



PQStat Software
Statystyczne Oprogramowanie Obliczeniowe

Podręcznik Użytkownika - PQStat

Barbara Więckowska

COPYRIGHT ©2010-2023 PQSTAT SOFTWARE

Wszelkie prawa zastrzeżone

Do wersji 1.8.6
P7909040423

www.pqstat.pl

Spis treści

1	WYMAGANIA SPRZĘTOWE	6
2	INSTALACJA PQSTAT	6
3	USTAWIENIA	7
4	OBSŁUGA PROGRAMU	10
4.1	PRACA Z ARKUSZAMI DANYCH	12
4.1.1	DOŁĄCZANIE, USUWANIE I EKSPORTOWANIE ARKUSZY DANYCH	12
4.1.2	UMIESZCZANIE DANYCH W ARKUSZU	12
4.1.3	OKNO ARKUSZA DANYCH	14
4.1.4	WŁAŚCIWOŚCI ZMIENNEJ	15
4.1.5	EDYCJA DANYCH	17
4.1.6	SORTOWANIE DANYCH	19
4.1.7	ZAMIANA DANYCH SUROWYCH NA TABELĘ KONTYNGENCJI	20
4.1.8	ZAMIANA TABELI KONTYNGENCJI NA DANÉ SUROWE	20
4.1.9	FORMUŁY	21
4.1.10	GENEROWANIE DANYCH	24
4.1.11	BRAKI DANYCH	25
4.1.12	TRANSFORMACJE	28
4.1.13	STANDARYZACJA BEZPOŚREDNIA I POŚREDNIA	32
4.1.14	SYMULACJA PRÓBKOWANIA	33
4.1.15	MACIERZ PODOBIENSTWA	36
4.2	PRACA Z ARKUSZAMI WYNIKÓW (RAPORTAMI)	46
4.3	MENU I USTAWIENIA JĘZYKOWE PROGRAMU	47
4.3.1	JĘZYK	47
4.3.2	MENU	47
5	ORGANIZACJA PRACY Z PROGRAMEM	48
5.1	FORMA ORGANIZACJI DANYCH	48
5.2	OBSZAR ROBOCZY ARKUSZA	50
5.2.1	Organizowanie zmiennych w zestawie	50
5.2.2	Aktywacja/dezaktywacja przypadków	51
5.2.3	Zaznaczenie obszaru	51
5.2.4	Zapisane zaznaczenie	52
5.2.5	Filtr danych	53
5.3	WIELOKROTNE WYKONYWANIE ANALIZ	56
5.4	ORGANIZOWANIE RAPORTÓW W ZESTAWY	56
5.5	INFORMACJE ZWRACANE DO RAPORTU	56
5.6	OZNACZANIE WYNIKÓW ISTOTNYCH STATYSTYCZNIE	57
6	WYKRESY	58
6.1	GALERIA WYKRESÓW	58
6.1.1	Wykresy kolumnowe	59
6.1.2	Wykresy błędów estymatora	64
6.1.3	Wykresy typu ramka-wąsy	65
6.1.4	Wykresy punktowe	67
6.1.5	Wykresy liniowe	69
6.2	LOKALNE TECHNIKI WYGŁADZANIA LINIOWEGO	70
6.2.1	LOWESS	70
6.2.2	Wygładzanie jądrowe	71
6.3	ESTYMACJA JĄDROWA	73
6.3.1	Jednowymiarowy estymator jądrowy	73
6.4	WYKRES BLANDA-ALTMANA	76
6.5	Macierze korelacji	79

7	MOC TESTU I LICZNOŚĆ PRÓBY	87
7.1	Niezbędna liczność próby	87
7.2	Moc i liczność dla testu	89
7.2.1	Test t-Studenta dla pojedynczej próby	91
7.2.2	Test t-Studenta dla grup zależnych	94
7.2.3	Test t-Studenta dla grup niezależnych	95
7.2.4	Test chi-kwadrat wariancji pojedynczej próby	96
7.2.5	Test chi-kwadrat dwóch wariancji Fishera-Snedecora	98
7.2.6	Test chi-kwadrat zgodności	99
7.2.7	Test chi-kwadrat (RxC)	99
7.2.8	Test dwóch niezależnych proporcji, Chi-kwadrat (2x2)	101
7.2.9	ANOVA dla grup niezależnych	102
7.2.10	Test dla jednej proporcji	103
8	ANALIZY OPISOWE	104
8.1	SKALE POMIAROWE	104
8.2	TABELE	105
8.2.1	TABELE LICZNOŚCI I ROZKŁADY EMPIRYCZNE DANYCH	105
8.2.2	RAPORT TABELI	110
8.2.3	ANALIZY DLA TABEL KONTYNGENCJI	115
8.3	STATYSTYKI OPISOWE	119
8.3.1	MIARY POŁOŻENIA	120
8.3.2	MIARY TENDENCJI CENTRALNEJ	120
8.3.3	INNE MIARY POŁOŻENIA	121
8.3.4	MIARY ROZPROSZENIA	121
8.3.5	INNE ATRYBUTY ROZKŁADU	123
8.4	ZESTAWIENIA OPISOWE	125
9	ROZKŁADY PRAWDOPODOBIENSTWA	129
9.1	CIĄGŁE ROZKŁADY PRAWDOPODOBIENSTWA	131
9.2	KALKULATOR FUNKCJI DYSTRYBUCJI	134
10	TESTOWANIE HIPOTEZ	137
10.0.1	ESTYMACJA PUNKTOWA I PRZEDZIAŁOWA	137
10.0.2	WERYFIKACJA HIPOTEZ STATYSTYCZNYCH	137
11	TESTY NORMALNOŚCI ROZKŁADU	140
11.0.1	Testy normalności jednowymiarowej	140
11.0.2	Testy normalności wielowymiarowej	149
12	PORÓWNANIE - 1 GRUPA	153
12.1	TESTY PARAMETRYCZNE	154
12.1.1	Test t-Studenta dla pojedynczej próby	154
12.1.2	Test chi-kwadrat wariancji pojedynczej próby	157
12.2	TESTY NIEPARAMETRYCZNE	160
12.2.1	Test Wilcoxon (rangowanych znaków)	160
12.2.2	Test chi-kwadrat zgodności	162
12.2.3	Testy dla jednej proporcji	166
13	PORÓWNANIE - 2 GRUPY	170
13.1	TESTY PARAMETRYCZNE	171
13.1.1	Test Fishera-Snedecora	171
13.1.2	Test t-Studenta dla grup niezależnych	172
13.1.3	Test t-Studenta z korektą Cochran-Coxa	173
13.1.4	Test t-Studenta dla grup zależnych	176
13.2	TESTY NIEPARAMETRYCZNE	180
13.2.1	Test U Manna-Whitneya	180

13.2.2	Test Wilcoxona (kolejności par)	184
13.2.3	Testy chi-kwadrat	187
13.2.4	Test chi-kwadrat dla dużych tabel	189
13.2.5	Test Fishera dla tabel dużych tabel	191
13.2.6	Test chi-kwadrat dla małych tabel	194
13.2.7	Poprawki testu chi-kwadrat dla małych tabel	197
13.2.8	Test chi-kwadrat dla trendu	198
13.2.9	Relatywne Ryzyko i Iloraz Szans	202
13.2.10	Test Z dla dwóch niezależnych proporcji	204
13.2.11	Test McNemara, test wewnętrznej symetrii Bowkera	208
13.2.12	Test Z dla dwóch zależnych proporcji	214
14	PORÓWNANIE - WIĘCEJ NIŻ 2 GRUPY	218
14.1	TESTY PARAMETRYCZNE	219
14.1.1	ANOVA dla grup niezależnych	219
14.1.2	Kontrasty i testy POST-HOC	220
14.1.3	ANOVA dla grup niezależnych z korektą F^* i F''	228
14.1.4	Test Browna-Forsythea i test Levene'a	234
14.1.5	ANOVA powtarzanych pomiarów	235
14.1.6	Sferyczność Mauchly'a	239
14.1.7	ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA	240
14.2	TESTY NIEPARAMETRYCZNE	246
14.2.1	ANOVA Kruskala-Wallis'a	246
14.2.2	Test Jonckheere-Terpstra dla trendu	252
14.2.3	Test wariancji rang Conover	255
14.2.4	ANOVA Friedmana	258
14.2.5	Test Page dla trendu	264
14.2.6	ANOVA Durbin'a (brakujących danych)	267
14.2.7	ANOVA Skillings-Mack (brakujących danych)	272
14.2.8	Test chi-kwadrat dla wielowymiarowych tabel kontyngencji	274
14.2.9	ANOVA Q-Cochrana	276
15	WIELOKROTNE PORÓWNANIA	280
16	HOTELLING T-KWADRAT	284
16.1	Test T-kwadrat Hotellinga dla pojedynczej próby	284
16.2	Test T-kwadrat Hotellinga dla grup zależnych	288
16.3	Test T-kwadrat Hotellinga dla grup niezależnych	291
16.4	Test T-kwadrat Hotellinga dla grup niezależnych z korektą dla różnych wariancji	293
16.5	Test Box'a równości macierzy kowariancji	297
17	JEDNOCZYNNIKOWA MANOVA	299
18	ANALIZA DLA WARSTW	305
18.1	Metoda Mantela-Haenszela dla wielu tabel 2x2	305
18.1.1	Iloraz szans Mantela-Haenszela	305
18.1.2	Relatywne ryzyko Mantela-Haenszela	310
19	KORELACJA	312
19.1	TESTY PARAMETRYCZNE	313
19.1.1	WSPÓŁCZYNNIKI KORELACJI LINIOWEJ	313
19.1.2	Istotność współczynnika korelacji Pearsona	314
19.1.3	Istotność współczynnika nachylenia prostej	314
19.1.4	Porównanie współczynników korelacji	318
19.1.5	Porównanie nachylenia prostych regresji	319
19.2	TESTY NIEPARAMETRYCZNE	321
19.2.1	WSPÓŁCZYNNIKI KORELACJI MONOTONICZNEJ	321

19.2.2	Test t do sprawdzania istotności współczynnika korelacji rangowej Spearmana	322
19.2.3	Test do sprawdzania istotności współczynnika korelacji \tilde{r} Kendalla	324
19.2.4	WSPÓŁCZYNNIKI TABEL KONTYNGENCJI I ICH ISTOTNOŚĆ STATYSTYCZNA	326
20	ANALIZA ZGODNOŚCI	332
20.1	TESTY PARAMETRYCZNE	333
20.1.1	Współczynnik korelacji wewnątrzklasowej i test badający jego istotność	333
20.2	TESTY NIEPARAMETRYCZNE	338
20.2.1	Współczynnik zgodności Kendalla i test badający jego istotność	338
20.2.2	Współczynnik Kappa Cohena i test badający jego istotność	341
20.2.3	Współczynnik Kappa Fleissa i test badający jego istotność	346
21	TESTY DIAGNOSTYCZNE	351
21.1	OCENA TESTU DIAGNOSTYCZNEGO	351
21.2	KRZYWA ROC	355
21.2.1	Wybór optymalnego punktu odcięcia	358
21.2.2	Porównywanie krzywych ROC	363
22	MODELE WIELOWYMIAROWE	370
23	DOPASOWANIE GRUP	370
23.1	PRZYGOTOWANIE ZMIENNYCH DO ANALIZY W MODELACH WIELOWYMIAROWYCH	381
23.1.1	Kodowanie zmiennych w modelach wielowymiarowych	381
23.1.2	Interakcje	384
23.2	LINIOWA REGRESJA WIELORAKA	384
23.2.1	Weryfikacja modelu	386
23.2.2	Więcej informacji o zmiennych w modelu	389
23.2.3	Analiza reszt modelu	390
23.2.4	Przykład dla regresji wielorakiej	391
23.2.5	Predykcja na podstawie modelu i walidacja zbioru testowego	400
23.3	PORÓWNYWANIE MODELI LINIOWEJ REGRESJI WIELORAKIEJ	404
23.4	REGRESJA LOGISTYCZNA	408
23.4.1	Iloraz Szans	410
23.4.2	Weryfikacja modelu	411
23.4.3	Przykłady dla regresji logistycznej	419
23.4.4	Predykcja na podstawie modelu i walidacja zbioru testowego	430
23.5	PORÓWNYWANIE MODELI REGRESJI LOGISTYCZNEJ	435
23.6	ANOVA CZYNNIKOWA - GLM	447
23.7	ANCOVA	459
24	Efekt mediacji	466
25	REDUKCJA WYMIARÓW I GRUPOWANIE	468
25.1	ANALIZA SKŁADOWYCH GŁÓWNYCH	469
25.1.1	Interpretacja współczynników związanych z analizą	470
25.1.2	Interpretacja graficzna	471
25.1.3	Kryteria redukcji wymiarów	472
25.1.4	Wyznaczenie składowych głównych	473
25.1.5	Zasadność stosowania Analizy składowych głównych	473
25.2	ANALIZA SKUPIEŃ	480
25.2.1	Metody hierarchiczne	480
25.2.2	Metoda k-średnich	485

26 ANALIZA PRZEŻYCIA	489
26.1 TABELE PRZEŻYCIA	490
26.2 KRZYWE KAPLANA-MEIERA	493
26.3 PORÓWNYWANIE KRZYWYCH PRZEŻYCIA	496
26.3.1 Różnice w krzywych przeżycia	497
26.3.2 Trend w krzywych przeżycia	498
26.3.3 Krzywe przeżycia dla warstw	499
26.4 REGRESJA PROPORCJONALNEGO HAZARDU COX'A	505
26.4.1 Iloraz Hazardu (HR)	507
26.4.2 Weryfikacja modelu	508
26.4.3 Analiza reszt modelu	510
26.5 PORÓWNYWANIE MODELI REGRESJI PH COX'A	511
27 META-ANALIZA	519
27.1 Wprowadzenie	519
27.2 Efekt podsumowujący	523
27.3 Wagi poszczególnych badań	527
27.4 Badanie heterogeniczności	527
27.5 Badanie wrażliwości	529
27.6 Badanie asymetrii	530
27.7 Meta-analiza skumulowana	534
27.8 Porównanie grup	535
27.9 Meta-regresja	543
28 ANALIZA RZETELNOŚCI	548
29 ZESTAWIENIA TESTÓW	554
30 KREATOR	558
31 INNE UWAGI	559
31.1 FORMATY PLIKÓW	559



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
- Zainstalowany system operacyjny: Windows® 2000/XP/Vista/7/8

2 INSTALACJA PQSTAT

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

Po uruchomieniu instalator wyświetli ekran powitalny, aby rozpocząć proces instalacji przechodzimy "Dalej". Instalacja aplikacji wymaga zaakceptowania wyświetlonej w oknie umowy licencyjnej programu, jeżeli zgadzasz się z postanowieniami licencji zaznacz opcję "Akceptuję warunki umowy" i przejdź "Dalej". W przeciwnym przypadku wybierz "Nie akceptuję warunków umowy" i kliknij przycisk Anuluj, aby zakończyć instalację.

Następne okno umożliwia zmianę docelowego miejsca instalacji oraz sprawdza ilość dostępnego miejsca, zalecane jest instalowanie aplikacji w domyślnie podanej lokalizacji.

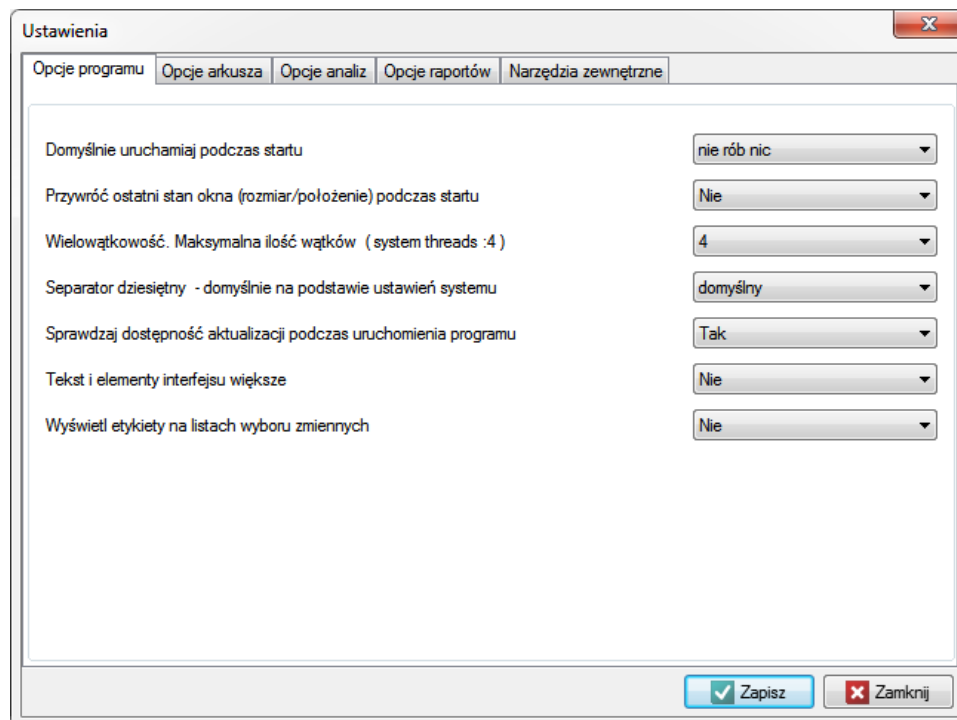
Przechodząc "Dalej" mamy możliwość wyboru pełnej instalacji Oprogramowania bądź Oprogramowania bez przykładowych plików danych. Pliki danych wykorzystywane są w pomocy programu.

Kolejne okno informuje i daje możliwość zmiany nazwy skrótu, który utworzony zostanie w Menu Start systemu Windows.

Przechodząc "Dalej" mamy możliwość, aby instalator utworzył ikonę do programu na pulpicie oraz na pasku szybkiego uruchamiania, zaznacz interesujące Cię opcje i przejdź "Dalej".

Kolejne okno jest ostatnim przed rozpoczęciem instalacji i wyświetla podsumowanie wybranych opcji instalacji. Aby rozpocząć instalację uruchamiamy przycisk "Instaluj".

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 sprawdzenia 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 odpowiedź do nazw wyświetlanych w tych listach. Opcja ta przyniesie pożądany skutek, gdy nazwy zmiennych posiadają przypisane etykiety.

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 (g_2 , czy b_2 Pearson'a), gdzie rozkład normalny charakteryzuje zerowa kurtoza g_2 lub b_2 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 alfabetyczny, 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 odpowiadającym 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 .
- 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>

4 OBSŁUGA PROGRAMU

Podstawa zarządzania dokumentami w programie opiera się na projektach. Każdy projekt stanowi oddzielny plik.

Projekt jest obiektem o znaczeniu podobnym do skoroszytu (zeszytu), w którego skład wchodzi 3 podstawowe elementy:

1. Arkusze danych (w tym arkusze map i macierzy) - liczba arkuszy w danym projekcie jest ograniczona do 1000,
2. Arkusze wyników (raporty) - liczba raportów w odpowiednim arkuszu danych jest ograniczona do 2000,
3. Zarządca projektu - pozwala na zmianę nazwy arkuszy z danymi oraz wynikami, tworzenie własnych opisów i notatek oraz eksport.

Jednocześnie można pracować na maksymalnie 255 projektach. Pierwszy projekt wraz z pustym arkuszem danych jest otwierany automatycznie po uruchomieniu programu, jeśli ta opcja jest ustawiona w oknie [Ustawienia programu](#).

Kolejne projekty mogą być tworzone poprzez menu

- Plik→Nowy projekt (Ctrl+N)




- Plik→Nowy arkusz (Ctrl+D)



Utworzone projekty (pliki z rozszerzeniem [pqs](#), [pqx](#)) otwieramy przez:

- Plik→Otwórz projekt (Ctrl+O)



- Plik→Otwórz przykłady  - dotyczy przykładów dołączanych do programu,
- przeciągnięcie pliku projektu na okno programu,
- dwukrotne kliknięcie pliku projektu.

Projekt można zapisać przez:

- menu Plik→Zapisz (Ctrl+S)



- Plik→Zapisz jako...



- przycisk Zapisz w [Zarządcy Projektu](#).

Zapisanie projektu powoduje zapisanie wszystkich składowych projektu do pliku z rozszerzeniem [pqs](#) lub [pqx](#).

Projekt zamykamy przez:

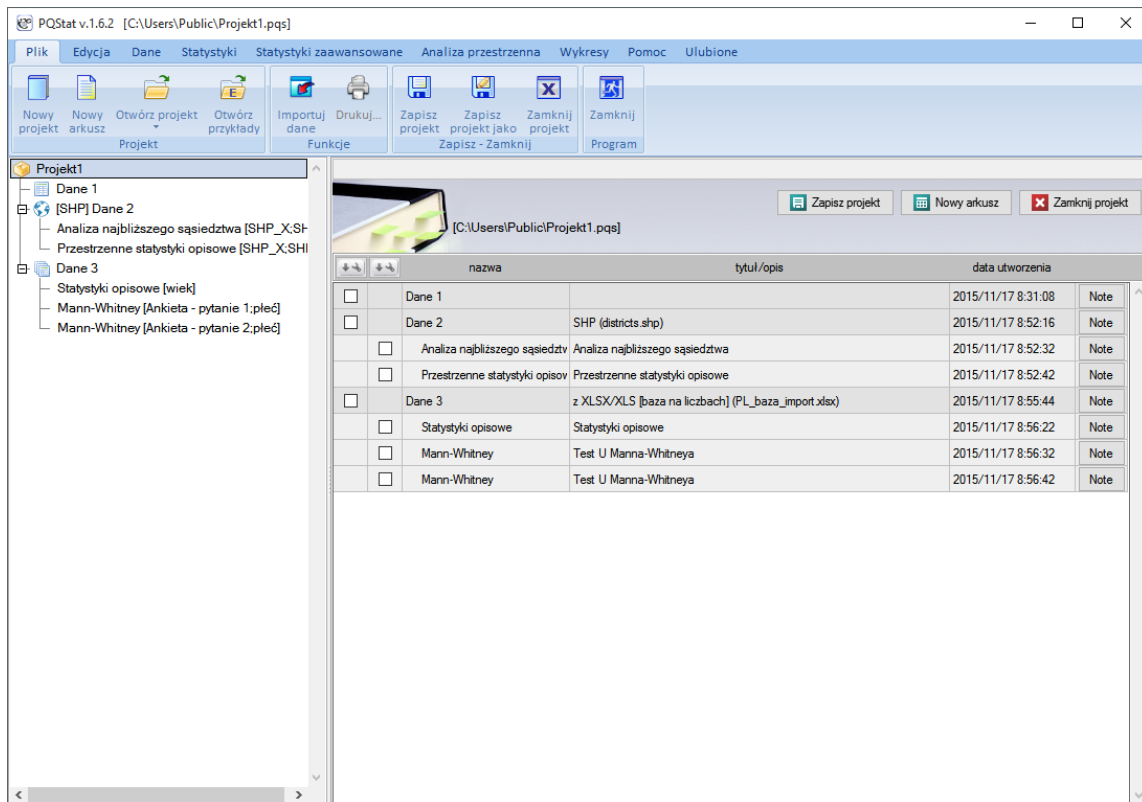
- menu Plik→Zamknij projekt



- przycisk Zamknij projekt w [Zarządcy Projektu](#).

Do łatwej nawigacji w projekcie służy Zarządca Projektu wyświetlany po wybraniu odpowiedniego projektu. W oknie tym można zapisać jak i usunąć wybrany projekt, dodać lub usunąć arkusz danych, usunąć

arkusz wyników jak też dodać notatki. Nazwa projektu jest jednocześnie nazwą pliku projektu (pqqs/pqpx).



The screenshot shows the PQStat v.1.6.2 interface. The main window displays a table of data for 'Projekt1'. The table has columns for 'nazwa', 'tytuł/opis', 'data utworzenia', and 'Note'. The data is as follows:

	nazwa	tytuł/opis	data utworzenia	Note
<input type="checkbox"/>	Dane 1		2015/11/17 8:31:08	Note
<input type="checkbox"/>	Dane 2	SHP (districts.shp)	2015/11/17 8:52:16	Note
<input type="checkbox"/>	Analiza najbliższego sąsiedztwa	Analiza najbliższego sąsiedztwa	2015/11/17 8:52:32	Note
<input type="checkbox"/>	Przestrzenne statystyki opisow	Przestrzenne statystyki opisowe	2015/11/17 8:52:42	Note
<input type="checkbox"/>	Dane 3	z XLSX/XLS [baza na liczbach] (PL_baza_import.xlsx)	2015/11/17 8:55:44	Note
<input type="checkbox"/>	Statystyki opisowe	Statystyki opisowe	2015/11/17 8:56:22	Note
<input type="checkbox"/>	Mann-Whitney	Test U Manna-Whitneya	2015/11/17 8:56:32	Note
<input type="checkbox"/>	Mann-Whitney	Test U Manna-Whitneya	2015/11/17 8:56:42	Note


4.1 PRACA Z ARKUSZAMI DANYCH

Najważniejszym elementem każdego projektu jest arkusz danych. Każdy otwarty projekt musi posiadać przynajmniej jeden arkusz z danymi.


4.1.1 DOŁĄCZANIE, USUWANIE I EKSPORTOWANIE ARKUSZY DANYCH

Pierwszy pusty arkusz danych jest otwierany automatycznie wraz z nowym projektem.

Kolejne arkusze danych mogą być dołączane do projektu przez:

- menu Plik→Nowy arkusz (Ctrl+D) 
- przycisk Nowy arkusz w [Zarządcy Projektu](#).

Usunięcie arkusza danych możliwe jest przez:

- menu kontekstowe Usuń arkusz (Shift+Del) na nazwie arkusza danych w [Drzewie nawigacji](#),
- przycisk  →Usuń w [Zarządcy Projektu](#), gdy zaznaczony jest arkusz/arkusze.

Należy jednak pamiętać, że jeśli do arkusza danych dołączone są raporty lub mapa, to usuwając arkusz danych usuwa się jednocześnie wszystkie przypisane do niego raporty/mapę.

Arkusze możemy opisać w [Zarządcy Projektu](#) dodając nazwę, wpisując tytuł lub dłuższą notatkę.

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

4.1.2 UMIESZCZANIE DANYCH W ARKUSZU

Arkusze danych po utworzeniu są puste. Użytkownik sam wprowadza dane, kopiuje wcześniej przygotowane dane z innego arkusza danych lub importuje. Ilość danych, jaką mieści arkusz jest ograniczona do 4 milionów wierszy i 1000 kolumn. W każdej komórce może się znajdować maksymalnie 40 znaków.

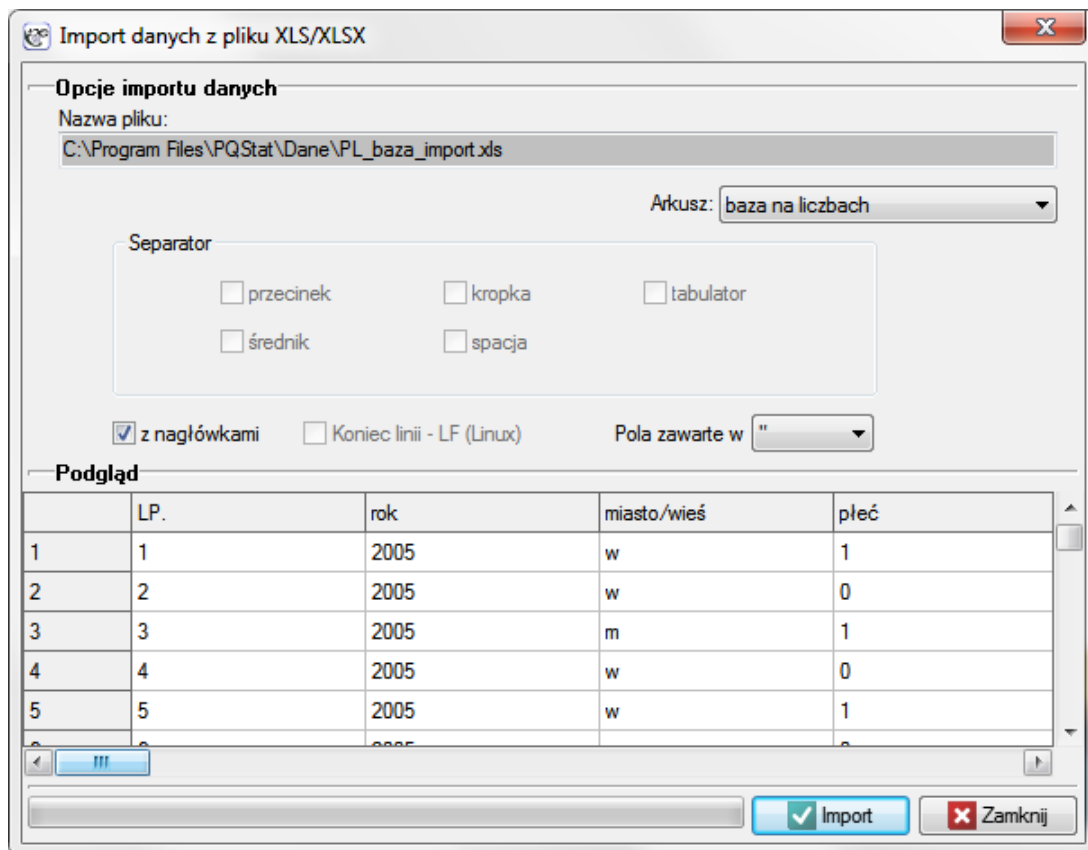
Import danych

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 ...





W oknie importu mamy możliwość podglądu importowanych danych i wcześniejszego sprawdzenia wyniku importu w zależności od ustalonych opcji dotyczących sposobu interpretacji danych. Aby uniknąć błędnej interpretacji znaków narodowych należy zwrócić szczególną uwagę na poprawność wyświetlenia tych znaków w oknie podglądu. W przypadku dużych plików okno podglądu zawiera tylko początkową część danych z pliku.

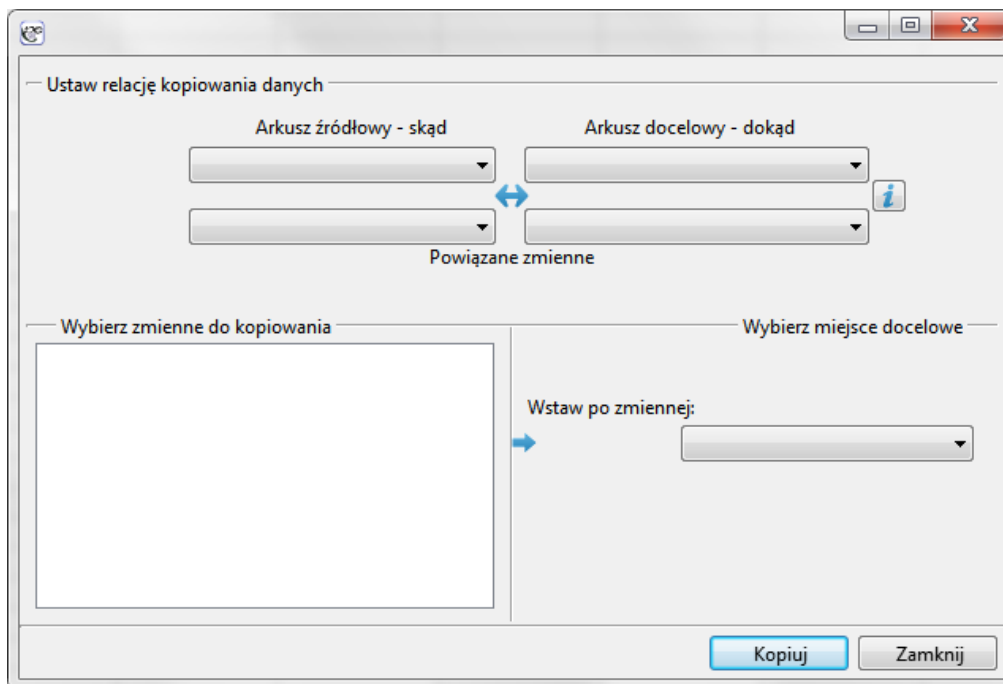
Uwaga!

W programach Microsoft Office Excel 2000-2007 domyślne kodowanie znaków to Windows-1250. Import danych z dokumentów Microsoft Excel dotyczy tylko wartości komórek, nie ma możliwości importu formatowania oraz 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...

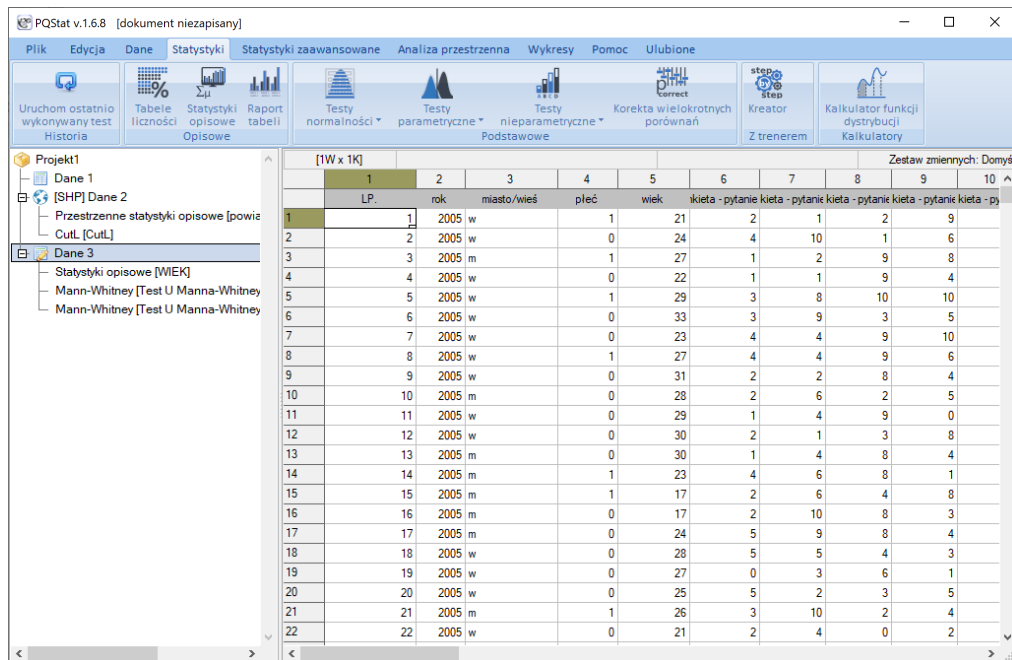




By zbudować relację, należy wskazać arkusz danych, z którego dokonujemy kopiowania i arkusz danych, w którym umieścimy skopiowane dane. Oba te arkusze muszą posiadać ten sam klucz tzn. zmienną, której wartości identyfikują każdy wiersz w arkuszu. Wymagane jest aby klucz dla arkusza źródłowego był unikalny. Powiązanie odbywa się według relacji jeden-do-wielu, czyli jeden wiersz z arkusza źródłowego może być powiązany z wieloma wierszami z arkusza docelowego. Klucze obu arkuszy należy wybrać jako Powiązane zmienne. Dla tak ustawionej relacji wskazujemy zmienne do kopiowania i kolumnę po której mają zostać umieszczone przekopiowane zmienne.

4.1.3 OKNO ARKUSZA DANYCH

Wiersze i kolumny arkusza danych oznaczone są poprzez kolejne liczby naturalne. Każdej kolumnie w miejscu oznaczonym szarym kolorem można nadać własny nagłówek. Na górze arkusza znajduje się Pasek komunikatów. Jest to miejsce, w którym wyświetlane są bieżące informacje dla użytkownika. Lewa część paska informuje o wielkości zaznaczonego obszaru [liczba wierszy, liczba kolumn], środkowa wyświetla wartość znajdującą się w zaznaczonej komórce, a prawa część przeznaczona jest na informacje dla użytkownika dotyczące min. wykonywanej analizy statystycznej.

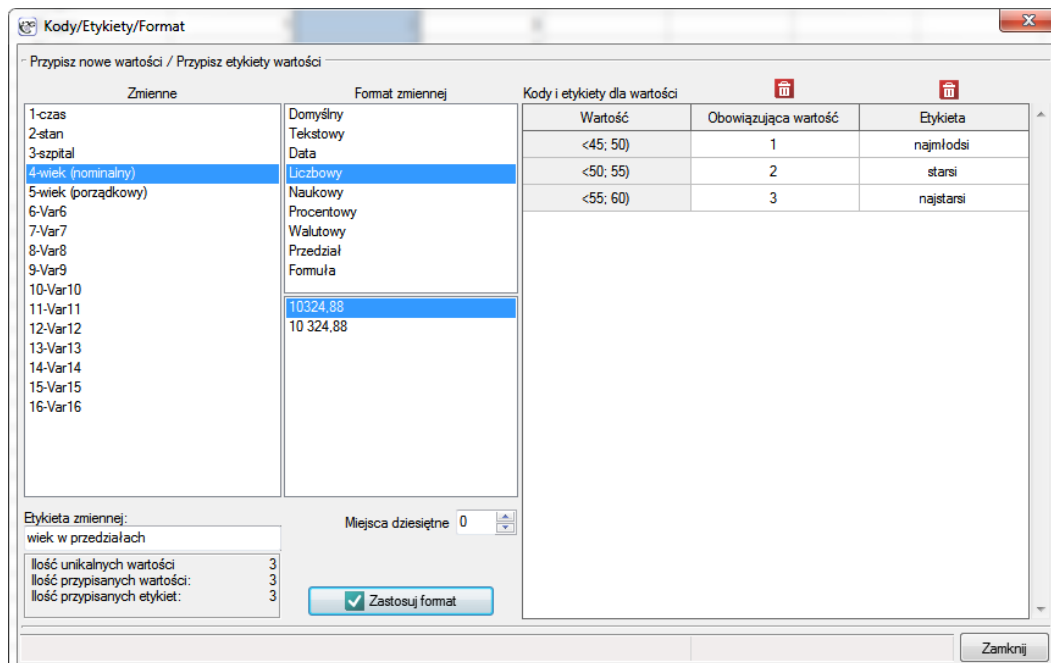


	1	2	3	4	5	6	7	8	9	10
	LP.	rok	miasto/wieś	płeć	wiek	kieta - pytanie kieta	- pytanie kieta	- pytanie kieta	- pytanie kieta	- pytanie kieta
1	1	2005	w	1	21	2	1	2	9	
2	2	2005	w	0	24	4	10	1	6	
3	3	2005	m	1	27	1	2	9	8	
4	4	2005	w	0	22	1	1	9	4	
5	5	2005	w	1	29	3	8	10	10	
6	6	2005	w	0	33	3	9	3	5	
7	7	2005	w	0	23	4	4	9	10	
8	8	2005	w	1	27	4	4	9	6	
9	9	2005	w	0	31	2	2	8	4	
10	10	2005	m	0	28	2	6	2	5	
11	11	2005	w	0	29	1	4	9	0	
12	12	2005	w	0	30	2	1	3	8	
13	13	2005	m	0	30	1	4	8	4	
14	14	2005	m	1	23	4	6	8	1	
15	15	2005	m	1	17	2	6	4	8	
16	16	2005	m	0	17	2	10	8	3	
17	17	2005	m	0	24	5	9	8	4	
18	18	2005	w	0	28	5	5	4	3	
19	19	2005	w	0	27	0	3	6	1	
20	20	2005	w	0	25	5	2	3	5	
21	21	2005	m	1	26	3	10	2	4	
22	22	2005	w	0	21	2	4	0	2	

4.1.4 WŁAŚCIWOŚCI ZMIENNEJ

Dla każdej kolumny arkusza możemy ustawić jej właściwości takie jak kody, etykiety i format. Ustawienie właściwości zmiennej jest możliwe po

- wybraniu Właściwości zmiennej → Kody/Etykiety/Format z menu kontekstowego na numerze znajdującym się powyżej nagłówka kolumny,
- po podwójnym kliknięciu na numerze znajdującym się powyżej nagłówka kolumny - o ile tak zaznaczono w ustawieniach programu (odpowiednia akcja podwójnego kliknięcia).



Zmienne	Format zmiennej	Kody i etykiety dla wartości		
		Wartość	Obowiązująca wartość	Etykieta
1-czas	Domyślny			
2-stan	Tekstowy			
3-szpital	Data	<45; 50)	1	najmłodszy
4-wiek (nominalny)	Liczbowy	<50; 55)	2	starszy
5-wiek (porządkowy)	Naukowy	<55; 60)	3	najstarszy
6-Var6	Procentowy			
7-Var7	Walutowy			
8-Var8	Przedział			
9-Var9	Formuła			
10-Var10				
11-Var11	10324.88			
12-Var12	10 324.88			
13-Var13				
14-Var14				
15-Var15				
16-Var16				

Etykieta zmiennej: wiek w przedziałach

Miejsca dziesiętne: 0

ilość unikalnych wartości: 3
 ilość przypisanych wartości: 3
 ilość przypisanych etykiet: 3

Zastosuj format

Zamknij

Kody i etykiety dla wartości – są możliwe do przypisania dla każdej wartości, która występuje w danej kolumnie.

Obowiązująca wartość - Wypełniając kody ustawiamy wartości jakie mają obowiązywać w danej kolumnie. W rezultacie wartości pobierane do obliczeń zostaną zmienione (w tle) na odpowiednio przypisane kody (obowiązujące wartości).

Etykieta - Wartości jakie wpisane zostaną w polu Etykieta wykorzystywane są w raportach i na wykresach dla ustalonego przez użytkownika opisu wyników.

Etykieta zmiennej - jest przypisywana do nagłówka danej kolumny. Zwykle jest to krótki opis zawartości zmiennej. Ustawiona etykieta zmiennej wykorzystywana jest zamiast nagłówka kolumny (nazwy zmiennej) w raportach i na wykresach dla czytelniejszego opisu wyników. Wykorzystanie etykiety zmiennej jest opcjonalne i zależy od [ustawień](#) programu.

Format zmiennej

Każda komórka arkusza (również nagłówki kolumny) może zawierać najwyżej 40 znaków. Dozwolone są również teksty zawierające 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 interpretowane 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

- **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 ustawiony jest standardowo **format domyślny**.

Cały wiersz nagłówkowy ma na stałe ustawiony format tekstowy. Dla pozostałej części arkusza istnieje możliwość ustawienia przez użytkownika zdefiniowanych formatów. Formatowaniu nie podlega pojedyncza komórka, ale cała kolumna (za wyjątkiem jej nagłówka).

W arkuszu można określić **szerokość kolumny** przy użyciu myszy. W tym celu należy przy pomocy wskaźnika myszy przeciągnąć linię dzielącą kolumny, zwężając lub rozszerzając kolumnę znajdującą się po lewej stronie wybranej linii.

Dodatkowo w każdej komórce arkusza można ustalić inny kolor tła (po uprzednim zaznaczeniu zmienianego obszaru). Służy do tego:






- menu Plik → Kolor wypełnienia ...
- polecenie Kolor komórki w menu kontekstowym komórki.

4.1.5 EDYCJA DANYCH

Zaznaczanie spójnego obszaru arkusza może się odbywać przy pomocy myszy lub klawiatury (Klawisze strzałek + Shift). Podczas zaznaczenia na bieżąco na pasku komunikatów jest wyświetlany jego rozmiar (liczba wierszy i kolumn). Cały arkusz można zaznaczyć w prosty sposób myszą klikając w lewy górny róg arkusza lub wybierając z menu Edycja → Zaznacz wszystko (Ctrl+A). Całe wiersze lub całe kolumny zaznaczamy wybierając ich nagłówki.

Kopiowanie lub przenoszenie komórek dokonuje się poprzez polecenia kopiuj, wytnij i wklej.

Polecenia kopiuj, wytnij i wklej są dostępne w kilku miejscach:

- w menu Edycja,
- w menu kontekstowym komórek,
- na pasku narzędziowym   ,
- w menu kontekstowym wierszy i kolumn,
- przez klawisze skrótów: kopiuj (Ctrl+C), wytnij (Ctrl+X), i wklej (Ctrl+V).

Usuwanie danych z komórek można wykonać poprzez menu Edycja→Usuń (Del)


Cofanie ostatnio wykonywanej operacji można wykonać poprzez menu Edycja→Cofnij (Ctrl+Z). Standardowo Program pamięta 10 ostatnich operacji dotyczących 5000 komórek w każdej z nich. Ustawienia te można zmienić w oknie [Ustawienia](#). Należy zaznaczyć jednak, iż zwiększanie tych wartości niesie za sobą większe użycie pamięci komputera przez program.

Wstawianie i usuwanie wierszy i kolumn

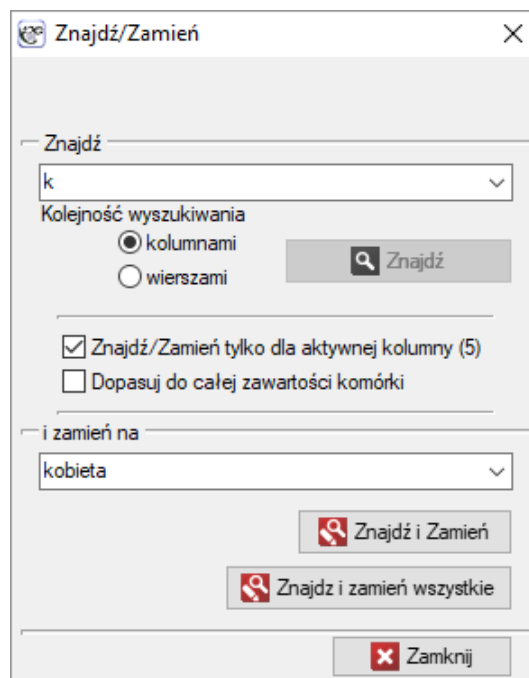
Można wstawić puste wiersze lub kolumny powyżej lub po lewej stronie istniejącego wiersza lub kolumny. Spowoduje to przesunięcie komórek w dół lub w prawo. Aby wstawić wiersz/wiersze należy zaznaczyć ten wiersz/wiersze powyżej którego chcemy wstawić nowy wiersz/wiersze, a następnie wybrać z menu kontekstowego na numerze zaznaczonego wiersza Wstaw wiersz. W analogiczny sposób dokonuje się wstawianie kolumn.

Wiersze i kolumny można również usuwać poprzez ich zaznaczenie i wybranie w menu kontekstowym na numerze wiersza lub kolumny Usuń wiersz/Usuń kolumnę.

Znajdowanie/zamiana wartości komórki

By wyszukać lub zastąpić całą zawartość komórki inną wartością używa się okna Znajdź/Zamień, które wywołujemy poprzez menu Edycja→Znajdź/Zamień (Ctrl+F) .

Górna część okna Znajdź/Zamień służy wyszukiwaniu a dolna zamianie wartości komórki.



Aby wyszukać dane, w górnej części okna należy wpisać szukany ciąg znaków oraz zaznaczyć kolejność wyszukiwania i wybrać przycisk Znajdź.

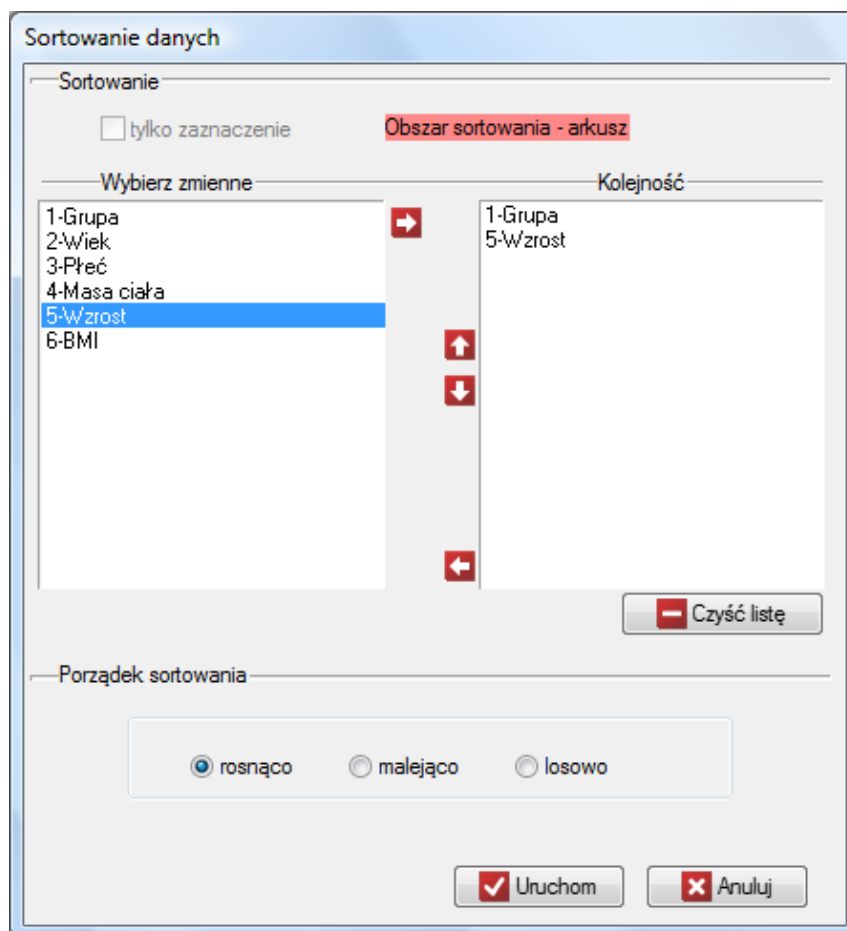
Aby znaleźć i zastąpić całą zawartość komórki inną wartością należy wypełnić górną i dolną część okna. Górną część okna wypełniamy tak jak w przypadku wyszukiwania danych. W dolnej części okna wpisujemy dane, które mają zastąpić znalezione i wybieramy przycisk Znajdź i Zamień lub Znajdź i zamień

wszystkie, gdy chcemy zastąpić wszystkie wystąpienia wyszukanych danych. Wyszukiwanie jak i zamiana odbywa się w trybie bezpośredniego podglądu wykonywanych operacji na arkuszu.

4.1.6 SORTOWANIE DANYCH



Opcje sortowania dostępne są po wybraniu z menu Dane→Sortuj... lub opcji Sortuj... z menu kontekstowego na numerze znajdującym się powyżej nagłówka kolumny. Zwykle sortujemy cały arkusz danych (takie jest ustawienie domyślnie dotyczące sortowania), jeśli jednak sortowanie rozpoczniemy od zaznaczenia fragmentu danych, wówczas w oknie sortowania będziemy mieli możliwość ograniczenia obszaru sortowania tylko do zaznaczenia.



W oknie sortowania za pomocą strzałek przenosimy z pola Wybierz zmienne do pola Kolejność te zmienne, według których chcemy sortować dane, a następnie wybieramy Porządek sortowania i potwierdzamy wybór przyciskiem Uruchom.

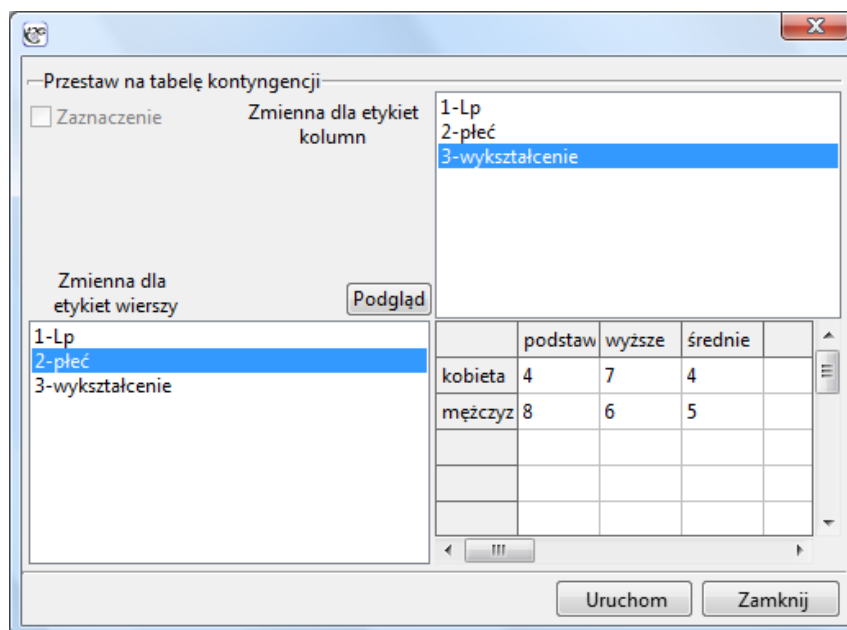
Sortować możemy według co najwyżej 3 kolumn (zmiennych). Jeśli sortujemy dane według więcej niż jednej zmiennej to sortowanie odbywa się w takiej kolejności, w jakiej znajdują się zmienne umieszczone w polu Kolejność.

4.1.7 ZAMIANA DANYCH SUROWYCH NA TABELĘ KONTYNGENCJI



Operacja zamiany **danych surowych** na **tabelę kontyngencji** dostępna jest po wybraniu z menu Dane→Przetaw na tabelę... Zwykle dla tej operacji dostępny jest cały arkusz danych (takie jest ustawienie domyślnie), jeśli jednak przekształcanie rozpoczniemy od zaznaczenia fragmentu danych, wówczas w oknie przekształcania danych będziemy mieli możliwość ograniczenia dostępnego obszaru tylko do zaznaczenia.

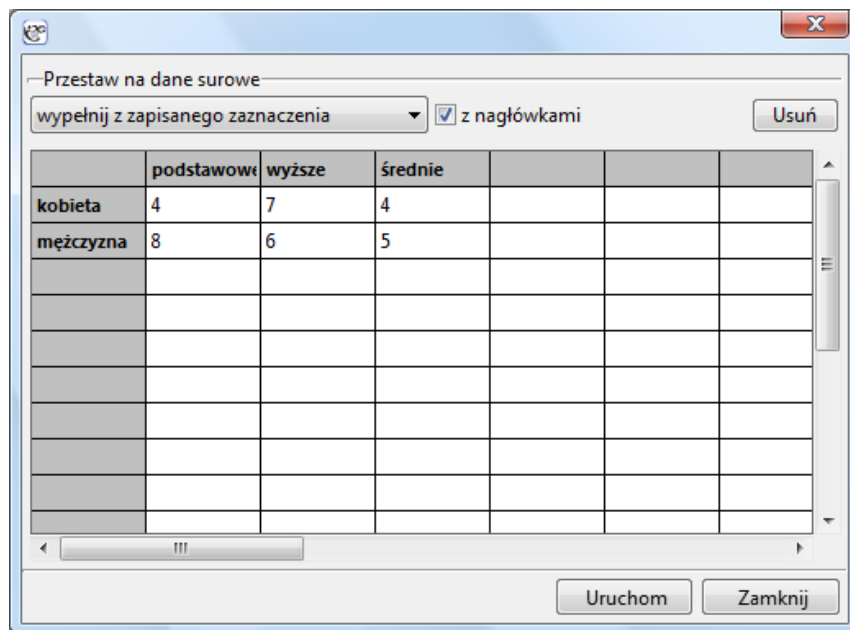
Tabelę kontyngencji projektujemy wybierając zmienne tworzące etykiety wierszy i kolumn. Jeśli podgląd tabeli zgodny jest z oczekiwanym wynikiem, wówczas dokonany wybór potwierdzamy przyciskiem Uruchom. Zwrócony wynik umieszczony zostanie w nowym arkuszu.



4.1.8 ZAMIANA TABELI KONTYNGENCJI NA DANE SUROWE



Operacja zamiany **tabeli kontyngencji** na **dane surowe** dostępna jest po wybraniu z menu Dane→Przetaw na dane surowe... W oknie przekształcania danych wpisujemy odpowiednie licznosci oraz nagłówki wierszy i kolumn i dokonany wybór potwierdzamy przyciskiem Uruchom. Zwrócony wynik umieszczony zostanie w nowym arkuszu.



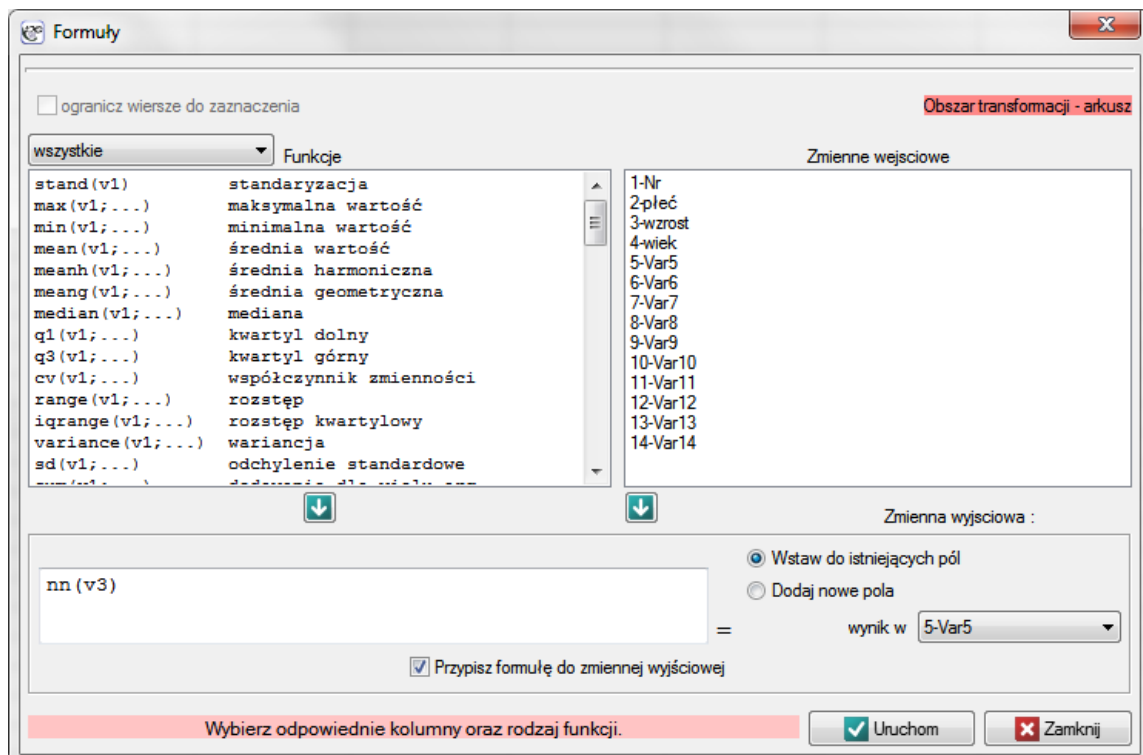
Jeśli przekształcamy tabelę, która znajduje się w arkuszu danych, wówczas przed przystąpieniem do zamiany tej tabeli na dane surowe zaznaczamy ją (z nagłówkami lub bez). Wówczas w oknie przekształcania danych tabela ta zostanie umieszczona automatycznie. Możliwe jest także wykorzystanie innych tabel oznaczonych jako [zapisane zaznaczenie](#).

4.1.9 FORMUŁY



Definiowanie formuły jest sposobem na przeliczenie danych, w wyniku którego uzyskujemy nowe wartości dla zmiennych.

Okno definiowania formuł wywołujemy poprzez Dane→Formuły...



Formuły przypisane do danej zmiennej arkusza jako format tej zmiennej są zapamiętywane wraz z arkuszem danych. Ich wynik zostaje automatycznie przeliczony, gdy zmieniona zostanie którakolwiek z danych wejściowych. Przypisania formuły można dokonać w oknie Formuły... lub poprzez ustawienie Formatu kolumny (Ctrl+F10).

Budowanie formuł

Formuły wprowadzamy w polu edycji

- Zmienne, do których odwołują się formuły, wprowadzamy podając ich numery np. $v1+v2$
- Wartości tekstowe wpisujemy używając apostrofu np. 'dom'
- Funkcje wprowadzamy poprzez dwukrotne kliknięcie na nazwie wybranej funkcji - wówczas nazwa ta pojawi się w polu edycji formuły, lub samodzielnie wpisujemy nazwę w polu edycji, przy czym wielkość liter w nazwie funkcji nie ma znaczenia. Argumenty funkcji podajemy w nawiasie stosując składnię podaną w opisie funkcji.

Wyniki formuł

Wyniki formuł zostaną wyświetlone w wybranej kolumnie.

Jeżeli wśród argumentów funkcji wystąpią wartości, których funkcja nie potrafi zinterpretować to program wyświetli komunikat pytający czy pominąć zmienne niezinterpretowane. Udzielenie odpowiedzi twierdzącej spowoduje przeliczenie formuły z pominięciem danych niezinterpretowanych. W przypadku udzielenia odpowiedzi przeczącej formuła zwróci błąd (NA). Przykładowo dla wartości w kolumnach $v1$, $v2$ i $v3$ odpowiednio: 1, 2, 'ada', funkcja sumowania $\text{sum}(v1; v2; v3)$ zwróci wynik 3 - gdy pominiemy niezinterpretowaną wartość 'ada', lub zwróci NA - gdy nie pominiemy tej wartości w obliczeniach.

Wartość pusta (brak danych) zostanie zwrócona tylko wtedy, gdy wszystkie argumenty wykorzystywane w formule są puste.

Liczbę wierszy biorących udział w formule możemy ograniczyć zaznaczając w arkuszu danych odpowiedni zakres wierszy i wybierając w oknie formuły 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.

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 kwartyła dolnego,
q3(v1,...) - zwraca wartość kwartyła górnego,
cv(v1,...) - zwraca wartość współczynnika zmienności,
range(v1,...) - zwraca wartość rozstępu,
iqrang(v1,...) - zwraca wartość rozstępu kwartyłowego,
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 4.1.4). 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 odpowiadającą podanemu czasowi,

minute(v1;) - zwraca minutę odpowiadającą podanemu czasowi,

second(v1;) - zwraca sekundę odpowiadającą 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,

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.

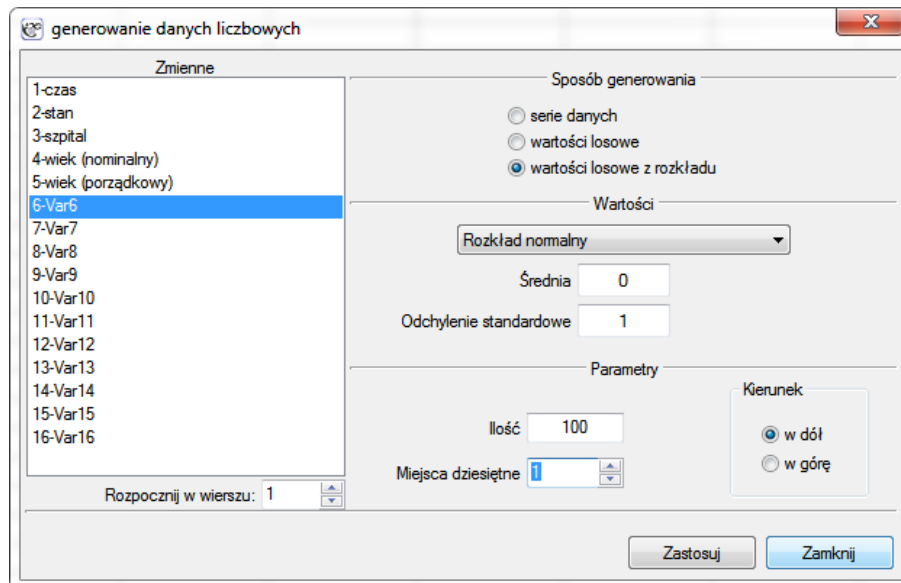
4.1.10 GENEROWANIE DANYCH



Istnieją dwie metody generowania danych:

1. Pierwsza metoda korzysta z prostego przeciągnięcia zawartości z zaznaczonych komórek na komórki sąsiednie przy pomocy wskaźnika myszy. Metoda ta pozwala na generowanie takich samych wartości (tekstowych lub liczbowych) w sąsiednich kolumnach bądź wierszach. Generowanie rozpoczynamy od wyboru komórki z odpowiednim wpisem, a następnie chwytamy wskaźnikiem myszy obrazującym znak + prawy dolny narożnik tej komórki i przytrzymując przeciągamy przez komórki, które chcemy wypełnić. Przeciąganie pojedynczej komórki możemy wykonać w dowolnym kierunku (w górę w dół w lewo i w prawo). Możliwe jest również przeciąganie różnych wartości umieszczonych w jednej kolumnie (w lewo bądź w prawo) lub w jednym wierszu (w górę bądź w dół).
2. Druga metoda pozwala na generowanie danych liczbowych w kolumnach jako serii danych, wartości losowych oraz wartości losowych z odpowiedniego rozkładu danych.

By wygenerować dane liczbowe należy zaznaczyć komórkę, od której chcemy zacząć wypełnianie arkusza i wywołać okno generowania danych liczbowych poprzez menu Dane→Generuj...



Zaczynamy od wyboru zmiennej, w której będą umieszczone wygenerowane dane. W części środkowej okna w zależności od ustawień wybranego wyżej sposób generowania danych ustawiamy:

- Dla generowania serii danych:
 - Wartość początkowa - pierwsza wartość, która ma być wygenerowana,
 - Krok - wartość, o jaką mają się różnić kolejne wygenerowane dane.
- Dla generowania wartości losowych:
 - Przedział (od) - początek przedziału, z którego będą losowane wartości,
 - Przedział (do) - koniec przedziału, z którego będą losowane wartości.
- Dla generowania wartości losowych z rozkładu wybieramy rodzaj rozkładu (Rozkład normalny, Rozkład Chi-kwadrat) i wpisujemy jego parametry.

Ilość wygenerowanych danych zależy od wartości, jaką użytkownik wpisze w polu ilość, a precyzja od ustawienia pola miejsca dziesiętne. Dane zostaną wpisane począwszy od aktywnej komórki w dół bądź w górę - w zależności od wybranej opcji. Ostatecznie potwierdzamy dokonany wybór przyciskiem Zastosuj.

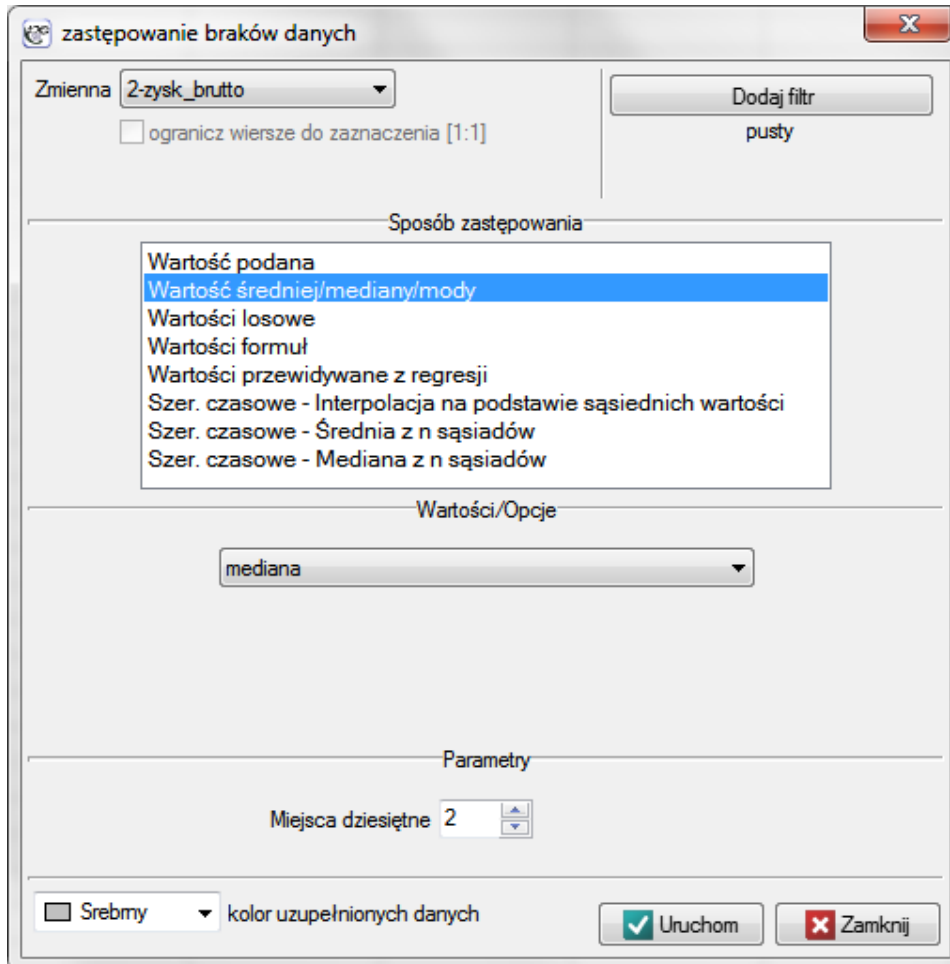
4.1.11 BRAKI DANYCH



W badaniach bardzo często napotykamy na braki danych, jest to naturalne szczególnie dla danych ankietowych. Bywają sytuacje, w których braki danych wnoszą wartościową informację. Przykładowo: liczba braków danych w odpowiedzi na pytanie dotyczące sympatii do partii politycznych daje pogląd o liczbie niezdecydowanych osób, które nie darzą sympatią (lub nie przyznają się że darzą sympatią) określonych ugrupowań politycznych. Niewielkie licznosci braków danych nie stanowią problemu w analizach statystycznych. Duża ich liczba może jednak poddawać pod wątpliwość rzetelność przeprowadzonych badań. Warto już na początku pracy zadbać by było ich jak najmniej. Oczywiście najlepiej jest dotrzeć do informacji o rzeczywistej wartości, która powinna być wpisana w miejsce braku danych, jednak nie zawsze jest to możliwe.

Sposób w jaki szacowane są brakujące dane zależy przede wszystkim od charakteru danych. W programie zaproponowano kilka sposobów imputacji braków danych dla poszczególnych zmiennych.

Okno z ustawieniami opcji zastępowania braków danych wywołujemy poprzez menu Dane→Braki danych...



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;
- losowymi z przedziału podanego przez użytkownika;
- wyliczonymi z funkcji użytkownika - opcja ta pozwala na wykorzystanie danych z innych zmiennych by móc przewidzieć brakującą wartość w wybranej kolumnie;
- wyliczonymi z modelu regresji - opcja ta pozwala na przewidywanie wartości braków danych na podstawie modelu regresji wielorakiej (zasada działania regresji wielorakiej opisana została w dziale [23.2 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 można było odróżnić dane, które zostały imputowane od danych rzeczywistych, miejsca zastąpione oznaczane są wybranym kolorem.

PRZYKŁAD 4.1. (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 nieznanymi (kodowanych jako 0). Wyliczymy więc medianę zysku brutto oddzielnie dla książek autorów znanych i nieznanymi. 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 tysięcy dolarów.

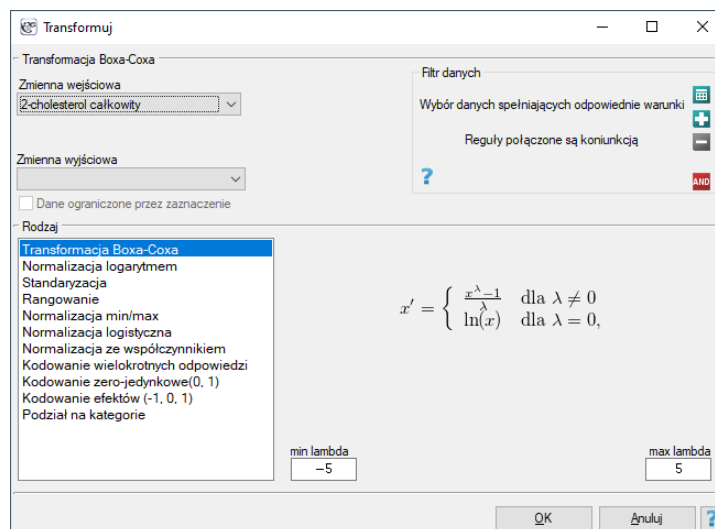
Innym sposobem zastępowania braków jest np. skorzystanie z modelu regresji. Wybieramy arkusz danych "Wstaw z regresji" i ponownie wybieramy w oknie Braków danych zmienną dotyczącą zysku brutto

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.

4.1.12 TRANSFORMACJE



Okno transformacji wywołujemy poprzez Dane → Transformuj...



Transformacja danych to ich przekształcenie w taki sposób, by spełniały określone kryteria np. spełniały kryteria normalności rozkładu czy też rozciągały się w określonym przedziale.

Transformacja Boxa-Coxa

Transformacja Boxa-Coxa wprowadzona przez Boxa i Coxa w roku 1964[24] 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:

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

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:

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

gdzie:

n - liczność próby,

sd_{pop} - odchylenie standardowe populacji.

Uwaga!

Jeśli przed transformacją Boxa-Coxa wykorzystano normalizację min-max, wówczas po transformacji Boxa-Coxa można powrócić do poprzedniego przedziału ponownie używając tej transformacji.

Transformacja logarytmiczna

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

$$x' = \ln x \quad (3)$$

Standaryzacja

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

$$x' = \frac{x - \bar{x}}{sd} \quad (4)$$

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 związaną**. Rangą związaną 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 związana wyliczona jako średnia arytmetyczna z liczb: 7 i 8, a dla zmiennej o wartości 8.6 ranga związana 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. 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).

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

Normalizacja logistyczna

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

$$x' = \frac{e^x}{1 + e^x} \quad (6)$$

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 .

$$x' = \frac{x}{\sqrt{x^2 + \alpha}} \quad (7)$$

Zwiększenie wartości współczynnika α tworzy wykres o bardziej łagodnym zboczach.

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 odpowiedzi jest rozbijana na k nowych zmiennych. Przy czym należy podać jaki znak (lub zestaw 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 23.1 PRZYGOTOWANIE ZMIENNYCH DO ANALIZY W MODELACH WIELOWYMIAROWYCH.

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 23.1 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.

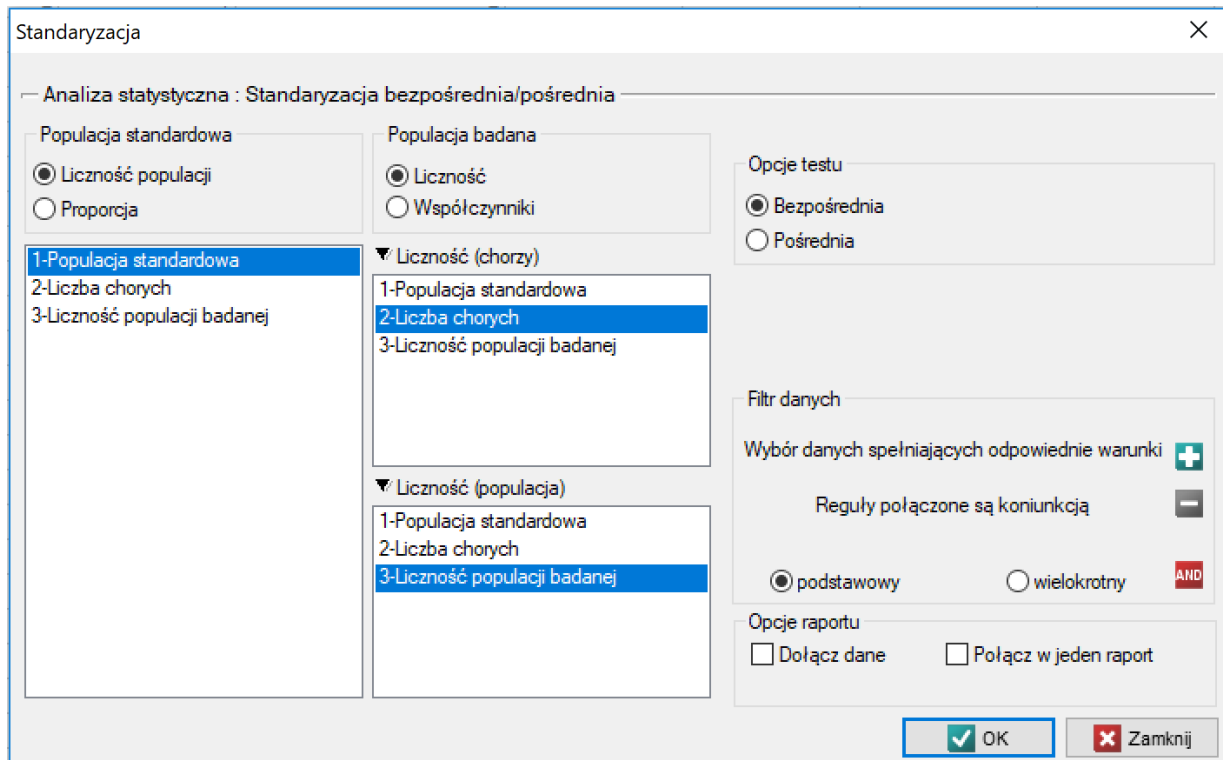
PRZYKŁAD 4.2. (plik: normalizacja.pqs)

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 kilkakrotnie zmieniając wartość współczynnika α .
- f) Dokonaj standaryzacji wszystkich danych, które są opisane rozkładem normalnym.
- g) Przekształć zmienną obrazującą jak zmieniała 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.

4.1.13 STANDARYZACJA BEZPOŚREDNIA I POŚREDNIA

Okno bezpośredniej i pośredniej standaryzacji współczynników epidemiologicznych wywołujemy poprzez Dane→Standaryzacja...



Standaryzacji pośredniej i bezpośredniej podlegają współczynniki częstości np. współczynniki chorobowości. Bezpośrednie porównanie surowych współczynników pomiędzy różnymi populacjami, zwykle zamieszkującymi rozmaite rejony geograficzne i różniącymi się czynnikami mogącymi wpływać na wartość współczynników (np. wiekiem), może maskować występujące różnice. By zniwelować wpływ struktury populacji co do tych cech wnikających, porównania można dokonać na bazie współczynników standaryzowanych. Standaryzacja oferuje mechanizm „odfiltrowania” wpływu znanego czynnika wnikającego (np. wieku) i 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 dokonywane-go wyboru, m.in.:

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

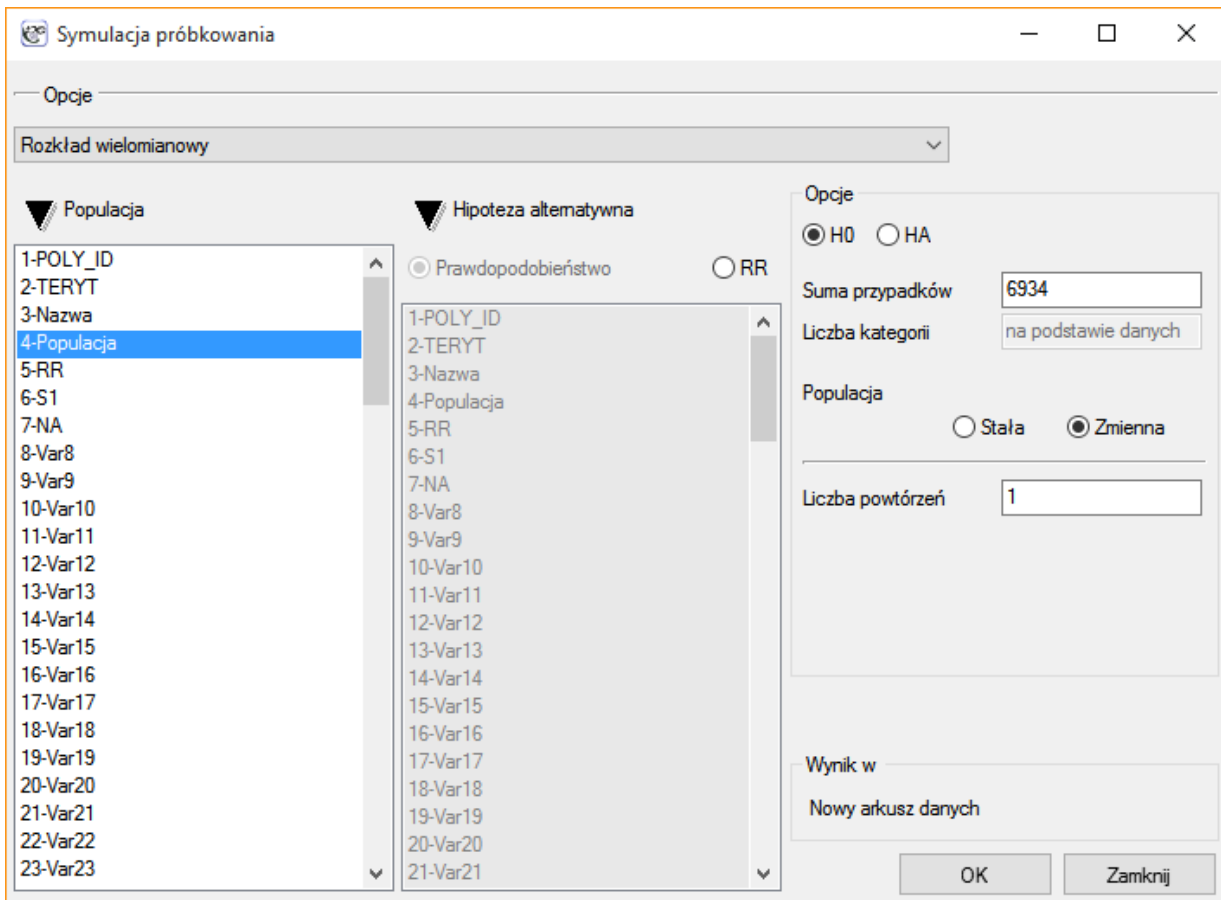
Rodzaje standaryzacji:

- **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.

4.1.14 SYMULACJA PRÓBKOWANIA



Okno próbkowania wywołujemy poprzez Dane→Symulacja próbkowania ...



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, tożsamy jest z rozkładem dla H0.

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

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


PRZYKŁAD 4.3. (plik symulacje.pqs)

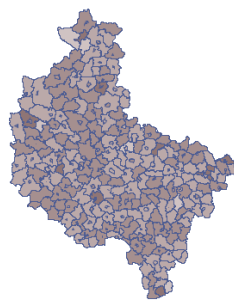
Jako podstawę symulacji użyto populację Wielkopolski w roku 2013, która wynosiła wg GUS $N = 3467016$ osób. Województwo podzielone jest na 315 gmin. Gminy różnią się znacznie liczbą mieszkańców. Najliczniejsza gmina (stolica województwa) 548028 mieszkańców, najmniej liczna 1454 mieszkańców, mediana i 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")

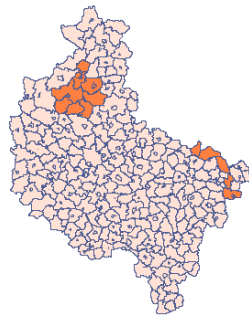
Ad 1. Należy zaznaczyć, że równomierne rozłożenie 6934 chorych w sposób losowy nie oznacza podobnej liczby chorych w każdej gminie. Wiadomo, że gminy o większej liczbie narażonych powinny mieć odpowiednią większą liczbę chorych niż te o mniejszej liczbie mieszkańców. Interesujące jest więc takie rozłożenie chorych, by współczynnik liczby chorych do liczby mieszkańców był względnie stały. Oznacza to przyjęcie hipotezy zerowej H_0 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 (podejrzec 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.

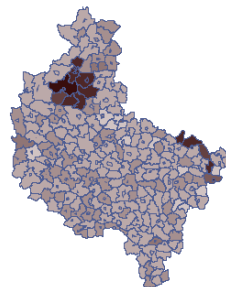


Ad 2. W arkuszu "Klastery" podano, podobnie jak w poprzednim zadaniu, liczność dla populacji badanej. Tym razem oczekujemy wyższej częstości w niektórych gminach (wskazanych na mapie) więc dodatkowo w kolejnej kolumnie arkusza przedstawiono wartość względnego ryzyka dla poszczególnych gmin ustawiając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



Odpowiednie próbkowanie wymaga od nas wybrania hipotezy alternatywnej HA (poprzez wybranie kolumny z relatywnym ryzykiem) i zmiennej populacji (poprzez wskazanie kolumny z licznoscią populacji gmin). Wylosowane na podstawie tych zalożeń 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.

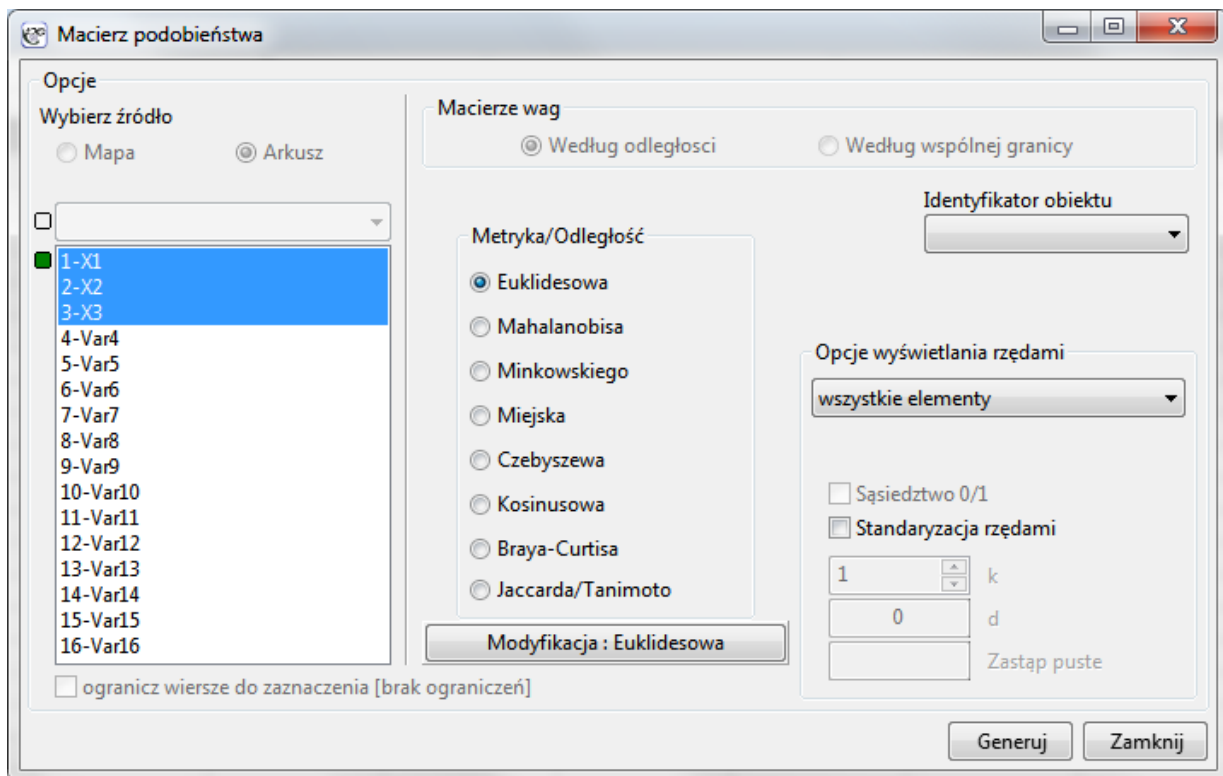


4.1.15 MACIERZ PODOBIEŃSTWA



Wzajemne relacje między obiektami mogą być wyrażone przez ich odległości lub bardziej ogólnie poprzez niepodobieństwo. Czym dalej od siebie znajdują się obiekty, tym bardziej są do siebie niepodobne, im bliżej natomiast, tym podobieństwo między nimi jest większe. Badać można odległość obiektów pod względem wielu cech np. gdy porównywane obiekty to miasta ich podobieństwo możemy definiować między innymi w oparciu o: długość drogi je łączącej, gęstości zaludnienia, PKB przypadającym na mieszkańca, emisję zanieczyszczeń, przeciętne ceny nieruchomości itd. Mając tak wiele różnych cech badacz tak musi dobrać miarę odległości, by najlepiej obrazowała rzeczywiste podobieństwo obiektów.

Okno z ustawieniami opcji macierzy podobieństwa wywołujemy poprzez menu Dane→Macierz podobieństwa...



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: $d(x, z) \leq d(x, y) + d(y, z)$.

Uwaga!

Metryki powinny być wyliczane dla cech o tych samych zakresach wartości. Gdyby tak nie było to cechy o wyższych zakresach miałyby większy wpływ na uzyskany wynik podobieństwa niż te o niższych zakresach. Przykładowo, wyliczając podobieństwo osób możemy oprzeć je na takich cechach jak min. masa ciała i wiek. Wówczas masa ciała w kilogramach, w zakresie od 40 do 150 kg, będzie miała większy wpływ na wynik niż wiek w latach, w zakresie od 18 do 90 lat. By wpływ każdej cechy na uzyskany wynik podobieństwa był zrównoważony powinniśmy każdą z nich **znormalizować/wystandaryzować** przed przystąpieniem do analizy. Chcąc natomiast samodzielnie zdecydować o wielkości tego wpływu, po zastosowaniu standaryzacji, wskazując rodzaj metryki należy wpisać nadane przez siebie 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:

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

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 diagonalu). 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 wartości ujemne powinniśmy 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 między nimi nie jest duża $d(\text{tekst}_1, \text{tekst}_2) = 1 - 0,784465 = 0,215535$.

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, powinniśmy 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.

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

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

$$BC = 1 - d(x_1, x_2) \quad (9)$$

Jaccarda

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

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

Odległość Jaccarda wyraża się wzorem:

$$d(x_1, x_2) = 1 - J. \quad (10)$$

gdzie:

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

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 oznaczmy 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 nierozdzielne 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

Najbardziej podobni są osobniki 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 4.4. (plik: podobienstwoLokali.pqs)

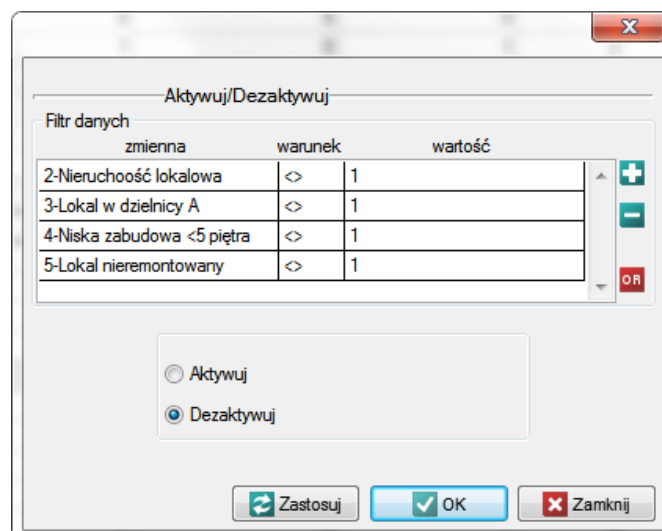
W procedurach wyceny nieruchomości, zarówno ze względów merytorycznych jak też prawnych, kwestia podobieństwa pełni ważną rolę. Jest na przykład zasadniczą przesłanką umożliwiającą grupowanie obiektów i przypisywanie do odpowiedniego segmentu.

Założmy, 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**).

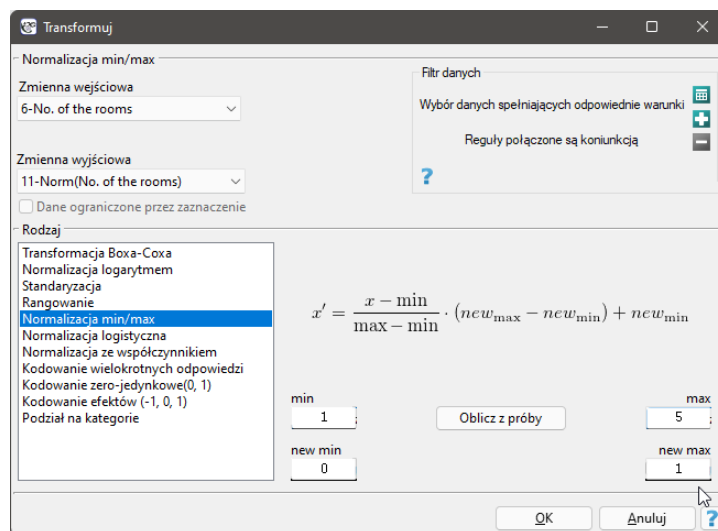
W rezultacie wyłoniono 11 lokali (lokal 10, 12, 17, 35, 88, 101, 105, 122, 130, 132, 135) pasujących do tego segmentu (spełniających wszystkie 4 warunki).

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.

	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", "Wiek budynku", "Dystansu do centrum" i "Odległości przystanku".



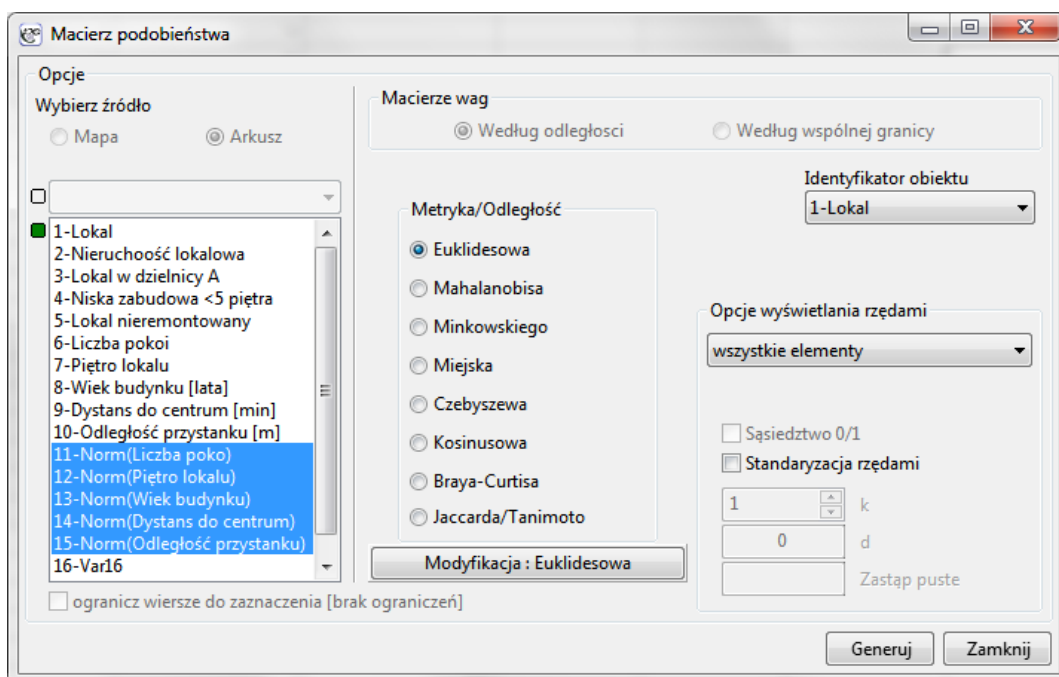
Znormalizowane dane przedstawia poniższa tabela.

	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 wyznaczmy 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:

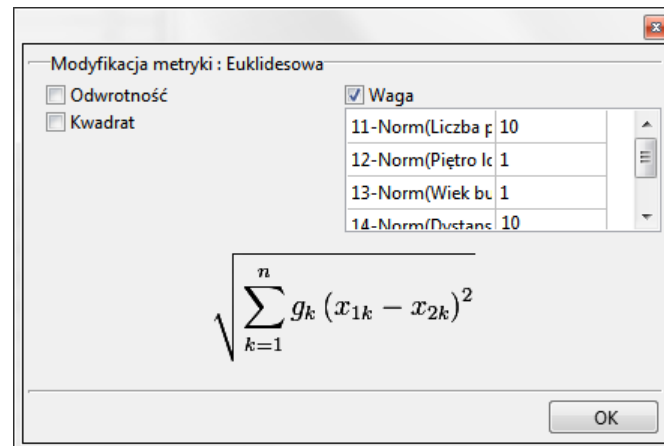
- (1) W pierwszej macierzy znajdują się odległości euklidesowe wyliczone na podstawie równoważnie traktowanych pięciu cech;
- (2) W drugiej macierzy znajdują się odległości euklidesowe, w budowie których największe znaczenie będzie miała liczba pokoi i dystans do centrum dzielnicy.

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



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:

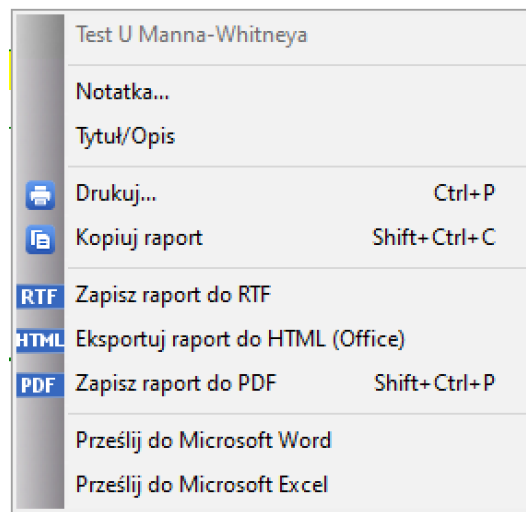
Euklidesowa	Poszukiwany	...	Euklidesowa z wagami	Poszukiwany	...
Poszukiwany	0	...	Poszukiwany	0	...
Lokal 10	1.10	...	Lokal 10	3.35	...
Lokal 12	1.31	...	Lokal 12	3.84	...
Lokal 17	1.04	...	Lokal 17	1.44	...
Lokal 35	0.96	...	Lokal 35	2.86	...
Lokal 88	1.23	...	Lokal 88	2.78	...
Lokal 101	1.38	...	Lokal 101	3.45	...
Lokal 105	1.18	...	Lokal 105	3.37	...
Lokal 122	1.12	...	Lokal 122	3.39	...
Lokal 130	1.32	...	Lokal 130	3.43	...
Lokal 132	1.24	...	Lokal 132	1.24	...
Lokal 135	0.92	...	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.

4.2 PRACA Z ARKUSZAMI WYNIKÓW (RAPORTAMI)

Raport jest elementem projektu służącym do przechowywania wyników wykonanych analiz statystycznych. Jest on dodawany do projektu automatycznie i przypisywany do aktywnego arkusza danych w chwili zakończenia wybranej procedury statystycznej. Raport nie podlega edycji, za wyjątkiem wykresu i tytułu. Edycję wykresu uruchamiamy dwukrotnym kliknięciem myszy lub poprzez menu kontekstowe prawego przycisku myszy. Edycja tytułu wykonywana jest w [Zarządcy Projektu](#) poprzez dodanie lub zmianę opisu.

Najważniejsze operacje dotyczące raportu mogą być wykonywane poprzez menu kontekstowe prawego przycisku myszy w oknie raportu:




- **Drukowanie**

Opcje drukowania dostępne są poprzez:

- menu kontekstowe,
- menu Plik→Drukuj...

- **Eksportowanie i przesyłanie raportów**

Raporty utworzone w PQStat można eksportować do pliku w formacie *.rtf, *.html i *.pdf. Można je również przesłać do dokumentów programu Word lub Excel.

Jeśli eksportu dokonujemy w [Zarządcy Projektu](#), to raporty możemy umieścić w oddzielnych plikach lub w jednym wspólnym pliku. W tym celu zaznaczamy wybrane raporty a następnie wybieramy przycisk  i eksportujemy do pliku lub plików o wybranym formacie. Eksportu poszczególnych raportów możemy dokonać oddzielnie poprzez menu kontekstowe w oknie raportu.

- **Opisywanie raportów**

Raporty możemy opisać w [Zarządcy Projektu](#) lub oknie raportu dodając tytuł lub notatkę.

- **Edycja wykresów**

Edycja wykresu dotycząca jego opcji ogólnych i właściwych dostępna jest poprzez menu kontekstowe w oknie raportu.

- **Kopiowanie raportów**

Za pomocą schowka systemowego można również przenosić wyniki analiz do innych aplikacji np. Word, Excel.

- **Usuwanie raportów**

Usunięcie raportu możliwe jest poprzez:

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

Należy jednak pamiętać, że jeśli do raportu dołączone są warstwy mapy, to usuwając raport usuwa się jednocześnie wszystkie przypisane do niego warstwy.

Zmiana kolejności raportów jest możliwa poprzez menu kontekstowe prawego przycisku myszy Przenieś w górę (Ctrl+Up) lub Przenieś w dół (Ctrl+Down) na nazwie raportu w **Drzewie nawigacji**.

Dodanie informacji do nazwy raportu w Drzewie nawigacji takich jak:

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

jest możliwe po wybraniu odpowiedniej opcji w oknie **ustawień programu**.

4.3 MENU I USTAWIENIA JĘZYKOWE PROGRAMU

4.3.1 JĘZYK

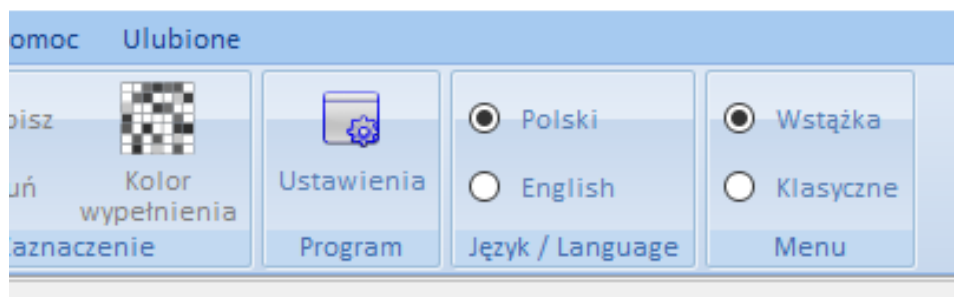
Zarówno interfejs programu jak i tworzone raporty mogą być wyświetlane w języku polskim jak i języku angielskim. Zmiana wybranego języka nie wymaga ponownego uruchomienia programu i możliwa jest poprzez wybranie Language/ Język z menu Edycja. Raporty otwierane po dokonaniu zmiany języka będą automatycznie przetłumaczone (za wyjątkiem nazwy procedury, która stanowi opis i podlega edycji użytkownika).

4.3.2 MENU

Menu programu może być wyświetlone w formie **klasycznej** lub jako **menu wstążkowe**. Sposób wyświetlania menu można przełączyć wskazując menu Edycja a następnie wybrany rodzaj menu.

Ulubione

Menu Ulubione umożliwia szybki dostęp do często wykonywanych statystyk i innych operacji. Do projektowania na własne potrzeby menu służy okno wyszukiwarki Ulubionych - za jego pośrednictwem możemy wyszukać dowolną pozycję menu a następnie dodać ją do menu lub z niego usunąć.



5 ORGANIZACJA PRACY Z PROGRAMEM

Wszystkie procedury analiz statystycznych dostępne są w menu Statystyka.

5.1 FORMA ORGANIZACJI DANYCH

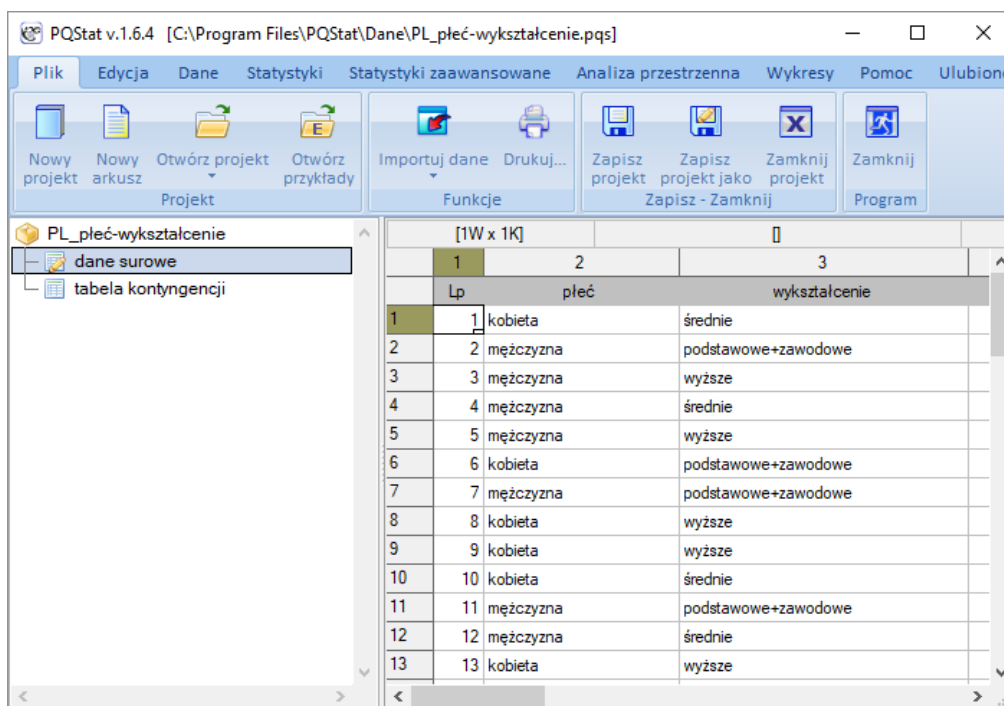
Forma organizacji danych zależy od procedur statystycznych, jakie użytkownik planuje wykonać.

Analiza statystyczna danych może dotyczyć danych zebranych w postaci tabeli kontyngencji lub danych w postaci surowej. Przy czym istnieje możliwość zamiany danych:

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

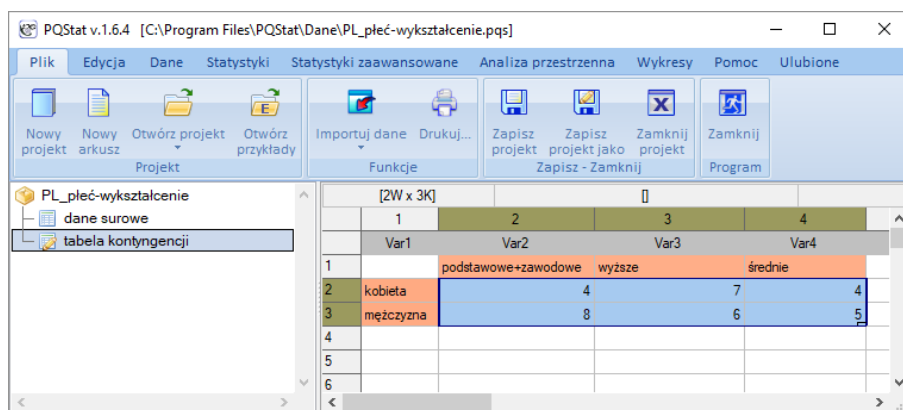
PRZYKŁAD 5.1. Dane surowe (plik płęć-wykształcenie.pqs)



	1	2	3
	Lp	płeć	wykształcenie
1	1	kobieta	średnie
2	2	mężczyzna	podstawowe+zawodowe
3	3	mężczyzna	wyższe
4	4	mężczyzna	średnie
5	5	mężczyzna	wyższe
6	6	kobieta	podstawowe+zawodowe
7	7	mężczyzna	podstawowe+zawodowe
8	8	kobieta	wyższe
9	9	kobieta	wyższe
10	10	kobieta	średnie
11	11	mężczyzna	podstawowe+zawodowe
12	12	mężczyzna	średnie
13	13	kobieta	wyższe

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

PRZYKŁAD 5.2. Tabela kontyngencji (plik płęć-wykształcenie.pqs)



The screenshot shows the PQStat v.1.6.4 software interface. The main window displays a contingency table with the following data:

	1	2	3	4	
	Var1	Var2	Var3	Var4	
1		podstawowe+zawodowe	wyzsze	średnie	
2	kobieta		4	7	4
3	mężczyzna		8	6	5
4					
5					
6					

5.2 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:

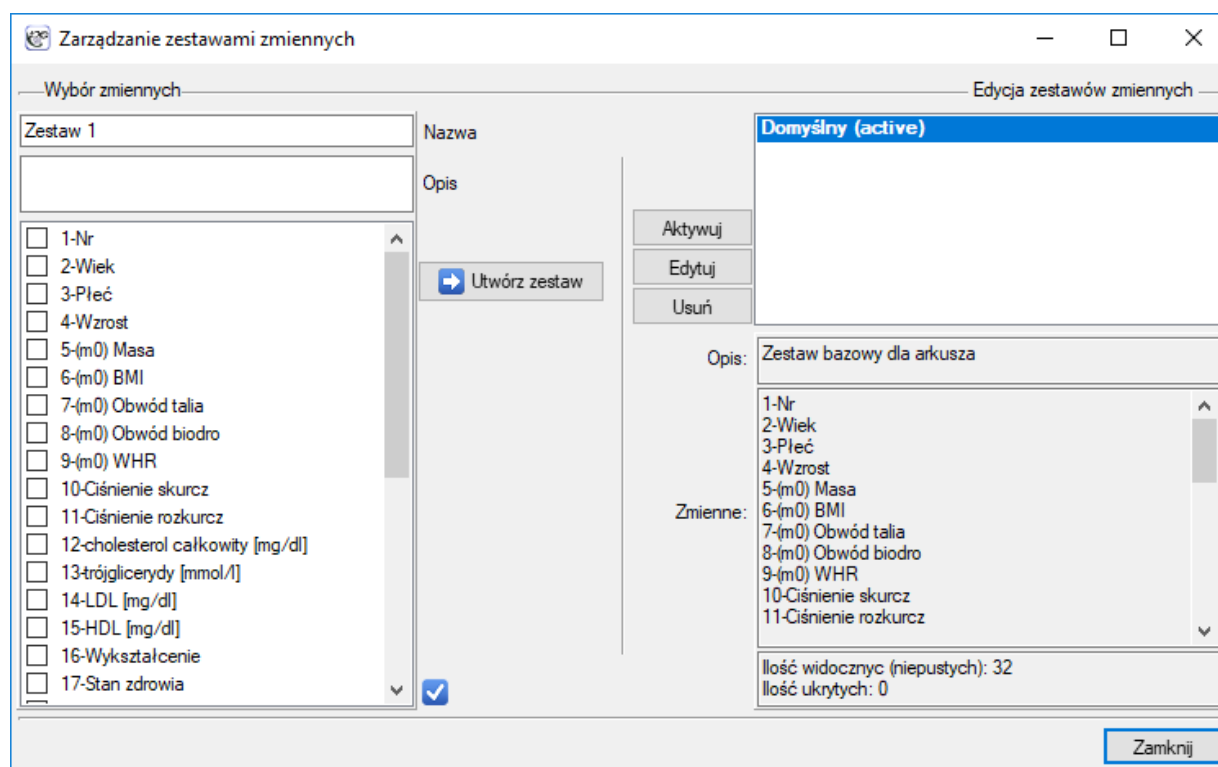
5.2.1 Organizowanie zmiennych w zestawy

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.

Edycja zestawów zmiennych

By edytować zestaw w oknie Zarządzania zestawami zmiennych wybieram zestaw z listy zestawów i potwierdzam 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 potwierdzam wybór przyciskiem Zapisz zmiany.

Aktywowanie wybranego zestawu zmiennych

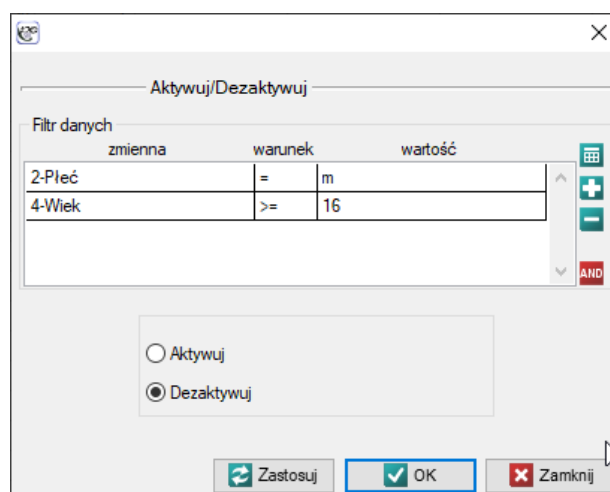
Aktywowanie zestawu dokonywane jest w oknie Zarządzania zestawami zmiennych lub poprzez menu kontekstowe w górnej części arkusza danych na nazwie wybranego zestawu.

5.2.2 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 5.3. (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. Pamiętajmy: aby wykonać to zadanie poprawnie wszystkie warunki filtru powinny być połączone koniunkcją (informuje o tym znak ). 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 

Więcej przykładów wykorzystania filtru znajdziesz [tutaj](#).

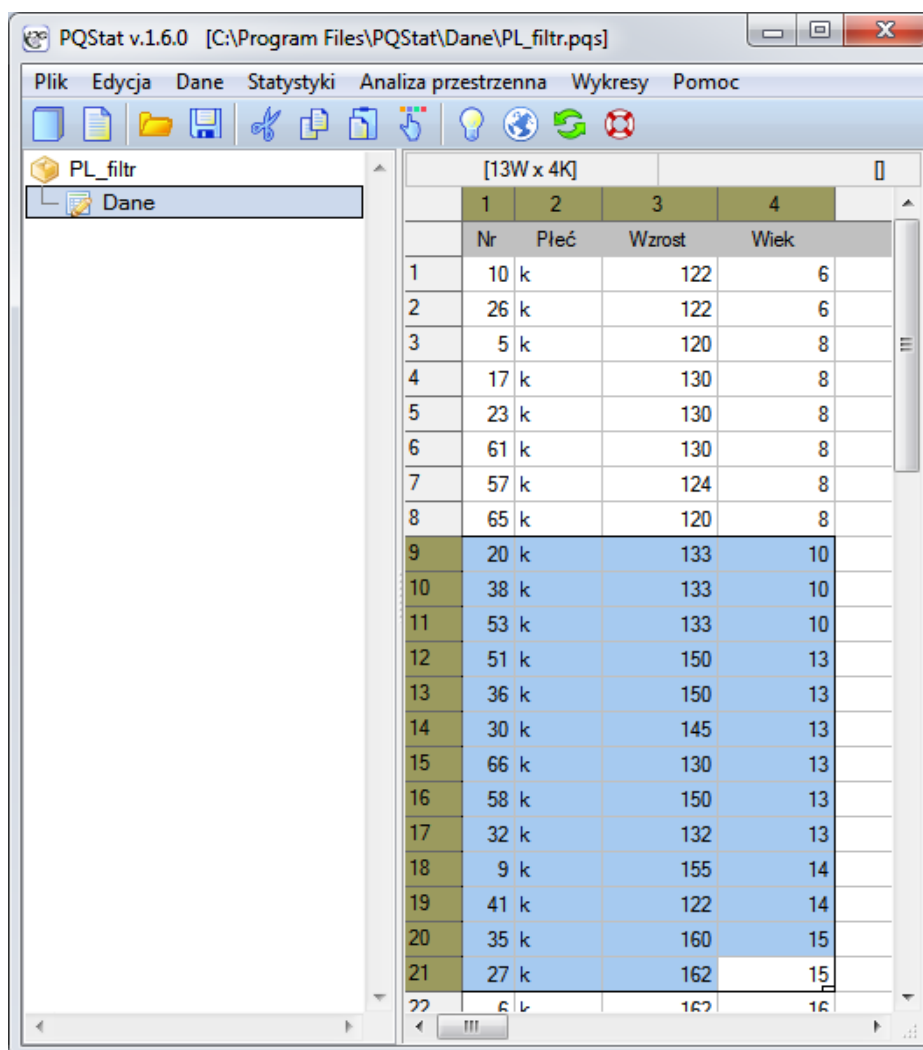
By aktywować wszystkie przypadki należy wybrać menu Edycja → Aktywuj wszystkie

5.2.3 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 5.4. (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 następnie 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.



	1	2	3	4
	Nr	Płeć	Wzrost	Wiek
1	10	k	122	6
2	26	k	122	6
3	5	k	120	8
4	17	k	130	8
5	23	k	130	8
6	61	k	130	8
7	57	k	124	8
8	65	k	120	8
9	20	k	133	10
10	38	k	133	10
11	53	k	133	10
12	51	k	150	13
13	36	k	150	13
14	30	k	145	13
15	66	k	130	13
16	58	k	150	13
17	32	k	132	13
18	9	k	155	14
19	41	k	122	14
20	35	k	160	15
21	27	k	162	15
22	6	k	162	16

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) i potwierdzamy wybór przyciskiem OK.

Zawężenie obszaru roboczego arkusza danych poprzez zaznaczenie spójnego fragmentu danych powoduje wyświetlenie w oknie analizy komunikatu:

Dane ograniczone przez zaznaczenie

5.2.4 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 5.5. (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 Usun zaznaczenia.

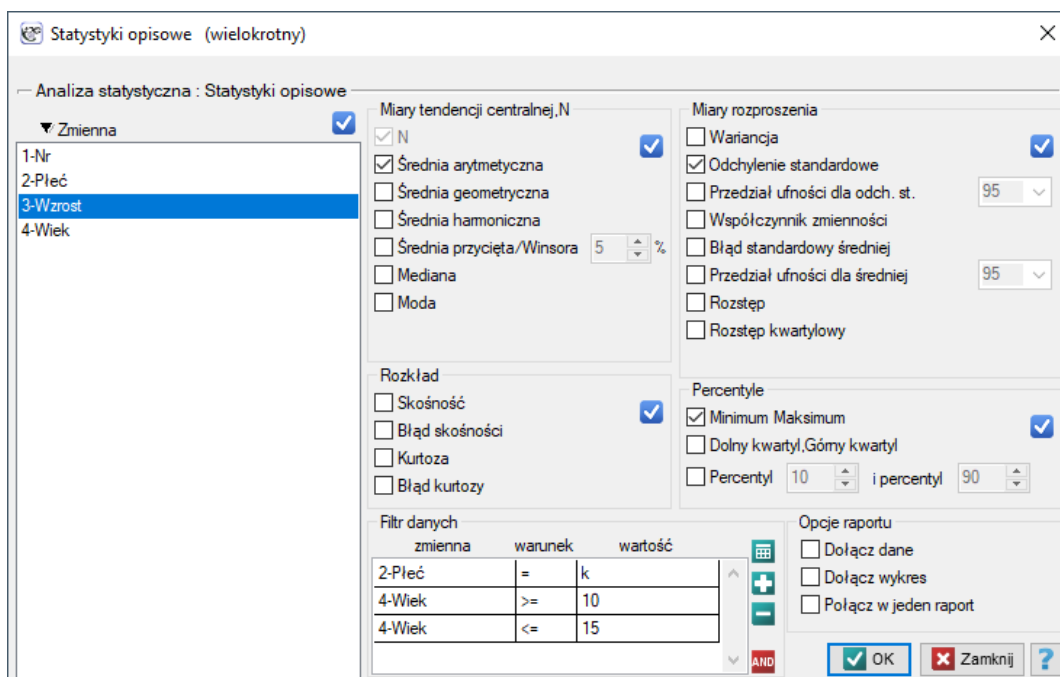
5.2.5 Filtr danych




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.

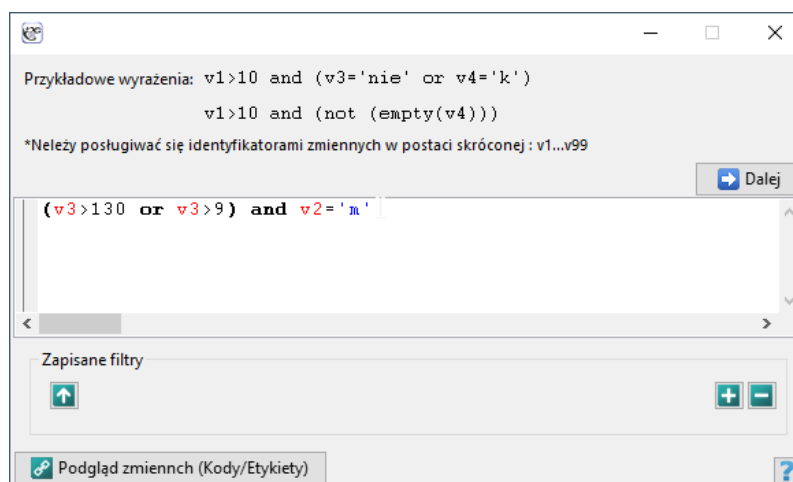
- **Filtr podstawowy** pozwala wybrać jedną, szczególną podgrupę. Wybór może być dokonany na dwa sposoby:
 - **Sposób automatyczny** – w ten sposób możemy nałożyć jeden warunek na zebrane dane używając przycisku , lub kilka warunków wciskając ponownie . Warunki te połączone będą alternatywą lub koniunkcją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

PRZYKŁAD 5.6. Filtr podstawowy automatyczny (plik filtr.pqs)

Chcemy wyznaczyć **statystyki opisowe** dla wzrostu dziewczynek w wieku pomiędzy 10 a 15-tym 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 filtr dla zmiennej wiek, jako warunek wybieramy znak \geq , a jako wartość wpisujemy 10. W podobny sposób dodajemy warunek wieku ≤ 15 . Pamiętajmy: aby wykonać to zadanie poprawnie wszystkie warunki filtru powinny być połączone koniunkcją (informuje o tym znak ). 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 przelącując 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 warunek z podanych jest prawdziwy,



xor - operator alternatywy rozłącznej - sprawdza czy dokładnie jeden warunek z podanych 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 .

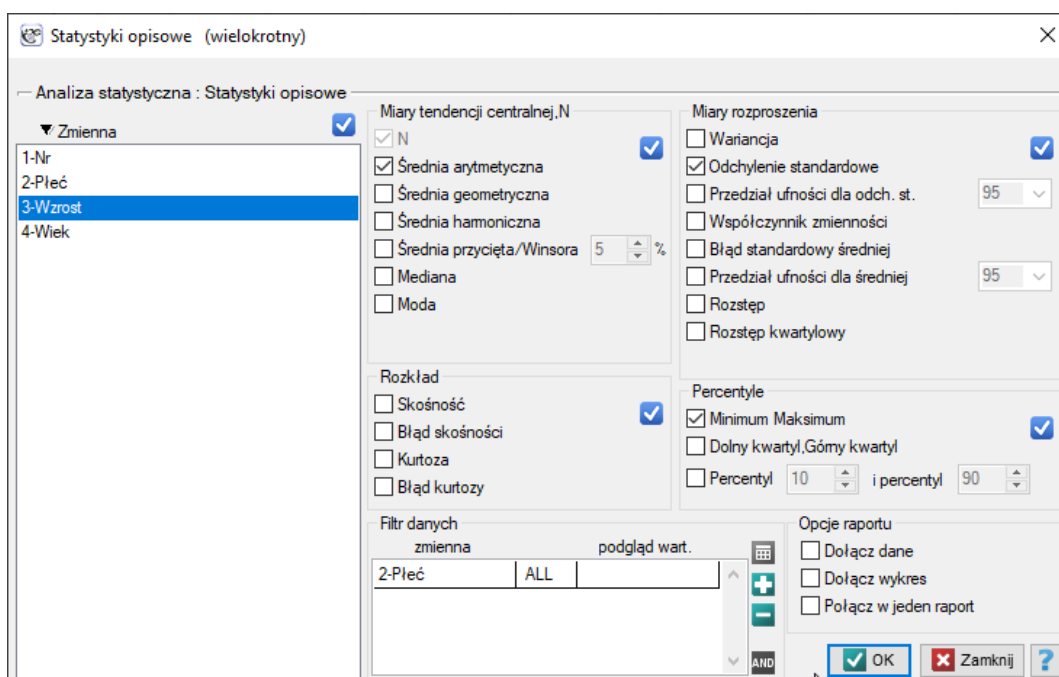
PRZYKŁAD 5.7. Filtr podstawowy zaawansowany (plik filtr.pqs)

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 $(v3>130 \text{ or } v4>9)$ 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 5.8. 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.



5.3 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 'textsDrzewie 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.

5.4 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).

5.5 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 filtra 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.

5.6 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.

 poziom istotności

6 WYKRESY

Program PQStat oferuje wykresy kolumnowe, wykresy błędów, wykresy typu ramka-wąsy, punktowe, liniowo-punktowe.

Okno z ustawieniami opcji wykresów wywołujemy poprzez menu Wykresy.

Zmiana podstawowych parametrów wykresu jest możliwa bezpośrednio w oknie wykresu. Jeśli natomiast:

- chcemy zmienić parametry ogólne wykresu takie jak: tytuły, tła, osie, linie siatki czy legendę – wybieramy zakładkę Ogólne opcje wykresu;
- chcemy zmienić wygląd samego rysowanego obiektu np. kształt, styl, kolory – wybieramy zakładkę Właściwe opcje wykresu;
- chcemy narysować dodatkowe elementy na wykresie np. dorysować linię – wybieramy zakładkę Inne.

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

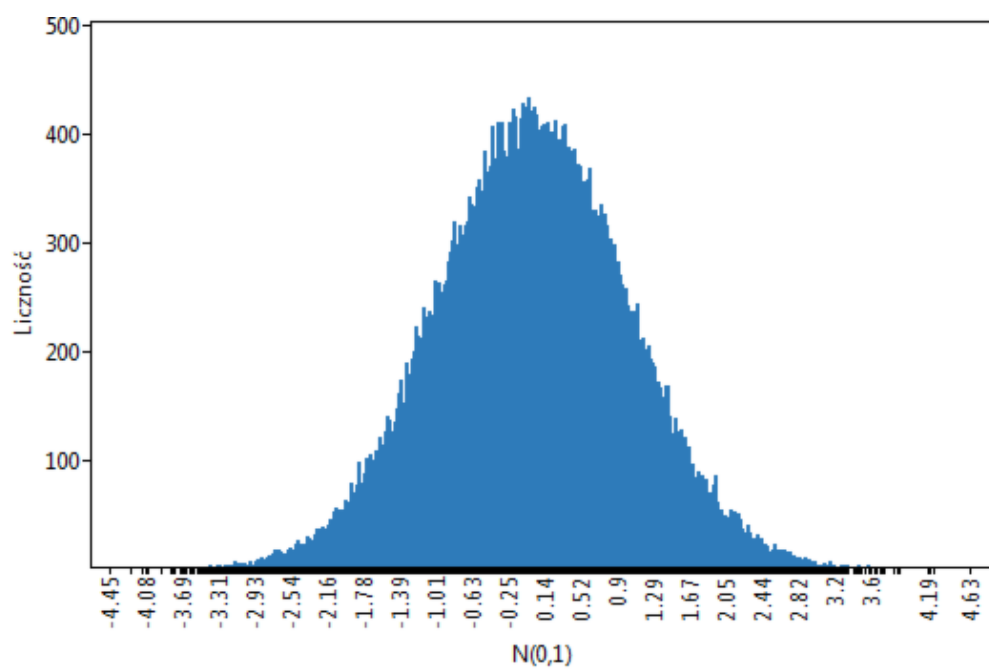
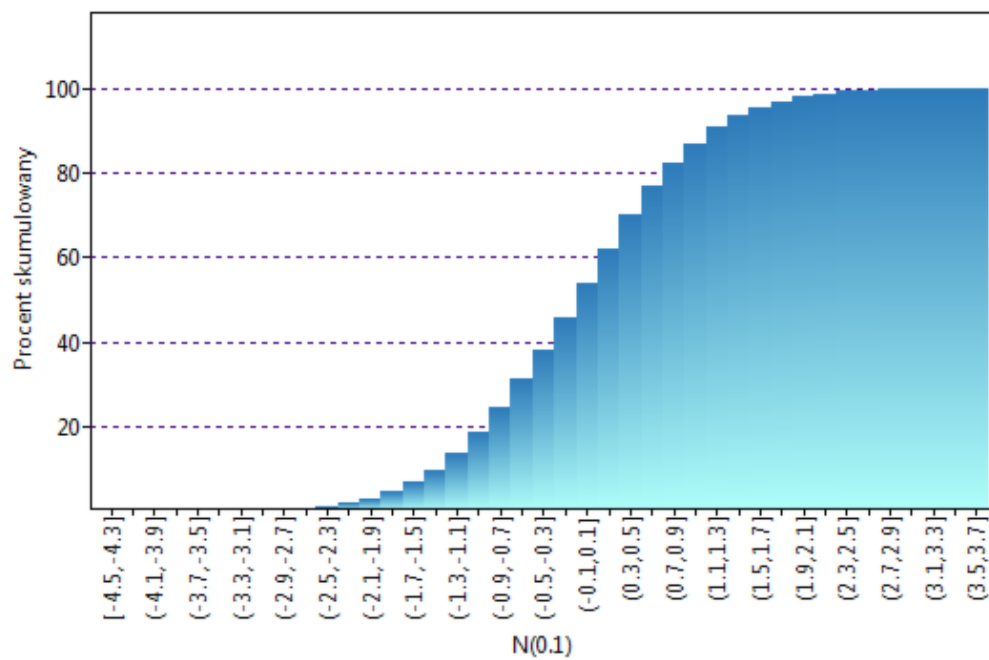
Wykres zwracany jest do raportu, gdzie może być:

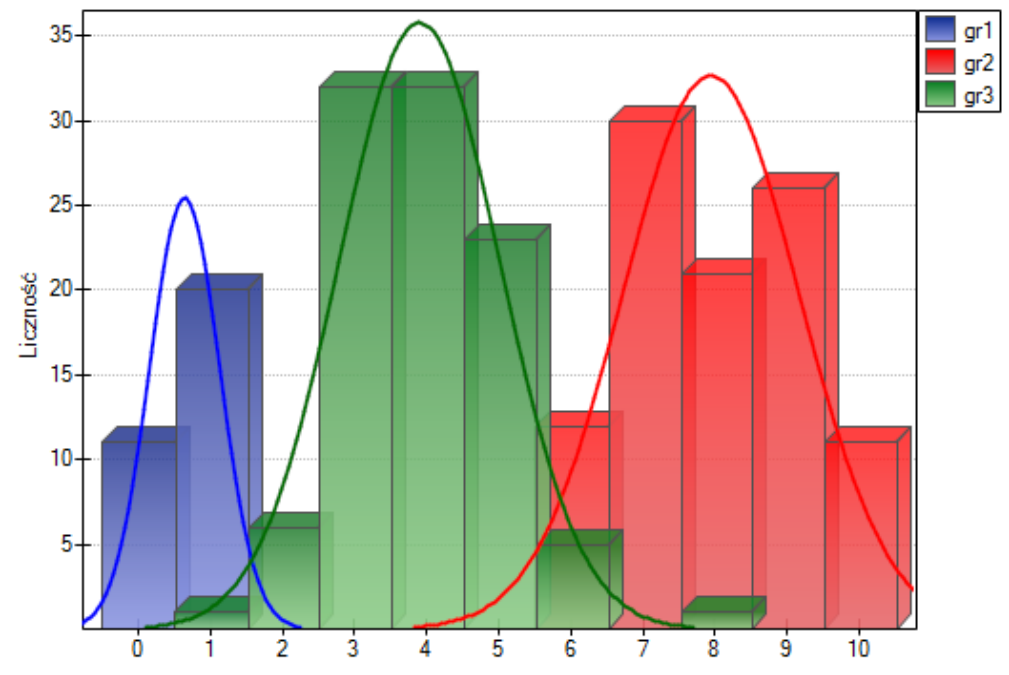
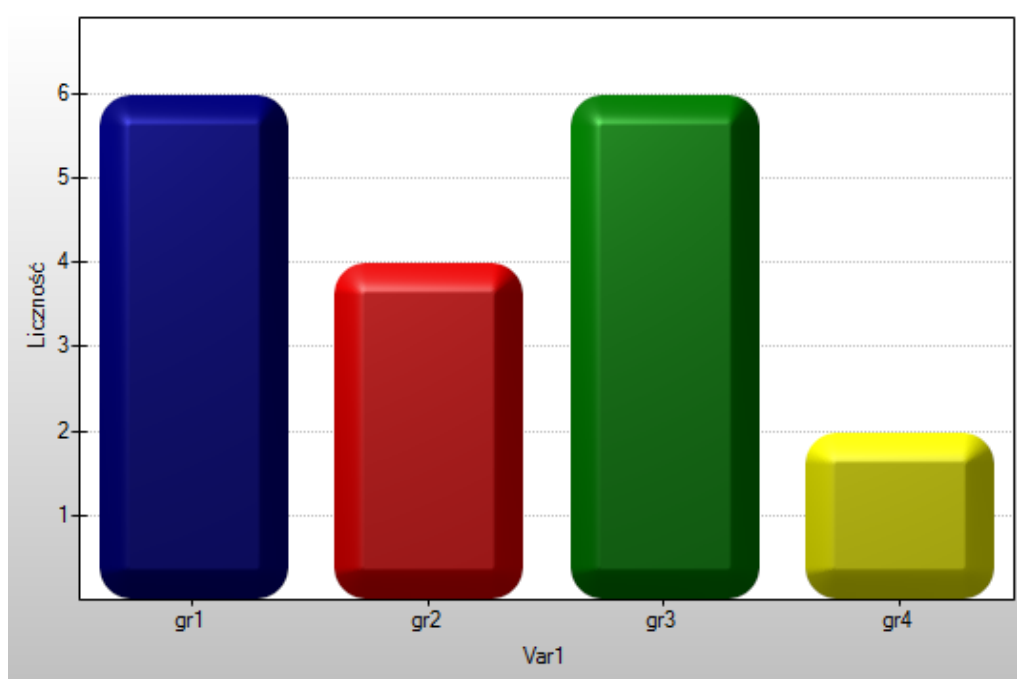
- zapisany – opcja Zapisz wykres jako... z menu kontekstowego;
- wydrukowany – opcja Drukuj wykres z menu kontekstowego;
- skopiowany – opcja Kopiuj wykres z menu kontekstowego;
- poddany edycji – dotyczy to Ogólnych opcji wykresu i Właściwych opcji wykresu. By edytować wykres wystarczy podwójne kliknięcie myszy na wykresie, lub wybranie opcji Edytuj wykres z menu kontekstowego. W oknie edycji wykresu możliwe jest również zapisanie wykresu w wysokiej rozdzielczości.

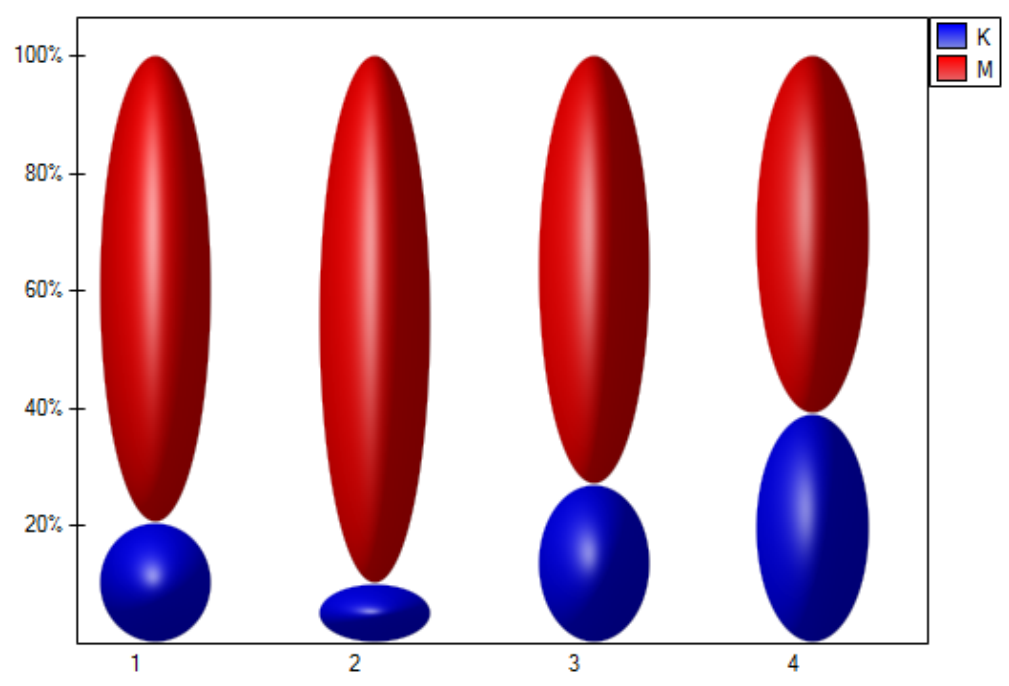
6.1 GALERIA WYKRESÓW

W zależności od rodzaju analizy mamy do wyboru:

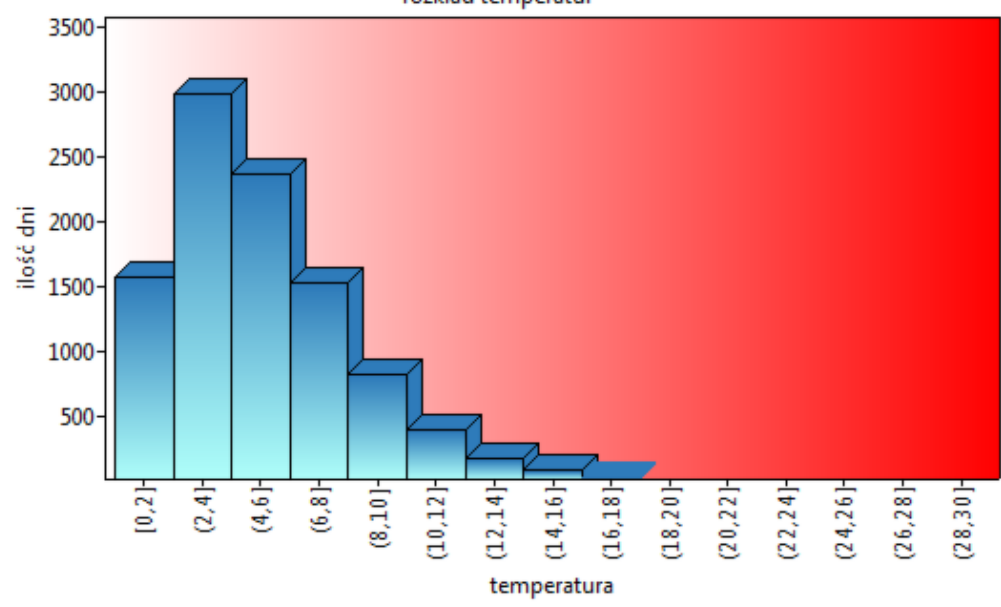
6.1.1 Wykresy kolumnowe

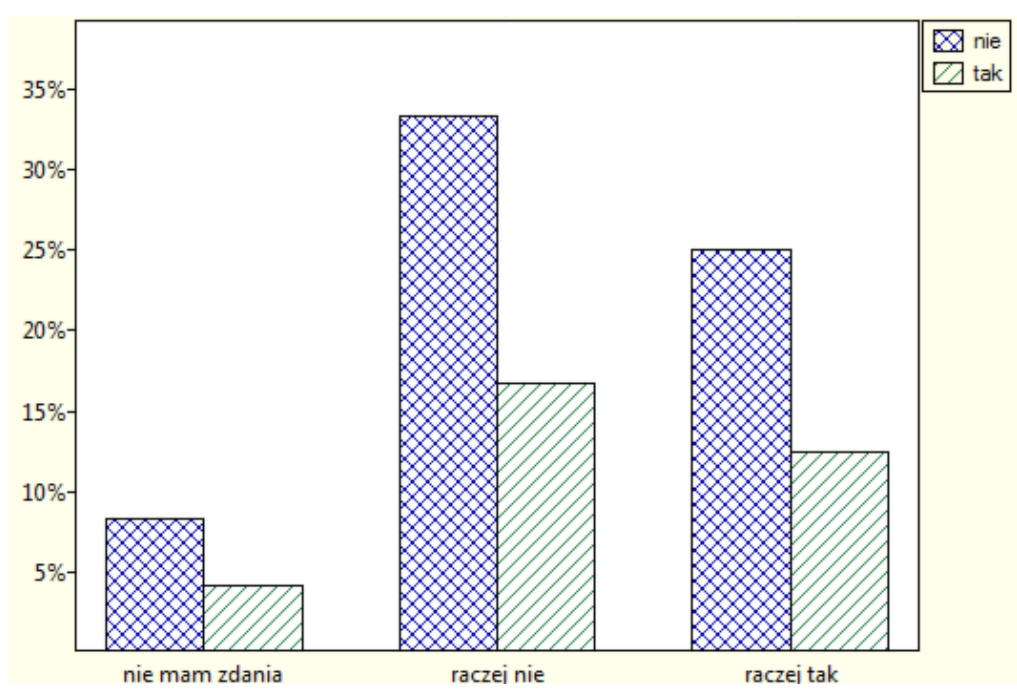
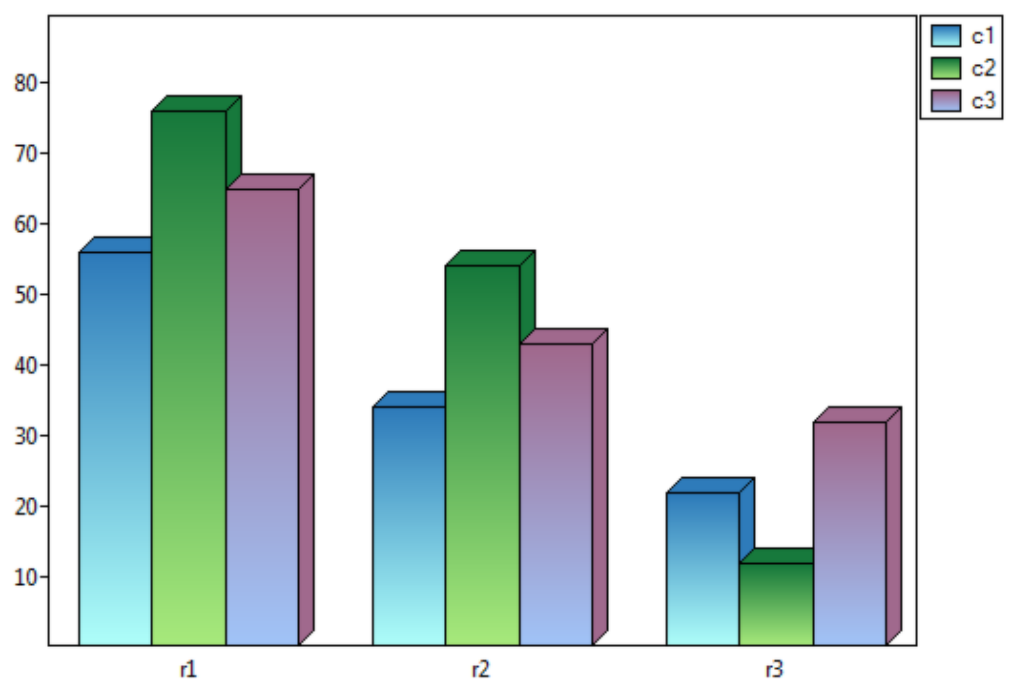


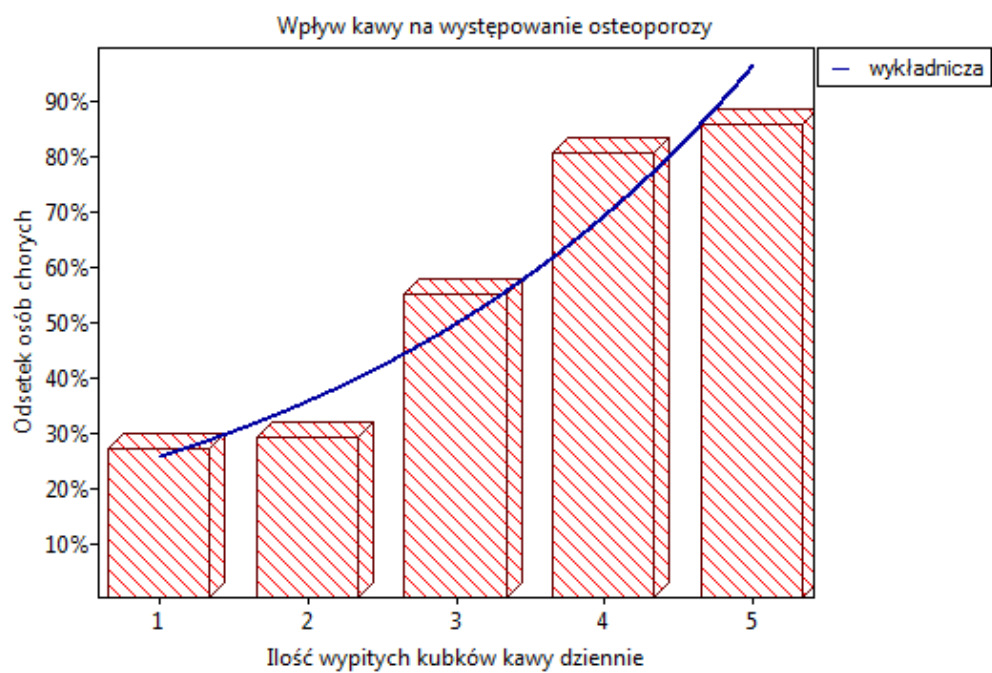
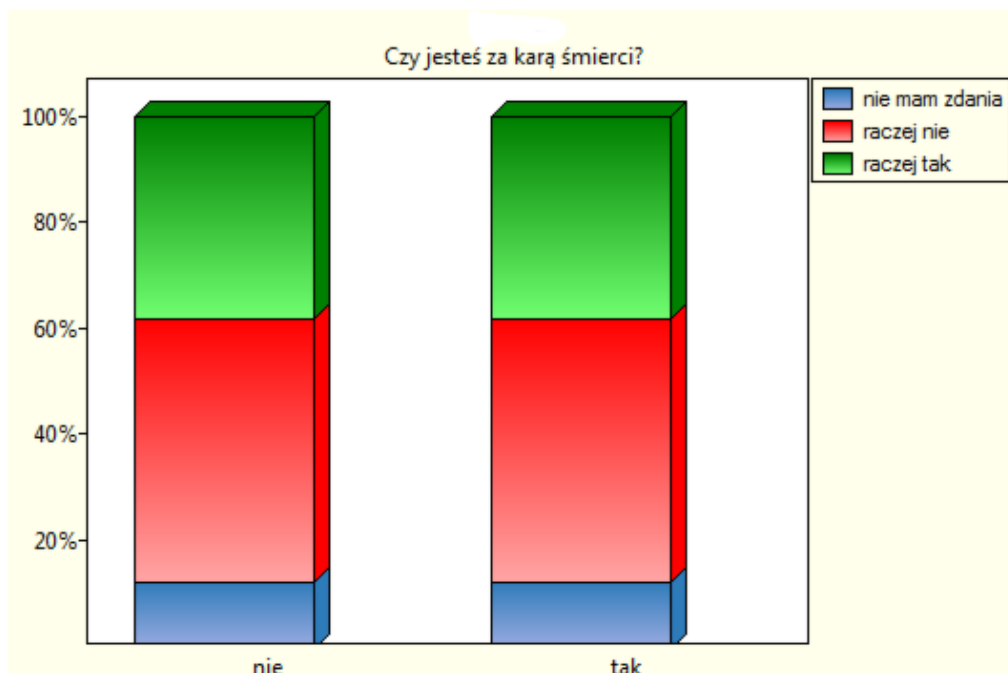




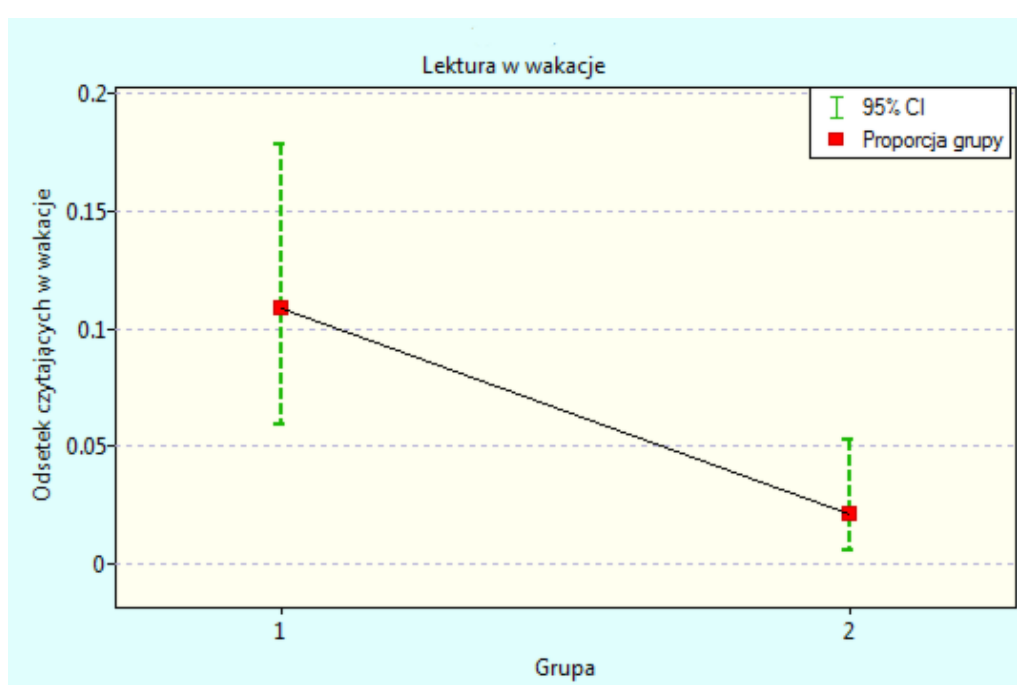
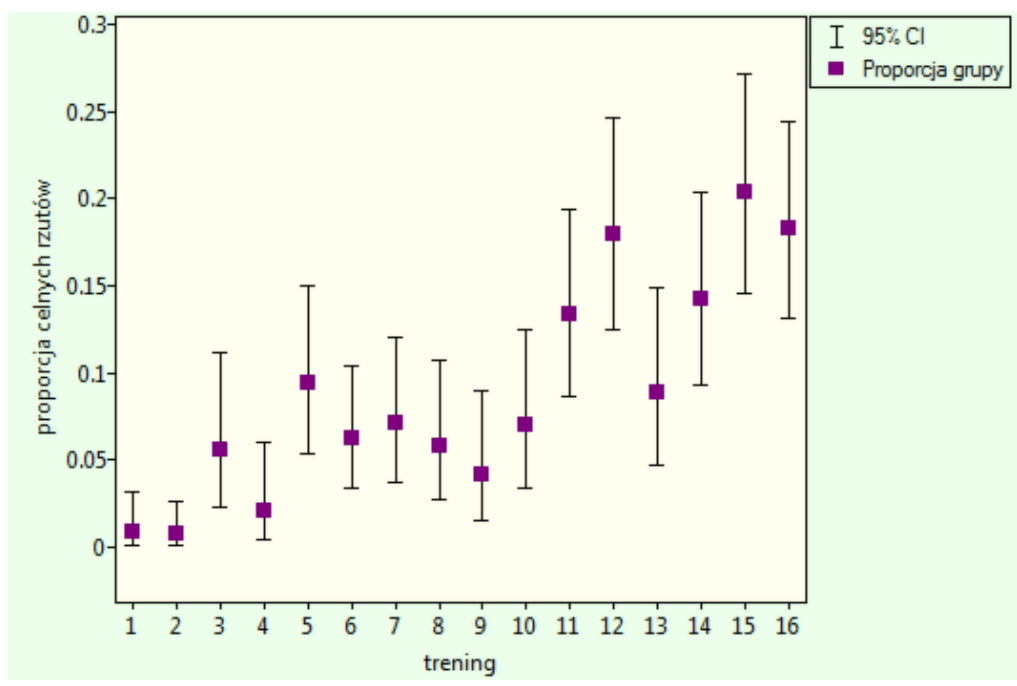
rozkład temperatur

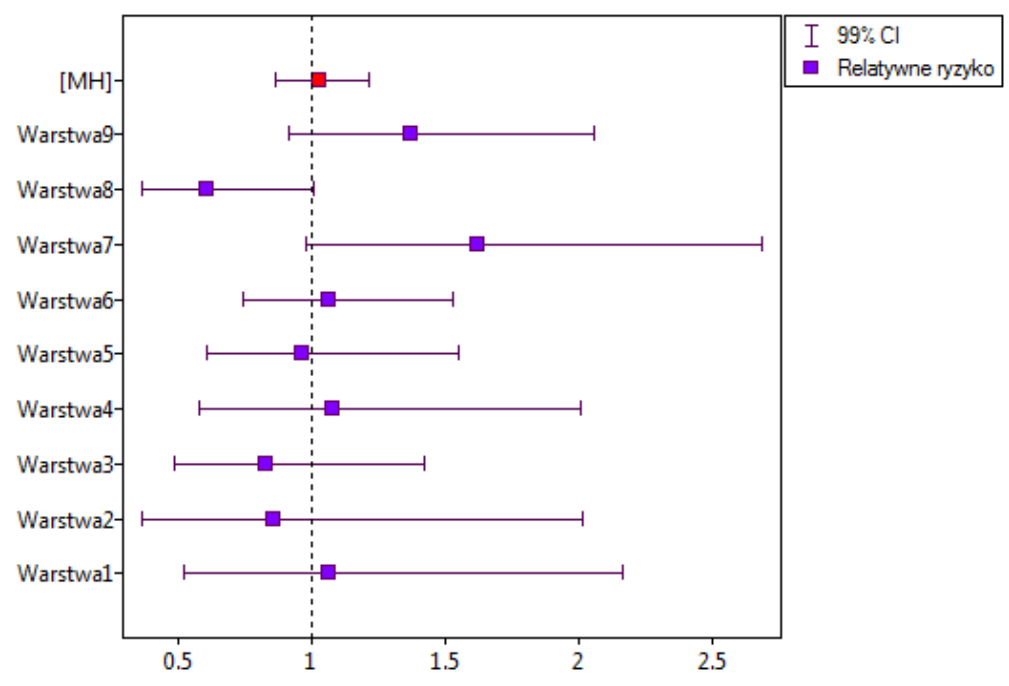




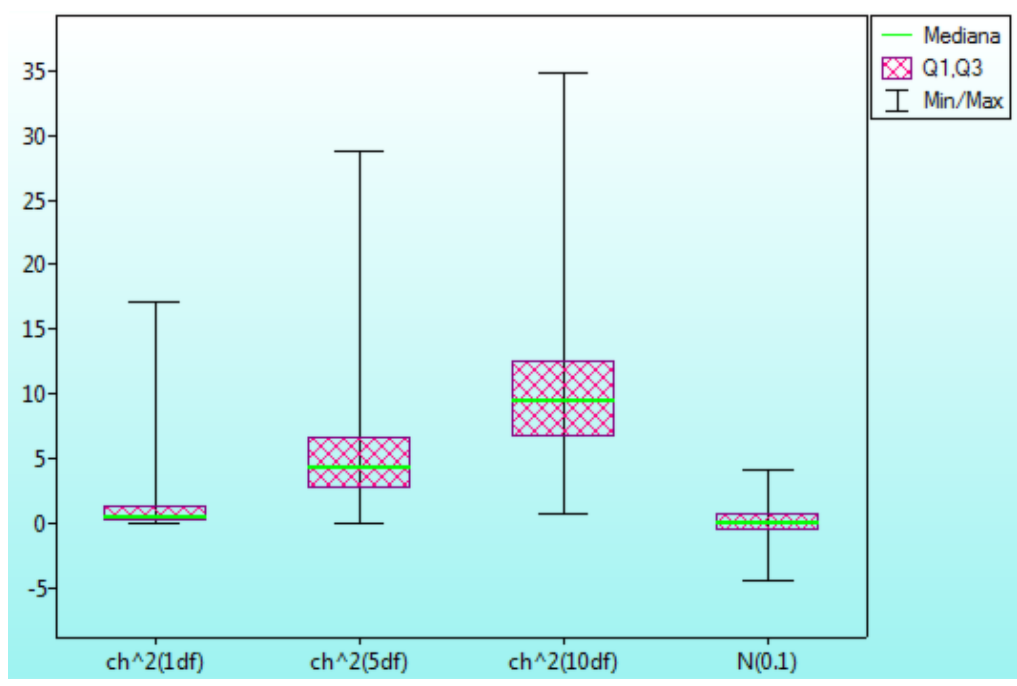


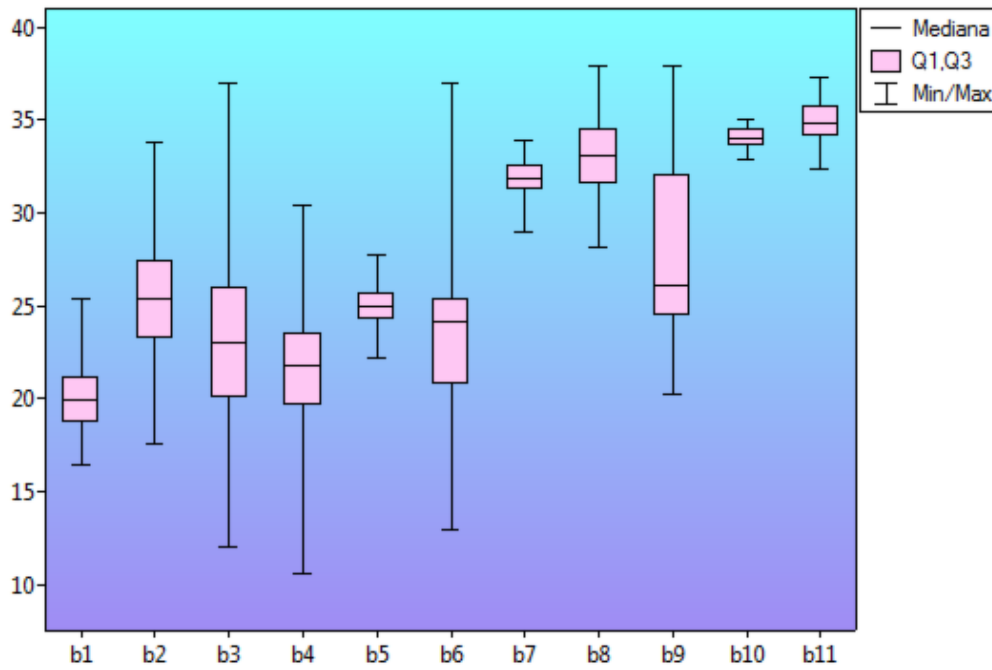
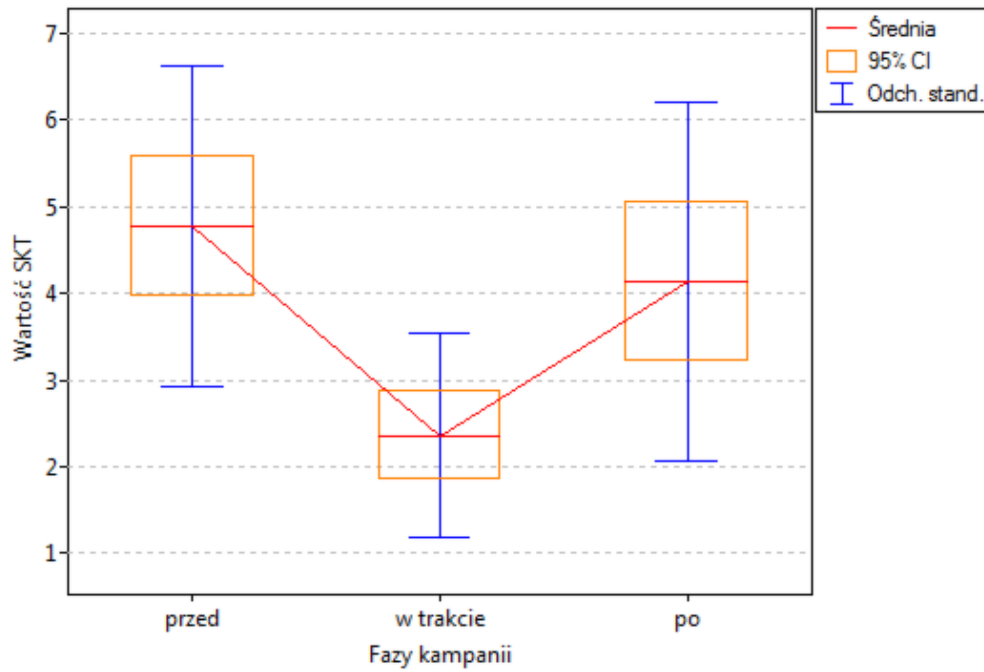
6.1.2 Wykresy błędów estymatora



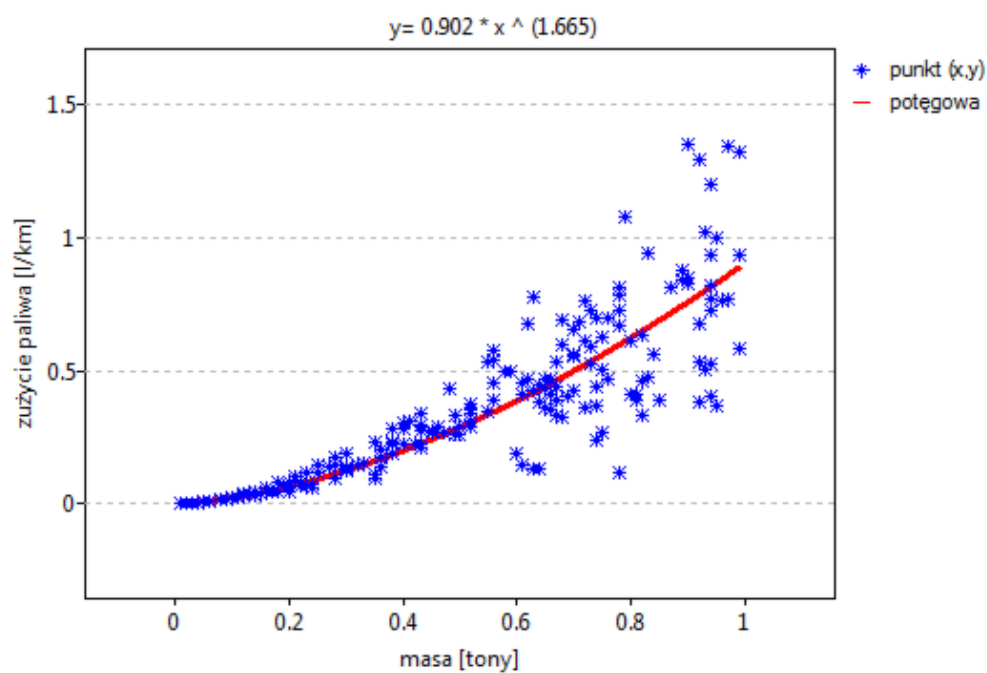
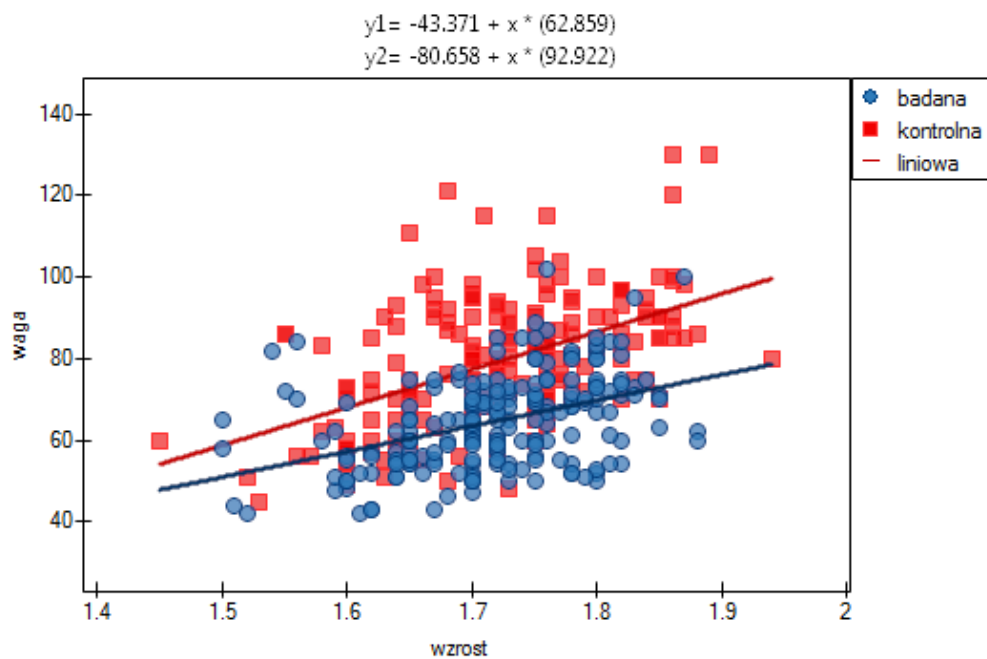


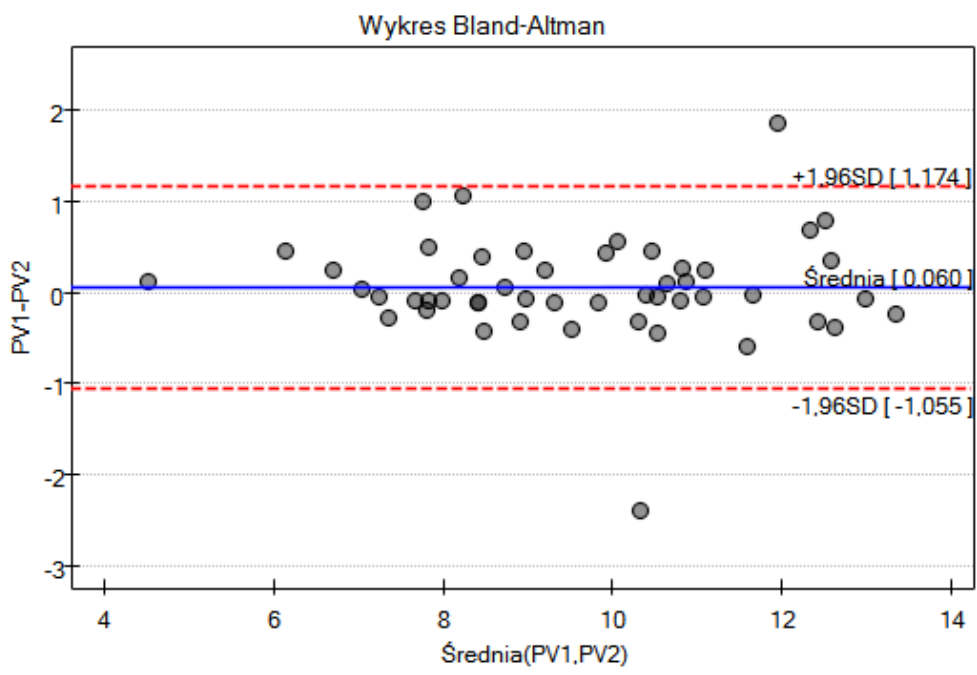
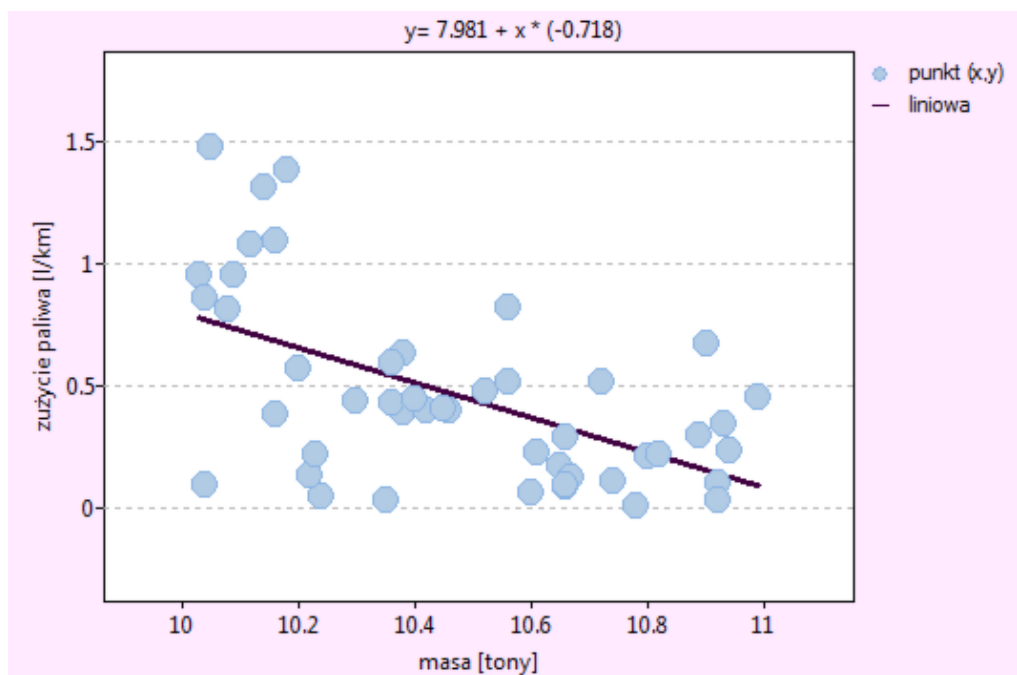
6.1.3 Wykresy typu ramka-wąsy



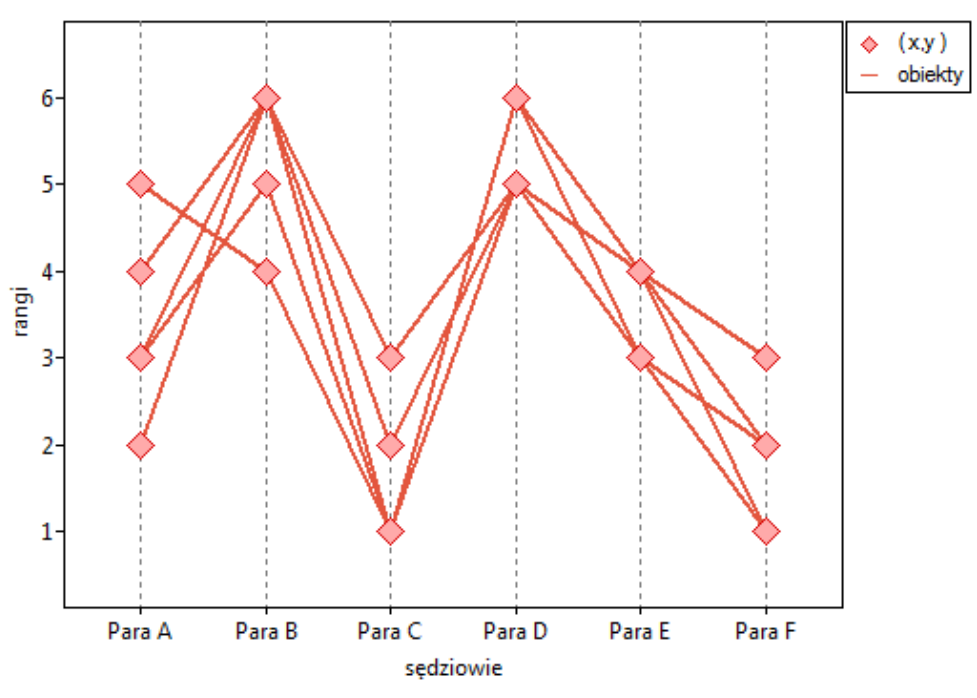
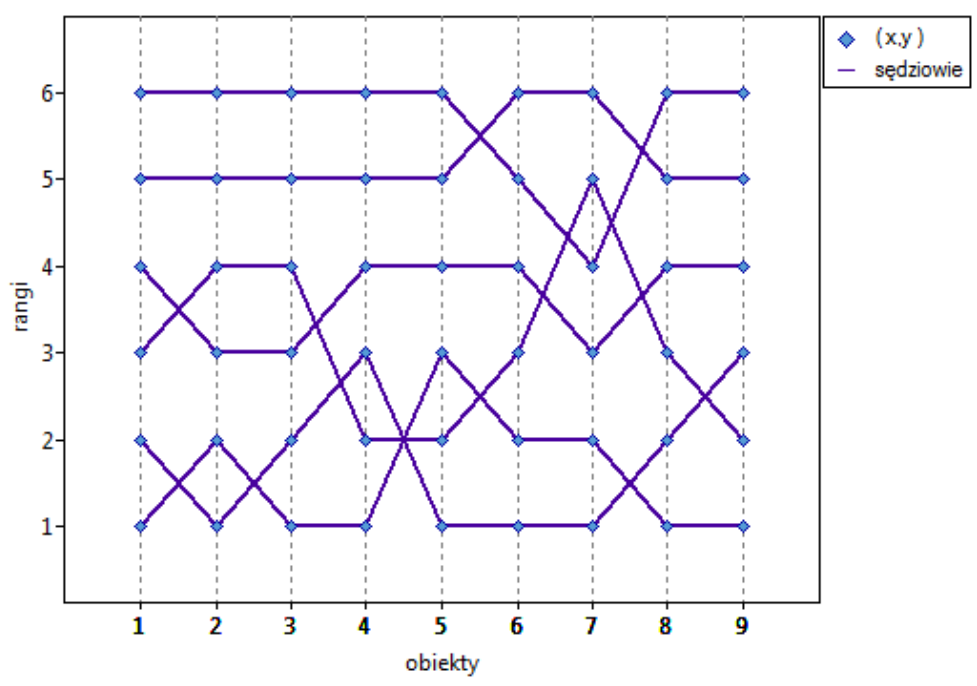


6.1.4 Wykresy punktowe





6.1.5 Wykresy liniowe



6.2 LOKALNE TECHNIKI WYGŁADZANIA LINIOWEGO

6.2.1 LOWESS

LOWESS (ang. locally weighted scatterplot smoothing) znana również pod nazwą LOESS (ang. locally estimated scatterplot smoothing) jest jedną z wielu „nowoczesnych” metod modelowania opartych na metodzie najmniejszych kwadratów. LOESS łączy w sobie prostotę regresji liniowej z elastycznością regresji nieliniowej. Oszacowanie regresji lokalnej (locally weighted regression) zostało niezależnie wprowadzone na kilku różnych polach pod koniec XIX i na początku XX wieku (Henderson, 1916[77]; Schiaparelli, 1866[146]). W literaturze statystycznej metodę tę wprowadzono niezależnie z różnych punktów widzenia pod koniec lat siedemdziesiątych min. Cleveland, 1979[37]. Metoda wykorzystana w programie PQstat opiera się na tej właśnie pracy. 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 .

Algorytm postępowania:

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.

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

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.

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.

6.2.2 Wygładzanie jądrowe

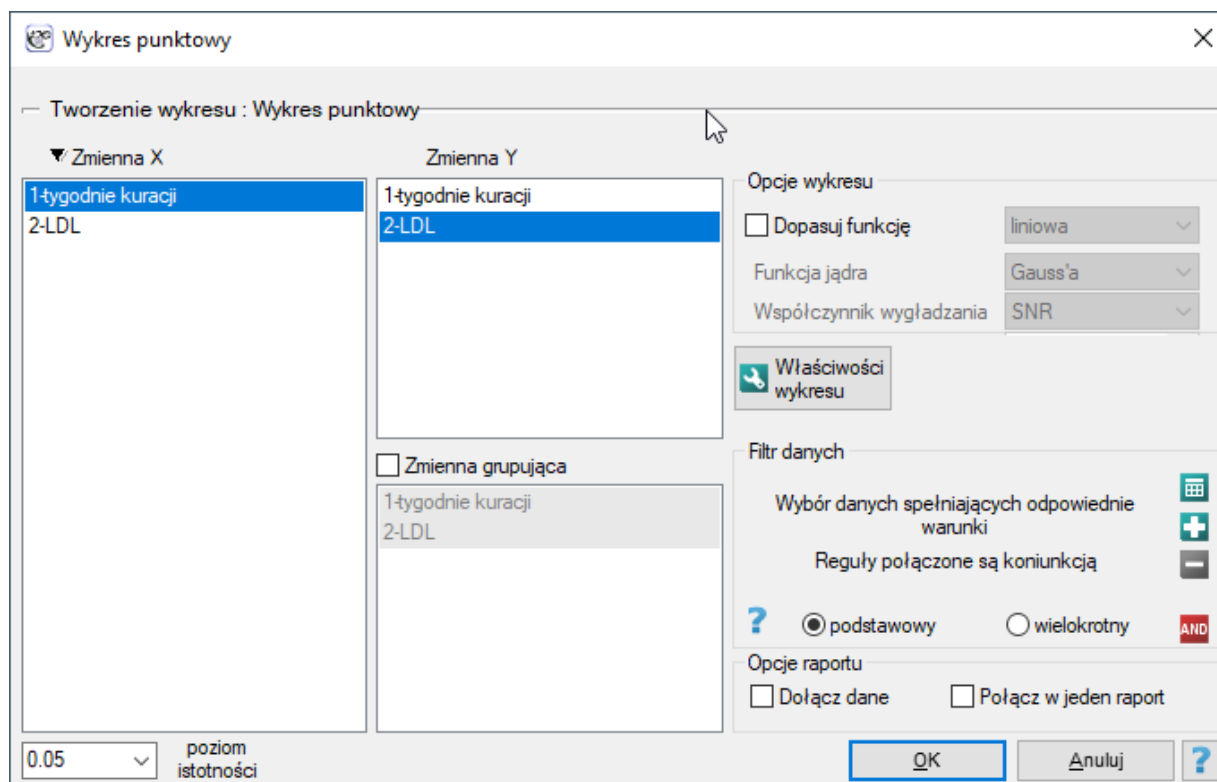
Oszacowanie funkcji regresji metodą wygładzania jądrowego jest czasem nazywane oszacowaniem Nadaraya-Watson (Nadaraya, 1964[122]; Watson, 1964[166]). Szacowanie jądra jest średnią ważoną obserwacji w oknie wygładzania:

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

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

Parametr wygładzania h (ang. bandwidth) ma decydujący wpływ na uzyskany estymator. Im wyższa wartość parametru wygładzania, tym stopień wygładzenia jest większy. Do wyboru mamy możliwość dowolnego wyboru parametru wygładzania poprzez ustawienie wartości użytkownika. Możliwy jest również automatyczny dobór tej wielkości poprzez metodę SNR, SROT lub OS. Znacznie mniejszy wpływ na uzyskany wynik ma funkcja jądra. Do wyboru mamy jądro: Gaussa, w postaci funkcji jednostajnej (prostokąt), trójkątnej, Epanechnikova, quartic lub biweight (czwartego stopnia). Opis poszczególnych wielkości parametru wygładzania i funkcji jądra można znaleźć we wspomnianym rozdziale.

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.

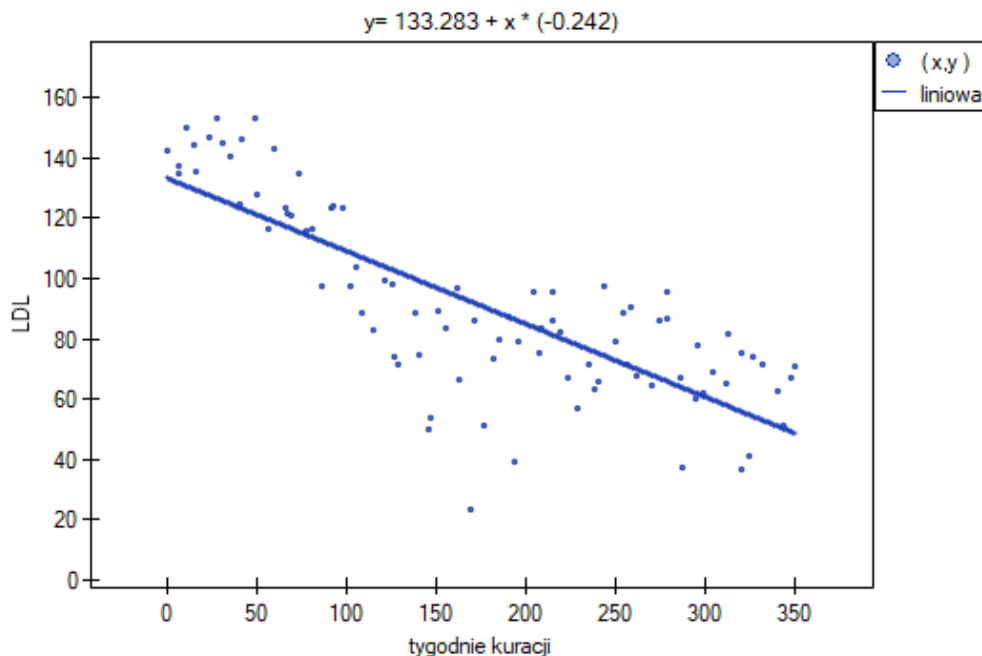


PRZYKŁAD (19.2) c.d. (plik LDL tygodnie.pqs)

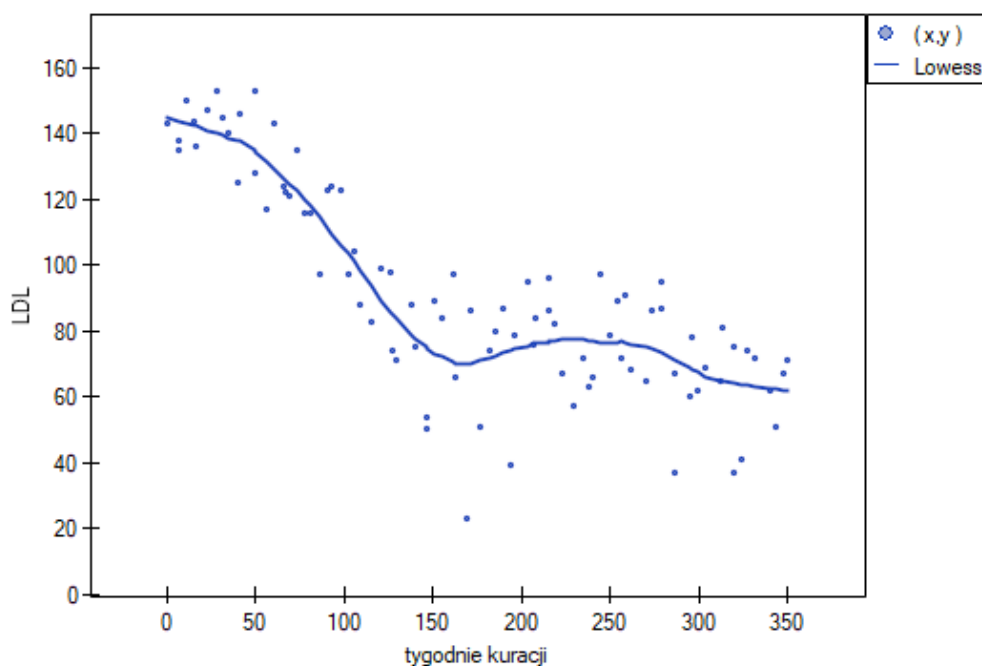
Badano skuteczność nowej terapii, której celem jest obniżenie poziomu cholesterolu we frakcji LDL.

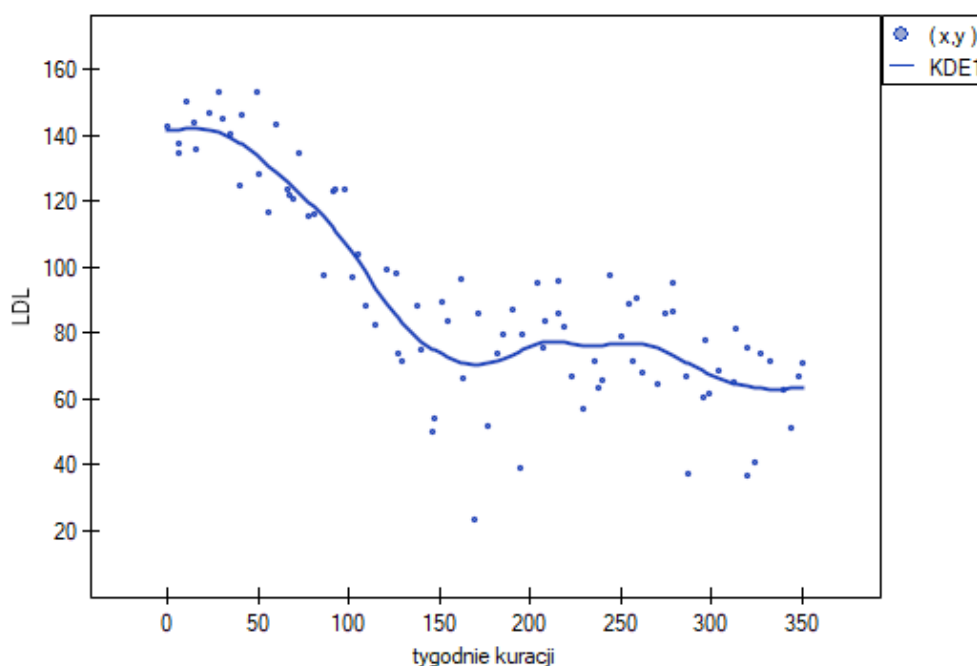
Przebadano 88 osób na różnym etapie kuracji. Sprawdźmy, czy wraz z upływem czasu stosowania kuracji (czas w tygodniach) poziom cholesterolu LDL spada i się stabilizuje.

Wyniki przedstawiono początkowo dopasowując linię prostą wskazującą kierunek badanej zależności.



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.



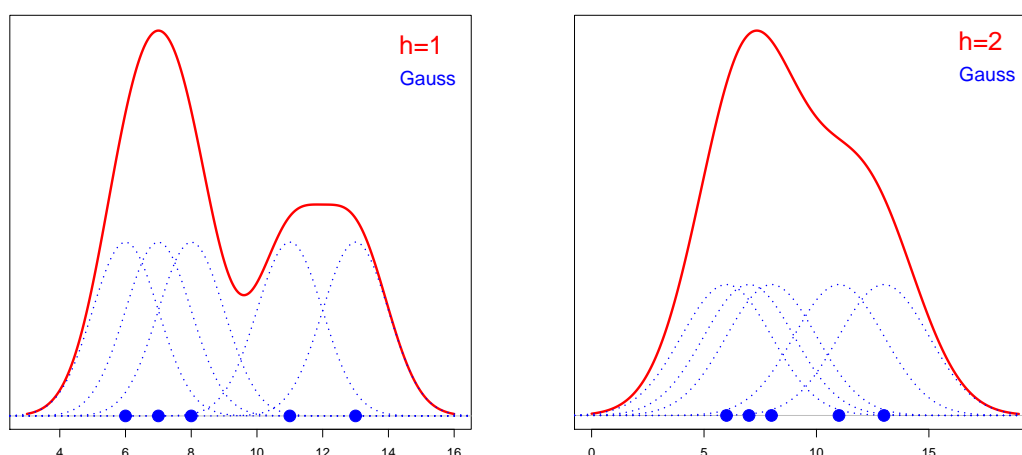


Obydwa sposoby tzn. zarówno LOWESS jak i wygładzanie jądrowe dały zbliżone wyniki i znacznie lepiej opisały dane, wskazując na początkowy spadek LDL a następnie jego stabilizację na poziomie bliskim 70 mg/dl.

6.3 ESTYMACJA JĄDROWA

6.3.1 Jednowymiarowy estymator jądrowy

Jednowymiarowy jądrowy estymator gęstości pozwala na przybliżenie gęstości rozkładu danych tworząc wygładzoną krzywą gęstości w sposób nieparametryczny. Dzięki niemu uzyskuje się lepszą estymację gęstości niż daje tradycyjny histogram, którego kolumny tworzą funkcję schodkową.



Estymator jądrowy definiowany jest w oparciu o odpowiednio wygładzone jądro $K_h(t_i)$. Parametr wygładzania h (*ang. bandwidth*) ma decydujący wpływ na uzyskany estymator. Im wyższa wartość parametru wygładzania, tym stopień wygładzenia jest większy.

Dla każdego punktu x z zakresu określonego przez dane wyznacza się gęstość czyli podaje wartość estymatora jądrowego w tym punkcie. Estymator ten powstaje poprzez zsumowanie wartości funkcji jąder $K_h(t_i)$ w tym punkcie:

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

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

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

Współczynniki wygładzania

Użytkownika - daje możliwość wybrania dowolnego współczynnika wygładzania wskazanego przez użytkownika, przy czym współczynnik ten musi być dodatni.

Użytkownika skalowane - jest ustalane tak, by można było zmieniać funkcję jądra pozostając przy wygładzeniu jakie wybrane zostało wcześniej dla jądra Gaussa. W praktyce wybierając inną funkcję niż Gaussa współczynnik wygładzania zostaje przeskalowany (Scott, D. W. 1992[147]), przez co wygładzenie pozostaje na podobnym poziomie jaki był dla funkcji Gaussa. Daje to wygodę przełączania się pomiędzy różnymi jądrami bez rozważania skalowania parametru wygładzania. Przeliczenia dotyczące skalowania dokonywane są w oparciu o odchylenie standardowe:

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

SNR - współczynnik wygładzania budowany na podstawie metody Silverman'a (Silverman B. W. 1986 [152] str 45 i 47) i korekty Jones'a (Jones M. C. i inni 1996[87]) wykorzystującej odchylenie standardowe z próby a nie z populacji - jak zaproponował Silverman:

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

Dla jądra innego niż Gaussa, współczynnik wygładzania podlega skalowaniu (Scott D. W., 1992[147])

SROT - współczynnik wygładzania budowany na podstawie metody Silverman'a (Silverman B. W. 1986 [152] str. 48) z korektą Jones'a (Jones M. C. i inni 1996[87]):

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

Dla jądra innego niż Gaussa, współczynnik wygładzania podlega skalowaniu (Scott D. W., 1992[147])

OS - współczynnik wygładzania budowany na podstawie metody Terrell'a i Scott'a (Terrell G. R. i Scott D. W. 1985[161], Terrell G. R. 1990[160] str. 470):

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

Dla jądra innego niż Gaussa, współczynnik wygładzania podlega skalowaniu (Scott D. W., 1992[147])

Funkcja jądra

Funkcja jądra w mniejszym zakresie niż parametr wygładzania wpływa na uzyskaną wartość estymatora jądrowego. Jądro jest funkcją gęstości prawdopodobieństwa budowaną wokół każdego punktu danych x_i . Zwykle jest to funkcja symetryczna osiągająca maksimum w punkcie x_i , a zmniejszająca swoje wartości wraz z oddalaniem się (wzrostem odległości d_i) od tego punktu. Odległość od analizowanego punktu jest modyfikowana przez parametr wygładzania h zgodnie z wzorem: $t_i = \frac{d_i}{h}$.

W zależności od potrzeb funkcja jądra może przyjmować postać funkcji:

Gausa

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

jednostajnej (prostokąt)

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

trójkątnej

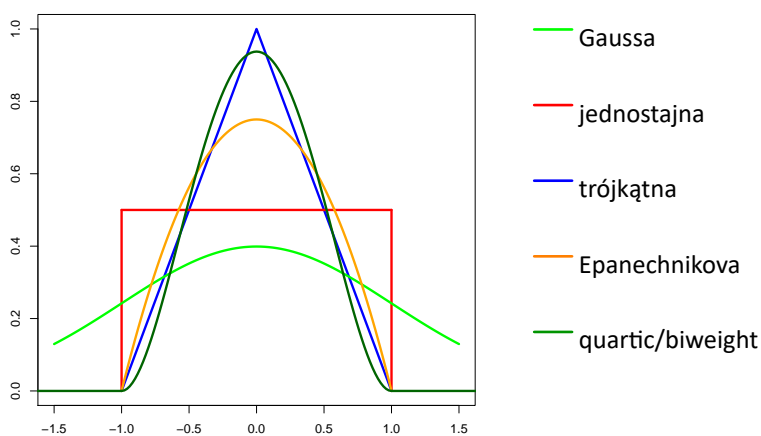
$$K_h(t_i) = \begin{cases} \frac{1-t_i}{h} & \text{jeżeli } t_i < 1 \\ 0 & \text{jeżeli } t_i \geq 1 \end{cases}$$

Epanechnikova

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

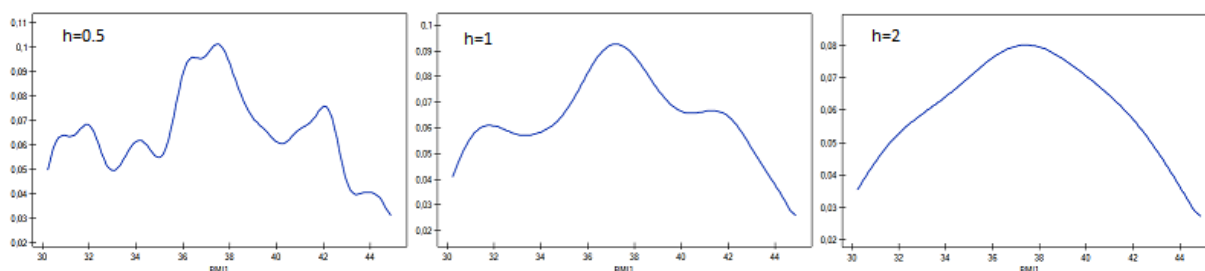
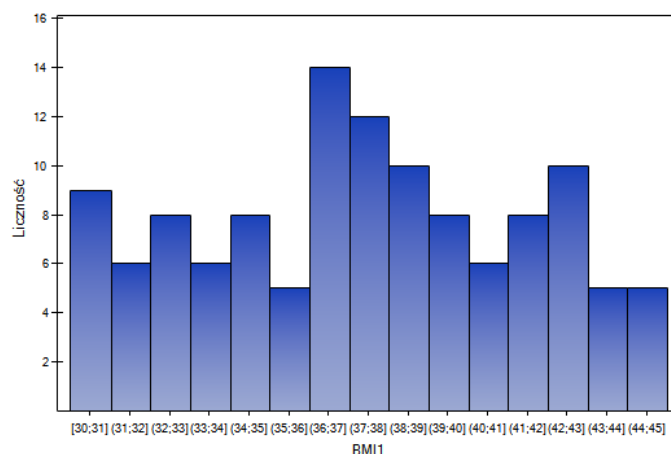
quartic lub biweight (czwartego stopnia)

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

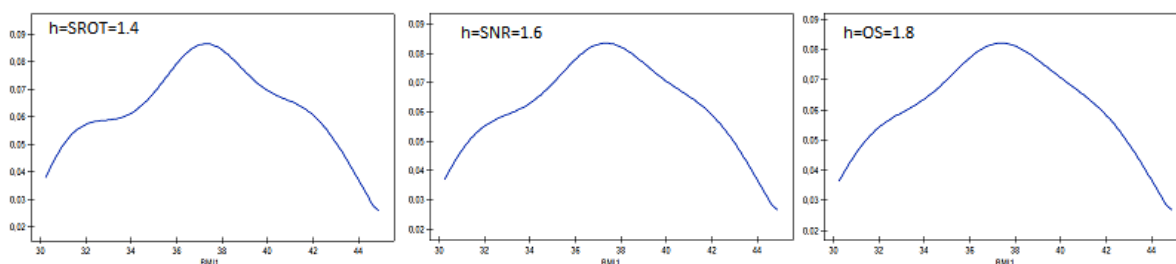


PRZYKŁAD 6.1. (plik BMI.pqs)

Wyliczono wartości współczynnika wagowo-wzrostowego BMI1 dla pewnej grupy osób otyłych. Przedstawiono ich rozkład za pomocą histogramu z podziałem wartości co 1 jednostkę BMI. Dane zobrazowano również za pomocą jądrowego estymatora gęstości wybierając Gaussowską funkcję jądra i ustawiając współczynniki wygładzania odpowiednio: 0.5, 1, 2.



Współczynniki wygładzania estymatora jądrowego sugerowane przez metody SROT, SNR i OS osiągają wielkości pomiędzy 1.4 a 2.

**6.4 WYKRES BLANDA-ALTMANA**

Jak zauważył Bland i Altman (1986[21], 1999[8]) w medycynie klinicznej pomiary wykonywane na żywym organizmie ulegają ciągłym zmianom a ich prawdziwa wartość jest nieznaną (np. ciśnienie tętnicze krwi), co wymaga ciągłego doskonalenia i tworzenia nowych, lepszych narzędzi ich pomiaru. Zwykle, kiedy tworzona jest nowa metoda, porównuje się jej wyniki z inną, uznaną metodą tzw. gold standard. W tym celu bada się zgodność nowej metody z metodą dotychczas stosowaną. Oczywiście, nie można oczekiwać, że nowa metoda da dokładnie ten sam wynik co metoda dotychczas wykorzystywana, ale badaczowi zależy na sprawdzeniu na ile różnią się uzyskane wyniki. By zastąpić starą metodę metodą

nową, uzyskana różnica pomiędzy wynikami obu metod powinna być na tyle mała, by nie stwarzać problemu w interpretacji klinicznej. Na przykład w pomiarach ciśnienia tętniczego krwi różnica wielkości 20mmHg będzie na tyle duża, że nie może zostać uznana za dozwolony błąd, ponieważ może rzutować na zmianę decyzji o podjęciu leczenia. Metody statystyczne nie odpowiedzą na pytanie jak duża różnica metod jest dozwolona, by można było uznać metody za zgodne, ale odpowiednia ilustracja graficzna uzyskanych różnic i możliwych granic zmienności ułatwi podjęcie decyzji badaczowi.

Wykres Blanda-Altmana jest wykresem punktowym, gdzie:

Oś X - średnia pomiarów dla porównywanych metod;

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, wówczas występuje przesunięcie, które nazywamy ang. bias, czyli 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.

Uwaga1!

Zdefiniowany powyżej przedział zgodności nie jest tym samym co przedział ufności.

Uwaga2!

Nie ma wymogu, by dane miały rozkład normalny a jedynie by rozkład różnic nie odbiegał znacznie od rozkładu normalnego. Możemy sprawdzić czy różnice mają rozkład normalny korzystając z [testów](#) badających zgodność z rozkładem normalnym lub interpretacji graficznej tego rozkładu.

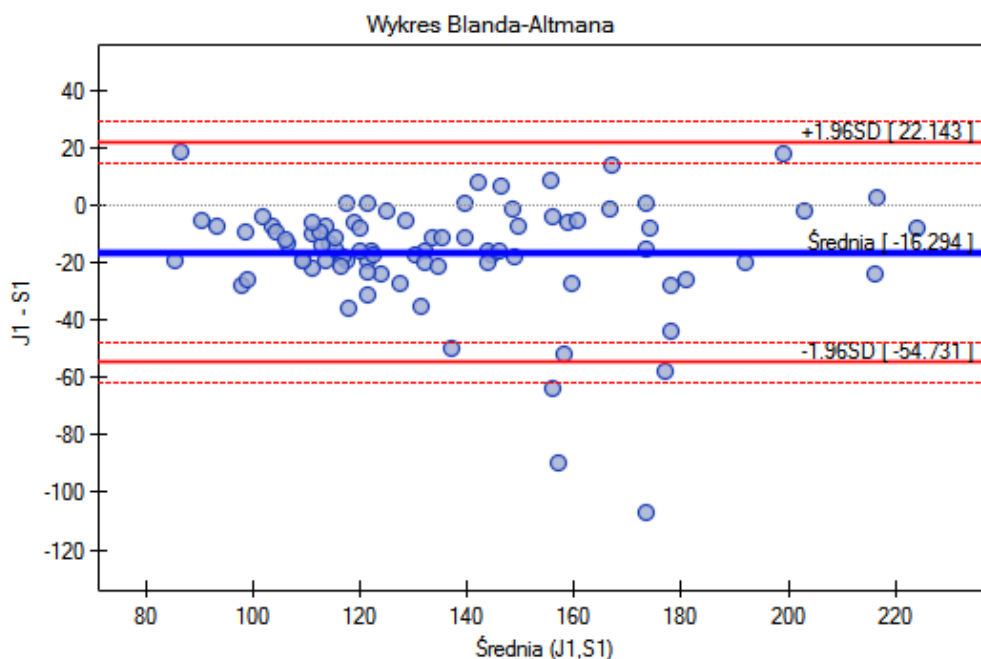
Precyzja granic przedziału zgodności (ang. Precision of the limits of agreement) - jest to przedział ufności dla granic, a więc przedział dokładności z jaką wyznaczamy granice bazując na reprezentatywnej próbie. Czym liczniejsza będzie próba i czym mniejsza będzie wariancja różnic, tym uzyskana precyzja będzie większa.

PRZYKŁAD 6.2. (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

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 poziomą 0). Rozpiętość przedziału zgodności wynosi aż 76.9mmHG.



Dla osób z nadciśnieniem (ciśnienie skurczowe ≥ 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 w badaniu zgodności metod. Metoda o większej powtarzalności jest bardziej precyzyjna. Jeśli pomiary jednej z porównywanych metod nie są powtarzalne (tzn. powtarzane 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 jeden pomiar dla każdego obiektu (pacjenta), w badaniach mających na celu szacowanie zgodności zaleca się wykonywanie pomiarów kilkakrotnie. Takie podejście daje możliwość uwzględnienia powtarzalności uzyskanych wyników w badaniach nad zgodnością metod.

Uwaga!

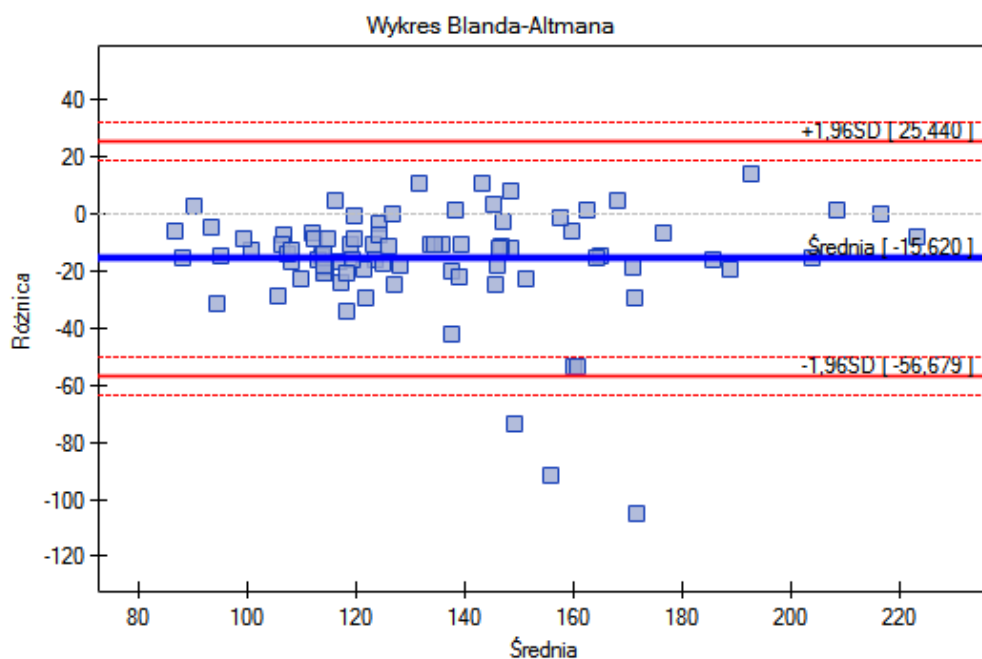
Jako powtarzane pomiary rozumiemy tutaj pomiary wykonywane niezależnie na tych samych obiektach, w takich samych warunkach.

PRZYKŁAD (6.2) c.d. (plik ciśnieniomierz.pqs)

W porównaniu zgodność 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 przy pomocy ciśnieniomierza półautomatycznego i 3 pomiary wykonane przy pomocy ciśnieniomierza klasycznego. 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. Nie uwzględniając bowiem stopnia powtarzalności zakładamy, że powtarzalność jest 100 procentowa, co w rzeczywistych warunkach jest niemal niemożliwe.



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

6.5 Macierze korelacji

Kiedy interesuje nas korelacja pomiędzy wieloma zmiennymi, wygodnym sposobem jej zobrazowania jest przedstawienie współczynników korelacji w postaci tabeli. W zależności od skali na jakiej zebrano dane, w programie PQStat mamy do wyboru współczynniki:

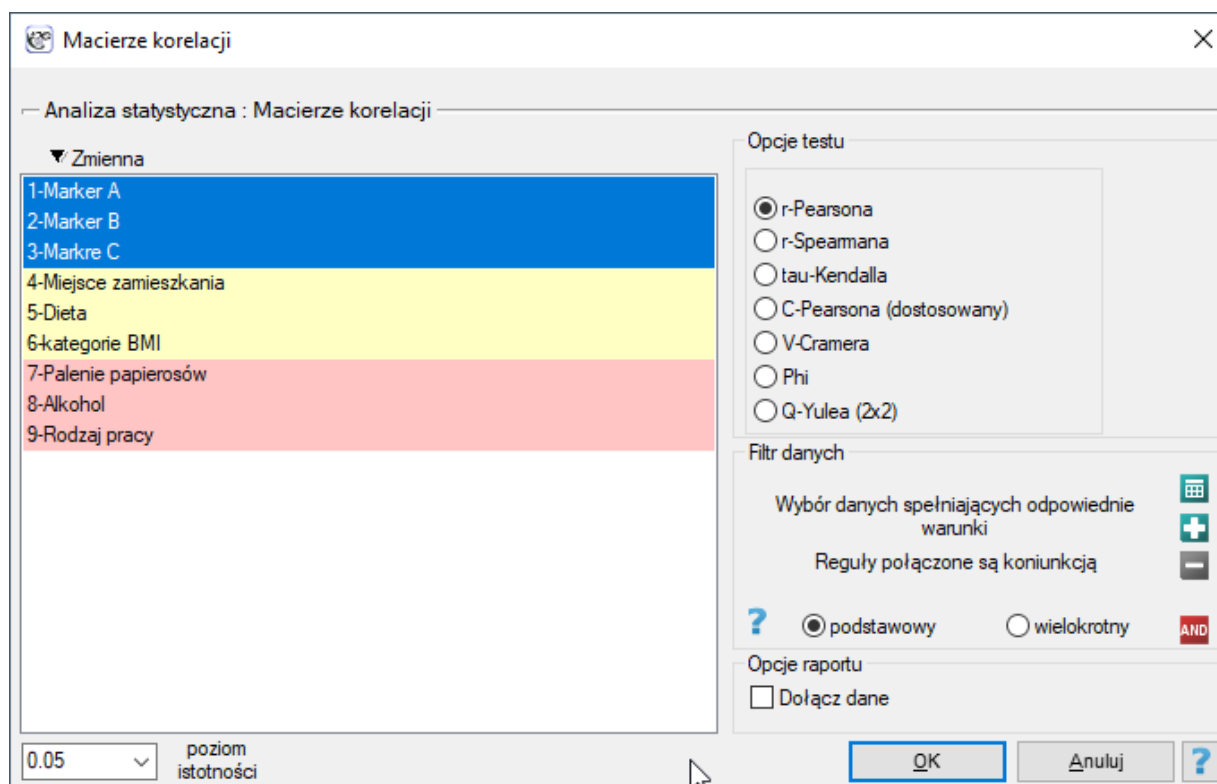
- **r-Pearsona** (skala interwałowa)
- **r-Spearmana** (skala porządkowa lub silniejsza)
- **tau-Kendalla** (skala porządkowa lub silniejsza)
- **C-Pearsona** (skala nominalna lub silniejsza)

- **V-Cramera** (skala nominalna lub silniejsza)
- **Phi** (skala nominalna lub silniejsza)
- **Q-Yulea** (skala nominalna lub silniejsza)

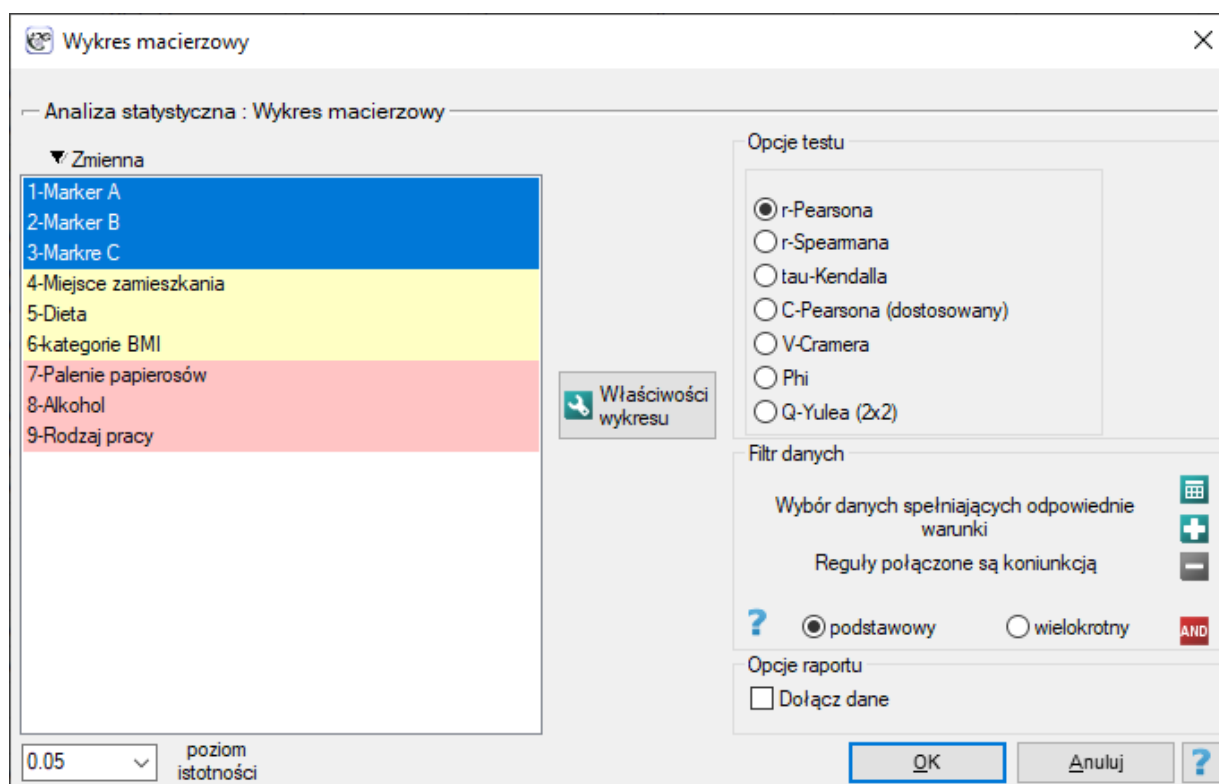
W wyniku analizy powstają dwie macierze, tzn. macierz współczynników korelacji i macierz wartości p dla testu określającego istotność statystyczną danego współczynnika (dla włóczników r-Pearsona, r-Pearsona i tau-Kendalla były to dedykowane im testy, dla skali nominalnej test chi-kwadrat). W macierzy współczynników korelacji, w miejscu przecięcia się dwóch zmiennych podany jest współczynnik ich korelacji, a jego wartość p znajduje się w odpowiadającym mu miejscu w drugiej macierzy. Kolor komórki w macierzy współczynników jest stopniowany od niebieskiego (ujemne współczynniki) do czerwonego (dodatnie współczynniki).

Analiza wyznacza korelację dla każdej pary zmiennych, dlatego braki danych pomijane są parami. Jeśli chcemy przeprowadzić analizę pomijając baki danych występujące w innych zmiennych (tzn. nie tych, które wchodzi w skład danej pary), wówczas powinniśmy to zrobić korzystając z [filtru zaawansowanego](#).

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



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



PRZYKŁAD 6.3. (plik markery inne.pqs)

Przedstawiono fragment większego badania dotyczącego choroby nowotworowej. Zabrane dane dotyczą grupy 100 osób. W badaniu mierzono min. 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 zakłada konieczność wcześniejszego sprawdzenia wzajemnych korelacji zmiennych. Na potrzeby dalszych analiz:

- (1) Sprawdzimy korelację wewnątrz każdej z tych grup.
- (2) Sprawdzimy korelację między wszystkimi zmiennymi.

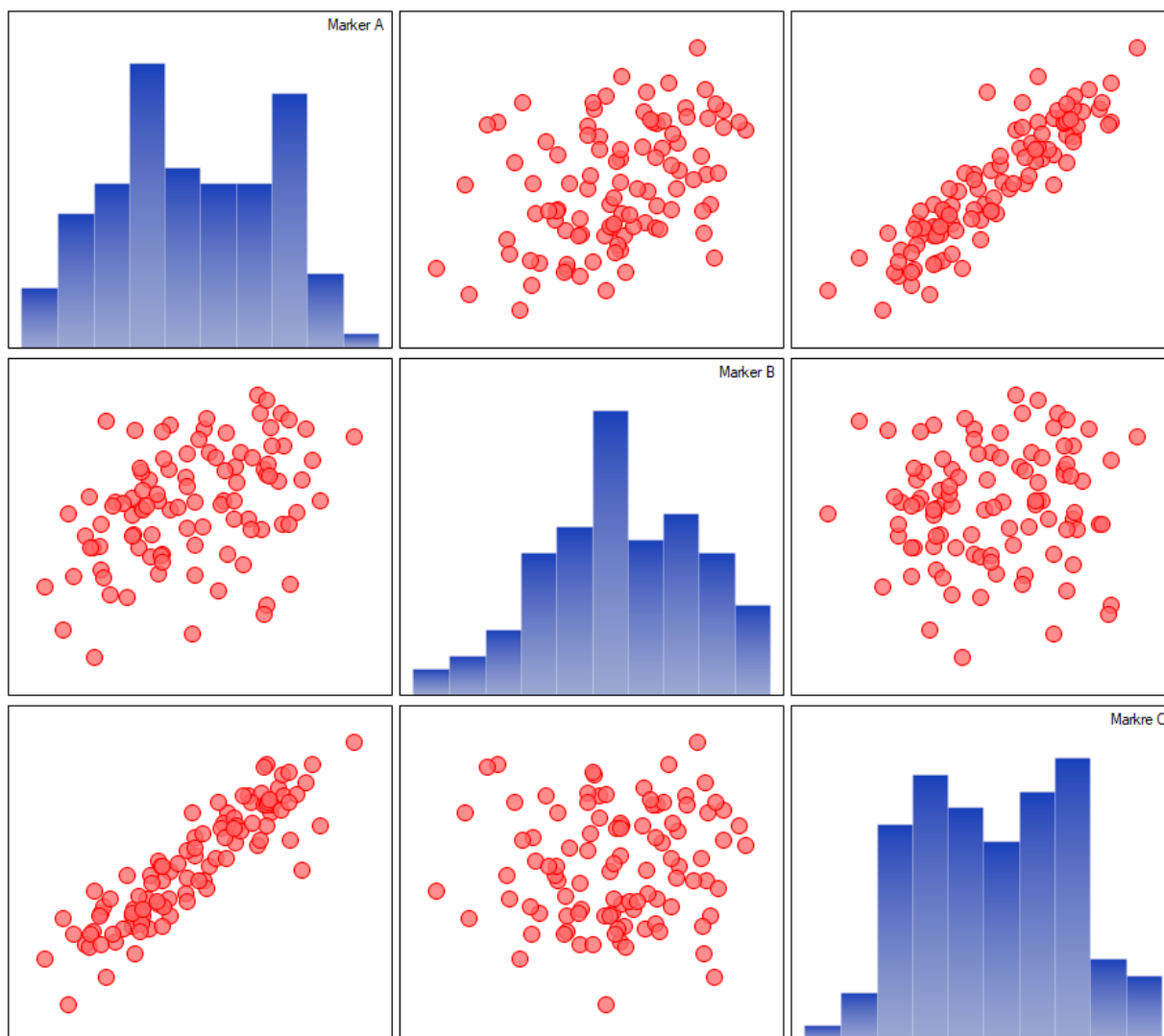
(1)

Dla skali interwałowej, przy założeniu normalności rozkładu korelację możemy sprawdzić współczynnikiem korelacji liniowej Pearsona. Najsilniejsza korelacja dotyczy markera A i markera C ($r=0.8995$, $p<0.0001$) naj słabsza i nieistotna statystycznie 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.

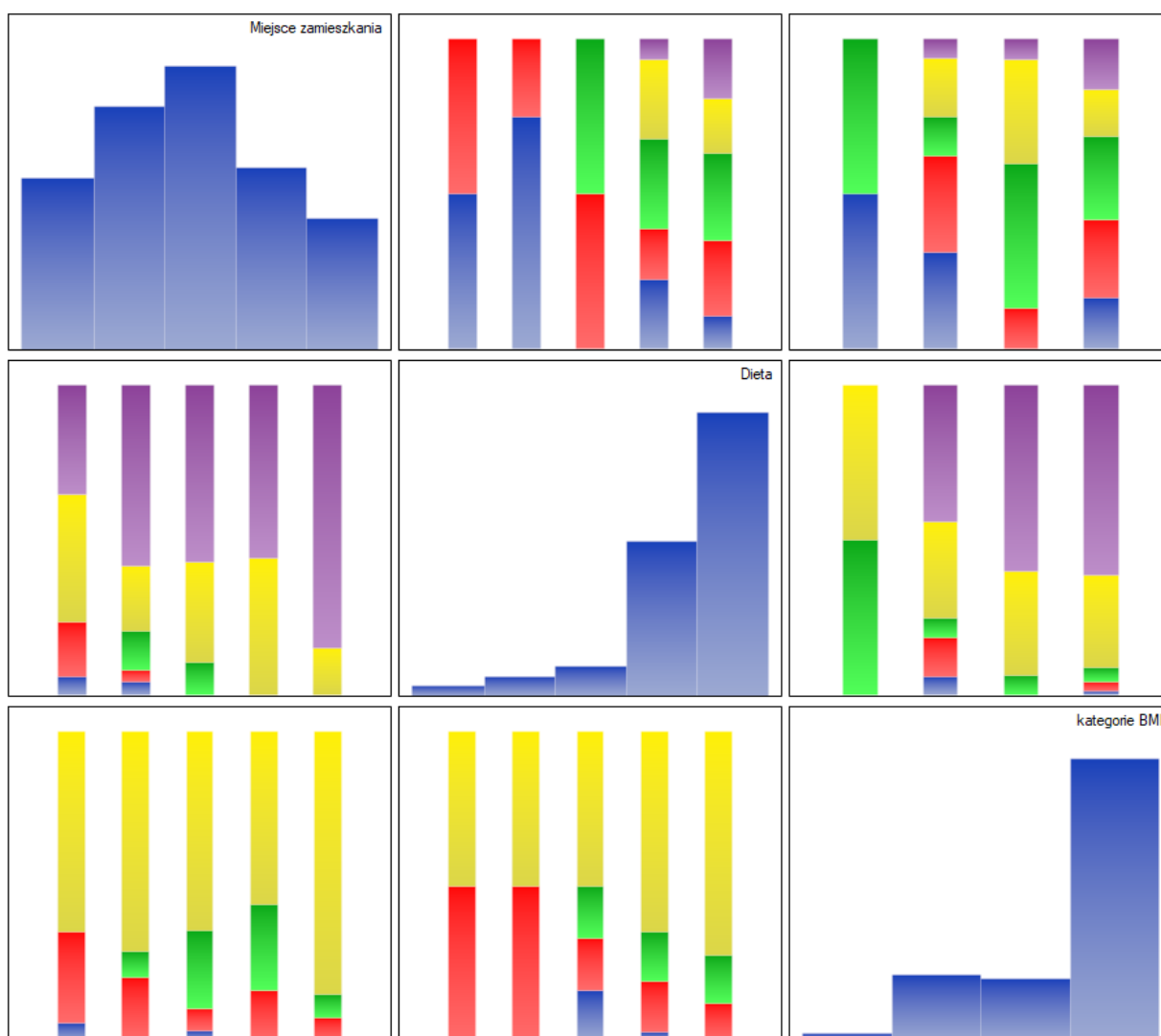


Dla skali porządkowej korelację sprawdzimy przy pomocy współczynnika korelacji Spearmana. Jedyna istotna korelacja dotyczy diety i miejsca zamieszkania ($r=0.2634$, $p=0.0081$).

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.



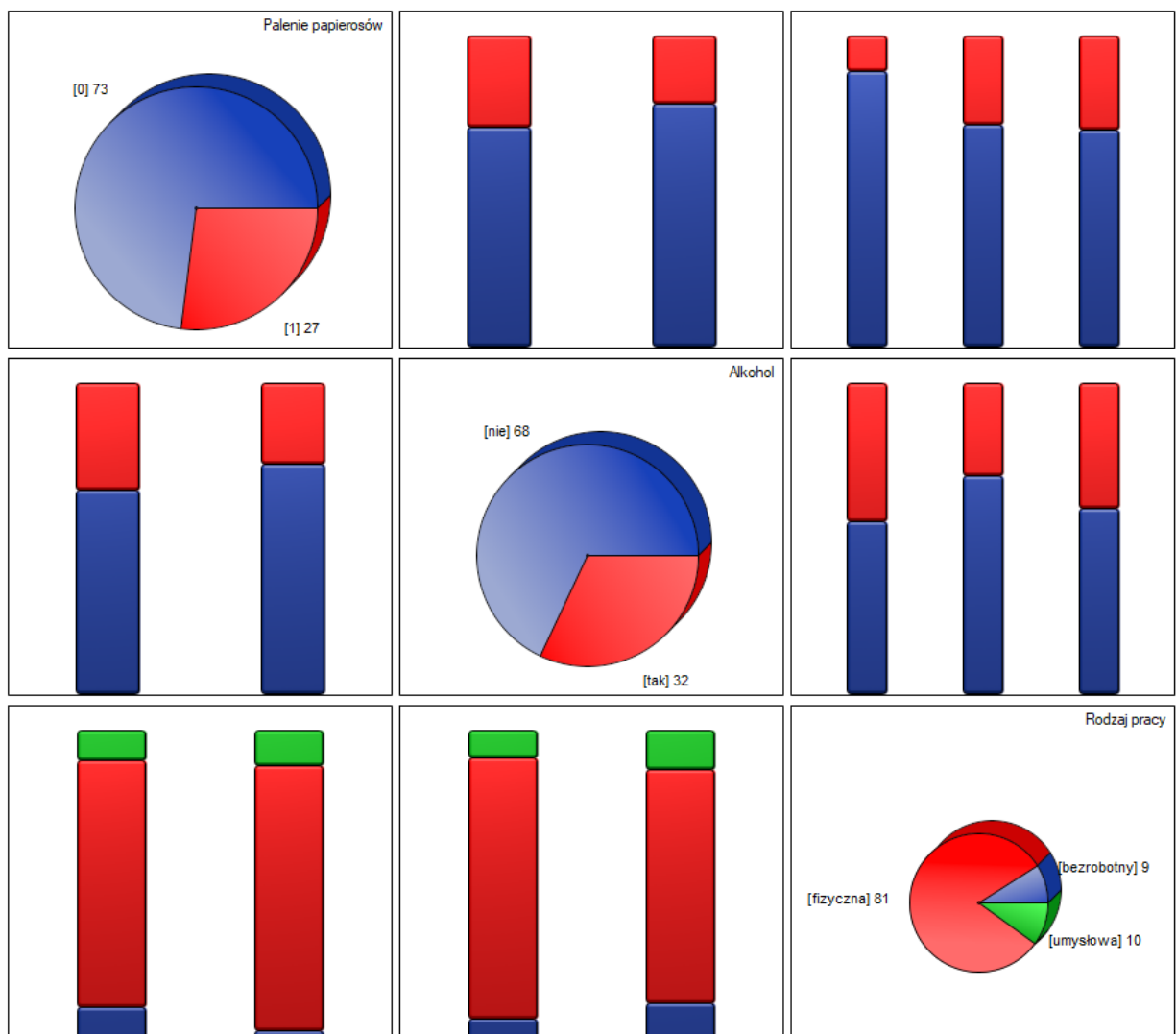
Dla skali nominalnej korelację sprawdzimy przy pomocy współczynnika C-Pearsona skorygowanego na

wielkość tabeli. Nie uzyskaliśmy istotnych statystycznie korelacji.

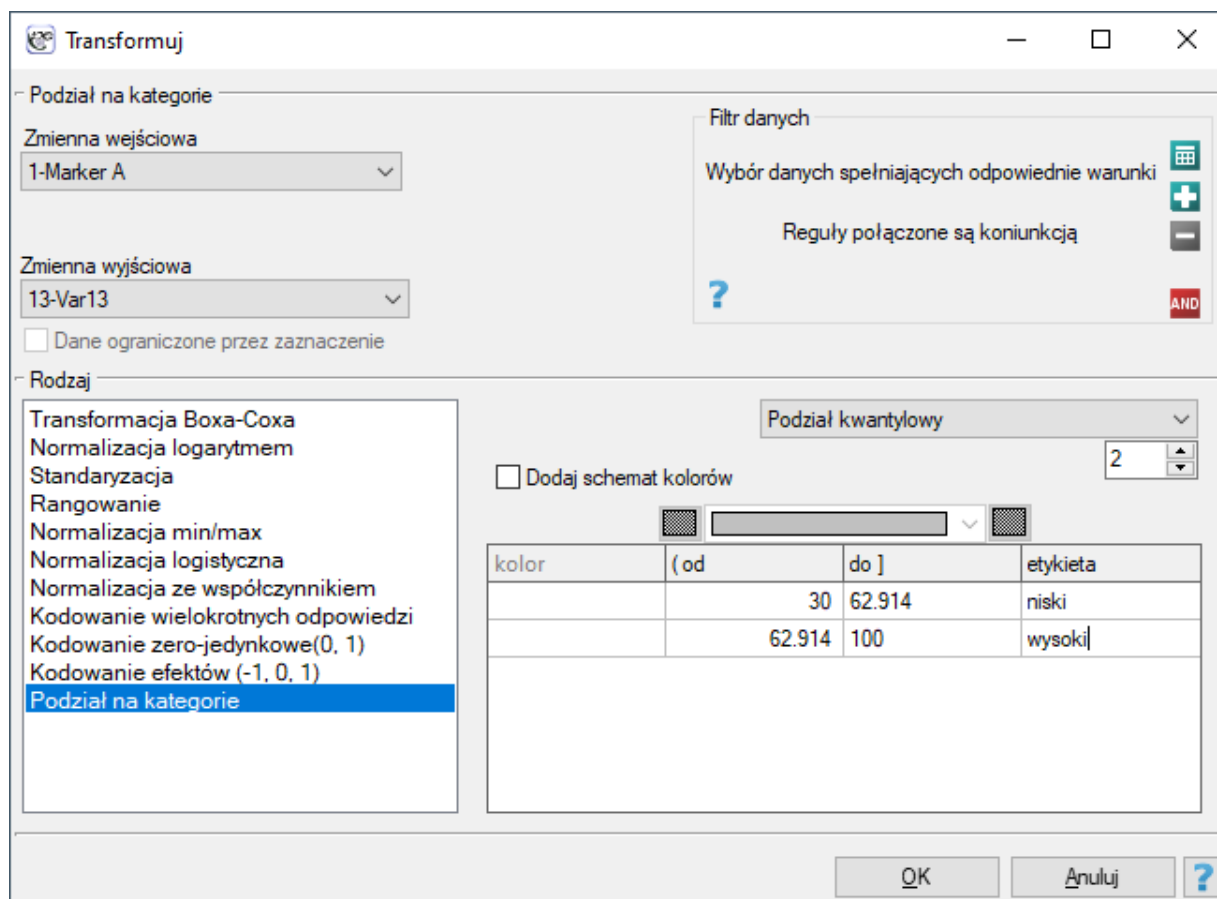
C-Pearsona (do			
	Palenie papi	Alkohol	Rodzaj pracy
Palenie papieros		0.1116	0.1589
Alkohol	0.1116		0.1504
Rodzaj pracy	0.1589	0.1504	

P-value			
	Palenie papi	Alkohol	Rodzaj pracy
Palenie papieros		0.4284	0.5277
Alkohol	0.4284		0.5645
Rodzaj pracy	0.5277	0.5645	

Ewentualne 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.



(2) Najłatwiejszym sposobem wyznaczania 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.



Transformuj

Podział na kategorie

Zmienna wejściowa
1-Marker A

Zmienna wyjściowa
13-Var13

Dane ograniczone przez zaznaczenie

Rodzaj

- Transformacja Boxa-Coxa
- Normalizacja logarytmem
- Standaryzacja
- Rangowanie
- Normalizacja min/max
- Normalizacja logistyczna
- Kodowanie wielokrotnych odpowiedzi
- Kodowanie zero-jedynkowe(0, 1)
- Kodowanie efektów (-1, 0, 1)
- Podział na kategorie**

Podział kwantylowy: 2

Dodaj schemat kolorów

kolor	(od	do]	etykieta
	30	62.914	niski
	62.914	100	wysoki

OK Anuluj ?

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.

Macierze korelacji

Analiza statystyczna : Macierze korelacji

Zmienna

- 1-Marker A
- 2-Marker B
- 3-Markre C
- 4-Miejsce zamieszkania
- 5-Dieta
- 6-kategorie BMI
- 7-Palenie papierosów
- 8-Alkohol
- 9-Rodzaj pracy
- 10-Range Quanty
- 11-Range Quanty
- 12-Range Quanty

Wpisz numery zmiennych

Filtruj zmienne

Właściwości zmiennych (kody/etykiety)

Opcje testu

- r-Pearsona
- r-Spearmana
- tau-Kendalla
- C-Pearsona (dostosowany)
- V-Cramera
- Phi
- Q-Yulea (2x2)

Kody/Etykiety/Format

Przypisz nowe wartości / Przypisz etykiety wartości

Zmienne

- 1-Marker A
- 2-Marker B
- 3-Markre C
- 4-Miejsce zamieszkania
- 5-Dieta
- 6-kategorie BMI
- 7-Palenie naniemsów

Etykieta zmiennej:

Liczba unikalnych wartości: 5

Liczba przypisanych wartości: 5

Liczba przypisanych etykiet: 5

Kody i etykiety dla wartości

Wartość	Obowiązująca wartość	Etykieta
Istotne	2	Istotne
Mało istotne	1	Nieistotne
Nieistotne	1	Nieistotne
Umiarkowanie istotne	2	Istotne
Zdecydowanie istotne	2	Istotne

Dodaj wartość niewystępującą

Zamknij

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

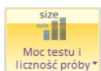
C-Pearsona (do

	Miejsce zam	Dieta	kategorie BM	Palenie papier	Alkohol	Rodzaj pracj	Range Quan	Range Quan	Range Quan
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 papier	Alkohol	Rodzaj pracj	Range Quan	Range Quan	Range Quan
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	

7 MOC TESTU I LICZNOŚĆ PRÓBY



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

7.1 Niezbędna liczność próby

Dla marginesu błędu proporcji i średniej

Ze względu na to, że badanie całej populacji najczęściej nie jest ani praktyczne ani możliwe, wybiera się jej podzbiór - próbę. Próba jest oczywiście odpowiednio mniej liczna niż populacja, ale powinna ją dobrze odzwierciedlać. Jednym z kluczowych aspektów przy planowaniu badania, obok losowości próby, jest założenie jej rozmiaru. Rozmiar powinien być tak dobrany, by wnioskowanie na temat populacji było prawdziwe.

Jeśli zainteresowani jesteśmy, by proporcje pewnych cech, lub ich wartości średnie, wyliczone dla próby odzwierciedlały proporcje lub wartości średnie w populacji z możliwie małym błędem, możemy odpowiednio oszacować niezbędną do tego celu liczność próby.

Liczność próby dla oszacowania proporcji populacji

Zakładając możliwość popełnienia błędnie szacunku wielkości E , możemy wyznaczyć niezbędną liczność próby n_0 - dla nieznanego 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 licznosc będzie tak zwiększona, by była wystarczająca dla każdej możliwej proporcji, zatem wykorzystana zostanie wartość $p = 0.5$.

W przypadku, gdy znamy licznosc populacji (a w szczególności - gdy licznosc jest relatywnie mała w stosunku do n_0 , tzn. gdy $n_0/N > 5\%$) powinniśmy stosować tzw. korektę skończonej populacji (*FPC*) (Lenth (2001)[99], Armitage i Colton (2009)[12]) daną wzorem:

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

Licznosc próby dla oszacowania sredniej populacji Zakładając mozliwosc popełnienia błędu szacunku wielkości E , możemy wyznaczyć niezbędną licznosc próby n_0 - dla nieznanego 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 licznosc populacji (a w szczególności - gdy licznosc jest relatywnie mała w stosunku do n_0 , tzn. gdy $n_0/N > 5\%$) powinniśmy stosować tzw. korektę skończonej populacji (*FPC*) daną wzorem:

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

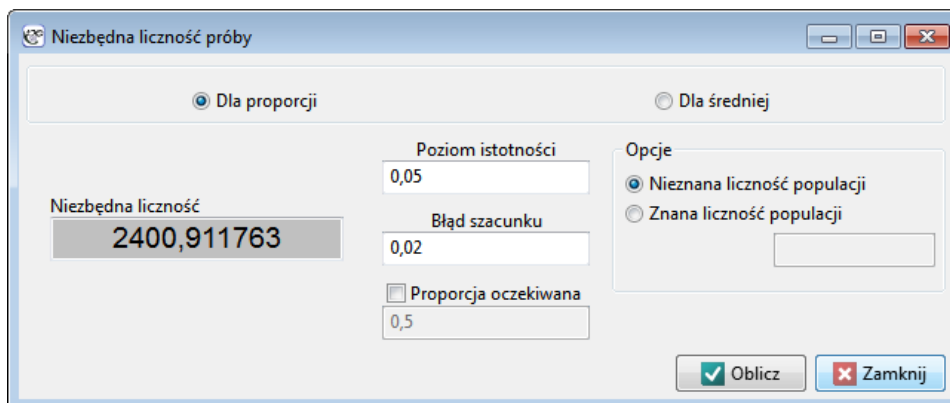
Okno z ustawieniami opcji Niezbędnej licznosci próby wywołujemy poprzez menu Statystyki zaawansowane → Moc testu i licznosc próby → Niezbędna licznosc próby.

PRZYKŁAD 7.1. Szacowanie proporcji

Populacja: Uprawnieni do głosowania na prezydenta Polski.

Zainteresowani jesteście poparciem dla poszczególnych kandydatów.

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



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%.

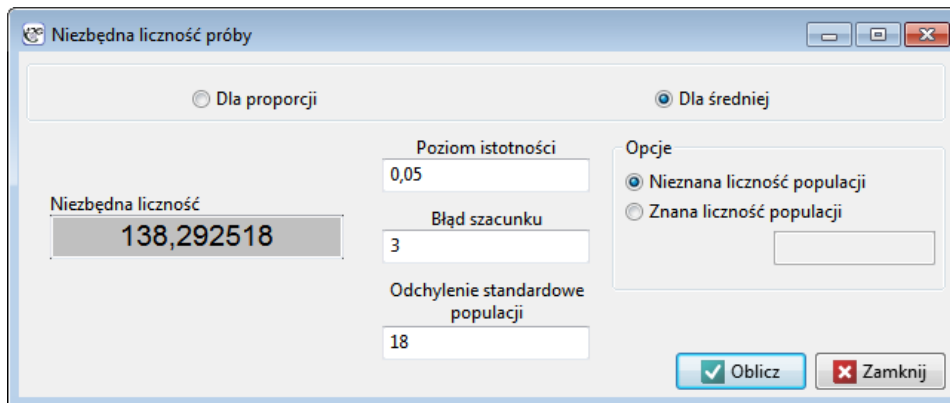
Przy wyborze wielkości dopuszczanego błędu należy zwrócić uwagę na to, czy nie zachodzi taka sytuacja, iż w wyborach biorą udział również kandydaci o niewielkim poparciu (na granicy założonego błędu szacunku). Jeśli tak jest, warto zmniejszyć wartość zakładanego błędu - konsekwencją zmniejszenia błędu będzie wówczas wzrost niezbędnej liczebności próby.

PRZYKŁAD 7.2. Szacowanie wartości średniej

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

Zainteresowani jesteście średnią masą ciała tych ludzi.

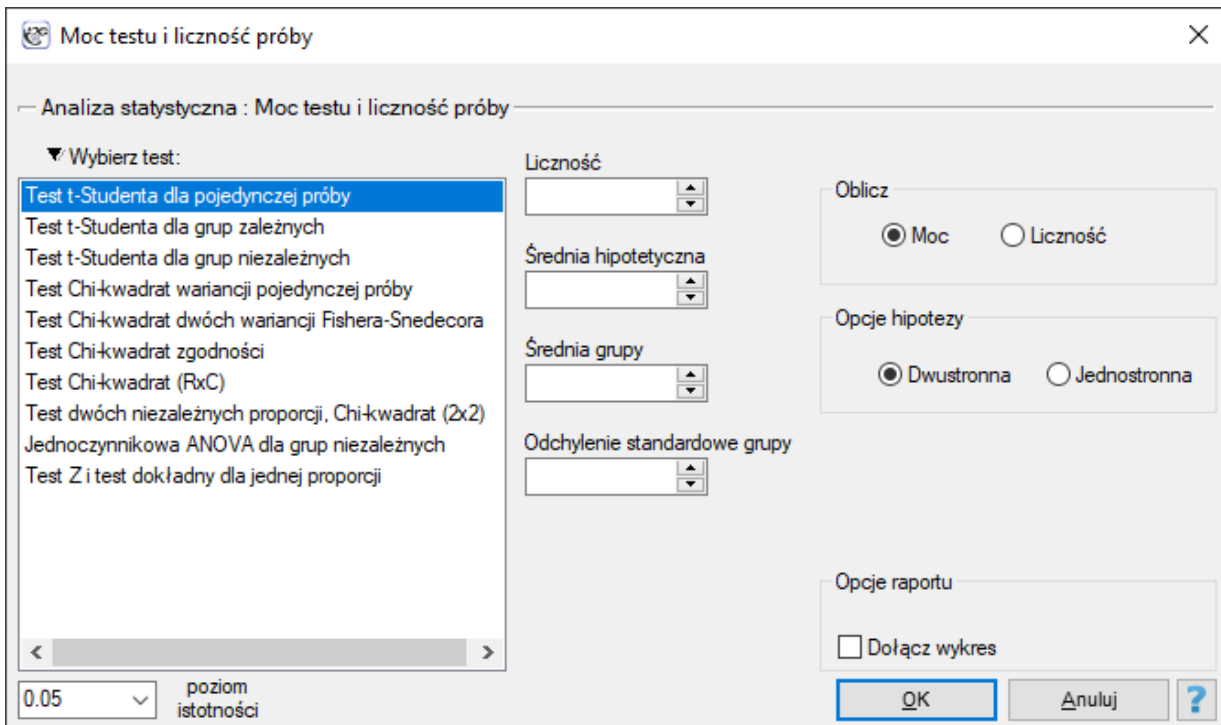
Ile osób wybrać by średnia masa obarczona była błędem co najwyżej $3kg$? Wiemy, że populacyjne odchylenie standardowe masy ciała tych osób wynosi $18kg$.



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.

7.2 Moc i liczebność dla testu

Okno z ustawieniami opcji mocy testu i niezbędnej liczebności próby dla tego testu wywołujemy poprzez menu Statystyka zaawansowana → Moc testu i liczebność próby → Moc i liczebność dla testu.



Analiza mocy jest bezpośrednio związana z testowaniem hipotez, a więc z konkretnymi testami statystycznymi. Testy różnią się swoją mocą. Jedne testy są silniejsze, inne słabsze. W związku z tym faktem, jeśli do rozwiązania danego problemu statystycznego mamy do dyspozycji kilka testów, lepiej jest wybrać ten test, który ma większą moc. Taki test jest silniejszy, więc łatwiej odrzuci hipotezę zerową, a zatem łatwiej będzie nam udowodnić hipotezę alternatywną - co jest przecież celem badacza.

Moc testu statystycznego jest to zdolność testu do wykrywania różnic, związków, zgodności i wszelkiego rodzaju zależności, które opisujemy w hipotezie alternatywnej. Językiem fachowym mocą testu nazywamy prawdopodobieństwo przyjęcia hipotezy alternatywnej gdy ona jest w rzeczywistości prawdziwa.

Sprawdzenie mocy testu może być wykonywane a priori tzn. przed zebraniem danych do badania właściwego, ale często to recenzenci prac lub my sami już w trakcie analiz czyli a' posteriori tzn. po zebraniu właściwej próby jesteśmy zainteresowani jaka jest moc analiz, które wykonujemy. Jeśli moc testu jest mała, wówczas uzyskane wyniki mogą być niejednoznaczne, jeśli duża - możemy się spodziewać, że w przyszłości trudno będzie innym badaczom uzyskać wyniki odmienne tzn. podważyć nasze rezultaty. Na przykład gdy wykazujemy przy pomocy testu o mocy 80%, że dwie grupy studentów różnią się od siebie istotnie statystycznie pod względem liczby poprawnie rozwiązanych zadań, oznacza to, że gdy inni badacze będą powtarzać ten eksperyment w takich samych warunkach jak my, to w 80% losowych prób takiej wielkości jak nasza, stosując ten sam test i przyjmując ten sam poziom istotności, również udowodnią hipotezę alternatywną.

Moc testu określana jest wzorem:

$$Moc = 1 - \beta,$$

gdzie:

β - błąd II rodzaju, czyli prawdopodobieństwo przyjęcia hipotezy zerowej gdy jest ona fałszywa.

Moc testu związana jest bezpośrednio z licznością próby n - czym liczniejsza próba, tym większa moc, tzn. czym więcej zbierzemy studentów by przeprowadzić badanie, tym łatwiej będzie przekonać, że wykryte różnice między grupami nie są dziełem przypadku, ale rzeczywiście występują między populacjami. Stąd w tym samym podejściu badacz może być zainteresowany wyznaczeniem **niezbędnej liczności**

próby dla danego testu statystycznego przy zachowaniu jego mocy na zadanym poziomie.

W programie PQStat możemy obliczyć moc testu zadając liczność próby, lub obliczyć liczność próby zadając moc testu jaką chcemy osiągnąć. Niestety zarówno moc testu jak też liczność próby, oprócz tego że są związane z sobą wzajemnie, zależą również od innych dodatkowych informacji dotyczących zebranej próby, które należy ustalić, są to:

- Wielkość efektu ES (ang. effect size) jaki uznajemy za ważny. Im większy efekt, tym większa moc i tym mniejsza próba jest potrzebna by tę moc uzyskać. Na przykład uzyskując różnicę pięciu poprawnie rozwiązanych zadań łatwiej będzie przekonać, że w rzeczywistości grupy różnią się (będziemy mieli większą moc argumentu) niż wtedy, gdy próbowalibyśmy przekonać o rzeczywistej przewadze jednej z grup studenckich podając, że różnią się zaledwie jednym poprawnie rozwiązaniem.
Do wyznaczenia mocy lub niezbędnej liczności zwykle badany efekt musi zostać zestandaryzowany, dlatego w wielu sytuacjach konieczne jest podanie dodatkowych informacji np. odchylenia standardowego, współczynnika korelacji i innych współczynników, które pozwalają zestandaryzować taki efekt.
- Poziomem istotności statystycznej α (błąd I rodzaju) - im większy poziom istotności, tym większa moc. Niestety poziom istotności, to ta część analizy, na którą mamy tylko pozorny wpływ, to znaczy niewiele jest sytuacji kiedy można ją zmieniać, a jeśli już dozwolona jest zmiana, to polegająca na zmniejszeniu α (patrz [Wielokrotne porównania](#)). Zwykle przyjmujemy $\alpha = 0,05$.
- Kierunek hipotezy tzn. hipoteza dwustronna (równość w hipotezie zerowej) lub jednostronna (znak $<$ lub $>$ w hipotezie zerowej). Hipoteza jednostronna daje większą moc, ale jest znacznie rzadziej wybierana, gdyż w rzeczywistych sytuacjach stosując test statystyczny rzadko zakładamy, że np. studenci z pierwszej populacji nie mają szansy pokonać studentów z drugiej populacji, często natomiast na stracie dajemy równe szanse obu grupom.

Przed przystąpieniem do wyznaczania mocy testu należy znać zasady jego stosowania, rozumieć jego hipotezy i umieć określić wielkość efektu, a jeśli dysponujemy danymi z badania np. pilotażowego, to również wykonać ten test.

7.2.1 Test t-Studenta dla pojedynczej próby

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności [testu t-Studenta dla pojedynczej próby](#) warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy:

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

Zadana wielkość efektu, to w tym przypadku wielkość różnicy między zadaną średnią populacji badanej a hipotetyczną wartością średnią.

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o niecentralny rozkład t-Studenta.

PRZYKŁAD 7.3. 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 jakie spodziewamy się uzyskać w planowanym badaniu - my zakładamy (na bazie doświadczeń pracowników tej 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

Uzyskana niezbędna liczność, która pozwoli na udowodnienie, że efekt przekraczający 12 godzin jest istotny statystycznie wynosi 34 przesyłki.

Odp 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).

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

Uzyskana niezbędna liczność, która pozwoli na udowodnienie, że efekt przekraczający 6 godzin jest istotny statystycznie wynosi 128 przesyłek.

- **Badanie 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 jednocześnie niezmienną pozostałe założenia?

Odp a) Moc przeprowadzonej analizy wynosi zaledwie 0.400302.

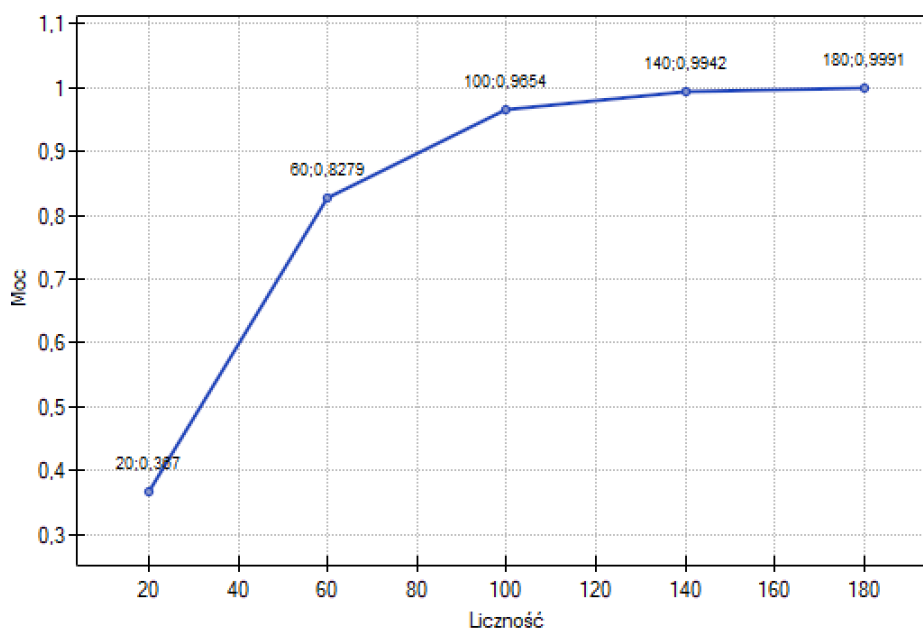
Moc testu i liczność próby	
Czas analizy	0,00 sek.
Poziom istotności	0,05
Test t-Studenta dla pojedynczej próby	
Hipoteza	dwustronna
Liczność - zadana	22
Średnia hipotetyczna - zadana	3
Średnia grupy - zadana	3,7273
Odchylenie standardowe - zadane	1,9069
Wielkość efektu (różnica średnich)	0,7273
Moc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

Wiemy stąd, że wiele losowych prób o licznosci równej 22 (około 60% takich prób) nie będzie prowadziło do potwierdzenia hipotezy alternatywnej.

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

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 będzie się zmieniała moc analizy przy zmieniającej się licznosci próby i niezmiennych pozostałych założeniach możemy zobaczyć na wykresie.



7.2.2 Test t-Studenta dla grup zależnych

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności testu *t*-Studenta dla grup zależnych warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy:

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	

Zadana wielkość efektu, to w tym przypadku wielkość różnicy między średnimi jaką spodziewamy się uzyskać w populacji.

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o niecentralny rozkład *t*-Studenta.

PRZYKŁAD 7.4. 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. Przy czym 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 znajdują się osoby zdyscyplinowane by przestrzegać 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

Uzyskana niezbędna liczność to 199 osoby, gdy hipoteza jest dwustronna (tzn. zakładamy, że na skutek diety BMI może spaść lub wzrosnąć) lub 156 osób, gdy hipoteza jest jednostronna (tzn. zakłada jedynie spadek BMI).

Gdybyśmy założyli, że grupa będzie bardziej zdyscyplinowana i odchylenie standardowe różnicy wyniesie 1.5 jednostki BMI, wówczas próba będzie mogła być nieco mniejsza tzn. 73 osoby w badaniu hipotezy dwustronnej i 58 osób przy hipotezie jednostronnej.

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

7.2.3 Test t-Studenta dla grup niezależnych

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności [testu t-Studenta dla grup niezależnych](#) warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy:

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

***Uwaga!**

Gdy znane są licznosci, wówczas 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 licznosci nie są znane wówczas stosowany jest uproszczony wzór Cohena (Cohen, J. (1988)[45])

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


gdzie: n_1, n_2 - licznosci w pierwszej i drugiej próbie,

sd_1, sd_2 - odchylenia standardowe w pierwszej i drugiej próbie.

Zadana wielkość efektu, to w tym przypadku wielkość różnicy między średnimi jaką spodziewamy się uzyskać pomiędzy populacjami.

Moc testu oraz niezbędna licznosc próby wyliczana jest w oparciu o niecentralny rozkład t-Studenta.

PRZYKŁAD 7.5. Badamy mężczyzn chorych na chorobę X i zdrowych. Chcemy sprawdzić (na poziomie istotności 0.05) czy chorzy różnią się od zdrowych poziomem cholesterolu HDL. Przy czym 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 licznosci obu grup to 1, ponieważ zakładaliśmy równoliczne grupy. Dysponujemy danymi z badania pilotażowego, stąd 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 licznosc 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 licznosci - 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
Licznosc	571 / 571	855 / 428

Uzyskana niezbędna licznosc to 571 osób, gdy grupy są równoliczne (tzn. zakładamy, że na skutek $n_1/n_2=1$) lub $n_1 = 855$ a $n_2 = 428$, gdy nie są równoliczne (tzn. zakładamy stosunek licznosci $n_1/n_2=2$)

7.2.4 Test chi-kwadrat wariancji pojedynczej próby

Przed przystąpieniem do wyznaczania mocy lub niezbędnej licznosci testu chi-kwadrat wariancji pojedynczej próby warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej licznosci próby potrzebujemy:

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

Zadana wielkość efektu, to w tym przypadku iloraz odchylenia standardowego populacji badanej i odchylenia hipotetycznego.

Moc testu oraz niezbędna licznosc próby wyliczana jest w oparciu o [rozkład chi-kwadrat](#).

PRZYKŁAD 7.6. 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 tą wartość wpisujemy w polu dotyczącym odchylenia standardowego grupy butelek, którą będziemy badać.

Jeśli próba będzie liczyła 20 butelek, to uzyskana moc przy hipotezie dwustronnej wyniesie zaledwie 0.25, a zakładając hipotezę jednostronną 0.34. Są to wartości niskie bo mniej niż 35% losownych prób tej wielkości pozwoli wykryć przekroczenie rozrzutu o 0.2 ml.

Moc testu i licznosc 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

Trzeba uznać, że 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 i wyliczamy niezbędną licznosc, która w tym przypadku wyniesie 115 w hipotezie dwustronnej i 92 w jednostronnej.

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 - zadane	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

7.2.5 Test chi-kwadrat dwóch wariancji Fishera-Snedecora

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności [testu Fishera-Snedecora](#) warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy:

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

Zadana wielkość efektu, to w tym przypadku iloraz odchylenia standardowego populacji pierwszej i drugiej.

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o [rozkład F-Snedecora](#).

PRZYKŁAD 7.7. 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 wartość 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 odchyień)	0,8333	0,8333
Moc przybliżona	0,8015	0,8006
Liczność	239 / 239	363 / 182

Uzyskana wielkość próby dla każdego urządzenia wynosi $n_1=n_2=239$, przy założeniu równolicznych grup (tzn. stosunek $n_1/n_2=1$) i $n_1=363$ oraz $n_2=182$ przy założeniu dwukrotnie większej próby dla pierwszego urządzenia (tzn. stosunek $n_1/n_2=2$).

7.2.6 Test chi-kwadrat zgodności

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności [testu chi-kwadrat zgodności](#) warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy[45]:

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

Zadana wielkość efektu, czyli ϕ to w tym przypadku pierwiastek z ilorazu statystyki testu chi-kwadrat i liczności grupy badanej:

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

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o niecentralny rozkład chi-kwadrat.

7.2.7 Test chi-kwadrat (RxC)

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności [testu chi-kwadrat dla tabel RxC](#) warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy[45][128]:

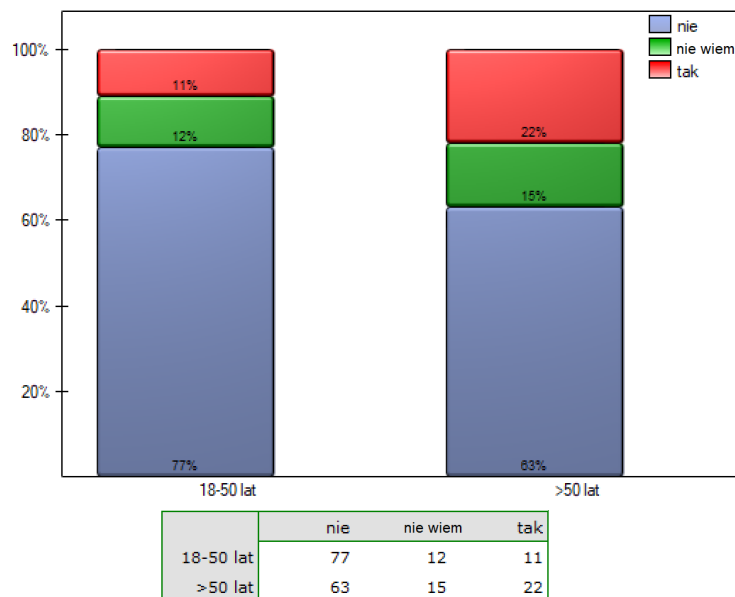
Moc testu	Liczność grupy
Liczność grupy Liczba kategorii (wierszy) Liczba kategorii (kolumn)	Moc testu
Wielkość efektu ϕ	
Poziom istotności α	

Zadana wielkość efektu, czyli ϕ to w tym przypadku pierwiastek z ilorazu statystyki testu chi-kwadrat i liczności grupy badanej:

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

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o niecentralny rozkład chi-kwadrat.

PRZYKŁAD 7.8. 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 losowo wybranej próbie 200 osób zadano pytanie: „Czy antybiotyki zwalczają wirusy?”. Respondenci mieli możliwość wskazania jednej z trzech odpowiedzi: „nie” lub „nie wiem”, „tak”. Wyniki badania pilotażowego przygotowano do publikacji. Poniżej zamieszczono fragment opisu umieszczonego w pracy:



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, iż 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

Wyznamy współczynnik ϕ dla uzyskanego testu (menu: Chi-kwadrat, Fisher, OR/RR→Współczynniki korelacji...→Phi). Uzyskujemy $\phi=0.1643$.

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

Uzyskana moc w tej analizie jest niewielka i wynosi 0.5368, co wydaje się potwierdzać obawy o niewystarczająco licznej próbie.

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

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

Uzyskujemy informację, że potrzebna będzie próba wielkości 357 ankietowanych. Ponieważ jest to dopiero badanie pilotażowe, planujemy zwiększyć grupę badaną do 357 osób w badaniu właściwym. Jednak 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 >50 lat (12.5% vs 25.9%).

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

Przed przystąpieniem do wyznaczania mocy lub niezbędnej licznosci testu chi-kwadrat dla tabel 2x2 i testu Z dla dwóch niezależnych proporcji warto zapoznać się z zasadami ich stosowania.

Do wyznaczenia mocy testu i niezbędnej licznosci próby potrzebujemy[35]:

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

Zadana wielkość efektu, to różnica między wyróżnionymi proporcjami.

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o [rozkład normalny](#).

PRZYKŁAD 7.9. 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 licznosci - 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

Zakładając, że grupy będą równoliczne, do każdej z nich trzeba zebrać 635 osób.

7.2.9 ANOVA dla grup niezależnych

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności [ANOVA dla grup niezależnych](#) warto zapoznać się z zasadami jej stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy:

Moc testu	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

Zadana wielkość efektu, to w tym przypadku RMSSE czyli miara standaryzowana stosowana w ANOVA do opisanego ogólnego poziomu efektu w populacji.

Moc testu oraz niezbędna liczność próby wyliczana jest w oparciu o niecentralny rozkład F-Snedecora.

PRZYKŁAD 7.10. Badano parametr FVC oraz 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 odchyłeń standardowych. Uzyskana wielkość próby dla każdej z grup badanych to 33 osoby, przy założeniu 80% mocy testu.

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

7.2.10 Test dla jednej proporcji

Przed przystąpieniem do wyznaczania mocy lub niezbędnej liczności **testu dla jednej proporcji** warto zapoznać się z zasadami jego stosowania.

Do wyznaczenia mocy testu i niezbędnej liczności próby potrzebujemy:

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

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 wyliczana jest w oparciu o **rozkład normalny**, gdy korzystamy z testu asymptotycznego lub rozkład dwumianowy, gdy korzystamy z testu dokładnego.

PRZYKŁAD 7.11. Zapytano 10 wylosowanych rodzin posiadających dzieci w wieku poniżej 10 lat i zamieszkujących w Poznaniu o planowaną przyszłość edukacyjną swoich dzieci. Spośród nich 6 rodziny 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 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. Uzyskana wielkość próby to 194 rodzin - gdy badana hipoteza jest dwustronna lub 151 rodzin- gdy hipoteza jest jednostronna.

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

8 ANALIZY OPISOWE

Dane zebrane przez badacza w pierwszej kolejności powinny zostać opisane. W zależności od sposobu dokonywania pomiarów (od skali pomiarowej) do opisu zmiennej użyjemy różnych miar.

8.1 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 i różnica tych elementów 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.

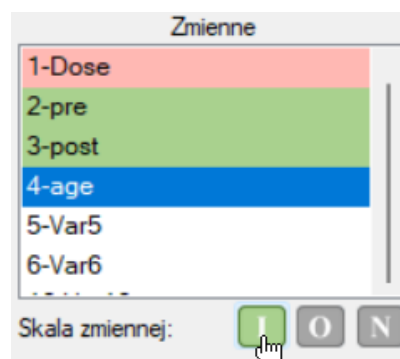
Przykład: płeć, kraj zamieszkania itp.

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 zalecane jest przyporządkowanie skali pomiarowej 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.



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.

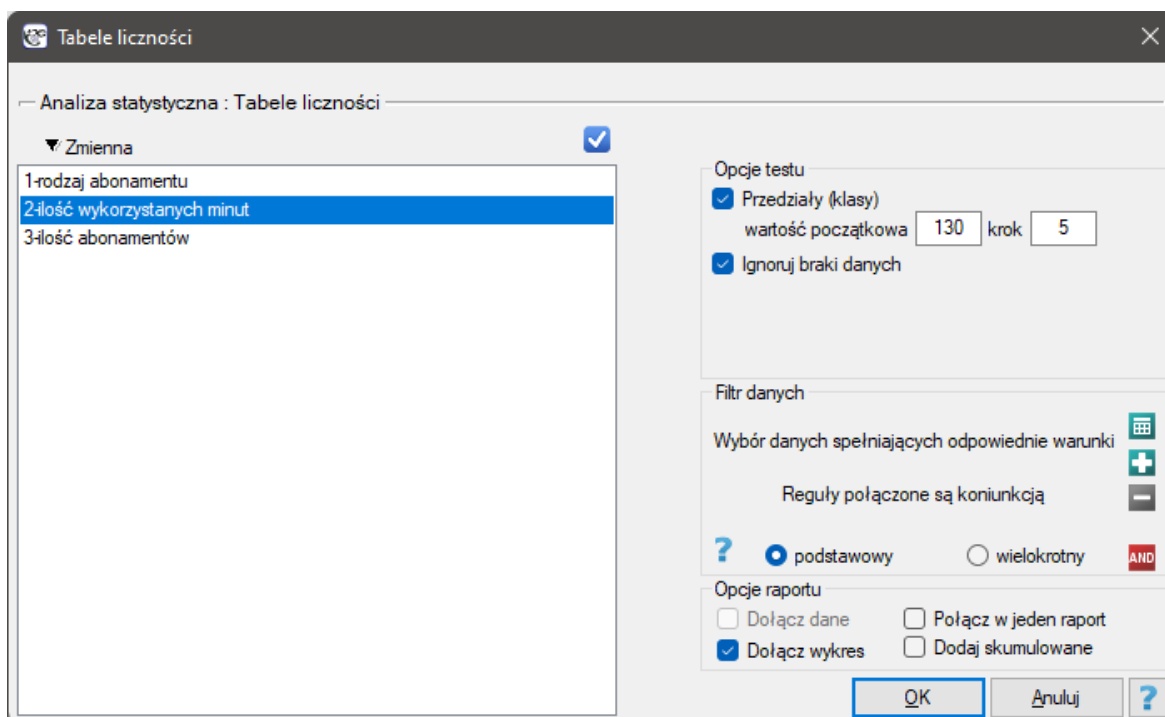
Skala porządkowa która ma możliwych wiele zobiektywizowanych kategorii to dane ilościowe np. skala jakości życia SF-36 (od 0 do 100 punktów), ale jeśli kategorii jest tak niewiele, że można je opisać tekstem np. wykształcenie (podstawowe, średnie, wyższe), to są to dane jakościowe.

8.2 TABELLE

8.2.1 TABELLE LICZNOŚCI I ROZKŁADY EMPIRYCZNE DANYCH

Podstawą badań statystycznych jest określenie **rozkładu empirycznego** tzn. zaobserwowanego w próbie rozkładu cechy. Określenie empirycznego rozkładu polega na przyporządkowaniu kolejnym wartościom przyjmowanym przez cechę częstości ich występowania. Rozkład taki można przedstawić w postaci **tabeli liczności** lub w postaci wykresu (histogramu). Dla małych zbiorów danych tabele liczności mogą prezentować wszystkie dane - tzw. szeregi rozdzielcze punktowe, w przypadku większych zbiorów danych tworzy się tzw. szeregi rozdzielcze przedziałowe.

Aby przedstawić rozkład danych w postaci tabeli należy wyświetlić okno Tabele liczności poprzez wybranie menu Statystyka→Analizy opisowe→Tabele liczności.



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 występują w analizowanej kolumnie puste komórki, to mogą być wliczane do analizy bądź pomijane. Wynik dokonanej analizy znajdzie się w raporcie dołączonym do arkusza danych, dla których analiza została wykonana.

Dodatkowo, jeśli chcemy by dane zostały zobrazowane za pomocą wykresu kolumnowego lub histogramu, wówczas w oknie Tabel licznosci zaznaczamy opcję Dołącz wykres.

PRZYKŁAD 8.1. (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,
- liczby wykorzystanych darmowych minut,
- liczby zarejestrowanych na danego klienta abonamentów (nie dotyczy firm).

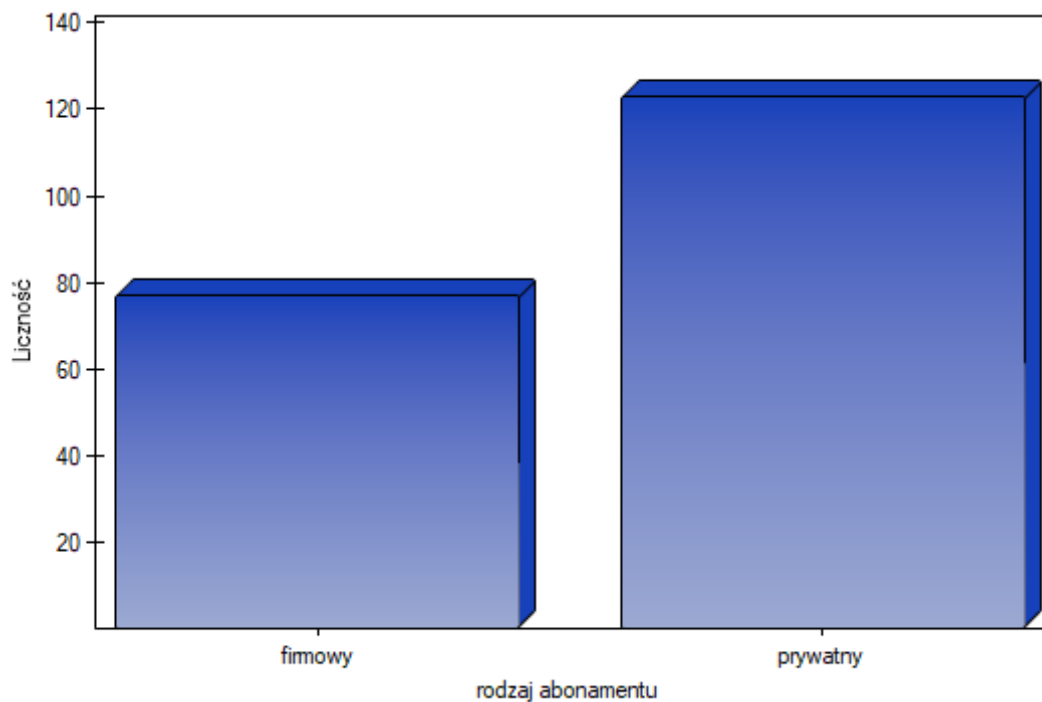
Chcemy przedstawić rozkład:

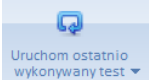
1. rodzaju abonamentu,
2. liczby wykorzystanych darmowych minut,
3. liczby zarejestrowanych abonamentów na osoby prywatne.

Uruchamiamy okno Tabele licznosci.

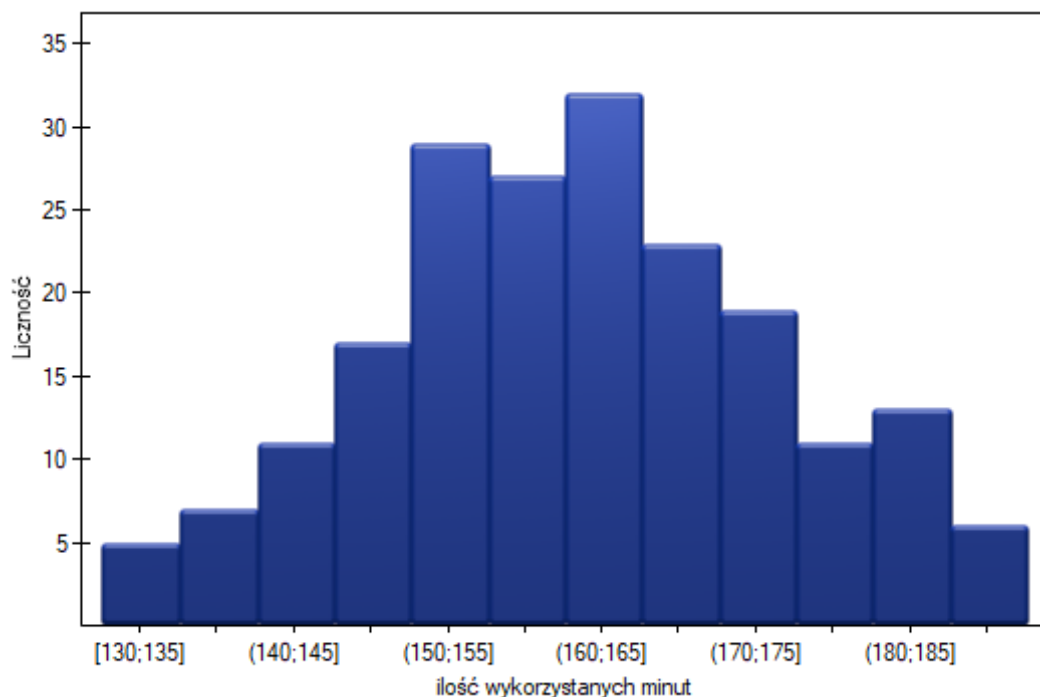
1. Wybieramy Zmienną do analizy: "rodzaj abonamentu" i zaznaczamy opcję oraz Dołącz wykres. Następnie potwierdzamy wybrane ustawienia przyciskiem OK i uzyskujemy wynik w postaci raportu:


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:

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:

Tabele liczności		
Analizowane zmienne	ilość abonamen	
Filtr danych:	rodzaj abonam	
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 abonam	
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 8.2. (plik nawozy.pqs)

Przeprowadzono doświadczenie, w którym badano stan mikrobiologiczny gleby pod uprawą żywiczy 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:

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%

8.2.2 RAPORT TABELI

Przy pomocy raportu tabeli można przygotować jednoczesne podsumowanie bardzo wielu danych w postaci tabel dwudzielczych (tabel dwóch cech). Na przykład w postaci tabeli 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 oraz wyznaczyć tabelę liczności [oczekiwanych](#). Ponadto dla takich tabel możliwe jest automatyczne podsumowania w postaci wykresu kolumnowego. Okno z ustawieniami opcji Raportu tabel wywołujemy poprzez menu Statystyka→Analizy opisowe→Raport tabeli

Raport tabeli (wielokrotny) X

Analiza statystyczna : Raport tabeli

▼ Zmienna 1

- 1-Nr
- 2-Płeć
- 3-Wiek (w latach)
- 4-Miejsce zamieszkania
- 5-Warunki socjalno-bytowe
- 6-Wykształcenie
- 7-Stan cywilny
- 8-Var7
- 9-Var8
- 10-Var9
- 11-Var10
- 12-Var11
- 13-Var12
- 14-Var13
- 15-Var14
- 16-Var15
- 17-Var16
- 18-Var18
- 19-Var19

▼ Zmienna 2

- 1-Nr
- 2-Płeć
- 3-Wiek (w latach)
- 4-Miejsce zamieszkania
- 5-Warunki socjalno-bytowe
- 6-Wykształcenie
- 7-Stan cywilny
- 8-Var7
- 9-Var8
- 10-Var9
- 11-Var10
- 12-Var11
- 13-Var12
- 14-Var13
- 15-Var14
- 16-Var15
- 17-Var16
- 18-Var18
- 19-Var19

Opcje testu

% wiersza

% kolumny

% sumy

Oczekiwane

Filtr danych

Wybór danych spełniających odpowiednie warunki +

Reguły połączone są koniunkcją -

podstawowy wielokrotny AND

Opcje raportu

Połącz w jeden raport

Dołącz wykres

PRZYKŁAD 8.3. (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 tabeli dla każdej pary cech, czyli 8 tabel dla wszystkich par i odpowiadające im wykresy. Poniżej przedstawiono tylko zestawienie względem płci:

[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	

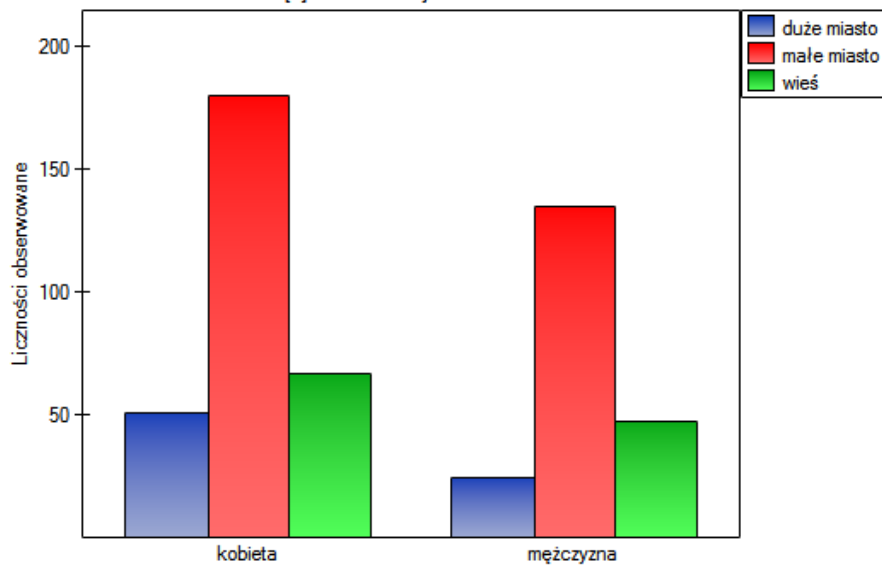
[3] Płeć <> Wykształcenie

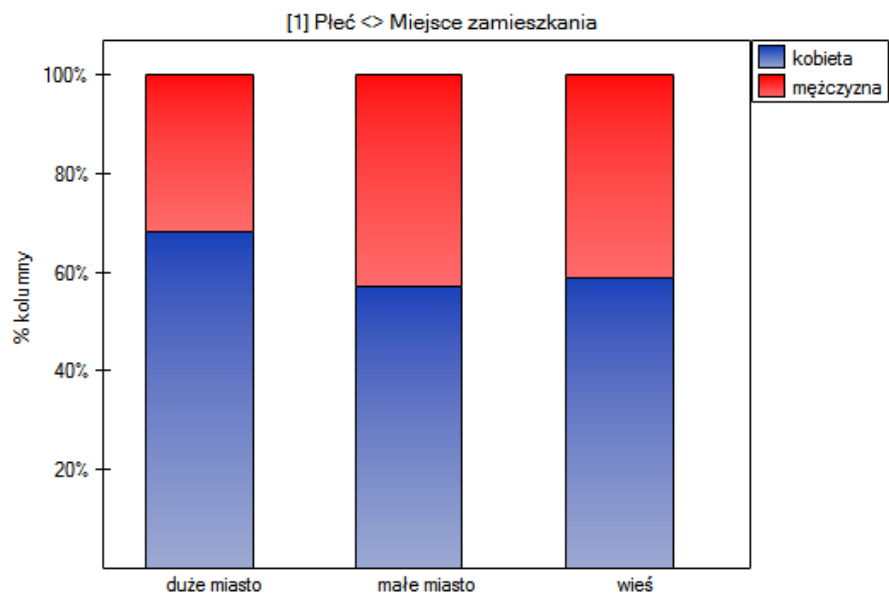
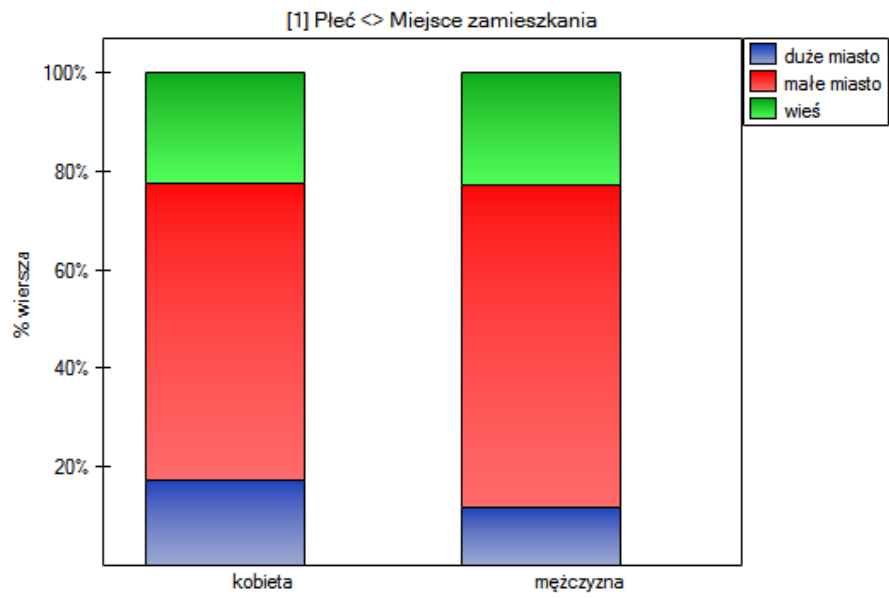
Dane	podstawowe	wyższe	zawodowe	średnie	Razem
kobieta	90	43	59	106	298
mężczyzna	33	39	76	58	206
Razem	123	82	135	164	504
% wiersza	podstawowe	wyższe	zawodowe	średnie	
kobieta	30,2%	14,43%	19,8%	35,57%	
mężczyzna	16,02%	18,93%	36,89%	28,16%	
% kolumny	podstawowe	wyższe	zawodowe	średnie	
kobieta	73,17%	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	

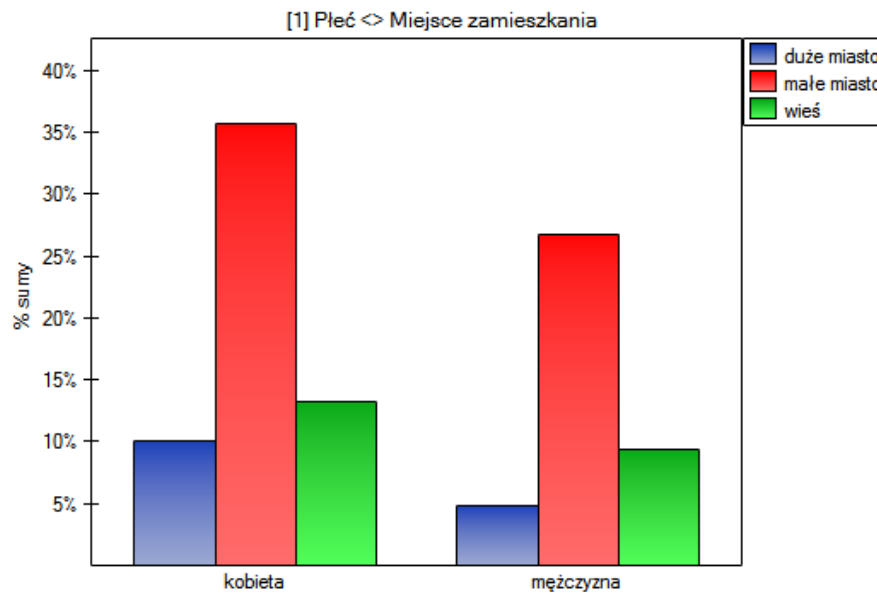
[4] Płeć <> Stan cywilny

Dane	mężatka/zon	panna/kawał	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/zon	panna/kawał	rozwódka/ro:	wdowa/w	
kobieta	37,58%	8,05%	5,37%	48,99%	
mężczyzna	63,59%	5,34%	8,74%	22,33%	
% kolumny	mężatka/zon	panna/kawał	rozwódka/ro:	wdowa/w	
kobieta	46,09%	68,57%	47,06%	76,04%	
mężczyzna	53,91%	31,43%	52,94%	23,96%	
% sumy	mężatka/zon	panna/kawał	rozwódka/ro:	wdowa/w	
kobieta	22,22%	4,76%	3,17%	28,97%	
mężczyzna	25,99%	2,18%	3,57%	9,13%	
Oczekiw.:	mężatka/zon	panna/kawał	rozwódka/ro:	wdowa/w	
kobieta	143,68	20,69	20,1	113,52	
mężczyzna	99,32	14,31	13,9	78,48	

[1] Płeć <> Miejsce zamieszkania







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

8.2.3 ANALIZY DLA TABEL KONTYNGENCJI

Analizy dla tabel kontyngencji mogą być wyliczane na podstawie danych zebranych w [tabelach kontyngencji](#) lub bezpośrednio tzn. na podstawie danych w postaci [surowej](#). Przy czym istnieje możliwość [transformacji](#) danych z tabeli kontyngencji do postaci surowej lub odwrotnie.

PRZYKŁAD 8.4. (plik płeć-wykształcenie.pqs)

Rozpatrzmy próbkę 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.

Chi-kwadrat (RxC)

Analiza statystyczna : Test Chi-kwadrat (RxC)

▼ Zmienna 1

1-Lp
2-pleć
3-wykształcenie

▼ Zmienna 2

1-Lp
2-pleć
3-wykształcenie

Tabela kontyngencji Dane surowe

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny **AND**

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

Dołącz wartości procentowe wiersze

0,05 poziom istotności

OK Zamknij

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 i dane z zaznaczenia zostaną wyświetlone.

Chi-kwadrat (RxC)

Analiza statystyczna : Test Chi-kwadrat (RxC)

wypełnij z zapisanego zaznaczenia

	c1	c2	c3
r1	4	7	4
r2	8	6	5

Tabela kontyngencji Dane surowe

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

Dołącz wartości procentowe wiersze+kolumny

0,05 poziom istotności

OK Zamknij

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

Warunek Cochra

Jest to podstawowy warunek stosowania wielu testów statystycznych opartych na tabelach kontyngencji np. testu chi-kwadrat. Warunek ten zakłada duże **liczności oczekiwane**. Według interpretacji Cochra na 1952[40] żadna z liczności oczekiwanych nie może być < 1 oraz nie więcej niż 20% liczności oczekiwanych może być < 5 . Informacja o spełnieniu (bądź nie spełnieniu) 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:

- 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

W raporcie wynikowym można również umieścić podstawowe podsumowanie tabel:

- **Tabelę 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(8.1)).

Tabela 8.1. 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} ($i = 1, 2, \dots, r; j = 1, 2, \dots, c$) przedstawiają częstość występowania poszczególnych kategorii dla obu cech.

By tabela taka była zwrócona przez program należy w oknie testu wybrać opcję dołącz analizowane dane. Dla danych z przykładu (8.4) tabela kontyngencji liczności obserwowanych przedstawia się następująco:

Dane :	podstawov	wyższe	średnie
kobieta	4	7	4
mężczyźni	8	6	5

- **Tabelę kontyngencji licznosci oczekiwanych** – dla każdej tabeli kontyngencji licznosci obserwowanych można utworzyć odpowiadającą jej tabelę **licznosci oczekiwanych**: E_{ij} (tabela(8.2)).

Tabela 8.2. Tabela kontyngencji $r \times c$ licznosci oczekiwanych

Licznosci 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}

gdzie:

$$E_{11} = \frac{\sum_{i=1}^r O_{i1} \times \sum_{j=1}^c O_{1j}}{n}, E_{12} = \frac{\sum_{i=1}^r O_{i2} \times \sum_{j=1}^c O_{1j}}{n}, E_{1c} = \frac{\sum_{i=1}^r O_{ic} \times \sum_{j=1}^c O_{1j}}{n}$$

$$E_{21} = \frac{\sum_{i=1}^r O_{i1} \times \sum_{j=1}^c O_{2j}}{n}, E_{22} = \frac{\sum_{i=1}^r O_{i2} \times \sum_{j=1}^c O_{2j}}{n}, E_{2c} = \frac{\sum_{i=1}^r O_{ic} \times \sum_{j=1}^c O_{2j}}{n}$$

$$E_{r1} = \frac{\sum_{i=1}^r O_{i1} \times \sum_{j=1}^c O_{rj}}{n}, E_{r2} = \frac{\sum_{i=1}^r O_{i2} \times \sum_{j=1}^c O_{rj}}{n}, E_{rc} = \frac{\sum_{i=1}^r O_{ic} \times \sum_{j=1}^c O_{rj}}{n}.$$

Dla danych z przykładu (8.4) tabela kontyngencji licznosci oczekiwanych przedstawia się następująco:

Oczekiw.:	podstawov	wyższe	średnie
kobieta	5.29	5.74	3.97
mężczyźni	6.71	7.26	5.03

- **Tabelę kontyngencji wartosci procentowych wyliczanych z sumy kolumn.** Dla danych z przykładu (8.4) tabela ta przedstawia się następująco:

kolumny:	podstawov	wyższe	średnie
kobieta	33.33%	53.85%	44.44%
mężczyźni	66.67%	46.15%	55.56%

- **Tabelę kontyngencji wartosci procentowych wyliczanych z sumy wierszy.** Dla danych z przykładu (8.4) tabela ta przedstawia się następująco:

wiersze:	podstawov	wyższe	średnie
kobieta	26.67%	46.67%	26.67%
mężczyźni	42.11%	31.58%	26.32%

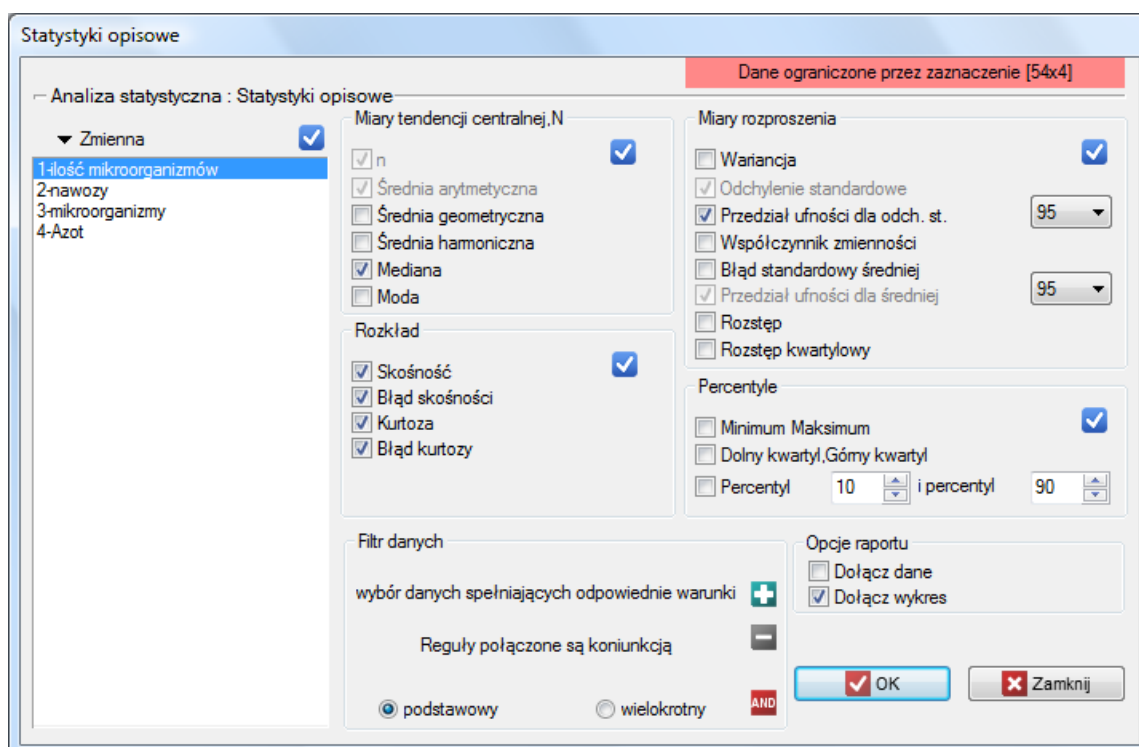
- Tabelę kontyngencji wartości procentowych wyliczanych z sumy całkowitej wierszy i kolumn. Dla danych z przykładu (8.4) tabela ta przedstawia się następująco:

sumy:		podstawowy	wyższe	średnie
kobieta	11,76%	20,59%	11,76%	
mężczyźni	23,53%	17,65%	14,71%	

8.3 STATYSTYKI OPISOWE

Celem stosowania metod statystyki opisowej jest podsumowanie zbioru danych poprzez pewne charakterystyki np. poprzez wartość średniej, mediany czy odchylenia standardowego, oraz wyciągnięcie pewnych podstawowych wniosków i uogólnień na temat zbioru.

Aby wyznaczyć statystyki opisowe dla danych zgromadzonych w arkuszu należy wyświetlić okno Statystyki opisowe poprzez wybranie menu Statystyka → Analizy opisowe → Statystyki opisowe.



W oknie tym wybieramy zmienną do analizy oraz opcje analizy i zaznaczamy interesujące nas miary statystyk opisowych. Przy czym zaznaczać można pojedyncze statystyki lub grupy statystyk wybierając przycisk . Dokonany wybór potwierdzamy przyciskiem OK. Wynik dokonanej analizy znajdzie się w raporcie dołączonym do arkusza danych, dla których analiza została wykonana.

Dodatkowo, jeśli chcemy by dane zostały zobrazowane za pomocą wykresu ramka-wąsy, wówczas w oknie Statystyk opisowych zaznaczamy opcję Dołącz wykres.

8.3.1 MIARY POŁOŻENIA

8.3.2 MIARY TENDENCJI CENTRALNEJ

Miary tendencji centralnej są to tzw. miary przeciętne charakteryzujące średni lub typowy poziom wartości cechy.

Średnia arytmetyczna (*ang. arithmetic mean*) wyraża się wzorem:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n},$$

gdzie x_i to kolejne wartości zmiennej a n - licznosc próby.

Średnia arytmetyczna jest stosowana dla [skali interwałowej](#). Dla próby przyjmuje się ją oznaczać przez \bar{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. Przy czym, gdy uzyskana na podstawie przeliczeń liczba pomiarów przeznaczonych do usunięcia nie będzie liczbą całkowitą, wówczas 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 zmniejszonym zbiorze wartości. Jeśli zdecydujemy się na obliczanie średniej Winsora przycinając np. 5% pomiarów, to wówczas 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 na liczbę pomiarów przeznaczonych do zamiany nie uzyskamy liczby całkowitej, wówczas zaokrąglamy w dół do najbliższej całkowitej. **Średnia geometryczna** (*ang. geometric mean*) wyraża się wzorem:

$$\bar{x}_G = \sqrt[n]{x_1 x_2 \dots x_n} = \sqrt[n]{\prod_{i=1}^n x_i}.$$

Średnia ta jest stosowana dla [skali interwałowej](#), gdy zmienna ma rozkład logarytmiczno-normalny (logarytm zmiennej ma [rozkład normalny](#)).

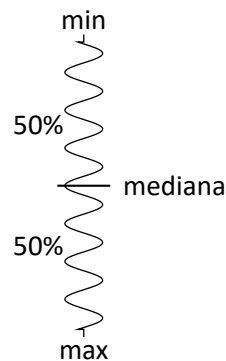
Średnia harmoniczna (*ang. harmonic mean*) wyraża się wzorem:

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}.$$

Średnia ta 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.



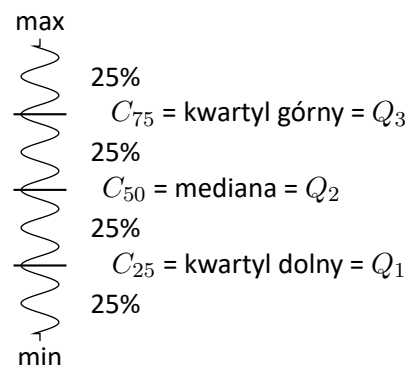
Mediana może być stosowana w [skali interwałowej](#) oraz [porządkowej](#).

Moda (ang. mode)

Moda – jest to wartość, która występuje najczęściej wśród uzyskanych pomiarów. Moda może być stosowana w każdej [skali pomiarowej](#).

8.3.3 INNE MIARY POŁOŻENIA

kwartyle (ang. *quartiles*), **decyle** (ang. *deciles*), **centyle** (ang. *centiles*)



Kwartyle (Q_1, Q_2, Q_3) dzielą uporządkowany szereg na 4 równe części, decyle ($D_i, i = 1, 2, \dots, 9$) na 10 równych części a centyle (percentyle: $C_i, i = 1, 2, \dots, 99$) na 100 równych części. Drugi kwartył, piąty decyl i pięćdziesiąty centyl są równe medianie. Miary te mogą być stosowane w [skali interwałowej](#) oraz [porządkowej](#).

8.3.4 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 formuły pozwalające wyznaczyć wartości, które charakteryzują rozrzut danych.

Miary rozproszenia są liczone tylko dla [skali interwałowej](#), ponieważ bazują one na odległościach między punktami.

Rozstęp (ang. *range*) wyraża się wzorem:

$$I = \max x_i - \min x_i,$$

gdzie x_i to wartości badanej zmiennej

$$IQR = \text{rozstęp kwartylowy} = Q_3 - Q_1,$$

gdzie Q_1, Q_3 to dolny i górny kwartył.

Rozstępy dla skali procentylowej (decylowej, centylowej).

Rozstępy między percentylami to jedna z miar rozproszenia i określa procent wszystkich obserwacji, których wartość znajduje się pomiędzy wybranymi percentylami.

Wariancja (*ang. variance*) – mierzy stopień rozproszenia pomiarów wokół średniej arytmetycznej

wariancja z próby:

$$sd^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1},$$

gdzie x_i to kolejne wartości zmiennej a \bar{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 a μ to średnia arytmetyczna tych wartości, N - liczność populacji.

Wariancja jest zawsze dodatnia, ale nie jest wyrażona w tych samych jednostkach co wyniki pomiarów.

Odchylenie standardowe (*ang. standard deviation*) – mierzy stopień rozproszenia pomiarów wokół średniej arytmetycznej.

odchylenie standardowe z próby:

$$sd = \sqrt{sd^2},$$

odchylenie standardowe z populacji:

$$\sigma = \sqrt{\sigma^2}.$$

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 na ocenę stopnia jednorodności badanej zbiorowości. Wyraża się wzorem:

$$V = \frac{sd}{\bar{x}} 100\%,$$

gdzie sd to odchylenie standardowe, \bar{x} to średnia arytmetyczna.

Jest to wielkość niemianowana. Pozwala on na ocenę 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 określają stopień dokładności z jaką możemy określić wartości parametrów populacji na podstawie wyznaczenia ich estymatorów dla próby.

Błąd średniej arytmetycznej (*ang. standard error of the mean*) wyraża się następującym wzorem:

$$SEM = \text{błąd standardowy średniej} = \frac{sd}{\sqrt{n}}.$$

Uwaga!

Na podstawie błędu standardowego estymatora z próby można określić **przedział ufności** dla parametru populacji.

8.3.5 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 $[-1, 1]$, chociaż może w przypadku szczególnie dużej asymetrii znaleźć się poza tym przedziałem. Wartości dodatnie świadczą o występowaniu skośności prawostronnej (o dłuższym prawym "ogonie") wartości ujemne zaś o skośności lewostronnej (o dłuższym lewym "ogonie"). Skośność wyraża się wzorem:

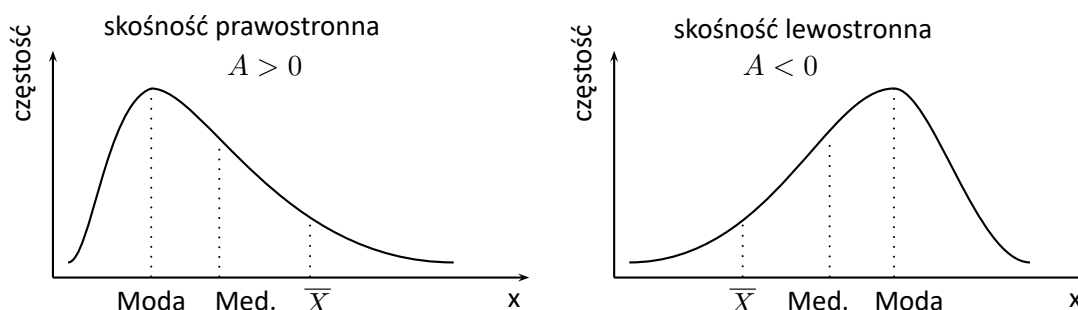
$$A = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{sd} \right)^3,$$

gdzie:

x_i – kolejne wartości zmiennej,

\bar{x} , sd – odpowiednio średnia arytmetyczna i odchylenie standardowe x_i ,

n – licznosc próby.



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 wartość kurtozy jest mniejsza od zera, tym badany rozkład jest bardziej spłaszczony niż rozkład normalny. Kurtoza wyraża się wzorem:

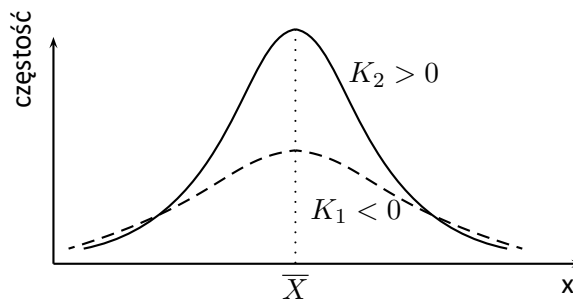
$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{sd} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)},$$

gdzie:

x_i – kolejne wartości zmiennej,

\bar{x} , sd – odpowiednio średnia arytmetyczna i odchylenie standardowe x_i ,

n – liczność próby.

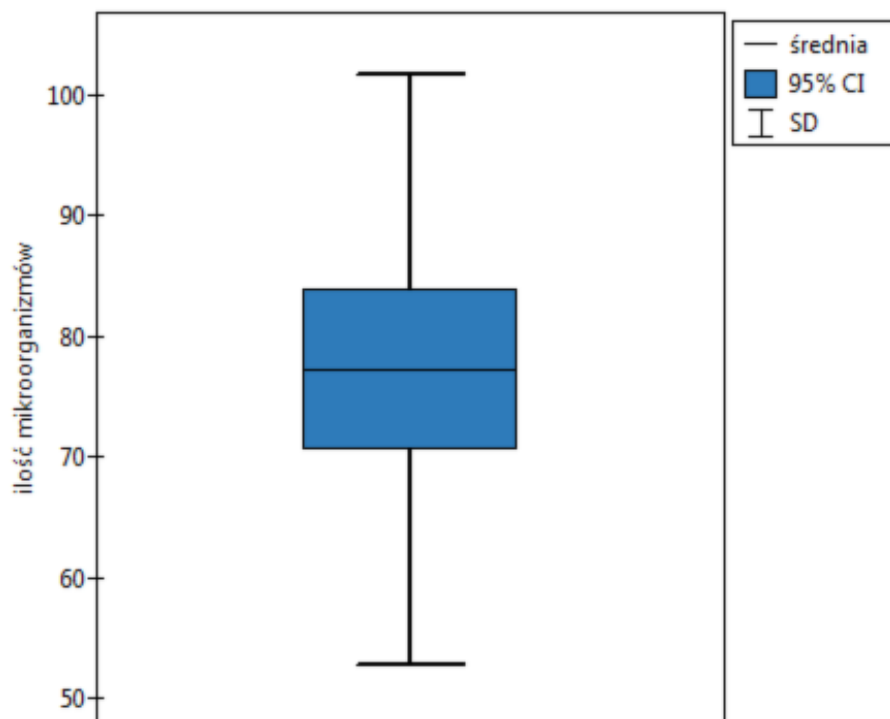


PRZYKŁAD 8.5. (plik nawozy.pqs)

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



8.4 ZESTAWIENIA OPISOWE

Zestawienia opisowe to szybkie przygotowanie raportu przedstawiającego opis danych, który jest gotowy do umieszczenia bezpośrednio w pracy naukowej. Jest to narzędzie ułatwiające zestawienie wszystkich form podstawowego opisu danych w jednym miejscu.

Okno z ustawieniami opcji testu Zestawień opisowych wywołujemy poprzez menu Statystyka → Analizy opisowe → Zestawienia opisowe

Zestawienia opisowe

Analiza statystyczna : Zestawienia opisowe

Dla statystyk opisowych

Dla licznosci i procentow

Zmienna grupujaca

1-wiek
2-wiek przedzialy
3-poziom bolu
4-poziom bolu przedzialy
5-palenie papierosow
6-metoda leczenia

1-wiek
2-wiek przedzialy
3-poziom bolu
4-poziom bolu przedzialy
5-palenie papierosow
6-metoda leczenia

1-wiek
2-wiek przedzialy
3-poziom bolu
4-poziom bolu przedzialy
5-palenie papierosow

Opcje testu

$\Sigma\mu$

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguly połączone są koniunkcją

? podstawowy wielokrotny AND

Opcje raportu

Połącz w jeden raport

0.05 poziom istotności

OK Anuluj ?

Zestawienia opisowe

Statystyki opisowe i normalność rozkładu

Funkcja	Miejsca dzi	Prefiks	Funkcja	Miejsca dzi	Prefiks
Średnia	2	±	SD	2	
Mediana	2	[Q1; Q3	2]
Wartość p (S-W)	2			2	

Licznosc i procent

Funkcja	Miejsca dzi	Prefiks	Funkcja	Miejsca dzi	Prefiks
Licznosc	0	(Procent	2)

Braki danych

Funkcja	Miejsca dzi	Prefiks	Funkcja	Miejsca dzi	Prefiks
Licznosc	0	(Procent	2)

Inne

Dodaj kolumnę: Razem

OK

W zależności od sposobu dokonywania pomiarów do opisu zmiennej używamy najczęściej jednej z trzech typów miar:

- średnia \pm odchylenie standardowe, w skrócie $\bar{x} \pm sd$
- mediana [kwartył dolny; kwartył 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 potrzeby 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.

PRZYKŁAD 8.6. (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 \pm SD	66.38 \pm 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%)

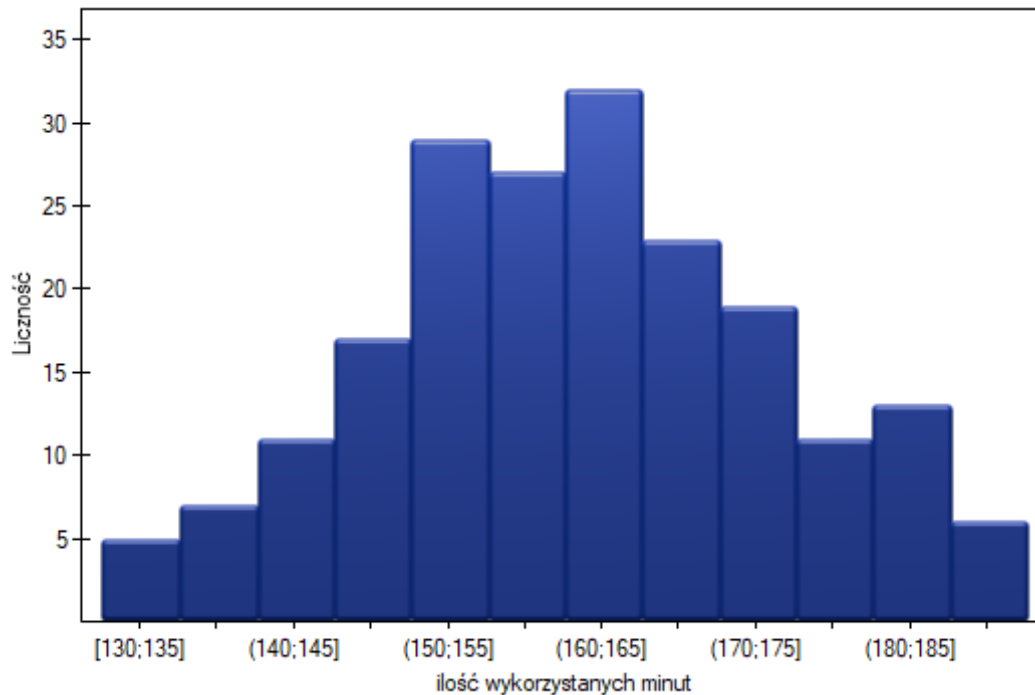
Zestawienia opisowe		metoda leczenia	
		A(N=44)	B(N=56)
wiek	Średnia ± SD	66.27±17.7	66.46±15.29
	Mediana [Q1; Q3]	64 [56; 74.5]	65.5 [54.75; 76.25]
wiek przedziały	>=70	17(38.64%)	23(41.07%)
	60-69	11(25%)	11(19.64%)
	50-59	9(20.45%)	14(25%)
	40-49	5(11.36%)	7(12.5%)
	<=39	2(4.55%)	1(1.79%)
poziom bólu 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%)

9 ROZKŁADY PRAWDOPODOBIENSTWA

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 (8.1), plik: rozklad.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.



Wykres taki możemy uzyskać zaznaczając w oknie Tabeli licznosci opcję Dołącz wykres.

Rozkład teoretyczny danych zwany również **rozkładem prawdopodobieństwa** graficznie przedstawiany jest zwykle przy pomocy wykresu liniowego. Taka linia opisana jest funkcją (modelem matematycznym) i zwana **funkcją gęstości rozkładu**. Odpowiednim rozkładem teoretycznym można zastąpić rozkład empiryczny.

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 rozkład o średniej 161.15 i odchyleniu standardowym równym 13.03 prezentują dane na temat liczby wykorzystanych darmowych minut (*przykład (8.1), plik: rozklad.pqs*).

9.1 CIĄGŁE ROZKŁADY PRAWDOPODOBIENSTWA

- **Rozkład normalny** (*ang. normal distribution*) nazywany również rozkładem Gaussa lub krzywą dzwonową, jest jednym z najważniejszych rozkładów w statystyce. Ma interesujące własności matematyczne i bardzo często występuje w naturze. Oznacza się go zwykle symbolem: $N(\mu, \sigma)$.

Funkcja gęstości jest zdefiniowana jako:

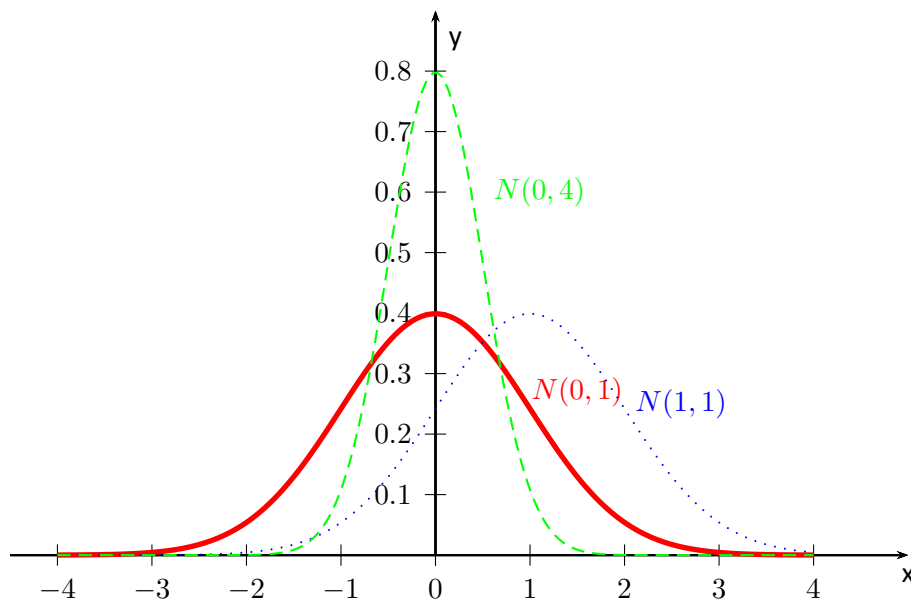
$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),$$

gdzie:

$$-\infty < x < +\infty,$$

μ – wartość oczekiwana populacji (której miarą jest średnia),

σ – odchylenie standardowe.



Rozkład normalny jest rozkładem symetrycznym względem prostej prostopadłej do osi odciętych i przechodzącej przez punkt wyznaczający średnią, modę oraz medianę.

Rozkład normalny o średniej $\mu = 0$ i $\sigma = 1$ ($N(0, 1)$), to tzw. **rozkład normalny standaryzowany**.

- **Rozkład t-Studenta** (*ang. t-Student distribution*) ma kształt podobny do standaryzowanego rozkładu normalnego lecz ma dłuższe "ogony". Wraz ze wzrostem liczby stopni swobody df (*ang. degree of freedom*) kształt rozkładu t-Studenta przybliża się do kształtu rozkładu normalnego.

Funkcja gęstości jest zdefiniowana jako:

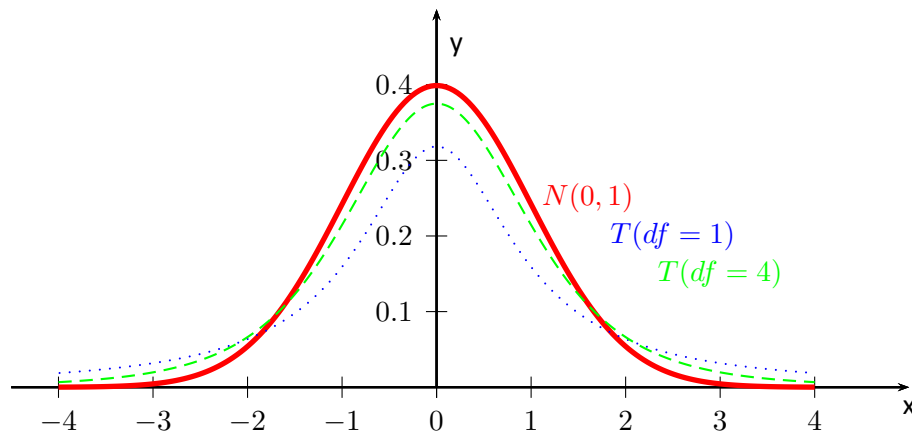
$$f(x, df) = \frac{\Gamma(\frac{df+1}{2})}{\Gamma(\frac{df}{2})\sqrt{df}\pi} \left(1 + \frac{x^2}{df}\right)^{-\frac{df+1}{2}},$$

gdzie:

$$-\infty < x < +\infty,$$

df – stopnie swobody (liczność próby pomniejszona o liczbę ograniczeń w określonych obliczeniach),

Γ to funkcja Gamma.



- **Rozkład χ^2** (*ang. Chi-square distribution*), jest rozkładem prawoskośnym o kształcie zależnym od liczby stopni swobody df . Wraz ze wzrostem liczby stopni swobody kształt rozkładu χ^2 przybliża się do kształtu rozkładu normalnego.

Funkcja gęstości jest zdefiniowana jako:

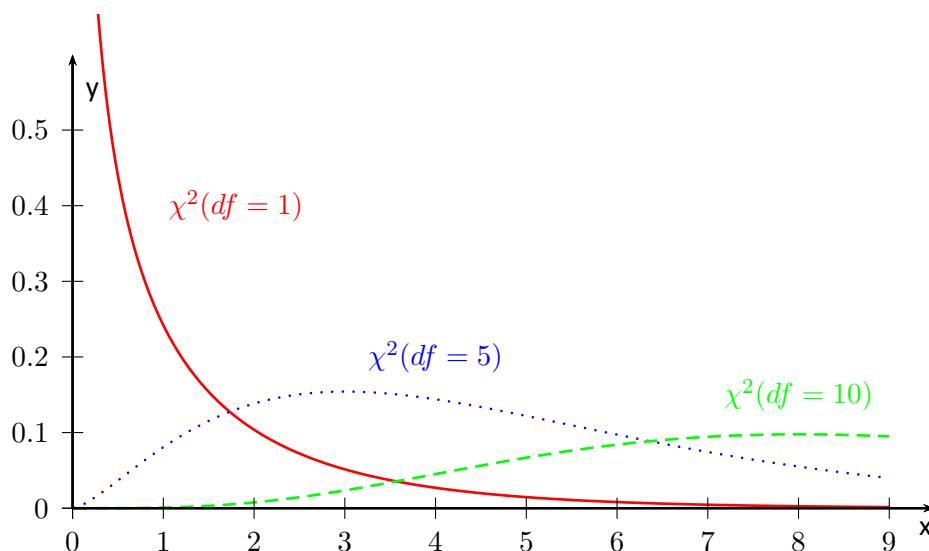
$$f(x, df) = \frac{1}{2^{\frac{df}{2}} \Gamma(\frac{df}{2})} x^{\frac{df}{2}-1} e^{-\frac{x}{2}},$$

gdzie:

$x > 0$,

df – stopnie swobody (liczność próby pomniejszona o liczbę ograniczeń w określonych obliczeniach),

Γ to funkcja Gamma.



- **Rozkład F Snedecora** (*ang. Fisher-Snedecor distribution*), jest rozkładem o dłuższym prawym "ogonie" i kształcie zależnym od liczby stopni swobody df_1 i df_2 .

Funkcja gęstości jest zdefiniowana jako:

$$F(x, df_1, df_2) = \frac{\sqrt{\frac{(df_1 x)^{df_1} d_2^{df_2}}{(df_1 x + df_2)^{df_1 + df_2}}}}{xB\left(\frac{df_1}{2}, \frac{df_2}{2}\right)},$$

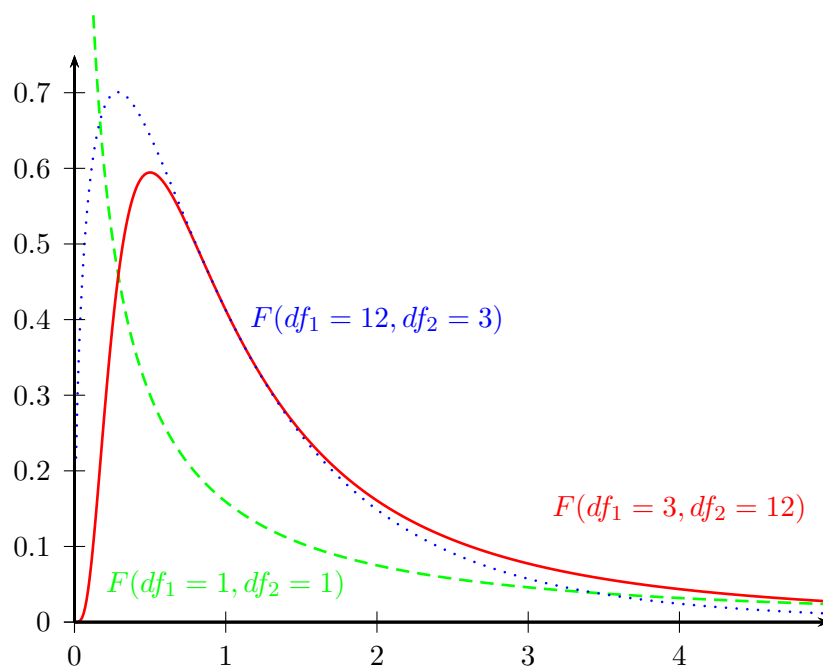
gdzie:

$x > 0$,

df_1, df_2 – 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

$F(df_1, df_2)$),

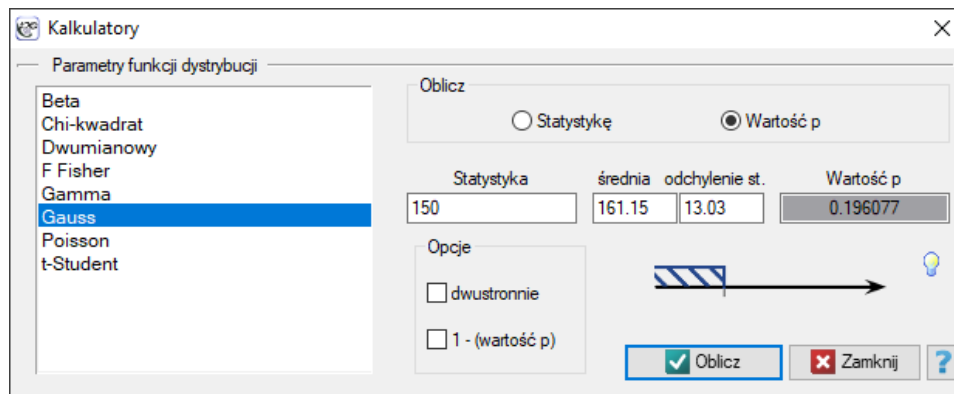
B to funkcja Beta.



9.2 KALKULATOR FUNKCJI DYSTRYBUCJI



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 jest możliwe wyliczanie wartości pola pod krzywą (Wartość p) danego rozkładu na podstawie Statystyki, jak też wyznaczanie wartości Statystyki na podstawie Wartości p . Okno Kalkulatora funkcji dystrybucji uruchamiamy poprzez wybranie menu Statystyka → Kalkulatory → Kalkulator funkcji dystrybucji



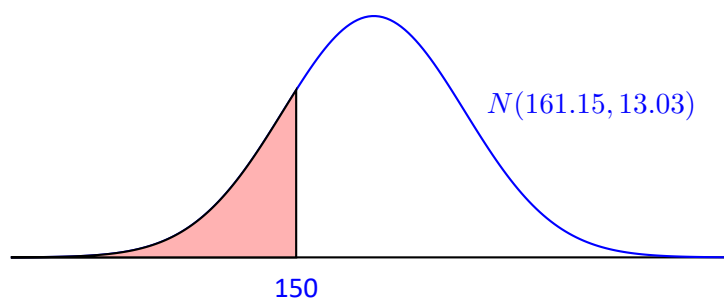
PRZYKŁAD 9.1. Kalkulator funkcji dystrybucji

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 $\bar{x} = 161.15min.$ 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 $[\bar{x} - sd, \bar{x} + sd] = [148.12min., 174.18min.]$,
4. ilość minut spoza przedziału $\bar{x} \pm sd$.

Uruchamiamy okno Kalkulatora funkcji dystrybucji, wybieramy rozkład Gaussa i wpisujemy średnią $\bar{x} = 161.15min.$ i odchylenie st. $sd = 13.03min.$ oraz zaznaczamy, że będziemy wyliczać Wartość p .

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.

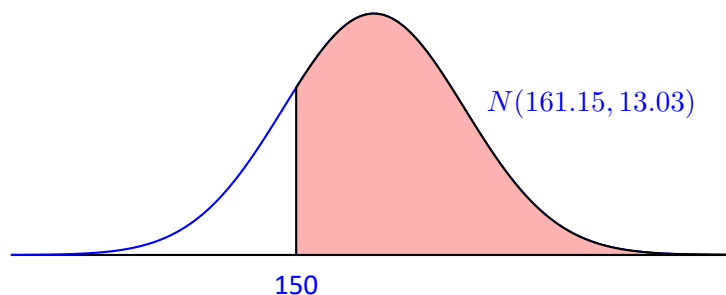


Uzyskana Wartość p wynosi 0.193961.

Uwaga!

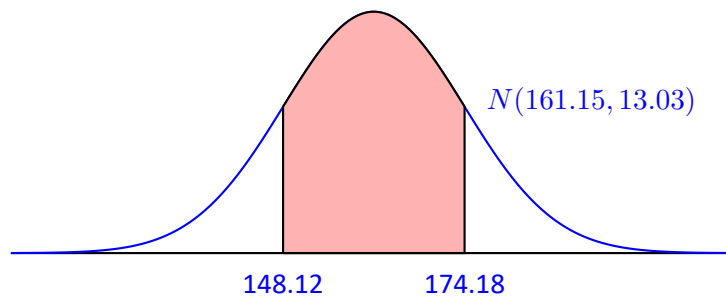
Podobne obliczenia możemy wykonać na podstawie rozkładu empirycznego. Wystarczy wówczas przy pomocy okna Tabele licznosci wyznaczyć procent klientów wykorzystujących 150 minut lub mniej (patrz przykład (8.1), plik: rozklad.pqs). W badanej 200 osobowej próbie klientów wykorzystujących 150 minut lub mniej jest 40, co stanowi 20% próby a zatem szukane prawdopodobieństwo wynosi $P = 0.2$.

2. By wyliczyć na podstawie rozkładu normalnego (Gausa) 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.



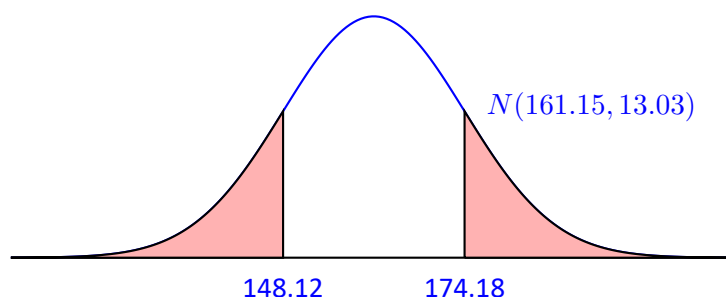
Uzyskana Wartość p wynosi 0.806039.

3. By wyliczyć na podstawie rozkładu normalnego (Gausa) jakie jest prawdopodobieństwo, że klient którego wylosujemy wykorzystał minuty z przedziału $[\bar{x} - sd, \bar{x} + sd] = [148.12min., 174.18min.]$, w polu Statystyka wpisujemy jedną z końcowych wartości przedziału, a następnie zaznaczamy opcję dwustronnie. Wybrane ustawienia potwierdzamy przyciskiem Oblicz.



Uzyskana Wartość p wynosi 0.682689.

4. By wyliczyć na podstawie rozkładu normalnego (Gausa) jakie jest prawdopodobieństwo, że klient którego wylosujemy wykorzystał minuty spoza przedziału $[\bar{x} - sd, \bar{x} + sd] = [148.12min., 174.18min.]$, 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.

10 TESTOWANIE HIPOTEZ

Proces uogólnienia 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.

10.0.1 ESTYMACJA PUNKTOWA I PRZEDZIAŁOWA

W praktyce najczęściej nie znamy **parametrów** (charakterystyk) dla całej populacji a dysponujemy jedynie próbką wylosowaną z tej populacji. **Estymatory punktowe** to charakterystyki otrzymane z próby losowej. Dokładność takiego estymatora określona jest przez jego **błąd standardowy**. Prawdziwe parametry populacji znajdują się w okolicy wyznaczonego estymatora punktowego. Np. parametr populacyjny **średnia arytmetyczna** μ znajduje się w okolicy estymatora z próby jakim jest \bar{x} .

Znając teoretyczny rozkłady estymatorów próby możemy z określonym z góry **poziomem ufności** ($1 - \alpha$) szacować wartości parametrów populacji na podstawie określenia estymatorów próby losowej. Zabieg taki nosi nazwę **estymacji przedziałowej**, a przedział określa się mianem **przedziału ufności** (*ang. confidence interval*), samo α zaś **poziomem istotności** (*ang. significance level*).

Najczęściej przyjmuje się poziom istotności równy 0.05, 0.01 lub 0.001.

10.0.2 WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Weryfikacja hipotez statystycznych wykonywana jest w kilku krokach:

krok 1 Postawienie hipotez, które będą weryfikowane za pomocą **testów** statystycznych.

Każdy test statystyczny podaje postać ogólną hipotezy zerowej – \mathcal{H}_0 (*ang. null hypothesis*) i 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 tak musi sformułować hipotezę, by była zgodna z rzeczywistością i wymogami **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) w badanej populacji mógłby być większy, wówczas hipoteza alternatywna powinna być dwustronna, tzn. nie zakładamy kierunku:

\mathcal{H}_1 : odsetek kobiet i mężczyzn prowadzących działalność gospodarczą w badanej populacji jest różny.

Może się zdarzyć (są to bardzo rzadkie przypadki), że mamy pewność, iż znamy kierunek w hipotezie alternatywnej. Wówczas można zastosować jednostronną hipotezę alternatywną.

krok 2 Sprawdzenie, 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 1!

Wybór testu statystycznego wiąże się przede wszystkim z wyborem odpowiedniej **skali pomiarowej** (interwałowej, porządkowej, nominalnej) jaką reprezentują analizowane dane, oraz z wyborem modelu badania (model zależny, model niezależny)

Pomiary danej cechy nazywamy **zależnymi (powiązаными)**, gdy są wykonane kilkakrotnie dla tych samych obiektów. Gdy pomiary danej cechy dokonane są na obiektach należących do różnych grup, wówczas mówimy że są to grupy **niezależnych (niepowiązanych)** pomiarów.

Przykłady badań w grupach zależnych:

Badanie masy ciała pacjentów przed i po terapii odchudzającej, badanie reakcji na bodziec tej samej grupy obiektów ale w dwóch różnych warunkach (np. w nocy i w dzień), badanie zgodności oceny zdolności kredytowej określonej przez 2 różne banki ale na tej samej grupie klientów itp.

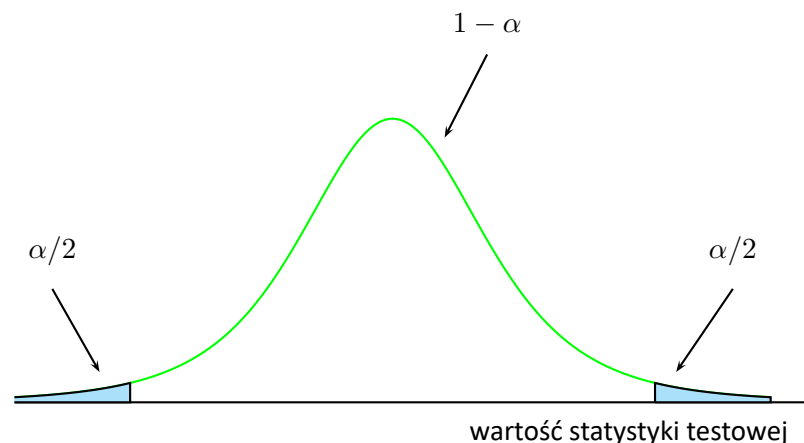
Przykłady badań w grupach niezależnych:

Badanie masy ciała w grupie pacjentów zdrowych i grupie pacjentów chorych, badanie efektywności nawożenia kilku różnych rodzajów nawozów, badanie wielkości PKB dla różnych krajów itp.

Uwaga 2!

Zawarty w oknie **Kreatora** graf ułatwia wybór odpowiedniego testu statystycznego.

Statystyka testowa wybranego testu wyliczana zgodnie z jej wzorem podlega odpowiedniemu dla niej rozkładowi teoretycznemu.



Program wylicza wartość **statystyki testowej**, oraz **wartość p** dla tej statystyki (czyli część pola pod krzywą, która odpowiada wartości statystyki testowej). Wartość p pozwala wybrać spośród hipotezy zerowej i alternatywnej tę bardziej prawdopodobną. Przy czym zawsze zakładamy prawdziwość hipotezy zerowej, a zebrane w danych dowody mają dostarczyć wystarczającej liczby argumentów przeciwko tej hipotezie:

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Zwykle wybiera się **poziom istotności** $\alpha = 0.05$, zgadzając się, że w 5% sytuacji odrzucimy hipotezę zerową gdy jest ona prawdziwa. W szczególnych przypadkach można wybrać inny poziom istotności np. 0.01 lub 0.001.

Uwaga!

Test statystyczny może nie być zgodny z rzeczywistością w dwóch przypadkach:

		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),

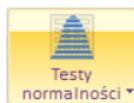
β = błąd drugiego rodzaju (prawdopodobieństwo przyjęcia hipotezy \mathcal{H}_0 , gdy jest ona fałszywa).

Moc testu to $1 - \beta$.

Wartości α i β są z sobą związane. Przyjętą praktyką jest ustalenie z góry poziomu istotności α i minimalizacja β poprzez zwiększanie rozmiaru próby.

krok 3 Opis wyników weryfikacji hipotez.

11 TESTY NORMALNOŚCI ROZKŁADU



11.0.1 Testy normalności jednowymiarowej

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.

Podstawowy warunek stosowania testów normalności rozkładu:

- pomiar na skali **interwałowej**.

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.

Wyznaczoną na podstawie statystyki testowej wybranego testu, wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!!!

Badanie normalności rozkładu można przeprowadzać dla zmiennych lub dla różnic wyznaczonych na podstawie dwóch zmiennych.

Test Kołmogorova-Smirnova (ang. Kolmogorov-Smirnov test for normality)

Test zaproponowany przez Kolmogorova (1933)[93] jest testem stosunkowo konserwatywnym (trudniej przy jego użyciu udowodnić brak normalności rozkładu). Opiera swoje działanie na wyznaczeniu odległości pomiędzy dystrybucją 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 możemy sprawdzić, czy rozkład zgodny jest z rozkładem zdefiniowanym przez podaną średnią i odchylenie standardowe.

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 (*ang. Lilliefors test for normality*)

Test zaproponowany przez Lillieforsa (1967, 1969, 1973)[102][103][104]. Jest on poprawką testu Kołmogorova-Smirnova, gdy nie znana jest wartość średnia (μ) i odchylenie standardowe (σ) dla populacji, z której pochodzi próba. Uznawany jest za nieco mniej konserwatywny od testu Kołmogorova-Smirnova.

Statystyka testowa D wyznaczana jest na podstawie tej samej formuły, z której korzysta test Kołmogorova-Smirnova, ale podlega rozkładowi Lillieforsa.

Test Shapiro-Wilka (*ang. Shapiro-Wilk test for normality*)

Zaproponowany przez Shapiro oraz Wilka (1965)[148] dla mało licznych grup, a następnie zaadoptowany dla grup liczniejszych (do 5000 obiektów) przez Roystona (1992)[141][142]. Test ten charakteryzuje stosunkowo wysoka moc, co ułatwia dowodzenie braku normalności rozkładu. Ideę działania testu przedstawia wykres [Q-Q plot](#).

Statystyka testowa Shapiro-Wilka ma postać:

$$W = \frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

gdzie:

a_i - współczynniki wyznaczone w oparciu o wartości oczekiwane dla statystyk uporządkowanych (ordered statistics), przypisanych wag oraz macierzy kowariancji,
 \bar{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 licznosciach $n \in \ll 4; 12 \gg$:

$$g(W) = -\ln(\gamma - \ln(1 - W)),$$

$$\gamma = 0.459n - 2.273,$$

$$\mu = -0.0006714n^3 + 0.025054n^2 - 0.39978n + 0.5440,$$

$$\sigma = \exp(-0.0020322n^3 + 0.062767n^2 - 0.77857n + 1.3822);$$

- dla prób dużych o licznosciach $n \in \ll 12; 5000 \gg$:

$$g(W) = \ln(1 - W),$$

$$\mu = 0.0038915u^3 - 0.083751u^2 - 0.31082u - 1.5851,$$

$$\sigma = \exp(0.0030302u^2 - 0.082676u - 0.4803),$$

$$u = \ln(n).$$

Test D'Agostino-Pearsona (*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 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]).

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).

Statystyka testowa ma postać:

$$Z_A = \delta \ln \left(\frac{Y}{\alpha} + \sqrt{\frac{Y^2}{\alpha^2} + 1} \right),$$

gdzie:

$$Y = \sqrt{(b_1)} \sqrt{\frac{(n+1)(n+3)}{6(n-2)}},$$

$$\sqrt{(b_1)} = \frac{m_3}{m_2^{(3/2)}},$$

$$m_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^k}{n},$$

$$\beta(\sqrt{(b_1)}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},$$

$$W^2 = -1 + \sqrt{2(\beta(\sqrt{(b_1)}) - 1)},$$

$$\delta = \frac{1}{\sqrt{\ln W}},$$

$$\alpha = \sqrt{\frac{2}{W^2-1}}.$$

Statystyka Z ma asymptotycznie (dla dużych licznosci) 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.

Statystyka testowa ma postać:

$$Z_K = \frac{\left(1 - \frac{2}{9H}\right) - \left(\frac{1 - \frac{2}{9H}}{1 + x \sqrt{\frac{2}{H^2-4}}}\right)^{1/3}}{\sqrt{\frac{2}{9H}}},$$

gdzie:

$$E(b_2) = \frac{3(n-1)}{n+1},$$

$$b_2 = \frac{m_4}{m_2^2},$$

$$\text{var}(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)},$$

$$x = \frac{b_2 - E(b_2)}{\sqrt{\text{var}(b_2)}},$$

$$\sqrt{\beta(b_2)} = \frac{6(n^2-5n+2)}{(n+7)(n+9)} \sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-5)'}}$$

$$H = 6 + \frac{8}{\sqrt{\beta(b_2)}} \left(\frac{2}{\sqrt{\beta(b_2)}} + \sqrt{1 + \frac{4}{\beta(b_2)}} \right).$$

Statystyka Z ma asymptotycznie (dla dużych liczności) **rozkład normalny**.

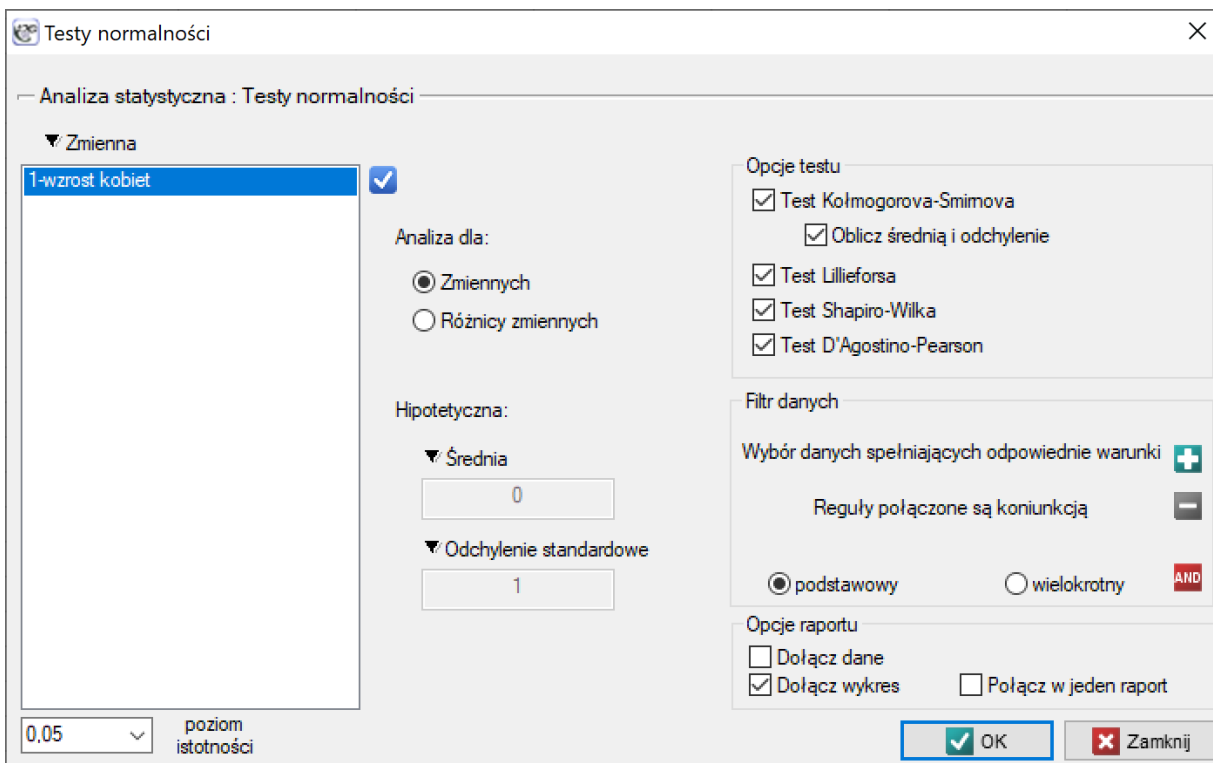
Wykres Kwantyl-Kwantyl (ang. Q-Q plot, Quantile-Quantile plot)

Wykres typu Kwantyl-Kwantyl wykorzystywany jest do przedstawienia zgodności dwóch rozkładów. W przypadku badania zgodności z rozkładem normalny, 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 schemat:

- dane rozłożone na linii, lecz kilka punktów mocno odbiega od linii
 - występują wartości odstające w danych
- punkty po lewej stronie wykresu znajdują się powyżej linii, a po prawej poniżej linii
 - rozkład charakteryzuje większa obecność wartości odległych od średniej niż jest w rozkładzie normalnym (ujemna kurtoza)
- punkty po lewej stronie wykresu znajdują się poniżej linii, a po prawej powyżej linii
 - rozkład charakteryzuje mniejsza obecność wartości odległych od średniej niż jest w rozkładzie normalnym (dodatnia kurtoza)
- punkty po lewej i po prawej stronie wykresu znajdują się powyżej linii
 - rozkład prawostronnie skośny (dodatnia skośność);
- punkty po lewej i po prawej stronie wykresu znajdują się poniżej linii
 - rozkład lewostronnie skośny (ujemna skośność).

Okno z ustawieniami opcji testów normalności wywołujemy poprzez menu Statystyka→Testy normalności→Normalność jednowymiarowa lub poprzez [Kreator](#).



PRZYKŁAD 11.1. (plik Gauss.pqs)

- **Wzrost kobiet**

Założmy, że wzrost kobiet jest taką cechą, dla której wartość przeciętna wynosi 168cm. 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.

Hipotezy:

\mathcal{H}_0 : rozkład wzrostu kobiet w badanej populacji jest rozkładem normalnym,

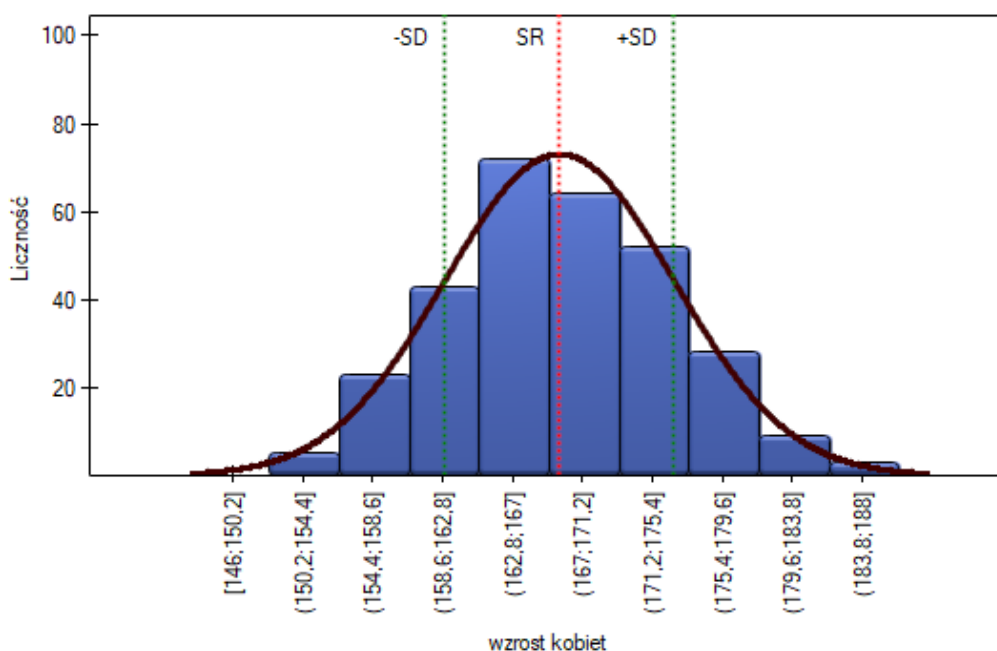
\mathcal{H}_1 : rozkład wzrostu kobiet w badanej populacji jest różny od rozkładu normalnego.

Ponieważ nie znamy średniej ani odchylenia standardowego dla wzrostu kobiet, a jedynie mamy przypuszczenia co do tych wielkości, będą one wyznaczone z próby.

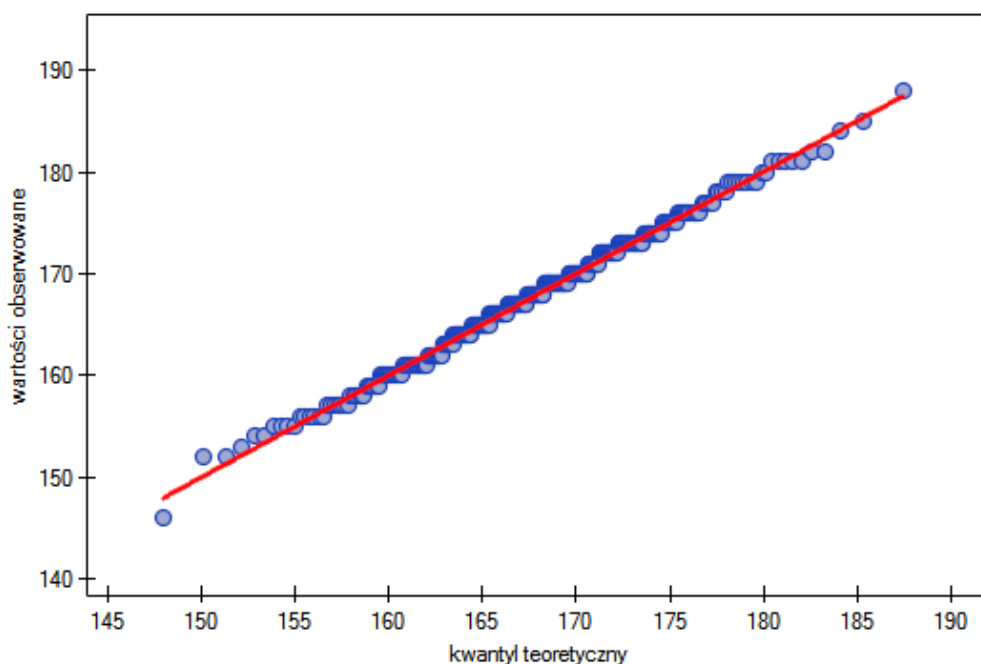
Testy normalności	
Czas analizy	0,15 sek.
Analizowane zmienne	wzrost kobiet
Poziom istotności	0,05
Liczność - grupy	300
Średnia grupy	167,726666667
Odchylenie standardowe grupy	6,886462574
Test Kołmogorova-Smirnova	
Statystyka D	0,042331501
Stopnie swobody	300
Wartość p	0,639671197
Test Lillieforsa	
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 wyznaczone testy wskazują na brak odstępstwa od rozkładu normalnego, ponieważ ich wartości p znajdują się powyżej standardowego poziomu istotności $\alpha = 0.05$. 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.



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.

- **Dochód**

Założmy, 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), ale będą osoby zarabiające bardzo dużo (prezisi firm) których jest stosunkowo niewielu. By sprawdzić czy dochód osób w badanym kraju ma rozkład normalny zebrano

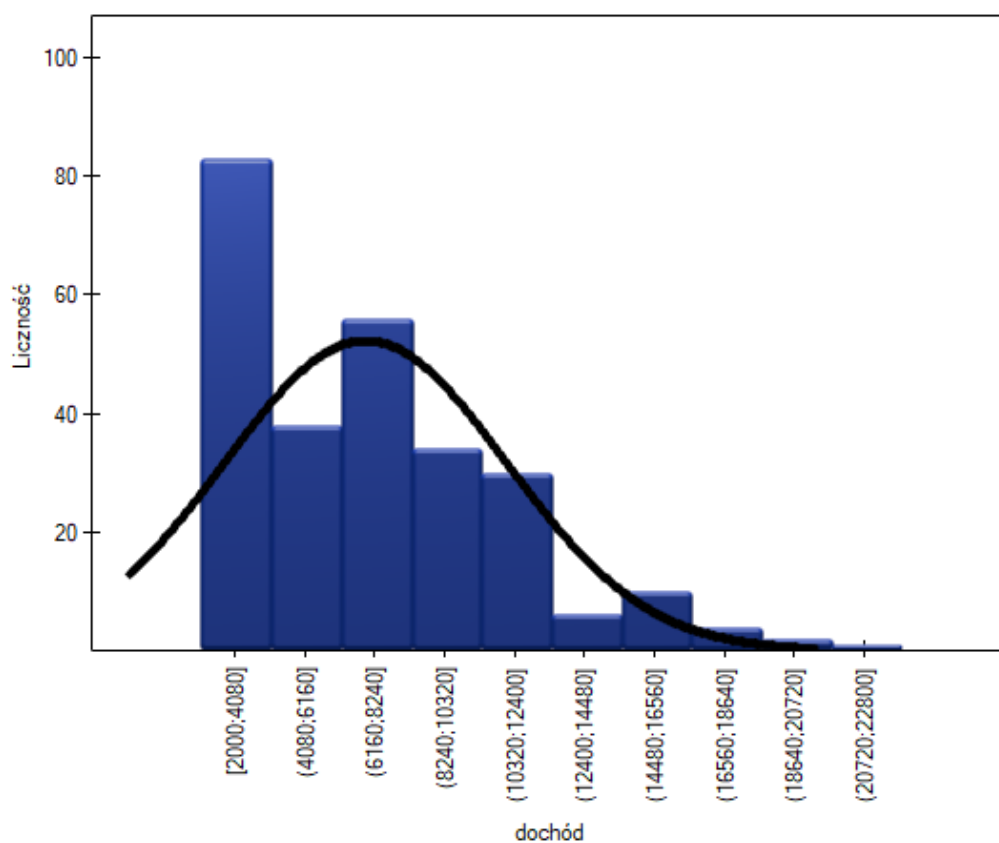
informację o dochodach 264 losowo wybranych osób.

Hipotezy:

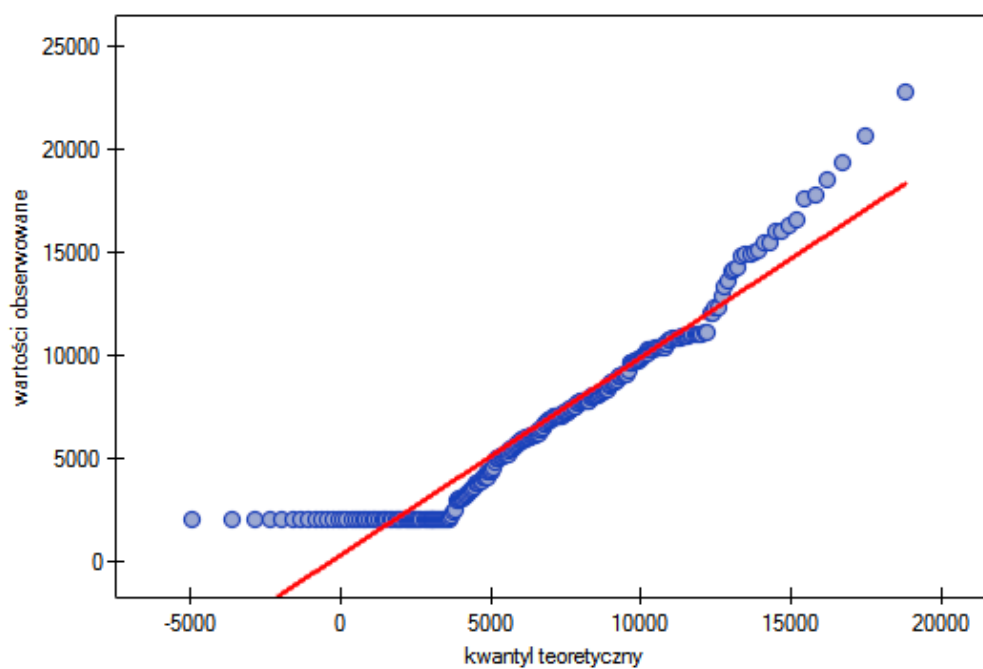
- \mathcal{H}_0 : rozkład dochodów osób w badanej populacji jest rozkładem normalnym,
 \mathcal{H}_1 : rozkład dochodów osób w badanej populacji jest różny od rozkładu normalnego.

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

Rozkład nie jest rozkładem normalnym, o czym świadczą wszystkie wyniki testów badających normalność rozkładu ($p < \alpha$). Dodatnia i istotna statystycznie ($p < \alpha$) wartość skośności świadczy o zbyt długim prawym ogonie funkcji. Rozkład funkcji jest również bardziej smukły od rozkładu normalnego, ale nie jest to istotna statystycznie różnica (test kurtozy).



Na wykresie kwartył-kwartył odstępstwo od rozkładu normalnego jest obrazowane poprzez skośność prawostronną, czyli położenie znacznie powyżej linii początkowych i końcowych punktów wykresu.



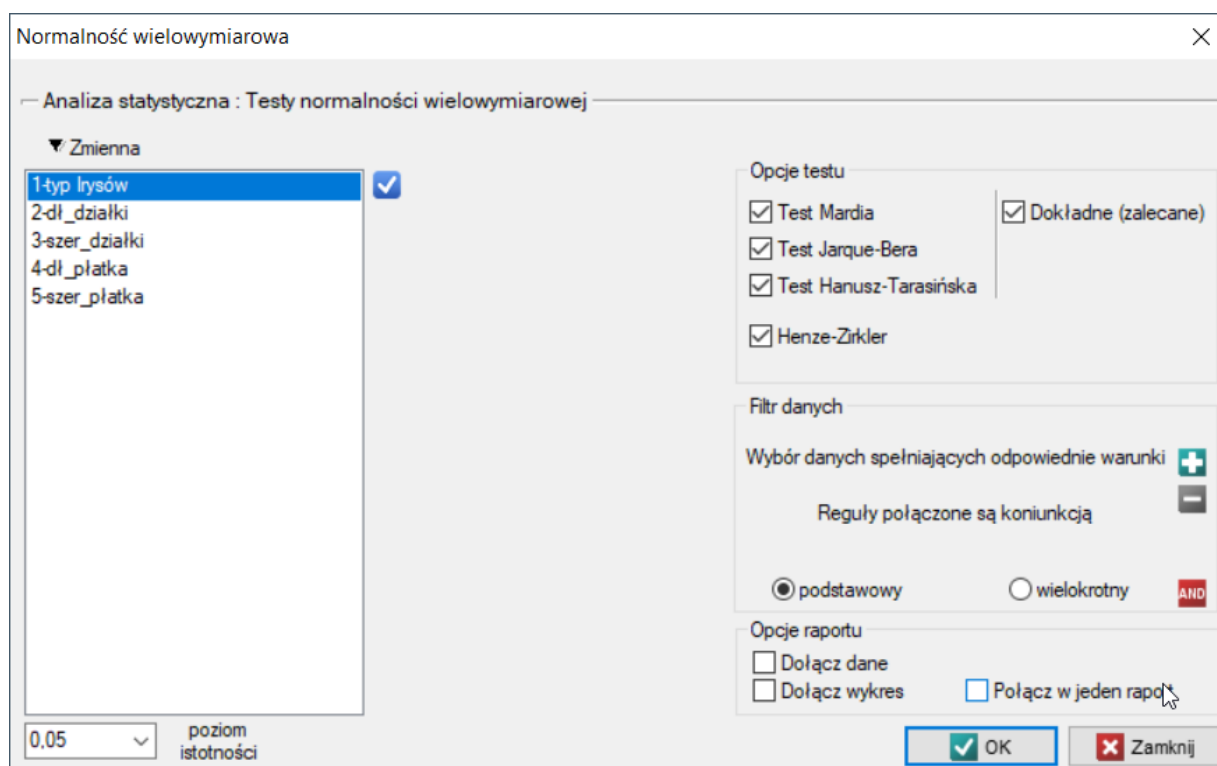
W rezultacie zebrane dane nie świadczą o zgodności rozkładu dochodów z rozkładem normalnym.

11.0.2 Testy normalności wielowymiarowej

Wiele metod analizy wielowymiarowej, w tym MANOVA, testy Hotellinga czy też modele regresji opierają się na założeniu normalności wielowymiarowej. Jeśli zbiór zmiennych charakteryzuje wielowymiarowy rozkład normalny, to można założyć, że 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.

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.

Okno z ustawieniami opcji testu wielowymiarowej normalności rozkładu wywołujemy poprzez menu Statystyka → Testy normalności → Normalność wielowymiarowa.



Test Mardia dla wielowymiarowej normalności rozkładu (ang. Mardia's test for multivariate normality)

Test zaproponowany przez Mardia w roku 1970 [114] i zmodyfikowany w roku 1974 [115] bada normalność rozkładu analizując oddzielnie rozmiar wielowymiarowej skośności i wielowymiarowej kurtozy. Jarque i Bera [86] zaproponowali złączenie tych dwóch miar Mardia w jeden test. Podobny sposób łączenia w jeden test informacji o skośności i kurtozie oferuje metoda Hanusz i Tarasińskiej [76].

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

$$m_{ij} = (X_i - \bar{X})^T S^{-1} (X_j - \bar{X}),$$

$$S = \frac{1}{n} \sum_{j=1}^n (X_i - \bar{X})(X_i - \bar{X})^T,$$

\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 $(n/(n-1))^3$ i kurtoza przez $(n/(n-1))^2$.

Hipotezy:

\mathcal{H}_0 : rozkład danych w populacji = wielowymiarowy rozkład normalny,

\mathcal{H}_1 : rozkład danych w populacji \neq wielowymiarowy rozkład normalny,

- **Mardia test skośności:** Gdy próba pochodzi z populacji o wielowymiarowym rozkładzie normalnym (hipoteza zerowa), to statystyka testowa ma postać (Mardia, 1970):

$$\chi^2(M) = \frac{n}{6} skew$$

lub z poprawką dokładnych momentów dla grup o mniejszych licznosciach (<20) (Mardia, 1974):

$$\chi_c^2(M) = \frac{n(n+1)(n+3)(k+1)}{6n(n+1)(k+1)-6} skew$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) rozkład χ^2 z $df = f = \frac{k(k+1)(k+2)}{6}$ stopniami swobody.

- **Mardia test kurtozy:** Gdy próba pochodzi z populacji o wielowymiarowym rozkładzie normalnym (hipoteza zerowa), to 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 licznosci) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p dla obu testów tzn. testu skośności i testu kurtozy porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha$ dla przynajmniej jednego testu \implies odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha$ dla obu zastosowanych testów \implies nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Test Jarque-Bera dla wielowymiarowej normalności rozkładu (ang. Jarque-Bera test for multivariate normality)

Test Jarque i Bera (1987) [86] 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)

$$\chi_c^2(JB) = \chi_c^2(M) + (Z_c(M))^2$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) rozkład χ^2 z $df = f + 1$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Test Hanusz-Tarasińska dla wielowymiarowej normalności rozkładu (ang. Hanusz-Tarasinska test for multivariate normality)

Test Zofii Hanusz i Joanny Tarasińskiej (2014) [76] bazuje na statystyce skośności i kurtozy testu Mardia. Statystyka testowa ma postać:

$$t_c(HT) = \frac{Z_c(M)}{\sqrt{\frac{\chi_c^2(M)}{f}}}$$

Statystyka testowa ma rozkład *t-Studenta* z $df = f$ stopniami swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Test Henze-Zirklera dla wielowymiarowej normalności rozkładu (ang. Henze-Zirkler test for multivariate normality)

Henze i Zirklera (1990) [78] zaproponowali test badający wielowymiarową normalność rozkładu rozszerzając pracę Baringhausa i Henzego nad empiryczną funkcją charakterystyczną [54]. W literaturze jest to test uznawany za jeden z najsilniejszych testów poświęconych wielowymiarowemu rozkładowi normalnemu (Thode 2002) [162]. Statystyka testowa ma postać:

$$Z(HZ)_\beta = n(4I_E + D_{n,\beta}I_{E^c})$$

$$I_E \text{ oraz } I_{E^c} \text{ to funkcje indyktorowe zależne od osobliwości macierzy kowariancji,}$$

$$D_{n,\beta} = \frac{1}{n^2} \sum \exp\left(\frac{-\beta^2 \|Y_j - Y_k\|^2}{2}\right) + (1 + 2\beta^2)^{-p/2} - 2(1 + \beta^2)^{-p/2} \sum \exp\left(\frac{-\beta^2 \|Y_j\|^2}{2(1+\beta^2)}\right)$$

$$Y_i = S^{1/2}(X_i - \bar{X})$$

$$\beta_* = 2^{-1/2} \left(\frac{n(2k+1)}{4}\right)^{1/(k+4)} \text{ - optymalna wartość parametru } \beta$$

Statystyka $Z(HZ)_\beta$ ma asymptotycznie (dla dużych licznosci) 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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

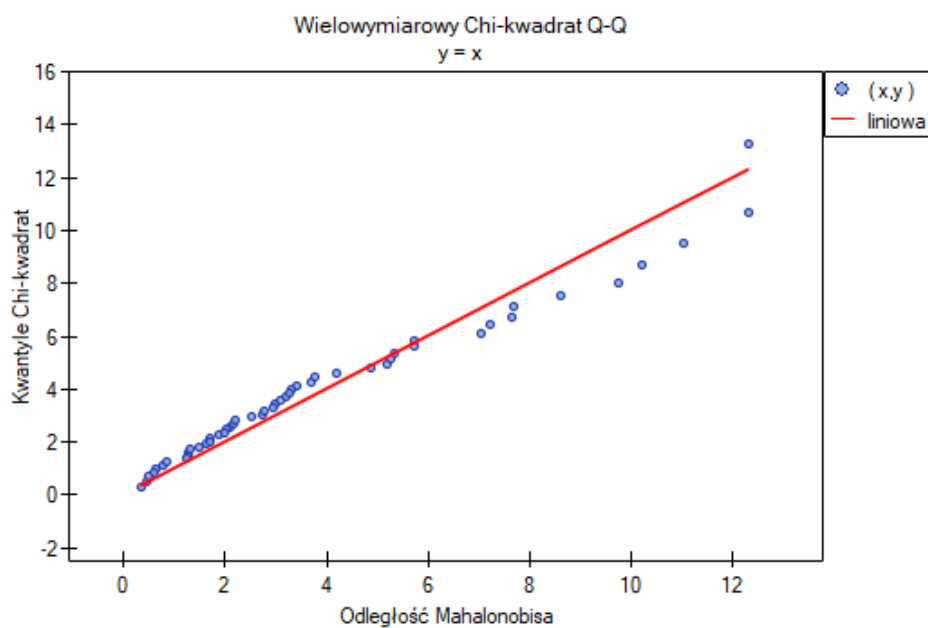
PRZYKŁAD 11.2. (plik Irysy.pqs)

Badamy normalność rozkładu dla klasycznego zestawu danych R.A. Fishera 1936 [58]. Plik znajduje się w pomocy dołączonej do programu i zawiera pomiary długości i szerokości płatków i działek kielicha dla 3 odmian kwiatu irysa. Analiza zostanie przeprowadzona oddzielnie dla każdej odmiany.

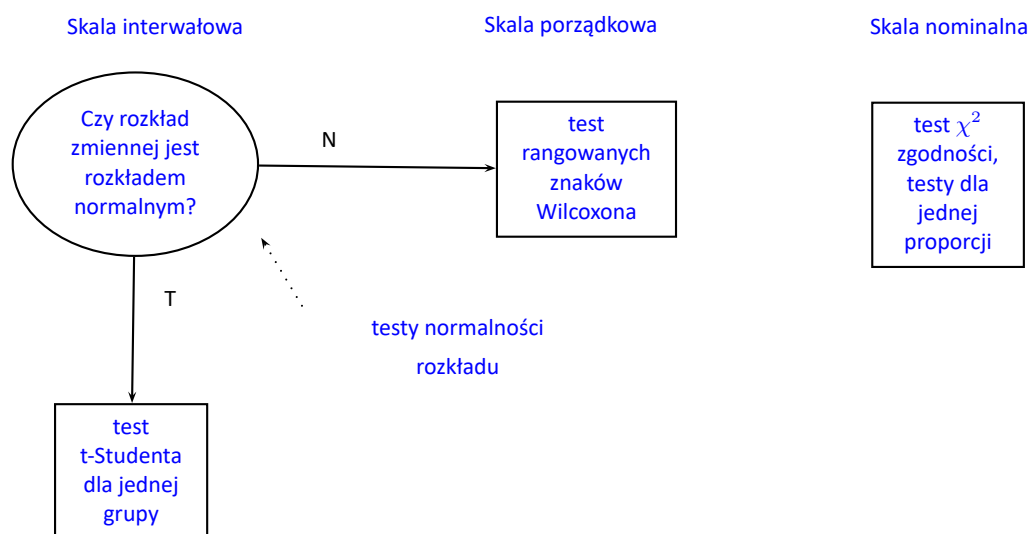
W oknie analizy zaznaczamy wszystkie testy oraz wykres, a w celu powtórzenia analizy dla każdej odmiany irysa ustawiamy filtr wielokrotny. Wszystkie wyniki zwrócimy do tego samego arkusza, dlatego zaznaczamy opcję Połącz w jeden raport.

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 swobody	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 swobody	21	21	21
Wartość p	0,05	0,1595	0,1109
Test Hanusz-Tarasińska			
Statystyka t	1,8578	-0,0967	0,6546
Stopnie swobody	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 potwierdzają normalność rozkładu dla odmiany versicolor i virginica. W przypadku odmiany setosa wyniki testów są na pograniczu istotności statystycznej, przy czym test Mardia dla Kurtozy i test Henze-Zirkler wskazują na odstępstwa od wielowymiarowego rozkładu normalnego. Możemy obserwować takie odstępstwa również na pierwszym wykresie, gdzie wraz narastaniem odległości Mahalanobisa punkty znajdują się coraz dalej od prostej.



12 PORÓWNANIE - 1 GRUPA



12.1 TESTY PARAMETRYCZNE

12.1.1 Test t -Studenta dla pojedynczej próby

Test t -Studenta dla pojedynczej próby (*ang. single-sample t test*) służy do weryfikacji hipotezy, że badana próba o średniej \bar{x} pochodzi z populacji dla której średnia μ to zadana wartość.

Podstawowe warunki stosowania:

- pomiar na skali [interwałowej](#),
- [normalność rozkładu](#) badanej cechy.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: \mu = \mu_0, \\ \mathcal{H}_1 &: \mu \neq \mu_0,\end{aligned}$$

gdzie:

μ – średnia cechy w populacji reprezentowanej przez próbę,
 μ_0 – zadana wartość.

Statystyka testowa ma postać:

$$t = \frac{\bar{x} - \mu_0}{sd} \sqrt{n},$$

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 wartości \$p\$](#) porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Uwaga!

Gdy próba jest duża i znane jest odchylenie standardowe z populacji wówczas statystykę testową można wyznaczyć z wzoru:

$$t = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}.$$

Tak wyznaczona statystyka testowa ma [rozkład normalny](#). Przy $n \rightarrow \infty$ 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.

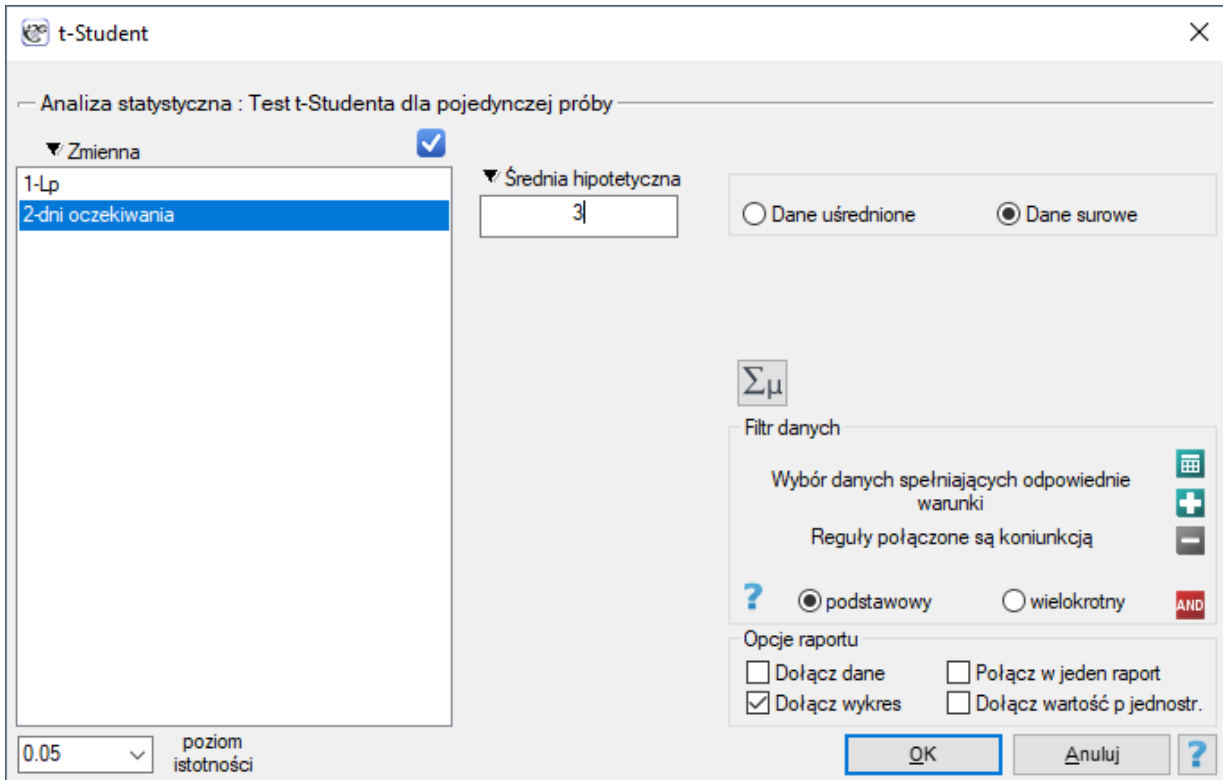
Standaryzowana wielkość efektu

Współczynnik **d-Cohena** określa jak dużą częśćią występującej zmienności jest różnica między średnimi.

$$d = \left| \frac{\bar{x} - \mu_0}{sd} \right|$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [45] wskaźnikami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Okno z ustawieniami opcji testu t -Studenta dla pojedynczej próby wywołujemy poprzez menu Statystyka → Testy parametryczne → t -Student lub poprzez [Kreator](#).


Uwaga!

Obliczenia mogą bazować na danych w postaci **surowych** rekordów lub danych uśrednionych tzn. średniej arytmetycznej, odchyleniu standardowym i liczności próby.

PRZYKŁAD 12.1. (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, 4, 4, 4, 4, 4, 4, 5, 5, 6, 6, 6, 7, 7).

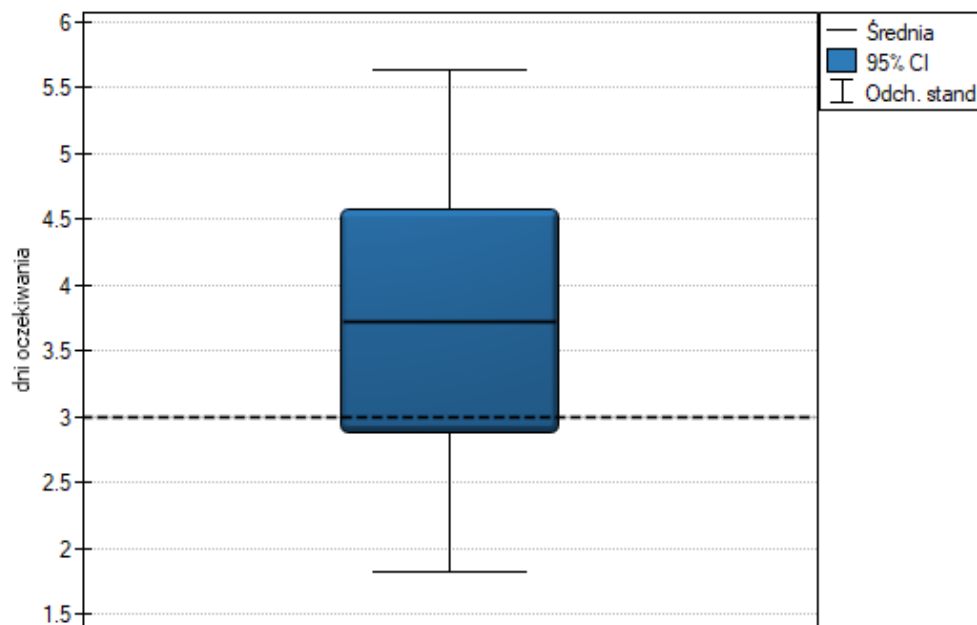
Liczba dni oczekiwania na przesyłkę w badanej populacji spełnia założenie **normalności** rozkładu.

Hipotezy:

- \mathcal{H}_0 : średnia liczba dni oczekiwania na dostarczenie przesyłki przez analizowaną firmę kurierską wynosi 3,
- \mathcal{H}_1 : średnia liczba dni oczekiwania na dostarczenie przesyłki przez analizowaną firmę kurierską jest różna od 3.

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

Podsumowanie	
Grupa	dni oczekiwania
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



Porównując wartość $p = 0.088074$ testu t -Studenta z poziomem istotności $\alpha = 0.05$ stwierdzamy, że 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 $\bar{x} = 3.727$ a odchylenie standardowe $sd = 1.907$.

12.1.2 Test chi-kwadrat wariancji pojedynczej próby

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. Przy czym hipotezy mogą dotyczyć zarówno wariancji jak i równoważnie odchylenia standardowego.

Podstawowe warunki stosowania:

- pomiar na skali **interwałowej**,
- **normalność rozkładu** badanej cechy.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: \sigma = \sigma_0, \\ \mathcal{H}_1 &: \sigma \neq \sigma_0,\end{aligned}$$

gdzie:

σ - odchylenie cechy w populacji reprezentowanej przez próbę,
 σ_0 – zadana wartość.

Statystyka testowa ma postać:

$$t = \frac{(n-1)sd^2}{\sigma_0^2},$$

gdzie:

sd – odchylenie standardowe z próby,
 n – liczebność próby.

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 :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Przy czym, jeśli wartość odchylenia standardowego jest mniejsza niż wartość zadana, to wartość p wyliczana jest jako podwojoną wartość pola pod krzywą rozkładu chi-kwadrat na lewo od odpowiedniej wartości krytycznej, a jeśli jest większa niż wartość zadana, jest to podwojoną wartość odpowiedniego pola po prawej stronie.

Okno z ustawieniami opcji testu chi-kwadrat wariancji pojedynczej próby wywołujemy poprzez menu Statystyka→Testy parametryczne→Chi-kwadrat dla wariancji.

Chi-kwadrat dla wariancji

Analiza statystyczna : Test Chi-kwadrat wariancji pojedynczej próby

Zmienna

1-objętość syropu

Odchylenie hipotetyczne

Dane uśrednione Dane surowe

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

0,05 poziom istotności

Uwaga!

Obliczenia mogą bazować na danych w postaci **surowych** rekordów lub danych uśrednionych tzn. odchyleniu standardowym i liczności próby.

PRZYKŁAD 12.2. (dozownik.pqs)

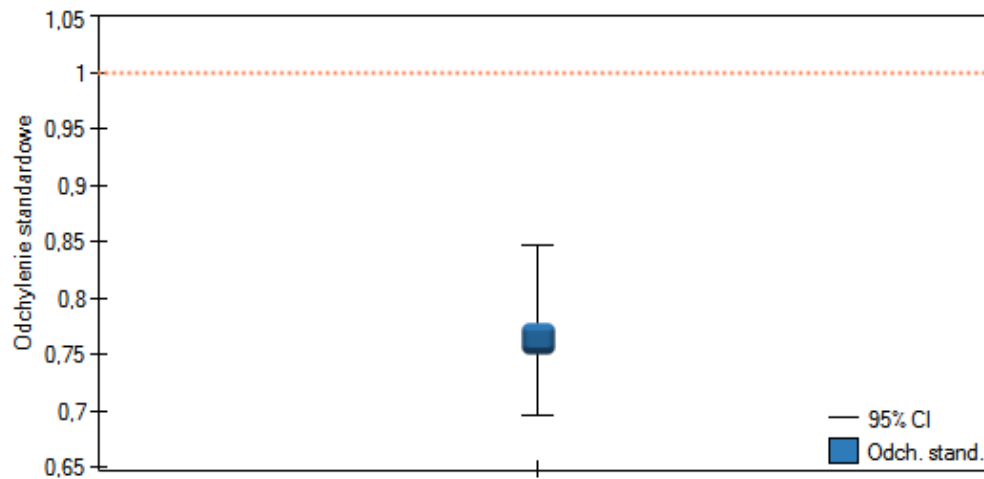
Przed przystąpieniem do produkcji kolejnej partii pewnego syropu przeciwkaszlowego dokonano kontrolnych pomiarów objętości syropu nalewanego do butelek. Z dokumentacji technicznej urządzenia dozującego wynika, że dopuszczalny rozrzut objętości syropu mierzony za pomocą odchylenia standardowego wynosi 1ml. Należy sprawdzić, czy testowane urządzenie działa prawidłowo.

Sprawdzono (testem Lillieforsa) **normalność** rozkładu objętości syropu nalewanego do butelek używając wynik zgodny z tym rozkładem. Analiza dotycząca odchylenia standardowego może być więc przeprowadzona testem chi-kwadrat dla wariancji.

Hipotezy:

- \mathcal{H}_0 : odchylenie standardowe objętości syropu nalewanego przez urządzenie dozujące wynosi 1ml,
- \mathcal{H}_1 : odchylenie standardowe objętości syropu nalewanego przez urządzenie dozujące jest różne od 1ml.

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



Porównując wartość $p = 0.000001$ testu χ^2 z poziomem istotności $\alpha = 0.05$ stwierdzamy, że rozrzut urządzenia dozującego jest inny niż 1ml. 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.

12.2 TESTY NIEPARAMETRYCZNE

12.2.1 Test Wilcoxona (rangowanych znaków)

Test Wilcoxona rangowanych znaków (*ang. Wilcoxon signed-ranks test*) znany również pod nazwą testu Wilcoxona dla pojedynczej próby, Wilcoxon (1945, 1949)[170]. Test ten służy do weryfikacji hipotezy, że badana próba pochodzi z populacji, dla której mediana (θ) to znana wartość.

Podstawowe warunki stosowania:

- pomiar na [skali porządkowej](#) lub [interwałowej](#).

Hipotezy dotyczą równości sumy rang dodatnich i ujemnych lub są upraszczane do median:

$$\begin{aligned}\mathcal{H}_0 &: \theta = \theta_0, \\ \mathcal{H}_1 &: \theta \neq \theta_0.\end{aligned}$$

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 .

Porównujemy [wartość \$p\$](#) z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Uwaga!

W zależności od wielkości próby statystyka testowa przyjmuje inną postać:

- dla małej liczności próby

$$T = \min \left(\sum R_-, \sum R_+ \right),$$

gdzie:

$\sum R_+$ i $\sum R_-$ to odpowiednio: suma [rang](#) dodatnich i suma rang ujemnych.

Statystyka ta podlega rozkładowi Wilcoxona

- dla próby o dużej liczności

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}},$$

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 $(\sum t^3 - \sum t) / 48 = 0$).

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład [normalny](#).

Poprawka na ciągłość testu Wilcoxona (Marascuilo and McSweeney (1977)[113])

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:

$$Z = \frac{\left| T - \frac{n(n+1)}{4} \right| - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}$$

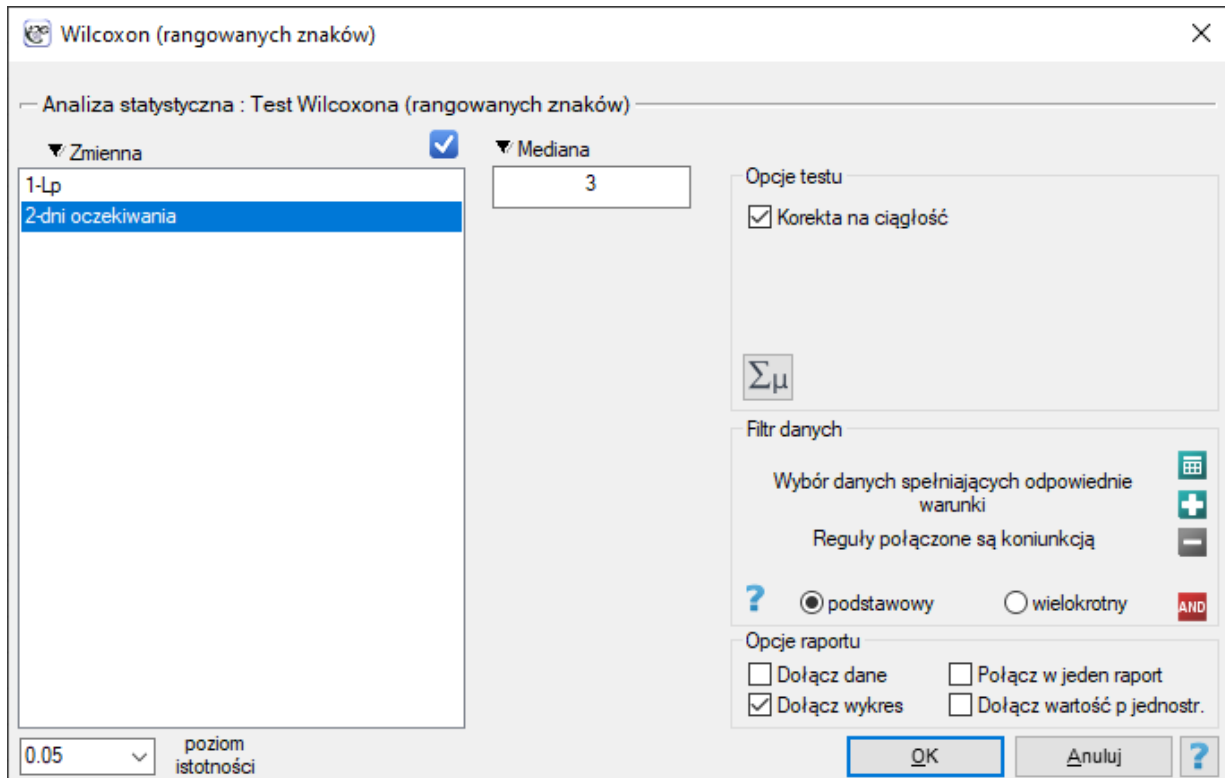
Standaryzowana wielkość efektu

Rozkład statystyki testu Wilcoxona jest aproksymowany przez rozkłady normalny, który można przekształcić na wielkość efektu $r = |Z/n|$ [65] by następnie uzyskać wartość **d-Cohena** zgodnie ze standardową konwersją stosowaną przy meta-analizach:

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [45] wskaźnikami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Okno z ustawieniami opcji testu Wilcoxona (rangowanych znaków) wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Wilcoxon (rangowanych znaków) lub poprzez [Kreator](#).



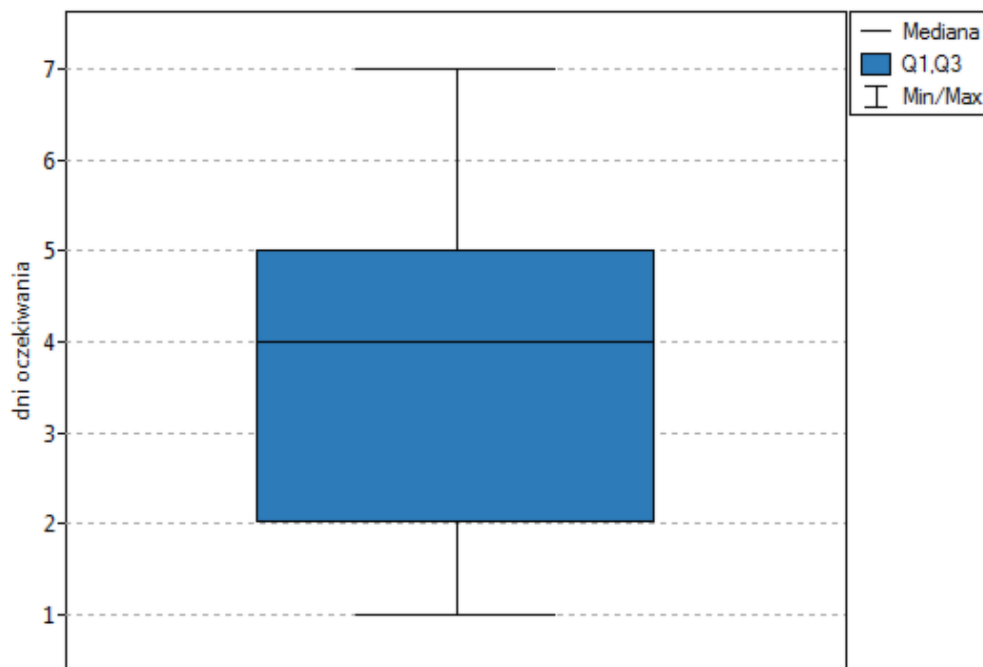
PRZYKŁAD 12.1 c.d. (plik kurier.pqs)

Hipotezy:

- \mathcal{H}_0 : mediana liczby dni oczekiwania na dostarczenie przesyłki przez analizowaną firmę kurierską wynosi 3
- \mathcal{H}_1 : mediana liczby dni oczekiwania na dostarczenie przesyłki przez analizowaną firmę kurierską jest różna od 3

Test Wilcozona (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 oczekiwania
Liczność	22
Mediana	4
Dolny kwartyl	2
Górny kwartyl	5



Porównując wartość $p = 0.123212$ testu Wilcozona opartego o statystykę T z poziomem istotności $\alpha = 0.05$ stwierdzamy, że nie mamy podstaw by odrzucić hipotezę zerową mówiącą, że zwykle liczba 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 Wilcozona opartego o statystykę Z lub Z z poprawką na ciągłość.

12.2.2 Test chi-kwadrat zgodności

Test χ^2 zgodności (dobroci dopasowania) (*ang. Chi-square goodness-of-fit test*) nazywany jest również testem χ^2 dla pojedynczej próby, przeznaczony jest do testowania zgodności wartości obserwowanych dla r ($r \geq 2$) kategorii X_1, X_2, \dots, 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, \dots, 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 $E_i = np_i$. Utworzona tabela może przyjąć jedną z poniższych postaci:

Kategorie X_i	O_i	E_i	Kategorie X_i	O_i	p_i
X_1	O_1	E_1	X_1	O_1	p_1
X_2	O_2	E_2	X_2	O_2	p_2
...
X_r	O_r	E_r	X_r	O_r	p_r

Podstawowe warunki stosowania:

- pomiar na [skali nominalnej](#) - ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- duże licznosci oczekiwane według [interpretacji Cochran](#) (1952)[40],
- suma licznosci obserwowanych powinna być taka sama jak suma licznosci oczekiwanych, a suma wszystkich prawdopodobienstw p_i powinna wynosić 1.

Hipotezy:

$$\mathcal{H}_0 : O_i = E_i \text{ dla wszystkich kategorii,}$$

$$\mathcal{H}_1 : O_i \neq E_i \text{ dla przynajmniej jednej kategorii.}$$

Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i}.$$

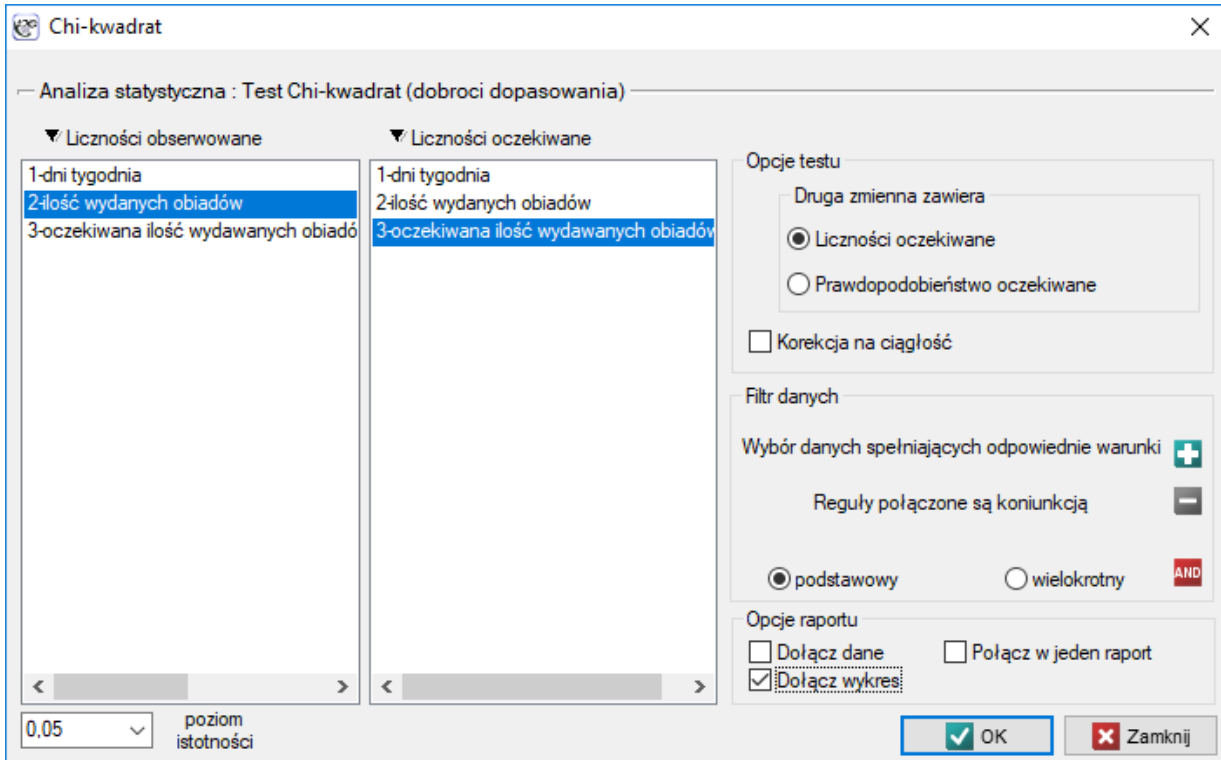
Statystyka ta ma asymptotycznie (dla dużych licznosci oczekiwanych) [rozkład \$\chi^2\$](#) z liczbą stopni swobody wyznaczaną według wzoru: $df = (r - 1)$.

Wyznaczoną na podstawie wartości [statystyki](#) i rozkładu χ^2 [wartość \$p\$](#) porównujemy z poziomem istotności α :

$$\text{jeżeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1,$$

$$\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0.$$

Okno z ustawieniami opcji testu Chi-kwadrat zgodności wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat lub poprzez [Kreator](#).



Chi-kwadrat

Analiza statystyczna : Test Chi-kwadrat (dobroci dopasowania)

Licznosci obserwowane

Licznosci oczekiwane

Opcje testu

Druga zmienna zawiera

Licznosci oczekiwane

Prawdopodobienstwo oczekiwane

Korekcja na ciaglosc

Filtr danych

Wybor danych speiniajacych odpowiednie warunki

Reguly polaczone sa koniunkcja

podstawowy wielokrotny

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

0,05 poziom istotności

PRZYKŁAD 12.3. (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 $E_i = 150 \cdot \frac{1}{5} = 30$.

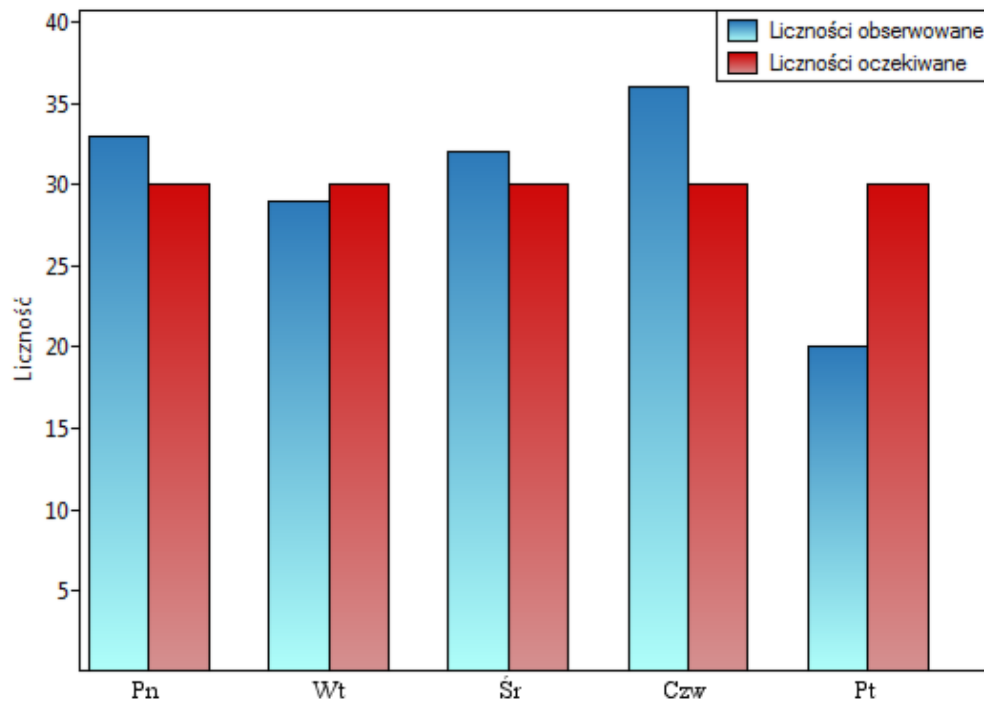
dni tygodnia	ilość wydanych obiadów	oczekiwana ilość wydawanych obiadów
poniedziałek	33	30
wtorek	29	30
środa	32	30
czwartek	36	30
piątek	20	30

Postawiono hipotezy:

\mathcal{H}_0 : liczba wydawanych obiadów w badanej stołówce szkolnej w kolejnych dniach tygodnia jest zgodna z oczekiwaną liczbą wydawanych obiadów w tych dniach

\mathcal{H}_1 : liczba wydawanych obiadów w badanej stołówce szkolnej w kolejnych dniach tygodnia nie jest zgodna z oczekiwaną liczbą wydawanych obiadów w tych dniach

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



Wartość p z rozkładu χ^2 dla 4 stopni swobody wynosi 0.287297. Zatem na poziomie istotności $\alpha = 0.05$ możemy powiedzieć, że nie mamy 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](#). Ta poprawka jest używana by ograniczyć wielkość popełnionego błędu pierwszego rodzaju, gdy porównujemy wartości obserwowane i oczekiwane pomiędzy wybranymi dniami np:

Pt \iff Pn,

Pt \iff Wt,

Pt \iff Śr,

Pt \iff 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. Poziom istotności dla pojedynczego porównania zgodnie z poprawką Bonferroniego wynosi dla naszego przykładu $\alpha = \frac{0.05}{4} = 0.0125$.

Należy jednak pamiętać, że redukując α dla każdego porównania zmniejszamy również **moc testu**.

12.2.3 Testy dla jednej proporcji

Testy dla jednej proporcji stosujemy, gdy mamy do uzyskania dwa możliwe wyniki (jeden z nich to wynik wyróżniony o licznosci m) i wiemy, jak często te wyniki pojawiają się w próbie (znamy proporcję p). 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 licznosci. Testy te służą do weryfikacji hipotezy, że proporcja w populacji z której pochodzi próba to zadana wartość.

Podstawowe warunki stosowania:

- pomiar na skali **nominalnej** - ewentualne uporządkowanie kategorii nie jest brane pod uwagę.

Dodatkowy warunek dla testu Z dla jednej proporcji:

- duża licznosc (według interpretacji Marascuilo i McSweeney (1977)[113] każda z wartości $np > 5$ i $n(1 - p) > 5$).

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: p = p_0, \\ \mathcal{H}_1 &: p \neq p_0,\end{aligned}$$

gdzie:

p – prawdopodobieństwo (wyróżniona proporcja) w populacji,
 p_0 – prawdopodobieństwo oczekiwane (proporcja oczekiwana).

Test Z dla jednej proporcji

Test Z dla jednej proporcji (*ang. Z test for one proportion*).

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 – licznosc wartości wyszczególnionych w próbie,
 n – licznosc 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 bez korekcy na ciągłość jak i z tą korekcją ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Test dwumianowy

Test dwumianowy (*ang. Binominal test for one proportion*) 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 $1 - p_0$.

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.

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

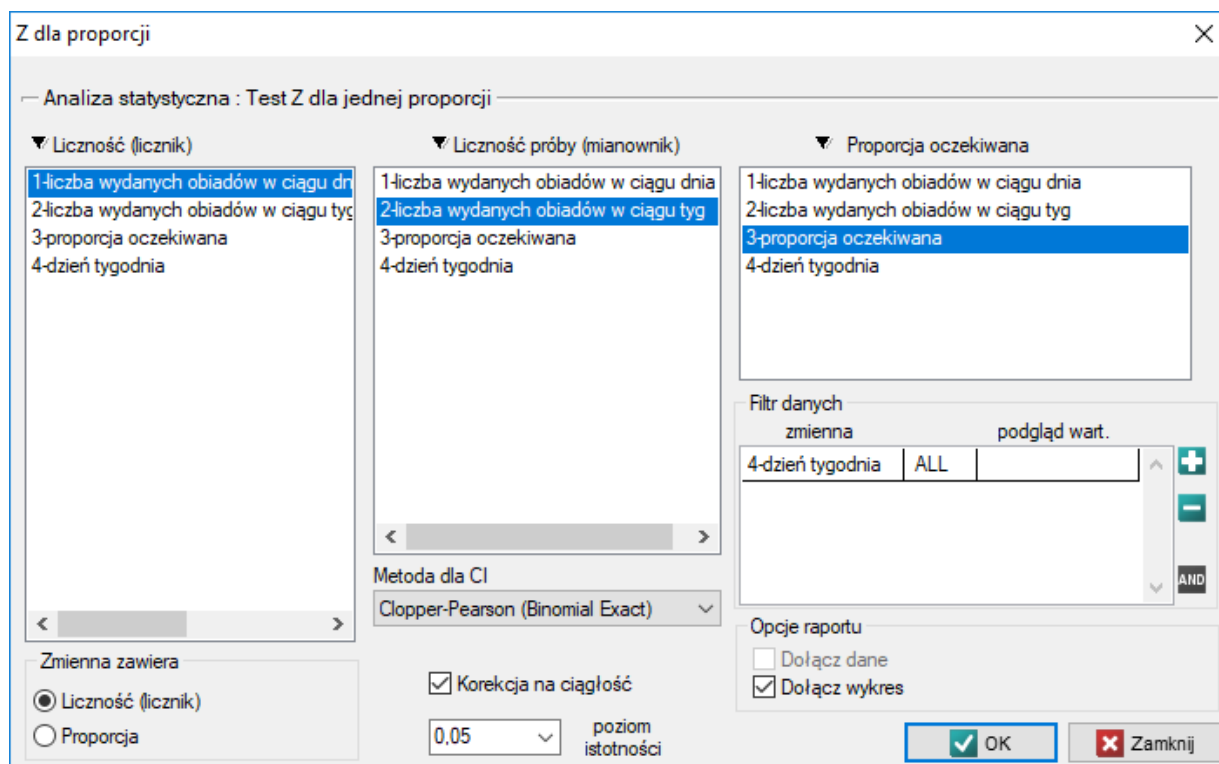
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

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)[173] a także Agresti i Coull (1998)[5]. Przedziały Cloppera i Pearsona (1934)[38] są dokładniejsze dla prób o mniejszej liczności.

Porównanie metod budowania przedziałów dla proporcji można znaleźć w pracy Brown L.D i innych (2001)[29].

Okno z ustawieniami opcji testu Z dla jednej proporcji wywołujemy poprzez menu Statystyka→Testy nieparametryczne→Z dla proporcji.



PRZYKŁAD 12.3 c.d. (plik obiady.pqs)

Założmy, 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

Przy ustawianiu opcji analizy włączamy **filtr** wybierając odpowiedni dzień tygodnia - czyli piątek. Brak ustawienia filtru nie generuje błędu a jedynie wyliczenie kolejnych statystyk dla kolejnych dni tygodnia.

Hipotezy:

\mathcal{H}_0 : w piątek, w stołówce szkolnej wydaje się $\frac{1}{5}$

spośród obiadów wydawanych w ciągu tygodnia w tej stołówce,

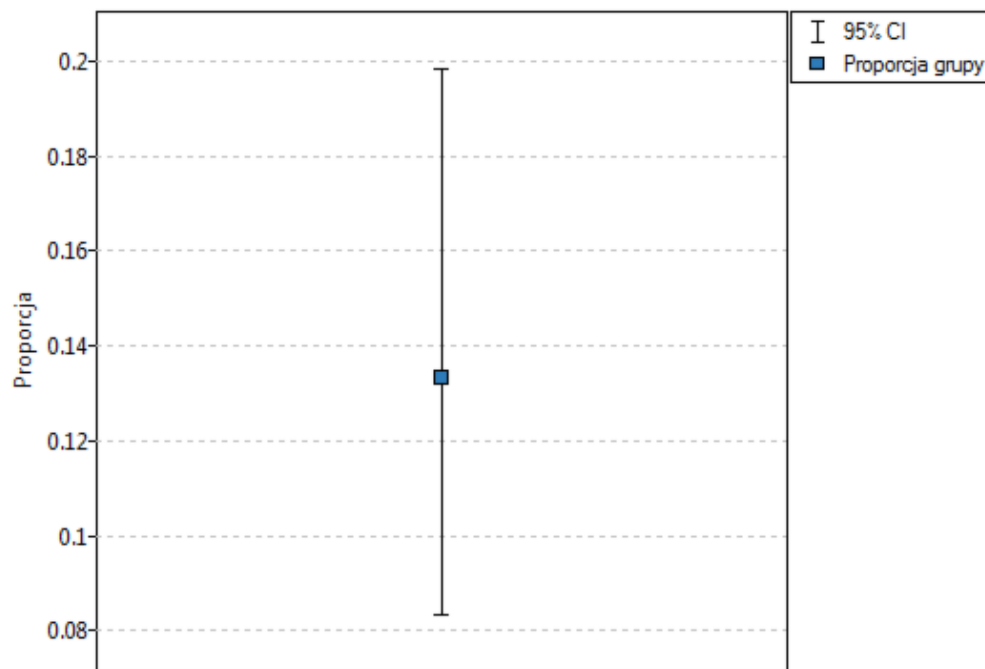
\mathcal{H}_1 : w piątek, w stołówce szkolnej wydaje się istotnie więcej lub mniej niż $\frac{1}{5}$

spośród obiadów wydawanych w ciągu tygodnia w tej stołówce.

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 :

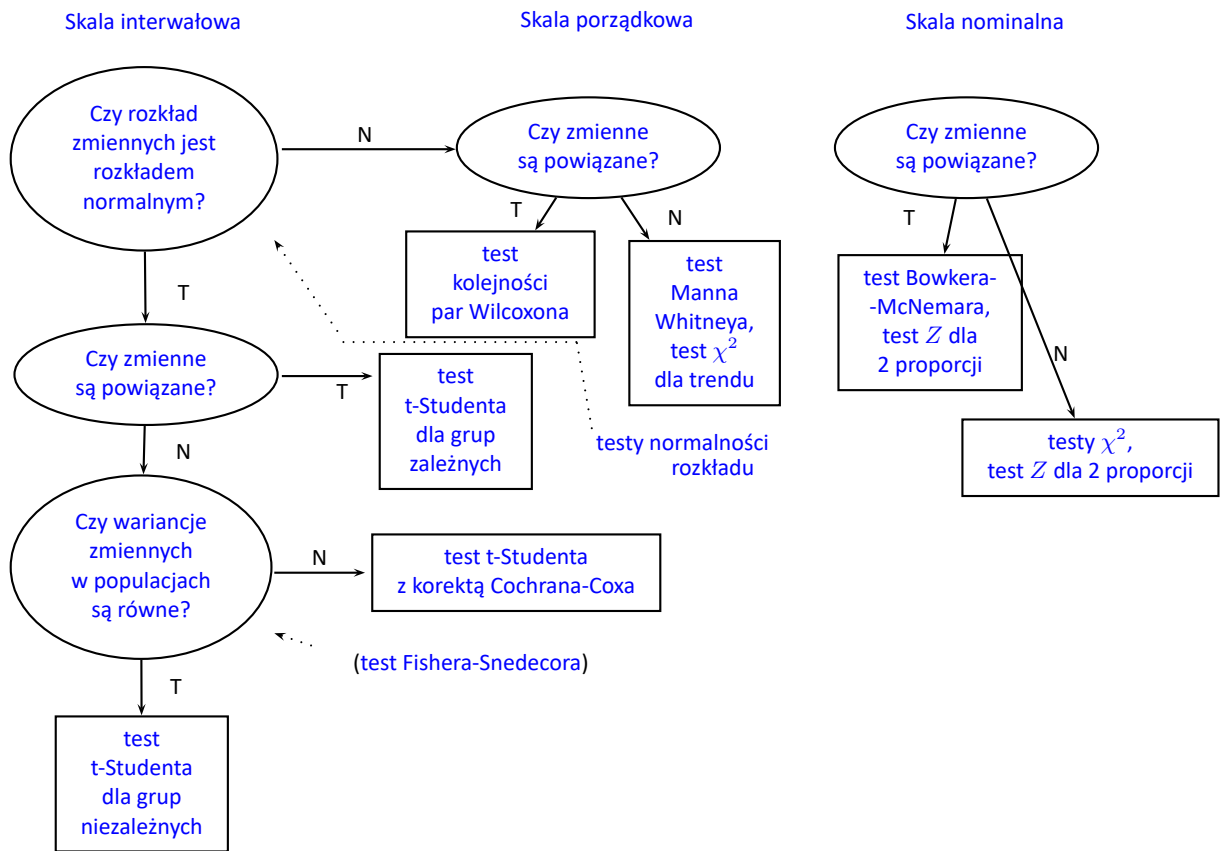
v.1	v.2	v.3
20	150	0,2



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.

Na podstawie testu Z bez poprawki na ciągłość ($p=0.041227$) jak i na podstawie dokładnej wartości prawdopodobieństwa wyliczonego z rozkładu dwumianowego ($p=0.044711$) mogliśmy przyjąć (na poziomie istotności $\alpha = 0.05$), że w piątek wydaje się statystycznie mniej niż $\frac{1}{5}$ obiadów wydawanych przez cały tydzień. Po zastosowaniu poprawki na ciągłość jednak nie udaje się odrzucić hipotezy zerowej ($p=0.052479$).

13 PORÓWNANIE - 2 GRUPY



13.1 TESTY PARAMETRYCZNE

13.1.1 Test Fishera-Snedecora

Test Fishera-Snedecora (*ang. F-Snedecor test*) opiera się na zmiennej F sformułowanej przez Fishera (1924), a jej rozkład opisał Snedecor. Test ten służy do weryfikacji hipotezy o równości **wariancji** badanej zmiennej w dwóch populacjach.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- **normalność rozkładu** badanej zmiennej w obu populacjach,
- **model niezależny**.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: \sigma_1^2 = \sigma_2^2, \\ \mathcal{H}_1 &: \sigma_1^2 \neq \sigma_2^2,\end{aligned}$$

gdzie:

σ_1^2, σ_2^2 – wariancje badanej zmiennej w pierwszej i drugiej populacji.

Statystyka testowa ma postać:

$$F = \frac{sd_1^2}{sd_2^2},$$

gdzie:

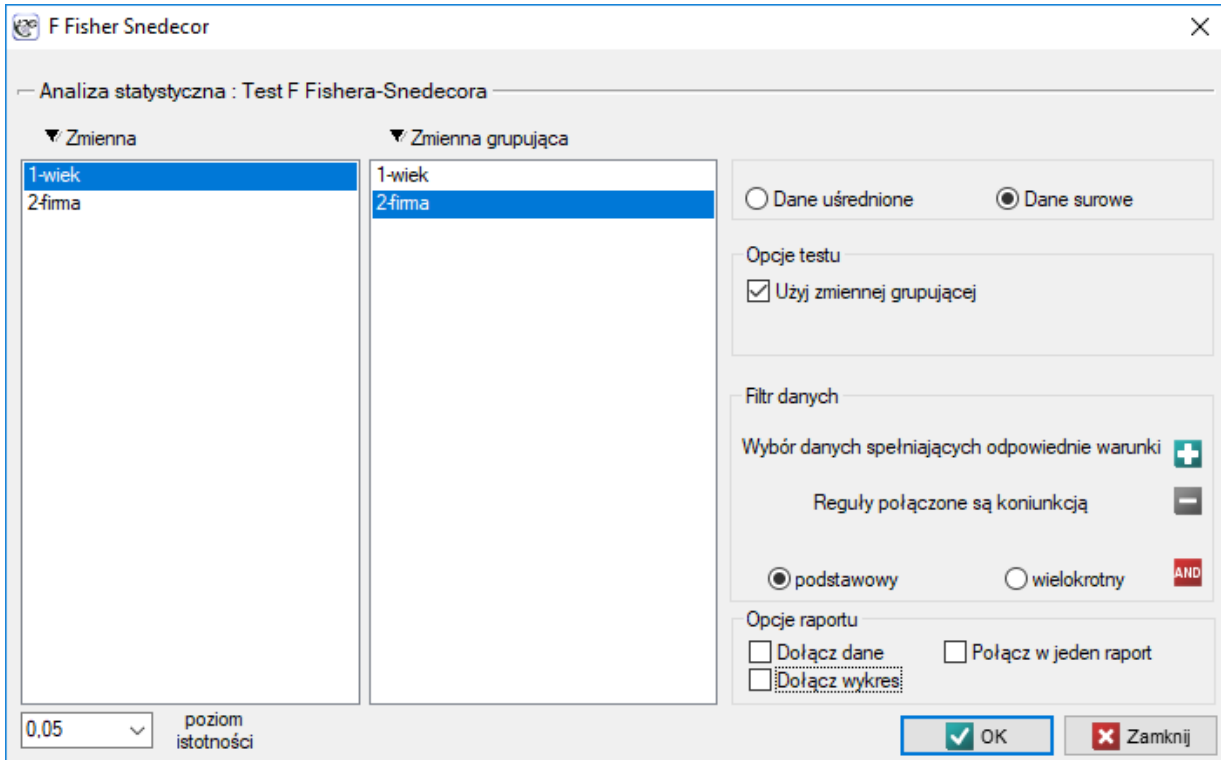
sd_1^2, sd_2^2 – wariancje badanej zmiennej w próbach wybranych losowo z pierwszej i drugiej populacji.

Statystyka ta podlega **rozkładowi F Snedecora** z $n_1 - 1$ i $n_2 - 1$ stopniami swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Okno z ustawieniami opcji testu F Fishera-Snedecora wywołujemy poprzez menu Statystyka→Testy parametryczne→F Fisher Snedecor.


Uwaga!

Obliczenia mogą bazować na danych w postaci **surowych** rekordów lub danych uśrednionych tzn. odchyleniach standardowych i liczności prób.

13.1.2 Test *t*-Studenta dla grup niezależnych

Test *t*-Studenta dla grup niezależnych (*ang. t test for independent groups*) służy do weryfikacji hipotezy o równości **średnich** badanej zmiennej w dwóch populacjach.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- **normalność rozkładu** badanej zmiennej w obu populacjach,
- **model niezależny**,
- **równość wariancji** badanej zmiennej obu populacji.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : \mu_1 &= \mu_2, \\ \mathcal{H}_1 : \mu_1 &\neq \mu_2. \end{aligned}$$

gdzie:

μ_1, μ_2 – średnie badanej zmiennej w pierwszej i drugiej populacji.

Statystyka testowa ma postać:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}},$$

gdzie:

\bar{x}_1, \bar{x}_2 – średnie w pierwszej i drugiej próbie,

n_1, n_2 – licznosci w pierwszej i drugiej próbie,

sd_1^2, sd_2^2 – **wariancje** w pierwszej i drugiej próbie.

Statystyka testowa ma **rozkład t-Studenta** z $df = n_1 + n_2 - 2$ stopniami swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

- **wspólne odchylenie standardowe** wyraża się wzorem:

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

- **błąd standardowy różnicy średnich** wyraża się wzorem:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}.$$

Standaryzowana wielkość efektu

Współczynnik **d-Cohena** określa jak dużą częśćią występującej zmienności jest różnica między średnimi.

$$d = \left| \frac{\bar{x}_1 - \bar{x}_2}{SD_p} \right|$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [45] wskazówkami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

13.1.3 Test t-Studenta z korektą Cochran-Coxa

Poprawka Cochran-Coxa dotyczy **testu t-Studenta dla grup niezależnych** (ang. *Cochran-Cox adjustment*), (1957)[42] i jest wyliczana wówczas, gdy wariancje badanych zmiennych w obu populacjach są różne.

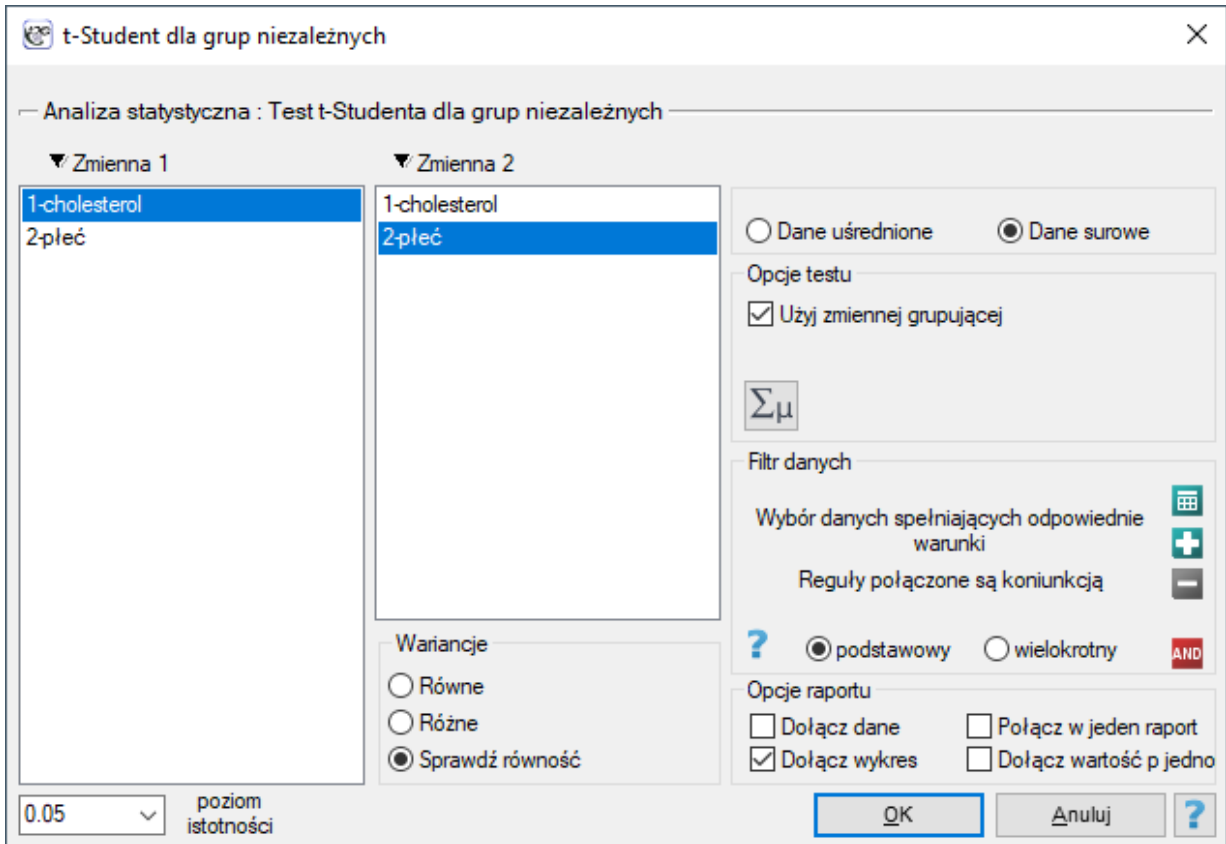
Statystyka testowa ma postać:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}}.$$

Statystyka testowa ma **rozkład t-Studenta** z liczbą stopni swobody zaproponowaną przez Satterthwaite (1946)[144] i wyliczaną z wzoru:

$$df = \frac{\left(\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2} \right)^2}{\left(\frac{sd_1^2}{n_1} \right)^2 \cdot \frac{1}{(n_1-1)} + \left(\frac{sd_2^2}{n_2} \right)^2 \cdot \frac{1}{(n_2-1)}}.$$

Okno z ustawieniami opcji testu t-Studenta dla grup niezależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → t-Student dla grup niezależnych lub poprzez **Kreator**.



Gdy w oknie testu w opcji dotyczącej wariacji wybierzemy:

- równe, wówczas zostanie wyliczony test t -Studenta dla grup niezależnych,
- różne, wówczas zostanie wyliczony test t -Studenta z korektą Cochran-Coxa,
- sprawdź równość, wówczas zostanie wyliczony test Fishera-Snedecora a na podstawie jego wyniku i ustawionego poziomu istotności zostanie wybrany i wyliczony test t -Studenta dla grup niezależnych z bądź bez poprawki Cochran-Coxa.

Uwaga!

Obliczenia mogą bazować na danych w postaci **surowych** rekordów lub danych uśrednionych tzn. średnich arytmetycznych, odchyleniach standardowych i liczności prób.

PRZYKŁAD 13.1. (plik cholesterol.pqs)

Z populacji kobiet i z populacji mężczyzn w wieku powyżej 40 roku życia wylosowano po 500 osób. Badanie dotyczyło oceny ryzyka chorób sercowo-naczyniowych. Wśród badanych parametrów znajduje się wartość cholesterolu całkowitego. Celem tego badania będzie porównanie mężczyzn i kobiet co do tej wartości. Chcemy bowiem wykazać, że te populacje różnią się już na poziomie cholesterolu całkowitego a nie tylko w obrębie cholesterolu rozbitego na frakcje.

Rozkład wieku w obu grupach jest rozkładem normalnym (zostało to sprawdzone [testem Lillieforsa](#)). Średnia wartość cholesterolu w grupie mężczyzn wyniosła $\bar{x}_1 = 201.1$ a odchylenie standardowe $sd_1 = 47.6$, w grupie kobiet odpowiednio $\bar{x}_2 = 191.5$ i $sd_2 = 43.5$. [Test Fishera-Snedecora](#) wskazuje na

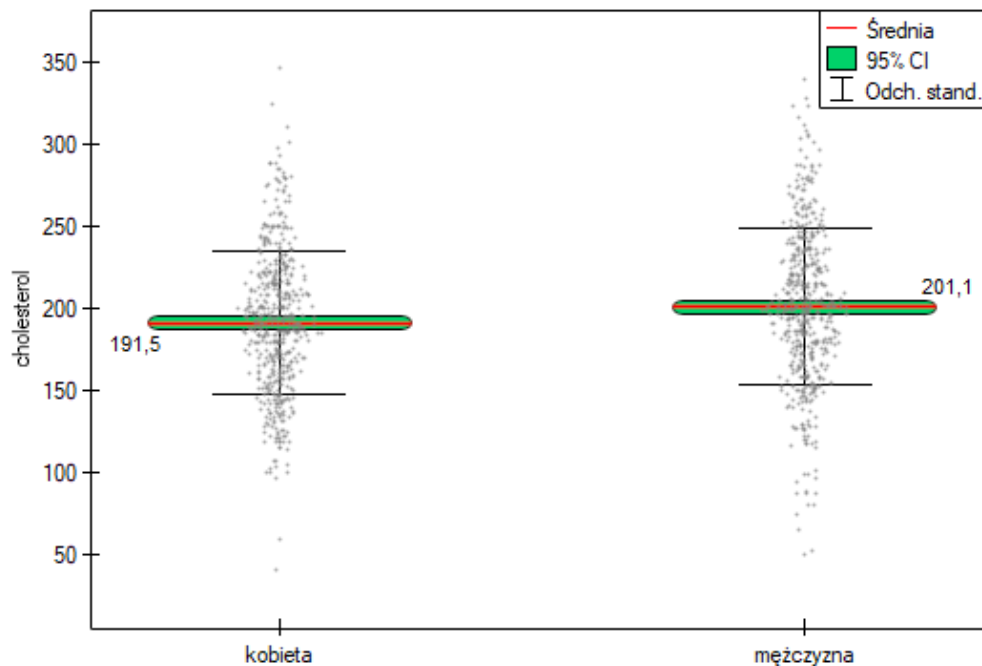
niewielkie lecz istotne statystycznie ($p = 0.0434$) różnice w wariancjach. W analizie zastosowany został test t-Studenta z poprawką Cochran-Coxa.

Hipotezy:

- \mathcal{H}_0 : średni cholesterol całkowity populacji kobiet jest równy
średniemu cholesterolowi całkowitemu populacji mężczyzn,
- \mathcal{H}_1 : średni cholesterol całkowity populacji kobiet jest różny
od średniego cholesterolu całkowitego populacji mężczyzn.

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



Porównując wartość $p = 0.0009$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że kobiety i mężczyźni w Polsce różnią się istotnie statystycznie wartością cholesterolu całkowitego. Przeciętny Polak, który ukończył 40 rok życia ma wyższy cholesterol całkowity od przeciętnej Polki prawie o 10 jednostek.

13.1.4 Test t-Studenta dla grup zależnych

Test t -Studenta dla grup zależnych (*ang. t test for dependent groups*) stosuje się w sytuacji gdy pomiarów badanej zmiennej dokonujemy dwukrotnie w różnych warunkach (przy czym zakładamy, że wariancje zmiennej w obu pomiarach są sobie bliskie). Interesuje nas różnica pomiędzy parami pomiarów ($d_i = x_{1i} - x_{2i}$). Różnica ta wykorzystywana jest do weryfikacji hipotezy o tym, że **średnia** dla niej (dla różnicy) w badanej populacji wynosi 0.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- **normalność rozkładu** różnicy pomiarów d_i ,
- **model zależny**.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: \mu_0 = 0, \\ \mathcal{H}_1 &: \mu_0 \neq 0,\end{aligned}$$

gdzie:

μ_0 , – średnia różnic d_i w populacji.

Statystyka testowa ma postać:

$$t = \frac{\bar{d}}{sd_d} \sqrt{n},$$

gdzie:

\bar{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.

Wyznaczoną na podstawie *statystyki testowej wartość p* porównujemy z α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

- odchylenie standardowe różnicy wyraża się wzorem:

$$sd_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}},$$

- błąd standardowy średniej różnic wyraża się wzorem:

$$SEM_d = \frac{SD_d}{\sqrt{n}}.$$

Standaryzowana wielkość efektu

Współczynnik **d-Cohena** określa jak dużą częśćią występującej zmienności jest różnica między średnimi, biorąc pod uwagę skorelowanie zmiennych.

$$d = \frac{dz}{\sqrt{1 - r_p}}$$

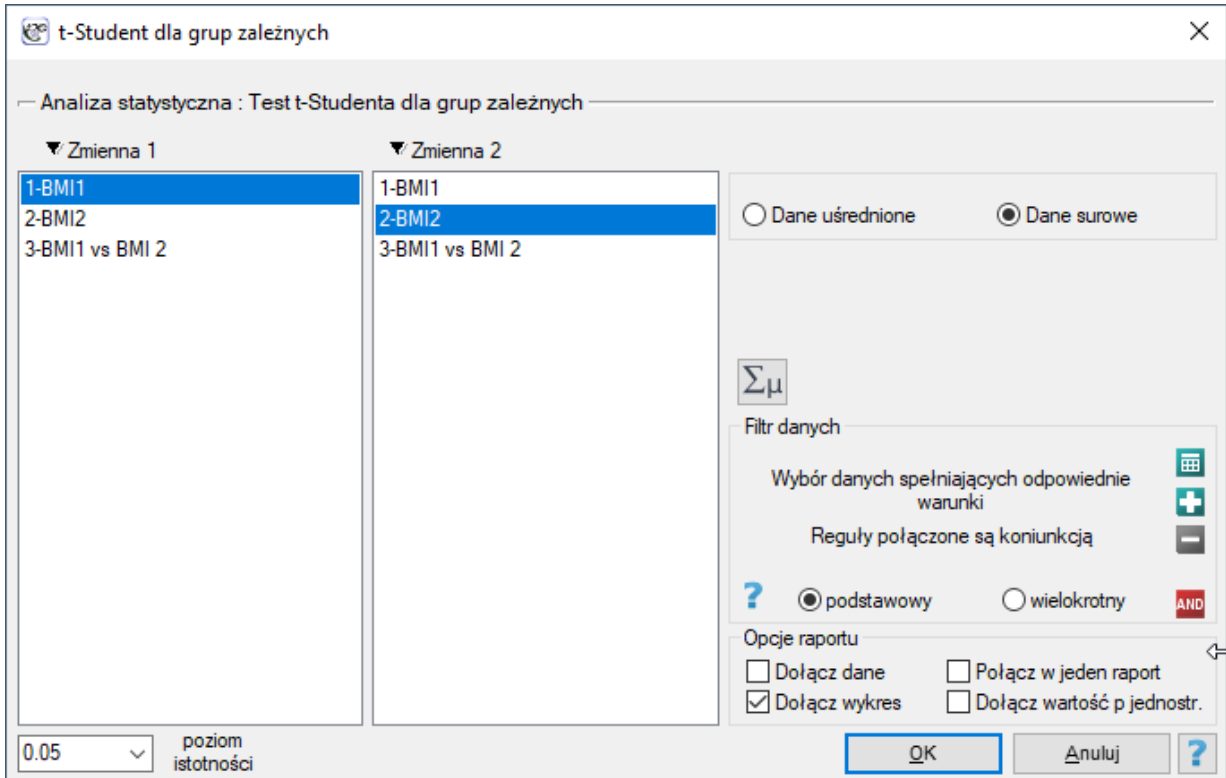
gdzie:

$$dz = \left| \frac{\bar{d}}{sd_d} \right|,$$

r_p - współczynnik korelacji liniowej Pearsona.

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [45] wskazówkami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Okno z ustawieniami opcji testu *t-Studenta* dla grup zależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → *t-Student* dla grup zależnych lub poprzez [Kreator](#).


Uwaga!

Obliczenia mogą bazować na danych w postaci **surowych** rekordów lub danych uśrednionych tzn. średniej różnic, odchyleniu standardowym różnic i liczności próby.

PRZYKŁAD 13.2. (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 stosowania diety. By sprawdzić skuteczność diety porównano uzyskane wyniki pomiarów BMI.

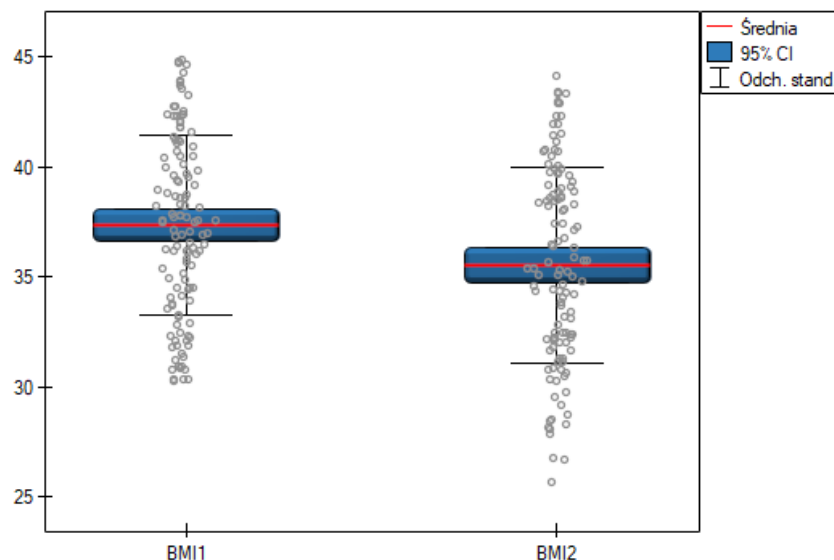
Hipotezy:

\mathcal{H}_0 : średnie wartości BMI nie zmieniają się na skutek stosowania diety

\mathcal{H}_1 : średnie wartości BMI zmieniają się na skutek stosowania diety

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



Porównując wartość $p < 0.0001$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że średni poziom BMI zmienił się istotnie. Przed stosowaniem diety był wyższy średnio o niecałe 2 jednostki.

W badaniu możliwe było stosowanie testu t-Studenta dla grup zależnych, ponieważ rozkład różnicy pomiędzy parami pomiarów był rozkładem normalnym (test Lillieforsa, wartość $p = 0.0837$).

13.2 TESTY NIEPARAMETRYCZNE

13.2.1 Test U Manna-Whitneya

Test U Manna-Whitneya (*ang. Mann-Whitney U test*) znany jest również jako test Wilcoxon-Manna-Whitneya (*ang. Wilcoxon Mann-Whitney test*), Mann i Whitney (1947)[109] oraz Wilcoxon (1949)[172]. Test ten służy do weryfikacji hipotezy o braku przesunięcia porównywanych rozkładów tzn. najczęściej nieistotności różnic pomiędzy medianami badanej zmiennej w dwóch populacjach (przy czym zakładamy, że rozkłady zmiennej są sobie bliskie - porównanie wariancji rang można sprawdzić testem dla rang Conovera).

Podstawowe warunki stosowania:

- pomiar na **skali porządkowej** lub **interwałowej**,
- **model niezależny**.

Hipotezy dotyczą równości średnich rang dla porównywanych populacji lub są upraszczane do median:

$$\begin{aligned}\mathcal{H}_0 &: \phi_1 = \phi_2, \\ \mathcal{H}_1 &: \phi_1 \neq \phi_2,\end{aligned}$$

gdzie:

ϕ_1, ϕ_2 to rozkłady badanej zmiennej w pierwszej i drugiej populacji.

Wyznaczamy wartość statystyki testowej, a na jej podstawie **wartość p** , którą porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Uwaga!

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 to liczności prób, R_1, R_2 to sumy **rang** dla prób.

Standardowo interpretacji podlega mniejsza z wartości U lub U' .

Statystyka ta podlega rozkładowi Manna-Whitneya i nie zawiera poprawki na rangi wiązane. Wartość dokładnego prawdopodobieństwa z rozkładu Manna-Whitneya wyliczana jest z dokładnością do części setnej ułamka.

- Dla próby o dużej liczności

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 \sum (t^3 - t)}{12(n_1 + n_2)(n_1 + n_2 - 1)}},$$

gdzie:

U można zastąpić przez U' ,

t – liczba przypadków wchodzących w skład **rangi wiązanej**.

Wzór na statystykę testową Z zawiera poprawkę na rangi wiązane. Poprawka ta jest stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych poprawka ta nie jest wyliczana, gdyż wówczas $\frac{n_1 n_2 \sum (t^3 - t)}{12(n_1 + n_2)(n_1 + n_2 - 1)} = 0$)

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Poprawka na ciągłość testu Manna-Whitneya (Marascuilo and McSweeney (1977)[113])

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:

$$Z = \frac{|U - \frac{n_1 n_2}{2}| - 0.5}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 \sum (t^3 - t)}{12(n_1 + n_2)(n_1 + n_2 - 1)}}$$

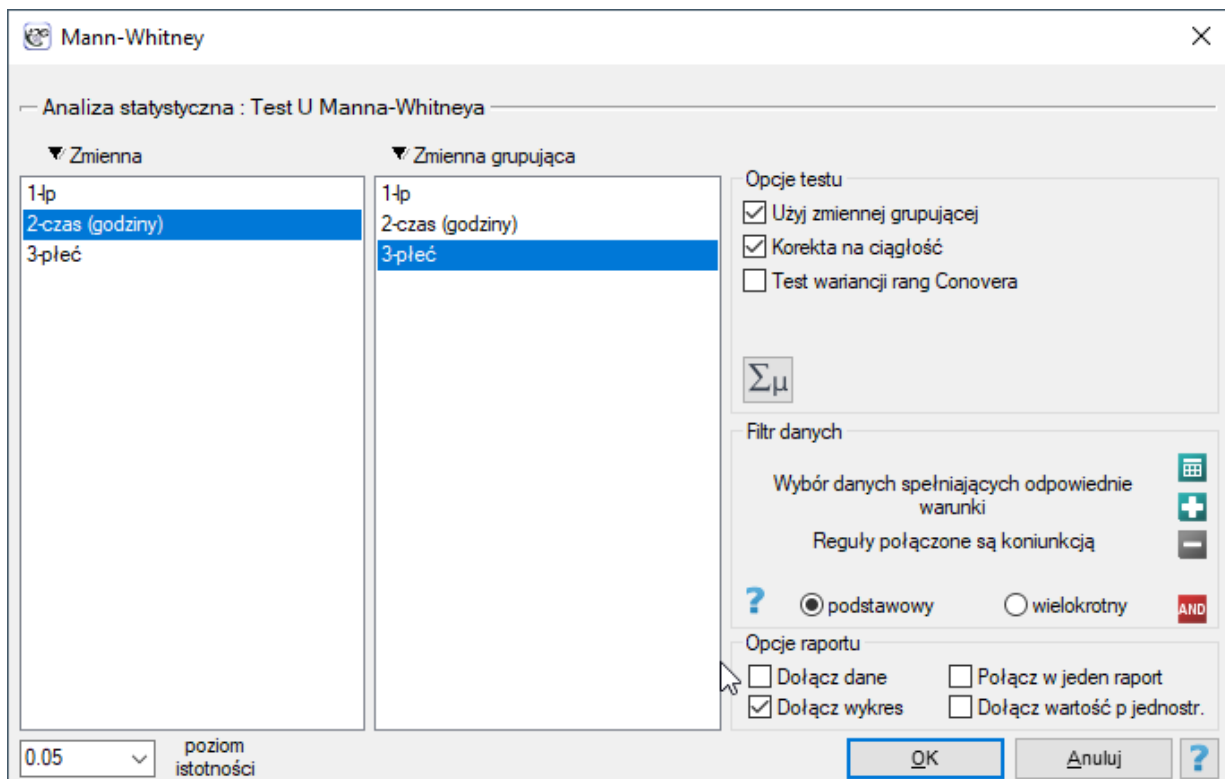
Standaryzowana wielkość efektu

Rozkład statystyki testu Manna-Whitneya jest aproksymowany przez rozkłady normalny, który można przekształcić na wielkość efektu $r = |Z / (n_1 + n_2)|$ [65] by następnie uzyskać wartość **d-Cohena** zgodnie ze standardową konwersją stosowaną przy meta-analizach:

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [45] wskaźnikami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Okno z ustawieniami opcji testu U Manna-Whitneya wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Mann-Whitney lub poprzez [Kreator](#).



PRZYKŁAD 13.3. (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:

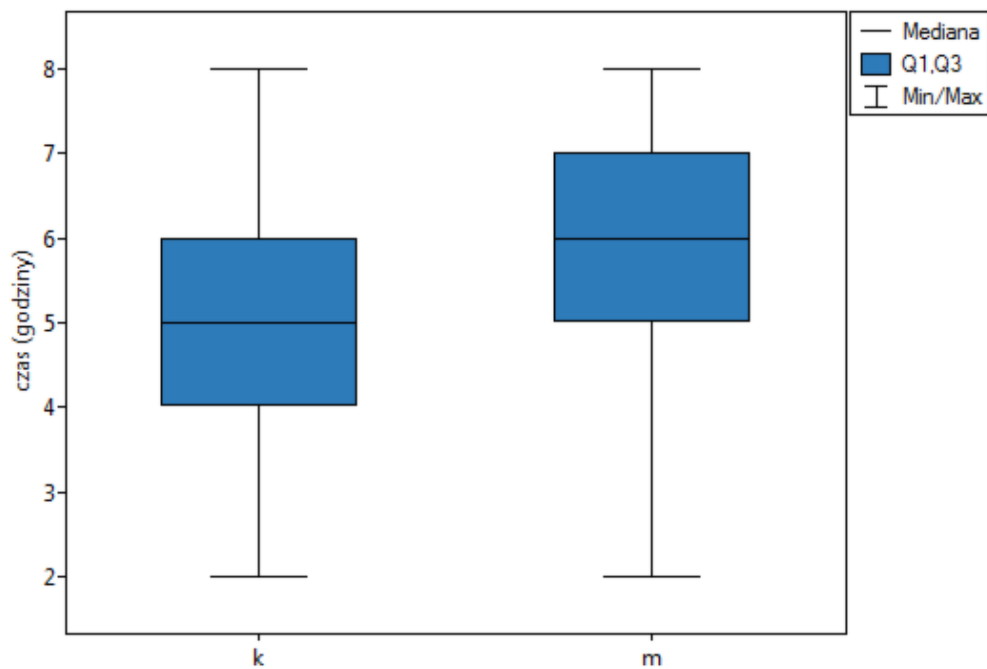
(czas, płeć): (2, k) (2, m) (2, m) (3, k) (3, k) (3, k) (3, k) (3, m) (3, m) (4, k) (4, k) (4, k) (4, k) (4, m) (4, m) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, k) (5, m) (5, m) (5, m) (5, m) (6, k) (6, k) (6, k) (6, k) (6, k) (6, k) (6, m) (6, m) (6, m) (6, m) (6, m) (6, m) (6, m) (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).

Hipotezy:

- \mathcal{H}_0 : mediana czasu spędzanego przed komputerem jest taka sama w populacji studentek i studentów badanej uczelni,
 \mathcal{H}_1 : mediana czasu spędzanego przed komputerem jest inna dla populacji studentek i dla studentów badanej uczelni.

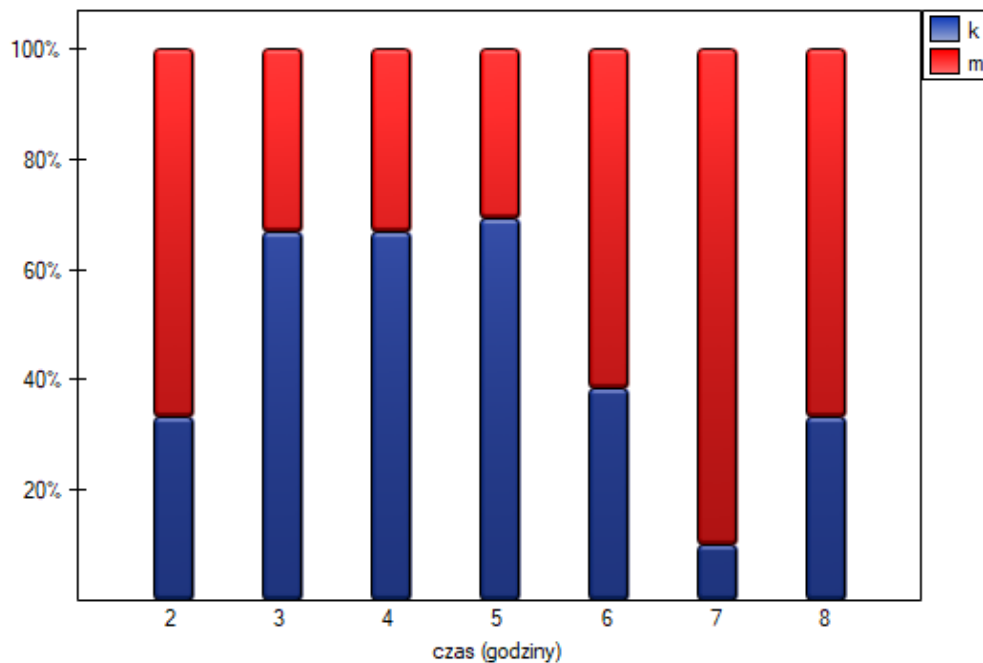
Test U Manna-Whitneya	
Czas analizy	0.20 sek.
Analizowane zmienne	czas (godziny);płeć
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Korekta na ciągłość	Tak
Zmienna grupująca	płeć
Statystyka U	225.5
Statystyka U'	499.5
Wartość p dwustronne (dokładne)	0.0149
Statystyka 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 kwartył	4	5
Górny kwartył	6	7



Na podstawie przyjętego poziomu $\alpha = 0.05$ i statystyki Z testu Manna-Whitneya bez poprawki na ciągłość ($p=0.0154$) jak i z tą poprawką $p = 0.0158$, jak też na podstawie dokładnej statystyki U ($p=0.0149$) możemy przyjąć, że istnieją ważne statystycznie różnice pomiędzy studentkami a studentami matematyki w ilości czasu spędzanego przed komputerem. Różnice te polegają na tym, że studentki spędzają mniej czasu przed komputerem niż studenci. Opisać je można na podstawie mediany, kwartyli oraz wartości największej i najmniejszej, które widzimy również na wykresie typu ramka-wąsy. Innym sposobem opisu różnic jest przedstawienie czasu spędzonego przed komputerem na podstawie tabeli licznosci i procentów (które uruchamiamy w oknie analizy ustawiając statystyki opisowe $\Sigma\mu$) lub na podstawie wykresu kolumnowego.

Licznosc(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



13.2.2 Test Wilcoxona (kolejności par)

Test kolejności par Wilcoxona (*ang. Wilcoxon matched-pairs test*), znany również pod nazwą testu Wilcoxona dla grup zależnych, Wilcoxon (1945[169],1949[171]). Stosuje się go w sytuacji gdy pomiarów badanej zmiennej dokonujemy dwukrotnie w różnych warunkach. Jest on rozszerzeniem na dwie zależne próby **testu rangowanych znaków Wilcoxona** (przeznaczonego dla jednej próby). Interesuje nas różnica pomiędzy parami pomiarów badanej cechy ($d_i = x_{1i} - x_{2i}$) dla każdego z i badanych obiektów. Różnica ta wykorzystywana jest do weryfikacji hipotezy o tym, że **mediana** dla niej (dla różnicy) w badanej populacji wynosi 0.

Podstawowe warunki stosowania:

- pomiar na **skali porządkowej** lub **interwałowej**,
- **model zależny**.

Hipotezy dotyczą równości sumy rang dodatnich i ujemnych lub są upraszczane do median:

$$\begin{aligned}\mathcal{H}_0 &: \theta_0 = 0, \\ \mathcal{H}_1 &: \theta_0 \neq 0,\end{aligned}$$

gdzie:

θ_0 to - mediana d_i w populacji.

Wyznaczamy wartość statystyki testowej, a na jej podstawie **wartość p** , którą porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Uwaga!

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 Wilcoxon'a i nie zawiera poprawki na rangi wiązane.

- Dla próby o dużej liczności

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}},$$

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 jest stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych poprawka ta nie jest wyliczana, gdyż wówczas $\frac{\sum t^3 - \sum t}{48} = 0$).

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Poprawka na ciągłość testu Wilcoxon'a (Marascuilo and McSweeney (1977)[113])

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:

$$Z = \frac{\left| T - \frac{n(n+1)}{4} \right| - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}.$$

Uwaga!

Od wersji 1.8.0 mediana wyliczana dla kolumny różnica obejmuje wszystkie pary wyników za wyjątkiem tych, których różnica wynosi 0.

Standaryzowana wielkość efektu

Rozkład statystyki testu Wilcoxon'a jest aproksymowany przez rozkład normalny, który można przekształcić na wielkość efektu $r = |Z/n|$ [65] by następnie uzyskać wartość **d-Cohena** zgodnie ze standardową konwersją stosowaną przy meta-analizach:

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Przy interpretacji efektu badacze często posługują się ogólnymi, określonymi przez Cohena [45] wskaźnikami definiującymi małą (0.2), średnią (0.5) i dużą (0.8) wielkość efektu.

Okno z ustawieniami opcji testu Wilcoxon'a dla grup zależnych wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Wilcoxon (kolejności par) lub poprzez [Kreator](#).

Wilcoxon (kolejności par)

Analiza statystyczna : Test Wilcoxona dla grup zależnych

Zmienna 1

1-Lp
2-poziom bólu przed kuracją
3-poziom bólu po kuracji

Zmienna 2

1-Lp
2-poziom bólu przed kuracją
3-poziom bólu po kuracji

Opcje testu

Korekta na ciągłość

$\Sigma\mu$

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

? podstawowy wielokrotny **AND**

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres Dołącz wartość p jednostr.

0.05 poziom istotności

OK **Anuluj** ?

PRZYKŁAD 13.4. (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.

Hipotezy:

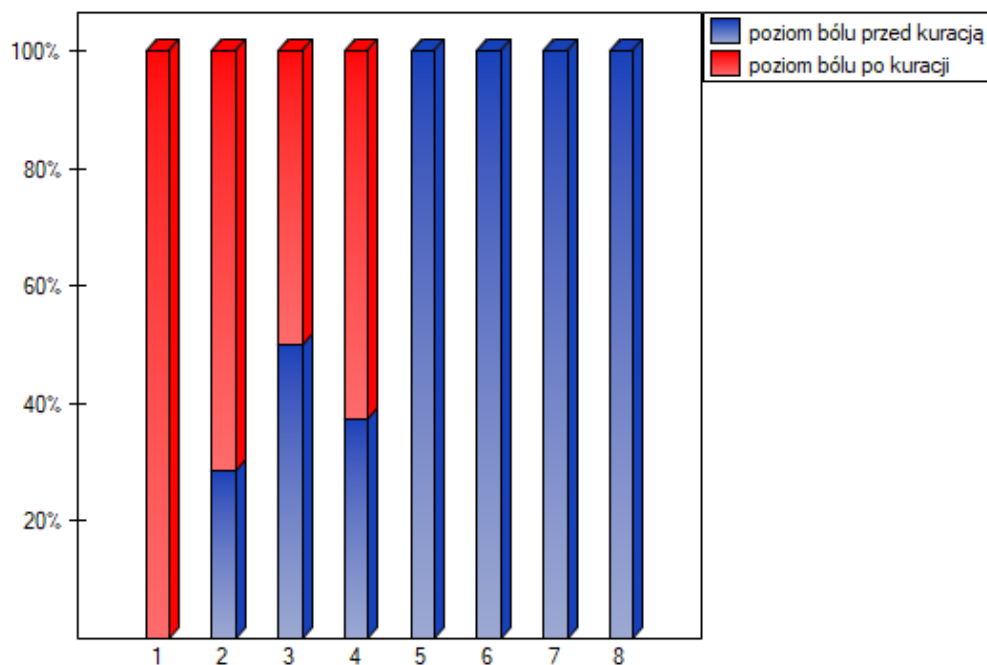
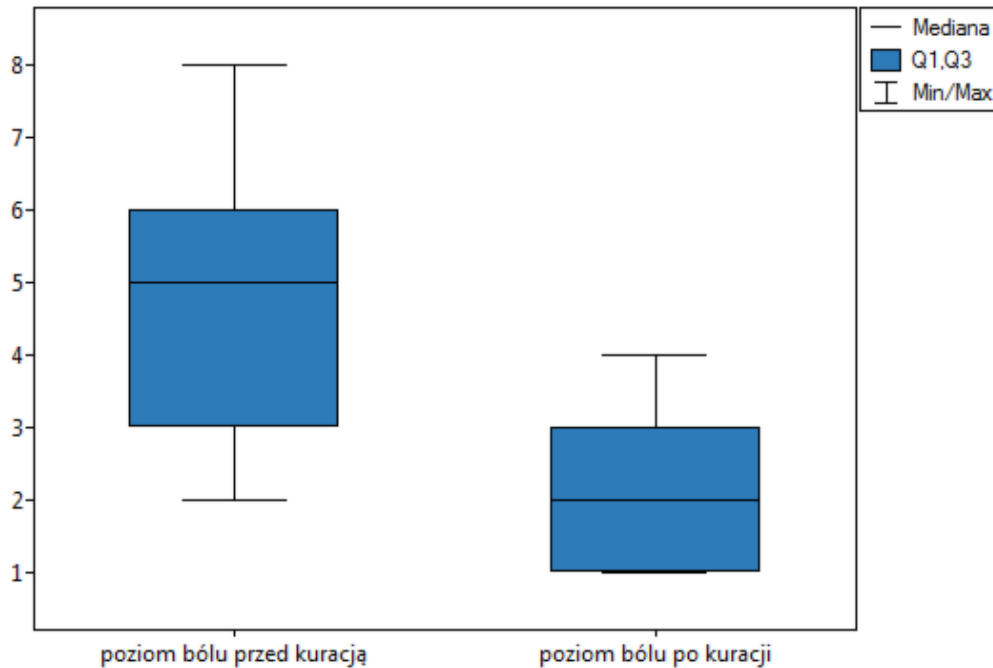
\mathcal{H}_0 : w badanej populacji mediana różnic pomiędzy poziomem bólu przed i po miesiącu kuracji wynosi 0,

\mathcal{H}_1 : w badanej populacji mediana różnic pomiędzy poziomem bólu przed i po miesiącu kuracji jest różna od 0.

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 związanych)	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 kwartył	3	1
Górny kwartył	6	3
Suma rang dod.	3,5	
Suma rang ujen	186,5	

Porównując wartość $p = 0.000211$ testu Wilcoxona opartego o statystykę T z poziomem istotności $\alpha = 0.05$ stwierdzamy, że istnieje ważna statystycznie różnica w poziomie odczuwanego bólu pomiędzy dwoma badaniami. Różnica ta polega na tym, że poziom bólu spadł (suma rang ujemnych jest znacznie większa niż suma rang dodatnich). Taką samą decyzję podjęlibyśmy również na podstawie wartości $p = 0.00021$ lub $p = 0.000229$ testu Wilcoxona opartego o statystykę Z lub Z z poprawką na ciągłość. Różnice możemy zaobserwować na wykresie typu ramka-wąsy lub wykresie kolumnowym.



13.2.3 Testy chi-kwadrat

Testy te opierają się na danych zebranych w postaci tabeli kontyngencji 2 cech, cechy X i cechy Y, z których pierwsza ma r a druga c kategorii, a więc powstała tabela ma r wierszy i c kolumn. Z tego

względu możemy mówić o teście chi-kwadrat 2x2 (dla tabel o dwóch wierszach i dwóch kolumnach) lub o teście chi-kwadrat RxC (o wielu wierszach i kolumnach). (patrz tabela (8.1)).

Szczegółowe informacje na temat testu chi-kwadrat dwóch cech możemy przeczytać tutaj:

[test chi-kwadrat 2x2](#)

[test chi-kwadrat RxC](#).

Podstawowe warunki stosowania:

- pomiar na [skali nominalnej](#) - ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- [model niezależny](#).

Dodatkowy warunek dla testu χ^2 :

- duże [liczności oczekiwane](#) według [interpretacji Cochra](#) (1952)[40].

- **Hipotezy w brzmieniu ogólnym:**

$$\mathcal{H}_0 : O_{ij} = E_{ij} \text{ dla wszystkich kategorii,}$$

$$\mathcal{H}_1 : O_{ij} \neq E_{ij} \text{ dla przynajmniej jednej kategorii,}$$

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.

Wyznaczoną [wartość \$p\$](#) porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Dodatkowo

- Oprócz testu chi-kwadrat może zająć konieczność wyznaczenia innego, pokrewnego testu. W przypadku gdy warunek Cochra nie jest spełniony można wyznaczyć:
 - [test Fishera dla tabel RxC](#)
 - [test Fishera dla tabel 2x2](#)
 - [test chi-kwadrat z poprawką Yatesa](#)
 - [test mid-p dla tabel 2x2](#).
- W przypadku gdy uzyskamy tabelę Rx2, i kategorie R można uporządkować, możliwe jest wyznaczenie trendu:
 - [test chi-kwadrat dla trendu dla tabel Rx2](#)
- W przypadku, gdy na podstawie testu wykonanego dla tabeli większej niż 2x2 stwierdzimy występowanie istotnych zależności lub różnic, wówczas można wykonać wielokrotne porównania wraz z odpowiednią korektą porównań wielokrotnych po to, by zlokalizować umiejscowienie tych zależności/różnic. Korekta taka może być dokonana automatycznie, gdy tabela ma wiele kolumn. Wówczas w oknie opcji testu należy zaznaczyć Wielokrotne porównania kolumn (RxC).

- W przypadku, gdy chcemy opisać siłę związku między cechą X i cechą Y możemy wyznaczyć: [miary zależności](#)
- W przypadku, gdy chcemy opisać dla tabel 2x2 wielkość wpływu czynnika ryzyka możemy wyznaczyć: [Iloraz szans \(OR\) i relatywne ryzyko \(RR\)](#).

13.2.4 Test chi-kwadrat dla dużych tabel

Test ten opiera się na danych zebranych w postaci tabeli kontyngencji 2 cech (X, Y), z których pierwsza ma możliwe r kategorii X_1, X_2, \dots, X_r a druga c kategorii Y_1, Y_2, \dots, Y_c (patrz tabela (8.1)).

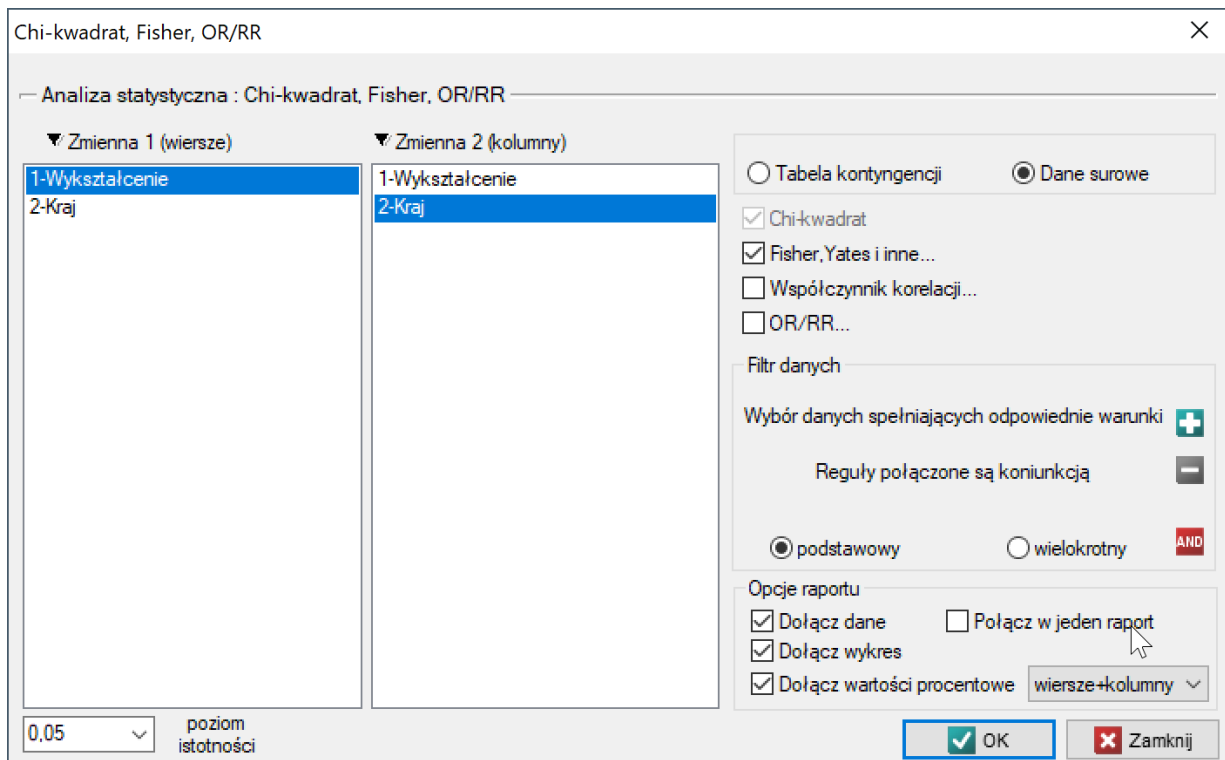
Test χ^2 dla tabel $r \times c$ znany jest również pod nazwą testu χ^2 Pearsona (*ang. Pearson's Chi-square test*), Karl Pearson 1900. Test ten jest rozszerzeniem na 2 cechy testu χ^2 ([dobroci dopasowania](#)). Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Statystyka ta ma asymptotycznie (dla dużych licznosci oczekiwanych) [rozkład \$\chi^2\$](#) z liczbą stopni swobody wyznaczaną według wzoru: $df = (r - 1)(c - 1)$.

Wyznaczoną na podstawie [statystyki testowej wartość \$p\$](#) porównujemy z poziomem istotności α .

Okno z ustawieniami opcji testu Chi-kwadrat (RxC) wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR lub poprzez [Kreator](#).



Chi-kwadrat, Fisher, OR/RR

Analiza statystyczna : Chi-kwadrat, Fisher, OR/RR

Zmienna 1 (wiersze): 1-Wykształcenie, 2-Kraj

Zmienna 2 (kolumny): 1-Wykształcenie, 2-Kraj

Tabela kontyngencji Dane surowe

Chi-kwadrat

Fisher, Yates i inne...

Współczynnik korelacji...

OR/RR...

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

Dołącz wartości procentowe

0,05 poziom istotności

PRZYKŁAD 13.5. (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:

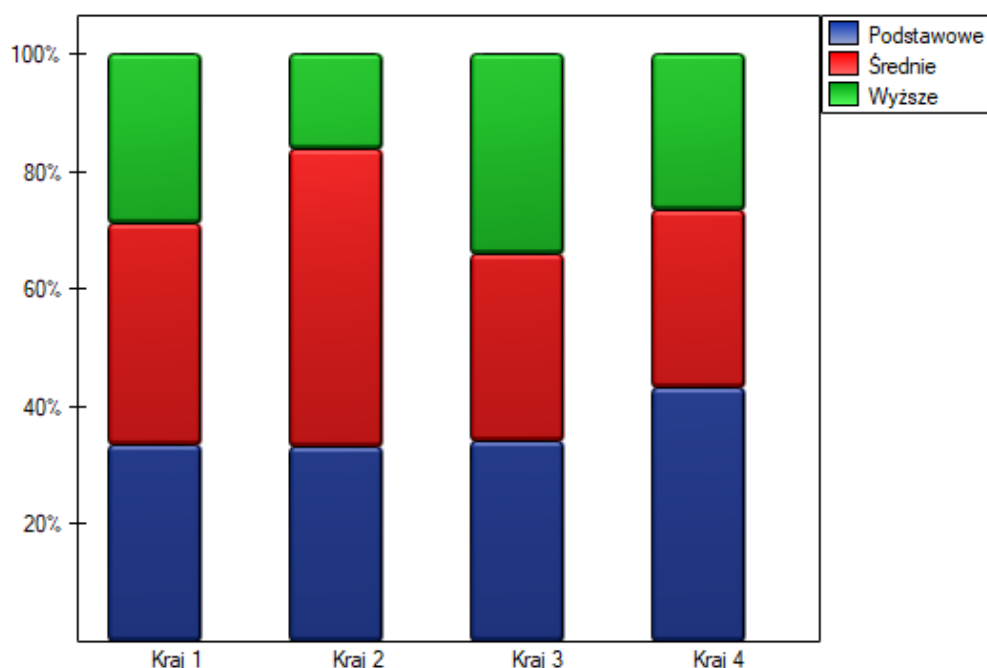
Dane :		✓Kraj			
I Wykształcenie		Kraj 4	Kraj 3	Kraj 2	Kraj 1
Podstawowe		65	52	51	50
Średnie		45	48	78	56
Wyższe		40	52	25	43

Na podstawie tej próby chcielibyśmy się dowiedzieć, czy w badanej populacji istnieje zależność pomiędzy wykształceniem a krajem zamieszkania.

Hipotezy:

- \mathcal{H}_0 : nie istnieje zależność pomiędzy wykształceniem a krajem zamieszkania w badanej populacji,
 \mathcal{H}_1 : istnieje zależność pomiędzy wykształceniem a krajem zamieszkania w badanej populacji.

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 Cochra	spełniony
Statystyka chi-kwadrat Pearsona	23,6104
Stopnie swobody	6
Wartość p	0,0006



Warunek Cochraana jest spełniony.

Wartość $p = 0.0006$. Zatem na poziomie istotności $\alpha = 0.05$ możemy powiedzieć, że istnieje zależność pomiędzy krajem zamieszkania a wykształceniem w badanej populacji.

Jeśli interesują nas dokładniejsze informacje na temat wykrytych zależności, uzyskamy je wyznaczając porównania wielokrotne poprzez opcje Fisher, Yates i inne... a następnie Wielokrotne porównania kolumn (RxC) i jedną z poprawek np. Benjamini-Hochberg

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.

13.2.5 Test Fishera dla tabel dużych tabel

Test Fishera dla tabel $r \times c$ zwany jest również testem Fishera-Freemana-Haltona (*ang. Fisher-Freeman-Halton test*), Freeman G.H., Halton J.H. (1951)[62]. Test ten jest rozszerzeniem na tabele $r \times c$ **testu dokładnego Fishera**. Określa dokładne prawdopodobieństwo wystąpienia konkretnego rozkładu liczb w tabeli przy znanym n i ustalonych sumach brzegowych.

Jeśli sumy brzegowe wierszy zdefiniujemy jako:

$$W_i = \sum_{j=1}^c O_{ij},$$

gdzie:

O_{ij} – **liczności obserwowane** w tabeli kontyngencji,

a sumy brzegowe kolumn jako:

$$K_i = \sum_{i=1}^r O_{ij}.$$

To przy ustalonych sumach brzegowych, dla różnych układów wartości obserwowanych oznaczonych jako U_{ij} wyznaczamy prawdopodobieństwa P :

$$P = \frac{D^{-1} \prod_{j=1}^c K_j!}{U_{1j}! U_{2j}! \dots U_{rj}},$$

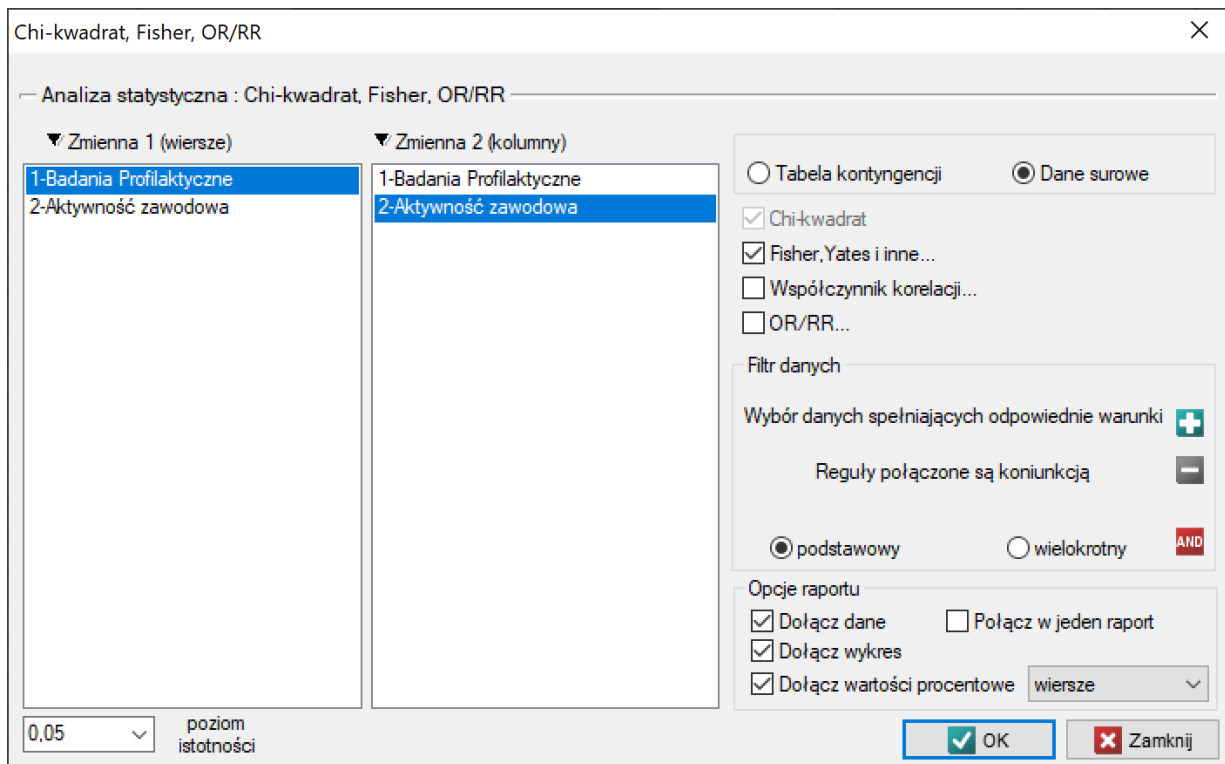
gdzie

$$D = \frac{(W_1 + W_2 + \dots + W_r)!}{W_1! W_2! \dots W_r!}.$$

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}

Porównujemy dokładną **wartość p** z poziomem istotności α .

Okno z ustawieniami opcji testu dokładny Fishera (RxC) wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher OR/RR lub poprzez **Kreator**.



Info.

Procedura obliczania wartości p dla tego testu bazuje na algorytmie opublikowanym w pracy Mehta (1986)[118].

PRZYKŁAD 13.6. (plik praca-profilaktyka.pqs)

W populacji osób zamieszkujących na obszarach wiejskich gminy Komorniki badano czy wykonywanie badań profilaktyki zdrowia jest uzależnione od rodzaju aktywność 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.

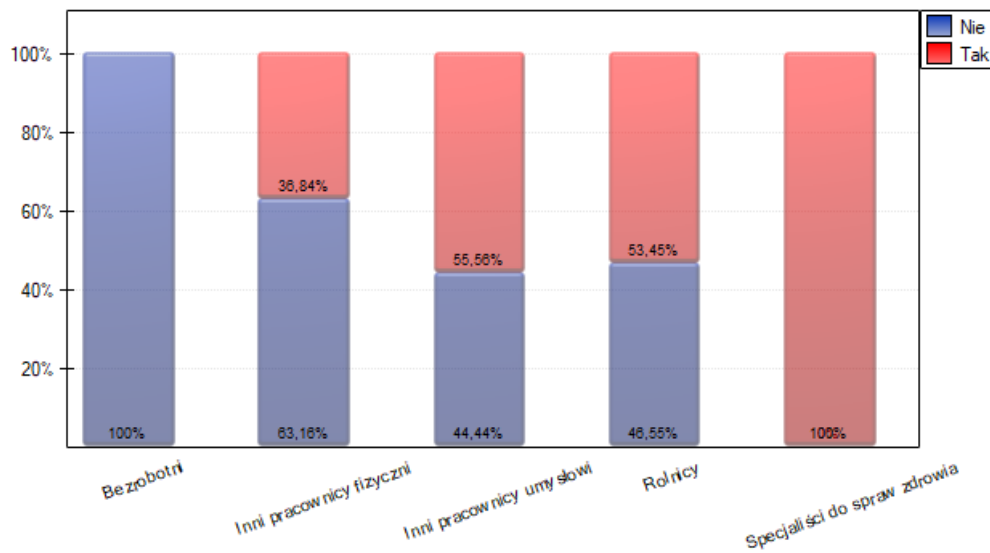
Dane :					Aktywność zawodowa
Badania profilaktyczne	Bezrobotni	Inni pracownicy fizyczni	Inni pracownicy umysłowi	Rolnicy	Specjaliści do spraw zdrowia
Nie	17	12	4	27	0
Tak	0	7	5	31	10

Hipotezy:

H_0 : nie istnieje zależność pomiędzy wykonywaniem badań profilaktycznych a rodzajem wykonywanej pracy przez mieszkańców obszarów wiejskich gminy Komorniki,

H_1 : istnieje zależność pomiędzy wykonywaniem badań profilaktycznych a rodzajem wykonywanej pracy przez mieszkańców obszarów wiejskich gminy Komorniki.

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 Cochra	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



Warunek Cochra nie jest spełniony, przez co nie powinniśmy stosować testu chi-kwadrat.

Wartość $p < 0.0001$. Zatem na poziomie istotności $\alpha = 0.05$ możemy powiedzieć, że istnieje zależność pomiędzy wykonywaniem badań profilaktycznych a rodzajem wykonywanej pracy przez mieszkańców obszarów wiejskich gminy Komorniki.

Jeśli interesują nas dokładniejsze informacje na temat wykrytych zależności, uzyskamy je wyznaczając porównania wielokrotne poprzez opcje Fisher, Yates i inne... a następnie Wielokrotne porównania kolumn (RxC) i jedną z poprawek np. Benjamini-Hochberg.

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 wykonało badania), 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 powoduje, że te trzy grupy nie różnią się od siebie istotnie statystycznie. Część wartości p uzyskanych w tabeli oznaczona jest gwiazdką, oznacza ona te wyniki które powstały poprzez użycie testu dokładnego Fishera wraz z poprawką Benjaminiego-Hochberga, wartości nie oznaczone są wynikiem testu chi-kwadrat wraz z poprawką Benjaminiego-Hochberga, przy którym założenia Cochrańa były spełnione.

13.2.6 Test chi-kwadrat dla małych tabel

Test ten opiera się na danych zebranych w postaci tabeli kontyngencji 2 cech (X, Y), z których każda ma możliwe 2 kategorie X_1, X_2 oraz Y_1, Y_2 (patrz tabela (8.1)).

Test χ^2 dla tabel 2×2 (ang. *Pearson's Chi-square test*), Karl Pearson 1900. Test ten jest zawężeniem testu χ^2 dla tabel $r \times c$.

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 licznosci oczekiwanych) rozkład χ^2 z jednym stopniem swobody.

Okno z ustawieniami opcji testu Chi-kwadrat i jego poprawek wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR) lub poprzez [Kreator](#).

Chi-kwadrat, Fisher, OR/RR

Analiza statystyczna : Chi-kwadrat, Fisher, OR/RR

	c1	c2
r1	50	40
r2	20	60

wypełnij z zapisanego zaznaczenia

Tabela kontyngencji
 Dane surowe

Chi-kwadrat
 Fisher, Yates i inne...
 Współczynnik korelacji...
 OR/RR...

Opcje raportu

Dołącz dane
 Połącz w jeden raport
 Dołącz wykres
 Dołącz wartości procentowe
 wiersze+kolumny

0,05 poziom istotności

OK Zamknij

PRZYKŁAD 13.7. (plik płęć-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 tabeli kontyngencji:

Liczności obserwowane O_{ij}		zdawalność egzaminu		
		tak	nie	suma
płeć	k	50	40	90
	m	20	60	80
	suma	70	100	170

Hipotezy:

\mathcal{H}_0 : nie istnieje zależność pomiędzy płcią a zdawalnością egzaminu w badanej populacji,

\mathcal{H}_1 : istnieje zależność pomiędzy płcią a zdawalnością egzaminu w badanej populacji.

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 Cochra	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	
		tak	nie
k	i	37,06	52,94
	m	32,94	47,06

Dane :		✓Tabela ko	
		tak	nie
k	i	50	40
	m	20	60

% wiersza		✓Tabela ko	
		tak	nie
k	i	55,56%	44,44%
	m	25%	75%

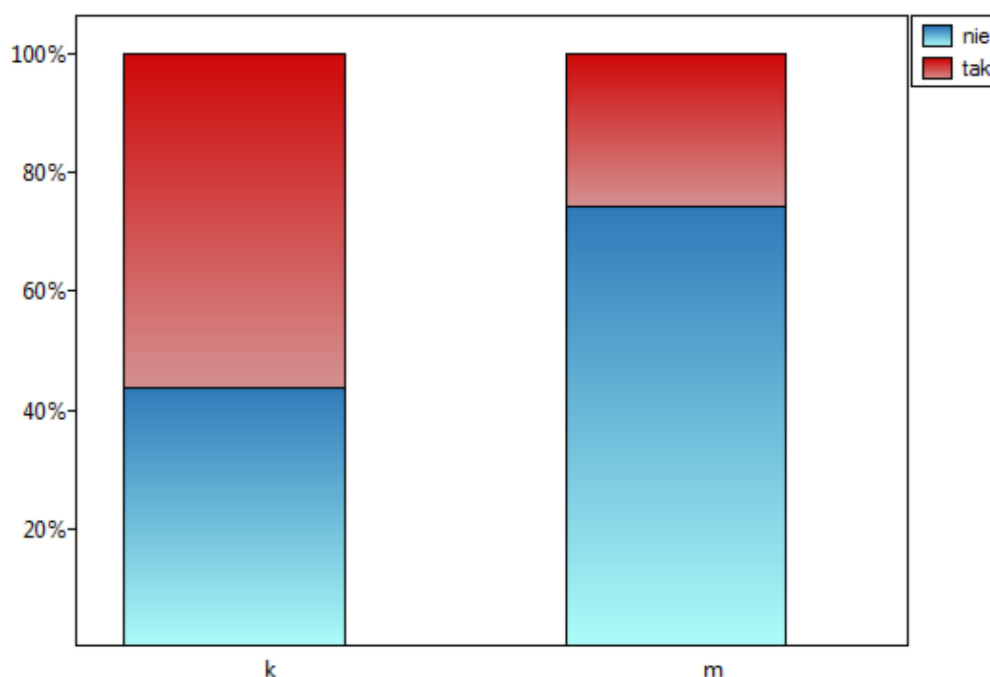


Tabela licznosci oczekiwanych nie zawiera wartosci mniejszych niz 5. Warunek Cochra jest spełniony.

Przy przyjetym poziomie istotnosci $\alpha = 0.05$ wszystkie wykonane testy potwierdzily prawdziwosc hipotezy alternatywnej:

- test chi-kwadrat, wartosc $p = 0.000053$,
- test chi-kwadrat z poprawka Yatesa, wartosc $p = 0.000103$,
- test dokladny Fishera, wartosc $p = 0.000083$,
- test mid-p, wartosc $p = 0.000054$.

13.2.7 Poprawki testu chi-kwadrat dla małych tabel

Testy te opierają się na danych zebranych w postaci tabeli kontyngencji 2 cech (X, Y), z których każda ma możliwe 2 kategorie X_1, X_2 oraz Y_1, Y_2 (patrz tabela (8.1)).

Test chi-kwadrat z poprawką Yatesa na ciągłość

Test χ^2 z poprawką Yatesa (*ang. Chi-square test with Yates correction*), Frank Yates (1934)[175] jest testem 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 .

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 dla tabel 2×2

Test Fishera dla tabel 2×2 nazywany jest również testem dokładnym Fishera (*ang. Fisher exact test*), R. A. Fisher (1934[56], 1935[57]). Test ten określa dokładne prawdopodobieństwo wystąpienia konkretnego rozkładu liczb w tabeli przy znanym n i ustalonych sumach brzegowych.

$$P = \frac{\binom{O_{11}+O_{21}}{O_{11}} \binom{O_{12}+O_{22}}{O_{12}}}{\binom{O_{11}+O_{12}+O_{21}+O_{22}}{O_{11}+O_{12}}}.$$

Przy znanych sumach brzegowych, dla różnych układów wartości obserwowanych wyznaczamy prawdopodobieństwa P . Dokładny poziom istotności p jest sumą tych prawdopodobieństw, które są mniejsze lub równe badanemu prawdopodobieństwu.

Test mid-p

mid-p jest korektą testu dokładnego Fishera. Ta zmodyfikowana wartość p jest rekomendowana przez wielu statystyków (Lancaster 1961[96], Anscombe 1981[9], Pratt i Gibbons 1981[134], Plackett 1984[133], Miettinen 1985[119] i Barnard 1989[15], Rothman 2008[139]) jako metoda zmniejszenia konserwyzm testu dokładnego Fishera. W rezultacie testem mid-p szybciej odrzucimy hipotezę zerową niż dokładnym testem Fishera. Dla dużych prób wartość p otrzymana przy pomocy testu χ^2 z poprawką Yatesa i test Fishera dają zbliżone wyniki, natomiast wartość p testu χ^2 bez korekcji koresponduje z wartością mid-p.

Wartość p mid-p wyznaczana jest przez przekształcenie wartości prawdopodobieństwa dla testu dokładnego Fishera. Jednostronna wartość p wyznaczana jest ze wzoru:

$$PI(\text{mid-p}) = PI(\text{Fisher}) - 0.5 \cdot P_{\text{punktu(tabeli zadanej)}},$$

gdzie:

$PI(\text{mid-p})$ - wartość jednostronna p testu mid-p

$PI(\text{Fisher})$ - wartość jednostronna p testu dokładnego Fishera

a dwustronna wartość p jest definiowana jako podwojona wartość mniejszego z jednostronnych prawdopodobieństw:

$$PII(\text{mid-p}) = 2PI(\text{mid-p}),$$

gdzie:

$PII(\text{mid-p})$ - wartość dwustronna p testu mid-p.

Okno z ustawieniami opcji testu Chi-kwadrat i jego poprawek wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR) lub poprzez [Kreator](#).

13.2.8 Test chi-kwadrat dla trendu

Test χ^2 dla trendu, nazywany również testem dla trendu Cochrańa-Armitage (ang. *Cochran-Armitage test for trend*)[41][10], służy do weryfikacji hipotezy o istnieniu trendu w proporcjach dla poszczególnych kategorii badanej zmiennej (cechy). Opiera się na danych zebranych w postaci **tabeli kontyngencji 2 cech**, z których pierwsza ma możliwe r uporządkowanych kategorii: X_1, X_2, \dots, X_r a druga 2 kategorie G_1, G_2 (tabela (13.1)).

Tabela 13.1. Tabela kontyngencji $r \times 2$ licznosci 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	$C_1 = \sum_{i=1}^r O_{i1}$	$C_2 = \sum_{i=1}^r O_{i2}$	$n = C_1 + C_2$

Podstawowe warunki stosowania:

- pomiar na **skali porządkowej** lub **interwałowej**,
- **model niezależny** (druga cecha, to 2 niezależne grupy).

Niech p_1, p_2, \dots, p_r oznaczają proporcje $p_1 = \frac{O_{11}}{W_1}, p_2 = \frac{O_{21}}{W_2}, \dots, p_r = \frac{O_{r1}}{W_r}$.

Hipotezy:

- \mathcal{H}_0 : W badanej populacji nie istnieje trend w proporcjach p_1, p_2, \dots, p_r ,
 \mathcal{H}_1 : W badanej populacji istnieje trend w proporcjach p_1, p_2, \dots, p_r .

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.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

- jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji testu Chi-kwadrat dla trendu wywołujemy poprzez menu Statystyka \rightarrow Testy nieparametryczne \rightarrow Chi-kwadrat, Fisher, OR/RR, opcja Fisher, Yates i inne... \rightarrow Chi-kwadrat dla trendu.

Chi-kwadrat, Fisher, OR/RR

Analiza statystyczna : Chi-kwadrat, Fisher, OR/RR


Zmienna 1 (wiersze)
1-Wykształcenie
2-Statut palenia


Zmienna 2 (kolumny)
1-Wykształcenie
2-Statut palenia


Tabela kontyngencji Dane surowe

Chi-kwadrat
 Fisher, Yates i inne...
 Współczynnik korelacji...
 OR/RR...

Filtr danych

Wybór danych spełniających odpowiednie warunki 

Reguły połączone są koniunkcją 

podstawowy wielokrotny 

Opcje raportu

Dołącz dane Połącz w jeden raport
 Dołącz wykres
 Dołącz wartości procentowe wiersze

0,05 poziom istotności

PRZYKŁAD 13.8. (plik palenie-wykształcenie.pqs)

Sprawdzamy czy palenie papierosów związane jest z wykształceniem mieszkańców pewnej wsi. Wylosowano próbę 122 osób. Dane zapisano w pliku.

Zakładamy, że zależność może być dwójakiego typu tzn. czym bardziej wykształceni ludzie, tym częściej palą lub czym bardziej wykształceni ludzie, tym rzadziej palą. Poszukujemy zatem rosnącego lub malejącego trendu.

Przed przystąpieniem do analizy musimy przygotować dane, tzn. musimy wskazać kolejność w jakiej mają się pojawiać kategorie wykształcenia. W tym celu z właściwości zmiennej Wykształcenie wybieramy Kody/Etykiety/Format... i nadajemy kolejność podając kolejne liczby naturalne. Przypisujemy również etykiety.

Wartość	Obowiązująca wartość	Etykieta
gimnazjalne	2	gimnazjalne
podstawowe	1	podstawowe
wyższe	5	wyższe
zasadnicze zawodowe	3	zasadnicze zawodowe
średnie	4	średnie

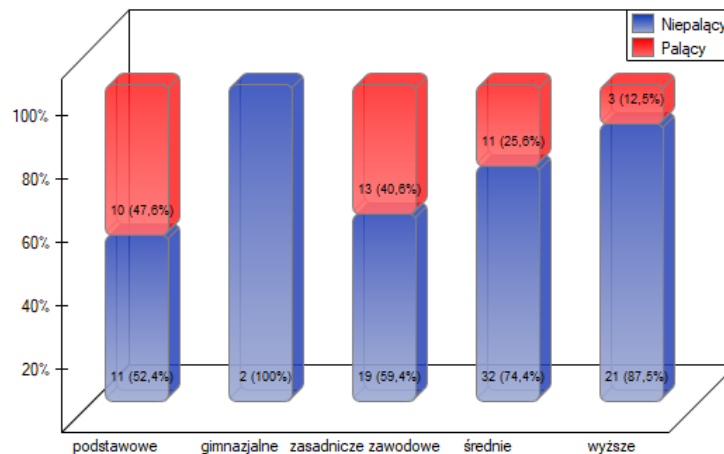
Hipotezy:

\mathcal{H}_0 : w populacji mieszkańców wsi nie istnieje trend wzrostu/spadku liczby palących wraz ze wzrostem wykształcenia,

\mathcal{H}_1 : w populacji mieszkańców wsi istnieje trend wzrostu/spadku liczby palących wraz ze wzrostem wykształcenia.

Statystyka chi-kwadrat dla trendu	6,8098
Stopnie swobody	1
Wartość p	0,0091

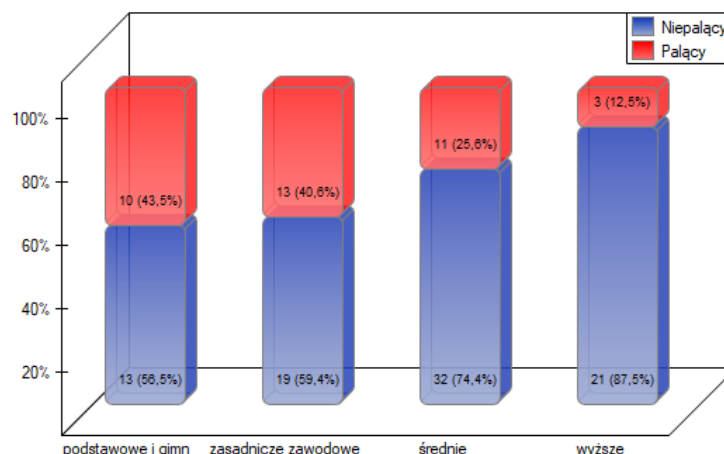
Wartość $p = 0.0091$, co w porównaniu z poziomem istotności $\alpha=0.05$ świadczy o prawdziwości hipotezy alternatywnej mówiącej o występowaniu trendu.



Jak wynika z wykresu czym osoby są bardziej wykształcone, tym rzadziej palą. Jednak wynik uzyskany przez osoby o wykształceniu gimnazjalnym odbiega od tego trendu. Ponieważ wykształcenie gimnazjalne dotyczy tylko dwóch osób, nie miało to dużego wpływu na rysujący się trend. Ze względu na bardzo małą licznosc tej grupy postanowiono analizę powtórzyć dla połączonych kategorii wykształcenia podstawowego i gimnazjalnego.

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

Uzyskano ponownie niewielką wartość $p = 0.0078$ i potwierdzenie istotnego statystycznie trendu.



PRZYKŁAD 13.9. (plik widzowie.pqs)

Z powodu spadku oglądalności pewnego serialu telewizyjnego przeprowadzono badanie opinii widzów tego serialu. W tym celu przepytano 100 osób, które rozpoczęły oglądanie serialu w ostatnim czasie i 300, które oglądają systematycznie serial od początku. Zapytano ich między innymi o ocenę stopnia zaabsorbowania widza losami bohaterów. Wyniki zapisano w tabeli poniżej:

Stopień zaciekawienia	grupa		suma
	grupa nowych widzów	grupa stałych widzów	
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" ale przedstawia się następująco:

Stopień zaciekawienia	grupa		suma
	grupa nowych widzów	grupa stałych widzów	
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%

Hipotezy:

\mathcal{H}_0 : w populacji widzów serialu nie istnieje trend w proporcjach p_1, p_2, p_3, p_4, p_5 ,

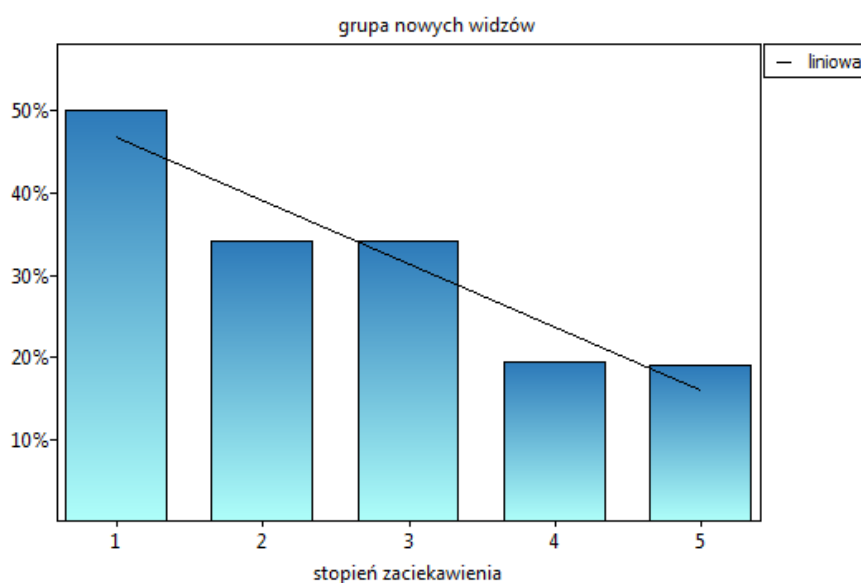
\mathcal{H}_1 : w populacji widzów serialu istnieje trend w proporcjach p_1, p_2, p_3, p_4, p_5 .

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%



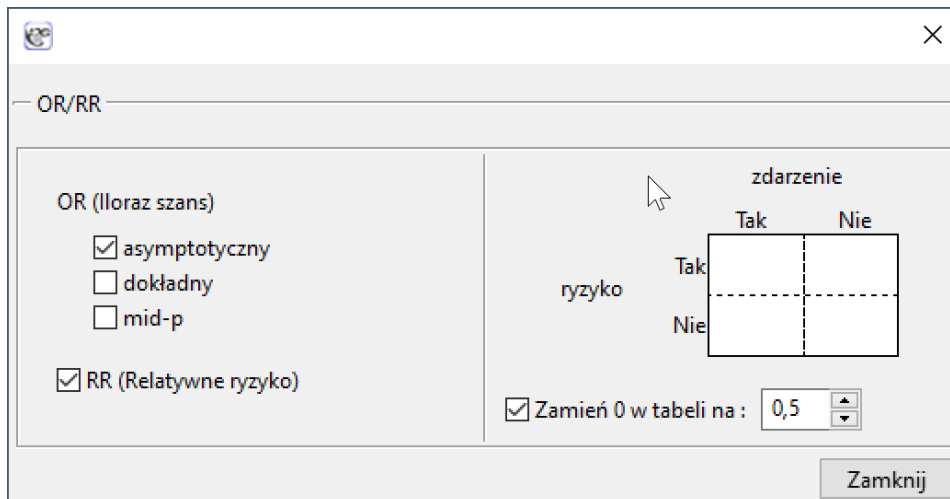
Wartość $p = 0.000436$, co w porównaniu z poziomem istotności $\alpha=0.05$ świadczy o prawdziwości hipotezy alternatywnej mówiącej o występowaniu trendu w proporcjach p_1, p_2, \dots, p_5 . Jak wynika z tabeli kontyngencji wartości procentowych wyliczanych z sumy kolumn, jest to trend malejący (im grupa widzów jest bardziej zainteresowana losami bohaterów serialu, tym mniejszą jej część stanowią nowi widzowie).

13.2.9 Relatywne Ryzyko i Iloraz Szans

Określenie Szansy lub Ryzyka wystąpienia badanego zjawiska na podstawie narażenia na czynnik mogący je wywoływać szacujemy na podstawie danych zebranych w tabeli kontyngencji 2×2 . Na przykład możemy sprawdzać jak palenie papierosów wpływa na zachorowanie:

	chorzy	zdrowi
palący	O_{11}	O_{12}
niepalący	O_{21}	O_{22}

Okno z możliwością wyznaczenia tych miar wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR wybierając OR/RR... lub poprzez [Kreator](#).



Jeśli przeprowadzone badanie jest badaniem typu **przypadek-kontrola**, wówczas dla takiej tabeli wylicza się **Iloraz Szans** na to, że badane zjawisko wystąpi. Zwykle są to badania retrospektywne – czyli takie, w których badacz sam decyduje o wielkości próby, w której badane zjawisko występuje i próby kontrolnej, wolnej od tego zjawiska.

Jeśli natomiast badanie jest badaniem **kohortowym**, wówczas dla takiej tabeli wylicza się **Relatywne Ryzyko** na to, że badane zjawisko wystąpi. Zwykle są to badania prospektywne – czyli takie, w których badacz dba o to, by warunki eksperymentu pozwalały na to by struktura występowania badanego zjawiska w próbie była zbliżona do struktury tego zjawiska w populacji.

Iloraz Szans (tabela 2×2)

Dla wyznaczenia Ilorazu Szans (*ang. Odds Ratio – OR*) wylicza się szansę bycia przypadkiem w grupie narażonej na czynnik ryzyka i szansę bycia przypadkiem w grupie nie narażonej na czynnik ryzyka zgodnie z wzorami:

$$szansa_{ryzyko=tak} = \frac{O_{11}/(O_{11} + O_{12})}{O_{21}/(O_{21} + O_{22})} = \frac{O_{11}}{O_{21}}$$

$$\text{szansa}_{\text{ryzyko=nie}} = \frac{O_{21}/(O_{21} + O_{22})}{O_{22}/(O_{21} + O_{22})} = \frac{O_{21}}{O_{22}}.$$

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

Test ten służy do weryfikacji hipotezy o tym, że szansa na wystąpienie badanego zjawiska jest taka sama w grupie narażonej i w grupie nienarażonej na czynnik ryzyka.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : OR &= 1, \\ \mathcal{H}_1 : OR &\neq 1. \end{aligned}$$

Statystyka testowa wyraża się wzorem:

$$z = \frac{\ln(OR)}{SE},$$

gdzie:

$$SE = \sqrt{\frac{1}{O_{11}} + \frac{1}{O_{12}} + \frac{1}{O_{21}} + \frac{1}{O_{22}}} - \text{błąd standardowy logarytmu Ilorazu Szans.}$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Uwaga!

Często w interpretacji istotności Ilorazu Szans posługujemy się wyznaczonym dla niego przedziałem ufności. Badamy wówczas, czy przedział ten zawiera wartość 1.

Relatywne Ryzyko (tabela 2×2)

Dla badania kohortowego można wyznaczać **Ryzyko** wystąpienia badanego zjawiska (ponieważ próbka powinna zbliżyć się strukturą zjawiska, które interesuje badacza do populacji, z której została pobrana) oraz wyliczać Relatywne Ryzyko (*ang. Relative Risk – RR*).

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

Test ten służy do weryfikacji hipotezy o tym, że ryzyko wystąpienia badanego zjawiska jest takie samo w grupie narażonej i w grupie nienarażonej na czynnik ryzyka.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : RR &= 1, \\ \mathcal{H}_1 : RR &\neq 1. \end{aligned}$$

Statystyka testowa wyraża się wzorem:

$$z = \frac{\ln(RR)}{SE},$$

gdzie:

$$SE = \sqrt{\frac{1}{O_{11}} - \frac{1}{O_{11}+O_{12}} + \frac{1}{O_{21}} - \frac{1}{O_{21}+O_{22}}} - \text{ błąd standardowy logarytmu Relatywnego Ryzyka.}$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Uwaga!

Często w interpretacji istotności Relatywnego Ryzyka posługujemy się wyznaczonym dla niego przedziałem ufności. Badamy wówczas, czy przedział ten zawiera wartość 1.

Uwaga!

Gdy w tabeli danych znajdują się zera, wówczas wyliczenie ilorazu szans lub relatywnego ryzyka może być niemożliwe. W takiej sytuacji, aby zagwarantować możliwość wyznaczenia ilorazu szans, w oknie analizy można zaznaczyć opcję Zamień 0 w tabeli na: . Wybór tej opcji (korekty na ciągłość) powoduje dodanie do wszystkich komórek tabeli zadanej wartości.

13.2.10 Test Z dla dwóch niezależnych proporcji

Test Z dla dwóch niezależnych proporcji stosujemy w podobnych sytuacjach jak test $\chi^2 (2 \times 2)$, tzn. gdy mamy 2 niezależne próby o licznosci n_1 i n_2 , w których możemy uzyskać 2 możliwe wyniki badanej cechy (jeden z nich to wynik wyróżniony o licznosci m_1 - w pierwszej próbie i m_2 - w drugiej próbie). Dla prób tych możemy również wyznaczyć wyróżnione proporcje $p_1 = \frac{m_1}{n_1}$ i $p_2 = \frac{m_2}{n_2}$. Test ten służy do weryfikacji hipotezy, że wyróżnione proporcje P_1 i P_2 w populacjach, z których pochodzą próby są sobie równe.

Podstawowe warunki stosowania:

- pomiar na **skali nominalnej** - ewentualne uporządkowanie kategorii nie jest brane pod uwagę
- **model niezależny**,
- duża licznosc.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : P_1 &= P_2, \\ \mathcal{H}_1 : P_1 &\neq P_2, \end{aligned}$$

gdzie:

P_1, P_2 frakcja dla pierwszej i drugiej populacji.

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 licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

W programie oprócz różnicy proporcji wyliczana jest wartość wskaźnika NNT i/lub NNH.

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. NNT wyliczane jest z wzoru:

$$NNT = \frac{1}{|p_1 - p_2|}$$

i jest cytowane wtedy, gdy różnica $p_1 - p_2$ jest dodatnia.

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. NNH wyliczane jest w ten sam sposób co NNT, ale jest cytowane wtedy, gdy różnica $p_1 - p_2$ jest ujemna.

Przedział ufności – im węższy przedział ufności, tym bardziej precyzyjne oszacowanie. Jeśli w przedziale ufności zawarte jest 0 dla różnicy proporcji, a ∞ dla NNT i/lub NNH, to jest wskazanie do tego, by dany wynik traktować jako nieistotny statystycznie.

Uwaga!

Przedziały ufności dla różnicy dwóch niezależnych proporcji od wersji PQStat 1.3.0 estymowane są w oparciu o metodę Newcomba-Wilsona (Bender (2001)[18], Newcombe (1998)[123], Wilson (1927)[173]). W poprzednich wersjach były estymowane w oparciu o metodę Walda.

Uzasadnienie zmiany:

Przedziały ufności oparte o klasyczną metodę Walda są odpowiednie dla dużych rozmiarów próbek i różnicy proporcji dalekiej od 0 lub 1. Dla małych próbek i różnicy proporcji bliskiej tym skrajnym wartościom, w wielu sytuacjach praktycznych, metoda Walda może prowadzić do wyników niewiarygodnych (Newcombe 1998[123], Miettinen 1985[120], Beal 1987[17], Wallenstein 1997[164]). Porównanie i przeanalizowanie wielu metod, które mogą być używane zamiast prostej metody Walda, można znaleźć w pracy Newcombe (1998)[123]. Sugerowaną, odpowiednią również dla skrajnych wartości proporcji, jest rozszerzona na przedziały dla różnicy dwóch niezależnych proporcji, metoda opublikowana po raz pierwszy przez Wilsona (1927)[173].

Uwaga!

Przedział ufności dla NNT i/lub NNH wyliczany jest jako odwrotność przedziału dla proporcji, zgodnie ze sposobem zaproponowanym przez Altmana (Altman (1998)[6]).

Okno z ustawieniami opcji testu Z dla dwóch niezależnych proporcji wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Z dla dwóch niezależnych proporcji.

Z dla dwóch niezależnych proporcji

Dane ograniczone przez zaznaczenie [1x4]

Analiza statystyczna : Test Z dla dwóch niezależnych proporcji

▼ Liczność (licznik) ▼ Liczność próby (mianownik) ▼ Liczność (licznik) Liczność próby (mianownik)

2-Var2	2-Var2	2-Var2	2-Var2
3-Var3	3-Var3	3-Var3	3-Var3
4-Var4	4-Var4	4-Var4	4-Var4
5-Var5	5-Var5	5-Var5	5-Var5

Filtr danych

Wybór danych spełniających odpowiednie warunki +

Reguły połączone są koniunkcją -

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane

Dołącz wykres

Zmienna zawiera

Liczność (licznik)

Proporcja

Korekcja na ciągłość

0,05 poziom istotności

PRZYKŁAD (13.7) c.d. (plik płęć-egzamin.pqs)

Wiemy, że $\frac{50}{90} = 55.56\%$ z wszystkich kobiet w próbie zdaje pozytywnie egzamin i $\frac{20}{80} = 25.00\%$ z wszystkich mężczyzn w próbie zdaje egzamin pozytywnie. 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

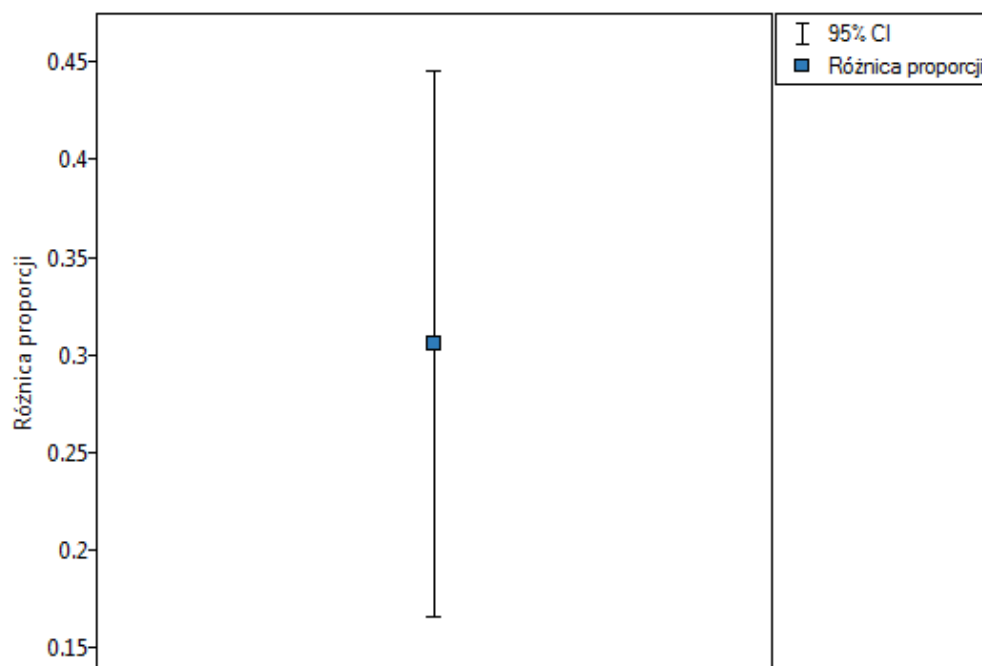
Hipotezy:

- \mathcal{H}_0 : proporcja mężczyzn, uzyskujących pozytywny wynik egzaminu jest taka sama jak kobiet uzyskujących pozytywny wynik egzaminu w badanej populacji,
- \mathcal{H}_1 : proporcja mężczyzn, uzyskujących pozytywny wynik egzaminu jest inna niż proporcja kobiet uzyskujących pozytywny wynik egzaminu w badanej populacji.

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

**Uwaga!**

Ponieważ w arkuszu danych znajduje się więcej informacji, przed rozpoczęciem analizy należy zaznaczyć odpowiedni obszar (dane bez nagłówek). W oknie testu natomiast wybrać opcję mówiącą o zawartości zmiennej (liczność (licznik) lub proporcja). Różnica proporcji wyróżnionych w próbie to 30.56%, a 95% przedział ufności dla niej (15.90%, 43.35%) nie zawiera 0.

Na podstawie testu Z bez poprawki na ciągłość ($p=0.000053$) jak też z poprawką na ciągłość ($p=0.0001$), na poziomie istotności $\alpha=0.05$ (podobnie jak w przypadku testu dokładnego Fishera, jego poprawki mid-p, testu χ^2 i testu χ^2 z poprawką Yatesa) przyjmujemy hipotezę alternatywną. Zatem proporcja mężczyzn, uzyskujących pozytywny wynik egzaminu jest inna niż proporcja kobiet uzyskujących pozytywny wynik egzaminu w badanej populacji. Istotnie częściej ten egzamin zdają kobiety ($\frac{50}{90} = 55.56\%$ z wszystkich kobiet w próbie zdało egzamin) niż mężczyźni ($\frac{20}{80} = 25.00\%$ z wszystkich mężczyzn w próbie zdało egzamin).

PRZYKŁAD 13.10.

Założmy, że choroba ma śmiertelność 100% bez leczenia, a terapia zmniejsza śmiertelność do 50% –

jest to wynik 20 letnich badań. Chcemy wiedzieć jak wiele osób będzie trzeba leczyć, aby zapobiec w ciągu 20 lat 1 śmierci. By odpowiedzieć na to pytanie pobrano dwie 100 osobowe próby z populacji osób chorych. W próbie nieleczonych mamy 100 chorych pacjentów, wiemy, że bez leczenia wszyscy oni umrą. W próbie leczonych mamy również 100 pacjentów, z których 50 przeżyje.

Chorzy – nie leczeni		Chorzy – leczeni	
liczność licznik	liczność próby (mianownik)	liczność licznik	liczność próby (mianownik)
100	100	50	100

Wyliczymy wskaźnik NNT.

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

Różnica pomiędzy proporcjami jest istotna statystycznie ($p < 0.000001$), ale nas interesuje wskaźnik NNT – wynosi on 2, czyli stosowanie leczenia u 2 chorych przez 20 lat zapobiegnie 1 śmierci. Wyliczony 95% przedział ufności należy zaokrąglić do wartości całkowitych, co daje NNT od 2 do 3 chorych.

PRZYKŁAD 13.11. Wartość pewnej różnicy proporcji w badaniu porównującym skuteczność leku 1 vs lek 2 wynosiła różnica (95%CI)=-0.08 (-0.27 do 0.11). Ta ujemna różnica proporcji sugeruje, że lek 1 był mniej skuteczny niż lek 2, jego zastosowanie naraziło więc chorych na ryzyko. Ponieważ różnica proporcji jest ujemna, to wyznaczoną odwrotność nazywamy NNH, a ponieważ przedział ufności zawiera nieskończoność NNH(95%CI)= 2.5 (NNH 3.7 to ∞ to NNT 9.1) i przechodzi z NNH do NNT, należy uznać że uzyskany wynik nie jest istotny statystycznie (Altman (1998)[6]).

13.2.11 Test McNemara, test wewnętrznej symetrii Bowkera

Podstawowe warunki stosowania:

- pomiar na **skali nominalnej** - ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- **model zależny**.

Test McNemara

Test McNemara (*ang. McNemar test*), McNemar (1947)[117]. Test ten służy do weryfikacji hipotezy o zgodności pomiędzy wynikami dwukrotnych pomiarów $X^{(1)}$ i $X^{(2)}$ cechy X (pomiędzy dwiema zmiennymi zależnymi $X^{(1)}$ i $X^{(2)}$). Badana cecha może mieć tylko 2 kategorie (oznaczone przez nas (+) i (-)). Test McNemara można wyliczać na podstawie **danych surowych** albo z wykonanej na podstawie danych surowych **tabeli kontyngencji** o wymiarach 2×2 .

Tabela 13.2. Tabela kontyngencji 2×2 dla licznosci obserwowanych zmiennych zależnych

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	$O_{11} + O_{21}$	$O_{12} + O_{22}$	$n = O_{11} + O_{12} + O_{21} + O_{22}$

Hipotezy:

$$\mathcal{H}_0 : O_{12} = O_{21},$$

$$\mathcal{H}_1 : O_{12} \neq O_{21}.$$

Statystyka testowa ma postać:

$$\chi^2 = \frac{(O_{12} - O_{21})^2}{O_{12} + O_{21}}.$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) rozkład χ^2 z jednym stopniem swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\text{jeżeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1,$$

$$\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0.$$

Poprawka na ciągłość testu McNemara

Poprawka ta jest testem bardziej konserwatywny od testu McNemara (trudniej niż test McNemara odrzuca hipotezę zerową). Ma ona zapewnić możliwość przyjmowania przez statystykę testową wszystkich wartości liczb rzeczywistych zgodnie z założeniem rozkładu χ^2 . Część źródeł podaje, że poprawkę na ciągłość powinno się wykonywać zawsze, natomiast część uznaje, że tylko wtedy, gdy licznosci w tabeli są małe.

Statystyka testowa testu McNemara z poprawką na ciągłość ma postać:

$$\chi^2 = \frac{(|O_{12} - O_{21}| - 1)^2}{O_{12} + O_{21}}.$$

Dokładny test McNemara

Powszechną, ogólną zasadą ważności asymptotycznego testu McNemara chi-kwadrat jest **warunek Ruffinich**, czyli to, że liczba niezgodnych par jest większa niż 10: $O_{12} + O_{21} \geq 10$ [143] gdy warunek ten nie jest spełniony, wówczas powinniśmy bazować na dokładnych wartościach prawdopodobieństwa tego testu [55] Dokładna wartość prawdopodobieństwa testu oparta jest o rozkład dwumianowy i jest testem konserwatywnym, dlatego obok dokładnej wartości testu McNemara podano również polecaną wartość dokładną mid-p testu McNemara.

Iloraz szans na zmianę wyniku

Jeśli przeprowadzone zostało 2 krotnie badanie tej samej cechy na tych samych obiektach – wówczas dla takiej tabeli wylicza się **iloraz szans** na zmianę wyniku (z (+) na (-) i odwrotnie).

Szansa zmiany wyniku z (+) na (-) wynosi O_{12} , a szansa zmiany wyniku z (-) na (+) wynosi O_{21} . Iloraz szans (*ang. odds ratio – OR*) to:

$$OR = \frac{O_{12}}{O_{21}}.$$

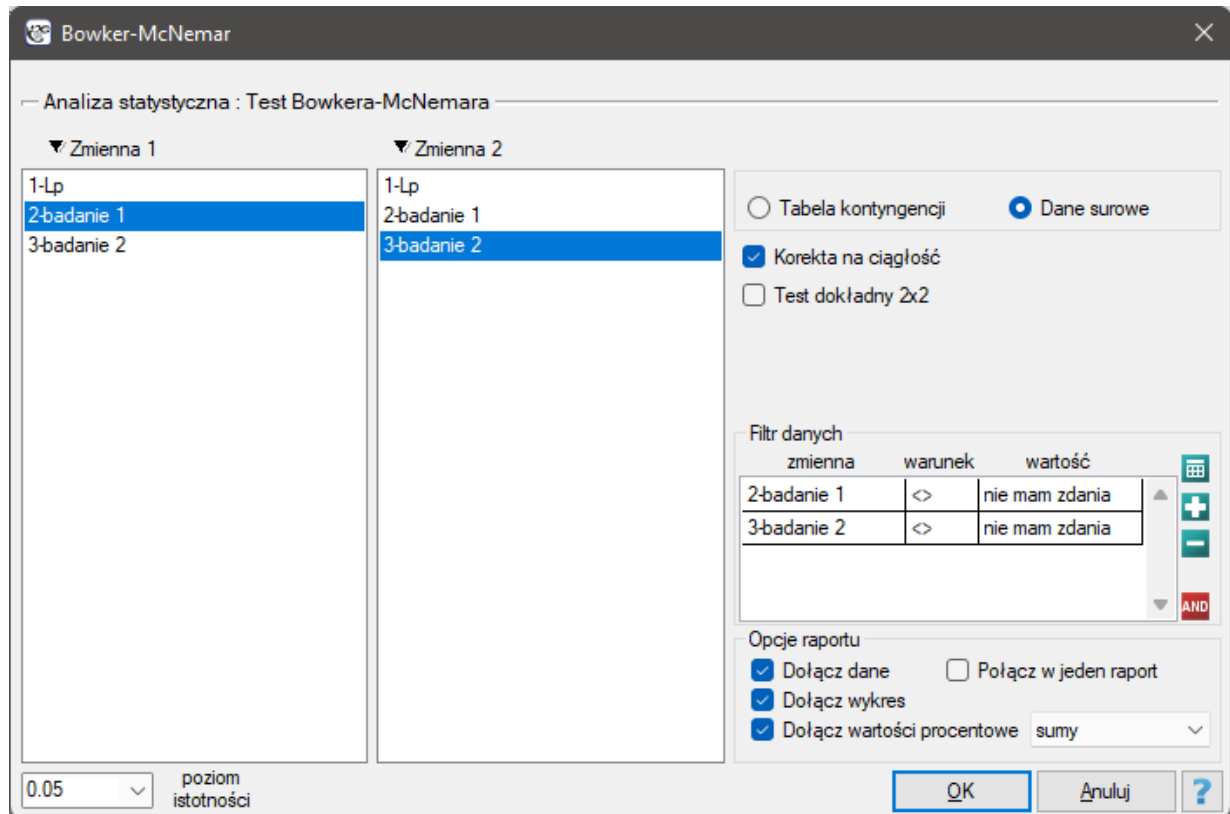
Przedział ufności dla ilorazu szans buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{1}{O_{12}} + \frac{1}{O_{21}}}.$$

Uwaga!

Dodatkowo, dla prób o niewielkich licznosciach, można wyznaczyć dokładny zakres przedziału ufności dla ilorazu Szans[101].

Okno z ustawieniami opcji testu Bowkera-McNemara wywołujemy poprzez menu Statystyka→Testy nieparametryczne→Bowker-McNemar lub poprzez [Kreator](#).

**Test wewnętrznej symetrii Bowkera**

Test wewnętrznej symetrii Bowkera (*ang. Bowker test of internal symmetry*), Bowker (1948)[23]. Test ten jest rozszerzeniem testu McNemara na 2 zmienne o więcej niż dwóch kategoriach ($c > 2$). Służy do weryfikacji hipotezy o symetryczności wyników dwukrotnych pomiarów $X^{(1)}$ i $X^{(2)}$ cechy X (symetryczności 2 zmiennych zależnych $X^{(1)}$ i $X^{(2)}$). Badana cecha może mieć więcej niż 2 kategorie. Test wewnętrznej symetrii Bowker można wyliczać na podstawie [danych surowych](#) albo z wykonanej na podstawie danych surowych [tabeli kontyngencji](#) o wymiarach $c \times c$.

Tabela 13.3. Tabela kontyngencji $c \times c$ dla licznosci obserwowanych zmiennych zależnych

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:

$$\mathcal{H}_0 : O_{ij} = O_{ji},$$

$$\mathcal{H}_1 : O_{ij} \neq O_{ji} \text{ dla przynajmniej jednej pary } O_{ij}, O_{ji},$$

gdzie $j \neq i, j \in 1, 2, \dots, c, i \in 1, 2, \dots, c$, zatem O_{ij} i O_{ji} to liczności symetrycznych par w tabeli $c \times c$

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 $df = \frac{c(c-1)}{2}$.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\text{jeżeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1,$$

$$\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0.$$

PRZYKŁAD 13.12. (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 dokonano 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.

	badanie 1	badanie 2
pozytywnie	negatywnie	negatywnie
negatywnie	nie mam zdania	negatywnie
pozytywnie	negatywnie	negatywnie
negatywnie	negatywnie	negatywnie

	negatywnie	pozytywnie	nie mam zdania
negatywnie	50	4	3
pozytywnie	44	54	5
nie mam zdania	35	18	37

Hipotezy:

\mathcal{H}_0 : liczba studentów którzy zmienili swoją opinię jest taka sama dla każdej z możliwych symetrycznych zmian opinii,

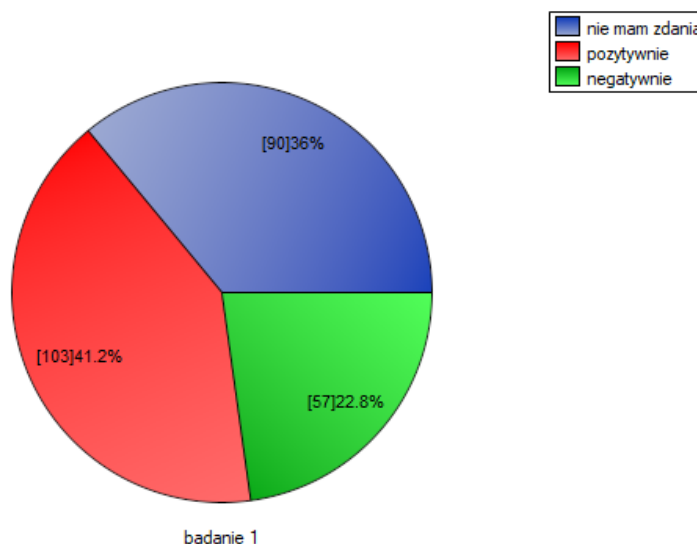
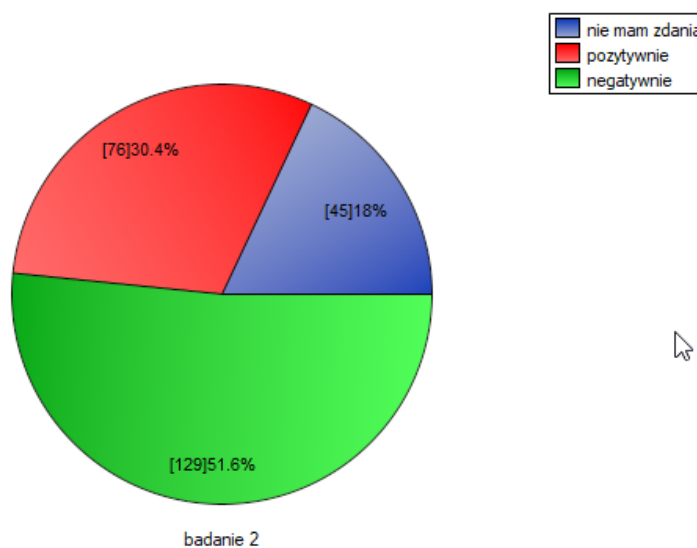
\mathcal{H}_1 : liczba studentów którzy zmienili swoją opinię jest inna dla przynajmniej jednej z możliwych (symetrycznych) zmian opinii,

gdzie np. zmiana opinii z pozytywnej na negatywną jest symetryczna względem zmiany opinii z negatywnej na pozytywną.

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 zdę	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 zdę	pozytywnie	negatywnie	
nie mam zdania	14.8%	7.2%	14%	
pozytywnie	2%	21.6%	17.6%	
negatywnie	1.2%	1.6%	20%	



Porównując wartość p dla testu Bowkera $p < 0.0001$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że opinie studentów zmieniły się. Z tabeli wynika, że istotnie więcej było tych studentów, którzy zmienili swoją opinię na negatywną po egzaminie niż tych którzy zmienili ją na pozytywną, oraz wielu studentów oceniających przed egzaminem wykładowcę pozytywnie po egzaminie nie wyrażało już takiego zdania.

Gdybyśmy ograniczyli nasze badanie do osób mających zdefiniowany pogląd na temat wykładowcy (tzn. oceniają tylko pozytywnie lub negatywnie), to moglibyśmy wykorzystać test McNemara:

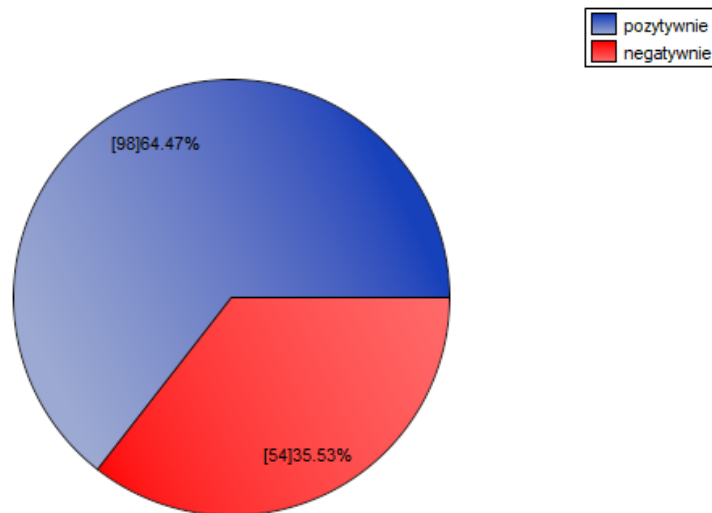
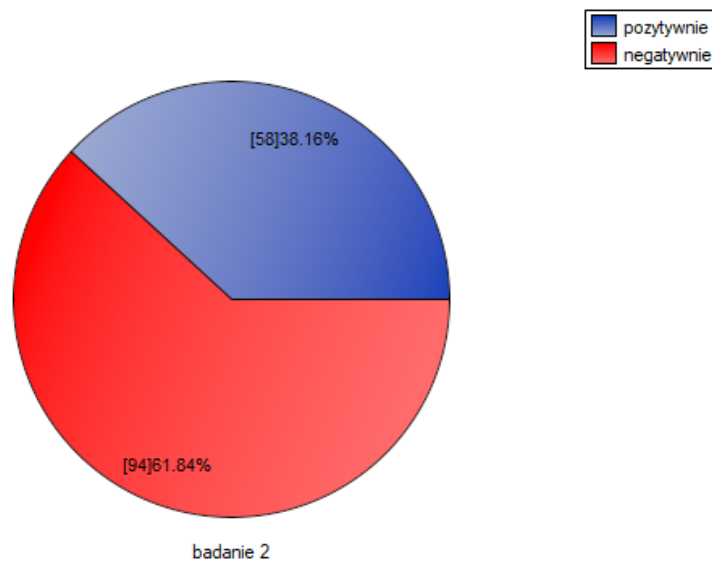
Hipotezy:

- \mathcal{H}_0 : liczba studentów którzy zmienili swoją opinię z negatywnej na pozytywną jest taka sama jak tych, którzy zmienili ją z pozytywnej na negatywną,
 \mathcal{H}_1 : liczba studentów którzy zmienili swoją opinię z negatywnej na pozytywną jest inna niż tych, którzy zmienili ją z pozytywnej na negatywną.

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 Ruffach	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%	



Porównując wartość p dla testu McNemara $p < 0.0001$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że opinie studentów zmieniły się. 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 $\left(\frac{44}{4}\right)$ razy większa niż z negatywnej na pozytywną (szansa zmiany opinii w przeciwną stronę to: $\left(\frac{4}{44}\right)$).

13.2.12 Test Z dla dwóch zależnych proporcji

Test Z dla dwóch zależnych proporcji stosujemy w podobnych sytuacjach jak test **Test McNemara**, tzn. gdy mamy 2 zależne grupy pomiarów ($X^{(1)}$ i $X^{(2)}$), w których możemy uzyskać 2 możliwe wyniki badanej cechy ((+)(-)).

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	$O_{11} + O_{21}$	$O_{12} + O_{22}$	$n = O_{11} + O_{12} + O_{21} + O_{22}$

Dla grup tych możemy również wyliczyć wyróżnione proporcje $p_1 = \frac{O_{11}+O_{12}}{n}$ i $p_2 = \frac{O_{11}+O_{21}}{n}$. Test ten służy do weryfikacji hipotezy, że wyróżnione proporcje P_1 i P_2 w populacji, z której pochodzi próba są sobie równe.

Podstawowe warunki stosowania:

- pomiar na **skali nominalnej** - ewentualne uporządkowanie kategorii nie jest brane pod uwagę,
- **model zależny**,
- duża licznosc.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 : P_1 - P_2 &= 0, \\ \mathcal{H}_1 : P_1 - P_2 &\neq 0,\end{aligned}$$

gdzie:

P_1, P_2 frakcja dla pierwszego i drugiego pomiaru.

Statystyka testowa ma postać:

$$Z = \frac{p_1 - p_2}{\sqrt{O_{21} + O_{12}}} \cdot n,$$

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Uwaga!

Przedział ufności dla różnicy dwóch zależnych proporcji estymowany jest w oparciu o metodę Newcoma-Wilsona.

Okno z ustawieniami opcji testu Z dla dwóch zależnych proporcji wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Z dla dwóch zależnych proporcji.

Z dla dwóch zależnych proporcji

Analiza statystyczna : Test Z dla dwóch zależnych proporcji

▼ O[11] ▼ O[12] ▼ O[21] O[22]

1-Var1
2-Var2
3-Var3
4-Var4

1-Var1
2-Var2
3-Var3
4-Var4

1-Var1
2-Var2
3-Var3
4-Var4

1-Var1
2-Var2
3-Var3
4-Var4

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny **AND**

Opcje raportu

Dołącz dane

Dołącz wykres

0,05 poziom istotności

PRZYKŁAD (13.12) c.d. (plik opinia.pqs)

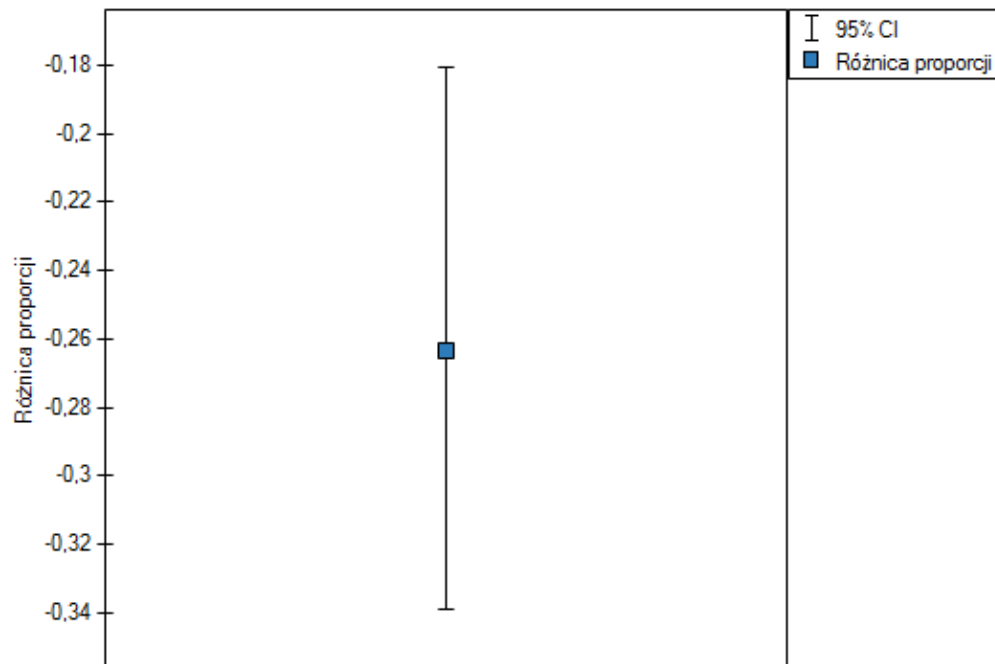
Gdy ograniczymy nasze badanie do osób mających zdefiniowany pogląd na temat wykładowcy (tzn. oceniają tylko pozytywnie lub negatywnie), to takich studentów uzyskamy 152. Dane do obliczeń to: $O_{11} = 50$, $O_{12} = 4$, $O_{21} = 44$, $O_{22} = 54$. Wiemy, że $\frac{50+4}{152} = 35.53\%$ studentów przed egzaminem wyrażało negatywną opinię. Po egzaminie odpowiedni procent wynosił $\frac{50+44}{152} = 61.84\%$.

Hipotezy:

- \mathcal{H}_0 : brak różnicy pomiędzy liczbą negatywnych ocen dotyczących wykładowcy przed egzaminem i po egzaminie,
- \mathcal{H}_1 : istnieje różnica pomiędzy liczbą negatywnych ocen dotyczących wykładowcy przed egzaminem i po egzaminie.

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

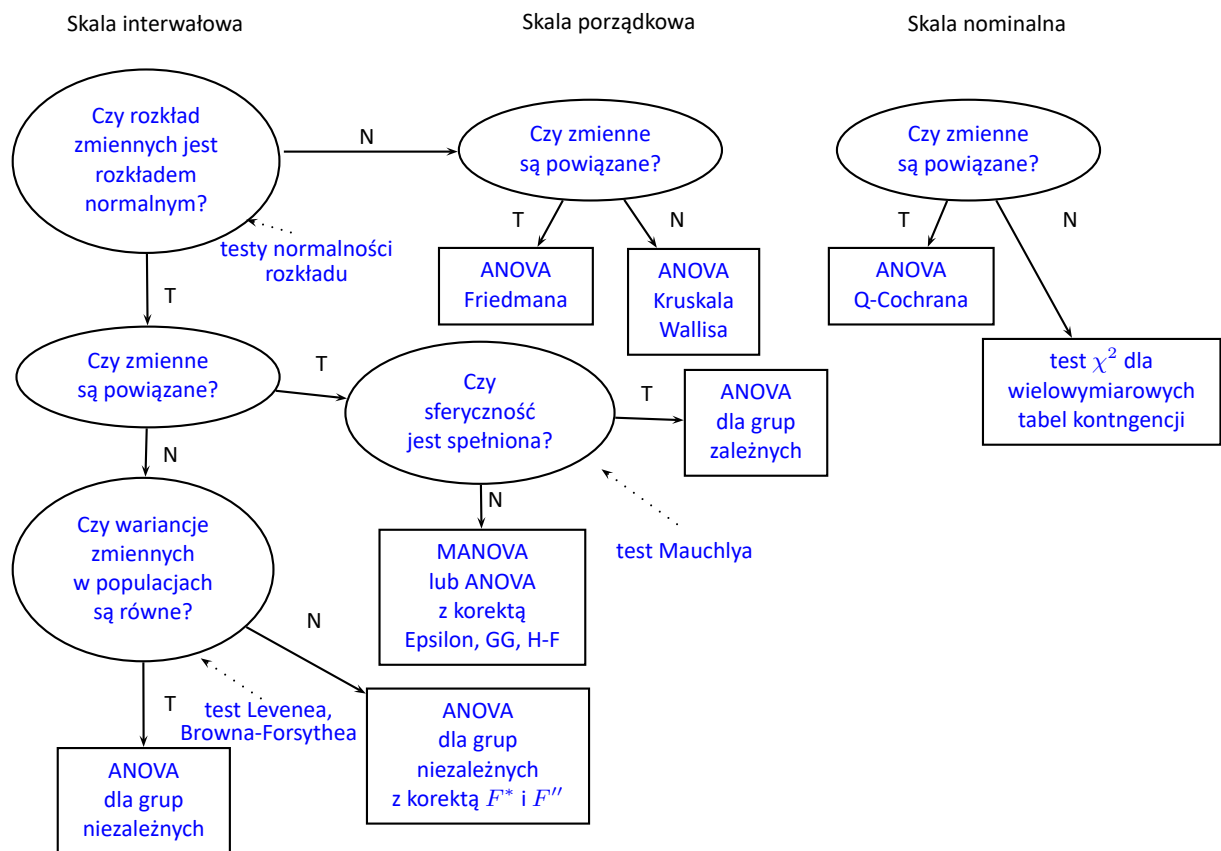
Dane :				
	v.1	v.2	v.3	v.4
	50	4	44	54



Różnica proporcji wyróżnionych w próbie to 26.32%, a 95% przedział ufności dla niej (18.07%, 33.88%) nie zawiera 0.

Na podstawie testu Z ($p=0.0001$), na poziomie istotności $\alpha=0.05$ (podobnie jak w przypadku testu McNemara) przyjmujemy hipotezę alternatywną. Zatem proporcja negatywnych ocen przed egzaminem jest inna niż proporcja negatywnych ocen po egzaminie. Po egzaminie istotnie częściej wykładowca jest oceniany negatywnie.

14 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 zwiększilibyśmy znacznie założony poziom α . Tego błędu unikniemy stosując odpowiednie testy ANOVA (*ang. Analysis of Variance*) oraz dedykowane im kontrasty, czy też testy POST-HOC.

14.1 TESTY PARAMETRYCZNE

14.1.1 ANOVA dla grup niezależnych

Jednoczynnikowa analiza wariancji (ANOVA) dla grup niezależnych (*ang. one-way analysis of variance*) zaproponowana przez Ronalda Fishera, służy do weryfikacji hipotezy o równości **średnich** badanej zmiennej w kilku ($k \geq 2$) populacjach.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- **normalność rozkładu** badanej zmiennej w każdej populacji,
- **model niezależny**,
- **równość wariancji** badanej zmiennej wszystkich populacji.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \mu_1 = \mu_2 = \dots = \mu_k, \\ \mathcal{H}_1 &: \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, \dots, k), \end{aligned}$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_k$ – średnie badanej zmiennej w populacjach, z których pobrano próby.

Statystyka testowa ma postać:

$$F = \frac{MS_{BG}}{MS_{WG}},$$

gdzie:

$$MS_{BG} = \frac{SS_{BG}}{df_{BG}} \text{ – średnia kwadratów między grupami,}$$

$$MS_{WG} = \frac{SS_{WG}}{df_{WG}} \text{ – średnia kwadratów wewnątrz grup,}$$

$$SS_{BG} = \sum_{j=1}^k \frac{(\sum_{i=1}^{n_j} x_{ij})^2}{n_j} - \frac{(\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij})^2}{N} \text{ – suma kwadratów między grupami,}$$

$$SS_{WG} = SS_T - SS_{BG} \text{ – suma kwadratów wewnątrz grup,}$$

$$SS_T = \left(\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}^2 \right) - \frac{(\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij})^2}{N} \text{ – całkowita suma kwadratów,}$$

$$df_{BG} = k - 1 \text{ – stopnie swobody (między grupami),}$$

$$df_{WG} = df_T - df_{BG} \text{ – stopnie swobody (wewnątrz grup),}$$

$$df_T = N - 1 \text{ – całkowite stopnie swobody,}$$

$$N = \sum_{j=1}^k n_j,$$

$$n_j \text{ – liczności prób dla } (j = 1, 2, \dots, k),$$

$$x_{ij} \text{ – wartości zmiennej w próbach dla } (i = 1, 2, \dots, n_j), (j = 1, 2, \dots, k).$$

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Wielkość efektu - cząstkowa η^2

Wielkość ta określa proporcję wariancji wyjaśnionej do wariancji całkowitej związanej z danym czynnikiem. Zatem w modelu jednoczynnikowej ANOVA dla grup niezależnych wskazuje jaka część wewnątrz-osobowej zmienności wyników może być przypisana badanemu czynnikowi wyznaczającemu grupy niezależne.

$$\eta^2 = \frac{SS_{BG}}{SS_{BG} + SS_{res}}$$

14.1.2 Kontrasty i testy POST-HOC

Analiza wariancji daje informację tylko o tym, czy między populacjami występują istotne statystycznie różnice. Nie mówi ona, które populacje różnią się między sobą. By uzyskać wiedzę o różnicach dla fragmentu naszego złożonego układu stosujemy **kontrasty** (gdy dokonujemy wcześniej zaplanowanych i zwykle tylko wybranych porównań - tzw. a priori), lub procedury porównań wielokrotnych czyli **testy POST-HOC** (gdy po wykonanej analizie wariancji szukamy różnic, zwykle pomiędzy wszystkimi parami).

Liczba wszystkich możliwych porównań prostych wyliczana jest z wzoru:

$$c = \binom{k}{2} = \frac{k(k-1)}{2}$$

Hipotezy:

Przykład 1 - **porównania proste** (porównanie pomiędzy sobą 2 wybranych średnich):

$$\begin{aligned} \mathcal{H}_0 : \mu_1 &= \mu_2, \\ \mathcal{H}_1 : \mu_1 &\neq \mu_2. \end{aligned}$$

Przykład 2 - **porównania złożone** (porównanie kombinacji wybranych średnich):

$$\begin{aligned} \mathcal{H}_0 : \mu_1 &= \frac{\mu_2 + \mu_3}{2}, \\ \mathcal{H}_1 : \mu_1 &\neq \frac{\mu_2 + \mu_3}{2}. \end{aligned}$$

By można było zdefiniować wybrane hipotezy należy dla każdej średniej przypisać wartość kontrastu c_j , ($j = 1, 2, \dots, k$). 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 wynosi 0.

Przykład 1: $c_1 = 1, c_2 = -1, c_3 = 0, \dots, c_k = 0$.

Przykład 2: $c_1 = 2, c_2 = -1, c_3 = -1, c_4 = 0, \dots, c_k = 0$.

Wyboru właściwej hipotezy możemy dokonać:

- (i) Porównując różnicę wybranych średnich z **najmniejszą istotną różnicą (NIR)** wyznaczoną odpowiednim testem POST-HOC :

jeżeli różnica średnich $\geq NIR \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli różnica średnich $< NIR \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

- (ii) Porównując wyznaczoną na podstawie **statystyki testowej** odpowiedniego testu POST-HOC **wartość p z α** :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ 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:

$$t = \frac{\sum_{j=1}^k c_j \bar{x}_j}{\sqrt{\left(\sum_{j=1}^k \frac{c_j^2}{n_j}\right) MS_{WG}}}.$$

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.

(ii) Statystyka testowa ma postać:

$$F = \frac{\left(\sum_{j=1}^k c_j \bar{x}_j\right)^2}{(k-1) \left(\sum_{j=1}^k \frac{c_j^2}{n_j}\right) MS_{WG}}.$$

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:

$$q = \sqrt{2} \frac{\sum_{j=1}^k c_j \bar{x}_j}{\sqrt{\left(\sum_{j=1}^k \frac{c_j^2}{n_j}\right) MS_{WG}}}.$$

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)[106]. Inne programy lub strony internetowe mogą wyliczać nieco inne wartości niż PQStat, gdyż mogą bazować na mniej precyzyjnych lub bardziej restrykcyjnych algorytmach (Copenhagen i Holland (1988), Gleason (1999)).

Test dla trendu.

Test badający istnienie trendu może być wyliczony 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.

Liczba grup	Trend	Kontrast									
		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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Okno z ustawieniami opcji jednoczynnikowej ANOVA dla grup niezależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → ANOVA dla grup niezależnych lub poprzez **Kreator**.

ANOVA dla grup niezależnych

Analiza statystyczna : Jednoczynnikowa ANOVA dla grup niezależnych

Zmienna: 1-wiek, 2-firma

Zmienna grupująca: 1-wiek, 2-firma

Opcje testu

- Użyj zmiennej grupującej
- POST-HOC: Fisher LSD
- Kontrasty
- Dla trendu: Fisher LSD

$\Sigma\mu$

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny

Opcje raportu

- Dołącz dane
- Dołącz wykres
- Połącz w jeden raport

0.05 poziom istotności

OK Anuluj ?

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, założmy, że podzieliłiśmy 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 wydechow. 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 :

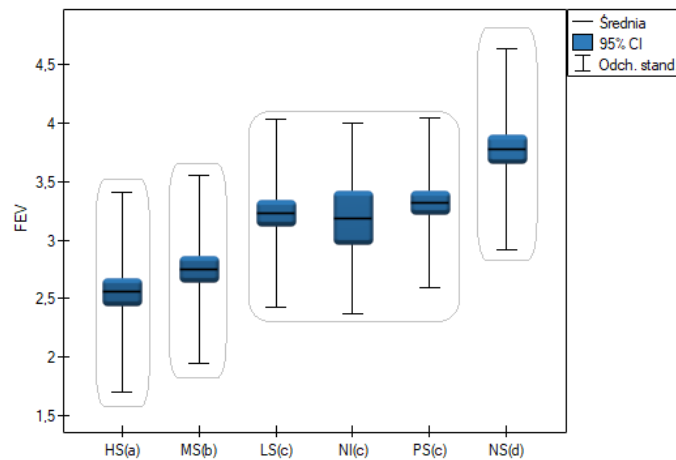
Statystyk	HS	MS	LS	NI	PS	NS
HS		2,434752	8,276928	4,920204	9,41745	15,066567
MS	2,434752		5,842176	3,380331	6,982697	12,631814
LS	8,276928	5,842176		0,314585	1,140522	6,789639
NI	4,920204	3,380331	0,314585		1,035914	4,60873
PS	9,41745	6,982697	1,140522	1,035914		5,649117
NS	15,066567	12,631814	6,789639	4,60873	5,649117	
Wartość p	HS	MS	LS	NI	PS	NS

uzyskujemy podział na grupy jednorodne:

Jednorodni	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 przeprowadzenia 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 go grupy o najniższej średniej ważonej średniej ważonej, a dalsze litery alfabetu kolejno do grup o coraz wyższych średnich.



PRZYKŁAD 14.1. (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 zależy 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 założenie równości wariancji, uzyskując w obydwu testach $p > 0.05$.

Hipotezy:

\mathcal{H}_0 : średnia wieku dla pracowników wszystkich badanych firm przewozowych jest taka sama,

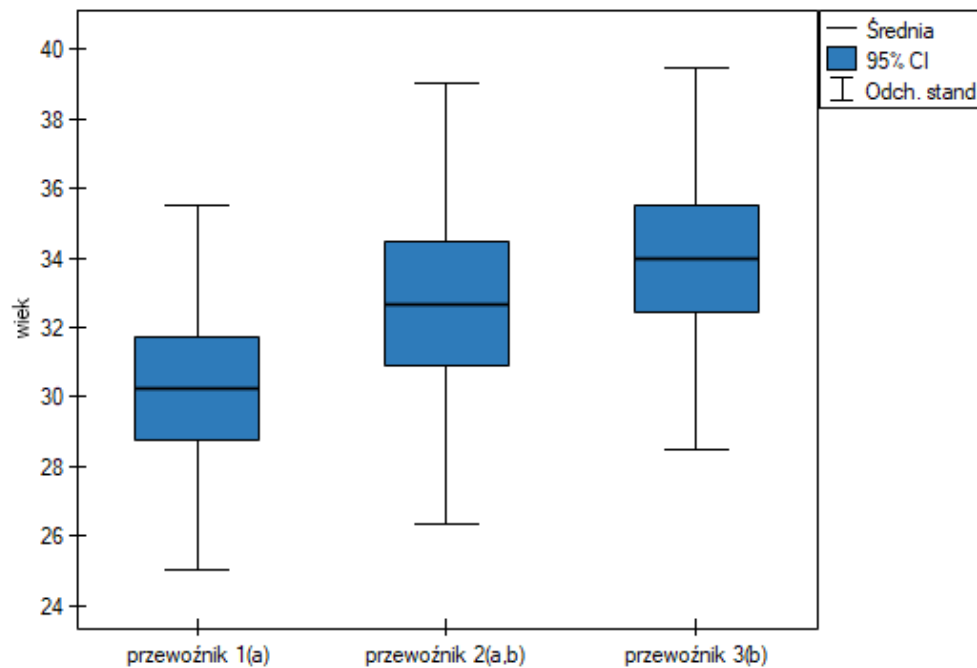
\mathcal{H}_1 : przynajmniej 2 średnie są różne.

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

Porównując wartość $p = 0.0051$ jednoczynnikowej analizy wariancji z poziomem istotności $\alpha = 0.05$ stwierdzamy, że średni wiek pracowników tych firm przewozowych jest różny. Na podstawie wyniku samej ANOVA nie możemy odpowiedzieć sobie na pytanie, które grupy różnią się pod względem wieku. By uzyskać taką wiedzę wykorzystany zostanie jeden z testów POST-HOC, np. **test Tukeya**. W tym celu wznawiamy analizę przyciskiem  i w oknie opcji testu wybieramy Tukey HSD oraz dołączmy wykres.

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		*	*

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.



Dokładny opis danych możemy przedstawić wybierając w oknie analizy statystyki opisowe Σ_{μ}

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

14.1.3 ANOVA dla grup niezależnych z korektą F^* i F''

Poprawki F^* (Brown-Forsythe, 1974[32]) oraz F'' (Welch, 1951[167]) dotyczą ANOVA dla grup niezależnych i są wyliczane wówczas, gdy nie jest spełnione założenie równości wariancji.

Statystyka testowa ma postać:

$$F^* = \frac{SS_{BG}}{\sum_{j=1}^k \left(1 - \frac{n_j}{n} sd_j^2\right)},$$

$$F'' = \frac{\frac{\sum_{j=1}^k w_j (\bar{x}_j - \tilde{x})}{k-1}}{1 + \frac{2(k-2)}{k^2-1} \sum_{j=1}^k h_j},$$

gdzie:

sd_j – odchylenie standardowe grupy j ,

$w_j = \frac{n_j}{sd_j^2}$ – waga grupy j ,

\tilde{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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Testy POST-HOC

Wprowadzenie do **kontrastów i testów POST-HOC** przeprowadzone zostało w rozdziale 14.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test T2 Tamhane

Dla porównań prostych i złożonych zarówno równolicznych jak i różnolicznych grup, gdy wariancje różnią się istotnie (Tamhane A. C., 1977[156]).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha_{Sidak}, 1, df_v}} \cdot \sqrt{\left(\sum_{j=1}^k \frac{c_j^2 s d_j^2}{n_j} \right)},$$

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,

$$\alpha_{Sidak} = 1 - (1 - \alpha)^{(1/k)},$$

$$df_v = \frac{\left(\sum_{j=1}^k \frac{c_j^2 s d_j^2}{n_j} \right)^2}{\sum_{j=1}^k \frac{c_j^4 s d_j^4}{n_j^2 (n_j - 1)}}$$

(ii) Statystyka testowa ma postać:

$$t = \frac{\left(\sum_{j=1}^k c_j \bar{x}_j \right)^2}{\sqrt{\left(\sum_{j=1}^k \frac{c_j^2 s d_j^2}{n_j} \right)}}$$

Statystyka ta podlega **rozkładowi t-Studenta** z df_v stopniami swobody, a **wartość p** jest korygowana o **liczbę możliwych porównań prostych**.

Test BF (Brown-Forsythe)

Dla porównań prostych zarówno równolicznych jak i różnolicznych grup, gdy wariancje różnią się istotnie (Brown M. B. i Forsythe A. B. (1974)[31]).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \sqrt{F_{\alpha, k-1, df_v}} \cdot \sqrt{(k-1) \left(\sum_{j=1}^k \frac{c_j^2 s d_j^2}{n_j} \right)},$$

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:

$$F = \frac{\left(\sum_{j=1}^k c_j \bar{x}_j \right)^2}{(k-1) \left(\sum_{j=1}^k \frac{c_j^2 s d_j^2}{n_j} \right)}$$

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, gdy wariancje różnią się istotnie (Games P. A. i Howell J. F. 1976[66]).

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = \frac{q_{\alpha, k, df_v} \cdot \sqrt{\left(\sum_{j=1}^k \frac{c_j^2 s_d^2}{n_j} \right)}}{\sqrt{2}},$$

gdzie:

q_{α, k, df_v} - to **wartość krytyczna** (statystyka) rozkładu studentyzowanego rozstępu dla danego poziomu istotności α oraz k i df_v stopni swobody.

(ii) Statystyka testowa ma postać:

$$q = \sqrt{2} \frac{\sum_{j=1}^k c_j \bar{x}_j}{\sqrt{\left(\sum_{j=1}^k \frac{c_j^2 s_d^2}{n_j} \right)}}.$$

Statystyka ta podlega rozkładowi studentyzowanego rozstępu z k i df_v stopniami swobody.

Test dla trendu.

Test badający istnienie trendu może być wyliczony 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**). Budując odpowiednie kontrasty można badać dowolny rodzaj trendu np. liniowy, kwadratowy, sześcienny, itd. Tabela przykładowych wartości kontrastów dla wybranych trendów znajduje się w opisie **testu dla trendu** dla ANOVA bez korekty dla różnych wariancji.

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 **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 na podstawie **statystyki testowej** wartość p porównujemy z α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Okno z ustawieniami opcji jednoczynnikowej ANOVA dla grup niezależnych z korektą F^* i F'' wywołujemy poprzez menu Statystyka → Testy parametryczne → ANOVA dla grup niezależnych lub poprzez **Kreator**.

ANOVA dla grup niezależnych

Analiza statystyczna : Jednoczynnikowa ANOVA dla grup niezależnych

Zmienna: 1-czas, 2-wykształcenie

Zmienna grupująca: 1-czas, 2-wykształcenie

Opcje testu

- Użyj zmiennej grupującej
- POST-HOC Tamhane's T2
- Kontrasty Dla trendu Tamhane's T2

podstawowe	1
zawodowe	2
średnie	3
wyższe	4

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny

Opcje raportu

- Dołącz dane
- Dołącz wykres
- Połącz w jeden raport

0.05 poziom istotności

OK Anuluj ?

PRZYKŁAD 14.2. (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.

Hipotezy:

- H_0 : średni czas poszukiwania pracy jest taki sam dla każdej kategorii wykształcenia,
- H_1 : przynajmniej jedna kategoria wykształcenia (jedna populacja) charakteryzuje się innym średnim czasem poszukiwania pracy.

Ze względu na różnice dotyczące wariancji pomiędzy poszczególnymi populacjami (dla testu Levene wartość $p = 0.0001$, a dla testu Brown-Forsythe 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

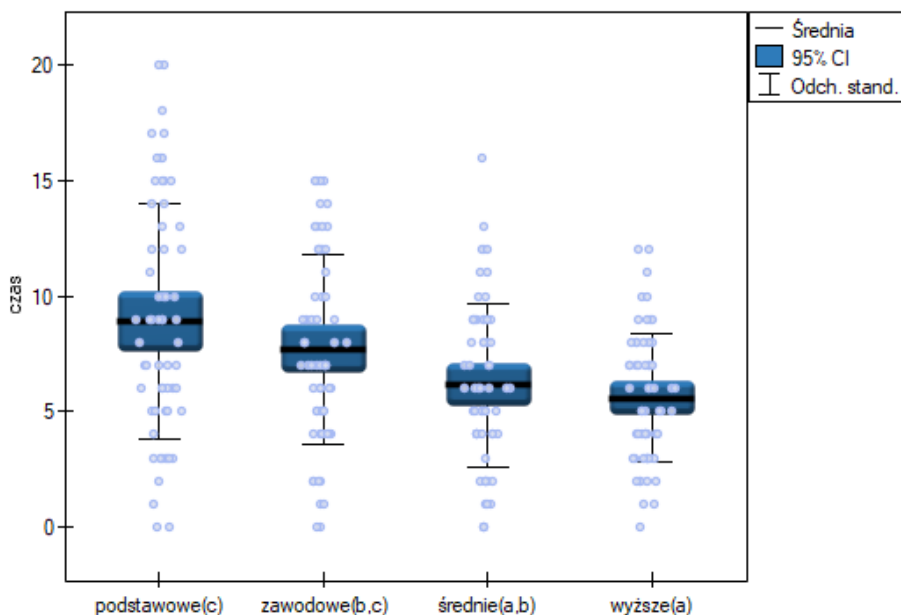
analizę przeprowadzamy przy włączonej korekcie różnych wariancji. Uzyskany wynik skorygowanej statystyki F jest przedstawiony poniżej.


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

Porównując wartość $p < 0.0001$ (dla testu F^*) oraz wartość $p < 0.0001$ (dla testu F'') z poziomem istotności $\alpha = 0.05$ stwierdzamy, że średni czas poszukiwania pracy różni się w zależności od posiadanego wykształcenia. Wykonując jeden z testów POST-HOC, dedykowany porównaniu grup o różnych wariancjach, dowiadujemy się których kategorii wykształcenia dotyczą stwierdzone różnice:

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 czym bardziej wykształcona osoba, tym mniej czasu zajmuje jej poszukiwanie pracy.



By móc sprawdzić tak postawioną hipotezę, należy podjąć analizę dla trendu. W tym celu wznawiamy analizę przyciskiem  i w oknie opcji testu wybieramy: metodę Tamhane's T2, opcję Kontrasy (i ustawiamy odpowiedni kontrast) lub opcję Dla trendu (i wskazujemy kolejność kategorii wykształcenia podając kolejne liczby naturalne).

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 istnieje rzeczywiście w populacji z której pochodzi próba.

14.1.4 Test Browna-Forsythea i test Levene'a

Obydwa testy: **test Levene'a** (*ang. Levene test*), Levene (1960) [100] i **test Browna-Forsythea** (*ang. Brown-Forsythe test*), Brown i Forsythe (1974) [30], służą do weryfikacji hipotezy o równości **wariancji** badanej zmiennej w kilku ($k \geq 2$) populacjach.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- **normalność rozkładu** badanej zmiennej w każdej populacji,
- **model niezależny**.

Hipotezy:

$$\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \sigma_j^2 \text{ są sobie równe } (j = 1, 2, \dots, k),$$

gdzie:

$\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ – wariacje badanej zmiennej w populacjach, z których pobrano próby.

Analiza polega na wyznaczeniu bezwzględnego odchylenia wyników pomiarowych od średniej (w teście Levene'a) lub od mediany (w teście Browna-Forsythea), w każdej z badanych grup. Owo bezwzględne odchylenie stanowi dane, które zostają poddane dokładnie tej samej procedurze, którą wykonuje się dla **analizy wariancji dla grup niezależnych**. Stąd statystyka testowa przyjmuje postać:

$$F = \frac{MS_{BG}}{MS_{WG}},$$

Statystyka ta podlega **rozkładowi F Snedecora** z df_{BG} i df_{WG} stopniami swobody.

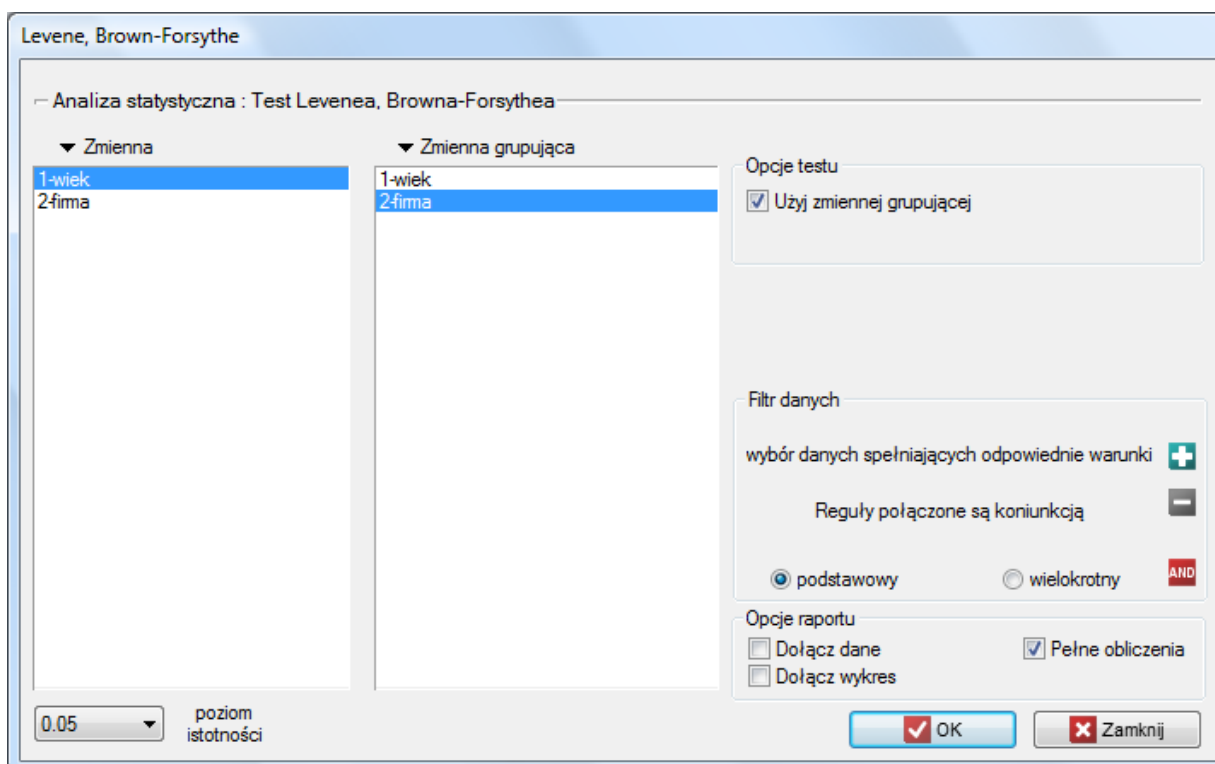
Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

Test Browna-Forsythea jest mniej wrażliwy niż test Levene'a na niespełnienie założenia dotyczącego normalności rozkładu.

Okno z ustawieniami opcji testu Levene'a, Browna-Forsythea wywołujemy poprzez menu Statystyka → Testy parametryczne → Levene, Brown-Forsythe.



14.1.5 ANOVA powtarzanych pomiarów

Jednoczynnikowa analiza wariancji dla powtarzanych pomiarów, czyli ANOVA dla grup zależnych (*ang. single-factor repeated-measures analysis of variance*) stosuje się w sytuacji, gdy pomiarów badanej zmiennej dokonujemy kilkakrotnie ($k \geq 2$) w różnych warunkach (przy czym zakładamy, że wariancje różnic pomiędzy wszystkimi parami pomiarów są sobie bliskie).

Test ten służy do weryfikacji hipotezy o równości **średnich** badanej zmiennej w kilku ($k \geq 2$) populacjach.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- **normalność rozkładu** dla wszystkich zmiennych będących różnicą par pomiarowych (lub normalność badanej zmiennej dla każdego pomiaru),
- **model zależny**.

Hipotezy:

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \dots = \mu_k,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, \dots, k),$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_k$ – średnie badanej zmiennej w kolejnych pomiarach z badanej populacji.

Statystyka testowa ma postać:

$$F = \frac{MS_{BC}}{MS_{res}}$$

gdzie:

$$MS_{BC} = \frac{SS_{BC}}{df_{BC}} - \text{średnia kwadratów między pomiarami,}$$

$$MS_{res} = \frac{SS_{res}}{df_{res}} - \text{średnia kwadratów dla reszt,}$$

$$SS_{BC} = \sum_{j=1}^k \left(\frac{(\sum_{i=1}^n x_{ij})^2}{n} \right) - \frac{(\sum_{j=1}^k \sum_{i=1}^n x_{ij})^2}{N} - \text{suma kwadratów między pomiarami,}$$

$$SS_{res} = SS_T - SS_{BS} - SS_{BC} - \text{suma kwadratów dla reszt,}$$

$$SS_T = \left(\sum_{j=1}^k \sum_{i=1}^n x_{ij}^2 \right) - \frac{(\sum_{j=1}^k \sum_{i=1}^n x_{ij})^2}{N} - \text{całkowita suma kwadratów,}$$

$$SS_{BS} = \sum_{i=1}^n \left(\frac{(\sum_{j=1}^k x_{ij})^2}{k} \right) - \frac{(\sum_{j=1}^k \sum_{i=1}^n x_{ij})^2}{N} - \text{suma kwadratów między obiektami,}$$

$$df_{BC} = k - 1 - \text{stopnie swobody (między pomiarami),}$$

$$df_{res} = df_T - df_{BC} - df_{BS} - \text{stopnie swobody (dla reszt),}$$

$$df_T = N - 1 - \text{całkowite stopnie swobody,}$$

$$df_{BS} = n - 1 - \text{stopnie swobody (między obiektami),}$$

$$N = nk,$$

n – liczność próby,

x_{ij} – wartości zmiennej dla i obiektów ($i = 1, 2, \dots, n$) w j pomiarach ($j = 1, 2, \dots, k$).

Statystyka ta podlega [rozkładowi F Snedecora](#) z df_{BC} i df_{res} stopniami swobody.

Wyznaczoną na podstawie [statystyki testowej wartości \$p\$](#) porównujemy z α :

$$\text{jeżeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1,$$

$$\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0.$$

Wielkość efektu - cząstkowa η^2

Wielkość ta określa proporcję wariancji wyjaśnionej do wariancji całkowitej związanej z danym czynnikiem. Zatem w modelu powtarzanych pomiarów wskazuje jaka część wewnątrzsobowej zmienności wyników może być przypisana powtarzanim pomiarom zmiennej.

$$\eta^2 = \frac{SS_{BC}}{SS_{BC} + SS_{res}}$$

Testy POST-HOC

Wprowadzenie do **kontrastów i testów POST-HOC** przeprowadzone zostało w rozdziale 14.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test LSD Fishera

Dla porównań prostych i złożonych (liczność w poszczególnych pomiarach zawsze jest taka sama).

Hipotezy:

Przykład - **porównania proste** (porównanie pomiędzy sobą 2 wybranych średnich):

$$\mathcal{H}_0 : \mu_j = \mu_{j+1},$$

$$\mathcal{H}_1 : \mu_j \neq \mu_{j+1}.$$

(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:

$$t = \frac{\sum_{j=1}^k c_j \bar{x}_j}{\sqrt{\left(\sum_{j=1}^k \frac{c_j^2}{n}\right) MS_{res}}}.$$

Statystyka ta podlega **rozkładowi t-Studenta** z df_{res} stopniami swobody.

Uwaga!

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 (liczność w poszczególnych pomiarach zawsze jest taka sama).

(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:

$$F = \frac{\left(\sum_{j=1}^k c_j \bar{x}_j\right)^2}{(k-1) \left(\sum_{j=1}^k \frac{c_j^2}{n}\right) MS_{res}}$$

Statystyka ta podlega rozkładowi F Snedecora z df_{BC} i df_{ref} stopniami swobody.

Test Tukeya.

Dla porównań prostych (liczność w poszczególnych pomiarach zawsze jest taka sama).

(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 danego poziomu istotności α oraz df_{res} i k stopni swobody.

(ii) Statystyka testowa ma postać:

$$q = \sqrt{2} \frac{\sum_{j=1}^k c_j \bar{x}_j}{\sqrt{\left(\sum_{j=1}^k \frac{c_j^2}{n}\right) MS_{res}}}$$

Statystyka ta podlega rozkładowi studentyzowanego rozstępu df_{res} i k stopniami swobody.

Info.

Algorytm obliczania wartości p i statystyki rozkładu studentyzowanego rozstępu w PQStat bazuje na pracy Lunda (1983)[106]. Inne programy lub strony internetowe mogą wyliczać nieco inne wartości niż PQStat, gdyż mogą bazować na mniej precyzyjnych lub bardziej restrykcyjnych algorytmach (Copenhagen i Holland (1988), Gleason (1999)).

Test dla trendu.

Test badający istnienie trendu może być wyliczony 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 Test LSD Fishera. Budując odpowiednie kontrasty można badać dowolny rodzaj trendu np. liniowy, kwadratowy, sześcienny, itd. Tabela przykładowych wartości kontrastów dla wybranych trendów znajduje się w opisie testu dla trendu dla ANOVA zmiennych niezależnych.

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.

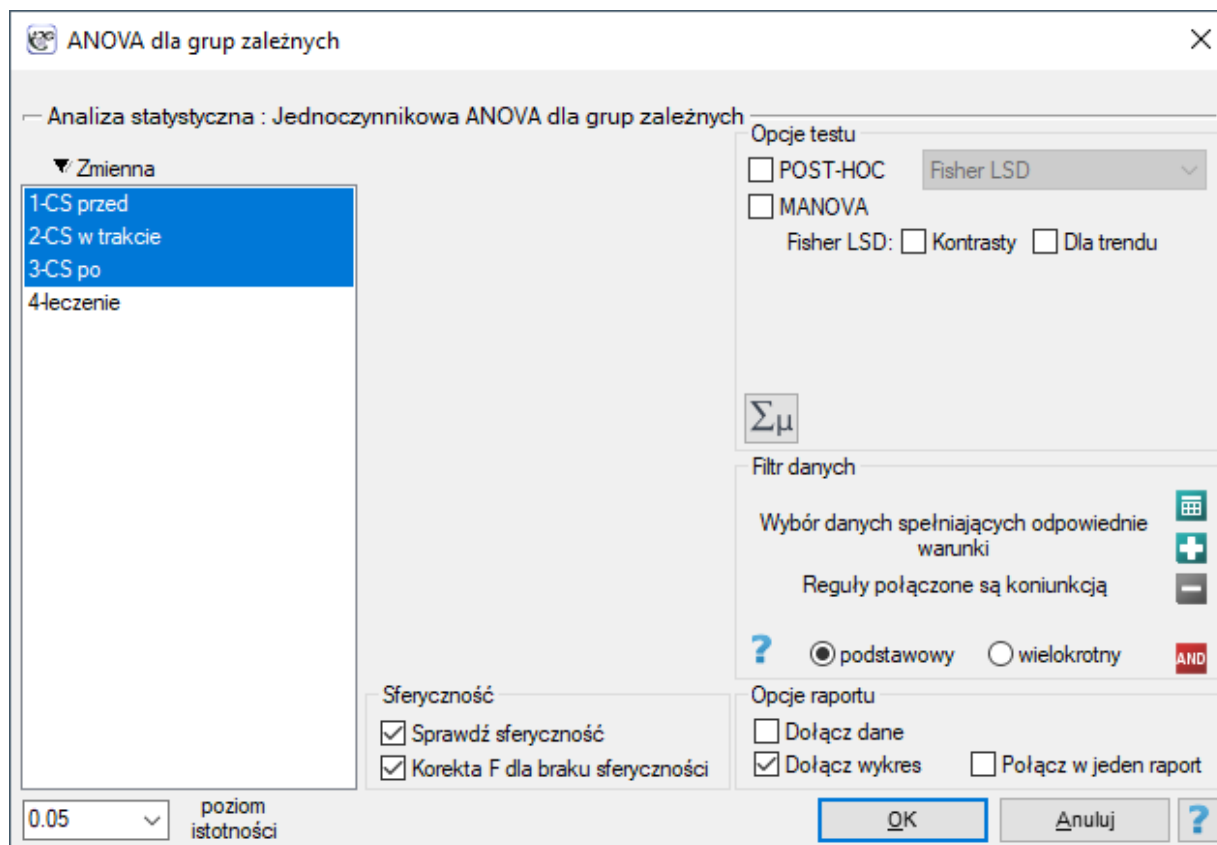
Analiza przeprowadzana jest w oparciu o kontrast liniowy, czyli wskazanym według naturalnego uporządkowania grupom przypisane są odpowiednie wartości kontrastu i wyliczona zostaje statystyka Test LSD Fishera.

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji jednoczynnikowej ANOVA dla grup zależnych wywołujemy poprzez menu Statystyka \rightarrow Testy parametryczne \rightarrow ANOVA dla grup zależnych lub poprzez [Kreator](#).



PATRZ PRZYKŁAD (14.3)

14.1.6 Sferyczność Mauchly'a

Założenie sferyczności jest podobne ale silniejsze niż założenie równości wariancji. Jest ono spełnione, jeśli wariancje dla różnic pomiędzy parami powtarzanych pomiarów są takie same. Zwykle w zastępstwie założenia o sferyczności rozważa się prostszy, ale bardziej rygorystyczny warunek symetrii połączonej (ang. *compound symmetry*). Można tak postąpić ponieważ spełnienie warunku symetrii połączonej pociąga za sobą spełnienie założenia sferyczności.

Warunek symetrii połączonej zakłada, symetrię w macierzy kowariancji, a zatem równość wariancji zmiennych (elementów głównej przekątnej macierzy kowariancji) oraz równość kowariancji (elementów poza główną przekątną macierzy kowariancji).

Naruszenie założenia sferyczności lub symetrii połączonej zmniejsza w nieuzasadniony sposób konserwatywność testu F (ułatwia odrzucenie hipotezy zerowej).

Dla sprawdzenia założenia sferyczności używa się testu Mauchly'a (1940)[116]. Istotność wyniku ($p \leq \alpha$)

oznacza tu naruszenie założenia sferyczności.

Podstawowe warunki stosowania:

- pomiar na [skali interwałowej](#),
- wielowymiarowy rozkład normalny lub [normalność rozkładu](#) każdej badanej zmiennej,
- [model zależny](#).

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \sigma_{diff(1)} = \sigma_{diff(2)} = \dots = \sigma_{diff(s)}, \\ \mathcal{H}_1 &: \text{nie wszystkie } \sigma_{diff(i)} \text{ są sobie równe } (i = 1, 2, \dots, s), \end{aligned}$$

gdzie:

$\sigma_{diff(i)}$ - populacyjna wariancja różnic pomiędzy i -tą parą powtarzanych pomiarów,
 s - liczba par.

Wartość W Mauchly'a 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.

Statystyka ta ma asymptotycznie (dla dużych licznosci) [rozkład \$\chi^2\$](#) z $df = \frac{k(k-1)}{2} - 1$ stopniami swobody.

Wyznaczoną na podstawie [statystyki testowej wartości \$p\$](#) porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

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.

PATRZ PRZYKŁAD (14.3) Poprawki Epsilon i MANOVA dotyczą [ANOVA powtarzanych pomiarów](#) i są wyliczone wówczas, gdy nie jest spełnione założenie sferyczności lub wariancje różnic pomiędzy wszystkimi parami pomiarów nie są sobie bliskie.

14.1.7 ANOVA powtarzanych pomiarów z korektą Epsilon i MANOVA

Korekta braku sferyczności

Stopień w jakim sferyczność jest spełniona jest reprezentowany przez wartość W w teście Mauchly'a, ale również przez wartości **Epsilon** (ε) wyliczone wraz z korektami. $\varepsilon = 1$ oznacza ściśle przestrzeganie warunku sferyczności. Im mniejsza od 1 jest wartość Epsilon, tym założenie sferyczności jest bardziej naruszone. Dolna granica jaką Epsilon może osiągnąć to $\frac{1}{k-1}$.

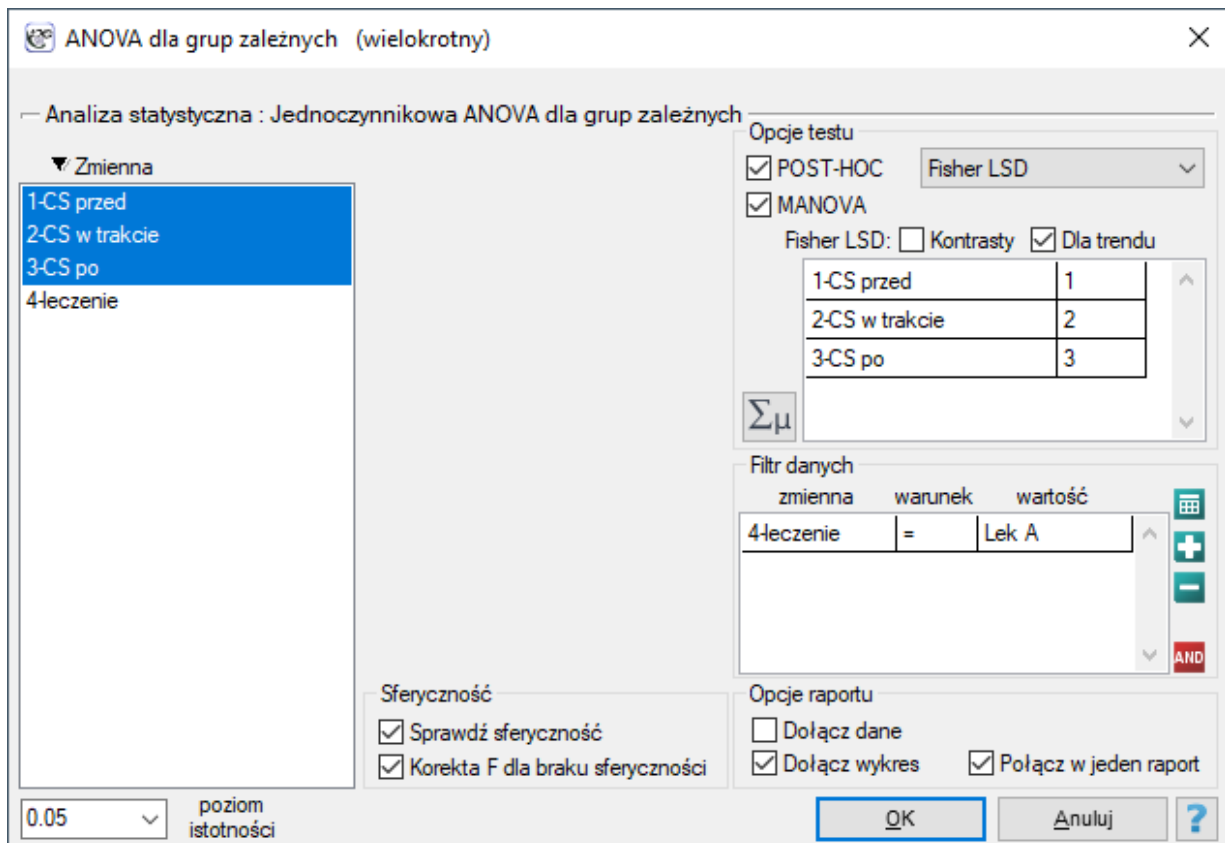
W celu zminimalizowania skutków braku sferyczności można wykorzystać trzy korekty zmieniające liczbę stopni swobody przy testowaniu z rozkładu F. Najprostszą lecz najślabszą jest **korekta**

dolnej granicy Epsilon. Nieco silniejszą lecz również konserwatywną jest korekta Greenhouse-Geisser (1959)[70]. Największą mocą charakteryzuje się korekta Huynh-Feldt (1976)[84]. Przy znacznym naruszeniu sferyczności najbardziej wskazanym rozwiązaniem jest jednak wykonanie analizy, która tego założenie nie wymaga, czyli MANOVA.

Podjęcie wielowymiarowe - MANOVA

MANOVA (ang. *multivariate analysis of variance*) czyli wielowymiarowa ANOVA jest analizą nie zakładającą sferyczności. W przypadku niespełnienia tego założenia jest to metoda najefektywniejsza, więc powinna być wybierana w zastępstwie analizy wariancji dla powtarzanych pomiarów. Omówienie tej metody można znaleźć w rozdziale [jednoczynnikowa MANOVA](#). Zastosowanie jej do pomiarów powtarzanych (bez czynnika grup niezależnych) ogranicza jej działanie do danych będących różnicą sąsiadujących pomiarów i zapewnia testowanie tej samej hipotezy, co ANOVA dla zmiennych zależnych.

Okno z ustawieniami opcji jednoczynnikowej ANOVA dla grup zależnych z korektą Epsilon i MANOVA wywołujemy poprzez menu Statystyka→Testy parametryczne→ANOVA dla grup zależnych lub poprzez [Kreator](#).



ANOVA dla grup zależnych (wielokrotny)

Analiza statystyczna : Jednoczynnikowa ANOVA dla grup zależnych

Zmienna

- 1-CS przed
- 2-CS w trakcie
- 3-CS po
- 4-leczenie

Opcje testu

POST-HOC Fisher LSD

MANOVA

Fisher LSD: Kontrasty Dla trendu

1-CS przed	1
2-CS w trakcie	2
3-CS po	3

$\Sigma\mu$

Filtr danych

zmienna	warunek	wartość
4-leczenie	=	Lek A

Sferyczność

Sprawdź sferyczność

Korekta F dla braku sferyczności

Opcje raportu

Dołącz dane

Dołącz wykres Połącz w jeden raport

0.05 poziom istotności

OK Anuluj ?

PRZYKŁAD 14.3. (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.

Hipotezy dla leczonych lekiem A:

- \mathcal{H}_0 : średnia wartość ciśnienia skurczowego krwi jest taka sama na każdym etapie leczenia - dla leczonych lekiem A,
 \mathcal{H}_1 : przynajmniej na jednym etapie leczenia lekiem A średnia wartość ciśnienia skurczowego jest inna.

Hipotezy dla leczonych lekiem B brzmią analogicznie.


Ponieważ dane mają rozkład normalny, analizę rozpoczynamy od sprawdzenia założenia o 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 ($W = 0.68$, wartość $p = 0.0063$).

Macierz kowariancji			
	CS przed	CS w trakcie	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 trakcie	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

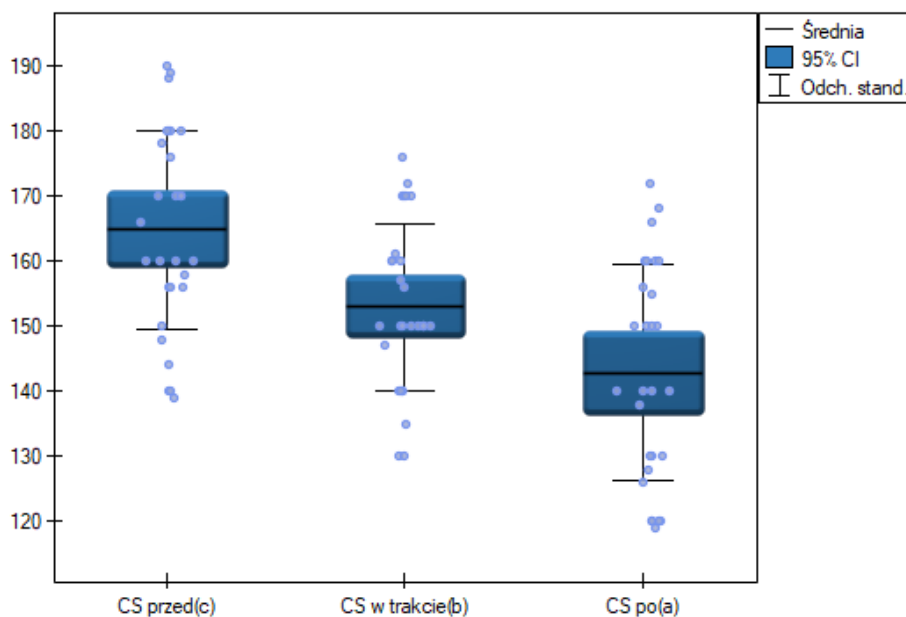
Wznawiamy analizę  i w oknie opcji testu wybieramy filtr podstawowy, by wykonać ANOVA powtarzanych pomiarów - dla leczonych lekiem A, a następnie korektę tej analizy i statystykę MANOVA - dla leczonych lekiem B.

Wyniki dla leczonych lekiem A:

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

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

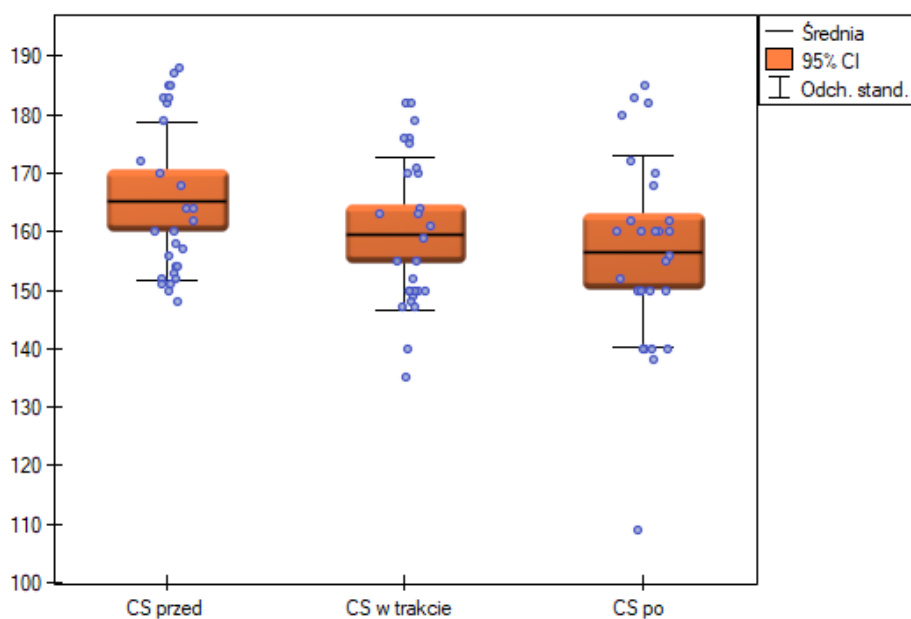


wskazują na istotne (na poziomie istotności $\alpha = 0.05$) różnice pomiędzy średnimi wartościami ciśnienia skurczowego (wartość $p < 0.0001$ dla ANOVA powtarzanych pomiarów). Ponad 66% wewnątrzsobowego zróżnicowania wyników może być tłumaczona zastosowaniem leku A ($\eta = 0.66$). Różnice dotyczą wszystkich porównywanych etapów leczenia (wynik POST-HOC). Istotny jest również trend spadku ciśnienia skurczowego na skutek leczenia (wartość $p < 0.0001$). Zatem Lek A możemy uznać za lek 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

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



wskazują na brak istotnych różnic pomiędzy średnimi wartościami ciśnienia skurczowego, zarówno wtedy, gdy korzystamy z poprawek epsilon, jak i Lambda Wilksa (MANOVA). Zaledwie niecałe 17% wewnątrzsobowego zróżnicowania wyników może być tłumaczone zastosowaniem leku B ($\eta = 0.17$).

14.2 TESTY NIEPARAMETRYCZNE

14.2.1 ANOVA Kruskala-Wallis

Jednoczynnikowa analiza wariancji dla rang Kruskala-Walisa, czyli ANOVA Kruskala-Wallis (ang. *Kruskal-Wallis one-way analysis of variance by ranks*) opisana przez Kruskala (1952)[94] oraz Kruskala i Wallisa (1952)[95] jest rozszerzeniem testu [U-Manna-Whitneya](#) na więcej niż dwie populacje. Test ten służy do weryfikacji hipotezy o braku przesunięcia porównywanych rozkładów tzn. najczęściej nieistotności różnic pomiędzy medianami badanej zmiennej w kilku ($k \geq 2$) populacjach (przy czym zakładamy, że rozkłady zmiennej są sobie bliskie - porównanie wariancji rang można sprawdzić testem dla rang [Conovera](#)).

Dodatkowe analizy:

- możliwe jest testowanie trendu w ułożeniu badanych grup poprzez wykonanie [testu Jonckheere-Terpstra dla trendu](#).

Podstawowe warunki stosowania:

- pomiar na [skali porządkowej](#) lub [interwałowej](#),
- [model niezależny](#).

Hipotezy dotyczą równości średnich rang dla kolejnych populacji lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0 &: \phi_1 = \phi_2 = \dots = \phi_k, \\ \mathcal{H}_1 &: \text{nie wszystkie } \phi_j \text{ są sobie równe } (j = 1, 2, \dots, k), \end{aligned}$$

gdzie:

$\phi_1, \phi_2, \dots, \phi_k$ to rozkłady badanej zmiennej w populacjach, z których pobrano próby.

Statystyka testowa ma postać:

$$H = \frac{1}{C} \left(\frac{12}{N(N+1)} \sum_{j=1}^k \left(\frac{(\sum_{i=1}^{n_j} R_{ij})^2}{n_j} \right) - 3(N+1) \right),$$

gdzie:

$$N = \sum_{j=1}^k n_j,$$

n_j – licznosci prób dla ($j = 1, 2, \dots, k$),

R_{ij} – rangi przypisane do wartości zmiennej, dla ($i = 1, 2, \dots, n_j$), ($j = 1, 2, \dots, k$),

$$C = 1 - \frac{\sum (t^3 - t)}{N^3 - N} \text{ – korekta na } \text{rangi wiązane},$$

t – liczba przypadków wchodzących w skład rangi wiązanej.

Wzór na statystykę testową H 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 $C = 1$).

Statystyka H ma asymptotycznie (dla dużych licznosci) [rozkład \$\chi^2\$](#) z liczbą stopni swobody wyznaczaną według wzoru: $df = (k - 1)$.

Wyznaczoną na podstawie [statystyki testowej wartość \$p\$](#) porównujemy z α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Testy POST-HOC

Wprowadzenie do **kontrastów i testów POST-HOC** przeprowadzone zostało w rozdziale dotyczącym jednoczynnikowej analizy wariancji.

Test Dunna

Dla porównań prostych, zarówno równolicznych jak i różnolicznych grup.

Test Dunna (Dunn 1964[51]) zawiera poprawkę na rangi wiązane (Zar 2010[178]) i jest testem korygowanym ze względu na wielokrotne testowanie. Najczęściej wykorzystuje się tu korektę Bonferroniego lub Sidaka, chociaż dostępne są również inne, nowsze korekty opisane szerzej w dziale **Wielokrotne porównania**.

Przykład - **porównania proste** (porównanie pomiędzy sobą 2 wybranych median / średnich rang):

$$\begin{aligned}\mathcal{H}_0 &: \theta_j = \theta_{j+1}, \\ \mathcal{H}_1 &: \theta_j \neq \theta_{j+1}.\end{aligned}$$

- (i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = Z_{\alpha(\text{corrected})} \sqrt{\left(\frac{N(N+1)}{12} - \frac{\sum(t^3 - t)}{12(N-1)} \right) \sum_{j=1}^k \frac{c_j^2}{n_j}},$$

gdzie:

t – liczba przypadków wchodzących w skład rangi wiązanej

$Z_{\alpha(\text{corrected})}$ - 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:

$$Z = \frac{\sum_{j=1}^k c_j \bar{R}_j}{\sqrt{\left(\frac{N(N+1)}{12} - \frac{\sum(t^3 - t)}{12(N-1)} \right) \sum_{j=1}^k \frac{c_j^2}{n_j}}},$$

gdzie:

\bar{R}_j – średnia rang j -tej grupy, dla $(j = 1, 2, \dots, k)$,

Wzór na statystykę testową Z zawiera poprawkę na rangi wiązane. Poprawka ta jest stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych poprawka ta nie jest wyliczana, ponieważ $\sum(t^3 - t) = 0$).

Statystyka ta ma asymptotycznie (dla dużych licznosci 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**[46], stosowany dla porównań prostych zarówno równolicznych jak i różnolicznych grup.

- (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:

$$S^2 = \frac{1}{N-1} \left(\sum_{j=1}^k \sum_{i=1}^{n_j} R_{ij}^2 - N \frac{(N+1)^2}{4} \right)$$

$F_{\alpha,1,N-k}$ 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:

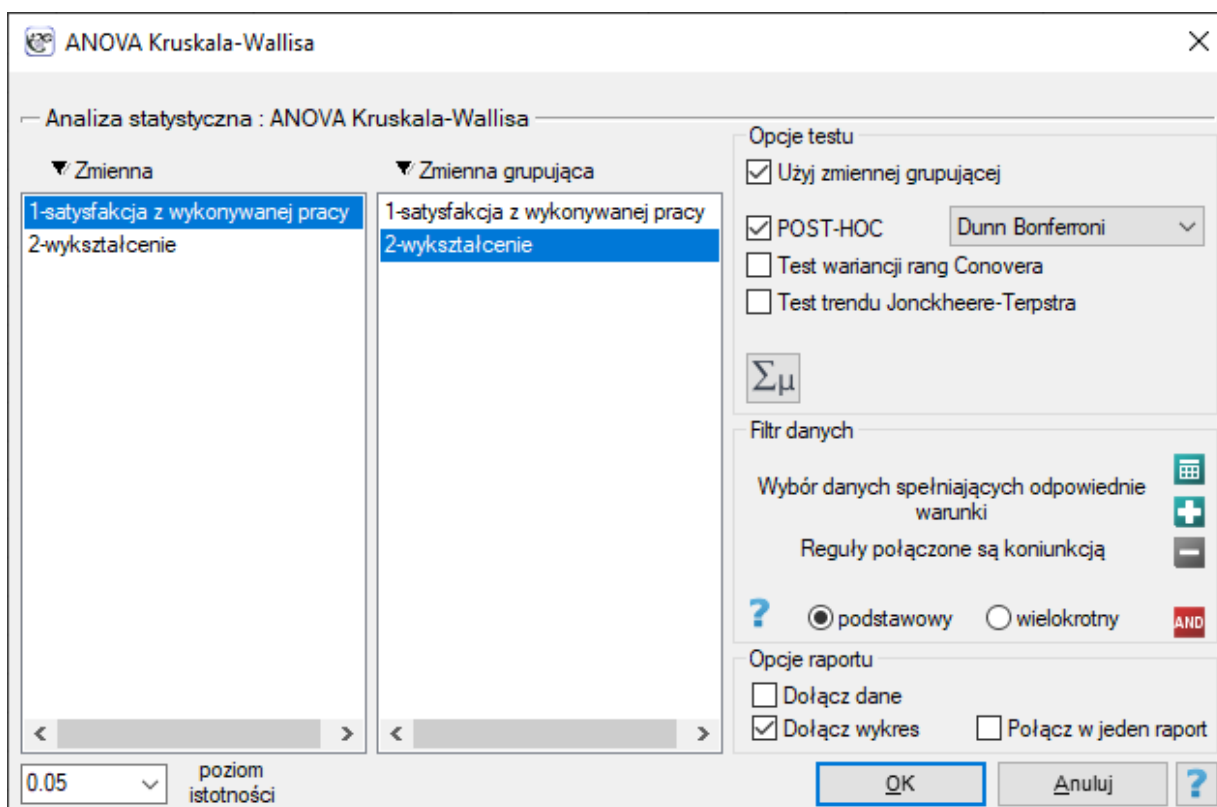
$$t = \frac{\sum_{j=1}^k c_j \bar{R}_j}{\sqrt{S^2 \frac{N-1-H}{N-k} \sum_{j=1}^k \frac{c_j^2}{n_j}}}$$

gdzie:

\bar{R}_j – średnia rang j -tej grupy, dla ($j = 1, 2, \dots, k$),

Statystyka ta podlega rozkładowi t -Studenta z $N - k$ stopniami swobody.

Okno z ustawieniami opcji ANOVA Kruskala-Wallisa wywołujemy poprzez menu Statystyka → Testy nieparametryczne → ANOVA Kruskala-Wallisa lub poprzez Kreator.



PRZYKŁAD 14.4. (satysfakcjaZpracy.pqs)

Przepymano 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 nie zmienia się dla poszczególnych kategorii wykształcenia.

Hipotezy:

- \mathcal{H}_0 : poziom satysfakcji z pracy jest taki sam dla każdej kategorii wykształcenia,
 \mathcal{H}_1 : przynajmniej jedna kategoria wykształcenia (jedna populacja) charakteryzuje się innym poziomem satysfakcji z pracy.

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

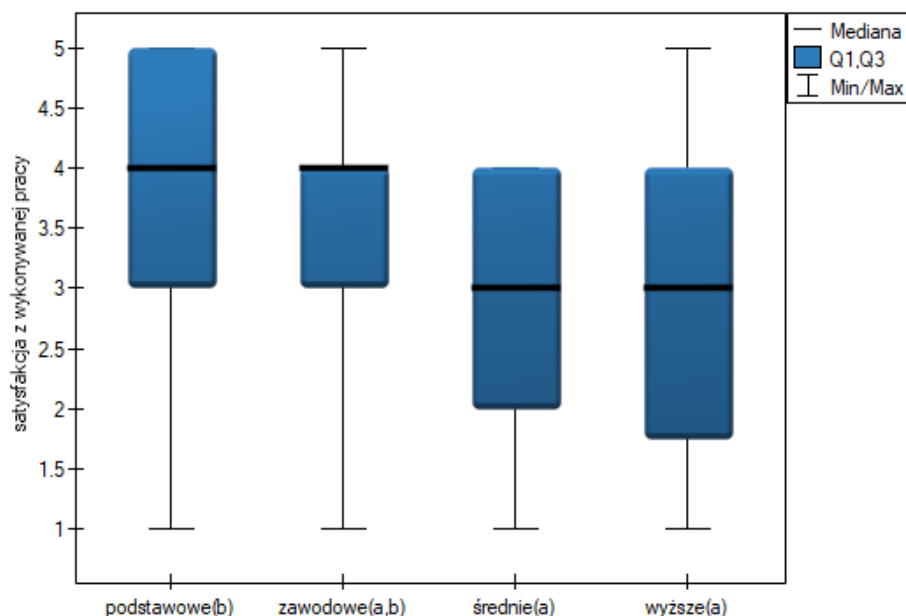
Uzyskana wartość $p = 0.001034$ świadczy o istotnej różnicy poziomu satysfakcji pomiędzy porównywanymi kategoriami wykształcenia. Przeprowadzona analiza POST-HOC Dunna z korektą Bonferroniego wskazuje, że istotne różnice dotyczą osób z wykształceniem podstawowym i średnim oraz z wykształceniem podstawowym i wyższym. Nieco więcej różnic możemy potwierdzić wybierając silniejszy 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	*	*		

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 możemy zobaczyć grupy jednorodne wyznaczone

przez test POST-HOC. Jeśli zdecydujemy się na przedstawienie wyników Dunna z korektą Bonferroniego zobaczymy dwie grupy jednorodne, które nie są zupełnie odrębne, tzn. grupę (a) - osoby słabiej oceniające satysfakcję z pracy i grupę (b) - osoby lepiej oceniające tę satysfakcję. Wykształcenie zawodowe przynależy do obydwu tych grup, co oznacza, że osoby z tym wykształceniem oceniają satysfakcję z pracy dość różnorodnie. Ten sam opis grup jednorodnych możemy znaleźć w wynikach testów POST-HOC.

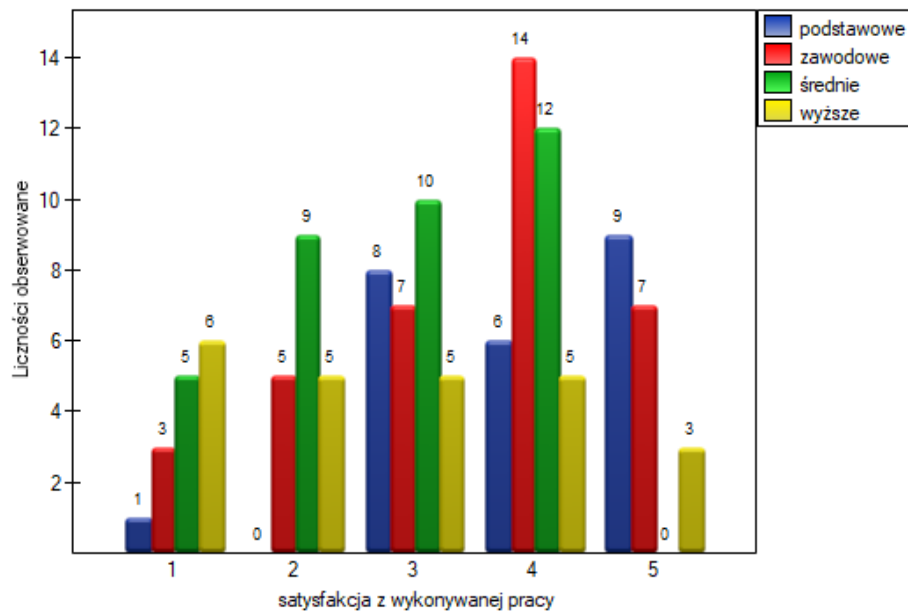


Dokładny opis danych możemy przedstawić wybierając w oknie analizy statystyki opisowe $\Sigma\mu$ i wskazując na dodanie do opisu licznosci i procentów.

Podsumowanie				
Grupa	podstawowe	zawodowe	średnie	wyższe
Licznosc	24	36	36	24
Mediana	4	4	3	3
Dolny kwartyl	3	3	2	1.75
Gorny kwartyl	5	4	4	4

Licznosc(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 licznosci.



14.2.2 Test Jonckheere-Terpstra dla trendu

Test Jonckheere-Terpstra dla trendu (*ang. the Jonckheere-Terpstra test for ordered alternatives*) opisany niezależnie przez Jonckheere (1954) [88] i Terpstra (1952)[159] może być wyliczany w takiej samej sytuacji jak ANOVA Kruskala-Wallis, gdyż bazuje na tych samych założeniach. Test Jonckheere-Terpstra inaczej jednak ujmuje hipotezę alternatywną - wskazując w niej na istnienie trendu dla kolejnych populacji.

Hipotezy są upraszczane do median:

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \dots = \theta_k,$$

$$\mathcal{H}_1 : \theta_1 \geq \theta_2 \geq \dots \geq \theta_k, \text{ z co najmniej jedną nierównośćią ścisłą}$$

Uwaga!

Określenie: "z co najmniej jedną nierównośćią ścisłą" zapisane w hipotezie alternatywnej tego testu oznacza, że co najmniej mediana jednej populacji powinna być większa niż mediana innej populacji w kolejności określonej.

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 - licznosci prób dla ($j = 1, 2, \dots, k$).

Uwaga!

By można było przeprowadzić analizę trendu, należy wskazać oczekiwaną kolejność populacji przypisując im kolejne liczby naturalne.

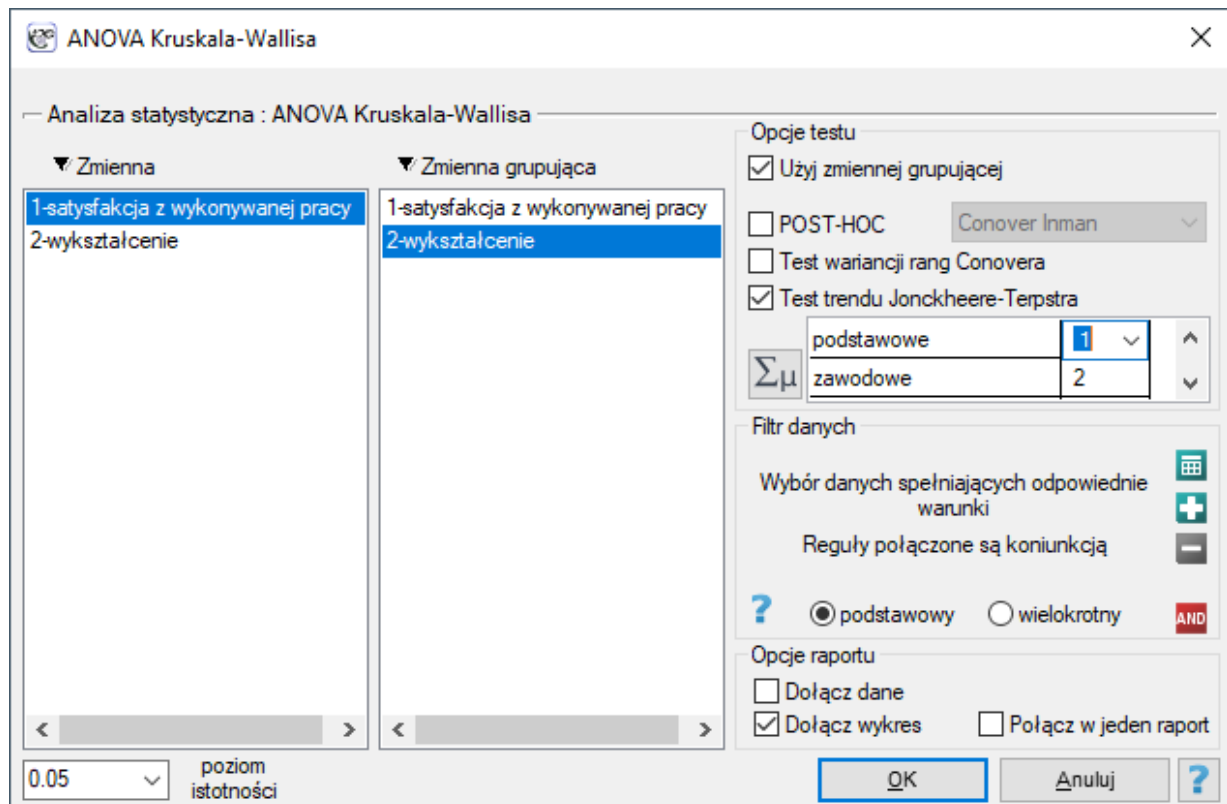
Wzór na statystykę testową Z zawiera poprawkę na rangi wiązane. Poprawka ta jest stosowana, gdy rangi wiązane występują (gdy nie ma rang wiązanych wzór na statystykę testową sprowadza się do oryginalnej formuły Jonckheere-Terpstra nie zawierającej tej poprawki).

Statystyka Z ma asymptotycznie (dla dużych licznosci) [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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji testu Jonckheere-Terpstra wywołujemy poprzez menu Statystyka \rightarrow Testy nieparametryczne \rightarrow ANOVA Kruskala-Wallisa lub poprzez [Kreator](#).




PRZYKŁAD (14.4) c.d. (plik *satysfakcjaZpracy.pqs*) Podejrzewa się, że osoby lepiej wykształcone mają wysokie wymagania zawodowe, co może zmniejszać poziom satysfakcji z pierwszej pracy, która często takich wymagań nie spełnia. Dlatego też warto przeprowadzić analizę trendu.

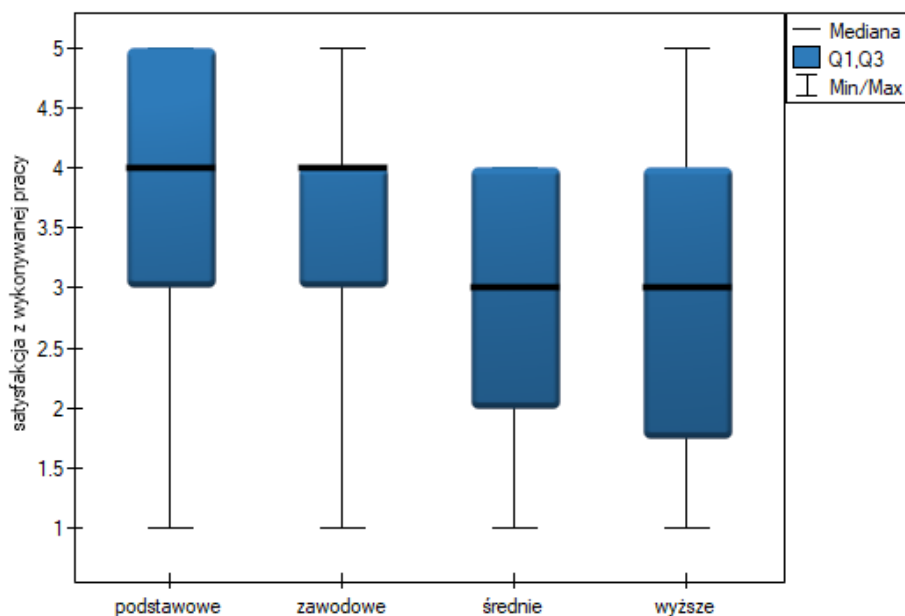
Hipotezy:

\mathcal{H}_0 : brak wskazanego trendu w poziomie satysfakcji z pierwszej pracy,

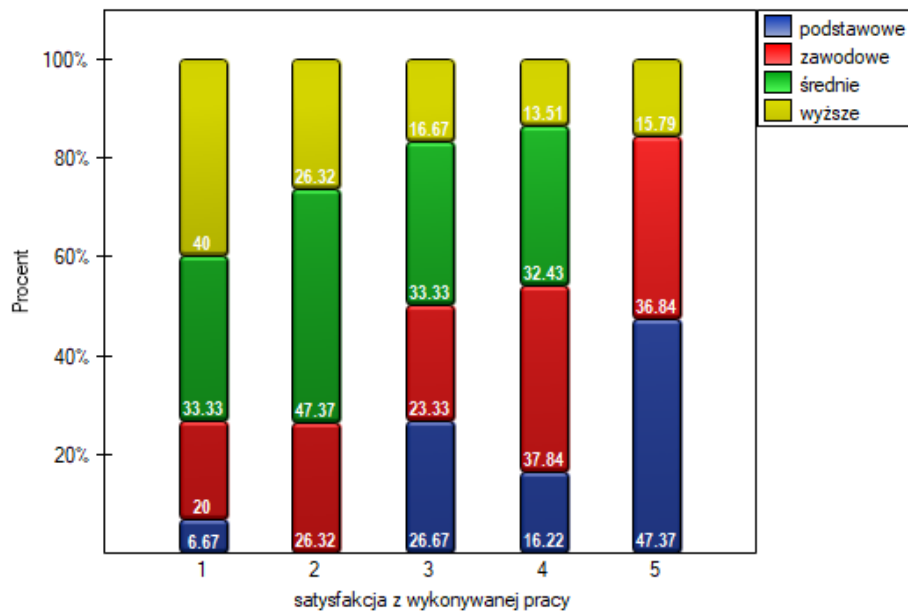
\mathcal{H}_1 : istnieje wskazany trend w poziomie satysfakcji z pierwszej pracy.

W tym celu wznawiamy analizę przyciskiem , zaznaczamy opcję Test trendu Jonckheere-Terpstra i kolejnym kategoriom wykształcenia przypisujemy kolejne liczby naturalne.

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



Uzyskana jednostronna wartość $p = 0.000045$ i jest mniejsza niż zadany poziom istotności $\alpha = 0.05$, co przemawia na rzecz rzeczywistego występującego trendu zgodnego z oczekiwaniami badacza. Istnienie tego trendu możemy również potwierdzić przedstawiając procentowy rozkład uzyskanych odpowiedzi.



14.2.3 Test wariancji rang Conover

Test kwadratów rang Conovera służy, podobnie jak test Fishera-Snedecora (dla $k = 2$), test Levene i test Browna-Forsythea (dla $k \geq 2$) do weryfikacji hipotezy podobnym zróżnicowaniu badanej zmiennej w kilku populacjach. Jest on nieparametrycznym odpowiednikiem wskazanych wyżej testów, przez to nie zakłada normalności rozkładu danych i opiera się na rangach [46]. Jednak test ten bada zróżnicowanie, a więc odległości do średniej, dlatego podstawowym warunkiem jego stosowania jest:

- pomiar na **skali interwałowej**,

Hipotezy:

- \mathcal{H}_0 : rozproszenie danych w porównywanych populacjach jest takie samo
- \mathcal{H}_1 : przynajmniej dwie populacje różnią się wielkością rozproszenia danych.

Statystyka testowa ma postać:

$$\chi^2 = \frac{1}{D^2} \left(\sum_{j=1}^k \frac{S_j^2}{n_j} - N \bar{S}^2 \right)$$

gdzie:

$$N = n_1 + n_2 + \dots + n_k,$$

n_j - licznosc w poszczególnych grupach,

S_j - suma kwadratów rang w j -tej grupie,

$$\bar{S} = \frac{1}{N} \sum_{j=1}^k S_j - \text{średnia wszystkich kwadratów rang,}$$

$$D^2 = \frac{1}{N-1} \left(\sum_{i=1}^N R_i^2 - N \bar{S}^2 \right),$$

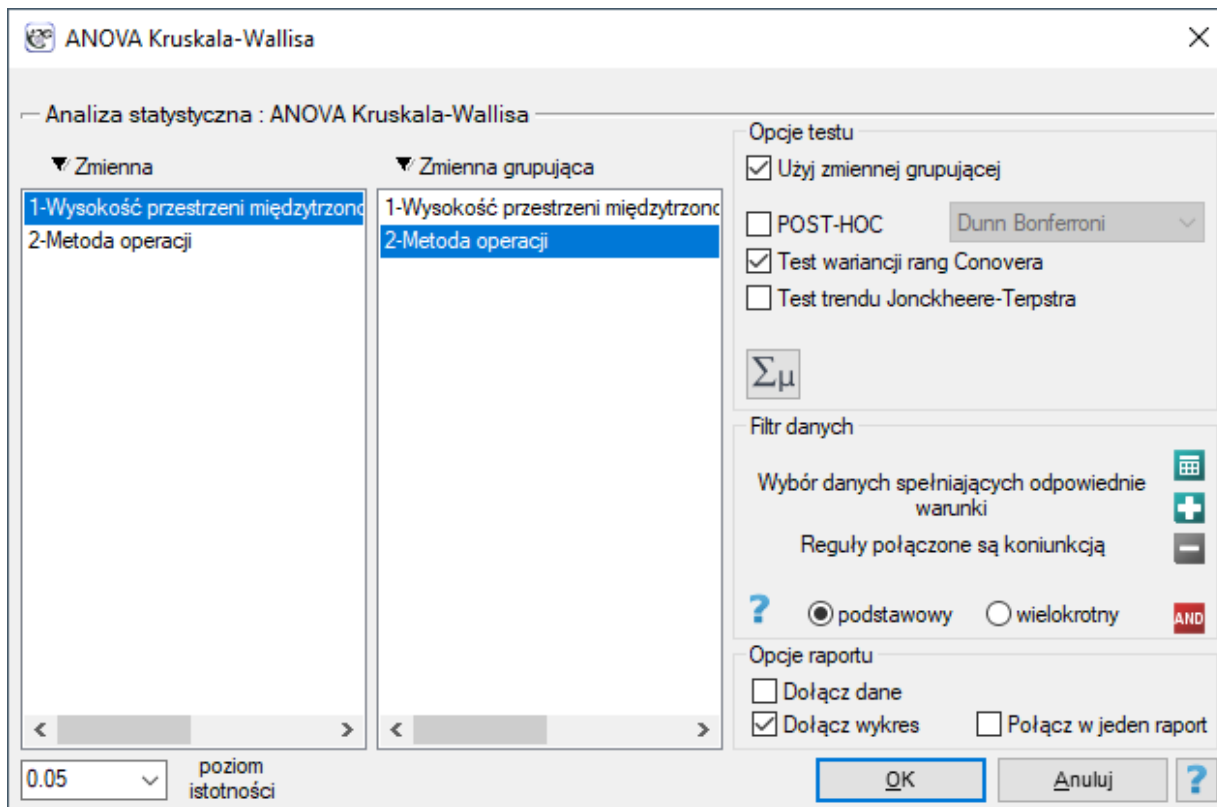
R_i -rangi dla wartości przedstawiających odległość pomiaru od średniej danej grupy.

Statystyka ta ma **rozkład χ^2** z $k - 1$ stopniem swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

- jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
- jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji testu wariacji rang Conovera wywołujemy poprzez menu Statystyka → Testy nieparametryczne → ANOVA Kruskala-Wallis, opcja testu wariacji rang Conovera lub Statystyka → Testy nieparametryczne → Mann-Whitney, opcja testu wariacji rang Conovera.



PRZYKŁAD 14.5. (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 min. pod względem wysokości przestrzeni międzytrzonowej (WPMT) przed operacją. Podobieństwo powinno dotyczyć nie tylko wartości przeciętnych ale również zróżnicowania grup.

Sprawdzono rozkład danych.

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 w przypadku dwóch metod operacji WPMT wykazuje odstępstwa od normalności, których przyczyną w dużej mierze jest skośność danych. Dalsza analiza porównawcza przeprowadzona zostanie przy pomocy testu Kruskala-Wallisa, by porównać czy poziom WPMT różni się pomiędzy metodami oraz testu Conovera, który wskaże, czy w każdej metodzie rozpiętość wyników WPMT jest podobna. Hipotezy dla testu wariancji Conovera:

\mathcal{H}_0 : różnorodność (zakres) WPMT jest taka sama dla każdej metody operacji,

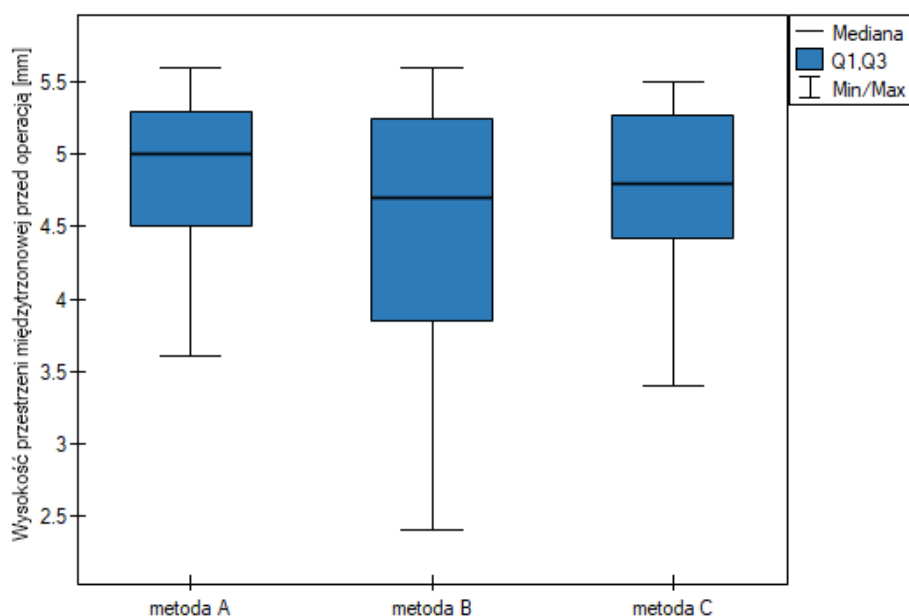
\mathcal{H}_1 : różnorodność (zakres) WPMT jest wyższa/nizsza przynajmniej dla jednej metody operacji.

Hipotezy dla testu Kruskala-Wallisa :

\mathcal{H}_0 : poziom WPMT jest taki sam dla każdej metody operacji,

\mathcal{H}_1 : poziom WPMT jest wyższy/nizszy przynajmniej dla jednej metody operacji.

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



Najpierw interpretacji podlega wartość testu wariancji Conovera, która wskazuje na istotne statystycznie różnice w zakresach porównywanych grup ($p=0.0022$). Z wykresu możemy wnioskować, że różnice dotyczą głównie grupy 3. Ponieważ wykryto różnice w zakresie WPMT, to interpretacja wyniku testu Kruskala-Wallisa, porównującego poziom WPMT dla tych metod, powinna być ostrożna, gdyż test ten jest wrażliwy na niejednorodność wariancji. Mimo iż test Kruskala-Wallisa nie wykazał istotnych różnic ($p=0.2057$), to zaleca się by chorych o niskich WPMT (którzy zostali przypisani głównie do operacji metodą B) rozmieścić bardziej równomiernie tzn. by sprawdzić, czy nie można im zaproponować wykonania operacji metodą A lub C. Po ponownym przydziale chorych należy powtórzyć analizę.

14.2.4 ANOVA Friedmana

Analiza wariancji powtarzanych pomiarów dla rang Friedmana, czyli ANOVA Friedmana (*ang. Friedman repeated measures analysis of variance by ranks*) opisana została przez Friedmana (1937)[64]. Test ten stosuje się w sytuacji, gdy pomiarów badanej zmiennej dokonujemy kilkakrotnie ($k \geq 2$) w różnych warunkach. Stosowana jest również, 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.

Iman Davenport (1980[85]) pokazał, że w wielu przypadkach statystyka Friedmana jest nadmiernie konserwatywna i dokonał pewnej jej modyfikacji. Modyfikacja ta jest nieparametrycznym odpowiednikiem ANOVA powtarzanych pomiarów co sprawia, że jest obecnie rekomendowana do stosowania w zastępstwie tradycyjnej statystyki Friedmana.

Dodatkowe analizy:

- możliwe jest uwzględnienie braków danych poprzez opcje Akceptuj braki danych, wyliczając ANOVA Durбина lub ANOVA Skillings-Mack;
- możliwe jest testowanie trendu w ułożeniu badanych grup poprzez wykonanie testu Page dla trendu.

Podstawowe warunki stosowania:

- pomiar na skali porządkowej lub interwałowej,
- model zależny.

Hipotezy dotyczą równości sumy rang dla kolejnych pomiarów lub są upraszczane do median:

$$\begin{aligned} \mathcal{H}_0 &: \theta_1 = \theta_2 = \dots = \theta_k, \\ \mathcal{H}_1 &: \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2, \dots, k), \end{aligned}$$

gdzie:

$\theta_1, \theta_2, \dots, \theta_k$ mediany badanej cechy w kolejnych pomiarach z badanej populacji.

Wyznacza się dwie statystyki testowe: statystykę Friedmana i modyfikację Imana-Davenport tej statystyki.

Statystyka Friedmana ma postać:

$$T_1 = \frac{1}{C} \left(\frac{12}{nk(k+1)} \left(\sum_{j=1}^k \left(\sum_{i=1}^n R_{ij} \right)^2 \right) - 3n(k+1) \right),$$

gdzie:

n – liczność próby,

R_{ij} – rangi przypisane kolejnym pomiarom ($j = 1, 2, \dots, k$), oddzielnie dla każdego z badanych obiektów ($i = 1, 2, \dots, n$),

$C = 1 - \frac{\sum(t^3 - t)}{n(k^3 - k)}$ – korekta na rangi wiązane,

t – liczba przypadków wchodzących w skład rangi wiązanej.

Modyfikacja Imana-Davenport statystyki Friedmana ma postać:

$$T_2 = \frac{(n_j - 1)T_1}{n_j(k - 1) - T_1}$$

Wzór na statystykę T_1 i T_2 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 $C = 1$).

Statystyka T_1 ma asymptotycznie (dla dużych licznosci) 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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Testy POST-HOC

Wprowadzenie do **kontrastów i testów POST-HOC** przeprowadzone zostało w rozdziale 14.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Dunna

Stosowany dla porównań prostych (liczność w poszczególnych pomiarach zawsze jest taka sama).

Test Dunna (Dunn 1964[51]) jest testem korygowanym ze względu na wielokrotne testowanie. Najczęściej wykorzystuje się tu korektę **Bonferroniego** lub **Sidaka**, chociaż dostępne są również inne, nowsze korekty opisane szerzej w dziale **Wielokrotne porównania**.

Przykład - **porównania proste** (porównanie pomiędzy sobą 2 wybranych median / średnich rang):

$$\begin{aligned}\mathcal{H}_0 &: \theta_j = \theta_{j+1}, \\ \mathcal{H}_1 &: \theta_j \neq \theta_{j+1}.\end{aligned}$$

(i) Wartość najmniejszej istotnej różnicy wyliczana jest z wzoru:

$$NIR = Z_{\alpha(\text{corrected})} \sqrt{\frac{k(k+1)}{6n}},$$

gdzie:

$Z_{\alpha(\text{corrected})}$ - 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:

$$Z = \frac{\sum_{j=1}^k c_j \bar{R}_j}{\sqrt{\frac{k(k+1)}{6n}}},$$

gdzie:

\bar{R}_j - średnia rang j -tego pomiaru, dla $(j = 1, 2, \dots, k)$,

Statystyka ta ma asymptotycznie (dla dużych licznosci próby) **rozkład normalny**, a **wartość p** jest korygowana o **liczbę możliwych porównań prostych** zgodnie z wybraną poprawką.

Test Conover-Inman

Nieparametryczny odpowiednik **LSD Fishera**[46], 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,

$F_{\alpha,1,df_2}$ 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:

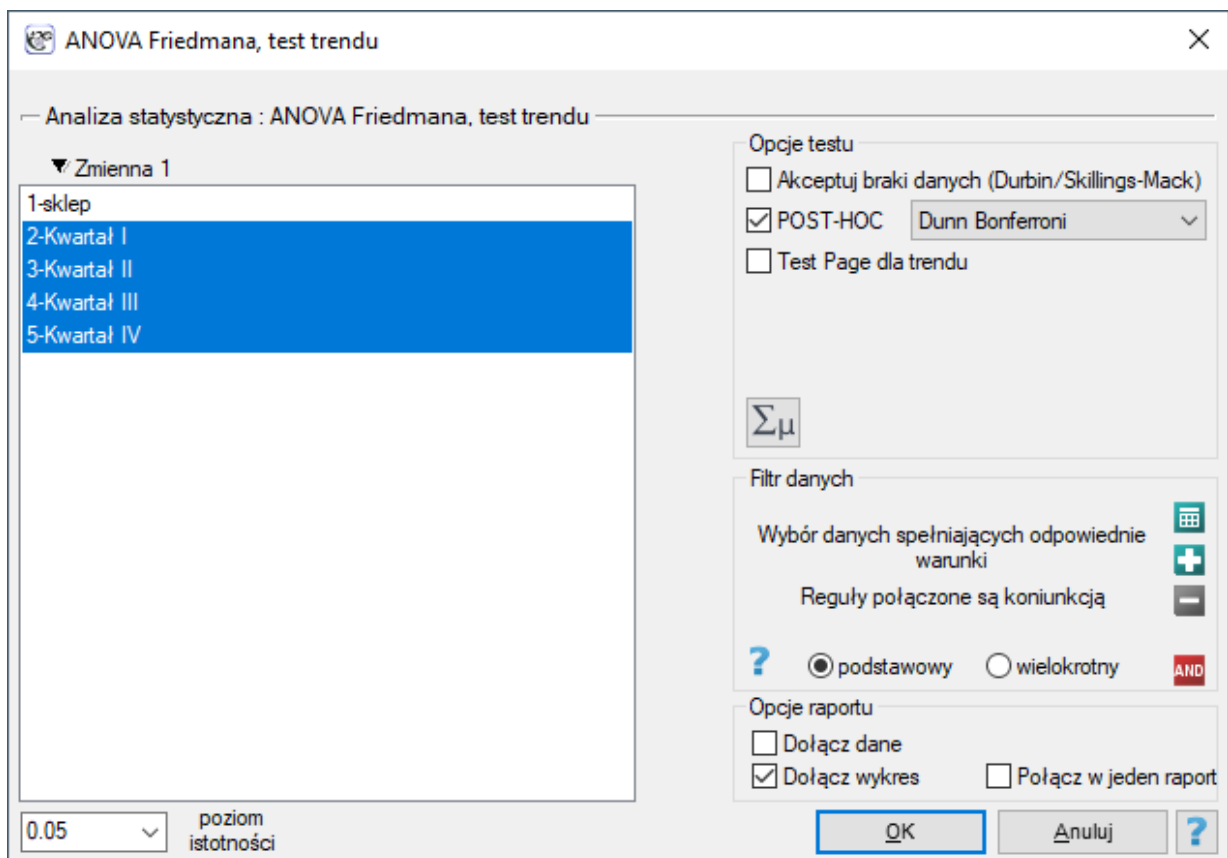
$$t = \frac{\sum_{j=1}^k c_j R_j}{\sqrt{\frac{2(n_j A - \sum_{j=1}^k R_j^k)}{(n_j - 1)(k - 1)}}},$$

gdzie:

R_j – suma rang j -tego pomiaru, dla ($j = 1, 2, \dots, k$),

Statystyka ta podlega **rozkładowi t -Studenta** z df_2 stopniami swobody.

Okno z ustawieniami opcji ANOVA Friedmana wywołujemy poprzez menu Statystyka → Testy nieparametryczne → ANOVA Friedmana (możliwość braków danych) lub poprzez [Kreator](#).



PRZYKŁAD 14.6. (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 intensywne 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	1279	1350	1490
SK13	1271	1417	1583	1513
SK14	1436	1310	1357	1468

Hipotezy:

- \mathcal{H}_0 : brak jest istotnej różnicy pomiędzy wielkością sprzedaży dla porównywanych kwartałów w populacji reprezentowanej przez zebraną próbę,
 \mathcal{H}_1 : różnica w sprzedaży dla populacji reprezentowanej przez zebraną próbę pomiędzy przynajmniej dwoma kwartałami jest istotna.

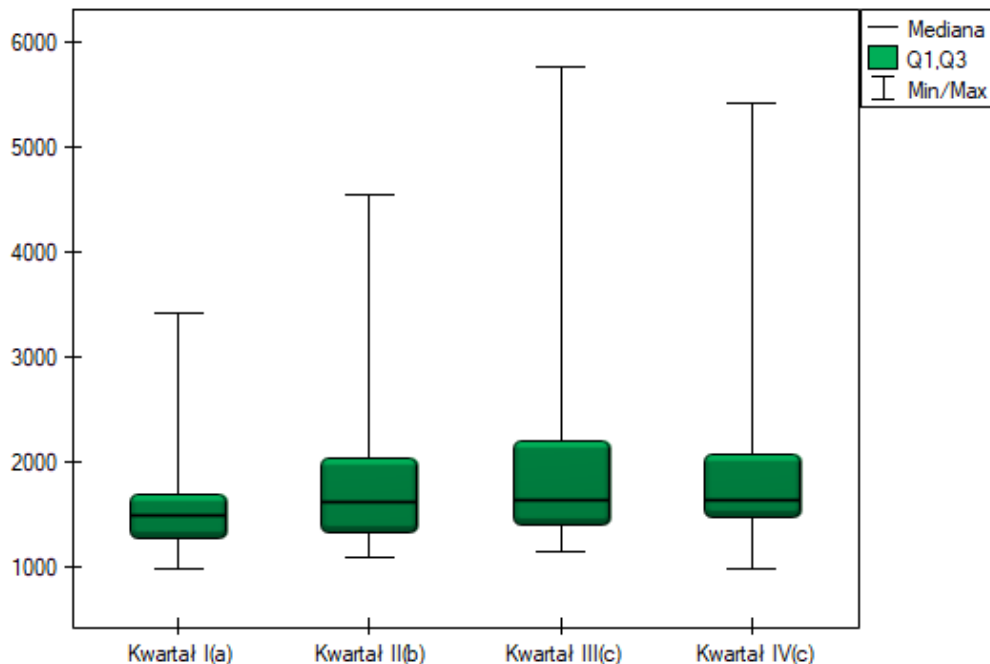
ANOVA Friedmana, test trendu	
Czas analizy	0.12 sek.
Analizowane zmienne	Kwartał I;Kwartał II;Kwartał
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Akceptuj braki danych (Durbin/Skillings-Mack)	Nie
C	350
A	420
Statystyka T1 Friedman	23.9143
Stopnie swobody	3
Wartość 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

Porównując wartość p testu Friedmana (jak i wartość p korekty Iman-Davenport testu Friedmana) z poziomem istotności $\alpha = 0.05$, stwierdzamy, że 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.

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(t)	Kwartał IV(b)
A	*	*		
B		*	*	*

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 przedstawiliśmy grupy jednorodnie wyznaczone testem Conover-Iman.



Dokładny opis danych możemy przedstawić wybierając w oknie analizy statystyki opisowe Σu .

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 kwartył	1272.25	1326.75	1398.25	1473.5
Górny kwartył	1683.75	2026.25	2193.5	2068.25

Gdyby dane były opisane skalą porządkową o niewielu kategoriach, warto by było przedstawić je również w licznosciach i procentach. W naszym przykładzie nie byłaby to dobra metoda opisu.

14.2.5 Test Page dla trendu

Test Page dla trendu (*ang. the Page test for ordered alternative*) opisany w roku 1963 przez Page E. B. [129] może być wyliczany w takiej samej sytuacji jak ANOVA Friedmana, gdyż bazuje na tych samych założeniach. Test Page inaczej jednak ujmuje hipotezę alternatywną - wskazując w niej na istnienie trendu w kolejnych pomiarach.

Hipotezy dotyczą równości sumy rang dla kolejnych pomiarów lub są upraszczane do median:

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \dots = \theta_k,$$

$$\mathcal{H}_1 : \theta_1 \geq \theta_2 \geq \dots \geq \theta_k, \text{ z co najmniej jedną nierównością ścisłą}$$

Uwaga!

Określenie: "z co najmniej jedną nierównością ścisłą" zapisane w hipotezie alternatywnej tego testu oznacza, że co najmniej jedna mediana powinna być większa niż mediana innej grupy pomiarów w kolejności określonej.

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).

Uwaga!

By można było przeprowadzić analizę trendu, należy wskazać oczekiwane uporządkowanie pomiarów przypisując kolejne liczby naturalne kolejnym grupom pomiarowym. Liczby te w analizie traktowane są jako wagi c_1, c_2, \dots, c_k .

Wzór na statystykę testową Z nie zawiera poprawki na rangi wiązane, przez co staje się nieco bardziej konserwatywny, gdy rangi wiązane występują. Jednakże stosowanie korekty na rangi wiązane w przypadku tego testu nie jest rekomendowane.

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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Okno z ustawieniami opcji testu Page wywołujemy poprzez menu Statystyka → Testy nieparametryczne → ANOVA Friedmana (możliwość braków danych) lub poprzez [Kreator](#).

ANOVA Friedmana, test trendu

Analiza statystyczna : ANOVA Friedmana, test trendu

▼ Zmienna 1

- 1-sklep
- 2-Kwartał I
- 3-Kwartał II
- 4-Kwartał III
- 5-Kwartał IV

0.05 poziom istotności

Opcje testu

Akceptuj braki danych (Durbin/Skillings-Mack)

POST-HOC Dunn Bonferroni

Test Page dla trendu

2-Kwartał I	1
3-Kwartał II	2
4-Kwartał III	3
5-Kwartał IV	4

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny

Opcje raportu

Dołącz dane

Dołącz wykres Połącz w jeden raport

OK Anuluj ?

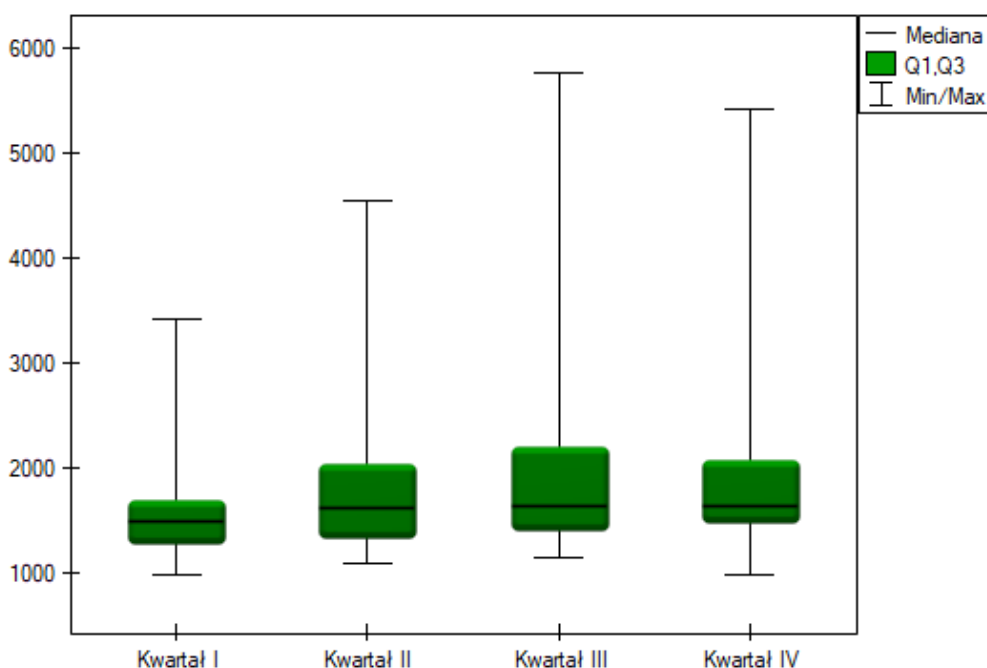
PRZYKŁAD (14.6) c.d. (plik *baton.pqs*) Oczekiwanym skutkiem prowadzenia przez firmę intensywnej kampanii reklamowej jest stały wzrost sprzedaży oferowanego batonu.

Hipotezy:

\mathcal{H}_0 : brak wskazanego trendu w sprzedaży batonu,

\mathcal{H}_1 : istnieje wskazany trend w sprzedaży batonu.

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



Porównując jednostronną wartość $p = 0,000004$ z poziomem istotności $\alpha = 0.05$, stwierdzamy, że kampania przyniosła oczekiwany trend wzrostu sprzedaży produktu.

14.2.6 ANOVA Durbina (brakujących danych)

Analiza wariancji powtarzanych pomiarów dla rang Durbina została zaproponowana przez Durbina (1951)[52]. Test ten stosuje się w sytuacji, gdy pomiarów badanej zmiennej dokonujemy kilkakrotnie - czyli w podobnej sytuacji w jakiej stosowana jest ANOVA Friedmana. Oryginalny test Durbina i test Friedmana dają ten sam wynik w sytuacji, gdy dysponujemy kompletnym zestawem danych. Test Durbina ma jednak pewną przewagę - można go również wyliczać dla niekompletnego zestawu danych. Przy czym braki danych nie mogą być zlokalizowane dowolnie, ale dane muszą tworzyć tzw. zbalansowany i niekompletny blok, czyli:

- liczba pomiarów dla każdego obiektu wynosi k ($k \leq t$),
- każdy pomiar dokonywany jest na r obiektach ($r \leq b$),
- liczba obiektów dla których wykonano jednocześnie tą samą parę pomiarów jest stała i wynosi λ .

gdzie:

t – łączna liczba rozpatrywanych pomiarów,

b – łączna liczba badanych obiektów

Podstawowe warunki stosowania:

- pomiar na [skali porządkowej](#) lub [interwałowej](#),
- [model zależny](#).

Hipotezy dotyczą równości sumy rang dla kolejnych pomiarów (R_j) lub są upraszczane do median (θ_j):

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \dots = \theta_t,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2, \dots, t),$$

Wyznacza się dwie statystyki testowe o następującej postaci:

$$T_1 = \frac{(t-1) \left[\sum_{j=1}^t R_j^2 - tC \right]}{A - C},$$

$$T_2 = \frac{T_1 / (t-1)}{(b(k-1) - T_1) / (bk - b - t + 1)},$$

gdzie:

R_j – suma rang dla kolejnych pomiarów ($j = 1, 2, \dots, t$),

R_{ij} – rangi przypisane kolejnym pomiarom, oddzielnie dla każdego z badanych obiektów ($i = 1, 2, \dots, b$),

$A = \sum_{i=1}^b \sum_{j=1}^t R_{ij}^2$ – 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. Ma ona asymptotycznie (dla dużych licznosci) **rozkład χ^2** z $df = t - 1$ stopniami swobody.

Statystyka T_2 to odpowiednik korekty Iman-Davenport ANOVA Friedmana, więc 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[46].

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Testy POST-HOC

Wprowadzenie do **kontrastów i testów POST-HOC** przeprowadzone zostało w rozdziale 14.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Conover-Inman

Stosowany dla porównań prostych (licznosc w poszczególnych pomiarach zawsze jest taka sama).

Hipotezy:

Przykład - **porównania proste** (porównanie pomiędzy sobą 2 wybranych median / sum rang):

$$\mathcal{H}_0 : \theta_j = \theta_{j+1},$$

$$\mathcal{H}_1 : \theta_j \neq \theta_{j+1}.$$

(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:

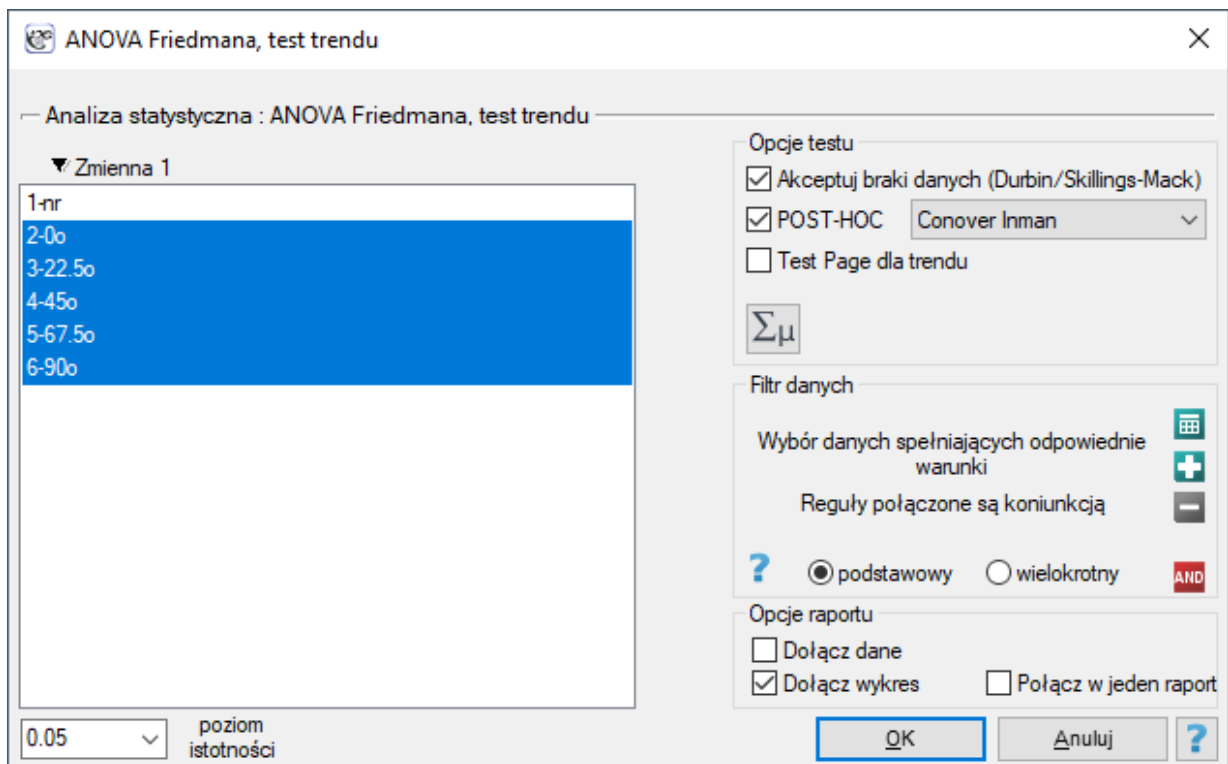
$t_{1-\alpha/2, bk-b-t+1}$ - to **wartość krytyczna** (statystyka) **rozkładu t -Studenta** dla poziomu istotności α i $df = bk - b - t + 1$ stopni swobody.

(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.

Okno z ustawieniami opcji ANOVA Durбина wywołujemy poprzez menu Statystyka→Testy nieparametryczne→ANOVA Friedmana (możliwość braków danych).



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.

PRZYKŁAD 14.7. (plik mirror.pqs)

Przeprowadzono eksperyment wśród 20 pacjentów szpitala psychiatrycznego (Ogilvie 1965 [126]). Eksperyment ten polegał na odrysowaniu linii prostych według zaprezentowanego wzoru. Wzór przedstawiał 5 linii rysowanych pod różnym kątem (0° , 22.5° , 45° , 67.5° , 90°) względem wskazanego środka. Zadaniem pacjentów było odwzorowanie linii mając zasłoniętą dłoń. Jako wynik eksperymentu zapisano czas w jakim pacjent kreślił daną linię. W idealnym przypadku każdy pacjent kreśliłby linię pod każdym kątem, jednak upływający czas i zmęczenie miałyby znaczny wpływ na wydajność pracy. Ponadto trudno jest utrzymać zainteresowanie pacjenta i chęć współpracy przez dłuższy czas. W związku z tym projekt zaplanowano i przeprowadzono w zbalansowanych i niekompletnych blokach. Każdy z 20 pacjentów wyrysowywał linię pod dwoma kontami (możliwych kątów było pięć). W ten sposób każdy kąt wyrysowywany był ośmiokrotnie. Czas w jakim każdy pacjent wyrysowywał linię pod zadaniem kątem zapisano w tabeli.

nr pacjenta	0°	22.5°	45°	67.5°	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

Chcemy sprawdzić, czy czas jaki został poświęcony na wyrysowanie poszczególnych linii jest zupełnie losowy, czy też są linie, których wyrysowywanie zajęło więcej lub mniej czasu.

Hipotezy:

\mathcal{H}_0 : brak jest istotnej różnicy pomiędzy czasem poświęconym na wyrysowanie poszczególnych linii przez pacjentów,

\mathcal{H}_1 : przynajmniej jedna linia rysowana jest w czasie krótszym/dłuższym.

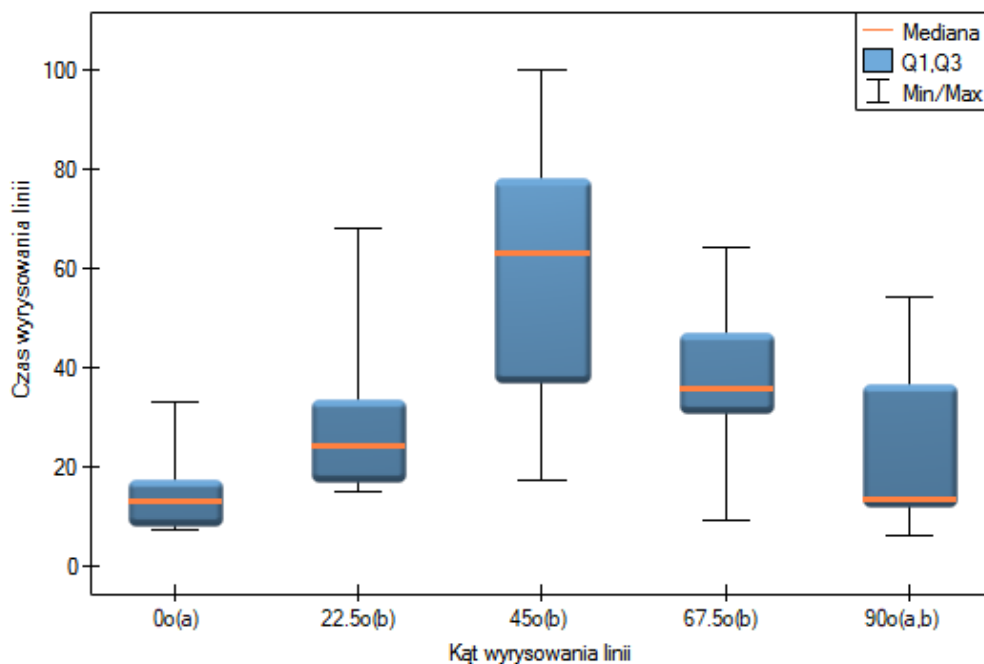
ANOVA Friedmana (możliwość braków danych)	
Czas analizy	0,06 sek.
Analizowane zmienne	0o;22.5o;45o;67.5o;90o
Poziom istotności	0,05
Akceptuj braki danych (Durbin/Skillings-Mack)	Tak
C	90
A	100
Statystyka T1 Durbin (F)	10,4
Stopnie swobody	4
Wartość p	0,034203
Statystyka T2 Durbin (F)	4,333333
Stopnie swobody	4 / 16
Wartość p	0,014541
Statystyka Skillings-Mack	10,4
Stopnie swobody	4
Wartość p	0,034203

Porównując wartość $p = 0,014541$ dla statystyki T_2 (lub wartość $p = 0.034203$ dla statystyki T_1) z poziomem istotności $\alpha = 0.05$ stwierdzamy, że linie nie są rysowane w tym samym czasie. 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°, 45° oraz 67.5°.

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		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		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		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.



14.2.7 ANOVA Skillings-Mack (brakujących danych)

Analiza wariancji powtarzanych pomiarów dla rang Skillings-Mack została zaproponowana przez Skillings'a i Mack'a w roku 1981 [153]. Jest to test, który może być wykorzystywany w przypadku występowania braków danych, ale braki te nie muszą występować w żadnym szczególnym układzie. Każdy obiekt musi mieć jednak przynajmniej dwie obserwacje. Jeśli nie ma rang wiązanych a braki nie występują jest tożsamy z ANOVA Friedmana, a jeśli braki danych występują w układzie zbalansowanym odpowiada wynikom ANOVA Durбина.

Podstawowe warunki stosowania:

- pomiar na [skali porządkowej](#) lub [interwałowej](#),
- [model zależny](#).

Hipotezy dotyczą równości sumy rang dla kolejnych pomiarów (R_j) lub są upraszczane do median (θ_j):

$$\begin{aligned} \mathcal{H}_0 &: \theta_1 = \theta_2 = \dots = \theta_k, \\ \mathcal{H}_1 &: \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2, \dots, k), \end{aligned}$$

Statystyka testowa ma postać:

$$\chi^2 = A \Sigma_0^{-1} A^T$$

gdzie:

$$A = (A_1, A_2, \dots, A_{k-1})$$

$$A_j = \sum_{i=1}^n \sqrt{\frac{12}{s_i+1}} \left(R_{ij} - \frac{s_i+1}{2} \right),$$

s_i - liczba obserwacji dla i -tego obiektu,

R_{ij} - rangi przypisane kolejnym pomiarom ($j = 1, 2, \dots, k$), oddzielnie dla każdego z badanych obiektów ($i = 1, 2, \dots, n$), przy czym rangi dla braków danych równe są średniej randze dla danego obiektu,

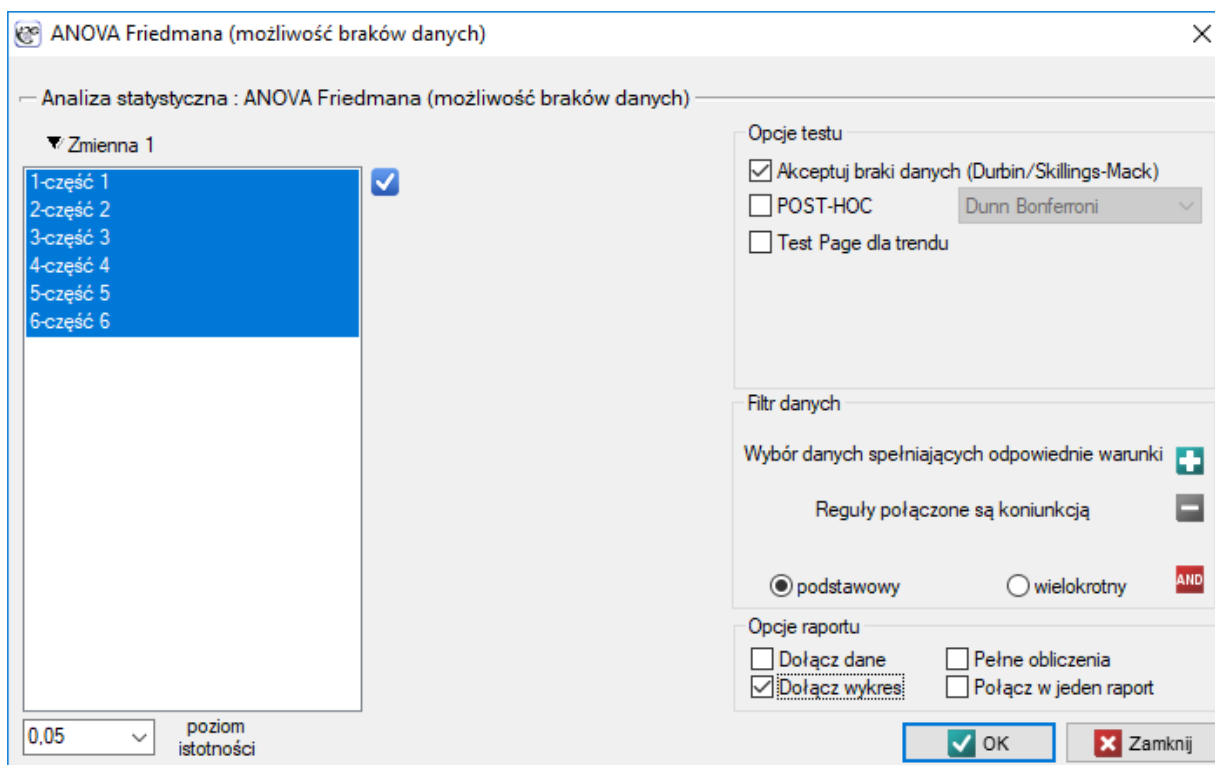
Σ_0 - macierz wyznaczająca kowariancje dla A przy prawdziwości \mathcal{H}_0 [153].

Gdy każda para pomiarów występuje równocześnie dla przynajmniej jednej obserwacji, statystyka ta ma asymptotycznie (dla dużych licznosci) [rozkład \$\chi^2\$](#) z $k - 1$ stopniami swobody.

Wyznaczoną na podstawie [statystyki testowej wartość \$p\$](#) porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Okno z ustawieniami opcji ANOVA Skillings-Mack wywołujemy poprzez menu Statystyka \rightarrow Testy nieparametryczne (kat. uporządkowane) \rightarrow ANOVA Friedmana (możliwość braków danych).


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.

PRZYKŁAD 14.8. (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. Ankieta składała się z 6 pytań dotyczących sposobu ilustrowania sześciu wyszczególnionych partii materiału. Studenci mogli dokonać oceny w skali pięciostopniowej, gdzie 1-zupełnie niezrozumiały sposób przedstawiana materiału, 5 - bardzo jasny i ciekawy sposób ilustrowania materiału. Uzyskane w ten sposób dane okazały się nie być kompletne ze względu na brak odpowiedzi studentów na pytania dotyczące tej części materiału, na której byli nieobecni. W 30-osobowej grupie wypełniającej ankietę, tylko 15 osób udzieliło kompletu odpowiedzi. Wykonanie analizy nie uwzględniającej braków danych (w tym przypadku analizy Friedmana) będzie miało ograniczoną moc poprzez tak drastyczne obcięcie liczności grupy i nie doprowadzi do wykrycia istotnych różnic. Braki danych nie były zaplanowane i nie występują w bloku zbalansowanym, a więc nie można wykonać tego zadania przy użyciu analizy Durbina wraz z jego testem POST-HOC.

Hipotezy:

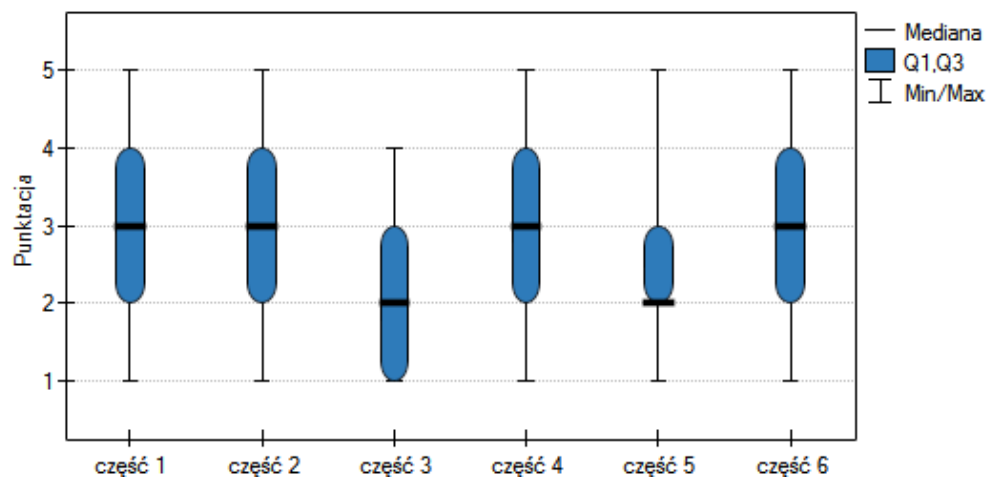
\mathcal{H}_0 : brak jest istotnej różnicy w ocenach poszczególnych części materiału przez studentów,

\mathcal{H}_1 : przynajmniej jedna część materiału jest oceniana przez studentów inaczej.

Wyniki analizy ANOVA Skillings-Mack przedstawia następujący raport:

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

Uzyskaną wartość p należy traktować ostrożnie ze względu na możliwe rangi wiązane. Jednak w przypadku tego badania wartość $p = 0.006739$ znajduje się znacznie poniżej przyjętego poziomu istotności $\alpha = 0.05$, co świadczy o występowaniu istotnych różnic. Różnice w odpowiedziach można obserwować na wykresie, nie ma jednak możliwości przeprowadzenia analizy POST-HOC dla tego testu.



14.2.8 Test chi-kwadrat dla wielowymiarowych tabel kontyngencji

Test χ^2 dla wielowymiarowych tabel kontyngencji (*ang. Chi-square test for multidimensional contingency tables*) jest rozszerzeniem testu χ^2 dla tabel ($R \times C$) na więcej niż dwie cechy.

Podstawowe warunki stosowania:

- pomiar na **skali nominalnej** - ewentualne uporządkowanie kategorii nie jest brane pod uwagę,

- model niezależny,
- duże licznosci oczekiwane według interpretacji Cochra (1952)[40].

Hipotezy:

$$\mathcal{H}_0 : O_{ij\dots} = E_{ij\dots} \text{ dla wszystkich kategorii,}$$

$$\mathcal{H}_1 : O_{ij\dots} \neq E_{ij\dots} \text{ dla przynajmniej jednej kategorii,}$$

gdzie:

$O_{ij\dots}$ i $E_{ij\dots}$ to licznosci obserwowane w tabeli kontyngencji i odpowiadajace im licznosci oczekiwane.

Statystyka testowa ma postac:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \sum \dots \sum \frac{(O_{ij\dots} - E_{ij\dots})^2}{E_{ij\dots}}.$$

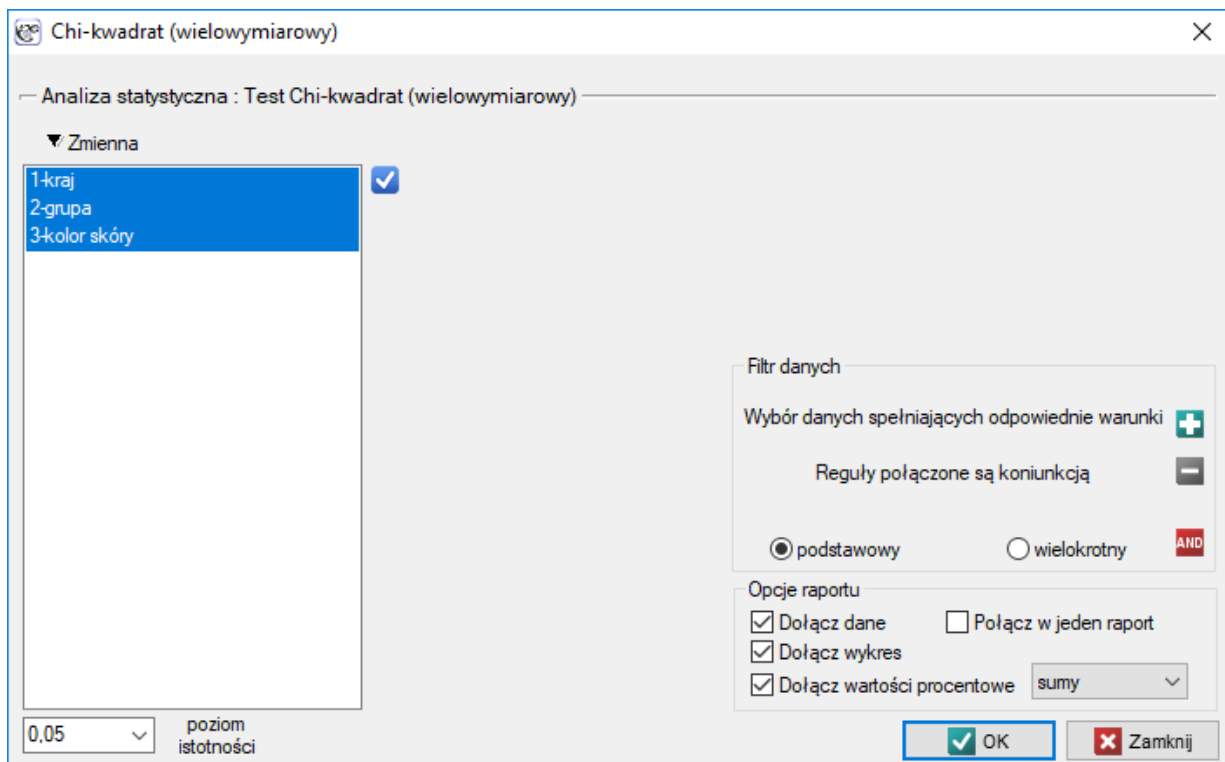
Statystyka ta ma asymptotycznie (dla duzych licznosci oczekiwanych) rozklad χ^2 z liczbą stopni swobody wyznaczaną według wzoru: $df = (r - l)(c - 1)(l - 1) + (r - l)(c - 1) + (r - 1)(l - 1) + (c - 1)(l - 1)$ - dla tabeli o 3 wymiarach.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

$$\text{jezeli } p \leq \alpha \implies \text{odrzucamy } \mathcal{H}_0 \text{ przyjmujac } \mathcal{H}_1,$$

$$\text{jezeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucic } \mathcal{H}_0.$$

Okno z ustawieniami opcji wielowymiarowego testu chi-kwadrat wywołujemy poprzez menu Statystyka → Testy nieparametryczne (kat. nieuporzadkowane) → chi-kwadrat (wielowymiarowy) lub poprzez Kreator.



Uwaga!

Test ten jest możliwy do wyliczenia tylko na podstawie danych surowych.

14.2.9 ANOVA Q-Cochrana

Analiza wariancji Q-Cochrana oparta na teście Q-Cochrana (*ang.* *Q-Cochran test*) opisana została przez Cochrana (1950)[39]. Test ten jest rozszerzeniem testu McNemara do $k \geq 2$ grup zależnych. Służy do weryfikacji hipotezy o symetryczności pomiędzy wynikami kilkukrotnych pomiarów $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ cechy X . Badana cecha może mieć tylko 2 wartości, do których (dla potrzeb analizy) przypisywane są liczby 0 i 1.

Podstawowe warunki stosowania:

- pomiar na **skali nominalnej** (zmiennie dychotomiczne, czyli zmiennie o 2 kategoriach),
- **model zależny**.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \text{wszystkie "niezgodne" licznosci obserwowane są sobie równe,} \\ \mathcal{H}_1 &: \text{nie wszystkie "niezgodne" licznosci obserwowane są równe,} \end{aligned}$$

gdzie:

"niezgodne" licznosci obserwowane, to **licznosci obserwowane** wyliczone, gdy wartość badanej cechy jest różna w kolejnych pomiarach.

Statystyka testowa ma postać:

$$Q = \frac{(k-1)(kC - T^2)}{kT - R}$$

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 licznosci) **rozkład χ^2** z liczbą stopni swobody wyznaczaną według wzoru: $df = k - 1$.

Wyznaczoną na podstawie **statystyki testowej wartości p** porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Testy POST-HOC

Wprowadzenie do **kontrastów i testów POST-HOC** przeprowadzone zostało w rozdziale 14.1.2 dotyczącym jednoczynnikowej analizy wariancji.

Test Dunna

Stosowany dla porównań prostych (licznosc w poszczególnych pomiarach zawsze jest taka sama).

Hipotezy:

Przykład - **porównania proste** (dla różnicy proporcji 1 wybranej pary pomiarów):

$$\begin{aligned} \mathcal{H}_0 &: \text{wybrane "niezgodne" licznosci obserwowane są sobie równe,} \\ \mathcal{H}_1 &: \text{wybrane "niezgodne" licznosci obserwowane są różne,} \end{aligned}$$

(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:

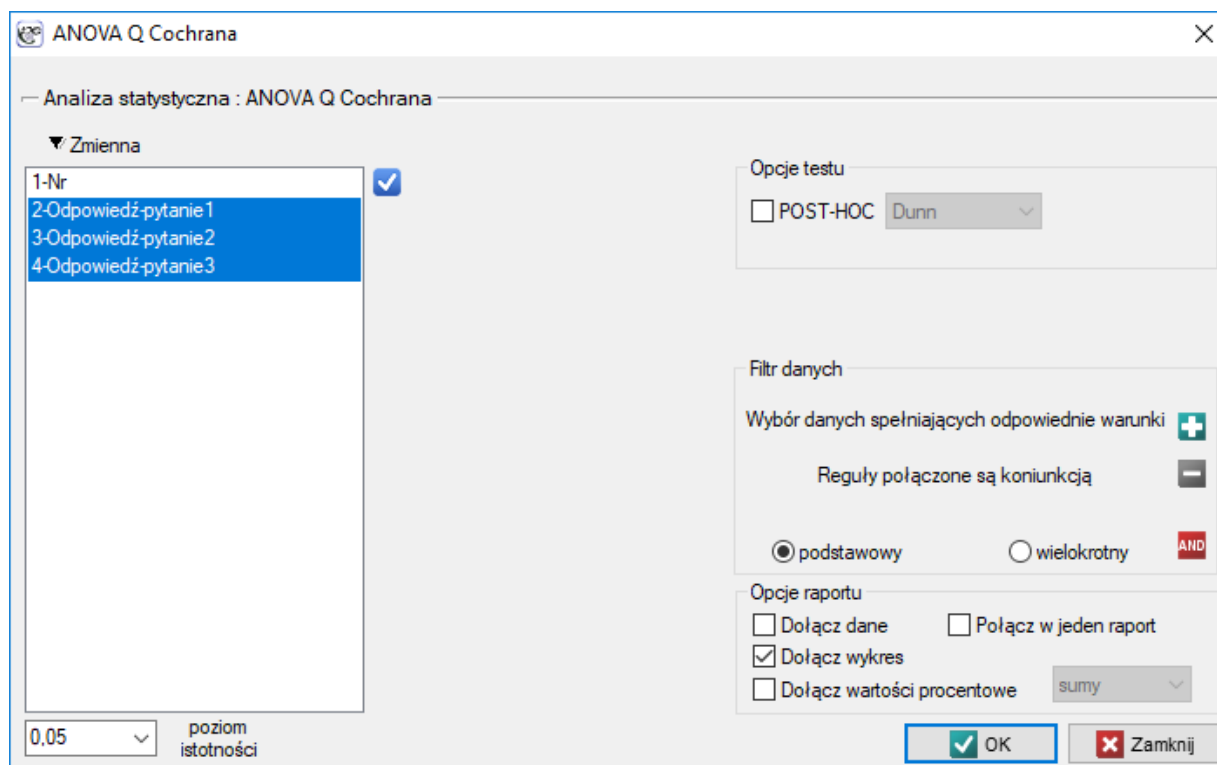
$$Z = \frac{\sum_{j=1}^k c_j p_j}{\sqrt{2 \frac{kT - R}{n^2 k(k-1)}}}$$

gdzie:

p_j – proporcja j -tego pomiaru ($j = 1, 2, \dots, k$),

Statystyka ta ma asymptotycznie (dla dużych licznosci) **rozkład normalny**, a **wartość p** jest korygowana o **liczbę możliwych porównań prostych c** .

Okno z ustawieniami opcji ANOVA Q Cochрана wywołujemy poprzez menu Statystyka → Testy nieparametryczne (kat. nieuporządkowane) → ANOVA Q Cochрана lub poprzez [Kreator](#).



ANOVA Q Cochрана

Analiza statystyczna : ANOVA Q Cochрана

Zmienna

1-Nr

2-Odpowiedź-pytanie1

3-Odpowiedź-pytanie2

4-Odpowiedź-pytanie3

Opcje testu

POST-HOC Dunn

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

Dołącz wartości procentowe sumy

0,05 poziom istotności

OK Zamknij

Uwaga!

Test ten jest możliwy do wyliczenia tylko na podstawie **danych surowych**.

PRZYKŁAD 14.9. (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 daje odpowiedzi na 3 pytania


zawarte w teście. Następnie sprawdzamy poprawność tych odpowiedzi (osoba może odpowiedzieć poprawnie lub błędnie). Wyniki zawiera poniższa tabela:

Nr	Odpowiedź-pytanie1	Odpowiedź-pytanie2	Odpowiedź-pytanie3
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
17	błędna	poprawna	błędna
18	błędna	poprawna	błędna
19	błędna	błędna	błędna
20	poprawna	poprawna	błędna

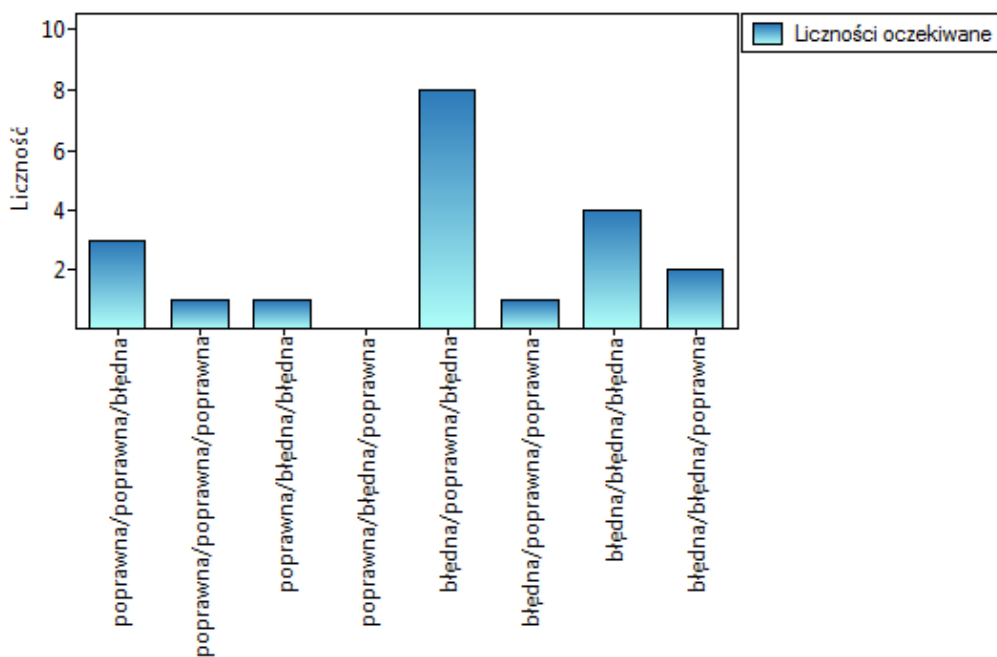
Hipotezy:

- \mathcal{H}_0 : poszczególne pytania uzyskały taką samą liczbę poprawnych odpowiedzi udzielonych przez osoby reprezentujące badaną populację,
 \mathcal{H}_1 : liczba poprawnych odpowiedzi udzielonych przez osoby reprezentujące badaną populację jest inna w poszczególnych pytaniach testu.

ANOVA Q Cochraha	
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

Porównując wartość $p = 0.007699$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że pytanie 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)			
	Odpowiedź:	Odpowiedź:	Odpowiedź:
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	



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).

15 WIELOKROTNE PORÓWNANIA

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, to oznacza, 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 (zwiększanie) wartości p testów. Najbardziej znaną korektą jest korekta Bonferroniego[2], która jest jednocześnie najbardziej konserwatywna. Korekta Sidaka (1967)[151] jest nieco bardziej liberalna. Obie korekty doczekały się kilku sekwencyjnych ulepszeń zwiększających ich moc. W programie wykorzystano wstępującą krokową procedurę Holma (1979)[80] oraz zstępującą krokową procedurę Hochberga (1988)[79]. Największą mocą wśród zaproponowanych korekt charakteryzuje się zmieniona przez Benjamini (1995)[19] procedura Hochberga, która nie kontroluje bezpośrednio błędu α ale minimalizuje oczekiwany odsetek fałszywych różnic które występują wśród różnic wykrytych.

Jeżeli liczbę testowanych hipotez oznaczymy przez c , wówczas korekty dotyczące wielokrotnych porównań będzie można opisać następująco:

- **Korekta Bonferroniego**

Polega na pomnożeniu każdego z prawdopodobieństw testowych przez całkowitą liczbę przeprowadzonych badań (lub podzieleniu poziomu istotności przez tą liczbę).

Korekta wartości p :

$$P_{(Bonferroni,i)} = p_i \cdot c$$

Korekta poziomu istotności α :

$$\alpha_{(Bonferroni,i)} = \frac{\alpha_i}{c}$$

- **Korekta Sidaka**

Korekta ta ma większą moc niż korekta Bonferroniego (dlatego jest coraz chętniej wykorzystywana).

Korekta wartości p :

$$P_{(Sidak,i)} = 1 - (1 - p_i)^c$$

Korekta poziomu istotności α :

$$\alpha_{(Sidak,i)} = 1 - (1 - \alpha_i)^{1/c}$$

- **Korekta Bonferroniego-Holma**

Polega na wykorzystaniu kilkukrokowej procedury Holma do poprawki Bonferroniego. Procedura ta rozpoczyna się od posortowania w sposób rosnący kolejnych wartości p_i (p_1, p_2, \dots, 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, które testowane są po pierwszej nieistotnej statystycznie wartości p_i , są również nieistotne.

Korekta wartości p :

$$P_{(Bonferroni,i)} = p_i \cdot (c - i + 1)$$

- **Korekta Sidaka-Holma**

Polega na wykorzystaniu kilkukrokowej procedury Holma do poprawki Sidaka. Procedura ta rozpoczyna się od posortowania w sposób rosnący kolejnych wartości p_i (p_1, p_2, \dots, p_c). Następnie zastosowana jest korekta Sidaka do każdej kolejnej wartości p_i (przy odpowiednim zmniejszeniu liczby hipotez jaka została do przetestowania). W rezultacie wszystkie hipotezy, które testowane

są po pierwszej nieistotnej statystycznie wartości p_i , są również nieistotne.

Korekta wartości p :

$$p_{(Sidak,i)} = 1 - (1 - p_i)^{c-i+1}$$

- **Korekta Bonferroniego-Hochberga**

Polega na wykorzystaniu kilkukrokowej procedury Hochberga do poprawki Bonferroniego. Procedura ta rozpoczyna się od posortowania w sposób malejący kolejnych wartości p_i (p_c, p_{c-1}, \dots, 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, które testowane są po pierwszej nieistotnej statystycznie wartości p_i , są również nieistotne.

Korekta wartości p :

$$p_{(Bonferroni,i)} = p_i \cdot (c - i + 1)$$

- **Korekta Sidaka-Hochberg**

Polega na wykorzystaniu kilkukrokowej procedury Hochberga do poprawki Sidaka. Procedura ta rozpoczyna się od posortowania w sposób malejący kolejnych wartości p_i (p_c, p_{c-1}, \dots, p_1). Następnie zastosowana jest korekta Sidaka do każdej kolejnej wartości p_i (przy odpowiednim zmniejszeniu liczby hipotez jaka została do przetestowania). W rezultacie wszystkie hipotezy, które testowane są po pierwszej nieistotnej statystycznie wartości p_i , są również nieistotne.

Korekta wartości p :

$$p_{(Sidak,i)} = 1 - (1 - p_i)^{c-i+1}$$

- **Korekta Benjamini-Hochberg**

Polega na wykorzystaniu kilkukrokowej procedury Hochberga do poprawki Benjamini będącej zmodyfikowaną wersją poprawki Bonferroniego. Procedura ta rozpoczyna się od posortowania w sposób malejący kolejnych wartości p_i (p_c, p_{c-1}, \dots, 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, które testowane są po pierwszej nieistotnej statystycznie wartości p_i , są również nieistotne.

Korekta wartości p :

$$p_{(BH,i)} = p_i \frac{c}{i}$$

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.

Korekta wielokrotnych porównań

Analiza statystyczna : Korekta wielokrotnych porównań

Zmienna

1-wartość p (nieskorygowana) ✓

0,05 poziom istotności

Opcje testu

Bonferroni
 Bonferroni-Holm
 Bonferroni-Hochberg
 Sidak
 Sidak-Holm
 Sidak-Hochberg
 Benjamini-Hochberg

Filtr danych

Wybór danych spełniających odpowiednie warunki +

Reguły połączone są koniunkcją -

podstawowy wielokrotny AND

Opcje raportu

Dołącz wykres Połącz w jeden raport

OK Zamknij

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 stanowią 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.

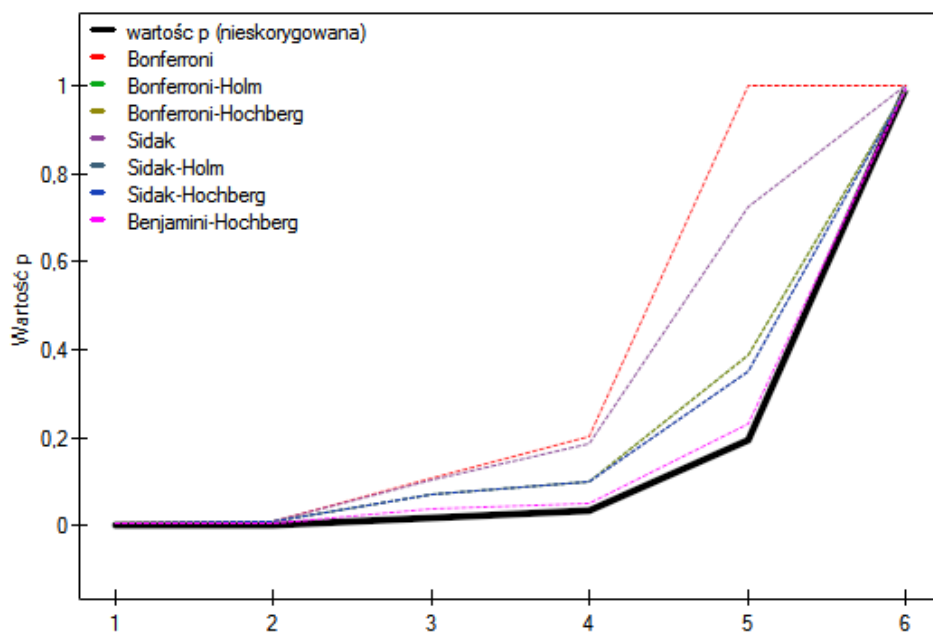
Przykład (14.4) c.d. (satysfakcjaZpracy.pqs)

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



W rezultacie uznano za istotne statystycznie różnice dotyczące satysfakcji z pracy dla dwóch par porównań (wyszktałcenie podstawowe vs średnie i wyszktałcenie podstawowe vs wyższe). Tylko przy zastosowaniu korekty Benjamini-Hochberg różnice można zlokalizować w trzech parach porównań.

INNE PRZYKŁADY: przykład (16.1), przykład (16.2)

16 HOTELLING T-KWADRAT

Analiza Hotellinga zaproponowana została przez Hotellinga (1931)[81] (1947)[82] i Mahalanobisa (1930[107], 1936[108]) jest rozszerzeniem testów t-Studenta do dziedziny wielu zmiennych. W rezultacie jednoczesnej analizie możemy poddawać wiele różnych cech badanych obiektów.

16.1 Test T-kwadrat Hotellinga dla pojedynczej próby

Służy do weryfikacji hipotezy, że k zmiennych w badanej populacji X_1, X_2, \dots, X_k charakteryzuje się średnimi zadanymi przez badacza.

Podstawowe warunki stosowania:

- pomiar na [skali interwałowej](#),
- wielowymiarowy rozkład normalny lub [normalność rozkładu](#) każdej badanej zmiennej.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \mu = \mu_0, \\ \mathcal{H}_1 &: \text{nie wszystkie } \mu_j \text{ są równe zadanym wartościom,} \end{aligned}$$

gdzie:

$\mu = (\mu_1, \mu_2, \dots, \mu_k)$ - średnie zmiennych w populacji reprezentowanej przez próbę,
 $\mu_0 = (\mu_{01}, \mu_{02}, \dots, \mu_{0k})$ - zadane przez badacza wartości średnich.

Statystyka testowa ma postać:

$$F = \frac{n - k}{k(n - 1)} T^2$$

gdzie:

$n = n_1 = n_2 = \dots = n_k$ - licznosci poszczególnych zmiennych w próbie,
 $T^2 = n(\bar{x} - \mu_0)^T S^{-1}(\bar{x} - \mu_0)$ - pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznosciach),
 $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ - 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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności średnich:

$$\bar{x}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha, df_1, df_2}} \cdot SE_j$$

lub przedziały z poprawką Bonferroniego, w celu sprawdzenia czy znajduje się w nich zadana wartość. Jeśli bowiem zadana wartość znajduje się w wyznaczonym przedziale to oznacza, że w rzeczywistości średnia danej zmiennej może być równa tej zadanej wartości. Stosując tą 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.

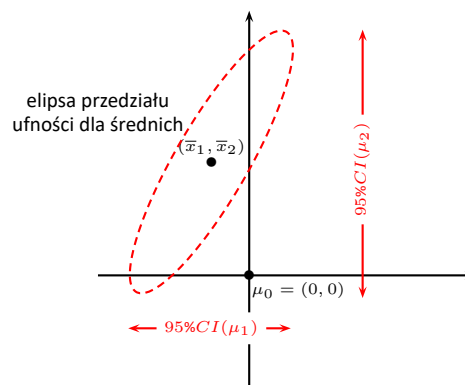
Wykonujemy wówczas porównania testem *t*-Studenta dla pojedynczej próby oddzielnie dla poszczególnych zmiennych. Niestety, nie uwzględnimy tym samym wzajemnych powiązań, ale uzyskane wartości *p* testu *t*-Studenta możemy skorygować w dziale [Wielokrotne porównania](#).

Uwaga!

Zasada działania testu Hotellinga jest tożsama z budową "wielowymiarowej elipsy" przedziałów ufności wokół centrum wyznaczonego przez średnie. Przez co, 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.



Okno z ustawieniami opcji testu Hotellinga dla pojedynczej próby wywołujemy poprzez menu Statystyka → Testy parametryczne → T-kwadrat Hotellinga

T-kwadrat Hotellinga

Analiza statystyczna : Test Hotellinga T-kwadrat dla pojedynczej próby

▼ Zmienna

- 1-LP
- 2-Płeć
- 3-Sport
- 4-RBC
- 5-WBC
- 6-Hc
- 7-Hg
- 8-Ferr
- 9-SSF
- 10-%Bfat
- 11-LBM
- 12-Wzrost
- 13-Masa ciała

▼ Średnia hipotetyczna

4-RBC	4.8
7-Hg	15
8-Ferr	100

Opcje testu

Testy jednowymiarowe (t-Studenta)

Filtr danych

zmienna	warunek	wartość
3-Sport	=	wodne

Opcje raportu

Dołącz dane Statystyki opisowe

Dołącz wykres Połącz w jeden raport

0,05 poziom istotności

PRZYKŁAD 16.1. (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 [$\mu\text{g/l}$].

Chcemy wiedzieć, na ile bliskie oczekiwany 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[$\mu\text{g/l}$].

Hipotezy:

\mathcal{H}_0 : średnie wartości RBC, Hg i Ferr w badanej populacji sportowców wynoszą odpowiednio: 4.8, 15, 100,

\mathcal{H}_1 : przynajmniej jeden z badanych parametrów ma średnią w populacji inną od zadanej wartości.

Ponieważ arkusz danych zawiera informacje o badanych parametrach dla większej grupy sportowców, w oknie analizy osoby, które uprawiają sporty wodne wskazujemy poprzez filtr danych.

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

Porównując wartość $p = 0.000708$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że 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	+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 przedziału ufności ale również przedział dla hemoglobiny.

Przedziały ufności [Benferroni]			
	Średnia	-95% CI	+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 przekopiarować te wartości do jednej kolumny nowego arkusza danych i z menu Statystyka wybrać Korektę wielokrotnych porównań.

Testy t-Studenta						
	Liczność	-95% CI	+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

16.2 Test T-kwadrat Hotellinga dla grup zależnych

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, \dots, X_k a drugi przez Y_1, Y_2, \dots, Y_k , wówczas weryfikujemy hipotezę że populacyjne średnie zmiennych z pierwszego pomiaru są takie same jak z pomiaru drugiego. Równoważnie gdy wyznaczmy różnice pomiędzy parami pomiarów d_1, d_2, \dots, d_k , hipoteza wskaże, że średnie dla różnic w badanej populacji wynoszą 0.

Podstawowe warunki stosowania:

- pomiar na [skali interwałowej](#),
- wielowymiarowy rozkład normalny lub [normalność rozkładu](#) każdej badanej zmiennej,
- [model zależny](#).

Hipotezy:

$$\mathcal{H}_0 : \mu_0 = 0,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \mu_{j0} \text{ są równe zero,}$$

gdzie:

$\mu_0 = (\mu_{01}, \mu_{02}, \dots, \mu_{0k})$ - populacyjne średnie różnic pomiaru pierwszego i drugiego.

Statystyka testowa ma postać:

$$F = \frac{n - k}{k(n - 1)} T^2$$

gdzie:

$n = n_1 = n_2 = \dots = n_k$ - licznosci poszczególnych różnic w próbie,

$T^2 = n(\bar{x} - \bar{y})^T S^{-1}(\bar{x} - \bar{y})$ - pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznosciach),

$\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ - średnie zmiennych w próbie dla pierwszego pomiaru,

$\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$ - 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 :

$$\text{jeżeli } p \leq \alpha \implies \text{odrzucaamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1,$$

$$\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0.$$

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednocześnie przedziały ufności dla różnic średnich:

$$\bar{x}_j - \bar{y}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha, df_1, df_2}} \cdot SE_{(diff)_j}$$

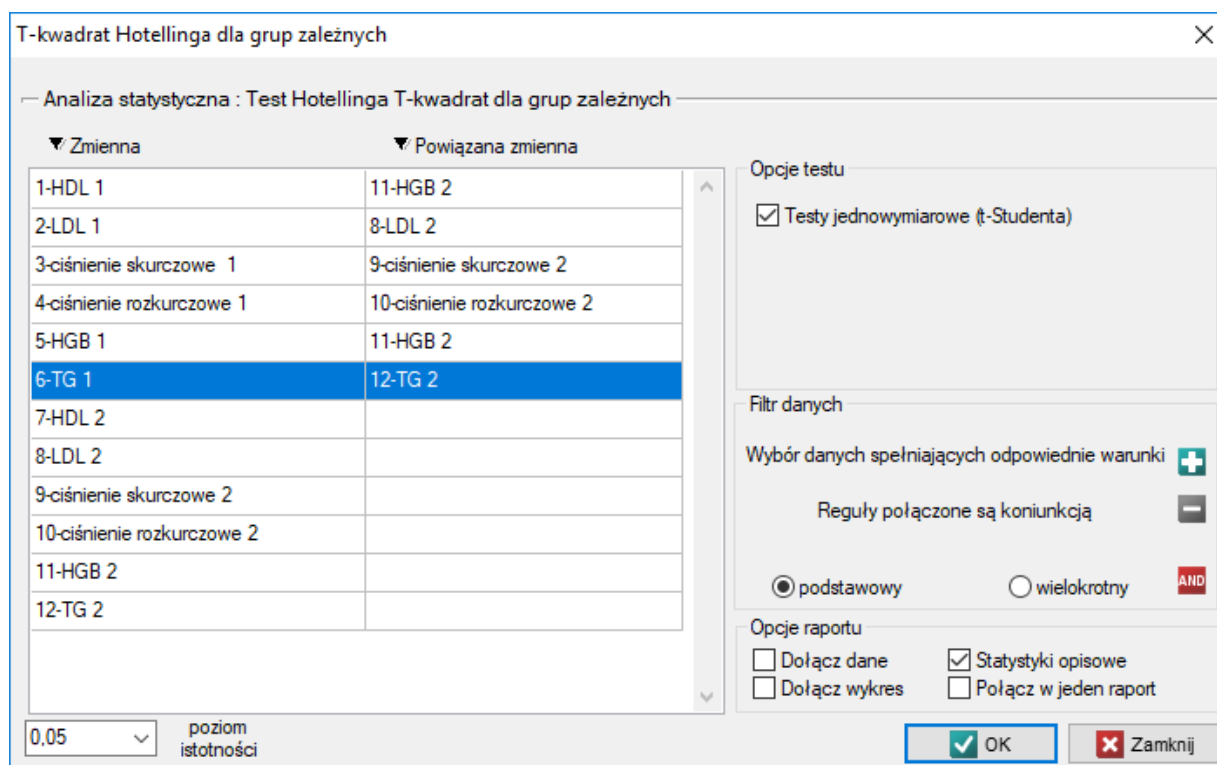
lub przedziały z poprawką Bonferroniego, w celu sprawdzenia czy znajduje się w nich wartość 0. Jeśli bowiem różnica może wynosić 0 to oznacza, że w rzeczywistości różnica pomiędzy badanymi wartościami może nie istnieć. Stosując tą metodę należy pamiętać, że wyznaczone przedziały nie uwzględniają powiązań pomiędzy zmiennymi towarzyszącymi (które uwzględni test Hotellinga) a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice możemy również zastosować podejście jednowymiarowe. Wykonujemy wówczas 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**.

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 co, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi) często nie jesteśmy w stanie uzyskać tożsamych wyników.

Okno z ustawieniami opcji testu Hotellinga dla grup zależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → T-kwadrat Hotellinga dla grup zależnych



T-kwadrat Hotellinga dla grup zależnych

Analiza statystyczna : Test Hotellinga T-kwadrat dla grup zależnych

Zmienna	Powiązana zmienna
1-HDL 1	11-HGB 2
2-LDL 1	8-LDL 2
3-ciśnienie skurczowe 1	9-ciśnienie skurczowe 2
4-ciśnienie rozkurczowe 1	10-ciśnienie rozkurczowe 2
5-HGB 1	11-HGB 2
6-TG 1	12-TG 2
7-HDL 2	
8-LDL 2	
9-ciśnienie skurczowe 2	
10-ciśnienie rozkurczowe 2	
11-HGB 2	
12-TG 2	

Opcje testu

Testy jednowymiarowe (t-Studenta)

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny **AND**

Opcje raportu

Dołącz dane Statystyki opisowe

Dołącz wykres Połącz w jeden raport

0,05 poziom istotności

PRZYKŁAD 16.2. (plik nadcisnienie.pqs) W grupie osób chorujących na nadciśnienie badano wpływ zastosowanego leczenia na zmiany wskaźników takich jak: cholesterol we frakcji HDL i LDL, hemoglobinę (HGB), trójglicerydy (TG) oraz wartości ciśnienia skurczowego i rozkurczowego krwi. Pomiarów od 44 pacjentów pobrano dwukrotnie (przed leczeniem i po 3 miesiącach stosowania leczenia). Następnie porównano uzyskane wyniki.

Hipotezy:

- \mathcal{H}_0 : średnie wartości badanych parametrów nie zmieniają się na skutek leczenia
 \mathcal{H}_1 : w badanej populacji przynajmniej jeden parametr zmienia się na skutek leczenia

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

Porównując wartość $p = 0.00024$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że średnie wartości badanych parametrów nie pozostają na tym samym poziomie (ich 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, a 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 s1	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

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 s1	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

Podjęcie 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

16.3 Test T-kwadrat Hotellinga dla grup niezależnych

Służy do weryfikacji hipotezy o równości średnich badanych k zmiennych X_1, X_2, \dots, X_k z populacji pierwszej i średnich tych samych k zmiennych Y_1, Y_2, \dots, Y_k z populacji drugiej.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- 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 licznosciach - gdy warunek nie jest pełniony, powinna zostać wyznaczona **korekta**.

Hipotezy:

$$\mathcal{H}_0 : \mu_x = \mu_y,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \mu_{xj} = \mu_{yj},$$

gdzie:

$\mu_x = (\mu_{x1}, \mu_{x2}, \dots, \mu_{xk})$ - średnie zmiennych w populacji pierwszej,

$\mu_y = (\mu_{y1}, \mu_{y2}, \dots, \mu_{yk})$ - średnie zmiennych w populacji drugiej.

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)

$T^2 = (\bar{x} - \bar{y})^T \left(S \left(\frac{1}{n_x} + \frac{1}{n_y} \right) \right)^{-1} (\bar{x} - \bar{y})$ - pierwotna statystyka testowa Hotellinga o rozkładzie χ^2 (zalecana dla prób o dużych licznościach),

$\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ - średnie zmiennych w próbie pierwszej,

$\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$ - 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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności dla różnic średnich

$$\bar{x}_j - \bar{y}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha, df_1, df_2}} \cdot SE_{(diff)_j}$$

lub przedziały z poprawką Bonferroniego, w celu sprawdzenia czy znajduje się w nich wartość 0. Jeśli bowiem różnica może wynosić 0 to oznacza, że w rzeczywistości różnica pomiędzy badanymi wartościami może nie istnieć. Stosując tą 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. Wykonujemy wówczas 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](#).

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 co, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi) często nie jesteśmy w stanie uzyskać tożsamych wyników.

Okno z ustawieniami opcji testu Hotellinga dla grup niezależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → T-kwadrat Hotellinga dla grup niezależnych.

T-kwadrat Hotellinga dla grup niezależnych

Analiza statystyczna : Test Hotellinga T-kwadrat dla grup niezależnych

Zmienne

- 1-LP
- 2-Płeć
- 3-Sport
- 4-RBC
- 5-WBC
- 6-Hc
- 7-Hg
- 8-Ferr
- 9-SSF
- 10-%Bfat
- 11-LBM
- 12-Wzrost
- 13-Masa ciała

Zmienna grupująca

- 1-LP
- 2-Płeć
- 3-Sport
- 4-RBC
- 5-WBC
- 6-Hc
- 7-Hg
- 8-Ferr
- 9-SSF
- 10-%Bfat
- 11-LBM
- 12-Wzrost
- 13-Masa ciała

Wariancje/Kowariancje

Sprawdź równość (BOX)

Opcje testu

Korekta dla różnych kowariancji

Testy jednowymiarowe (t-Studenta)

korekta Cochran-Coxa

sprawdź równość wariancji

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane Statystyki opisowe

Dołącz wykres Połącz w jeden raport

0,05 poziom istotności

PATRZ PRZYKŁAD (16.3)

16.4 Test T-kwadrat Hotellinga dla grup niezależnych z korektą dla różnych wariancji

Poprawka dotyczy testu [T-kwadrat Hotellinga dla grup niezależnych](#) i jest wyliczana wówczas, gdy wariancje badanych zmiennych w obu populacjach są różne.

Statystyka testowa ma postać:

$$F = \frac{n - k}{k(n - 1)} T^2$$

gdzie:

$$n = n_x + n_y - 1,$$

n_x, n_y - licznosci pierwszej i drugiej próby (licznosci poszczególnych zmiennych równe) ,

$T^2 = (\bar{x} - \bar{y})^T \left(\frac{S_x}{n_x} + \frac{S_y}{n_y} \right)^{-1} (\bar{x} - \bar{y})$ - pierwotna statystyka testowa Hotellinga o rozkładzie

χ^2 (zalecana dla prób o dużych licznosciach),

$\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ - średnie zmiennych w próbie pierwszej,

$\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$ - 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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Gdy po wykonanej analizie szukamy zmiennych, których dotyczą różnice, wyznaczamy jednoczesne przedziały ufności dla różnic średnich

$$\bar{x}_j - \bar{y}_j \pm \sqrt{\frac{k(n-1)}{n-k} F_{\alpha, df_1, df_2}} \cdot SE_{(diff)_j}$$

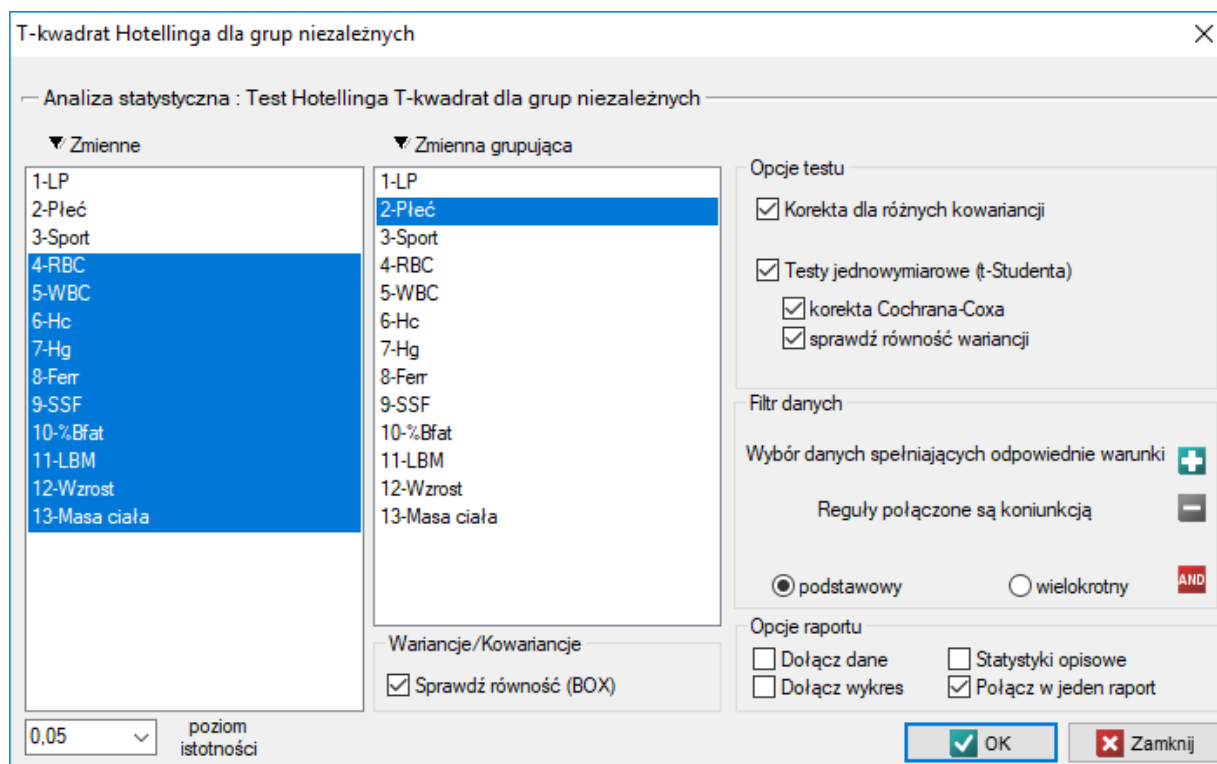
lub przedziały z poprawką Bonferroniego, w celu sprawdzenia czy znajduje się w nich wartość 0. Jeśli bowiem różnica może wynosić 0 to oznacza, że w rzeczywistości różnica pomiędzy badanymi wartościami może nie istnieć. Stosując tą metodę należy pamiętać, że wyznaczone przedziały nie uwzględniają powiązań pomiędzy zmiennymi (które uwzględnia test Hotellinga) a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice możemy również zastosować podejście jednowymiarowe. Wykonujemy wówczas porównania testem *t*-Studenta z korektą Cochran-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](#).

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 co, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi) często nie jesteśmy w stanie uzyskać tożsamych wyników.

Okno z ustawieniami opcji testu Hotellinga dla grup niezależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → T-kwadrat Hotellinga dla grup niezależnych.



PRZYKŁAD 16.3. (plik sport.pqs) Przebadano grupę sportowców by uzyskać informację o parametrach zdrowotnych takich jak:

RBC - Liczba czerwonych krwinek,	SSF - Suma fałdów skórno-tłuszczowych [mm],
WBC - Liczba białych krwinek,	% Bfat - Zawartość tłuszczu w organizmie,
Hc - Hematokryt [%],	LBM - Beztłuszczowa masa ciała [kg],
Hg - Hemoglobina [g/dl],	Wzrost [cm],
Ferr - Ferrytyna [μg/l],	Masa ciała [kg].

Chcemy wiedzieć, czy kobiety i mężczyźni, uprawiający zawodowo analizowane rodzaje sportu, różnią się poziomami tych parametrów.

Hipotezy:

- \mathcal{H}_0 : średnie analizowanych parametrów są takie same dla kobiet i mężczyzn uprawiających analizowane rodzaje dyscypliny sportu,
 \mathcal{H}_1 : przynajmniej jeden parametr ma inną średnią wartość dla porównywanych populacji.

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

Porównując wartość $p < 0.000001$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że mężczyźni i kobiety w badanej populacji mają inne wartości średnie analizowanych parametrów. Według jednoczesnych przedziałów ufności (oraz według przedziałów z korektą Bonferroniego) różnice dotyczą:

- RBC - średnio kobiety uzyskują niższe wartości,
- Hc - średnio kobiety uzyskują niższe wartości,
- Hg - średnio kobiety uzyskują niższe wartości,
- Ferr - średnio kobiety uzyskują niższe wartości,
- SSF - średnio mężczyźni uzyskują niższe wartości,

% Bfat - średnio mężczyźni uzyskują niższe wartości,
LBM - średnio kobiety uzyskują niższe wartości.
Pozostałe parametry czyli WBC, wzrost i masa ciała nie różnią się istotnie.

Przedziały ufności [jednoczesne]				
	różnica śr	-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 śr	-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

Podejście jednowymiarowe, choć nie uwzględniające wielokrotnego testowania ani wzajemnych powiązań, ze względu na swoją prostotę wykorzystywane jest najczęściej. Jego wyniki wskazują że różnice dotyczą wszystkich badanych parametrów za wyjątkiem WBC. Przy czym wyniki testu t-Studenta interpretujemy po uwzględnieniu korekty Cochran-Coxa lub bez tej korekty, w zależności od spełnienia założenia równości wariancji (wynik testu Fishera-Snedecora). Chcąc uwzględnić wielokrotne testowania należałoby w tym przypadku zastosować jedną z korekt wartości p, opisanych w dziale [Wielokrotne porównania](#).

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 c	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

16.5 Test Box'a równości macierzy kowariancji

Test ten służy do porównania dwóch lub więcej ($m \geq 2$) macierzy kowariancji opisujących niezależne populacje.

Podstawowe warunki stosowania:

- pomiar na **skali interwałowej**,
- wielowymiarowy rozkład normalny w badanych populacjach lub **normalność rozkładu** każdej badanej zmiennej w każdej populacji,
- **model niezależny**.

Hipotezy:

$$\mathcal{H}_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_m,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \Sigma_j \text{ są sobie równe } (j = 1, 2, \dots, m),$$

gdzie:

$\Sigma_1, \Sigma_2, \dots, \Sigma_m$ - populacyjne macierze kowariancji.

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,

$n = n_1 = n_2 = \dots = n_k$ - 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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Uwaga!

Należy pamiętać, że test Box'a jest szczególnie wrażliwy na złamanie założenia normalności rozkładu.

Test Box'a jest wyliczany opcjonalnie w [Hotellingu](#) dla grup niezależnych lub w analizie [MANOVA](#).

PATRZ PRZYKŁAD (16.3)

17 JEDNOCZYNNIKOWA MANOVA

Wcześniej zalecamy zapoznanie się z analizą [T-kwadrat Hotelling'a](#).

Jednoczynnikowa MANOVA (ang. *multivariate analysis of variance*), jest rozszerzeniem [jednoczynnikowej ANOVA dla grup niezależnych](#). Służy do weryfikacji hipotezy o równości średnich badanych k zmiennych w kilku ($m \geq 2$) populacjach.

Pozostanie przy metodzie ANOVA polega na wielokrotnym (oddzielnie dla każdej zmiennej) porównaniu ($m \geq 2$) populacji, nie uwzględniając wzajemnego skorelowania zmiennych. Analiza typu MANOVA natomiast bada różnice pomiędzy populacjami jednocześnie dla wielu zmiennych, uwzględniając przy tym ich skorelowanie. Ponadto podejście MANOVA jest stosowane jako alternatywa dla [ANOVA dla grup zależnych](#), gdyż nie wymaga spełnienia założenia sferyczności.

Podstawowe 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 licznosciach.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \mu_1 = \mu_2 = \dots = \mu_m, \\ \mathcal{H}_1 &: \text{nie wszystkie } \mu_i \text{ są sobie równe,} \end{aligned}$$

gdzie:

$$\begin{aligned} \mu_i &= (\mu_{i1}, \mu_{i2}, \dots, \mu_{ik}) - \text{średnie zmiennych w } i\text{-tej populacji,} \\ (i &= 1, 2, \dots, m), \\ (j &= 1, 2, \dots, k). \end{aligned}$$

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 natomiast jest najmniej konserwatywnym z trzech zaproponowanych testów. Pionierskie prace nad rozwojem tych technik rozpoczęli: Wilks (1932)[174], Pillai(1955)[132], Lawley(1938)[97], Hotelling(1951)[83], i Roy(1939)[140].

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, z których pierwsza związana jest z analizowaną hipotezą i oznaczana przez H (w tym przypadku jest to 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

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, b, c - współczynniki zależne od liczby analizowanych zmiennych i liczby porównywanych populacji.

Ślad Hotelling-Lawley

Ślad Hotelling-Lawley definiowany jest następująco:

$$T_0 = \text{trace}(HE^{-1})$$

Statystyka testowa ma postać:

$$F = \frac{T_0^2}{s} \frac{df_2}{df_1}$$

gdzie:

$$df_1 = s(2t + s + 1), df_2 = 2(su + 1),$$

s, t, u - współczynniki zależne od liczby analizowanych zmiennych i liczby porównywanych populacji.

Ślad Pillai-Bartlett

Ślad Pillai-Bartlett definiowany jest następująco:

$$V = \text{trace}(H(H + E)^{-1})$$

Statystyka testowa ma postać:

$$F = \frac{V}{s - V} \frac{df_2}{df_1}$$

gdzie:

$$df_1 = s(2t + s + 1), df_2 = s(2u + s + 1),$$

s, t, u - 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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Wielkość efektu - cząstkowa η^2

Wielkość ta określa proporcję wariancji wyjaśnionej do wariancji całkowitej związanej z danym czynnikiem. W modelu jednoczynnikowej MANOVA dla grup niezależnych wskazuje jaka część wewnątrzgrupowej zmienności wyników może być przypisana badanemu czynnikowi wyznaczającemu grupy niezależne.

$$\eta^2 = \frac{F \cdot df_1}{F \cdot df_1 + df_2}$$

Wielkość efektu - kontrasty, analiza jednowymiarowa

Gdy wykonywana analiza ma porównać wybrane populacje, lub wybrany układ populacji, wówczas przeprowadzamy analizę kontrastów. Analiza ta jest analogiczna do [kontrastów](#) w analizie jednowymiarowej ale uwzględnia wzajemne powiązanie zmiennych.

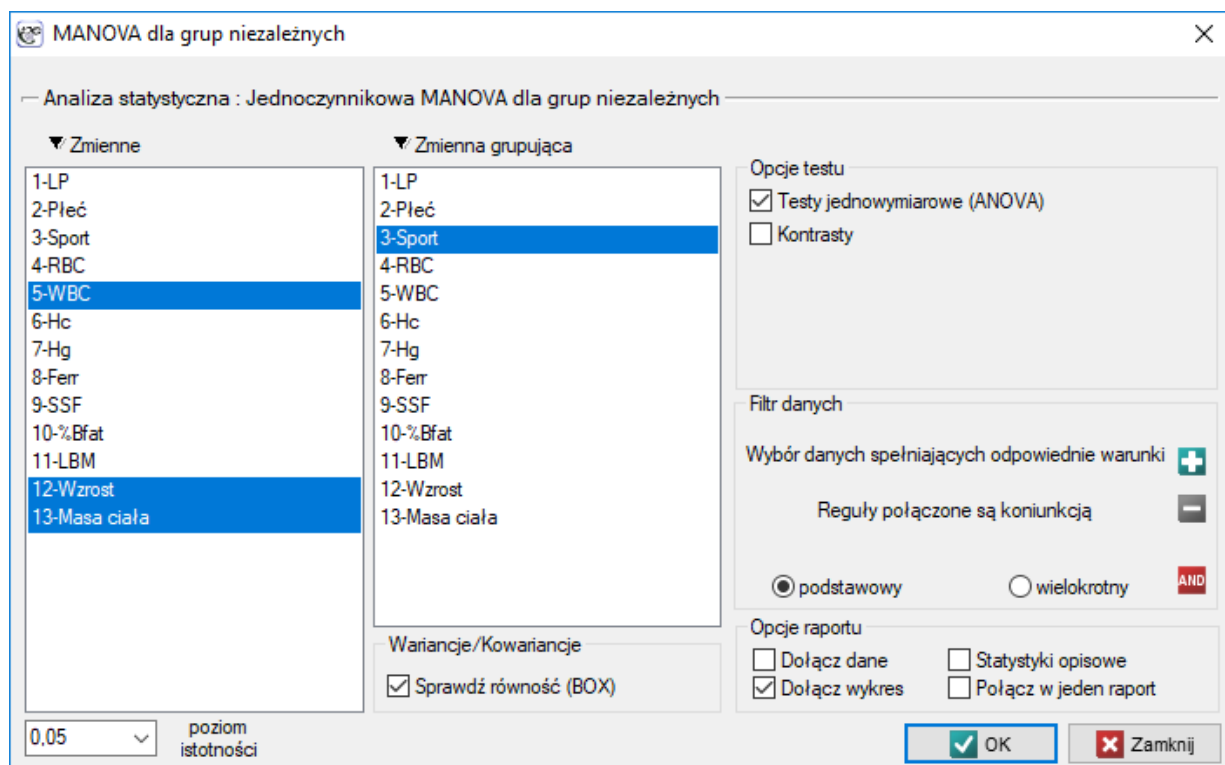
Dla wielkości efektu można także wyznaczyć **jednoczesne przedziały ufności** lub przedziały ufności z poprawką Bonferroniego. Stosując te przedziały należy jednak pamiętać, że nie uwzględniają one powiązań pomiędzy poszczególnymi zmiennymi (które uwzględnia MANOVA) a jedynie wielokrotne testowanie.

Szukając zmiennych, których dotyczą różnice, możemy również zastosować podejście jednowymiarowe. Wykonujemy wówczas 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**.

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 co, stosując analizę jednowymiarową (nie uwzględniającą wzajemnych powiązań między zmiennymi) często nie jesteśmy w stanie uzyskać tożsamych wyników.

Okno z ustawieniami opcji MANOVA dla grup niezależnych wywołujemy poprzez menu Statystyka → Testy parametryczne → Jednoczynnikowa MANOVA dla grup niezależnych



MANOVA dla grup niezależnych

Analiza statystyczna : Jednoczynnikowa MANOVA dla grup niezależnych

Zmienne

- 1-LP
- 2-Płeć
- 3-Sport
- 4-RBC
- 5-WBC
- 6-Hc
- 7-Hg
- 8-Ferr
- 9-SSF
- 10-%Bfat
- 11-LBM
- 12-Wzrost
- 13-Masa ciała

Zmienna grupująca

- 1-LP
- 2-Płeć
- 3-Sport
- 4-RBC
- 5-WBC
- 6-Hc
- 7-Hg
- 8-Ferr
- 9-SSF
- 10-%Bfat
- 11-LBM
- 12-Wzrost
- 13-Masa ciała

Opcje testu

- Testy jednowymiarowe (ANOVA)
- Kontrasty

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny **AND**

Opcje raportu

- Dołącz dane
- Statystyki opisowe
- Dołącz wykres
- Połącz w jeden raport

0,05 poziom istotności

PRZYKŁAD 17.1. (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:

- 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 Box'a ($p=0.63025$) pozwala na wyliczenie Analiz typu MANOVA.

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

Istotność współczynników: Lambda Wilksa, śladu Hotelling-Lawley oraz śladu Pillai-Bartlett pozwala twierdzić, że badane populacje sportowców różnią się co do tych parametrów. By określić 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

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 uwagę na 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 dwu 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, -1 dla bieżni i dyscyplin określanych jako wodne.

Kontrasty dla analizy wielowymiarowej:	
Kontrast dla grupy [zespolowe]	2
Kontrast dla grupy [bieznia]	-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	-15,009595	3,608457
Waga	-11,818715	-24,740085	1,102659

Przedziały ufności efektu [Bonferroni]			
Grupa	Efekt	-95% CI	+95% CI
WBC	1,373722	0,006633	2,740812
Wzrost	-5,700571	-13,6164	2,215257
Waga	-11,818715	-22,806261	-0,831169

W rezultacie uzyskana istotność współczynników: Lambda Wilksa, śladu Hotelling-Lawley oraz śladu Pillai-Bartlett ($p=0.00592$) pozwala twierdzić, że sportowcy uprawiający sporty wymagające dużego wysiłku różnią się co do tych parametrów 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.

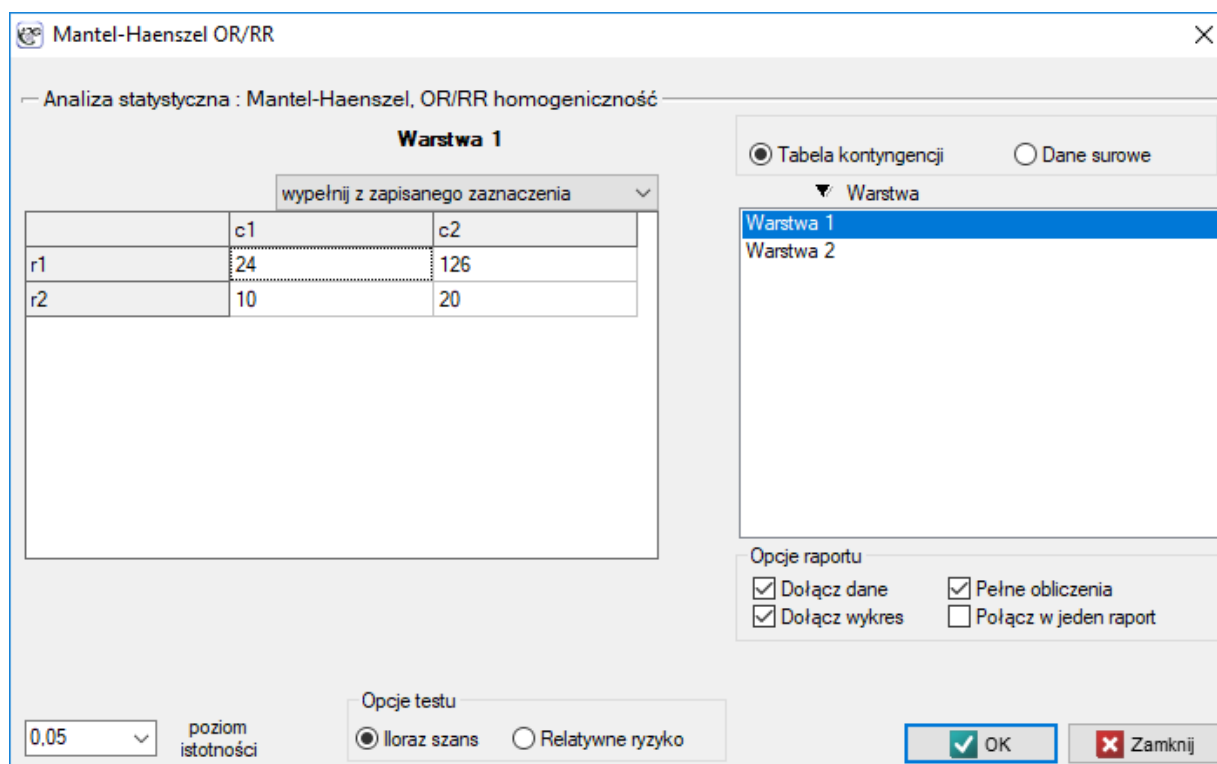
18 ANALIZA DLA WARSTW

18.1 Metoda Mantela-Haenszela dla wielu tabel 2x2

Metoda Mantela-Haenszela dla tabel 2×2 (*ang. Mantel-Haenszel method for 2×2 tables*) zaproponowana została przez Mantela i Haenszela (1959)[110] a następnie rozszerzona przez Mantela (1963)[111]. Szerszy przegląd rozwoju tych metod został przeprowadzony m.in. przez Newman (2001)[124]. Metoda Mantela-Haenszela może być używana między innymi do analizy tabel 2×2 występujących w kilku ($w \geq 2$) warstwach tworzonych przez tzw. zmienną uwikłaną (*ang. confounding*). Dla kolejnych warstw ($s = 1, \dots, w$) budowane są tabele kontyngencji 2×2 licznosci obserwowanych:

Liczności obserwowane s -tej warstwy ($O_{ij}^{(s)}$)		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	$O_{11}^{(s)} + O_{21}^{(s)}$	$O_{12}^{(s)} + O_{22}^{(s)}$	$n^{(s)} = O_{11}^{(s)} + O_{12}^{(s)} + O_{21}^{(s)} + O_{22}^{(s)}$

Okno z ustawieniami opcji metod Mantel–Haenszel OR/RR wywołujemy poprzez menu Statystyka→Analiza dla warstw→Mantel–Haenszel OR/RR.



Analiza statystyczna : Mantel-Haenszel, OR/RR homogeniczność

Warstwa 1

wypełnij z zapisanego zaznaczenia

	c1	c2
r1	24	126
r2	10	20

Tabela kontyngencji Dane surowe

▼ Warstwa

Warstwa 1
Warstwa 2

Opcje raportu

Dołącz dane Pełne obliczenia
 Dołącz wykres Połącz w jeden raport

Opcje testu

Iloraz szans Relatywne ryzyko

0,05 poziom istotności

18.1.1 Iloraz szans Mantela-Haenszela

Jeśli wszystkie tabele (tworzone przez poszczególne warstwy) są homogeniczne (warunek ten można sprawdzić przy pomocy testu χ^2 homogeniczności OR), wówczas 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 na uwzględnienie wkładu (wagi), jaki do budowy wspólnego ilorazu szans wnosi każda warstwa (im bardziej liczna warstwa, tym większy ma wpływ na powstały iloraz szans).

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:

$$R = \sum_{s=1}^w \frac{O_{11}^{(s)} \cdot O_{22}^{(s)}}{n^{(s)}},$$

$$S = \sum_{s=1}^w g^{(s)}.$$

Przedział ufności dla $\log OR_{MH}$ wyznacza się na podstawie błędu standardowego (RGB – Robins–Breslow–Greenland [135][136]) wyliczonego ze wzoru:

$$SE_{MH} = \sqrt{\frac{T}{2R^2} + \frac{U+Y}{2RS} + \frac{W}{2S^2}},$$

gdzie:

$$T = \sum_{s=1}^w T^{(s)}, \quad T^{(s)} = \frac{O_{11}^{(s)} \cdot O_{22}^{(s)} \cdot (O_{11}^{(s)} + O_{22}^{(s)})}{(n^{(s)})^2},$$

$$U = \sum_{s=1}^w U^{(s)}, \quad U^{(s)} = \frac{O_{21}^{(s)} \cdot O_{12}^{(s)} \cdot (O_{11}^{(s)} + O_{22}^{(s)})}{(n^{(s)})^2},$$

$$Y = \sum_{s=1}^w Y^{(s)}, \quad Y^{(s)} = \frac{O_{11}^{(s)} \cdot O_{22}^{(s)} \cdot (O_{21}^{(s)} + O_{12}^{(s)})}{(n^{(s)})^2},$$

$$W = \sum_{s=1}^w W^{(s)}, \quad W^{(s)} = \frac{O_{21}^{(s)} \cdot O_{12}^{(s)} \cdot (O_{21}^{(s)} + O_{12}^{(s)})}{(n^{(s)})^2}.$$

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 licznosciach, tzn. gdy są spełnione obydwie 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 $s = 1, 2, \dots, w$,
- $\max(0, O_{11}^{(s)} - O_{22}^{(s)}) \geq 5$ dla wszystkich warstw $s = 1, 2, \dots, w$.

Kiedy występują wartości zerowe w tabeli stosowana jest korekta na ciągłość (powiększenie licznosci o wartość 0,5), zarówno dla licznosci obserwowanych jak i dla licznosci oczekiwanych.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: OR_{MH} = 1, \\ \mathcal{H}_1 &: OR_{MH} \neq 1. \end{aligned}$$

Statystyka testowa ma postać:

$$\chi_{MH}^2 = \frac{\left(\sum_{s=1}^w O_{11}^{(s)} - \sum_{s=1}^w E_{11}^{(s)}\right)^2}{V},$$

gdzie:

$E_{11}^{(s)} = \frac{(O_{11}^{(s)} + O_{21}^{(s)})(O_{11}^{(s)} + O_{12}^{(s)})}{n^{(s)}}$ to wartości oczekiwane w pierwszej komórce tabeli kontyngencji, dla poszczególnych warstw $s = 1, 2, \dots, w$,

$$V = \sum_{s=1}^w V^{(s)},$$

$$V^{(s)} = \frac{(O_{11}^{(s)} + O_{12}^{(s)})(O_{21}^{(s)} + O_{22}^{(s)})(O_{11}^{(s)} + O_{21}^{(s)})(O_{12}^{(s)} + O_{22}^{(s)})}{(n^{(s)})^2(n^{(s)} - 1)}.$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) **rozkład χ^2** z jednym stopniem swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Test χ^2 homogeniczności dla OR

Test χ^2 homogeniczności dla OR (*ang. Chi-square test of homogeneity for OR*) służy do weryfikacji hipotezy o tym, 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:

$$\begin{aligned} \mathcal{H}_0 &: OR_{MH} = OR^{(s)}, \text{ dla wszystkich warstw } s = 1, 2, \dots, w, \\ \mathcal{H}_1 &: OR_{MH} \neq OR^{(s)}, \text{ dla przynajmniej jednej warstwy.} \end{aligned}$$

Statystyka testowa (Breslow–Day (1980)[25], Tarone (1985)[26][158]) ma postać:

$$\chi^2 = \sum_{s=1}^w \frac{(O_{11}^{(s)} - E^{(s)})^2}{Var^{(s)}} - \frac{\left(\sum_{s=1}^w O_{11}^{(s)} - \sum_{s=1}^w E^{(s)}\right)^2}{\sum_{s=1}^w Var^{(s)}}$$

gdzie:

$E^{(s)}$ jest rozwiązaniem równania kwadratowego (solution to the quadratic equation):

$$\frac{E^{(s)}(O_{22}^{(s)} - O_{11}^{(s)} + E^{(s)})}{(O_{11}^{(s)} + O_{21}^{(s)} - E^{(s)})(O_{11}^{(s)} + O_{12}^{(s)} - E^{(s)})} = OR_{MH},$$

$$Var^{(s)} = \left(\frac{1}{E^{(s)}} + \frac{1}{O_{22}^{(s)} - O_{11}^{(s)} + E^{(s)}} + \frac{1}{O_{11}^{(s)} + O_{21}^{(s)} - E^{(s)}} + \frac{1}{O_{11}^{(s)} + O_{12}^{(s)} - E^{(s)}} \right)^{-1}.$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) **rozkład χ^2** z liczbą stopni swobody wyliczaną według wzoru: $df = w - 1$.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

PRZYKŁAD 18.1. (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

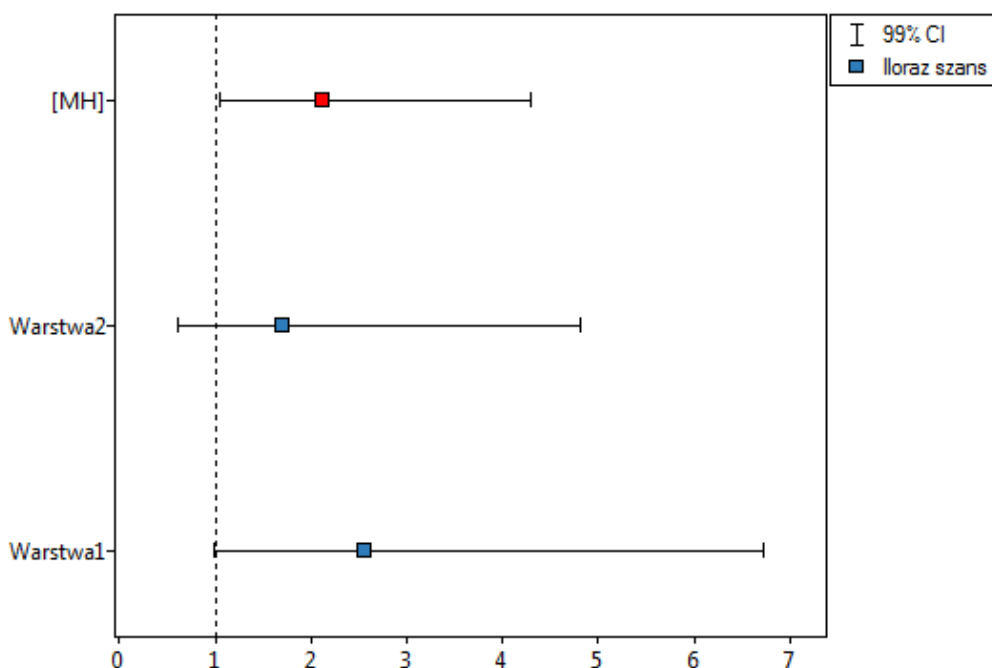
Szansa wystąpienia przeciwciał leptospirozy mieszkańców miasta i wsi jest taka sama ($OR=1$). Chcemy wiedzieć, jaka będzie szansa wystąpienia przeciwciał leptospirozy, gdy w analizie weźmiemy pod uwagę również płeć. Podzielimy więc próbę na 2 warstwy ze względu na płeć (są one zapamiętane w pliku jako **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

Płeć jest powiązana z obydwojema czynnikami (z występowaniem przeciwciał leptospirozy i z miejscem zamieszkania w Zachodnich Indiach). Jest to czynnik wnikający, 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]	
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



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]). Tabele są homogeniczne ($p=0.465049$). Możemy zatem posłużyć się wyliczonym ilorazem szans wspólnym dla obu tabel ($OR_{MH}[95\%CI]=2.13[1.24, 3.65]$). W rezultacie uzyskany wynik wskazuje, że szansa wystąpienia przeciwiał leptospirozy jest istotnie większa dla osób zamieszkujących tereny wiejskie ($p=0.005169$).

18.1.2 Relatywne ryzyko Mantela-Haenszela

Jeśli wszystkie tabele (tworzone przez poszczególne warstwy) są homogeniczne (warunek ten można sprawdzić przy pomocy testu χ^2 homogeniczności RR), wówczas 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. Zastosowanie ważonej metody zaproponowanej przez Mantela i Haenszela pozwala na uwzględnienie wkładu (wagi), jaki do budowy wspólnego Relatywnego Ryzyka wnosi każda warstwa (im bardziej liczna warstwa, tym większy ma wpływ na powstałe relatywne ryzyko).

Wagi dla każdej warstwy wyznacza się zgodnie z wzorem:

$$g^{(s)} = \frac{O_{21}^{(s)} (O_{11}^{(s)} + O_{12}^{(s)})}{n^{(s)}},$$

a relatywne ryzyko Mantela-Haenszela:

$$RR_{MH} = \frac{R}{S},$$

gdzie:

$$R = \sum_{s=1}^w \frac{O_{11}^{(s)} (O_{21}^{(s)} + O_{22}^{(s)})}{n^{(s)}},$$

$$S = \sum_{s=1}^w g^{(s)}.$$

Przedział ufności dla $\log RR_{MH}$ wyznacza się na podstawie błędu standardowego wyliczonego ze wzoru:

$$SE_{MH} = \sqrt{\frac{V}{RS}},$$

gdzie:

$$V = \sum_{s=1}^w V^{(s)},$$

$$V^{(s)} = \frac{(O_{11}^{(s)} + O_{12}^{(s)}) (O_{21}^{(s)} + O_{22}^{(s)}) (O_{11}^{(s)} + O_{21}^{(s)}) - (O_{11}^{(s)} * O_{21}^{(s)} * n^{(s)})}{(n^{(s)})^2}.$$

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 licznosciach w tabeli kontyngencji.

Hipotezy:

$$\mathcal{H}_0 : RR_{MH} = 1,$$

$$\mathcal{H}_1 : RR_{MH} \neq 1.$$

Statystyka testowa ma postać:

$$\chi_{MH}^2 = \frac{\left(\sum_{s=1}^w O_{11}^{(s)} - \sum_{s=1}^w E_{11}^{(s)}\right)^2}{V},$$

gdzie:

$E_{11}^{(s)} = \frac{(O_{11}^{(s)} + O_{21}^{(s)})(O_{11}^{(s)} + O_{12}^{(s)})}{n^{(s)}}$ to wartości oczekiwane w pierwszej komórce tabeli kontyngencji, dla poszczególnych warstw $s = 1, 2, \dots, w$.

Statystyka ta ma asymptotycznie (dla dużych licznosci) rozkład χ^2 z jednym stopniem swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ 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 o tym, że zmienna tworząca warstwy jest efektem modyfikującym, tzn. wpływa ona na wyznaczone relatywne ryzyko w taki sposób, że jest on znacząco inne dla poszczególnych warstw.

Hipotezy:

\mathcal{H}_0 : $RR_{MH} = RR^{(s)}$, dla wszystkich warstw $s = 1, 2, \dots, w$,
 \mathcal{H}_1 : $RR_{MH} \neq RR^{(s)}$, dla przynajmniej jednej warstwy.

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:

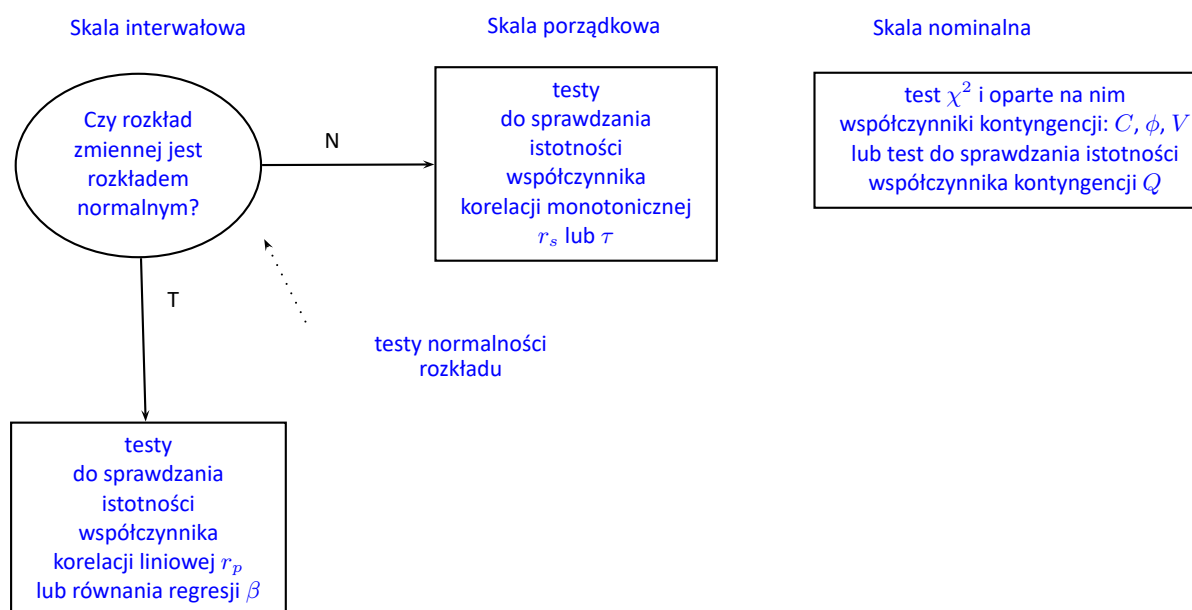
$$v^{(s)} = \left(\frac{O_{12}^{(s)}}{O_{11}^{(s)}(O_{11}^{(s)} + O_{12}^{(s)})} + \frac{O_{22}^{(s)}}{O_{21}^{(s)}(O_{21}^{(s)} + O_{22}^{(s)})} \right)^{-1}.$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) rozkład χ^2 z liczbą stopni swobody wyliczaną według wzoru: $df = w - 1$.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

19 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.

19.1 TESTY PARAMETRYCZNE

19.1.1 WSPÓŁCZYNNIKI KORELACJI LINIOWEJ

Współczynnik korelacji liniowej Pearsona r_p (ang. *Pearson product-moment correlation coefficient*, *Pearson (1896,1900)*) jest wykorzystywany do badania siły związku liniowego pomiędzy cechami. Można go wyznaczać dla skali **interwałowej**, o ile brak jest odstających pomiarów, a rozkład reszt lub badanych cech jest **rozkładem normalnym**.

$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

gdzie:

x_i, y_i - kolejne wartości cechy X i Y ,

\bar{x}, \bar{y} - średnie z wartości cechy X i cechy Y ,

n - liczność próby.

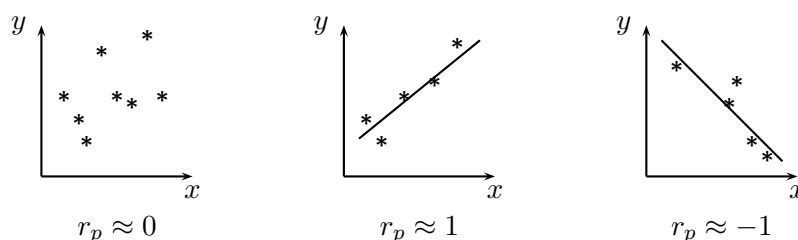
Uwaga!

R_p oznacza współczynnik korelacji Pearsona populacji, natomiast r_p w próbie.

Wartość $r_p \in \langle -1; 1 \rangle$ interpretujemy w następujący sposób:

- $r_p \approx 1$ 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 korelacji liniowej przyjmuje wartość równą lub bardzo bliską zero wówczas nie istnieje liniowa zależność między badanymi parametrami (ale może istnieć związek nieliniowy).

Wykres 19.1. Interpretacja graficzna współczynnika r_p .



Gdy jedna z badanych cech jest stała (niezależnie od zmian drugiej cechy) to nie są one związane zależnością. 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 też badana zależność wyraźnie przyjmuje kształt inny niż liniowy.

Współczynnik determinacji – r_p^2 . Wyraża procent zmienności zmiennej zależnej tłumaczony zmiennością zmiennej niezależnej.

Tworzony model korelacji przedstawia zależność liniową postaci:

$$Y = \beta X + \alpha.$$

Współczynniki β i α równania regresji liniowej możemy wyznaczyć z wzorów:

$$\beta = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \alpha = \bar{y} - \beta\bar{x}.$$

19.1.2 Istotność współczynnika korelacji Pearsona

Test do sprawdzania istotności współczynnika korelacji liniowej Pearsona (*ang. test of significance for a Pearson product-moment correlation coefficient*) służy do weryfikacji hipotezy o braku zależności liniowej pomiędzy badanymi cechami populacji i opiera się na współczynniku korelacji liniowej Pearsona wyliczonym dla próby. Im wartość r_p jest bliższa 0, tym słabszą zależnością związane są badane cechy.

Podstawowe warunki stosowania:

- pomiar na [skali interwałowej](#),
- [normalność rozkładu](#) badanych cech w populacji lub normalność reszt modelu.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : R_p &= 0, \\ \mathcal{H}_1 : R_p &\neq 0. \end{aligned}$$

Statystyka testowa ma postać:

$$t = \frac{r_p}{SE},$$

$$\text{gdzie } SE = \sqrt{\frac{1 - r_p^2}{n - 2}}.$$

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 wartości p](#) porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

19.1.3 Istotność współczynnika nachylenia prostej

Test t do sprawdzania istotności współczynników równania regresji liniowej

Test ten służy do weryfikacji hipotezy o braku zależności liniowej pomiędzy badanymi cechami populacji i opiera się na współczynniku nachylenia prostej wyliczonym dla próby. Im wartość β będzie bliższa 0, tym słabszą zależność dopasowana prosta przedstawia.

Podstawowe warunki stosowania:

- pomiar na [skali interwałowej](#),
- [normalność rozkładu](#) badanych cech w populacji lub normalność reszt modelu.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : \beta &= 0, \\ \mathcal{H}_1 : \beta &\neq 0. \end{aligned}$$

Statystyka testowa ma postać:

$$t = \frac{\beta}{SE}$$

gdzie:

$$SE = \frac{s_{yx}}{sd_x \sqrt{n-1}},$$

$$s_{yx} = sd_y \sqrt{\frac{n-1}{n-2}(1-r^2)},$$

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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Predykcja polega na przewidywaniu wartości jednej ze zmiennych (najczęściej zmiennej zależnej y_0) na podstawie wartości innej zmiennej (najczęściej zmiennej niezależnej x_0). Dokładność wyznaczonej wartości określają obliczone dla niej przedziały predykcji.

- **Interpolacja** polega na przewidywaniu wartości zadanej zmiennej leżącej wewnątrz obszaru, dla którego wykonaliśmy model regresji. Interpolacja jest więc 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. Podobnie jak w interpolacji zakłada się ciągłość funkcji wyrażającej zależność obu zmiennych.

Analiza reszt modelu - wyjaśnienie w module [Liniowa Regresja Wieloraka](#).

Okno z ustawieniami opcji zależności liniowej Pearsona wywołujemy poprzez menu Statystyka → Testy parametryczne → zależność liniowa (r-Pearsona) lub poprzez [Kreator](#).

zależność liniowa (r Pearsona)

Analiza statystyczna : Zależność liniowa Pearsona

▼ Zmienna niezależna (X) ▼ Zmienna zależna (Y)

1-wiek	1-wiek
2-wzrost	2-wzrost

0,05 poziom istotności

Opcje testu

Predykcja

Zmienna niezależna (X) 6

Zmienna zależna (Y)

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane Analiza reszt

Dołącz wykres Połącz w jeden raport

PRZYKŁAD 19.1. (plik wiek-wzrost.pqs)

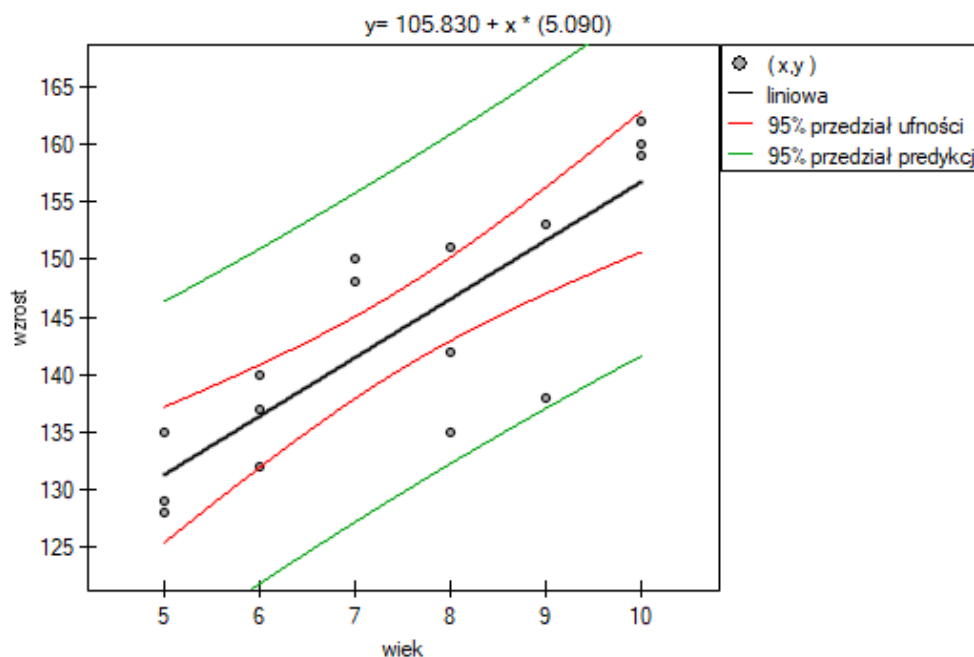
Wśród uczniów pewnej szkoły baletowej badano zależność pomiędzy wiekiem a wzrostem. W tym celu pobrano próbę obejmującą szesnaścioro dzieci i zapisano dla nich następujące wyniki pomiaru tych cech:

(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).

Hipotezy:

- \mathcal{H}_0 : nie istnieje zależność liniowa pomiędzy wiekiem a wzrostem dla populacji dzieci badanej szkoły,
- \mathcal{H}_1 : istnieje zależność liniowa pomiędzy wiekiem a wzrostem dla populacji dzieci badanej szkoły.

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
r ²	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



Porównując wartość $p=0.000069$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że 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 136.37cm.

19.1.4 Porównanie współczynników korelacji

Test t do sprawdzania równości współczynników korelacji liniowej Pearsona pochodzących z 2 niezależnych populacji

Test ten służy do weryfikacji hipotezy o równości dwóch współczynników korelacji liniowej Pearsona (R_{p1}, R_{p2}).

Podstawowe warunki stosowania:

- współczynniki r_{p1} i r_{p2} pochodzą z 2 prób pobranych z niezależnych populacji,
- współczynniki r_{p1} i r_{p2} badają zależność tych samych cech X i Y ,
- znane są liczebności obu prób n_1 i n_2 .

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: R_{p1} = R_{p2}, \\ \mathcal{H}_1 &: R_{p1} \neq R_{p2}. \end{aligned}$$

Statystyka testowa ma postać:

$$t = \frac{z_{r_{p_1}} - z_{r_{p_2}}}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}},$$

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

19.1.5 Porównanie nachylenia prostych regresji

Test t do sprawdzania równości współczynników regresji liniowej pochodzących z 2 niezależnych populacji

Test ten służy do weryfikacji hipotezy o równości dwóch **współczynników regresji liniowej** β_1 i β_2 w badanych populacjach.

Podstawowe 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 licznosci 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}

Hipotezy:

$$\mathcal{H}_0 : \beta_1 = \beta_2,$$

$$\mathcal{H}_1 : \beta_1 \neq \beta_2.$$

Statystyka testowa ma postać:

$$t = \frac{\beta_1 - \beta_2}{\sqrt{\frac{s_{yx_1}^2}{sd_{x_1}^2(n_1-1)} + \frac{s_{yx_2}^2}{sd_{x_2}^2(n_2-1)}}},$$

gdzie:

$$s_{yx_1} = sd_{y_1} \sqrt{\frac{n_1 - 1}{n_1 - 2} (1 - r_{p_1}^2)},$$

$$s_{yx_2} = sd_{y_2} \sqrt{\frac{n_2 - 1}{n_2 - 2} (1 - r_{p_2}^2)}.$$

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy H_0 przyjmując H_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić H_0 .

Okno z ustawieniami opcji porównania współczynników zależności wywołujemy poprzez menu Statystyka → Testy parametryczne → porównanie współczynników zależności.

porównanie współczynników zależności

Analiza statystyczna : Porównanie współczynników zależności

wypełnij z zapisanego zaznaczenia ▼

	G1	G2
współczynnik nachylenia a	2	3
współczynnik korelacji Pearsona r	0.6	0.67
liczność próby n	60	72
odch. st. dla X	0.441	0.556
odch. st. dla Y	0.575	0.239

Opcje raportu
 Dołącz dane

0.05 ▼ poziom istotności

OK Zamknij

19.2 TESTY NIEPARAMETRYCZNE

19.2.1 WSPÓŁCZYNNIKI KORELACJI MONOTONICZNEJ

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. Związek pomiędzy 2 cechami przedstawia monotoniczny spadek jeżeli wzrostowi jednej cechy towarzyszy spadek drugiej cechy.

Współczynnik korelacji rangowej Spearmana r_s (ang. *Spearman's rank-order correlation coefficient*) jest wykorzystywany do badania siły związku monotonicznego pomiędzy cechami X i Y . Wyznacza się go dla skali [porządkowej](#) lub [interwałowej](#). 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 licznosc d_i .

Wzór ten ulega pewniej 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:

$$\Sigma_X = \frac{n^3 - n - T_X}{12}, \quad \Sigma_Y = \frac{n^3 - n - T_Y}{12},$$

$$T_X = \sum_{i=1}^s (t_{i(X)}^3 - t_{i(X)}), \quad T_Y = \sum_{i=1}^s (t_{i(Y)}^3 - t_{i(Y)}),$$

t – liczba przypadków wchodzących w skład rangi wiązanej.

Poprawka na rangi wiązane powinna być stosowana, gdy rangi wiązane występują. Gdy nie ma rang wiązanych poprawka redukuje się i sprowadza wzór do postaci opisanej wcześniej równaniem.

Uwaga!

R_s oznacza współczynnik korelacji rangowej Spearmana populacji, natomiast r_s w próbie.

Wartość $r_s \in \langle -1; 1 \rangle$ interpretujemy w następujący sposób:

- $r_s \approx 1$ 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 korelacji rangowej Spearmana przyjmuje wartość równą lub bardzo bliską zeru, wówczas nie istnieje monotoniczna zależność między badanymi parametrami (ale może istnieć związek niemonotoniczny np. sinusoidalny).

Współczynnik korelacji tau Kendalla $\tilde{\tau}$ (ang. *Kendall's tau correlation coefficient, Kendall (1938)[90]*) jest wykorzystywany do badania siły związku monotonicznego pomiędzy cechami. Wyznacza się go dla skali [porządkowej](#) lub [interwałowej](#). Wartość współczynnika korelacji tau Kendalla $\tilde{\tau}$ 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),

$$T_X = \sum_{i=1}^s (t_{i(X)}^2 - t_{i(X)}), T_Y = \sum_{i=1}^s (t_{i(Y)}^2 - t_{i(Y)}),$$

t – 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 $T_Y = 0$).

Uwaga!

τ oznacza współczynnik korelacji Kendalla w populacji, natomiast $\tilde{\tau}$ w próbie.

Wartość $\tilde{\tau} \in \langle -1; 1 \rangle$ interpretujemy w następujący sposób:

- $\tilde{\tau} \approx 1$ 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 korelacji $\tilde{\tau}$ przyjmuje wartość równą lub bardzo bliską zeru wówczas nie istnieje monotoniczna zależność między badanymi parametrami (ale może istnieć związek niemonotoniczny np. sinusoidalny).

Współczynnik r_s Spearmana a współczynnik $\tilde{\tau}$ Kendalla

- dla skali interwałowej z normalnością rozkładu obu cech wartość r_s daje rezultaty bliższe wartości r_p natomiast wartości $\tilde{\tau}$ 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 .

19.2.2 Test t do sprawdzania istotności współczynnika korelacji rangowej Spearmana

Test t do sprawdzania istotności współczynnika korelacji rangowej Spearmana (*ang. Test of significance for Spearman's rank-order correlation coefficient*) służy do weryfikacji hipotezy o braku zależności monotonicznej pomiędzy badanymi cechami populacji i opiera się na współczynniku korelacji rangowej Spearmana wyliczonym dla próby. Im wartość r_s jest bliższa 0, tym słabszą zależnością monotoniczną związane są badane cechy.

Podstawowe warunki stosowania:

- pomiar na **skali porządkowej** lub **interwałowej**.

Hipotezy:

$$\mathcal{H}_0 : R_s = 0,$$

$$\mathcal{H}_1 : R_s \neq 0.$$

Statystyka testowa ma postać:

$$t = \frac{r_s}{SE},$$

gdzie $SE = \sqrt{\frac{1 - r_s^2}{n - 2}}$.

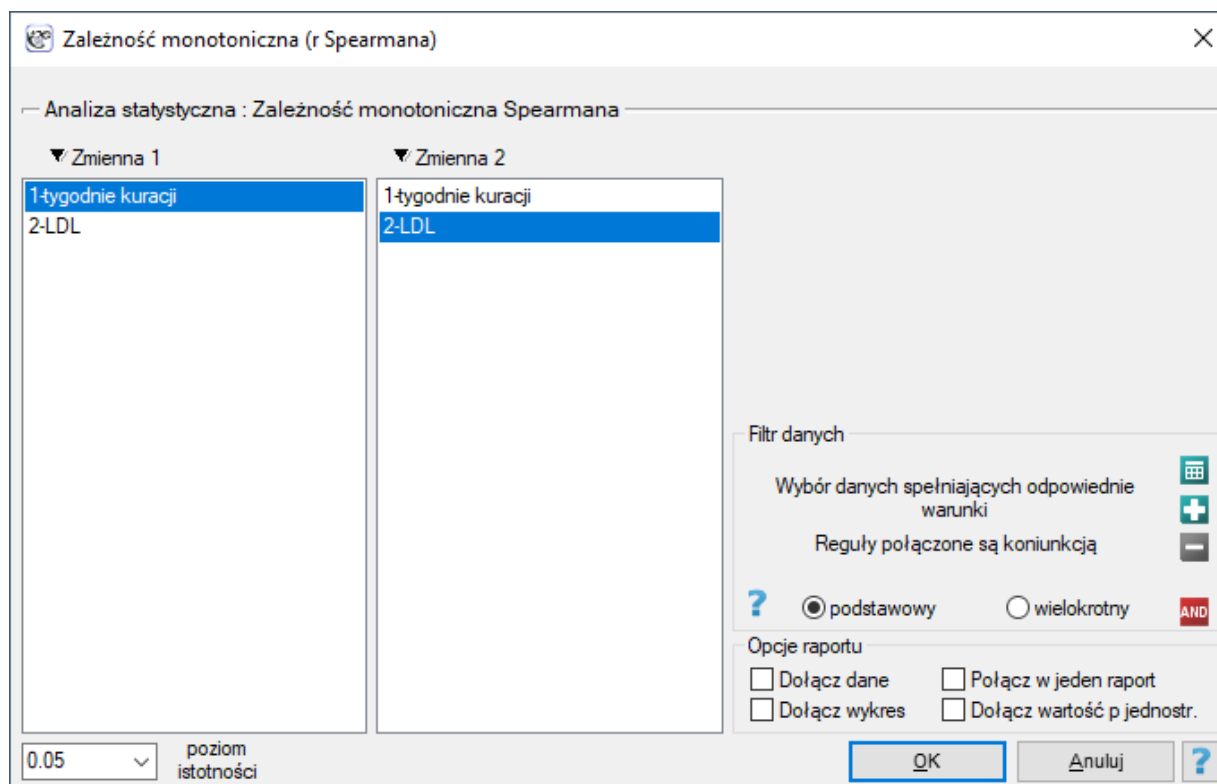
Wartość statystyki testowej nie może być wyznaczona gdy $r_s = 1$ lub $r_s = -1$ albo, gdy $n < 3$.

Statystyka testowa ma **rozkład t-Studenta** z $n - 2$ stopniami swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji zależności monotonicznej Spearmana wywołujemy poprzez menu Statystyka → Testy nieparametryczne → zależność monotoniczna (r-Spearmana) lub poprzez **Kreator**.



PRZYKŁAD 19.2. (plik LDL tygodnie.pqs)

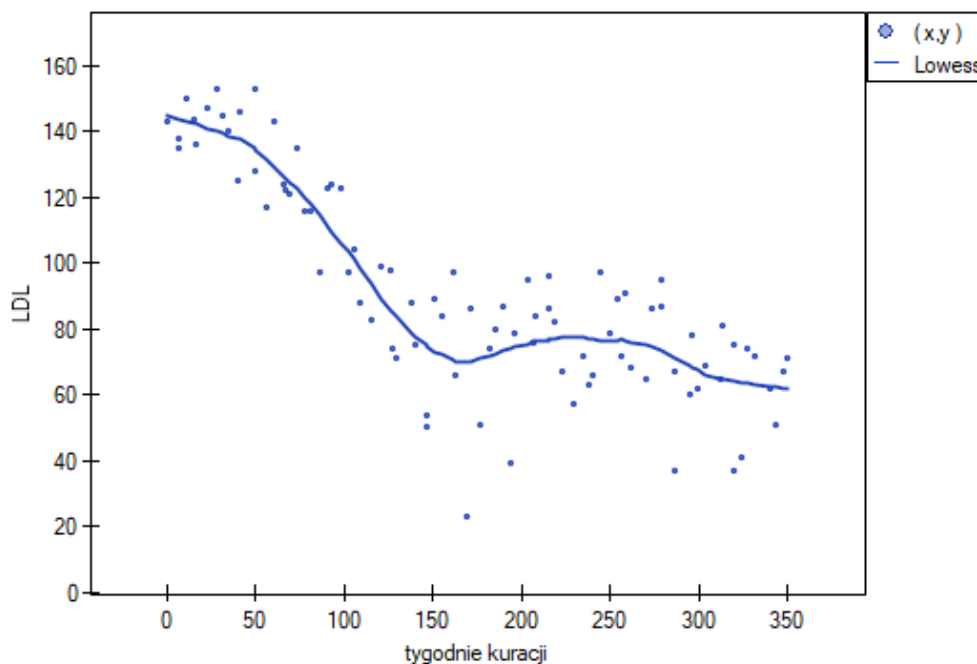
Badano skuteczność nowej terapii, której celem jest obniżenie poziomu cholesterolu we frakcji LDL. Przebadano 88 osób na różnym etapie kuracji. Sprawdzimy, czy wraz z upływem czasu stosowania kuracji (czas w tygodniach) poziom cholesterolu LDL spada i się stabilizuje.

Hipotezy:

\mathcal{H}_0 : W populacji nie istnieje zależność monotoniczna pomiędzy czasem kuracji a poziomem LDL,
 \mathcal{H}_1 : W populacji istnieje zależność monotoniczna pomiędzy czasem kuracji a poziomem LDL.

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

Porównując wartość $p < 0.0001$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że 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.



19.2.3 Test do sprawdzania istotności współczynnika korelacji $\tilde{\tau}$ Kendalla

Test do sprawdzania istotności współczynnika korelacji $\tilde{\tau}$ Kendalla (*ang. Test of significance for Kendall's tau correlation coefficient*) służy do weryfikacji hipotezy o braku zależności monotonicznej pomiędzy badanymi cechami populacji i opiera się na współczynniku korelacji Kendalla wyliczonym dla próby. Im wartość $\tilde{\tau}$ jest bliższa 0, tym słabszą zależnością monotoniczną związane są badane cechy.

Podstawowe warunki stosowania:

- pomiar na [skali porządkowej](#) lub [interwałowej](#).

Hipotezy:

$$\mathcal{H}_0 : \tau = 0,$$

$$\mathcal{H}_1 : \tau \neq 0.$$

Statystyka testowa ma postać:

$$Z = \frac{3\tilde{\tau}\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}.$$

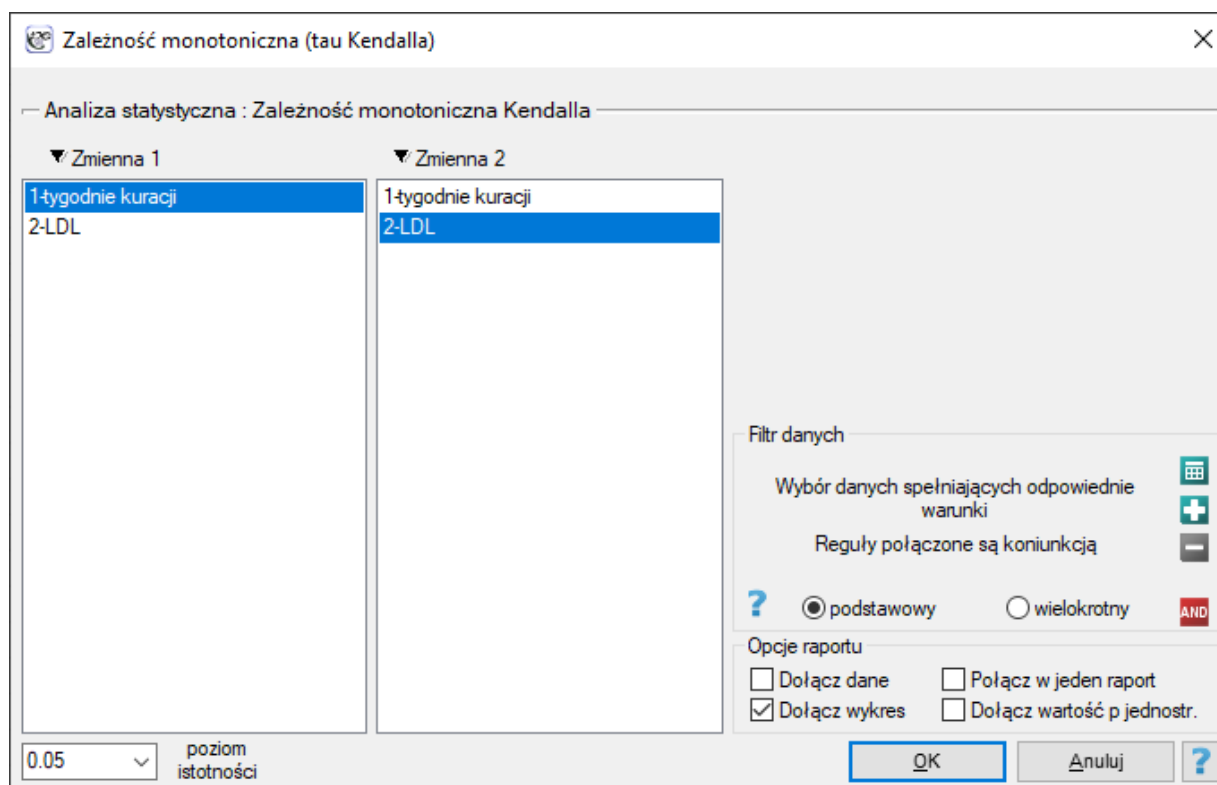
Statystyka testowa ma asymptotycznie (dla dużych licznosci) [rozkład normalny](#).

Wyznaczoną na podstawie [statystyki testowej wartość p](#) porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji zależności monotonicznej Kendalla wywołujemy poprzez menu Statystyka → Testy nieparametryczne → zależność monotoniczna (tau-Kendalla) lub poprzez [Kreator](#).



PRZYKŁAD (19.2) c.d. (plik LDL tygodnie.pqs)

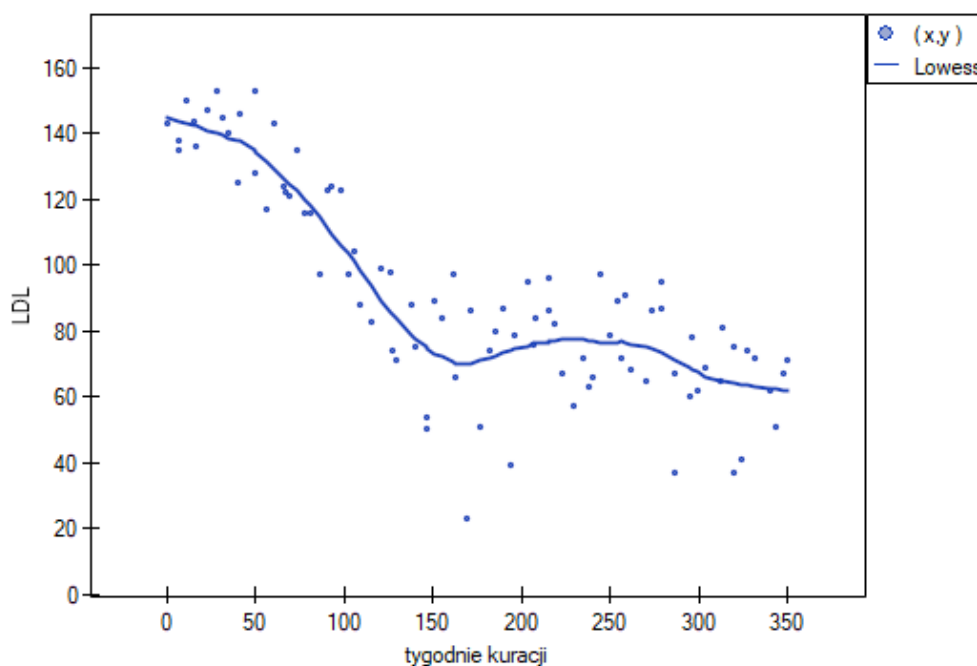
Hipotezy:

\mathcal{H}_0 : W populacji nie istnieje zależność monotoniczna pomiędzy czasem kuracji a poziomem LDL,

\mathcal{H}_1 : W populacji istnieje zależność monotoniczna pomiędzy czasem kuracji a poziomem LDL.

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

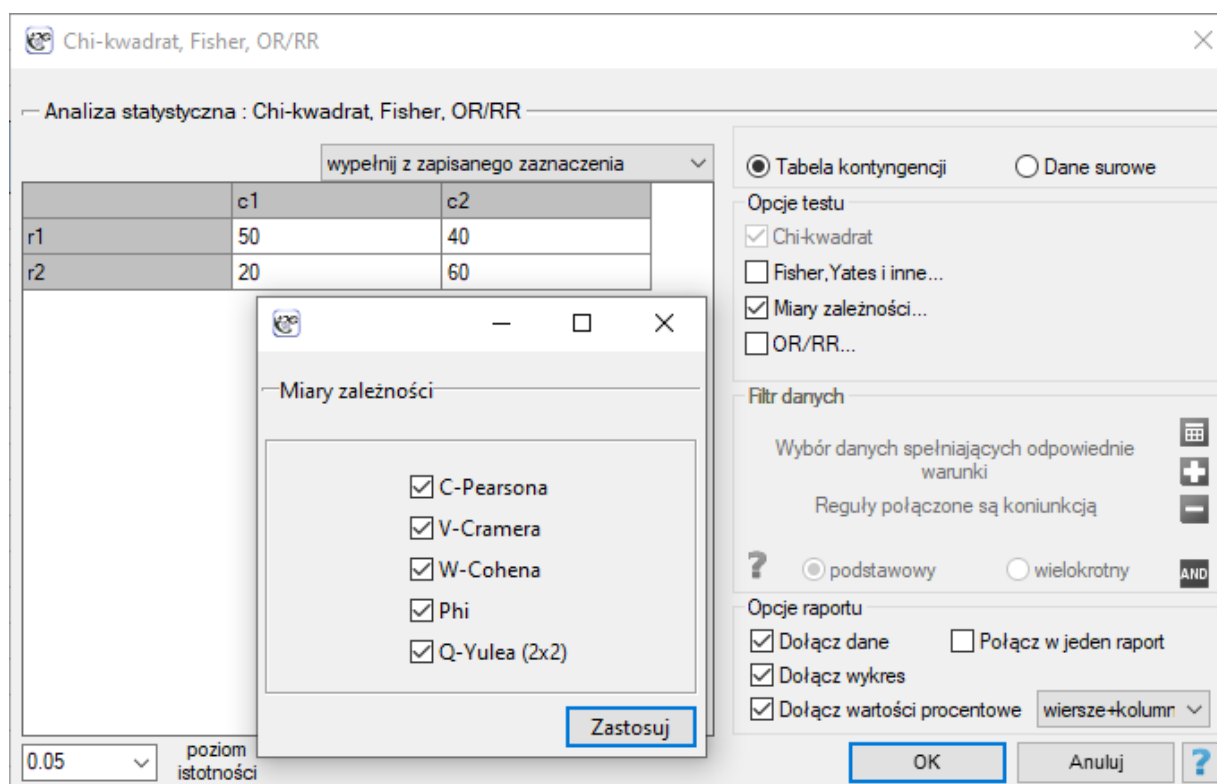
Porównując wartość $p < 0.0001$ z poziomem istotności $\alpha = 0.05$ stwierdzamy, że istnieje ważna statystycznie monotoniczna zależność pomiędzy czasem kuracji a poziomem LDL. Zależność ta jest początkowo malejąca, a po 150 tygodniach zaczyna się stabilizować. Współczynnik korelacji monotonicznej Kendalla, a zatem siła związku monotonicznego dla tej zależności jest dość wysoki i wynosi $\hat{\tau} = -0.5975$. Wykres wyrysowano dopasowując krzywą poprzez lokalne techniki wygładzania liniowego typu LOWESS.



19.2.4 WSPÓŁCZYNNIKI TABEL KONTYNGENCJI I ICH ISTOTNOŚĆ STATYSTYCZNA

Współczynniki kontyngencji są wyliczane dla [danych w postaci surowej](#) lub danych zebranych w [tabelę kontyngencji](#) (patrz tabela(8.1)).

Okno z ustawieniami opcji miar zależności dla tabel wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Chi-kwadrat, Fisher, OR/RR → Współczynniki korelacji... lub poprzez [Kreator](#).



Współczynnik kontyngencji Q -Yulea

Współczynnik kontyngencji Q -Yulea (*Yule's Q contingency coefficient*), Yule (1900)[177], jest miarą zależności, która może być wyznaczana dla tabel kontyngencji 2×2

$$Q = \frac{O_{11}O_{22} - O_{12}O_{21}}{O_{11}O_{22} + O_{12}O_{21}},$$

gdzie:

$O_{11}, O_{12}, O_{21}, O_{22}$ - licznosci obserwowane w tabeli kontyngencji.

Wartość współczynnika Q mieści się w przedziale $< -1; 1 >$. Im wartość ta jest bliższa 0, tym siła związku pomiędzy badanymi cechami jest mniejsza, a im bliższa -1 lub $+1$, tym siła badanego związku jest większa (ze względu na błędy w interpretacji ujemnej wartości współczynnika, wyniki tego współczynnika w programie PQStat przedstawiane są wówczas również jako wartość bezwzględna). Wadą tego współczynnika jest to, iż jest mało odporny na małe licznosci obserwowane (gdy jakaś z licznosci obserwowanych wynosi 0, to współczynnik może błędnie wskazywać całkowitą zależność cech).

Istotność statystyczną wyznaczonego współczynnika kontyngencji Q -Yulea określamy testem Z .

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: Q = 0, \\ \mathcal{H}_1 &: Q \neq 0. \end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{Q}{\sqrt{\frac{1}{4}(1 - Q^2)^2 \left(\frac{1}{O_{11}} + \frac{1}{O_{12}} + \frac{1}{O_{21}} + \frac{1}{O_{22}} \right)}}.$$

Statystyka testowa ma asymptotycznie (dla dużych licznosci n) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Współczynnik kontyngencji ϕ

Współczynnik kontyngencji ϕ (*ang. phi contingency coefficient*) jest miarą zależności polecaną szczególnie dla tabel kontyngencji 2×2 , chociaż możliwą do wyznaczenia dla dowolnych tabel.

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

gdzie:

χ^2 – wartość statystyki testu χ^2 ,

n – całkowita licznosc w tabeli kontyngencji.

Wartość współczynnika ϕ mieści się w przedziale $< 0; 1 >$. Im wartość ta jest bliższa 0, tym siła związku pomiędzy badanymi cechami jest mniejsza, a im bliższa wartości 1 tym większa.

Współczynnik kontyngencji ϕ uznaje się za **istotny statystycznie** jeśli **wartość p** wyznaczona na podstawie statystyki testu χ^2 i rozkładu χ^2 (wyznaczonego dla tej tabeli) jest równa bądź mniejsza niż **poziom istotności α** .

Współczynnik kontyngencji V -Cramera

Współczynnik kontyngencji V -Cramera (*ang. Cramer's V contingency coefficient*), Cramer (1946)[48], jest rozszerzeniem współczynnika ϕ na tabele kontyngencji $r \times c$.

$$V = \sqrt{\frac{\chi^2}{n(w' - 1)}},$$

gdzie:

χ^2 – wartość statystyki testu χ^2 ,

n – całkowita licznosc w tabeli kontyngencji,

w' – jest mniejszą z dwóch wartości r i c .

Wartość współczynnika V mieści się w przedziale $< 0; 1 >$. Im wartość ta jest bliższa 0, tym siła związku pomiędzy badanymi cechami jest mniejsza, a im bliższa +1, tym siła badanego związku jest większa. Wartość współczynnika V zależy również od wielkości tabeli, stąd nie powinno się stosować tego współczynnika do porównywania tabel kontyngencji o różnych wielkościach.

Współczynnik kontyngencji V uznaje się za **istotny statystycznie** jeśli **wartość p** wyznaczona na podstawie statystyki testu χ^2 i rozkładu χ^2 (wyznaczonego dla tej tabeli) jest równa bądź mniejsza niż **poziom istotności α** .

Współczynnik kontyngencji W -Cohena

Współczynnik kontyngencji W -Cohena (*ang. Cohen's w contingency coefficient*), Cohen (1988)[45], jest modyfikacją współczynnika V -Cramera i jest możliwy do wyliczenia dla tabel $r \times c$.

$$W = \sqrt{\frac{\chi^2}{n(w' - 1)}} \sqrt{w' - 1},$$

gdzie:

χ^2 – wartość statystyki testu χ^2 ,

n – całkowita licznosc w tabeli kontyngencji,

w' – jest mniejszą z dwóch wartości r i c .

Wartość współczynnika W mieści się w przedziale $< 0; \max W >$, gdzie $\max W = \sqrt{w' - 1}$ (dla tabel, w których co najmniej jedna zmienna zawiera tylko dwie kategorie wartość współczynnika W mieści się w przedziale $< 0; 1 >$). Im wartość ta jest bliższa 0, tym siła związku pomiędzy badanymi cechami jest mniejsza, a im bliższa maksymalnej wartości, tym siła badanego związku jest większa. Wartość współczynnika W zależy od wielkości tabeli, stąd nie powinno się stosować tego współczynnika do porównywania tabel kontyngencji o różnych wielkościach.

Współczynnik kontyngencji W uznaje się za **istotny statystycznie** jeśli **wartość p** wyznaczona na podstawie statystyki testu χ^2 i rozkładu χ^2 (wyznaczonego dla tej tabeli) jest równa bądź mniejsza niż **poziom istotności α** .

Współczynnik kontyngencji C -Pearsona

Współczynnik kontyngencji C -Pearsona (*ang. Pearson's C contingency coefficient*) jest miarą zależności wyznaczaną dla tabel kontyngencji $r \times c$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

gdzie:

χ^2 – wartość statystyki testu χ^2 ,

n – całkowita licznosc w tabeli kontyngencji.

Wartość współczynnika C mieści się w przedziale $< 0; 1)$. Im wartość ta jest bliższa 0, tym siła związku pomiędzy badanymi cechami jest mniejsza, a im dalsza od 0, tym siła badanego związku jest większa. Ponieważ wartość współczynnika C zależy również od wielkości tabeli (im większa tabela, tym wartość C może być bliższa 1), dlatego wyznacza się górną granicę jaką dla danej wielkości tabeli współczynnik C może osiągnąć:

$$C_{max} = \sqrt{\frac{w' - 1}{w'}},$$

gdzie:

w' – jest mniejszą z dwóch wartości r i c .

Niewygodną konsekwencją uzależnienia wartości C od wielkości tabeli jest brak możliwości porównywania wartości współczynnika C wyznaczonego dla różnych wielkości tabel kontyngencji. Nieco lepszą miarą w takim przypadku jest dostosowana do wielkości tabeli wielkość współczynnika kontyngencji C_{adj}

$$C_{adj} = \frac{C}{C_{max}}.$$

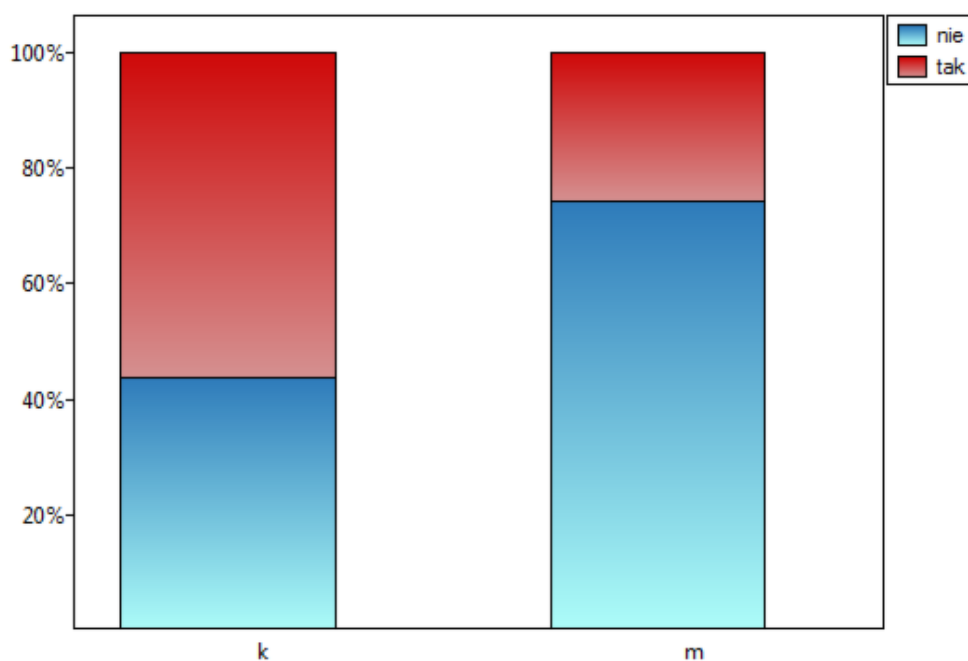
Współczynnik kontyngencji C uznaje się za **istotny statystycznie** jeśli **wartość p** wyznaczona na podstawie statystyki testu χ^2 i rozkładu χ^2 (wyznaczonego dla tej tabeli) jest równa bądź mniejsza niż **poziom istotności α** .

PRZYKŁAD 19.3. (plik płęć-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 tabeli 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	
Czas analizy	0.10 sek.
Analizowane zmienne	Tabela kontyngencji
Liczność - niezinterpretowane	0
Liczność - braki danych	0
Poziom istotności	0.05
Liczność	170
Warunek Cochra	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



Wartość statystyki testowej wynosi $\chi^2 = 16.33$ a wyznaczona dla niej wartość $p = 0.00005$. Uzyskany wynik wskazuje na istnienie zależności statystycznej pomiędzy płcią a zdawalnością egzaminu w badanej populacji.

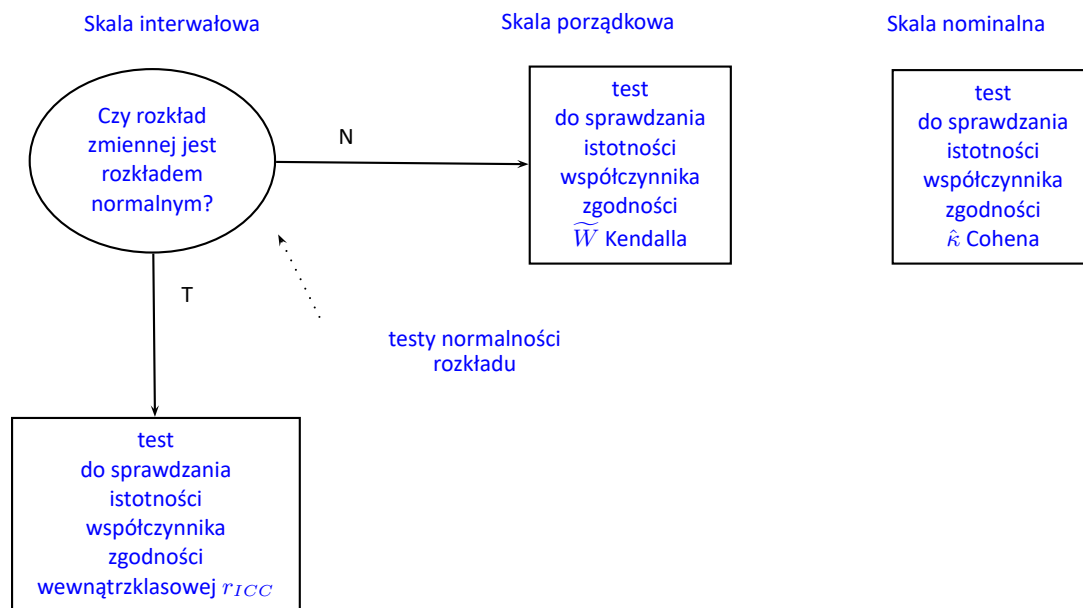
Wartość współczynników opartych o test χ^2 , a zatem siła związku między badanymi cechami to:

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.

20 ANALIZA ZGODNOŚCI



20.1 TESTY PARAMETRYCZNE

20.1.1 Współczynnik korelacji wewnątrzklasowej i test badający jego istotność

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 i typu zgodności. W zależności od występującej w danych zmienności możemy rozróżnić 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 inny zestaw k sędziów może być wylosowany.

Współczynnik ICC jest wówczas wyznaczany w modelu losowym **ANOVA dla grup niezależnych**. Na pytanie o rzetelność ocen pojedynczego sędziego odpowiada ICC(1,1) dany wzorem:

$$ICC(1, 1) = \frac{MS_{WG} - MS_{BG}}{MS_{WG} + (k - 1)MS_{BG}}.$$

Do oszacowania wiarygodności wyników będących średnią ocen sędziowskich (dla k sędziów) wyznacza się ICC(1,k) dany wzorem:

$$ICC(1, k) = \frac{MS_{WG} - MS_{BG}}{MS_{WG}},$$

gdzie:

MS_{WG} - średnia kwadratów wewnątrz grup,

MS_{BG} - średnia kwadratów między obiektami.

Model 2 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 możemy szacować: zgodność bezwzględną tzn. jeśli sędziowie są zgodni bezwzględnie, to wydają dokładnie takie same oceny np. perfekcyjnie zgodne będą takie oceny wydane przez parę sędziów (2,2), (5,5), (8,8); lub spójność tzn. sędziowie mogą używać innych zakresów wartości ale poza tym przesunięciem nie powinno być różnic by zachować spójność werdyktu np. perfekcyjnie spójne będą takie oceny wydane przez parę sędziów (2,5), (5,8), (8,11).

Zgodność bezwzględna

Na pytanie o rzetelność ocen pojedynczego sędziego odpowiada ICC(2,1) dany wzorem:

$$ICC(2, 1) = \frac{MS_{BS} - MS_{res}}{MS_{BS} + (k - 1)MS_{res} + \frac{k}{n}(MS_{BC} - MS_{res})}.$$

Do oszacowania wiarygodności wyników będących średnią ocen sędziowskich (dla k sędziów) wyznacza się ICC(2,k) dany wzorem:

$$ICC(2, k) = \frac{MS_{BS} - MS_{res}}{MS_{BS} + (MS_{BC} - MS_{res})/n},$$

gdzie:

MS_{BC} - średnia kwadratów między sędziami,

MS_{BS} - średnia kwadratów między obiektami,

MS_{res} - średnia kwadratów dla reszt.

Spójność

Na pytanie o rzetelność ocen pojedynczego sędziego odpowiada ICC(2,1) dany wzorem:

$$ICC(2, 1) = \frac{MS_{BS} - MS_{res}}{MS_{BS} + (k - 1)MS_{res}},$$

Do oszacowania wiarygodności wyników będących średnią ocen sędziowskich (dla k sędziów) wyznacza się ICC(2,k) dany wzorem:

$$ICC(2, k) = \frac{MS_{BS} - MS_{res}}{MS_{BS}},$$

gdzie:

MS_{BS} - średnia kwadratów między obiektami,

MS_{res} - średnia kwadratów dla reszt.

Uwaga!

Czasami istnieje konieczność rozważania **modelu 3** [150], tzn. wybierany jest zestaw k sędziów i każdy z nich ocenia wszystkie n losowych obiektów. Wynik zgodności dotyczy tylko tych konkretnych k sędziów. Współczynnik ICC jest wówczas wyznaczany w modelu mieszanym (ponieważ losowość dotyczy tylko obiektów a nie dotyczy sędziów). Ponieważ ignorujemy zmienność dotyczącą sędziów, badamy spójność (a nie absolutną zgodność) i zastosowanie mogą mieć 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.

Uwaga!

Wartość $ICC \in < -1; 1 >$ interpretujemy w następujący sposób:

- $ICC \approx 1$ oznacza silną zgodność w ocenie poszczególnych obiektów przez sędziów, co ma odzwierciedlenie w dużej wariancji między obiektami (znaczej różnicy średnich między n obiektami) i małej wariancji między ocenami sędziowskimi (niewielkiej różnicy średnich ocen wyznaczonych dla k sędziów);
- $ICC \approx -1$ negatywny współczynnik korelacji wewnątrzklasowej, jest traktowany w ten sam sposób jak $r_{ICC} \approx 0$;
- $ICC \approx 0$ 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).

Test F do sprawdzania istotności współczynnika korelacji wewnątrzklasowej

Podstawowe 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](#).

Hipotezy:

$$\mathcal{H}_0 : ICC = 0$$

$$\mathcal{H}_1 : ICC \neq 0 \quad (ICC = 1)$$

Statystyka testowa ma postać:

$$F = \frac{MS_{BS}}{MS_{res}} - \text{w modelu zaleźnym}$$

lub

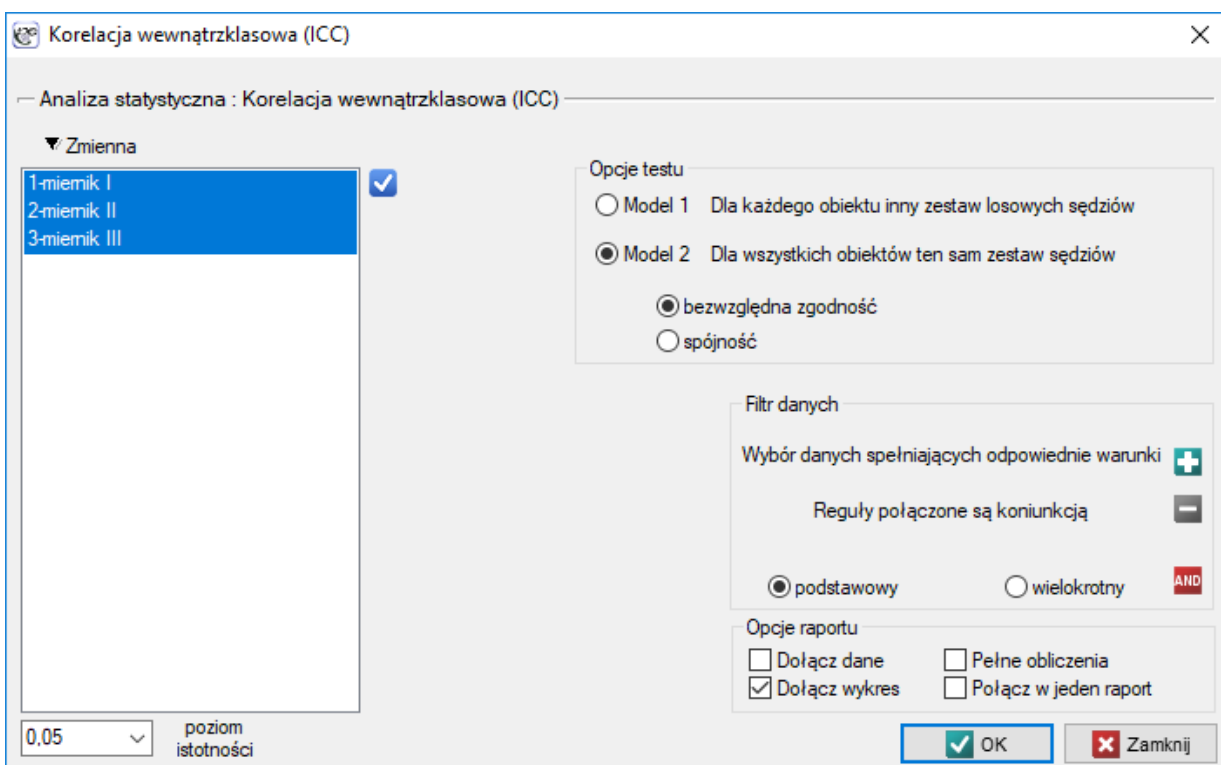
$$F = \frac{MS_{WG}}{MS_{BG}} - \text{w modelu niezależnym.}$$

Statystyka ta podlega [rozkładowi F Snedecora](#) ze zdefiniowaną w modelu liczbą stopni swobody.

Wyznaczoną na podstawie [statystyki testowej wartość p](#) porównujemy z α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Okno z ustawieniami opcji korelacji wewnątrzklasowej (ICC) wywołujemy poprzez menu Statystyka → Testy parametryczne → Korelacja wewnątrzklasowa (ICC) lub poprzez [Kreator](#).



PRZYKŁAD 20.1. (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 ($p < 0.000001$). Rzetelność pomiaru dokonana miernikiem analogowym wynosi $ICC(2, 1) = 0.45$, natomiast rzetelność pomiaru będącego średnią pomiarów dokonanych przez 3 mierniki analogowe jest nieco wyższa i wynosi $ICC(2, k) = 0.71$. Niepokojąco niska jest jednak dolna granica 95% przedziału ufności dla tych współczynników.

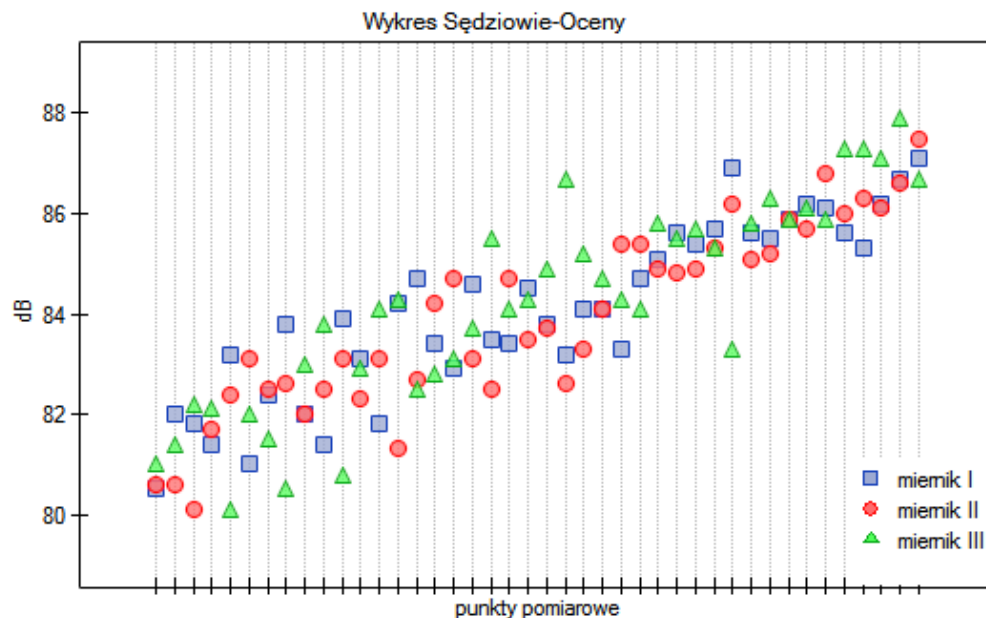
Korelacja wewnątrzklasowa (ICC)	
Czas analizy	0.06 sek.
Analizowane zmienne	miernik I miernik II miernik III
Poziom istotności	0.05
Model 2 Dla wszystkich obiektów ten sam zestaw sędziów	bezwzględna zgodność
Stopnie swobody [objekty]	41
Stopnie swobody [res]	82
Średnia kwadratów [pomiędzy sędziami]	105.151667
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

Podobna analiza przeprowadzona dla mierników cyfrowych dała lepsze rezultaty. Model ponownie jest istotny statystycznie, ale współczynniki ICC oraz ich przedziały ufności są znacznie wyżej niż dla mierników analogowych, a więc uzyskana zgodność bezwzględna jest wyższa $ICC(2, 1) = 0.73$, $ICC(2, k) = 0.89$.

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

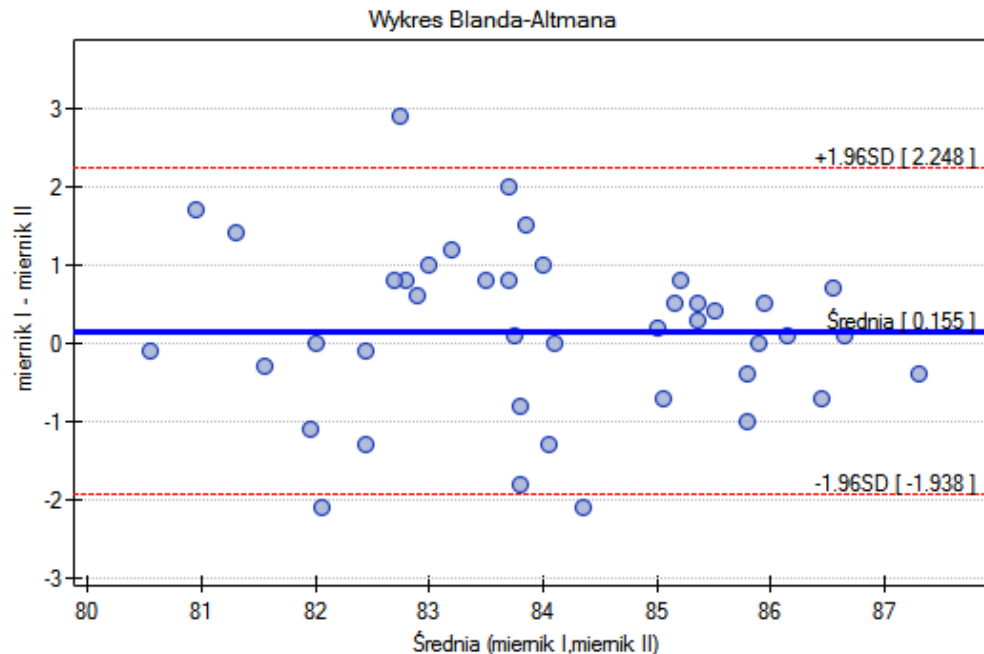
Dlatego ostatecznie w zakładzie pracy wykorzystywane będą mierniki cyfrowe.

Zgodność uzyskanych wyników dla mierników cyfrowych przedstawiono na wykresie punktowym, gdzie każdy punkt pomiarowy opisany jest wartością natężenia dźwięku uzyskaną dla poszczególnych mierników.



Przedstawiając wykres dla uprzednio posortowanych danych zgodnie ze średnią wartością natężenia dźwięku, można sprawdzić czy stopień zgodności rośnie lub spada wraz z 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][22] 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.

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.

20.2 TESTY NIEPARAMETRYCZNE

20.2.1 Współczynnik zgodności Kendalla i test badający jego istotność

Współczynnik zgodności \widetilde{W} Kendalla (*ang. Kendall's Coefficient of Concordance*) opisany w pracy Kendalla i Babingtona-Smitha (1939)[91] oraz Wallisa (1939)[165] 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.

Współczynnik zgodności Kendalla wyznacza się dla skali **porządkowej** lub **interwałowej**, a jego wartość wylicza się według wzoru:

$$\widetilde{W} = \frac{12U - 3n^2k(k+1)^2}{n^2k(k^2-1) - nC},$$

gdzie:

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 ($j = 1, 2, \dots, k$), oddzielnie dla każdego z sędziów ($i = 1, 2, \dots, n$),

$C = \sum (t^3 - t)$ – 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 $C = 0$).

Uwaga!

W oznacza współczynnik zgodności Kendalla w populacji, natomiast \widetilde{W} w próbie.

Wartość $W \in \langle 0; 1 \rangle$ interpretujemy w następujący sposób:

- $\widetilde{W} \approx 1$ oznacza silną zgodność w ocenie poszczególnych obiektów przez sędziów;
- $\widetilde{W} \approx 0$ oznacza brak zgodności w ocenie poszczególnych obiektów przez sędziów.

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 danych:

$$\bar{r}_s = \frac{n\widetilde{W} - 1}{n - 1}$$

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 chi-kwadrat do sprawdzania istotności współczynnika zgodności Kendalla i wartość statystyki testowej ANOVA Friedmana jest taka sama.

Test chi-kwadrat do sprawdzania istotności współczynnika zgodności Kendalla

Podstawowe warunki stosowania:

- pomiar na [skali porządkowej](#) lub [interwałowej](#).

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: W = 0 \\ \mathcal{H}_1 &: W \neq 0 \end{aligned}$$

Statystyka testowa ma postać:

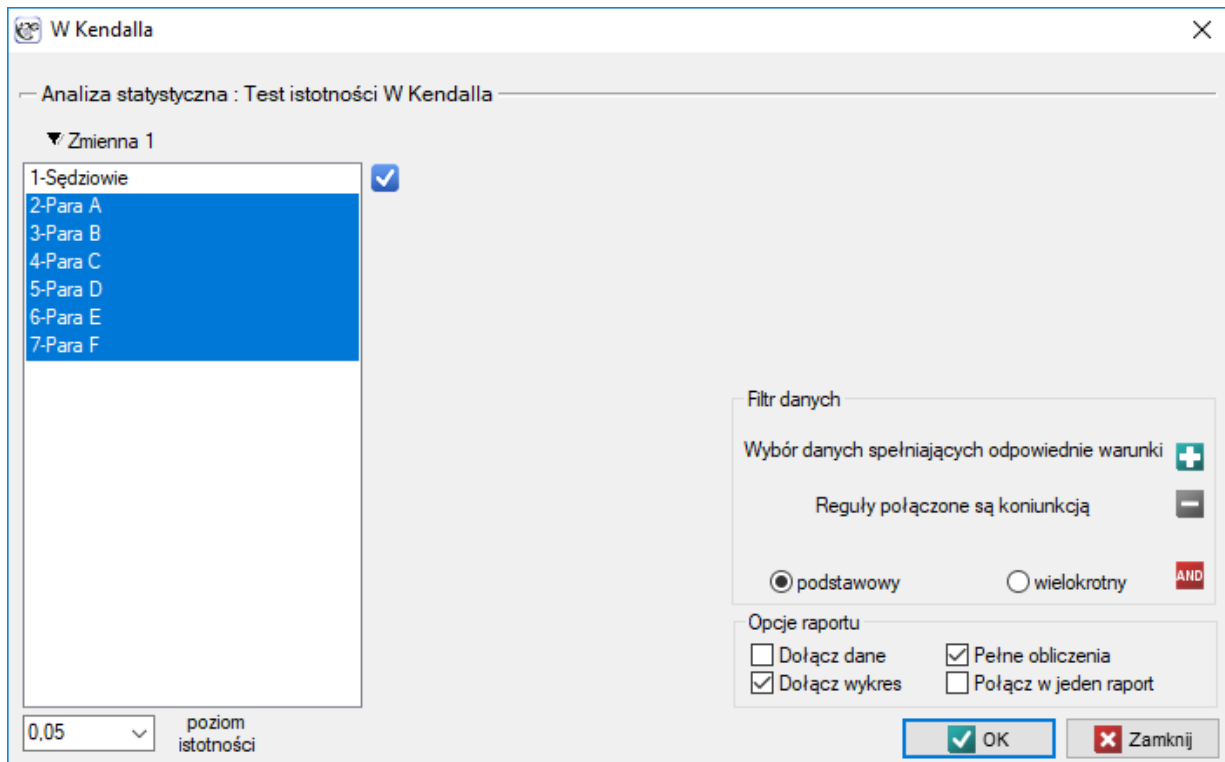
$$\chi^2 = n(k - 1)\widetilde{W}$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) [rozkład \$\chi^2\$](#) z liczbą stopni swobody wyliczaną z wzoru: $df = k - 1$.

Wyznaczoną na podstawie [statystyki testowej wartość \$p\$](#) porównujemy z α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Okno z ustawieniami opcji testu istotności W Kendalla wywołujemy poprzez menu Statystyka \rightarrow Testy nieparametryczne $\rightarrow W$ Kendalla lub poprzez [Kreator](#).



PRZYKŁAD 20.2. (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). Sprawdźmy, czy rangi przypisane przez sędziów są zgodne:

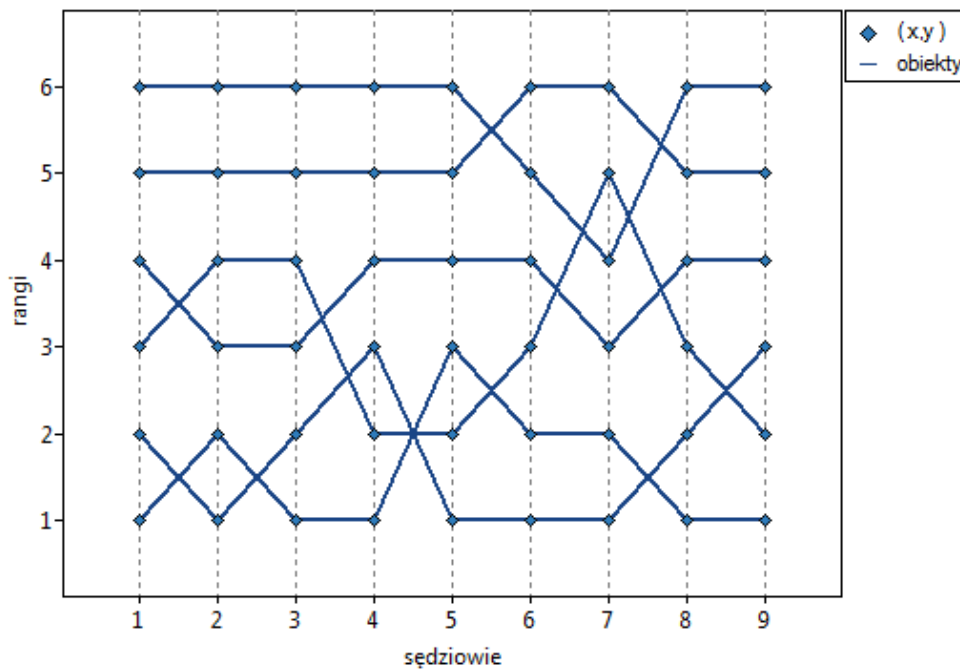
Sędziowie	Para A	Para B	Para C	Para D	Para E	Para F
S1	3	6	2	5	4	1
S2	4	6	1	5	3	2
S3	4	6	2	5	3	1
S4	2	6	3	5	4	1
S5	2	6	1	5	4	3
S6	3	5	1	6	4	2
S7	5	4	1	6	3	2
S8	3	6	2	5	4	1
S9	2	6	3	5	4	1

Hipotezy:

- \mathcal{H}_0 : brak zgodności pomiędzy rankingami 9 sędziów w populacji reprezentowanej przez zebraną próbę,
 \mathcal{H}_1 : rankingi 9 sędziów w populacji reprezentowanej przez zebraną próbę są zgodne.

Test istotności W Kendalla	
Czas analizy	0.04 sek.
Analizowane zmienne	Para A,Para B,Para C,Para D,Para E,Para
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

Porównując wartość $p < 0.000001$ z poziomem istotności $\alpha = 0.05$, stwierdziliśmy, że 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.



20.2.2 Współczynnik Kappa Cohena i test badający jego istotność

Współczynnik Kappa Cohena (*ang. Cohen's Kappa*), Cohen J. (1960)[43], określa stopień zgodności dwukrotnych pomiarów tej samej zmiennej w różnych warunkach. Pomiaru tej samej zmiennej może dokonać dwóch różnych sędziów (odtworzalność) lub jeden sędzia może dokonać pomiaru dwukrotnie (powtarzalność). Współczynnik $\hat{\kappa}$ wyznacza się dla zależnych zmiennych kategoryalnych, a 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. Współczynnik $\hat{\kappa}$ można wyliczać na podstawie danych surowych albo z wykonanej na podstawie danych surowych tabeli kontyngencji o wymiarach $c \times c$.

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 $< 0; 1$).

Kappa nieważona

Wyluczana jest dla danych, których kategorii nie da się uporządkować np. dane pochodzą od pacjentów, których dzielimy ze względu na rodzaj choroby jaka została zdiagnozowana, a chorób tych nie można uporządkować np. zapalenie płuc (1), zapalenie oskrzeli (2) i 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 kiedy kategorie danych mogą być posortowane np. dane pochodzą od pacjentów, których dzielimy ze względu na stopień zmian chorobowych na: brak zmian (1), zmiany łagodne (2), podejrzenie raka (3), rak (4), budować można zgodność wystawionych ocen przez dwóch radiologów uwzględniając możliwość sortowania. Za bardziej niezgodne pary ocen mogą być wówczas uznane oceny $\{(1), (4)\}$ niż $\{(1), (2)\}$. By tak było, by kolejność kategorii wpływała na wynik zgodności, wyznaczać należy ważoną Kappę.

Przydzielane wagi mogą mieć postać liniową lub kwadratową.

- **Wagi liniowe** (Cicchetti, 1971[36]) - wyluczane według wzoru:

$$w_{ij} = 1 - \frac{|i - j|}{c - 1}.$$

Czym większe oddalenie od przekątnej macierzy tym mniejsza waga, przy czym wagi maleją proporcjonalnie. Przykładowe wagi dla macierzy wielkości 5x5 przedstawia tabela:

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** (Cohen, 1968[44]) - wyluczane według wzoru:

$$w_{ij} = 1 - \frac{(i - j)^2}{(c - 1)^2}.$$

Czym większe oddalenie od przekątnej macierzy tym mniejsza waga, przy czym wagi maleją wolniej w bliższej odległości od przekątnej i szybciej w odległości dalszej. Przykładowe wagi dla macierzy wielkości 5x5 przedstawia tabela:

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

Wagi kwadratowe cieszą się większym zainteresowaniem ze względu na praktyczną interpretację współczynnika Kappa, który w tym przypadku tożsamy jest ze współczynnikiem korelacji wewnątrzklasowej [60].

By wyznaczyć zgodność współczynnikiem Kappa dane przedstawia się w postaci tabeli licznosci obserwowanych O_{ij} (13.3), dalej tą tabelę 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:

$$P_o = \sum_{i=1}^c \sum_{j=1}^c w_{ij} p_{ij},$$

$$P_e = \sum_{i=1}^c \sum_{j=1}^c w_{ij} p_{i.} p_{.j},$$

$p_{i.}, p_{.j}$ - sumy końcowe kolumn i wierszy tabeli kontyngencji prawdopodobieństw.

Uwaga!

$\hat{\kappa}$ oznacza współczynnik zgodności w próbie, natomiast κ w populacji.

Błąd standardowy dla Kappa wyraża się wzorem:

$$SE_{\hat{\kappa}} = \frac{1}{(1 - P_e)\sqrt{n}} \sqrt{\sum_{i=1}^c \sum_{j=1}^c p_{i.} p_{.j} [w_{ij} - (\bar{w}_{i.} + \bar{w}_{.j})]^2 - P_e^2}$$

gdzie:

$$\bar{w}_{i.} = \sum_{j=1}^c p_{.j} w_{ij},$$

$$\bar{w}_{.j} = \sum_{i=1}^c p_{i.} w_{ij}.$$

Test Z do sprawdzania istotności współczynnika Kappa Cohena ($\hat{\kappa}$) (ang. *The Z test of significance for the Cohen's Kappa*) Fleiss (2003[61]) 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.

Podstawowe warunki stosowania:

- pomiar na **skali nominalnej** (Kappa Cohena - nieważona) i **skali porządkowej** (Kappa ważona).

Hipotezy:

$$\mathcal{H}_0 : \kappa = 0,$$

$$\mathcal{H}_1 : \kappa \neq 0.$$

Statystyka testowa ma postać:

$$Z = \frac{\hat{\kappa}}{SE_{\kappa_{distr}}},$$

gdzie:

$$SE_{\kappa_{distr}} = \frac{1}{(1 - P_e)\sqrt{n}} \sqrt{\sum_{i=1}^c \sum_{j=1}^c p_{ij} [w_{ij} - (\bar{w}_{i.} + \bar{w}_{.j})(1 - \hat{\kappa})]^2 - [\hat{\kappa} - P_e(1 - \hat{\kappa})]^2}.$$

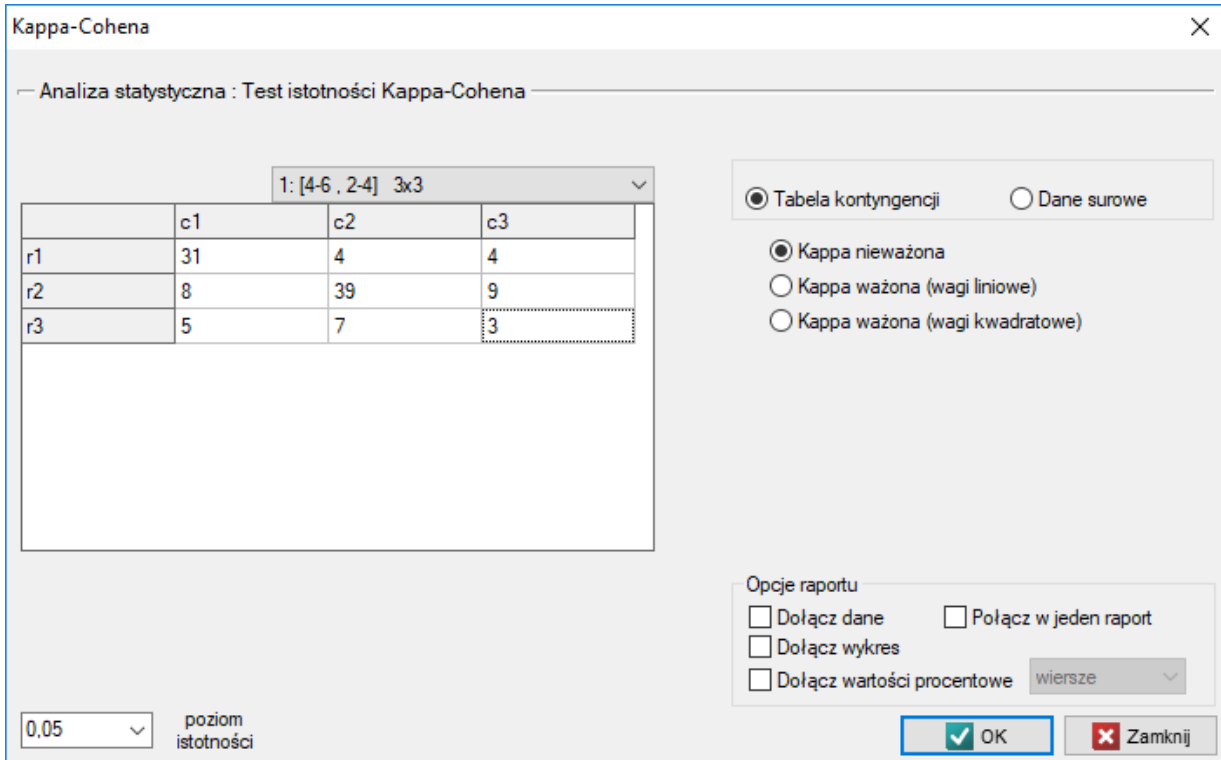
Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji testu istotności Kappa-Cohena wywołujemy poprzez menu Statystyka → Testy nieparametryczne → Kappa-Cohena lub poprzez [Kreator](#).



Kappa-Cohena

Analiza statystyczna : Test istotności Kappa-Cohena

1: [4-6, 2-4] 3x3

	c1	c2	c3
r1	31	4	4
r2	8	39	9
r3	5	7	3

0,05 poziom istotności

Tabela kontyngencji
 Dane surowe

Kappa nieważona
 Kappa ważona (wagi liniowe)
 Kappa ważona (wagi kwadratowe)

Opcje raportu

Dołącz dane
 Połącz w jeden raport
 Dołącz wykres
 Dołącz wartości procentowe wiersze

OK Zamknij

PRZYKŁAD 20.3. (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

Hipotezy:

$$\mathcal{H}_0 : \kappa = 0,$$

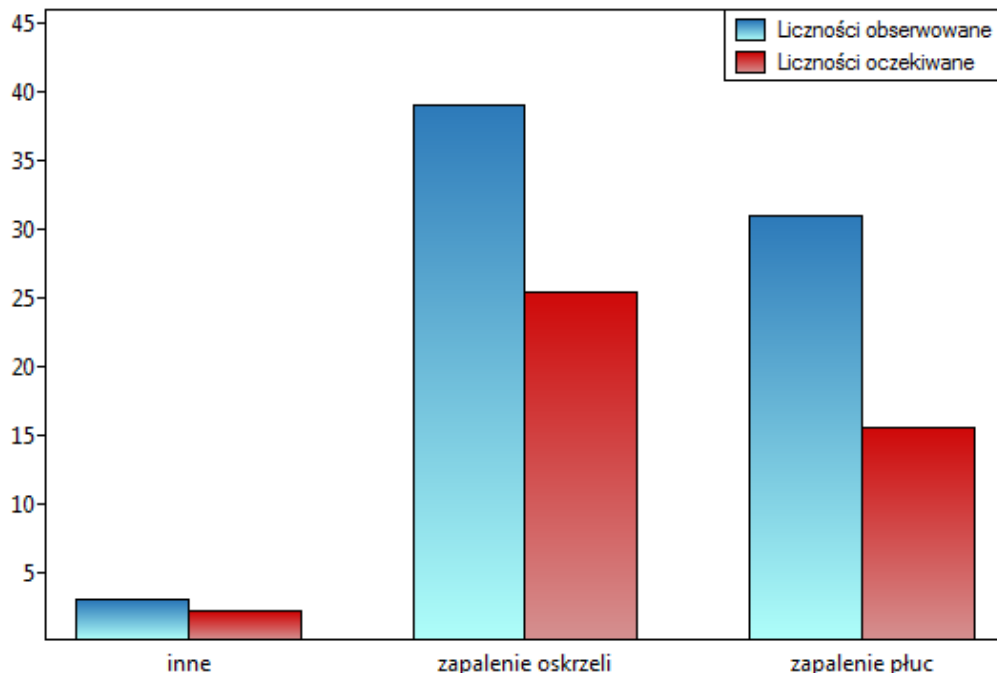
$$\mathcal{H}_1 : \kappa \neq 0.$$

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. w korekcji o tę zgodność, która pojawia się dla przypadkowego rozłożenia danych w tabeli).

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
Błąd 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%



Zgodność wyrażona współczynnikiem $\hat{\kappa} = 44,58\%$ jest mniejsza niż ta nie skorygowana o szanse na zgodność.

Wartość $p < 0.000001$. Wynik taki, na poziomie istotności $\alpha = 0.05$, świadczy o zgodności opinii tych 2 lekarzy.

PRZYKŁAD 20.4. (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.

	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

Hipotezy:

$$\mathcal{H}_0 : \kappa = 0,$$

$$\mathcal{H}_1 : \kappa \neq 0.$$

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

Ponieważ dane koncentrują się głównie na głównej przekątnej macierzy i w jej bliskim sąsiedztwie, współczynnik ważony wagami liniowymi jest niższy ($\hat{\kappa} = 0.390194$) niż współczynnik wyznaczony dla wag kwadratowych ($\hat{\kappa} = 0.418658$). W obu sytuacjach jest to wynik istoty statystycznie (na poziomie istotności $\alpha = 0.05$), wartość p wynosi odpowiednio: 0.000007 i 0.000431.

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 widoczny będzie bardziej, gdy wykorzystamy wagi kwadratowe (współczynnik Kappa drastycznie spadnie) niż przy wykorzystaniu wag liniowych.

20.2.3 Współczynnik Kappa Fleissa i test badający jego istotność

Współczynnik ten określa zgodność pomiarów prowadzonych przez kilku sędziów (Fleiss, 1971[59]) i jest rozszerzeniem współczynnika [Kappa Cohena](#), pozwalającego na badanie zgodności jedynie dwóch sędziów. Przy czym, należy zaznaczyć że 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. Zatem w każdym wierszu tabeli podano x_{ij} czyli liczbę sędziów wydających określone w danej kolumnie opinie.

Współczynnik Kappa ($\hat{\kappa}$) wyraża się wtedy wzorem:

$$\hat{\kappa} = \frac{P_o - P_e}{1 - P_e},$$

gdzie:

$$P_o = \frac{1}{kn(k-1)} \sum_{i=1}^n \sum_{j=1}^c x_{ij} - kn,$$

$$P_e = \sum_{i=1}^c q_j^2,$$

$$q_j = \frac{1}{kn} \sum_{i=1}^n x_{ij}.$$

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.

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[61]) służy do weryfikacji hipotezy o zgodności ocen kilku sędziów i opiera się na współczynniku $\hat{\kappa}$ wyliczonym dla próby.

Podstawowe warunki stosowania:

- pomiar na [skali nominalnej](#) - ewentualne uporządkowanie kategorii nie jest brane pod uwagę.

Hipotezy:

$$\mathcal{H}_0 : \kappa = 0,$$

$$\mathcal{H}_1 : \kappa \neq 0.$$

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 :

- jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

Wyznaczanie współczynnika Kappa Fleissa zbliżone jest 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[61]). Dlatego, jako podsumowanie każdej warstwy można wyznaczyć zgodność sędziowską (współczynnik Kappa) podsumowującą każdą możliwą ocenę z osobna.

Okno z ustawieniami opcji testu istotności Kappa-Fleissa wywołujemy poprzez menu Statystyka \rightarrow Testy nieparametryczne \rightarrow Kappa-Fleissa

Kappa Fleissa

Analiza statystyczna : Kappa Fleissa

wypełnij z zapisanego zaznaczenia

	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

Tabela kontyngencji Dane surowe
 Podsumuj zgodność każdej oceny

Opcje raportu

Dołącz dane
 Dołącz wykres
 Dołącz wartości procentowe wiersze

0.05 poziom istotności

OK Zamknij

PRZYKŁAD 20.5. (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.

	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

Hipotezy:

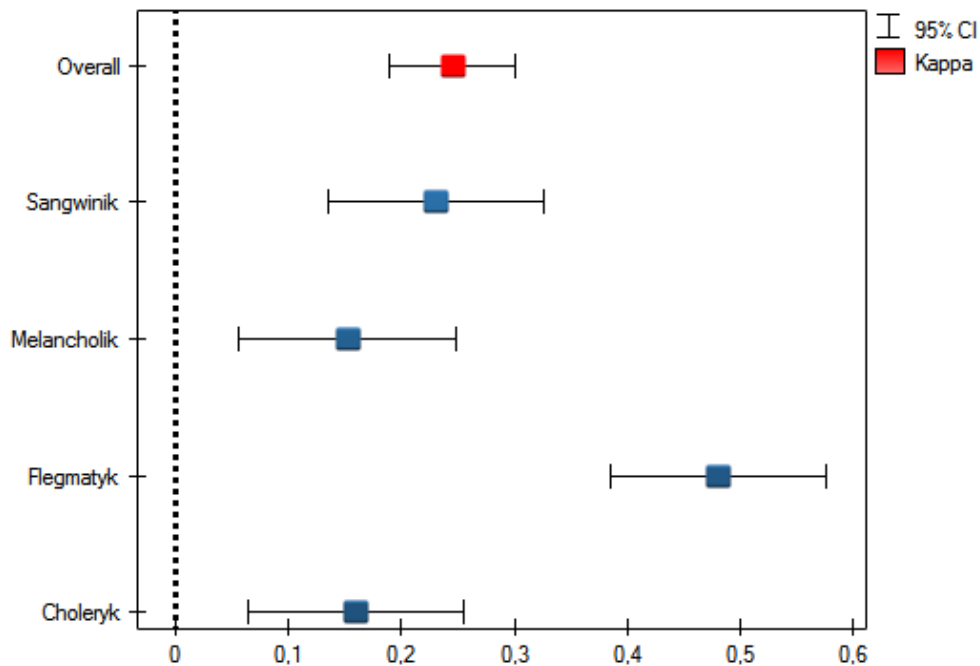
$$\mathcal{H}_0 : \kappa = 0,$$

$$\mathcal{H}_1 : \kappa \neq 0.$$

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

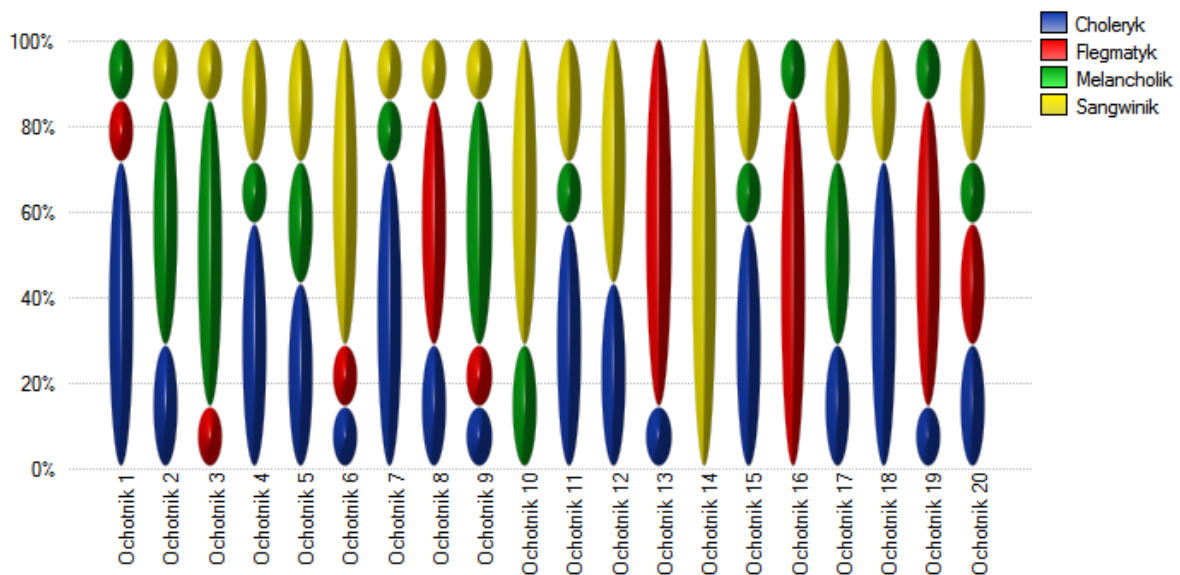
Obserwujemy nieduży współczynnik Kappa = 0.244918, lecz istotny statystycznie ($p < 0.000001$), co oznacza nieprzypadkową zgodność ocen sędziowskich. Istotna zgodność dotyczy każdej oceny, czego potwierdzeniem jest raport podsumowujący zgodność dla każdej warstwy (dla każdej oceny) oraz wykres prezentujący poszczególne współczynniki Kappa i Kappę podsumowującą całość.

Podsumowanie zgodności				
	Choleryk	Flegmatyk	Melancholii	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 (asymptoty:	0,001084	<0.000001	0,001867	0,000002



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.

21 TESTY DIAGNOSTYCZNE

21.1 OCENA TESTU DIAGNOSTYCZNEGO

Założmy, 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.

- **Czułość i swoistość testu diagnostycznego**

Każdy test diagnostyczny może w niektórych przypadkach uzyskać wyniki różne od wyników rzeczywistych, na przykład test diagnostyczny na podstawie otrzymanych parametrów klasyfikuje pacjenta do grupy osób chorych na daną chorobę, bądź zdrowych. W rzeczywistości liczba osób zakwalifikowanych do powyższych grup przez test może się różnić od liczby osób rzeczywiście zdrowych i rzeczywiście chorych.

Stosowane są dwie miary oceny trafności testu diagnostycznego. Są to:

Czułość (*ang. sensitivity*) – opisuje zdolność wykrywania osób rzeczywiście chorych (posiadających daną cechę). Jeśli więc badamy grupę osób chorych, to czułość daje nam informacje jaki procent z nich ma pozytywny wynik testu.

$$\text{czułość} = \frac{TP}{TP + FN}$$

Przedział ufności budowany jest w oparciu o metodę [Cloppera-Pearsona](#) dla pojedynczej proporcji.

Swoistość (*ang. specificity*) – opisuje zdolność wykrywania osób rzeczywiście zdrowych (bez danej cechy). Jeśli więc badamy grupę osób zdrowych, to swoistość daje nam informacje jaki procent z nich ma negatywny wynik testu.

$$\text{swoistość} = \frac{TN}{FP + TN}$$

Przedział ufności budowany jest w oparciu o metodę [Cloppera-Pearsona](#) dla pojedynczej proporcji.

- **Wartości predykcyjne dodatnie i ujemne oraz współczynnik chorobowości**

Wartość predykcyjna dodatnia (ang. *positive predictive value* – *PPV*) – prawdopodobieństwo, że osobnik miał 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.

$$PPV = \frac{TP}{TP + FP}$$

Przedział ufności budowany jest w oparciu o metodę **Cloppera-Pearsona** dla pojedynczej proporcji.

Wartość predykcyjna ujemna (ang. *negative predictive value* – *NPV*) – prawdopodobieństwo, że osobnik nie miał 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.

$$NPV = \frac{TN}{FN + TN}$$

Przedział ufności budowany jest w oparciu o metodę **Cloppera-Pearsona** dla pojedynczej proporcji.

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.

$$\text{Współczynnik chorobowości} = \frac{TP + FN}{n}$$

Przedział ufności budowany jest w oparciu o metodę **Cloppera-Pearsona** dla pojedynczej proporcji.

- **Iloraz wiarygodności wyniku dodatniego i iloraz wiarygodności wyniku ujemnego**

Iloraz wiarygodności wyniku dodatniego (ang. *likelihood ratio of positive test* – *LR₊*) – 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 na to, że pozytywny wynik testu otrzyma osoba z grupy chorych do szansy, że ten sam efekt będzie obserwowany wśród osób zdrowych.

$$LR_+ = \frac{\text{czułość}}{1 - \text{swoistość}} = \frac{TP(TP + FN)}{FP(FP + TN)}$$

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 (ang. *likelihood ratio of negative test* – *LR₋*) – jest to iloraz dwóch szans: szansy na to, że negatywny wynik testu otrzyma osoba z grupy chorych do szansy, że ten sam efekt będzie obserwowany wśród osób zdrowych.

$$LR_- = \frac{1 - \text{czułość}}{\text{swoistość}} = \frac{FN(TP + FN)}{TN(FP + TN)}$$

Przedział ufności dla *LR₋* buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{\text{czułość}}{FN} + \frac{1 - \text{swoistość}}{TN}}$$

- **Dokładność**

Dokładność (*ang. Accuracy – Acc*) – 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.

$$Acc = \frac{TP + TN}{n}$$

Przedział ufności budowany jest w oparciu o metodę **Cloppera-Pearsona** dla pojedynczej proporcji.

- **Diagnostyczny iloraz szans**

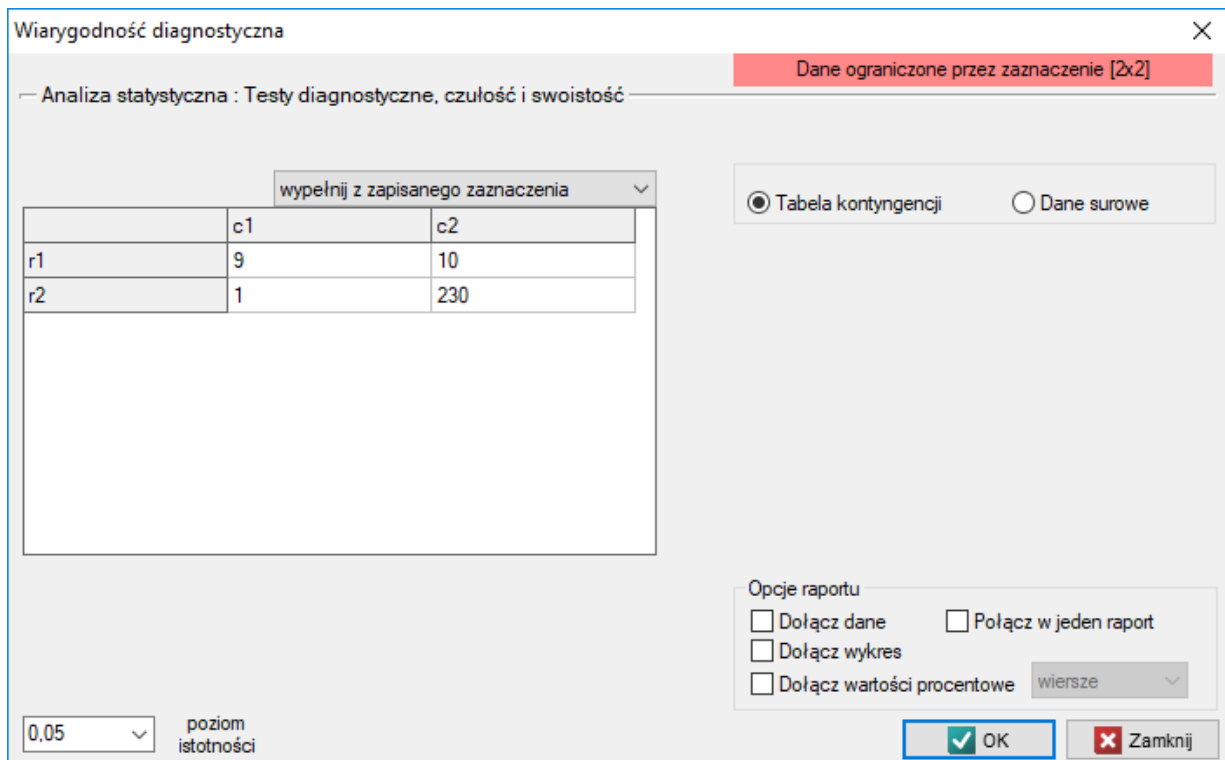
Diagnostyczny iloraz szans (*ang. Diagnostic Odds Ratio – DOR*) – jest to iloraz dwóch szans: szansy na pozytywny wynik testu osoby chorej do szansy na pozytywny wynik testu osoby zdrowej.

$$DOR = \frac{TP/FN}{FP/TN}$$

Przedział ufności dla *DOR* buduje się w oparciu o błąd standardowy:

$$SE = \sqrt{\frac{1}{TP} + \frac{1}{FN} + \frac{1}{FP} + \frac{1}{TN}}$$

Okno z ustawieniami opcji wiarygodności diagnostycznej wywołujemy poprzez menu Statystyka zaawansowana → Testy diagnostyczne → Wiarygodność diagnostyczna



Wiarygodność diagnostyczna

Dane ograniczone przez zaznaczenie [2x2]

Analiza statystyczna : Testy diagnostyczne, czułość i swoistość

wypełnij z zapisanego zaznaczenia

	c1	c2
r1	9	10
r2	1	230

Tabela kontyngencji Dane surowe

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

Dołącz wartości procentowe wiersze

0,05 poziom istotności

PRZYKŁAD 21.1. (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.

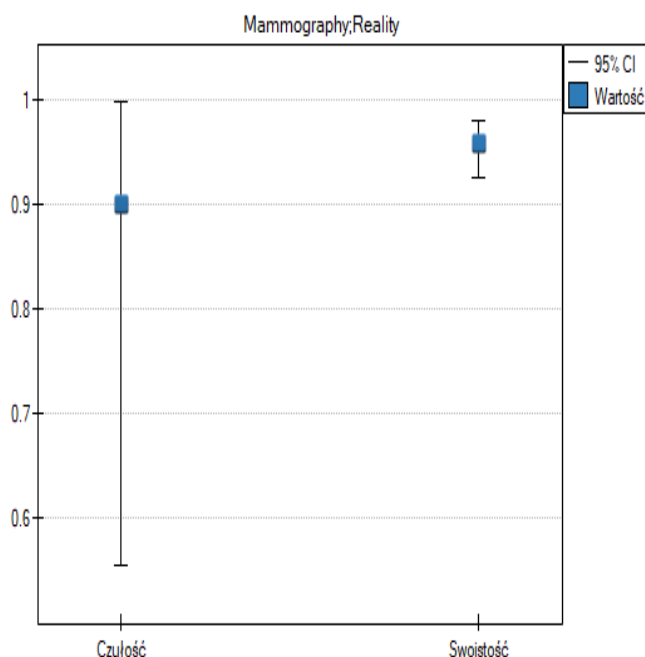
Liczności obserwowane		Rzeczywistość (badanie histopatologiczne)		
		choroba (+)	brak choroby (-)	Suma
test diagnostyczny	wynik pozytywny (+)	9	10	19
	wynik negatywny (-)	1	230	231
	Suma	10	240	250

Wyznamy 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%



- 90% kobiet chorych na raka piersi zostało poprawnie zdiagnozowanych, czyli uzyskało pozytywny wynik mammografii;
- 95.83% kobiet zdrowych (nie chorujących na raka piersi) zostało poprawnie zdiagnozowanych, czyli uzyskało negatywny wynik mammografii;
- 4 kobiety na 100 przebadanych cierpi 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 (nie chorujących na raka piersi);
- Szansa na to, że negatywny wynik mammografii otrzyma kobieta rzeczywiście chora na raka stanowi 10.43% szansyna to, że negatywny wynik mammografii otrzyma kobieta rzeczywiście zdrowa (nie chorujących na raka piersi);
- Kobieta poddająca się mammografii (bez względu na uzyskany wynik) może być pewna postawionej diagnozy w 96.50%;
- Szansy na pozytywny wynik testu u kobiety która rzeczywiście choruje na raka piersi jest 207 razy większa od szansy na taki wynik kobiety zdrowej.

21.2 KRZYWA ROC

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 wartości prawdziwie dodatnich (TP), prawdziwie ujemnych (TN), fałszywie dodatnich (FP) i fałszywie ujemnych (FN). 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 szansę na wystąpienie badanego zjawiska lub odwrotnie: wraz ze wzrostem wartości zmiennej diagnostycznej maleją szansę 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.

Założmy, 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, \dots, 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).

stymulanta		Rzeczywistość	
		(+)	(–)
zmienna diagnostyczna	$x_i \geq x_{cat}$ (+)	TP	FP
	$x_i < x_{cat}$ (–)	FN	TN

destymulanta		Rzeczywistość	
		(+)	(–)
zmienna diagnostyczna	$x_i \leq x_{cat}$ (+)	TP	FP
	$x_i > x_{cat}$ (–)	FN	TN

Na podstawie tych wartości każdy punkt odcięcia x_{cat} może być dalej opisany za pomocą **czułości** i **swoistości** oraz wartości predykcyjnych dodatnich (PPV), wartości predykcyjnych ujemnych (NPV), ilorazu wiarygodności wyniku dodatniego (LR_+), ilorazu wiarygodności wyniku ujemnego (LR_-) i dokładności (Acc).

Uwaga!

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, to 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 z wzorami Bayesa:

$$PPV_{revised} = \frac{\text{Czułość} \cdot P_{apriori}}{\text{Czułość} \cdot P_{apriori} + (1 - \text{Swoistość}) \cdot (1 - P_{apriori})}$$

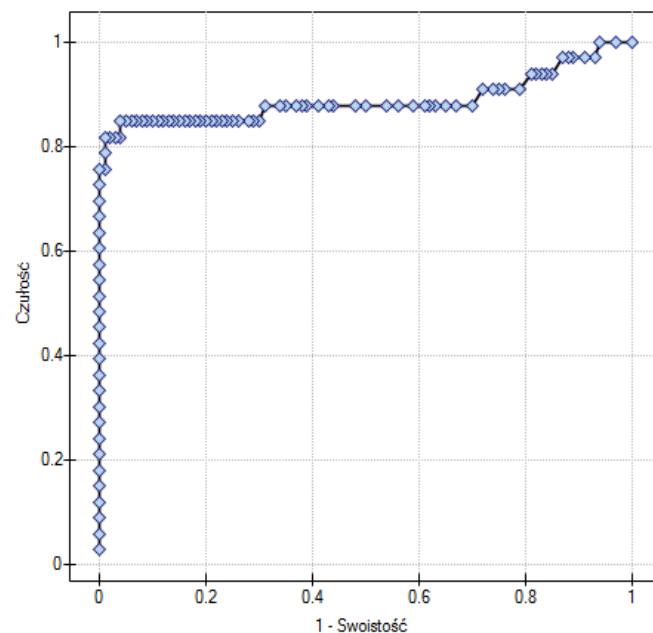
$$NPV_{revised} = \frac{\text{Swoistość} \cdot (1 - P_{apriori})}{\text{Swoistość} \cdot (1 - P_{apriori}) + (1 - \text{Czułość}) \cdot P_{apriori}}$$

gdzie:

$P_{apriori}$ - zadany przez użytkownika współczynnik chorobowości, tzw. pre-test probability of disease

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}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_k	czułość _k	swoistość _k	PPV _k	NPV _k	LR _{+k}	LR _{-k}	Acc _k	PPV _{revk}	NPV _{revk}

Krzywa ROC powstaje na podstawie wyznaczonych wartości czułości i swoistości. Na osi odciętych umieszczona jest $x = 1 - \text{swoistość}$, a na osi rzędnych $y = \text{czułość}$. Uzyskane punkty są ze sobą połączone. Powstała w ten sposób krzywa, a w szczególności pole pod nią, obrazuje jakość klasyfikacyjną analizowanej zmiennej diagnostycznej. Gdy krzywa ROC pokrywa się z przekątną $y = x$, to decyzja podejmowana na podstawie zmiennej diagnostycznej jest tak samo dobra jak losowy podział badanych obiektów na grupy (+) i (-).



AUC (*ang. area under curve*) – wielkość pola pod krzywą ROC mieści się w przedziale $< 0; 1 >$. Im większe jest pole, tym dokładniej zaklasyfikujemy obiekty do grupy (+) i (-) na podstawie analizowanej zmiennej diagnostycznej. Zatem z tym lepszym skutkiem ta zmienna diagnostyczna może być wykorzystywana jako klasyfikator. Pole AUC , błąd SE_{AUC} i przedział ufności dla AUC wyliczane są w oparciu:

- ★ o metodę nieparametryczną **DeLong** (DeLong E.R. i inni 1988[50], Hanley J.A. i Hajian-Tilaki K.O. 1997[73]) - **rekomendowane**,
- ★ o metodę nieparametryczną **Hanley-McNeil** (Hanley J.A. i McNeil M.D. 1982[74]),
- ★ o metodę **Hanley-McNeil** zakładającą dwu-ujemny rozkład wykładniczy (Hanley J.A. i McNeil M.D. 1982[74]) - wyliczną tylko wtedy, gdy grupy (+) i (–) są równoliczne.

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:

$$\begin{aligned}\mathcal{H}_0 &: AUC = 0.5, \\ \mathcal{H}_1 &: AUC \neq 0.5.\end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{AUC - 0.5}{SE_{0.5}},$$

gdzie:

$$SE_{0.5} = \sqrt{\frac{n_{(+)} + n_{(-)} + 1}{12n_{(+)}n_{(-)}}},$$

$n_{(+)}$ – liczność grupy (+), w której dane zjawisko rzeczywiście występuje,

$n_{(-)}$ – liczność grupy (–), w której dane zjawisko rzeczywiście nie występuje.

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Dodatkowo, gdy przyjmiemy że parametr diagnostyczny tworzy wysokie pole (AUC), możemy wybrać **optymalny punkt odcięcia**.

21.2.1 Wybór optymalnego punktu odcięcia

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.

Program PQStat wybór optymalnego punktu odcięcia umożliwia poprzez:

- **Metodę stycznej (indeks kosztów)** – wyliczany w oparciu o czułość, swoistość, koszty błędnych decyzji i współczynnik chorobowości.

Błędy jakie można popełnić przydzielając badane objekty 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, ...), to wówczas w polu koszt FP i w polu koszt FN wpisujemy tą samą dodatnią wartość – zwykle 1. Jeśli natomiast uznamy, że jeden rodzaj błędu

jest obarczony większym kosztem niż drugi, wówczas przypiszemy mu odpowiednio większą wagę.

Optymalna wartość odcięcia obliczana jest na podstawie czułości, swoistości i przy użyciu wielkości m – nachylenia stycznej do krzywej ROC. Kąt nachylenia m jest określany w odniesieniu do dwóch wartości: 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. W rezultacie 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[179]).

Wybrany w ten sposób optymalny punkt odcięcia zmiennej diagnostycznej zostanie ostatecznie zaznaczony na wykresie krzywej ROC.

- **Indeks Youdena** – Konceptyjnie jest to maksymalna odległość pomiędzy linią będącą przekątną kwadratu o boku 1 a punktem krzywej ROC [176]. Indeks ten obliczany jest z wzoru:

$$d = \text{Czułość} + \text{Swoistość} - 1$$

Wybrany w ten sposób optymalny punkt odcięcia zmiennej diagnostycznej zostanie ostatecznie zaznaczony na wykresie krzywej ROC.

- **Odległość do lewego górnego rogu** – Konceptyjnie 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. Indeks ten obliczany jest z wzoru:

$$d = \sqrt{(1 - \text{Czułość})^2 + (1 - \text{Swoistość})^2}$$

Wybrany w ten sposób optymalny punkt odcięcia zmiennej diagnostycznej zostanie ostatecznie zaznaczony na wykresie krzywej ROC.

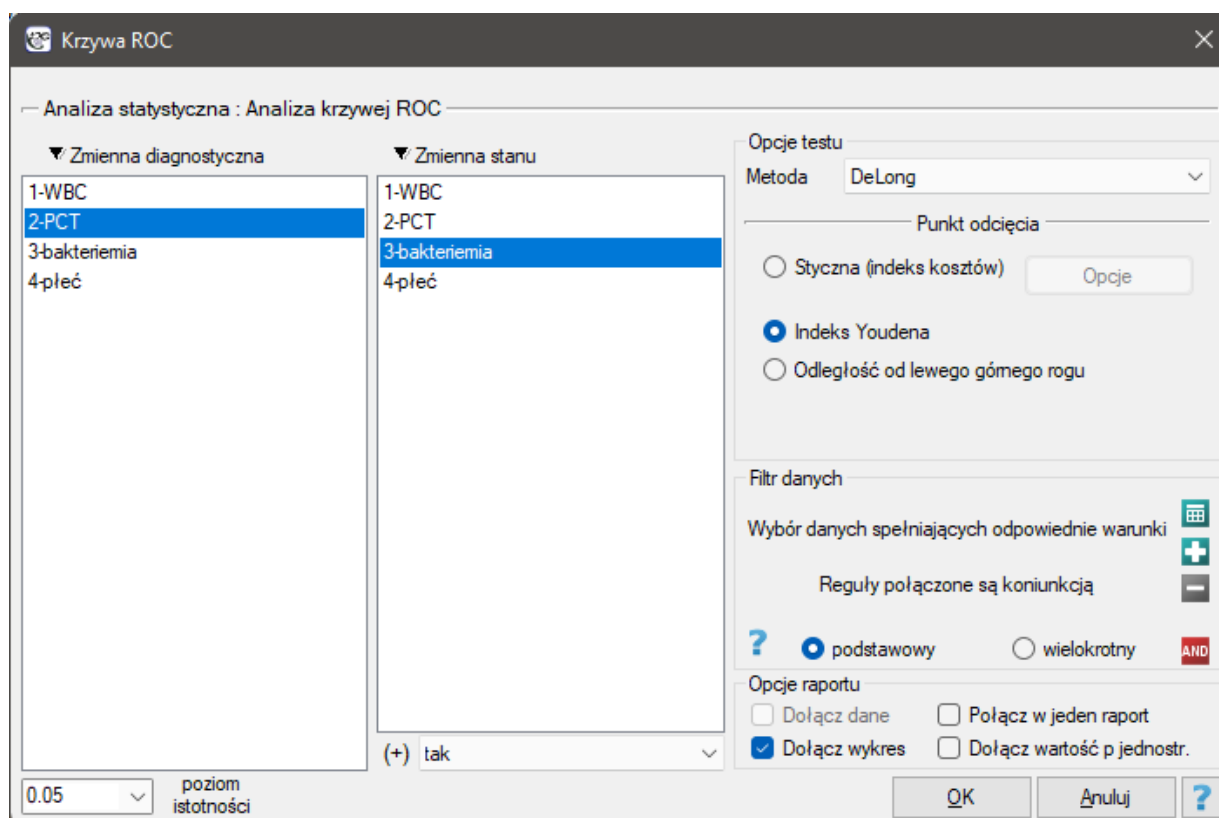
- **Wykres kosztów** – prezentuje wyliczone wartości błędnej diagnozy wraz z ich kosztami. Wartości te wyliczane są zgodnie z wzorem:

$$\text{koszt} = \text{koszt}_{FP} \cdot FP + \text{koszt}_{FN} \cdot FN$$

Zaznaczony na wykresie punkt, to minimum powyższej funkcji.

- **Wykres przecięcia czułości i swoistości** – pozwala na zlokalizowanie punktu, w którym wartość czułości i swoistości jest jednocześnie największa.

Okno z ustawieniami opcji analizy ROC wywołujemy poprzez menu Statystyka zaawansowana → Testy diagnostyczne → Krzywa ROC.


PRZYKŁAD 21.2. (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/ μ l, 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 $> 39^{\circ}\text{C}$ przedstawia poniższy fragment tabeli:

WBC	PCT	bakteriemia	pleć
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

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 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. Jeśli tak, jaki jest optymalny punkt odcięcia powyżej którego możemy uznać badanie za "pozytywne" – wykrywające bakterie.

By sprawdzić, czy PCT jest rzeczywiście przydatny w rozpoznawaniu bakteriemii wyliczymy wielkość pola pod krzywą ROC i zweryfikujemy hipotezę, że:

$$H_0 : \text{pole pod zbudowaną krzywą ROC} = 0.5,$$

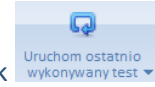
$$H_1 : \text{pole pod zbudowaną krzywą ROC} \neq 0.5.$$

Ponieważ bakterie towarzyszy podwyższony poziom PCT, to wskaźnik ten uznajemy za stymulantę. W zmiennej stanu musimy określić, która wartość znajdująca się w kolumnie bakterie 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;bakterie
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	<0.00000001

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

Wyliczona wielkość pola pod krzywą ROC wynosi $AUC = 0.889$. Zatem na podstawie przyjętego poziomu $\alpha = 0.05$, w oparciu o uzyskaną wartość $p < 0.000001$ wnioskujemy, że rozpoznawanie bakteriemii przy użyciu wskaźnika PCT jest istotnie korzystniejsze niż losowy podział pacjentów na 2 grupy: chorujących na bakterię i nie chorujących. Wracamy więc do analizy (przycisk



), by wyznaczyć optymalny punkt odcięcia.

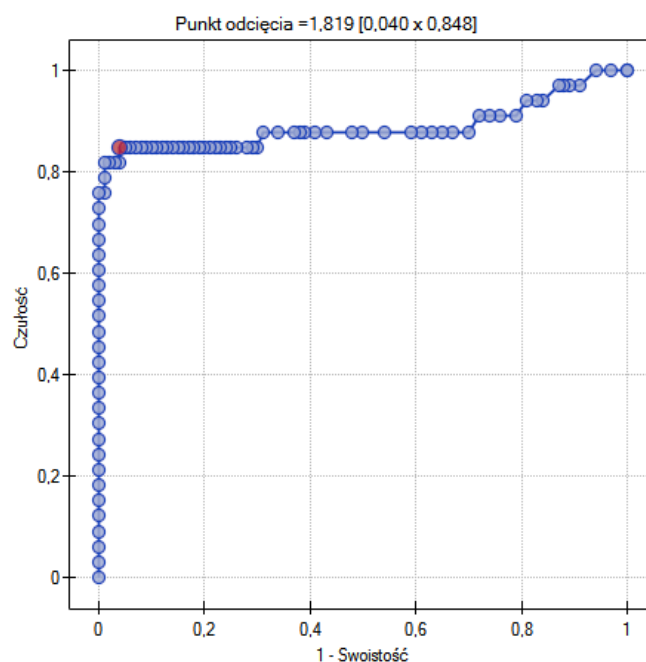
Algorytm poszukiwania optymalnego punktu odcięcia uwzględnia koszty błędnych decyzji i współczynnik chorobowości:

- (1) koszt FN - błędna diagnoza, to koszt uznania, że pacjent nie choruje na bakterię mimo, że rzeczywiście jest on chory (koszty decyzji fałszywie ujemnej)
- (2) koszt FP - błędna diagnoza, to koszt uznania, że pacjent choruje na bakterię mimo, że rzeczywiście na nią nie choruje (koszty decyzji fałszywie dodatniej)

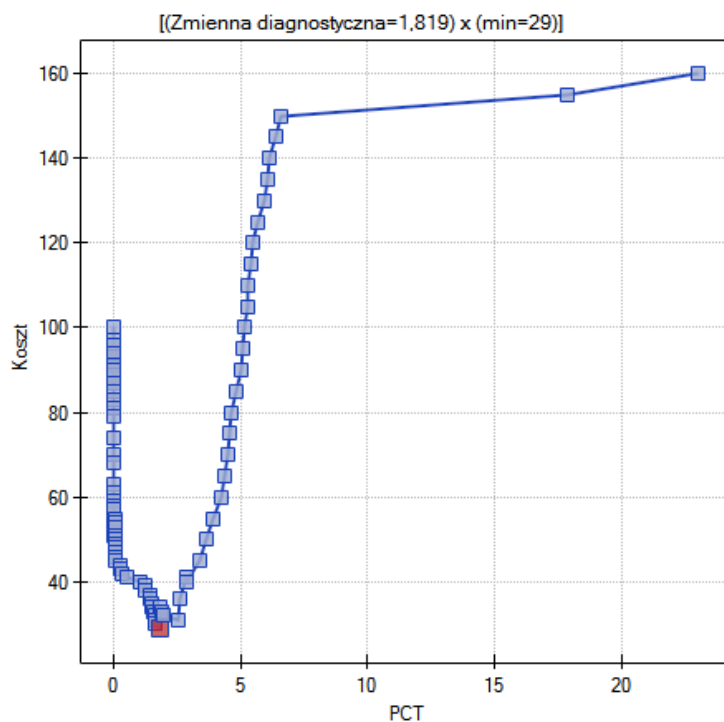
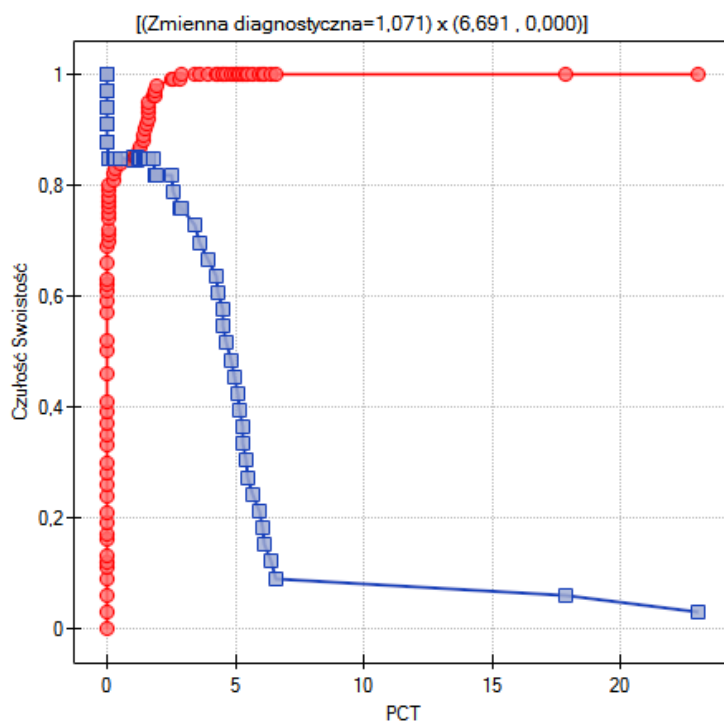
Ponieważ koszty FN są znacznie poważniejsze niż koszty FP, to w polu pierwszym wpisujemy wartość większą niż w polu drugim. Uznaliśmy, że będzie to wartość 5.

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 liczbę 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



Wyznaczony optymalny punkt odcięcia PCT to 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:



Analiza wykresu kosztów wskazuje, że minimum kosztów błędnych decyzji przypada na $PCT=1.819$. Natomiast wartość czułości i swoistości jest podobna dla $PCT=1.071$

21.2.2 Porównywanie krzywych ROC

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.

- **Model zależny** – porównywane krzywe ROC powstają na bazie pomiarów dokonanych na tych samych obiektach.

Hipotezy:

$$\mathcal{H}_0 : AUC_1 = AUC_2,$$

$$\mathcal{H}_1 : AUC_1 \neq AUC_2.$$

Statystyka testowa ma postać:

$$Z = \frac{|AUC_1 - AUC_2|}{SE_{AUC_1 - AUC_2}},$$

gdzie:

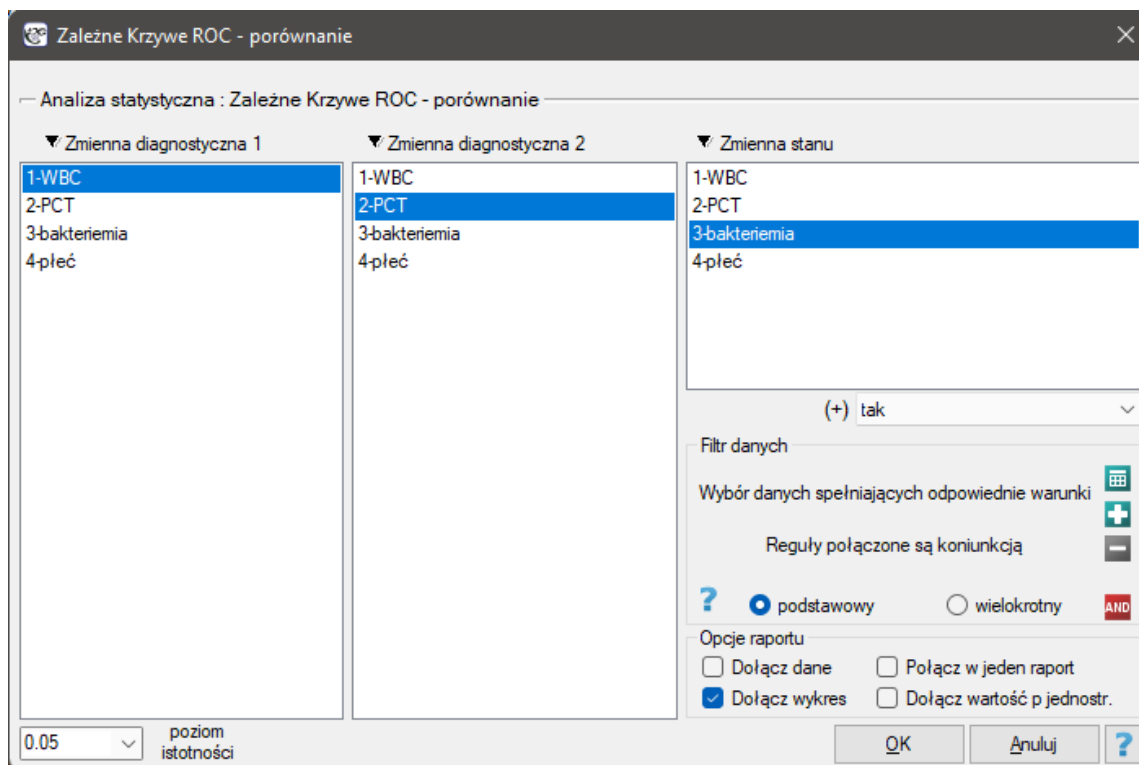
AUC_1 , AUC_2 i błąd standardowy różnicy pól $SE_{AUC_1 - AUC_2}$ wyliczane są w oparciu o metodę nieparametryczną zaproponowaną przez **DeLong** (DeLong E.R. i inni 1988[50], Hanley J.A. i Hajian-Tilaki K.O. 1997[73])

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji porównywania zależnych krzywych ROC wywołujemy poprzez menu Statystyka zaawansowana → Testy diagnostyczne → Zależne Krzywe ROC – porównywanie.



- **Model niezależny** – porównywane krzywe ROC powstają na bazie pomiarów dokonanych na różnych obiektach.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: AUC_1 = AUC_2, \\ \mathcal{H}_1 &: AUC_1 \neq AUC_2.\end{aligned}$$

Statystyka testowa (Hanley J.A. i McNeil M.D. 1983[75]) 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:

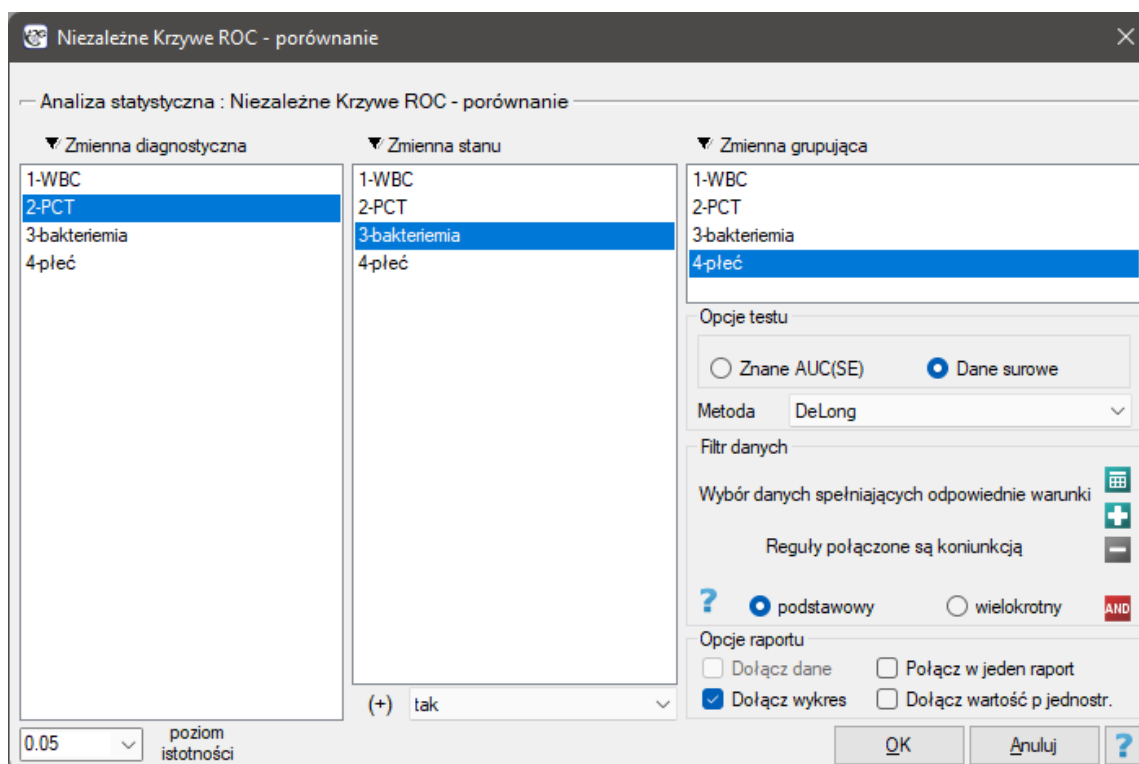
- ★ o metodę nieparametryczną **DeLong** (DeLong E.R. i inni 1988[50], Hanley J.A. i Hajian-Tilaki K.O. 1997[73]) - **rekomendowane**,
- ★ o metodę nieparametryczną **Hanley-McNeil** (Hanley J.A. i McNeil M.D. 1982[74]).

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Okno z ustawieniami opcji porównywania niezależnych krzywych ROC wywołujemy poprzez menu Statystyka zaawansowana → Testy diagnostyczne → Niezależne Krzywe ROC – porównywanie.



PRZYKŁAD (21.2) c.d. (plik bakteriemia.pqs)

Wykonamy 2 porównania:

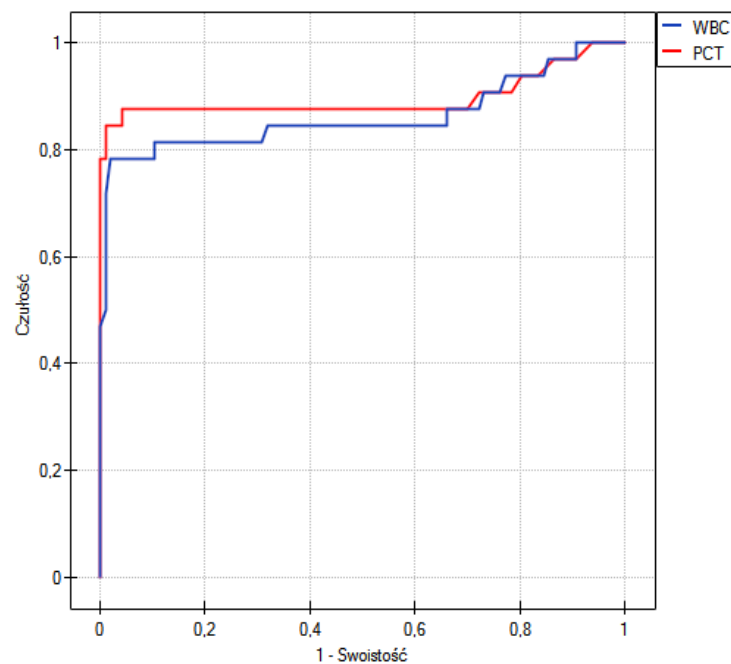
- 1) Zbudujemy 2 krzywe ROC, by porównać wartość diagnostyczną parametrów WBC i PCT;
- 2) Zbudujemy 2 krzywe ROC, by porównać wartość diagnostyczną parametru PCT dla chłopców i dziewczynek.

ad1) Zarówno parametr WBC jak i PCT jest stymulantą (wysokie wartości tych parametrów towarzyszą bakteriemii). Porównując wartość diagnostyczną tych parametrów weryfikujemy hipotezy:

\mathcal{H}_0 : pole pod krzywą ROC dla WBC = pole pod krzywą ROC dla PCT,

\mathcal{H}_1 : pole pod krzywą ROC dla WBC \neq pole pod krzywą ROC dla PCT.

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



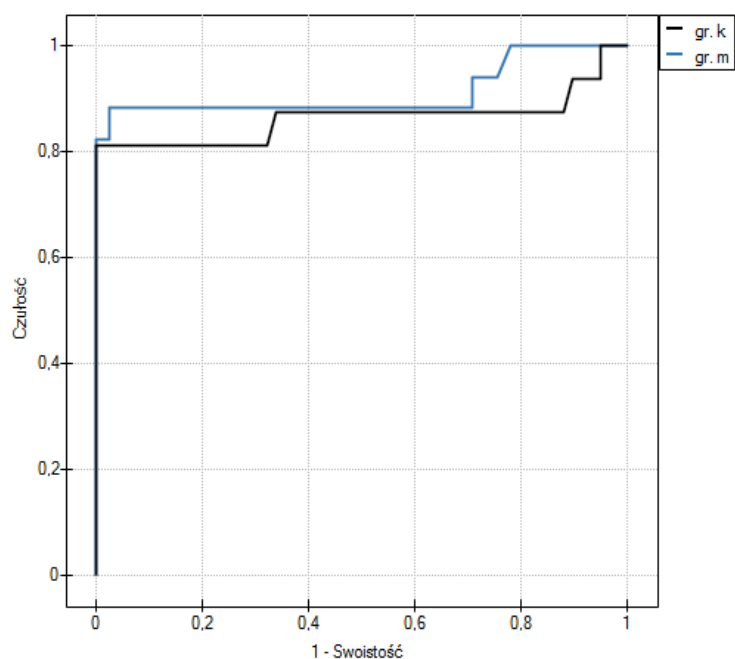
Wyliczone wielkości pól to $AUC_{WBC} = 0.8613$, $AUC_{PCT} = 0.8956$. Na podstawie przyjętego poziomu $\alpha = 0.05$, w oparciu o uzyskaną wartość $p=0.130321915$ wnioskujemy, że nie możemy wskazać, który z parametrów WBC czy PCT jest lepszy w rozpoznawaniu bakteriemii.

ad2) Parametr PCT jest stymulantą (jego wysokie wartości towarzyszą bakteriemii). Porównując jego wartość diagnostyczną dla dziewczynek i chłopców weryfikujemy hipotezy:

\mathcal{H}_0 : pole pod krzywą ROC dla $PCT_k =$ pole pod krzywą ROC dla PCT_m ,

\mathcal{H}_1 : pole pod krzywą ROC dla $PCT_k \neq$ pole pod krzywą ROC dla PCT_m .

Niezależne Krzywe ROC - porównanie	
Analizowane zmienne	PCT
	bakteriemia
Liczność / niezinterpretowane	0
Liczność / braki danych	1
Poziom istotności	0.05
Zmienna grupująca	pleć
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



Wyliczone wielkości pól to $AUC_k = 0.8649$, $AUC_m = 0.9118$. Zatem na podstawie przyjętego poziomu $\alpha = 0.05$, w oparciu o uzyskaną wartość $p=0.637176453$ wnioskujemy, że nie możemy wybrać płci, dla której parametr PCT jest lepszy w rozpoznawaniu bakteriemii.

22 MODELE WIELOWYMIAROWE

23 DOPASOWANIE GRUP

Dlaczego dokonuje się dopasowania grup?

Odpowiedzi na to pytanie jest bardzo wiele. Posłużymy się przykładem sytuacji medycznych.

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 byłoby nieetycznym wykonanie badania z pełną randomizacją ponieważ oznaczałoby ono, 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. 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/nie palenia 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 bada i kontrolna jest podobna 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.

Sposoby oceny podobieństwa:

Dwa pierwsze wymienione sposoby opierają się na dopasowaniu grup poprzez **dopasowanie wyników skłonności** (*ang. Propensity Score Matching, PSM*). Ten rodzaj dopasowania został zaproponowany przez Rosenbauma i Rubina [137]. W praktyce jest to technika doboru grupy kontrolnej (osób nie leczonych 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 skłonności, z angielskiego nazywany *Propen-*

sity 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.

Metody:

- Zadane prawdopodobieństwo – czyli podaną dla każdej badanej osoby wartość z przedziału od 0 do 1, określającą prawdopodobieństwo przynależenia do grupy leczonej czyli Propensity Score. Takie prawdopodobieństwo może zostać wyznaczone wcześniej różnymi metodami. Na przykład w modelu regresji logistycznej, poprzez sieci neuronowe lub wiele innych metod. Jeśli osoba z grupy, z której losujemy kontrole 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 wyznaczenia automatycznie 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, tzn. na podobieństwie jednych zmiennych może badaczowi zależeć bardziej, podczas gdy podobieństwo innych jest mniej ważne. Jednak w przypadku wyboru macierzy odległości/niepodobieństwa zalecana jest duża ostrożność. Wiele cech i wiele sposobów wyznaczenia 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 niepodobieństwa/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 tym przypadku osoby leczone i nieleczone. 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 określonej przez niepodobieństwo, najbardziej niepodobne obiekty to te odległe o 1, a najbardziej podobne, to te odległe o 0. Wybierając więc tę metodą 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 wiąż 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 musi uwzględniać fakt, że ta sama nieleczona osoba może znajdować się w wielu dopasowanych zestawach.

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 połączona 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 zbiorów, 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, to 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 zawsze 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 zostają z założeniem hipotezy zerowej, a więc nie wykazują istotnych różnic ze względu na zmniejszoną liczebność.

Dla porównania zmiennych ciągłych wyznaczamy standaryzowaną różnicę średnich:

$$d = \frac{(\bar{x}_{leczona} - \bar{x}_{kontrolna})}{\sqrt{\frac{sd_{leczona}^2 + sd_{kontrolna}^2}{2}}}$$

gdzie:

$\bar{x}_{leczona}, \bar{x}_{kontrolna}$ - to średnia wartość zmiennej w grupie leczonej i średnia wartość zmiennej w grupie kontrolnej,

$sd_{leczona}^2, sd_{kontrolna}^2$ - to 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:

$\hat{p}_{leczona}, \hat{p}_{kontrolna}$ - to częstość wartości opisanej jako 1 w grupie leczonej i częstość wartości opisanej jako 1 w grupie kontrolnej.

Zmienne o wielu kategoriach powinniśmy rozbić w analizie regresji logistycznej na zmienne fikcyjne o dwóch kategoriach i sprawdzając dopasowania obu grup wyznaczać dla nich standaryzowaną różnicę częstości.

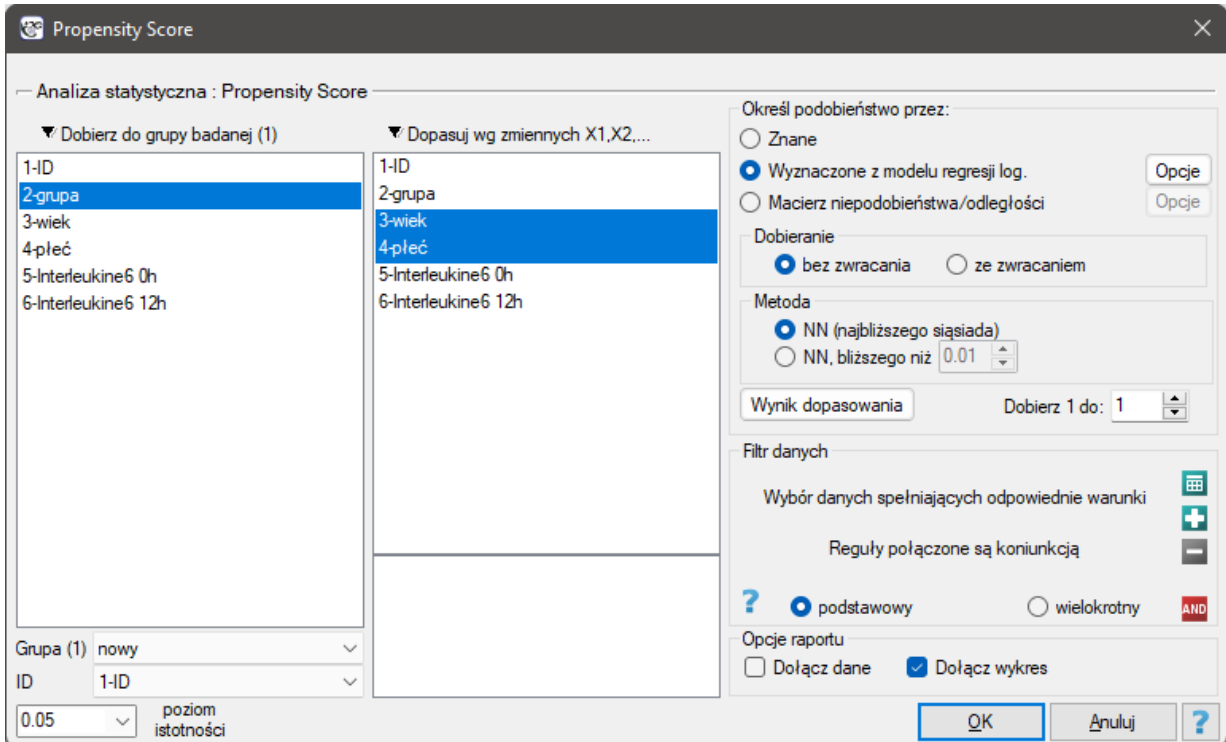
Uwaga!

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)[125]. 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.

Uwaga!

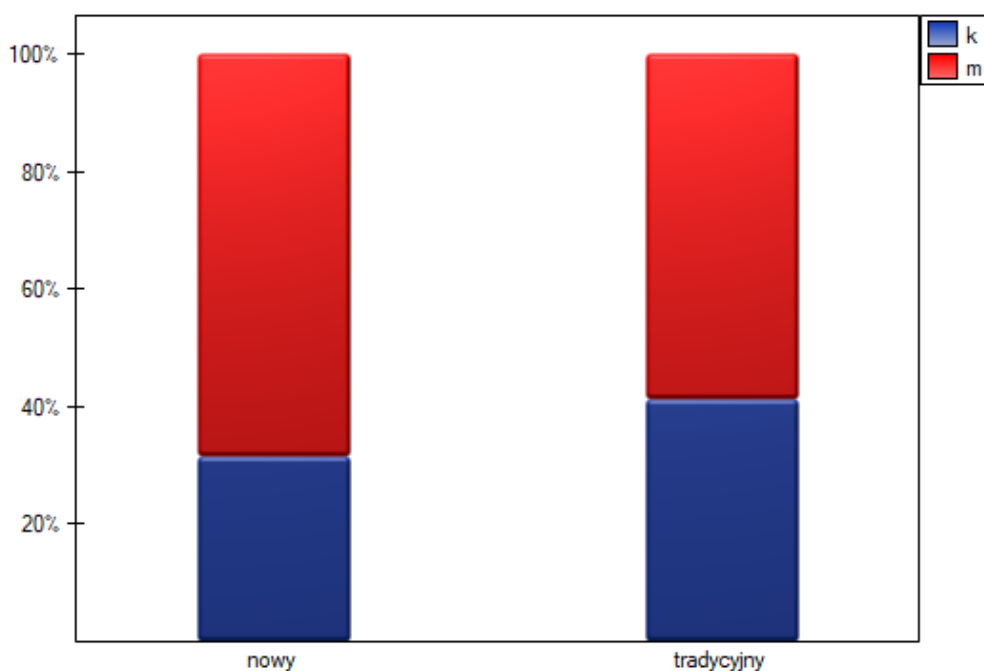
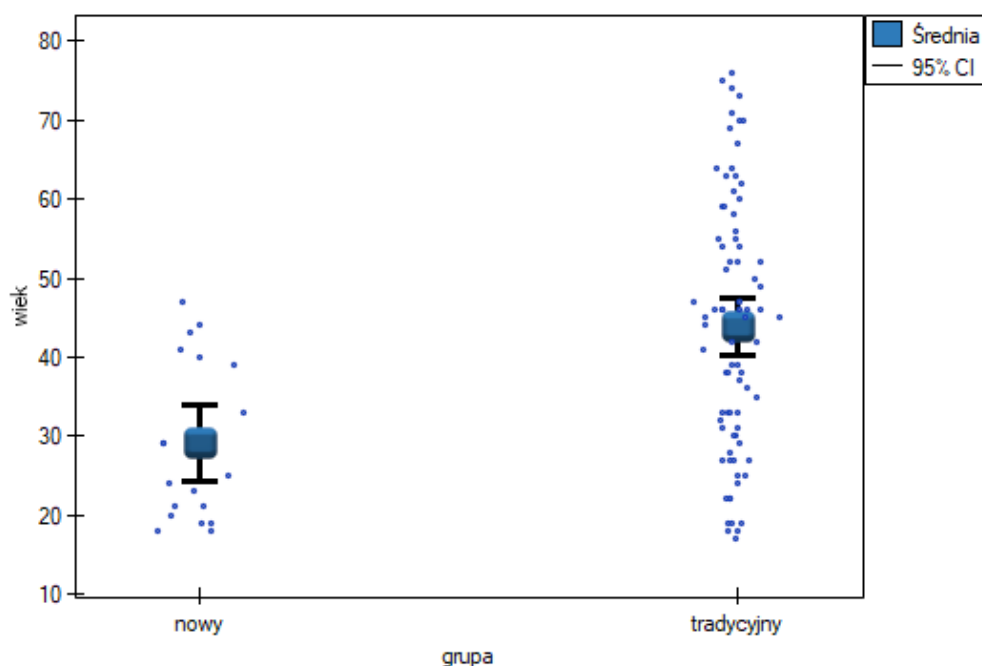
Uzyskane w raportach dopasowanie 1:1 oznacza podsumowanie dotyczące grupy badanej i odpowiadającej jej grupy kontrolnej uzyskanej w pierwszym dopasowaniu, dopasowanie 1:2 oznacza podsumowanie dotyczące grupy badanej i odpowiadającej jej grupy kontrolnej uzyskanej w pierwszym + drugim dopasowaniu (czyli nie dotyczy grupy badanej i odpowiadającej jej grupy kontrolnej uzyskanej tylko w drugim dopasowaniu), itd.

Okno z ustawieniami opcji dopasowania grup wywołujemy poprzez menu Statystyki zaawansowane → Modele wielowymiarowe → Dopasowanie grup



PRZYKŁAD 23.1. (plik dopasowanie.pl)

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 obydwie one 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 obniżyć poziom cytokin. Grupy były dość podobne co do struktury płci, ale nie identyczne.

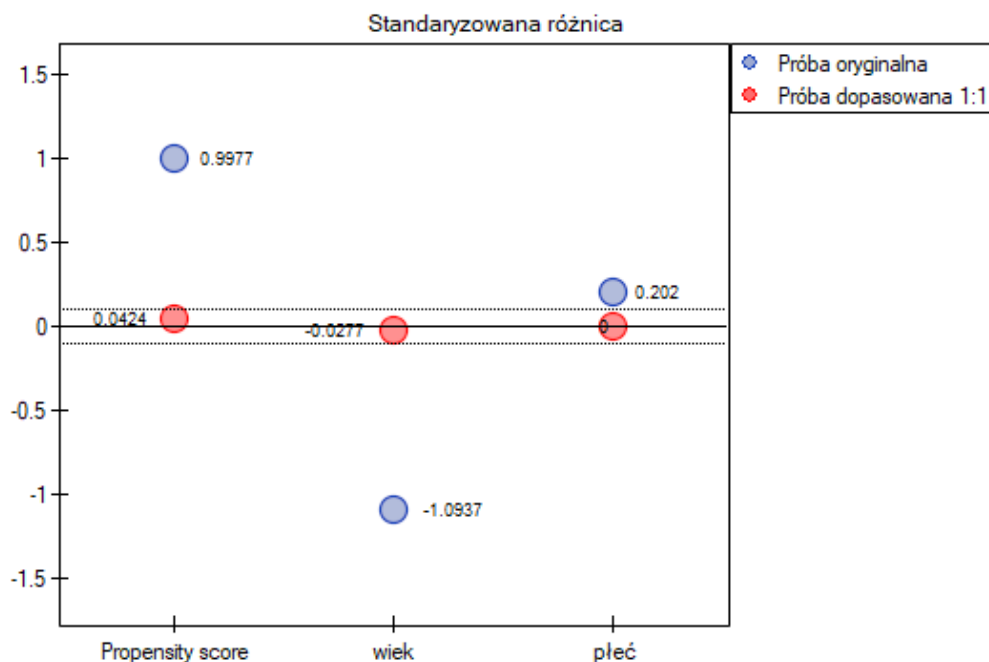


Gdyby przeprowadzono planowane badanie na tak wybranych grupach pacjentów, to nowy sposób miałby łatwiejsze zadanie do wykonania, gdyż młodsze organizmy lepiej mogłyby reagować na leczenie. Warunki eksperymentu nie byłyby równe dla obydwu sposobów, co mogłoby zafałszować wyniki analiz i wyciągane wnioski. 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. Pamiętajmy, 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.

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).



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 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). Standaryzowana różnica między grupami określona przez Propensity Score wynosi 0.0424, czyli mieści

sie w wyznaczonym przedziale. 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) a standaryzowana różnica między średnimi wynosi -0.0277. W przypadku płci dopasowanie jest idealne, tzn. odsetek kobiet i mężczyzn jest taki sam w obu grupach (standaryzowana różnica odsetków przedstawiona w tabeli i na wykresie wynosi teraz 0). Tak przygotowane dane możemy zwrócić do arkusza i podać 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.

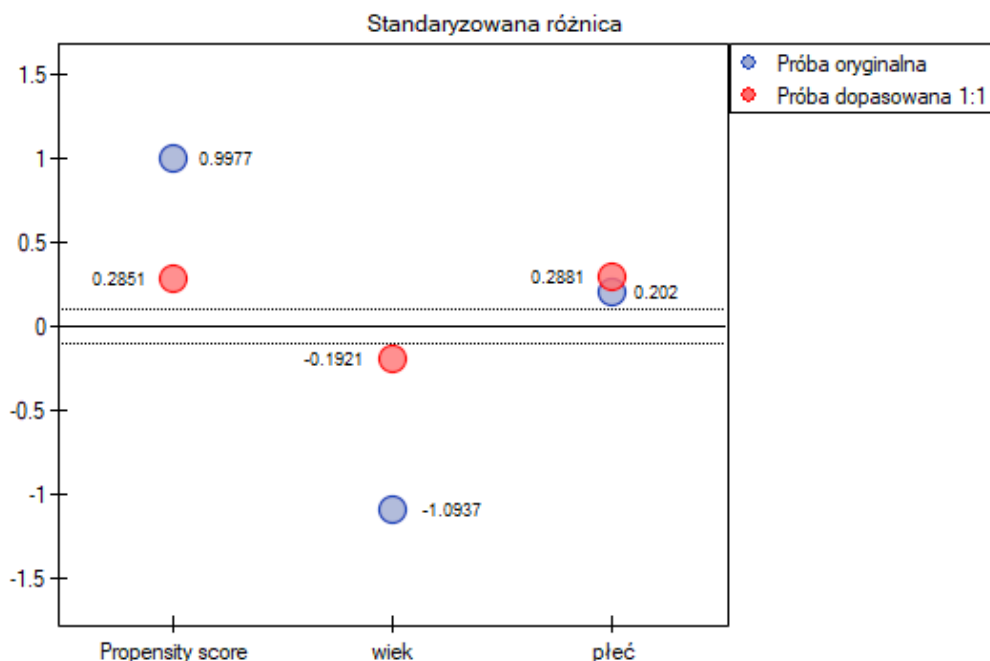
- Załóżmy, że chcielibyśmy uzyskać takie grupy, których Propensity Score (tzn. skłonność do wzięcia udziału w badaniu) różni się nie więcej niż ...
Jak ustalić tę wartość? Można zerknąć na raport z wcześniejszej analizy, gdzie podana jest najmniejsza i największa odległość między losowanymi obiektami.

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	pleć	ID	Propensity	wiek	pleć
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 z 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, to 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.

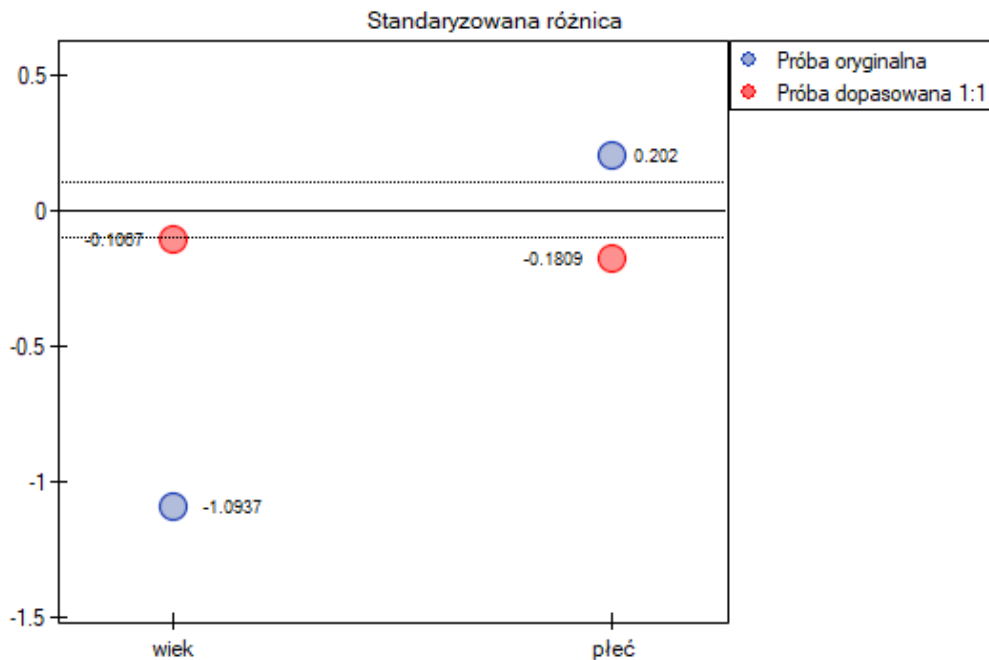


- Załóżmy, że chcielibyśmy uzyskać takie pary (osób leczonych nową metodą i osób leczonych tradycyjnie)

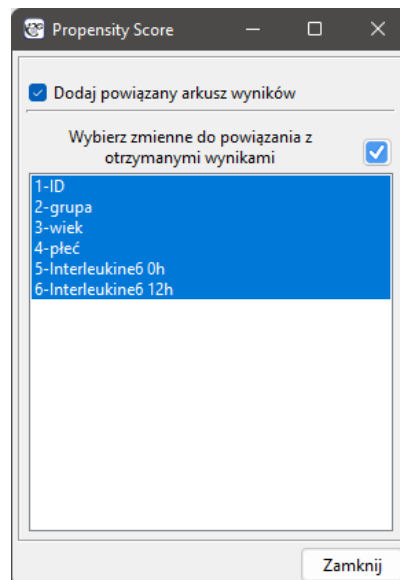
cyjnie), które są tej samej płci i których wiek nie różni się więcej niż 3 lata. W losowaniu opartym o Propensity Score nie mieliśmy tego typu możliwości wpływania na zakres zgodności każdej ze zmiennych. Do tego celu skorzystamy z innej metody, nie opartej na Propensity Score, ale bazującej na macierzach 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.

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, że 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.



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 znajdują się dane dotyczące osób leczonych nowym sposobem oraz dopasowanych do nich osób leczonych tradycyjnie.

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.

23.1 PRZYGOTOWANIE ZMIENNYCH DO ANALIZY W MODELACH WIELOWYMIAROWYCH

23.1.1 Kodowanie zmiennych w modelach wielowymiarowych

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.

$k = 2$ 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.

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).

$k = 2$ 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, \dots, X_{k-1} o kodowaniu zero-jedynkowym zostaną umieszczone w modelu regresji, wówczas zostaną dla nich wyliczone współczynniki b_1, b_2, \dots, 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	Wykształcenie	Zakodowane wykształcenie		
			zawodowe	średnie	wyższe
k	1	podstawowe	0	0	0
k	1	podstawowe	0	0	0
k	1	podstawowe	0	0	0
m	0	zawodowe	1	0	0
m	0	zawodowe	1	0	0
k	1	zawodowe	1	0	0
k	1	zawodowe	1	0	0
m	0	średnie	0	1	0
m	0	średnie	0	1	0
k	1	średnie	0	1	0
m	0	średnie	0	1	0
k	1	wyższe	0	0	1
m	0	wyższe	0	0	1
k	1	wyższe	0	0	1
m	0	wyższe	0	0	1
m	0	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 pół tysiąca złoty 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 złoty 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 złoty 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, \dots, X_{k-1} o kodowaniu efektu zostaną umieszczone w modelu regresji, wówczas zostaną dla nich wyliczone współczynniki b_1, b_2, \dots, 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.

Płeć	Zakodowana płeć	Region zamieszkania	Zakodowany region			
			zachodni	wschodni	północny	południowy
k	1	centralny	-1	-1	-1	-1
k	1	centralny	-1	-1	-1	-1
k	1	centralny	-1	-1	-1	-1
m	-1	zachodni	1	0	0	0
m	-1	zachodni	1	0	0	0
k	1	zachodni	1	0	0	0
k	1	zachodni	1	0	0	0
m	-1	wschodni	0	1	0	0
m	-1	wschodni	0	1	0	0
k	1	wschodni	0	1	0	0
m	-1	wschodni	0	1	0	0
k	1	północny	0	0	1	0
m	-1	północny	0	0	1	0
k	1	południowy	0	0	0	1
m	-1	południowy	0	0	0	1
m	-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 za-

robki 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.

23.1.2 Interakcje

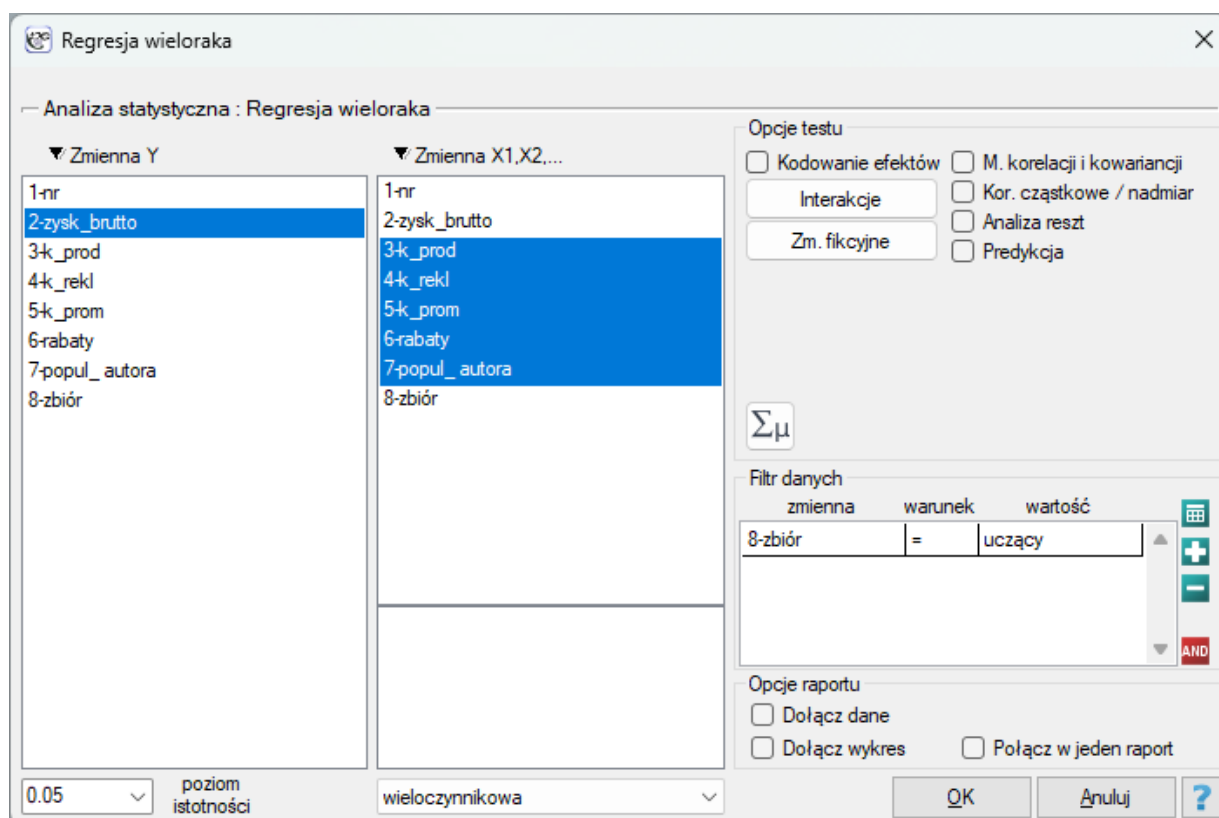
Interakcje rozważane są w modelach wielowymiarowych a ich występowanie oznacza, że wpływ zmiennej niezależnej (X_1) na zmienną zależną (Y) jest inny, w zależności od poziomu kolejnej zmiennej niezależnej (X_2) lub szeregu kolejnych zmiennych niezależnych. By można było rozważyć interakcje w modelach 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.

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 $X_1 * X_2 * X_3$.

23.2 LINIOWA REGRESJA WIELORAKA

Okno z ustawieniami opcji Regresji wielorakiej wywołujemy poprzez menu Statystyka zaawansowana → Modele wielowymiarowe → Regresja wieloraka



Budowany model regresji wielorakiej pozwala na zbadanie wpływu wielu zmiennych niezależnych (X_1, X_2, \dots, 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 ona występowanie liniowego związku pomiędzy badanymi zmiennymi. Liniowy model regresji wielorakiej przyjmuje postać:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon.$$

gdzie:

Y - zmienna zależna, objaśniana przez model,

X_1, X_2, \dots, X_k - zmienne niezależne, objaśniające,

$\beta_0, \beta_1, \beta_2, \dots, \beta_k$ - parametry,

ϵ - składnik losowy (reszta modelu).

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. Wielkość tego błędu wyliczana jest ze wzoru:

$$SE_b = \sqrt{\frac{1}{n - (k + 1)} e^T e (X^T X)^{-1}},$$

gdzie:

$e = Y - \hat{Y}$ to wektor **reszt modelu** (różnica pomiędzy rzeczywistymi wartościami zmiennej zależnej Y a wartościami \hat{Y} przewidywanymi na podstawie modelu).

Zmienne fikcyjne i interakcje w modelu

Omówienie przygotowania zmiennych fikcyjnych i interakcji przedstawiono w rozdziale [23.1 Przygotowanie zmiennych do analizy w modelach wielowymiarowych](#).

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 [71].

23.2.1 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.

Hipotezy:

$$\mathcal{H}_0 : \beta_i = 0,$$

$$\mathcal{H}_1 : \beta_i \neq 0.$$

Wyliczmy statystykę testową według wzoru:

$$t = \frac{b_i}{SE_{b_i}}$$

Statystyka testowa ma **rozkład t-Studenta** z $n - k$ stopniami swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

- **Jakość zbudowanego modelu** liniowej regresji wielorakiej możemy ocenić kilkoma miarami.

- **Błąd standardowy estymacji** – jest miarą dopasowania modelu:

$$SE_e = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - (k + 1)}}.$$

Miara ta opiera się na resztach modelu $e_i = y_i - \hat{y}_i$, czyli rozbieżności pomiędzy rzeczywistymi wartościami zmiennej zależnej y_i w próbie a wartościami zmiennej zależnej \hat{y}_i wyliczonej na podstawie zbudowanego modelu. Najlepiej byłoby, gdyby różnica ta była jak najbliższa zeru dla wszystkich badanych obiektów próby. Zatem, aby model był dobrze dopasowany, błąd standardowy estymacji (SE_e) wyrażony jako wariancja e_i , 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, \dots, 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. 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,

R_{SS} – resztowa suma kwadratów.

Współczynnik determinacji wyliczamy z wzoru:

$$R^2 = \frac{E_{SS}}{T_{SS}}.$$

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 licznosc próby, bywają sytuacje, w których może być obciążona pewnym błędem. Dalego też wyznacza się poprawianą wartość tego parametru:

$$R_{adj}^2 = R^2 - \frac{k(1 - R^2)}{n - (k + 1)}.$$

- **Kryteria informacyjne** opierają się na entropii informacji niesionej przez model (niepewności modelu) tzn. szacują utraconą informację, gdy dany model jest używany do opisu badanego zjawiska. Powinniśmy zatem wybierać model o minimalnej wartości danego kryterium informacyjnego.

AIC , $AICc$ i BIC jest rodzajem kompromisu pomiędzy dobrocią dopasowania i złożonością. Drugi element sumy we wzorach na kryteria informacyjne (tzw. funkcja straty lub kary) mierzy prostotę modelu. Zależy on od liczby zmiennych w modelu (k) i licznosci próby (n). W obu przypadkach element ten rośnie wraz ze wzrostem liczby zmiennych i wzrost ten jest tym szybszy im mniejsza jest liczba obserwacji.

Kryterium informacyjne nie jest jednak miarą absolutną, tzn. jeśli wszystkie porównywane modele źle opisują rzeczywistość w kryterium informacyjnym nie ma sensu szukać ostrzeżenia.

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

Założmy, że wyznaczyliśmy AIC dla trzech modeli $AIC_1=100$, $AIC_2=101.4$, $AIC_3=110$. Wówczas można wyznaczyć względną wiarygodność dla modelu. Wiarygodność ta jest względna, gdyż wyznaczana jest względem innego modelu, najczęściej tego o najmniejszej wartości AIC. Wyznaczamy ją wg 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 (a dokładnie $\exp((100 - 101.4)/2) = 0.497$). Porównując model 3 do modelu pierwszego powiemy, że prawdopodobieństwo, iż zminimalizuje on utratę informacji stanowi niewielką część prawdopodobieństwa, że zrobi to model 1 (a dokładnie $\exp((100 - 110)/2) = 0.007$).

Poprawione kryterium informacyjne Akaikiego

$$AIC_c = AIC + \frac{2(k+3)(k+4)}{n-k}$$

Poprawka kryterium Akaikiego dotyczy wielkości próby, przez co jest to miara rekomendowana również dla prób o małych liczebnościach. **Bayesowskie kryterium informacyjne Schwartza** (*ang. Bayes Information Criterion* lub *Schwarz criterion*)

$$BIC = n \cdot \ln \frac{RSS}{n} + (k+1) \ln n + (constant)$$

gdzie, stałą można pominąć, ponieważ jest taka sama w każdym z porównywanych modeli. Podobnie jak poprawione kryterium Akaikiego 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 odchyłać (co do wartości bezwzględnej) od prognoz.

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i|$$

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.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right|$$

– Istotność statystyczna wszystkich zmiennych w modelu

Podstawowym narzędziem szacującym istotność wszystkich zmiennych w modelu jest test analizy wariancji (test F). Test ten weryfikuje jednocześnie 3 równoważne hipotezy:

$$\begin{aligned} \mathcal{H}_0 : & \text{wszystkie } \beta_i = 0, & \mathcal{H}_1 : & \text{istnieje } \beta_i \neq 0; \\ \mathcal{H}_0 : & R^2 = 0, & \mathcal{H}_1 : & R^2 \neq 0; \\ \mathcal{H}_0 : & \text{liniowość związku,} & \mathcal{H}_1 : & \text{brak związku liniowego.} \end{aligned}$$

Statystyka testowa ma postać:

$$F = \frac{E_{MS}}{R_{MS}}$$

gdzie:

$$E_{MS} = \frac{E_{SS}}{df_E} - \text{średnia kwadratów wyjaśniona przez model,}$$

$$R_{MS} = \frac{R_{SS}}{df_R} - \text{resztowa średnia kwadratów,}$$

$$df_E = k, df_R = n - (k + 1) - \text{odpowiednie stopnie swobody.}$$

Statystyka ta podlega [rozkładowi F Snedecora](#) z df_E i df_R stopniami swobody.

Wyznaczoną na podstawie [statystyki testowej](#) [wartość p](#) porównujemy z α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha & \implies \text{odrzucaamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha & \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

23.2.2 Więcej informacji o zmiennych w modelu

- **Standaryzowane** b_1, b_2, \dots, 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](#) $r_p \in < -1; 1 >$. Współczynnikiem tym badamy korelację dla każdej pary zmiennych, nie uwzględniając 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 $< -1; 1 >$ 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 $< 0; 1 >$ i oznacza stosunek wyłączonej 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 $< -1; 1 >$ 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 skorelowanie z pozostałymi zmiennymi w modelu). Kwadrat tego współczynnika to **współczynnik determinacji semicząstkowej** – należy do przedziału $< 0; 1 >$ i oznacza stosunek wyłączonej zmienności danej zmiennej niezależnej X_i do całkowitej zmienności zmiennej zależnej Y . Im wartość tych współczynników znajduje się bliżej zera, tym bardziej bezużyteczną informację niesie badana zmienna, czyli jest ona nadmiarowa.

- **R-kwadrat** ($R^2 \in < 0; 1 >$) - wyraża on 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** ($VIF \in < 1; \infty$) - określa jak bardzo wariancja szacowanego współczynnika regresji jest zwiększona z powodu współliniowości. Im bliżej wartości 1, tym mniejsza współliniowość i tym mniejszy jej wpływ na wariancję współczynnika. Przyjmuje się, że silna współliniowość występuje, gdy współczynnik $VIF > 5$ [149]. Jeśli współczynnik inflacji wariancji wynosi 5 ($\sqrt{5} = 2.2$), oznacza to, że błąd standardowy dla współczynnika tej zmiennej jest 2.2 razy większy niż w przypadku, gdyby ta zmienna miała zerową korelację z innymi zmiennymi X_i .
- **Tolerancja** = $1 - R^2 \in < 0; 1 >$ - wyraża on 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**
Porównanie tych dwóch modeli dokonujemy:
 - **testem F**, w sytuacji gdy z modelu usuwamy jedną zmienną lub więcej niż jedną zmienną (patrz porównywanie modeli),
 - **testem t-Studenta**, gdy z modelu usuwamy tylko jedną zmienną. Jest to ten sam test, którym badamy istotność poszczególnych zmiennych w modelu.

W przypadku usunięcia tylko jednej zmiennej wyniki obu tych testów są tożsame.

Jeśli różnica pomiędzy porównywanymi modelami jest istotna statystycznie (wartość $p \leq \alpha$), wówczas model pełny jest istotnie lepszy niż model zredukowany. To oznacza, że badana zmienna nie jest nadmiarowa, wywiera ona istotny wpływ na dany model i nie powinna być z niego usuwana.

- **Wykresy rozrzutu**
Wykresy te pozwalają dokonać subiektywnej oceny liniowości związku pomiędzy zmiennymi i zidentyfikować punkty odstające. Dodatkowo wykresami rozrzutu możemy posłużyć się w analizie reszt modelu.

23.2.3 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, które przekraczają 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. Jeśli wśród przypadków oddalonych o więcej niż 3 odchylenia znajdzie się przypadek o największej wartości Mahalanobisa, wówczas jako najbardziej odstający zostanie on zaznaczony pogrubioną czcionką.

- **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:

- reszty względem wartości przewidywanych
- kwadrat reszty względem wartości przewidywanych
- reszty względem wartości obserwowanych
- kwadrat reszty względem wartości obserwowanych

- **Autokorelacja reszt modelu**

Aby zbudowany model można było uznać za poprawny, wartości reszt nie powinny być ze sobą skorelowane (dla wszystkich par e_i, e_j). Założenie to możemy sprawdzić wyliczając statystykę testu Durбина-Watsona

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2},$$

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;
- Jeżeli $d > d_{U,\alpha}$ – błędy nie są dodatnio skorelowane;
- Jeżeli $d_{L,\alpha} < d < d_{U,\alpha}$ – wynik testu jest niejednoznaczny.

Aby sprawdzić ujemną autokorelację na poziomie istotności α , sprawdzamy położenie wartości $4 - d$ w stosunku do górnej ($d_{U,\alpha}$) i dolnej ($d_{L,\alpha}$) wartości krytycznej:

- Jeżeli $4 - d < d_{L,\alpha}$ – błędy są ujemnie skorelowane;
- Jeżeli $4 - d > d_{U,\alpha}$ – błędy nie są ujemnie skorelowane;
- Jeżeli $d_{L,\alpha} < 4 - d < d_{U,\alpha}$ – wynik testu jest niejednoznaczny.

Wartości krytyczne testu Durбина-Watsona dla poziomu istotności $\alpha = 0.05$ znajdują się na stronie internetowej www.pqstat.pl – źródło tablic: Savina i White (1977)[145]

23.2.4 Przykład dla regresji wielorakiej

PRZYKŁAD 23.2. (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:

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.

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 licznosci próby $n \geq 50+8k$	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

Na podstawie oszacowanej wartości współczynnika b , związek pomiędzy zyskiem brutto a wszystkimi zmiennymi niezależnymi możemy opisać równaniem:

$$zysk_{brutto} = 4.18 + 2.56(k_{prod}) + 2(k_{rekl}) + 4.67(k_{prom}) + 1.42(rabaty) + 10.15(popul_{autora}) + [8.09]$$

Uzyskane współczynniki interpretujemy następująco:

- Jeśli koszt produkcji wzrośnie o 1 tysiąc dolarów, to zysk brutto wzrośnie o około 2.56 tysiące 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 tysiące dolarów, przy założeniu, że pozostałe zmienne się nie zmieniają;
- Jeśli suma udzielonych rabatów wzrośnie o 1 tysiąc dolarów, to zysk brutto wzrośnie o około 1.42 tysiące dolarów, przy założeniu, że pozostałe zmienne się nie zmieniają;
- 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 w modelu popularność autora przyjmujemy jako wartość 0 i 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 wywiera istotny wpływ na otrzymany zysk. Jednocześnie, dla tych zmiennych standaryzowane współczynniki b są największe.

Dodatkowo, model jest dobrze dopasowany o czym świadczy: mały błąd standardowy estymacji $SE_e = 8.086501$, wysoka wartość współczynnika determinacji wielorakiej $R^2 = 0.850974$ i poprawionego współczynnika determinacji wielorakiej $R^2_{adj} = 0.829059$ oraz wynik testu F analizy wariancji: $p < 0.000001$.

Na podstawie interpretacji dotychczasowych wyników możemy przypuszczać, że część zmiennych nie wywiera istotnego wpływu na zysk i może być zbędna.

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

Najbardziej spójną informację, pozwalającą znaleźć te zmienne w modelu, które są zbędne (nadmiarowe) niesie analiza korelacji cząstkowej i semicząstkowej i nadmiarowości:

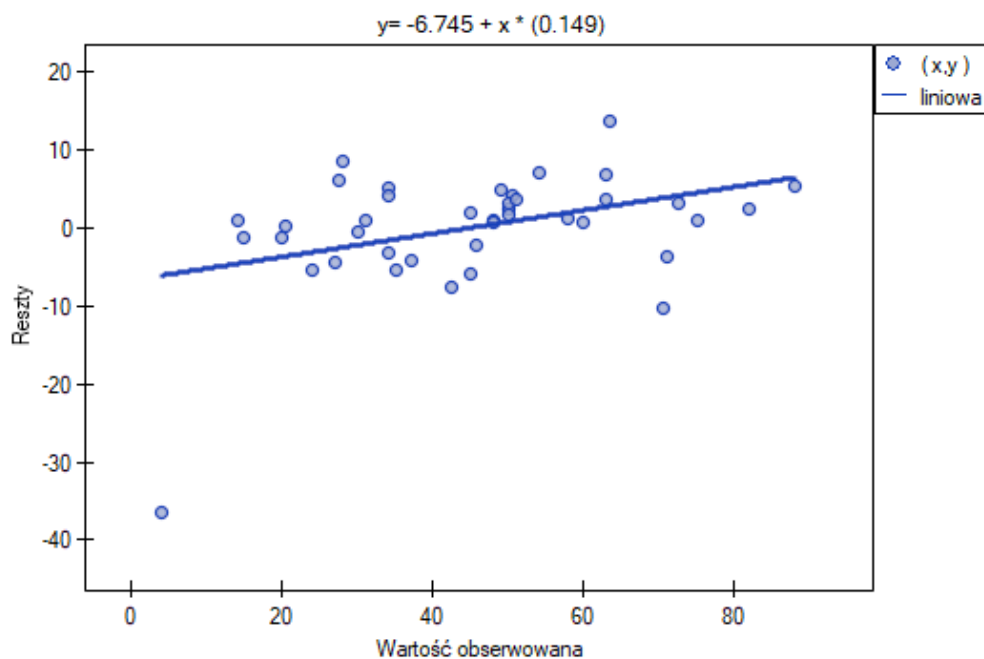
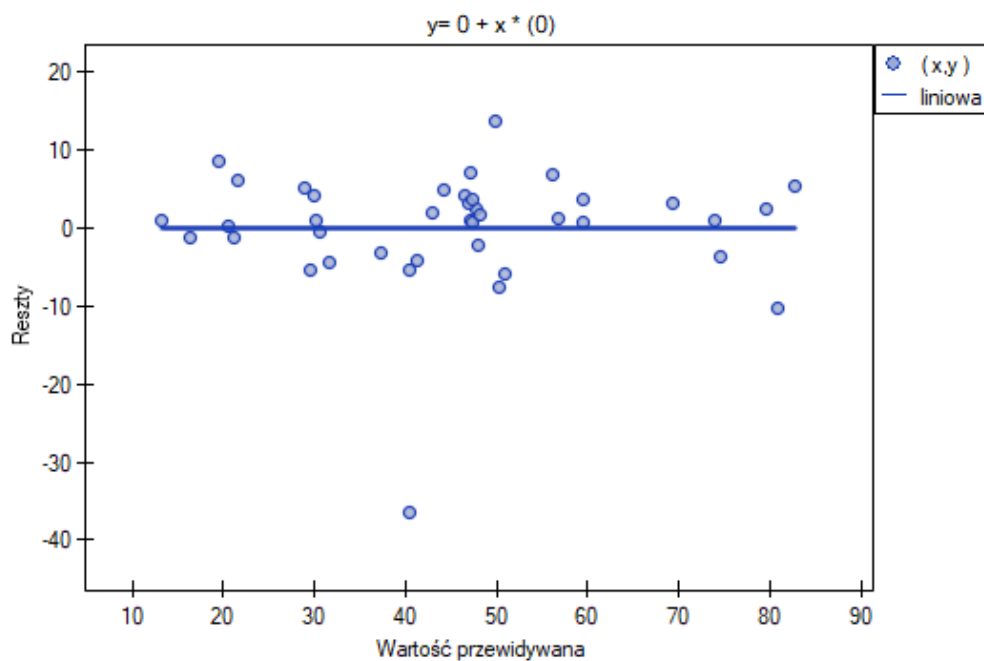
Kor. cząst. semi.						
	cząstkowe	semicząstk	tolerancja	R ²	Stat. t	Wartość p
k_prod	0.658793	-0.338045	0.639209	0.360791	5.106033	0.000013
k_rekl	0.690424	-0.368438	0.637949	0.362051	5.565106	0.000003
k_prom	0.164797	-0.0645	0.948099	0.051901	0.974245	0.336816
rabaty	0.171176	-0.06707	0.987951	0.012049	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.

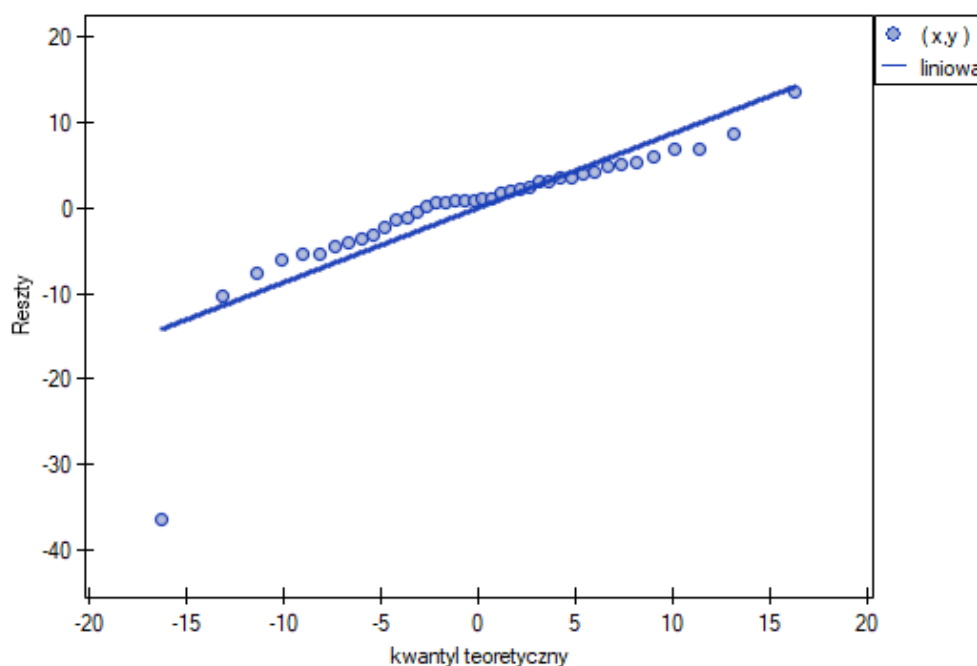
Na koniec przeprowadzimy analizę reszt. Fragment tej analizy znajduje się poniżej:

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	-36.467863	-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 resz względem obserwowanych lub przewidywanych wartości zmiennej Y .



Ten odstający punkt zaburza założenie dotyczące homoskedastyczności. Założenie homoskedastyczności było by 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 Lilieforsa wynosi $p = 0.016415$):



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 = $(8+6+0.33+1.6) = 15.93$ tysiące 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:

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 licznosci próby $n \geq 50+8k$	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 Schwarzera	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 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.

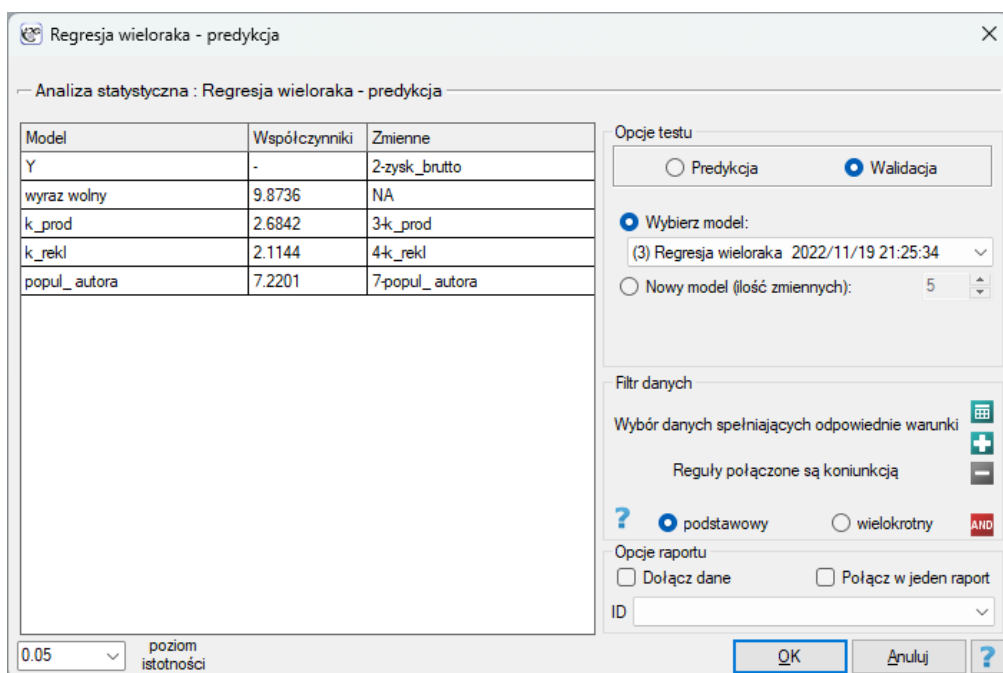
Na koniec należy jeszcze zauważyć, że jest to tylko model wstępny. W badaniu właściwym należałoby zebrać więcej danych. Liczba zmiennych w modelu jest bowiem zbyt mała w stosunku do liczby ocenianych książek tzn. $n < 50 + 8k$.

23.2.5 Predykcja na podstawie modelu i walidacja zbioru testowego

Walidacja

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 zostanie modelu do danych, które nie były wykorzystywane w estymacji modelu. Jeśli podsumowanie w oparciu o dane uczące będzie satysfakcjonujące tzn. wyznaczone 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.

Okno z ustawieniami opcji walidacji wywołujemy poprzez menu Statystyki zaawansowane → Modele wielowymiarowe → Regresja wieloraka - predykcja/walidacja.



By dokonać walidacji należy wskazać model, na podstawie którego chcemy jej dokonać. Walidacji możemy dokonać 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.

Predykcja

Najczęściej ostatnim etapem analizy regresji jest wykorzystanie zbudowanego i uprzednio zweryfikowanego modelu do predykcji.

- Predykcja dla jednego obiektu może być wykonywana wraz z budową modelu, czyli w oknie analizy Statystyki zaawansowane → Modele wielowymiarowe → Regresja wieloraka ,
- Predykcja dla większej grupy nowych danych jest wykonywana poprzez menu Statystyki zaawansowane → Modele wielowymiarowe → Regresja wieloraka - predykcja/walidacja.

By dokonać predykcji należy wskazać model, na podstawie którego chcemy jej dokonać. Predykcji możemy dokonać 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 predykcji.

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.

PRZYKŁAD 23.2 c.d. (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 pozycji książkowych. By korzystać z odpowiednich zbiorów ustawiamy każdorazowo filtr danych.

Regresja wieloraka - predykcja/walidacja		Współczynniki modelu	
Analizowane zmienne	k_prod	wyraz wolny	9.8735582
	k_rekl	k_prod	2.6842296
	popul_ autora	k_rekl	2.1143989
	zysk_brutto	popul_ autora	7.2201236
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 Schwarzera	142.0086223		
MAE (średni błąd bezwzględny)	3.6739289		
MPE (średni błąd procentowy)	-0.0171597		
MAPE (średni bezwzględny błąd procentowy)	0.0941638		

Dla zbioru uczącego wartości opisujące jakość dopasowania modelu są bardzo wysokie: skorygowane $R^2 = 0.93$ a średni błąd prognozy (MAE) wynosi 3.8 tys. dolarów.

Regresja wieloraka - predykcja/walidacja		Współczynniki modelu	
Analizowane zmienne	k_prod	wyraz wolny	9.8735582
	k_rekl	k_prod	2.6842296
	popul_ autora	k_rekl	2.1143989
	zysk_brutto	popul_ autora	7.2201236
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 Schwarzera	96.9979976		
MAE (średni błąd bezwzględny)	5.8739802		
MPE (średni błąd procentowy)	0.0088586		
MAPE (średni bezwzględny błąd procentowy)	0.1354043		

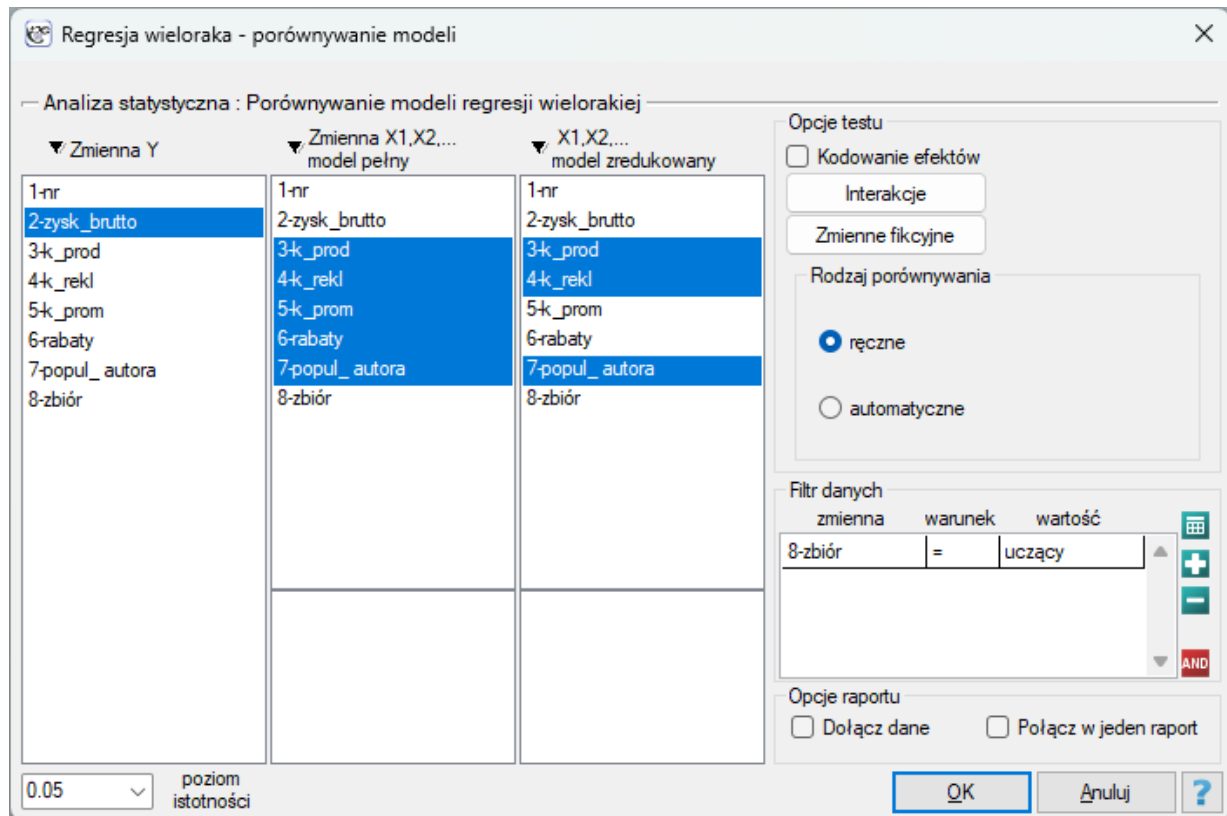
Dla zbioru testowego wartości opisujące jakość dopasowania modelu są nieco niższe niż dla zbioru uczącego: Skorygowane $R^2 = 0.80$ a średni błąd prognozy (MAE) wynosi 5.9 tys. dolarów. Ponieważ wynik walidacji na zbiorze testowym jest prawie tak dobry jak na zbiorze uczącym, użyjemy modelu do predykcji. W tym celu skorzystamy z danych trzech nowych pozycji książkowych dopisanych na końcu zbioru. Wybierzemy opcję Predykcja, ustawiany filtr na nowy zbiór danych i użyjemy naszego modelu do tego by przewidzieć zysk brutto dla tych książek. dorobić

Predykcja				
Pred. Y	-95% CI (pu)	+95% CI (pi)	-95% CI (wz)	+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.

23.3 PORÓWNYWANIE MODELI LINIOWEJ REGRESJI WIELORAKIEJ

Okno z ustawieniami opcji porównywania modeli wywołujemy poprzez menu Statystyka → Modele wielowymiarowe → Regresja wieloraka – porównywanie modeli



Liniiowa 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 może pominąć te cechy, które w wiarygodny sposób mogłyby opisać badane zjawisko. Bowiem 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.

Nie istnieje jedna prosta reguła statystyczna, która decydowałaby o liczba zmiennych niezbędnych w modelu. Najczęściej w porównaniu posługujemy się miarami dopasowania modelu takimi jak: R_{adj}^2 – 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 . Test ten służy do weryfikacji hipotezy, że dopasowanie obu porównywanych modeli jest tak samo dobre.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: R_{FM}^2 = R_{RM}^2, \\ \mathcal{H}_1 &: R_{FM}^2 \neq R_{RM}^2, \end{aligned}$$

gdzie:

R_{FM}^2, R_{RM}^2 – współczynniki determinacji wielorakiej w porównywanych modelach (pełnym i zredukowanym).

Statystyka testowa ma postać:

$$F = \frac{R_{FM}^2 - R_{RM}^2}{k_{FM} - k_{RM}} \cdot \frac{n - k_{FM} - 1}{1 - R_{FM}^2},$$

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 :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Jeśli porównywane modele nie różnią się istotnie, to 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.

W programie PQStat porównywanie modeli możemy 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.

- **Automatyczne** porównywanie modeli jest wykonywane w kilku krokach:

- krok 1 Zbudowanie modelu z wszystkich zmiennych.
- krok 2 Usunięcie jednej zmiennej z modelu. Usuwana zmienna to ta, która ze statystycznego punktu widzenia wnosi do aktualnego modelu najmniej informacji.
- krok 3 Porównanie modelu pełnego i zredukowanego.
- krok 4 Usunięcie kolejnej zmiennej z modelu. Usuwana zmienna to ta, która ze statystycznego punktu widzenia wnosi do aktualnego modelu najmniej informacji.
- krok 5 Porównanie modelu wcześniejszego i nowo zredukowanego.

...

W ten sposób powstaje wiele, coraz mniejszych modeli. Ostatni model zawiera tylko 1 zmienną niezależną.

W rezultacie, każdy model jest opisany miarami dopasowania (R_{adj}^2 , SE_e , AIC, AICc, BIC), a kolejno powstające (sąsiednie) modele są porównywane testem F. Model, który zostanie ostatecznie zaznaczony jako statystycznie optymalny, to model o największym R_{adj}^2 , najmniejszym kryterium informacyjnym i najmniejszym SE_e . Ponieważ jednak żadna z metod statystycznych nie potrafi w pełni odpowiedzieć na pytanie który z modeli jest najlepszy, to badacz, na podstawie uzyskanych wyników, powinien wybrać model.

PRZYKŁAD (23.2) c.d. (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 optymalną (ze statystycznego punktu widzenia) liczbę zmiennych. Do tej analizy użyjemy danych zbioru uczącego.

- **Ręczne** porównywanie modeli.

Na podstawie zbudowanego wcześniej modelu pełnego możemy podejrzewać, że zmienne: koszty promocji bezpośredniej, suma udzielonych rabatów mają niewielki wpływ na budowany model (tzn. te zmienne nie pomagają przewidzieć wielkości zysku). Sprawdźmy 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 Akaikiego	172.7149
AICc - skorygowane kryterium Akaikiego	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 Akaikiego	170.863
AICc - skorygowane kryterium Akaikiego	173.263
BIC - bayesowskie kryterium Schwarza	184.9963
F - porównanie modeli	0.9379
DF1	2
DF2	34
Wartość p	0.4013

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

Okazuje się, że nie ma podstaw by uważać, że model pełny jest lepszy niż model zredukowany (wartość p testu F służącego porównywaniu modeli wynosi $p = 0.4013$). Dodatkowo, model zredukowany jest nieco lepiej dopasowany niż model pełny (dla modelu zredukowanego $R_{adj}^2 = 0.8296$, dla modelu pełnego $R_{adj}^2 = 0.8291$) oraz ma mniejsze, czyli korzystniejsze wartości kryteriów informacyjnych AIC, AICc i BIC.

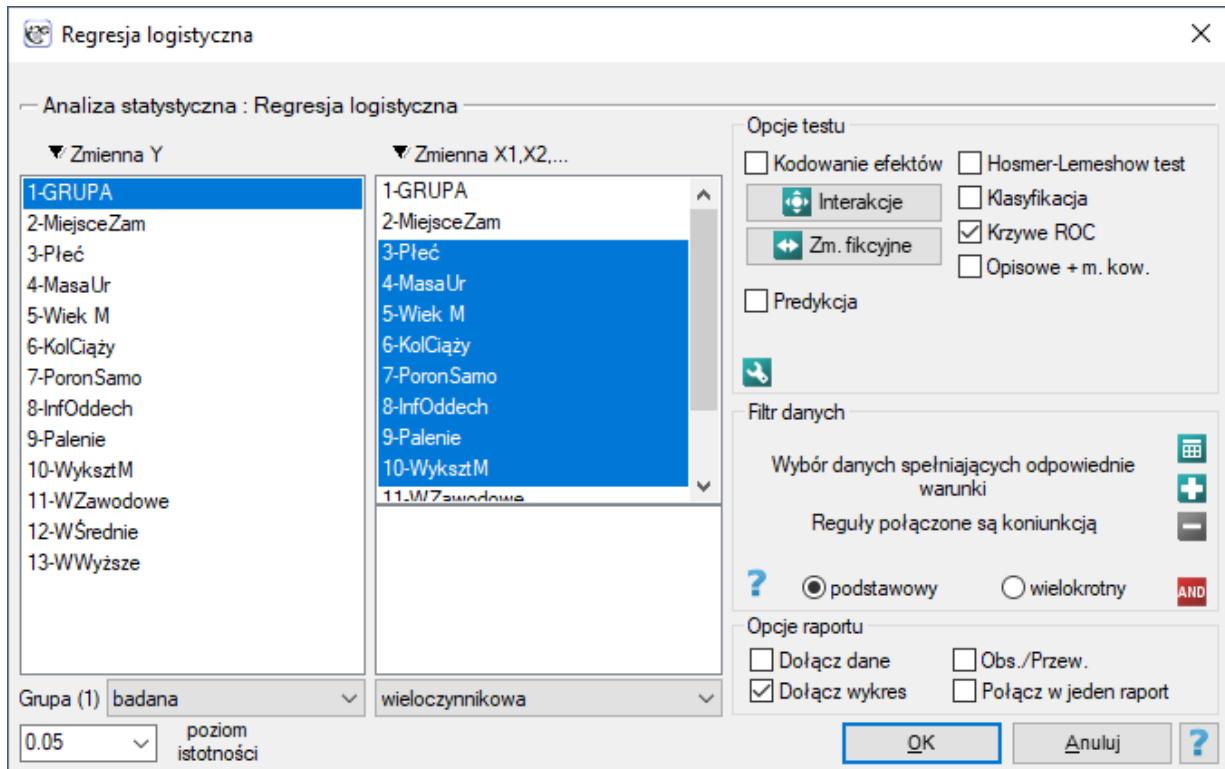
- **Automatyczne** porównywanie modeli.

W przypadku automatycznego porównywania modeli uzyskaliśmy bardzo podobne wyniki. Najlepszym modelem jest model o największym współczynniku R_{adj}^2 , 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.

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 posiadającą 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.

23.4 REGRESJA LOGISTYCZNA

Okno z ustawieniami opcji Regresji logistycznej wywołujemy poprzez menu Statystyka zaawansowana → Modele wielowymiarowe → Regresja logistyczna



Budowany model regresji logistycznej (podobnie jak liniowej regresji wielorakiej) pozwala na zbadanie wpływu wielu zmiennych niezależnych (X_1, X_2, \dots, X_k) na jedną zmienną zależną (Y). Tym razem jednak zmienna zależna przyjmuje jedynie dwie wartości, np. chory/zdrowy, niewyptalalny/wyptalalny itp.

Owe dwie wartości kodowane są jako (1)/(0) gdzie:

- (1) –wartość wyróżniona –posiadanie danej cechy
- (0) –brak danej cechy.

Funkcja, na której oparty jest model regresji logistycznej wylicza nie dwupoziomową zmienną Y , a prawdopodobieństwo przyjęcia przez tą zmienną wyróżnionej wartości:

$$P(Y = 1|X_1, X_2, \dots, X_k) = \frac{e^Z}{1 + e^Z},$$

gdzie:

$P(Y = 1|X_1, X_2, \dots, X_k)$ –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, \dots, X_k –zmiennie niezależne, objaśniające,
 $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ –parametry.

Zmienne fikcyjne i interakcje w modelu

Omówienie przygotowania zmiennych fikcyjnych i interakcji przedstawiono w rozdziale 23.1 [Przygotowanie zmiennych do analizy w modelach wielowymiarowych](#).

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 .

Logitem nazywamy przekształcenie tego modelu do postaci:

$$\ln\left(\frac{P}{1-P}\right) = Z.$$

Macierze biorące udział w równaniu, dla próby o licznosci n , zapisujemy następująco:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix},$$

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 **metodę największej wiarygodności** czyli poprzez poszukiwanie maksimum funkcji wiarygodności L (w programie użyto algorytm iteracyjny 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. Wielkość tego błędu wyliczana jest ze wzoru:

$$SE_b = \sqrt{\text{diag}(H^{-1})_b}$$

gdzie:

$\text{diag}(H^{-1})$ to główna przekątna macierzy kowariancji.

Uwaga!

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 [131] mówiące, iż liczba obserwacji powinna być dziesięciokrotnie większa lub równa stosunkowi liczby zmiennych niezależnych (v) i mniejszej z proporcji licznosci (p) opisanych z zmiennej zależnej (tzn. proporcji chorych lub zdrowych), czyli $n \geq 10v/p$.

Uwaga!

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.

Uwaga!

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).

23.4.1 Iloraz Szans

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).

Iloraz Szans - wzór ogólny

Program PQStat wylicza jednostkowy iloraz Szans. Jego modyfikacja, na podstawie ogólnego wzoru, umożliwi zmianę interpretacji uzyskanego wyniku.

Iloraz szans na wystąpienie stanu wyróżnionego w ogólnym przypadku jest wyliczany jako iloraz dwóch szans. Zatem dla zmiennej niezależnej X_1 dla Z wyrażonego zależnością liniową wyliczamy:

szansę dla kategorii pierwszej:

$$Szansa(1) = \frac{P(1)}{1 - P(1)} = e^Z(1) = e^{\beta_0 + \beta_1 X_1(1) + \beta_2 X_2 + \dots + \beta_k X_k},$$

szansę dla kategorii drugiej:

$$Szansa(2) = \frac{P(2)}{1 - P(2)} = e^Z(2) = e^{\beta_0 + \beta_1 X_1(2) + \beta_2 X_2 + \dots + \beta_k X_k}.$$

Iloraz Szans dla zmiennej X_1 wyraża się wówczas wzorem:

$$\begin{aligned}
 OR_1(2)/(1) &= \frac{Szansa(2)}{Szansa(1)} = \frac{e^{\beta_0 + \beta_1 X_1(2) + \beta_2 X_2 + \dots + \beta_k X_k}}{e^{\beta_0 + \beta_1 X_1(1) + \beta_2 X_2 + \dots + \beta_k X_k}} \\
 &= e^{\beta_0 + \beta_1 X_1(2) + \beta_2 X_2 + \dots + \beta_k X_k - [\beta_0 + \beta_1 X_1(1) + \beta_2 X_2 + \dots + \beta_k X_k]} \\
 &= e^{\beta_1 X_1(2) - \beta_1 X_1(1)} = e^{\beta_1 [X_1(2) - X_1(1)]} = \\
 &= (e^{\beta_1})^{[X_1(2) - X_1(1)]}.
 \end{aligned}$$

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:

$$OR = (e^{\beta_1})^1,$$

który mówi o ile zmieni się szansa na wystąpienie wyróżnionej wartości gdy wiek zmieni się o 1 rok.

Iloraz szans wyliczony dla niesąsiadujących kategorii zmiennej wiek np. 25 lat i 30 lat będzie pięcioletnim Ilorazem Szans, ponieważ różnica $X_1(2) - X_1(1) = 30 - 25 = 5$. Wówczas otrzymamy pięcioletni Iloraz Szans:

$$OR = (e^{\beta_1})^5,$$

który mówi o ile zmieni się szansa na wystąpienie wyróżnionej wartości gdy wiek zmieni się o 5 lat.

Uwaga!

Jeśli analizę przeprowadzamy dla modelu innego niż liniowy, lub uwzględniamy interakcję, wówczas na podstawie ogólnego wzoru możemy wyliczyć odpowiedni Ilorazu Szans zmieniając formułę wyrażającą Z .

23.4.2 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.

Hipotezy:

$$\begin{aligned}
 \mathcal{H}_0 : \beta_i = 0, & \quad \text{lub równoważnie:} \quad \mathcal{H}_0 : OR_i = 1, \\
 \mathcal{H}_1 : \beta_i \neq 0. & \quad \mathcal{H}_1 : OR_i \neq 1.
 \end{aligned}$$

Statystykę testową testu Walda wyliczamy według wzoru:

$$\chi^2 = \left(\frac{b_i}{SE_{b_i}} \right)^2$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) rozkład χ^2 z 1 stopniem swobody. Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned}
 \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\
 \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.
 \end{aligned}$$

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, które opierają się 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** opierają się na entropii informacji niesionej przez model (niepewności modelu) tzn. szacują utraconą informację, gdy dany model jest używany do opisu badanego zjawiska. Powinniśmy zatem wybierać model o minimalnej wartości danego kryterium informacyjnego.

AIC , $AICc$ i BIC jest rodzajem kompromisu pomiędzy dobrocią dopasowania i złożonością. Drugi element sumy we wzorach na kryteria informacyjne (tzw. funkcja straty lub kary) mierzy prostotę modelu. Zależy on od liczby zmiennych w modelu (k) i liczności próby (n). W obu przypadkach element ten rośnie wraz ze wzrostem liczby zmiennych i wzrost ten jest tym szybszy im mniejsza jest liczba obserwacji.

Kryterium informacyjne nie jest jednak miarą absolutną, tzn. jeśli wszystkie porównywane modele źle opisują rzeczywistość w kryterium informacyjnym nie ma sensu szukać ostrzeżenia.

- Kryterium informacyjne Akaikego (*ang. Akaike information criterion*)

$$AIC = -2 \ln L_{FM} + 2k,$$

Jest to kryterium asymptotyczne - odpowiednie dla dużych prób.

- Poprawione kryterium informacyjne Akaikego

$$AICc = AIC + \frac{2k(k+1)}{n-k-1},$$

Poprawka kryterium Akaikego dotyczy wielkości próby, przez co jest to miara rekomendowana również dla prób o małych licznosciach.

- Bayesowskie kryterium informacyjne Schwartza (*ang. Bayes Information Criterion* lub *Schwarz criterion*)

$$BIC = -2 \ln L_{FM} + k \ln(n),$$

Podobnie jak poprawione kryterium Akaikego uwzględnia wielkość próby.

- **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 $< 0; 1$), gdzie wartości bliskie 1 oznaczają doskonałe dopasowanie modelu, 0 –zupełny brak dopasowania. Współczynnik R_{Pseudo}^2 wyliczamy z wzoru:

$$R_{Pseudo}^2 = 1 - \frac{\ln L_{FM}}{\ln L_0}.$$

Ponieważ współczynnik R_{Pseudo}^2 nie przyjmuje wartości 1 i jest wrażliwy na liczbę zmiennych w modelu, wyznacza się jego poprawioną wartość:

$$R_{Nagelkerke}^2 = \frac{1 - e^{-(2/n)(\ln L_{FM} - \ln L_0)}}{1 - e^{(2/n) \ln L_0}} \quad \text{lub} \quad R_{Cox-Snell}^2 = 1 - e^{\frac{(-2 \ln L_0) - (-2 \ln L_{FM})}{n}}.$$

- **Istotność statystyczna wszystkich zmiennych w modelu**

Podstawowym narzędziem szacującym istotność wszystkich zmiennych w modelu jest **test ilorazu wiarygodności**. Test ten weryfikuje hipotezę:

$$\begin{aligned}\mathcal{H}_0 &: \text{wszystkie } \beta_i = 0, \\ \mathcal{H}_1 &: \text{istnieje } \beta_i \neq 0.\end{aligned}$$

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 :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

- **Test Hosmera-Lemeshowa** – Test ten, 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, wówczas można założyć, że zbudowano dobrze dopasowany model.

Do obliczeń najpierw obserwacje są dzielone na G podgrup – zwykle na decyle ($G = 10$).

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: O_g = E_g \text{ dla wszystkich kategorii,} \\ \mathcal{H}_1 &: O_g \neq E_g \text{ dla przynajmniej jednej kategorii.}\end{aligned}$$

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 :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

- **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). Powstała w ten sposób krzywa, a w szczególności pole pod nią, obrazuje jakość klasyfikacyjną modelu. Gdy krzywa ROC pokrywa się z przekątną $y = x$, to decyzja o przyporządkowaniu przypadku do wybranej klasy (1) lub (0) podejmowana na podstawie modelu jest tak samo dobra jak losowy podział badanych przypadków do tych grup. Jakość klasyfikacyjna modelu jest dobra, gdy krzywa znajduje się znacznie powyżej przekątnej $y = x$, czyli gdy pole pod krzywą ROC jest znacznie większe niż pole pod prostą $y = x$, zatem większe niż 0.5

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: AUC = 0.5, \\ \mathcal{H}_1 &: AUC \neq 0.5.\end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{AUC - 0.5}{SE_{0.5}},$$

gdzie:

$SE_{0.5}$ – błąd pola.

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_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.

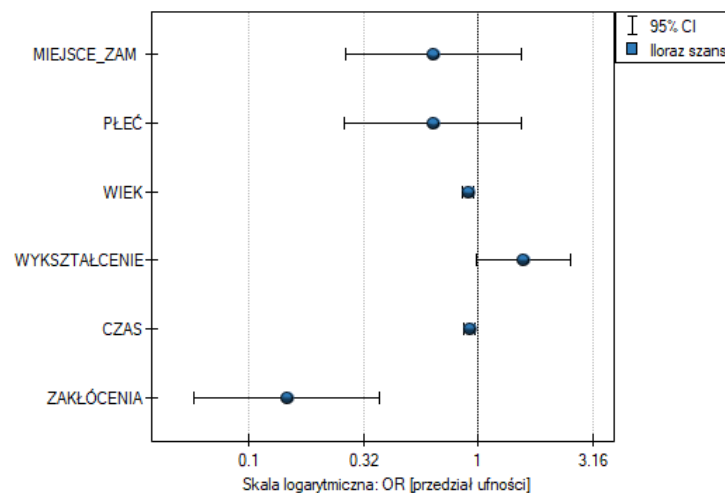
• Klasyfikacja

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.

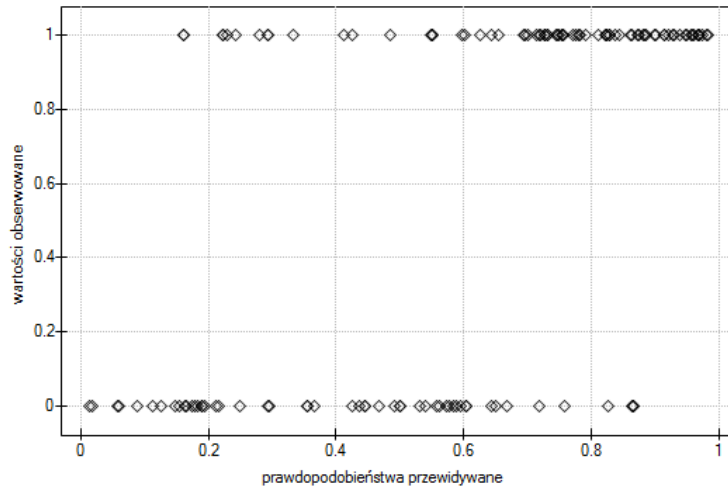
Wykresy w regresji logistycznej

- **Iloraz szans \pm przedział ufności** - to wykres przedstawiający OR wraz z 95% przedziałem ufności dla wyniku każdej zmiennej zwróconej w zbudowanym modelu. Dla zmiennych kategoryalnych, linia na poziomie 1 wskazuje wartość ilorazu szans dla kategorii referencyjnej.

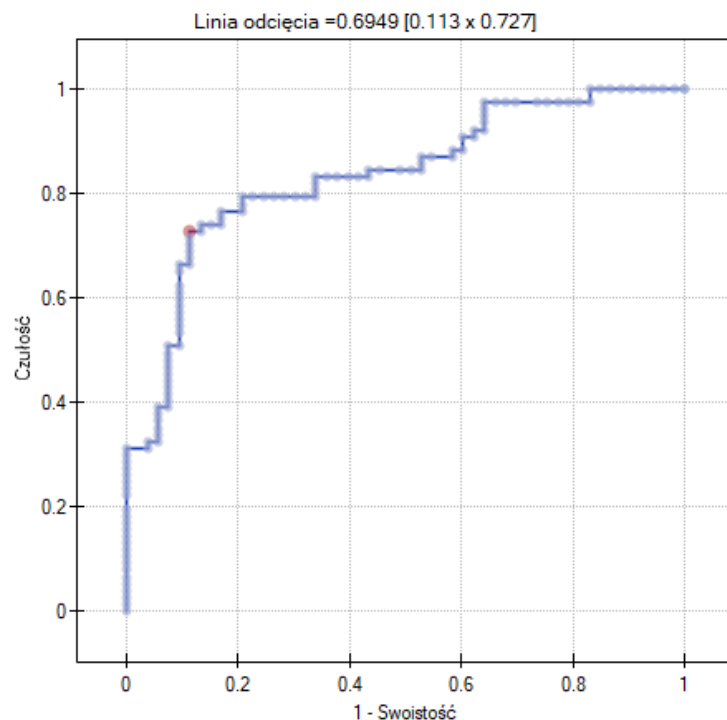


- **Wartości obserwowane / Prawdopodobieństwo oczekiwane** - to wykres przedstawiający wyniki przewidywanego dla każdej osoby prawdopodobieństwa wystąpienia zdarzenia (oś

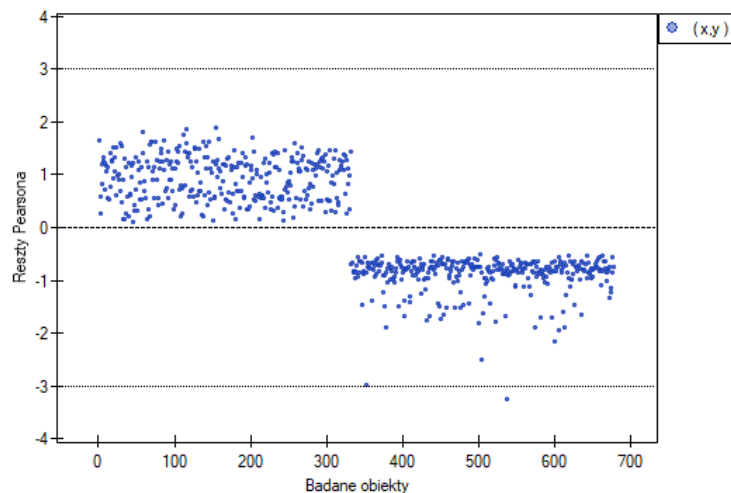
X) oraz wartości prawdziwej, czyli wystąpienia zdarzenia (wartość 1 na osi Y) lub braku zdarzenia (wartość 0 na osi Y). Jeśli model bardzo dobrze prognozuje, to przy lewej stronie wykresu punkty będą się kumulowały w dolnej części, a przy prawej stronie wykresu w górnej części.



- **Krzywa ROC** - to wykres zbudowany w oparciu o wartość zmiennej zależnej oraz przewidywane prawdopodobieństwo wystąpienia zdarzenia.



- **Wykres reszt Pearsona** - to 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. Reszta Pearsona to 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 lub większe niż 3 sugerują, zbyt duże odchylenie danego obiektu.



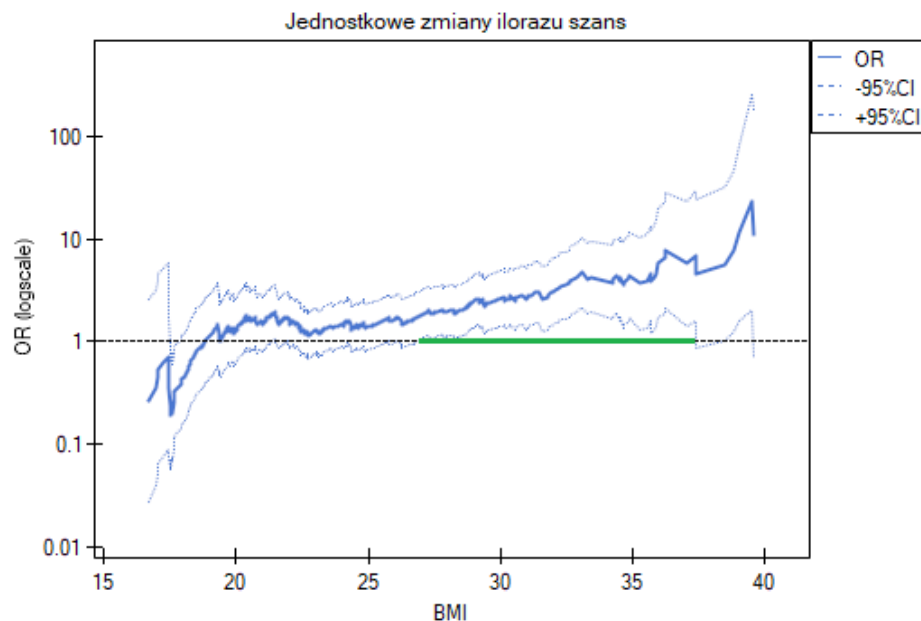
- **Jednostkowe zmiany ilorazu szans** - to wykres przedstawiający serie ilorazów szans wraz z przedziałem ufności, wyznaczone dla każdego z możliwych punktów odcięcia zmiennej umieszczonej na osi X. Umożliwia on użytkownikowi wybór jednego, dobrego punktu odcięcia i następnie zbudowania na tej podstawie nowej dwuwartościowej zmiennej, przy której zostanie osiągnięty odpowiednio wysoki lub niski iloraz szans. Wykres dedykowany jest dla oceny zmiennych ciągłych w analizie jednoczynnikowej, tzn. gdy wybrana jest tylko jedna zmienna niezależna.
- **Profile ilorazu szans** - to wykres przedstawiający serie ilorazów szans wraz z przedziałem ufności, wyznaczone dla wskazanej wielkości okna tzn. porównujące częstości wewnątrz okna z częstościami umieszczonymi na zewnątrz okna. Umożliwia to użytkownikowi wybór kilku kategorii na jakie chce podzielić badaną zmienną i przyjęcie najkorzystniejszej kategorii referencyjnej. 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. Nie ma jednej dobrej dla każdej analizy wielkości okna, wielkość tę należy ustalać indywidualnie dla każdej zmiennej. Wielkość okna wskazuje liczbę niepowtarzalnych wartości zmiennej X znajdujących się w oknie. Im szersze okno, tym większe uogólnienie wyników i gładzsza funkcja ilorazu szans. Im węższe okno, tym bardziej szczegółowe wyniki, przez co bardziej rozchwiany iloraz szans. Do wykresu dodana jest krzywa ukazująca wygładzoną (metodą Lowess) wartość ilorazu szans. Ustawiając bliski 0 współczynnik wygładzania uzyskamy krzywą przylegającą ściśle do wyznaczonego ilorazu szans, natomiast ustawiając współczynnik wygładzania bliżej 1 dostaniemy większe uogólnienie ilorazu szans, a więc bardziej gładką i mniej przyległą do ilorazu szans krzywą. Wykres dedykowany jest dla oceny zmiennych ciągłych w analizie jednoczynnikowej, tzn. gdy wybrana jest tylko jedna zmienna niezależna.

PRZYKŁAD 23.3. (plik Profile OR.pqs)

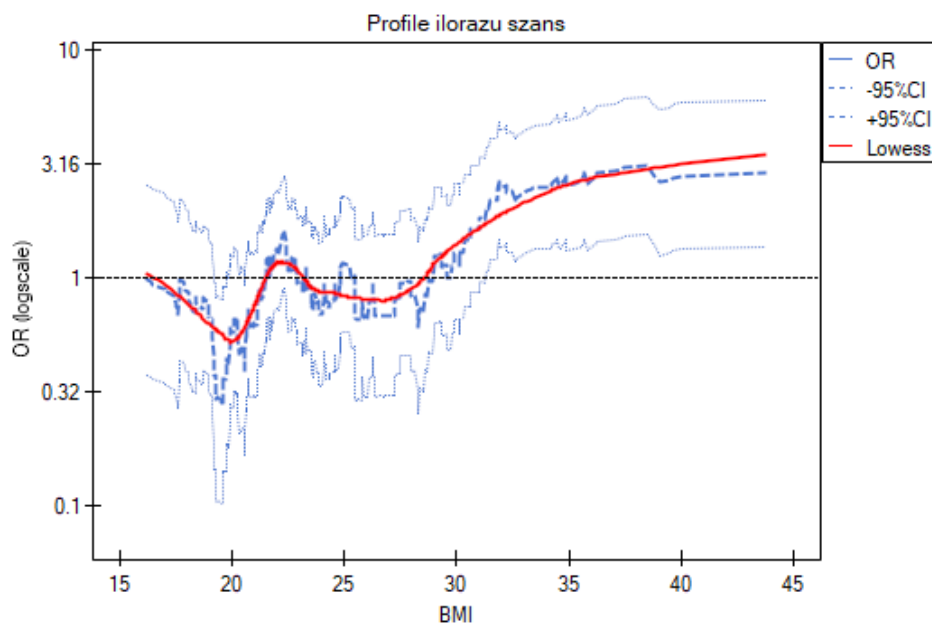
Badamy ryzyko występowania choroby A i choroby B w zależności od BMI pacjenta. Ponieważ BMI jest zmienną ciągłą, to jej umieszczenie w modelu skutkuje wyznaczeniem jednostkowego ilorazu szans wyznaczającego liniowy trend wzrostu lub spadku ryzyka. Nie wiemy czy model liniowy będzie dobrym modelem dla 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 i na tej podstawie zdecydować o sposobie w jaki powinniśmy przygotować zmienną do analizy. Do tego celu posłużą wykresy jednostkowych zmiany 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.



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, tzn. taką która da istotne ilorazy szans.

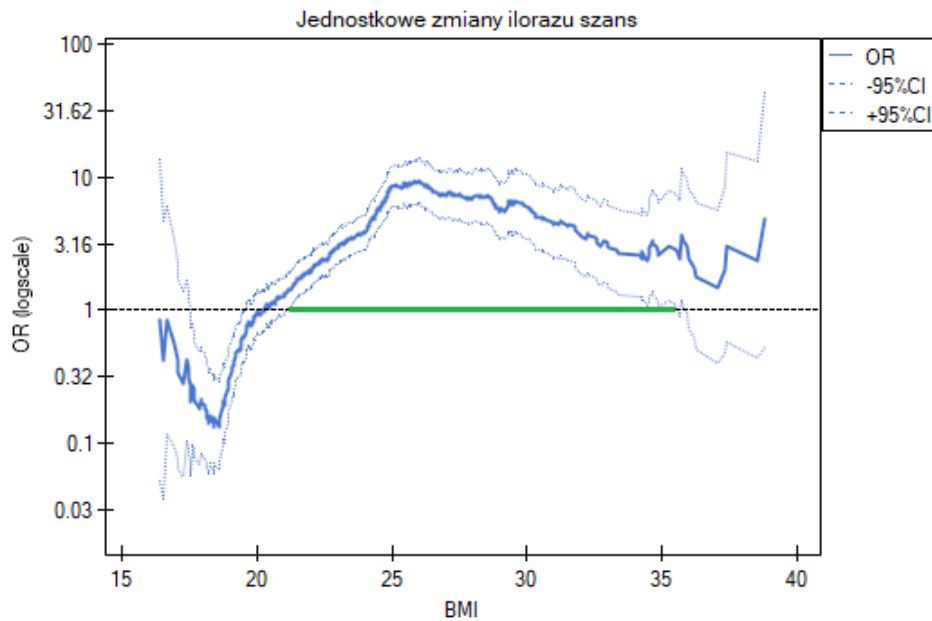


Podsumowując, można skorzystać z podziału BMI na dwie wartości (np. odnieść osoby z BMI powyżej 30 do tych z BMI poniżej tej granicy, wówczas $OR[95\%CI]=[1.41, 4.90]$, $p=0.0024$)

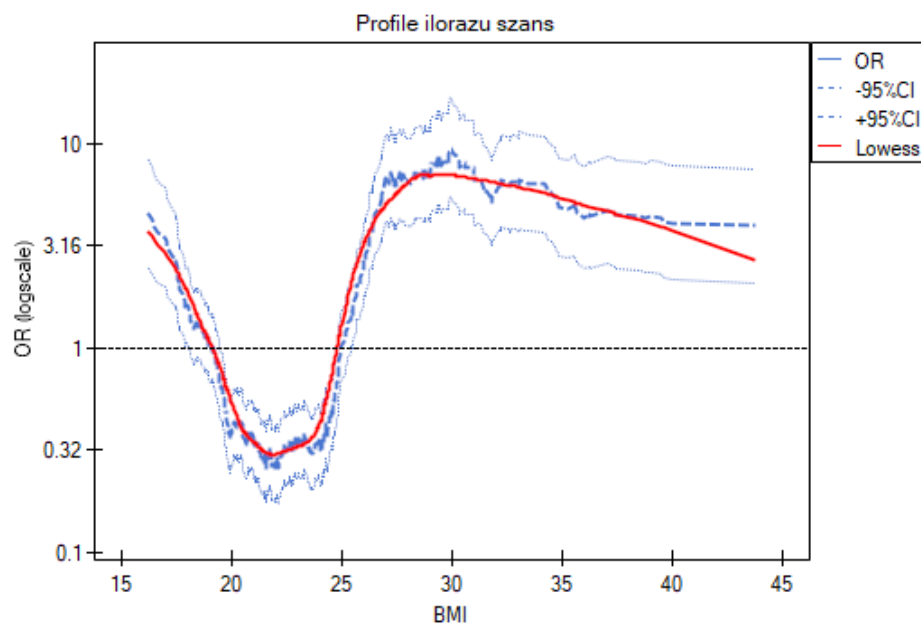
lub pozostać przy jednostkowym ilorazie szans, wskazującym stały wzrost ryzyka choroby przy wzroście BMI o jednostkę ($OR[95\%CI]=1.07[1.02, 1.13], p=0.0052$).

- Choroba B

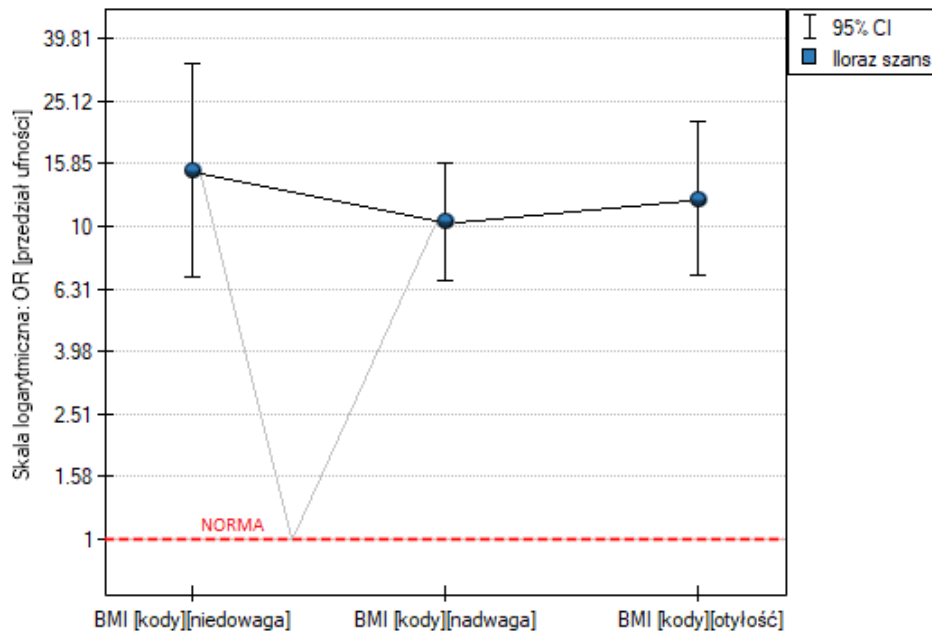
Jednostkowe zmiany ilorazu szans pokazują, że gdy punkt odcięcia BMI wybierzemy gdzieś między 22 a 35, to uzyskamy istotny statystycznie i dodatni iloraz szans pokazujący, że osoby mające BMI powyżej tej wartości mają istotnie wyższe ryzyko choroby niż osoby poniżej tej wartości.



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.



Podsumowując, mimo, że zależność dla jednostkowego ilorazu szans, czyli zależność liniowa jest istotna statystycznie, to nie warto budować takiego właśnie modelu. Znacznie lepiej podzielić BMI na kategorie. Podział pokazujący najlepiej kształt tej zależności, to podział wykorzystujący dwie lub trzy kategorie BMI, 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.

23.4.3 Przykłady dla regresji logistycznej

PRZYKŁAD 23.4. (plik wada.pqs)

Przeprowadzono badanie mające na celu identyfikację czynników ryzyka pewnej rzadko występującej wady wrodzonej u dzieci. W badaniu wzięto 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. Kolejne 9 zmiennych, to zmienne niezależne:

- MiejsceZam (2=miasto/1=wieś),
- Płeć (1=mężczyzna/0=kobieta),
- MasaUr (w kilogramach z dokładnością do 0.5kg),
- WiekM (w latach),
- KolCiąży (dziecko z której ciąży),
- PoronSamo (1=tak/0=nie),

InfOddech (1=tak/0=nie),

Palenie (1=tak/0=nie),

Wykształcenie (1=podstawowe lub niższe/2=zawodowe/3=średnie/4=wyższe).

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 licznosci próby $n \geq 10(k+1)$	Tak
Założenie licznosci próby $n \geq 10v/p$	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 Schwarzera	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 szans	-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
KolCięży	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

Jakość dopasowania modelu nie jest wysoka ($R_{Pseudo}^2 = 0.11$, $R_{Nagelkerke}^2 = 0.19$ i $R_{Cox-Snell}^2 = 0.14$). Jednocześnie model jest istotny statystycznie (wartość $p < 0.0001$ testu ilorazu wiarygodności), a zatem część zmiennych niezależnych znajdujących się w modelu jest istotna statystycznie. Wynik testu Hosmera-Lemeshowa wskazuje na brak istotności ($p = 0.2753$). Przy czym, w przypadku testu Hosmera-Lemeshowa pamiętamy o tym, że brak istotności jest pożądany, bo wskazuje na podobieństwo licznosci obserwowanych i prawdopodobieństwa przewidywanego.

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:

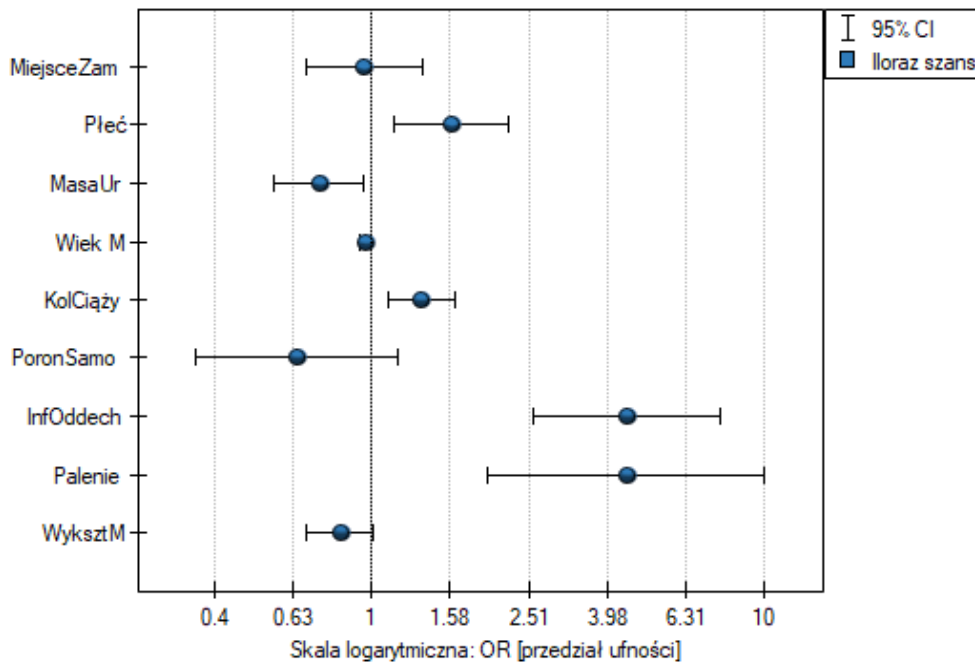
Płeć: $p = 0.0063$,
 MasaUr: $p = 0.0188$,
 KolCięży: $p = 0.0035$,
 InfOddech: $p < 0.0001$,
 Palenie: $p = 0.0003$.

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 KolCięży: $OR[95\%CI] = 1.34[1.10; 1.63]$ –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 infekcje 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 wystąpienia wady u dziecka.

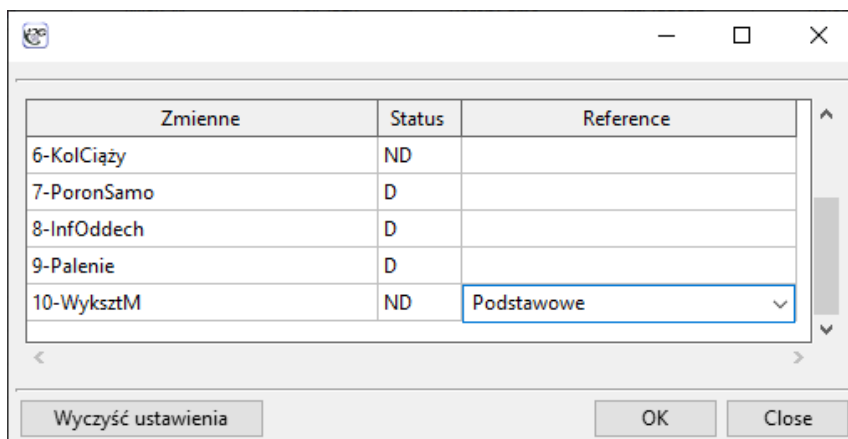
W przypadku zmiennych nieistotnych statystycznie przedział ufności dla Ilorazu Szans zawiera jedynekę co oznacza, że zmienne te nie zwiększają ani nie zmniejszają szansy na wystąpienie badanej wady. Nie można więc interpretować uzyskanego ilorazu w podobny sposób jak dla zmiennych istotnych statystycznie.

Wpływ poszczególnych zmiennych niezależnych na występowanie wady możemy również opisać przy pomocy wykresu dotyczącego ilorazu szans:



PRZYKŁAD 23.4 c.d. (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.

Regresja logistyczna	
Czas analizy	0.32 sek.
Analizowane zmienne	MiejsceZam;Płeć;MasaUr;Wii
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
KolCięży	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
WykszTM[Zawoc	-0.6823	0.3448	-1.3581	-0.0065	3.9152	0.0479	0.5055	0.2571	0.9936
WykszTM[Średni	-0.8715	0.3327	-1.5236	-0.2195	6.8633	0.0088	0.4183	0.2179	0.8029
WykszTM[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]

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 na urodzenie dziecka z wadą jaką ma matka z wykształceniem podstawowym;
- średnim szansa wystąpienia wady u dziecka stanowi 0.42 część szansy na urodzenie dziecka z wadą jaką ma matka z wykształceniem podstawowym;
- wyższym szansa wystąpienia wady u dziecka stanowi 0.45 część szansy na urodzenie dziecka z wadą jaką ma matka z wykształceniem podstawowym.

PRZYKŁAD 23.5. (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).

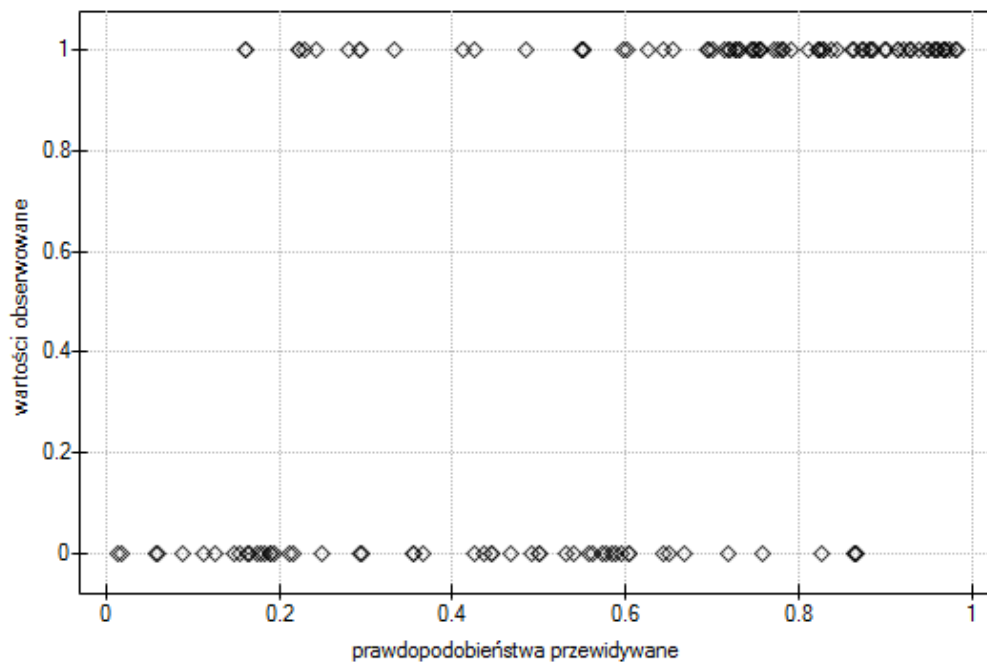
Na bazie wszystkich zmiennych zbudowano model regresji logistycznej, gdzie jako stan wyróżniony zmiennej ROZWIĄZANIE wybrano "tak".

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 licznosci próby $n \geq 10(k+1)$	Tak
Założenie licznosci próby $n \geq 10v/p$	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

Jakość jego dopasowania opisują współczynniki: $R_{Pseudo}^2 = 0.27$, $R_{Nagelkerke}^2 = 0.41$ i $R_{Cox-Snell}^2 = 0.30$. Na wystarczającą jakość dopasowania wskazuje również wynik testu Hosmera-Lemeshowa ($p = 0.1725$). Cały model jest istotny statystycznie o czym mówi wynik testu ilorazu wiarygodności ($p < 0.0001$).

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
PLEĆ	-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

Wartości obserwowane i prawdopodobieństwo przewidywane możemy zobaczyć na wykresie:



W modelu zmienne, które w sposób istotny wpływają na wynik to:

WIEK: $p = 0.0014$,

CZAS: $p = 0.0012$,

ZAKŁÓCENIA: $p = 0.0001$.

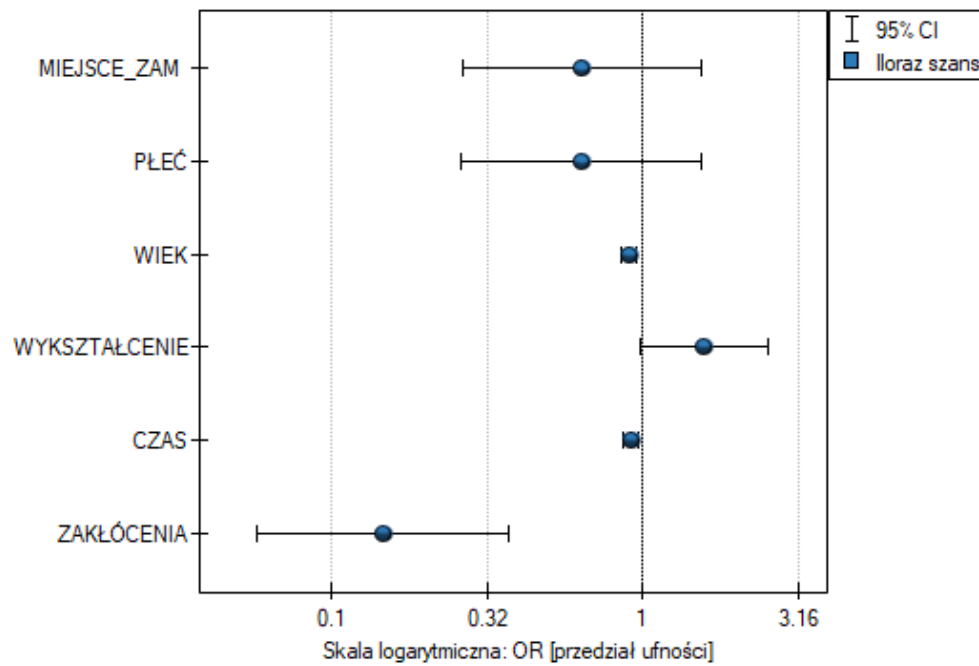
Przy czym, im osoba rozwiązująca jest młodsza, czas rozwiązywania krótszy i brak jest czynnika zakłócającego, tym większe prawdopodobieństwo poprawnego rozwiązania:

WIEK: $OR[95\%CI] = 0.90[0.85; 0.96]$,

CZAS: $OR[95\%CI] = 0.91[0.87; 0.97]$,

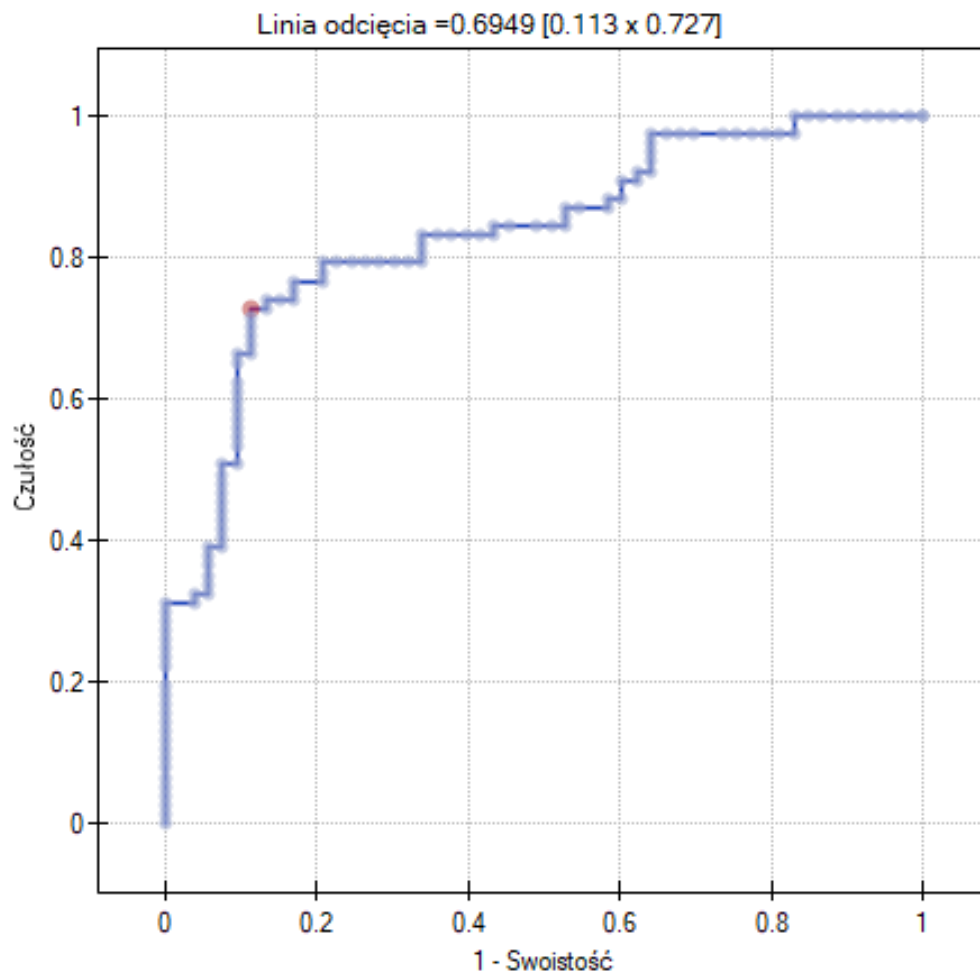
ZAKŁÓCENIA: $OR[95\%CI] = 0.15[0.06; 0.37]$.

Uzyskane wyniki Ilorazu Szans przedstawiono na poniższym wykresie:



Jeśli model miałby zostać użyty do prognozowania, to należy przyrzeć się jakości klasyfikacji. Wyliczamy w tym celu krzywe ROC.

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



Rezultat wydaje się zadowalający. Pole pod krzywą wynosi $AUC = 0.83$ i jest istotnie większe niż 0.5 ($p < 0.0001$), więc 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 nie co mniej, bo 73.85% przypadków zaklasyfikowanych poprawnie, ale uzyskamy dzięki niej więcej poprawnie zaklasyfikowanych wartości "tak" jest 83.12%, choć mniej poprawnie zaklasyfikowanych wartości "nie" jest 60.38%.

Klasyfikacja		w. obserwowana		Klasyfikacja		w. obserwowana	
w. przewidywana		1	0	w. przewidywana	1	0	
	1	56	6		64	21	
	0	21	47		13	32	
Linia odcięcia	0.6949			Linia odcięcia	0.5		
% poprawnych	79.23%			% poprawnych	73.85%		
Czułość (% pop)	72.73%			Czułość (% pop)	83.12%		
-95% CI	61.38%			-95% CI	72.86%		
+95% CI	82.26%			+95% CI	90.69%		
Swoistość (% pr)	88.68%			Swoistość (% pr)	60.38%		
-95% CI	76.97%			-95% CI	46%		
+95% CI	95.73%			+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](#).

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 (prawdopodobieństwo uzyskania odpowiedzi "tak" pod warunkiem określenia wartości zmiennych zależnych):

$$\begin{aligned}
 P(Y = tak | MIEJSCEZAM, PE, WIEK, WYKSZTACENIE, CZAS, ZAKCENIA) &= \\
 &= \frac{e^{7.23 - 0.45MIEJSCEZAM - 0.45PŁEĆ - 0.1WIEK + 0.46WYKSZTAŁCENIE - 0.09CZAS - 1.92ZAKŁÓCENIA}}{1 + e^{7.23 - 0.45MIEJSCEZAM - 0.45PŁEĆ - 0.1WIEK + 0.46WYKSZTAŁCENIE - 0.09CZAS - 1.92ZAKŁÓCENIA}} = \\
 &= \frac{e^{7.231 - 0.453 \cdot 1 - 0.455 \cdot 1 - 0.101 \cdot 50 + 0.456 \cdot 1 - 0.089 \cdot 20 - 1.924 \cdot 1}}{1 + e^{7.231 - 0.453 \cdot 1 - 0.455 \cdot 1 - 0.101 \cdot 50 + 0.456 \cdot 1 - 0.089 \cdot 20 - 1.924 \cdot 1}}
 \end{aligned}$$

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

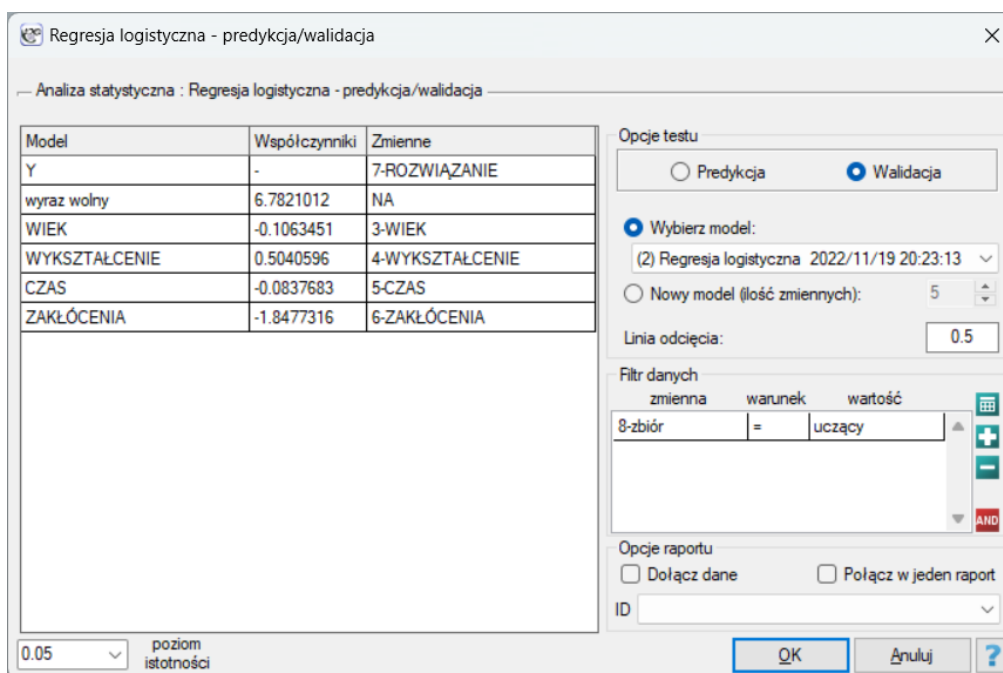
Uzyskane prawdopodobieństwo rozwiązania zadania wynosi 0.1215, więc na podstawie punktu odcięcia 0.5 przewidziany wynik to 0 –czyli zadanie nie rozwiązane poprawnie.

23.4.4 Predykcja na podstawie modelu i walidacja zbioru testowego

Walidacja

Walidacja modelu to sprawdzenie jego jakości. W pierwszej kolejności wykonywana jest na danych, na których model był zbudowany (**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. wyznaczone 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.

Okno z ustawieniami opcji walidacji wywołujemy poprzez menu Statystyki zaawansowane → Modele wielowymiarowe → Regresja logistyczna - predykcja/walidacja.



Model	Współczynniki	Zmienne
Y	-	7-ROZWIĄZANIE
wyraz wolny	6.7821012	NA
WIEK	-0.1063451	3-WIEK
WYKSZTAŁCENIE	0.5040596	4-WYKSZTAŁCENIE
CZAS	-0.0837683	5-CZAS
ZAKŁÓCENIA	-1.8477316	6-ZAKŁÓCENIA

By dokonać walidacji należy wskazać model, na podstawie którego chcemy jej dokonać. Walidacji możemy dokonać 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 nie zbudowanego 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.

Predykcja

Najczęściej ostatnim etapem analizy regresji jest wykorzystanie zbudowanego i uprzednio zweryfikowanego modelu do predykcji.

- Predykcja dla jednego obiektu może być wykonywana wraz z budową modelu, czyli w oknie analizy Statystyki zaawansowane → Modele wielowymiarowe → Regresja logistyczna ,
- Predykcja dla większej grupy nowych danych jest wykonywana poprzez menu Statystyki zaawansowane → Modele wielowymiarowe → Regresja logistyczna - predykcja/walidacja.

By dokonać predykcji należy wskazać model, na podstawie którego chcemy jej dokonać. Predykcji możemy dokonać 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 nie zbudowanego 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 predykcji.

Na podstawie nowych danych wyznaczana jest wartość prawdopodobieństwa przewidywanego przez model a następnie predykcja wystąpienia zdarzenia (1) lub jego braku (0). Punkt odcięcia, na podstawie którego wykonywana jest klasyfikacja to domyślnie wartość 0.5. Użytkownik może zmienić tę wartość na dowolną wartość z przedziału (0 – 1) np. wartość sugerowaną przez krzywą ROC.

PRZYKŁAD 23.5 c.d. (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),

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.

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 Akaikiego	139.0822737
AICc - skorygowane kryterium Akaikiego	139.4022737
BIC - bayesowskie kryterium Schwarzera	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

Dla zbioru uczącego wartości opisujące jakość dopasowania modelu nie są bardzo wysokie $PseudoR^2 = 0.26$ a $R^2(Nagelkerke) = 0.39$, ale już jakość jego predykcji jest zadowalająca (AUC[95%CI]=0.82[0.75, 0.90], czułość =82%, swoistość 60%).

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 Akaikiego	68.7731337
AICc - skorygowane kryterium Akaikiego	69.5004065
BIC - bayesowskie kryterium Schwarzera	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

Dla zbioru testowego wartości opisujące jakość dopasowania modelu są nawet wyższe niż dla danych uczących $PseudoR^2 = 0.65$ a $R^2(Nagelkerke) = 0.90$. Jakość predykcji dla danych testowych jest wciąż zadowalająca (AUC[95%CI]=0.82[0.71, 0.93], czułość =73%, swoistość 64%), dlatego użyjemy modelu do predykcji. 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).

Regresja logistyczna - predykcja/walidacja		Współczynniki modelu	
Analizowane zmienne	WIEK	wyraz wolny	6.7821012
	WYKSZTAŁCENIE	WIEK	-0.1063451
	CZAS	WYKSZTAŁCENI	0.5040596
	ZAKŁÓCENIA	CZAS	-0.0837683
Filtr danych	zbiór=nowy	ZAKŁÓCENIA	-1.8477316
Liczność / niezinterpretowane	0	Predykcja	
Liczność / braki danych	0	Pred. prawdopo	Pred. Y
Poziom istotności	0.05	0.1744817	0
Liczność	3	0.6393883	1
Predykcja		0.962964	1
Linia odcięcia	0.5		
liczba 0	1		
liczba 1	2		
procent 0	33.333%		
procent 1	66.667%		

Okazują 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 młodziej 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.

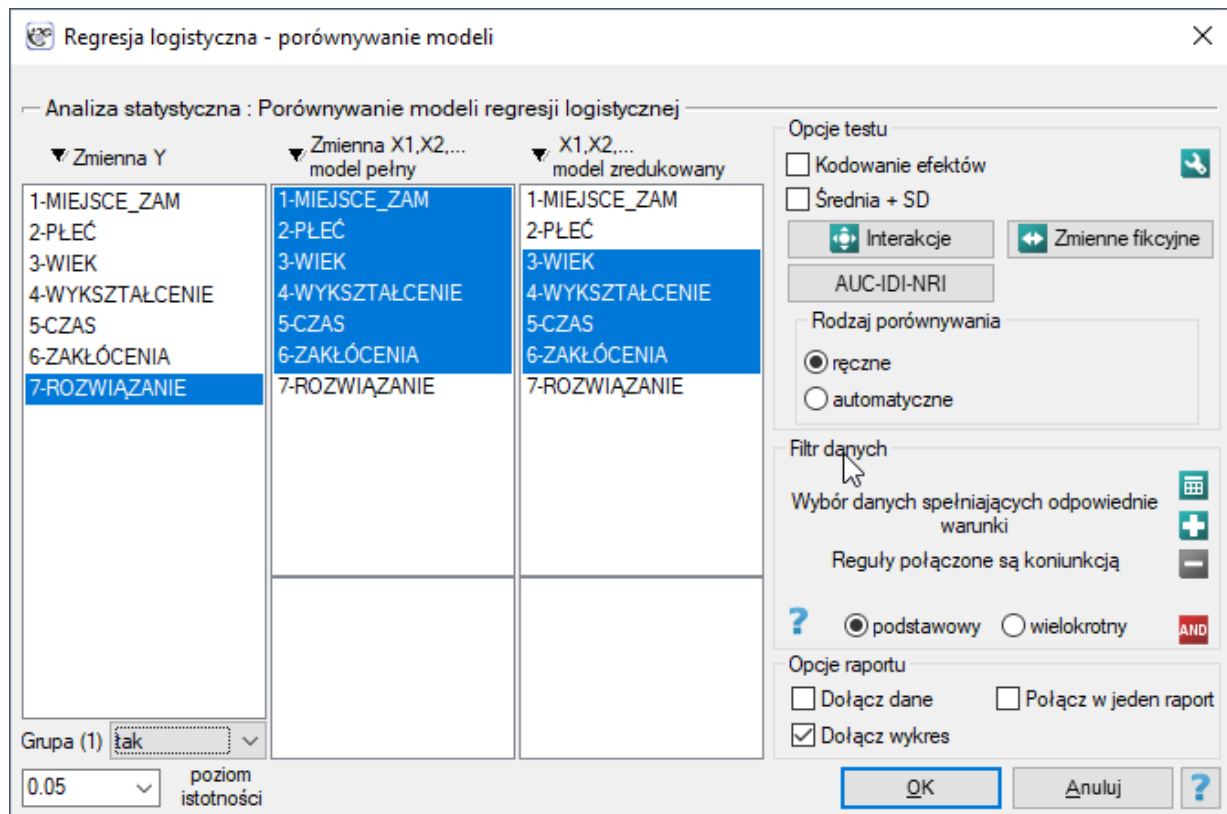
Gdybyśmy chcieli postawić prognozę na podstawie innego modelu (np. uzyskanego podczas innego badania naukowego: $ROZWIĄZANIE=6-0.1*WIEK+0.5*WYKSZT-0.1*CZAS-2*ZAKŁÓCENIA$) - wystarczy, że w oknie analizy wybierzemy nowy model, ustawimy jego współczynniki i prognozę dla wybranych osób można powtórzyć w oparciu o ten model.

Regresja logistyczna - predykcja/walidacja		Współczynniki modelu	
Analizowane zmienne	WIEK	wyraz wolny	6
	WYKSZTAŁCENIE	wiek	-0.1
	CZAS	wykształcenie	0.5
	ZAKŁÓCENIA	czas	-0.1
Filtr danych	zbiór=nowy	zakłócenia	-2
Liczność / niezinterpretowane	0	Predykcja	
Liczność / braki danych	0	Pred. prawdopo	Pred. Y
Poziom istotności	0.05	0.0758582	0
Liczność	3	0.3775407	0
Predykcja		0.9241418	1
Linia odcięcia	0.5		
liczba 0	2		
liczba 1	1		
procent 0	66.667%		
procent 1	33.333%		

Tym razem, zgodnie z prognozą nowego modelu, przewidywania dla pierwszej i drugiej osoby są negatywne, a trzeciej pozytywne.

23.5 PORÓWNYWANIE MODELI REGRESJI LOGISTYCZNEJ

Okno z ustawieniami opcji porównywania modeli wywołujemy poprzez menu Statystyka → Modele wielowymiarowe → Regresja logistyczna – porównywanie modeli



Ze względu na możliwość jednoczesnej analizy wielu zmiennych niezależnych w jednym modelu regresji logistycznej, podobnie jak w liniowej regresji wielorakiej, istnieje 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.

Porównując modele z różną liczbą zmiennych niezależnych zwracamy uwagę na dopasowanie modelu oraz kryteria informacyjne. Dla każdego modelu wyliczamy również maksimum funkcji wiarygodności, które następnie porównujemy przy użyciu testu ilorazu wiarygodności.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: L_{FM} = L_{RM}, \\ \mathcal{H}_1 &: L_{FM} \neq L_{RM}, \end{aligned}$$

gdzie:

L_{FM}, L_{RM} – maksimum funkcji wiarygodności w porównywanych modelach (pełnym i zredukowanym).

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 licznosci) 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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

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. Jeśli porównywane modele nie różnią się istotnie, to 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 jest istotna statystycznie oznacza to, że jeden z nich (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 w oparciu o analizowane zmienne niezależne. Gdy znanych jest już wiele zmiennych (czynników) zwiększających ryzyko wystąpienia zdarzenia, wówczas ważnym kryterium dla nowego kandydata na czynnik ryzyka jest poprawa skuteczności predykcji po dołączeniu tego czynnika do modelu. Dla ustalenia uwagi posłużmy się przykładem. Założmy, że badamy czynniki ryzyka wystąpienia choroby wieńcowej. Do znanych czynników ryzyka tej choroby należą wiek, wartości ciśnienia skurczowego i rozkurczowego, otyłość, cholesterol czy też palenie. Badacze są jednak zainteresowani jak bardzo włączenie poszczególnych czynników do modelu regresji pozwoli na znaczną poprawę oszacowywania ryzyka wystąpienia choroby. Czynniki ryzyka dołączone do modelu będą miały znaczenie prognostyczne, jeśli nowy i większy model (zawierający te czynniki) będzie wykazywał lepszą wartość prognostyczną niż model ich pozbawiony. Wartość prognostyczna modelu wynika z wyznaczonej wartości przewidywanego prawdopodobieństwa wystąpienia zdarzenia, w tym przypadku choroby wieńcowej. Wartość ta jest wyznaczana na podstawie modelu dla każdej badanej osoby. Im bliższe wartości 1 jest przewidywane prawdopodobieństwo, tym bardziej prawdopodobna jest choroba. Na bazie prawdopodobieństwa przewidywanego można wyznaczyć i porównać pomiędzy różnymi modelami wartość pola AUC pod krzywą ROC a także współczynnik NRI i IDI .

Zmiana pola pod krzywą ROC

Krzywa ROC w modelach regresji logistycznej zbudowana jest w oparciu o klasyfikację przypadków do grupy doświadczającej zdarzenia lub nie, oraz przewidywane prawdopodobieństwo zmiennej zależnej P . Czym większe pole pod krzywą, tym trafniej prawdopodobieństwo wyznaczone przez model przewiduje rzeczywiste wystąpienie zdarzenia. Jeśli porównujemy modele zbudowane w oparciu o większą lub mniejszą liczbę czynników prognostycznych, to porównując wielkość pola pod krzywą możemy sprawdzić, czy dołożenie czynników poprawiło znacząco predykcję modelu. Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: AUC_{FM} = AUC_{RM}, \\ \mathcal{H}_1 &: AUC_{FM} \neq AUC_{RM}.\end{aligned}$$

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Poprawa rekasyfikacji netto

Miara ta oznaczana jest skrótem NRI (*ang. Net Reclassification Improvement*). NRI skupia się na tabeli rekasyfikacji opisującej przesunięcie wartości prawdopodobieństwa w górę lub w dół po

dołożeniu nowego czynnika do modelu. Wyznacza się go na bazie dwóch oddzielnych współczynników, tzn. współczynnika wyznaczonego oddzielnie dla obiektów doświadczających zdarzenia (1) i oddzielnie dla tych niedoświadczających zdarzenia (0). *NRI* może być wyznaczany przy zadanym podziale prawdopodobieństwa przewidywanego na kategorie (*NRI* kategoryalny) lub bez konieczności wyznaczania kategorii (*NRI* ciągły).

- ***NRI* kategoryalny** wymaga arbitralnego wyznaczenia podziału wartości prawdopodobieństwa przewidywanego z modelu. Podanych punktów podziału może być maksymalnie 9, a więc przewidywanych kategorii maksymalnie 10. Najczęściej jednak stosuje się jeden lub dwa punkty podziałowe. Przy czym należy zauważyć, że wartości kategoryalnego *NRI* mogą być z sobą porównywane tylko wtedy, jeśli bazowały na tych samych punktach podziału. Dla zobrazowania sytuacji ustalmy dwa przykładowe punkty podziału prawdopodobieństwa: 0.1 i 0.3. Jeśli badana osoba w modelu "starym" (mniejszym) uzyskała prawdopodobieństwo poniżej 0.1, a w "nowym modelu" (zwiększonym o nowy potencjalny czynnik ryzyka) prawdopodobieństwo znajdujące się pomiędzy 0.1 a 0.3, to znaczy, że osoba ta została reklasyfikowana w górę (tabela, sytuacja 1). Jeśli wartość prawdopodobieństwa z obu modeli mieści się w tym samym przedziale, wówczas osoba nie została reklasyfikowana (tabela, sytuacja 2), natomiast jeśli prawdopodobieństwo z modelu "nowego" będzie niższe niż w modelu "starym", oznacza to reklasyfikację w dół badanej osoby (tabela, sytuacja 3).

modele regresji	sytuacja 1		sytuacja 2		sytuacja 3	
prawdop. przewidywane	"stary"	"nowy"	"stary"	"nowy"	"stary"	"nowy"
[0.3 do 1]					⊕	
[0.1; 0.3)		⊕	⊕	⊕		⊕
[0; 0.1)	⊕					

- ***NRI* ciągły** nie wymaga arbitralnego wyznaczenia kategorii, ponieważ każda, nawet najmniejsza zmiana prawdopodobieństwa w górę lub w dół w stosunku do prawdopodobieństwa wyznaczonego w "starym modelu", traktowana jest jako przejście do kolejnej kategorii. Kategorii jest więc nieskończenie wiele, tak jak wiele jest możliwych zmian.

Uwaga!

Stosowanie *NRI* ciągłego nie wymaga arbitralnego definiowania punktów podziału prawdopodobieństwa, jednak nawet niewielkie zmiany ryzyka (nie mające odzwierciedlenia w klinicznych obserwacjach) mogą wpływać na zwiększenie lub zmniejszenie tego współczynnika. Kategoryalny współczynnik *NRI* pozwala na odzwierciedlenie tylko ważnych dla badacza zmian polegających na przekroczeniu zadanych wartości ryzyka wystąpienia zdarzenia (wartości prawdopodobieństwa przewidywanego).

By wyznaczyć *NRI* definiujemy:

$$\hat{P}_{up,events} = \frac{\#events_{up}}{\#events}$$

$$\hat{P}_{down,events} = \frac{\#events_{down}}{\#events}$$

$$\hat{P}_{up,nonevents} = \frac{\#nonevents_{up}}{\#nonevents}$$

$$\hat{P}_{down,nonevents} = \frac{\#nonevents_{down}}{\#nonevents}$$

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 nie doświadczającej zdarzenia, u których nastąpiła zmiana prawdopodobieństwa przewidywanego o przynajmniej jedną kategorię w górę,

$\#nonevents_{down}$ - liczba obiektów z grupy nie doświadczającej zdarzenia, u których nastąpiła zmiana prawdopodobieństwa przewidywanego o przynajmniej jedną kategorię w dół,

$\#nonevents$ - liczba obiektów w grupie nie doświadczającej zdarzenia.

Ogólny współczynnik NRI oraz współczynniki wyrażające procentową zmianę klasyfikacji jest wyznaczany z wzoru:

$$NRI = (\hat{p}_{up,events} - \hat{p}_{down,events}) - (\hat{p}_{up,nonevents} - \hat{p}_{down,nonevents})$$

$$NRI_{events} = \hat{p}_{up,events} - \hat{p}_{down,events}, \quad NRI_{nonevents} = \hat{p}_{down,nonevents} - \hat{p}_{up,nonevents},$$

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 wyrażony jako suma współczynników NRI_{events} i $NRI_{nonevents}$, przez co jest współczynnikiem domyślnie ważonym przez częstość zdarzenia i nie może być interpretowany jako procent.

Współczynniki NRI_{events} należą do przedziału od -1 do 1 (od -100% do 100%), a ogólny współczynnik NRI do przedziału od -2 do 2. Wartości dodatnie współczynników świadczą o korzystnej reklasyfikacji, a ujemne o niekorzystnej reklasyfikacji na skutek dołożenia nowej zmiennej do modelu.

Test Z do sprawdzania istotności współczynnika NRI

Przy pomocy tego testu badamy czy zmiana klasyfikacji wyrażona współczynnikiem NRI była istotna.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: NRI = 0, \\ \mathcal{H}_1 &: NRI \neq 0. \end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{NRI}{SE(NRI)}$$

gdzie:

$$SE(NRI) = \left[\left(\frac{\#events_{up} + \#events_{down}}{\#events^2} - \frac{(\#events_{up} + \#events_{down})^2}{\#events^3} \right) + \left(\frac{\#nonevents_{down} + \#nonevents_{up}}{\#nonevents^2} - \frac{(\#nonevents_{down} + \#nonevents_{up})^2}{\#nonevents^3} \right) \right]^{1/2}$$

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej** **wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Zintegrowana poprawa dyskryminacji

Miara ta oznaczana jest skrótem *IDI* (ang. *Integrated Discrimination Improvement*). Współczynnik *IDI* pokazuje różnicę pomiędzy wartością średniej zmiany prawdopodobieństwa przewidywanego pomiędzy grupą obiektów doświadczających zdarzenia a grupą obiektów, które zdarzenia nie doświadczyły.

$$IDI = \overline{p(diff)}_{events} - \overline{p(diff)}_{nonevents}$$

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*

Przy pomocy tego testu badamy czy różnica pomiędzy wartością średniej zmiany prawdopodobieństwa przewidywanego pomiędzy grupą obiektów doświadczających zdarzenia a obiektami nie doświadczającymi zdarzenia, wyrażona współczynnikiem *IDI*, była istotna.

Hipotezy:

$$\mathcal{H}_0 : IDI = 0,$$

$$\mathcal{H}_1 : IDI \neq 0.$$

Statystyka testowa ma postać:

$$Z = \frac{IDI}{SE(IDI)}$$

gdzie:

$$SE(IDI) = \sqrt{\frac{sd(diff)_{events}^2}{\#events} + \frac{sd(diff)_{nonevents}^2}{\#nonevents}}$$

Statystyka *Z* ma asymptotycznie (dla dużych liczności) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość *p*** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

W programie PQStat porównywanie modeli możemy 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.

- **Automatyczne** porównywanie modeli jest wykonywane w kilku krokach:

krok 1 Zbudowanie modelu z wszystkich zmiennych.

krok 2 Usunięcie jednej zmiennej z modelu. Usuwana zmienna to ta, która ze statystycznego punktu widzenia wnosi do aktualnego modelu najmniej informacji.

krok 3 Porównanie modelu pełnego i zredukowanego.

krok 4 Usunięcie kolejnej zmiennej z modelu. Usuwana zmienna to ta, która ze statystycznego punktu widzenia wnosi do aktualnego modelu najmniej informacji.

krok 5 Porównanie modelu wcześniejszego i nowo zredukowanego.

...

W ten sposób powstaje wiele, coraz mniejszych modeli. Ostatni model zawiera tylko 1 zmienną niezależną.

PRZYKŁAD 23.5 c.d. (plik zadanie.pqs)

W eksperymencie badającym umiejętność koncentracji, dla 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),

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 podstawie tych zmiennych nie możemy z sukcesem dokonywać klasyfikacji). Sprawdzimy czy, ze statystycznego punktu widzenia, model pełny jest lepszy niż model po usunięciu tych dwóch zmiennych.

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 Akaikiego	140.7082
AICc - skorygowane kryterium Akaikiego	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 Akaikiego	139.0823
AICc - skorygowane kryterium Akaikiego	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. Wald:	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_ZAM	-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	
WYKSZTAŁCENIE	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	
ZAKŁÓCENIA	-1.924	0.475056	-2.855092	-0.992908	16.402912	0.000051	0.146022	0.057551	0.370498	

Model 2									
	wsp. b	błąd b	-95% CI	+95% CI	stat. Wald:	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

Wynik testu ilorazu wiarygodności ($p = 0.3051$) wskazuje, że 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.

Porównania obu modeli pod względem zdolności do klasyfikacji możemy dokonać porównując krzywe ROC dla tych modeli, wartość NRI i IDI. W tym celu wybieramy odpowiednią opcję w oknie analizy. Uzyskany raport, podobnie jak wcześniejszy, wskazuje, że modele nie różnią się jakością predykcji tzn. wartości p dla porównania krzywych ROC oraz do oceny wskaźników NRI i IDI są nieistotne statystycznie. Decydujemy zatem pominąć płeć i miejsce zamieszkania w ostatecznym modelu.

Porównanie - Model 1 vs Model 2	
Porównanie krzywe ROC [Różnica AUC]	
AUC Model 1	0.8346
SE(AUC)	0.0355
-95% CI	0.765
+95% CI	0.9042
AUC Model 2	0.8231
SE(AUC)	0.0368
-95% CI	0.751
+95% CI	0.8952
AUC1-AUC2 [DeLong's method]	0.0115
SE(AUC1-AUC2)	0.0109
-95% CI	0
+95% CI	0.0328
Statystyka Z	1.0601
Wartość p	0.2891
IDI Zintegrowana poprawa dyskryminacji Model 1 vs	0.014
-95% CI	-0.0067
+95% CI	0.0348
Statystyka Z	1.3229
Wartość p	0.1859
NRI Poprawa reklasyfikacji netto (ciągła) Model 1 vs	0.1946
-95% CI	-0.1507
+95% CI	0.5399
Statystyka Z	1.1043
Wartość p	0.2695
NRI (1)	-0.013
NRI (0)	0.2075
NRI Poprawa reklasyfikacji netto (kategorialna) Mod	0
-95% CI	-0.0823
+95% CI	0.0823
Statystyka Z	0
Wartość p	1
NRI (1)	0
NRI (0)	0
Punkty odcięcia prawdopodobieństwa	0.5
Chi-kwadrat - porównanie	2.374
Stopnie swobody	2
Wartość p	0.3051

- **Automatyczne** porównywanie modeli.

W przypadku automatycznego porównywania modeli uzyskaliśmy bardzo podobne wyniki. Najlepszym modelem jest model zbudowany na podstawie zmiennych niezależnych: WIEK, WYKSZTAŁCENIE, CZAS rozwiązywania, ZAKŁUCENIA.

Na podstawie powyższych analiz, ze statystycznego punktu widzenia, optymalnym modelem jest model zawierający 4 najważniejsze zmienne niezależne: WIEK, WYKSZTAŁCENIE, CZAS rozwiązywania, ZAKŁUCENIA. Dokładną jego analizę możemy przeprowadzić w module Regresja Logistyczna. Jednak ostateczna decyzja, który model wybrać należy do eksperymentatora.

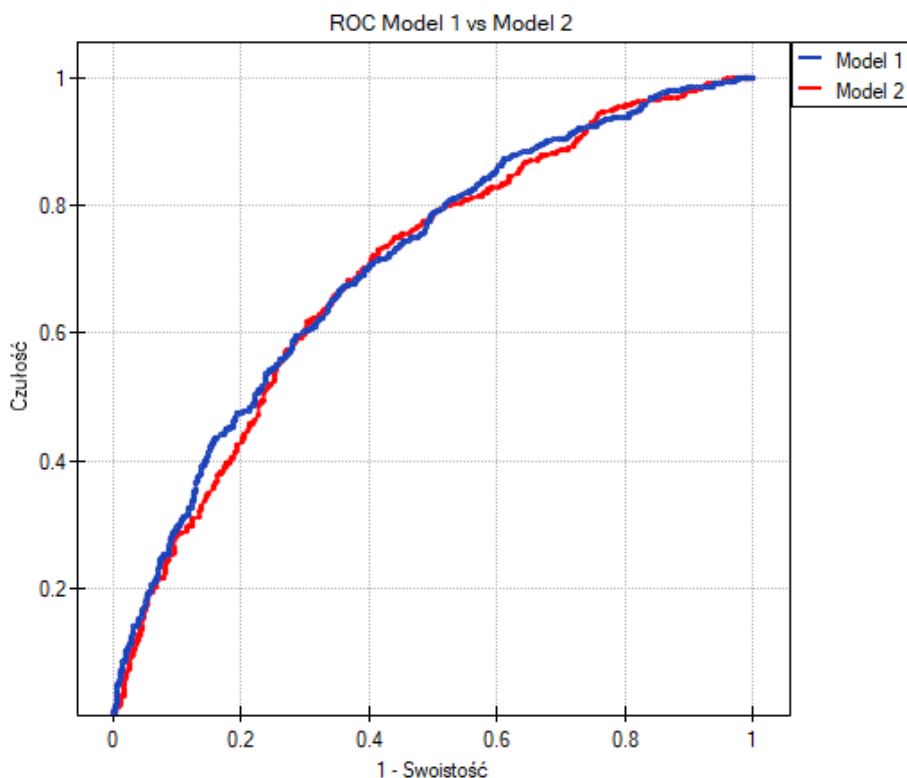
PRZYKŁAD 23.6. Przykład(plik Serce.pqs)

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.

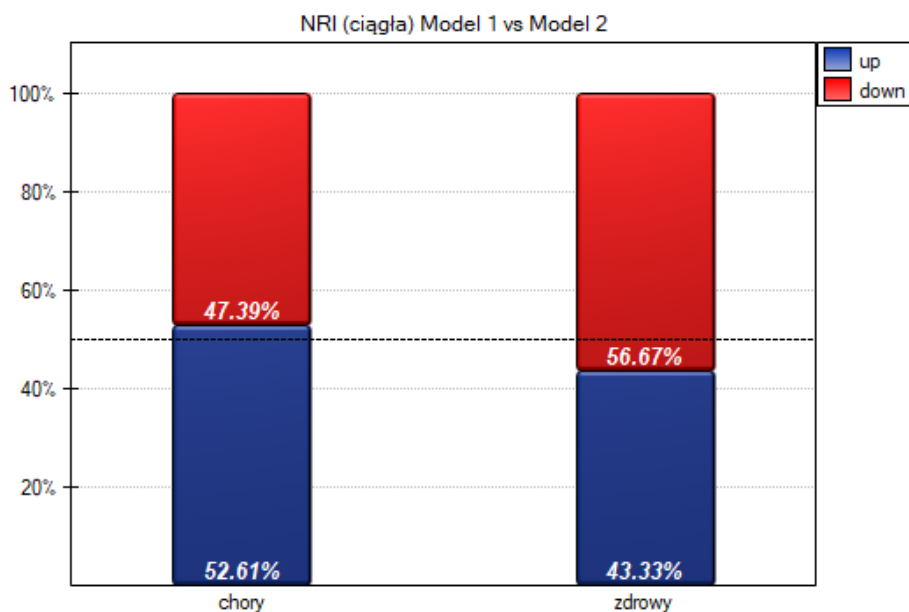
Porównujemy model regresji logistycznej opisujący ryzyko choroby serca na podstawie wszystkich badanych zmiennych z modelem pozbawionym informacji o paleniu. W oknie analizy zaznaczamy opcje związane z oceną predykcji, czyli krzywą ROC oraz współczynniki NRI. Dodatkowo wskazujemy, by w raporcie znalazły się wszystkie proponowane wykresy.

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	
-95% CI	0.0142
+95% CI	0.0078
Statystyka Z	0.0206
Wartość p	4.3486
NRI Poprawa reklasyfikacji netto (ciągła) Model 1 vs	
-95% CI	0.1855
+95% CI	0.0805
Statystyka Z	0.2905
Wartość p	3.4634
NRI (chory)	0.0522
NRI (zdrowy)	0.1333

Analiza raportu wskazuje na ważne różnice w predykcji na skutek dodania do modelu informacji o paleniu, chociaż nie są one istotne w opisie krzywej ROC ($p=0.057$).



Ciągłe wartości współczynnika IDI i NRI wskazują na istotną statystycznie i korzystną zmianę (wartości tych współczynników są dodatnie, a wartości $p < 0.05$). Prognoza dla osób z chorobą serca poprawiła się o ponad 5% a osób bez tej choroby o ponad 13% ($NRI(\text{chory}) = 0.0522$, $NRI(\text{zdrowy}) = 0.1333$) na skutek uwzględnienia informacji o paleniu.

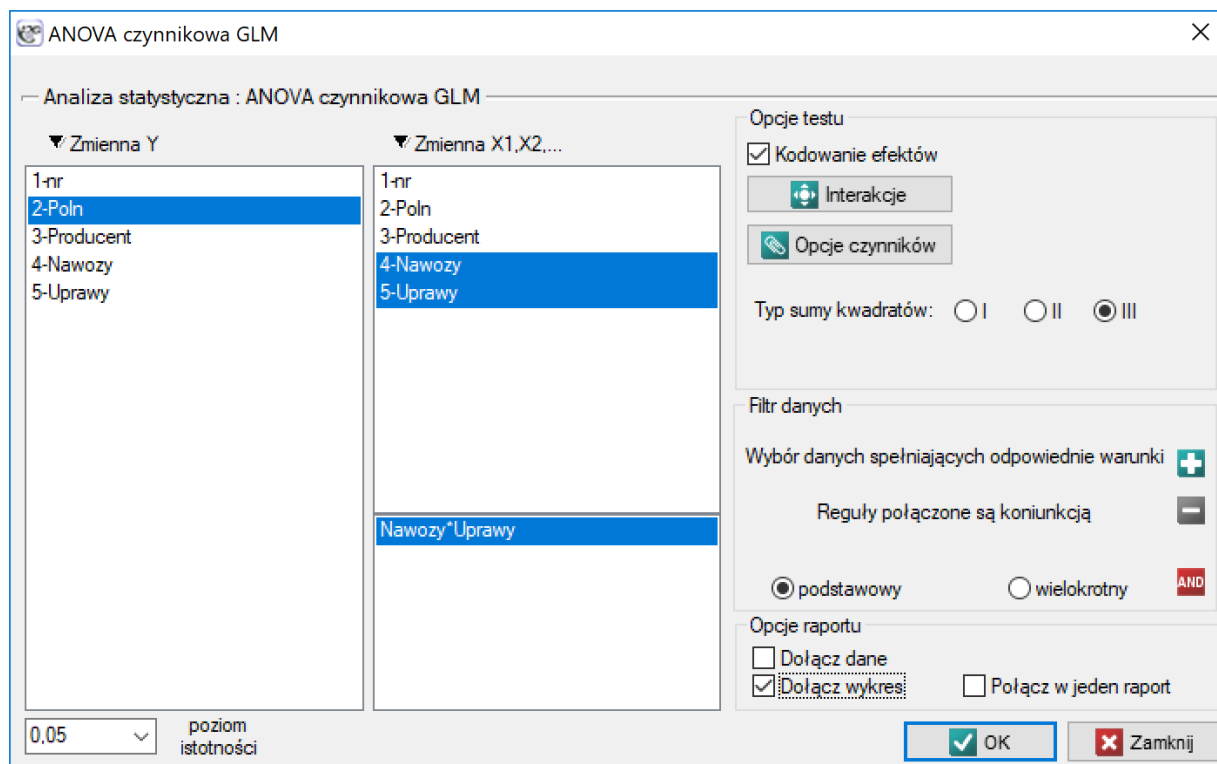


Wnioski wyciągnięte na bazie NRI widzimy również na wykresie. Wzrost prognozowanego przez model prawdopodobieństwa choroby widzimy u osób chorych (więcej osób zostało przeklasyfikowanych w górę niż w dół 52.61% vs 47.39%) natomiast spadek prawdopodobieństwa dotyczy w większym stopniu osób zdrowych (więcej osób zostało przeklasyfikowanych w dół niż w górę 56.67% vs 43.33%).

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.

23.6 ANOVA CZYNNIKOWA - GLM

Okno z ustawieniami opcji ANOVA czynnika GLM wywołujemy poprzez menu Statystyka→Modele wielowymiarowe→ANOVA czynnika GLM



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.

Przykład

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 badanych osobach mamy pewne 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 wnikającej * porównanie dochodu dla osób z różnym wykształceniem (podstawowym, zawodowym, średnim, wyższym) przy uwzględnieniu płci jako zmiennej wnikającej	zależność dochodu (Y) od płci (X₁) i wykształcenia (X₂) * płeć jako kolumna o dwóch kategoriach * wykształcenie rozbite na zmienne fikcyjne obie zmienne są wzajemnie wnikające
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 (X₁), od wykształcenia (X₂) oraz od interakcji płci i wykształcenia (X₁ * X₂) * płeć jako kolumna o dwóch kategoriach * wykształcenie rozbite na zmienne fikcyjne * interakcja to iloczyn płci i wykształcenia zmienne w modelu są wzajemnie wnikające

Analiza GLM może być wykorzystana w każdym z powyższych przypadków, ponieważ jednak analiza regresji wielorakiej podobnie jak jednoczynnikowa ANOVA zostały omówione w oddzielnych rozdziałach, w tym rozdziale przedstawimy wykorzystanie GLM w ANOVA wieloczynnikowa.

ANOVA czynnikowa jest takim rodzajem analizy wariancji, w którym możemy wykorzystać zarówno jedną 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.

Wpływ czynników wnikających

Pomimo, że wszystkie czynniki biorące udział w analizie są wobec siebie uwikłane, to ich wpływ na istotność poszczególnych czynników można kontrolować. Istnieją trzy sposoby, przy pomocy których badając istotność poszczególnych czynników można uwzględniać wpływ zmiennych wnikających. Zależą one od sposobu wyznaczania sumy kwadratów:

- **Sumy kwadratów typu I**

Sumy kwadratów typu I zależą od kolejności w jakiej w modelu znajdują się poszczególne czynniki. Ten rodzaj sumy kwadratów powoduje, że istotność czynnika który interpretujemy jest korygowana o te zmienne, których kolejność w modelu była wcześniejsza, pozostałe zmienne w modelu wpływają jedynie pośrednio na wynik analizy. 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 wnikające wykorzystywane są wprost tylko czynniki: $A, B, A * B, C$.

Sumy kwadratów dla czynnika $A * C$ wlicza się wówczas następująco:

$$SS(A * C) = SS(A, B, A * B, C, A * C, B * C, A * B * C, D) - SS(A, B, A * B, C)$$

Stosowanie sumy kwadratów typu I

Wskazania: Kiedy badanie jest w pełni zbalansowane, z równymi lub proporcjonalnymi licznosciami poszczególnych kategorii, również wtedy, gdy występują interakcje.

Przeciwwskazania: Kiedy badanie jest niezbalansowane (różne licznosci poszczególnych kategorii) i/lub występują interakcje.

- **Sumy kwadratów typu II**

Ten rodzaj sumy kwadratów powoduje, że istotność czynnika który interpretujemy jest korygowana o te zmienne, których rząd jest taki sam lub niższy, pozostałe zmienne w modelu wpływają jedynie pośrednio na wynik analizy. 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 (poprzez sumy kwadratów dla błędu) ale jako zmienne wklajające 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:

$$SS(A * C) = SS(A, B, A * B, C, A * C, B * C, A * B * C, D) - SS(A, B, A * B, C, B * C, D)$$

Stosowanie sumy kwadratów typu II

Wskazania: Kiedy badanie jest w pełni zbalansowane, z równymi lub proporcjonalnymi licznosciami poszczególnych kategorii, również wtedy, gdy występują interakcje.

Przeciwwskazania: Kiedy badanie jest niezbalansowane (różne licznosci poszczególnych kategorii) i/lub występują interakcje.

- **Sumy kwadratów typu III**

Zalecamy stosować ten rodzaj kodowania, gdy wybrane jest [kodowania efektów](#).

Ten rodzaj sumy kwadratów powoduje, że istotność czynnika który interpretujemy jest korygowana o wszystkie pozostałe zmienne w modelu. 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 (poprzez sumy kwadratów dla błędu) a jako zmienne wklajające 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:

$$SS(A * C) = SS(A, B, A * B, C, A * C, B * C, A * B * C, D) - SS(A, B, A * B, C, B * C, A * B * C, D)$$

Stosowanie sumy kwadratów typu III

Wskazania: Kiedy badanie jest zbalansowane lub niezbalansowane, również wtedy, gdy występują interakcje.

Przeciwwskazania: Kiedy badanie zawiera podklasy o brakujących obserwacjach.

W PQStat domyślnie wybrane są sumy kwadratów typu III, ze względu na ich uniwersalność. Domyślnie zaznaczona jest też opcja kodowanie efektów opisana w rozdziale [23.1](#). Należy pamiętać, że wybór odpowiedniego kodowania wpływa zarówno na interpretację współrzędnych modelu jak i na istotność poszczególnych czynników w ANOVA czynnika - szczególnie przy niezbalansowanych układach.

Podstawowe warunki stosowania:

- pomiar na [skali interwałowej](#),
- próby pochodzą z populacji o rozkładzie normalnym ([normalność](#) zmiennych bądź reszt modelu),
- [model niezależny](#),
- [równość wariancji](#) badanych zmiennych porównywanych populacji.

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 kategorii uzyskamy osiem:

- 1) kategoria żeńska z wykształceniem podstawowym,
- 2) żeńska z wykształceniem zawodowym,
- 3) żeńska z wykształceniem średnim,
- 4) żeńska z wykształceniem wyższym,
- 5) kategoria męska z wykształceniem podstawowym,
- 6) męska z wykształceniem zawodowym,
- 7) męska z wykształceniem średnim,
- 8) męska z wykształceniem wyższym.

Analiza typu ANOVA i modele regresji traktowane są równoważnie, i w ogólnym przypadku ich hipotezy są zbieżne. Hipotezy dla efektów głównych A i B i efektu interakcji $A * B$ przedstawimy w obu tych ujęciach. W interpretacji tych hipotez należy pamiętać, że hipotezy dla danych czynników korygowane są o te z pozostałych czynników, które dana analiza uwzględnia.

Podejście ANOVA

Hipotezy dla czynnika A :

$$\begin{aligned} \mathcal{H}_0 &: \mu_1 = \mu_2 = \dots = \mu_a, \\ \mathcal{H}_1 &: \text{nie wszystkie } \mu_i \text{ są sobie równe } (i = 1, 2, \dots, a), \end{aligned}$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_a$ – średnie czynnika A dla poszczególnych jego kategorii.

Hipotezy dla czynnika B :

$$\begin{aligned} \mathcal{H}_0 &: \mu_1 = \mu_2 = \dots = \mu_b, \\ \mathcal{H}_1 &: \text{nie wszystkie } \mu_j \text{ są sobie równe } (j = 1, 2, \dots, b), \end{aligned}$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_b$ – średnie czynnika B dla poszczególnych jego kategorii.

Hipotezy dla interakcji czynników $A * B$:

$$\begin{aligned} \mathcal{H}_0 &: \mu_1 = \mu_2 = \dots = \mu_{ab}, \\ \mathcal{H}_1 &: \text{nie wszystkie } \mu_k \text{ są sobie równe } (k = 1, 2, \dots, ab), \end{aligned}$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_{ab}$ – średnie interakcji czynników $A * B$ dla poszczególnych ich kategorii.

Podejście regresyjne

Podejście modelowe zakłada działanie modelu regresji

$$Y = \mu + \alpha_i A_i + \beta_j B_j + (\alpha\beta)_k B_k + \epsilon.$$

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).

Hipotezy dla czynnika A :

$$\mathcal{H}_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a = 0,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \alpha_i = 0,$$

Hipotezy dla czynnika B :

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \beta_j = 0,$$

Hipotezy dla interakcji czynników $A * B$:

$$\mathcal{H}_0 : (\alpha\beta)_1 = (\alpha\beta)_2 = \dots = (\alpha\beta)_{ab} = 0,$$

$$\mathcal{H}_1 : \text{nie wszystkie } (\alpha\beta)_k = 0,$$

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**. Dokładny opis kodowania można znaleźć w rozdziale Przygotowanie zmiennych do analizy w modelach wielowymiarowych. Domyślnie program wybiera kodowanie efektów. Odznaczenie tej opcji jest równoważne z wybraniem kodowania zero-jedynkowego.

Uwaga!

W przypadku stosowaniu sumy kwadratów typu III, gdy występują interakcje, wskazane jest stosowanie kodowania efektów.

PRZYKŁAD 23.7. (plik plon.pqs)

Aby zwiększyć plon roślin uprawnych, opracowuje się nawozy według coraz nowszych technologii. Na 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) H₀: 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).

Ponadto, choć jest to w tym przypadku mniej interesujące, sprawdzimy czy:

2) H₀: Średnie plony uzyskane w gospodarstwie 1 są takie same jak w gospodarstwie drugim (niezależnie od mieszanki stosowanego nawozu).

Równoważnie hipotezy te można zapisać korzystając z podejścia regresyjnego:

1) H₀: Współczynniki określające zmianę uzyskanego plonu przy zmianie stosowanego nawożenia są zerowe (niezależnie od gospodarstwa prowadzącego uprawę).

2) H₀: Współczynnik określający zmianę uzyskanego plonu przy zmianie gospodarstwa prowadzącego uprawę jest zerowy (niezależnie od mieszanki stosowanego nawozu).

W drugiej kolejności stosując GLM sprawdzimy czy:

3) H₀: Średnie plony uzyskane z uprawy poszczególnych zbóż są takie same gdy stosujemy różny sposób nawożenia.

Hipotezy 1) i 2)

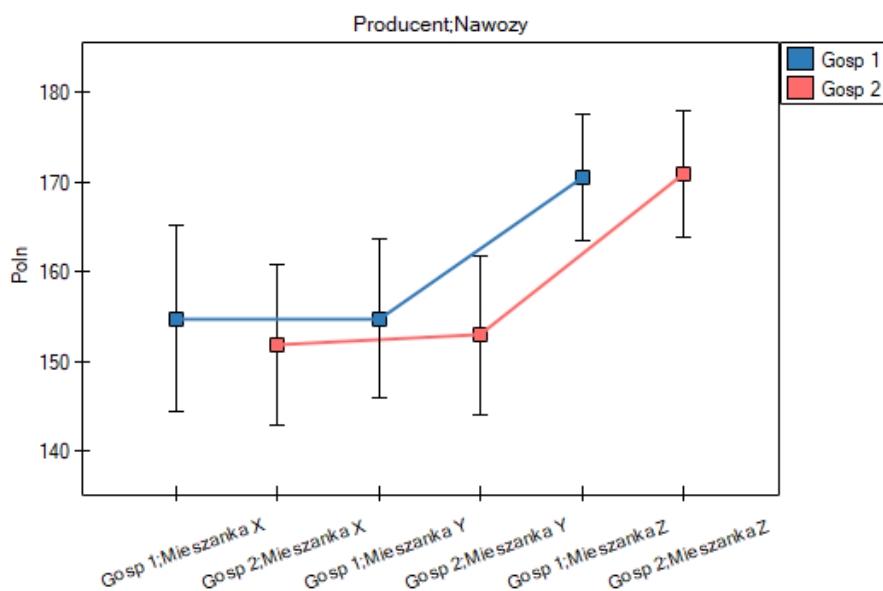
Podejście ANOVA

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					

Obserwujemy istotne statystycznie różnice pomiędzy plonem uzyskanym przy zastosowaniu różnych mieszanek nawozów ($p=0.000026$). Zastosowana mieszanka nawozów tłumaczy zmienność w uzyskanym plonie w około 12% o czym świadczy wartość cząstkowej Eta-kwadrat. Uzyskane plony nie zależą natomiast od tego w jakim gospodarstwie prowadzono uprawy ($p=0.667289$, Eta-kwadrat cząstkowe = 0,1%).

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. Dokładne wartości średnich możemy odczytać z tabeli statystyk opisowych.



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

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 plony 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 analizy 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

Podjęcie 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.

Poziomy czynnika Producent	2
Licznosc 0 [Gosp 1] REF	84
Licznosc 1 [Gosp 2]	84
Poziomy czynnika Nawozy	3
Licznosc 0 [Mieszanka X] REF	56
Licznosc 1 [Mieszanka Y]	56
Licznosc 2 [Mieszanka Z]	56

Model								
	wsp. b	bład b	-95% CI	+95% CI	stat. t	wartość p	stand. b	bład 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 (przypominam, że średnie wynosiły odpowiednio (153.285714 – dla mieszanki X, 170.696429 – dla mieszanki Z). Porównując gospodarstwa sprawa jest prosta, gdyż mamy do porównania tylko dwa gospodarstwa i uzyskany wynik jest wynikiem porównania gospodarstwa 2 z gospodarstwem 1, które to stanowiło kategorię odniesienia. Tym razem uzyskana różnica była niewielka (-1.464286) i nieistotna statystycznie (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	bład b	-95% CI	+95% CI	stat. t	wartość p	stand. b	bład 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:

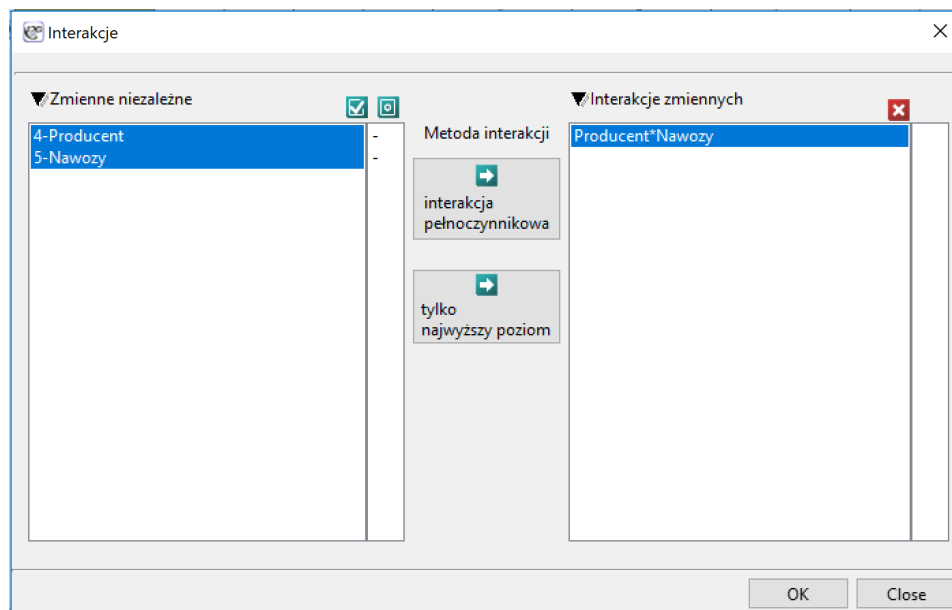
$$plon = 154.017857 + 0.571429 \cdot mieszankaY + 17.410714 \cdot mieszankaZ - 1.464286 \cdot gosp2$$

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.

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 tego 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.



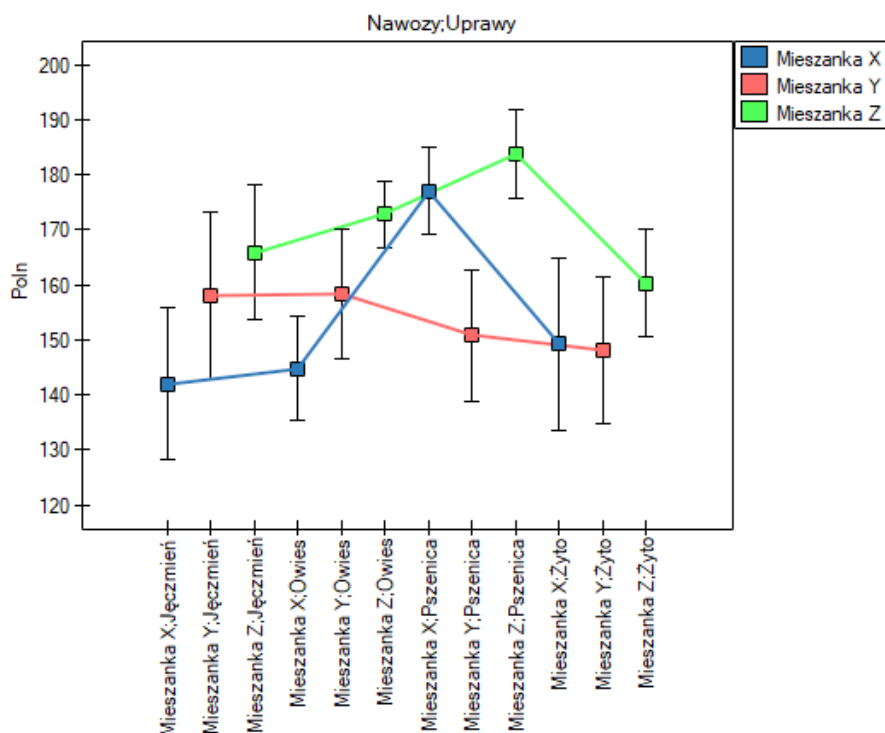
Test efektów międzygrupowych

	SS	MS	DF	F	p	Eta-kwadrat
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					

Uzyskany wynik potwierdził nasze przypuszczenia o braku istotnej interakcji (p=0,921377). 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.



Wykonamy analizę która oprócz efektów głównych uwzględni 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					

Ponieważ interakcje w zbudowanym modelu są istotne statystycznie ($p=0.003066$), to właśnie model z interakcjami powinniśmy stosować i opis uzyskanych wyników skupić właśnie na tej interakcji.

W podejściu ANOVA hipoteza odnosząca się do interakcji dotyczy wszystkich możliwych par średnich, tzn.:

H_0 : Średnie plony uzyskane przy nawożeniu pszenicy mieszanką X są takie same jak przy nawożeniu pszenicy mieszanką Y i takie same jak przy nawożeniu pszenicy mieszanką Z i takie same jak przy nawożeniu żyta mieszanką X i takie same jak przy nawożeniu żyta mieszanką Y i takie same jak przy nawożeniu żyta mieszanką Z i takie same jak przy nawożeniu owsa mieszanką X i takie same jak przy nawożeniu owsa mieszanką Y i takie same jak przy nawożeniu owsa mieszanką Z i takie same jak przy nawożeniu jęczmienia mieszanką X i takie same jak przy nawożeniu jęczmienia mieszanką Y i takie same jak przy nawożeniu jęczmienia mieszanką Z.

W podejściu regresyjnym powiemy, że:

H_0 : 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.

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

Natomiast mieszanka X i mieszanka Y uzyskują gorsze plony od mieszanki Z i dodatkowo zachodzi między nimi efekt interakcji. Przejawia się ona 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.

kodowanie zero - jedynkowe

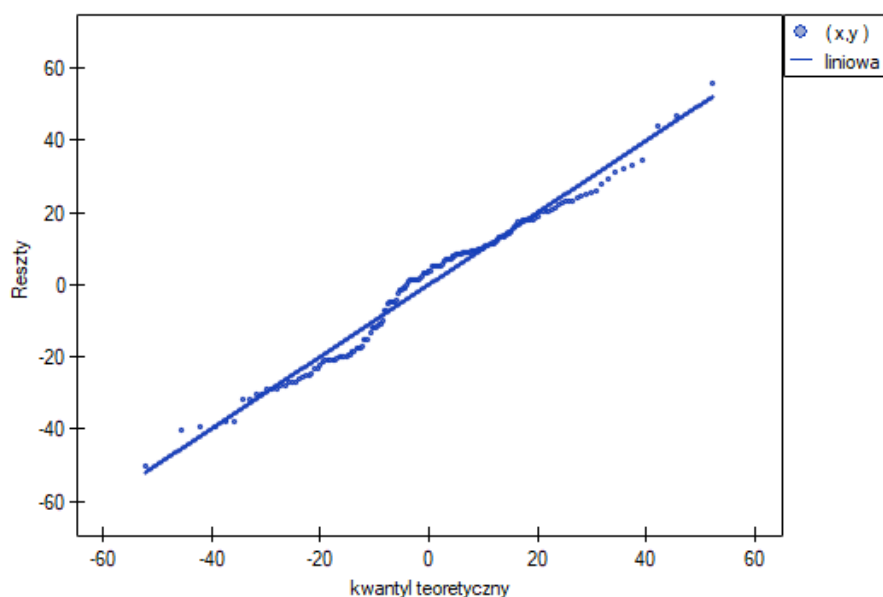
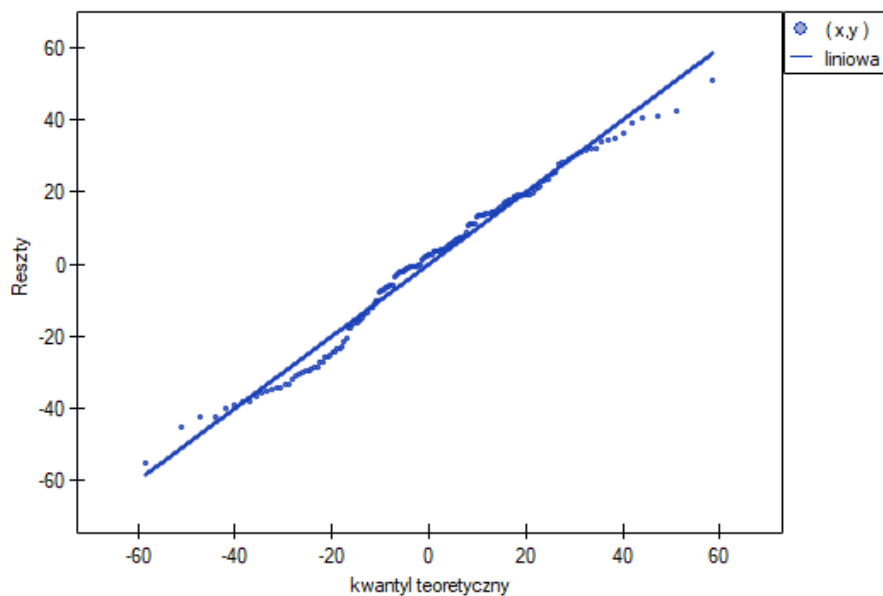
Model	wsp. b
w. wolny	142
Nawozy[1]	16
Nawozy[2]	23,9285
Uprawy[1]	2,78571
Uprawy[2]	35,0714
Uprawy[3]	7,28571
Nawozy[1]*Uprawy[1]	-2,3571
Nawozy[1]*Uprawy[2]	-42,285
Nawozy[1]*Uprawy[3]	-17,071
Nawozy[2]*Uprawy[1]	4,07142
Nawozy[2]*Uprawy[2]	-17,285
Nawozy[2]*Uprawy[3]	-12,857

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 wariacji oraz wizualnym określeniu normalności reszt modelu.



Wykres normalności reszt modelu typu Q-Q dla pierwszej oraz dla drugiej analizy przedstawia reszty modelu dobrze rozłożone wokół prostej, co świadczy o dobrym dopasowaniu reszt do rozkładu normalnego. Porównaniu wariacji 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 wariacji.

Jednorodność wariacji

	Test Levene'a		Browna-Forsythe'a	
	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

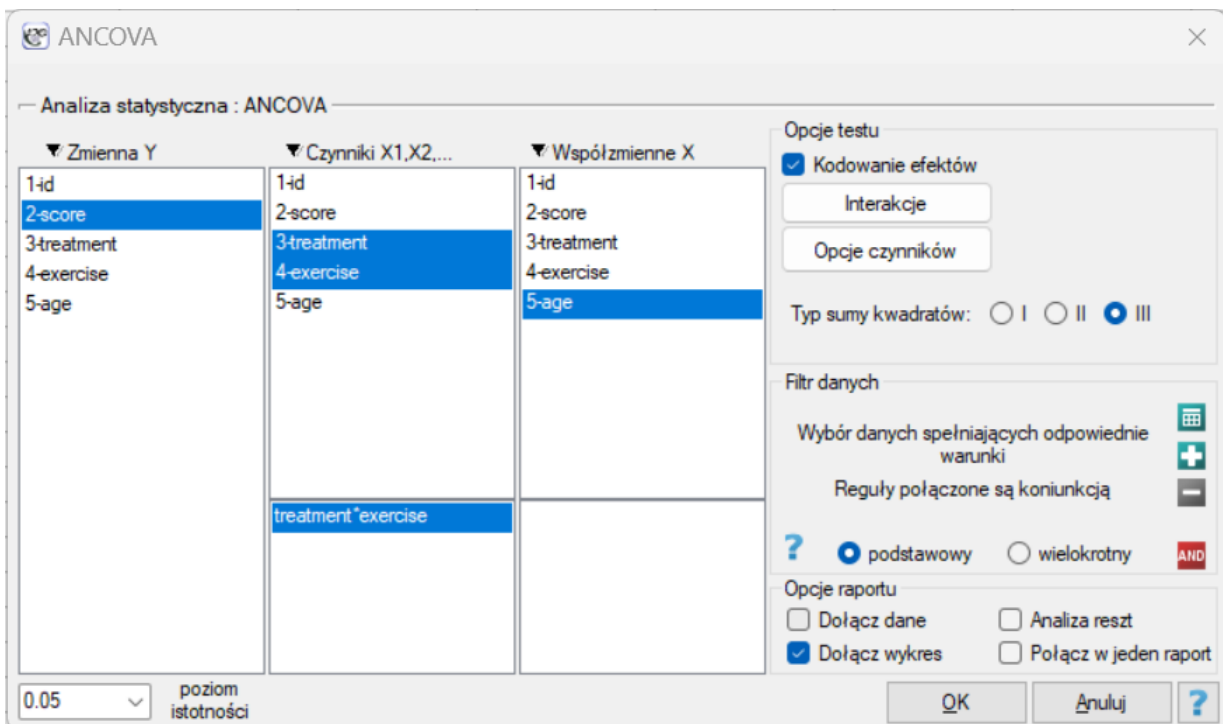
	Test Levene'a		Brown-Forsythe	
	Statystyka	Wartość p	Statystyka	Wartość p
Nawozy	2,954767	0,054854	3,223106	0,042352
Uprawy	2,383389	0,071266	2,298122	0,079466
Nawozy*Uprawy	3,266295	0,00048	1,754733	0,066416

23.7 ANCOVA

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 błędów, 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 kategoryjna (np. warunki leczenia) i zmienna ciągła (np. wiek) są zaangażowane w jedną analizę określającą wynik leczenia.

Okno z ustawieniami opcji ANCOVA wywołujemy poprzez menu Statystyki zaawansowane → Modele wielowymiarowe → ANCOVA



Uwaga!!!

Sposób uwzględniania badanych czynników i zmiennych wnikających opisany jest w rozdziale dotyczącym wieloczynnikowej ANOVA ([Wpływ czynników wnikających](#)). Polecanym sposobem jest wybór Sumy kwadratów typu III oraz [kodowania efektów](#)

Podstawowe 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 wnikającą a zmienną zależną) dla każdego możliwego poziomu czynnika.

Uwaga!

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 wnikającymi. Istotny statystycznie wynik oznacza złamanie założenia równych nachyleń, ponieważ istotna staje się interakcja, a więc różne nachylenia prostych.

Hipotezy ANCOVA dla pojedynczego czynnika A :

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \dots = \mu_a,$$

$$\mathcal{H}_1 : \text{nie wszystkie } \mu_i \text{ są sobie równe } (i = 1, 2, \dots, a),$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_a$ - oczekiwane średnie czynnika A dla poszczególnych jego kategorii.

Hipotezy ANCOVA dla interakcji czynników $A * B$:

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \dots = \mu_{ab},$$

$$\mathcal{H}_1 : \text{nie wszystkie } \mu_k \text{ są sobie równe } (k = 1, 2, \dots, ab),$$

gdzie:

$\mu_1, \mu_2, \dots, \mu_{ab}$ - oczekiwane średnie interakcji czynników $A * B$ dla poszczególnych ich kategorii.

PRZYKŁAD 23.8. (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 niezależna) czy cholesterol po leczeniu różnił się w zależności od dawki leku.

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

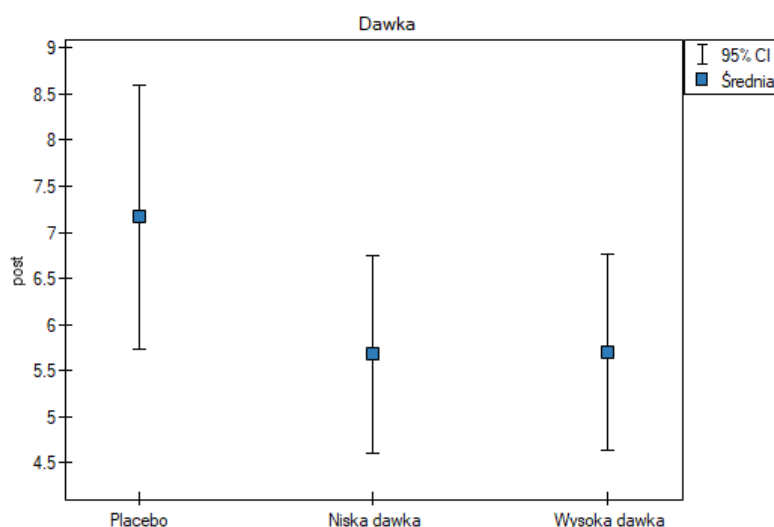
Niestety badacz nie uzyskał potwierdzenia różnic pomiędzy wynikami.

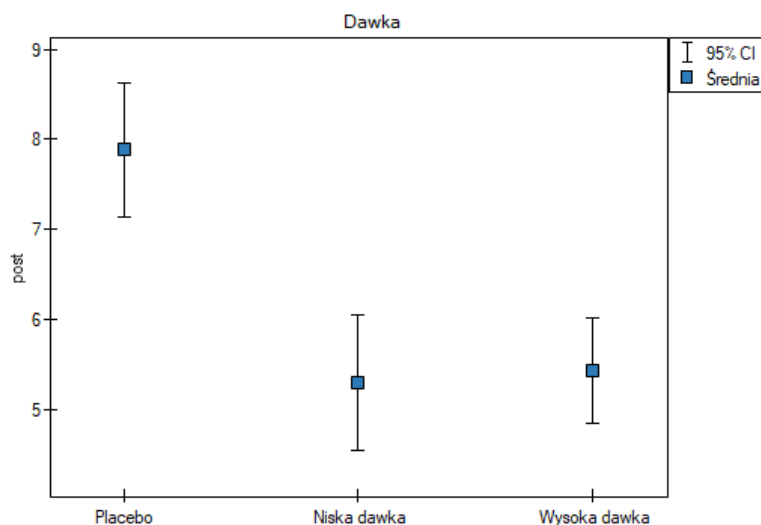
Wyobraźmy sobie, że badacz, zdał sobie sprawę, że to, czy dany lek zmieni poziom cholesterolu może być związane z wyjściowym poziomem cholesterolu oraz z wiekiem pacjenta. Z tego względu zdecydował się wykonać jednoczynnikową ANCOVA (czynnik to dawka leku) uwzględniającą jako współzmienną poziom cholesterolu przed leczeniem i wiek.

Tym razem wynik ANCOVA wskazał na występowanie istotnych różnic pomiędzy poziomem cholesterolu po zastosowaniu różnych dawek leku ($p=0.00003$):

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					

Uwzględnienie poziomu cholesterolu przed badaniem zmniejszyło uzyskiwane błędy dla średnich i zawężyło przedziały ufności. By wyświetlić obserwowane lub oczekiwane średnie, wybieram odpowiednie ustawienia poprzez Opcje czynników, do tego zaznaczam wykres błędów. Pierwszy wykres przedstawia obserwowane średnie wraz z przedziałem ufności, tzn. nie uwzględniające 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:





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 (wybrany 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.

POST-HOC Fisher LSD			
	NIR [F]	Statystyka [Wartość p [F]
Dawka			
Placebo;Niska dawka	1.021278	3.008137	0.005921
Placebo;Wysoka dawka	0.91139	3.31434	0.002804
Niska dawka;Wysoka dawka	0.944449	-0.054517	0.956957

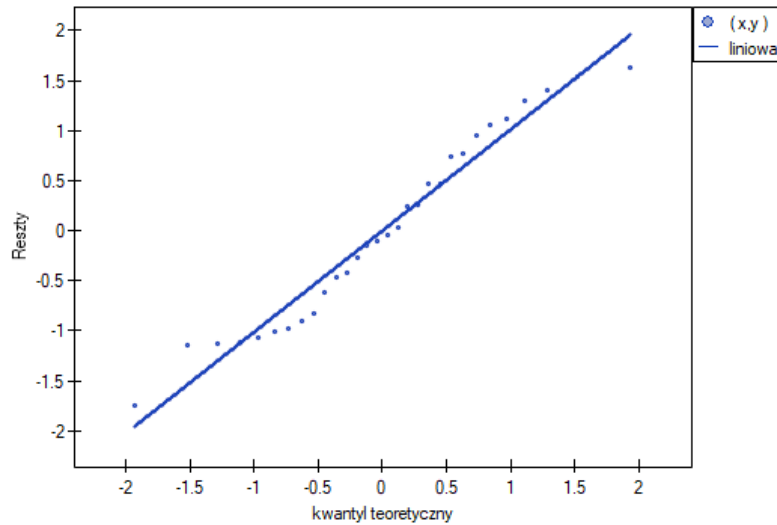
Grupy Jednorodne		
Dawka	A	B
Placebo(b)		*
Niska dawka(a)	*	
Wysoka dawka(a)	*	

Pozostały do sprawdzenia założenia ANCOVA. Jednorodność wariancji i stałość nachleń prostych regresji potwierdzono przy pomocy testów

Homogeniczność nachleń regresji	
DF1	4
DF2	21
p-value	0.80795

Jednorodność wariancji				
	Test Levene'a		Test Browna-Forsythe'a	
	Statystyka F	Wartość p	Statystyka F	Wartość p
Dawka	0.692964	0.508767	0.332599	0.719954

Normalność rozkładu reszty oceniono wizualnie wyrysowując wykresy Q-Q:



PRZYKŁAD 23.9. (plik stres.pqs)

Przykład pochodzi z pakietu Datarium R-Cran.

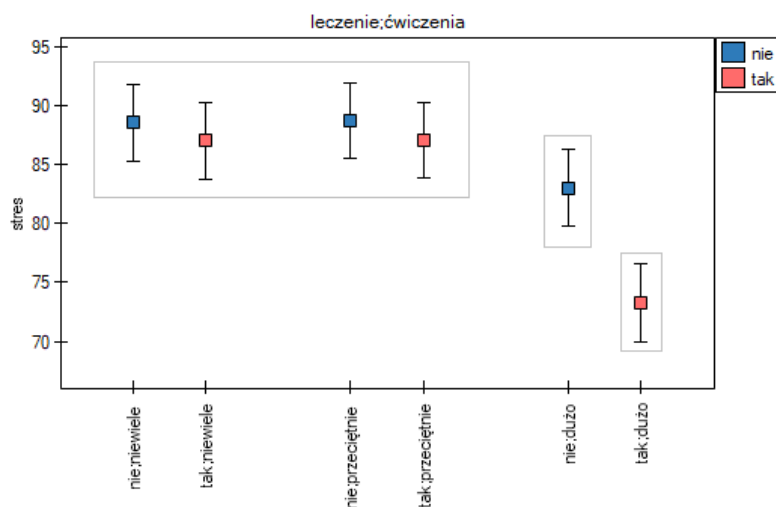
Badacze chcą ocenić wpływ nowego leczenia i ćwiczeń na redukcję stresu po uwzględnieniu różnic w wieku. Wartość pomiaru stresu to interwałowa zmienna wynikowa Y . Ze względu na to, że zmienne "leczenie" i "ćwiczenia" mają odpowiednio 2 i 3 kategorie, przeprowadzimy dwukierunkową ANCOVA w celu określenia, 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 ustawiam "stres", jako czynniki "leczenie" i "ćwiczenia" oraz dodaję 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					

Uzyskany wynik pokazuje, że wpływ leczenia na stres zmienia się w zależności od intensywności wykonywania ćwiczeń - wskazuje na to istotna interakcja obu tych zmiennych ($p=0.016409$). Wyrysujemy wykres przedstawiający oczekiwane średnie poziomy stresu dla każdej z sześciu podgrup, na jakie interakcja podzieliła nasze dane oraz wyznaczmy testy post-hoc.



leczenie;ćwiczenia	A	B	C
nie;niewiele(b)		*	
nie;przeciętnie(b)		*	
nie;duzo(c)			*
tak;niewiele(b)		*	
tak;przeciętnie(b)		*	
tak;duzo(a)	*		

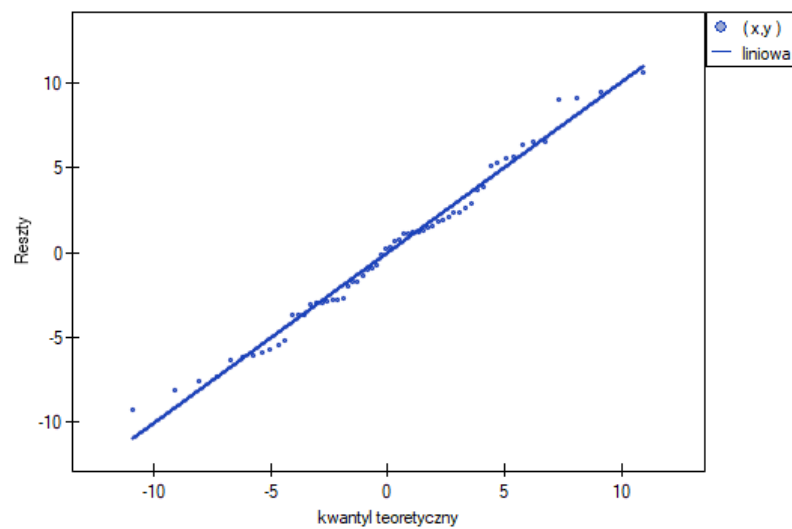
Zgodnie z wynikami testu post-hoc, możemy mówić o trzech różnych grupach jednorodnych: (B) grupa osób o wysokim poziomie stresu, to grupa ćwicząca niewiele lub przeciętnie (bez względu na to czy są to osoby leczone czy nie), (C) grupa osób o niższym poziomie stresu, to grupa ćwicząca dużo i nieleczona, (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
duzo	20	78.148878	1.19011	75.761819	80.535936
leczenie*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;duzo	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;duzo	10	73.275476	1.65137	69.963247	76.587704

Założenia dotyczące równości wariancji, nachyleń linii regresji oraz normalności reszt modelu są spełnione

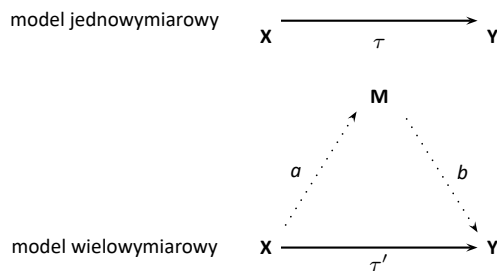
Homogeniczność nachleń regresji	
DF1	5
DF2	48
p-value	0.98578

	Test Levene'a		Test Browna-Forsythe'a	
	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



24 Efekt mediacji

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).



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

Test Sobela (1982)[154], test Aroiana (1947)[7] spopularyzowany przez Barona i Kennyego [16] oraz test Goodmana (1960)[69] 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:

$$\mathcal{H}_0 : \tau = \tau'$$

$$\mathcal{H}_1 : \tau \neq \tau',$$

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 :

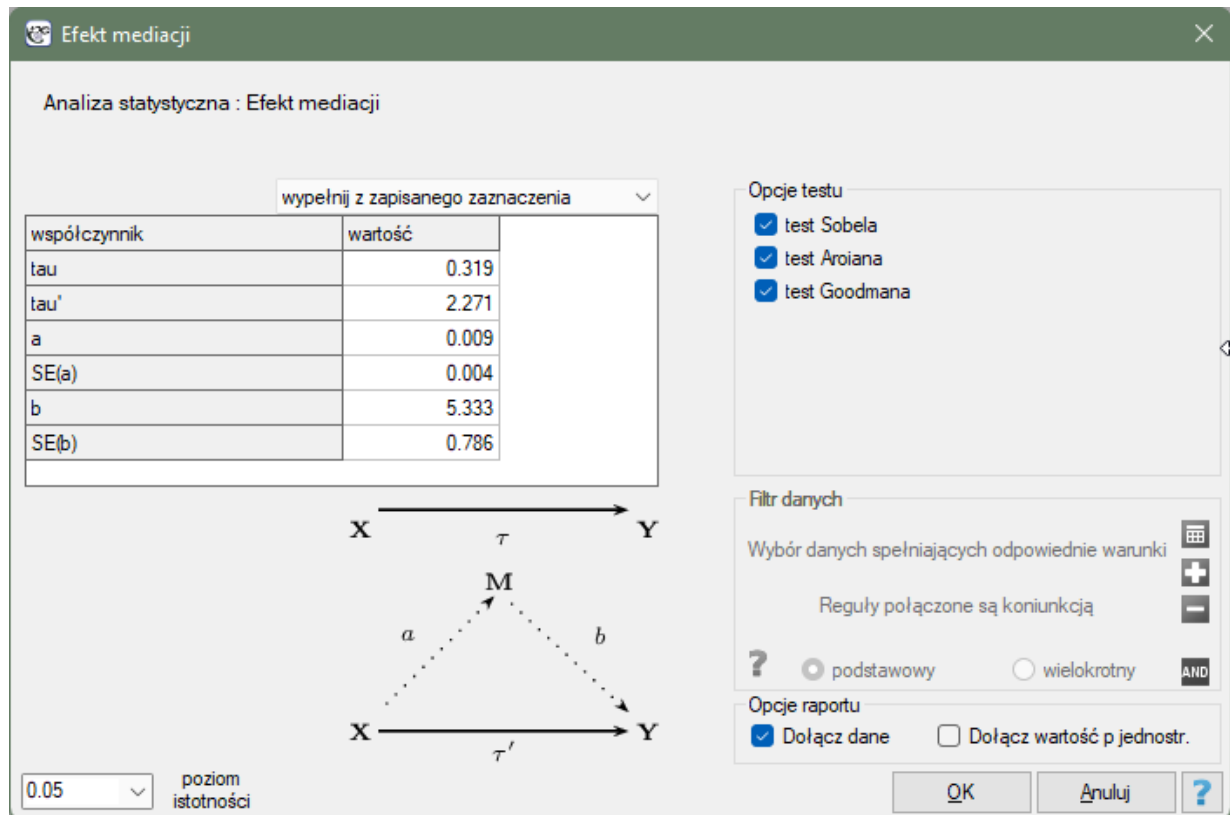
jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

Test Sobela, jak też test Aroiana i Goodmana, są to testy bardzo konserwatywne i przeznaczone jedynie dla dużych prób (przekraczających 100 elementów).

Okno Analizy efektu mediacji wywołujemy poprzez:

Statystyki zaawansowane \rightarrow Modele wielowymiarowe \rightarrow Efekt mediacji.



Analiza statystyczna : Efekt mediacji

wypełnij z zapisanego zaznaczenia

współczynnik	wartość
tau	0.319
tau'	2.271
a	0.009
SE(a)	0.004
b	5.333
SE(b)	0.786

Opcje testu

- test Sobela
- test Aroiana
- test Goodmana

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

? podstawowy wielokrotny AND

Opcje raportu

- Dołącz dane
- Dołącz wartość p jednostr.

0.05 poziom istotności

OK Anuluj ?

The path diagram shows a mediation model with variables X, M, and Y. A solid arrow points from X to Y with coefficient τ . A solid arrow points from X to M with coefficient a . A solid arrow points from M to Y with coefficient b . A solid arrow points from X to Y with coefficient τ' .

PRZYKŁAD 24.1. Na podstawie pracy Mimar Sinan Fine (2017) [121].

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 częstotliwości 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 iż **b=5.333**, a błąd $SE_b=0.786$

Różnica między współczynnikami to $\tau - \tau' = a \cdot b = 0.048$. 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 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		Dane :	
tau-tau'	-1.952	tau [Y=tau*X+c]	0.319
Efekt mediacji	-6.119122	tau' [Y=tau'*X+c]	2.271
Test Sobela		a [M=a*X+c]	0.009
Statystyka Z	2.135636	SE(a)	0.004
Wartość p dwustronne	0.032709	b [Y=tau'*X+b*]	5.333
Test Aroiana		SE(b)	0.786
Statystyka Z	2.115041		
Wartość p dwustronne	0.034426		
Test Goodmana			
Statystyka Z	2.156845		
Wartość p dwustronne	0.031018		

Na podstawie współczynników a i b oraz ich błędów standardowych wyznaczony zostaje wynik testów Sobela ($p = 0.0327$), Aroiana ($p = 0.0344$) i Goodmana ($p = 0.0310$). Uzyskane wartości p wskazują na istotne statystycznie znaczenie mediatora. Potwierdziliśmy więc, że częstość spożywania alkoholu wpływa na związek wieku z ciśnieniem rozkurczowym na tyle zauważalnie, że warto wyjaśnić dlaczego ten efekt występuje.

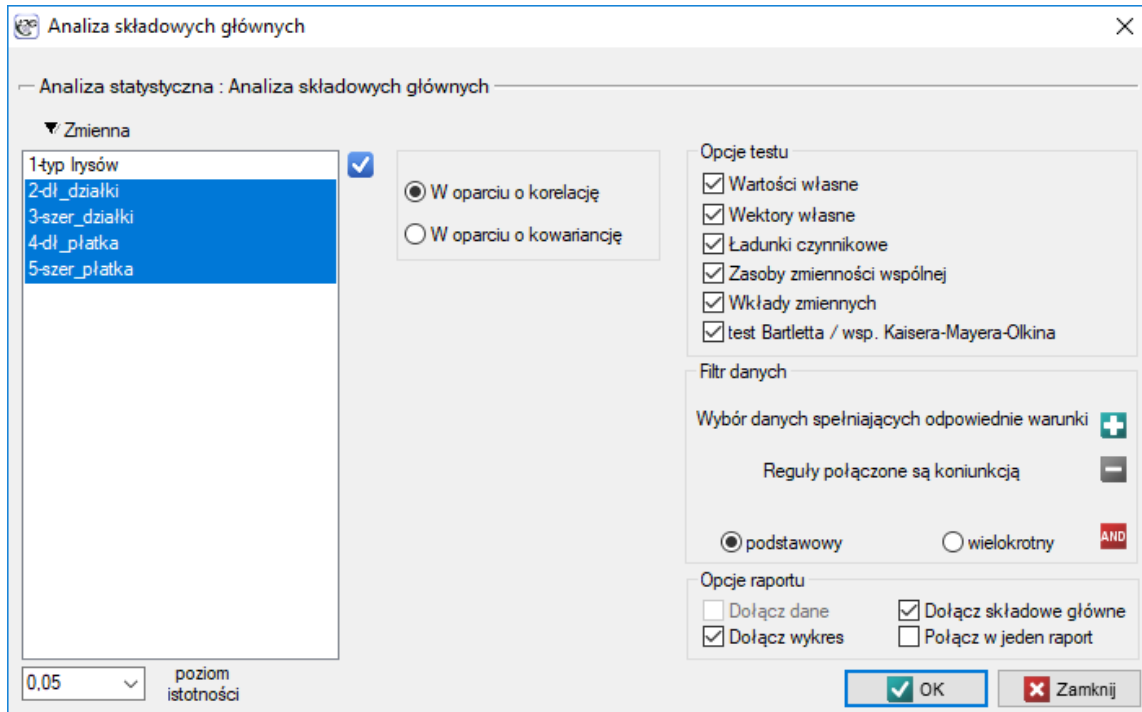
25 REDUKCJA WYMIARÓW I GRUPOWANIE

Wraz ze wzrostem liczby zmiennych poddawanych analizie statystycznej rośnie ich dokładność, ale rośnie też stopień komplikacji i trudności w interpretacji uzyskanych wyników. Zbyt duża liczba zmiennych niesie za sobą ryzyko wzajemnej ich korelacji. A więc informacja wnoszona przez niektóre zmienne może być redundantna, tzn. część zmiennych może nie wносить nowej informacji do analizy, ale powtarzać informacje wniesione już przez inne zmienne. Potrzeba redukcji wymiarowości (redukcji liczby zmiennych) zrodziła całą grupę analiz poświęconej tej kwestii takich jak: analiza czynnikowa, analiza składowych głównych, analiza skupień, czy analiza dyskryminacyjna. Metody te pozwalają na wykrywanie powiązań między zmiennymi. Na podstawie tych powiązań można wyodrębnić grupy podobnych zmiennych, a do dalszych analiz wybrać tylko jednego przedstawiciela każdej grupy (jedną zmienną), albo nową zmienną której wartości są wyliczane w oparciu o pozostałe zmienne w grupie. W ten sposób mamy pewność, że informacje niesione przez każdą grupę są w analizie uwzględnione. Przy niewielkiej utracie informacji możemy więc zbiór p zmiennych zredukować do zbioru k zmiennych gdzie $k < p$

Podobnie jak grupowaniem zmiennych możemy być zainteresowani grupowaniem obiektów. Mając do dyspozycji informację o pewnych cechach obiektów jesteśmy często w stanie wyodrębnić grupy obiektów podobnych pod względem tych cech np. na podstawie informacji o kwotach przeznaczanych przez klientów pewnej sieci sklepów na poszczególne artykuły, można dokonać podziału klientów w ten sposób, by wyodrębnić segmenty klientów o podobnych preferencjach zakupowych. W konsekwencji, swoją ofertę/reklamę można przygotować nie dla ogółu kupujących, ale oddzielnie dla każdego segmentu, tak by precyzyjniej trafić w potrzeby potencjalnego klienta.

25.1 ANALIZA SKŁADOWYCH GŁÓWNYCH

Okno z ustawieniami opcji Analizy składowych głównych wywołujemy poprzez menu Statystyki → Modele wielowymiarowe → Analiza składowych głównych.



Analiza składowych głównych polega na wyznaczeniu zupełnie nowych zmiennych (**składowych głównych**) będących kombinacją liniową zmiennych obserwowanych (pierwotnych). Dokładna analiza składowych głównych umożliwia wskazanie tych zmiennych pierwotnych, które mają duży wpływ na wygląd poszczególnych składowych głównych czyli tych, które tworzą grupę jednorodną. Składowa główna jest wówczas reprezentantem tej grupy. Kolejne składowe są wzajemnie ortogonalne (nieskorelowane) a ich liczba (k) jest mniejsza lub równa liczbie zmiennych pierwotnych (p).

Poszczególne składowe główne są kombinacją liniową zmiennych pierwotnych:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

gdzie:

X_1, X_2, \dots, X_p – zmienne pierwotne,
 $a_{i1}, a_{i2}, \dots, a_{ip}$ – współczynniki i -tej składowej głównej

Każda składowa główna wyjaśnia pewną część zmienności zmiennych pierwotnych. W naturalny sposób opiera się więc na takich miarach zmienności jak kowariancja (gdy zmienne pierwotne są podobnej wielkości i są wyrażone w tych samych jednostkach) lub korelacja (gdy założenia potrzebne do stosowania kowariancji nie są spełnione).

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,

$a_i = (a_{i1}, a_{i2}, \dots, a_{ip})$ – wektor własny odpowiadający i -tej wartości własnej,

M – macierz wariancji lub kowariancji zmiennych pierwotnych X_1, X_2, \dots, X_p ,

I – macierz identycznościowa (1 na głównej przekątnej, 0 poza nią).

25.1.1 Interpretacja współczynników związanych z analizą

Każda składowa główna opisana jest poprzez:

Wartość własną

Wartość własna daje informacje o tym, jaka część całkowitej zmienności jest tłumaczona przez daną składową główną. Pierwsza składowa główna tłumaczy największą część wariancji, druga składowa tłumaczy największą część tej wariancji, która nie została wytłumaczona przez poprzednią składową, kolejna składowa tłumaczy największą część tej wariancji, która nie została wytłumaczona przez poprzednie składowe. W rezultacie każda kolejna składowa główna tłumaczy coraz mniejszą część wariancji, czyli kolejne wartości własne są coraz mniejsze.

Całkowita wariancja jest sumą wartości własnych, co pozwala, dla każdej składowej wyliczyć procent zmienności przez nią definiowany

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \cdot 100\%$$

W rezultacie dla kolejnych składowych można wyliczyć również, skumulowaną zmienność i skumulowany procent zmienności.

Wektor własny

Wektor własny odzwierciedla wpływ poszczególnych zmiennych pierwotnych na daną składową główną. Zawiera współczynniki $a_{i1}, a_{i2}, \dots, a_{ip}$ kombinacji liniowej wyznaczającej składową. Przy czym znak tych współczynników wskazuje kierunek wpływu i jest przypadkowy co nie zmienia wartości niesionej informacji.

Ładunki czynnikowe

Ładunki czynnikowe, podobnie jak współczynniki zawarte w wektorze własnym, odzwierciedlają wpływ poszczególnych zmiennych na daną składową główną. Są to wartości obrazujące jaką część wariancji danej składowej stanowi zmienne pierwotne. Gdy analiza oparta jest na macierzy korelacji wartości te interpretujemy 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

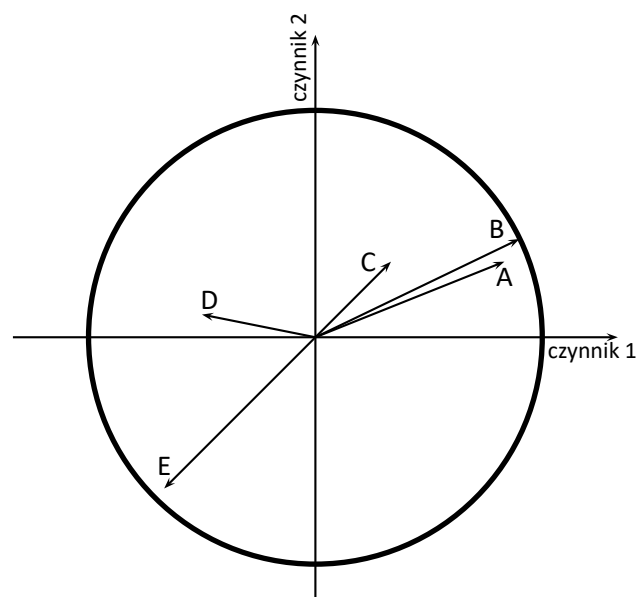
Bazują na współczynnikach determinacji pomiędzy zmiennymi pierwotnymi a daną składową główną. Wskazują jaki procent zmienności danej zmiennej pierwotnej może być tłumaczony zmiennością pierwszych kilku składowych głównych. Na przykład: wynik dotyczący drugiej zmiennej zawarty w kolumnie dotyczącej czwartej składowej głównej mówi o tym, jaki procent zmienności drugiej zmiennej może być tłumaczony zmiennością pierwszych czterech składowych głównych.

25.1.2 Interpretacja graficzna

Wiele informacji, które niosą współczynniki zwracane w tabelach można przedstawić na jednym wykresie. Umiejętność czytania wykresów pozwala na szybką interpretację wielu aspektów przeprowadzonej analizy. Wykresy zbierają w jednym miejscu informację dotyczącą wzajemnych relacji pomiędzy składowymi, zmiennymi pierwotnymi i przypadkami. Dają całościowy obraz analizy składowych głównych, przez co są bardzo dobrym jej podsumowaniem.

Wykres ładunków czynnikowych

Wykres przedstawia wektory połączone z początkiem układu współrzędnych, które to reprezentują zmienne pierwotne. Wektory te są 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 wkład zmiennej pierwotnej w budowę składowych większy. W przypadku analizy opartej na macierzy korelacji ładunki są korelacjami pomiędzy zmiennymi pierwotnymi a składowymi, wówczas punkty wpadają do koła jednostkowego. Dzieje się tak dlatego, że współczynnik korelacji nie może przekroczyć jedynki. W rezultacie, im dana zmienna pierwotna leży bliżej brzegu tego koła, tym lepsza jest jej reprezentacja przez przedstawione główne składowe.

Znak współrzędnych końca wektora czyli znak ładunku czynnikowego - wskazuje na dodatnią lub ujemną korelację zmiennej pierwotnej i składowych głównych tworzących układ współrzędnych. Jeżeli rozpatrujemy łącznie obie osie (2 składowe), wówczas zmienne pierwotne mogą być kwalifikowane do jednej z czterech kategorii, zależnie od kombinacji znaków (+/-) ich ładunków czynnikowych.

Kąt między wektorami wskazuje na skorelowanie zmiennych pierwotnych:

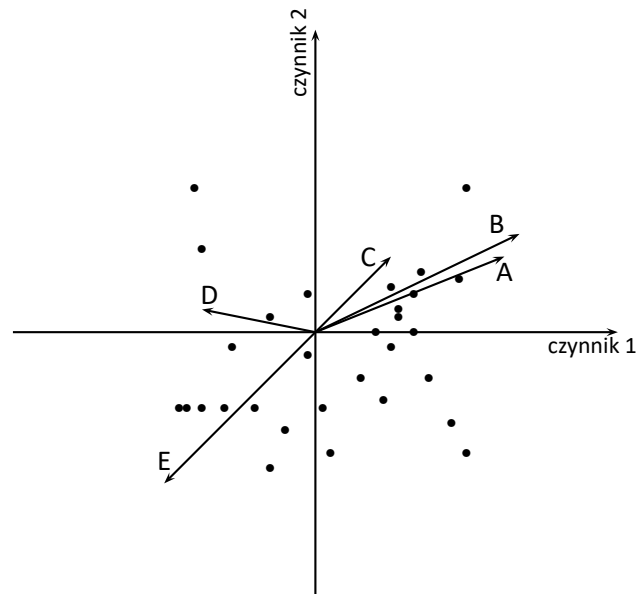
$0 < \alpha < 90^0$: im kąt pomiędzy wektorami reprezentującymi zmienne pierwotne jest mniejszy tym silniejsza jest dodatnia korelacja pomiędzy tymi zmiennymi.

$\alpha = 90^0$ - wektory te są prostopadłe, czyli zmienne pierwotne nie są skorelowane.

$90^{\circ} < \alpha < 180^{\circ}$ - im kąt pomiędzy wektorami reprezentującymi zmienne pierwotne jest większy, tym silniejsza jest ujemna korelacja pomiędzy tymi zmiennymi.

Biplot

Wykres przedstawia 2 serie danych umieszczone w układzie współrzędnych wyznaczonych przez 2 składowe główne. Serię pierwszą na wykresie stanowią dane z wykresu pierwszego (czyli wektory zmiennych pierwotnych) a serię drugą punkty przedstawiające poszczególne przypadki.



Współrzędne punktów powinny być interpretowane 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 punkty znajdują się od średniej. Przy czym, jeśli na wykresie znajdują się obserwacje nietypowe - odstające, to mogą one zaburzać analizę i powinny być usunięte a analiza przeprowadzona ponownie.

Odległości między punktami wskazują na podobieństwo przypadków - im bliżej (w znaczeniu odległości euklidesowej) zlokalizowane są względem siebie punkty, tym bardziej podobne informacje niosą porównywane przypadki.

Rzuty prostopadłe punktów na wektory interpretujemy tak samo jak współrzędne punktów, czyli rzuty na osie z tym, że interpretacja dotyczy nie składowych głównych a zmiennych pierwotnych. Wartości umieszczone po stronie końca wektora są większe od średniej wartości zmiennej pierwotnej a wartości umieszczone na przedłużeniu wektora ale w kierunku przeciwnym są wartościami mniejszymi od średniej.

25.1.3 Kryteria redukcji wymiarów

Nie istnieje jedno uniwersalne kryterium wyboru liczby składowych głównych. Dobrze jest więc, aby przy wyborze kierować się kilkoma metodami.

Procent wyjaśnionej wariancji

Liczba składowych głównych, jaką badacz powinien przyjąć zależy od tego, w jaki stopniu reprezentują one zmienne pierwotne, czyli zawartej w nich wariancji zmiennych pierwotnych. Wszystkie składowe główne niosą 100% wariancji zmiennych pierwotnych. Jeśli suma wariancji dla kilku

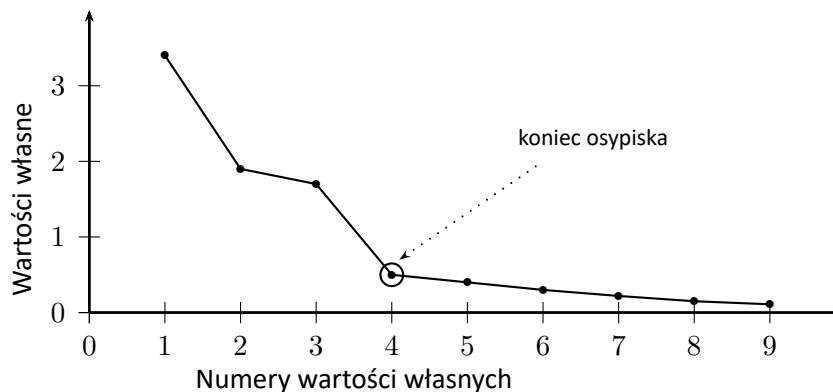
pierwszych składowych stanowi znaczną część całkowitej wariancji zmiennych pierwotnych, wówczas te składowe główne mogą w zadowalającym stopniu zastąpić zmienne pierwotne. Przyjmuje się, że wariancja ta powinna zostać odzwierciedlona w składowych głównych w ponad 80 procentach.

Kryterium Kaisera

Kryterium Kaisera mówi o tym, że składowe główne, które chcemy pozostawić do interpretacji powinny mieć przynajmniej taką samą wariancję jak dowolna wystandaryzowana zmienna pierwotna. W związku z tym, że wariancja każdej wystandaryzowanej zmiennej pierwotnej wynosi 1, to według kryterium Kaisera ważne są tylko składowe główne, których wartość własna przekracza lub jest bliska wartości 1.

Wykres osypiska

Na wykresie tym przedstawione jest tempo spadku wartości własnych, czyli procentu wyjaśnionej wariancji.



Moment na wykresie, w którym proces ten stabilizuje się i linia malejąca przechodzi w poziomą to tzw. koniec osypiska (koniec osypywania się informacji o zmiennych pierwotnych, jaką niosą składowe główne). Składowe znajdujące się na prawo od punktu kończącego osypisko reprezentują znikomą wariancję i przedstawiają w większości losowy szum.

25.1.4 Wyznaczenie składowych głównych

Gdy zdecydujemy ilu składowych głównych będziemy potrzebować, możemy przystąpić do ich wygenerowania. W przypadku składowych głównych powstałych w oparciu o macierz korelacji składowe główne są wyliczane jako kombinacja liniowa wystandaryzowanych zmiennych pierwotnych. Jeśli natomiast składowe główne powstały w oparciu o macierz kowariancji, wówczas są wyliczane jako kombinacja liniowa wycentrowanych względem średniej zmiennych pierwotnych.

Otrzymane w ten sposób składowe główne stanowią nowe zmienne o pewnych zaletach. Przede wszystkim zmienne te nie są współliniowe. Zwykle jest ich mniej niż zmiennych pierwotnych, czasem znacznie mniej, a niosą one tyle samo lub nieznacznie mniej informacji niż zmienne pierwotne. Są więc zmiennymi, które z powodzeniem mogą być wykorzystane w większości analiz wielowymiarowych.

25.1.5 Zasadność stosowania Analizy składowych głównych

Jeśli zmienne nie są skorelowane (współczynnik korelacji Pearsona jest bliski 0), to wówczas przeprowadzanie analizy składowych głównych nie ma sensu. W takiej sytuacji bowiem każda zmienna stanowi już odrębną składową.

Test Bartletta

Test ten wykorzystywany jest do weryfikacji hipotezy o tym, że współczynniki korelacji pomiędzy zmiennymi są zerowe (czyli macierz korelacji jest macierzą jednostkową).

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: M = I, \\ \mathcal{H}_1 &: M \neq I.\end{aligned}$$

gdzie:

M – macierz wariancji lub kowariancji zmiennych pierwotnych X_1, X_2, \dots, X_p ,

I – macierz jednostkowa (1 na głównej przekątnej, 0 poza nią).

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 – licznosc (liczba przypadków),

λ_i – i -ta wartość własna.

Statystyka ta ma asymptotycznie (dla dużych licznosci oczekiwanych) rozkład χ^2 z $p(p - 1)/2$ stopniami swobody.

Wyznaczoną na podstawie [statystyki testowej](#) wartość p porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Współczynnik Kaisera-Mayera-Olkina

Współczynnik ten wykorzystywany jest do sprawdzania stopnia skorelowania zmiennych pierwotnych, czyli siły dowodów świadczących na rzecz sensowności przeprowadzania analizy składowych głównych.

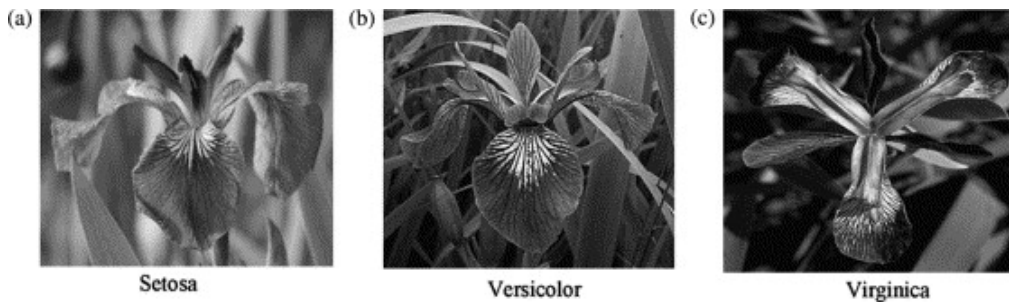
$$KMO = \frac{\sum_{i \neq j}^p \sum_{j \neq i}^p r_{ij}^2}{\sum_{i \neq j}^p \sum_{j \neq i}^p r_{ij}^2 + \sum_{i \neq j}^p \sum_{j \neq i}^p \hat{r}_{ij}^2},$$

r_{ij} – współczynnik korelacji pomiędzy i -tą a j -tą zmienną,

\hat{r}_{ij} – współczynnik korelacji cząstkowej pomiędzy i -tą a j -tą zmienną.

Wartość współczynnika Kaisera należy do przedziału $< 0, 1 >$, gdzie wartości niskie świadczą o braku podstaw do przeprowadzania analizy składowych głównych, a wartości wysokie są przesłanką do przeprowadzania tej analizy.

PRZYKŁAD 25.1. (plik: iris.pqs) Ten klasyczny zestaw danych pierwotnie ukazał się w pracy R.A. Fishera 1936[58], 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 na wskazanie tych pomiarów (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.

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_pł
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

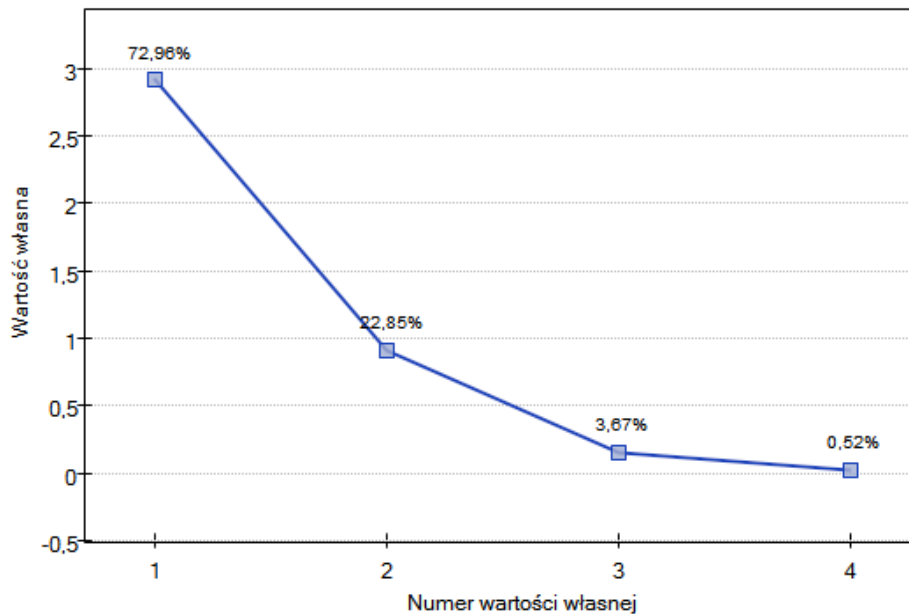
Wartość p statystyki Bartletta wskazuje na prawdziwość hipotezy mówiącej o istotnej różnicy pomiędzy uzyskaną macierzą korelacji a macierzą jednostkową, czyli znacznym skorelowaniu zmiennych. Uzyskany współczynnik KMO jest natomiast przeciętny i wynosi 0.54. Wskazania do przeprowadzenia analizy składowych głównych uznajemy za wystarczające.

Pierwszym wynikiem tej analizy, na który należy zwrócić szczególną uwagę są wartości własne:

Wartości własne				
Numer	Wartość	% warian	Wartość kumul.	% 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 lub nawet jedna składowa główna w dobry sposób opisują 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 w interpretacji jest tylko jedna składowa główna, gdyż tylko dla pierwszej składowej wartość własna jest większa niż 1. Patrząc jednak na wykres osypiska można wysnuć wniosek, że linia spadkowa przechodzi w poziomą dopiero od 3 składowej głównej.



Zgodnie z tym wskazaniem dwie pierwsze składowe wnoszą istotną informację. Łącznie dwie pierwsze składowe wyjaśniają sporo bo 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, to zmienna szerokość działki zostałaby odzwierciedlona w niewielkiej części.

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 i dla każdej z analizowanych zmiennych przekraczają 90%, czyli wariancja każdej zmiennej jest reprezentowana przy użyciu tych 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

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. Ułożmy je kolejno według wielkości tego wpływu:

1. Długość płatka jest ujemnie skorelowana z pierwszą składową, a zatem czym 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 najslabiej skorelowana z pierwszą składową i 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 jest w składowej drugiej najwyższy.

Każda składowa główna wyznacza homogeniczną grupę zmiennych pierwotnych. Pierwszą składową nazwiemy "rozmiar płatka", gdyż najbardziej znaczącymi dla niej zmiennymi są te zmienne, które niosą informacje o płatku, choć trzeba zaznaczyć, że długość działki również wpływa znacznie na wartość tej składowej. W interpretacji pamiętamy, że czym 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:

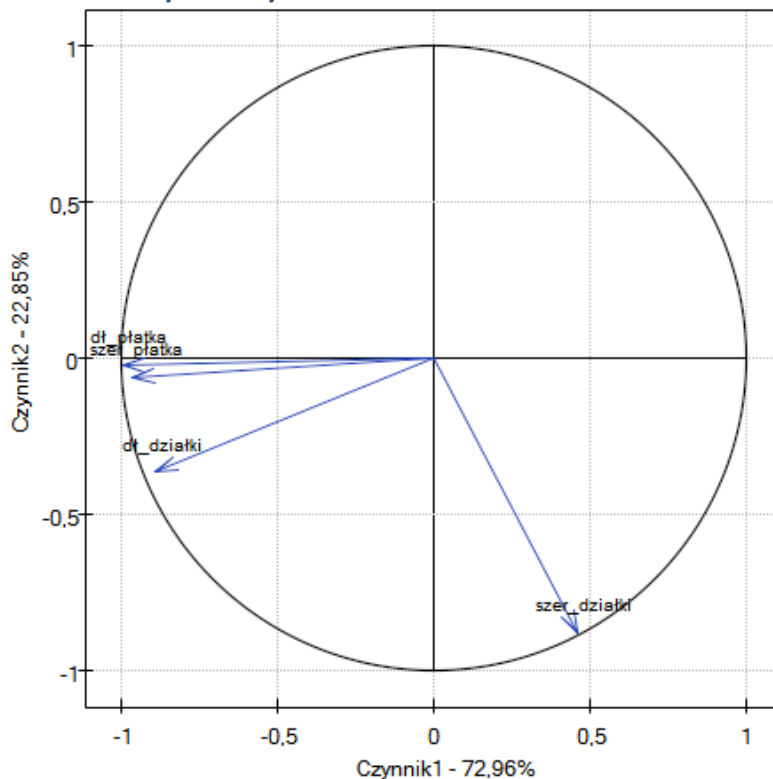
Składowe główne			
Czynnik1	Czynnik2	Czynnik3	Czynnik4
2,257141	-0,47842	0,12728	0,024088
2,074013	0,671883	0,233826	0,102663
2,356335	0,340766	-0,04405	0,028282
2,291707	0,5954	-0,09098	-0,06573
2,381863	-0,64467	-0,01568	-0,03580
2,068701	-1,48420	-0,02687	0,006586
2,435868	-0,04748	-0,33435	-0,03665
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 wcześniejszych czterech 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% wariacji i 22.85% wariacji co daje łącznie 95.81% wariacji zmiennych pierwotnych

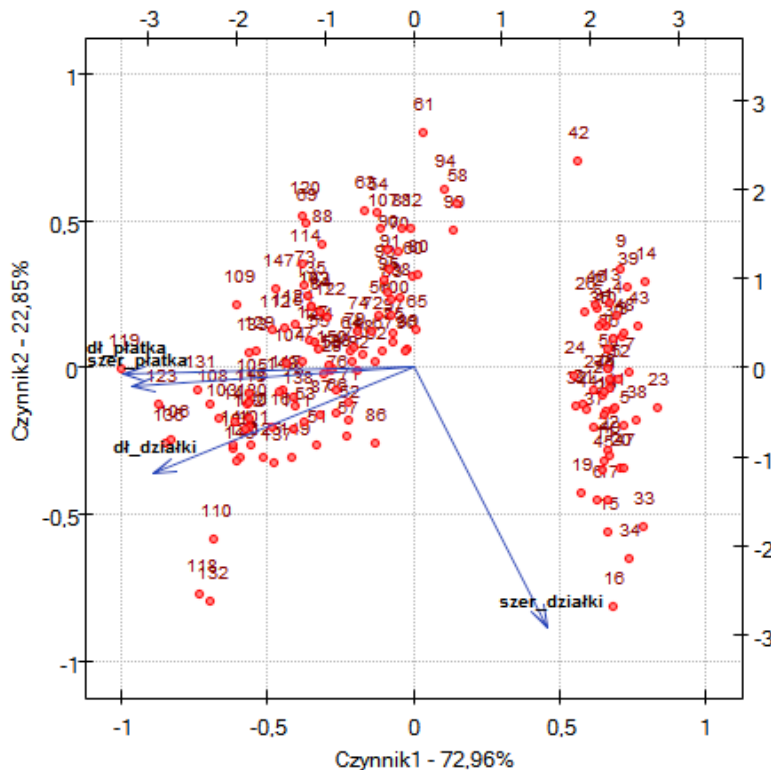
Wektory reprezentujące zmienne pierwotne niemalże sięgają brzegów koła jednostkowego (koła o promieniu 1), a zatem wszystkie one są bardzo dobrze reprezentowane przez dwie pierwsze

składowe główne tworzące układ współrzędnych.

Kąt pomiędzy wektorami obrazującymi długość płatków, szerokość płatków 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, która jest reprezentowana 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 pierwotnych, który jest bliski kątowni 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 reprezentowana jest dobrze przez drugą składową.

Biplot



Biplot przedstawia dwie serie danych rozpięte na dwóch pierwszych składowych. Jedną serią są wektory zmiennych pierwotnych, które przedstawione były na poprzednim wykresie, a drugą serią są 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 punktu 33 i 34 jest niewielka. Oba punkty mają sporo większą niż przeciętna wartość pierwszej i sporo mniejszą niż przeciętna wartość drugiej składowej. Wartość przeciętna, a zatem średnia arytmetyczna obu składowych wynosi 0, więc jest to środek układu współrzędnych. Pamiętając, że pierwsza składowa to głównie rozmiar płatków a druga to 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 punktów. Jest to kwiat o ujemnej pierwszej składowej i dodatniej lecz nie wysokiej drugiej składowej. Zatem jest to 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).

25.2 ANALIZA SKUPIEŃ

Analiza skupień to szereg metod dzielenia obiektów, bądź cech (zmiennych) na podobne grupy. Zasadniczo metody te dzieli się na dwie klasy: metody hierarchiczne i niehierarchiczne jak metoda k-średnich. W swoich algorytmach obie metody wykorzystują macierz podobieństwa, by na jego podstawie tworzyć skupienia.

Grupowanie obiektów i grupowanie zmiennych odbywa się w analizie skupień dokładnie w ten sam sposób. W rozdziale tym metody grupowania opisane zostaną na przykładzie grupowania obiektów.

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.

25.2.1 Metody hierarchiczne

Metody hierarchicznej analizy skupień polegają na budowaniu hierarchii skupień poczynając od tych najmniejszych (złożonych z pojedynczych obiektów), a kończąc na tych największych (złożonych z maksymalnej liczby obiektów). Skupienia tworzone są na bazie macierzy podobieństwa obiektów.

PROCEDURA AGLOMERACYJNA

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.

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.

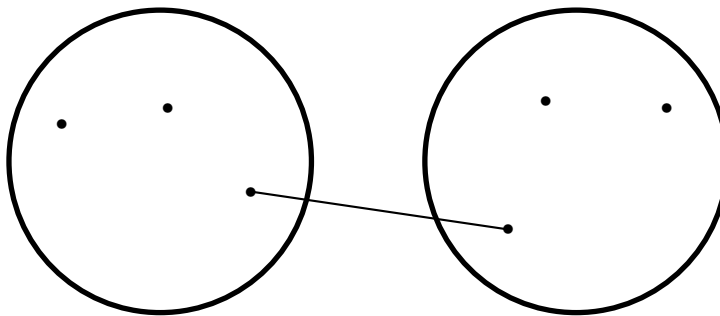
Uwaga!

By zwiększyć wpływ wybranych zmiennych na elementy macierzy podobieństwa, należy wskazać odpowiednie wagi przy ustawianiu sposobu definiowania odległości pamiętając jednocześnie o wystandaryzowaniu danych.

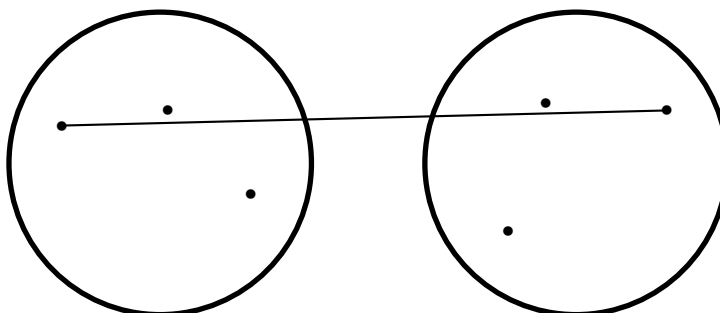
Np. Dla osób chcących zaopiekować się psem pogrupowanie psów zgodnie z wielkością, umaszczeniem, długością ogona, charakterem, rasą itp. ułatwi dokonanie wyboru. Jednak, identyczne traktowanie wszystkich cech może spowodować umieszczenie zupełnie niepodobnych psów w jednej grupie. Natomiast dla większości 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.

Metody wiązania obiektów i skupień

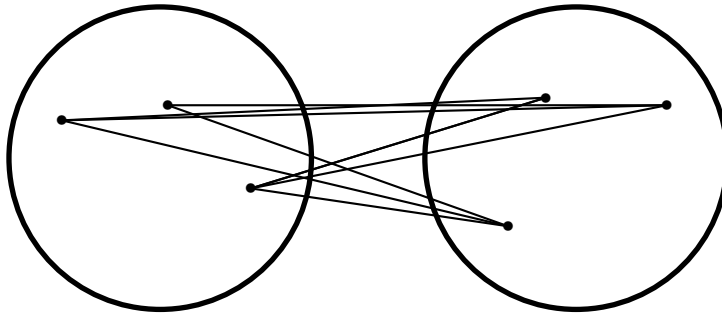
- Metoda pojedynczego wiązania (najbliższe sąsiedztwo) - odległość pomiędzy skupieniami określona jest poprzez odległość tych obiektów każdego skupienia, które znajdują się najbliżej siebie.



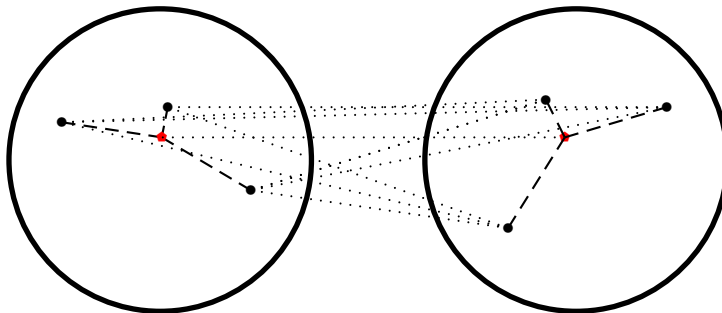
- Metoda pełnego wiązania (najdalsze sąsiedztwo) - odległość pomiędzy skupieniami określona jest poprzez odległość tych obiektów każdego skupienia, które znajdują się najdalej siebie.



- 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 liczebnościach.
- Metoda Warda - opiera się na zasadzie analizy wariancji - wylicza różnicę między sumami kwadratów odchyień odległości poszczególnych obiektów od środka ciężkości skupień, do których te obiekty należą. Metoda ta wybierana jest najczęściej ze względu na jej dość uniwersalny charakter.



Wynik analizy skupień prowadzonej metodą hierarchiczną przedstawia się przy pomocy dendogramu. **Dendogram** jest formą drzewa wskazującego związki pomiędzy poszczególnymi obiektami uzyskane z analizy macierzy podobieństwa. Poziom odcięcia dendogramu 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 dendogramie.

Okno z ustawieniami opcji hierarchicznej analizy skupień wywołujemy poprzez menu Statystyki zaawansowane → Grupowanie i Redukcja → Hierarchiczna analiza skupień.

Hierarchiczna analiza skupień

Analiza statystyczna : Hierarchiczna analiza skupień

Zmienna

- 1-tyt Irysów
- 2-dł_działki
- 3-szer_działki
- 4-dł_płatka
- 5-szer_płatka

Identyfikator obiektu

1-tyt Irysów

Opcje testu

Metoda wiązania Standaryzuj

Pojedyncze wiązanie

Odległość Euklidesowa - wszystkich elementów

Poziom odcięcia skupień (%) 50

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane+skupiska Połącz w jeden raport

Dołącz wykres Przebieg wiązania

OK Zamknij

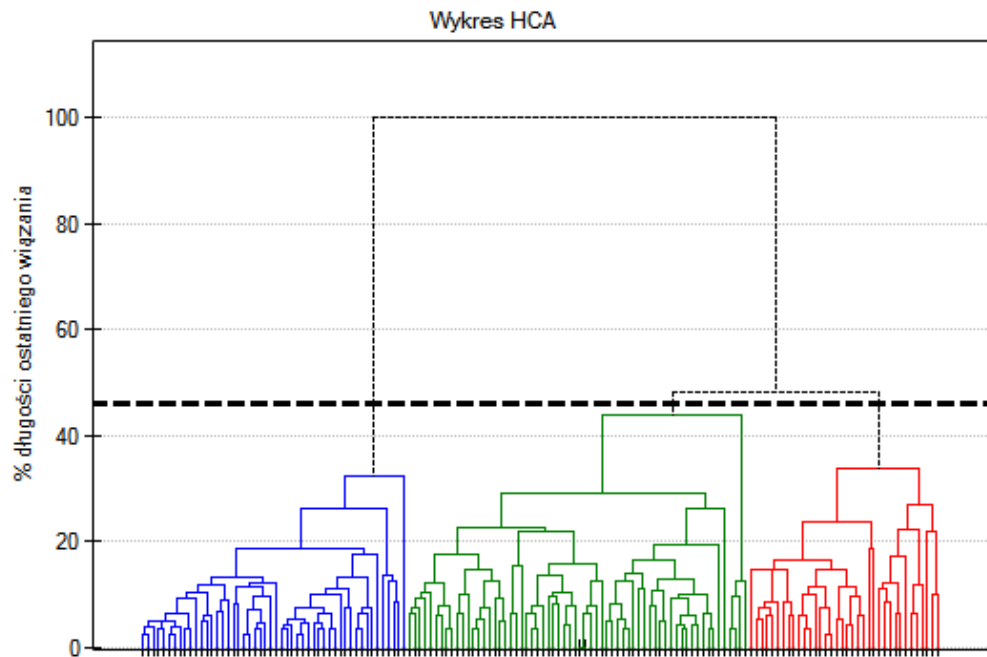
Przykład (25.1) c.d. (plik iris.pqs)

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[58]). Ponieważ w tym zestawie danych znajduje się informacja o rzeczywistej odmianie każdego kwiatu, po przeprowadzonej analizie skupień istnieje możliwość określenia dokładności dokonanego podziału.

Przydzieleniu 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.

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](#).

Kappa-Cohena ✕

— Analiza statystyczna : Test istotności Kappa-Cohena

▼ Zmienna 1

- 1-typ Irysów
- 2-dł_działki
- 3-szer_działki
- 4-dł_płatka
- 5-szer_płatka
- 6-Skupienie

▼ Zmienna 2

- 1-typ Irysów
- 2-dł_działki
- 3-szer_działki
- 4-dł_płatka
- 5-szer_płatka
- 6-Skupienie

Tabela kontyngencji Dane surowe

Kappa nieważona

Kappa ważona (wagi liniowe)

Kappa ważona (wagi kwadratowe)

Filtr danych

Wybór danych spełniających odpowiednie warunki +

Reguły połączone są koniunkcją -

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane

Dołącz wykres

Dołącz wartości procentowe wiersze

0,05 ▼ poziom istotności

Dla przedstawionego przykładu obserwowaną zgodność przedstawia tabela:

Dane :	1	2	3
setosa	50	0	0
versicolor	0	50	0
virginica	0	14	36

Wnioskujemy z niej, że odmiana virginica może być mylona z versicolor, dlatego obserwujemy tu 14 błędnych zaklasyfikowań. Natomiast współczynnik zgodności Kappa jest istotny statystycznie i wynosi 0.86, co świadczy o dużej zgodności uzyskanych skupień z rzeczywistą odmianą kwiatów.

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

25.2.2 Metoda k-średnich

Metoda k-średnich (*ang. k-means*) wzoruje swoje działanie na algorytmie zaproponowanym początkowo przez Stuarta Lloyda i opublikowana w roku 1982 [105]. 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 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.

PROCEDURA AGLOMERACYJNA

1. Wybór centów 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.

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:

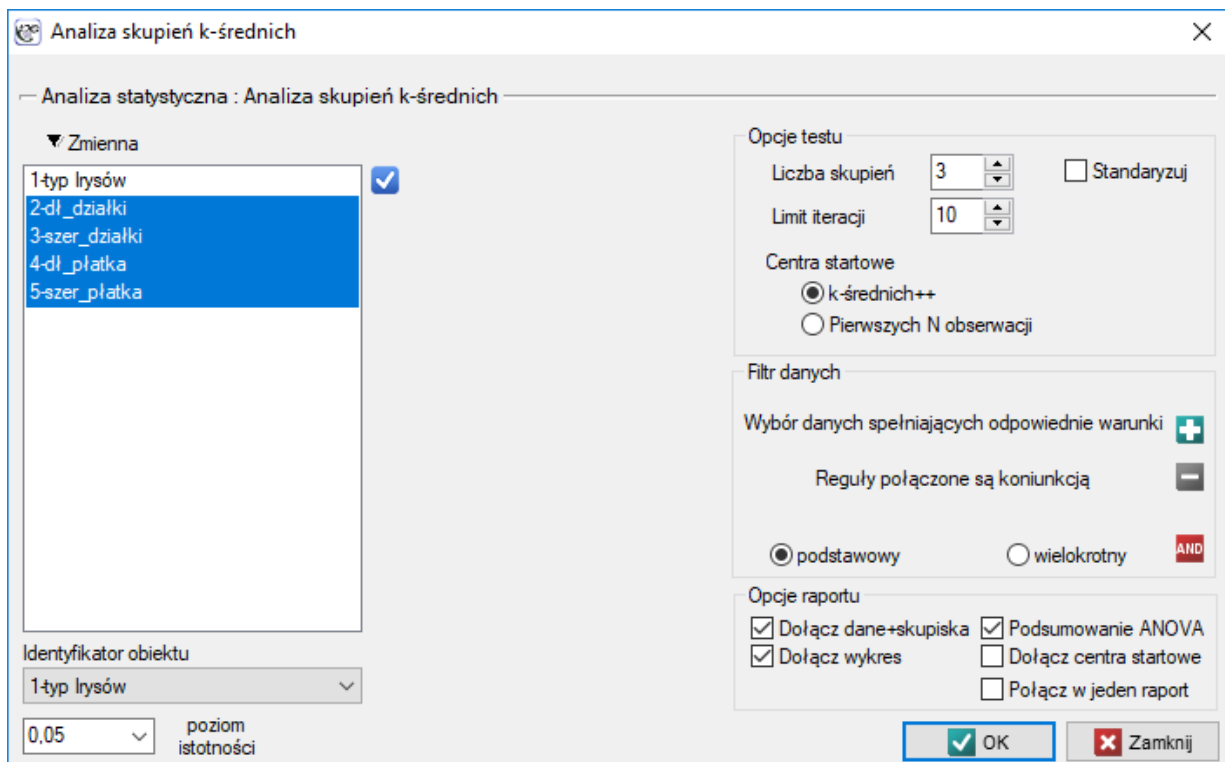
k-means++ - to optymalny dobór punktów startowych poprzez wykorzystanie algorytmu k-means++ zaproponowanego w 2007 roku przez David Arthur and Sergei Vassilvitskii [1]. Gwarantuje on uzyskanie optymalnego rozwiązania algorytmu k-średnich przy możliwie niewielu iteracjach. Algorytm wykorzystuje w swoim działaniu element losowości, dlatego uzyskiwane wyniki mogą się nieco różnić przy kolejnych uruchomieniach analizy. Jeśli dane nie formułują się w naturalne klastry, lub jeśli nie można skutecznie podzielić danych na rozłączne skupienia, użycie k-means++ spowoduje uzyskanie zupełnie innych wyników 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 kroku 2 i 3 algorytmu osiągnięta przez użytkownika liczbę iteracji.

Ze względu na sposób działania algorytmu analizy skupień k-średnich, naturalną jej konsekwencją jest porównanie powstałych skupienia przy pomocy jednoczynnikowej analizy wariancji (**ANOVA**) dla grup niezależnych.

Okno z ustawieniami opcji analizy skupień k-średnich wywołujemy poprzez menu Statystyki zaawansowane → Grupowanie i Redukcja → Analiza skupień k-średnich.



Przykład (25.1) c.d. (plik iris.pqs)

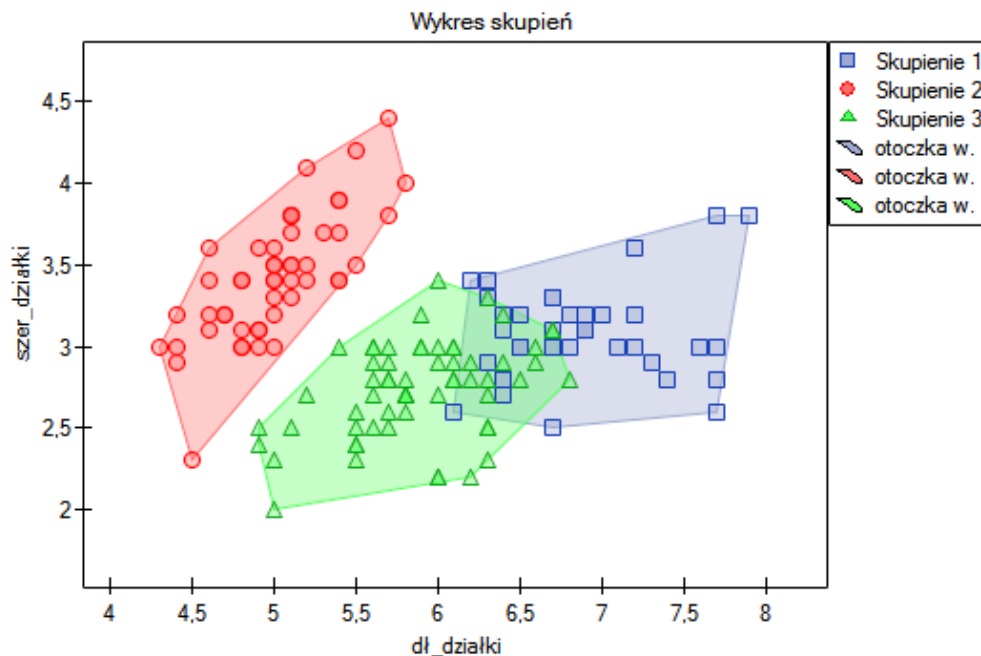
Analiza przeprowadzona zostanie na klasycznym zestawie danych, dotyczącym podziału kwiatów irysa na 3 odmiany na podstawie szerokości oraz długości płatków i działek kielicha (R.A. Fishera 1936[58]).

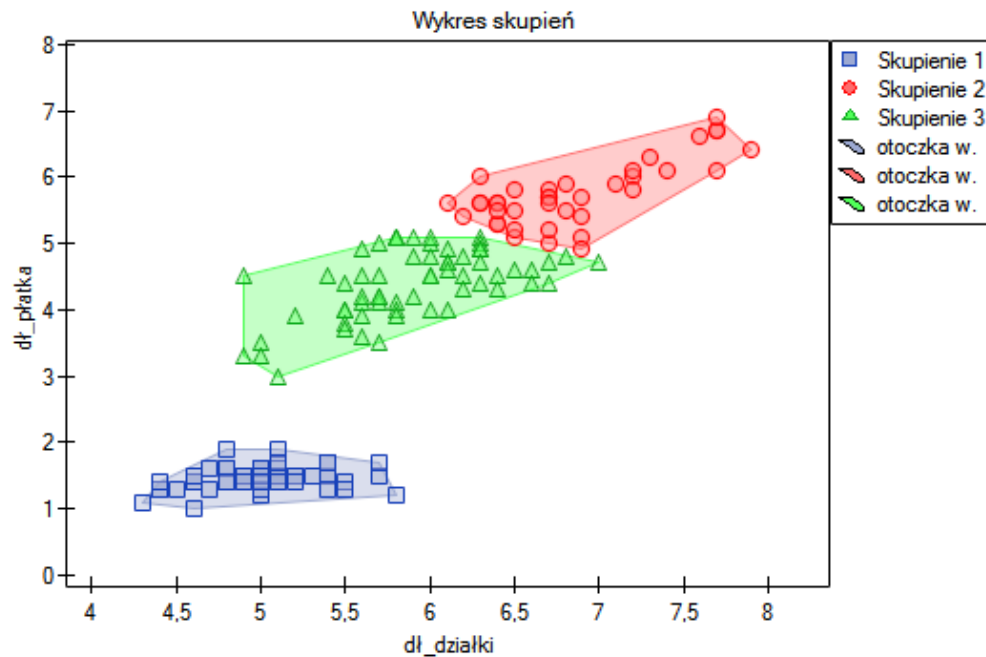
Przydzieleniu kwiatów do poszczególnych grup dokonujemy na podstawie kolumn od 2 do 5. Wskazujemy również, że badane kwiaty chcemy podzielić na 3 grupy. W rezultacie przeprowadzonej analizy utworzono 3 skupienia, które różnią się istotnie statystycznie w każdym badanym wymiarze (wyniki ANOVA), czyli dla szerokości płatka, długości płatka, szerokości kielicha jak i długości kielicha.

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

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

Różnicę można zaobserwować na wykresach, gdzie przedstawiamy w wybranych dwóch wymiarach przynależność poszczególnych punktów do skupienia:





Powtarzając analizę możemy uzyskać nieco inne wyniki, ale uzyskana zmienność nie będzie duża. Świadczy to o tym, że dane poddane analizie formują się w naturalne skupienia i przeprowadzanie analizy skupień jest w tym przypadku uzasadnione.

Uwaga 1!

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 2!

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](#).

26 ANALIZA PRZEŻYCIA

Analiza przeżycia stosowana często w medycynie, w innych dyscyplinach naukowych występuje pod takimi nazwami jak: analiza niezawodności (*ang. reliability analysis*), analiza trafności (*ang. duration analysis*), analiza historii zdarzeń (*ang. event history analysis*). Jej głównym celem jest ocena czasu przeżycia np. pacjentów po operacji – stosowanym narzędziem są tu tabele trwania życia i krzywe Kaplana-Meiera. Innym interesującym aspektem jest porównanie czasów przeżycia np. czasu przeżycia po zastosowaniu różnych metod leczenia – w tym celu wykorzystywane są metody porównania 2 lub więcej krzywych przeżycia. Opracowano również szereg metod (modeli regresji) pozwalających na badanie wpływu różnych zmiennych na czas przeżycia.

By ułatwić zrozumienie zagadnienia, podstawowe definicje podane zostaną na przykładzie opisującym długość życia pacjentów po przeszczepie serca:

Zdarzenie – jest to interesująca badacza zmiana np. śmierć;

Czas przeżycia – jest to okres pomiędzy stanem początkowym a wystąpieniem zdarzenia np. długość życia pacjenta po przeszczepie serca;

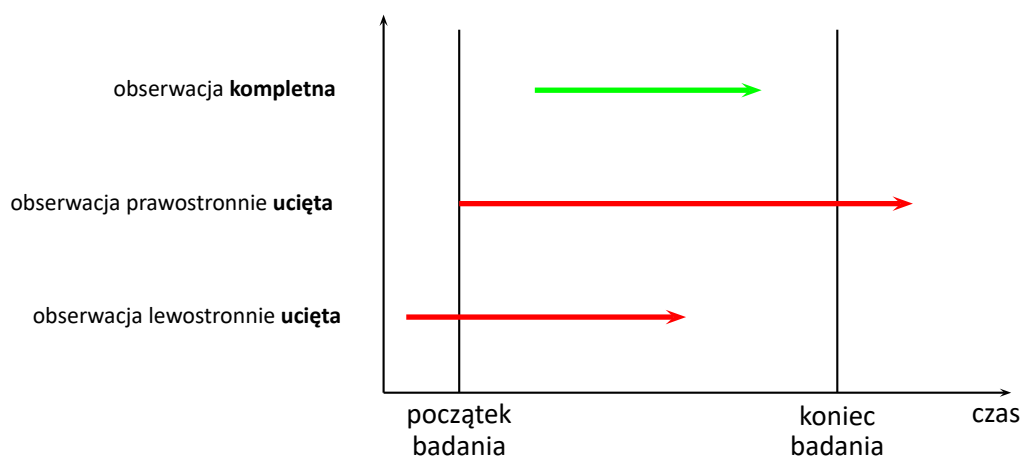
Uwaga!

W analizie należy zaznaczyć jedną kolumnę z wyliczonym czasem. Gdy dysponujemy dwoma punktami w czasie: punktem początkowym i końcowym przed analizą wyliczamy czas jaki upłynął między tymi punktami korzystając z formuł arkusza

Obserwacje ucięte (*ang. censored*) – to obserwacje, dla których mamy niepełne informacje o czasie przeżycia.

Obserwacje ucięte i kompletne - przykład dotyczący przeżycia po przeszczepie serca:

- **obserwacja kompletna** - znana jest data przeszczepu i data śmierci pacjenta, zatem można ustalić dokładny czas przeżycia po przeszczepie;
- **obserwacja ucięta prawostronnie** - nie jest znana data śmierci pacjenta (chory żyje w chwili ukończenia badania), więc nie można ustalić dokładnego czasu przeżycia;
- **obserwacja ucięta lewostronnie** - nie jest znana data przeszczepu serca, ale wiadomo, że było to przed rozpoczęciem niniejszego badania, zatem nie można ustalić dokładnego czasu przeżycia.

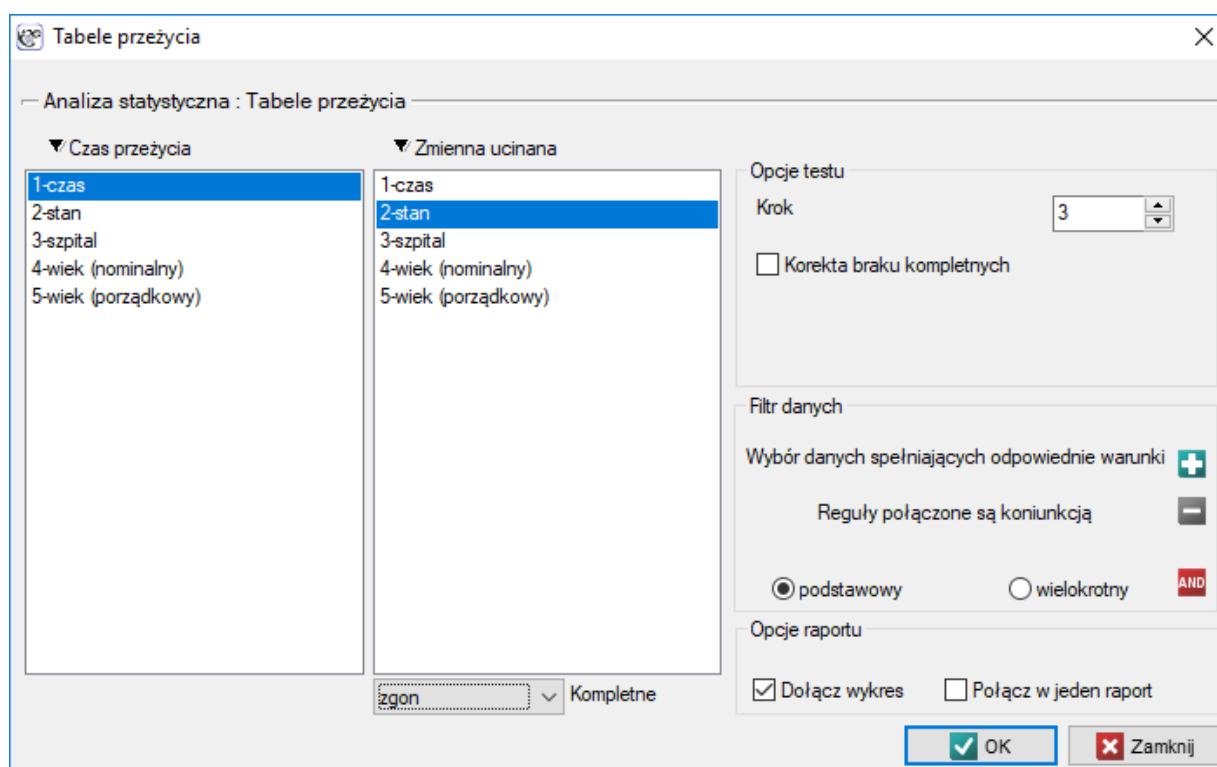


Uwaga!

Zakończenie badania oznacza zakończenie obserwacji pacjenta, nie zawsze jest to moment zakończenia badania dla wszystkich pacjentów. Zakończeniem badania może być moment w którym tracimy kontakt z pacjentem (nie znamy więc jego czasu przeżycia). Analogicznie rozpoczęcie badania nie musi dotyczyć wszystkich w tym samym czasie.

26.1 TABELE PRZEŻYCIA

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 można zdefiniować podając krok. Dla każdego przedziału program PQStat wylicza:

- **liczbę przypadków wchodzących** – liczba osób, które przeżyły do określonego przez przedział czasu;
- **liczbę przypadków uciętych** – liczba osób w danym przedziale, które są zdefiniowane jako przypadki ucięte;
- **liczbę przypadków zagrożonych** – liczba przypadków, które wchodzą do danego przedziału minus połowa liczby przypadków uciętych w danym przedziale;
- **liczbę przypadków kompletnych** – liczba osób, które doświadczyły zdarzenia (czyli zmarły) w danym przedziale;
- **proporcje przypadków kompletnych** – stosunek liczby przypadków kompletnych (zgonów) w danym przedziale do liczby przypadków zagrożonych w tym przedziale;

- **proporcje przypadków przeżywających** – oblicza się jako 1 minus proporcja przypadków kompletnych w danym przedziale;
- **skumulowane proporcje przeżywających (funkcja przeżycia)** – prawdopodobieństwo przeżycia do danego czasu – ponieważ żeby przeżyć kolejny przedział czasu, trzeba przeżyć wszystkie poprzednie, to prawdopodobieństwo to wyliczane jest jako iloczyn wszystkich poprzednich proporcji przypadków przeżywających;
 - ± błąd standardowy funkcji przeżycia;
- **gęstość prawdopodobieństwa** – oszacowane prawdopodobieństwo doświadczenia zdarzenia (zgonu) w danym przedziale obliczone w jednostce czasu;
 - ± błąd standardowy gęstości prawdopodobieństwa;
- **stopę hazardu** – prawdopodobieństwo (wyliczane na jednostkę czasu) tego, że przypadek, który przeżył do początku danego przedziału doświadczy w tym przedziale zdarzenia (zgonu);
 - ± błąd standardowy stopy hazardu.

Uwaga!

W przypadku braku obserwacji kompletnych w dowolnym przedziale czasu przeżycia istnieje możliwość zastosowania korekty. Zerową liczbę przypadków kompletnych zastępuje się wówczas wartością 0.5.

Interpretacja graficzna

Dla zobrazowania informacji uzyskanych dzięki tabelom przeżycia możemy posłużyć się kilkoma wykresami:

- wykresem funkcji przeżycia,
- wykresem gęstości prawdopodobieństwa,
- wykresem stopy hazardu.

PRZYKŁAD 26.1. (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 (<45lat; 60lat). Fragment zebranych danych przedstawia poniższa tabela:

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

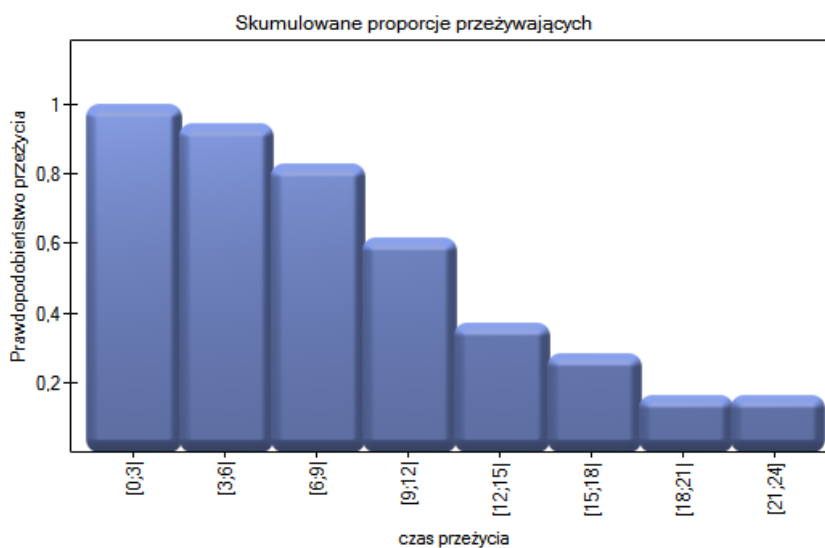
Dane kompletne w tej analizie to te, o których mamy pełną informację o długości życia po przeszczepie, czyli opisane jako "zgon" (są to 53 osoby, co stanowi 59.55%). Dane ucięte to te, o których tej informacji nie mamy ponieważ w chwili zakończenia badania żyją (jest to 36 osób, co stanowi 40.45%). Budujemy tablice trwania życia tych pacjentów tworząc 3 letnie przedziały czasowe:

Tabela trwania życia												
przedział	ucięte	kompletne	wchodzące	zagrożone	prop. kom	prop. prze	skumulow:	gęstość pr.	stopa haż	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		

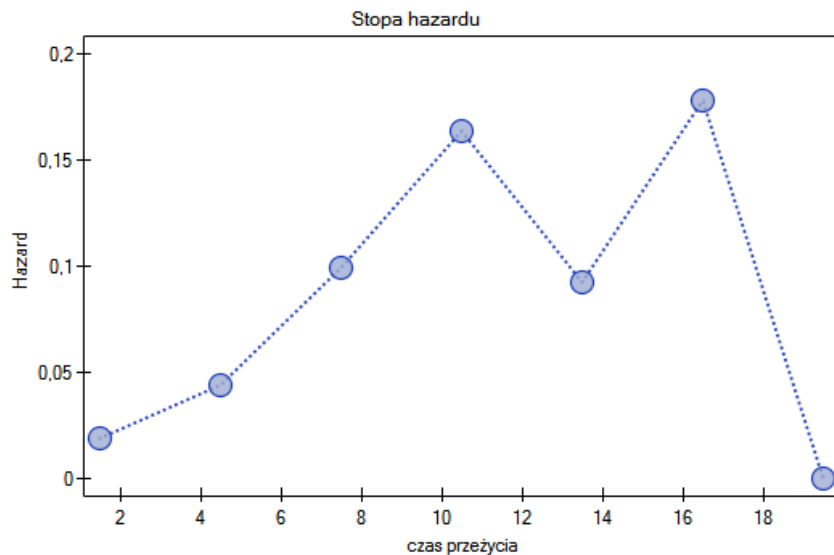
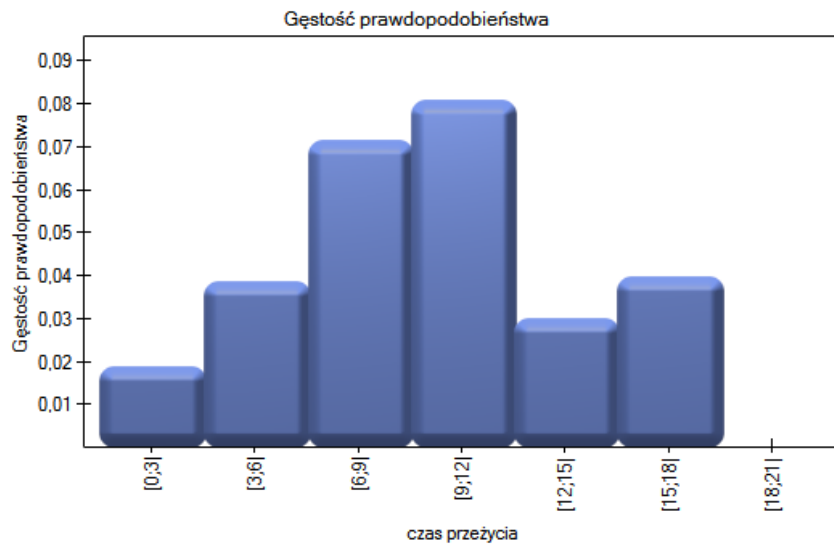
Dla każdego 3-letniego przedziału możemy zinterpretować uzyskane w tabeli wyniki - przykładowo dla osób, które żyją przynajmniej 9 lat od przeszczepu, a zatem weszły do przedziału [9;12):

- liczba osób, które przeżyły do 9 lat od przeszczepu to 39;
- jest 7 osób, o których wiemy, że żyły przynajmniej 9-12 lat - gdy informacje o nich były zbierane, ale nie wiemy czy żyły dłużej, gdyż po tym czasie wypadły z badania;
- liczbę 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 ± 0.02 - to prawdopodobieństwo zgonu dla każdego roku z przedziału 9-12
- 0.16 ± 0.04 - to warunkowe, roczne prawdopodobieństwo zgonu dla przypadku, który przeżył do początku przedziału 9-12;

Wyniki przedstawimy na kilku wykresach:



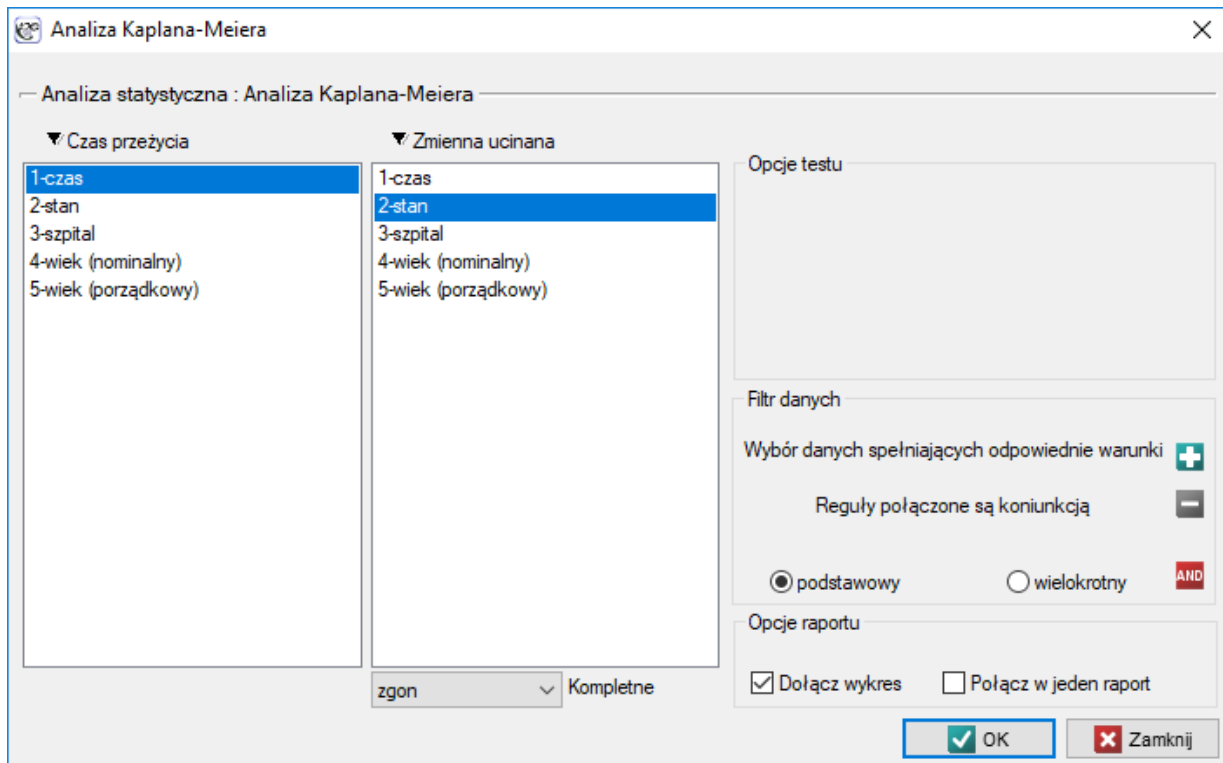
Prawdopodobieństwo przeżycia maleje wraz z upływem czasu jaki minął od przeszczepu. Nie obserwujemy jednak nagłego spadku funkcji przeżycia, a więc czasu w który gwałtownie wzrastałoby prawdopodobieństwo zgonu.



26.2 KRZYWE KAPLANA-MEIERA

Krzywe Kaplana-Meiera pozwalają na ocenę czasu przeżycia bez konieczności arbitralnego grupowania obserwacji, tak jak to jest w tabelach przeżycia. Estymator ten został wprowadzony przez Kaplana i Meiera (1958)[89].

Okno z ustawieniami opcji krzywej Kaplana-Meiera wywołujemy poprzez menu Statystyka→Analiza przeżycia→Analiza Kaplana-Meiera

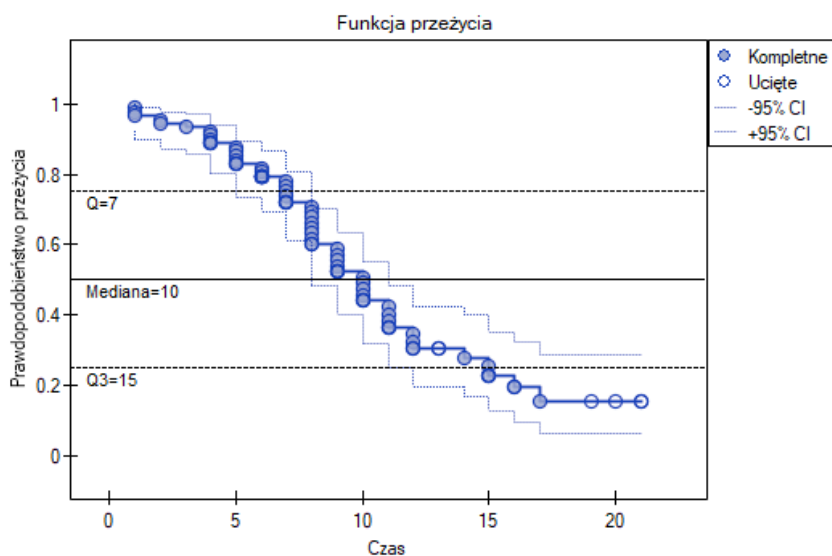


Podobnie jak dla tablic przeżycia, wyliczamy tutaj funkcję przeżycia, czyli prawdopodobieństwo przeżycia do danego czasu. Wykres funkcji przeżycia Kaplana-Meiera tworzy funkcja schodkowa. Na podstawie błędu standardowego (formuła Greenwooda) i transformacji logarytmicznej (log-log) zbudowane są przedziały 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 iż u połowy pacjentów nastąpi zgon w ciągu wskazanego czasu. Zarówno mediana jak i inne percentyle wyznaczone są jako najkrótszy czas przeżycia, dla którego funkcja przeżycia jest mniejsza lub równa danemu percentylowi. Dla mediany wyznaczany jest przedział ufności w oparciu o metodę "test-based" autorstwa Brookmeyer i Crowley (1982)[28]. **Średnia czasu przeżycia** jest wyznaczona jako pole pod krzywą przeżycia.

Dane dotyczące czasu przeżycia są zwykle mocno skośne, dlatego w analizie przeżycia, mediana jest lepszą miarą tendencji centralnej niż średnia.

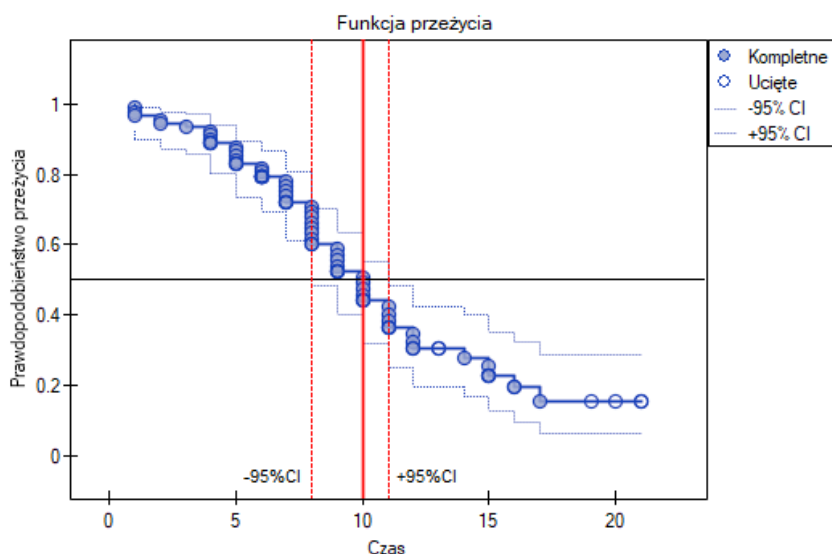
Przykład (26.1) c.d. (plik *przeszczep.pqs*)

Przedstawimy długość życia po przeszczepie wątroby przy pomocy krzywej Kaplana-Meiera



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 obciążony ryzykiem zgonu. Z wartości mediany wynika, że w ciągu 10 lat od przeszczepu oczekujemy, że u połowy pacjentów nastąpi zgon. Wartość tę zaznaczamy na wykresie rysując linię w punkcie 0.5 oznaczającym medianę. 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 linie na poziomie 0.5.

26.3 PORÓWNYWANIE KRZYWYCH PRZEŻYCIA

Funkcje przeżycia mogą być budowane oddzielnie dla różnych podgrup np. oddzielnie dla kobiet i mężczyzn, a następnie porównane między sobą. Takie porównanie może dotyczyć zarówno dwóch, jak i kilku krzywych.

Okno z ustawieniami opcji porównania krzywych przeżycia wywołujemy poprzez menu Statystyka → Analiza przeżycia → Porównanie grup

Porównanie grup ✕

— Analiza statystyczna : Porównanie grup —

▼ Czas przeżycia

- 1-czas
- 2-stan
- 3-szpital
- 4-wiek (nominalny)
- 5-wiek (porządkowy)

▼ Zmienna ucinana

- 1-czas
- 2-stan
- 3-szpital
- 4-wiek (nominalny)
- 5-wiek (porządkowy)

zgon ▼ Kompletnie

Opcje testu

LogRank ▼

Zmienna warstwy Dla trendu

- 1-czas
- 2-stan
- 3-szpital
- 4-wiek (nominalny)
- 5-wiek (porządkowy)

▼ Zmienna grupująca

- 1-czas
- 2-stan
- 3-szpital
- 4-wiek (nominalny)
- 5-wiek (porządkowy)

Filtr danych

Wybór danych spełniających odpowiednie warunki +

Reguły połączone są koniunkcją -

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane Pełne obliczenia

Dołącz wykres Połącz w jeden raport

0,05 ▼ poziom istotności

OK Zamknij

Porównania k krzywych przeżycia S_1, S_2, \dots, S_k w poszczególnych punktach czasu przeżycia t w programie możemy dokonać przy pomocy trzech testów:

Testu log-rank najbardziej znanego i szeroko stosowanego, nawiązującego do procedury Mantela-Heanszela dla wielu tabel 2x2 (Mantel-Heanszel 1959[110], Mantel 1966[112], Cox 1972[47]),

Uogólnienia Gehana testu Wilcoxona wywodzącego się od testu Wilcoxona (Breslow 1970, Gehan 1965[67][68]),

Testu Taron-Ware wywodzącego się od testu Wilcoxona (Tarone i Ware 1977[157]).

Wszystkie trzy testy działają w oparciu o tę samą statystykę testową, inne są tylko **wagi** w_j w poszczególnych punktach osi czasu, na których bazuje statystyka testowa.

Test log-rank: $w_j = 1$ - wszystkie punkty osi czasu posiadają tę samą wagę, co daje większy wpływ na uzyskany wynik późniejszym wartościom osi czasu;

Uogólnienie Gehana testu Wilcoxona: $w_j = n_j$ - momenty czasowe są ważone liczbą obserwacji w każdym z nich, a zatem przypisywane są większe wagi początkowym wartościom osi czasu;

Test Taron-Ware: $w_j = \sqrt{n_j}$ - momenty czasowe są ważone pierwiastkiem z liczby obserwacji w każdym z nich co powoduje, usytuowanie tego testu pomiędzy dwoma omówionymi wcześniej.

Ważnym warunkiem stosowania powyższych testów jest proporcjonalność hazardu. Hazard definiowany jako nachylenie krzywej przeżycia jest miarą tego, jak szybko następuje niepożądane zdarzenie. Złamanie założenia proporcjonalności hazardu choć nie dyskwalifikuje całkowicie powyższych testów, to niesie kilka niebezpieczeństw. Przede wszystkim położenie punktu przecięcia krzywych względem osi czasu ma decydujący wpływ na obniżenie mocy poszczególnych testów.

26.3.1 Różnice w krzywych przeżycia

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : S_1(t) &= S_2(t) = \dots = S_k(t), \quad \text{dla wszystkich } t, \\ \mathcal{H}_1 : &\text{nie wszystkie } S_i(t) \text{ są sobie równe.} \end{aligned}$$

W obliczeniach wykorzystano statystykę chi-kwadrat postaci:

$$\chi^2 = U'V^{-1}U$$

gdzie:

$$U_i = \sum_{j=1}^m w_j (d_{ij} - e_{ij})$$

V - macierz kowariancji o wymiarach $(k-1) \times (k-1)$

gdzie:

$$\text{diagonała: } \sum_{j=1}^m w_j^2 \frac{n_{ij}(n_j - n_{ij})d_j(n_j - d_j)}{n_j^2(n_j - 1)},$$

$$\text{poza diagonalą (off diagonal): } \sum_{j=1}^m w_j^2 \frac{n_{ij}n_{lj}d_j(n_j - d_j)}{n_j^2(n_j - 1)}$$

m - liczba momentów czasowych, w których nastąpiło niepożądane zdarzenie (zgon),

$d_j = \sum_{i=1}^k d_{ij}$ - obserwowana liczba niepożądanych zdarzeń (zgonów) w j -tym momencie czasowym,

d_{ij} – obserwowana liczba niepożądanych zdarzeń (zgonów) w i -tej grupie w j -tym momencie czasowym,

$e_{ij} = \frac{n_{ij}d_j}{n_j}$ – oczekiwana liczba niepożądanych zdarzeń (zgonów) 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 licznosci oczekiwanych) rozkład χ^2 z $df = k - 1$ stopniami swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Iloraz Hazardów

W teście Log-rank dla każdej grupy podawane są obserwowane wartości niepożądanych zdarzeń (zgonów) $O_i = \sum_{j=1}^m d_{ij}$ oraz odpowiednie wartości oczekiwane $E_i = \sum_{j=1}^m e_{ij}$.

Miarą opisującą wielkość obserwowanej różnicy między parą krzywych przeżycia jest Iloraz Hazardów (*ang. Hazard Ratio - HR*).

$$HR = \frac{O_1/E_1}{O_2/E_2}$$

Jeśli Iloraz Hazardów jest większy niż 1 np. $HR = 2$, to stopień narażenia na niepożądane zdarzenie w pierwszej grupie jest dwa razy większy niż w grupie drugiej. Odwrotna sytuacja jest gdy HR jest mniejsze niż jeden. Natomiast przy HR równym 1 obie grupy są narażone w tym samym stopniu.

Uwaga!

Przedział ufności dla HR wyliczany jest w oparciu o błąd standardowy logarytmu HR (Armitage i Berry 1994[11]).

26.3.2 Trend w krzywych przeżycia

Hipotezy:

\mathcal{H}_0 : W badanej populacji nie istnieje trend w położeniu krzywych S_1, S_2, \dots, S_k ,

\mathcal{H}_1 : W badanej populacji istnieje trend w położeniu krzywych S_1, S_2, \dots, S_k .

W obliczeniach wykorzystano statystykę chi-kwadrat postaci:

$$\chi^2 = \frac{(c'U)^2}{c'Ve}$$

gdzie:

$c = (c_1, c_2, \dots, c_k)$ – wektor wag dla porównywanych grup informujący o ich naturalnym porządku (najczęściej kolejne liczby naturalne).

Statystyka ta ma asymptotycznie (dla dużych licznosci oczekiwanych) rozkład χ^2 z jednym stopniem swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

By można było przeprowadzić analizę trendu w krzywych przeżycia, zmienna grupująca musi być zmienną liczbową, w której wartości liczb informują o naturalnym porządku grup. Liczby te w analizie traktowane są jako wagi c_1, c_2, \dots, c_k .

26.3.3 Krzywe przeżycia dla warstw

Często chcąc porównać czasy przeżycia dla dwóch lub więcej grup nie możemy zapomnieć o innych czynnikach, które mogą mieć wpływ na wynik tego porównania. Dostosowanie (korekcja) analizy o takie czynniki może być przydatna. Na przykład w badaniach domu opieki porównujących długość pobytu osób poniżej i powyżej 80 roku życia uzyskano istotną różnicę. Wiadomo jednak, że płeć ma silny związek z długością pobytu, a także wiekiem. Dlatego próbując ocenić wpływ wieku dobrym pomysłem byłaby stratyfikacja analizy ze względu na płeć.

Hipotezy dla różnic w krzywych przeżycia:

$$\mathcal{H}_0 : S_1^*(t) = S_2^*(t) = \dots = S_k^*(t), \text{ dla wszystkich } t,$$

$$\mathcal{H}_1 : \text{nie wszystkie } S_i^*(t) \text{ są sobie równe.}$$

Hipotezy dla analizy trendu w krzywych przeżycia:

$$\mathcal{H}_0 : \text{W badanej populacji nie istnieje trend w położeniu krzywych } S_1^*, S_2^*, \dots, S_k^*,$$

$$\mathcal{H}_1 : \text{W badanej populacji istnieje trend w położeniu krzywych } S_1^*, S_2^*, \dots, S_k^*.$$

gdzie $S_1^*(t), S_2^*(t), \dots, S_k^*(t)$ - to krzywe przeżycia po korekcji o zmienną wyznaczającą warstwy.

Obliczenia dla statystyk testowych bazują na formułach opisanych dla testów nie uwzględniających warstw z tą różnicą, że macierz U i V jest zastąpiona sumą macierzy $\sum_{l=1}^L U$ i $\sum_{l=1}^L V$. Sumowanie następuje po warstwach utworzonych przez zmienną, względem której dostosowujemy (korygujemy) analizę (adjusted) $l=1,2,\dots,L$

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Przykład (26.1) 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 zależały od szpitala, w którym dokonywano przeszczepu. Porównania krzywych przeżycia dla tych szpitali dokonamy w oparciu o wszystkie zaproponowane w programie testy służące temu porównaniu.

Hipotezy:

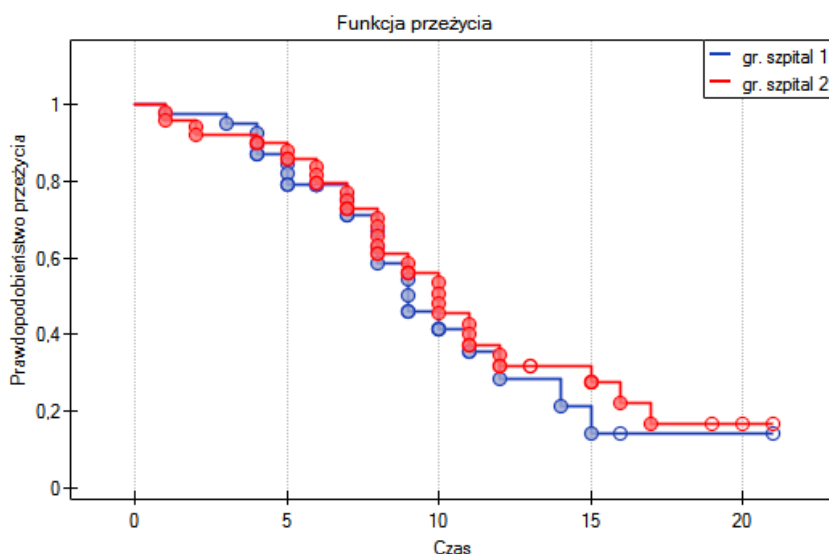
$$\mathcal{H}_0 : \text{krzywa przeżycia pacjentów szpitala 1} = \text{krzywa przeżycia pacjentów szpitala 2,}$$

$$\mathcal{H}_1 : \text{krzywa przeżycia pacjentów szpitala 1} \neq \text{krzywa przeżycia pacjentów szpitala 2.}$$

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



Na podstawie przyjętego poziomu $\alpha = 0.05$, w oparciu o uzyskaną wartość $p=0.6004$ dla testu log-rank ($p=0.6959$ dla Gehana i 0.6465 dla Tarona) wnioskujemy, że nie ma podstaw by odrzucić hipotezę \mathcal{H}_0 . Długość życia wyliczona dla pacjentów obu tych szpitali jest podobna.

Do tego samego wniosku dojdziemy porównując ryzyko zgonu dla tych szpitali poprzez wyznaczenie ilorazu tego ryzyka. Uzyskana oszacowana wartość $HR = 1.1499$, a 95% przedział ufności dla tej wartości zawiera jedynekę: $(0.6570, 2.0126)$.

Różnice dla wielu krzywych przeżycia

Przeszczepy wątroby dokonywane były u ludzi w różnym wieku. Wyróżniono 3 grupy wiekowe: $(45\text{lat}; 50\text{lat})$, $(50\text{lat}; 55\text{lat})$, $(55\text{lat}; 60\text{lat})$. Sprawdźmy, czy długość życia pacjentów po przeszczepie zależy od ich wieku w chwili dokonania przeszczepu.

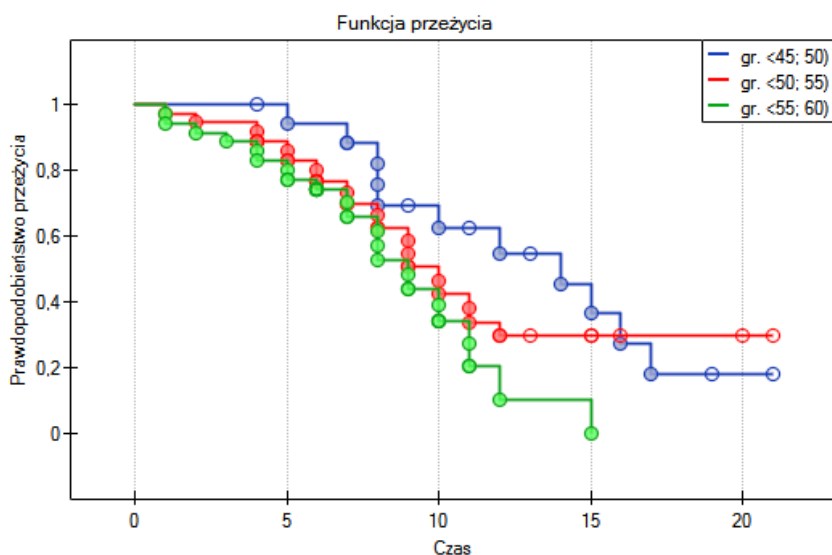
Hipotezy:

- \mathcal{H}_0 : krzywe przeżycia pacjentów w wieku (45lat; 50lat), (50lat; 55lat), (55lat; 60lat) 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



Na podstawie przyjętego poziomu $\alpha = 0.05$, w oparciu o uzyskaną wartość $p=0.0692$ w teście log-rank ($p=0.09279$ dla Gehana, $p=0.0779$ dla Tarona) wnioskujemy, że nie ma podstaw by odrzucić hipotezę \mathcal{H}_0 . Długość życia wyliczona dla pacjentów należących do porównywanych trzech grup wiekowych jest podobna. Choć należy zauważyć że wartości p są dość bliskie standardowemu poziomowi istotności 0.05.

Przeglądając wartości hazardu (ilorazu wartości obserwowanych i oczekiwanych niepożądanych zdarzeń) zauważamy, że z każdą kategorią wiekową są one nieco wyższe (0.68, 0.93, 1.43). Chociaż nie

wykryto istotnych statystycznie różnic między nimi, to możliwe jest, że znaleziony zostanie trend wzrostu wartości hazardu (trend w położeniu krzywych przeżycia).

Trend dla kilku krzywych przeżycia

Jeśli do testu wprowadzimy informację dotyczącą uporządkowania porównywanych kategorii (wykorzystamy zmienną wiek, w której przedziały wiekowe ponumerujemy odpowiednio 1, 2 i 3), wówczas będziemy mogli sprawdzić, czy istnieje trend w porównywanych krzywych. Będziemy badać 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

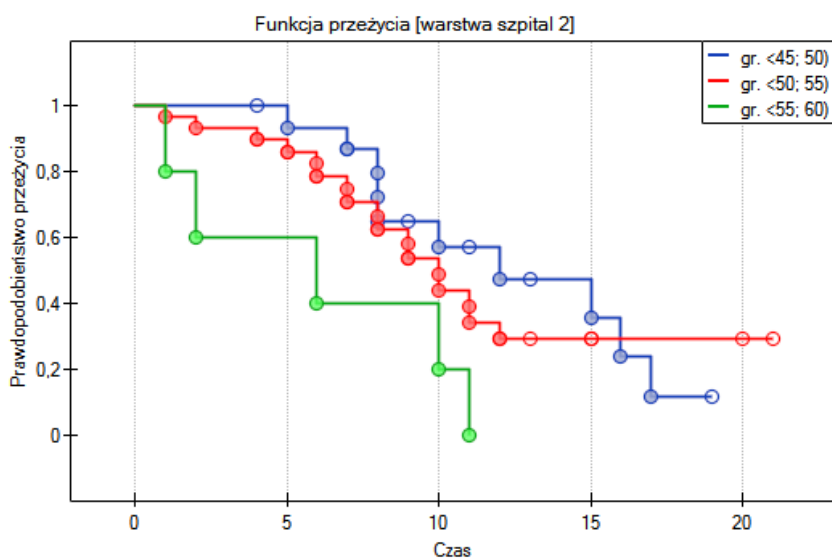
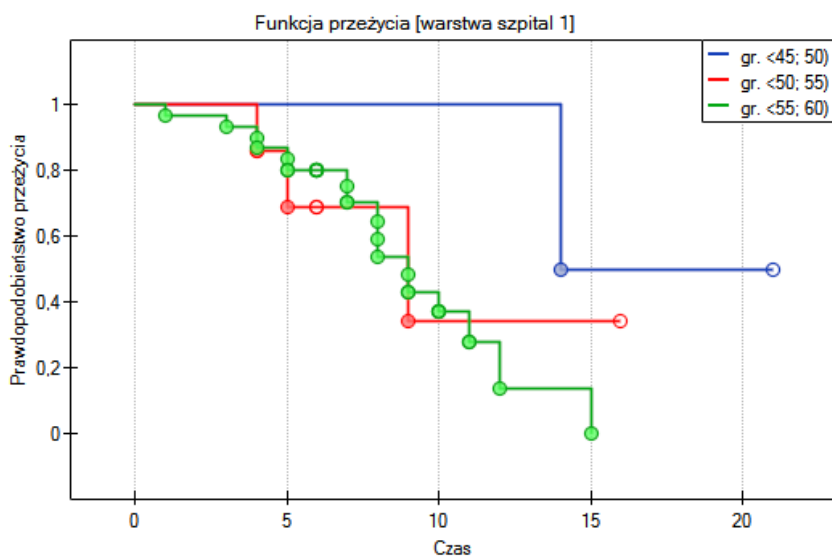
Na podstawie przyjętego poziomu $\alpha = 0.05$, w oparciu o uzyskaną wartość $p=0.0237$ w teście log-rank ($p=0.0317$ dla Gehana, $p=0.0241$ dla Tarona) wnioskujemy, że krzywe przeżycia ułożone są w pewnym trendzie. Najniżej na wykresie Kaplana-Meiera znajduje się krzywa dla osób w wieku (55 lat; 60 lat). Nad nią jest krzywa dla pacjentów w wieku (50 lat; 55 lat). Najwyżej zaś krzywa dla pacjentów w wieku (45 lat; 50 lat). Zatem czym starszy pacjent w chwili przeszczepu, tym mniejsze prawdopodobieństwo przeżycia określonego odcinka czasu.

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.

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



W raporcie najpierw przedstawiona jest analiza poszczególnych warstw, zarówno wyniki testów jak i wartości hazardu. W warstwie pierwszej trend wzrostu hazardu jest widoczny, choć nieistotny, trend o tym samym kierunku (wynik na pograniczu istotności statystycznej) obserwowany jest w warstwie drugiej. Kumulacja tych trendów we wspólnej analizie warstw pozwoliła uzyskać istotność trendu krzywych przeżycia. Zatem: 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 Tarona a nieistotny dla Gehana, co może wskazywać na to, że pojawiające się różnice w krzywych nie są tak widoczne w początkowych okresach czasu przeżycia co w okresach późniejszych. Przyglądając się ilorazowi hazardu dla porównywanych parami krzywych

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 mamy dla porównania krzywej dla najmłodszej grupy z krzywą dla grupy najstarszej 0.53, 95% przedział ufności dla tego ilorazu $\langle 0.26 ; 1.05 \rangle$ zawiera co prawda wartość 1, ale jest na pograniczu tej wartości, co może sugerować wystąpienie między odpowiadającymi im krzywymi istotnych różnic. By potwierdzić to przypuszczenie dociekliwy badacz, używając filtru danych w oknie analizy, może porównać krzywe parami.

Filtr danych		
zmienna	warunek	wartość
4-wiek (nominalny)	=	<45; 50
4-wiek (nominalny)	=	<50; 55

Należy jednak pamiętać by zastosować jedną z poprawek używanych przy wielokrotnych porównaniach i zmodyfikować poziom istotności. W tym przypadku dla poprawki Bonferroniego przy trzech porównaniach poziom istotności wyniesie 0.017. Dla uproszczenia rozważań posłużymy się tylko testem log-rank.

$\langle 45 \text{ lat}; 50 \text{ lat} \rangle$ vs $\langle 50 \text{ lat}; 55 \text{ lat} \rangle$

Wspólne dla warstw	
Statystyka Chi-kwadrat	0,588444784
Ilość stopni swobody	1
Wartość p	0,443021142

$\langle 45 \text{ lat}; 50 \text{ lat} \rangle$ vs $\langle 55 \text{ lat}; 60 \text{ lat} \rangle$

Wspólne dla warstw	
Statystyka Chi-kwadrat	8,944721898
Ilość stopni swobody	1
Wartość p	0,002782725

$\langle 50 \text{ lat}; 55 \text{ lat} \rangle$ vs $\langle 55 \text{ lat}; 60 \text{ lat} \rangle$

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.

26.4 REGRESJA PROPORCJONALNEGO HAZARDU COX'A

Okno z ustawieniami opcji Regresji Cox'a wywołujemy poprzez menu Statystyka→Analiza przeżycia→Regresja PH Cox'a

Regresja PH Cox'a

Analiza statystyczna : Regresja proporcjonalnego hazardu Cox'a

▼ Czas przeżycia

- 1-czas przeżycia (weeks)
- 2-status (0=ucięta, 1=kompletna)
- 3-pleć
- 4-log WBC
- 5-Rx

▼ Zmienna X1,X2,...

- 1-czas przeżycia (weeks)
- 2-status (0=ucięta, 1=kompletna)
- 3-pleć
- 4-log WBC
- 5-Rx

▼ Zmienna ucinana

- 1-czas przeżycia (weeks)
- 2-status (0=ucięta, 1=kompletna)
- 3-pleć
- 4-log WBC
- 5-Rx

1 Kompletne

0,05 poziom istotności

Opcje testu

Interakcje Zmienne fikcyjne

Kodowanie efektów Dołącz reszty

Dołącz hazard wynikowy

Średnia + SD

Zaawansowane

Filtr danych

Wybór danych spełniających odpowiednie warunki +

Reguły połączone są koniunkcją -

podstawowy wielokrotny AND

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

OK Zamknij

Regresja Cox'a, znana również jako model proporcjonalnego hazardu Cox'a (Cox PH model) - Cox D.R. (1972)[47], jest najszersze stosowanym podejściem regresyjnym w analizie przeżycia. Pozwala na zbadanie wpływu wielu zmiennych niezależnych (X_1, X_2, \dots, X_k) na czas przeżycia. Jest to podejście w pewnym sensie nieparametryczne, a więc obarczone niewieloma założeniami - stąd też wynika jego popularność. Nie musi być znana natura ani kształt funkcji hazardu czy przeżycia, a najważniejszy warunek to założenie, które dotyczy również większości parametrycznych modeli przeżycia czyli proporcjonalność hazardu.

Funkcja, na której oparty jest model proporcjonalnego hazardu Cox'a opisuje hazard wynikowy i jest produktem dwóch wielkości, z których tylko jedna jest zależna od czasu (t):

$$h(t, X_1, X_2, \dots, X_k) = h_0(t) \cdot \exp \left(\sum_{i=1}^k \beta_i X_i \right),$$

gdzie:

$h(t, X_1, X_2, \dots, X_k)$ –wynikowy hazard opisujący zmieniające się w czasie ryzyko zależne od innych czynników np. sposobu leczenia,

$h_0(t)$ –hazard bazowy, czyli hazard przy założeniu, że wszystkie zmienne objaśniające są równe zero,

$\sum_{i=1}^k \beta_i X_i$ –kombinacja (najczęściej liniowa) zmiennych niezależnych i parametrów modelu,

X_1, X_2, \dots, X_k –zmienne objaśniające, niezależne od czasu,

$\beta_1, \beta_2, \dots, \beta_k$ –parametry.

Zmienne fikcyjne i interakcje w modelu

Omówienie przygotowania zmiennych fikcyjnych i interakcji przedstawiono w rozdziale 23.1 [Przygotowanie zmiennych do analizy w modelach wielowymiarowych.](#)

Korekcja rang wiązanych w regresji Cox'a oparta jest na metodzie Breslow'a[27]

Model można sprowadzić do postaci liniowej:

$$\ln \left(\frac{h(t, X_1, X_2, \dots, X_k)}{h_0(t)} \right) = \sum_{i=1}^k \beta_i X_i.$$

Rozwiązaniem równania jest wówczas wektor ocen parametrów β_1, \dots, β_k nazywanych **współczynnikami regresji**:

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}.$$

Współczynniki te szacowane są poprzez tzw. **"częściową" metodę największej wiarygodności**. Jest to metoda "częściowa" ponieważ poszukiwanie maksimum funkcji wiarygodności L (w programie użyto algorytm iteracyjny Newton-Raphson) odbywa się tylko dla danych kompletnych, dane ucięte są w tym algorytmie uwzględniane, ale nie bezpośrednio.

Każdy współczynnik obarczony jest pewnym błędem szacunku. Wielkość tego błędu wyliczana jest ze wzoru:

$$SE_b = \sqrt{\text{diag}(H^{-1})_b}$$

gdzie:

$\text{diag}(H^{-1})$ to główna przekątna macierzy kowariancji.

Uwaga!

Budując model należy pamiętać, że liczba obserwacji powinna być dziesięciokrotnie większa lub równa stosunkowi liczby szacowanych parametrów modelu (k) i mniejszej z porcji licznosci uciętych lub kompletnych (p), czyli ($n \geq 10k/p$), Peduzzi P. i inni (1995)[130].

Uwaga!

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.

Uwaga!

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).

26.4.1 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 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 HR interpretujemy następująco:

- $HR > 1$ oznacza stymulujący wpływ badanej zmiennej niezależnej na wystąpienie niepożądanego zdarzenia, tj. mówi o ile wzrasta ryzyko na wystąpienie niepożądanego zdarzenia, gdy zmienna niezależna wzrasta o jeden poziom.
- $HR < 1$ oznacza destymulujący wpływ badanej zmiennej niezależnej na wystąpienie niepożądanego zdarzenia, tj. mówi o ile spada ryzyko na wystąpienie niepożądanego zdarzenia, gdy zmienna niezależna wzrasta o jeden poziom.
- $HR \approx 1$ oznacza, że badana zmienna niezależna nie ma wpływu na wystąpienie niepożądanego zdarzenia.

Uwaga!

Jeśli analizę przeprowadzamy dla modelu innego niż liniowy, lub uwzględniamy interakcję, wówczas analogicznie jak w modelu [regresji logistycznej](#) na podstawie ogólnego wzoru możemy wyliczyć odpowiedni HR zmieniając formułę będącą kombinacją zmiennych niezależnych.

26.4.2 Weryfikacja modelu

Istotność statystyczna poszczególnych zmiennych w modelu (istotność ilorazu hazardów)

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.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : \beta_i = 0, & \quad \text{lub równoważnie:} \quad \mathcal{H}_0 : HR_i = 1, \\ \mathcal{H}_1 : \beta_i \neq 0. & \quad \mathcal{H}_1 : HR_i \neq 1. \end{aligned}$$

Statystykę testową testu Walda wyliczamy według wzoru:

$$\chi^2 = \left(\frac{b_i}{SE_{b_i}} \right)^2$$

Statystyka ta ma asymptotycznie (dla dużych licznosci) [rozkład \$\chi^2\$](#) z 1 stopniem swobody . Wyznaczoną na podstawie [statystyki testowej wartość \$p\$](#) porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha & \implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha & \implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Jakość zbudowanego modelu

Dobry model powinien spełniać dwa podstawowe warunki: powinien być dobrze dopasowany i możliwie jak najprostszy. Jakość modelu proporcjonalnego hazardu Cox'a możemy ocenić kilkoma ogólnymi miarami, które opierają się na:

L_{FM} –maksimum funkcji wiarygodności modelu pełnego (z wszystkimi zmiennymi),

L_0 –maksimum funkcji wiarygodności modelu zawierającego jedynie wyraz wolny,

d –obserwowanej liczbie niepożądanych zdarzeń.

- **Kryteria informacyjne** opierają się na entropii informacji niesionej przez model (niepewności modelu) tzn. szacują utraconą informację, gdy dany model jest używany do opisu badanego zjawiska. Powinniśmy zatem wybierać model o minimalnej wartości danego kryterium informacyjnego.

AIC , $AICc$ i BIC jest rodzajem kompromisu pomiędzy dobrocią dopasowania i złożonością. Drugi element sumy we wzorach na kryteria informacyjne (tzw. funkcja straty lub kary)

mierzy prostotę modelu. Zależy on od liczby parametrów w modelu (k) i liczby obserwacji kompletnych (d). W obu przypadkach element ten rośnie wraz ze wzrostem liczby parametrów i wzrost ten jest tym szybszy im mniejsza jest liczba obserwacji.

Kryterium informacyjne nie jest jednak miarą absolutną, tzn. jeśli wszystkie porównywane modele źle opisują rzeczywistość w kryterium informacyjnym nie ma sensu szukać ostrzeżenia.

- Kryterium informacyjne Akaikego (*ang. Akaike information criterion*)

$$AIC = -2 \ln L_{FM} + 2k,$$

Jest to kryterium asymptotyczne - odpowiednie dla dużych prób.

- Poprawione kryterium informacyjne Akaikego

$$AIC_c = AIC + \frac{2k(k+1)}{d-k-1},$$

Poprawka kryterium Akaikego dotyczy wielkości próby (liczby zdarzeń niepożądanych), przez co jest to miara rekomendowana również dla prób o małych licznosciach.

- Bayesowskie kryterium informacyjne Schwartza (*ang. Bayes Information Criterion* lub *Schwarz criterion*)

$$BIC = -2 \ln L_{FM} + k \ln(d),$$

Podobnie jak poprawione kryterium Akaikego uwzględnia wielkość próby (liczbę zdarzeń niepożądanych) - Volinsky i Raftery 2000r[163]

- **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 $(0; 1)$, gdzie wartości bliskie 1 oznaczają doskonałe dopasowanie modelu, 0 – zupełny brak dopasowania. Współczynnik R_{Pseudo}^2 wyliczamy z wzoru:

$$R_{Pseudo}^2 = 1 - \frac{\ln L_{FM}}{\ln L_0}.$$

Ponieważ współczynnik R_{Pseudo}^2 nie przyjmuje wartości 1 i jest wrażliwy na liczbę zmiennych w modelu, wyznacza się jego poprawioną wartość:

$$R_{Nagelkerke}^2 = \frac{1 - e^{-(2/d)(\ln L_{FM} - \ln L_0)}}{1 - e^{(2/d) \ln L_0}} \quad \text{lub} \quad R_{Cox-Snell}^2 = 1 - e^{\frac{(-2 \ln L_0) - (-2 \ln L_{FM})}{d}}.$$

- **Istotność statystyczna wszystkich zmiennych w modelu**

Podstawowym narzędziem szacującym istotność wszystkich zmiennych w modelu jest **test ilorazu wiarygodności**. Test ten weryfikuje hipotezę:

$$\begin{aligned} \mathcal{H}_0 &: \text{wszystkie } \beta_i = 0, \\ \mathcal{H}_1 &: \text{istnieje } \beta_i \neq 0. \end{aligned}$$

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 licznosci) **rozkład χ^2** z k stopniami swobody.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

- **AUC - pole pod krzywą ROC** – Krzywa ROC – zbudowana w oparciu o informację o wystąpieniu zdarzenia lub jego braku oraz kombinację zmiennych niezależnych i parametrów modelu – pozwala na ocenę zdolności zbudowanego modelu regresji PH Coxa do klasyfikacji przypadków do dwóch grup: (1 – zdarzenie) i (0 – brak zdarzenia). Powstała w ten sposób krzywa, a w szczególności pole pod nią, obrazuje jakość klasyfikacyjną modelu. Gdy krzywa ROC pokrywa się z przekątną $y = x$, to decyzja o przyporządkowaniu przypadku do wybranej klasy (1) lub (0) podejmowana na podstawie modelu jest tak samo dobra jak losowy podział badanych przypadków do tych grup. Jakość klasyfikacyjna modelu jest dobra, gdy krzywa znajduje się znacznie powyżej przekątnej $y = x$, czyli gdy pole pod krzywą ROC jest znacznie większe niż pole pod prostą $y = x$, zatem większe niż 0.5

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: AUC = 0.5, \\ \mathcal{H}_1 &: AUC \neq 0.5.\end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{AUC - 0.5}{SE_{0.5}},$$

gdzie:

$SE_{0.5}$ – błąd pola.

Statystyka Z ma asymptotycznie (dla dużych licznosci) **rozkład normalny**.

Wyznaczoną na podstawie **statystyki testowej wartość p** porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Dodatkowo, dla krzywej ROC podawana jest proponowana wartość **punktu odcięcia** kombinacji zmiennych niezależnych i parametrów modelu.

26.4.3 Analiza reszt modelu

Analiza reszt modelu pozwala na weryfikację jego założeń. Głównym jej celem w regresji Cox'a jest lokalizacja wartości odstających i badanie proporcjonalności hazardu. Standardowo w modelach regresji reszty oblicza się jako różnice obserwowanych i przewidywanych przez model wartości zmiennej zależnej. Jednakże w przypadku występowania zmiennych uciętych taka koncepcja wyznaczania reszt nie jest odpowiednia. W programie można analizować reszty opisywane jako: Martingale, Deviance, Schoenfeld. Reszty te można wyrysowywać względem czasu lub zmiennych niezależnych.

Założenie proporcjonalności hazardu

Opracowano szereg graficznych metod pozwalających na ocenę adekwatności modelu proporcjonalnego hazardu (Lee i Wang 2003[98]). Najszerszej stosowane są metody bazujące na resztach modelu. Podobnie jak inne graficzne metody oceny proporcjonalności hazardu jest to metoda subiektywna. Aby założenie proporcjonalnego hazardu było spełnione reszty względem czasu nie powinny układać się w żaden wzór ale powinny być losowo rozłożone w okolicach wartości 0.

Martingale - reszty te mogą być interpretowane jako różnica w czasie $[0, t]$ pomiędzy obserwowaną licznością zdarzeń niepożądanych i przewidywaną przez model licznością tych zdarzeń. Wartość oczekiwana dla tych reszt to 0, ale ich wadą jest skośny rozkład utrudniający interpretację ich wykresu (zawierają się w przedziale od $-\infty$ do 1).

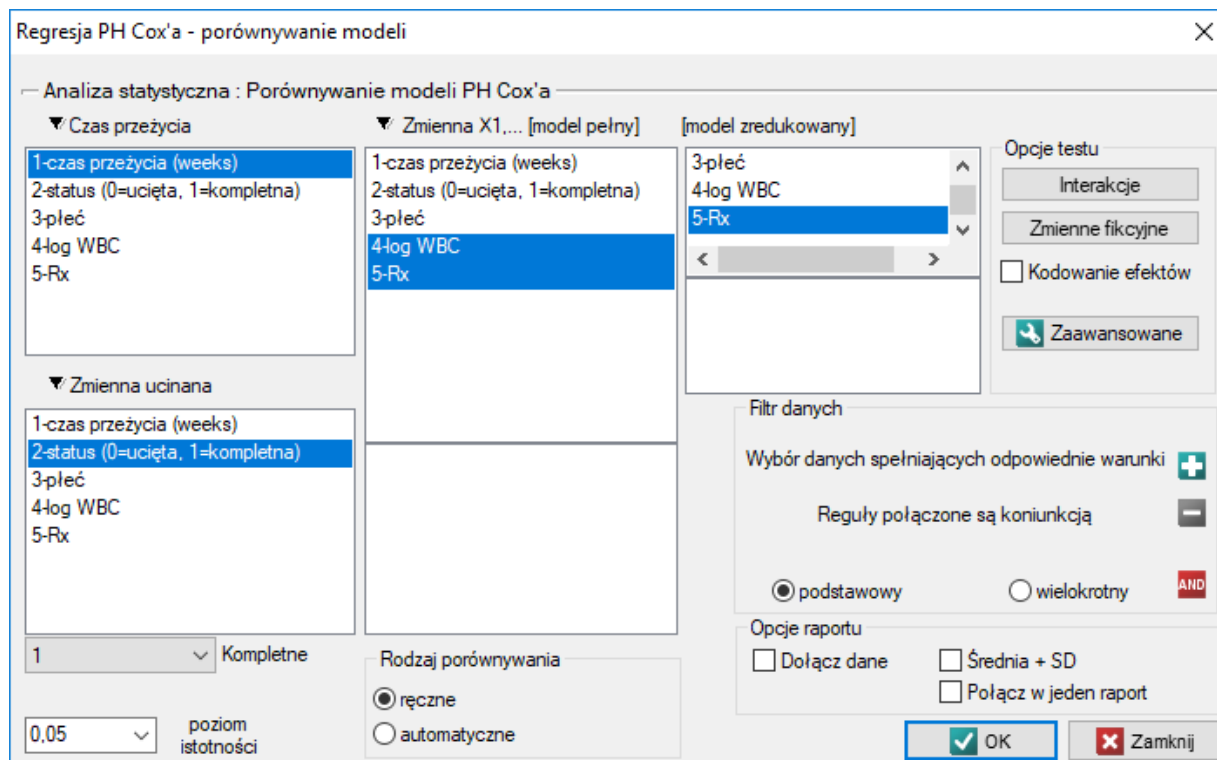
Deviance - podobnie jak Martingale asymptotycznie uzyskują wartość oczekiwaną równą 0, ale rozkładają się symetrycznie wokół zera z odchyleniem standardowym równym 1, gdy dopasowywany model jest odpowiedni. Wartość Deviance jest dodatnia, gdy obiekt badany przeżywa krócej niż oczekiwany na podstawie modelu czas, a ujemna gdy ten czas jest dłuższy. Analiza tych reszt jest wykorzystywana w badaniu proporcjonalności hazardu - ale jest to przede wszystkim narzędzie identyfikujące wartości odstające. W raporcie reszt, te z nich, które są oddalone o więcej niż 3 odchylenia standardowe od 0 oznaczone są kolorem czerwonym.

Schoenfeld - reszty te wyliczane są oddzielnie dla każdej zmiennej niezależnej i definiowane tylko dla obserwacji kompletnych. Dla każdej zmiennej niezależnej suma reszt Schoenfeld'a i ich wartość oczekiwana to 0. Zaletą przedstawiania reszt względem czasu dla każdej zmiennej jest możliwość zidentyfikowania zmiennej nie spełniającej w modelu założenia proporcjonalności hazardu. Jest to ta zmienna, której wykres reszt układa się w systematyczny wzór (najczęściej bada się tu liniową zależność reszt od czasu). Równomierne rozłożenie punktów względem wartości 0 świadczy o braku zależności reszt od czasu - czyli spełnieniu założenia proporcjonalności hazardu przez daną zmienną w modelu.

Gdy dla którejkolwiek zmiennej w modelu Cox'a nie jest spełnione założenie proporcjonalności hazardu, jednym z możliwych rozwiązań jest wykonanie analizy Cox'a oddzielnie dla każdego poziomu tej zmiennej.

26.5 PORÓWNYWANIE MODELI REGRESJI PH COX'A

Okno z ustawieniami opcji porównywania modeli wywołujemy poprzez menu Statystyka → Analiza przeżycia → Regresja PH Cox'a – porównywanie modeli



Ze względu na możliwość jednoczesnej analizy wielu zmiennych niezależnych w jednym modelu regresji Cox'a, istnieje 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.

Porównując modele z różną liczbą zmiennych niezależnych zwracamy uwagę na kryteria informacyjne (AIC , $AICc$, BIC) oraz współczynniki dopasowania modelu (R_{Pseudo}^2 , $R_{Nagelkerke}^2$, $R_{Cox-Snell}^2$). Modele porównywane są również przy użyciu testu ilorazu wiarygodności.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: L_{FM} = L_{RM}, \\ \mathcal{H}_1 &: L_{FM} \neq L_{RM},\end{aligned}$$

gdzie:

L_{FM} , L_{RM} – maksimum funkcji wiarygodności w porównywanych modelach (pełnym i zredukowanym).

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 licznosci) 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 :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Decyzję o tym, który model wybrać podejmujemy na podstawie wielkości: AIC , $AICc$, BIC , R_{Pseudo}^2 , $R_{Nagelkerke}^2$, $R_{Cox-Snell}^2$ oraz wyniku testu ilorazu wiarygodności porównującego kolejno powstające (sąsiednie) modele. Jeśli porównywane modele nie różnią się istotnie, to 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 jest istotna statystycznie oznacza to, że jeden z nich (ten z większą liczbą zmiennych, o większym R^2) jest istotnie lepszy niż drugi.

W programie PQStat porównywanie modeli możemy 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.

- **Automatyczne** porównywanie modeli jest wykonywane w kilku krokach:
 - krok 1 Zbudowanie modelu z wszystkich zmiennych.
 - krok 2 Usunięcie jednej zmiennej z modelu. Usuwana zmienna to ta, która ze statystycznego punktu widzenia wnosi do aktualnego modelu najmniej informacji.
 - krok 3 Porównanie modelu pełnego i zredukowanego.
 - krok 4 Usunięcie kolejnej zmiennej z modelu. Usuwana zmienna to ta, która ze statystycznego punktu widzenia wnosi do aktualnego modelu najmniej informacji.

krok 5 Porównanie modelu wcześniejszego i nowo zredukowanego.

...

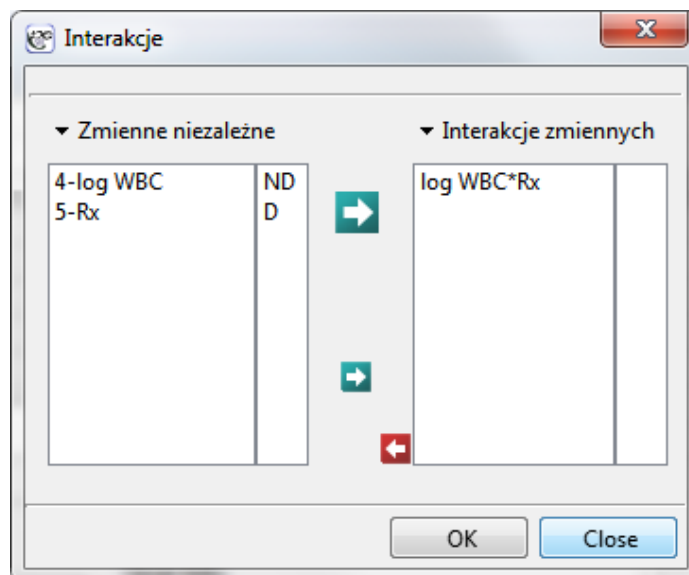
W ten sposób powstaje wiele, coraz mniejszych modeli. Ostatni model zawiera tylko 1 zmienną niezależną.

PRZYKŁAD 26.2. (plik: remisjaBiałaczka.pqs)

Analiza opiera się na danych dotyczących białaczki opisanych w pracy Freireich i innych 1963[63] i analizowanych dalej przez wielu autorów min. Kleinbaum i Klein 2005[92]. 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 jest nastąpieniem zdarzenia niekorzystnego - 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: grupy leczonej tradycyjnie (oznaczonej jako 1 i czasami nazywanej "grupą placebo") i grupy leczonej nową metodą (oznaczonej jako 0). Zebrano informację o płci pacjentów (1=mężczyzna, 0=kobieta) oraz o wartościach wskaźnika określającego liczbę 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 wnikających (confounder) i interakcji. W analizie uwagę skupimy na zmiennej "Rx (1=placebo, 0=new treatment)", zmienną "log WBC" umieścimy w modelu jako możliwy czynnik wnikający (modyfikujący uzyskany efekt). By ocenić ewentualny wpływ interakcji "Rx" i "log WBC" rozważymy także trzecią zmienną będącą iloczynem zmiennych wchodzących w skład interakcji. Zmienną tę dołączymy do modelu wybierając w oknie analizy przycisk Interakcje i dokonując odpowiednich ustawień.



Budujemy trzy modele Coxa:

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 wnikającą "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 interakcji tych zmiennych: "Rx × 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*Rx	-0,342195	0,5197406	-1,360867	0,6764778	0,4334850	0,5102838			

Zmienna mówiąca o interakcji "Rx" i "log WBC", zawarta w modelu C jest w nim nieistotna ($p=0.5103$) według testu Walda. Możemy więc uznać za zbyt dalsze rozważanie w modelu interakcji tych dwóch zmiennych. Podobne wyniki uzyskamy porównując testem Ilorazu Wiarygodności model C z modelem B. Porównanie to możemy wykonać wybierając menu Regresja Cox'a – porównywanie modeli, uzyskamy wówczas wynik nieistotny ($p=0.5134$), co oznacza, że model C (model z interakcją) jest NIE-istotnie lepszy niż model B (model bez interakcji).

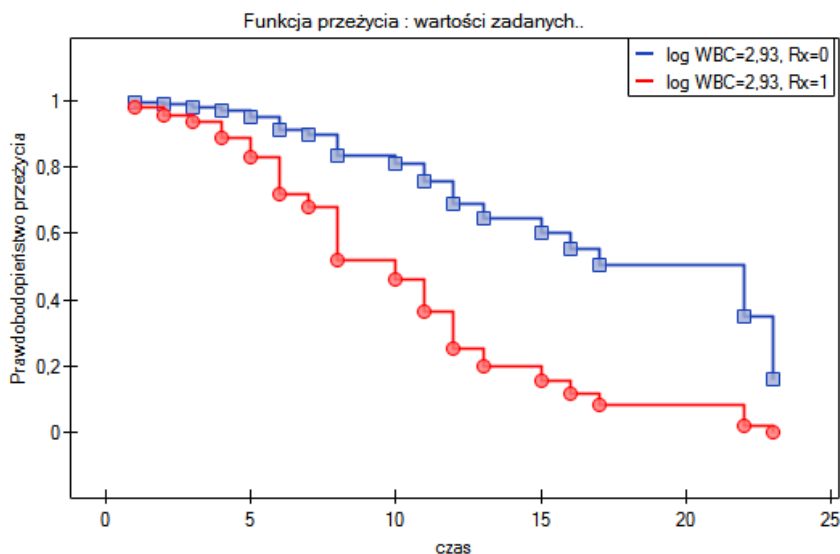
Chi-kwadrat - porównanie modeli	0,427079874
Ilość stopni swobody	1
Wartość p	0,513425299

Odrzucamy więc model C przechodząc do rozważania modelu B i modelu A.

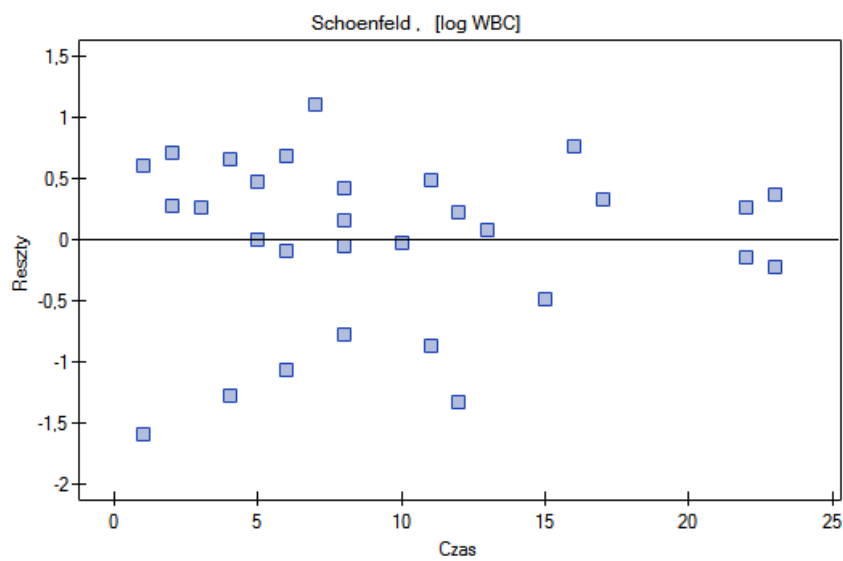
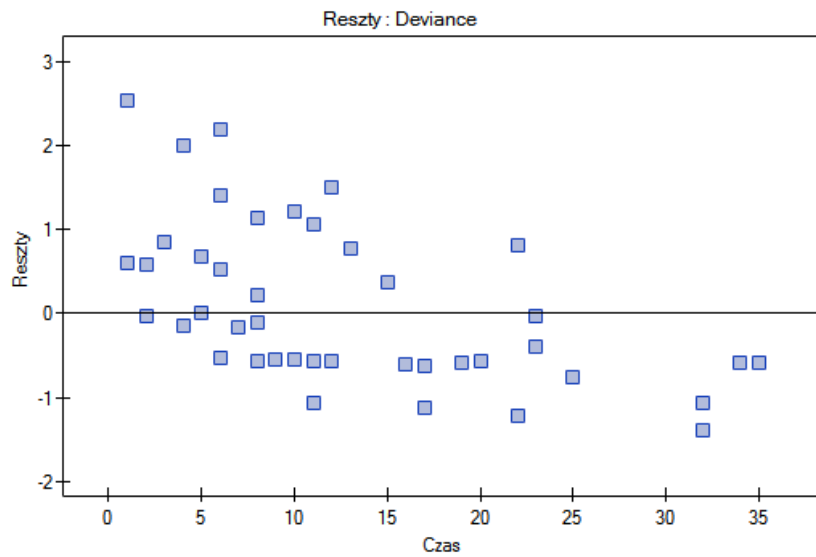
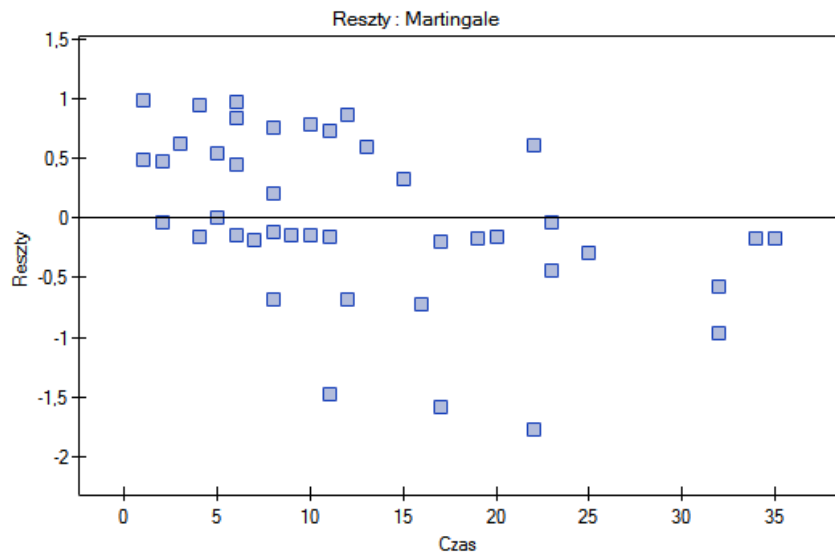
HR dla "Rx" w modelu B wynosi 3.65, co oznacza, że hazard dla "grupy placebo" jest około 3.6 większy niż dla grupy pacjentów leczonych nową metodą. Model A zawiera tylko zmienną "Rx", przez co nazywany jest zwykle modelem "surowym" (ang. crude model) ponieważ ignoruje efekt potencjalnych zmiennych wnikających. W modelu tym HR dla "Rx" wynosi 4.52 i jest sporo większe niż w modelu B. Przyjrzyjmy się jednak nie tylko punktowym wartościom estymatora HR ale również 95% przedziałowi ufności dla tych estymatorów. Przedział dla "Rx" w modelu A ma szerokość 8.06 (10.09 minus 2.03) a w modelu B jest węższy: 6.74 (8.34 minus 1.60). Dlatego model B daje bardziej precyzyjną estymację HR niż model A. By ostatecznie zdecydować, który model (model A czy model B) będzie lepszy w oszacowaniu efektu leczenia ("Rx") ponownie wykonujemy analizę porównawczą modeli w module Regresja Cox'a – porównywanie modeli. Tym razem test Ilorazu Wiarygodności daje wynik istotny ($p<0.0001$), co ostatecznie potwierdza wyższość modelu B. Jest to model o najniższej wartości kryteriów informacyjnych ($AIC=148.6$, $AICc=149$, $BIC=151.4$) i wysokich wartościach dopasowania modelu ($Pseudo R^2_{McFadden} = 0.2309$, $R^2_{Nagelkerke} = 0.7662$, $R^2_{Coxa-Snella} = 0.7647$).

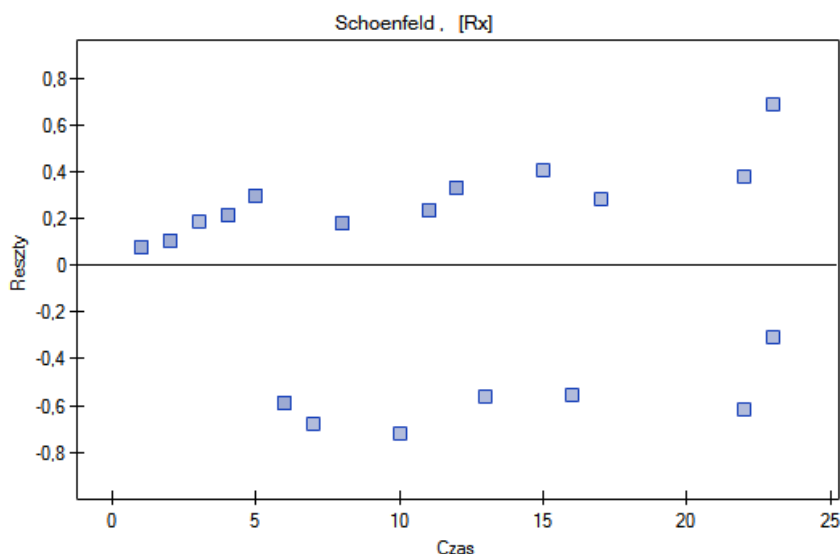
Porównywanie modeli PH Cox'a	
Czas analizy	0,11 sek.
Analizowane zmienne	czas przeżycia (tygodnie);st:
Poziom istotności	0,05
Zmienna grupująca	Rx(0;1)
Liczba zmiennych w modelu 1	2
Kryterium zbieżności zostało osiągnięte	
-2 Log Likelihood	144,558519789
AIC - kryterium Akaikiego	148,558519789
AICc - skorygowane kryterium Akaikiego	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 Akaikiego	174,759244379
AICc - skorygowane kryterium Akaikiego	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

Dopełnieniem powyższej analizy jest przedstawienie dla modelu B krzywych przeżycia (pozostawania w remisji) obu grup: leczonej nowym lekiem i tradycyjnie, skorygowanych o wpływ "log WBC". Na wykresie obserwujemy różnice między grupami, występujące w poszczególnych punktach osi czasu. By wyrysować takie krzywe, po wybraniu opcji Dołącz wykres zaznaczamy opcję Funkcja przeżycia: w podgrupach ... a następnie, by szybko zbudować wykres dwóch krzywych wybieram Szybkie podgrupy i wskazuję zmienną Rx. Opcja Zaawansowane podgrupy pozwala na zbudowanie dowolnej liczby dowolnie zdefiniowanych krzywych.



Na koniec ocenimy założenia regresji Cox'a analizując reszty modelu względem czasu.





Nie obserwujemy odstających punktów, aczkolwiek reszty Martingale i Deviance są coraz niższe dla coraz dłuższego czasu. Natomiast reszty Schoenfeld'a przedstawiają symetryczny rozkład względem czasu. W przypadku reszt Schoenfeld'a analizę wykresu można wesprzeć różnymi testami mogącymi ocenić, czy punkty wykresu reszt układają się w pewien wzór np. zależność liniową. By tej analizy dokonać należy przekopiować reszty Schoenfeld'a oraz czas do arkusza danych i przetestować szukany rodzaj zależności. Wynik takiego testu dla każdej zmiennej wskazuje na spełnienie założenia proporcjonalności hazardu przez daną zmienną w modelu - gdy jest nieistotny statystycznie lub złamanie tego założenia - gdy jest istotny statystycznie. W rezultacie zmienną łamiącą założenia regresji proporcjonalnego hazardu Cox'a można wyłączyć z modelu. W przypadku zmiennych "Log WBC" i "Rx" symetryczny rozkład reszt sugeruje spełnienie założenia proporcjonalności hazardu przez te zmienne. Potwierdzeniem tego może być sprawdzenie zależności np. liniowej Pearsona, lub monotonicznej Spearmana dla tych reszt i czasu.

Dalej możemy dodać do modelu zmienną płeć. Musimy jednak postępować ostrożnie, gdyż na podstawie różnych źródeł wiadomo, że płeć może wpływać na funkcję przeżycia (pozostawania w remisji) dotyczącą białaczki w ten sposób, że funkcje przeżycia mogą układać się nieproporcjonalnie względem siebie wzdłuż osi czasu. Wykonujemy więc model Cox'a dla trzech zmiennych: zmiennej "Płeć", "Rx" i "log WBC". Zanim dokonamy interpretacji współczynników modelu sprawdzimy reszty Schoenfeld'a. Reszty przedstawiamy na wykresach, a wyniki reszt Schoenfeld'a oraz czasu przekopujemy z raportu do nowego arkusza, gdzie sprawdzimy występowanie zależności monotonicznej Spearmana. W wyniku uzyskujemy wartość $p=0.0259$ (dla zależności czasu i reszt Schoenfeld'a dla płci), $p=0.6192$ (dla zależności czasu i reszt Schoenfeld'a dla log WBC) i $p=0,1490$ (dla zależności czasu i reszt Schoenfeld'a dla Rx) co potwierdza naruszenie założenia proporcjonalności hazardu przez zmienną płeć. Model Cox'a zbudujemy więc oddzielnie dla kobiet i mężczyzn. W tym celu analizę wykonamy dwukrotnie z włączonym filtrem danych. Za pierwszym razem filtr będzie wskazywał płeć żeńską (0), a za drugim razem płeć męską (1).

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

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

27 META-ANALIZA

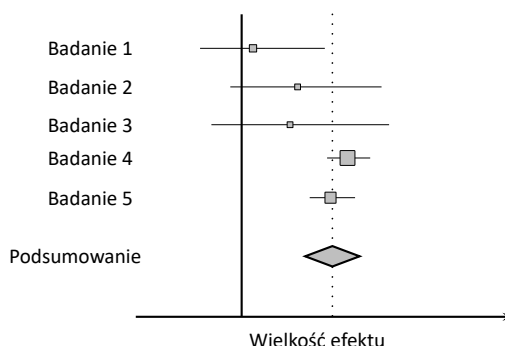
Liczba pojawiających się prac naukowych w ostatniej dekadzie bardzo wzrosła. Jest to związane z szeregiem korzyści, ale utrudnia nadążenie za wciąż pojawiającymi się nowymi informacjami. Jeśli na przykład lekarz zastosował by nowy sposób leczenia dla swoich pacjentów na podstawie przeczytanej pracy naukowej mógłby popełnić błąd. Błąd może wynikać z faktu opublikowania całego szeregu innych prac, które przeczą skuteczności tego leczenia. Aby decyzja podjęta przez lekarza obarczona była jak najmniejszym błędem, powinien on przeczytać większość prac naukowych, jakie ukazały się w tym temacie. W rezultacie ciągnęła konieczność przeglądania coraz to większych zbiorów literatury zajęłoby lekarzowi tak dużo czasu, że mogłoby go zabraknąć na leczenie pacjentów. Meta-analiza pozwala na szybkie dokonanie takiego przeglądu ponieważ jest rezultatem przeprowadzonego przez jej autora szerokiego przeglądu literatury i wykonaniu statystycznego podsumowania opisanych tam wyników.

Meta-analiza w programie PQStat wykonywana jest na podstawie następujących miar:

- Różnica średnich,
- d Cohen,
- g Hedges,
- Iloraz średnich,
- Iloraz szans (OR),
- Relatywne ryzyko (RR),
- Różnica ryzyka (RD),
- Współczynnik Pearsona,
- AUC dla krzywej ROC,
- Proporcja.

27.1 Wprowadzenie

Najbardziej znanym obrazem związanym z meta-analizą jest wykres leśny przedstawiający wyniki poszczególnych badań wraz z ich podsumowaniem.



Aby można było dokonać wspólnego podsumowania wybranej literatury musi być ona spójna w opisie, a podane tam miary muszą być tożsame.

By dana praca naukowa mogła być wykorzystana w meta-analizie powinna opisywać:

Efekt końcowy czyli pewnego rodzaju miarę statystyczną wskazującą na uzyskany w pracy wynik (efekt). W rzeczywistości mogą to być różnego rodzaju wielkości np. różnica pomiędzy wartościami średnimi, iloraz szans, relatywne ryzyko itp.

Błąd efektu czyli **SE** (*ang. Standard Error*) pozwalający określić precyzję przeprowadzonego badania. Owa precyzja nadaje **wagę badania**. Im mniejszy jest błąd (SE), tym większa jest precyzja danego badania i tym przypisana waga będzie wyższa, przez co dane badanie będzie miało większy wpływ na wyniki meta-analizy.

Liczność jest to liczba obiektów na jakiej przeprowadzone zostało badanie.

Uwaga!

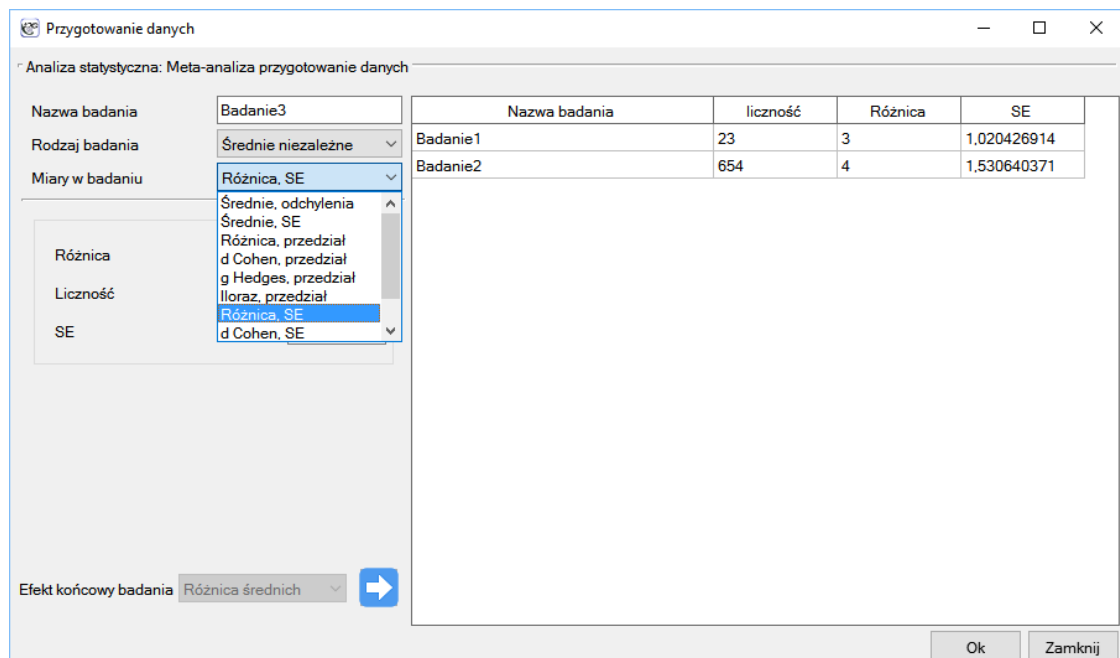
Często zdarza się, że praca naukowa nie zawiera wszystkich wymienionych wyżej elementów, wówczas należy szukać w niej danych, na podstawie których wyliczenie tych miar będzie możliwe.

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 w przypadku kiedy w danej pracy nie podano wprost tych trzech miar.

Okno przygotowania danych wywołujemy poprzez menu:

Statystyki zaawansowane → Meta-analiza → Przygotowanie danych.



Nazwa badania	liczność	Różnica	SE
Badanie1	23	3	1,020426914
Badanie2	654	4	1,530640371

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 Błąd efektu i Liczność zostają wyliczone w oparciu o miary pozyskane z odpowiedniej pracy naukowej. Miary zawarte w badaniach na podstawie których mogą być wyliczone efekty końcowe, przedstawia poniższa tabela:

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 Hedges, przedział	c
	Iloraz, przedział	e
	Różnica, SE	a
	d Cohen, SE	b
	g Hedges, SE Iloraz, SE	c e
Średnie zależne	Średnie, odchylenia	a, b, c
	Średnie, SE	a, b, c
	Różnica, przedział	a
	d Cohen, przedział	b
	g Hedges, przedział	c
	Różnica, odchylenie	a, b, c
	Różnica, SE	a, b, c
	d Cohen, SE g Hedges, SE	b 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 Hedges, przedział	b, c, f
	OR, SE	b, c, f
	RR, SE	g
	RD, SE	h
	d Cohen, SE g Hedges, SE	b, c, f b, c, f
Korelacja	Współczynnik	i
	Współczynnik, przedział	i
	Współczynnik, SE	i
Krzywa ROC	AUC, SE	j
	AUC, przedział	j
Jedna proporcja	Wyróżniona licznosc	k
	Wyróżniona proporcja	k

gdzie poszczególne efekty końcowe, to:

- a** - Różnica średnich
- b** - d Cohen
- c** - g Hedges
- 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[38].

PRZYKŁAD 27.1. 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 mieć możliwość wyliczenia 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 gdy podane są odpowiednie licznosci w czterech kategoriach tzn. tabela 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: licznosci: (palący i chorujący)=100, (palący i niechorujący)=73, (niepalący i chorujący)=80, (niepalący i niechorujący)=70,

Badanie 2: licznosci: (palący i chorujący)=182, (palący i niechorujący)=172, (niepalący i chorujący)=180, (niepalący i niechorujący)=172,

Badanie 3: licznosci: (palący i chorujący)=157, (palący i niechorujący)=132, (niepalący i chorujący)=125, (niepalący i niechorujący)=201,

Badanie 4: licznosci: (palący i chorujący)=19, (palący i niechorujący)=15, (niepalący i chorujący)=35, (niepalący i niechorujący)=20,

Badanie 5: licznosc: 278, RR[95%CI]=1.03[0.85-1.25],

Badanie 6: licznosc: 560, RR[95%CI]=1.21[1.05-1.40],

Badanie 7: licznosc: 1207, RR[95%CI]=1.04[0.93-1.15],


Badanie 8: licznosc: 214, RR[95%CI]=1.15[0.95-1.40],

Badanie 9: licznosc: 285, RR[95%CI]=1.36[1.03-1.79],

Badanie 10: licznosc: 1968, RR=1.17, SE(lnRR)=0.0437,

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):

Analiza statystyczna: Meta-analiza przygotowanie danych		Analiza statystyczna: Meta-analiza przygotowanie danych		Analiza statystyczna: Meta-analiza przygotowanie danych	
Nazwa badania	Badanie 1	Nazwa badania	Badanie 5	Nazwa badania	Badanie 10
Rodzaj badania	Tabele	Rodzaj badania	Tabele	Rodzaj badania	Tabele
Miary w badaniu	Tabela 2x2	Miary w badaniu	RR, przedział	Miary w badaniu	RR, SE
narażeni - zdarzenie	100	RR	1,03	RR	0,8546
narażeni - brak zdarzenia	73	Liczność	278	Liczność	229
nie narażeni - zdarzenie	80	-95%CI	0,85	SE(lnRR)	0,090359
nie narażeni - brak zdarzenia	70	+95%CI	1,25		
Efekt końcowy badania	Relatywne ryzyko (RI)	Efekt końcowy badania	Relatywne ryzyko (RI)	Efekt końcowy badania	Relatywne ryzyko (RI)

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.

Wartość p , a więc 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. Co więcej, zupełnie mała wielkość efektu może być istotna statystycznie w badaniu dużym, a całkiem duża wielkość efektu może być nieistotna w badaniu małym. Fakt ten związany jest z mocą testów statystycznych. Badając istotność statystyczną sprawdzamy czy efekt w ogóle istnieje, tzn. czy jest różny od zera, a nie czy jest on na tyle duży by miał przełożenie na pożądane skutki. Na przykład fakt, że lek obniża ciśnienie istotnie statystycznie o 1mmHg nie spowoduje że będzie stosowany, gdyż z klinicznego punktu widzenia 1 mmHg jest zbyt małą wielkością. Meta-analiza skupia swoje działanie na wielkości poszczególnych efektów a nie na ich istotności statystycznej. W rezultacie nie ma większego znaczenia, czy prace wykorzystywane w meta-analizie wskazują na istotność statystyczną danego efektu czy też nie.

W programie PQStat, dla każdego badania wyliczana jest istotność statystyczna podana dla stosunku efektu i błędu tego efektu. Jest to podejście asymptotyczne, oparte na rozkładzie normalnym i dedykowane dla licznych grup. Jeśli w cytowanym badaniu do sprawdzenia istotności statystycznej użyto innego testu, uzyskane wyniki mogą się nieco różnić.

27.2 Efekt podsumowujący

W wyniku działania meta-analizy najbardziej pożądanym jej elementem jest podsumowanie zebranych badań, czyli podanie wspólnego efektu M . Takie podsumowanie może odbywać się na dwa sposoby, poprzez wyznaczenie efektu stałego lub efektu zmiennego.

Efekt stały

Wyliczając efekt stały zakładamy, że wszystkie badania w meta-analizie dzielą jeden wspólny prawdziwy efekt. Jeśli więc każde badanie dotyczyłoby tej samej populacji np. tego samego kraju, to chcąc podsumować meta-analizę efektem stałym zakładamy, że w każdym z tych badań prawdziwy (populacyjny) efekt będzie taki sam. Co za tym idzie, wszystkie czynniki, które mogłyby zaburzać wielkość tego efektu są takie same. Na przykład, jeśli na uzyskany efekt może wpływać wiek lub płeć badanych, to w każdym badaniu owe czynniki są podobne. Tak więc różnice w uzyskanych efektach dla poszczególnych badań wynikają tylko z błędów próbkowania (błędów wewnętrznego każdego badania) - czyli wielkości SE .

Efekt stały oszacowuje populacyjny efekt - prawdziwy efekt dla każdego z badań.

Przedział ufności wokół efektu stałego (szerokość rombu w 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 będzie inny dla każdej populacji. Jeśli więc każde badanie dotyczyłoby innego kraju, to chcąc podsumować meta-analizę efektem zmiennym zakładamy, że niektóre czynniki, które mogłyby zaburzać wielkość tego efektu mogą mieć różne wielkości w poszczególnych krajach. Na przykład, jeśli na uzyskany efekt (np. średni wzrost dzietności) może wpływać poziom wykształcenia badanych czy zamożność danego kraju, a owe kraje różnią się tymi czynnikami, to w rezultacie prawdziwy efekt (średni wzrost dzietności) będzie nieco inny w każdym z tych krajów. Tak więc różnice w uzyskanych efektach dla poszczególnych badań wynikają z błędu próbkowania (błędu wewnątrz każdego badania) - czyli wielkości SE , oraz z różnic między badanymi populacjami (wariancji między badaniami - **heterogeniczności** badań) - czyli T^2 . Owa heterogeniczność nie może być zbyt duża, zbyt duże zróżnicowanie pomiędzy badanymi populacjami wskazuje brak podstaw do wspólnego podsumowania.

Efekt zmienny oszacowuje średnią ważoną z prawdziwych (populacyjnych) efektów poszczególnych badań.

Przedział ufności wokół efektu zmiennego (szerokość rombu w wykresie leśnym) zależy od poszczególnych SE oraz od T^2 .

Przedział ufności a przedział predykcji

95% przedział ufności (szerokość rombu w wykresie leśnym) - oznacza, że w 95% przypadkach 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.

Okno z ustawieniami opcji meta-analzy wywołujemy poprzez menu:

Statystyki zaawansowane → Meta-analiza → Podsumowanie.

Meta-analiza podsumowanie

Analiza statystyczna : Meta-analiza podsumowanie

Efekt końcowy: Relatywne ryzyko (RR)

Nazwa badania	Liczność
1-Nazwa badania	1-Nazwa badania
2-Liczność	2-Liczność
3-RR	3-RR
4-SE(lnRR)	4-SE(lnRR)
5-Kolejność publikowania	5-Kolejność publikowania

Relatywne ryzyko (RR)	Błąd standardowy (SE (ln))
1-Nazwa badania	1-Nazwa badania
2-Liczność	2-Liczność
3-RR	3-RR
4-SE(lnRR)	4-SE(lnRR)
5-Kolejność publikowania	5-Kolejność publikowania

0,05 poziom istotności Zm. sortowania: 1-Nazwa badania

Opcje testu

Efekt stały Efekt zmienny

Badanie heterogeniczności
 Badanie wrażliwości
 Badanie asymetrii
 Meta-analiza skumulowana

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny **AND**

Opcje raportu

Dołącz dane Połącz w jeden raport
 Dołącz wykres

W oknie tym, w zależności od wybranego Efektu końcowego można dokonać podsumowania meta-analizy i przeprowadzić podstawowe analizy pozwalające na sprawdzenie jej założeń takich jak: heterogeniczność, błąd publikacyjny (badanie wrażliwości, asymetrii) oraz przeprowadzić meta-analizę skumulowaną.

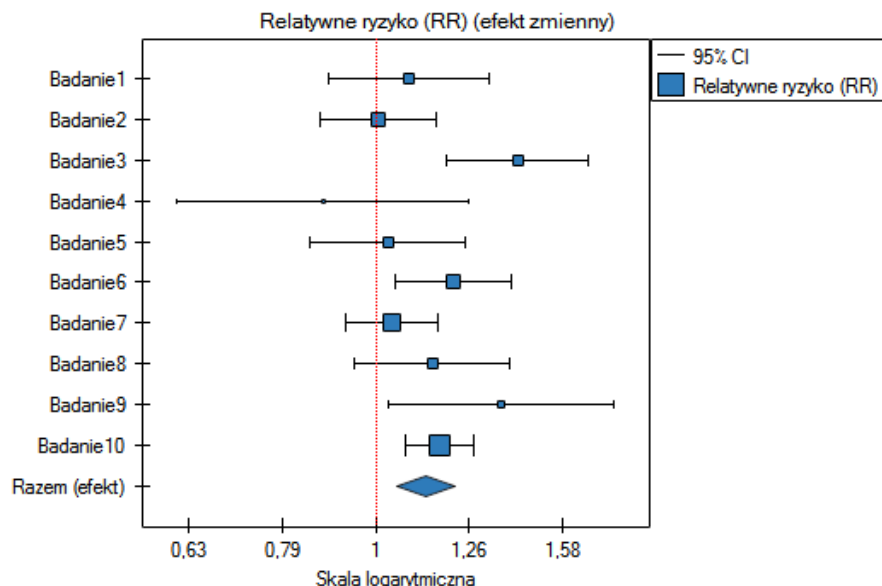
PRZYKŁAD 27.2. (plik MetaanalizaRR.pqs)

Badano ryzyko choroby X dla osób palących i dla niepalących. W niektórych pracach naukowych wskazywano, że ryzyko wystąpienia choroby X jest większe dla palących, a w niektórych pracach nie udowodniono takiego związku. By ustalić czy palenie papierosów ma wpływa na występowanie choroby X zaplanowano przeprowadzenie meta-analizy. Wykonano dokładny przegląd literatury dotyczącej tego tematu i na tej podstawie wytypowano 10 prac naukowych do meta-analizy. Prace te dysponowały danymi, na podstawie których możliwe było wyliczenie relatywnego ryzyka (tzn. ryzyka zachorowania dla palących w stosunku do ryzyka zachorowania dla niepalących) i możliwe było ustalenie błędu jakim jest obciążone podane relatywne ryzyko (tzn. precyzji 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.

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	Relatynw	SE (ln)	-95%CI	+95%CI	Statystyk	Wartość p	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%
Badanie10	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. Uzyskane relatywne ryzyko dla efektu podsumowującego wraz z 95% przedziałem ufności znajduje się powyżej wartości jeden: $RR[95\%CI]=1.13[1.05-1.22]$. 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](#)).

27.3 Wagi poszczególnych badań

Waga w_i badania zależy od obserwowanej zmienności.

Dla efektu stałego zmienność wynika tylko z błędu próbkowania (błędu wewnątrz każdego badania) - czyli wielkości SE :

$$w_i = \frac{1}{SE^2}$$

Dla efektu zmiennego zmienność wynika z błędu próbkowania (błędu wewnątrz każdego badania) - czyli wielkości SE , oraz z różnic pomiędzy badaniami - czyli obserwowanej wariancji T^2 :

$$w_i = \frac{1}{SE^2 + T^2}$$

Na podstawie wag przypisanych poszczególnym badaniom wyznacza się **udział** danego badania w całej analizie. Jest to odsetek jaki stanowi waga danego badania w stosunku do łącznej wagi wszystkich włączonych badań.

27.4 Badanie heterogeniczności

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. nieduże.** 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 starszych, więc w pracach opartych na danych pochodzących głównie od osób młodych uzyskiwany efekt może znacznie odbiegać od prac przeprowadzanych na osobach starszych. Podzielenie zebranych prac na bardziej jednorodnych podgrup pozwoli na dobre oszacowanie efektu sumarycznego dla każdej z tych podgrup oddzielnie.

Badanie heterogeniczności ma na celu sprawdzenie czy zmienność pomiędzy badaniami jest zerowa.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0 : \tau^2 &= 0, \\ \mathcal{H}_1 : \tau^2 &\neq 0, \end{aligned}$$

gdzie:

τ^2 - to wariancja prawdziwych (populacyjnych) efektów poszczególnych badań.

Statystyka testowa ma postać:

$$Q = T^2W + k - 1$$

gdzie:

T^2 - to wariancja obserwowanych efektów,

W - współczynnik wyliczony na podstawie wag przypisanych do poszczególnych badań,

k - liczba badań.

Statystyka ta ma asymptotycznie (dla dużych licznosci) 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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,

jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Uwaga!

- Jeśli uzyskamy wynik jest istotny statystycznie - jest to mocna sugestia, by zrezygnować z wspólnego podsumowywania wszystkich zebranych badań.
- Jeśli uzyskany wynik jest nieistotny statystycznie - możemy podsumować badania wspólnym efektem. Przy czym sugeruje się, by podsumowania dokonywać zwykle efektem zmiennym - zgodnie z poniższym uzasadnieniem.

Uzasadnienie wyboru efektu zmiennego:

Podsumowanie badania efektem zmiennym bierze pod uwagę zmienność pomiędzy badaniami (T^2), natomiast podsumowanie efektem stałym nie bierze pod uwagę tej zmienności. Jeśli jednak T^2 jest małe, to wynik działania modelu z efektem stałym będzie bliski wynikom działania modelu z efektem zmiennym, a gdy $T^2 = 0$, oba modele dadzą dokładnie ten sam wynik.

Dodatkowymi miarami opisującymi heterogeniczność są współczynniki I^2 i H^2 :

$$I^2 = \frac{H^2 - 1}{H^2}, H^2 = \frac{Q}{k - 1}.$$

Współczynnik I^2 określa procent obserwowanej wariancji, jaki wynika z rzeczywistej różnicy w wielkości badanych efektów (graficznie, odzwierciedla stopień zazębienia 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 $I^2 = 0$, wówczas cała obserwowana wariancja wielkości efektów jest "falszywa", więc jeśli w wyznaczonym wokół współczynnika I^2 przedziale ufności znajdzie się wartość 0, uzyskaną wariancję można uznać za nieistotną statystycznie. Natomiast czym wartość I^2 jest bliższa 100%, tym bardziej należy się zastanowić nad rezygnacją z 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, wówczas uzyskaną wariancję można uznać za nieistotną statystycznie, a im wyższa wartość H^2 , tym większa heterogeniczność badań.

Przykład (27.2) 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%

Uzyskano istotny statystycznie wynik statystyki Q ($p=0.04284$). Wariancja obserwowanych efektów nie jest zerowa ($T2=0.0058$), a współczynnik $I2=48.27\%$, wskazuje na przeciętną heterogeniczność badań. Jedynie przedział ufności dla współczynnika $H2$ uznaje za nieistotną zmienność pomiędzy badaniami (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.

27.5 Badanie wrażliwości

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** będącego skutkiem usunięcia danego badania. Badania powinny być na tyle zbliżone, by usunięcie jednego z nich nie zmieniało całkowicie interpretacji efektu podsumowującego.

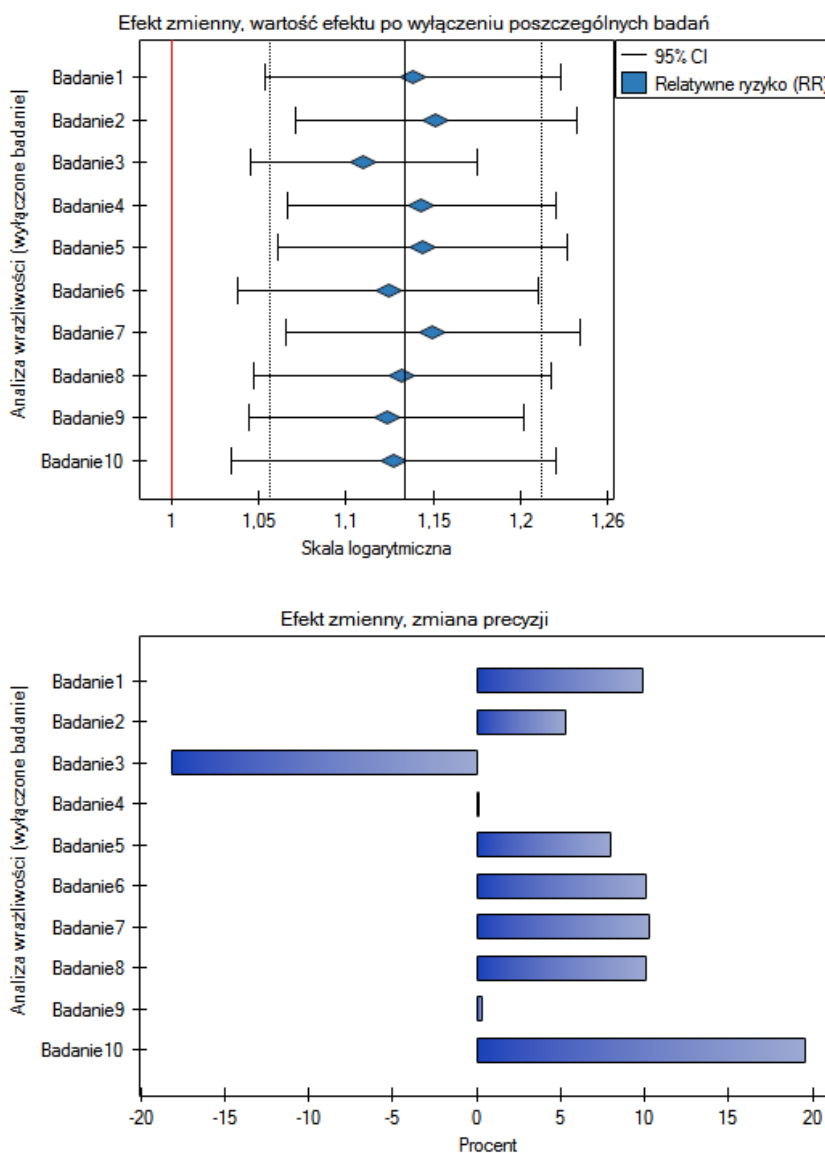
Przypisana do danego badania wartość **udziału pozostałych** definiuje odsetek jaki stanowi łączna waga pozostałych w analizie badań po wykluczeniu danego badania. **Zmiana precyzji** informuje natomiast jak zmieni się precyzja efektu podsumowującego (szerokość przedziału ufności), gdy dane badanie zostanie wykluczone z analizy.

Dobrym zilustrowaniem analizy wrażliwości jest wykres leśny wielkości efektu i wykres zmian precyzji, przy wyłączeniu poszczególnych badań.

Przykład (27.2) 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, jednakże nie uwzględniającego poszczególnych badań, wciąż pozostają istotne statystycznie. Zastrzeżenie budzi jedynie badanie 3, po jego wyłączeniu precyzja uzyskanego podsumowania wzrasta, przedział ufności dla efektu końcowego jest wówczas węższy o około 18%.

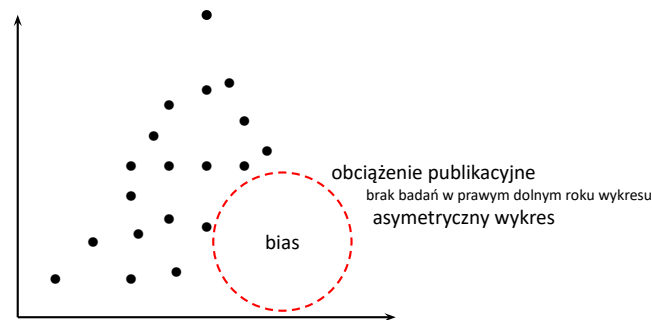


Do tych samych wniosków prowadzi analiza wykresów. Największy przedział i najkorzystniejszą zmianę precyzji uzyskamy po wyłączeniu badania 3.

27.6 Badanie asymetrii

Symetria w uzyskanych efektach 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 przeprowadzanych meta-analiz publikowana jest mimo występującej zdiagnozowanej asymetrii. Badania takie wymagają jednakże dobrego uzasadnienia takiego postępowania.

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, zlokalizowane są blisko siebie i wyznaczają środek lejka, 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 niż przedstawienie samej wielkości badania. Błąd efektu jest bowiem miarą wskazującą na jakość badania i niosącą w sobie również informację o jego liczności.

Test Egger'a

Ponieważ interpretacja wykresu lejkowego jest zawsze subiektywna, pomocą może być wykorzystanie współczynnika Egger'a (Egger 1997[53]) - wyrazu wolnego w dopasowanej prostej regresji. Współczynnik ten bazuje na korelacji pomiędzy odwrotnością błędu standardowego i stosunkiem wielkości efektu do jego błędu. Czym bardziej oddalona od 0 wartość współczynnika, tym większa asymetria. Kierunek współczynnika określa rodzaj asymetrii: wartość dodatnia wraz z dodatnim przedziałem ufności dla niej 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 Egger'a powinien być używany tylko przy dużym zróżnicowaniu wielkości badań i występowaniu badania o średniej wielkości.

Uwaga!

Przy niewielu badaniach (małej liczbie k) trudno jest osiągnąć istotny wynik mimo widocznej asymetrii.

Hipotezy:

$$\mathcal{H}_0 : b = 0,$$

$$\mathcal{H}_1 : b \neq 0,$$

gdzie:

b - to wyraz wolny w równaniu regresji Egger'a.

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 :

jeżeli $p \leq \alpha \implies$ odrzucamy \mathcal{H}_0 przyjmując \mathcal{H}_1 ,
 jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić \mathcal{H}_0 .

Badanie liczby "Fail-safe"

Nfs Rosenthal'a - liczba "Fail-safe" opisana przez Rosenthal'a (1979)[138] określa liczbę prac nie wskazujących na występowanie efektu (np. różnica średnich równa 0, iloraz szans równy 1 itp.) jaka jest potrzebna by zredukować efekt podsumowujący z istotnego statystycznie do statystycznie nieistotnego.

$$Nfs = \frac{\left(\sum_{i=1}^k Z_i\right)^2}{Z_c^2} - k$$

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.

Rosenthal (1984)[127] określił liczbę prac będącą **punktem odcięcia** jako $5k + 10$. Wyznaczając iloraz Nfs i punktu odcięcia uzyskujemy **współczynnik(fs)**. Według interpretacji Rosenthal'a, jeśli współczynnik(fs) będzie większy niż 1, to prawdopodobieństwo obciążenia publikacyjnego jest minimalne.

Nfs Orwin'a - liczba "Fail-safe" opisana przez Orwin'a (1983) określa liczbę prac o średnim efekcie wskazanym przez badacza \overline{M}_{fs} , jaka jest potrzebna by zredukować efekt podsumowujący do pożądanej wielkości M_d wskazanej przez badacza.

$$Nfs = k \frac{M - M_d}{M_d - \overline{M}_{fs}}$$

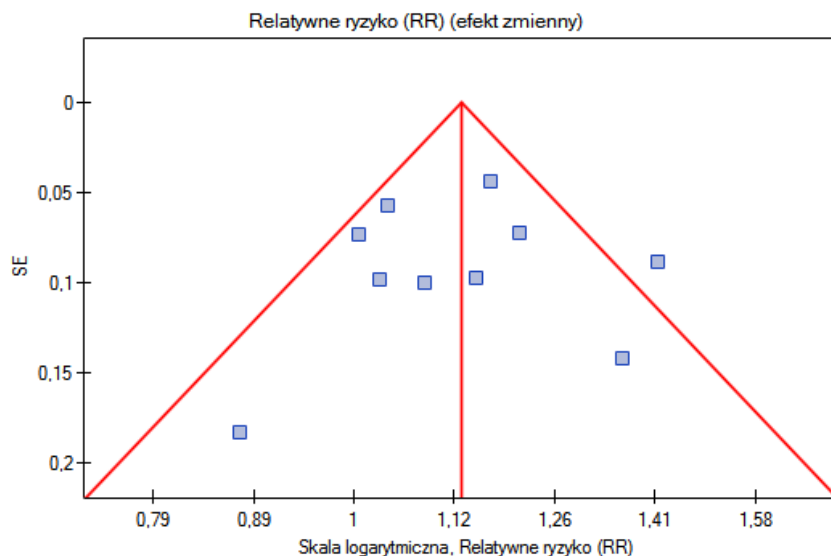
gdzie:

M - efekt podsumowujący uzyskany w meta-analizie.

Przykład (27.2) 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	
Pożądaný efekt	1,12
Średni efekt	1,11
N(fs)	11

Wynik testu Eggera nie są istotne statystycznie ($p=0.9137$), co wskazuje na brak obciążenia publikacyjnego.



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. Na podstawie wykresu również nie mamy zasadniczych zastrzeżeń co do doboru badań, jedyna wątpliwość dotyczy badania trzeciego.

Liczba publikacji "Fail-safe" wyznaczona metodą Rosenthal'a jest duża i wynosi 72. Zatem, jeśli efekt podsumowujący (wspólne dla wszystkich badań relatywne ryzyko) miałby być nieistotny statystycznie (palenie papierosów miałyby 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 Orwin'a 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. Natomiast zmniejszając wielkość relatywnego ryzyka z 1.11 do 1.10 już tylko 5 prac potrzeba by wspólne relatywne ryzyko wyniosło 1.12.

27.7 Meta-analiza skumulowana

Typowym celem przeprowadzenia meta-analizy skumulowanej jest pokazanie jak zmienił się efekt od czasu przeprowadzenia/opublikowania ostatniej meta-analizy dotyczącej danego tematu, lub jak zmienił się na przestrzeni ostatnich lat. Wówczas chronologicznie (zgodnie z linią czasu) dodawane są kolejne badania i każdorazowo wyliczany efekt podsumowujący. Równie ważna jest skumulowana analiza w badaniu zmiany 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**.

W zależności od celu, któremu ma służyć kumulacja należy wybrać zmienną wg której poszczególne badania będą sortowane, czyli ustalić kolejność dodawania badań do podsumowania meta-analizą. Może to być dowolna zmienna liczbowa.

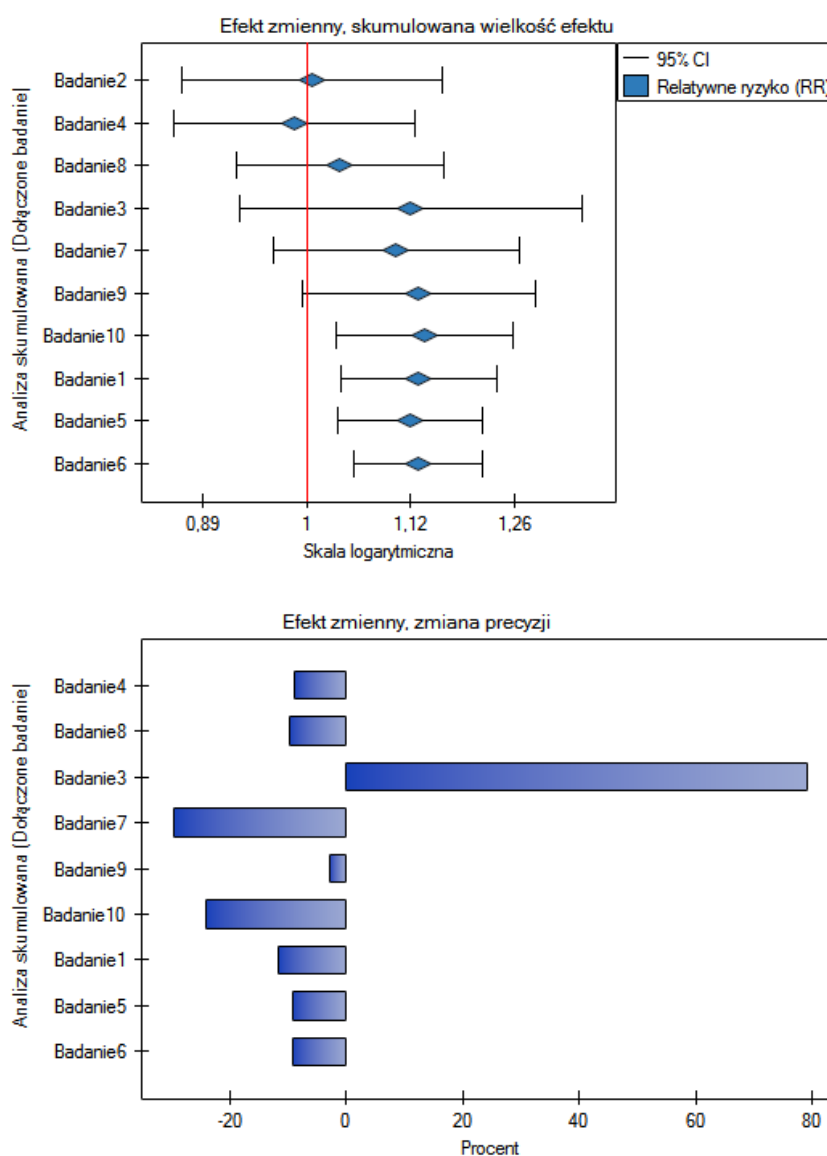
Przypisana do danego badania wartość **udziału skumulowanego** definiuje odsetek jaki stanowi łączna waga włączonych do analizy badań tzn. danego badania i badań go poprzedzających. **Zmiana precyzji** informuje natomiast jak zmieni się precyzja efektu podsumowującego (szerokość przedziału ufności), gdy dane badanie zostanie dołączone do badań go poprzedzających.

Dobrym zilustrowaniem analizy skumulowanej jest wykres leśny wielkości efektu i wykres zmian precyzji, przy dołączaniu poszczególnych badań.

Przykład (27.2) c.d. (plik *MetaanalizaRR.pqs*) Badając wpływ palenia papierosów na wystąpienie choroby X sprawdzono jak uzyskane wyniki ewaluowały w czasie. W tym celu, w oknie analizy wybrano opcję Meta-analiza skumulowana i zmienną według której kolejne prace będą włączane do meta-analizy, oraz wskazano wykres leśny (skumulowany) i wykres słupkowy (skumulowany).

Skumulowana										
Nazwa	Liczność	Relatywny	SE (ln)	-95%CI	+95%CI	Statystyk	Wartość p	Udział sku	Zmiana pre	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 dodając do wcześniejszych prac Badanie 10, a następnie również kolejne badania. Zasadniczo dodanie kolejnych prac zwiększało precyzję uzyskanego relatywnego ryzyka, za wyjątkiem sytuacji, w której dodano Badanie 3. Przedział ufności podsumowującego relatywnego ryzyka poszerzył się wówczas o 79.15%. Efekt ten widzimy w tabeli oraz na dołączonych wykresach. W rezultacie należy się zastanowić nad wykluczeniem Badania 3 z przeprowadzonej meta-analizy.



27.8 Porównanie grup

Bywają sytuacje, w których zebrane dane dotyczą tego samego efektu, są wykonywane na tej samej populacji, ale w nieco innych warunkach. Załóżmy że część badań była przeprowadzona w warunkach A, a część w warunkach B. Wówczas interesujące może być porównanie uzyskanych dla każdej grupy efektów podsumowujących. Wykazanie różnic pomiędzy efektami podsumowującymi może być głównym celem meta-analizy i wówczas niewskazane jest podsumowanie jednym wspólnym efektem obu podgrup jednocześnie. Jeśli jednak badacz zdaje sobie sprawę że badania przeprowadzane były w różnych warunkach, ale celowe wydaje się wspólne podsumowanie wszystkich badań, wówczas wykazując brak istotnych statystycznie (lub klinicznie) różnic, może dokonać wspólnego podsumowania uwzględniając jednocześnie ów podział na podgrupy A i B, czyli wyznaczyć wspólne podsumowanie w korekcji o różne warunki eksperymentu. Np. W kraju A panuje nieco inny klimat niż w kraju B. Dysponujemy szeregiem badań dotyczących kraju A i szeregiem badań dotyczących kraju B. Jeśli naszą populacją badaną jest roślinność tych dwóch krajów, to możemy sprawdzić, czy warunki klimatyczne mają wpływ na uzyskane efekty badań dla poszczególnych krajów. Analiza porównawcza wyznaczonych w ten sposób

podgrup pozwoli ocenić czy klimat ma zasadniczy wpływ na uzyskiwane wyniki, czy też nie i czy można rzeczywiście wyniki badań obejmujące te dwa kraje podsumować jednym wspólnym efektem, czy też powinniśmy wyznaczać oddzielne podsumowania dla każdego kraju. Innym przykładem może być sytuacja, gdy część badań to badania, w których przeprowadzona była randomizacja ale w części nie mamy pełnej randomizacji, wówczas możemy podzielić badania na podgrupy, by następnie sprawdzić czy badania bez randomizacji dają na tyle zbliżone efekty do badań z randomizacją by móc je włączyć do dalszej, wspólnej analizy.

Badanie heterogeniczności grup

Porównania grup możemy 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.

- **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 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 (tzn. populacyjnym) efektem podsumowującym. W praktyce jest to sprawdzenie czy wariancja efektów podsumowujących grupy jest zerowa, czyli badanie heterogeniczności grup. Opis i interpretacja wyników analizy heterogeniczności znajduje się w rozdziale [Badanie heterogeniczności](#), z tym że w przypadku porównania grup heterogeniczność dotyczy efektów podsumowujących porównywanych grup, nie poszczególnych badań, a wynik uzależniony jest od wybranego efektu podsumowującego.

Hipotezy:

$$\begin{aligned}\mathcal{H}_0 &: \tau^2 = 0, \\ \mathcal{H}_1 &: \tau^2 \neq 0,\end{aligned}$$

gdzie:

τ^2 - to wariancja prawdziwych (populacyjnych) efektów podsumowujących porównywanych grup.

Wyznaczoną na podstawie [statystyki testowej](#) wartość p porównujemy z poziomem istotności α :

$$\begin{aligned}\text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0.\end{aligned}$$

Jeśli uzyskamy wynik jest istotny statystycznie (wynik **statystyki Q**, **współczynnika I^2** lub **współczynnika H^2**), jest to mocna sugestia, by zrezygnować z wspólnego podsumowywania porównywanych grup.

Badanie heterogeniczności w grupach

Dodatkową opcją analizy jest możliwość przeanalizowania każdej grupy oddzielnie pod kątem heterogeniczności, zgodnie z opisem w rozdziale [Badanie heterogeniczności](#). Uzyskane wyniki (w szczególności wyznaczona wariancja T^2) ułatwiają podjęcie decyzji co do sposobu porównania grup tzn. wyboru efektu zmiennego (oddzielne T^2) lub efektu zmiennego (wspólne T^2).

Wspólne podsumowanie grup

W sytuacji, w której na podstawie wyników porównania grup, różnice uzyskane pomiędzy efektami podsumowującymi poszczególne grupy są niewielkie i nieistotne, można dokonać wspólnego podsumowania tych grup. Podsumowanie odbywa się w korekcji o podział na wskazane grupy. Jeśli na przykład badania dzieliliśmy na podstawie różnych warunków prowadzonego eksperymentu, wówczas wspólne podsumowanie dokonane zostanie w korekcji o różne warunki eksperymentu. Rezultat wspólnego podsumowania zależy od obserwowanych różnic (od wariacji pomiędzy badaniami i pomiędzy grupami) czyli od wyboru efektu podsumowującego (czy jest to efekt stały, czy zmienny (oddzielne T^2), czy też zmienny (wspólne T^2)).

Dobrym zilustrowaniem wspólnego podsumowania grup w meta-analizie jest wykres leśny przedstawiający wyniki poszczególnych badań wraz z podsumowaniem każdej grupy i wspólnym podsumowaniem grup.

Porównanie ANOVA

Porównanie ANOVA jest dodatkową opcją porównania grup, jest to nieco inna metoda porównania niż porównanie poprzez badanie heterogeniczności grup (oparta na innym modelu matematycznym). Obie metody dają jednakże zbieżne wyniki, co do porównania grup. W przypadku porównania grup metodą ANOVA obserwowana wariancja zostaje rozbita na wariancję między grupami i wariancję wewnątrz grup. Wariancja wewnątrz grup rozbita zostaje następnie na wariancję każdej grupy oddzielnie. W rezultacie wyznaczone zostają następujące statystyki Q :

- 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) - bada tę część wariancji łącznej, która odnosi się do grupy drugiej, czyli wariancję pomiędzy badaniami znajdującymi się wewnątrz grupy drugiej,
- ...
- Statystyka Q (grupa g) - bada tę część wariancji łącznej, która odnosi się do grupy ostatniej, czyli wariancję pomiędzy badaniami znajdującymi się wewnątrz grupy ostatniej,
- Statystyka Q (wewnątrz grup) = Q (grupa 1) + Q (grupa 2) + ... + Q (grupa g) - bada tę część wariancji łącznej, która odnosi się do wnętrza poszczególnych grup, czyli wariancję badań wewnątrz poszczególnych grup,
- **Statystyka Q (między grupami)** - bada tę część wariancji łącznej, która odnosi się do różnic między grupami, czyli wariancję pomiędzy grupami (**wynik tożsamy z badaniem heterogeniczności grup**),
- Statystyka Q (łączna) - bada wariancję pomiędzy wszystkimi badaniami.

Każda z powyższych statystyk Q ma [rozkład \$\chi^2\$](#) z odpowiednią dla niej liczbą stopni swobody.

Okno z ustawieniami opcji porównania grup dla meta-analizy wywołujemy poprzez menu: Statystyki zaawansowane → Meta-analiza → Porównanie grup.

Meta-analiza porównanie grup

Analiza statystyczna : Meta-analiza porównanie grup

Efekt końcowy: Relatywne ryzyko (RR)

Nazwa badania	Liczność
1-Nazwa badania	1-Nazwa badania
2-Liczność	2-Liczność
3-RR	3-RR
4-SE(lnRR)	4-SE(lnRR)
5-lata palenia	5-lata palenia
6-warunki eksperymentu	6-warunki eksperymentu

Relatywne ryzyko (RR)	Błąd standardowy (SE (ln))
1-Nazwa badania	1-Nazwa badania
2-Liczność	2-Liczność
3-RR	3-RR
4-SE(lnRR)	4-SE(lnRR)
5-lata palenia	5-lata palenia
6-warunki eksperymentu	6-warunki eksperymentu

0,05 poziom istotności

Zm. grupująca: 1-Nazwa badania

Opcje testu

Efekt stały

Efekt zmienny (oddzielne T2)

Efekt zmienny (wspólne T2)

Badanie heterogeniczności w grupach

Wspólne podsumowanie grup

Porównanie ANOVA

Filtr danych

Wybór danych spełniających odpowiednie warunki

Reguły połączone są koniunkcją

podstawowy wielokrotny **AND**

Opcje raportu

Dołącz dane Połącz w jeden raport

Dołącz wykres

PRZYKŁAD 27.3. (plik MetaanalizaRR.pqs)

Badano ryzyko choroby X dla osób palących i dla niepalących. By ustalić czy czas pozostawania w nałogu ma wpływa na występowanie choroby X zaplanowano przeprowadzenie meta-analzy. Wykonano dokładny przegląd literatury dotyczącej tego tematu i na tej podstawie wytypowano 17 prac naukowych dysponujących opisem relatywnego ryzyka i jego błędu (tzn. 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-analzy i zapisano w pliku.

Celem przeprowadzania meta-analzy 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-analzy 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.

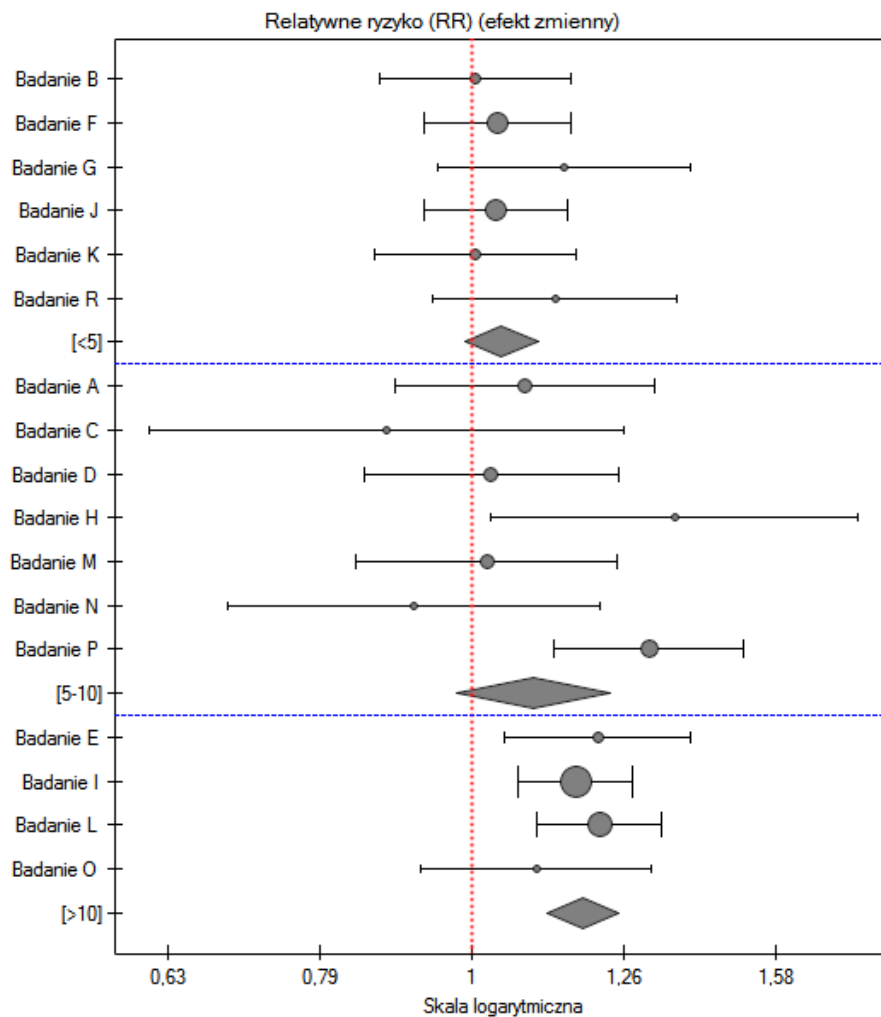
Meta-analiza porównanie grup	
Czas analizy	0,07 sek.
Analizowane zmienne	Nazwa badania Liczność RR SE(lnRR) lata palenia
Poziom istotności	0,05
Zmienna grupująca	lata palenia
Efekt	Zmienny oddzielne T2
Efekt końcowy	Relatywne ryzyko (RR)
Liczba grup	3
Badanie heterogeniczności	
Porównanie grup	
Statystyka 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%

Grupy różnią się istotnie statystycznie ($p=0.0092$), co obserwujemy nie tylko na bazie testu heterogeniczności, ale również współczynnika H2 (współczynnik wraz z przedziałem ufności znajduje się powyżej wartości jeden) oraz I2 (78%, to wysoka heterogeniczność). Dlatego zebrane prace nie zostaną podsumowane wspólnym efektem a jedynie poprzez oddzielne podsumowanie każdej grupy.

Dane :										
Nazwa	Liczność	Relatywne	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.



Dodatkowo sprawdzona została 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

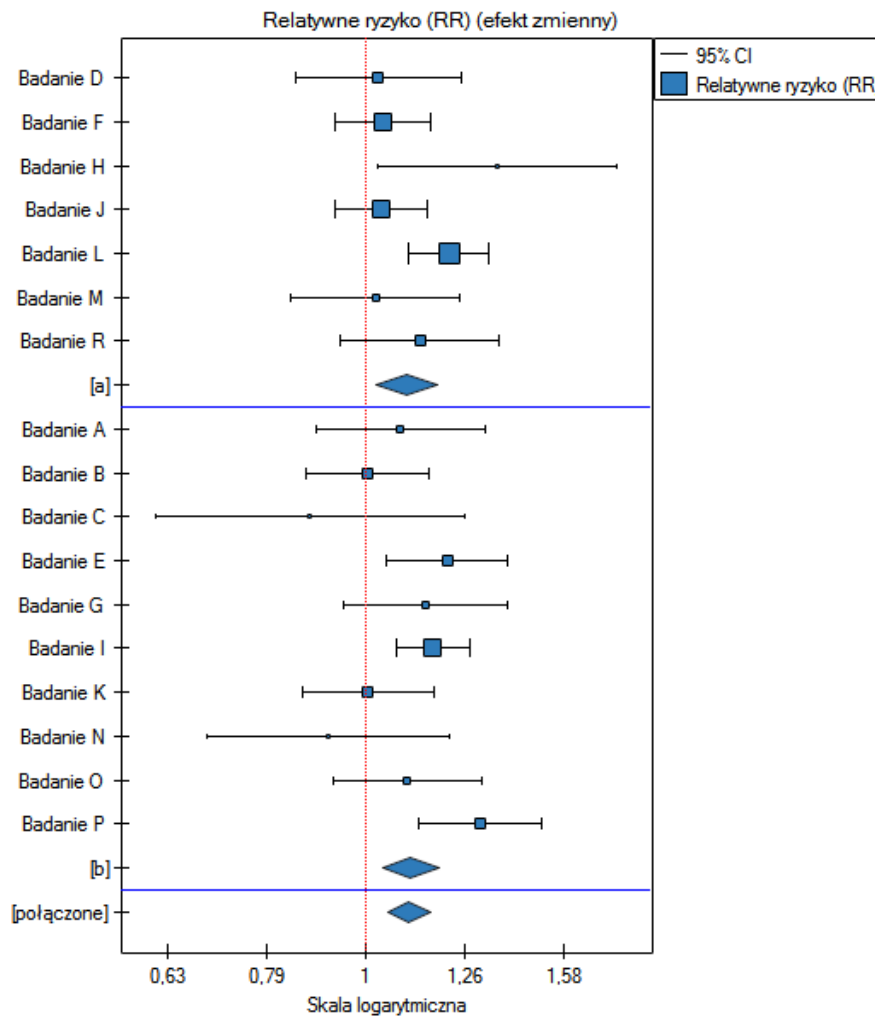
Badanie heterogeniczności			
<5	Wartości	-95%CI	+95%CI
T2	0	0	0,004108
H2	0,454570	0,115348	1,791383
I2	0%	0%	44,18%
5-10	Wartości	-95%CI	+95%CI
T2	0,011432	0	0,043367
H2	1,954872	0,826799	4,622070
I2	48,85%	0%	78,36%
>10	Wartości	-95%CI	+95%CI
T2	0	0	0,004504
H2	0,351844	0,053874	2,297807
I2	0%	0%	56,48%

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 w korekcji 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

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	Relatywna	SE (ln)	-95%CI	+95%CI	Statystyka	Wartość p	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			



27.9 Meta-regresja

Analiza meta-regresji jest przeprowadzana w analogiczny sposób 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, \dots, X_k) czyli wyjaśniające. Podobnie jak w tradycyjnych modelach regresji, zmienne niezależne mogą wchodzić w interakcje a te, które są opisane skalą nominalną mogą podlegać specjalnemu kodowaniu (więcej informacji na ten temat można znaleźć w rozdziale [Przygotowanie zmiennych do analizy w modelach wielowymiarowych](#)). Liczba zmiennych niezależnych powinna być niewielka, mniejsza niż liczba prac na podstawie których przeprowadza się badanie ($n \geq k + 1$).

Meta-regresję możemy 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 wielkość tego efektu są takie same, za wyjątkiem czynników badanych jako zmienne niezależne w modelu (X_1, X_2, \dots, X_k). Jest to sytuacja występująca bardzo rzadko w prawdziwych badaniach, ponieważ wymaga w pełni kontrolowanych warunków, co w przypadku różnych badań, prowadzonych przez różne ośrodki i różnych badaczy jest prawie niemożliwe. Wykorzystanie efektu stałego miało by uzasadnienie na przykład w sytuacji, gdy wszystkie badania przeprowadza jeden ośrodek, na tej samej

populacji, przy zmianie tylko tych warunków które opisuje badana cecha. Na przykład, jeśli chcielibyśmy sprawdzić jak wpływa zmiana temperatury na zmianę opisaną w poszczególnych badaniach wielkości relatywnego ryzyka występowania choroby, wówczas wszystkie badania powinny być przeprowadzone na tej samej populacji w dokładnie tych samych warunkach za wyjątkiem zmiany temperatury, która stanowi zmienną niezależną X w modelu.

- **Efekt zmienny** wybieramy wtedy, gdy zakładamy, że badania mogą przedstawiać nieco różniące się populacje tzn. czynniki, które mogłyby zaburzać wielkość badanego efektu nie we wszystkich pracach są opisane (można założyć, że są podobne, ale nie muszą być dokładnie takie same). Każda praca podaje wielkości czynników, którymi jesteśmy zainteresowani, które biorą udział w budowaniu modelu jako zmienne niezależne (X_1, X_2, \dots, X_k). Wykorzystanie efektu zmiennego jest częste, ponieważ poszczególne badania przeprowadzane są najczęściej przez różne ośrodki w nieco innych warunkach, interesująca zmienność dotyczy tylko tych warunków które opisują podane w badaniu czynniki np. temperatura, która stanowić będzie zmienną niezależną X w modelu.

Weryfikacja modelu

- **Istotność statystyczna poszczególnych zmiennych w modelu.**

Na podstawie współczynnika oraz jego błędu możemy wnioskować czy zmienna niezależna, dla której ten współczynnik został oszacowany wywiera istotny wpływ na efekt końcowy. W tym celu testujemy hipotezy:

$$\begin{aligned} \mathcal{H}_0 &: \beta_i = 0, \\ \mathcal{H}_1 &: \beta_i \neq 0. \end{aligned}$$

Wyliczmy statystykę testową według wzoru:

$$Z = \frac{b_i}{SE_{b_i}}$$

Statystyka testowa ma [rozkład normalny](#).

Wyznaczoną na podstawie [statystyki testowej](#) [wartość p](#) porównujemy z poziomem istotności α :

$$\begin{aligned} \text{jeżeli } p \leq \alpha &\implies \text{ odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha &\implies \text{ nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

- **Jakość zbudowanego modelu** liniowej regresji wielorakiej możemy ocenić kilkoma miarami.
 - **Współczynnik R^2** - jest miarą dopasowania modelu. Wyraża on procent zmienności pomiędzy efektami badań tłumaczony przez model. Wartość tego współczynnika mieści się w przedziale $< 0; 1 >$, gdzie 1 oznacza doskonałe dopasowanie modelu, 0 - zupełny brak dopasowania. W jego wyznaczeniu posługujemy się następującą równością:

$$R^2 = T_{(modelu)}^2 + T_{(total)}^2,$$
 gdzie:
 - $T_{(modelu)}^2$ - wariancja między badaniami wyjaśniona przez model,
 - $T_{(total)}^2$ - całkowita wariancja między badaniami.
 - **Współczynnik I^2** - określa procent obserwowanej wariancji, jaki wynika z rzeczywistej różnicy w wielkości badanych efektów.

Uwaga!

Dokładne przedstawienie opisywanej przez współczynniki wariancji można znaleźć w dziale [Badanie heterogeniczności](#)

- **Istotność statystyczna wszystkich zmiennych w modelu**

Podstawowym narzędziem szacującym istotność wszystkich zmiennych w modelu jest ANOVA wyznaczająca Q (modelu).

$$\mathcal{H}_0 : \text{wszystkie } \beta_i = 0,$$

$$\mathcal{H}_1 : \text{istnieje } \beta_i \neq 0.$$

Wykorzystując podejście ANOVA, obserwowaną wariancję pomiędzy badaniami rozbija się na wariancję tłumaczoną przez model i wariancję reszt (nie tłumaczoną 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.

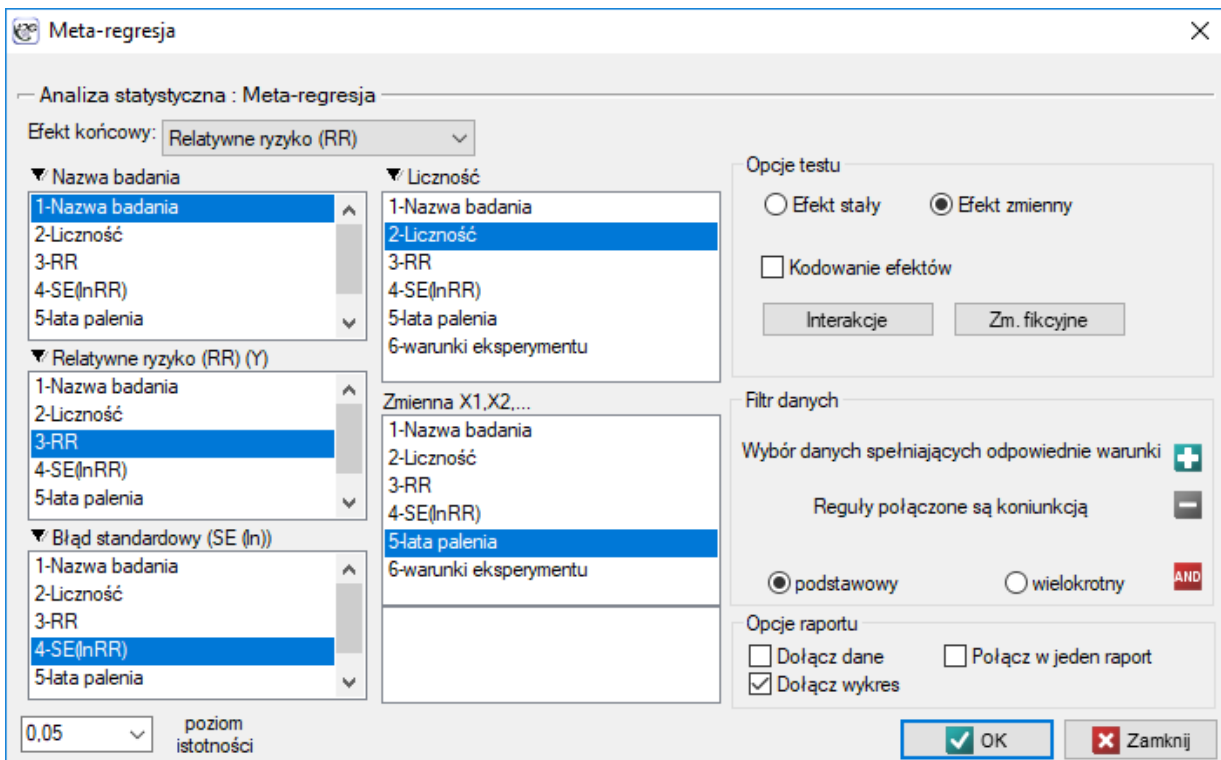
Każda z powyższych statystyk Q ma rozkład χ^2 z odpowiednią dla niej liczbą stopni swobody.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

$$\text{jeżeli } p \leq \alpha \implies \text{odrzucaamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1,$$

$$\text{jeżeli } p > \alpha \implies \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0.$$

Okno z ustawieniami opcji porównania grup dla meta-analizy wywołujemy poprzez menu: Statystyki zaawansowane → Meta-analiza → Porównanie grup.



Przykład (27.3) c.d. (plik *MetaanalizaRR.pqs*) Badano ryzyko choroby X dla osób palących i dla niepalących. By ustalić czy czas pozostawania w nałogu ma wpływa 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 (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 papierosów zbudowano dwa modele regresji. W pierwszym modelu zmienną grupującą Lata palenia potraktowano jak zmienną ciągłą. W modelu drugim ustalono, że zmienna Lata palenia traktowana będzie jako zmienna kategoryjalna (fikcyjna) 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	Statystyka	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	Statystyka	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

W obu modelach potwierdzono istotny związek pomiędzy czasem palenia a wielkością relatywnego ryzyka wystąpienia choroby X. W modelu pierwszym ustalono, że wraz ze upływem czasu palenia (przejdźciem do kolejnej grupy lat palenia) logarytm relatywnego ryzyka choroby X zwiększy się o 0.06139. Do podobnych wniosków prowadzi analiza wyników modelu drugiego. W tym przypadku wyniki rozpatrujemy w odniesieniu do grupy palących krócej niż 5 lat. Logarytm relatywnego ryzyka dla palących od 5 do 10 lat wzrasta o 0.06663 (w stosunku do palących krócej niż 5 lat), a dla palących dłużej niż dziesięć lat wzrasta aż o 0.12182 (w stosunku do palących krócej niż 5 lat).

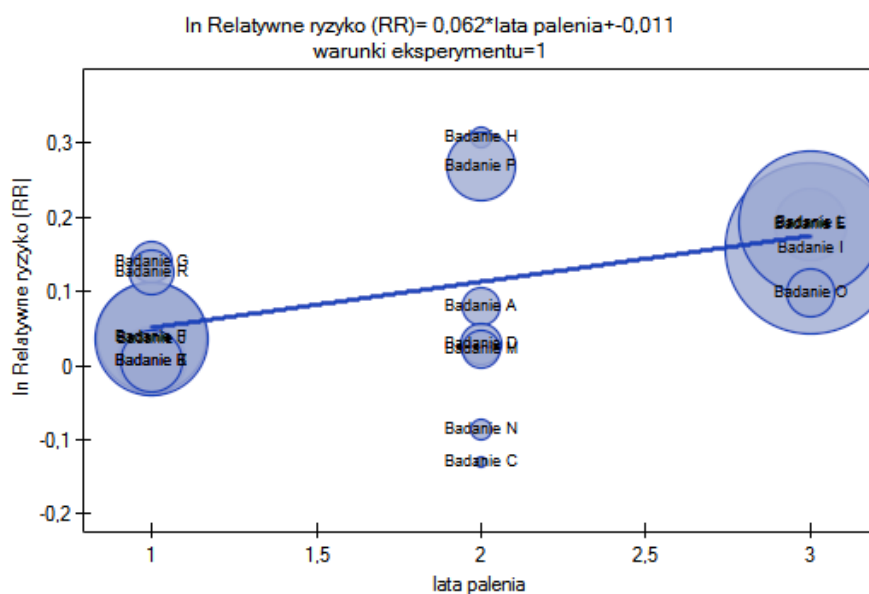
Ponieważ część badań prowadzona była według innych kryteriów (w innych warunkach) uzyskane wyniki obu modeli skorygowano różne warunki prowadzenia badań.

Model						
	wsp. b	błąd b	-95% CI	+95% CI	Statystyka	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

Model						
	wsp. b	błąd b	-95% CI	+95% CI	Statystyka	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 to jaką metodologię (kryteria włączenia/wyłączenia osób) stosowano by przeprowadzić badania. Uzyskaną zależność dla modelu pierwsze-

go, przy założeniu prowadzenia badań w warunkach "a" (wskazanych jako warunki pierwsze) przedstawia wykres.



28 ANALIZA RZETELNOŚCI

Analiza rzetelności (*ang. reliability analysis*) związana jest najczęściej z budową skal złożonych, w szczególności skal sumarycznych, czyli składających się z wielu pojedynczych pozycji (pytań). Rzetelność skali – rozumiana jako jej wewnętrzna spójność, daje nam informację o tym, na ile dana skala mierzy to, co mierzy; czyli w jakim stopniu poszczególne pozycje skali mierzą to, co cała skala.

Aby można było mówić o rzetelności skali, poszczególne pozycje skali powinny mierzyć ten sam konstrukt – korelacja pomiędzy poszczególnymi pozycjami skali powinna być wysoka. Założenie to możemy sprawdzić wyliczając macierz [współczynników korelacji Pearsona](#). 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[49]. Mierzy on stosunek wariancji poszczególnych pozycji do wariancji całej skali (sumy tych pozycji) i jest wyliczany ze wzoru:

$$\alpha_C = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k sd_i^2}{sd_t^2} \right),$$

gdzie:

k – liczba pozycji w skali,

sd_i^2 – wariancja pozycji i ,

sd_t^2 – wariancja sumy wszystkich pozycji.

Standaryzowany współczynnik rzetelności $\alpha_{standard}$ jest wyliczany jako:

$$\alpha_{standard} = \frac{k\bar{r}_p}{1 + (k-1)\bar{r}_p},$$

gdzie:

\bar{r}_p – średnia współczynników korelacji Pearsona wszystkich $(k(k-1)/2)$ pozycji skali.

Współczynnik α może przyjmować dowolne wartości mniejsze lub równe 1, w tym również wartości ujemne, jednakże tylko wartości dodatnie mają sens. Jeżeli wszystkie pozycje skali są idealnie rzetelne, wówczas współczynnik rzetelności wynosi 1.

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.

Rzetelność połówkowa (*ang. split-half reliability*)

Polega na losowym podzieleniu pozycji skali na dwie połowy i analizie korelacji między tymi połówkami. Odbywa się to przy użyciu współczynnika rzetelności połówkowej Spearmana-Browna opublikowanego niezależnie przez Spearmana (1910)[155] i Browna (1910)[33]:

$$r_{SH} = \frac{2r_p^*}{1 + r_p^*},$$

gdzie:

r_p^* – współczynnik korelacji Pearsona pomiędzy połówkami skali.

Jeżeli dwie losowo wybrane połówki są idealnie skorelowane, wówczas $r_{SH} = 1$.

Inny wzór na współczynnik rzetelności połówkowej zaproponował Guttman (1945)[72]:

$$r_{SHG} = 2 \left(1 - \frac{sd_{t1}^2 + sd_{t2}^2}{sd_t^2} \right),$$

gdzie:

sd_{t1}^2, sd_{t2}^2 – wariancja pierwszej i drugiej połowy skali,

sd_t^2 – wariancja sumy wszystkich pozycji.

Uwaga!

Przyjmuje się, że aby można było uznać skalę za rzetelną, współczynniki rzetelności skali ($\alpha_C, \alpha_{standard}, r_{SH}, r_{SHG}$) powinny być większe niż 0.6 i mniejsze niż 1.

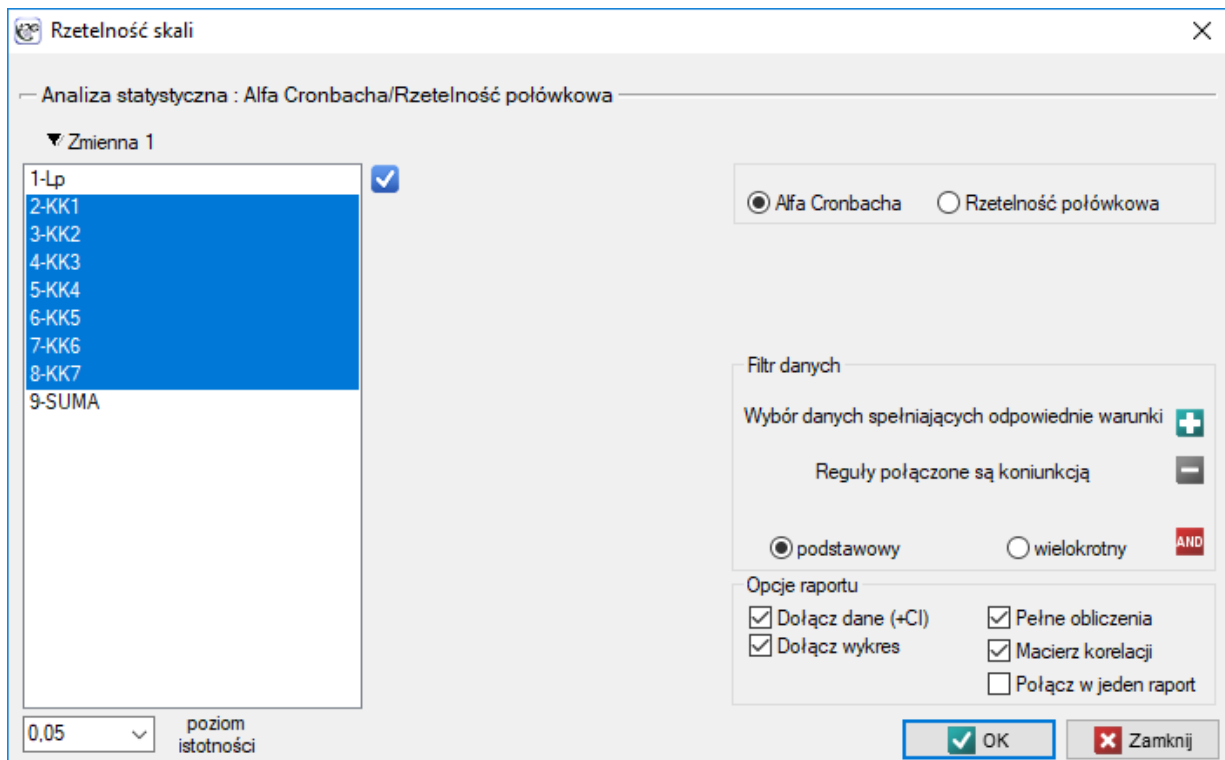
Błąd standardowy pomiaru (*ang. standard error of measurement*) wyznacza się dla skali rzetelnej zgodnie z wzorem:

$$SEM = sd_t \sqrt{1 - \alpha_C} \quad \text{– dla współczynnika rzetelności Cronbacha}$$

lub

$$SEM = sd_t \sqrt{1 - r_{SH}} \quad \text{– dla współczynnika rzetelności połówkowej}$$

Okno z ustawieniami opcji rzetelności skali wywołujemy poprzez menu Statystyka → Rzetelność skali.



PRZYKŁAD 28.1. (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 uzyskać 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

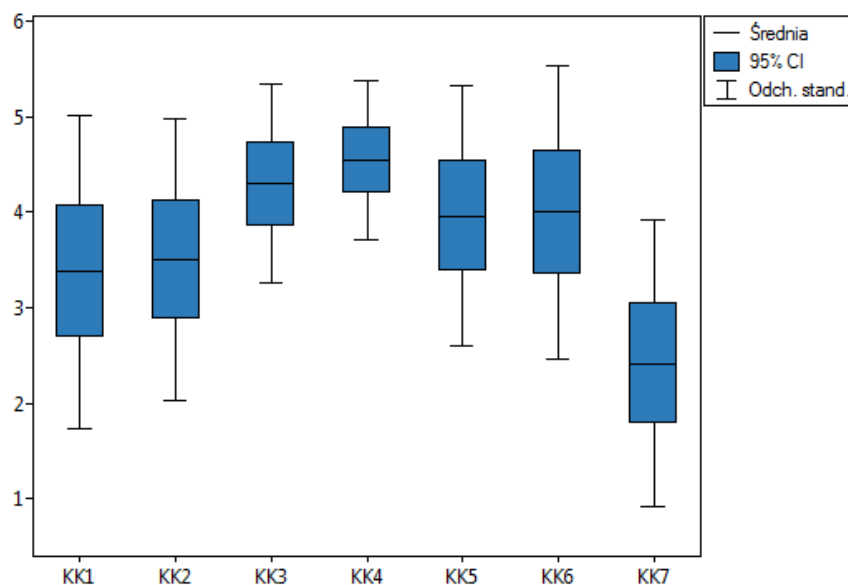
Chcemy sprawdzić dokładność zastosowanego narzędzia oceny kompetencji (zastosowanej skali). W tym celu przeprowadzona zostanie analiza rzetelności.

Macierz korelacji wskazuje, że ostatnia pozycja jest najslabiej skorelowana z pozostałymi. Podejrzewamy zatem, że pozycja ta nie mierzy tego samego konstruktów, 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

Skala kompetencji okazała się być skalą rzetelną. Współczynnik Alfa Cronbacha wynosi 0.736805, a średni współczynnik korelacji pomiędzy pozycjami skali 0.31847.

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



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 najniższa 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).

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

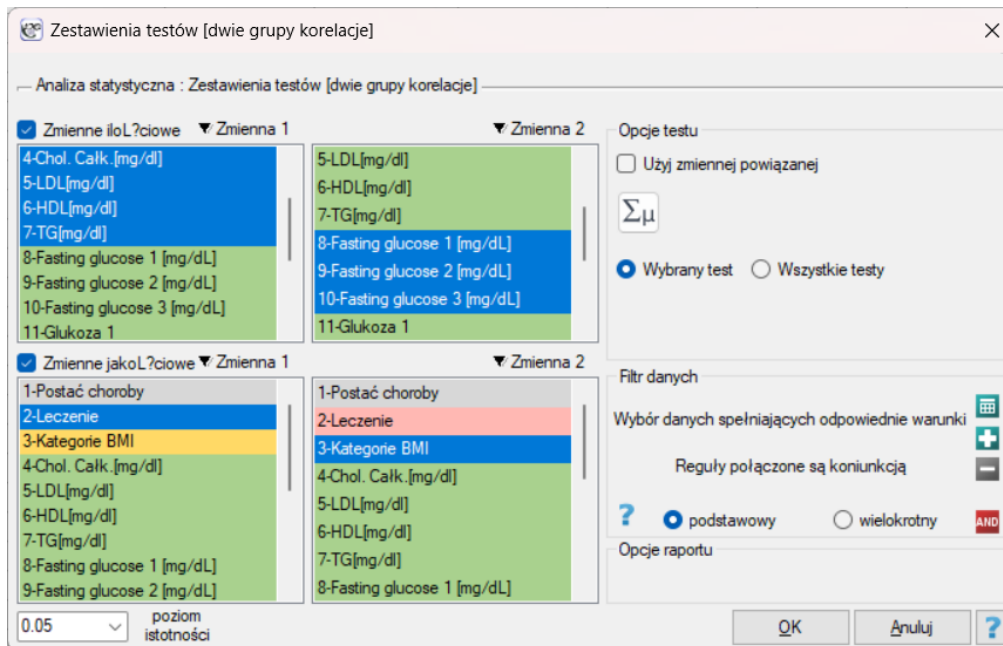
Współczynnik rzetelności połówkowej Spearmana-Browna wynosi 0.857705, a Guttmana 0.856531. Połówki są dość dobrze skorelowane – współczynnik korelacji połówek wynosi 0.750862. Jednak wartość współczynnika Alfa Cronbacha jest zbyt niska dla drugiej połówki (0.416958). Połówka ta zawiera pozycje KK7, która wykazuje słabą korelację z pozostałymi pozycjami skali. Po usunięciu tej pozycji i ponownym wykonaniu analizy, wszystkie współczynniki rzetelności rzeczywiście są wyższe.

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

29 ZESTAWIENIA TESTÓW

By przyspieszyć pracę poszczególne testy możemy wykonywać w zestawach. Dane ilościowe będą mogły być dodatkowo opisane przy pomocy średnich, median, itp., a dane jakościowe przy pomocy licznosci i procentów.

Okno z ustawieniami opcji Zestwień testów wywołujemy poprzez menu Statystyka → Zestawy → Zestawienia testów a następnie wybraną grupę analiz.



Do dyspozycji mamy:

1. Porównanie dwóch grup zależnych:
 - Test t-Studenta dla grup zależnych
 - Test Wilcozona dla grup zależnych
 - Test Bowkera-McNemara
 - Test normalności rozkładu Kołmogorow-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 Kołmogorow-Smirnov (lub inny zaproponowany przez użytkownika)
 - i inne ...
3. Korelację:
 - Zależność liniowa Pearsona

- Zależność monotoniczna Spearmana
- Chi-kwadrat, Fisher, wsp. korelacji
- Test normalności rozkładu Kołmogorow-Smirnov (lub inny zaproponowany przez użytkownika)
- i inne ...

W programie dla każdej z analizowanych zmiennych, w zależności od tego czy są to zmienne ilościowe czy jakościowe, możemy zwrócić wyniki:

- **wybranych testów** - dobranych w sposób automatyczny wg reguły opisanej poniżej zwróconej w raporcie tabeli;
- **wszystkich testów** i towarzyszących im współczynników oraz miar bez względu na to, czy są spełnione minimalne warunki ich stosowania.

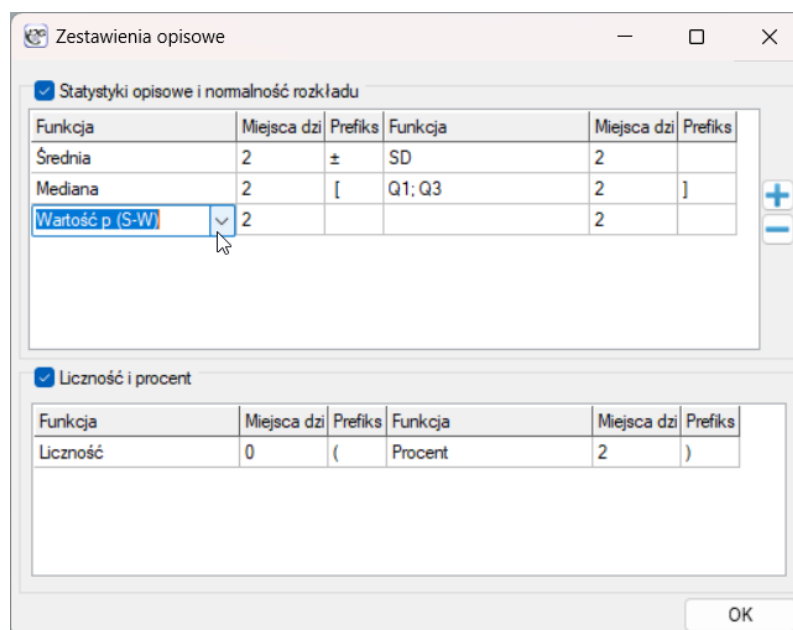
Uwagi dotyczą automatycznego wyboru testu przez program

Uwaga 1!

Jeśli przed analizą użytkownik nie opíše poszczególnych zmiennych odpowiednimi **skalami**, to dane ilościowe będą traktowane jak skala interwałowa, a dane jakościowe jak skala nominalna.

Uwaga 2!

Badanie normalności rozkładu opiera się na wynikach testu normalności wybranych przez użytkownika podczas ustawiania statystyk opisowych.



Uwaga 3!

Jeśli użytkownik zdecyduje nie wskazywać testów oceniających normalność rozkładu w oknie tych statystyk, wówczas zostanie ona sprawdzona w oparciu o test Kolmogorowa-Smirnowa. Analizy które proponujemy są odporne na niewielkie odstępstwa od rozkładu normalnego, a test Kolmogorowa-Smirnowa jest najbardziej konserwatywny spośród dostępnych testów, przez co wykazujemy brak normalności rozkładu tylko przy dużej różnicy badanego rozkładu od rozkładu normalnego. W takiej sytuacji normalność rozkładu badamy (1) dla porównania dwóch grup niezależnych w oparciu dane w każdej grupie, (2) dla porównania dwóch grup zależnych w oparciu o różnicę pomiarów, (3) dla korelacji w oparciu o reszty modelu.

PRZYKŁAD 29.1. (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.

Note

Zestawienia testów [dwie grupy niezależne]					
Zmienna	Miary	Postać choroby[ostra]	Postać choroby[przewlekła]	Wartość p	test
Chol. Całk.[mg/dl]	Średnia ± SD	160.12±23.77	160.82±30.31	0.5433	(M-W)
	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]	Średnia ± SD	86.83±20	82.91±27.81	0.2241	(M-W)
	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]	Średnia ± SD	53.27±9.7	58.99±9.97	0.0339	(t-st)
	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]	Średnia ± SD	103.25±40.09	93.36±35.59	0.3557	(M-W)
	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	Lek A	19(79.17%)	21(61.76%)	0.1583	(chi2)
	lek B	5(20.83%)	13(38.24%)		
Kategorie BMI	norma	6(25%)	32(94.12%)	<0.0001	(chi2 trendu)
	nadwaga	8(33.33%)	0(0%)		
	otyłość	10(41.67%)	2(5.88%)		
Estradiol 1	norma	7(29.17%)	22(64.71%)	0.0077	(chi2)
	powyżej normy	17(70.83%)	12(35.29%)		
Estradiol 2	norma	22(91.67%)	31(91.18%)	0.6822	(chi2 Yates)
	powyżej normy	2(8.33%)	3(8.82%)		
Estradiol 3	norma	21(87.5%)	31(91.18%)	0.988	(chi2 Yates)
	powyżej normy	3(12.5%)	3(8.82%)		

W efekcie otrzymujemy zarówno opis każdej grupy oraz dobrany do porównania test statystyczny.

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 Cochran-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 Cochran 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.”

30 KREATOR

Kreator jest narzędziem ułatwiającym początkującemu użytkownikowi poruszanie się po podstawowych statystykach zawartych w programie. Zawiera podpowiedzi 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 z zadaną przez siebie wartością wartości pomiarów dla 1 populacji. Populację tę reprezentują dane zebrane w 1 kolumnie w postaci surowej lub skumulowane do tabeli licznosci.

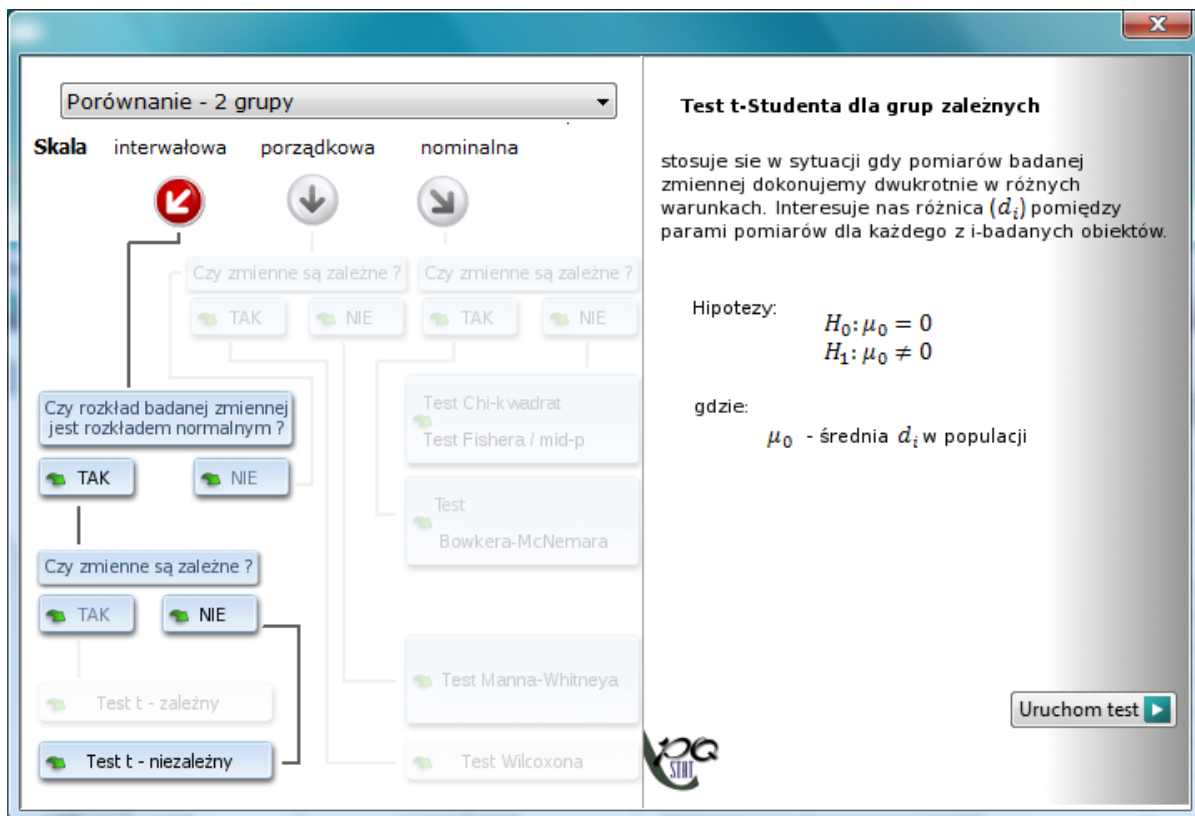
Porównanie - 2 grupy - by porównać między sobą wartości pomiarów dla 2 populacji. Populacje te reprezentowane są przez dane zebrane w postaci surowej w 2 kolumny lub skumulowane do tabeli kontyngencji.

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 kolumny lub skumulowane do tabeli kontyngencji.

Zgodność - by zbadać zgodność uzyskanych kolejnych kilku 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**).



Użytkownik przemieszcza się po grafie wybierając prawidłową odpowiedź na postawione pytania. Po przejściu wybranej przez siebie ścieżki grafu ma możliwość wykonania tego testu, który zgodnie z udzielonymi odpowiedziami jest odpowiednim testem do rozwiązania stojącego przed nim problemu statystycznego.

31 INNE UWAGI

31.1 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 to reprezentacji wszystkich obiektów w programie PQStat; zalecane zastosowanie na systemach z małą ilością pamięci operacyjnej.

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 b_2 and $\sqrt{b_1}$* . *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

- [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] Bland J.M., Altman D.G. (1986), *Statistical methods for assessing agreement between two methods of clinical measurement*. *Lancet* 327 (8476): 307–10
- [22] Bland J.M., Altman D.G. (1999), *Measuring agreement in method comparison studies*. *Statistical Methods in Medical Research* 8 (2): 135–60
- [23] Bowker A.H. (1948), *Test for symmetry in contingency tables*. *Journal of the American Statistical Association*, 43, 572-574
- [24] Box G. E. , Cox D. R. (1964), *An analysis of transformations*. *Journal of the Royal Statistical Society, Series B* 26: 211–252
- [25] 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
- [26] Breslow N.E. (1996), *Statistics in epidemiology: the case-control study*, *Journal of the American Statistical Association*, 91, 14–28
- [27] Breslow N.E. (1974), *Covariance analysis of censored survival data*. *Biometrics*, 30(1):89–99
- [28] Brookmeyer R. and Crowley J. (1982a), *A confidence interval for the median survival time*. *Biometrics* 38, 29-41
- [29] Brown L.D., Cai T.T., DasGupta A. (2001), *Interval Estimation for a Binomial Proportion*. *Statistical Science*, Vol. 16, no. 2, 101-133
- [30] Brown M.B., Forsythe A. B. (1974a), *Robust tests for equality of variances*. *Journal of the American Statistical Association*, 69,364-367
- [31] Brown M. B., Forsythe A. B. (1974), *The ANOVA and multiple comparisons for data with heterogeneous variances*. *Biometrics*, 30, 719-724
- [32] 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
- [33] Brown W. (1910), *Some experimental results in the correlation of mental abilities*. *British Journal of Psychology*, 3, 296-322
- [34] Hochberg Y. (1988), *A sharper Bonferroni procedure for multiple tests of significance*. *Biometrika* 75, 800–803
- [35] Chow S.C., Shao J., and Wang H. (2008). *Sample Size Calculations in Clinical Research, Second Edition*. Chapman and Hall/CRC. Boca Raton, Florida.
- [36] Cicchetti D. and Allison T. (1971), *A new procedure for assessing reliability of scoring eeg sleep recordings*. *American Journal EEG Technology*, 11, 101-109
- [37] Cleveland, W. S. (1979), *Robust Locally Weighted Regression and Smoothing Scatterplots*. *Journal of the American Statistical Association*. 74

- [38] Clopper C. and Pearson S. (1934), *The use of confidence or fiducial limits illustrated in the case of the binomial*. *Biometrika* 26: 404-413
- [39] Cochran W.G. (1950), *The comparison of percentages in matched samples*. *Biometrika*, 37, 256-266
- [40] Cochran W.G. (1952), *The chi-square goodness-of-fit test*. *Annals of Mathematical Statistics*, 23, 315-345,
- [41] Cochran W. G. (1954), *Some methods for strengthening the common chi-square tests*. *Biometrics*, 10(4) 17-45 1
- [42] Cochran W.G. and Cox G.M. (1957), *Experimental designs (2nd ed.)*. New York: John Wiley and Sons.
- [43] Cohen J. (1960), *A coefficient of agreement for nominal scales*. *Educational and Psychological Measurement*, 10, 3746
- [44] Cohen J. (1968), *Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit*. *Psychological Bulletin*, 70, 213-220
- [45] Cohen J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey
- [46] Conover W. J. (1999), *Practical nonparametric statistics (3rd ed.)*. John Wiley and Sons, New York
- [47] Cox D.R. (1972), *Regression models and life tables*. *Journal of the Royal Statistical Society*, B34:187-220
- [48] Cramér H. (1946), *Mathematical models of statistics*. Princeton, NJ: Princeton University Press.
- [49] Cronbach L.J. (1951), *Coefficient alpha and the internal structure of tests*. *Psychometrika*, 16(3), 297-334
- [50] 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
- [51] Dunn O. J. (1964), *Multiple comparisons using rank sums*. *Technometrics*, 6: 241–252
- [52] Durbin J. (1951), *Incomplete blocks in ranking experiments*. *British Journal of Statistical Psychology*, 4: 85–90
- [53] 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
- [54] Epps T.W., Pulley L.B. (1983), *A test for normality based on the empirical characteristic function*. *Biometrika*. 1983;70:723–726
- [55] 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.
- [56] Fisher R.A. (1934), *Statistical methods for research workers (5th ed.)*. Edinburgh: Oliver and Boyd
- [57] Fisher R.A. (1935), *The logic of inductive inference*. *Journal of the Royal Statistical Society, Series A*, 98, 39-54
- [58] Fisher R.A. (1936), *The use of multiple measurements in taxonomic problems*. *Annals of Eugenics* 7 (2): 179–188

- [59] Fleiss J.L. (1971), *Measuring nominal scale agreement among many raters. Psychological Bulletin*, 76 (5): 378–382
- [60] Fleiss J.L., Cohen J. (1973), *The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability. Educational and Psychological Measurement*, 33, 613-619
- [61] Fleiss J.L., Levin B., Paik M.C. (2003), *Statistical methods for rates and proportions. 3rd ed. (New York: John Wiley) 598-626*
- [62] 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
- [63] Freireich E.O., Gehan E., Frei E., Schroeder L.R., Wolman I.J., et al. (1963), *The effect of 6-mercaptopurine on the duration of steroid induced remission in acute leukemia. Blood*, 21: 699–716
- [64] 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
- [65] 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.
- [66] 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
- [67] Gehan E. A. (1965a), *A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. Biometrika*, 52:203—223
- [68] Gehan E. A. (1965b), *A Generalized Two-Sample Wilcoxon Test for Doubly-Censored Data. Biometrika*, 52:650—653
- [69] Goodman L. A. (1960), *On the exact variance of products. Journal of the American Statistical Association*, 55, 708-713
- [70] Greenhouse S. W., Geisser S. (1959), *On methods in the analysis of profile data. Psychometrika*, 24, 95–112
- [71] Green S.B. (1991), *How many subjects does it take to do a regression analysis? Multivariate Behavioral Research*, 26, 499-510
- [72] Guttman L. (1945), *A basic for analyzing test-retest reliability. Psychometrika*, 10, 255-282
- [73] 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
- [74] 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
- [75] 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
- [76] Hanusz Z., Tarasińska J. (2014), *On multivariate normality tests using skewness and kurtosis, Colloquium Biometricum* 44, 139-148
- [77] Henderson, R. (1916), *Note on graduation by adjusted average. Transactions of the Actuarial Society of America*, 17:43–48

- [78] Henze N., Zirkler B. (1990), *A class of invariant consistent tests for multivariate normality*. *Comm. Statist. Theory Methods*. 1990;19:3595–3617
- [79] Hochberg Y. (1988), *A Sharper Bonferroni Procedure for Multiple Tests of Significance*. *Biometrika* 75 (4): 800–802
- [80] Holm S. (1979), *A simple sequentially rejective multiple test procedure*. *Scandinavian Journal of Statistics* 6, 65–70
- [81] Hotelling H. (1931), *The generalization of Student's ratio*. *Annals of Mathematical Statistics* 2 (3): 360–378
- [82] 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
- [83] 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
- [84] 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
- [85] Iman R. L., Davenport J. M. (1980), *Approximations of the critical region of the friedman statistic*, *Communications in Statistics* 9, 571–595
- [86] Jarque C. M., Bera A. K., (1987)., *A test for normality of Observations and Regression Residuals*, *International Statistical Review* 55, 163-172
- [87] 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
- [88] Jonckheere A. R. (1954), *A distribution-free k-sample test against ordered alternatives*. *Biometrika*, 41: 133–145
- [89] Kaplan E.L., Meier P. (1958), *Nonparametric estimation from incomplete observations*. *Journal of the American Statistical Association*, 53:457-481
- [90] Kendall M.G. (1938), *A new measure of rank correlation*. *Biometrika*, 30, 81-93.
- [91] Kendall M.G., Babington-Smith B. (1939), *The problem of m rankings*. *Annals of Mathematical Statistics*, 10, 275-287
- [92] Kleinbaum D. G., Klein M. (2005), *Survival Analysis: A Self-Learning Text, Second Edition (Statistics for Biology and Health)*
- [93] Kolmogorov A.N. (1933), *Sulla deterrnazione empirica di una legge di distribuzione*. *Giorn-de1'Inst. Ital. degli. Art.*, 4, 89-91
- [94] Kruskal W.H. (1952), *A nonparametric test for the several sample problem*. *Annals of Mathematical Statistics*, 23, 525-540
- [95] Kruskal W.H., Wallis W.A. (1952), *Use of ranks in one-criterion variance analysis*. *Journal of the American Statistical Association*, 47, 583-621
- [96] Lancaster H.O. (1961), *Significance tests in discrete distributions*. *Journal of the American Statisti-cal Association* 56:223-234
- [97] Lawley D. N. (1938), *A generalization of Fisher's z-test*. *Biometrika* 30: 180–187

- [98] Lee E. T., Wang J. W. (2003), *Statistical Methods for Survival Data Analysis*, ed. third, Wiley
- [99] Lenth, R. V., (2001), *Some Practical Guidelines for Effective Sample Size Determination*. *The American Statistician*, 55(3), 187-193
- [100] 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
- [101] 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.
- [102] 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
- [103] 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
- [104] 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
- [105] Lloyd S. P. (1982), *Least squares quantization in PCM*. *IEEE Transactions on Information Theory* 28 (2): 129–137
- [106] Lund R.E., Lund J.R. (1983), *Algorithm AS 190, Probabilities and Upper Quantiles for the Studentized Range*. *Applied Statistics*; 34
- [107] Mahalanobis P. C. (1930), *On tests and measures of group divergence*. *Journal of the Asiatic Society of Bengal* 26: 541–588
- [108] Mahalanobis P. C. (1936), *On the generalized distance in statistics*. *National Institute of Science of India* 12: 49–55
- [109] 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*, 18, 504
- [110] 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
- [111] Mantel N. (1963), *Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure*. *J. Am. Statist. Assoc.*, 58, 690-700
- [112] Mantel N. (1966), *Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration*. *Cancer Chemotherapy Reports*, 50:163—170
- [113] Marascuilo L.A. and McSweeney M. (1977), *Nonparametric and distribution-free method for the social sciences*. Monterey, CA: Brooks/Cole Publishing Company
- [114] Mardia K. V. (1970), *Measures of multivariate skewness and kurtosis with applications*, *Biometrika* 57, 519-530
- [115] Mardia K. V. (1974), *Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies*, *Sankhy B* 36, 115-128
- [116] Mauchly J. W. (1940), *Significance test for sphericity of n-variate normal population*. *Annals of Mathematical Statistics*, 11, 204-209.

- [117] McNemar Q. (1947), *Note on the sampling error of the difference between correlated proportions or percentages*. *Psychometrika*, 12, 153-157
- [118] Mehta C.R. and Patel N.R. (1986), *Algorithm 643. FEXACT: A Fortran subroutine for Fisher's exact test on unordered $r \times c$ contingency tables*. *ACM Transactions on Mathematical Software*, 12, 154–161
- [119] Miettinen O.S. (1985), *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*. John Wiley and Sons, New York
- [120] Miettinen O.S. and Nurminen M. (1985), *Comparative analysis of two rates*. *Statistics in Medicine* 4: 213-226
- [121] 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
- [122] Nadaraya, E. A. (1964), *On Estimating Regression*. *Theory of Probability and Its Applications*. 9 (1): 141–2
- [123] Newcombe R.G. (1998), *Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods*. *Statistics in Medicine* 17: 873-890
- [124] Newman S.C.(2001), *Biostatistical Methods in Epidemiology*. 2nd ed. New York: John Wiley
- [125] 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.
- [126] Ogilvie J. C. (1965), *Paired comparison models with tests for interaction*. *Biometrics* 21(3): 651-64
- [127] Orwin R. G. (1983), *A Fail-SafeN for Effect Size in Meta-Analysis*. *J Educ Behav Stat*, 8(2):157-159
- [128] 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
- [129] 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
- [130] 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
- [131] 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
- [132] Pillai K. C. (1955), *Some new test criteria in multivariate analysis*. *Annals of Mathematical Statistics* 26: 117–121
- [133] Plackett R.L. (1984), *Discussion of Yates' "Tests of significance for 2x2 contingency tables"*. *Journal of Royal Statistical Society Series A* 147:426-463
- [134] Pratt J.W. and Gibbons J.D. (1981), *Concepts of Nonparametric Theory*. Springer-Verlag, New York
- [135] 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

- [136] 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
- [137] Rosenbaum P.R., Rubin D.B. (1983a), *The central role of the propensity score in observational studies for causal effects*. *Biometrika*; 70:41–55
- [138] Rosenthal R. (1979), *The “file drawer problem” and tolerance for null results*. *Psychological Bulletin*, 5, 638-641
- [139] Rothman K.J., Greenland S., Lash T.L. (2008), *Modern Epidemiology*, 3rd ed. (Lippincott Williams and Wilkins) 221–225
- [140] Roy S. N. (1939), *p-statistics or some generalizations in analysis of variance appropriate to multivariate problems*. *Sankhya* 4: 381–396
- [141] Royston P. (1992), *Approximating the Shapiro–Wilk W-test for non-normality*”. *Statistics and Computing* 2 (3): 117–119
- [142] Royston P. (1993b), *A toolkit for testing for non-normality in complete and censored samples*. *Statistician* 42: 37–43
- [143] Rufibach K. (2010), *Assessment of paired binary data; Skeletal Radiology volume 40, pages1–4*
- [144] Satterthwaite F.E. (1946), *An approximate distribution of estimates of variance components*. *Biometrics Bulletin*, 2, 1 10-1 14
- [145] 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
- [146] 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
- [147] Scott D. W., (1992), *Multivariate Density Estimation. Theory, Practice and Visualization*. New York: Wiley.
- [148] Shapiro S.S. and Wilk M.B. (1965), *An analysis of variance test for normality (complete samples)*. *Biometrika* 52 (3–4): 591–611
- [149] Sheather S.J. (2009), *A modern approach to regression with R*. New York, NY: Springer
- [150] Shrout P.E., and Fleiss J.L (1979), *Intraclass correlations: uses in assessing rater reliability*. *Psychological Bulletin*, 86, 420-428
- [151] Šidák Z. K. (1967), *Rectangular Confidence Regions for the Means of Multivariate Normal Distributions*. *Journal of the American Statistical Association*, 62 (318): 626–633
- [152] Silverman B. W., (1986), *Density estimation for statistics and data analysis*, London: Chapman and Hall
- [153] 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
- [154] Sobel M. E. (1982). *Asymptotic confidence intervals for indirect effects in structural equation models*. *Sociological Methodology* 13: 290–312
- [155] Spearman C. (1910), *Correlation calculated from faulty data*. *British Journal of Psychology*, 3, 271-295

- [156] Tamhane A. C. (1977), *Multiple comparisons in model I One-Way ANOVA with unequal variances. Communications in Statistics, A6 (1), 15-32*
- [157] Tarone R. E., Ware J. (1977), *On distribution-free tests for equality of survival distributions. Biometrika, 64(1):156-160*
- [158] Tarone R.E. (1985), *On heterogeneity tests based on efficient scores. Biometrika 72, 91–95*
- [159] 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*
- [160] Terrell G. R. (1990), *The maximal smoothing principle in density estimation. Journal of the American Statistical Association 85, 470–477*
- [161] Terrell G.R., Scott D. W. (1985), *Oversmoothed nonparametric density estimates. Journal of the American Statistical Association 80, 209-214*
- [162] Thode H. C. (2002), *Testing For Normality. CRC Press; 2002. 506 s.*
- [163] Volinsky C.T., Raftery A.E. (2000) , *Bayesian information criterion for censored survival models. Biometrics, 56(1):256–262*
- [164] Wallenstein S. (1997), *A non-iterative accurate asymptotic confidence interval for the difference between two Proportions. Statistics in Medicine 16: 1329-1336*
- [165] Wallis W.A. (1939), *The correlation ratio for ranked data. Journal of the American Statistical Association, 34,533-538*
- [166] Watson, G. S. (1964), *Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A. 26 (4): 359–372*
- [167] Welch B. L. (1951), *On the comparison of several mean values: an alternative approach. Biometrika 38: 330–336*
- [168] Wilcoxon F. (1945), *Individual comparisons by ranking methods. Biometrics, 1,80-83*
- [169] Wilcoxon F. (1945), *Individual comparisons by ranking methods. Biometrics, 1, 80-83*
- [170] Wilcoxon F. (1949), *Some rapid approximate statistical procedures. Stamford, CT: Stamford Research Laboratories, American Cyanamid Corporation*
- [171] Wilcoxon F. (1949), *Some rapid approximate statistical procedures. Stamford, CT: Stamford Research Laboratories, American Cyanamid Corporation*
- [172] Wilcoxon F. (1949), *Some rapid approximate statistical procedures. Stamford, CT: Stamford Research Laboratories, American Cyanamid Corporation*
- [173] Wilson E.B. (1927), *Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association: 22(158):209-212*
- [174] Wilks S.S. (1932), *Certain generalizations in the analysis of variance. Biometrika 24: 471–494*
- [175] Yates F. (1934), *Contingency tables involving small numbers and the chi-square test. Journal of the Royal Statistical Society, 1,2 17-235*
- [176] Youden W.J. (1950), *Index for rating diagnostic tests. Cancer. 3: 32–35*

- [177] Yule G. (1900), *On the association of the attributes in statistics: With illustrations from the material of the childhood society, and c. Philosophical Transactions of the Royal Society, Series A, 194,257-3 19*
- [178] Zar J. H., (2010), *Biostatistical Analysis (Fifth Edition). Pearson Educational*
- [179] Zweig M.H., Campbell G. (1993), *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39:561-577*