowi t-Studenta z df_{WG} stopniami swobody.

Test Scheffego
Dla porównań prostych zarówno równolicznych jak i różnolicznych grup, gdy wariancje nie różnią się istotnie.

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,df_{BG},df_{WG}}} \cdot \sqrt{(k-1)\left(\sum_{j=1}^{k} \frac{c_j^2}{n_j}\right) MS_{WG}},$$

gdzie $F_{\alpha,df_{BG},df_{WG}}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz df_{BG} i df_{WG} stopni swobody.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

(ii) Statystyka testowa ma postać:

$$F = \frac{\left(\sum_{j=1}^{k} c_j \overline{x}_j\right)^2}{(k-1)\left(\sum_{j=1}^{k} \frac{c_j^2}{n_j}\right) MS_{WG}}.$$

Statystyka ta podlega rozkładowi F Snedecora z df_{BG} i df_{WG} stopniami swobody.

Test Tukeya
Dla porównań prostych zarówno równolicznych jak i różnolicznych grup, gdy wariancje nie różnią się istotnie.

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \frac{\sqrt{2} \cdot q_{\alpha, df_{WG}, k} \cdot \sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n_j}\right) MS_{WG}}}{2},$$

gdzie $q_{\alpha, df_{WG}, k}$ to wartość krytyczna (statystyka) rozkładu studentyzowanego rozstępu dla zadanego poziomu istotności α oraz df_{WG} i k stopni swobody.

(ii) Statystyka testowa ma postać:

$$q = \sqrt{2} \frac{\sum_{j=1}^{k} c_j \overline{x}_j}{\sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n_j}\right) MS_{WG}}}.$$

Statystyka ta podlega rozkładowi studentyzowanego rozstępu z df_{WG} i k stopniami swobody.

Info. Algorytm obliczania wartości p i statystyki rozkładu studentyzowanego rozstępu w PQStat bazuje na pracy Lunda (1983)[110]. Inne programy lub strony internetowe mogą wyliczać nieco inne wartości niż PQStat, gdyż mogą bazować na mniej precyzyjnych lub bardziej restrykcyjnych algorytmach (Copenhaver i Holland (1988), Gleason (1999)).

Test dla trendu
Test badający istnienie trendu może być wyliczany w takiej samej sytuacji jak ANOVA dla zmiennych niezależnych, gdyż bazuje na tych samych założeniach, inaczej jednak ujmuje hipotezę alternatywną – wskazując w niej na istnienie trendu wartości średnich dla kolejnych populacji. Analiza trendu w ułożeniu średnich oparta jest na kontrastach LSD Fishera. Budując odpowiednie kontrasty, można badać dowolny rodzaj trendu, np. liniowy, kwadratowy, sześcienny itd. Poniżej znajduje się tabela przykładowych wartości kontrastów dla wybranych trendów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

		Kontrast									
Liczba grup	Trend	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
3	liniowy	-1	0	1							
	kwadratowy	1	-2	1							
4	liniowy	-3	-1	1	3						
	kwadratowy	1	-1	-1	1						
	sześcienny	-1	3	-3	1						
5	liniowy	-2	-1	0	1	2					
	kwadratowy	2	-1	-2	-1	2					
	sześcienny	-1	2	0	-2	1					
6	liniowy	-5	-3	-1	1	3	5				
	kwadratowy	5	-1	-4	-4	-1	5				
	sześcienny	-5	7	4	-4	-7	5				
7	liniowy	-3	-2	-1	0	1	2	3			
	kwadratowy	5	0	-3	-4	-3	0	5			
	sześcienny	-1	1	1	0	-1	-1	1			
8	liniowy	-7	-5	-3	-1	1	3	5	7		
	kwadratowy	7	1	-3	-5	-5	-3	1	7		
	sześcienny	-7	5	7	3	-3	-7	-5	7		
9	liniowy	-4	-3	-2	-1	0	1	2	3	4	
	kwadratowy	28	7	-8	-17	-20	-17	-8	7	28	
	sześcienny	-14	7	13	9	0	-9	-13	-7	14	
10	liniowy	-9	-7	-5	-3	-1	1	3	5	7	9
	kwadratowy	6	2	-1	-3	-4	-4	-3	-1	2	6
	sześcienny	-42	14	35	31	12	-12	-31	-35	-14	42

Trend liniowy
Trend liniowy, tak jak pozostałe trendy, możemy analizować, wpisując odpowiednie wartości kontrastów. Jeśli jednak znany jest kierunek trendu liniowego, wystarczy skorzystać z opcji Trend liniowy i wskazać oczekiwaną kolejność populacji, przypisując im kolejne liczby naturalne. Analiza przeprowadzana jest w oparciu o kontrast liniowy, czyli wskazanym według naturalnego uporządkowania grupom przypisane są odpowiednie wartości kontrastu i wyliczona zostaje statystyka LSD Fishera.

Przy znanym oczekiwanym kierunku trendu hipoteza alternatywna jest jednostronna i interpretacji podlega jednostronna wartość p. Interpretacja dwustronnej wartości p oznacza, że badacz nie zna (nie zakłada) kierunku ewentualnego trendu.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

ANOVA dla grup niezależnych z korektą F^* i F''

KIEDY STOSOWAĆ TEN TEST:

Korekty F^* (Brown-Forsythe, 1974[33]) oraz F'' (Welch, 1951[172]) to warianty ANOVA dla grup niezależnych stosowane wtedy, gdy wariancje badanej zmiennej w porównywanych populacjach różnią się

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

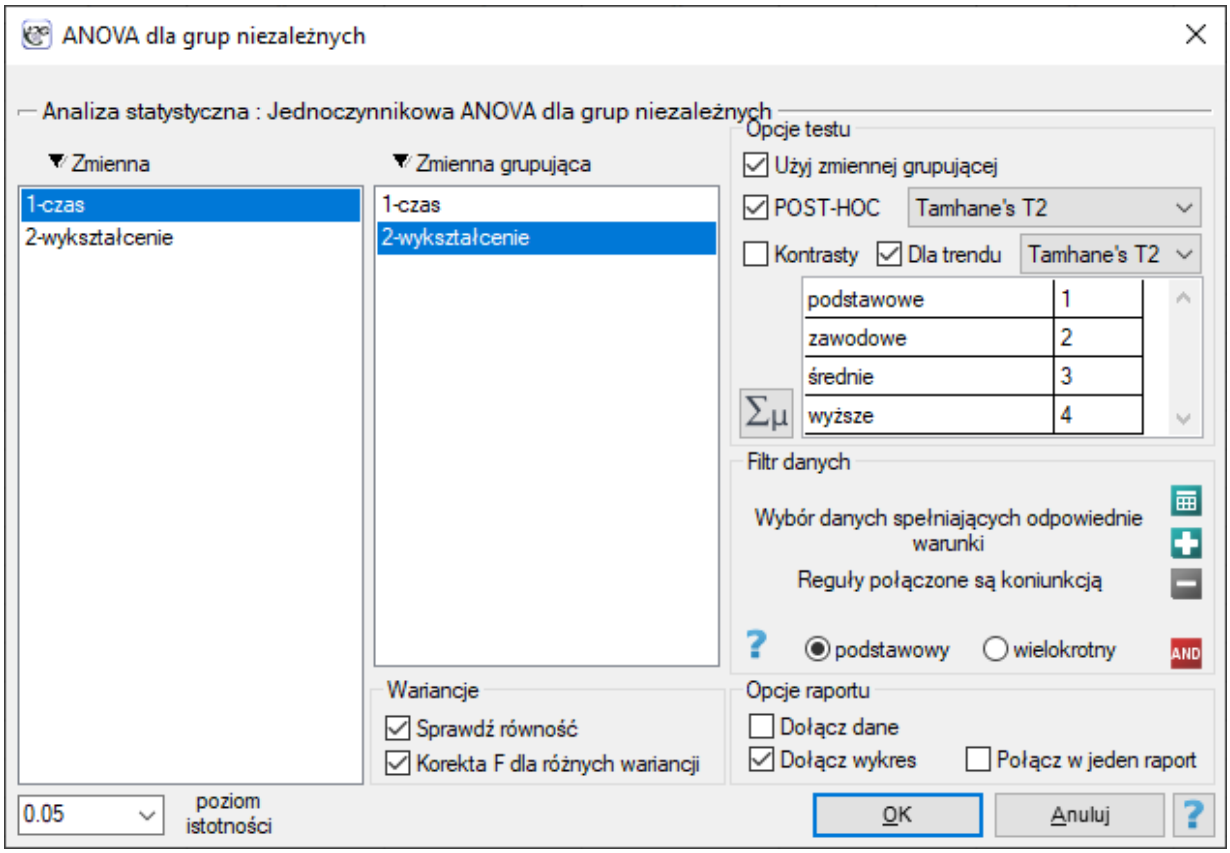
istotnie. Cel testu jest taki sam jak dla wariantu podstawowego – różni się jedynie sposób wyliczenia statystyki testowej i stopni swobody (zob. *Dla zainteresowanych*).

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanej zmiennej w każdej populacji,
- model niezależny,
- istotnie różne wariancje badanej zmiennej między populacjami (w przeciwnym razie stosuje się ANOVA dla grup niezależnych bez korekty).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → ANOVA dla grup niezależnych (lub poprzez Kreator).



Krok 1. Wskaż zmienną grupującą oraz zmienną wynikową.

Krok 2. Sprawdź w oknie analizy założenie równości wariancji (test Levene'a lub Browna-Forsythe'a).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krok 3. Jeśli wariancje różnią się istotnie, uruchom analizę z korektą F^*/F'' oraz dobierz odpowiedni test POST-HOC (np. Tamhane's T2) dla grup o różnych wariancjach.

Krok 4. Uruchom analizę.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 17.2. (plik bezrobocie.pqs)
Jest wiele czynników regulujących czas poszukiwania pracy w dobie kryzysu gospodarczego. Jednym z najważniejszych może być poziom wykształcenia. Przykładowe dane dotyczące wykształcenia oraz czasu (w miesiącach) pozostawania bezrobotnym zebrano w pliku. Chcemy sprawdzić, czy istnieją różnice w średnim czasie poszukiwania pracy dla poszczególnych kategorii wykształcenia.

Ze względu na różnice dotyczące wariancji pomiędzy poszczególnymi populacjami (dla testu Levene'a wartość $p = 0.0001$, a dla testu Browna-Forsythe'a wartość $p = 0.0002$):

Równość wariancji - Brown-Forsythe	
Statystyka F	7.0103
Wartość p	0.0002
Równość wariancji - Levene	
Statystyka F	7.4708
Wartość p	0.0001

analizę przeprowadzamy przy włączonej korekcie różnych wariancji. Uzyskany wynik skorygowanej statystyki F jest przedstawiony poniżej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednoczynnikowa ANOVA dla grup niezależnych	
Czas analizy	0.33 sek.
Analizowane zmienne	czas;wykształcenie
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Zmienna grupująca	wykształcenie
Eta-kwadrat	0.1001
Całkowita suma kwadratów (SS[T])	4226.3607
Suma kwadratów między grupami (SS[BG])	423.2131
Suma kwadratów wewnątrz grup (SS[WG])	3803.1475
Średnia kwadratów między grupami (MS[BG])	141.071
Średnia kwadratów wewnątrz grup (MS[WG])	15.8464
Stopnie swobody między grupami (df[BG])	3
Stopnie swobody wewnątrz grup (df[WG])	240
Całkowite stopnie swobody (df[T])	243
Statystyka F	8.9024
Wartość p	<0.0001
Korekta F dla różnych wariancji - F* (Brown-Forsythe	
Statystyka F	8.9024
Stopnie swobody	3 / 203.3014
Wartość p	<0.0001
Korekta F dla różnych wariancji - F'' (Welch)	
Statystyka F	8.6631
Stopnie swobody	3 / 130.5548
Wartość p	<0.0001

$p < 0.0001$ (dla testu F^*) oraz $p < 0.0001$ (dla testu F''), obie istotne statystycznie, $\alpha = 0.05$

Średni czas poszukiwania pracy różni się w zależności od posiadanego wykształcenia.

Wykonując jeden z testów POST-HOC dedykowanych porównaniu grup o różnych wariancjach (zob. *Dla zainteresowanych*), dowiadujemy się, których kategorii wykształcenia dotyczą stwierdzone różnice:

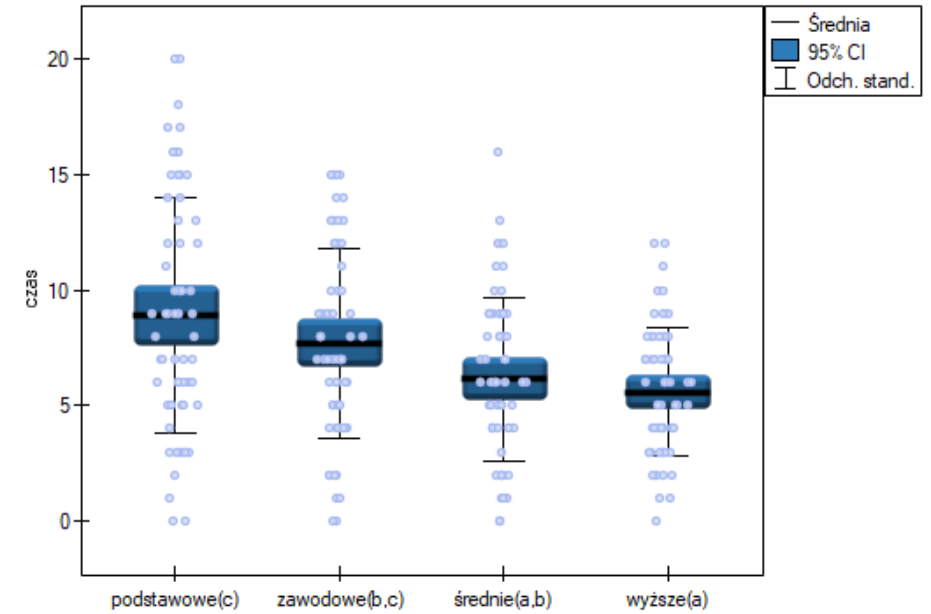
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

POST-HOC (Tamhane's T2)

	podstawowe	zawodowe	średnie	wyższe
Różnica średn				
podstawowe		1.2131	2.7869	3.3443
zawodowe	1.2131		1.5738	2.1311
średnie	2.7869	1.5738		0.5574
wyższe	3.3443	2.1311	0.5574	
NIR	podstawowe	zawodowe	średnie	wyższe
podstawowe		2.2457	2.1376	2.0038
zawodowe	2.2457		1.8608	1.7024
średnie	2.1376	1.8608		1.551
wyższe	2.0038	1.7024	1.551	
Statystyka T	podstawowe	zawodowe	średnie	wyższe
podstawowe		-1.4462	-3.4947	-4.4866
zawodowe	-1.4462		-2.2633	-3.3564
średnie	-3.4947	-2.2633		-0.9623
wyższe	-4.4866	-3.3564	-0.9623	
Wartość p	podstawowe	zawodowe	średnie	wyższe
podstawowe		0.6251	0.0041	0.0001
zawodowe	0.6251		0.1433	0.0066
średnie	0.0041	0.1433		0.9158
wyższe	0.0001	0.0066	0.9158	
Jednorodne	podstawowe	zawodowe(b	średnie(a,b)	wyższe(a)
A			*	*
B		*	*	
C	*	*		

Najmniejsza istotna różnica (NIR) wyznaczona dla każdej pary porównań nie jest taka sama (mimo, że liczności grup są sobie równe), ponieważ nie są równe wariancje. Odnosząc wartość NIR do uzyskanych różnic wartości średnich, uzyskamy ten sam rezultat co, porównując wartość p z poziomem istotności $\alpha = 0.05$. Różnice dotyczą wykształcenia podstawowego i wyższego, wykształcenia podstawowego i średniego oraz wykształcenia zawodowego i wyższego. Powstałe grupy jednorodne zazębiają się. Generalnie jednak, spoglądając na wykres, możemy oczekiwać, że im bardziej wykształcona osoba, tym mniej czasu zajmuje jej poszukiwanie pracy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



By móc sprawdzić tak postawioną hipotezę, podejmujemy analizę dla trendu. W tym celu wznawiamy analizę przyciskiem i w oknie opcji testu wybieramy metodę Tamhane's T2 oraz opcję Dla trendu, wskazując kolejność kategorii wykształcenia kolejnymi liczbami naturalnymi.

Trend liniowy (Fisher LSD)	
Numer porządkowy dla [podstawowe]	1
Numer porządkowy dla [zawodowe]	2
Numer porządkowy dla [średnie]	3
Numer porządkowy dla [wyższe]	4
Statystyka F	25.9283
Wartość p dwustronne	<0.0001
Wartość p jednostronne	<0.0001

W zależności od tego, czy kierunek zależności pomiędzy wykształceniem a czasem poszukiwania pracy jest nam znany, wykorzystujemy jednostronną lub dwustronną wartość p. Obie te wartości są mniejsze niż zadany poziom istotności – przewidywany przez nas trend został potwierdzony, czyli na poziomie istotności $\alpha = 0.05$ możemy powiedzieć, że ów trend rzeczywiście istnieje w populacji, z której pochodzi próba.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_k, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, ..., k). \end{aligned}$$

Wzory:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Statystyka testowa ma postać:

$$F^* = \frac{SS_{BG}}{\sum_{j=1}^{k}\left(1-\frac{n_j}{n}sd_j^2\right)},$$

$$F'' = \frac{\frac{\sum_{j=1}^{k} w_j(\overline{x}_j-\widetilde{x})}{k-1}}{1+\frac{2(k-2)}{k^2-1}\sum_{j=1}^{k} h_j},$$

gdzie:
sd_j – odchylenie standardowe grupy j,
$w_j = \frac{n_j}{sd_j^2}$ – waga grupy j,
$\widetilde{x}$ – średnia ważona,
$h_j = \frac{\left(1-\frac{w_j}{\sum_{j=1}^{k} w_j}\right)^2}{n_j-1}$.

Statystyka ta podlega rozkładowi F Snedecora z $k-1$ i skorygowanymi df_{WG_k} stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Testy POST-HOC

Wprowadzenie do kontrastów i testów POST-HOC znajduje się w rozdziale 17.1.2 dotyczącym jednoczynnikowej analizy wariancji. Poniższe metody, w odróżnieniu od testów LSD Fishera, Scheffego i Tukeya, stosuje się, gdy wariancje badanej zmiennej różnią się istotnie.

Test T2 Tamhane
Dla porównań prostych i złożonych zarówno równolicznych jak i różnolicznych grup (Tamhane A. C., 1977[160]).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha_{Sidak},1,df_v}} \cdot \sqrt{\left(\sum_{j=1}^{k}\frac{c_j^2 sd_j^2}{n_j}\right)},$$

gdzie $F_{\alpha_{Sidak},1,df_v}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zmodyfikowanego poziomu istotności α_{Sidak} oraz dla stopni swobody 1 i df_v odpowiednio,
$\alpha_{Sidak} = 1-(1-\alpha)^{(1/k)}$,

$$df_v = \frac{\left(\sum_{j=1}^{k}\frac{c_j^2 sd_j^2}{n_j}\right)^2}{\sum_{j=1}^{k}\frac{c_j^4 sd_j^4}{n_j^2(n_j-1)}}$$

(ii) Statystyka testowa ma postać:

$$t = \frac{\left(\sum_{j=1}^{k} c_j\overline{x}_j\right)^2}{\sqrt{\left(\sum_{j=1}^{k}\frac{c_j^2 sd_j^2}{n_j}\right)}}.$$

Statystyka ta podlega rozkładowi t-Studenta z df_v stopniami swobody, a wartość p jest korygowana o liczbę możliwych porównań prostych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test BF (Brown-Forsythe)
Dla porównań prostych zarówno równolicznych jak i różnolicznych grup (Brown M. B. i Forsythe A. B., 1974[32]).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,k-1,df_v}} \cdot \sqrt{(k-1)\left(\sum_{j=1}^{k} \frac{c_j^2 sd_j^2}{n_j}\right)},$$

gdzie $F_{\alpha,k-1,df_v}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz $k-1$ i df_v stopni swobody.

(ii) Statystyka testowa ma postać:

$$F = \frac{\left(\sum_{j=1}^{k} c_j \overline{x}_j\right)^2}{(k-1)\left(\sum_{j=1}^{k} \frac{c_j^2 sd_j^2}{n_j}\right)}.$$

Statystyka ta podlega rozkładowi F Snedecora z $k-1$ i df_v stopniami swobody.

Test GH (Games-Howell)
Dla porównań prostych zarówno równolicznych jak i różnolicznych grup (Games P. A. i Howell J. F., 1976[68]).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \frac{q_{\alpha,k,df_v} \cdot \sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2 sd_j^2}{n_j}\right)}}{\sqrt{2}},$$

gdzie q_{α,k,df_v} to wartość krytyczna (statystyka) rozkładu studentyzowanego rozstępu dla zadanego poziomu istotności α oraz k i df_v stopni swobody.

(ii) Statystyka testowa ma postać:

$$q = \sqrt{2}\frac{\sum_{j=1}^{k} c_j \overline{x}_j}{\sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2 sd_j^2}{n_j}\right)}}.$$

Statystyka ta podlega rozkładowi studentyzowanego rozstępu z k i df_v stopniami swobody.

Test dla trendu
Test badający istnienie trendu może być wyliczany w takiej samej sytuacji jak ANOVA dla grup niezależnych z korektą F^* i F'', gdyż bazuje na tych samych założeniach, inaczej jednak ujmuje hipotezę alternatywną – wskazując w niej na istnienie trendu wartości średnich dla kolejnych populacji. Analiza trendu w ułożeniu średnich oparta jest na kontrastach T2 Tamhane. Tabela przykładowych wartości kontrastów dla wybranych trendów znajduje się w opisie testu dla trendu dla ANOVA bez korekty dla różnych wariancji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Trend liniowy analizujemy w oparciu o kontrast liniowy, przypisując grupom uporządkowanym w naturalny sposób odpowiednie wartości kontrastu i wyliczając statystykę T2 Tamhane. Przy znanym oczekiwanym kierunku trendu hipoteza alternatywna jest jednostronna i interpretacji podlega jednostronna wartość p; interpretacja dwustronnej wartości p oznacza, że badacz nie zna (nie zakłada) kierunku ewentualnego trendu. Wyznaczoną wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test Browna-Forsythe'a i test Levene'a

KIEDY STOSOWAĆ TEN TEST:

Test Levene'a (Levene, 1960[104]) oraz test Browna-Forsythe'a (Brown i Forsythe, 1974[31]) pozwalają sprawdzić, czy rozrzut (wariancja) badanej cechy jest taki sam w co najmniej dwóch populacjach. Wynik tego testu jest zwykle sprawdzany jako warunek wstępny przed wykonaniem ANOVA dla grup niezależnych – wskazuje, czy należy zastosować wariant z korektą dla różnych wariancji.

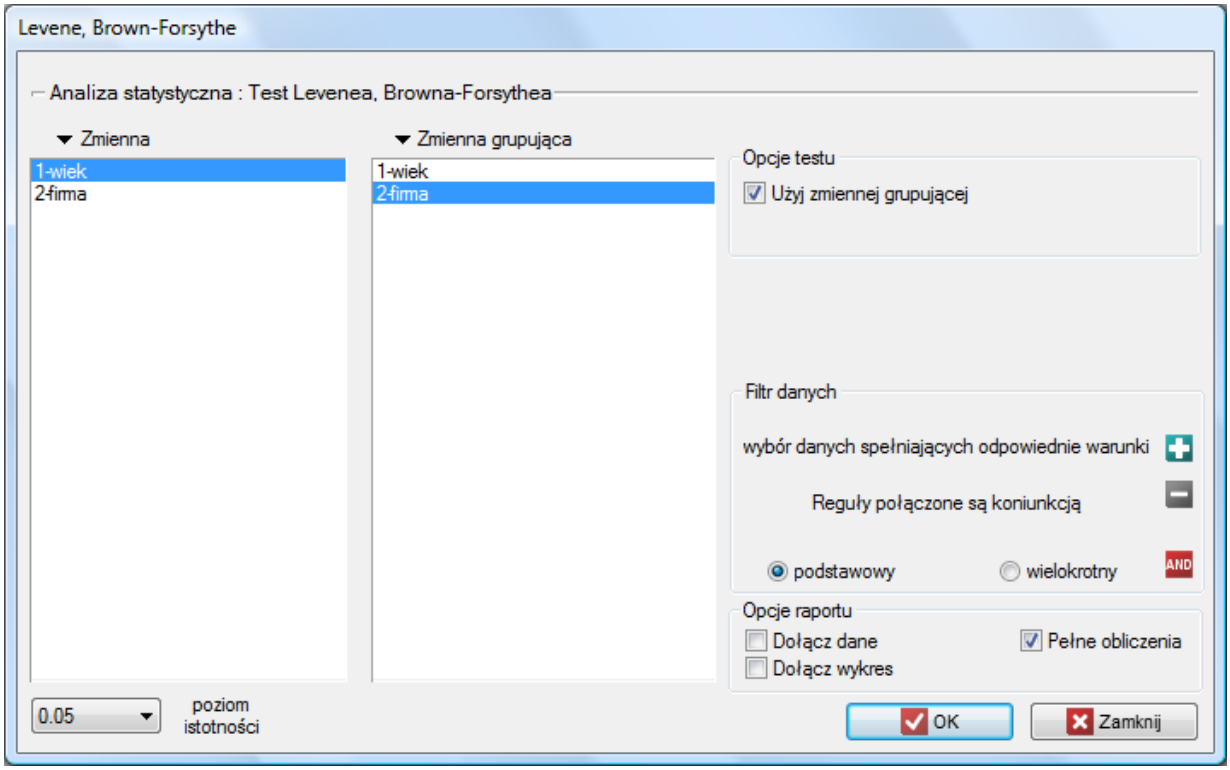
WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanej zmiennej w każdej populacji,
- model niezależny.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → Levene, Brown-Forsythe.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Uwaga! Test Browna-Forsythe'a jest mniej wrażliwy niż test Levene'a na niespełnienie założenia dotyczącego normalności rozkładu.

WYKONANIE I INTERPRETACJA

Testy te w praktyce wykonuje się jako element szerszej analizy – przykład ich zastosowania i interpretacji znajduje się w rozdziale ANOVA dla grup niezależnych z korektą F^* i F'' (Przykład 17.2).

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \sigma_j^2 \text{ są sobie równe } (j = 1, 2, ..., k), \end{aligned}$$

gdzie σ_1^2,σ_2^2,...,σ_k^2 – wariancje badanej zmiennej w populacjach, z których pobrano próby.

Procedura: analiza polega na wyznaczeniu bezwzględnego odchylenia wyników pomiarowych od średniej (w teście Levene'a) lub od mediany (w teście Browna-Forsythe'a) w każdej z badanych grup. To bezwzględne odchylenie stanowi dane, które poddaje się dokładnie tej samej procedurze, co przy analizie wariancji dla grup niezależnych.

Wzory: statystyka testowa przyjmuje postać:

$$F = \frac{MS_{BG}}{MS_{WG}},$$

Statystyka ta podlega rozkładowi F Snedecora z df_{BG} i df_{WG} stopniami swobody.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

ANOVA powtarzanych pomiarów

KIEDY STOSOWAĆ TEN TEST:

Jednoczynnikowa analiza wariancji dla powtarzanych pomiarów, czyli ANOVA dla grup zależnych, pozwala sprawdzić, czy średnia wartość badanej cechy jest taka sama przy co najmniej trzech pomiarach wykonanych na tych samych obiektach w różnych warunkach (np. przed leczeniem, w trakcie i po leczeniu). Test wskazuje jedynie, czy takie różnice w ogóle istnieją – nie mówi, między którymi konkretnie pomiarami występują; do tego służą testy POST-HOC, opisane niżej.

WARUNKI STOSOWANIA

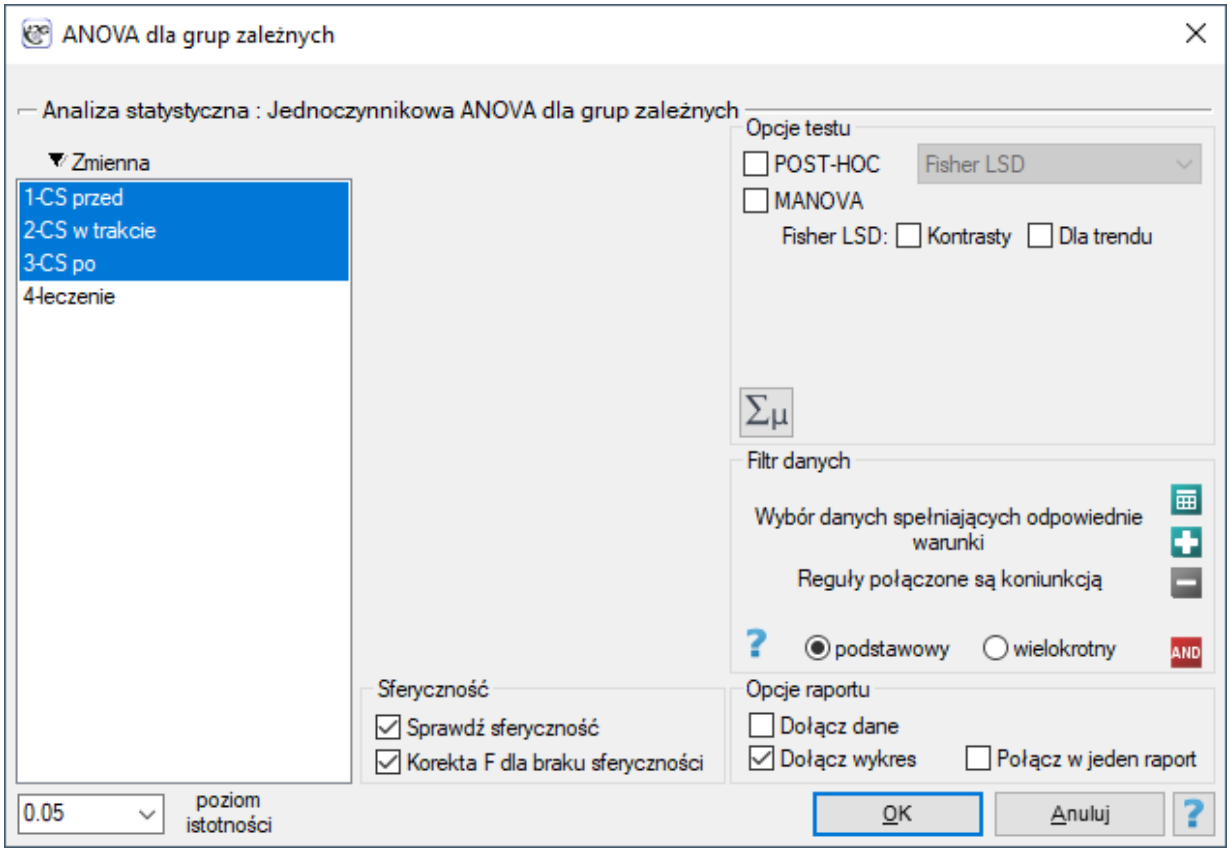
- pomiar na skali interwałowej,
- normalność rozkładu dla wszystkich zmiennych będących różnicą par pomiarowych (lub normalność badanej zmiennej dla każdego pomiaru),
- model zależny,
- sferyczność: wariancje różnic pomiędzy wszystkimi parami pomiarów powinny być sobie bliskie.

Założenie sferyczności sprawdza się testem Mauchly'a. Gdy nie jest ono spełnione, stosuje się ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → ANOVA dla grup zależnych (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienne odpowiadające kolejnym pomiarom.

Krok 2. Sprawdź założenie sferyczności (test Mauchly'a).

Krok 3. Uruchom analizę.

WYKONANIE I INTERPRETACJA

Interpretacja wyniku ANOVA powtarzanych pomiarów zależy od wcześniejszej oceny założenia sferyczności, dlatego pełny przykład wykonania i interpretacji tego testu – łącznie z sytuacją, w której założenie to jest naruszone – znajduje się w rozdziale ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA (patrz Przykład 17.3).

DEFINICJE MIAR

Cząstkowa η^2 – miara wielkości efektu określająca, jaka część wewnątrzosobowej zmienności wyników wynika z powtarzanych pomiarów badanej zmiennej, a jaka pozostaje niewyjaśniona. Im wyższa wartość, tym silniejszy jest związek pomiędzy pomiarem a wynikiem badanej cechy.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_k, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, ..., k), \end{aligned}$$

gdzie μ_1,μ_2,...,μ_k – średnie badanej zmiennej w kolejnych pomiarach z badanej populacji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wzory:

Statystyka testowa ma postać:

$$F = \frac{MS_{BC}}{MS_{res}},$$

gdzie:

$MS_{BC} = \frac{SS_{BC}}{df_{BC}}$ – średnia kwadratów między pomiarami,

$MS_{res} = \frac{SS_{res}}{df_{res}}$ – średnia kwadratów dla reszt,

$SS_{BC} = \sum_{j=1}^{k}\left(\frac{\left(\sum_{i=1}^{n} x_{ij}\right)^2}{n}\right) - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{N}$ – suma kwadratów między pomiarami,

$SS_{res} = SS_T - SS_{BS} - SS_{BC}$ – suma kwadratów dla reszt,

$SS_T = \left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}^2\right) - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{N}$ – całkowita suma kwadratów,

$SS_{BS} = \sum_{i=1}^{n}\left(\frac{\left(\sum_{j=1}^{k} x_{ij}\right)^2}{k}\right) - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{N}$ – suma kwadratów między obiektami,

$df_{BC} = k - 1$ – stopnie swobody (między pomiarami),

$df_{res} = df_T - df_{BC} - df_{BS}$ – stopnie swobody (dla reszt),

$df_T = N - 1$ – całkowite stopnie swobody,

$df_{BS} = n - 1$ – stopnie swobody (między obiektami),

$N = nk$,

n – liczność próby,

x_{ij} – wartości zmiennej dla i obiektów $(i = 1, 2, ...n)$ w j pomiarach $(j = 1, 2, ...k)$.

Statystyka ta podlega rozkładowi F Snedecora z df_{BC} i df_{res} stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Cząstkowa η^2

$$\eta^2 = \frac{SS_{BC}}{SS_{BC} + SS_{res}}$$

Testy POST-HOC

Wprowadzenie do kontrastów i testów POST-HOC znajduje się w rozdziale 17.1.2 dotyczącym jednoczynnikowej analizy wariancji dla grup niezależnych. Poniższe metody stosuje się analogicznie dla powtarzanych pomiarów (liczność w poszczególnych pomiarach jest zawsze taka sama).

Test LSD Fishera

Dla porównań prostych i złożonych.

Przykład – **porównania proste** (porównanie pomiędzy sobą 2 wybranych średnich):

$$\begin{aligned} \mathcal{H}_0 &: \quad \mu_j = \mu_{j+1}, \\ \mathcal{H}_1 &: \quad \mu_j \neq \mu_{j+1}. \end{aligned}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,1,df_{res}}} \cdot \sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}},$$

gdzie $F_{\alpha,1,df_{res}}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz dla stopni swobody odpowiednio: 1 i df_{res}.

(ii) Statystyka testowa ma postać:

$$t = \frac{\sum_{j=1}^{k} c_j \overline{x}_j}{\sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}}}.$$

Statystyka ta podlega rozkładowi t-Studenta z df_{res} stopniami swobody. Dla kontrastów wielkość $\sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}}$ zastąpiona jest błędem kontrastu $SE_{contrast}$, a stopnie swobody to df_{BS}.

Test Scheffego
Dla porównań prostych.

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,df_{BC},df_{res}}} \cdot \sqrt{(k-1)\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}},$$

gdzie $F_{\alpha,df_{BC},df_{res}}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz df_{BC} i df_{res} stopni swobody.

(ii) Statystyka testowa ma postać:

$$F = \frac{\left(\sum_{j=1}^{k} c_j \overline{x}_j\right)^2}{(k-1)\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}}.$$

Statystyka ta podlega rozkładowi F Snedecora z df_{BC} i df_{res} stopniami swobody.

Test Tukeya
Dla porównań prostych.

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \frac{\sqrt{2} \cdot q_{\alpha,df_{res},k} \cdot \sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}}}{2},$$

gdzie $q_{\alpha,df_{res},k}$ to wartość krytyczna (statystyka) rozkładu studentyzowanego rozstępu dla zadanego poziomu istotności α oraz df_{res} i k stopni swobody.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

(ii) Statystyka testowa ma postać:

$$q = \sqrt{2}\frac{\sum_{j=1}^{k} c_j \overline{x}_j}{\sqrt{\left(\sum_{j=1}^{k} \frac{c_j^2}{n}\right) MS_{res}}}.$$

Statystyka ta podlega rozkładowi studentyzowanego rozstępu z df_{res} i k stopniami swobody.

Info. Algorytm obliczania wartości p i statystyki rozkładu studentyzowanego rozstępu w PQStat bazuje na pracy Lunda (1983)[110]. Inne programy lub strony internetowe mogą wyliczać nieco inne wartości niż PQStat, gdyż mogą bazować na mniej precyzyjnych lub bardziej restrykcyjnych algorytmach (Copenhaver i Holland (1988), Gleason (1999)).

Test dla trendu
Test badający istnienie trendu może być wyliczany w takiej samej sytuacji jak ANOVA dla zmiennych zależnych, gdyż bazuje na tych samych założeniach, inaczej jednak ujmuje hipotezę alternatywną – wskazując w niej na istnienie trendu wartości średnich w kolejnych pomiarach. Analiza trendu w ułożeniu średnich oparta jest na kontrastach Testu LSD Fishera. Tabela przykładowych wartości kontrastów dla wybranych trendów znajduje się w opisie testu dla trendu dla ANOVA zmiennych niezależnych.

Trend liniowy analizujemy w oparciu o kontrast liniowy, przypisując pomiarom uporządkowanym w naturalny sposób odpowiednie wartości kontrastu i wyliczając statystykę LSD Fishera. Przy znanym oczekiwanym kierunku trendu hipoteza alternatywna jest jednostronna i interpretacji podlega jednostronna wartość p; interpretacja dwustronnej wartości p oznacza, że badacz nie zna (nie zakłada) kierunku ewentualnego trendu. Wyznaczoną wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Sferyczność Mauchly'ego

KIEDY STOSOWAĆ TEN TEST:

Test Mauchly'ego (1940[120]) sprawdza założenie sferyczności – warunek podobny, ale silniejszy niż równość wariancji, wymagany przed wykonaniem ANOVA powtarzanych pomiarów. Założenie to jest spełnione, jeśli wariancje różnic pomiędzy wszystkimi parami powtarzanych pomiarów są sobie bliskie. Istotny wynik testu ($p \leq \alpha$) oznacza naruszenie założenia sferyczności i konieczność zastosowania odpowiedniej korekty (zob. ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA).

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- wielowymiarowy rozkład normalny lub normalność rozkładu każdej badanej zmiennej,
- model zależny.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zamiast założenia sferyczności rozważa się czasem prostszy, ale bardziej rygorystyczny warunek symetrii złożonej (*ang. compound symmetry*), zakładający symetrię macierzy kowariancji – równość wariancji zmiennych oraz równość kowariancji między nimi. Spełnienie warunku symetrii złożonej pociąga za sobą spełnienie założenia sferyczności.

WYKONANIE I INTERPRETACJA

Test ten w praktyce wykonuje się jako element szerszej analizy – przykład jego zastosowania i interpretacji znajduje się w rozdziale ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA (Przykład 17.3).

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \sigma_{diff(1)} = \sigma_{diff(2)} = ... = \sigma_{diff(s)}, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \sigma_{diff(i)} \text{ są sobie równe } (i = 1, 2, ..., s), \end{aligned}$$

gdzie $\sigma_{diff(i)}$ – populacyjna wariancja różnic pomiędzy i-tą parą powtarzanych pomiarów, s – liczba par.

Wzory: wartość W Mauchly'ego definiowana jest następująco:

$$W = \frac{\prod_{j=1}^{k-1} \lambda_j}{\left(\frac{1}{k-1}\sum_{j=1}^{k-1} \lambda_j\right)^{k-1}}.$$

Statystyka testowa ma postać:

$$\chi^2 = (f-1)(n-1)\ln W,$$

gdzie:
$f = \frac{2(k-1)^2+(k-1)+2}{6(k-1)(n-1)}$,
λ_j – wartość własna oczekiwanej macierzy kowariancji,
k – liczba analizowanych zmiennych.

Interpretacja: statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $df = \frac{k(k-1)}{2} - 1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Wartość $W \approx 1$ oznacza spełnienie założenia sferyczności. W interpretacji wyników tego testu należy jednak pamiętać, że jest on wrażliwy na złamanie założenia normalności rozkładu.

ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA

KIEDY STOSOWAĆ TEN TEST:

Korekty Epsilon oraz podejście wielowymiarowe MANOVA dotyczą ANOVA powtarzanych pomiarów i stosuje się je wówczas, gdy nie jest spełnione założenie sferyczności, czyli gdy wariancje różnic pomiędzy wszystkimi parami pomiarów nie są sobie bliskie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu dla wszystkich zmiennych będących różnicą par pomiarowych,
- model zależny,
- brak sferyczności (w przeciwnym razie wystarczy ANOVA powtarzanych pomiarów bez korekty).

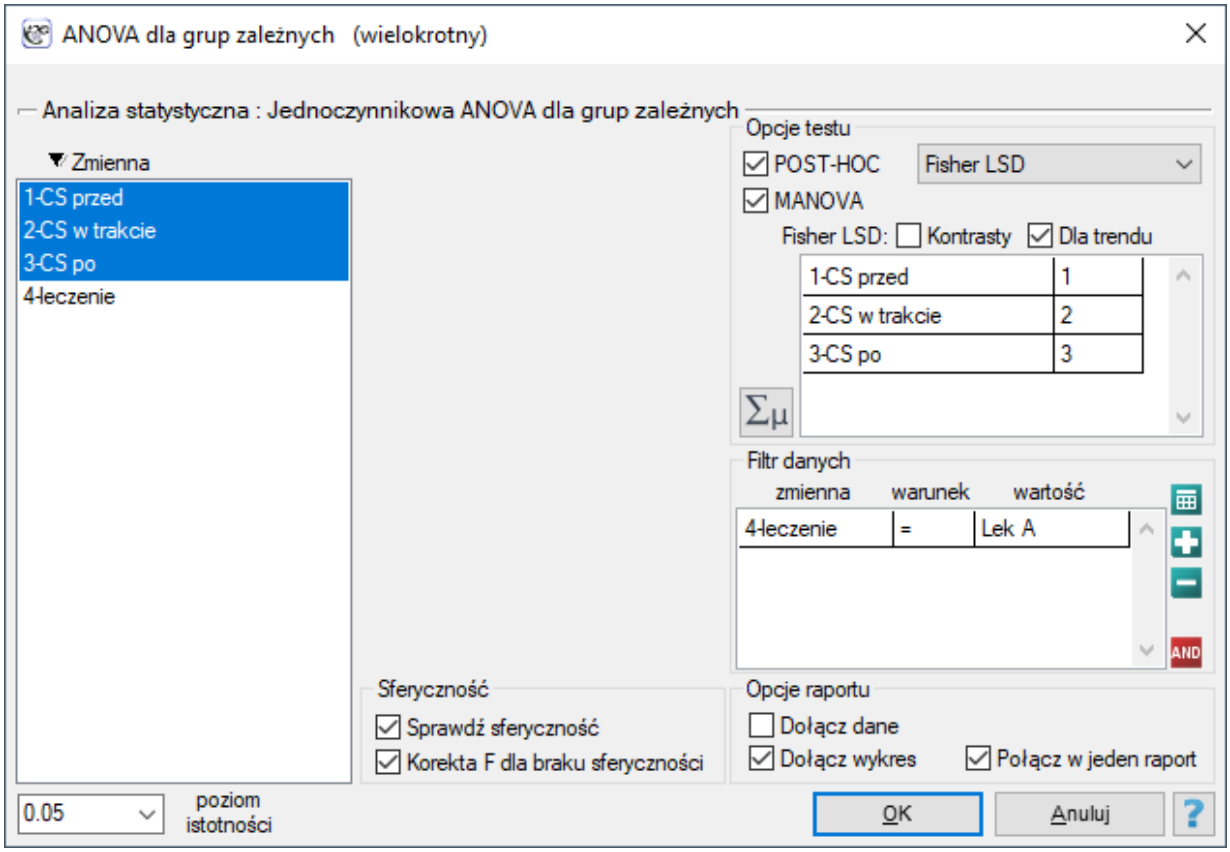
Stopień, w jakim sferyczność jest spełniona, reprezentuje wartość W w teście Mauchly'a oraz wartości **Epsilon** (ε) wyliczane wraz z korektami. $\varepsilon = 1$ oznacza ścisłe przestrzeganie warunku sferyczności – im mniejsza od 1 jest wartość Epsilon, tym założenie sferyczności jest bardziej naruszone (dolna granica, jaką Epsilon może osiągnąć, to $\frac{1}{k-1}$).

Dostępne są trzy korekty zmieniające liczbę stopni swobody testu F: najprostsza, lecz najsłabsza **korekta dolnej granicy Epsilon**; nieco silniejsza, choć również konserwatywna **korekta Greenhouse-Geisser** (1959)[72]; oraz najsilniejsza **korekta Huynh-Feldt** (1976)[88]. Przy znacznym naruszeniu sferyczności najbardziej wskazane jest jednak wykonanie analizy, która tego założenia nie wymaga, czyli **MANOVA** (*ang. multivariate analysis of variance*) – w zastosowaniu do pomiarów powtarzanych (bez czynnika grup niezależnych) testuje ona tę samą hipotezę, co ANOVA dla zmiennych zależnych, bazując na różnicach sąsiadujących pomiarów. Szersze omówienie tej metody znajduje się w rozdziale jednoczynnikowa MANOVA.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → ANOVA dla grup zależnych (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Sprawdź założenie sferyczności testem Mauchly'a (najlepiej filtrując dane wg grup, jeśli badanie ma więcej niż jedną grupę).

Krok 2. Jeśli sferyczność jest naruszona, wybierz w oknie opcji odpowiednią korektę Epsilon lub statystykę MANOVA.

Krok 3. Uruchom analizę.

WYKONANIE I INTERPRETACJA

Przykład 17.3. (plik ciśnienie.pqs)
Badano skuteczność dwóch metod leczenia nadciśnienia. W tym celu zebrano próbę 56 chorych i przydzielono ich losowo do dwóch grup: grupy leczonej lekiem A i grupy leczonej lekiem B. Pomiaru ciśnienia skurczowego w każdej grupie dokonano trzykrotnie: przed leczeniem, w trakcie leczenia i po 6 miesiącach prowadzenia kuracji. Sprawdzimy, czy w każdej z grup średnia wartość ciśnienia skurczowego zmienia się w kolejnych etapach leczenia.

Ponieważ dane mają rozkład normalny, analizę rozpoczynamy od sprawdzenia założenia sferyczności – testowanie wykonujemy dla każdej grupy osobno, wykorzystując filtr wielokrotny. Na niespełnienie założenia sferyczności przez grupę leczoną lekiem B wskazują zarówno obserwowane wartości macierzy kowariancji i korelacji, jak i wynik testu Mauchly'a ($W = 0.68$, $p = 0.0063$).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Macierz kowariancji			
	CS przed	CS w trak	CS po
CS przed	233,06349	139,33333	167,35978
CS w trakcie	139,33333	164,74074	175,48148
CS po	167,35978	175,48148	275,28571

Macierz korelacji			
	CS przed	CS w trak	CS po
CS przed	1	0,7110782	0,6607271
CS w trakcie	0,7110782	1	0,8240222
CS po	0,6607271	0,8240222	1

ZAŁOŻENIE SFERYCZNOŚCI	
Sferyczność Mauchly	
W	0.677
Statystyka Chi-kwadrat	10.1439
Stopnie swobody	2
Wartość p	0.0063

Wznawiamy analizę i w oknie opcji testu wybieramy filtr podstawowy, by wykonać ANOVA powtarzanych pomiarów dla leczonych lekiem A, a następnie korektę tej analizy i statystykę MANOVA dla leczonych lekiem B.

Wyniki dla leczonych lekiem A:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednoczynnikowa ANOVA dla grup zależnych	
Czas analizy	0.13 sek.
Analizowane zmienne	CS przed;CS w trakcie;CS p
Liczność - niezinterpretowane	0
Liczność - braki danych (wierszami)	0
Poziom istotności	0.05
Filtr danych	leczenie=Lek A
Eta-kwadrat	0.6639
Całkowita suma kwadratów (SS[T])	24960.9524
Suma kwadratów między pomiarami (SS[BC])	6787.5238
Suma kwadratów między obiektami (SS[BS])	14736.9524
Suma kwadratów dla reszt (SS[RES])	3436.4762
Stopnie swobody między pomiarami (df[BC])	2
Stopnie swobody [obiekty]	27
Stopnie swobody [res]	54
Całkowite stopnie swobody (df[T])	83
Średnia kwadratów [pomiędzy sędziami]	3393.7619
Średnia kwadratów [pomiędzy obiektami]	545.8131
Średnia kwadratów [res]	63.6384
Statystyka F	53.3288
Wartość p	<0.0001

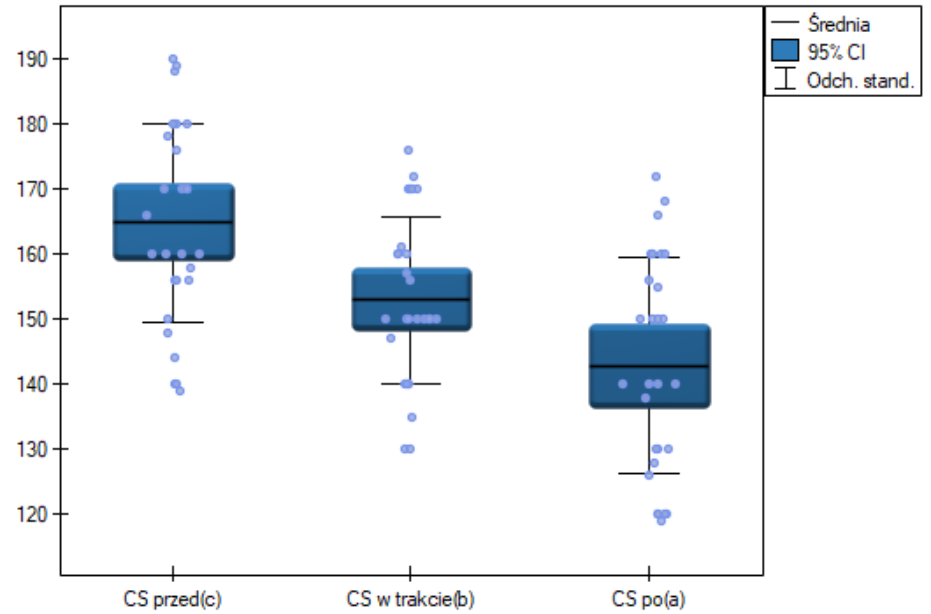
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

POST-HOC (Fisher LSD)

	CS przed	CS w trakcie	CS po
Różnica średn			
CS przed		11.7857	22
CS w trakcie	11.7857		10.2143
CS po	22	10.2143	
NIR	CS przed	CS w trakcie	CS po
CS przed		4.2745	4.2745
CS w trakcie	4.2745		4.2745
CS po	4.2745	4.2745	
Statystyka t	CS przed	CS w trakcie	CS po
CS przed		5.5279	10.3187
CS w trakcie	5.5279		4.7908
CS po	10.3187	4.7908	
Wartość p	CS przed	CS w trakcie	CS po
CS przed		<0.0001	<0.0001
CS w trakcie	<0.0001		<0.0001
CS po	<0.0001	<0.0001	
Jednorodne	CS przed(c)	CS w trakcie	CS po(a)
A			*
B		*	
C	*		

Trend liniowy	
Numer porządkowy dla [CS przed]	1
Numer porządkowy dla [CS w trakcie]	2
Numer porządkowy dla [CS po]	3
Statystyka F	106.4765
Wartość p dwustronne	<0.0001
Wartość p jednostronne	<0.0001

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p < 0.0001$ dla leczonych lekiem A (istotne statystycznie, $\alpha = 0.05$)

Wartości ciśnienia skurczowego różnią się istotnie między etapami leczenia lekiem A.

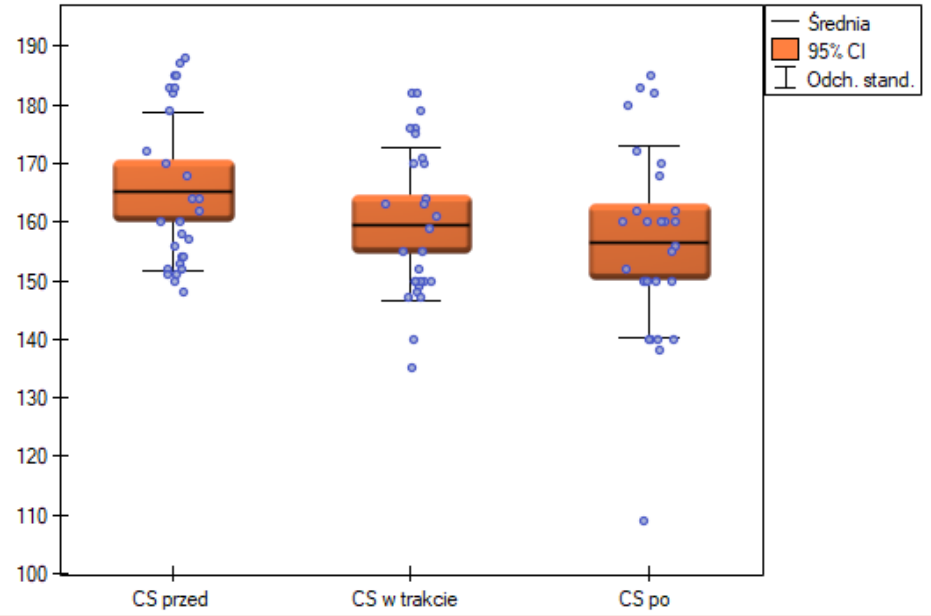
Ponad 66% wewnątrzosobowego zróżnicowania wyników można wytłumaczyć zastosowaniem leku A ($\eta^2 = 0.66$). Różnice dotyczą wszystkich porównywanych etapów leczenia (wynik POST-HOC). Istotny jest również trend spadku ciśnienia skurczowego na skutek leczenia ($p < 0.0001$) – lek A możemy więc uznać za skuteczny.

Wyniki dla leczonych lekiem B:

Korekta - GG (Greenhouse-Geisser)	
GG Epsilon	0.7558
df1	1.5117
df2	40.8149
Wartość p	0.0561
Korekta - HF (Huyhn-Feldt)	
HF Epsilon	0.7911
df1	1.5822
df2	42.7183
Wartość p	0.0536
Korekta - dolnej granicy Epsilon	
Dolna granica Epsilon	0.5
df1	1
df2	27
Wartość p	0.0768

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

MANOVA	
Wilk's Lambda	0.8331
Eta-kwadrat	0.1669
df1	2
df2	26
Statystyka F	2.6049
Wartość p	0.0931
Hotelling-Lawley Trace	0.2004
Eta-kwadrat	0.1669
df1	2
df2	26
Statystyka F	2.6049
Wartość p	0.0931
Pillai-Bartlett Trace	0.1669
Eta-kwadrat	0.1669
df1	2
df2	26
Statystyka F	2.6049
Wartość p	0.0931



$p > 0.05$ dla leczonych lekiem B, zarówno dla korekt epsilon jak i dla statystyki Lambda Wilksa (nieistotne statystycznie, $\alpha = 0.05$)

Brak istotnych różnic wartości ciśnienia skurczowego między etapami leczenia lekiem B.

Zaledwie niecałe 17% wewnątrzosobowego zróżnicowania wyników można wytłumaczyć zastosowaniem leku B ($\eta^2 = 0.17$).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

17.2 TESTY NIEPARAMETRYCZNE

ANOVA Kruskala-Wallisa

KIEDY STOSOWAĆ TEN TEST:

Jednoczynnikowa analiza wariancji dla rang Kruskala-Wallisa [98][99] to nieparametryczny odpowiednik ANOVA dla grup niezależnych, będący rozszerzeniem testu U Manna-Whitneya na więcej niż dwie populacje. Pozwala sprawdzić, czy rozkład badanej cechy (najczęściej: jej mediana) jest taki sam w co najmniej trzech niezależnych od siebie populacjach – stosuje się go zamiast ANOVA dla grup niezależnych, gdy nie są spełnione warunki tego testu.

WARUNKI STOSOWANIA

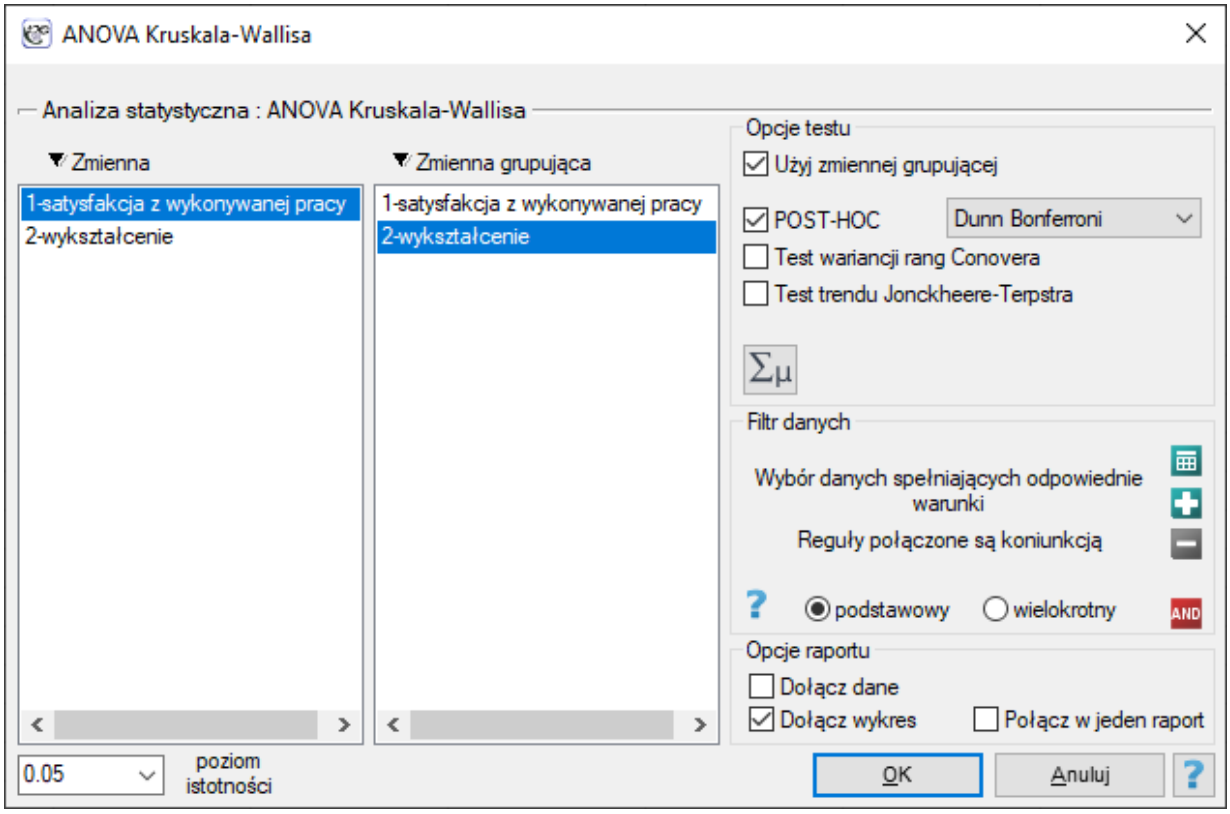
- pomiar na skali porządkowej lub interwałowej,
- model niezależny.

Test zakłada, że rozkłady badanej zmiennej w porównywanych populacjach są sobie zbliżone kształtem – równość wariancji rang można sprawdzić testem dla rang Conovera. Jeśli interesuje nas istnienie trendu w ułożeniu badanych grup, można dodatkowo wykonać test Jonckheere-Terpstra dla trendu.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → ANOVA Kruskala-Wallisa (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienną grupującą oraz zmienną wynikową.

Krok 2. Jeśli test wykaże istotną różnicę, dobierz test POST-HOC, np. Dunn lub Conover-Iman, wraz z odpowiednią korektą (np. Bonferroni).

Krok 3. Uruchom analizę.

WYKONANIE I INTERPRETACJA

Przykład 17.4. (plik satysfakcjaZpracy.pqs)
Przepytano grupę 120 osób, dla których wykonywane zajęcie jest ich pierwszą pracą uzyskaną po otrzymaniu odpowiedniego wykształcenia. Ankietowani oceniali satysfakcję z wykonywanej pracy w pięciostopniowej skali, gdzie: 1 – praca niesatysfakcjonująca, 2 – praca dająca niewielką satysfakcję, 3 – praca dająca przeciętny poziom satysfakcji, 4 – praca dająca dość dużą satysfakcję, 5 – praca bardzo satysfakcjonująca. Sprawdzimy, czy poziom deklarowanej satysfakcji z pracy różni się dla poszczególnych kategorii wykształcenia.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ANOVA Kruskala-Wallisa	
Czas analizy	0.21 sek.
Analizowane zmienne	satysfakcja z wykonywanej p
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Zmienna grupująca	wykształcenie
Stopnie swobody	3
Statystyka H (z poprawką dla rang wiązanych)	16.1954
Wartość p	0.001

$p = 0.001034$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Poziom satysfakcji z pracy różni się istotnie statystycznie pomiędzy kategoriami wykształcenia.

Przeprowadzona analiza POST-HOC Dunna z korektą Bonferroniego wskazuje, że istotne różnice dotyczą osób z wykształceniem podstawowym i średnim oraz z wykształceniem podstawowym i wyższym. Nieco więcej różnic potwierdza silniejszy test POST-HOC Conover-Iman.

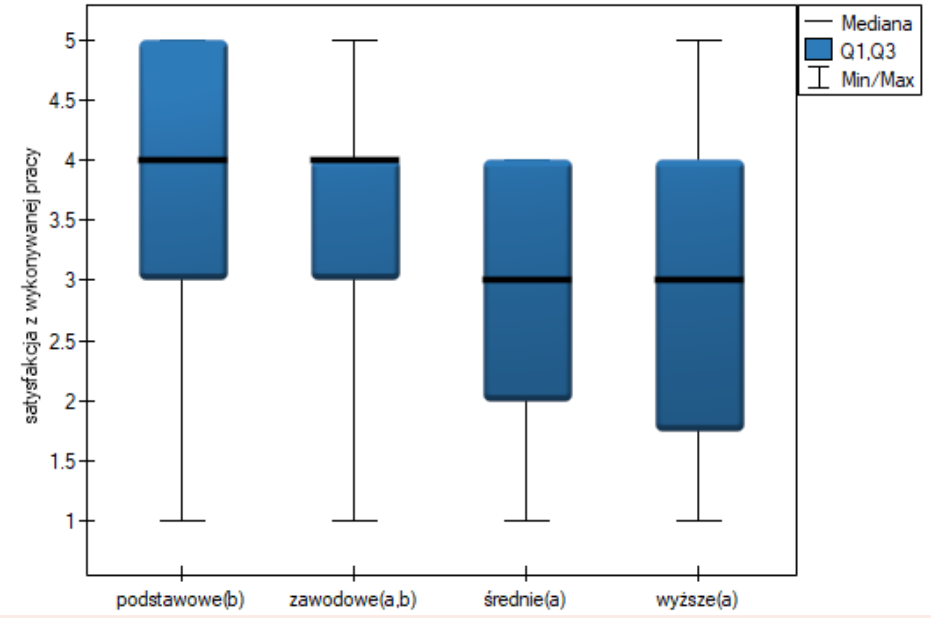
POST-HOC (Dunn Bonferroni)				
	podstawowe	zawodowe	średnie	wyższe
Różnica dla śr				
podstawowe		11.5833	30.4306	30.5208
zawodowe	11.5833		18.8472	18.9375
średnie	30.4306	18.8472		0.0903
wyższe	30.5208	18.9375	0.0903	
NIR	podstawowe	zawodowe	średnie	wyższe
podstawowe		23.5125	23.5125	25.7566
zawodowe	23.5125		21.0302	23.5125
średnie	23.5125	21.0302		23.5125
wyższe	25.7566	23.5125	23.5125	
Statystyka	podstawowe	zawodowe	średnie	wyższe
podstawowe		1.2997	3.4145	3.1263
zawodowe	1.2997		2.3644	2.1249
średnie	3.4145	2.3644		0.0101
wyższe	3.1263	2.1249	0.0101	
Wartość p	podstawowe	zawodowe	średnie	wyższe
podstawowe		1	0.0038	0.0106
zawodowe	1		0.1084	0.2016
średnie	0.0038	0.1084		1
wyższe	0.0106	0.2016	1	
Jednorodne	podstawowe	zawodowe(a	średnie(a)	wyższe(a)
A		*	*	*
B	*	*		

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

POST-HOC (Conover-Iman)

	podstawowe	zawodowe	średnie	wyższe
Różnica dla śr				
podstawowe		11.5833	30.4306	30.5208
zawodowe	11.5833		18.8472	18.9375
średnie	30.4306	18.8472		0.0903
wyższe	30.5208	18.9375	0.0903	
NIR	podstawowe	zawodowe	średnie	wyższe
podstawowe		16.6173	16.6173	18.2034
zawodowe	16.6173		14.863	16.6173
średnie	16.6173	14.863		16.6173
wyższe	18.2034	16.6173	16.6173	
Statystyka	podstawowe	zawodowe	średnie	wyższe
podstawowe		1.3806	3.627	3.3208
zawodowe	1.3806		2.5116	2.2572
średnie	3.627	2.5116		0.0108
wyższe	3.3208	2.2572	0.0108	
Wartość p	podstawowe	zawodowe	średnie	wyższe
podstawowe		0.17	0.0004	0.0012
zawodowe	0.17		0.0134	0.0259
średnie	0.0004	0.0134		0.9914
wyższe	0.0012	0.0259	0.9914	
Jednorodne	podstawowe	zawodowe(b	średnie(a)	wyższe(a)
A			*	*
B	*	*		

Na wykresie przedstawiającym mediany i kwartyle widać grupy jednorodne wyznaczone przez test POST-HOC. Dla wyniku Dunna z korektą Bonferroniego widoczne są dwie grupy jednorodne, które nie są całkowicie odrębne: grupa (a) – osoby słabiej oceniające satysfakcję z pracy i grupa (b) – osoby lepiej ją oceniające. Wykształcenie zawodowe przynależy do obu tych grup, co oznacza, że osoby z tym wykształceniem oceniają satysfakcję z pracy dość różnorodnie.



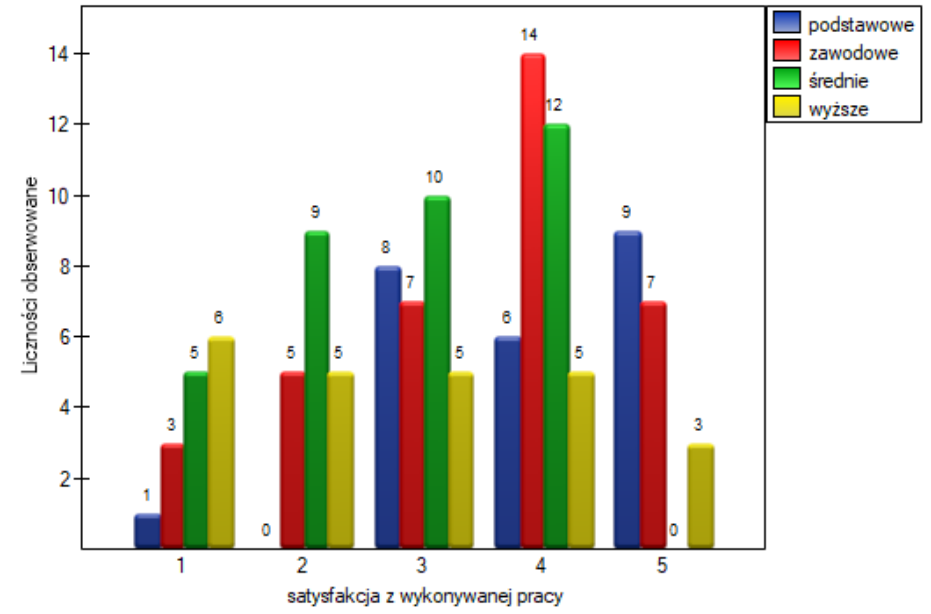
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dokładny opis danych możemy przedstawić, wybierając w oknie analizy statystyki opisowe $\Sigma\mu$ i wskazując na dodanie do opisu liczności i procentów.

Podsumowanie				
Grupa	podstawowe	zawodowe	średnie	wyższe
Liczność	24	36	36	24
Mediana	4	4	3	3
Dolny kwartyl	3	3	2	1.75
Górny kwartyl	5	4	4	4

Liczność(Procent)				
Grupa	podstawowe	zawodowe	średnie	wyższe
1	1 (4.17%)	3 (8.33%)	5 (13.89%)	6 (25%)
2	0 (0%)	5 (13.89%)	9 (25%)	5 (20.83%)
3	8 (33.33%)	7 (19.44%)	10 (27.78%)	5 (20.83%)
4	6 (25%)	14 (38.89%)	12 (33.33%)	5 (20.83%)
5	9 (37.5%)	7 (19.44%)	0 (0%)	3 (12.5%)
razem	24	36	36	24

Rozkład odpowiedzi możemy też przedstawić na wykresie liczności.



DLA ZAINTERESOWANYCH

Hipotezy: dotyczą równości średnich rang dla kolejnych populacji lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0: &\quad \phi_1 = \phi_2 = ... = \phi_k, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \phi_j \text{ są sobie równe } (j = 1, 2, ..., k), \end{aligned}$$

gdzie $\phi_1, \phi_2, ...\phi_k$ to rozkłady badanej zmiennej w populacjach, z których pobrano próby.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wzory: statystyka testowa ma postać:

$$H = \frac{1}{C}\left(\frac{12}{N(N+1)}\sum_{j=1}^{k}\left(\frac{\left(\sum_{i=1}^{n_j} R_{ij}\right)^2}{n_j}\right) - 3(N+1)\right),$$

gdzie:
$N = \sum_{j=1}^{k} n_j$,
n_j – liczności prób dla $(j = 1, 2, ...k)$,
R_{ij} – rangi przypisane do wartości zmiennej, dla $(i = 1, 2, ...n_j)$, $(j = 1, 2, ...k)$,
$C = 1 - \frac{\sum(t^3 - t)}{N^3 - N}$ – korekta na rangi wiązane,
t – liczba przypadków wchodzących w skład rangi wiązanej. Poprawka C jest stosowana, gdy rangi wiązane występują (w przeciwnym razie $C = 1$).

Statystyka H ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody $df = (k-1)$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Testy POST-HOC

Wprowadzenie do kontrastów i testów POST-HOC znajduje się w rozdziale 17.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Dunna
Dla porównań prostych, zarówno równolicznych jak i różnolicznych grup (Dunn, 1964[52]). Zawiera poprawkę na rangi wiązane (Zar, 2010[187]) i jest korygowany ze względu na wielokrotne testowanie – najczęściej korektą Bonferroniego lub Sidaka (inne korekty opisano w dziale Wielokrotne porównania).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = Z_{\alpha(corrected)}\sqrt{\left(\frac{N(N+1)}{12} - \frac{\sum(t^3 - t)}{12(N-1)}\right)\sum_{j=1}^{k}\frac{c_j^2}{n_j}},$$

gdzie t to liczba przypadków wchodzących w skład rangi wiązanej, a $Z_{\alpha(corrected)}$ to wartość krytyczna (statystyka) rozkładu normalnego dla poziomu istotności α skorygowanego o liczbę możliwych porównań prostych c zgodnie z wybraną poprawką.

(ii) Statystyka testowa ma postać:

$$Z = \frac{\sum_{j=1}^{k} c_j \overline{R}_j}{\sqrt{\left(\frac{N(N+1)}{12} - \frac{\sum(t^3-t)}{12(N-1)}\right)\sum_{j=1}^{k}\frac{c_j^2}{n_j}}},$$

gdzie $\overline{R}_j$ – średnia rang j-tej grupy, dla $(j = 1, 2, ...k)$. Statystyka ta ma asymptotycznie (dla dużych liczności próby) rozkład normalny, a wartość p jest korygowana o liczbę możliwych porównań prostych c zgodnie z wybraną poprawką.

Test Conover-Inman
Nieparametryczny odpowiednik LSD Fishera[47], stosowany dla porównań prostych, zarówno równolicznych jak i różnolicznych grup.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,1,N-k}} \cdot \sqrt{S^2 \frac{N-1-H}{N-k} \sum_{j=1}^{k} \frac{c_j^2}{n_j}},$$

gdzie $S^2 = \frac{1}{N-1}\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j} R_{ij}^2 - N\frac{(N+1)^2}{4}\right)$, a $F_{\alpha,1,N-k}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz dla stopni swobody odpowiednio: 1 i $N-k$.

(ii) Statystyka testowa ma postać:

$$t = \frac{\sum_{j=1}^{k} c_j \overline{R}_j}{\sqrt{S^2 \frac{N-1-H}{N-k} \sum_{j=1}^{k} \frac{c_j^2}{n_j}}},$$

gdzie $\overline{R}_j$ – średnia rang j-tej grupy, dla $(j = 1, 2, ...k)$. Statystyka ta podlega rozkładowi t-Studenta z $N-k$ stopniami swobody.

Test Jonckheere-Terpstra dla trendu

KIEDY STOSOWAĆ TEN TEST:

Test Jonckheere-Terpstra dla trendu, opisany niezależnie przez Jonckheere (1954)[92] i Terpstra (1952)[163], można wykonać w tej samej sytuacji, co ANOVA Kruskala-Wallisa, ponieważ bazuje na tych samych założeniach. W odróżnieniu od ANOVA Kruskala-Wallisa nie sprawdza on jednak dowolnych różnic między grupami, lecz to, czy mediany kolejnych, uporządkowanych populacji rosną lub maleją zgodnie z oczekiwanym kierunkiem.

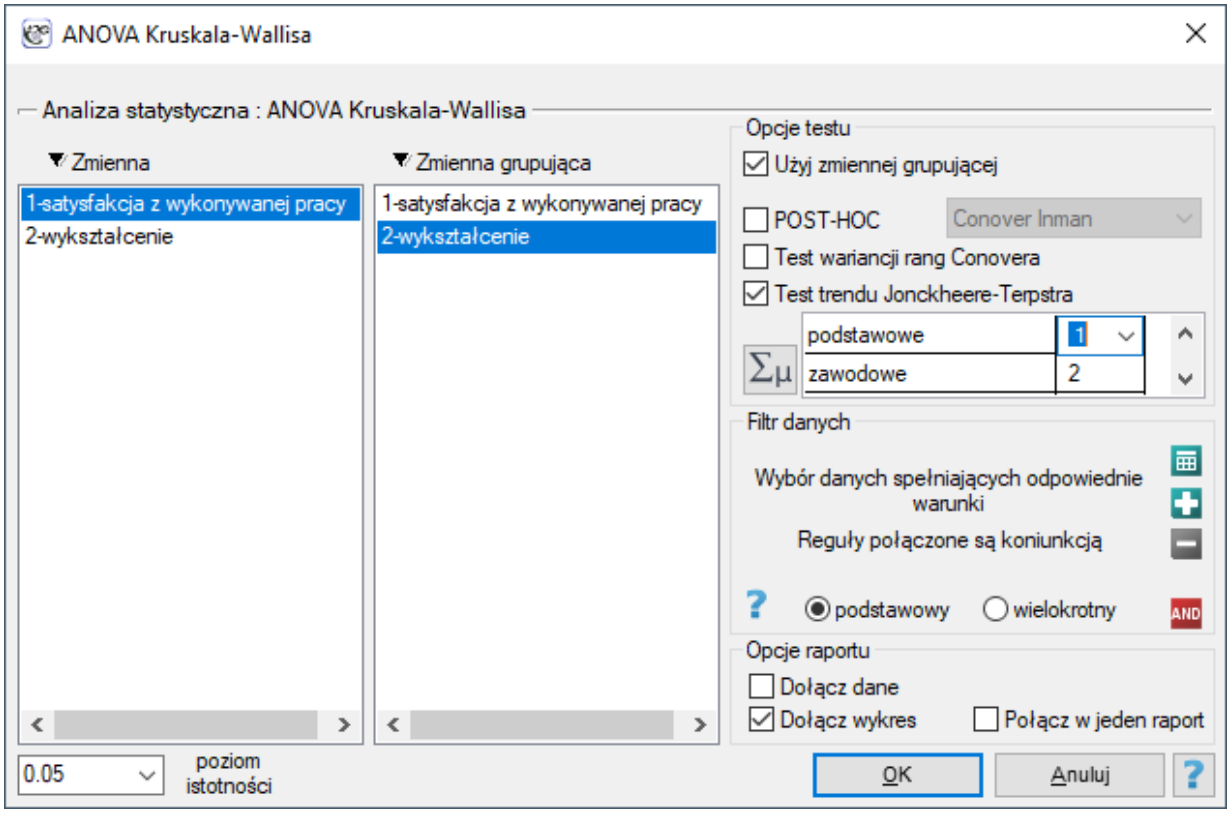
WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej,
- model niezależny,
- znana lub zakładana kolejność porównywanych populacji.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → ANOVA Kruskala-Wallisa (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wznów analizę przyciskiem .

Krok 2. Zaznacz opcję Test trendu Jonckheere-Terpstra.

Krok 3. Wskaż oczekiwaną kolejność populacji, przypisując im kolejne liczby naturalne.

Krok 4. Uruchom analizę.

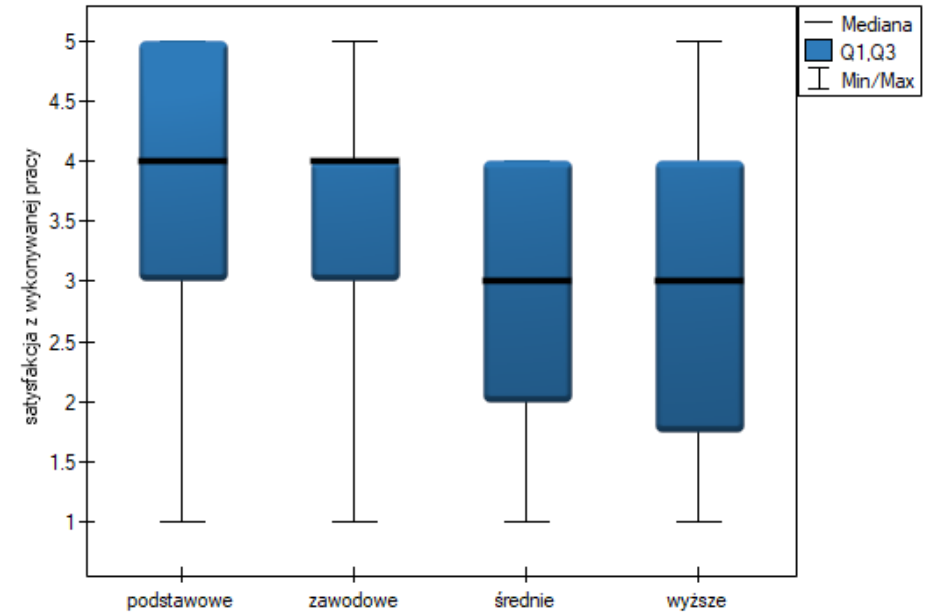
Uwaga! Określenie „z co najmniej jedną nierównością ścisłą" w hipotezie alternatywnej tego testu oznacza, że mediana co najmniej jednej populacji powinna być większa niż mediana innej populacji, w oczekiwanej kolejności.

WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 17.4 (plik satysfakcjaZpracy.pqs): podejrzewa się, że osoby lepiej wykształcone mają wyższe wymagania zawodowe, co może zmniejszać poziom satysfakcji z pierwszej pracy, która często takich wymagań nie spełnia. Sprawdzimy więc, czy poziom satysfakcji z pierwszej pracy zmienia się zgodnie z oczekiwanym kierunkiem wraz z poziomem wykształcenia.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test trendu Jonckheere-Terpstra	
Numer porządkowy dla [podstawowe]	1
Numer porządkowy dla [zawodowe]	2
Numer porządkowy dla [średnie]	3
Numer porządkowy dla [wyższe]	4
Statystyka L	1856
Statystyka Z	3,915719
Wartość p jednostronne	0,000045
Wartość p dwustronne	0,00009

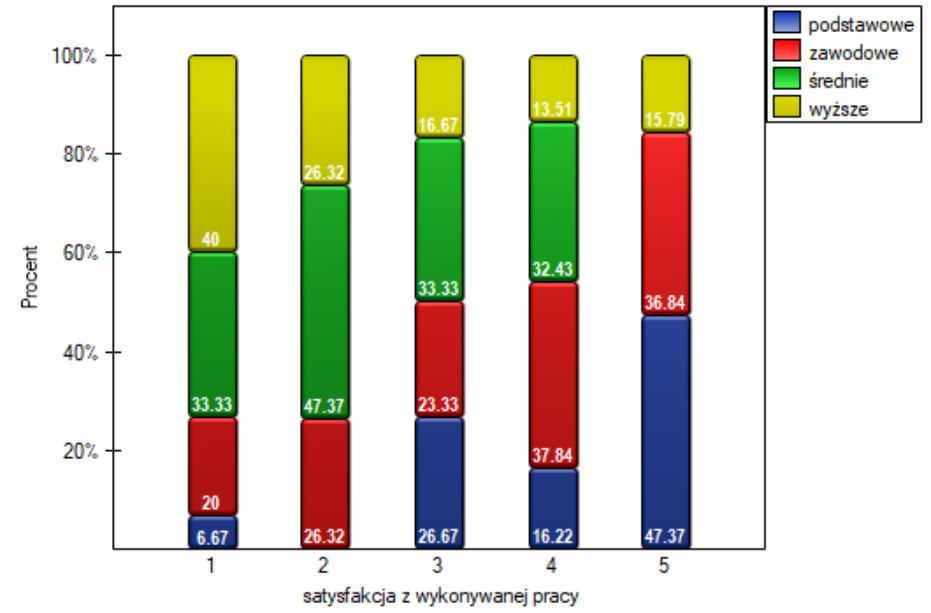


$p = 0.000045$ **(jednostronna, istotna statystycznie,** $\alpha = 0.05$**)**

Poziom satysfakcji z pierwszej pracy rzeczywiście zmienia się zgodnie z oczekiwanym kierunkiem trendu.

Istnienie tego trendu możemy również potwierdzić, przedstawiając procentowy rozkład uzyskanych odpowiedzi.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



DLA ZAINTERESOWANYCH

Hipotezy (upraszczane do median):

$$\begin{aligned}\mathcal{H}_0: &\quad \theta_1 = \theta_2 = ... = \theta_k,\\ \mathcal{H}_1: &\quad \theta_1 \geq \theta_2 \geq ... \geq \theta_k, \text{ z co najmniej jedną nierównością ścisłą}\end{aligned}$$

Wzory: statystyka testowa ma postać:

$$Z = \frac{L - \left[\frac{N^2 - \sum_{j=1}^{k} n_j^2}{4}\right]}{SE}$$

gdzie:
L – suma wartości l_{ij} uzyskanych dla każdej pary porównywanych populacji,
l_{ij} – liczba wyników wyższych niż zadana wartość w grupie występującej w następnej kolejności,
$SE = \sqrt{\frac{A}{72} + \frac{B}{36N(N-1)(N-2)} + \frac{C}{8N(N-1)}}$,
$A = N(N-1)(2N+5) - \sum_{j=1}^{k} n_j(n_j-1)(2n_j+5) - \sum_{l=1}^{g} t_l(t_l-1)(2t_l+5)$,
$B = \sum_{j=1}^{k} n_j(n_j-1)(n_j-2) \cdot \sum_{l=1}^{g} t_l(t_l-1)(t_l-2)$,
$C = \sum_{j=1}^{k} n_j(n_j-1) \cdot \sum_{l=1}^{g} t_l(t_l-1)$,
g – liczba grup różnych rang wiązanych,
t_l – liczba przypadków wchodzących w skład rangi wiązanej,
$N = \sum_{j=1}^{k} n_j$,
n_j – liczności prób dla $(j = 1, 2, ...k)$.

Interpretacja: wzór na statystykę testową Z zawiera poprawkę na rangi wiązane. Poprawka ta jest stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych, wzór sprowadza się do oryginalnej

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

formuły Jonckheere-Terpstra bez tej poprawki). Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Przy znanym oczekiwanym kierunku trendu hipoteza alternatywna jest jednostronna i interpretacji podlega jednostronna wartość p. Interpretacja dwustronnej wartości p oznacza, że badacz nie zna (nie zakłada) kierunku ewentualnego trendu.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test wariancji rang Conovera

KIEDY STOSOWAĆ TEN TEST:

Test kwadratów rang Conovera pozwala sprawdzić, czy rozproszenie (zróżnicowanie) badanej cechy jest podobne w kilku populacjach – podobnie jak test Fishera-Snedecora (dla dwóch populacji) czy testy Levene'a i Browna-Forsythe'a (dla co najmniej dwóch populacji). Jest ich nieparametrycznym odpowiednikiem: nie zakłada normalności rozkładu danych i opiera się na rangach[47]. Stosuje się go zwykle jako uzupełnienie ANOVA Kruskala-Wallisa, gdyż test ten bada zróżnicowanie danych (odległości od średniej), a nie ich poziom.

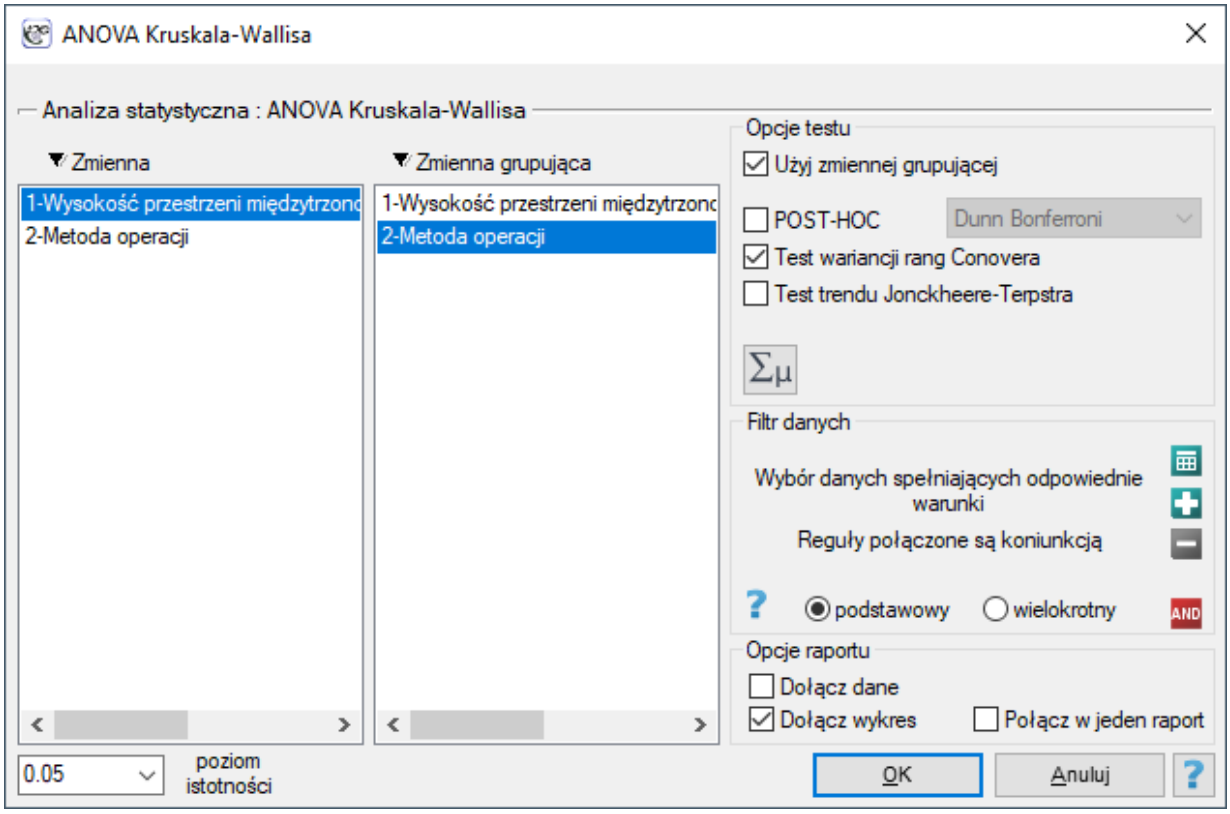
WARUNKI STOSOWANIA

- pomiar na skali interwałowej.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → ANOVA Kruskala-Wallisa, opcja test wariancji rang Conovera, lub Statystyka → Testy nieparametryczne → Mann-Whitney, opcja test wariancji rang Conovera.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

PRZYKŁAD 17.5. (plik metodaOperacji.pqs)
Przygotowano chorych do operacji kręgosłupa. Chorzy będą operowani jedną z trzech metod. Dokonano wstępnego przydziału każdej osoby chorej do poszczególnych typów operacji. Na późniejszym etapie zamierzamy porównywać stan chorych po przebytych operacjach, dlatego zależy nam, by grupy chorych były porównywalne – powinny być podobne między innymi pod względem wysokości przestrzeni międzytrzonowej (WPMT) przed operacją. Podobieństwo powinno dotyczyć nie tylko wartości przeciętnych, ale również zróżnicowania grup.

Sprawdzono rozkład danych.

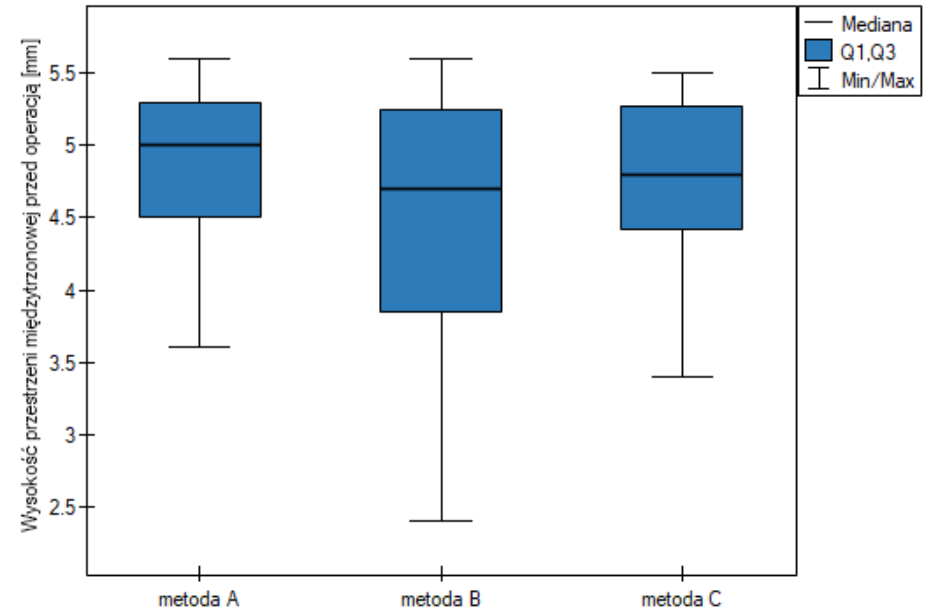
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Normalność jednowymiarowa	ID1	ID2	ID3
Czas analizy	0.12 sek.	0.12 sek.	0.15 sek.
Analizowane zmienne	Wysokość prze	Wysokość prze	Wysokość prze
Liczność - niezinterpretowane	0	0	0
Liczność - braki danych	0	0	0
Poziom istotności	0.05	0.05	0.05
Filtr danych	Metoda operac	Metoda operac	Metoda operac
Liczność - grupy	21	55	46
Średnia grupy	4.8762	4.5218	4.7957
Odchylenie standardowe grupy	0.5718	0.8315	0.5296
Test Shapiro-Wilka			
Statystyka W	0.9178	0.931	0.9309
Statystyka Z	1.4157	2.6875	2.3618
Wartość p	0.0784	0.0036	0.0091
Test D'Agostino-Pearson			
Statystyka K-kwadrat	2.9229	4.1294	3.6689
Stopnie swobody	2	2	2
Wartość p	0.2319	0.1269	0.1597
Test skośności			
Skośność	-0.8514	-0.6489	-0.6467
Statystyka Z	-1.7051	-1.9901	-1.8381
Wartość p	0.0882	0.0466	0.0661
Test kurtozy (g2)			
Kurtoza	-0.0954	-0.3252	0.2169
Statystyka Z	0.124	-0.4109	0.5389
Wartość p	0.9013	0.6811	0.59

Okazuje się, że dla dwóch metod operacji WPMT wykazuje odstępstwa od normalności, spowodowane w dużej mierze skośnością danych. Dalszą analizę porównawczą przeprowadzimy przy pomocy testu Kruskala-Wallisa, by sprawdzić, czy poziom WPMT różni się pomiędzy metodami, oraz testu Conovera, który wskaże, czy w każdej metodzie rozpiętość wyników WPMT jest podobna.

ANOVA Kruskala-Wallisa	
Czas analizy	0.12 sek.
Analizowane zmienne	Wysokość przestrzeni międz
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Zmienna grupująca	Metoda operacji
Stopnie swobody	2
Statystyka H (z poprawką dla rang wiązanych)	3.1627
Wartość p	0.2057
Test wariancji rang Conovera	
Statystyka Chi-kwadrat	12.2712
Stopnie swobody	2
Wartość p	0.0022

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p = 0.0022$ dla testu wariancji Conovera (istotne statystycznie, $\alpha = 0.05$), $p = 0.2057$ dla testu Kruskala-Wallisa (nieistotne statystycznie)

Zakresy WPMT różnią się istotnie między metodami operacji, natomiast poziom WPMT – nie.

Najpierw interpretacji podlega wynik testu wariancji Conovera, który wskazuje na istotne statystycznie różnice w zakresach porównywanych grup. Z wykresu możemy wnioskować, że różnice dotyczą głównie grupy 3. Ponieważ wykryto różnice w zakresie WPMT, interpretacja wyniku testu Kruskala-Wallisa (porównującego poziom WPMT dla tych metod) powinna być ostrożna, gdyż test ten jest wrażliwy na niejednorodność wariancji. Mimo że test Kruskala-Wallisa nie wykazał istotnych różnic, zaleca się, by chorych o niskich wartościach WPMT (przypisanych głównie do operacji metodą B) rozmieścić bardziej równomiernie, tzn. sprawdzić, czy nie można im zaproponować operacji metodą A lub C. Po ponownym przydziale chorych analizę należy powtórzyć.

DLA ZAINTERESOWANYCH

Hipotezy:

$\mathcal{H}_0$: rozproszenie danych w porównywanych populacjach jest takie samo,
$\mathcal{H}_1$: przynajmniej dwie populacje różnią się wielkością rozproszenia danych.

Wzory: Statystyka testowa ma postać:

$$\chi^2 = \frac{1}{D^2}\left(\sum_{j=1}^{k}\frac{S_j^2}{n_j} - N\overline{S}^2\right)$$

gdzie:
$N = n_1 + n_2 + ... + n_k$,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

n_j – liczność w poszczególnych grupach,
S_j – suma kwadratów rang w j-tej grupie,
$\overline{S} = \frac{1}{N}\sum_{j=1}^{k} S_j$ – średnia wszystkich kwadratów rang,
$D^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} R_i^4 - N\overline{S}^2\right)$,
R_i – rangi dla wartości przedstawiających odległość pomiaru od średniej danej grupy.

Statystyka ta ma rozkład χ^2 z $k-1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

ANOVA Friedmana

KIEDY STOSOWAĆ TEN TEST:

Analiza wariancji powtarzanych pomiarów dla rang Friedmana (Friedman, 1937[66]) to nieparametryczny odpowiednik ANOVA powtarzanych pomiarów. Stosuje się ją, gdy pomiarów badanej cechy dokonujemy kilkukrotnie na tych samych obiektach, a nie są spełnione warunki testu parametrycznego – także wtedy, gdy dysponujemy rankingami pochodzącymi od różnych sędziów i zależy nam na ocenie ich zgodności. Modyfikacja Imana-Davenport (1980[89]) tej statystyki jest obecnie rekomendowana w miejsce tradycyjnej statystyki Friedmana, która bywa nadmiernie konserwatywna.

WARUNKI STOSOWANIA

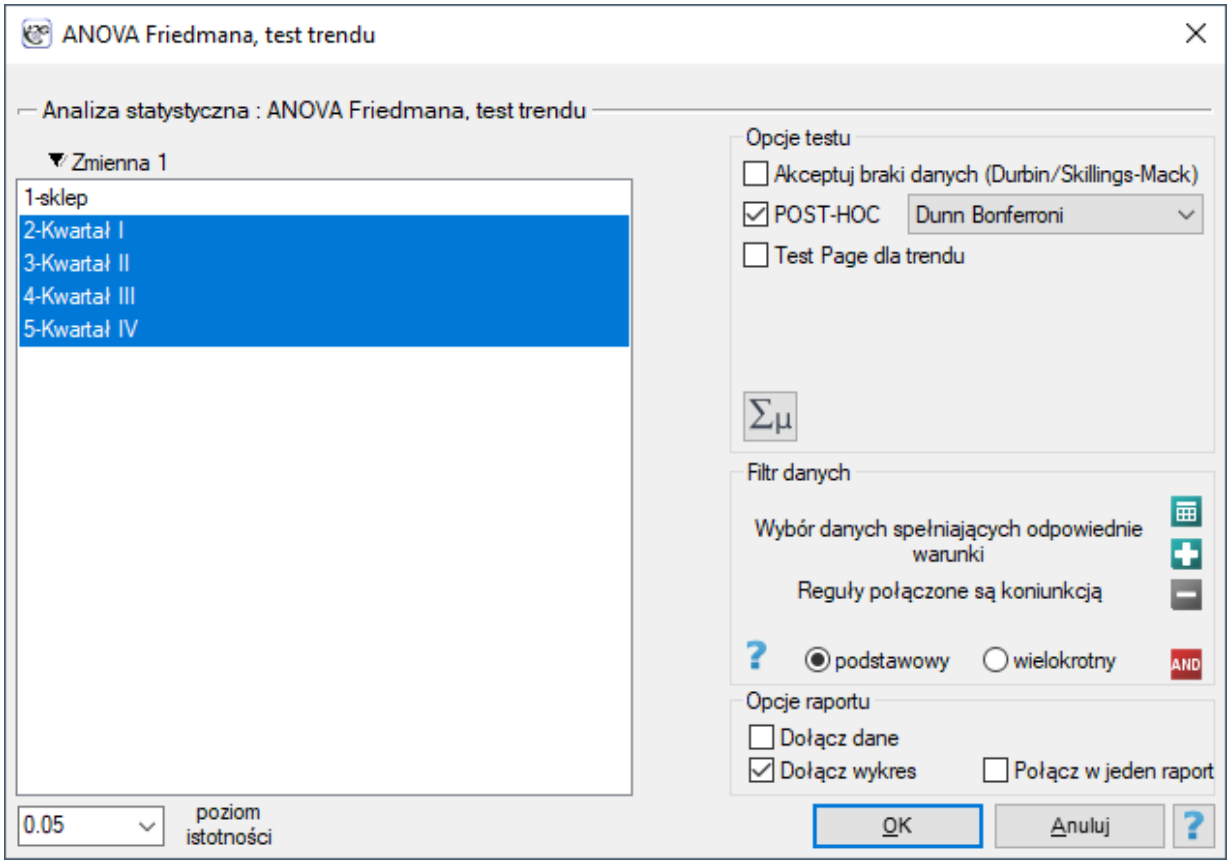
- pomiar na skali porządkowej lub interwałowej,
- model zależny.

Gdy w danych występują braki, można zamiast tego wyliczyć ANOVA Durbina (braki zbalansowane) lub ANOVA Skillings-Mack (braki w dowolnym układzie). Jeśli interesuje nas istnienie trendu w ułożeniu badanych grup, można dodatkowo wykonać test Page'a dla trendu.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → ANOVA Friedmana (możliwość braków danych) (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienne odpowiadające kolejnym pomiarom.

Krok 2. Jeśli test wykaże istotną różnicę, dobierz test POST-HOC, np. Dunn lub Conover-Iman, wraz z odpowiednią korektą (np. Bonferroni).

Krok 3. Uruchom analizę.

WYKONANIE I INTERPRETACJA

Przykład 17.6. (plik baton.pqs)
Badano kwartalną wielkość sprzedaży pewnego batonu czekoladowego w 14 losowo wybranych marketach. Badanie rozpoczęto w styczniu, a zakończono w grudniu. W czasie drugiego kwartału trwała intensywna billboardowa kampania reklamowa tego produktu. Sprawdzimy, czy kampania miała wpływ na wielkość sprzedaży reklamowanego batonu.

sklep	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
SK1	3415	4556	5772	5432
SK2	1593	1937	2242	2794
SK3	1976	2056	2240	2085
SK4	1526	1594	1644	1705
SK5	1538	1634	1866	1769
SK6	983	1086	1135	1177
SK7	1050	1209	1245	977
SK8	1861	2087	2054	2018
SK9	1714	2415	2361	2424
SK10	1320	1621	1624	1551
SK11	1276	1377	1522	1412
SK12	1263	1275	1350	1490
SK13	1271	1417	1583	1513
SK14	1436	1310	1357	1468

Copyright © 2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

ANOVA Friedmana, test trendu	
Czas analizy	0.12 sek.
Analizowane zmienne	Kwartał I;Kwartał II;Kwartał
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Akceptuj braki danych (Durbin/Skillings-Mack)	Nie
C	350
A	420
Statystyka T1 Friedman	23.9143
Stopnie swobody	3
Wartość p	<0.0001
Statystyka T2 Iman-Davenport	17.1896
Stopnie swobody	3 / 39
Wartość p	<0.0001
Statystyka Skillings-Mack	23.9143
Stopnie swobody	3
Wartość p	<0.0001

$p < 0.05$ **dla testu Friedmana i dla korekty Iman-Davenport, istotne statystycznie,** $\alpha = 0.05$

Sprzedaż batonu nie jest taka sama w każdym kwartale.

Wykonana analiza POST-HOC Dunna z korektą Bonferroniego wskazuje na różnice wielkości sprzedaży dotyczące kwartału I i III oraz I i IV, a analogiczna analiza przeprowadzona silniejszym testem Conover-Iman wskazuje na różnice pomiędzy wszystkimi kwartałami za wyjątkiem kwartału III i IV.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

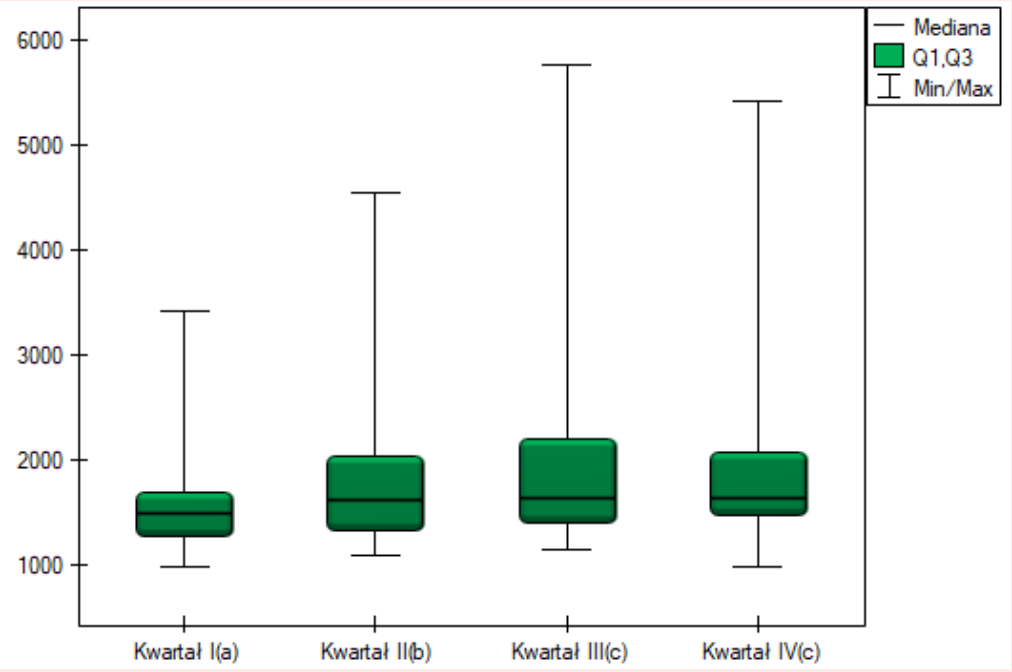
POST-HOC (Dunn Bonferroni)				
	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Różnica dla śr				
Kwartał I		1.0714	2.1429	1.9286
Kwartał II	1.0714		1.0714	0.8571
Kwartał III	2.1429	1.0714		0.2143
Kwartał IV	1.9286	0.8571	0.2143	
NIR	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Kwartał I		1.2873	1.2873	1.2873
Kwartał II	1.2873		1.2873	1.2873
Kwartał III	1.2873	1.2873		1.2873
Kwartał IV	1.2873	1.2873	1.2873	
Statystyka	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Kwartał I		2.1958	4.3916	3.9524
Kwartał II	2.1958		2.1958	1.7566
Kwartał III	4.3916	2.1958		0.4392
Kwartał IV	3.9524	1.7566	0.4392	
Wartość p	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Kwartał I		0.1686	0.0001	0.0005
Kwartał II	0.1686		0.1686	0.4739
Kwartał III	0.0001	0.1686		1
Kwartał IV	0.0005	0.4739	1	
Jednorodne	Kwartał I(a)	Kwartał II(a,	Kwartał III(b	Kwartał IV(b
A	*	*		
B		*	*	*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

POST-HOC (Conover-Iman)

	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Różnica dla su				
Kwartał I		15	30	27
Kwartał II	15		15	12
Kwartał III	30	15		3
Kwartał IV	27	12	3	
NIR	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Kwartał I		9.4095	9.4095	9.4095
Kwartał II	9.4095		9.4095	9.4095
Kwartał III	9.4095	9.4095		9.4095
Kwartał IV	9.4095	9.4095	9.4095	
Statystyka	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Kwartał I		3.2244	6.4489	5.804
Kwartał II	3.2244		3.2244	2.5795
Kwartał III	6.4489	3.2244		0.6449
Kwartał IV	5.804	2.5795	0.6449	
Wartość p	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Kwartał I		0.0026	<0.0001	<0.0001
Kwartał II	0.0026		0.0026	0.0138
Kwartał III	<0.0001	0.0026		0.5228
Kwartał IV	<0.0001	0.0138	0.5228	
Jednorodne	Kwartał I(a)	Kwartał II(b	Kwartał III(c	Kwartał IV(c
A	*			
B		*		
C			*	*

Na wykresie przedstawiono grupy jednorodne wyznaczone testem Conover-Iman.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dokładny opis danych możemy przedstawić, wybierając w oknie analizy statystyki opisowe $\Sigma\mu$.

Podsumowanie				
Grupa	Kwartał I	Kwartał II	Kwartał III	Kwartał IV
Liczność	14	14	14	14
Mediana	1481	1607.5	1634	1628
Minimum	983	1086	1135	977
Maksimum	3415	4556	5772	5432
Dolny kwartyl	1272.25	1326.75	1398.25	1473.5
Górny kwartyl	1683.75	2026.25	2193.5	2068.25

Gdyby dane były opisane skalą porządkową o niewielu kategoriach, warto by było przedstawić je również w licznościach i procentach – w tym przykładzie nie byłaby to jednak dobra metoda opisu.

DLA ZAINTERESOWANYCH

Hipotezy: dotyczą równości sumy rang dla kolejnych pomiarów lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0: &\quad \theta_1 = \theta_2 = ... = \theta_k, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2, ..., k), \end{aligned}$$

gdzie $\theta_1, \theta_2, ...\theta_k$ – mediany badanej cechy w kolejnych pomiarach z badanej populacji.

Wzory: wyznacza się dwie statystyki testowe: statystykę Friedmana

$$T_1 = \frac{1}{C}\left(\frac{12}{nk(k+1)}\left(\sum_{j=1}^{k}\left(\sum_{i=1}^{n} R_{ij}\right)^2\right) - 3n(k+1)\right),$$

oraz jej modyfikację Imana-Davenport

$$T_2 = \frac{(n_j - 1)T_1}{n_j(k-1) - T_1},$$

gdzie:
n – liczność próby,
R_{ij} – rangi przypisane kolejnym pomiarom $(j = 1, 2, ...k)$, oddzielnie dla każdego z badanych obiektów $(i = 1, 2, ...n)$,
$C = 1 - \frac{\sum(t^3 - t)}{n(k^3 - k)}$ – korekta na rangi wiązane,
t – liczba przypadków wchodzących w skład rangi wiązanej. Poprawka C jest stosowana, gdy rangi wiązane występują (w przeciwnym razie $C = 1$).

Statystyka T_1 ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $df = k - 1$ stopniami swobody. Statystyka T_2 podlega rozkładowi F Snedecora z $df_1 = k-1$ i $df_2 = (n_j-1)(k-1)$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Testy POST-HOC

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Wprowadzenie do kontrastów i testów POST-HOC znajduje się w rozdziale 17.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Dunna
Stosowany dla porównań prostych (liczność w poszczególnych pomiarach zawsze jest taka sama). Test Dunna (Dunn, 1964[52]) jest korygowany ze względu na wielokrotne testowanie – najczęściej korektą **Bonferroniego** lub **Sidaka** (inne korekty opisano w dziale Wielokrotne porównania).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = Z_{\alpha(corrected)}\sqrt{\frac{k(k+1)}{6n}},$$

gdzie $Z_{\alpha(corrected)}$ to wartość krytyczna (statystyka) rozkładu normalnego dla poziomu istotności α skorygowanego o liczbę możliwych porównań prostych c zgodnie z wybraną poprawką.

(ii) Statystyka testowa ma postać:

$$Z = \frac{\sum_{j=1}^{k} c_j \overline{R}_j}{\sqrt{\frac{k(k+1)}{6n}}},$$

gdzie $\overline{R}_j$ – średnia rang j-tego pomiaru, dla $(j = 1, 2, ...k)$. Statystyka ta ma asymptotycznie (dla dużych liczności próby) rozkład normalny, a wartość p jest korygowana o liczbę możliwych porównań prostych zgodnie z wybraną poprawką.

Test Conover-Inman
Nieparametryczny odpowiednik LSD Fishera[47], stosowany dla porównań prostych (liczność w poszczególnych pomiarach zawsze jest taka sama).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha,1,df_2}} \cdot \sqrt{\frac{2\left(n_j A - \sum_{j=1}^{t} R_j^k\right)}{(n_j - 1)(k-1)}},$$

gdzie $A = \sum_{i=1}^{n_j}\sum_{j=1}^{k} R_{ij}^2$ – suma kwadratów dla rang, a $F_{\alpha,1,df_2}$ to wartość krytyczna (statystyka) rozkładu F Snedecora dla zadanego poziomu istotności α oraz dla stopni swobody odpowiednio: 1 i df_2.

(ii) Statystyka testowa ma postać:

$$t = \frac{\sum_{j=1}^{k} c_j R_j}{\sqrt{\frac{2\left(n_j A - \sum_{j=1}^{t} R_j^k\right)}{(n_j - 1)(k-1)}}},$$

gdzie R_j – suma rang j-tego pomiaru, dla $(j = 1, 2, ...k)$. Statystyka ta podlega rozkładowi t-Studenta z df_2 stopniami swobody.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test Page dla trendu

KIEDY STOSOWAĆ TEN TEST:

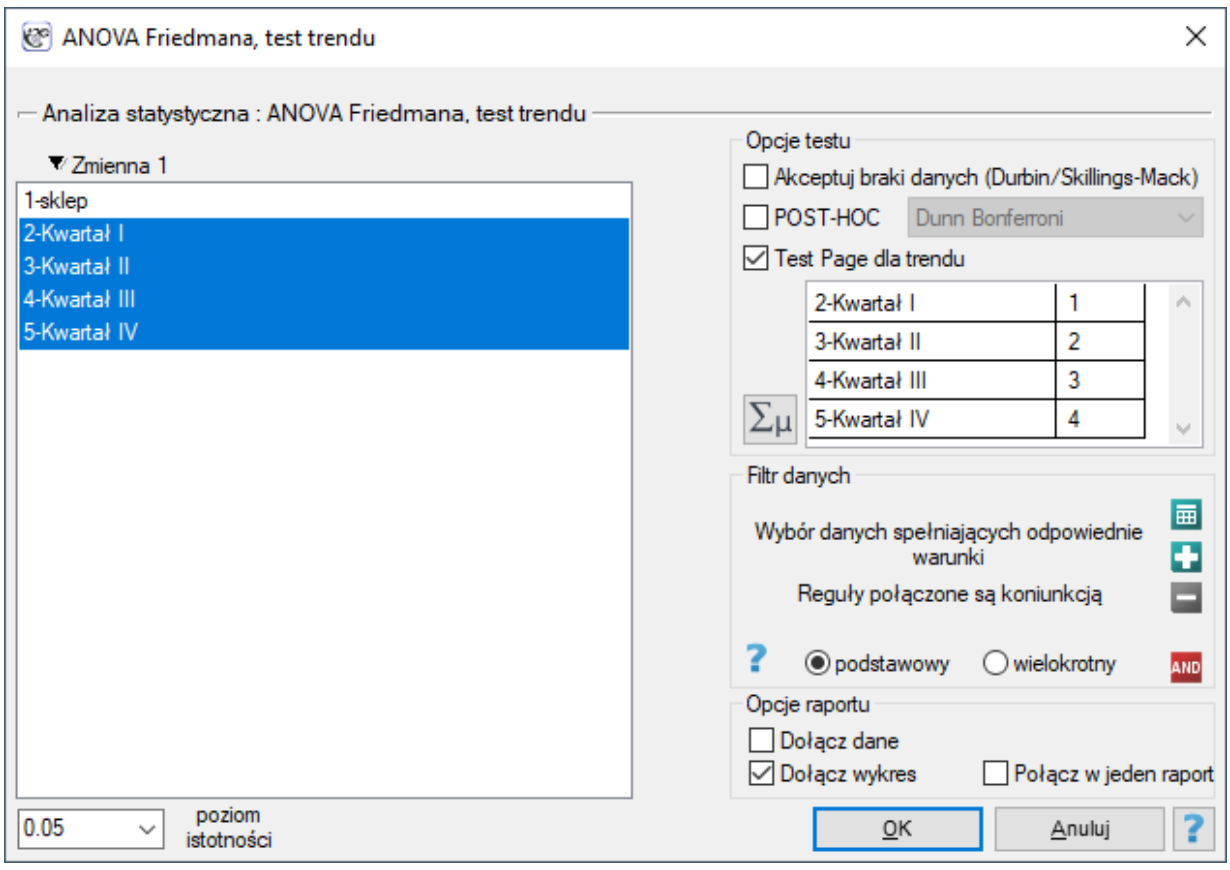
Test Page dla trendu, opisany w 1963 roku przez Page'a[133], można wykonać w tej samej sytuacji, co ANOVA Friedmana, ponieważ bazuje na tych samych założeniach. W odróżnieniu od ANOVA Friedmana nie sprawdza on jednak dowolnych różnic między pomiarami, lecz to, czy mediany kolejnych pomiarów rosną lub maleją zgodnie z oczekiwanym kierunkiem.

WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej,
- model zależny,
- znana lub zakładana kolejność pomiarów.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → ANOVA Friedmana (możliwość braków danych) (lub poprzez Kreator).



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krok 1. Wskaż zmienne odpowiadające kolejnym pomiarom.

Krok 2. Zaznacz opcję Test Page dla trendu.

Krok 3. Wskaż oczekiwaną kolejność pomiarów, przypisując im kolejne liczby naturalne.

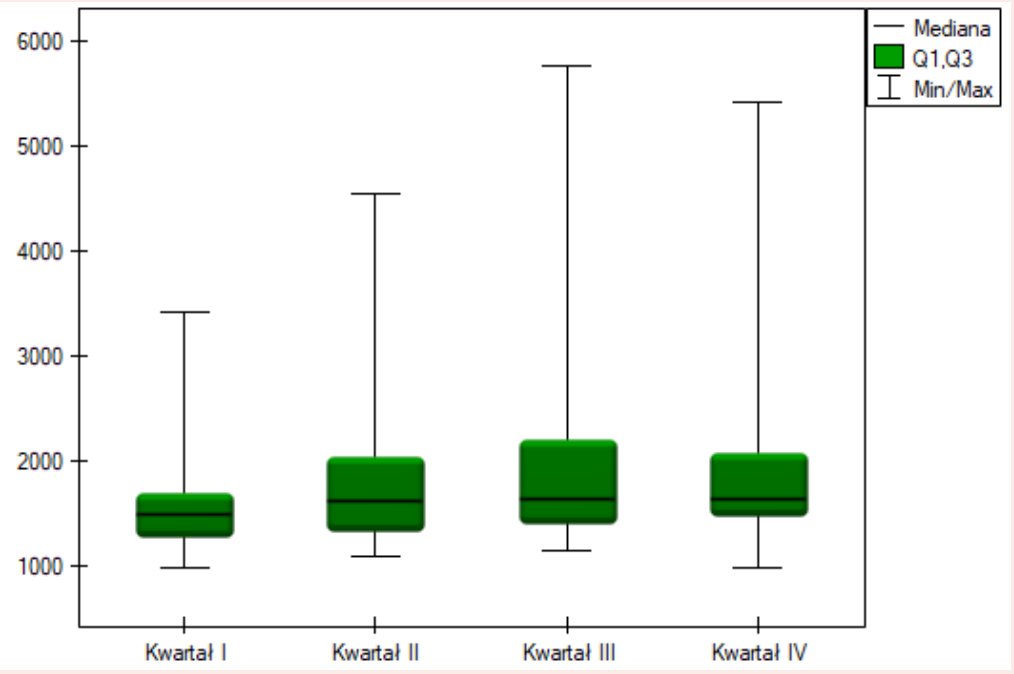
Krok 4. Uruchom analizę.

Uwaga! Określenie „z co najmniej jedną nierównością ścisłą" w hipotezie alternatywnej tego testu oznacza, że co najmniej jedna mediana powinna być większa niż mediana innej grupy pomiarów, w oczekiwanej kolejności.

WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 17.6 (plik baton.pqs): oczekiwanym skutkiem prowadzenia przez firmę intensywnej kampanii reklamowej jest stały wzrost sprzedaży oferowanego batonu. Sprawdzimy więc, czy sprzedaż batonu zmienia się zgodnie z oczekiwanym kierunkiem w kolejnych kwartałach.

Test Page dla trendu	
Numer porządkowy dla [Kwartał I]	1
Numer porządkowy dla [Kwartał II]	2
Numer porządkowy dla [Kwartał III]	3
Numer porządkowy dla [Kwartał IV]	4
Statystyka L	398
Statystyka Z	4,443936
Wartość p jednostronne	0,000004
Wartość p dwustronne	0,000009



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$p = 0.000004$ **(jednostronna, istotna statystycznie,** $\alpha = 0.05$**)**

Kampania reklamowa przyniosła oczekiwany trend wzrostu sprzedaży produktu.

DLA ZAINTERESOWANYCH

Hipotezy: dotyczą równości sumy rang dla kolejnych pomiarów lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0: &\quad \theta_1 = \theta_2 = ... = \theta_k, \\ \mathcal{H}_1: &\quad \theta_1 \geq \theta_2 \geq ... \geq \theta_k, \text{ z co najmniej jedną nierównością ścisłą} \end{aligned}$$

Wzory: statystyka testowa ma postać:

$$Z = \frac{L - \left[\frac{nk(k+1)^2}{4}\right]}{\sqrt{\frac{n(k^3-k)^2}{144(k-1)}}}$$

gdzie:
$L = \sum_{j=1}^{k} R_j c_j$,
R_j – suma rang j-tego pomiaru,
c_j – waga dla j-tego pomiaru informująca o naturalnym porządku tego pomiaru wśród innych pomiarów (wagi to kolejne liczby naturalne).

Komentarz: wzór na statystykę testową Z nie zawiera poprawki na rangi wiązane, przez co staje się nieco bardziej konserwatywny, gdy rangi wiązane występują. Stosowanie korekty na rangi wiązane w przypadku tego testu nie jest jednak rekomendowane. Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Interpretacja: przy znanym oczekiwanym kierunku trendu hipoteza alternatywna jest jednostronna i interpretacji podlega jednostronna wartość p. Interpretacja dwustronnej wartości p oznacza, że badacz nie zna (nie zakłada) kierunku ewentualnego trendu.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

ANOVA Durbina (brakujących danych)

KIEDY STOSOWAĆ TEN TEST:

Analiza wariancji powtarzanych pomiarów dla rang Durbina (Durbin, 1951[53]) stosuje się w tej samej sytuacji, co ANOVA Friedmana – gdy pomiarów badanej zmiennej dokonujemy kilkukrotnie na tych samych obiektach. Test Durbina ma jednak przewagę: można go wyliczać również dla niekompletnego zestawu danych, o ile braki tworzą tzw. zbalansowany i niekompletny blok (każdy obiekt ma tę samą liczbę pomiarów, a każda para pomiarów występuje wspólnie tak samo często).

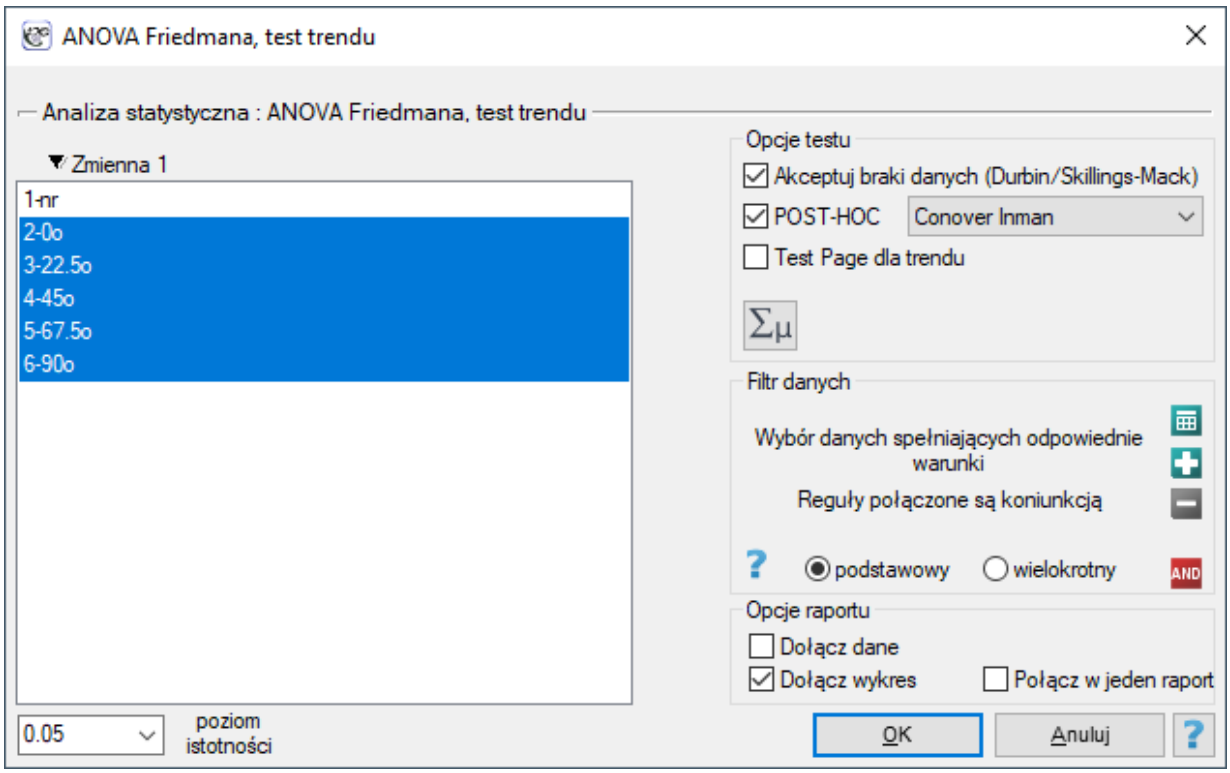
WARUNKI STOSOWANIA

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- pomiar na skali porządkowej lub interwałowej,
- model zależny,
- braki danych (jeśli występują) tworzą zbalansowany i niekompletny blok: każdy obiekt ma tę samą liczbę k pomiarów, każdy pomiar dotyczy tej samej liczby r obiektów, a liczba obiektów, dla których jednocześnie wykonano tę samą parę pomiarów, jest stała.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → ANOVA Friedmana (możliwość braków danych).



Krok 1. Wskaż zmienne odpowiadające kolejnym pomiarom.

Krok 2. Zaznacz opcję Akceptuj braki danych, jeśli w danych występują braki.

Krok 3. Jeśli test wykaże istotną różnicę, dobierz test POST-HOC Conover-Inman.

Krok 4. Uruchom analizę.

Uwaga! By rekordy, w których występują braki danych, były brane pod uwagę, wymagane jest zaznaczenie opcji Akceptuj braki danych. Jako braki danych traktowane są puste komórki oraz komórki o wartościach nieliczbowych. W analizie biorą udział tylko rekordy zawierające więcej niż jedną wartość liczbową.

WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przykład 17.7. (plik mirror.pqs)
Przeprowadzono eksperyment wśród 20 pacjentów szpitala psychiatrycznego (Ogilvie, 1965[130]), polegający na odrysowaniu linii prostych według zaprezentowanego wzoru, przedstawiającego 5 linii rysowanych pod różnym kątem (0°, 22.5^o, 45°, 67.5^o, 90°) względem wskazanego środka – zadaniem pacjentów było odwzorowanie linii mając zasłoniętą dłoń, a wynikiem był czas kreślenia danej linii. Ze względu na zmęczenie i ograniczoną chęć współpracy pacjentów, projekt zaplanowano w zbalansowanym i niekompletnym bloku: każdy z 20 pacjentów rysował linię pod dwoma z pięciu możliwych kątów, dzięki czemu każdy kąt wyrysowywano ośmiokrotnie. Chcemy sprawdzić, czy czas poświęcony na wyrysowanie poszczególnych linii jest zupełnie losowy, czy też są linie, których wyrysowywanie zajmuje więcej lub mniej czasu. Czas (w sekundach), w jakim każdy pacjent wyrysowywał linię pod zadanym kątem, zapisano w tabeli.

nr pacjenta	0°	22.5^o	45°	67.5^o	90°
1	7	15			
2	20		72		
3	8			26	
4	33				36
5	7	16			
6		68	67		
7		33		64	
8		34			12
9	10		96		
10		29	59		
11			17	9	
12			100		15
13	16			32	
14		19		32	
15			36	39	
16				44	54
17	16				38
18		17			12
19			37		11
20				56	6

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ANOVA Friedmana (możliwość braków danych)	
Czas analizy	0,06 sek.
Analizowane zmienne	0o;22.5o;45o;67.5o;90o
Poziom istotności	0,05
Akceptuj braki danych (Durbin/Skillings-Mack)	Tak
C	90
A	100
Statystyka T1 Durbin (F)	10,4
Stopnie swobody	4
Wartość p	0,034203
Statystyka T2 Durbin (F)	4,333333
Stopnie swobody	4 / 16
Wartość p	0,014541
Statystyka Skillings-Mack	10,4
Stopnie swobody	4
Wartość p	0,034203

$p = 0.014541$ **dla statystyki** T_2 **(lub** $p = 0.034203$ **dla** T_1**), istotne statystycznie,** $\alpha = 0.05$

Czas poświęcony na wyrysowanie poszczególnych linii nie jest taki sam.

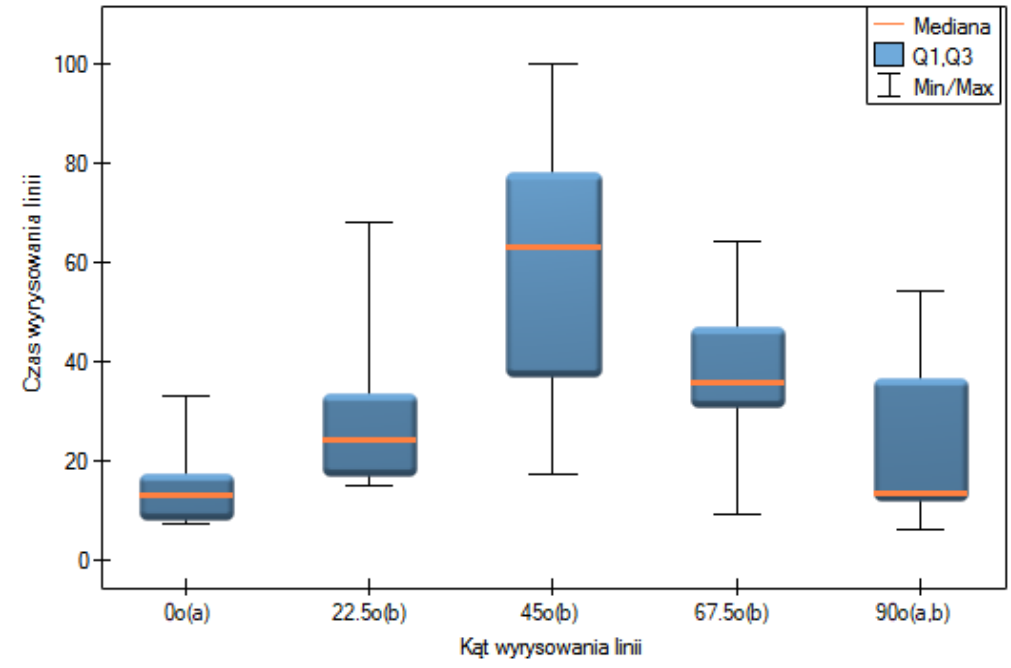
Wykonana analiza POST-HOC wskazuje na różnice czasu poświęconego na narysowanie linii pod kątem 0° – jest ona rysowana szybciej niż linie pod kątem 22.5^o, 45° oraz 67.5^o.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

POST-HOC (Conover-Iman)

	0o	22.5o	45o	67.5o	90o
Różnica dla su					
0o		5	6	6	3
22.5o	5		1	1	2
45o	6	1		0	3
67.5o	6	1	0		3
90o	3	2	3	3	
NIR	0o	22.5o	45o	67.5o	90o
0o		4.6445	4.6445	4.6445	4.6445
22.5o	4.6445		4.6445	4.6445	4.6445
45o	4.6445	4.6445		4.6445	4.6445
67.5o	4.6445	4.6445	4.6445		4.6445
90o	4.6445	4.6445	4.6445	4.6445	
Statystyka	0o	22.5o	45o	67.5o	90o
0o		2.2822	2.7386	2.7386	1.3693
22.5o	2.2822		0.4564	0.4564	0.9129
45o	2.7386	0.4564		0	1.3693
67.5o	2.7386	0.4564	0		1.3693
90o	1.3693	0.9129	1.3693	1.3693	
Wartość p	0o	22.5o	45o	67.5o	90o
0o		0.0365	0.0146	0.0146	0.1898
22.5o	0.0365		0.6542	0.6542	0.3749
45o	0.0146	0.6542		1	0.1898
67.5o	0.0146	0.6542	1		0.1898
90o	0.1898	0.3749	0.1898	0.1898	
Jednorodne	0o(a)	22.5o(b)	45o(b)	67.5o(b)	90o(a,b)
A	*				*
B		*	*	*	*

Na wykresie zaznaczono grupy jednorodne wskazane przez test POST-HOC.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Hipotezy: dotyczą równości sumy rang dla kolejnych pomiarów (R_j) lub są upraszczane do median (θ_j):

$$\begin{aligned}\mathcal{H}_0: &\quad \theta_1 = \theta_2 = ... = \theta_t,\\ \mathcal{H}_1: &\quad \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2, ..., t),\end{aligned}$$

gdzie t – łączna liczba rozpatrywanych pomiarów, b – łączna liczba badanych obiektów.

Wzory: wyznacza się dwie statystyki testowe:

$$T_1 = \frac{(t-1)\left[\sum_{j=1}^{t} R_j^2 - tC\right]}{A - C},$$

$$T_2 = \frac{T_1/(t-1)}{(b(k-1) - T_1)/(bk - b - t + 1)},$$

gdzie:
R_j – suma rang dla kolejnych pomiarów $(j = 1, 2, ...t)$,
R_{ij} – rangi przypisane kolejnym pomiarom, oddzielnie dla każdego z badanych obiektów $(i = 1, 2, ...b)$,
$A = \sum_{i=1}^{b}\sum_{j=1}^{t} R_{ij}^2$ – suma kwadratów dla rang,
$C = \frac{bk(k+1)^2}{4}$ – współczynnik korekcji. Wzór na statystykę T_1 i T_2 zawiera poprawkę na rangi wiązane.

W przypadku danych kompletnych statystyka T_1 jest tożsama z testem Friedmana i ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $df = t - 1$ stopniami swobody. Statystyka T_2 to odpowiednik korekty Iman-Davenport ANOVA Friedmana, podlega rozkładowi F Snedecora z $df_1 = t - 1$ i $df_2 = bk - b - t + 1$ stopniami swobody – uznaje się ją obecnie za bardziej precyzyjną niż statystykę T_1 i rekomenduje jej stosowanie[47].

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Testy POST-HOC

Wprowadzenie do kontrastów i testów POST-HOC znajduje się w rozdziale 17.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Conover-Inman
Stosowany dla porównań prostych (liczność w poszczególnych pomiarach zawsze jest taka sama).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = t_{1-\alpha/2, bk-b-t+1}\sqrt{\frac{(A-C)2r}{bk-b-t+1}\left(1 - \frac{T_1}{b(k-1)}\right)},$$

gdzie $t_{1-\alpha/2,bk-b-t+1}$ to wartość krytyczna (statystyka) rozkładu t-Studenta dla poziomu istotności α i $df = bk - b - t + 1$ stopni swobody.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

(ii) Statystyka testowa ma postać:

$$t = \frac{\sum_{j=1}^{k} c_j R_j}{\sqrt{\frac{(A-C)2r}{bk-b-t+1}\left(1 - \frac{T_1}{b(k-1)}\right)}}.$$

Statystyka ta podlega rozkładowi t-Studenta z $df = bk - b - t + 1$ stopniami swobody.

ANOVA Skillings-Mack (brakujących danych)

KIEDY STOSOWAĆ TEN TEST:

Analiza wariancji powtarzanych pomiarów dla rang Skillings-Mack (Skillings i Mack, 1981[157]) stosuje się w tej samej sytuacji, co ANOVA Friedmana, gdy jednak w danych występują braki, które nie tworzą żadnego szczególnego układu (wymagane są jedynie co najmniej dwie obserwacje dla każdego obiektu). Gdy braków nie ma i nie ma rang wiązanych, test jest tożsamy z ANOVA Friedmana; gdy braki występują w układzie zbalansowanym, daje te same wyniki co ANOVA Durbina.

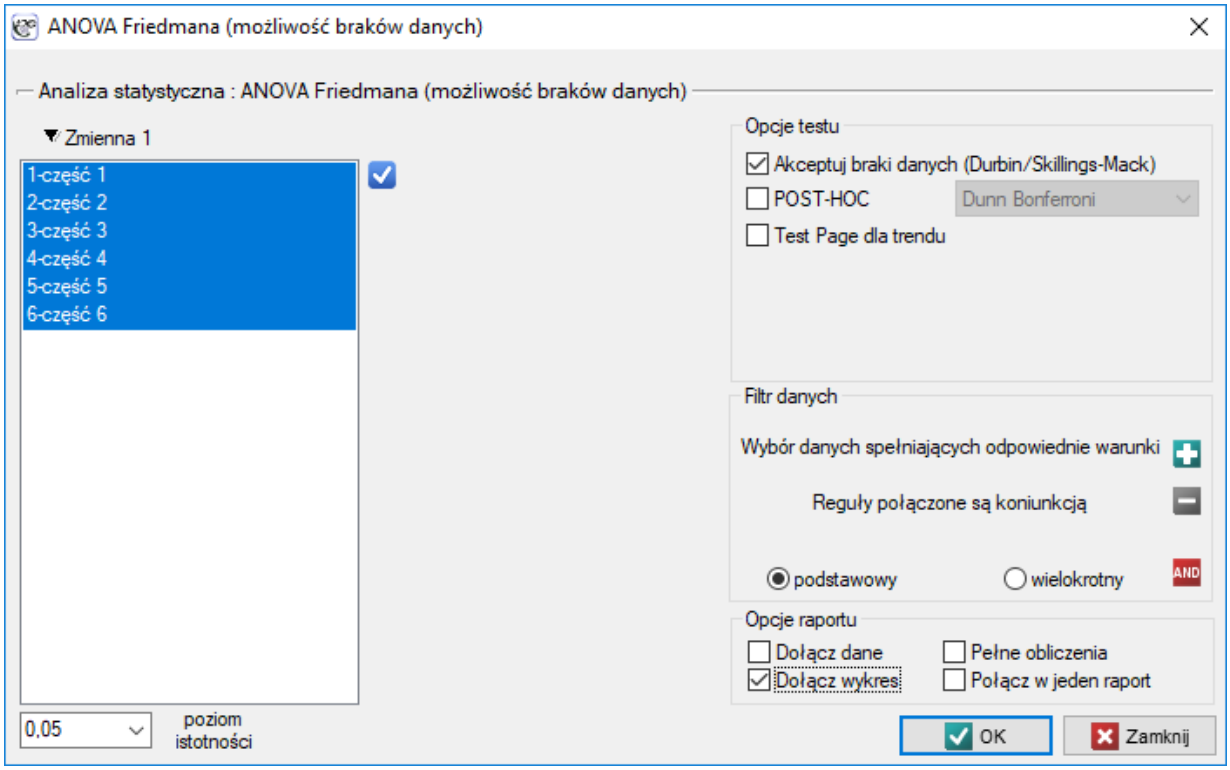
WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej,
- model zależny,
- każdy obiekt ma przynajmniej dwie obserwacje.

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Testy nieparametryczne (kat. uporządkowane) $\rightarrow$ ANOVA Friedmana (możliwość braków danych).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienne odpowiadające kolejnym pomiarom.

Krok 2. Zaznacz opcję Akceptuj braki danych.

Krok 3. Uruchom analizę.

Uwaga! By rekordy, w których występują braki danych, były brane pod uwagę, wymagane jest zaznaczenie opcji Akceptuj braki danych. Jako braki danych traktowane są puste komórki oraz komórki o wartościach nieliczbowych. W analizie biorą udział tylko rekordy zawierające więcej niż jedną wartość liczbową. Dla tego testu nie ma możliwości przeprowadzenia analizy POST-HOC.

WYKONANIE I INTERPRETACJA

Przykład 17.8. (plik ankietyzacja.pqs)
Pewien nauczyciel akademicki, chcąc poprawić sposób prowadzenia zajęć, postanowił zweryfikować swoje umiejętności dydaktyczne. W kilku losowo wybranych grupach studenckich, podczas ostatnich zajęć, prosił o wypełnienie krótkiej anonimowej ankiety złożonej z 6 pytań dotyczących sposobu ilustrowania sześciu wyszczególnionych partii materiału (skala pięciostopniowa, gdzie 1 – zupełnie niezrozumiały sposób przedstawiania materiału, 5 – bardzo jasny i ciekawy sposób ilustrowania materiału). Uzyskane dane okazały się niekompletne: w 30-osobowej grupie wypełniającej ankietę tylko 15 osób udzieliło kompletu odpowiedzi. Analiza nieuwzględniająca braków danych (ANOVA Friedmana) miałaby w tym przypadku ograniczoną moc ze względu na drastyczne obcięcie liczności grupy. Braki danych nie były też zaplanowane i nie tworzą zbalansowanego bloku, więc nie można również wykorzystać ANOVA Durbina. Sprawdzimy, czy poziom oceny poszczególnych części materiału różni się między sobą.

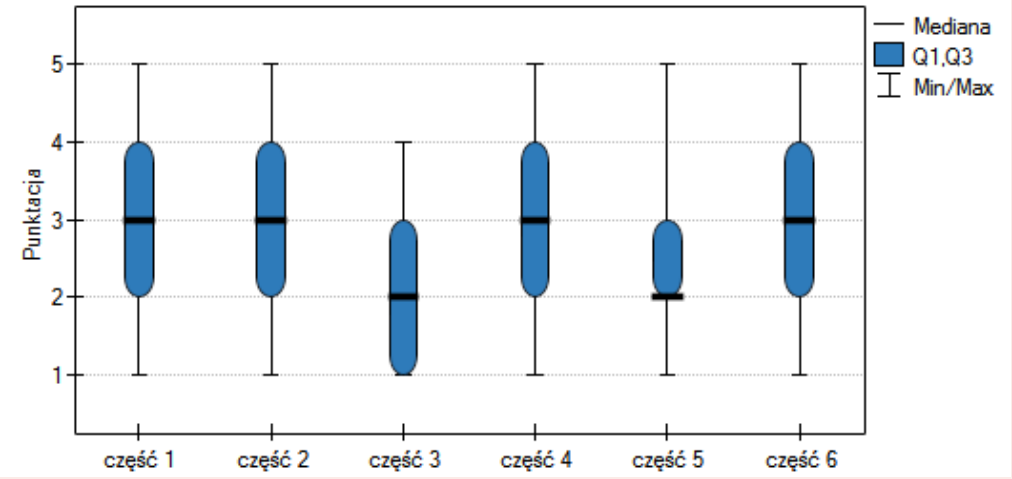
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ANOVA Friedmana (możliwość braków danych)	
Czas analizy	0,11 sek.
Analizowane zmienne	część 1
	część 2
	część 3
	część 4
	część 5
	część 6
Poziom istotności	0,05
Akceptuj braki danych (Durbin/Skillings-Mack)	Tak
C	2205
A	2086,5
Statystyka T1 Friedman	NA
Stopnie swobody	5
Wartość p	NA
Statystyka T2 Iman-Davenport	NA
Stopnie swobody	5 / 145
Wartość p	NA
Statystyka Skillings-Mack	16,037016
Stopnie swobody	5
Wartość p	0,006739

$p = 0.006739$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Oceny poszczególnych części materiału różnią się istotnie statystycznie.

Uzyskaną wartość p należy traktować ostrożnie ze względu na możliwe rangi wiązane – w tym przypadku znajduje się ona jednak znacznie poniżej przyjętego poziomu istotności. Różnice w odpowiedziach można obserwować na wykresie; dla tego testu nie ma jednak możliwości przeprowadzenia analizy POST-HOC.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Hipotezy: dotyczą równości sumy rang dla kolejnych pomiarów (R_j) lub są upraszczane do median (θ_j):

$$\begin{aligned} \mathcal{H}_0: &\quad \theta_1 = \theta_2 = ... = \theta_k, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2, ..., k), \end{aligned}$$

Wzory:

Statystyka testowa ma postać:

$$\chi^2 = A\Sigma_0^{-1}A^T,$$

gdzie:
$A = (A_1, A_2, ..., A_{k-1})$,
$A_j = \sum_{i=1}^{n} \sqrt{\frac{12}{s_i+1}} \left(R_{ij} - \frac{s_i+1}{2}\right)$,
s_i – liczba obserwacji dla i-tego obiektu,
R_{ij} – rangi przypisane kolejnym pomiarom ($j = 1, 2, ...k$), oddzielnie dla każdego z badanych obiektów ($i = 1, 2, ...n$), przy czym rangi dla braków danych są równe średniej randze dla danego obiektu,
Σ_0 – macierz wyznaczająca kowariancje dla A przy prawdziwości $\mathcal{H}_0$[157].

Gdy każda para pomiarów występuje równocześnie dla przynajmniej jednej obserwacji, statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $k-1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test chi-kwadrat dla wielowymiarowych tabel kontyngencji

KIEDY STOSOWAĆ TEN TEST:

Test χ^2 dla wielowymiarowych tabel kontyngencji *(ang. Chi-square test for multidimensional contingency tables)* stosujemy, gdy chcemy sprawdzić zależność pomiędzy więcej niż dwiema cechami naraz – jest to rozszerzenie testu χ^2 dla tabel ($R \times C$) na trzy i więcej cech.

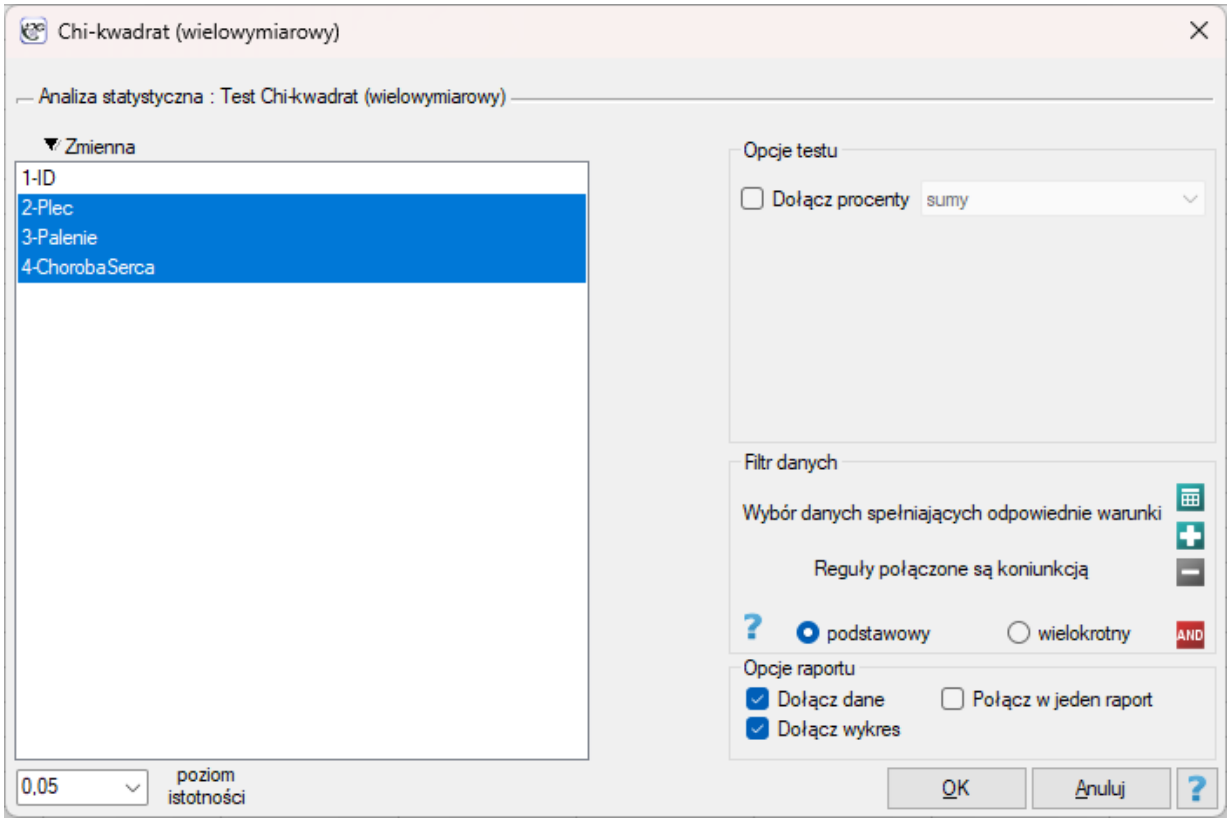
WARUNKI STOSOWANIA

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- model niezależny,
- duże liczności oczekiwane według interpretacji Cochrana (1952)[41].

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne (kat. nieuporządkowane) → chi-kwadrat (wielowymiarowy) (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Krok 1. Wskaż wszystkie badane cechy (trzy lub więcej).

Krok 2. Uruchom analizę.

Uwaga! Test ten jest możliwy do wyliczenia tylko na podstawie danych surowych.

WYKONANIE I INTERPRETACJA

Przykład 17.9. (plik choroba_serca.pqs)
Zbadano próbę 200 osób pod kątem trzech cech: płci (X_1: K, M), palenia papierosów (X_2: tak, nie) oraz występowania choroby serca (X_3: tak, nie). Celem badania było sprawdzenie, czy pomiędzy tymi trzema cechami w badanej populacji istnieje zależność.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dane :

	Plec	Palenie	ChorobaSer	obs.	oczekiw.
1	K	tak	tak	11	9,12135
2	K	tak	nie	23	28,10865
3	K	nie	tak	9	15,86865
4	K	nie	nie	59	48,90135
5	M	tak	tak	25	8,76365
6	M	tak	nie	14	27,00635
7	M	nie	tak	4	15,24635
8	M	nie	nie	55	46,98365

Test Chi-kwadrat (wielowymiarowy)	
Analizowane zmienne	Plec
	Palenie
	ChorobaSerca
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0,05
Liczność	200
Statystyka Chi-kwadrat	52,382342
Stopnie swobody	4
Wartość p	<0,000001

$\chi^2 = 52.382342$, $df = 4$, $p < 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

W badanej populacji istnieje istotna statystycznie zależność pomiędzy płcią, paleniem papierosów a występowaniem choroby serca.

Liczności oczekiwane widoczne w oknie danych są wystarczająco duże (warunek Cochrana spełniony), więc wynik testu jest wiarygodny. Porównanie liczności obserwowanych z oczekiwanymi pokazuje, gdzie zależność jest najsilniejsza: wśród palących mężczyzn choroba serca występuje znacznie częściej, niż wynikałoby z braku zależności (25 obserwowanych przypadków wobec 8,76 oczekiwanych) — niemal trzykrotnie częściej niż oczekiwano. Odwrotną sytuację widać u niepalących mężczyzn: choroba serca występuje u nich znacznie rzadziej, niż wynikałoby z braku zależności (4 obserwowane przypadki wobec 15,25 oczekiwanych). U kobiet odchylenia liczności obserwowanych od oczekiwanych są wyraźnie mniejsze. Zależność dotyczy więc przede wszystkim mężczyzn – to u nich palenie papierosów wiąże się z wyraźnie podwyższonym ryzykiem choroby serca.

DLA ZAINTERESOWANYCH

Hipotezy:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$$\mathcal{H}_0: \quad O_{ij\ldots} = E_{ij\ldots} \text{ dla wszystkich kategorii,}$$
$$\mathcal{H}_1: \quad O_{ij\ldots} \neq E_{ij\ldots} \text{ dla przynajmniej jednej kategorii,}$$

gdzie $O_{ij\ldots}$ i $E_{ij\ldots}$ to liczności obserwowane w tabeli kontyngencji i odpowiadające im liczności oczekiwane.

Wzory:

Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum\ldots\sum\frac{(O_{ij\ldots} - E_{ij\ldots})^2}{E_{ij\ldots}}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z liczbą stopni swobody wyznaczaną, dla tabeli o 3 wymiarach, według wzoru:

$$df = (r-l)(c-1)(l-1) + (r-l)(c-1) + (r-1)(l-1) + (c-1)(l-1).$$

ANOVA Q-Cochrana

KIEDY STOSOWAĆ TEN TEST:

Analiza wariancji Q-Cochrana (Cochran, 1950[40]) jest rozszerzeniem testu McNemara na $k \geq 2$ grup zależnych. Pozwala sprawdzić, czy wyniki kilkukrotnych pomiarów cechy dwuwartościowej (np. odpowiedź poprawna/błędna) są symetryczne, czyli czy proporcja jednej z wartości jest taka sama we wszystkich pomiarach.

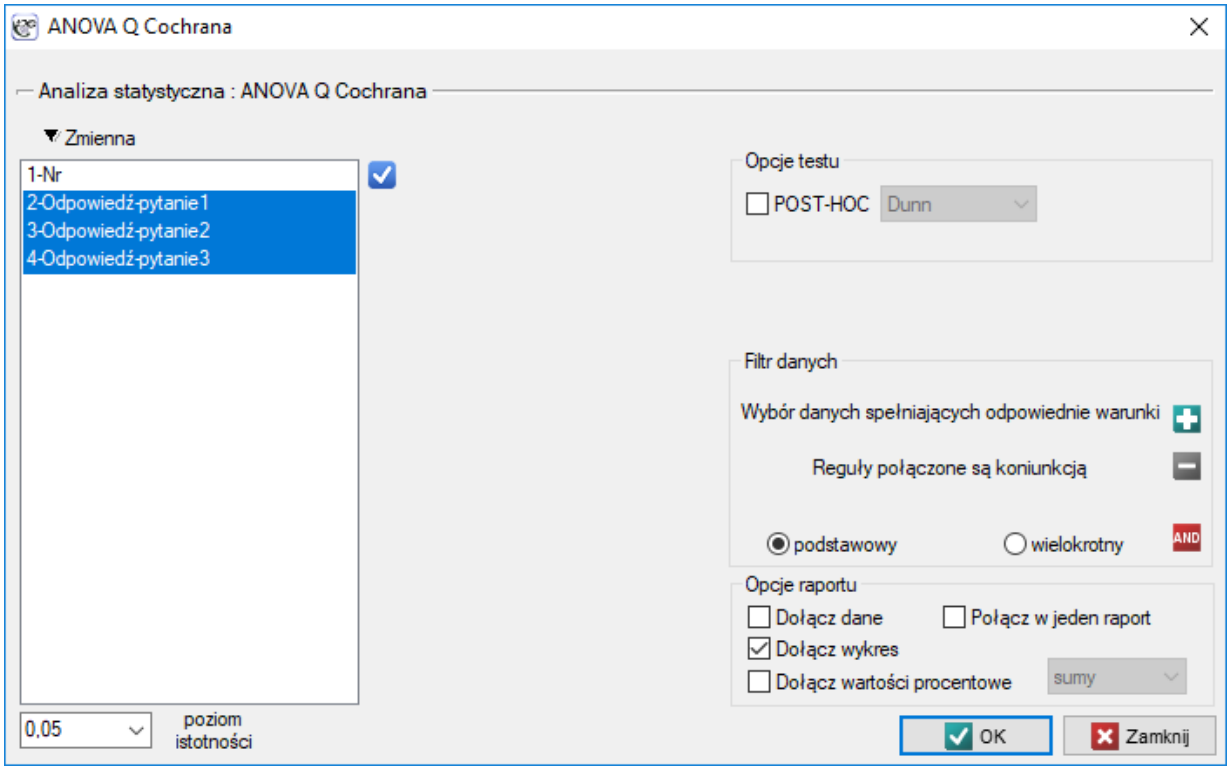
WARUNKI STOSOWANIA

- pomiar na skali nominalnej (zmienne dychotomiczne, czyli o 2 kategoriach, zakodowane jako 0 i 1),
- model zależny.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne (kat. nieuporządkowane) → ANOVA Q-Cochrana (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krok 1. Wskaż zmienne odpowiadające kolejnym pomiarom (zakodowane jako 0/1).

Krok 2. Jeśli test wykaże istotną różnicę, dobierz test POST-HOC Dunn.

Krok 3. Uruchom analizę.

Uwaga! Test ten można wyliczyć wyłącznie na podstawie danych surowych.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 17.10. (plik test.pqs)
Chcemy porównać trudność 3 pytań testowych. W tym celu z badanej populacji osób, do których adresowany jest test, wybieramy 20-osobową próbę. Każda osoba z próby udziela odpowiedzi na 3 pytania zawarte w teście, po czym sprawdzamy poprawność tych odpowiedzi (odpowiedź poprawna lub błędna). Wyniki zawiera poniższa tabela.

Nr	Odpowiedź – pytanie 1	Odpowiedź – pytanie 2	Odpowiedź – pytanie 3
1	poprawna	poprawna	błędna
2	błędna	poprawna	błędna
3	poprawna	poprawna	poprawna
4	błędna	poprawna	błędna
5	błędna	poprawna	błędna
6	błędna	poprawna	poprawna
7	błędna	błędna	błędna
8	błędna	poprawna	błędna
9	poprawna	poprawna	błędna
10	błędna	poprawna	błędna
11	błędna	błędna	błędna
12	błędna	błędna	poprawna
13	błędna	poprawna	błędna
14	błędna	błędna	poprawna
15	poprawna	błędna	błędna
16	błędna	błędna	błędna

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

ANOVA Q Cochrana	
Czas analizy	0.03 sek.
Analizowane zmienne	Odpowiedź-pytanie1,Odpowiedź-pytanie2,Odpowie
Poziom istotności	0.05
Liczność	20
Ilość stopni swobody	2
Statystyka Q	9.733333
Wartość p	0.007699

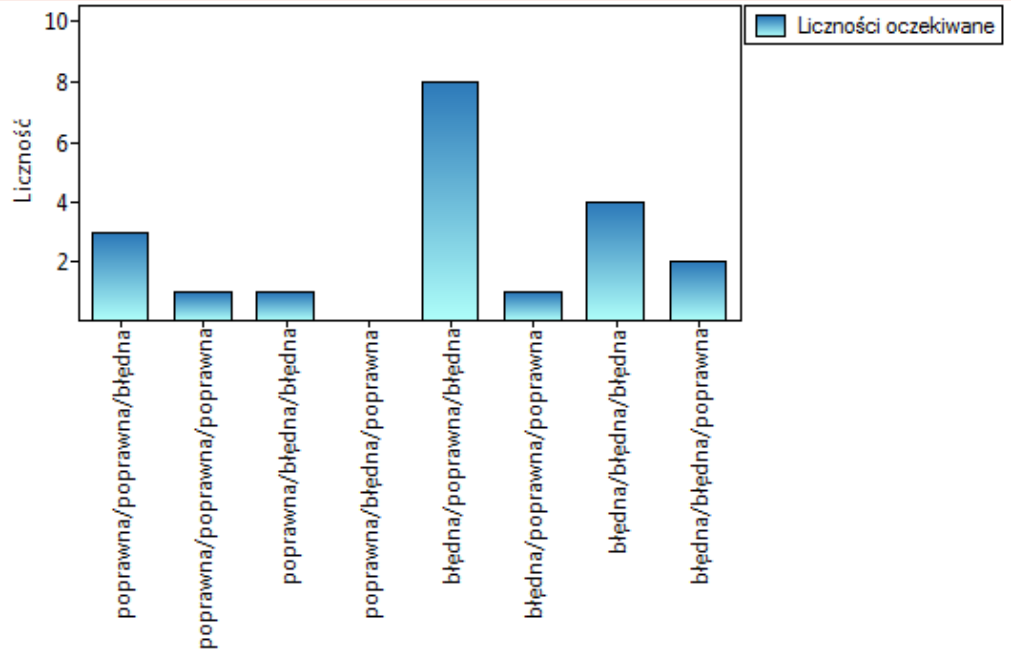
$p = 0.007699$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Pytania testowe wykazują różny stopień trudności.

Wznawiamy analizę przyciskiem , by wykonać test POST-HOC, i w oknie opcji testu wybieramy POST-HOC Dunn.

POST-HOC (Dunn)			
	Odpowied:	Odpowied:	Odpowied:
Różnica średnich			
Odpowiedź-pytanie1		0.4	0.05
Odpowiedź-pytanie2	0.4		0.45
Odpowiedź-pytanie3	0.05	0.45	
NIR			
Odpowiedź-pytanie1		0.4641	0.4641
Odpowiedź-pytanie2	0.4641		0.4641
Odpowiedź-pytanie3	0.4641	0.4641	
Statystyka Z			
Odpowiedź-pytanie1		2.52982	0.31623
Odpowiedź-pytanie2	2.52982		2.84605
Odpowiedź-pytanie3	0.31623	2.84605	
Wartość p			
Odpowiedź-pytanie1		0.03424	1
Odpowiedź-pytanie2	0.03424		0.01328
Odpowiedź-pytanie3	1	0.01328	

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Wykonana analiza POST-HOC wskazuje, że różnice dotyczą pytania 2. i 1. oraz 2. i 3. Różnica ta polega na tym, że pytanie 2. jest łatwiejsze niż 1. i 3. (liczba poprawnych odpowiedzi jest tu wyższa).

DLA ZAINTERESOWANYCH

Hipotezy:

$\mathcal{H}_0$: wszystkie „niezgodne" liczności obserwowane są sobie równe,
$\mathcal{H}_1$: nie wszystkie „niezgodne" liczności obserwowane są równe,

gdzie „niezgodne" liczności obserwowane to liczności obserwowane wyliczone, gdy wartość badanej cechy jest różna w kolejnych pomiarach.

Wzory:
Statystyka testowa ma postać:

$$Q = \frac{(k-1)\left(kC - T^2\right)}{kT - R}$$

gdzie:
$T = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}$,
$R = \sum_{i=1}^{n} \left(\sum_{j=1}^{k} x_{ij}\right)^2$,
$C = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} x_{ij}\right)^2$,
x_{ij} – wartość j-tego pomiaru dla i-tego obiektu (czyli 0 lub 1).

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody $df = k - 1$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Testy POST-HOC

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wprowadzenie do kontrastów i testów POST-HOC znajduje się w rozdziale 17.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Dunna
Stosowany dla porównań prostych (liczność w poszczególnych pomiarach zawsze jest taka sama).

Przykład – **porównania proste** (dla różnicy proporcji jednej wybranej pary pomiarów).
Hipotezy:

$\mathcal{H}_0$: wybrane „niezgodne" liczności obserwowane są sobie równe,
$\mathcal{H}_1$: wybrane „niezgodne" liczności obserwowane są różne,

(i) Wartość najmniejszej istotnej różnicy wyliczana jest ze wzoru:

$$NIR = Z_{\frac{\alpha}{c}} \sqrt{2\frac{kT - R}{n^2 k(k-1)}},$$

gdzie $Z_{\frac{\alpha}{c}}$ to wartość krytyczna (statystyka) rozkładu normalnego dla poziomu istotności α skorygowanego o liczbę możliwych porównań prostych c.

(ii) Statystyka testowa ma postać:

$$Z = \frac{\sum_{j=1}^{k} c_j p_j}{\sqrt{2\frac{kT-R}{n^2 k(k-1)}}},$$

gdzie p_j – proporcja j-tego pomiaru $(j = 1, 2, ...k)$. Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład normalny, a wartość p jest korygowana o liczbę możliwych porównań prostych c.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

18 WIELOKROTNE PORÓWNANIA

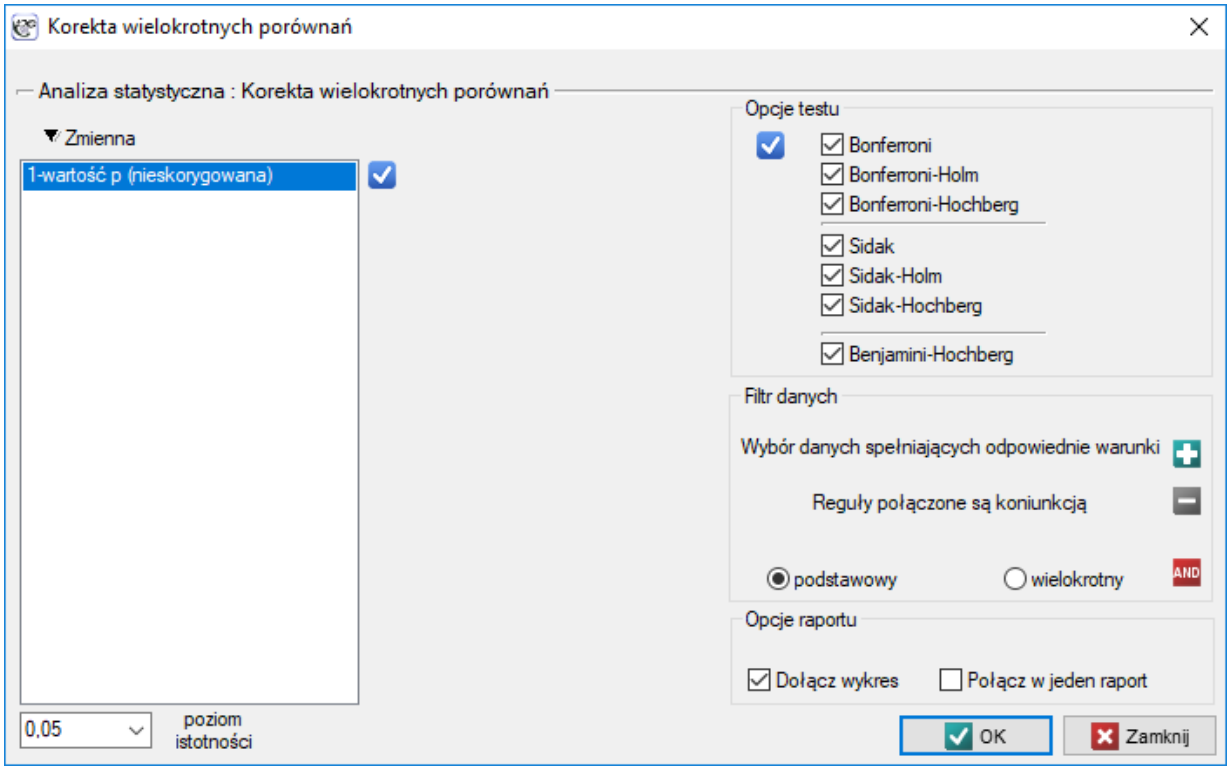
KIEDY STOSOWAĆ:

Jednoczesne testowanie wielu hipotez (tzw. **rodziny hipotez**) pociąga za sobą niebezpieczeństwo wzrostu błędu α, co stanowi główny problem w dziedzinie wielokrotnych porównań. Gdy błąd α wzrasta, oznacza to, że zbyt często odrzucamy hipotezę zerową, gdy jest ona prawdziwa – czyli zbyt często wskazujemy na istnienie różnic, podczas gdy w rzeczywistości różnice nie występują. Aby chronić przed wzrostem α, jedną ze strategii jest poprawienie (zmniejszenie) poziomu α lub odpowiednie poprawienie (zwiększenie) wartości p testów.

Uwaga! Rodzina hipotez może być zdefiniowana na wiele sposobów. Najczęściej są to hipotezy w ramach procedury POST-HOC, czyli wykonywania wielu testów w ramach jednoczesnego porównania kilku badanych grup. Taką rodzinę stanowić mogą również testy wykonywane w ramach analizy typu Hotelling. Rodziny hipotez występują również w wielu analizach geograficznych. Wszędzie tam, gdzie w ramach ogólnej hipotezy analizujemy wiele pomniejszych hipotez, korekta wielokrotnych porównań może mieć zastosowanie.

JAK USTAWIĆ ANALIZĘ

By wykonać korektę wielokrotnych porównań, do jednej kolumny arkusza danych wprowadzamy kolejne wartości p. Okno z ustawieniami opcji wielokrotnych porównań wywołujemy poprzez menu Statystyka→Korekta wielokrotnych porównań.



WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Kontynuacja przykładu 17.4 (plik satysfakcjaZpracy.pqs): w badaniu tym sprawdzano, czy poziom satysfakcji z wykonywanej pracy jest taki sam dla czterech kategorii wykształcenia. Rodzinę hipotez tworzyły tu hipotezy wynikające z porównania wszystkich grup parami. By porównać wszystkie 4 grupy, utworzono 6 par porównań – w każdym przypadku hipoteza zerowa dotyczyła braku różnic w poziomie satysfakcji w analizowanej parze. By wykorzystać kilka zaproponowanych korekt wielokrotnych porównań, analizę przeprowadzono przy pomocy nieskorygowanego testu POST-HOC Dunna.

POST-HOC (Dunn nieskorygowany)*

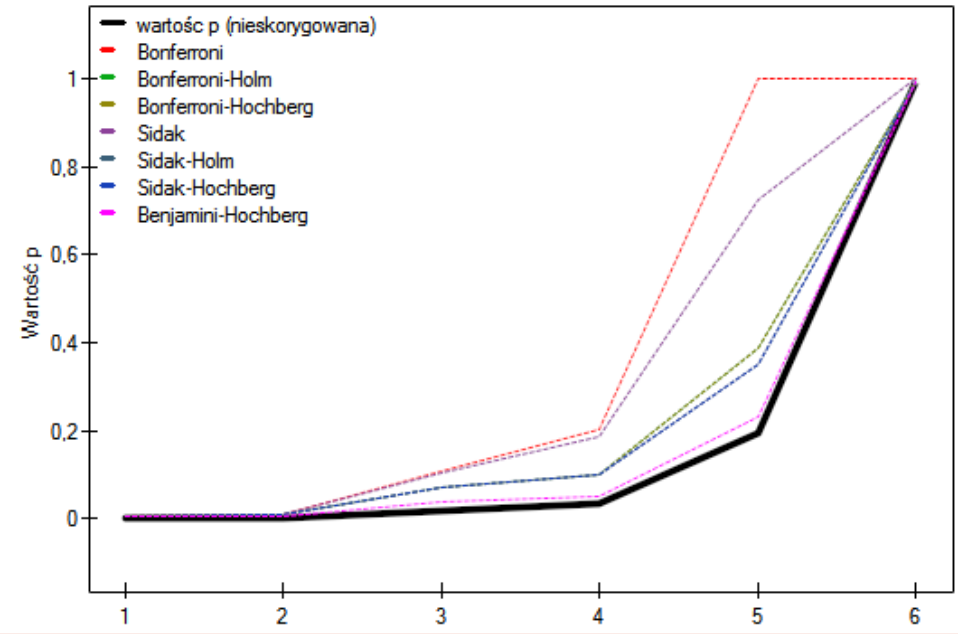
	podstawow	zawodowe	średnie	wyższe
Wartość p				
podstawow		0,193695	0,000639	0,00177
zawodowe	0,193695		0,018059	0,033594
średnie	0,000639	0,018059		0,991918
wyższe	0,00177	0,033594	0,991918	

Uzyskane w ten sposób wartości p podano jako dane do korekty wielokrotnych porównań, otrzymując następujące wyniki:

Korekta wielokrotnych porównań

Czas analizy	0,04 sek.
Liczność	6
Poziom istotności	0,05

Zmienna: wartość p (nieskorygowana)	Bonferroni	Bonferroni - Holm	Bonferroni - Hochberg	Sidak	Sidak - Holm	Sidak - Hochberg	Benjamini - Hochberg
0,193695	1	0,38739	0,38739	0,725213	0,349872	0,349872	0,232434
0,000639	0,003834	0,003834	0,003834	0,003828	0,003828	0,003828	0,003834
0,00177	0,01062	0,00885	0,00885	0,010573	0,008819	0,008819	0,00531
0,018059	0,108354	0,072236	0,072236	0,103578	0,070303	0,070303	0,036118
0,033594	0,201564	0,100782	0,100782	0,185375	0,097434	0,097434	0,050391
0,991918	1	0,991918	0,991918	1	0,991918	0,991918	0,991918



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

istotne statystycznie różnice dla 2 (lub 3, przy korekcie Benjamini-Hochberg) z 6 par porównań, $\alpha = 0.05$

Różnice dotyczą satysfakcji z pracy dla wykształcenia podstawowego vs średniego oraz podstawowego vs wyższego.

Tylko przy zastosowaniu korekty Benjamini-Hochberg różnice można zlokalizować w trzech parach porównań.

Inne przykłady zastosowania korekty wielokrotnych porównań: Przykład 19.1 oraz Przykład 19.2 (analiza typu Hotelling).

DLA ZAINTERESOWANYCH

Jeżeli liczbę testowanych hipotez oznaczymy przez c, wówczas korekty dotyczące wielokrotnych porównań można opisać następująco:

Korekta Bonferroniego

Najbardziej znana i jednocześnie najbardziej konserwatywna z korekt[2]. Polega na pomnożeniu każdego z prawdopodobieństw testowych przez całkowitą liczbę przeprowadzonych badań (lub podzieleniu poziomu istotności przez tę liczbę).

Wzór – korekta wartości p:

$$p_{(Bonferroni,i)} = p_i \cdot c$$

Wzór – korekta poziomu istotności α:

$$\alpha_{(Bonferroni,i)} = \frac{\alpha_i}{c}$$

Korekta Sidaka

Korekta Sidaka (1967)[155] ma większą moc niż korekta Bonferroniego, dlatego jest coraz chętniej wykorzystywana.

Wzór – korekta wartości p:

$$p_{(Sidak,i)} = 1 - (1 - p_i)^c$$

Wzór – korekta poziomu istotności α:

$$\alpha_{(Sidak,i)} = 1 - (1 - \alpha_i)^{1/c}$$

Korekta Bonferroniego-Holma

Polega na wykorzystaniu kilkukrokowej procedury Holma (1979)[84] do poprawki Bonferroniego. Procedura ta rozpoczyna się od posortowania w sposób rosnący kolejnych wartości p_i $(p_1, p_2, ...p_c)$. Następnie zastosowana jest korekta Bonferroniego do każdej kolejnej wartości p_i (przy odpowiednim zmniejszeniu liczby hipotez, jaka została do przetestowania). W rezultacie wszystkie hipotezy testowane po pierwszej nieistotnej statystycznie wartości p_i są również nieistotne.

Wzór – korekta wartości p:

$$p_{(Bonferroni,i)} = p_i \cdot (c - i + 1)$$

Korekta Sidaka-Holma

Polega na wykorzystaniu kilkukrokowej procedury Holma do poprawki Sidaka, analogicznie do korekty

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Bonferroniego-Holma powyżej.

Wzór – korekta wartości p:

$$p_{(Sidak,i)} = 1 - (1 - p_i)^{c-i+1}$$

Korekta Bonferroniego-Hochberga

Polega na wykorzystaniu kilkukrokowej procedury Hochberga (1988)[83] do poprawki Bonferroniego. Procedura ta rozpoczyna się od posortowania w sposób malejący kolejnych wartości p_i $(p_c, p_{c-1}, ..., p_1)$. Następnie zastosowana jest korekta Bonferroniego do każdej kolejnej wartości p_i (przy odpowiednim zmniejszeniu liczby hipotez, jaka została do przetestowania). W rezultacie wszystkie hipotezy testowane po pierwszej nieistotnej statystycznie wartości p_i są również nieistotne.

Wzór – korekta wartości p:

$$p_{(Bonferroni,i)} = p_i \cdot (c - i + 1)$$

Korekta Sidaka-Hochberga

Polega na wykorzystaniu kilkukrokowej procedury Hochberga do poprawki Sidaka, analogicznie do korekty Bonferroniego-Hochberga powyżej.

Wzór – korekta wartości p:

$$p_{(Sidak,i)} = 1 - (1 - p_i)^{c-i+1}$$

Korekta Benjamini-Hochberga

Polega na wykorzystaniu kilkukrokowej procedury Hochberga do poprawki Benjamini (1995)[19], będącej zmodyfikowaną wersją poprawki Bonferroniego. Procedura ta rozpoczyna się od posortowania w sposób malejący kolejnych wartości p_i $(p_c, p_{c-1}, ..., p_1)$. Następnie zastosowana jest korekta Benjamini do każdej kolejnej wartości p_i (przy odpowiednim zmniejszeniu liczby hipotez, jaka została do przetestowania). W rezultacie wszystkie hipotezy testowane po pierwszej nieistotnej statystycznie wartości p_i są również nieistotne. Spośród zaproponowanych korekt charakteryzuje się największą mocą – nie kontroluje bezpośrednio błędu α, ale minimalizuje oczekiwany odsetek fałszywych różnic wśród różnic wykrytych.

Wzór – korekta wartości p:

$$p_{(BH,i)} = p_i \frac{c}{i}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

19 HOTELLING T-KWADRAT

Analiza Hotellinga, zaproponowana przez Hotellinga (1931)[85] (1947)[86] i Mahalanobisa (1930[111], 1936[112]), jest rozszerzeniem testów t-Studenta na dziedzinę wielu zmiennych. W rezultacie jednoczesnej analizie możemy poddawać wiele różnych cech badanych obiektów.

Test T-kwadrat Hotellinga dla pojedynczej próby

KIEDY STOSOWAĆ TEN TEST:

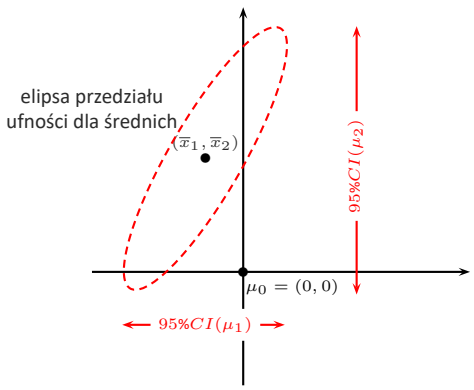
Test pozwala sprawdzić, czy k zmiennych w badanej populacji $X_1, X_2, ..., X_k$ charakteryzuje się średnimi zadanymi przez badacza – jest wielowymiarowym rozszerzeniem testu t-Studenta dla pojedynczej próby.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- wielowymiarowy rozkład normalny lub normalność rozkładu każdej badanej zmiennej.

Uwaga! Zasada działania testu Hotellinga jest tożsama z budową wielowymiarowej elipsy przedziałów ufności wokół centrum wyznaczonego przez średnie. Przez to, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi), często nie jesteśmy w stanie uzyskać tożsamych wyników.

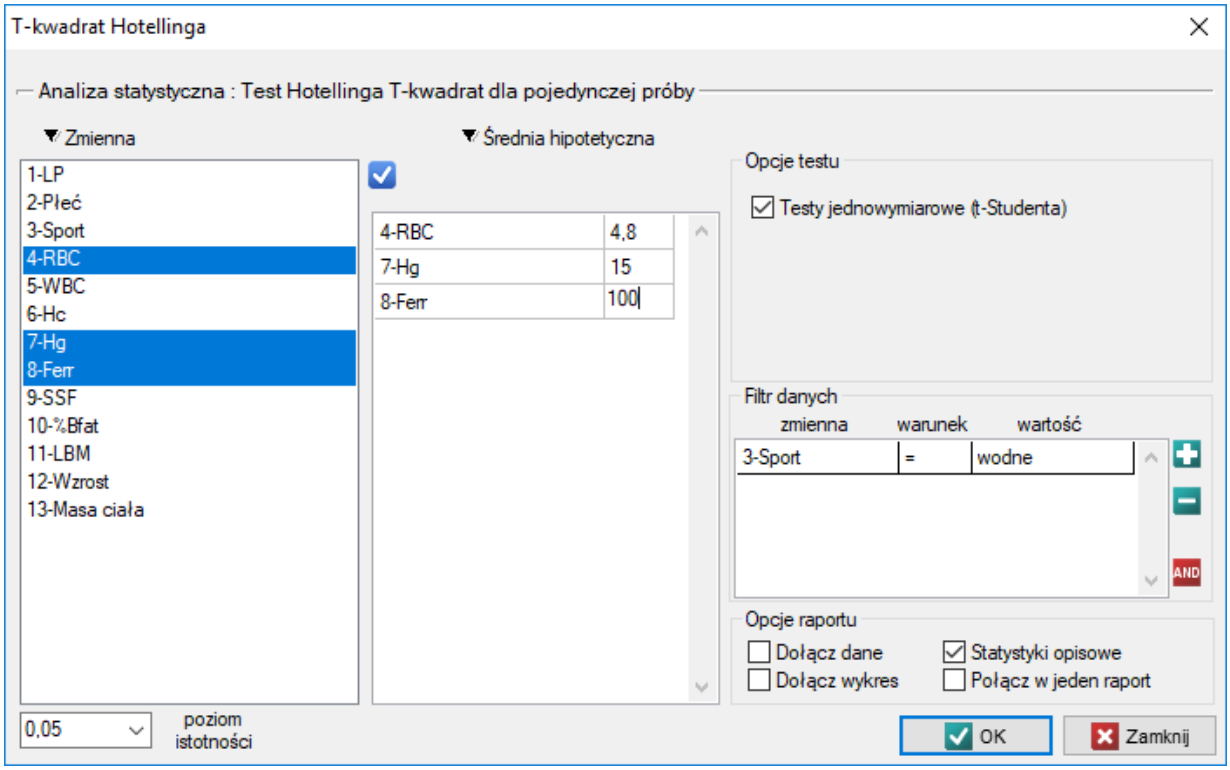
Przykład – interpretacja elipsy testu Hotellinga dla dwóch zmiennych
Zadany punkt opisany przez wartości średnie ($\mu_0 = (0, 0)$) znajduje się poza elipsą, co oznacza, że test Hotellinga odrzuca hipotezę $\mathcal{H}_0$, ale stosując podejście jednowymiarowe (przedziały ufności dla każdej zmiennej oddzielnie), nie jesteśmy w stanie odrzucić hipotezy $\mathcal{H}_0$, by wskazać zmienną, której dotyczą różnice.



Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → T-kwadrat Hotellinga.



WYKONANIE I INTERPRETACJA

Przykład 19.1. (plik sport.pqs)
Przebadano grupę sportowców, by uzyskać informację między innymi o takich parametrach zdrowotnych jak:
RBC – liczba czerwonych krwinek,
Hg – hemoglobina [g/dl],
Ferr – ferrytyna [µg/l].
Chcemy wiedzieć, na ile bliskie oczekiwanym przez badaczy wartościom są średnie poziomy RBC, hemoglobiny i ferrytyny dla sportowców uprawiających tzw. sporty „wodne”. Oczekiwane średnie to:
RBC = 4.8,
Hg = 15 [g/dl],
Ferr = 100 [µg/l].

Ponieważ arkusz danych zawiera informacje o badanych parametrach dla większej grupy sportowców, w oknie analizy osoby, które uprawiają sporty wodne, wskazujemy poprzez filtr danych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test Hotellinga T-kwadrat dla pojedynczej próby	
Czas analizy	0,08 sek.
Analizowane zmienne	RBC;Hg;Ferr
Poziom istotności	0,05
Filtr danych	Sport=wodne
Liczność	59
T-kwadrat	20,361767
Statystyka F	6,553212
Stopnie swobody 1	3
Stopnie swobody 2	56
Wartość p	0,000708

$p = 0.000708$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Średnie poziomy badanych parametrów różnią się od wartości zadanej.

Różnic możemy poszukiwać w wyznaczonych jednoczesnych przedziałach ufności lub w przedziałach z korektą Bonferroniego. Dla ferrytyny jednoczesny 95% przedział ufności znajduje się poniżej zadanej wartości, co świadczy o niższej populacyjnej wartości ferrytyny niż zadana przez badaczy. Przedział dla RBC i hemoglobiny zawiera zadane wartości, co wskazuje na brak istotnych statystycznie różnic.

Przedziały ufności [jednoczesne]			
	Średnia	-95% CI d	+95% CI (
RBC	4,691525	4,538874	4,844177
Hg	14,630508	14,232219	15,028798
Ferr	76,40678	60,422862	92,390697

Nieco węższe przedziały uzyskamy, wykorzystując poprawkę Bonferroniego – wówczas nie tylko przedział dla ferrytyny znajduje się poniżej zadanej wartości, ale również przedział dla hemoglobiny.

Przedziały ufności [Benferroni]			
	Średnia	-95% CI d	+95% CI (
RBC	4,691525	4,56323	4,819821
Hg	14,630508	14,295766	14,965251
Ferr	76,40678	62,973117	89,840442

Podejście jednowymiarowe, ze względu na swoją prostotę, wykorzystywane jest najczęściej. Możemy tu wybrać mniej konserwatywne korekty wielokrotnych porównań niż poprawka Bonferroniego lub Sidaka, uzyskując w ten sposób różnice dotyczące wszystkich badanych parametrów. By wykonać korektę wartości p testu t-Studenta, należy przekopiować te wartości do jednej kolumny nowego arkusza danych i z menu Statystyka wybrać Korektę wielokrotnych porównań.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Testy t-Studenta						
	Liczność	-95% CI d	+95% CI (	Statystyka	Stopnie sw	Wartość p
RBC	59	4,587359	4,795692	-2,08451	58	0,041527
Hg	59	14,358723	14,902294	-2,72133	58	0,00857
Ferr	59	65,499659	87,313901	-4,32992	58	0,00006

Korekta wielokrotnych porównań							
Czas analizy	0,05 sek.						
Liczność	3						
Poziom istotności	0,05						
Zmienna: Wartość p (t-Studenta)	Bonferroni	Bonferroni - Holm	Bonferroni - Hochberg	Sidak	Sidak - Holm	Sidak - Hochberg	Benjamini - Hochberg
0,041527	0,124581	0,041527	0,041527	0,119479	0,041527	0,041527	0,041527
0,00857	0,02571	0,01714	0,01714	0,02549	0,017067	0,017067	0,012855
0,00006	0,00018	0,00018	0,00018	0,00018	0,00018	0,00018	0,00018

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu = \mu_0, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_j \text{ są równe zadanym wartościom,} \end{aligned}$$

gdzie:
$\mu = (\mu_1, \mu_2, ..., \mu_k)$ – średnie zmiennych w populacji reprezentowanej przez próbę,
$\mu_0 = (\mu_{01}, \mu_{02}, ..., \mu_{0k})$ – zadane przez badacza wartości średnich.

Wzory: statystyka testowa ma postać:

$$F = \frac{n-k}{k(n-1)} T^2$$

gdzie:
$n = n_1 = n_2 = ... = n_k$ – liczności poszczególnych zmiennych w próbie,
$T^2 = n(\overline{x} - \mu_0)^T S^{-1} (\overline{x} - \mu_0)$ – pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznościach),
$\overline{x} = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_k)$ – średnie zmiennych w próbie,
S – macierz kowariancji.

Statystyka ta podlega rozkładowi F Snedecora z k i $n-k$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Jednoczesne przedziały ufności

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności średnich:

$$\overline{x}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha, df_1, df_2}} \cdot SE_j$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

lub przedziały z poprawką Bonferroniego, w celu sprawdzenia, czy znajduje się w nich zadana wartość. Jeśli zadana wartość znajduje się w wyznaczonym przedziale, oznacza to, że w rzeczywistości średnia danej zmiennej może być równa tej zadanej wartości. Stosując tę metodę, należy jednak pamiętać, że wyznaczone przedziały nie uwzględniają powiązań pomiędzy poszczególnymi zmiennymi (które uwzględnia test Hotellinga), a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice, możemy również zastosować podejście jednowymiarowe, wykonując porównania testem t-Studenta dla pojedynczej próby oddzielnie dla poszczególnych zmiennych. Niestety, nie uwzględnimy tym samym wzajemnych powiązań, ale uzyskane wartości p testu t-Studenta możemy skorygować w dziale Wielokrotne porównania.

Test T-kwadrat Hotellinga dla grup zależnych

KIEDY STOSOWAĆ TEN TEST:

Stosuje się w sytuacji, gdy pomiarów badanych k zmiennych dokonujemy dwukrotnie w różnych warunkach (przy czym zakładamy, że wariancje zmiennych w obu pomiarach są sobie bliskie). Jeśli pierwszy pomiar oznaczymy przez $X_1, X_2, ..., X_k$, a drugi przez $Y_1, Y_2, ..., Y_k$, weryfikujemy hipotezę, że populacyjne średnie zmiennych z pierwszego pomiaru są takie same jak z pomiaru drugiego.

WARUNKI STOSOWANIA

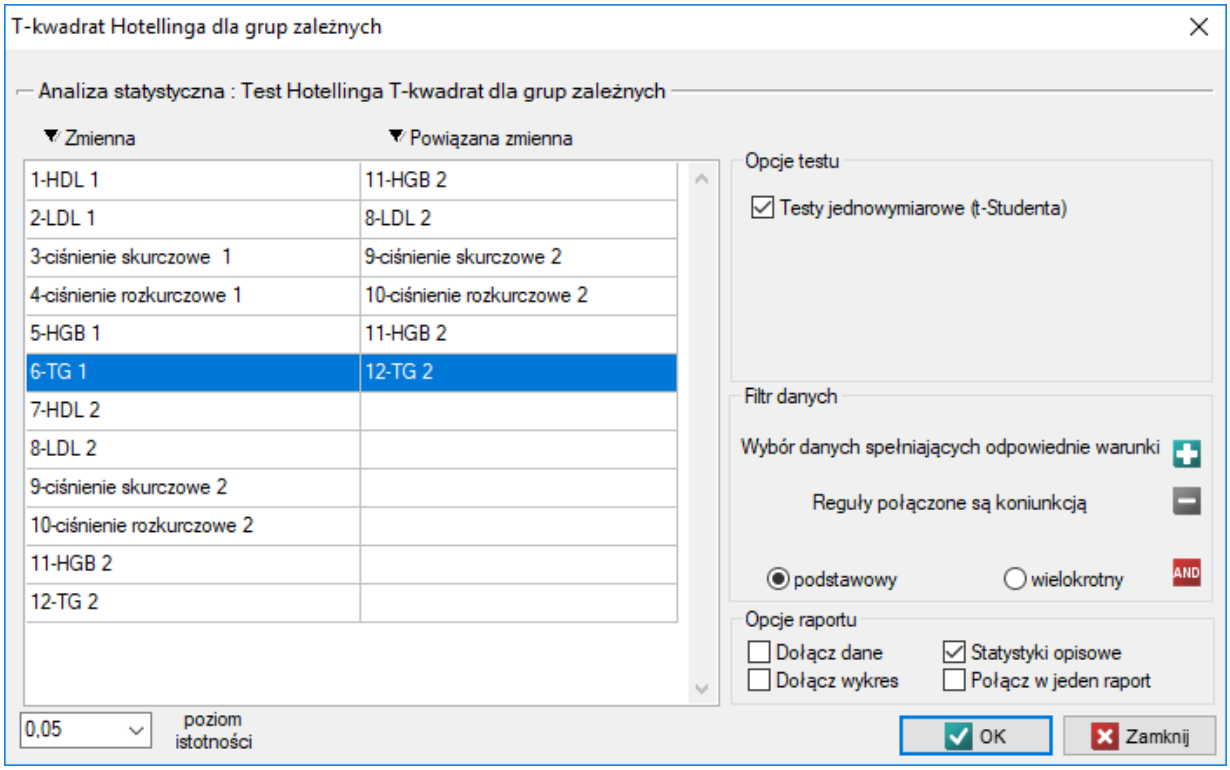
- pomiar na skali interwałowej,
- wielowymiarowy rozkład normalny lub normalność rozkładu każdej badanej zmiennej,
- model zależny.

Uwaga! Zasada działania testu Hotellinga jest tożsama z budową wielowymiarowej elipsy przedziałów ufności wokół centrum wyznaczonego przez średnie różnic (patrz przykład interpretacji elipsy testu Hotellinga dla pojedynczej próby). Przez to, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi), często nie jesteśmy w stanie uzyskać tożsamych wyników.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → T-kwadrat Hotellinga dla grup zależnych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

PRZYKŁAD 19.2. (plik nadcisnienie.pqs)
W grupie osób chorujących na nadciśnienie badano wpływ zastosowanego leczenia na zmiany wskaźników takich jak: cholesterol we frakcji HDL i LDL, hemoglobina (HGB), trójglicerydy (TG) oraz wartości ciśnienia skurczowego i rozkurczowego krwi. Pomiary od 44 pacjentów pobrano dwukrotnie (przed leczeniem i po 3 miesiącach stosowania leczenia). Następnie porównano uzyskane wyniki.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test Hotellinga T-kwadrat dla grup zależnych	
Czas analizy	0,09 sek.
Analizowane zmienne	HDL 1
	LDL 1
	ciśnienie skurczowe 1
	ciśnienie rozkurczowe 1
	HGB 1
	TG 1
	HDL 2
	LDL 2
	ciśnienie skurczowe 2
	ciśnienie rozkurczowe 2
	HGB 2
	TG 2
Poziom istotności	0,05
Liczność	44
T-kwadrat	39,19982
Statystyka F	5,773617
Stopnie swobody 1	6
Stopnie swobody 2	38
Wartość p	0,00024

$p = 0.00024$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Średnie wartości badanych parametrów zmieniają się na skutek leczenia (różnica jest istotnie różna od 0).

Przedziały ufności dla ciśnienia skurczowego oraz rozkurczowego znajdują się powyżej wartości 0, co świadczy o istotnym obniżeniu tych parametrów na skutek leczenia. Przedziały dla pozostałych parametrów zawierają wartość 0, więc nie mamy dowodów na ich zmianę na skutek leczenia.

Przedziały ufności [Jednoczesne]			
	różnica śre	-95% CI	+95% CI
HDL 1/HDL	0,019091	-0,199511	0,237693
LDL 1/LDL	0,095682	-0,347687	0,53905
ciśnienie sł	8,022727	1,111148	14,934306
ciśnienie rc	5,795455	0,205321	11,385588
HGB 1/HGE	0,232273	-0,174718	0,639264
TG 1/TG 2	-0,110227	-0,518751	0,298297

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przedziały ufności [Benferroni]			
	różnica śre	-95% CI	+95% CI
HDL 1/HDL	0,019091	-0,132306	0,170488
LDL 1/LDL	0,095682	-0,211383	0,402746
ciśnienie sł	8,022727	3,235961	12,809493
ciśnienie rc	5,795455	1,923885	9,667024
HGB 1/HGE	0,232273	-0,049598	0,514143
TG 1/TG 2	-0,110227	-0,39316	0,172705

Podejście jednowymiarowe, ze względu na swoją prostotę, wykorzystywane jest najczęściej. Stosując to podejście wraz z korektą wielokrotnych porównań, również uznamy, że różnice dotyczą jedynie wartości ciśnienia. By wykonać korektę wartości p testu t-Studenta, należy przekopiować te wartości do jednej kolumny nowego arkusza danych i z menu Statystyka wybrać Korektę wielokrotnych porównań.

Test t-Studenta						
	Liczność	-95% CI dl	+95% CI c	Statystyka	Stopnie sw	Wartość p
HDL 1/HDL	44	-0,091299	0,129481	0,348767	43	0,728966
LDL 1/LDL	44	-0,128212	0,319576	0,86184	43	0,393556
ciśnienie sł	44	4,532491	11,512963	4,63561	43	0,000033
ciśnienie rc	44	2,972527	8,618382	4,140259	43	0,000159
HGB 1/HGE	44	0,026749	0,437797	2,279164	43	0,02768
TG 1/TG 2	44	-0,316525	0,096071	-1,07754	43	0,287247

Korekta wielokrotnych porównań							
Czas analizy	0,04 sek.						
Liczność	6						
Poziom istotności	0,05						
Zmienna: Wartość p (t-Studenta)	Bonferroni	Bonferroni - Holm	Bonferroni - Hochberg	Sidak	Sidak - Holm	Sidak - Hochberg	Benjamini - Hochberg
0,728966	1	0,861741	0,728966	0,999604	0,728966	0,728966	0,728966
0,393556	1	0,861741	0,728966	0,950256	0,637909	0,632226	0,472267
0,000033	0,000198	0,000198	0,000198	0,000198	0,000198	0,000198	0,000198
0,000159	0,000954	0,000795	0,000795	0,000954	0,000795	0,000795	0,000477
0,02768	0,16608	0,11072	0,11072	0,155003	0,106207	0,106207	0,05536
0,287247	1	0,861741	0,728966	0,86889	0,637909	0,632226	0,430871

DLA ZAINTERESOWANYCH

Definicja: równoważnie, gdy wyznaczymy różnice pomiędzy parami pomiarów $d_1, d_2, ..., d_k$, hipoteza wskaże, że średnie dla różnic w badanej populacji wynoszą 0.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_0 = 0, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_{j0} \text{ są równe zero,} \end{aligned}$$

gdzie $\mu_0 = (\mu_{01}, \mu_{02}, ..., \mu_{0k})$ – populacyjne średnie różnic pomiaru pierwszego i drugiego.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Wzory: statystyka testowa ma postać:

$$F = \frac{n-k}{k(n-1)}T^2$$

gdzie:
$n = n_1 = n_2 = ... = n_k$ – liczności poszczególnych różnic w próbie,
$T^2 = n(\overline{x} - \overline{y})^T S^{-1}(\overline{x} - \overline{y})$ – pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznościach),
$\overline{x} = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_k)$ – średnie zmiennych w próbie dla pierwszego pomiaru,
$\overline{y} = (\overline{y}_1, \overline{y}_2, ..., \overline{y}_k)$ – średnie zmiennych w próbie dla drugiego pomiaru,
S – macierz kowariancji różnic pomiaru pierwszego i drugiego.

Statystyka ta podlega rozkładowi F Snedecora z k i $n-k$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Jednoczesne przedziały ufności

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności dla różnic średnich:

$$\overline{x}_j - \overline{y}_j \pm \sqrt{\frac{k(n-1)}{n-k}F_{\alpha,df_1,df_2}} \cdot SE_{(diff)_j}$$

lub przedziały z poprawką Bonferroniego, w celu sprawdzenia, czy znajduje się w nich wartość 0. Jeśli różnica może wynosić 0, oznacza to, że w rzeczywistości różnica pomiędzy badanymi wartościami może nie istnieć. Stosując tę metodę, należy pamiętać, że wyznaczone przedziały nie uwzględniają powiązań pomiędzy zmiennymi towarzyszącymi (które uwzględnia test Hotellinga), a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice, możemy również zastosować podejście jednowymiarowe, wykonując porównania testem t-Studenta dla grup zależnych oddzielnie dla poszczególnych zmiennych. Niestety, nie uwzględnimy tym samym wzajemnych powiązań, ale uzyskane wartości p testu t-Studenta możemy skorygować w dziale Wielokrotne porównania.

Test T-kwadrat Hotellinga dla grup niezależnych

KIEDY STOSOWAĆ TEN TEST:

Test pozwala sprawdzić równość średnich badanych k zmiennych $X_1, X_2, ..., X_k$ z populacji pierwszej i średnich tych samych k zmiennych $Y_1, Y_2, ..., Y_k$ z populacji drugiej.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,

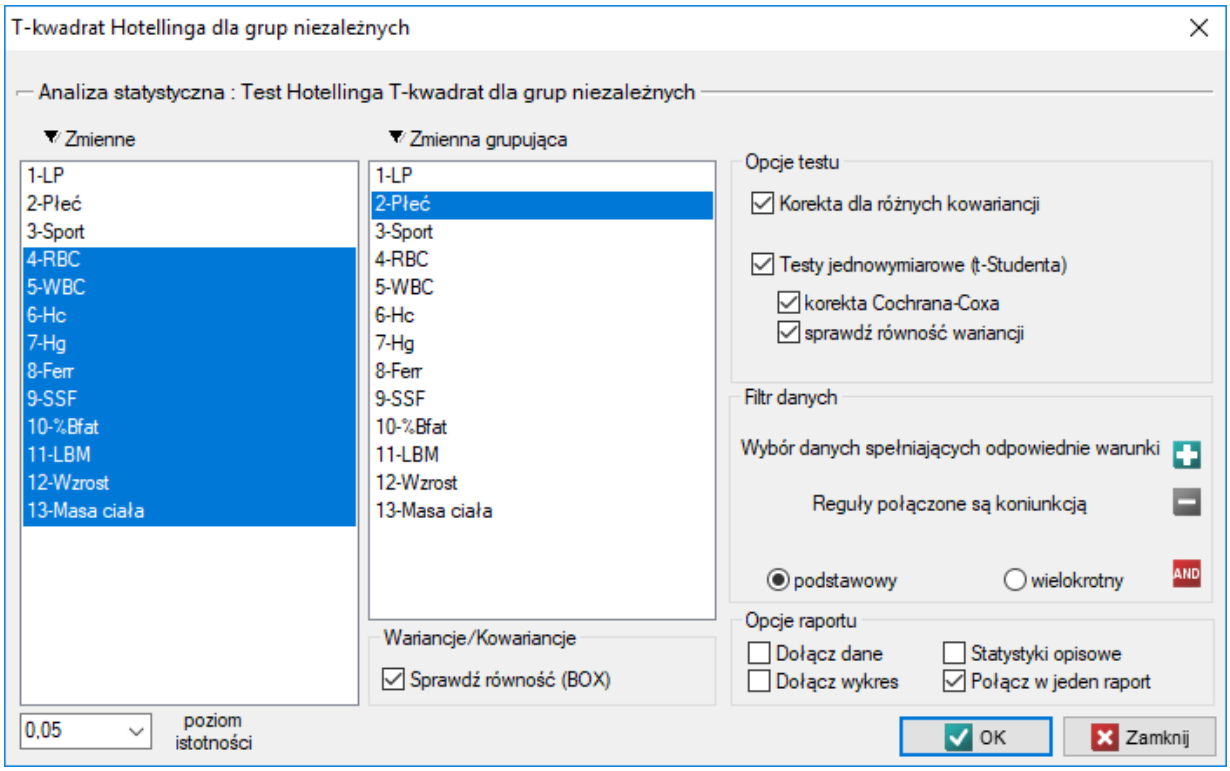
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- wielowymiarowy rozkład normalny w obu populacjach lub normalność rozkładu każdej badanej zmiennej w obu populacjach,
- model niezależny,
- równość wariancji badanych zmiennych obu populacji lub równość macierzy kowariancji – warunek szczególnie istotny w przypadku grup o różnych licznościach; gdy warunek nie jest spełniony, powinna zostać wyznaczona korekta.

Uwaga! Zasada działania testu Hotellinga jest tożsama z budową wielowymiarowych elips przedziałów ufności wokół centrów wyznaczonych przez średnie (patrz przykład interpretacji elipsy testu Hotellinga dla pojedynczej próby). Przez to, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi), często nie jesteśmy w stanie uzyskać tożsamych wyników.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → T-kwadrat Hotellinga dla grup niezależnych.



WYKONANIE I INTERPRETACJA

Przykład 19.3. (plik sport.pqs)
Przebadano grupę sportowców, by uzyskać informację o parametrach zdrowotnych takich jak:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

RBC – liczba czerwonych krwinek,
WBC – liczba białych krwinek,
Hc – hematokryt [%],
Hg – hemoglobina [g/dl],
Ferr – ferrytyna [µg/l],
SSF – suma fałdów skórno-tłuszczowych [mm],
% Bfat – zawartość tłuszczu w organizmie,
LBM – beztłuszczowa masa ciała [kg],
Wzrost [cm],
Masa ciała [kg].

Chcemy wiedzieć, czy kobiety i mężczyźni uprawiający zawodowo analizowane rodzaje sportu różnią się poziomami tych parametrów.

Ze względu na różnice dotyczące kowariancji analizowanych parametrów (wartość p testu Boxa < 0.000001), analizę przeprowadzamy przy włączonej korekcie różnych macierzy kowariancji.

BOX	
Stopnie swobody 1	55
Stopnie swobody 2	67477,108211
Statystyka M Boxa	260,302979
Statystyka F	4,394912
Wartość p	<0.000001

Uzyskany wynik skorygowanej statystyki Hotellinga jest przedstawiony poniżej.

Test Hotellinga T-kwadrat dla grup niezależnych	
Czas analizy	0,08 sek.
Analizowane zmienne	RBC
	WBC
	Hc
	Hg
	Ferr
	SSF
	%Bfat
	LBM
	Wzrost
	Masa ciała
	Płeć
Poziom istotności	0,05
Zmienna grupująca	Płeć
Liczność	151
Korekta dla różnych kowariancji	Tak
T-kwadrat	1048,269666
Statystyka F	98,495136
Stopnie swobody 1	10
Stopnie swobody 2	146,896444
Wartość p	<0.000001

$p < 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Mężczyźni i kobiety w badanej populacji mają inne wartości średnie analizowanych parametrów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Według jednoczesnych przedziałów ufności (oraz według przedziałów z korektą Bonferroniego) różnice dotyczą: RBC, Hc, Hg, Ferr oraz LBM (średnio kobiety uzyskują niższe wartości), a także SSF i % Bfat (średnio mężczyźni uzyskują niższe wartości). Pozostałe parametry, czyli WBC, wzrost i masa ciała, nie różnią się istotnie.

Przedziały ufności [jednoczesne]

	różnica śre	-95% CI d	+95% CI (
RBC	-0,614035	-0,867942	-0,360128
WBC	0,351732	-0,831448	1,534912
Hc	-4,96117	-6,915348	-3,006992
Hg	-2,004401	-2,685508	-1,323294
Ferr	-35,94450:	-64,79191:	-7,097089
SSF	43,953517	25,780916	62,126118
%Bfat	10,164473	7,273499	13,055448
LBM	-17,64161	-23,55338:	-11,72983'
Wzrost	-4,145087	-11,08362:	2,793454
Waga	-7,092418	-16,70593:	2,521096

Przedziały ufności [Benferroni]

	różnica śre	-95% CI d	+95% CI (
RBC	-0,614035	-0,87343	-0,35464
WBC	0,351732	-0,856277	1,559741
Hc	-4,96117	-6,956879	-2,965461
Hg	-2,004401	-2,699977	-1,308825
Ferr	-35,94450:	-65,45032:	-6,438684
SSF	43,953517	25,366351	62,540684
%Bfat	10,164473	7,207175	13,121771
LBM	-17,64161	-23,68889:	-11,59432:
Wzrost	-4,145087	-11,23407:	2,9439
Waga	-7,092418	-16,91733:	2,732503

Ze względu na swoją prostotę najczęściej stosuje się podejście jednowymiarowe, choć nie uwzględnia ono wielokrotnego testowania ani wzajemnych powiązań między zmiennymi. Jego wyniki wskazują, że różnice dotyczą wszystkich badanych parametrów za wyjątkiem WBC, przy czym wyniki testu t-Studenta interpretujemy po uwzględnieniu korekty Cochrana-Coxa lub bez tej korekty, w zależności od spełnienia założenia równości wariancji (wynik testu Fishera-Snedecora). Chcąc uwzględnić wielokrotne testowanie, należałoby w tym przypadku zastosować jedną z korekt wartości p, opisanych w dziale Wielokrotne porównania.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Testy t-Studenta

	Liczność	-95% CI d	+95% CI (	Statystyka	Stopnie sw	Wartość p
RBC	151	-0,72358	-0,50449	-11,07615	149	<0.000001
WBC	151	-0,173743	0,877207	1,322665	149	0,187973
Hc	151	-5,814318	-4,108022	-11,49078	149	<0.000001
Hg	151	-2,301872	-1,70693	-13,31463	149	<0.000001
Ferr	151	-48,08041	-23,80859	-5,852617	149	2,9656E-8
SSF	151	35,396283	52,510752	10,149626	149	-8,8818E-1
%Bfat	151	8,800804	11,528142	14,728746	149	<0.000001
LBM	151	-20,12450	-15,15871	-14,04006	149	<0.000001
Wzrost	151	-7,133263	-1,15691	-2,741052	149	0,006874
Waga	151	-11,20206	-2,982775	-3,410201	149	0,000835

Testy t-Studenta CC

	Liczność	-95% CI d	+95% CI (	Statystyka	Stopnie sw	Wartość p
RBC	151	-0,725877	-0,502193	-10,86252	149	<0.000001
WBC	151	-0,168795	0,872259	1,335282	149	0,183834
Hc	151	-5,82134	-4,101	-11,40334	149	<0.000001
Hg	151	-2,304198	-1,704604	-13,21847	149	<0.000001
Ferr	151	-48,68136	-23,20764	-5,596774	149	1,8161E-7
SSF	151	35,930029	51,977005	10,863974	149	<0.000001
%Bfat	151	8,887741	11,441206	15,79257	149	<0.000001
LBM	151	-20,25231	-15,03090	-13,40394	149	<0.000001
Wzrost	151	-7,201795	-1,088378	-2,683352	149	0,00826
Waga	151	-11,33011	-2,854721	-3,313785	149	0,00122

F Fishera-Snedecora

	Statystyka	Wartość p
RBC	0,63312	0,048854
WBC	1,246044	0,351424
Hc	0,838017	0,444019
Hg	0,845643	0,467539
Ferr	0,319929	0,000001
SSF	8,255157	<0.000001
%Bfat	9,189042	<0.000001
LBM	0,303421	4,2149E-7
Wzrost	0,605913	0,030927
Waga	0,502903	0,003151

DLA ZAINTERESOWANYCH

Hipotezy:

$$\mathcal{H}_0: \quad \mu_x = \mu_y,$$
$$\mathcal{H}_1: \quad \text{nie wszystkie } \mu_{xj} = \mu_{yj},$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

gdzie:
$\mu_x = (\mu_{x1}, \mu_{x2}, ..., \mu_{xk})$ – średnie zmiennych w populacji pierwszej,
$\mu_y = (\mu_{y1}, \mu_{y2}, ..., \mu_{yk})$ – średnie zmiennych w populacji drugiej.

Wzory: statystyka testowa ma postać:

$$F = \frac{n-k}{k(n-1)} T^2$$

gdzie:
$n = n_x + n_y - 1$,
n_x, n_y – liczności pierwszej i drugiej próby (liczności poszczególnych zmiennych takie same),
$T^2 = (\overline{x} - \overline{y})^T \left(S \left(\frac{1}{n_x} + \frac{1}{n_y} \right) \right)^{-1} (\overline{x} - \overline{y})$ – pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznościach),
$\overline{x} = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_k)$ – średnie zmiennych w próbie pierwszej,
$\overline{y} = (\overline{y}_1, \overline{y}_2, ..., \overline{y}_k)$ – średnie zmiennych w próbie drugiej,
S – macierz kowariancji wspólna (ang. pooled) dla obu prób.

Statystyka ta podlega rozkładowi F Snedecora z k i $n-k$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Jednoczesne przedziały ufności

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności dla różnic średnich:

$$\overline{x}_j - \overline{y}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha, df_1, df_2}} \cdot SE_{(diff)_j}$$

lub przedziały z poprawką Bonferroniego, w celu sprawdzenia, czy znajduje się w nich wartość 0. Jeśli różnica może wynosić 0, oznacza to, że w rzeczywistości różnica pomiędzy badanymi wartościami może nie istnieć. Stosując tę metodę, należy pamiętać, że wyznaczone przedziały nie uwzględniają powiązań pomiędzy zmiennymi (które uwzględnia test Hotellinga), a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice, możemy również zastosować podejście jednowymiarowe, wykonując porównania testem t-Studenta dla grup niezależnych oddzielnie dla poszczególnych zmiennych. Niestety, nie uwzględnimy tym samym wzajemnych powiązań, ale uzyskane wartości p testu t-Studenta możemy skorygować w dziale Wielokrotne porównania.

Test T-kwadrat Hotellinga dla grup niezależnych z korektą dla różnych wariancji

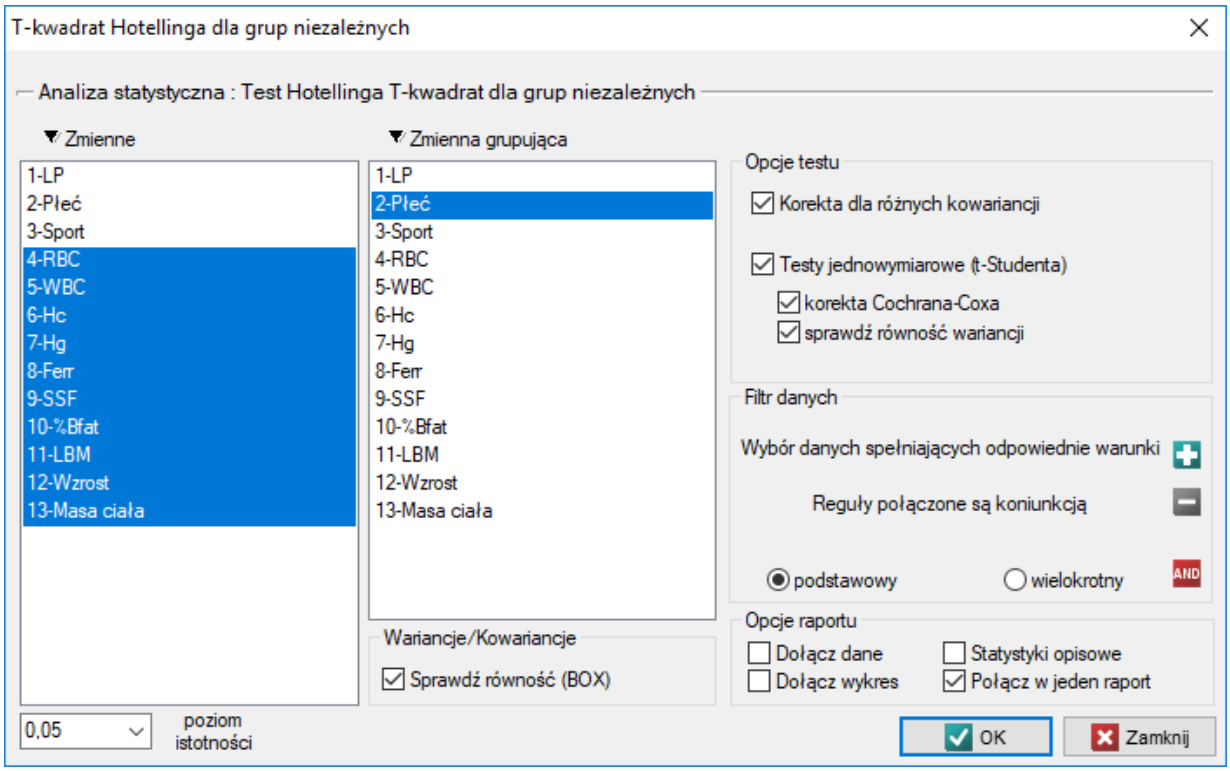
KIEDY STOSOWAĆ TEN TEST:

Poprawka dotyczy testu T-kwadrat Hotellinga dla grup niezależnych i jest wyliczana wówczas, gdy wariancje badanych zmiennych w obu populacjach są różne.

JAK USTAWIĆ ANALIZĘ

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Korekta jest opcją dostępną w tym samym oknie co test T-kwadrat Hotellinga dla grup niezależnych.



WYKONANIE I INTERPRETACJA

Patrz Przykład 19.3 powyżej – ze względu na istotny wynik testu Boxa, analizę tam przeprowadzono właśnie z zastosowaniem tej korekty.

DLA ZAINTERESOWANYCH

Wzory: statystyka testowa ma postać:

$$F = \frac{n-k}{k(n-1)}T^2$$

gdzie:
$n = n_x + n_y - 1$,
n_x, n_y – liczności pierwszej i drugiej próby (liczności poszczególnych zmiennych równe),
$T^2 = (\overline{x}-\overline{y})^T \left(\frac{S_x}{n_x} + \frac{S_y}{n_y}\right)^{-1} (\overline{x}-\overline{y})$ – pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznościach),
$\overline{x} = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_k)$ – średnie zmiennych w próbie pierwszej,
$\overline{y} = (\overline{y}_1, \overline{y}_2, ..., \overline{y}_k)$ – średnie zmiennych w próbie drugiej,
S_x, S_y – macierz kowariancji dla pierwszej i drugiej próby.

Statystyka ta podlega rozkładowi F Snedecora z k i m stopniami swobody (gdzie m to liczba stopni swobody skorygowana ze względu na różne wariancje).

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Jednoczesne przedziały ufności

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności dla różnic średnich:

$$\overline{x}_j - \overline{y}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha,df_1,df_2}} \cdot SE_{(diff)_j}$$

lub przedziały z poprawką Bonferroniego, w celu sprawdzenia, czy znajduje się w nich wartość 0. Jeśli różnica może wynosić 0, oznacza to, że w rzeczywistości różnica pomiędzy badanymi wartościami może nie istnieć. Stosując tę metodę, należy pamiętać, że wyznaczone przedziały nie uwzględniają powiązań pomiędzy zmiennymi (które uwzględnia test Hotellinga), a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice, możemy również zastosować podejście jednowymiarowe, wykonując porównania testem t-Studenta z korektą Cochrana-Coxa oddzielnie dla poszczególnych zmiennych. Niestety, nie uwzględnimy tym samym wzajemnych powiązań, ale uzyskane wartości p testu t-Studenta możemy skorygować w dziale Wielokrotne porównania.

Test Box'a równości macierzy kowariancji

KIEDY STOSOWAĆ TEN TEST:

Test służy do porównania dwóch lub więcej ($m \geq 2$) macierzy kowariancji opisujących niezależne populacje.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- wielowymiarowy rozkład normalny w badanych populacjach lub normalność rozkładu każdej badanej zmiennej w każdej populacji,
- model niezależny.

Uwaga! Test Boxa jest szczególnie wrażliwy na złamanie założenia normalności rozkładu.

JAK USTAWIĆ ANALIZĘ

Test Boxa jest wyliczany opcjonalnie w Hotellingu dla grup niezależnych lub w analizie MANOVA.

WYKONANIE I INTERPRETACJA

PATRZ PRZYKŁAD 19.3.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \Sigma_1 = \Sigma_2 = ... = \Sigma_m, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \Sigma_j \text{ są sobie równe } (j = 1, 2, ..., m), \end{aligned}$$

gdzie $\Sigma_1, \Sigma_2, ..., \Sigma_m$ – populacyjne macierze kowariancji.

Wzory: statystyka testowa ma postać:

$$F = \frac{M}{b}$$

gdzie:
$M = (n - m) \ln |S| - \sum_{j=1}^{m} (n_j - 1) \ln |S_j|$,
S – wspólna (ang. pooled) macierz kowariancji,
S_j – macierz kowariancji dla j-tej próby,
$b = \frac{df_1}{1 - c_1 - \frac{df_1}{df - 2}}$,
$df_1 = \frac{k(k+1)(m-1)}{2}$,
$df_2 = \frac{df_1 + 2}{|c_2 - c_1^2|}$,
k – liczba analizowanych zmiennych,
$n = n_1 = n_2 = ... = n_k$ – liczności poszczególnych zmiennych w próbie.

Statystyka ta podlega rozkładowi F Snedecora z df_1 i df_2 stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

20 JEDNOCZYNNIKOWA MANOVA

Jednoczynnikowa MANOVA dla grup niezależnych

KIEDY STOSOWAĆ TEN TEST:

Jednoczynnikowa MANOVA (*ang. multivariate analysis of variance*) jest rozszerzeniem jednoczynnikowej ANOVA dla grup niezależnych. Pozwala sprawdzić równość średnich badanych k zmiennych w kilku ($m \geq 2$) populacjach. W odróżnieniu od wielokrotnego (oddzielnie dla każdej zmiennej) stosowania ANOVA, MANOVA bada różnice pomiędzy populacjami jednocześnie dla wielu zmiennych, uwzględniając przy tym ich wzajemne skorelowanie. Podejście to jest też stosowane jako alternatywa dla ANOVA dla grup zależnych, gdyż nie wymaga spełnienia założenia sferyczności. Zalecamy wcześniejsze zapoznanie się z analizą T-kwadrat Hotellinga.

WARUNKI STOSOWANIA

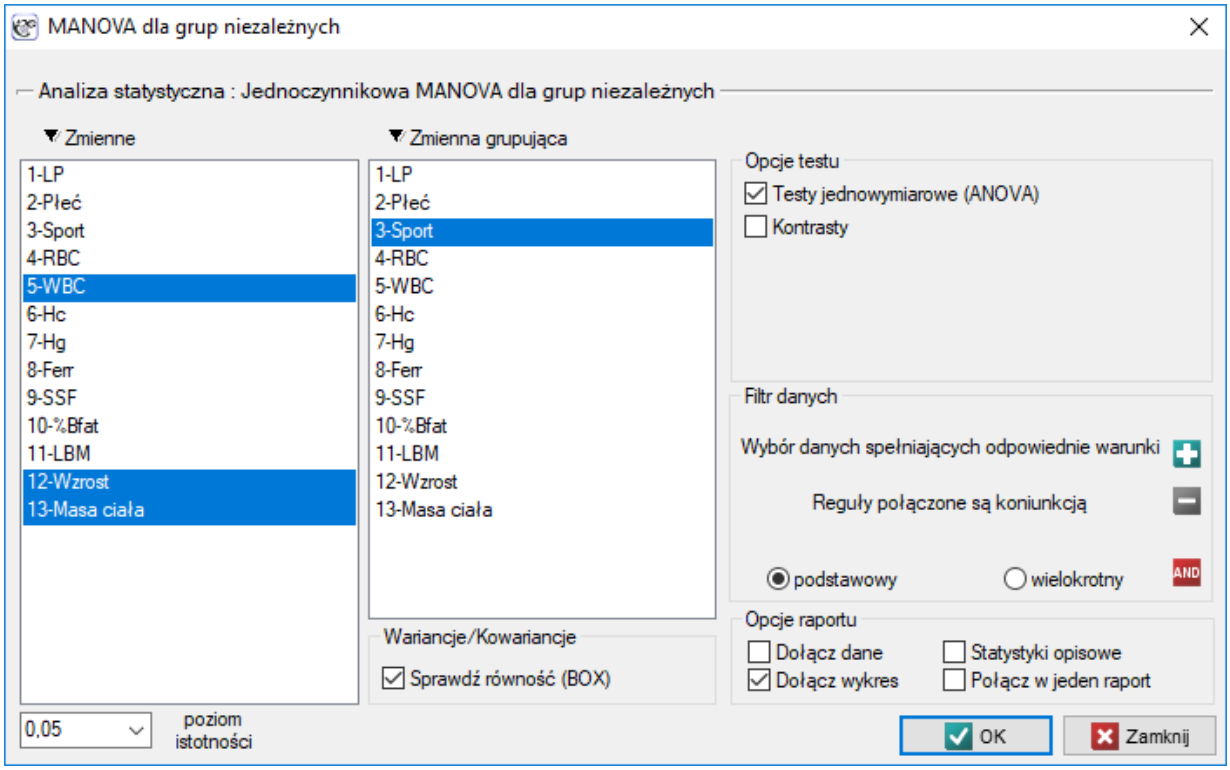
- pomiar na skali interwałowej,
- wielowymiarowy rozkład normalny w każdej populacji lub normalność rozkładu każdej badanej zmiennej w każdej populacji,
- model niezależny,
- równość macierzy kowariancji lub równość wariancji badanych zmiennych dla porównywanych populacji – warunek szczególnie istotny w przypadku grup o różnych licznościach.

Uwaga! Zasada działania MANOVA (podobnie jak testów Hotellinga) jest tożsama z budową wielowymiarowych elips przedziałów ufności wokół centrów wyznaczonych przez średnie (patrz przykład interpretacji elipsy testu Hotellinga dla pojedynczej próby). Przez to, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi), często nie jesteśmy w stanie uzyskać tożsamych wyników.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → Jednoczynnikowa MANOVA dla grup niezależnych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

Przykład 20.1. (plik sport.pqs)
Przebadano grupę sportowców, by uzyskać informację o parametrach zdrowotnych, m.in. takich jak:
WBC – liczba białych krwinek,
Wzrost [cm],
Masa ciała [kg].

Chcemy wiedzieć:

1. czy uprawiający zawodowo trzy rodzaje sportów: „gry zespołowe” (jak: koszykówka, siatkówka itp.), „bieżnia” (jak: 100m, 400m itp.), „wodne” (jak: pływanie, wioślarstwo itp.), różnią się poziomami tych parametrów;

2. czy uprawiający sporty wymagające dużego wysiłku, takie jak „bieżnia” i „wodne”, różnią się poziomami tych parametrów od uprawiających „gry zespołowe”.

Ad. 1)
Hipotezy:

$\mathcal{H}_0$: średnie analizowanych parametrów są takie same dla sportowców uprawiających poszczególne rodzaje sportu,
$\mathcal{H}_1$: przynajmniej jeden parametr ma inną średnią wartość dla porównywanych populacji.

Wynik testu Boxa ($p = 0.63025$) pozwala na wyliczenie analizy typu MANOVA.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednoczynnikowa MANOVA dla grup niezależnych	
Czas analizy	0,17 sek.
Analizowane zmienne	WBC;Wzrost;Waga;Sport
Poziom istotności	0,05
Zmienna grupująca	Sport
Liczność	151
Liczba grup	3
Liczba zmiennych w modelu	3
Wilk's Lambda	0,892927
Eta-kwadrat	0,055052
df1	6
df2	292
Statystyka F	2,835288
Wartość p	0,010672
Hotelling-Lawley Trace	0,117386
Eta-kwadrat	0,055439
df1	6
df2	290
Statystyka F	2,836819
Wartość p	0,010645
Pillai-Bartlett Trace	0,10933
Eta-kwadrat	0,054665
df1	6
df2	294
Statystyka F	2,833475
Wartość p	0,010705
BOX	
Stopnie swobody 1	12
Stopnie swobody 2	93631,926908
Statystyka M Boxa	10,140583
Statystyka F	0,819757
Wartość p	0,63025

istotne Lambda Wilksa, ślad Hotelling-Lawley oraz ślad Pillai-Bartlett (istotne statystycznie, $\alpha = 0.05$)

Badane populacje sportowców różnią się co do analizowanych parametrów.

By określić, których zmiennych dotyczą różnice, przeprowadzamy analizę jednowymiarową ANOVA.

Jednoczynnikowa ANOVA			
	WBC	Wzrost	Waga
SS[BG]	15,968207	554,06012	1262,2410
SS[WG]	383,48487	12857,218	24771,693
df[BG]	2	2	2
df[WG]	148	148	148
Statystyka	3,08134	3,188905	3,770668
Wartość p	0,048854	0,044064	0,02528

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ANOVA-POST-HOC (Fisher LSD) [WBC]			
	zespołowe	bieżnia	wodne
Wartość p			
zespołowe		0,058516	0,020444
bieżnia	0,058516		0,772989
wodne	0,020444	0,772989	

ANOVA-POST-HOC (Fisher LSD) [Wzrost]			
	zespołowe	bieżnia	wodne
Wartość p			
zespołowe		0,017512	0,571842
bieżnia	0,017512		0,051325
wodne	0,571842	0,051325	

ANOVA-POST-HOC (Fisher LSD) [Waga]			
	zespołowe	bieżnia	wodne
Wartość p			
zespołowe		0,00763	0,074656
bieżnia	0,00763		0,280787
wodne	0,074656	0,280787	

Do uzyskanych wyników należy podejść ostrożnie. Choć wskazują one na istotne różnice we wszystkich porównywanych parametrach, zwracają wartość p na pograniczu istotności statystycznej (dla WBC $p = 0.04885$, dla wzrostu $p = 0.04406$, dla wagi $p = 0.02528$). Dodatkowo, przy ich interpretacji należy pamiętać, że nie uwzględniają one ani wzajemnego skorelowania parametrów, ani wielokrotnego testowania. Uwzględnienie wielokrotnego testowania wymagałoby w tym przypadku zastosowania jednej z korekt wartości p, opisanych w dziale Wielokrotne porównania.

Ad. 2)
Hipotezy:

$\mathcal{H}_0$: średnie analizowanych parametrów dla „gier zespołowych”
nie różnią się od odpowiednich średnich sportowców z dwóch pozostałych grup,
$\mathcal{H}_1$: przynajmniej jeden parametr ma inną średnią wartość
dla porównywanych populacji.

By sprawdzić prawdziwość powyższych hipotez, w oknie analizy ustawiamy odpowiedni kontrast. Jako wartość kontrastu wpisujemy 2 dla sportów zespołowych, a -1 dla bieżni i dyscyplin określanych jako wodne.

Copyright ©2010-2026 U-smile Stats − Wszelkie prawa zastrzeżone

Kontrasty dla analizy wielowymiarowej:	
Kontrast dla grupy [zespołowe]	2
Kontrast dla grupy [bieżnia]	-1
Kontrast dla grupy [wodne]	-1
Wilk's Lambda	0,918374
Eta-kwadrat	0,081626
df1	3
df2	146
Statystyka F	4,325515
Wartość p	0,005924
Hotelling-Lawley Trace	0,08888
Eta-kwadrat	0,081626
df1	3
df2	146
Statystyka F	4,325515
Wartość p	0,005924
Pillai-Bartlett Trace	0,081626
Eta-kwadrat	0,081626
df1	3
df2	146
Statystyka F	4,325515
Wartość p	0,005924

Przedziały ufności efektu [jednoczesne]			
Grupa	Efekt	-95% CI	+95% CI
WBC	1,373722	-0,233977	2,981422
Wzrost	-5,700571	-15,00959	3,608457
Waga	-11,81871	-24,74008	1,102659

Przedziały ufności efektu [Benferroni]			
Grupa	Efekt	-95% CI	+95% CI
WBC	1,373722	0,006633	2,740812
Wzrost	-5,700571	-13,6164	2,215257
Waga	-11,81871	-22,80626	-0,831169

$p = 0.00592$ dla Lambda Wilksa, śladu Hotelling-Lawley oraz śladu Pillai-Bartlett (istotne statystycznie, $\alpha = 0.05$)

Sportowcy uprawiający sporty wymagające dużego wysiłku różnią się analizowanymi parametrami od uprawiających gry zespołowe.

W przedziałach jednoczesnych nie obserwujemy tych różnic, natomiast na bazie przedziałów Bonferroniego możemy stwierdzić, że różnica dotyczy wagi i WBC. Wartości WBC są wyższe w grupie sportów zespołowych, a masa ciała jest istotnie niższa w tej grupie.

DEFINICJE MIAR

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Cząstkowa η^2 – miara wielkości efektu określająca proporcję wariancji wyjaśnionej do wariancji całkowitej związanej z danym czynnikiem. W modelu jednoczynnikowej MANOVA dla grup niezależnych wskazuje, jaka część wewnątrzosobowej zmienności wyników może być przypisana badanemu czynnikowi wyznaczającemu grupy niezależne.

Kontrasty (analiza jednowymiarowa) – gdy wykonywana analiza ma porównać wybrane populacje lub wybrany układ populacji, przeprowadzamy analizę kontrastów, analogiczną do kontrastów w analizie jednowymiarowej, ale uwzględniającą wzajemne powiązanie zmiennych.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_m, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_i \text{ są sobie równe,} \end{aligned}$$

gdzie:
$\mu_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{ik})$ – średnie zmiennych w i-tej populacji,
$(i = 1, 2, ..., m)$,
$(j = 1, 2, ..., k)$.

Wyjaśnienie:
W analizach MANOVA posługujemy się kilkoma współczynnikami. Najszerzej znanym jest Lambda Wilksa. Ślad Pillai-Bartlett jest najbardziej konserwatywny, ale stosunkowo odporny w przypadku naruszenia założeń stosowania MANOVA i preferowany przy mało licznych próbach. Ślad Hotelling-Lawley jest natomiast najmniej konserwatywny z trzech zaproponowanych testów. Pionierskie prace nad rozwojem tych technik rozpoczęli: Wilks (1932)[179], Pillai (1955)[136], Lawley (1938)[101], Hotelling (1951)[87] i Roy (1939)[144].

Statystyki testowe bazują na macierzach sum kwadratów i iloczynów mieszanych ($SSCP$ – ang. *Sums of Squares and Cross Products*). Macierz całkowita $T = SSCP$ zostaje rozbita na dwie macierze: pierwsza związana jest z analizowaną hipotezą i oznaczana przez H (macierz międzygrupowych sum kwadratów i iloczynów mieszanych), a druga dotyczy reszt (błędów) i jest oznaczana przez E (macierz wewnątrzgrupowych sum kwadratów i iloczynów mieszanych).

Lambda Wilksa – wzory:
Wartość Lambdy definiowana jest następująco:

$$\Lambda = \frac{|E|}{|H + E|}$$

Statystyka testowa ma postać:

$$F = \frac{1 - \Lambda^{\frac{1}{b}}}{\Lambda^{\frac{1}{b}}} \frac{df_2}{df_1}$$

gdzie $df_1 = k(m-1)$, $df_2 = ab-c$, a a, b, c to współczynniki zależne od liczby analizowanych zmiennych i liczby porównywanych populacji.

Ślad Hotelling-Lawley – wzory:
Ślad Hotelling-Lawley definiowany jest następująco:

$$T_0 = trace(HE^{-1})$$

Statystyka testowa ma postać:

$$F = \frac{T_0^2}{s} \frac{df_2}{df_1}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie $df_1 = s(2t + s + 1)$, $df_2 = 2(su + 1)$, a s, t, u to współczynniki zależne od liczby analizowanych zmiennych i liczby porównywanych populacji.

Ślad Pillai-Bartlett – wzory:
Ślad Pillai-Bartlett definiowany jest następująco:

$$V = trace(H(H + E)^{-1})$$

Statystyka testowa ma postać:

$$F = \frac{V}{s - V}\frac{df_2}{df_1}$$

gdzie $df_1 = s(2t+s+1)$, $df_2 = s(2u+s+1)$, a s, t, u to współczynniki zależne od liczby analizowanych zmiennych i liczby porównywanych populacji.

Każda z powyższych statystyk testowych podlega rozkładowi F Snedecora z df_1 i df_2 stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Cząstkowa η^2 – wzór:

$$\eta^2 = \frac{F \cdot df_1}{F \cdot df_1 + df_2}$$

Kontrasty (analiza jednowymiarowa)
Dla wielkości efektu można także wyznaczyć **jednoczesne przedziały ufności** lub przedziały ufności z poprawką Bonferroniego. Stosując te przedziały, należy jednak pamiętać, że nie uwzględniają one powiązań pomiędzy poszczególnymi zmiennymi (które uwzględnia MANOVA), a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice, możemy również zastosować podejście jednowymiarowe, wykonując porównania ANOVA dla grup niezależnych oddzielnie dla poszczególnych zmiennych. Niestety, nie uwzględnimy tym samym wzajemnych powiązań, ale uzyskane na podstawie ANOVA wartości p możemy skorygować w dziale Wielokrotne porównania.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

21 ANALIZA DLA WARSTW

21.1 Metoda Mantela-Haenszela dla wielu tabel 2x2

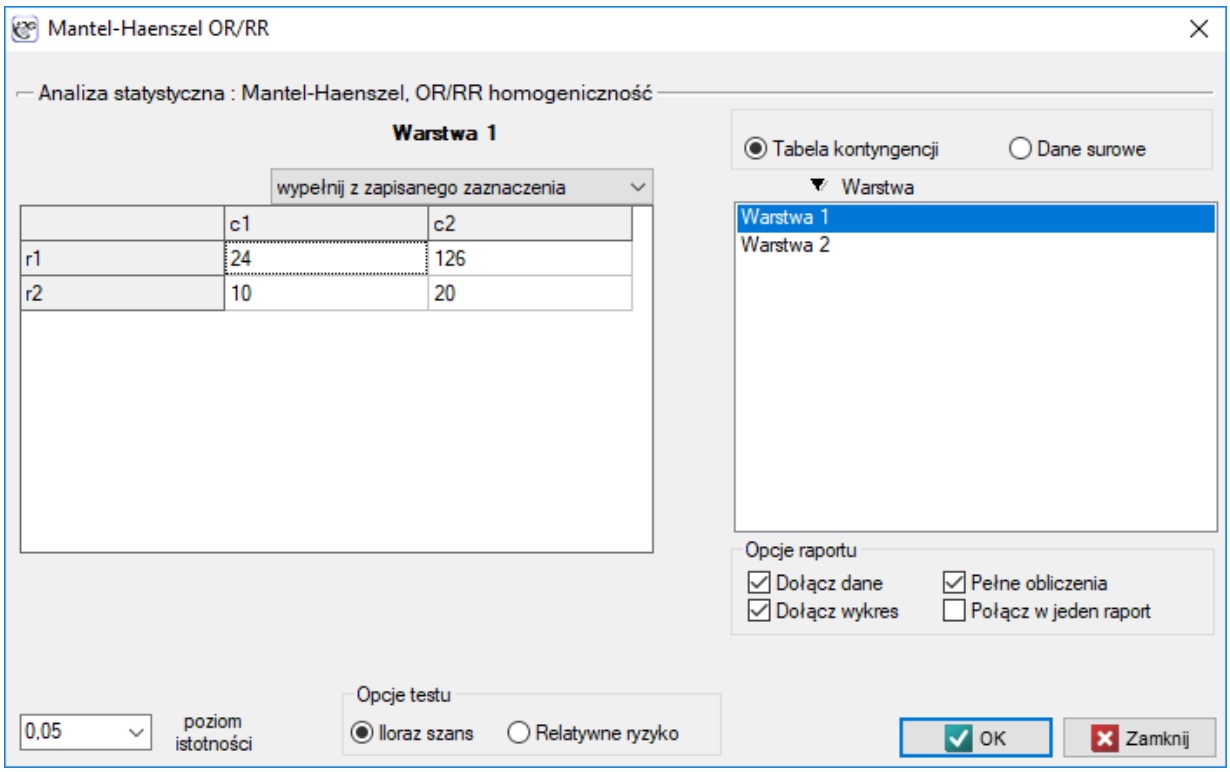
KIEDY STOSOWAĆ:

Metoda Mantela-Haenszela dla tabel 2×2 (*ang. Mantel-Haenszel method for* 2×2 *tables*), zaproponowana przez Mantela i Haenszela (1959)[114] a następnie rozszerzona przez Mantela (1963)[115] (szerszy przegląd rozwoju tych metod przeprowadził m.in. Newman (2001)[128]), może być używana m.in. do analizy tabel 2×2 występujących w kilku ($w \geq 2$) warstwach tworzonych przez tzw. zmienną uwikłaną (*ang. confounding*). Dla kolejnych warstw ($s = 1, ..., w$) budowane są tabele kontyngencji 2×2 liczności obserwowanych:

Liczności obserwowane s-tej warstwy $\left(O_{ij}^{(s)}\right)$		badane zjawisko		
		występuje	nie występuje	Suma
czynnik ryzyka	występuje	$O_{11}^{(s)}$	$O_{12}^{(s)}$	$O_{11}^{(s)} + O_{12}^{(s)}$
	nie występuje	$O_{21}^{(s)}$	$O_{22}^{(s)}$	$O_{21}^{(s)} + O_{22}^{(s)}$
	Suma	$O_{11}^{(s)} + O_{21}^{(s)}$	$O_{12}^{(s)} + O_{22}^{(s)}$	$n^{(s)} = O_{11}^{(s)} + O_{12}^{(s)} + O_{21}^{(s)} + O_{22}^{(s)}$

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji metod Mantel−Haenszel OR/RR wywołujemy poprzez menu Statystyka→Analiza dla warstw→Mantel−Haenszel OR/RR.



Copyright ©2010-2026 U-smile Stats − Wszelkie prawa zastrzeżone

Iloraz szans Mantela-Haenszela

KIEDY STOSOWAĆ:

Jeśli wszystkie tabele (tworzone przez poszczególne warstwy) są homogeniczne, na podstawie tych tabel można wyznaczyć wspólny iloraz szans wraz z przedziałami ufności. Taki iloraz szans jest średnią ważoną ilorazów szans wyznaczonych dla poszczególnych warstw. Zastosowanie ważonej metody zaproponowanej przez Mantela i Haenszela pozwala uwzględnić wkład (wagę), jaki do budowy wspólnego ilorazu szans wnosi każda warstwa (im bardziej liczna warstwa, tym większy ma wpływ na powstały iloraz szans).

Wyznaczonemu w ten sposób ilorazowi szans towarzyszą dwa testy istotności. **Test χ^2 Mantela-Haenszela dla** OR_{MH} sprawdza, czy wspólny iloraz szans istotnie różni się od 1, czyli czy w ogóle istnieje zależność między badanymi cechami po uwzględnieniu warstw. **Test χ^2 homogeniczności dla** OR sprawdza z kolei założenie leżące u podstaw całej metody – czy iloraz szans jest rzeczywiście zbliżony we wszystkich warstwach, czy też zmienna tworząca warstwy istotnie modyfikuje siłę zależności (wówczas łączenie warstw w jeden wspólny iloraz szans byłoby nieuprawnione).

DEFINICJE MIAR

Iloraz szans Mantela-Haenszela – ważona średnia ilorazów szans wyznaczonych dla poszczególnych warstw, uwzględniająca liczność każdej warstwy.

WYKONANIE I INTERPRETACJA

Przykład 21.1. (plik leptospiroza.pqs)
Przedstawiona tabela pokazuje hipotetyczne wyniki ankiety przeprowadzonej w celu wykrycia czynników ryzyka występowania leptospirozy wśród mieszkańców miasta i wsi Zachodnich Indii (gdzie obszar wiejski traktowany jest jako czynnik narażenia)[20]. Występowanie przeciwciał leptospirozy jest pośrednim dowodem na obecność zakażenia.

Liczności obserwowane O_{ij}		przeciwciała leptospirozy	
		tak	nie
miejsce zamieszkania	wieś	60	140
	miasto	60	140

Szansa wystąpienia przeciwciał leptospirozy mieszkańców miasta i wsi jest taka sama (OR=1). Chcemy wiedzieć, jaka będzie szansa wystąpienia przeciwciał leptospirozy, gdy w analizie weźmiemy pod uwagę również płeć. Podzielimy więc próbę na 2 warstwy ze względu na płeć (są one zapamiętane w pliku jako zaznaczony zakres):

Liczności obserwowane dla mężczyzn		przeciwciała leptospirozy	
		tak	nie
miejsce zamieszkania	wieś	36	14
	miasto	50	50

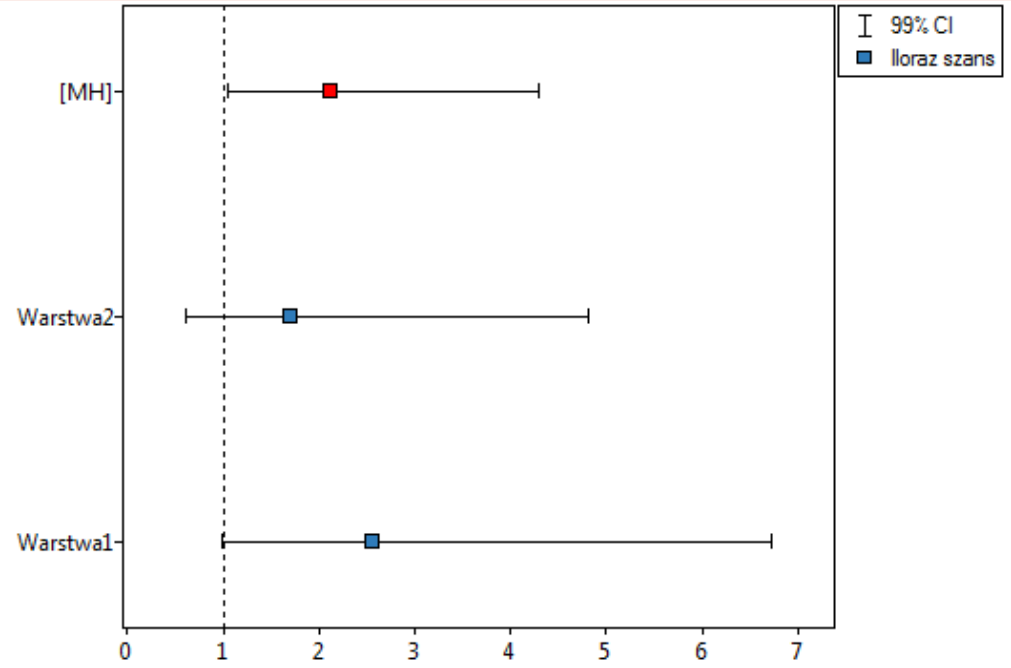
Liczności obserwowane dla kobiet		przeciwciała leptospirozy	
		tak	nie
miejsce zamieszkania	wieś	24	126
	miasto	10	90

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Płeć jest powiązana z obydwoma czynnikami (z występowaniem przeciwciał leptospirozy i z miejscem zamieszkania w Zachodnich Indiach). Jest to czynnik wikłający, którego zignorowanie może prowadzić do błędnych wyników.

Mantel-Haenszel, OR/RR homogeniczność	
Czas analizy	0.01 sek.
Analizowane zmienne	Tabela kontyngencji
Poziom istotności	0.05
Liczność	400
Warstwa 1	
Iloraz szans	2.571429
-95% CI dla ilorazu szans	1.237622
+95% CI dla ilorazu szans	5.342701
Statystyka dla ilorazu szans	2.531365
Wartość p	0.011362
Warstwa 2	
Iloraz szans	1.714286
-95% CI dla ilorazu szans	0.781346
+95% CI dla ilorazu szans	3.76117
Statystyka dla ilorazu szans	1.344493
Wartość p	0.178789
Iloraz szans [MH]	2.126374
-95% CI dla ilorazu szans [MH]	1.244338
+95% CI dla ilorazu szans [MH]	3.63363
Ilość stopni swobody	1
Statystyka dla ilorazu szans [MH]	7.81939
Wartość p	0.005169
Homogeniczność ilorazu szans	
Ilość stopni swobody	1
Statystyka	0.548611
Wartość p	0.458886

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



OR_{MH}[95%CI]=2.13[1.24, 3.65], test χ^2 Mantela-Haenszela: $p = 0.005169$ (istotne statystycznie, $\alpha = 0.05$)

Szansa wystąpienia przeciwciał leptospirozy jest istotnie większa dla osób zamieszkujących tereny wiejskie.

Zanim odczytamy wspólny iloraz szans, sprawdzamy test χ^2 homogeniczności dla OR – tabele są homogeniczne ($p = 0.465049$), a więc łączenie warstw w jeden wspólny iloraz szans jest uprawnione. Szansa wystąpienia przeciwciał leptospirozy jest większa dla mieszkańców wsi, zarówno dla kobiet (OR[95%CI]=2.57[1.24, 5.34]), jak i dla mężczyzn (OR[95%CI]=1.71[0.78, 3.76]). Test χ^2 Mantela-Haenszela potwierdza, że tak wyznaczony wspólny iloraz szans jest istotny statystycznie.

DLA ZAINTERESOWANYCH

Wzory:

Wagi dla każdej warstwy wyznacza się zgodnie z wzorem:

$$g^{(s)} = \frac{O_{21}^{(s)} \cdot O_{12}^{(s)}}{n^{(s)}},$$

a **iloraz szans Mantela-Haenszela**:

$$OR_{MH} = \frac{R}{S},$$

gdzie:

$$R = \sum_{s=1}^{w} \frac{O_{11}^{(s)} \cdot O_{22}^{(s)}}{n^{(s)}},$$

$$S = \sum_{s=1}^{w} g^{(s)}.$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przedział ufności dla $\log OR_{MH}$ wyznacza się na podstawie błędu standardowego (RGB – Robins-Breslow-Greenland[139][140]) wyliczonego ze wzoru:

$$SE_{MH} = \sqrt{\frac{T}{2R^2} + \frac{U+Y}{2RS} + \frac{W}{2S^2}},$$

gdzie:

$$T = \sum_{s=1}^{w} T^{(s)}, \qquad T^{(s)} = \frac{O_{11}^{(s)} \cdot O_{22}^{(s)} \cdot \left(O_{11}^{(s)} + O_{22}^{(s)}\right)}{\left(n^{(s)}\right)^2},$$

$$U = \sum_{s=1}^{w} U^{(s)}, \qquad U^{(s)} = \frac{O_{21}^{(s)} \cdot O_{12}^{(s)} \cdot \left(O_{11}^{(s)} + O_{22}^{(s)}\right)}{\left(n^{(s)}\right)^2},$$

$$Y = \sum_{s=1}^{w} Y^{(s)}, \qquad Y^{(s)} = \frac{O_{11}^{(s)} \cdot O_{22}^{(s)} \cdot \left(O_{21}^{(s)} + O_{12}^{(s)}\right)}{\left(n^{(s)}\right)^2},$$

$$W = \sum_{s=1}^{w} W^{(s)}, \qquad W^{(s)} = \frac{O_{21}^{(s)} \cdot O_{12}^{(s)} \cdot \left(O_{21}^{(s)} + O_{12}^{(s)}\right)}{\left(n^{(s)}\right)^2}.$$

Test χ^2 Mantela-Haenszela dla OR_{MH}

Test χ^2 Mantela-Haenszela (*ang. Mantel-Haenszel Chi-square test*) służy do weryfikacji hipotezy o istotności wyznaczonego ilorazu szans (OR_{MH}) i powinien być wyliczany przy dużych licznościach, tzn. gdy są spełnione obydwa warunki tzw. „reguły 5”:

- $\min(O_{11}^{(s)} + O_{12}^{(s)}, O_{11}^{(s)} + O_{21}^{(s)}) - \sum_{s=1}^{w} E_{11}^{(s)} \geq 5$ dla wszystkich warstw $s = 1, 2, ..., w$,
- $\max(0, O_{11}^{(s)} - O_{22}^{(s)}) \geq 5$ dla wszystkich warstw $s = 1, 2, ..., w$.

Kiedy w tabeli występują wartości zerowe, stosowana jest korekta na ciągłość (powiększenie liczności o 0,5), zarówno dla liczności obserwowanych, jak i dla liczności oczekiwanych.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad OR_{MH} = 1, \\ \mathcal{H}_1 &: \quad OR_{MH} \neq 1. \end{aligned}$$

Wzory:

Statystyka testowa ma postać:

$$\chi^2_{MH} = \frac{\left(\sum_{s=1}^{w} O_{11}^{(s)} - \sum_{s=1}^{w} E_{11}^{(s)}\right)^2}{V},$$

gdzie:

$E_{11}^{(s)} = \dfrac{\left(O_{11}^{(s)} + O_{21}^{(s)}\right)\left(O_{11}^{(s)} + O_{12}^{(s)}\right)}{n^{(s)}}$ to wartości oczekiwane w pierwszej komórce tabeli kontyngencji dla poszczególnych warstw $s = 1, 2, ..., w$,

$$V = \sum_{s=1}^{w} V^{(s)},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$$V^{(s)} = \frac{\left(O_{11}^{(s)} + O_{12}^{(s)}\right)\left(O_{21}^{(s)} + O_{22}^{(s)}\right)\left(O_{11}^{(s)} + O_{21}^{(s)}\right)\left(O_{12}^{(s)} + O_{22}^{(s)}\right)}{\left(n^{(s)}\right)^2\left(n^{(s)} - 1\right)}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z jednym stopniem swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test χ^2 homogeniczności dla OR
Test χ^2 homogeniczności dla OR (*ang. Chi-square test of homogeneity for OR*) służy do weryfikacji hipotezy, że zmienna tworząca warstwy jest efektem modyfikującym, tzn. wpływa ona na wyznaczany iloraz szans w taki sposób, że jest on znacząco inny dla poszczególnych warstw.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad OR_{MH} = OR^{(s)}, \text{ dla wszystkich warstw } s = 1, 2, ..., w, \\ \mathcal{H}_1: &\quad OR_{MH} \neq OR^{(s)}, \text{ dla przynajmniej jednej warstwy.} \end{aligned}$$

Wzory:
Statystyka testowa (Breslow-Day (1980)[26], Tarone (1985)[27][162]) ma postać:

$$\chi^2 = \sum_{s=1}^{w} \frac{\left(O_{11}^{(s)} - E^{(s)}\right)^2}{Var^{(s)}} - \frac{\left(\sum_{s=1}^{w} O_{11}^{(s)} - \sum_{s=1}^{w} E^{(s)}\right)^2}{\sum_{s=1}^{w} Var^{(s)}}$$

gdzie $E^{(s)}$ jest rozwiązaniem równania kwadratowego:

$$\frac{E^{(s)}\left(O_{22}^{(s)} - O_{11}^{(s)} + E^{(s)}\right)}{\left(O_{11}^{(s)} + O_{21}^{(s)} - E^{(s)}\right)\left(O_{11}^{(s)} + O_{12}^{(s)} - E^{(s)}\right)} = OR_{MH},$$

$$Var^{(s)} = \left(\frac{1}{E^{(s)}} + \frac{1}{O_{22}^{(s)} - O_{11}^{(s)} + E^{(s)}} + \frac{1}{O_{11}^{(s)} + O_{21}^{(s)} - E^{(s)}} + \frac{1}{O_{11}^{(s)} + O_{12}^{(s)} - E^{(s)}}\right)^{-1}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody $df = w - 1$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Relatywne ryzyko Mantela-Haenszela

KIEDY STOSOWAĆ:

Jeśli wszystkie tabele (tworzone przez poszczególne warstwy) są homogeniczne, na podstawie tych tabel można wyznaczyć wspólne relatywne ryzyko wraz z przedziałami ufności. Takie relatywne ryzyko jest średnią ważoną wartości relatywnego ryzyka wyznaczonego dla poszczególnych warstw, z uwzględnieniem wkładu (wagi), jaki do budowy wspólnego relatywnego ryzyka wnosi każda warstwa.

Podobnie jak w przypadku ilorazu szans, wyznaczonemu w ten sposób relatywnemu ryzyku towarzyszą dwa testy istotności. **Test χ^2 Mantela-Haenszela dla RR_{MH}** sprawdza, czy wspólne relatywne ryzyko istotnie różni się od 1. **Test χ^2 homogeniczności dla RR** sprawdza z kolei, czy relatywne ryzyko jest

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

zbliżone we wszystkich warstwach, czy też zmienna tworząca warstwy istotnie modyfikuje jego wartość (wówczas łączenie warstw w jedno wspólne relatywne ryzyko byłoby nieuprawnione).

DEFINICJE MIAR

Relatywne ryzyko Mantela-Haenszela – ważona średnia wartości relatywnego ryzyka wyznaczonych dla poszczególnych warstw, uwzględniająca liczność każdej warstwy.

WYKONANIE I INTERPRETACJA

Wykonanie i interpretacja są analogiczne do ilorazu szans Mantela-Haenszela – patrz Przykład 21.1 powyżej: przed odczytaniem wspólnego relatywnego ryzyka sprawdzamy test χ^2 homogeniczności dla RR, a jego istotność statystyczną potwierdzamy testem χ^2 Mantela-Haenszela dla RR_{MH}.

DLA ZAINTERESOWANYCH

Wzory:
Wagi dla każdej warstwy wyznacza się zgodnie z wzorem:

$$g^{(s)} = \frac{O_{21}^{(s)}\left(O_{11}^{(s)} + O_{12}^{(s)}\right)}{n^{(s)}},$$

a **relatywne ryzyko Mantela-Haenszela**:

$$RR_{MH} = \frac{R}{S},$$

gdzie:

$$R = \sum_{s=1}^{w} \frac{O_{11}^{(s)}\left(O_{21}^{(s)} + O_{22}^{(s)}\right)}{n^{(s)}},$$

$$S = \sum_{s=1}^{w} g^{(s)}.$$

Przedział ufności dla log RR_{MH} wyznacza się na podstawie błędu standardowego wyliczonego ze wzoru:

$$SE_{MH} = \sqrt{\frac{V}{RS}},$$

gdzie:

$$V = \sum_{s=1}^{w} V^{(s)},$$

$$V^{(s)} = \frac{\left(O_{11}^{(s)} + O_{12}^{(s)}\right)\left(O_{21}^{(s)} + O_{22}^{(s)}\right)\left(O_{11}^{(s)} + O_{21}^{(s)}\right) - \left(O_{11}^{(s)} * O_{21}^{(s)} * n^{(s)}\right)}{\left(n^{(s)}\right)^2}.$$

Test χ^2 Mantela-Haenszela dla RR_{MH}
Test χ^2 Mantela-Haenszela (*ang. Mantel-Haenszel Chi-square test*) służy do weryfikacji hipotezy o istotności wyznaczonego relatywnego ryzyka (RR_{MH}) i powinien być wyliczany przy dużych licznościach w tabeli kontyngencji.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad RR_{MH} = 1, \\ \mathcal{H}_1: &\quad RR_{MH} \neq 1. \end{aligned}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wzory:
Statystyka testowa ma postać:

$$\chi^2_{MH} = \frac{\left(\sum_{s=1}^{w} O_{11}^{(s)} - \sum_{s=1}^{w} E_{11}^{(s)}\right)^2}{V},$$

gdzie $E_{11}^{(s)} = \frac{\left(O_{11}^{(s)}+O_{21}^{(s)}\right)\left(O_{11}^{(s)}+O_{12}^{(s)}\right)}{n^{(s)}}$ to wartości oczekiwane w pierwszej komórce tabeli kontyngencji dla poszczególnych warstw $s = 1, 2, ..., w$.

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z jednym stopniem swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test χ^2 homogeniczności dla RR
Test χ^2 homogeniczności dla RR (*ang. Chi-square test of homogeneity for* RR) służy do weryfikacji hipotezy, że zmienna tworząca warstwy jest efektem modyfikującym, tzn. wpływa ona na wyznaczane relatywne ryzyko w taki sposób, że jest ono znacząco inne dla poszczególnych warstw.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad RR_{MH} = RR^{(s)}, \text{ dla wszystkich warstw } s = 1, 2, ..., w, \\ \mathcal{H}_1: &\quad RR_{MH} \neq RR^{(s)}, \text{ dla przynajmniej jednej warstwy.} \end{aligned}$$

Wzory:
Statystyka testowa oparta o ważoną metodę najmniejszych kwadratów ma postać:

$$\chi^2 = \sum_{s=1}^{w} v^{(s)} \left(\ln(RR^{(s)}) - \ln(RR_{MH})\right)^2$$

gdzie:

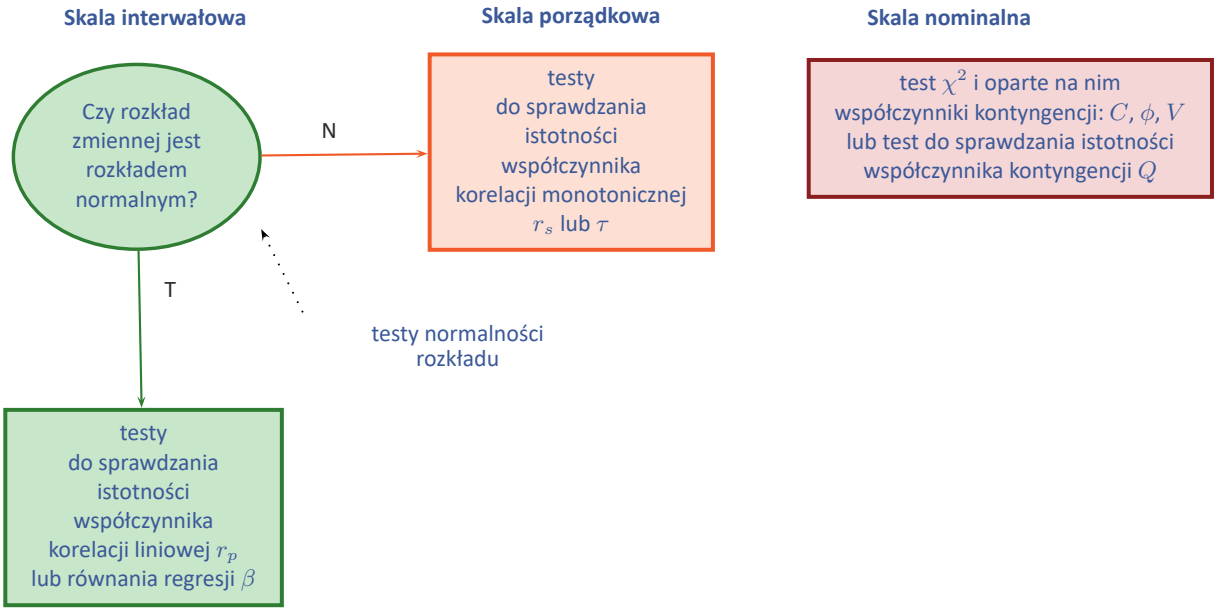
$$v^{(s)} = \left(\frac{O_{12}^{(s)}}{O_{11}^{(s)}\left(O_{11}^{(s)}+O_{12}^{(s)}\right)} + \frac{O_{22}^{(s)}}{O_{21}^{(s)}\left(O_{21}^{(s)}+O_{22}^{(s)}\right)}\right)^{-1}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody $df = w - 1$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

22 KORELACJA



Współczynniki korelacji są jedną z miar statystyki opisowej, która reprezentuje stopień korelacji (zależności) pomiędzy 2 lub więcej cechami (zmiennymi). Wybór konkretnego współczynnika zależy w głównej mierze od skali, na której dokonano pomiarów. Wyznaczenie go stanowi jeden z pierwszych etapów pracy nad analizą korelacji. Następnie istotność statystyczną otrzymanych współczynników można analizować przy pomocy testów statystycznych.

Uwaga! Zależność pomiędzy zmiennymi nie zawsze obrazuje ich związek przyczynowo-skutkowy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

22.1 TESTY PARAMETRYCZNE

Współczynniki korelacji liniowej

KIEDY STOSOWAĆ:

Współczynnik korelacji liniowej Pearsona r_p *(ang. Pearson product-moment correlation coefficient, Pearson (1896, 1900))* służy do badania siły związku liniowego pomiędzy dwiema cechami.

WARUNKI STOSOWANIA

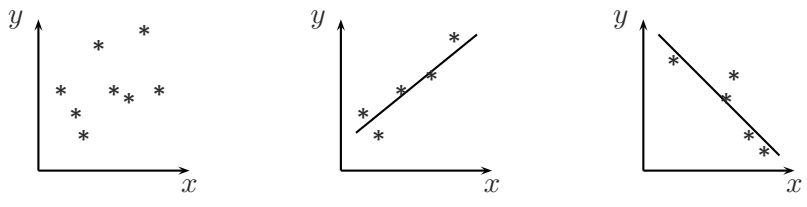
- pomiar na skali interwałowej,
- brak obserwacji odstających,
- normalność rozkładu badanych cech lub normalność reszt modelu.

DEFINICJE MIAR

Współczynnik korelacji liniowej Pearsona r_p – określa siłę i kierunek związku liniowego pomiędzy dwiema cechami interwałowymi. Wartość $r_p \in \langle -1; 1 \rangle$ interpretujemy następująco:

- $r_p \approx 1$ oznacza silną dodatnią zależność liniową, tj. punkty pomiarowe leżą blisko linii prostej, a wzrostowi zmiennej niezależnej odpowiada wzrost zmiennej zależnej;
- $r_p \approx -1$ oznacza silną ujemną zależność liniową, tj. punkty pomiarowe leżą blisko linii prostej, lecz wzrostowi zmiennej niezależnej odpowiada spadek zmiennej zależnej;
- gdy współczynnik przyjmuje wartość równą lub bardzo bliską zeru, nie istnieje liniowa zależność między badanymi cechami (może jednak istnieć związek nieliniowy).

Wykres 22.1. Interpretacja graficzna współczynnika r_p.



$r_p \approx 0$ $\quad$ $r_p \approx 1$ $\quad$ $r_p \approx -1$

Gdy jedna z badanych cech jest stała (niezależnie od zmian drugiej cechy), obie cechy nie są związane zależnością, a współczynnika r_p nie można wyznaczyć.

Uwaga! Błędem jest wyznaczanie współczynnika korelacji, gdy w próbie występują obserwacje odstające, które mogą całkowicie przekłamać wartość i znak współczynnika korelacji Pearsona, gdy próba jest wyraźnie niejednorodna, bądź gdy badana zależność wyraźnie przyjmuje kształt inny niż liniowy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Współczynnik determinacji r_p^2 – wyraża procent zmienności zmiennej zależnej tłumaczony zmiennością zmiennej niezależnej.

Tworzony model korelacji przedstawia zależność liniową postaci:

$$Y = \beta X + \alpha,$$

gdzie **współczynniki** β **i** α **równania regresji liniowej** to nachylenie i wyraz wolny prostej regresji.

DLA ZAINTERESOWANYCH

Wzory:

$$r_p = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

gdzie:
x_i, y_i – kolejne wartości cechy X i Y,
$\overline{x}, \overline{y}$ – średnie z wartości cechy X i cechy Y,
n – liczność próby.

R_p oznacza współczynnik korelacji Pearsona populacji, natomiast r_p w próbie.

Współczynniki równania regresji liniowej wyznacza się z wzorów:

$$\beta = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \qquad \alpha = \overline{y} - \beta\overline{x}.$$

Istotność współczynnika korelacji Pearsona

KIEDY STOSOWAĆ TEN TEST:

Test do sprawdzania istotności współczynnika korelacji liniowej Pearsona *(ang. test of significance for a Pearson product-moment correlation coefficient)* pozwala sprawdzić, czy pomiędzy badanymi cechami populacji istnieje zależność liniowa. Opiera się na współczynniku korelacji liniowej Pearsona r_p wyliczonym dla próby – im wartość r_p jest bliższa 0, tym słabszą zależnością związane są badane cechy.

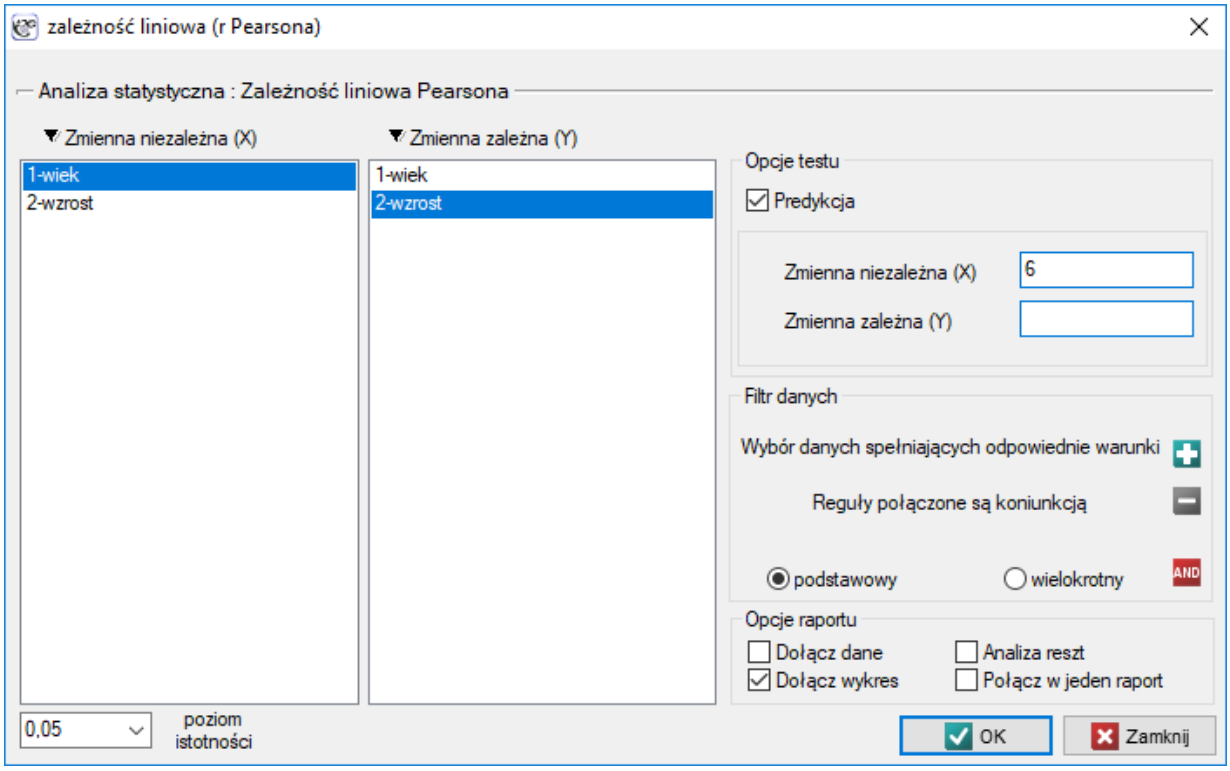
WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanych cech w populacji lub normalność reszt modelu.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → zależność liniowa (r-Pearsona) (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



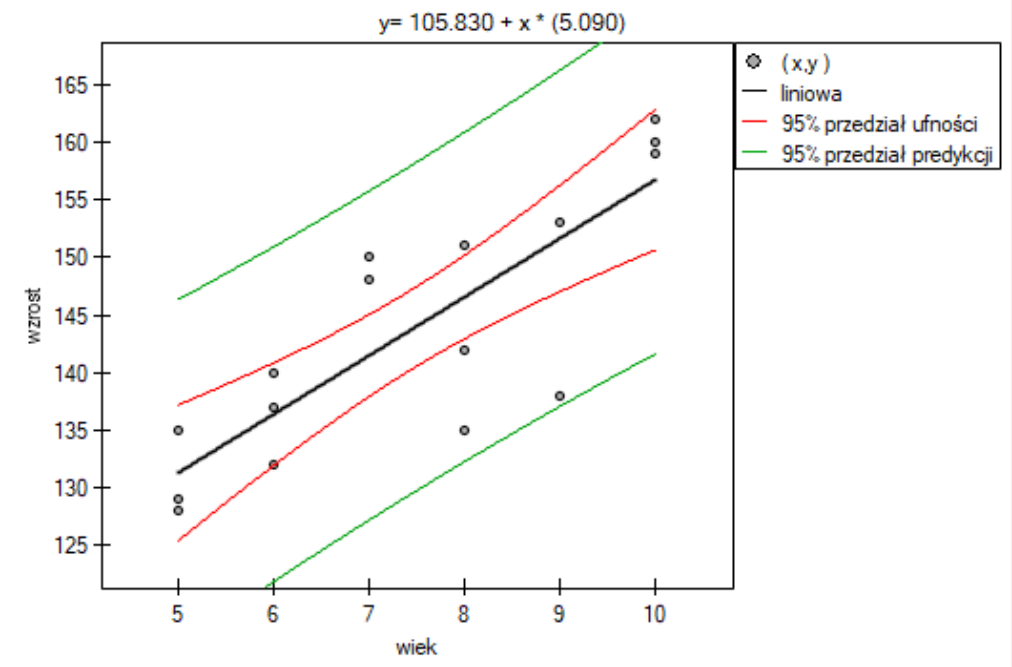
WYKONANIE I INTERPRETACJA

PRZYKŁAD 22.1. (plik wiek-wzrost.pqs)
Wśród uczniów pewnej szkoły baletowej badano zależność pomiędzy wiekiem a wzrostem. W tym celu pobrano próbę obejmującą szesnaścioro dzieci i zapisano dla nich następujące wyniki pomiaru tych cech:
(wiek, wzrost): (5, 128) (5, 129) (5, 135) (6, 132) (6, 137) (6, 140) (7, 148) (7, 150) (8, 135) (8, 142) (8, 151) (9, 138) (9, 153) (10, 159) (10, 160) (10, 162).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zależność liniowa Pearsona	
Czas analizy	0.01 sek.
Analizowane zmienne	wiek,wzrost
Poziom istotności	0.05
Liczność - ilość par	16
Nazwa grupy	wiek
Średnia grupy	7.4375
Odchylenie standardowe grupy	1.8246
Nazwa grupy	wzrost
Średnia grupy	143.6875
Odchylenie standardowe grupy	11.187605
Odchylenie standardowe reszt	6.456417
r	0.830153
r2	0.689154
Błąd dla r	0.149008
-95% CI dla współczynnika r	0.568316
+95% CI dla współczynnika r	0.939318
Statystyka t dla r	5.571206
Ilość stopni swobody	14
Wartość p	0.000069
a - współczynnik nachylenia	5.090113
Błąd dla współczynnika a	0.913646
-95% CI dla współczynnika a	3.130536
+95% CI dla współczynnika a	7.049689
statystyka t dla a	5.571206
Ilość stopni swobody	14
Wartość p	0.000069
b - współczynnik przecięcia z Y	105.829787
Błąd dla współczynnika b	6.984318
-95% CI dla współczynnika b	90.849915
+95% CI dla współczynnika b	120.809659
statystyka t dla b	15.152487
Ilość stopni swobody	14
Wartość p	<0.000001
predykcja wartości Y dla X = 6	136.370463
-95% CI dla predykcji Y	121.821348
+95% CI dla predykcji Y	150.919578

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$p = 0.000069$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Istnieje zależność liniowa pomiędzy wiekiem a wzrostem dla populacji dzieci badanej szkoły.

Zależność ta jest wprost proporcjonalna, tzn. wraz ze wzrostem wieku dzieci rośnie wysokość ciała. Współczynnik korelacji liniowej Pearsona, a zatem siła związku liniowego pomiędzy wiekiem a wzrostem, wynosi $r_p = 0.8302$. Współczynnik determinacji $r_p^2 = 0.6892$ oznacza, że ok. 69% zmienności wzrostu jest tłumaczona zmiennością wieku.

Z równania regresji postaci:

$$wzrost = 5.09 \cdot wiek + 105.83$$

można wyliczyć predykcyjną wartość dla dziecka w wieku np. 6 lat. Przewidywany wzrost takiego dziecka wynosi 136.37 cm.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad R_p = 0, \\ \mathcal{H}_1: &\quad R_p \neq 0. \end{aligned}$$

Wzory:

Statystyka testowa ma postać:

$$t = \frac{r_p}{SE},$$

gdzie $SE = \sqrt{\frac{1 - r_p^2}{n - 2}}$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wartość statystyki testowej nie może być wyznaczona, gdy $r_p = 1$ lub $r_p = -1$ albo gdy $n < 3$. Statystyka testowa ma rozkład t-Studenta z $n - 2$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Istotność współczynnika nachylenia prostej

KIEDY STOSOWAĆ TEN TEST:

Test t do sprawdzania istotności współczynników równania regresji liniowej pozwala sprawdzić, czy pomiędzy badanymi cechami populacji istnieje zależność liniowa, opierając się na współczynniku nachylenia prostej β wyliczonym dla próby – im wartość β jest bliższa 0, tym słabszą zależność dopasowana prosta przedstawia. W regresji z jedną zmienną niezależną test ten jest równoważny testowi istotności współczynnika korelacji Pearsona – daje dokładnie tę samą wartość p.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu badanych cech w populacji lub normalność reszt modelu.

WYKONANIE I INTERPRETACJA

Ze względu na równoważność z testem istotności współczynnika korelacji Pearsona przy jednej zmiennej niezależnej, wykonanie i interpretacja tego testu pokrywa się z Przykładem 22.1 powyżej.

Analiza reszt modelu – wyjaśnienie w module Liniowa Regresja Wieloraka.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad \beta = 0, \\ \mathcal{H}_1 &: \quad \beta \neq 0. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$t = \frac{\beta}{SE}$$

gdzie:

$$SE = \frac{s_{yx}}{sd_x\sqrt{n-1}},$$

$$s_{yx} = sd_y\sqrt{\tfrac{n-1}{n-2}(1-r^2)},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

sd_x, sd_y – odchylenie standardowe wartości cechy X i cechy Y.

Wartość statystyki testowej nie może być wyznaczona, gdy $r_p = 1$ lub $r_p = -1$ albo gdy $n < 3$. Statystyka testowa ma rozkład t-Studenta z $n - 2$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Predykcja

KIEDY STOSOWAĆ:

Predykcja polega na przewidywaniu wartości jednej ze zmiennych (najczęściej zmiennej zależnej y_0) na podstawie wartości innej zmiennej (najczęściej zmiennej niezależnej x_0), z wykorzystaniem wyznaczonego równania regresji liniowej. Dokładność wyznaczonej wartości określają obliczone dla niej przedziały predykcji.

DEFINICJE MIAR

Interpolacja – polega na przewidywaniu wartości zadanej zmiennej leżącej wewnątrz obszaru, dla którego wykonaliśmy model regresji. Interpolacja jest z reguły procedurą bezpieczną – zakłada się tu jedynie ciągłość funkcji wyrażającej zależność obu zmiennych.

Ekstrapolacja – polega na przewidywaniu wartości zadanej zmiennej leżącej poza obszarem, dla którego zbudowaliśmy model regresji. W przeciwieństwie do interpolacji, ekstrapolacja bywa często zabiegiem ryzykownym i dokonuje się jej jedynie w niewielkiej odległości od obszaru, dla którego powstał model regresji.

JAK USTAWIĆ ANALIZĘ

W oknie analizy zależności liniowej Pearsona zaznaczamy przycisk Predykcja i wskazujemy wartości zmiennej niezależnej, dla których chcemy przeprowadzić predykcję.

WYKONANIE I INTERPRETACJA

Patrz Przykład 22.1 powyżej – przewidywany wzrost 6-letniego dziecka wyliczony na podstawie równania regresji jest właśnie wynikiem takiej predykcji (interpolacji, ponieważ wiek 6 lat mieści się w zakresie danych, na podstawie których zbudowano model).

Porównanie współczynników korelacji

KIEDY STOSOWAĆ TEN TEST:

Test t do sprawdzania równości współczynników korelacji liniowej Pearsona pochodzących z 2 niezależnych populacji pozwala sprawdzić, czy dwa współczynniki korelacji liniowej Pearsona (R_{p_1}, R_{p_2}), wyznaczone w dwóch różnych populacjach, są sobie równe.

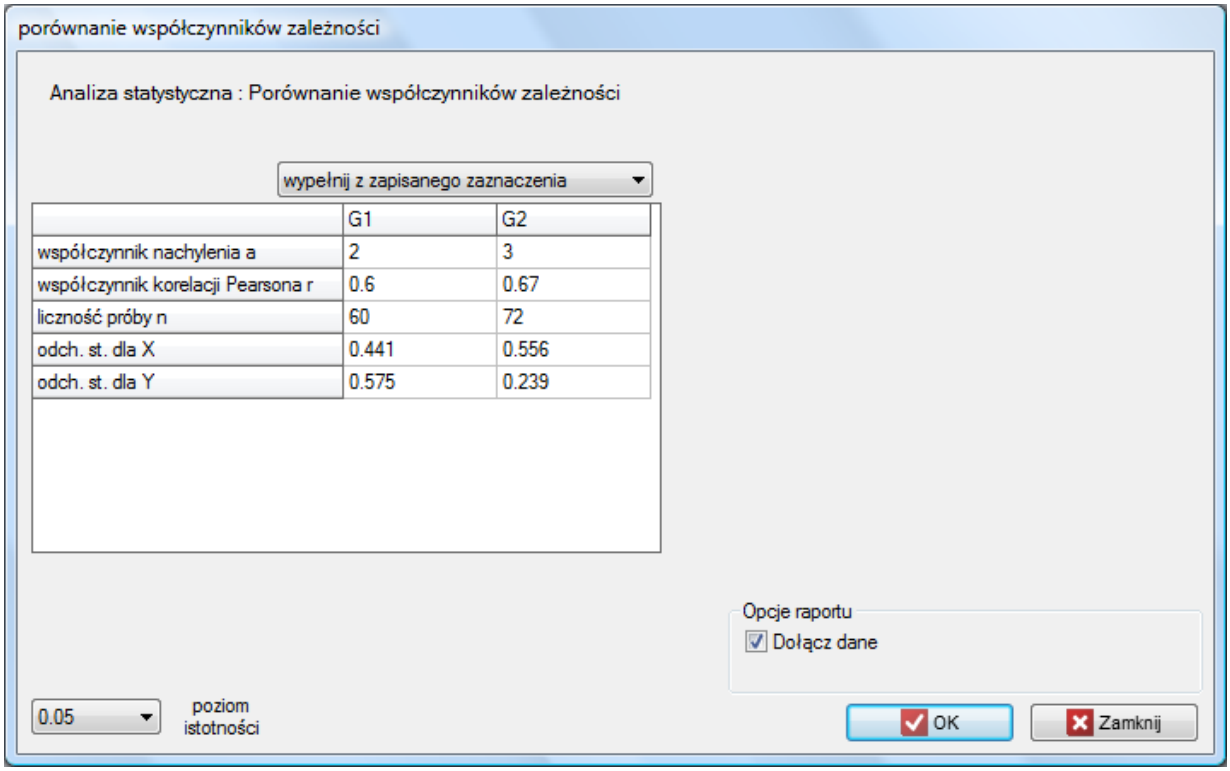
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

WARUNKI STOSOWANIA

- współczynniki r_{p_1} i r_{p_2} pochodzą z 2 prób pobranych z niezależnych populacji,
- współczynniki r_{p_1} i r_{p_2} badają zależność tych samych cech X i Y,
- znane są liczności obu prób n_1 i n_2.

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Testy parametryczne $\rightarrow$ porównanie współczynników zależności.



DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad R_{p_1} = R_{p_2}, \\ \mathcal{H}_1: &\quad R_{p_1} \neq R_{p_2}. \end{aligned}$$

Wzory: statystyka testowa ma postać:

$$t = \frac{z_{r_{p_1}} - z_{r_{p_2}}}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie:
$z_{r_{p_1}} = \frac{1}{2} \ln\left(\frac{1+r_{p_1}}{1-r_{p_1}}\right)$,
$z_{r_{p_2}} = \frac{1}{2} \ln\left(\frac{1+r_{p_2}}{1-r_{p_2}}\right)$.

Statystyka testowa ma rozkład t-Studenta z $n_1 + n_2 - 4$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Porównanie nachylenia prostych regresji

KIEDY STOSOWAĆ TEN TEST:

Test t do sprawdzania równości współczynników regresji liniowej pochodzących z 2 niezależnych populacji pozwala sprawdzić, czy dwa współczynniki regresji liniowej β_1 i β_2, wyznaczone w dwóch badanych populacjach, są sobie równe.

WARUNKI STOSOWANIA

- współczynniki β_1 i β_2 pochodzą z 2 prób pobranych z niezależnych populacji,
- współczynniki β_1 i β_2 badają zależność tych samych cech X i Y,
- znane są liczności obu prób n_1 i n_2,
- znane są odchylenia standardowe dla wartości obu cech w obu próbach: sd_{x_1}, sd_{y_1} i sd_{x_2}, sd_{y_2},
- znane są współczynniki korelacji liniowej Pearsona obu prób: r_{p_1} i r_{p_2}.

JAK USTAWIĆ ANALIZĘ

Okno ustawień jest wspólne z testem porównania współczynników korelacji (patrz zrzut ekranu powyżej) – w oknie porównanie współczynników zależności wskazujemy wówczas współczynniki nachylenia prostych regresji.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad \beta_1 = \beta_2, \\ \mathcal{H}_1 &: \quad \beta_1 \neq \beta_2. \end{aligned}$$

Wzory: statystyka testowa ma postać:

$$t = \frac{\beta_1 - \beta_2}{\sqrt{\frac{s^2_{yx_1}}{sd^2_{x_1}(n_1-1)} + \frac{s^2_{yx_2}}{sd^2_{x_1}(n_2-1)}}},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie:

$$s_{yx_1} = sd_{y_1}\sqrt{\frac{n_1-1}{n_1-2}(1-r^2_{p_1})},$$

$$s_{yx_2} = sd_{y_2}\sqrt{\frac{n_2-1}{n_2-2}(1-r^2_{p_2})}.$$

Statystyka testowa ma rozkład t-Studenta z n_1+n_2-4 stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

22.2 TESTY NIEPARAMETRYCZNE

Współczynniki korelacji monotonicznej

KIEDY STOSOWAĆ:

Zależność monotoniczna może być opisywana jako monotoniczny wzrost lub monotoniczny spadek. Związek pomiędzy 2 cechami przedstawia monotoniczny wzrost, jeżeli wzrostowi jednej cechy towarzyszy wzrost drugiej cechy, a monotoniczny spadek, jeżeli wzrostowi jednej cechy towarzyszy spadek drugiej cechy. **Współczynnik korelacji rangowej Spearmana** r_s *(ang. Spearman's rank-order correlation coefficient)* oraz **współczynnik korelacji tau Kendalla** $\tilde{\tau}$ *(ang. Kendall's tau correlation coefficient, Kendall (1938)[94])* służą do badania siły takiego związku pomiędzy dwiema cechami X i Y.

WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej.

DEFINICJE MIAR

Współczynnik korelacji rangowej Spearmana r_s – wartość $r_s \in \langle -1; 1 \rangle$ interpretujemy następująco:

- $r_s \approx 1$ oznacza silną dodatnią zależność monotoniczną (rosnącą), tj. wzrostowi zmiennej niezależnej odpowiada wzrost zmiennej zależnej;
- $r_s \approx -1$ oznacza silną ujemną zależność monotoniczną (malejącą), tj. wzrostowi zmiennej niezależnej odpowiada spadek zmiennej zależnej;
- gdy współczynnik przyjmuje wartość równą lub bardzo bliską zeru, nie istnieje monotoniczna zależność między badanymi cechami (może jednak istnieć związek niemonotoniczny, np. sinusoidalny).

R_s oznacza współczynnik korelacji rangowej Spearmana populacji, natomiast r_s w próbie.

Współczynnik korelacji tau Kendalla $\tilde{\tau}$ – wartość $\tilde{\tau} \in \langle -1; 1 \rangle$ interpretujemy następująco:

- $\tilde{\tau} \approx 1$ oznacza silną zgodność uporządkowania rang (zależność monotoniczną rosnącą), tj. wzrostowi zmiennej niezależnej odpowiada wzrost zmiennej zależnej;
- $\tilde{\tau} \approx -1$ oznacza silną niezgodność uporządkowania rang (zależność monotoniczną malejącą), tj. wzrostowi zmiennej niezależnej odpowiada spadek zmiennej zależnej;
- gdy współczynnik przyjmuje wartość równą lub bardzo bliską zeru, nie istnieje monotoniczna zależność między badanymi cechami (może jednak istnieć związek niemonotoniczny, np. sinusoidalny).

τ oznacza współczynnik korelacji Kendalla w populacji, natomiast $\tilde{\tau}$ w próbie.

Współczynnik r_s Spearmana a współczynnik $\tilde{\tau}$ Kendalla

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– dla skali interwałowej z normalnością rozkładu obu cech wartość r_s daje rezultaty bliższe wartości r_p, natomiast wartość $\tilde{\tau}$ może znacznie różnić się od r_p,

– wartość $\tilde{\tau}$ jest mniejsza bądź równa wartości r_p,

– $\tilde{\tau}$ jest nieobciążonym estymatorem parametru populacji τ, podczas gdy wartość r_s nie jest estymatorem nieobciążonym parametru R_s.

DLA ZAINTERESOWANYCH

Wzory:
Wartość współczynnika korelacji rangowej Spearmana wylicza się według wzoru:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

gdzie:
$d_i = R_{x_i} - R_{y_i}$ – różnica rang dla cechy X i cechy Y,
n – liczność d_i.

Wzór ten ulega modyfikacji, gdy występują rangi wiązane:

$$r_s = \frac{\Sigma_X + \Sigma_Y - \sum_{i=1}^{n} d_i^2}{2\sqrt{\Sigma_X \Sigma_Y}},$$

gdzie $\Sigma_X = \frac{n^3-n-T_X}{12}$, $\Sigma_Y = \frac{n^3-n-T_Y}{12}$, $T_X = \sum_{i=1}^{s}(t_{i_{(X)}}^3 - t_{i_{(X)}})$, $T_Y = \sum_{i=1}^{s}(t_{i_{(Y)}}^3 - t_{i_{(Y)}})$, a t to liczba przypadków wchodzących w skład rangi wiązanej. Poprawka na rangi wiązane powinna być stosowana, gdy rangi wiązane występują – w przeciwnym razie wzór sprowadza się do postaci opisanej wcześniejszym równaniem.

Wartość współczynnika korelacji tau Kendalla wylicza się według wzoru:

$$\tilde{\tau} = \frac{2(n_C - n_D)}{\sqrt{n(n-1)-T_X}\sqrt{n(n-1)-T_Y}},$$

gdzie:
n_C – liczba par obserwacji, dla których wartości rang dla cechy X jak i dla cechy Y zmieniają się w tym samym kierunku (liczba par zgodnych),
n_D – liczba par obserwacji, dla których wartości rang dla cechy X zmieniają się w innym kierunku niż dla cechy Y (liczba par niezgodnych),
$T_X = \sum_{i=1}^{s}(t_{i_{(X)}}^2 - t_{i_{(X)}})$, $T_Y = \sum_{i=1}^{s}(t_{i_{(Y)}}^2 - t_{i_{(Y)}})$, a t to liczba przypadków wchodzących w skład rangi wiązanej.

Wzór na współczynnik $\tilde{\tau}$ zawiera poprawkę na rangi wiązane. Poprawka ta powinna być stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych, poprawka nie jest wyliczana, gdyż wówczas $T_X = 0$ i $T_Y = 0$).

Test t do sprawdzania istotności współczynnika korelacji rangowej Spearmana

KIEDY STOSOWAĆ TEN TEST:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

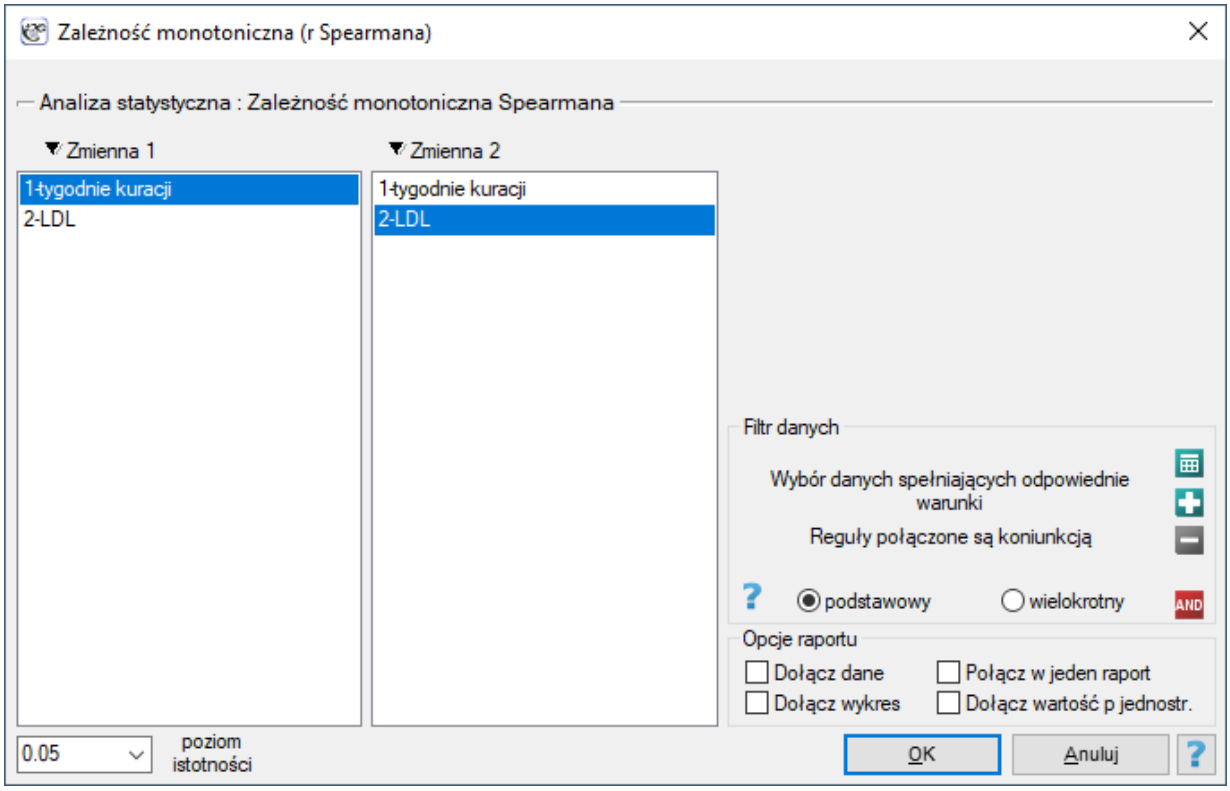
Test t do sprawdzania istotności współczynnika korelacji rangowej Spearmana *(ang. test of significance for Spearman's rank-order correlation coefficient)* pozwala sprawdzić, czy pomiędzy badanymi cechami populacji istnieje zależność monotoniczna. Opiera się na współczynniku korelacji rangowej Spearmana r_s wyliczonym dla próby – im wartość r_s jest bliższa 0, tym słabszą zależnością monotoniczną związane są badane cechy.

WARUNKI STOSOWANIA

– pomiar na skali porządkowej lub interwałowej.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → zależność monotoniczna (r-Spearmana) (lub poprzez Kreator).



WYKONANIE I INTERPRETACJA

Przykład 22.2. (plik LDL tygodnie.pqs)
Badano skuteczność nowej terapii, której celem jest obniżenie poziomu cholesterolu we frakcji LDL. Przebadano 88 osób na różnym etapie kuracji. Sprawdzimy, czy wraz z upływem czasu stosowania kuracji (czas w tygodniach) poziom cholesterolu LDL spada i się stabilizuje.

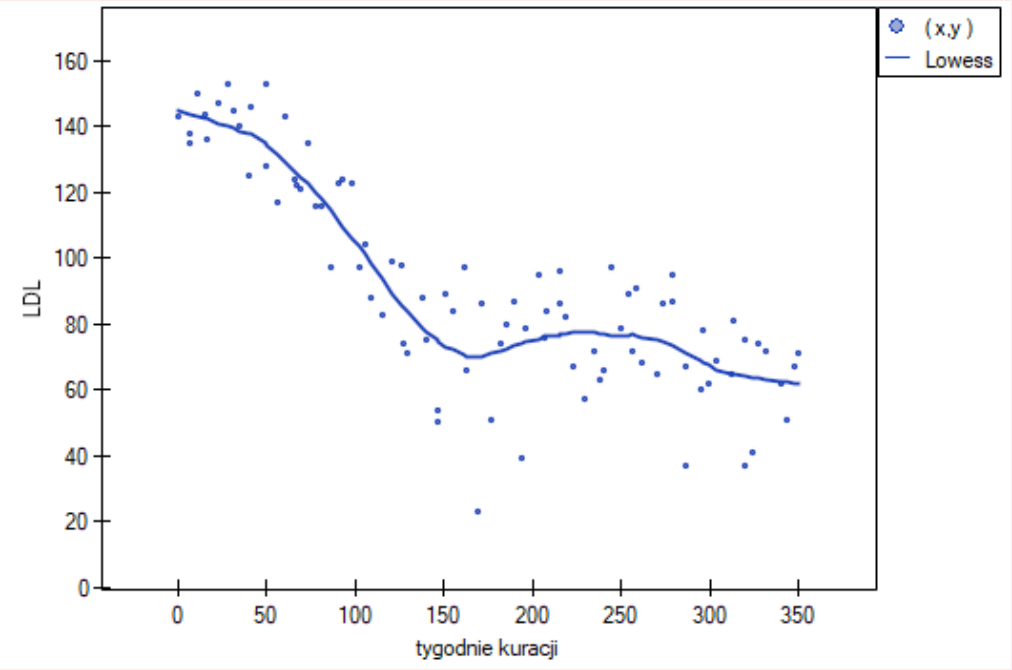
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Zależność monotoniczna Spearmana	
Czas analizy	0.12 sek.
Analizowane zmienne	tygodnie kuracji;LDL
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Liczność - ilość par	88
r	-0.7806
Błąd dla r	0.0674
-95% CI dla współczynnika r	-0.8527
+95% CI dla współczynnika r	-0.6794
Statystyka t dla r	-11.5823
Stopnie swobody	86
Wartość p dwustronne	<0.0001

$p < 0.0001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Istnieje ważna statystycznie monotoniczna zależność pomiędzy czasem kuracji a poziomem LDL.

Zależność ta jest początkowo malejąca, a po 150 tygodniach zaczyna się stabilizować. Współczynnik korelacji monotonicznej Spearmana, a zatem siła związku monotonicznego dla tej zależności jest dość wysoki i wynosi $r_s = -0.7806$. Wykres wyrysowano, dopasowując krzywą poprzez lokalne techniki wygładzania liniowego typu LOWESS.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad R_s = 0, \\ \mathcal{H}_1: &\quad R_s \neq 0. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$t = \frac{r_s}{SE},$$

gdzie $SE = \sqrt{\frac{1 - r_s^2}{n - 2}}$.

Wartość statystyki testowej nie może być wyznaczona, gdy $r_s = 1$ lub $r_s = -1$ albo gdy $n < 3$. Statystyka testowa ma rozkład t-Studenta z $n - 2$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test do sprawdzania istotności współczynnika korelacji $\tilde{\tau}$ Kendalla

KIEDY STOSOWAĆ TEN TEST:

Test do sprawdzania istotności współczynnika korelacji $\tilde{\tau}$ Kendalla *(ang. test of significance for Kendall's tau correlation coefficient)* pozwala sprawdzić, czy pomiędzy badanymi cechami populacji istnieje zależność monotoniczna. Opiera się na współczynniku korelacji Kendalla wyliczonym dla próby – im wartość $\tilde{\tau}$ jest bliższa 0, tym słabszą zależnością monotoniczną związane są badane cechy.

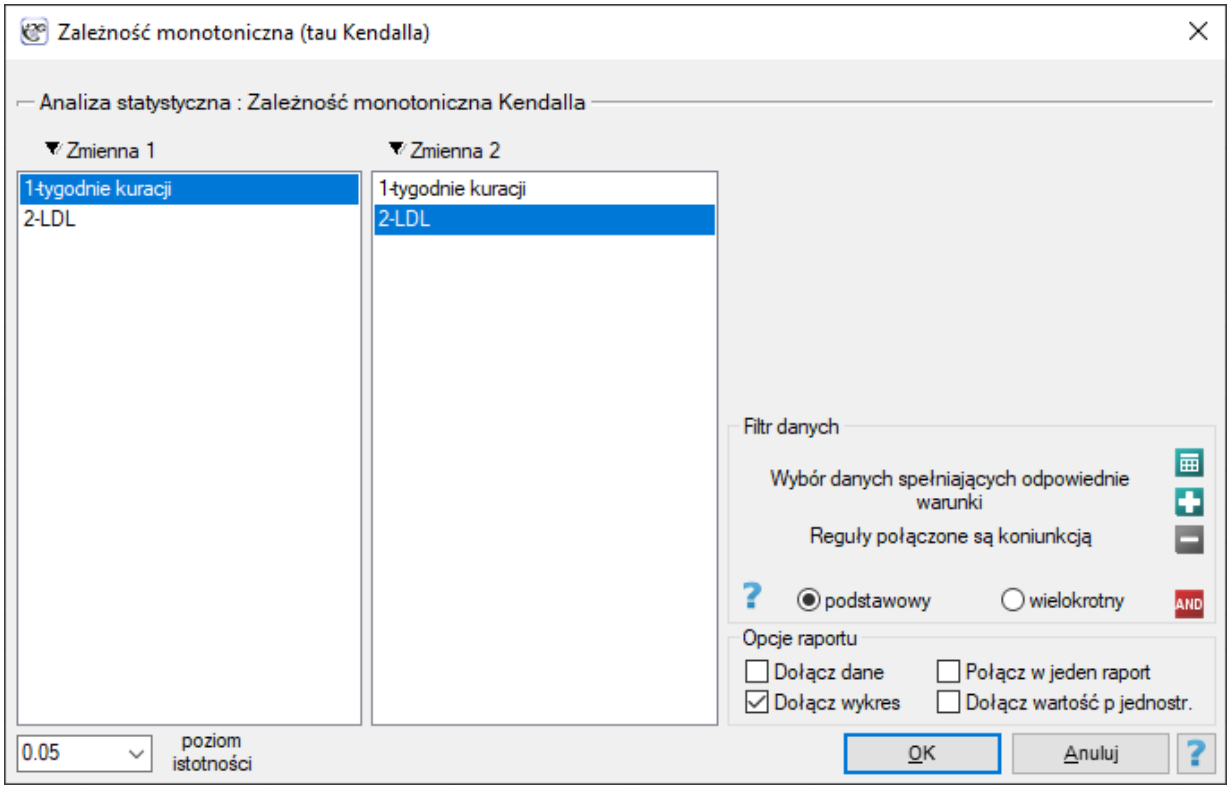
WARUNKI STOSOWANIA

- pomiar na skali porządkowej lub interwałowej.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → zależność monotoniczna (tau-Kendalla) (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 22.2 (plik LDL tygodnie.pqs): sprawdzimy tę samą zależność między czasem kuracji a poziomem LDL, tym razem współczynnikiem tau Kendalla.

Zależność monotoniczna Kendalla	
Czas analizy	0.14 sek.
Analizowane zmienne	tygodnie kuracji;LDL
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Liczność - ilość par	88
tau	-0.5975
Statystyka Z dla tau	-8.2435
Wartość p dwustronne	<0.0001

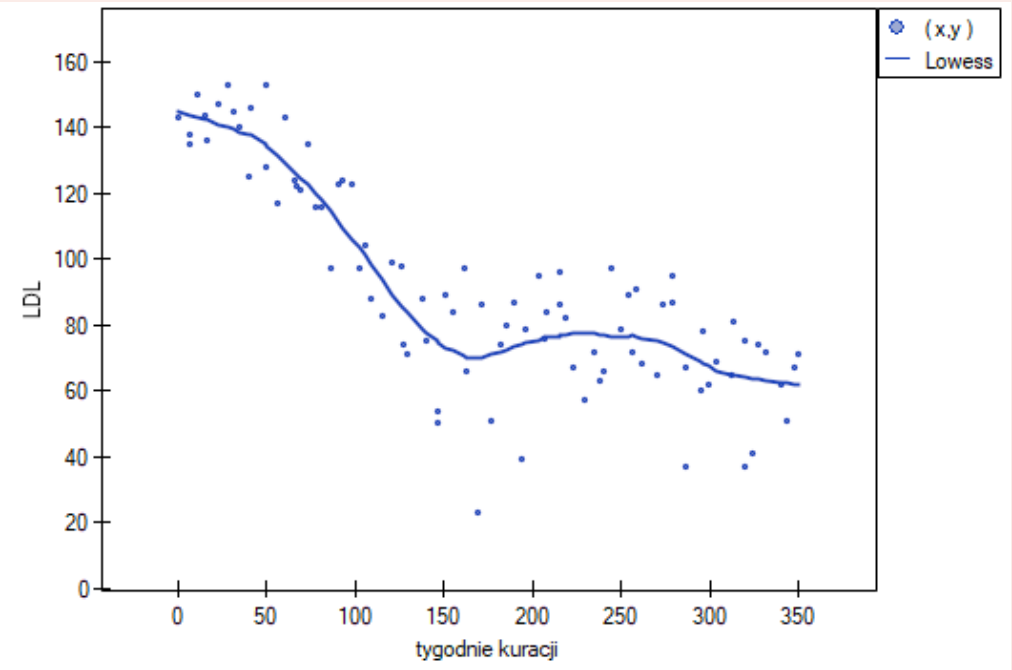
$p < 0.0001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Istnieje ważna statystycznie monotoniczna zależność pomiędzy czasem kuracji a poziomem LDL.

Zależność ta jest początkowo malejąca, a po 150 tygodniach zaczyna się stabilizować. Współczynnik korelacji monotonicznej Kendalla, a zatem siła związku monotonicznego dla tej zależności jest dość wysoki

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

i wynosi $\tilde{\tau} = -0.5975$. Wykres wyrysowano, dopasowując krzywą poprzez lokalne techniki wygładzania liniowego typu LOWESS.



DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \tau = 0, \\ \mathcal{H}_1: &\quad \tau \neq 0. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{3\tilde{\tau}\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}.$$

Statystyka testowa ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Współczynniki tabel kontyngencji i ich istotność statystyczna

KIEDY STOSOWAĆ:

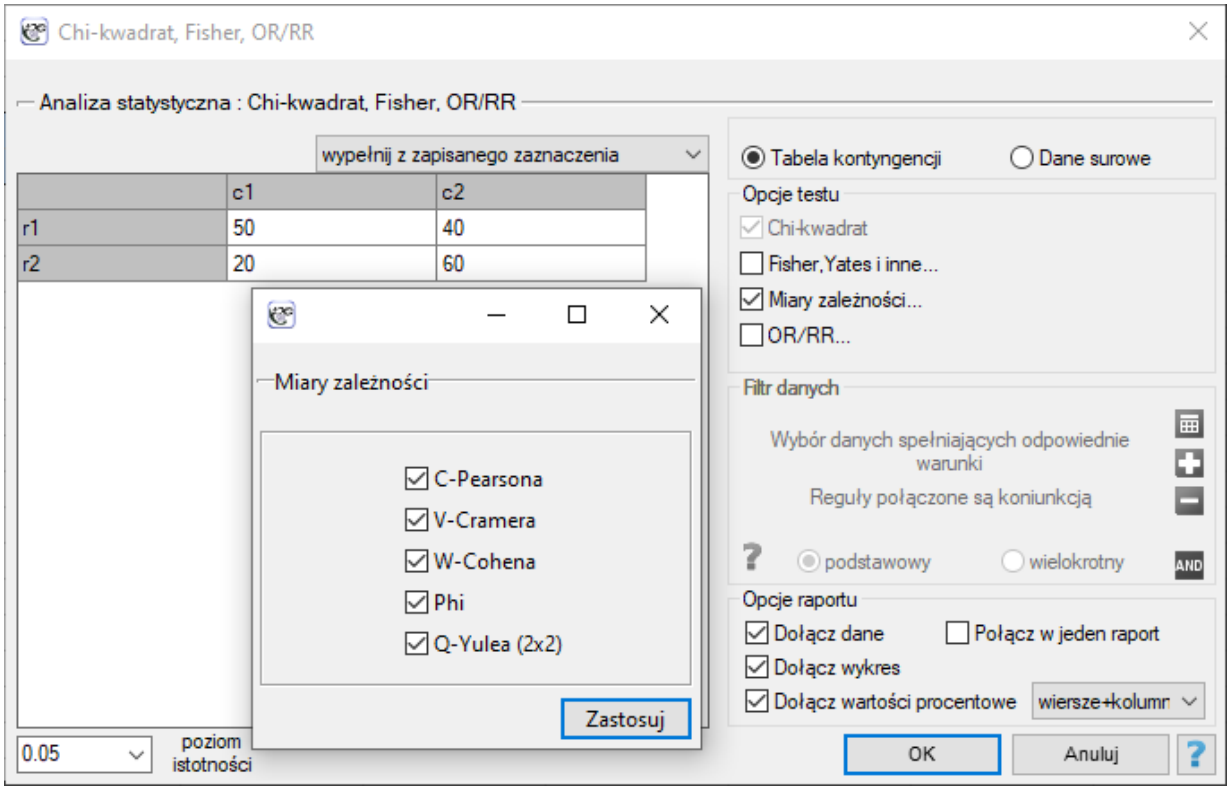
Współczynniki kontyngencji pozwalają ocenić siłę zależności pomiędzy dwiema cechami zapisanymi na skali nominalnej. Wyliczane są dla danych w postaci surowej lub danych zebranych w tabelę kontyngen-

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

cji.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR → Współczynniki korelacji... (lub poprzez Kreator).



DEFINICJE MIAR

Współczynnik kontyngencji Q-Yulea *(ang. Yule's Q contingency coefficient, Yule (1900)[186])* – miara zależności wyznaczana dla tabel kontyngencji 2×2. Wartość Q mieści się w przedziale $\langle -1; 1 \rangle$: im bliższa 0, tym siła związku jest mniejsza, a im bliższa -1 lub $+1$, tym większa (ze względu na możliwe błędy w interpretacji wartości ujemnej, PQStat przedstawia wynik dodatkowo jako wartość bezwzględną). Współczynnik jest mało odporny na małe liczności obserwowane – gdy jedna z nich wynosi 0, może błędnie wskazywać całkowitą zależność cech.

Współczynnik kontyngencji ϕ *(ang. phi contingency coefficient)* – miara zależności polecana szczególnie dla tabel 2×2, choć możliwa do wyznaczenia dla dowolnych tabel. Wartość ϕ mieści się w przedziale $\langle 0; 1 \rangle$: im bliższa 0, tym siła związku jest mniejsza, a im bliższa 1, tym większa.

Współczynnik kontyngencji V-Cramera *(ang. Cramer's V contingency coefficient, Cramer (1946)[49])* – rozszerzenie współczynnika ϕ na tabele $r \times c$. Wartość V mieści się w przedziale $\langle 0; 1 \rangle$: im bliższa 0, tym siła związku jest mniejsza, a im bliższa 1, tym większa. Wartość V zależy od wielkości tabeli, więc nie powinno się go stosować do porównywania tabel o różnych wielkościach.

Współczynnik kontyngencji W-Cohena *(ang. Cohen's w contingency coefficient, Cohen (1988)[46])* – modyfikacja współczynnika V-Cramera, możliwa do wyliczenia dla tabel $r \times c$. Wartość W mieści się w

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

przedziale $\langle 0; \max W \rangle$ (dla tabel, w których co najmniej jedna zmienna ma tylko dwie kategorie, przedział ten to $\langle 0; 1 \rangle$). Podobnie jak V, zależy od wielkości tabeli.

Współczynnik kontyngencji C-Pearsona *(ang. Pearson's C contingency coefficient)* – miara zależności dla tabel $r \times c$. Wartość C mieści się w przedziale $\langle 0; 1 \rangle$, a jej górna granica zależy od wielkości tabeli, dlatego wyznacza się także wersję dostosowaną do wielkości tabeli C_{adj}, umożliwiającą porównania między tabelami różnej wielkości.

Uwaga! Współczynniki ϕ, V-Cramera, W-Cohena i C-Pearsona uznaje się za istotne statystycznie, jeśli wartość p wyznaczona na podstawie statystyki testu χ^2 i rozkładu χ^2 (wyznaczonego dla danej tabeli) jest równa bądź mniejsza niż poziom istotności α. Istotność współczynnika Q-Yulea sprawdza się natomiast osobnym testem Z (zob. *Dla zainteresowanych*).

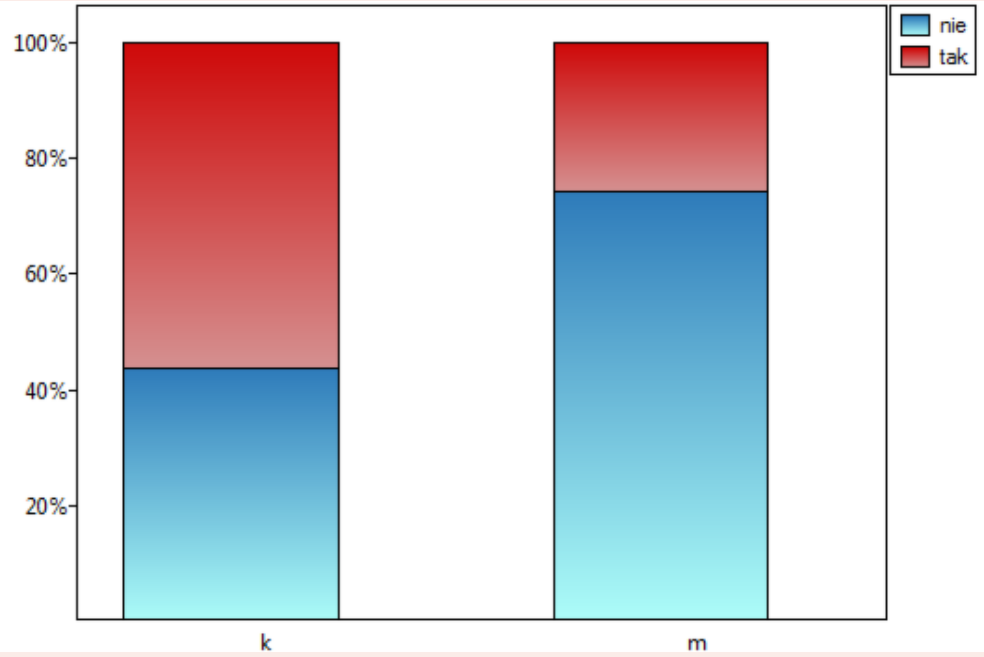
WYKONANIE I INTERPRETACJA

Przykład 22.3. (plik płeć-egzamin.pqs)
Rozpatrzmy próbę składającą się z 170 osób ($n = 170$), dla których badamy 2 cechy (X=płeć, Y=zdawalność egzaminu). Każda z tych cech występuje w dwóch kategoriach (X_1=k, X_2=m, Y_1=tak, Y_2=nie). Na podstawie tej próby chcielibyśmy się dowiedzieć, czy w badanej populacji istnieje zależność pomiędzy płcią a zdawalnością egzaminu. Rozkład danych przedstawia tabela kontyngencji:

Liczności obserwowane O_{ij}		zdawalność egzaminu		
		tak	nie	suma
płeć	k	50	40	90
	m	20	60	80
	suma	70	100	170

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Chi-kwadrat, Fisher, OR/RR	
Czas analizy	0.10 sek.
Analizowane zmienne	Tabela kontyngencji
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Liczność	170
Warunek Cochrana	spełniony
Statystyka chi-kwadrat Pearsona	16.3254
Stopnie swobody	1
Wartość p	0.0001
Miary zależności	
C-Pearsona	0.296
C-Pearsona (max)	0.7071
C-Pearsona (dostosowane)	0.4186
V-Cramera	0.3099
W-Cohen	0.3099
W-Cohen (max)	1
Phi	0.3099
Q-Yulea	0.5789
Statystyka Z Yulea	5.212
Wartość p (asymptotyczne) Yulea	<0.0001



$\chi^2 = 16.33$, $p = 0.00005$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Istnieje zależność statystyczna pomiędzy płcią a zdawalnością egzaminu w badanej populacji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wartość współczynników opartych o test χ^2, a zatem siła związku między badanymi cechami, wynosi:

Współczynnik kontyngencji C_{adj}-Pearsona = 0.42.
Współczynnik kontyngencji V-Cramera = ϕ = W-Cohena = 0.31.

Współczynnik kontyngencji Q-Yulea = 0.58, a wartość p wykonanego testu Z, podobnie jak poziom istotności testu χ^2, wskazuje na istotność statystyczną badanego związku.

DLA ZAINTERESOWANYCH

Wzór – współczynnik kontyngencji Q-Yulea:

$$Q = \frac{O_{11}O_{22} - O_{12}O_{21}}{O_{11}O_{22} + O_{12}O_{21}},$$

gdzie $O_{11}, O_{12}, O_{21}, O_{22}$ – liczności obserwowane w tabeli kontyngencji.

Istotność statystyczną wyznaczonego współczynnika Q-Yulea określa się testem Z.
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad Q = 0, \\ \mathcal{H}_1: &\quad Q \neq 0. \end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{Q}{\sqrt{\frac{1}{4}(1-Q^2)^2(\frac{1}{O_{11}} + \frac{1}{O_{12}} + \frac{1}{O_{21}} + \frac{1}{O_{22}})}}.$$

Statystyka testowa ma asymptotycznie (dla dużych liczności n) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Wzór – współczynnik kontyngencji ϕ:

$$\phi = \sqrt{\frac{\chi^2}{n}},$$

gdzie χ^2 – wartość statystyki testu χ^2, n – całkowita liczność w tabeli kontyngencji.

Wzór – współczynnik kontyngencji V-Cramera:

$$V = \sqrt{\frac{\chi^2}{n(w'-1)}},$$

gdzie χ^2 – wartość statystyki testu χ^2, n – całkowita liczność w tabeli, a w' – mniejsza z dwóch wartości r i c.

Wzór – współczynnik kontyngencji W-Cohena:

$$W = \sqrt{\frac{\chi^2}{n(w'-1)}}\sqrt{w'-1},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie χ^2 – wartość statystyki testu χ^2, n – całkowita liczność w tabeli, a w' – mniejsza z dwóch wartości r i c.

Wzór – współczynnik kontyngencji C-Pearsona:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

gdzie χ^2 – wartość statystyki testu χ^2, n – całkowita liczność w tabeli. Górną granicę, jaką dla danej wielkości tabeli współczynnik C może osiągnąć, wyznacza wzór:

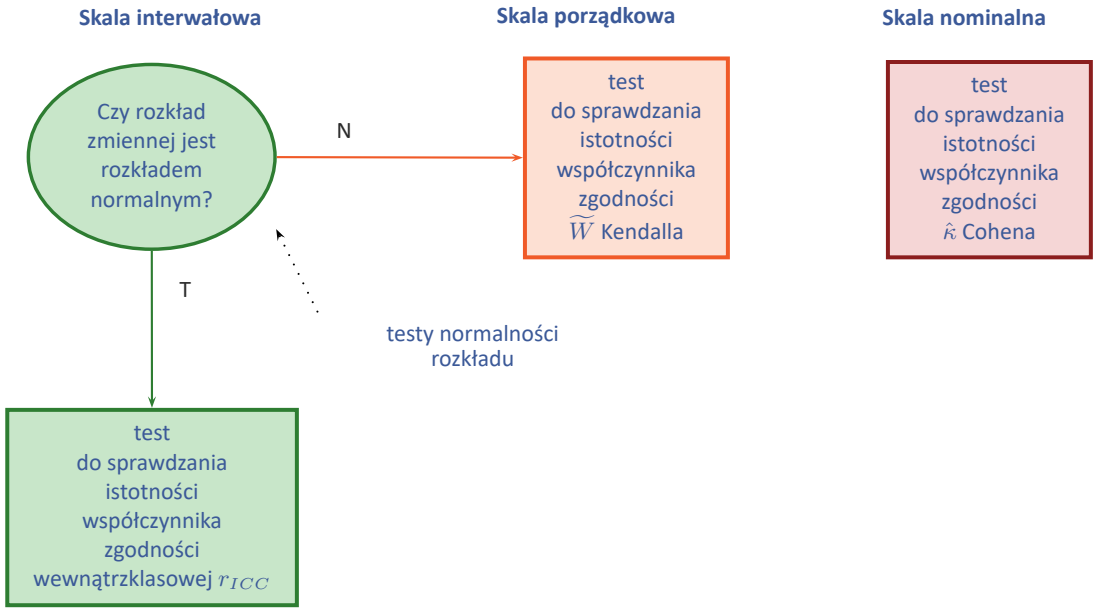
$$C_{max} = \sqrt{\frac{w' - 1}{w'}},$$

gdzie w' – mniejsza z dwóch wartości r i c. Wersja dostosowana do wielkości tabeli:

$$C_{adj} = \frac{C}{C_{max}}.$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

23 ANALIZA ZGODNOŚCI



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

23.1 TESTY PARAMETRYCZNE

Współczynnik korelacji wewnątrzklasowej i test badający jego istotność

KIEDY STOSOWAĆ:

Współczynnik korelacji wewnątrzklasowej *(ang. ICC – intraclass correlation coefficient)* stosuje się w sytuacji, gdy pomiarów badanej zmiennej dokonuje kilku „sędziów" ($k \geq 2$). Mierzy on siłę **sędziowskiej rzetelności**, czyli stopień, w jakim ich oceny są zgodne. Ponieważ można go wyznaczać w kilku różnych sytuacjach, rozróżnia się kilka jego odmian, zależnych od modelu badania i typu zgodności.

WARUNKI STOSOWANIA

- pomiar na skali interwałowej,
- normalność rozkładu dla wszystkich zmiennych będących różnicą par pomiarowych (lub normalność badanej zmiennej dla każdego pomiaru),
- dla modelu 1 – model niezależny, dla modelu 2 / 3 – model zależny.

DEFINICJE MIAR

W zależności od występującej w danych zmienności rozróżnia się 2 główne modele badań i 2 typy zgodności.

Model 1 – dla każdego z losowo wybranych n ocenianych obiektów wybierany jest losowo zestaw k sędziów z populacji sędziów, przy czym dla każdego obiektu może być wylosowany inny zestaw sędziów. Współczynnik ICC jest wówczas wyznaczany w modelu losowym ANOVA dla grup niezależnych. ICC(1,1) odpowiada na pytanie o rzetelność oceny pojedynczego sędziego, a ICC(1,k) – o rzetelność wyniku będącego średnią ocen k sędziów.

Model 2 – wybierany jest losowo zestaw k sędziów z populacji sędziów i każdy z nich ocenia wszystkie n losowych obiektów. Współczynnik ICC jest wówczas wyznaczany w modelu losowym ANOVA dla grup zależnych. W zależności od poszukiwanego typu zgodności można szacować:

- **zgodność bezwzględną** – sędziowie są zgodni bezwzględnie, jeśli wydają dokładnie takie same oceny, np. perfekcyjnie zgodne będą oceny wydane przez parę sędziów (2,2), (5,5), (8,8);
- **spójność** – sędziowie mogą używać innych zakresów wartości, ale poza tym przesunięciem nie powinno być różnic, by zachować spójność werdyktu, np. perfekcyjnie spójne będą oceny wydane przez parę sędziów (2,5), (5,8), (8,11).

Podobnie jak w modelu 1, ICC(2,1) odpowiada na pytanie o rzetelność oceny pojedynczego sędziego, a ICC(2,k) – o rzetelność wyniku będącego średnią ocen k sędziów (wzory różnią się w zależności od wybranego typu zgodności).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

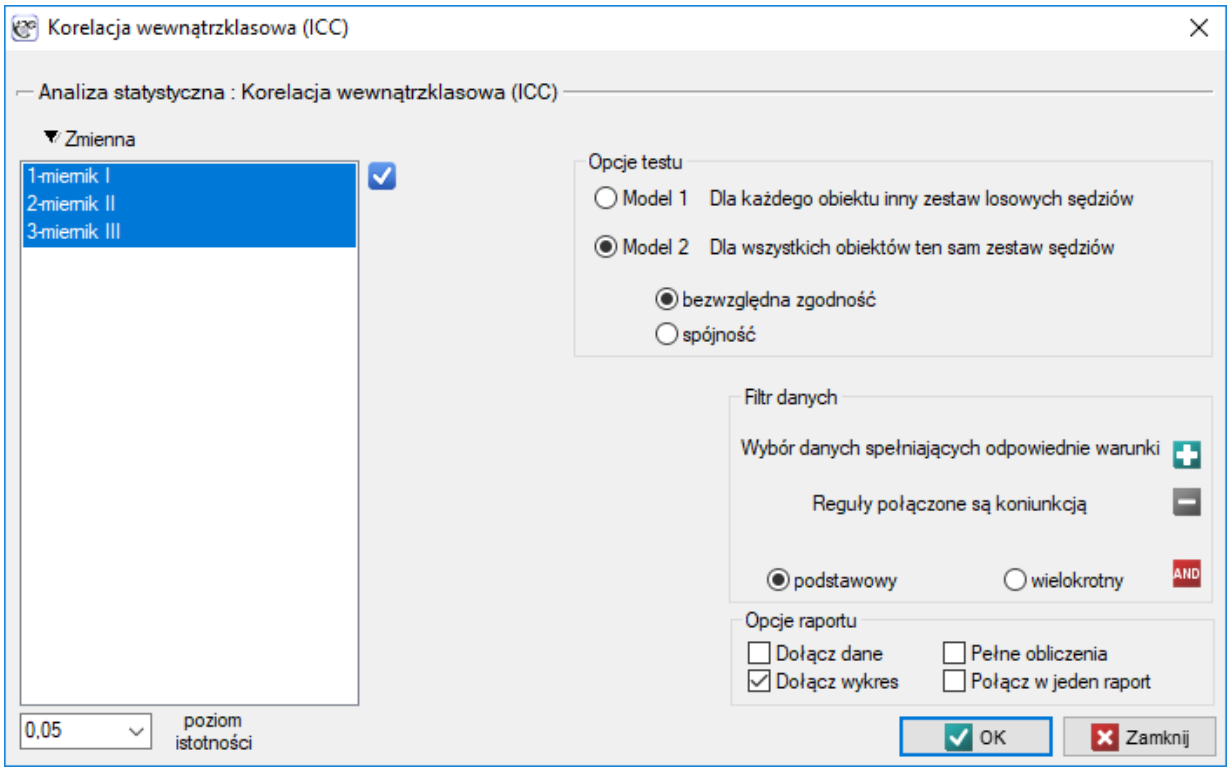
Uwaga! Czasami istnieje konieczność rozważenia **modelu 3**[154], tzn. wybierany jest zestaw k sędziów i każdy z nich ocenia wszystkie n losowych obiektów, przy czym wynik zgodności dotyczy tylko tych konkretnych k sędziów. Współczynnik ICC jest wówczas wyznaczany w modelu mieszanym (losowość dotyczy tylko obiektów, a nie sędziów). Ponieważ ignorujemy zmienność dotyczącą sędziów, badamy spójność (a nie absolutną zgodność) – zastosowanie mają wtedy współczynniki z modelu drugiego: ICC(2,1) i ICC(2,k), gdyż są one tożsame z pożądanymi w tym przypadku współczynnikami ICC(3,1) i ICC(3,k) przy założeniu braku interakcji obiektów i sędziów.

Interpretacja – wartość $ICC \in \langle -1; 1 \rangle$ interpretujemy następująco:

- $ICC \approx 1$ oznacza silną zgodność w ocenie poszczególnych obiektów przez sędziów, co ma odzwierciedlenie w dużej wariancji między obiektami (znacznej różnicy średnich między n obiektami) i małej wariancji między ocenami sędziowskimi (niewielkiej różnicy średnich ocen wyznaczonych dla k sędziów);
- $ICC \approx -1$ – ujemny współczynnik korelacji wewnątrzklasowej jest traktowany w ten sam sposób jak $ICC \approx 0$;
- $ICC \approx 0$ oznacza brak zgodności w ocenie poszczególnych obiektów przez sędziów, co ma odzwierciedlenie w małej wariancji między obiektami (niewielkiej różnicy średnich między n obiektami) i dużej wariancji między ocenami sędziowskimi (znaczącej różnicy średnich ocen wyznaczonych dla k sędziów).

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy parametryczne → Korelacja wewnątrzklasowa (ICC) (lub poprzez Kreator).



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

WYKONANIE I INTERPRETACJA

PRZYKŁAD 23.1. (plik natężenie dźwięku.pqs)
By skutecznie dbać o słuch pracowników zakładu pracy, w pierwszej kolejności należy rzetelnie oszacować natężenie dźwięku w poszczególnych miejscach przebywania osób. W pewnym zakładzie pracy postanowiono przed wyborem miernika natężenia dźwięku (sonografu) przeprowadzić eksperyment. Pomiarów natężenia dźwięku dokonano w 42 losowo wybranych na terenie zakładu punktach pomiarowych przy pomocy 3 wylosowanych sonografów analogowych i 3 losowo wybranych sonografów cyfrowych. Fragment zebranych pomiarów przedstawia poniższa tabela.

miernik I	miernik II	miernik III
84.1	84.1	84.7
85.7	85.3	85.3
84.5	83.5	84.3
86.2	86.1	87.1
85.6	84.8	85.5
80.5	80.6	81
81.4	81.7	82.1
85.6	86	87.3
83.2	82.4	80.1
83.5	82.5	85.5
82.4	82.5	81.5

By sprawdzić, który typ urządzenia (analogowy czy cyfrowy) lepiej zrealizuje postawione przed nim zadanie, należy wyznaczyć współczynnik ICC w modelu 2, badając bezwzględną zgodność. Typ miernika o wyższym wskaźniku ICC będzie charakteryzował się bardziej wiarygodnymi pomiarami, przez co będzie w przyszłości wykorzystywany na terenie zakładu.

Analiza przeprowadzona dla mierników analogowych wskazuje na istotną zgodność pomiarów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Korelacja wewnątrzklasowa (ICC)	
Czas analizy	0.06 sek.
Analizowane zmienne	miernik I
	miernik II
	miernik III
Poziom istotności	0.05
Model 2 Dla wszystkich obiektów ten sam zestaw sędziów	bezwzględna zgodność
Stopnie swobody [obiekty]	41
Stopnie swobody [res]	82
Średnia kwadratów [pomiędzy sędziami]	105.151667
Średnia kwadratów [pomiędzy obiektami]	8.902625
Średnia kwadratów [res]	0.796626
ICC(2,1) dla pojedynczego sędziego	0.451593
-95% CI dla ICC(2,1)	0.030782
+95% CI dla ICC(2,1)	0.730383
ICC(2,k) dla średniej k sędziów	0.711848
-95% CI dla ICC(2,k)	0.08699
+95% CI dla ICC(2,k)	0.890434
Statystyka F	11.175413
Wartość p	<0.000001

$p < 0.000001$ **dla mierników analogowych (istotne statystycznie,** $\alpha = 0.05$**)**

Rzetelność pomiaru dokonana miernikiem analogowym wynosi $ICC(2,1) = 0.45$, a rzetelność pomiaru będącego średnią pomiarów trzech mierników analogowych wynosi $ICC(2,k) = 0.71$.

Niepokojąco niska jest jednak dolna granica 95% przedziału ufności dla tych współczynników. Podobna analiza przeprowadzona dla mierników cyfrowych dała lepsze rezultaty.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

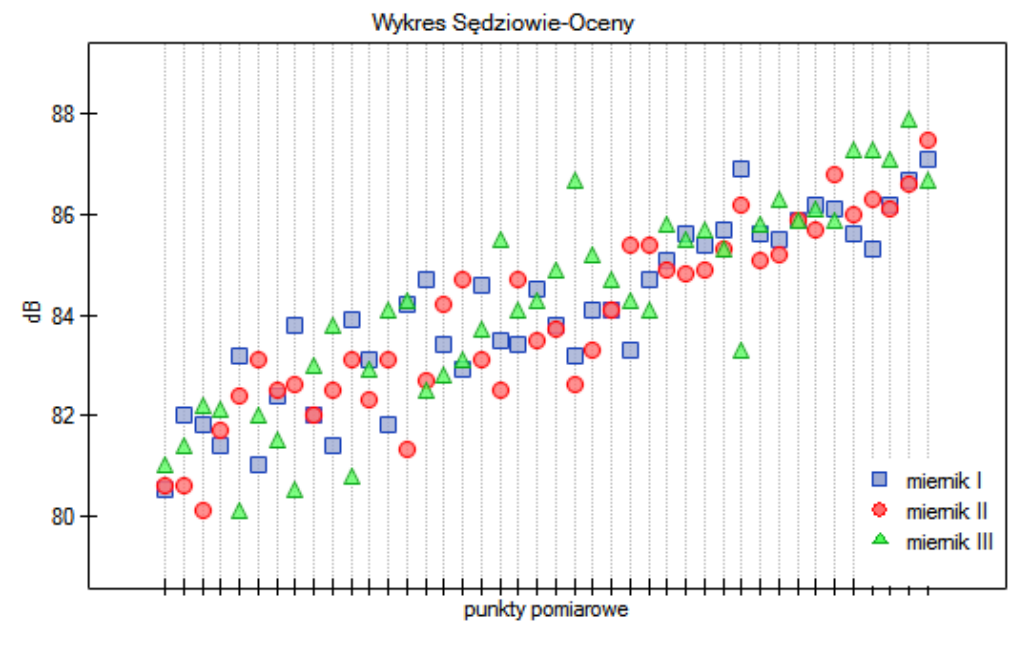
Korelacja wewnątrzklasowa (ICC)	
Czas analizy	0.14 sek.
Analizowane zmienne	miernik I
	miernik II
	miernik III
Poziom istotności	0.05
Model 2 Dla wszystkich obiektów ten sam zestaw sędziów	bezwzględna zgodność
Stopnie swobody [obiekty]	41
Stopnie swobody [res]	82
Średnia kwadratów [pomiędzy sędziami]	0.915952
Średnia kwadratów [pomiędzy obiektami]	8.700186
Średnia kwadratów [res]	0.947009
ICC(2,1) dla pojedynczego sędziego	0.731985
-95% CI dla ICC(2,1)	0.601269
+95% CI dla ICC(2,1)	0.834036
ICC(2,k) dla średniej k sędziów	0.891226
-95% CI dla ICC(2,k)	0.818968
+95% CI dla ICC(2,k)	0.937796
Statystyka F	9.187012
Wartość p	<0.000001

$p < 0.000001$ **dla mierników cyfrowych (istotne statystycznie,** $\alpha = 0.05$**)**

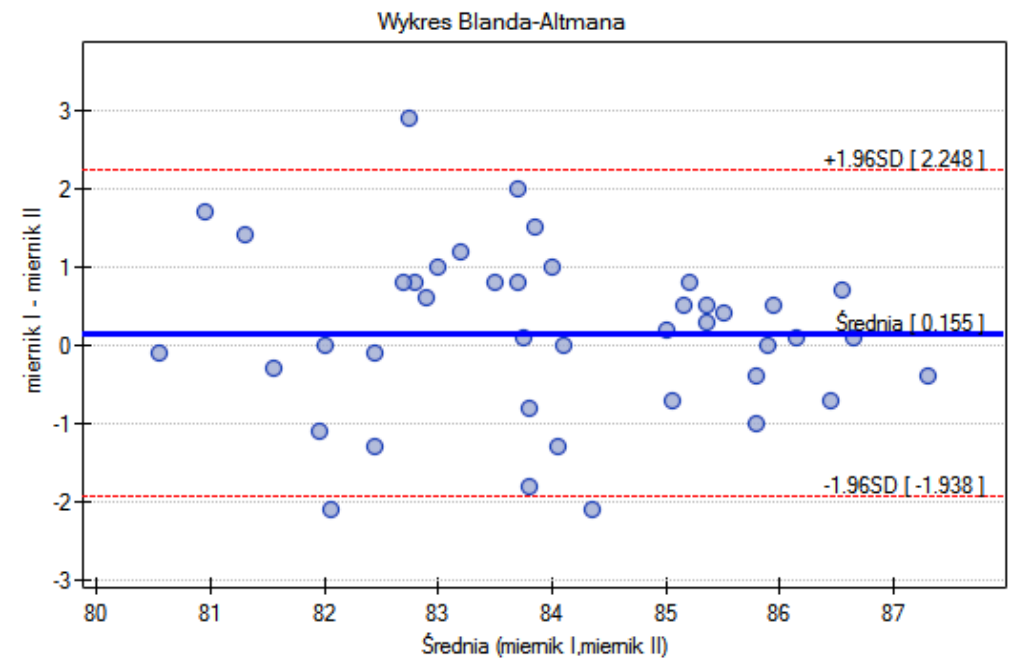
Współczynniki ICC oraz ich przedziały ufności są znacznie wyższe niż dla mierników analogowych: $ICC(2,1) = 0.73$, $ICC(2,k) = 0.89$.

Dlatego ostatecznie w zakładzie pracy wykorzystywane będą mierniki cyfrowe. Zgodność uzyskanych wyników dla mierników cyfrowych przedstawiono na wykresie punktowym, gdzie każdy punkt pomiarowy opisany jest wartością natężenia dźwięku uzyskaną dla poszczególnych mierników.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Przedstawiając wykres dla uprzednio posortowanych danych zgodnie ze średnią wartością natężenia dźwięku, można sprawdzić, czy stopień zgodności rośnie lub spada wraz ze wzrostem natężenia dźwięku. W przypadku naszych danych nieco wyższą zgodność (bliskość położenia punktów na wykresie) obserwować można przy wysokich wartościach natężenia dźwięku. Podobnie, zgodność uzyskanych wyników obserwować można na wykresach Blanda-Altmana[8][23] budowanych oddzielnie dla każdej pary mierników. Wykres dla miernika I i miernika II przedstawiono poniżej.



Tu również obserwujemy wyższą zgodność (punkty koncentrują się blisko osi poziomej $y = 0$) dla wyższych wartości natężenia dźwięku.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uwaga! Gdyby badaczowi nie zależało na oszacowaniu rzeczywistej wartości natężenia dźwięku na terenie zakładu, ale chciałby wskazać miejsca, gdzie ten poziom jest wyższy niż w innych miejscach, lub sprawdzić, czy poziom hałasu zmienia się w czasie, wówczas wystarczającym modelem byłby model 2 badający spójność.

DLA ZAINTERESOWANYCH

Model 1 – wzory:
ICC(1,1):

$$ICC(1,1) = \frac{MS_{WG} - MS_{BG}}{MS_{WG} + (k-1)MS_{BG}}.$$

ICC(1,k):

$$ICC(1,k) = \frac{MS_{WG} - MS_{BG}}{MS_{WG}},$$

gdzie:
MS_{WG} – średnia kwadratów wewnątrz grup,
MS_{BG} – średnia kwadratów między obiektami.

Model 2 – zgodność bezwzględna – wzory:
ICC(2,1):

$$ICC(2,1) = \frac{MS_{BS} - MS_{res}}{MS_{BS} + (k-1)MS_{res} + \frac{k}{n}(MS_{BC} - MS_{res})}.$$

ICC(2,k):

$$ICC(2,k) = \frac{MS_{BS} - MS_{res}}{MS_{BS} + (MS_{BC} - MS_{res})/n},$$

gdzie:
MS_{BC} – średnia kwadratów między sędziami,
MS_{BS} – średnia kwadratów między obiektami,
MS_{res} – średnia kwadratów dla reszt.

Model 2 – spójność – wzory:
ICC(2,1):

$$ICC(2,1) = \frac{MS_{BS} - MS_{res}}{MS_{BS} + (k-1)MS_{res}},$$

ICC(2,k):

$$ICC(2,k) = \frac{MS_{BS} - MS_{res}}{MS_{BS}},$$

gdzie:
MS_{BS} – średnia kwadratów między obiektami,
MS_{res} – średnia kwadratów dla reszt.

Test F do sprawdzania istotności współczynnika korelacji wewnątrzklasowej
Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad ICC = 0 \\ \mathcal{H}_1 &: \quad ICC \neq 0 \quad (ICC = 1) \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$F = \frac{MS_{BS}}{MS_{res}}$$ – w modelu zależnym,

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

lub

$$F = \frac{MS_{WG}}{MS_{BG}}$$ – w modelu niezależnym.

Statystyka ta podlega rozkładowi F Snedecora ze zdefiniowaną w modelu liczbą stopni swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

23.2 TESTY NIEPARAMETRYCZNE

Współczynnik zgodności Kendalla i test badający jego istotność

KIEDY STOSOWAĆ:

Współczynnik zgodności $\widetilde{W}$ Kendalla *(ang. Kendall's Coefficient of Concordance)*, opisany w pracy Kendalla i Babingtona-Smitha (1939)[95] oraz Wallisa (1939)[170], stosuje się w sytuacji, gdy dysponujemy rankingami pochodzącymi z różnych źródeł (od różnych sędziów) i dotyczącymi kilku ($k \geq 2$) obiektów, a zależy nam na ocenie zgodności tych rankingów. Często używa się go do mierzenia siły **sędziowskiej rzetelności**, czyli stopnia, w jakim oceny sędziów są zgodne.

WARUNKI STOSOWANIA

– pomiar na skali porządkowej lub interwałowej.

DEFINICJE MIAR

Współczynnik zgodności $\widetilde{W}$ Kendalla – wartość $W \in \langle 0; 1 \rangle$ interpretujemy następująco:

- $\widetilde{W} \approx 1$ oznacza silną zgodność w ocenie poszczególnych obiektów przez sędziów;
- $\widetilde{W} \approx 0$ oznacza brak zgodności w ocenie poszczególnych obiektów przez sędziów.

W oznacza współczynnik zgodności Kendalla w populacji, natomiast $\widetilde{W}$ w próbie.

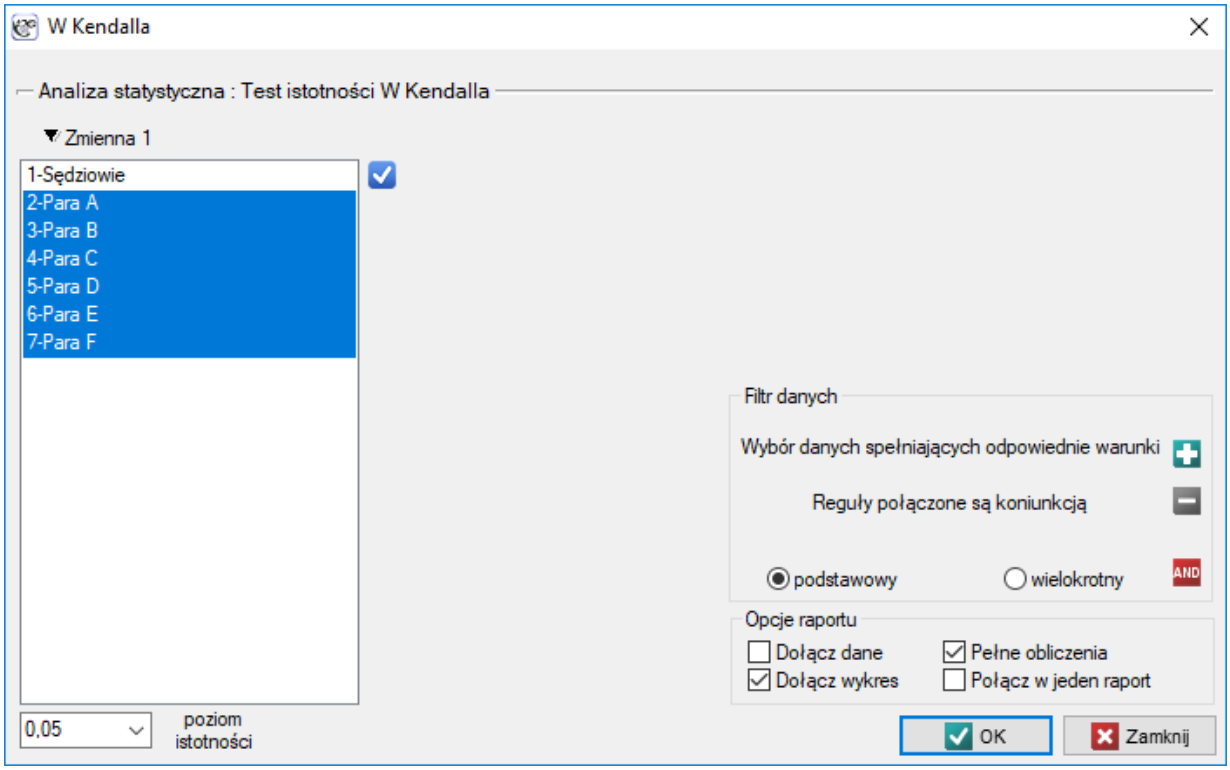
Uwaga! Współczynnik zgodności $\widetilde{W}$ Kendalla a współczynnik r_s Spearmana: gdy wyliczymy wartość współczynnika korelacji Spearmana r_s dla wszystkich możliwych par rankingów, to średni współczynnik r_s (oznaczony przez $\bar{r}_s$) jest funkcją liniową wartości współczynnika $\widetilde{W}$ wyliczonego na podstawie tych samych danych (zob. *Dla zainteresowanych*).

Współczynnik zgodności $\widetilde{W}$ Kendalla a ANOVA Friedmana: współczynnik zgodności $\widetilde{W}$ Kendalla i ANOVA Friedmana bazują na tym samym modelu matematycznym. W rezultacie wartość statystyki testowej testu χ^2 do sprawdzania istotności współczynnika zgodności Kendalla i wartość statystyki testowej ANOVA Friedmana są takie same.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → W Kendalla (lub poprzez Kreator).



WYKONANIE I INTERPRETACJA

Przykład 23.2. (plik sędziowie.pqs)
W systemie szóstkowym oceny par tanecznych 9 sędziów punktuje m.in. wrażenie artystyczne. Sędziowie rozpoczynają wystawianie oceny od porównania zawodników względem siebie i ustawienia ich na określonym miejscu (tworzą ich ranking). Sprawdzimy, czy rangi przypisane przez sędziów są zgodne:

Sędziowie	Para A	Para B	Para C	Para D	Para E	Para F
S1	3	6	2	5	4	1
S2	4	6	1	5	3	2
S3	4	6	2	5	3	1
S4	2	6	3	5	4	1
S5	2	6	1	5	4	3
S6	3	5	1	6	4	2
S7	5	4	1	6	3	2
S8	3	6	2	5	4	1
S9	2	6	3	5	4	1

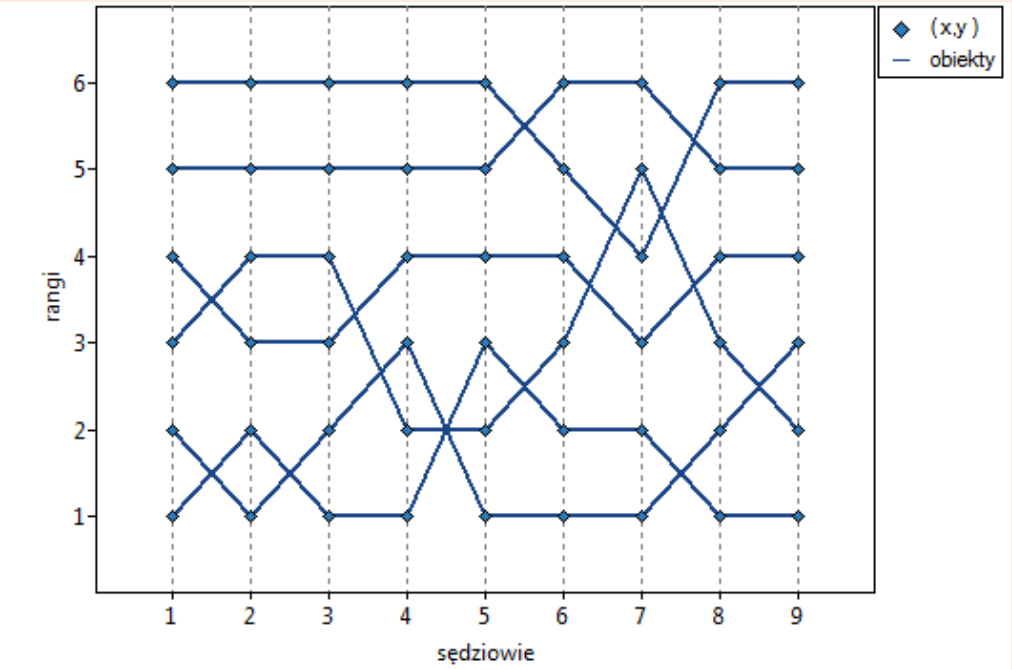
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test istotności W Kendalla	
Czas analizy	0.04 sek.
Analizowane zmienne	Para A,Para B,Para C,Para D,Para E,Par
Poziom istotności	0.05
Ilość stopni swobody	5
Współczynnik zgodności Kendalla	0.83351
Średni współczynnik korelacji Spearmana	0.812698
Statystyka Chi2 (z poprawką dla rang wiązanych)	37.507937
Wartość p	<0.000001

$p < 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Oceny sędziów są statystycznie zgodne.

Siła tej zgodności jest wysoka i wynosi $\widetilde{W} = 0.83351$, podobnie jak średni współczynnik korelacji monotonicznej Spearmana $\bar{r}_s = 0.81270$. Wynik ten możemy przedstawić na wykresie, na którym oś X reprezentuje kolejnych sędziów. Wówczas im częściej przecinają się linie (które powinny być równoległe do osi X, gdy zgodność jest pełna), tym słabszą zgodność reprezentują oceny sędziów.



DLA ZAINTERESOWANYCH

Wzory:

Wartość współczynnika zgodności Kendalla wylicza się według wzoru:

$$\widetilde{W} = \frac{12U - 3n^2k(k+1)^2}{n^2k(k^2-1) - nC},$$

gdzie:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

n – liczba różnych zbiorów ocen (liczba sędziów),

k – liczba rangowanych obiektów,

$U = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} R_{ij} \right)^2$,

R_{ij} – rangi przypisane kolejnym obiektom $(j = 1, 2, ...k)$, oddzielnie dla każdego z sędziów $(i = 1, 2, ...n)$,

$C = \sum (t^3 - t)$ – korekta na rangi wiązane,

t – liczba przypadków wchodzących w skład rangi wiązanej.

Wzór na współczynnik zawiera poprawkę na rangi wiązane C. Poprawka ta jest stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych, poprawka ta nie jest wyliczana, gdyż wówczas $C = 0$).

Zależność pomiędzy średnim współczynnikiem korelacji Spearmana a współczynnikiem zgodności Kendalla wyraża się wzorem:

$$\bar{r}_s = \frac{n\widetilde{W} - 1}{n - 1}.$$

Test χ^2 do sprawdzania istotności współczynnika zgodności Kendalla
Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad W = 0 \\ \mathcal{H}_1 &: \quad W \neq 0 \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$\chi^2 = n(k-1)\widetilde{W}.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody wyliczaną z wzoru $df = k - 1$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Współczynnik Kappa Cohena i test badający jego istotność

KIEDY STOSOWAĆ:

Współczynnik Kappa Cohena *(ang. Cohen's Kappa)* (Cohen J., 1960[44]) określa stopień zgodności dwukrotnych pomiarów tej samej zmiennej w różnych warunkach. Pomiaru tej samej zmiennej może dokonać dwóch różnych sędziów (odtwarzalność) lub jeden sędzia może dokonać pomiaru dwukrotnie (powtarzalność). Współczynnik $\hat{\kappa}$ wyznacza się dla zależnych zmiennych kategorialnych, na podstawie danych surowych albo z wykonanej na ich podstawie tabeli kontyngencji o wymiarach $c \times c$.

WARUNKI STOSOWANIA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– pomiar na skali nominalnej (Kappa Cohena – nieważona) lub skali porządkowej (Kappa ważona).

DEFINICJE MIAR

Współczynnik Kappa ($\hat{\kappa}$) – jego wartość zawiera się w przedziale od -1 do 1. Wartość 1 oznacza pełną zgodność, wartość 0 oznacza zgodność na poziomie takim samym, jaki powstałby dla losowego rozłożenia danych w tabeli kontyngencji. Poziom pomiędzy 0 a -1 jest w praktyce niewykorzystywany – ujemna wartość $\hat{\kappa}$ oznacza zgodność na poziomie mniejszym niż powstała dla losowego rozłożenia danych w tabeli kontyngencji. $\hat{\kappa}$ oznacza współczynnik zgodności w próbie, natomiast κ w populacji.

W zależności od potrzeb można wyznaczać Kappę nieważoną (czyli Kappę Cohena) lub Kappę ważoną. Przydzielane wagi (w_{ij}) odnoszą się do poszczególnych komórek tabeli kontyngencji: na przekątnej wynoszą 1, a poza przekątną należą do przedziału $\langle 0; 1)$.

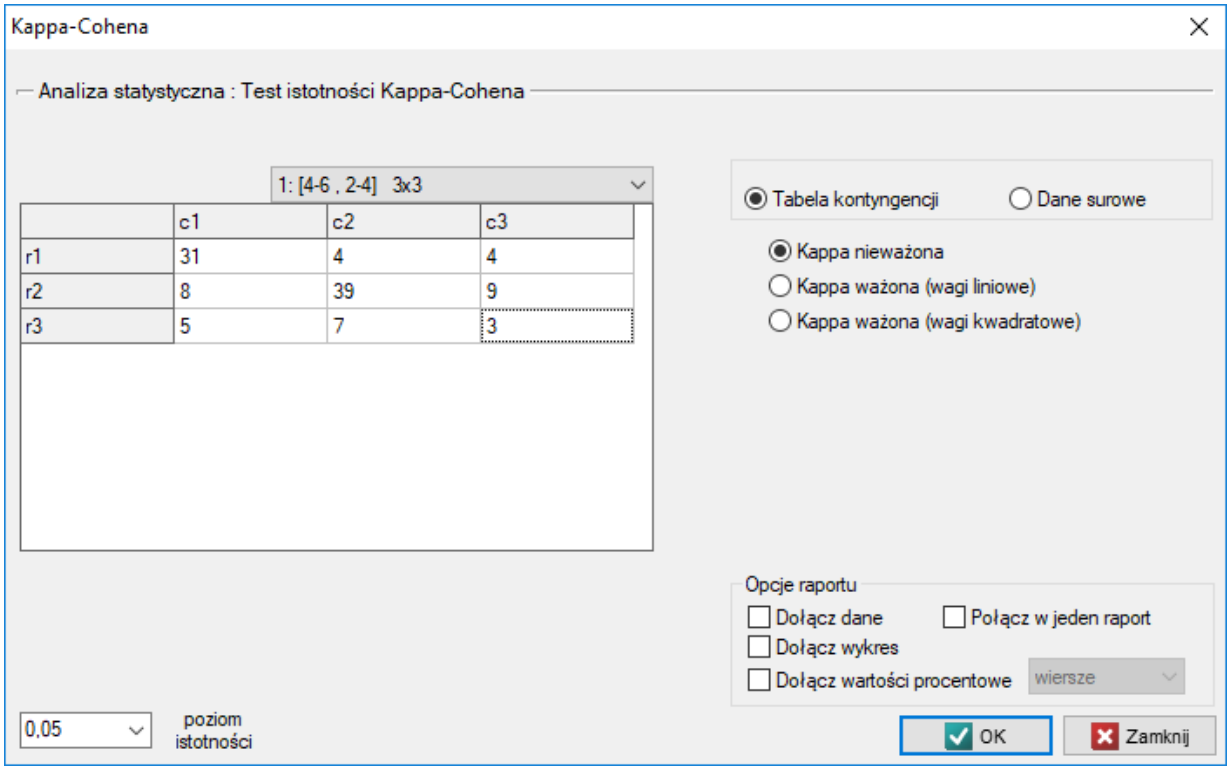
Kappa nieważona – wyliczana jest dla danych, których kategorii nie da się uporządkować, np. dane pochodzą od pacjentów, których dzielimy ze względu na rodzaj zdiagnozowanej choroby, a chorób tych nie można uporządkować (zapalenie płuc (1), zapalenie oskrzeli (2), inne (3)). W takiej sytuacji można sprawdzać zgodność diagnoz wystawionych przez dwóch lekarzy, stosując nieważoną Kappę, czyli Kappę Cohena. Niezgodność par $\{(1), (3)\}$ oraz $\{(1), (2)\}$ traktowana będzie równoważnie, a więc wagi poza przekątną macierzy wag będą zerowane.

Kappa ważona – w sytuacji, gdy kategorie danych mogą być posortowane (np. dane pochodzą od pacjentów, których dzielimy ze względu na stopień zmian chorobowych: brak zmian (1), zmiany łagodne (2), podejrzenie raka (3), rak (4)), zgodność ocen wystawionych przez dwóch radiologów można budować z uwzględnieniem możliwości sortowania. Za bardziej niezgodne pary ocen mogą być wówczas uznane oceny $\{(1), (4)\}$ niż $\{(1), (2)\}$. By kolejność kategorii wpływała na wynik zgodności, wyznaczać należy ważoną Kappę. Przydzielane wagi mogą mieć postać liniową (Cicchetti, 1971[37]) lub kwadratową (Cohen, 1968[45]) – im większe oddalenie od przekątnej macierzy, tym mniejsza waga, przy czym wagi liniowe maleją proporcjonalnie, a wagi kwadratowe wolniej w bliższej odległości od przekątnej i szybciej w odległości dalszej. Wagi kwadratowe cieszą się większym zainteresowaniem ze względu na praktyczną interpretację współczynnika Kappa, który w tym przypadku jest tożsamy ze współczynnikiem korelacji wewnątrzklasowej[61].

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Kappa-Cohena (lub poprzez Kreator).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

PRZYKŁAD 23.3. (plik diagnoza.pqs)
Badamy zgodność diagnozy postawionej przez 2 lekarzy. W tym celu pobieramy próbę 110 pacjentów szpitala dziecięcego. Lekarze przyjmują pacjentów w sąsiednich gabinetach. Każdy z pacjentów jest najpierw badany przez lekarza A, a następnie przez lekarza B. Diagnozy postawione przez lekarzy przedstawia poniższa tabela.

	zapalenie płuc	zapalenie oskrzeli	inne
zapalenie płuc	31	4	4
zapalenie oskrzeli	8	39	9
inne	5	7	3

Moglibyśmy badać zgodność diagnozy poprzez zwykły procent wartości zgodnych. W naszym przykładzie zgodną diagnozę lekarze postawili dla 73 pacjentów (31+39+3=73), co stanowi 66.36% badanej grupy. Współczynnik Kappa wprowadza korekcję tej wartości o szanse na zgodność (tzn. koryguje o tę zgodność, która pojawia się dla przypadkowego rozłożenia danych w tabeli).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

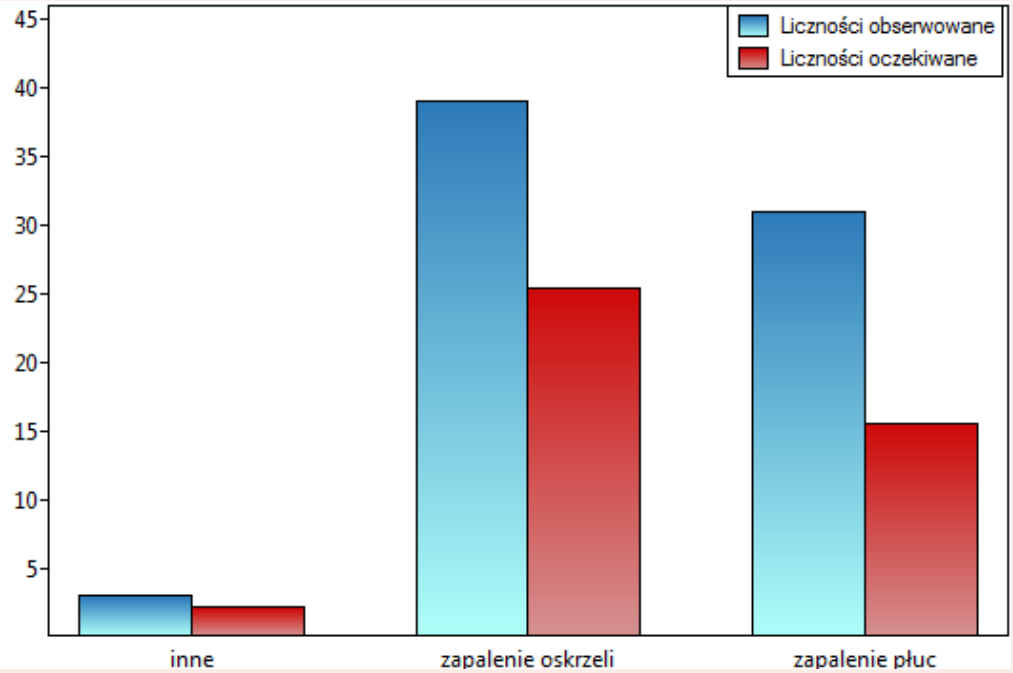
Test istotności Kappa-Cohena	
Czas analizy	0.03 sek.
Analizowane zmienne	Lekarz 1;Lekarz 2
Poziom istotności	0.05
Liczność - ilość par	110
Współczynnik Kappa	0.4458061
Błąd dla Kappa	0.068029
-95% CI dla współczynnika Kappa	0.3124717
+95% CI dla współczynnika Kappa	0.5791405
Blad dla rozkładu Kappa	0.0723248
Statystyka Z	6.163942
Wartość p (asymptotyczne)	<0.0000001

Dane :

	inne	zapalenie	zapalenie
inne	3	7	5
zapalenie	9	39	8
zapalenie	4	4	31

sumy:

	inne	zapalenie	zapalenie
inne	2.73%	6.36%	4.55%
zapalenie	8.18%	35.45%	7.27%
zapalenie	3.64%	3.64%	28.18%



$\hat{\kappa} = 0.4458$, $p < 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Diagnozy dwóch lekarzy są statystycznie zgodne.

Zgodność wyrażona współczynnikiem $\hat{\kappa} = 44.58\%$ jest mniejsza niż ta nieskorygowana o szanse na zgodność.

PRZYKŁAD 23.4. (plik radiologia.pqs)
W obrazie radiologicznym oceniano uszkodzenie wątroby w następujących kategoriach: brak zmian (1), zmiany łagodne (2), podejrzenie raka (3), rak (4). Oceny dokonywało dwóch niezależnych radiologów bazując na grupie 70 pacjentów. Chcemy sprawdzić zgodność postawionej diagnozy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

	brak zmian	zmiany łagodne	podejrzenie raka	rak
brak zmian	23	4	3	0
zmiany łagodne	2	12	5	2
podejrzenie raka	5	4	4	0
rak	1	2	1	2

Ze względu na to, że diagnoza wystawiona jest na skali porządkowej, właściwą miarą zgodności byłby ważony współczynnik Kappa.

Test istotności Kappa-Cohena	
Czas analizy	0,06 sek.
Analizowane zmienne	Tabela kontyngencji
Poziom istotności	0,05
Liczność - ilość par	70
Typ	ważona (wagi liniowe)
Współczynnik Kappa	0,390194
Błąd dla Kappa	0,090462
-95% CI dla współczynnika Kappa	0,212893
+95% CI dla współczynnika Kappa	0,567496
Błąd dla rozkładu Kappa	0,086638
Statystyka Z	4,503732
Wartość p (asymptotyczne)	0,000007

Test istotności Kappa-Cohena	
Czas analizy	0,06 sek.
Analizowane zmienne	Tabela kontyngencji
Poziom istotności	0,05
Liczność - ilość par	70
Typ	ważona (wagi kwadratowe)
Współczynnik Kappa	0,418658
Błąd dla Kappa	0,110945
-95% CI dla współczynnika Kappa	0,201208
+95% CI dla współczynnika Kappa	0,636107
Błąd dla rozkładu Kappa	0,118935
Statystyka Z	3,520052
Wartość p (asymptotyczne)	0,000431

$\hat{\kappa} = 0.390194$ dla wag liniowych ($p = 0.000007$), $\hat{\kappa} = 0.418658$ dla wag kwadratowych ($p = 0.000431$) – obie wartości istotne statystycznie, $\alpha = 0.05$

Diagnozy dwóch radiologów są statystycznie zgodne, niezależnie od zastosowanych wag.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Ponieważ dane koncentrują się głównie na głównej przekątnej macierzy i w jej bliskim sąsiedztwie, współczynnik ważony wagami liniowymi jest niższy niż współczynnik wyznaczony dla wag kwadratowych. Gdyby w ocenach istniała duża niezgodność dotycząca dwóch skrajnych przypadków i para (brak zmian i rak), znajdująca się w prawym górnym rogu tabeli, występowała zdecydowanie częściej, np. 15 razy, wówczas taki duży brak zgodności byłby bardziej widoczny przy wykorzystaniu wag kwadratowych (współczynnik Kappa drastycznie by spadł) niż przy wykorzystaniu wag liniowych.

DLA ZAINTERESOWANYCH

Wagi liniowe – wzór:

$$w_{ij} = 1 - \frac{|i-j|}{c-1}.$$

Przykładowe wagi dla macierzy wielkości 5x5:

1	0.75	0.5	0.25	0
0.75	1	0.75	0.5	0.25
0.5	0.75	1	0.75	0.5
0.25	0.5	0.75	1	0.75
0	0.25	0.5	0.75	1

Wagi kwadratowe – wzór:

$$w_{ij} = 1 - \frac{(i-j)^2}{(c-1)^2}.$$

Przykładowe wagi dla macierzy wielkości 5x5:

1	0.9375	0.75	0.4375	0
0.9375	1	0.9375	0.75	0.4375
0.75	0.9375	1	0.9375	0.75
0.4375	0.75	0.9375	1	0.9375
0	0.4375	0.75	0.9375	1

Procedura:
By wyznaczyć zgodność współczynnikiem Kappa, dane przedstawia się w postaci tabeli liczności obserwowanych O_{ij} (**??**), którą następnie przekształca się w tabelę kontyngencji prawdopodobieństw $p_{ij} = O_{ij}/n$.

Współczynnik Kappa ($\hat{\kappa}$) wyraża się wtedy wzorem:

$$\hat{\kappa} = \frac{P_o - P_e}{1 - P_e},$$

gdzie:
$P_o = \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} p_{ij}$,
$P_e = \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} p_{i.} p_{.i}$,
$p_{i.}$, $p_{.i}$ – sumy końcowe kolumn i wierszy tabeli kontyngencji prawdopodobieństw.

Błąd standardowy dla Kappa wyraża się wzorem:

$$SE_{\hat{\kappa}} = \frac{1}{(1-P_e)\sqrt{n}} \sqrt{\sum_{i=1}^{c} \sum_{j=1}^{c} p_{i.} p_{.j} [w_{ij} - (\overline{w}_{i.} + (\overline{w}_{.j})]^2 - P_e^2}$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

gdzie:
$\overline{w}_{i.} = \sum_{j=1}^{c} p_{.j} w_{ij}$,
$\overline{w}_{.j} = \sum_{i=1}^{c} p_{i.} w_{ij}$.

Test Z do sprawdzania istotności współczynnika Kappa Cohena ($\hat{\kappa}$) *(ang. The Z test of significance for the Cohen's Kappa)* (Fleiss, 2003[62]) służy do weryfikacji hipotezy o zgodności wyników dwukrotnych pomiarów $X^{(1)}$ i $X^{(2)}$ cechy X i opiera się na współczynniku $\hat{\kappa}$ wyliczonym dla próby.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad \kappa = 0, \\ \mathcal{H}_1 &: \quad \kappa \neq 0. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{\hat{\kappa}}{SE_{\kappa_{distr}}},$$

gdzie:

$$SE_{\kappa_{distr}} = \frac{1}{(1-P_e)\sqrt{n}} \sqrt{\sum_{i=1}^{c}\sum_{j=1}^{c} p_{ij}[w_{ij} - (\overline{w}_{i.} + \overline{w}_{.j})(1-\hat{\kappa})]^2 - [\hat{\kappa} - P_e(1-\hat{\kappa})]^2}.$$

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Współczynnik Kappa Fleissa i test badający jego istotność

KIEDY STOSOWAĆ:

Współczynnik Kappa Fleissa określa zgodność pomiarów prowadzonych przez kilku sędziów (Fleiss, 1971[60]) i jest rozszerzeniem Kappy Cohena, pozwalającego na badanie zgodności jedynie dwóch sędziów. Każdy z n losowo wybranych obiektów może być oceniany przez inny losowy zestaw k sędziów. Analiza opiera się na danych przekształconych do tabeli o n wierszach i c kolumnach, gdzie c stanowi liczbę możliwych kategorii, do których sędziowie przydzielają badane obiekty.

WARUNKI STOSOWANIA

- pomiar na skali nominalnej – ewentualne uporządkowanie kategorii nie jest brane pod uwagę.

DEFINICJE MIAR

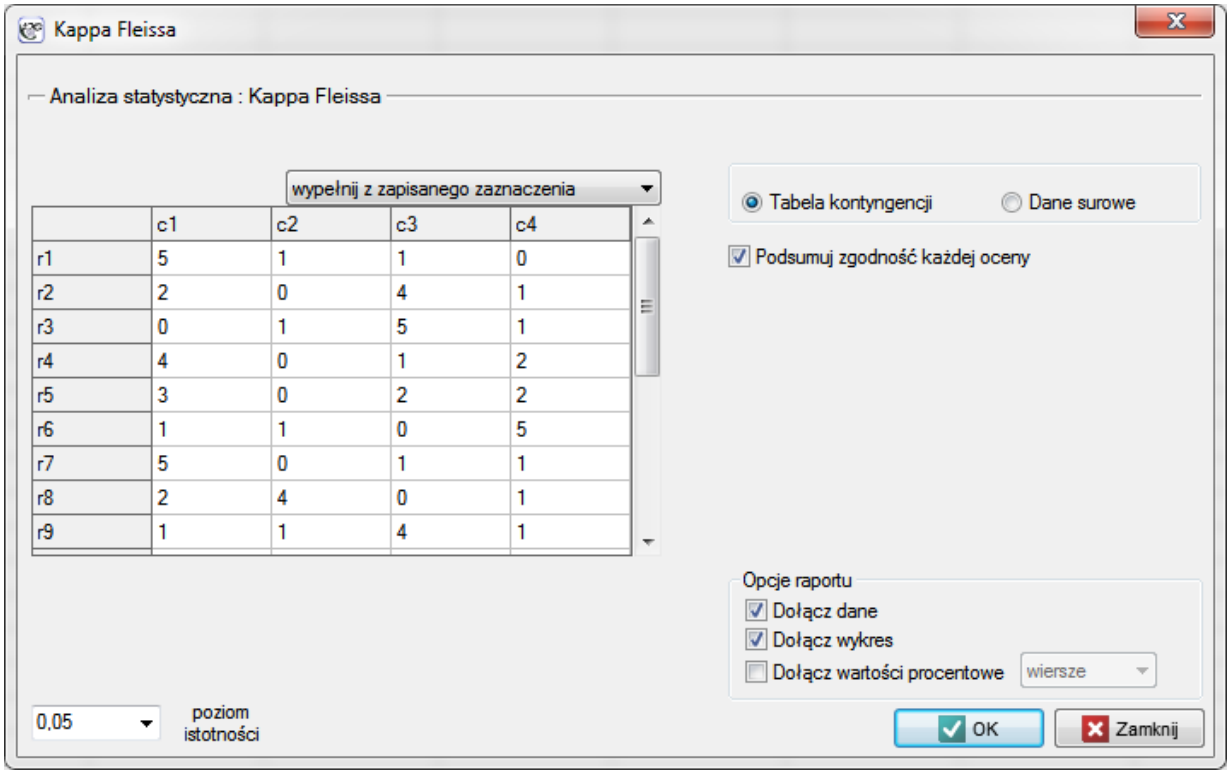
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Współczynnik Kappa ($\hat{\kappa}$) – wartość $\hat{\kappa} = 1$ oznacza pełną zgodność sędziów, natomiast $\hat{\kappa} = 0$ oznacza zgodność, jaka powstałaby, gdyby opinie sędziów wydane były w sposób losowy. Wartości ujemne Kappa wskazują natomiast na zgodność mniejszą niż na poziomie losowym.

Uwaga! Wyznaczanie współczynnika Kappa Fleissa jest zbliżone koncepcyjnie do metody Mantela-Haenszela. Wyznaczona Kappa jest ogólną miarą podsumowującą zgodność wszystkich ocen sędziowskich i może być wyznaczona jako Kappa utworzona z poszczególnych warstw, którymi są konkretne oceny sędziowskie (Fleiss, 2003[62]). Dlatego, jako podsumowanie każdej warstwy, można wyznaczyć zgodność sędziowską (współczynnik Kappa) podsumowującą każdą możliwą ocenę z osobna.

JAK USTAWIĆ ANALIZĘ

Statystyka → Testy nieparametryczne → Kappa-Fleissa.



	c1	c2	c3	c4
r1	5	1	1	0
r2	2	0	4	1
r3	0	1	5	1
r4	4	0	1	2
r5	3	0	2	2
r6	1	1	0	5
r7	5	0	1	1
r8	2	4	0	1
r9	1	1	4	1

WYKONANIE I INTERPRETACJA

Przykład 23.5. (plik temperament.pqs)
20 ochotników bierze udział w zabawie mającej na celu ustalenie typu osobowości badanych. Każdy z ochotników dysponuje oceną wystawioną przez 7 różnych obserwatorów (najczęściej osób z bliskiego otoczenia lub rodziny). Każdy z obserwatorów został zapoznany z podstawowymi cechami opisującymi temperament w poszczególnych typach osobowości: choleryk, flegmatyk, melancholik, sangwinik. Badamy zgodność obserwatorów w przypisywaniu typów osobowości. Fragment danych przedstawia poniższa tabela.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

	Choleryk	Flegmatyk	Melancholik	Sangwinik
Ochotnik 1	5	1	1	0
Ochotnik 2	2	0	4	1
Ochotnik 3	0	1	5	1
Ochotnik 4	4	0	1	2
Ochotnik 5	3	0	2	2
Ochotnik 6	1	1	0	5

Kappa Fleissa	
Czas analizy	0,17 sek.
Analizowane zmienne	Var1;Var2
Poziom istotności	0,05
Liczba sędziów	7
Liczba kategorii ocen	4
Liczba obiektów	20
Łączna liczba wydanych ocen	140
Współczynnik Kappa	0,244918
Błąd dla Kappa	0,028632
-95% CI dla współczynnika Kappa	0,188801
+95% CI dla współczynnika Kappa	0,301035
Statystyka Z	8,554087
Wartość p (asymptotyczne)	<0.000001

$\hat{\kappa} = 0.244918$, $p < 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

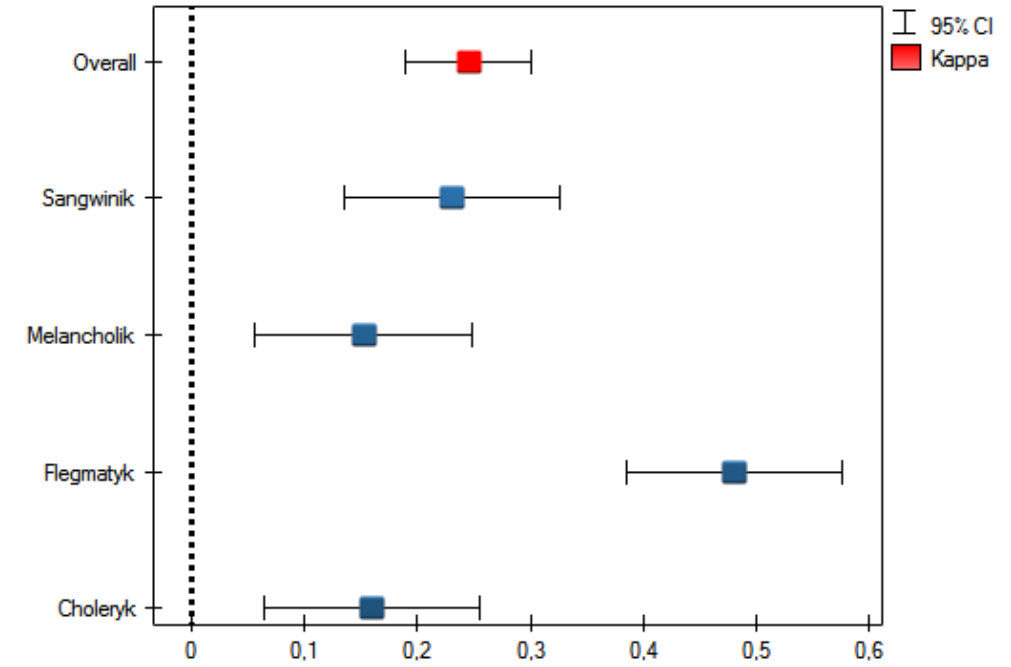
Obserwujemy niewielki, lecz statystycznie istotny współczynnik Kappa, co oznacza nieprzypadkową zgodność ocen sędziowskich.

Istotna zgodność dotyczy każdej oceny, czego potwierdzeniem jest raport podsumowujący zgodność dla każdej warstwy (dla każdej oceny) oraz wykres prezentujący poszczególne współczynniki Kappa i Kappę podsumowującą całość.

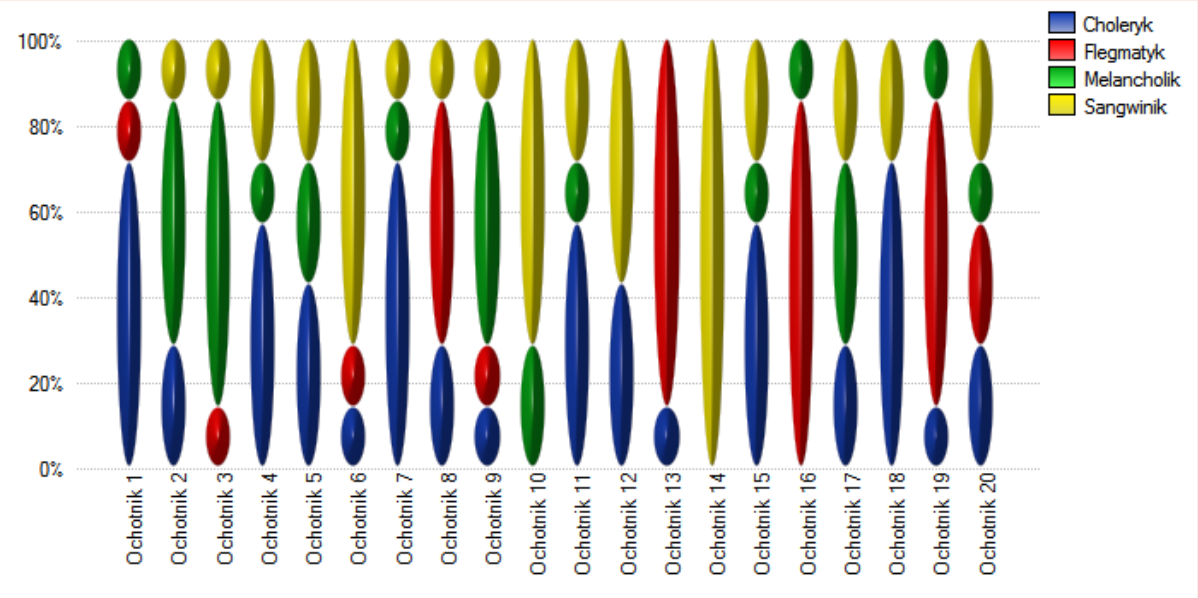
Podsumowanie zgodności

	Choleryk	Flegmatyk	Melancholil	Sangwinik
Współczynnik Kappa	0,159454	0,479952	0,151786	0,23
Błąd dla Kappa	0,048795	0,048795	0,048795	0,048795
-95% CI dla współczyni	0,063818	0,384315	0,056149	0,134364
+95% CI dla współczyr	0,255091	0,575588	0,247422	0,325636
Statystyka Z	3,267838	9,836088	3,110681	4,713597
Wartość p (asymptotyc	0,001084	<0.000001	0,001867	0,000002

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Ciekawy może być fakt, że najwyższa zgodność dotyczy oceny flegmatyków (Kappa=0.479952). Przy niewielkiej liczbie obserwowanych osób warto również wykonać wykres obrazujący, w jaki sposób obserwatorzy oceniali każdą z nich.



W tym przypadku tylko osoba nr 14 uzyskała jednoznaczną ocenę typu osobowości – sangwinik. Osoby nr 13 i 16 ocenione były jako typ flegmatyk przez 6 obserwatorów (na 7 możliwych). W przypadku pozostałych osób panowała nieco mniejsza zgodność ocen. Najtrudniejszy do zdefiniowania typ osobowości wydaje się cechować ostatnią osobę, która uzyskała najbardziej różnorodny zestaw ocen.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Współczynnik Kappa ($\hat{\kappa}$) wyraża się wzorem:

$$\hat{\kappa} = \frac{P_o - P_e}{1 - P_e},$$

gdzie:
$P_o = \frac{1}{kn(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{c} x_{ij} - kn$,
$P_e = \sum_{i=1}^{c} q_j^2$,
$q_j = \frac{1}{km} \sum_{i=1}^{n} x_{ij}$,
x_{ij} – liczba sędziów przydzielających i-temu obiektowi kategorię j.

Dla współczynnika $\hat{\kappa}$ można wyznaczyć błąd standardowy SE, który pozwala na zbadanie istotności statystycznej i wyznaczenie asymptotycznych przedziałów ufności.

Test Z do sprawdzania istotności współczynnika Kappa Fleissa ($\hat{\kappa}$) *(ang. The Z test of significance for the Fleiss's Kappa)* (Fleiss, 2003[62]) służy do weryfikacji hipotezy o zgodności ocen kilku sędziów i opiera się na współczynniku $\hat{\kappa}$ wyliczonym dla próby.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad \kappa = 0, \\ \mathcal{H}_1 &: \quad \kappa \neq 0. \end{aligned}$$

Wzory: statystyka testowa ma postać:

$$Z = \frac{\hat{\kappa}}{SE},$$

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

24 TESTY DIAGNOSTYCZNE

Ocena testu diagnostycznego

KIEDY STOSOWAĆ:

Załóżmy, że przy pomocy testu diagnostycznego badamy występowanie danej cechy (najczęściej choroby) i znamy rzeczywistość (tzw. gold-standard), czyli wiemy, czy ta cecha rzeczywiście występuje u badanych osób. Na podstawie tych informacji możemy zbudować tabelę kontyngencji 2×2:

Liczności obserwowane		Rzeczywistość (gold-standard)		
		choroba **(+)**	brak choroby **(−)**	**Suma**
test diagnostyczny	wynik pozytywny **(+)**	TP	FP	TP+FP
	wynik negatywny **(−)**	FN	TN	FN+TN
	Suma	TP+FN	FP+TN	n=TP+FP+FN+TN

gdzie:
TP – wyniki prawdziwie dodatnie (*ang. true positive*),
FP – wyniki fałszywie dodatnie (*ang. false positive*),
FN – wyniki fałszywie ujemne (*ang. false negative*),
TN – wyniki prawdziwie ujemne (*ang. true negative*).
Dla takiej tabeli możemy wyliczyć podane niżej miary.

DEFINICJE MIAR

Czułość *(ang. sensitivity)* – opisuje zdolność wykrywania osób rzeczywiście chorych (posiadających daną cechę). Jeśli więc badamy grupę osób chorych, to czułość daje nam informację, jaki procent z nich ma pozytywny wynik testu.

Swoistość *(ang. specificity)* – opisuje zdolność wykrywania osób rzeczywiście zdrowych (bez danej cechy). Jeśli więc badamy grupę osób zdrowych, to swoistość daje nam informację, jaki procent z nich ma negatywny wynik testu.

Wartość predykcyjna dodatnia (PPV) *(ang. positive predictive value)* – prawdopodobieństwo, że osobnik ma chorobę, mając pozytywny wynik testu. Jeśli więc badana osoba otrzymała pozytywny wynik testu, to PPV daje jej informację, na ile może być pewna, że cierpi na daną chorobę.

Wartość predykcyjna ujemna (NPV) *(ang. negative predictive value)* – prawdopodobieństwo, że osobnik nie ma choroby, mając negatywny wynik testu. Jeśli więc badana osoba otrzymała negatywny wynik testu, to NPV daje jej informację, na ile może być pewna, że nie cierpi na daną chorobę. Wartości predykcyjne dodatnie i ujemne są zależne od rozpowszechnienia choroby (od współczynnika chorobowości).

Współczynnik chorobowości *(ang. prevalence)* – prawdopodobieństwo wystąpienia choroby w populacji, dla której przeprowadzony był test diagnostyczny.

Iloraz wiarygodności wyniku dodatniego (LR_+) *(ang. likelihood ratio of positive test)* – miara ta pozwala na porównywanie dopasowania wyników kilku testów do tzw. gold-standard i nie jest zależna od rozpowszechnienia choroby. Jest to iloraz dwóch szans: szansy, że pozytywny wynik testu otrzyma osoba z grupy chorych, do szansy, że ten sam wynik będzie obserwowany wśród osób zdrowych.

Iloraz wiarygodności wyniku ujemnego (LR_-) *(ang. likelihood ratio of negative test)* – iloraz dwóch

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

szans: szansy, że negatywny wynik testu otrzyma osoba z grupy chorych, do szansy, że ten sam wynik będzie obserwowany wśród osób zdrowych.

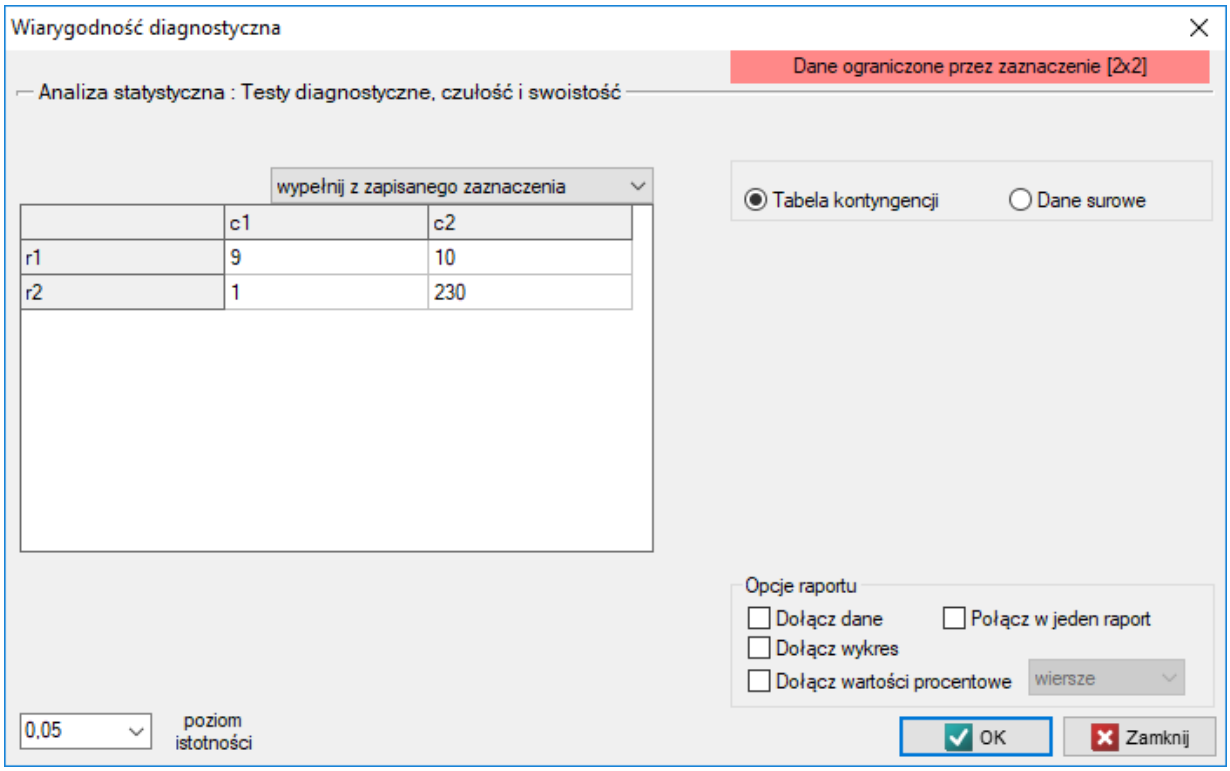
Dokładność (Acc) *(ang. Accuracy)* – prawdopodobieństwo prawidłowej diagnozy przy wykorzystaniu testu diagnostycznego. Jeśli więc badana osoba otrzymała pozytywny lub negatywny wynik testu, to Acc daje jej informację o tym, na ile może być pewna postawionej diagnozy.

Diagnostyczny iloraz szans (DOR) *(ang. Diagnostic Odds Ratio)* – iloraz dwóch szans: szansy na pozytywny wynik testu osoby chorej do szansy na pozytywny wynik testu osoby zdrowej.

Uwaga! Przedziały ufności dla czułości, swoistości, PPV, NPV, Acc oraz współczynnika chorobowości budowane są w oparciu o metodę Cloppera-Pearsona dla pojedynczej proporcji.

JAK USTAWIĆ ANALIZĘ

Statystyka zaawansowana → Testy diagnostyczne → Wiarygodność diagnostyczna.



WYKONANIE I INTERPRETACJA

Przykład 24.1. (plik mammografia.pqs)
Mammografia jest jednym z najpowszechniej stosowanych testów przesiewowych pozwalających na wykrycie raka piersi. Poniższe badanie zostało przeprowadzone na grupie 250 tzw. „bezobjawowych" kobiet w wieku od 40 do 50 lat. Mammografia może wykryć ognisko raka mniejsze niż 5 mm, ale również pozwala stwierdzić zmiany, które nie są jeszcze guzkiem, a jedynie zmianą struktury tkanek. Poniżej przedstawiono przykładowy wynik badania mammograficznego.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Liczności obserwowane		Rzeczywistość (badanie histopatologiczne)		
		choroba **(+)**	brak choroby **(−)**	**Suma**
test diagnostyczny	wynik pozytywny **(+)**	9	10	19
	wynik negatywny **(−)**	1	230	231
	Suma	10	240	250

Wyznaczymy wartości pozwalające dokonać oceny przeprowadzonego testu diagnostycznego.

Testy diagnostyczne, czułość i swoistość	
Analizowane zmienne	Tabela kontyngencji
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Czułość	0.9
-95% CI	0.554984
+95% CI	0.997733
Swoistość	0.958333
-95% CI	0.92471
+95% CI	0.979841
Wartości predykcyjne dodatnie (PPV)	0.473684
-95% CI	0.244475
+95% CI	0.711357
Wartości predykcyjne ujemne (NPV)	0.995671
-95% CI	0.976118
+95% CI	0.999902
Iloraz wiarygodności wyniku dodatniego (PLR)	21.6
-95% CI	11.378651
+95% CI	41.003107
Iloraz wiarygodności wyniku ujemnego (NLR)	0.104348
-95% CI	0.016251
+95% CI	0.670016
Dokładność (ACC)	0.956
-95% CI	0.922637
+95% CI	0.977834
Współczynnik chorobowości	0.04
-95% CI	0.019346
+95% CI	0.072329
Diagnostyczny iloraz szans (DOR)	207
-95% CI	23.852775
+95% CI	1796.394772

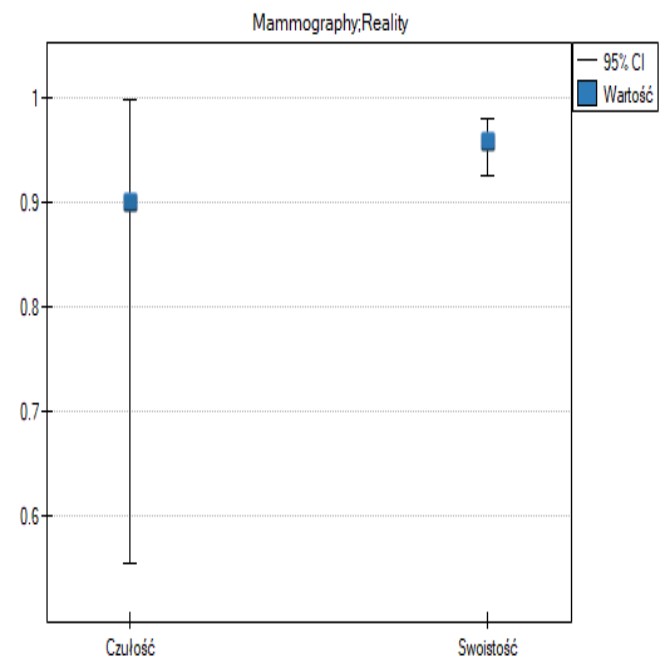
Dane :

	A	B
A	9	10
B	1	230

wiersze:

	A	B
A	47.37%	52.63%
B	0.43%	99.57%

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



czułość=90%, swoistość=95.83%, PPV=47.37%, NPV=99.57%

Test mammograficzny wykazuje wysoką czułość i swoistość, jednak ze względu na niski współczynnik chorobowości pozytywny wynik testu nie daje jeszcze pewności rozpoznania.

Interpretacja poszczególnych miar:

- 90% kobiet chorych na raka piersi zostało poprawnie zdiagnozowanych, czyli uzyskało pozytywny wynik mammografii;
- 95.83% kobiet zdrowych (niechorujących na raka piersi) zostało poprawnie zdiagnozowanych, czyli uzyskało negatywny wynik mammografii;
- 4 kobiety na 100 przebadanych cierpią z powodu raka piersi;
- kobieta uzyskująca pozytywny wynik mammografii może być w 47.37% pewna, że ma raka piersi;
- kobieta uzyskująca negatywny wynik mammografii może być w 99.57% pewna, że nie ma raka piersi;
- szansa na to, że pozytywny wynik mammografii otrzyma kobieta rzeczywiście chora na raka, jest 21.60 razy większa niż szansa, że pozytywny wynik mammografii otrzyma kobieta rzeczywiście zdrowa;
- szansa na to, że negatywny wynik mammografii otrzyma kobieta rzeczywiście chora na raka, stanowi 10.43% szansy na to, że negatywny wynik mammografii otrzyma kobieta rzeczywiście zdrowa;
- kobieta poddająca się mammografii (bez względu na uzyskany wynik) może być pewna postawionej diagnozy w 96.50%;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- szansa na pozytywny wynik testu u kobiety, która rzeczywiście choruje na raka piersi, jest 207 razy większa od szansy na taki wynik u kobiety zdrowej.

DLA ZAINTERESOWANYCH

Czułość

$$\text{czułość} = \frac{TP}{TP+FN}$$

Swoistość

$$\text{swoistość} = \frac{TN}{FP+TN}$$

Wartość predykcyjna dodatnia

$$PPV = \frac{TP}{TP+FP}$$

Wartość predykcyjna ujemna

$$NPV = \frac{TN}{FN+TN}$$

Współczynnik chorobowości

$$\text{Współczynnik chorobowości} = \frac{TP+FN}{n}$$

Iloraz wiarygodności wyniku dodatniego

$$LR_+ = \frac{\text{czułość}}{1-\text{swoistość}} = \frac{TP\,(TP+FN)}{FP\,(FP+TN)}$$

Przedział ufności dla LR_+ buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{1-\text{czułość}}{TP} + \frac{\text{swoistość}}{FP}}.$$

Iloraz wiarygodności wyniku ujemnego

$$LR_- = \frac{1-\text{czułość}}{\text{swoistość}} = \frac{FN\,(TP+FN)}{TN\,(FP+TN)}$$

Przedział ufności dla LR_- buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{\text{czułość}}{FN} + \frac{1-\text{swoistość}}{TN}}.$$

Dokładność

$$Acc = \frac{TP+TN}{n}$$

Diagnostyczny iloraz szans

$$DOR = \frac{TP/FN}{FP/TN}$$

Przedział ufności dla DOR buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{1}{TP} + \frac{1}{FN} + \frac{1}{FP} + \frac{1}{TN}}.$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krzywa ROC

KIEDY STOSOWAĆ:

Testem diagnostycznym posługujemy się, by odróżnić obiekty z daną cechą (oznaczone jako **(+)**, np. osoby chore) od obiektów bez danej cechy (oznaczone jako **(−)**, np. osoby zdrowe). Aby test diagnostyczny mógł być uznany za wartościowy, powinien dawać stosunkowo niewielką liczbę błędnych klasyfikacji. Jeśli test opiera się na zmiennej dychotomicznej, wówczas właściwym narzędziem do oceny jego jakości jest analiza tabeli kontyngencji 2×2. Najczęściej jednak testy diagnostyczne opierają się na zmiennych ciągłych lub o uporządkowanych kategoriach. W takiej sytuacji właściwym środkiem oceny zdolności testu do rozróżnienia **(+)** i **(−)** są krzywe ROC *(ang. Receiver Operating Characteristic)*.

Często obserwuje się, że wraz ze wzrostem wartości zmiennej diagnostycznej rosną szanse na wystąpienie badanego zjawiska, lub odwrotnie: wraz ze wzrostem wartości zmiennej diagnostycznej maleją szanse na wystąpienie badanego zjawiska. Wówczas przy użyciu krzywych ROC dokonuje się wyboru optymalnego punktu odcięcia, czyli pewnej wartości zmiennej diagnostycznej, która najlepiej dzieli badaną zbiorowość na dwie grupy: **(+)**, w której występuje dane zjawisko, i **(−)**, w której dane zjawisko nie występuje. Kiedy w oparciu o badania przeprowadzone na tych samych obiektach są zbudowane dwie lub więcej krzywych ROC, można dokonać porównania tych krzywych pod kątem jakości klasyfikacji.

WARUNKI STOSOWANIA

- pomiar zmiennej diagnostycznej na skali porządkowej lub interwałowej,
- znana rzeczywista przynależność każdego obiektu do grupy **(+)** lub **(−)** (tzw. gold-standard).

DEFINICJE MIAR

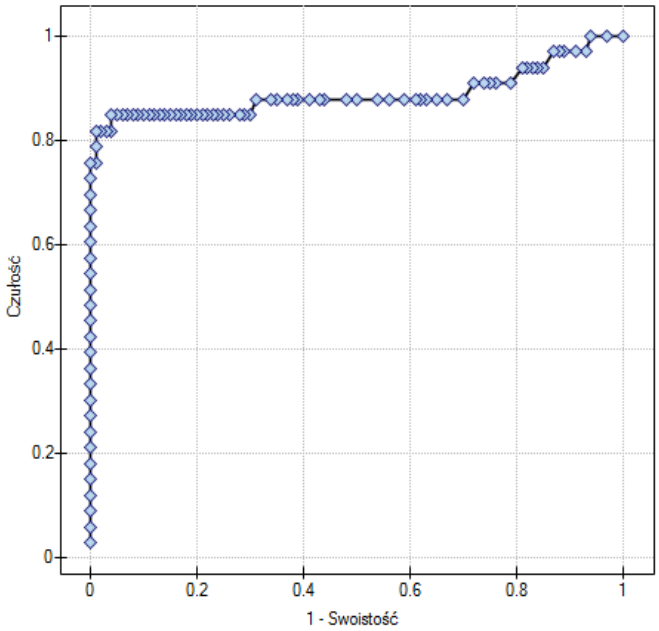
Załóżmy, że dysponujemy n-elementową próbą, w której każdy obiekt uzyskuje jedną z k wartości zmiennej diagnostycznej. Każda z uzyskanych wartości zmiennej diagnostycznej $x_1, x_2, ...x_k$ staje się potencjalnym punktem odcięcia x_{cat}. Jeśli zmienna diagnostyczna to:

- **stymulanta** (wraz ze wzrostem jej wartości rosną szanse na wystąpienie badanego zjawiska), to wartości większe lub równe punktowi odcięcia ($x_i \geq x_{cat}$) zaliczamy do grupy **(+)**;
- **destymulanta** (wraz ze wzrostem jej wartości maleją szanse na wystąpienie badanego zjawiska), to wartości mniejsze lub równe punktowi odcięcia ($x_i \leq x_{cat}$) zaliczamy do grupy **(+)**.

Dla każdego z k punktów odcięcia wyznaczamy wartości prawdziwie dodatnie (TP), prawdziwie ujemne (TN), fałszywie dodatnie (FP) i fałszywie ujemne (FN), a na ich podstawie czułość, swoistość oraz pozostałe miary opisane w rozdziale Ocena testu diagnostycznego.

AUC *(ang. area under curve)* – wielkość pola pod krzywą ROC, mieści się w przedziale $\langle 0; 1 \rangle$. Krzywa ROC powstaje na podstawie wyznaczonych wartości czułości i swoistości dla kolejnych punktów odcięcia – na osi odciętych umieszczona jest $x = 1-$swoistość, a na osi rzędnych $y =$czułość, a uzyskane punkty są ze sobą połączone.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Im większe jest pole AUC, tym dokładniej zaklasyfikujemy obiekty do grupy **(+)** i **(−)** na podstawie analizowanej zmiennej diagnostycznej, a więc z tym lepszym skutkiem ta zmienna diagnostyczna może być wykorzystywana jako klasyfikator. Gdy krzywa ROC pokrywa się z przekątną $y = x$ (czyli $AUC = 0.5$), decyzja podejmowana na podstawie zmiennej diagnostycznej jest tak samo dobra jak losowy podział badanych obiektów na grupy **(+)** i **(−)**.

Uwaga! Pole AUC, błąd SE_{AUC} i przedział ufności dla AUC wyliczane są w oparciu o:

- ⋆ metodę nieparametryczną **DeLong** (DeLong E.R. i inni, 1988[51]; Hanley J.A. i Hajian-Tilaki K.O., 1997[75]) – **rekomendowane**,
- ⋆ metodę nieparametryczną **Hanley-McNeil** (Hanley J.A. i McNeil M.D., 1982[76]),
- ⋆ metodę **Hanley-McNeil** zakładającą dwu-ujemny rozkład wykładniczy (Hanley J.A. i McNeil M.D., 1982[76]) – wyliczaną tylko wtedy, gdy grupy **(+)** i **(−)** są równoliczne.

WYKONANIE I INTERPRETACJA

Test istotności pola AUC jest zawsze wykonywany łącznie z wyborem optymalnego punktu odcięcia – patrz Przykład 24.2 w kolejnym rozdziale.

DLA ZAINTERESOWANYCH

By klasyfikacja była lepsza niż losowy podział obiektów do dwóch klas, pole pod krzywą ROC powinno być istotnie większe niż pole pod prostą $y = x$, czyli niż 0.5.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad AUC = 0.5, \\ \mathcal{H}_1: &\quad AUC \neq 0.5. \end{aligned}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{AUC - 0.5}{SE_{0.5}},$$

gdzie:
$SE_{0.5} = \sqrt{\frac{n_{(+)}+n_{(-)}+1}{12 n_{(+)} n_{(-)}}}$,
$n_{(+)}$ – liczność grupy **(+)**, w której dane zjawisko rzeczywiście występuje,
$n_{(-)}$ – liczność grupy **(−)**, w której dane zjawisko rzeczywiście nie występuje.

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Dodatkowo, gdy przyjmiemy, że parametr diagnostyczny tworzy wysokie pole (AUC), możemy wybrać optymalny punkt odcięcia.

Poprawione wartości predykcyjne – wzory:
Program PQStat na podstawie posiadanej próby wylicza współczynnik chorobowości. Wyliczony współczynnik chorobowości będzie odzwierciedlał występowanie badanego zjawiska (choroby) w populacji, gdy są to badania przesiewowe obejmujące dużą próbę reprezentującą populację. Gdy na badania skierowane są tylko osoby z podejrzeniem choroby, wyliczony dla nich współczynnik chorobowości może być znacznie wyższy od tego współczynnika w populacji. Ponieważ zarówno wartość predykcyjna dodatnia, jak i ujemna zależy od współczynnika chorobowości, znając a priori ten współczynnik dla populacji, możemy się nim posłużyć, by wyliczyć dla każdego punktu odcięcia x_{cat} poprawione wartości predykcyjne zgodnie ze wzorami Bayesa:

$$PPV_{revised} = \frac{\text{Czułość} \cdot P_{apriori}}{\text{Czułość} \cdot P_{apriori} + (1 - \text{Swoistość}) \cdot (1 - P_{apriori})}$$

$$NPV_{revised} = \frac{\text{Swoistość} \cdot (1 - P_{apriori})}{\text{Swoistość} \cdot (1 - P_{apriori}) + (1 - \text{Czułość}) \cdot P_{apriori}}$$

gdzie $P_{apriori}$ – zadany przez użytkownika współczynnik chorobowości, tzw. pre-test probability of disease.

Zestawienie:
Wyznaczone dla każdego punktu odcięcia miary zestawia się w tabeli:

x_{cat}	**czułość**	**swoistość**	**PPV**	**NPV**	**LR$_+$**	**LR$_-$**	**Acc**	**PPV$_{rev}$**	**NPV$_{rev}$**
x_1	czułość$_1$	swoistość$_1$	PPV_1	NPV_1	LR_{+1}	LR_{-1}	Acc_1	PPV_{rev1}	NPV_{rev1}
x_2	czułość$_2$	swoistość$_2$	PPV_2	NPV_2	LR_{+2}	LR_{-2}	Acc_2	PPV_{rev2}	NPV_{rev2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	czułość$_k$	swoistość$_k$	PPV_k	NPV_k	LR_{+k}	LR_{-k}	Acc_k	PPV_{revk}	NPV_{revk}

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wybór optymalnego punktu odcięcia

KIEDY STOSOWAĆ:

Ten poszukiwany punkt to pewna wartość zmiennej diagnostycznej, która optymalnie dzieli badaną zbiorowość na dwie grupy: **(+)**, w której występuje dane zjawisko, i **(−)**, w której dane zjawisko nie występuje. Wybór optymalnego punktu odcięcia nie jest łatwy, gdyż wymaga fachowej wiedzy z zakresu tematu badań. Na przykład innego punktu odcięcia będzie wymagał test użyty w badaniach przesiewowych obejmujących dużą grupę osób, np. w badaniu mammograficznym, a innego w badaniach inwazyjnych przeprowadzanych, by potwierdzić wcześniejsze podejrzenie, np. w histopatologii. Stosując zaawansowany aparat matematyczny, możemy znaleźć taki punkt, tzw. cut-off, który będzie najkorzystniejszy z matematycznego punktu widzenia.

DEFINICJE MIAR

Program PQStat wybór optymalnego punktu odcięcia umożliwia poprzez:

Metodę stycznej (indeks kosztów) – wyliczaną w oparciu o czułość, swoistość, koszty błędnych decyzji i współczynnik chorobowości. Błędy, jakie można popełnić, przydzielając badane obiekty do grupy **(+)** i do grupy **(−)**, to wyniki fałszywie dodatnie (FP) i wyniki fałszywie ujemne (FN). Jeśli popełnienie tych błędów jest tak samo kosztowne (koszty etyczne, finansowe itd.), to wówczas w polu koszt FP i w polu koszt FN wpisujemy tę samą dodatnią wartość – zwykle 1. Jeśli natomiast uznamy, że jeden rodzaj błędu jest obarczony większym kosztem niż drugi, przypiszemy mu odpowiednio większą wagę. Optymalna wartość odcięcia obliczana jest przy użyciu wielkości m – nachylenia stycznej do krzywej ROC, określanego w odniesieniu do kosztów błędnych decyzji i współczynnika chorobowości.

Indeks Youdena – koncepcyjnie jest to maksymalna odległość pomiędzy linią będącą przekątną kwadratu o boku 1 a punktem krzywej ROC[181].

Odległość do lewego górnego rogu – koncepcyjnie jest to minimalna odległość pomiędzy lewym górnym rogiem kwadratu o boku 1 (czyli miejscem, w którym czułość i swoistość może być najwyższa) a punktem krzywej ROC.

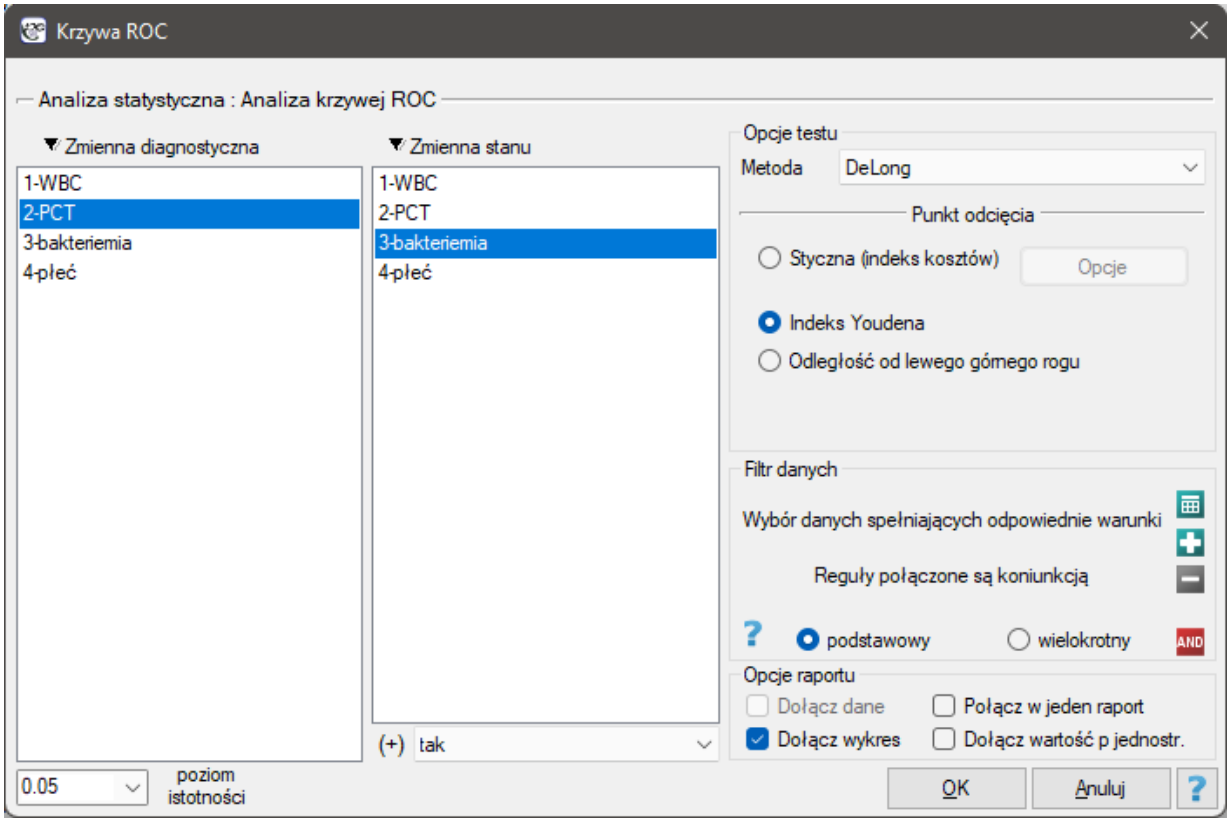
Wykres kosztów – prezentuje wyliczone wartości błędnej diagnozy wraz z ich kosztami. Zaznaczony na wykresie punkt to minimum funkcji kosztu.

Wykres przecięcia czułości i swoistości – pozwala na zlokalizowanie punktu, w którym wartość czułości i swoistości jest jednocześnie największa.

JAK USTAWIĆ ANALIZĘ

Statystyka zaawansowana → Testy diagnostyczne → Krzywa ROC.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

Przykład 24.2. (plik bakteriemia.pqs)
Utrzymująca się wysoka gorączka u niemowlęcia lub małego dziecka bez ustalonych wyraźnych przyczyn jest wskazówką do przeprowadzenia badań w kierunku bakteriemii. Za najbardziej przydatne i wiarygodne parametry służące do przesiewowej diagnostyki i monitorowania zakażeń bakteryjnych uważa się wskaźniki:

WBC – liczba białych krwinek (*ang. white blood cells*),

PCT – prokalcytonina (*ang. procalcitonin*).

Przyjmuje się, że u zdrowego niemowlęcia i małego dziecka WBC nie powinno przekraczać 15 tys/µl, a PCT powinno być niższe niż 0.5 ng/ml. Przykładowe wartości tych wskaźników dla 136 dzieci do 3. roku życia z utrzymującą się gorączką $> 39^0C$ przedstawia poniższy fragment tabeli:

WBC	PCT	bakteriemia	płeć
11,3	0,023	nie	k
11	0,022	nie	k
6,7	0,009	nie	k
5,9	0,004	nie	k
6,1	0,006	nie	k
12,5	0,031	nie	k
4,9	0,002	nie	k
6,9	0,011	nie	k
11,6	0,025	nie	k
20,9	5,919	tak	k
20,8	6,405	tak	k

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednym z możliwych sposobów analizy wskaźnika PCT jest przekształcenie go w zmienną dychotomiczną przez wybranie punktu odcięcia (np. $x_{cat} = 0.5$ ng/ml), powyżej którego badanie jest uznane za „pozytywne". Jak dobry jest taki podział, wskaże wartość czułości i swoistości. Chcemy jednak wykorzystać bardziej kompleksowe podejście, czyli wyliczyć czułość i swoistość nie tylko dla jednej wartości, ale dla każdej uzyskanej w próbie wartości PCT – czyli zbudować krzywą ROC. Na podstawie uzyskanych w ten sposób informacji chcemy sprawdzić, czy wskaźnik PCT jest rzeczywiście przydatny w rozpoznawaniu bakteriemii, a jeśli tak, jaki jest optymalny punkt odcięcia, powyżej którego możemy uznać badanie za „pozytywne" – wykrywające bakteriemię.

Ponieważ bakteriemii towarzyszy podwyższony poziom PCT, wskaźnik ten uznajemy za stymulantę. W zmiennej stanu musimy określić, która wartość znajdująca się w kolumnie bakteriemia określa jej obecność – tutaj wybieramy wartość „tak". W raporcie oprócz wyniku testu statystycznego możemy znaleźć dokładny opis każdego z możliwych punktów odcięcia.

Analiza krzywej ROC	
Czas analizy	0,20 sek.
Analizowane zmienne	PCT;bakteriemia
Liczność - braki danych	3
Poziom istotności	0,05
Liczność	133
Liczność STAN + (tak)	33
Liczność STAN - (nie)	100
Kierunek zmiennej diagnostycznej	stymulanta
Współczynnik chorobowości	0,248120301
-95% CI	0,177358632
+95% CI	0,330437258
DeLong's method	
AUC	0,889242424
SE(AUC)	0,048124092
-95% CI	0,794920921
+95% CI	0,9835639
Statystyka Z	6,691374893
Wartość p	<0.000000001

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PCT	STAN -	STAN +	FP	TP	TN	FN	Czułość	Swoistość	PPV	NPV	LR+	LR-	ACC
0,001	3	0	100	33	0	0	1	0	0,2481203	NA	1	NA	0,2481203
0,002	3	0	97	33	3	0	1	0,03	0,2538461	1	1,0309278	0	0,2706766
0,003	3	1	94	33	6	0	1	0,06	0,2598425	1	1,0638297	0	0,2932330
0,004	2	0	91	32	9	1	0,9696969	0,09	0,2601626	0,9	1,0656010	0,3367003	0,3082706
0,005	1	0	89	32	11	1	0,9696969	0,11	0,2644628	0,9166666	1,0895471	0,2754820	0,3233082
0,006	1	0	88	32	12	1	0,9696969	0,12	0,2666666	0,9230769	1,1019283	0,2525252	0,3308270
0,007	3	1	87	32	13	1	0,9696969	0,13	0,2689075	0,9285714	1,1145942	0,2331002	0,3383458
0,008	1	0	84	31	16	2	0,9393939	0,16	0,2695652	0,8888888	1,1183261	0,3787878	0,3533834
0,009	2	0	83	31	17	2	0,9393939	0,17	0,2719298	0,8947368	1,1317999	0,3565062	0,3609022
0,011	2	1	81	31	19	2	0,9393939	0,19	0,2767857	0,9047619	1,1597456	0,3189792	0,3759398
0,012	3	0	79	30	21	3	0,9090909	0,21	0,2752293	0,875	1,1507479	0,4329004	0,3834586
0,013	2	0	76	30	24	3	0,9090909	0,24	0,2830188	0,8888888	1,1961722	0,3787878	0,4060150
0,014	2	0	74	30	26	3	0,9090909	0,26	0,2884615	0,8965517	1,2285012	0,3496503	0,4210526
0,015	2	1	72	30	28	3	0,9090909	0,28	0,2941176	0,9032258	1,2626262	0,3246753	0,4360902
0,016	3	0	70	29	30	4	0,8787878	0,3	0,2929292	0,8823529	1,2554112	0,4040404	0,4436090
0,017	2	0	67	29	33	4	0,8787878	0,33	0,3020833	0,8918918	1,3116236	0,3673094	0,4661654
0,018	2	0	65	29	35	4	0,8787878	0,35	0,3085106	0,8974358	1,3519813	0,3463203	0,4812030
0,019	2	0	63	29	37	4	0,8787878	0,37	0,3152173	0,9024390	1,3949013	0,3276003	0,4962406

$AUC = 0.889$, $p < 0.000001$ **(istotne statystycznie,** $\alpha = 0.05$**)**

Rozpoznawanie bakteriemii przy użyciu wskaźnika PCT jest istotnie korzystniejsze niż losowy podział pacjentów na 2 grupy.

Wracamy więc do analizy (przycisk), by wyznaczyć optymalny punkt odcięcia. Algorytm poszukiwania optymalnego punktu odcięcia uwzględnia koszty błędnych decyzji i współczynnik chorobowości:

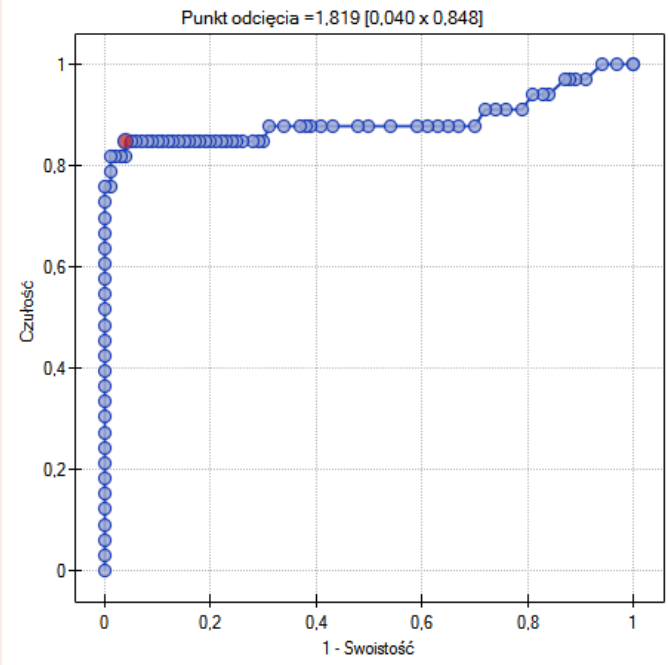
(1) koszt FN – błędna diagnoza, to koszt uznania, że pacjent nie choruje na bakteriemię, mimo że rzeczywiście jest chory (koszty decyzji fałszywie ujemnej);

(2) koszt FP – błędna diagnoza, to koszt uznania, że pacjent choruje na bakteriemię, mimo że rzeczywiście na nią nie choruje (koszty decyzji fałszywie dodatniej).

Ponieważ koszty FN są znacznie poważniejsze niż koszty FP, w polu pierwszym wpisujemy wartość większą niż w polu drugim – uznaliśmy, że będzie to wartość 5. Wartość PCT ma być wykorzystywana w badaniach przesiewowych, nie podajemy więc populacyjnego współczynnika chorobowości (współczynnika chorobowości a priori), który jest bardzo niski, ale pozostajemy przy współczynniku estymowanym z próby – postępujemy tak, by nie przesunąć punktu odcięcia wartości PCT zbyt wysoko i nie zwiększyć liczby fałszywie ujemnych wyników.

Dla punktu odcięcia	
Metoda punktu odcięcia	Styczna (indeks kosztów)
Punkt odcięcia	1.819
Koszt FN - błędna diagnoza	5
Koszt FP - błędna diagnoza	1

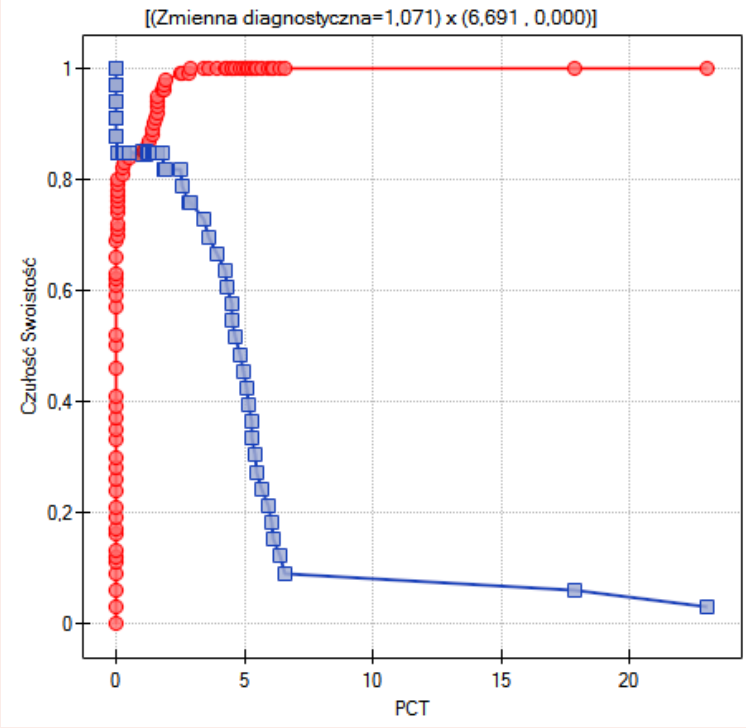
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



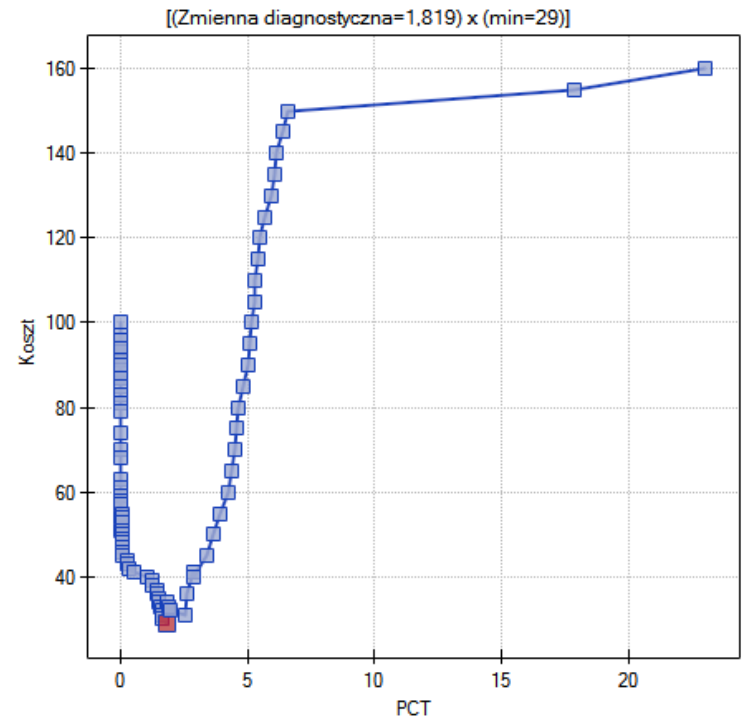
optymalny punkt odcięcia PCT = 1.819

Dla tego punktu czułość=0.85, a swoistość=0.96.

Innym sposobem wyboru punktu odcięcia jest analiza wykresu kosztów i wykresu przecięcia czułości:



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Analiza wykresu kosztów wskazuje, że minimum kosztów błędnych decyzji przypada na PCT=1.819. Natomiast wartość czułości i swoistości jest podobna dla PCT=1.071.

DLA ZAINTERESOWANYCH

Metoda stycznej (indeks kosztów) – wzór:
Za optymalny punkt odcięcia uznana zostaje taka wartość zmiennej diagnostycznej, przy której wyrażenie:

$$\text{Czułość} - m \cdot (1 - \text{Swoistość})$$

osiąga minimum (Zweig M.H., 1993[188]). Kąt nachylenia m jest określany w odniesieniu do kosztów błędnych decyzji i współczynnika chorobowości. Standardowo koszty błędnych decyzji są równe 1, a współczynnik chorobowości estymowany jest z próby. Znając a priori współczynnik chorobowości ($P_{apriori}$) i koszty błędnych decyzji, użytkownik może wpływać na wartość m, a tym samym na wyszukiwanie optymalnego punktu odcięcia.

Indeks Youdena – wzór:

$$d = \text{Czułość} + \text{Swoistość} - 1$$

Odległość do lewego górnego rogu – wzór:

$$d = \sqrt{(1 - \text{Czułość})^2 + (1 - \text{Swoistość})^2}$$

Wykres kosztów – wzór:
Wartości błędnej diagnozy wyliczane są zgodnie ze wzorem:

$$koszt = koszt_{FP} \cdot FP + koszt_{FN} \cdot FN$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Porównywanie krzywych ROC

KIEDY STOSOWAĆ:

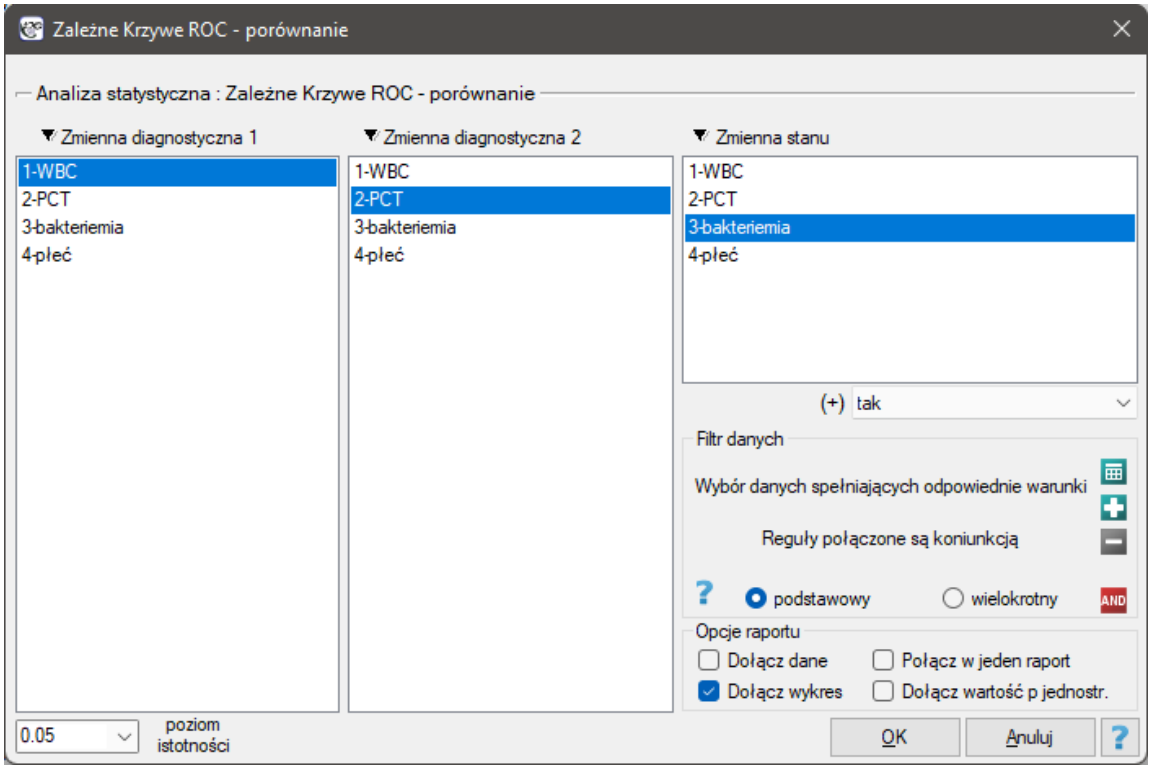
Bardzo często celem badań jest porównanie wielkości pola pod krzywą ROC (AUC_1) z polem pod inną krzywą ROC (AUC_2). Krzywa ROC o większym polu pozwala zwykle na dokładniejszą klasyfikację obiektów. Metody służące porównaniu pól zależne są od modelu badania.

WARUNKI STOSOWANIA

- **model zależny** – porównywane krzywe ROC powstają na bazie pomiarów dokonanych na tych samych obiektach,
- **model niezależny** – porównywane krzywe ROC powstają na bazie pomiarów dokonanych na różnych obiektach.

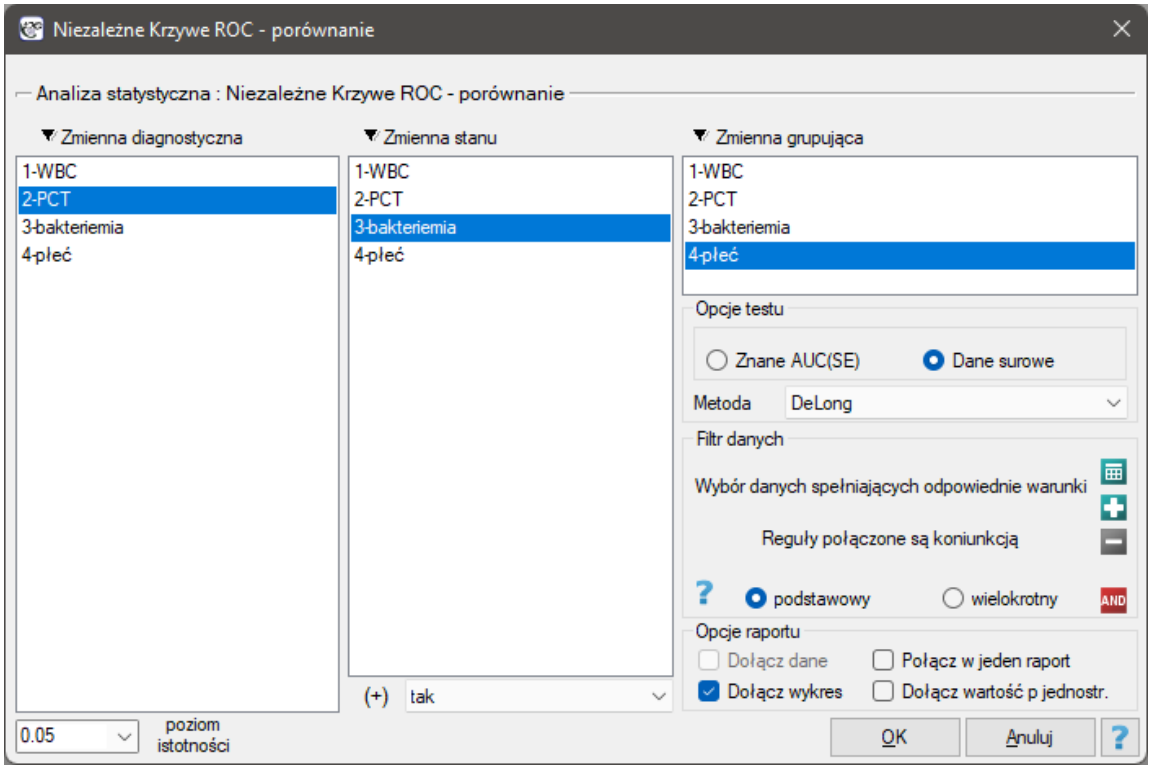
JAK USTAWIĆ ANALIZĘ

Dla modelu zależnego: Statystyka zaawansowana → Testy diagnostyczne → Zależne Krzywe ROC – porównywanie.



Dla modelu niezależnego: Statystyka zaawansowana → Testy diagnostyczne → Niezależne Krzywe ROC – porównywanie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

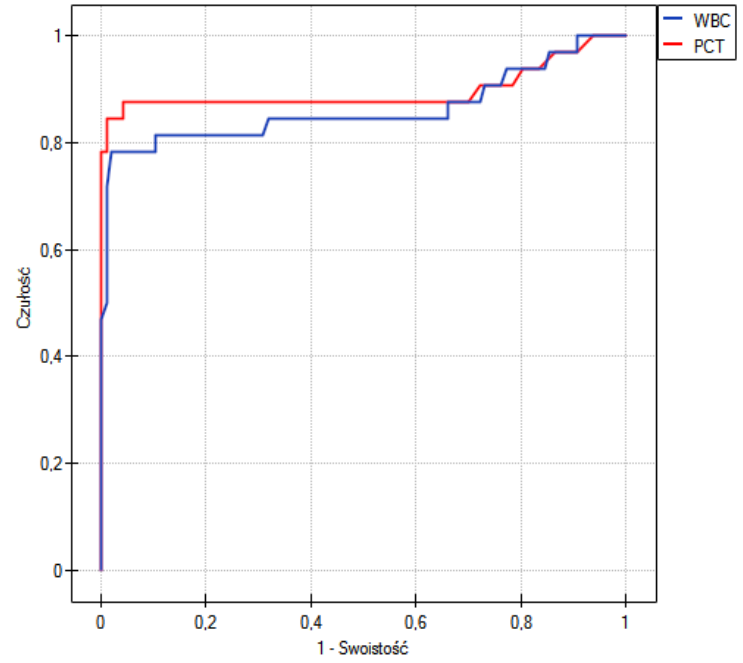
Kontynuacja przykładu 24.2 (plik bakteriemia.pqs): wykonamy 2 porównania – (1) zbudujemy 2 krzywe ROC, by porównać wartość diagnostyczną parametrów WBC i PCT (model zależny – te same dzieci); (2) zbudujemy 2 krzywe ROC, by porównać wartość diagnostyczną parametru PCT dla chłopców i dziewczynek (model niezależny – różne dzieci).

(1) WBC vs PCT (model zależny)
Zarówno parametr WBC, jak i PCT jest stymulantą (wysokie wartości tych parametrów towarzyszą bakteriemii).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zależne Krzywe ROC - porównanie	
Czas analizy	0,55 sek.
Analizowane zmienne	WBC;PCT;bakteriemia
Liczność - braki danych	7
Poziom istotności	0,05
Zmienna grupująca	bakteriemia
Liczność	129
Liczność STAN + (tak)	32
Liczność STAN - (nie)	97
Zmienna WBC	
Kierunek zmiennej diagnostycznej	stymulanta
AUC	0,86130799
SE(AUC)	0,051727687
-95% CI	0,759923577
+95% CI	0,96269238
Zmienna PCT	
Kierunek zmiennej diagnostycznej	stymulanta
AUC	0,895618557
SE(AUC)	0,049011126
-95% CI	0,79955852
+95% CI	0,991678596
DeLong's method	
AUC1-AUC2	0,034310567
SE(AUC1-AUC2)	0,022679669
-95% CI	0
+95% CI	0,078761905
Statystyka Z	1,512833666
Wartość p	0,130321915

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$AUC_{WBC} = 0.8613$, $AUC_{PCT} = 0.8956$, $p = 0.130322$ **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Nie możemy wskazać, który z parametrów – WBC czy PCT – jest lepszy w rozpoznawaniu bakteriemii.

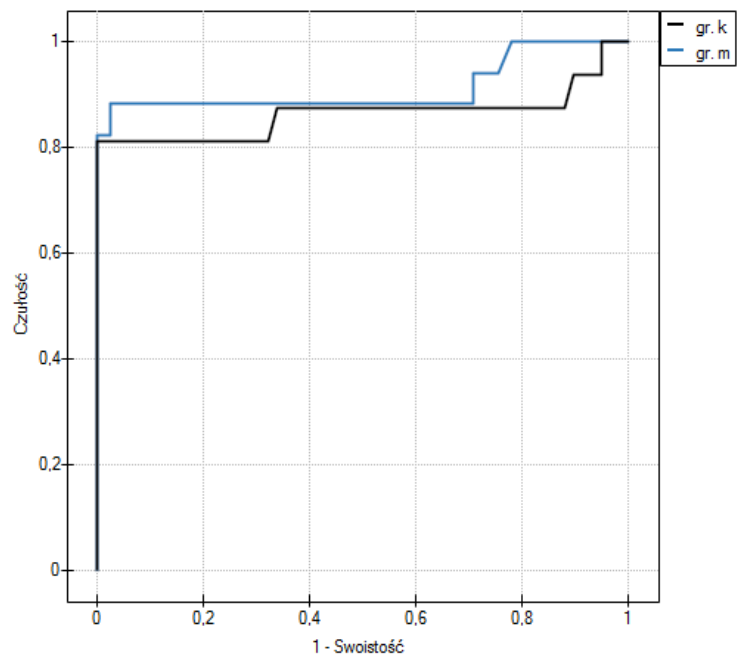
(2) PCT: dziewczynki vs chłopcy (model niezależny)

Parametr PCT jest stymulantą (jego wysokie wartości towarzyszą bakteriemii); porównujemy jego wartość diagnostyczną osobno dla dziewczynek i chłopców.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Niezależne Krzywe ROC - porównanie	
Analizowane zmienne	PCT
	bakteriemia
Liczność / niezinterpretowane	0
Liczność / braki danych	1
Poziom istotności	0.05
Zmienna grupująca	płeć
Kierunek zmiennej diagnostycznej	stymulanta
Metoda	DeLong
Nazwa grupy	0
Liczność	58
Liczność STAN + (tak)	17
Liczność STAN - (nie)	41
AUC	0.9118
SE(AUC)	0.0599
-95% CI	0.7943
+95% CI	1
Nazwa grupy	1
Liczność	75
Liczność STAN + (tak)	16
Liczność STAN - (nie)	59
AUC	0.8649
SE(AUC)	0.0792
-95% CI	0.7098
+95% CI	1
AUC1-AUC2	0.0468
SE(AUC1-AUC2)	0.0993
Statystyka Z	0.4717
Wartość p	0.6372

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$AUC_k = 0.8649$, $AUC_m = 0.9118$, $p = 0.637176$ **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Nie możemy wskazać płci, dla której parametr PCT jest lepszy w rozpoznawaniu bakteriemii.

DLA ZAINTERESOWANYCH

Hipotezy (wspólne dla obu modeli):

$$\begin{aligned} \mathcal{H}_0: &\quad AUC_1 = AUC_2, \\ \mathcal{H}_1: &\quad AUC_1 \neq AUC_2. \end{aligned}$$

Model zależny – wzór:
Statystyka testowa ma postać:

$$Z = \frac{|AUC_1 - AUC_2|}{SE_{AUC_1 - AUC_2}},$$

gdzie AUC_1, AUC_2 i błąd standardowy różnicy pól $SE_{AUC_1 - AUC_2}$ wyliczane są w oparciu o metodę nieparametryczną zaproponowaną przez **DeLong** (DeLong E.R. i inni, 1988[51]; Hanley J.A. i Hajian-Tilaki K.O., 1997[75]).

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Model niezależny – wzór:
Statystyka testowa (Hanley J.A. i McNeil M.D., 1983[77]) ma postać:

$$Z = \frac{|AUC_1 - AUC_2|}{\sqrt{SE_{AUC_1}^2 - SE_{AUC_2}^2}},$$

gdzie AUC_1, AUC_2 i błędy standardowe pól SE_{AUC_1}, SE_{AUC_2} wyliczane są w oparciu o:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- ⋆ metodę nieparametryczną **DeLong** (DeLong E.R. i inni, 1988[51]; Hanley J.A. i Hajian-Tilaki K.O., 1997[75]) – **rekomendowane**,
- ⋆ metodę nieparametryczną **Hanley-McNeil** (Hanley J.A. i McNeil M.D., 1982[76]).

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Dla obu modeli: wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

25 MODELE WIELOWYMIAROWE

Dopasowanie grup

KIEDY STOSOWAĆ:

Jeśli szacujemy efekt leczenia na podstawie **eksperymentu z pełną randomizacją**, to przypisując losowo osoby do grupy leczonej i nieleczonej, tworzymy **grupy podobne pod względem możliwych czynników zakłócających**. Podobieństwo grup wynika z samego losowego przydziału. W takich badaniach możemy badać czysty (niezależny od czynników zakłócających) wpływ metody leczenia na wynik eksperymentu. W takim przypadku inne, poza losowym, dopasowanie grup nie jest konieczne.

Możliwość błędu pojawia się, gdy różnica w wyniku leczenia między leczonymi i nieleczonymi grupami może być spowodowana nie przez samo leczenie, ale przez czynnik, który skłonił do wzięcia udziału w leczeniu. Dzieje się tak wtedy, gdy z pewnych względów randomizacja nie jest możliwa, np. jest to badanie obserwacyjne lub ze względów etycznych nie możemy przypisać leczenia dowolnie. Wówczas sztuczne dopasowanie grup może mieć zastosowanie. Na przykład, jeśli osoby, które przydzielamy do grupy leczonej, to osoby zdrowsze, a osoby, które są w grupie kontrolnej, to osoby o większym nasileniu choroby, wówczas to nie sam sposób leczenia, ale stan pacjenta przed leczeniem może mieć wpływ na wynik eksperymentu. Gdy zobaczymy taką nierównowagę grup, dobrze jest, kiedy możemy zdecydować się na randomizację – w ten sposób problem zostaje rozwiązany, gdyż losowanie osób do grup powoduje, że stają się one podobne. Można sobie jednak wyobrazić inną sytuację. Tym razem grupą, którą jesteśmy zainteresowani, nie będą osoby poddane leczeniu, lecz osoby palące, a grupą kontrolną osoby niepalące, a analizy będą miały na celu wykazanie niekorzystnego wpływu palenia na występowanie raka płuc. Wówczas, chcąc sprawdzić czy palenie rzeczywiście wpływa na zwiększenie ryzyka zachorowania, nieetyczne byłoby wykonanie badania z pełną randomizacją, ponieważ oznaczałoby to, że losowo wybrane do grupy ryzyka osoby zmusimy do palenia. Rozwiązaniem tej sytuacji jest ustalenie grupy narażonej, czyli wylosowanie pewnej liczby osób spośród osób, które już palą, a następnie dobór grupy kontrolnej złożonej z niepalących. Grupa kontrolna powinna być dobrana, ponieważ zostawiając dobór przypadkowi, możemy uzyskać grupę niepalącą, która jest młodsza od palących tylko ze względu na fakt, że w naszym kraju palenie staje się mniej modne, a więc automatycznie wśród osób niepalących jest wiele osób młodych. Grupa kontrolna powinna być wylosowana z osób niepalących, ale tak, by była **jak najbardziej podobna do grupy leczonej**. W ten sposób przybliżamy się do badania czystego (niezależnego od wybranych czynników zakłócających, takich jak chociażby wiek) wpływu palenia/niepalenia na wynik eksperymentu, czyli w tym przypadku wystąpienie nowotworu płuc. **Taki dobór możemy wykonać właśnie poprzez zaproponowane w programie dopasowanie**.

Jedną z głównych zalet kontrolowanego przez badacza dopasowania jest upodobnienie się grupy kontrolnej do grupy leczonej, ale jest to też największa wada tej metody. Jest to zaleta, ponieważ nasze badanie coraz bardziej przypomina badanie randomizowane – w badaniu randomizowanym grupa badana i kontrolna są podobne pod względem niemalże wszystkich cech, również tych, których nie badamy (losowy podział zapewnia nam to podobieństwo). W wyniku zastosowania dopasowania kontrolowanego przez badacza grupa leczona i kontrolna stają się podobne pod względem tylko wybranych cech.

WARUNKI STOSOWANIA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- cechy, względem których dokonujemy dopasowania, powinny być zmierzone dla wszystkich obiektów w obu grupach,
- grupa, z której losujemy dopasowanie (zwykle grupa kontrolna), powinna być wystarczająco liczna i zawierać obiekty rzeczywiście podobne do grupy leczonej.

DEFINICJE MIAR

Dwa pierwsze wymienione niżej sposoby opierają się na dopasowaniu grup poprzez **dopasowanie wyników skłonności** *(ang. Propensity Score Matching, PSM)*. Ten rodzaj dopasowania został zaproponowany przez Rosenbauma i Rubina[141]. W praktyce jest to technika doboru grupy kontrolnej (osób nieleczonych lub leczonych minimalnie/standardowo) do grupy leczonej na podstawie prawdopodobieństwa opisującego skłonność badanych do przypisania leczenia w zależności od obserwowanych zmiennych towarzyszących. Wynik prawdopodobieństwa opisującego skłonność, z angielskiego nazywany *Propensity Score*, jest wynikiem równoważącym, dlatego w wyniku doboru grupy kontrolnej do grupy leczonej rozkład zmierzonych zmiennych towarzyszących staje się bardziej podobny między osobami leczonymi i nieleczonymi. Trzecia metoda nie wyznacza prawdopodobieństwa dla każdej osoby, ale wyznacza macierz odległości/niepodobieństwa, która wskazuje obiekty najbliższe/najbardziej podobne pod względem wielu wybranych cech.

Zadane prawdopodobieństwo – czyli podana dla każdej badanej osoby wartość z przedziału od 0 do 1, określająca prawdopodobieństwo przynależenia do grupy leczonej, czyli Propensity Score. Takie prawdopodobieństwo może zostać wyznaczone wcześniej różnymi metodami, np. w modelu regresji logistycznej, poprzez sieci neuronowe lub wiele innych metod. Jeśli osoba z grupy, z której losujemy kontrolę, uzyska Propensity Score podobne do tego, jaki uzyskała osoba z grupy leczonej, wówczas może ona wejść do analizy, ponieważ obie te osoby są podobne pod względem cech, jakie były rozważane przy wyznaczaniu Propensity Score.

Wyznaczone z modelu regresji logistycznej – ponieważ regresja logistyczna to najczęściej stosowana metoda doboru, program PQStat daje możliwość automatycznego wyznaczenia w oknie analizy wartości Propensity Score w oparciu o tę metodę. Dopasowanie przebiega dalej przy wykorzystaniu uzyskanego w ten sposób Propensity Score.

Macierz podobieństwa/odległości – na podstawie tej opcji nie jest wyznaczana wartość Propensity Score, ale budowana jest macierz wskazująca odległość każdej osoby z grupy leczonej do osoby z grupy kontrolnej. Użytkownik może zadać warunki brzegowe, np. może wskazać, że osoba dobrana do danej osoby z grupy badanej nie może się od niej różnić wiekiem bardziej niż o 3 lata i musi być tej samej płci. Odległości w budowanej macierzy wyznaczone są w oparciu o dowolną metrykę lub sposób opisujący niepodobieństwo. Ten sposób doboru grupy kontrolnej do leczonej jest bardzo elastyczny – oprócz dowolnego wyboru sposobu wyznaczania odległości/niepodobieństwa, w wielu metrykach pozwala na wskazywanie wag określających to, jak ważne są dla badacza poszczególne zmienne. Jednak w przypadku wyboru macierzy odległości/niepodobieństwa zalecana jest duża ostrożność – wiele cech i wiele sposobów wyznaczania odległości wymaga wcześniejszego standaryzowania lub normalizowania danych, ponadto wybór odwrotności odległości lub podobieństwa (zamiast niepodobieństwa) może skutkować wyszukiwaniem najbardziej odległych i niepodobnych obiektów, podczas gdy standardowo stosujemy te metody do wyszukiwania obiektów podobnych. Jeśli badacz nie ma określonych powodów zmiany metryki, standardowo zalecane jest korzystanie z odległości statystycznej, czyli metryki Mahalanobisa – jest ona najbardziej uniwersalna, nie wymaga wcześniejszej standaryzacji danych i jest odporna na skorelowanie zmiennych. Dokładniejszy opis dostępnych w programie odległości i miar niepodobień-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

stwa/podobieństwa oraz sposób interpretacji uzyskanych wyników można znaleźć w dziale Macierz podobieństwa.

W praktyce istnieje wiele metod wskazujących, jak blisko znajdują się porównywane obiekty. W programie zaproponowane są dwie:

Metoda najbliższego sąsiada – jest standardowym sposobem doboru obiektów, nie tylko takich o podobnym Propensity Score, ale również takich, których odległość/niepodobieństwo znajdujące się w macierzy jest najmniejsze.

Metoda najbliższego sąsiada, bliższego niż... – działa w ten sam sposób co metoda najbliższego sąsiada, z tą różnicą, że dopasowane mogą zostać jedynie obiekty znajdujące się odpowiednio blisko. Granicę tej bliskości wyznaczamy, podając wielkość opisującą próg, za którym znajdują się już obiekty tak niepodobne do badanych, że nie chcemy dać im szansy na dołączenie do nowo budowanej grupy kontrolnej. W przypadku gdy analiza opiera się na Propensity Score lub na macierzy określanej przez niepodobieństwo, najbardziej niepodobne obiekty to te odległe o 1, a najbardziej podobne to te odległe o 0 – wybierając więc tę metodę, należy podać wartość bliższą 0, gdy dobieramy bardziej restrykcyjnie, lub bliższą 1, gdy próg ten będzie umieszczony dalej. Gdy zamiast niepodobieństwa w macierzy wyznaczamy odległości, wówczas wielkość minimalna to wciąż 0, ale wielkość maksymalna nie jest z góry określona.

Dopasować możemy bez zwracania obiektów już wylosowanych lub ze zwracaniem tych obiektów ponownie do grupy, z której losujemy.

Dopasowanie bez zwracania – w przypadku stosowania dopasowywania bez zwracania, gdy nieleczona osoba została wybrana do dopasowania do danej leczonej osoby, ta nieleczona osoba nie jest już dostępna do rozważenia jako potencjalne dopasowanie dla kolejnych osób leczonych. W rezultacie każda nieleczona osoba jest zawarta w co najwyżej jednym dopasowanym zestawie.

Dopasowanie ze zwracaniem – dopasowanie ze zwracaniem pozwala na uwzględnienie danej nieleczonej osoby więcej niż raz w jednym dopasowanym zestawie. Kiedy stosuje się dopasowanie ze zwracaniem, dalsze analizy, a w szczególności oszacowanie wariancji, muszą uwzględniać fakt, że ta sama nieleczona osoba może znajdować się w wielu dopasowanych zestawach.

Uwaga! W przypadku gdy nie da się jednoznacznie dobrać osoby nieleczonej do leczonej, ze względu na to, że w grupie, z której wybieramy, mamy więcej osób tak samo dobrze pasujących do osoby leczonej, wówczas dołączona zostaje jedna z tych osób wybrana w sposób losowy. Dla wznowionej analizy domyślnie ustawiony jest stały *seed*, więc wyniki powtórzonego losowania będą te same, jednak gdy analizę wykonamy na nowo, *seed* zostaje zmieniony i wynik losowania może być inny.

W przypadku gdy nie da się dobrać osoby nieleczonej do leczonej, ze względu na to, że w grupie, z której wybieramy, nie ma już osób do dołączenia (np. osoby pasujące zostały już dołączone do innych osób leczonych lub zbiór, z którego wybieramy, nie ma osób podobnych), wówczas osoba ta pozostaje bez pary.

Najczęściej dokonuje się dopasowania **1:1**, tzn. dla jednej osoby leczonej dobiera się jedną osobę nieleczoną. Jednak jeśli oryginalna grupa kontrolna, z której dokonujemy losowania, jest wystarczająco duża i potrzebujemy wylosować więcej osób, można wybrać dopasowanie **1:k**, gdzie k wskazuje liczbę osób, która powinna zostać dopasowana do każdej osoby leczonej.

Ocena dopasowania

Po dopasowaniu grupy kontrolnej do grupy leczonej wyniki takiego dopasowania możemy zwrócić do arkusza, tzn. uzyskać nową grupę kontrolną. Nie należy jednak zakładać, że stosując dopasowanie, zawsze uzyskamy satysfakcjonujące wyniki – w wielu sytuacjach grupa, z której losujemy, nie posiada wystarczającej liczby takich obiektów, które są wystarczająco podobne do grupy leczonej. Dlatego za-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

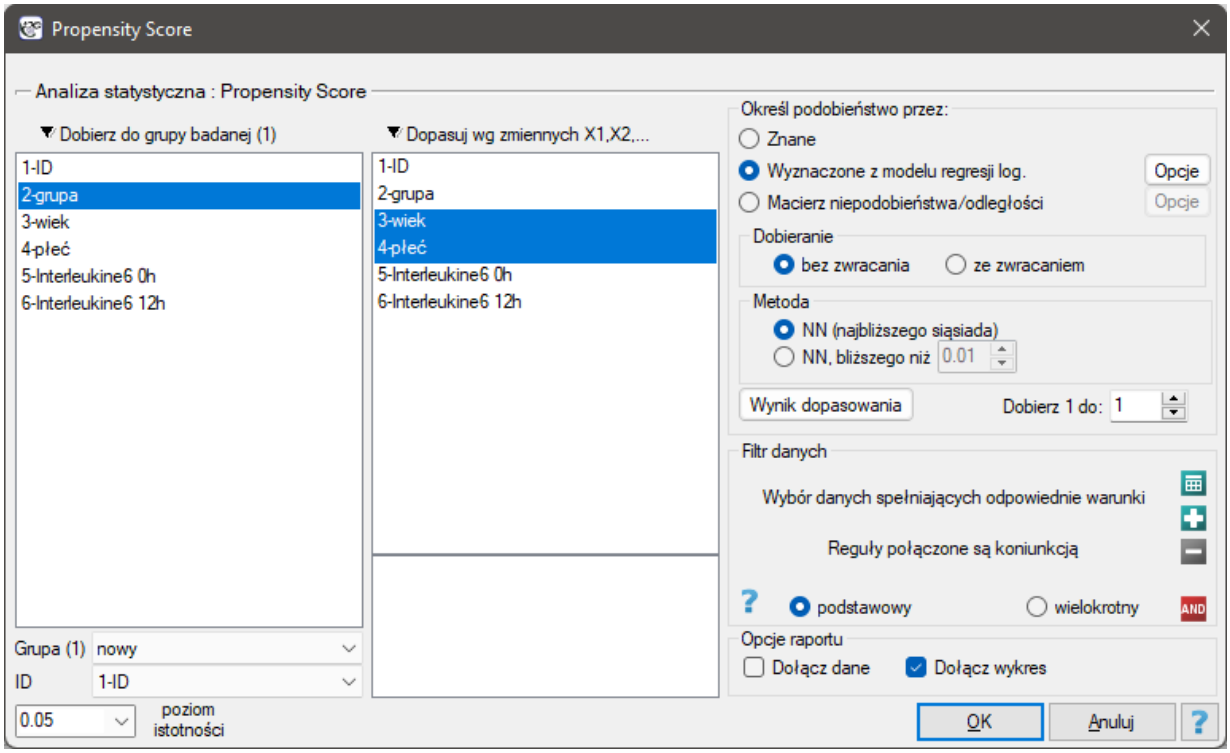
wsze wykonane dopasowanie należy ocenić. Istnieje wiele metod oceny dopasowania grup. W programie wykorzystano metody opierające się na standaryzowanej różnicy grup, szerzej opisywane m.in. w pracach P.C. Austina[13][14]. Takie podejście pozwala na porównanie względnej równowagi zmiennych mierzonych w różnych jednostkach, a na jego wynik nie ma wpływu wielkość próby. Zrezygnowano z oszacowania zgodności przy pomocy testów statystycznych, gdyż dobrana grupa kontrolna jest zwykle dużo mniejsza niż oryginalna grupa kontrolna, przez co uzyskiwane wartości p testów porównujących grupę badaną do mniejszej grupy kontrolnej częściej pozostają przy założeniu hipotezy zerowej, a więc nie wykazują istotnych różnic ze względu na zmniejszoną liczność.

Uwaga! Chociaż nie ma powszechnie uzgodnionego kryterium określającego, jaki próg znormalizowanej różnicy można zastosować do wskazania istotnej nierównowagi, wskazówką może być standaryzowana różnica mniejsza niż 0.1 (zarówno w ocenie średnich, jak i częstości)[129]. Dlatego, by uznać, że grupy są dobrze dobrane, powinniśmy obserwować standaryzowane różnice położone blisko wartości 0, a najlepiej, by nie wychodziły poza przedział od -0.1 do 0.1. Graficznie wyniki te przedstawiamy na wykresie punktowym. Ujemne różnice świadczą o niższych średnich/częstościach w grupie leczonej, dodatnie – w grupie kontrolnej.

Uzyskane w raportach dopasowanie 1:1 oznacza podsumowanie dotyczące grupy badanej i odpowiadającej jej grupy kontrolnej uzyskanej w pierwszym dopasowaniu, dopasowanie 1:2 oznacza podsumowanie dotyczące grupy badanej i odpowiadającej jej grupy kontrolnej uzyskanej w pierwszym + drugim dopasowaniu (czyli nie dotyczy grupy badanej i odpowiadającej jej grupy kontrolnej uzyskanej tylko w drugim dopasowaniu) itd.

JAK USTAWIĆ ANALIZĘ

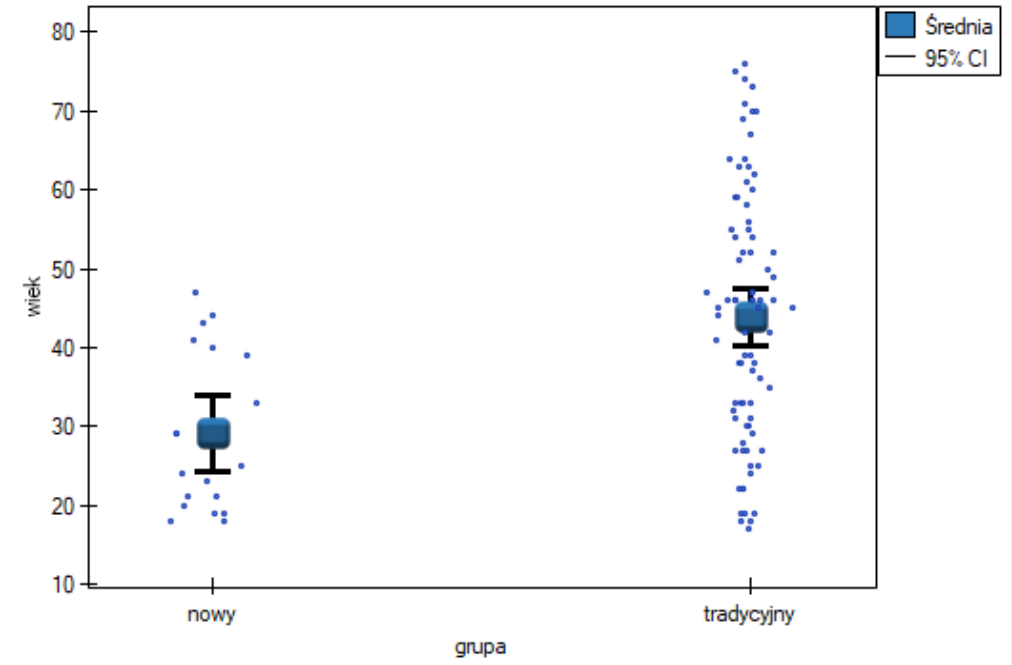
Statystyki zaawansowane → Modele wielowymiarowe → Dopasowanie grup.



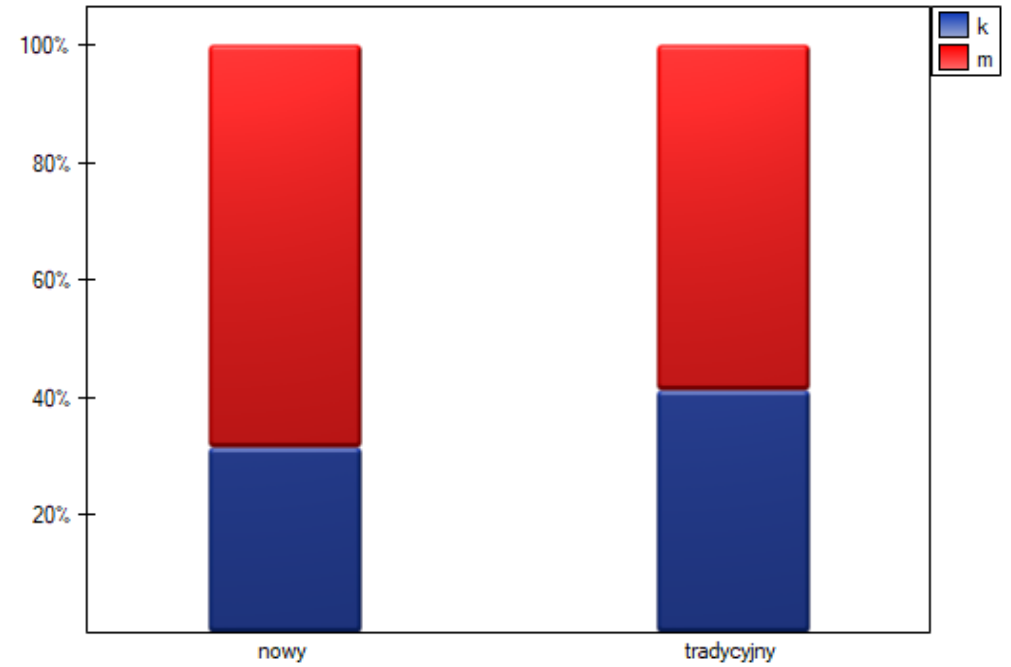
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

WYKONANIE I INTERPRETACJA

PRZYKŁAD 25.1. (plik dopasowanie.pqs)
Chcemy porównać dwa sposoby leczenia pacjentów po wypadkach: sposób tradycyjny i nowy sposób leczenia. Poprawne działanie obu sposobów leczenia powinno być obserwowane w obniżającym się poziomie wybranych cytokin. By porównać skuteczność tych dwóch sposobów leczenia, obydwa powinny zostać przeprowadzone na pacjentach, którzy są dość podobni – wtedy będziemy mieli pewność, że ewentualne różnice w skuteczności tych metod będą wynikały z samego oddziaływania leczenia, a nie z innych różnic między pacjentami przydzielonymi do różnych grup. Badanie jest przeprowadzone a posteriori, to znaczy bazuje na danych zebranych od pacjentów z historii leczenia. Dlatego badacze nie mieli wpływu na przypisanie pacjentów do grupy leczonej nowym lekiem i grupy leczonej tradycyjnie. Zauważono, że tradycyjny sposób leczenia był przepisywany głównie pacjentom starszym, podczas gdy nowy sposób leczenia – pacjentom w młodszym wieku, u których łatwiej jest obniżać poziom cytokin. Grupy były dość podobne co do struktury płci, ale nie identyczne.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Gdyby przeprowadzono planowane badanie na tak wybranych grupach pacjentów, to nowy sposób miałby łatwiejsze zadanie do wykonania, gdyż młodsze organizmy lepiej mogłyby reagować na leczenie. Warunki eksperymentu nie byłyby równe dla obydwu sposobów, co mogłoby zafałszować wyniki analiz i wyciągane wnioski. Dlatego zdecydowano się dobrać grupę leczoną tradycyjnie tak, by była podobna do grupy badanej leczonej nowym sposobem.

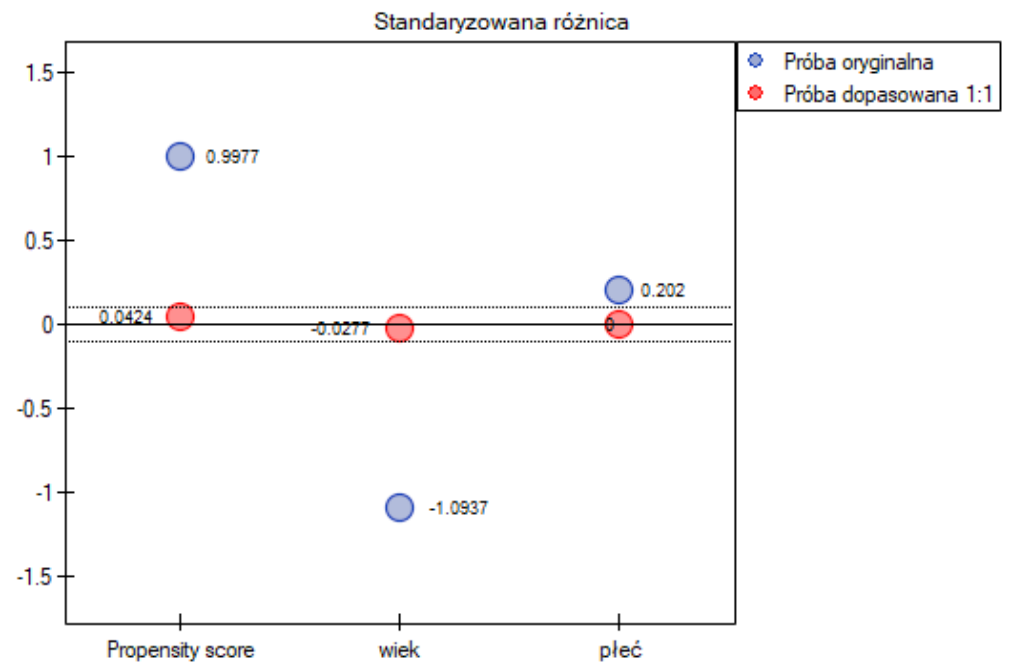
Dopasowania planujemy dokonać względem dwóch cech, tzn. względem wieku i płci. Grupa leczona tradycyjnie jest większa (80 osób) od grupy leczonej nowym lekiem (19 osób), dlatego jest duża szansa na to, że uda się dobrać grupy tak, by były podobne. Losowego doboru dokonujemy poprzez algorytm modelu regresji logistycznej zaszyty w PSM. Pamiętamy, by płeć była zakodowana liczbowo, gdyż w analizie regresji logistycznej biorą udział jedynie wartości liczbowe. Jako metodę wybieramy najbliższe sąsiedztwo. Chcemy, by ta sama osoba nie mogła zostać wybrana dwukrotnie, więc wybieramy losowanie bez zwracania. Spróbujemy dopasowania 1:1, czyli dla każdej osoby leczonej nowym lekiem dopasujemy jedną osobę leczoną tradycyjnie. Pamiętajmy przy tym, że dobór jest losowy, a więc zależy od losowej wartości *seed* ustawionej przez nasz komputer, więc losowanie przeprowadzone przez czytelnika może się różnić od wartości przedstawionych tutaj.

Podsumowanie doboru obejrzymy w tabelach i na wykresach.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Próba oryginalna					
	badana	kontrolna	różnica	SE różnicy	std. różnica
Propensity	0.3073	0.1645	0.1428	0.1431	0.9977
wiek	29.1053	43.9125	-14.8072	13.5386	-1.0937
płeć	0.6842	0.5875	0.0967	0.4788	0.202
Dopasowanie 1:1					
Propensity	0.3073	0.3015	0.0057	0.1353	0.0424
wiek	29.1053	29.3684	-0.2632	9.493	-0.0277
płeć	0.6842	0.6842	0	0.4648	0

W próbie oryginalnej średnia wieku jest o ponad 14 lat wyższa u pacjentów leczonych tradycyjnie (różnica między średnimi wynosi 14.8072), natomiast struktura płci różni się o niecałe 10% (0.0967). Znacznie mniejsze różnice obserwujemy pomiędzy pacjentami leczonymi nowym sposobem i dopasowanymi do nich osobami leczonymi tradycyjnie. Najwięcej informacji o jakości dopasowania uzyskamy na podstawie różnic standaryzowanych (ostatnia kolumna tabeli i wykres).



standaryzowana różnica Propensity Score = 0.0424, standaryzowana różnica wieku = -0.0277, standaryzowana różnica odsetka płci = 0

Po dopasowaniu grupy stają się dobrze zbalansowane – wszystkie standaryzowane różnice mieszczą się w przedziale od -0.1 do 0.1.

Linia na poziomie 0 oznacza równowagę grup (różnicę między grupami równą 0). Gdy grupy znajdują się w równowadze co do zadanych cech, wówczas wszystkie punkty na wykresie znajdują się blisko tej linii, tzn. w okolicach przedziału od -0.1 do 0.1. W przypadku próby oryginalnej (kolor niebieski) widzimy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

znaczne oddalenie Propensity Score – jak wiemy, ten brak dopasowania jest skutkiem głównie niedopasowania wieku (jego standaryzowana różnica znajduje się w dużej odległości od 0), a w mniejszym stopniu niedopasowania płci. Dokonując dopasowania, uzyskaliśmy grupy bardziej podobne do siebie (kolor czerwony na wykresie). Wiek obu grup jest już podobny – grupa leczona tradycyjnie różni się od grupy leczonej nowym sposobem średnio o niecały rok (różnica między średnimi przedstawiona w tabeli to 0.2632). W przypadku płci dopasowanie jest idealne, tzn. odsetek kobiet i mężczyzn jest taki sam w obu grupach. Tak przygotowane dane możemy zwrócić do arkusza i poddać planowanym przez siebie analizom.

Przyglądając się uzyskanemu przed chwilą podsumowaniu, można zauważyć, że mimo dobrego zbalansowania grup i dobrania wielu osób idealnie, znajdują się pojedyncze osoby, które nie są do siebie tak podobne, jak moglibyśmy oczekiwać.

badana				kontrolna 1			
ID	Propensity	wiek	płeć	ID	Propensity	wiek	płeć
1	0.3671	20	0	80	0.3566	27	1
2	0.1745	33	0	5	0.1745	33	0
3	0.1575	41	1	68	0.1575	41	1
4	0.1681	40	1	7	0.1745	33	0
5	0.4689	21	1	56	0.4496	22	1
6	0.3853	19	0	3	0.3853	19	0
7	0.4689	21	1	11	0.4496	22	1
8	0.4116	24	1	66	0.4227	17	0
9	0.527	18	1	24	0.527	18	1
10	0.3929	25	1	46	0.3929	25	1
11	0.2239	29	0	59	0.2376	28	0
12	0.1792	39	1	27	0.1909	38	1
13	0.3853	19	0	44	0.3566	27	1
14	0.105	47	1	65	0.105	47	1
15	0.129	44	1	79	0.129	44	1
16	0.138	43	1	10	0.1435	36	0
17	0.4305	23	1	16	0.4496	22	1
18	0.3218	29	1	33	0.3218	29	1
19	0.4038	18	0	71	0.3051	30	1

Czasami oprócz uzyskania grup dobrze zbalansowanych, badaczom zależy na dokładnym określeniu sposobu doboru poszczególnych osób, tzn. uzyskaniu większego wpływu na podobieństwo obiektów co do wartości Propensity Score lub na podobieństwo obiektów co do wartości konkretnych cech. Wówczas, jeśli grupa, z której losujemy, jest wystarczająco liczna, analiza może przynieść korzystniejsze z punktu widzenia badacza efekty, ale gdy w grupie, z której losujemy, zabraknie obiektów spełniających nasze kryteria, wówczas dla części osób nie uda się znaleźć dopasowania spełniającego nasze warunki.

Załóżmy, że chcielibyśmy uzyskać takie grupy, których Propensity Score (tzn. skłonność do wzięcia udziału w badaniu) różni się nie więcej niż zadana wartość. Jak ustalić tę wartość? Można zerknąć na raport z wcześniejszej analizy, gdzie podana jest najmniejsza i największa odległość między losowanymi obiektami.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

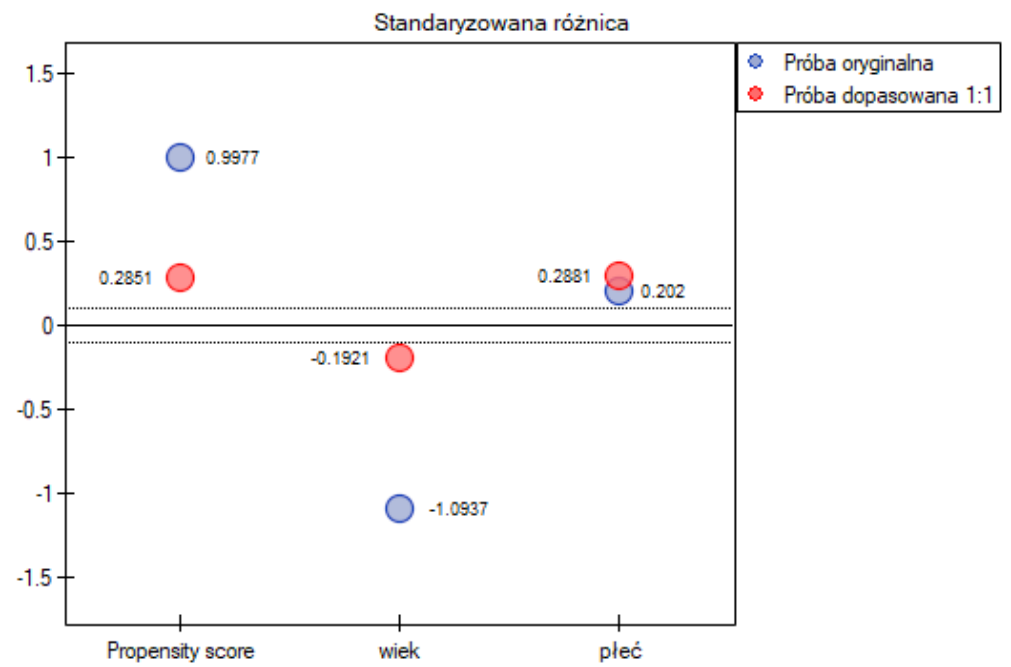
Minimum odległości	0
Maksimum odległości	0.5183

W naszym przypadku obiekty najbliższe sobie różnią się o min=0, a najdalsze o max=0.5183. Spróbujemy więc sprawdzić, jaki dobór uzyskamy, gdy będziemy dopasowywać do osób leczonych nową metodą takie osoby leczone tradycyjnie, których Propensity Score będzie bardzo bliskie, np. mniejsze od 0.01.

badana				kontrolna 1			
ID	Propensity	wiek	płeć	ID	Propensity	wiek	płeć
1	0.3671	20	0	brak dop.			
2	0.1745	33	0	7	0.1745	33	0
3	0.1575	41	1	68	0.1575	41	1
4	0.1681	40	1	5	0.1745	33	0
5	0.4689	21	1	brak dop.			
6	0.3853	19	0	3	0.3853	19	0
7	0.4689	21	1	brak dop.			
8	0.4116	24	1	brak dop.			
9	0.527	18	1	14	0.527	18	1
10	0.3929	25	1	46	0.3929	25	1
11	0.2239	29	0	brak dop.			
12	0.1792	39	1	brak dop.			
13	0.3853	19	0	brak dop.			
14	0.105	47	1	65	0.105	47	1
15	0.129	44	1	79	0.129	44	1
16	0.138	43	1	10	0.1435	36	0
17	0.4305	23	1	66	0.4227	17	0
18	0.3218	29	1	33	0.3218	29	1
19	0.4038	18	0	brak dop.			

Widzimy, że tym razem nie udało się dobrać całej grupy. Porównując Propensity Score dla poszczególnych par (leczonych nowym lekiem i leczonych tradycyjnie), widzimy, że różnice są naprawdę niewielkie. Jednak ze względu na to, że dobrana grupa jest znacznie mniejsza, podsumowując cały proces doboru, musimy zauważyć, że zarówno Propensity Score, wiek, jak i płeć nie znalazły się wystarczająco blisko linii na poziomie 0. Nasza chęć poprawy sytuacji nie doprowadziła do pożądanego skutku, a uzyskane grupy nie są wystarczająco dobrze zbalansowane.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



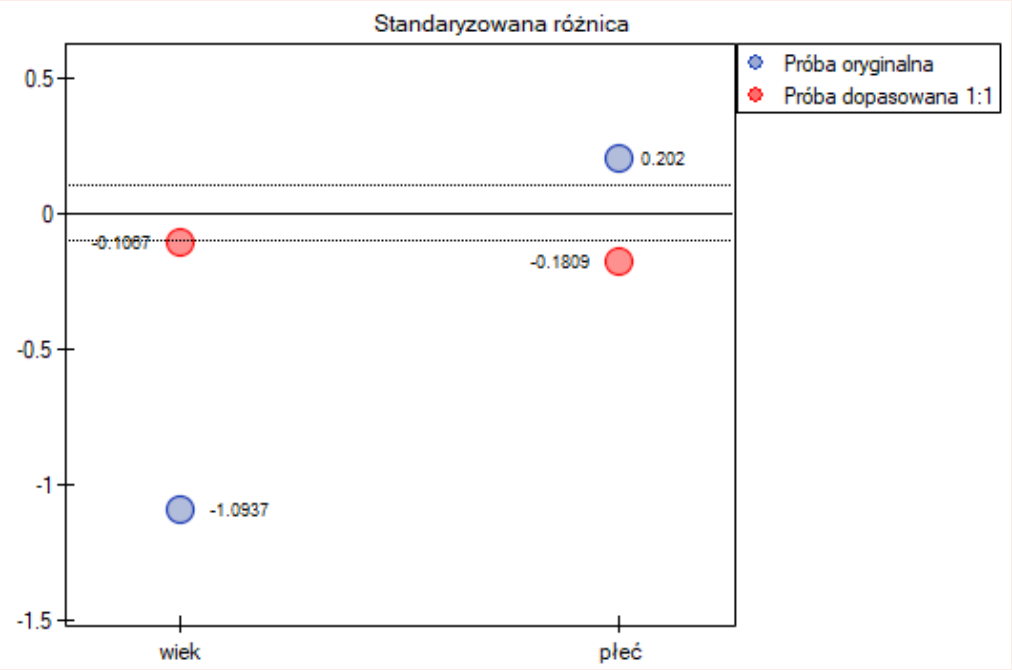
Wariant 2: macierz odległości Mahalanobisa

Załóżmy, że chcielibyśmy uzyskać takie pary (osób leczonych nową metodą i osób leczonych tradycyjnie), które są tej samej płci i których wiek nie różni się więcej niż 3 lata. W losowaniu opartym o Propensity Score nie mieliśmy tego typu możliwości wpływania na zakres zgodności każdej ze zmiennych. Do tego celu skorzystamy z innej metody, nieopartej na Propensity Score, ale bazującej na macierzach odległości/niepodobieństwa. Po wybraniu przycisku Opcje wybieramy zaproponowaną macierz odległości statystycznej Mahalanobisa i ustawiamy dopasowanie sąsiedztwa na maksymalną odległość równą 3 dla wieku i równą 0 dla płci. W efekcie dla dwóch osób nie udało się znaleźć dopasowania, ale pozostałe dopasowania spełniają zadane kryteria.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

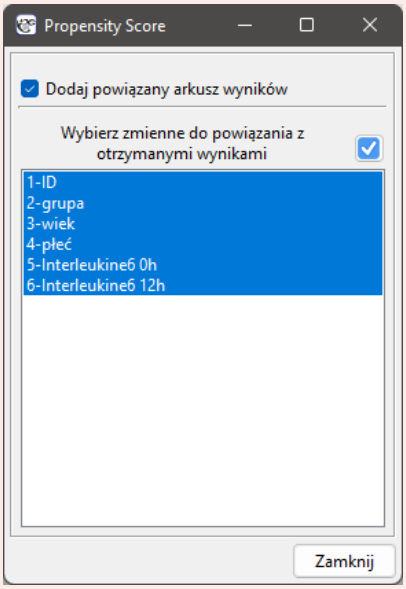
badana				kontrolna 1			
ID	Propensity	wiek	płeć	ID	Propensity	wiek	płeć
1		20	0	3		19	0
2		33	0	7		33	0
3		41	1	68		41	1
4		40	1	53		38	1
5		21	1	11		22	1
6		19	0	66		17	0
7		21	1	16		22	1
8		24	1	46		25	1
9		18	1	24		18	1
10		25	1	44		27	1
11		29	0	59		28	0
12		39	1	38		38	1
13		19	0	brak dop.			
14		47	1	65		47	1
15		44	1	43		45	1
16		43	1	77		42	1
17		23	1	56		22	1
18		29	1	33		29	1
19		18	0	brak dop.			

Podsumowując całościowo przeprowadzone losowanie, musimy zauważyć, że mimo iż spełnia ono nasze założenia, to uzyskane grupy nie są tak dobrze zbalansowane, jak to było w naszym pierwszym losowaniu przeprowadzonym w oparciu o Propensity Score. Punkty oznaczone kolorem czerwonym, przedstawiające jakość dopasowania pod względem wieku i jakość dopasowania pod względem płci, odbiegają nieco od linii identyczności ustalonej na poziomie 0, co oznacza, że średnia różnica wieku i struktury płci jest obecnie większa niż w pierwszym dopasowaniu.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

To od badacza zależy, który sposób przygotowania danych będzie dla niego bardziej korzystny. Ostatecznie, gdy decyzja zostanie podjęta, dane można zwrócić do nowego arkusza. By tego dokonać, wracamy do wybranego przez siebie raportu i w drzewie projektu pod prawym przyciskiem wybieramy menu Powtórz analizę. W oknie tej samej analizy wskazujemy przycisk Wynik dopasowania i określamy, które jeszcze zmienne będą zwrócone do nowego arkusza.



W efekcie otrzymamy nowy arkusz danych, w którym obok siebie znajdą się dane dotyczące osób leczonych nowym sposobem oraz dopasowanych do nich osób leczonych tradycyjnie.

DLA ZAINTERESOWANYCH

Wzory

Dla porównania zmiennych ciągłych wyznaczamy standaryzowaną różnicę średnich:

$$d = \frac{(\bar{x}_{leczona} - \bar{x}_{kontrolna})}{\sqrt{\frac{sd^2_{leczona} + sd^2_{kontrolna}}{2}}}$$

gdzie:

$\bar{x}_{leczona}$, $\bar{x}_{kontrolna}$ – średnia wartość zmiennej w grupie leczonej i średnia wartość zmiennej w grupie kontrolnej,

$sd^2_{leczona}$, $sd^2_{kontrolna}$ – wariancja w grupie leczonej i wariancja w grupie kontrolnej.

Dla porównania zmiennych binarnych (o dwóch kategoriach, zwykle 0 i 1) wyznaczamy standaryzowaną różnicę częstości:

$$d = \frac{(\hat{p}_{leczona} - \hat{p}_{kontrolna})}{\sqrt{\frac{\hat{p}_{leczona}(1-\hat{p}_{leczona}) + \hat{p}_{kontrolna}(1-\hat{p}_{kontrolna})}{2}}}$$

gdzie:

$\hat{p}_{leczona}$, $\hat{p}_{kontrolna}$ – częstość wartości opisanej jako 1 w grupie leczonej i częstość wartości opisanej jako 1 w grupie kontrolnej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Procedura: zmienne o wielu kategoriach powinniśmy rozbić w analizie regresji logistycznej na zmienne fikcyjne o dwóch kategoriach i sprawdzając dopasowanie obu grup, wyznaczać dla nich standaryzowaną różnicę częstości.

Wielowymiarowe modele regresji dają możliwość badania wpływu wielu zmiennych niezależnych (wielu czynników) oraz ich interakcji na jedną zmienną zależną. Poprzez modele wielowymiarowe możliwe jest również budowanie jednocześnie wielu uproszczonych modeli – modeli jednowymiarowych (jednoczynnikowych). Informacja o tym, jaki model chcemy budować (wieloczynnikowy czy jednoczynnikowy) widoczna jest w oknie wybranej analizy. Przy jednoczesnym wyborze wielu zmiennych niezależnych w oknie analizy możliwe jest dokonanie wyboru modelu.

Przygotowanie zmiennych do analizy w modelach wielowymiarowych

Kodowanie zmiennych w modelach wielowymiarowych

KIEDY STOSOWAĆ:

Problemem w przygotowaniu danych do analizy wielowymiarowej jest odpowiednie zakodowanie zmiennych nominalnych i porządkowych. Jest to ważny element przygotowania danych do analizy, gdyż ma zasadniczy wpływ na interpretację współczynników modelu. Zmienne nominalne lub porządkowe dzielą analizowane obiekty na dwie lub więcej kategorii, przy czym zmienne dychotomiczne (o dwóch kategoriach, $k = 2$) wystarczy tylko odpowiednio zakodować, a zmienne o wielu kategoriach ($k > 2$) rozbić na zmienne fikcyjne *(ang. dummy variable)* o dwóch kategoriach oraz zakodować.

JAK USTAWIĆ ANALIZĘ

$k = 2$ Jeśli zmienna jest dychotomiczna, badacz sam decyduje o sposobie, w jaki wprowadzi dane ją reprezentujące, może więc wprowadzić dowolne kody liczbowe, np. 0 i 1. W programie można zmienić własne kodowanie na kodowanie efektu, zaznaczając tę opcję w oknie wybranej analizy wielowymiarowej. Kodowanie takie powoduje zastąpienie mniejszej wartości wartością -1, a wartości większej wartością 1.

$k > 2$ Jeśli zmienna ma wiele kategorii, to w oknie wybranej analizy wielowymiarowej wybieramy przycisk Zmienne fikcyjne i ustawiamy kategorię referencyjną/bazową dla tych zmiennych, które chcemy rozbić na zmienne fikcyjne. Zmienne te będą zakodowane zero-jedynkowo, chyba że w oknie analizy zostanie wybrana opcja kodowanie efektu – wówczas kodowane będą jako -1, 0 i 1.

DEFINICJE MIAR

Kodowanie zero-jedynkowe (dummy coding) jest wykorzystywane, by przy pomocy modeli wielowymiarowych odpowiedzieć na pytanie: jak wyniki (Y), w każdej analizowanej kategorii, różnią się od wyników kategorii referencyjnej. Kodowanie to polega na przypisaniu wartości 0 lub 1 do każdej kategorii danej zmiennej. Kategoria zakodowana jako 0 jest wówczas **kategorią referencyjną** (reference).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$k = 2$ Gdy kodowana zmienna jest dychotomiczna, wówczas umieszczając ją w modelu regresji, uzyskamy wyliczony dla niej współczynnik (b_i). Współczynnik ten jest odniesieniem wartości zmiennej zależnej Y dla kategorii 1 do kategorii referencyjnej (w korekcji o pozostałe zmienne w modelu).

$k > 2$ Gdy analizowana zmienna ma więcej niż dwie kategorie, wówczas k kategorii jest reprezentowanych przez $k-1$ zmiennych fikcyjnych (dummy variables) o kodowaniu zero-jedynkowym. Tworząc zmienne o kodowaniu zero-jedynkowym, wybiera się kategorię, dla której nie tworzy się zmiennej fikcyjnej. Kategoria ta traktowana jest w modelach jako kategoria odniesienia (gdyż w każdej zmiennej zakodowanej w sposób zero-jedynkowy odpowiadają jej wartości 0).

Gdy tak uzyskane zmienne $X_1, X_2, ..., X_{k-1}$ o kodowaniu zero-jedynkowym zostaną umieszczone w modelu regresji, wówczas zostaną dla nich wyliczone współczynniki $b_1, b_2, ..., b_{k-1}$.

b_1 to odniesienie wyników Y (dla kodów 1 w X_1) do kategorii referencyjnej (w korekcji o pozostałe zmienne w modelu);

b_2 to odniesienie wyników Y (dla kodów 1 w X_2) do kategorii referencyjnej (w korekcji o pozostałe zmienne w modelu);

...

b_{k-1} to odniesienie wyników Y (dla kodów 1 w X_{k-1}) do kategorii referencyjnej (w korekcji o pozostałe zmienne w modelu).

Przykład

Zakodujemy zgodnie z kodowaniem zero-jedynkowym zmienną płeć o dwóch kategoriach (płeć męską wybierzemy jako kategorię referencyjną) i zmienną wykształcenie o 4 kategoriach (wykształcenie podstawowe wybierzemy jako referencyjne).

Płeć	Zakodowana płeć
k	1
k	1
k	1
m	0
m	0
k	1
k	1
m	0
m	0
k	1
m	0
k	1
m	0
k	1
m	0
m	0
...	...

Wykształcenie	Zakodowane wykształcenie		
	zawodowe	średnie	wyższe
podstawowe	0	0	0
podstawowe	0	0	0
podstawowe	0	0	0
zawodowe	1	0	0
zawodowe	1	0	0
zawodowe	1	0	0
zawodowe	1	0	0
średnie	0	1	0
średnie	0	1	0
średnie	0	1	0
średnie	0	1	0
wyższe	0	0	1
wyższe	0	0	1
wyższe	0	0	1
wyższe	0	0	1
wyższe	0	0	1
...	...	...	...

Budując na podstawie zmiennych fikcyjnych, w modelu regresji wielorakiej, moglibyśmy chcieć sprawdzić, jak zmienne te wpływają na pewną zmienną zależną, np. Y = wysokość zarobków (wyrażoną w tysiącach złotych). W wyniku takiej analizy dla każdej zmiennej fikcyjnej uzyskamy przykładowe współczynniki:

– dla płci istotny statystycznie współczynnik $b_i = -0.5$ – co oznacza, że średnie zarobki kobiet są o

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

pół tysiąca złotych niższe niż mężczyzn, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla wykształcenia zawodowego istotny statystycznie współczynnik $b_i = 0.6$ – co oznacza, że średnie zarobki osób z wykształceniem zawodowym są o 0.6 tysiąca złotych wyższe niż dla osób z wykształceniem podstawowym, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla wykształcenia średniego istotny statystycznie współczynnik $b_i = 1$ – oznacza, że średnie zarobki osób z wykształceniem średnim są o tysiąc złotych wyższe niż dla osób z wykształceniem podstawowym, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla wykształcenia wyższego istotny statystycznie współczynnik $b_i = 1.5$ – co oznacza, że średnie zarobki osób z wykształceniem wyższym są o 1.5 tysiąca wyższe niż dla osób z wykształceniem podstawowym, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie.

Kodowanie efektów (effect coding) jest wykorzystywane, by przy pomocy modeli wielowymiarowych odpowiedzieć na pytanie: jak wyniki (Y), w każdej analizowanej kategorii, różnią się od wyników średniej (nieważonej) uzyskanej z próby. Kodowanie to polega na przypisaniu wartości -1 lub 1 do każdej kategorii danej zmiennej. Kategoria zakodowana jako -1 jest wówczas **kategorią bazową** (base).

$k = 2$ Gdy kodowana zmienna jest dychotomiczna, wówczas umieszczając ją w modelu regresji, uzyskamy wyliczony dla niej współczynnik (b_i). Współczynnik ten jest odniesieniem Y dla kategorii 1 do nieważonej średniej ogólnej (w korekcji o pozostałe zmienne w modelu).

Gdy analizowana zmienna ma więcej niż dwie kategorie, wówczas k kategorii jest reprezentowanych przez $k-1$ zmiennych fikcyjnych o kodowaniu efektu. Tworząc zmienne o kodowaniu efektu, wybiera się kategorię, dla której nie tworzy się oddzielnej zmiennej. Kategoria ta traktowana jest w modelach jako kategoria bazowa (gdyż w każdej zmiennej zapisanej poprzez kodowanie efektu odpowiadają jej wartości -1).

Gdy tak uzyskane zmienne $X_1, X_2, ..., X_{k-1}$ o kodowaniu efektu zostaną umieszczone w modelu regresji, wówczas zostaną dla nich wyliczone współczynniki $b_1, b_2, ..., b_{k-1}$.

b_1 to odniesienie wyników Y (dla kodów 1 w X_1) do nieważonej średniej ogólnej (w korekcji o pozostałe zmienne w modelu);

b_2 to odniesienie wyników Y (dla kodów 1 w X_2) do nieważonej średniej ogólnej (w korekcji o pozostałe zmienne w modelu);

...

b_{k-1} to odniesienie wyników Y (dla kodów 1 w X_{k-1}) do nieważonej średniej ogólnej (w korekcji o pozostałe zmienne w modelu).

Przykład

Zakodujemy przy pomocy kodowania efektu zmienną płeć o dwóch kategoriach (płeć męską wybierzemy jako kategorię bazową) i zmienną wskazującą region zamieszkania na terenie analizowanego kraju. Wyróżniono 5 regionów: północny, południowy, wschodni, zachodni i centralny – region centralny wybierzemy jako bazowy.

Copyright ©2010-2026 U-smile Stats − Wszelkie prawa zastrzeżone

Płeć	Zakodowana płeć
k	1
k	1
k	1
m	-1
m	-1
k	1
k	1
m	-1
m	-1
k	1
m	-1
k	1
m	-1
k	1
m	-1
m	-1
...	...

Region zamieszkania	Zakodowany region			
	zachodni	wschodni	północny	południowy
centralny	-1	-1	-1	-1
centralny	-1	-1	-1	-1
centralny	-1	-1	-1	-1
zachodni	1	0	0	0
zachodni	1	0	0	0
zachodni	1	0	0	0
zachodni	1	0	0	0
wschodni	0	1	0	0
wschodni	0	1	0	0
wschodni	0	1	0	0
wschodni	0	1	0	0
północny	0	0	1	0
północny	0	0	1	0
południowy	0	0	0	1
południowy	0	0	0	1
południowy	0	0	0	1
...	...	...	...	...

Budując na podstawie zmiennych fikcyjnych, w modelu regresji wielorakiej, moglibyśmy chcieć sprawdzić jak zmienne te wpływają na pewną zmienną zależną np. Y = wysokość zarobków (wyrażoną w tysiącach złotych). W wyniku takiej analizy dla każdej zmiennej fikcyjnej uzyskamy przykładowe współczynniki:
– dla płci istotny statystycznie współczynnik $b_i = -0.5$ – co oznacza, że średnie zarobki kobiet są o pół tysiąca złotych niższe niż średnie zarobki w kraju, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla regionu zachodniego istotny statystycznie współczynnik $b_i = 0.6$ – co oznacza, że średnie zarobki osób zamieszkujących na zachodzie kraju są o 0.6 tysiąca złotych wyższe niż średnie zarobki w kraju, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla regionu wschodniego istotny statystycznie współczynnik $b_i = -1$ – oznacza, że średnie zarobki osób zamieszkujących na wschodzie kraju są o tysiąc złotych niższe niż średnie zarobki w kraju, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla regionu północnego istotny statystycznie współczynnik $b_i = 0.4$ – co oznacza, że średnie zarobki osób zamieszkujących na północy są o 0.4 tysiąca wyższe niż średnie zarobki w kraju, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie;
– dla regionu południowego nieistotny statystycznie współczynnik $b_i = 0.1$ – co oznacza, że średnie zarobki osób zamieszkujących na południu nie różnią się istotnie od średnich zarobków w kraju, przy założeniu że pozostałe zmienne w modelu pozostają na stałym poziomie.

Interakcje

Interakcje rozważane są w modelach wielowymiarowych, a ich występowanie oznacza, że wpływ zmiennej niezależnej (X_1) na zmienną zależną (Y) jest inny w zależności od poziomu kolejnej zmiennej niezależnej (X_2) lub szeregu kolejnych zmiennych niezależnych. By można było rozważać interakcje w modelach wielowymiarowych, należy wskazać zmienne mówiące o prawdopodobnych interakcjach, czyli iloczyny odpowiednich zmiennych. W tym celu wybieramy przycisk Interakcje w oknie wybranej analizy wielowymiarowej. W oknie ustawiania interakcji, z wciśniętym przyciskiem CTRL, wskazujemy zmienne, które mają tworzyć interakcje, i przenosimy je do sąsiedniej listy przy pomocy strzałki. Uruchamiając przycisk OK, uzyskujemy odpowiednie kolumny w arkuszu danych.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Uwaga! W analizie interakcji wybór odpowiedniego kodowania zmiennych dychotomicznych pozwala na uniknięcie przeparametryzowania związanego z interakcjami. Przeparametryzowanie powoduje, że efekty niższego rzędu dla zmiennych dychotomicznych są redundantne względem uwikłanych interakcji wyższego rzędu. W rezultacie uwzględnienie w modelu interakcji wyższego rzędu niweluje efekt interakcji rzędów niższych, nie pozwalając na ich prawidłową ocenę. By uniknąć przeparametryzowania w modelu, w którym występują interakcje zmiennych dychotomicznych, zaleca się wybierać opcję kodowanie efektów.

W modelach z interakcjami należy pamiętać o odpowiednim ich „przycinaniu", tak by usuwając efekty główne, usunąć również efekty rzędów wyższych, które są od nich zależne. To znaczy: jeśli w modelu mamy następujące zmienne (efekty główne): X_1, X_2, X_3 i interakcje: $X_1 * X_2$, $X_1 * X_3$, $X_2 * X_3$, $X_1 * X_2 * X_3$, to usuwając z modelu zmienną X_1, musimy usunąć również te interakcje, w których ona występuje, czyli: $X_1 * X_2$, $X_1 * X_3$ oraz $X_1 * X_2 * X_3$.

25.1 LINIOWA REGRESJA WIELORAKA

Liniowa regresja wieloraka

KIEDY STOSOWAĆ:

Budowany model regresji wielorakiej pozwala na zbadanie wpływu wielu zmiennych niezależnych (X_1, X_2, ..., X_k) na jedną zmienną zależną (Y). Najczęściej wykorzystywaną odmianą regresji wielorakiej jest liniowa regresja wieloraka. Jest ona rozszerzeniem modeli regresji liniowej opartej o współczynnik korelacji liniowej Pearsona. Zakłada występowanie liniowego związku pomiędzy badanymi zmiennymi. Liniowy model regresji wielorakiej przyjmuje postać:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon,$$

gdzie:
Y – zmienna zależna, objaśniana przez model,
$X_1, X_2, \ldots X_k$ – zmienne niezależne, objaśniające,
$\beta_0, \beta_1, \beta_2, \ldots \beta_k$ – parametry,
ϵ – składnik losowy (reszta modelu).

WARUNKI STOSOWANIA

- pomiar zmiennej zależnej Y na skali interwałowej,
- liniowy związek pomiędzy zmienną zależną a zmiennymi niezależnymi,
- odpowiednio duża liczba obserwacji.

Uwaga! Budując model, należy pamiętać, że liczba obserwacji musi być duża, to znaczy powinna spełniać założenie: $n \geq 50 + 8k$, gdzie k to liczba zmiennych objaśniających w modelu[73].

DEFINICJE MIAR

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zmienne fikcyjne i interakcje w modelu – omówienie przygotowania zmiennych fikcyjnych i interakcji przedstawiono w rozdziale 25.0.2 Przygotowanie zmiennych do analizy w modelach wielowymiarowych.

Współczynniki regresji – jeśli model został stworzony w oparciu o próbę o liczności n, powyższe równanie można przedstawić w postaci macierzowej:

$$Y = X\beta + \epsilon,$$

gdzie: $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$

Rozwiązaniem równania jest wówczas wektor ocen parametrów $\beta_0, \beta_1, \dots, \beta_k$, nazywanych **współczynnikami regresji**:

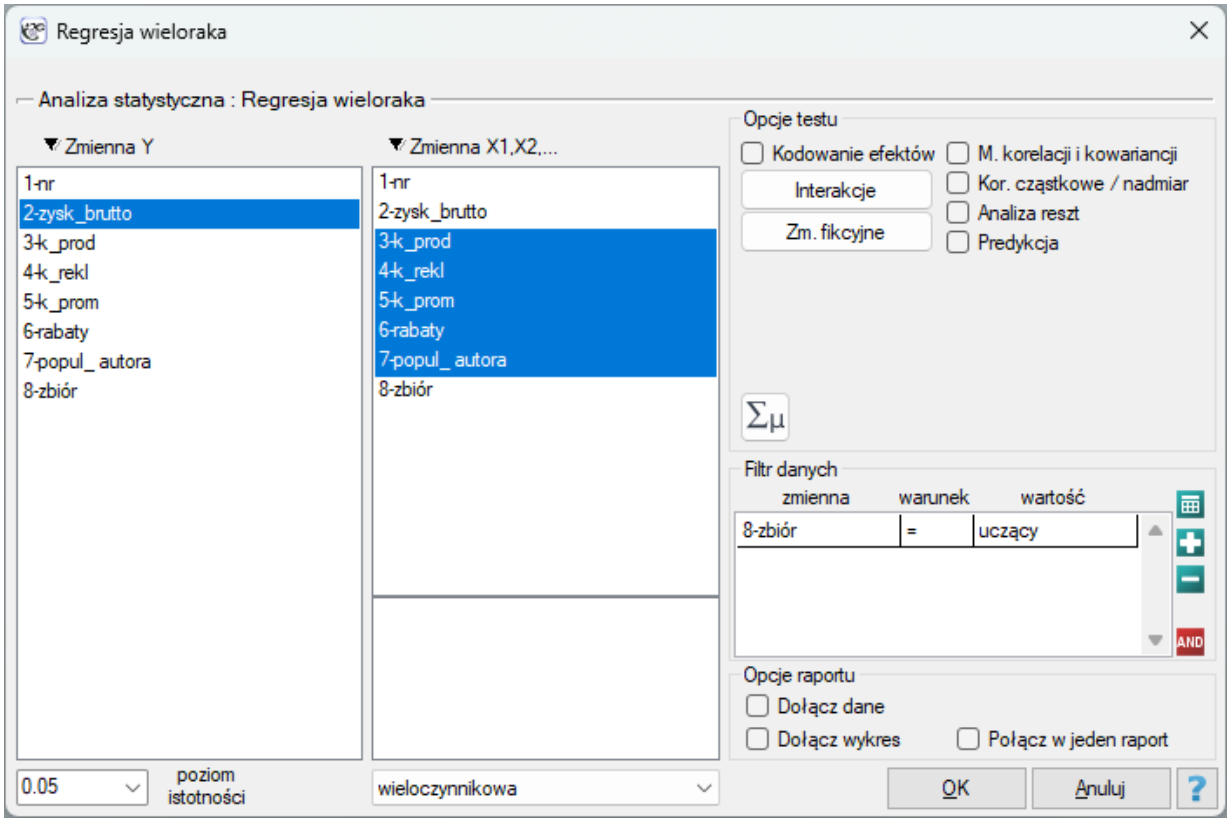
$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}.$$

Współczynniki te szacowane są poprzez klasyczną **metodę najmniejszych kwadratów**. Na podstawie tych wartości możemy wnioskować o wielkości wpływu zmiennej niezależnej (dla której ten współczynnik został oszacowany) na zmienną zależną – podają, o ile jednostek zmieni się zmienna zależna, gdy zmienną niezależną zmienimy o 1 jednostkę. Każdy współczynnik obarczony jest pewnym błędem szacunku, wyliczanym ze wzoru $SE_b = \sqrt{\frac{1}{n-(k+1)} e^T e (X^T X)^{-1}}$, gdzie $e = Y - \widehat{Y}$ to wektor **reszt modelu** (różnica pomiędzy rzeczywistymi wartościami zmiennej zależnej Y a wartościami $\widehat{Y}$ przewidywanymi na podstawie modelu).

JAK USTAWIĆ ANALIZĘ

Statystyka zaawansowana → Modele wielowymiarowe → Regresja wieloraka.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Weryfikacja modelu

Istotność statystyczna poszczególnych zmiennych w modelu – na podstawie współczynnika oraz jego błędu szacunku możemy wnioskować, czy zmienna niezależna, dla której ten współczynnik został oszacowany, wywiera istotny wpływ na zmienną zależną. W tym celu posługujemy się testem t-Studenta.

Jakość zbudowanego modelu liniowej regresji wielorakiej możemy ocenić kilkoma miarami:

- **Błąd standardowy estymacji** – jest miarą dopasowania modelu, opartą na resztach modelu $e_i = y_i - \widehat{y}_i$, czyli rozbieżności pomiędzy rzeczywistymi wartościami zmiennej zależnej y_i w próbie a wartościami zmiennej zależnej $\widehat{y}_i$ wyliczonymi na podstawie zbudowanego modelu. Aby model był dobrze dopasowany, błąd standardowy estymacji powinien być jak najmniejszy.
- **Współczynnik korelacji wielorakiej** $R = \sqrt{R^2} \in \langle 0; 1 \rangle$ – określa siłę oddziaływania zespołu zmiennych $X_1, X_2, \ldots X_k$ na zmienną zależną Y.
- **Współczynnik determinacji wielorakiej** R^2 – jest miarą dopasowania modelu. Wartość tego współczynnika mieści się w przedziale $\langle 0; 1 \rangle$, gdzie 1 oznacza doskonałe dopasowanie modelu, 0 – zupełny brak dopasowania. Wyraża on procent zmienności zmiennej zależnej tłumaczony przez model. Ponieważ wartość współczynnika R^2 zależy od dopasowania modelu, ale jest również wrażliwa na liczbę zmiennych w modelu i liczność próby, wyznacza się też jego poprawianą wartość R^2_{adj}.
- **Kryteria informacyjne** (AIC, $AICc$, BIC) – opierają się na entropii informacji niesionej przez model (niepewności modelu), tzn. szacują utraconą informację, gdy dany model jest używany do opisu badanego zjawiska. Powinniśmy wybierać model o minimalnej wartości danego kryterium informacyjnego. Kryterium informacyjne nie jest jednak miarą absolutną – jeśli wszystkie porównywane modele źle opisują rzeczywistość, w kryterium informacyjnym nie ma sensu szukać ostrzeżenia.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **Analiza błędów dla prognoz ex post** (MAE, MPE, $MAPE$) – opisują trafność prognozy, informując, o ile średnio uzyskiwane realizacje zmiennej zależnej będą się odchylać od prognoz.
- **Istotność statystyczna wszystkich zmiennych w modelu** – podstawowym narzędziem szacującym istotność wszystkich zmiennych w modelu jest test analizy wariancji (test F).

Więcej informacji o zmiennych w modelu:

- **Standaryzowane** $b_1, b_2, \ldots, b_k$ – w odróżnieniu od parametrów surowych (które w zależności od opisywanej zmiennej są wyrażone w różnych jednostkach miary i nie mogą być bezpośrednio porównywane), standaryzowane oceny parametrów modelu pozwalają porównywać wkład poszczególnych zmiennych w wyjaśnienie zmienności zmiennej zależnej Y.
- **Macierz korelacji** – zawiera informacje o sile związku pomiędzy poszczególnymi zmiennymi, czyli współczynnik korelacji Pearsona $r_p \in \langle -1; 1 \rangle$, wyznaczany dla każdej pary zmiennych, bez uwzględniania wpływu pozostałych zmiennych w modelu.
- **Macierz kowariancji** – podobnie jak macierz korelacji, zawiera informacje o związku liniowym pomiędzy poszczególnymi zmiennymi, przy czym wartość ta nie jest wystandaryzowana.
- **Współczynnik korelacji cząstkowej** – należy do przedziału $\langle -1; 1 \rangle$ i jest miarą korelacji pomiędzy konkretną zmienną niezależną X_i (uwzględniając jej skorelowanie z pozostałymi zmiennymi w modelu) a zmienną zależną Y (uwzględniając jej skorelowanie z pozostałymi zmiennymi w modelu). Kwadrat tego współczynnika to **współczynnik determinacji cząstkowej** – należy do przedziału $\langle 0; 1 \rangle$ i oznacza stosunek wyłącznej zmienności danej zmiennej niezależnej X_i do tej zmienności zmiennej zależnej Y, która nie została wyjaśniona przez pozostałe zmienne w modelu. Im wartość tych współczynników znajduje się bliżej 0, tym bardziej bezużyteczną informację niesie badana zmienna, czyli jest ona nadmiarowa.
- **Współczynnik korelacji semicząstkowej** – należy do przedziału $\langle -1; 1 \rangle$ i jest miarą korelacji pomiędzy konkretną zmienną niezależną X_i (uwzględniając jej skorelowanie z pozostałymi zmiennymi w modelu) a zmienną zależną Y (NIE uwzględniając jej skorelowania z pozostałymi zmiennymi w modelu). Kwadrat tego współczynnika to **współczynnik determinacji semicząstkowej** – należy do przedziału $\langle 0; 1 \rangle$ i oznacza stosunek wyłącznej zmienności danej zmiennej niezależnej X_i do całkowitej zmienności zmiennej zależnej Y.
- **R-kwadrat (**$R^2 \in \langle 0; 1 \rangle$**)** – wyraża procent zmienności danej zmiennej niezależnej X_i tłumaczony przez pozostałe zmienne niezależne. Im bliżej wartości 1, tym silniej badana zmienna związana jest liniowo z pozostałymi zmiennymi niezależnymi, co może oznaczać, że jest ona zmienną nadmiarową.
- **Współczynnik inflacji wariancji (**$VIF \in \langle 1; \infty)$**)** – określa, jak bardzo wariancja szacowanego współczynnika regresji jest zwiększona z powodu współliniowości. Im bliżej wartości 1, tym mniejsza współliniowość. Przyjmuje się, że silna współliniowość występuje, gdy $VIF > 5$[153].
- **Tolerancja** $= 1 - R^2 \in \langle 0; 1 \rangle$ – wyraża procent zmienności danej zmiennej niezależnej X_i NIE tłumaczony przez pozostałe zmienne niezależne. Im wartość tolerancji jest bliższa 0, tym silniej badana zmienna związana jest liniowo z pozostałymi zmiennymi niezależnymi, co może oznaczać, że jest ona zmienną nadmiarową.
- **Porównanie modelu pełnego z modelem po usunięciu danej zmiennej** – dokonujemy testem F (gdy z modelu usuwamy więcej niż jedną zmienną) lub testem t-Studenta (gdy usuwamy tylko

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

jedną zmienną – jest to ten sam test, którym badamy istotność poszczególnych zmiennych w modelu; wyniki obu testów są wówczas tożsame). Jeśli różnica pomiędzy porównywanymi modelami jest istotna statystycznie, model pełny jest istotnie lepszy niż model zredukowany, co oznacza, że badana zmienna nie jest nadmiarowa.

- **Wykresy rozrzutu** – pozwalają dokonać subiektywnej oceny liniowości związku pomiędzy zmiennymi i zidentyfikować punkty odstające, a także posłużyć w analizie reszt modelu.

Analiza reszt modelu
By otrzymać poprawny model regresji, powinniśmy sprawdzić podstawowe założenia dotyczące reszt modelu:

- **Obserwacje odstające** – badając reszty modelu, szybko można uzyskać wiedzę na temat wartości odstających. Obserwacje takie mogą bardzo zaburzyć równanie regresji, ponieważ mają duży wpływ na wartości współczynników tego równania. Jeśli dana reszta e_i jest oddalona o więcej niż 3 odchylenia standardowe od wartości średniej, wówczas można podejrzewać, że usunięcie tej obserwacji przyczyni się do poprawy modelu. **Odległość Cooka** opisuje wielkość zmian współczynników regresji powstałą na skutek pominięcia danego przypadku (w programie zaznaczone pogrubioną czcionką są odległości Cooka dla przypadków przekraczających 50 percentyl statystyki rozkładu Fishera-Snedecora $F(0.5, k+1, n-k-1)$). **Odległość Mahalanobisa** dedykowana jest do wykrywania obserwacji odstających – wysokie wartości świadczą o znacznym oddaleniu danego przypadku od centrum zmiennych niezależnych.
- **Normalność rozkładu reszt modelu** – założenie to sprawdzamy wizualnie przy pomocy wykresu Q-Q rozkładu normalnego. Duża różnica między rozkładem reszt a rozkładem normalnym może zaburzać ocenę istotności współczynników poszczególnych zmiennych modelu.
- **Homoskedastyczność (stałość wariancji)** – by sprawdzić, czy istnieją obszary, gdzie wariancja reszt modelu jest zwiększona lub zmniejszona, posługujemy się wykresami reszt (lub kwadratu reszt) względem wartości przewidywanych i obserwowanych.
- **Autokorelacja reszt modelu** – aby zbudowany model można było uznać za poprawny, wartości reszt nie powinny być ze sobą skorelowane. Założenie to możemy sprawdzić statystyką testu Durbina-Watsona.

WYKONANIE I INTERPRETACJA

Przykład 25.2. (plik wydawca.pqs)
Pewien wydawca książek chciał się dowiedzieć, jaki wpływ na zysk brutto ze sprzedaży mają takie zmienne jak: koszty produkcji, koszty reklamy, koszty promocji bezpośredniej, suma udzielonych rabatów, popularność autora. W tym celu przeanalizował 40 pozycji wydanych w ciągu ostatniego roku (zbiór uczący). Fragment danych przedstawia poniższy rysunek:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

nr	zysk_brutto	k_prod	k_rekl	k_prom	rabaty	popul_ autora
1	58	7.9	9	0.38	1.8	1
2	63	10.1	10	0.59	2.4	0
3	27	3	7	0.7	1.7	0
4	35	6	3	0.21	2.6	1
5	34	6.6	2.1	0.13	2.2	0
6	48	10.7	1	0.08	2.1	1
7	14	2.7	0.7	0.06	0.3	0
8	63.5	12	5	0.56	1.7	0

Pięć pierwszych zmiennych wyrażonych jest w tysiącach dolarów – są to więc zmienne zebrane na skali interwałowej. Natomiast ostatnia zmienna, popularność autora, to zmienna dychotomiczna, gdzie 1 oznacza autora znanego, 0 oznacza autora nieznanego. Na podstawie uzyskanej wiedzy wydawca planuje przewidzieć zysk brutto z kolejnej wydawanej książki znanego autora. Koszty, jakie zamierza ponieść, to: koszty produkcji ≈ 11, koszty reklamy ≈ 13, koszty promocji bezpośredniej ≈ 0.5, suma udzielonych rabatów ≈ 0.5.

Budujemy model liniowej regresji wielorakiej dla zbioru uczącego, wybierając: zysk brutto – jako zmienną zależną Y, koszty produkcji, koszty reklamy, koszty promocji bezpośredniej, suma udzielonych rabatów, popularność autora – jako zmienne niezależne X_1, X_2, X_3, X_4, X_5. W rezultacie wyliczone zostaną współczynniki równania regresji oraz miary pozwalające ocenić jakość modelu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja wieloraka	
Analizowane zmienne	zysk_brutto
	k_prod
	k_rekl
	k_prom
	rabaty
	popul_ autora
Filtr danych	zbiór=uczący
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Założenie liczności próby n ≥ 50+8k	Nie
Liczność	40
Liczba zmiennych w modelu	5
R	0.9225
R2	0.851
R2 skorygowane	0.8291
Błąd standardowy estymacji	8.0865
Resztowa suma kwadratów	2223.3112
Całkowita suma kwadratów	14918.9978
Wyjaśniona suma kwadratów	12695.6865
F	38.8298
Wartość p	<0.0001
AIC - kryterium Akaikego	172.7149
AICc - skorygowane kryterium Akaikego	177.0786
BIC - bayesowskie kryterium Schwarza	190.226
MAE (średni błąd bezwzględny)	4.621
MPE (średni błąd procentowy)	-21.483%
MAPE (średni bezwzględny błąd procentowy)	31.575%

Model

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	wartość p	stand. b	błąd stand	średnia	odch. std.
w. wolny	4.175186	4.772779	-5.524268	13.874639	0.874791	0.387825				
k_prod	2.560709	0.501507	1.541525	3.579894	5.106033	0.000013	0.422818	0.082807	7.4675	3.229463
k_rekl	1.998235	0.359065	1.268527	2.727943	5.565106	0.000003	0.461287	0.082889	5.85	4.515046
k_prom	4.668238	4.791644	-5.069555	14.40603	0.974245	0.336816	0.066242	0.067993	0.5145	0.277534
rabaty	1.423171	1.404811	-1.431749	4.278091	1.013069	0.318183	0.067478	0.066608	1.605	0.927348
popul_ autora	10.153717	2.782567	4.498861	15.808574	3.649047	0.000874	0.261561	0.071679	0.55	0.503831

$SE_e = 8.086501$, $R^2 = 0.850974$, $R^2_{adj} = 0.829059$, **test F:** $p < 0.000001$

Model jest dobrze dopasowany i istotny statystycznie.

Na podstawie oszacowanej wartości współczynnika b, związek pomiędzy zyskiem brutto a wszystkimi zmiennymi niezależnymi możemy opisać równaniem:

$$zysk_{brutto} = 4.18 + 2.56(k_{prod}) + 2(k_{rekl}) + 4.67(k_{prom}) + 1.42(rabaty) + 10.15(popul_{autora}) + [8.09]$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uzyskane współczynniki interpretujemy następująco:

- jeśli koszt produkcji wzrośnie o 1 tysiąc dolarów, to zysk brutto wzrośnie o około 2.56 tysiąca dolarów, przy założeniu że pozostałe zmienne się nie zmienią;
- jeśli koszt reklamy wzrośnie o 1 tysiąc dolarów, to zysk brutto wzrośnie o około 2 tysiące dolarów, przy założeniu że pozostałe zmienne się nie zmienią;
- jeśli koszt promocji bezpośredniej wzrośnie o 1 tysiąc dolarów, to zysk brutto wzrośnie o około 4.67 tysiąca dolarów, przy założeniu że pozostałe zmienne się nie zmienią;
- jeśli suma udzielonych rabatów wzrośnie o 1 tysiąc dolarów, to zysk brutto wzrośnie o około 1.42 tysiąca dolarów, przy założeniu że pozostałe zmienne się nie zmienią;
- jeśli książka została napisana przez autora znanego (oznaczonego przez 1), to w modelu popularność autora przyjmujemy jako wartość 1 i otrzymujemy równanie:

$$zysk_{brutto} = 14.33 + 2.56(k_{prod}) + 2(k_{rekl}) + 4.67(k_{prom}) + 1.42(rabaty)$$

Jeśli natomiast książka została napisana przez autora nieznanego (oznaczonego przez 0), to otrzymujemy równanie:

$$zysk_{brutto} = 4.18 + 2.56(k_{prod}) + 2(k_{rekl}) + 4.67(k_{prom}) + 1.42(rabaty)$$

Wynik testu t-Studenta uzyskany dla każdej zmiennej wskazuje, że tylko koszt produkcji, koszt reklamy oraz popularność autora wywierają istotny wpływ na otrzymany zysk. Jednocześnie, dla tych zmiennych standaryzowane współczynniki b są największe.

Na podstawie interpretacji dotychczasowych wyników możemy przypuszczać, że część zmiennych nie wywiera istotnego wpływu na zysk i może być zbyteczna. Aby model był dobrze sformułowany, interwałowe zmienne niezależne powinny być silnie skorelowane ze zmienną zależną i stosunkowo słabo pomiędzy sobą. Możemy to sprawdzić, wyliczając macierz korelacji i macierz kowariancji:

Korelacje

	zysk_brutt	k_prod	k_rekl	k_prom	rabaty	popul_ aut
zysk_brutto	1	0.770685	0.794843	0.071095	0.131924	0.553803
k_prod	0.770685	1	0.558843	-0.079792	0.092951	0.340624
k_rekl	0.794843	0.558843	1	0.119889	0.056708	0.326878
k_prom	0.071095	-0.079792	0.119889	1	-0.056478	-0.049327
rabaty	0.131924	0.092951	0.056708	-0.056478	1	0.010427
popul_ autora	0.553803	0.340624	0.326878	-0.049327	0.010427	1

Kowariancja

	zysk_brutt	k_prod	k_rekl	k_prom	rabaty	popul_ aut
zysk_brutto	382.53840	48.679353	70.190897	0.385914	2.392782	5.457308
k_prod	48.679353	10.429429	8.14859	-0.071517	0.278372	0.554231
k_rekl	70.190897	8.14859	20.385641	0.150231	0.237436	0.74359
k_prom	0.385914	-0.071517	0.150231	0.077025	-0.014536	-0.006897
rabaty	2.392782	0.278372	0.237436	-0.014536	0.859974	0.004872
popul_ autora	5.457308	0.554231	0.74359	-0.006897	0.004872	0.253846

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Najbardziej spójną informację, pozwalającą znaleźć te zmienne w modelu, które są zbędne (nadmiarowe), niesie analiza korelacji cząstkowej i semicząstkowej oraz nadmiarowości:

Kor. cząst. semi.

	cząstkowe	semicząstk	tolerancja	R^2	Stat. t	Wartość p
k_prod	0.658793	-0.338045	0.639209	0.360791	5.106033	0.000013
k_rekl	0.690424	-0.368438	0.637949	0.362051	5.565106	0.000003
k_prom	0.164797	-0.0645	0.948099	0.051901	0.974245	0.336816
rabaty	0.171176	-0.06707	0.987951	0.012049	1.013069	0.318183
popul_ autora	0.53049	-0.241585	0.85309	0.14691	3.649047	0.000874

Wartości współczynników korelacji cząstkowej i semicząstkowej wskazują, że najmniejszy wkład w budowany model mają: koszt promocji bezpośredniej i suma udzielonych rabatów. Jednak są to zmienne najmniej skorelowane z pozostałymi w modelu, o czym świadczy niska wartość R^2 i wysoka wartość tolerancji. Ostatecznie, ze statystycznego punktu widzenia, modele bez tych zmiennych nie byłyby modelami gorszymi niż model obecny (patrz wynik testu t-Studenta dla porównywania modeli). To od decyzji badacza zależy, czy pozostawi ten model, czy zbuduje nowy model pozbawiony kosztów promocji bezpośredniej i sumy udzielonych rabatów. My pozostawiamy model obecny.

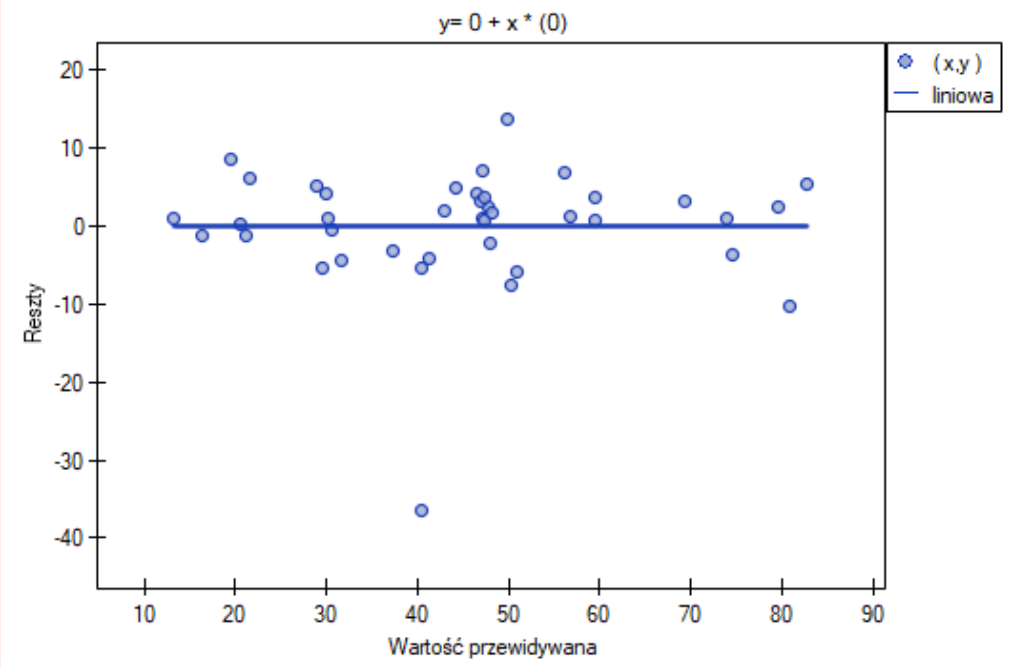
Na koniec przeprowadzimy analizę reszt. Fragment tej analizy znajduje się poniżej:

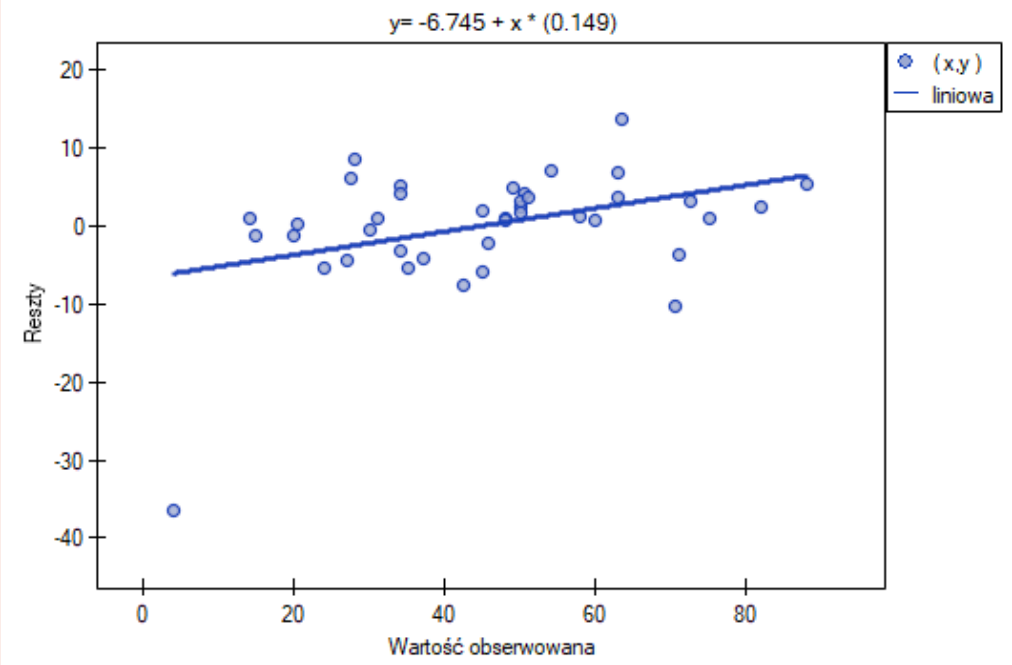
Analiza reszt

	wartość pr	reszta	reszta star	<=-3sd	(-3sd;2sd]	(-2sd;sd]	(-sd;sd)	[sd;2sd)	[2sd;3sd)	>=3sd
1	56.87826	1.12174	0.138718				*			
2	56.190571	6.809429	0.842074				*			
3	31.532115	-4.532115	-0.560454				*			
4	40.368438	-5.368438	-0.663877				*			
5	29.010008	4.989992	0.617077				*			
6	47.088847	0.911153	0.112676				*			
7	13.194911	0.805089	0.09956				*			
8	49.928477	13.571523	1.678293					*		
9	46.501478	3.998522	0.494469				*			
10	47.776747	2.223253	0.274934				*			
11	47.989623	-2.289623	-0.283141				*			
12	79.536147	2.463853	0.304687				*			
13	29.921482	4.078518	0.504361				*			
14	30.206017	0.793983	0.098186				*			
15	46.840918	3.159082	0.390661				*			
16	40.467863	-36.46786	-4.509721	*						
17	47.110045	6.889955	0.852032				*			
18	50.197864	-7.697864	-0.95194			*				
19	43.035765	1.964235	0.242903				*			
20	44.211462	4.788538	0.592164				*			
21	29.452408	-5.452408	-0.67426				*			

Możemy zauważyć, że jedna z reszt modelu jest obserwacją odstającą – jest oddalona o więcej niż 3 odchylenia standardowe od wartości średniej. Jest to obserwacja o numerze 16. Obserwację tę możemy łatwo znaleźć, kreśląc wykres reszt względem obserwowanych lub przewidywanych wartości zmiennej Y.

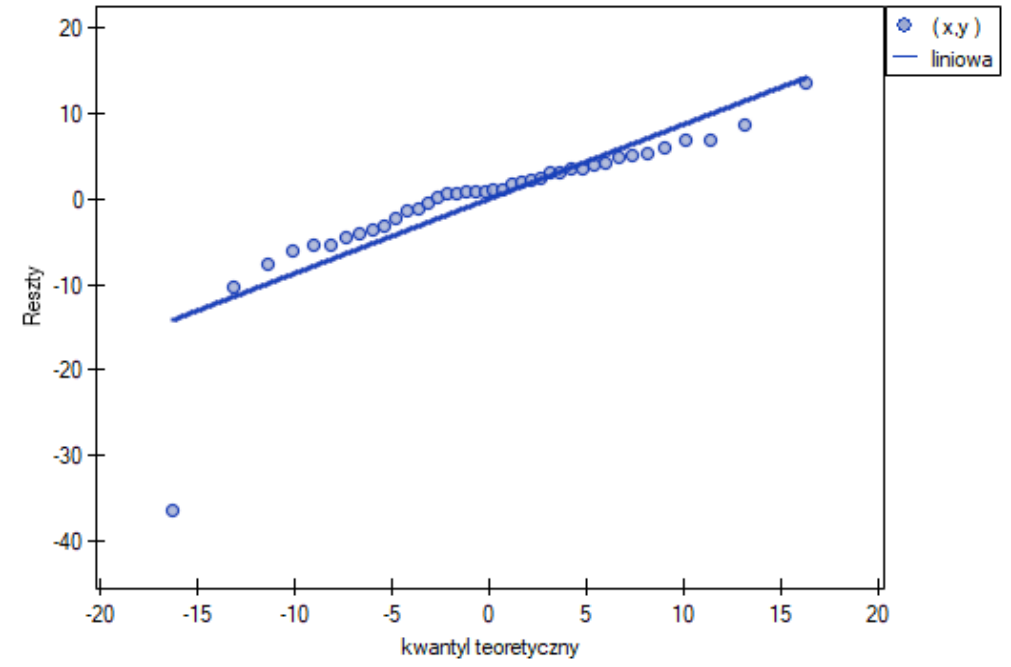
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





Ten odstający punkt zaburza założenie dotyczące homoskedastyczności – byłoby ono spełnione (tzn. wariancja reszt opisana na osi Y byłaby podobna, gdy przechodzimy wzdłuż osi X), gdybyśmy ten punkt odrzucili. Dodatkowo, rozkład reszt nieco odbiega od rozkładu normalnego (wartość p testu Lillieforsa wynosi $p = 0.016415$):

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Filtr danych

zmienna	warunek	wartość
8-zbiór	=	uczący
1-nr	<>	16

Przyglądając się dokładniej punktowi odstającemu (pozycja 16 w danych do zadania), widzimy, że książka ta jako jedyna wykazuje wyższe koszty niż zysk brutto (zysk brutto = 4 tysiące dolarów, suma kosztów = (8+6+0.33+1.6) = 15.93 tysiąca dolarów).

Uzyskany model możemy poprawić, usuwając z niego punkt odstający. Wymaga to ponownego przeprowadzenia analizy z włączonym filtrem wykluczającym punkt odstający.

Filtr danych

zmienna	warunek	wartość
1-nr	<>	16

W rezultacie uzyskaliśmy bardzo podobny model, ale obarczony mniejszym błędem i lepiej dopasowany:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja wieloraka	
Analizowane zmienne	zysk_brutto
	k_prod
	k_rekl
	k_prom
	rabaty
	popul_ autora
Filtr danych	zbiór=uczący
	i nr<>16
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Założenie liczności próby n ≥ 50+8k	Nie
Liczność	39
Liczba zmiennych w modelu	5
R	0.9699
R2	0.9408
R2 skorygowane	0.9318
Błąd standardowy estymacji	4.8633
Resztowa suma kwadratów	780.4997
Całkowita suma kwadratów	13173.1708
Wyjaśniona suma kwadratów	12392.6711
F	104.7939
Wartość p	<0.0001
AIC - kryterium Akaikego	128.8585
AICc - skorygowane kryterium Akaikego	133.3585
BIC - bayesowskie kryterium Schwarza	146.167
MAE (średni błąd bezwzględny)	3.5963
MPE (średni błąd procentowy)	-1.478%
MAPE (średni bezwzględny błąd procentowy)	9.137%

Model

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	wartość p	stand. b	błąd stand	średnia	odch. std.
w. wolny	6.892628	2.891394	1.010044	12.775213	2.383843	0.023041				
k_prod	2.678933	0.301989	2.064531	3.293335	8.870952	0	0.47057	0.053046	7.453846	3.27051
k_rekl	2.080912	0.216204	1.641042	2.520781	9.624775	0	0.511207	0.053114	5.846154	4.574002
k_prom	1.920214	2.903129	-3.986247	7.826675	0.661429	0.512929	0.028828	0.043584	0.519231	0.279524
rabaty	1.325426	0.844957	-0.393651	3.044503	1.568632	0.126274	0.066878	0.042635	1.605128	0.93947
popul_ autora	7.382624	1.710653	3.902274	10.862973	4.315676	0.000136	0.199191	0.046155	0.564103	0.502356

$$zysk_{brutto} = 6.89 + 2.68(k_{prod}) + 2.08(k_{rekl}) + 1.92(k_{prom}) + 1.33(rabaty) + 7.38(popul_{autora}) + [4.86]$$

Ostatecznie zbudowany model wykorzystamy do predykcji. Na podstawie przewidywanych nakładów w wysokości: koszty produkcji ≈ 11 tysięcy dolarów, koszty reklamy ≈ 13 tysięcy dolarów, koszty promocji bezpośredniej ≈ 0.5 tysiąca dolarów, suma udzielonych rabatów ≈ 0.5 tysiąca dolarów oraz faktu, że

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

jest to autor znany (popularność autora ≈ 1), wyliczamy przewidywany zysk brutto wraz z przedziałem ufności:

Predykcja dla 3-k_prod	11
Predykcja dla 4-k_rekl	13
Predykcja dla 5-k_prom	0.5
Predykcja dla 6-rabaty	0.5
Predykcja dla 7-popul_ autora	1
predykcja wartości Y	72.418189
-95% CI (punktu)	61.856275
+95% CI (punktu)	82.980103
-95% CI (wartości oczekiwanej)	68.722994
+95% CI (wartości oczekiwanej)	76.113384

Przewidziany zysk wynosi 72 tysiące dolarów.

Na koniec należy jeszcze zauważyć, że jest to tylko model wstępny. W badaniu właściwym należałoby zebrać więcej danych, gdyż liczba zmiennych w modelu jest zbyt mała w stosunku do liczby ocenianych książek, tzn. $n < 50 + 8k$.

DLA ZAINTERESOWANYCH

Test t-Studenta dla istotności poszczególnych zmiennych
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \beta_i = 0, \\ \mathcal{H}_1: &\quad \beta_i \neq 0. \end{aligned}$$

Wzory: statystyka testowa ma postać:

$$t = \frac{b_i}{SE_{b_i}}.$$

Statystyka testowa ma rozkład t-Studenta z $n - k$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Błąd standardowy estymacji

$$SE_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n - (k+1)}}.$$

Współczynnik determinacji wielorakiej
W jego wyznaczeniu posługujemy się następującą równością:

$$T_{SS} = E_{SS} + R_{SS},$$

gdzie:
T_{SS} – całkowita suma kwadratów,
E_{SS} – suma kwadratów wyjaśniona przez model,

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

R_{SS} – resztowa suma kwadratów.
Współczynnik determinacji wyliczamy z wzoru:

$$R^2 = \frac{E_{SS}}{T_{SS}}.$$

Poprawiona wartość tego parametru:

$$R^2_{adj} = R^2 - \frac{k(1-R^2)}{n-(k+1)}.$$

Kryteria informacyjne
AIC, $AICc$ i BIC są rodzajem kompromisu pomiędzy dobrocią dopasowania i złożonością. Drugi element sumy we wzorach na kryteria informacyjne (tzw. funkcja straty lub kary) mierzy prostotę modelu – zależy od liczby zmiennych w modelu (k) i liczności próby (n).

Kryterium informacyjne Akaikego *(ang. Akaike information criterion)*

$$AIC = n \cdot \ln \frac{R_{SS}}{n} + 2(k+1) + (constant)$$

gdzie stałą można pominąć, ponieważ jest taka sama w każdym z porównywanych modeli. Jest to kryterium asymptotyczne – odpowiednie dla dużych prób, tzn. gdy $\frac{n}{k+2} > 40$. Przy małych próbach ma tendencję do preferowania modeli z dużą liczbą zmiennych.

Przykład interpretacji porównania wielkości AIC
Załóżmy, że wyznaczyliśmy AIC dla trzech modeli: AIC_1=100, AIC_2=101.4, AIC_3=110. Wówczas można wyznaczyć względną wiarygodność każdego modelu, wyznaczaną względem innego modelu, najczęściej tego o najmniejszej wartości AIC, według wzoru: $e^{(AIC_{min}-AIC_i)/2}$. Porównując model 2 do modelu pierwszego, powiemy, że prawdopodobieństwo, iż zminimalizuje on utratę informacji, stanowi około połowę prawdopodobieństwa, że zrobi to model 1 (dokładnie $\exp((100-101.4)/2) = 0.497$). Porównując model 3 do modelu pierwszego, powiemy, że to prawdopodobieństwo stanowi niewielką część prawdopodobieństwa dla modelu 1 (dokładnie $\exp((100-110)/2) = 0.007$).

Poprawione kryterium informacyjne Akaikego

$$AICc = AIC + \frac{2(k+3)(k+4)}{n-k}$$

Poprawka kryterium Akaikego dotyczy wielkości próby, przez co jest to miara rekomendowana również dla prób o małych licznościach.

Bayesowskie kryterium informacyjne Schwarza *(ang. Bayes Information Criterion, Schwarz criterion)*

$$BIC = n \cdot \ln \frac{R_{SS}}{n} + (k+1) \ln n + (constant)$$

Podobnie jak poprawione kryterium Akaikego, BIC uwzględnia wielkość próby.

Analiza błędów dla prognoz ex post
MAE (średni błąd bezwzględny) *(ang. mean absolute error)* – trafność prognozy określona przez MAE informuje, o ile średnio uzyskiwane realizacje zmiennej zależnej będą się odchylać (co do wartości bezwzględnej) od prognoz.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

MPE (średni błąd procentowy) *(ang. mean percentage error)* – informuje, jaki średni procent realizacji zmiennej zależnej stanowią błędy prognozy.

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{e_i}{y_i}$$

MAPE (średni bezwzględny błąd procentowy) *(ang. mean absolute percentage error)* – informuje o średniej wielkości błędów prognoz wyrażonych w procentach rzeczywistych wartości zmiennej zależnej. MAPE pozwala porównać dokładność prognoz uzyskanych na bazie różnych modeli.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right|$$

Test F – istotność statystyczna wszystkich zmiennych w modelu
Test ten weryfikuje jednocześnie 3 równoważne hipotezy:

$$\begin{array}{llll} \mathcal{H}_0: & \text{wszystkie } \beta_i = 0, & \mathcal{H}_1: & \text{istnieje } \beta_i \neq 0; \\ \mathcal{H}_0: & R^2 = 0, & \mathcal{H}_1: & R^2 \neq 0; \\ \mathcal{H}_0: & \text{brak związku liniowego}, & \mathcal{H}_1: & \text{liniowość związku}. \end{array}$$

Wzory: statystyka testowa ma postać:

$$F = \frac{E_{MS}}{R_{MS}}$$

gdzie:
$E_{MS} = \dfrac{E_{SS}}{df_E}$ – średnia kwadratów wyjaśniona przez model,
$R_{MS} = \dfrac{R_{SS}}{df_R}$ – resztowa średnia kwadratów,
$df_E = k$, $df_R = n - (k+1)$ – odpowiednie stopnie swobody.
Statystyka ta podlega rozkładowi F Snedecora z df_E i df_R stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Odległość Cooka i odległość Mahalanobisa – opisane powyżej w części praktycznej.

Autokorelacja reszt – statystyka Durbina-Watsona

$$d = \frac{\sum_{t=2}^{n}\left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{n} e_t^2}.$$

Interpretacja: aby sprawdzić dodatnią autokorelację na poziomie istotności α, sprawdzamy położenie statystyki d w stosunku do górnej ($d_{U,\alpha}$) i dolnej ($d_{L,\alpha}$) wartości krytycznej:

- jeżeli $d < d_{L,\alpha}$ – błędy są dodatnio skorelowane;
- jeśli $d > d_{U,\alpha}$ – błędy nie są dodatnio skorelowane;
- jeśli $d_{L,\alpha} < d < d_{U,\alpha}$ – wynik testu jest niejednoznaczny.

Aby sprawdzić ujemną autokorelację na poziomie istotności α, sprawdzamy położenie wartości $4 - d$ w stosunku do górnej i dolnej wartości krytycznej:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- jeżeli $4 - d < d_{L,\alpha}$ – błędy są ujemnie skorelowane;
- jeśli $4 - d > d_{U,\alpha}$ – błędy nie są ujemnie skorelowane;
- jeśli $d_{L,\alpha} < 4 - d < d_{U,\alpha}$ – wynik testu jest niejednoznaczny.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Predykcja na podstawie modelu i walidacja zbioru testowego

Walidacja i predykcja dla regresji wielorakiej

KIEDY STOSOWAĆ:

Walidacja modelu to sprawdzenie jego jakości. W pierwszej kolejności wykonywana jest na danych, na których model był zbudowany (tzw. **zbiór uczący**), czyli zwracana jest w raporcie opisującym uzyskany model. By można było z większą pewnością osądzić, na ile model nadaje się do prognozy nowych danych, ważnym elementem walidacji jest zastosowanie modelu do danych, które nie były wykorzystywane w estymacji modelu. Jeśli podsumowanie w oparciu o dane uczące będzie satysfakcjonujące, tzn. wyznaczane błędy, współczynniki R^2 i kryteria informacyjne będą na zadowalającym nas poziomie, a podsumowanie w oparciu o nowe dane (tzw. **zbiór testowy**) będzie równie korzystne, wówczas z dużym prawdopodobieństwem można uznać, że taki model nadaje się do predykcji. Dane testujące powinny pochodzić z tej samej populacji, z której były wybrane dane uczące. Często jest tak, że przed przystąpieniem do budowy modelu zbieramy dane, a następnie w sposób losowy dzielimy je na zbiór uczący, czyli dane które posłużą do budowy modelu, i zbiór testowy, czyli dane które posłużą do dodatkowej walidacji modelu.

Predykcja – najczęściej ostatnim etapem analizy regresji jest wykorzystanie zbudowanego i uprzednio zweryfikowanego modelu do predykcji.

DEFINICJE MIAR

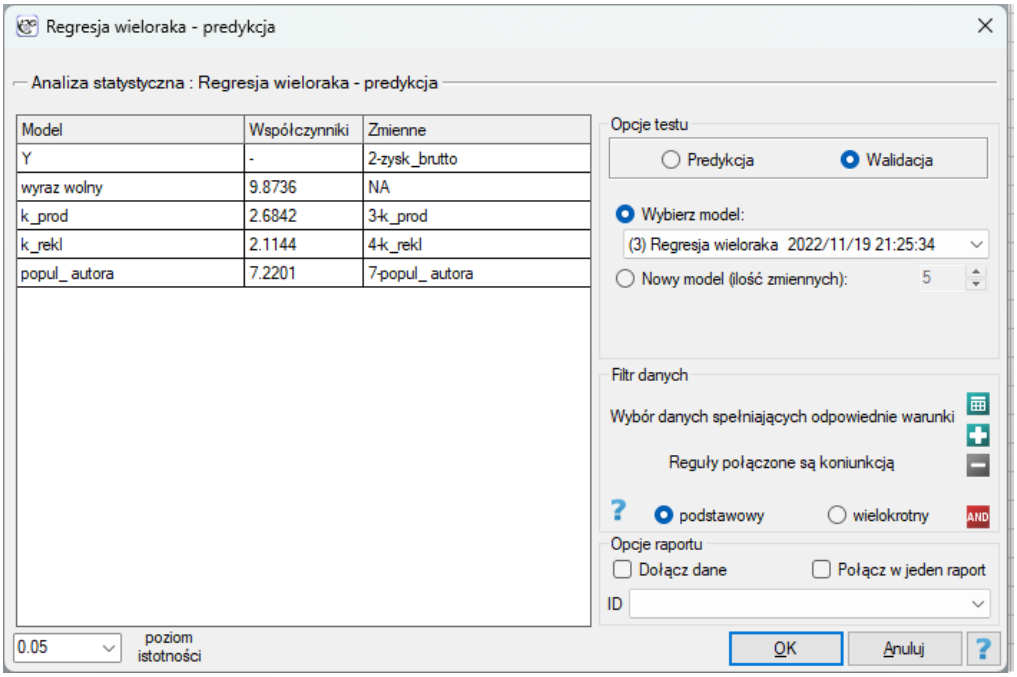
Oszacowana wartość wyliczana jest z pewnym błędem. Dlatego też dodatkowo dla przewidzianej przez model wartości wyznaczane są granice wynikające z błędu:

- dla wartości oczekiwanej wyznaczane są **granice ufności**,
- dla pojedynczego punktu wyznaczane są **granice predykcji**.

JAK USTAWIĆ ANALIZĘ

Statystyki zaawansowane → Modele wielowymiarowe → Regresja wieloraka – predykcja/walidacja.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



By dokonać walidacji lub predykcji, należy wskazać model, na podstawie którego chcemy jej dokonać. Możemy dokonać jej na bazie:

- zbudowanego w PQStat modelu regresji wielorakiej – wystarczy wybrać model spośród modeli przypisanych do danego arkusza, a liczba zmiennych i współczynniki modelu zostaną ustawione automatycznie; zbiór testowy powinien się znaleźć w tym samym arkuszu co zbiór uczący;
- modelu niezbudowanego w programie PQStat, ale uzyskanego z innego źródła (np. opisanego w przeczytanej przez nas pracy naukowej) – w oknie analizy należy podać liczbę zmiennych oraz wpisać współczynniki dotyczące każdej z nich.

W oknie analizy należy wskazać te nowe zmienne, które powinny zostać wykorzystane do walidacji lub predykcji. Predykcja dla jednego obiektu może być wykonywana wraz z budową modelu, czyli w oknie analizy Statystyki zaawansowane→Modele wielowymiarowe→Regresja wieloraka, natomiast predykcja dla większej grupy nowych danych jest wykonywana poprzez menu Statystyki zaawansowane→Modele wielowymiarowe→Regresja wieloraka – predykcja/walidacja.

WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 25.2 (plik wydawca.pqs): do przewidywania zysku brutto ze sprzedaży książek wydawca zbudował model regresji w oparciu o zbiór uczący pozbawiony pozycji 16 (czyli 39 książek). W modelu znalazły się: koszty produkcji, koszty reklamy i popularność autora (1=autor popularny, 0=nie). Zbudujemy raz jeszcze ten model w oparciu o zbiór uczący, a następnie, by się upewnić, że model będzie działał poprawnie, zwalidujemy go na testowym zbiorze danych. Jeśli model przejdzie tę próbę, to będziemy go stosować do predykcji dla nowych pozycji książkowych. By korzystać z odpowiednich zbiorów, ustawiamy każdorazowo filtr danych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja wieloraka - predykcja/walidacja	
Analizowane zmienne	k_prod
	k_rekl
	popul_ autora
	zysk_brutto
Filtr danych	zbiór=uczący
	i nr<>16
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	39
Walidacja	
R (korelacja Y i pred Y)	0.9673344
R2 (korelacja Y i pred Y)	0.9357359
R2 skorygowane (korelacja Y i pred Y)	0.9302275
Błąd standardowy estymacji	4.9180776
Resztowa suma kwadratów	846.5620515
Całkowita suma kwadratów	13173.1707692
Wyjaśniona suma kwadratów	12326.6084199
F	169.8758305
Wartość p	0
AIC - kryterium Akaikego	128.0272524
AICc - skorygowane kryterium Akaikego	130.4978406
BIC - bayesowskie kryterium Schwarza	142.0086223
MAE (średni błąd bezwzględny)	3.6739289
MPE (średni błąd procentowy)	-0.0171597
MAPE (średni bezwzględny błąd procentowy)	0.0941638

Współczynniki modelu	
wyraz wolny	9.8735582
k_prod	2.6842296
k_rekl	2.1143989
popul_ autora	7.2201236

zbiór uczący: $R^2_{adj} = 0.93$, $MAE = 3.8$ **tys. dolarów**

Model jest bardzo dobrze dopasowany do danych, na których został zbudowany.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja wieloraka - predykcja/walidacja	
Analizowane zmienne	k_prod
	k_rekl
	popul_ autora
	zysk_brutto
Filtr danych	zbiór=testowy
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	20
Walidacja	
R (korelacja Y i pred Y)	0.913034
R2 (korelacja Y i pred Y)	0.8336311
R2 skorygowane (korelacja Y i pred Y)	0.802437
Błąd standardowy estymacji	8.0620017
Resztowa suma kwadratów	1039.9339387
Całkowita suma kwadratów	5864.992
Wyjaśniona suma kwadratów	5751.2377808
F	29.4954006
Wartość p	0.0000009
AIC - kryterium Akaikego	87.0236039
AICc - skorygowane kryterium Akaikego	92.6236039
BIC - bayesowskie kryterium Schwarza	96.9979976
MAE (średni błąd bezwzględny)	5.8739802
MPE (średni błąd procentowy)	0.0088586
MAPE (średni bezwzględny błąd procentowy)	0.1354043

Współczynniki modelu	
wyraz wolny	9.8735582
k_prod	2.6842296
k_rekl	2.1143989
popul_ autora	7.2201236

zbiór testowy: $R^2_{adj} = 0.80$**,** $MAE = 5.9$ **tys. dolarów**

Jakość dopasowania na zbiorze testowym jest nieco niższa niż na zbiorze uczącym, ale wciąż zadowalająca – model nadaje się do predykcji.

W tym celu skorzystamy z danych trzech nowych pozycji książkowych dopisanych na końcu zbioru. Wybierzemy opcję Predykcja, ustawimy filtr na nowy zbiór danych i użyjemy naszego modelu do tego, by przewidzieć zysk brutto dla tych książek.

Predykcja				
Pred. Y	-95% CI (pu	+95% CI (pı	-95% CI (wa	+95% CI (w
74.1071	63.6621	84.5521	71.0391	77.1751
24.8392	14.4985	35.1799	22.1477	27.5307
58.7365	48.4804	68.9926	56.3908	61.0822

Okazuje się, że najwyższy zysk brutto (pomiędzy 64 a 85 tys. dolarów) jest prognozowany dla pierwszej, najbardziej reklamowanej i najdrożej wydanej książki popularnego autora.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

25.2 PORÓWNYWANIE MODELI LINIOWEJ REGRESJI WIELORAKIEJ

Porównywanie modeli regresji wielorakiej

KIEDY STOSOWAĆ:

Liniowa regresja wieloraka daje możliwość jednoczesnej analizy wielu zmiennych niezależnych. Pojawia się więc problem wyboru optymalnego modelu. W natłoku informacji, jakie niesie zbyt duży model, istnieje możliwość zagubienia ważnych informacji. Zbyt mały model może natomiast pominąć te cechy, które w wiarygodny sposób mogłyby opisać badane zjawisko – nie liczba zmiennych w modelu, ale ich jakość decyduje o jakości modelu. W wyborze zmiennych niezależnych niezbędna jest wiedza i doświadczenie związane z badanym zjawiskiem. Należy pamiętać, by w modelu znajdowały się zmienne silnie skorelowane ze zmienną zależną i słabo skorelowane między sobą.

DEFINICJE MIAR

Nie istnieje jedna prosta reguła statystyczna, która decydowałaby o liczbie zmiennych niezbędnych w modelu. Najczęściej w porównaniu posługujemy się miarami dopasowania modelu, takimi jak: R^2_{adj} – poprawiona wartość współczynnika determinacji wielorakiej (im wyższa wartość, tym lepiej dopasowany model), SE_e – błąd standardowy estymacji (im niższa wartość, tym lepiej dopasowany model), lub kryteria informacyjne AIC, AICc, BIC (im niższa wartość, tym lepszy model). W tym celu można również wykorzystać test F oparty o współczynnik determinacji wielorakiej R^2, służący do weryfikacji hipotezy, że dopasowanie obu porównywanych modeli jest tak samo dobre.

Jeśli porównywane modele nie różnią się istotnie, powinniśmy wybrać ten z mniejszą liczbą zmiennych – brak różnicy oznacza bowiem, że zmienne, które są w modelu pełnym, a nie ma ich w modelu zredukowanym, nie wnoszą istotnej informacji. Jeśli natomiast różnica w jakości dopasowania modeli jest istotna statystycznie, oznacza to, że jeden z nich (ten z większą liczbą zmiennych, o większym R^2 lub mniejszej wartości kryterium informacyjnego) jest istotnie lepszy niż drugi.

Ręczne porównywanie modeli polega na zbudowaniu 2 modeli:

- pełnego – modelu z większą liczbą zmiennych,
- zredukowanego – modelu z mniejszą liczbą zmiennych; model taki powstaje z modelu pełnego po usunięciu zmiennych, które z punktu widzenia badanego zjawiska są zbędne.

Wybór zmiennych niezależnych w porównywanych modelach, a następnie wybór lepszego modelu na podstawie uzyskanych wyników porównania, należy do badacza.

Automatyczny dobór modelu – program sam buduje kolejne modele i wskazuje najlepszy według wybranego kryterium.

Składniki modelu. Jednostką, którą metody automatyczne dodają do modelu lub z niego usuwają, jest składnik: pojedyncza zmienna, cała zmienna fikcyjna (wszystkie jej kolumny 0–1 naraz) albo interakcja. Obowiązuje zasada hierarchii: interakcja może znaleźć się w modelu tylko wtedy, gdy są w nim jej efekty główne, i tylko ona może zostać usunięta przed nimi.

Kryterium doboru wybieramy w oknie analizy:

- AIC, AICc lub BIC – kryteria informacyjne; lepszy jest model o mniejszej wartości kryterium. Ruch (dodanie lub usunięcie składnika) jest przyjmowany tylko wtedy, gdy obniża kryterium;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- p testu zagnieżdżonego – test F porównujący model przed zmianą i po zmianie, z zadanym progiem p. Składnik jest usuwany, gdy p jest większe od progu, a dodawany, gdy p jest mniejsze od progu. Ta sama wartość p jest pokazywana w raporcie jako porównanie sąsiednich modeli;
- błąd walidacji – błąd predykcji (RMSE) wyznaczony na obserwacjach, które nie brały udziału w dopasowaniu modelu (opis poniżej: Ocena modeli poza zbiorem uczącym). Ruch jest przyjmowany tylko wtedy, gdy zmniejsza ten błąd. Kryterium wymaga włączonej opcji Walidacja – wybranie go przy wyłączonej walidacji włącza walidację krzyżową.

Metody doboru:

- **eliminacja wstecz** – start od modelu ze wszystkimi zmiennymi; w każdym kroku usuwany jest ten składnik, którego usunięcie najbardziej obniża kryterium; koniec, gdy żadne usunięcie już go nie obniża;
- **selekcja w przód** – start od modelu zerowego (tylko wyraz wolny); w każdym kroku dodawany jest składnik najbardziej obniżający kryterium; koniec, gdy żadne dodanie już go nie obniża;
- **krokowa dwukierunkowa** – start od modelu zerowego; w każdym kroku rozważane są zarówno dodania, jak i usunięcia składników i wybierany jest najlepszy ruch; koniec, gdy żaden ruch nie poprawia kryterium;
- **wszystkie podzbiory** – dopasowywane są wszystkie modele zgodne z zasadą hierarchii (maksymalnie 15 składników), a w raporcie prezentowany jest najlepszy model każdej wielkości; metoda korzysta z kryteriów informacyjnych albo z błędu walidacji (kryterium p nie jest w niej dostępne).

Wszystkie modele są liczone na tym samym zbiorze obserwacji – wiersze z brakami w którejkolwiek ze zmiennych kandydujących są odrzucane raz, przed analizą. Dzięki temu kryteria informacyjne i testy porównawcze są między modelami porównywalne.

Raport zawiera: metodę i kryterium doboru, liczbę składników i dopasowanych modeli, kryteria informacyjne modelu zerowego, kolejne modele ścieżki doboru (każdy z miarami dopasowania R^2_{adj}, SE_e, AIC, AICc, BIC oraz tabelą współczynników) i porównanie każdej sąsiedniej pary modeli testem F (statystyka, stopnie swobody, wartość p). Najlepszy model według wybranego kryterium jest oznaczony w raporcie. Jeśli żaden składnik nie poprawia modelu zerowego, raport informuje, że najlepszy jest model zerowy. Przy metodzie wszystkich podzbiorów sąsiednie modele mogą nie być zagnieżdżone – wówczas test F nie jest wykonywany.

Uwaga! Wartości p i przedziały ufności w modelu wybranym automatycznie są zbyt optymistyczne, ponieważ ten sam zbiór danych posłużył do wyboru zmiennych i do oszacowania ich efektów. Dobór automatyczny warto traktować jako pomoc w poszukiwaniu modelu, a ostateczny wybór oprzeć na wiedzy merytorycznej. Wybrany model należy zbudować ponownie w oknie Regresja wieloraka i zweryfikować jego założenia.

Ocena modeli poza zbiorem uczącym

Miary dopasowania (R^2_{adj}, SE_e), kryteria informacyjne i testy są wyznaczane na tych samych obserwacjach, na których oszacowano współczynniki modelu. Taka ocena jest zbyt optymistyczna: model dopasowuje się także do przypadkowych cech próby, więc nowe obserwacje przewiduje zwykle gorzej, niż wynikałoby to z jego dopasowania, a im więcej zmiennych w modelu, tym większa jest ta różnica. Dobór oparty wyłącznie na zbiorze uczącym nie zawsze wskazuje więc model, który najlepiej sprawdzi się na nowych danych. Dlatego w oknie analizy można włączyć ocenę modeli na obserwacjach, które nie brały udziału w ich dopasowaniu (opcja Walidacja):

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **walidacja krzyżowa k-krotna** – obserwacje są losowo dzielone na k części o zbliżonej liczności. Każdy model jest dopasowywany k razy, za każdym razem z pominięciem jednej części, i służy do przewidzenia wartości zmiennej zależnej w pominiętej części. Każda obserwacja zostaje więc przewidziana dokładnie raz – przez model, który jej „nie widział". Modele prezentowane w raporcie (współczynniki, miary dopasowania, kryteria informacyjne, testy) są dopasowane na wszystkich obserwacjach, a miary walidacji są ich uzupełnieniem. Najczęściej stosuje się $k = 5$ lub $k = 10$;
- **losowy zbiór testowy** – zadany procent obserwacji jest losowo odkładany jako zbiór testowy. Wszystkie modele są dopasowywane wyłącznie na pozostałej części (zbiorze uczącym) – to jej dotyczą współczynniki, miary dopasowania, kryteria informacyjne i testy w raporcie – a oceniane na zbiorze testowym. Sposób ten jest prostszy i szybszy, ale wymaga większej próby: modele powstają na mniejszej liczbie obserwacji, a ocena opiera się na jednym, często niewielkim zbiorze.

Podział jest losowy, ale w pełni wyznaczony przez Ziarno: to samo ziarno daje zawsze ten sam podział, więc analizę można powtórzyć, a wszystkie modele jednej analizy są oceniane na identycznym podziale. Zmiana ziarna daje inny podział i nieco inne wartości miar – jeśli po zmianie ziarna zmienia się kolejność modeli, to różnice między nimi są mniejsze niż losowa zmienność podziału i nie należy przywiązywać do nich wagi.

Dla każdego modelu raport podaje wówczas trzy dodatkowe miary:

- **RMSE** – pierwiastek ze średniego kwadratu błędu predykcji, wyrażony w jednostkach zmiennej zależnej; odpowiednik SE_e wyznaczony poza zbiorem uczącym (im mniejsza wartość, tym lepszy model);
- **MAE** – średni błąd bezwzględny predykcji; mniej wrażliwy na pojedyncze duże błędy niż RMSE;
- R^2 **predykcji** – część zmienności zmiennej zależnej wyjaśniona przez przewidywania poza zbiorem uczącym; odpowiednik R^2, zwykle od niego mniejszy. Wartość ujemna oznacza, że model przewiduje nowe obserwacje gorzej niż sama średnia.

Punktem odniesienia jest **model zerowy**, który każdej obserwacji przypisuje średnią zmiennej zależnej z części uczącej – jego RMSE i MAE raport podaje w bloku doboru modelu. Tam też wskazany jest **najlepszy model wg walidacji**, czyli model o najmniejszym RMSE poza zbiorem uczącym. Jest on wskazywany niezależnie od wybranego kryterium doboru, dzięki czemu można sprawdzić, czy model wybrany np. przez AIC jest również tym, który najlepiej przewiduje nowe obserwacje. Ocena poza zbiorem uczącym działa także przy ręcznym porównywaniu modeli – raport podaje wtedy miary obu modeli.

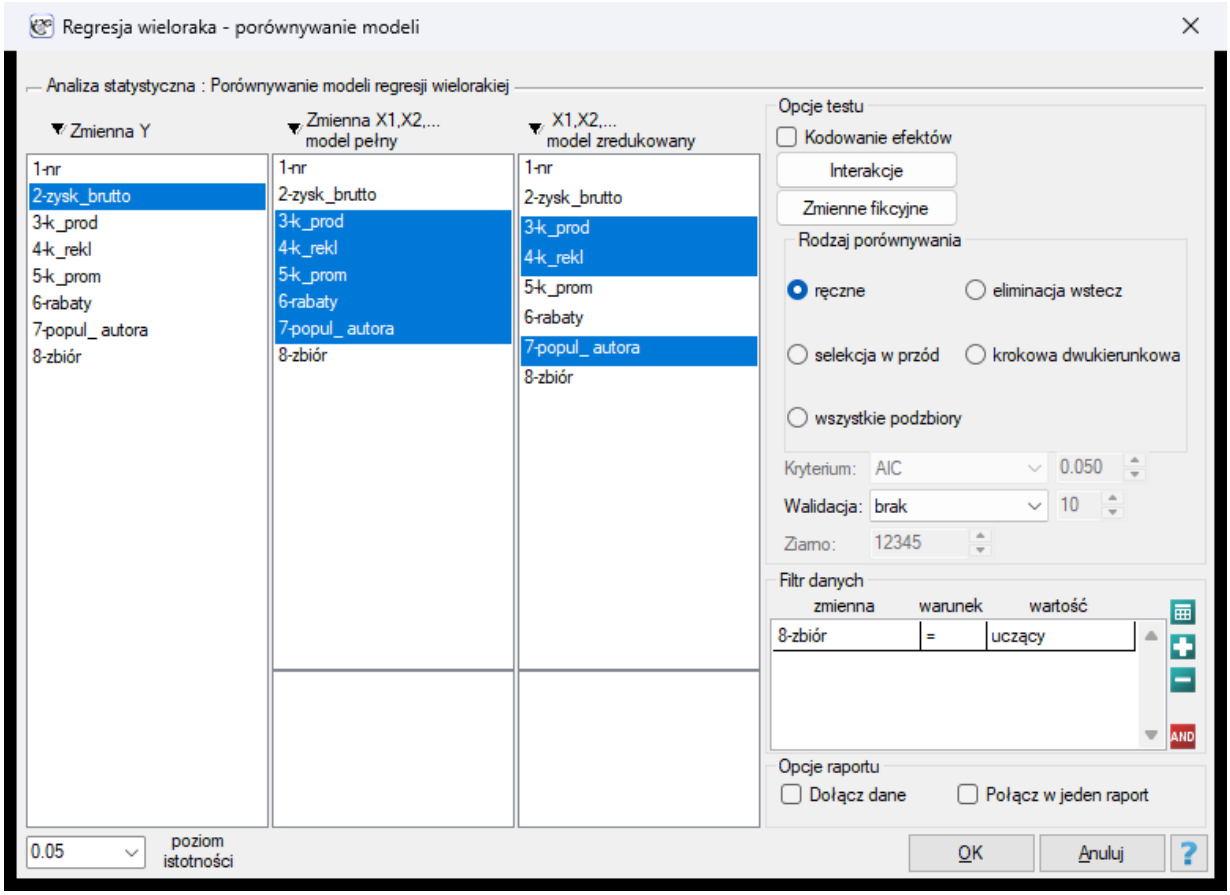
Jeśli jako Kryterium wybrano błąd walidacji, na błędzie poza zbiorem uczącym opierają się również kolejne kroki doboru i oznaczenie najlepszego modelu. Każdy rozważany model jest wtedy dopasowywany dodatkowo k razy, więc obliczenia trwają dłużej.

Uwaga! Gdy błąd walidacji jest kryterium doboru, błąd walidacji wybranego modelu jest nieco zaniżony – spośród wielu modeli wybrano ten, który na danym podziale wypadł najlepiej (raport przypomina o tym odpowiednią uwagą). Rzetelną ocenę ostatecznego modelu dają dopiero dane, które nie brały udziału ani w dopasowaniu, ani w wyborze modelu. Przy małych próbach różnice błędu walidacji między modelami bywają mniejsze niż zmienność wynikająca z losowego podziału – warto wtedy sprawdzić wynik dla kilku wartości ziarna.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele wielowymiarowe → Regresja wieloraka – porównywanie modeli.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



W oknie wskazujemy zmienną zależną Y oraz zmienne niezależne modelu pełnego i zredukowanego (w doborze automatycznym – zmienne kandydujące). Opcje doboru modelu:

- Rodzaj porównywania – ręczne albo jedna z metod doboru automatycznego: eliminacja wstecz, selekcja w przód, krokowa dwukierunkowa, wszystkie podzbiory,
- Kryterium – AIC, AICc, BIC, p (test LR / F) z progiem podawanym w polu obok albo błąd walidacji; w trybie ręcznym kryterium nie jest używane,
- Walidacja – brak (ocena tylko na zbiorze uczącym), krzyżowa, k = z liczbą podziałów w polu obok (od 2 do 100, domyślnie 10) albo zbiór testowy, % = z udziałem zbioru testowego w procentach obserwacji (od 5 do 90, domyślnie 30),
- Ziarno – ziarno losowania podziału danych (domyślnie 12345); to samo ziarno oznacza ten sam podział.

WYKONANIE I INTERPRETACJA

Kontynuacja przykładu 25.2 (plik wydawca.pqs): do przewidywania zysku brutto ze sprzedaży książek wydawca planuje brać pod uwagę takie zmienne jak: koszty produkcji, koszty reklamy, koszty promocji bezpośredniej, suma udzielonych rabatów, popularność autora. Nie wszystkie te zmienne muszą jednak wpływać znacząco na zysk. Spróbujemy wybrać taki model regresji liniowej, który będzie zawierał

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

optymalną (ze statystycznego punktu widzenia) liczbę zmiennych. Do tej analizy użyjemy danych zbioru uczącego.

Na podstawie zbudowanego wcześniej modelu pełnego możemy podejrzewać, że zmienne: koszty promocji bezpośredniej i suma udzielonych rabatów mają niewielki wpływ na budowany model. Sprawdzimy, czy ze statystycznego punktu widzenia model pełny jest lepszy niż model po usunięciu tych dwóch zmiennych.

Porównywanie modeli regresji wielorakiej	
Analizowane zmienne	zysk_brutto
	k_prod
	k_rekl
	k_prom
	rabaty
	popul_ autora
Filtr danych	zbiór=uczący
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	40
Liczba zmiennych w modelu 1	5
Błąd standardowy estymacji	8.0865
R	0.9225
R2	0.851
R2 skorygowane	0.8291
AIC - kryterium Akaikego	172.7149
AICc - skorygowane kryterium Akaikego	177.0786
BIC - bayesowskie kryterium Schwarza	190.226
Liczba zmiennych w modelu 2	3
Błąd standardowy estymacji	8.0725
R	0.918
R2	0.8428
R2 skorygowane	0.8296
AIC - kryterium Akaikego	170.863
AICc - skorygowane kryterium Akaikego	173.263
BIC - bayesowskie kryterium Schwarza	184.9963
F - porównanie modeli	0.9379
DF1	2
DF2	34
Wartość p	0.4013

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Model 1

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	wartość p	stand. b	błąd stand
w. wolny	4.175186	4.772779	-5.524268	13.874639	0.874791	0.387825		
k_prod	2.560709	0.501507	1.541525	3.579894	5.106033	0.000013	0.422818	0.082807
k_rekl	1.998235	0.359065	1.268527	2.727943	5.565106	0.000003	0.461287	0.082889
k_prom	4.668238	4.791644	-5.069555	14.40603	0.974245	0.336816	0.066242	0.067993
rabaty	1.423171	1.404811	-1.431749	4.278091	1.013069	0.318183	0.067478	0.066608
popul_ aut	10.153717	2.782567	4.498861	15.808574	3.649047	0.000874	0.261561	0.071679

Model 2

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	wartość p	stand. b	błąd stand
w. wolny	8.85096	3.277587	2.203704	15.498215	2.70045	0.010486		
k_prod	2.519802	0.492781	1.520396	3.519209	5.113431	0.000011	0.416063	0.081367
k_rekl	2.074037	0.350655	1.362876	2.785198	5.914752	0.000001	0.478786	0.080948
popul_ aut	9.921455	2.771625	4.30034	15.54257	3.579653	0.001007	0.255578	0.071397

test F: $p = 0.4013$ **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Nie ma podstaw, by uważać, że model pełny jest lepszy niż model zredukowany.

Dodatkowo, model zredukowany jest nieco lepiej dopasowany niż model pełny (dla modelu zredukowanego $R^2_{adj} = 0.8296$, dla modelu pełnego $R^2_{adj} = 0.8291$) oraz ma mniejsze, czyli korzystniejsze, wartości kryteriów informacyjnych AIC, AICc i BIC.

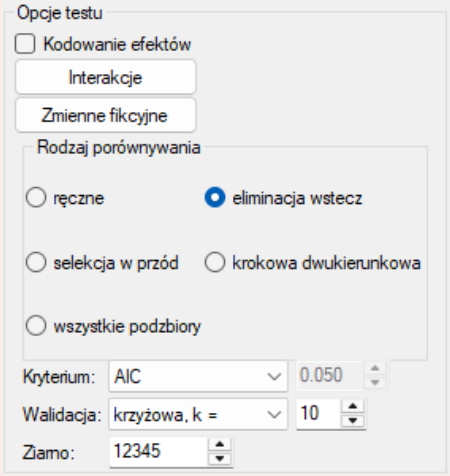
Automatyczny dobór modelu

Dobór automatyczny (eliminacja wstecz, kryterium AIC) daje bardzo podobne wyniki. Najlepszym modelem jest model o największym współczynniku R^2_{adj}, najmniejszych kryteriach informacyjnych i najmniejszym błędzie standardowym estymacji SE_e. Sugerowanym najlepszym modelem jest tu model zawierający tylko 3 zmienne niezależne: koszty produkcji, koszty reklamy, popularność autora.

Ocena poza zbiorem uczącym

Dotychczasowe porównania opierały się na tych samych 40 książkach, na których dopasowano modele. Sprawdzimy, czy wybór się potwierdzi, gdy modele będą oceniane na obserwacjach, które nie brały udziału w ich dopasowaniu. W oknie analizy zostawiamy eliminację wstecz i kryterium AIC, a w polu Walidacja wybieramy krzyżowa, k = i wpisujemy 10 podziałów (ziarno 12345).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



W bloku doboru modelu raport podaje sposób oceny, błąd modelu zerowego i model najlepszy według walidacji, a przy każdym modelu – trzy wiersze miar walidacji krzyżowej.

Ocena poza zbiorem uczącym	walidacja krzyżowa
Liczba podziałów (k)	10
Ziarno losowania podziału	12345
Model zerowy: RMSE	19.7867
Model zerowy: MAE	15.9297
Najlepszy model wg walidacji (RMSE)	Model 2

Liczba zmiennych w modelu 3 (najlepszy)	3
Błąd standardowy estymacji	8.0725
R	0.918
R2	0.8428
R2 skorygowane	0.8296
F - porównanie modeli Model 2 vs Model 3	0.928
DF1	1
DF2	35
Wartość p	0.342
AIC - kryterium Akaikego	170.863
AICc - skorygowane kryterium Akaikego	173.263
BIC - bayesowskie kryterium Schwarza	184.9963
Walidacja krzyżowa: RMSE	8.2794
Walidacja krzyżowa: MAE	4.9911
Walidacja krzyżowa: R2 predykcji	0.8249

		zbiór uczący			walidacja krzyżowa ($k = 10$)		
model	zmienne	SE_e	R^2_{adj}	AIC	RMSE	MAE	R^2 predykcji
zerowy	0				19.7867	15.9297	
Model 1	5	8.0865	0.8291	172.7149	8.3425	5.2976	0.8222
Model 2	4	8.0806	0.8293	171.8163	8.2309	5.1237	0.8270
Model 3	3	8.0725	0.8296	170.8630	8.2794	4.9911	0.8249

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Model 1 to model pełny, Model 2 powstał z niego po usunięciu kosztów promocji bezpośredniej, a Model 3 – po usunięciu dodatkowo sumy rabatów. Wszystkie trzy przewidują zysk wielokrotnie lepiej niż model zerowy (RMSE około 8.3 wobec 19.8), a R^2 predykcji (0.82–0.83) jest tylko nieznacznie niższe od R^2_{adj} – modele nie są więc przeuczone. Błąd poza zbiorem uczącym jest przy tym nieco większy niż błąd standardowy estymacji (dla Modelu 3: 8.28 wobec 8.07), czego należało się spodziewać.

Najmniejszy błąd walidacji ma Model 2 (RMSE $= 8.2309$), a wskazany przez AIC Model 3 jest tuż za nim (RMSE $= 8.2794$) i ma najmniejszy średni błąd bezwzględny (MAE $= 4.9911$). Różnice między tymi modelami są znikome w porównaniu z samą wielkością błędu, natomiast model pełny – choć na zbiorze uczącym jest najlepiej dopasowany (ma największe R^2) – nowe obserwacje przewiduje najsłabiej z całej trójki (RMSE $= 8.3425$). Usunięcie kosztów promocji bezpośredniej i sumy rabatów nie pogarsza więc przewidywań, a walidacja potwierdza wybór modelu z 3 zmiennymi.

walidacja krzyżowa ($k = 10$): RMSE $= 8.34$ (Model 1), 8.23 (Model 2), 8.28 (Model 3); model zerowy: 19.79

Mniejsze modele przewidują nowe obserwacje nie gorzej niż model pełny – ocena poza zbiorem uczącym potwierdza wybór modelu z 3 zmiennymi.

Na podstawie powyższych analiz, ze statystycznego punktu widzenia, optymalnym modelem jest model zawierający 3 najważniejsze zmienne niezależne: koszty produkcji, koszty reklamy, popularność autora. Jednak ostateczna decyzja, który model wybrać, należy do osoby posiadającej specjalistyczną wiedzę z zakresu badania – w tym przypadku wydawcy. Należy pamiętać, że wybrany model powinien zostać ponownie zbudowany, a jego założenia zweryfikowane w oknie Regresja wieloraka.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad R^2_{FM} = R^2_{RM}, \\ \mathcal{H}_1: &\quad R^2_{FM} \neq R^2_{RM}, \end{aligned}$$

gdzie R^2_{FM}, R^2_{RM} – współczynniki determinacji wielorakiej w porównywanych modelach (pełnym i zredukowanym).

Wzory: statystyka testowa ma postać:

$$F = \frac{R^2_{FM} - R^2_{RM}}{k_{FM} - k_{RM}} \cdot \frac{n - k_{FM} - 1}{1 - R^2_{FM}}.$$

Statystyka ta podlega rozkładowi F Snedecora z $df_1 = k_{FM} - k_{RM}$ i $df_2 = n - k_{FM} - 1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Ocena poza zbiorem uczącym
Niech $\hat{y}_{(-i)}$ oznacza wartość przewidzianą dla obserwacji i przez model dopasowany bez tej części danych, do której należy obserwacja i (w walidacji krzyżowej – bez jej podziału, przy zbiorze testowym – na zbiorze uczącym), a $\overline{y}_{(-i)}$ – średnią zmiennej zależnej w tej samej części uczącej. Miary wyznaczane

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

są łącznie dla wszystkich n_v ocenianych obserwacji (w walidacji krzyżowej – wszystkich obserwacji, przy zbiorze testowym – obserwacji zbioru testowego):

$$RMSE = \sqrt{\frac{1}{n_v}\sum_{i=1}^{n_v}\left(y_i - \hat{y}_{(-i)}\right)^2}, \qquad MAE = \frac{1}{n_v}\sum_{i=1}^{n_v}\left|y_i - \hat{y}_{(-i)}\right|,$$

$$R^2_{pred} = 1 - \frac{\sum_{i=1}^{n_v}\left(y_i - \hat{y}_{(-i)}\right)^2}{\sum_{i=1}^{n_v}\left(y_i - \overline{y}_{(-i)}\right)^2}.$$

Mianownik R^2_{pred} to suma kwadratów błędów modelu zerowego, dlatego R^2_{pred} pokazuje, o jaką część model zmniejsza błąd predykcji w stosunku do modelu zerowego.

Podział: obserwacje są losowo tasowane (generator liczb pseudolosowych o stanie początkowym równym ziarnu), a następnie – w walidacji krzyżowej – przydzielane kolejno do k części, których liczności różnią się najwyżej o 1; przy zbiorze testowym do zbioru testowego trafia zadany procent obserwacji (po zaokrągleniu do liczby całkowitej). Gdy k jest większe od liczby obserwacji, przyjmowane jest k równe liczbie obserwacji. Jeśli obserwacji jest zbyt mało, by wydzielić część danych do walidacji, program informuje o tym komunikatem.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

25.3 REGRESJA LOGISTYCZNA

Regresja logistyczna

KIEDY STOSOWAĆ:

Budowany model regresji logistycznej (podobnie jak liniowej regresji wielorakiej) pozwala na zbadanie wpływu wielu zmiennych niezależnych ($X_1, X_2, ..., X_k$) na jedną zmienną zależną (Y). Tym razem jednak zmienna zależna przyjmuje jedynie dwie wartości, np. chory/zdrowy, niewypłacalny/wypłacalny itp. Owe dwie wartości kodowane są jako (1)/(0), gdzie (1) to wartość wyróżniona – posiadanie danej cechy, a (0) – brak danej cechy.

Funkcja, na której oparty jest model regresji logistycznej, wylicza nie dwupoziomową zmienną Y, a prawdopodobieństwo przyjęcia przez tę zmienną wyróżnionej wartości:

$$P(Y = 1|X_1, X_2, ..., X_k) = \frac{e^Z}{1 + e^Z},$$

gdzie:
$P(Y = 1|X_1, X_2, ..., X_k)$ – prawdopodobieństwo przyjęcia wartości wyróżnionej (1) pod warunkiem uzyskania konkretnych wartości zmiennych niezależnych, tzw. prawdopodobieństwo przewidywane dla 1,
Z najczęściej wyrażone jest zależnością liniową: $Z = \beta_0 + \sum_{i=1}^{k} \beta_i X_i$,
$X_1, X_2, \ldots X_k$ – zmienne niezależne, objaśniające,
$\beta_0, \beta_1, \beta_2, \ldots \beta_k$ – parametry.

Logitem nazywamy przekształcenie tego modelu do postaci:

$$\ln\left(\frac{P}{1-P}\right) = Z.$$

WARUNKI STOSOWANIA

- zmienna zależna Y dychotomiczna, zakodowana jako (1)/(0),
- zmienne niezależne niewspółliniowe,
- odpowiednio duża liczba obserwacji.

Uwaga! Funkcja Z może być również opisana zależnością wyższego stopnia, np. kwadratową – do modelu wprowadzamy wówczas zmienną zawierającą kwadrat danej zmiennej niezależnej X_i^2.

Budując model, należy pamiętać, że liczba obserwacji powinna być dziesięciokrotnie większa lub równa liczbie szacowanych parametrów modelu ($n \geq 10(k + 1)$). Jednak coraz częściej stosuje się bardziej restrykcyjne kryterium zaproponowane przez P. Peduzzi i innych w roku 1996[135], mówiące, iż liczba obserwacji powinna być dziesięciokrotnie większa lub równa stosunkowi liczby zmiennych niezależnych (v) i mniejszej z proporcji liczności (p) opisanych przez zmienną zależną (tzn. proporcji chorych lub zdrowych), czyli $n \geq 10v/p$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Budując model, należy pamiętać, że zmienne niezależne nie powinny być współliniowe. W przypadku gdy występuje współliniowość, estymacja może być niepewna, a uzyskane wartości błędów bardzo wysokie. Zmienne współliniowe należy usunąć z modelu bądź zbudować z nich jedną zmienną niezależną, np. zamiast współliniowych zmiennych: wiek matki i wiek ojca, można zbudować zmienną wiek rodziców.

Kryterium zbieżności funkcji algorytmu iteracyjnego Newtona-Raphsona można kontrolować przy pomocy dwóch parametrów: limitu iteracji zbieżności (podaje maksymalną liczbę iteracji, w jakiej algorytm powinien osiągnąć zbieżność) i kryterium zbieżności (podaje wartość, poniżej której uzyskana poprawa estymacji uznana będzie za nieznaczną i algorytm zakończy działanie).

DEFINICJE MIAR

Zmienne fikcyjne i interakcje w modelu – omówienie przygotowania zmiennych fikcyjnych i interakcji przedstawiono w rozdziale 25.0.2 Przygotowanie zmiennych do analizy w modelach wielowymiarowych.

Współczynniki regresji – rozwiązaniem równania modelu jest wektor ocen parametrów $\beta_0, \beta_1, \ldots, \beta_k$, nazywanych **współczynnikami regresji** ($b_0, b_1, ..., b_k$), szacowanych poprzez **metodę największej wiarygodności**, czyli poprzez poszukiwanie maksimum funkcji wiarygodności L (w programie użyto algorytmu iteracyjnego Newton-Raphson). Na podstawie tych wartości możemy wnioskować o wielkości wpływu zmiennej niezależnej (dla której ten współczynnik został oszacowany) na zmienną zależną. Każdy współczynnik obarczony jest pewnym błędem szacunku, wyliczanym ze wzoru $SE_b = \sqrt{diag(H^{-1})_b}$, gdzie $diag(H^{-1})$ to główna przekątna macierzy kowariancji.

Jednostkowy Iloraz Szans – na podstawie współczynników, dla każdej zmiennej niezależnej w modelu, wylicza się łatwą w interpretacji miarę, jaką jest jednostkowy Iloraz Szans:

$$OR_i = e^{\beta_i}.$$

Otrzymany Iloraz Szans wyraża zmianę szansy na wystąpienie wyróżnionej wartości (1), gdy zmienna niezależna rośnie o 1 jednostkę. Wynik ten jest skorygowany o pozostałe zmienne niezależne znajdujące się w modelu w ten sposób, że zakłada, iż pozostają one na stałym poziomie, podczas gdy badana zmienna niezależna rośnie o jednostkę. Wartość OR interpretujemy następująco:

- $OR > 1$ oznacza stymulujący wpływ badanej zmiennej niezależnej na uzyskanie wyróżnionej wartości (1), tj. mówi, o ile wzrasta szansa na wystąpienie wyróżnionej wartości (1), gdy zmienna niezależna wzrasta o jeden poziom;
- $OR < 1$ oznacza destymulujący wpływ badanej zmiennej niezależnej na uzyskanie wyróżnionej wartości (1), tj. mówi, o ile spada szansa na wystąpienie wyróżnionej wartości (1), gdy zmienna niezależna wzrasta o jeden poziom;
- $OR \approx 1$ oznacza, że badana zmienna niezależna nie ma wpływu na uzyskanie wyróżnionej wartości (1).

U-smile – wykres czterech podklas (nie-zdarzenia: lepsza i gorsza predykcja; zdarzenia: gorsza i lepsza predykcja) z wybranym współczynnikiem (rLR, BA lub RB) na osi Y; rozmiar punktu odpowiada udziałowi podklasy I, linie klas są ciągłe przy istotnej poprawie (dla rLR), a pozioma linia odniesienia oznacza brak zmiany. Uśmiech oznacza poprawę predykcji względem modelu zerowego w obu klasach (patrz U-smile).

Iloraz Szans – wzór ogólny – program PQStat wylicza jednostkowy Iloraz Szans. Jego modyfikacja, na podstawie ogólnego wzoru, umożliwia zmianę interpretacji uzyskanego wyniku. Iloraz szans na wystą-

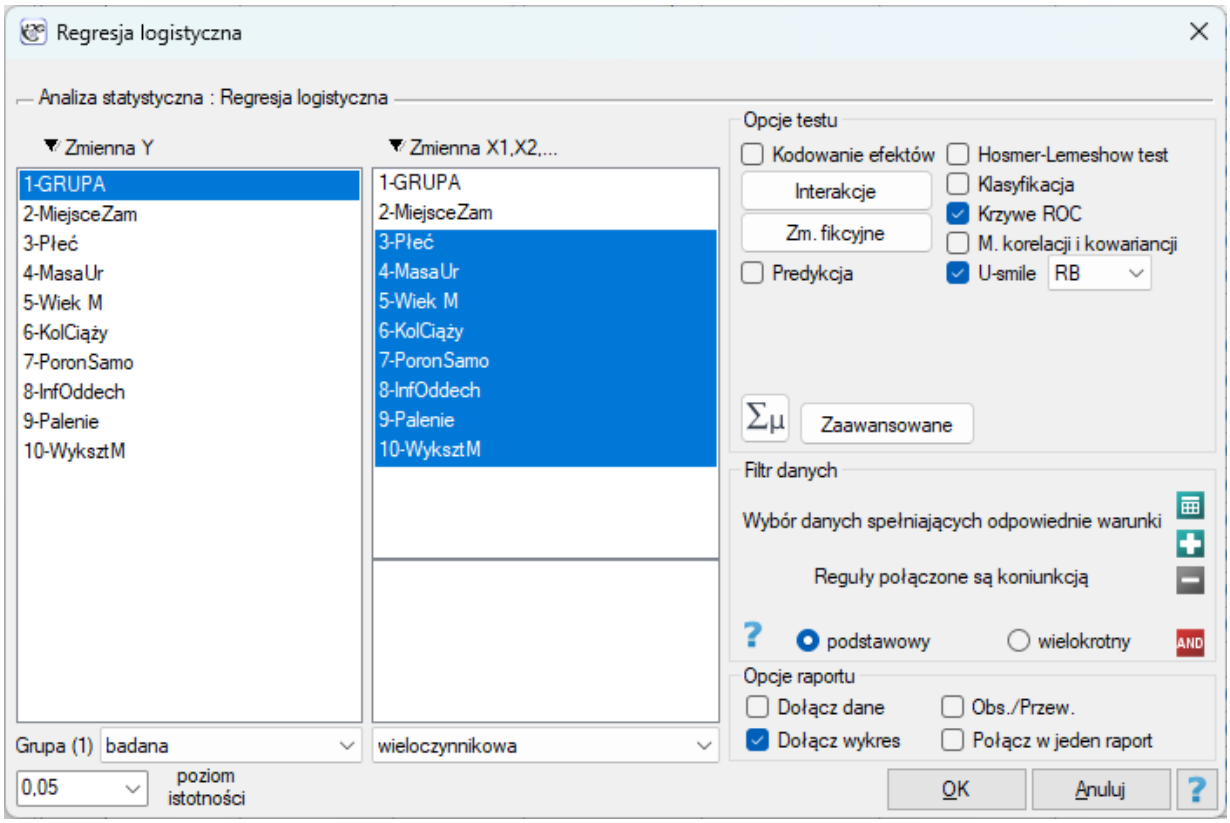
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

pienie stanu wyróżnionego w ogólnym przypadku jest wyliczany jako iloraz dwóch szans wyznaczonych dla dwóch kategorii porównywanej zmiennej niezależnej.

Uwaga! Jeśli analizę przeprowadzamy dla modelu innego niż liniowy lub uwzględniamy interakcję, wówczas na podstawie ogólnego wzoru możemy wyliczyć odpowiedni Iloraz Szans, zmieniając formułę wyrażającą Z.

JAK USTAWIĆ ANALIZĘ

Statystyka zaawansowana → Modele wielowymiarowe → Regresja logistyczna.



Weryfikacja modelu

Istotność statystyczna poszczególnych zmiennych w modelu (istotność ilorazu szans) – na podstawie współczynnika oraz jego błędu szacunku możemy wnioskować, czy zmienna niezależna, dla której ten współczynnik został oszacowany, wywiera istotny wpływ na zmienną zależną. W tym celu posługujemy się testem Walda.

Jakość zbudowanego modelu – dobry model powinien spełniać dwa podstawowe warunki: powinien być dobrze dopasowany i możliwie jak najprostszy. Jakość modelu regresji logistycznej możemy ocenić kilkoma miarami, opartymi na: L_{FM} – maksimum funkcji wiarygodności modelu pełnego (z wszystkimi zmiennymi), L_0 – maksimum funkcji wiarygodności modelu zawierającego jedynie wyraz wolny, n – liczności próby:

- **Kryteria informacyjne** (AIC, $AICc$, BIC) – opierają się na entropii informacji niesionej przez model. Powinniśmy wybierać model o minimalnej wartości danego kryterium informacyjnego.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- **Pseudo R^2** – tzw. McFadden R^2, jest miarą dopasowania modelu (odpowiednikiem współczynnika determinacji wielorakiej R^2 wyznaczanego dla liniowej regresji wielorakiej). Wartość tego współczynnika mieści się w przedziale $\langle 0; 1)$, gdzie wartości bliskie 1 oznaczają doskonałe dopasowanie modelu. Ponieważ współczynnik R^2_{Pseudo} nie przyjmuje wartości 1 i jest wrażliwy na liczbę zmiennych w modelu, wyznacza się jego poprawione wartości: $R^2_{Nagelkerke}$ i $R^2_{Cox-Snell}$.
- **Istotność statystyczna wszystkich zmiennych w modelu** – podstawowym narzędziem szacującym istotność wszystkich zmiennych w modelu jest test ilorazu wiarygodności.
- **Test Hosmera-Lemeshowa** – dla różnych podgrup danych porównuje obserwowane liczności występowania wartości wyróżnionej O_g i przewidywane prawdopodobieństwo E_g. Jeśli O_g i E_g są wystarczająco bliskie, można założyć, że zbudowano dobrze dopasowany model. Do obliczeń najpierw obserwacje są dzielone na G podgrup, zwykle na decyle ($G = 10$).
- **AUC – pole pod krzywą ROC** – krzywa ROC, zbudowana w oparciu o wartość zmiennej zależnej oraz przewidywane prawdopodobieństwo zmiennej zależnej P, pozwala na ocenę zdolności zbudowanego modelu regresji logistycznej do klasyfikacji przypadków do dwóch grup: (1) i (0). Dodatkowo, dla krzywej ROC podawana jest proponowana wartość **punktu odcięcia** prawdopodobieństwa przewidywanego oraz tabela podająca wielkość czułości i swoistości dla każdego możliwego punktu odcięcia.

Uwaga! Więcej możliwości w wyliczeniu punktu odcięcia daje moduł Krzywa ROC. Analizę przeprowadzamy na podstawie wartości obserwowanych i prawdopodobieństwa przewidywanego, które uzyskujemy w analizie regresji logistycznej.

Klasyfikacja – na podstawie wybranego punktu odcięcia prawdopodobieństwa przewidywanego można sprawdzić jakość klasyfikacji. Punkt odcięcia to domyślnie wartość 0.5. Użytkownik może zmienić tę wartość na dowolną wartość z przedziału $(0, 1)$, np. wartość sugerowaną przez krzywą ROC. W wyniku uzyskamy tabelę klasyfikacji oraz procent poprawnie zaklasyfikowanych przypadków, procent poprawnie zaklasyfikowanych (0) – swoistość, oraz procent poprawnie zaklasyfikowanych (1) – czułość.

U-smile – ocena modelu względem modelu zerowego – metoda U-smile (*Universal Smile Layout for Explanation*, Kubiak i in. 2024[182], Więckowska i in. 2025[183], 2026[184]) porównuje dwa modele przewidujące prawdopodobieństwo zdarzenia na tych samych obserwacjach. W oknie regresji logistycznej modelem odniesienia jest model zerowy (tylko wyraz wolny, czyli częstość zdarzeń w próbie), a modelem nowym – zbudowany model. Obserwacje dzieli się na cztery podklasy:

- nie-zdarzenia (0) z lepszą predykcją (prawdopodobieństwo zmalało),
- nie-zdarzenia (0) z gorszą predykcją,
- zdarzenia (1) z gorszą predykcją,
- zdarzenia (1) z lepszą predykcją (prawdopodobieństwo wzrosło).

Dla każdej podklasy wyznaczany jest:

- współczynnik jakościowy dotyczący liczby zmian I (udział obserwacji podklasy w klasie),
- oraz do wyboru jeden ze współczynników ilościowych dotyczących wielkości zmian:
  - rLR (względny iloraz wiarygodności) – znormalizowany do przedziału $\langle 0; 1\rangle$,

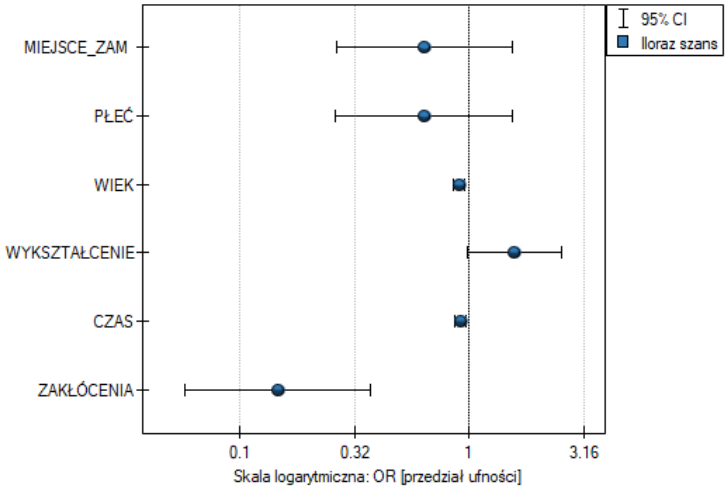
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- BA (*Brier Alteration*) – średnia zmiana kwadratu reszty,
- RB (*Relative Brier*) – zmiana względem błędu Briera modelu odniesienia.

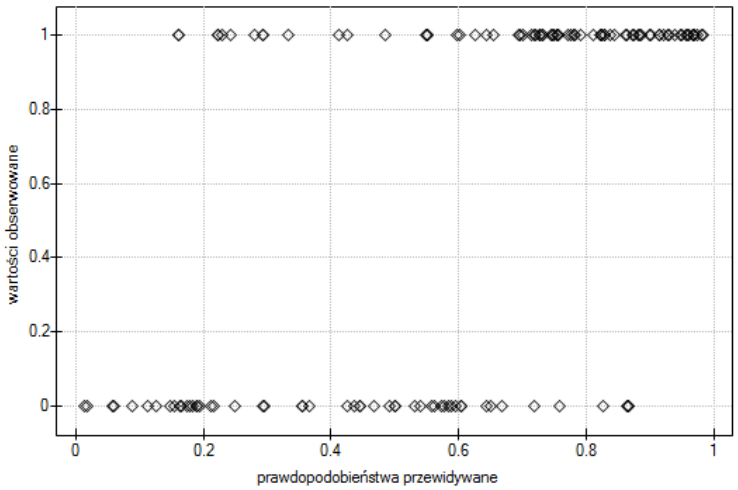
Współczynniki netto dla klas (poprawa minus pogorszenie) i **współczynnik ogólny** pokazują, o ile model poprawia predykcję osobno dla zdarzeń i nie-zdarzeń.
Istotność sprawdza test ilorazu wiarygodności (ogółem), a dla rLR dodatkowo testy dla każdej klasy oparte na rozkładzie gamma.
Wzory: dla zainteresowanych.

Wykresy w regresji logistycznej

- **Iloraz szans $\pm$ przedział ufności** – wykres przedstawiający OR wraz z 95% przedziałem ufności dla wyniku każdej zmiennej zwróconej w zbudowanym modelu. Dla zmiennych kategorialnych linia na poziomie 1 wskazuje wartość ilorazu szans dla kategorii referencyjnej.

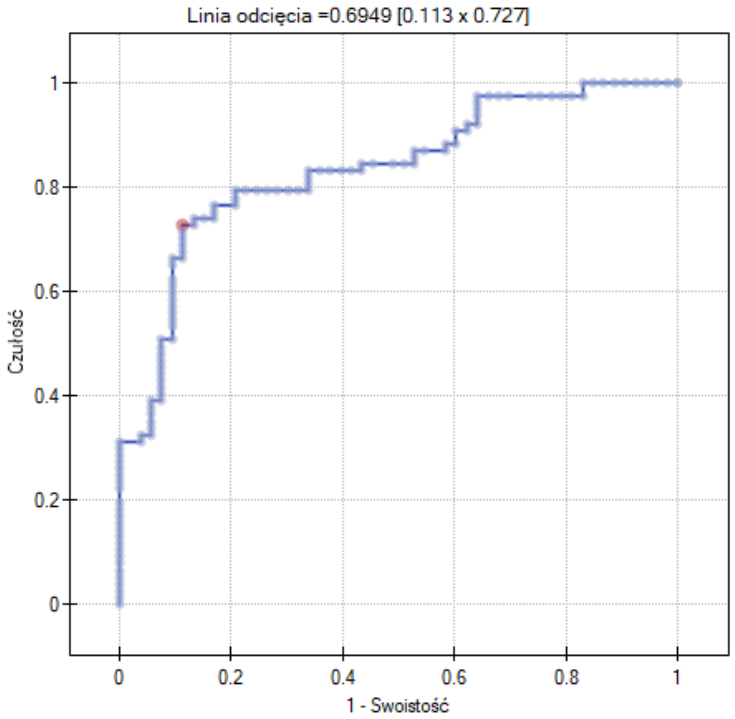


- **Wartości obserwowane / Prawdopodobieństwo oczekiwane** – wykres przedstawiający wyniki przewidywanego dla każdej osoby prawdopodobieństwa wystąpienia zdarzenia (oś X) oraz wartości prawdziwej, czyli wystąpienia zdarzenia (wartość 1 na osi Y) lub braku zdarzenia (wartość 0 na osi Y). Jeśli model bardzo dobrze prognozuje, to przy lewej stronie wykresu punkty będą się kumulowały w dolnej części, a przy prawej stronie – w górnej.

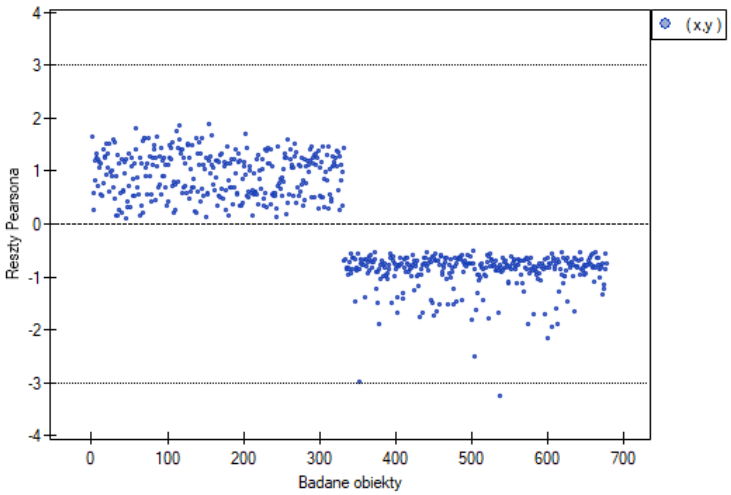


Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **Krzywa ROC** – wykres zbudowany w oparciu o wartość zmiennej zależnej oraz przewidywane prawdopodobieństwo wystąpienia zdarzenia.



- **Wykres reszt Pearsona** – wykres pozwalający ocenić, czy występują odstające dane. Reszty to różnice wartości obserwowanej i przewidywanego przez model prawdopodobieństwa. Wykresy surowych reszt z regresji logistycznej trudno interpretować, dlatego ujednolica się je, wyznaczając reszty Pearsona (surowa reszta podzielona przez pierwiastek kwadratowy z funkcji wariancji). Znak (dodatni lub ujemny) wskazuje, czy obserwowana wartość jest wyższa czy niższa niż wartość dopasowana do modelu, a wielkość wskazuje stopień odchylenia. Reszty Pearsona mniejsze niż -3 lub większe niż 3 sugerują zbyt duże odchylenie danego obiektu.



- **Jednostkowe zmiany ilorazu szans** – wykres przedstawiający serie ilorazów szans wraz z przedziałem ufności, wyznaczane dla każdego z możliwych punktów odcięcia zmiennej umieszczonej

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

na osi X. Umożliwia użytkownikowi wybór jednego, dobrego punktu odcięcia i następnie zbudowanie na tej podstawie nowej dwuwartościowej zmiennej. Wykres dedykowany jest dla oceny zmiennych ciągłych w analizie jednoczynnikowej.

- **Profile ilorazu szans** – wykres przedstawiający serie ilorazów szans wraz z przedziałem ufności, wyznaczane dla wskazanej wielkości okna, tzn. porównujące częstości wewnątrz okna z częstościami umieszczonymi na zewnątrz okna. Najlepiej sprawdza się, gdy poszukujemy funkcji U-kształtnej, tzn. o wysokim ryzyku przy niskich i przy wysokich wartościach badanej zmiennej, a o niskim – przy wartościach przeciętnych. Im szersze okno, tym większe uogólnienie wyników i gładsza funkcja ilorazu szans; im węższe okno, tym bardziej szczegółowe, ale bardziej rozchwiane wyniki. Wykres dedykowany jest dla oceny zmiennych ciągłych w analizie jednoczynnikowej.

- **Wykres U-smile** – na wykresie U-smile cztery podklasy tworzą „uśmiech", gdy model poprawia predykcję w obu klasach:

  – (oś Y) wielkość uśmiechu pokazuje rozmiar poprawy (do wyboru współczynniki rLR, BA, RB),

  – (oś X) od lewej strony przedstawia klasę braku zdarzeń (kolor niebieski), od prawej klasę zdarzeń (kolor czerwony); punkty skrajne to podklasa: poprawa predykcji, a punkty środkowe to podklasa: pogorszenie predykcji,

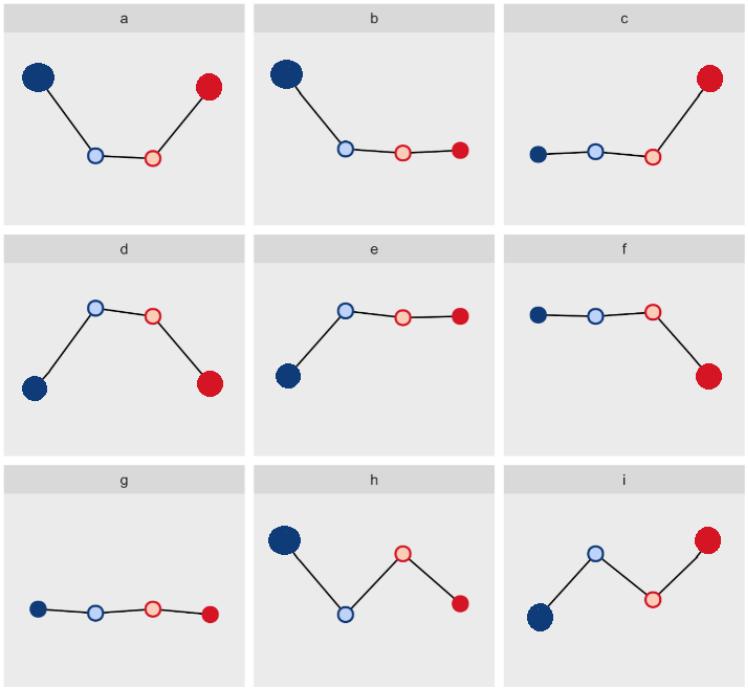
  – symetria (lewa strona wykresu podobna do prawej) oznacza, że poprawa/pogorszenie predykcji dotyczy obu klas w taki sam sposób,

  – wielkość punktu odpowiada współczynnikowi I, czyli wskazuje, jak duży odsetek obiektów znajduje się w każdej podklasie,

  – linia klasy jest ciągła, gdy poprawa w tej klasie jest istotna (tylko rLR).

  Przykłady możliwych kształtów wykresu U-smile. Poprawa przewidywania: (a) dla obu klas wyników, (b) tylko dla braku zdarzeń, (c) tylko dla zdarzeń. Pogorszenie przewidywania: (d) dla obu klas wyników, (e) tylko dla braku zdarzeń, (f) tylko dla zdarzeń. Okręgi leżące na mniej więcej stałym poziomie oznaczają brak poprawy lub pogorszenia przewidywania w porównaniu z modelem odniesienia (g). Zygzak wskazuje na poprawę przewidywania w jednej klasie wyników i pogorszenie przewidywania w drugiej (h i i).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

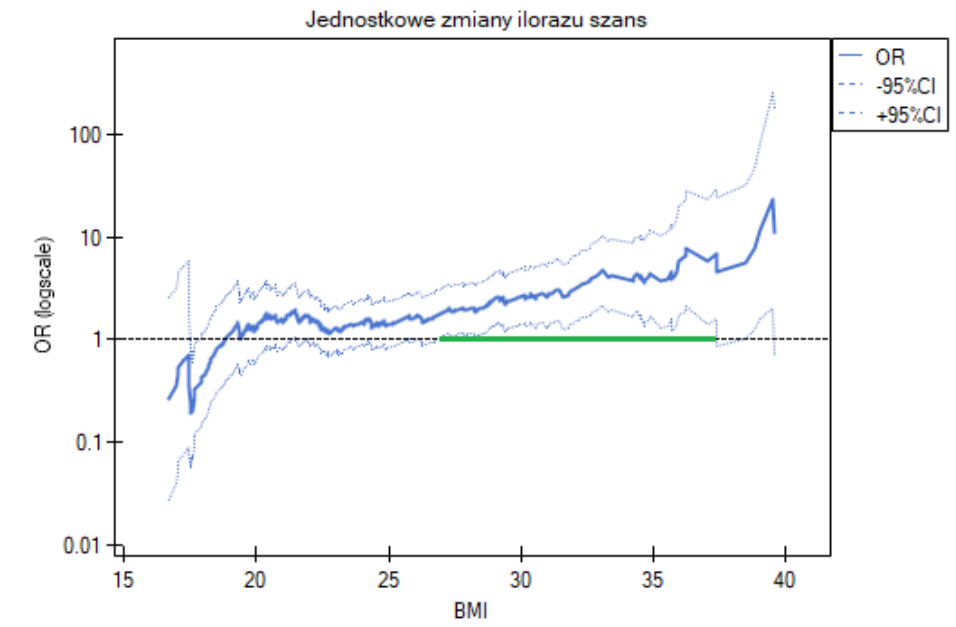


WYKONANIE I INTERPRETACJA

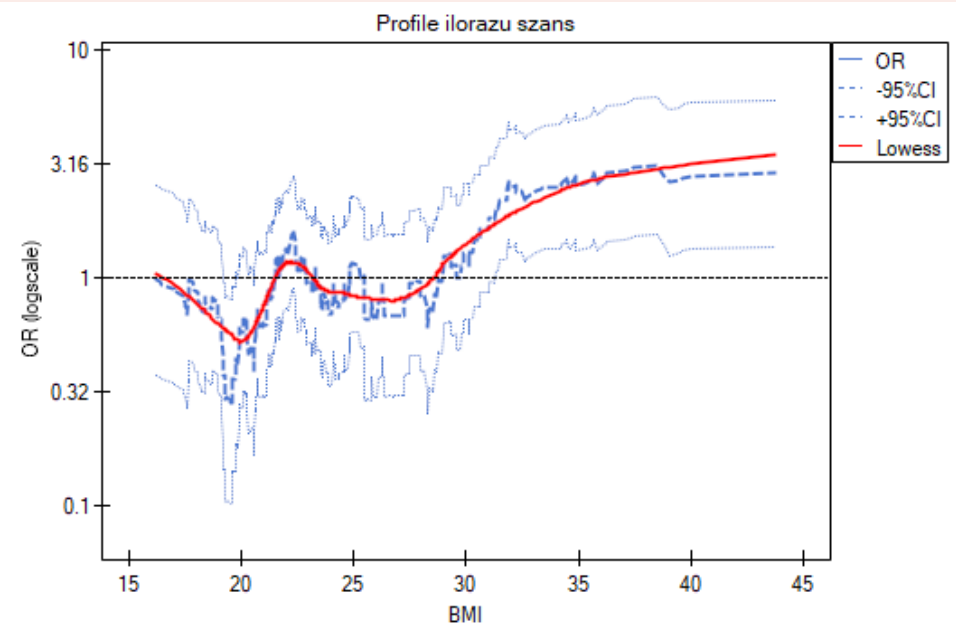
Przykład 25.3. (plik Profile OR.pqs)
Badamy ryzyko występowania choroby A i choroby B w zależności od BMI pacjenta. Ponieważ BMI jest zmienną ciągłą, jej umieszczenie w modelu skutkuje wyznaczeniem jednostkowego ilorazu szans wyznaczającego liniowy trend wzrostu lub spadku ryzyka. Nie wiemy, czy model liniowy będzie dobrym modelem dla analizy tego ryzyka, dlatego przed budowaniem wielowymiarowych modeli regresji logistycznej zbudujemy kilka modeli jednowymiarowych prezentujących tę zmienną na wykresach, by móc ocenić kształt badanej zależności. Do tego celu posłużą wykresy jednostkowych zmian ilorazu szans i profili ilorazu szans, przy czym dla profili wybierzemy okno o wielkości 100, ponieważ prawie każdy pacjent ma inne BMI, więc około 100 pacjentów znajdzie się w każdym oknie.

Choroba A
Jednostkowe zmiany ilorazu szans pokazują, że gdy punkt odcięcia BMI wybierzemy gdzieś między 27 a 37, to uzyskamy istotny statystycznie i dodatni iloraz szans, pokazujący że osoby mające BMI powyżej tej wartości mają istotnie wyższe ryzyko choroby niż osoby poniżej tej wartości.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Profile ilorazu szans pokazują, że czerwona krzywa znajduje się wciąż blisko jedynki, nieco wyżej jest tylko końcówka krzywej, co wskazuje, że może być trudno podzielić BMI na więcej niż 2 kategorie i wybrać dobrą kategorię referencyjną.



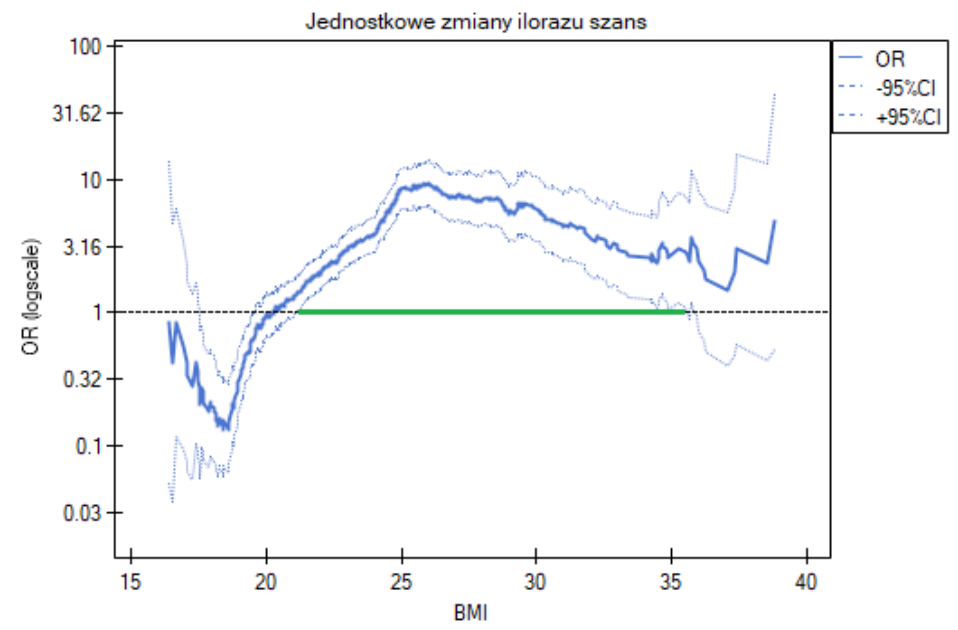
podział na 2 wartości: $OR[95\%CI] = [1.41, 4.90]$**,** $p = 0.0024$**; jednostkowy OR:** $OR[95\%CI] = 1.07[1.02, 1.13]$**,** $p = 0.0052$

Osoby z BMI powyżej 30 mają istotnie wyższe ryzyko choroby A; alternatywnie można pozostać przy jednostkowym ilorazie szans, wskazującym stały wzrost ryzyka przy wzroście BMI o jednostkę.

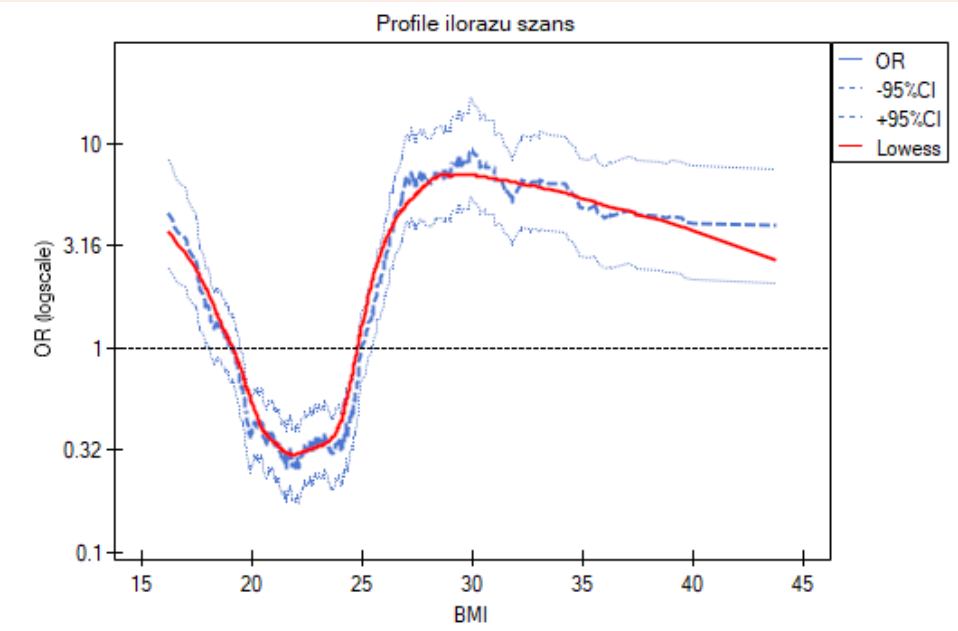
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Choroba B

Jednostkowe zmiany ilorazu szans pokazują, że gdy punkt odcięcia BMI wybierzemy gdzieś między 22 a 35, to uzyskamy istotny statystycznie i dodatni iloraz szans, pokazujący że osoby mające BMI powyżej tej wartości mają istotnie wyższe ryzyko choroby niż osoby poniżej tej wartości.

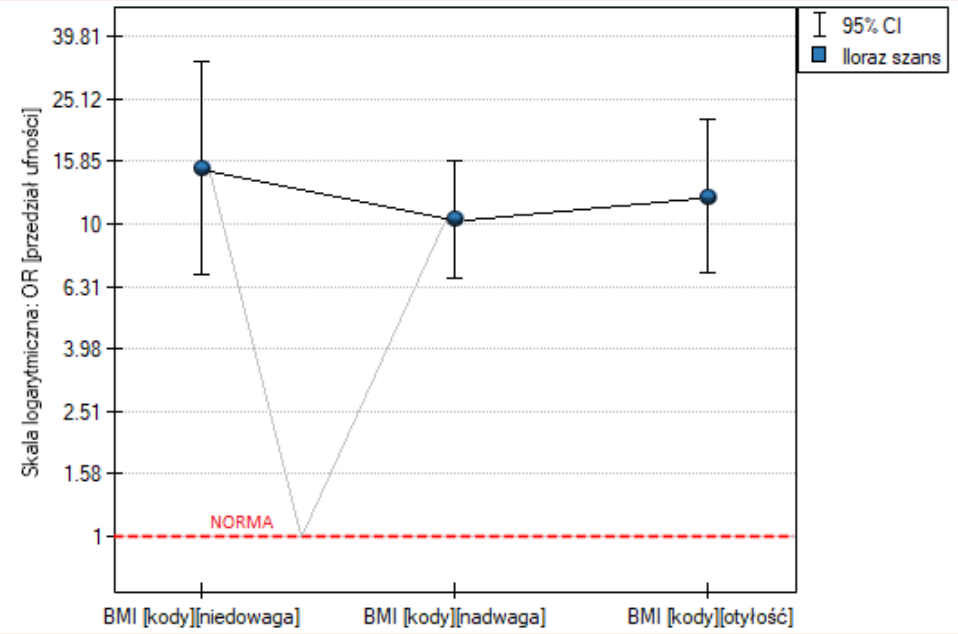


Profile ilorazu szans pokazują, że znacznie lepiej byłoby podzielić BMI na 2 lub 4 kategorie, przy czym kategorią referencyjną powinna być kategoria obejmująca BMI gdzieś pomiędzy 19 a 25, ponieważ to ta kategoria znajduje się najniżej i jest mocno oddalona od wyników dla BMI znajdujących się na lewo i na prawo od tego przedziału. Widzimy wyraźny kształt przypominający literę U, co oznacza, że ryzyko choroby jest wysokie przy niskim i przy wysokim BMI.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Podsumowując, mimo że zależność dla jednostkowego ilorazu szans (zależność liniowa) jest istotna statystycznie, nie warto budować takiego właśnie modelu. Znacznie lepiej podzielić BMI na kategorie, gdzie wartością odniesienia będzie przeciętne BMI. Wykorzystując standardowy podział BMI i ustanawiając kategorią odniesienia BMI w normie, uzyskamy ponad 15-krotnie wyższe ryzyko dla osób z niedowagą (OR[95%CI]=15.14[6.93, 33.10]), ponad dziesięciokrotnie dla osób z nadwagą (OR[95%CI]=10.35[6.74, 15.90]) i ponad dwunastokrotnie dla osób z otyłością (OR[95%CI]=12.22[6.94, 21.49]).



Na wykresie ilorazów szans norma BMI wskazana jest na poziomie 1, jako kategoria referencyjna. Dorysowaliśmy linie łączące uzyskane OR i również normę, tak by pokazać, że uzyskany kształt zależności jest tożsamy z wyznaczonym wcześniej poprzez profil ilorazu szans.

Przykład 25.4. (plik wada.pqs)
Przeprowadzono badanie mające na celu identyfikację czynników ryzyka pewnej rzadko występującej wady wrodzonej u dzieci. W badaniu wzięło udział 395 matek dzieci z tą wadą oraz 375 matek dzieci zdrowych. Zebrane dane to: miejsce zamieszkania, płeć dziecka, masa urodzeniowa dziecka, wiek matki, kolejność ciąży, przebyte poronienia samoistne, infekcje oddechowe, palenie tytoniu, wykształcenie matki.

Budujemy model regresji logistycznej, by sprawdzić, które zmienne mogą wywierać istotny wpływ na występowanie wady. Jako zmienną zależną ustawiamy kolumnę GRUPA, wartością wyróżnioną w tej zmiennej jako 1 jest grupa „badana", czyli matki dzieci z wadą wrodzoną. Kolejnych 9 zmiennych to zmienne niezależne: MiejsceZam (2=miasto/1=wieś), Płeć (1=mężczyzna/0=kobieta), MasaUr (w kilogramach, z dokładnością do 0.5 kg), WiekM (w latach), KolCiąży (dziecko z której ciąży), PoronSamo (1=tak/0=nie), InfOddech (1=tak/0=nie), Palenie (1=tak/0=nie), WyksztM (1=podstawowe lub niżej/2=zawodowe/3=średnie/4=wyższe).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja logistyczna	
Czas analizy	0.12 sek.
Analizowane zmienne	GRUPA;MiejsceZam;Płeć;Ma:
Liczność - niezinterpretowane	0
Liczność - braki danych	89
Poziom istotności	0.05
Założenie liczności próby n ≥ 10(k+1)	Tak
Założenie liczności próby n ≥ 10v/p	Tak
Liczność	678
Liczba zmiennych w modelu	9
Ilość iteracji dla zbieżności	895
Kryterium zbieżności zostało osiągnięte	
Liczność 0 (kontrolna)	346
Liczność 1 (badana)	332
Test ilorazu wiarygodności	
Log Likelihood	-418.0009
-2 Log Likelihood	836.0017
Log Likelihood (w. wolny)	-469.8092
-2 Log Likelihood (w. wolny)	939.6185
Statystyka Chi-kwadrat	103.6167
Stopnie swobody	9
Wartość p	<0.0001
AIC - kryterium Akaikego	854.0017
AICc - skorygowane kryterium Akaikego	854.2712
BIC - bayesowskie kryterium Schwarza	894.6741
Pseudo R2	0.1103
R2(Nagelkerke)	0.189
R2(Coxa-Snella)	0.1417
Hosmer-Lemeshow test	
Statystyka Chi-kwadrat	9.8557
Stopnie swobody	8
Wartość p	0.2753

Model

	wsp. b	błąd b	-95% CI	+95% CI	Wald stat.	wartość p	iloraz szan	-95% CI	+95% CI
w. wolny	1.473902	0.664907	0.170709	2.777096	4.913783	0.026643	4.366241	1.186146	16.072277
MiejsceZam	-0.040877	0.171507	-0.377024	0.29527	0.056807	0.811616	0.959947	0.685899	1.343489
Płeć	0.464687	0.170064	0.131368	0.798005	7.46616	0.006287	1.591515	1.140387	2.221106
MasaUr	-0.307868	0.131076	-0.564772	-0.050963	5.516717	0.018836	0.735013	0.56849	0.950314
Wiek M	-0.033758	0.018671	-0.070353	0.002837	3.268947	0.070603	0.966805	0.932064	1.002841
KolCiąży	0.293138	0.100444	0.096271	0.490005	8.517193	0.003518	1.340628	1.101058	1.632324
PoronSamo	-0.433693	0.303193	-1.02794	0.160554	2.046102	0.152596	0.648111	0.357743	1.174161
InfOddech	1.495785	0.277773	0.95136	2.040209	28.99738	<0.000001	4.462837	2.589229	7.69222
Palenie	1.490982	0.411868	0.683736	2.298227	13.104768	0.000295	4.441453	1.981266	9.956516
WyksztM	-0.183437	0.101185	-0.381755	0.014881	3.286588	0.069848	0.832404	0.682662	1.014992

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$R^2_{Pseudo} = 0.11$**, test ilorazu wiarygodności:** $p < 0.0001$**, test Hosmera-Lemeshowa:** $p = 0.2753$

Jakość dopasowania modelu nie jest wysoka, ale model jest istotny statystycznie, a wynik testu Hosmera-Lemeshowa (brak istotności jest tu pożądany) wskazuje na dobre dopasowanie liczności obserwowanych do przewidywanego prawdopodobieństwa.

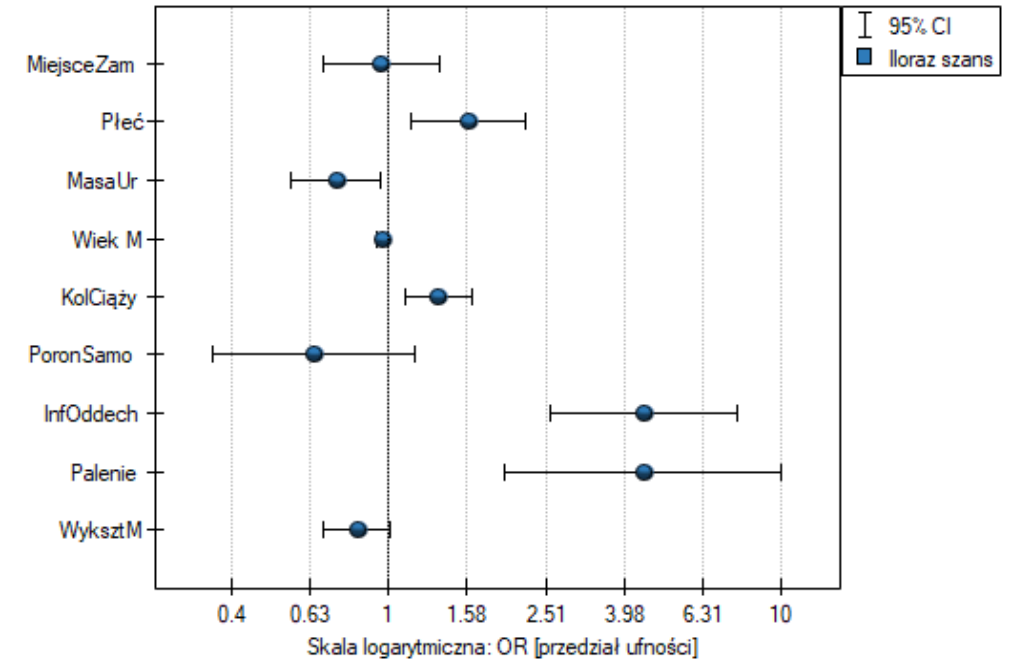
Interpretacja poszczególnych zmiennych w modelu zaczyna się od sprawdzenia ich istotności. W tym przypadku zmienne, które w istotny sposób są związane z występowaniem wady, to: Płeć ($p = 0.0063$), MasaUr ($p = 0.0188$), KolCiąży ($p = 0.0035$), InfOddech ($p < 0.0001$), Palenie ($p = 0.0003$).

Badana wada wrodzona jest wadą rzadką, ale szansa na jej wystąpienie zależy od wymienionych zmiennych w sposób opisany poprzez iloraz szans:

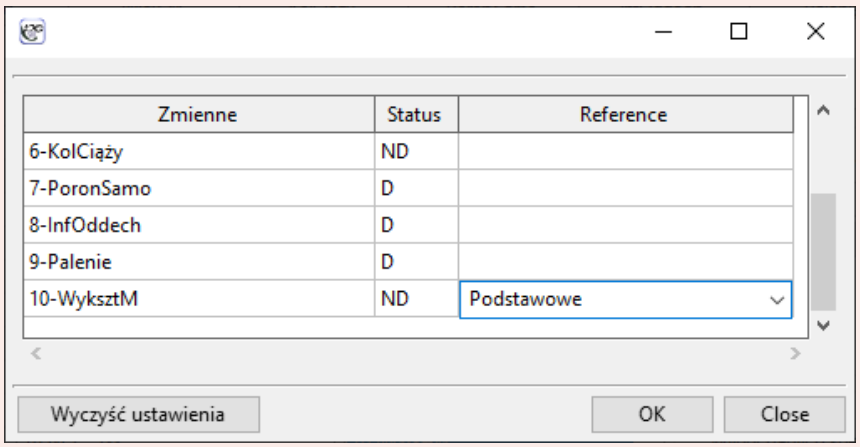
- zmienna Płeć: $OR[95\%CI] = 1.60[1.14; 2.22]$ – szansa wystąpienia wady u chłopca jest 1.6-krotnie większa niż u dziewczynki;
- zmienna MasaUr: $OR[95\%CI] = 0.74[0.57; 0.95]$ – im wyższa masa urodzeniowa, tym szansa wystąpienia wady u dziecka jest mniejsza;
- zmienna KolCiąży: $OR[95\%CI] = 1.34[1.10; 1.63]$ – szansa wystąpienia wady u dziecka wzrasta wraz z każdą kolejną ciążą 1.34-krotnie;
- zmienna InfOddech: $OR[95\%CI] = 4.46[2.59; 7.69]$ – szansa wystąpienia wady u dziecka, gdy matka w czasie ciąży przechodziła infekcję oddechową, jest 4.46-krotnie większa niż gdyby jej nie przechodziła;
- zmienna Palenie: $OR[95\%CI] = 4.44[1.98; 9.96]$ – matka paląca w czasie ciąży zwiększa 4.44-krotnie szansę na wystąpienie wady u dziecka.

W przypadku zmiennych nieistotnych statystycznie przedział ufności dla Ilorazu Szans zawiera jedynkę, co oznacza, że zmienne te nie zwiększają ani nie zmniejszają szansy na wystąpienie badanej wady. Nie można więc interpretować uzyskanego ilorazu w podobny sposób jak dla zmiennych istotnych statystycznie. Wpływ poszczególnych zmiennych niezależnych na występowanie wady możemy również opisać przy pomocy wykresu dotyczącego ilorazu szans:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Kontynuacja przykładu 25.4 (plik wada.pqs): zbudujemy raz jeszcze model regresji logistycznej, ale tym razem zmienną wykształcenie rozbijemy na zmienne fikcyjne (kodowanie zero-jedynkowe). Tracimy tym samym informację o uporządkowaniu kategorii wykształcenia, ale zyskujemy możliwość wnikliwszej analizy poszczególnych kategorii. Rozbicia na zmienne fikcyjne dokonujemy, wybierając w oknie analizy Zm. fikcyjne:



Wykształcenie podstawowe wybieramy jako kategorię odniesienia.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja logistyczna	
Czas analizy	0.32 sek.
Analizowane zmienne	MiejsceZam;Płeć;MasaUr;Wi
Liczność - braki danych	89
Poziom istotności	0.05
Kryterium zbieżności zostało osiągnięte	
Liczność	678
Liczba szacowanych parametrów	12
Liczność 0 (kontrolna)	346
Liczność 1 (badana)	332
Test ilorazu wiarygodności	
Log Likelihood	-416.06069
-2 Log Likelihood	832.12138
Log Likelihood (w. wolny)	-469.809235
-2 Log Likelihood (w. wolny)	939.618471
Statystyka Chi-kwadrat	107.49709
Ilość stopni swobody	11
Wartość p	<0.000001
Pseudo R2	0.114405
R2(Nagelkerke)	0.195521
R2(Coxa-Snella)	0.14662
Hosmer-Lemeshow test	
Statystyka Chi-kwadrat	6.720908
Ilość stopni swobody	8
Wartość p	0.567022

Model

	wsp. b	błąd b	-95% CI	+95% CI	stat. Walda	wartość p	iloraz szans	-95% CI	+95% CI
w. wolny	1.6651	0.6933	0.3062	3.024	5.7675	0.0163	5.2863	1.3582	20.5744
MiejsceZam	-0.0466	0.173	-0.3856	0.2925	0.0725	0.7878	0.9545	0.68	1.3398
Płeć	0.4381	0.171	0.1029	0.7733	6.5635	0.0104	1.5498	1.1084	2.1669
MasaUr	-0.2959	0.1316	-0.5538	-0.0381	5.06	0.0245	0.7438	0.5748	0.9626
Wiek M	-0.0341	0.0188	-0.071	0.0028	3.277	0.0703	0.9665	0.9315	1.0028
KolCiąży	0.2997	0.101	0.1017	0.4976	8.7992	0.003	1.3494	1.107	1.6448
PoronSamo	-0.4918	0.3069	-1.0933	0.1098	2.5675	0.1091	0.6116	0.3351	1.116
InfOddech	1.4878	0.2777	0.9436	2.032	28.7124	<0.0001	4.4274	2.5692	7.6294
Palenie	1.4579	0.4145	0.6455	2.2704	12.3713	0.0004	4.2971	1.907	9.6828
WykształM[Zawoc	-0.6823	0.3448	-1.3581	-0.0065	3.9152	0.0479	0.5055	0.2571	0.9936
WykształM[Średni	-0.8715	0.3327	-1.5236	-0.2195	6.8633	0.0088	0.4183	0.2179	0.8029
WykształM[Wyższ	-0.7908	0.3599	-1.4963	-0.0853	4.8272	0.028	0.4535	0.224	0.9182

W rezultacie zmienne opisujące wykształcenie stają się istotne statystycznie. Dopasowanie modelu nie ulega znacznej zmianie, ale zmienia się sposób interpretacji ilorazu szans dla wykształcenia:

Zmienna	$OR[95\%CI]$
Wykształcenie podstawowe	kategoria referencyjna
Wykształcenie zawodowe	$0.51[0.26; 0.99]$
Wykształcenie średnie	$0.42[0.22; 0.80]$
Wykształcenie wyższe	$0.45[0.22; 0.92]$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Szansa na wystąpienie badanej wady w każdej kategorii wykształcenia odnoszona jest zawsze do szansy wystąpienia wady przy wykształceniu podstawowym. Widzimy, że dla bardziej wykształconych matek iloraz szans jest niższy. Dla matki z wykształceniem zawodowym szansa wystąpienia wady u dziecka stanowi 0.51 część szansy dotyczącej matki z wykształceniem podstawowym; ze średnim – 0.42 część; z wyższym – 0.45 część.

PRZYKŁAD 25.5. (plik zadanie.pqs)
Przeprowadzono eksperyment mający na celu zbadanie umiejętności koncentracji grupy dorosłych podczas sytuacji niekomfortowych. W eksperymencie wzięło udział 190 osób (130 osób to zbiór uczący, 40 osób to zbiór testowy). Każda badana osoba dostała pewne zadanie, którego rozwiązanie wymagało skupienia uwagi. Podczas eksperymentu niektóre osoby zostały poddane działaniu czynnika zakłócającego, jakim była podwyższona temperatura powietrza do 32 stopni Celsjusza. Osoby biorące udział w eksperymencie zapytano dodatkowo o ich miejsce zamieszkania, płeć, wiek i wykształcenie. Czas na rozwiązanie zadania ograniczono do 45 minut. Dla osób, które skończyły przed czasem, odnotowano rzeczywisty czas poświęcony na rozwiązanie. Całość naszych obliczeń wykonamy tylko dla osób należących do zbioru uczącego.

Zmienna ROZWIĄZANIE (tak/nie) zawiera wynik eksperymentu, czyli informację o tym, czy zadanie zostało rozwiązane poprawnie czy też nie. Pozostałe zmienne, które mogły wpływać na wynik eksperymentu, to: MIEJSCEZAM (1=miasto/0=wieś), PŁEĆ (1=kobieta/0=mężczyzna), WIEK (w latach), WYKSZTAŁCENIE (1=podstawowe, 2=zawodowe, 3=średnie, 4=wyższe), CZAS rozwiązywania (w minutach), ZAKŁÓCENIA (1=tak/0=nie). Na bazie wszystkich zmiennych zbudowano model regresji logistycznej, gdzie jako stan wyróżniony zmiennej ROZWIĄZANIE wybrano „tak”.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

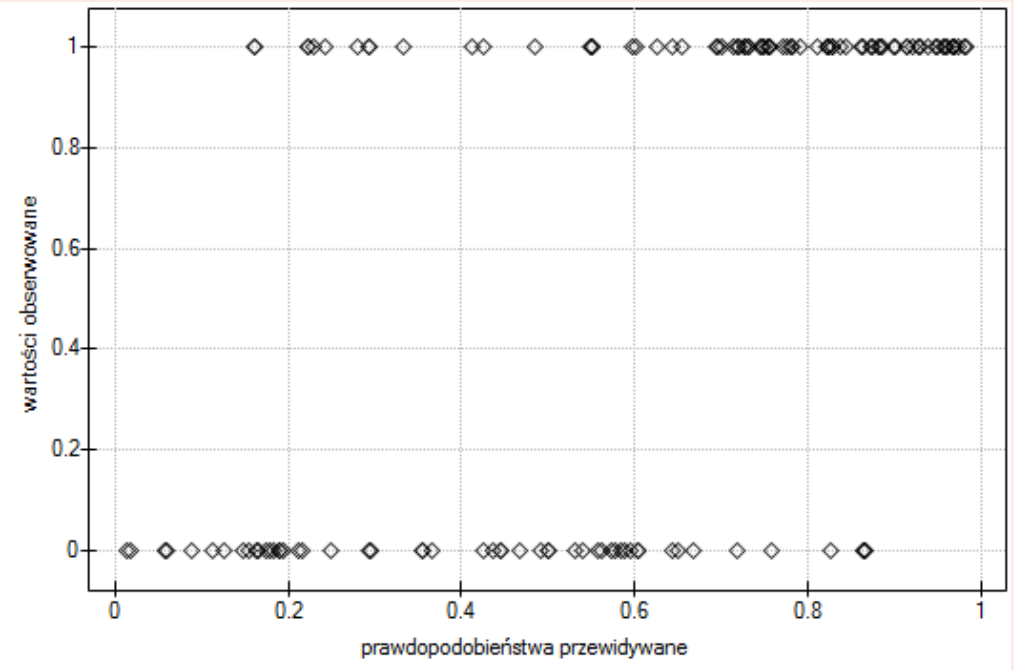
Regresja logistyczna	
Analizowane zmienne	ROZWIĄZANIE
	MIEJSCE_ZAM
	PŁEĆ
	WIEK
	WYKSZTAŁCENIE
	CZAS
	ZAKŁÓCENIA
Filtr danych	zbiór=uczący
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Założenie liczności próby n ≥ 10(k+1)	Tak
Założenie liczności próby n ≥ 10v/p	Nie
Liczność	130
Liczba zmiennych w modelu	6
Ilość iteracji dla zbieżności	158
Kryterium zbieżności zostało osiągnięte	
Liczność 0 (nie)	53
Liczność 1 (tak)	77
Test ilorazu wiarygodności	
Log Likelihood	-64.3541
-2 Log Likelihood	128.7082
Log Likelihood (w. wolny)	-87.881
-2 Log Likelihood (w. wolny)	175.762
Statystyka Chi-kwadrat	47.0537
Stopnie swobody	6
Wartość p	<0.0001
AIC - kryterium Akaikego	140.7082
AICc - skorygowane kryterium Akaikego	141.3912
BIC - bayesowskie kryterium Schwarza	157.9134
Pseudo R2	0.2677
R2(Nagelkerke)	0.4097
R2(Coxa-Snella)	0.3037
Hosmer-Lemeshow test	
Statystyka Chi-kwadrat	11.5486
Stopnie swobody	8
Wartość p	0.1725

$R^2_{Pseudo} = 0.27$, $R^2_{Nagelkerke} = 0.41$, $R^2_{Cox-Snell} = 0.30$; **test Hosmera-Lemeshowa:** $p = 0.1725$; **test ilorazu wiarygodności:** $p < 0.0001$

Model jest istotny statystycznie i wystarczająco dobrze dopasowany.

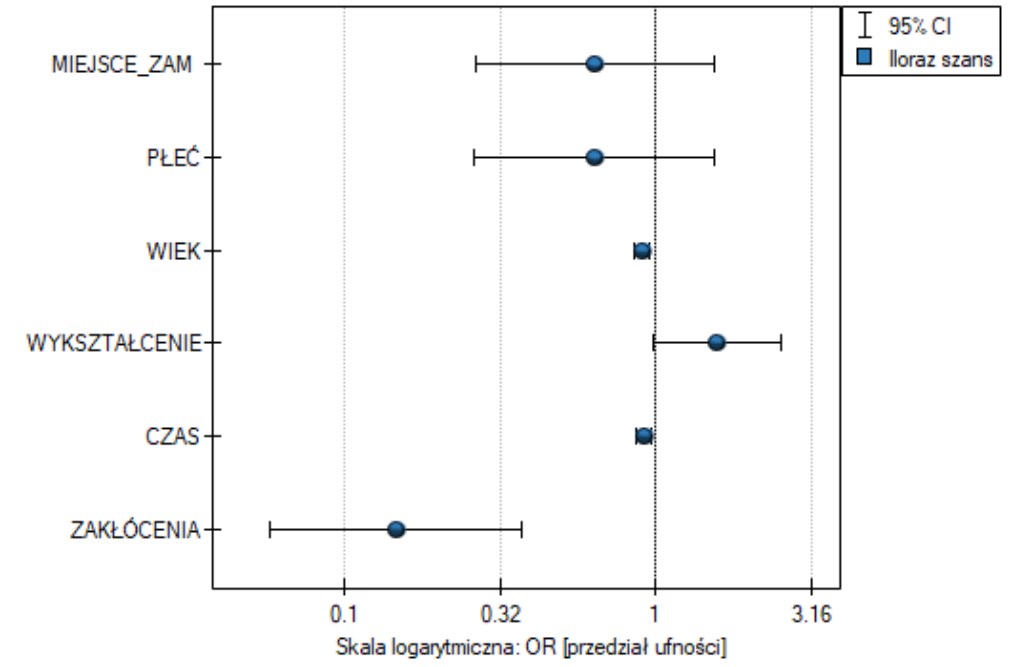
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Model	wsp. b	błąd b	-95% CI	+95% CI	stat. Walda	wartość p	iloraz szans	-95% CI	+95% CI
w. wolny	7.2306	1.8701	3.5652	10.896	14.9487	0.0001	1381.0527	35.3467	53959.8804
MIEJSCE_ZAM	-0.4532	0.4505	-1.3363	0.4298	1.0121	0.3144	0.6356	0.2628	1.5369
PŁEĆ	-0.4548	0.4513	-1.3393	0.4298	1.0155	0.3136	0.6346	0.262	1.5369
WIEK	-0.1009	0.0316	-0.1628	-0.039	10.2009	0.0014	0.904	0.8498	0.9618
WYKSZTAŁCENI	0.4559	0.2418	-0.018	0.9299	3.5552	0.0594	1.5776	0.9822	2.5341
CZAS	-0.0894	0.0276	-0.1435	-0.0353	10.4839	0.0012	0.9145	0.8663	0.9653
ZAKŁÓCENIA	-1.924	0.4751	-2.8551	-0.9929	16.4029	0.0001	0.146	0.0576	0.3705

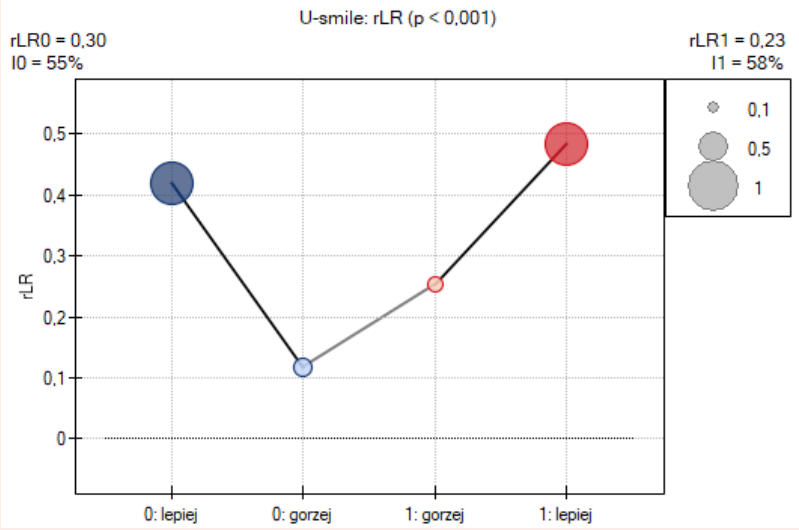


W modelu zmienne, które w sposób istotny wpływają na wynik, to: WIEK ($p = 0.0014$), CZAS ($p = 0.0012$), ZAKŁÓCENIA ($p = 0.0001$). Przy czym im osoba rozwiązująca jest młodsza, czas rozwiązywania krótszy i brak jest czynnika zakłócającego, tym większe prawdopodobieństwo poprawnego rozwiązania: WIEK $OR[95\%CI] = 0.90[0.85; 0.96]$, CZAS $OR[95\%CI] = 0.91[0.87; 0.97]$, ZAKŁÓCENIA $OR[95\%CI] = 0.15[0.06; 0.37]$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



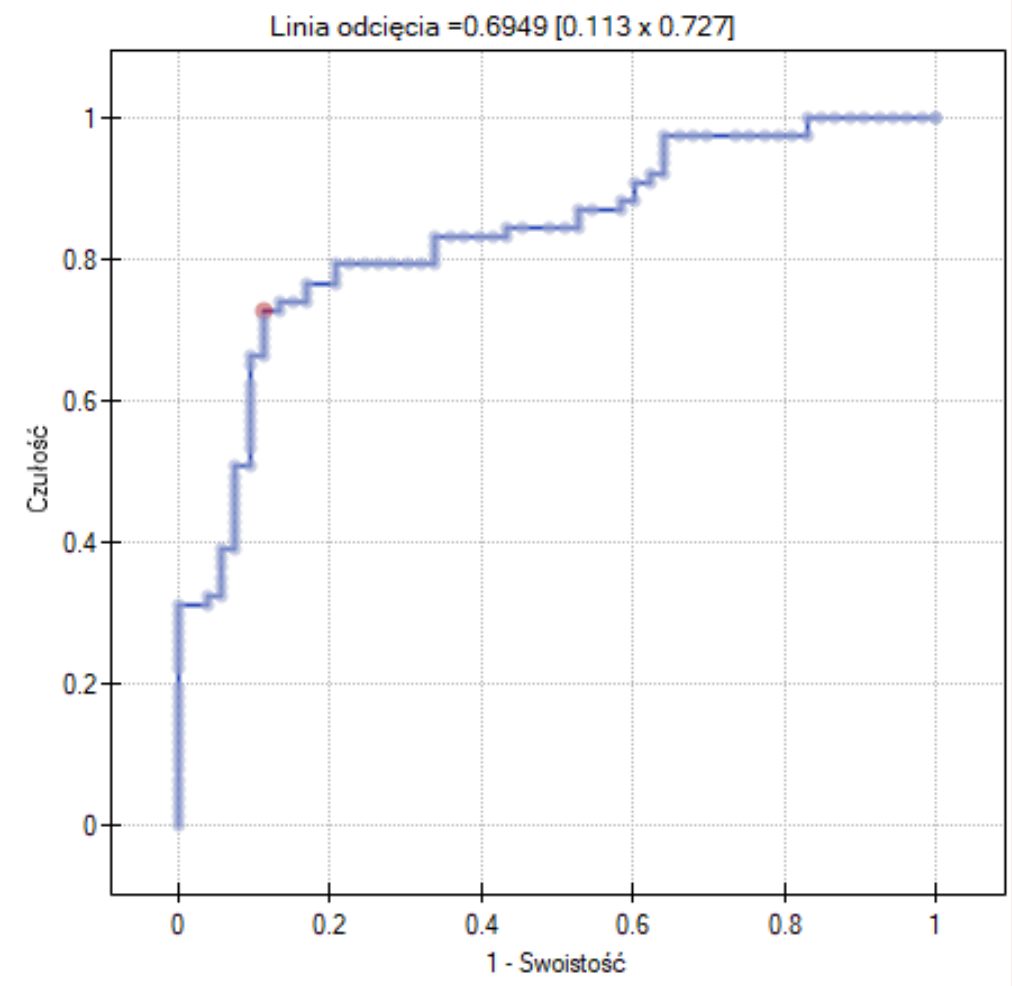
Jeśli model miałby zostać użyty do prognozowania, należy ocenić, czy będzie w tym lepszy niż klasyfikacja losowa. Narysujemy wykres U-smile na bazie współczynnika określającego jakość dopasowania rLR. Wykres się „uśmiecha". Wiarygodność prognozy wskazującej, że osoba rozwiąże zadanie (prawa część wykresu), jest 0.23 razy większa niż w modelu losowym i netto (tzn. po odjęciu osób, u których wiarygodność spadła) dotyczy 58% osób. Wykres jest symetryczny i podobne wyniki dotyczą lewej strony wykresu, czyli nierozwiązania zadania: $rLR_0 = 0.3$, $I_0 = 55\%$.



Jakość klasyfikacji ocenimy na bazie krzywej ROC.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krzywe ROC (DeLong's method)	
AUC	0.8346
SE(AUC)	0.0355
-95% CI	0.765
+95% CI	0.9042
Statystyka Z	6.4694
Wartość p	<0.0001
Linia odcięcia	0.6949



$AUC = 0.83$, $p < 0.0001$ **(istotnie większe niż 0.5)**

Na podstawie zbudowanego modelu można klasyfikować.

Proponowany punkt odcięcia dla krzywej ROC wynosi 0.6949 i jest nieco wyższy niż standardowo używany w regresji poziom 0.5. Klasyfikacja wyznaczona na bazie tego punktu odcięcia daje 79.23% przypadków zaklasyfikowanych poprawnie, z czego poprawnie zaklasyfikowanych wartości „tak” jest 72.73% (czułość), wartości „nie” jest 88.68% (swoistość). Klasyfikacja uzyskana na podstawie standardowej wartości daje nieco mniej, bo 73.85% przypadków zaklasyfikowanych poprawnie, ale uzyskamy dzięki niej

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

więcej poprawnie zaklasyfikowanych wartości „tak” (83.12%), choć mniej poprawnie zaklasyfikowanych wartości „nie” (60.38%).

Klasyfikacja		w. obserwowana	
w. przewidywana		1	0
	1	56	6
	0	21	47
Linia odcięcia	0.6949		
% poprawnych	79.23%		
Czułość (% pop	72.73%		
-95% CI	61.38%		
+95% CI	82.26%		
Swoistość (% p	88.68%		
-95% CI	76.97%		
+95% CI	95.73%		

Klasyfikacja		w. obserwowana	
w. przewidywana		1	0
	1	64	21
	0	13	32
Linia odcięcia	0.5		
% poprawnych	73.85%		
Czułość (% pop	83.12%		
-95% CI	72.86%		
+95% CI	90.69%		
Swoistość (% p	60.38%		
-95% CI	46%		
+95% CI	73.55%		

Na tym etapie możemy zakończyć analizę klasyfikacji, lub jeśli wynik nie jest wystarczający, bardziej wnikliwą analizę krzywej ROC możemy przeprowadzić w module Krzywa ROC.

Jednak pamiętajmy, że wszelkich ocen dokonywaliśmy jedynie na zbiorze uczącym. Jeśli model ma rzeczywiście służyć do analizy danych, na których się nie uczył, powinniśmy go ocenić na niezależnym zbiorze testowym. Zostało to zrobione w module Walidacja i predykcja dla regresji logistycznej, gdzie potwierdzono wyniki zarówno na wykresie U-smile, jak i na krzywej ROC.

Ponieważ uznaliśmy, że klasyfikacja na podstawie modelu jest zadowalająca, możemy wyliczyć prognozowaną wartość zmiennej zależnej dla dowolnie zadanych warunków. Sprawdźmy, jakie szanse na rozwiązanie zadania ma osoba, dla której: MIEJSCEZAM (1=miasto), PŁEĆ (1=kobieta), WIEK (50 lat), WYKSZTAŁCENIE (1=podstawowe), CZAS rozwiązywania (20 minut), ZAKŁÓCENIA (1=tak). W tym celu na podstawie wartości współczynnika b wyliczane jest prawdopodobieństwo przewidywane:

W rezultacie tych obliczeń program zwróci wynik:

Predykcja	
1-MIEJSCE_ZAM	1
2-PŁEĆ	1
3-WIEK	50
4-WYKSZTAŁCEI	1
5-CZAS	20
6-ZAKŁÓCENIA	1
Punkt odcięcia	0.5
pred. prawdop.	0.121512
Pred. Y	0

Uzyskane prawdopodobieństwo rozwiązania zadania wynosi 0.1215, więc na podstawie punktu odcięcia 0.5 przewidziany wynik to 0 – czyli zadanie nierozwiązane poprawnie.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Iloraz szans – wzór ogólny

Wzory:

Dla zmiennej niezależnej X_1, przy Z wyrażonym zależnością liniową, wyliczamy szansę dla kategorii pierwszej:

$$Szansa(1) = \frac{P(1)}{1-P(1)} = e^Z(1) = e^{\beta_0+\beta_1 X_1(1)+\beta_2 X_2+\ldots+\beta_k X_k},$$

oraz szansę dla kategorii drugiej:

$$Szansa(2) = \frac{P(2)}{1-P(2)} = e^Z(2) = e^{\beta_0+\beta_1 X_1(2)+\beta_2 X_2+\ldots+\beta_k X_k}.$$

Iloraz Szans dla zmiennej X_1 wyraża się wówczas wzorem:

$$\begin{aligned} OR_1(2)/(1) &= \frac{Szansa(2)}{Szansa(1)} = \frac{e^{\beta_0+\beta_1 X_1(2)+\beta_2 X_2+\ldots+\beta_k X_k}}{e^{\beta_0+\beta_1 X_1(1)+\beta_2 X_2+\ldots+\beta_k X_k}} \\ &= e^{\beta_0+\beta_1 X_1(2)+\beta_2 X_2+\ldots+\beta_k X_k-[\beta_0+\beta_1 X_1(1)+\beta_2 X_2+\ldots+\beta_k X_k]} \\ &= e^{\beta_1 X_1(2)-\beta_1 X_1(1)} = e^{\beta_1[X_1(2)-X_1(1)]} = \\ &= \left(e^{\beta_1}\right)^{[X_1(2)-X_1(1)]}. \end{aligned}$$

Przykład: jeśli zmienną niezależną jest wiek wyrażony w latach, to różnica pomiędzy sąsiadującymi kategoriami wieku, np. 25 lat i 26 lat, wynosi 1 rok $(X_1(2) - X_1(1) = 26 - 25 = 1)$. Wówczas otrzymamy jednostkowy Iloraz Szans $OR = \left(e^{\beta_1}\right)^1$, który mówi, o ile zmieni się szansa na wystąpienie wyróżnionej wartości, gdy wiek zmieni się o 1 rok. Iloraz szans wyliczony dla niesąsiadujących kategorii, np. 25 i 30 lat, będzie pięcioletnim Ilorazem Szans $OR = \left(e^{\beta_1}\right)^5$, ponieważ różnica $X_1(2) - X_1(1) = 30 - 25 = 5$.

Test Walda – istotność poszczególnych zmiennych

Hipotezy:

$$\begin{array}{ll} \mathcal{H}_0: & \beta_i = 0, \\ \mathcal{H}_1: & \beta_i \neq 0. \end{array} \quad \text{lub równoważnie:} \quad \begin{array}{ll} \mathcal{H}_0: & OR_i = 1, \\ \mathcal{H}_1: & OR_i \neq 1. \end{array}$$

Wzory:

Statystykę testową testu Walda wyliczamy według wzoru:

$$\chi^2 = \left(\frac{b_i}{SE_{b_i}}\right)^2.$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z 1 stopniem swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Kryterium informacyjne Akaikego

$$AIC = -2 \ln L_{FM} + 2k,$$

jest kryterium asymptotycznym – odpowiednim dla dużych prób.

Poprawione kryterium informacyjne Akaikego

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}.$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Bayesowskie kryterium informacyjne Schwarza

$$BIC = -2 \ln L_{FM} + k \ln(n).$$

Pseudo R^2

$$R^2_{Pseudo} = 1 - \frac{\ln L_{FM}}{\ln L_0}.$$

$$R^2_{Nagelkerke} = \frac{1 - e^{-(2/n)(\ln L_{FM} - \ln L_0)}}{1 - e^{(2/n) \ln L_0}} \quad \text{lub} \quad R^2_{Cox-Snell} = 1 - e^{\frac{(-2 \ln L_0) - (-2 \ln L_{FM})}{n}}.$$

Test ilorazu wiarygodności
Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \text{ wszystkie } \beta_i = 0, \\ \mathcal{H}_1 &: \text{ istnieje } \beta_i \neq 0. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$\chi^2 = -2 \ln(L_0/L_{FM}) = -2 \ln(L_0) - (-2 \ln(L_{FM})).$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z k stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test Hosmera-Lemeshowa
Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \ O_g = E_g \text{ dla wszystkich kategorii}, \\ \mathcal{H}_1 &: \ O_g \neq E_g \text{ dla przynajmniej jednej kategorii}. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$H = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - \frac{E_g}{N_g})},$$

gdzie N_g – liczba obserwacji w grupie g.
Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $G - 2$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Test istotności AUC
Interpretacja:
Gdy krzywa ROC pokrywa się z przekątną $y = x$, to decyzja o przyporządkowaniu przypadku do wybranej klasy podejmowana na podstawie modelu jest tak samo dobra jak losowy podział badanych przypadków do tych grup. Jakość klasyfikacyjna modelu jest dobra, gdy pole pod krzywą ROC jest znacznie większe niż 0.5.

Hipotezy:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

$$\mathcal{H}_0: \quad AUC = 0.5,$$
$$\mathcal{H}_1: \quad AUC \neq 0.5.$$

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{AUC - 0.5}{SE_{0.5}},$$

gdzie $SE_{0.5}$ – błąd pola. Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Współczynniki U-smile
Oznaczenia: $y_i \in \{0, 1\}$ – wynik obserwacji, p_i^{ref} i p_i – prawdopodobieństwa zdarzenia z modelu odniesienia i modelu nowego, n_0, n_1 – liczności nie-zdarzeń i zdarzeń, $n = n_0 + n_1$. Podklasy: 0^+ (nie-zdarzenia, $p_i < p_i^{ref}$), 0^- (nie-zdarzenia, $p_i > p_i^{ref}$), 1^- (zdarzenia, $p_i < p_i^{ref}$), 1^+ (zdarzenia, $p_i > p_i^{ref}$); obserwacje z $p_i = p_i^{ref}$ nie należą do żadnej podklasy.

Współczynniki I: $I_0^+ = n_0^+/n_0$, $I_0^- = n_0^-/n_0$, $I_1^- = n_1^-/n_1$, $I_1^+ = n_1^+/n_1$; netto $I_0 = I_0^+ - I_0^-$, $I_1 = I_1^+ - I_1^-$; ogółem $I = I_0 + I_1$.

Współczynniki BA i RB: dla reszt kwadratowych $r_i = (y_i - p_i)^2$ i $r_i^{ref} = (y_i - p_i^{ref})^2$ oraz różnic sum kwadratów w podklasie S: $\Delta SS_S = \sum_{i \in S} r_i^{ref} - \sum_{i \in S} r_i$

$$BA_0^+ = \frac{\Delta SS_{0^+}}{n_0}, \quad BA_0^- = \frac{-\Delta SS_{0^-}}{n_0}, \quad BA_1^- = \frac{-\Delta SS_{1^-}}{n_1}, \quad BA_1^+ = \frac{\Delta SS_{1^+}}{n_1}.$$

Współczynniki RB liczone są tak samo, ale z mianownikami $SS_0^{ref} = \sum_{y_i=0} r_i^{ref}$ i $SS_1^{ref} = \sum_{y_i=1} r_i^{ref}$. Netto: $BA_0 = BA_0^+ - BA_0^-$, $BA_1 = BA_1^+ - BA_1^-$; ogółem $BA = \frac{n_0}{n} BA_0 + \frac{n_1}{n} BA_1$ (analogicznie RB).

Współczynnik rLR: dla składników log-wiarygodności $\ell_i = y_i \ln p_i + (1 - y_i) \ln(1 - p_i)$ i ℓ_i^{ref} (analogicznie dla p_i^{ref}) oraz $\Delta_S = \sum_{i \in S} \ell_i^{ref} - \sum_{i \in S} \ell_i$:

$$LR_0^+ = -2\Delta_{0^+}, \quad LR_0^- = 2\Delta_{0^-}, \quad LR_1^- = 2\Delta_{1^-}, \quad LR_1^+ = -2\Delta_{1^+}, \quad LR_0 = LR_0^+ - LR_0^-, \quad LR_1 = LR_1^+ - LR_1^-$$

Wartości te są normalizowane przez maksymalny osiągalny iloraz wiarygodności klasy: $\max LR_0 = -2\sum_{y_i=0} \ell_i^{ref}$ (gdy $LR_0^+ \geq LR_0^-$; w przeciwnym razie z ℓ_i modelu nowego), analogicznie $\max LR_1$. Stąd $rLR_0^{\pm} = LR_0^{\pm} / \max LR_0$, $rLR_1^{\pm} = LR_1^{\pm} / \max LR_1$, $rLR_0 = LR_0 / \max LR_0$, $rLR_1 = LR_1 / \max LR_1$ oraz $rLR = (LR_0 + LR_1) / \max LR$, gdzie $\max LR = -2\sum_i \ell_i^{ref}$ (suma po obserwacjach należących do podklas).

Testy: ogółem – test ilorazu wiarygodności o statystyce $2\left|\sum_i \ell_i - \sum_i \ell_i^{ref}\right|$, mającej asymptotycznie rozkład χ^2 z df równym różnicy liczby parametrów porównywanych modeli (co najmniej 1). Dla rLR istotność zmiany w klasie sprawdza się z rozkładu gamma: $|LR_1| \sim \Gamma\left(\frac{df}{2}, \frac{1}{2c_1}\right)$ i $|LR_0| \sim \Gamma\left(\frac{df}{2}, \frac{1}{2c_0}\right)$, gdzie $c_1 = n_1/n$, $c_0 = 1 - c_1$ (parametry kształtu i tempa). Wyznaczone wartości p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, uznajemy zmianę predykcji za istotną.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Predykcja na podstawie modelu i walidacja zbioru testowego

Walidacja i predykcja dla regresji logistycznej

KIEDY STOSOWAĆ:

Walidacja modelu to sprawdzenie jego jakości. W pierwszej kolejności wykonywana jest na danych, na których model był zbudowany (tzw. **zbiór uczący**), czyli zwracana jest w raporcie opisującym uzyskany model. By można było z większą pewnością osądzić, na ile model nadaje się do prognozy nowych danych, ważnym elementem walidacji jest zastosowanie modelu do danych, które nie były wykorzystywane w estymacji modelu. Jeśli podsumowanie w oparciu o dane uczące będzie satysfakcjonujące, tzn. wyznaczane błędy, współczynniki R^2 i kryteria informacyjne będą na zadowalającym nas poziomie, a podsumowanie w oparciu o nowe dane (tzw. **zbiór testowy**) będzie równie korzystne, wówczas z dużym prawdopodobieństwem można uznać, że taki model nadaje się do predykcji. Dane testujące powinny pochodzić z tej samej populacji, z której były wybrane dane uczące. Często jest tak, że przed przystąpieniem do budowy modelu zbieramy dane, a następnie w sposób losowy dzielimy je na zbiór uczący, czyli dane, które posłużą do budowy modelu, i zbiór testowy, czyli dane, które posłużą do dodatkowej walidacji modelu.

Predykcja – najczęściej ostatnim etapem analizy regresji jest wykorzystanie zbudowanego i uprzednio zweryfikowanego modelu do predykcji. Na podstawie nowych danych wyznaczana jest wartość prawdopodobieństwa przewidywanego przez model, a następnie predykcja wystąpienia zdarzenia (1) lub jego braku (0).

DEFINICJE MIAR

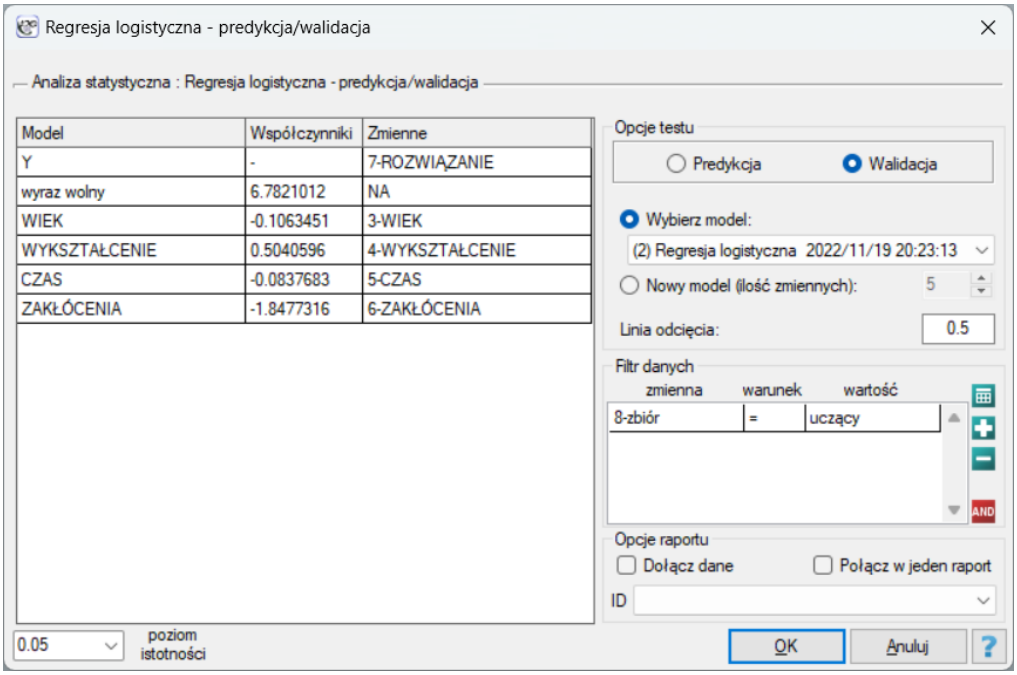
Predykcja wystąpienia zdarzenia (1) lub jego braku (0) opiera się na porównaniu wartości prawdopodobieństwa przewidywanego przez model z tzw. **punktem odcięcia**. Domyślnie punkt odcięcia ustawiony jest na wartość 0.5, jednak użytkownik może go zmienić na dowolną wartość z przedziału $(0-1)$, np. na wartość sugerowaną przez krzywą ROC.

U-smile na zbiorze walidacyjnym – gdy w trybie walidacji wskazano zmienną z obserwowanym wynikiem, można dodatkowo wyznaczyć współczynniki i wykres U-smile wybranego modelu względem modelu zerowego (częstości zdarzeń w zbiorze walidacyjnym). Odpowiada to ocenie modelu na danych niezależnych od danych uczących: uśmiech potwierdza, że model poprawia predykcję obu klas także poza próbą uczącą, a asymetria wskazuje klasę, dla której poprawa się nie utrzymuje.

JAK USTAWIĆ ANALIZĘ

Statystyki zaawansowane → Modele wielowymiarowe → Regresja logistyczna – predykcja/walidacja.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



By dokonać walidacji lub predykcji, należy wskazać model, na podstawie którego chcemy jej dokonać. Możemy dokonać jej na bazie:

- zbudowanego w PQStat modelu regresji logistycznej – wystarczy wybrać model spośród modeli przypisanych do danego arkusza, a liczba zmiennych i współczynniki modelu zostaną ustawione automatycznie; zbiór testowy powinien się znaleźć w tym samym arkuszu co zbiór uczący;
- modelu niezbudowanego w programie PQStat, ale uzyskanego z innego źródła (np. opisanego w przeczytanej przez nas pracy naukowej) – w oknie analizy należy podać liczbę zmiennych oraz wpisać współczynniki dotyczące każdej z nich.

W oknie analizy należy wskazać te nowe zmienne, które powinny zostać wykorzystane do walidacji lub predykcji. Predykcja dla jednego obiektu może być wykonywana wraz z budową modelu, czyli w oknie analizy Statystyki zaawansowane→Modele wielowymiarowe→Regresja logistyczna, natomiast predykcja dla większej grupy nowych danych jest wykonywana poprzez menu Statystyki zaawansowane→Modele wielowymiarowe→Regresja logistyczna – predykcja/walidacja.

WYKONANIE I INTERPRETACJA

Przykład 25.6. c.d. przykładu 25.5 (plik zadanie.pqs)
W eksperymencie badającym umiejętność koncentracji, dla grupy 130 osób zbioru uczącego, zbudowano model regresji logistycznej w oparciu o następujące zmienne:

zmienna zależna: ROZWIĄZANIE (tak/nie) – informacja o tym, czy zadanie zostało rozwiązane poprawnie czy też nie;

zmienne niezależne:

MIEJSCEZAM (1=miasto/0=wieś),

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PŁEĆ (1=kobieta/0=mężczyzna),
WIEK (w latach),
WYKSZTAŁCENIE (1=podstawowe, 2=zawodowe, 3=średnie, 4=wyższe),
CZAS rozwiązywania (w minutach),
ZAKŁÓCENIA (1=tak/0=nie).

Jednak tylko cztery zmienne: WIEK, WYKSZTAŁCENIE, CZAS rozwiązywania i ZAKŁÓCENIA, wnoszą istotne informacje do modelu. Zbudujemy model dla danych zbioru uczącego w oparciu o te cztery zmienne, a następnie, by się upewnić, że będzie działał poprawnie, zwalidujemy go na testowym zbiorze danych. Jeśli model przejdzie tę próbę, to będziemy go stosować do predykcji dla nowych osób. By korzystać z odpowiednich zbiorów, ustawiamy każdorazowo filtr danych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja logistyczna - predykcja/walidacja	
Analizowane zmienne	WIEK
	WYKSZTAŁCENIE
	CZAS
	ZAKŁÓCENIA
	ROZWIĄZANIE
Filtr danych	zbiór=uczący
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	130
Walidacja	
Test ilorazu wiarygodności	
Log Likelihood	-65.5411368
-2 Log Likelihood	131.0822737
Log Likelihood (w. wolny)	-87.8809897
-2 Log Likelihood (w. wolny)	175.7619793
Statystyka Chi-kwadrat	44.6797056
Stopnie swobody	4
Wartość p	4.6347E-9
AIC - kryterium Akaikego	139.0822737
AICc - skorygowane kryterium Akaikego	139.4022737
BIC - bayesowskie kryterium Schwarza	150.5524115
Pseudo R2 (McFadden)	0.2542057
R2(Nagelkerke)	0.3923627
R2(Coxa-Snella)	0.2908513
Hosmer-Lemeshow test	
Statystyka Chi-kwadrat	15.0231507
Stopnie swobody	8
Wartość p	0.0586969
Krzywe ROC	
AUC	0.8230826
SE(AUC)	0.036776
-95% CI	0.751003
+95% CI	0.8951622
Statystyka Z	6.2467167
Wartość p dwustronne	4.1917E-10
Linia odcięcia	0.5649961

zbiór uczący: $Pseudo\ R^2 = 0.26$, $R^2(Nagelkerke) = 0.39$; $AUC[95\%CI] = 0.82[0.75, 0.90]$, **czułość** $= 82\%$, **swoistość** $= 60\%$

Wartości opisujące jakość dopasowania modelu nie są bardzo wysokie, ale jakość jego predykcji jest już zadowalająca.

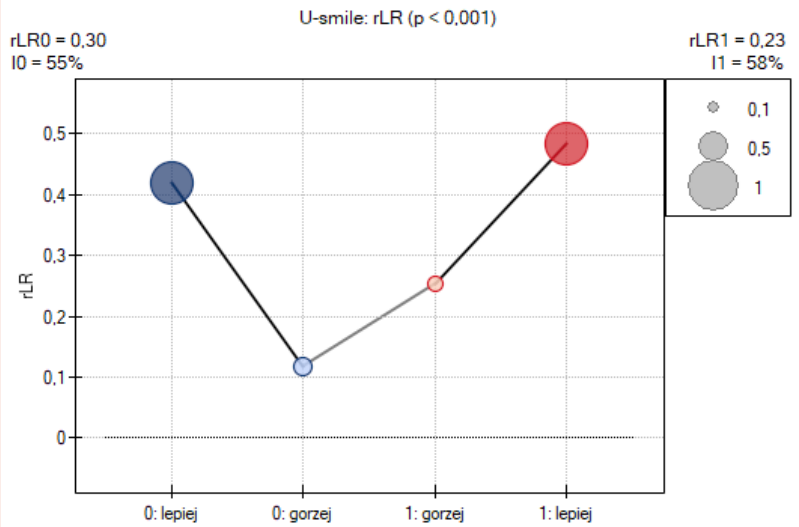
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

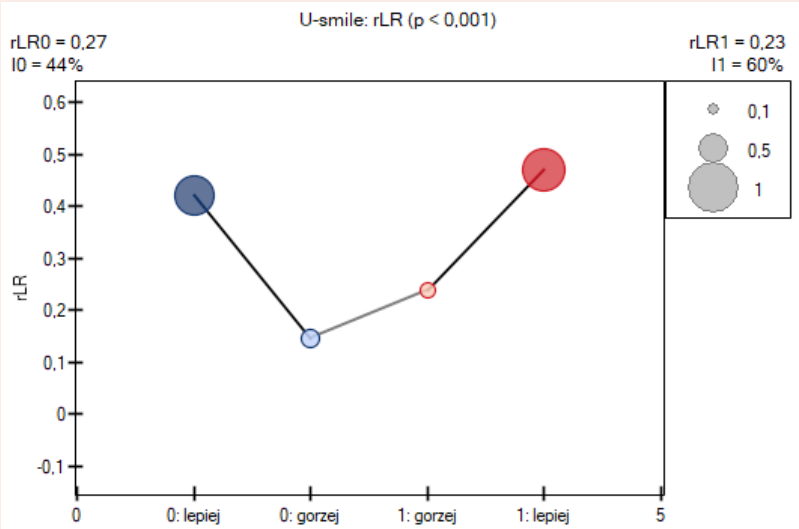
Regresja logistyczna - predykcja/walidacja	
Analizowane zmienne	WIEK
	WYKSZTAŁCENIE
	CZAS
	ZAKŁÓCENIA
	ROZWIĄZANIE
Filtr danych	zbiór=testowy
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	60
Walidacja	
Test ilorazu wiarygodności	
Log Likelihood	-30.3865669
-2 Log Likelihood	60.7731337
Log Likelihood (w. wolny)	-87.8809897
-2 Log Likelihood (w. wolny)	175.7619793
Statystyka Chi-kwadrat	114.9888456
Stopnie swobody	4
Wartość p	0
AIC - kryterium Akaikego	68.7731337
AICc - skorygowane kryterium Akaikego	69.5004065
BIC - bayesowskie kryterium Schwarza	77.150512
Pseudo R2 (McFadden)	0.6542305
R2(Nagelkerke)	0.9010184
R2(Coxa-Snella)	0.8528762
Hosmer-Lemeshow test	
Statystyka Chi-kwadrat	7.4731153
Stopnie swobody	8
Wartość p	0.4865496
Krzywe ROC	
AUC	0.8171429
SE(AUC)	0.0552507
-95% CI	0.7088534
+95% CI	0.9254323
Statystyka Z	4.160875
Wartość p dwustronne	0.0000317
Linia odcięcia	0.6132695

zbiór testowy: $Pseudo\, R^2 = 0.65$, $R^2(Nagelkerke) = 0.90$; $AUC[95\%CI] = 0.82[0.71, 0.93]$, **czułość** $= 73\%$, **swoistość** $= 64\%$

Wartości opisujące jakość dopasowania modelu na zbiorze testowym są nawet wyższe niż dla danych uczących, a jakość predykcji jest wciąż zadowalająca – model nadaje się do predykcji. Podobieństwo ocen na zbiorze uczącym i testowym widzimy również na wykresie U-smile, gdzie pierwsza rycina pochodzi ze zbioru uczącego, a druga ze zbioru testowego.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





W tym celu skorzystamy z danych trzech nowych osób dopisanych na końcu zbioru. Wybierzemy opcję Predykcja, ustawimy filtr na nowy zbiór danych i użyjemy naszego modelu do tego, by przewidzieć, czy dana osoba rozwiąże zadanie poprawnie (uzyska wartość 1) czy też niepoprawnie (uzyska wartość 0).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Regresja logistyczna - predykcja/walidacja	
Analizowane zmienne	WIEK
	WYKSZTAŁCENIE
	CZAS
	ZAKŁÓCENIA
Filtr danych	zbiór=nowy
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	3
Predykcja	
Linia odcięcia	0.5
liczba 0	1
liczba 1	2
procent 0	33.333%
procent 1	66.667%

Współczynniki modelu	
wyraz wolny	6.7821012
WIEK	-0.1063451
WYKSZTAŁCENI	0.5040596
CZAS	-0.0837683
ZAKŁÓCENIA	-1.8477316

Predykcja	
Pred. prawdopo	Pred. Y
0.1744817	0
0.6393883	1
0.962964	1

Okazuje się, że prognoza dla pierwszej osoby jest negatywna, a dla dwóch kolejnych – pozytywna. Prognoza dla 50-letniej kobiety z wykształceniem podstawowym, rozwiązującej test podczas zakłóceń w czasie 20 min, wynosi 0.17, co oznacza, że prognozujemy, iż rozwiąże ona zadanie niepoprawnie, podczas gdy prognoza dla kobiety o 20 lat młodszej jest już korzystna – prawdopodobieństwo rozwiązania przez nią zadania wynosi 0.64. Największe prawdopodobieństwo (równe 0.96) poprawnego rozwiązania ma trzecia kobieta, która rozwiązywała test w ciągu 10 minut i bez zakłóceń.

Gdybyśmy chcieli postawić prognozę na podstawie innego modelu (np. uzyskanego podczas innego badania naukowego: ROZWIĄZANIE$= 6 - 0.1\cdot$WIEK$+0.5\cdot$WYKSZT$-0.1\cdot$CZAS$-2\cdot$ZAKŁÓCENIA) – wystarczy, że w oknie analizy wybierzemy nowy model, ustawimy jego współczynniki i prognozę dla wybranych osób można powtórzyć w oparciu o ten model.

Regresja logistyczna - predykcja/walidacja	
Analizowane zmienne	WIEK
	WYKSZTAŁCENIE
	CZAS
	ZAKŁÓCENIA
Filtr danych	zbiór=nowy
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	3
Predykcja	
Linia odcięcia	0.5
liczba 0	2
liczba 1	1
procent 0	66.667%
procent 1	33.333%

Współczynniki modelu	
wyraz wolny	6
wiek	-0.1
wykształcenie	0.5
czas	-0.1
zakłócenia	-2

Predykcja	
Pred. prawdopo	Pred. Y
0.0758582	0
0.3775407	0
0.9241418	1

Tym razem, zgodnie z prognozą nowego modelu, przewidywania dla pierwszej i drugiej osoby są nega-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

tywne, a trzeciej – pozytywne.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

25.4 PORÓWNYWANIE MODELI REGRESJI LOGISTYCZNEJ

Porównywanie modeli regresji logistycznej

KIEDY STOSOWAĆ:

Ze względu na możliwość jednoczesnej analizy wielu zmiennych niezależnych w jednym modelu regresji logistycznej, podobnie jak w liniowej regresji wielorakiej, pojawia się problem wyboru optymalnego modelu. Wybierając zmienne niezależne, należy pamiętać, by w modelu znajdowały się zmienne silnie skorelowane ze zmienną zależną i słabo skorelowane między sobą.

Porównywanie modeli regresji logistycznej pozwala odpowiedzieć na pytanie, czy dołożenie (lub usunięcie) zmiennych niezależnych istotnie zmienia jakość modelu. Jeśli dwa modele nie różnią się istotnie, warto wybrać ten prostszy – z mniejszą liczbą zmiennych. Dodatkowo, gdy interesuje nas praktyczna przydatność nowego czynnika ryzyka, można porównać, o ile poprawia się zdolność modelu do przewidywania badanego zdarzenia po dołączeniu tego czynnika.

WARUNKI STOSOWANIA

- porównywane modele zbudowane są na tej samej próbie (na tych samych obserwacjach),
- w porównaniu opartym na teście ilorazu wiarygodności model zredukowany jest „zagnieżdżony” w modelu pełnym, tzn. powstaje z niego przez usunięcie części zmiennych niezależnych,
- w porównaniu wartości prognostycznej (krzywe ROC, NRI, IDI) dla obu modeli wyznaczono prawdopodobieństwo przewidywane dla tych samych obiektów.

DEFINICJE MIAR

Porównując modele z różną liczbą zmiennych niezależnych, zwracamy uwagę na dopasowanie modelu oraz na kryteria informacyjne. Decyzję o tym, który model wybrać, podejmujemy na podstawie wielkości R^2_{Pseudo}, $R^2_{Nagelkerke}$, $R^2_{Cox-Snell}$ oraz wyniku testu ilorazu wiarygodności, porównującego kolejno powstające (sąsiednie) modele. Dla każdego modelu wyliczane jest bowiem maksimum funkcji wiarygodności, a test ilorazu wiarygodności sprawdza, czy modele pełny i zredukowany różnią się istotnie pod tym względem. Jeśli porównywane modele nie różnią się istotnie, powinniśmy wybrać ten z mniejszą liczbą zmiennych – brak różnicy oznacza bowiem, że zmienne obecne w modelu pełnym, a nieobecne w modelu zredukowanym, nie wnoszą istotnej informacji. Jeśli natomiast różnica jest istotna statystycznie, oznacza to, że jeden z modeli (ten z większą liczbą zmiennych, o większym R^2 i mniejszej wartości kryterium informacyjnego AIC, AICc lub BIC) jest istotnie lepszy niż drugi.

Porównanie wartości prognostycznej modeli – budowane modele regresji pozwalają przewidzieć prawdopodobieństwo wystąpienia badanego zdarzenia na podstawie analizowanych zmiennych niezależnych. Gdy znanych jest już wiele czynników zwiększających ryzyko wystąpienia zdarzenia, ważnym kryterium dla nowego kandydata na czynnik ryzyka jest to, na ile poprawia on skuteczność predykcji po dołączeniu go do modelu. Przykładowo, badając czynniki ryzyka choroby wieńcowej (wiek, ciśnienie skurczowe i rozkurczowe, otyłość, cholesterol, palenie), możemy sprawdzić, czy dołączenie kolejnego

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

czynnika w istotny sposób poprawia oszacowanie ryzyka choroby. Wartość prognostyczna modelu wynika z wyznaczonego dla każdej badanej osoby prawdopodobieństwa przewidywanego – im bliższe wartości 1, tym bardziej prawdopodobne jest wystąpienie zdarzenia. Na podstawie prawdopodobieństwa przewidywanego można wyznaczyć i porównać między modelami wielkość pola AUC pod krzywą ROC, a także współczynniki NRI i IDI.

- **Zmiana pola pod krzywą ROC** – krzywa ROC w modelach regresji logistycznej budowana jest na podstawie klasyfikacji przypadków do grupy doświadczającej zdarzenia lub nie oraz przewidywanego prawdopodobieństwa zmiennej zależnej. Im większe pole pod krzywą, tym trafniej prawdopodobieństwo wyznaczone przez model przewiduje rzeczywiste wystąpienie zdarzenia. Porównując pole pod krzywą ROC dla modelu pełnego i zredukowanego, sprawdzamy więc, czy dołożenie czynników istotnie poprawiło predykcję modelu (metoda DeLonga, opisana szerzej w rozdziale Porównywanie krzywych ROC).

- **Poprawa reklasyfikacji netto (NRI)** *(ang. Net Reclassification Improvement)* – miara oparta na tabeli reklasyfikacji, opisującej przesunięcie wartości prawdopodobieństwa przewidywanego w górę lub w dół po dołożeniu nowego czynnika do modelu. Wyznacza się ją oddzielnie dla obiektów doświadczających zdarzenia i dla tych, które go nie doświadczyły. NRI może być wyznaczany dla zadanego podziału prawdopodobieństwa przewidywanego na kategorie (NRI kategorialny) lub bez konieczności wyznaczania kategorii (NRI ciągły).

  – NRI **kategorialny** wymaga arbitralnego wyznaczenia punktów podziału wartości prawdopodobieństwa przewidywanego (maksymalnie 9 punktów, czyli maksymalnie 10 kategorii; najczęściej stosuje się jeden lub dwa punkty). Wartości kategorialnego NRI można porównywać między sobą tylko wtedy, gdy bazują na tych samych punktach podziału. Przykładowo, dla punktów podziału 0.1 i 0.3: jeśli w modelu „starym" (mniejszym) osoba uzyskała prawdopodobieństwo poniżej 0.1, a w modelu „nowym" – pomiędzy 0.1 a 0.3, to została reklasyfikowana w górę. Jeśli wartość z obu modeli mieści się w tym samym przedziale, osoba nie została reklasyfikowana, a jeśli prawdopodobieństwo z modelu „nowego" jest niższe niż w modelu „starym", mamy do czynienia z reklasyfikacją w dół.

modele regresji	reklasyfikacja w górę		brak reklasyfikacji		reklasyfikacja w dół	
prawdop. przewidywane	"stary"	"nowy"	"stary"	"nowy"	"stary"	"nowy"
[0.3 do 1]					⊕	
[0.1; 0.3)		⊕	⊕	⊕		⊕
[0; 0.1)	⊕					

  – NRI **ciągły** nie wymaga arbitralnego wyznaczania kategorii – każda, nawet najmniejsza zmiana prawdopodobieństwa w górę lub w dół w stosunku do prawdopodobieństwa z modelu „starego" jest traktowana jako przejście do kolejnej kategorii.

Uwaga! Stosowanie NRI ciągłego nie wymaga arbitralnego definiowania punktów podziału prawdopodobieństwa, jednak nawet niewielkie zmiany ryzyka (niemające odzwierciedlenia w obserwacjach klinicznych) mogą wpływać na wartość tego współczynnika. NRI kategorialny odzwierciedla natomiast tylko te zmiany, które badacz uznał za ważne, definiując punkty podziału prawdopodobieństwa. Ogólny współczynnik NRI mieści się w przedziale od -2 do 2, a współczynniki cząstkowe NRI_{events} i $NRI_{nonevents}$ – w przedziale od -1 do 1 (od -100% do 100%). Wartości dodatnie świadczą o korzystnej reklasyfikacji, a ujemne – o niekorzystnej reklasyfikacji na skutek dołożenia nowej zmiennej do modelu. Istotność statystyczną uzyskanej wartości NRI sprawdza się dedykowanym testem Z.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **Zintegrowana poprawa dyskryminacji (**IDI**)** *(ang. Integrated Discrimination Improvement)* – współczynnik pokazujący różnicę pomiędzy średnią zmianą prawdopodobieństwa przewidywanego (między modelem „starym" i „nowym") w grupie obiektów doświadczających zdarzenia a grupą obiektów, które zdarzenia nie doświadczyły. Podobnie jak dla NRI, istotność statystyczną IDI sprawdza się dedykowanym testem Z.

U-smile – opisana w rozdziale Regresja logistyczna metoda U-smile może być użyta również do porównania kolejnych (sąsiednich) modeli: modelem odniesienia jest model wcześniejszy na liście, a modelem nowym – następny. Raport zawiera współczynniki rLR, BA lub RB (do wyboru w oknie AUC-IDI-NRI) dla czterech podklas, współczynniki netto dla klas, współczynnik ogólny, udziały I oraz testy istotności; dla każdej pary można dołączyć wykres U-smile. Wartości p testów zakładają modele zagnieżdżone – dla par niezagnieżdżonych (metoda wszystkich podzbiorów) należy je traktować orientacyjnie.

W programie PQStat porównywanie modeli można przeprowadzić ręcznie lub automatycznie.

- **Ręczne** porównywanie modeli polega na zbudowaniu 2 modeli:
  - pełnego – modelu z większą liczbą zmiennych,
  - zredukowanego – modelu z mniejszą liczbą zmiennych; model taki powstaje z modelu pełnego po usunięciu zmiennych, które z punktu widzenia badanego zjawiska są zbędne.

  Wybór zmiennych niezależnych w porównywanych modelach, a następnie wybór lepszego modelu na podstawie uzyskanych wyników porównania, należy do badacza.

- **Automatyczny dobór modelu** – program sam buduje kolejne modele i wskazuje najlepszy według wybranego kryterium.

  Składniki modelu. Jednostką, którą metody automatyczne dodają do modelu lub z niego usuwają, jest składnik: pojedyncza zmienna, cała zmienna fikcyjna (wszystkie jej kolumny 0–1 naraz) albo interakcja. Obowiązuje zasada hierarchii: interakcja może znaleźć się w modelu tylko wtedy, gdy są w nim jej efekty główne, i tylko ona może zostać usunięta przed nimi.

  Kryterium doboru wybieramy w oknie analizy:
  - AIC, AICc lub BIC – kryteria informacyjne; lepszy jest model o mniejszej wartości kryterium. Ruch (dodanie lub usunięcie składnika) jest przyjmowany tylko wtedy, gdy obniża kryterium;
  - p testu zagnieżdżonego – test ilorazu wiarygodności porównujący model przed zmianą i po zmianie, z zadanym progiem p. Składnik jest usuwany, gdy p jest większe od progu, a dodawany, gdy p jest mniejsze od progu. Ta sama wartość p jest pokazywana w raporcie jako porównanie sąsiednich modeli;
  - błąd walidacji – średnia dewiancja wyznaczona na obserwacjach, które nie brały udziału w dopasowaniu modelu (opis poniżej: Ocena modeli poza zbiorem uczącym). Ruch jest przyjmowany tylko wtedy, gdy zmniejsza ten błąd. Kryterium wymaga włączonej opcji Walidacja – wybranie go przy wyłączonej walidacji włącza walidację krzyżową.

  Metody doboru:
  - **eliminacja wstecz** – start od modelu ze wszystkimi zmiennymi; w każdym kroku usuwany jest ten składnik, którego usunięcie najbardziej obniża kryterium; koniec, gdy żadne usunięcie już go nie obniża;
  - **selekcja w przód** – start od modelu zerowego (tylko wyraz wolny); w każdym kroku dodawany jest składnik najbardziej obniżający kryterium; koniec, gdy żadne dodanie już go nie obniża;

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- **krokowa dwukierunkowa** – start od modelu zerowego; w każdym kroku rozważane są zarówno dodania, jak i usunięcia składników i wybierany jest najlepszy ruch; koniec, gdy żaden ruch nie poprawia kryterium;
- **wszystkie podzbiory** – dopasowywane są wszystkie modele zgodne z zasadą hierarchii (maksymalnie 15 składników), a w raporcie prezentowany jest najlepszy model każdej wielkości; metoda korzysta z kryteriów informacyjnych albo z błędu walidacji (kryterium p nie jest w niej dostępne).

Wszystkie modele są liczone na tym samym zbiorze obserwacji – wiersze z brakami w którejkolwiek ze zmiennych kandydujących są odrzucane raz, przed analizą. Dzięki temu kryteria informacyjne, testy ilorazu wiarygodności oraz porównania krzywych ROC, NRI i IDI są między modelami porównywalne.

Raport zawiera: metodę i kryterium doboru, liczbę składników i dopasowanych modeli, kryteria informacyjne modelu zerowego, kolejne modele ścieżki doboru (każdy z wartością $-2\ln L$, kryteriami AIC, AICc, BIC, współczynnikami R^2_{Pseudo}, $R^2_{Nagelkerke}$, $R^2_{Cox-Snell}$ oraz tabelą współczynników i ilorazów szans) i porównanie każdej sąsiedniej pary modeli testem ilorazu wiarygodności (statystyka χ^2, stopnie swobody, wartość p). Jeśli w oknie analizy wybrano ocenę wartości prognostycznej, dla każdej sąsiedniej pary modeli wyznaczane są także porównanie krzywych ROC, NRI i IDI – niezależnie od tego, czy modele są zagnieżdżone. Najlepszy model według wybranego kryterium jest oznaczony w raporcie. Jeśli żaden składnik nie poprawia modelu zerowego, raport informuje, że najlepszy jest model zerowy. Przy metodzie wszystkich podzbiorów sąsiednie modele mogą nie być zagnieżdżone – wówczas test ilorazu wiarygodności nie jest wykonywany.

Uwaga! Wartości p i przedziały ufności w modelu wybranym automatycznie są zbyt optymistyczne, ponieważ ten sam zbiór danych posłużył do wyboru zmiennych i do oszacowania ich efektów. Dobór automatyczny warto traktować jako pomoc w poszukiwaniu modelu, a ostateczny wybór oprzeć na wiedzy merytorycznej. Wybrany model należy zbudować ponownie w oknie Regresja logistyczna i ocenić jego jakość.

Ocena modeli poza zbiorem uczącym

Miary dopasowania, kryteria informacyjne i testy, a także krzywe ROC, NRI i IDI, wyznaczane są na tych samych obserwacjach, na których oszacowano współczynniki modeli. Taka ocena jest zbyt optymistyczna – model przewiduje nowe przypadki zwykle gorzej, niż wynikałoby to z jego dopasowania, a różnica rośnie wraz z liczbą zmiennych. Dlatego, tak jak w porównywaniu modeli regresji wielorakiej, w oknie analizy można włączyć ocenę modeli na obserwacjach, które nie brały udziału w ich dopasowaniu (opcja Walidacja): **walidację krzyżową k-krotną** (każdy model jest dodatkowo dopasowywany k razy, za każdym razem bez jednej z k części danych, i oceniany na pominiętej części) albo **losowy zbiór testowy** (modele są dopasowywane na części uczącej, a oceniane na odłożonej części testowej). Podział jest losowy, wyznaczony przez Ziarno, i **warstwowany względem zmiennej zależnej**: obserwacje ze zdarzeniem i bez zdarzenia są dzielone osobno, dzięki czemu w każdej części odsetek zdarzeń jest zbliżony.

Dla każdego modelu raport podaje wówczas trzy dodatkowe miary:

- **średnia dewiancja** ($-2\ln L/n$) – wartość $-2\ln L$ wyznaczona z prawdopodobieństw przewidzianych poza zbiorem uczącym, podzielona przez liczbę ocenianych obserwacji. Im mniejsza, tym lepszy model; miara silnie „karze” przewidywania pewne, a błędne. To ona jest **błędem walidacji** – według niej wskazywany jest najlepszy model wg walidacji i na niej opiera się kryterium doboru błąd walidacji;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **AUC** – pole pod krzywą ROC zbudowaną z prawdopodobieństw przewidzianych poza zbiorem uczącym; opisuje zdolność modelu do odróżniania nowych obiektów ze zdarzeniem od obiektów bez zdarzenia (0.5 – model bezużyteczny, 1 – rozróżnianie doskonałe);
- **wskaźnik Briera** – średni kwadrat różnicy między prawdopodobieństwem przewidywanym a rzeczywistym wynikiem (1 – zdarzenie, 0 – brak zdarzenia)[201]; im mniejszy, tym lepsze przewidywania.

Punktem odniesienia jest **model zerowy**, który każdej obserwacji przypisuje częstość zdarzenia z części uczącej – jego średnią dewiancję i wskaźnik Briera raport podaje w bloku doboru modelu (AUC takiego modelu wynosi 0.5). Tam też wskazany jest **najlepszy model wg walidacji** – niezależnie od wybranego kryterium doboru. Ocena poza zbiorem uczącym działa także przy ręcznym porównywaniu modeli.

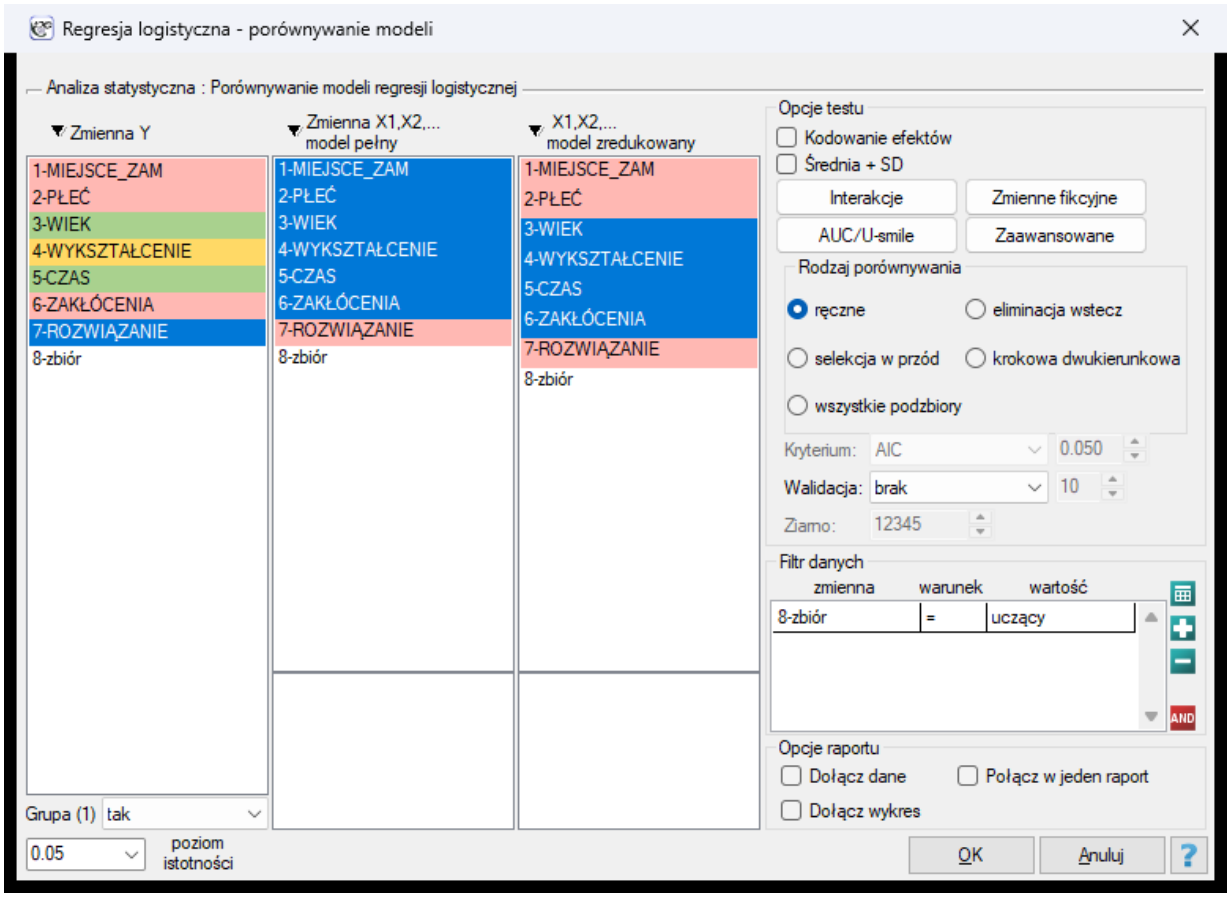
Przy włączonej walidacji również **porównanie wartości prognostycznej modeli** (krzywe ROC, NRI, IDI, U-smile) jest wyznaczane z prawdopodobieństw przewidzianych poza zbiorem uczącym: w walidacji krzyżowej dla wszystkich obserwacji, a przy zbiorze testowym – dla obserwacji zbioru testowego. Raport zaznacza to wierszem ROC, IDI, NRI, U-smile: z przewidywań walidacji. Wyniki te pokazują wtedy, czy dodatkowe zmienne poprawiają przewidywanie nowych przypadków, a nie tylko dopasowanie do danych, na których model zbudowano.

Uwaga! Gdy błąd walidacji jest kryterium doboru, błąd walidacji wybranego modelu jest nieco zaniżony – spośród wielu modeli wybrano ten, który na danym podziale wypadł najlepiej. Przy małej liczbie zdarzeń wyniki walidacji zależą od losowego podziału – warto wtedy sprawdzić je dla kilku wartości ziarna.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele wielowymiarowe → Regresja logistyczna – porównywanie modeli.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W oknie wskazujemy zmienną zależną Y z wartością oznaczającą zdarzenie oraz zmienne niezależne modelu pełnego i zredukowanego (w doborze automatycznym – zmienne kandydujące). Opcje doboru modelu:

- Rodzaj porównywania – ręczne albo jedna z metod doboru automatycznego,
- Kryterium – AIC, AICc, BIC, p (test LR / F) z progiem podawanym w polu obok albo błąd walidacji; w trybie ręcznym kryterium nie jest używane,
- Walidacja – brak (ocena tylko na zbiorze uczącym), krzyżowa, k = z liczbą podziałów w polu obok (od 2 do 100, domyślnie 10) albo zbiór testowy, % = z udziałem zbioru testowego w procentach obserwacji (od 5 do 90, domyślnie 30),
- Ziarno – ziarno losowania podziału danych (domyślnie 12345); to samo ziarno oznacza ten sam podział,
- AUC/U-smile – wybór porównań wartości prognostycznej (krzywe ROC, NRI, IDI, U-smile) i związanych z nimi wykresów.

WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Kontynuacja przykładu 25.5 (plik zadanie.pqs): w eksperymencie badającym umiejętność koncentracji, dla 130 osób zbioru uczącego, zbudowano model regresji logistycznej na podstawie następujących zmiennych:

zmienna zależna: ROZWIĄZANIE (tak/nie) – informacja o tym, czy zadanie zostało rozwiązane poprawnie czy też nie;

zmienne niezależne:

MIEJSCEZAM (1=miasto/0=wieś),
PŁEĆ (1=kobieta/0=mężczyzna),
WIEK (w latach),
WYKSZTAŁCENIE (1=podstawowe, 2=zawodowe, 3=średnie, 4=wyższe),
CZAS rozwiązywania (w minutach),
ZAKŁÓCENIA (1=tak/0=nie).

Sprawdzimy, czy wszystkie zmienne niezależne są w modelu niezbędne.

Ręczne porównywanie modeli

Na podstawie zbudowanego wcześniej modelu pełnego możemy podejrzewać, że zmienne MIEJSCEZAM i PŁEĆ mają niewielki wpływ na budowany model (tzn. na ich podstawie nie możemy z sukcesem dokonywać klasyfikacji). Sprawdzimy, czy ze statystycznego punktu widzenia model pełny jest lepszy niż model po usunięciu tych dwóch zmiennych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Porównywanie modeli regresji logistycznej	
Analizowane zmienne	ROZWIĄZANIE
	MIEJSCE_ZAM
	PŁEĆ
	WIEK
	WYKSZTAŁCENIE
	CZAS
	ZAKŁÓCENIA
Filtr danych	zbiór=uczący
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Liczność	130
Kryterium zbieżności zostało osiągnięte	
Liczba zmiennych w modelu 1	6
AIC - kryterium Akaikego	140.7082
AICc - skorygowane kryterium Akaikego	141.3912
BIC - bayesowskie kryterium Schwarza	157.9134
-2 Log Likelihood	128.7082
Pseudo R2	0.2677
R2(Nagelkerke)	0.4097
R2(Coxa-Snella)	0.3037
Kryterium zbieżności zostało osiągnięte	
Liczba zmiennych w modelu 2	4
AIC - kryterium Akaikego	139.0823
AICc - skorygowane kryterium Akaikego	139.4023
BIC - bayesowskie kryterium Schwarza	150.5524
-2 Log Likelihood	131.0823
Pseudo R2	0.2542
R2(Nagelkerke)	0.3924
R2(Coxa-Snella)	0.2909
Porównanie - Model 1 vs Model 2	
Chi-kwadrat - porównanie	2.374
Stopnie swobody	2
Wartość p	0.3051

Model 1

	wsp. b	błąd b	-95% CI	+95% CI	stat. Walda	wartość p	Iloraz szar	-95% CI	+95% CI
w. wolny	7.230601	1.870134	3.565206	10.895997	14.948697	0.00011	1381.0525	35.346728	53959.905
MIEJSCE_Z	-0.453242	0.450524	-1.336253	0.429769	1.012102	0.3144	0.635564	0.262829	1.536902
PŁEĆ	-0.454788	0.451304	-1.339327	0.429751	1.015501	0.313589	0.634582	0.262022	1.536875
WIEK	-0.100896	0.03159	-0.162812	-0.03898	10.200921	0.001404	0.904027	0.849751	0.96177
WYKSZTAŁ	0.455928	0.241805	-0.018	0.929857	3.555199	0.059359	1.577637	0.982161	2.534146
CZAS	-0.089395	0.027609	-0.143507	-0.035282	10.483921	0.001204	0.914484	0.866314	0.965333
ZAKŁÓCEN	-1.924	0.475056	-2.855092	-0.992908	16.402912	0.000051	0.146022	0.057551	0.370498

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

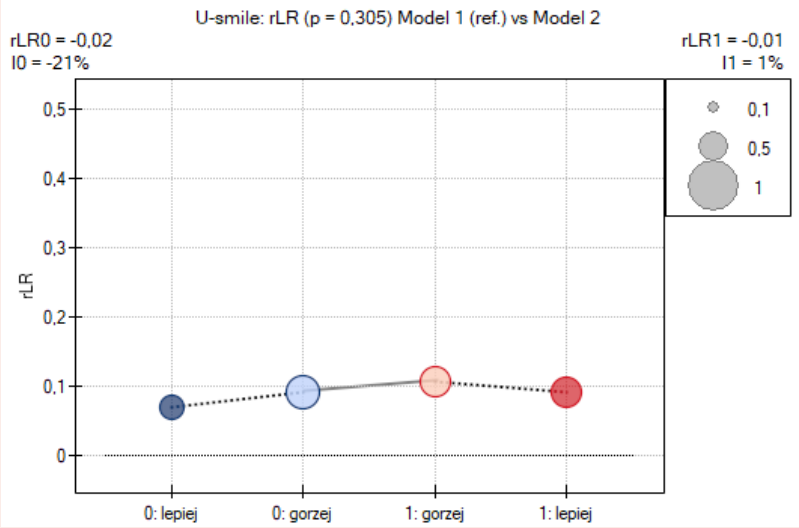
Model 2

	wsp. b	błąd b	-95% CI	+95% CI	stat. Walda	wartość p	Iloraz szar	-95% CI	+95% CI
w. wolny	6.782101	1.821888	3.211267	10.352936	13.857511	0.000197	881.91991	24.81049	31348.947
WIEK	-0.106345	0.031611	-0.168301	-0.044389	11.317897	0.000768	0.899114	0.845099	0.956582
WYKSZTAŁ	0.50406	0.23731	0.038941	0.969178	4.511625	0.033665	1.655428	1.039709	2.635777
CZAS	-0.083768	0.026838	-0.13637	-0.031166	9.741995	0.001801	0.919644	0.872519	0.969315
ZAKŁÓCEN	-1.847732	0.46198	-2.753195	-0.942268	15.99675	0.000063	0.157594	0.063724	0.389743

test ilorazu wiarygodności: $p = 0.3051$

Nie ma podstaw, by uważać, że model pełny jest lepszy niż model zredukowany. Zatem, przy nieznacznej utracie jakości modelu, miejsce zamieszkania i płeć mogą zostać pominięte.

Dobrym obrazem działania testu ilorazu wiarygodności jest wykres U-smile. Widzimy, że wiarygodność predykcji porównywanych modeli jest dość podobna – wykres prawie płaski (nieco „smutna mina”). Mniejszy model jest tylko nieistotnie statystycznie gorszy niż model większy – przerywana linia w obu częściach wykresu świadczy o nieistotności statystycznej pogorszenia zarówno w grupie osób, które nie rozwiązały zadania (lewa część wykresu), jak i w grupie osób, które rozwiązały zadanie (prawa część wykresu).



Porównania obu modeli pod względem zdolności do klasyfikacji można dokonać, porównując krzywe ROC dla tych modeli. Uzyskany raport, podobnie jak wcześniejszy, wskazuje, że modele nie różnią się jakością predykcji, tzn. wartości p zarówno dla porównania krzywych ROC, jak i dla wskaźników NRI i IDI są nieistotne statystycznie. Decydujemy zatem pominąć płeć i miejsce zamieszkania w ostatecznym modelu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Porównanie - Model 1 vs Model 2	
Porównanie krzywe ROC [Różnica AUC]	
AUC Model 1	0,834599
SE(AUC)	0,035518
-95% CI	0,764986
+95% CI	0,904213
AUC Model 2	0,823083
SE(AUC)	0,036776
-95% CI	0,751003
+95% CI	0,895162
AUC1-AUC2 [DeLong's method]	0,011517
SE(AUC1-AUC2)	0,010864
-95% CI	0
+95% CI	0,032809
Statystyka Z	1,06013
Wartość p	0,289085
Chi-kwadrat - porównanie modeli Model 1 vs Model 2	2,374039
Stopnie swobody	2
Wartość p	0,305129
U-smile Model 1 (ref.) vs Model 2	
Współczynnik	rLR
Obserwacje bez zmiany predykcji	0
rLR	
nie-zdarzenia (0), lepsza predykcja	0,069448
nie-zdarzenia (0), gorsza predykcja	0,090564
zdarzenia (1), gorsza predykcja	0,105699
zdarzenia (1), lepsza predykcja	0,090821
netto nie-zdarzenia (0)	-0,021116
netto zdarzenia (1)	-0,014878
ogółem	-0,018445
p (nie-zdarzenia (0))	0,17212
p (zdarzenia (1))	0,452516
p (LR ogółem)	0,305129
I (udział podklasy)	
nie-zdarzenia (0), lepsza predykcja	0,396226 (21)
nie-zdarzenia (0), gorsza predykcja	0,603774 (32)
zdarzenia (1), gorsza predykcja	0,493506 (38)
zdarzenia (1), lepsza predykcja	0,506494 (39)
netto nie-zdarzenia (0)	-0,207547
netto zdarzenia (1)	0,012987
ogółem	-0,19456

Ocena poza zbiorem uczącym

Powyższe porównania dotyczą tych samych 130 osób, na których dopasowano oba modele. Sprawdzimy, jak modele przewidują wynik osób, które nie brały udziału w ich dopasowaniu. W oknie analizy pozostawiamy ręczne porównywanie tych samych dwóch modeli, a w polu Walidacja wybieramy krzyżowa, k = i wpisujemy 10 podziałów (ziarno 12345).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Ocena poza zbiorem uczącym	walidacja krzyżowa
Liczba podziałów (k)	10
Ziarno losowania podziału	12345
Model zerowy: średnia dewiancja	1.3532
Model zerowy: wskaźnik Briera	0.2418
Najlepszy model wg walidacji (średnia dewiancja)	Model 2

Kryterium zbieżności zostało osiągnięte	
Liczba zmiennych w modelu 2	4
AIC - kryterium Akaikego	139.0823
AICc - skorygowane kryterium Akaikego	139.4023
BIC - bayesowskie kryterium Schwarza	150.5524
-2 Log Likelihood	131.0823
Pseudo R2	0.2542
R2(Nagelkerke)	0.3924
R2(Coxa-Snella)	0.2909
Walidacja krzyżowa: średnia dewiancja (-2 ln L / n)	1.0921
Walidacja krzyżowa: AUC	0.7932
Walidacja krzyżowa: wskaźnik Briera	0.1859

		zbiór uczący		walidacja krzyżowa ($k = 10$)		
model	zmienne	AIC	$-2 \ln L/n$	śr. dewiancja	AUC	Brier
zerowy	0			1.3532		0.2418
Model 1 (pełny)	6	140.7082	0.9901	1.1246	0.7856	0.1920
Model 2 (zreduk.)	4	139.0823	1.0083	1.0921	0.7932	0.1859

Oba modele przewidują wynik znacznie lepiej niż model zerowy (średnia dewiancja około 1.1 wobec 1.35). Wartości uzyskane poza zbiorem uczącym są przy tym wyraźnie gorsze niż wartości policzone na zbiorze uczącym (dla modelu pełnego 1.1246 wobec $128.7082/130 = 0.9901$) – ocena na danych uczących była zbyt optymistyczna. Co ważniejsze, kolejność modeli się odwraca: na zbiorze uczącym model pełny jest dopasowany lepiej (jak każdy model z większą liczbą zmiennych), natomiast poza zbiorem uczącym lepszy jest model zredukowany – ma mniejszą średnią dewiancję (1.0921 wobec 1.1246), większe pole AUC (0.7932 wobec 0.7856) i mniejszy wskaźnik Briera (0.1859 wobec 0.1920). Miejsce zamieszkania i płeć nie tylko nie poprawiają istotnie dopasowania, ale wręcz nieco pogarszają przewidywania dla nowych osób.

walidacja krzyżowa ($k = 10$): średnia dewiancja 1.1246 (model pełny) i 1.0921 (model zredukowany); model zerowy: 1.3532

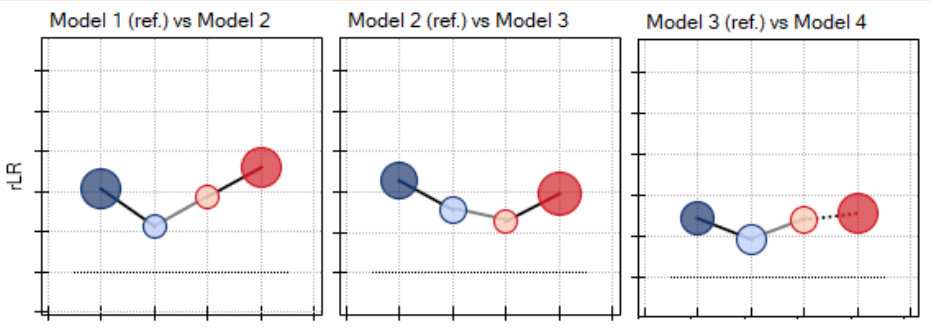
Najlepszym modelem według walidacji jest model zredukowany – pominięcie miejsca zamieszkania i płci poprawia przewidywanie poza zbiorem uczącym.

Automatyczny dobór modelu

Dobór automatyczny (dobór dwukierunkowy, kryterium AIC) daje bardzo podobne wyniki. Najlepszym modelem jest model zbudowany na podstawie zmiennych niezależnych: WIEK, WYKSZTAŁCENIE, CZAS rozwiązywania, ZAKŁÓCENIA. Na wykresach U-smile widać poszczególne kroki budowy modelu. W ostat-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

nim kroku dodano wykształcenie – widzimy, że poprawiło ono w sposób istotny statystycznie już tylko jedną część wykresu – tę dotyczącą osób, które nie rozwiązały zadania (linia ciągła).



Na podstawie powyższych analiz, ze statystycznego punktu widzenia, optymalnym modelem jest model zawierający 4 najważniejsze zmienne niezależne: WIEK, WYKSZTAŁCENIE, CZAS rozwiązywania, ZAKŁÓCENIA. Dokładną jego analizę możemy przeprowadzić w module Regresja Logistyczna. Jednak ostateczna decyzja, który model wybrać, należy do eksperymentatora.

PRZYKŁAD 25.7. (plik Serce.pqs)
Badano czynniki ryzyka pewnej choroby serca, takie jak wiek, BMI, palenie, cholesterol we frakcji LDL, cholesterol we frakcji HDL i nadciśnienie. Z punktu widzenia badacza interesujące było określenie, jak bardzo informacja o paleniu może poprawić predykcję występowania badanej choroby.

Porównujemy model regresji logistycznej opisujący ryzyko choroby serca na podstawie wszystkich badanych zmiennych z modelem pozbawionym informacji o paleniu. W oknie analizy zaznaczamy opcje związane z oceną predykcji, czyli krzywą ROC oraz współczynniki NRI. Dodatkowo wskazujemy, by w raporcie znalazły się wszystkie proponowane wykresy.

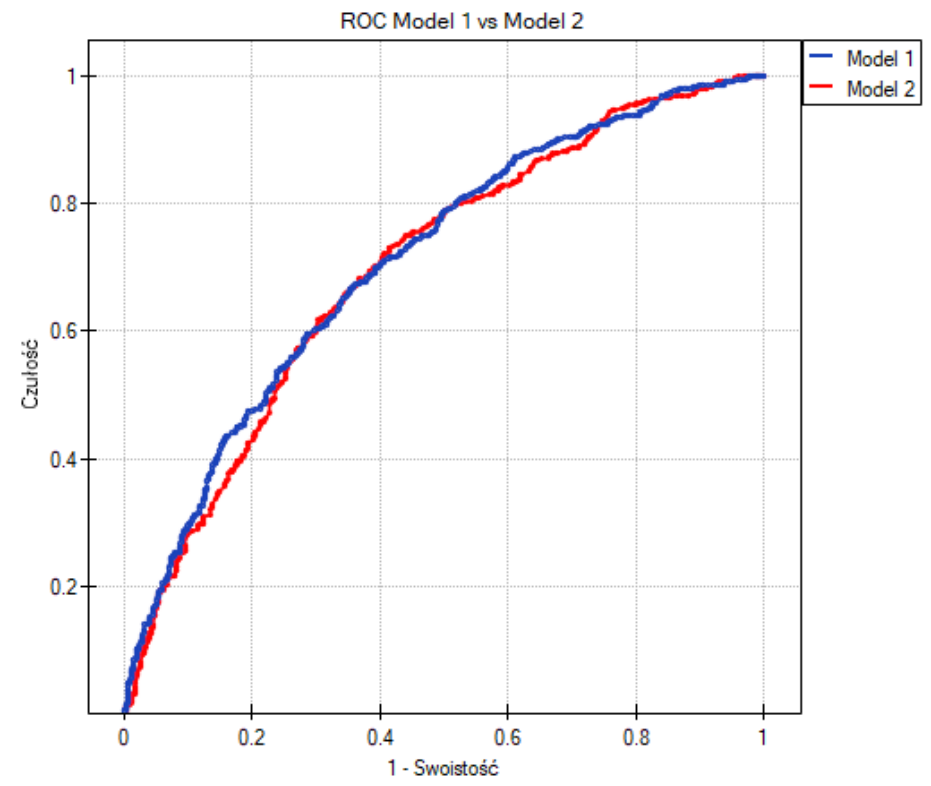
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Porównanie - Model 1 vs Model 2	
Porównanie krzywe ROC [Różnica AUC]	
AUC Model 1	0.709
SE(AUC)	0.0138
-95% CI	0.682
+95% CI	0.736
AUC Model 2	0.699
SE(AUC)	0.014
-95% CI	0.6716
+95% CI	0.7265
AUC1-AUC2 [DeLong's method]	0.01
SE(AUC1-AUC2)	0.0053
-95% CI	0
+95% CI	0.0203
Statystyka Z	1.9033
Wartość p	0.057
IDI Zintegrowana poprawa dyskryminacji Model 1 vs	0.0142
-95% CI	0.0078
+95% CI	0.0206
Statystyka Z	4.3486
Wartość p	<0.0001
NRI Poprawa reklasyfikacji netto (ciągła) Model 1 vs	0.1855
-95% CI	0.0805
+95% CI	0.2905
Statystyka Z	3.4634
Wartość p	0.0005
NRI (chory)	0.0522
NRI (zdrowy)	0.1333

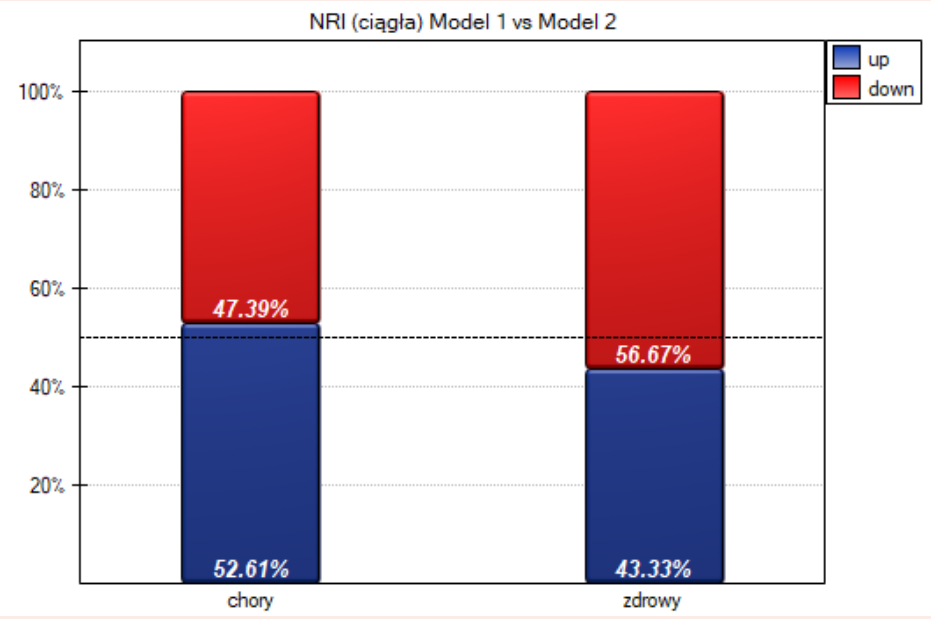
krzywa ROC: $p = 0.057$**;** NRI **i** IDI **ciągłe:** $p < 0.05$**,** $NRI(chory) = 0.0522$**,** $NRI(zdrowy) = 0.1333$

Dodanie do modelu informacji o paleniu wiąże się z ważną różnicą w predykcji, choć różnica ta nie jest istotna statystycznie w opisie krzywej ROC. Ciągłe wartości NRI i IDI wskazują natomiast na istotną statystycznie i korzystną zmianę: prognoza dla osób z chorobą serca poprawiła się o ponad 5%, a dla osób bez tej choroby – o ponad 13%.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



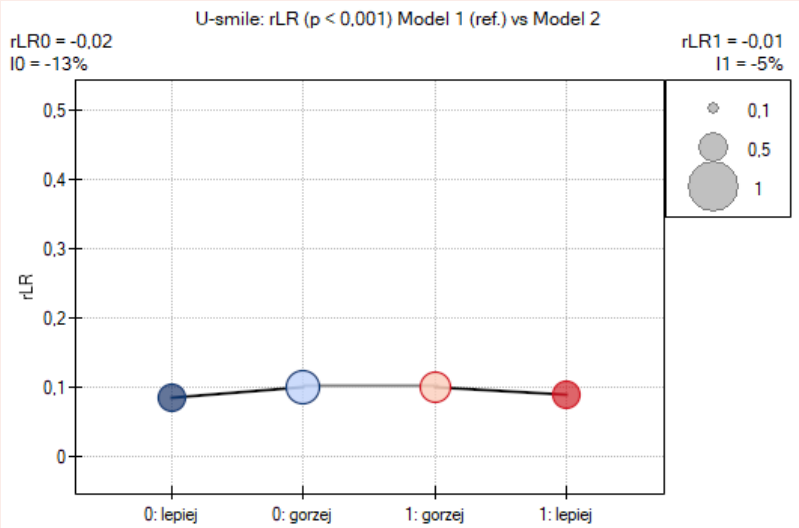
Wnioski wyciągnięte na podstawie NRI widzimy również na wykresie. Wzrost prognozowanego przez model prawdopodobieństwa choroby obserwujemy u osób chorych (więcej osób zostało przeklasyfikowanych w górę niż w dół: 52.61% vs 47.39%), natomiast spadek prawdopodobieństwa dotyczy w większym stopniu osób zdrowych (więcej osób zostało przeklasyfikowanych w dół niż w górę: 56.67% vs 43.33%).



Istotne statystycznie pogorszenie modelu na skutek usunięcia z niego „palenia” widać w obu gałęziach

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

wykresu U-smile i w tym, że wszystkie jego współczynniki są ujemne.



Istnieje też możliwość wyznaczenia NRI kategorialnego, ale w tym celu należałoby najpierw ustalić przyjęte w literaturze dotyczącej chorób serca punkty odcięcia prawdopodobieństwa wyznaczonego przez model.

DLA ZAINTERESOWANYCH

Test ilorazu wiarygodności
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad L_{FM} = L_{RM}, \\ \mathcal{H}_1: &\quad L_{FM} \neq L_{RM}, \end{aligned}$$

gdzie L_{FM}, L_{RM} – maksimum funkcji wiarygodności w porównywanych modelach (pełnym i zredukowanym).

Wzory:
Statystyka testowa ma postać:

$$\chi^2 = -2\ln(L_{RM}/L_{FM}) = -2\ln(L_{RM}) - (-2\ln(L_{FM}))$$

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z $df = k_{FM} - k_{RM}$ stopniami swobody, gdzie k_{FM} i k_{RM} to liczba szacowanych parametrów w porównywanych modelach.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Ocena poza zbiorem uczącym
Niech $\hat{p}_{(-i)}$ oznacza prawdopodobieństwo zdarzenia przewidziane dla obserwacji i przez model dopasowany bez tej części danych, do której należy obserwacja i (w walidacji krzyżowej – bez jej podziału, przy zbiorze testowym – na zbiorze uczącym), a y_i – wynik (1 – zdarzenie, 0 – brak zdarzenia). Dla n_v ocenianych obserwacji:

$$\text{średnia dewiancja} = -\frac{2}{n_v}\sum_{i=1}^{n_v}\left[y_i \ln \hat{p}_{(-i)} + (1-y_i)\ln\left(1-\hat{p}_{(-i)}\right)\right],$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

$$\text{wskaźnik Briera} = \frac{1}{n_v}\sum_{i=1}^{n_v}\left(y_i - \hat{p}_{(-i)}\right)^2.$$

AUC jest polem pod krzywą ROC wyznaczoną łącznie ze wszystkich prawdopodobieństw $\hat{p}_{(-i)}$. W modelu zerowym $\hat{p}_{(-i)}$ jest częstością zdarzenia w odpowiedniej części uczącej. Podział jest wykonywany osobno wśród obserwacji ze zdarzeniem i bez zdarzenia (warstwowanie), jak opisano w rozdziale Porównywanie modeli regresji wielorakiej.

Zmiana pola pod krzywą ROC
Hipotezy:

$$\begin{aligned}\mathcal{H}_0: &\quad AUC_{FM} = AUC_{RM},\\ \mathcal{H}_1: &\quad AUC_{FM} \neq AUC_{RM}.\end{aligned}$$

Sposób wyznaczania statystyki testowej, oparty o metodę DeLonga, został opisany w rozdziale Porównywanie krzywych ROC.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Poprawa reklasyfikacji netto (NRI)
Wzory:
Definiujemy:

$$\hat{p}_{up,events} = \frac{\#events_{up}}{\#events}, \quad \hat{p}_{down,events} = \frac{\#events_{down}}{\#events}, \quad \hat{p}_{up,nonevents} = \frac{\#nonevents_{up}}{\#nonevents}, \quad \hat{p}_{down,nonevents} = \frac{\#non}{\#no}$$

gdzie:
$\#events_{up}$ – liczba obiektów z grupy doświadczającej zdarzenia, u których nastąpiła zmiana prawdopodobieństwa przewidywanego o przynajmniej jedną kategorię w górę,

$\#events_{down}$ – liczba obiektów z grupy doświadczającej zdarzenia, u których nastąpiła zmiana prawdopodobieństwa przewidywanego o przynajmniej jedną kategorię w dół,

$\#events$ – liczba obiektów w grupie doświadczającej zdarzenia,

$\#nonevents_{up}$ – liczba obiektów z grupy niedoświadczającej zdarzenia, u których nastąpiła zmiana prawdopodobieństwa przewidywanego o przynajmniej jedną kategorię w górę,

$\#nonevents_{down}$ – liczba obiektów z grupy niedoświadczającej zdarzenia, u których nastąpiła zmiana prawdopodobieństwa przewidywanego o przynajmniej jedną kategorię w dół,

$\#nonevents$ – liczba obiektów w grupie niedoświadczającej zdarzenia.

Ogólny współczynnik NRI oraz współczynniki wyrażające procentową zmianę klasyfikacji wyznacza się ze wzoru:

$$NRI = (\hat{p}_{up,events} - \hat{p}_{down,events}) - (\hat{p}_{up,nonevents} - \hat{p}_{down,nonevents})$$

$$NRI_{events} = \hat{p}_{up,events} - \hat{p}_{down,events}, \quad NRI_{nonevents} = \hat{p}_{down,nonevents} - \hat{p}_{up,nonevents},$$

Interpretacja:
Współczynnik NRI_{events} może być interpretowany jako **procent netto** prawidłowo reklasyfikowanych osób, u których wystąpiło zdarzenie, a $NRI_{nonevents}$ jako **procent netto** prawidłowo reklasyfikowanych osób, u których nie wystąpiło zdarzenie. Ogólny współczynnik NRI jest sumą współczynników NRI_{events} i $NRI_{nonevents}$, przez co jest współczynnikiem domyślnie ważonym przez częstość zdarzenia i nie może być interpretowany jako procent.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Test Z do sprawdzania istotności współczynnika NRI
Hipotezy:

$$\begin{aligned}\mathcal{H}_0: &\quad NRI = 0,\\ \mathcal{H}_1: &\quad NRI \neq 0.\end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{NRI}{SE(NRI)}$$

gdzie:

$$\begin{aligned}SE(NRI) = \quad & [\left(\frac{\#events_{up}+\#events_{down}}{\#events^2} - \frac{(\#events_{up}+\#events_{down})^2}{\#events^3}\right) + \\ & + \left(\frac{\#nonevents_{down}+\#nonevents_{up}}{\#nonevents^2} - \frac{(\#nonevents_{down}+\#nonevents_{up})^2}{\#nonevents^3}\right)]^{1/2}\end{aligned}$$

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Zintegrowana poprawa dyskryminacji (IDI)
Wzory:

$$IDI = \overline{p(diff)}_{events} - \overline{p(diff)}_{nonevents}$$

gdzie:
$\overline{p(diff)}_{events}$ – średnia różnicy wartości prawdopodobieństwa przewidywanego między modelami regresji („starym” i „nowym”) dla obiektów, które doświadczyły zdarzenia,
$\overline{p(diff)}_{nonevents}$ – średnia różnicy wartości prawdopodobieństwa przewidywanego między modelami regresji („starym” i „nowym”) dla obiektów, które nie doświadczyły zdarzenia.

Test Z do sprawdzania istotności współczynnika IDI
Hipotezy:

$$\begin{aligned}\mathcal{H}_0: &\quad IDI = 0,\\ \mathcal{H}_1: &\quad IDI \neq 0.\end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{IDI}{SE(IDI)}$$

gdzie:

$$SE(IDI) = \sqrt{\frac{sd(diff)^2_{events}}{\#events} + \frac{sd(diff)^2_{nonevents}}{\#nonevents}}$$

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

ANOVA czynnikowa – GLM

KIEDY STOSOWAĆ:

Czynnikowa analiza wariancji GLM jest rozszerzeniem jednoczynnikowej analizy wariancji (ANOVA) dla grup niezależnych oraz liniowej regresji wielorakiej. Skrót GLM *(ang. general linear model)* czytamy jako Ogólny Model Liniowy. Analiza GLM polega zwykle na wykorzystaniu modeli regresji liniowej w wyliczaniu różnych złożonych porównań ANOVA.

ANOVA czynnikowa jest takim rodzajem analizy wariancji, w którym możemy wykorzystać zarówno jeden, jak i wiele czynników, by wyodrębnić porównywane grupy. W analizie mogą brać udział również takie zmienne, które są interakcją wskazanych czynników. Gdy ANOVA zawiera więcej czynników niż jeden, wówczas czynniki te są wobec siebie uwikłane. ANOVA czynnikowa wymaga, by czynniki dzieliły się na poszczególne kategorie (tj. niezależne populacje), np. czynnik A: płeć dzielimy na męską i żeńską, czynnik B: wykształcenie na podstawowe, zawodowe, średnie i wyższe. Interakcja czynnika $A * B$ jest również dzielona na kategorie – w tym przypadku uzyskamy ich osiem.

Przykład równoważnych analiz, które mogą być przeprowadzone poprzez GLM. Analizy zawarte w poszczególnych wierszach tabeli są równoważne w tym sensie, że ich wyniki są tożsame, choć nie muszą być identyczne. Badanie dotyczy *dochodu* pewnej grupy osób, o których mamy dodatkowe informacje typu płeć i wykształcenie.

dochód oraz...	ANOVA	Regresja wieloraka
płeć	**t-Studenta dla zmiennych niezależnych** ⋆ porównanie dochodu dla kobiet z dochodem dla mężczyzn	**zależność dochodu (Y) od płci (X)** ⋆ płeć jako kolumna o dwóch kategoriach
wykształcenie	**ANOVA jednoczynnikowa** ⋆ porównanie dochodu dla osób z różnym wykształceniem (podstawowym, zawodowym, średnim, wyższym)	**zależność dochodu (Y) od wykształcenia (X)** ⋆ wykształcenie rozbite na zmienne fikcyjne
płeć, wykształcenie	**ANOVA wieloczynnikowa** ⋆ porównanie dochodu dla kobiet z dochodem dla mężczyzn przy uwzględnieniu wykształcenia jako zmiennej wikłającej ⋆ porównanie dochodu dla osób z różnym wykształceniem (podstawowym, zawodowym, średnim, wyższym) przy uwzględnieniu płci jako zmiennej wikłającej	**zależność dochodu (Y) od płci (X_1) i wykształcenia (X_2)** ⋆ płeć jako kolumna o dwóch kategoriach ⋆ wykształcenie rozbite na zmienne fikcyjne obie zmienne są wzajemnie wikłające
płeć, wykształcenie, płeć*wykształcenie	**ANOVA wieloczynnikowa** ⋆ porównanie dochodu dla kobiet z dochodem dla mężczyzn przy uwzględnieniu innych zmiennych w modelu ⋆ porównanie dochodu dla osób z różnym wykształceniem (podstawowym, zawodowym, średnim, wyższym) przy uwzględnieniu innych zmiennych w modelu ⋆ porównanie dochodu dla osób z różnym wykształceniem i płcią (w interakcji) przy uwzględnieniu innych zmiennych w modelu	**zależność dochodu (Y) od płci (X_1), od wykształcenia (X_2) oraz od interakcji płci i wykształcenia ($X_1 * X_2$)** ⋆ płeć jako kolumna o dwóch kategoriach ⋆ wykształcenie rozbite na zmienne fikcyjne ⋆ interakcja to iloczyn płci i wykształcenia zmienne w modelu są wzajemnie wikłające

Analiza GLM może być wykorzystana w każdym z powyższych przypadków, jednak ponieważ analiza regresji wielorakiej, podobnie jak jednoczynnikowa ANOVA, zostały omówione w oddzielnych rozdziałach, w tym rozdziale przedstawimy wykorzystanie GLM w ANOVA wieloczynnikowej.

WARUNKI STOSOWANIA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- pomiar na skali interwałowej,
- próby pochodzą z populacji o rozkładzie normalnym (normalność zmiennych bądź reszt modelu),
- model niezależny,
- równość wariancji badanych zmiennych porównywanych populacji.

DEFINICJE MIAR

Wpływ czynników wikłających – pomimo że wszystkie czynniki biorące udział w analizie są wobec siebie uwikłane, ich wpływ na istotność poszczególnych czynników można kontrolować. Istnieją trzy sposoby, przy pomocy których, badając istotność poszczególnych czynników, można uwzględniać wpływ zmiennych wikłających. Zależą one od sposobu wyznaczania sumy kwadratów:

Sumy kwadratów typu I – zależą od kolejności, w jakiej w modelu znajdują się poszczególne czynniki. Ten rodzaj sumy kwadratów powoduje, że istotność interpretowanego czynnika jest korygowana o te zmienne, których kolejność w modelu była wcześniejsza, a pozostałe zmienne w modelu wpływają jedynie pośrednio na wynik analizy.
Stosowanie: wskazane, gdy badanie jest w pełni zbalansowane, z równymi lub proporcjonalnymi licznościami poszczególnych kategorii, również wtedy gdy występują interakcje. Przeciwwskazane, gdy badanie jest niezbalansowane i/lub występują interakcje.

Sumy kwadratów typu II – istotność interpretowanego czynnika jest korygowana o te zmienne, których rząd jest taki sam lub niższy, a pozostałe zmienne w modelu wpływają jedynie pośrednio na wynik analizy.
Stosowanie: wskazane, gdy badanie jest w pełni zbalansowane, z równymi lub proporcjonalnymi licznościami poszczególnych kategorii, również wtedy gdy występują interakcje. Przeciwwskazane, gdy badanie jest niezbalansowane i/lub występują interakcje.

Sumy kwadratów typu III – istotność interpretowanego czynnika jest korygowana o wszystkie pozostałe zmienne w modelu. Zalecamy stosować ten rodzaj sumy kwadratów, gdy wybrane jest kodowanie efektów.
Stosowanie: wskazane, gdy badanie jest zbalansowane lub niezbalansowane, również wtedy gdy występują interakcje. Przeciwwskazane, gdy badanie zawiera podklasy o brakujących obserwacjach.

Uwaga! W PQStat domyślnie wybrane są sumy kwadratów typu III, ze względu na ich uniwersalność. Domyślnie zaznaczona jest też opcja kodowanie efektów, opisana w rozdziale Przygotowanie zmiennych do analizy w modelach wielowymiarowych. Należy pamiętać, że wybór odpowiedniego kodowania wpływa zarówno na interpretację współczynników modelu, jak i na istotność poszczególnych czynników w ANOVA czynnikowej – szczególnie przy niezbalansowanych układach.

Podejście ANOVA a podejście regresyjne – analiza typu ANOVA i modele regresji traktowane są równoważnie, i w ogólnym przypadku ich hipotezy są zbieżne. W interpretacji tych hipotez należy pamiętać, że hipotezy dla danych czynników korygowane są o te z pozostałych czynników, które dana analiza uwzględnia. Podejście modelowe zakłada działanie modelu regresji:

$$Y = \mu + \alpha_i A_i + \beta_j B_j + (\alpha\beta)_k B_k + \epsilon,$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

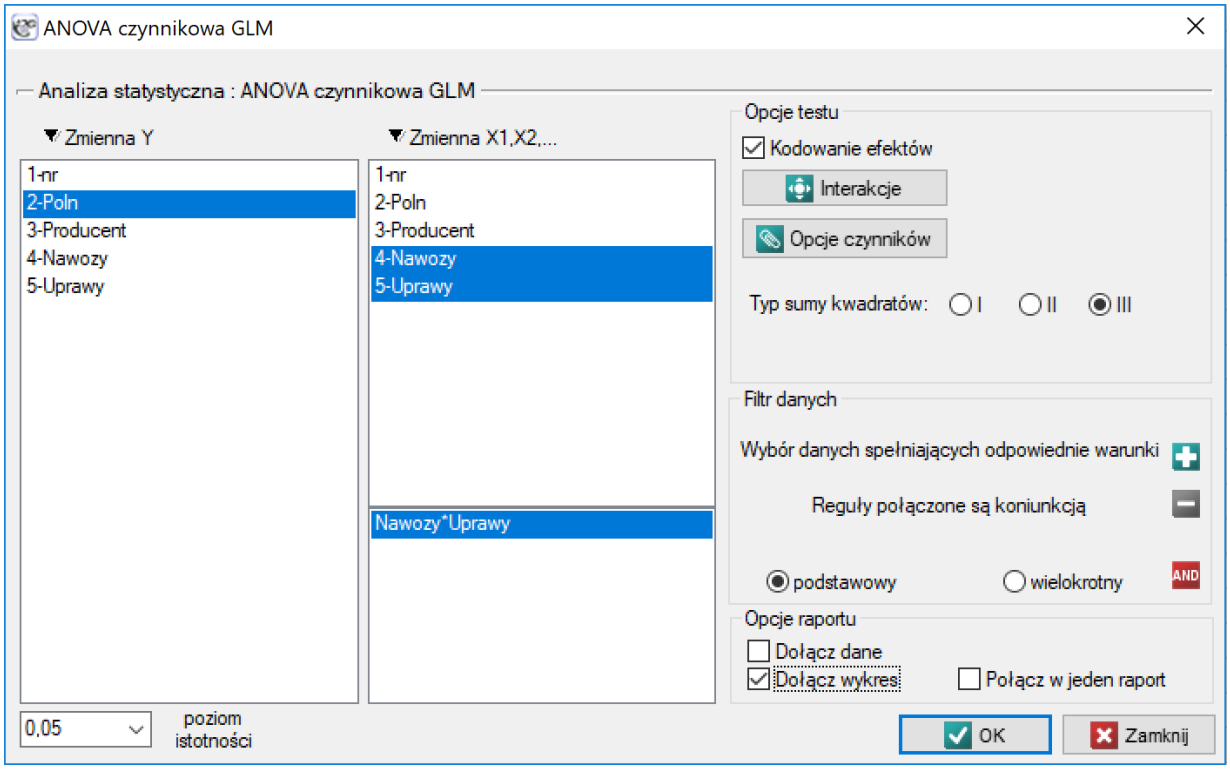
gdzie:
Y – zmienna zależna, objaśniana przez model,
μ – średnia ogólna zmiennej Y (o ile zastosowano kodowanie efektów),
A_i, B_j, AB_k – czynniki – zmienne niezależne, objaśniające,
$\alpha_i, \beta_i, \alpha\beta_i$ – parametry,
ϵ – składnik losowy (reszta modelu).

Kodowanie – uzyskiwane wyniki analiz (w szczególności budowanego modelu regresji) oraz interpretacja hipotez zależą również od sposobu kodowania. Program PQStat oferuje kodowanie **zero-jedynkowe** i kodowanie efektów. Domyślnie program wybiera kodowanie efektów – odznaczenie tej opcji jest równoważne z wybraniem kodowania zero-jedynkowego.

Uwaga! W przypadku stosowania sumy kwadratów typu III, gdy występują interakcje, wskazane jest stosowanie kodowania efektów.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele wielowymiarowe → ANOVA czynnikowa GLM.



WYKONANIE I INTERPRETACJA

Przykład 25.8. (plik plon.pqs)
Aby zwiększyć plon roślin uprawnych, opracowuje się nawozy według coraz nowszych technologii. Na

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

podstawie przeprowadzonego eksperymentu badacze chcą się dowiedzieć, która z trzech mieszanek nowych nawozów jest najbardziej skuteczna. Uprawy były prowadzone przez dwa różne gospodarstwa rolne i dotyczyły zasiewu pszenicy, żyta, owsa i jęczmienia. Plon podawano w % (w porównaniu do plonu uzyskanego bez nawożenia).

W pierwszej kolejności chcemy sprawdzić, czy: (1) średnie plony uzyskane przy zastosowaniu nawożenia mieszanką X są takie same jak uzyskane przy nawożeniu mieszanką Y i takie same jak przy nawożeniu mieszanką Z (niezależnie od gospodarstwa prowadzącego uprawę); (2) średnie plony uzyskane w gospodarstwie 1 są takie same jak w gospodarstwie drugim (niezależnie od mieszanki stosowanego nawozu). Równoważnie, w podejściu regresyjnym: (1) współczynniki określające zmianę uzyskanego plonu przy zmianie stosowanego nawożenia są zerowe; (2) współczynnik określający zmianę uzyskanego plonu przy zmianie gospodarstwa prowadzącego uprawę jest zerowy. W drugiej kolejności, stosując GLM, sprawdzimy czy: (3) średnie plony uzyskane z uprawy poszczególnych zbóż są takie same, gdy stosujemy różny sposób nawożenia.

Analizę przeprowadzimy, stosując trzeci typ sumy kwadratów i kodowanie efektów.

Test efektów międzygrupowych

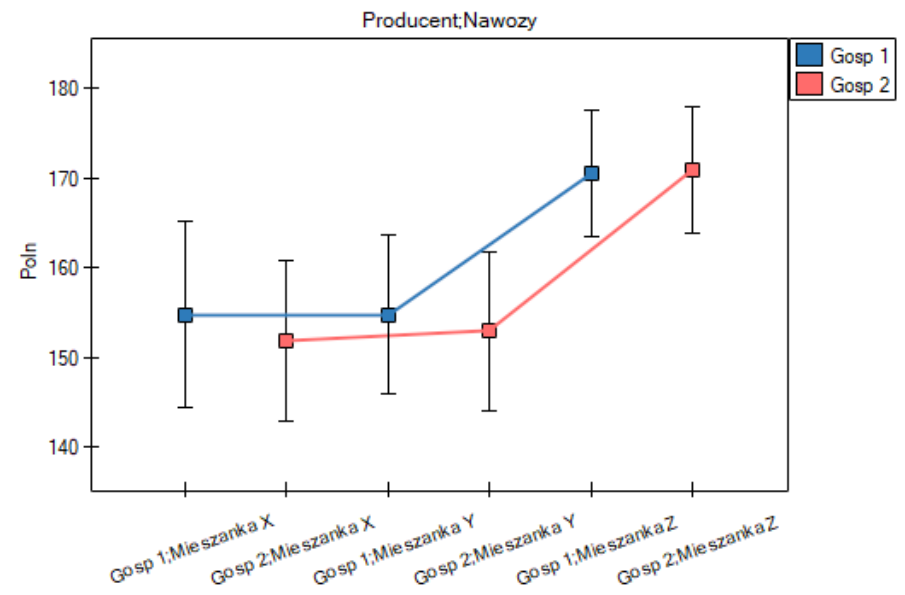
	SS	MS	DF	F	p	Eta-kwadrat cząstkowa
Wyraz wolny	4262167,1	4262167,1	1	8777,5923	<0.000001	0,981659
Nawozy	10957,726	5478,8630	2	11,28328	0,000026	0,120957
Producent	90,053571	90,053571	1	0,185458	0,667289	0,00113
within(błąd)	79634,071	485,57360	164			
total	90681,851					

mieszanka nawozów: $p = 0.000026$**, Eta-kwadrat cząstkowe** $\approx$ **12%; gospodarstwo:** $p = 0.667289$**, Eta-kwadrat cząstkowe = 0.1%**

Plon istotnie zależy od zastosowanej mieszanki nawozów, ale nie zależy od tego, w jakim gospodarstwie prowadzono uprawy.

Po wybraniu średnich obserwowanych lub oczekiwanych w oknie Opcji czynników, różnice te możemy przedstawić graficznie na wykresach obrazujących średnie plony przy stosowaniu poszczególnych mieszanek nawozów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Statystyki opisowe (obserwowane)						
Producent	liczność	Średnia	odch. stand	błąd stand.	-95% CI	+95% CI
Gosp 1	84	160,01190	23,783914	2,595038	154,85047	165,17333
Gosp 2	84	158,54761	22,930174	2,501887	153,57146	163,52377
Nawozy	liczność	Średnia	odch. stand	błąd stand.	-95% CI	+95% CI
Mieszanka X	56	153,28571	24,794218	3,313267	146,64577	159,92564
Mieszanka Y	56	153,85714	22,616509	3,022258	147,80040	159,91388
Mieszanka Z	56	170,69642	17,979705	2,402639	165,88143	175,51142

To, gdzie różnice są zlokalizowane, możemy sprawdzić, stosując testy post-hoc. Test post-hoc NIR Fishera wskazuje, iż najkorzystniejsze rezultaty przynosi stosowanie mieszanki Z – uzyskany plon stanowi średnio 170.7% plonu, który uzyskano by, nie stosując nawożenia. Pozostałe mieszanki nie różnią się istotnie statystycznie wielkością uzyskanego plonu. Ponieważ w modelu jednoczesnej analizie poddawano gospodarstwo, w którym prowadzone były uprawy, możemy powiedzieć, że przewaga mieszanki Z jest niezależna od tego, w którym gospodarstwie wykonano zasiew.

POST-HOC Fisher LSD			
	CD [F]	Statystyka [F]	Wartość p [F]
Nawozy: Mieszanak X;Mieszanak Y	8,222674	-0,137219	0,891026
Nawozy: Mieszanak X;Mieszanak Z	8,222674	-4,180884	0,000047
Nawozy: Mieszanak Y;Mieszanak Z	8,222674	-4,043665	0,000081

Podejście regresyjne

Analogiczną interpretację uzyskamy, posługując się modelem regresji, choć tutaj interpretacja jest nieco trudniejsza – trudność wynika z konieczności ustalenia sposobu kodowania i wyboru kategorii odniesienia. Przyjrzyjmy się najpierw wynikom otrzymanym przy kodowaniu zero-jedynkowym, które możemy uzyskać, odznaczając opcję kodowania efektów. Analiza automatycznie przyjęła alfabetycznie pierwszy poziom jako poziom odniesienia – dla nawozów poziomem tym była mieszanka X, dla gospodarstw było to gospodarstwo 1.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Poziomy czynnika Producent	2
Liczność 0 [Gosp 1] REF	84
Liczność 1 [Gosp 2]	84
Poziomy czynnika Nawozy	3
Liczność 0 [Mieszanka X] REF	56
Liczność 1 [Mieszanka Y]	56
Liczność 2 [Mieszanka Z]	56

Model

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	wartość p	stand. b	błąd stand.
w. wolny	154,01785	3,400188	147,30406	160,73164	45,296869	<0.000001		
Producent	-1,464286	3,400188	-8,178074	5,249502	-0,430648	0,667289	-0,031513	0,073176
Nawozy[1]	0,571429	4,164362	-7,651249	8,794106	0,137219	0,891026	0,011594	0,084496
Nawozy[2]	17,410714	4,164362	9,188037	25,633392	4,180884	0,000047	0,353268	0,084496

Analiza współczynników modelu przypomina analizę testów post-hoc, z tą różnicą, że porównujemy wyłącznie do kategorii odniesienia. Jeśli więc wszystkie mieszanki nawozów porównamy do mieszanki X, możemy zauważyć, że jedynie stosując mieszankę Z, uzyskano istotnie wyższe wyniki (p=0.000047) – wyniki te są wyższe o 17.410714 (średnie wynosiły odpowiednio 153.285714 dla mieszanki X i 170.696429 dla mieszanki Z). Porównując gospodarstwa, sprawa jest prosta, gdyż mamy do porównania tylko dwa gospodarstwa, a uzyskany wynik jest wynikiem porównania gospodarstwa 2 z gospodarstwem 1, które stanowiło kategorię odniesienia. Tym razem uzyskana różnica była niewielka (-1.464286) i nieistotna statystycznie (p=0.667289).

Stosując kodowanie efektów, również wybieramy kategorię odniesienia, ale wielkość współczynników i ich istotność nie jest odnoszona do wybranej kategorii odniesienia, ale do średniej ogólnej uzyskanego plonu, zapisanej w modelu jako wyraz wolny (159.279762).

Model

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	wartość p	stand. b	błąd stand.
w. wolny	159,27976	1,700094	155,92286	162,63665	93,688806	<0.000001		
Producent	-0,732143	1,700094	-4,089037	2,624751	-0,430648	0,667289	-0,031513	0,073176
Nawozy[1]	-5,422619	2,404296	-10,16998	-0,675254	-2,255388	0,025431	-0,190571	0,084496
Nawozy[2]	11,416667	2,404296	6,669302	16,164032	4,748445	0,000004	0,401225	0,084496

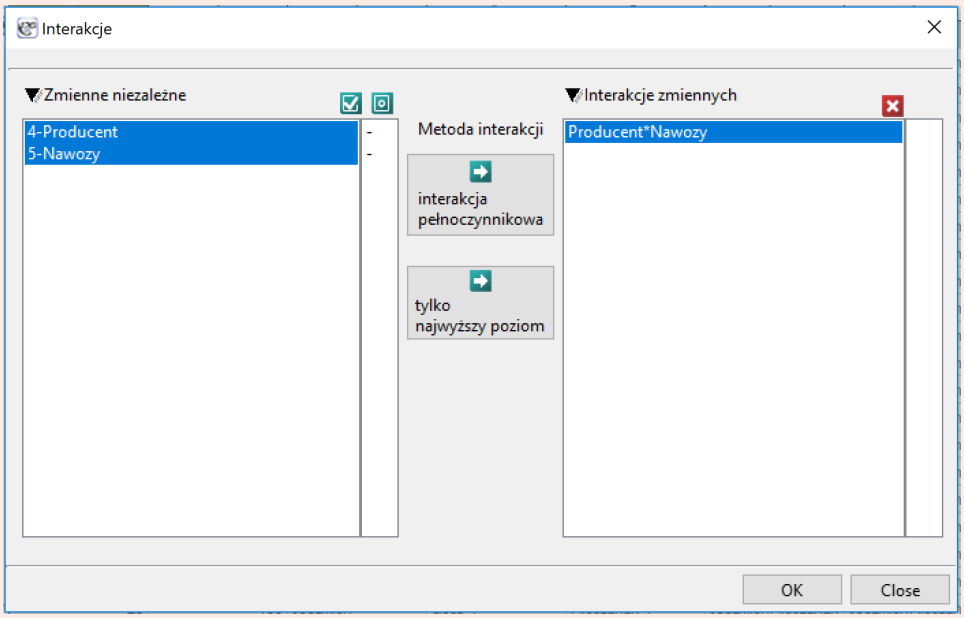
W porównaniu do średniej ogólnej znajdujemy sporo różnic: plon uzyskany przy nawożeniu mieszanką Y jest o 5.422619 niższy niż średnia ogólna, a mieszanką Z – o 11.416667 wyższy. Obie różnice są istotne statystycznie.

Niepodważalną zaletą budowania modelu regresyjnego jest możliwość wykorzystania jego formuły w przewidywaniu uzyskanych plonów. Zbudowane modele prezentują się następująco:
Dla kodowania zero-jedynkowego: $plon = 154.017857 + 0.571429 \cdot mieszankaY + 17.410714 \cdot mieszankaZ - 1.464286 \cdot gosp2$
Dla kodowania efektów: $plon = 159.279762 - 5.422619 \cdot mieszankaY + 11.416667 \cdot mieszankaZ - 0.732143 \cdot gosp2$

By móc zastosować wybrany model w prognozowaniu, należy udać się do menu Regresja wieloraka – predykcja i na podstawie nowych danych dokonać predykcji. Przy czym przygotowanie danych zależy od sposobu ich kodowania.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Na podstawie wszystkich uzyskanych wyników nie podejrzewamy, by wielkość plonu była zależna od interakcji między rodzajami stosowanych nawozów a gospodarstwem prowadzącym uprawy. Najczęściej występowanie interakcji widoczne jest na wykresie w postaci wyraźnie przecinających się linii – tu obie linie były prawie równoległe i na tyle bliskie sobie, że różnica między gospodarstwami była nieistotna statystycznie. Mimo że przecinające się linie najczęściej świadczą o występowaniu interakcji, należy pamiętać, że gdy linie znajdują się blisko siebie, ich przypadkowe przecięcie jest bardzo prawdopodobne, w efekcie czego interakcja nie będzie istotna statystycznie. Dla pewności sprawdzimy jednak, czy w naszym przypadku występuje interakcja. W tym celu obie zmienne wybierzemy raz jeszcze w oknie interakcji i przeniesiemy do listy interakcji umieszczonej po prawej stronie okna, a następnie powtórzymy analizę.



Test efektów międzygrupowych

	SS	MS	DF	F	p	Eta-kwadra
Wyraz wolny	4262167,1	4262167,1	1	8679,3182	<0.000001	0,981677
Producent	90,053571	90,053571	1	0,183382	0,669052	0,001131
Nawozy	10957,726	5478,8630	2	11,156953	0,000029	0,121065
Producent*Nawozy	80,464286	40,232143	2	0,081927	0,921377	0,00101
within(błąd)	79553,607	491,07164	162			
total	90681,851					

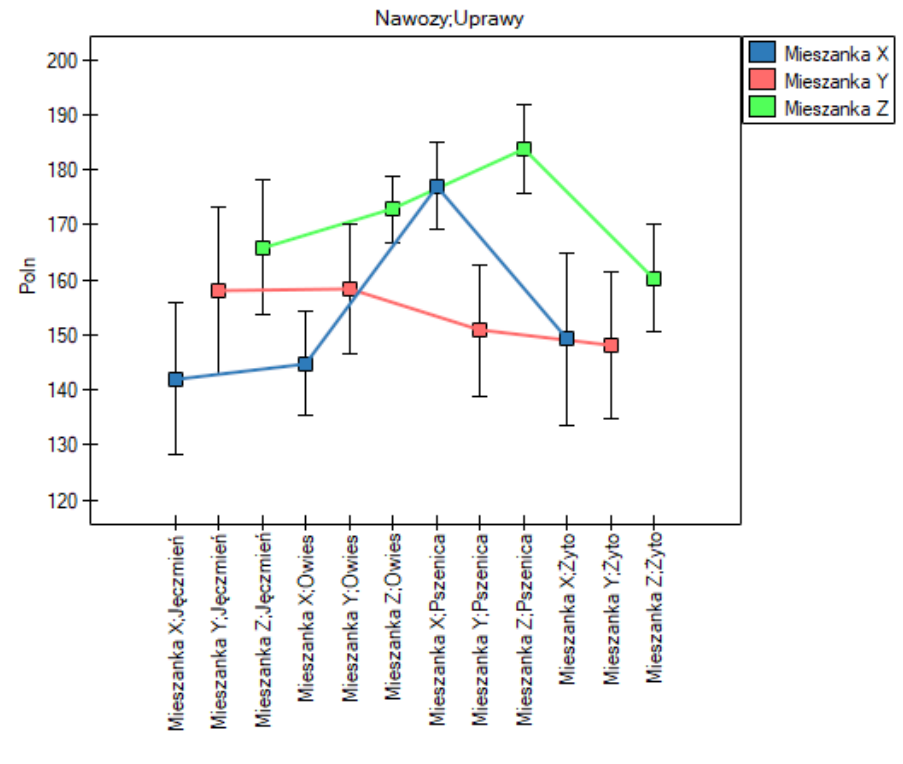
interakcja mieszanka*gospodarstwo: $p = 0.921377$ (nieistotne statystycznie)

Potwierdziliśmy brak istotnej interakcji – w tym wypadku zaleca się więc stosowanie modelu prostszego, tzn. pozbawionego interakcji.

Hipoteza 3)

Z odmienną sytuacją zetkniemy się, badając wielkość uzyskanego plonu w zależności od stosowanej dawki nawozu oraz w zależności od rodzaju uprawianego zboża.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Wykonamy analizę, która oprócz efektów głównych uwzględnia interakcje.

Test efektów międzygrupowych

	SS	MS	DF	F	p	Eta-kwadrat cząstkowa
Wyraz wolny	4262167,1	4262167,1	1	10483,111	<0.000001	0,985337
Nawozy	10957,726	5478,8630	2	13,475664	0,000004	0,147314
Uprawy	7851,1607	2617,0535	3	6,436834	0,000388	0,11015
Nawozy*Uprawy	8447,3214	1407,8869	6	3,462801	0,003066	0,117531
within(błąd)	63425,642	406,57463	156			
total	90681,851					

interakcja nawóz*rodzaj zboża: $p = 0.003066$ (istotne statystycznie)

Model z interakcjami powinniśmy stosować, a opis uzyskanych wyników skupić właśnie na tej interakcji.

W podejściu ANOVA hipoteza odnosząca się do interakcji dotyczy wszystkich możliwych par średnich (średnie plony uzyskane przy każdej kombinacji rodzaju zboża i mieszanki nawozu są takie same). W podejściu regresyjnym powiemy, że współczynniki określające zmianę uzyskanego plonu przy zmianie stosowanego nawożenia i zmianie rodzaju uprawy są zerowe.

Na podstawie wykresu (oraz średnich zapisanych w tabeli) widzimy, iż zdecydowanie najlepsze plony przynosi mieszanka Z, niezależnie od rodzaju uprawianego zboża.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Statystyki opisowe (obserwowane)						
Producent	liczność	Średnia	odch. stan	błąd stand.	-95% CI	+95% CI
Gosp 1	84	160,01190	23,783914	2,595038	154,85047	165,17333
Gosp 2	84	158,54761	22,930174	2,501887	153,57146	163,52377
Nawozy	liczność	Średnia	odch. stan	błąd stand.	-95% CI	+95% CI
Mieszanka X	56	153,28571	24,794218	3,313267	146,64577	159,92564
Mieszanka Y	56	153,85714	22,616509	3,022258	147,80040	159,91388
Mieszanka Z	56	170,69642	17,979705	2,402639	165,88143	175,51142

Natomiast mieszanka X i mieszanka Y uzyskują gorsze plony od mieszanki Z i dodatkowo zachodzi między nimi efekt interakcji. Przejawia się on tym, że uprawa pszenicy przynosi nietypowo wysoki plon w przypadku zastosowania mieszanki X w porównaniu do plonu pszenicy uzyskanego przy nawożeniu Y, podczas gdy uprawa jęczmienia i owsa lepiej plonują, gdy stosowana jest mieszanka Y. Dokładniej uzyskane różnice możemy sprawdzić, wykonując testy post-hoc. Fragment tego raportu zamieszczono poniżej:

POST-HOC Fisher LSD			
	CD [F]	Statystyka	Wartość p
Nawozy;Uprawy			
Mieszanak X;Jęczmień<>Mieszanak X;Owies	15,053977	-0,365524	0,715216
Mieszanak X;Jęczmień<>Mieszanak X;Pszenica	15,053977	-4,601849	0,000009
Mieszanak X;Jęczmień<>Mieszanak X;Żyto	15,053977	-0,955985	0,340559
Mieszanak X;Jęczmień<>Mieszanak Y;Jęczmień	15,053977	-2,099418	0,03739
Mieszanak X;Jęczmień<>Mieszanak Y;Owies	15,053977	-2,155652	0,032644
Mieszanak X;Jęczmień<>Mieszanak Y;Pszenica	15,053977	-1,152805	0,250754
Mieszanak X;Jęczmień<>Mieszanak Y;Żyto	15,053977	-0,815399	0,416087
Mieszanak X;Jęczmień<>Mieszanak Z;Jęczmień	15,053977	-3,139754	0,002023
Mieszanak X;Jęczmień<>Mieszanak Z;Owies	15,053977	-4,039505	0,000084
Mieszanak X;Jęczmień<>Mieszanak Z;Pszenica	15,053977	-5,473482	<0.000001
Mieszanak X;Jęczmień<>Mieszanak Z;Żyto	15,053977	-2,408707	0,017175
Mieszanak X;Owies<>Mieszanak X;Pszenica	15,053977	-4,236325	0,000039
Mieszanak X;Owies<>Mieszanak X;Żyto	15,053977	-0,590461	0,555736

Wynik testu post-hoc Fishera jest obszerny i potwierdza dużą i istotną statystycznie przewagę uzyskanego plonu przy stosowaniu mieszanki Z dla dowolnych upraw i mieszanki Y dla uprawy pszenicy.

Współczynniki modelu regresji możemy wykorzystać do prognozy poprzez menu regresja wieloraka – predykcja, pamiętając, by w zależności od wybranego modelu odpowiednio zakodować nowe dane.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

kodowanie zero - jedynkowe

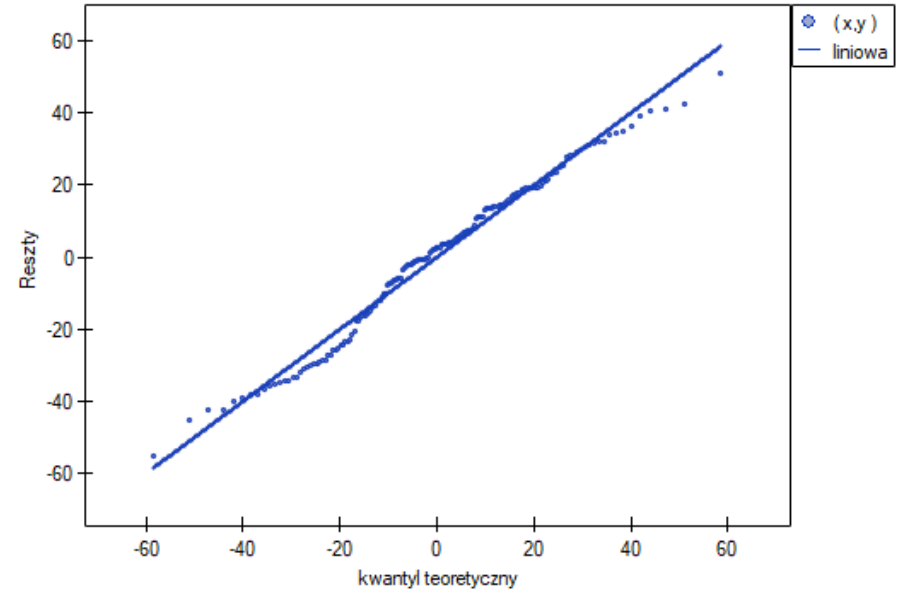
Model	wsp. b
w. wolny	142
Nawozy[1]	16
Nawozy[2]	23,9285
Uprawy[1]	2,78571
Uprawy[2]	35,0714
Uprawy[3]	7,28571
Nawozy[1]*Uprawy[1]	-2,3571
Nawozy[1]*Uprawy[2]	-42,285
Nawozy[1]*Uprawy[3]	-17,071
Nawozy[2]*Uprawy[1]	4,07142
Nawozy[2]*Uprawy[2]	-17,285
Nawozy[2]*Uprawy[3]	-12,857

kodowanie efektów

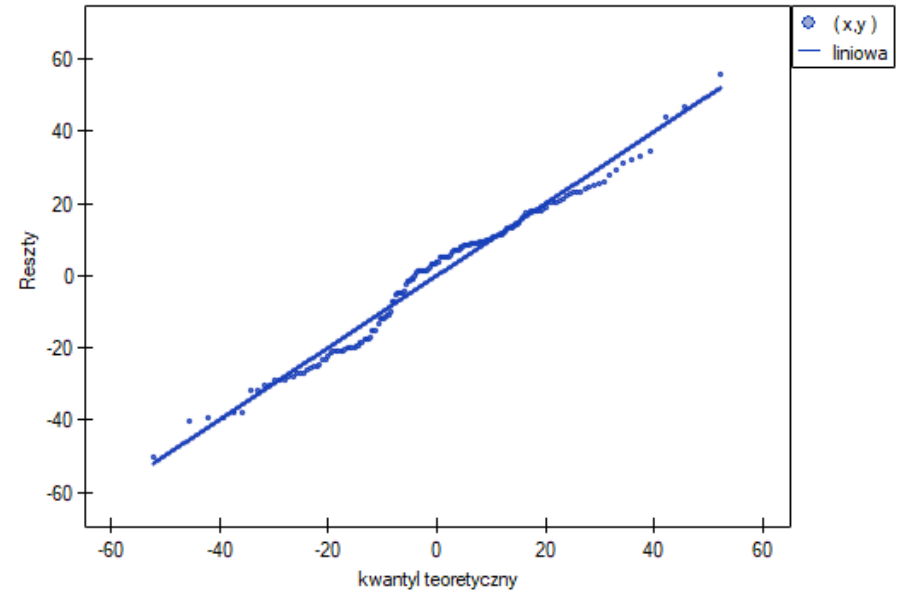
Model	wsp. b
w. wolny	159,2797
Nawozy[1]	-5,42261
Nawozy[2]	11,41666
Uprawy[1]	-0,61309
Uprawy[2]	11,24404
Uprawy[3]	-6,66071
Nawozy[1]*Uprawy[1]	5,184524
Nawozy[1]*Uprawy[2]	-14,3154
Nawozy[1]*Uprawy[3]	1,017857
Nawozy[2]*Uprawy[1]	2,702381
Nawozy[2]*Uprawy[2]	1,77381
Nawozy[2]*Uprawy[3]	-3,67857

Sprawdzenie założeń

Sprawdzenie głównych założeń będzie polegało na porównaniu wariancji oraz wizualnym określeniu normalności reszt modelu.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Wykres normalności reszt modelu typu Q-Q dla pierwszej oraz dla drugiej analizy przedstawia reszty modelu dobrze rozlokowane wokół prostej, co świadczy o dobrym dopasowaniu reszt do rozkładu normalnego. Porównaniu wariancji służy test Levene'a lub Browna-Forsythe'a. W przypadku tych testów możemy założyć, że uzyskane wyniki nie są jednoznaczne i są na pograniczu równości wariancji.

Jednorodność wariancji				
	Test Levene'a		Browna-Forsythea	
	Statystyka	Wartość p	Statystyka	Wartość p
Producent	0,126603	0,722432	0,219644	0,639926
Nawozy	2,954767	0,054854	3,223106	0,042352

Jednorodność wariancji				
	Test Levene'a		Browna-Forsythea	
	Statystyka	Wartość p	Statystyka	Wartość p
Nawozy	2,954767	0,054854	3,223106	0,042352
Uprawy	2,383389	0,071266	2,298122	0,079466
Nawozy*Uprawy	3,266295	0,00048	1,754733	0,066416

DLA ZAINTERESOWANYCH

Wzory – sumy kwadratów

Sumy kwadratów typu I

Na przykład: jeśli w modelu umieszczamy czynniki we wskazanej kolejności: A, B, $A*B$, C, $A*C$, $B*C$, $A*B*C$, D, wówczas istotność dla czynnika $A*C$ uwzględnia cały model (poprzez sumy kwadratów dla błędu), ale jako zmienne wikłające wykorzystywane są wprost tylko czynniki: A, B, $A*B$, C. Sumy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

kwadratów dla czynnika $A * C$ wylicza się wówczas następująco:

$$SS(A * C) = SS(A, B, A * B, C, A * C, B * C, A * B * C, D) - SS(A, B, A * B, C)$$

Sumy kwadratów typu II
Na przykład: jeśli w modelu umieszczamy czynniki: A, B, $A*B$, C, $A*C$, $B*C$, $A*B*C$, D, wówczas istotność dla czynnika $A * C$ uwzględnia cały model, ale jako zmienne wikłające wykorzystywane są wprost zmienne pierwszego rzędu: A, B, C, D oraz wszystkie pozostałe zmienne drugiego rzędu: $A*B$, $B * C$. Sumy kwadratów dla czynnika $A * C$ wylicza się wówczas następująco:

$$SS(A * C) = SS(A, B, A * B, C, A * C, B * C, A * B * C, D) - SS(A, B, A * B, C, B * C, D)$$

Sumy kwadratów typu III
Na przykład: jeśli w modelu umieszczamy czynniki: A, B, $A * B$, C, $A * C$, $B * C$, $A * B * C$, D, wówczas istotność dla zmiennej $A*C$ uwzględnia cały model, a jako zmienne wikłające wykorzystywane są wprost wszystkie czynniki za wyjątkiem badanego: A, B, $A * B$, C, $B * C$, $A * B * C$, D. Sumy kwadratów dla czynnika $A * C$ wylicza się wówczas następująco:

$$SS(A*C) = SS(A, B, A*B, C, A*C, B*C, A*B*C, D) - SS(A, B, A*B, C, B*C, A*B*C, D)$$

Hipotezy – podejście ANOVA
Hipotezy dla czynnika A:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_a, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_i \text{ są sobie równe } (i = 1, 2, ..., a), \end{aligned}$$

gdzie $\mu_1, \mu_2, ..., \mu_a$ – średnie czynnika A dla poszczególnych jego kategorii.

Hipotezy dla czynnika B:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_b, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, ..., b), \end{aligned}$$

gdzie $\mu_1, \mu_2, ..., \mu_b$ – średnie czynnika B dla poszczególnych jego kategorii.

Hipotezy dla interakcji czynników $A * B$:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_{ab}, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_k \text{ są sobie równe } (k = 1, 2, ..., ab), \end{aligned}$$

gdzie $\mu_1, \mu_2, ..., \mu_{ab}$ – średnie interakcji czynników $A * B$ dla poszczególnych ich kategorii.

Hipotezy – podejście regresyjne
Hipotezy dla czynnika A:

$$\begin{aligned} \mathcal{H}_0: &\quad \alpha_1 = \alpha_2 = ... = \alpha_a = 0, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \alpha_i = 0, \end{aligned}$$

Hipotezy dla czynnika B:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

$$\begin{aligned}\mathcal{H}_0: &\quad \beta_1 = \beta_2 = ... = \beta_b = 0,\\ \mathcal{H}_1: &\quad \text{nie wszystkie } \beta_j = 0\ ,\end{aligned}$$

Hipotezy dla interakcji czynników $A * B$:

$$\begin{aligned}\mathcal{H}_0: &\quad (\alpha\beta)_1 = (\alpha\beta)_2 = ... = (\alpha\beta)_{ab} = 0,\\ \mathcal{H}_1: &\quad \text{nie wszystkie } (\alpha\beta)_k = 0,\end{aligned}$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

ANCOVA

KIEDY STOSOWAĆ:

Analiza kowariancji (ANCOVA) jest metodą testowania hipotezy o równości średnich dwóch lub większej liczby populacji, w korekcji o inne ciągłe zmienne. Skutkiem tych korekt są efekty chętniej widziane przez badaczy niż te uzyskiwane poprzez ANOVA, tzn. węższe przedziały ufności i większa moc statystyczna.

Załóżmy, że przeprowadza się eksperyment w celu oceny efektów dwóch metod leczenia. Grupy, którym losowo przydzielono leczenie, różnią się nieco średnią wieku, która również wpływa na efekt leczenia. Różnice między grupami w osiągnięciach będą dość niejednoznaczne do zinterpretowania, ponieważ grupy różnią się zarówno pod względem wieku, jak i warunków leczenia. Analiza kowariancji dostarczy „skorygowanych średnich", które szacują wartość, jaką miałyby średnie wyniki, gdyby grupy były dokładnie takie same pod względem wieku. Jednocześnie zmienność wyników w obrębie grupy, wynikająca ze zmiennej (wiek), zostanie usunięta ze zmienności błędu, aby zwiększyć precyzję testu różnic między skorygowanymi średnimi.

Oznaczenie „analiza kowariancji" jest obecnie postrzegane jako anachroniczne przez niektórych metodologów badań i statystyków, ponieważ analiza ta nie jest odrębną analizą, ale wariantem ogólnego modelu liniowego (GLM). Jednak termin ten jest nadal użyteczny, ponieważ natychmiast przekazuje większości badaczy pojęcie, że zmienna kategorialna (np. warunki leczenia) i zmienna ciągła (np. wiek) są zaangażowane w jedną analizę określającą wynik leczenia.

WARUNKI STOSOWANIA

- pomiar zmiennej zależnej Y na skali interwałowej,
- próby pochodzą z populacji o rozkładzie normalnym (normalność zmiennych bądź reszt modelu),
- model niezależny względem czynników,
- równość wariancji reszt pomiędzy grupami uzyskanymi na bazie czynników,
- równość nachyleń linii regresji (współczynników regresji pomiędzy każdą zmienną wikłającą a zmienną zależną) dla każdego możliwego poziomu czynnika.

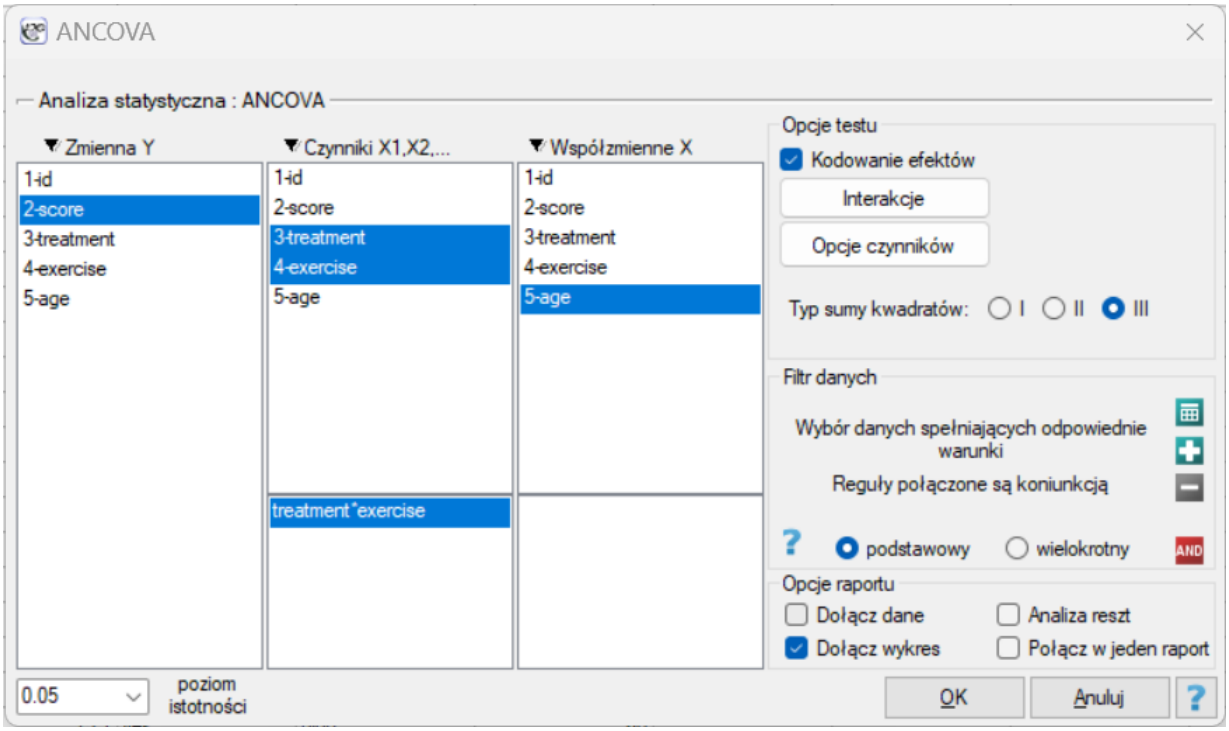
Uwaga! Sposób uwzględniania badanych czynników i zmiennych wikłających opisany jest w rozdziale dotyczącym wieloczynnikowej ANOVA (Wpływ czynników wikłających). Polecanym sposobem jest wybór Sumy kwadratów typu III oraz kodowania efektów.

Równość nachyleń linii regresji badana jest przy pomocy testu F porównującego model zawierający analizowane czynniki z takim samym modelem, ale powiększonym o interakcje z czynnikami wikłającymi. Istotny statystycznie wynik oznacza złamanie założenia równych nachyleń, ponieważ istotna staje się interakcja, a więc różne nachylenia prostych.

JAK USTAWIĆ ANALIZĘ

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Statystyki zaawansowane → Modele wielowymiarowe → ANCOVA.



WYKONANIE I INTERPRETACJA

Przykład 25.9. (plik lekCholesterol.pqs)
Wyobraźmy sobie, że badacz prowadził badanie nad nowym lekiem obniżającym poziom cholesterolu. Badanie było tak zaprojektowane, że dawka leku występowała na trzech poziomach: wysoka, niska i placebo. Badacz sprawdził (przy pomocy ANOVA niezależnej), czy cholesterol po leczeniu różnił się w zależności od dawki leku.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednoczynnikowa ANOVA dla grup niezależnych	
Analizowane zmienne	post
	Dawka
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0.05
Zmienna grupująca	Dawka
ANOVA dla grup niezależnych	
Eta-kwadrat	0.152326
Całkowita suma kwadratów (SS[T])	90.146667
Suma kwadratów między grupami (SS[BG])	13.731667
Suma kwadratów wewnątrz grup (SS[WG])	76.415
Średnia kwadratów między grupami (MS[BG])	6.865833
Średnia kwadratów wewnątrz grup (MS[WG])	2.830185
Stopnie swobody między grupami (df[BG])	2
Stopnie swobody wewnątrz grup (df[WG])	27
Całkowite stopnie swobody (df[T])	29
Statystyka F	2.425931
Wartość p	0.107421

Niestety, badacz nie uzyskał potwierdzenia różnic pomiędzy wynikami. Wyobraźmy sobie, że badacz zdał sobie sprawę, iż to, czy dany lek zmieni poziom cholesterolu, może być związane z wyjściowym poziomem cholesterolu oraz z wiekiem pacjenta. Z tego względu zdecydował się wykonać jednoczynnikową ANCOVA (czynnik to dawka leku), uwzględniającą jako współzmienną poziom cholesterolu przed leczeniem i wiek.

Test efektów międzygrupowych

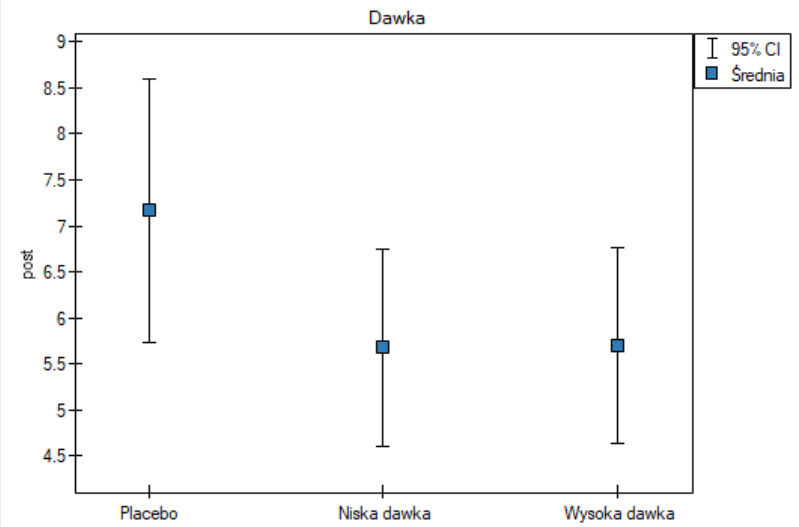
	SS	MS	DF	F	p	Eta-kwadrat
wyraz wolny	0.041927	0.041927	1	0.040259	0.842598	0.001608
Dawka	33.876847	16.938424	2	16.264512	0.00003	0.565437
pre	45.799793	45.799793	1	43.977603	0.000001	0.637564
wiek	1.217654	1.217654	1	1.169209	0.289885	0.044679
within(błąd)	26.035863	1.041435	25			
total	90.146667					

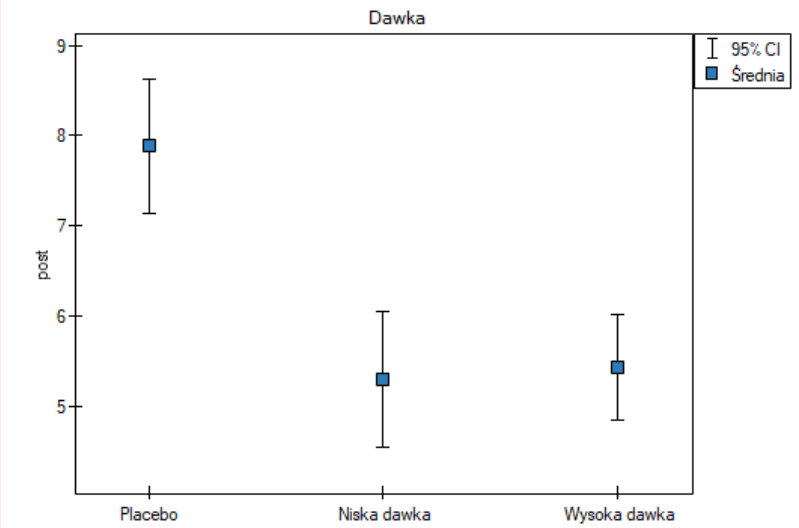
ANCOVA: $p = 0.00003$ (istotne statystycznie, $\alpha = 0.05$)

Po uwzględnieniu poziomu cholesterolu przed leczeniem i wieku, poziom cholesterolu po zastosowaniu różnych dawek leku różni się istotnie – w przeciwieństwie do wyniku zwykłej ANOVA.

Uwzględnienie poziomu cholesterolu przed badaniem zmniejszyło uzyskiwane błędy dla średnich i zawęziło przedziały ufności. By wyświetlić obserwowane lub oczekiwane średnie, wybieramy odpowiednie ustawienia poprzez Opcje czynników, zaznaczając wykres błędów. Pierwszy wykres przedstawia obserwowane średnie wraz z przedziałem ufności, tzn. nieuwzględniające wpływu wieku i poziomu cholesterolu przed leczeniem; drugi wykres to oczekiwane, na podstawie zbudowanego modelu, średnie wraz z przedziałami ufności, tzn. po uwzględnieniu oddziaływania tych dwóch współzmiennych:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





W efekcie, uwzględniając poziom cholesterolu przed leczeniem, badacz był w stanie wykazać skuteczność nowego sposobu leczenia. Poziom cholesterolu przed leczeniem i wiek tłumaczy w pewnym stopniu zmiany w poziomie cholesterolu po leczeniu, jednak pozostałą część zmian w 57% możemy przypisać zastosowanej dawce leku (cząstkowa Eta-kwadrat = 0.565437). Testy post-hoc (wybrane poprzez Opcje czynników) zasugerowały powstanie dwóch grup jednorodnych: grupy placebo i grupy pacjentów z lekiem, wskazując że podnoszenie dawki do wysokiej nie ma znaczenia, gdyż uzyskane poziomy cholesterolu będą podobne.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

POST-HOC Fisher LSD			
	NIR [F]	Statystyka [	Wartość p [F
Dawka			
Placebo;Niska dawka	1.021278	3.008137	0.005921
Placebo;Wysoka daw	0.91139	3.31434	0.002804
Niska dawka;Wysoka	0.944449	-0.054517	0.956957

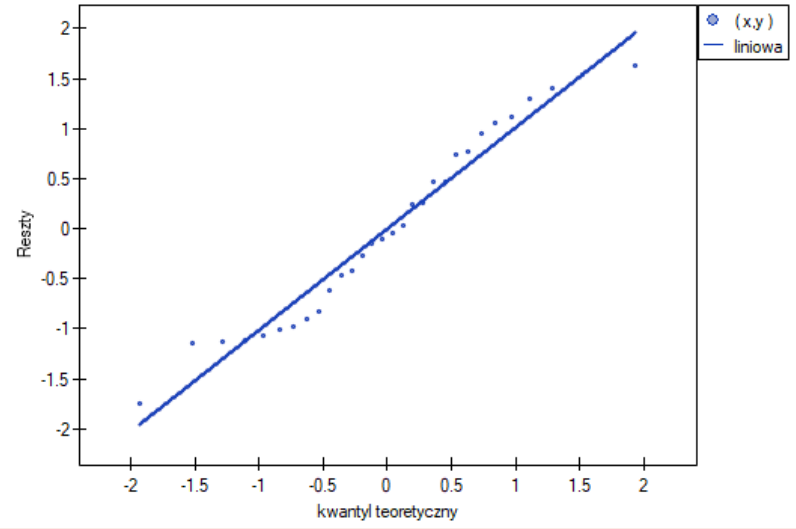
Grupy Jednorodne		
Dawka	A	B
Placebo(b)		*
Niska dawka(a)	*	
Wysoka dawka(a)	*	

Pozostały do sprawdzenia założenia ANCOVA. Jednorodność wariancji i stałość nachyleń prostych regresji potwierdzono przy pomocy testów:

Homogeniczność nacheleń regresji	
DF1	4
DF2	21
p-value	0.80795

Jednorodność wariancji				
	Test Levenea		Test Browna-Forsythea	
	Statystyka F	Wartość p	Statystyka F	Wartość p
Dawka	0.692964	0.508767	0.332599	0.719954

Normalność rozkładu reszt oceniono wizualnie, wyrysowując wykresy Q-Q:



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PRZYKŁAD 25.10. (plik stres.pqs)
Przykład pochodzi z pakietu Datarium R-Cran. Badacze chcą ocenić wpływ nowego leczenia i ćwiczeń na redukcję stresu, po uwzględnieniu różnic w wieku. Wartość pomiaru stresu to interwałowa zmienna wynikowa Y. Ze względu na to, że zmienne „leczenie" i „ćwiczenia" mają odpowiednio 2 i 3 kategorie, przeprowadzimy dwukierunkową ANCOVA w celu określenia, czy interakcja między ćwiczeniami i leczeniem, przy jednoczesnym uwzględnieniu wieku badanych, ma związek ze stresem.

Czynniki	
leczenie X_1	ćwiczenia X_2
tak	niewiele
	przeciętnie
	dużo
nie	niewiele
	przeciętnie
	dużo

W oknie analizy jako zmienną zależną ustawiamy „stres", jako czynniki „leczenie" i „ćwiczenia" oraz dodajemy interakcję tych dwóch zmiennych, współzmienna ciągła to „wiek".

Test efektów międzygrupowych

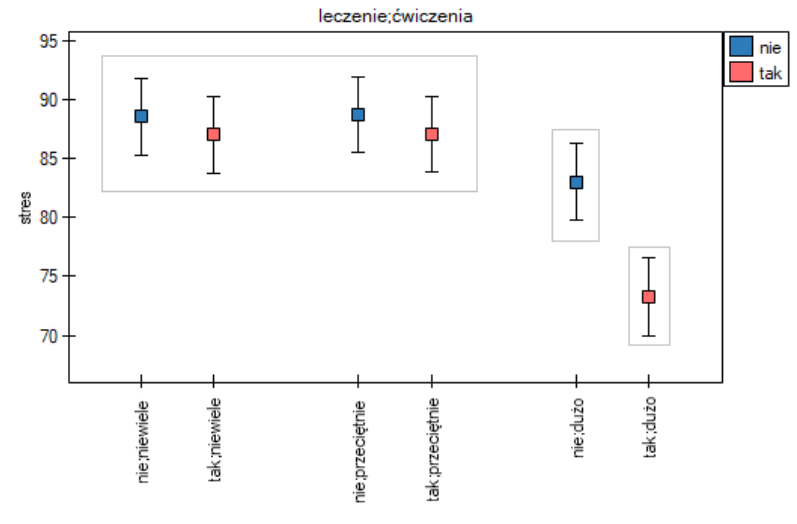
	SS	MS	DF	F	p	Eta-kwadrat
wyraz wolny	731.731365	731.731365	1	29.448395	0.000001	0.357174
leczenie	274.959179	274.959179	1	11.065682	0.001603	0.172724
ćwiczenia	1029.432561	514.716281	2	20.714663	<0.000001	0.438734
leczenie*ćwiczenia	220.937383	110.468692	2	4.445792	0.016409	0.143664
wiek	226.356257	226.356257	1	9.109666	0.003903	0.146671
within(błąd)	1316.939743	24.84792	53			
total	3888.267333					

interakcja leczenie*ćwiczenia: $p = 0.016409$ (istotne statystycznie, $\alpha = 0.05$)

Wpływ leczenia na stres zmienia się w zależności od intensywności wykonywania ćwiczeń.

Wyrysujemy wykres przedstawiający oczekiwane średnie poziomy stresu dla każdej z sześciu podgrup, na jakie interakcja podzieliła nasze dane, oraz wyznaczymy testy post-hoc.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



leczenie;ćwiczenia	A	B	C
nie;niewiele(b)		*	
nie;przeciętnie(b)		*	
nie;dużo(c)			*
tak;niewiele(b)		*	
tak;przeciętnie(b)		*	
tak;dużo(a)	*		

Zgodnie z wynikami testu post-hoc, możemy mówić o trzech różnych grupach jednorodnych: (B) grupa osób o wysokim poziomie stresu, to grupa ćwicząca niewiele lub przeciętnie (bez względu na to, czy są to osoby leczone czy nie); (C) grupa osób o niższym poziomie stresu, to grupa ćwicząca dużo i nieleczona; (A) grupa osób o najniższym poziomie stresu, to grupa ćwicząca dużo i leczona. Wartości poszczególnych średnich wraz z przedziałami ufności przedstawia tabela:

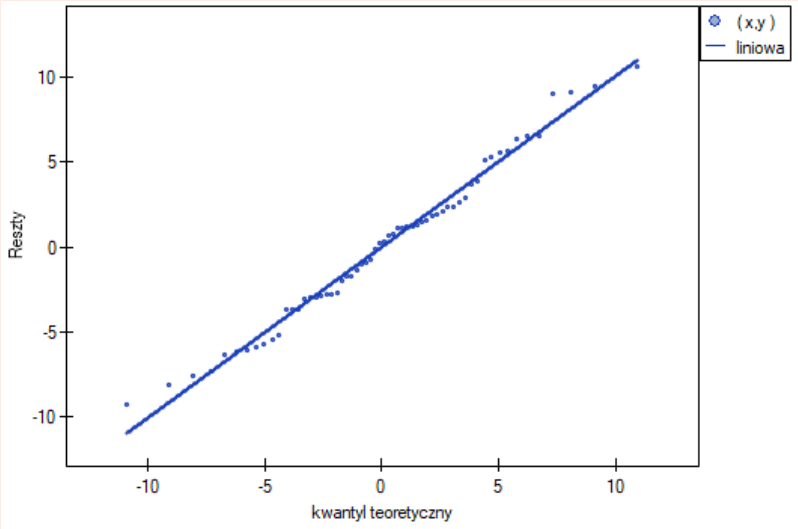
Statystyki opisowe (oczekiwane)					
leczenie	liczność	Średnia	błąd stand.	-95% CI	+95% CI
nie	30	86.736499	0.914163	84.90292	88.570078
tak	30	82.416835	0.914163	80.583256	84.250414
ćwiczenia	liczność	Średnia	błąd stand.	-95% CI	+95% CI
niewiele	20	87.743076	1.161136	85.414132	90.072019
przeciętnie	20	87.838047	1.118399	85.594823	90.081271
dużo	20	78.148878	1.19011	75.761819	80.535936
eczenie*ćwiczenia	liczność	Średnia	błąd stand.	-95% CI	+95% CI
nie;niewiele	10	88.507365	1.616619	85.264838	91.749892
nie;przeciętnie	10	88.679851	1.594777	85.481133	91.878569
nie;dużo	10	83.02228	1.612999	79.787013	86.257546
tak;niewiele	10	86.978786	1.603133	83.763308	90.194263
tak;przeciętnie	10	86.996243	1.577403	83.832372	90.160113
tak;dużo	10	73.275476	1.65137	69.963247	76.587704

Założenia dotyczące równości wariancji, nachyleń linii regresji oraz normalności reszt modelu są spełnione:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Homogeniczność nacheleń regresji	
DF1	5
DF2	48
p-value	0.98578

Jednorodność wariancji				
	Test Levenea		Test Browna-Forsythea	
	Statystyka F	Wartość p	Statystyka F	Wartość p
leczenie	0.334561	0.565225	0.304069	0.583461
ćwiczenia	0.323608	0.724854	0.353205	0.70396
leczenie*ćwiczenia	0.955209	0.453396	0.828725	0.534913



DLA ZAINTERESOWANYCH

Hipotezy – dla pojedynczego czynnika A:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_a, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_i \text{ są sobie równe } (i = 1, 2, ..., a), \end{aligned}$$

gdzie $\mu_1, \mu_2, ..., \mu_a$ – oczekiwane średnie czynnika A dla poszczególnych jego kategorii (skorygowane o współzmienne).

Hipotezy – dla interakcji czynników $A * B$:

$$\begin{aligned} \mathcal{H}_0: &\quad \mu_1 = \mu_2 = ... = \mu_{ab}, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } \mu_k \text{ są sobie równe } (k = 1, 2, ..., ab), \end{aligned}$$

gdzie $\mu_1, \mu_2, ..., \mu_{ab}$ – oczekiwane średnie interakcji czynników $A * B$ dla poszczególnych ich kategorii (skorygowane o współzmienne).

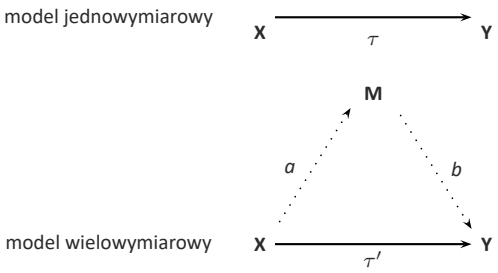
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Efekt mediacji

KIEDY STOSOWAĆ:

Analizę mediacji stosujemy wtedy, gdy chcemy sprawdzić, czy wpływ jednej zmiennej niezależnej (X) na zmienną wynikową (Y) zachodzi choćby częściowo „za pośrednictwem" trzeciej zmiennej, zwanej **mediatorem** (M). Innymi słowy, szukamy odpowiedzi na pytanie, czy związek X z Y można częściowo wyjaśnić tym, że X wpływa na M, a M z kolei wpływa na Y.

Baron i Kenny (1986)[16] zdefiniowali mediatora (M) jako zmienną wyjaśniającą w istotny sposób relację między zmienną niezależną (X) a zmienną wynikową (Y). W mediacji zakłada się, że związek między zmienną niezależną a zmienną zależną jest efektem pośrednim, który istnieje dzięki wpływowi trzeciej zmiennej (mediatora).



Najpierw budujemy **model jednowymiarowy**, w którym mediator w ogóle nie jest brany pod uwagę – opisuje on „surowy" związek X z Y współczynnikiem τ. Następnie budujemy **model wielowymiarowy**, w którym do wyjaśniania Y wykorzystujemy jednocześnie X i M – w tym modelu związek X z Y jest opisany współczynnikiem τ'. Jeśli po dołączeniu mediatora do modelu wpływ X na Y wyraźnie maleje (czyli τ' jest zauważalnie mniejsze niż τ), oznacza to, że część oddziaływania X na Y rzeczywiście przechodzi przez mediator M.

WARUNKI STOSOWANIA

- zmienne X, Y i M powinny być mierzone na skali umożliwiającej budowę modeli regresji liniowej (czyli mieć charakter ilościowy),
- konieczne jest wyznaczenie trzech modeli regresji: związku X z Y (bez mediatora), związku X i M z Y (z mediatorem) oraz związku X z M,
- testy oceniające istotność efektu mediacji (Sobela, Aroiana, Goodmana) są testami konserwatywnymi i wiarygodne wyniki dają dopiero przy dużej liczności próby (powyżej 100 obserwacji).

DEFINICJE MIAR

Porównując model jednowymiarowy (τ) z modelem wielowymiarowym uwzględniającym mediator (τ'), rozróżniamy trzy rodzaje efektów:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Efekt całkowity – to cały wpływ zmiennej X na zmienną Y, wyznaczony w modelu jednowymiarowym (bez uwzględnienia mediatora) i opisany współczynnikiem τ.

Efekt bezpośredni – to ta część wpływu X na Y, która pozostaje, gdy do modelu włączymy mediator M. Jest to współczynnik τ' – opisuje wpływ X na Y niezależny od pośrednictwa M.

Efekt pośredni (mediacyjny) – to ta część wpływu X na Y, która przechodzi „przez" mediator, czyli różnica pomiędzy efektem całkowitym a efektem bezpośrednim ($\tau - \tau'$). Wielkość tę można też wyrazić jako iloczyn współczynników opisujących kolejne odcinki drogi X→M→Y (współczynnika *a*, opisującego wpływ X na M, oraz współczynnika *b*, opisującego wpływ M na Y przy jednoczesnym uwzględnieniu X).

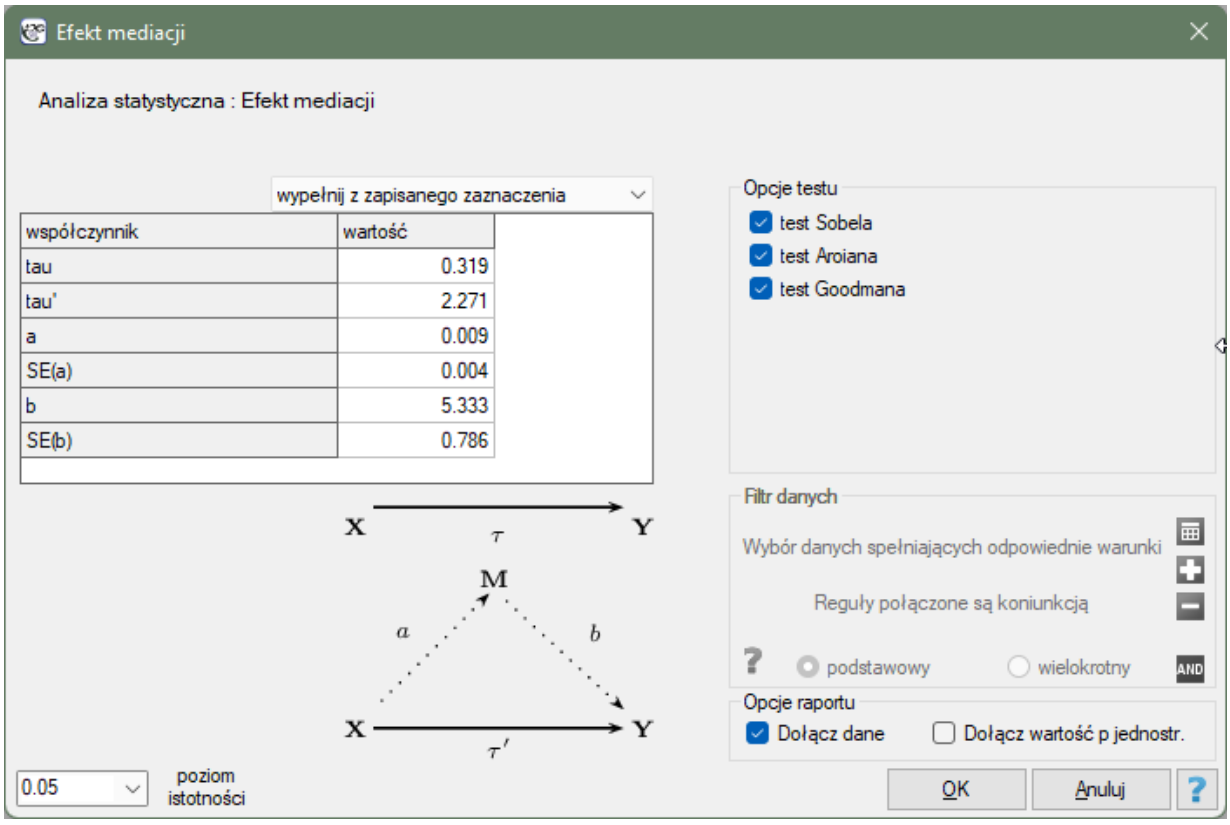
Aby ocenić, jak duża część związku X z Y jest „przejmowana" przez mediator, wyznacza się **procentowy efekt mediacji**, czyli udział efektu pośredniego w efekcie całkowitym. Im wyższa ta wartość, tym większego znaczenia nabiera mediator w wyjaśnianiu badanej zależności.

Do sprawdzenia, czy tak wyznaczony efekt pośredni jest istotny statystycznie (czyli czy zmniejszenie wpływu X na Y po uwzględnieniu mediatora nie jest przypadkowe), służą **testy Sobela (1982)[158], Aroiana (1947)[7] (spopularyzowany przez Barona i Kennyiego[16]) oraz Goodmana (1960)[71]**.

Uwaga! Test Sobela, jak też test Aroiana i Goodmana, są to testy bardzo konserwatywne i przeznaczone jedynie dla dużych prób (przekraczających 100 elementów).

JAK USTAWIĆ ANALIZĘ

Okno analizy efektu mediacji wywołujemy poprzez: Statystyki zaawansowane→Modele wielowymiarowe→Efekt mediacji.



Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

WYKONANIE I INTERPRETACJA

Przykład 25.11. Na podstawie pracy Mimar Sinan Fine (2017) [125].
Badanie obejmuje 300 osób dorosłych mieszkających w Stambule. Zmienna zależna Y to ciśnienie skurczowe, a zmienna niezależna X to wiek. Zmienna pośrednicząca M to częstotliwość spożywania alkoholu. Celem pracy jest zbadanie zależności między wiekiem a skurczowym ciśnieniem krwi oraz przedstawienie wpływu częstości spożywania alkoholu na tę zależność.

Zbudowano model jednowymiarowy, w którym nie uwzględniono potencjalnego mediatora:
$Y = 0.319 \cdot X + c$.
Wielkość wpływu zmiennej X (wiek) na zmienną Y (ciśnienie skurczowe) wyniosła **tau**=0.319.

Zbudowano model wielowymiarowy, w którym uwzględniono potencjalny mediator:
$Y = 2.271 \cdot X + 5.333 \cdot M + c$.
Wielkość wpływu zmiennej X (wiek) na zmienną Y (ciśnienie skurczowe) wyniosła **tau'**=2.271. Wiemy również z tego modelu, że **b**=5.333, a błąd SE_b=0.786.

Różnica między współczynnikami to **tau-tau'**=a*b=0.048. Efekt mediacji wynosi **(tau-tau')/tau**=(0.319-0.271)/0.371=0.15047, co oznacza, że M (częstość spożycia alkoholu) modyfikuje badany związek, zmniejszając współczynnik o około 15%.

Zbudowano jeszcze model jednowymiarowy badający wpływ zmiennej X na mediator:
$M = 0.009 \cdot X + c$.
Wiemy z tego modelu, że współczynnik **a**=0.009, a błąd SE_a=0.004. Wszystkie te informacje wprowadzamy w oknie analizy, uzyskując następujący raport:

Efekt mediacji	
tau-tau'	-1.952
Efekt mediacji	-6.119122
Test Sobela	
Statystyka Z	2.135636
Wartość p dwustronne	0.032709
Test Aroiana	
Statystyka Z	2.115041
Wartość p dwustronne	0.034426
Test Goodmana	
Statystyka Z	2.156845
Wartość p dwustronne	0.031018

Dane :	
tau [Y=tau*X+c	0.319
tau' [Y=tau'*X+	2.271
a [M=a*X+c]	0.009
SE(a)	0.004
b [Y=tau'*X+b*	5.333
SE(b)	0.786

Na podstawie współczynników a i b oraz ich błędów standardowych wyznaczony zostaje wynik testów Sobela, Aroiana i Goodmana.

Sobel $p = 0.0327$, Aroian $p = 0.0344$, Goodman $p = 0.0310$ (istotne statystycznie, $\alpha = 0.05$)

Częstość spożywania alkoholu istotnie pośredniczy w związku wieku z ciśnieniem skurczowym.

Uzyskane wartości p wskazują na istotne statystycznie znaczenie mediatora. Potwierdziliśmy więc, że częstość spożywania alkoholu wpływa na związek wieku z ciśnieniem skurczowym na tyle zauważalnie, że warto wyjaśniać, dlaczego ten efekt występuje.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Wzory:
Wielkość zmian określamy poprzez różnicę współczynników opisujących związek zmiennej X ze zmienną Y w modelu jednowymiarowym:
$Y = \tau \cdot X + c$
i w modelu wielowymiarowym, czyli uwzględniającym zmienną M:
$Y = \tau \cdot X + b \cdot M + c$.

Różnica:

$$\tau - \tau' = a \cdot b$$

Efekt mediacji:

$$\frac{\tau - \tau'}{\tau} \cdot 100\%$$

W rezultacie, gdy mediator (M) jest włączony do modelu regresji określającego związek zmiennej X i Y, wpływ zmiennej niezależnej τ jest zmniejszony do τ'.

Testy oceniające efekt mediacji
Interpretacja:
Test Sobela (1982)[158], test Aroiana (1947)[7] spopularyzowany przez Barona i Kennyiego [16] oraz test Goodmana (1960)[71] są testami, które określają, czy zmniejszenie wpływu zmiennej niezależnej na zmienną wynikową, po uwzględnieniu mediatora w modelu, jest znaczną redukcją, a zatem czy efekt mediacji jest istotny statystycznie.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \tau = \tau' \\ \mathcal{H}_1: &\quad \tau \neq \tau', \end{aligned}$$

Wzory:
Statystyka testowa dla testu Sobela ma postać:

$$Z = \frac{a \cdot b}{\sqrt{b^2 \cdot SE_a^2 + a^2 \cdot SE_b^2}}$$

Statystyka testowa dla testu Aroiana ma postać:

$$Z = \frac{a \cdot b}{\sqrt{b^2 \cdot SE_a^2 + a^2 \cdot SE_b^2 + SE_a^2 \cdot SE_b^2}}$$

Statystyka testowa dla testu Goodmana ma postać:

$$Z = \frac{a \cdot b}{\sqrt{b^2 \cdot SE_a^2 + a^2 \cdot SE_b^2 - SE_a^2 \cdot SE_b^2}}$$

Statystyki te mają asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α:

$$\begin{aligned} &\text{jeżeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ &\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

26 MODELE MIESZANE

Wprowadzenie: dane zależne, efekty stałe i efekty losowe

PO CO MODELE MIESZANE

Klasyczne regresje dostępne w programie – regresja wieloraka, regresja logistyczna i regresja Coxa – zakładają, że wszystkie obserwacje są od siebie niezależne. W wielu badaniach to założenie nie jest spełnione, bo obserwacje naturalnie układają się w grupy:

- **pomiary powtarzane** – ten sam pacjent badany kilkakrotnie (przed leczeniem, po tygodniu, po miesiącu), ta sama osoba testowana w kolejnych dniach;
- **dane klastrowe (skupione)** – pacjenci leczeni w tym samym ośrodku, uczniowie z tej samej klasy, zwierzęta z jednego miotu lub stada;
- **dane hierarchiczne** – pomiary pacjenta, pacjenci w ośrodku, ośrodki w regionie (kilka poziomów zagnieżdżenia);
- **zdarzenia nawracające i obiekty sparowane** – kolejne infekcje tego samego chorego, dwoje oczu tej samej osoby.

Obserwacje z tej samej grupy są do siebie bardziej podobne niż obserwacje z różnych grup: pacjent o wysokim ciśnieniu ma zwykle wysokie ciśnienie na każdej wizycie, a ośrodek o gorszych wynikach ma gorsze wyniki u większości swoich pacjentów. Jeśli tę zależność zignorujemy i policzymy zwykłą regresję, program potraktuje np. 10 pomiarów jednego pacjenta jak 10 niezależnych osób. Skutki są dwojakie:

- **błędy standardowe są zaniżone**, a wartości p zbyt małe – zbyt łatwo uznajemy efekt za istotny (tzw. pseudoreplikacja);
- nie wiemy, **jaka część zmienności** wynika z różnic między grupami (osobami, ośrodkami), a jaka ze zmienności wewnątrz grup, choć bywa to najciekawsza informacja w badaniu.

Model mieszany rozwiązuje oba problemy: uwzględnia strukturę grupową w obliczaniu błędów i jawnie szacuje, jak bardzo grupy różnią się między sobą.

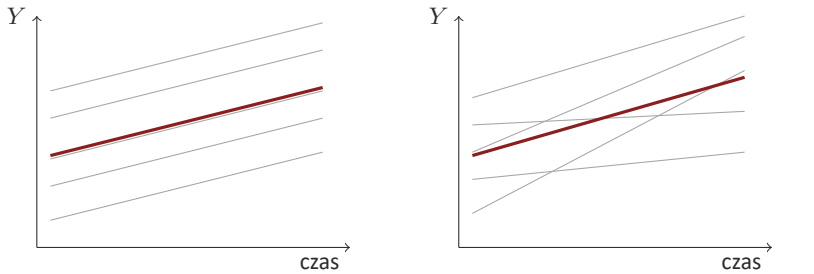
EFEKTY STAŁE I EFEKTY LOSOWE

Nazwa „model mieszany” pochodzi stąd, że model zawiera dwa rodzaje efektów:

- **efekty stałe** (ang. *fixed effects*) – zwykłe współczynniki regresji $\beta_0, \beta_1, \ldots$ przy zmiennych niezależnych. Opisują one populację jako całość i interpretuje się je tak samo jak w klasycznej regresji: o ile zmienia się Y, gdy X rośnie o jednostkę;
- **efekty losowe** (ang. *random effects*) – odchylenia poszczególnych grup od tego, co przewidują efekty stałe. Nie interesują nas pojedyncze wartości tych odchyleń dla każdej grupy, lecz ich **wariancja**, czyli miara tego, jak bardzo grupy się różnią. Zakładamy, że grupy w badaniu są losową próbą z większej populacji grup (pacjentów, ośrodków), a ich efekty mają rozkład normalny o średniej 0.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Najprostszy efekt losowy to **losowy wyraz wolny**: każda grupa ma własny poziom Y, przesunięty względem poziomu populacyjnego. Jeśli dodatkowo efekt jakiejś zmiennej (najczęściej czasu) różni się między grupami, wprowadzamy **losowe nachylenie** tej zmiennej. Poniższy rysunek pokazuje różnicę: po lewej grupy mają różne poziomy, ale ten sam trend; po prawej różnią się także tempem zmian.



losowy wyraz wolny | losowy wyraz wolny i losowe nachylenie

Gruba linia – efekty stałe (populacja); cienkie linie – poszczególne grupy.

TRZY MODELE MIESZANE W PROGRAMIE

Każda z klasycznych regresji ma w programie swój odpowiednik mieszany. Wybór zależy wyłącznie od typu zmiennej zależnej:

Zmienna zależna	Model klasyczny	Model mieszany	Efekty losowe
ciągła	regresja wieloraka	regresja wieloraka mieszana	wyraz wolny, nachylenia
dwuwartościowa (0/1)	regresja logistyczna	regresja logistyczna mieszana	wyraz wolny, nachylenia
czas i status	regresja PH Coxa	regresja Coxa z efektem losowym	frailty grupy

Regresja wieloraka mieszana zastępuje ANOVA dla grup zależnych, gdy pomiary są niekompletne, nierówno rozłożone w czasie lub gdy w modelu mają się znaleźć zmienne ciągłe. Regresja logistyczna mieszana pełni tę samą rolę wobec ANOVA Q-Cochrana. Regresja Coxa z efektem losowym jest odpowiednikiem regresji Coxa dla pacjentów z wielu ośrodków lub dla zdarzeń nawracających.

PRZYGOTOWANIE DANYCH: UKŁAD DŁUGI

Wszystkie trzy analizy wymagają danych w **układzie długim**: każdy wiersz arkusza to jedna obserwacja (jeden pomiar), a osobna kolumna – **zmienna grupująca** (ID) – mówi, do której grupy obserwacja należy. Zmienna grupująca może być liczbą, tekstem lub datą; program traktuje ją wyłącznie jako etykietę. Grupy mogą mieć różną liczbę obserwacji, a braki danych usuwane są pojedynczymi wierszami, nie całymi grupami.

układ szeroki (ANOVA dla grup zależnych)

Pacjent	Dzień 0	Dzień 1	Dzień 2
308	249.6	258.7	250.8
309	222.7	205.3	203.0
310	199.1	194.3	234.3

układ długi (modele mieszane)

Pacjent	Dzień	Reakcja
308	0	249.6
308	1	258.7
308	2	250.8
309	0	222.7
309	1	205.3
...	...	...

Dane zapisane w układzie szerokim (kolejne pomiary w kolejnych kolumnach, jak w ANOVA dla grup zależnych) trzeba przed analizą ułożyć w układ długi: kolumny z pomiarami kopiujemy jedną pod drugą do wspólnej kolumny wartości, obok wpisujemy identyfikator obiektu (powtórzony tyle razy, ile jest

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

pomiarów) oraz numer lub czas pomiaru. Przekształcenie to wykonuje automatycznie polecenie Dane→Przekształć: szeroka → długa... (zob. przekształcanie danych między postacią szeroką a długą); polecenie odwrotne Przekształć: długa → szeroka... pozwala wrócić do układu szerokiego. Polecenia Przestaw na dane surowe i Przestaw na tabelę służą do czego innego – zamieniają tabelę kontyngencji (liczności) na dane surowe i odwrotnie.

Uwaga! Zmienna niezależna, której chcemy nadać losowe nachylenie, musi zmieniać się wewnątrz grup (np. czas pomiaru, dawka podawana w kolejnych wizytach). Zmienna stała w obrębie grupy (np. płeć pacjenta, gdy grupą jest pacjent) może być tylko efektem stałym.

CO CZYTAĆ W RAPORCIE I W JAKIEJ KOLEJNOŚCI

Raporty trzech modeli mają wspólny układ. Zalecana kolejność czytania:

1. **Zbieżność i dopasowanie osobliwe.** Jeśli algorytm nie osiągnął zbieżności albo raport zgłasza dopasowanie osobliwe (wariancja efektu losowego równa praktycznie zeru lub korelacja efektów losowych równa ± 1), pozostałe liczby są niewiarygodne. Model należy uprościć: usunąć losowe nachylenie, wybrać kowariancję niezależne albo zmniejszyć liczbę zmiennych.

2. **ICC z modelu pustego.** Współczynnik korelacji wewnątrzklasowej mówi, jaka część całkowitej zmienności Y leży między grupami. ICC bliskie 0 oznacza, że obserwacje w grupie są prawie niezależne; ICC równe 0.4 oznacza, że 40% zmienności to różnice między grupami, a dwie obserwacje z tej samej grupy są skorelowane na poziomie 0.4. Nawet małe ICC nie jest jednak powodem, by porzucić model mieszany, jeśli układ badania jest hierarchiczny – model mieszany pozostaje wtedy poprawną metodą, a zwykła regresja nie.

3. **Test efektów losowych.** Test ilorazu wiarygodności (LR) porównuje model mieszany z modelem bez efektów losowych. Hipoteza zerowa (wariancja równa 0) leży na granicy dopuszczalnych wartości, dlatego program stosuje korektę wartości p (mieszankę rozkładów χ^2, w najprostszym przypadku dzielenie p przez 2) i informuje o tym w raporcie.

4. **Efekty losowe.** Odchylenie standardowe losowego wyrazu wolnego jest wyrażone w jednostkach Y (a w regresji logistycznej i Coxa – na skali logitowej lub logarytmu hazardu) i mówi, jak typowa grupa odbiega od średniej. Odchylenie standardowe losowego nachylenia mówi, jak bardzo grupy różnią się tempem zmian.

5. **Efekty stałe.** Czytane jak w klasycznej regresji, z jedną różnicą: w regresji logistycznej i Coxa iloraz szans i hazard ratio są **warunkowe** – opisują efekt zmiennej „w obrębie tej samej grupy" (dla tego samego pacjenta, tego samego ośrodka). Są one co do wartości bezwzględnej większe niż populacyjne OR i HR ze zwykłej regresji, która uśrednia po grupach, i nie należy ich porównywać wprost.

6. **Efekty losowe grup (BLUP).** Przewidywane odchylenia poszczególnych grup. Służą do wskazania grup nietypowych (pacjenta o odstającej trajektorii, ośrodka o wyższym ryzyku), a nie do testowania istotności. Wykres typu „caterpillar" (grupy uporządkowane według efektu, z przedziałami) pokazuje je najczytelniej.

ILE GRUP POTRZEBA

Wariancję efektu losowego szacuje się na podstawie liczby grup, nie liczby obserwacji. Program wymaga co najmniej 3 grup, ale przy mniej niż 10 grupach ostrzega, że oszacowanie wariancji jest niepewne. W

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

praktyce wiarygodne oszacowania zaczynają się od kilkunastu–kilkudziesięciu grup, a każde losowe nachylenie zwiększa te wymagania. Przy 3–5 ośrodkach rozsądniej jest potraktować ośrodek jako zmienną fikcyjną (efekt stały) niż jako efekt losowy.

OGRANICZENIA

Modele mieszane nie wchodzą do Zestawień ani do Kreatora. Dostępne są dwa poziomy zagnieżdżenia (np. pacjent w ośrodku) i struktury korelacji reszt AR(1) oraz wymienność w regresji wielorakiej mieszanej; w regresji Coxa z efektem losowym dostępny jest wyłącznie losowy wyraz wolny (frailty gamma).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

26.1 REGRESJA WIELORAKA MIESZANA

Regresja wieloraka mieszana (liniowy model mieszany)

KIEDY STOSOWAĆ:

Regresja wieloraka mieszana (liniowy model mieszany, ang. *linear mixed model*, LMM) jest rozszerzeniem liniowej regresji wielorakiej na dane, w których obserwacje nie są niezależne, lecz tworzą grupy: wielokrotne pomiary tego samego pacjenta, pacjenci leczeni w tym samym ośrodku, uczniowie z jednej klasy, zwierzęta z jednego miotu (zob. wprowadzenie do modeli mieszanych). Obserwacje z tej samej grupy są do siebie bardziej podobne niż obserwacje z różnych grup. Zwykła regresja tę zależność ignoruje, przez co zaniża błędy standardowe współczynników i zawyża istotność wyników. Model mieszany rozbija zmienność zmiennej zależnej na część między grupami i część wewnątrz grup, a każdy współczynnik otrzymuje właściwy błąd.

Model zawiera dwa rodzaje efektów:

- **efekty stałe** – zwykłe współczynniki regresji $\beta_0, \beta_1, \ldots, \beta_k$ opisujące populację,
- **efekty losowe** – odchylenia poszczególnych grup od populacji: losowy wyraz wolny (każda grupa ma własny poziom Y) i ewentualnie losowe nachylenia (efekt wybranej zmiennej X różni się między grupami).

Model z losowym wyrazem wolnym i losowym nachyleniem zmiennej X_1 ma postać:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \ldots + \beta_k X_{kij} + u_{0j} + u_{1j} X_{1ij} + \epsilon_{ij},$$

gdzie:
Y_{ij} – i-ta obserwacja w j-tej grupie,
$\beta_0, \beta_1, \ldots, \beta_k$ – efekty stałe,
u_{0j} – losowy wyraz wolny grupy j, u_{1j} – losowe nachylenie grupy j; zakłada się, że (u_{0j}, u_{1j}) mają rozkład normalny o średniej 0 i macierzy kowariancji Σ (wariancje τ_0^2, τ_1^2 i korelacja ρ_{01}),
ϵ_{ij} – reszta, o rozkładzie normalnym ze średnią 0 i wariancją σ^2.

Dla grupy j model opisuje więc własną prostą: jej wyraz wolny to $\beta_0 + u_{0j}$, a nachylenie $\beta_1 + u_{1j}$. Efekty stałe opisują prostą „przeciętnej" grupy, a wariancje efektów losowych – jak bardzo proste poszczególnych grup są rozrzucone wokół niej.

Model ten zastępuje ANOVA dla grup zależnych, gdy pomiary powtarzane są niekompletne lub nierówno rozmieszczone w czasie, gdy liczba pomiarów różni się między osobami, albo gdy oprócz czynnika powtarzanego w modelu mają się znaleźć zmienne ciągłe (np. wiek, dawka) lub zmienne opisujące grupę (np. płeć, ośrodek).

WARUNKI STOSOWANIA

- pomiar zmiennej zależnej Y na skali interwałowej,
- liniowy związek pomiędzy zmienną zależną a zmiennymi niezależnymi,

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- efekty losowe i reszty mają rozkład normalny; reszty mają stałą wariancję,
- grupy są od siebie niezależne – zależność występuje tylko wewnątrz grup,
- wystarczająca liczba grup – co najmniej 3, a do wiarygodnego oszacowania wariancji efektów losowych zwykle kilkanaście lub więcej,
- braki danych losowe (model nie wymaga kompletnych ani zbalansowanych pomiarów).

Uwaga! Dane muszą być przygotowane w układzie długim: każdy wiersz arkusza to jedna obserwacja, a osobna kolumna (zmienna grupująca, np. identyfikator pacjenta) wskazuje, do której grupy obserwacja należy. Dane w układzie szerokim (pomiary w kolejnych kolumnach) należy najpierw ułożyć w układ długi (zob. wprowadzenie).

DEFINICJE MIAR

Wariancje efektów losowych – τ_0^2 (losowy wyraz wolny) mówi, jak bardzo grupy różnią się poziomem Y; τ_1^2 (losowe nachylenie) – jak bardzo różnią się efektem zmiennej X_1. Wygodniej czyta się ich pierwiastki, czyli odchylenia standardowe, bo są wyrażone w jednostkach Y (lub Y na jednostkę X_1): typowa grupa odbiega od populacji o mniej więcej jedno odchylenie standardowe, a około 95% grup mieści się w przedziale ± 2 odchylenia.

Korelacja efektów losowych ρ_{01} – związek między poziomem grupy a jej nachyleniem. Korelacja ujemna oznacza, że grupy startujące wyżej zmieniają się wolniej. Uwaga: jej wartość zależy od tego, gdzie leży zero zmiennej X_1 – wyraz wolny to poziom przy $X_1 = 0$. Gdy X_1 to np. wiek 8–14 lat, warto go wycentrować (odjąć wartość środkową), aby korelacja opisywała poziom w środku zakresu, a nie hipotetyczny poziom w wieku 0.

Współczynnik korelacji wewnątrzklasowej (ICC) – liczony z **modelu pustego** (bez zmiennych niezależnych, tylko losowy wyraz wolny):

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2},$$

gdzie τ^2 to wariancja między grupami, a σ^2 wariancja wewnątrz grup (reszt). ICC to udział zmienności między grupami w zmienności całkowitej, a zarazem korelacja dwóch obserwacji z tej samej grupy. Przedział ufności ICC jest orientacyjny (metoda delta).

R^2 **brzegowe i warunkowe** (Nakagawa i Schielzeth 2013[196]) – R^2 brzegowe to udział wariancji wyjaśnianej przez same efekty stałe, R^2 warunkowe – przez efekty stałe i losowe razem. Różnica między nimi mówi, ile zmienności „tłumaczy” przynależność do grupy.

Test LR efektów losowych – test ilorazu wiarygodności porównujący model z danym efektem losowym i bez niego. Hipoteza zerowa (wariancja równa 0) leży na granicy przestrzeni parametrów, więc wartość p z rozkładu χ^2 jest korygowana (mieszanka rozkładów; dla jednego parametru p dzieli się przez 2). Program zawsze liczy ten test metodą ML, niezależnie od metody estymacji wybranej dla modelu głównego.

Efekty losowe grup (BLUP) – przewidywane wartości u_{0j}, u_{1j} dla każdej grupy (ang. *best linear unbiased predictor*). Są to oszacowania „ściągnięte” w stronę zera: grupa z małą liczbą obserwacji lub o dużym rozrzucie dostaje efekt bliższy 0, niż wynikałoby ze zwykłej średniej jej obserwacji. Służą do wskazania grup nietypowych i do predykcji dla konkretnej grupy, nie do testowania istotności.

Dopasowanie osobliwe (ang. *singular fit*) – sytuacja, gdy oszacowana wariancja któregoś efektu losowego jest praktycznie równa 0 albo korelacja efektów losowych wynosi ± 1. Model jest wtedy zbyt

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

złożony w stosunku do danych i należy go uprościć (usunąć losowe nachylenie lub wybrać kowariancję niezależne).

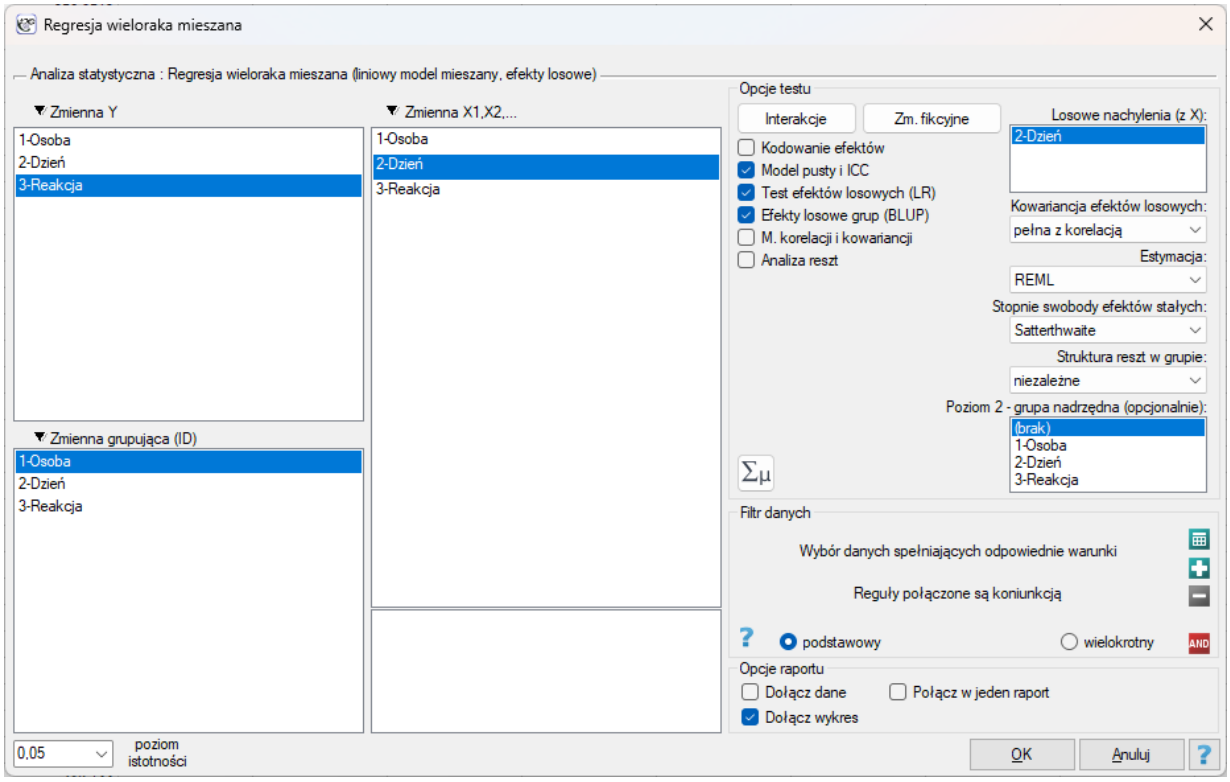
Stopnie swobody Satterthwaite'a – w modelu mieszanym nie ma jednej prostej liczby stopni swobody dla testu efektu stałego. Zmienna opisująca grupę (np. płeć pacjenta) ma w istocie tyle niezależnych obserwacji, ile jest grup, a zmienna zmieniająca się wewnątrz grup – więcej. Metoda Satterthwaite'a wyznacza dla każdego współczynnika osobne, zwykle niecałkowite stopnie swobody, dzięki czemu test t jest wiarygodny także przy małej liczbie grup.

Struktura reszt w grupie – domyślnie reszty ϵ_{ij} są niezależne, a cała korelacja wewnątrz grupy wynika z efektów losowych. Dla pomiarów w czasie można dodatkowo założyć, że reszty sąsiednich pomiarów są skorelowane: AR(1) (korelacja ρ maleje wykładniczo z odległością w kolejności wierszy grupy: $\rho, \rho^2, \rho^3, \ldots$) lub wymienność (CS) (jednakowa korelacja wszystkich par reszt w grupie).

Poziom 2 – drugi, nadrzędny poziom grupowania, np. pacjent (poziom 1) zagnieżdżony w ośrodku (poziom 2). Model dostaje wtedy dwa losowe wyrazy wolne, a raport – wariancję i ICC osobno dla każdego poziomu.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja wieloraka mieszana.



Okno analizy ma układ znany z regresji wielorakiej. W jego lewej części wskazujemy zmienne:

- Zmienna zależna Y – zmienna ciągła, która ma być wyjaśniana,
- Zmienna grupująca (ID) – kolumna identyfikująca grupę (pacjenta, ośrodek, miot); może być liczbowa lub tekstowa,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– Zmienne niezależne X1, X2, ... – predyktory; zmienne fikcyjne i interakcje przygotowuje się przyciskami Zm. fikcyjne i Interakcje tak jak w regresji wielorakiej (rozdział Przygotowanie zmiennych do analizy w modelach wielowymiarowych).

Prawa część okna zawiera opcje modelu mieszanego. Poniżej opisano, co zmienia każda z nich.

Efekty losowe

– Losowe nachylenia (z X) – lista wypełnia się zmiennymi wybranymi jako X. Zaznaczenie zmiennej oznacza, że jej efekt ma się różnić między grupami (np. tempo zmian w czasie różne dla każdego pacjenta). Losowy wyraz wolny jest zawsze w modelu i nie trzeba go wybierać. Zmienna z losowym nachyleniem musi zmieniać się wewnątrz grup; zmienna stała w grupie (np. płeć pacjenta) nie może mieć losowego nachylenia – program zgłosi wtedy błąd. Każde losowe nachylenie dodaje do modelu wariancję i (przy kowariancji pełnej) korelacje, więc wymaga większej liczby grup; rozsądnie jest zaczynać od jednego nachylenia, najczęściej dla czasu.

– Kowariancja efektów losowych – pełna z korelacją szacuje także korelację między losowym wyrazem wolnym a nachyleniem (oraz między nachyleniami); niezależne przyjmuje te korelacje równe 0. Wariant niezależne ma mniej parametrów, jest stabilniejszy przy małej liczbie grup i zalecany, gdy korelacja w modelu pełnym wychodzi bliska 0 albo ± 1 (dopasowanie osobliwe). Przy braku losowych nachyleń opcja nie ma znaczenia.

Estymacja

– Estymacja – REML (ograniczona największa wiarygodność, domyślnie) daje nieobciążone oszacowania wariancji i jest zalecana dla modelu końcowego; ML (największa wiarygodność) zaniża nieco wariancje, ale tylko wiarygodność ML pozwala porównywać kryteriami informacyjnymi modele różniące się efektami stałymi. Wybór metody nie zmienia w zauważalny sposób współczynników efektów stałych, zmienia natomiast wariancje efektów losowych, ich błędy i wartości $-2LL$, AIC, BIC. Zasada: dobór zmiennych – ML, raportowanie modelu – REML.

– Stopnie swobody efektów stałych – Satterthwaite (domyślnie) wyznacza dla każdego współczynnika własne stopnie swobody i test t; zalecane zawsze, a szczególnie przy małej liczbie grup. Test z Walda traktuje statystykę jak w dużej próbie (rozkład normalny, bez stopni swobody); daje mniejsze wartości p i jest odpowiedni tylko przy dużej liczbie grup.

– Struktura reszt w grupie – niezależne (domyślnie), AR(1) (kolejność wierszy) lub wymienność (CS). AR(1) wymaga, by wiersze każdej grupy były w arkuszu uporządkowane w czasie i mniej więcej równoodległe. W raporcie pojawia się parametr ρ. Wymienność razem z losowym wyrazem wolnym jest nieidentyfikowalna (obie opisują tę samą stałą korelację w grupie), więc ma sens głównie jako porównanie z modelem AR(1).

– Poziom 2 – grupa nadrzędna (opcjonalnie) – kolumna wskazująca grupę wyższego poziomu, w której zagnieżdżona jest zmienna grupująca (np. ośrodek dla pacjentów). Dodaje losowy wyraz wolny poziomu 2, a w raporcie osobną wariancję, ICC i tabelę BLUP tego poziomu.

Elementy raportu

– Model pusty i ICC – dodatkowo liczony jest model bez zmiennych niezależnych, z którego wyznacza się ICC z przedziałem ufności; warto zostawić włączone, bo ICC to pierwsza liczba, na którą należy spojrzeć.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- Test efektów losowych (LR) – testy ilorazu wiarygodności dla każdego składnika losowego osobno oraz dla wszystkich łącznie (model mieszany wobec zwykłej regresji wielorakiej). Wydłuża obliczenia, bo wymaga dopasowania dodatkowych modeli.
- Efekty losowe grup (BLUP) – dołącza tabelę przewidywanych efektów każdej grupy z błędami; przy setkach grup tabela jest długa, a te same wartości pokazuje wykres caterpillar.
- M. korelacji i kowariancji – macierze korelacji i kowariancji ocen efektów stałych.
- Analiza reszt – wartości przewidywane (warunkowe i populacyjne), reszty i reszty standaryzowane dla każdej obserwacji.
- Dołącz wykresy – wykres caterpillar efektów losowych grup oraz (przy losowym nachyleniu) trajektorie grup na tle prostej populacyjnej.

Pozostałe elementy okna (poziom istotności, filtr danych, łączenie raportów) działają tak jak w innych analizach.

WYNIKI I INTERPRETACJA

Liczności – liczba obserwacji, liczba grup oraz minimalna, medianowa i maksymalna liczba obserwacji na grupę. Przy mniej niż dziesięciu grupach raport ostrzega, że oszacowania wariancji efektów losowych są niepewne.

Estymacja – metoda (REML/ML), liczba iteracji, informacja o zbieżności oraz o **dopasowaniu osobliwym**. Brak zbieżności lub dopasowanie osobliwe oznacza, że model należy uprościć, zanim zaczniemy interpretować pozostałe wyniki.

Dopasowanie modelu – logarytm wiarygodności, $-2LL$, kryteria informacyjne AIC, AICc i BIC (im mniejsze, tym lepszy model) oraz R^2 brzegowe i warunkowe. Kryteria informacyjne wolno porównywać tylko między modelami liczonymi tą samą metodą; modele różniące się efektami stałymi – tylko przy ML.

Test LR efektów losowych (vs regresja wieloraka, ML) – czy uwzględnienie struktury grupowej jest w ogóle potrzebne. Istotny wynik oznacza, że zwykła regresja wieloraka nie jest dla tych danych poprawna. Wynik nieistotny przy hierarchicznym układzie badania nie zwalnia jednak z użycia modelu mieszanego.

Model pusty – wariancja między grupami τ^2, wariancja reszt σ^2 i **ICC** z przedziałem ufności.

Efekty stałe (warunkowe) – współczynnik b, jego błąd, przedział ufności, statystyka t ze stopniami swobody Satterthwaite'a (mogą być niecałkowite) lub statystyka z oraz wartość p. Interpretacja jak w regresji wielorakiej: b to zmiana Y na jednostkę X przy pozostałych zmiennych ustalonych, w obrębie tej samej grupy. Stopnie swobody podpowiadają, na jakim poziomie „działa" zmienna: liczba bliska liczbie grup oznacza zmienną opisującą grupy, liczba bliska liczbie obserwacji – zmienną zmieniającą się wewnątrz grup.

Efekty losowe – dla każdego składnika wariancja i odchylenie standardowe. Kolumny LR χ^2, df i p zawierają test, czy dany składnik jest potrzebny (model z nim wobec modelu bez niego, ML, z korektą granicy). Ostatni wiersz to wariancja reszt. Przy strukturze AR(1) lub CS raport podaje parametr ρ, a przy poziomie 2 – wariancję i ICC tego poziomu.

Korelacje efektów losowych – przy więcej niż jednym składniku macierz korelacji między losowym wyrazem wolnym a nachyleniami. Wartość bliska ± 1 sygnalizuje dopasowanie osobliwe.

Efekty losowe grup (BLUP) – przewidywane odchylenie każdej grupy od populacji wraz z błędem (dla wyrazu wolnego i każdego losowego nachylenia). Suma efektu stałego i BLUP daje własny współczyn-

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

nik grupy. Na wykresie caterpillar grupy, których przedział nie obejmuje zera, wyraźnie odbiegają od przeciętnej.

Analiza reszt – wartości przewidywane warunkowe (z efektami losowymi grupy, czyli „dla tego pacjenta") i populacyjne (same efekty stałe, czyli „dla przeciętnego pacjenta"), reszty oraz reszty standaryzowane; obserwacje o $|reszta| \geq 3$ są wyróżnione jako odstające.

Wykresy – Efekty losowe grup (BLUP $\pm$ błąd): grupy uporządkowane według wartości efektu, z przedziałem błędu i linią zera; osobny panel dla wyrazu wolnego i każdego nachylenia. Trajektorie grup: dla modelu z losowym nachyleniem cienkie proste dopasowane dla poszczególnych grup na tle grubej prostej populacyjnej – najbardziej obrazowe podsumowanie pomiarów powtarzanych.

WYKONANIE I INTERPRETACJA

Przykład 26.1. (plik deprywacjaSnu.pqs)
W badaniu wpływu niedoboru snu na sprawność psychomotoryczną (Belenky i in. 2003[190]) 18 osób przez 10 kolejnych dni mogło spać tylko 3 godziny na dobę. Każdego dnia mierzono im średni czas reakcji (w milisekundach) w teście czujności. Dane są w układzie długim: kolumna Osoba identyfikuje badanego, Dzień to numer dnia deprywacji (0–9), a Reakcja – czas reakcji. Łącznie 180 obserwacji, po 10 dla każdej osoby.

	1	2	3
	Osoba	Dzień	Reakcja
1	308	0	249,56
2	308	1	258,7047
3	308	2	250,8006
4	308	3	321,4398
5	308	4	356,8519
6	308	5	414,6901
7	308	6	382,2038
8	308	7	290,1486
9	308	8	430,5853
10	308	9	466,3535
11	309	0	222,7339
12	309	1	205,2658
13	309	2	202,9778
14	309	3	204,707
15	309	4	207,7161

Chcemy odpowiedzieć na trzy pytania: (1) o ile wydłuża się czas reakcji z każdym dniem bez snu, (2) czy ludzie różnią się odpornością na niedobór snu, czyli czy tempo pogarszania się reakcji jest u wszystkich takie samo, (3) które osoby reagują na deprywację najsilniej.

Gdybyśmy policzyli zwykłą regresję wieloraką czasu reakcji względem dnia, program potraktowałby 180 pomiarów jak 180 niezależnych osób, choć w istocie mamy 18 osób, każdą mierzoną dziesięciokrotnie. Dlatego budujemy model mieszany: Reakcja jako zmienną zależną Y, Osoba jako zmienną grupującą (ID), Dzień jako zmienną niezależną X_1. Ponieważ pytamy, czy tempo zmian różni się między osobami,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

zaznaczamy Dzień na liście Losowe nachylenia; kowariancję zostawiamy pełną z korelacją, estymację REML, stopnie swobody Satterthwaite. Włączamy Model pusty i ICC, Test efektów losowych (LR), Efekty losowe grup (BLUP) i Dołącz wykresy.

Regresja wieloraka mieszana (liniowy model mieszany, efekty losowe)	
Analizowane zmienne	Reakcja
	Dzień
	Osoba
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0,05
Zmienna grupująca	Osoba
Liczność	180
Liczba grup	18
Obserwacji na grupę (min / mediana / max)	10 / 10 / 10
Liczba zmiennych w modelu	1
Metoda estymacji	REML
Ilość iteracji	169
Zbieżność	osiągnięta
Dopasowanie osobliwe	nie
Log Likelihood (REML)	-871,814136
-2LL	1743,628272
Liczba parametrów	6
AIC - kryterium Akaikego	1755,628272
AICc - skorygowane kryterium Akaikego	1756,113821
BIC - bayesowskie kryterium Schwarza	1774,786013
R2 brzegowe (efekty stałe)	0,278652
R2 warunkowe (stałe + losowe)	0,799219
Wariancja reszt	654,941024
Odch. std. reszt	25,591816
Test LR efektów losowych (vs regresja wieloraka, ML)	
Chi-kwadrat	148,353712
stopnie swobody	3
Wartość p	<0,000001
Model pusty	
Wariancja między grupami	1278,338204
Wariancja reszt	1958,865115
ICC	0,39489
-95% CI	0,200158
+95% CI	0,589621

Krok 1 – czy model jest poprawnie dopasowany. Raport podaje 180 obserwacji w 18 grupach (po 10 obserwacji na grupę), zbieżność osiągniętą i brak dopasowania osobliwego. Można czytać dalej.

Krok 2 – ile zmienności leży między osobami (model pusty i ICC). W modelu pustym wariancja między grupami wynosi 1278.34, a wariancja reszt 1958.87, stąd ICC $= 0.395$ (95% CI: 0.200–0.590). Około 39% całej zmienności czasu reakcji to trwałe różnice między osobami (jedni mają z natury szybsze reakcje,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

inni wolniejsze), a dwa pomiary tej samej osoby są skorelowane na poziomie 0.39. To wyraźnie za dużo, by traktować pomiary jako niezależne.

Krok 3 – czy efekty losowe są potrzebne. Test LR efektów losowych wobec zwykłej regresji wielorakiej daje $\chi^2 = 148.35$, $df = 3$, $p < 0.000001$ – struktura grupowa jest niezbędna.

Efekty stałe (warunkowe)

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	df	wartość p
w. wolny	251,405105	6,824555	237,006565	265,803645	36,838314	17,000193	<0,000001
Dzień	10,467286	1,545789	7,205957	13,728615	6,771483	17,000087	0,000003

Krok 4 – efekty stałe (odpowiedź na pytanie 1). Wyraz wolny $b_0 = 251.41$ ms (95% CI: 237.01–265.80) to przeciętny czas reakcji w dniu 0, czyli przed deprywacją. Współczynnik przy dniu $b_1 = 10.47$ ms (95% CI: 7.21–13.73), $t = 6.77$, $df = 17.0$, $p = 0.000003$: każdy kolejny dzień bez snu wydłuża czas reakcji przeciętnej osoby o około 10.5 ms. Warto zwrócić uwagę na stopnie swobody: wynoszą 17 (18 osób minus 1), a nie 178, bo o tempie zmian informuje nas 18 osób, nie 180 pomiarów. Dla porównania zwykła regresja wieloraka dałaby ten sam współczynnik 10.47, ale z błędem 0.80 zamiast 1.55 – dwukrotnie zaniżonym.

Efekty losowe

	składnik	wariancja	odch. std.	LR chi2	stopnie swob	wartość p
Osoba	w. wolny	612,089596	24,740445	22,140971	2	0,000009
	Dzień	35,071683	5,922135	42,139299	2	<0,000001
Reszty		654,941024	25,591816			

test LR: model z

Korelacje efektów losowych

	w. wolny	Dzień
w. wolny	1	0,065551
Dzień	0,065551	1

Krok 5 – efekty losowe (odpowiedź na pytanie 2). Odchylenie standardowe losowego wyrazu wolnego wynosi 24.74 ms: typowa osoba w dniu 0 różni się od przeciętnej o około 25 ms, a większość osób (95%) mieści się w przedziale $251 \pm 2 \cdot 25$, czyli od około 200 do 300 ms. Odchylenie standardowe losowego nachylenia wynosi 5.92 ms/dzień: skoro przeciętne tempo to 10.5 ms/dzień, to u jednych osób reakcja wydłuża się zaledwie o 0–5 ms dziennie, a u innych o 15–20 ms. Test LR dla nachylenia ($\chi^2 = 42.14$, $df = 2$, $p < 0.000001$) potwierdza, że tempo pogarszania się reakcji istotnie różni się między osobami. Korelacja wyrazu wolnego z nachyleniem wynosi tylko 0.066 – to, jak szybko ktoś reaguje na początku, nie mówi nic o tym, jak szybko będzie się pogarszał. Wariancja reszt 654.94 (odchylenie 25.59 ms) to zmienność pomiarów tej samej osoby z dnia na dzień, niewyjaśniona przez jej własną prostą.

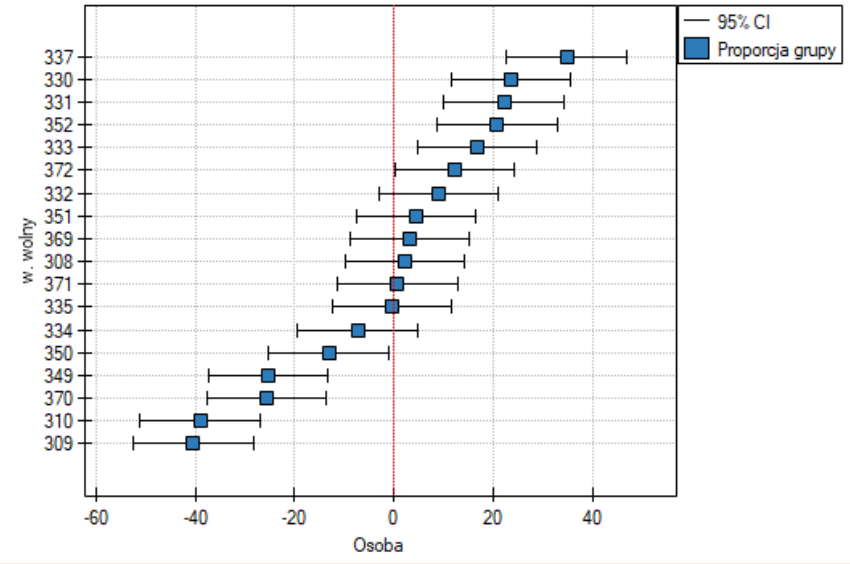
R^2 brzegowe $= 0.279$: sam dzień deprywacji wyjaśnia 28% zmienności czasu reakcji. R^2 warunkowe $= 0.799$: dzień razem z indywidualnymi poziomami i nachyleniami wyjaśnia 80%. Różnica (52 punkty procentowe) to zmienność przypisana różnicom między osobami.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

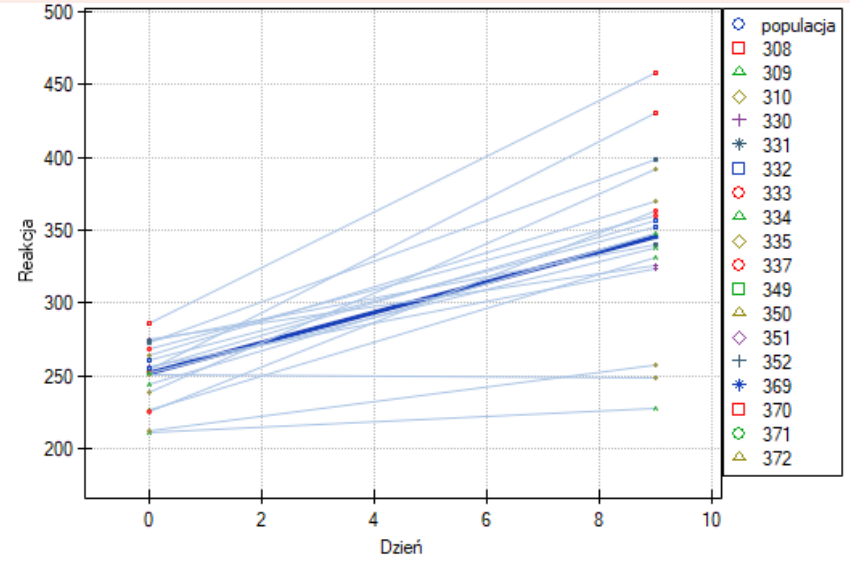
Efekty losowe grup (BLUP)

grupa	w. wolny	błąd	Dzień	błąd
308	2,258558	12,070837	9,198973	2,304837
309	-40,398571	12,070837	-8,619704	2,304837
310	-38,960242	12,070837	-5,44888	2,304837
330	23,690503	12,070837	4,814332	2,304837

Krok 6 – które osoby są nietypowe (odpowiedź na pytanie 3). Tabela BLUP podaje odchylenia każdej osoby od populacji. Osoba 308 ma wyraz wolny $+2.26$ (start przeciętny) i nachylenie $+9.20$: jej czas reakcji rośnie o $10.47 + 9.20 \approx 19.7$ ms dziennie, niemal dwa razy szybciej niż przeciętnie – to osoba najbardziej wrażliwa na brak snu. Osoby 309 i 310 startują nisko (odpowiednio -40.4 i -39.0 ms, czyli reagują szybko) i mają nachylenia -8.62 i -5.45: ich reakcja wydłuża się tylko o około 2 i 5 ms dziennie – są niemal odporne na deprywację. Te same informacje pokazują wykresy: caterpillar (osoby uporządkowane według efektu, z przedziałem błędu) i trajektorie grup (cienkie proste poszczególnych osób na tle grubej prostej populacyjnej).



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Dzień: $b = 10.47$ **ms (**$p = 0.000003$**), ICC** $= 0.395$**,** SD **nachylenia** $= 5.92$ **ms/dzień (LR** $p < 0.000001$**)**

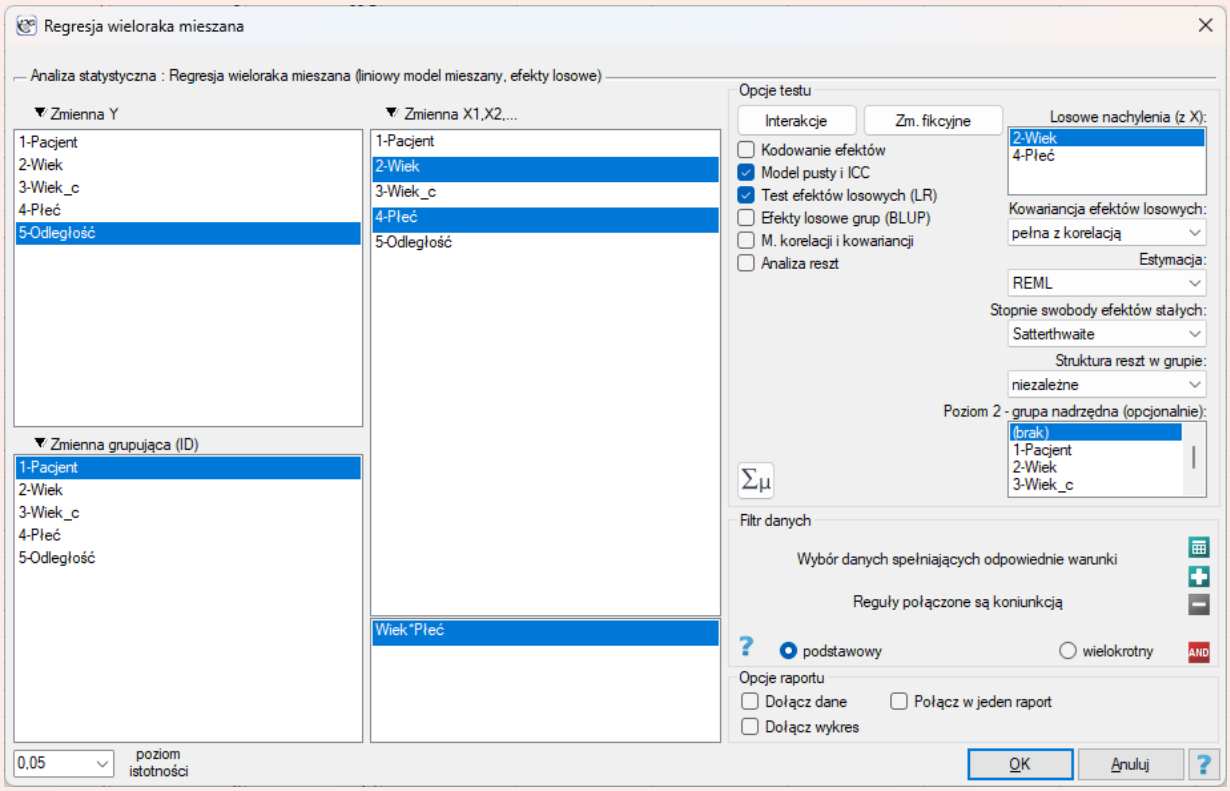
Każdy dzień deprywacji snu wydłuża czas reakcji średnio o 10.5 ms, ale tempo tego pogorszenia istotnie różni się między osobami: od niemal zerowego do około 20 ms na dzień.

Co zmieniają inne ustawienia. Ten sam model liczony metodą ML daje identyczne współczynniki efektów stałych, ale mniejsze wariancje efektów losowych (565.5 i 32.7 zamiast 612.1 i 35.1) – to znane obciążenie ML, dlatego do raportowania używamy REML. Wybór kowariancji niezależne usuwa korelację (i tak bliską zera): wariancje niemal się nie zmieniają, a AIC spada z 1755.63 do 1753.67, bo model ma o jeden parametr mniej – prostszy model jest tu równie dobry. Wybór Test z Walda zamiast Satterthwaite'a daje dla dnia $z = 6.77$ i $p < 0.000001$ zamiast $p = 0.000003$; przy 18 grupach różnica jest mała, ale przy kilku grupach test Walda byłby zbyt liberalny.

Przykład 26.2. (plik ortodoncja.pqs)
W klasycznym badaniu ortodontycznym (Potthoff i Roy 1964[197]) u 27 dzieci (16 chłopców i 11 dziewczynek) mierzono na zdjęciach rentgenowskich odległość między przysadką a szczeliną skrzydłowo-szczękową (w mm) w wieku 8, 10, 12 i 14 lat. Aby pokazać działanie modelu przy danych niezbalansowanych, z pełnego zbioru usunięto losowo 20 pomiarów – pozostało 88 obserwacji, od 2 do 4 na dziecko. Kolumny: Pacjent, Wiek, Wiek_c (wiek pomniejszony o 11 lat, czyli wycentrowany), Płeć (0 – chłopiec, 1 – dziewczynka) i Odległość. Pytamy, czy dziewczynki i chłopcy rosną w tym wymiarze w innym tempie.

Budujemy model z Odległość jako Y, Pacjent jako zmienną grupującą i zmiennymi niezależnymi Wiek, Płeć oraz interakcją Wiek$\times$Płeć utworzoną przyciskiem Interakcje. Losowe nachylenie nadajemy zmiennej Wiek (każde dziecko rośnie we własnym tempie); Płeć nie może mieć losowego nachylenia, bo jest stała w obrębie dziecka. Interakcja jest kluczem do pytania badawczego: jej współczynnik mówi, o ile nachylenie (tempo wzrostu) dziewczynek różni się od nachylenia chłopców.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Efekty stałe (warunkowe)

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	df	wartość p
w. wolny	15,641529	1,08197	13,407767	17,87529	14,456528	23,861652	<0,000001
Wiek	0,842648	0,095888	0,644103	1,041193	8,787821	22,618417	<0,000001
Płeć	2,111642	1,777548	-1,55245	5,775733	1,187952	24,58156	0,246204
Wiek*Płeć	-0,394188	0,156384	-0,716839	-0,071537	-2,520636	24,156631	0,018721

Model jest zbieżny, bez dopasowania osobliwego, mimo że część dzieci ma tylko 2 pomiary – model mieszany nie wymaga kompletnych danych. Efekty stałe: wyraz wolny 15.64 to odległość chłopca w wieku 0 lat (ekstrapolacja bez sensu klinicznego); wiek $b = 0.843$ mm/rok ($p < 0.000001$) to tempo wzrostu chłopców; płeć $b = 2.11$ ($p = 0.246$) to różnica dziewczynek względem chłopców w wieku 0 lat, znów pozbawiona sensu; interakcja $b = -0.394$ mm/rok ($t = -2.52$, $df = 24.2$, $p = 0.019$): dziewczynki rosną o 0.39 mm/rok wolniej niż chłopcy, czyli w tempie $0.843 - 0.394 = 0.449$ mm/rok. To odpowiedź na pytanie badawcze.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Efekty losowe

	składnik	wariancja	odch. std.	LR chi2	stopnie swob	wartość p
Pacjent	w. wolny	6,799717	2,607627	2,043173	2	0,256457
	Wiek	0,055702	0,236013	2,093144	2	0,24955
Reszty		1,546741	1,24368			

test LR: model z

Korelacje efektów losowych

	w. wolny	Wiek
w. wolny	1	-0,780067
Wiek	-0,780067	1

Efekty losowe: odchylenie standardowe wyrazu wolnego 2.61 mm i nachylenia 0.24 mm/rok, ale korelacja między nimi wynosi -0.78. Nie oznacza to, że dzieci „większe” rosną wolniej – wyraz wolny odnosi się do wieku 0, daleko poza zakresem danych (8–14 lat), i przy takim położeniu zera korelacja jest artefaktem geometrii prostych. Testy LR dla obu składników nie są istotne ($p = 0.256$ i $p = 0.250$): przy 27 dzieciach i 2–4 pomiarach na dziecko nie da się wiarygodnie odróżnić indywidualnych temp wzrostu, a model z samym losowym wyrazem wolnym byłby równie dobry.

Centrowanie wieku. Powtórzmy analizę, zastępując Wiek kolumną Wiek_c (wiek -11) zarówno wśród zmiennych X, w interakcji, jak i na liście losowych nachyleń. Dopasowanie modelu ($-2LL$, AIC) i współczynniki przy wieku oraz interakcji pozostają identyczne, zmienia się natomiast znaczenie wyrazu wolnego i płci: wyraz wolny 24.91 mm to odległość przeciętnego chłopca w wieku 11 lat, płeć $b = -2.22$ ($p = 0.006$) to różnica dziewczynek w wieku 11 lat (są mniejsze o 2.2 mm), a korelacja efektów losowych zmienia się z -0.78 na $+0.33$. Wniosek ogólny: gdy wyraz wolny ma być interpretowalny, a korelacja efektów losowych sensowna, zmienną czasu (wieku) należy wycentrować w środku zakresu obserwacji.

interakcja Wiek×Płeć: $b = -0.39$ mm/rok, $p = 0.019$

Dziewczynki rosną w tym wymiarze istotnie wolniej niż chłopcy (0.45 wobec 0.84 mm/rok); w wieku 11 lat są o 2.2 mm mniejsze.

Przykład 26.3. (plik ośrodkiPacjenci.pqs)
Dane symulowane: w 8 ośrodkach u 6 pacjentów w każdym wykonano po 5 pomiarów w czasie ($t = 0, 1, \ldots, 4$) pewnego wskaźnika Y, rejestrując też zmienną ciągłą X. Pacjenci są zagnieżdżeni w ośrodkach, więc dwa źródła zależności nakładają się: pomiary tego samego pacjenta są podobne, a pacjenci tego samego ośrodka też są do siebie podobni.

Wskazujemy Y, zmienną grupującą Pacjent, zmienne niezależne t i X, a w polu Poziom 2 – grupa nadrzędna wybieramy Ośrodek. Zmienna grupująca musi być jednoznaczna w całym arkuszu (pacjent 1 w ośrodku 1 i pacjent 1 w ośrodku 2 to różne osoby, więc muszą mieć różne identyfikatory).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ICC	0,620058
ICC poziomu 2 (Ośrodek)	0,429007
-95% CI	NA
+95% CI	NA

Efekty stałe (warunkowe)

	wsp. b	błąd b	-95% CI	+95% CI	stat. t	df	wartość p
w. wolny	10,753816	1,26561	7,775658	13,731973	8,496944	7,172563	0,000054
t	1,367874	0,069479	1,230825	1,504923	19,68763	190,021452	<0,000001
X	-2,070848	0,11767	-2,302893	-1,838803	-17,59876	198,300312	<0,000001

Efekty losowe

	składnik	wariancja	odch. std.	LR chi2	stopnie swob	wartość p
Pacjent	w. wolny	5,333983	2,309542			
	w. wolny po:	11,692803	3,419474			
Reszty		2,316493	1,522003			

Raport pokazuje trzy wariancje: pacjentów (5.33), ośrodków (11.69) i reszt (2.32). Model pusty daje ICC całkowite 0.620 (62% zmienności to różnice między pacjentami i ośrodkami razem) oraz ICC poziomu 2 równe 0.429 (43% zmienności to same różnice między ośrodkami). Efekty stałe: czas $b = 1.37$ ($p < 0.000001$), X: $b = -2.07$ ($p < 0.000001$). Stopnie swobody dla wyrazu wolnego wynoszą tylko 7.2 – o poziomie ogólnym informuje w istocie 8 ośrodków, dlatego jego przedział ufności jest szeroki (7.78–13.73), mimo 240 obserwacji. Osobna tabela BLUP poziomu 2 wskazuje ośrodki o nietypowym poziomie Y (np. ośrodek 1: -3.86).

Przykład 26.4. c.d. przykładu 26.1 (plik deprywacjaSnu.pqs)
Pomiary czasu reakcji wykonywano dzień po dniu, więc reszty sąsiednich dni mogą być skorelowane (zły dzień pociąga za sobą kolejny). Sprawdzimy to modelem z samym losowym wyrazem wolnym (bez losowego nachylenia) i strukturą reszt $\mathrm{AR}(1)$; wiersze każdej osoby są w arkuszu uporządkowane według dnia, co jest warunkiem tej opcji.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Struktura reszt w grupie	AR(1), rho = 0,653133
Log Likelihood (REML)	-868,095532
-2LL	1736,191064
Liczba parametrów	5
AIC - kryterium Akaikego	1746,191064
AICc - skorygowane kryterium Akaikego	1746,535891
BIC - bayesowskie kryterium Schwarza	1762,155848
R2 brzegowe (efekty stałe)	0,272839
R2 warunkowe (stałe + losowe)	0,574758
Wariancja reszt	1416,569316

Efekty losowe

	składnik	wariancja	odch. std.	LR chi2	stopnie swob	wartość p
Osoba	w. wolny	1005,752667	31,713604	3,283727	1	0,034985
Reszty		1416,569316	37,637339	49,248553	1	<0,000001

Parametr $\rho = 0.653$: reszty dwóch sąsiednich dni są skorelowane na poziomie 0.65, dni odległych o dwa – $0.65^2 \approx 0.43$ itd. Model AR(1) z losowym wyrazem wolnym ma $-2LL = 1736.19$ i AIC $= 1746.19$, podczas gdy ten sam model bez struktury reszt ma $-2LL = 1786.47$ i AIC $= 1794.47$ – poprawa jest ogromna. Co ciekawe, AIC modelu AR(1) jest też niższe niż modelu z losowym nachyleniem z przykładu 26.1 (1755.63; porównanie dozwolone, bo oba modele mają te same efekty stałe i liczone są metodą REML). Oba modele opisują to samo zjawisko innym językiem: losowe nachylenie mówi „każda osoba ma własne stałe tempo", AR(1) mówi „odchylenia od prostej utrzymują się przez kilka dni". Efekt dnia pozostaje niemal ten sam ($b = 10.47$, błąd 1.34).

DLA ZAINTERESOWANYCH

Estymacja: Parametry szacuje się metodą największej wiarygodności (ML) lub ograniczonej największej wiarygodności (REML; Laird i Ware 1982[193], Bates i in. 2015[189]). Kowariancja efektów losowych grupy jest parametryzowana jako $\sigma^2 \Lambda\Lambda^T$, gdzie Λ jest macierzą dolnotrójkątną (przy kowariancji niezależne – diagonalną). Dla ustalonego Λ efekty stałe i σ^2 mają postać zamkniętą (uogólniona metoda najmniejszych kwadratów), a profilowana dewiancja

$$d(\Lambda) = \sum_j \log\det A_j + (n-p)\left[1 + \log\frac{2\pi r^2}{n-p}\right] + \log\det(X^T W X)$$

(dla REML; dla ML bez ostatniego składnika i z n zamiast $n-p$) jest minimalizowana numerycznie metodą Neldera-Meada. $A_j = I + \Lambda^T Z_j^T Z_j \Lambda$, W jest odwrotnością macierzy kowariancji brzegowej, a r^2 ważoną resztową sumą kwadratów. Obliczenia są prowadzone blokowo, grupa po grupie. Przy strukturze reszt AR(1) lub CS iloczyny $Z_j^T Z_j$ i sumy kwadratów liczone są z wagami R_j^{-1}, gdzie R_j jest macierzą korelacji reszt grupy ($\rho^{|i-i'|}$ dla AR(1), stała ρ poza przekątną dla CS), a ρ jest dodatkowym parametrem optymalizacji. Drugi poziom zagnieżdżenia dodaje losowy wyraz wolny grupy nadrzędnej o wariancji $\tau^2_{(2)}$.

Stopnie swobody Satterthwaite'a: dla j-tego efektu stałego $df_j = 2\,\mathrm{Var}(\hat\beta_j)^2 / \mathrm{Var}[\mathrm{Var}(\hat\beta_j)]$, gdzie wariancja wariancji jest liczona metodą delta z hesjanu dewiancji względem parametrów wariancji (Kuznetsova i in. 2017[192]).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ICC i jego przedział ufności: $ICC = \tau^2/(\tau^2 + \sigma^2)$ z modelu pustego; przedział ufności metodą delta na podstawie hesjanu dewiancji względem (τ^2, σ^2).

R^2 **Nakagawy:** $R^2_{brzeg} = \sigma^2_f/(\sigma^2_f + \sigma^2_u + \sigma^2)$ oraz $R^2_{war} = (\sigma^2_f + \sigma^2_u)/(\sigma^2_f + \sigma^2_u + \sigma^2)$, gdzie σ^2_f jest wariancją predykcji z efektów stałych, a σ^2_u średnią wariancją efektów losowych (dla losowych nachyleń uśrednioną po wartościach X).

Test LR na granicy: statystyka $2(\ell_1 - \ell_0)$ dla hipotezy o zerowej wariancji ma asymptotycznie rozkład będący mieszanką 50:50 rozkładów χ^2_{q-1} i χ^2_q (Self i Liang 1987[198]); dla jednego parametru wartość p z rozkładu χ^2_1 dzieli się przez dwa.

Efekty losowe grup: $\hat{u}_j = \Lambda A_j^{-1} \Lambda^T Z_j^T (y_j - X_j \hat{\beta})$ z warunkową kowariancją $\sigma^2 \Lambda A_j^{-1} \Lambda^T$ (najlepsze liniowe nieobciążone predyktory, BLUP).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Walidacja i predykcja dla regresji wielorakiej mieszanej

KIEDY STOSOWAĆ:

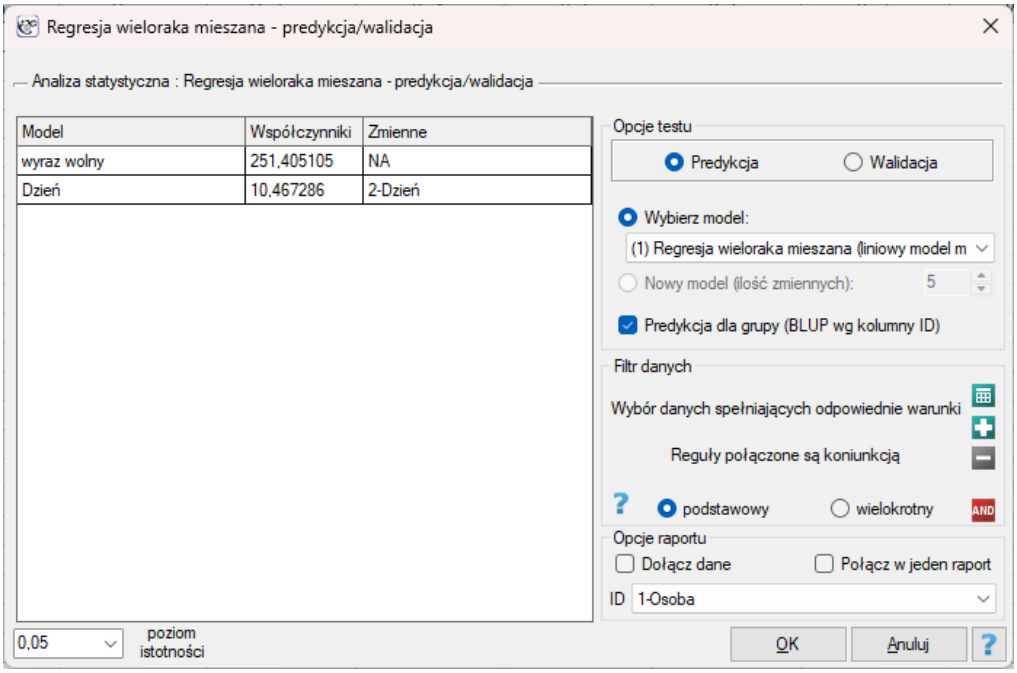
Zbudowany i zweryfikowany model mieszany można wykorzystać do **predykcji** wartości Y dla nowych danych oraz do **walidacji** na zbiorze testowym, tak jak w walidacji i predykcji dla regresji wielorakiej. Model mieszany pozwala jednak na dwa rodzaje predykcji:

- **predykcja populacyjna** – z samych efektów stałych (efekty losowe równe 0); odpowiada na pytanie „jakiej wartości spodziewamy się u przeciętnej, nowej grupy (pacjenta, ośrodka)";
- **predykcja dla grupy** – z efektami stałymi i efektem losowym (BLUP) konkretnej grupy, która brała udział w budowie modelu; odpowiada na pytanie „jakiej wartości spodziewamy się u tego pacjenta (w tym ośrodku)". Jest dokładniejsza, ale dostępna tylko dla grup znanych modelowi.

Dla nowych obserwacji, których identyfikator nie występował w danych uczących, program zawsze stosuje predykcję populacyjną.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja wieloraka mieszana – predykcja/walidacja.



Predykcja wymaga wskazania **modelu zapisanego w programie** (raportu regresji wielorakiej mieszanej przypisanego do arkusza), bo oprócz współczynników potrzebne są efekty losowe grup i macierz kowariancji; nie ma możliwości ręcznego wpisania współczynników. Po wybraniu modelu program wypełnia listę zmiennych, a my wskazujemy kolumny arkusza z nowymi wartościami X. Opcje:

- Predykcja dla grupy (BLUP wg kolumny ID) – po zaznaczeniu wskazujemy kolumnę z identyfikatorem grupy; dla identyfikatorów znanych modelowi predykcja uwzględnia BLUP grupy, dla pozostałych jest populacyjna. Bez tej opcji wszystkie predykcje są populacyjne;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– walidacja – gdy w nowych danych jest obserwowane Y, wskazujemy je, a raport zawiera miary dopasowania modelu do zbioru testowego.

Dane do predykcji umieszczamy w tym samym arkuszu co dane uczące (np. w kolejnych wierszach) i ograniczamy analizę filtrem.

WYNIKI I INTERPRETACJA

Raport podaje, jaki rodzaj predykcji zastosowano (populacyjna lub dla grup), liczbę obserwacji, dla których użyto BLUP grupy, oraz dla każdej obserwacji przewidywane Y z przedziałem ufności. Przedział predykcji populacyjnej jest szerszy niż przedział dla znanej grupy, bo obejmuje niepewność co do poziomu nowej grupy. W trybie walidacji dodatkowo wyznaczane są błędy predykcji i miary dopasowania.

Przykład 26.5. c.d. przykładu 26.1 (plik deprywacjaSnu.pqs)
Na podstawie modelu z losowym wyrazem wolnym i losowym nachyleniem dnia chcemy przewidzieć czas reakcji w dziesiątym dniu deprywacji (dzień $= 10$, czyli jeden dzień poza zakresem badania) dla osoby 308, dla osoby 309 oraz dla nowej, nieznanej osoby. Do arkusza dopisujemy trzy wiersze z wartością 10 w kolumnie Dzień i identyfikatorami 308, 309 i 999, po czym uruchamiamy predykcję z filtrem na te wiersze i włączoną opcją Predykcja dla grupy.

Predykcja	
Średnia Pred. Y	345,295525
Odchylenie standardowe Pred. Y	110,816079
Minimum Pred. Y	229,482352
Maximum Pred. Y	450,326257

Współczynniki modelu	
wyraz wolny	251,405105
Dzień	10,467286

Predykcja					
	Pred. Y	-95% CI (pu	+95% CI (pı	-95% CI (wa	+95% CI (w
308 (BLUP)	450,326257	390,035301	510,617213	371,678359	528,974155
309 (BLUP)	229,482352	169,191396	289,773308	150,834454	308,13025
999	356,077965	324,473111	387,682819	296,501466	415,654464

Dla nowej osoby (999) predykcja jest populacyjna: $251.41 + 10.47 \cdot 10 \approx 356$ ms. Dla osoby 308 program dodaje jej efekty losowe (wyraz wolny $+2.26$, nachylenie $+9.20$): $(251.41 + 2.26) + (10.47 + 9.20) \cdot 10 \approx 450$ ms, a dla osoby 309 (efekty -40.40 i -8.62): $(251.41 - 40.40) + (10.47 - 8.62) \cdot 10 \approx 230$ ms. Różnica między osobami 308 i 309 sięga 220 ms, choć populacyjnie obie „powinny” mieć 356 ms – to ilustruje, ile informacji wnoszą efekty losowe, gdy grupa jest znana. Przedział ufności predykcji populacyjnej jest zarazem najszerszy, bo obejmuje rozrzut między osobami.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Porównywanie modeli regresji wielorakiej mieszanej

KIEDY STOSOWAĆ:

Porównywanie modeli służy do wyboru takiego zestawu zmiennych niezależnych (efektów stałych), który opisuje dane najlepiej przy możliwie małej liczbie parametrów. Działa tak samo jak porównywanie modeli regresji wielorakiej: porównujemy model pełny z modelem zredukowanym (bez części zmiennych) albo pozwalamy programowi wybrać model automatycznie (dobór wsteczny, w przód, krokowy lub przegląd podzbiorów).

W modelach mieszanych obowiązuje jedna dodatkowa zasada: **struktura losowa jest wspólna dla wszystkich porównywanych modeli**. Dobór dotyczy wyłącznie efektów stałych, a losowy wyraz wolny, losowe nachylenia i typ kowariancji ustalone w oknie pozostają w każdym z modeli. Zmienna, której nadano losowe nachylenie, pozostaje w każdym porównywanym modelu jako efekt stały (nie można usunąć efektu stałego zmiennej, zachowując jej losowe nachylenie). O tym, czy same efekty losowe są potrzebne, decydują testy LR w raporcie regresji wielorakiej mieszanej.

DEFINICJE MIAR

Modele porównuje się na podstawie kryteriów informacyjnych AIC, AICc i BIC (im mniejsza wartość, tym lepszy model) oraz **testu ilorazu wiarygodności** (LR) dla modeli zagnieżdżonych: statystyka $2(\ell_P - \ell_Z)$ (ℓ_P, ℓ_Z – logarytm wiarygodności modelu pełnego i zredukowanego) ma rozkład χ^2 o liczbie stopni swobody równej liczbie usuniętych zmiennych. Nieistotny wynik oznacza, że usunięte zmienne nie wnoszą informacji i należy wybrać model prostszy; wynik istotny – że model pełny jest istotnie lepszy.

Uwaga! Kryteria informacyjne i test LR dla modeli różniących się efektami stałymi są poprawne tylko przy estymacji metodą największej wiarygodności (ML). Dlatego **dobór modeli jest zawsze prowadzony metodą ML**, niezależnie od ustawienia Współczynniki modeli. To ustawienie decyduje jedynie, jaką metodą (REML czy ML) mają być policzone tabele współczynników prezentowanych modeli – zalecane REML, bo daje nieobciążone wariancje.

Ocena modeli poza zbiorem uczącym

Kryteria informacyjne i test LR są wyznaczane na tych samych danych, na których dopasowano modele, przez co ocena modeli jest zbyt optymistyczna. Tak jak w porównywaniu modeli regresji wielorakiej, w oknie analizy można więc włączyć ocenę modeli na obserwacjach, które nie brały udziału w ich dopasowaniu (opcja Walidacja): **walidację krzyżową** k**-krotną** albo **losowy zbiór testowy**, z podziałem wyznaczonym przez Ziarno. Dla każdego modelu raport podaje wtedy RMSE, MAE i R^2 predykcji, a w bloku doboru modelu – błąd modelu zerowego i najlepszy model wg walidacji (o najmniejszym RMSE); jako kryterium doboru automatycznego można też wybrać błąd walidacji. W modelach mieszanych obowiązują dwie dodatkowe zasady:

- **dzielone są całe grupy** – wszystkie obserwacje jednej grupy (np. jednego pacjenta) trafiają do tej samej części danych. Obserwacje tej samej grupy są ze sobą skorelowane, więc rozdzielenie ich między część uczącą i ocenianą zaniżałoby błąd. Liczba podziałów k jest ograniczana do liczby grup, a udział zbioru testowego oznacza procent grup;
- **przewidywanie dla nowej grupy opiera się wyłącznie na efektach stałych** – efekt losowy grupy, której nie było w zbiorze uczącym, nie jest znany, dlatego przyjmuje się jego wartość oczekiwaną, czyli zero. Miary walidacji mówią więc, jak dobrze efekty stałe modelu przewidują obserwacje

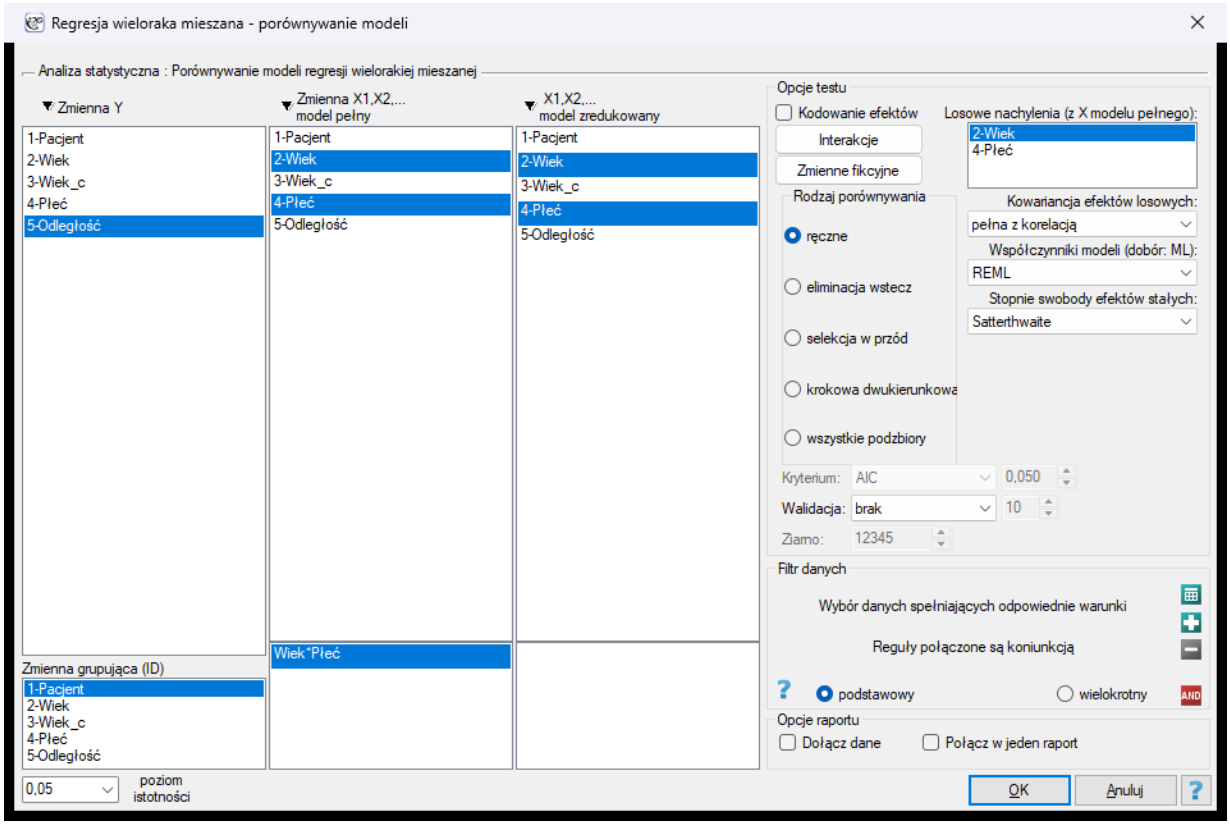
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

nowej grupy (np. nowego pacjenta), i odpowiadają R^2 brzegowemu, a nie warunkowemu. RMSE obejmuje także zmienność między grupami, jest więc większe niż odchylenie standardowe reszt modelu.

Modelem zerowym jest tu model z wyrazem wolnym, wspólną strukturą losową i zmiennymi, którym nadano losowe nachylenie (pozostają one w każdym modelu); jest on dopasowywany i oceniany tak samo jak pozostałe modele.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja wieloraka mieszana – porównywanie modeli.



W oknie wskazujemy zmienną zależną Y, zmienne modelu pełnego i zmienne modelu zredukowanego (lub kryterium doboru automatycznego), tak jak w porównywaniu modeli regresji wielorakiej. Opcje modelu mieszanego:

- Zmienna grupująca (ID) – kolumna identyfikująca grupę; wspólna dla wszystkich modeli,
- Losowe nachylenia (z X modelu pełnego) – zmienne, których efekt różni się między grupami; taka zmienna pozostaje w każdym porównywanym modelu,
- Kowariancja efektów losowych – niezależne lub pełna z korelacją, wspólna dla wszystkich modeli,
- Współczynniki modeli (dobór: ML) – REML lub ML dla tabel współczynników; sam dobór zawsze ML,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Stopnie swobody efektów stałych – Satterthwaite lub test z Walda w tabelach współczynników,
- Kryterium – kryterium doboru automatycznego: AIC, AICc, BIC, p (test LR / F) z progiem albo błąd walidacji,
- Walidacja – brak, krzyżowa, k = z liczbą podziałów (od 2 do 100, domyślnie 10) albo zbiór testowy, % = z udziałem zbioru testowego w procentach grup (od 5 do 90, domyślnie 30); dzielone są całe grupy,
- Ziarno – ziarno losowania podziału (domyślnie 12345); to samo ziarno oznacza ten sam podział.

WYNIKI I INTERPRETACJA

Raport zawiera ramkę z opisem wspólnej struktury losowej (zmienna grupująca, liczba grup, składniki losowe, kowariancja), tabelę **Kryteria informacyjne i testy LR** z wartościami $-2LL$ (ML), AIC, AICc, BIC, R^2 brzegowym i warunkowym każdego modelu oraz testem LR między kolejnymi modelami, a także tabele współczynników porównywanych modeli (efekty stałe warunkowe, z wariancjami efektów losowych każdego modelu). Jeśli dla któregoś modelu wystąpiło dopasowanie osobliwe, jest to zaznaczone w tabeli. Przy włączonej walidacji raport zawiera dodatkowo blok **Ocena poza zbiorem uczącym** (sposób podziału, ziarno, jednostka podziału, błąd modelu zerowego, najlepszy model wg walidacji), a przy każdym modelu – RMSE, MAE i R^2 predykcji.

Przykład 26.6. c.d. przykładu 26.2 (plik ortodoncja.pqs)
W modelu wzrostu odległości przysadka–szczelina skrzydłowo-szczękowa znalazły się: wiek, płeć i interakcja wiek×płeć. Sprawdzimy, czy interakcja (różne tempo wzrostu dziewczynek i chłopców) jest rzeczywiście potrzebna, porównując model pełny z modelem zredukowanym bez interakcji. Struktura losowa: losowy wyraz wolny i losowe nachylenie wieku, kowariancja pełna, wspólne dla obu modeli.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Liczba zmiennych w modelu 1	3
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood (ML)	348,704807
AIC - kryterium Akaikego	364,704807
AICc - skorygowane kryterium Akaikego	366,527592
BIC - bayesowskie kryterium Schwarza	384,523501
R2 brzegowe (efekty stałe)	0,465299
R2 warunkowe (stałe + losowe)	0,818385
Wariancja: w. wolny	6,799717
Wariancja: Wiek	0,055702
Wariancja: reszty	1,546741
Liczba zmiennych w modelu 2	2
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood (ML)	354,986352
AIC - kryterium Akaikego	368,986352
AICc - skorygowane kryterium Akaikego	370,386352
BIC - bayesowskie kryterium Schwarza	386,32771
R2 brzegowe (efekty stałe)	0,42409
R2 warunkowe (stałe + losowe)	0,811562
Wariancja: w. wolny	9,855281
Wariancja: Wiek	0,078336
Wariancja: reszty	1,568481
Chi-kwadrat - porównanie modeli Model 1 vs Model 2	6,281545
Stopnie swobody	1
Wartość p	0,0122

Model pełny ma $-2LL = 348.70$ (ML), AIC $= 364.70$, BIC $= 384.52$; model bez interakcji $-2LL = 354.99$, AIC $= 368.99$, BIC $= 386.33$. Test LR: $\chi^2 = 354.99 - 348.70 = 6.28$, $df = 1$, $p = 0.012$ – usunięcie interakcji istotnie pogarsza dopasowanie, a oba kryteria informacyjne wskazują model pełny. Zostajemy przy modelu z interakcją: dziewczynki i chłopcy rosną w różnym tempie. Gdybyśmy dodatkowo porównali model z samym wiekiem ($-2LL = 360.72$, AIC $= 372.72$), okazałoby się, że płeć wnosi informację tylko razem z interakcją – w modelu bez interakcji jej efekt to średnia różnica w całym zakresie wieku ($b = -2.04$, $p = 0.008$), która ukrywa fakt, że różnica ta rośnie z wiekiem.

test LR pełny vs bez interakcji: $\chi^2 = 6.28$, $df = 1$, $p = 0.012$**; AIC** 364.70 **vs** 368.99

Interakcja wiek×płeć pozostaje w modelu.

Ocena poza zbiorem uczącym. Sprawdzimy jeszcze, czy interakcja poprawia przewidywanie wzrostu u

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

dzieci, które nie brały udziału w dopasowaniu modelu. To samo porównanie powtarzamy, wybierając w polu Walidacja opcję krzyżowa, k = z 10 podziałami (ziarno 12345). Dzielonych jest 27 pacjentów, a nie 88 pomiarów – wszystkie pomiary jednego dziecka trafiają do tej samej części.

Ocena poza zbiorem uczącym	walidacja krzyżowa
Liczba podziałów (k)	10
Ziarno losowania podziału	12345
Jednostka podziału	cała grupa
Model zerowy: RMSE	2,463707
Model zerowy: MAE	1,842072
Najlepszy model wg walidacji (RMSE)	Model 1
Predykcja nowych grup: tylko efekty stałe	

	zbiór uczący	walidacja krzyżowa ($k = 10$)		
model	AIC	RMSE	MAE	R^2 predykcji
zerowy (sam wiek)		2.46	1.84	
Model 1 (z interakcją)	364.70	2.27	1.80	0.41
Model 2 (bez interakcji)	368.99	2.31	1.80	0.38

Model z interakcją ma mniejszy błąd poza zbiorem uczącym (RMSE $= 2.27$ mm wobec 2.31 mm) i większe R^2 predykcji (0.41 wobec 0.38); oba modele przewidują lepiej niż model zerowy, który – ze względu na losowe nachylenie wieku – zawiera sam wiek (RMSE $= 2.46$ mm). Walidacja potwierdza więc wynik testu LR i kryteriów informacyjnych, choć przewaga modelu z interakcją jest niewielka. Warto zauważyć, że R^2 predykcji (0.41) jest zbliżone do R^2 brzegowego modelu (0.47), a nie do warunkowego (0.82): dla nowego dziecka nie znamy jego efektu losowego, więc przewidywanie opiera się na samych efektach stałych.

walidacja krzyżowa ($k = 10$, całe grupy): RMSE $= 2.27$ (z interakcją), 2.31 (bez interakcji), 2.46 (model zerowy)

Najlepszym modelem według walidacji jest model z interakcją wiek$\times$płeć.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

26.2 REGRESJA LOGISTYCZNA MIESZANA

Regresja logistyczna mieszana (uogólniony model mieszany)

KIEDY STOSOWAĆ:

Regresja logistyczna mieszana (uogólniony liniowy model mieszany, ang. *generalized linear mixed model*, GLMM) jest rozszerzeniem regresji logistycznej na dane, w których obserwacje tworzą grupy: powtarzane oceny tego samego pacjenta (np. obecność objawu na kolejnych wizytach), pacjenci z tego samego ośrodka, zwierzęta z tego samego stada, uczniowie z jednej klasy (zob. wprowadzenie do modeli mieszanych). Wynik binarny (0/1) obserwacji z jednej grupy jest skorelowany, a zwykła regresja logistyczna, która tę korelację pomija, zaniża błędy standardowe i zawyża istotność.

Model opisuje logit prawdopodobieństwa zdarzenia:

$$\ln \frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)} = \beta_0 + \beta_1 X_{1ij} + \ldots + \beta_k X_{kij} + u_{0j} + u_{1j} X_{1ij},$$

gdzie:
Y_{ij} – wynik i-tej obserwacji w grupie j (0 lub 1),
$\beta_0, \beta_1, \ldots, \beta_k$ – efekty stałe,
u_{0j} – losowy wyraz wolny grupy j, u_{1j} – losowe nachylenie grupy j; zakłada się, że efekty losowe mają rozkład normalny o średniej 0 i macierzy kowariancji Σ na skali logitowej.

Losowy wyraz wolny oznacza, że każda grupa ma własne „bazowe" szanse zdarzenia: grupa o $u_{0j} = 1$ ma szanse $e^1 \approx 2.7$ razy większe niż grupa przeciętna, przy tych samych wartościach zmiennych. W przeciwieństwie do regresji wielorakiej mieszanej model nie ma wariancji reszt: na skali logitowej odpowiada jej stała $\pi^2/3 \approx 3.29$.

Model ten zastępuje ANOVA Q-Cochrana, gdy pomiary powtarzane są niekompletne lub gdy oprócz czynnika powtarzanego w modelu mają być zmienne ciągłe i zmienne opisujące grupę.

WARUNKI STOSOWANIA

- zmienna zależna dwuwartościowa,
- efekty losowe o rozkładzie normalnym na skali logitowej; grupy niezależne od siebie,
- co najmniej 3 grupy, a do wiarygodnego oszacowania wariancji zwykle kilkanaście lub więcej; przy losowych nachyleniach potrzeba też kilku obserwacji w każdej grupie,
- wystarczająca liczba zdarzeń (jedynek) i braków zdarzeń (zer) – jak w regresji logistycznej,
- braki danych losowe.

Uwaga! Dane muszą być w układzie długim: każdy wiersz to jedna obserwacja (jeden wynik 0/1), a osobna kolumna wskazuje grupę. Dane zagregowane (liczba zdarzeń i liczba prób w grupie) należy najpierw rozpisać na pojedyncze wiersze.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DEFINICJE MIAR

Warunkowy iloraz szans $OR = e^b$ – mówi, ile razy zmieniają się szanse zdarzenia, gdy X rośnie o jednostkę **w obrębie tej samej grupy** (u tego samego pacjenta, w tym samym ośrodku). Zwykła regresja logistyczna daje iloraz szans **populacyjny**, uśredniony po wszystkich grupach. Warunkowy OR jest co do wartości bezwzględnej zawsze większy (dalej od 1) niż populacyjny, a różnica rośnie wraz z wariancją efektów losowych; przy wariancji τ^2 populacyjny współczynnik wynosi w przybliżeniu $b/\sqrt{1 + 0.346\,\tau^2}$. Oba są poprawne, ale odpowiadają na różne pytania i nie należy ich porównywać wprost.

Wariancja efektów losowych τ^2 na skali logitowej i jej pierwiastek – odchylenie standardowe. Odchylenie standardowe losowego wyrazu wolnego równe 1 oznacza, że grupa o poziomie wyższym o jedno odchylenie ma szanse zdarzenia $e^1 \approx 2.7$ razy większe niż grupa przeciętna, a grupy skrajne (o ± 2 odchylenia) różnią się szansami $e^4 \approx 55$ razy.

ICC na skali logitowej – z modelu pustego:

$$ICC = \frac{\tau^2}{\tau^2 + \pi^2/3},$$

interpretowany jako korelacja utajonych skłonności do zdarzenia w obrębie grupy (udział zmienności między grupami w zmienności utajonej zmiennej ciągłej stojącej za wynikiem 0/1).

R^2 **brzegowe i warunkowe** (Nakagawa) – na skali logitowej, z $\pi^2/3$ jako wariancją reszt: udział wariancji wyjaśnianej przez efekty stałe oraz przez efekty stałe i losowe razem.

Aproksymacja wiarygodności – wiarygodność modelu wymaga scałkowania po efektach losowych, czego nie da się zrobić dokładnie. Laplace przybliża całkę jedną funkcją wokół jej maksimum – szybko i dostatecznie dokładnie przy większych grupach. Kwadratura Gaussa-Hermite'a (adaptacyjna, 7, 15 lub 25 węzłów) liczy całkę numerycznie w kilku punktach i jest dokładniejsza, zwłaszcza przy małych grupach (kilka obserwacji na grupę) i dużej wariancji; kosztem jest czas obliczeń. Kwadratura jest dostępna tylko dla modelu z samym losowym wyrazem wolnym; przy losowych nachyleniach program stosuje aproksymację Laplace'a. Wyniki różnych aproksymacji różnią się zwykle na trzecim miejscu znaczącym.

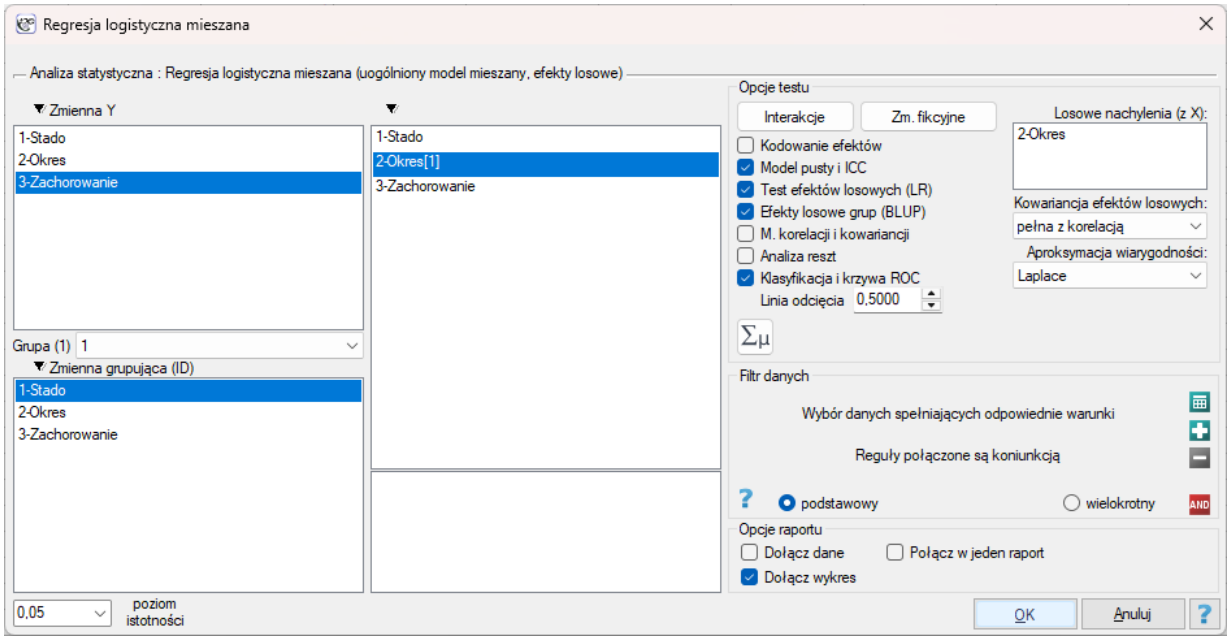
Test LR efektów losowych, BLUP, dopasowanie osobliwe – jak w regresji wielorakiej mieszanej; BLUP są odchyleniami grup na skali logitowej.

Predykcja warunkowa i populacyjna – prawdopodobieństwo przewidywane z efektami losowymi grupy („dla tego pacjenta") oraz z samych efektów stałych („dla przeciętnego pacjenta"). Klasyfikacja i krzywa ROC mogą być policzone dla obu; predykcja warunkowa jest zawsze dokładniejsza, bo korzysta z informacji o grupie, ale nie da się jej użyć dla nowych, nieznanych grup.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja logistyczna mieszana.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W oknie wskazujemy Zmienną zależną Y i w polu Grupa (1) wartość oznaczającą zdarzenie, Zmienną grupującą (ID) oraz Zmienne niezależne X1, X2, ...; zmienne fikcyjne i interakcje przygotowuje się jak w regresji logistycznej. Opcje modelu mieszanego:

- Losowe nachylenia (z X) – zmienne spośród X, których efekt ma się różnić między grupami; muszą zmieniać się wewnątrz grup. Losowy wyraz wolny jest zawsze w modelu. Losowe nachylenia w modelu logistycznym wymagają większej liczby obserwacji na grupę niż w modelu liniowym, bo każda obserwacja niesie tylko jeden bit informacji,
- Kowariancja efektów losowych – pełna z korelacją lub niezależne; ta druga jest stabilniejsza przy małej liczbie grup,
- Aproksymacja wiarygodności – Laplace (domyślnie, zawsze dostępna) albo kwadratura G-H z 7, 15 lub 25 węzłami (tylko losowy wyraz wolny). Zmiana aproksymacji zmienia nieznacznie wszystkie liczby w raporcie; do publikacji, przy małych grupach, warto użyć kwadratury z 15–25 węzłami,
- Model pusty i ICC, Test efektów losowych (LR), Efekty losowe grup (BLUP), M. korelacji i kowariancji, Analiza reszt – jak w regresji wielorakiej mieszanej,
- Klasyfikacja i krzywa ROC – czułość, swoistość, udział poprawnych klasyfikacji i AUC dla wskazanej linii odcięcia, osobno dla predykcji warunkowej i populacyjnej,
- Dołącz wykresy – wykres caterpillar efektów losowych grup.

WYNIKI I INTERPRETACJA

Liczności – liczba obserwacji, liczność $Y = 1$, liczba grup, minimum, mediana i maksimum obserwacji na grupę. Przy medianie poniżej 3 obserwacji na grupę aproksymacja Laplace'a jest najmniej dokładna – warto wtedy użyć kwadratury.

Aproksymacja i zbieżność – metoda (Laplace lub Gauss-Hermite z liczbą węzłów), liczba iteracji, zbieżność, dopasowanie osobliwe.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dopasowanie modelu – logarytm wiarygodności, $-2LL$, AIC, AICc, BIC (estymacja ML), R^2 brzegowe i warunkowe oraz – dla porównania – $-2LL$ zwykłej regresji logistycznej (bez efektów losowych) i $-2LL$ modelu pustego.

Test LR efektów losowych (vs regresja logistyczna) – czy struktura grupowa jest potrzebna; wartość p skorygowana dla hipotezy na granicy.

Model pusty – wariancja między grupami τ^2 i **ICC (skala logit)** z przedziałem ufności.

Efekty stałe (warunkowe) – b, błąd, przedział ufności, statystyka z, wartość p oraz **iloraz szans (warunkowy)** z przedziałem ufności. Test jest testem z Walda (bez stopni swobody), tak jak w regresji logistycznej.

Efekty losowe – wariancja i odchylenie standardowe każdego składnika na skali logitowej, testy LR potrzeby składników; ostatni wiersz to stała $\pi^2/3$. Przy więcej niż jednym składniku – macierz korelacji efektów losowych.

Efekty losowe grup (BLUP) – odchylenia grup od populacji na skali logitowej z błędem. Grupa o BLUP $= 0.5$ ma szanse zdarzenia $e^{0.5} \approx 1.65$ razy większe niż grupa przeciętna. Grupy z małą liczbą obserwacji lub bez zdarzeń mają BLUP ściągnięte w stronę zera.

Klasyfikacja i krzywa ROC – dla predykcji warunkowej (z BLUP) i populacyjnej: czułość, swoistość, udział poprawnych, AUC z błędem i przedziałem ufności. Różnica AUC między nimi pokazuje, ile zdolności rozróżniania wnosi znajomość grupy.

Analiza reszt – prawdopodobieństwo przewidywane warunkowe i populacyjne oraz reszty Pearsona dla każdej obserwacji.

WYKONANIE I INTERPRETACJA

Przykład 26.7. (plik stadaCBPP.pqs)
W badaniu zakaźnego zapalenia płuc i opłucnej bydła (CBPP) w Etiopii (Lesnoff i in. 2004[194]) obserwowano 15 stad zebu przez cztery kolejne okresy. Dla każdego zwierzęcia w każdym okresie zapisano, czy pojawił się u niego nowy przypadek choroby. Dane są w układzie długim: Stado (1–15), Okres (1–4) i Zachorowanie (1 – nowy przypadek, 0 – brak); 842 obserwacje, w tym 99 zachorowań, od 26 do 96 obserwacji na stado. Pytamy, czy zapadalność zmienia się w kolejnych okresach epidemii i jak bardzo stada różnią się między sobą.

Stado	Okres	Zachorowanie
1	1	1
1	1	1
1	1	0
1	1	0
1	1	0
1	1	0
1	1	0
1	1	0
1	1	0

Zwierzęta z jednego stada dzielą warunki, kontakty i odporność, więc ich wyniki nie są niezależne – stado jest naturalną zmienną grupującą. Budujemy model: Zachorowanie jako Y z wartością 1 w polu Grupa (1), Stado jako zmienną grupującą, Okres jako zmienną niezależną. Okres jest zmienną jakościową o

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

czterech kategoriach, więc przyciskiem Zm. fikcyjne zamieniamy go na trzy zmienne fikcyjne z okresem 1 jako kategorią odniesienia. Zostawiamy sam losowy wyraz wolny (okres nie ma losowego nachylenia), aproksymację Laplace, włączamy Model pusty i ICC, Test efektów losowych (LR), Efekty losowe grup (BLUP), Klasyfikacja i krzywa ROC oraz Dołącz wykresy.

Liczność	842
Liczność Y = 1 (1)	99
Liczba grup	15
Obserwacji na grupę (min / mediana / max)	26 / 61 / 96
Liczba zmiennych w modelu	3
Aproksymacja wiarygodności	Laplace
Ilość iteracji	292
Zbieżność	osiągnięta
Dopasowanie osobliwe	nie
Log Likelihood	-277,501942
-2LL	555,003883
Liczba parametrów	5
AIC - kryterium Akaikego	565,003883
AICc - skorygowane kryterium Akaikego	565,075653
BIC - bayesowskie kryterium Schwarza	588,682783
R2 brzegowe (efekty stałe)	0,088256
R2 warunkowe (stałe + losowe)	0,189838
Wariancja skali logit (π²/3)	3,289868
Test LR efektów losowych (vs regresja logistyczna)	
-2LL regresji logistycznej (bez efektów losowych)	569,009718
Chi-kwadrat	14,005835
stopnie swobody	1
Wartość p	0,000091
Model pusty	
-2LL modelu pustego	580,613947
Wariancja między grupami	0,659343
ICC (skala logit)	0,166956
-95% CI	0,036797
+95% CI	0,297114

Krok 1 – dopasowanie. 842 obserwacje, 99 zdarzeń, 15 grup (26/61/96 obserwacji na grupę), zbieżność osiągnięta, bez dopasowania osobliwego. $-2LL$ modelu wynosi 555.00 (AIC $= 565.00$), podczas gdy zwykła regresja logistyczna z tymi samymi zmiennymi ma $-2LL = 569.01$. Test LR efektów losowych: $\chi^2 = 14.01$, $df = 1$, $p = 0.00009$ (z korektą granicy) – stada istotnie różnią się zapadalnością i model bez efektu losowego byłby niepoprawny.

Krok 2 – ICC. W modelu pustym wariancja między stadami wynosi 0.659, stąd ICC $= 0.659/(0.659 + 3.29) = 0.167$ (95% CI: 0.037–0.297): około 17% utajonej zmienności skłonności do zachorowania to różnice między stadami.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Efekty stałe (warunkowe)	wsp. b	błąd b	-95% CI	+95% CI	stat. z	wartość p	iloraz szans	-95% CI	+95% CI
w. wolny	-1,398532	0,232472	-1,854169	-0,942895	-6,015916	<0,000001	0,246959	0,156583	0,389498
Okres[2]	-0,992332	0,306642	-1,59334	-0,391324	-3,23612	0,001212	0,370711	0,203246	0,676161
Okres[3]	-1,128672	0,326638	-1,76887	-0,488473	-3,455422	0,000549	0,323463	0,170526	0,613562
Okres[4]	-1,580314	0,427437	-2,418074	-0,742553	-3,697189	0,000218	0,20591	0,089093	0,475897

Krok 3 – efekty stałe. Wyraz wolny $b_0 = -1.399$ odpowiada prawdopodobieństwu zachorowania w okresie 1 w przeciętnym stadzie równemu $1/(1+e^{1.399}) = 0.198$. Warunkowe ilorazy szans dla kolejnych okresów względem okresu 1: okres 2 – $OR = 0.371$ (95% CI: 0.203–0.676, $p = 0.0012$), okres 3 – $OR = 0.324$ (0.171–0.614, $p = 0.0005$), okres 4 – $OR = 0.206$ (0.089–0.476, $p = 0.0002$). W tym samym stadzie szanse nowego zachorowania w okresie 2 są o 63% mniejsze niż w okresie 1, a w okresie 4 – o 79% mniejsze: epidemia wygasa. Słowo „warunkowy” przypomina, że porównujemy okresy wewnątrz stada; iloraz szans uśredniony po wszystkich stadach (ze zwykłej regresji logistycznej) byłby nieco bliższy 1.

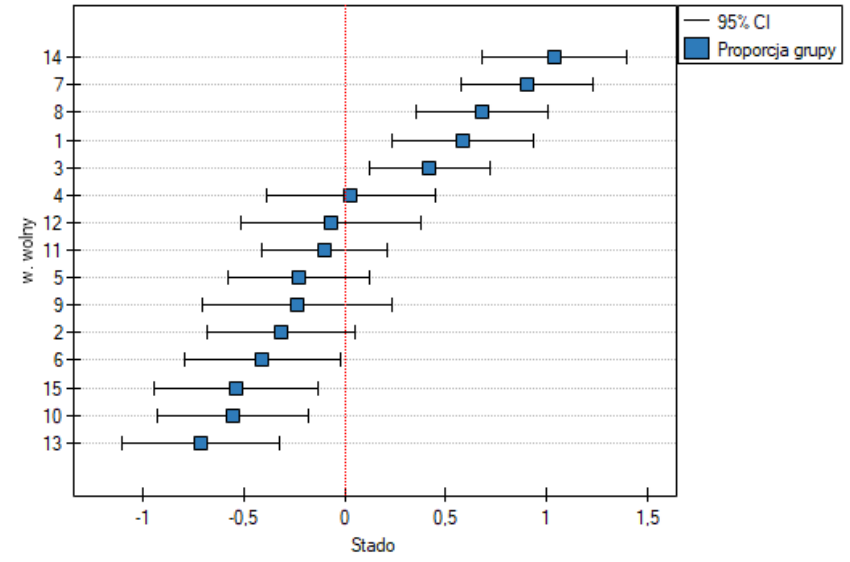
Efekty losowe	składnik	wariancja	odch. std.	LR chi2	stopnie swob	wartość p
Stado	w. wolny	0,4125	0,642262	14,005835	1	0,000091
Skala logit	π²/3	3,289868	1,813799			

Krok 4 – efekty losowe. Wariancja losowego wyrazu wolnego wynosi 0.413, odchylenie standardowe 0.642 na skali logitowej. Stado o poziomie wyższym o jedno odchylenie ma szanse zachorowania $e^{0.642} \approx 1.9$ razy większe niż stado przeciętne, a różnica między stadami skrajnymi (± 2 odchylenia) to $e^{2.57} \approx 13$-krotność szans. R^2 brzegowe $= 0.088$, warunkowe $= 0.190$: okres wyjaśnia niecałe 9% zmienności na skali logitowej, a razem ze stadem – 19%.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Efekty losowe grup (BLUP)

grupa	w. wolny	błąd
1	0,59002	0,348265
2	-0,298897	0,365558
3	0,406255	0,297318
4	0,039277	0,416388
5	-0,190016	0,350399
6	-0,400269	0,379956
7	0,889394	0,326471
8	0,599372	0,321089
9	-0,237655	0,461408
10	-0,540935	0,370679
11	-0,084637	0,309122
12	-0,064817	0,441138
13	-0,689925	0,384816
14	0,970717	0,355401
15	-0,530477	0,397703



Krok 5 – stada nietypowe. BLUP stada 1 wynosi $+0.590$ (błąd 0.348): jego szanse zachorowania są $e^{0.59} \approx 1.8$ razy większe niż przeciętnego stada; w okresie 1 daje to prawdopodobieństwo $1/(1 + e^{1.399-0.590}) = 0.308$ zamiast 0.198. Stado 2 ma BLUP -0.299 (szanse o 26% mniejsze). Wykres caterpillar porządkuje stada od najmniej do najbardziej zagrożonego; stada, których przedział nie obejmuje zera, wyraźnie odbiegają od przeciętnej. Warto pamiętać, że stada z małą liczbą zwierząt mają BLUP ściągnięte w stronę zera.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Klasyfikacja i krzywa ROC (linia odcięcia = 0,5)													
	czułość	-CI	+CI	swoistość	-CI	+CI	udział popra	AUC	SE(AUC)	-CI	+CI	stat. z	p
predykcja warui	0	NA	0,036576	1	0,995047	NA	0,882423	0,757807	0,027128	0,704637	0,810976	8,342252	<0,000001
predykcja popul	0	NA	0,036576	1	0,995047	NA	0,882423	0,677842	0,026898	0,625122	0,730562	5,754696	<0,000001

Krok 6 – klasyfikacja. Tabela klasyfikacji i ROC pokazuje dwie wersje: predykcję warunkową (z BLUP stada) i populacyjną (sam okres). AUC predykcji warunkowej jest wyższe – znajomość stada poprawia rozróżnianie zwierząt, które zachorują, od tych, które nie zachorują. Wersja populacyjna mówi, jak model sprawdzi się w nowym, nieznanym stadzie.

OR warunkowe: okres 2 – 0.37, okres 3 – 0.32, okres 4 – 0.21 (wszystkie $p < 0.002$); SD stad $= 0.64$ (LR $p = 0.00009$), ICC $= 0.17$

Zapadalność maleje w kolejnych okresach epidemii, a stada istotnie różnią się bazowym ryzykiem.

Co zmienia aproksymacja. Ten sam model z kwadraturą G-H, 25 węzłów daje $-2LL = 554.92$, wariancję stad 0.419 zamiast 0.413 i współczynniki różniące się na trzecim miejscu po przecinku (np. okres 4: $b = -1.579$ zamiast -1.580). Przy stadach liczących 26–96 zwierząt aproksymacja Laplace'a jest w pełni wystarczająca; gdyby grupy liczyły po 2–3 obserwacje, różnice byłyby większe i kwadratura stałaby się metodą z wyboru.

DLA ZAINTERESOWANYCH

Estymacja: Wiarygodność brzegowa wymaga scałkowania po efektach losowych, które nie ma postaci zamkniętej. Program stosuje aproksymację Laplace'a (Bates i in. 2015[189]): dla każdej grupy znajduje modę $\hat{u}_j$ penalizowanej log-wiarygodności

$$f_j(u) = -\sum_{i \in j} \left[y_{ij}\eta_{ij} - \ln(1 + e^{\eta_{ij}}) \right] + \frac{1}{2} u^T u, \qquad \eta_{ij} = x_{ij}^T \beta + z_{ij}^T \Lambda u,$$

i przybliża dewiancję przez $\sum_j \left[2 f_j(\hat{u}_j) + \ln\det(I + \Lambda^T Z_j^T W_j Z_j \Lambda) \right]$, gdzie W_j jest macierzą wag $p(1-p)$, a $\Sigma = \Lambda\Lambda^T$. Dla jednego składnika losowego całka może być liczona adaptacyjną kwadraturą Gaussa-Hermite'a wokół mody, z węzłami przesuniętymi i przeskalowanymi według mody i krzywizny każdej grupy. Parametry (θ, β) są optymalizowane łącznie metodą Neldera-Meada ze startem z regresji logistycznej.

Błędy standardowe: z odwrotności hesjanu dewiancji po wszystkich parametrach (θ, β), liczonego numerycznie w optimum, jak w pakiecie lme4.

ICC: $\tau^2/(\tau^2 + \pi^2/3)$ z modelu pustego, z przedziałem ufności metodą delta.

R^2 **Nakagawy:** jak w regresji wielorakiej mieszanej, z $\pi^2/3$ w miejsce wariancji reszt.

Test na granicy: statystyka $2(\ell_1 - \ell_0)$ dla hipotezy o zerowej wariancji ma rozkład będący mieszanką 50:50 rozkładów χ^2_{q-1} i χ^2_q (Self i Liang 1987[198]); dla jednego parametru wartość p z χ^2_1 dzieli się przez dwa.

Efekt populacyjny: przy losowym wyrazie wolnym o wariancji τ^2 współczynnik populacyjny (brzegowy) wynosi w przybliżeniu $\beta/\sqrt{1 + 0.346\,\tau^2}$ (Zeger i in. 1988[200]).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Walidacja i predykcja dla regresji logistycznej mieszanej

KIEDY STOSOWAĆ:

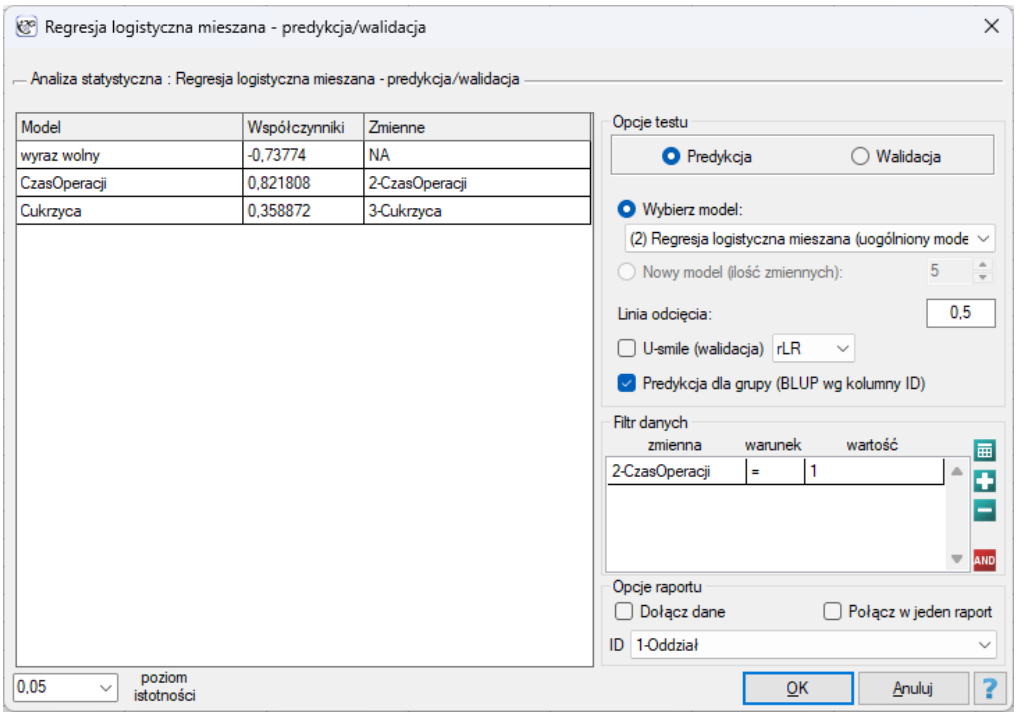
Zbudowany model można wykorzystać do **predykcji** prawdopodobieństwa zdarzenia dla nowych obserwacji i do **walidacji** na zbiorze testowym, jak w walidacji i predykcji dla regresji logistycznej. Tak jak w regresji wielorakiej mieszanej dostępne są dwa rodzaje predykcji:

- **populacyjna** – z samych efektów stałych; prawdopodobieństwo dla przeciętnej, nieznanej grupy,
- **dla grupy** – z efektem losowym (BLUP) grupy znanej modelowi; prawdopodobieństwo dla tego konkretnego pacjenta, ośrodka, stada.

Predykcja wystąpienia zdarzenia (1) lub jego braku (0) polega na porównaniu przewidywanego prawdopodobieństwa z **punktem odcięcia** (domyślnie 0.5).

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja logistyczna mieszana – predykcja/walidacja.



Wskazujemy model zapisany w programie (raport regresji logistycznej mieszanej przypisany do arkusza) – ręczne wpisanie współczynników nie jest możliwe, bo potrzebne są efekty losowe grup. Następnie wskazujemy kolumny z nowymi wartościami zmiennych, ewentualnie obserwowane Y (walidacja) oraz opcję Predykcja dla grupy (BLUP wg kolumny ID) z kolumną identyfikatora. Dla identyfikatorów nieznanych modelowi predykcja jest populacyjna.

WYNIKI I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Raport podaje rodzaj predykcji, liczbę obserwacji, dla których użyto BLUP grupy, oraz dla każdej obserwacji przewidywane prawdopodobieństwo i przewidywaną klasę. W trybie walidacji dodatkowo klasyfikację (czułość, swoistość, udział poprawnych) i AUC na zbiorze testowym.

Przykład 26.8. c.d. przykładu 26.9 (plik oddziałyZakażenia.pqs)
Na podstawie modelu zakażeń miejsca operowanego (czas operacji i cukrzyca jako efekty stałe, losowy wyraz wolny oddziału i losowe nachylenie czasu operacji) chcemy oszacować ryzyko zakażenia u pacjenta z cukrzycą, którego operacja trwała o jedno odchylenie standardowe dłużej niż przeciętna (CzasOperacji $= 1$, Cukrzyca $= 1$), gdyby był operowany na oddziale 1, na oddziale 2 albo na nowym oddziale, którego model nie zna.

Najpierw budujemy i zapisujemy model w oknie Regresja logistyczna mieszana: Zakażenie jako Y (Grupa $(1) = 1$), Oddział jako zmienną grupującą, CzasOperacji i Cukrzyca jako X, CzasOperacji na liście losowych nachyleń, kowariancja pełna, z włączoną opcją Efekty losowe grup (BLUP). Następnie dopisujemy do arkusza trzy wiersze z wartościami 1 i 1 w kolumnach CzasOperacji i Cukrzyca oraz identyfikatorami oddziałów 1, 2 i 99, ustawiamy filtr na te wiersze i w oknie predykcji wskazujemy zapisany model oraz opcję Predykcja dla grupy (BLUP wg kolumny ID) z kolumną Oddział.

Współczynniki modelu	
wyraz wolny	-0,73774
CzasOperacji	0,821808
Cukrzyca	0,358872

Predykcja				
	Pred. prawd	-95% CI	+95% CI	Pred. Y
1 (BLUP)	0,292478	0,101743	0,601388	0
2 (BLUP)	0,842557	0,631994	0,943427	1
99	0,608959	0,496041	0,711302	1

Efekty stałe modelu: wyraz wolny -0.738, czas operacji 0.822, cukrzyca 0.359; BLUP oddziału 1: wyraz wolny -1.145, nachylenie -0.182; oddziału 2: $+1.193$ i $+0.041$. Dla nowego oddziału (99) predykcja jest populacyjna: logit $= -0.738 + 0.822 + 0.359 = 0.443$, prawdopodobieństwo 0.61. Dla oddziału 1 program dodaje jego BLUP wyrazu wolnego i nachylenia (pomnożone przez czas operacji): logit $= 0.443 - 1.145 - 0.182 \cdot 1 = -0.884$, prawdopodobieństwo 0.29. Dla oddziału 2: logit $= 0.443 + 1.193 + 0.041 = 1.677$, prawdopodobieństwo 0.84. Ten sam pacjent ma więc ryzyko zakażenia od 29% do 84% w zależności od oddziału, a predykcja populacyjna (61%) jest tylko środkiem tego zakresu. Przy punkcie odcięcia 0.5 pacjent na oddziale 1 zostałby zaklasyfikowany jako „bez zakażenia”, na oddziałach 2 i nowym – jako „zakażenie”. W trybie walidacji (ze wskazaną kolumną Zakażenie) na zbiorze testowym raport podałby dodatkowo czułość, swoistość i AUC.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Porównywanie modeli regresji logistycznej mieszanej

KIEDY STOSOWAĆ:

Porównywanie modeli służy do wyboru zestawu zmiennych niezależnych (efektów stałych), tak jak w porównywaniu modeli regresji logistycznej: porównujemy model pełny z modelem zredukowanym lub korzystamy z doboru automatycznego (wsteczny, w przód, krokowy, podzbiory). **Struktura losowa jest wspólna dla wszystkich porównywanych modeli**: losowy wyraz wolny, losowe nachylenia i kowariancja ustawione w oknie obowiązują w każdym modelu, a zmienna z losowym nachyleniem pozostaje w każdym modelu jako efekt stały. Dobór dotyczy wyłącznie efektów stałych; o tym, czy losowe nachylenie jest potrzebne, decyduje test LR w raporcie regresji logistycznej mieszanej.

DEFINICJE MIAR

Modele porównuje się kryteriami informacyjnymi AIC, AICc, BIC (im mniejsze, tym lepiej), pseudo-R^2 oraz R^2 brzegowym i warunkowym, a modele zagnieżdżone – testem ilorazu wiarygodności: $2(\ell_P - \ell_Z)$ (ℓ_P, ℓ_Z – logarytm wiarygodności modelu pełnego i zredukowanego) ma rozkład χ^2 o liczbie stopni swobody równej liczbie usuniętych zmiennych. Wiarygodność jest zawsze aproksymowana tą samą metodą (Laplace lub kwadratura) we wszystkich modelach, więc jest porównywalna.

Ocena modeli poza zbiorem uczącym

Kryteria informacyjne i test LR są wyznaczane na tych samych danych, na których dopasowano modele, przez co ocena modeli jest zbyt optymistyczna. Tak jak w porównywaniu modeli regresji logistycznej, w oknie analizy można więc włączyć ocenę modeli na obserwacjach, które nie brały udziału w ich dopasowaniu (opcja Walidacja): **walidację krzyżową k-krotną** albo **losowy zbiór testowy**, z podziałem wyznaczonym przez Ziarno. Dla każdego modelu raport podaje wtedy średnią dewiancję ($-2 \ln L/n$; to ona jest błędem walidacji), AUC i wskaźnik Briera, a w bloku doboru modelu – błąd modelu zerowego i najlepszy model wg walidacji; jako kryterium doboru automatycznego można też wybrać błąd walidacji. W modelach mieszanych obowiązują dwie dodatkowe zasady:

- **dzielone są całe grupy** – wszystkie obserwacje jednej grupy (np. jednego oddziału) trafiają do tej samej części danych. Obserwacje tej samej grupy są ze sobą skorelowane, więc rozdzielenie ich między część uczącą i ocenianą zaniżałoby błąd. Liczba podziałów k jest ograniczana do liczby grup, a udział zbioru testowego oznacza procent grup (podział nie jest tu warstwowany względem zmiennej zależnej);
- **przewidywanie dla nowej grupy opiera się wyłącznie na efektach stałych** – efekt losowy grupy, której nie było w zbiorze uczącym, nie jest znany, dlatego przyjmuje się jego wartość oczekiwaną, czyli zero. Miary walidacji mówią więc, jak dobrze efekty stałe modelu przewidują zdarzenie w nowej grupie (np. na nowym oddziale).

Modelem zerowym jest tu model z wyrazem wolnym, wspólną strukturą losową i zmiennymi, którym nadano losowe nachylenie (pozostają one w każdym modelu); jest on dopasowywany i oceniany tak samo jak pozostałe modele.

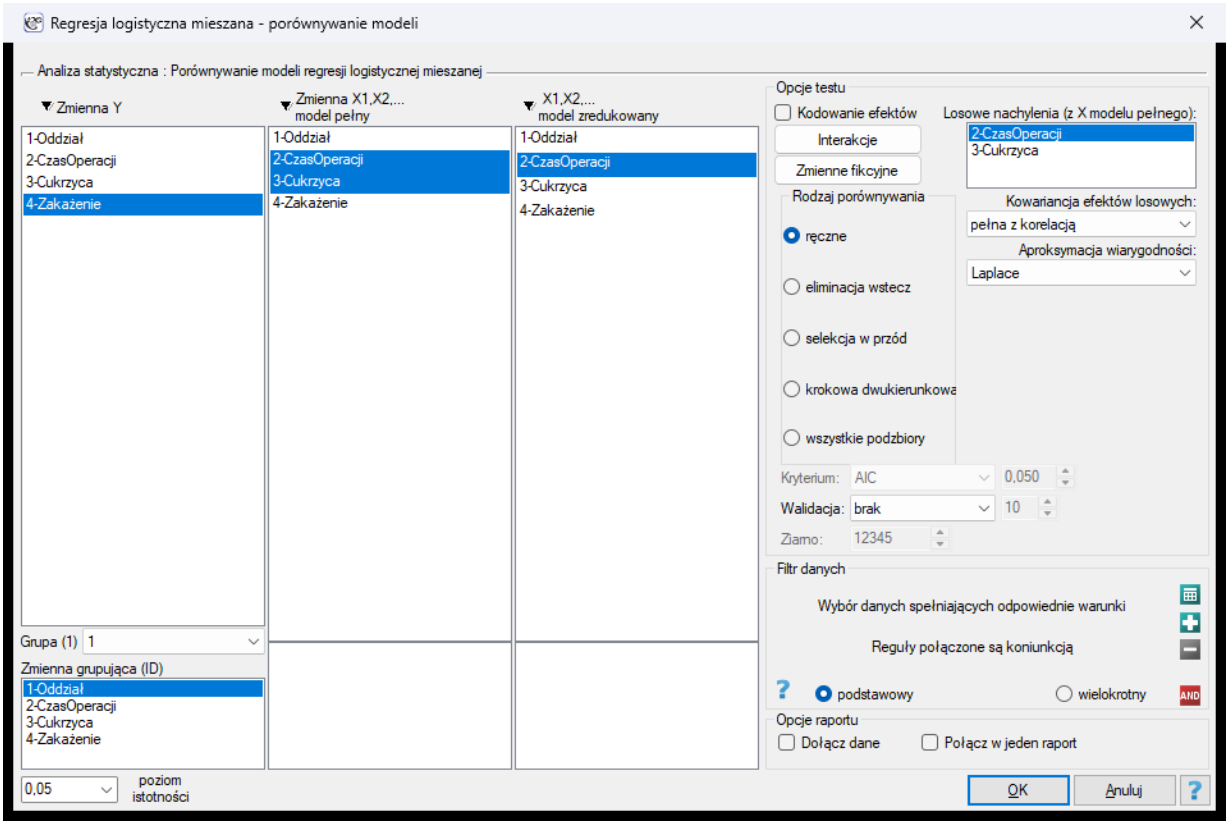
Uwaga! Współczynniki modelu mieszanego są warunkowe – opisują efekt zmiennych wewnątrz grupy. Prawdopodobieństwo wyznaczone z samych efektów stałych odpowiada grupie „przeciętnej” (o efekcie losowym równym zero) i przy dużej zmienności między grupami jest bardziej skrajne niż częstość zdarzenia uśredniona po grupach. Miary zależne od kalibracji (średnia dewiancja, wskaźnik Briera) są przez

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

to mniej korzystne, niż gdyby efekt grupy był znany; wszystkie modele są jednak oceniane w ten sam sposób, więc ich porównanie pozostaje miarodajne.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja logistyczna mieszana – porównywanie modeli.



Wskazujemy zmienną zależną Y z wartością zdarzenia (Grupa (1)), zmienne modelu pełnego i zredukowanego (lub kryterium doboru) oraz opcje modelu mieszanego: Zmienna grupująca (ID), Losowe nachylenia (z X modelu pełnego), Kowariancja efektów losowych i Aproksymacja wiarygodności – wszystkie wspólne dla porównywanych modeli. Pod listą metod doboru znajdują się pola: Kryterium (AIC, AICc, BIC, p (test LR / F) z progiem albo błąd walidacji), Walidacja (brak, krzyżowa, k = z liczbą podziałów albo zbiór testowy, % = z udziałem zbioru testowego w procentach grup) i Ziarno losowania podziału (to samo ziarno oznacza ten sam podział).

WYNIKI I INTERPRETACJA

Raport zawiera ramkę z opisem wspólnej struktury losowej, tabelę **Kryteria informacyjne i testy LR** z $-2LL$, AIC, AICc, BIC, R^2 brzegowym i warunkowym każdego modelu i testem LR między kolejnymi modelami oraz tabele współczynników modeli z warunkowymi ilorazami szans i wariancją efektów losowych każdego modelu. Przy włączonej walidacji raport zawiera dodatkowo blok **Ocena poza zbiorem uczącym** (sposób podziału, ziarno, jednostka podziału, błąd modelu zerowego, najlepszy model wg walidacji), a przy każdym modelu – średnią dewiancję, AUC i wskaźnik Briera.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PRZYKŁAD 26.9. (plik oddziałyZakażenia.pqs)
Dane symulowane: w 40 oddziałach chirurgicznych obserwowano po 25 pacjentów po operacji. Dla każdego pacjenta zapisano, czy wystąpiło zakażenie miejsca operowanego (Zakażenie: 1 – tak), standaryzowany czas trwania operacji (CzasOperacji: 0 oznacza czas przeciętny, 1 – o jedno odchylenie standardowe dłuższy) oraz obecność cukrzycy (Cukrzyca: 1 – tak). Łącznie 1000 pacjentów, 404 zakażenia. Pacjenci tego samego oddziału dzielą personel, procedury i warunki sali, więc oddział jest zmienną grupującą. Ponieważ wpływ czasu operacji może zależeć od oddziału (na jednych oddziałach długie operacje są rutyną, na innych sygnałem powikłań), czas operacji dostaje losowe nachylenie. Pytamy, czy cukrzyca wnosi coś ponad czas operacji.

Oddział	CzasOperacji	Cukrzyca	Zakażenie
1	0,343	0	0
1	0,004	0	0
1	0,029	1	0
1	-0,393	0	0
1	-0,793	0	0
1	-0,312	0	0
1	-0,346	0	0
1	-0,305	1	0
1	-1,786	0	0
1	0,587	1	0
1	1,636	0	0
1	-0,645	1	0
1	0,619	1	0
1	0,236	0	1
1	0,847	0	0
1	-0,574	0	0
1	1,118	1	0
1	-1,540	0	0
1	0,438	1	0

Ustawiamy Zakażenie jako Y z wartością 1 w polu Grupa (1), Oddział jako zmienną grupującą, model pełny: CzasOperacji i Cukrzyca, model zredukowany: sam CzasOperacji. Na liście Losowe nachylenia (z X modelu pełnego) zaznaczamy CzasOperacji (zmienna ta pozostanie w obu modelach), kowariancja pełna z korelacją, aproksymacja Laplace.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Liczba zmiennych w modelu 1	2
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood (ML)	1121,206672
AIC - kryterium Akaikego	1133,206672
AICc - skorygowane kryterium Akaikego	1133,291264
BIC - bayesowskie kryterium Schwarza	1162,653204
R2 brzegowe (efekty stałe)	0,125476
R2 warunkowe (stałe + losowe)	0,395786
Wariancja: w. wolny	1,324926
Wariancja: CzasOperacji	0,152284
Wariancja: skali logit (pi^2/3)	3,289868
Liczba zmiennych w modelu 2	1
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood (ML)	1126,474737
AIC - kryterium Akaikego	1136,474737
AICc - skorygowane kryterium Akaikego	1136,535099
BIC - bayesowskie kryterium Schwarza	1161,013513
R2 brzegowe (efekty stałe)	0,11888
R2 warunkowe (stałe + losowe)	0,392954
Wariancja: w. wolny	1,330535
Wariancja: CzasOperacji	0,159713
Wariancja: skali logit (pi^2/3)	3,289868
Chi-kwadrat - porównanie modeli Model 1 vs Model 2	5,268065
Stopnie swobody	1
Wartość p	0,02172

Model 1

	wsp. b	błąd b	-95% CI	+95% CI	stat. z	wartość p	OR (warunk	-95% CI	+95% CI
w. wolny	-0,73774	0,217101	-1,163251	-0,312229	-3,398135	0,000678	0,478193	0,312469	0,731814
CzasOperacji	0,821808	0,113309	0,599728	1,043889	7,25284	<0,000001	2,27461	1,821623	2,840241
Cukrzyca	0,358872	0,156732	0,051684	0,666061	2,289726	0,022037	1,431714	1,053043	1,946554

Model 2

	wsp. b	błąd b	-95% CI	+95% CI	stat. z	wartość p	OR (warunk	-95% CI	+95% CI
w. wolny	-0,55852	0,202124	-0,954676	-0,162365	-2,763258	0,005723	0,572055	0,384937	0,850131
CzasOperacji	0,82035	0,113983	0,596946	1,043753	7,197097	<0,000001	2,271294	1,816563	2,839856

Model pełny ma $-2LL = 1121.21$, AIC $= 1133.21$, BIC $= 1162.65$; model zredukowany $-2LL = 1126.47$, AIC $= 1136.47$, BIC $= 1161.01$. Test LR: $\chi^2 = 1126.47 - 1121.21 = 5.27$, $df = 1$, $p = 0.022$ – usunięcie cukrzycy istotnie pogarsza dopasowanie, AIC wskazuje model pełny, a BIC (surowiej karzący za parametry) – model zredukowany, ale różnica jest niewielka (1.6). W tabeli współczynników modelu pełnego cukrzyca ma warunkowy $OR = 1.43$ (95% CI: 1.05–1.95, $p = 0.022$): u pacjentów tego samego

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

oddziału, przy tym samym czasie operacji, cukrzyca zwiększa szanse zakażenia o 43%. Czas operacji: $OR = 2.27$ (1.82–2.84) na jedno odchylenie standardowe. Zostajemy przy modelu pełnym – wynik testu LR jest zgodny z testem Walda, a efekt cukrzycy ma sens kliniczny.

Struktura losowa. Oba modele mają wariancję losowego wyrazu wolnego około 1.33 (odchylenie 1.15 na skali logitowej – oddziały bardzo różnią się bazowym ryzykiem) i wariancję losowego nachylenia około 0.15 (odchylenie 0.39: efekt czasu operacji waha się między oddziałami od około $0.82 - 0.78 = 0.04$ do $0.82 + 0.78 = 1.60$ na skali logitowej). Test LR nachylenia w raporcie regresji logistycznej mieszanej daje $p = 0.095$, więc dowód na różne nachylenia jest słaby; model z samym losowym wyrazem wolnym ma AIC $= 1133.11$, praktycznie równe modelowi z nachyleniem. Wersja z kowariancją niezależne (korelacja i tak wynosi -0.06) ma najniższe AIC $= 1131.24$ i byłaby najoszczędniejszym wyborem.

test LR pełny vs bez cukrzycy: $\chi^2 = 5.27$, $df = 1$, $p = 0.022$; AIC 1133.21 vs 1136.47

Cukrzyca pozostaje w modelu ($OR = 1.43$); czas operacji jest najsilniejszym czynnikiem ($OR = 2.27$ na odchylenie standardowe).

Ocena poza zbiorem uczącym. Sprawdzimy jeszcze, czy cukrzyca poprawia przewidywanie zakażeń na oddziałach, które nie brały udziału w dopasowaniu modelu. To samo porównanie powtarzamy, wybierając w polu Walidacja opcję krzyżowa, k = z 10 podziałami (ziarno 12345). Dzielonych jest 40 oddziałów – w każdym z 10 kroków modele są dopasowywane na 36 oddziałach, a oceniane na 4 pozostałych.

Ocena poza zbiorem uczącym	walidacja krzyżowa
Liczba podziałów (k)	10
Ziarno losowania podziału	12345
Jednostka podziału	cała grupa
Model zerowy: średnia dewiancja	1,28897
Model zerowy: wskaźnik Briera	0,224097
Najlepszy model wg walidacji (średnia dewiancja)	Model 1
Predykcja nowych grup: tylko efekty stałe	

	zbiór uczący	walidacja krzyżowa ($k = 10$)		
model	AIC	śr. dewiancja	AUC	Brier
Model 1 (czas operacji, cukrzyca)	1133.21	1.2843	0.6677	0.2227
Model 2 (czas operacji)	1136.47	1.2890	0.6657	0.2241

Czas operacji ma losowe nachylenie, więc pozostaje w każdym modelu – model zerowy jest tu tym samym modelem co Model 2 i stąd jednakowe wartości w raporcie. Model z cukrzycą ma mniejszą średnią dewiancję (1.2843 wobec 1.2890), nieco większe AUC (0.6677 wobec 0.6657) i mniejszy wskaźnik Briera (0.2227 wobec 0.2241), jest więc wskazany jako najlepszy według walidacji. Poprawa jest jednak niewielka: cukrzyca jest czynnikiem istotnym statystycznie, ale przewidywanie zakażeń na nowym oddziale poprawia w skromnym stopniu – o ryzyku decyduje przede wszystkim czas operacji i sam oddział, którego efektu losowego dla nowego oddziału nie znamy.

walidacja krzyżowa ($k = 10$, całe oddziały): średnia dewiancja 1.2843 (z cukrzycą) i 1.2890 (bez cukrzycy); AUC 0.668 i 0.666

Najlepszym modelem według walidacji jest model pełny, choć zysk z dodania cukrzycy w przewidywaniu dla nowych oddziałów jest niewielki.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

26.3 REGRESJA PH COXA Z EFEKTEM LOSOWYM

Regresja PH Coxa z efektem losowym (frailty)

KIEDY STOSOWAĆ:

Model Coxa z efektem losowym jest rozszerzeniem regresji proporcjonalnych hazardów Coxa na dane, w których obserwacje tworzą grupy: pacjenci leczeni w tym samym ośrodku, zwierzęta z jednego miotu, zdarzenia nawracające u tego samego pacjenta (kolejne infekcje, kolejne hospitalizacje), oczy lub nerki tego samego chorego (zob. wprowadzenie do modeli mieszanych). Czasy przeżycia w obrębie grupy są ze sobą powiązane, bo grupa dzieli nieobserwowane czynniki ryzyka: pacjent ma swoją odporność, ośrodek – swoją jakość opieki. Zwykła regresja Coxa tę zależność pomija, przez co zaniża błędy standardowe współczynników i zawyża istotność.

Efekt losowy, zwany w analizie przeżycia *frailty* (kruchością), to nieobserwowany czynnik w_j wspólny dla całej grupy j, który mnoży hazard każdej jej obserwacji:

$$h_{ij}(t) = w_j \, h_0(t) \, \exp(\beta_1 X_{1ij} + \ldots + \beta_k X_{kij}),$$

gdzie:
$h_{ij}(t)$ – hazard i-tej obserwacji w grupie j,
$h_0(t)$ – hazard bazowy,
$\beta_1, \ldots, \beta_k$ – współczynniki (efekty stałe),
w_j – frailty grupy j; zakłada się rozkład gamma o średniej 1 i wariancji θ (dzielone frailty gamma, ang. *shared gamma frailty*).

Grupy o $w_j > 1$ mają wyższe ryzyko, niż wynika z ich zmiennych, grupy o $w_j < 1$ – niższe. Wariancja θ mówi, jak bardzo grupy różnią się ryzykiem; $\theta = 0$ oznacza brak zależności w grupach i model sprowadza się do zwykłej regresji Coxa. Frailty jest odpowiednikiem losowego wyrazu wolnego z regresji wielorakiej mieszanej, tyle że działa multiplikatywnie na hazard; losowe nachylenia nie są w tym modelu dostępne.

WARUNKI STOSOWANIA

- proporcjonalność hazardów, jak w regresji Coxa,
- grupy niezależne od siebie; zależność występuje tylko wewnątrz grup,
- co najmniej 3 grupy, a do wiarygodnego oszacowania θ zwykle kilkanaście lub więcej,
- wystarczająca liczba zdarzeń (obserwacji kompletnych); grupy bez zdarzeń wnoszą mało informacji o θ,
- brak silnej współliniowości zmiennych niezależnych.

Uwaga! Dane muszą być przygotowane w układzie długim: każdy wiersz arkusza to jedna obserwacja (czas, status, zmienne niezależne), a osobna kolumna wskazuje grupę (np. identyfikator ośrodka lub pacjenta). Przy zdarzeniach nawracających każdy epizod to osobny wiersz z czasem liczonym od początku tego epizodu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DEFINICJE MIAR

Wariancja frailty θ – miara zróżnicowania ryzyka między grupami. Frailty ma średnią 1, więc przy $\theta = 0.4$ odchylenie standardowe w_j wynosi $\sqrt{0.4} \approx 0.63$: typowa grupa ma hazard o 60% wyższy lub niższy, niż wynikałoby z jej zmiennych. Błąd θ jest orientacyjny (z profilu wiarygodności) i nie służy do testowania – do tego jest test LR.

τ **Kendalla** – miara zależności czasów przeżycia dwóch obserwacji z tej samej grupy, wynikająca z rozkładu gamma:

$$\tau = \frac{\theta}{\theta + 2}.$$

$\tau = 0$ to niezależność; $\tau = 0.17$ oznacza umiarkowaną zgodność: obserwacja z grupy, w której inna obserwacja przeżyła długo, sama ma większe szanse przeżyć długo.

Test LR: $\theta = 0$ – test ilorazu wiarygodności porównujący model z frailty ze zwykłą regresją Coxa. Hipoteza leży na granicy przestrzeni parametrów, więc wartość p z rozkładu χ^2_1 jest dzielona przez dwa (korekta granicy), o czym raport informuje.

Warunkowy hazard ratio $HR = e^b$ – ile razy zmienia się hazard, gdy X rośnie o jednostkę **w obrębie tej samej grupy** (przy tym samym frailty). Populacyjny HR ze zwykłej regresji Coxa jest uśredniony po grupach i co do wartości bezwzględnej bliższy 1; nie należy porównywać ich wprost.

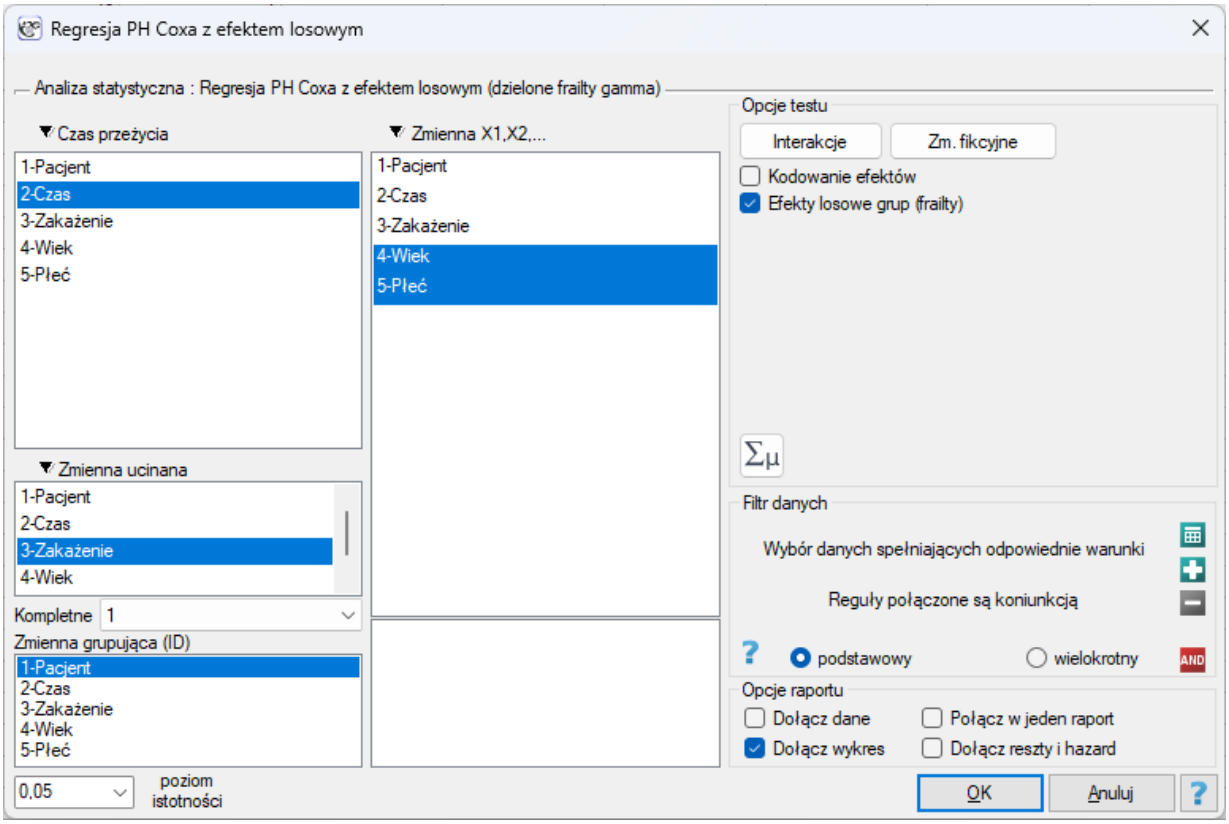
Efekty losowe grup – oszacowane w_j dla każdej grupy oraz log w_j z błędem. Grupy z małą liczbą zdarzeń mają w_j ściągnięte w stronę 1 (nie wiemy o nich dość, by uznać je za nietypowe). Nie testuje się ich istotności; służą do wskazania grup o podwyższonym lub obniżonym ryzyku.

Zintegrowana wiarygodność – logarytm wiarygodności modelu po scałkowaniu po frailty; na jej podstawie liczone są kryteria informacyjne i testy LR (zarówno test $\theta = 0$, jak i test łącznego wpływu zmiennych niezależnych).

JAK USTAWIĆ ANALIZĘ

Statystyka $\rightarrow$ Modele mieszane $\rightarrow$ Regresja PH Coxa z efektem losowym.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Okno ma układ okna regresji Coxa. Wskazujemy Czas przeżycia, Zmienną ucinaną z wartością oznaczającą zdarzenie w polu Kompletne oraz Zmienne niezależne X1, X2, ...; zmienne fikcyjne i interakcje przygotowuje się tak samo jak w regresji Coxa. Opcje modelu z efektem losowym:

- Zmienna grupująca (ID) – kolumna wskazująca grupę obserwacji (ośrodek, pacjent, miot); liczbowa lub tekstowa. Jest obowiązkowa – bez niej model nie różni się od regresji Coxa,
- Efekty losowe grup (frailty) – dołącza do raportu tabelę oszacowanych w_j dla każdej grupy; przy setkach grup tabela jest długa,
- Dołącz reszty i hazard – reszty martyngałowe i dewiancyjne uwzględniające frailty oraz hazard wynikowy $w_j H_i$ każdej obserwacji; służą do oceny dopasowania i wskazania obserwacji odstających,
- Dołącz wykresy – krzywe przeżycia grup (frailty).

Wiązania czasów są rozstrzygane metodą Breslowa, tak jak w regresji Coxa, więc przy $\theta = 0$ wyniki obu analiz są identyczne.

WYNIKI I INTERPRETACJA

Liczności – liczba obserwacji, kompletnych (zdarzeń) i uciętych oraz liczba grup. Raport przypomina też warunek liczności $n \geq 10k/p$ znany z regresji Coxa.

Dopasowanie modelu – logarytm zintegrowanej wiarygodności modelu, ta sama wielkość dla modelu z samym efektem losowym (bez zmiennych) oraz test ilorazu wiarygodności dla łącznego wpływu zmiennych niezależnych. Kryteria informacyjne AIC, AICc, BIC i pseudo-R^2 jak w regresji Coxa.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Efekt losowy (frailty gamma) – zmienna grupująca, liczba grup, wariancja frailty θ z orientacyjnym błędem (profil), τ Kendalla oraz **test LR:** $\theta = 0$ **(vs regresja Coxa)** z korektą granicy. Istotny wynik oznacza, że struktury grupowej nie wolno pominąć. Podana jest też liczba iteracji EM.

Współczynniki – b, błąd, przedział ufności, statystyka Walda, wartość p oraz **hazard ratio (warunkowy)** $HR = e^b$ z przedziałem ufności. Błędy standardowe pochodzą z penalizowanej macierzy informacji i uwzględniają obecność efektu losowego, dlatego są większe niż w zwykłej regresji Coxa.

Funkcje bazowe – skumulowany hazard i przeżycie dla grupy przeciętnej ($w_j = 1$): przy średnich wartościach zmiennych oraz bazowe, jak w regresji Coxa.

Efekty losowe grup (frailty gamma) – dla każdej grupy: w, log w z błędem, liczba zdarzeń, liczność i skumulowany hazard grupy. Grupa z $w = 1.4$ ma hazard o 40% wyższy niż przeciętna grupa o tych samych zmiennych; grupa z $w = 0.6$ – o 40% niższy. Grupy z wieloma zdarzeniami przy małym skumulowanym hazardzie oczekiwanym dostają $w > 1$, grupy bez zdarzeń mimo długiej obserwacji – $w < 1$.

Wykresy – Krzywe przeżycia grup (frailty): krzywa przeżycia dla przeciętnej grupy ($w = 1$) oraz dla wybranych grup z ich frailty; im wyższe w, tym niżej leży krzywa.

WYKONANIE I INTERPRETACJA

Przykład 26.10. (plik nerkiCewnik.pqs)
W badaniu McGilchrista i Aisbett (1991)[195] obserwowano 38 pacjentów dializowanych z założonym cewnikiem. Rejestrowano czas (w dniach) do zakażenia w miejscu wprowadzenia cewnika; po wyleczeniu zakładano cewnik ponownie, więc każdy pacjent ma dwa epizody. Cewnik bywał usuwany z innych przyczyn – takie obserwacje są ucięte. Dane w układzie długim: Pacjent (1–38), Czas (dni), Zakażenie (1 – zakażenie, 0 – obserwacja ucięta), Wiek (lata) i Płeć (0 – mężczyzna, 1 – kobieta). Łącznie 76 obserwacji, 58 zakażeń. Pytamy, czy wiek i płeć wpływają na ryzyko zakażenia i czy pacjenci różnią się indywidualną podatnością.

Pacjent	Czas	Zakażenie	Wiek	Płeć
1	8	1	28	0
1	16	1	28	0
2	23	1	48	1
2	13	0	48	1
3	22	1	32	0
3	28	1	32	0
4	447	1	31	1
4	318	1	32	1
5	30	1	10	0
5	12	1	10	0
6	24	1	16	1
6	245	1	17	1

Dwa epizody tego samego pacjenta nie są niezależne – pacjent ma swoją odporność, higienę, choroby współistniejące. Gdybyśmy policzyli zwykłą regresję Coxa, potraktowalibyśmy 76 epizodów jak 76 różnych osób. Budujemy więc model z efektem losowym: Czas jako czas przeżycia, Zakażenie jako zmienną ucinaną z wartością 1 jako Kompletne, Wiek i Płeć jako zmienne niezależne, Pacjent jako zmienną grupującą. Włączamy Efekty losowe grup (frailty) i Dołącz wykresy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Model	Cox z efektem losowym (fra
Założenie liczności próby n ≥ 10k/p	Nie
Liczba zmiennych w modelu	2
Ilość iteracji dla zbieżności	96
Kryterium zbieżności zostało osiągnięte	
Liczność	76
Kompletne (1)	58
Ucięte (0)	18
Test ilorazu wiarygodności	
Log Likelihood	-182,053359
-2 Log Likelihood	364,106718
Log Likelihood (w. wolny)	-187,945526
-2 Log Likelihood (w. wolny)	375,891051
Statystyka Chi-kwadrat	11,784333
Stopnie swobody	2
Wartość p	0,002761
AIC - kryterium Akaikego	370,106718
AICc - skorygowane kryterium Akaikego	370,551163
BIC - bayesowskie kryterium Schwarza	376,288047
Pseudo R2 (McFadden)	0,03135
R2 (Nagelkerke)	0,184149
R2 (Coxa-Snella)	0,183867
Efekt losowy (frailty gamma)	
Zmienna grupująca	Pacjent
Liczba grup	38

Efekt losowy (frailty gamma)	
Zmienna grupująca	Pacjent
Liczba grup	38
Wariancja frailty (theta)	0,397184
Błąd theta (profil)	0,234386
tau Kendalla (zależność w grupie)	0,165688
Test LR: theta = 0 (vs regresja Coxa)	
Statystyka Chi-kwadrat	5,207469
Wartość p (korekta granicy)	0,011245
Log Likelihood (Cox bez efektu losowego)	-184,657094
Iteracje EM	42

Krok 1 – czy pacjenci różnią się podatnością. Wariancja frailty $\theta = 0.397$ (błąd 0.234), τ Kendalla $= 0.166$. Test LR $\theta = 0$: $\chi^2 = 5.21$, $p = 0.011$ (z korektą granicy) – zwykła regresja Coxa nie jest dla tych danych poprawna, pacjenci istotnie różnią się nieobserwowaną podatnością na zakażenie. Odchylenie standardowe frailty $\sqrt{0.397} \approx 0.63$: typowy pacjent ma hazard zakażenia o około 60% wyższy lub niższy, niż wynika z jego wieku i płci.

Krok 2 – dopasowanie. Zintegrowana log-wiarygodność modelu wynosi -182.05, modelu z samym frailty (bez zmiennych) -187.95; test LR łącznego wpływu zmiennych: $\chi^2 = 11.78$, $df = 2$, $p = 0.0028$. AIC $= 370.11$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

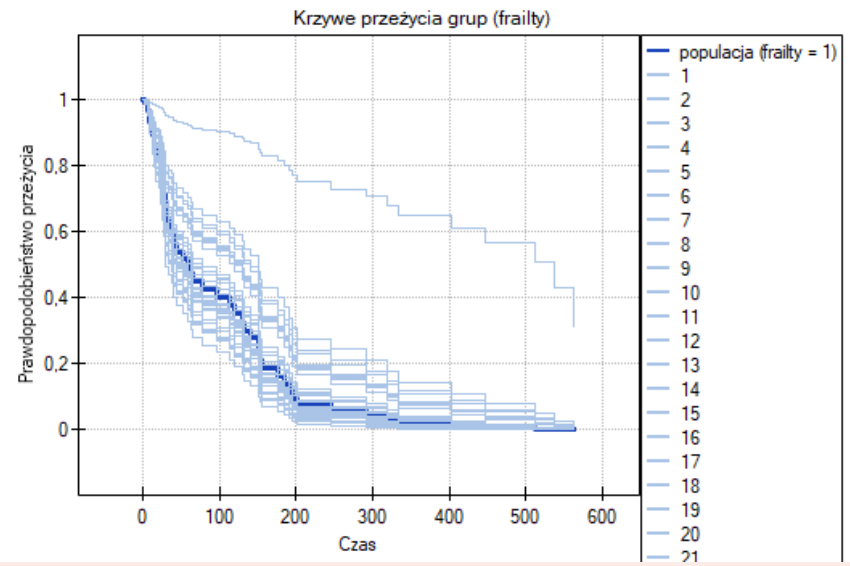
Model									
	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	Hazard ratio	-95% CI	+95% CI
Wiek	0,005462	0,01158	-0,017234	0,028158	0,222456	0,637175	1,005477	0,982913	1,028558
Płeć	-1,55613	0,444803	-2,427928	-0,684332	12,239291	0,000468	0,210951	0,088219	0,504427

Krok 3 – współczynniki. Wiek: $b = 0.0055$, $HR = 1.006$ (95% CI: 0.983–1.029), $p = 0.637$ – wiek nie wpływa na ryzyko zakażenia. Płeć: $b = -1.556$ (błąd 0.445), $HR = 0.211$ (95% CI: 0.088–0.504), $p = 0.0005$. Hazard zakażenia u kobiet jest o 79% niższy niż u mężczyzn – przy porównaniu w obrębie tej samej „podatności", czyli warunkowo. Błąd współczynnika płci (0.445) jest większy niż dałaby zwykła regresja Coxa, bo model uwzględnia, że informacja z dwóch epizodów tego samego pacjenta częściowo się powtarza.

Efekty loso

grupa	w	log(w)	błąd log(w)	zdarzenia	n	skum. hazar
1	1,435732	0,361675	0,478199	2	2	0,628911
2	1,282004	0,248424	0,533957	1	2	0,226203
3	1,102199	0,097307	0,493495	2	2	1,581103
4	0,574402	-0,554426	0,51423	2	2	5,347371
5	1,145763	0,136071	0,506425	2	2	1,42526

Krok 4 – pacjenci nietypowi. Tabela efektów losowych grup: pacjent 1 ma $w = 1.44$ ($\log w = 0.36$, błąd 0.48) – dwa zakażenia przy niewielkim oczekiwanym skumulowanym hazardzie (0.63) oznaczają podatność o 44% wyższą od przeciętnej. Pacjent 4 ma $w = 0.57$: również dwa zakażenia, ale przy dużym skumulowanym hazardzie (5.35), czyli po długiej obserwacji – jego podatność jest o 43% niższa od przeciętnej. Błędy $\log w$ (około 0.5) są duże, bo każdy pacjent ma tylko dwa epizody; frailty pojedynczych pacjentów należy traktować orientacyjnie.



Wykres krzywych przeżycia grup pokazuje krzywą dla przeciętnego pacjenta ($w = 1$) oraz krzywe pacjentów o wyższym i niższym frailty: pacjent 1 traci „przeżycie bez zakażenia" szybciej, pacjent 4 – wolniej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$\theta = 0.40$**, test LR** $\theta = 0$**:** $p = 0.011$**; płeć:** $HR = 0.21$ **(**$p = 0.0005$**); wiek:** $HR = 1.01$ **(**$p = 0.64$**)**

Pacjenci istotnie różnią się podatnością na zakażenie, a po jej uwzględnieniu kobiety mają hazard zakażenia o 79% niższy niż mężczyźni; wiek nie ma znaczenia.

DLA ZAINTERESOWANYCH

Estymacja: Parametry szacuje się algorytmem EM na profilowanej wiarygodności częściowej (Klein 1992[191], Therneau i Grambsch 2000[199]). W kroku M dopasowywana jest regresja Coxa z przesunięciem (offsetem) $\log w_j$ i wyznaczany skumulowany hazard bazowy Breslowa $H_0(t)$; w kroku E frailty grupy jest zastępowane wartością oczekiwaną

$$w_j = \frac{1/\theta + D_j}{1/\theta + A_j}, \qquad A_j = \sum_{i \in j} H_0(t_i)\exp(\beta^T X_i),$$

gdzie D_j to liczba zdarzeń w grupie, a A_j jej skumulowany hazard. Wariancja θ jest wyznaczana przez maksymalizację zintegrowanej log-wiarygodności

$$\ell = \sum_{i:\delta_i=1} [\eta_i - \log S_0(t_i)] + \sum_j \ell_j + D,$$

$$\ell_j = D_j \log\theta + \log\Gamma(1/\theta + D_j) - \log\Gamma(1/\theta) - (1/\theta + D_j)\log(1 + \theta A_j),$$

która przy $\theta \to 0$ przechodzi w wiarygodność częściową Coxa (D – łączna liczba zdarzeń). Zbieżność przyjmuje się, gdy zmiana ℓ jest mniejsza niż 10^{-9}.

Błędy standardowe: z penalizowanej macierzy informacji dla parametrów $(\beta, \log w_1, \ldots, \log w_G)$, gdzie kara wynika z rozkładu gamma; blok β jest odwracany przez dopełnienie Schura.

τ **Kendalla:** dla dzielonego frailty gamma $\tau = \theta/(\theta + 2)$.

Test na granicy: statystyka $2(\ell_{frailty} - \ell_{Cox})$ ma asymptotycznie rozkład będący mieszanką 50:50 rozkładu zdegenerowanego w zerze i χ^2_1 (Self i Liang 1987[198]), dlatego wartość p z χ^2_1 dzieli się przez dwa.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Porównywanie modeli regresji PH Coxa z efektem losowym

KIEDY STOSOWAĆ:

Porównywanie modeli służy do wyboru zestawu zmiennych niezależnych, tak jak w porównywaniu modeli regresji PH Coxa: porównujemy model pełny z modelem zredukowanym lub korzystamy z doboru automatycznego. Efekt losowy (frailty gamma dla wskazanej zmiennej grupującej) jest **wspólny dla wszystkich porównywanych modeli** – w każdym z nich wariancja θ jest szacowana na nowo, ale struktura grupowa pozostaje ta sama. Dobór dotyczy wyłącznie zmiennych niezależnych; o tym, czy efekt losowy jest potrzebny, decyduje test LR $\theta = 0$ w raporcie regresji Coxa z efektem losowym.

DEFINICJE MIAR

Modele porównuje się na podstawie zintegrowanej log-wiarygodności, kryteriów informacyjnych AIC, AICc, BIC (im mniejsze, tym lepiej) oraz testu ilorazu wiarygodności dla modeli zagnieżdżonych: $2(\ell_P - \ell_Z)$ (ℓ_P, ℓ_Z – zintegrowana log-wiarygodność modelu pełnego i zredukowanego) ma rozkład χ^2 o liczbie stopni swobody równej liczbie usuniętych zmiennych.

Ocena modeli poza zbiorem uczącym

Kryteria informacyjne i test LR są wyznaczane na tych samych danych, na których dopasowano modele, przez co ocena modeli jest zbyt optymistyczna. Tak jak w porównywaniu modeli regresji PH Coxa, w oknie analizy można więc włączyć ocenę modeli na obserwacjach, które nie brały udziału w ich dopasowaniu (opcja Walidacja): **walidację krzyżową k-krotną** albo **losowy zbiór testowy**, z podziałem wyznaczonym przez Ziarno. Dla każdego modelu raport podaje wtedy $-2 \ln PL$ na zdarzenie (błąd walidacji; im mniejszy, tym lepiej) i współczynnik zgodności C Harrella, a w bloku doboru modelu – błąd modelu zerowego (bez zmiennych) i najlepszy model wg walidacji; jako kryterium doboru automatycznego można też wybrać błąd walidacji. W modelach z efektem losowym obowiązują dwie dodatkowe zasady:

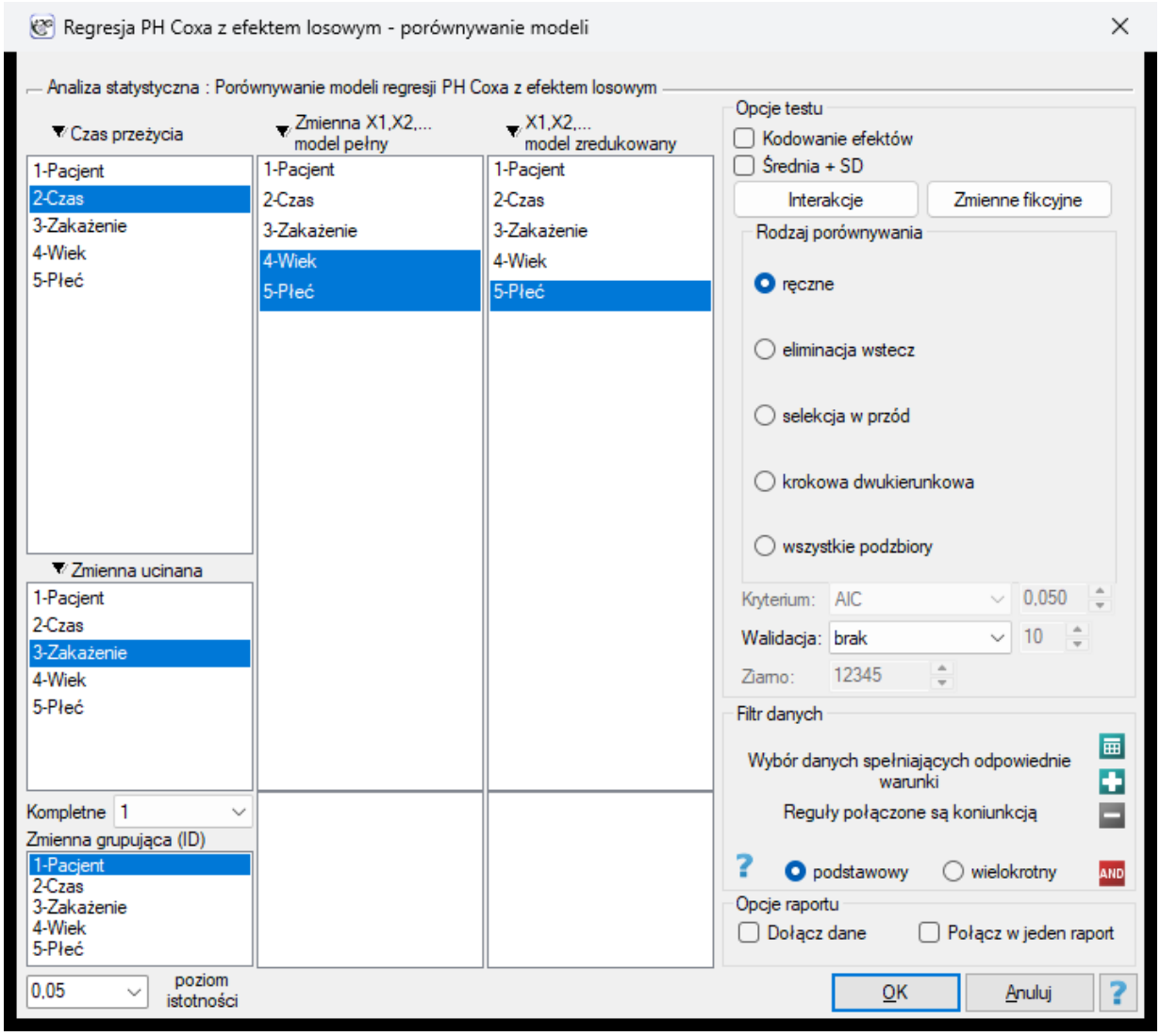
- **dzielone są całe grupy** – wszystkie obserwacje jednej grupy (np. jednego pacjenta) trafiają do tej samej części danych. Liczba podziałów k jest ograniczana do liczby grup, a udział zbioru testowego oznacza procent grup (podział nie jest tu warstwowany względem statusu);
- **przewidywanie dla nowej grupy opiera się wyłącznie na efektach stałych** – efekt losowy grupy, której nie było w zbiorze uczącym, nie jest znany, dlatego przyjmuje się jego wartość oczekiwaną (frailty $= 1$).

Uwaga! Hazard ratio w modelu z efektem losowym są warunkowe – porównują obserwacje o tej samej wartości efektu losowego – i są zwykle silniejsze niż efekty uśrednione po populacji. Zastosowane do nowej grupy, dla której przyjmuje się frailty $= 1$, dają przewidywania bardziej skrajne, niż wynikałoby to z danych, co wiarygodność częściowa poza zbiorem uczącym wyraźnie „karze". Przy dużej wariancji θ i małej liczbie grup błąd walidacji modeli ze zmiennymi może więc być większy niż błąd modelu zerowego, mimo że zmienne są istotne, a C Harrella wyraźnie przekracza 0.5. Błąd walidacji służy tu przede wszystkim do porównywania modeli między sobą.

JAK USTAWIĆ ANALIZĘ

Statystyka → Modele mieszane → Regresja PH Coxa z efektem losowym – porównywanie modeli.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Wskazujemy czas przeżycia, zmienną ucinaną z wartością Kompletne, zmienne modelu pełnego i modelu zredukowanego (albo kryterium doboru automatycznego) oraz Zmienną grupującą (ID), wspólną dla wszystkich modeli. Pod listą metod doboru znajdują się pola: Kryterium (AIC, AICc, BIC, p (test LR / F) z progiem albo błąd walidacji), Walidacja (brak, krzyżowa, k = z liczbą podziałów albo zbiór testowy, % = z udziałem zbioru testowego w procentach grup) i Ziarno losowania podziału (to samo ziarno oznacza ten sam podział).

WYNIKI I INTERPRETACJA

Raport zawiera tabelę dopasowania porównywanych modeli (zintegrowana log-wiarygodność, AIC, AICc, BIC, θ każdego modelu) z testem LR między nimi oraz tabele współczynników modeli z warunkowymi hazard ratio. Przy włączonej walidacji raport zawiera dodatkowo blok **Ocena poza zbiorem uczącym** (sposób podziału, ziarno, jednostka podziału, błąd modelu zerowego, najlepszy model wg walidacji), a przy każdym modelu – $-2 \ln PL$ na zdarzenie i C Harrella.

PRZYKŁAD 26.11. c.d. przykładu 26.10 (plik nerkiCewnik.pqs)

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

W modelu ryzyka zakażenia cewnika wiek okazał się nieistotny. Porównujemy model pełny (wiek i płeć) z modelem zredukowanym (sama płeć); efekt losowy pacjenta w obu.

Liczba zmiennych w modelu 1	2
Kryterium zbieżności nie zostało osiągnięte	
Wariancja frailty (theta)	0,397184
Błąd theta (profil)	0,234387
Tau Kendalla	0,165688
-2 Log Likelihood	364,106718
AIC - kryterium Akaikego	370,106718
AICc - skorygowane kryterium Akaikego	370,551163
BIC - bayesowskie kryterium Schwarza	376,288047
Pseudo R2 (McFadden)	NA
R2 (Nagelkerke)	NA
R2 (Coxa-Snella)	NA
Liczba zmiennych w modelu 2	1
Kryterium zbieżności nie zostało osiągnięte	
Wariancja frailty (theta)	0,387508
Błąd theta (profil)	0,231994
Tau Kendalla	0,162306
-2 Log Likelihood	364,328423
AIC - kryterium Akaikego	368,328423
AICc - skorygowane kryterium Akaikego	368,546605
BIC - bayesowskie kryterium Schwarza	372,449309
Pseudo R2 (McFadden)	NA
R2 (Nagelkerke)	NA
R2 (Coxa-Snella)	NA
Chi-kwadrat - porównanie modeli Model 1 vs Model 2	0,221705
Stopnie swobody	1
Wartość p	0,637744

Model 1

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	HR (warunkc	-95% CI	+95% CI
Wiek	0,005462	0,01158	-0,017234	0,028158	0,222456	0,637175	1,005477	0,982913	1,028558
Płeć	-1,55613	0,444803	-2,427928	-0,684332	12,239291	0,000468	0,210951	0,088219	0,504427

Model 2

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	HR (warunkc	-95% CI	+95% CI
Płeć	-1,534128	0,440713	-2,397909	-0,670348	12,117491	0,0005	0,215644	0,090908	0,511531

Model pełny: log-wiarygodność -182.05, AIC $= 370.11$; model z samą płcią: -182.16, AIC $= 368.33$. Test LR: $\chi^2 = 0.22$, $df = 1$, $p = 0.64$ – usunięcie wieku nie pogarsza dopasowania, a AIC modelu prostszego jest niższe. Wybieramy model z samą płcią: $HR = 0.216$ (95% CI: 0.091–0.512), $\theta = 0.388$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dla kontrastu model z samym wiekiem ma log-wiarygodność -187.86 i AIC $= 379.72$, niewiele lepiej niż model bez zmiennych (-187.95): wiek sam nic nie wnosi.

test LR pełny vs bez wieku: $\chi^2 = 0.22$, $p = 0.64$; AIC 370.11 vs 368.33

Wiek można usunąć z modelu; pozostaje płeć i efekt losowy pacjenta.

Ocena poza zbiorem uczącym. To samo porównanie powtarzamy, wybierając w polu Walidacja opcję krzyżowa, k = z 10 podziałami (ziarno 12345). Dzielonych jest 38 pacjentów – oba epizody tego samego pacjenta trafiają do tej samej części danych.

Ocena poza zbiorem uczącym	walidacja krzyżowa
Liczba podziałów (k)	10
Ziarno losowania podziału	12345
Jednostka podziału	cała grupa
Model zerowy: -2 ln PL / zdarzenie	8,295639
Najlepszy model wg walidacji (-2 ln PL / zdarzenie)	model zerowy
Predykcja nowych grup: efekty stałe, frailty = 1	

	zbiór uczący	walidacja krzyżowa ($k = 10$)	
model	AIC	$-2 \ln PL$ na zdarzenie	C Harrella
zerowy		8.2956	
Model 1 (wiek, płeć)	370.11	8.5322	0.6723
Model 2 (płeć)	368.33	8.4458	0.6186

Model z samą płcią ma mniejszy błąd walidacji niż model z wiekiem i płcią (8.4458 wobec 8.5322) – ocena poza zbiorem uczącym potwierdza, że wiek można usunąć. Jednocześnie najmniejszy błąd ma model zerowy (8.2956), choć płeć jest w modelu istotna, a C Harrella obu modeli (0.62–0.67) pokazuje, że porządkują one nowych pacjentów lepiej niż losowo. To sytuacja opisana w uwadze powyżej: efekt płci jest warunkowy (silny: $HR = 0.216$ przy $\theta = 0.388$), próba mała (38 pacjentów, po 3–4 w każdej ocenianej części), a dla nowego pacjenta przyjmuje się frailty $= 1$, przez co przewidywania są zbyt skrajne. Wynik ten nie podważa znaczenia płci, ale przypomina, że przewidywania tego modelu dla nowych pacjentów – bez znajomości ich efektu losowego – są niepewne.

walidacja krzyżowa ($k = 10$, całe grupy): $-2 \ln PL$ na zdarzenie $= 8.53$ (wiek i płeć), 8.45 (płeć), 8.30 (model zerowy)

Walidacja potwierdza usunięcie wieku; przy tak małej próbie żaden z modeli nie przewiduje jednak czasu do zakażenia u nowych pacjentów lepiej niż model zerowy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

27 REDUKCJA WYMIARÓW I GRUPOWANIE

Wraz ze wzrostem liczby zmiennych poddawanych analizie statystycznej rośnie dokładność opisu badanego zjawiska, ale rośnie też stopień komplikacji i trudność interpretacji uzyskanych wyników. Zbyt duża liczba zmiennych niesie za sobą ryzyko, że część z nich jest ze sobą skorelowana, a więc niesie informację częściowo powtórzoną, już zawartą w innych zmiennych. Potrzeba redukcji liczby zmiennych (redukcji wymiarowości) zrodziła całą grupę metod poświęconych temu zagadnieniu, takich jak: analiza czynnikowa, analiza składowych głównych, analiza skupień czy analiza dyskryminacyjna. Metody te pozwalają wykryć powiązania między zmiennymi, a na ich podstawie wyodrębnić grupy zmiennych podobnych do siebie i do dalszych analiz wybrać tylko jednego przedstawiciela każdej grupy albo nową zmienną, której wartości są wyliczane na podstawie pozostałych zmiennych z grupy. W ten sposób mamy pewność, że informacja niesiona przez każdą grupę zmiennych jest w analizie uwzględniona – przy niewielkiej utracie informacji zbiór p zmiennych można zredukować do zbioru k zmiennych, gdzie $k < p$.

Podobnie jak grupowaniem zmiennych, możemy być zainteresowani grupowaniem obiektów. Mając informację o pewnych cechach obiektów, jesteśmy często w stanie wyodrębnić grupy obiektów podobnych pod względem tych cech, np. na podstawie kwot przeznaczanych przez klientów pewnej sieci sklepów na poszczególne artykuły można dokonać podziału klientów na segmenty o podobnych preferencjach zakupowych. W konsekwencji swoją ofertę lub reklamę można przygotować nie dla ogółu kupujących, ale oddzielnie dla każdego segmentu, tak by precyzyjniej trafić w potrzeby potencjalnego klienta.

Analiza składowych głównych

KIEDY STOSOWAĆ:

Analiza składowych głównych *(ang. Principal Component Analysis, PCA)* służy do redukcji liczby zmiennych opisujących badane obiekty. Zamiast wielu, często wzajemnie skorelowanych zmiennych pierwotnych, wyznacza się niewielką liczbę zupełnie nowych zmiennych – **składowych głównych**, będących ich kombinacją liniową. Kolejne składowe główne są wzajemnie nieskorelowane (ortogonalne), a ich liczba (k) jest mniejsza lub równa liczbie zmiennych pierwotnych (p).

Dokładna analiza składowych głównych pozwala wskazać, które zmienne pierwotne mają duży wpływ na daną składową, czyli tworzą razem z nią grupę jednorodną – składowa główna staje się wówczas reprezentantem tej grupy. Uzyskane w ten sposób nowe zmienne są zwykle liczniejsze mniej niż zmienne pierwotne, a niosą tyle samo lub tylko nieznacznie mniej informacji. Co więcej, ponieważ nie są ze sobą skorelowane, mogą być z powodzeniem wykorzystane w większości dalszych analiz wielowymiarowych.

WARUNKI STOSOWANIA

- pomiar zmiennych na skali interwałowej,
- co najmniej kilka zmiennych pierwotnych poddanych analizie,
- istotne skorelowanie zmiennych pierwotnych – jeśli zmienne nie są ze sobą skorelowane (współczynniki korelacji Pearsona bliskie 0), przeprowadzanie analizy składowych głównych

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

nie ma sensu, gdyż wówczas każda zmienna stanowiłaby już odrębną składową; skorelowanie zmiennych można ocenić testem Bartletta oraz współczynnikiem Kaisera-Mayera-Olkina (KMO), opisanymi poniżej.

DEFINICJE MIAR

Poszczególne składowe główne są kombinacją liniową zmiennych pierwotnych $X_1, X_2, ..., X_p$. Każda z nich wyjaśnia pewną część zmienności zmiennych pierwotnych i w naturalny sposób opiera się na takich miarach zmienności, jak kowariancja (gdy zmienne pierwotne są podobnej wielkości i wyrażone w tych samych jednostkach) lub korelacja (gdy warunki potrzebne do stosowania kowariancji nie są spełnione). Każdą składową główną opisują poniższe wielkości.

Wartość własna

Informuje, jaka część całkowitej zmienności zmiennych pierwotnych jest tłumaczona przez daną składową główną. Pierwsza składowa tłumaczy największą część wariancji, druga – największą część tej wariancji, która nie została wytłumaczona przez pierwszą, kolejna – największą część wariancji nie wytłumaczonej przez poprzednie składowe. W efekcie każda kolejna składowa główna tłumaczy coraz mniejszą część wariancji, czyli kolejne wartości własne są coraz mniejsze. Ponieważ całkowita wariancja jest sumą wartości własnych, dla każdej składowej można wyliczyć procent tłumaczonej przez nią zmienności, a dla kolejnych składowych – zmienność skumulowaną i skumulowany procent zmienności.

Wektor własny

Odzwierciedla wpływ poszczególnych zmiennych pierwotnych na daną składową główną – zawiera współczynniki kombinacji liniowej wyznaczającej tę składową. Znak tych współczynników wskazuje kierunek wpływu i jest przypadkowy, co nie zmienia wartości niesionej informacji.

Ładunki czynnikowe

Podobnie jak wektor własny, odzwierciedlają wpływ poszczególnych zmiennych pierwotnych na daną składową główną – są to wartości pokazujące, jaką część wariancji danej składowej stanowią zmienne pierwotne. Gdy analiza jest oparta na macierzy korelacji, ładunki te interpretuje się jako współczynniki korelacji pomiędzy zmiennymi pierwotnymi a daną składową główną.

Wkłady zmiennych

Bazują na współczynnikach determinacji pomiędzy zmiennymi pierwotnymi a daną składową główną. Wskazują, jaki procent zmienności danej składowej głównej może być tłumaczony zmiennością poszczególnych zmiennych pierwotnych.

Zasoby zmienności wspólnej

Również bazują na współczynnikach determinacji pomiędzy zmiennymi pierwotnymi a daną składową główną, ale wskazują odwrotną zależność – jaki procent zmienności danej zmiennej pierwotnej może być tłumaczony zmiennością pierwszych kilku składowych głównych. Na przykład wynik dotyczący drugiej zmiennej w kolumnie czwartej składowej głównej mówi o tym, jaki procent zmienności drugiej zmiennej może być tłumaczony zmiennością pierwszych czterech składowych głównych.

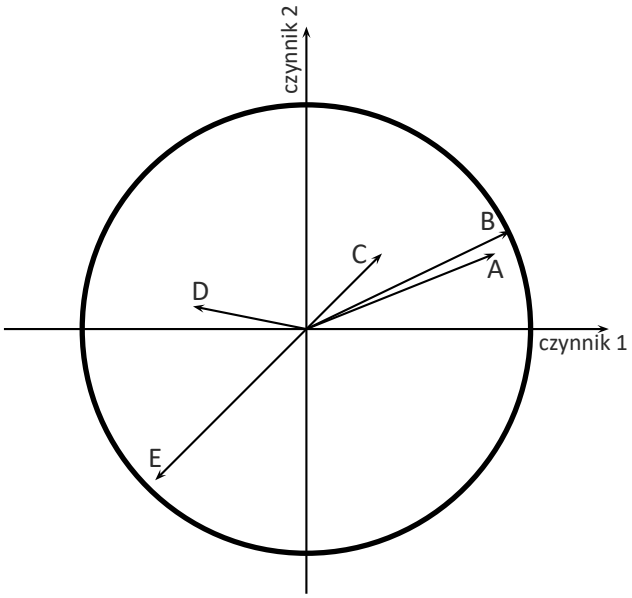
Interpretacja graficzna

Wiele informacji niesionych przez współczynniki zwracane w tabelach można przedstawić na jednym wykresie. Wykresy zbierają w jednym miejscu informacje o wzajemnych relacjach pomiędzy składowymi, zmiennymi pierwotnymi i przypadkami, przez co dają całościowy obraz analizy i są jej bardzo dobrym

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

podsumowaniem.

Wykres ładunków czynnikowych przedstawia wektory połączone z początkiem układu współrzędnych, reprezentujące zmienne pierwotne, umieszczone na płaszczyźnie wyznaczonej przez dwie wybrane składowe główne.



Współrzędne końca wektora to odpowiadające im ładunki czynnikowe zmiennych.

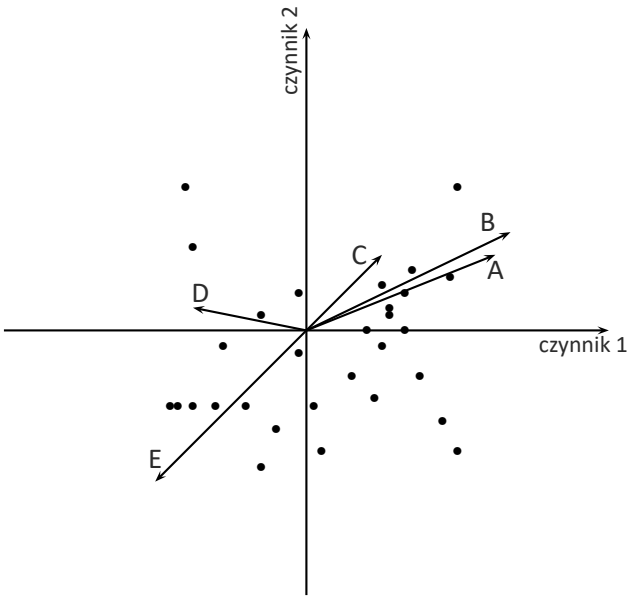
Długość wektora reprezentuje zasób informacyjny zmiennej pierwotnej, jaki niosą składowe główne wyznaczające układ współrzędnych – im dłuższy wektor, tym większy wkład zmiennej pierwotnej w budowę składowych. W przypadku analizy opartej na macierzy korelacji ładunki są korelacjami pomiędzy zmiennymi pierwotnymi a składowymi, więc punkty wpadają do koła jednostkowego (współczynnik korelacji nie może przekroczyć jedynki). Im dana zmienna pierwotna leży bliżej brzegu tego koła, tym lepsza jest jej reprezentacja przez przedstawione składowe główne.

Znak współrzędnych końca wektora czyli znak ładunku czynnikowego, wskazuje na dodatnią lub ujemną korelację zmiennej pierwotnej i składowych głównych tworzących układ współrzędnych. Rozpatrując łącznie obie osie (2 składowe), zmienne pierwotne można zakwalifikować do jednej z czterech kategorii, w zależności od kombinacji znaków $(+/-)$ ich ładunków czynnikowych.

Kąt między wektorami wskazuje na skorelowanie zmiennych pierwotnych:
$0 < \alpha < 90^0$: im kąt pomiędzy wektorami reprezentującymi zmienne pierwotne jest mniejszy, tym silniejsza jest dodatnia korelacja pomiędzy tymi zmiennymi.
$\alpha = 90^0$ – wektory są prostopadłe, czyli zmienne pierwotne nie są skorelowane.
$90^0 < \alpha < 180^0$ – im kąt pomiędzy wektorami reprezentującymi zmienne pierwotne jest większy, tym silniejsza jest ujemna korelacja pomiędzy tymi zmiennymi.

Biplot przedstawia dwie serie danych umieszczone w układzie współrzędnych wyznaczonym przez dwie składowe główne: pierwszą serię stanowią dane z wykresu ładunków czynnikowych (wektory zmiennych pierwotnych), a drugą – punkty przedstawiające poszczególne przypadki.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



Współrzędne punktów interpretujemy jak wartości zestandaryzowane, tzn. współrzędne dodatnie wskazują na wartość wyższą od średniej wartości składowej głównej, ujemne – na wartość niższą, a im wyższa wartość bezwzględna, tym dalej punkt znajduje się od średniej. Jeśli na wykresie znajdują się obserwacje nietypowe (odstające), mogą one zaburzać analizę i powinny zostać usunięte, a analiza przeprowadzona ponownie.

Odległości między punktami wskazują na podobieństwo przypadków – im bliżej (w sensie odległości euklidesowej) znajdują się względem siebie punkty, tym bardziej podobne informacje niosą porównywane przypadki.

Rzuty prostopadłe punktów na wektory interpretuje się tak samo jak współrzędne punktów, z tą różnicą, że dotyczą one nie składowych głównych, a zmiennych pierwotnych. Wartości po stronie końca wektora są większe od średniej wartości zmiennej pierwotnej, a wartości na przedłużeniu wektora w kierunku przeciwnym są mniejsze od średniej.

Kryteria redukcji wymiarów

Nie istnieje jedno uniwersalne kryterium wyboru liczby składowych głównych, dlatego dobrze jest kierować się kilkoma metodami jednocześnie.

Procent wyjaśnionej wariancji

Wszystkie składowe główne razem niosą 100% wariancji zmiennych pierwotnych. Jeśli suma wariancji kilku pierwszych składowych stanowi znaczną część całkowitej wariancji, wówczas te składowe mogą w zadowalającym stopniu zastąpić zmienne pierwotne. Przyjmuje się, że wariancja powinna zostać odzwierciedlona w wybranych składowych głównych w ponad 80%.

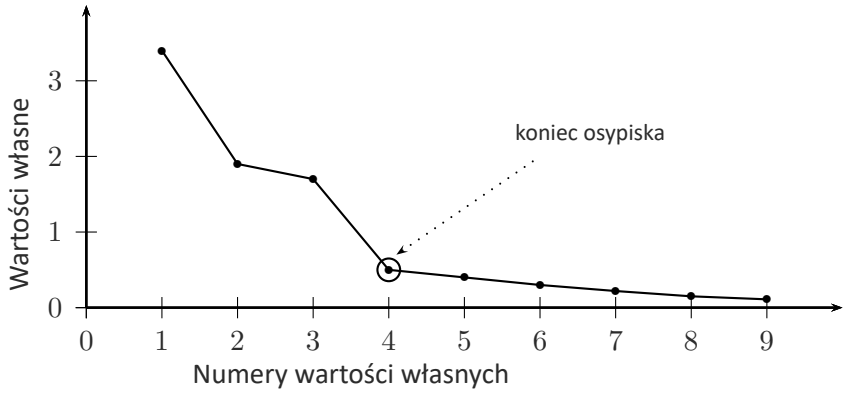
Kryterium Kaisera

Zgodnie z tym kryterium do interpretacji pozostawia się te składowe główne, które mają przynajmniej taką wariancję, jak dowolna wystandaryzowana zmienna pierwotna (czyli wariancję równą 1). Innymi słowy, ważne są tylko składowe, których wartość własna przekracza lub jest bliska wartości 1.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wykres osypiska

Przedstawia tempo spadku wartości własnych (procentu wyjaśnionej wariancji) dla kolejnych składowych.



Moment na wykresie, w którym proces ten stabilizuje się i linia malejąca przechodzi w poziomą, to tzw. koniec osypiska. Składowe znajdujące się na prawo od tego punktu reprezentują znikomą wariancję i przedstawiają w większości losowy szum.

Wyznaczenie składowych głównych

Gdy zdecydujemy, ilu składowych głównych będziemy potrzebować, możemy przystąpić do ich wygenerowania. Składowe główne powstałe w oparciu o macierz korelacji są wyliczane jako kombinacja liniowa wystandaryzowanych zmiennych pierwotnych, natomiast te powstałe w oparciu o macierz kowariancji – jako kombinacja liniowa zmiennych pierwotnych wycentrowanych względem średniej.

Sprawdzanie zasadności stosowania analizy

Zanim przystąpimy do wyznaczania i interpretowania składowych głównych, warto sprawdzić, czy analiza w ogóle ma sens, czyli czy zmienne pierwotne są ze sobą wystarczająco skorelowane. Służą do tego test Bartletta oraz współczynnik Kaisera-Mayera-Olkina (KMO).

Test Bartletta

Weryfikuje hipotezę, że współczynniki korelacji pomiędzy zmiennymi są zerowe (czyli macierz korelacji jest macierzą jednostkową). Istotny statystycznie wynik testu (wartość p niższa niż przyjęty poziom istotności α) świadczy o tym, że zmienne pierwotne są znacząco skorelowane, a więc przeprowadzenie analizy składowych głównych ma sens.

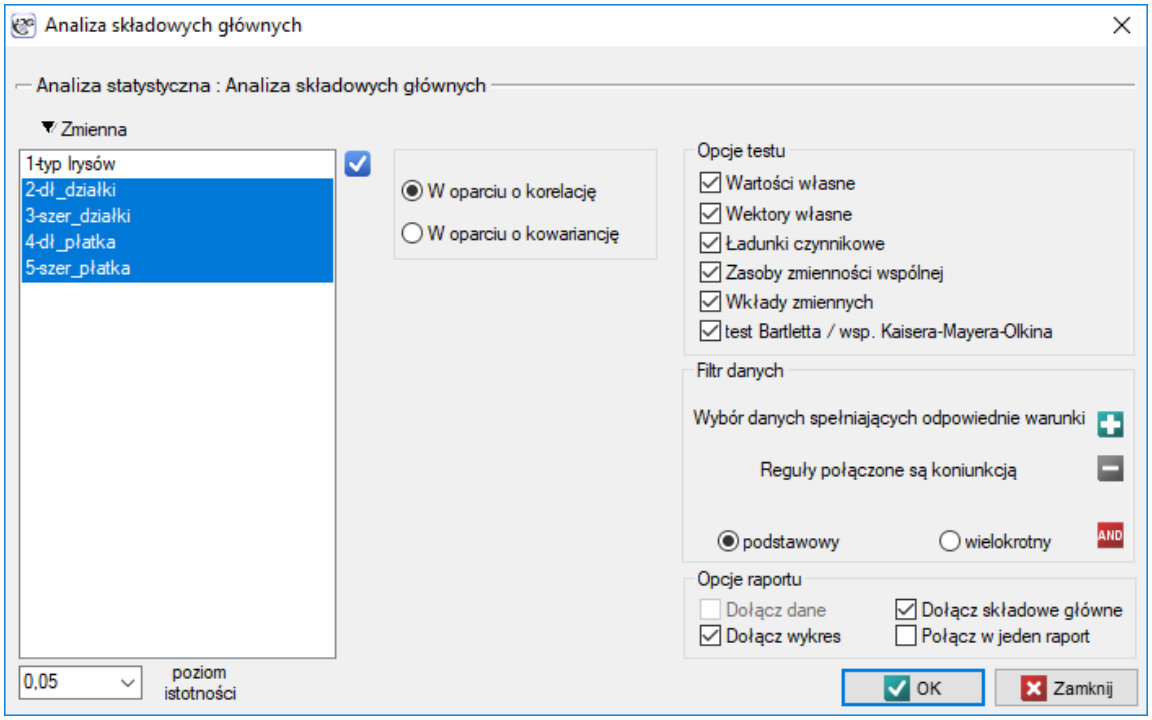
Współczynnik Kaisera-Mayera-Olkina (KMO)

Ocenia stopień skorelowania zmiennych pierwotnych, czyli siłę przesłanek za sensownością przeprowadzania analizy składowych głównych. Wartość współczynnika należy do przedziału $< 0, 1 >$ – wartości niskie świadczą o braku podstaw do przeprowadzania analizy, a wartości wysokie są przesłanką do jej przeprowadzenia.

JAK USTAWIĆ ANALIZĘ

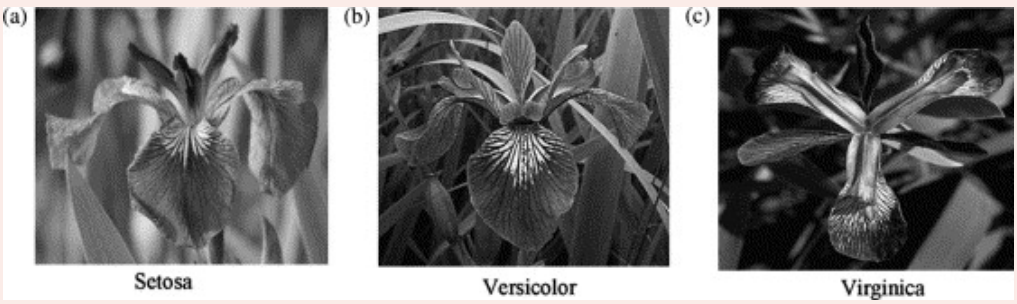
Okno z ustawieniami opcji Analizy składowych głównych wywołujemy poprzez menu Statystyki → Modele wielowymiarowe → Analiza składowych głównych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

Przykład 27.1. (plik: iris.pqs)
Ten klasyczny zestaw danych pierwotnie ukazał się w pracy R.A. Fishera 1936[59], gdzie przedstawiona została analiza dyskryminacyjna. Plik zawiera pomiary (w centymetrach) długości i szerokości płatków i działek kielicha dla 3 odmian kwiatu irysa. Poddane badaniu gatunki to setosa, versicolor i virginica. Interesujące jest określenie sposobu rozróżniania tych gatunków na bazie uzyskanych pomiarów.



Rycina pochodzi z pracy Lee i innych (2006r): "Application of a noisy data classification technique to determine the occurrence of flashover in compartment fires"

Analiza składowych głównych pozwoli wskazać te pomiary (długości i szerokości płatków i działek kielicha), które dają badaczowi najwięcej informacji o obserwowanych kwiatach.

Pierwszym etapem, wykonywanym jeszcze przed przystąpieniem do wyznaczania i analizowania składowych, jest sprawdzenie celowości przeprowadzenia tej analizy. Zaczynamy więc od wyznaczenia macierzy korelacji zmiennych oraz przeanalizowania uzyskanych korelacji przy wykorzystaniu testu Bartletta i współczynnika KMO.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Macierz korelacji

Zmienna	dł_działki	szer_dzia	dł_płatka	szer_płat
dł_działki	1	-0,11757	0,871754	0,817941
szer_dzia	-0,11757	1	-0,42844	-0,36612(
dł_płatka	0,871754	-0,42844	1	0,962865
szer_płat	0,817941	-0,36612(	0,962865	1

Analiza składowych głównych	
Czas analizy	0,34 sek.
Analizowane zmienne	dł_działki;szer_działki;dł_
Poziom istotności	0,05
Analiza dla macierzy korelacji	
Test Bartletta	
Statystyka Chi-kwadrat	706,959243
Ilość stopni swobody	0
Wartość p	<0.000001
Współczynnik Keisera-Mayera-Olkina	
KMO	0,540077

Test Bartletta istotny statystycznie; KMO $= 0.54$

Zmienne pierwotne są znacząco skorelowane (macierz korelacji istotnie różni się od macierzy jednostkowej), a przeciętna wartość współczynnika KMO wystarczająco potwierdza sensowność przeprowadzenia analizy składowych głównych.

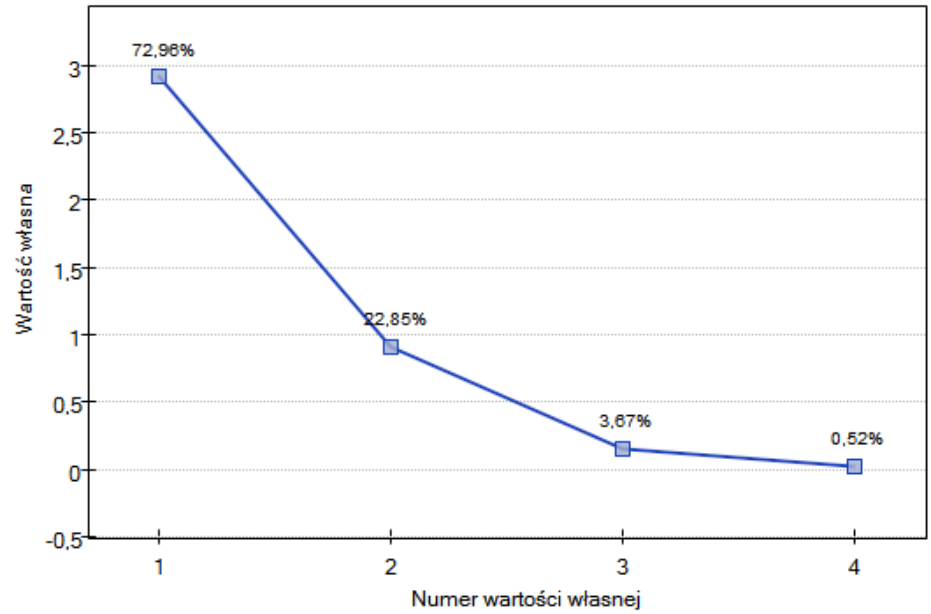
Pierwszym wynikiem tej analizy, na który należy zwrócić szczególną uwagę, są wartości własne:

Wartości własne

Numer	Wartość \	% warian	Wartość s	% skumul
1	2,918498	72,96244	2,918498	72,96244
2	0,91403	22,85076	3,832528	95,81320
3	0,146757	3,668922	3,979285	99,48212
4	0,020715	0,517871	4	100

Uzyskane wartości własne wskazują, że dwie, a nawet jedna składowa główna w dobry sposób opiszą nasze dane. Wartość własna dla pierwszej składowej wynosi 2.92, a procent wyjaśnionej przez nią wariancji to 72.96%. Druga składowa wyjaśnia już znacznie mniej wariancji, bo 22.85%, a jej wartość własna to 0.91. Według kryterium Kaisera wystarczająca do interpretacji jest tylko jedna składowa główna, gdyż tylko dla pierwszej wartość własna jest większa niż 1. Patrząc jednak na wykres osypiska, można wysnuć wniosek, że linia spadkowa przechodzi w poziomą dopiero od trzeciej składowej głównej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Zgodnie z tym wskazaniem, dwie pierwsze składowe wnoszą istotną informację – łącznie wyjaśniają aż 95.81% wariancji (patrz kolumna % skumulowany).

Zasoby zmienności wspólnej dla pierwszej składowej są wysokie dla wszystkich zmiennych pierwotnych, za wyjątkiem zmiennej szerokość działki, dla której wynoszą 21.17%. Oznacza to, że gdybyśmy pozostali przy interpretacji tylko pierwszej składowej, zmienna szerokość działki zostałaby odzwierciedlona w niewielkim stopniu.

Zasoby zmienności wspólnej %

Zmienna	Czynnik1	Czynnik2	Czynnik3	Czynnik4
dł_działki	79,24004	92,25986	99,85857	100
szer_dzia	21,17313	99,09193	99,9684	100
dł_płatka	98,31816	98,37299	98,66944	100
szer_płat	93,11843	93,52803	99,43209	100

Dla dwóch pierwszych składowych zasoby zmienności wspólnej są już na podobnym, bardzo wysokim poziomie – dla każdej z analizowanych zmiennych przekraczają 90%, czyli wariancja każdej zmiennej jest reprezentowana przy użyciu tych dwóch składowych w ponad 90%. Zebrawszy całą tę wiedzę, zdecydowano się na wyodrębnienie i interpretację 2 składowych.

By przyjrzeć się dokładniej powiązaniu składowych głównych i zmiennych pierwotnych, czyli długości i szerokości płatków i działek kielicha, interpretujemy: wektory własne, ładunki czynnikowe oraz wkłady zmiennych.

Wektory własne

Zmienna	Czynnik1	Czynnik2	Czynnik3	Czynnik4
dł_działki	-0,52106	-0,37741	0,719566	0,261286
szer_dzia	0,269347	-0,92329	-0,24438	-0,12351
dł_płatka	-0,58041	-0,02449	-0,14212	-0,80144
szer_płat	-0,56485	-0,06694	-0,63427	0,523597

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Ładunki czynnikowe

Zmienna	Czynnik1	Czynnik2	Czynnik3	Czynnik4
dł_działki	-0,89016	-0,36083	0,275658	0,037606
szer_dzia	0,460143	-0,88271	-0,09362	-0,01777
dł_płatka	-0,99155	-0,02341	-0,05444	-0,11535
szer_płat	-0,96497	-0,064	-0,24298	0,07536

Wkład zmiennych %

Zmienna	Czynnik1	Czynnik2	Czynnik3	Czynnik4
dł_działki	27,15096	14,24440	51,77757	6,827052
szer_dzia	7,254804	85,24748	5,972245	1,525463
dł_płatka	33,68793	0,059984	2,01999	64,23208
szer_płat	31,90629	0,448123	40,23019	27,41539

Poszczególne zmienne pierwotne w różny sposób oddziałują na pierwszą składową główną. Ułóżmy je kolejno według wielkości tego wpływu:

1. Długość płatka jest ujemnie skorelowana z pierwszą składową, a zatem im dłuższy płatek, tym niższe wartości tej składowej. Wektor własny dla długości płatka jest największy w tej składowej i wynosi -0.58. Jego ładunek czynnikowy informuje, że korelacja pomiędzy pierwszą składową główną a długością płatka jest bardzo wysoka i wynosi -0.99, co daje 33.69% wkładu w pierwszą składową;

2. Szerokość płatka ma nieco tylko mniejszy wpływ na pierwszą składową i jest z nią również ujemnie skorelowana;

3. Długość działki interpretujemy podobnie jak poprzednie dwie zmienne, ale jej wpływ na pierwszą składową jest mniejszy;

4. Szerokość działki jest najsłabiej skorelowana z pierwszą składową, a znak tej korelacji jest dodatni.

Druga składowa reprezentuje głównie zmienną pierwotną szerokość działki – pozostałe zmienne pierwotne są w niej odzwierciedlone w niewielkim stopniu. Wektor własny, ładunek czynnikowy oraz wkład zmiennej szerokość działki są w składowej drugiej najwyższe.

Każda składowa główna wyznacza homogeniczną grupę zmiennych pierwotnych. Pierwszą składową nazwiemy „rozmiar płatka”, gdyż najbardziej znaczącymi dla niej zmiennymi są te niosące informacje o płatku, choć trzeba zaznaczyć, że długość działki również znacznie wpływa na wartość tej składowej. W interpretacji pamiętamy, że im większe są wartości tej składowej, tym mniejsze są płatki. Drugą składową nazwiemy natomiast „szerokość działki”, gdyż tylko szerokość działki w większym stopniu odzwierciedlona jest w drugiej składowej. Przy czym im większe są wartości tej składowej, tym węższa jest działka.

Ostatecznie składowe wygenerujemy, wybierając w oknie analizy opcję: Dołącz składowe główne. Fragment uzyskanego wyniku przedstawiamy poniżej:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

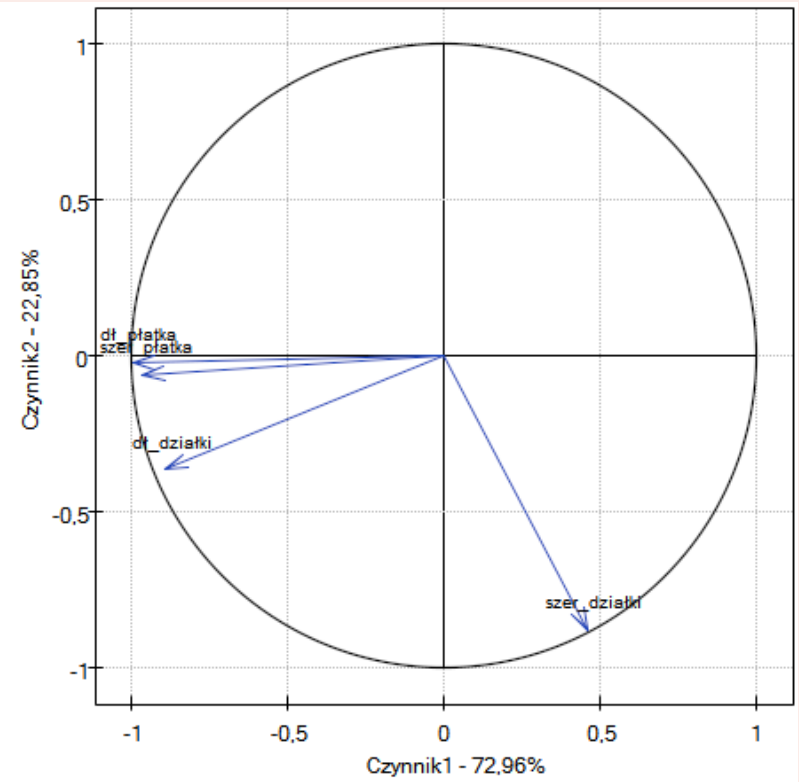
Składowe główne			
Czynnik1	Czynnik2	Czynnik3	Czynnik4
2,257141	-0,47842	0,12728	0,024088
2,074013	0,671883	0,233826	0,102663
2,356335	0,340766	-0,04405	0,028282
2,291707	0,5954	-0,09098	-0,06573
2,381863	-0,64467	-0,01568	-0,03580
2,068701	-1,48420	-0,02687	0,006586
2,435868	-0,04748	-0,33435	-0,03665
2,225392	-0,22240	0,088399	-0,02453
2,326845	1,111604	-0,14459	-0,02677
2,177035	0,467448	0,252918	-0,03976
2,159077	-1,04020	0,267784	0,016676
2,318364	-0,13263	-0,09344	-0,13303
2,211044	0,726243	0,23014	0,002417
2,624309	0,958296	-0,18019	-0,01915
2,191399	-1,85384	0,471322	0,194082

Aby dwie pierwsze składowe można było wykorzystać zamiast czterech wcześniejszych zmiennych pierwotnych, ostatecznie przeklejamy je do arkusza danych. Dalsze planowane przez siebie statystyki badacz może teraz przeprowadzić na dwóch nowych, nieskorelowanych zmiennych.

Analiza wykresów dwóch pierwszych składowych

Analiza wykresów nie dość, że doprowadzi badacza do tych samych wniosków co analiza tabel, to da możliwość wnikliwszej oceny uzyskanych wyników.

Wykres ładunków czynnikowych



Wykres przedstawia dwie pierwsze składowe główne, które reprezentują 72.96% i 22.85% wariancji, co daje łącznie 95.81% wariancji zmiennych pierwotnych.

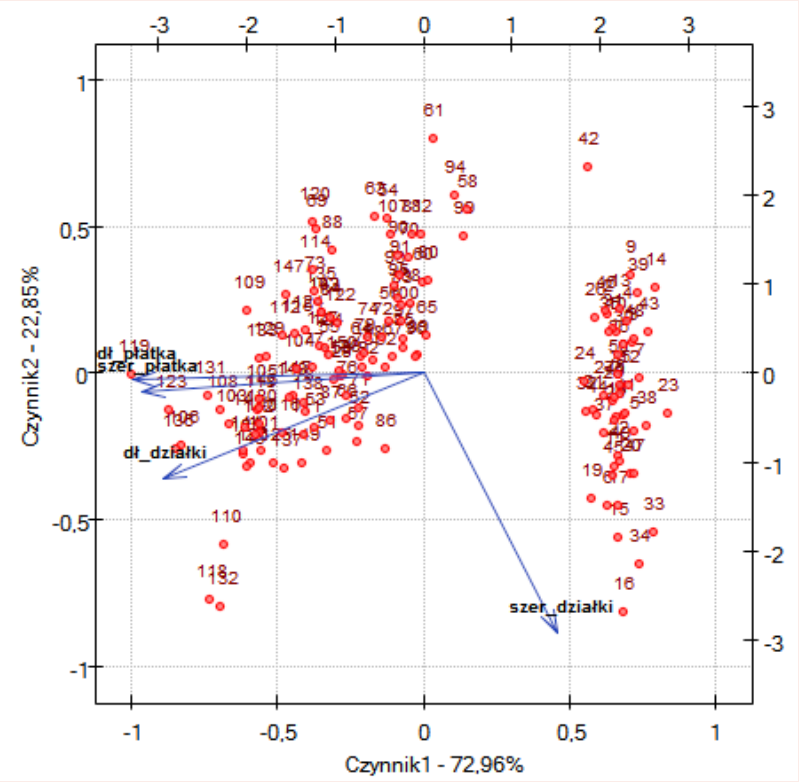
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wektory reprezentujące zmienne pierwotne nieomalże sięgają brzegów koła jednostkowego (koła o promieniu 1), a zatem wszystkie są bardzo dobrze reprezentowane przez dwie pierwsze składowe główne tworzące układ współrzędnych.

Kąt pomiędzy wektorami obrazującymi długość płatka, szerokość płatka i długość działki jest niewielki, co oznacza duże skorelowanie tych zmiennych. Skorelowanie zaś tych zmiennych ze składowymi tworzącymi układ jest negatywne – wektory znajdują się w III ćwiartce układu, przy czym wyższe wartości współrzędnych końca wektora obserwowane są dla składowej pierwszej niż drugiej. Takie ułożenie wektorów świadczy o tym, że tworzą one grupę jednorodną, reprezentowaną głównie przez składową pierwszą.

Zupełnie inny kierunek wskazuje wektor dotyczący szerokości działki, który jest w niewielkim stopniu skorelowany z pozostałymi zmiennymi pierwotnymi, o czym świadczy kąt nachylenia do pozostałych zmiennych, bliski kątowi prostemu. Skorelowanie tego wektora z pierwszą składową jest pozytywne i niezbyt wysokie (niska wartość pierwszej współrzędnej końca wektora), a z drugą składową – negatywne i wysokie (wysoka wartość drugiej współrzędnej końca wektora). Wnioskujemy z tego, że szerokość działki jako jedyna zmienna pierwotna jest dobrze reprezentowana przez drugą składową.

Biplot



Biplot przedstawia dwie serie danych rozpięte na dwóch pierwszych składowych. Jedną serią są wektory zmiennych pierwotnych, przedstawione już na poprzednim wykresie, a drugą – punkty niosące informacje o poszczególnych kwiatach, przy czym wartości drugiej serii odczytujemy na górnej osi X i prawej osi Y. Sposób interpretacji wektorów, czyli pierwszej serii, omówiliśmy przy poprzednim wykresie. By zrozumieć interpretację punktów, skupimy się na kwiatach o numerze 33 i 34 oraz o numerze 109.

Kwiaty o numerze 33 i 34 są do siebie podobne – odległość między nimi jest niewielka. Oba punkty

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

mają sporo większą niż przeciętna wartość pierwszej i sporo mniejszą niż przeciętna wartość drugiej składowej. Wartość przeciętna, czyli średnia arytmetyczna obu składowych, wynosi 0, więc jest to środek układu współrzędnych. Pamiętając, że pierwsza składowa to głównie rozmiar płatków, a druga – szerokość działki, możemy powiedzieć, że kwiaty 33 i 34 mają niewielkie płatki i dużą szerokość działki. Natomiast kwiat 109 reprezentuje punkt znacznie oddalony od pozostałych dwóch. Jest to kwiat o ujemnej pierwszej składowej i dodatniej, lecz niewysokiej, drugiej składowej, a więc kwiat o stosunkowo dużych płatkach i szerokości działki nieco tylko mniejszej niż przeciętna.

Podobne informacje uzyskamy, rzutując punkty na linie przedłużające wektory zmiennych pierwotnych. Przykładowo kwiat 33 wykazuje się dużą szerokością działki (wysokie i dodatnie wartości rzutowania na zmienną pierwotną szerokość działki), ale niskimi wartościami pozostałych zmiennych pierwotnych (ujemne wartości rzutowania na przedłużenie wektorów obrazujących pozostałe zmienne pierwotne).

DLA ZAINTERESOWANYCH

Wzory:
Poszczególne składowe główne są kombinacją liniową zmiennych pierwotnych:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + ... + a_{in}X_p$$

gdzie:
$X_1, X_2, ..., X_p$ – zmienne pierwotne,
$a_{i1}, a_{i2}, ..., a_{ip}$ – współczynniki i-tej składowej głównej.

Obliczenia matematyczne pozwalające wyodrębnić składowe główne sprowadzają się do wyznaczenia wartości własnych i odpowiadających im wektorów własnych z równania macierzowego postaci:

$$(M - \lambda I)a = 0$$

gdzie:
λ – wartości własne,
$a_i = (a_{i1}, a_{i2}, ..., a_{ip})$ – wektor własny odpowiadający i-tej wartości własnej,
M – macierz wariancji lub kowariancji zmiennych pierwotnych $X_1, X_2, ..., X_p$,
I – macierz identycznościowa (1 na głównej przekątnej, 0 poza nią).

Procent zmienności wyjaśniany przez i-tą składową główną wyraża się wzorem:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + ... + \lambda_p} \cdot 100\%.$$

Test Bartletta

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \quad M = I, \\ \mathcal{H}_1 &: \quad M \neq I. \end{aligned}$$

gdzie:
M – macierz wariancji lub kowariancji zmiennych pierwotnych $X_1, X_2, ..., X_p$,
I – macierz jednostkowa (1 na głównej przekątnej, 0 poza nią).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Wzory:
Statystyka testowa ma postać:

$$\chi^2 = -\left(n - 1 - \frac{2p+5}{6}\right)\sum_{i=1}^{k} \ln \lambda_i,$$

gdzie:
p – liczba zmiennych pierwotnych,
n – liczność (liczba przypadków),
λ_i – i-ta wartość własna.

Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z $p(p-1)/2$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α:

$$\begin{array}{lcl} \text{jeżeli } p \leq \alpha & \Longrightarrow & \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha & \Longrightarrow & \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{array}$$

Współczynnik Kaisera-Mayera-Olkina – wzór:

$$KMO = \frac{\sum_{i\neq j}^{p}\sum_{j\neq i}^{p} r_{ij}^2}{\sum_{i\neq j}^{p}\sum_{j\neq i}^{p} r_{ij}^2 + \sum_{i\neq j}^{p}\sum_{j\neq i}^{p} \hat{r}_{ij}^2},$$

gdzie:
r_{ij} – współczynnik korelacji pomiędzy i-tą a j-tą zmienną,
$\hat{r}_{ij}$ – współczynnik korelacji cząstkowej pomiędzy i-tą a j-tą zmienną.

27.1 ANALIZA SKUPIEŃ

Analiza skupień to szukanie w danych grup podobnych do siebie obiektów (np. klientów, pacjentów, gatunków) – tak, by obiekty wewnątrz jednej grupy (**skupienia**) były do siebie jak najbardziej podobne, a obiekty z różnych skupień jak najbardziej się różniły. Wyróżnia się dwie główne rodziny metod: **metody hierarchiczne**, budujące całą hierarchię możliwych podziałów, oraz metody niehierarchiczne, spośród których najpopularniejsza jest **metoda k-średnich**. Obie rodziny metod opierają swoje działanie na macierzy podobieństwa, czyli na informacji o tym, jak bardzo poszczególne obiekty są do siebie podobne (lub jak bardzo się różnią).

Dokładnie w ten sam sposób, w jaki grupuje się obiekty (np. osoby badane), można grupować również zmienne (cechy) – np. by wskazać zmienne niosące podobną informację. W tym rozdziale metody grupowania opisane zostaną na przykładzie grupowania obiektów.

Metody hierarchiczne

KIEDY STOSOWAĆ:

Hierarchiczną analizę skupień stosuje się, gdy chcemy zobaczyć nie tylko jeden, ostateczny podział

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

obiektów na grupy, ale całą hierarchię możliwych podziałów – od pojedynczych obiektów, poprzez coraz większe skupienia, aż po jedno skupienie zawierające wszystkie obiekty. Metoda buduje tę hierarchię krok po kroku, łącząc w każdym kroku dwa najbardziej do siebie podobne (najbliższe) obiekty lub skupienia, na podstawie macierzy podobieństwa. Dzięki temu badacz sam decyduje później, na jakim poziomie „przeciąć" powstałą hierarchię, by uzyskać interesującą go liczbę skupień.

WARUNKI STOSOWANIA

- możliwość wyliczenia odległości/podobieństwa pomiędzy obiektami – dobór odpowiedniej miary zależy od skali pomiarowej analizowanych zmiennych (patrz Macierz podobieństwa),
- standaryzacja zmiennych przed wyliczeniem macierzy podobieństwa, by żadna z nich (np. wyrażona dużo większymi liczbami) nie zdominowała pozostałych przy budowaniu skupień,
- brak rażąco nietypowych obserwacji (odstających), które mogłyby w istotny sposób zniekształcić odległości pomiędzy pozostałymi obiektami.

Uwaga! Chcąc zapewnić zrównoważony wpływ wszystkich zmiennych na elementy macierzy podobieństwa, dane należy poddać standaryzacji wybierając odpowiednią opcję w oknie analizy. Brak standaryzacji daje większy wpływ na uzyskany wynik zmiennym wyrażonym większymi liczbami.

DEFINICJE MIAR

Podobieństwo obiektów

W toku prac związanych z analizą skupień zasadniczą rolę odgrywają miary podobieństwa lub odległości. Wzajemne podobieństwo obiektów umieszczane jest w macierzy podobieństwa. Duża różnorodność metod wyznaczania odległości/niepodobieństwa między obiektami pozwala na wybór takich miar, które najlepiej odzwierciedlają rzeczywiste relacje. Szerzej miary odległości i podobieństwa opisane są w dziale Macierz podobieństwa.

Analiza skupień opiera swoje działania na wyszukiwaniu skupień wewnątrz macierzy podobieństwa. Macierz taka jest budowana w trakcie wykonywania analizy skupień. By analiza skupień przyniosła pożądane skutki, przy wyborze sposobu wyliczania odległości należy pamiętać, że większe wartości liczbowe w macierzy podobieństwa mają wskazywać na większe zróżnicowanie obiektów, a wartości mniejsze na ich podobieństwo.

Uwaga! By zwiększyć wpływ wybranych zmiennych na elementy macierzy podobieństwa,

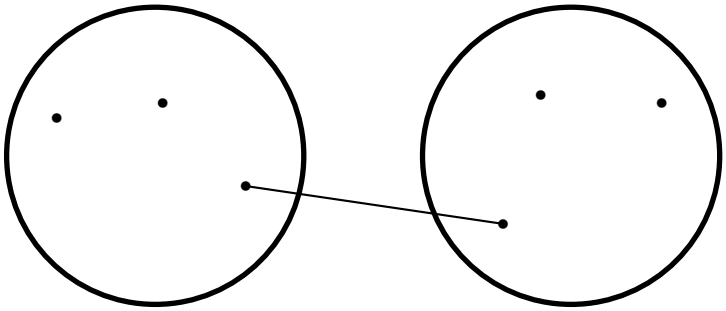
należy wskazać odpowiednie wagi przy ustawianiu sposobu definiowania odległości pamiętając jednocześnie o wystandaryzowaniu danych.

Np. Dla osób chcących zaopiekować się psem pogrupowanie psów zgodnie z wielkością, umaszczeniem, długością ogona, charakterem, rasą itp. ułatwi dokonanie wyboru. Jednak, identyczne traktowanie wszystkich cech może spowodować umieszczenie zupełnie niepodobnych psów w jednej grupie. Natomiast dla większości z nas ważniejsza jest wielkość psa i jego charakter niż długość jego ogona, dlatego w grupowaniu należałoby ustawić miary podobieństwa tak, by to właśnie wielkość i charakter miały największe znaczenie w budowaniu skupień.

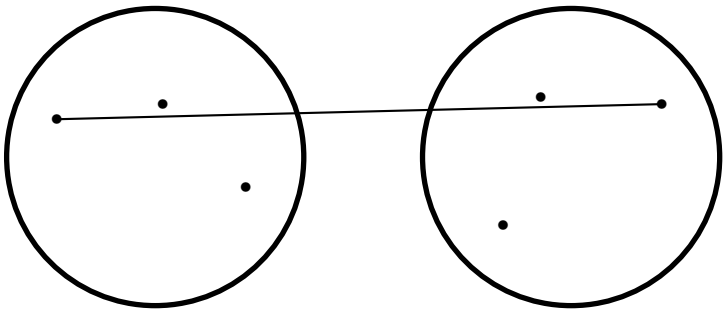
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Metody wiązania obiektów i skupień

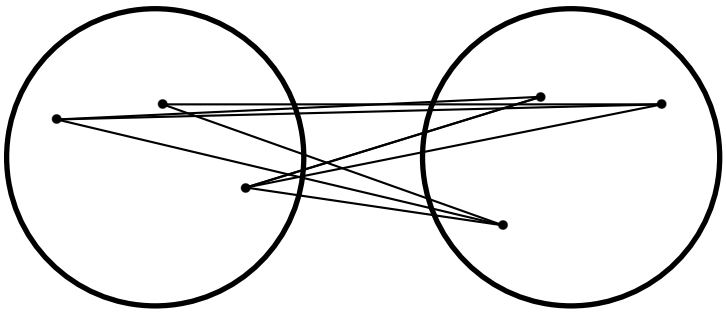
- Metoda pojedynczego wiązania (najbliższe sąsiedztwo) – odległość pomiędzy skupieniami określona jest poprzez odległość tych obiektów każdego skupienia, które znajdują się najbliżej siebie.



- Metoda pełnego wiązania (najdalsze sąsiedztwo) – odległość pomiędzy skupieniami określona jest poprzez odległość tych obiektów każdego skupienia, które znajdują się najdalej siebie.



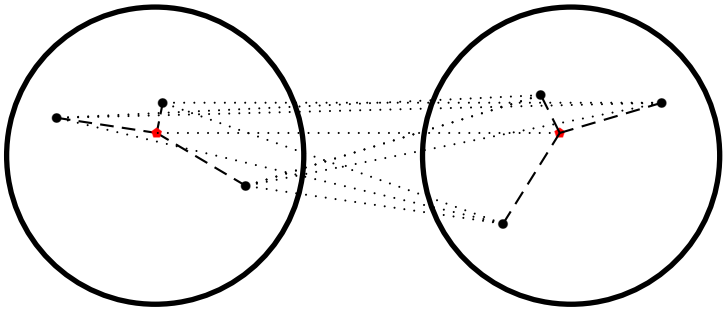
- Metoda średnich połączeń – odległość pomiędzy skupieniami określona jest poprzez średnią odległość pomiędzy wszystkimi parami obiektów zlokalizowanych w obrębie dwóch różnych skupień.



- Metoda średnich połączeń ważonych – analogicznie do metody średnich połączeń polega na wyliczeniu średniej odległości, ale średnia ta ważona jest poprzez liczbę elementów każdego skupienia. W rezultacie powinniśmy wybierać tę metodę, gdy oczekujemy uzyskać skupienia o podobnych licznościach.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Metoda Warda – opiera się na zasadzie analizy wariancji – wylicza różnicę między sumami kwadratów odchyleń odległości poszczególnych obiektów od środka ciężkości skupień, do których te obiekty należą. Metoda ta wybierana jest najczęściej ze względu na jej dość uniwersalny charakter.



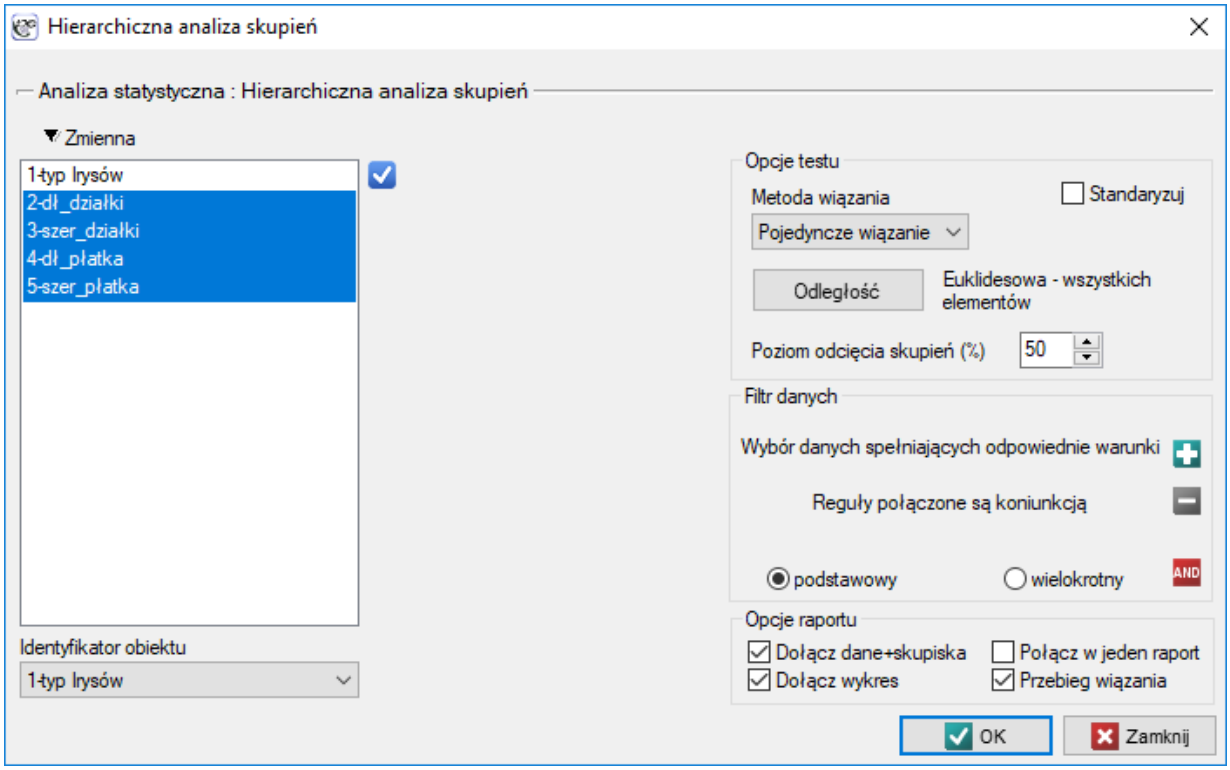
Dendrogram

Wynik analizy skupień prowadzonej metodą hierarchiczną przedstawia się przy pomocy dendrogramu. **Dendrogram** jest formą drzewa wskazującego związki pomiędzy poszczególnymi obiektami uzyskane z analizy macierzy podobieństwa. Poziom odcięcia dendrogramu decyduje o liczbie skupień, na które chcemy dzielić zgromadzone obiekty. Wybór sposobu odcięcia określamy podając w procentach długość wiązania, przy którym nastąpi cięcie, gdzie 100% stanowi długość ostatniego i jednocześnie najdłuższego wiązania w dendrogramie.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji hierarchicznej analizy skupień wywołujemy poprzez menu Statystyki zaawansowane→Grupowanie i Redukcja→Hierarchiczna analiza skupień.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

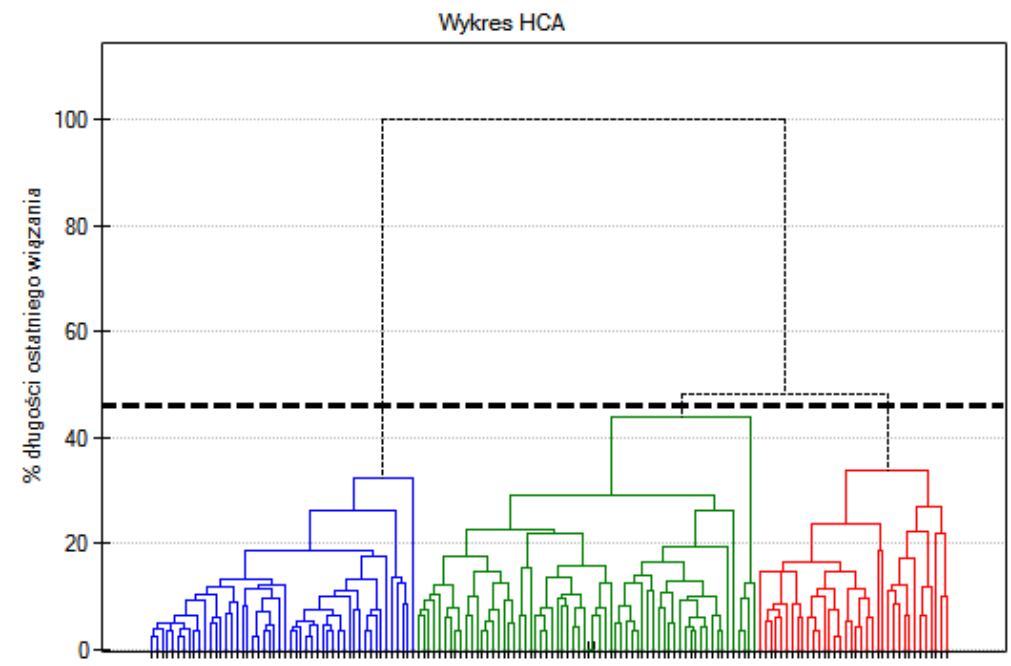
Przykład (27.1) c.d. *(plik iris.pqs)*
Analiza przeprowadzona zostanie na klasycznym zestawie danych, dotyczącym podziału kwiatów irysa na 3 odmiany na podstawie szerokości oraz długości płatków i działek kielicha (R.A. Fishera 1936[59]). Ponieważ w tym zestawie danych znajduje się informacja o rzeczywistej odmianie każdego kwiatu, po przeprowadzonej analizie skupień istnieje możliwość określenia dokładności dokonanego podziału.

Przydziału kwiatów do poszczególnych grup dokonujemy na podstawie kolumn od 2 do 5. Wybieramy sposób wyliczania odległości np. odległość Euklidesową i metodę wiązania np. średnią. Podanie poziomu odcięcia skupień pozwoli na takie odcięcie dendrogramu by powstały skupienia – w przypadku tej analizy chcemy uzyskać 3 skupienia i by to osiągnąć zmieniamy poziom odcięcia na 45%. Do raportu dołączamy również dane+skupienia.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

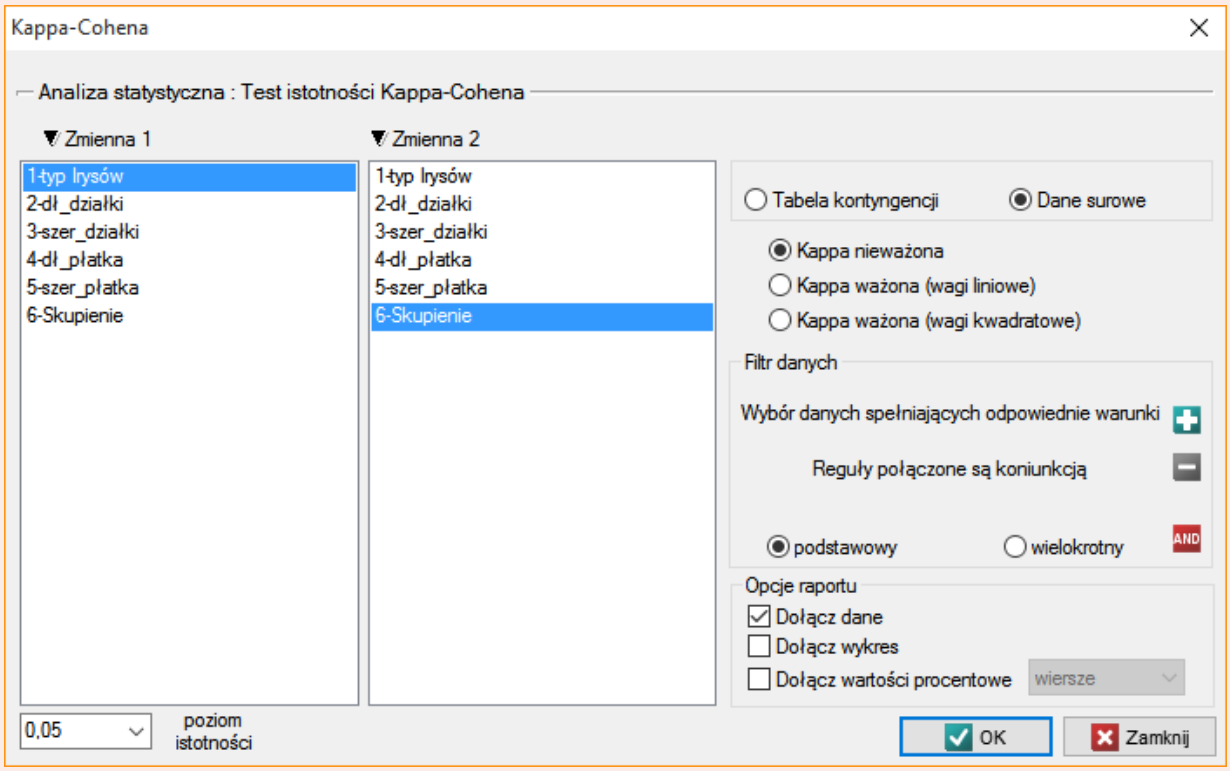
Hierarchiczna analiza skupień	
Czas analizy	0,59 sek.
Analizowane zmienne	dł_działki szer_działki dł_płatka szer_płatka
Liczność	150
Standaryzacja	Nie
Metoda wiązania	Średnia
Poziom odcięcia skupień (%)	45
Odległość	Euklidesowa - wszystkich ele
Liczba skupień	3
Liczba elementów skupienia 1	50
Liczba elementów skupienia 2	64
Liczba elementów skupienia 3	36

Na wykresie typu dendrogram, przedstawiono kolejność wiązań i ich długość.



By zbadać czy wyodrębnione skupienia stanowią 3 rzeczywiste odmiany kwiatów irysa, możemy skopiować kolumnę zawierającą informację o przynależności do skupień z raportu i wkleić do arkusza danych. Podobnie jak skupienia, odmiany opisane są również liczbowo poprzez Kody/Etykiety/Format, dlatego z łatwością można przeprowadzić analizę zgodności. Zgodność naszych wyników z rzeczywistą przynależnością danego kwiatu do odpowiedniego gatunku sprawdzimy metodą Kappa Cohena.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Dla przedstawionego przykładu obserwowaną zgodność przedstawia tabela:

Dane :

	1	2	3
setosa	50	0	0
versicolor	0	50	0
virginica	0	14	36

Kappa Cohena $= 0.86$ **(istotna statystycznie)**

Odmiana virginica może być mylona z versicolor, dlatego obserwujemy tu 14 błędnych zaklasyfikowań. Sam współczynnik zgodności Kappa jest jednak istotny statystycznie i wynosi 0.86, co świadczy o dużej zgodności uzyskanych skupień z rzeczywistą odmianą kwiatów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Test istotności Kappa-Cohena	
Czas analizy	0,13 sek.
Analizowane zmienne	typ Irysów;Skupienie
Poziom istotności	0,05
Liczność - ilość par	150
Typ	Kappa nieważona
Współczynnik Kappa	0,86
Błąd dla Kappa	0,035135
-95% CI dla współczynnika Kappa	0,791137
+95% CI dla współczynnika Kappa	0,928863
Błąd dla rozkładu Kappa	0,056976
Statystyka Z	15,094173
Wartość p (asymptotyczne)	<0.000001

Metoda k-średnich

KIEDY STOSOWAĆ:

Metodę k-średnich *(ang. k-means)* stosuje się, gdy z góry wiadomo (lub zakłada się), na ile grup (k) chcemy podzielić obiekty, a interesuje nas przede wszystkim jeden, ostateczny podział – bez budowania całej hierarchii, jak w metodach hierarchicznych. Metoda wzoruje swoje działanie na algorytmie zaproponowanym początkowo przez Stuarta Lloyda i opublikowanym w roku 1982 [109]. Obiekty przenoszone są pomiędzy skupieniami tak, by zróżnicowanie obiektów wewnątrz każdego skupienia było jak najmniejsze, a odległości między skupieniami jak największe.

WARUNKI STOSOWANIA

- pomiar zmiennych umożliwiający wyliczenie odległości euklidesowej pomiędzy obiektami (algorytm bazuje na macierzy odległości euklidesowych),
- z góry założona liczba k poszukiwanych skupień,
- standaryzacja zmiennych, by żadna z nich nie zdominowała pozostałych w budowie skupień (analogicznie jak w metodach hierarchicznych).

DEFINICJE MIAR

W metodzie tej obiekty dzieli się na z góry założoną liczbę k skupień. Początkowe skupienia poprawia się w toku wykonywania procedury aglomeracji, przenosząc pomiędzy nimi obiekty tak by zróżnicowanie obiektów wewnątrz skupienia było jak najmniejsze a odległości skupień jak największe. Algorytm działa w oparciu o macierz odległości euklidesowych pomiędzy obiektami, a parametrami niezbędnymi

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

w procedurze aglomeracji metody k-średnich są: centra startowe i kryterium stopu. Centra startowe są to obiekty od których algorytm rozpocznie budowę skupień, a kryterium stopu to określenie sposobu zatrzymania algorytmu.

Centra startowe
Wybór centrów startowych ma zasadniczy wpływ na zbieżność algorytmu k-średnich dla uzyskania odpowiednich skupień. Centra startowe mogą być wybrane na dwa sposoby:

k-means++ – optymalny dobór punktów startowych, wykorzystujący algorytm k-means++ zaproponowany w 2007 roku przez Davida Arthura i Sergeia Vassilvitskiego [1]. Gwarantuje on uzyskanie optymalnego rozwiązania algorytmu k-średnich przy możliwie niewielu iteracjach. Algorytm wykorzystuje element losowości, dlatego uzyskiwane wyniki mogą się nieco różnić przy kolejnych uruchomieniach analizy. Jeśli dane nie formułują się w naturalne klastery, użycie k-means++ da zupełnie inne wyniki w kolejnych uruchomieniach analizy k-średnich. Duża powtarzalność wyników świadczy natomiast o możliwości dokonania dobrego podziału na oddzielne skupienia.

n pierwszych – pozwala by użytkownik wskazał punkty będące centrami startowymi poprzez ustawienie tych obiektów na pierwszych miejscach w tabeli danych.

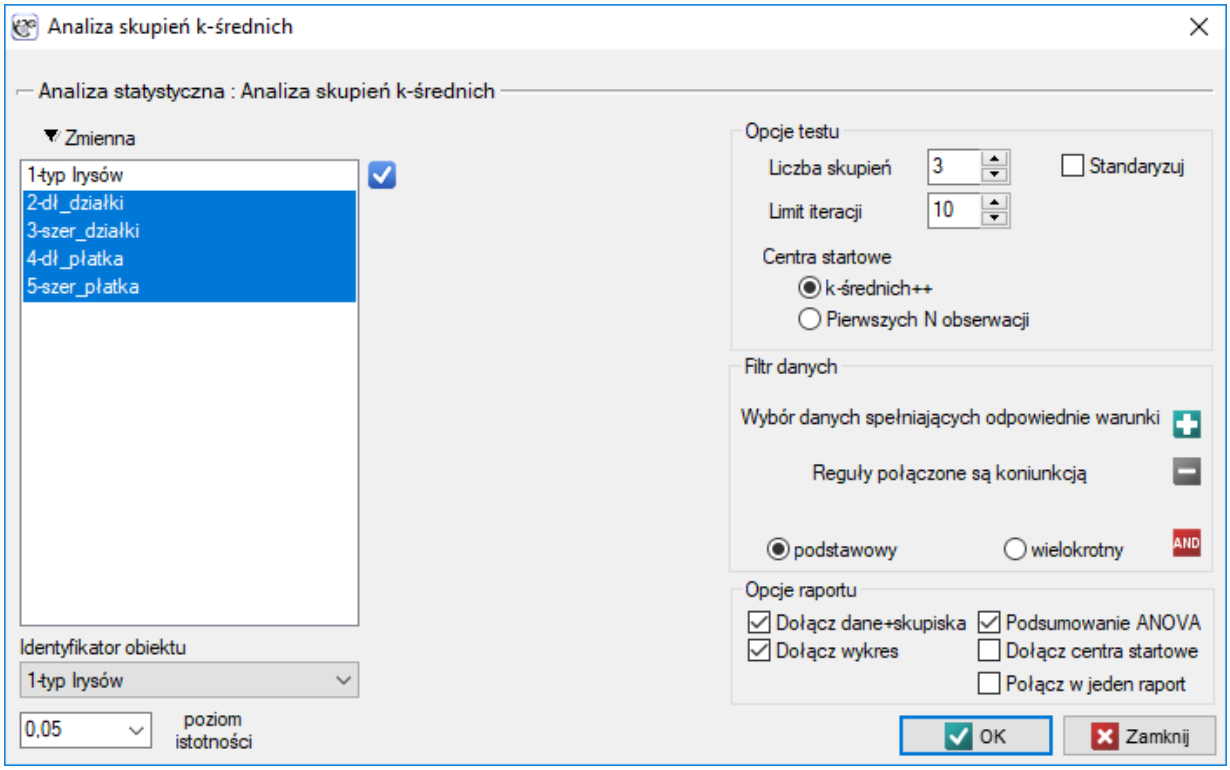
Kryterium stopu jest to moment, w którym nie zmienia się przynależność punktów do klas lub liczba powtórzeń kroków procedury aglomeracji osiągnie zadaną przez użytkownika liczbę iteracji.

Ze względu na sposób działania algorytmu analizy skupień k-średnich, naturalną jej konsekwencją jest porównanie powstałych skupień przy pomocy jednoczynnikowej analizy wariancji (**ANOVA**) dla grup niezależnych.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji analizy skupień k-średnich wywołujemy poprzez menu Statystyki zaawansowane→Grupowanie i Redukcja→Analiza skupień k-średnich.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

Przykład (27.1) c.d. *(plik iris.pqs)*
Analiza przeprowadzona zostanie na klasycznym zestawie danych, dotyczącym podziału kwiatów irysa na 3 odmiany na podstawie szerokości oraz długości płatków i działek kielicha (R.A. Fishera 1936[59]).

Przydziału kwiatów do poszczególnych grup dokonujemy na podstawie kolumn od 2 do 5. Wskazujemy również, że badane kwiaty chcemy podzielić na 3 grupy.

Analiza skupień k-średnich	
Czas analizy	0,10 sek.
Analizowane zmienne	
Poziom istotności	0,05
Liczność	150
Standaryzacja	Nie
Limit iteracji	10
Centra startowe na podstawie	k-średnich++
Liczba skupień	3
Liczba elementów skupienia 1	39
Liczba elementów skupienia 2	50
Liczba elementów skupienia 3	61

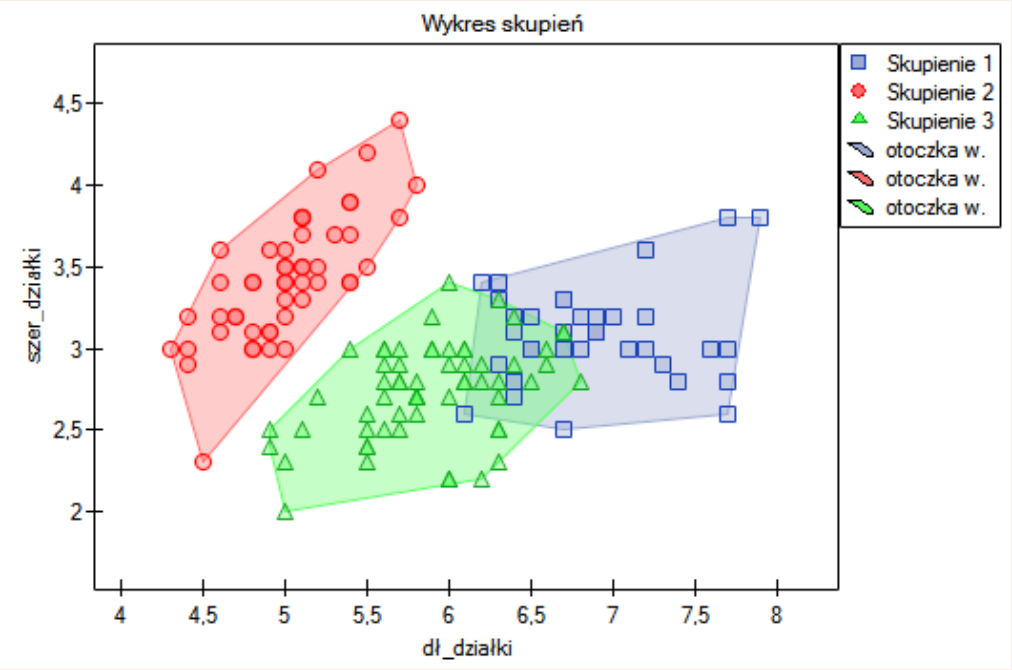
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jednoczynnikowa ANOVA

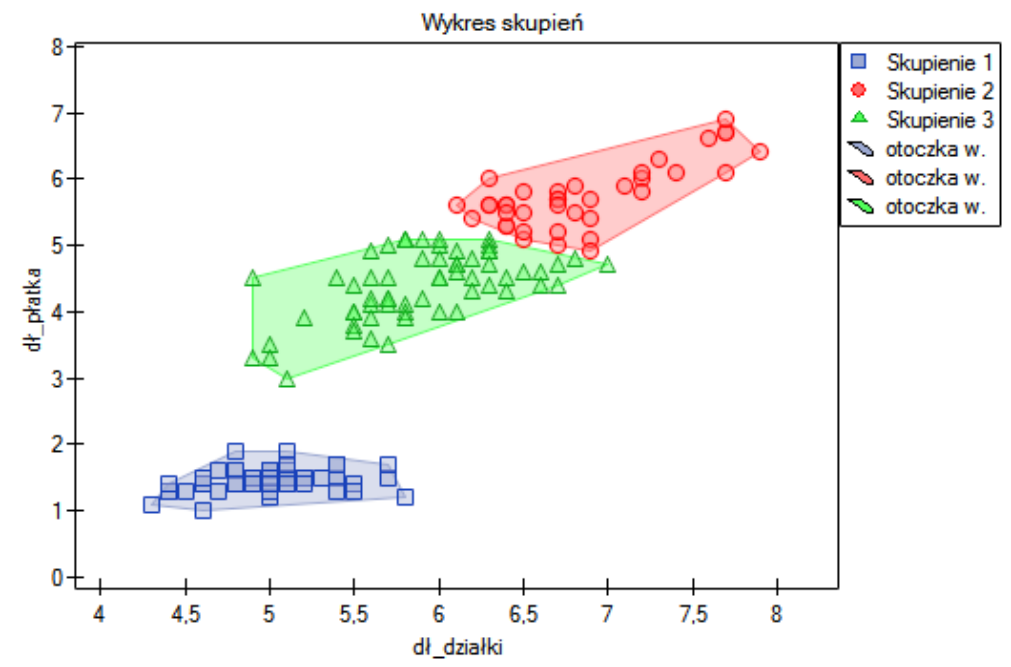
	dł_działki	szer_działk	dł_płatka	szer_płatka
SS[BG]	74,9796	12,9894	437,2549	77,2911
SS[WG]	27,1887	15,3176	27,0705	9,2788
df[BG]	2	2	2	2
df[WG]	147	147	147	147
Statystyka F	202,6943	62,3283	1187,2034	612,2429
Wartość p	<0.0001	<0.0001	<0.0001	<0.0001

Wynik ANOVA istotny statystycznie dla każdego z 4 wymiarów

W rezultacie przeprowadzonej analizy utworzono 3 skupienia, które różnią się istotnie statystycznie w każdym badanym wymiarze, czyli dla szerokości płatka, długości płatka, szerokości kielicha jak i długości kielicha. Różnicę można zaobserwować na wykresach, gdzie przedstawiamy w wybranych dwóch wymiarach przynależność poszczególnych punktów do skupienia:



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Powtarzając analizę możemy uzyskać nieco inne wyniki, ale uzyskana zmienność nie będzie duża. Świadczy to o tym, że dane poddane analizie formułują się w naturalne skupienia i przeprowadzanie analizy skupień jest w tym przypadku uzasadnione.

Uwaga! Po kilkukrotnym przeprowadzeniu analizy możemy wybrać najbardziej interesujący nas wynik, a następnie ustawić te dane, które stanowią centra startowe na początku arkusza – wówczas analiza przeprowadzona w oparciu o centra startowe wybierane jako N pierwszych obserwacji będzie dawała stale ten wybrany przez nas efekt.

Uwaga! By dowiedzieć się czy skupienia stanowią 3 rzeczywiste odmiany kwiatów irysa, możemy skopiować informację o przynależności do skupienia z raportu i wkleić do arkusza danych. Zgodność naszych wyników z rzeczywistą przynależnością danego kwiatu do odpowiedniego gatunku możemy sprawdzić analogicznie jak dla hierarchicznej analizy skupienia.

DLA ZAINTERESOWANYCH

Procedura aglomeracyjna – metody hierarchiczne

1. Postępując zgodnie z wskazaną **metodą wiązania**, algorytm znajduje w macierzy **podobieństwa** parę podobnych obiektów i łączy je w skupienie;
2. Wymiar macierzy podobieństwa zostaje zredukowany o jeden (dwa obiekty zastąpiono jednym) a odległości znajdujące się w macierzy wyliczone są ponownie;
3. Kroki 2-3 są powtarzane aż do uzyskania jednego skupienia zawierającego wszystkie obiekty.

Procedura aglomeracyjna – metoda k-średnich

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1. Wybór centrów startowych;
2. Bazując na macierzy podobieństwa algorytm przypisuje każdy obiekt do najbliższego centrum;
3. Dla uzyskanych skupień wyznacza się poprawione centra;
4. Kroki 2-3 są powtarzane aż do momentu spełnienia kryterium stopu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

28 ANALIZA PRZEŻYCIA

Analiza przeżycia pozwala ocenić, jak długo trwa do wystąpienia interesującego badacza zdarzenia – np. jak długo żyją pacjenci po operacji, jak długo trwa remisja choroby, po jakim czasie następuje awaria urządzenia. Poza medycyną (skąd pochodzi nazwa) wykorzystywana jest też w innych dziedzinach, gdzie funkcjonuje pod nazwami: analiza niezawodności (*ang. reliability analysis*), analiza trafności (*ang. duration analysis*) czy analiza historii zdarzeń (*ang. event history analysis*).

Podstawowym narzędziem opisowym są tu **tabele trwania życia** i **krzywe Kaplana-Meiera**, pozwalające oszacować, jaka część badanych „przeżywa" do danego momentu czasu. Gdy interesuje nas porównanie czasu przeżycia pomiędzy grupami (np. między różnymi metodami leczenia), stosuje się **porównanie krzywych przeżycia**. Wpływ wielu zmiennych jednocześnie na czas przeżycia bada się natomiast za pomocą **modeli regresji**, przede wszystkim regresji proporcjonalnego hazardu Coxa.

Podstawowe pojęcia analizy przeżycia wyjaśnimy na przykładzie długości życia pacjentów po przeszczepie serca:

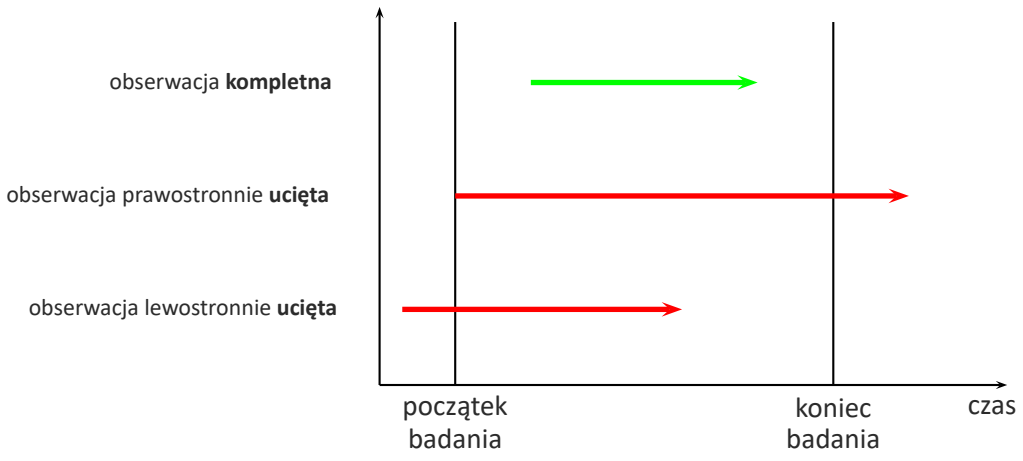
Zdarzenie — interesująca badacza zmiana, np. śmierć.

Czas przeżycia — okres, jaki upłynął między stanem początkowym a wystąpieniem zdarzenia, np. długość życia pacjenta po przeszczepie serca.

Uwaga! Do analizy trzeba przygotować jedną kolumnę z gotowym, wyliczonym czasem. Gdy dysponujemy dwoma datami: początkową i końcową, czas jaki upłynął między nimi wyliczamy przed analizą, korzystając z formuł arkusza.

Obserwacje ucięte (*ang. censored*) — obserwacje, dla których nie mamy pełnej informacji o czasie przeżycia. Rozróżniamy tu:

- **obserwację kompletną** – znana jest zarówno data przeszczepu, jak i data śmierci pacjenta, więc można wyznaczyć dokładny czas przeżycia po przeszczepie;
- **obserwację uciętą prawostronnie** – nie jest znana data śmierci pacjenta (chory żyje jeszcze w chwili zakończenia badania), więc nie można ustalić dokładnego czasu przeżycia;
- **obserwację uciętą lewostronnie** – nie jest znana data przeszczepu serca, choć wiadomo, że nastąpił on przed rozpoczęciem badania, więc również w tym przypadku nie można ustalić dokładnego czasu przeżycia.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uwaga! Zakończenie badania oznacza zakończenie obserwacji danego pacjenta – nie zawsze jest to ten sam moment dla wszystkich badanych. Zakończeniem obserwacji może być np. moment, w którym tracimy kontakt z pacjentem (nie znamy więc jego dalszego czasu przeżycia). Analogicznie, rozpoczęcie badania nie musi dotyczyć wszystkich pacjentów w tym samym czasie.

Tabele przeżycia

KIEDY STOSOWAĆ:

Tabele trwania życia (tabele przeżycia) to najprostszy sposób podsumowania czasu przeżycia badanej grupy – oś czasu dzielimy na równe przedziały (np. co 3 lata) i dla każdego z nich sprawdzamy, ilu badanych do niego dotrwało, ilu zmarło, a ilu zostało utraconych z obserwacji (ucięte). Ich zaletą jest łatwość interpretacji, wadą – konieczność arbitralnego podziału czasu na przedziały.

DEFINICJE MIAR

Dla każdego przedziału czasu program PQStat wylicza:

- **liczbę przypadków wchodzących** – liczbę osób, które przeżyły do początku danego przedziału;
- **liczbę przypadków uciętych** – liczbę osób w danym przedziale zdefiniowanych jako przypadki ucięte;
- **liczbę przypadków zagrożonych** – liczbę przypadków wchodzących do danego przedziału pomniejszoną o połowę liczby przypadków uciętych w tym przedziale;
- **liczbę przypadków kompletnych** – liczbę osób, które doświadczyły zdarzenia (np. zmarły) w danym przedziale;
- **proporcję przypadków kompletnych** – stosunek liczby przypadków kompletnych (zgonów) w danym przedziale do liczby przypadków zagrożonych w tym przedziale;
- **proporcję przypadków przeżywających** – obliczaną jako 1 minus proporcja przypadków kompletnych w danym przedziale;
- **skumulowaną proporcję przeżywających (funkcję przeżycia)** – prawdopodobieństwo przeżycia do danego czasu; ponieważ by przeżyć kolejny przedział, trzeba przeżyć wszystkie poprzednie, prawdopodobieństwo to wylicza się jako iloczyn wszystkich poprzednich proporcji przypadków przeżywających

  ± błąd standardowy funkcji przeżycia;
- **gęstość prawdopodobieństwa** – oszacowane prawdopodobieństwo doświadczenia zdarzenia (zgonu) w danym przedziale, w przeliczeniu na jednostkę czasu

  ± błąd standardowy gęstości prawdopodobieństwa;
- **stopę hazardu** – prawdopodobieństwo (w przeliczeniu na jednostkę czasu), że przypadek, który przeżył do początku danego przedziału, doświadczy zdarzenia (zgonu) właśnie w tym przedziale

  ± błąd standardowy stopy hazardu.

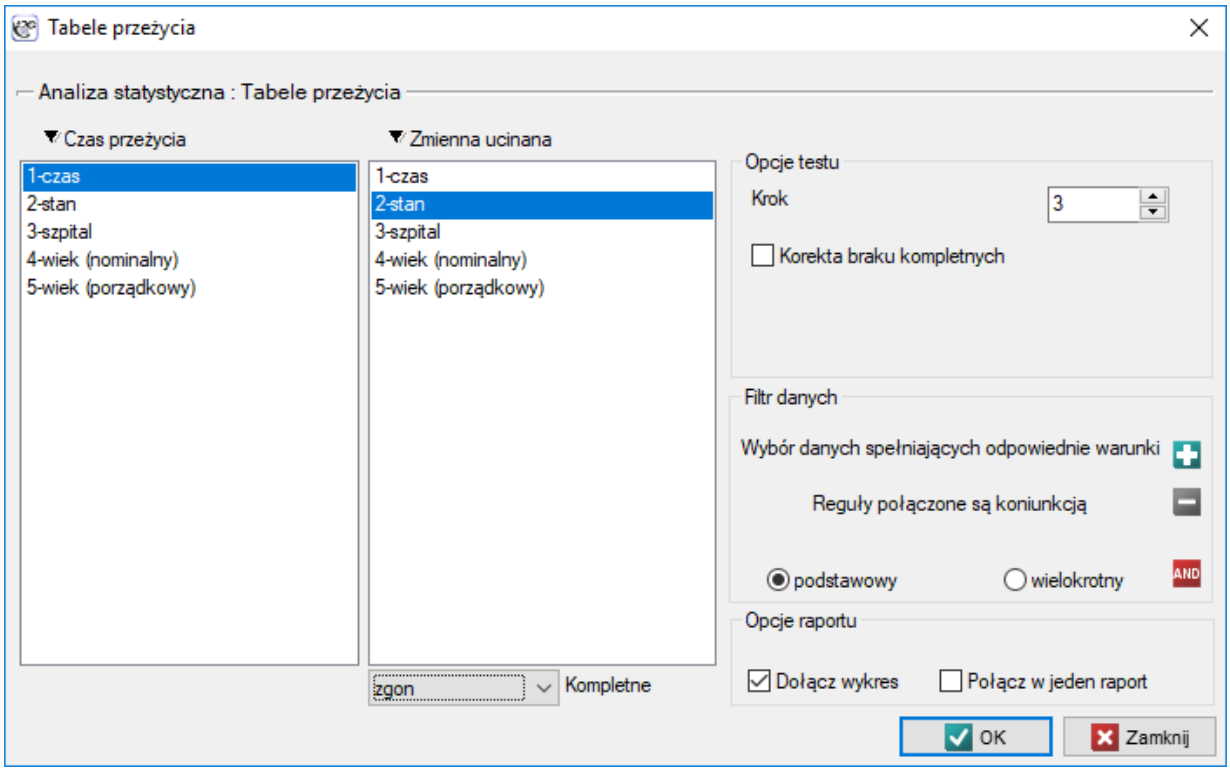
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Uwaga! Gdy w którymś przedziale czasu nie ma żadnej obserwacji kompletnej, program umożliwia zastosowanie korekty – zerowa liczba przypadków kompletnych zastępowana jest wówczas wartością 0.5.

Uzyskane informacje można zilustrować trzema wykresami: wykresem funkcji przeżycia, wykresem gęstości prawdopodobieństwa oraz wykresem stopy hazardu.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji tabel trwania życia wywołujemy poprzez menu Statystyka→Analiza przeżycia→Tabele przeżycia.



Tabele trwania życia tworzy się dla podanych przez badacza przedziałów czasu przeżycia o równej rozpiętości – przedziały te definiuje się, podając ich krok.

WYKONANIE I INTERPRETACJA

Przykład 28.1. (plik przeszczep.pqs)
Badano długość życia chorych po przeszczepie wątroby. Przez okres 21 lat poddano obserwacji grupę 89 pacjentów, których wiek w chwili przeszczepu należał do przedziału $\langle 45\text{lat}; 60\text{lat}\rangle$. Fragment zebranych danych przedstawia poniższa tabela:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

czas	stan	szpital	wiek (nominalny)	wiek (porządkowy)
14	zgon	szpital 1	<45; 50)	1
21	żyje	szpital 1	<45; 50)	1
4	zgon	szpital 1	<50; 55)	2
4	żyje	szpital 1	<50; 55)	2
5	zgon	szpital 1	<50; 55)	2
6	żyje	szpital 1	<50; 55)	2
6	żyje	szpital 1	<50; 55)	2
9	zgon	szpital 1	<50; 55)	2
16	żyje	szpital 1	<50; 55)	2

Dane kompletne w tej analizie to te, o których mamy pełną informację o długości życia po przeszczepie, czyli opisane jako „zgon" (są to 53 osoby, co stanowi 59.55%). Dane ucięte to te, o których tej informacji nie mamy, ponieważ w chwili zakończenia badania żyją (jest to 36 osób, co stanowi 40.45%). Budujemy tablice trwania życia tych pacjentów, tworząc 3-letnie przedziały czasowe:

Tabela trwania życia

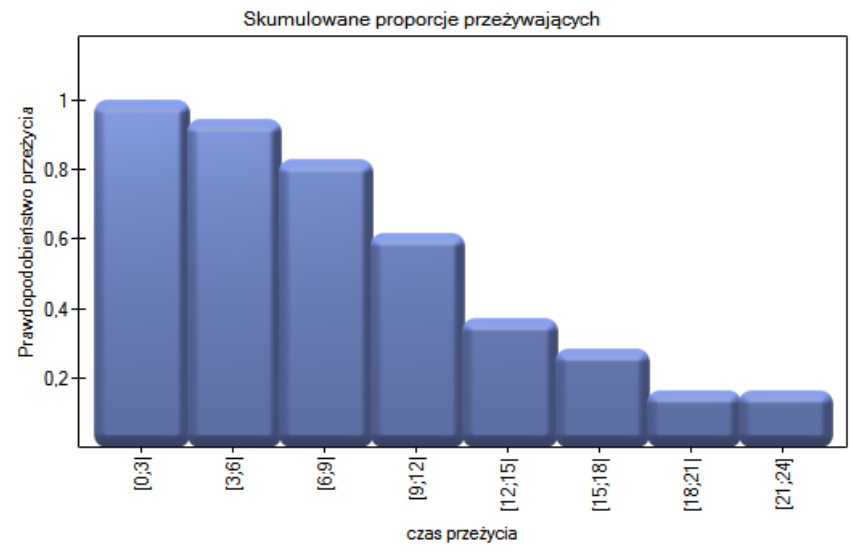
przedział	ucięte	kompletne	wchodzące	zagrożone	prop. kom	prop. prze	skumulowa	gęstość pr	stopa haza	błąd std. s	błąd std. g	błąd std. s
[0;3)	0	5	89	89	0,05618	0,94382	1	0,018727	0,019268	0	0,008136	0,008613
[3;6)	5	10	84	81,5	0,122699	0,877301	0,94382	0,038602	0,043573	0,024408	0,011477	0,01375
[6;9)	14	16	69	62	0,258065	0,741935	0,828014	0,071227	0,098765	0,040436	0,015727	0,024419
[9;12)	7	14	39	35,5	0,394366	0,605634	0,614333	0,080757	0,163743	0,05493	0,018283	0,042422
[12;15)	3	4	18	16,5	0,242424	0,757576	0,372061	0,030066	0,091954	0,060381	0,013965	0,045538
[15;18)	3	4	11	9,5	0,421053	0,578947	0,281864	0,03956	0,177778	0,060276	0,017265	0,08567
[18;21)	2	0	4	3	0	1	0,163185	0	0	0,057065	0	0
[21;24]	2	0	2	1	0	1	0,163185			0,057065		

Dla każdego 3-letniego przedziału możemy zinterpretować uzyskane w tabeli wyniki – przykładowo dla osób, które żyją przynajmniej 9 lat od przeszczepu, a zatem weszły do przedziału [9;12):

- liczba osób, które przeżyły do 9 lat od przeszczepu to 39;
- jest 7 osób, o których wiemy, że żyły przynajmniej 9-12 lat – informacje o nich były zbierane, ale nie wiemy, czy żyły dłużej, gdyż po tym czasie wypadły z badania;
- liczba osób zagrożonych zgonem w tym przedziale to 36;
- jest 14 osób, o których wiemy, że zmarły od 9 do 12 lat po przeszczepie;
- 39.4% narażonych pacjentów zmarło 9-12 lat po przeszczepie;
- 60.6% narażonych pacjentów żyło 9-12 lat po przeszczepie;
- procent przeżywających do 9 lat po przeszczepie to 61.4% $\pm$ 5%;
- 0.08 $\pm$ 0.02 – to prawdopodobieństwo zgonu dla każdego roku z przedziału 9-12;
- 0.16 $\pm$ 0.04 – to warunkowe, roczne prawdopodobieństwo zgonu dla przypadku, który przeżył do początku przedziału 9-12.

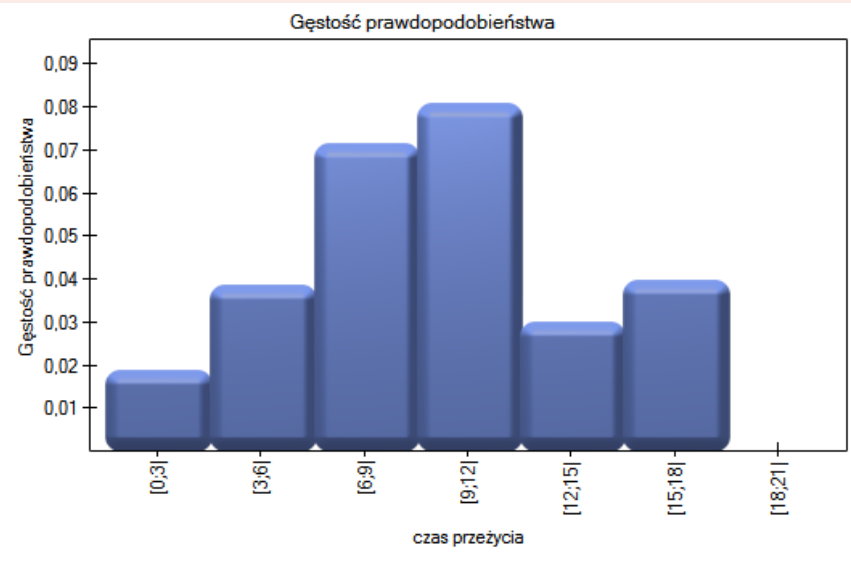
Wyniki przedstawimy na kilku wykresach:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

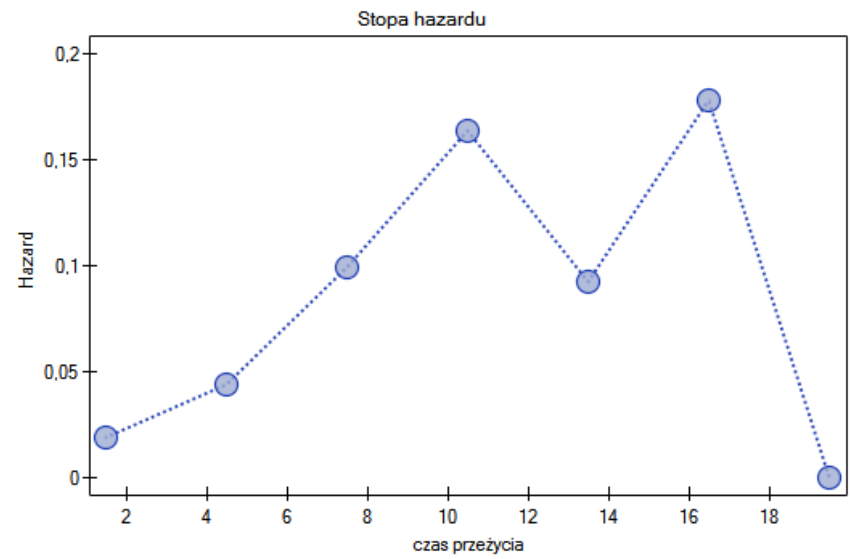


Funkcja przeżycia opada stopniowo, bez gwałtownego załamania

Prawdopodobieństwo przeżycia maleje wraz z upływem czasu, jaki minął od przeszczepu, ale nie obserwujemy okresu, w którym gwałtownie wzrastałoby prawdopodobieństwo zgonu.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krzywe Kaplana-Meiera

KIEDY STOSOWAĆ:

Krzywe Kaplana-Meiera pozwalają na ocenę czasu przeżycia bez konieczności arbitralnego grupowania obserwacji w przedziały, tak jak dzieje się to w tabelach przeżycia. Estymator ten wprowadzili Kaplan i Meier (1958)[93]. Wykres funkcji przeżycia Kaplana-Meiera ma postać funkcji schodkowej.

DEFINICJE MIAR

Podobnie jak dla tablic przeżycia, wyliczamy tutaj **funkcję przeżycia**, czyli prawdopodobieństwo przeżycia do danego czasu. Na podstawie błędu standardowego (formuła Greenwooda) i transformacji logarytmicznej (log-log) budowane są przedziały ufności wokół tej krzywej.

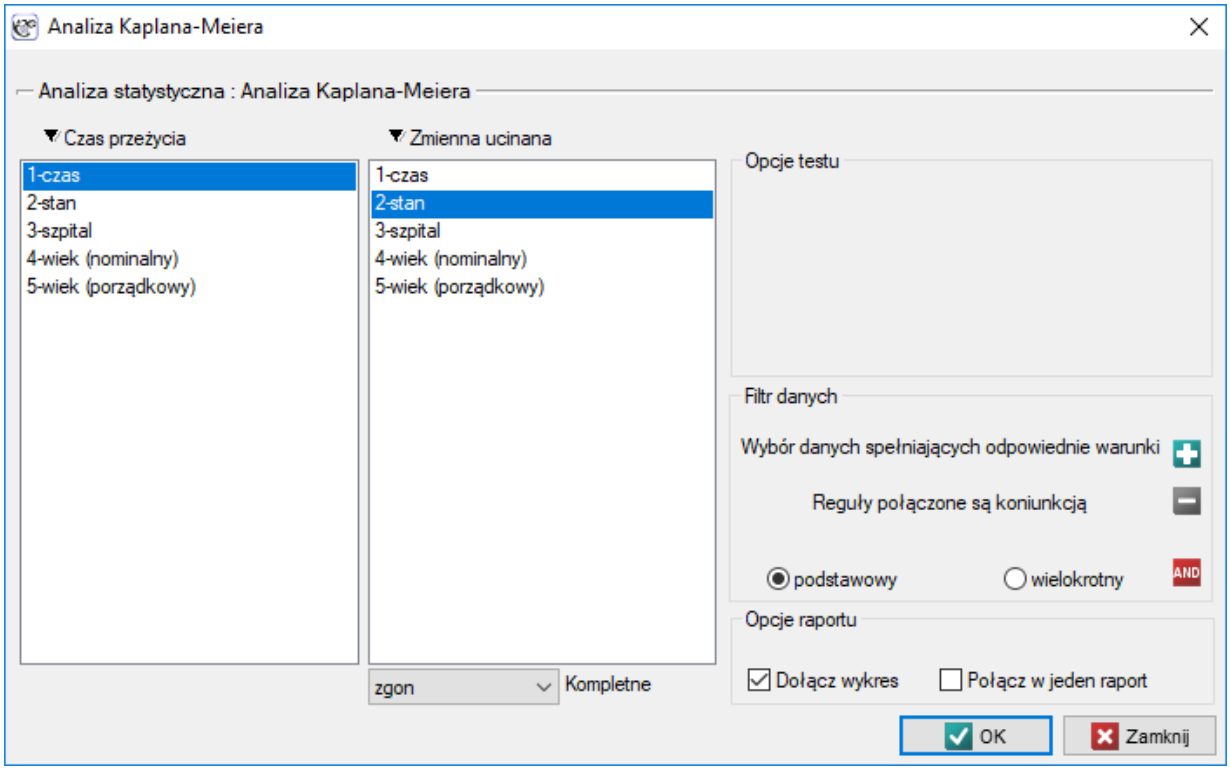
Punkt czasu, przy którym wartość funkcji przyjmuje 0.5, to **mediana czasu przeżycia**. Mediana ta wskazuje na 50% ryzyko wystąpienia zgonu, czyli przewiduje, że u połowy pacjentów zgon nastąpi w ciągu wskazanego czasu. Zarówno mediana, jak i inne percentyle wyznaczane są jako najkrótszy czas przeżycia, dla którego funkcja przeżycia jest mniejsza lub równa danemu percentylowi. Dla mediany wyznaczany jest przedział ufności metodą „test-based" Brookmeyera i Crowley (1982)[29].

Średnia czasu przeżycia wyznaczana jest jako pole pod krzywą przeżycia. Dane dotyczące czasu przeżycia są zwykle mocno skośne, dlatego w analizie przeżycia mediana jest zwykle lepszą miarą tendencji centralnej niż średnia.

JAK USTAWIĆ ANALIZĘ

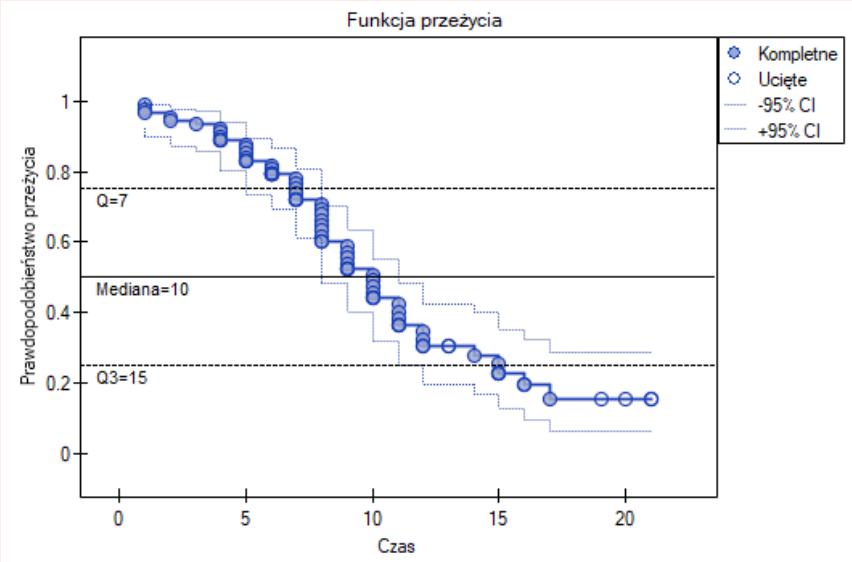
Okno z ustawieniami opcji krzywej Kaplana-Meiera wywołujemy poprzez menu Statystyka→Analiza przeżycia→Analiza Kaplana-Meiera.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

Przykład (28.1) c.d. *(plik przeszczep.pqs)*
Przedstawimy długość życia po przeszczepie wątroby przy pomocy krzywej Kaplana-Meiera:



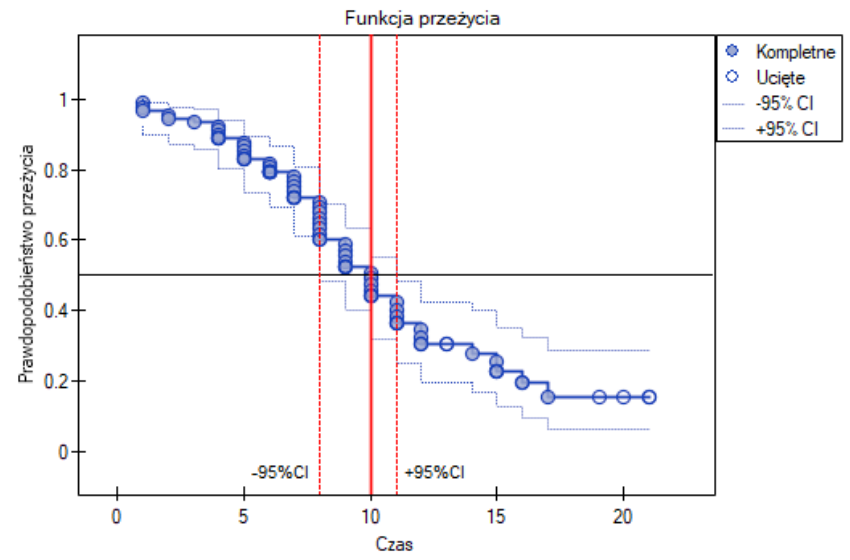
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Analiza Kaplana-Meiera	
Czas analizy	0,10 sek.
Analizowane zmienne	czas;stan
Zmienna ucinana	stan(zgon;żyje)
Liczność	89
Kompletne	zgon
Liczność	53
Procent	59,55%
Ucięte	żyje
Liczność	36
Procent	40,45%
Czas przeżycia	
Dolny kwartyl	7
Mediana	10
Górny kwartyl	15
Średnia	10,954902442

Funkcja przeżycia nie opada gwałtownie zaraz po przeszczepie. Wnioskujemy stąd, że początkowy okres po transplantacji nie jest szczególnie obarczony ryzykiem zgonu.

Mediana czasu przeżycia $= 10$ lat

W ciągu 10 lat od przeszczepu oczekujemy, że zgon nastąpi u połowy pacjentów. Wartość tę zaznaczamy na wykresie, rysując linię w punkcie 0.5 odpowiadającym medianie. W podobny sposób zaznaczamy na wykresie kwartyle.



Przedział ufności dla mediany możemy zobrazować na wykresie, wyrysowując linie pionowe w oparciu o przedział ufności wokół krzywej oraz linię na poziomie 0.5.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Porównywanie krzywych przeżycia

KIEDY STOSOWAĆ:

Funkcje przeżycia można budować oddzielnie dla różnych podgrup, np. oddzielnie dla kobiet i mężczyzn, a następnie porównać je między sobą. Takie porównanie może dotyczyć zarówno dwóch, jak i kilku krzywych jednocześnie, może też badać, czy krzywe układają się w naturalny trend (np. wraz z wiekiem), oraz uwzględniać (korygować) analizę o wpływ dodatkowego czynnika (warstwy).

Porównania k krzywych przeżycia $S_1, S_2, \ldots, S_k$ w programie można dokonać przy pomocy trzech testów:

testu log-rank – najbardziej znanego i szeroko stosowanego, nawiązującego do procedury Mantela-Haenszela dla wielu tabel 2x2 (Mantel-Haenszel 1959[114], Mantel 1966[116], Cox 1972[48]),
uogólnienia Gehana testu Wilcoxona – wywodzącego się od testu Wilcoxona (Breslow 1970, Gehan 1965[69][70]),
testu Tarone-Ware – wywodzącego się od testu Wilcoxona (Tarone i Ware 1977[161]).

Wszystkie trzy testy opierają się na tej samej statystyce testowej – różnią je jedynie **wagi** w_j przypisywane poszczególnym punktom osi czasu:

test log-rank: $w_j = 1$ – wszystkie punkty osi czasu mają tę samą wagę, co daje większy wpływ na wynik późniejszym wartościom osi czasu;
uogólnienie Gehana testu Wilcoxona: $w_j = n_j$ – momenty czasowe ważone są liczbą obserwacji w każdym z nich, a więc większą wagę mają wartości początkowe osi czasu;
test Tarone-Ware: $w_j = \sqrt{n_j}$ – momenty czasowe ważone są pierwiastkiem z liczby obserwacji, co sytuuje ten test pomiędzy dwoma powyższymi.

WARUNKI STOSOWANIA

- **proporcjonalność hazardu** – hazard, definiowany jako nachylenie krzywej przeżycia, jest miarą tego, jak szybko następuje niepożądane zdarzenie; złamanie tego założenia nie dyskwalifikuje całkowicie powyższych testów, ale obniża ich moc, tym bardziej im wcześniej na osi czasu krzywe się przecinają.

DEFINICJE MIAR

Iloraz Hazardów (*ang. Hazard Ratio* – HR) opisuje wielkość obserwowanej różnicy między parą krzywych przeżycia, na podstawie liczby zdarzeń obserwowanych i oczekiwanych w każdej z grup:

$$HR = \frac{O_1/E_1}{O_2/E_2}$$

Jeśli $HR > 1$ (np. $HR = 2$), stopień narażenia na niepożądane zdarzenie w pierwszej grupie jest odpowiednio większy niż w grupie drugiej (dwukrotnie dla $HR = 2$); jeśli $HR < 1$, sytuacja jest odwrotna; $HR \approx 1$ oznacza, że obie grupy są narażone w podobnym stopniu.

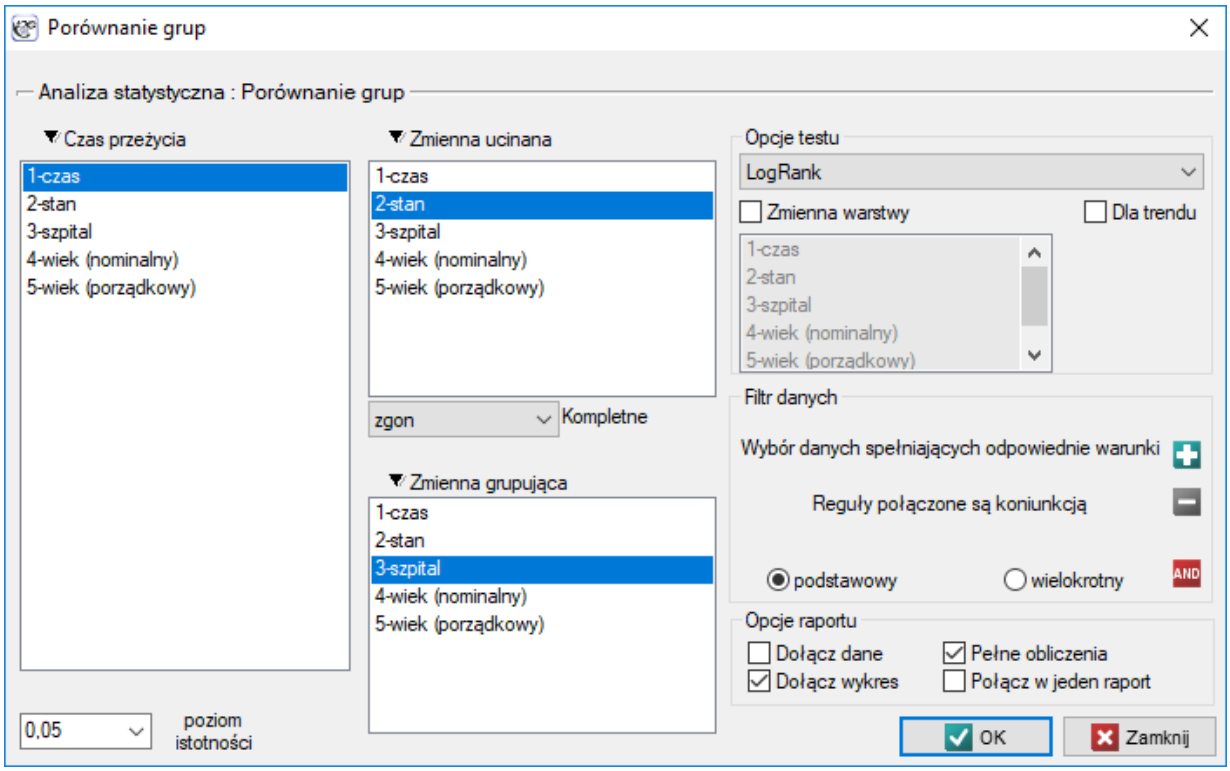
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Uwaga! Przedział ufności dla HR wyliczany jest w oparciu o błąd standardowy logarytmu HR (Armitage i Berry 1994[11]).

Program pozwala też sprawdzić, czy krzywe układają się w **trend** zgodny z naturalnym porządkiem porównywanych grup (np. kolejnych kategorii wiekowych), oraz porównać krzywe **z uwzględnieniem warstw**, czyli po skorygowaniu o dodatkowy czynnik mogący wpływać na wynik porównania (np. porównując czas przeżycia względem wieku, oddzielnie w warstwach wyznaczonych przez płeć).

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji porównania krzywych przeżycia wywołujemy poprzez menu Statystyka→Analiza przeżycia→Porównanie grup.



WYKONANIE I INTERPRETACJA

Przykład (28.1) c.d. *(plik przeszczep.pqs)*
Różnice dla dwóch krzywych przeżycia

Przeszczepy wątroby dokonywane były w dwóch różnych szpitalach. Sprawdzimy, czy długość życia pacjentów po przeszczepie zależała od szpitala, w którym dokonano przeszczepu. Porównania krzywych przeżycia dla tych szpitali dokonamy w oparciu o wszystkie zaproponowane w programie testy służące temu porównaniu.
Hipotezy:

$\mathcal{H}_0$: krzywa przeżycia pacjentów szpitala 1 = krzywa przeżycia pacjentów szpitala 2,
$\mathcal{H}_1$: krzywa przeżycia pacjentów szpitala 1 $\neq$ krzywa przeżycia pacjentów szpitala 2.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

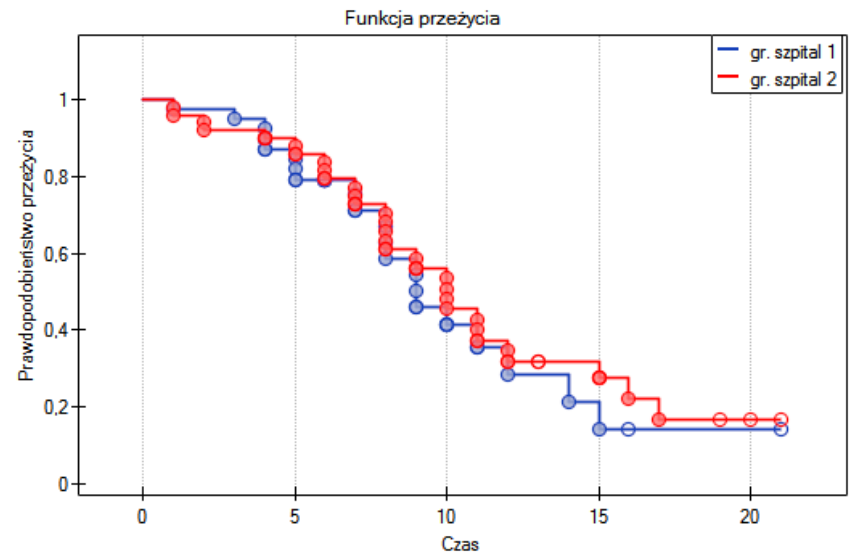
Porównanie grup	
Czas analizy	0,16 sek.
Analizowane zmienne	czas;stan
Poziom istotności	0,05
Zmienna grupująca	szpital(szpital 1;szpital 2)
Liczność	89
Kompletne	zgon
Ucięte	żyje
Test: LogRank	
Statystyka Chi-kwadrat	0,274351308
Ilość stopni swobody	1
Wartość p	0,600427701

Logrank

grupa	obs.	ocz.	obs./ocz.
szpital 1	21	19,257088	1,0905075
szpital 2	32	33,742911	0,9483473

Logrank

grupa	hazard r.	-95%CI	+95%CI
szpital 1<=	1,1499031	0,6569948	2,0126144



$p = 0.6004$ **dla testu log-rank (**$p = 0.6959$ **dla Gehana,** $p = 0.6465$ **dla Tarone-Ware);** $HR = 1.1499$**, 95%CI:** $\langle$**0.6570; 2.0126**$\rangle$

Brak podstaw do odrzucenia $\mathcal{H}_0$ – długość życia pacjentów obu szpitali jest podobna (przedział ufności dla HR zawiera wartość 1).

Różnice dla wielu krzywych przeżycia

Przeszczepy wątroby dokonywane były u ludzi w różnym wieku. Wyróżniono 3 grupy wiekowe: $\langle 45\text{lat}; 50\text{lat}\rangle$, $\langle 50\text{lat}; 55\text{lat}\rangle$, $\langle 55\text{lat}; 60\text{lat}\rangle$. Sprawdzimy, czy długość życia pacjentów po przeszczepie zależy od ich wieku w chwili dokonania przeszczepu.
Hipotezy:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$\mathcal{H}_0$: krzywe przeżycia pacjentów w wieku $\langle 45\text{lat}; 50\text{lat})$, $\langle 50\text{lat}; 55\text{lat})$, $\langle 55\text{lat}; 60\text{lat})$ są podobne,

$\mathcal{H}_1$: przynajmniej jedna krzywa przeżycia, spośród powyższych 3 krzywych, różni się od pozostałych.

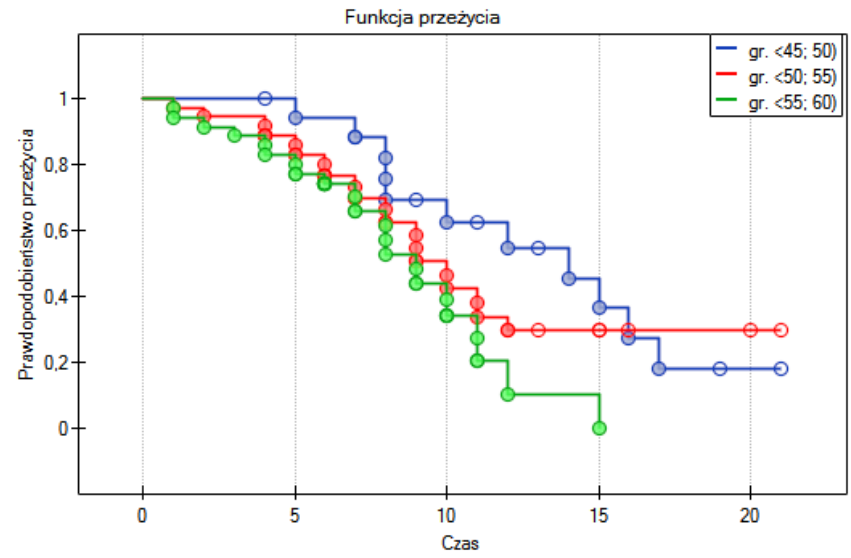
Porównanie grup	
Czas analizy	0,16 sek.
Analizowane zmienne	czas;stan
Poziom istotności	0,05
Zmienna grupująca	wiek (nominalny)(<45; 50);‹
Liczność	89
Kompletne	zgon
Ucięte	żyje
Test: LogRank	
Statystyka Chi-kwadrat	5,342329381
Ilość stopni swobody	2
Wartość p	0,069171615

Logrank

grupa	obs.	ocz.	obs./ocz.
<45; 50)	11	16,120092	0,6823782
<50; 55)	20	21,490397	0,9306482
<55; 60)	22	15,389510	1,4295451

Logrank

grupa	hazard r.	-95%CI	+95%CI
<45; 50)<	0,7332289	0,3843909	1,3986403
<45; 50)<	0,4773394	0,2373928	0,9598135
<50; 55)<	0,6510100	0,3383317	1,2526581



$p = 0.0692$ dla testu log-rank ($p = 0.09279$ dla Gehana, $p = 0.0779$ dla Tarone-Ware)

Brak podstaw do odrzucenia $\mathcal{H}_0$ – długość życia pacjentów z trzech grup wiekowych jest podobna, choć wartości p są dość bliskie poziomowi istotności 0.05.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przeglądając wartości hazardu (ilorazu wartości obserwowanych i oczekiwanych niepożądanych zdarzeń), zauważamy, że z każdą kolejną kategorią wiekową są one nieco wyższe: $\langle 0.68, 0.93, 1.43 \rangle$. Choć różnice te nie są istotne statystycznie, mogą sugerować istnienie trendu wzrostu wartości hazardu wraz z wiekiem.

Trend dla kilku krzywych przeżycia

Wprowadzając do testu informację o uporządkowaniu porównywanych kategorii (wykorzystując zmienną wiek, w której przedziały wiekowe ponumerowano odpowiednio 1, 2 i 3), sprawdzimy, czy w porównywanych krzywych istnieje trend.
Hipotezy:

$\mathcal{H}_0$: brak trendu w krzywych czasu przeżycia pacjentów po przeszczepie (trendu zależnego od wieku pacjentów w chwili przeszczepu),
$\mathcal{H}_1$: czym starsi są pacjenci w momencie dokonania przeszczepu, tym większe/mniejsze jest prawdopodobieństwo ich przeżycia określonego odcinka czasu.

Dla trendu:	
Statystyka Chi-kwadrat	5,113469574
Ilość stopni swobody	1
Wartość p	0,023740797

$p = 0.0237$ dla testu log-rank ($p = 0.0317$ dla Gehana, $p = 0.0241$ dla Tarone-Ware)

Krzywe przeżycia ułożone są w istotnym statystycznie trendzie: czym starszy pacjent w chwili przeszczepu, tym mniejsze prawdopodobieństwo przeżycia określonego odcinka czasu.

Najniżej na wykresie Kaplana-Meiera znajduje się krzywa dla osób w wieku $\langle$55 lat; 60 lat), nad nią krzywa dla pacjentów w wieku $\langle$50 lat; 55 lat), a najwyżej krzywa dla pacjentów w wieku $\langle$45 lat; 50 lat).

Krzywe przeżycia dla warstw

Sprawdzimy teraz, czy obserwowany wcześniej trend jest niezależny od szpitala, w którym dokonano przeszczepu. W tym celu jako zmienną warstwa wybierzemy szpital.

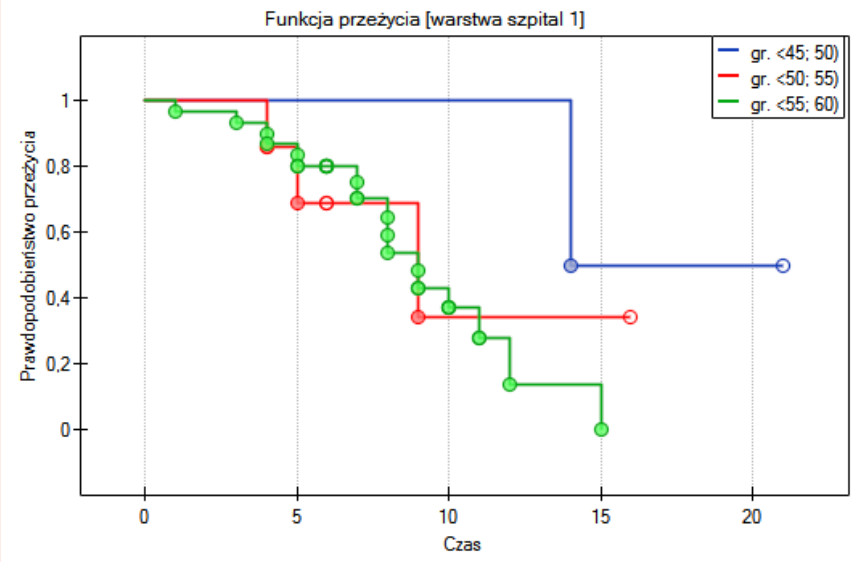
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

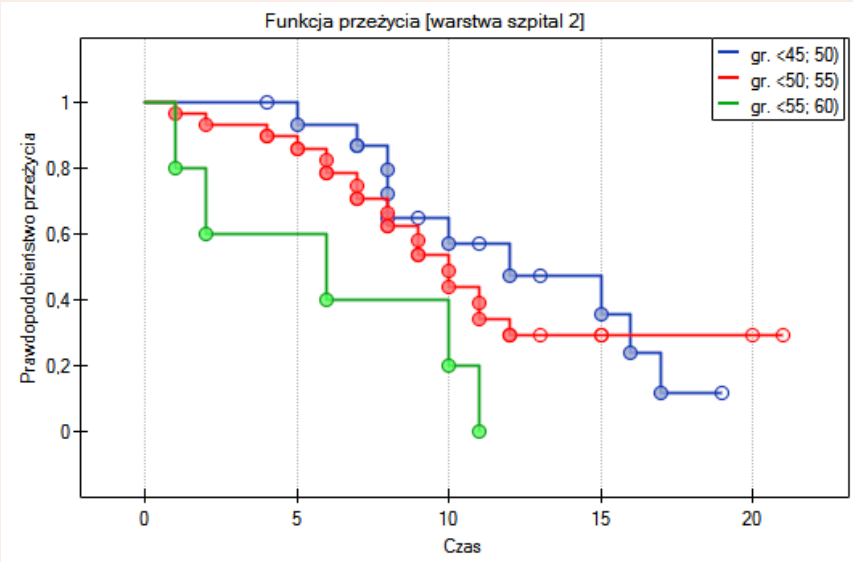
Porównanie grup	
Czas analizy	0,21 sek.
Analizowane zmienne	czas;stan
Poziom istotności	0,05
Zmienna grupująca	wiek (porządkowy)(1;2;3)
Zmienna warstwy	szpital
Liczność	89
Kompletne	zgon
Ucięte	żyje
Test: LogRank	
Warstwa: szpital 1	
Statystyka Chi-kwadrat	2,413735014
Ilość stopni swobody	2
Wartość p	0,299132845
Dla trendu:	
Statystyka Chi-kwadrat	2,357088823
Ilość stopni swobody	1
Wartość p	0,124714733
Warstwa: szpital 2	
Statystyka Chi-kwadrat	5,542657238
Ilość stopni swobody	2
Wartość p	0,062578806
Dla trendu:	
Statystyka Chi-kwadrat	3,028300936
Ilość stopni swobody	1
Wartość p	0,081823659
Wspólne dla warstw	
Statystyka Chi-kwadrat	6,259357261
Ilość stopni swobody	2
Wartość p	0,043731849
Dla trendu:	
Statystyka Chi-kwadrat	5,374392333
Ilość stopni swobody	1
Wartość p	0,020434458

Logrank

Warstwa	Grupa	Obs.	Ocz.	Obs./Ocz.
szpital 1	<45; 50)	1	3,1484541	0,3176161
szpital 1	<50; 55)	3	3,3935138	0,8840394
szpital 1	<55; 60)	17	14,458031	1,1758170
szpital 2	<45; 50)	10	12,506498	0,7995843
szpital 2	<50; 55)	17	17,490412	0,9719610
szpital 2	<55; 60)	5	2,0030892	2,4961444
Wspólne	<45; 50)	11	15,654952	0,7026530
Wspólne	<50; 55)	20	20,883926	0,9576743
Wspólne	<55; 60)	22	16,461121	1,3364824

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





W raporcie najpierw przedstawiona jest analiza poszczególnych warstw, zarówno wyniki testów, jak i wartości hazardu. W warstwie pierwszej widoczny jest trend wzrostu hazardu, choć nieistotny statystycznie; trend o tym samym kierunku (na pograniczu istotności) obserwowany jest w warstwie drugiej. Kumulacja tych trendów we wspólnej analizie warstw pozwoliła uzyskać istotność trendu krzywych przeżycia.

Analiza w warstwach: trend istotny niezależnie od szpitala

Czym starszy pacjent w chwili przeszczepu, tym mniejsze prawdopodobieństwo przeżycia określonego odcinka czasu – niezależnie od szpitala dokonującego przeszczepu.

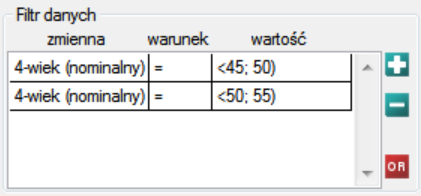
Analiza porównawcza krzywych przeżycia w korekcji o warstwy daje wynik istotny dla testu log-rank i Tarone-Ware, a nieistotny dla Gehana, co może wskazywać na to, że pojawiające się różnice w krzywych nie są tak widoczne w początkowych okresach czasu przeżycia, jak w okresach późniejszych. Przyglądając się ilorazowi hazardu dla porównywanych parami krzywych

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Logrank

warstwa	grupa	hazard r.	-95%CI	+95%CI
szpital 1	1<=>2	0,3592782	0,0775124	1,6652928
szpital 1	1<=>3	0,2701238	0,0798335	0,9139877
szpital 1	2<=>3	0,7518511	0,2305103	2,4522984
szpital 2	1<=>2	0,8226505	0,3981102	1,6999158
szpital 2	1<=>3	0,3203277	0,0720764	1,4236263
szpital 2	2<=>3	0,3893849	0,0902500	1,6800069
wspólne	1<=>2	0,7337077	0,3810310	1,4128167
wspólne	1<=>3	0,5257480	0,2631977	1,0502028
wspólne	2<=>3	0,7165633	0,3755802	1,3671193

możemy zlokalizować istotne różnice. Najmniejszy iloraz hazardu uzyskano dla porównania krzywej dla najmłodszej grupy z krzywą dla grupy najstarszej: 0.53, 95% przedział ufności ⟨0.26; 1.05⟩ zawiera wprawdzie wartość 1, ale jest na pograniczu tej wartości, co może sugerować istotne różnice między odpowiadającymi im krzywymi. By potwierdzić to przypuszczenie, można użyć filtru danych w oknie analizy i porównać krzywe parami.



Uwaga! Przy porównaniach parami należy zastosować jedną z poprawek stosowanych przy wielokrotnych porównaniach i zmodyfikować poziom istotności. W tym przypadku dla poprawki Bonferroniego przy trzech porównaniach poziom istotności wyniesie 0.017. Dla uproszczenia posłużymy się tylko testem log-rank.

⟨45 lat; 50 lat) vs ⟨50 lat; 55 lat)

Wspólne dla warstw	
Statystyka Chi-kwadrat	0,588444784
Ilość stopni swobody	1
Wartość p	0,443021142

⟨45 lat; 50 lat) vs ⟨55 lat; 60 lat)

Wspólne dla warstw	
Statystyka Chi-kwadrat	8,944721898
Ilość stopni swobody	1
Wartość p	0,002782725

⟨50 lat; 55 lat) vs ⟨55 lat; 60 lat)

Wspólne dla warstw	
Statystyka Chi-kwadrat	2,241246614
Ilość stopni swobody	1
Wartość p	0,134372609

Zgodnie z oczekiwaniem istotne statystycznie różnice dotyczą tylko krzywych przeżycia dla najmłodszej i najstarszej grupy wiekowej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DLA ZAINTERESOWANYCH

Różnice w krzywych przeżycia
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad S_1(t) = S_2(t) = ... = S_k(t), \quad \text{dla wszystkich } t, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } S_i(t) \text{ są sobie równe.} \end{aligned}$$

Wzory:

Statystyka testowa (chi-kwadrat):

$$\chi^2 = U' V^{-1} U$$

gdzie:

$U_i = \sum_{j=1}^{m} w_j(d_{ij} - e_{ij})$,

V – macierz kowariancji o wymiarach $(k-1) \times (k-1)$:

diagonala: $\sum_{j=1}^{m} w_j^2 \frac{n_{ij}(n_j - n_{ij}) d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$,

poza diagonalą: $\sum_{j=1}^{m} w_j^2 \frac{n_{ij} n_{lj} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$,

m – liczba momentów czasowych, w których nastąpiło niepożądane zdarzenie (zgon),

$d_j = \sum_{i=1}^{k} d_{ij}$ – obserwowana liczba niepożądanych zdarzeń w j-tym momencie czasowym,

d_{ij} – obserwowana liczba niepożądanych zdarzeń w i-tej grupie w j-tym momencie czasowym,

$e_{ij} = \frac{n_{ij} d_j}{n_j}$ – oczekiwana liczba niepożądanych zdarzeń w i-tej grupie w j-tym momencie czasowym,

$n_j = \sum_{i=1}^{k} n_{ij}$ – liczba narażonych w j-tym momencie czasowym.

Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z $df = k-1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Trend w krzywych przeżycia
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{w badanej populacji nie istnieje trend w położeniu krzywych } S_1, S_2, ..., S_k, \\ \mathcal{H}_1: &\quad \text{w badanej populacji istnieje trend w położeniu krzywych } S_1, S_2, ..., S_k. \end{aligned}$$

Wzory:

Statystyka testowa (chi-kwadrat):

$$\chi^2 = \frac{(c'U)^2}{c'Vc}$$

gdzie $c = (c_1, c_2, ..., c_k)$ – wektor wag dla porównywanych grup informujący o ich naturalnym porządku (najczęściej kolejne liczby naturalne). Statystyka ta ma asymptotycznie (dla dużych liczności oczekiwanych) rozkład χ^2 z jednym stopniem swobody.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Uwaga! By można było przeprowadzić analizę trendu w krzywych przeżycia, zmienna grupująca musi być zmienną liczbową, w której wartości informują o naturalnym porządku grup – liczby te traktowane są w analizie jako wagi $c_1, c_2, ..., c_k$.

Krzywe przeżycia dla warstw
Hipotezy dla różnic w krzywych przeżycia:

$$\begin{aligned} \mathcal{H}_0: &\quad S_1^*(t) = S_2^*(t) = ... = S_k^*(t), \quad \text{dla wszystkich } t, \\ \mathcal{H}_1: &\quad \text{nie wszystkie } S_i^*(t) \text{ są sobie równe.} \end{aligned}$$

Hipotezy dla analizy trendu w krzywych przeżycia:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{w badanej populacji nie istnieje trend w położeniu krzywych } S_1^*, S_2^*, ..., S_k^*, \\ \mathcal{H}_1: &\quad \text{w badanej populacji istnieje trend w położeniu krzywych } S_1^*, S_2^*, ..., S_k^*, \end{aligned}$$

gdzie $S_1^*(t), S_2^*(t), ..., S_k^*(t)$ – krzywe przeżycia po korekcji o zmienną wyznaczającą warstwy. Obliczenia dla statystyk testowych bazują na formułach opisanych dla testów nieuwzględniających warstw, z tą różnicą, że macierz U i V zastąpiona jest sumą macierzy $\sum_{l=1}^{L} U$ i $\sum_{l=1}^{L} V$, sumowaną po warstwach utworzonych przez zmienną korygującą ($l = 1, 2, ..., L$).

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Regresja proporcjonalnego hazardu Coxa

KIEDY STOSOWAĆ:

Regresja Coxa, znana też jako model proporcjonalnego hazardu Coxa (Cox PH model) – Cox D.R. (1972)[48], jest najszerzej stosowanym podejściem regresyjnym w analizie przeżycia. Pozwala zbadać jednoczesny wpływ wielu zmiennych niezależnych (X_1, X_2, ..., X_k) na czas przeżycia, np. wpływ sposobu leczenia, wieku i innych czynników rokowniczych na długość życia. Jest to podejście w pewnym sensie nieparametryczne – nie wymaga znajomości kształtu funkcji hazardu ani przeżycia, co jest źródłem jego popularności.

WARUNKI STOSOWANIA

- **proporcjonalność hazardu** – wpływ każdej zmiennej niezależnej na hazard musi być stały w czasie (zob. *Analiza reszt modelu* poniżej);
- brak silnej współliniowości zmiennych niezależnych;
- odpowiednio duża liczba obserwacji względem liczby szacowanych parametrów.

Uwaga! Liczba obserwacji powinna być co najmniej dziesięciokrotnie większa niż stosunek liczby szacowanych parametrów modelu (k) do mniejszej z proporcji liczności uciętych lub kompletnych (p), czyli

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

$n \geq 10k/p$ (Peduzzi i in. 1995[134]).

Uwaga! Zmienne niezależne nie powinny być współliniowe – współliniowość prowadzi do niepewnej estymacji i bardzo wysokich wartości błędów.

DEFINICJE MIAR

Funkcja, na której oparty jest model, opisuje hazard wynikowy jako iloczyn dwóch wielkości, z których tylko jedna zależy od czasu (t):

$$h(t, X_1, X_2, ..., X_k) = h_0(t) \cdot \exp\left(\sum_{i=1}^{k} \beta_i X_i\right),$$

gdzie $h(t, X_1, ..., X_k)$ to wynikowy hazard (zmieniające się w czasie ryzyko, zależne od innych czynników, np. sposobu leczenia), $h_0(t)$ to hazard bazowy (hazard przy założeniu, że wszystkie zmienne objaśniające są równe zero), a $\sum_{i=1}^{k} \beta_i X_i$ to kombinacja (najczęściej liniowa) zmiennych niezależnych $X_1, ..., X_k$ i parametrów modelu $\beta_1, ..., \beta_k$, nazywanych **współczynnikami regresji**. Omówienie przygotowania zmiennych fikcyjnych i interakcji przedstawiono w rozdziale 25.0.2 Przygotowanie zmiennych do analizy w modelach wielowymiarowych.

Współczynniki szacowane są tzw. **„częściową" metodą największej wiarygodności** (algorytm iteracyjny Newtona-Raphsona) – „częściową", ponieważ poszukiwanie maksimum funkcji wiarygodności L odbywa się wprost tylko dla danych kompletnych, natomiast dane ucięte są uwzględniane pośrednio. Korekcja rang wiązanych oparta jest na metodzie Breslowa[28].

Uwaga! Kryterium zbieżności algorytmu iteracyjnego Newtona-Raphsona można kontrolować dwoma parametrami: limitem liczby iteracji oraz kryterium zbieżności (wartością, poniżej której poprawa estymacji uznawana jest za nieznaczącą, co kończy działanie algorytmu).

Iloraz Hazardu (HR) – dla każdej zmiennej niezależnej wyliczany jest jednostkowy Iloraz Hazardu (*ang. Hazard Ratio* - HR):

$$HR_i = e^{\beta_i}.$$

Wyraża on zmianę ryzyka niepożądanego zdarzenia, gdy dana zmienna niezależna rośnie o 1 jednostkę, przy pozostałych zmiennych modelu utrzymanych na stałym poziomie. Interpretacja:

- $HR > 1$ – stymulujący wpływ zmiennej: o ile wzrasta ryzyko wystąpienia zdarzenia, gdy zmienna niezależna rośnie o jeden poziom;
- $HR < 1$ – destymulujący wpływ zmiennej: o ile spada ryzyko wystąpienia zdarzenia, gdy zmienna niezależna rośnie o jeden poziom;
- $HR \approx 1$ – zmienna nie ma wpływu na wystąpienie zdarzenia.

Uwaga! Jeśli analizę przeprowadzamy dla modelu innego niż liniowy lub uwzględniamy interakcję, wówczas – analogicznie jak w modelu regresji logistycznej – odpowiedni HR wyliczamy, zmieniając formułę kombinacji zmiennych niezależnych.

Weryfikacja modelu – na podstawie współczynnika oraz jego błędu szacunku wnioskujemy, czy dana zmienna niezależna wywiera istotny wpływ na zmienną zależną; posługujemy się w tym celu testem Walda, weryfikującym hipotezę $\mathcal{H}_0 : \beta_i = 0$ (równoważnie $HR_i = 1$) wobec $\mathcal{H}_1 : \beta_i \neq 0$ ($HR_i \neq 1$).

Jakość całego zbudowanego modelu ocenia się natomiast kilkoma ogólnymi miarami:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **kryteria informacyjne** (AIC, $AICc$, BIC) – opierają się na entropii informacji niesionej przez model, czyli szacują utraconą informację, gdy dany model opisuje badane zjawisko; wybieramy model o **minimalnej** wartości danego kryterium. Są rodzajem kompromisu między dobrocią dopasowania a złożonością modelu – rosną wraz z liczbą parametrów modelu, tym szybciej, im mniejsza jest liczba obserwacji. $AICc$ to poprawka AIC zalecana przy małych próbach, a BIC dodatkowo silniej karze złożoność modelu (Volinsky i Raftery 2000[167]). Kryterium informacyjne nie jest miarą absolutną – jeśli wszystkie porównywane modele źle opisują rzeczywistość, w żadnym z kryteriów nie znajdziemy tego ostrzeżenia.

- **Pseudo R**^{2} (McFadden R^2) – miara dopasowania modelu, odpowiednik współczynnika determinacji wielorakiej R^2 z liniowej regresji wielorakiej; przyjmuje wartości z przedziału $\langle 0; 1)$, gdzie wartości bliskie 1 oznaczają dobre dopasowanie, a 0 – całkowity brak dopasowania. Ze względu na wrażliwość na liczbę zmiennych w modelu wyznacza się też jego poprawione wersje: $R^2_{Nagelkerke}$ i $R^2_{Cox-Snell}$.

- **test ilorazu wiarygodności** – podstawowe narzędzie oceny istotności wszystkich zmiennych modelu łącznie; weryfikuje hipotezę $\mathcal{H}_0$: wszystkie $\beta_i = 0$ wobec $\mathcal{H}_1$: istnieje $\beta_i \neq 0$.

- **AUC zależne od czasu** – w analizie przeżycia program wyznacza **AUC zależne od czasu** (Heagerty i in. 2000[82], Blanche i in. 2013[21]): dla każdego czasu t, w którym wystąpiło zdarzenie badane, obserwacje dzielone są na **przypadki** (zdarzenie do czasu t) i **kontrole** (obserwowane dłużej niż t, bez zdarzenia), a markerem jest kombinacja liniowa zmiennych i parametrów modelu. $AUC(t)$ to prawdopodobieństwo, że losowo wybrany przypadek ma wyższą wartość markera niż losowo wybrana kontrola; cenzurowanie uwzględniane jest przez ważenie obserwacji odwrotnością prawdopodobieństwa cenzurowania (IPCW). Interpretacja jest taka jak dla klasycznego AUC: 0.5 odpowiada rozróżnianiu losowemu, 1 – doskonałemu, a wartości powyżej ok. 0.7–0.8 uznaje się zwykle za dobrą zdolność dyskryminacyjną. Ponieważ $AUC(t)$ jest funkcją czasu, w raporcie podawane jest średnie $AUC(t)$ oraz jego wartości w kwartylach i na końcu czasów zdarzeń, a wykres $AUC(t)$ względem czasu (opcja Krzywa ROC w oknie Dołącz wykres) pokazuje, czy zdolność modelu do przewidywania zdarzeń jest stabilna dla różnych horyzontów prognozy, czy też zmienia się z czasem. Dla najpóźniejszych czasów, gdy w zbiorze kontroli pozostaje już niewiele obserwacji, estymator $AUC(t)$ staje się niestabilny – skoków na samym końcu wykresu nie należy nadinterpretować. Dla $AUC(t)$ nie wyznacza się punktu odcięcia ani testu istotności.

Analiza reszt modelu pozwala zweryfikować założenia modelu – przede wszystkim zlokalizować wartości odstające i ocenić proporcjonalność hazardu (opracowano szereg graficznych metod pozwalających na ocenę adekwatności modelu proporcjonalnego hazardu – Lee i Wang 2003[102]). Ze względu na obserwacje ucięte reszt nie liczy się tu w standardowy sposób (jako różnicę wartości obserwowanej i przewidywanej), lecz jako jeden z trzech rodzajów: **Martingale**, **Deviance** lub **Schoenfeld**, które można wyrysować względem czasu lub zmiennych niezależnych.

Martingale – interpretowane jako różnica w czasie $[0, t]$ pomiędzy obserwowaną a przewidywaną przez model licznością zdarzeń niepożądanych. Wartość oczekiwana wynosi 0, ale rozkład jest skośny (od $-\infty$ do 1), co utrudnia interpretację wykresu.

Deviance – podobnie jak Martingale, asymptotycznie mają wartość oczekiwaną 0, ale przy odpowiednio dopasowanym modelu rozkładają się symetrycznie wokół zera z odchyleniem standardowym 1. Wartość dodatnia oznacza, że badany obiekt przeżył krócej niż przewidywał model, ujemna – że dłużej. To przede wszystkim narzędzie identyfikujące wartości odstające (w raporcie reszty oddalone o ponad 3 odchylenia standardowe od 0 oznaczane są kolorem czerwonym).

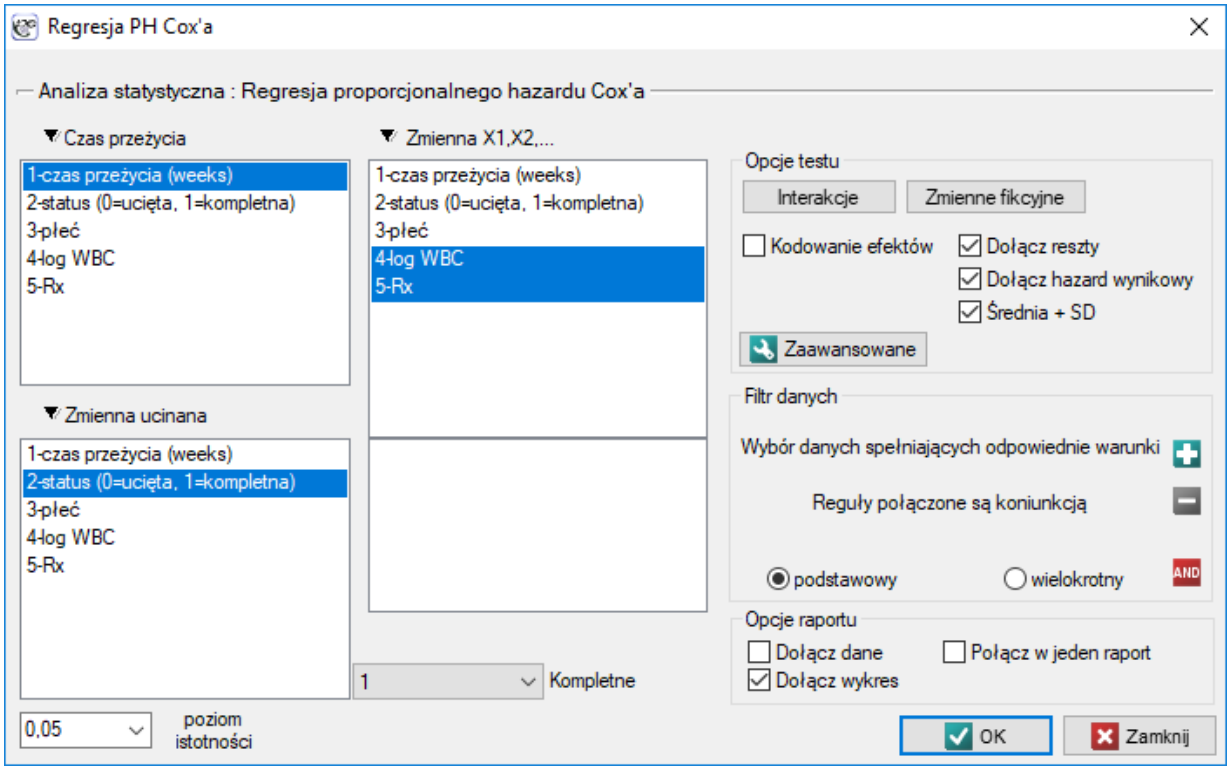
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Schoenfeld – wyliczane oddzielnie dla każdej zmiennej niezależnej, tylko dla obserwacji kompletnych; ich suma i wartość oczekiwana wynoszą 0. Pozwalają zidentyfikować zmienną łamiącą założenie proporcjonalności hazardu – jest to ta, której wykres reszt względem czasu układa się w systematyczny wzór (najczęściej bada się liniową zależność reszt od czasu). Równomierne rozłożenie punktów wokół 0 świadczy o spełnieniu założenia proporcjonalności hazardu przez daną zmienną.

Uwaga! Gdy dla którejś zmiennej modelu nie jest spełnione założenie proporcjonalności hazardu, rozwiązaniem może być wykonanie analizy Coxa oddzielnie dla każdego poziomu tej zmiennej (stratyfikacja) lub rozszerzenie modelu o odpowiednią zmienną zależną od czasu.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji Regresji Coxa wywołujemy poprzez menu Statystyka→Analiza przeżycia→Regresja PH Coxa.



DLA ZAINTERESOWANYCH

Błąd standardowy współczynnika

$$SE_b = \sqrt{diag(H^{-1})_b},$$

gdzie $diag(H^{-1})$ to główna przekątna macierzy kowariancji.

Test Walda dla pojedynczego współczynnika

Hipotezy: $\mathcal{H}_0 : \beta_i = 0$ wobec $\mathcal{H}_1 : \beta_i \neq 0$ (równoważnie $\mathcal{H}_0 : HR_i = 1$ wobec $\mathcal{H}_1 : HR_i \neq 1$).

Wzory: Statystyka testowa:

$$\chi^2 = \left(\frac{b_i}{SE_{b_i}}\right)^2$$

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

ma asymptotycznie (dla dużych liczności) rozkład χ^2 z 1 stopniem swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Kryteria informacyjne (oparte na L_{FM} – maksimum funkcji wiarygodności modelu pełnego, L_0 – maksimum funkcji wiarygodności modelu z samym wyrazem wolnym, k – liczbie parametrów, d – liczbie zdarzeń niepożądanych):

$$AIC = -2\ln L_{FM} + 2k, \qquad AICc = AIC + \frac{2k(k+1)}{d-k-1}, \qquad BIC = -2\ln L_{FM} + k\ln(d).$$

Pseudo R^2

$$R^2_{Pseudo} = 1 - \frac{\ln L_{FM}}{\ln L_0}, \qquad R^2_{Nagelkerke} = \frac{1 - e^{-(2/d)(\ln L_{FM} - \ln L_0)}}{1 - e^{(2/d)\ln L_0}}, \qquad R^2_{Cox-Snell} = 1 - e^{\frac{(-2\ln L_0) - (-2\ln L_{FM})}{d}}.$$

Test ilorazu wiarygodności dla wszystkich zmiennych modelu
Hipotezy: $\mathcal{H}_0$: wszystkie $\beta_i = 0$ wobec $\mathcal{H}_1$: istnieje $\beta_i \neq 0$. **Wzory:** Statystyka testowa:

$$\chi^2 = -2\ln(L_0/L_{FM}) = -2\ln(L_0) - (-2\ln(L_{FM}))$$

ma asymptotycznie (dla dużych liczności) rozkład χ^2 z k stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

AUC zależne od czasu
Dla markera $M = \sum_{i=1}^{k} b_i X_i$ i czasu t, w którym wystąpiło zdarzenie badane, estymator typu *cumulative/dynamic* z wagami IPCW (Heagerty i in. 2000[82], Blanche i in. 2013[21]) ma postać:

$$AUC(t) = \frac{\sum_i \sum_j w_i(t)\, w_j(t) \left[\mathbf{1}(M_i > M_j) + \frac{1}{2}\mathbf{1}(M_i = M_j)\right]}{\sum_i w_i(t) \sum_j w_j(t)},$$

gdzie sumowanie po i obejmuje przypadki ($T_i \leq t$, $\delta_i = 1$) z wagą $w_i(t) = 1/\hat{G}(T_i^-)$, a po j – kontrole ($T_j > t$) z wagą $w_j(t) = 1/\hat{G}(t)$; $\hat{G}$ jest estymatorem Kaplana-Meiera rozkładu cenzurowania (za „zdarzenie" uznawane jest w nim ucięcie obserwacji). $AUC(t)$ wyliczane jest we wszystkich unikalnych czasach zdarzeń badanych, a podawane w raporcie średnie $AUC(t)$ jest średnią arytmetyczną tych wartości.

Porównywanie modeli regresji PH Coxa

KIEDY STOSOWAĆ:

Ze względu na możliwość jednoczesnej analizy wielu zmiennych niezależnych w jednym modelu regresji Coxa, pojawia się problem wyboru optymalnego modelu. Wybierając zmienne niezależne, należy pamiętać, by w modelu znajdowały się zmienne silnie związane z czasem przeżycia i słabo skorelowane między sobą. Porównywanie modeli z różną liczbą zmiennych pozwala ocenić, czy dodanie (lub usunięcie) danej zmiennej istotnie poprawia model.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DEFINICJE MIAR

Porównując modele, zwracamy uwagę na kryteria informacyjne (AIC, $AICc$, BIC) oraz współczynniki dopasowania modelu (R^2_{Pseudo}, $R^2_{Nagelkerke}$, $R^2_{Cox-Snell}$), a także na wynik **testu ilorazu wiarygodności**, weryfikującego, czy maksimum funkcji wiarygodności modelu pełnego (L_{FM}) różni się istotnie od modelu zredukowanego (L_{RM}).

Decyzję o wyborze modelu podejmujemy na podstawie wszystkich powyższych wielkości łącznie. Jeśli porównywane modele nie różnią się istotnie, powinniśmy wybrać ten z mniejszą liczbą zmiennych – oznacza to bowiem, że zmienne obecne w modelu pełnym, a nieobecne w zredukowanym, nie wnoszą istotnej informacji. Jeśli różnica jest istotna statystycznie, oznacza to, że model z większą liczbą zmiennych (o większym R^2) jest istotnie lepszy.

W programie PQStat porównywanie modeli można przeprowadzić na dwa sposoby:

- **ręcznie** – badacz sam buduje model **pełny** (z większą liczbą zmiennych) i **zredukowany** (powstały z pełnego po usunięciu zmiennych zbędnych z punktu widzenia badanego zjawiska), a następnie na podstawie wyników porównania wybiera lepszy z nich;
- **automatycznie** – program sam buduje kolejne modele i wskazuje najlepszy według wybranego kryterium (opis poniżej).

Automatyczny dobór modelu

Składniki modelu. Jednostką, którą metody automatyczne dodają do modelu lub z niego usuwają, jest składnik: pojedyncza zmienna, cała zmienna fikcyjna (wszystkie jej kolumny 0–1 naraz) albo interakcja. Obowiązuje zasada hierarchii: interakcja może znaleźć się w modelu tylko wtedy, gdy są w nim jej efekty główne, i tylko ona może zostać usunięta przed nimi.

Kryterium doboru wybieramy w oknie analizy:

- AIC, AICc lub BIC – kryteria informacyjne; lepszy jest model o mniejszej wartości kryterium. Ruch (dodanie lub usunięcie składnika) jest przyjmowany tylko wtedy, gdy obniża kryterium;
- p testu zagnieżdżonego – test ilorazu wiarygodności porównujący model przed zmianą i po zmianie, z zadanym progiem p. Składnik jest usuwany, gdy p jest większe od progu, a dodawany, gdy p jest mniejsze od progu. Ta sama wartość p jest pokazywana w raporcie jako porównanie sąsiednich modeli;
- błąd walidacji – wartość $-2 \ln PL$ na zdarzenie wyznaczona poza zbiorem uczącym (opis poniżej: Ocena modeli poza zbiorem uczącym). Ruch jest przyjmowany tylko wtedy, gdy zmniejsza ten błąd. Kryterium wymaga włączonej opcji Walidacja – wybranie go przy wyłączonej walidacji włącza walidację krzyżową.

Metody doboru:

- **eliminacja wstecz** – start od modelu ze wszystkimi zmiennymi; w każdym kroku usuwany jest ten składnik, którego usunięcie najbardziej obniża kryterium; koniec, gdy żadne usunięcie już go nie obniża;
- **selekcja w przód** – start od modelu zerowego (bez zmiennych, sam hazard bazowy); w każdym kroku dodawany jest składnik najbardziej obniżający kryterium; koniec, gdy żadne dodanie już go nie obniża;

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- **krokowa dwukierunkowa** – start od modelu zerowego; w każdym kroku rozważane są zarówno dodania, jak i usunięcia składników i wybierany jest najlepszy ruch; koniec, gdy żaden ruch nie poprawia kryterium;
- **wszystkie podzbiory** – dopasowywane są wszystkie modele zgodne z zasadą hierarchii (maksymalnie 15 składników), a w raporcie prezentowany jest najlepszy model każdej wielkości; metoda korzysta z kryteriów informacyjnych albo z błędu walidacji (kryterium p nie jest w niej dostępne). Przy wielu zmiennych obliczenia mogą trwać dłużej, bo liczba dopasowywanych modeli rośnie wykładniczo.

Wszystkie modele są liczone na tym samym zbiorze obserwacji – wiersze z brakami w którejkolwiek ze zmiennych kandydujących są odrzucane raz, przed analizą. Dzięki temu kryteria informacyjne i testy ilorazu wiarygodności są między modelami porównywalne.

Raport zawiera: metodę i kryterium doboru, liczbę składników i dopasowanych modeli, kryteria informacyjne modelu zerowego, kolejne modele ścieżki doboru (każdy z wartością $-2\ln L$, kryteriami AIC, $AICc$, BIC, współczynnikami R^2_{Pseudo}, $R^2_{Nagelkerke}$, $R^2_{Cox-Snell}$ oraz tabelą współczynników i ilorazów hazardu) i porównanie każdej sąsiedniej pary modeli testem ilorazu wiarygodności (statystyka χ^2, stopnie swobody, wartość p). Najlepszy model według wybranego kryterium jest oznaczony w raporcie. Jeśli żaden składnik nie poprawia modelu zerowego, raport informuje, że najlepszy jest model zerowy. Przy metodzie wszystkich podzbiorów sąsiednie modele mogą nie być zagnieżdżone – wówczas test nie jest wykonywany.

Uwaga! Wartości p i przedziały ufności w modelu wybranym automatycznie są zbyt optymistyczne, ponieważ ten sam zbiór danych posłużył do wyboru zmiennych i do oszacowania ich efektów. Dobór automatyczny warto traktować jako pomoc w poszukiwaniu modelu, a ostateczny wybór oprzeć na wiedzy merytorycznej. Wybrany model należy zbudować ponownie w oknie Regresja PH Coxa i zweryfikować założenie proporcjonalności hazardów.

Ocena modeli poza zbiorem uczącym

Kryteria informacyjne, współczynniki dopasowania i testy są wyznaczane na tych samych obserwacjach, na których oszacowano współczynniki modeli, przez co ocena modeli jest zbyt optymistyczna – tym bardziej, im więcej zmiennych zawiera model. Dlatego, tak jak w porównywaniu modeli regresji wielorakiej, w oknie analizy można włączyć ocenę modeli na obserwacjach, które nie brały udziału w ich dopasowaniu (opcja Walidacja): **walidację krzyżową k-krotną** (każdy model jest dodatkowo dopasowywany k razy, za każdym razem bez jednej z k części danych) albo **losowy zbiór testowy** (modele są dopasowywane na części uczącej, a oceniane na odłożonej części testowej). Podział jest losowy, wyznaczony przez Ziarno, i **warstwowany względem statusu**: obserwacje kompletne i ucięte są dzielone osobno, dzięki czemu w każdej części udział zdarzeń jest zbliżony.

Dla każdego modelu raport podaje wówczas dwie dodatkowe miary:

- $-2\ln PL$ **na zdarzenie** – logarytm wiarygodności częściowej wyznaczony poza zbiorem uczącym, pomnożony przez -2 i podzielony przez liczbę zdarzeń. Przy zbiorze testowym jest to wiarygodność częściowa obserwacji testowych dla współczynników oszacowanych na zbiorze uczącym, a w walidacji krzyżowej – wiarygodność częściowa z walidacji krzyżowej według Verweija i van Houwelingena[168]. Im mniejsza wartość, tym lepszy model. To ona jest **błędem walidacji** – według niej wskazywany jest najlepszy model wg walidacji i na niej opiera się kryterium doboru błąd walidacji;
- **C Harrella** – współczynnik zgodności[79]: odsetek porównywalnych par obserwacji ocenianych

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

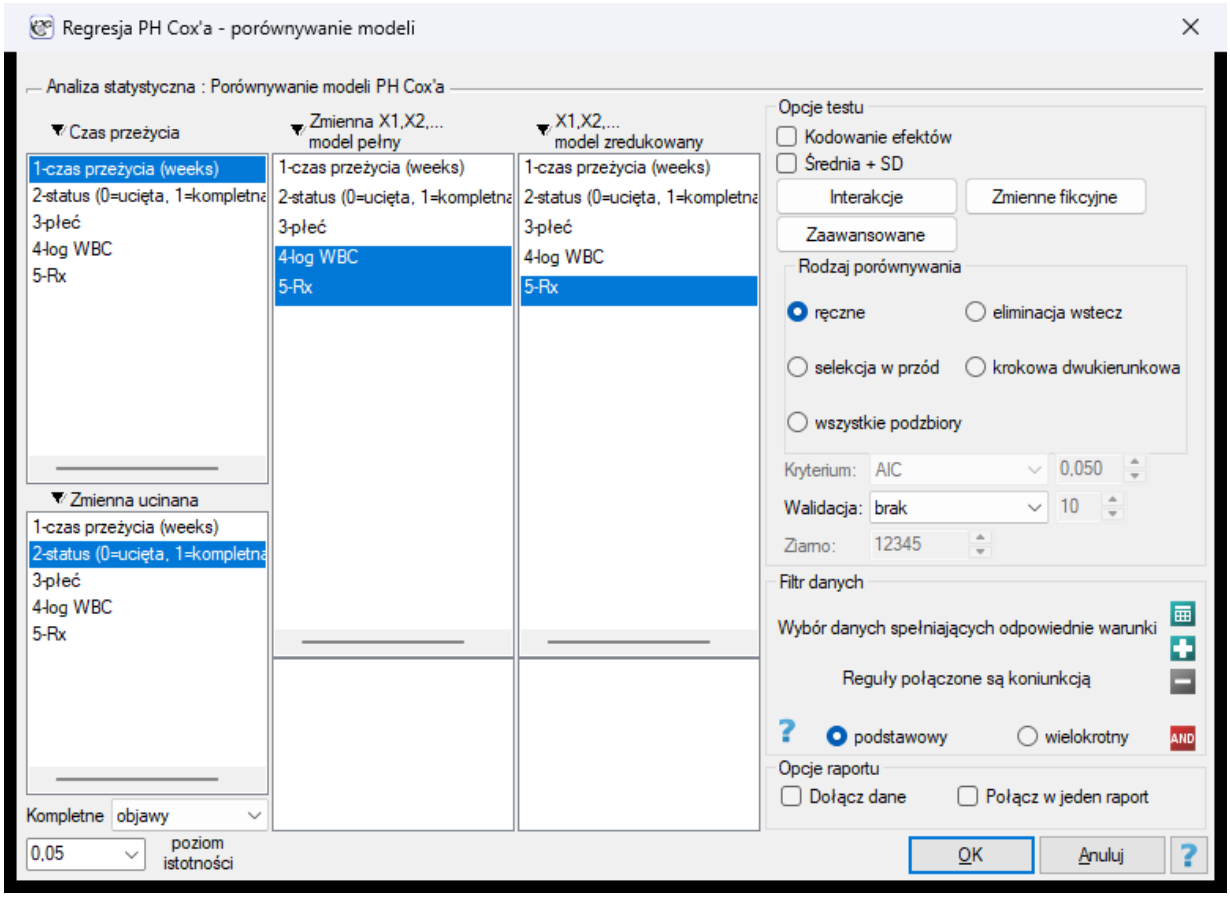
poza zbiorem uczącym, w których obserwacja o wyższym przewidywanym ryzyku doświadczyła zdarzenia wcześniej (0.5 – uporządkowanie losowe, 1 – uporządkowanie doskonałe).

Punktem odniesienia jest **model zerowy** (bez zmiennych – jednakowe ryzyko wszystkich obserwacji); jego błąd walidacji raport podaje w bloku doboru modelu wraz ze wskazaniem **najlepszego modelu wg walidacji**, niezależnie od wybranego kryterium doboru. Ocena poza zbiorem uczącym działa także przy ręcznym porównywaniu modeli.

Uwaga! Gdy błąd walidacji jest kryterium doboru, błąd walidacji wybranego modelu jest nieco zaniżony – spośród wielu modeli wybrano ten, który na danym podziale wypadł najlepiej. Przy małej liczbie zdarzeń wyniki walidacji zależą od losowego podziału – warto wtedy sprawdzić je dla kilku wartości ziarna.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji porównywania modeli wywołujemy poprzez menu Statystyka→Analiza przeżycia→Regresja PH Coxa – porównywanie modeli.



W oknie wskazujemy czas przeżycia, zmienną ucinaną z wartością Kompletne oraz zmienne niezależne modelu pełnego i zredukowanego (w doborze automatycznym – zmienne kandydujące). Opcje doboru modelu:

– Rodzaj porównywania – ręczne albo jedna z metod doboru automatycznego,

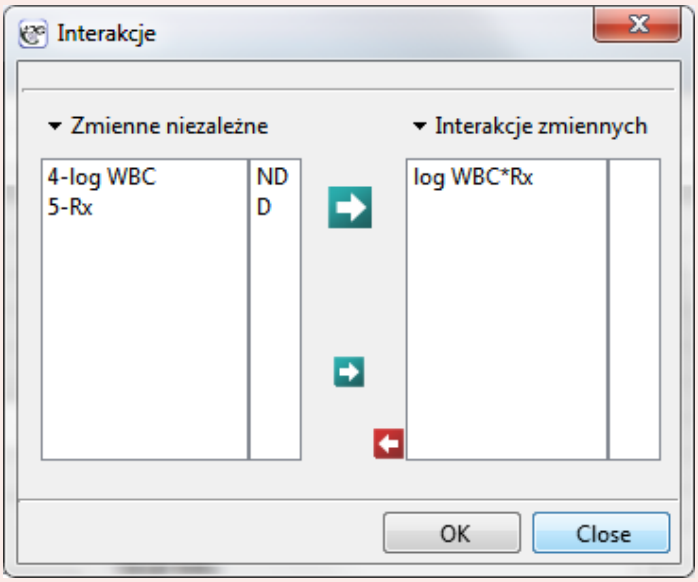
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

- Kryterium – AIC, AICc, BIC, p (test LR / F) z progiem podawanym w polu obok albo błąd walidacji; w trybie ręcznym kryterium nie jest używane,
- Walidacja – brak (ocena tylko na zbiorze uczącym), krzyżowa, k = z liczbą podziałów w polu obok (od 2 do 100, domyślnie 10) albo zbiór testowy, % = z udziałem zbioru testowego w procentach obserwacji (od 5 do 90, domyślnie 30),
- Ziarno – ziarno losowania podziału danych (domyślnie 12345); to samo ziarno oznacza ten sam podział.

WYKONANIE I INTERPRETACJA

Przykład 28.2. (plik: remisjaBiałaczka.pqs)
Analiza opiera się na danych dotyczących białaczki opisanych w pracy Freireich i innych 1963[65] i analizowanych dalej przez wielu autorów, m.in. Kleinbaum i Klein 2005[96]. Dane zawierają informację o czasie (w tygodniach) pozostawania w remisji, aż do momentu wycofania pacjenta z badania z powodu wyjścia z remisji (nawrotu objawów) lub ucięcia informacji o pacjencie. Wyjście z remisji to zdarzenie niekorzystne – traktowane jest jako obserwacja **kompletna**. Obserwacja jest **ucięta**, jeśli pacjent pozostaje w badaniu do końca i remisja nie nastąpi, lub jeśli opuści badanie.
Pacjenci przydzieleni zostali do dwóch grup: leczonej tradycyjnie (oznaczonej jako 1, czasami nazywanej „grupą placebo") i leczonej nową metodą (oznaczonej jako 0). Zebrano informację o płci pacjentów (1=mężczyzna, 0=kobieta) oraz o wartościach wskaźnika liczby białych krwinek oznaczonego jako „log WBC", który jest znanym czynnikiem prognostycznym.

Celem badania jest określenie wpływu sposobu leczenia na czas pozostawania w remisji przy uwzględnieniu możliwych czynników wikłających (*confounder*) i interakcji. W analizie uwagę skupimy na zmiennej „Rx" (1=placebo, 0=nowe leczenie); zmienną „log WBC" umieścimy w modelu jako możliwy czynnik wikłający. By ocenić ewentualny wpływ interakcji „Rx" i „log WBC", rozważymy także trzecią zmienną będącą ich iloczynem, dołączaną do modelu przyciskiem Interakcje.



Budujemy trzy modele Coxa:

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Model A zawiera tylko zmienną „Rx”

Model	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Wald	Wartość p	Hazard rat	-95% CI	+95% CI
Rx	1,5091913	0,4095644	0,7064599	2,3119228	13,578263	0,0002288	4,5230719	2,0268035	10,093814

Model B zawiera zmienną „Rx” i potencjalną zmienną wikłającą „log WBC”

Model	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Wald	Wartość p	Hazard rat	-95% CI	+95% CI
log WBC	1,6043432	0,3293283	0,9588716	2,2498148	23,732115	0,0000011	4,9745910	2,6087511	9,4859783
Rx	1,2940672	0,4221039	0,4667586	2,1213757	9,3988516	0,0021712	3,6475921	1,5948164	8,3426073

Model C zawiera zmienną „Rx”, zmienną „log WBC” oraz potencjalny efekt ich interakcji: „Rx $\times$ log WBC”

Model	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Wald	Wartość p	Hazard rat	-95% CI	+95% CI
log WBC	1,8027879	0,4467169	0,9272388	2,6783371	16,286373	0,0000544	6,0665375	2,5275205	14,560859
Rx	2,3549391	1,6810211	-0,939801	5,6496801	1,9625151	0,1612445	10,537488	0,3907052	284,20054
log WBC*F	-0,342195	0,5197406	-1,360867	0,6764778	0,4334850	0,5102838			

Zmienna interakcji „Rx” i „log WBC” zawarta w modelu C jest w nim nieistotna ($p = 0.5103$) według testu Walda. Podobny wynik uzyskamy, porównując testem ilorazu wiarygodności model C z modelem B w module Regresja Coxa – porównywanie modeli: wynik nieistotny ($p = 0.5134$) oznacza, że model C (z interakcją) nie jest istotnie lepszy od modelu B (bez interakcji).

Chi-kwadrat - porównanie modeli	0,427079874
Ilość stopni swobody	1
Wartość p	0,513425299

Porównanie modelu C i B: $p = 0.5134$ (nieistotne)

Odrzucamy model C – interakcja „Rx” $\times$ „log WBC” nie wnosi istotnej informacji. Przechodzimy do porównania modelu B i A.

HR dla „Rx” w modelu B wynosi 3.65, co oznacza, że hazard dla „grupy placebo” jest około 3.6 razy większy niż dla grupy leczonej nową metodą. Model A, zawierający tylko zmienną „Rx”, nazywany jest zwykle modelem „surowym” (*ang. crude model*), ponieważ ignoruje efekt potencjalnych zmiennych wikłających – w tym modelu HR dla „Rx” wynosi już 4.52, czyli sporo więcej niż w modelu B. Ważny jest tu też 95% przedział ufności dla tych estymatorów: dla „Rx” w modelu A ma on szerokość 8.06 (10.09 minus 2.03), a w modelu B jest węższy: 6.74 (8.34 minus 1.60) – model B daje więc bardziej precyzyjną estymację HR niż model A. By ostatecznie zdecydować, który z nich lepiej oszacowuje efekt leczenia, ponownie porównujemy je testem ilorazu wiarygodności.

Porównanie modelu B i A: $p < 0.0001$ (istotne)

Model B jest istotnie lepszy niż model A – ma najniższe wartości kryteriów informacyjnych (AIC=148.6, AICc=149, BIC=151.4) i wysokie wartości dopasowania (Pseudo $R^2_{McFadden} = 0.2309$, $R^2_{Nagelkerke} = 0.7662$, $R^2_{Cox-Snell} = 0.7647$).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Porównywanie modeli PH Cox'a	
Czas analizy	0,11 sek.
Analizowane zmienne	czas przeżycia (tygodnie);sta
Poziom istotności	0,05
Zmienna grupująca	Rx(0;1)
Liczba zmiennych w modelu 1	2
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood	144,558519789
AIC - kryterium Akaikego	148,558519789
AICc - skorygowane kryterium Akaikego	149,002964233
BIC - bayesowskie kryterium Schwarza	151,360914552
Pseudo R2 (McFadden)	0,230949395
R2 (Nagelkerke)	0,766193449
R2 (Coxa-Snella)	0,764737344
Liczba zmiennych w modelu 2	1
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood	172,759244379
AIC - kryterium Akaikego	174,759244379
AICc - skorygowane kryterium Akaikego	174,902101522
BIC - bayesowskie kryterium Schwarza	176,160441761
Pseudo R2 (McFadden)	0,080921681
R2 (Nagelkerke)	0,398474704
R2 (Coxa-Snella)	0,397717426
Chi-kwadrat - porównanie modeli	28,20072459
Ilość stopni swobody	1
Wartość p	0,000000109

Ocena poza zbiorem uczącym. Sprawdzimy jeszcze, czy przewaga modelu B utrzyma się dla pacjentów, którzy nie brali udziału w dopasowaniu modeli. Porównanie modeli B i A powtarzamy, wybierając w polu Walidacja opcję krzyżowa, k = z 10 podziałami (ziarno 12345). W raporcie model B (pełny) jest oznaczony jako Model 1, a model A (zredukowany) – jako Model 2.

Ocena poza zbiorem uczącym	walidacja krzyżowa
Liczba podziałów (k)	10
Ziarno losowania podziału	12345
Model zerowy: -2 ln PL / zdarzenie	8.1196
Najlepszy model wg walidacji (-2 ln PL / zdarzenie)	Model 1

	zbiór uczący	walidacja krzyżowa ($k = 10$)	
model	AIC	$-2 \ln PL$ na zdarzenie	C Harrella
zerowy		8.1196	
Model B (Rx, log WBC)	148.5585	6.7100	0.8871
Model A (Rx)	174.7592	7.6269	0.7258

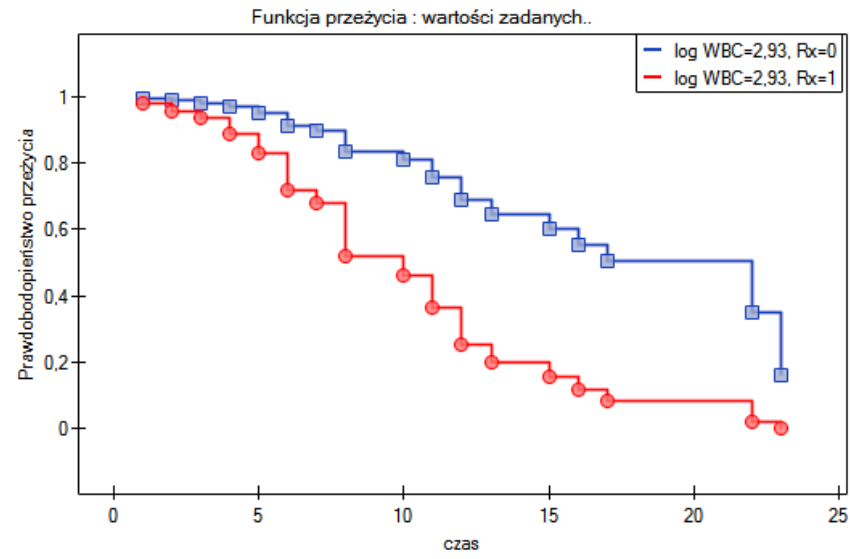
Model A, zawierający tylko sposób leczenia, przewiduje czas pozostawania w remisji lepiej niż model zerowy (7.6269 wobec 8.1196), ale dopiero dodanie „log WBC" wyraźnie zmniejsza błąd walidacji (do 6.7100) i podnosi zgodność przewidywań z rzeczywistą kolejnością zdarzeń z $C = 0.7258$ do $C = 0.8871$. Przewaga modelu B nie jest więc wynikiem dopasowania do danych, na których go zbudowano – utrzymuje się także dla pacjentów spoza zbioru uczącego.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

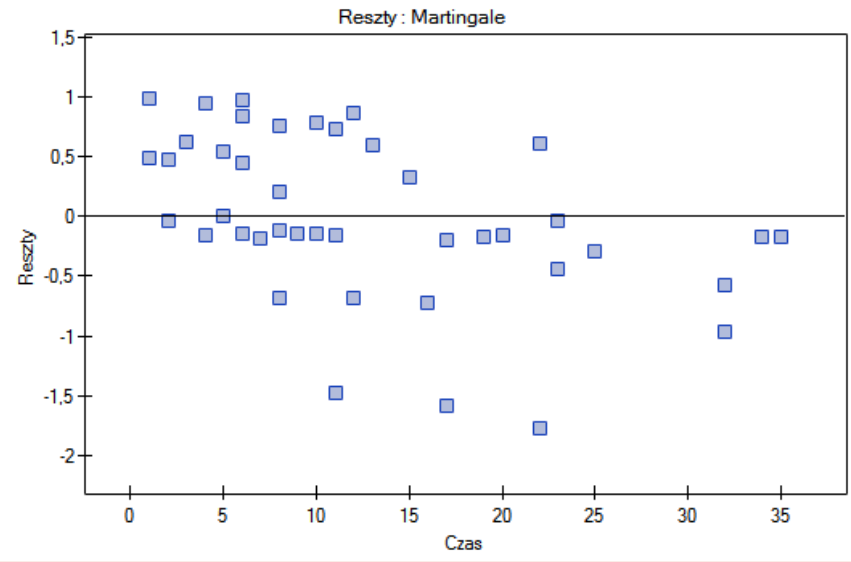
walidacja krzyżowa ($k = 10$): $-2 \ln PL$ na zdarzenie $= 6.71$ (model B), 7.63 (model A), 8.12 (model zerowy); C Harrella: 0.89 i 0.73

Najlepszym modelem według walidacji jest model B – „log WBC" poprawia przewidywanie także poza zbiorem uczącym.

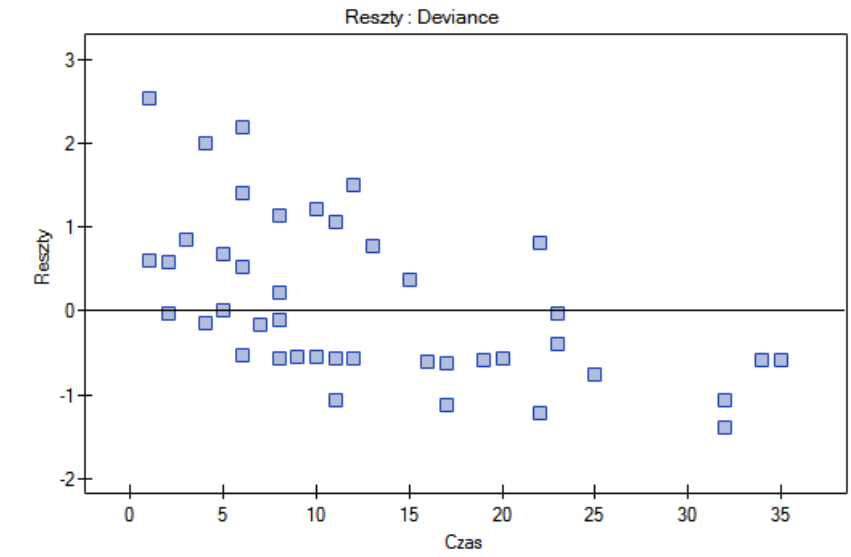
Dopełnieniem analizy jest przedstawienie dla modelu B krzywych przeżycia (pozostawania w remisji) obu grup leczenia, skorygowanych o wpływ „log WBC". By je wyrysować, po wybraniu opcji Dołącz wykres zaznaczamy Funkcja przeżycia: w podgrupach..., a następnie – by szybko zbudować wykres dwóch krzywych – wybieramy Szybkie podgrupy i wskazujemy zmienną Rx. Opcja Zaawansowane podgrupy pozwala zbudować dowolną liczbę dowolnie zdefiniowanych krzywych.

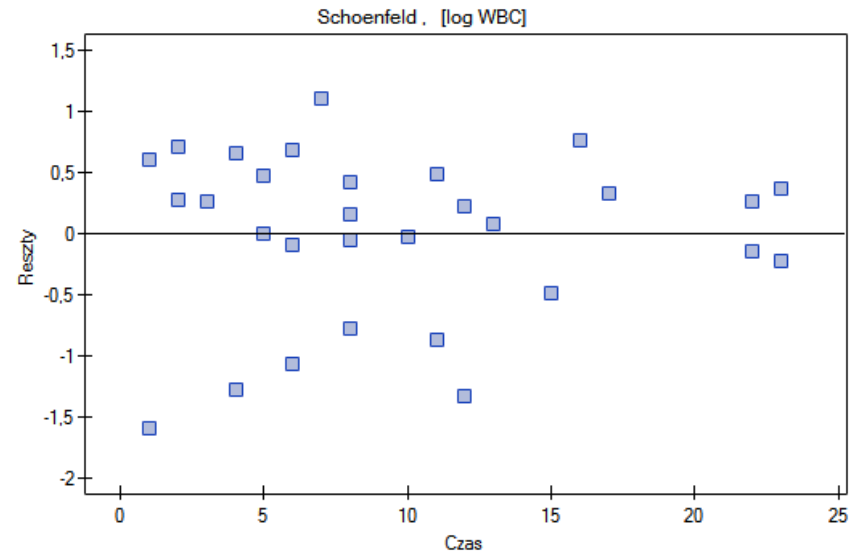


Na koniec ocenimy założenia regresji Coxa, analizując reszty modelu względem czasu.

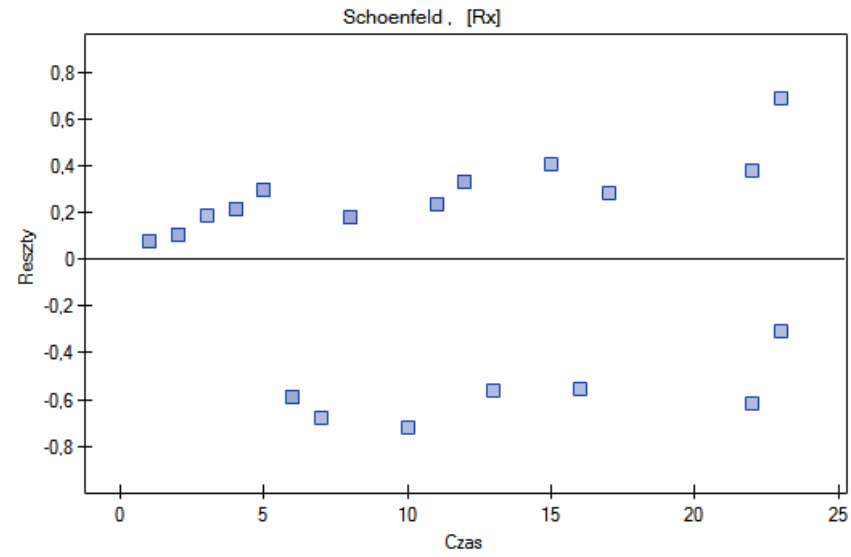


Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Nie obserwujemy punktów odstających, choć reszty Martingale i Deviance są coraz niższe dla coraz dłuższego czasu. Reszty Schoenfelda mają symetryczny rozkład względem czasu. W przypadku reszt Schoenfelda analizę wykresu można wesprzeć testami sprawdzającymi, czy punkty wykresu reszt układają się w pewien wzór, np. zależność liniową – w tym celu przekopiowujemy reszty Schoenfelda oraz czas do arkusza danych i testujemy szukaną zależność. Wynik takiego testu dla każdej zmiennej wskazuje na spełnienie założenia proporcjonalności hazardu (gdy jest nieistotny statystycznie) lub na jego złamanie (gdy jest istotny) – taką zmienną można wówczas wyłączyć z modelu. Dla zmiennych „log WBC" i „Rx" symetryczny rozkład reszt sugeruje spełnienie założenia proporcjonalności hazardu; potwierdzeniem może być sprawdzenie zależności liniowej Pearsona lub monotonicznej Spearmana dla tych reszt i czasu.

Dalej możemy dodać do modelu zmienną płeć – ostrożnie, bo z literatury wiadomo, że płeć może wpływać na funkcję przeżycia w remisji białaczki w sposób nieproporcjonalny względem czasu. Budujemy więc model Coxa dla trzech zmiennych: „Płeć", „Rx" i „log WBC", a przed interpretacją współczynników sprawdzamy reszty Schoenfelda – przekopiowując je wraz z czasem do nowego arkusza i sprawdzając zależność monotoniczną Spearmana. Uzyskujemy $p = 0.0259$ (dla płci), $p = 0.6192$ (dla log WBC) i $p = 0.1490$ (dla Rx), co potwierdza naruszenie założenia proporcjonalności hazardu przez zmienną płeć.

Reszty Schoenfelda vs czas: $p = 0.0259$ dla płci (istotne)

Zmienna „Płeć" łamie założenie proporcjonalności hazardu – model Coxa budujemy oddzielnie dla kobiet i mężczyzn, wykonując analizę dwukrotnie z włączonym filtrem danych (najpierw płeć żeńska, potem męska).

Dla kobiet

Model

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Wald	Wartość p	Hazard rat	-95% CI	+95% CI
log WBC	1,1701250	0,4985684	0,1929488	2,1473011	5,5082668	0,0189267	3,2223954	1,2128207	8,5617207
Rx	0,2667231	0,5659162	-0,842452	1,3758985	0,2221350	0,6374179	1,3056788	0,4306531	3,9586323

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dla mężczyzn

Model

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Wald	Wartość p	Hazard rat	-95% CI	+95% CI
log WBC	1,6389171	0,5190378	0,6216215	2,6562126	9,9704762	0,0015907	5,1495899	1,8619449	14,242245
Rx	1,8590474	0,7291016	0,4300345	3,2880603	6,5013690	0,0107791	6,4176205	1,5373105	26,790850

DLA ZAINTERESOWANYCH

Ocena poza zbiorem uczącym

Niech $\ell(\beta)$ oznacza logarytm wiarygodności częściowej (z poprawką Breslowa dla czasów wiązanych) wyznaczony dla współczynników β. W walidacji krzyżowej wykorzystywana jest wiarygodność częściowa z walidacji krzyżowej Verweija i van Houwelingena[168]:

$$CVPL = \sum_{j=1}^{k} \left[\ell\left(\hat{\beta}_{(-j)}\right) - \ell_{(-j)}\left(\hat{\beta}_{(-j)}\right) \right],$$

gdzie $\hat{\beta}_{(-j)}$ to współczynniki oszacowane bez j-tej części danych, ℓ – logarytm wiarygodności częściowej wszystkich obserwacji, a $\ell_{(-j)}$ – obserwacji bez j-tej części. Błąd walidacji to $-2\,CVPL/d$, gdzie d jest liczbą zdarzeń. Przy zbiorze testowym błąd walidacji wynosi $-2\,\ell_T(\hat{\beta}_U)/d_T$, gdzie ℓ_T jest logarytmem wiarygodności częściowej obserwacji testowych, $\hat{\beta}_U$ – współczynnikami oszacowanymi na zbiorze uczącym, a d_T – liczbą zdarzeń w zbiorze testowym. W modelu zerowym przyjmuje się $\beta = 0$.

Współczynnik C Harrella[79] jest ilorazem liczby par zgodnych do liczby par porównywalnych; pary tworzone są wewnątrz każdej ocenianej części danych (w walidacji krzyżowej – wewnątrz każdego podziału, dla przewidywań modelu dopasowanego bez tego podziału) i zliczane łącznie. Para jest porównywalna, gdy obserwacja o krótszym czasie jest kompletna, a zgodna – gdy ta obserwacja ma wyższe przewidywane ryzyko (para o jednakowym przewidywanym ryzyku liczona jest jako połowa pary zgodnej).

Regresja Coxa ze zmiennymi zależnymi od czasu

KIEDY STOSOWAĆ:

Podstawowym założeniem modelu Coxa jest **proporcjonalność hazardu**: wpływ każdej zmiennej niezależnej na hazard musi być stały w czasie, czyli Iloraz Hazardów HR_i nie zmienia się wraz z upływem czasu. Gdy to założenie jest złamane dla którejś ze zmiennych (przesłanką może być np. nielosowy, systematyczny wzór reszt Schoenfelda względem czasu – patrz Analiza reszt modelu), jednym z rozwiązań – obok stratyfikacji analizy (osobne modele budowane dla poszczególnych poziomów tej zmiennej) – jest rozszerzenie modelu o dodatkowy człon opisujący, jak efekt zmiennej zmienia się w czasie.

DEFINICJE MIAR

Rozszerzenie modelu polega na dodaniu członu będącego iloczynem badanej zmiennej niezależnej X_j i pewnej funkcji czasu $g(t)$:

$$h(t, X_1, ..., X_k) = h_0(t) \cdot \exp\left(\sum_{i=1}^{k} \beta_i X_i + \delta \cdot X_j \cdot g(t) \right),$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie iloczyn $X_j \cdot g(t)$ nazywany jest **zmienną zależną od czasu** (*ang. time-dependent covariate*). W efekcie hazard cząstkowy związany ze zmienną X_j przestaje być stały i zależy od czasu:

$$HR_j(t) = e^{\beta_j + \delta \cdot g(t)}.$$

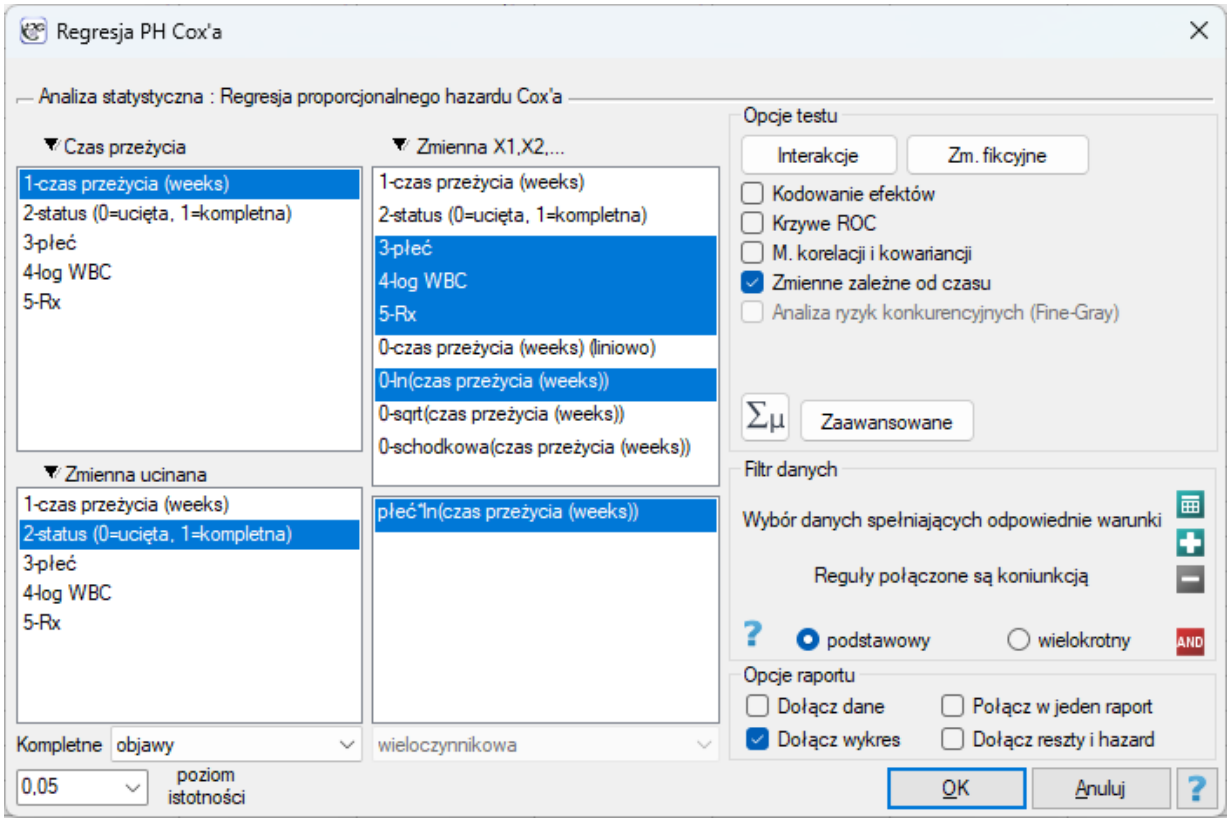
Test Walda dla współczynnika δ pozwala sprawdzić, czy efekt zmiennej X_j rzeczywiście zmienia się w czasie: $\delta = 0$ oznacza, że efekt X_j jest stały (założenie proporcjonalności hazardu spełnione), a $\delta \neq 0$ – że efekt zmienia się w czasie (założenie złamane).

W programie PQStat funkcję $g(t)$ można wybrać spośród czterech dodatkowych pseudo-zmiennych czasu, z którymi buduje się iloczyn z dowolną zmienną niezależną modelu:

- **czas (liniowo)** – $g(t) = t$,
- **ln(czas)** – $g(t) = \ln(t)$,
- **sqrt(czas)** – $g(t) = \sqrt{t}$,
- **schodkowa(czas)** – $g(t)$ jest funkcją skokową (Heaviside'a), przyjmującą wartość 0 przed ustalonym progiem czasu i 1 po jego przekroczeniu.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji regresji Coxa ze zmiennymi zależnymi od czasu wywołujemy poprzez menu Statystyka→Analiza przeżycia→Regresja PH Coxa – to samo okno, co dla standardowej regresji Coxa, w którym zaznaczamy dodatkowo opcję Zmienne zależne od czasu. Po jej zaznaczeniu i wskazaniu zmiennej definiującej czas przeżycia, na liście zmiennych do budowy interakcji (przycisk Interakcje) pojawiają się cztery dodatkowe pseudo-zmienne czasu wymienione powyżej – funkcję $g(t)$ definiujemy więc, wybierając jedną z nich.

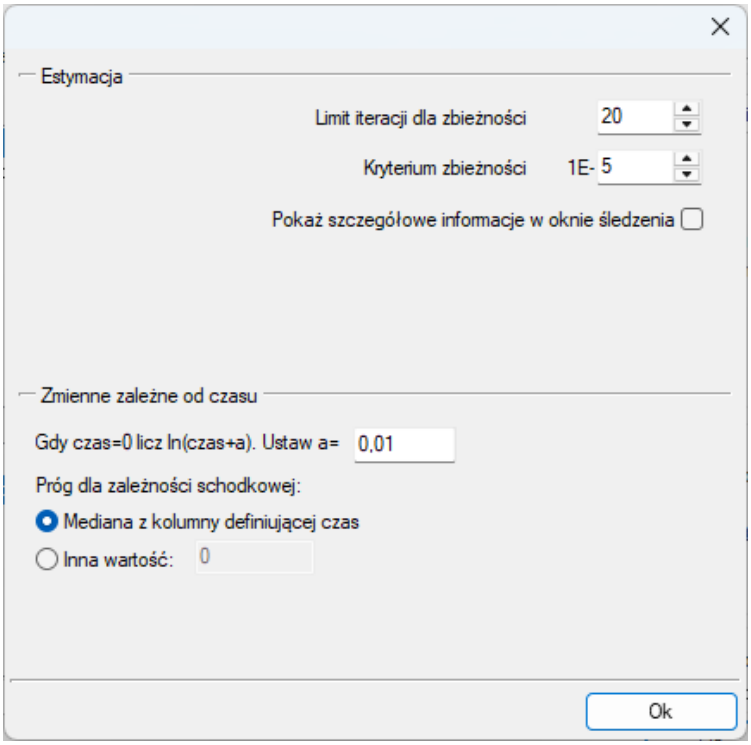


Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Uwaga! Opcja Zmienne zależne od czasu oraz opcja Analiza ryzyk konkurencyjnych (Fine-Gray) wzajemnie się wykluczają.

W oknie ustawień zaawansowanych (przycisk z ikoną koła zębatego), obok znanych już parametrów zbieżności algorytmu Newtona-Raphsona, dostępne są dodatkowo dwa parametry związane ze zmiennymi zależnymi od czasu:

- wartość a używana do wyliczenia $\ln(t+a)$ dla obserwacji, dla których czas przeżycia wynosi 0 (domyślnie funkcja logarytmiczna nie jest wówczas określona),
- próg dla funkcji schodkowej – może nim być mediana zmiennej definiującej czas (opcja domyślna) albo dowolna inna, wskazana przez badacza wartość.



WYKONANIE I INTERPRETACJA

Przykład (28.2) c.d. *(plik remisjaBiałaczka.pqs)*

W przykładzie dotyczącym remisji białaczki analiza reszt Schoenfelda wykazała naruszenie założenia proporcjonalności hazardu przez zmienną płeć (1=mężczyzna, 0=kobieta; istotna zależność monotoniczna reszt od czasu, p=0.0259), podczas gdy dla zmiennych Rx i log WBC założenie to było spełnione. Zanim zbudujemy model rozszerzony, przyjrzyjmy się, jak to naruszenie wygląda w danych i do czego prowadzi w zwykłym modelu Coxa.

Krok 1. Krzywe Kaplana-Meiera dla płci

Porównanie krzywych przeżycia (pozostawania w remisji) kobiet i mężczyzn w module Porównanie grup daje wynik nieistotny: test log-rank $\chi^2 = 0.56$, $p = 0.4555$, iloraz hazardów $HR_{m/k} = 0.78$ (95%CI: $\langle 0.38;\ 1.60 \rangle$).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

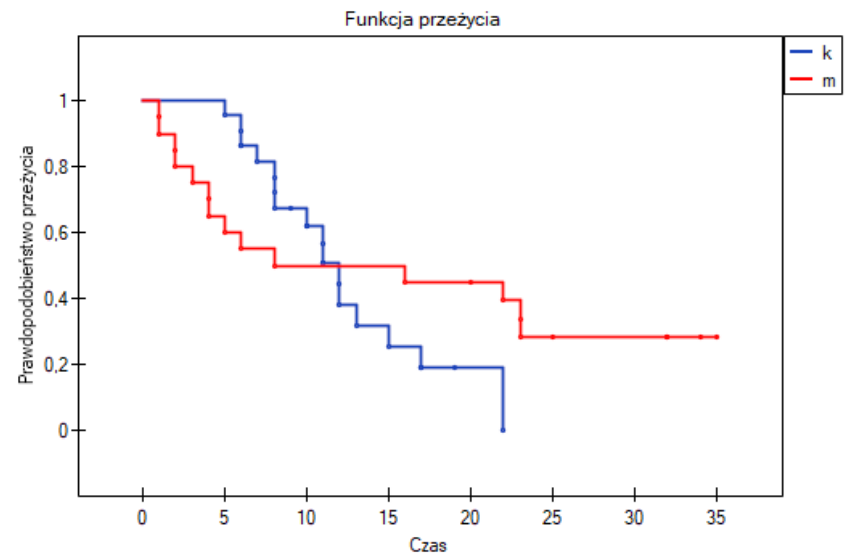
Porównanie grup	
Analizowane zmienne	czas przeżycia (weeks)
	status (0=ucięta, 1=komplet
Liczność / niezinterpretowane	0
Liczność / braki danych	0
Poziom istotności	0,05
Zmienna grupująca	płeć
Liczność	42
Kompletne	objawy
Ucięte	brak objawów
Test: LogRank	
Statystyka Chi-kwadrat	0,556864
Stopnie swobody	1
Wartość p	0,455526

Logrank

Grupa	Obs.	Ocz.	Obs./Ocz.
k	16	14,156938	1,130188
m	14	15,843062	0,883668

Logrank

Grupa	Hazard r.	-95%CI	+95%CI
m/k	0,781877	0,381796	1,601199



Wykres pokazuje jednak, dlaczego test „nie widzi" różnicy: krzywe **przecinają się** około 11.–12. tygodnia. W pierwszych tygodniach mężczyźni wychodzą z remisji znacznie szybciej (ich krzywa leży niżej), ale ci, którzy przetrwają ten okres, pozostają w remisji dłużej niż kobiety – po 12. tygodniu krzywa dla mężczyzn leży już wyżej i stabilizuje się na poziomie ok. 0.28, podczas gdy krzywa dla kobiet spada do 0. Przecinające się krzywe to klasyczny obraz **złamania założenia proporcjonalności hazardu**: hazard mężczyzn względem kobiet nie jest stały, lecz zmienia kierunek w czasie. Test log-rank, który przypisuje wszystkim punktom czasu tę samą wagę, sumuje różnice o przeciwnych znakach – i traci moc.

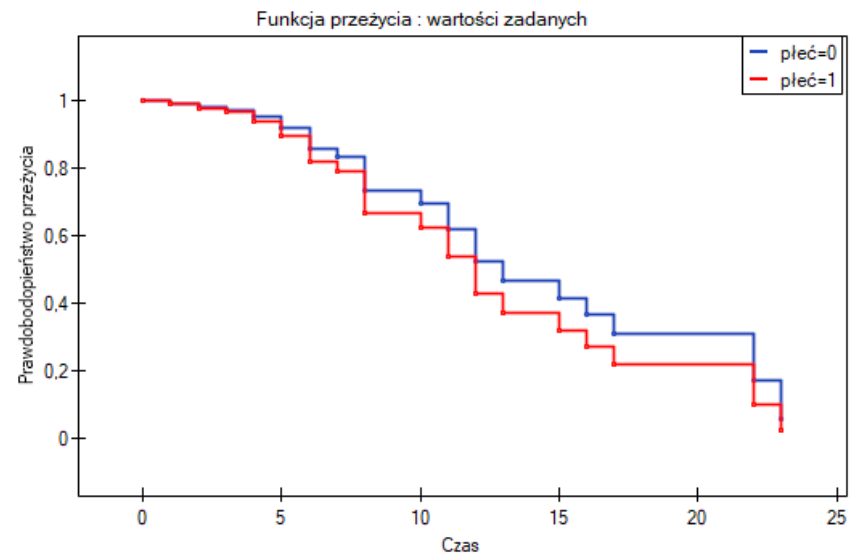
Krok 2. Model Coxa ignorujący złamanie założenia

Ten sam efekt „znoszenia się" obserwujemy w zwykłym modelu Coxa z trzema zmiennymi: płeć, log WBC i Rx.

Model

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	Hazard ratio	-95% CI	+95% CI
płeć	0,263171	0,449435	-0,617706	1,144048	0,342879	0,558172	1,301049	0,53918	3,13945
log WBC	1,593619	0,329996	0,946839	2,240399	23,32127	0,000001	4,921527	2,577549	9,397077
Rx	1,390877	0,456646	0,495867	2,285886	9,277232	0,00232	4,018371	1,641922	9,834394

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Współczynnik przy zmiennej płeć jest nieistotny ($HR = 1.30$, 95%CI: $\langle 0.54; 3.14 \rangle$, $p = 0.5582$), a wyrysowane z modelu krzywe przeżycia dla obu płci (przy średnich pozostałych zmiennych) leżą blisko siebie i – z założenia modelu – nie mogą się przeciąć. Model uśrednia wczesny, bardzo niekorzystny efekt bycia mężczyzną z późniejszym efektem korzystnym i wychodzi mu „nic". Wniosek, że płeć nie ma znaczenia dla czasu remisji, byłby więc błędny – jest artefaktem niespełnionego założenia.

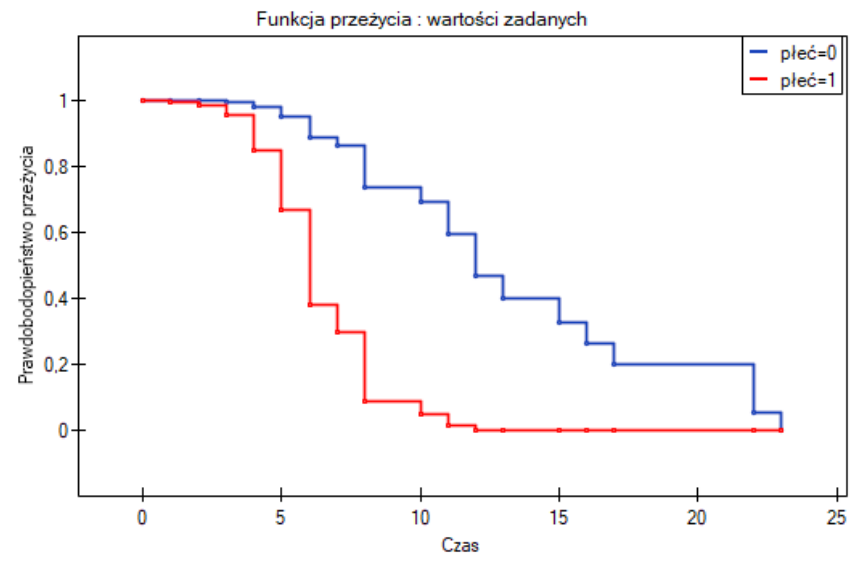
Krok 3. Model Coxa ze zmienną zależną od czasu

Zamiast dzielić analizę na dwa oddzielne modele (osobno dla kobiet i dla mężczyzn), budujemy jeden wspólny model, w którym płeć wchodzi w interakcję z funkcją czasu. Ponieważ interesuje nas nie to, *czy*, ale *jak z upływem czasu* zmienia się wpływ płci, wybieramy funkcję logarytmiczną $g(t) = \ln(t)$ – efekt zmiennych klinicznych w analizie przeżycia zmienia się zwykle najszybciej na początku obserwacji i stabilizuje się z czasem, co dobrze oddaje logarytm. W oknie analizy zaznaczamy opcję Zmienne zależne od czasu, a w oknie Interakcje definiujemy iloczyn płeć×ln(czas). Model zawiera więc cztery człony: płeć, log WBC, Rx oraz płeć×ln(czas).

Model

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	Hazard ratio	-95% CI	+95% CI
płeć	4,057084	1,944145	0,24663	7,867537	4,354826	0,036904	57,805481	1,279706	2611,12651
log WBC	1,409521	0,335513	0,751927	2,067115	17,649162	0,000027	4,093995	2,121084	7,901993
Rx	0,985126	0,478369	0,04754	1,922711	4,240899	0,039461	2,678149	1,048689	6,839476
płeć x ln(czas pı	-1,911986	0,921656	-3,718398	-0,105574	4,303601	0,038032			

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Współczynnik interakcji płeć×ln(czas) wynosi $\delta = -1.912$ ($SE = 0.922$, $\chi^2_{Walda} = 4.30$, $p = 0.0380$) i jest istotny statystycznie – potwierdza to formalnie, że efekt płci zmienia się w czasie, czyli to, co wcześniej sugerowały reszty Schoenfelda i przecinające się krzywe Kaplana-Meiera. Istotny stał się też sam współczynnik przy zmiennej płeć: $\beta_{pe} = 4.057$ ($p = 0.0369$).

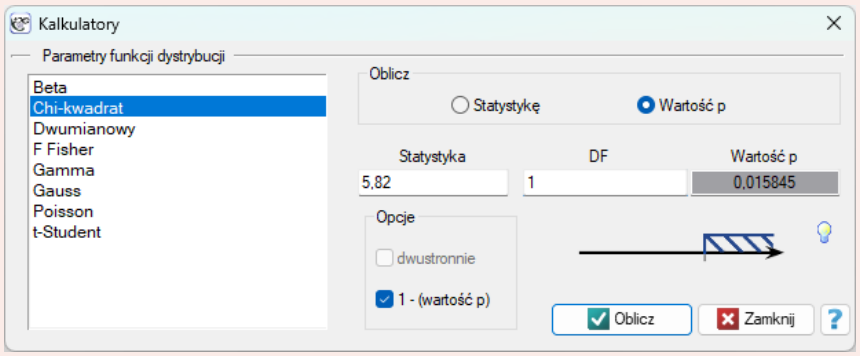
Interpretując te dwa współczynniki, trzeba pamiętać, że β_{pe} opisuje efekt płci tylko w chwili $t = 1$ (bo $\ln(1) = 0$), a nie efekt uśredniony w całym okresie obserwacji. Hazard mężczyzn względem kobiet w dowolnym tygodniu t wyliczamy ze wzoru $HR_{pe}(t) = e^{\beta_{pe} + \delta \cdot \ln(t)} = e^{4.057 - 1.912 \cdot \ln(t)}$:

t (tygodnie)	1	2	4	8	12	16	23
$HR_{pe}(t)$	57.8	15.4	4.1	1.1	0.50	0.29	0.14

Na początku obserwacji ryzyko wyjścia z remisji jest u mężczyzn wielokrotnie wyższe niż u kobiet, ale przewaga ta szybko maleje; po około 8 tygodniach ($t^* = e^{4.057/1.912} \approx 8.4$) hazardy obu płci się zrównują, a później kierunek zależności się odwraca – mężczyźni, którzy pozostali w remisji, mają już ryzyko nawrotu kilkukrotnie *niższe* niż kobiety. Dokładnie taki obraz pokazywały krzywe Kaplana-Meiera z kroku 1 – model ze zmienną zależną od czasu nie tylko go potwierdza, ale pozwala go zmierzyć, z jednoczesną korekcją o log WBC i Rx.

Model rozszerzony jest też po prostu lepszy od modelu z kroku 2: ma wyższą wiarygodność ($-2 \ln L = 138.4$ wobec 144.2), niższe kryteria informacyjne ($AIC = 146.4$ wobec 150.2; $BIC = 152.0$ wobec 154.4) i wyższe współczynniki dopasowania (Pseudo $R^2_{McFadden} = 0.264$ wobec 0.233). Test ilorazu wiarygodności porównujący oba modele (różnica $-2 \ln L$ równa 5.82 przy 1 stopniu swobody) daje $p = 0.0158$, czyli dodanie członu zależnego od czasu istotnie poprawia model.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



$\delta_{pe\times\ln(t)} = -1.91$, $p = 0.0380$; $HR_{pe}(t)$ **od 57.8 (**$t = 1$**) przez 1.0 (**$t \approx 8.4$**) do 0.14 (**$t = 23$**)**

Efekt płci nie jest stały w czasie: mężczyźni mają znacznie wyższe ryzyko wyjścia z remisji w pierwszych tygodniach, a znacznie niższe później. Zwykły model Coxa (i test log-rank) uśredniają te przeciwne efekty do nieistotnego „braku różnicy”; model ze zmienną zależną od czasu ujawnia i mierzy tę zmianę.

DLA ZAINTERESOWANYCH

Test Walda dla współczynnika interakcji ze zmienną czasu
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \delta = 0 \quad (\text{efekt } X_j \text{ jest stały w czasie} - \text{założenie proporcjonalności hazardu spełnione}), \\ \mathcal{H}_1: &\quad \delta \neq 0 \quad (\text{efekt } X_j \text{ zmienia się w czasie} - \text{złamanie założenia proporcjonalności hazardu}). \end{aligned}$$

Statystyka testowa oraz sposób wnioskowania są analogiczne jak dla testu Walda pozostałych współczynników modelu Coxa.

Analiza ryzyk konkurencyjnych – model Fine’a-Graya

KIEDY STOSOWAĆ:

W wielu badaniach obserwowane jednostki mogą doświadczyć jednego z kilku, wzajemnie wykluczających się rodzajów zdarzeń niepożądanych. Przykładowo pacjent onkologiczny może umrzeć z powodu choroby podstawowej, ale może też umrzeć z zupełnie innej przyczyny (np. sercowo-naczyniowej), zanim dojdzie do zdarzenia będącego przedmiotem badania. Wystąpienie takiego alternatywnego zdarzenia nazywane jest **zdarzeniem konkurencyjnym** (*ang. competing risk*) – uniemożliwia ono zaobserwowanie zdarzenia badanego, a jednocześnie nie jest zwykłym cenzurowaniem, ponieważ dostarcza dodatkowej informacji (wiemy, dlaczego obserwacja się zakończyła).

Zastosowanie standardowego modelu Coxa, w którym zdarzenia konkurencyjne potraktowane zostałyby jak zwykłe cenzurowanie, prowadzi do obciążonych oszacowań – standardowa funkcja $1 - S(t)$ przecenia wówczas prawdziwe ryzyko zdarzenia badanego, ponieważ część osób nigdy by go nie doświadczyła (zmarły wcześniej z innej przyczyny). Właściwym narzędziem analizy w takiej sytuacji jest **skumu-**

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

lowana funkcja zapadalności (*ang. cumulative incidence function, CIF*) oraz model regresji **Fine'a-Graya** (Fine i Gray 1999)[63], oparty na tzw. **subdystrybucyjnym hazardzie** (*ang. subdistribution hazard*).

DEFINICJE MIAR

Skumulowana funkcja zapadalności (CIF) wyraża, dla zdarzenia badanego, prawdopodobieństwo jego wystąpienia do czasu t, przy uwzględnieniu faktu, że zdarzenia konkurencyjne mogą temu zapobiec. Suma CIF dla zdarzenia badanego, CIF dla zdarzeń konkurencyjnych oraz funkcji przeżycia $S(t)$ w każdym punkcie czasu wynosi 1.

Model subdystrybucyjnego hazardu Fine'a-Graya definiuje subdystrybucyjny hazard $h^*(t)$ analogicznie do tego, jak model Coxa opisuje standardowy hazard, opiera się jednak na innym niż standardowo zbiorze ryzyka: obiekty, u których wystąpiło zdarzenie konkurencyjne, **nie są usuwane** ze zbioru ryzyka – pozostają w nim (z odpowiednią wagą) aż do końca obserwacji, symulując sytuację, w której nigdy nie mogłyby doświadczyć zdarzenia badanego, ale nadal są obserwowane. Wagi te (*ang. inverse probability of censoring weighting, IPCW*) wyliczane są w oparciu o odwrotność estymatora Kaplana-Meiera rozkładu cenzurowania. Parametry modelu szacowane są – podobnie jak w standardowym modelu Coxa – metodą największej wiarygodności (algorytm iteracyjny Newtona-Raphsona), z tą różnicą, że funkcja wiarygodności uwzględnia opisane powyżej wagi IPCW.

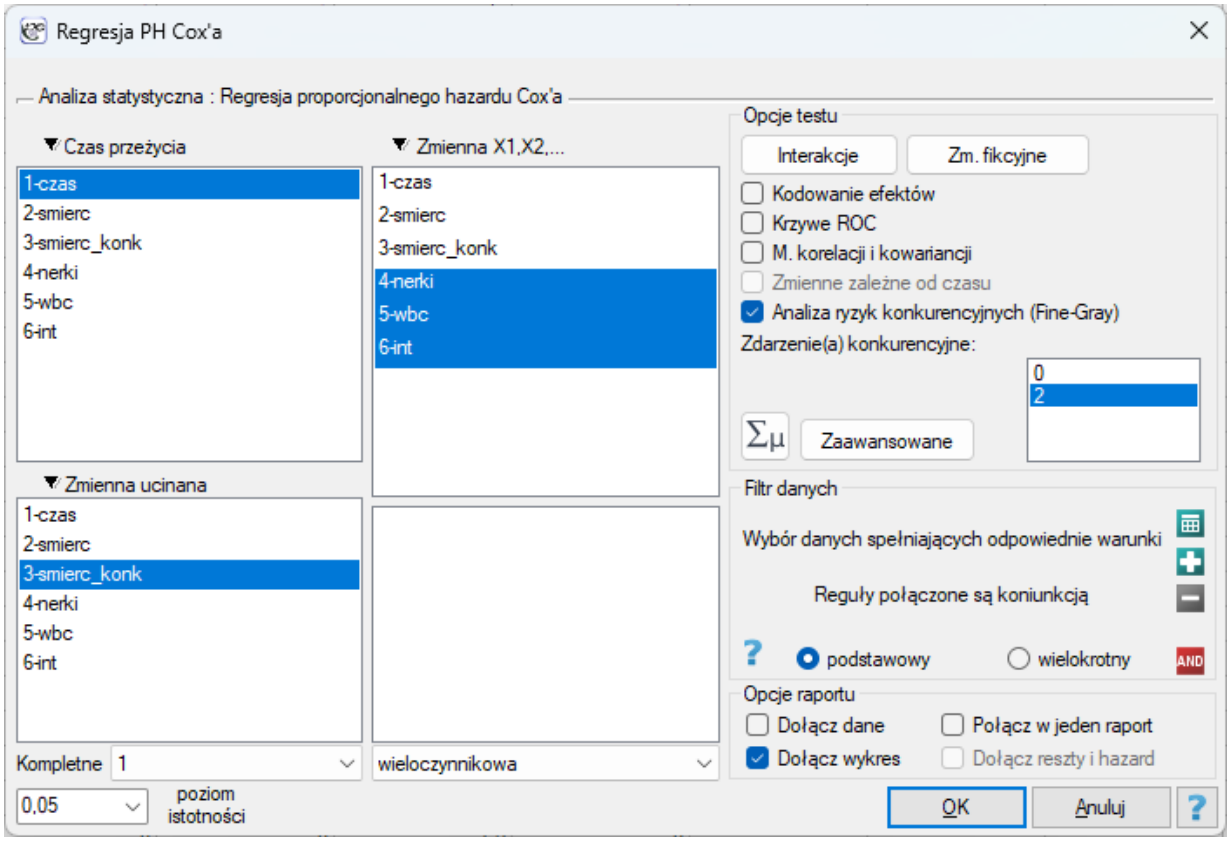
Subdystrybucyjny Iloraz Hazardów (sHR) – dla każdej zmiennej niezależnej wyliczany jest subdystrybucyjny iloraz hazardów (*ang. subdistribution Hazard Ratio* – sHR). W odróżnieniu od standardowego HR modelu Coxa (opisującego wpływ zmiennej na tempo występowania zdarzenia badanego wyłącznie u osób nadal „zagrożonych" tym zdarzeniem), sHR opisuje wpływ zmiennej niezależnej bezpośrednio na skumulowaną funkcję zapadalności $CIF(t)$, co czyni go łatwiejszym do interpretacji klinicznej (np. do prognozowania) w kontekście ryzyk konkurencyjnych. Kierunek interpretacji ($sHR > 1$, $sHR < 1$, $sHR \approx 1$) jest analogiczny jak dla standardowego HR opisanego wcześniej, a istotność poszczególnych współczynników weryfikowana jest również testem Walda.

Uwaga! Model Fine'a-Graya jest osobnym modelem od standardowego modelu Coxa liczonego na tych samych danych, w którym zdarzenia konkurencyjne po prostu wykluczono lub pominięto (uznano za brak zdarzenia) – oba modele odpowiadają na różne pytania badawcze i mogą prowadzić do odmiennych, a czasem wręcz przeciwstawnych wniosków (opisywany w literaturze efekt zmiany kierunku, *ang. sign reversal*) dla tej samej zmiennej niezależnej.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji modelu Fine'a-Graya wywołujemy poprzez menu Statystyka→Analiza przeżycia→Regresja PH Coxa – to samo okno, co dla standardowej regresji Coxa, w którym zaznaczamy dodatkowo opcję Analiza ryzyk konkurencyjnych (Fine-Gray). Po jej zaznaczeniu, na podstawie wartości zmiennej wskazanej jako zmienna ucinana, pojawia się lista Zdarzenie(a) konkurencyjne:, na której wskazujemy – spośród wartości innych niż kod zdarzenia badanego – jedną lub więcej wartości oznaczających zdarzenia konkurencyjne. Wartości statusu niewybrane ani jako zdarzenie badane, ani jako zdarzenie konkurencyjne pozostają prawdziwym cenzurowaniem.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Uwaga! Z uwagi na inny sposób liczenia reszt modelu (reszty Martingale, Deviance i Schoenfelda nie są zdefiniowane dla CIF), opcja Dołącz reszty i hazard jest dla modelu Fine'a-Graya niedostępna, podobnie jak jednoczesne zaznaczenie opcji Zmienne zależne od czasu. Dostępna jest natomiast opcja Krzywa ROC, czyli AUC zależne od czasu – w wersji uwzględniającej zdarzenia konkurencyjne (patrz niżej).

W raporcie z analizy Fine'a-Graya liczności obserwacji kompletnych i uciętych są rozbite dodatkowo o informację, jaki kod (wartość zmiennej statusu) im odpowiadał, a liczba obserwacji uciętych jest dodatkowo podzielona na poszczególne zdarzenia konkurencyjne. Kolumna standardowego HR jest zaś zastąpiona kolumną sHR (*Fine-Gray*).

Analogicznie do standardowego modelu Coxa (opcja Dołącz wykres), dla modelu Fine'a-Graya program pozwala wyrysować **wykresy skumulowanej funkcji zapadalności (CIF)**:

- **bazową CIF** – skumulowaną zapadalność przy średnich wartościach kowariantów,
- **CIF dla wybranych wartości zmiennych** – analogicznie jak funkcję przeżycia w standardowym modelu Coxa, dla wskazanych przez badacza wartości zmiennych niezależnych,
- **CIF dla podgrup** – oddzielne krzywe CIF budowane dla poszczególnych kategorii wskazanej zmiennej grupującej (opcje Szybkie podgrupy / Zaawansowane podgrupy).

Uwaga! W przeciwieństwie do funkcji przeżycia Kaplana-Meiera, krzywa CIF nie dąży zwykle do 0 ani do 1 – jej wartość graniczna to szacowane prawdopodobieństwo, że dana jednostka kiedykolwiek doświadczy zdarzenia badanego, zanim dojdzie u niej do zdarzenia konkurencyjnego.

AUC zależne od czasu dla ryzyk konkurencyjnych. Zdolność modelu Fine'a-Graya do rozróżniania pacjentów, u których wystąpi zdarzenie badane, oceniamy – tak jak dla standardowego modelu Coxa – za

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

pomocą AUC zależnego od czasu, z jedną istotną różnicą w definicji kontroli. Przypadkami w chwili t są wyłącznie osoby, które do czasu t doświadczyły **zdarzenia badanego**; kontrolami są natomiast zarówno osoby obserwowane dłużej niż t bez żadnego zdarzenia, jak i osoby, u których do czasu t wystąpiło **zdarzenie konkurencyjne** – one bowiem z pewnością nie doświadczą już zdarzenia badanego, a dobry model powinien przypisywać im niższe ryzyko. Interpretacja $AUC(t)$ pozostaje bez zmian: 0.5 – rozróżnianie losowe, 1 – doskonałe; wykres $AUC(t)$ względem czasu (opcja Krzywa ROC w oknie Dołącz wykres) pokazuje, jak zmienia się jakość prognozy zdarzenia badanego dla kolejnych horyzontów czasu.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 28.3. (plik: szpiczak_konkurencyjne.pqs)
W badaniu obejmującym 445 pacjentów ze zdiagnozowanym szpiczakiem mnogim porównywano skuteczność nowego leku (zmienna int: 0=leczenie standardowe, 1=nowy lek) w wydłużaniu czasu przeżycia (zmienna czas, w dniach). U pacjentów odnotowywano również wyjściową liczbę leukocytów (wbc) oraz obecność niewydolności nerek (nerki) – powikłania często towarzyszącego szpiczakowi mnogiemu i uznawanego za niekorzystny czynnik rokowniczy.

Zgon pacjenta mógł nastąpić z powodu progresji choroby nowotworowej (zdarzenie badane) albo z zupełnie innej przyczyny, np. sercowo-naczyniowej lub infekcyjnej, niezwiązanej bezpośrednio ze szpiczakiem (zdarzenie konkurencyjne). Informację tę koduje zmienna smierc_konk: 0 – pacjent żyje na koniec obserwacji (cenzurowanie), 1 – zgon z powodu szpiczaka, 2 – zgon z innej przyczyny. W danych dostępna jest też uproszczona zmienna smierc (zgon z jakiejkolwiek przyczyny, bez rozróżnienia powodu), wykorzystywana zwykle do standardowego modelu Coxa.

Model standardowy (nieprawidłowy dla tych danych)

Dla porównania zbudujemy najpierw standardowy model Coxa oparty na zmiennej smierc, w którym zgon z innej przyczyny niż szpiczak (zdarzenie konkurencyjne) błędnie potraktowany jest tak samo jak zgon z powodu szpiczaka (zdarzenie badane):

Model

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	Hazard ratio	-95% CI	+95% CI
nerki	2,432497	0,242623	1,956965	2,908028	100,51764	<0,000001	11,387278	7,077816	18,320641
wbc	0,053077	0,015474	0,022749	0,083405	11,765627	0,000603	1,054511	1,02301	1,086982
int	0,779894	0,294228	0,203217	1,356571	7,025902	0,008034	2,181242	1,225338	3,882858

Model Fine'a-Graya

Następnie budujemy właściwy dla tych danych model Fine'a-Graya, wskazując jako zmienną ucinaną smierc_konk, jako zdarzenie badane kod 1 (zgon z powodu szpiczaka) oraz zaznaczając na liście Zdarzenie(a) konkurencyjne: kod 2 (zgon z innej przyczyny).

Liczność	445
Kompletne (1)	43
konkurencyjne:	
(2)	45
Ucięte (0)	357

Model

	Wsp. b	Błąd b	-95% CI	+95% CI	Stat. Walda	Wartość p	sHR (Fine-Gr	-95% CI	+95% CI
nerki	3,654627	0,349788	2,969055	4,340198	109,16315	<0,000001	38,653086	19,473498	76,722769
wbc	0,057522	0,019021	0,020241	0,094804	9,145211	0,002494	1,059209	1,020448	1,099443
int	-0,406406	0,50775	-1,401578	0,588765	0,64065	0,423475	0,66604	0,246208	1,801763

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Uzyskany sHR dla zmiennej int wynosi $sHR = 0.67$ (95%CI: $\langle 0.25; 1.80 \rangle$), $p = 0.4235$, co oznacza brak podstaw, by twierdzić, że nowy lek istotnie wpływa na ryzyko zgonu z powodu szpiczaka mnogiego, gdy poprawnie uwzględni się konkurencyjne ryzyko zgonu z innej przyczyny. Inny obraz daje model standardowy (błędnie ignorujący rozróżnienie przyczyn zgonu), w którym HR dla int wynosi $HR = 2.18$, $p = 0.008034$ – sugerując istotny, niekorzystny wpływ nowego leku na ryzyko zgonu z jakiejkolwiek przyczyny. To dobra ilustracja opisanego wcześniej efektu zmiany kierunku (*sign reversal*): lek być może nieznacznie zwiększa ryzyko zgonu z innych przyczyn (np. w wyniku działań niepożądanych), co w modelu standardowym miesza się z ryzykiem zgonu z powodu samej choroby nowotworowej, sugerując pozorny, niekorzystny wpływ leku na przebieg szpiczaka – podczas gdy model Fine'a-Graya pokazuje, że lek nie ma istotnego wpływu na ryzyko zgonu specyficznie z powodu szpiczaka.

$HR_{int} = 2.18$, $p = 0.008034$ (model standardowy) wobec $sHR_{int} = 0.67$, $p = 0.4235$ (model Fine'a-Graya)

Nowy lek pozornie zwiększa ogólne ryzyko zgonu, ale nie zwiększa istotnie ryzyka zgonu specyficznie z powodu szpiczaka – efekt widoczny w modelu standardowym znika po poprawnym uwzględnieniu zdarzeń konkurencyjnych.

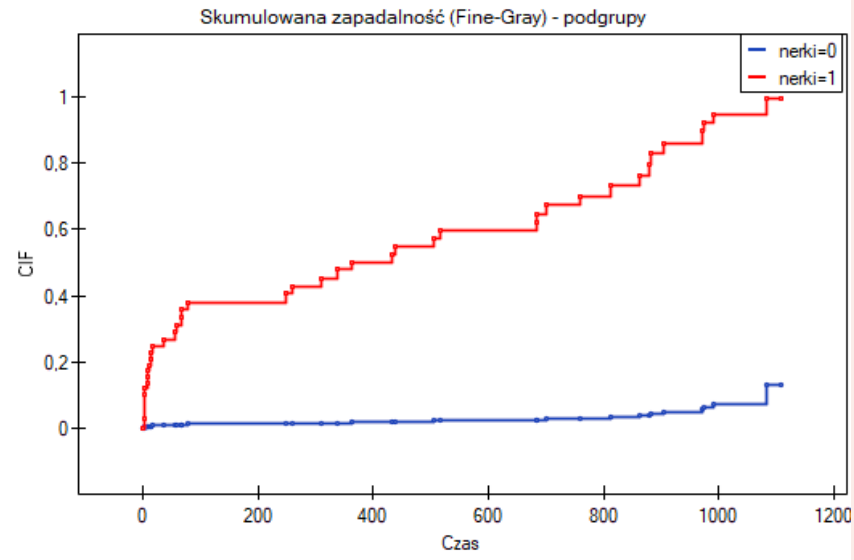
Odwrotną sytuację obserwujemy dla niewydolności nerek (nerki) – tu oba modele zgodnie wskazują na silny, istotny statystycznie związek z ryzykiem zgonu, przy czym w modelu Fine'a-Graya efekt jest jeszcze wyraźniejszy: $sHR = 38.65$ (95%CI: $\langle 19.47; 76.72 \rangle$), $p < 0.000001$, wobec $HR = 11.39$ (95%CI: $\langle 7.08; 18.32 \rangle$), $p < 0.000001$ w modelu standardowym. Niewydolność nerek jest więc nie tylko istotnym czynnikiem ryzyka zgonu z jakiejkolwiek przyczyny, ale przede wszystkim silnie – i jeszcze silniej, niż sugerowałby model standardowy – związana z ryzykiem zgonu z powodu progresji samego szpiczaka, co jest zgodne z jej dobrze znaną rolą niekorzystnego czynnika rokowniczego w tej chorobie (uszkodzenie nerek towarzyszy zwykle zaawansowanej postaci szpiczaka).

$sHR_{nerki} = 38.65$, 95%CI: $\langle 19.47; 76.72 \rangle$, $p < 0.000001$ (model Fine'a-Graya) wobec $HR_{nerki} = 11.39$, $p < 0.000001$ (model standardowy)

Niewydolność nerek jest silnie i konsekwentnie związana z ryzykiem zgonu z powodu szpiczaka – efekt jest jeszcze wyraźniejszy po poprawnym uwzględnieniu konkurencyjnego ryzyka zgonu z innej przyczyny.

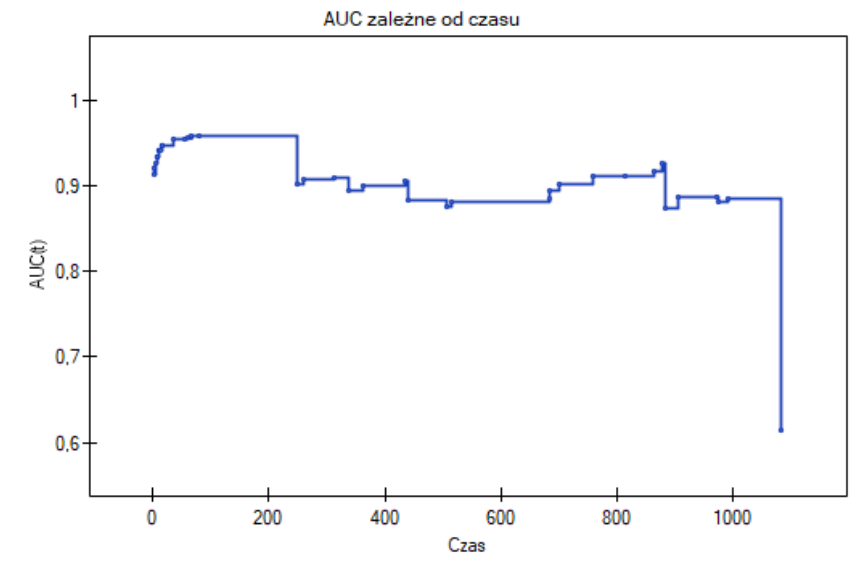
Na zakończenie porównamy krzywe CIF zgonu z powodu szpiczaka w grupach wyróżnionych przez zmienną nerki:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Krzywa CIF dla pacjentów z niewydolnością nerek (nerki=1) rośnie znacznie szybciej i osiąga znacznie wyższe wartości niż krzywa dla pacjentów bez tego powikłania (nerki=0) – już po około 100 dniach obserwacji skumulowane prawdopodobieństwo zgonu z powodu szpiczaka przekracza w tej grupie 0.4, a pod koniec obserwacji zbliża się do 1. U pacjentów bez niewydolności nerek skumulowane prawdopodobieństwo zgonu z powodu szpiczaka pozostaje przez cały okres obserwacji niskie (poniżej 0.15). Wykres potwierdza więc liczbowy wynik sHR: niewydolność nerek wiąże się w tych danych z istotnie wyższym ryzykiem zgonu z powodu szpiczaka mnogiego.

Na koniec ocenimy zdolność zbudowanego modelu Fine'a-Graya do rozróżniania pacjentów, którzy umrą z powodu szpiczaka, wybierając w oknie Dołącz wykres opcję Krzywa ROC, czyli wykres AUC zależnego od czasu:



Wartości $AUC(t)$ są przez cały okres obserwacji wysokie. W pierwszych tygodniach $AUC(t)$ rośnie od około 0.91 do około 0.96 i utrzymuje się na tym poziomie przez pierwsze ok. 250 dni, a następnie – w miarę jak wśród kontroli ubywa pacjentów bez zdarzenia i przybywa zgonów z innych przyczyn – obniża

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

się do poziomu 0.88–0.92, na którym pozostaje niemal do końca obserwacji. Oznacza to, że model (a w praktyce przede wszystkim zmienna nerki) bardzo dobrze odróżnia pacjentów, którzy umrą z powodu szpiczaka do danego czasu t, od tych, którzy do tego czasu z tego powodu nie umrą – i to niezależnie od horyzontu prognozy. Gwałtowny spadek $AUC(t)$ w ostatnim punkcie (ok. 1080 dni) jest artefaktem estymacji: w tak późnym czasie w zbiorze ryzyka pozostaje już tylko kilka obserwacji, a estymator oparty na garstce par przypadek–kontrola staje się niestabilny; wartości z samego końca osi czasu nie należy więc interpretować.

$AUC(t) \approx 0.88$–0.96 dla t od kilku do ok. 1000 dni

Model Fine'a-Graya bardzo dobrze i stabilnie w czasie rozróżnia pacjentów zagrożonych zgonem z powodu szpiczaka – jakość prognozy nie zależy istotnie od przyjętego horyzontu czasu.

DLA ZAINTERESOWANYCH

Skumulowana funkcja zapadalności (CIF)

$$CIF(t) = P(T \leq t,\, \varepsilon = 1) = \sum_{t_i \leq t} S(t_{i-1}) \cdot h_1(t_i),$$

gdzie $S(t_{i-1})$ to ogólna (uwzględniająca wszystkie przyczyny łącznie) funkcja przeżycia tuż przed t_i, a $h_1(t_i)$ to hazard specyficzny dla zdarzenia badanego w t_i.

Model subdystrybucyjnego hazardu Fine'a-Graya

$$h^*(t, X_1, ..., X_k) = h_0^*(t) \cdot \exp\left(\sum_{i=1}^{k} \beta_i X_i\right)$$

Wagi IPCW wykorzystywane do utrzymania w zbiorze ryzyka obiektów po wystąpieniu zdarzenia konkurencyjnego wyliczane są w oparciu o odwrotność estymatora Kaplana-Meiera rozkładu cenzurowania $\hat{G}(t)$:

$$w_i(t) = \frac{\hat{G}(t)}{\hat{G}(\min(T_i, t))},$$

gdzie estymator $\hat{G}$ liczony jest tak jak estymator Kaplana-Meiera funkcji przeżycia, ale z zamienionymi rolami – za „zdarzenie" uznawane jest tu prawdziwe cenzurowanie, natomiast zarówno zdarzenie badane, jak i zdarzenia konkurencyjne traktowane są jako cenzurowanie tego pomocniczego estymatora.

Subdystrybucyjny Iloraz Hazardów (sHR)

$$sHR_i = e^{\beta_i}.$$

AUC zależne od czasu dla ryzyk konkurencyjnych

Formuła jest taka sama jak dla standardowego modelu Coxa (patrz Regresja proporcjonalnego hazardu Coxa), zmienia się jedynie zbiór kontroli i ich wagi: obok osób z $T_j > t$ (waga $1/\hat{G}(t)$) kontrolami są także osoby, u których do czasu t wystąpiło zdarzenie konkurencyjne ($T_j \leq t$, $\varepsilon_j = 2$), z wagą $1/\hat{G}(T_j^-)$. Estymator $\hat{G}$ jest tym samym estymatorem Kaplana-Meiera rozkładu cenzurowania, który służy do wyznaczania wag IPCW modelu (Blanche i in. 2013[21]).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

29 META-ANALIZA

Liczba pojawiających się prac naukowych w ostatniej dekadzie bardzo wzrosła. Ułatwia to dostęp do wiedzy, ale utrudnia nadążenie za wciąż pojawiającymi się nowymi informacjami. Gdyby na przykład lekarz zastosował nowy sposób leczenia swoich pacjentów na podstawie jednej przeczytanej pracy naukowej, mógłby popełnić błąd – inne, równie wiarygodne prace mogą przeczyć skuteczności tego leczenia. Aby decyzja lekarza była jak najlepiej uzasadniona, powinien on znać większość prac naukowych, jakie ukazały się na dany temat, a ciągłe przeglądanie coraz większych zbiorów literatury zajęłoby mu tyle czasu, że zabrakłoby go na leczenie pacjentów. **Meta-analiza** pozwala szybko dokonać takiego przeglądu, ponieważ jest statystycznym podsumowaniem wyników opisanych w wielu niezależnych pracach na ten sam temat.

Meta-analiza w programie PQStat wykonywana jest na podstawie następujących miar (**efektów końcowych**):

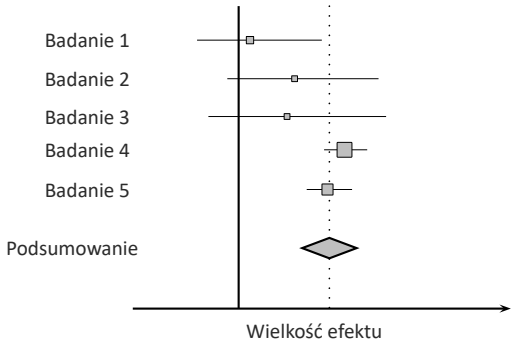
- **Różnica średnich**,
- **d Cohen**,
- **g Hedges**,
- **Iloraz średnich**,
- **Iloraz szans (OR)**,
- **Relatywne ryzyko (RR)**,
- **Różnica ryzyka (RD)**,
- **Współczynnik Pearsona**,
- **AUC dla krzywej ROC**,
- **Proporcja**.

Wprowadzenie do meta-analizy i przygotowanie danych

KIEDY STOSOWAĆ:

Meta-analizę stosuje się, gdy chcemy w sposób statystyczny podsumować wyniki wielu niezależnych prac naukowych dotyczących tego samego zagadnienia. Najbardziej znanym obrazem związanym z meta-analizą jest **wykres leśny**, przedstawiający wyniki poszczególnych badań wraz z ich wspólnym podsumowaniem.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Aby można było dokonać wspólnego podsumowania wybranej literatury, musi być ona spójna w opisie, a podane w niej miary muszą być tożsame.

WARUNKI STOSOWANIA

Aby dana praca naukowa mogła być wykorzystana w meta-analizie, powinna umożliwiać wyznaczenie:

- **Efektu końcowego** – czyli miary statystycznej wskazującej na uzyskany w pracy wynik. W praktyce mogą to być różnego rodzaju wielkości, np. różnica średnich, iloraz szans, relatywne ryzyko itp.
- **Błędu efektu**, czyli **SE** *(ang. Standard Error)* – określającego precyzję przeprowadzonego badania. Im mniejszy jest błąd (SE), tym większa jest precyzja danego badania, a co za tym idzie – tym większa jest jego **waga** i wpływ na wynik meta-analizy.
- **Liczności**, czyli liczby obiektów, na jakiej przeprowadzono badanie – wymaganej w niektórych sytuacjach.

Uwaga! Często zdarza się, że praca naukowa nie zawiera wszystkich wymienionych wyżej elementów wprost. Wówczas należy poszukać w niej danych, na podstawie których wyliczenie tych miar będzie możliwe.

Uwaga! Program PQStat obliczenia związane z meta-analizą wykonuje na danych zawierających: Efekt końcowy, Błąd efektu i w niektórych sytuacjach Liczność. Zaleca się, aby przed przystąpieniem do meta-analizy wprowadzić dane dotyczące każdej publikacji w oknie przygotowania danych – jest to szczególnie przydatne, gdy w danej pracy nie podano wprost tych trzech miar.

DEFINICJE MIAR

Wartość p (istotność statystyczna) nie jest bezpośrednio wykorzystywana w meta-analizie. Ta sama wielkość efektu może być istotna statystycznie w badaniu dużym, a nieistotna w badaniu opartym na małej liczności – i odwrotnie: mała wielkość efektu może być istotna statystycznie w badaniu dużym, a duża wielkość efektu może być nieistotna w badaniu małym. Wynika to z mocy testów statystycznych. Badając istotność statystyczną, sprawdzamy jedynie, czy efekt w ogóle istnieje (czy jest różny od zera), a nie czy jest na tyle duży, by miał przełożenie na pożądane skutki – np. fakt, że lek obniża ciśnienie istotnie statystycznie o 1 mmHg, nie spowoduje, że będzie stosowany, gdyż z klinicznego punktu widzenia jest to zbyt mała wielkość. Meta-analiza skupia się przede wszystkim na wielkości poszczególnych efektów,

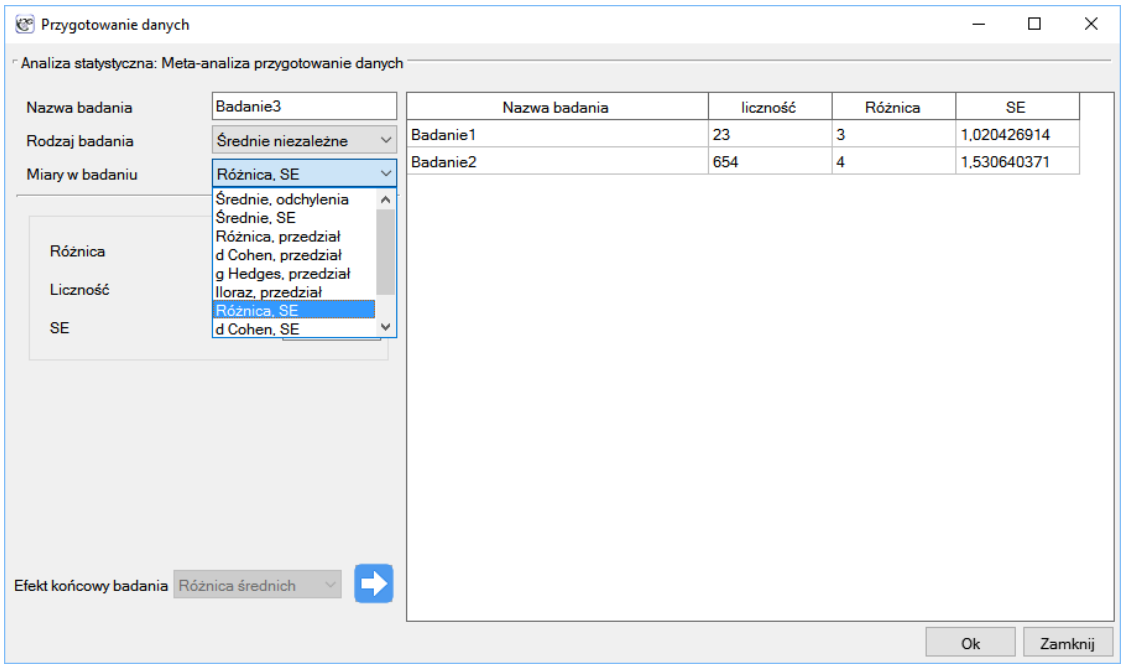
Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

a nie na ich istotności statystycznej, dlatego nie ma większego znaczenia, czy prace wykorzystywane w meta-analizie same w sobie wskazują na istotność statystyczną danego efektu, czy nie.

W programie PQStat dla każdego badania wyliczana jest jednak istotność statystyczna, wyznaczona na podstawie stosunku efektu do jego błędu. Jest to podejście asymptotyczne, oparte na rozkładzie normalnym, dedykowane dla licznych grup. Jeśli w cytowanym badaniu do sprawdzenia istotności statystycznej użyto innego testu, uzyskane wyniki mogą się nieco różnić.

JAK USTAWIĆ ANALIZĘ

Okno przygotowania danych do meta-analizy wywołujemy poprzez menu:
Statystyki zaawansowane→Meta-analiza→Przygotowanie danych.



W oknie przygotowania danych do meta-analizy badacz najpierw podaje nazwę wprowadzanego badania. Nazwa ta powinna jednoznacznie identyfikować badanie, ponieważ będzie je opisywać we wszystkich wynikach meta-analizy, w tym również na wykresach. Pożądany Efekt końcowy badania, jego błąd oraz liczność zostają wyliczone na podstawie miar pozyskanych z odpowiedniej pracy naukowej. Miary zawarte w badaniach, na podstawie których mogą być wyliczone efekty końcowe, przedstawia poniższa tabela:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Rodzaj badania	Miary w badaniu	Efekt końcowy
Średnie niezależne	Średnie, odchylenia	**a**, **b**, **c**, **e**
	Średnie, SE	**a**, **b**, **c**, **e**
	Różnica, przedział	**a**
	d Cohen, przedział	**b**
	g Hadges, przedział	**c**
	Iloraz, przedział	**e**
	Różnica, SE	**a**
	d Cohen, SE	**b**
	g Hadges, SE	**c**
	Iloraz, SE	e
Średnie zależne	Średnie, odchylenia	**a**, **b**, **c**
	Średnie, SE	**a**, **b**, **c**
	Różnica, przedział	**a**
	d Cohen, przedział	**b**
	g Hadges, przedział	**c**
	Różnica, odchylenie	**a**, **b**, **c**
	Różnica, SE	**a**, **b**, **c**
	d Cohen, SE	**b**
	g Hadges, SE	**c**
Średnia vs zadana	Średnia, odchylenie	**a**
	Średnia, SE	**a**
Średnia - podsumowanie	Średnia, odchylenie	**d**
	Średnia, SE	**d**
Tabele	Tabela 2x2	**b**, **c**, **f**, **g**, **h**
	OR, przedział	**b**, **c**, **f**
	RR, przedział	**g**
	RD, przedział	**h**
	d Cohen, przedział	**b**, **c**, **f**
	g Hadges, przedział	**b**, **c**, **f**
	OR, SE	**b**, **c**, **f**
	RR, SE	**g**
	RD, SE	**h**
	d Cohen, SE	**b**, **c**, **f**
	g Hadges, SE	**b**, **c**, **f**
Korelacja	Współczynnik	**i**
	Współczynnik, przedział	**i**
	Współczynnik, SE	**i**
Krzywa ROC	AUC, SE	**j**
	AUC, przedział	**j**
Jedna proporcja	Wyróżniona liczność	**k**
	Wyróżniona proporcja	**k**

gdzie poszczególne efekty końcowe, to:

a - Różnica średnich

b - d Cohen

c - g Hadges

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

d - Średnia

e - Iloraz średnich

f - Iloraz szans (OR)

g - Relatywne ryzyko (RR)

h - Różnica ryzyka (RD)

i - Współczynnik Pearsona

j - AUC (krzywa ROC)

k - Proporcja

Uwaga! Przy wyznaczaniu błędu współczynników takich jak OR lub RR i innych opartych na tabelach, gdy występują zerowe wartości w tabelach, lub przy wyznaczaniu błędu proporcji, gdy proporcja wynosi 0 lub 1, stosowana jest korekta na ciągłość wykorzystująca współczynnik zwiększenia równy 0,5. Przedział ufności dla proporcji wyznaczany jest zgodnie z dokładną metodą Cloppera-Pearsona[39].

WYKONANIE I INTERPRETACJA

Przykład 29.1. Jesteśmy zainteresowani wpływem palenia papierosów na ryzyko wystąpienia choroby X. Chcemy przeprowadzić meta-analizę, dla której efektem końcowym będzie relatywne ryzyko (RR). Przy takim założeniu prace wytypowane do meta-analizy muszą umożliwiać wyliczenie **RR** i jego **błędu**.

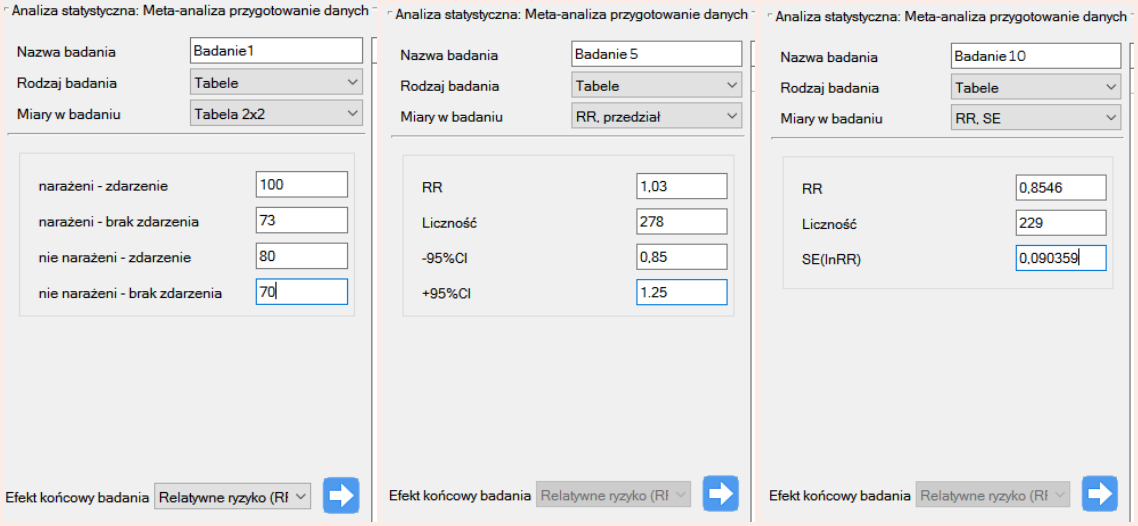
Krok 1. Na podstawie opisu efektu końcowego (patrz powyższa tabela) ustalono, że relatywne ryzyko (opisane jako wynik **g**) jest możliwe do wyznaczenia w programie PQStat w trzech sytuacjach, tzn. jeżeli w danej pracy naukowej podano: RR i przedział ufności dla niego, lub RR wraz z błędem, lub odpowiednie liczności w czterech kategoriach tzn. tabelę 2x2.

Krok 2. Wytypowano do meta-analizy 10 prac, które spełniały kryteria włączenia do badań oraz miały możliwość wyznaczenia relatywnego ryzyka (patrz krok 1). Potrzebne dane zawarte w wytypowanych pracach to:

Badanie 1: liczności: (palący i chorujący)=100, (palący i niechorujący)=73, (niepalący i chorujący)=80, (niepalący i niechorujący)=70,
Badanie 2: liczności: (palący i chorujący)=182, (palący i niechorujący)=172, (niepalący i chorujący)=180, (niepalący i niechorujący)=172,
Badanie 3: liczności: (palący i chorujący)=157, (palący i niechorujący)=132, (niepalący i chorujący)=125, (niepalący i niechorujący)=201,
Badanie 4: liczności: (palący i chorujący)=19, (palący i niechorujący)=15, (niepalący i chorujący)=35, (niepalący i niechorujący)=20,
Badanie 5: liczność: 278, RR[95%CI]=1.03[0.85-1.25],
Badanie 6: liczność: 560, RR[95%CI]=1.21[1.05-1.40],
Badanie 7: liczność: 1207, RR[95%CI]=1.04[0.93-1.15],
Badanie 8: liczność: 214, RR[95%CI]=1.15[0.95-1.40],
Badanie 9: liczność: 285, RR[95%CI]=1.36[1.03-1.79],
Badanie 10: liczność: 1968, RR=1.17, SE(lnRR)=0.0437,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Krok 3. Korzystając z okna przygotowania badań do meta-analizy, wprowadzono dane do arkusza. Pierwsze cztery badania wprowadzamy, wybierając tabele; badania od piątego do dziewiątego wprowadzamy, wybierając RR i przedział; ostatnie badanie podaje wszystkie niezbędne dane, tzn. RR i SE. Jako Efekt końcowy badania ustawiamy Relatywne ryzyko (RR):



Każde wprowadzone badanie przenosimy do okna po prawej stronie. Przyciskiem OK przenosimy przygotowane badania do arkusza danych. Na podstawie zawartych w arkuszu informacji o poszczególnych badaniach można przystąpić do wykonywania meta-analizy.

Efekt podsumowujący – efekt stały i efekt zmienny

KIEDY STOSOWAĆ:

Najbardziej pożądanym elementem meta-analizy jest podsumowanie zebranych badań, czyli wyznaczenie wspólnego efektu M. Takie podsumowanie można wykonać na dwa sposoby – wyznaczając efekt stały lub efekt zmienny.

DEFINICJE MIAR

Efekt stały – wyliczając efekt stały, zakładamy, że wszystkie badania w meta-analizie dzielą jeden wspólny prawdziwy efekt. Gdyby więc każde badanie dotyczyło tej samej populacji (np. tego samego kraju), to podsumowując meta-analizę efektem stałym, zakładamy, że w każdym z tych badań prawdziwy (populacyjny) efekt jest taki sam, a co za tym idzie – także wszystkie czynniki, które mogłyby zaburzać wielkość tego efektu, są takie same (np. wiek czy płeć badanych są w każdym badaniu podobne). Różnice w uzyskanych efektach dla poszczególnych badań wynikają wtedy tylko z błędu próbkowania (błędu wewnętrznego każdego badania), czyli z wielkości SE.

Efekt stały oszacowuje populacyjny efekt – prawdziwy efekt wspólny dla każdego z badań. Przedział ufności wokół efektu stałego (szerokość rombu na wykresie leśnym) zależy tylko od poszczególnych SE.

Efekt zmienny – wyliczając efekt zmienny, zakładamy, że każde badanie reprezentuje nieco inną populację, przez co prawdziwy (populacyjny) efekt jest inny dla każdej z nich. Gdyby więc każde badanie dotyczyło innego kraju, to podsumowując meta-analizę efektem zmiennym, zakładamy, że niektóre czynniki mogące zaburzać wielkość tego efektu (np. poziom wykształcenia czy zamożność) różnią się między kra-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

jami, przez co i prawdziwy efekt (np. średni wzrost dzietności) jest nieco inny w każdym z nich. Różnice w uzyskanych efektach dla poszczególnych badań wynikają wtedy z błędu próbkowania (błędu wewnątrz każdego badania), czyli z wielkości SE, oraz z różnic między badanymi populacjami, czyli z heterogeniczności badań (T^2). Heterogeniczność ta nie może być zbyt duża – zbyt duże zróżnicowanie badanych populacji wskazuje na brak podstaw do wspólnego podsumowania.

Efekt zmienny oszacowuje średnią ważoną z prawdziwych (populacyjnych) efektów poszczególnych badań. Przedział ufności wokół efektu zmiennego (szerokość rombu na wykresie leśnym) zależy od poszczególnych SE oraz od T^2.

Przedział ufności a przedział predykcji

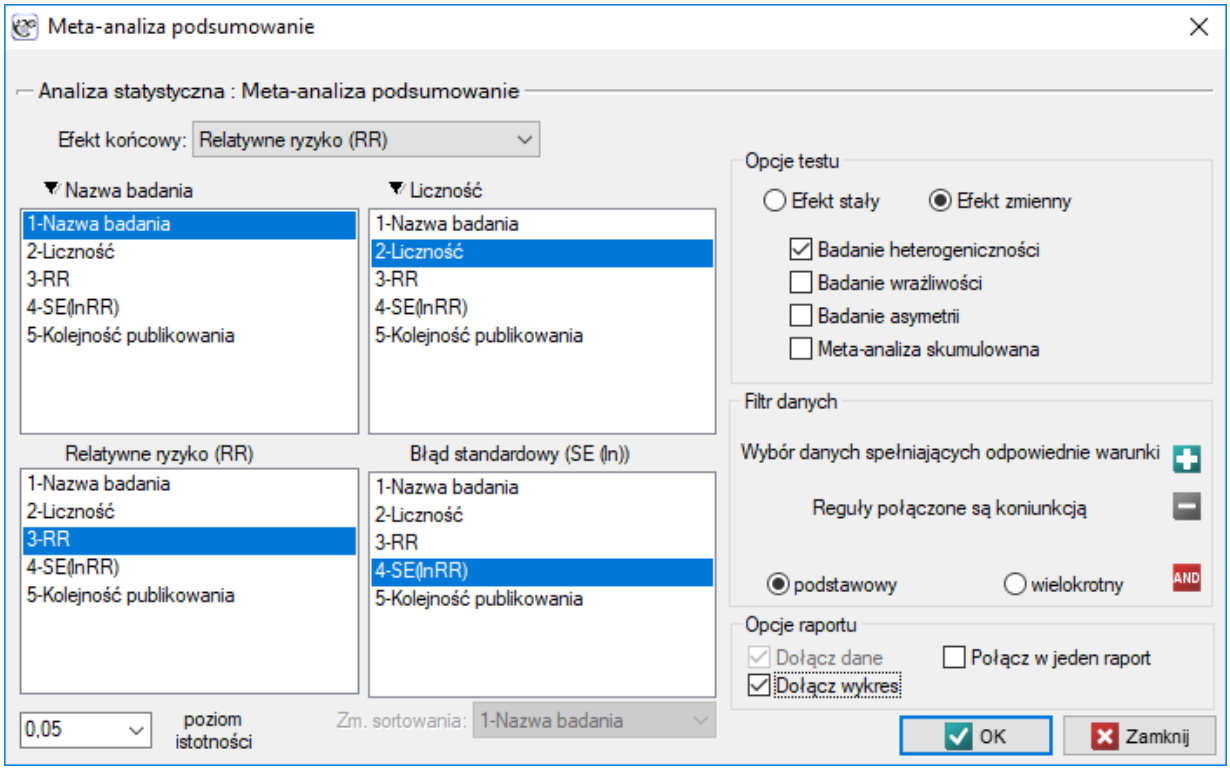
> **95% przedział ufności** (szerokość rombu na wykresie leśnym) – oznacza, że w 95% przypadków takich meta-analiz podsumowujący efekt zmienny wpadnie do wyznaczonego przez romb przedziału.
>
> **95% przedział predykcji** – oznacza, że w 95% przypadków prawdziwy (populacyjny) efekt nowego badania wpadnie do wyznaczonego przedziału.

Waga badania i jego udział – waga w_i badania zależy od obserwowanej zmienności: dla efektu stałego wynika ona tylko z błędu próbkowania (SE) danego badania, a dla efektu zmiennego – dodatkowo z różnic pomiędzy badaniami, czyli z obserwowanej wariancji T^2 (im mniejszy błąd, tym większa waga i większy wpływ badania na wynik meta-analizy). Na podstawie wag przypisanych poszczególnym badaniom wyznacza się **udział** danego badania w całej analizie, czyli odsetek, jaki stanowi waga danego badania w stosunku do łącznej wagi wszystkich włączonych badań.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji meta-analizy wywołujemy poprzez menu:
Statystyki zaawansowane→Meta-analiza→Podsumowanie.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W oknie tym, w zależności od wybranego efektu końcowego, można dokonać podsumowania meta-analizy oraz przeprowadzić podstawowe analizy pozwalające sprawdzić jej założenia, takie jak heterogeniczność czy błąd publikacyjny (badanie wrażliwości, badanie asymetrii), a także przeprowadzić meta-analizę skumulowaną.

WYKONANIE I INTERPRETACJA

PRZYKŁAD 29.2. (plik MetaanalizaRR.pqs)
Badano ryzyko choroby X dla osób palących i niepalących. W niektórych pracach naukowych wskazywano, że ryzyko wystąpienia choroby X jest większe dla palących, a w niektórych nie udowodniono takiego związku. By ustalić, czy palenie papierosów ma wpływ na występowanie choroby X, zaplanowano przeprowadzenie meta-analizy. Wykonano dokładny przegląd literatury i na tej podstawie wytypowano 10 prac naukowych, które dysponowały danymi pozwalającymi wyliczyć relatywne ryzyko (tzn. ryzyko zachorowania dla palących w stosunku do ryzyka zachorowania dla niepalących) oraz jego błąd (precyzję danego badania). Dane przygotowano do meta-analizy i zapisano w pliku.

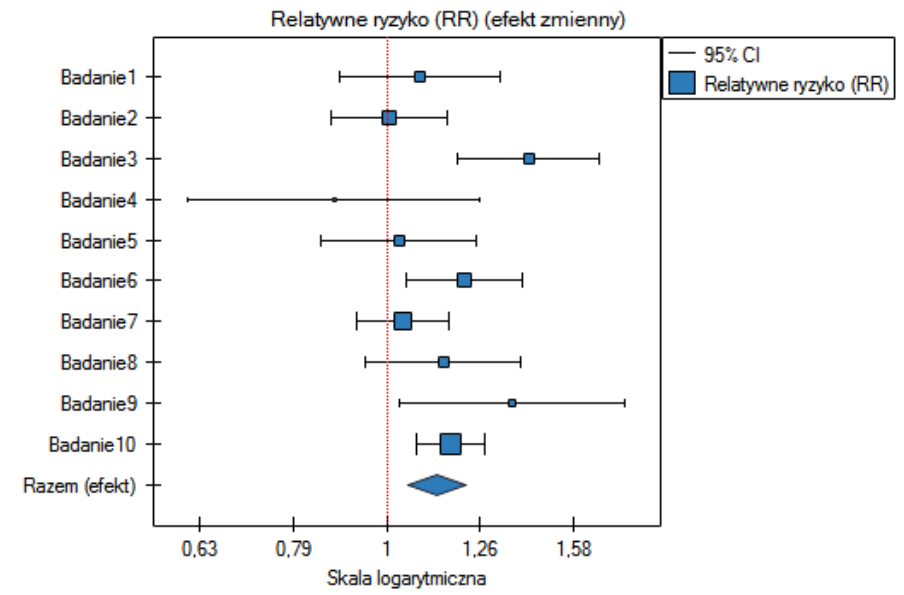
Ze względu na to, że prace włączone do meta-analizy pochodziły z różnych ośrodków i obejmowały nieco inne populacje, podsumowania dokonano, wybierając efekt zmienny. Jako efekt końcowy wybrano relatywne ryzyko oraz przedstawiono wyniki na wykresie leśnym.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Meta-analiza podsumowanie	
Czas analizy	0,08 sek.
Analizowane zmienne	Nazwa badania
	Liczność
	RR
	SE(lnRR)
Poziom istotności	0,05
Efekt	Zmienny
Efekt końcowy	Relatywne ryzyko (RR)
Liczba prac	10
Przedział predykcji efektu zmiennego	
-95%	0,93066352
+95%	1,375432703

Dane :

Nazwa	Liczność	Relatywn	SE (ln)	-95%CI	+95%CI	Statystyk	Wartość	Wariancja	Wagi	Udział
Badanie1	323	1,083815	0,100264	0,890450	1,319169	0,802748	0,422120	0,010052	62,93093	8,41%
Badanie2	706	1,005398	0,073377	0,870721	1,160907	0,073375	0,941507	0,005384	89,11270	11,91%
Badanie3	615	1,416802	0,088553	1,191056	1,685335	3,934372	0,000083	0,007841	73,10376	9,77%
Badanie4	89	0,878151	0,183328	0,613083	1,257821	-0,70876	0,478472	0,033609	25,35055	3,39%
Badanie5	278	1,03	0,098000	0,849999	1,248117	0,301618	0,762942	0,009604	64,76022	8,65%
Badanie6	560	1,21	0,072363	1,050000	1,394380	2,634199	0,008433	0,005236	90,30199	12,07%
Badanie7	1207	1,04	0,057037	0,929999	1,163010	0,687630	0,491685	0,003253	110,0022	14,7%
Badanie8	214	1,15	0,097478	0,95	1,392105	1,433765	0,151639	0,009502	65,19076	8,71%
Badanie9	285	1,36	0,141801	1,03	1,795728	2,168415	0,030127	0,020107	38,54288	5,15%
Badanie1(	1968	1,17	0,0437	1,073960	1,274627	3,592763	0,000327	0,001909	129,0798	17,25%
Razem	6245	1,131399	0,036554	1,053175	1,215433	3,377304	0,000732			



Wyniki czterech badań (badania 3, 6, 9 i 10) wskazują na istotnie wyższe ryzyko zachorowania dla palących. Wynik podsumowujący przeprowadzoną meta-analizę również jest istotny statystycznie i potwierdza ten sam efekt.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

RR[95%CI]=1.13[1.05-1.22]

Palenie papierosów zwiększa ryzyko wystąpienia choroby X (efekt podsumowujący jest istotny statystycznie).

Niestety przedział predykcji efektu zmiennego jest szerszy: [0.93-1.38], co oznacza, że w 95% przypadków prawdziwe, populacyjne relatywne ryzyko uzyskane w kolejnych badaniach może być zarówno większe, jak i mniejsze niż jeden.

Uwaga! Przed interpretacją wyników należy sprawdzić spełnienie założeń meta-analizy. W tym przypadku należy się zastanowić nad wykluczeniem badania trzeciego (patrz wyniki analizy wrażliwości, analizy asymetrii, meta-analizy skumulowanej oraz założenie heterogeniczności).

DLA ZAINTERESOWANYCH

Waga badania

Wzory:

Dla efektu stałego zmienność wynika tylko z błędu próbkowania (błędu wewnątrz każdego badania), czyli z wielkości SE:

$$w_i = \frac{1}{SE^2}$$

Dla efektu zmiennego zmienność wynika z błędu próbkowania (błędu wewnątrz każdego badania), czyli z wielkości SE, oraz z różnic pomiędzy badaniami, czyli z obserwowanej wariancji T^2:

$$w_i = \frac{1}{SE^2 + T^2}$$

Badanie heterogeniczności

KIEDY STOSOWAĆ:

Trudno oczekiwać, by każde badanie kończyło się uzyskaniem dokładnie tej samej wielkości efektu – w naturalny sposób uzyskane w różnych pracach wyniki będą nieco inne. Badanie heterogeniczności ma ustalić, na ile pojawiające się różnice pomiędzy uzyskanymi w różnych pracach efektami mają wpływ na budowany w meta-analizie efekt sumaryczny. **Efekt sumaryczny dobrze podsumowuje wyniki uzyskane w poszczególnych pracach, jeśli różnice między poszczególnymi efektami są naturalne, tzn. niewielkie.** Duże różnice w obserwowanych efektach mogą świadczyć o niejednorodności badań i konieczności wydzielenia bardziej homogenicznych podgrup, np. podzielenia zebranych prac na kilka podgrup względem dodatkowego czynnika. Dla przykładu: dany lek inaczej działa na osoby młodsze, a inaczej na starsze, więc w pracach opartych głównie na danych od osób młodych uzyskiwany efekt może znacznie odbiegać od prac przeprowadzanych na osobach starszych. Podzielenie zebranych prac na bardziej jednorodne podgrupy pozwala na lepsze oszacowanie efektu sumarycznego dla każdej z tych podgrup oddzielnie.

Badanie heterogeniczności sprawdza, czy zmienność pomiędzy badaniami jest zerowa.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

DEFINICJE MIAR

Dodatkowymi miarami opisującymi heterogeniczność są współczynniki I^2 i H^2. **Współczynnik** I^2 określa procent obserwowanej wariancji, jaki wynika z rzeczywistej różnicy w wielkości badanych efektów (graficznie odzwierciedla stopień zazębiania się przedziałów ufności poszczególnych badań). Ze względu na to, że mieści się on pomiędzy 0% a 100%, podlega prostej interpretacji i jest chętnie stosowany: jeśli $I^2 = 0$, cała obserwowana wariancja wielkości efektów jest „fałszywa", więc jeśli w wyznaczonym wokół współczynnika I^2 przedziale ufności znajdzie się wartość 0, uzyskaną wariancję można uznać za nieistotną statystycznie. Im wartość I^2 jest bliższa 100%, tym bardziej należy się zastanowić nad rezygnacją ze wspólnego podsumowania badań. Przyjmuje się, że $I^2 \approx 25\%$ oznacza słabą, $I^2 \approx 50\%$ średnią, a $I^2 \approx 75\%$ silną heterogeniczność badań. **Współczynnik** H^2 rozpatruje się natomiast w odniesieniu do wartości 1 – jeśli przedział ufności dla H^2 zawiera wartość 1, uzyskaną wariancję można uznać za nieistotną statystycznie, a im wyższa wartość H^2, tym większa heterogeniczność badań.

Uwaga! Jeśli badanie heterogeniczności daje wynik istotny statystycznie, jest to mocna sugestia, by zrezygnować ze wspólnego podsumowywania wszystkich zebranych badań. Jeśli wynik jest nieistotny statystycznie, badania można podsumować wspólnym efektem – przy czym sugeruje się, by podsumowania dokonywać zwykle efektem zmiennym. Podsumowanie efektem zmiennym bierze bowiem pod uwagę zmienność pomiędzy badaniami (T^2), a podsumowanie efektem stałym jej nie uwzględnia. Jeśli jednak T^2 jest małe, wynik modelu z efektem stałym będzie bliski wynikom modelu z efektem zmiennym, a gdy $T^2 = 0$, oba modele dadzą dokładnie ten sam wynik.

JAK USTAWIĆ ANALIZĘ

Opcja Badanie heterogeniczności jest dostępna w oknie Podsumowanie meta-analizy (patrz powyżej).

WYKONANIE I INTERPRETACJA

Przykład (29.2) c.d. *(plik MetaanalizaRR.pqs)*
Badając wpływ palenia papierosów na wystąpienie choroby X, sprawdzono założenie dotyczące heterogeniczności badań. W tym celu w oknie analizy wybrano opcję Badanie heterogeniczności.

Badanie heterogeniczności	
Statystyka Q	17,397991523
Stopnie swobody	9
Wartość p	0,042836225

Badanie heterogeniczności			
	Wartości	-95%CI	+95%CI
T2	0,005837	0	0,018776
H2	1,933110	0,933891	4,001445
I2	48,27%	0%	75,01%

$p = 0.04284$ **dla statystyki Q (istotne statystycznie,** $\alpha = 0.05$**)**

Wariancja obserwowanych efektów nie jest zerowa ($T^2 = 0.0058$), a współczynnik $I^2 = 48.27\%$ wskazuje na przeciętną heterogeniczność badań.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jedynie przedział ufności dla współczynnika H^2 uznaje zmienność pomiędzy badaniami za nieistotną (przedział dla tego współczynnika to [0.93-4.00]). Mając na uwadze powyższe wyniki, należy się zastanowić, czy zebrane prace mogą być podsumowane jednym wspólnym efektem końcowym (wspólnym relatywnym ryzykiem), czy też warto wyznaczyć bardziej homogeniczną grupę prac i przeprowadzić analizę ponownie.

DLA ZAINTERESOWANYCH

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \tau^2 = 0, \\ \mathcal{H}_1: &\quad \tau^2 \neq 0, \end{aligned}$$

gdzie:
τ^2 – wariancja prawdziwych (populacyjnych) efektów poszczególnych badań.

Wzory:
Statystyka testowa ma postać:

$$Q = T^2 W + k - 1$$

gdzie:
T^2 – wariancja obserwowanych efektów,
W – współczynnik wyliczony na podstawie wag przypisanych poszczególnym badaniom,
k – liczba badań.

Statystyka ta ma asymptotycznie (dla dużych liczności) rozkład χ^2 z liczbą stopni swobody wyliczaną według wzoru $df = k - 1$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Wzory (współczynniki I^2 i H^2):
Współczynniki I^2 i H^2 wyraża się wzorami:

$$I^2 = \frac{H^2 - 1}{H^2}, \qquad H^2 = \frac{Q}{k-1}.$$

Badanie wrażliwości

KIEDY STOSOWAĆ:

Efekt podsumowujący badania może się zmieniać w zależności od tego, które badania włączymy do analizy, a które z niej wyłączymy. Powinnością badacza jest sprawdzenie, jak wrażliwa jest analiza na zmianę kryteriów doboru badań. Sprawdzenie wrażliwości pozwala ustalić **zmiany efektu podsumowującego** będące skutkiem usunięcia danego badania. Badania powinny być na tyle zbliżone, by usunięcie jednego z nich nie zmieniało całkowicie interpretacji efektu podsumowującego.

DEFINICJE MIAR

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Przypisana do danego badania wartość **udziału pozostałych** definiuje odsetek, jaki stanowi łączna waga pozostałych w analizie badań po wykluczeniu danego badania. **Zmiana precyzji** informuje natomiast, jak zmieni się precyzja efektu podsumowującego (szerokość przedziału ufności), gdy dane badanie zostanie wykluczone z analizy. Dobrym zilustrowaniem analizy wrażliwości jest wykres leśny wielkości efektu oraz wykres zmian precyzji przy wyłączaniu poszczególnych badań.

JAK USTAWIĆ ANALIZĘ

Opcja Badanie wrażliwości jest dostępna w oknie Podsumowanie meta-analizy (patrz powyżej).

WYKONANIE I INTERPRETACJA

Przykład (29.2) c.d. *(plik MetaanalizaRR.pqs)*
Badając wpływ palenia papierosów na wystąpienie choroby X, sprawdzono wrażliwość analizy na wyłączanie poszczególnych badań. W tym celu w oknie analizy wybrano opcję Badanie wrażliwości oraz wskazano wykres leśny (wrażliwości) i wykres słupkowy (wrażliwości).

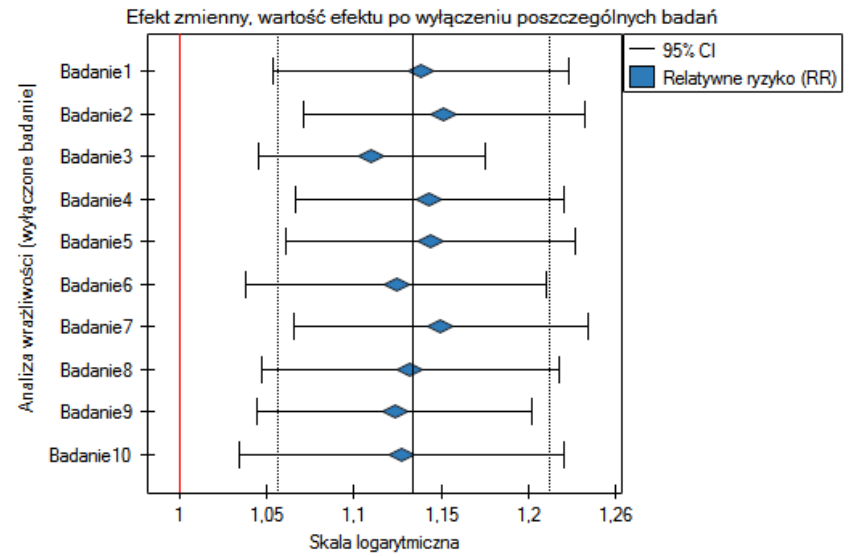
Wrażliwość									
Usunięte badania	Liczność	Relatywn	SE (ln)	-95%CI	+95%CI	Statystyk	Wartość p	Udział po:	Zmiana p
Badanie1	5922	1,136027	0,039985	1,050397	1,228638	3,189633	0,001424	91,59%	9,85%
Badanie2	5539	1,149589	0,037893	1,067302	1,238220	3,678815	0,000234	88,09%	5,34%
Badanie3	5630	1,106810	0,030595	1,042389	1,175212	3,316849	0,000910	90,23%	-18,14%
Badanie4	6156	1,141116	0,036296	1,062758	1,225251	3,636923	0,000275	96,61%	0,15%
Badanie5	5967	1,141652	0,039118	1,057392	1,232627	3,386533	0,000707	91,35%	8%
Badanie6	5685	1,121108	0,040591	1,035371	1,213945	2,816305	0,004857	87,93%	10,06%
Badanie7	5038	1,147909	0,039710	1,061955	1,240820	3,473717	0,000513	85,3%	10,24%
Badanie8	6031	1,129640	0,040301	1,043844	1,222488	3,024700	0,002488	91,29%	10,1%
Badanie9	5960	1,120299	0,037048	1,041833	1,204675	3,066137	0,002168	94,85%	0,36%
Badanie10	4277	1,124368	0,043956	1,031555	1,225532	2,666749	0,007658	82,75%	19,55%

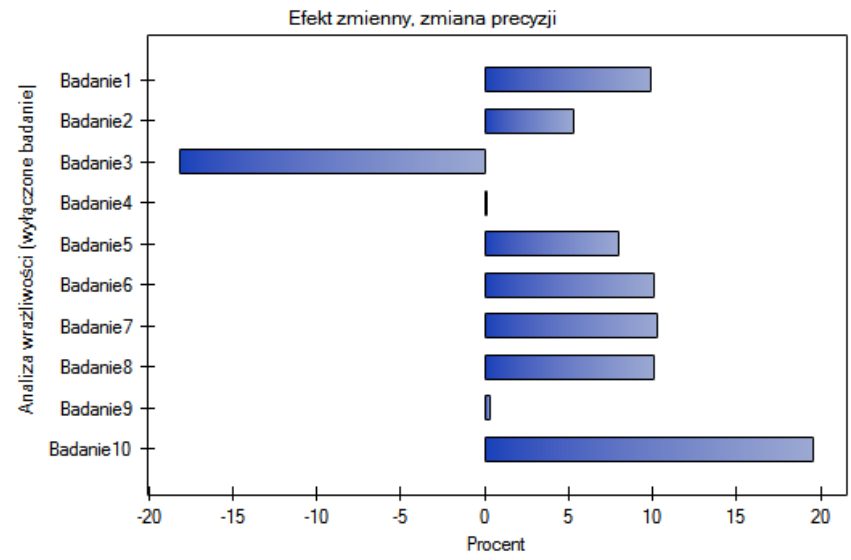
Wyniki wspólnego relatywnego ryzyka, nawet nieuwzględniające poszczególnych badań, wciąż pozostają istotne statystycznie.

Wyłączenie badania 3

Precyzja uzyskanego podsumowania wzrasta – przedział ufności dla efektu końcowego jest wówczas węższy o około 18%.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone





Do tych samych wniosków prowadzi analiza wykresów – najwęższy przedział i najkorzystniejszą zmianę precyzji uzyskamy po wykluczeniu badania 3.

Badanie asymetrii (obciążenie publikacyjne)

KIEDY STOSOWAĆ:

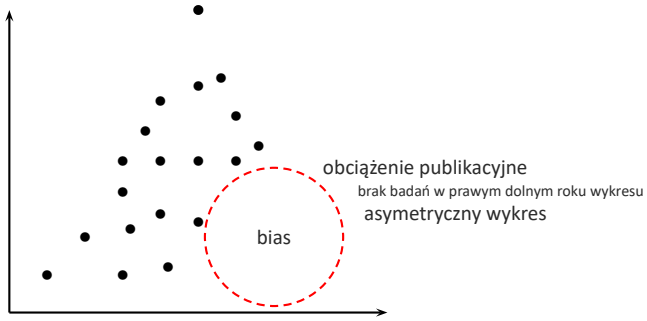
Symetria uzyskanych efektów zwykle świadczy o braku **obciążenia publikacyjnego**. Należy jednak pamiętać, że wiele obiektywnych czynników może zakłócić symetrię – np. często nie są publikowane badania o nieistotnych statystycznie efektach lub badania małe, co znacznie utrudnia dotarcie do takich wyników. Jednocześnie brak jest dostatecznie wszechstronnych i uniwersalnych narzędzi statystycznych do detekcji asymetrii, w konsekwencji znaczna część meta-analiz publikowana jest mimo zdiagnozowa-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

nej asymetrii; badania takie wymagają jednak dobrego uzasadnienia takiego postępowania.

DEFINICJE MIAR

Wykres lejkowy



Standardowym sposobem sprawdzenia obciążenia publikacyjnego w postaci asymetrii jest wykres lejkowy, pokazujący zależność pomiędzy wielkością badania (oś Y) a wielkością efektu podsumowującego (oś X). Zakłada się, że duże badania (umieszczone w górnej części wykresu) w prawidłowo dobranym zestawie są zlokalizowane blisko siebie i wyznaczają środek leja, a badania mniejsze znajdują się niżej oraz są bardziej zróżnicowane i symetrycznie rozłożone. Zamiast wielkości badania na osi Y można przedstawić błąd efektu w danym badaniu, co jest lepszym rozwiązaniem – błąd efektu jest bowiem miarą wskazującą na jakość badania i niosącą w sobie również informację o jego liczności.

Test Eggera – ponieważ interpretacja wykresu lejkowego jest zawsze subiektywna, pomocny może być współczynnik Eggera (Egger, 1997[54]) – wyraz wolny w dopasowanej prostej regresji. Współczynnik ten bazuje na korelacji pomiędzy odwrotnością błędu standardowego a stosunkiem wielkości efektu do jego błędu. Im bardziej oddalona od 0 wartość współczynnika, tym większa asymetria. Kierunek współczynnika określa rodzaj asymetrii: wartość dodatnia wraz z dodatnim przedziałem ufności wskazuje na zbyt wysoką wielkość efektu w małych badaniach, a ujemna wraz z ujemnym przedziałem ufności – na zbyt niską wielkość efektu w małych badaniach.

Uwaga! Test Eggera powinien być używany tylko przy dużym zróżnicowaniu wielkości badań i występowaniu badania o średniej wielkości. Przy niewielu badaniach (małej liczbie k) trudno jest osiągnąć istotny wynik mimo widocznej asymetrii.

Badanie liczby „Fail-safe" – liczba prac, jaka byłaby potrzebna, by podważyć uzyskany efekt podsumowujący. **Nfs Rosenthala** (Rosenthal, 1979[142]) określa liczbę prac niewskazujących na występowanie efektu (np. różnica średnich równa 0, iloraz szans równy 1 itp.), jaka jest potrzebna, by zredukować efekt podsumowujący z istotnego statystycznie do nieistotnego. Rosenthal (1984[131]) określił liczbę prac będącą **punktem odcięcia** jako $5k + 10$. Wyznaczając iloraz Nfs i punktu odcięcia, uzyskujemy **współczynnik(fs)** – jeśli jest on większy niż 1, prawdopodobieństwo obciążenia publikacyjnego jest minimalne. **Nfs Orwina** (Orwin, 1983[131]) określa natomiast liczbę prac o średnim efekcie wskazanym przez badacza, jaka jest potrzebna, by zredukować efekt podsumowujący do pożądanej wielkości wskazanej przez badacza.

JAK USTAWIĆ ANALIZĘ

Opcja Badanie asymetrii jest dostępna w oknie Podsumowanie meta-analizy (patrz powyżej).

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

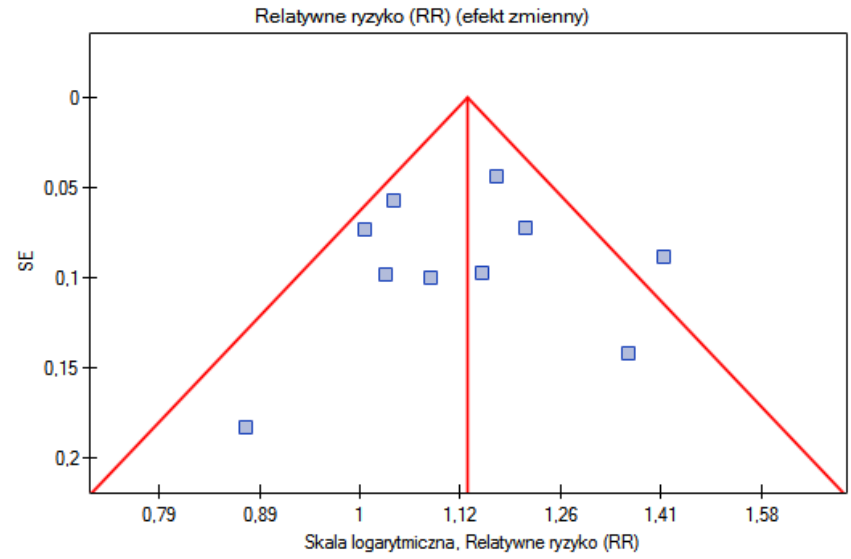
WYKONANIE I INTERPRETACJA

Przykład (29.2) c.d. *(plik MetaanalizaRR.pqs)*
Badając wpływ palenia papierosów na wystąpienie choroby X, sprawdzono założenie dotyczące asymetrii badań, a więc obciążenia publikacyjnego. W tym celu w oknie analizy wybrano opcję Badanie asymetrii oraz wskazano wykres lejkowy.

Badanie asymetrii	
Współczynnik b Egger	-0,141881052
SE(b)	1,268015989
-95% CI dla współczynnika b	-3,065931078
+95% CI dla współczynnika b	2,782168973
statystyka t dla b	-0,111892163
Stopnie swobody	8
Wartość p	0,91366544
Liczba publikacji "Fail-safe"	
Metoda Rosenthal'a	
N(fs)	72
Punkt odcięcia N(fs)	60
Współczynnik(fs)	1,2
Metoda Orwin'a	
Pożądany efekt	1,12
Średni efekt	1,11
N(fs)	11

Test Eggera: $p = 0.9137$ **(nieistotne statystycznie,** $\alpha = 0.05$**)**

Brak obciążenia publikacyjnego – rozkład badań na wykresie lejkowym jest w praktyce symetryczny.



Punkty przedstawiające poszczególne badania są rozłożone symetrycznie na wykresie lejkowym. Co prawda jedno badanie znajduje się poza obrębem trójkąta (Badanie 3), ale jest blisko jego brzegów,

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

więc również na podstawie wykresu nie mamy zasadniczych zastrzeżeń co do doboru badań – jedyna wątpliwość dotyczy badania trzeciego.

Liczba publikacji „Fail-safe" wyznaczona metodą Rosenthala jest duża i wynosi 72. Zatem, jeśli efekt podsumowujący (wspólne dla wszystkich badań relatywne ryzyko) miałby być nieistotny statystycznie (palenie papierosów miałoby nie wpływać na ryzyko choroby X), należałoby dołączyć do zebranych prac jeszcze 72 prace, w których relatywne ryzyko wynosi jeden. Uzyskany efekt można więc uznać za stabilny, gdyż nie będzie można łatwo (niewielką liczbą prac) podważyć uzyskanego efektu.

Uzyskane, wspólne relatywne ryzyko wynosi RR=1.13. Wykorzystując metodę Orwina, sprawdzono, jak wiele prac o relatywnym ryzyku równym 1.11 potrzeba włączyć, by wspólne relatywne ryzyko spadło do wielkości 1.12. Uzyskany rezultat to 11 prac. Zmniejszając natomiast wielkość relatywnego ryzyka z 1.11 do 1.10, już tylko 5 prac potrzeba, by wspólne relatywne ryzyko wyniosło 1.12.

DLA ZAINTERESOWANYCH

Test Eggera
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad b = 0, \\ \mathcal{H}_1: &\quad b \neq 0, \end{aligned}$$

gdzie:
b – wyraz wolny w równaniu regresji Eggera.

Wzory:
Statystyka testowa ma postać:

$$t = \frac{b}{SE(b)}$$

gdzie:
$SE(b)$ – błąd standardowy wyrazu wolnego.

Statystyka testowa ma rozkład t-Studenta z $k-2$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Nfs Rosenthala

$$Nfs = \frac{\left(\sum_{i=1}^{k} Z_i\right)^2}{Z_c^2} - k$$

gdzie:
Z_i – wartość statystyki testowej (o rozkładzie normalnym) danego badania,
Z_c – wartość krytyczna rozkładu normalnego dla zadanego poziomu istotności,
k – liczba badań w meta-analizie.

Nfs Orwina

$$Nfs = k\frac{M - M_d}{M_d - \overline{M_{fs}}}$$

gdzie:
M – efekt podsumowujący uzyskany w meta-analizie,

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

M_d – pożądana wielkość efektu wskazana przez badacza,
$\overline{M_{fs}}$ – średni efekt prac dodawanych do analizy, wskazany przez badacza.

Meta-analiza skumulowana

KIEDY STOSOWAĆ:

Typowym celem przeprowadzenia meta-analizy skumulowanej jest pokazanie, jak zmienił się efekt od czasu przeprowadzenia/opublikowania ostatniej meta-analizy dotyczącej danego tematu, lub jak zmieniał się on na przestrzeni ostatnich lat. Wówczas chronologicznie (zgodnie z linią czasu) dodawane są kolejne badania i każdorazowo wyliczany jest efekt podsumowujący. Równie ważna jest skumulowana analiza badająca zmianę efektu podsumowującego w zależności od wielkości wpływu wybranego, dodatkowego czynnika – wówczas badania są sortowane zgodnie z wielkością tego czynnika i **dla kolejno dodawanych badań wyliczany jest wspólny efekt podsumowujący**.

DEFINICJE MIAR

W zależności od celu kumulacji należy wybrać zmienną, według której poszczególne badania będą sortowane, czyli ustalić kolejność dodawania badań do podsumowania meta-analizą – może to być dowolna zmienna liczbowa. Przypisana do danego badania wartość **udziału skumulowanego** definiuje odsetek, jaki stanowi łączna waga włączonych do analizy badań, tzn. danego badania i badań je poprzedzających. **Zmiana precyzji** informuje natomiast, jak zmieni się precyzja efektu podsumowującego (szerokość przedziału ufności), gdy dane badanie zostanie dołączone do badań je poprzedzających. Dobrym zilustrowaniem analizy skumulowanej jest wykres leśny wielkości efektu oraz wykres zmian precyzji przy dołączaniu poszczególnych badań.

JAK USTAWIĆ ANALIZĘ

Opcja Meta-analiza skumulowana jest dostępna w oknie Podsumowanie meta-analizy (patrz powyżej).

WYKONANIE I INTERPRETACJA

Przykład (29.2) c.d. *(plik MetaanalizaRR.pqs)*
Badając wpływ palenia papierosów na wystąpienie choroby X, sprawdzono, jak uzyskane wyniki ewoluowały w czasie. W tym celu, w oknie analizy wybrano opcję Meta-analiza skumulowana oraz zmienną, według której kolejne prace będą włączane do meta-analizy, a także wskazano wykres leśny (skumulowany) i wykres słupkowy (skumulowany).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Skumulowana

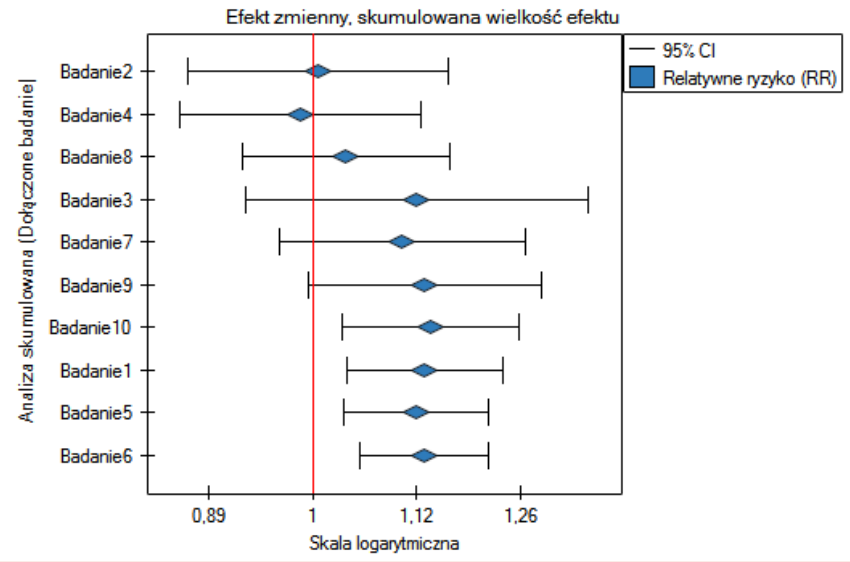
Nazwa	Liczność	Relatywn	SE (ln)	-95%CI	+95%CI	Statystyk	Wartość p	Udział sku	Zmiana prec	Wartość s
Badanie2	706	1,005398	0,073377	0,870721	1,160907	0,073375	0,941507	11,91%	NA	1
Badanie4	795	0,986786	0,068123	0,863448	1,127743	-0,19524	0,845197	15,29%	-8,92%	2
Badanie8	1009	1,037731	0,058633	0,925073	1,164109	0,631674	0,527599	24,01%	-9,56%	3
Badanie3	1624	1,121852	0,096799	0,927982	1,356224	1,187830	0,234900	33,77%	79,15%	4
Badanie7	2831	1,104040	0,069520	0,963406	1,265203	1,423713	0,154529	48,47%	-29,53%	5
Badanie9	3116	1,132179	0,065868	0,995057	1,288196	1,884739	0,059465	53,62%	-2,87%	6
Badanie10	5084	1,138658	0,049819	1,032730	1,255452	2,606410	0,009149	70,87%	-24,02%	7
Badanie1	5407	1,131981	0,044258	1,037927	1,234559	2,801062	0,005093	79,28%	-11,71%	8
Badanie5	5685	1,121108	0,040591	1,035371	1,213945	2,816305	0,004857	87,93%	-9,18%	9
Badanie6	6245	1,131399	0,036554	1,053175	1,215433	3,377304	0,000732	100%	-9,14%	10

Wraz z pojawianiem się nowych prac uzyskiwany efekt podsumowujący zyskiwał na sile, a jego istotność uzyskano po dodaniu do wcześniejszych prac Badania 10, a następnie również kolejnych badań.

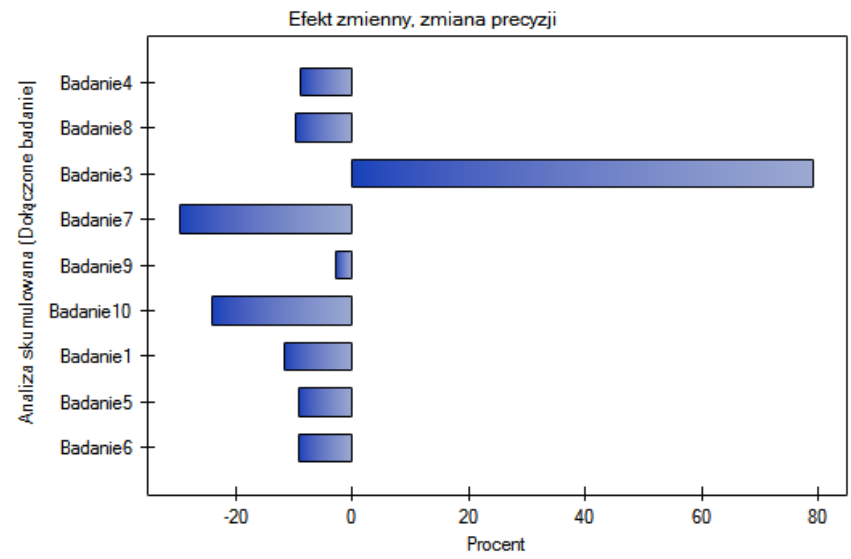
Dodanie Badania 3

Wyjątkowo osłabia precyzję podsumowania – przedział ufności podsumowującego relatywnego ryzyka poszerzył się wówczas o 79.15%.

Zasadniczo dodanie kolejnych prac zwiększało precyzję uzyskanego relatywnego ryzyka, za wyjątkiem sytuacji, w której dodano Badanie 3. Efekt ten widzimy w tabeli oraz na dołączonych wykresach.



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



W rezultacie należy się zastanowić nad wykluczeniem Badania 3 z przeprowadzonej meta-analizy.

Porównanie grup

KIEDY STOSOWAĆ:

Bywają sytuacje, w których zebrane dane dotyczą tego samego efektu, wykonywanego na tej samej populacji, ale w nieco innych warunkach – np. część badań przeprowadzono w warunkach A, a część w warunkach B. Wówczas interesujące może być porównanie efektów podsumowujących uzyskanych dla każdej grupy. Wykazanie różnic pomiędzy efektami podsumowującymi może być głównym celem meta-analizy – wówczas niewskazane jest podsumowanie obu podgrup jednym wspólnym efektem. Jeśli jednak badacz zdaje sobie sprawę, że badania przeprowadzano w różnych warunkach, ale celowe wydaje się wspólne podsumowanie wszystkich badań, to po wykazaniu braku istotnych statystycznie (lub klinicznie) różnic może dokonać wspólnego podsumowania z uwzględnieniem podziału na podgrupy A i B, czyli wyznaczyć wspólne podsumowanie skorygowane o różne warunki eksperymentu.

Na przykład: w kraju A panuje nieco inny klimat niż w kraju B. Dysponując szeregiem badań dotyczących obu krajów i badając roślinność tych krajów, możemy sprawdzić, czy warunki klimatyczne mają wpływ na uzyskane efekty badań. Analiza porównawcza wyznaczonych w ten sposób podgrup pozwoli ocenić, czy klimat ma zasadniczy wpływ na uzyskiwane wyniki, i czy wyniki badań obejmujące oba kraje można podsumować jednym wspólnym efektem, czy też należy wyznaczać oddzielne podsumowania dla każdego kraju. Innym przykładem może być podział badań na te z pełną randomizacją i bez niej, by sprawdzić, czy badania bez randomizacji dają na tyle zbliżone efekty do badań z randomizacją, by móc je włączyć do dalszej, wspólnej analizy.

DEFINICJE MIAR

Porównania grup można dokonać, wybierając jako efekt podsumowujący: efekt stały, efekt zmienny (oddzielne T^2) lub efekt zmienny (wspólne T^2), gdzie T^2 to wariancja obserwowanych efektów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **Efekt stały** wybieramy wtedy, gdy zakładamy, że badania wewnątrz każdej grupy dzielą jeden wspólny prawdziwy (tzn. populacyjny) efekt.
- **Efekt zmienny (oddzielne T^2)** wybieramy wtedy, gdy zakładamy, że badania wewnątrz każdej grupy reprezentują nieco inne populacje oraz że grupy różnią się wariancją pomiędzy badaniami.
- **Efekt zmienny (wspólne T^2)** wybieramy wtedy, gdy zakładamy, że badania wewnątrz każdej grupy reprezentują nieco inne populacje, ale wariancja pomiędzy badaniami jest taka sama, bez względu na grupę, do której przynależą.

W rezultacie każda grupa zostaje opisana oddzielnie wybranym efektem podsumowującym. Celem głównym jest porównanie grup, czyli ustalenie, czy porównywane grupy różnią się prawdziwym (populacyjnym) efektem podsumowującym – w praktyce jest to sprawdzenie, czy wariancja efektów podsumowujących grupy jest zerowa, czyli **badanie heterogeniczności grup**. Opis i interpretacja wyników tej analizy odpowiada rozdziałowi Badanie heterogeniczności, z tym że w przypadku porównania grup heterogeniczność dotyczy efektów podsumowujących porównywanych grup, a nie poszczególnych badań, a wynik jest uzależniony od wybranego efektu podsumowującego. Jeśli wynik jest istotny statystycznie (statystyki Q, współczynnika I^2 lub współczynnika H^2), jest to mocna sugestia, by zrezygnować ze wspólnego podsumowywania porównywanych grup.

Dodatkowe możliwości analizy porównania grup:

Badanie heterogeniczności w grupach Możliwość przeanalizowania każdej grupy oddzielnie pod kątem heterogeniczności (zgodnie z opisem w rozdziale Badanie heterogeniczności). Uzyskane wyniki, w szczególności wyznaczona wariancja T^2, ułatwiają wybór efektu zmiennego (oddzielne T^2) lub efektu zmiennego (wspólne T^2) do porównania grup.

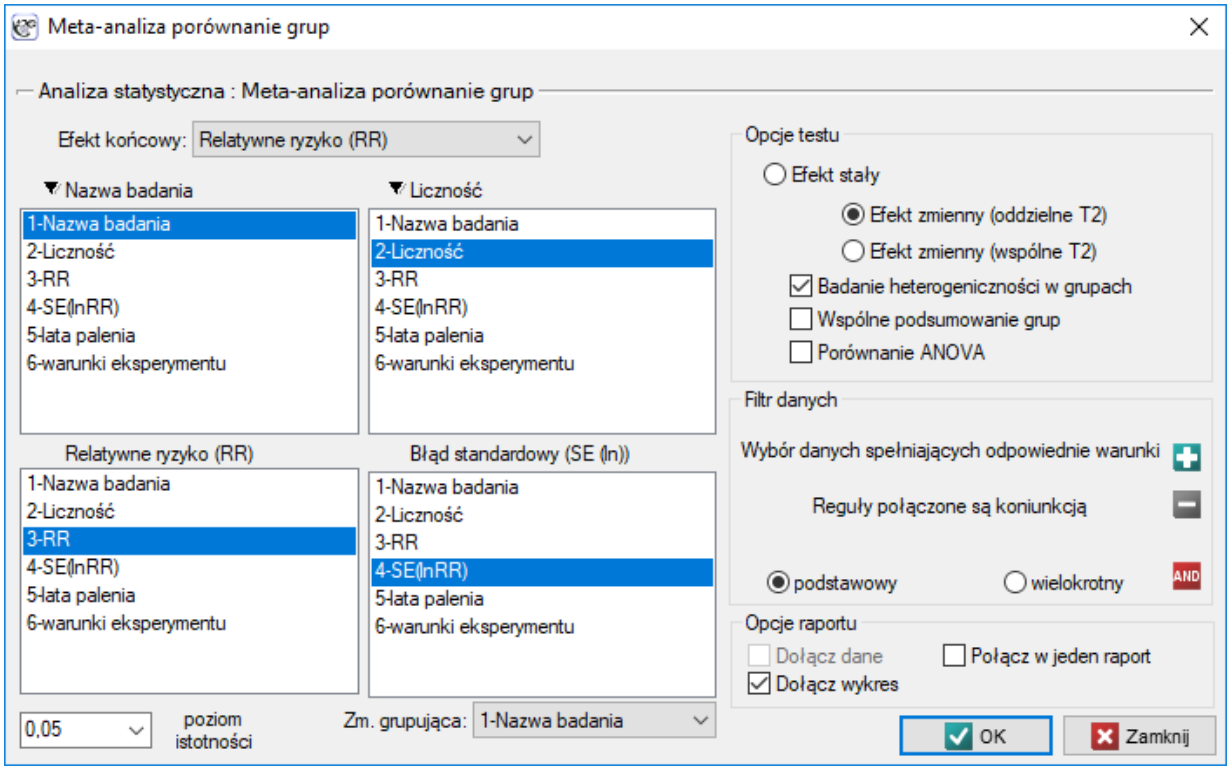
Wspólne podsumowanie grup Jeśli różnice uzyskane pomiędzy efektami podsumowującymi poszczególne grupy są niewielkie i nieistotne, można dokonać wspólnego podsumowania tych grup, skorygowanego o podział na wskazane grupy. Rezultat wspólnego podsumowania zależy od obserwowanych różnic (wariacji pomiędzy badaniami i pomiędzy grupami), czyli od wyboru efektu podsumowującego. Dobrym zilustrowaniem wspólnego podsumowania grup jest wykres leśny przedstawiający wyniki poszczególnych badań wraz z podsumowaniem każdej grupy i wspólnym podsumowaniem grup.

Porównanie ANOVA Dodatkowa opcja porównania grup, oparta na innym modelu matematycznym niż porównanie poprzez badanie heterogeniczności grup, choć obie metody dają zbieżne wyniki. Obserwowana wariancja zostaje rozbita na wariancję między grupami i wariancję wewnątrz grup, a ta – na wariancję każdej grupy oddzielnie.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji porównania grup dla meta-analizy wywołujemy poprzez menu: Statystyki zaawansowane→Meta-analiza→Porównanie grup.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



WYKONANIE I INTERPRETACJA

PRZYKŁAD 29.3. (plik MetaanalizaRR.pqs)
Badano ryzyko choroby X dla osób palących i niepalących. By ustalić, czy czas pozostawania w nałogu ma wpływ na występowanie choroby X, zaplanowano przeprowadzenie meta-analizy. Wykonano dokładny przegląd literatury i na tej podstawie wytypowano 17 prac naukowych dysponujących opisem relatywnego ryzyka i jego błędu (precyzji danego badania). Ze względu na to, że badania dotyczyły różnego czasu palenia, wyodrębniono 3 grupy badań:

(1) badania dotyczące palących dłużej niż 10 lat,

(2) badania dotyczące palących od 5 do 10 lat,

(3) badania dotyczące palących krócej niż 5 lat.

Dodatkowo dokonano podziału na dwa różne warunki przeprowadzanych badań (różne kryteria włączenia/wyłączenia osób). Dane przygotowano do meta-analizy i zapisano w pliku. Celem przeprowadzania meta-analizy było porównanie grup wiekowych. Dodatkowo sprawdzono, czy różne warunki eksperymentu przełożyły się na różnice w uzyskanym relatywnym ryzyku.

Ze względu na to, że prace włączone do meta-analizy pochodziły z różnych ośrodków i obejmowały nieco inne populacje, podsumowania dokonano, wybierając efekt zmienny (oddzielne T2). Jako efekt końcowy wybrano relatywne ryzyko oraz przedstawiono wyniki na wykresie leśnym.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Meta-analiza porównanie grup	
Czas analizy	0,07 sek.
Analizowane zmienne	Nazwa badania
	Liczność
	RR
	SE(lnRR)
	lata palenia
Poziom istotności	0,05
Zmienna grupująca	lata palenia
Efekt	Zmienny oddzielne T2
Efekt końcowy	Relatywne ryzyko (RR)
Liczba grup	3
Badanie heterogeniczności	
Porównanie grup	
Statystyka Q	9,374994454
Stopnie swobody	2
Wartość p	0,009209707

Badanie heterogeniczności			
Porównanie grup	Wartości	-95%CI	+95%CI
T2	0,004566	0,000571	0,017379
H2	4,687497	1,461467	15,03463
I2	78,67%	31,58%	93,35%

$p = 0.0092$ dla heterogeniczności grup (istotne statystycznie, $\alpha = 0.05$)

Grupy różnią się istotnie statystycznie – widać to nie tylko w teście heterogeniczności, ale również we współczynniku H^2 (wraz z przedziałem ufności powyżej wartości jeden) i współczynniku I^2 (78%, czyli wysoka heterogeniczność).

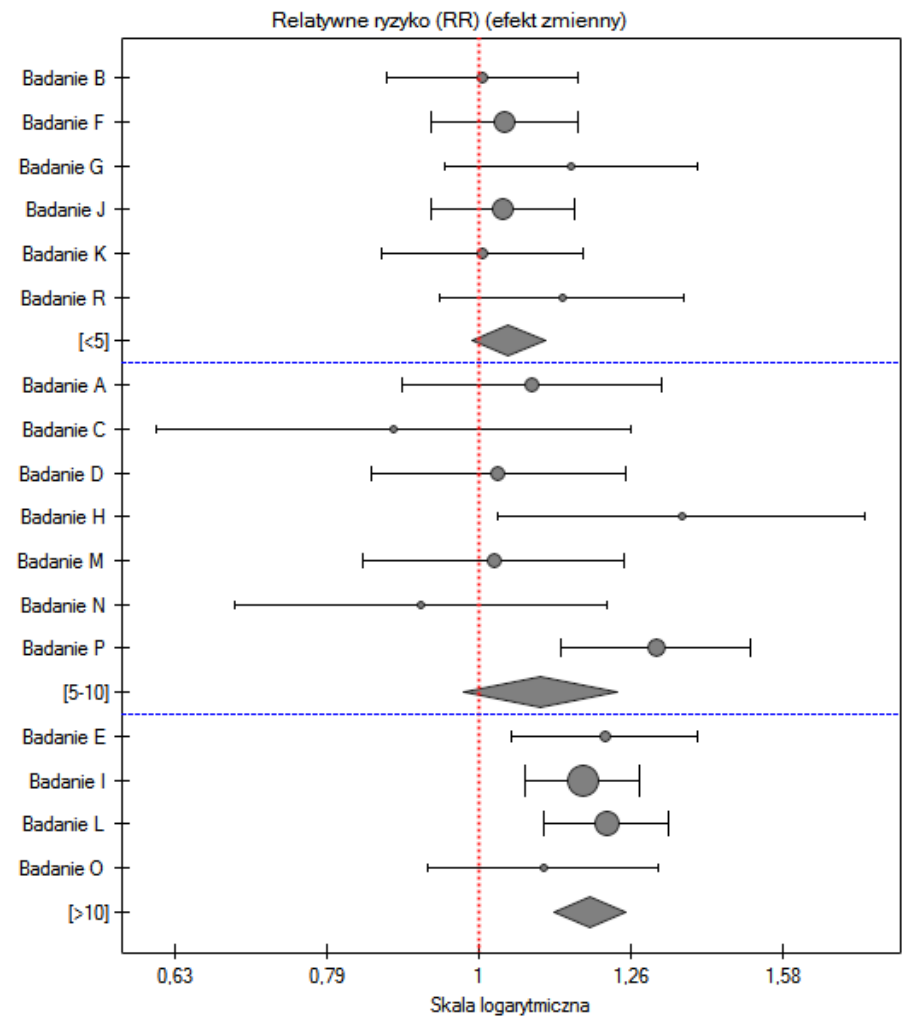
Dlatego zebrane prace nie zostaną podsumowane wspólnym efektem, a jedynie poprzez oddzielne podsumowanie każdej grupy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dane :										
Nazwa	Liczność	Relatywn	SE (ln)	-95%CI	+95%CI	Statystyk	Wartość p	Wariancja	Wagi	Udział
Grupa:<5										
Badanie B	706	1,005398	0,073377	0,870721	1,160907	0,073375	0,941507	0,005384	185,7253	15,47%
Badanie F	1207	1,04	0,057037	0,929999	1,163010	0,687630	0,491685	0,003253	307,3826	25,6%
Badanie G	214	1,15	0,097478	0,95	1,392105	1,433765	0,151639	0,009502	105,2393	8,77%
Badanie J	1155	1,036486	0,055583	0,929503	1,155783	0,644735	0,519098	0,003089	323,6752	26,96%
Badanie K	626	1,006050	0,077672	0,863982	1,171479	0,077668	0,938091	0,006032	165,7550	13,81%
Badanie R	270	1,134749	0,094163	0,943512	1,364748	1,342469	0,179443	0,008866	112,7797	9,39%
Razem	4178	1,046546	0,028860	0,988990	1,107452	1,576386	0,114936			
Grupa:5-10										
Badanie A	323	1,083815	0,100264	0,890450	1,319169	0,802748	0,422120	0,010052	46,54255	16,34%
Badanie C	89	0,878151	0,183328	0,613083	1,257821	-0,70876	0,478472	0,033609	22,20142	7,79%
Badanie D	278	1,03	0,098000	0,849999	1,248117	0,301618	0,762942	0,009604	47,53562	16,69%
Badanie H	285	1,36	0,141801	1,03	1,795728	2,168415	0,030127	0,020107	31,70536	11,13%
Badanie M	250	1,023215	0,101149	0,839206	1,247571	0,226895	0,820505	0,010231	46,15984	16,2%
Badanie N	150	0,916363	0,143949	0,691094	1,215060	-0,60675	0,544014	0,020721	31,10010	10,92%
Badanie P	481	1,308106	0,073059	1,133587	1,509493	3,676208	0,000236	0,005337	59,62897	20,93%
Razem	1856	1,097000	0,059247	0,976730	1,232079	1,562579	0,118151			
Grupa:>10										
Badanie E	560	1,21	0,072363	1,050000	1,394380	2,634199	0,008433	0,005236	190,9672	15,1%
Badanie I	1968	1,17	0,0437	1,073960	1,274627	3,592763	0,000327	0,001909	523,6451	41,39%
Badanie L	1583	1,213419	0,048516	1,103350	1,334467	3,987152	0,000066	0,002353	424,8383	33,58%
Badanie O	414	1,102678	0,089246	0,925726	1,313455	1,095199	0,273429	0,007964	125,5514	9,92%
Razem	4525	1,183450	0,028116	1,119999	1,250496	5,990682	0,000000			

Wykres leśny również przedstawia podsumowanie każdej grupy i nie zawiera wspólnego podsumowania grup.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Dodatkowo sprawdzono homogeniczność w każdej z grup, by upewnić się co do możliwości ich oddzielnego podsumowania.

Badanie heterogeniczności	
<5	
Statystyka Q	2,27285084
Stopnie swobody	5
Wartość p	0,810245983
5-10	
Statystyka Q	11,729232984
Stopnie swobody	6
Wartość p	0,068289157
>10	
Statystyka Q	1,055532164
Stopnie swobody	3
Wartość p	0,787818159

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Badanie heterogeniczności			
<5	Wartości	-95%CI	+95%CI
T2	0	0	0,004108
H2	0,454570	0,115348	1,791383
I2	0%	0%	44,18%
5-10	Wartości	-95%CI	+95%CI
T2	0,011432	0	0,043367
H2	1,954872	0,826799	4,622070
I2	48,85%	0%	78,36%
>10	Wartości	-95%CI	+95%CI
T2	0	0	0,004504
H2	0,351844	0,053874	2,297807
I2	0%	0%	56,48%

Wyniki porównania dotyczącego różnych warunków prowadzenia badań wskazują natomiast na brak istotnego wpływu tych warunków na efekt podsumowujący. W tym przypadku możliwy jest do wyliczenia wspólny efekt końcowy skorygowany o różne warunki przeprowadzania badań, czyli wspólne podsumowanie obu grup.

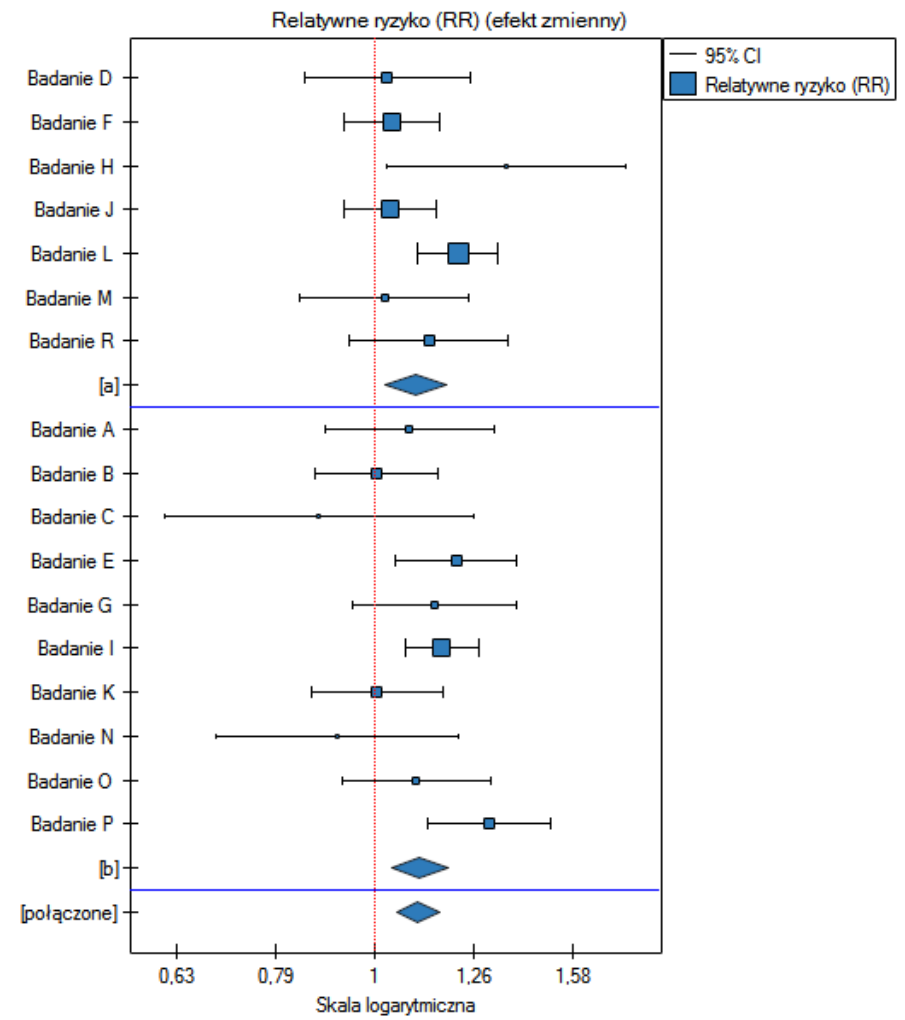
Meta-analiza porównanie grup	
Czas analizy	0,09 sek.
Analizowane zmienne	Nazwa badania
	Liczność
	RR
	SE(lnRR)
	warunki eksperymentu
Poziom istotności	0,05
Zmienna grupująca	warunki eksperymentu
Efekt	Zmienny oddzielne T2
Efekt końcowy	Relatywne ryzyko (RR)
Liczba grup	2
Badanie heterogeniczności	
a	
Statystyka Q	9,492828171
Stopnie swobody	6
Wartość p	0,14769954
b	
Statystyka Q	14,588588591
Stopnie swobody	9
Wartość p	0,102874491
Porównanie grup	
Statystyka Q	0,043648303
Stopnie swobody	1
Wartość p	0,834509314

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Heterogeneity analysis			
1	Values	-95%CI	+95%CI
T2	0,003096	0	0,014651
H2	1,582138	0,666774	3,754133
I2	36,79%	0%	73,36%
2	Values	-95%CI	+95%CI
T2	0,004013	0	0,015507
H2	1,620954	0,773008	3,399046
I2	38,31%	0%	70,58%
Comparison groups	Values	-95%CI	+95%CI
T2	0	NA	NA
H2	0,043648	NA	NA
I2	0%	NA	NA

Dane :										
Nazwa	Liczność	Relatywn	SE (ln)	-95%CI	+95%CI	Statystył	Wartość	Wariancja	Wagi	Udział
Grupa:a										
Badanie D	278	1,03	0,098000	0,849999	1,248117	0,301618	0,762942	0,009604	78,73376	10,06%
Badanie F	1207	1,04	0,057037	0,929999	1,163010	0,687630	0,491685	0,003253	157,4758	20,11%
Badanie H	285	1,36	0,141801	1,03	1,795728	2,168415	0,030127	0,020107	43,09494	5,5%
Badanie J	1155	1,036486	0,055583	0,929503	1,155783	0,644735	0,519098	0,003089	161,6443	20,64%
Badanie L	1583	1,213419	0,048516	1,103350	1,334467	3,987152	0,000066	0,002353	183,4612	23,43%
Badanie M	250	1,023215	0,101149	0,839206	1,247571	0,226895	0,820505	0,010231	75,02986	9,58%
Badanie R	270	1,134749	0,094163	0,943512	1,364748	1,342469	0,179443	0,008866	83,58588	10,67%
Razem	5028	1,100975	0,035736	1,026499	1,180855	2,691843	0,007105			
Grupa:b										
Badanie A	323	1,083815	0,100264	0,890450	1,319169	0,802748	0,422120	0,010052	71,08887	8,03%
Badanie B	706	1,005398	0,073377	0,870721	1,160907	0,073375	0,941507	0,005384	106,4032	12,02%
Badanie C	89	0,878151	0,183328	0,613083	1,257821	-0,70876	0,478472	0,033609	26,57924	3%
Badanie E	560	1,21	0,072363	1,050000	1,394380	2,634199	0,008433	0,005236	108,1032	12,21%
Badanie G	214	1,15	0,097478	0,95	1,392105	1,433765	0,151639	0,009502	73,98606	8,36%
Badanie I	1968	1,17	0,0437	1,073960	1,274627	3,592763	0,000327	0,001909	168,8162	19,07%
Badanie K	626	1,006050	0,077672	0,863982	1,171479	0,077668	0,938091	0,006032	99,53309	11,24%
Badanie N	150	0,916363	0,143949	0,691094	1,215060	-0,60675	0,544014	0,020721	40,42786	4,57%
Badanie O	414	1,102678	0,089246	0,925726	1,313455	1,095199	0,273429	0,007964	83,48098	9,43%
Badanie P	481	1,308106	0,073059	1,133587	1,509493	3,676208	0,000236	0,005337	106,9340	12,08%
Razem	5531	1,112317	0,033607	1,041410	1,188053	3,167291	0,001538			
Podsumowanie										
a	5028	1,100975	0,035736	1,026499	1,180855	2,691843	0,007105	0,001277	783,0259	46,93%
b	5531	1,112317	0,033607	1,041410	1,188053	3,167291	0,001538	0,001129	885,3530	53,07%
Razem	10559	1,106980	0,024482	1,055116	1,161393	4,151398	0,000033			

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



DLA ZAINTERESOWANYCH

Hipotezy dla badania heterogeniczności grup:

$$\begin{aligned} \mathcal{H}_0: &\quad \tau^2 = 0, \\ \mathcal{H}_1: &\quad \tau^2 \neq 0, \end{aligned}$$

gdzie:
τ^2 – wariancja prawdziwych (populacyjnych) efektów podsumowujących porównywanych grup.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Porównanie ANOVA
Procedura:
Wykorzystując podejście ANOVA, obserwowaną wariancję łączną rozbija się na wariancję między grupami i wariancję wewnątrz grup, tę zaś na wariancję każdej grupy oddzielnie. W rezultacie wyznaczone zostają następujące statystyki Q:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Statystyka Q (grupa 1) – bada tę część wariancji łącznej, która odnosi się do grupy pierwszej, czyli wariancję pomiędzy badaniami znajdującymi się wewnątrz grupy pierwszej,
- Statystyka Q (grupa 2) – analogicznie dla grupy drugiej,
- ...
- Statystyka Q (grupa g) – analogicznie dla grupy ostatniej,
- Statystyka Q (wewnątrz grup) $= Q$ (grupa 1) $+ Q$ (grupa 2) $+ \ldots +$ Q (grupa g) – bada tę część wariancji łącznej, która odnosi się do wnętrza poszczególnych grup,
- **Statystyka Q (między grupami)** – bada tę część wariancji łącznej, która odnosi się do różnic między grupami (wynik tożsamy z badaniem heterogeniczności grup),
- Statystyka Q (łączna) – bada wariancję pomiędzy wszystkimi badaniami.

Każda z powyższych statystyk Q ma rozkład χ^2 z odpowiednią dla niej liczbą stopni swobody.

Meta-regresja

KIEDY STOSOWAĆ:

Meta-regresja jest przeprowadzana w sposób analogiczny do analizy regresji opisanej w dziale Liniowa regresja wieloraka. W przypadku meta-regresji badanymi obiektami są poszczególne badania: ich wyniki (np. ilorazy szans, relatywne ryzyka, różnice średnich) stanowią zmienną zależną Y, czyli wyjaśnianą, natomiast dodatkowe warunki przeprowadzania tych badań stanowią zmienne niezależne (X_1, X_2, ..., X_k), czyli wyjaśniające. Podobnie jak w tradycyjnych modelach regresji, zmienne niezależne mogą wchodzić w interakcje, a te opisane skalą nominalną mogą podlegać specjalnemu kodowaniu (więcej informacji w rozdziale Przygotowanie zmiennych do analizy w modelach wielowymiarowych).

Meta-regresję można przeprowadzić, wybierając efekt stały lub efekt zmienny.

- **Efekt stały** wybieramy wtedy, gdy zakładamy, że badania przedstawiają jeden wspólny prawdziwy efekt w taki sposób, że wszystkie czynniki, które mogłyby zaburzać jego wielkość, są takie same – za wyjątkiem czynników badanych jako zmienne niezależne w modelu. Jest to sytuacja rzadka w praktyce, ponieważ wymaga w pełni kontrolowanych warunków (trudnych do spełnienia, gdy badania prowadzą różne ośrodki i różni badacze). Miałaby ona uzasadnienie np. wtedy, gdyby wszystkie badania przeprowadzał jeden ośrodek, na tej samej populacji, przy zmianie tylko tych warunków, które opisuje badana cecha.
- **Efekt zmienny** wybieramy wtedy, gdy zakładamy, że badania mogą przedstawiać nieco różniące się populacje, tzn. nie wszystkie czynniki mogące zaburzać wielkość badanego efektu są opisane we wszystkich pracach (można założyć, że są podobne, ale nie muszą być identyczne). Wykorzystanie efektu zmiennego jest częstsze, ponieważ poszczególne badania przeprowadzane są zwykle przez różne ośrodki w nieco innych warunkach.

WARUNKI STOSOWANIA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– liczba zmiennych niezależnych powinna być niewielka – mniejsza niż liczba prac, na podstawie których przeprowadza się badanie ($n \geq k + 1$).

DEFINICJE MIAR

Istotność statystyczna poszczególnych zmiennych w modelu – na podstawie współczynnika oraz jego błędu wnioskujemy, czy zmienna niezależna, dla której ten współczynnik oszacowano, wywiera istotny wpływ na efekt końcowy.

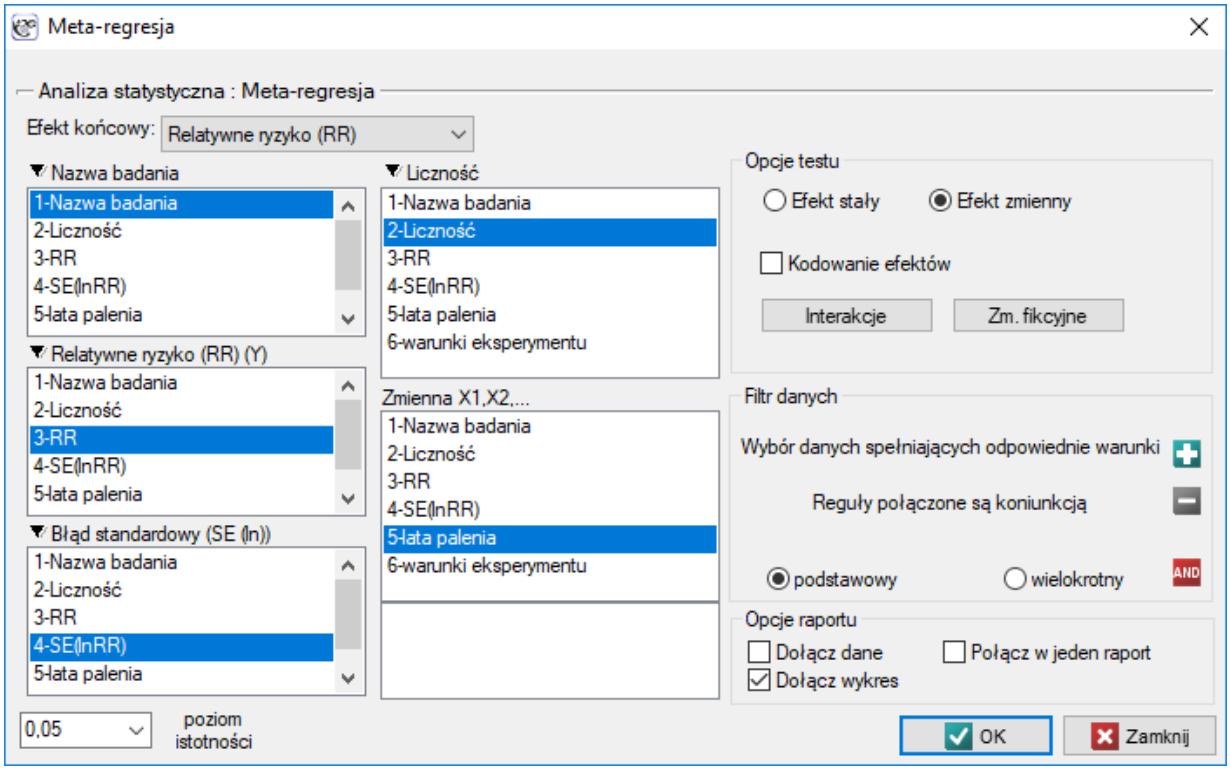
Jakość zbudowanego modelu meta-regresji oceniamy kilkoma miarami:

- **Współczynnik** R^2 – miara dopasowania modelu, wyrażająca procent zmienności pomiędzy efektami badań tłumaczony przez model. Wartość tego współczynnika mieści się w przedziale $< 0; 1 >$, gdzie 1 oznacza doskonałe dopasowanie modelu, a 0 – zupełny brak dopasowania.
- **Współczynnik** I^2 – określa procent obserwowanej wariancji, jaki wynika z rzeczywistej różnicy w wielkości badanych efektów (por. Badanie heterogeniczności).

Istotność statystyczna wszystkich zmiennych w modelu sprawdzana jest podejściem ANOVA, wyznaczającym statystykę Q(modelu) – obserwowaną wariancję pomiędzy badaniami rozbija się wówczas na wariancję tłumaczoną przez model i wariancję reszt (nietłumaczoną przez model).

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji meta-regresji wywołujemy poprzez menu:
Statystyki zaawansowane→Meta-analiza→Meta-regresja.



Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

WYKONANIE I INTERPRETACJA

Przykład (29.3) c.d. *(plik MetaanalizaRR.pqs)*
Badano ryzyko choroby X dla osób palących i niepalących. By ustalić, czy czas pozostawania w nałogu ma wpływ na występowanie choroby X oraz czy różne warunki eksperymentu przełożyły się na różnice w uzyskanym relatywnym ryzyku, wykonano meta-analizę porównującą wyodrębnione grupy badań. Na podstawie porównania grup badań udało się ustalić, że ostatnia grupa (palących najdłużej, tzn. dłużej niż 10 lat) wskazuje na związek pomiędzy paleniem a występowaniem choroby X, natomiast dla grup, w których czas palenia był krótszy, nie udało się uzyskać istotnego efektu. Zauważono jednak, że efekt systematycznie rośnie wraz z upływem lat palenia. By sprawdzić hipotezę o istotnym zwiększeniu ryzyka choroby X wraz z upływem lat palenia, zbudowano dwa modele regresji: w pierwszym zmienną grupującą Lata palenia potraktowano jako zmienną ciągłą, a w drugim jako zmienną kategorialną (fikcyjną) z grupą odniesienia palącą krócej niż 5 lat. Dane przygotowano do meta-analizy i zapisano w pliku.

Ze względu na to, że prace włączone do meta-analizy pochodziły z różnych ośrodków i obejmowały nieco inne populacje, meta-regresję wykonano, wybierając efekt zmienny. Jako efekt końcowy wybrano relatywne ryzyko oraz przedstawiono wyniki na wykresie.

Model

	wsp. b	błąd b	-95% CI	+95% CI	Statystyk	Wartość p
w. wolny	-0,01422	0,044721	-0,10188	0,073425	-0,31814	0,7503744
lata palenia	0,061387	0,020258	0,021681	0,101094	3,030167	0,0024441

Model

	wsp. b	błąd b	-95% CI	+95% CI	Statystyk	Wartość p
w. wolny	0,046212	0,030357	-0,01328	0,105712	1,522243	0,1279481
lata palenia[2]	0,066634	0,051263	-0,03383	0,167109	1,299850	0,1936521
lata palenia[3]	0,121823	0,043046	0,037454	0,206193	2,830053	0,0046540

Istotny związek pomiędzy czasem palenia a wielkością relatywnego ryzyka w obu modelach

Wraz z upływem czasu palenia (przejściem do kolejnej grupy lat palenia) logarytm relatywnego ryzyka choroby X zwiększa się o 0.06139. W modelu drugim, w odniesieniu do palących krócej niż 5 lat, logarytm relatywnego ryzyka wzrasta o 0.06663 dla palących od 5 do 10 lat oraz aż o 0.12182 dla palących dłużej niż 10 lat.

Ponieważ część badań prowadzono według innych kryteriów (w innych warunkach), uzyskane wyniki obu modeli skorygowano o różne warunki prowadzenia badań.

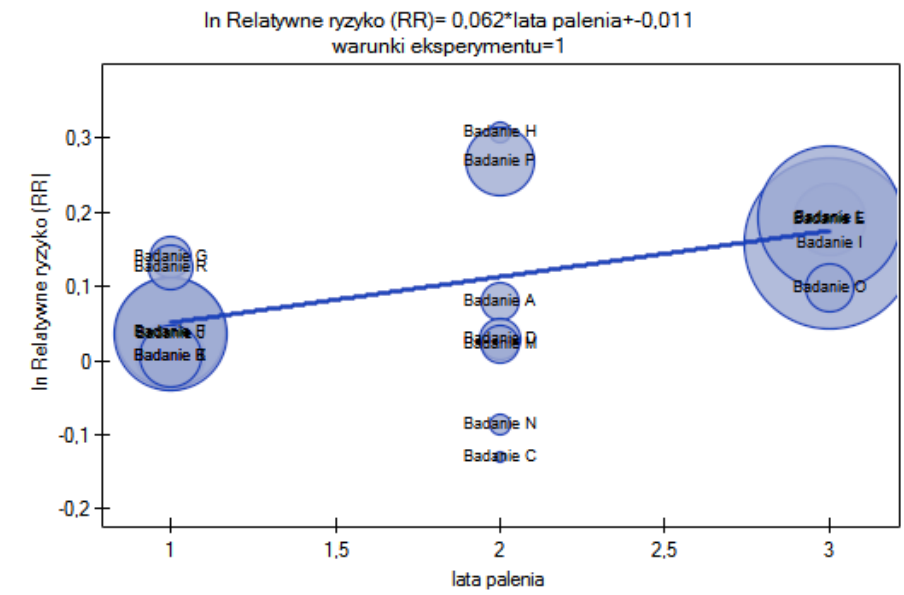
Model

	wsp. b	błąd b	-95% CI	+95% CI	Statystyk	Wartość p
w. wolny	-0,00265	0,068790	-0,13748	0,132168	-0,03863	0,9691846
lata palenia	0,062078	0,022243	0,018481	0,105676	2,790826	0,0052573
warunki eks	-0,00853	0,039584	-0,08611	0,069052	-0,21551	0,8293636

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Model						
	wsp. b	błąd b	-95% CI	+95% CI	Statystyk	Wartość p
w. wolny	0,059451	0,066746	-0,07136	0,190272	0,890703	0,3730884
lata palenia[2]	0,065415	0,053635	-0,03970	0,170538	1,219631	0,2226045
lata palenia[3]	0,123049	0,047253	0,030435	0,215663	2,604060	0,0092126
warunki ekspery	-0,00891	0,041757	-0,09075	0,072928	-0,21348	0,8309512

Przeprowadzona korekcja nie zmieniła zasadniczej tendencji, a więc można uznać, że ryzyko wystąpienia choroby X wzrasta wraz z upływem lat palenia, bez względu na przyjętą metodologię (kryteria włączenia/wyłączenia osób). Uzyskaną zależność dla modelu pierwszego, przy założeniu prowadzenia badań w warunkach „a" (wskazanych jako warunki pierwsze), przedstawia wykres.



DLA ZAINTERESOWANYCH

Istotność statystyczna poszczególnego współczynnika
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \beta_i = 0, \\ \mathcal{H}_1: &\quad \beta_i \neq 0. \end{aligned}$$

Wzory:
Statystyka testowa ma postać:

$$Z = \frac{b_i}{SE_{b_i}}$$

Statystyka testowa ma rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Współczynnik R^2

$$R^2 = T^2_{(modelu)} + T^2_{(total)},$$

gdzie:
$T^2_{(modelu)}$ – wariancja między badaniami wyjaśniona przez model,
$T^2_{(total)}$ – całkowita wariancja między badaniami.

Istotność statystyczna wszystkich zmiennych w modelu
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{wszystkie } \beta_i = 0, \\ \mathcal{H}_1: &\quad \text{istnieje } \beta_i \neq 0. \end{aligned}$$

Procedura:
Wykorzystując podejście ANOVA, obserwowaną wariancję pomiędzy badaniami rozbija się na wariancję tłumaczoną przez model i wariancję reszt (nietłumaczoną przez model). W rezultacie wyznaczone zostają następujące statystyki Q:

- Statystyka Q (reszty) – bada tę część wariancji łącznej, która nie jest tłumaczona przez model,
- Statystyka Q (modelu) – bada tę część wariancji łącznej, która jest tłumaczona przez model,
- Statystyka Q (łączna) – bada wariancję pomiędzy wszystkimi badaniami.

Każda z powyższych statystyk Q ma rozkład χ^2 z odpowiednią dla niej liczbą stopni swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α: gdy $p \leq \alpha$, odrzucamy $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$; gdy $p > \alpha$, nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

30 ANALIZA RZETELNOŚCI

Rzetelność skali

KIEDY STOSOWAĆ:

Analizę rzetelności (*ang. reliability analysis*) wykonuje się najczęściej wtedy, gdy budujemy skalę złożoną, czyli kwestionariusz lub test składający się z wielu pojedynczych pozycji (pytań), które sumujemy do jednego wyniku. Chcemy wówczas sprawdzić, czy poszczególne pytania rzeczywiście mierzą to samo – innymi słowy, na ile możemy ufać, że suma punktów uzyskanych ze skali dobrze oddaje badaną cechę (np. kompetencje, poziom lęku, satysfakcję z pracy). Mówimy wtedy o **rzetelności** skali, rozumianej jako jej wewnętrzna spójność.

WARUNKI STOSOWANIA

- skala złożona z co najmniej kilku pozycji (pytań), które są sumowane do jednego wyniku ogólnego,
- poszczególne pozycje skali powinny mierzyć ten sam konstrukt, czyli korelacja pomiędzy nimi powinna być wysoka – założenie to można sprawdzić, wyliczając macierz współczynników korelacji Pearsona pomiędzy pozycjami skali.

DEFINICJE MIAR

Pomimo, że w zagadnieniu rzetelności mogą być wykorzystywane różne miary zgodności, najczęściej używaną techniką jest tutaj współczynnik α-Cronbacha i tzw. rzetelność połówkowa.

Współczynnik α-Cronbacha (*ang. Cronbach's alpha*) został tak nazwany po raz pierwszy przez Cronbacha w 1951[50]. Mierzy on, w uproszczeniu, stosunek zmienności poszczególnych pozycji skali do zmienności całej skali (czyli sumy tych pozycji) – innymi słowy pokazuje, jak silnie pojedyncze pytania „trzymają się razem". Współczynnik α może przyjmować dowolne wartości mniejsze lub równe 1, w tym również wartości ujemne, jednak sensowne są tylko wartości dodatnie. Jeżeli wszystkie pozycje skali są idealnie rzetelne, wówczas współczynnik rzetelności wynosi 1. Obok zwykłego współczynnika α_C wyliczana jest też jego wersja standaryzowana ($\alpha_{standard}$), oparta na średniej korelacji pomiędzy pozycjami skali.

Pomocą w ocenie przydatności konkretnych pozycji skali służą:

- wartość współczynnika α_C obliczonego po usunięciu określonej pozycji skali,
- wartość odchylenia standardowego skali obliczonego po usunięciu określonej pozycji skali,
- wartość średnia skali obliczona po usunięciu określonej pozycji skali,
- wartość korelacji określonej pozycji i sumy pozycji pozostałych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Jeśli po usunięciu jakiejś pozycji współczynnik α_C wyraźnie rośnie, jest to sygnał, że dana pozycja słabo pasuje do reszty skali i warto rozważyć jej usunięcie lub przeredagowanie.

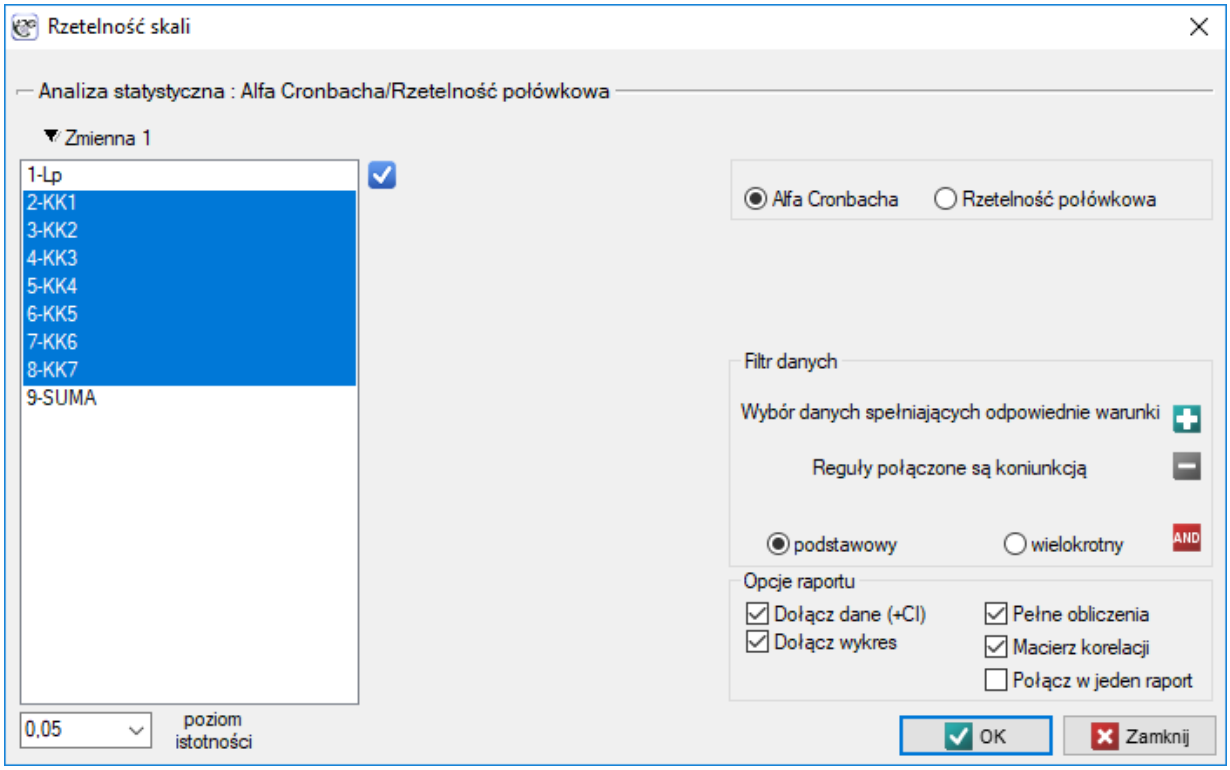
Rzetelność połówkowa (*ang. split-half reliability*) to inne podejście do tego samego zagadnienia. Polega ono na losowym podzieleniu pozycji skali na dwie połowy i sprawdzeniu, na ile te dwie połówki są ze sobą zgodne (skorelowane). Wykorzystuje się do tego współczynnik rzetelności połówkowej Spearmana-Browna, opublikowany niezależnie przez Spearmana (1910)[159] i Browna (1910)[34], oraz alternatywny wzór zaproponowany przez Guttmana (1945)[74]. Jeżeli dwie losowo wybrane połówki skali są idealnie skorelowane, współczynnik rzetelności połówkowej wynosi 1.

Uwaga! Przyjmuje się, że aby można było uznać skalę za rzetelną, współczynniki rzetelności skali (α_C, $\alpha_{standard}$, r_{SH}, r_{SHG}) powinny być większe niż 0.6 i mniejsze niż 1.

Dodatkowo, dla skali uznanej za rzetelną, wyznacza się **błąd standardowy pomiaru** (*ang. standard error of measurement*, SEM), który informuje, jak dużego rozrzutu wyniku pomiaru możemy się spodziewać z powodu niedoskonałej rzetelności skali – im niższa rzetelność, tym większy błąd pomiaru.

JAK USTAWIĆ ANALIZĘ

Okno z ustawieniami opcji rzetelności skali wywołujemy poprzez menu Statystyka→Rzetelność skali.



WYKONANIE I INTERPRETACJA

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

PRZYKŁAD 30.1. (plik skala.pqs)
Pomocą w ocenie przydatności przyszłych pracowników pewnej firmy konsultingowej służy między innymi opracowana w firmie „skala kompetencji". Oprócz rozmowy kwalifikacyjnej kandydaci wypełniają kwestionariusz zawierający między innymi pytania tworzące „skalę kompetencji". Skala zawiera 7 pytań. Za każde z pytań można zyskać od 1 do 5 punktów, gdzie 1 oznacza najniższą, a 5 najwyższą możliwą notę. Łącznie uzyskać można 35 punktów. Punkty w tej skali uzyskane przez 24 kandydatów zawiera poniższa tabela:

Lp	KK1	KK2	KK3	KK4	KK5	KK6	KK7	SUMA
1	3	3	5	5	5	5	1	27
2	5	4	4	3	3	5	1	25
3	5	5	3	5	3	2	1	24
4	1	2	5	5	5	5	2	25
5	4	5	5	5	5	5	1	30
6	4	4	5	5	5	5	3	31
7	1	1	5	5	5	5	2	24
8	5	5	5	5	3	5	3	31
9	3	2	2	5	4	2	1	19
10	3	4	3	4	4	2	1	21
11	4	4	3	4	4	4	4	27
12	1	1	3	4	1	1	3	16
13	3	3	4	5	5	5	1	26
14	4	5	5	5	5	5	2	31
15	1	4	4	4	1	4	4	22
16	1	4	5	5	5	5	1	26
17	5	5	5	5	5	5	2	32
18	5	3	5	5	3	5	4	30
19	1	1	2	2	2	1	4	13
20	5	5	5	5	5	5	5	35
21	5	3	5	5	5	5	1	29
22	5	5	5	5	5	1	5	31
23	2	1	5	3	2	4	1	18
24	5	5	5	5	5	5	5	35

Chcemy sprawdzić dokładność zastosowanego narzędzia oceny kompetencji (zastosowanej skali). W tym celu przeprowadzona zostanie analiza rzetelności.

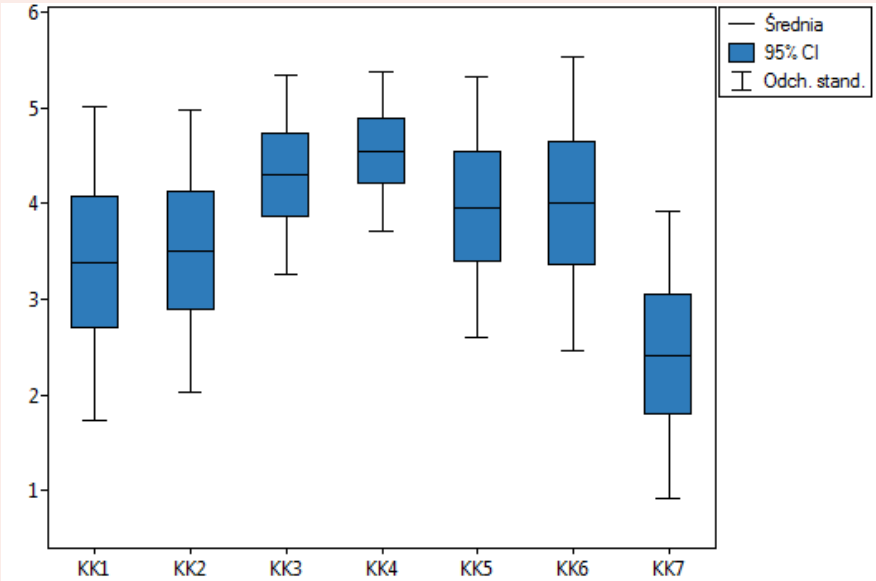
Macierz korelacji wskazuje, że ostatnia pozycja jest najsłabiej skorelowana z pozostałymi. Podejrzewamy zatem, że pozycja ta nie mierzy tego samego konstruktu, co pozostałe pozycje.

Korelacje

	KK2	KK3	KK4	KK5	KK6	KK7
KK1	0.712	0.265	0.355	0.338	0.225	0.164
KK2	.	0.326	0.443	0.378	0.269	0.196
KK3	.	.	0.512	0.498	0.735	0.058
KK4	.	.	.	0.67	0.409	-0.049
KK5	.	.	.	.	0.478	-0.14
KK6	.	.	.	.	.	-0.151

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Usunięta pozycja	KK7
Średnia skali po usunięciu pozycji	23.666667
Odchylenie standardowe skali po usunięciu pozycji	5.73067
Korelacja pozycji usuniętej i sumy pozostałych	0.026954
Alfa Cronbacha skali po usunięciu pozycji	0.803619
Liczność	24
Liczba pozycji	7
Średnia skali	26.083333
Odchylenie standardowe skali	5.96305
Alfa Cronbacha skali	0.736805
Błąd standardowy pomiaru	3.059199
Średnia korelacji pomiędzy parami pozycji	0.31847
Alfa Cronbacha standaryzowana	0.765863



Współczynnik Alfa Cronbacha $= 0.736805$

Skala kompetencji okazała się być skalą rzetelną. Średni współczynnik korelacji pomiędzy pozycjami skali wynosi 0.31847.

Dokładniejsza analiza każdej pozycji wskazuje, że wszystkie one, za wyjątkiem ostatniej pozycji, w podobnym stopniu wpływają na rzetelność całej skali. Korelacja pozycji KK7 z pozostałymi pozycjami skali jest najsłabsza i wynosi 0.026954. Po usunięciu ze skali pozycji KK7 wartość współczynnika Cronbacha wzrosłaby do 0.803619.

Podobne wnioski możemy wysunąć na podstawie analizy rzetelności połówkowej wykonanej na pozycjach losowo podzielonych na 2 części (KK1, KK3, KK5) (KK2, KK4, KK6, KK7).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Alfa Cronbacha/Rzetelność połówkowa	
Czas analizy	0.04 sek.
Analizowane zmienne	KK1,KK3,KK5,KK2,KK4,KK6,KK7
Poziom istotności	0.05
Liczność	24
Średnia skali	26.083333
Odchylenie standardowe skali	5.96305
Współczynnik korelacji połówek	0.750862
Rzetelność połówkowa	0.857705
Błąd standardowy pomiaru	2.24938
Rzetelność połówkowa Guttmana	0.856531
Pierwsza połowa	
Liczba pozycji	3
Nazwy pozycji	KK1,KK3,KK5
Średnia	11.625
Odchylenie standardowe	3.076029
Alfa Cronbacha	0.607122
Druga połowa	
Liczba pozycji	4
Nazwy pozycji	KK2,KK4,KK6,KK7
Średnia	14.458333
Odchylenie standardowe	3.296628
Alfa Cronbacha	0.416958

Rzetelność połówkowa: Spearmana-Browna $= 0.857705$**, Guttmana** $= 0.856531$

Połówki są dość dobrze skorelowane – współczynnik korelacji połówek wynosi 0.750862. Wartość współczynnika Alfa Cronbacha jest jednak zbyt niska dla drugiej połówki (0.416958), ponieważ zawiera ona pozycję KK7, słabo skorelowaną z pozostałymi pozycjami skali.

Po usunięciu tej pozycji i ponownym wykonaniu analizy, wszystkie współczynniki rzetelności rzeczywiście są wyższe.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Alfa Cronbacha/Rzetelność połówkowa	
Czas analizy	0.05 sek.
Analizowane zmienne	KK1,KK3,KK5,KK2,KK4,KK6
Poziom istotności	0.05
Liczność	24
Średnia skali	23.666667
Odchylenie standardowe skali	5.73067
Współczynnik korelacji połówek	0.822933
Rzetelność połówkowa	0.902867
Błąd standardowy pomiaru	1.786032
Rzetelność połówkowa Guttmana	0.902251
Pierwsza połowa	
Liczba pozycji	3
Nazwy pozycji	KK1,KK3,KK5
Średnia	11.625
Odchylenie standardowe	3.076029
Alfa Cronbacha	0.607122
Druga połowa	
Liczba pozycji	3
Nazwy pozycji	KK2,KK4,KK6
Średnia	12.041667
Odchylenie standardowe	2.92633
Alfa Cronbacha	0.586418

DLA ZAINTERESOWANYCH

Wzory

Współczynnik α-Cronbacha:

$$\alpha_C = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} sd_i^2}{sd_t^2}\right),$$

gdzie:
k – liczba pozycji w skali,
sd_i^2 – wariancja pozycji i,
sd_t^2 – wariancja sumy wszystkich pozycji.

Standaryzowany współczynnik rzetelności:

$$\alpha_{standard} = \frac{k\overline{r_p}}{1 + (k-1)\overline{r_p}},$$

gdzie:
$\overline{r_p}$ – średnia współczynników korelacji Pearsona wszystkich ($k(k-1)/2$) pozycji skali.

Rzetelność połówkowa Spearmana-Browna:

$$r_{SH} = \frac{2r_p^*}{1 + r_p^*},$$

gdzie:
r_p^* – współczynnik korelacji Pearsona pomiędzy połówkami skali.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Rzetelność połówkowa Guttmana:

$$r_{SHG} = 2\left(1 - \frac{sd_{t1}^2 + sd_{t2}^2}{sd_t^2}\right),$$

gdzie:
sd_{t1}^2, sd_{t2}^2 – wariancja pierwszej i drugiej połowy skali,
sd_t^2 – wariancja sumy wszystkich pozycji.

Błąd standardowy pomiaru (SEM), wyznaczany dla skali rzetelnej:

$SEM = sd_t\sqrt{1 - \alpha_C}$ – dla współczynnika rzetelności Cronbacha

lub

$SEM = sd_t\sqrt{1 - r_{SH}}$ – dla współczynnika rzetelności połówkowej

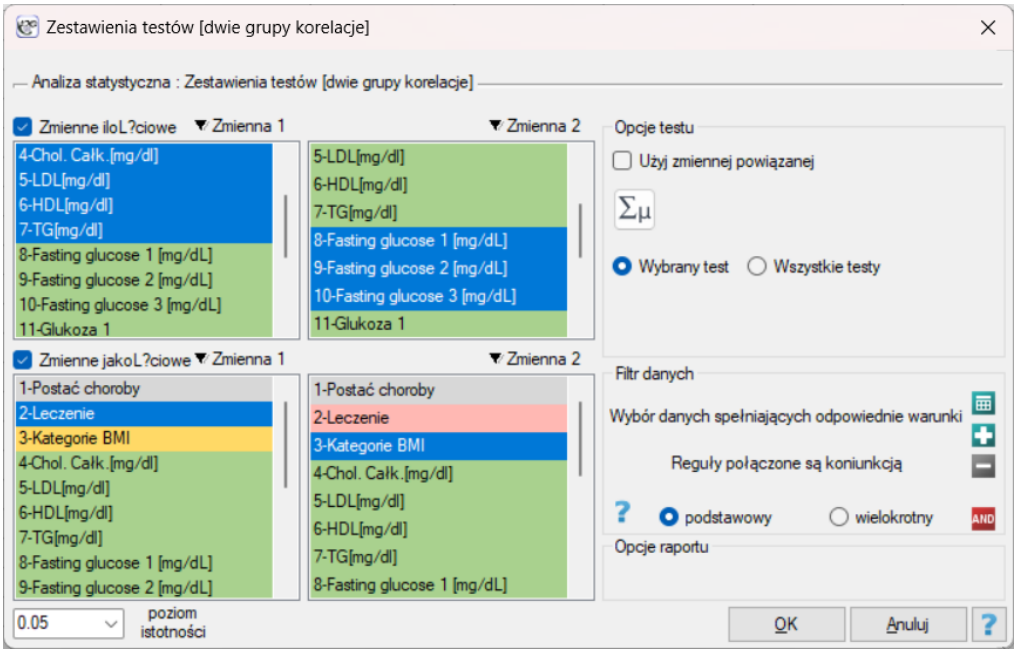
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

31 ZESTAWIENIA TESTÓW

Zamiast wykonywać kolejne testy pojedynczo, program pozwala uruchomić od razu zestaw kilku analiz naraz, dobranych do rodzaju porównywanych grup. Dane ilościowe zostaną przy okazji opisane średnimi, medianami itp., a dane jakościowe – licznościami i procentami.

JAK URUCHOMIĆ ZESTAWIENIE

Okno z ustawieniami opcji Zestawień testów wywołujemy poprzez menu Statystyka→Zestawy→Zestawienia testów, a następnie wybieramy odpowiednią grupę analiz.



Do wyboru mamy cztery grupy zestawień:

1. Porównanie dwóch grup zależnych:
   - Test t-Studenta dla grup zależnych
   - Test Wilcoxona dla grup zależnych
   - Test Bowkera-McNemara
   - Test normalności rozkładu Kołmogorov-Smirnov (lub inny zaproponowany przez użytkownika)
   - i inne ...
2. Porównanie dwóch grup niezależnych:
   - Test t-Studenta dla grup niezależnych
   - Test U Manna-Whitneya
   - Chi-kwadrat, Fisher, OR/RR
   - Test normalności rozkładu Kołmogorov-Smirnov (lub inny zaproponowany przez użytkownika)

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- i inne ...

3. Porównanie wielu grup niezależnych (zob. opis poniżej):
   - ANOVA dla grup niezależnych i ANOVA z korektą Welcha
   - ANOVA Kruskala-Wallisa
   - Chi-kwadrat i test Fishera dla dużych tabel
   - Test Levene'a równości wariancji, test normalności rozkładu (jak wyżej)
   - testy post-hoc (Tukey HSD, Tamhane T2, Dunn, porównania par tabel) i grupy jednorodne
   - i inne ...
4. Korelację:
   - Zależność liniowa Pearsona
   - Zależność monotoniczna Spearmana
   - Chi-kwadrat, Fisher, wsp. korelacji
   - Test normalności rozkładu Kołmogorov-Smirnov (lub inny zaproponowany przez użytkownika)
   - i inne ...

CO ZWRACA PROGRAM

Dla każdej analizowanej zmiennej, w zależności od tego czy jest ona ilościowa czy jakościowa, program może zwrócić w raporcie wyniki:

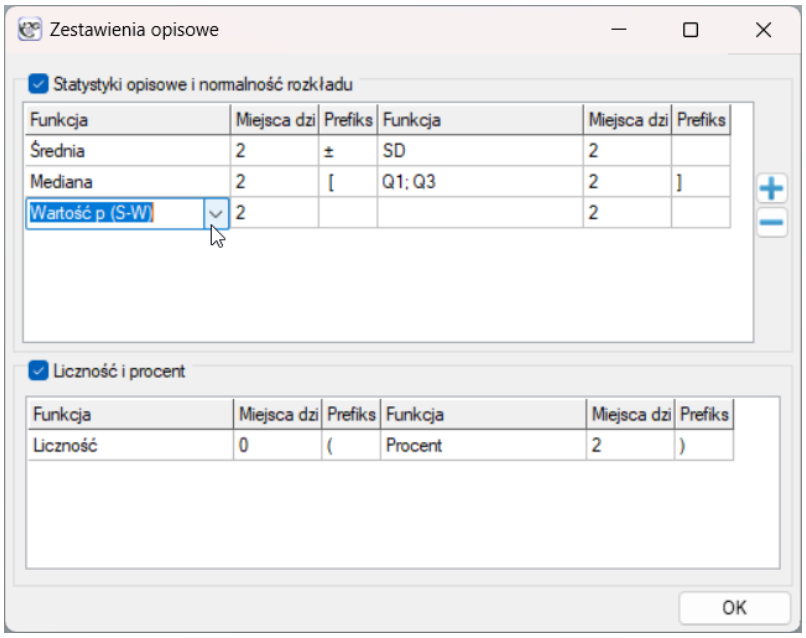
- **wybranego testu** – dobranego automatycznie wg reguły opisanej poniżej;
- **wszystkich testów** i towarzyszących im współczynników oraz miar, niezależnie od tego, czy spełnione są minimalne warunki ich stosowania.

JAK PROGRAM DOBIERA TEST AUTOMATYCZNIE

– Jeśli przed analizą użytkownik nie opisze poszczególnych zmiennych odpowiednimi skalami, dane ilościowe są traktowane jak skala interwałowa, a dane jakościowe – jak skala nominalna.

– Badanie normalności rozkładu opiera się na wyniku testu normalności wybranego przez użytkownika podczas ustawiania statystyk opisowych.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



– Jeśli użytkownik nie wskaże żadnego testu oceniającego normalność rozkładu w oknie tych statystyk, normalność zostanie sprawdzona testem Kołmogorova-Smirnowa.

Uwaga! Analizy, które proponujemy, są odporne na niewielkie odstępstwa od rozkładu normalnego, a test Kołmogorova-Smirnowa jest najbardziej konserwatywny spośród dostępnych testów normalności – brak normalności rozkładu wykazujemy więc dopiero przy dużej różnicy badanego rozkładu od rozkładu normalnego. Normalność rozkładu badamy: (1) dla porównania dwóch grup niezależnych – w oparciu o dane w każdej grupie, (2) dla porównania dwóch grup zależnych – w oparciu o różnicę pomiarów, (3) dla korelacji – w oparciu o reszty modelu, (4) dla porównania wielu grup niezależnych – w oparciu o dane w każdej grupie (warunek musi być spełniony we wszystkich grupach).

WYKONANIE I INTERPRETACJA

Przykład 31.1. (plik zestawienia.pqs)
Chcemy dokonać automatycznego porównania dwóch niezależnych grup: chorych przewlekle z chorującymi na ostrą postać choroby. Porównania dokonujemy w oparciu o dane opisane na skali interwałowej: Chol Całk., LDL, HDL, TG, w oparciu o cechę porządkową BMI, oraz o dane nominalne: Leczenie, Estradiol 1, Estradiol 2, Estradiol 3. W tym celu wybieramy menu Zestawy → Zestawienia testów [dwie grupy niezależne] i zaznaczamy zmienną grupującą: Leczenie, następnie wybieramy dane interwałowe (w części dotyczącej zmiennych ilościowych) oraz dane nominalne i porządkowe (w części dotyczącej zmiennych jakościowych). Zaznaczamy opcję Wybrany test i wykonujemy analizę.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Note

Zestawienia testów [dwie grupy niezależne]

Zmienna	Miary	Postać choroby[ostra]	Postać choroby[przewlekła	Wartość p	test
Chol. Całk.[mg/dl]				0.5433	(M-W)
	Średnia ± SD	160.12±23.77	160.82±30.31		
	Mediana [Q1; Q3]	165.2 [135.38; 178.53]	158.5 [141.25; 170.33]		
	Wartość p (S-W)	0.14	<0.01		
LDL[mg/dl]				0.2241	(M-W)
	Średnia ± SD	86.83±20	82.91±27.81		
	Mediana [Q1; Q3]	91.35 [69.55; 102.35]	79.75 [64.83; 94.33]		
	Wartość p (S-W)	0.31	0.01		
HDL[mg/dl]				0.0339	(t-st)
	Średnia ± SD	53.27±9.7	58.99±9.97		
	Mediana [Q1; Q3]	52.5 [48.28; 56.7]	58 [52.88; 67.55]		
	Wartość p (S-W)	0.52	0.8		
TG[mg/dl]				0.3557	(M-W)
	Średnia ± SD	103.25±40.09	93.36±35.59		
	Mediana [Q1; Q3]	95.45 [73.48; 119.35]	86.8 [64.65; 116.18]		
	Wartość p (S-W)	0.08	0.05		

Zmienna	Kategorie	Postać choroby[ostra]	Postać choroby[przewlekła	Wartość p	test
Leczenie				0.1583	(chi2)
	Lek A	19(79.17%)	21(61.76%)		
	lek B	5(20.83%)	13(38.24%)		
Kategorie BMI				<0.0001	(chi2 trendu)
	norma	6(25%)	32(94.12%)		
	nadwaga	8(33.33%)	0(0%)		
	otyłość	10(41.67%)	2(5.88%)		
Estradiol 1				0.0077	(chi2)
	norma	7(29.17%)	22(64.71%)		
	powyżej normy	17(70.83%)	12(35.29%)		
Estradiol 2				0.6822	(chi2 Yates)
	norma	22(91.67%)	31(91.18%)		
	powyżej normy	2(8.33%)	3(8.82%)		
Estradiol 3				0.988	(chi2 Yates)
	norma	21(87.5%)	31(91.18%)		
	powyżej normy	3(12.5%)	3(8.82%)		

W efekcie otrzymujemy zarówno opis każdej grupy, jak i dobrany do porównania test statystyczny.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zasada doboru testu – z Notatki raportu

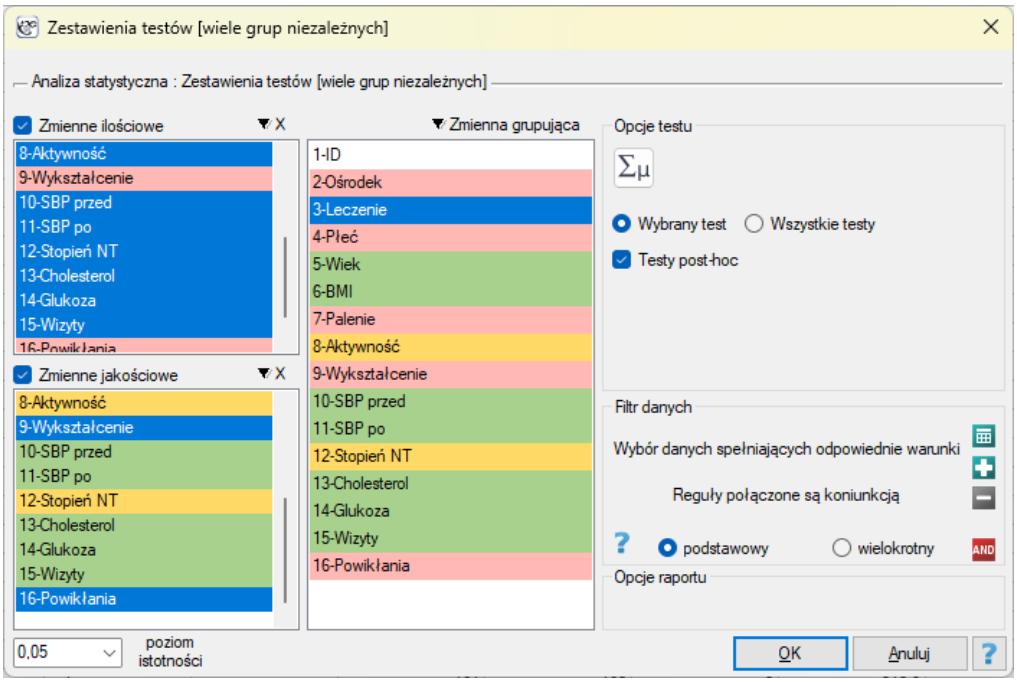
W Notatce, znajdującej się na górze raportu, umieszczono opis informujący o zasadzie doboru testu statystycznego odpowiedniego dla danej analizy:

”**Zmienne ilościowe**:
Dla skali interwałowej z normalnością rozkładu wyznaczono test t-studenta (t-st) lub jego poprawkę Cochrana-Coxa (C-C), gdy wariancje grup różniły się. Dla skali interwałowej, gdy warunek normalności rozkładu nie był spełniony, podobnie jak dla skali porządkowej wyznaczono test Manna-Whitneya (M-W). Normalność rozkładu danych badano testem Shapiro-Wilka a równość wariancji testem Fishera-Snedecora (F-S). Jeśli nie oznaczono skali dla analizowanych zmiennych, przyjęto, że dane pochodzą ze skali interwałowej.

Zmienne jakościowe:
Dla skali nominalnej wyznaczono test chi-kwadrat (chi2), a gdy warunek Cochrana nie był spełniony test dokładny Fishera (Fisher dokładny) lub, w przypadku tabel 2x2 o rozmiarze próby większym niż 40, poprawkę Yatesa (chi2-Yates). Dla skali porządkowej wyznaczono test chi-kwadrat dla trendu. Jeśli nie oznaczono skali dla analizowanych zmiennych, przyjęto, że dane pochodzą ze skali nominalnej.”

ZESTAWIENIE DLA WIELU GRUP NIEZALEŻNYCH

Zestawienie Statystyka→Zestawy→Zestawienia testów [wiele grup niezależnych] porównuje jednocześnie dwie lub więcej grup wyznaczonych przez zmienną grupującą (np. kilka ośrodków, kilka schematów leczenia, kategorie wieku). Dla dwóch grup daje wyniki zgodne z zestawieniem dla dwóch grup niezależnych, jednak testami przeznaczonymi dla wielu grup (ANOVA zamiast testu t-Studenta itd.), dlatego przy dokładnie dwóch grupach wygodniej jest korzystać z zestawienia dla dwóch grup.



W oknie zestawienia wskazujemy zmienną grupującą oraz – tak jak w pozostałych zestawieniach –

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

zmienne ilościowe i jakościowe, a następnie wybieramy:

- Wybrany test albo Wszystkie testy – w drugim przypadku dla zmiennych ilościowych raport zawiera kolumny z wynikiem testu normalności w grupach, testów Levene'a i Browna-Forsythe'a, ANOVA, ANOVA Welcha, ANOVA F^* Browna-Forsythe'a, ANOVA Kruskala-Wallisa, testu Conovera równości wariancji rang oraz miary wielkości efektu η^2 i ϵ^2; dla zmiennych jakościowych – wyniki testu chi-kwadrat i dokładnego Fishera dla tabel $r \times c$ oraz, dla skali porządkowej, ANOVA Kruskala-Wallisa dla kodów kategorii;
- Testy post-hoc – po zaznaczeniu tej opcji dla zmiennych, dla których test główny okazał się istotny, a grup jest co najmniej trzy, wykonywane są porównania wielokrotne: test Tukeya HSD po ANOVA, test T2 Tamhane po ANOVA Welcha, test Dunna z korektą Benjaminiego-Hochberga po ANOVA Kruskala-Wallisa, a dla zmiennych jakościowych porównania par grup testem chi-kwadrat lub dokładnym Fishera z korektą Benjaminiego-Hochberga. W kolumnie post-hoc raport wypisuje pary grup różniące się istotnie, a w kolumnach grup jednorodnych (GJ) grupy, które nie różnią się istotnie, otrzymują wspólną literę.

Miary opisowe dla każdej grupy zawierają domyślnie także liczność N; podobnie jak w innych zestawieniach ich zakres ustawiamy w oknie statystyk opisowych.

PRZYKŁAD 31.2. (plik zestawienia_wiele_grup.pqs)
Plik zawiera dane 1000 pacjentów z nadciśnieniem leczonych w czterech ośrodkach trzema schematami: Standard, Nowy lek i Placebo. Chcemy sprawdzić, czy schematy leczenia różnią się pod względem cech pacjentów i wyników leczenia. Porównania dokonujemy w oparciu o dane ilościowe: Wiek, BMI, SBP przed, SBP po (ciśnienie skurczowe przed leczeniem i po nim), Cholesterol, Glukoza i Wizyty (liczba wizyt), o cechy porządkowe: Aktywność i Stopień NT, oraz o dane nominalne: Płeć, Palenie i Powikłania. Skale zmiennych porządkowych i nominalnych są w pliku już ustawione.
W tym celu wybieramy menu Statystyka→Zestawy→Zestawienia testów [wiele grup niezależnych], jako zmienną grupującą wskazujemy Leczenie, w części dotyczącej zmiennych ilościowych wybieramy wymienione zmienne interwałowe, a w części dotyczącej zmiennych jakościowych – zmienne porządkowe i nominalne. Zaznaczamy opcje Wybrany test i Testy post-hoc i wykonujemy analizę.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Note

Zestawienia testów [wiele grup niezależnych]

Zmienna	Miary	Leczenie[Nowy lek	Leczenie[Placebo]	Leczenie[Stand	Wartość p	test	post-hoc	GJ Nowy lek	GJ Placebo	GJ Standard
Wiek					0,421784	(K-W)	brak	A	A	A
	Średnia ± SD	56,38±10,77	56,41±10,44	55,37±11,19						
	Mediana [Q1; Q3]	56 [49; 64]	57 [49; 64]	56 [47; 63]						
	Wartość p (S-W)	0,41	0,07	0,01						
	N	333	295	372						
BMI					0,427722	(K-W)	brak	A	A	A
	Średnia ± SD	27,91±3,94	28,26±4,34	28,09±4,11						
	Mediana [Q1; Q3]	28,1 [25,4; 30,6]	28,7 [25,1; 31,4]	27,9 [25,3; 31,						
	Wartość p (S-W)	0,42	0,03	0,12						
	N	329	287	364						
Aktywność					NA	(K-W)	brak	A	A	A
	Średnia ± SD	NA±NA	NA±NA	NA±NA						
	Mediana [Q1; Q3]	NA [NA; NA]	NA [NA; NA]	NA [NA; NA]						
	Wartość p (S-W)	NA	NA	NA						
	N	0	0	0						
SBP przed					0,516973	(K-W)	brak	A	A	A
	Średnia ± SD	152,28±13,38	152,44±13,95	153,05±13,2						
	Mediana [Q1; Q3]	152 [143; 161]	153 [143; 162]	154 [144; 162]						
	Wartość p (S-W)	0,05	0,01	0,02						
	N	333	295	372						
SBP po					<0,000001	(K-W)	Nowy lek-P	A	B	C
	Średnia ± SD	135,61±16,93	149,46±16,86	142,83±15,8						
	Mediana [Q1; Q3]	134 [123; 146]	149 [138; 161]	144 [132; 153]						
	Wartość p (S-W)	0,02	0,57	0,28						
	N	333	295	372						
Stopień NT					0,629838	(K-W)	brak	A	A	A
	Średnia ± SD	1,32±0,52	1,32±0,53	1,35±0,53						
	Mediana [Q1; Q3]	1 [1; 2]	1 [1; 2]	1 [1; 2]						
	Wartość p (S-W)	<0,01	<0,01	<0,01						
	N	333	295	372						
Cholesterol					0,076908	(ANOVA)	brak	A	A	A
	Średnia ± SD	206,64±33,75	202,16±31,3	207,75±31,96						
	Mediana [Q1; Q3]	207,6 [181,45; 22	201,3 [182,8; 223,9]	209 [183,15; 2						
	Wartość p (S-W)	0,21	0,81	0,17						
	N	320	287	359						
Glukoza					0,13009	(K-W)	brak	A	A	A
	Średnia ± SD	100,34±16,97	100,97±16,66	103,42±19,79						
	Mediana [Q1; Q3]	98,8 [87,9; 109,4	99,1 [88,6; 112,75]	101,1 [89,4; 11						
	Wartość p (S-W)	<0,01	<0,01	<0,01						
	N	326	291	365						
Wizyty					0,081599	(K-W)	brak	A	A	A
	Średnia ± SD	3,06±1,74	2,98±1,55	3,24±1,72						
	Mediana [Q1; Q3]	3 [2; 4]	3 [2; 4]	3 [2; 4]						
	Wartość p (S-W)	<0,01	<0,01	<0,01						
	N	333	295	372						

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zmienna	Kategorie	Leczenie[Nowy lek	Leczenie[Placebo]	Leczenie[Stand	Wartość p	test	post-hoc	GJ Nowy lek	GJ Placebo	GJ Standard
Ośrodek					0,30118	(chi2)	brak	A	A	A
	Zachód	68(20,42%)	67(22,71%)	93(25%)						
	Wschód	91(27,33%)	82(27,8%)	95(25,54%)						
	Północ	98(29,43%)	69(23,39%)	109(29,3%)						
	Południe	76(22,82%)	77(26,1%)	75(20,16%)						
Płeć					0,509218	(chi2)	brak	A	A	A
	M	164(49,25%)	132(44,75%)	172(46,24%)						
	K	169(50,75%)	163(55,25%)	200(53,76%)						
Palenie					0,712116	(chi2)	brak	A	A	A
	tak	96(28,83%)	82(27,8%)	114(30,65%)						
	nie	237(71,17%)	213(72,2%)	258(69,35%)						
Wykształceni					0,44659	(chi2)	brak	A	A	A
	wyższe	91(27,33%)	86(29,15%)	104(27,96%)						
	średnie	150(45,05%)	142(48,14%)	186(50%)						
	podstawowe	92(27,63%)	67(22,71%)	82(22,04%)						
Powikłania					0,03746	(chi2)	Placebo-St	A+B	A	B
	tak	32(9,61%)	48(16,27%)	44(11,83%)						
	nie	301(90,39%)	247(83,73%)	328(88,17%)						

W raporcie dla każdej zmiennej otrzymujemy opis trzech grup (liczność, średnią z odchyleniem standardowym lub medianę z kwartylami, a dla zmiennych jakościowych liczności i procenty), nazwę dobranego testu i jego wartość p. Dla zmiennych o rozkładzie zbliżonym do normalnego w każdej grupie (np. Wiek, BMI, SBP przed) dobierana jest ANOVA (lub ANOVA Welcha, gdy test Levene'a wykaże różnicę wariancji), dla zmiennych o rozkładzie skośnym (Glukoza, Wizyty) oraz dla zmiennych porządkowych – ANOVA Kruskala-Wallisa, a dla zmiennych nominalnych – test chi-kwadrat. Ponieważ pacjentów przydzielono do schematów losowo, cechy wyjściowe (Wiek, BMI, SBP przed, Płeć, Palenie) nie powinny różnić się istotnie między grupami; różnic spodziewamy się natomiast w ciśnieniu po leczeniu (SBP po) i w częstości powikłań. Dla zmiennych z istotnym wynikiem kolumna post-hoc wskazuje, które pary schematów różnią się istotnie (dla SBP po – wszystkie trzy pary, z najniższym ciśnieniem po nowym leku), a kolumny grup jednorodnych (GJ) oznaczają wspólną literą schematy, których nie da się odróżnić.
Ten sam plik pozwala wykonać zestawienie dla dwóch grup niezależnych (zmienna grupująca Płeć lub Palenie) oraz zestawienie dla wielu grup według ośrodka (Ośrodek, cztery grupy) lub stopnia nadciśnienia (Stopień NT, trzy grupy).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Zasada doboru testu dla wielu grup – z Notatki raportu

W Notatce na górze raportu program opisuje przyjętą regułę doboru testu:

”**Zmienne ilościowe**:
Dla skali interwałowej z normalnością rozkładu we wszystkich grupach i równością wariancji wyznaczono jednoczynnikową ANOVA dla grup niezależnych (ANOVA), a gdy wariancje grup różniły się – ANOVA z korektą Welcha (Welch). Dla skali interwałowej, gdy warunek normalności rozkładu nie był spełniony, podobnie jak dla skali porządkowej wyznaczono ANOVA Kruskala-Wallisa (K-W). Normalność rozkładu badano testem Shapiro-Wilka, a równość wariancji testem Levene’a. Jeśli nie oznaczono skali dla analizowanych zmiennych, przyjęto, że dane pochodzą ze skali interwałowej.

Zmienne jakościowe:
Dla skali nominalnej wyznaczono test chi-kwadrat (chi2), a gdy warunek Cochrana nie był spełniony test dokładny Fishera (Fisher dokładny) lub, w przypadku tabel 2x2 o rozmiarze próby większym niż 40, poprawkę Yatesa (chi2-Yates). Dla skali porządkowej wyznaczono ANOVA Kruskala-Wallisa (K-W) dla kodów kategorii oraz test Conovera równości wariancji rang. Jeśli nie oznaczono skali dla analizowanych zmiennych, przyjęto, że dane pochodzą ze skali nominalnej.

Testy post-hoc:
Wyznaczono tylko dla co najmniej 3 grup i istotnego testu głównego: Tukey HSD (ANOVA), Tamhane T2 (Welch), Dunn z korektą Benjaminiego-Hochberga (K-W), porównania par grup testem chi-kwadrat lub dokładnym Fishera z korektą Benjaminiego-Hochberga (tabele RxC). W kolumnie post-hoc wypisano pary grup różniące się istotnie, a w kolumnach grup jednorodnych (GJ) grupy nieróżniące się istotnie mają wspólną literę.”

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

32 KREATOR

Kreator jest narzędziem ułatwiającym początkującym użytkownikom poruszanie się po podstawowych statystykach zawartych w programie. Zawiera podpowiedzi dotyczące warunków, jakie należy sprawdzić przed wybraniem odpowiedniego testu statystycznego. Końcowym krokiem kreatora jest wybór odpowiedniego testu statystycznego i uruchomienie okna opcji tego testu.

Kreator można uruchomić poprzez:
- Statystyka→Kreator,
- przycisk  na pasku narzędzi.

Uruchomione okno kreatora zawiera pole wyboru rodzaju analizy, którą użytkownik chce przeprowadzić. Użytkownik może wybrać:

Porównanie - 1 grupa - by porównać wartości pomiarów dla 1 populacji z zadaną przez siebie wartością. Populację tę reprezentują dane zebrane w 1 kolumnie w postaci surowej lub skumulowane do tabeli liczności.

Porównanie - 2 grupy - by porównać między sobą wartości pomiarów dla 2 populacji. Populacje te reprezentowane są przez dane zebrane w postaci surowej w 2 kolumnach lub skumulowane do tabeli kontyngencji.

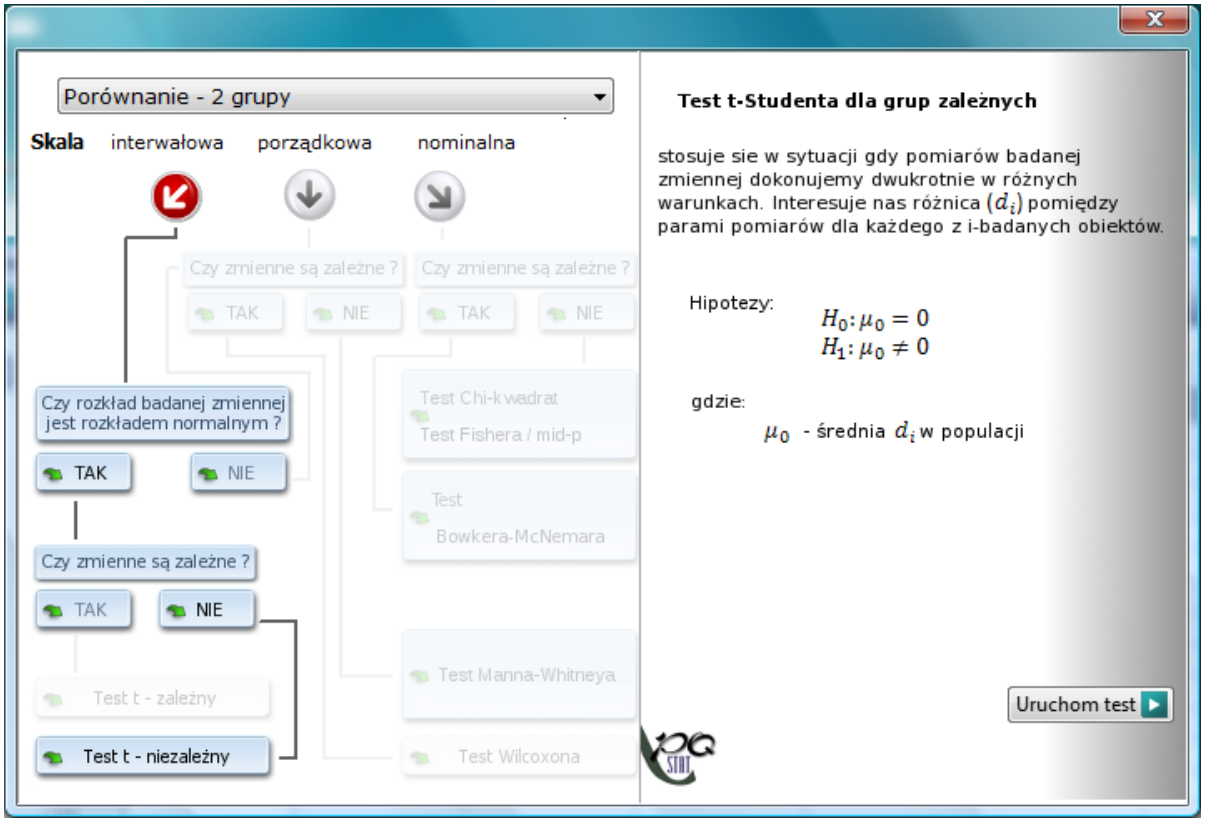
Porównanie - więcej niż 2 grupy - by porównać między sobą wartości pomiarów dla kilku populacji. Populacje te reprezentowane są przez dane zebrane w postaci surowej w kilku kolumnach.

Korelacja - by sprawdzić występowanie zależności między 2 parametrami populacji. Cechy te reprezentowane są przez dane zebrane w postaci surowej w 2 kolumnach lub skumulowane do tabeli kontyngencji.

Zgodność - by zbadać zgodność kolejnych kilku uzyskanych wyników. Cechy te reprezentowane są przez dane zebrane w postaci surowej w kilku kolumnach lub skumulowane do tabeli kontyngencji.

Po wybraniu rodzaju analizy uruchamia się graf podzielony zgodnie ze skalą, na jakiej dokonano pomiaru badanych cech populacji (skala interwałowa, porządkowa, nominalna).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Użytkownik przemieszcza się po grafie, wybierając prawidłową odpowiedź na postawione pytania. Po przejściu wybranej przez siebie ścieżki grafu ma możliwość wykonania testu, który zgodnie z udzielonymi odpowiedziami jest odpowiedni do rozwiązania stojącego przed nim problemu statystycznego.

33 INNE UWAGI

33.1 FORMATY PLIKÓW

PQS - domyślny format zapisu danych dla plików ℙℚStat; reprezentuje wszystkie obiekty utworzone za pomocą ℙℚStat.

PQX - format plików ℙℚStat w standardzie XML (kodowanie UTF-8); może zostać użyty do reprezentacji wszystkich obiektów w programie ℙℚStat; zalecany na systemach z małą ilością pamięci operacyjnej.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Literatura

[1] Arthur D., Vassilvitskii S. (2007). *k-means++: the advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics Philadelphia, PA, USA. 1027–1035*

[2] Abdi H. (2007), *Bonferroni and Sidak corrections for multiple comparisons", in N.J. Salkind (ed.): Encyclopedia of Measurement and Statistics. Thousand Oaks, CA: Sage*

[3] D'Agostino R.B. and Pearson E.S. (1973), *Tests of departure from normality. Empirical results for the distribution of b2 and sqrt(b1). Biometrika, 60, 613-622*

[4] D'Agostino R.B., Belanger A., D'Agostino Jr.R B. (1990), *A suggestion for using powerful and informative tests of normality. American Statistician, 44, 3 16-321*

[5] Agresti A., Coull B.A. (1998), *Approximate is better than "exact" for interval estimation of binomial proportions. American Statistics 52: 119-126*

[6] Altman D.G., (1998), *Confidence intervals for the number needed to treat. BMJ. 317(7168): 1309–1312*

[7] Aroian, L. A. (1947), *The probability function of the product of two normally distributed variables. Annals of Mathematical Statistics, 18, 265-271.*

[8] Bland J.M., Altman D.G. (1999), *Measuring agreement in method comparison studies. Statistical Methods in Medical Research 8:135-160.*

[9] Anscombe F.J. (1981), *Computing in Statistical Science through APL. Springer-Verlag, New York*

[10] Armitage P. (1955), *Tests for Linear Trends in Proportions and Frequencies. Biometrics. 11 (3): 375–386*

[11] Armitage P., Berry G., (1994), *Statistical Methods in Medical Research (3rd edition); Blackwell*

[12] Armitage P., Colton T., (2009), *Encyclopedia of Biostatistics. John Wiley and Sons.*

[13] Austin P.C., (2009), *The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Med Decis Making; 29(6):661-77*

[14] Austin P.C., (2011), *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research 46, 3: 399–424*

[15] Barnard G.A. (1989), *On alleged gains in power from lower p-values. Statistics in Medicine 8:1469-1477*

[16] Baron R. M., Kenny D. A. (1986), *The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. Journal of Personality and Social Psychology, 51, 1173-1182.*

[17] Beal S.L. (1987), *Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. Biometrics 43: 941-950*

[18] Bender R. (2001), *Calculating confidence intervals for the number needed to treat. Controlled Clinical Trials 22:102–110*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[19] Benjamini Y. and Hochberg Y. (1995), *Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57, 289–300*

[20] Betty R. Kirkwood and Jonathan A. C. Sterne (2003), *Medical Statistics (2nd ed.). Meassachusetts: Blackwell Science, 177—188, 240—248*

[21] Blanche P., Dartigues J.F., Jacqmin-Gadda H. (2013), *Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Statistics in Medicine, 32(30):5381–5397*

[22] Bland J.M., Altman D.G. (1986), *Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327 (8476): 307–10*

[23] Bland J.M., Altman D.G. (1999), *Measuring agreement in method comparison studies. Statistical Methods in Medical Research 8 (2): 135–60*

[24] Bowker A.H. (1948), *Test for symmetry in contingency tables. Journal of the American Statistical Association, 43, 572-574*

[25] Box G. E. , Cox D. R. (1964), *An analysis of transformations. Journal of the Royal Statistical Society, Series B 26: 211–252*

[26] Breslow N.E., Day N.E. (1980), *Statistical Methods in Cancer Research: Vol. I - The Analysis of Case-Control Studies. Lyon: International Agency for Research on Cancer*

[27] Breslow N.E. (1996), *Statistics in epidemiology: the case-control study', Journal of the American Statistical Association, 91, 14—28*

[28] Breslow N.E. (1974), *Covariance analysis of censored survival data. Biometrics, 30(1):89–99*

[29] Brookmeyer R. and Crowley J. (1982a), *A confidence interval for the median survival time. Biometrics 38, 29-41*

[30] Brown L.D., Cai T.T., DasGupta A. (2001), *Interval Estimation for a Binomial Proportion. Statistical Science, Vol. 16, no. 2, 101-133*

[31] Brown M.B., Forsythe A. B. (1974a), *Robust tests for equality of variances. Journal of the American Statistical Association, 69,364-367*

[32] Brown M. B., Forsythe A. B. (1974), *The ANOVA and multiple comparisons for data with heterogeneous variances. Biometrics, 30, 719-724*

[33] Brown M. B., Forsythe A. B. (1974), *The small sample behavior of some statistics which test the equality of several means. Technometrics, 16, 385-389*

[34] Brown W. (1910), *Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296-322*

[35] Hochberg Y. (1988), *A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800–803*

[36] Chow S.C., Shao J., and Wang H. (2008). *Sample Size Calculations in Clinical Research, Second Edition. Chapman and Hall/CRC. Boca Raton, Florida.*

[37] Cicchetti D. and Allison T. (1971),*A new procedure for assessing reliability of scoring eeg sleep recordings. American Journal EEG Technology, 11, 101-109*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[38] Cleveland, W. S. (1979),*Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association. 74*

[39] Clopper C. and Pearson S. (1934), *The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26: 404-413*

[40] Cochran W.G. (1950), *The comparison ofpercentages in matched samples. Biometrika, 37, 256-266*

[41] Cochran W.G. (1952), *The chi-square goodness-of-fit test. Annals of Mathematical Statistics, 23, 315-345*,

[42] Cochran W. G. (1954), *Some methods for strengthening the common chi-square tests. Biometrics, 10(4) 17-45 1*

[43] Cochran W.G. and Cox G.M. (1957), *Experimental designs (2nd 4.). New York: John Wiley and Sons.*

[44] Cohen J. (1960), *A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 10,3746*

[45] Cohen J. (1968), *Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 70, 213-220*

[46] Cohen J. (1988), *Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Hillsdale, New Jersey*

[47] Conover W. J. (1999), *Practical nonparametric statistics (3rd ed). John Wiley and Sons, New York*

[48] Cox D.R. (1972), *Regression models and life tables. Journal of the Royal Statistical Society, B34:187-220*

[49] Cramkr H. (1946), *Mathematical models of statistics. Princeton, NJ: Princeton University Press.*

[50] Cronbach L.J. (1951), *Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334*

[51] DeLong E.R., DeLong D.M., Clarke-Pearson D.L., (1988), *Comparing the areas under two or more correlated receiver operating curves: A nonparametric approach. Biometrics 44:837-845*

[52] Dunn O. J. (1964), *Multiple comparisons using rank sums. Technometrics, 6: 241–252*

[53] Durbin J. (1951), *Incomplete blocks in ranking experiments. British Journal of Statistical Psychology, 4: 85–90*

[54] Egger M., Smith G. D., Schneider M., Minder C. (1997), *Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109):629-634*

[55] Epps T.W., Pulley L.B. (1983), *A test for normality based on the empirical characteristic function. Biometrika. 1983;70:723–726*

[56] Fagerland M. W., Lydersen S., and Laake P. (2013), *The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional, BMC Med Res Methodol; 13: 91.*

[57] Fisher R.A. (1934), *Statistical methods for research workers (5th ed.). Edinburgh: Oliver and Boyd*

[58] Fisher R.A. (1935), *The logic of inductive inference. Journal of the Royal Statistical Society, Series A, 98,39-54*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[59] Fisher R.A. (1936), *The use of multiple measurements in taxonomic problems. Annals of Eugenics 7 (2): 179–188*

[60] Fleiss J.L. (1971), *Measuring nominal scale agreement among many raters. Psychological Bulletin, 76 (5): 378–382*

[61] Fleiss J.L., Cohen J. (1973), *The equivalence of weighted kappa and the intraclass correlation coeffcient as measure of reliability. Educational and Psychological Measurement, 33, 613-619*

[62] Fleiss J.L., Levin B., Paik M.C. (2003), *Statistical methods for rates and proportions. 3rd ed. (New York: John Wiley) 598-626*

[63] Fine J.P., Gray R.J. (1999), *A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association, 94(446):496–509*

[64] Freeman G.H. and Halton J.H. (1951), *Note on an exact treatment of contingency, goodness of fit and other problems of significance. Biometrika 38:141-149*

[65] Freireich E.O., Gehan E., Frei E., Schroeder L.R., Wolman I.J., et al. (1963), *The effect of 6-mercaptopmine on the duration of steroid induced remission in acute leukemia. Blood, 21: 699–716*

[66] Friedman M. (1937), *The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32,675-701*

[67] Fritz C.O., Morris P.E., Richler J.J.(2012),*Effect size estimates: Current use, calculations, and interpretation. Journal of Experimental Psychology: General., 141(1):2–18.*

[68] Games P. A., Howell J. F. (1976), *Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. Journal of Educational Statistics, 1, 113-125*

[69] Gehan E. A. (1965a), *A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. Biometrika, 52:203—223*

[70] Gehan E. A. (1965b), *A Generalized Two-Sample Wilcoxon Test for Doubly-Censored Data. Biometrika, 52:650—653*

[71] Goodman L. A. (1960), *On the exact variance of products. Journal of the American Statistical Association, 55, 708-713*

[72] Greenhouse S. W., Geisser S. (1959), *On methods in the analysis of profile data. Psychometrika, 24, 95–112*

[73] Green S.B. (1991), *How many subjects does it take to do a regression analysis? Multivariate Behavioral Research, 26, 499-510*

[74] Guttman L. (1945), *A basic for analyzing test-retest reliabilit. Psychometrika, 10, 255-282*

[75] Hanley J.A. i Hajian-Tilaki K.O. (1997), *Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update. Academic radiology 4(1):49-58*

[76] Hanley J.A. i McNeil M.D. (1982), *The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29-36*

[77] Hanley J.A. i McNeil M.D. (1983), *A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148: 839-843*

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

[78] Hanusz Z., Tarasińska J. (2014), *On multivariate normality tests using skewness and kurtosis, Colloquium Biometricum 44, 139-148*

[79] Harrell F.E., Lee K.L., Mark D.B. (1996), *Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 15: 361-387*

[80] Henderson, R. (1916), *Note on graduation by adjusted average.Transactionsof the Actuarial Society of America, 17:43–48*

[81] Henze N., Zirkler B. (1990), *A class of invariant consistent tests for multivariate normality. Comm. Statist. Theory Methods. 1990;19:3595–3617*

[82] Heagerty P.J., Lumley T., Pepe M.S. (2000), *Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics, 56(2):337–344*

[83] Hochberg Y. (1988), *A Sharper Bonferroni Procedure for Multiple Tests of Significance. Biometrika 75 (4): 800–802*

[84] Holm S. (1979), *A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70*

[85] Hotelling H. (1931), *The generalization of Student's ratio. Annals of Mathematical Statistics 2 (3): 360–378*

[86] Hotelling, H. (1947), *Multivariate Quality Control. In C. Eisenhart, M. W. Hastay, and W. A. Wallis, eds. Techniques of Statistical Analysis. New York: McGraw-Hill*

[87] Hotelling H. (1951), *A generalized t 2 test and measurement of multivariate dispersion. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability 1: 23–41*

[88] Huynh H., Feldt L. S. (1976), *Estimation of the Box correction for degrees of freedom from sample data in randomized block and split=plot designs. Journal of Educational Statistics, 1, 69–82*

[89] Iman R. L., Davenport J. M. (1980), *Approximations of the critical region of the friedman statistic, Communications in Statistics 9, 571–595*

[90] Jarque C. M., Bera A. K., (1987)., *A test for normality of Observations and Regression Residuals, International Statistical Review 55, 163-172*

[91] Jones M. C., Marron J. S., Sheather S. J., (1996)., *A brief survey of bandwidth selection for density estimation. J. Amer. Statist. Assoc. 91 401–407*

[92] Jonckheere A. R. (1954), *A distribution-free k-sample test against ordered alternatives. Biometrika, 41: 133–145*

[93] Kaplan E.L., Meier P. (1958), *Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53:457-481*

[94] Kendall M.G. (1938), *A new measure of rank correlation. Biometrika, 30, 81-93.*

[95] Kendall M.G., Babington-Smith B. (1939), *The problem of m rankings. Annals of Mathematical Statistics, 10, 275-287*

[96] Kleinbaum D. G., Klein M. (2005), *Survival Analysis: A Self-Learning Text, Second Edition (Statistics for Biology and Health)*

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

[97] Kolmogorov A.N. (1933), *Sulla deterrninazione empirica di una legge di distribuzione. Giornde1l'Inst. Ital. degli. Art., 4, 89-91*

[98] Kruskal W.H. (1952), *A nonparametric test for the several sample problem. Annals of Mathematical Statistics, 23, 525-540*

[99] Kruskal W.H., Wallis W.A. (1952), *Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47, 583-621*

[100] Lancaster H.O. (1961), *Significance tests in discrete distributions. Journal of the American Statistical Association 56:223-234*

[101] Lawley D. N. (1938), *A generalization of Fisher's z-test. Biometrika 30: 180–187*

[102] Lee E. T., Wang J. W. (2003), *Statistical Methods for Survival Data Analysis, ed. third, Wiley*

[103] Lenth, R. V., (2001), *Some Practical Guidelines for Effective Sample Size Determination. The American Statistician, 55(3), 187-193*

[104] Levene H. (1960), *Robust tests for the equality of variance. In I. Olkin (Ed.) Contributions to probability and statistics (278-292). Palo Alto, CA: Stanford University Press*

[105] Liddell F.D.K. (1983) *Simplified exact analysis of case-referent studies; matched pairs; dichotomous exposure. Journal of Epidemiology and Community Health; 37:82-84.*

[106] Lilliefors H.W. (1967), *On the Kolmogorov-Smimov test for normality with mean and variance unknown. Journal of the American Statistical Association, 62,399-402*

[107] Lilliefors H.W. (1969), *On the Kolmogorov-Smimov test for the exponential distribution with mean unknown. Journal of the American Statistical Association, 64,387-389*

[108] Lilliefors H.W. (1973), *The Kolmogorov-Smimov and other distance tests for the gamma distribution and for the extreme-value distribution when parameters must be estimated. Department of Statistics, George Washington University, unpublished manuscript*

[109] Lloyd S. P. (1982), *Least squares quantization in PCM. IEEE Transactions on Information Theory 28 (2): 129–137*

[110] Lund R.E., Lund J.R. (1983), *Algorithm AS 190, Probabilities and Upper Quantiles for the Studentized Range. Applied Statistics; 34*

[111] Mahalanobis P. C. (1930), *On tests and measures of group divergence. Journal of the Asiatic Society of Bengal 26: 541–588*

[112] Mahalanobis P. C. (1936), *On the generalized distance in statistics. National Institute of Science of India 12: 49–55*

[113] Mann H. and Whitney D. (1947), *On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics, 1 8 , 5 0 4*

[114] Mantel N. and Haenszel W. (1959), *Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22,719-748*

[115] Mantel N. (1963), *Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. J. Am. Statist. Assoc., 58, 690-700*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[116] Mantel N. (1966), *Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration. Cancer Chemotherapy Reports, 50:163—170*

[117] Marascuilo L.A. and McSweeney M. (1977), *Nonparametric and distribution-free method for the social sciences. Monterey, CA: Brooks/Cole Publishing Company*

[118] Mardia K. V. (1970), *Measures of multivariate skewness and kurtosis with applications, Biometrica 57, 519-530*

[119] Mardia K. V. (1974), *Applications of some measuresof multivariate skewness and kurtosis for testing normality and robustness studies, Sankhay B 36, 115-128*

[120] Mauchly J. W. (1940), *Significance test for sphericity of n-variate normal population. Annals of Mathematical Statistics, 11, 204-209.*

[121] McNemar Q. (1947), *Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12, 153-157*

[122] Mehta C.R. and Patel N.R. (1986), *Algorithm 643. FEXACT: A Fortran subroutine for Fisher's exact test on unordered r*c contingency tables. ACM Transactions on Mathematical Software, 12, 154–161*

[123] Miettinen O.S. (1985), *Theoretical Epidemiology: Principles of Occurrence Research in Medicine. John Wiley and Sons, New York*

[124] Miettinen O.S. and Nurminen M. (1985), *Comparative analysis of two rates. Statistics in Medicine 4: 213-226*

[125] Mimar S.F. (2017), *The Mediation Analysis With the Sobel Test and the Percentile Bootstrap, International Journal of Management and Applied Science, Volume-3, Issue-2*

[126] Nadaraya, E. A. (1964), *On Estimating Regression. Theory of Probability and Its Applications. 9 (1): 141–2*

[127] Newcombe R.G. (1998), *Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods. Statistics in Medicine 17: 873-890*

[128] Newman S.C.(2001), *Biostatistical Methods in Epidemiology. 2nd ed. New York: John Wiley*

[129] Normand S.L. T., Landrum M.B., Guadagnoli E., Ayanian J.Z., Ryan T.J., Cleary P.D., McNeil B.J. (2001), *Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: A matched analysis using propensity scores. Journal of Clinical Epidemiology; 54:387–398.*

[130] Ogilvie J. C. (1965), *Paired comparison models with tests for interaction. Biometrics 21(3): 651-64*

[131] Orwin R. G. (1983), *A Fail-SafeN for Effect Size in Meta-Analysis. J Educ Behav Stat, 8(2):157-159*

[132] Oyeyemi G.M. , Adewara A.A., Adebola F.B. and Salau S.I. (2010), *On the Estimation of Power and Sample Size in Test of Independence, Asian Journal of Mathematics and Statistics, 3(3): 139-146*

[133] Page E. B. (1963), *Ordered hypotheses for multiple treatments: A significance test for linear ranks. Journal of the American Statistical Association 58 (301): 216–30*

[134] Peduzzi P., Concato J., Feinstein A.R., Holford T.R. (1995), *Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. Journal of Clinical Epidemiology, 48:1503-1510*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[135] Peduzzi P., Concato J., Kemper E., Holford T.R., Feinstein A.R. (1996), *A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology; 49(12):1373-9*

[136] Pillai K. C. (1955), *Some new test criteria in multivariate analysis. Annals of Mathematical Statistics 26: 117–121*

[137] Plackett R.L. (1984), *Discussion of Yates' "Tests of significance for 2x2 contingency tables". Journal of Royal Statistical Society Series A 147:426-463*

[138] Pratt J.W. and Gibbons J.D. (1981), *Concepts of Nonparametric Theory. Springer-Verlag, New York*

[139] Robins, J., Breslow, N., and Greenland S. (1986), *Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. Biometrics 42, 311–323*

[140] Robins, J., Greenland S. and Breslow, N.E. (1986), *A general estimator for the variance of the Mantel–Haenszel odds ratio. American Journal of Epidemiology 124, 719–723*

[141] Rosenbaum P.R., Rubin D.B. (1983a), *The central role of the propensity score in observational studies for causal effects. Biometrika; 70:41–55*

[142] Rosenthal R. (1979), *The "file drawer problem" and tolerance for null results. Psychological Bulletin, 5, 638-641*

[143] Rothman K.J., Greenland S., Lash T.L. (2008), *Modern Epidemiology, 3rd ed. (Lippincott Williams and Wilkins) 221−225*

[144] Roy S. N. (1939), *p-statistics or some generalizations in analysis of variance appropriate to multivariate problems. Sankhya 4: 381–396*

[145] Royston P. (1992), *Approximating the Shapiro–Wilk W-test for non-normality". Statistics and Computing 2 (3): 117–119*

[146] Royston P. (1993b), *A toolkit for testing for non-normality in complete and censored samples. Statistician 42: 37–43*

[147] Rufibach K. (2010), *Assessment of paired binary data; Skeletal Radiology volume 40, pages1–4*

[148] Satterthwaite F.E. (1946), *An approximate distribution of estimates of variance components. Biometrics Bulletin, 2, 1 10-1 14*

[149] Savin N.E. and White K.J. (1977), *The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. Econometrica 45, 1989-1996*

[150] Schiaparelli, G. V.(1866), *Sul modo di ricavare la vera espressione delle leggidelta natura dalle curve empiricae.Effemeridi Astronomiche di Milano perl'Arno, 857:3–56*

[151] Scott D. W., (1992), *Multivariate Density Estimation. Theory, Practice and Visualization. New York: Wiley.*

[152] Shapiro S.S. and Wilk M.B. (1965), *An analysis of variance test for normality (complete samples). Biometrika 52 (3–4): 591–611*

[153] Sheather S.J. (2009), *A modern approach to regression with R. New York, NY: Springer*

[154] Shrout P.E., and Fleiss J.L (1979), *Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 86, 420-428*

Copyright ©2010-2026 U-smile Stats − Wszelkie prawa zastrzeżone

[155] Šidák Z. K. (1967), *Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. Journal of the American Statistical Association, 62 (318): 626–633*

[156] Silverman B. W., (1986), *Density estimation for statistics and data analysis, London: Chapman and Hall*

[157] Skillings J.H., Mack G.A. (1981) *On the use of a Friedman-type statistic in balanced and unbalanced block designs. Technometrics, 23:171–177*

[158] Sobel M. E. (1982). *Asymptotic confidence intervals for indirect effects in structural equation models. Sociological Methodology 13: 290–312*

[159] Spearman C. (1910), *Correlation calculated from faulty data. British Journal of Psychology, 3, 271-295*

[160] Tamhane A. C. (1977), *Multiple comparisons in model I One-Way ANOVA with unequal variances. Communications in Statistics, A6 (1), 15-32*

[161] Tarone R. E., Ware J. (1977), *On distribution-free tests for equality of survival distributions. Biometrica, 64(1):156-160*

[162] Tarone R.E. (1985), *On heterogeneity tests based on efficient scores. Biometrika 72, 91–95*

[163] Terpstra T. J. (1952), *The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. Indagationes Mathematicae, 14: 327–333*

[164] Terrell G. R. (1990), *The maximal smoothing principle in density estimation. Journal of the American Statistical Association 85, 470–477*

[165] Terrell G.R., Scott D. W. (1985), *Oversmoothed nonparametric density estimates. Journal of the American Statistical Association 80, 209-214*

[166] Thode H. C. (2002), *Testing For Normality. CRC Press; 2002. 506 s.*

[167] Volinsky C.T., Raftery A.E. (2000) , *Bayesian information criterion for censored survival models. Biometrics, 56(1):256–262*

[168] Verweij P.J.M., van Houwelingen H.C. (1993), *Cross-validation in survival analysis. Statistics in Medicine 12(24): 2305-2314*

[169] Wallenstein S. (1997), *A non-iterative accurate asymptotic confidence interval for the difference between two Proportions. Statistics in Medicine 16: 1329-1336*

[170] Wallis W.A. (1939), *The correlation ratio for ranked data. Journal of the American Statistical Association, 34,533-538*

[171] Watson, G. S. (1964), *Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A. 26 (4): 359–372*

[172] Welch B. L. (1951), *On the comparison of several mean values: an alternative approach. Biometrika 38: 330–336*

[173] Wilcoxon F. (1945), *Individual comparisons by ranking methods. Biometries, 1,80-83*

[174] Wilcoxon F. (1945), *Individual comparisons by ranking methods. Biometries, 1, 80-83*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[175] Wilcoxon F. (1949), *Some rapid approximate statistical procedures. Stamford, CT: Stamford Research Laboratories, American Cyanamid Corporation*

[176] Wilcoxon F. (1949), *Some rapid approximate statistical procedures. Stamford, CT: Stamford Research Laboratories, American Cyanamid Corporation*

[177] Wilcoxon F. (1949), *Some rapid approximate statistical procedures. Stamford, CT: Stamford Research Laboratories, American Cyanamid Corporation*

[178] Wilson E.B. (1927), *Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association: 22(158):209-212*

[179] Wilks S.S. (1932), *Certain generalizations in the analysis of variance. Biometrika 24: 471–494*

[180] Yates F. (1934), *Contingency tables involving small numbers and the chi-square test. Journal of the Royal Statistical Society, 1,2 17-235*

[181] Youden W.J. (1950), *Index for rating diagnostic tests. Cancer. 3: 32–35*

[182] Kubiak K.B., Więckowska B., Jodłowska-Siewert E., Guzik P. (2024), *Visualising and quantifying the usefulness of new predictors stratified by outcome class: The U-smile method. PLOS ONE 19(5): e0303276*

[183] Więckowska B., Kubiak K.B., Guzik P. (2025), *Evaluating the three-level approach of the U-smile method for imbalanced binary classification. PLOS ONE 20(4): e0321661*

[184] Więckowska B., Guzik P. (2026), *Usmile likelihood evaluation provides robust threshold free assessment of binary classification models for balanced and imbalanced datasets. Scientific Reports 16: 10000*

[185] Friedman J., Hastie T., Tibshirani R. (2010), *Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33(1):1-22*

[186] Yule G. (1900), *On the association of the attributes in statistics: With illustrations from the material ofthe childhood society, and c. Philosophical Transactions of the Royal Society, Series A, 194,257-3 19*

[187] Zar J. H., (2010), *Biostatistical Analysis (Fifth Edition). Pearson Educational*

[188] Zweig M.H., Campbell G. (1993), *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39:561-577*

[189] Bates D., Mächler M., Bolker B., Walker S. (2015), *Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67(1):1-48*

[190] Belenky G., Wesensten N.J., Thorne D.R., Thomas M.L., Sing H.C., Redmond D.P., Russo M.B., Balkin T.J. (2003), *Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. Journal of Sleep Research 12:1-12*

[191] Klein J.P. (1992), *Semiparametric estimation of random effects using the Cox model based on the EM algorithm. Biometrics 48:795-806*

[192] Kuznetsova A., Brockhoff P.B., Christensen R.H.B. (2017), *lmerTest package: tests in linear mixed effects models. Journal of Statistical Software 82(13):1-26*

[193] Laird N.M., Ware J.H. (1982), *Random-effects models for longitudinal data. Biometrics 38:963-974*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[194] Lesnoff M., Laval G., Bonnet P., Abdicho S., Workalemahu A., Kifle D., Peyraud A., Lancelot R., Thiaucourt F. (2004), *Within-herd spread of contagious bovine pleuropneumonia in Ethiopian highlands. Preventive Veterinary Medicine 64:27-40*

[195] McGilchrist C.A., Aisbett C.W. (1991), *Regression with frailty in survival analysis. Biometrics 47:461-466*

[196] Nakagawa S., Schielzeth H. (2013), *A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4:133-142*

[197] Potthoff R.F., Roy S.N. (1964), *A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51:313-326*

[198] Self S.G., Liang K.Y. (1987), *Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 82:605-610*

[199] Therneau T.M., Grambsch P.M. (2000), *Modeling Survival Data: Extending the Cox Model. Springer, New York*

[200] Zeger S.L., Liang K.Y., Albert P.S. (1988), *Models for longitudinal data: a generalized estimating equation approach. Biometrics 44:1049-1060*

[201] Brier G.W. (1950), *Verification of forecasts expressed in terms of probability. Monthly Weather Review 78:1-3*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone