KORZYSTANIE Z OKIENKOWEJ WERSJI TOOLBOXA BASICSTAT

BASICSTAT wersja 4.2,   Scilab wersja 6.1.1

Polecenia okienkowego toolboxa “Basicstat” (w podziale na 3 grupy jak w menu)

strona

STATYSTYKA OPISOWA I INNE OBLICZENIA

Nadawanie rang elementom ciągów 12
Obliczanie współczynników korelacji i regresja 13
Obliczanie wartości krytycznych - test znaków 15
Podstawowe statystyki - 1 próba 15
Regresja logistyczna 17
Rozkład Dunnetta i Tukeya 20
Tabele i wykresy 21
Tworzenie tablicy kontyngencji 27
Wieloczynnikowa regresja liniowa 29
Wyznaczanie częstości w klasach 32
Wyznaczanie wskaźnika GFR na podstawie poziomu kreatyniny oraz wieku i płci 36

TESTY PARAMETRYCZNE

Anova jednoczynnikowa i dwuczynnikowa 37
Test Browna Forsythea równości wariancji 39
Testy post hoc: Bonferroni-Dunnett-Tukey 39
Test proporcji (frakcji) dla 1 próby 40
Test proporcji (frakcji) dla 2 prób 40
Test średniej dla 1 próby 42
Test średnich dla 2 prób 42
Test wariancji dla 1 lub 2 prób 46
Test istotności współczynników korelacji 46
Wyznaczanie mocy testów i liczebności prób 49

TESTY NIEPARAMETRYCZNE

Miary asocjacji - współczynniki i testy istotności 54
Test chi kwadrat dla tablic kontyngencji (test niezależności i jednorodności) 55
Test chi kwadrat zgodności rozkładu danych z zadanym rozkładem 55
Test chi kwadrat zgodności z rozkładami ciągłymi 56
Test chi kwadrat zgodności z rozkładami dyskretnymi 60
Test dokładny Fishera dla tablicy kontyngencji 2×2 62
Test Kruskala-Wallisa 63
Test Kruskala-Wallisa dla tablic kontyngencji 65
Test McNemara dla tablicy 2×2 danych sparowanych 66
Test normalności Shapiro-Wilka 66
Test rangowanych znaków dla prób sparowanych 68
Test serii Walda-Wolfowitza losowości próby 70
Test serii Walda-Wolfowitza jednorodności dwóch populacji 70
Test U Manna-Whitneya 72
Test znaków dla prób sparowanych i dychotomicznych 72

1 Najważniejsze informacje o obiektach Scilaba niezbędne do korzystania z toolboxa "Basicstat"

Uwaga

Toolbox "Basicstat" można zainstalować w Scilabie począwszy od wersji 6.1.0 na komputerze z zainstalowanym 64 bitowym systemem Windows od wersji Windows 7 wzwyż. Wymaga on zainstalowania toolboxa "Distfun".

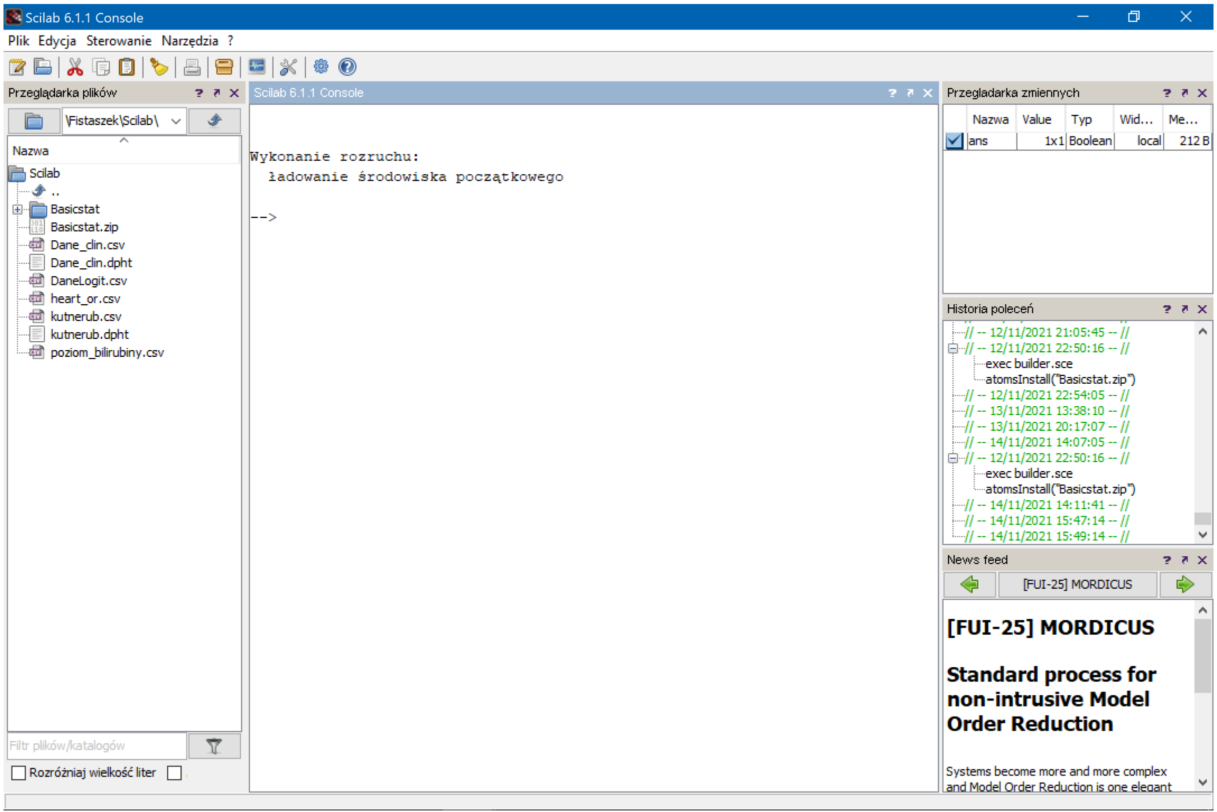
Toolbox ten został stworzony głównie z myślą o studentach uczelni medycznych początkowych lat jako pomoc w nauczaniu elementów statystki. Program (środowisko) Scilab jest darmowy i każdy może go zainstalować na swoim komputerze, uniezależniając się w ten sposób od komercyjnych wersji programów statystycznych. Proponowany toolbox jest dostosowany do treści podręcznika autorstwa dr Anny Baranowskiej "ELEMENTY STATYSTYKI dla studentów uczelni medycznych. Nowoczesne ujęcie z opisem obliczeń w programach Excel, R i Statistica" i zawiera on głównie różne testy statystyczne, w większości nieparametryczne, (w sumie jest ich 24) oraz pewną liczbę programów pomocniczych (w sumie jest ich 11). Po wpisaniu w odpowiednie okienka danych i ewentualnie wybraniu odpowiednich opcji i naciśnięciu odpowiedniego klawisza wyświetlane są wyniki w odpowiednich okienkach i polach oraz wykresy czy tabele, jeśli były dostępne takie opcje i je wybraliśmy. Wyniki te wyświetlane są w "czasie rzeczywistym", tj. w czasie nie dłuższym niż 2 sekundy na średniej klasy komputerze dla typowych danych używanych w dydaktyce, tj. rzędu kilkudziesięciu i tylko dla takich danych były one testowane. W wielu przypadkach otrzymamy wyniki w takim czasie również dla większych danych ale pewne przypadki z większymi danymi mogą wymagać nieakceptowalnie długiego czasu. Zanim przejdziemy do opisu poszczególnych okienek, opracowanie to zaczniemy od podania niezbędnych informacji o Scilabie, tak aby osoby spotykające się po raz pierwszy z tym programem, mogły korzystać z naszego toolboxa. Nie będą to jednak informacje wyczerpujące i opisujące wszystkie możliwe postaci opisywanych obiektów. Pełniejszych opisów należy szukać w innych źródłach.

1.1 Instalacja toolboxa "Basicstat"

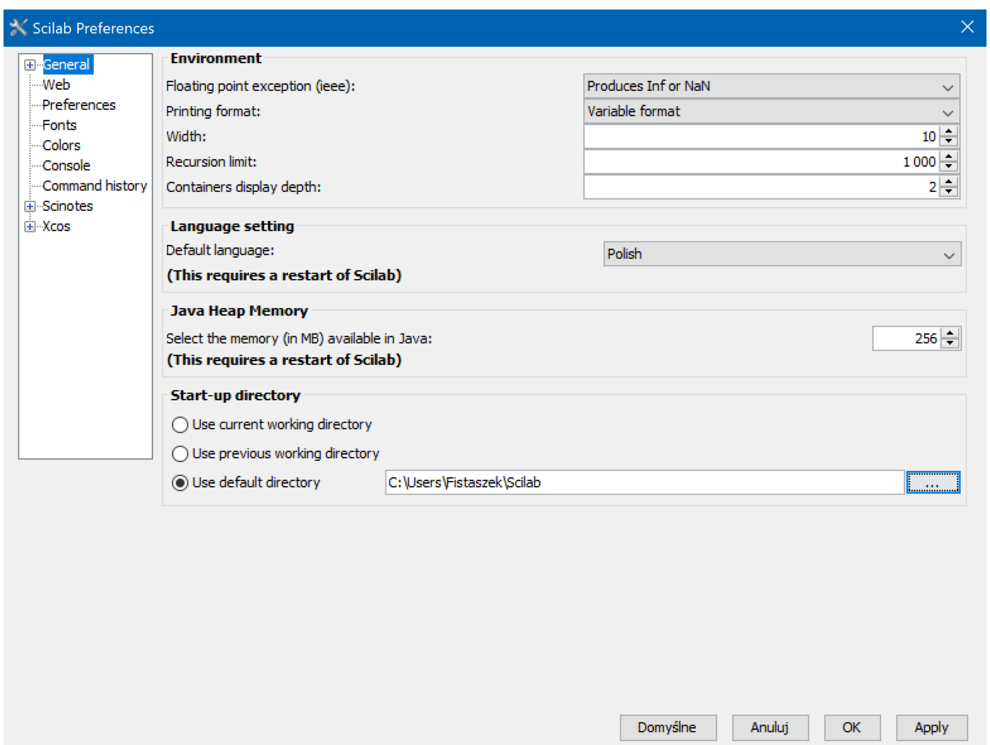
Aby zainstalować w scilabie toolbox "Basicstat" należy wykonać następujące czynności:

1. Kopiujemy plik `Basicstat.zip` i plik `scilab.ini` do katalogu roboczego Scilaba. Domyślnym katalogiem jest katalog `Documents` na dysku C, np. `"C:\Users\Fistaszek\Documents"` (oczywiście, nazwa "Fistaszek" jest tutaj nazwą umowną i każdy użytkownik będzie miał własną nazwę). Aby sprawdzić, który katalog jest aktualnym katalogiem roboczym Scilaba, piszemy w wierszu poleceń `pwd` i naciskamy `Enter`. Polecamy jednak utworzyć specjalny katalog dla plików Scilaba w katalogu `"C:\Users\Fistaszek\"` nazywając go, np. `scilab`, i tam umieścić oba pliki. Następnie w pasku menu konsoli należy wybrać `Edycja->Preferences`, kliknąć w `General` i u dołu zaznaczyć 'przełącznik' "Use default directory" (zob. Rys. 2) oraz kliknąć w '...' po prawej stronie tego 'przełącznika'. Po czym znaleźć i wybrać utworzony katalog `Scilab` i na koniec kliknąć w "OK" aby zamknąć okno ustawień. **Uwaga 1: Nie rozpakowujemy pliku Basicstat.zip. Natomiast plik scilab.ini wypakowujemy z pliku scilab_ini.zip**. **Uwaga 2: Instalacja toolboxa wymaga połączenia internetowego** (łączy się ze stroną internetową Scilaba).
2. W konsoli w wierszu poleceń wpisujemy `atomsInstall("Basicstat.zip")` (gdy Scilab nie został zainstalowany na dysku C, to wpisujemy `atomsInstall("Basicstat.zip","user")`) i naciskamy klawisz `Enter`.
3. Jeśli instalacja przebiegła pomyślnie, to w konsoli powinniśmy uzyskać komunikat taki jak na Rys. 3. Następnie zamykamy Scilaba i jeszcze raz go uruchamiamy. Po ponownym uruchomieniu informacja, że załadowany jest nasz toolbox powinna wyglądać jak na Rys. 4. Zwróćmy uwagę na dodanie elementu "Basicstat" na pasku menu. Kliknecie w niego otwiera okno z menu okienkowej wersji toolboxa.

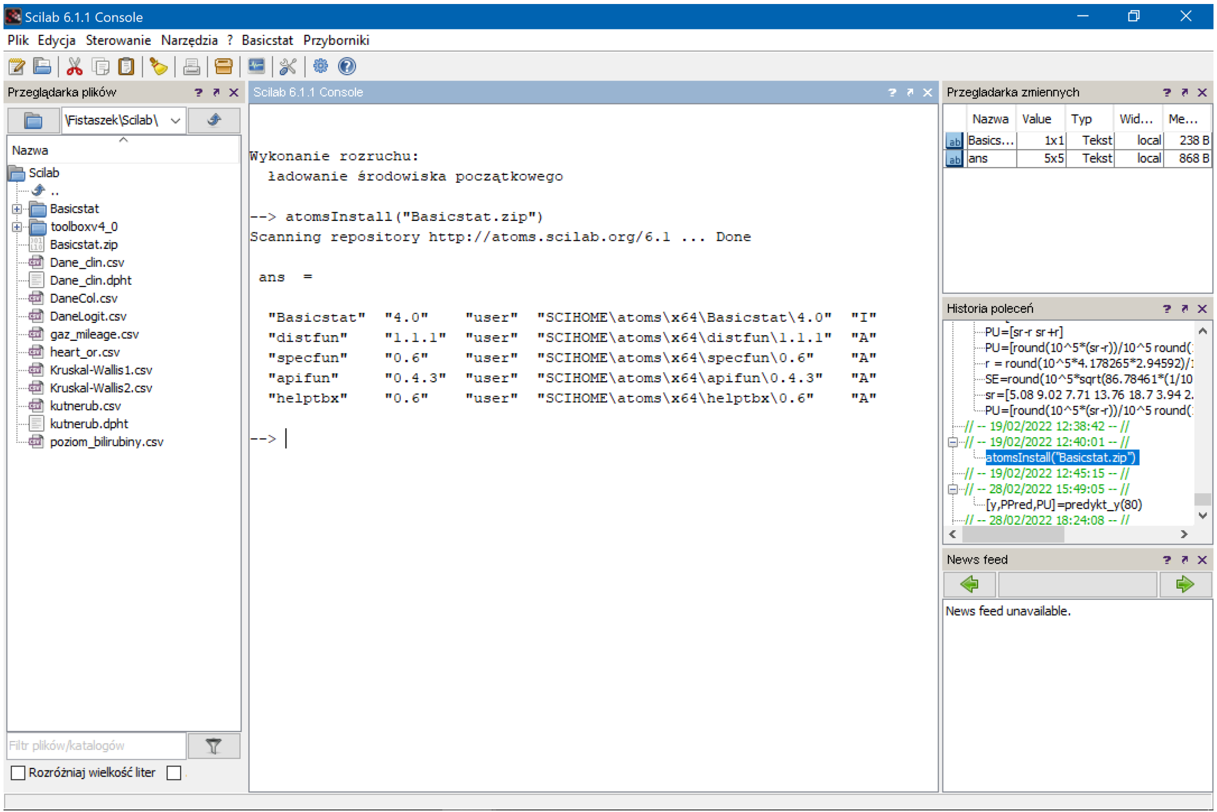
Aby skorzystać z systemu okienkowego toolboxa, najpierw klikamy w element "Basicstat" na pasku menu ekranu z konsolą Scilaba i otwieramy w ten sposób okienko, w którym znajdują się trzy rozwijane menus o nazwach: "Statystyka opisowa i inne obliczenia", "Testy parametryczne" i "Testy nieparametryczne". Z każdego z tych menus możemy wybrać potrzebne nam zadanie do wykonania przez najechanie myszką na odpowiedni element menu i kliknięcie.



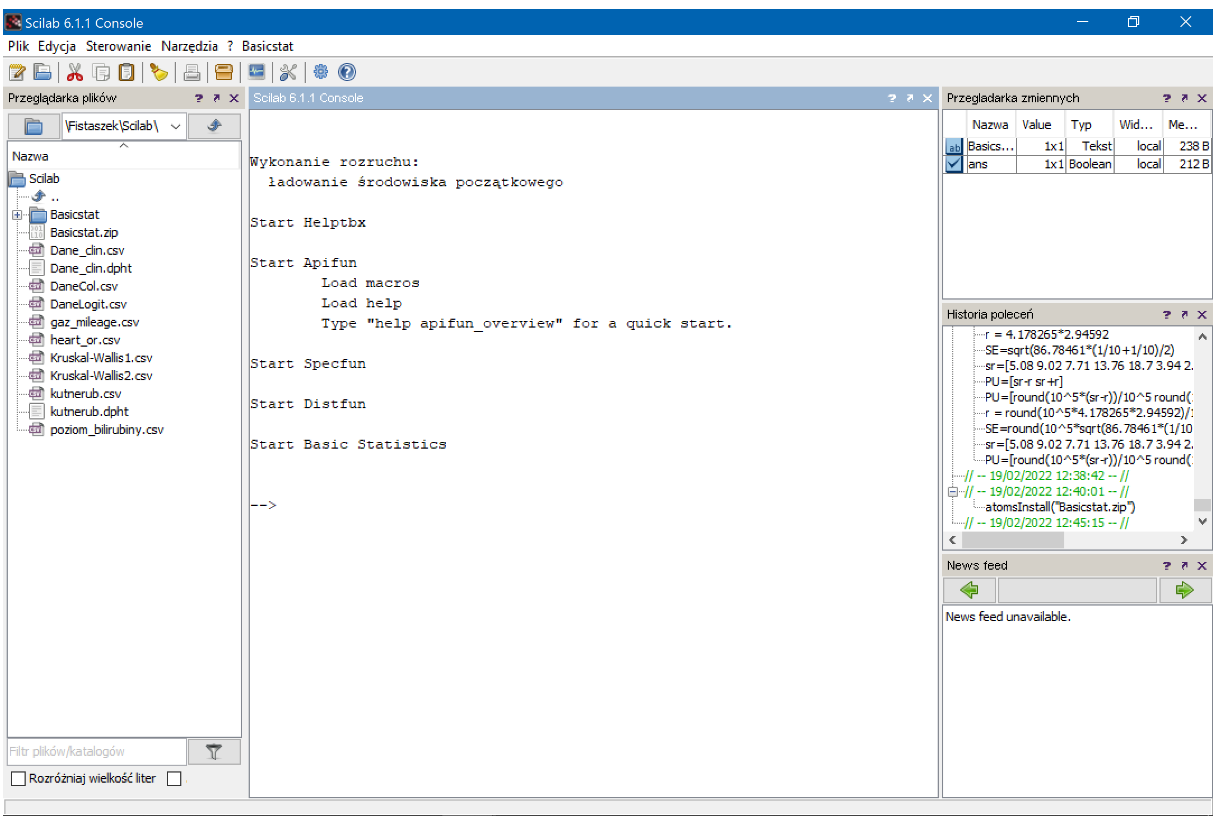
Rys. 1: Interfejs programu Scilab – okna (od lewej do prawej): okno przeglądarki plików, okno konsoli, okno przeglądarki zmiennych, okno historii poleceń i okno wiadomości strony internetowej Scilaba. Aby na górnej taśmie było pełne menu, okno konsoli musi być aktywne. O aktywności tego okna świadczy bladoniebieski górny brzeg okna konsoli.



Rys. 2: Ustawienie katalogu "Scilab" jako katalog roboczy programu Scilab



Rys. 3: Okno konsoli Scilaba po pomyślnym zainstalowaniu toolboxa "Basicstat"



Rys. 4: Okno konsoli Scilaba po ponownym uruchomieniu i załadowaniu funkcji toolboxa "Basicstat"

Najważniejsze obiekty Scilaba, których będziemy używać, korzystając z toolboxa[a] “Basicstat”, to

- **liczby i wartości logiczne**,
- **zmienne**,
- **łańcuchy**,
- **wektory**,
- **tablice**,
- **funkcje**.

[a] Inna nazwa dla toolboxa w Scilabie, to atom lub moduł.

Liczby i wartości logiczne

W Scilabie liczby mogą mieć różną postać. Wypiszmy zatem różne postacie liczb akceptowanych przez Scilab:

$$120 \quad 97.593211 \quad -5.27 \quad .4281 \quad 3578. \quad 2.54e{-}157 \quad -3e12$$

Jak widzimy, w Scilabie do zapisu liczb niecałkowitych używa się **kropki** dziesiętnej, a nie przecinka. Zauważmy, że .4281 3578. to skrócony zapis liczb 0.4281 3578.0.
Natomiast zapis

$$2.54e{-}157 \quad \text{oznacza liczbę} \quad 2.54 \cdot 10^{-157}, \qquad \text{a} \quad -3e12 \quad \text{oznacza liczbę} \quad -3 \cdot 10^{12}$$

W Scilabie wartościami logicznymi są %T lub %t (prawda) i %F lub %f (fałsz). W istocie, zamiast %T możemy użyć dowolnej liczby różnej od zera, a zamiast %F możemy użyć liczby 0.

Zmienne

W Scilabie zmiennym będziemy nadawać nazwy. Będziemy używać nazw, które będą ciągiem liter (małych lub dużych), liczb całkowitych i podkreślnika _. W nazwach zmiennych dopuszczalne są jeszcze inne znaki a pewne znaki nie są dopuszczalne. Jednak nie będziemy ich tutaj szczegółowo wymieniać. Podkreślmy jednak, że nazwy zmiennych **muszą zaczynać się od litery**. Przykłady nazw zmiennych:

a x5 abc ZYZ test_2proporcji Cena

Zmiennym możemy nadawać różne wartości, takie jak wartości liczbowe czy wartości logiczne. Inne obiekty, np. wektory czy tablice, też mogą być przyjmowane przez zmienne jako ich wartości. Nadawanie wartości zmiennych polega na przykład na wpisaniu w wierszu polecenia konsoli nazwy zmiennej, znaku =, a po nim np. liczby, nazwy obiektu lub wypisaniu odpowiedniego wyrażenia zawierającego np. nazwy innych zmiennych, wartości liczbowych, nazw funkcji z odpowiednimi argumentami itp. Następnie możemy wstawić na końcu średnik lub nie. Jeśli wstawimy średnik, to nie będzie w konsoli wyświetlona wartość zmiennej, której nadajemy tę wartość. A jeśli nie wstawimy średnika, to wartość tej zmiennej będzie wypisana w konsoli. Oczywiście, zamiast wpisywania polecenia w wierszu poleceń, możemy polecenia skopiować z innych źródeł,np. edytora, pliku pdf czy innego i wkleić do wiersza poleceń konsoli. Możemy też cały ciąg poleceń czy danych wypisać w edytorze Scilaba zwanym SciNotes, a następnie po zapisaniu ich w pliku o wybranej nazwie wykonać te polecenia lub wczytać wypisane dane, korzystając z menu tego edytora. Przykłady nadawania wartości zmiennym:

a=5; x5=5*a+7; A=[1 2 3 4]; B=[1 2;3 4];

Wyżej, nadaną wartością zmiennej A jest wektor o 4 elementach (współrzędnych) liczbowych, a zmiennej B została nadana tablica (macierz) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ o dwóch wierszach i dwóch kolumnach.

Łańcuchy

W Scilabie używa się obiektów zwanych łańcuchami. Nie będziemy podawać ogólnej postaci łańcuchów. Dla naszych celów wystarczy, gdy ograniczymy się do łańcuchów w postaci ciągów znaków takich jak wcześniej omówione nazwy zmiennych wziętych w apostrofy lub cudzysłów. Przy czym tym razem te ciągi znaków nie muszą zaczynać się od litery i mogą w nich też występować kropki, przecinki, znaki działań arytmetycznych czy logicznych. Przykładami łańcuchów są:

'a' "a_b" 'Janek' "234.3759" 'abra_kadabra100'

Łańcuchy mogą być nadawane jako wartości zmiennym, na przykład

a='5'; x5="5*a+7"; A=['k' 'm' 'm' 'm' 'k']; B=["a" "b"; "c" "d"];

Wektory

Podstawowym obiektem w Scilabie jest wektor[a]. Wektor składa się z dowolnej skończonej liczby elementów (w innym kontekście elementy te nazywa się współrzędnymi) wziętych w nawiasy kwadratowe. Elementy wektora oddziela się spacjami lub przecinkami. My będziemy używać wektorów, których elementami będą albo liczby albo łańcuchy.

Uwaga

Wszystkie elementy wektora muszą mieć ten sam typ, tj. wszystkie elementy muszą być albo liczbami albo łańcuchami. Nie można zatem użyć wektora, którego część elementów będzie liczbami i część elementów łańcuchami.

Przykłady zmiennych, których wartości są wektorami:

u=[0 1 3 8 18 6 4 4 3 0] v=[0.4 1.4 3.8 7.8 10.7 10.7 7.2 3.5 1.2 0.3] t=['a', 'a', 'b', 'a', 'b']

Szczególnym przypadkiem wektora (macierzy) jest pusty wektor (pusta macierz) postaci [ ]. Będzie on nam często przydatny jako argument funkcji w pewnych specjalnych sytuacjach.

Gdy wektor ma dużo elementów i chcemy, aby cały był widoczny w nie za wielkiej konsoli albo w oknie edytora, to można go "złamać" i zapisać w kilku linijkach. Ale wtedy musimy użyć trzech kropek (w istocie wystarczą dwie), aby wskazać, że w następnej linii są kolejne elementy tego wektora. Po wpisaniu trzech kropek i naciśnięciu `Enter` kursor zmienia swój kształt na > i Scilab czeka na wpisanie dalszych elementów wektora. Jeśli nie wstawimy tych kropek, to albo otrzymamy inny obiekt (tablicę, gdy w każdym wierszu będzie jednakowa liczba elementów) lub Scilab zasygnalizuje nam błąd. Przykład wektora zapisanego w kilku wierszach:

```
x=[12.07 12.02 12.00 12.01 11.98 11.96 12.04 12.05 12.01 11.97 ...
12.03 12.03 12.00 12.04 11.96 12.02 12.06 12.00 12.02 11.91 ...
12.05 11.98 11.91 12.01 12.06 12.02 12.05 11.90 12.07 11.98 ...
12.02 12.11 12.00 11.99 11.95 11.98 12.05 12.00 12.10 12.04 ...
12.06 12.04 11.99 12.06 11.99 12.07 11.96 11.97 12.00 11.97 ...
12.09 11.99 11.95 11.99 11.99 11.96 11.94 12.03 12.09 12.03 ...
11.99 12.00 12.05 12.04 12.05 12.01 11.97 11.93 12.00 11.97 ...
12.13 12.07 12.00 11.96 11.99 11.97 12.05 11.94 11.99 12.02 ...
11.95 11.99 11.91 12.06 12.03 12.06 12.05 12.04 12.03 11.98];
```

[a] W istocie podstawowym obiektem Scilaba jest tablica zwana macierzą. Natomiast wektor, a nawet liczba, to szczególny przypadek macierzy.

Tablice

Podstawowymi obiektami w Scilabie są tablice zwane **macierzami**. Aby móc korzystać z toolboxa "Basicstat", wystarczy wiedzieć co to jest macierz liczbowa. Mówiąc dość opisowo, przez macierz rozmiaru $m \times n$ (czytaj: m na n) będziemy rozumieli prostokątną tablicę liczb, która ma m wierszy i n kolumn. Liczby te będziemy nazywali elementami macierzy. Tablicom nadaje się nazwy i nazwy te traktuje się jak zmienne. Przykładami macierzy będą

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 0 & -1 \\ -1 & 4 & 1 & -5 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 4 & 1 \end{bmatrix} \quad w = \begin{bmatrix} 1 & 3 & 5 & 7 \end{bmatrix} \quad d = [5.3]$$

A jest macierzą 2×2, B macierzą 2×4, C macierzą 2×3, w macierzą 1×4 (wektorem wierszem), a d macierzą

1×1, czyli liczbą. Zauważmy, że na przecięciu dowolnego wiersza i dowolnej kolumny macierzy musi znajdować się liczba. Ponadto, wszystkie wiersze mają tę samą liczbę elementów oraz wszystkie kolumny mają też tę samą liczbę elementów, ale liczba elementów w wierszu nie musi być równa liczbie elementów w kolumnie. Jeśli jeden z wcześniej wymienionych warunków nie jest spełniony, to taki obiekt nie jest macierzą i Scilab sygnalizuje błąd.

Dlatego, gdy $D = \begin{bmatrix} 2 & 3 \\ 1 & -41 \end{bmatrix}$ $F = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $G = \begin{bmatrix} & 4 & \\ 5 & & 2 \\ 1 & & -1 \end{bmatrix}$,

to D jest macierzą 2×2, F jest macierzą 2×1, czyli wektorem (kolumną), a G **nie jest macierzą**.
Podane tu przykładowe wektory i macierze w Scilabie możemy zapisać następująco:

1. A=[1 2; -1 4] lub A=[1, 2; -1, 4], czyli liczby wypisywane po średniku są już elementami nowego wiersza

2. B=[1 2 0 -1; -1 4 1 -5]

3. C=[1 2 0; -1 4 1]

4. w=[1 3 5 7]

5. d=[5.3] lub po prostu d=5.3

6. D=[2 3; 1 -41]

7. F=[1; 0]

W poleceniu w edytorze SciNotes czy innym, po wypisaniu wiersza zamiast wpisywać 3 kropki, możemy przejść do nowego wiersza naciskając `Enter`. I to co napiszemy w nowym wierszu po wklejeniu do konsoli i naciśnięciu `Enter`, aby wykonał polecenia, będzie dla Scilaba nowym wierszem. Np., macierz B możemy utworzyć pisząc
B=[1 2 0 -1
-1 4 1 -5]

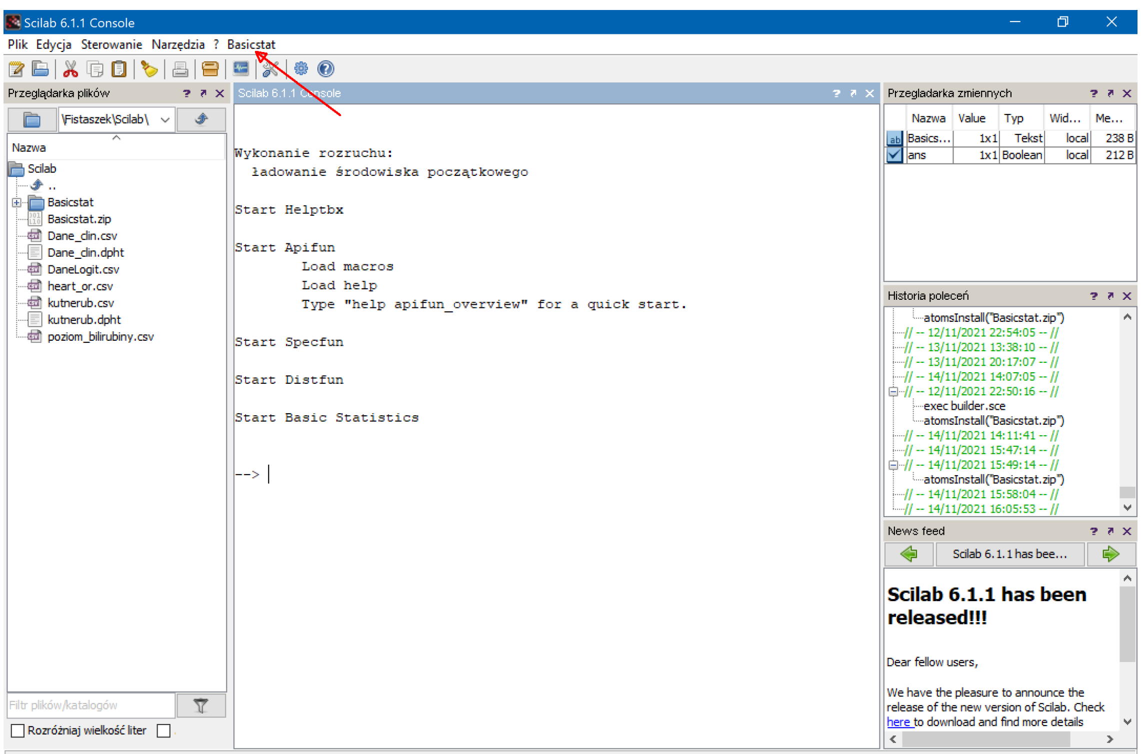
Wpisywanie danych mających postać wektora lub tablicy (macierzy)

- Choć w Scilabie elementy wektora można rozdzielać przecinkami i z zapisem przechodzić do nowego wiersza po uprzednim wstawieniu 3 kropek ..., to mocno zachęcamy do rozdzielania tych elementów spacjami i zapisywania ich w jednym wierszu, nawet jak jest dość długi. Używanie spacji do oddzielania elementów wektora ma tę zaletę przy wpisywaniu lub wklejaniu w okienka, że takie wiersze są automatycznie 'zawijane' (tj. gdy ich długość jest większa niż szerokość okienka, to automatycznie przechodzą do nowego wiersza bez tzw. "twardego znaku" nowego wiersza, który powstaje przy naciśnięciu klawisza 'Enter', i użycie którego jest niedozwolone). A jeśli wysokość okienka jest za mała, aby pomieścić w nim wszystkie tak utworzone wiersze, to automatycznie z prawej strony okienka pojawi się suwak. Dzięki temu będziemy mogli łatwo widzieć wszystkie dane i w przypadku, gdy w danych znajduje się jakiś błąd, łatwiej będzie go zauważyć.
- Choć w Scilabie elementy wektora znajdują się wewnątrz nawiasów kwadratowych, to wpisując wektory do okienek danych nawiasy te możemy pominąć, np. wpisując prawdopodobieństwa 0.05, 0.5 i 0.95 w odpowiednie okienko okna 'Podstawowe statystyki - 1 próba' w celu obliczenia kwantyli możemy je wpisać następująco: 0.05 0.5 0.95 lub [0.05 0.5 0.95].
- Wpisując wektory z danymi 'typu kategorialnego', np. wektor ["a" "b" "c"] lub ['a' 'b' 'c'] w okienko danych nie musimy używać ani cudzysłowów, ani apostrofów i taki wektor możemy wpisać następująco: [a b c] (ponieważ program pobierając takie dane sam dołoży takie znaki).
- Jeśli w Scilabie znajduje się jakiś wektor wcześniej wprowadzony, np. wczytaliśmy go z konsoli lub wygenerowaliśmy jakieś dane generatorem liczb pseudolosowych i znamy jego nazwę, a chcemy go użyć w okienku jako dane, to wystarczy wpisać w okienku danych jego nazwę, np. jeśli to jest wektor o nazwie v, to wpisujemy w okienku pojedynczą literę v i nic więcej.
- Zamiast wpisywać w okienko wektory danych możemy je kopiować z innych źródeł, a następnie je wklejać w okienko danych. Zwracamy uwagę, że na przykład kopiując dane z pliku pdf takie znaki jak apostrofy, znak sim (znak ~) czy - (znak minus) bardzo często po wklejeniu ma niewłaściwy kod i sygnalizowany jest błąd. Aby poprawić ten błąd, należy takie znaki ręcznie skasować i ponownie je wpisać, korzystając z klawiatury.

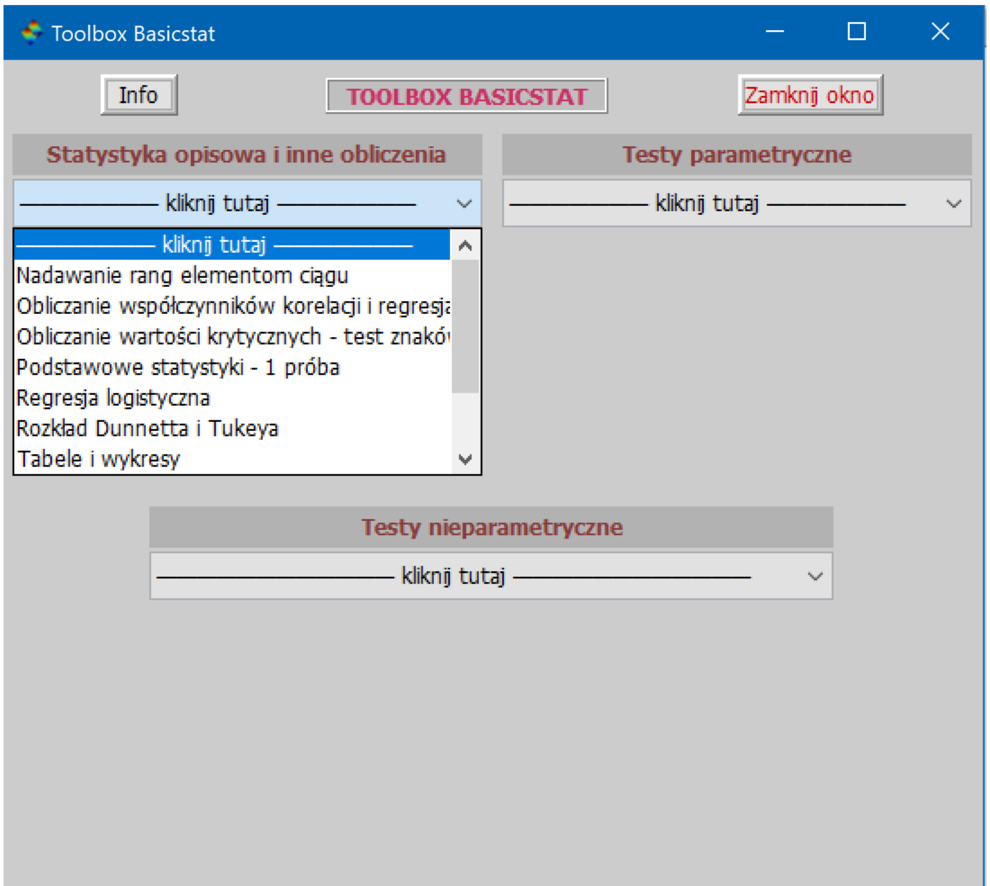
- Czasami choć bardzo rzadko zdarza się, że po wpisaniu w odpowiednie okienko danych i kliknięciu w odpowiedni klawisz, obliczenia nie są wykonywane lub sygnalizowany jest błąd danych, mimo że wcześniej dokładnie dla tych samych danych obliczenia były wykonywane i otrzymywaliśmy odpowiednie wyniki. Wtedy często pomaga zamknięcie całego programu Scilab i ponowne jego uruchomienie. Prawdopodobnie, w samym programie Scilab są jakieś błędy, ponieważ nie jest to program tak wszechstronnie przetestowany jak programy komercyjne. Pewien znany szwedzki matematyk, specjalista od matematyki obliczeniowej w jednej ze swoich książek napisał: “Jak coś jest za darmo, to nie należy od tego zbyt dużo oczekiwać”.
- Dwa okna: ’Test chi kwadrat dla tablic kontyngencji’ i ’Miary asocjacji - współczynniki i testy istotności’ zawierają okienka, w które należy wpisać tablice (kontyngencji).
  1. Tablicę kontyngencji możemy wpisać w okienko w formacie Scilaba: wpisujemy nawias kwadratowy otwierający, następnie wpisujemy kolejne elementy pierwszego wiersza oddzielane spacją i kończymy ten wiersz średnikiem. Wpisujemy w ten sam sposób drugi wiersz i kończymy go średnikiem, gdy nie jest to ostatni wiersz, a gdy jest to ostatni wiersz, to kończymy go nawiasem kwadratowym zamykającym. Wpisywanie nawiasów kwadratowych można pominąć.
  2. Drugi dopuszczalny sposób wpisania takiej tablicy do okienka jest następujący: wpisujemy kolejne elementy pierwszego wiersza oddzielane spacją i kończymy ten wiersz naciśnięciem klawisza ’Enter’, po czym wpisujemy kolejne wiersze w ten sam sposób. Pomiędzy wierszami danych nie powinno być pustych wierszy, czyli po wpisaniu każdego wiersza możemy tylko raz nacisnąć klawisz ’Enter’.

2 Wprowadzenie

Aby uruchomić menu okienkowej wersji Basicstat należy kliknąć w napis ’Basicstat’ na pasku menu okna konsoli wskazany strzałką na Rys. 5 i otrzymamy menu widoczne na Rys. 6.



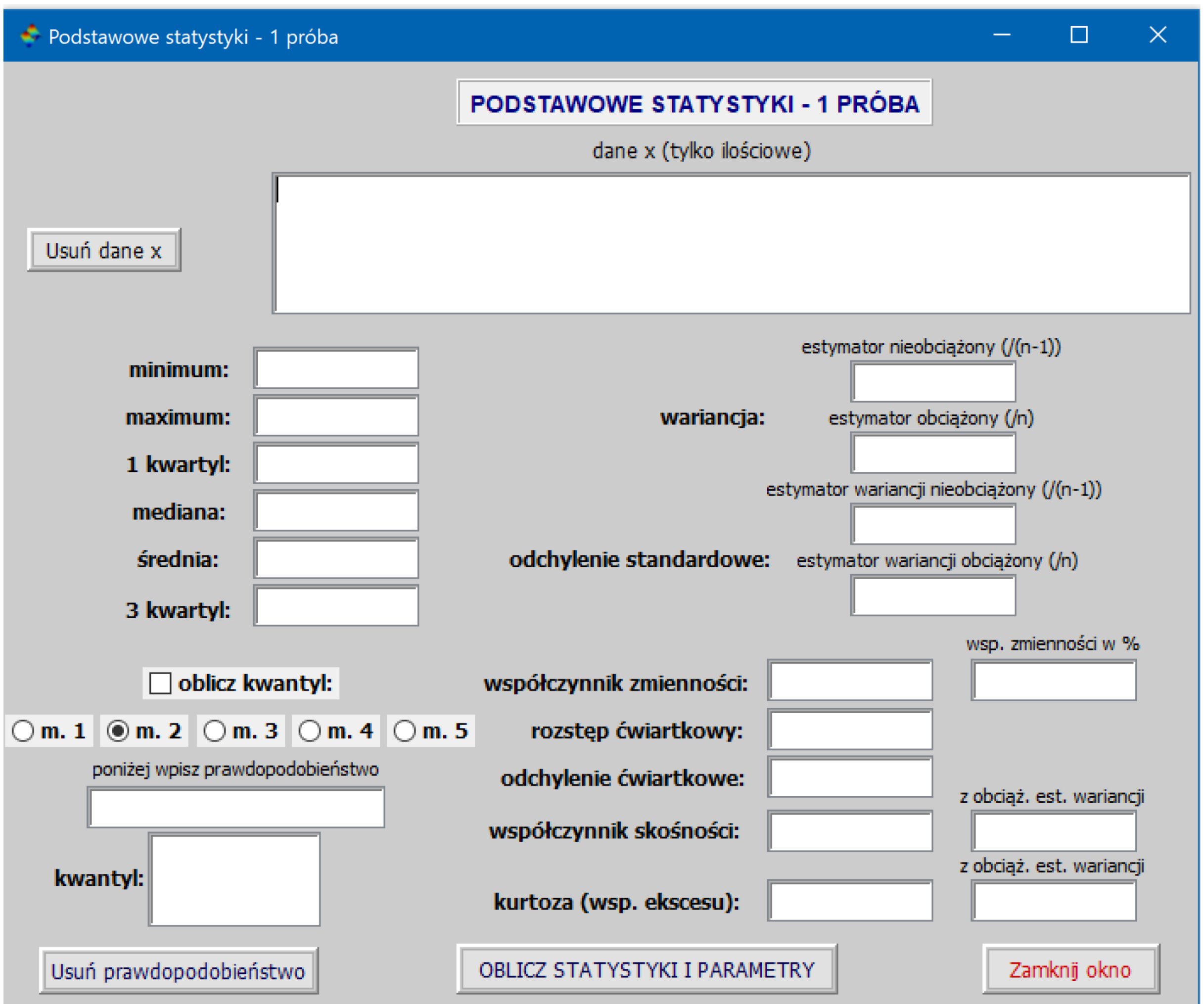
Rys. 5: Menu konsoli Scilaba uzupełnione o pozycję Basicstat



Rys. 6: Menu toolboxa Basicstat

Zostało ono podzielone na 3 grupy: “Statystyka opisowa i inne obliczenia”, “Testy parametryczne” i “Testy nieparametryczne”. Jak widzimy, w grupie “Statystyka opisowa i inne obliczenia” mamy 10 poleceń. Natomiast w grupie “Testy parametryczne” mamy 10 poleceń a w grupie “Testy nieparametryczne” mamy 15 poleceń. Kliknięcie w każde z 35 poleceń otwiera jedno nowe okno. Np. kliknięcie w polecenie ’Podstawowe statystyki - 1 próba’ otwiera okno widoczne na Rys. 7.
Widzimy, że u góry mamy okienko do wpisania danych. U dołu po lewej możemy zaznaczyć obliczanie kwantyli (z próby) oraz zaznaczyć jedną z 5 metod obliczania kwantyli oraz okienko do wpisywania prawdopodobieństw (można wpisać więcej niż jedno rozdzielając je spacją). Kliknięcie w ’OBLICZ STATYSTYKI I PARAMETRY’ spowoduje wykonanie obliczeń i wypełnienie wszystkich pozostałych okienek odpowiednimi wynikami. Omówimy po kolei jak należy korzystać ze wszystkich 35 okien, ale zanim to zrobimy podamy parę uwag ogólnych, a następnie przypomnimy postać wektora danych, które najczęściej będziemy wpisywać w odpowiednie okienka. Tylko trzy testy będą wymagały wpisania tablicy (macierzy) danych. Zwróćmy uwagę na poniższe informacje.

- Wybierając z menu polecenia nie możemy tego samego okna otworzyć 2 razy.
- W wielu oknach wyświetlając wyniki obliczeń równolegle wyświetlane są wyniki w konsoli Scilaba. Warto na nie też popatrzeć, ponieważ w konsoli czasami wyświetlane są ważne komunikaty, które nie są wyświetlane w okienkach, w tym komunikaty o błędach.

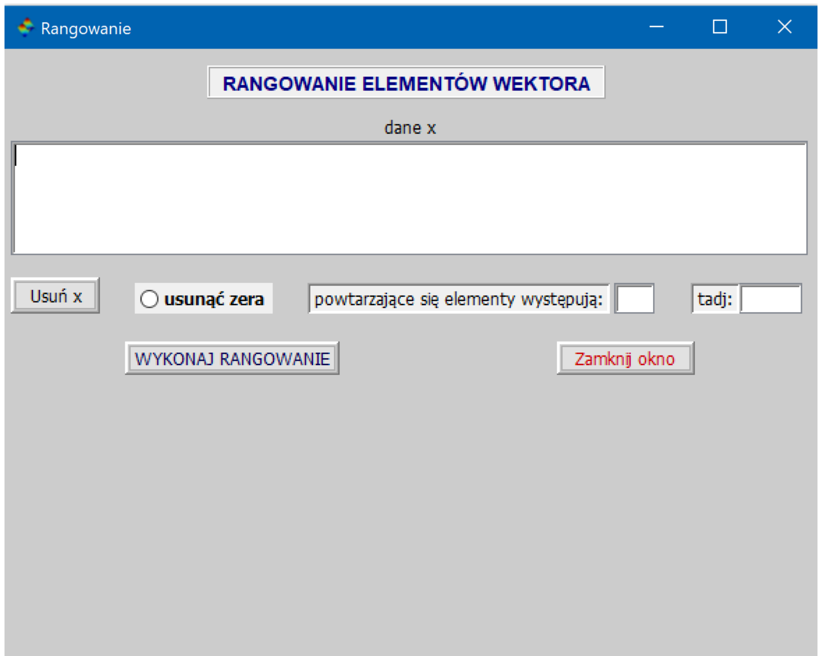


Rys. 7: Okno polecenia ’Podstawowe statystyki - 1 próba’

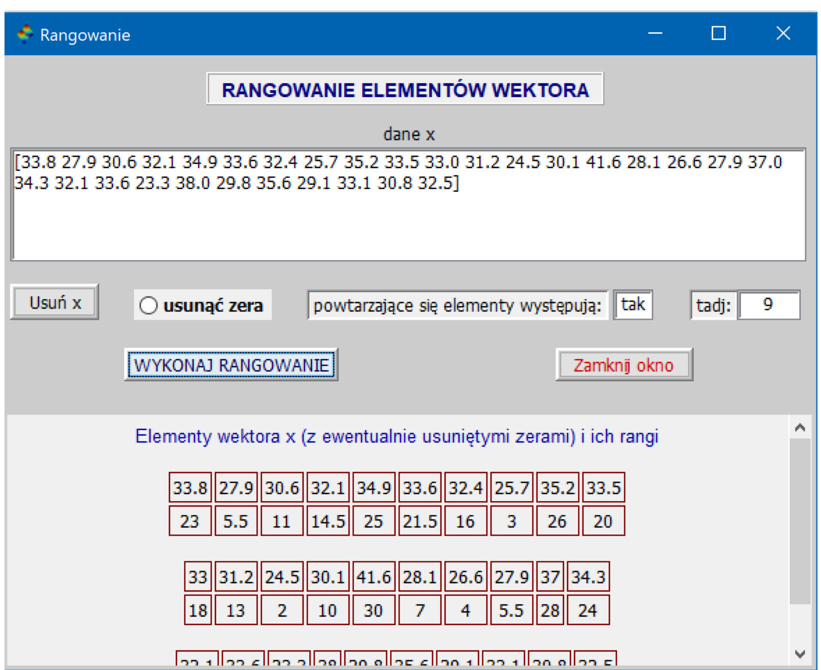
3 Grupa: “Statystyka opisowa i inne obliczenia”

Polecenie: ‘Nadawanie rang elementom ciągów’

1. Po wybraniu polecenia ‘Nadawanie rang elementom ciągów’ otrzymamy okno jak na Rys. 8.
2. Po wpisaniu danych w okienko ’dane x’, ewentualnym wybraniu opcji **usunąć zera** i kliknięciu w ’WYKONAJ RANGOWANIE’ otrzymamy wyniki jak na Rys. 9. Zauważmy, że w odpowiednich oknach ukazała się informacja o powtarzających się rangach i poprawka ’tadj’ (zob. ’Uwaga o poprawce ’tadj’ na następnej stronie).



Rys. 8: Okno RANGOWANIE ELEMENTÓW WEKTORA



Rys. 9: Okno RANGOWANIE ELEMENTÓW WEKTORA z wynikami obliczeń

Uwaga o poprawce 'tadj'

Poprawka 'tadj' na powtarzające się wartości (ties) równa jest $\left(\sum_{i=1}^{k}(t_i^3 - t_i)\right)/2$, gdzie k jest liczbą różnych grup równych wartości w ciągu statystyk pozycyjnych a t_i liczbą równych wartości w i-tej grupie. Poprawka `tadj` występuje bezpośrednio we wzorze na statystykę (przybliżenie rozkładem normalnym) w teście rangowanych znaków (Wilcoxona). Mamy też następujący związek `tadj` z poprawką dla statystyki (przybliżenie rozkładem normalnym) w teście U Manna-Whitneya (n_A, n_B - odpowiednio liczebności prób `x` i `y`):

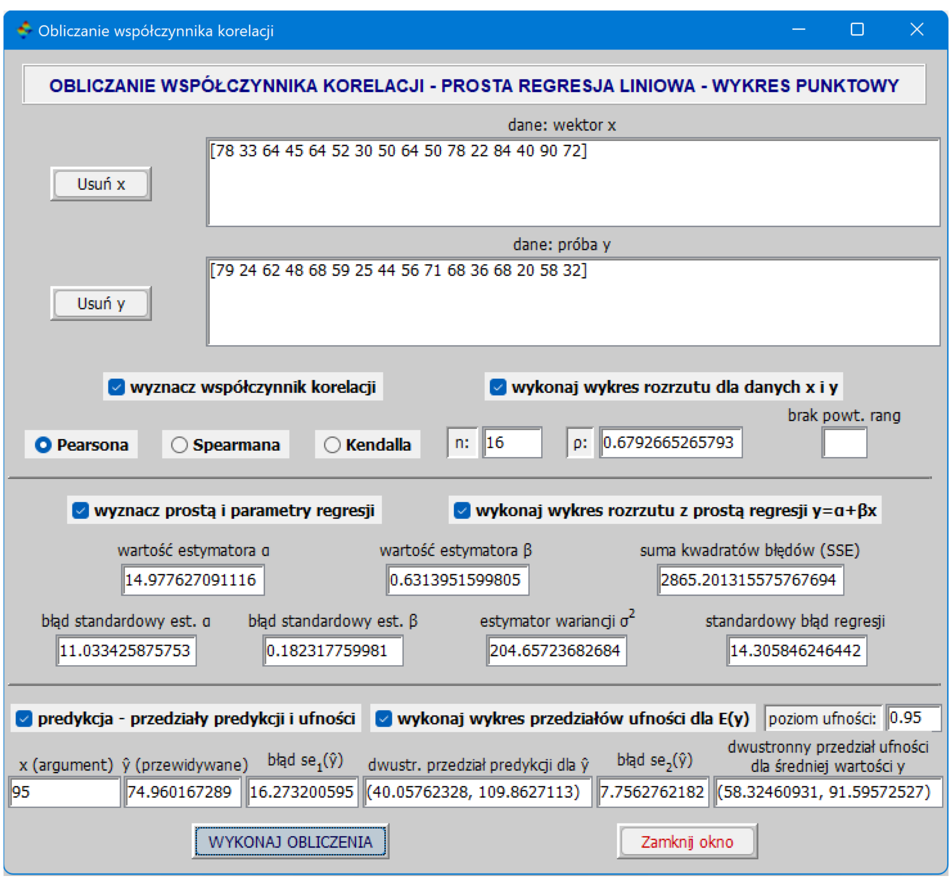
$$\frac{\sum_{i=1}^{k}(t_i^3 - t_i)}{(n_A + n_B)(n_A + n_B + 1)} \quad \left(= \frac{2}{(n_A + n_B)(n_A + n_B + 1)} \times \mathtt{tadj}\right)$$

Polecenie: 'Obliczanie współczynników korelacji i regresja'

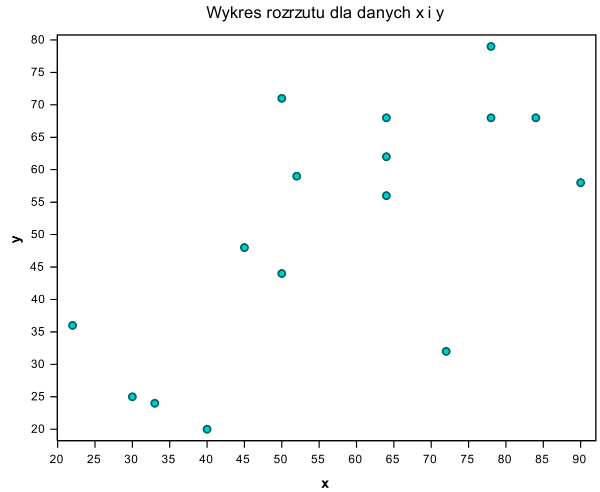
1. Po wybraniu polecenia 'Obliczanie współczynników korelacji i regresja', a następnie wpisaniu danych x i y, zaznaczeniu **wyznacz współczynnik korelacji**, **wykonaj wykres rozrzutu dla danych x i y**, pozostawienia domyślnego wyboru współczynnika korelacji **Pearsona**, zaznaczeniu **wyznacz prostą i parametry regresji** i **wykonaj wykres rozrzutu z prostą regresji** $\boldsymbol{y = \alpha + \beta x}$, zaznaczeniu **predykcja - przedziały predykcji i ufności** oraz **wykonaj wykres przedziałów ufności dla E(y)** i **wpisaniu w polu** 'x (argument)' liczby 95, otrzymamy okno jak na Rys. 10 oraz wykresy jak na Rys. 11, Rys. 12 i Rys. 13. W polu 'Brak powt. rang' będzie wpisane słowo 'tak' lub 'nie' tylko przy wybraniu opcji współczynnika **Spearmana** lub **Kendalla**. Wpisanie w tym polu słowa 'tak' oznacza, że brak jest takich samych rang.

   (1) Uwaga: Jeśli zaznaczamy opcję **predykcja - przedziały predykcji i ufności**, to nie należy zapominać o wpisanie w polu 'x (argument)' argumentu x, czyli wybranej przez siebie liczby x.

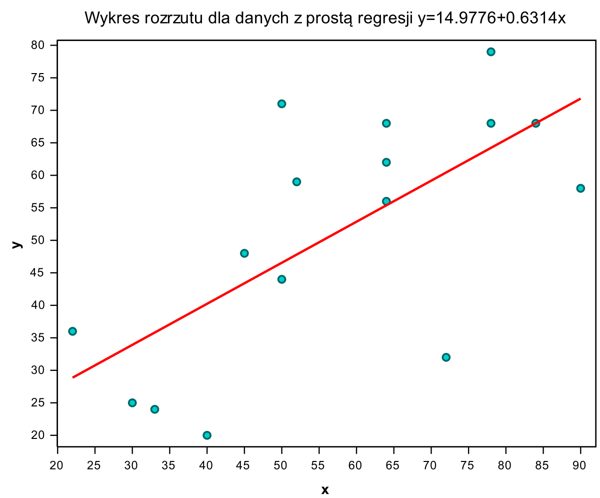
   (2) Rys. 13 – przedziały ufności dla średniej E(y), $y = \alpha + \beta x$ dla danego x: przedział na osi y między współrzędnymi y punktów przecięcia brązowych wykresów z pionową prostą przecinającą oś x w punkcie x.



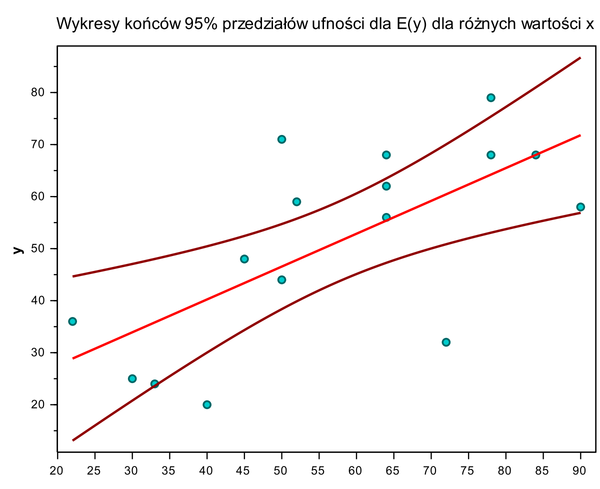
Rys. 10: Okno Obliczanie współczynników korelacji i regresja z wynikami obliczeń



Rys. 11: Wykres rozrzutu dla danych x i y



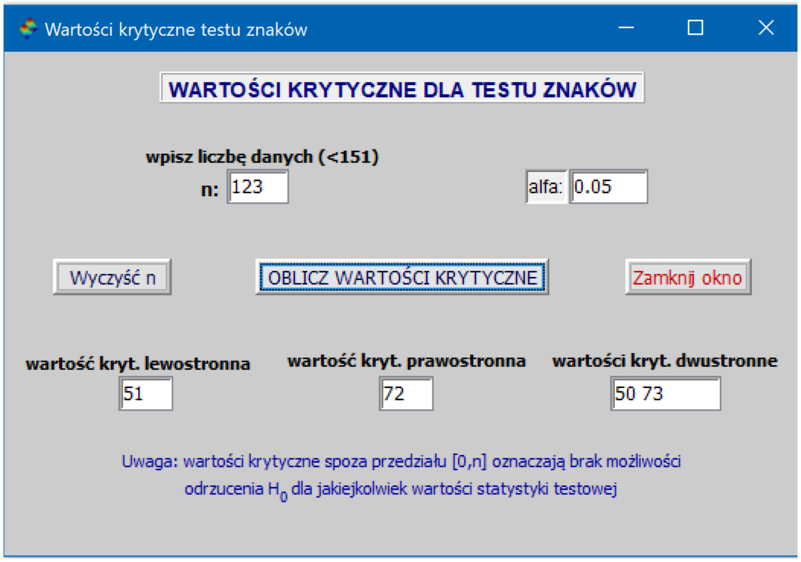
Rys. 12: Wykres rozrzutu danych z prostą regresji



Rys. 13: Wykres 95% przedziałów ufności dla wartości oczekiwanych E(y)

Polecenie: 'Obliczanie wartości krytycznych - test znaków'

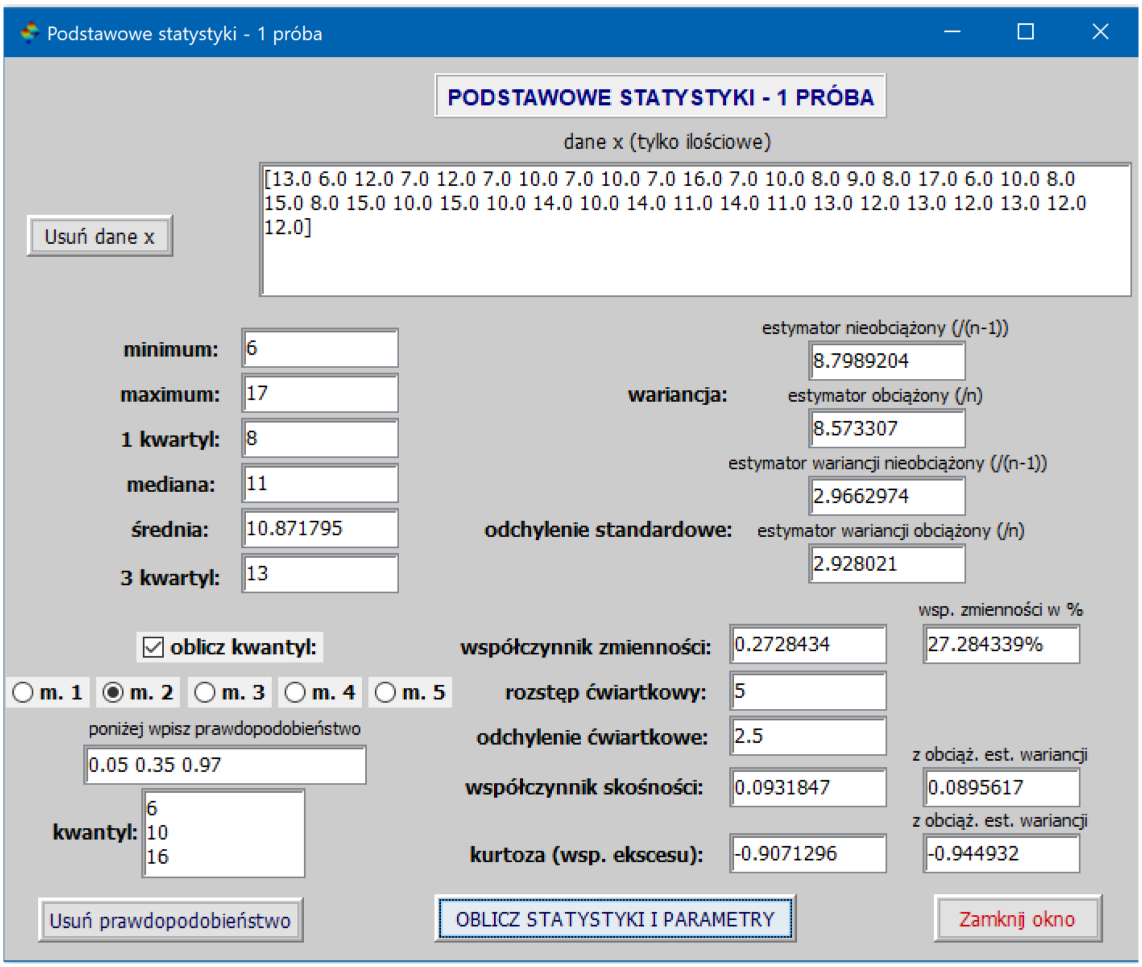
Po wybraniu polecenia 'Wartości krytyczne testu znaków' i wpisaniu w okienko **n:** liczby danych nie większej niż 150 oraz dokonaniu ewentualnej zmiany w polu 'alfa:' wartości parametru alfa i kliknięciu w klawisz 'OBLICZ WARTOŚCI KRYTYCZNE' otrzymamy wyniki jak na Rys. 14.



Rys. 14: Okno 'Wartości krytyczne testu znaków' z wynikami obliczeń

Polecenie: 'Podstawowe statystyki - 1 próba'

1. Wpisujemy dane w okienko 'Dane x'. Jeśli chcemy obliczyć kwantyl lub kwantyle, to zaznaczamy opcję **oblicz kwantyl** i wybieramy jedną z 5 metod obliczania kwantyli.
   (a) Opcja **m. 1** da nam kwantyl taki jak przy wyborze typu 1 w programie R i typu 3 w programie Statistica.
   (b) Opcja **m. 2** da nam kwantyl taki jak przy wyborze typu 2 w programie R, opcji domyślnej w programie Statistica i programie SAS.
   (c) Opcja **m. 3** da nam kwantyl taki jak przy wyborze typu 5 w programie R.
   (d) Opcja **m. 4** da nam kwantyl taki jak przy wyborze typu 6 w programie R, typu 2 w programie Statistica i użyciu funkcji PERCENTYL.PRZEDZ.ZAMK w programie Excel.
   (e) Opcja **m. 5** da nam kwantyl taki jak przy wyborze typu domyślnego w programie R i użyciu funkcji PERCENTYL.PRZEDZ.OTW w programie Excel.
2. Po wpisaniu danych x, zaznaczeniu **oblicz kwantyl:** i wpisaniu prawdopodobieństw 0.05 0.35 0.97 oraz kliknięciu w 'OBLICZ STATYSTYKI I PARAMETRY' otrzymujemy wyniki takie jak na Rys. 15.



Rys. 15: Okno ’Podstawowe statystyki - 1 próba’ z wynikami obliczeń

Polecenie: 'Regresja logistyczna'

Po kliknięciu w polecenie 'Regresja logistyczna' otwiera nam się okno, które pozwala otworzyć plik z rozszerzeniem .csv z zapisanymi w nim danymi. Plik z danymi ma zawierać dane zapisane w kolumnach, elementy których nie brane są w apostrofy czy cudzysłowy i oddzielone są od siebie przecinkami, a pierwsza kolumna ma zawierać nazwy kolumn (też oddzielone przecinkami). Kolumn może być wiele i mogą to być zarówno kolumny o elementach liczbowych jak i elementach typu kategorialnego. Wśród tych kolumn jedna kolumna ma być kolumną dychotomiczną, tj. (poza nazwą) ma zawierać tylko dwa różne (powtarzające się) elementy. Nazwijmy ją kolumną y. Jej elementy mogą być liczbowe (np. 0 i 1) lub kategorialne (np. tak, nie). Jeśli y jest kolumną z elementami typu kategorialnego, to oprócz wpisania jej nazwy w polu 'nazwa kolumny y' trzeba jej nazwę wpisać też w polu 'poniżej wpisz nazwy kolumn typu kategorialnego'. Oczywiście, wszystkie inne kolumny typu kategorialnego wybrane i wpisane w polu 'poniżej wpisz wybrane nazwy kolumn $x_1, x_2, \dots, x_n$' trzeba też wpisać w polu 'poniżej wpisz nazwy kolumn typu kategorialnego'. Natomiast w polu 'poniżej wpisz wybrane nazwy kolumn $x_1, x_2, \dots, x_n$' trzeba też wpisać wszystkie inne nazwy kolumn wybrane do modelu tak, aby razem z nazwą kolumny y tworzyły model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Wtedy zamiast y będzie wybrana prawdziwa nazwa kolumny, a zamiast $x_1, x_2, \dots, x_n$ będą nazwy kolumn wybranych do modelu. Aby było łatwiej zorientować się, jakie są nazwy kolumn i jakiego typu mają elementy, po kliknięciu w przycisk 'Wybierz plik .csv z danymi' i wybraniu odpowiedniego pliku, jego 3 pierwsze wiersze (o różnych kolorach) wpisane zostaną w polu o takiej nazwie. Trzy radiowe przyciski pozwalają nam wybrać 3 metody kodowania elementów w kolumnach, które nazwaliśmy $x_1, x_2, \dots, x_n$, i które są typu kategorialnego (tzw. 'dummy variables' - niektóre profesjonalne pakiety statystyczne mają do wyboru więcej metod). Wybrana kolumna y, jeśli jest typu kategorialnego, zawsze jest kodowana w jeden sposób z użyciem 0 i 1, jeśli jej elementami nie są 0 i 1. O kolejności decyduje porządek alfabetyczny, np. elementowi 'nie' nadawane jest 0 a elementowi 'tak' 1 - kolejność alfabetyczna będzie odwrócona, gdy zaznaczymy kratkę 'kolejność: najpierw 1'. Jeśli dana kolumna typu kategorialnego zawiera k różnych elementów, nazwijmy je $e_1, e_2, \dots, e_k$, to kolumna ta jest zastępowana k-1 kolumnami z odpowiednio dobranymi wartościami liczbowymi. Np., gdy dana kolumna ma 4 różne elementy e_1, e_2, e_3, e_4 i ich kolejność alfabetyczna jest właśnie taka, to każdemu elementowi e_i tej kolumny zostaną przyporządkowane 3 liczby, odpowiednio w wymienionych 3 kolumnach, zgodnie z regułą podaną w poniższej tabelce:

		ref1			refn			efekt		
e_1	$\longrightarrow$	0	0	0	1	0	0	1	0	0
e_2	$\longrightarrow$	1	0	0	0	1	0	0	1	0
e_3	$\longrightarrow$	0	1	0	0	0	1	0	0	1
e_4	$\longrightarrow$	0	0	1	0	0	0	-1	-1	-1

Natomiast, gdy dana kolumna ma 3 różne elementy e_1, e_2, e_3 i ich kolejność alfabetyczna jest właśnie taka, to każdemu elementowi e_i tej kolumny zostaną przyporządkowane 2 liczby, każda w osobnej kolumnie, zgodnie z regułą podaną w poniższej tabelce:

		ref1		refn		efekt	
e_1	$\longrightarrow$	0	0	1	0	1	0
e_2	$\longrightarrow$	1	0	0	1	0	1
e_3	$\longrightarrow$	0	1	0	0	-1	-1

Te nowe kolumny zastępujące daną kolumnę typu kategorialnego "dziedziczą" nazwę tej kolumny, przy czym, gdy kolumna typu kategorialnego ma co najmniej 3 różne elementy, to do tej dziedziczonej nazwy dodaje się wzięte w nawiasy kolejne cyfry zaczynając od 1, gdy wybrane jest kodowanie 'refn' lub 'efekt', i zaczynając od 2, gdy wybrane jest kodowanie 'ref1' ('ref1' jest kodowaniem domyślnym).

Ponadto, mamy jeszcze pole 'wpisz pary "nazwa1*nazwa2" (interakcja między zmiennymi)', w które możemy wpisać nazwy tych zmiennych $x_1, x_2, \dots, x_n$, dla których chcemy mieć model z interakcją, np. taki model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \beta_{n+1} x_1 * x_2 + \beta_{n+2} x_2 * x_n$$

Jeśli zaznaczymy kratkę 'zapisz w pliku funkcję predykcji y', to w domyślnym katalogu Scilaba w pliku o nazwie 'predykt_nazwakoly.sci' (zamiast nazwakoly będzie 'prawdziwa' nazwa wybranej kolumny y) zostanie zapisana funkcja o nazwie 'predykt_nazwakoly', której możemy użyć do obliczania wartości y dla danych $x_1,\ x_2,\ \dots\ , x_n$.

Jeśli jednak niektóre zmienne są typu kategorialnego, to:

1. jako wartości tej zmiennej możemy użyć tylko 0 lub 1, gdy taka zmienna przyjmuje tylko 2 różne wartości;

2. jako wartości k-1 zmiennych $e_1, \ldots, e_{k-1}$ (lub $e_2, \ldots, e_k$, jeżeli użyto opcji kodowania 'refn' albo 'efekt') możemy użyć tylko kombinacji zer i jedynki zgodnie z opisem w powyższych tabelkach (lub podobnych tabelkach dla $k > 4$), gdy taka zmienna przyjmuje k różnych wartości; np., gdy zmienna o przykładowej nazwie Typ przyjmuje 3 wartości: e_1, e_2, e_3 (tutaj k=3, a nie 4, jak w przykładowej tabelce z kodowaniem), to przy opcji 'ref1' jako parę argumentów $(Typ(2), Typ(3))$ weźmiemy odpowiednio (1,0), gdy Typ= e_2 i (0,1), gdy Typ= e_3; natomiast dla opcji 'refn' lub 'efekt' jako parę argumentów $(Typ(1), Typ(2))$ weźmiemy odpowiednio (1,0), gdy Typ= e_1 i (0,1), gdy Typ= e_2.

Jeśli dokonamy opisanych tutaj wpisów i wyborów, to klikamy w 'WYKONAJ OBLICZENIA' i otrzymamy wyniki, które zawierają odpowiednie współczynniki $\beta_0,\ \beta_1,\ \ldots\ ,\beta_m$ wraz z wynikami testów tych współczynników i przedziałów ufności oraz inne parametry. Ponadto, otrzymamy też wyestymowane ilorazy szans wraz z przedziałami ufności.

Uwaga

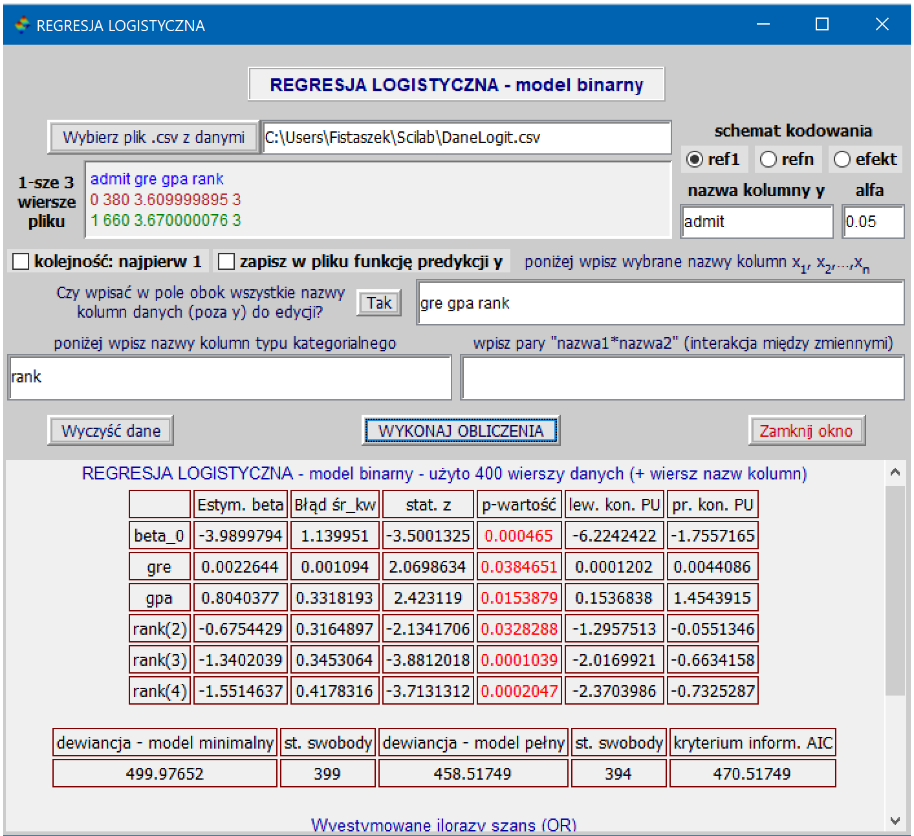
1. Dla zmiennych (włączając dodane kolumny związane z kodowaniem w przypadku zmiennych typu kategorialnego) będących elementami par wypisanych w polu 'wpisz pary "nazwa1*nazwa2" (interakcja między zmiennymi)' nie oblicza się ilorazów szans i odpowiadających im przedziałów ufności.
2. W polu 'wpisz pary "nazwa1*nazwa2" (interakcja między zmiennymi)' nie można wpisać więcej niż iloczyn dwóch zmiennych, czyli do modelu nie można włączać interakcji dla więcej niż dwóch zmiennych, np. "nazwa1*nazwa2*nazwa3". Jeśli jest taka potrzeba, to należy skorzystać z profesjonalnych pakietów statystycznych.
3. Podobnie jak w przypadku wieloczynnikowej regresji liniowej - zobacz opis tej regresji na str. 29 - można dokonywać na wiele sposobów (automatycznego) dopasowywania modelu. Nie zaimplementowaliśmy tutaj żadnej z takich metod, a gdy zachodzi taka potrzeba, to należy skorzystać z profesjonalnych pakietów statystycznych.

Rozpatrzmy dwa przykłady. W pierwszym z nich, dane są zapisane w pliku o nazwie `DaneLogit.csv`, który ma 4 kolumny, i którego pierwsze 4 wiersze mają postać:

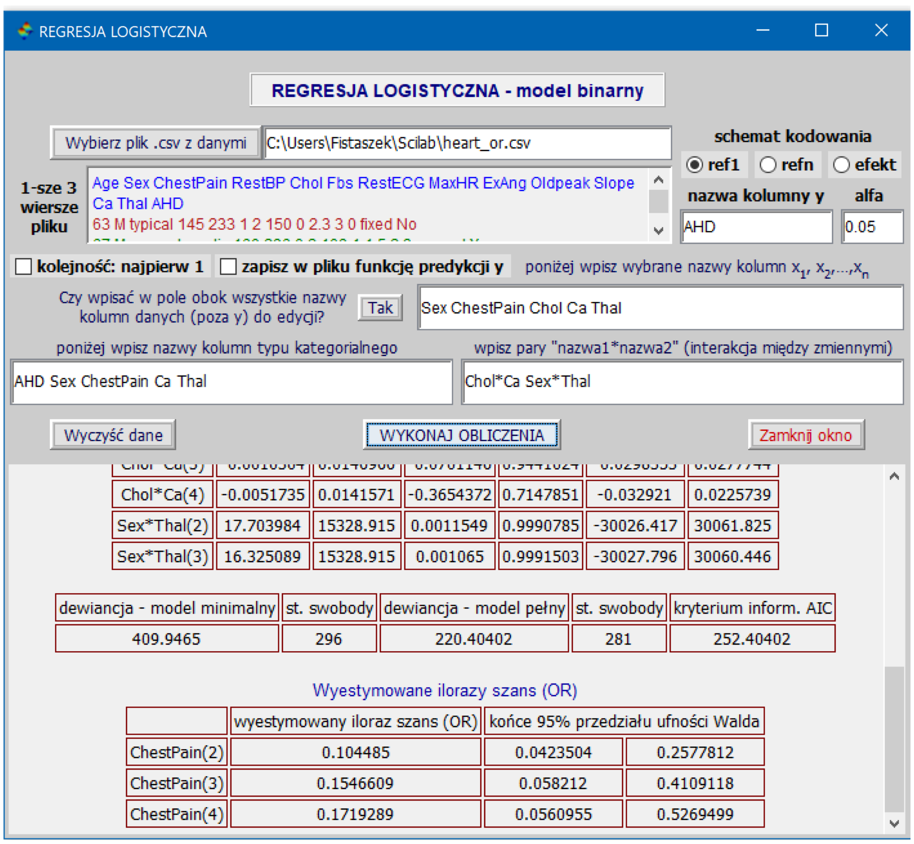
admit	gre	gpa	rank
0	380	3.609999895	3
1	660	3.670000076	3
1	800	4	1

Tutaj zmienną zależną y jest admit. Jest ona zmienną dychotomiczną - przyjmuje 2 wartości i tymi wartościami są 0 i 1 - była ona wcześniej zmienną typu kategorialnego ale jej wartości zostały zamienione na 0 i 1 i nie musimy jej nazwy zapisywać w polu 'poniżej wpisz nazwy kolumn typu kategorialnego' (wpisanie niczego by nie zmieniło). Natomiast należy ją wpisać w polu 'nazwa kolumny y'. Widzimy też, że zmienna rank, choć jej wartości są liczbami naturalnymi, to pełnią one rolę nazw kategorii, a zatem jest ona typu kategorialnego. Po wpisaniu nazwy admit w wymienionym polu klikamy w przycisk 'Tak' i pozostałe nazwy zostaną automatycznie wpisane w odpowiednim polu. Wpisujemy nazwę 'rank' (bez apostrofów) w polu 'poniżej wpisz nazwy kolumn typu kategorialnego', klikamy w przycisk 'WYKONAJ OBLICZENIA' i otrzymujemy wyniki jak na Rys. 16.

Jako drugi przykład weźmy dane z pliku `heart_or.csv` o 14 kolumnach, którego 4 pierwsze wiersze podane zostały w opisie polecenia 'Tworzenie tablicy kontyngencji' na str. 27. Po otwarciu tego pliku wybieramy opcję kodowania 'refn'. Zmienną dychotomiczną niezależną jest zmienna AHD - jest ona typu kategorialnego, zatem wybieramy ją jako zmienną y wpisując w polu 'nazwa kolumny y' i w polu 'poniżej wpisz nazwy kolumn typu kategorialnego', Do modelu dodajemy zmienne o nazwach Sex, ChestPain, Chol, Ca i Thal (jest to nasz wybór, można wybrać inne jeśli są powody ku temu) wpisując je w polu 'poniżej wpisz wybrane nazwy kolumn $x_1, x_2, \ldots, x_n$'. Ponieważ zmienne Sex, ChestPain, Ca i Thal są typu kategorialnego, to wpisujemy je też w polu 'poniżej wpisz nazwy kolumn typu kategorialnego'. Do interakcji wybieramy Chol, Ca, Sex i Thal, wpisując Chol*Ca Sex*Thal w polu 'wpisz pary "nazwa1*nazwa2" (interakcja między zmiennymi)'. Klikamy w przycisk 'WYKONAJ OBLICZENIA' i otrzymujemy komunikat, że w niektórych wierszach brakuje danych (w ich miejsce wpisane jest NA). Po zatwierdzeniu zgody na usunięcie tych wierszy otrzymujemy wyniki jak na Rys. 17. Zauważmy, że ilorazy szans OR zostały obliczone tylko dla ChestPain(1), ChestPain(2) i ChestPain(3), ponieważ tylko ChestPain nie został wybrany do interakcji.



Rys. 16: Wyniki obliczeń dla regresji logistycznej dla danych w pliku DaneLogit.csv



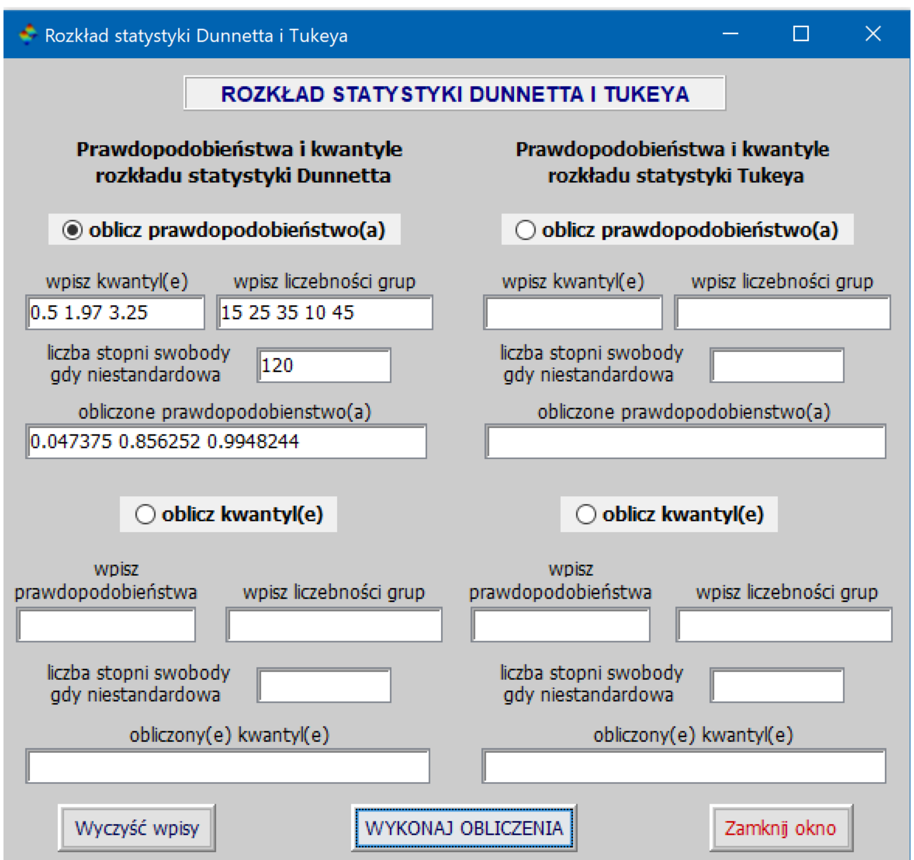
Rys. 17: Wyniki obliczeń dla regresji logistycznej dla danych w pliku heart_or.csv

Polecenie: ’Rozkład Dunnetta i Tukeya’

1. Wybierając polecenie ’Rozkład Dunnetta i Tukeya’ możemy obliczyć prawdopodobieństwa i kwantyle dla tych rozkładów (za każdym razem do obliczenia można wybrać tylko jeden z 4 rodzajów parametrów).
2. Różne liczebności prób są uwzględniane w obliczeniach. Np. wpisując dla rozkładu Dunnetta w polu ’wpisz kwantyl(e)’ 0.5 1.97 3.25, w polu ’wpisz liczebności grup’ 5 grup 15 25 35 10 45, a w polu ’liczba stopni swobody’ 120, otrzymamy wyniki takie jak na Rys. 18.

Uwaga

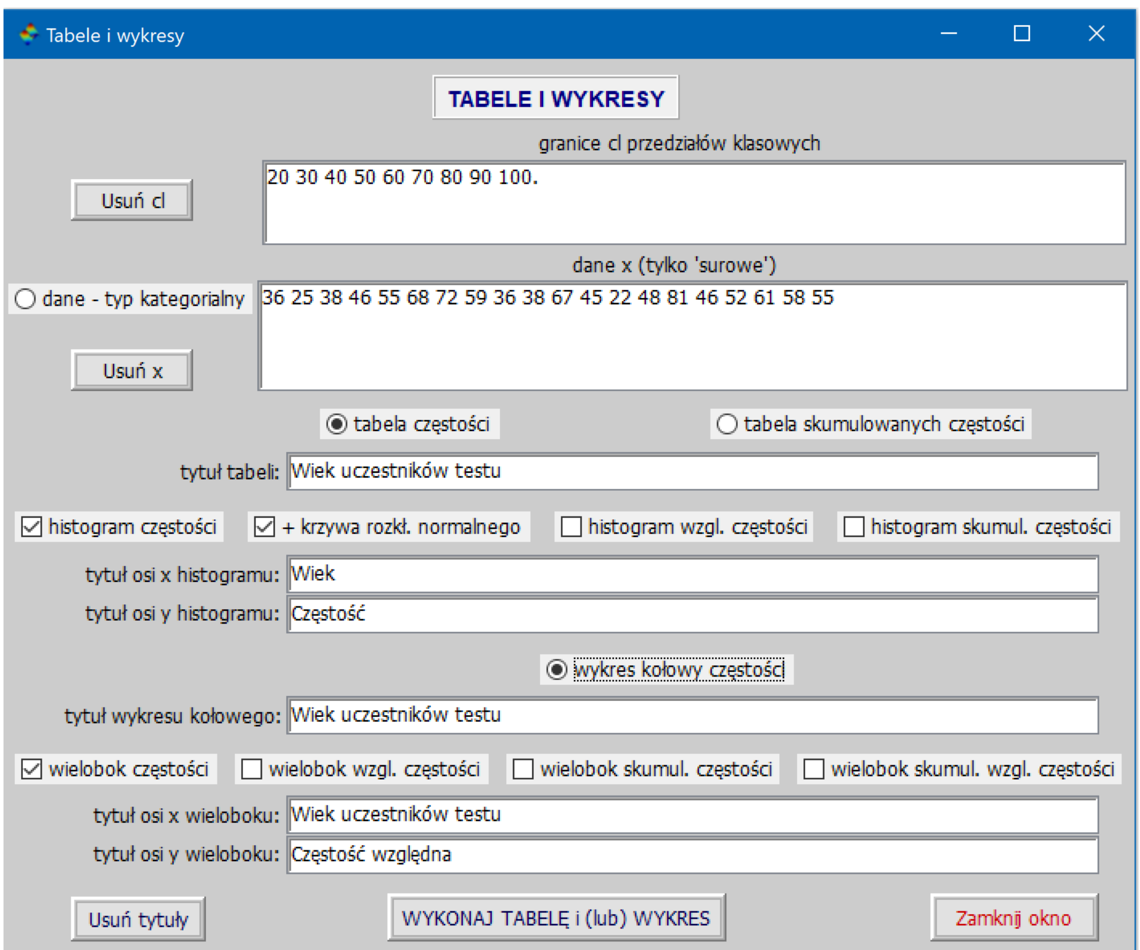
W oknie ’Rozkład Dunnetta i Tukeya’ znajdują się 4 pola oznaczone napisem: ’liczba stopni swobody gdy niestandardowe’ (w pierwszym z nich wpisana jest liczba 120). Standardowa liczba stopni swobody, to suma podanych liczebności grup minus liczba grup i jest wyliczana na podstawie wektora liczebności w polu ’wpisz liczebności grup’. Dla jednoczynnikowej anovy i następujących po niej post hoc testów wykorzystywana przy obliczaniu kwantyli i prawdopodobieństw liczba stopni swobody jest standardowa. Gdy wykonujemy jednak test dwuczynnikowej anovy, a następnie test ’post hoc’, to przy obliczaniu kwantyli i prawdopodobieństwa za liczbę stopni swobody bierzemy liczbę stopni swobody błędu z tablicy anovy, a taką liczbę stopni swobody nie zawsze oblicza się tak jak ’standardową liczbę stopni swobody’. Dlatego mamy możliwość wpisania tej liczby, jeśli np. ręcznie chcielibyśmy wykonać post hoc test metodą Tukeya lub Dunnetta.



Rys. 18: Okno ’Rozkład Dunnetta i Tukeya’ z przykładowymi wynikami obliczeń

Polecenie: 'Tabele i wykresy'

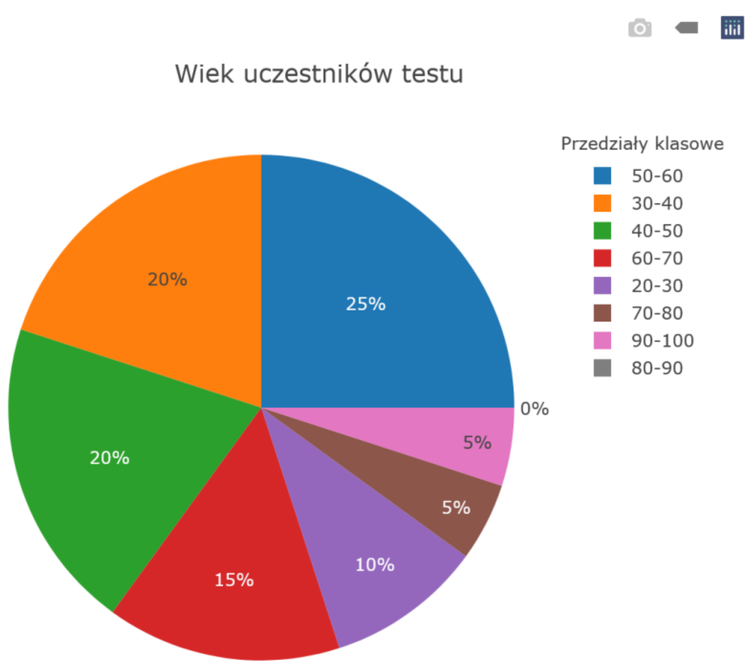
1. Po otwarciu okna 'Tabele i wykresy' wpisujemy dane "surowe" x w okienko o tej nazwie. Jeśli są to dane ilościowe i nie będziemy traktować ich jako dane kategorialne, to aby otrzymać dla nich histogram, musimy wpisać przedziały klasowe w okienko o nazwie 'granice cl przedziałów klasowych'.
2. Nie można wybrać wszystkich możliwych tabeli i wykresów na raz. Możemy wybrać jedną tabelę, jeden histogram, wykres kołowy i wielobok częstości. Możemy też wpisać w odpowiednie rubryki tytuł tabeli i tytuł wykresu kołowego oraz tytuły osi x i y histogramu i wieloboku częstości.
3. Tabele otrzymujemy w formacie HTML i otwierane są domyślną przeglądarką internetową. Narysowanie wykresu kołowego wymaga połączenia internetowego. Jest on też w formacie HTML i jest otwierany przeglądarką. Można go zapisać w formacie png klikając w ikonkę aparatu fotograficznego.
4. Histogram i wielobok częstości otrzymujemy w specjalnym okienku Scilaba i możemy go wyeksportować do kilku formatów graficznych, np. do pdf lub png.
5. Następnie można wybrać inne rodzaje tabeli, histogramu czy wieloboku częstości. Pamiętajmy jednak, że jeśli jakiś rodzaj wykresu już został wykonany, to nie można go drugi raz wykonać, jeśli nie zamkniemy już istniejącego wykresu.
6. Po wpisaniu danych, przedziałów klasowych i wybraniu odpowiedniej tabeli, wykresu, histogramu i wieloboku, tak jak na Rys. 19, otrzymamy tabelę, wykres, histogram i wielobok, takie jak na Rys. 20–23.



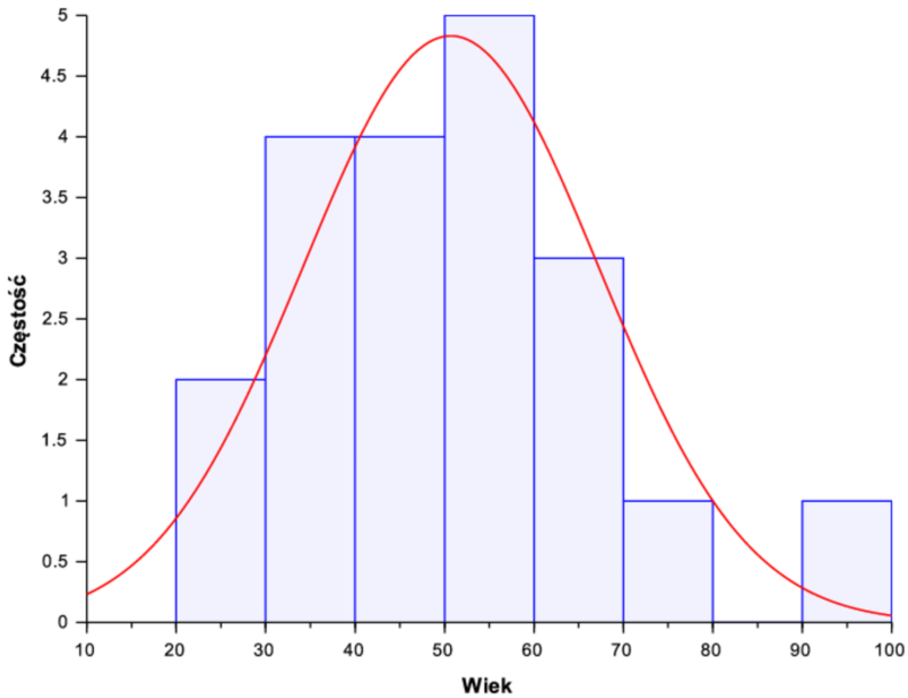
Rys. 19: Okno 'Tabele i wykresy' z wpisanymi danymi i dokonanym wyborem tabeli i wykresów

Wiek uczestników testu			
Klasy	Częstość	Częst. wzgl.	Częst. wzgl. w %
20-30	2	0.1	10
30-40	4	0.2	20
40-50	4	0.2	20
50-60	5	0.25	25
60-70	3	0.15	15
70-80	1	0.05	5
80-90	0	0	0
90-100	1	0.05	5
Suma:	20	1	100

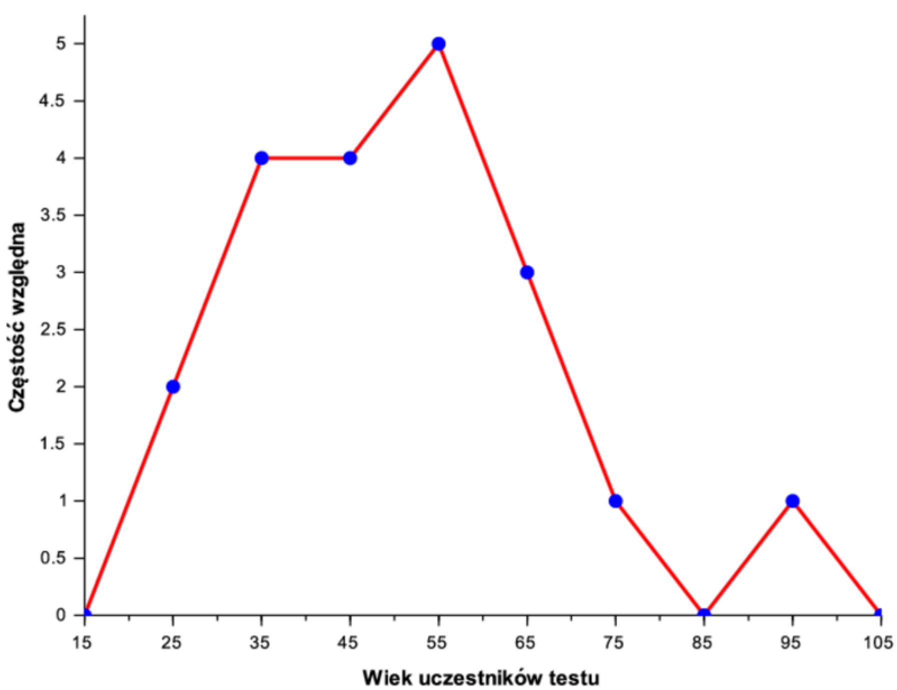
Rys. 20: Tabela częstości



Rys. 21: Wykres kołowy częstości



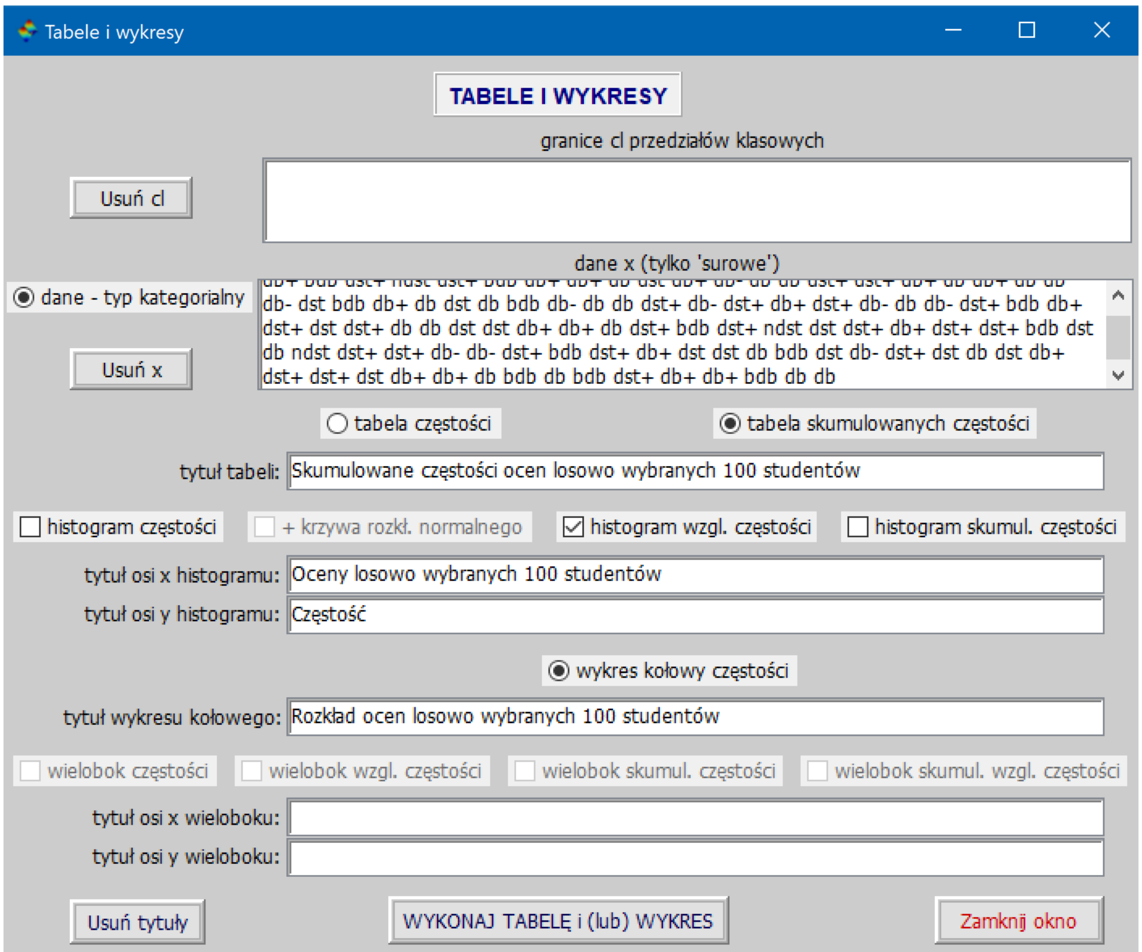
Rys. 22: Histogram częstości



Rys. 23: Okno 'Wielobok częstości'

Polecenie: ’Tabele i wykresy’ – kontynuacja dla danych typu kategorialnego

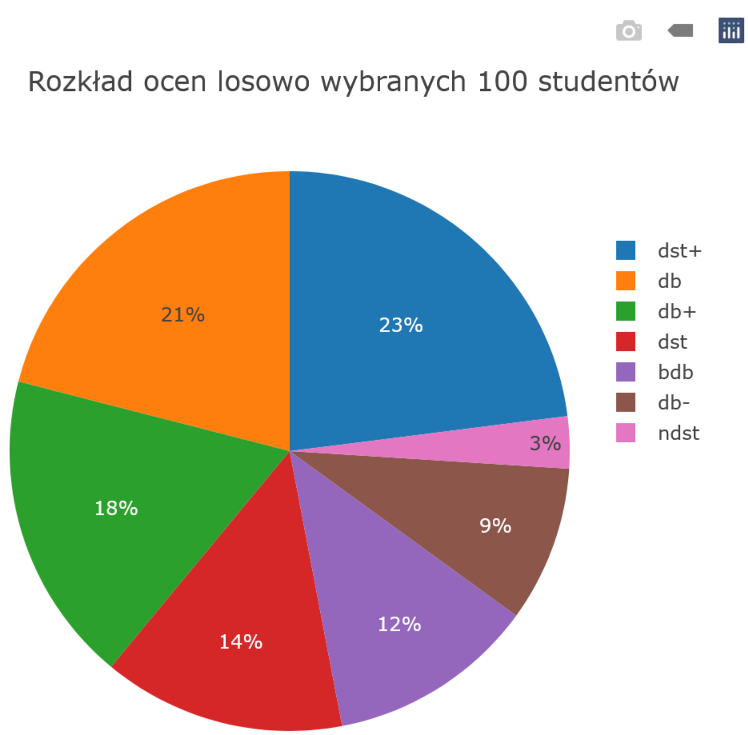
1. Jeśli mamy dane kategorialne, to wpisujemy je w okienko ’dane x’, przy czym nie musimy elementów wektora danych brać w cudzysłów lub w apostrofy. Ale musimy zaznaczyć opcję ’dane - typ kategorialny’. Okienko ’granice cl przedziałów klasowych’ zostawiamy puste.
2. Możemy wybrać odpowiednią tabelę, wykres kołowy i jeden z histogramów. Jednak zamiast histogramu otrzymamy wykres słupkowy, ponieważ ze względów technicznych (brak miejsca) nie daliśmy osobnego wyboru o nazwie ’wykres słupkowy’.
3. Po wpisaniu wektora ocen stu losowo wybranych studentów otrzymamy okno takie jak na Rys. 24. Natomiast tabela, wykres kołowy i wykres słupkowy wyglądają tak jak na Rys. 25–27.



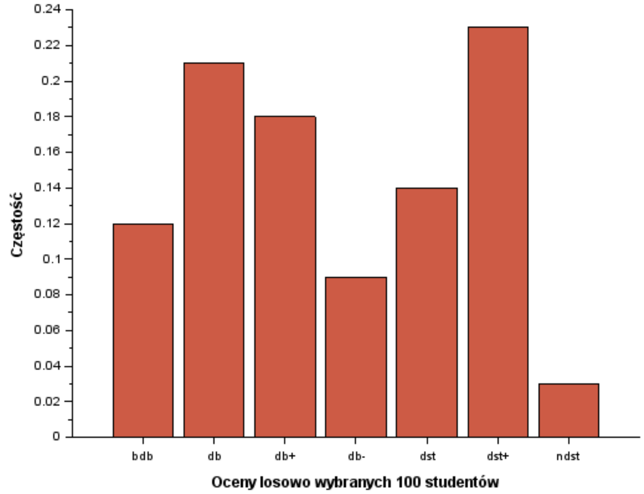
Rys. 24: Okno ’Tabele i wykresy’ z wpisanymi danymi i dokonanym wyborem tabeli i wykresów

Skumulowane częstości ocen losowo wybranych 100 studentów					
Klasy	Częstość	Skum. częstość	Częst. wzgl.	Skum. Częst. wzgl.	Częst. wzgl. w %
db	21	21	0.21	0.21	21
bdb	12	33	0.12	0.33	12
db+	18	51	0.18	0.51	18
db-	9	60	0.09	0.6	9
dst	14	74	0.14	0.74	14
dst+	23	97	0.23	0.97	23
ndst	3	100	0.03	1	3
Suma:	100		1		100

Rys. 25: Tabela częstości



Rys. 26: Wykres kołowy częstości



Rys. 27: Wykres słupkowy częstości ocen 100 losowo wybranych studentów

Polecenie: 'Tworzenie tablicy kontyngencji'

Po kliknięciu w polecenie 'Tworzenie tablicy kontyngencji' otwiera nam się okno do tworzenia tablicy kontyngencji. Tablicę taką tworzy się, korzystając z danych 'surowych' zapisanych w pliku z rozszerzeniem .csv. W istocie, dokonuje się tam obliczanie liczebności elementów należących do klas wyznaczonych przez wartości dwóch (lub więcej kolumn). Na przykład, mając dwie kolumny (zapiszmy je w wierszach aby wykorzystać całą szerokość strony),

A	A	B	A	C	C	B	B	C	A	C	A	B	C	A	B	B	C
D	F	F	E	D	E	D	F	F	E	D	F	E	F	E	E	D	E

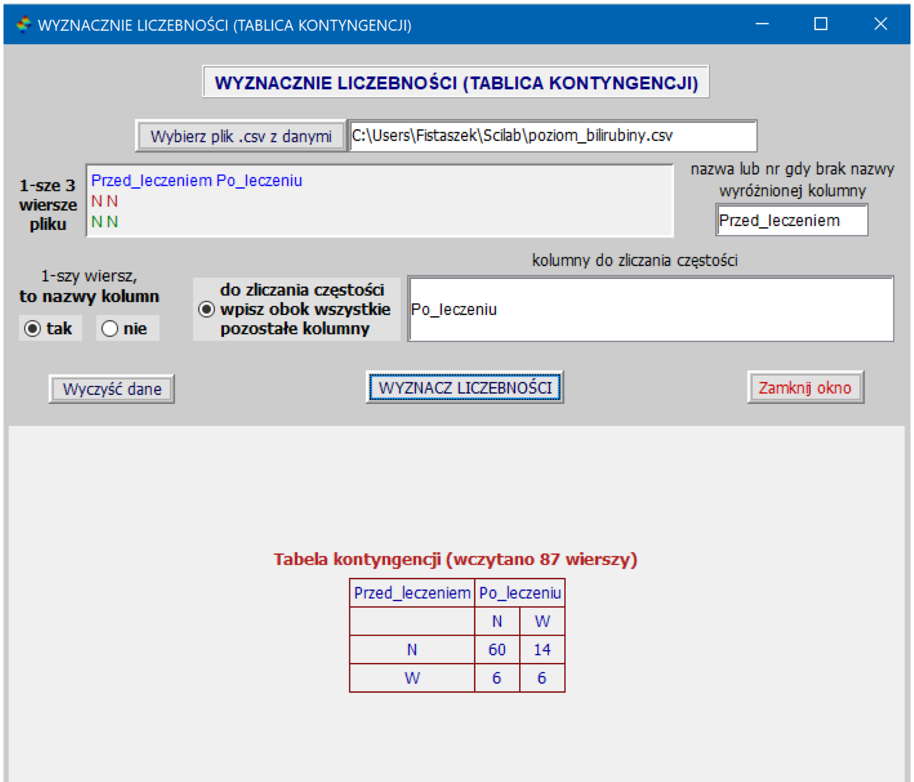
możemy policzyć ile mamy par postaci {A,D}, {A,E}, {A,F}, {B,D}, {B,E}, {B,F}, {C,D}, {C,E}, {C,F}. Przy czym literami A, B i C będą oznaczone wiersze tablicy, w której będą wypisane liczebności takich par, a literami C, D i F będą oznaczone kolumny tej tablicy. Aby otrzymać taką tablicę liczebności, należy wybrać plik z zapisanymi danymi, klikając w przycisk 'Wybierz plik .csv z danymi', następnie w polu 'nazwa lub nr gdy brak nazwy wyróżnionej kolumny' wpisujemy nazwę lub numer wyróżnionej kolumny - elementy tej kolumny będą tworzyć nazwy wierszy tablicy liczebności, a następnie w polu 'kolumny do zliczania częstości' wpisać nazwę lub numer kolumny, której elementy wraz z elementami wyróżnionej kolumny będą tworzyć pary i dla których będziemy obliczać częstość występowania takich par w danych.
Weźmy jako przykład plik `poziom_bilirubiny.csv`, gdzie jako dane mamy poziom bilirubiny przed i po leczeniu (N - normalny, W - wysoki) dla 86 pacjentów. Leczona była pewna choroba eksperymentalnym lekiem. Potrzebna nam jest tablica kontyngencji rozmiaru 2x2 aby przeprowadzić test McNemary i sprawdzić czy skutkiem ubocznym stosowania tego leku jest podniesienie poziomu bilirubiny. Kolumny danych mają nazwy `Przed_leczeniem` i `Po_leczeniu`. Najpierw otwieramy plik `poziom_bilirubiny.csv` w sposób opisany wyżej, a po jego otwarciu wpisujemy nazwę `Przed_leczeniem` w polu 'nazwa lub nr gdy brak nazwy wyróżnionej kolumny', zaznaczamy radiowy przycisk 'tak' pod napisem '1-szy wiersz, to nazwy kolumn' i klikamy w przycisk radiowy 'do zliczania częstości wpisz obok wszystkie pozostałe kolumny'. Automatycznie zostanie wpisana nazwa kolumny `Po_leczeniu`. Następnie klikamy w przycisk 'WYZNACZ LICZEBNOŚCI' i otrzymamy tabelę częstości (kontyngencji) jak na Rys. 28.
Jako drugi przykład weźmy sobie plik `heart_or.csv` o 14 kolumnach, którego 4 pierwsze wiersze maja postać:

Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
63	1	typical	145	233	1	2	150	0	2.3	3	0	fixed	No
67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3	normal	Yes
67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2	reversable	Yes

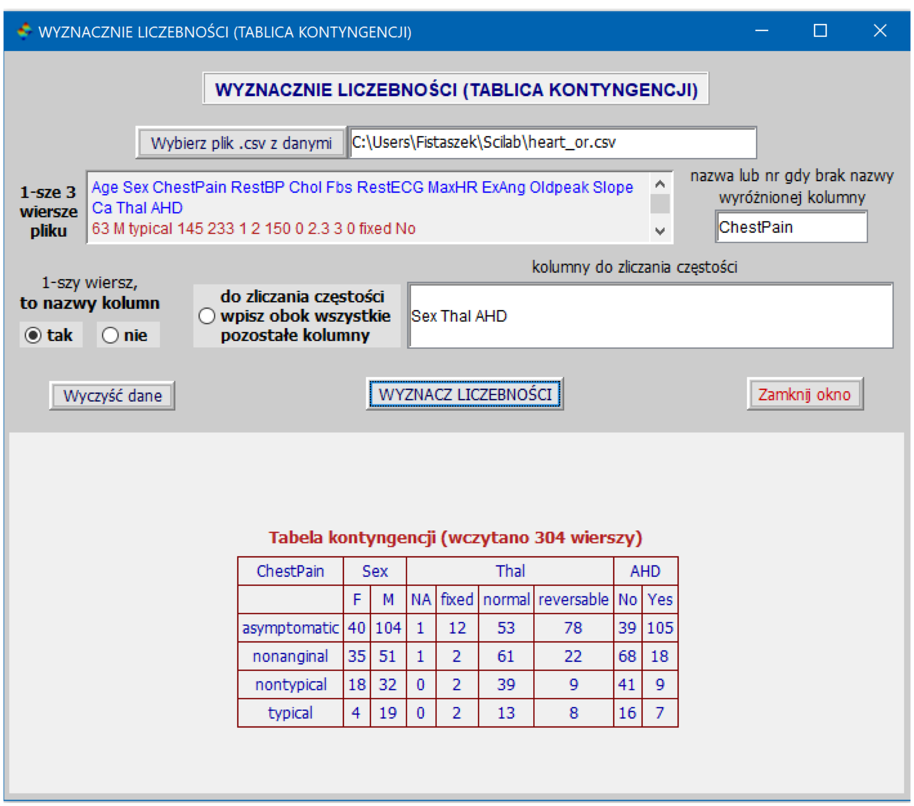
Otwieramy ten plik jak poprzednio a jako kolumnę wyróżnioną bierzemy kolumnę o nazwie 'ChestPain'. Jako kolumny do zliczania wybierzemy kilka kolumn, czyli utworzymy tablicę zawierającą kilka tablic kontyngencji. Nie będziemy wybierać kolumn, które mają, jak nie wszystkie, to prawie wszystkie elementy różne, bo wtedy z góry wiadomo, że wszystkie (lub prawie wszystkie) będą miały liczebność 1. Wybierzmy zatem jako kolumny do zliczania kolumny 'Sex', 'Thal' i 'AHD'. Skopiujemy te nazwy z pola '1-sze 3 wiersze pliku' zaznaczając je po kolei i wkleimy do pola 'kolumny do zliczania częstości' (nie klikamy w radiowy przycisk 'do obliczania częstości wpisz obok wszystkie pozostałe kolumny', ponieważ musielibyśmy usuwać z niego zbyt dużo nazw kolumn). Następnie klikamy w 'WYZNACZ LICZEBNOŚCI' i otrzymujemy tabelę jak na Rys. 29.

Uwaga

1. Zwróćmy uwagę na pole '1-sze 3 wiersze pliku' (Rys. 29), na którym część nazw w 1-szym wierszu przeszła do drugiej linii i aby uwidocznić cały ten wiersz trzeba było użyć suwaka po prawej stronie.
2. Korzystając z polecenia 'Tworzenie tablicy kontyngencji' możemy też policzyć liczebności różnych elementów w jednej kolumnie wybierając jako kolumnę wyróżnioną tę kolumnę i wpisując tę samą kolumnę w polu 'kolumny do zliczania częstości'.
3. Jeśli liczba kolumn w tablicy liczebności jest duża i tablica ta jest szersza niż szerokość okna, to wynik zapisuje się w formacie 'html' i pokazuje się w domyślnej przeglądarce internetowej.



Rys. 28: Wyniki zliczania par różnych elementów w kolumnach Przed_leczeniem i Po_leczeniu



Rys. 29: Wyniki zliczania par różnych elementów w kolumnach ChestPain i Sex, ChestPain i Thal oraz ChestPain i AHD

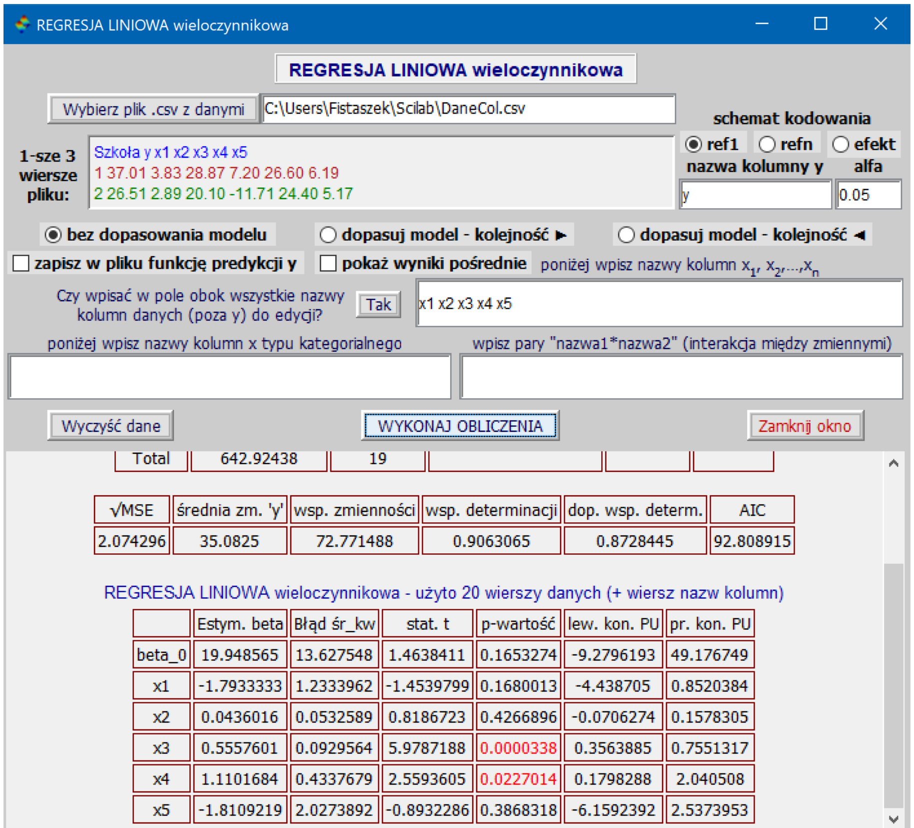
Polecenie: ’Wieloczynnikowa regresja liniowa’

Okno wywołane tym poleceniem różni się od okna polecenia ’Regresja logistyczna’ tylko tym, że występują dodatkowo 3 przyciski radiowe oznaczone tekstami: ’bez dopasowania modelu’, ’dopasuj model - kolejność ▶’ i ’dopasuj model - kolejność ◀’ oraz brak jest ’kratki’ oznaczonej ’kolejność: najpierw 1’. Brak tej ostatniej wynika z tego, że zmienna y (pełni tę samą rolę co y w regresji logistycznej) musi być zmienną o wartościach liczbowych i nie może być zmienną typu kategorialnego. Nie wymaga zatem kodowania. Wszystko pozostałe, co powiedzieliśmy o poleceniu ’Regresja logistyczna’, ma zastosowanie do polecenia ’Wieloczynnikowa regresja liniowa’. Dlatego nie będziemy tutaj o tym mówić a kierujemy czytelnika do omówienia polecenia ’Regresja logistyczna’ na str. 17. Omówimy tutaj (dość pobieżnie) tylko 2 przyciski radiowe ’dopasuj model - kolejność ▶’ i ’dopasuj model - kolejność ◀’, ponieważ wybierając przycisk ’bez dopasowania modelu’ otrzymamy oszacowanie parametrów regresji zgodnie z modelem regresji określonym przez nas przez wpisanie w odpowiednie pola odpowiednich nazw.
Wybierając zatem opcję ’dopasuj model - kolejność ▶’, wybieramy metodę ’wprzód’, zwaną po angielsku metodą *forward*, która wybór modelu zaczyna od najprostszego modelu $y = \beta_0$ a następnie dodaje jedną zmienną, nazwijmy ją x_i, wybraną ze zbioru $x_1, \dots, x_n$ tak, że dla niej model $y = \beta_0 + \beta_1 x_i$ daje najmniejszą p-wartość w teście istotności β_1 i wartość bezwzględna statystyki testowej $|t_1|$ dla parametru β_1 spełnia nierówność $|t_1| \geq 1$. Po dodaniu do modelu jednej zmiennej proces ten jest kontynuowany i wśród pozostałych zmiennych nie włączonych do modelu szukana jest następna zmienna o opisanych własnościach. Proces dopasowywania modelu kończy się, gdy zostały wyczerpane już wszystkie zmienne lub dla żadnej zmiennej x_k z pozostałych zmiennych nie spełniony jest warunek: $|t_j| \geq 1$, gdzie t_j jest wartością statystyki testowej w teście istotności współczynnika beta znajdującego się przy zmiennej x_k. Podane warunki nie są wszystkimi warunkami, które uwzględnia się przy dopasowywaniu modelu. Np. przestrzega się zasady nazywanej po angielsku “principle of marginality”: np., jeśli włączony został do modelu ’istotny statystycznie’ człon ’zmienna1*zmienna2’ (interakcja), to do modelu muszą być włączone też obie zmienne zmienna1 i zmienna2, nawet gdy same te zmienne nie są ’istotne statystycznie’ (tj., t test istotności daje dla nich statystykę testową o wartości bezwzględnej mniejszej od 1). Albo gdy mamy zmienną typu kategorialnego, dla której liczba różnych wartości jest większa od 2 i jedna z wprowadzonych (*dummy*) zmiennych z dodaną cyfrą w nawiasach jest istotna, to pozostałe takie zmienne różniące się tylko dodaną cyfrą w nawiasach też są włączane do modelu niezależnie od tego czy współczynniki beta przy nich są istotne statystycznie czy nie.
Jeśli wybierzemy opcję ’dopasuj model - kolejność ◀’, to metoda dopasowywania nazywa się metodą ’wstecz’ (po angielsku metodą *backward*). Różni się ona od metody *forward* tym, że zaczynamy od modelu pełnego (czyli takiego, jak przy wyborze opcji ’bez dopasowania modelu’), a następnie po kolei usuwa się po jednej zmiennej, np. usuwamy zmienną x_i, dla której t test istotności stojącego przy niej współczynnika beta daje największą p-wartość. Ale musi być spełniony warunek, że dla tej zmiennej w teście t istotności jej współczynnika beta, wartość bezwzględna statystyki testowej jest większa lub równa 1. Proces ten powtarza się tak długo dopóki jest co usuwać przy spełnieniu wymienionego warunku.

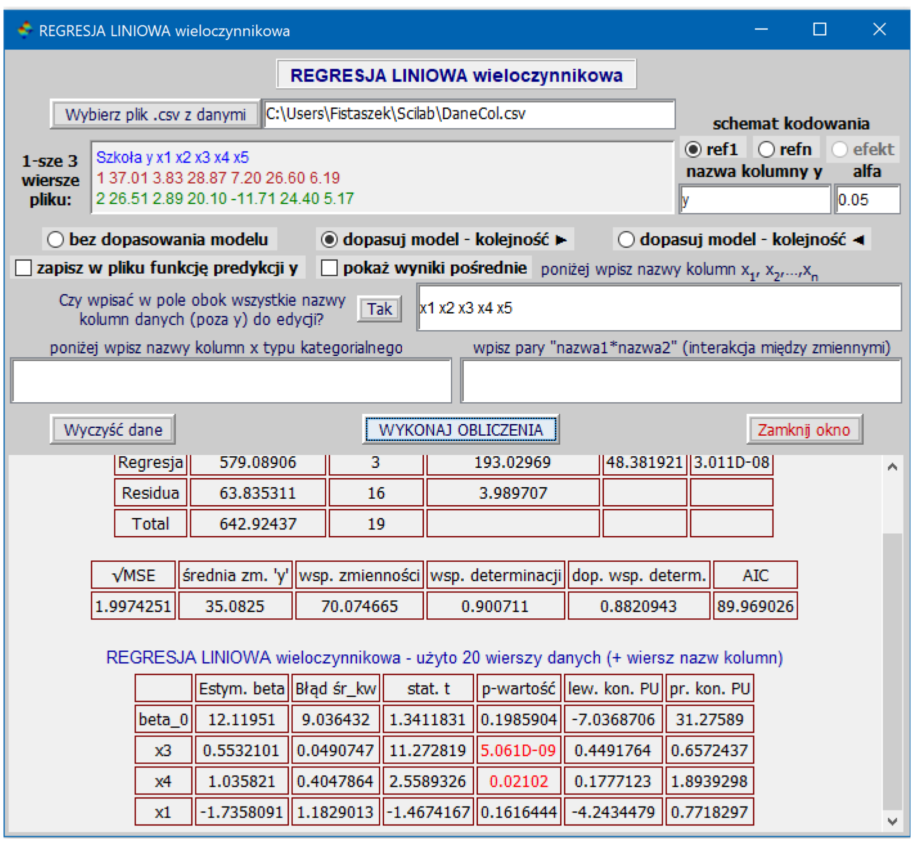
Uwaga

Wielu statystyków przestrzega przed korzystaniem z wyżej opisanych metod i innych tego typu ponieważ często nie prowadzą one do modelu optymalnego przy zadanym kryterium optymalności. Aby otrzymać model optymalny, należałoby wykonać regresję dla wszystkich możliwych modeli. Jeśli ograniczymy się do modeli bez interakcji, to mając p zmiennych niezależnych (tylko zmienna y jest zależna), mamy 2^p różnych modeli. Dla większych p, np. p=30, mamy 1073741824 modeli i tyle trzeba byłoby zbadać, co praktycznie czyni zadanie niewykonalnym w rozsądnym czasie. Jeśli już są polecane jakieś metody dopasowywania modeli, to wśród nich wymienia się bardziej skomplikowane metody, takie jak metodę ’lasso’ lub metody oparte na podejściu bayesowskim.

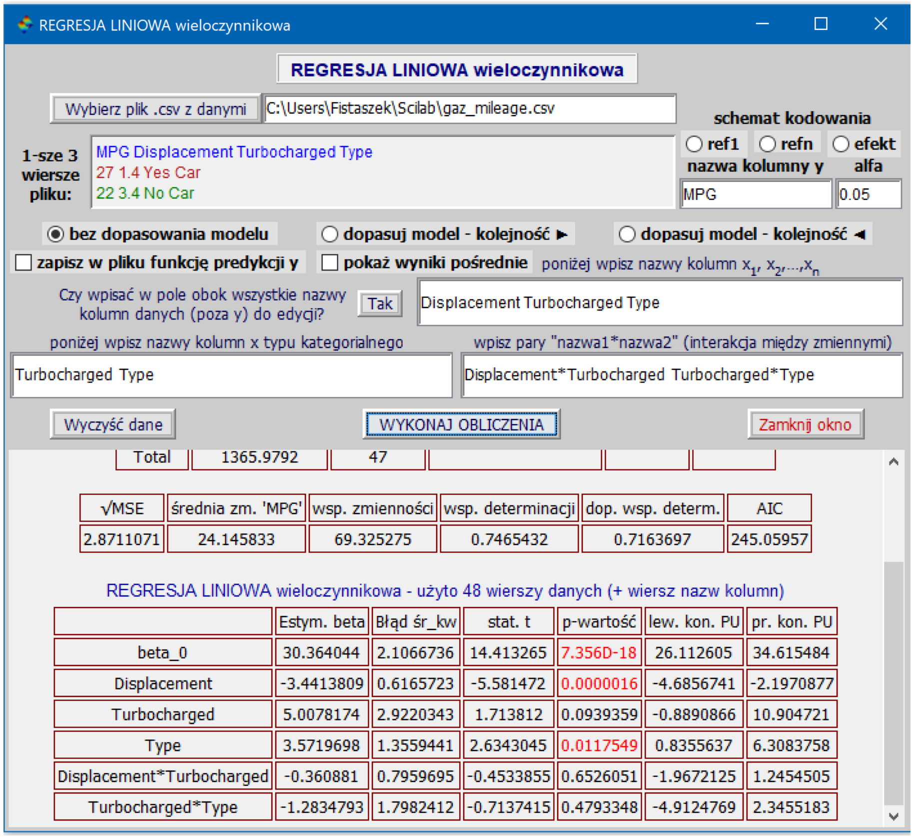
Jako przykład użycia okna ’Wieloczynnikowa regresja liniowa’ weźmy dane z pliku `DaneCol.csv`. Po wybraniu modelu przez wpisanie nazw zmiennych w pola ’nazwa kolumny y’ oraz ’poniżej wpisz nazwy kolumn $x_1, x_2, \dots, x_n$’ i wybraniu odpowiednich opcji otrzymamy wyniki, jak na Rys. 30. Natomiast gdy zamiast opcji ’bez dopasowania modelu’ zaznaczymy opcję ’dopasuj model - kolejność ▶’, to otrzymamy wyniki, jak na Rys. 40. Zwróćmy uwagę na to, że zmienna ’Szkoła’ pełni tutaj rolę pomocniczą, porządkową i ją pomijamy.
Natomiast wybierając dane zawarte w pliku `gaz_mileage.csv` i, jak poprzednio, raz nie wybierając modelu, a drugi raz wybierając model metodą ’wstecz’, otrzymamy wyniki, odpowiednio jak na Rys. 32 i Rys. 33.



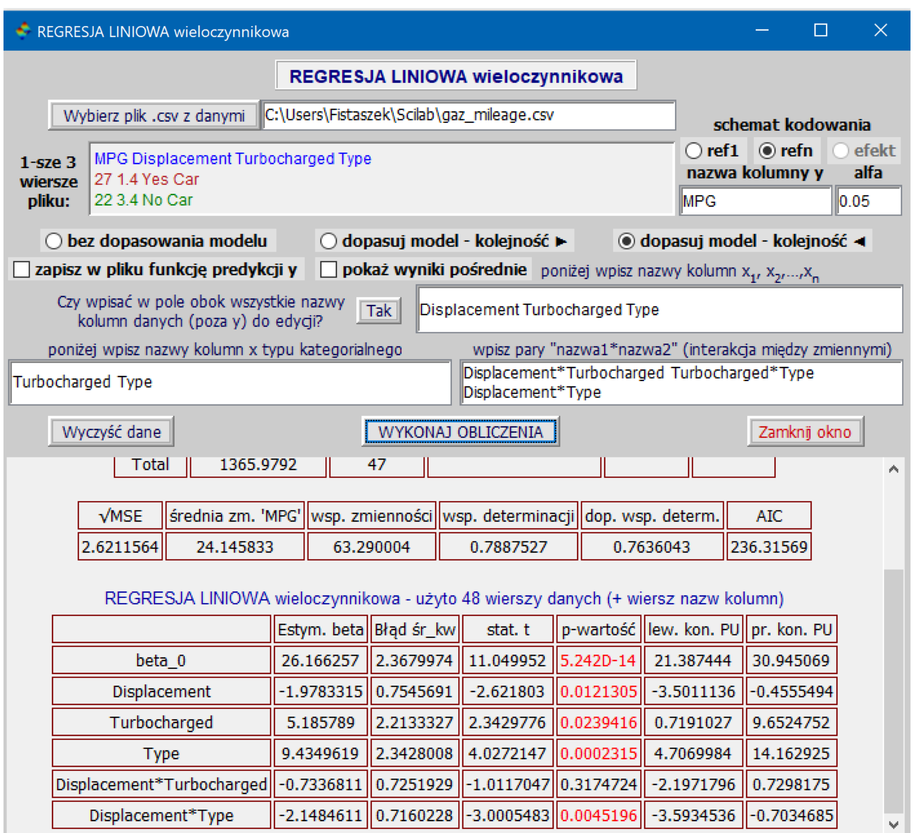
Rys. 30: Wyniki obliczeń dla wieloczynnikowej regresji liniowej dla danych w pliku DaneCol.csv



Rys. 31: Wyniki obliczeń dla regresji logistycznej dla danych w pliku DaneCol.csv z opcją 'dopasuj model: kolejność ►'



Rys. 32: Wyniki obliczeń dla wieloczynnikowej regresji liniowej dla danych w pliku gaz_mileage.csv



Rys. 33: Wyniki obliczeń dla regresji logistycznej dla danych w pliku gaz_mileage.csv z opcją 'dopasuj model: kolejność ◄'

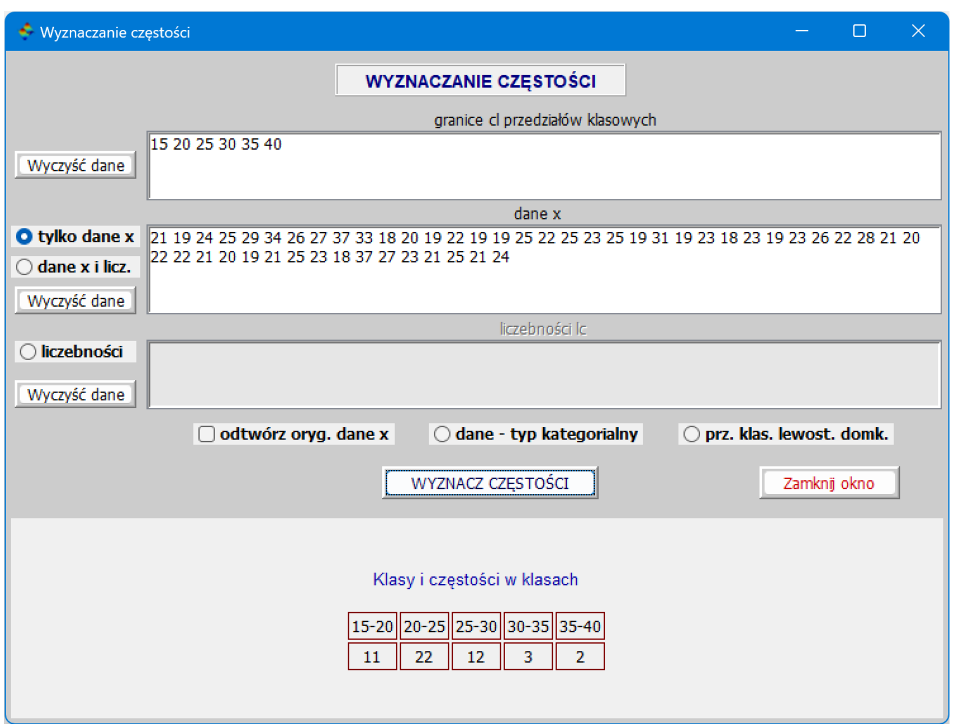
Polecenie: ’Wyznaczanie częstości w klasach’

Z menu ’Basicstat’ wybieramy ’Wyznaczanie częstości w klasach’. w otwartym w ten sposób oknie możemy wykonać wiele różnych obliczeń wyznaczających liczebności zarówno klas jak i liczebności specjalnie wybranych podgrup zbioru danych. Możemy też na podstawie wektora danych x oraz wektora liczebności poszczególnych danych podanych jako współrzędne wektora liczebności lc, odtworzyć wektor oryginalny, tj. otrzymać wektor, w którym każda współrzędna wektora x jest powtórzona tyle razy ile wynosi jej liczebność. Ze względu na możliwość wykonywania różnych zadań, przed wykonaniem poszczególnego zadania należy zacząć od wyboru odpowiednich przycisków dla tego zadania. Spowoduje to, automatyczną deaktywacją pewnych opcji (przycisków) nieodpowiednich dla wybranego zadania oraz aktywację odpowiednich opcji dla tego zadania. Zacznijmy od podstawowych zadań, które możemy wykonać w tym oknie.

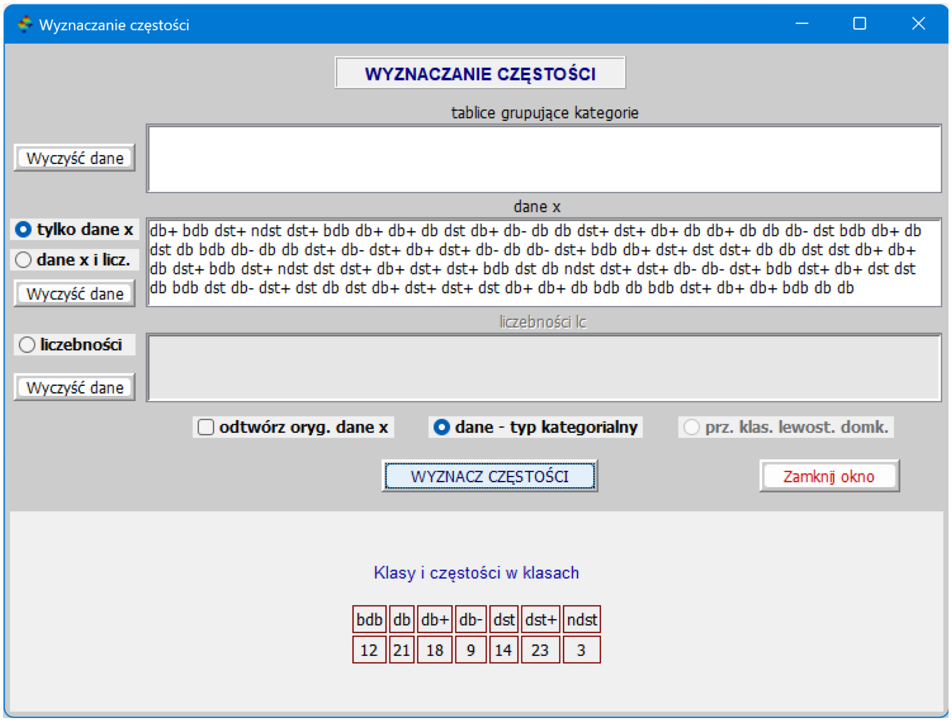
1. Najprostszym zadaniem do wykonania jest wyznaczenie liczebności klas po podaniu granic przedziałów klasowych i (surowych) danych x w postaci wektora liczbowego. Po wybraniu przycisku radiowego ’dane x’ i przycisku ’prz. kl. lewost. domkn.’, jeśli chcemy aby przedziały klasowe były lewostronnie domknięte, wpisujemy granice przedziałów klasowych w pole ’granice cl przedziałów klasowych’, a dane x w pole oznaczone ’dane x (tylko surowe)’. Następnie klikamy w ’WYZNACZ CZĘSTOŚCI’ i otrzymujemy wynik jak na Rys. 34.

2. Gdy mamy do czynienia z danymi typu kategorialnego, to wybieramy przycisk ’dane - typ kategorialny’, wpisujemy je w okienko oznaczone ’dane x (tylko surowe)’ (nie musimy brać w cudzysłów elementów wektora danych typu kategorialnego) i klikamy w ’WYZNACZ CZĘSTOŚCI’, aby otrzymać wynik jak na Rys. 35.

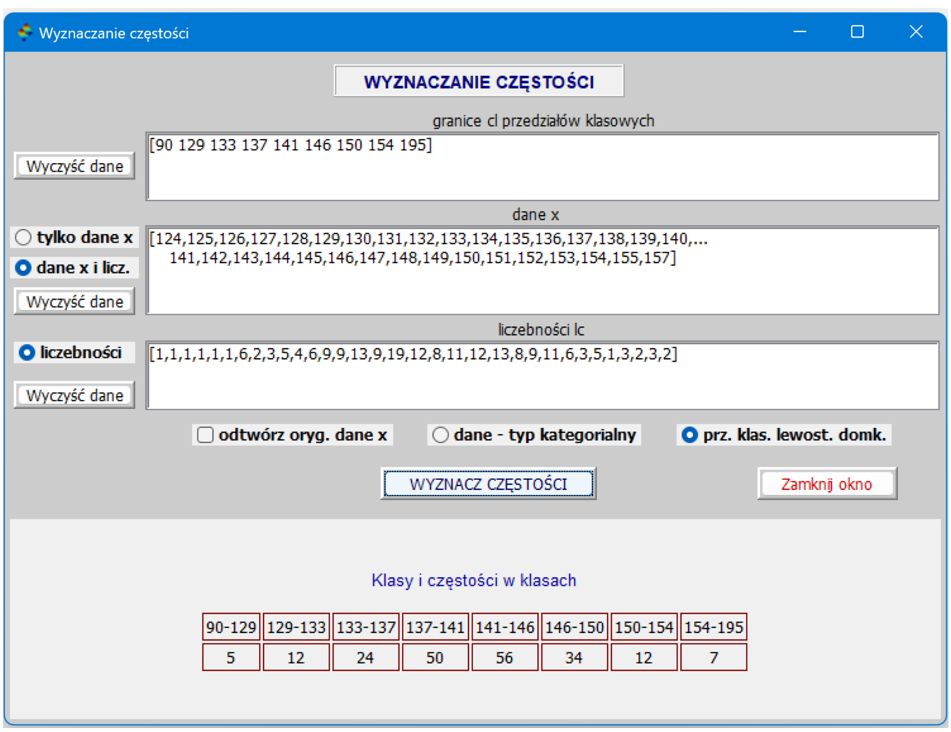
3. Kolejnym, bardziej skomplikowanym zadaniem, jakie można wykonać w oknie ’WYZNACZANIE CZĘSTOŚCI’, jest wyznaczenie częstości, gdy mamy dane (liczbowe) x, a oprócz nich mamy wektor, którego współrzędne są liczebnościami poszczególnych elementów danych x. Aby zliczyć częstości takich danych w zadanych klasach , zaznaczamy przycisk ’liczebności’ i wpisujemy w pole ’granice cl przedziałów klasowych’ granice przedziałów klasowych, w pole ’wartości w, dla których niżej podane są liczebności’ wpisuje się dane x, a w pole ’liczebności lc’ wpisujemy liczebności i klikamy w przycisk ’WYZNACZ CZĘSTOŚCI’. Dla przykładowych danych otrzymamy wynik jak na Rys. 36.



Rys. 34: Okno ’Wyznaczanie częstości w klasach’ z wpisanymi danymi ilościowymi i wyznaczonymi częstościami



Rys. 35: Okno ’Wyznaczanie częstości w klasach’ z danymi typu kategorialnego i wyznaczonymi częstościami



Rys. 36: Okno ’Wyznaczanie częstości w klasach’ z danymi typu kategorialnego i wyznaczonymi częstościami

Polecenie: ’Wyznaczanie częstości w klasach’ (kontynuacja)

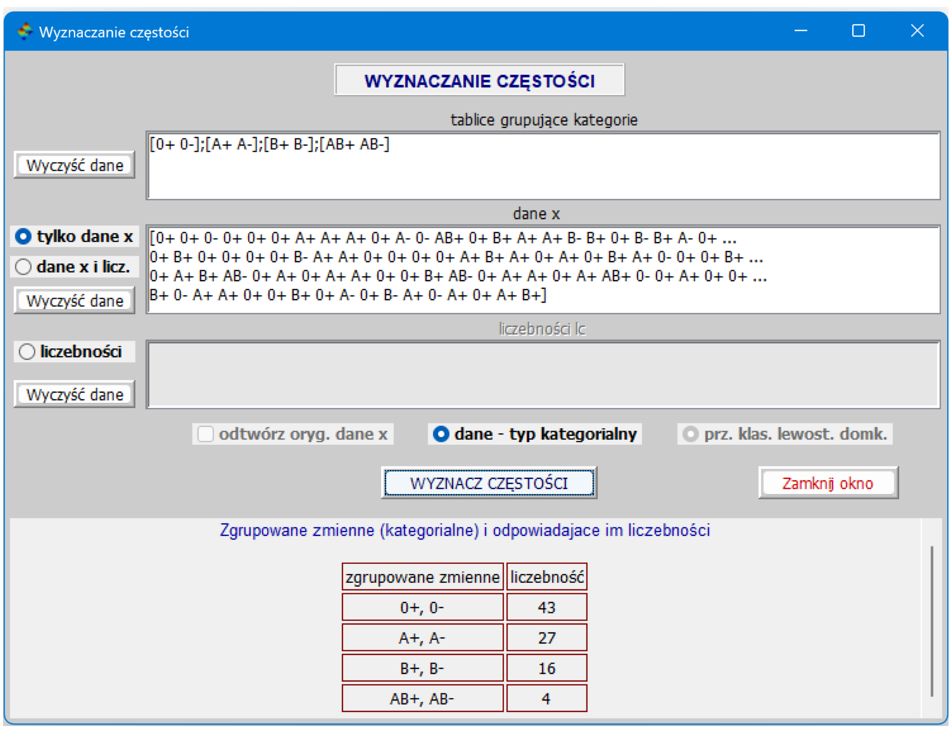
Jeśli nasze dane x traktujemy jako dane typu kategorialnego - wybieramy wtedy przycisk ’dane - typ kategorialny - ,to możemy również obliczać liczebności, wpisując w polu ’tablice grupujące kategorie’ (wcześniej nazywało się to pole ’granice przedziałów klasowych cl’) w nawiasach kwadratowych grupę wybranych przez nas nazw zmiennych występujących w polu ’dane x’. Nazwy wybranych zmiennych oddzielamy spacjami (można też przecinkami, ale jak jest ich dużo i po przecinkach nie ma spacji, to wiersze nie będą zawijane). Takich grup może być więcej i każdą z nich bierzemy w nawiasy kwadratowe i grupy te oddzielamy średnikami.

4. Na przykład, mamy dane grupy krwi pewnej liczby osób. Interesują nas liczebności osób, które mają poszczególne grupy krwi, ale nie ma dla nas znaczenia czynnik Rh. Wtedy w oknie ’WYZNACZANIE CZĘSTOŚCI’ wybieramy przyciski ’tylko dane x’ oraz ’dane - typ kategorialny’, w pola ’tablice grupujące kategorie’ i ’dane x’ wpisujemy dane jak na Rys. 37 i klikamy w ’WYZNACZ CZĘSTOŚCI’, otrzymując liczebności grup.

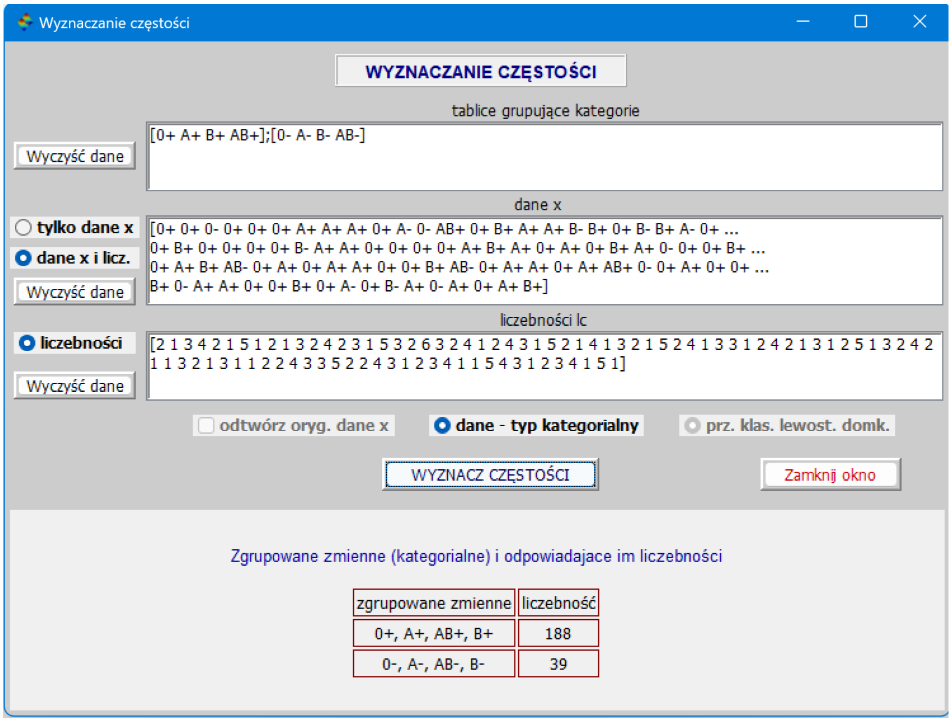
5. Jeśli mamy dane jak wyżej oraz dodatkowo mamy liczebności odpowiadające poszczególnym danym x i interesuje nas liczba osób z czynnikiem Rh+ i z czynnikiem Rh-, to możemy to zrobić wybierając przyciski oraz wpisując dane, jak na Rys. 38.

6. Jeśli mamy dane liczbowe, to można je też potraktować jako dane kategorialne i policzyć liczebności w grupach wybranych zmiennych ze zbioru danych x.

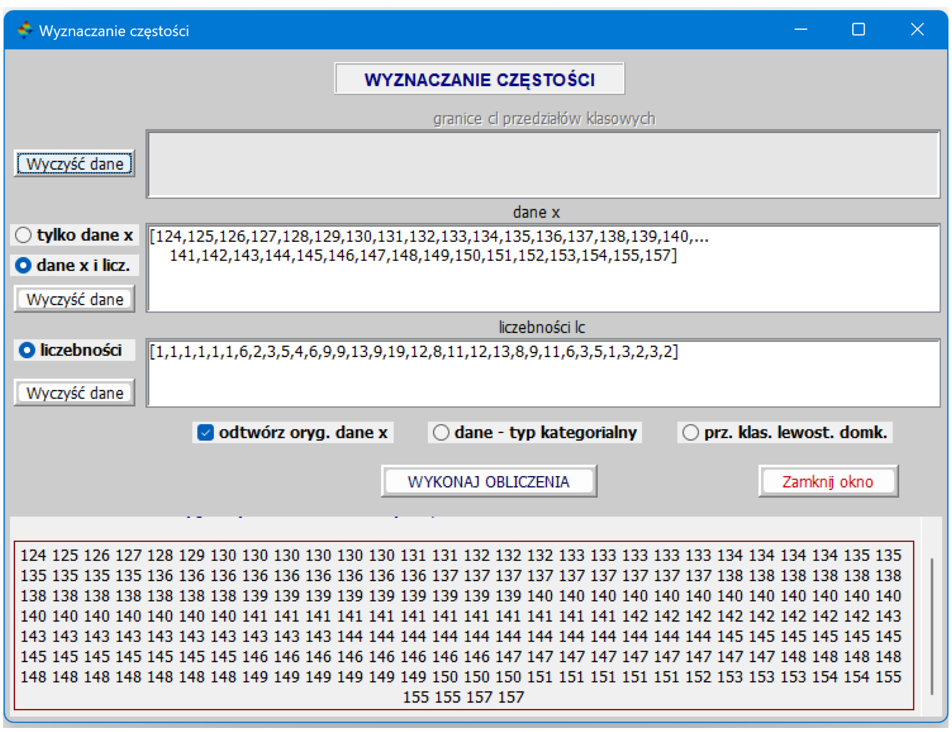
Na koniec pokażemy przykład z danymi liczbowymi, w którym odtwarza się wektor na podstawie danych x oraz ich liczebności w polu ’liczebności lc’ - zob. Rys. 39. Taki wektor może być przydatny, gdy chcemy przeprowadzić test chi kwadrat, porównując nasze dane z rozkładem ciągłym lub wykładniczy i musimy wyznaczyć parametry tych rozkładów, ponieważ ich nie znamy. Forma podanego wyniku jest łatwa do zaznaczenia i skopiowania.



Rys. 37: Okno ’Wyznaczanie częstości w klasach’ z danymi typu kategorialnego i wyznaczonymi częstościami w poszczególnych grupach



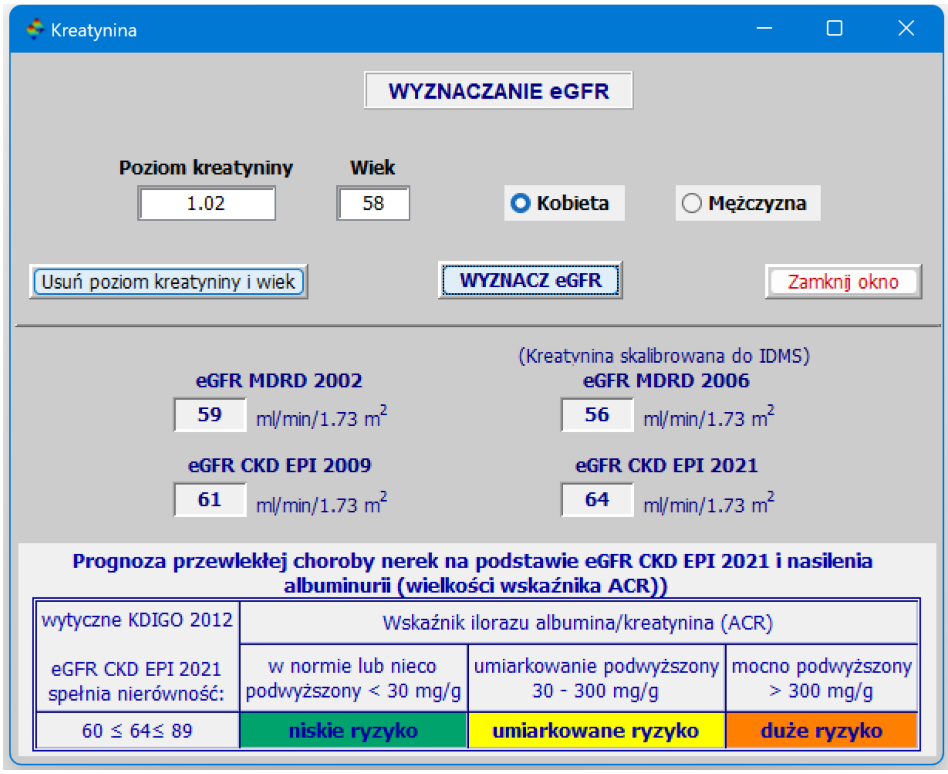
Rys. 38: Okno 'Wyznaczanie częstości w klasach' z danymi typu kategorialnego i wyznaczonymi częstościami w poszczególnych grupach



Rys. 39: Okno 'Wyznaczanie częstości w klasach' z danym wektorem x typu ilościowego i wynikiem z wektorem x rozszerzonym w ten sposób, że każda jego współrzędna została powtórzona tyle razy ile wynosi jej liczebność

Polecenie: 'Wyznaczanie wskaźnika GFR na podstawie poziomu kreatyniny oraz wieku i płci'

Aby wyznaczyć wskaźnik GFR, w pole 'Poziom kreatyniny' wpisujemy poziom kreatyniny (w mg/dl) a w pole 'Wiek' wpisujemy wiek (w pełnych latach) i wybieramy płeć klikając w odpowiedni przycisk radiowy. Otrzymamy dwie pary wyników - jedna para, ta z lewej strony, to wyniki otrzymane wg starszych procedur badawczych, ale ciągle używanych w niektórych laboratoriach, a druga para, to wyniki otrzymane wg nowszych procedur rekomendowanych przez przez międzynarodową grupę KDIGO. Na przykład, wpisując w pole 'Poziom kreatyniny' poziom kreatyniny równy 1.02 a w pole 'Wiek' liczbę 58 i wybierając płeć 'Kobieta', otrzymamy wyniki jak na Rys. 40. U dołu otrzymamy też tabelkę z oceną ryzyka przewlekłej choroby nerek. Ale ocena ta wymaga znajomości wskaźnik ilorazu albumina/kreatynina (ACR).



Rys. 40: Okno 'Wyznaczanie wskaźnika GFR na podstawie poziomu kreatyniny oraz wieku i płci'

4 Grupa "Testy parametryczne"

Polecenie: 'Anova jednoczynnikowa i dwuczynnikowa'

Uwaga:

Gdy wykonujemy test dwuczynnikowej anovy bez powtórzeń, tj. gdy liczebności wszystkich prób wziętych z populacji wyznaczonych przez każdą możliwą parę poziomów jednego i drugiego czynnika są równe 1, to takiego testu nie da się przeprowadzić dla modelu $\boldsymbol{A} \sim \boldsymbol{B} * \boldsymbol{C}$, ponieważ wtedy nie można odróżnić efektu interakcji od zmienności resztowej. Natomiast można taki test przeprowadzić dla modelu addytywnego, tj. dla modelu $\boldsymbol{A} \sim \boldsymbol{B} + \boldsymbol{C}$.

Zanim przejdziemy do omówienia polecenia 'Anova jednoczynnikowa i dwuczynnikowa' opiszemy najpierw sposób przygotowania danych i objaśnimy wybór modelu anovy. Zrobimy to na przykładzie dwuczynnikowej anovy. Niech zmienna R zależy od pierwszego czynnika o poziomach A i B i drugiego czynnika o poziomach D1, D2 i D3. Poniżej, znajduje się Tabela 1 z danymi, czyli wartościami zmiennej R, a pod tabelą objaśnienia argumentów q i S.

Zawartość pliku:

```
R, T, D
17, A, D1
21, A, D1
49, A, D1
54, A, D1
64, A, D2
48, A, D2
34, A, D2
63, A, D2
62, A, D3
72, A, D3
61, A, D3
91, A, D3
33, B, D1
37, B, D1
40, B, D1
16, B, D1
41, B, D2
64, B, D2
34, B, D2
64, B, D2
56, B, D3
62, B, D3
57, B, D3
72, B, D3
```

Tabela 1: Tabela z danymi dla dwuczynnikowej anovy

	Czynnik II		
Czynnik I	**D1**	**D2**	**D3**
A	17	64	62
	21	48	72
	49	34	61
	54	63	91
B	33	41	56
	37	64	62
	40	34	57
	16	64	72

Model anovy dla tego przykładu może mieć postać:

`'R~T*D'`	*Anova dwuczynnikowa z interakcją*
`'R~T+D'`	*Anova dwuczynnikowa bez interakcji*
`'R~T'`	*Anova jednoczynnikowa (ignorowany czynnik D)*
`'R~D'`	*Anova jednoczynnikowa (ignorowany czynnik T)*

Dla przeprowadzenia testu dwuczynnikowej anovy należy:
utworzyć plik z danymi o strukturze podanej obok i zapisać go w katalogu roboczym Scilaba z nazwą z rozszerzeniem .csv. **Uwaga:** Kolejność kolumn jest dowolna.
Dla przeprowadzenia testu jednoczynnikowej anovy należy:
utworzyć plik o strukturze podanej obok ale bez elementów trzeciej kolumny oraz poprzedzających je przecinków i zapisać go w katalogu roboczym Scilaba z nazwą z rozszerzeniem .csv.

1. Klikając w polecenie 'Anova jednoczynnikowa i dwuczynnikowa' z grupy 'Testy parametryczne' głównego menu, otwieramy okno anovy a następnie klikając w 'Wybierz plik .csv z danymi, wybieramy odpowiedni plik. Po wybraniu pliku automatycznie zostaną wpisane w odpowiednie pola nazwy zmiennej zależnej i jednej lub dwóch zmiennych niezależnych w zależności od tego czy dane mają 2 kolumny (anova jednoczynnikowa), czy 3 kolumny (anova dwuczynnikowa).

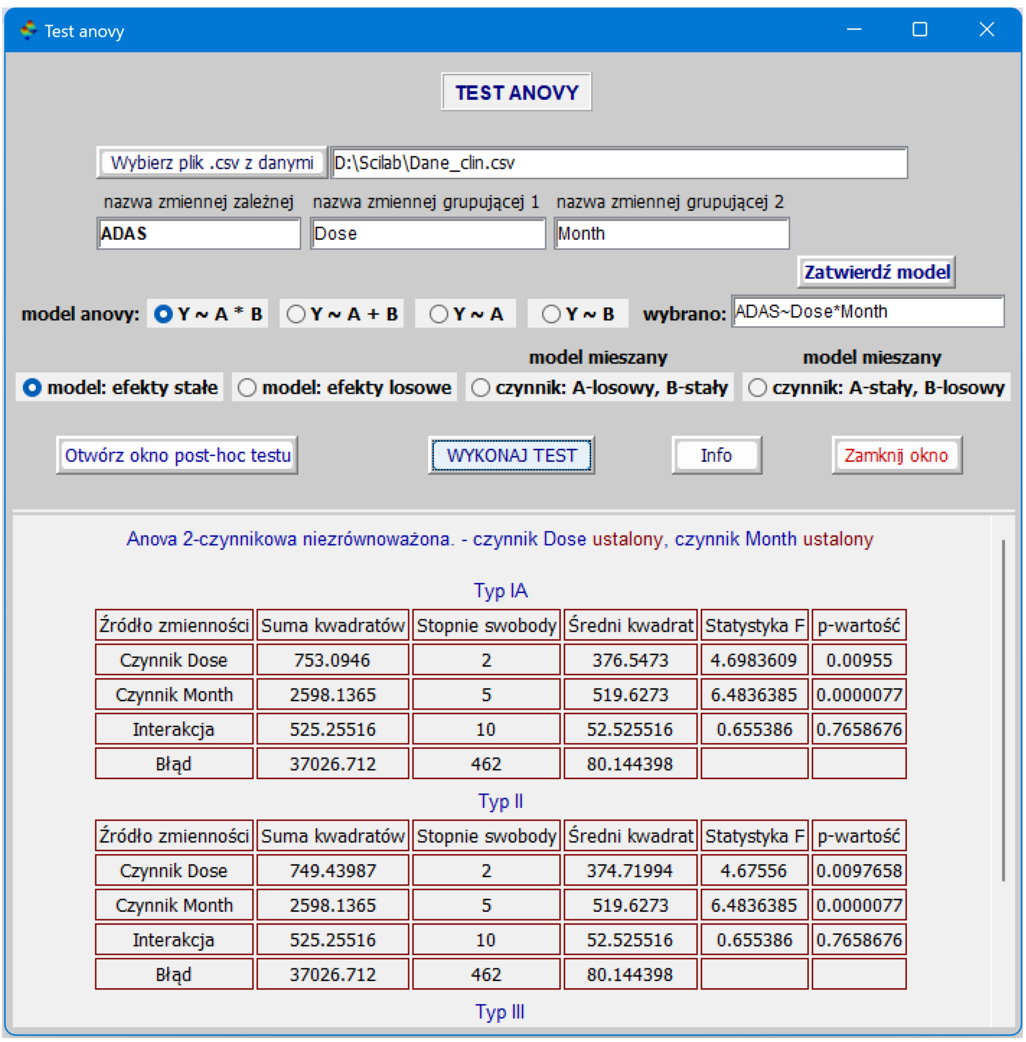
2. Zmienna numeryczna będzie wyróżniona tłustym drukiem i wpisana do pola 'nazwa zmiennej zależnej'. Niestety, gdy dla zmiennych grupujących zostały (lub została) użyte liczby (cyfry), to one też zostaną wyróżnione tłustym drukiem. Wtedy należy sprawdzić, która zmienna przyjmująca wartości liczbowe jest zmienną niezależną i jeśli automatyczny wpis jest nieprawidłowy, to należy ręcznie poprawić wpisy na prawidłowe. Najlepiej zrobić to wpisując w polu 'wybrano:' poprawny model, a następnie kliknąć w przycisk 'Zatwierdź model'. Wtedy

automatycznie zostaną wprowadzone poprawne nazwy w pola: 'nazwa zmiennej zależnej', 'nazwa zmiennej grupującej 1' i 'nazwa zmiennej grupującej 2' - właśnie z tych pól pobierane są wartości do wykonania testu anovy.

3. Następnie wybieramy odpowiedni model anovy, zaznaczając odpowiednią opcję oraz wybieramy model z odpowiednimi efektami - stałymi (oba czynniki, a dokładniej, w obu czynnikach wszystkie poziomy są ustalone), losowymi (oba czynniki są losowe, tj. wszystkie poziomy w obu czynnikach są losowe), czynnik A losowy a B ustalony (model mieszany) lub czynnik A ustalony a B losowy (model mieszany). Zwróćmy uwagę na to, że dla modelu addytywnego **y A+B** oraz dla jednoczynnikowej anovy obliczenia, a zatem i tabele anovy, są takie same dla wszystkich czterech wymienionych tu modeli z odpowiednimi efektami. A dla modelu z interakcją **y A*B**, w przypadku modeli mieszanych podawane tabele z wynikami anovy dotyczą tzw. modelu mieszanego ograniczonego (restricted). Jeśli chcemy uzyskać tabelę z wynikami anovy dla modelu mieszanego nieograniczonego (unrestricted), to trzeba zaznaczyć przycisk model: efekty losowe.

4. Otrzymamy wynik testu jak na Rys. 41. Ponieważ wybraliśmy dane, w których liczebności grup nie są równe, to otrzymaliśmy 3 tablice anovy wykonane trzema metodami (najbardziej w praktyce przydatne). Jeśli liczebności grup są jednakowe, to otrzymujemy tylko jedną tablicę anovy.

5. Jeśli wykonaliśmy już test anovy i chcemy wykonać post hoc testu, to klikamy w przycisk 'Otwórz okno post-hoc testu'. Wtedy wraz z otwarciem okna, przesłane zostaną do niego wszystkie niezbędne dane do wykonania testów post hoc.



Typ IA

Źródło zmienności	Suma kwadratów	Stopnie swobody	Średni kwadrat	Statystyka F	p-wartość
Czynnik Dose	753.0946	2	376.5473	4.6983609	0.00955
Czynnik Month	2598.1365	5	519.6273	6.4836385	0.0000077
Interakcja	525.25516	10	52.525516	0.655386	0.7658676
Błąd	37026.712	462	80.144398		

Typ II

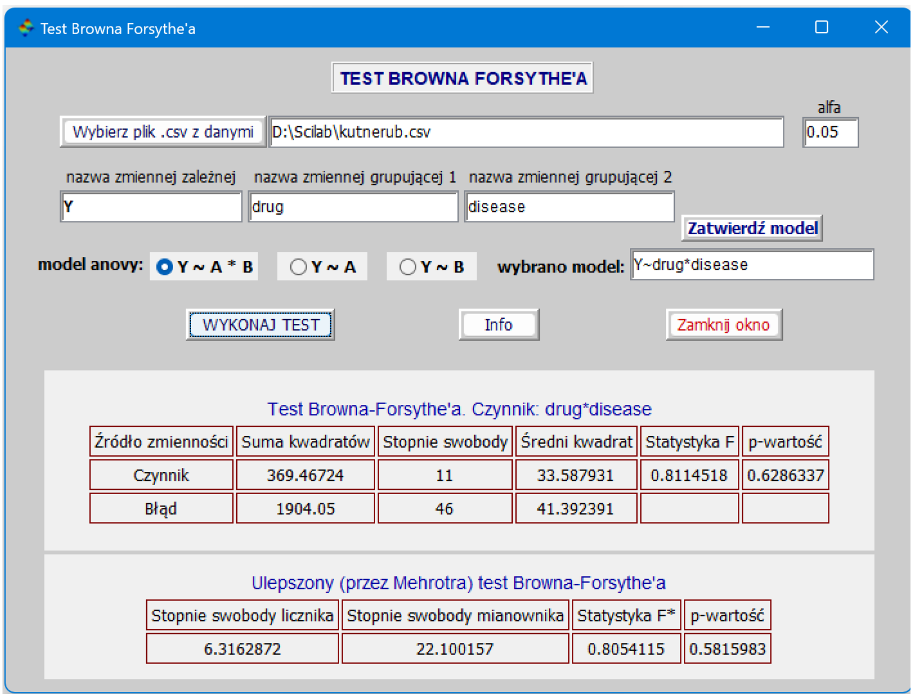
Źródło zmienności	Suma kwadratów	Stopnie swobody	Średni kwadrat	Statystyka F	p-wartość
Czynnik Dose	749.43987	2	374.71994	4.67556	0.0097658
Czynnik Month	2598.1365	5	519.6273	6.4836385	0.0000077
Interakcja	525.25516	10	52.525516	0.655386	0.7658676
Błąd	37026.712	462	80.144398		

Rys. 41: Okno 'Test Anovy' z wybranymi opcjami i wynikiem testu

Polecenie: ’Test Browna Forsythe’a równości wariancji’

1. Postępujemy podobnie jak w poprzednim opisie testu anovy. Testujemy równość wariancji w populacjach, z których pochodzą grupy, których równość średnich testujemy w teście anovy. W przypadku anovy jednoczynnikowej poziomy czynnika wyznaczają odpowiednie grupy i test równości wariancji w populacjach, które te grupy reprezentują przeprowadza się nie mając wyboru, bo jedyny udostępniony przycisk, to przycisk **Y~A**. W przypadku anovy dwuczynnikowej mamy grupy powstałe w wyniku kombinacji poziomów jednego i drugiego czynnika. Aby testować równość wariancji w populacjach, z których takie grupy pochodzą, wybieramy przycisk **Y~A*B**. Ponadto, mamy jeszcze grupy wyznaczone przez poziomy pierwszego czynnika i aby testować równość wariancji w populacjach reprezentowanych przez te grupy wybieramy przycisk **Y~A**. A do testowania równości wariancji w populacjach, z których pochodzą grupy wyznaczone przez poziomy drugiego, wybieramy przycisk **Y~B**. Dla przykładowych danych z dwoma czynnikami otrzymamy wynik testu jak na Rys. 42.

2. Jeśli w polu **wybrano model:** model jest niepoprawny, to zmieniamy go ręcznie i zatwierdzamy klikając w przycisk **Zatwierdź model**. W polu ’alfa’ możemy też zmienić poziom istotności testu na inny niż 0.05.

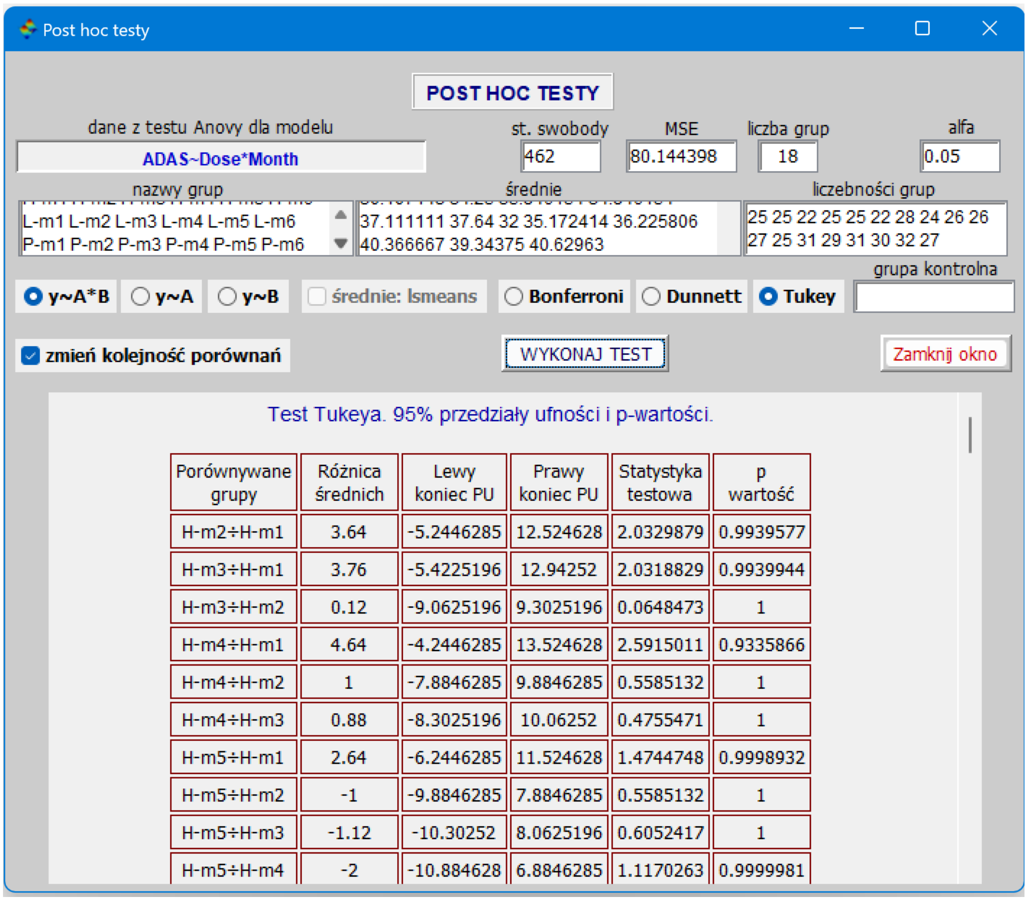


Rys. 42: Okno ’Test Browna Forsythe’a’ z wybranymi opcjami i wynikiem testu

Polecenie: ’Test post hoc: Bonferroni-Dunnett-Tukey’

1. Klikając w oknie Anovy w przycisk ’Otwórz okno post-hoc testu’, jesteśmy praktycznie gotowi do wykonania testu. Jedyne co nam pozostaje do zrobienia, to wybór jednego z 3 testów: Bonferroniego, Dannetta (dla tego testu trzeba wpisać w pole ’grupa kontrolna’ nazwę grupy kontrolnej) lub Tukeya. Trzeba też wybrać grupy, na podstawie których będziemy testować równość średnich w populacjach, na podstawie średnich w grupach, które w sposób losowy wybrano z tych populacji. Do tego służą 3 przyciski. Jeśli przeprowadzony został test anovy jednoczynnikowej, to aktywny jest tylko przycisk **Y~A**. W przypadku anovy dwuczynnikowej, gdy wykonaliśmy test anovy wybierając przycisk **Y~A*B**, to w oknie ’POST HOC TEST’ mamy do wyboru aktywne przyciski **Y~A*B**, **Y~A** i **Y~B**. Wybierając jakikolwiek z tych przycisków na podstawie wartości w polu ’nazwy grup’ i w polu ’średnie’ widzimy, w których populacjach będziemy testować równość średnich. Gdy w oknie anovy wybraliśmy model **Y~A+B**, to aktywne mamy przyciski **Y~A** i **Y~B** a gdy w teście anovy wybrano model **Y~A** lub **Y~B**, to w oknie ’POST HOC TEST’ aktywny jest tylko przycisk **Y~A**. W polu ’dane z testu Anovy dla modelu’ widzimy z jakiego modelu anovy przekazane zostały dane dla testów post hoc.

2. Dostępne jest też pole wyboru **średnie lsmeans**. Są to średnie, które pojawiły się w komputerowym programie statystycznym SAS i lansowane są przez wielu statystyków. Pole to jest aktywne tylko wtedy, gdy takie średnie są różne od “zwykłych” średnich arytmetycznych.
3. Po dokonaniu wyżej wymienionych wyborów oraz ewentualnej zmianie poziomu alfa istotności testu, klikamy w ’WYKONAJ TEST’. Otrzymujemy wynik testu (18 grup, 17 porównań) jak na Rys. 43.
4. Zwróćmy uwagę na to, że jeśli nazwy grup, średnie i liczebności nie mieszczą się w swoich polach w dwóch wierszach, to za pomocą suwaków (są widoczne z boku pola jak na rysunku po najechaniu myszką) można je przesuwać w górę lub w dół. Ponadto, jeśli chcemy, aby odejmować średnie w odwrotnym porządku, co skutkuje zmianą znaku różnicy i symetrią względem zera końców przedziałów ufności„ to zaznaczamy pole wyboru **zmień kolejność porównań**.



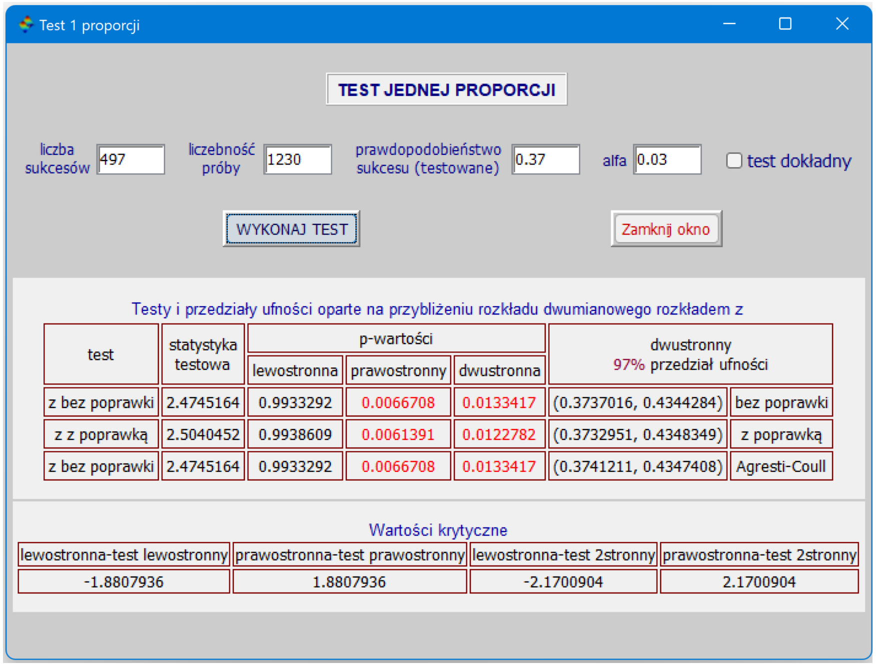
Rys. 43: Okno ’Test post hoc: Bonferroni-Dunnett-Tukey’ z wybranymi opcjami i wynikiem testu

Polecenie: ’Test proporcji (frakcji) dla 1 próby’

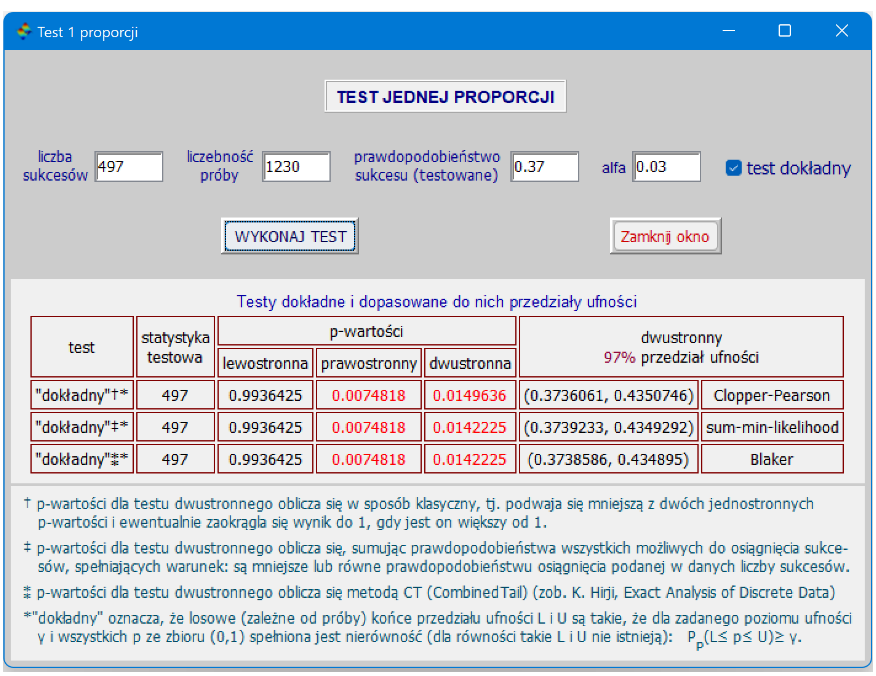
Po wpisaniu w pola liczby sukcesów, liczebności próby, prawdopodobieństwa (wartość domyślna to 0.5) i ewentualnej zmianie poziomu istotności alfa oraz kliknięciu w ’WYKONAJ TEST’ otrzymamy wyniki jak na Rys. 44 i Rys. 45.

Polecenie: ’Test proporcji (frakcji) dla 2 prób’

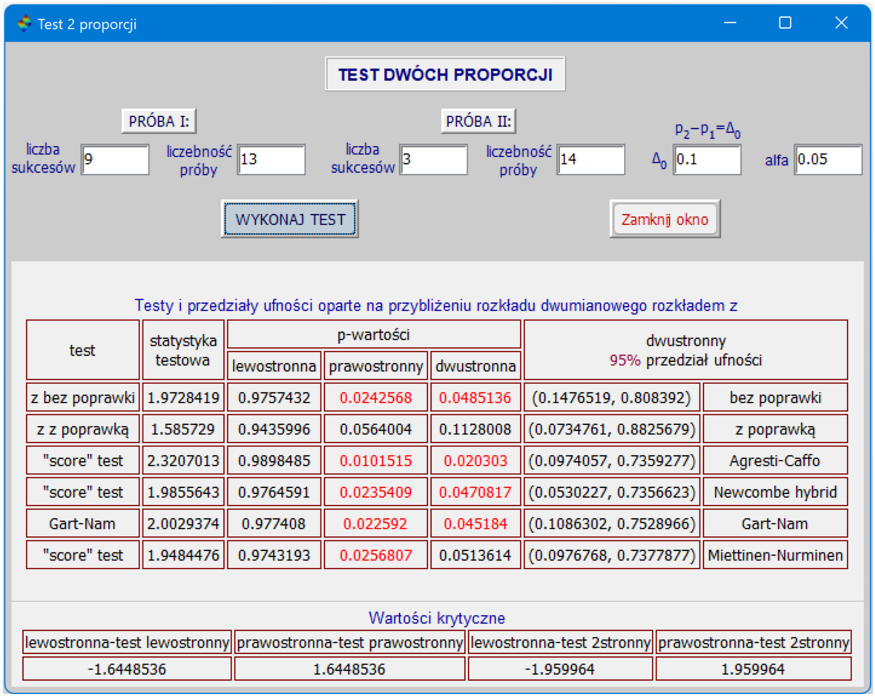
Wpisujemy w 4 polach odpowiednio liczbę sukcesów i liczebność I próby oraz liczbę sukcesów i liczebność II próby a w polu ’Δ_0’ liczbę Δ_0 (testujemy hipotezę $H_0 : p_1 - p_2 = \Delta_0$) lub zostawiamy wartość domyślną $\Delta_0 = 0$ dla $H_0 : p_1 = p_2$. Zmieniamy ewentualnie poziom istotności alfa i klikamy w ’WYKONAJ TEST’. Otrzymamy wynik jak na Rys. 46.



Rys. 44: Okno 'Test jednej proporcji' (testy przybliżone) z wpisanymi danymi i wynikami testów



Rys. 45: Okno 'Test jednej proporcji' (testy dokładne) z wpisanymi danymi i wynikami testów



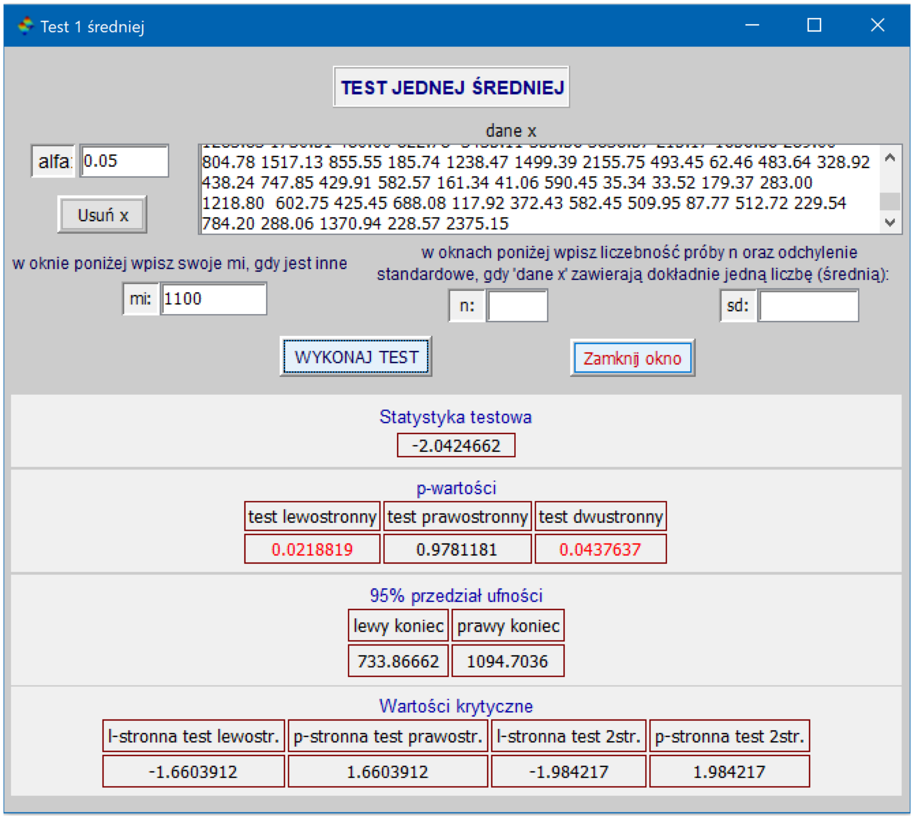
Rys. 46: Okno 'Test proporcji (frakcji) dla 2 prób' z wpisanymi danymi i wynikiem testu

Polecenie: 'Test średniej dla 1 próby'

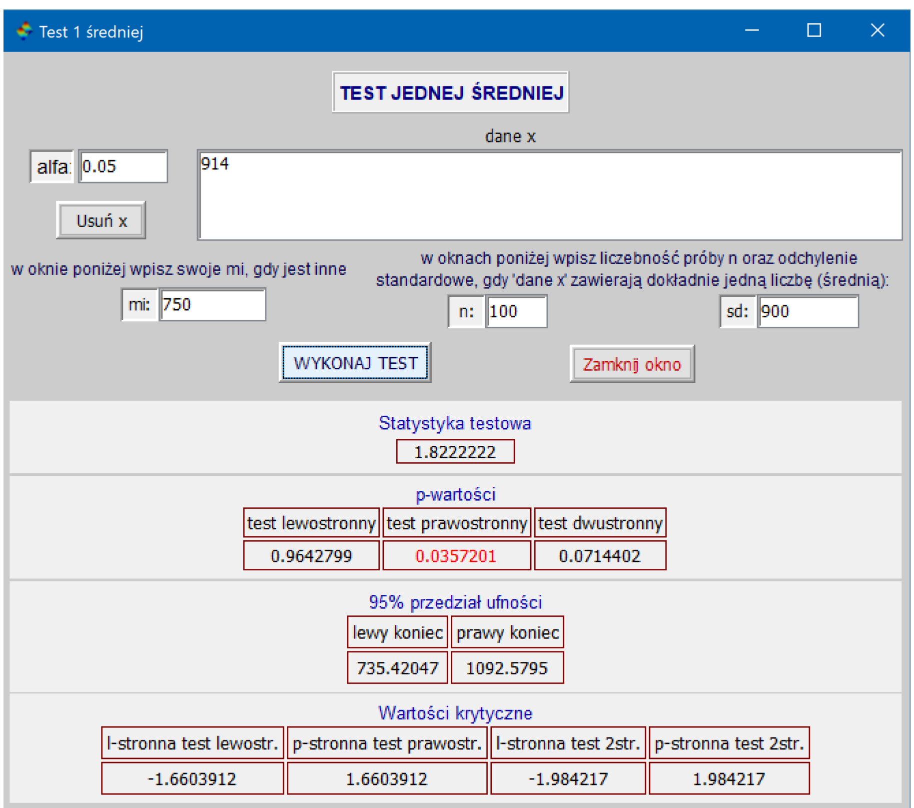
1. Po kliknięciu w polecenie 'Test średniej dla 1 próby' otworzy nam się okno, w którym możemy wykonać test hipotezy, że średnia z próby równa jest zadanej wartości mi. Na przykład wpisując dane w polu o nazwie 'dane x' i wpisując w polu 'mi:' średnią, z którą chcemy porównać średnią dla wpisanych danych oraz klikając w 'WYKONAJ TEST', otrzymamy wynik jak na Rys. 47.
2. W oknie tym możemy też wykonać test, nie mając danych źródłowych a mając tylko średnią, liczebność próby i odchylenie standardowe. Wpisujemy wtedy naszą średnią w polu dla danych, liczebność próby w polu oznaczonym 'n:', a odchylenie standardowe w polu oznaczonym 'sd:'. Wpisując na przykład średnią równą 914, średnią mi=750, liczebność n=100 i odchylenie standardowe sd=900 i klikając w 'WYKONAJ TEST', otrzymamy wynik testu z hipotezą H_0 : średnia μ w populacji skąd wzięta jest próba = 750 jak na Rys. 48.

Polecenie: 'Test średnich dla 2 prób'

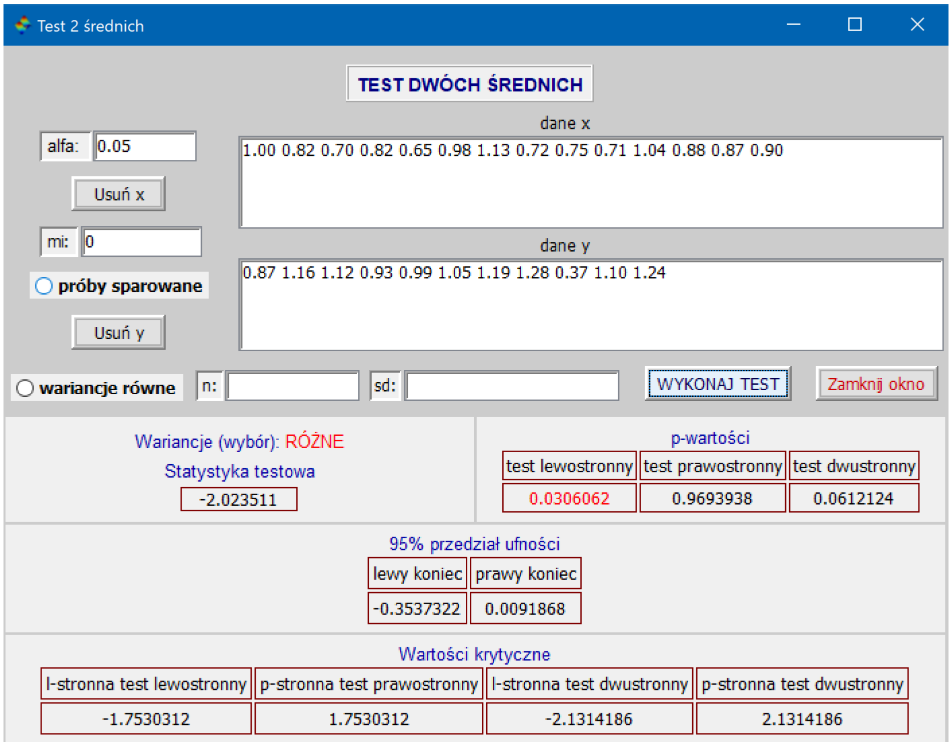
1. Po kliknięciu w polecenie 'Test średnich dla 2 prób' otworzy nam się okno, w którym możemy wykonać test hipotezy, że różnica średnich z 2 prób równa jest zadanej wartości mi. W szczególności, gdy wpiszemy w polu 'mi:' liczbę 0, to będziemy testować równość średnich. Na przykład wpisując dane w polu o nazwie 'dane x' i w polu 'dane y' i wpisując w polu 'mi:' liczbę 0 i niczego więcej ani nie zaznaczając, ani nie wpisując, a klikając w 'WYKONAJ TEST', otrzymamy wynik jak na Rys. 49.
2. Możemy też zaznaczyć opcję, że wariancje są równe i nie zmieniając niczego co poprzednio wpisaliśmy, otrzymamy wynik jak na Rys. 50.
3. Możemy taki test wykonać bez danych źródłowych, wpisując nasze średnie odpowiednio w okienko 'dane x' i okienko 'dane y'. Ale wtedy musimy znać liczebności obu prób i odchylenia standardowe w obu próbach. Jeśli wpiszemy w odpowiednie pola takie dane jak na Rys. 51, to otrzymamy widoczny na tym rysunku wynik.
4. W omawianym oknie możemy też wykonać test dla prób sparowanych, jeśli obie próby są równoliczne. Rys. 52 pokazuje otrzymane wyniki takiego testu.



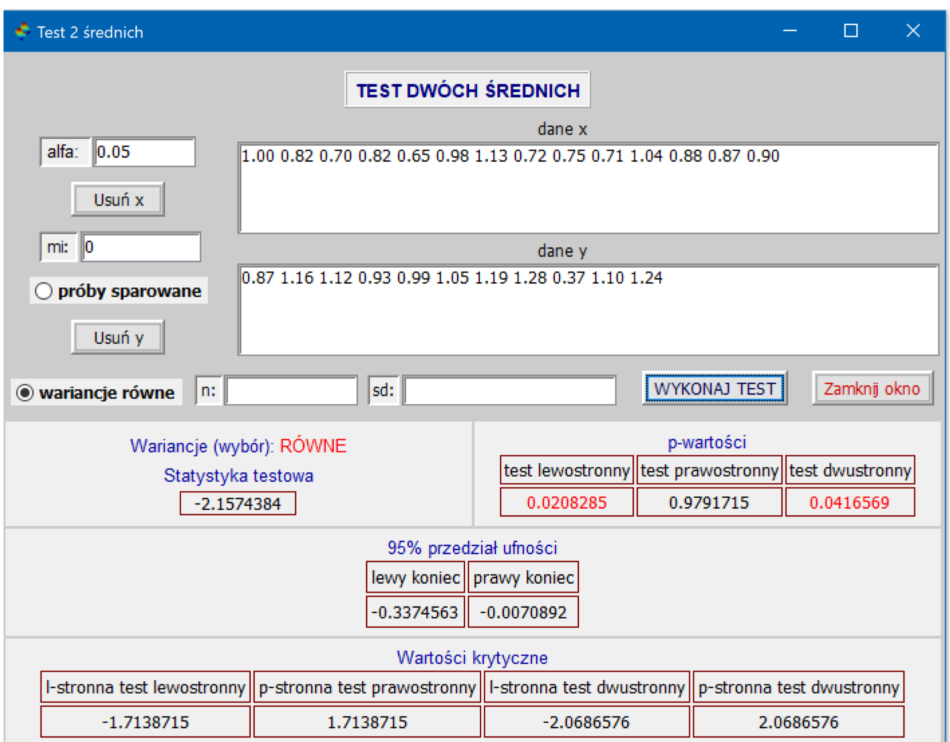
Rys. 47: Okno 'Test średniej dla 1 próby' z wpisanymi danymi i wynikiem testu



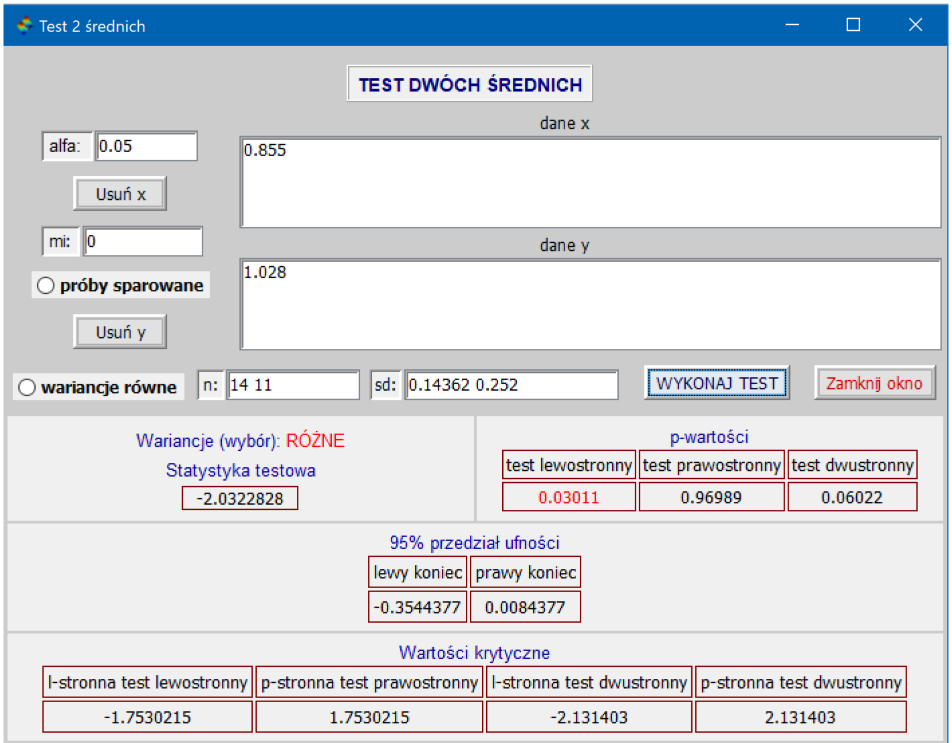
Rys. 48: Okno 'Test średniej dla 1 próby' z wpisanymi danymi i wynikiem testu



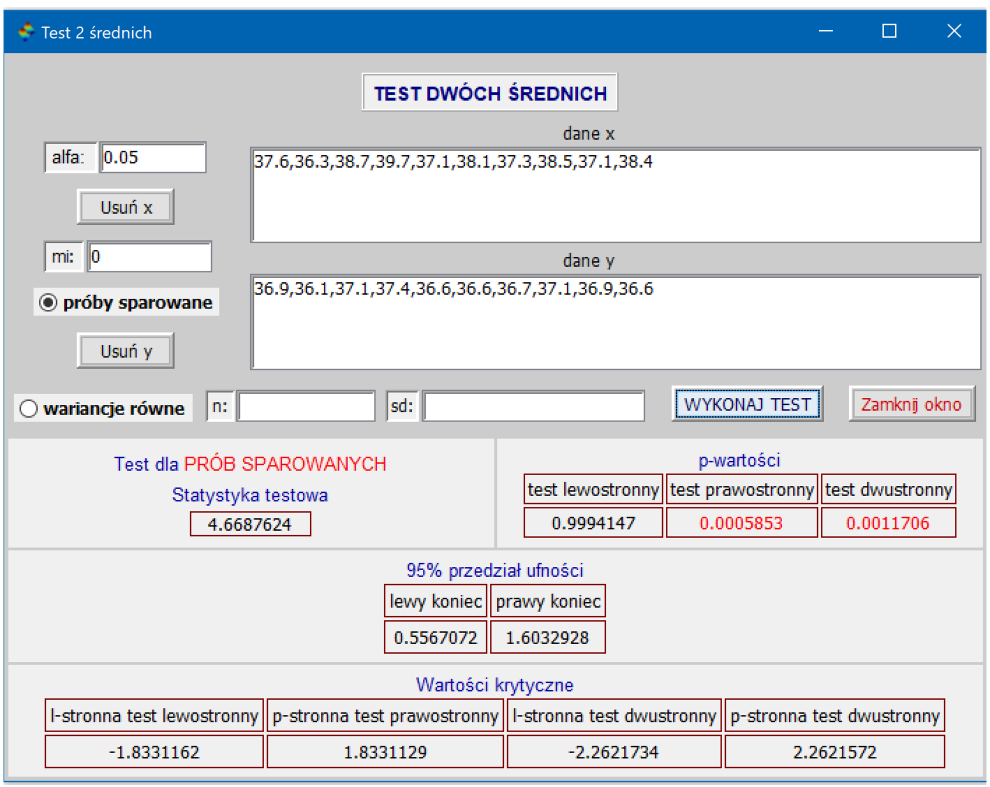
Rys. 49: Okno 'Test średnich dla 2 prób' z wpisanymi danymi przy założeniu, że wariancje są różne i wynikiem testu



Rys. 50: Okno 'Test średnich dla 2 prób' z wpisanymi danymi przy założeniu, że wariancje są równe i wynikiem testu



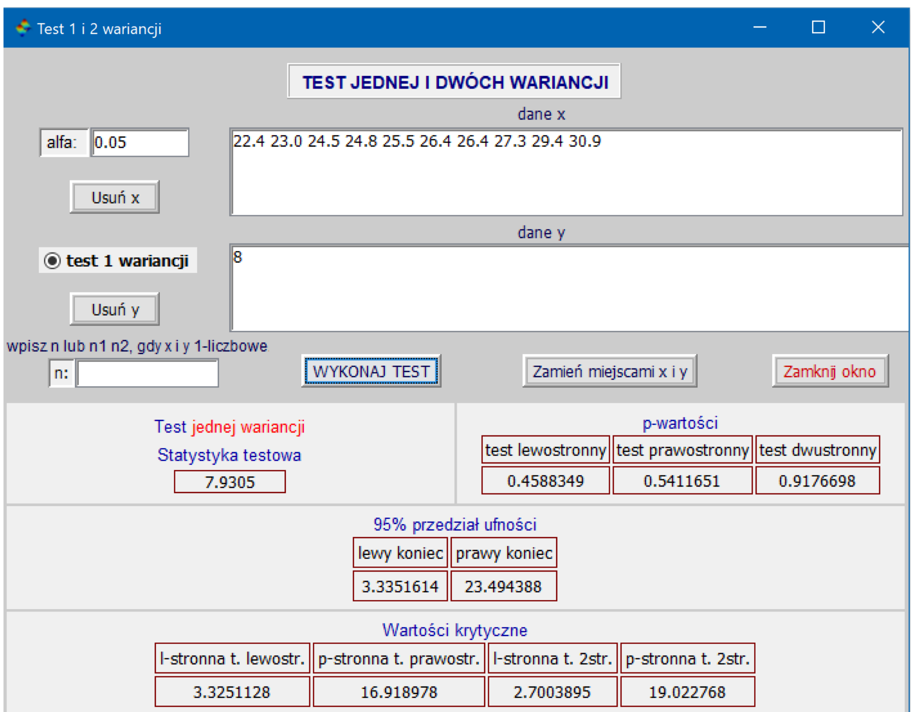
Rys. 51: Okno 'Test średnich dla 2 prób' z wynikiem testu dwóch średnich bez danych źródłowych z wpisanymi średnimi, liczebnościami prób i odchyleniami standardowymi dla obu prób



Rys. 52: Okno 'Test średnich dla 2 prób' dla danych sparowanych i wynikiem testu

Polecenie: ’Test wariancji dla 1 lub 2 prób’

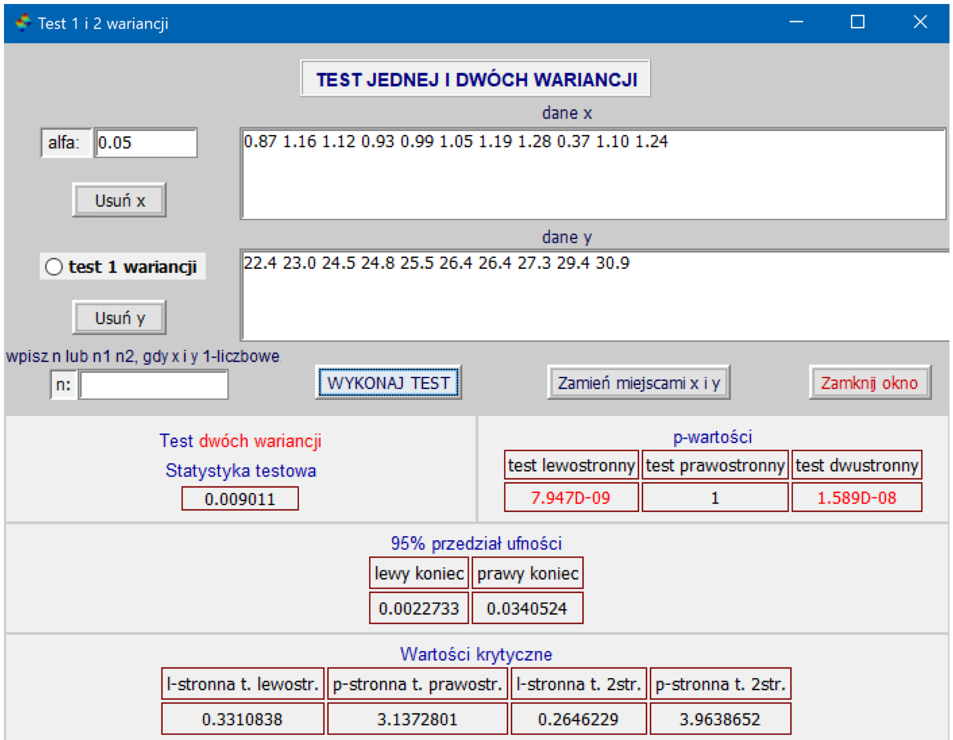
1. Klikając w polecenie ’Test wariancji dla 1 i 2 prób’ otrzymamy okno, które pozwoli nam wykonać zarówno test równości 2 wariancji, jak i test równości wariancji zadanej liczbie. Gdy wpiszemy wektor danych x do okienka ’dane x’, a do okienka ’dane y’ nieujemną liczbę oraz zaznaczymy opcję ’test 1 wariancji’ i klikniemy w ’WYKONAJ TEST’, to otrzymamy wynik testu równości wariancji danych x wpisanej liczbie y (zob. Rys. 53).
2. Wpisując wektor danych x do okienka ’dane x’ i wektor danych y do okienka ’dane y’ oraz ewentualnie zmieniając poziom istotności alfa i klikając w ’WYKONAJ TEST’ otrzymamy wynik testu równości wariancji danych x i wariancji danych y jak na Rys. 54.
3. Oba testy można też wykonać, nie mając danych ’surowych’, a mając za to wartość (lub wartości) wariancji oraz liczebność lub liczebności prób. Dokładniej, jeśli do obu okien danych wpiszemy po jednej liczbie i zaznaczymy opcję ’Test jednej wariancji’, to w polu ’n:’ trzeba wpisać liczebność próby, dla której mamy obliczoną wariancję. Klikając wtedy w ’WYKONAJ TEST’ otrzymamy wynik testu równości wpisanej wariancji i wpisanej liczby w polu ’dane y’. Gdy opcja ’test 1 wariancji nie będzie zaznaczona, to w polu ’n:’ trzeba wpisać dwie liczby: liczebność próby, z której pochodzi wariancja wpisana w polu ’dane x’ i liczebność próby, z której pochodzi wariancja wpisana w polu ’dane y. Klikając w ’WYKONAJ TEST’ otrzymamy wynik taki, jak np. na Rys. 55.



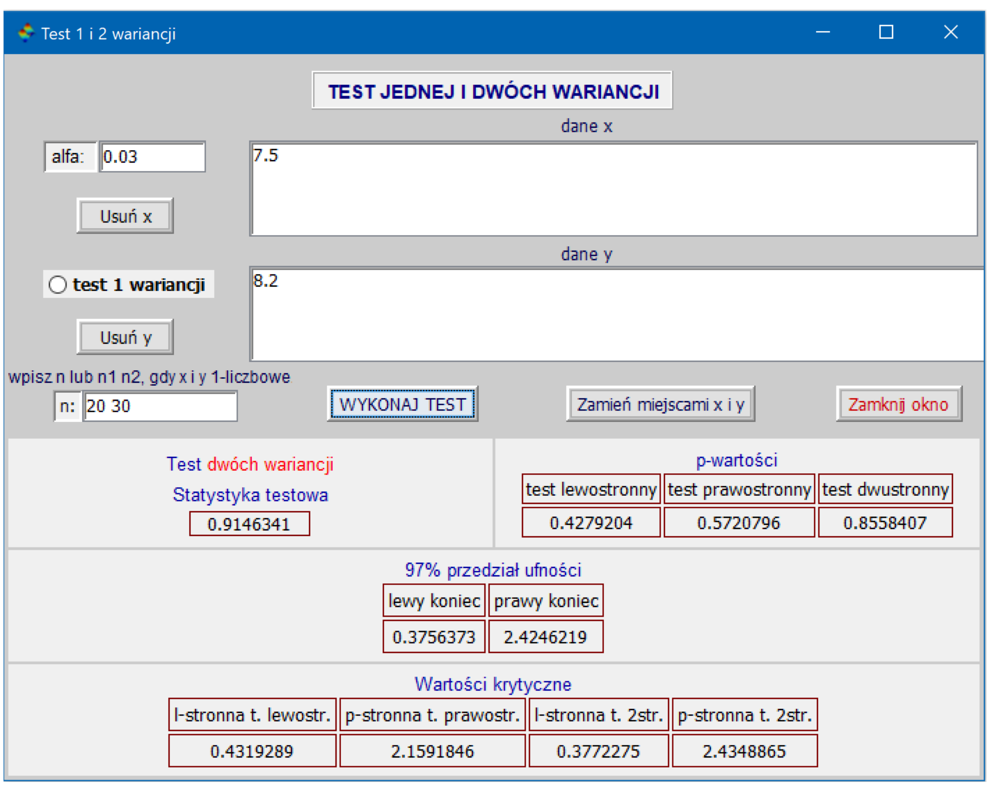
Rys. 53: Okno ’Test wariancji dla 1 lub 2 prób’ z wynikiem testu równości wariancji z próby x zadanej liczbie y

Polecenie: ’Test istotności współczynników korelacji’

1. Okno ’Test istotności współczynników korelacji’ daje nam możliwość zarówno obliczenia jednego z trzech współczynników korelacji, jak i przeprowadzenia testu istotności otrzymanego współczynnika korelacji. Aby obliczyć jeden z trzech współczynników korelacji, wpisujemy w okienka oznaczone ’dane x’ i ’dane y’ odpowiednio jedne i drugie dane (próby muszą być równoliczne), wybieramy jedną z opcji z nazwą współczynnika (Pearsona, Spearmana lub Kendalla) i klikamy w ’WYZNACZ WSPÓŁCZYNNIK KORELACJI’. Wtedy dany współczynnik zostanie obliczony i jego wartość zostanie wpisana w polu ’ρ :’ wraz z liczebnością próby w polu ’n:’ i (tylko dla współczynnika korelacji Spearmana i Kendalla) zaznacza kratkę ’brak powt. rang’ gdy wśród rang nie ma powtarzających się rang (ang. ties) (wtedy dla prób spełniających odpowiednie ograniczenia na wielkość próby: n<1291 dla testu Spearmana i n<161 dla testu Kendalla) wykonywane są dokładne testy istotności współczynnika korelacji Spearmana i Kendalla. W przypadku współczynnika korelacji Spearmana i



Rys. 54: Okno 'Test wariancji dla 1 lub 2 prób' z wynikiem testu równości wariancji z próby x i próby y



Rys. 55: Okno 'Test wariancji dla 1 lub 2 prób' z wynikiem testu wariancji x i wariancji y bez znajomości 'surowych danych' dla obu prób

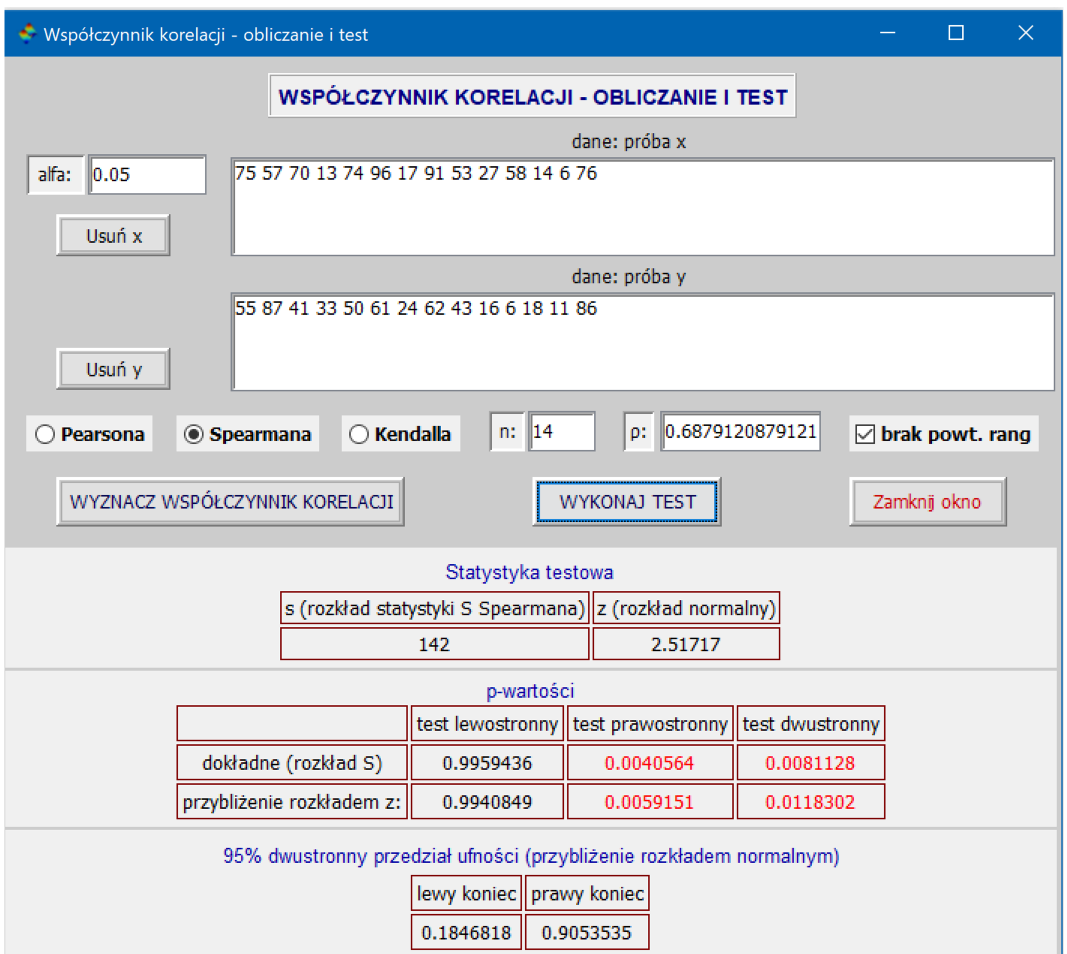
Kendalla kratka ta pozostaje niezaznaczoną, gdy wśród rang są powtarzające się rangi.

2. Jeśli chcemy wykonać test obliczonego współczynnika korelacji, to klikając w ’WYKONAJ TEST’ program pobierze wpisane automatycznie po obliczeniu współczynnika wyniki z dwóch pól, sprawdzi czy jest zaznaczona kratka ’brak powt. rang’ i wykona na ich podstawie test istotności oraz wyświetli wyniki, jak np. na Rys. 56.

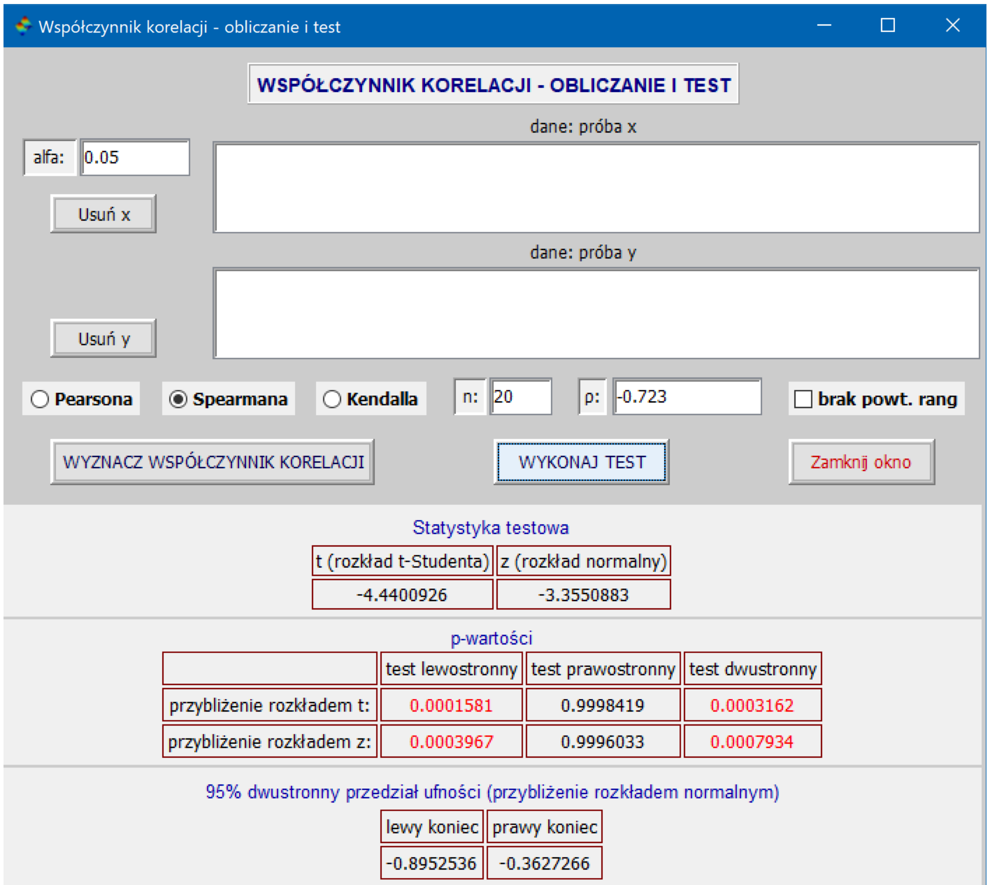
3. Ponieważ przy wykonywaniu testu istotności współczynnika korelacji program pobiera zawartości pól ’ρ :’, ’n:’ i informację o zaznaczeniu (lub nie) kratki ’brak powt. rang’, to test istotności współczynnika korelacji można przeprowadzić znając współczynnik korelacji i liczebności prób n. Aby go wykonać, wpisujemy współczynnik korelacji i liczebność n w pola ’ρ :’ i ’n:’, gdy wiemy, że nie było powtarzających się rang, to zaznaczamy kratkę ’brak powt. rang’ i klikamy w ’WYKONAJ TEST’. Otrzymamy wyniki testu, jak np. na Rys. 57.

Uwaga:

Jeśli wpisujemy sami współczynnik korelacji, to musi on być dość dokładny (co najmniej 4 miejsca po przecinku) - w przeciwnym razie test będzie mało dokładny.



Rys. 56: Okno z wynikiem obliczania współczynnika korelacji i testu istotności tego współczynnika



Rys. 57: Okno z wynikiem testu istotności współczynnika korelacji dla wpisanych danych n i ρ

Polecenie: 'Wyznaczanie mocy testów i liczebności prób'

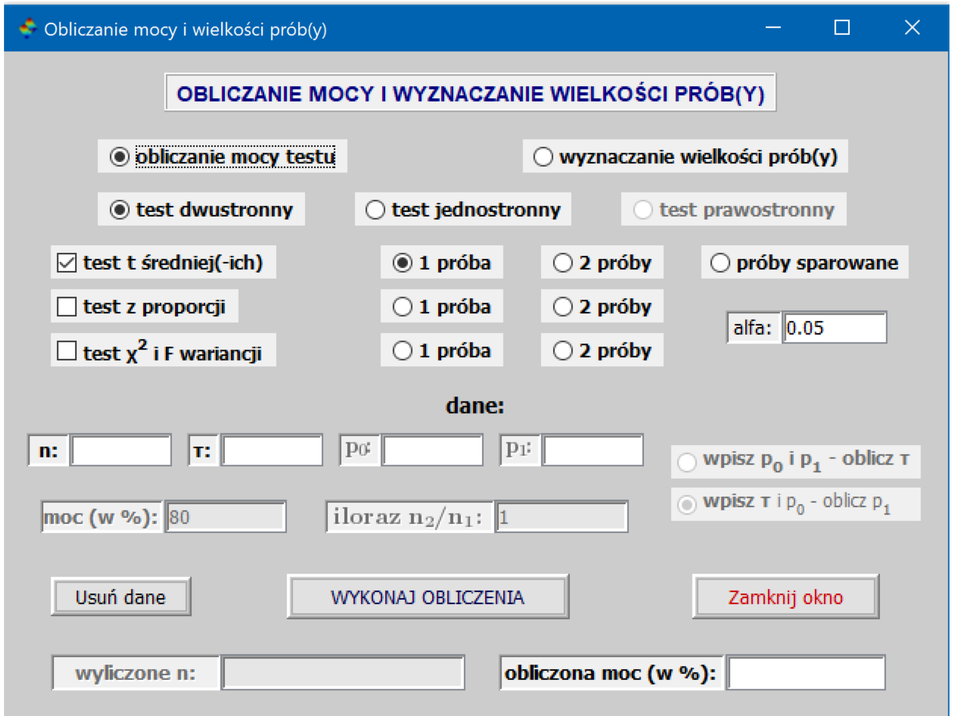
Klikając w polecenie 'Wyznaczanie mocy testów i liczebności prób' otwieramy okno, w którym mamy do wyboru obliczenie mocy dla jednego z trzech testów parametrycznych dla jednej lub dwóch prób lub wyznaczenia wielkości próby lub prób dla wspomnianych testów. Ponieważ te 3 testy dość istotnie różnią się od siebie, to w innych pakietach statystycznych są dla nich osobne funkcje lub okienka. My umieściliśmy je w jednym oknie, co powoduje, że korzystanie z tego okna jest nieco bardziej skomplikowane, a w konsekwencji opis korzystania z niego jest dłuższy i bardziej rozbudowany. Okno to ma postać taką jak na Rys. 58. Aby obliczyć moc testu lub wyznaczyć wielkość próby lub wielkości prób należy postępować wg następującej kolejności:

1. Klikając w odpowiedni przełącznik wybieramy albo obliczanie mocy albo wyznaczanie wielkości prób(y).
2. Następnie wybieramy jeden z trzech testów: test t średniej(-ich) lub test z proporcji, lub test χ^2 i F wariancji.
3. W następnym kroku wybieramy odpowiedni przełącznik decydując czy ma być to test dwustronny czy jednostronny (dla wymienionych w punkcie 2 dwóch ostatnich testów: test lewostronny lub prawostronny).
4. Ostatnią czynnością jest wpisanie danych w aktywne pola (te, które są oznaczone wyraźną a nie szarą czcionką) okienka i kliknięcie w przycisk 'WYKONAJ OBLICZENIA'.

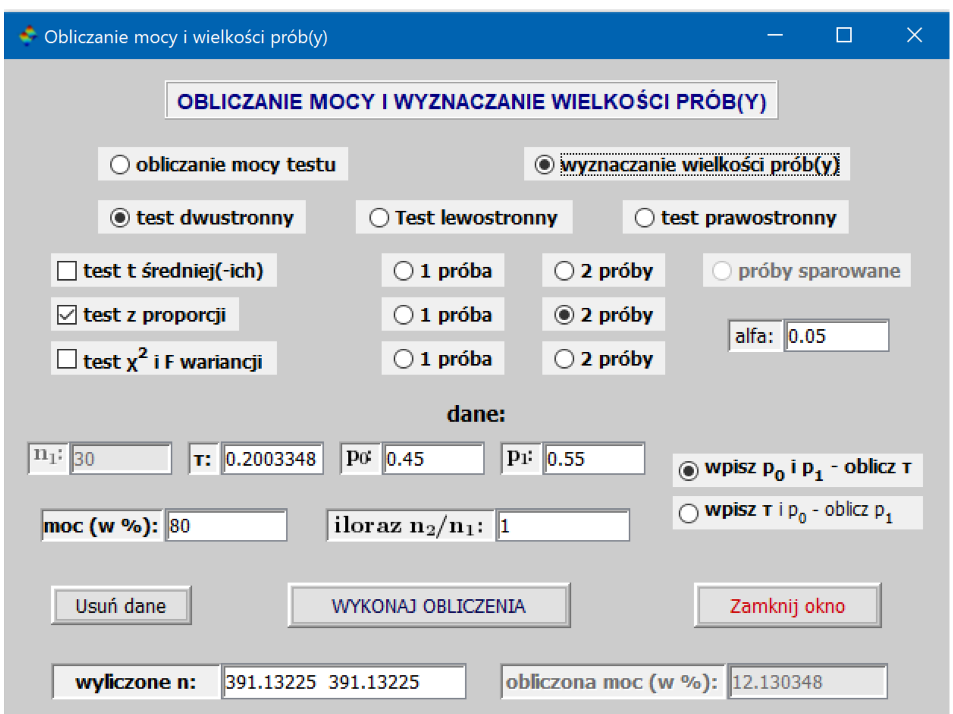
Zwróćmy uwagę na to, że w zależności od wyboru testu odpowiednie pola są aktywowane a te nieużywane w danym teście są dezaktywowane. Wyjaśnijmy dokładniej, co oznaczają dane, które należy wpisać.

- **n:** jest liczebnością próby a $\mathbf{n_1}$: liczebnością pierwszej próby. Minimalne wartości jakie możemy wpisać, to 2.
- $\boldsymbol{\tau}$ **:**, to wprowadzona przez Covena tzw. 'wielkość efektu' (siła relacji, ang. *"effect size"*) - liczba, inaczej rozumiana dla każdego z 3 testów, którą dokładniej objaśnimy później.
- $\mathbf{p_0}$: i $\mathbf{p_1}$: (tylko dla testu proporcji) - wielkości występujące w hipotezie $H_0 : p = p_0$ i obliczanej mocy dla $p = p_1$ $(\neq p_0)$, gdy hipoteza H_0 jest fałszywa a prawdziwa jest hipoteza alternatywna z $p = p_1$ $(\neq p_0)$ (np. dla testu dwustronnego prawdziwa jest hipoteza $H_a : p \neq p_0$ dla $p = p_1$).

- **moc (w %):** - żądana moc testu przy wyznaczaniu wielkości próby lub prób.
- iloraz $\mathbf{n_2/n_1}$ - (tylko gdy test dla 2 prób) służy do obliczenia wielkości n_2 2-giej próby (n_2 = wpisany iloraz $\cdot\, n_1$).



Rys. 58: Okno polecenia ’Wyznaczanie mocy testów i liczebności prób’



Rys. 59: Okno z wynikiem wyznaczania liczebności prób aby test 2 proporcji miał moc 80% z wcześniej obliczoną mocą tego testu, gdy próby mają po 30 elementów

kontynuacja: Polecenie: 'Wyznaczanie mocy testów i liczebności prób'

W opisywanym oknie mamy różne wybory rodzaju testu (dwustronny i jednostronny) dla różnych testów. A mianowicie, dla testu średniej lub średnich mamy wybór z dwóch możliwych testów: dwustronny i jednostronny bez wyszczególnienia, o który z dwóch jednostronnych testów chodzi: lewostronny czy prawostronny. Przy wyborze 'test jednostronny' otrzymamy moc lub wielkość próby tylko dla jednego z dwóch wymienionych (dla drugiego wielkości próby czy prób dla żądanej dużej mocy, np. 80% czy większej, nie da się wyznaczyć). I to użytkownik musi umieć zidentyfikować, dla którego z dwóch możliwych testów jednostronnych program obliczył moc lub wyznaczył wielkość próby. Zacznijmy zatem od testu średniej lub średnich. Daną "wielkość efektu" z naszego okna oznaczamy literą τ a literą δ oznaczamy parametr niecentralności rozkładu t-Studenta, który ma postać $\delta = \sqrt{n}\tau/\sigma$. Aby go obliczyć potrzebne jest nam odchylenie standardowe σ, którego na ogół nie znamy. Aby sobie poradzić z nieznajomością odchylenia standardowego σ, za jednostkę miary dla τ weźmiemy odchylenie standardowe σ. A zatem $\tau = 0.5$ będzie oznaczać $\tau = 0.5\sigma$. Np., gdy $\sigma = 10$ mm Hg, to $\tau = 0.5$ wyrażone w mm Hg będzie równe 5 mm Hg. Wstawiając takie τ do wzoru na δ, po uproszczeniu (podzieleniu licznika i mianownika przez σ otrzymamy $\delta = \sqrt{n}\tau$, gdzie teraz τ jest liczbą "bezwymiarową" i taką wstawiamy w polu oznaczonym symbolem $\boldsymbol{\tau}$.

Dla testu jednej średniej wstawione w polu $\boldsymbol{\tau}$ **:** τ jest wartością różnicy $\mu - \mu_0$ wyrażoną w jednostkach σ przy założeniu, że $H_0 : \mu = \mu_0$ jest fałszywa a prawdziwa jest hipoteza alternatywna dla $\mu = \mu_0 + \tau$. Natomiast przy teście dla 2 średnich $\boldsymbol{\tau}$ jest wartością różnicy $\mu_2 - \mu_1$ wyrażoną w jednostkach σ przy założeniu, że $H_0 : \mu_1 = \mu_2$ jest fałszywa a prawdziwa jest hipoteza alternatywna dla $\mu_2 = \mu_1 + \tau$. Program po wczytaniu τ bierze jej wartość bezwzględną i przy wyborze testu jednostronnego zawsze wylicza moc lub wyznacza wielkość próby lub prób dla testu prawostronnego. Jeśli jednak interesuje nas moc testu gdy prawdziwa jest hipoteza alternatywna $\mu = \mu_0 + \tau$ (jedna próba) lub $\mu_2 = \mu_1 + \tau$, gdzie $\tau < 0$, to racjonalny będzie test lewostronny i jego moc lub liczebność próby przy zadanej mocy będą takie same jak dla testu prawostronnego, gdy $\tau > 0$, które otrzymamy przy wyborze przełącznika 'Test jednostronny'.

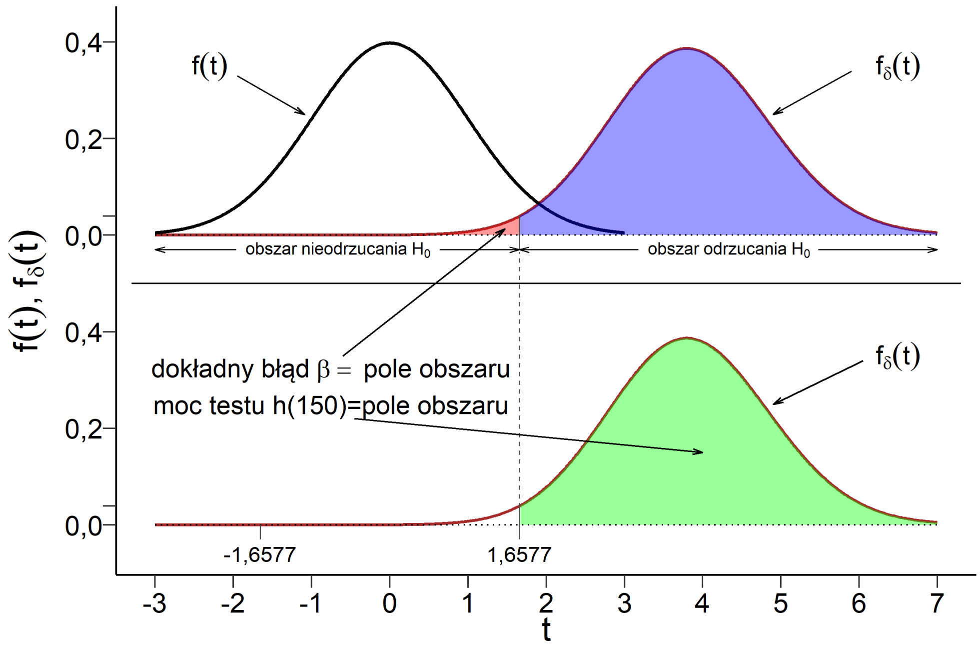
Podsumowując, jeśli pytamy się o moc testu średniej lub średnich dla $\tau < 0$, to wyniki obliczeń przy wyborze 'Test dwustronny' są wynikami dla testu lewostronnego, a dla $\tau > 0$ są wynikami dla testu prawostronnego. Oczywiście, możemy wpisać jako daną $\tau < 0$, ale otrzymamy dokładnie to samo, co dla $\tau > 0$, ponieważ zostanie wykonany test prawostronny dla $|\tau|$.

Aby zobaczyć, że tylko dla jednego z testu dwustronnego da się wyznaczyć liczebność prób(y) dającą zadaną dużą moc rozważmy test prawostronny dla 1 próby zilustrowany na Rys. 60. Wykonując ten rysunek, wzięliśmy $\tau = 0.3472$, otrzymując parametr niecentralności rozkładu t-Studenta z 120 stopniami swobody równy $\delta = \sqrt{121}\cdot 0.3472 = 3.8194$. Zauważmy, że jeśli $\mu_0 = 140$ mm Hg a $\sigma = 28.8$ mm Hg, to $\tau = 0.3472\cdot 28.8 \approx 10$ i zakładamy, że hipoteza $H_a : \mu > \mu_0$ jest prawdziwa, a dokładniej, że jest ona prawdziwa dla $\mu = \mu_0 + \tau = 140 + 10 = 150$ mm Hg.

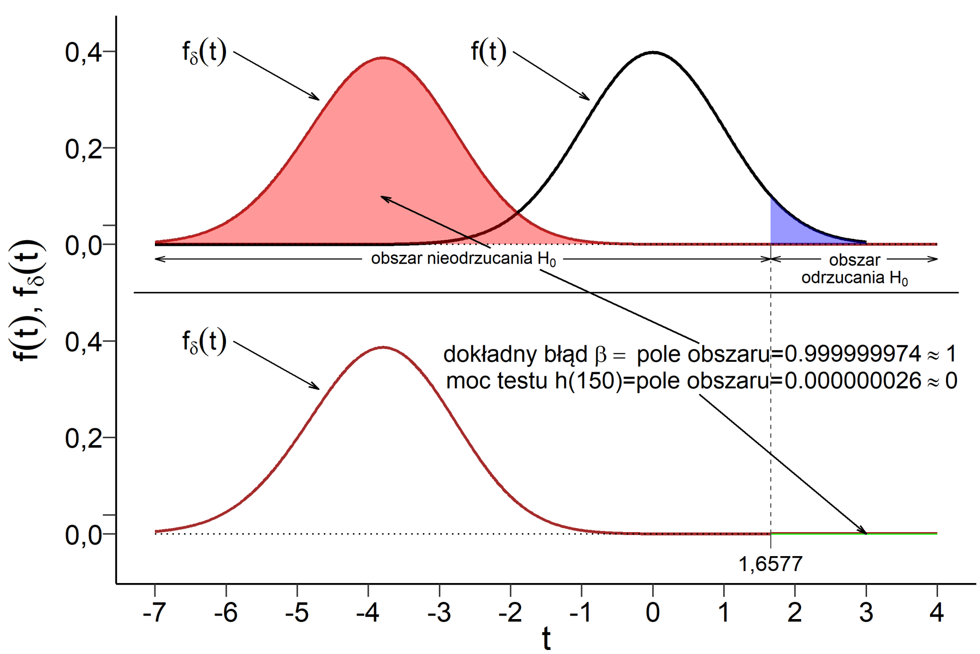
Wiemy, że błąd II rodzaju β, to prawdopodobieństwo nieodrzucenia hipotezy H_0, gdy jest ona fałszywa (tak, jak założyliśmy), a moc testu, to prawdopodobieństwo odrzucenia takiej hipotezy. Dla poziomu istotności $\alpha = 0.05$, liczebności próby $n = 121$ i testu prawostronnego, prawdopodobieństwo to jest równe prawdopodobieństwu, że wartość statystyki testowej T będzie mniejsza od wartości krytycznej $t^{kr}_{0.05,120} = 1.6577$. Ponieważ założyliśmy, że hipoteza H_0 jest fałszywa, to statystyka T nie ma rozkładu t-Studenta a ma niecentralny rozkład t-Studenta z parametrem niecentralności δ i musimy tego rozkładu użyć, aby znaleźć to prawdopodobieństwo. Z wykresu widzimy, że jest ono równe obszarowi, który jest zaznaczony kolorem czerwonym i że jest ono małe. Natomiast moc testu, gdy prawdziwa średnia $\mu = 150$ mm Hg, jest równa 1-β i jest duża (= 96.626246), a zwiększając liczebność można ją jeszcze zwiększać, ponieważ przy ustalonym τ zwiększa się parametr niecentralności $\delta = \sqrt{n}\cdot\tau$.

A teraz spójrzmy na Rys. 61, który pokazuje co się dzieje, gdy weźmiemy $\tau = -0.3472$, a pozostałe dane pozostawimy bez zmian i zamiast testu lewostronnego, który dla takich danych jest testem odpowiednim, wybierzemy test prawostronny. Ponieważ dla takich danych parametr niecentralności δ rozkładu t-Studenta będzie ujemny ($\delta = -3.8194$), to wykres gęstości f_δ będzie przesunięty w lewo względem wykresu gęstości f testu t-Studenta a obszar odrzucania hipotezy zerowej pozostanie taki jaki był poprzednio, tj. na lewo od wartości krytycznej $t^{kr}_{0.05,120} = 1.6577$. Prawdopodobieństwo, że wartość statystyki testowej T znajdzie się w tym obszarze obliczamy korzystając z rozkładu statystyki T, którym jest rozkład, którego gęstością jest f_δ. Na rysunku widzimy, że prawdopodobieństwo to jest równe obszarowi zaznaczonemu kolorem czerwonym i jest bardzo duże, a jego dopełnienie do 1 jest bardzo małe, czyli moc takiego testu jest bardzo mała. Zwiększając liczebność próby n prawdopodobieństwo będzie się zmniejszać, bo wykres f_δ będzie się przesuwać w lewo.

W przypadku pozostałych testów mamy podobną własność, tj. tylko dla jednego z dwóch testów jednostronnych, wybranych prawidłowo w podobny sposób jak opisaliśmy to dla testu średnich, można wyznaczyć liczebność prób lub próby dla z góry zadanej dużej mocy. Często taką mocą jest moc 0.8 (80%) czy 0.9 (90%).



Rys. 60: Ilustracja wyznaczania mocy prawostronnego testu średniej dla jednej próby, gdy jest to właściwie wybrany test ($\tau > 0$)



Rys. 61: Ilustracja wyznaczania mocy prawostronnego testu średniej dla jednej próby, gdy właściwym testem, który powinien być wybrany jest test lewostronny ($\tau < 0$)

kontynuacja: Polecenie: 'Wyznaczanie mocy testów i liczebności prób'

Obliczając moc testu proporcji lub przy zadanej mocy wyznaczając liczebności prób lub liczebność próby też korzystamy z "wielkości efektu" τ wpisywanego w polu oznaczonym symbolem $\boldsymbol{\tau}$ **:**. Ponieważ testujemy równość proporcji p i p_0 (lub p_1 i p_0 gdy mamy 2 próby), to wydaje się naturalnym wziąć zwykłą różnicę $\tau = p - p_0$ ($\tau = p_1 - p_0$) jako *wielkość efektu* bez wprowadzania specjalnych jednostek, jak to było w przypadku testu średniej (średnich), ponieważ żadne nieznane parametry nie występują w rozkładzie statystyki testowej Z, gdy hipoteza zerowa $H_0 : p = p_0$ ($H_0 : p_1 = p_0$) nie jest prawdziwa a prawdziwa jest hipoteza H_a dla $p = p_0 + \tau$ ($p_1 = p_0 + \tau$). Wtedy rozkładem statystyki testowej jest rozkład normalny ze zmienioną średnią i wariancją. Jednak w przypadku testu proporcji pojawia się inny problem, który nie występował w testach średniej. Okazuje się bowiem, że moc testu nie zależy tylko od różnicy $\tau = p - p_0$ ($\tau = p_1 - p_0$) ale też od wartości p i p_0 (p_1 i p_0). Aby się uniezależnić od tej zależności korzysta się z przekształcenia $2\arcsin(p)$ i zamiast wypisanych wcześniej zwykłych różnic proporcji jako *wielkości efektu* używa się $\tau = 2\arcsin(p) - 2\arcsin(p_0)$ ($\tau = 2\arcsin(p_1) - 2\arcsin(p_0)$). **Uwaga: Możemy wpisywać $\boldsymbol{\tau} > \mathbf{0}$ i $\boldsymbol{\tau} < \mathbf{0}$.**

Ponieważ brak jest intuicji jak różnica proporcji p_1 i p_0 (w przypadku testu jednej proporcji należy w miejsce p_1 wstawić p) przekłada się na różnicę $2\arcsin(p_1) - 2\arcsin(p_0)$ lub jakie należy wpisać τ aby test wykrywał różnicę proporcji $p_1 - p_0$, w omawianym oknie wprowadziliśmy dwa przełączniki:

1. **wpisz $\boldsymbol{p_0}$ i $\boldsymbol{p_1}$ - oblicz $\boldsymbol{\tau}$**, który wybieramy, gdy nie wiemy jaką wartość τ wpisać w polu $\boldsymbol{\tau}$ **:**, a wiemy, dla jakiej różnicy $p_1 - p_0$ proporcji chcemy obliczyć moc testu lub jaka różnica $p_1 - p_0$ proporcji, jeśli taka istnieje, ma być wykrywana przez test a fałszywa hipoteza zerowa odrzucana z zadanym prawdopodobieństwem (mocą);
2. **wpisz $\boldsymbol{\tau}$** i p_0 oblicz p_1, który wybieramy, gdy wiemy jaką wartość τ wpisać, znamy p_0 i chcemy wiedzieć jakiej różnicy $p_1 - p_0$ odpowiada wartość *wielkości efektu* τ. Zauważmy, że dla tego przełącznika tłusty druk został użyty tylko dla **wpisz $\boldsymbol{\tau}$**. Zrobiono tak, aby zaznaczyć, że wpisywanie p_0 jest opcjonalne i jak je wpiszemy, to p_1 będzie obliczone, a jak nie wpiszemy p_0, to p_1 nie zostanie obliczone.

Zajmijmy się teraz testami wariancji dla jednej i dwóch prób.

1. Obliczamy moc testu lub wyznaczamy liczebność próby dla jednej wariancji przy wyborze jednego z trzech typów testu: dwustronny, lewostronny i prawostronny. Hipoteza zerowa ma postać: $H_0 : \sigma^2 = \sigma_0^2$. Zakładamy, że hipoteza H_0 jest fałszywa, a prawdziwa jest, w zależności od typu wybranego testu, jedna z hipotez alternatywnych: $H_a : \sigma^2 \neq \sigma_0^2$ lub $H_a : \sigma^2 < \sigma_0^2$, lub $H_a : \sigma^2 > \sigma_0^2$ dla $\sigma^2 = \sigma_1^2$, $\sigma_1^2 \neq \sigma_0^2$. *Wielkością efektu* τ dla testu jednej wariancji jest iloraz
$$\frac{\sigma_1^2}{\sigma_0^2}$$
Z takiego wyboru *wielkości efektu* wynika, że jest ona zawsze dodatnia i jeśli chcemy wybrać test jednostronny aby wyznaczyć liczebność próby dla ustalonej mocy, to właściwym wyborem testu jednostronnego jest test lewostronny, gdy *wielkość efektu* τ spełnia nierówność $\tau < 1$ i test prawostronny, gdy *wielkość efektu* τ spełnia nierówność $\tau > 1$.

2. Teraz obliczamy moc testu lub wyznaczamy liczebność próby dla dwóch wariancji przy wyborze jednego z trzech typów testu: dwustronny, lewostronny i prawostronny. Hipoteza zerowa ma postać: $H_0 : \sigma_1^2 = \sigma_2^2$. Zakładamy, że hipoteza H_0 jest fałszywa, a prawdziwa jest, w zależności od typu wybranego testu, jedna z hipotez alternatywnych: $H_a : \sigma_1^2 \neq \sigma_2^2$ lub $H_a : \sigma_1^2 < \sigma_2^2$, lub $H_a : \sigma_1^2 > \sigma_2^2$ dla $\sigma_1^2 = \sigma_3^2$, $\sigma_3^2 \neq \sigma_2^2$. *Wielkością efektu* τ dla testu dwóch wariancji jest iloraz
$$\frac{\sigma_3^2}{\sigma_2^2}$$
Z takiego wyboru *wielkości efektu* wynika, że jest ona zawsze dodatnia i jeśli chcemy wybrać test jednostronny aby wyznaczyć liczebność próby dla ustalonej mocy, to właściwym wyborem testu jednostronnego jest test lewostronny, gdy *wielkość efektu* τ spełnia nierówność $\tau < 1$ i test prawostronny, gdy *wielkość efektu* τ spełnia nierówność $\tau > 1$.

Zauważmy, że w przypadku testów średniej lub średnich oraz testów proporcji im *wielkości efektu* τ są bliższe zera, tym większe muszą być liczebności prób aby uzyskać dużą moc.
Natomiast w przypadku testów wariancji im *wielkości efektu* τ są bliższe jedynki, tym większe muszą być liczebności prób aby uzyskać dużą moc.

Polecenie: ’Miary asocjacji - współczynniki i testy istotności’

Aby obliczyć 4 współczynniki asocjacji: Gamma Goodmana-Kruskala, tau b Kendalla i współczynników korelacji ρ Pearsona oraz ρ Spearmana i wykonać testy istotności tych obliczonych współczynników, należy:

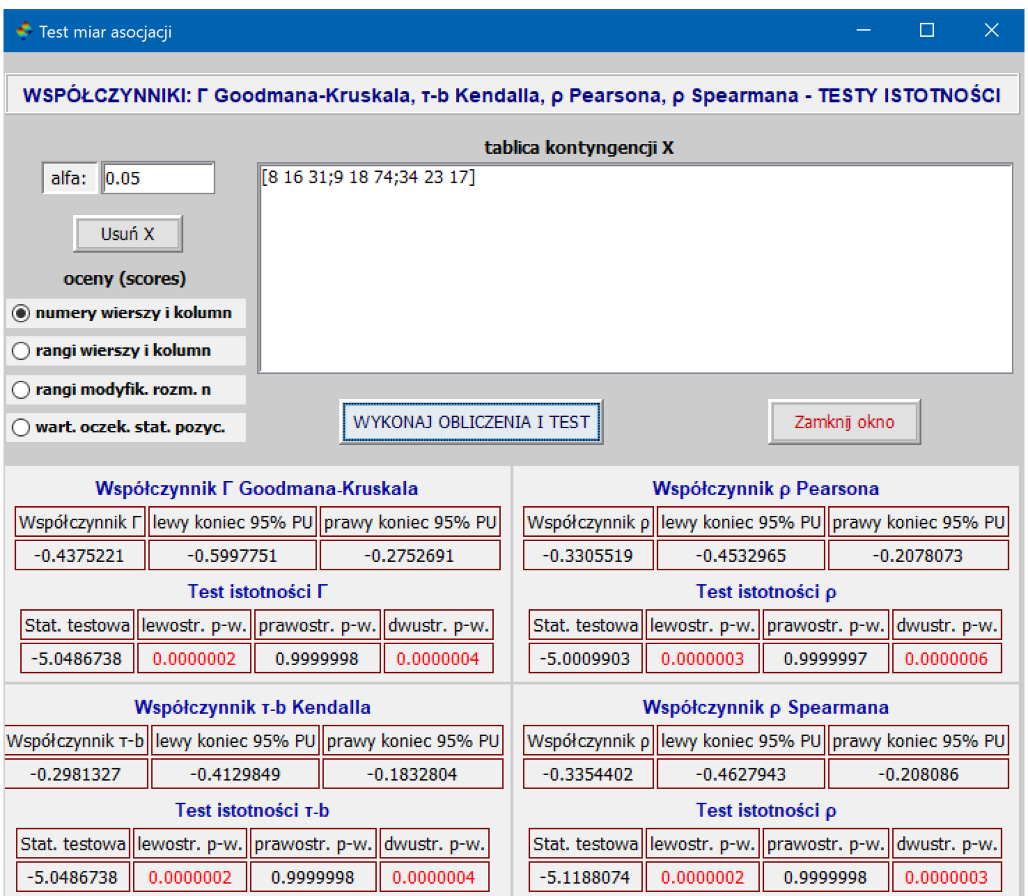
1. wpisać w okienko ’tablica kontyngencji X’ tablicę kontyngencji, np. taką jak na Rys. 62;
2. wybrać opcję ’oceny (scores)’ jeśli ma być inna niż kolejne numery wierszy i kolumn podanej tablicy kontyngencji (**Uwaga:** Wybór którejkolwiek z 4 opcji wpływa tylko na wartość współczynnika ρ korelacji Pearsona);
3. po ewentualnej zmianie poziomu istotności alfa kliknąć w przycisk ’WYKONAĆ OBLICZENIA I TEST’.

Otrzymamy 4 współczynniki asocjacji i przedziały ufności dla nich oraz statystyki testowe i p-wartości.

Uwagi o przedziałach ufności i statystykach testowych

1. Użyte tutaj przy konstrukcji przedziałów ufności dla współczynników korelacji ρ Pearsona oraz ρ Spearmana oraz w testach istotności tych współczynników asymptotyczne błędy standardowe (ASE) nie wymagają założenia ciągłości i normalności rozkładów obu zmiennych losowych.
2. Statystyki testowe oraz odpowiednie p-wartości w testach współczynników Gamma Goodmana-Kruskala i tau b Kendalla są jednakowe (i równe statystykom testowym i odpowiednim p-wartościom nie obliczanych tutaj współczynników tau-c Stuarta, Somersa D(R|C) oraz Somersa D(C|R)).
3. W przypadku tablicy kontyngencji rozmiaru 2×2 współczynnik Gamma Goodmana-Kruskala jest współczynnikiem Q Yula (nazywany też czasami współczynnikiem Q Kendalla).
4. Uzyskiwane w tym oknie wyniki są zgodne z wynikami, jakie daje pakiet statystyczny SAS.

Przykładowe wyniki obliczeń 4 współczynników służących za miary asocjacji widzimy na Rys. 62.



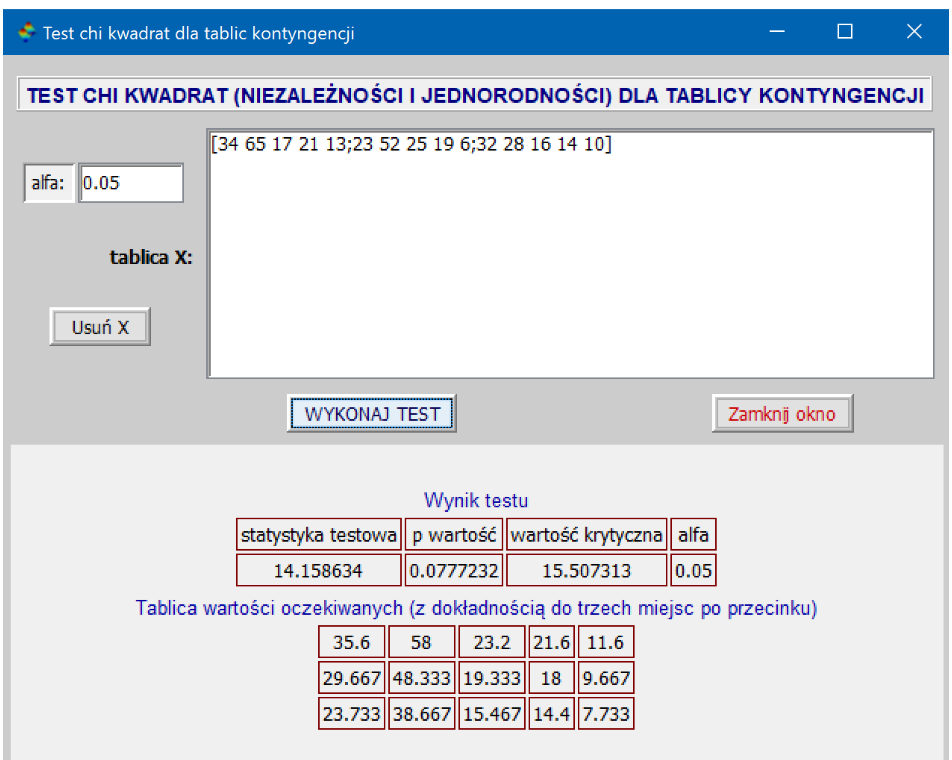
Rys. 62: Okno z obliczonymi współczynnikami, przedziałami ufności i wynikami testów istotności

Polecenie: ’Test chi kwadrat dla tablic kontyngencji’ (test niezależności i jednorodności)

1. Po otwarciu okna ’Test chi kwadrat dla tablic kontyngencji’ wpisujemy do okienka tablicę kontyngencji i klikamy w ’WYKONAJ TEST’.
2. Tablicę kontyngencji możemy wpisać w okienko w formacie Scilaba: wpisujemy nawias kwadratowy otwierający, następnie wpisujemy kolejne elementy pierwszego wiersza oddzielane spacją i kończymy ten wiersz średnikiem. Wpisujemy w ten sam sposób drugi wiersz i kończymy go średnikiem, gdy nie jest to ostatni wiersz, a gdy jest to ostatni wiersz, to kończymy go nawiasem kwadratowym zamykającym. Wpisywanie nawiasów kwadratowych można pominąć. Przykład tak wpisanej tablicy wraz z wynikiem testu widzimy na Rys. 63.
3. Drugi dopuszczalny sposób wpisania takiej tablicy do okienka jest następujący: wpisujemy kolejne elementy pierwszego wiersza oddzielane spacją i kończymy ten wiersz naciśnięciem klawisza ’Enter’, po czym wpisujemy kolejne wiersze w ten sam sposób. Pomiędzy wierszami danych nie powinno być pustych wierszy, czyli po wpisaniu każdego wiersza możemy tylko raz nacisnąć klawisz ’Enter’. Przykład tak wpisanej tablicy wraz z wynikiem testu widzimy na Rys. 64.

Uwaga (kryterium Cochrana)

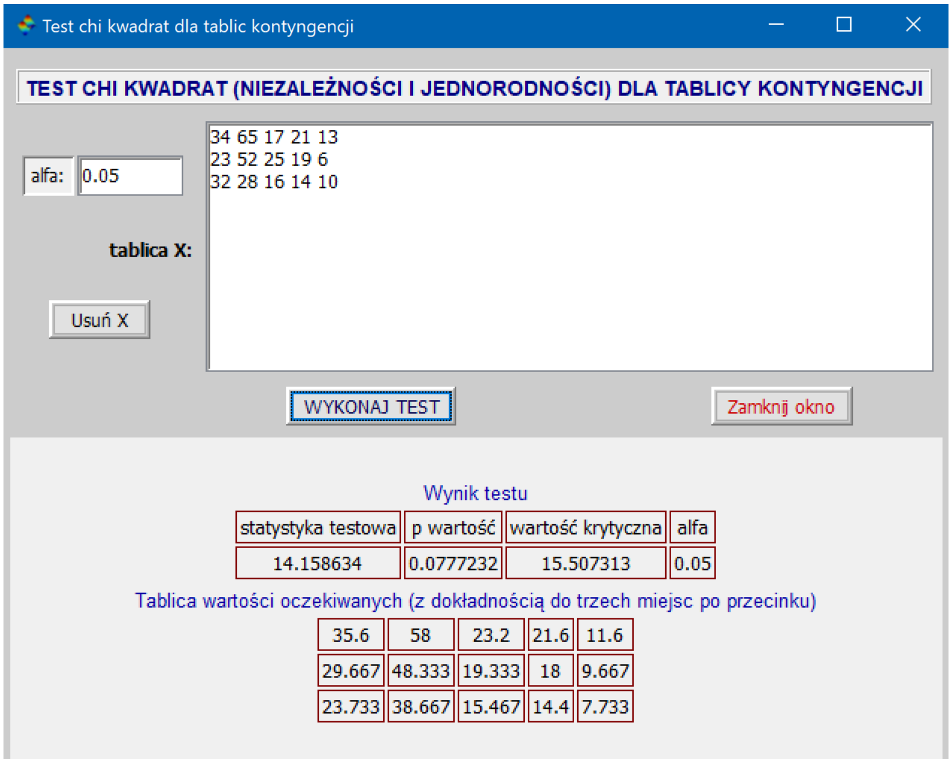
Jeśli któraś z obliczonych wartości oczekiwanych o_{ij} jest mniejsza od 1 lub wartości oczekiwanych o_{ij}, takich że $1 \le o_{ij} < 5$ jest więcej niż 20%, to podawane jest ostrzeżenie o niespełnieniu kryterium Cochrana.



Rys. 63: Okno z wynikiem testu chi kwadrat dla tablic kontyngencji

Polecenie: ’Test chi kwadrat zgodności rozkładu z zadanym rozkładem’

1. Tutaj danymi są wektory a nie tablice. Dlatego wpisujemy je tak jak w wielu wcześniejszych oknach oprócz okna testu chi kwadrat dla tablic kontyngencji. Wektor danych **xocz** jest w istocie wektorem wartości oczekiwanych. Tutaj mamy też możliwość ustalenia minimalnej wielkości wartości oczekiwanych. Jeśli któraś z wartości



Rys. 64: Okno z wynikiem testu chi kwadrat dla tablic kontyngencji

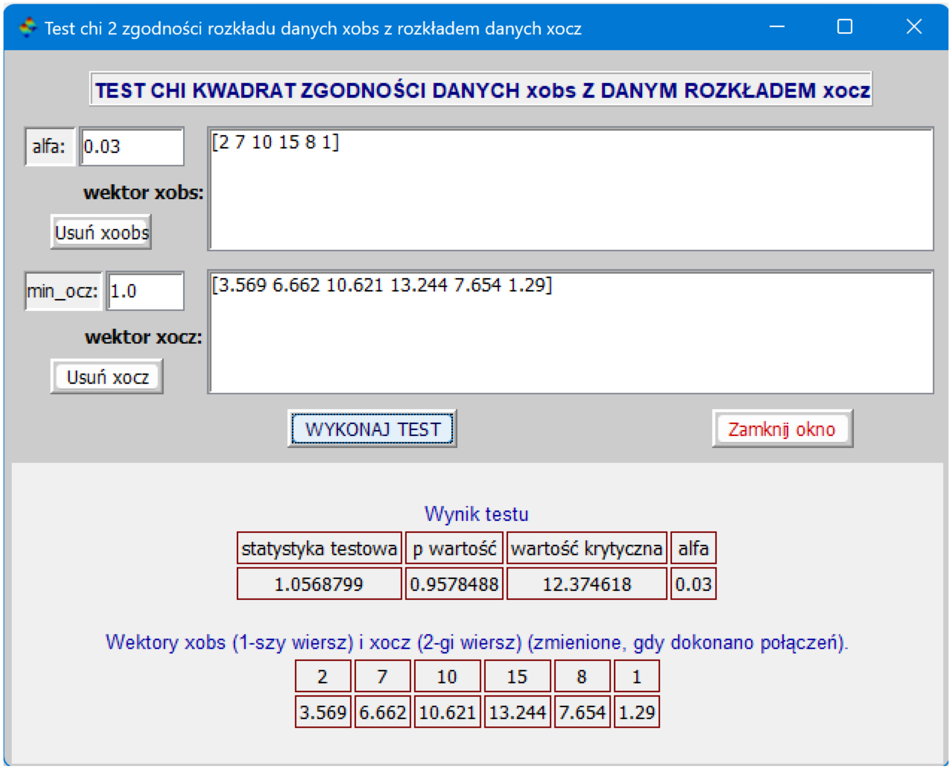
oczekiwanych jest mniejsza niż podane ograniczenie w polu 'min_xocz:', to komórka, w której ta wartość się znajduje jest łączona z sąsiednią komórką i jednocześnie łączy się odpowiednie komórki wektora xobs. Po połączeniu sprawdza się, czy wartość oczekiwana uzyskana przez zsumowanie sąsiednich wartości oczekiwanych jest większa od podanego ograniczenia. Jeśli nie, to proces łączenia sąsiednich wartości wektora xocz powtarza się tak długo, aż suma będzie większa od ograniczenia lub pozostaną tylko 2 liczebności.

2. Zwróćmy uwagę na to, że elementami wektora xobs są liczebności i muszą to być liczby całkowite nieujemne. Natomiast elementy wektora xocz nie muszą być całkowite. Na Rys. 65 widzimy wynik testu, gdzie nie było potrzeby łączenia sąsiednich komórek, a na Rys. 66 mamy dane, dla których trzeba było połączyć dwie pierwsze i dwie ostatnie komórki.

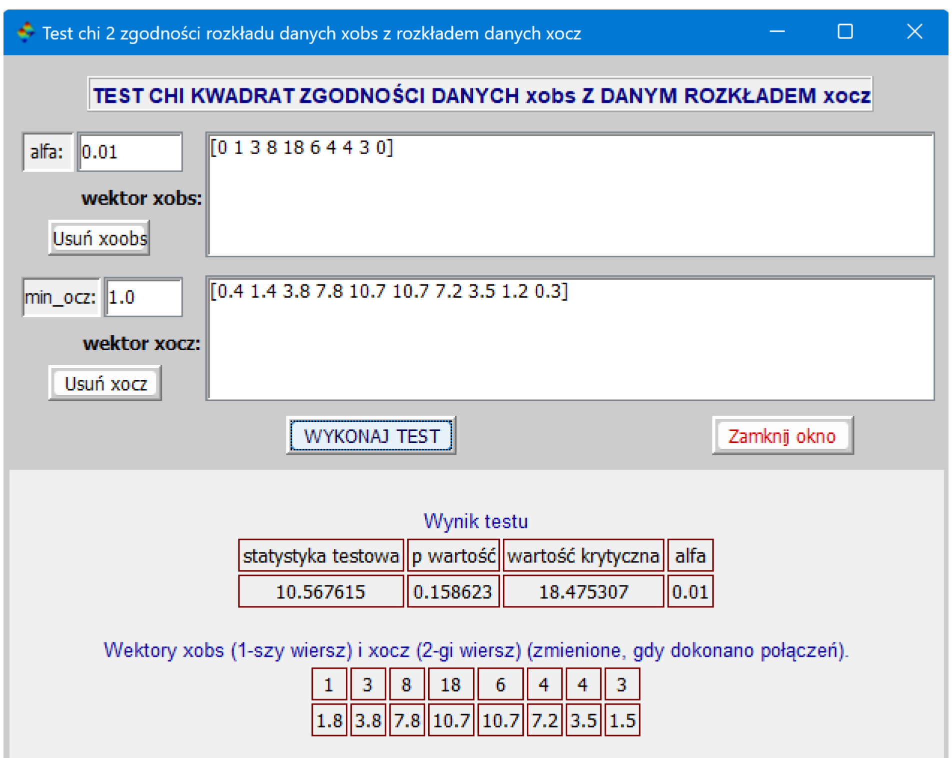
Polecenie: 'Test chi kwadrat zgodności z rozkładami ciągłymi'

1. Okno 'Test chi kwadrat zgodności z rozkładami ciągłymi' pozwala nam testować hipotezę, że zaobserwowane częstości zmiennej są zgodne z oczekiwanymi częstościami zmiennej mającej jeden z trzech rozkładów ciągłych: normalny, wykładniczy lub jednostajny. Aby wykonać jeden z tych trzech testów trzeba wpisać granice przedziałów klasowych w pole 'granice cl przedziałów klasowych' oraz dane surowe lub liczebności przedziałów klasowych w pole 'dane x'. Jeśli wpisujemy dane 'surowe', to trzeba zaznaczyć opcję 'dane x, to dane 'surowe' zamiast liczebności'. Następnie zaznaczamy jeden z trzech typów rozkładu. Jeśli zaznaczyliśmy rozkład wykładniczy, to do wykonania testu potrzebna jest jeszcze średnia a przy wyborze rozkładu normalnego oprócz średniej potrzebne jest odchylenie standardowe. Jeśli znamy te parametry, to wpisujemy je w odpowiednie pola i wybieramy przycisk **0** pod 'liczba estym. parametrów'. Jeśli nie znamy jednego (lub dwóch parametrów, gdy wybraliśmy rozkład normalny), to pozostawiamy puste pola dla tych parametrów, pod 'liczba estym. parametrów' wybieramy przycisk **1** dla rozkładu wykładniczego i przycisk **2** dla normalnego i klikamy w 'WYKONAJ TEST'. Danych pogrupowanych, tj. gdy mamy liczebności jako dane x, powinno się używać, gdy znamy parametry rozkładu ciągłego. Estymacja będzie możliwa tylko wtedy, gdy wśród granic przedziałów klasowym nie występuje żaden z rodzajów nieskończoności.

2. W oknie tym mamy jeszcze jedną opcję 'liczebn. rozkł. ciągłego równe'. Można ją zaznaczyć tylko wtedy, gdy mamy dane 'surowe'. Wtedy po wykonaniu czynności opisanych w poprzednim punkcie i kliknięciu w



Rys. 65: Okno z wynikiem testu chi kwadrat zgodności rozkładu z zadanym rozkładem



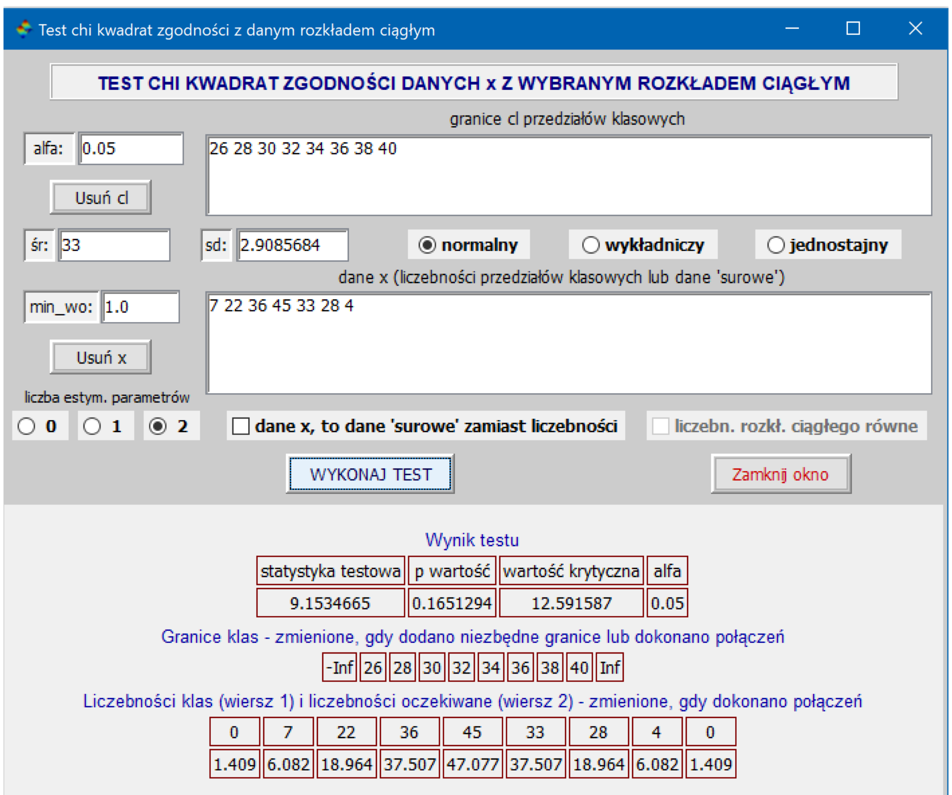
Rys. 66: Okno z testem chi kwadrat zgodności rozkładu z zadanym rozkładem, gdzie niektóre liczebności zostały połączone

’WYKONAJ TEST’ program wyznaczy nowe przedziały klasowe, dla których wartości oczekiwane liczebności obliczone dla rozkładu, który wybraliśmy do porównania, będą wszystkie jednakowe. Liczba tych przedziałów będzie taka sama jak liczba podanych wcześniej przedziałów, ale ich długości będą tak dobrane, aby wszystkie oczekiwane liczebności były równe. I dla tych nowych przedziałów wyliczy liczebności naszych danych.

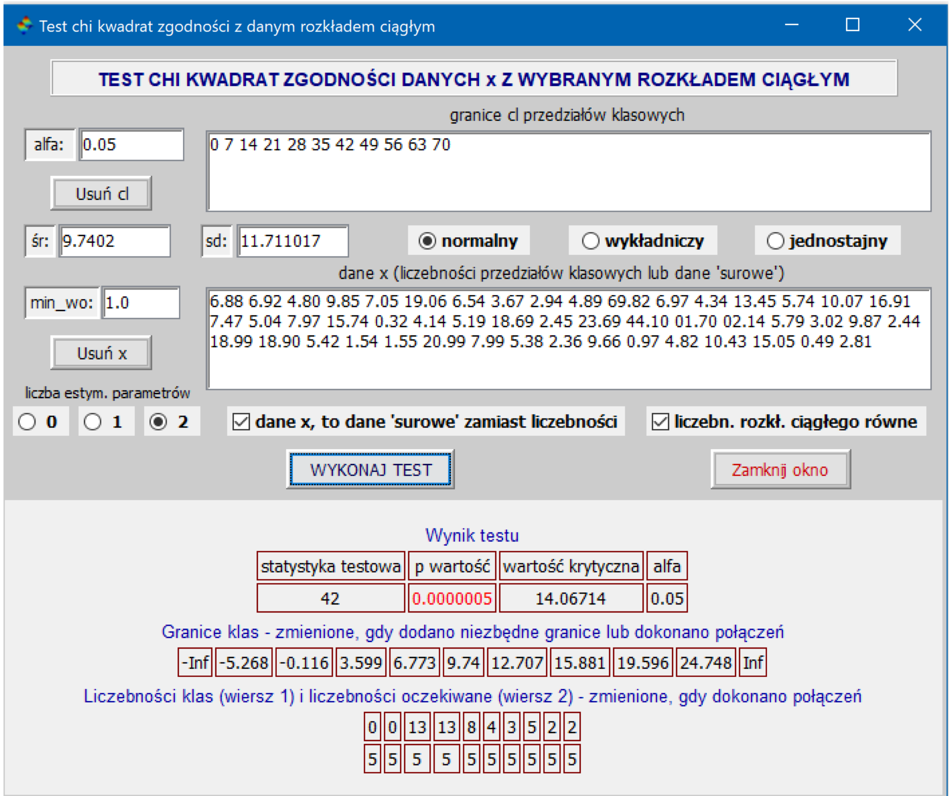
3. Kolejno na Rys. 67, 68 i 69 mamy wyniki testu zgodności danych z rozkładem normalnym dla danych przedziałów klasowych i danych liczebności w tych przedziałach, testu zgodności z rozkładem normalnym dla danych surowych i opcji ’liczebn. rozkł. ciągłego równe’ oraz dla testu zgodności z rozkładem wykładniczym i danych surowych i opcji ’liczebn. rozkł. ciągłego równe’.

Uwaga 1: Jeśli dla porównania z rozkładem normalnym lub wykładniczym wybieramy pod “liczba estym. parametrów” przycisk **0**, tj. sami mamy wpisać parametry rozkładu, to często w polach ’śr:’ i ’sd:’ zostają parametry po innych testach i wtedy łatwo zapomnieć je zmienić, Dlatego po tabelach z wynikami testu wypisany jest komunikat:
Proszę sprawdzić, że parametry śr i sd zostały poprawnie wprowadzone.

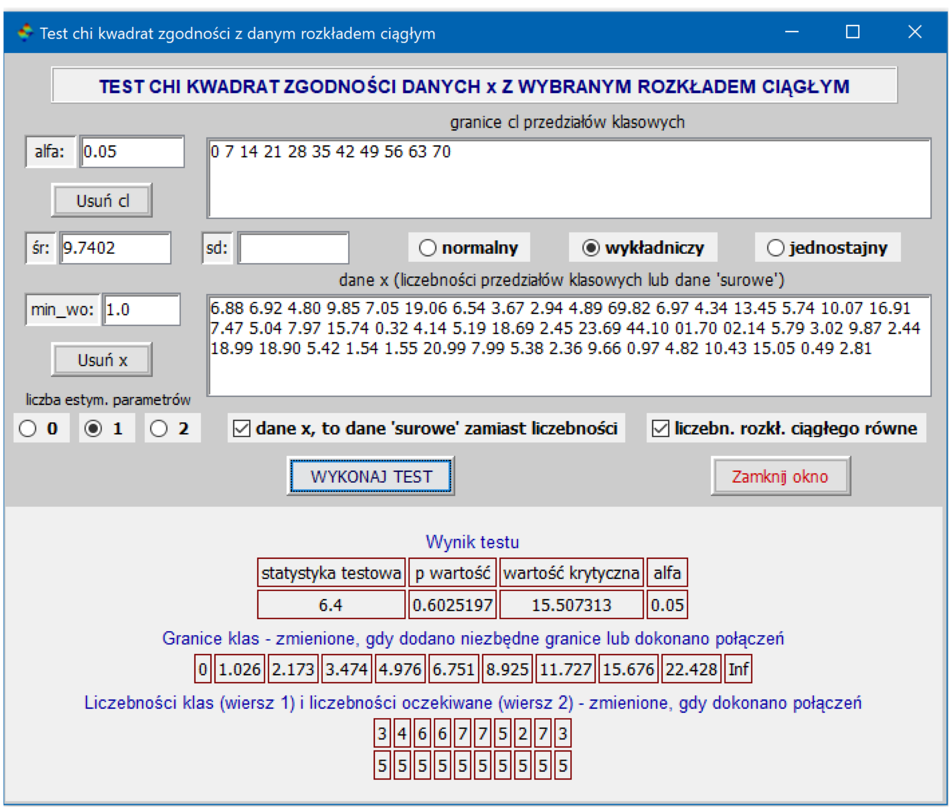
Uwaga 2: W opisanym teście dla podanych granic cl przedziałów klasowych bierze się przedziały lewostronnie domknięte. Jeśli chcemy, aby zostały użyte przedziały prawostronnie domknięte, to do podanych granic przedziałów klasowych należy dodać małą liczbę dodatnią epsilon, nie mniejszą niż 10^{-14}, np. epsilon=10^{-13}. Jednocześnie dokładnie to samo trzeba zrobić w polu ’dane x’, jeśli wpisujemy tam liczebności obliczone dla przedziałów lewostronnie domkniętych. Np. w sytuacji jak na Rys. 67, zamiast 26 28 30 32 34 36 38 40 i 7 22 36 45 33 28 4 możemy wpisać $[26\ 28\ 30\ 32\ 34\ 36\ 38\ 40] + 10^{-13}$ i $[7\ 22\ 36\ 45\ 33\ 28\ 4] + 10^{-13}$, koniecznie biorąc najpierw granice przedziałów klasowych i podane liczebności w nawiasy kwadratowe.



Rys. 67: Okno z wynikiem testu chi kwadrat zgodności danych z rozkładem normalnym



Rys. 68: Okno z wynikiem testu chi kwadrat zgodności danych z rozkładem normalnym



Rys. 69: Okno z wynikiem testu chi kwadrat zgodności danych z rozkładem wykładniczym

Polecenie: ’Test chi kwadrat zgodności z rozkładami dyskretnymi’

Użycie tego okna wymaga dodatkowych wyjaśnień. Dlatego nasz opis zaczniemy od przykładów.

Przykład użycia danych dla testu zgodności z rozkładem Poissona

I. Poniżej, w pierwszym wierszu podane są liczby dziennych wezwań do wypadków karetki pogotowia ratunkowego w pewnym mieście, a w drugim liczby wskazujące ile było dni w ciągu 3 miesięcy (90 dni) z taką liczbą wezwań.

0	1	2	3	4	5	6	7	8	9	10
0	0	5	14	15	23	16	9	3	3	2

Chcemy przetestować hipotezę H_0, że wezwania karetki pogotowia w tym mieście mają rozkład Poissona.

Uwaga 1: w pole dane x okna z testem wpisujemy tylko drugi wiersz.

Pierwszy wiersz tylko pomaga nam poprawnie wpisać drugi wiersz, podając jako jego elementy występujące na początku liczebności równe 0 odpowiadające liczbom zdarzeń 0,1,. . . (o ile takie występują (tutaj zaznaczone kolorem czerwonym), nawet wtedy gdy w podawanych danych są one pominięte.

Przykłady użycia danych dla testu zgodności z rozkładem dwumianowym

II. W pewnej małej i zamkniętej społeczności zbadano rozkład liczby synów w rodzinach z siedmiorgiem dzieci. Dane zawarte są niżej w tabelce. Pierwszy wiersz, to liczba synów a drugi, to częstość. Chcemy przetestować hipotezę H_0, że liczba synów w rodzinie z siedmiorgiem dzieci ma rozkład dwumianowy z $p = 0.5$.

0	1	2	3	4	5	6	7
0	6	14	25	21	22	9	1

III. Jedna z części karty pacjenta w pewnym szpitalu ma 12 rubryk. Z archiwum wzięto losowo 100 kart z taką częścią, aby sprawdzić w nich liczbę błędnie wypełnionych rubryk. Tabelka poniżej zawiera w pierwszym wierszu liczby błędnie wypełnionych rubryk a drugi wiersz liczbę kart z podaną w pierwszym wierszu liczbą błędnie wypełnionych rubryk. Chcemy przetestować czy rozkład błędów w wypełnianiu rubryk pasuje do rozkładu dwumianowego.

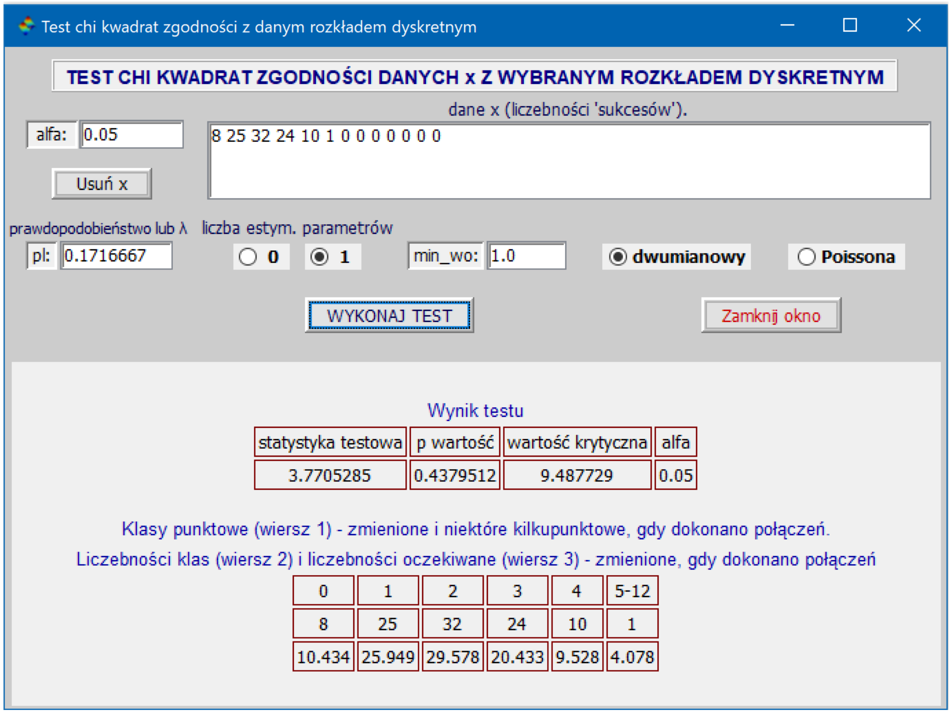
0	1	2	3	4	5	6	7	8	9	10	11	12
8	25	32	24	10	1	0	0	0	0	0	0	0

Uwaga 2: w pole dane x okna z testem wpisujemy tylko drugi wiersz.

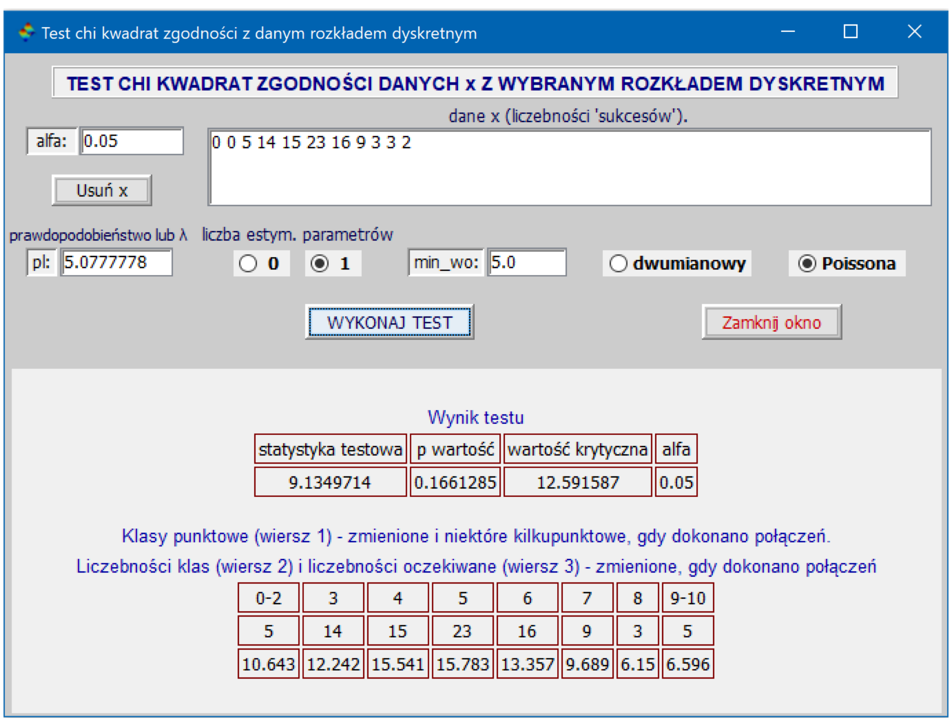
Pierwszy wiersz tylko pomaga nam poprawnie wpisać drugi wiersz, podając jako jego elementy występujące zarówno na początku jak i na końcu liczebności równe 0 odpowiadające liczbom zdarzeń z zakresu 0,1,. . . n (o ile takie występują, tutaj zaznaczone kolorem czerwonym). W podawanych danych takie wartości są często pomijane. Powtórzmy, aby poprawnie wykonać test, musimy w danych x podać jego wszystkie elementy, również te o liczebnościach równych 0. Łącznie podajemy $n + 1$ liczb, ponieważ numerujemy je od zera, gdzie n jest liczbą niezależnych prób w schemacie Bernoulliego.

1. Okno ’Test chi kwadrat zgodności z rozkładami dyskretnymi’ pozwala nam testować hipotezę, że zaobserwowane częstości zmiennej są zgodne z oczekiwanymi częstościami zmiennej mającej jeden z dwóch rozkładów dyskretnych: dwumianowy lub Poissona. Aby wykonać jeden z tych testów trzeba wpisać liczebności ’sukcesów’ w okienko ‘dane x (lub zdarzeń w przypadku rozkładu Poissona) zgodnie z Uwagą 1 lub Uwagą 2. Następnie zaznaczamy jeden z dwóch typów rozkładu. Dla każdego z obu typów rozkładów do wykonania testu potrzebny jest jeden parametr: prawdopodobieństwo dla rozkładu dwumianowego i średnia dla rozkładu Poissona. Jeśli znamy ten parametr, to wpisujemy go w polu ’pl:’ (p - prawdopodobieństwo w rozkładzie dwumianowym, l - lambda w rozkładzie Poissona) i zaznaczamy opcję **0** w polu ’Liczba estym. parametrów’. Jeśli go nie znamy, to pozostawiamy puste pole dla tego parametru i klikając w ’WYKONAJ TEST’, program wyestymuje ten parametr na podstawie danych x, zaznaczy odpowiednią liczbę estymowanych parametrów oraz wyświetli wynik testu.

2. W oknie tym mamy też (jak w poprzednich oknach) poziom istotności testu alfa oraz liczbę ograniczającą wielkość wartości oczekiwanych, którą możemy zmieniać w zakresie od 0.25 do 5. Kolejno na Rys. 70 i Rys. 71 mamy wyniki testu zgodności danych z wyżej podanego przykładu III z rozkładem dwumianowym oraz dla testu zgodności danych z podanego na początku przykładu I z rozkładem Poissona.



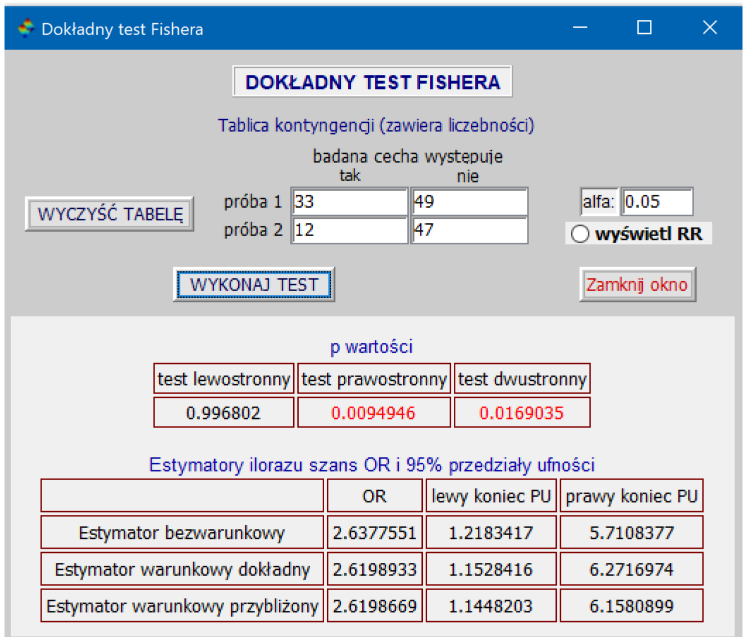
Rys. 70: Okno z wynikiem testu chi kwadrat zgodności danych z rozkładem dwumianowym



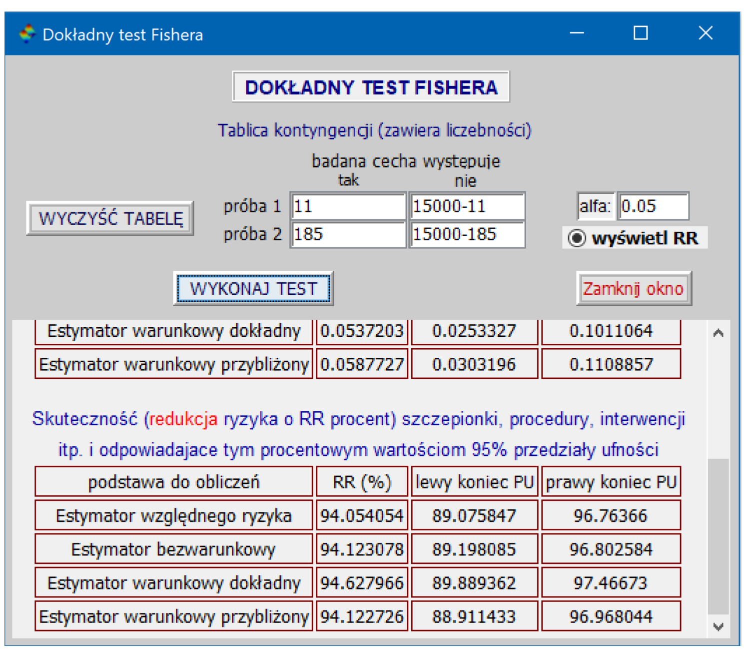
Rys. 71: Okno z wynikiem testu chi kwadrat zgodności danych z rozkładem Poissona

Polecenie: 'Test dokładny Fishera dla tablicy kontyngencji 2×2'

Po otwarciu okna wpisujemy 4 liczebności w odpowiednie pola. Możemy zmienić poziom istotności alfa oraz zaznaczyć opcję RR. RR to (R)edukcja (R)yzyka równa 1 minus względne ryzyko, gdzie względne ryzyko (angielski skrót RR) to iloraz prawdopodobieństwa wystąpienia danego skutku w grupie eksperymentalnej, w której zastosowano określoną interwencję, i tego prawdopodobieństwa w grupie kontrolnej. Klikamy w 'WYKONAJ TEST', otrzymując wynik testu. Np., gdy w odpowiednie pola wpiszemy dane 33, 49, 12, 47, to otrzymamy wynik taki, jak na Rys. 72. Na Rys. 73 widzimy obliczoną skuteczność szczepionki Moderna, gdy z grupy 15000 zaszczepionych zachorowało na Covid-19 11 osób a z grupy kontrolnej 15000 osób, którym podawano placebo, zachorowało na Covid-19 185 osób.



Rys. 72: Okno z wynikiem dokładnego testu Fishera



Rys. 73: Okno z wynikiem skuteczności szczepionki Moderna (94.054054%) - dane z 30.11.2020

Uwaga

Obliczanie warunkowego dokładnego estymatora ilorazu szans i odpowiadającego mu przedziału ufności dla większych elementów tablicy `x` jest zadaniem trudnym i wymaga użycia przybliżeń logarytmów silni liczb naturalnych, skalowań pośrednich wyników i rozwiązywania nieliniowych równań. Różne programy statystyczne różnie sobie z tym radzą a czasami i nie radzą. Np. dla `x=[3 1;1 3]` program R zamiast poprawnej wartości prawego końca 95% przedziału ufności równego 626.24363 (można "ręcznie" wypisać równanie i rozwiązać, używając np. programu Mathematica) daje niepoprawną wartość 621.93375. Podobnie, wątpliwości budzą prawe końce 95% przedziału ufności w R dla `x=[1029 1;1 616]` i `x=[400 1;150 400]`. Dlatego jest naturalnym, że wyniki otrzymane przez różne programy mogą się różnić. Brak dokładnego estymatora ilorazu szans i odpowiadającego mu przedziału ufności w argumencie wyjściowym funkcji `fisherexact_test` oznacza, że funkcja ta nie poradziła sobie z zadaniem ich wyznaczenia.

Polecenie: 'Test Kruskala-Willisa'

Po kliknięciu w polecenie 'Test Kruskala-Willisa' otwiera nam się okno, w którym możemy otworzyć plik z rozszerzeniem .csv z zapisanymi w nim danymi. Ale zanim otworzymy plik najpierw musimy zaznaczyć odpowiednie opcje powyżej przycisku 'Wybierz plik .csv z danymi' (o opcjach tych piszemy niżej). Dopuszcza się zapisywanie danych na 2 sposoby:

1. w dwóch kolumnach, gdzie jedna kolumna reprezentuje zmienną grupującą, a druga odpowiadające jej wartości liczbowe. Pozwala to użyć plików z danymi przygotowanymi dla jednoczynnikowej Anovy. Bowiem test Kruskala-Wallisa jest nieparametrycznym odpowiednikiem jednoczynnikowej Anovy, która jest testem parametrycznym. Dane takie mogą np. wyglądać następująco:

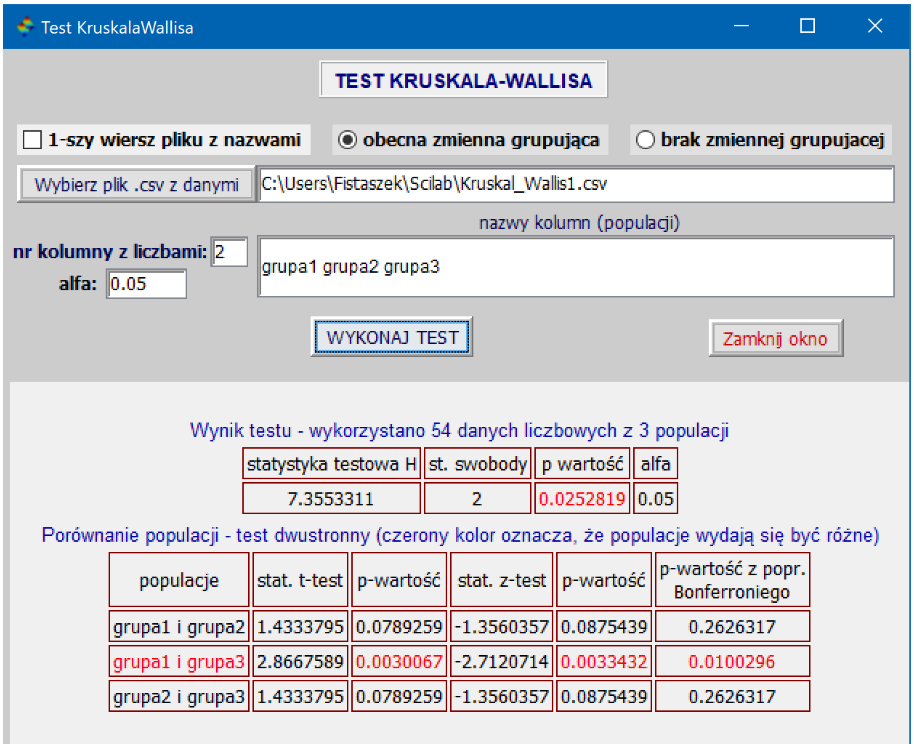
   ```
   grupa1,3
   grupa2,2
   grupa4,5
   grupa3,4
   ...
   ```

   Wiersze w takim pliku mogą być uporządkowane, np. najpierw dane z grupa1, potem z grupa2, itd., ale nie jest to konieczne. Pierwszy wiersz w takim pliku może zawierać nazwy kolumn, ale nie jest to konieczne. Jeśli taki wiersz występuje, to trzeba wybrać opcję '1-szy wiersz pliku z nazwami'. Dla tego typu pliku trzeba też wybrać opcję 'obecna zmienna grupująca'. Po otwarciu takiego pliku, w polu 'nr kolumny z liczbami' zostanie automatycznie wpisany numer kolumny, w której znajdują się wartości liczbowe. Jeśli jest on niepoprawny, to należy go 'ręcznie' zmienić. W polu 'nazwy kolumn (populacji)' zostaną wpisane wszystkie nazwy grup. Jeśli chcemy użyć do testowania tylko niektóre grupy, to trzeba edytować to pole, czyli ręcznie usunąć niepotrzebne nazwy. Następnie klikamy w przycisk 'WYKONAJ TEST' i otrzymujemy wyniki jak na Rys. 74.

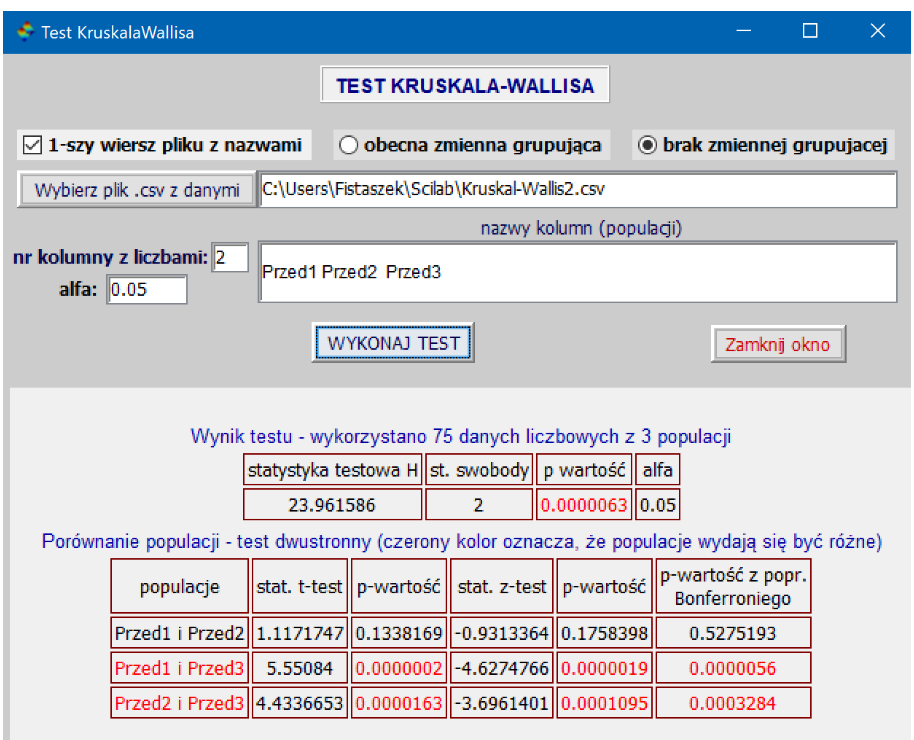
2. Druga akceptowalna postać danych w pliku .csv, to wiersz z nazwami kolumn (obowiązkowy) i kolumny zawierające liczby. Dane takie mogą np. wyglądać następująco:

   ```
   Przed1,Po1,Przed2,Po2,Przed3,Po3
   350,136,292,51,213,11
   142,62,302,42,296,13
   190,106,261,49,312,9
   ...
   ```

   Podobnie jak w opisie w pkt. 1, przed otwarciem pliku z danymi najpierw trzeba zaznaczyć odpowiednie opcje umieszczone powyżej przycisku 'Wybierz plik .csv z danymi'. Dla pliku z takimi danymi musi być zaznaczona opcja '1-szy wiersz pliku z nazwami' i opcja 'brak zmiennej grupującej'. Po otwarciu takiego pliku, w polu 'nr kolumny z liczbami' zostanie automatycznie wpisany numer 2 i nie należy go zmieniać. W polu 'nazwy kolumn (populacji)' zostaną wpisane wszystkie nazwy kolumn. Jeśli chcemy użyć do testowania tylko niektóre kolumny, to trzeba edytować to pole, czyli ręcznie usunąć niepotrzebne nazwy. Po usunięciu nazw 'Po1', 'Po2' i 'Po3' i kliknięciu w przycisk 'WYKONAJ TEST' otrzymamy wyniki jak na Rys. 75.



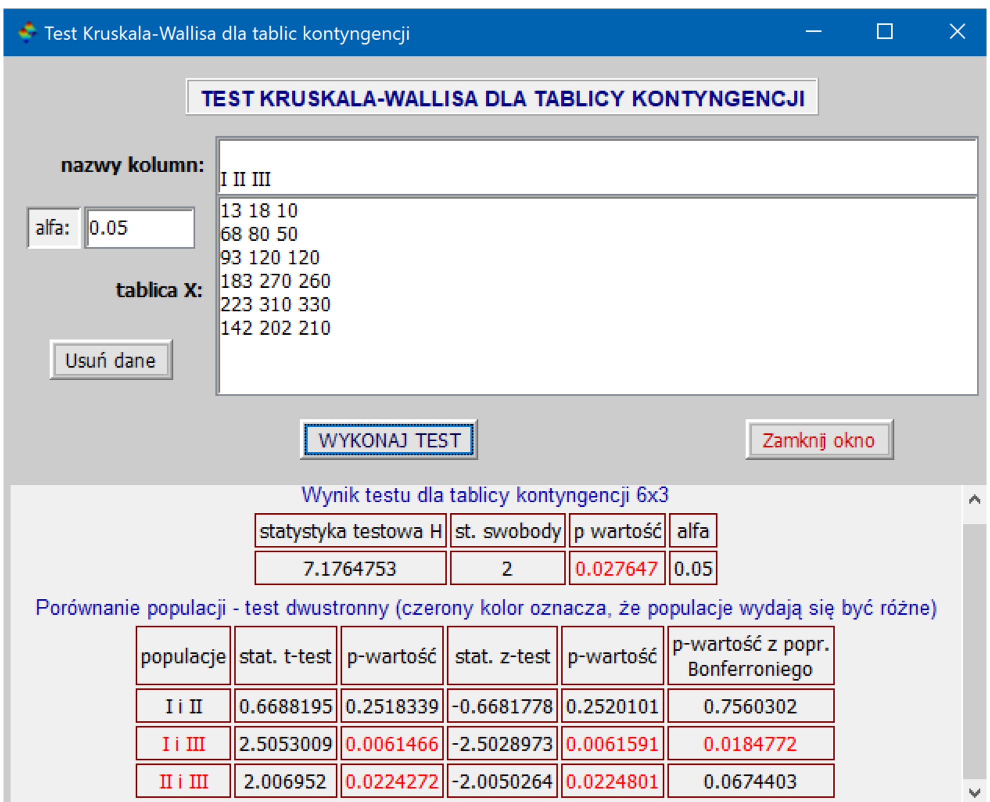
Rys. 74: Wyniki obliczeń dla testu Kruskala-Wallisa dla danych w pliku Kruskal-Wallis1.csv



Rys. 75: Wyniki obliczeń dla testu Kruskala-Wallisa dla danych w pliku Kruskal-Wallis1.csv

Polecenie: 'Test Kruskala-Willisa dla tablic kontyngencji'

Klikając w polecenie 'Test Kruskala-Willisa dla tablic kontyngencji' otwieramy okno, w którym testujemy czy badana cecha w populacjach, z których wzięto dane w kolumnach tablicy, ma jednakowy rozkład - dokładniej, zakłada się, że cecha ta ma jednakowy rozkład, tylko mediany (lub średnie) mogą się różnić i właśnie testuje się ich równość. Test ten można stosować, jeśli wiersze są uporządkowane, tj. istnieje naturalny kierunek, wg którego możemy je uporządkować. Np. każdy wiersz zawiera liczebności tej samej oceny w różnych grupach. Wiemy też, która ocena jest lepsza a która gorsza. Możemy zatem uporządkować wiersze, wypisując je w kolejności od tych zawierających liczebności dla oceny najgorszej do tych zawierających liczebności dla oceny najwyższej (lub w odwrotnej kolejności). Jest to ważne, bo sumuje się wszystkie liczebności w danym wierszu i tyle nadaje się jednakowych rang. Na przykład, jak suma liczebności w danym wierszu wynosi 100, to temu wierszowi odpowiada 100 rang, każda równa 50.5 (średnia kolejnych rang od 1 do 100). A gdy suma liczebności w następnym wierszu wynosi 120, to jemu odpowiada 120 rang, każda równa 160.5 (100 plus średnia kolejnych rang od 1 do 120). I tak dalej. W takim przypadku łatwo też wylicza się poprawki na obecność jednakowych rang (ties). Jako przykład weźmy wagi noworodków w trzech dużych regionach pewnego (rozwijającego się) kraju. Weźmiemy wagi tylko tych noworodków, które po urodzeniu uzyskały co najwyżej 3 punkty w skali Adgara. Wagi te podzielimy na 6 grup (kategorii): 500-999 gramów, 1000-1499 gramów, . . . , 2500-2999 gramów, 3000 gramów lub więcej. Wpisujemy liczebności dla poszczególnych 3 grup (3 regionów) w polu 'tablica X' - reguły wpisywania liczebności (tj. tablic) w tym polu są dokładnie takie same jak wcześniej opisane reguły wpisywania tablic (macierzy) w odpowiednich polach dla testu chi kwadrat dla tablic kontyngencji na str. 55. Nazwy regionów możemy wpisać w polu 'nazwy kolumn' lub nie. Gdy nie wpiszemy, to automatycznie nadane zostaną im jako nazwy kolejne numery w rzymskiej numeracji. Następnie klikamy w przycisk 'WYKONAJ TEST' i otrzymujemy wyniki jak na Rys. 76.



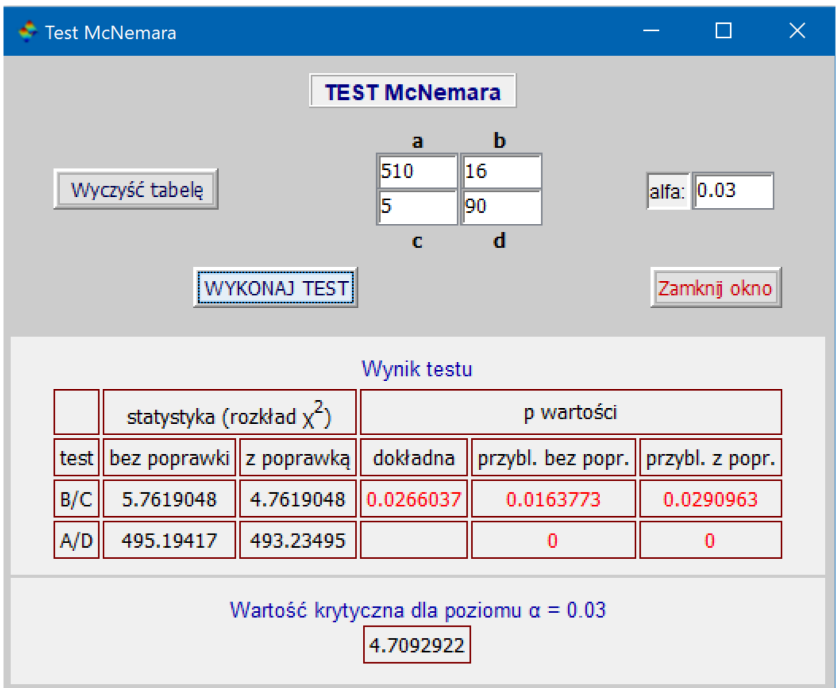
Rys. 76: Wyniki obliczeń dla testu Kruskala-Wallisa dla wpisanej w polu 'tablica X' tablicy liczebności

Polecenie: 'Test McNemara dla tablicy 2×2 danych sparowanych'

Po kliknięciu w polecenie 'Test McNemara dla tablicy 2×2 danych sparowanych' i otwarciu okna wpisujemy 4 liczebności w odpowiednie pola i po ewentualnej zmianie poziomu istotności alfa klikamy w 'WYKONAJ TEST', otrzymując wynik testu. Chcemy np. przetestować na poziomie alfa = 0.03 hipotezę, że obie metody leczenia mają te samą skuteczność.

	Pacjenci leczeni metodą II	
Pacjenci leczeni metodą I	przeżyli 5 lat	zmarli w ciągu 5 lat
przeżyli 5 lat	510	16
zmarli w ciągu 5 lat	5	90

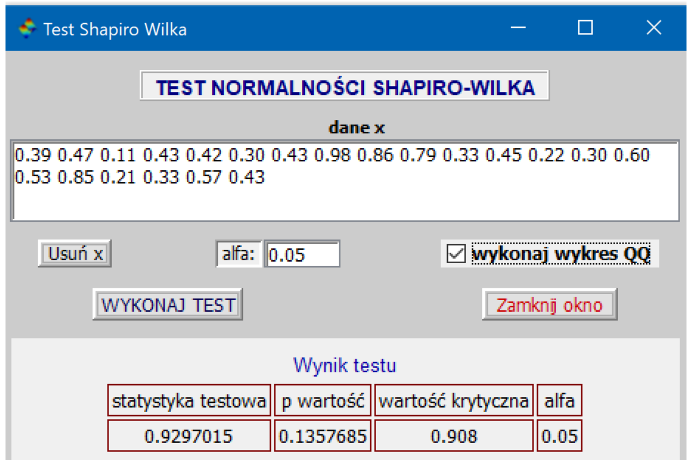
Wpisując 4 liczebności z powyższej tabeli w odpowiednie pola otrzymamy wynik jak na Rys. 77.



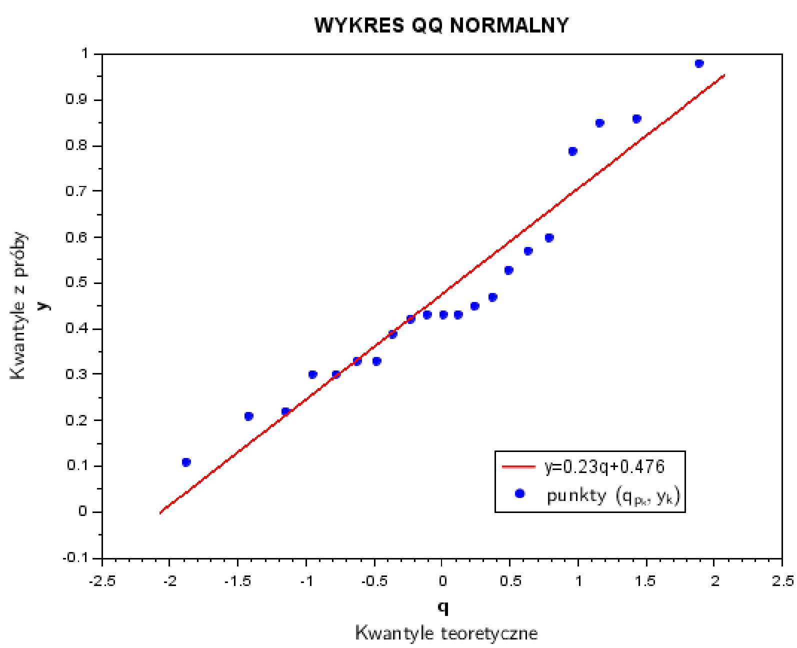
Rys. 77: Okno z wynikiem testu McNemara dla danych w powyższej tabelce

Polecenie: 'Test normalności Shapiro-Wilka'

Aby wykonać test normalności Shapiro-Wilka dla próby x, której elementy są typu ilościowego, należy wpisać elementy wektorów danych x w okienko 'dane x', zmienić wartość poziomu istotności alfa, jeśli ma być inny niż 0.05, zaznaczyć opcję 'wykonaj wykres QQ', jeśli chcemy uzyskać taki wykres i kliknąć w 'WYKONAJ TEST'. Po wykonaniu testu, np. dla danych x=[0.39 0.47 0.11 0.43 0.42 0.30 0.43 0.98 0.86 0.79 0.33 0.45 0.22 0.30 0.60 0.53 0.85 0.21 0.33 0.57 0.43] z zaznaczoną opcją 'Wykonaj wykres QQ', otrzymamy wyniki jak na Rys. 78 i QQ wykres jak na Rys. 79.



Rys. 78: Okno z wynikami testu normalności Shapiro-Wilka dla wyżej wypisanych danych



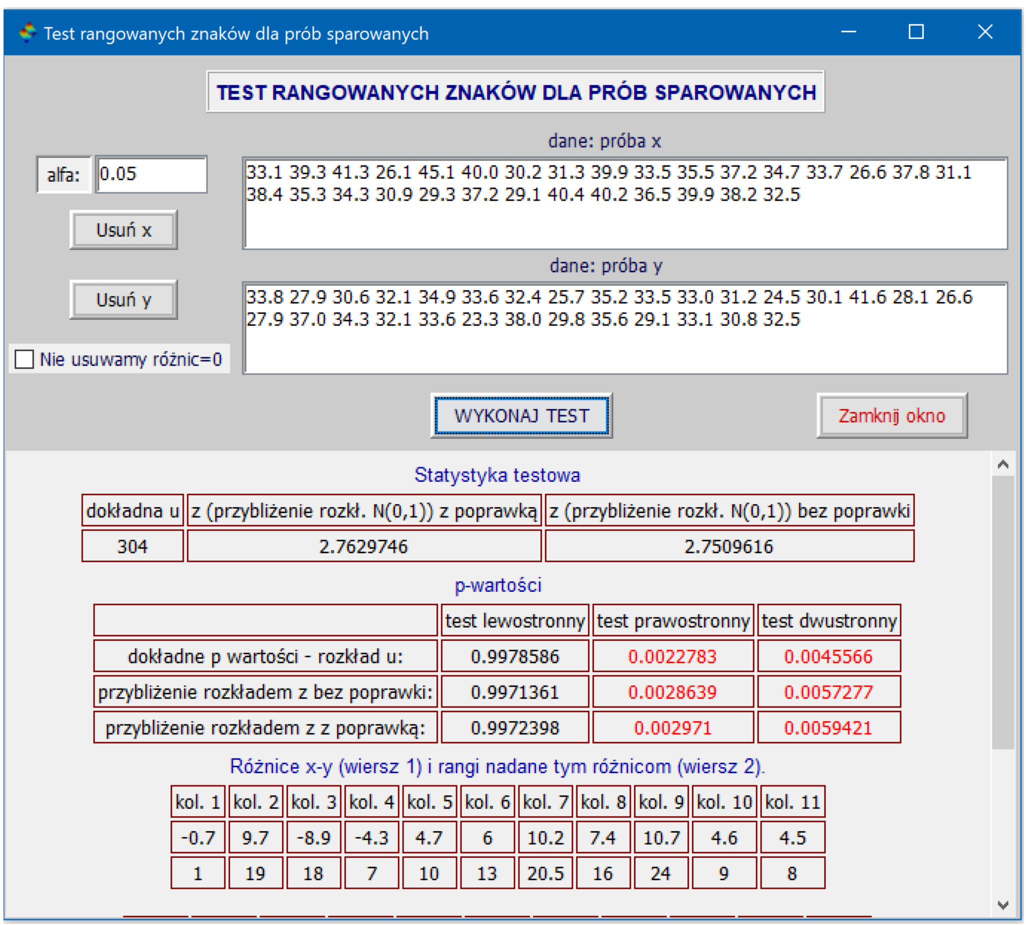
Rys. 79: QQ wykres dla wyżej wypisanych danych

Polecenie: 'Test rangowanych znaków dla prób sparowanych'

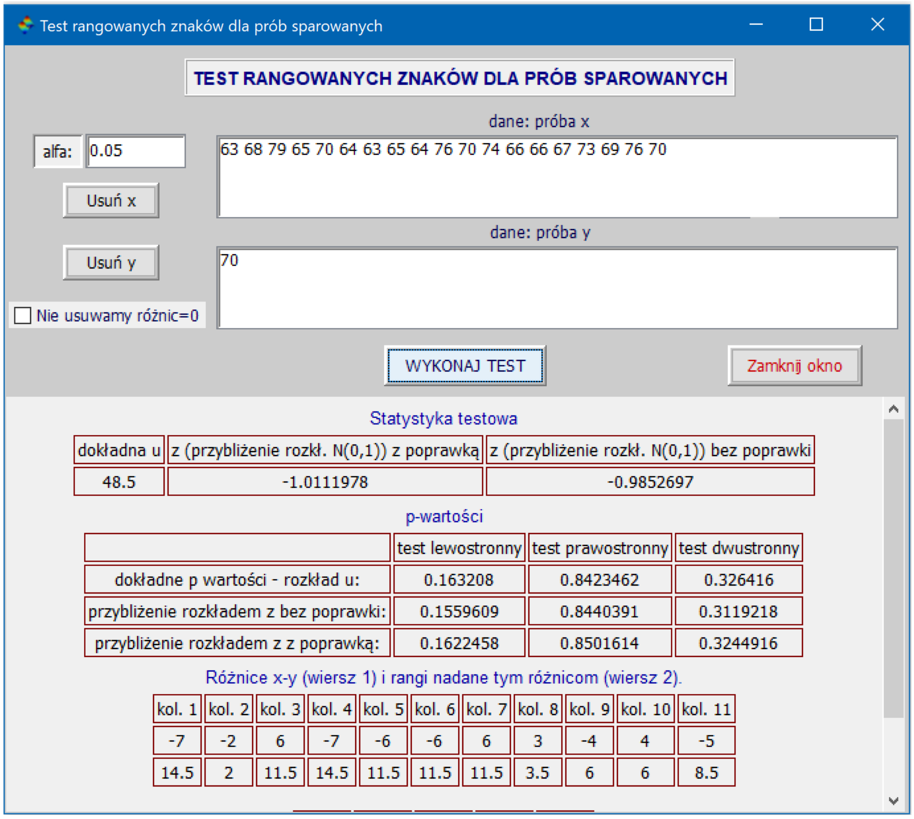
Po kliknięciu w wyżej wymienione polecenie i otwarciu okna wpisujemy w okienko 'dane: próba x' i okienko 'dane: próba y' odpowiednio równej długości wektory x i y o elementach liczbowych. Jeśli chcemy porównać położenie elementów liczbowych wektora x z zadaną wartością mediany, to w okienku 'dane: próba y' wpisujemy wartość tej mediany. Możemy jeszcze zaznaczyć opcję 'nie usuwamy różnic=0' - skutkuje to niemożnością wykonania dokładnego testu, jeśli wśród obliczonych różnic będą zera. Po ewentualnej zmianie poziomu istotności alfa klikamy w 'WYKONAJ TEST', otrzymując wynik testu. Np. dla danych x=[33.1 39.3 41.3 26.1 45.1 40.0 30.2 31.3 39.9 33.5 35.5 37.2 34.7 33.7 26.6 37.8 31.1 38.4 35.3 34.3 30.9 29.3 37.2 29.1 40.4 40.2 36.5 39.9 38.2 32.5], y=[33.8 27.9 30.6 32.1 34.9 33.6 32.4 25.7 35.2 33.5 33.0 31.2 24.5 30.1 41.6 28.1 26.6 27.9 37.0 34.3 32.1 33.6 23.3 38.0 29.8 35.6 29.1 33.1 30.8 32.5] otrzymamy wynik taki, jak na Rys. 80. A dla danych x=[63 68 79 65 70 64 63 65 64 76 70 74 66 66 67 73 69 76 70] i mediany równej 70 (wystąpią 3 zerowe różnice) otrzymamy wynik taki jak na Rys. 81, gdy nie zaznaczymy opcji 'nie usuwamy różnic=0' i wynik jak na Rys. 82, gdy tę opcję zaznaczymy.

Uwaga:

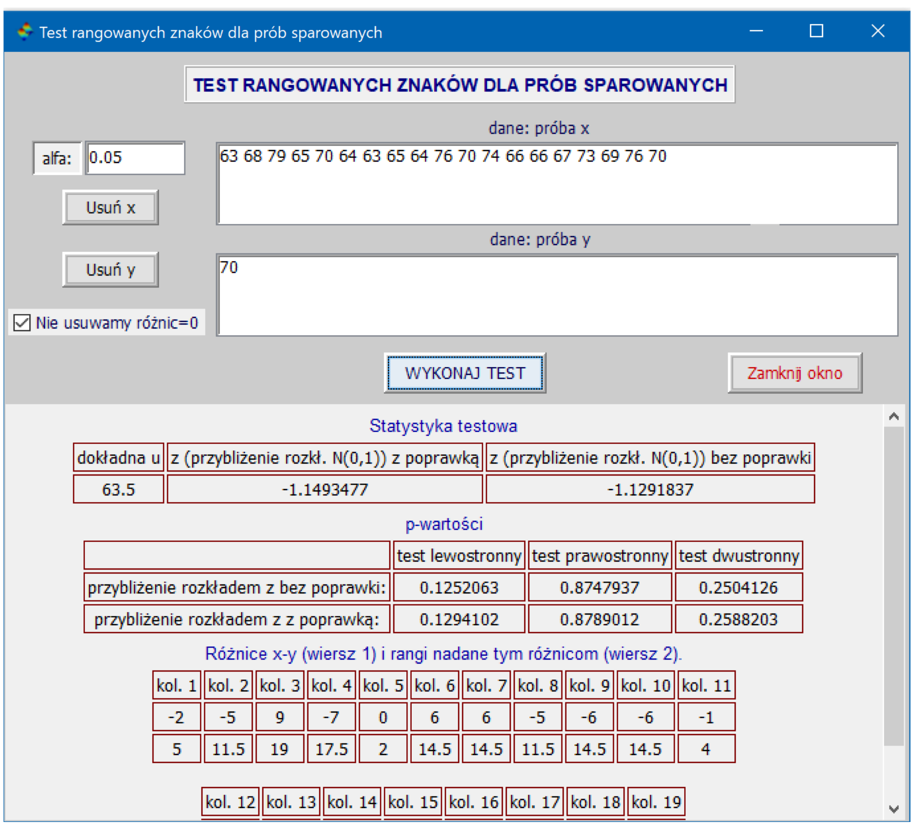
Test dokładny jest wykonywany, gdy zostały usunięte zera (lub gdy ich nie było) i liczebność próby n spełnia nierówność: $n \leq 80$. Dzięki wykorzystaniu tzw. 'shift' algorytmu jest on wykonywany nawet wtedy, gdy występują powtarzające się rangi (ang. ties). Gdy w różnicach x-y występują zera i ich nie usuwamy, to podajemy wartość statystyki dla testu dokładnego ale test taki nie jest wykonywany i w wynikach brak jest dla niego p-wartości.



Rys. 80: Okno z wynikami testu rangowanych znaków dla podanych danych w okienkach 'dane: próba x' i 'dane: próba y'



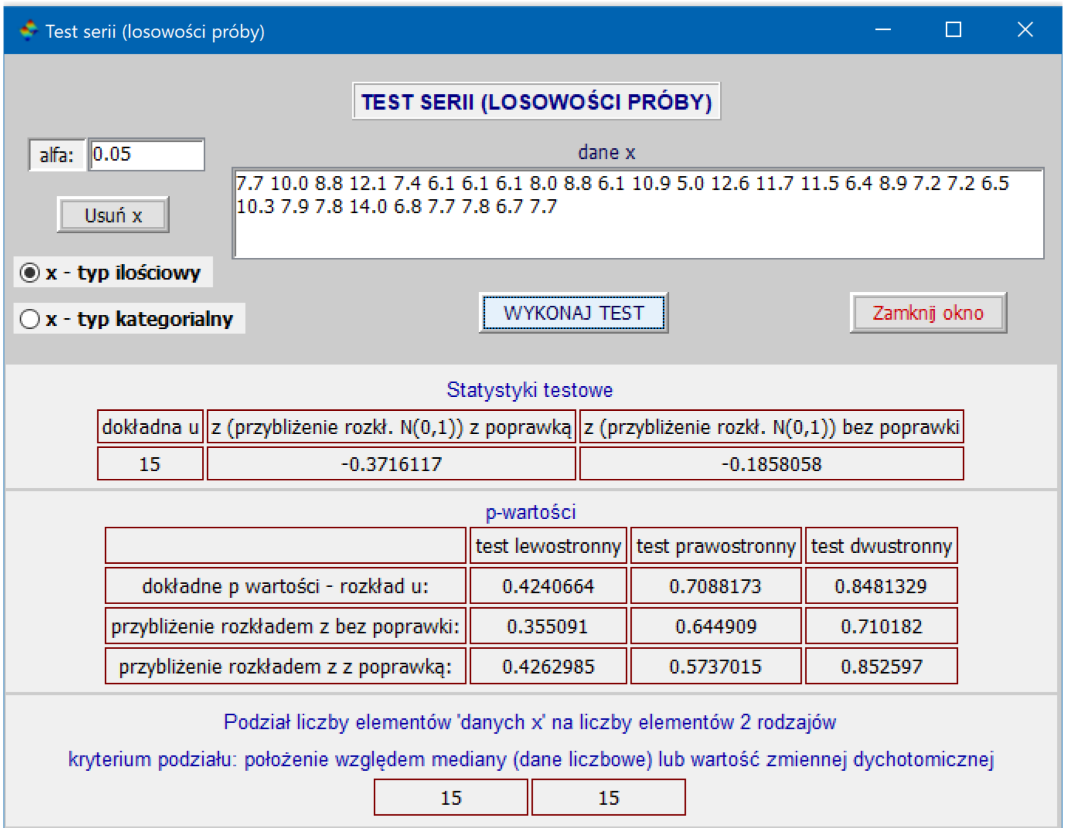
Rys. 81: Okno z wynikami testu rangowanych znaków bez zaznaczonej opcji 'nie usuwamy różnic=0'



Rys. 82: Okno z wynikami testu rangowanych znaków z zaznaczoną opcją 'nie usuwamy różnic=0'

Polecenie: 'Test serii Walda-Wolfowitza losowości próby'

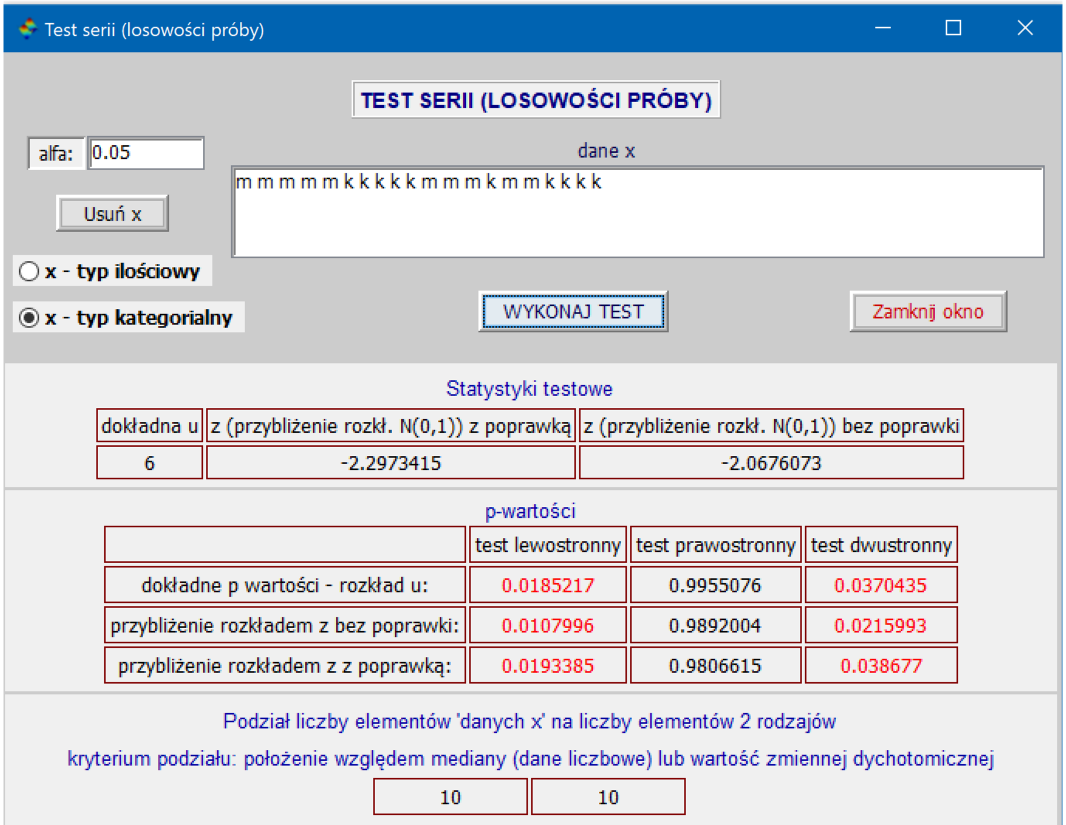
Po kliknięciu w polecenie 'Test serii Walda-Wolfowitza losowości próby' i otwarciu okna wpisujemy w okienko 'dane x' wektor danych x i zaznaczamy opcję 'x - typ ilościowy', gdy x jest wektorem danych liczbowych lub opcję 'x - typ kategorialny', gdy wektor danych jest wektorem o elementach 'łańcuchowych'. I choć elementy łańcuchowe są ciągami znaków wziętymi w cudzysłów lub apostrofy, to wpisując je nie musimy ich brać w cudzysłów, ponieważ program po ich wczytaniu sam to zrobi, jeśli zaznaczymy opcję 'x - typ kategorialny'. Po ewentualnej zmianie poziomu istotności alfa klikamy w 'WYKONAJ TEST', otrzymując wynik testu. Przykładowe wyniki testu dla danych typu ilościowego i danych typu kategorialnego przedstawiają odpowiednio Rys. 83 i Rys. 84.



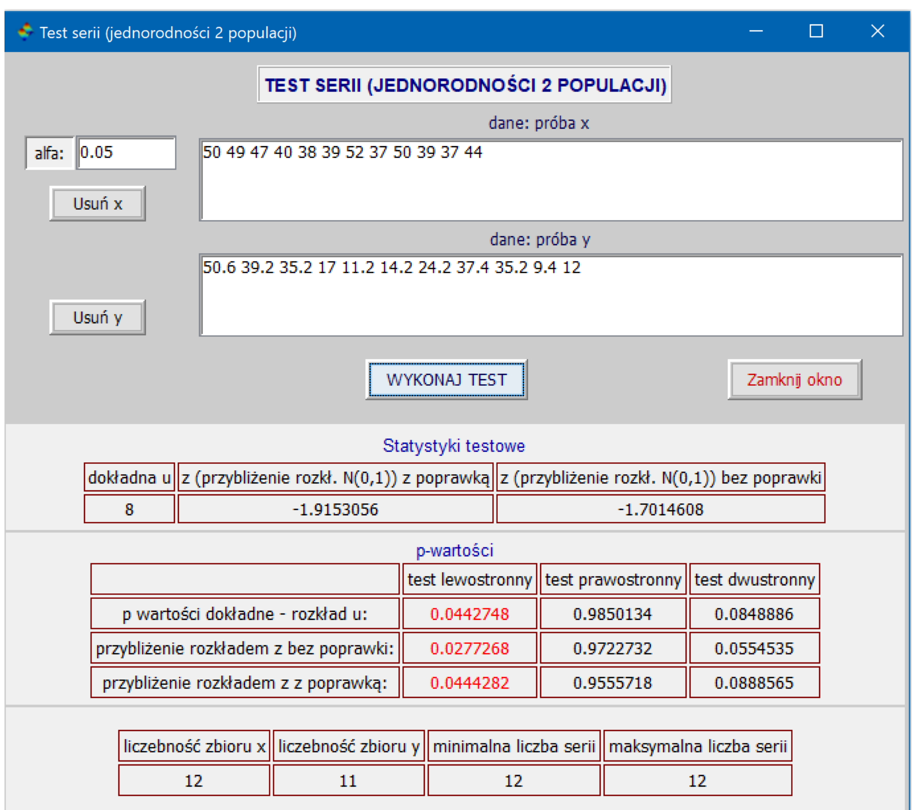
Rys. 83: Okno z wynikami testu serii Walda-Wolfowitza losowości próby dla danych typu ilościowego

Polecenie: 'Test serii Walda-Wolfowitza jednorodności dwóch populacji'

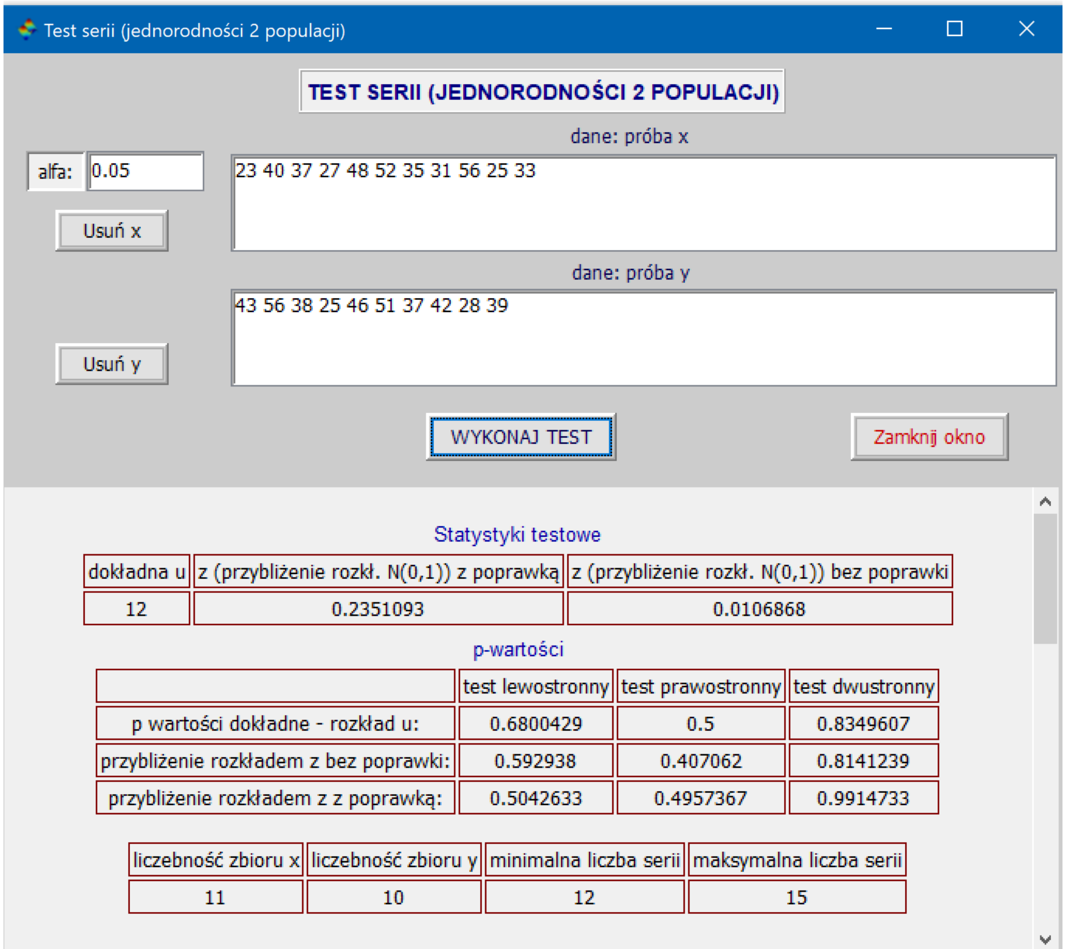
Po kliknięciu w polecenie 'Test serii Walda-Wolfowitza jednorodności dwóch populacji' i otwarciu okna wpisujemy odpowiednio w okienko 'dane: próba x' i 'dane: próba y' wektory danych x i danych y o elementach liczbowych i po ewentualnej zmianie poziomu istotności alfa klikamy w 'WYKONAJ TEST', otrzymując wynik testu. Jeśli wśród danych obu prób znajdują się dane o takich samych wartościach, to porządkując połączony ciąg, wspólną wartość możemy wziąć najpierw z próby x a potem z próby y. Ale możemy to zrobić w odwrotnej kolejności. Może to prowadzić do różnych wartości statystyk, a w konsekwencji do braku konkluzji, gdy jedna statystyka daje podstawę do odrzucenia hipotezy H_0 a druga nie daje takiej podstawy. Na Rys. 85 mamy wynik testu dla danych, gdzie obie próby nie maja wspólnych elementów. Natomiast na Rys. 86 mamy wynik dla danych, które mają 3 wspólne elementy {25,37,56}, co daje 4 różne liczby serii (dlatego na dole okna wyników mamy różne wartości dla minimalnej i maksymalnej liczby serii, które to we wcześniejszym przykładzie były jednakowe). Taka sytuacja skutkuje 4 różnymi wartościami statystyki. Wszystkie one jednak nie dają podstaw do odrzucenia hipotezy zerowej, czyli prowadzą do jednoznacznej konkluzji.



Rys. 84: Okno z wynikami testu serii Walda-Wolfowitza losowości próby dla danych typu kategorialnego



Rys. 85: Okno z wynikami testu serii Walda-Wolfowitza jednorodności dwóch populacji



Rys. 86: Okno z wynikami testu serii Walda-Wolfowitza jednorodności dwóch populacji z wieloma wartościami statystyki

Polecenie: 'Test U Manna-Whitneya'

Aby wykonać test równości median w próbach, których elementy są typu ilościowego, należy wpisać elementy wektorów danych x i y odpowiednio w okienko 'dane x' i okienko 'dane y', zmienić wartość poziomu istotności alfa, jeśli ma być inny niż 0.05 i kliknąć w 'WYKONAJ TEST'. Gdy w rangach występują powtarzające się wartości, to w polu 'brak powt. rang' program wstawi wartość %F (FAŁSZ) a w polu 'tadj' odpowiednią wartość służącą do obliczenia poprawki na występowanie powtarzających się wartości (ang. ties), którą uwzględnia się podczas obliczania statystyki testowej (zob. poniższą uwagę). Rys. 87 pokazuje przykładowy wynik testu. Jednym z wyników jest tabela z rangami. W jednej z jej kolumn pojawia się kolor niebieski użyty dla zaznaczenia pierwszego elementu wektora y.

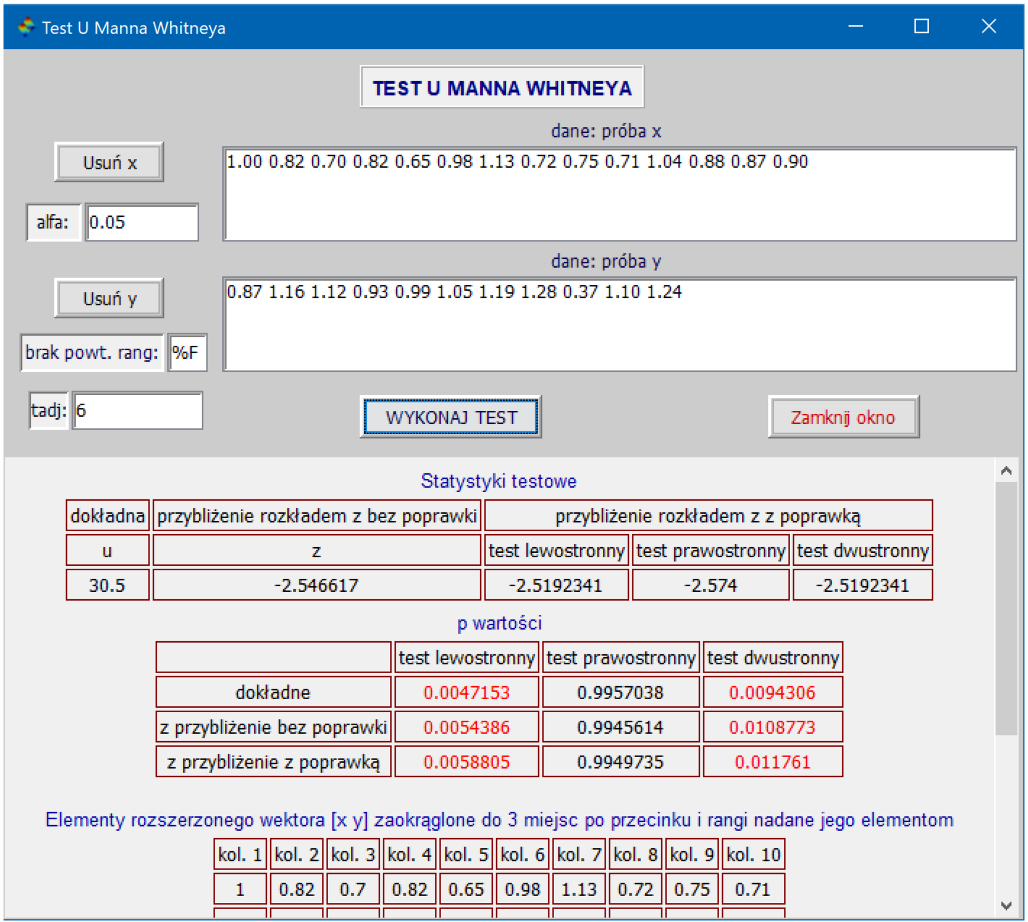
Uwaga o poprawce 'tadj'

Poprawka 'tadj' równa jest $\left(\sum_{i=1}^{k}(t_i^3 - t_i)\right)/2$, gdzie k jest liczbą różnych grup równych wartości w ciągu statystyk pozycyjnych a t_i liczbą równych wartości w i-tej grupie. Związek `tadj` z poprawką dla statystyki (przybliżenie rozkładem normalnym) w teście U Manna-Whitneya (n_A, n_B - odpowiednio liczebności prób `x` i `y`):

$$\frac{\sum_{i=1}^{k}(t_i^3 - t_i)}{(n_A+n_B)(n_A+n_B+1)} \quad \left(= \frac{2}{(n_A+n_B)(n_A+n_B+1)} \times \mathtt{tadj}\right)$$

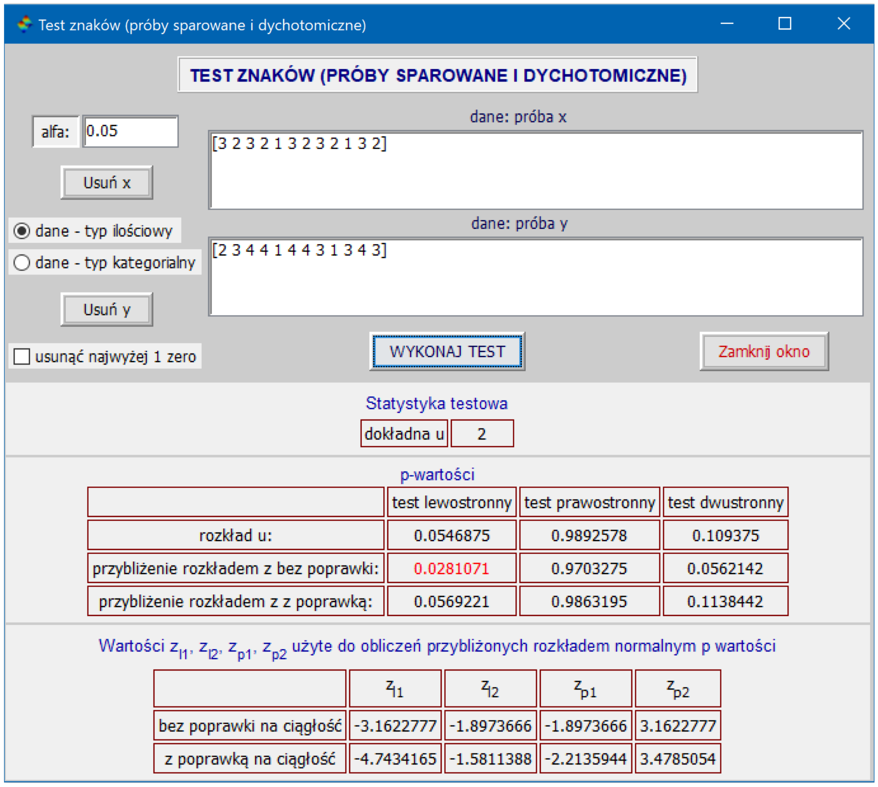
Polecenie: 'Test znaków dla prób sparowanych i dychotomicznych'

Po wybraniu z menu tego testu otrzymamy okno, w którym możemy wykonać test znaków zarówno dla dwóch zmiennych typu ilościowego jak i dla jednej zmiennej typu kategorialnego, która przyjmuje tylko 2 różne wartości

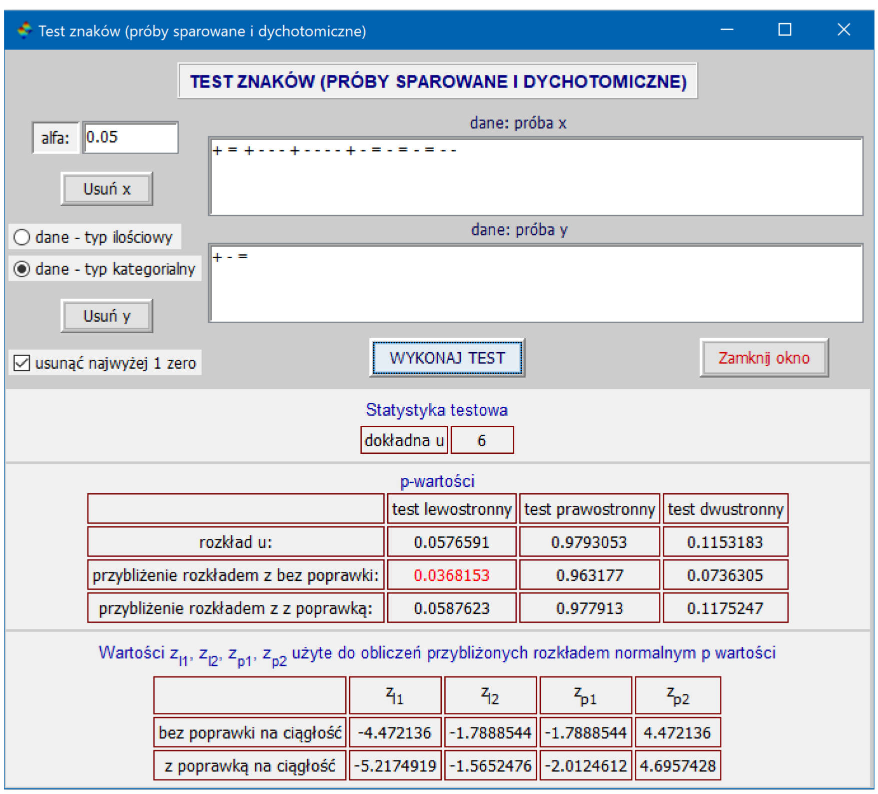


Rys. 87: Okno z wynikami testu U Manna-Whitneya

(zmienną przyjmującą tylko 2 różne wartości nazywamy zmienną dychotomiczną). Dla zmiennej dychotomicznej będziemy dopuszczać (jako wyjątek) przyjmowanie trzeciej wartości, ale co najwyżej kilku takich wartości, które będziemy traktować jako brak rozstrzygnięcia, do której z dwóch 'regularnych' wartości jej wartość należy. Jeśli mamy do czynienia z danymi x i y typu ilościowego, to wpisujemy je odpowiednio do okienka 'dane: próba x' i okienka 'dane: próba y' i zaznaczamy opcję 'dane - typ ilościowy'. Dla takich danych oblicza się różnice x-y odpowiednich elementów i zwykle usuwa się wszystkie tak otrzymane różnice zerowe. Można jednak zaznaczyć opcję 'usunąć najwyżej jedno zero' i wtedy, jeśli liczba zer jest nieparzysta, usuwa się jedno zero, i jeśli pozostały jeszcze jakieś zera to jedną ich połowę zalicza się do różnic dodatnich, a drugą do ujemnych. Jeśli mamy do czynienia z danymi typu kategorialnego, to dane te wpisuje się do okienka 'dane: próba x' a do okienka 'dane: próba y' wpisujemy tylko 3 wartości w kolejności: dwie 'regularne' wartości, które ta zmienna przyjmuje i trzecią (nawet gdy ona nie występuje w próbie x), którą traktujemy jako brak rozstrzygnięcia (remis, element neutralny). Wpisując do okienek elementy wektorów danych typu kategorialnego nie musimy brać ich w cudzysłów lub używać apostrofów. Program po ich pobraniu sam dopisze im cudzysłowy. Ale musimy zaznaczyć opcję 'dane - typ kategorialny'. Możemy też zaznaczyć opcję 'usunąć najwyżej jedno zero' i wtedy ta trzecia wartość będzie traktowana tak jak jest traktowane zero dla danych typu ilościowego. Po kliknięciu w 'WYKONAJ TEST' otrzymamy wyniki. Wyjaśnijmy jeszcze rolę wyświetlanych wartości z_{l1}, z_{l2}, z_{p1} i z_{p2}. A mianowicie, jeśli użyjemy ich we wzorach $\Phi(z_{l2})-\Phi(z_{l1})$ i $\Phi(z_{p2})-\Phi(z_{p1})$, to otrzymamy p-wartości z-przybliżeń w kolejności: lewostronną i prawostronną. Zwróćmy uwagę, że dla rozkładów dyskretnych suma lewej i prawej p-wartości nie musi być równa 1. Dlatego, na ogół nie można otrzymać p-wartości prawostronnej odejmując od 1 lewostronną p-wartość. Rys. 88 i Rys. 89 ilustrują odpowiednio użycie danych typu ilościowego i typu kategorialnego.



Rys. 88: Okno z wynikami testu znaków dla danych typu ilościowego



Rys. 89: Okno z wynikami testu znaków dla danych typu kategorialnego

Rozwiązania zadań z przykładów z podręcznika autorstwa Anny Baranowskiej "ELEMENTY STATYSTYKI dla studentów uczelni medycznych. Nowoczesne ujęcie z opisem obliczeń w programach Excel, R i Statistica" z użyciem środowiska obliczeniowego SCILAB i okienkowego toolboxa Basicstat

Uwaga

- W odpowiednie okienka trzeba wypisać wszystkie elementy wektorów danych, mimo że w niżej wypisanych danych podanych jest tylko kilka początkowych i końcowych danych a pomiędzy nimi znajdują się trzy kropki.
- Dla wielu przykładów aby uzyskać potrzebny wynik, trzeba skorzystać z konsoli Scilaba i tam wpisać kilka prostych poleceń. Aby nie pomijać takich przykładów, dajemy takie opisy ich rozwiązań. Pamiętajmy jednak, że kopiując z pliku .pdf polecenia, często takie znaki jak apostrofy, cudzysłowy, znak sim (~) czy znaki minus zastępowane są podobnymi znakami o innych kodach i uruchomiając te polecenia, Scilab sygnalizuje błędy. Wtedy takie 'podejrzane' znaki w poleceniach należy usunąć i ponownie wpisać je ręcznie.

Aby uruchomić menu okienkowej wersji Basicstat należy kliknąć w napis 'Basicstat' na pasku menu okna konsoli wskazany strzałką na Rys. 5 na str. 10.

Przykład 1.1.1

```
// Tabela 1.1 w podręczniku
Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Tabele i wykresy'. W okienko 'Granice Cl
przedziałów klasowych' wpisujemy granice przedziałów klasowych [100 110 ... 160 170] a w okienko
'Dane x' wpisujemy dane [117 136 ... 114 121]. Wybieramy 'Tabela częstości' klikając w odpowiedni prze-
łącznik. W polu 'tytuł tabeli' wpisujemy: Rozkład częstości i względny rozkład częstości ciśnienia
tętniczego krwi w badanej grupie. Klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując tabelę.
// Rys. 1.1 (histogram) w podręczniku
Likwidujemy zaznaczenie 'Tabela częstości'. Zaznaczamy 'histogram częstości' i '+krzywa rozkł. normalnego'.
W polu 'tytuł osi x histogramu' wpisujemy: Ciśnienie skurczowe w mg Hg a w polu 'tytuł osi y histogramu'
wpisujemy: Liczba pracowników i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując histogram.
// Rys 1.2 (histogram) w podręczniku
Zaznaczamy 'histogram wzgl. częstości'. Pole 'tytuł osi x histogramu' pozostawiamy bez zmian a w polu
'tytuł osi y histogramu' wpisujemy: Liczba pracowników i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES'
otrzymując histogram.
// Rys 1.7 (wielobok częstości) w podręczniku
Likwidujemy zaznaczenie 'histogram wzgl. częstości'. Zaznaczamy 'wielobok częstości'. W polu 'tytuł osi x
wieloboku' wpisujemy: Ciśnienie tętnicze skurczowe a w polu 'tytuł osi y wieloboku'
wpisujemy: Częstość i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując wielobok
```

Przykład 1.1.2

```
Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Tabele i wykresy'.
// Tabela 1.2 w podręczniku
W okienko 'Dane x' wpisujemy dane [0+ 0+ ... A+ B+] i zaznaczamy przełącznik 'Dane - tryb kategorialny'.
Wybieramy 'Tabela częstości' klikając w odpowiedni przełącznik. W polu 'tytuł tabeli' wpisujemy: Rozkład
grup krwi w badanej grupie. Klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując tabelę.
// Rys 1.3 (wykres słupkowy) w podręczniku
Likwidujemy zaznaczenie 'Tabela częstości' a zaznaczamy 'histogram częstości' (dla danych kategorialnych
otrzymamy wykres słupkowy zamiast histogramu!). W polu 'tytuł osi x wieloboku' wpisujemy: Grupa krwi i
czynnik RH a w polu 'tytuł osi y wieloboku' wpisujemy: Częstość i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES'
otrzymując wykres słupkowy.
// Rys 1.4 (wykres słupkowy) w podręczniku
Zaznaczamy 'histogram wzgl. częstości' (dla danych kategorialnych otrzymamy wykres słupkowy zamiast
histogramu!). Pole 'tytuł osi x wieloboku' pozostawiamy bez zmian a w polu 'tytuł osi y wieloboku'
wpisujemy: Częstość względna i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując wykres słupkowy.
```

Przykład 1.1.3

```
Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Tabele i wykresy'.
// Tabela 1.3 w podręczniku
W okienko 'Granice Cl przedziałów klasowych' wpisujemy granice przedziałów klasowych
```

[0.5 1.5 ... 5.5 6.5], a w okienko 'Dane x' wpisujemy dane [3 3 4 ... 4 3 4]. Wybieramy 'Tabela częstości' klikając w odpowiedni przełącznik. W polu 'tytuł tabeli' wpisujemy: Rozkłady częstości i skumulowane rozkłady częstości czasów reakcji organizmu na lek obniżający temperaturę ciała. Klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując tabelę.

*// **Rys 1.5** (histogram):*

Likwidujemy zaznaczenie 'Tabela częstości'. Zaznaczamy 'histogram skumul. częstości'. W polu 'tytuł osi x histogramu' wpisujemy: Czas w godzinach a w polu 'tytuł osi y histogramu' wpisujemy: Liczba pacjentów i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując histogram.

*// **Rys 1.6:** (wykres kołowy):*

Likwidujemy zaznaczenie 'Tabela częstości'. Zaznaczamy opcję 'wykres kołowy częstości'. W polu 'tytuł wykresu kołowego' wpisujemy: Wykres kołowy rozkładu częstości czasu potrzebnego po zażyciu leku do obniżenia temperatury ciała o co najmniej 1°C i klikamy w 'WYKONAJ TABELĘ i (lub) WYKRES' otrzymując wykres kołowy.

Przykład 1.2.1

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [4.09 3.79 ... 3.72 2.97] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polu 'średnia' mamy wynik.

Przykład 1.2.2

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [2.70 2.85 ... 4.05 4.09] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polu 'mediana' mamy wynik.

Przykład 1.2.3

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [3 5 ... 4 3] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'średnia' i 'mediana' mamy wynik.

Przykład 1.2.4

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [3 6 ... 45 56] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'średnia' i 'mediana' mamy wyniki dla danych w wierszu A. Usuwamy dane A i w ich miejsce wpisujemy [13 13 ... 40 42] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'średnia' i 'mediana' mamy wyniki dla danych w wierszu B.

Przykład 1.2.5

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [3 6 ... 45 56] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'wariancja' i 'odchylenie standardowe' mamy wyniki dla danych w wierszu A. Usuwamy dane A i w ich miejsce wpisujemy [13 13 ... 40 42] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'wariancja' i 'odchylenie standardowe' mamy wyniki dla danych w wierszu B.

Przykład 1.2.6

Parametrów tych nie można policzyć w oknie Basicstat. Najpierw korzystamy z funkcji "wekciord" toolboxa 'Basicstat'. W konsoli Scilaba wpisujemy: `x=[1 2 3 4 5 6]; k=[7 13 21 12 6 1];` i naciskamy Enter. Następnie wpisujemy w konsoli `daneoryg=wekciord(x,k)` i naciskamy Enter. Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy słowo daneoryg i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'średnia', 'wariancja' i 'odchylenie standardowe' mamy wyniki.

Przykład 1.2.7

Parametrów tych nie można policzyć w oknie Basicstat i postępujemy jak w Przykładzie 1.2.6. W konsoli Scilaba wpisujemy: `x=[105 115 125 135 145 155 165]; k=[6 14 18 13 5 3 1];` i naciskamy Enter (x - środki przedziałów). Następnie wpisujemy w konsoli `daneoryg=wekciord(x,k)` i naciskamy Enter. Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy słowo daneoryg i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'średnia', 'wariancja' i 'odchylenie standardowe' mamy wyniki.

Przykład 1.2.8

Parametrów tych nie można policzyć w oknie Basicstat i postępujemy jak w Przykładzie 1.2.6. W konsoli Scilaba wpisujemy (x-środki przedziałów): `x=[400 800 1250 1750 2250 2750 3250 3750 4250 4750 6000];` i naciskamy Enter oraz wpisujemy `k=[261 1012 2124 4649 14268 56471 143999 119365 35000 4667 427];` i naciskamy Enter. Następnie wpisujemy w konsoli `daneoryg=wekciord(x,k)` i naciskamy Enter. Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy słowo daneoryg i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polach 'średnia', 'wariancja' i 'odchylenie standardowe' mamy wyniki.

Przykład 1.2.9

Wpisujemy w konsoli Scilaba kolejno poniższe wiersze i po każdym z nich naciskamy Enter:

```
sA=53.85;osA=12.37;sB=43.17;
osB=12.73;VA=round(10000*osA/sA)/10000;
VB=round(10000*osB/sB)/1000;
100*[VA VB]
```

Przykład 1.2.10

Wpisujemy w konsoli Scilaba kolejno poniższe wiersze i po każdym z nich naciskamy Enter:

```
sA=104.59;osA=21.74;sB=79.24;
osB=7.22;VA=round(10000*osA/sA)/10000;
VB=round(10000*osB/sB)/10000;
100*[VA VB]
```

Przykład 1.2.11

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [2 7 ... 9 1], zaznaczamy kratkę 'oblicz kwantyl' oraz wybieramy przełącznik 'm. 2' a w polu 'poniżej wpisz prawdopodobieństwo' wpisujemy 0.05 0.2 0.25 ... 0.75 0.95 i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polu 'kwantyl' otrzymamy obliczone kwantyle.

Przykład 1.2.12

Postępujemy dokładnie tak samo jak w Przykładzie 1.2.11 z tą różnicą, że w polu 'dane x (tylko ilościowe)' wpisujemy [90,85 ... 68,84] a w polu 'poniżej wpisz prawdopodobieństwo' wpisujemy [0.5 0.6 0.7 0.8 0.9].

Przykład 1.2.13

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [2,7,...,9,1] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W odpowiednich polach znajdujemy wyniki.

Przykład 1.2.14

Parametrów tych nie można policzyć w oknie Basicstat. Najpierw skorzystamy z funkcji "wekciord" toolboxa 'Basicstat'. W konsoli Scilaba wpisujemy: a=[1 2 ... 5 6]; ka=[1 3 ... 9 6]; i naciskamy Enter, następnie w konsoli Scilaba wpisujemy: kb=[6 9 ... 2 2]; i naciskamy Enter, następnie w konsoli Scilaba wpisujemy: c=[1 2 ... 5]; kc=[1 3 ... 9 6]; i naciskamy Enter. Po czym kolejno wpisujemy w konsoli Scilaba: da=wekciord(a,ka); i naciskamy Enter, db=wekciord(a,kb); i naciskamy Enter oraz dc=wekciord(c,kc); i naciskamy Enter. Na koniec z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba' i w polu 'dane x (tylko ilościowe)' wpisujemy da (tylko te 2 litery) i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY' i z odpowiedniego pola odczytujemy wartość współczynnika skośności. Następnie, po zastąpieniu da przez db w polu 'dane x (tylko ilościowe)', klikamy w 'OBLICZ STATYSTYKI I PARAMETRY' i z odpowiedniego pola odczytujemy wartość współczynnika skośności. Na koniec, po zastąpieniu db przez dc w polu 'dane x (tylko ilościowe)', klikamy w 'OBLICZ STATYSTYKI I PARAMETRY' i z odpowiedniego pola odczytujemy wartość współczynnika skośności.

Przykład 1.2.15

Postępujemy analogicznie jak w Przykładzie 1.2.14, z tym, że zamiast trzech par danych mamy jedną parę danych c=[1 2 ... 4 5]; kc=[3 5 ... 5 3]; i wartość kurtozy odczytujemy z innego pola. **Uwaga:** Jeśli wcześniej wykonaliśmy obliczenia dla Przykładu 1.2.14, to kurtozę odczytamy z odpowiedniego pola dla części c tego przykładu.

Przykład 1.2.16

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [117 136 ... 114 121] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W polu 'kurtoza' znajdujemy wynik.

Przykład 1.2.17

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'obliczanie współczynników korelacji i regresja'. W polu 'dane: wektor x' wpisujemy [75 76 70 ... 98 82 65] a w polu 'dane: próba y' wpisujemy [130 133 ... 156 124]. Zaznaczamy kratki 'Wyznacz współczynnik korelacji' i 'Wykonaj wykres rozrzutu danych dla danych x i y' oraz przełącznik 'Pearsona' i klikamy w 'WYKONAJ OBLICZENIA'. Współczynnik korelacji znajdziemy w polu ρ:.

Przykład 2.2.11

Należy w konsoli Scilaba wpisać: f10=factorial(10); f2=factorial(2); f3=factorial(3); f1=factorial(1) i nacisnąć Enter. Następnie wpisać lp=f10/(f2*f3*f2*f2*f1) i nacisnąć Enter.

Przykład 2.2.12

Należy w konsoli Scilaba wpisać k=nchoosek(3,2) i nacisnąć Enter.

Przykład 3.1.4

Wartości tej dystrybuanty nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej k=0:6; p=distfun_binocdf(k,6,0.5) i naciskamy Enter.

Przykład 3.1.6

Wartości oczekiwanej nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej x=0:10; p=[0.003 0.002 ... 0.1 0.014]; sum(x.*p) i naciskamy Enter.

Przykład 3.1.8

Wartości oczekiwanej, wariancji i odchylenia standardowego nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `x=15:21; p=[0.10 0.15 ... 0.05]; sr=sum(x.*p)` i naciskamy Enter, wpisujemy `v=sum(x.^2.*p)-m^2` i naciskamy Enter, wpisujemy `sd=sqrt(v)` i naciskamy Enter.

Przykład 3.1.9

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `n=8; k=3; p=0.25; p=distfun_binopdf(3,8,0.25)` i naciskamy Enter.

Przykład 3.1.10

Wartości oczekiwanej, wariancji i odchylenia standardowego nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `n=2000; p=1/80;  sr=n*p` i naciskamy Enter, wpisujemy `v=n*p*(1-p)` i naciskamy Enter, wpisujemy `sd=sqrt(v)` i naciskamy Enter.

Przykład 3.1.11

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `n=10; k=8; p=0.8; p=distfun_binocdf(k,n,0.8)` i naciskamy Enter

Przykład 3.1.12

Prawdopodobieństw nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `pa=distfun_poisspdf(5,8)` i naciskamy Enter, wpisujemy `pb=distfun_poisscdf(2,8,%f)` i naciskamy Enter, wpisujemy `pc=distfun_poisscdf(7,8)-distfun_poisscdf(4,8)` i naciskamy Enter.

Przykład 3.1.13

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `m1=1024/(16*32);lb=3*m1; p=1-distfun_poisscdf(0,lb,%f)` i naciskamy Enter.

Przykład 3.2.2

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `p=distfun_unifcdf(1.5,0,2)-distfun_unifcdf(0.5,0,2)` i naciskamy Enter.

Przykład 3.3.1

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `p=distfun_normcdf(200,177.4,7.7)` i naciskamy Enter.

Przykład 3.3.2

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `p=1-distfun_normcdf(4000,3370.1,580.02)` i naciskamy Enter.

Przykład 3.3.4

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej `q=distfun_norminv([0.25,0.75],125,7.5)` i naciskamy Enter.

Przykład 3.3.5

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej kolejno `p1=100*(distfun_normcdf(1,0,1)-distfun_normcdf(-1,0,1))` i naciskamy Enter, `p2=100*(distfun_normcdf(2,0,1)-distfun_normcdf(-2,0,1))` i naciskamy Enter oraz `100*(distfun_normcdf(3,0,1)-distfun_normcdf(-3,0,1))` i naciskamy Enter.

Przykład 3.4.4

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Wpisujemy w niej kolejno `p1=distfun_normcdf(125,120,9)-distfun_normcdf(120,120,9)` i naciskamy Enter, `p2=distfun_normcdf(125,120,9/sqrt(36))-distfun_normcdf(120,120,9/sqrt(36))` i naciskamy Enter.

Przykład 3.4.5

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba.
Z rozkładu dwumianowego: Wpisujemy w konsoli `m=30;n=50;p=0.7;distfun_binocdf(m-1,n,p,%f)` i naciskamy Enter.
Przybliżenie rozkładem normalnym: Wpisujemy w konsolę `t=n*p;d=0.5; r=sqrt(t*(1-p)); a=(n-t+d)/r;b=(m-t-d)/r; distfun_normcdf(a,0,1)-distfun_normcdf(b,0,1)` i naciskamy Enter.

Przykład 3.4.6

Prawdopodobieństwa nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba.
Z rozkładu dwumianowego: Wpisujemy w konsolę `distfun_poisscdf(70,50,%f)` i naciskamy Enter.
Przybliżenie rozkładem normalnym: Wpisujemy w konsolę `a=0; b=70; l=50; m=50; r=sqrt(l);c=(b-l)/r;d=(a-l)/r; 1-(distfun_normcdf(c,0,1)-distfun_normcdf(d,0,1))` i naciskamy Enter.

Przykład 3.5.3

Sprawdźmy, że $F^{kr}_{\alpha,(m,n)} = 1/F^{kr}_{1-\alpha,(n,m)}$ dla $\alpha = 0{,}1$, $m = 9$ i $n = 5$.

Porównania nie można wykonać w oknie Basicstat i trzeba go wykonać w konsoli Scilaba. Porównajmy zatem wpisując w konsoli Scilaba [distfun_finv(0.1,5,9) distfun_finv(0.9,9,5) 1/distfun_finv(0.9,9,5)] i naciskając Enter.

Przykład 4.1.1 i **Przykład R.4.1.2**

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Podstawowe statystyki - 1 próba'. W polu 'dane x (tylko ilościowe)' wpisujemy [24.55 22.77 ... 23.05 21.37] i klikamy w 'OBLICZ STATYSTYKI I PARAMETRY'. W odpowiednich polach znajdujemy wyniki.

Przykład 4.1.8

Przedział ufności otrzymamy przy okazji testu 1 średniej. Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy [81 74 ... 79 81] i klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu w okienku (polu) wyników znajdziemy przedział ufności.

Przykład 4.1.9

Przedział ufności otrzymamy przy okazji testu 2 średnich. Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [3.99 4.78 ... 1.97 2.40] a w polu 'dane y' wpisujemy [2.76 0.16 ... -1.99 4.33], zaznaczamy przełącznik 'wariancje równe' i klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu w okienku (polu) wyników znajdziemy przedział ufności.

Przykład 4.1.10

Przedział ufności otrzymamy przy okazji testu 2 średnich. Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [246 249 ... 238 214] a w polu 'dane y' wpisujemy [159 147 ... 144 165], w polu 'alfa:' wpisujemy 0.03 (nie zaznaczamy przełącznika 'Wariancje równe'!) i klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu w okienku (polu) wyników znajdziemy przedział ufności.

Przykład 4.1.11

Postępujemy dokładnie jak w Przykładzie 4.1.9 tylko nie zaznaczamy przełącznika 'Wariancje równe'.

Przykład 4.1.12

Z menu 'Testy parametryczne' wybieramy 'Test proporcji (frakcji) dla 1 próby'. W polu 'liczba sukcesów' wpisujemy 16, w polu 'liczebność próby' wpisujemy 162 a w polu 'prawdopodobieństwo sukcesu' wpisujemy 16/162. Kratkę 'test dokładny' pozostawiamy niezaznaczoną. Klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu w okienku (polu) wyników znajdziemy przedział ufności.

Przykład 4.1.13

Z menu 'Testy parametryczne' wybieramy 'Test proporcji (frakcji) dla 2 prób'. W pola 'liczba sukcesów' i 'liczebność próby' dla 'PRÓBA I' wpisujemy odpowiednio 9 i 75 a w pola 'liczba sukcesów' i 'liczebność próby' dla 'PRÓBA II' wpisujemy odpowiednio 7 i 87. W polu 'Δ_0' pozostawiamy domyślną wartość 0. Klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu w polu wyników znajdziemy przedział ufności.

Przykład 4.1.14

Z menu 'Testy parametryczne' wybieramy 'Test wariancji dla 1 i 2 prób'. Wpisujemy (bez przecinków na końcu!): w polu 'dane x' [15.3 17.0 ... 16.8 16.3], w polu 'dane y' variance(x), w polu 'alfa:' 0.03 i zaznaczamy przełącznik 'test 1 wariancji'. Klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu w okienku wyników znajdujemy przedział ufności dla wariancji. Przedziału ufności dla odchylenia standardowego nie można policzyć w oknie Basicstat i obliczenia trzeba wykonać w konsoli Scilaba. Kopiujemy końce przedziałów ufności dla wariancji i bierzemy pierwiastki kwadratowe z nich otrzymując przedział ufności dla odchylenia standardowego, czyli wpisujemy w konsoli [sqrt(0.5623604) sqrt(2.3946423)] i naciskamy Enter.

Przykład 4.1.15

Z menu 'Testy parametryczne' wybieramy 'Test wariancji dla 1 i 2 prób'. W polu 'dane x' wpisujemy 0.025, w polu 'dane y' wpisujemy 0.036, w polu 'alfa:' wpisujemy 0.05 i w polu 'n:' wpisujemy [12 15]. Upewniamy się, że przełącznik 'Test jednej wariancji' nie jest zaznaczony i klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu okienka wyników znajdujemy przedział ufności dla ilorazu wariancji.

Przykład 4.1.16

Z menu 'Testy parametryczne' wybieramy 'Test istotności współczynników korelacji'. W polu 'dane: próba x' wpisujemy [75 76 ... 82 65], w polu 'dane: próba y' wpisujemy [130 133 ... 156 124], w polu 'alfa:' wpisujemy 0.05, wybieramy przełącznik 'Pearsona' i klikamy w 'WYZNACZ WSPÓŁCZYNNIK KORELACJI'. Następnie klikamy w 'WYKONAJ TEST'. W odpowiednim miejscu okienka (pola) wyników znajdujemy przedział ufności dla współczynnika korelacji.

Przykład 4.1.17

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Obliczanie współczynników korelacji i regresja'. W polu 'dane: wektor x' wpisujemy [20 20 ... 65 65] a w polu 'dane: próba y' wpisujemy [69.4 65.0 ... 79.0,82.9]. Zaznaczamy kratkę 'Wyznacz prostą i parametry regresji' i klikamy w 'WYKONAJ OBLICZENIA'. Estymatory parametrów znajdziemy w odpowiednich polach.

Przykład 4.1.18

Z menu 'Statystyka opisowa i inne obliczenia' wybieramy 'Obliczanie współczynników korelacji i regresja'. W polu 'dane: próba x' wpisujemy [14.0 46.6 ... 48.0 21.8] a w polu 'dane: próba y' wpisujemy [17.9 11.7 ... 12.2 18.2]. Zaznaczamy kratkę 'Wyznacz prostą i parametry regresji' oraz kratkę 'Wykonaj wykres rozrzutu z prostą regresji y=α+βx' i klikamy w 'WYKONAJ OBLICZENIA'.

Przykład 4.1.19

Zobacz Przykład 4.1.18 (nie zaznaczamy kratki 'Wykonaj wykres rozrzutu z prostą regresji y=α+βx')

Przykład 4.2.1

Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy [101.6 86.9 ... 78.0 90.5] a w polu 'mi:' wpisujemy 101 i klikamy w 'WYKONAJ TEST'. Test lewostronny.

Przykład 4.2.2

Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy 145, w polu 'mi:' wpisujemy 140, w polu 'n' wpisujemy 121, a w polu 'sd' wpisujemy 28.8 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.3

Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy 0.88, w polu 'mi:' wpisujemy 0.77, w polu 'n' wpisujemy 36, a w polu 'sd' wpisujemy 0.32 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.4

Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy 4400, w polu 'mi:' wpisujemy 4700, w polu 'n' wpisujemy 25, a w polu 'sd' wpisujemy 500 a w polu 'alfa:' wpisujemy 0.03 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.5

Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy [36.14 34.88 ... 36.89 34.83], w polu 'mi:' wpisujemy 35.5 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.6

Z menu 'Testy parametryczne' wybieramy 'Test średniej dla 1 próby'. W polu 'dane x' wpisujemy [130.34 126.69 ... 109.20 124.59], w polu 'mi:' wpisujemy 130 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.7

Z menu 'Testy parametryczne' wybieramy 'Test proporcji (frakcji) dla 1 próby'. W polu 'liczba sukcesów' wpisujemy 78, w polu 'liczebność próby' wpisujemy 160, a w polu 'prawdopodobieństwo sukcesu' wpisujemy 0.38. Nie zaznaczamy kratki 'test dokładny' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.8

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [30 25 23 ... 13 26 21], w polu 'dane y' wpisujemy [22 24 25 ... 30 19 42], zaznaczamy przełącznik 'wariancje równe' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.9

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [22 25 25 ... 21 23 26], w polu 'dane y' wpisujemy [6 6 7 ... 7 5 6], a w polu 'mi:' wpisujemy 16 (nie zaznaczamy przełącznika 'wariancje równe'!) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.10

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [32 30 31 ... 31 34 30], w polu 'dane y' wpisujemy [14 20 ... 21 15], w polu 'mi:' wpisujemy 10 (nie zaznaczamy przełącznika 'wariancje równe'!) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.11

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [1.00 0.82 ... 0.87 0.90], w polu 'dane y' wpisujemy [0.87 1.16 ... 1.10 1.24], w polu 'mi:' ma być 0 (nie zaznaczamy przełącznika 'wariancje równe'!) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.12

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [7.5 7.0 8.0 ... 8.0 7.5 7.0], w polu 'dane y' wpisujemy [7.5 5.5 6.5 ... 6.5 7.0 6.0], w polu 'mi:' wpisujemy 0.5, zaznaczamy przełącznik 'wariancje równe' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.13

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy 400, w polu 'dane y' wpisujemy 35, w polu 'mi:' wpisujemy 300, w polu 'n:' wpisujemy [40 50], a w polu 'sd:' wpisujemy [240 20] (nie zaznaczamy przełącznika 'wariancje równe'!) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.14

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [37.6 36.3 ... 37.1 38.4], w polu 'dane y' wpisujemy [36.9 36.1 ... 36.9 36.6], w polu 'alfa:' wpisujemy 0.01, zaznaczamy przełącznik 'próby sparowane' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.15

Z menu 'Testy parametryczne' wybieramy 'Test średnich dla 2 prób'. W polu 'dane x' wpisujemy [240 252 ... 245 217], w polu 'dane y' wpisujemy [235 232 ... 242 213], w polu 'alfa:' wpisujemy 0.01, zaznaczamy przełącznik 'próby sparowane' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.16

Z menu 'Testy parametryczne' wybieramy 'Test proporcji (frakcji) dla 2 prób'. W polach 'liczba sukcesów' i 'liczebność próby' dla 'PRÓBA I:' wpisujemy odpowiednio 300-12 i 300, a w polach 'liczba sukcesów' i 'liczebność próby' dla 'PRÓBA II:' wpisujemy odpowiednio 250-25 i 250 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.17

Z menu 'Testy parametryczne' wybieramy 'Test proporcji (frakcji) dla 2 prób'. W polach 'liczba sukcesów' i 'liczebność próby' dla 'PRÓBA I:' wpisujemy odpowiednio 200 i 230, a w polach 'liczba sukcesów' i 'liczebność próby' dla 'PRÓBA II:' wpisujemy odpowiednio 225 i 270 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.18

Z menu 'Testy parametryczne' wybieramy 'Test wariancji dla 1 i 2 prób'. W polu 'dane x' wpisujemy 144, w polu 'dane y' wpisujemy 100, w polu 'alfa:' wpisujemy 0.05, w polu 'n:' wpisujemy [10 10]. Upewniamy się, że przełącznik 'test jednej wariancji' nie jest zaznaczony i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.19

Z menu 'Testy parametryczne' wybieramy 'Test wariancji dla 1 i 2 prób'. W polu 'dane x' wpisujemy [22 25 ... 23 26], w polu 'dane y' wpisujemy [6 6 7 ... 7 5 6], w polu 'alfa:' wpisujemy 0.05. Upewniamy się, że przełącznik 'test jednej wariancji' nie jest zaznaczony i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.20

Z menu 'Testy parametryczne' wybieramy 'Test wariancji dla 1 i 2 prób'. W polu 'dane x' wpisujemy [7.5 7.0 ... 7.5 7.0], w polu 'dane y' wpisujemy [7.5 5.5 ... 7.0 6.0], w polu 'alfa:' wpisujemy 0.05. Upewniamy się, że przełącznik 'test jednej wariancji' nie jest zaznaczony i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.21

Z menu 'Testy parametryczne' wybieramy 'Test istotności współczynników korelacji'. W polu 'dane: próba x' wpisujemy [75 76 ... 82 65], w polu 'dane: próba y' wpisujemy [130 133 ... 156 124], w polu 'alfa:' wpisujemy 0.05, wybieramy przełącznik 'Pearsona' i klikamy w 'WYZNACZ WSPÓŁCZYNNIK KORELACJI'. Następnie klikamy w 'WYKONAJ TEST'.

Przykład 4.2.22

Z menu 'Testy parametryczne' wybieramy 'Test istotności współczynników korelacji'. W polu 'dane: próba x' wpisujemy [14.0 ... 21.8], w polu 'dane: próba y' wpisujemy [17.9 11.7 ... 12.2 18.2], w pole 'alfa:' wpisujemy 0.05, wybieramy przełącznik 'Pearsona' i klikamy w 'WYZNACZ WSPÓŁCZYNNIK KORELACJI'. Następnie klikamy w 'WYKONAJ TEST'. Test został wykonany dla hipotezy $H_0 : \rho = 0$. Jeśli chcemy testować hipotezę $H_0 : \rho = -0.7$, to możemy wykorzystać 95% przedział ufności: -0.7 nie należy do tego przedziału, a ponieważ -0.7 znajduje się po prawej stronie tego przedziału, to możemy być w 95% pewni, że $\rho < -0.7$. Jeśli jednak chcemy wykonać test taki jak w podręczniku, to w konsoli Scilaba wpisujemy: r=-0.8668119411195; n=30; (r bierzemy z okna 'Test istotności współczynników korelacji') a następnie wpisujemy: z=(sqrt(n-3)/2)*log((1+r)*(1+0.7)/((1-r)*(1-0.7))) i naciskamy Enter. Po czym wpisujemy p=distfun_normcdf(z,0,1) i naciskamy Enter.

Przykład 4.2.23

```
// Albo wpisujemy w konsoli:
x=[46 23 44 ... 41 39 30]; qqwykres(x); i Enter // Albo otrzymujemy go przy okazji testu: Shapiro-Wilka
```

Z menu 'Testy nieparametryczne' wybieramy 'Test normalności Shapiro-Wilka' i w polu 'dane x' wpisujemy [46 23 44 ... 41 39 30], zaznaczamy kratkę 'wykonaj wykres QQ' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.24

Z menu 'Testy nieparametryczne' wybieramy 'Test normalności Shapiro-Wilka' i w polu 'dane x' wpisujemy [3.99 4.78 ... 1.97 2.40], w polu 'alfa:' wpisujemy 0.01 oraz zaznaczamy kratkę 'wykonaj wykres QQ' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.25

Z menu 'Testy nieparametryczne' wybieramy 'Test normalności Shapiro-Wilka' i w polu 'dane x' wpisujemy [7.5 7.0 ... 7.5 7.0]. Zaznaczamy kratkę 'wykonaj wykres QQ' (alfa ma być 0.05) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.26

Z menu 'Testy nieparametryczne' wybieramy 'Test normalności Shapiro-Wilka' i w polu 'dane x' wpisujemy [0.51 0.87 ... 1.83 2.12]. W polu 'alfa:' wpisujemy 0.01 oraz zaznaczamy kratkę 'wykonaj wykres QQ' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.27

Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat zgodności rozkładu danych z zadanym rozkładem'. W polu 'wektor xo:' wpisujemy [40 35 15 10], w polu 'wektor xg;' wpisujemy 100*[0.37 ... 0.08], a w polu 'alfa:' wpisujemy 0.03 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.28

// Część (a) Poniżej opis zaczyna się od wyboru odpowiednich przycisków. Jest to ważne ponieważ one aktywują
// odpowiednie pola i wyświetlają nad nimi właściwe informacje o tych polach.
Z menu 'Basicstat' wybieramy 'Wyznaczanie częstości w klasach'. Ponieważ mamy dane ilościowe,to przycisk 'dane - typ kategorialny' ma być wyłączony. Mamy też dane dla wzrostu i odpowiadające im częstości. Dlatego wybieramy przycisk radiowy 'liczebności' lub 'wartości w' (jeśli wybierze się jeden z nich, to drugi włącza się automatycznie). Ponieważ w podręczniku były wybrane przedziały klasowe lewostronnie domknięte, wybieramy przycisk 'prz, klas. lewost. domk.'. Następnie w pole 'granice cl przedziałów klasowych' wpisujemy (lub wklejamy) [90 129 133 137 141 146 150 154 195], w pole 'wartości w, dla których niżej podane są liczebności' wpisujemy [124 125 126 127 ... 153 154 155 157] a wektor [1 1 1 1 1 1 6 2 ... 6 3 5 1 3 2 3 2] wpisujemy w pole 'liczebności lc' i klikamy w 'WYZNACZ CZĘSTOŚCI'. Następnie obliczamy wartości oczekiwane na podstawie siatek centylowych. Ponieważ z siatek wynika,że w rozpatrywanych przedziałach wzrostu znajduje się (w \%) [3 7 15 25 25 15 7 3] chłopców, a mamy 200 chłopców to dzieląc te liczby przez 100 i mnożąc przez 200, czyli mnożąc ten wektor przez 2 otrzymamy oczekiwane częstości [6 14 30 50 50 30 14 6]. Z menu 'Basicstat' wybieramy 'Test chi kwadrat zgodności rozkładu danych z zadanym rozkładem' i w pole 'xobs' wpisujemy częstości otrzymane we wcześniej otwartym oknie 'WYZNACZANIE CZĘSTOŚCI' a w pole 'xocz' wpisujemy wyliczone przez nas oczekiwane częstości. I klikamy w 'WYKONAJ TEST'.
// Część (b): Korzystamy z danych wczytanych w części (a) i obliczonego tam xo, dlatego trzeba najpierw
// wykonać część (a), jeśli nie była wcześniej wykonana. Będziemy musieli też wyestymować średnią i odchylenie
// standardowe rozkładu normalnego. Będzie ono dokładniejsze, jeśli użyjemy oryginalnego wektora x.
Z menu 'Basicstat' wybieramy 'Wyznaczanie częstości w klasach' (przyciski 'wartości w' i 'liczebności' zostaną automatycznie włączone. W pole 'liczebności lc' wpisujemy [1 1 1 1 1 1...1 3 2 3 2] (są to liczebności odpowiadające poszczególnym wartościom wzrostu chłopców), a w pole 'wartości w, dla których niżej podane są liczebności' wpisujemy [124 125 126 127...153 154 155 157] i klikamy w 'WYKONAJ OBLICZENIA' i kopiujemy wynik. Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat zgodności z rozkładami ciągłymi'. W polu 'granice cl przedziałów klasowych' wpisujemy [90 129 133 137 141 146 150 154 195], w polu 'dane x (liczebności przedziałów klasowych lub dane "surowe"' wklejamy skopiowany wynik, wybieramy przełączniki 'normalny' i 'liczba estym. parametrów' 2 oraz zaznaczamy kratkę 'dane x, to dane "surowe" zamiast liczebności', a w polu 'alfa:' wpisujemy 0.01. I klikamy w przycisk 'WYKONAJ TEST'. Zaznaczając dodatkowo kratkę 'liczebn. rozkł. ciągłego równe i klikając w przycisk 'WYKONAJ TEST', otrzymamy rozwiązanie sposobem opisanym drobniejszym drukiem na str. 201 podrecznika.

Przykład 4.2.29

// Część (a)
Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat dla tablic kontyngencji'. W polu 'tablica X:' wpisujemy tablicę [78 55;65 28], a w polu 'alfa;' wpisujemy 0.1 i klikamy w 'WYKONAJ TEST'.
// Część (b)
Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat dla tablic kontyngencji'. W polu 'tablica X:' wpisujemy tablicę [53 25;24 41], a w polu 'alfa;' wpisujemy 0.03 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.30

Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat dla tablic kontyngencji'. W polu 'tablica X:' wpisujemy tablicę [35 55 30;40 100 50;25 50 15], (w polu 'alfa;' ma być 0.05) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.31

Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat dla tablic kontyngencji'. W polu 'tablica X:' wpisujemy tablicę [9 11 7;25 22 20;6 7 13], (w polu 'alfa;' ma być 0.05) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.32

Z menu 'Testy nieparametryczne' wybieramy 'Test chi kwadrat dla tablic kontyngencji'. W polu 'tablica X:' wpisujemy tablicę [200 225;30 45], (w polu 'alfa;' ma być 0.05) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.33

Z menu 'Testy nieparametryczne' wybieramy 'Test Walda-Wolfowitza losowości próby'. W okienko 'Dane x' wpisujemy wektor [6 6 7 ... 7 5 6], (w polu 'alfa;' ma być 0.05) zaznaczamy przełącznik 'x - typ ilościowy' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.34

Z menu 'Testy nieparametryczne' wybieramy 'Test Walda-Wolfowitza losowości próby'. W polu 'dane x' wpisujemy wektor [69.4 65.0 ... 79.0 82.9], (w polu 'alfa;' ma być 0.05) zaznaczamy przełącznik 'x - typ ilościowy' i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.35

Z menu 'Testy nieparametryczne' wybieramy 'Test Walda-Wolfowitza jednorodności 2 populacji'. W polu 'dane: próba x' wpisujemy wektor [8.2 10.9 ... 9.1 9.1], w polu 'dane: próba y' wpisujemy wektor [8.6 6.8 ... 5.0 9.5] (w polu 'alfa;' ma być 0.05) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.36

Z menu 'Testy nieparametryczne' wybieramy 'Test znaków dla prób sparowanych i dychotomicznych'. W polu 'dane: próba x' wpisujemy wektor [33.1 39.3 ... 38.2 32.5], w polu 'dane: próba y' wpisujemy wektor [33.8 27.9 ... 30.8 28.2] (w polu 'alfa:' ma być 0.05) i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.37

Z menu 'Testy nieparametryczne' wybieramy 'Test znaków dla prób sparowanych i dychotomicznych'. W polu 'dane: próba x' wpisujemy wektor [- + + ... + - +], w polu 'dane: próba y' wpisujemy wektor [+ - =], w polu 'alfa:' wpisujemy 0.01 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.38

Z menu 'Testy nieparametryczne' wybieramy 'Test rangowanych znaków dla prób sparowanych'. W polu 'dane: próba x' wpisujemy wektor [50 49 ... 37 44], w polu 'dane: próba y' wpisujemy wektor [37 37 ... 40 36], w polu 'alfa:' wpisujemy 0.05 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.39

Z menu 'Testy nieparametryczne' wybieramy 'Test rangowanych znaków dla prób sparowanych'. W polu 'dane: próba x' wpisujemy wektor [4.9 5.9 4.4 ... 2.5 2.3 2.5], w polu 'dane: próba y' wpisujemy wektor [2.5 3.1 2.1 ... 2.0 2.5 4.1], w polu 'alfa:' wpisujemy 0.05 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.40

Z menu 'Testy nieparametryczne' wybieramy 'Test U Manna-Whitneya'. W polu 'dane: próba x' wpisujemy wektor [8.2 9.4 9.6 ... 16.1 17.6 21.5], w polu 'dane: próba y' wpisujemy wektor [4.2 5.2 5.8 ... 11.2 11.3 11.5], w polu 'alfa:' wpisujemy 0.05 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.41

Z menu 'Testy nieparametryczne' wybieramy 'Test U Manna-Whitneya'. W polu 'dane: próba x' wpisujemy wektor [22.4 23.0 24.5 ... 27.3 29.4 30.9], w polu 'dane: próba y' wpisujemy wektor [18.5 21.5 23.3 ... 29.5 32.5 33.9], w polu 'alfa:' wpisujemy 0.05 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.42

Z menu 'Testy nieparametryczne' wybieramy 'Test U Manna-Whitneya'. W polu 'dane: próba x' wpisujemy wektor [145 190 360 ... 370 475 200], w polu 'dane: próba y' wpisujemy wektor [90 130 165 ... 180 150 115], w polu 'alfa:' wpisujemy 0.025 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.43

Z menu 'Testy nieparametryczne' wybieramy 'Test McNemara dla tablicy 2x2 danych sparowanych'. W polach **a**, **b**, **c**, **d** wpisujemy kolejno 81, 48, 23 i 21, w polu 'alfa:' wpisujemy 0.01 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.44

Z menu 'Testy nieparametryczne' wybieramy 'Test McNemara dla tablicy 2x2 danych sparowanych'. W polach **a**, **b**, **c**, **d** wpisujemy kolejno 125, 3, 7 i 10, w polu 'alfa:' wpisujemy 0.01 i klikamy w 'WYKONAJ TEST'.

Przykład 4.2.45

Z menu 'Testy nieparametryczne' wybieramy 'Test McNemara dla tablicy 2x2 danych sparowanych'. W polach **a**, **b**, **c**, **d** wpisujemy kolejno 400, 50, 25 i 25, w polu 'alfa:' wpisujemy 0.01 i klikamy w 'WYKONAJ TEST'.

Przykład 4.3.1

```
// Kolumny Niepalący, Byli palący, itd. nazywamy krótko NP, BP, MP i DP. Zapisujemy dane w pliku tekstowym:
// 1-szy wiersz to: Papierosy,Testeron  2-gi to: NP,7.88  trzeci to: NP,8.24 itd. i ostatni 39-ty to: DP,3.89
// tj. każdy wiersz to jedna z nazw NP,BP,MP lub DP, przecinek i kolejna liczba z odpowiedniej kolumny tabeli.
// Zapisujemy ten plik w katalogu roboczym Scilaba (scilab, jeśli taki utworzyliśmy) pod nazwą Testosteron.csv
```

Z menu 'Testy parametryczne' wybieramy 'Anova jednoczynnikowa i dwuczynnikowa'. Klikamy w 'Wybierz plik .csv z danymi' i wybieramy plik Testosteron.csv. W polu **wybrano:** ma być wpisany Testosteron~Papierosy. Jeśli nie jest, to wpisujemy tam taki model i klikamy w 'Zatwierdź model', a następnie klikamy w 'WYKONAJ TEST'.

Przykład 4.3.2

// Gdy nie ma pliku "Testosteron.csv" tworzymy go jak w Przykładzie 4.3.1

Z menu 'Testy parametryczne' wybieramy 'Test Browna-Forsythe'a równości wariancji'. Dalej postępujemy dokładnie tak jak z testem anovy opisanym w Przykładzie 4.3.1. Jeśli jedna lub obie zmienne grupujące mają wartości liczbowe, to pojawia się informacja o tym oraz podany jest sposób postępowania.

Wykresy QQ otrzymamy najłatwiej, wpisując w konsoli Scilaba kolejno poniższe polecenia i po każdym z nich naciskając Enter: S = csvRead('Testosteron.csv',',','.', 'string'); xnp=strtod(S(S(:,1)=="NP",2)); qqwykres(xnp); xbp=strtod(S(S(:,1)=="BP",2)); qqwykres(xbp); xmp=strtod(S(S(:,1)=="MP",2)); qqwykres(xmp); xdp=strtod(S(S(:,1)=="DP",2)); qqwykres(xdp);

Przykład 4.3.3

// Gdy nie ma pliku "Testosteron.csv" tworzymy go jak w Przykładzie 4.3.1 a w oknie 'Test Anovy' otwieramy
// plik "Testosteron.csv" i wykonujemy test anovy tak jak opisaliśmy w Przykładzie 4.3.1.

Gdy wykonaliśmy anovę dla tego przykładu i nie zamknęliśmy okna z wynikami, to klikamy w przycisk 'Otwórz okno post-hoc testu' i przechodzimy do okna 'Post hoc testy'. W oknie tym wybieramy przełącznik 'Dunnett', a w pole 'grupa kontrolna' wpisujemy NP i klikamy w przycisk 'WYKONAJ TEST'. Gdy okno z wynikami zostało zamknięte lub nie wykonaliśmy testu anovy, to postępujemy tak jak opisaliśmy wyżej w pierwszych 2 linijkach.

Przykład 4.3.4

Postępujemy jak w opisie Przykładu 4.3.3, tylko zamiast przełącznika 'Dunnett' wybieramy 'Bonferroni'.

// Uwaga: Zawartość pola 'grupa kontrolna' jest ignorowana.
// W niektórych porównaniach zmieniona jest kolejność odejmowania, co powoduje, że różnice mają zmieniony znak
// z ujemnego na dodatni lub odwrotnie; skutkuje to zmianą granic przedziałów ufności - lewa granica zmienia znak
// i staje się prawą a prawa zmienia znak i staje się lewą granicą przedziału ufności. Tak będzie się też działo
// w innych przykładach. Można też próbować manipulować przełącznikiem ***zmień kolejność porównań*** *i raz*
// test wykonać przy zaznaczonym i odczytać część wyników dokładnie takich (pomijając ich kolejność) jak w
// podręczniku a potem go wyłączyć oraz wykonać test i znaleźć pozostałe wyniki takie jak w podręczniku.

Przykład 4.3.5

Postępujemy jak w opisie Przykładu 4.3.3, tylko zamiast przełącznika 'Dunnett' wybieramy 'Tukey'.

// Uwaga: Zawartość pola 'grupa kontrolna' jest ignorowana. Zobacz też komentarz z Przykładu 4.3.3.

Przykład 4.3.6

// Oznaczmy 3 kolumny krótko: N, P i W (od niski, prawidłowy i wysoki). Oznaczmy 2 wiersze też jedną
// literą: M i S (od młodsi i starsi). Zapisujemy dane w pliku tekstowym: 1-szy wiersz to:
// Cisnienie,BMI,WIEK, 2-gi to: 118.7,N,M 11-ty to: 119.8,N,M 12-ty to: 133.5,P,M
// 21-szy to: 127.4,P,M 22-gi to: 122.2, W,M 31-szy to: 148.3,W,M 32-gi to: 146.7,N,S
// 41-szy to: 122.7,N,S 42-gi to: 127.3,P,S 51-szy to 129.7,P,S 52-gi to: 141.5,W,S 61-szy to:
// 145.8,W,S Tj. każdy wiersz to kolejna liczba, nazwa kolumny (1 litera N P lub W), przecinek i nazwa
// wiersza (1 litera M lub S). Zapisujemy ten plik w katalogu roboczym Scilaba z nazwą Cisnienie.csv

Z menu 'Testy parametryczne' wybieramy 'Test Browna-Forsythe'a równości wariancji'. Klikamy w 'Wybierz plik .csv z danymi' i wybieramy plik Ciśnienie.csv. Sprawdzamy czy wybrany jest model anovy **A~B*C** (jeśli nie, to taki model wybieramy). Sprawdzamy czy w polu **wybrano:** wpisany jest model Cisnienie~WIEK*BMI (jeśli nie, to wpisujemy tam taki model i zatwierdzamy klikając w przycisk 'Zatwierdź model') i klikamy w 'WYKONAJ TEST'.

Przykład 4.3.7

// Gdy nie ma pliku "Cisnienie.csv" tworzymy go jak w Przykładzie 4.3.6

Z menu 'Testy parametryczne' wybieramy 'Anova jednoczynnikowa i dwuczynnikowa'. Klikamy w 'Wybierz plik .csv z danymi' i wybieramy plik Ciśnienie.csv. Sprawdzamy czy włączony jest przycisk **A~B*C** (jeśli nie, to go włączamy) i sprawdzamy, czy w polu **wybrano:** wpisany jest model Cisnienie~WIEK*BMI (jeśli nie, to wpisujemy tam taki model i zatwierdzamy go klikając w przycisk 'Zatwierdź model'). Następnie sprawdzamy, czy włączony jest przycisk **model: efekty stałe** (jeśczonymli nie, to go włączamy) i klikamy w 'WYKONAJ TEST'.

Przykład 4.3.8

// Gdy nie ma pliku "Cisnienie.csv" tworzymy go jak w Przykładzie 4.3.6 i wykonujemy test anovy jak w
// Przykładzie 4.3.7. Gdy (wcześniej) wykonaliśmy test anovy dla tego przykładu i nie zamknęliśmy okna, to:

klikamy w przycisk 'Otwórz okno post-hoc testu', klikamy w przycisk **y~A** oraz zaznaczamy przycisk 'Tukey'. Następnie klikamy w 'WYKONAJ TEST'. Otrzymamy wyniki dla różnicy M (młodsi) i S (starsi) (a nie w odwrotnej kolejności, jak w podręczniku). Dlatego różnica i końce przedziałów ufności mają wartości ujemne. Ale możemy włączyć przełącznik \textbf{zmień kolejność danych} i klikając w przycisk 'WYKONAJ TEST' otrzymamy wynik taki jak w podręczniku. Następnie, wybieramy przełącznik **y~B** i klikamy w 'WYKONAJ TEST'. Na koniec wybieramy przełączniki **y~A*B** oraz **zmień kolejność porównań** i przy nie zmienionym, tj. włączonym przełączniku 'Tukey' klikamy w 'WYKONAJ TEST'. Otrzymujemy wyniki takie jak w podręczniku tylko w zmienionej kolejności.

// Uwaga: Zawartość pola 'grupa kontrolna' jest ignorowana.