

JEDNA ZMIENNA

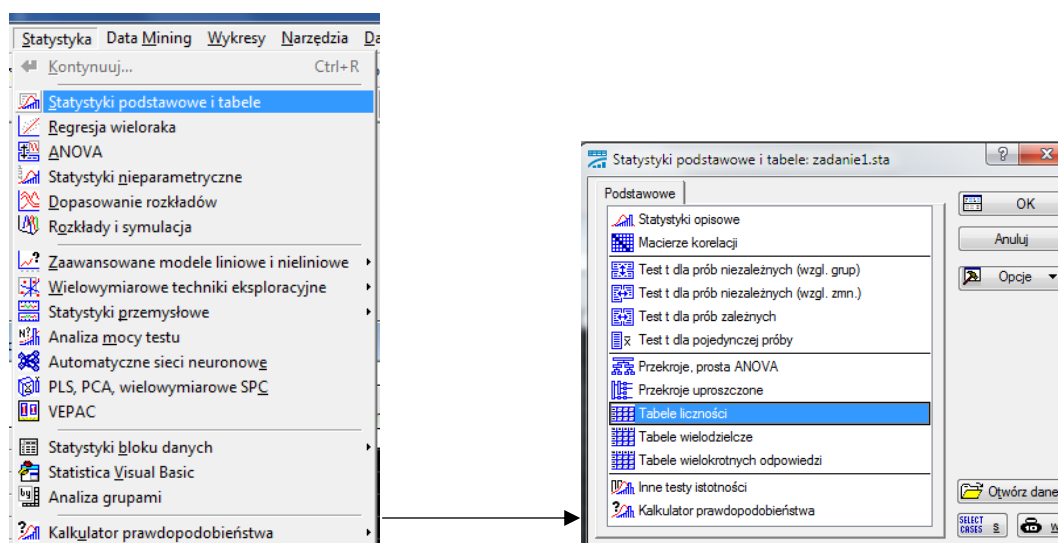
- Przygotowanie pliku (nazwy zmiennych, export plików .xlsx, selekcja przypadków);
- Graficzna prezentacja danych;
- Statystyki opisowe;
- Dane odstające;
- Weryfikacja normalności rozkładu cechy;
- Transformacja danych;
- Testowanie hipotezy dotyczącej wartości przeciętnej dla jednej zmiennej;
- Analiza mocy testu dla jednej próby. Minimalna liczność próby.

Wartość p-value jest to najmniejszy poziom istotności, przy którym możemy odrzucić hipotezę zerową. Pakiet statystyczny wylicza tę wartość na podstawie statystyki testowej. Decyzję podejmujemy następująco:

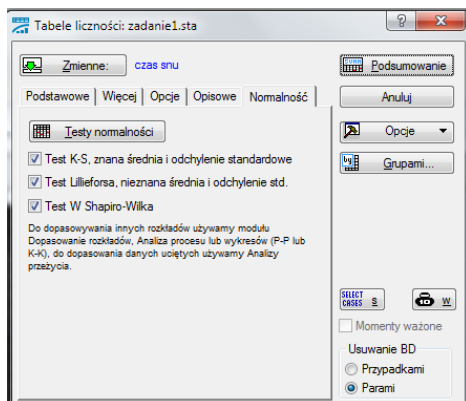
- Jeśli $p \leq \alpha$, to hipotezę zerową odrzucamy
- Jeśli $p > \alpha$, to nie mamy podstaw do odrzucenia hipotezy zerowej

Jak sprawdzić normalność rozkładu?

- **Testy statystyczne:**
 - **Test Kołmogorowa – Smirnowa** (wymaga znajomości wartości oczekiwanej i wariancji dla populacji).
 - **Test Kołmogorowa – Smirnowa z poprawką Lileforse’a** (nie wymaga znajomości wartości oczekiwanej i wariancji dla populacji).
 - **Test Shapiro – Wilka** - stosowany zwykle dla małych prób. Test preferowany.
- **Wykresy normalności**



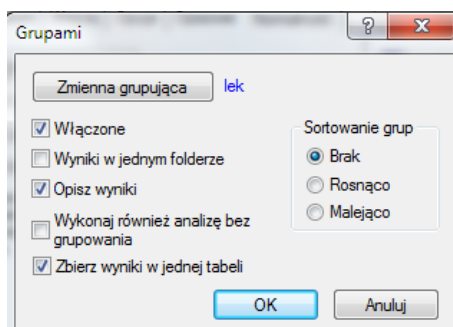
Test Shapiro–Wilka i inne testy normalności znajdziemy wybierając kolejno **Statystyka** → **statystyki podstawowe i tabele** → **tabele licznosci**.



Wybieramy zmienną do analizy (**zmienna**), a następnie przechodzimy do karty **normalność** i zaznaczamy **test W Shapiro-Wilka**.

Wybranie przycisku **testy normalności** spowoduje przeprowadzenie analizy, w której znajdziemy wartość statystyki testowej oraz wartość *p*.

Jeśli analizę rozkładu normalności chcemy przeprowadzić grupami, wybieramy dodatkowo zmienną grupującą (przycisk po lewej stronie **grupami**).



Wskazujemy **zmienną grupującą**, odznaczamy (jeśli nie potrzebujemy) **wykonaj również analizę bez grupowania**, zaznaczamy opcję **zbierz wyniki w jednej tabeli**. Dwa ostatnie kroki nie są niezbędne, ale ułatwiają czytanie wyników.

Wykresy normalności przedstawia wartości obserwowane (oś x) oraz wartości oczekiwane przy założeniu normalności rozkładu (oś y). Jeśli rozkład zmiennej jest normalny, przybiera formę linii prostej.

W pakiecie *Statistica* rozkład normalności znajdziemy w wielu miejscach. Kilka z nich:

- **Statystyka** → **statystyki podstawowe i tabele** → **statystyki opisowe** → karta **podstawowe** → **wykres 1**
- **Statystyka** → **statystyki podstawowe i tabele** → **statystyki opisowe** → karta **więcej** → **W1**
- **Statystyka** → **statystyki podstawowe i tabele** → **statystyki opisowe** → karta **w. prawd. i rozrzutu** → **wykres normalności**
- **Statystyka** → **statystyki podstawowe i tabele** → **tabele licznosci** → **opisowe**

Co zrobić, jeśli rozkład cechy nie jest normalny?

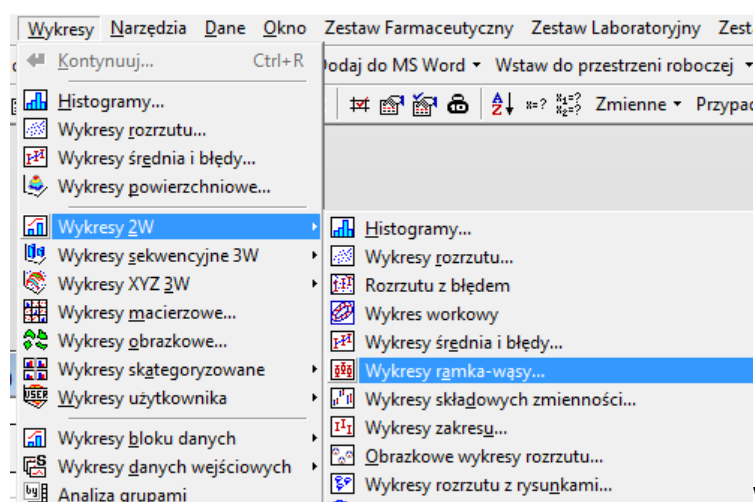
- Transformować dane;
- Stosować test niezależne od rozkładu – test nieparametryczne;
- Dla dużych prób stosować testy parametryczne, bowiem Centralne Twierdzenie Graniczne mówi, że nawet jeśli zmienna, którą mierzymy nie ma rozkładu normalnego, to statystyka testowa (oparta o średnią), dla dostatecznie dużych n ma rozkład normalny.

Dane odstające

Metoda oparta na odstępach kwartylowych: zakres danych nieodstających określa przedział:

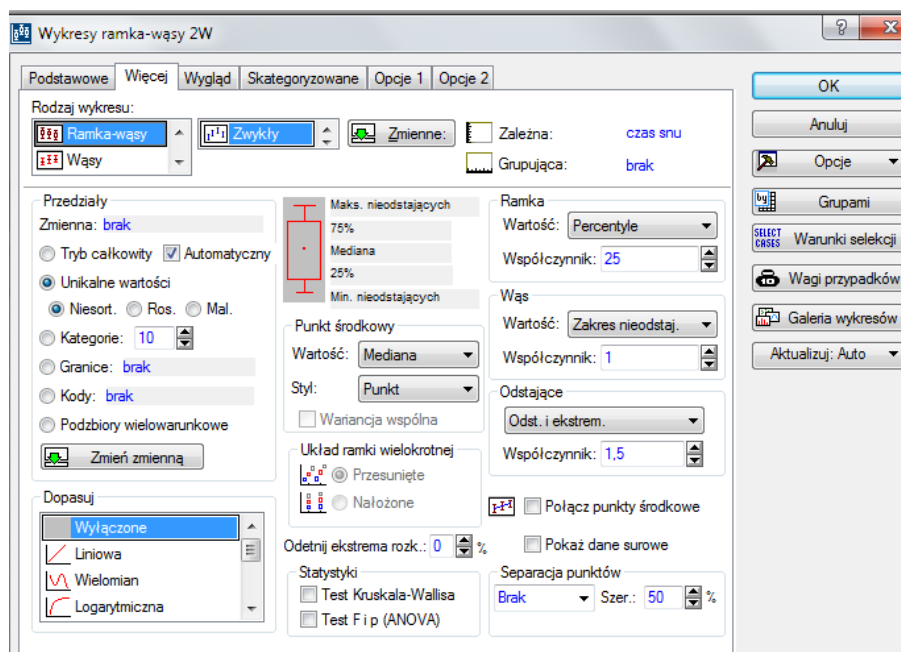
$$(Q_1 - 1,5 \cdot (Q_3 - Q_1); Q_3 + 1,5 \cdot (Q_3 - Q_1)) \quad (*)$$

Wybieramy kolejno **wykresy** → **wykresy 2W** → **wykresy ramka-wąsy**.



Wskazujemy zmienną do analizy.

Wszystkie domyślne ustawienia zostawiamy: punkt środkowy – mediana, ramka -kwartyle (percentyle 25%), współczynnik nieodstających 1,5. Po zatwierdzeniu przyciskiem **ok** otrzymamy wykres z zaznaczonymi punktami odstającymi. Uwaga: jeśli zakres nieodstających liczony ze wzoru (*) jest poniżej wartości najmniejszej (lub/i powyżej wartości największej), to program przyjmuje zakres nieodstających jako wartość minimum lub/i maksimum.

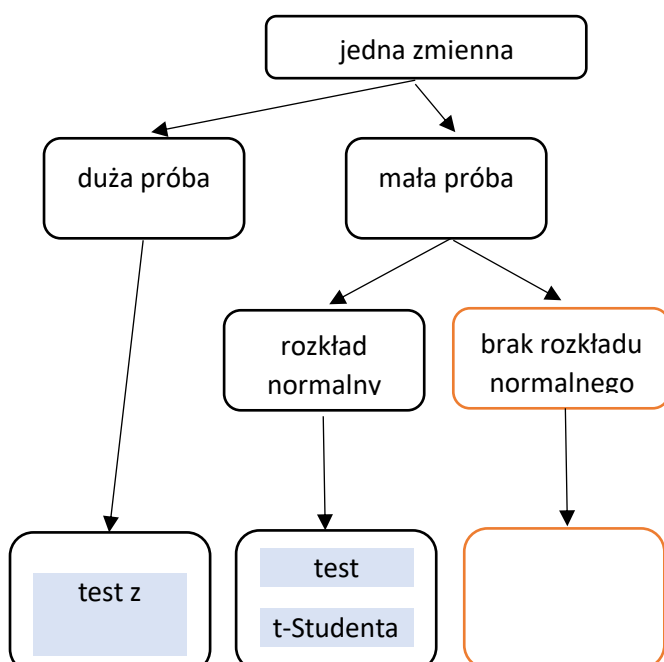


Metoda oparta na zasadzie trzech sigm.

Wykres ramka-wąsy wymaga zmiany ustawień: punkt środkowy – średnia, ramka – odchylenie standardowe, wąs - odchylenie standardowe - współczynnik 3, odstające wyłączone. Należy zaznaczyć **pokaż dane surowe**.

Test Grubbsa (wymaga normalności rozkładu). Wybieramy kolejno **statystyka** → **statystyki podstawowe i tabele** → **statystyki opisowe**. Na karcie **odporne** zaznaczamy **test Grubbsa na dane odstające**, wybieramy **zmienne**, a następnie **podsumowanie**. Jeśli test Grubbsa wykaże istnienie danej odstającej, to jest nią ta wartość spośród wartości minimalnej i maksymalnej, która leży dalej od wartości średniej.

Testowanie hipotez dotyczących wartości oczekiwanej dla jednej zmiennej.



UWAGA Ponieważ rozkład t-Studenta dla dużych licznosci jest w przybliżeniu rozkładem normalnym, test z (zarówno dla jednej zmiennej jak i par powiązanych) obliczeniowo daje takie same wyniki jak test t. W pakiecie Statistica zatem obydwa testy znajdują się w opcji test t.

Testowanie hipotez – BŁĘDY

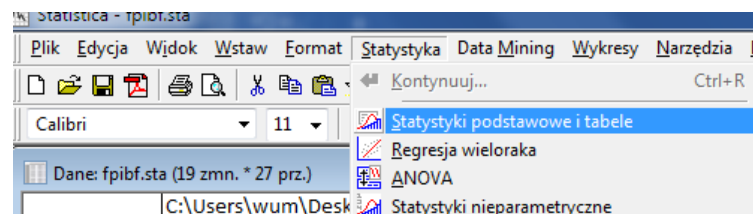
- **BŁĄD I RODZAJU** – odrzucenie hipotezy zerowej, gdy jest prawdziwa;
- **BŁĄD II RODZAJU** – przyjęcie hipotezy zerowej, gdy jest fałszywa.

$\alpha = P(\text{odrzuć hipotezy zerowej, gdy jest prawdziwa})$

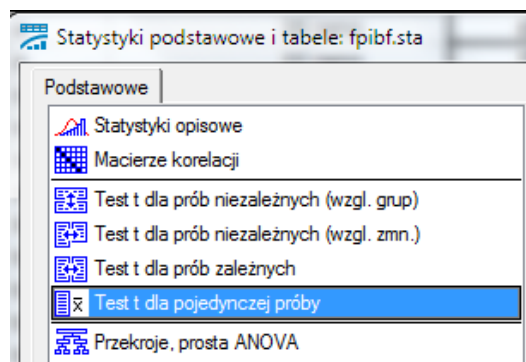
$\beta = P(\text{przyjęcie hipotezy zerowej, gdy jest fałszywa})$

$1 - \beta = P(\text{odrzuć hipotezy zerowej, gdy jest fałszywa}) \rightarrow \text{MOC TESTU}$

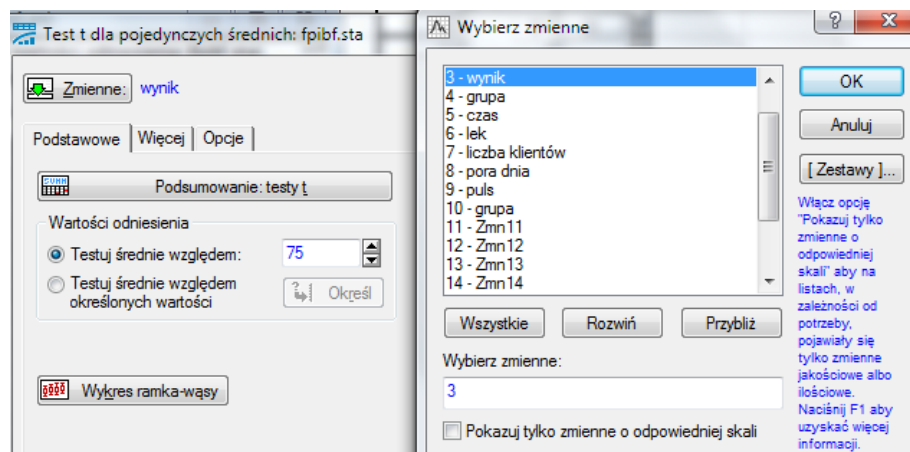
Test t-studenta dla jednej zmiennej



W zakładce **Statystyka** wybieramy **statystyki podstawowe i tabele**, a dalej **test t dla pojedynczej próby**.

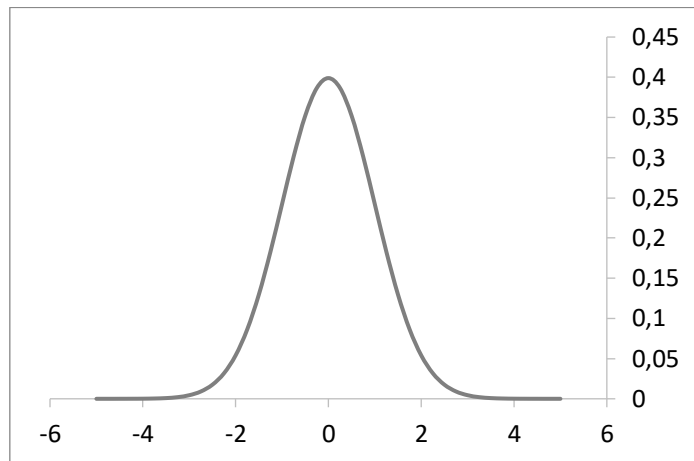


Należy wskazać **zmienne** do analizy oraz określić stałą z hipotezy zerowej (**testuj średnie względem ...**).



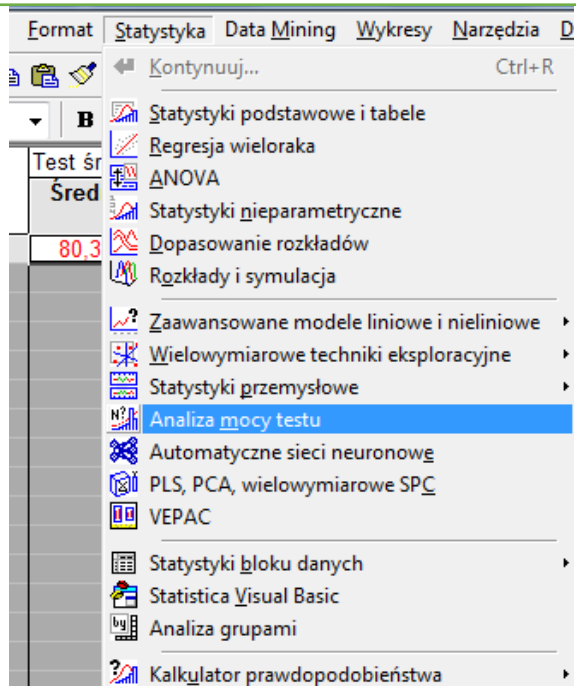
W **podsumowaniu** uzyskamy, oprócz podstawowych danych o próbie, wartości statystyki testowej i wartość krytyczną p dla testu z hipotezą alternatywną obustronną.

W zależności od kierunku hipotezy jednostronnej wartość krytyczna wynosi $\frac{p}{2}$ lub $1 - \frac{p}{2}$.

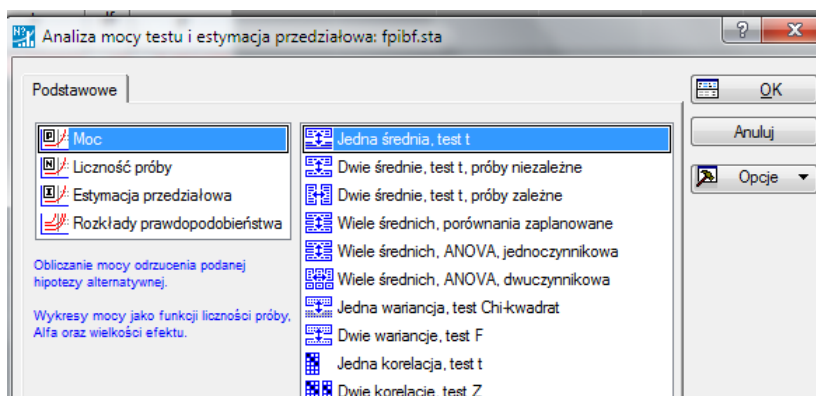


Jak uzyskać wartość krytyczną dla testu jednostronnego?

OBLICZANIE MOCY TESTU



W zakładce **statystyka** wybieramy **analizę mocy testu**, następnie wskazujemy rodzaj testu (**jedna średnia test t**).



Do analizy potrzebujemy kolejno: liczności próby (N), wartości względem której testujemy średnią (Mi0), średnią z próby (Mi), poziom istotności (alfa) oraz odchylenie standardowe z próby (sigma).

Należy wybrać odpowiednio do zadania hipotezę zerową, która określona jest tutaj przez zaprzeczenie hipotezy alternatywnej.

Zasady prezentowania statystyk opisowych.

Dla danych o rozkładzie normalnym (zbliżonym do rozkładu normalnego) prezentujemy średnią i odchylenie standardowe, co zapisujemy: *mean (SD)*, NIE $\text{mean} \pm SD$, oraz licznosc próby.

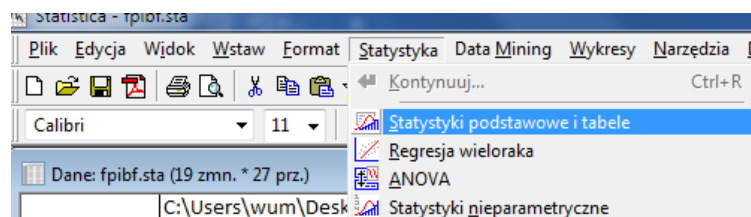
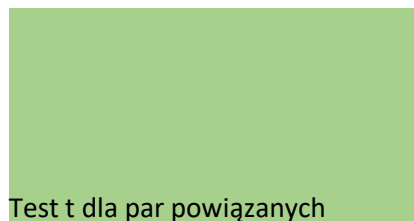
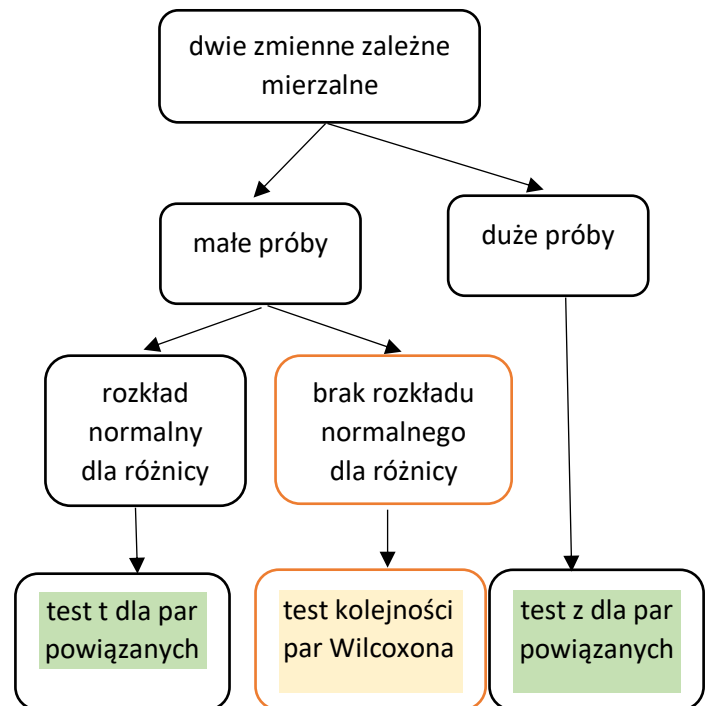
! *Przedstawiając statystykę opisową nie podajemy SE, SE- błąd standardowy - to parametr dotyczący nie surowych danych, lecz średniej: $SE = \frac{SD}{\sqrt{n}}$*

Dla danych, których rozkład nie jest zbliżony do rozkładu normalnego, podajemy medianę (*median*), dolną i górną wartość percentyli (*the lower and upper boundaries of interpercentile range*) lub ich różnicę, czyli rozstęp percentylowy (*interpercentile range*), zakres danych (*minimum and maximum values of range*) i licznosc próby.

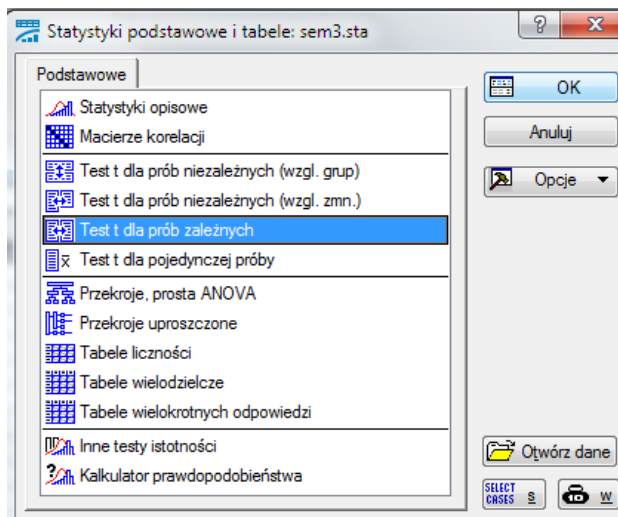
Statystyka opisowa może być prezentowana w postaci tabeli **lub** rysunków.

DWIE ZMIENNE

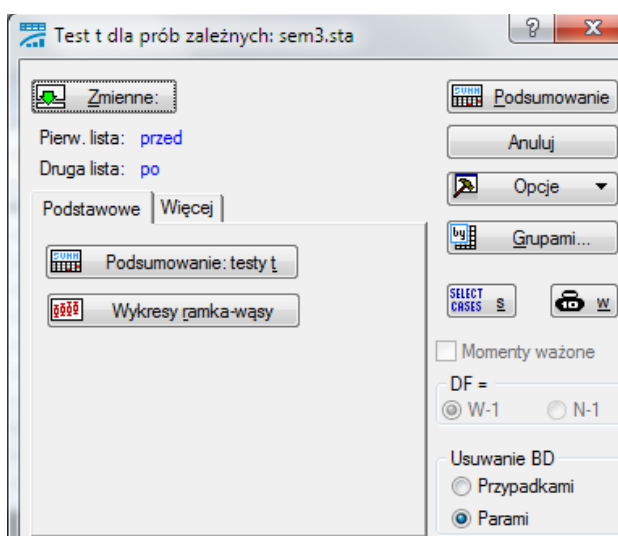
- Testowanie hipotez dla dwóch zmiennych zależnych;
- Analiza mocy testu dla dwóch zmiennych zależnych. Minimalna liczność próby;
- Testowanie hipotez dla dwóch zmiennych niezależnych;
- Analiza mocy testu dla dwóch zmiennych niezależnych. Minimalna liczność próby.



W zakładce **Statystyka** wybieramy **statystyki podstawowe i tabele**, a dalej **test t dla prób zależnych**.



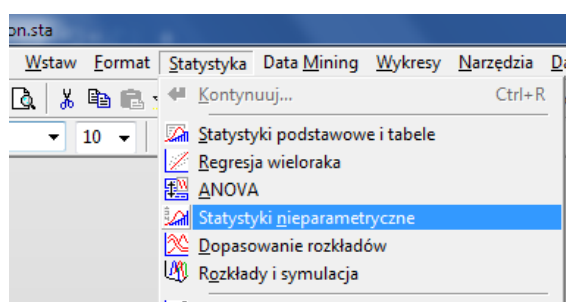
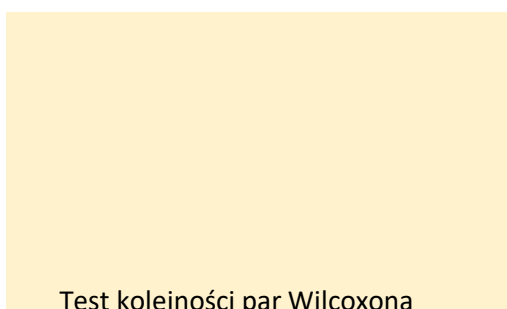
Należy wskazać dwie **zmienne** do analizy.



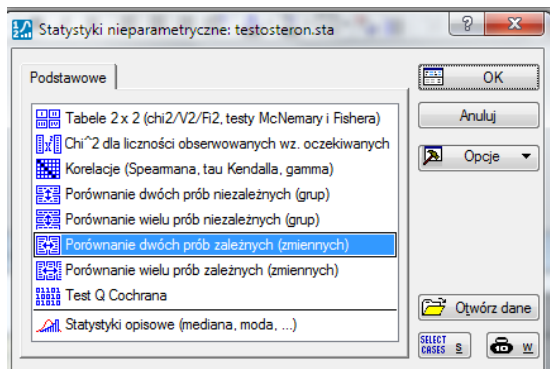
Podsumowanie oprócz podstawowych parametrów próby (liczności, średnie, odchylenia standardowe z próby) podaje wartość statystyki testowej oraz wartość krytyczną dla hipotezy alternatywnej obustronnej.

ANALIZA MOCY TESTU wymaga wybrania testu **dwie średnie, test t, próby zależne** i podania: liczności próby (N), wartości średniej pierwszej grupy (M_{i1}), średniej drugiej grupy (M_{i2}), poziomu istotności (α) oraz odchylenia standardowego z pierwszej próby (σ_1) oraz drugiej (σ_2), jak również współczynnika korelacji Pearsona (r_0).

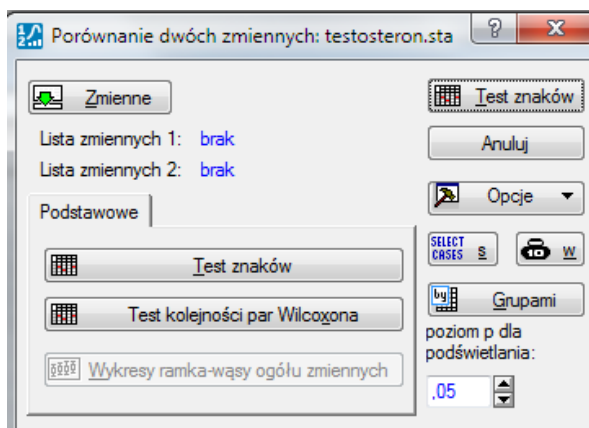
Należy wybrać odpowiednio do zadania hipotezę zerową, która określona jest tutaj przez zaprzeczenie hipotezy alternatywnej.



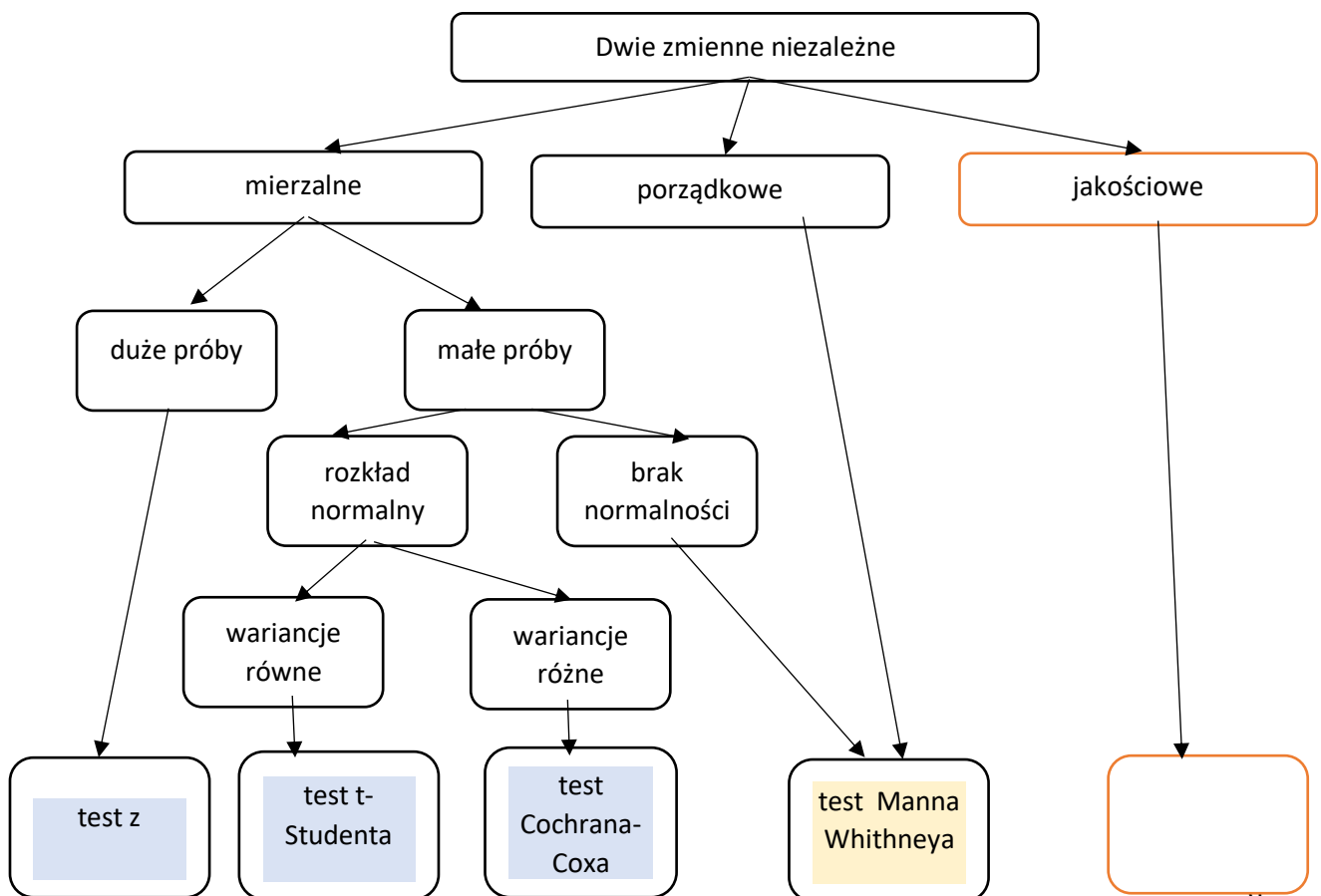
W zakładce **Statystyka** wybieramy **statystyki nieparametryczne**, a dalej **porównanie dwóch prób zależnych**.



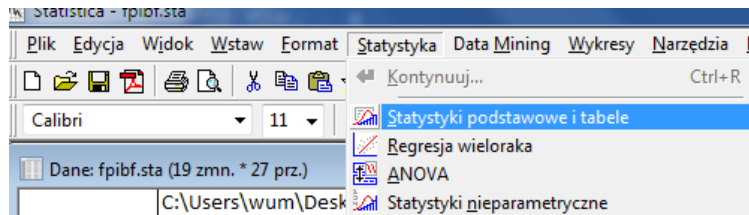
Należy wskazać **zmienne** do analizy.



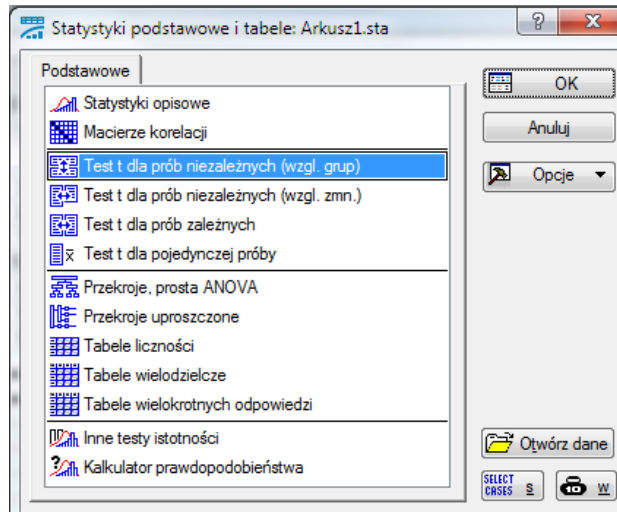
Test kolejności par Wilcoxona podaje oprócz liczności próby, wartość statystyki testowej oraz wartość p dla testu obustronnego.



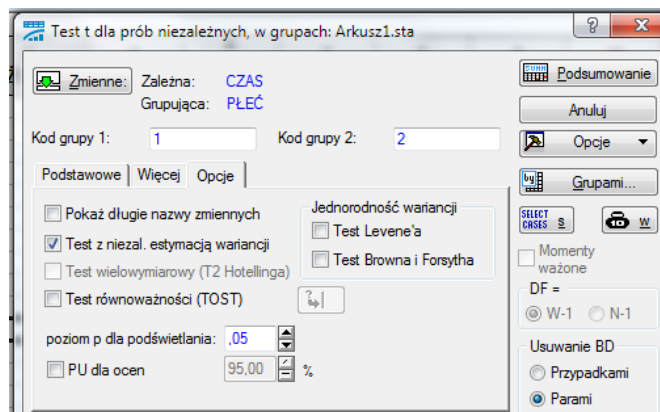
Test dla prób niezależnych



W zakładce **Statystyka** wybieramy **statystyki podstawowe i tabele**, a dalej **test t dla prób niezależnych**. W zależności od sposobu zapisywania zmiennych wybieramy opcję **względem grup** (preferowana) lub **względem zmiennych**.



W kolejnym kroku należy wskazać **zmienne** do analizy i w zakładce **opcje** zaznaczyć **test z niezależną estymacją wariancji**.



Na podstawie **testu dotyczącego równości wariancji**, wiemy jaki test zastosować: test t czy Cochran-Coxa. Jeśli $p >$ poziomowi istotności, to przyjmujemy, że wariancje są równe i stosujemy **test t**, w przeciwnym przypadku **test Cochran-Coxa**.

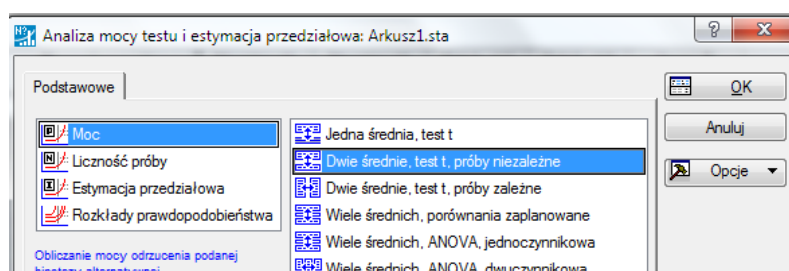
Uwaga: dla małych prób trzeba sprawdzać normalność

W podsumowaniu otrzymamy tabelę złożoną z trzech modułów.

Grupa 1 wz. Grupy 2	Testy dla prób niezależnych (tydzień2.sta)							
	Uwaga: Zmienne traktowane są jako niezależne próby.							
	Średnia Grupa 1	Średnia Grupa 2	t	df	p	t oddz. est. war.	df	p dwustronny
cytokiny vs. cytokiny	41,54545	31,00000	1,821056	20	0,083593	1,821056	19,78033	0,083760

Grupa 1 wz. Grupy 2	Testy dla prób niezależnych (tydzień2.sta)					
	Uwaga: Zmienne traktowane są jako niezależne próby.					
	N ważnych Grupa 1	N ważnych Grupa 2	Odch.std Grupa 1	Odch.std Grupa 2	iloraz F Wariancje	p Wariancje
cytokiny vs. cytokiny	11	11	14,27840	12,84523	1,235592	0,744462

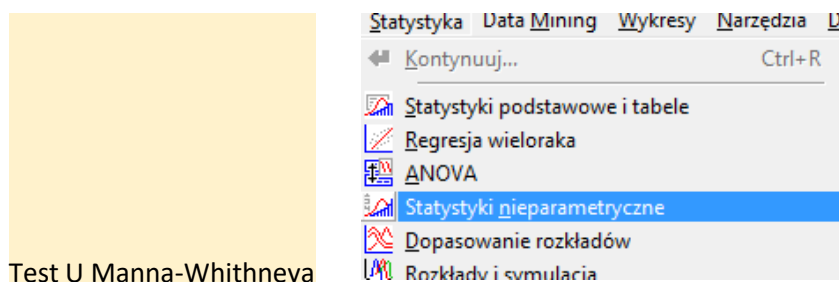
ANALIZA MOCY TESTU wymaga wybrania testu **dwie średnie, test t, próby niezależne**



W kolejnym kroku należy podać: wartość średniej pierwszej grupy (M_1), średniej drugiej grupy (M_2), licznosc pierwszej grupy (N_1), drugiej grupy (N_2), poziom istotności (alfa) oraz tzw. wspólne odchylenie z populacji (sigma) liczone wzorem:

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

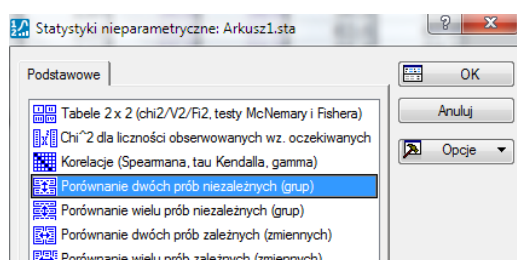
Należy wybrać odpowiednio do zadania hipotezę zerową, która określona jest tutaj przez zaprzeczenie hipotezy alternatywnej.



Test U Manna-Whitneya

W przypadku, gdy dane nie podlegają rozkładowi normalnemu, stosujemy test Manna-Whitneya.

W zakładce **statystyka** wybieramy **statystyki nieparametryczne**, dalej **porównanie dwóch prób niezależnych (grup)**.



Wskazujemy **zmienną zależną** i **grupującą**. W wyniku przeprowadzonego testu otrzymujemy tabelę, w której mamy wartość krytyczną testu dla dużej próby (p), skorygowaną wartość krytyczną, gdy występują rangi wiązane (p) oraz wartość krytyczną dla małej próby ($2*1$ str. dokł. p).

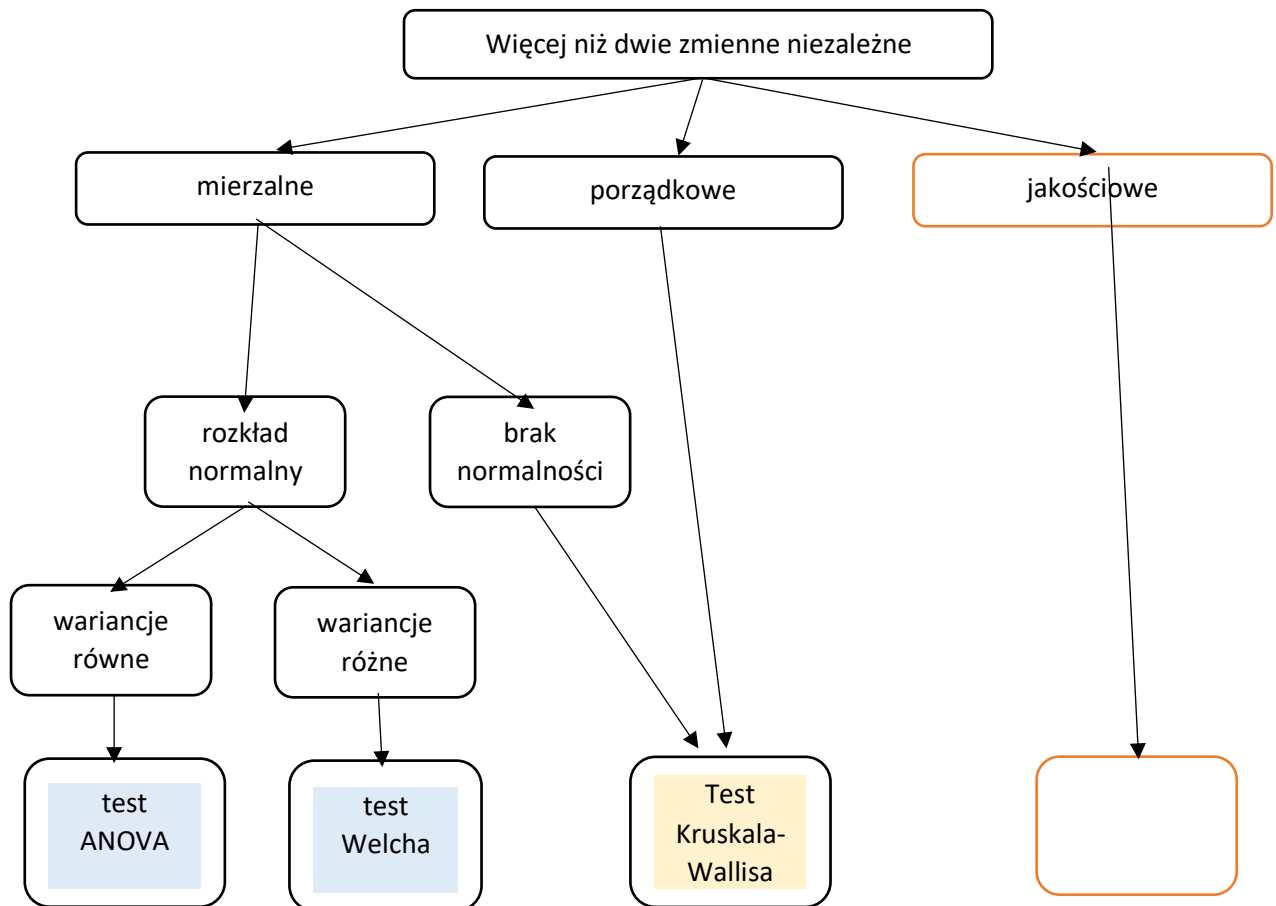
Sum.rang Grupa 1	Sum.rang Grupa 2	U	Z	p	Z popraw.	p	N ważn. Grupa 1	N ważn. Grupa 2	2*1str. dokł. p
103,5000	127,5000	37,50000	-1,19710	0,231267	-1,20774	0,227149	11	10	0,223024

Gdy nie dysponujemy danymi surowymi, a mamy tylko podstawowe parametry prób (średnia, wariancja z próby, licznosc) możemy skorzystać z *opcji inne testy istotności* → *różnica między średnimi*. WYMAGANA JEST NORMALNOŚĆ ROZKŁADU.

Zasady prezentowania wyników testów statystycznych:

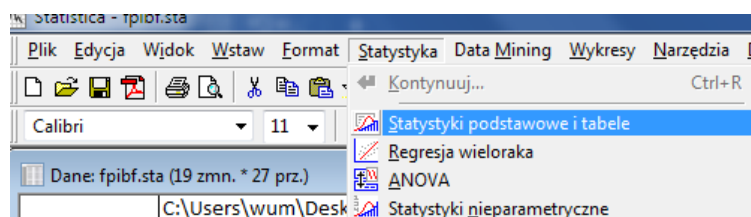
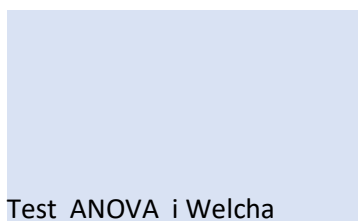
- Podajemy jaki typ hipotezy (jednostronna czy obustronna) testujemy;
- Wartość p podajemy używając równości, np. $p = 0,03$, a nie $p < 0,05$. Nie stosujemy zapisu NS. Zawsze podajemy konkretną wartość p . Dla p mniejszych niż 0,001 możemy użyć zapisu $p < 0,001$.

TRZY ZMIENNE

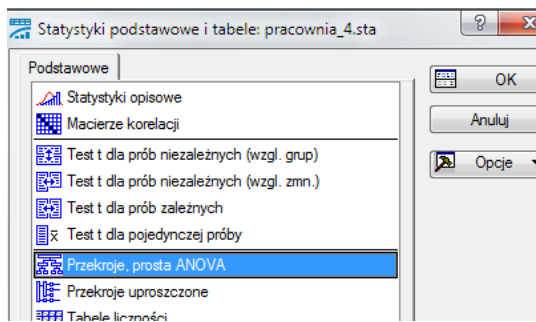


UWAGA: dla równolicznych grup założenie normalności i jednorodności wariancji nie jest niezbędne. **NALEŻY JEDNAK UPEWNIĆ SIĘ, ŻE NIE MA DANYCH ODSTAJĄCYCH I ŻE NIE ISTNIEJE KORELACJA MIĘDZY ŚREDNIMI I ODCHYLENIEM STANDARDOWYM.**

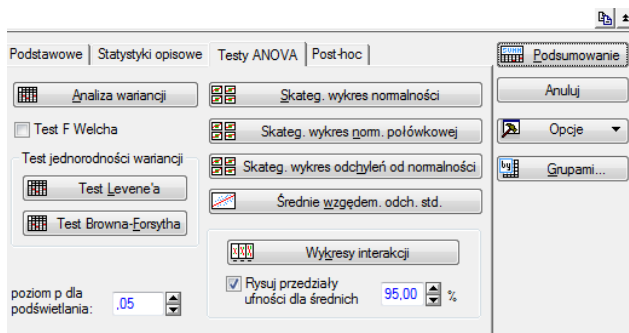
anova → założenia → średnie wzgl. odch. stand.



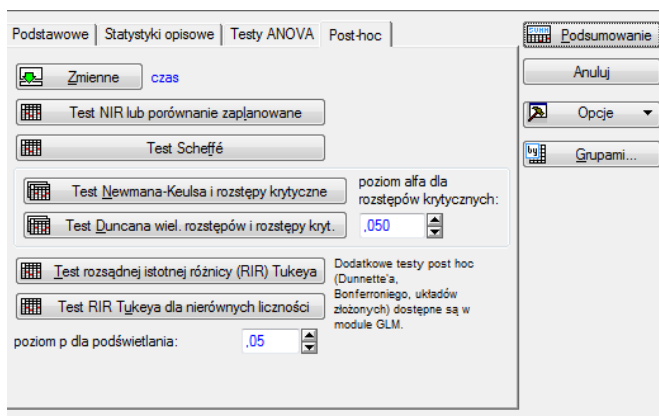
W zakładce **Statystyka** wybieramy **statystyki podstawowe i tabele**, a dalej **przekroje, prosta ANOVA**.



W kolejnym kroku podajemy **zmienne** do analizy: **zależną** i **grupującą**.



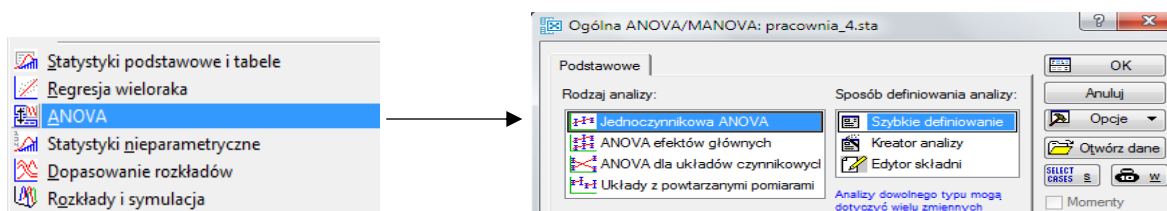
Wybieramy zakładkę **Testy ANOVA**, gdzie znajdziemy test sprawdzający jednorodność wariancji. W wypadku, gdy ostatni test wykaże brak równości wariancji, należy doznaczyć **test F Welcha** i wybrać **analizę wariancji**.



Jeśli analiza wariancji odrzuci hipotezę o braku różnic pomiędzy średnimi w grupach, należy przeprowadzić test post-hoc.

UWAGA: należy sprawdzić na początku założenie o normalności rozkładu w grupach.

ANALIZĘ MOCY TESTU najwygodniej jest przeprowadzić wybierając w zakładce **statystyka** opcję **ANOVA**. Następnie wybieramy **jednoczynnikową ANOVĘ** i wskazujemy zmienne.



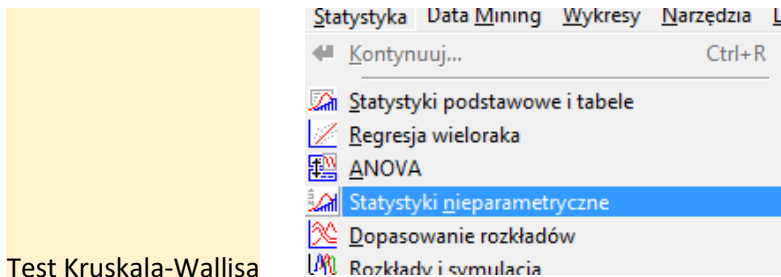
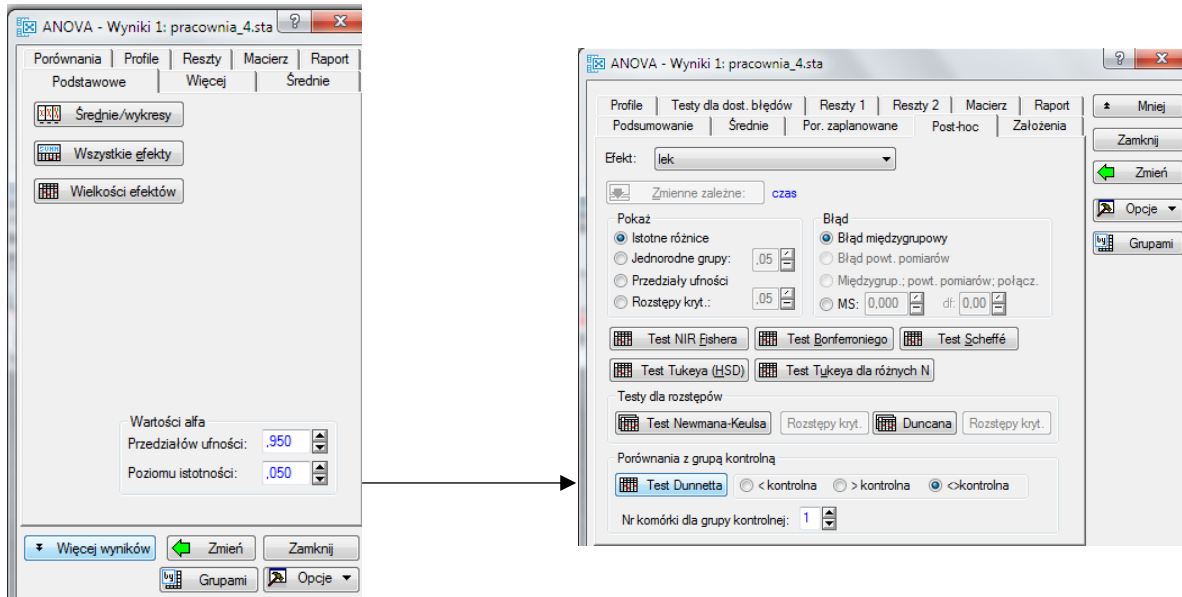
W celu obliczenia mocy testu wybieramy opcję **wielkości efektów**

UWAGA: Korzystając z opcji **analiza mocy testu** w zakładce **statystyka** wymagane jest podanie RMSSE Root Mean Square Standardized Effect.

Dla grup równolicznych RMSSE można obliczyć jako $\sqrt{\frac{F}{n_i}}$.

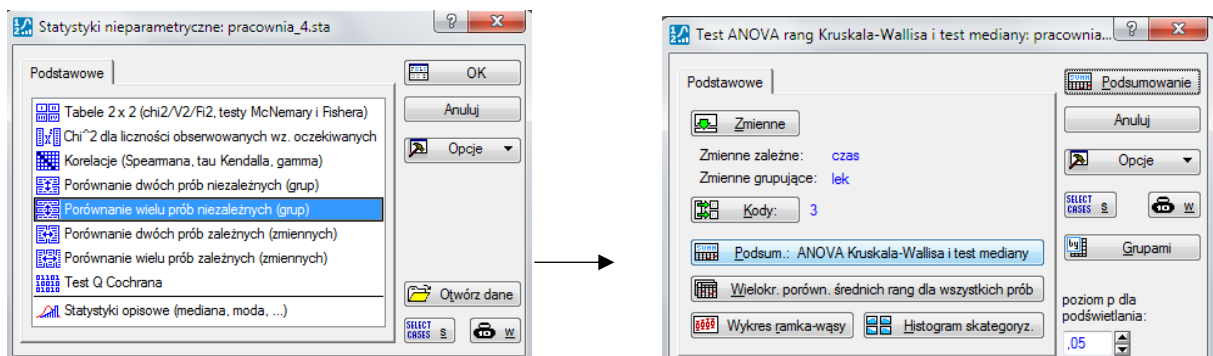
Opcja ta daje możliwość wyliczenia minimalnych licznosci grup dla testu anova.

Dalej znajduje się **test Dunnetta** do porównań z grupą kontrolną. Należy wybrać opcję **więcej wyników**, a następnie zakładkę **post-hoc**.



Test Kruskala-Wallisa

W przypadku, gdy dane w grupach nie podlegają rozkładowi normalnemu, stosujemy test Kruskala-Wallisa. W zakładce **statystyka** wybieramy **statystyki nieparametryczne**, dalej **porównanie wielu prób niezależnych (grup)**. Wskazujemy zmienne do analizy, a następnie wybieramy **Podsum.: ANOVA Kruskala-Wallisa i test mediany**.



Test Dunna (post-hoc) znajdziemy wybierając **wielokr. porówn. średnich rang dla wszystkich prób**.

REGRESJA PROSTA, KORELACJA

- Regresja liniowa – metoda najmniejszych kwadratów;
- Współczynnik korelacji Pearsona;
- Współczynnik korelacji Spearmana.

Regresja liniowa prosta zakłada, że zmienna Y, która jest zmienną zależną mierzalną, jest funkcją liniową zmiennej X, która jest zmienną niezależną również mierzalną, tzn.

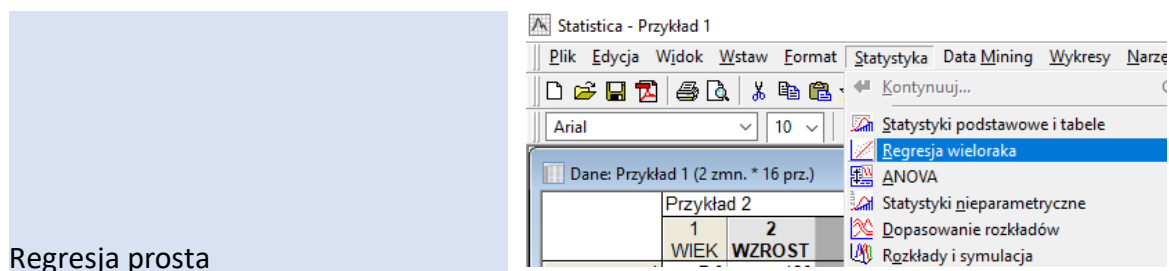
$$y_i = b_0 + b_1x + e_i,$$

gdzie e_i jest składnikiem losowym, nazywanym resztą.

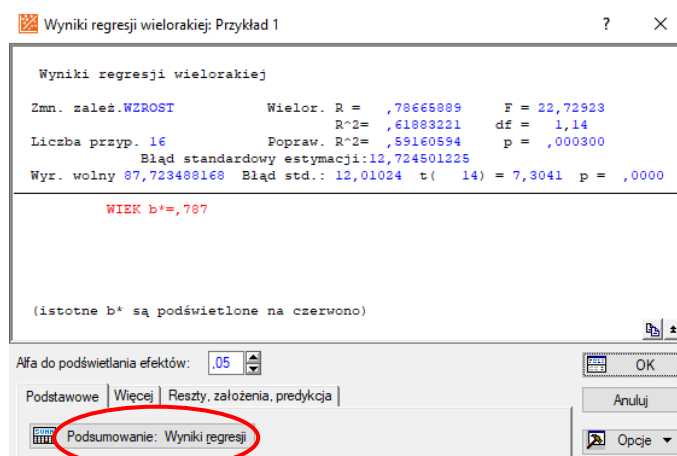
Procedura konstrukcji regresji liniowej prostej:

1. Rysujemy wykres rozrzutu punktów, aby sprawdzić czy zależność jest liniowa;
2. Wyznaczamy parametry modelu (współczynniki prostej regresji wraz z błędami estymacji, współczynnik determinacji oraz błąd standardowy estymacji).
3. Sprawdzamy założenia analizy regresji;

UWAGA: w razie nie spełnienia założeń, poprawiamy model (transformujemy zmienne, usuwamy dane odstające, itd.) i procedurę powtarzamy.



W zakładce **Statystyka** wybieramy **regresję wieloraką**. W kolejnym kroku wybieramy **zmienne** do analizy: zmienną **zależną** i **niezależną**.



Aby otrzymać potrzebne informacje wybieramy **Podsumowanie: Wyniki regresji**. W skrócie pojawią się wówczas dwie karty. Na pierwszej z nich odczytamy wyniki testu F, który sprawdza czy regresja jest właściwym modelem, na

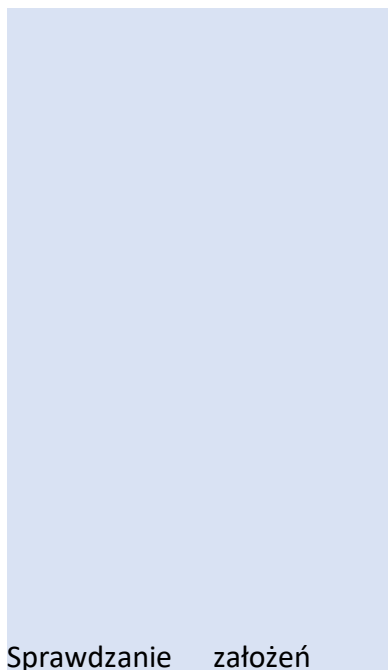
drugiej współczynniki prostej regresji (w kolumnie b) wraz z błędami estymacji (kolumna Bł. std. z b), współczynnik determinacji oraz błąd standardowy estymacji.

Założenia modelu:

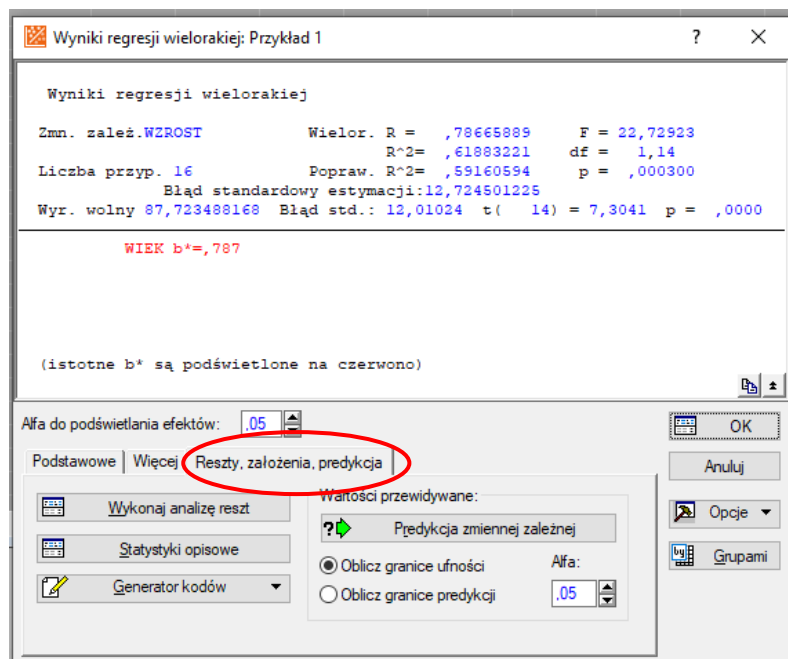
- Model jest liniowy względem parametrów – należy najpierw przestudiować wykres rozrzutu punktów.

W wypadku oczywistej nieliniowości należy transformować zmienną/zmienne **lub** zastosować techniki nieliniowe.

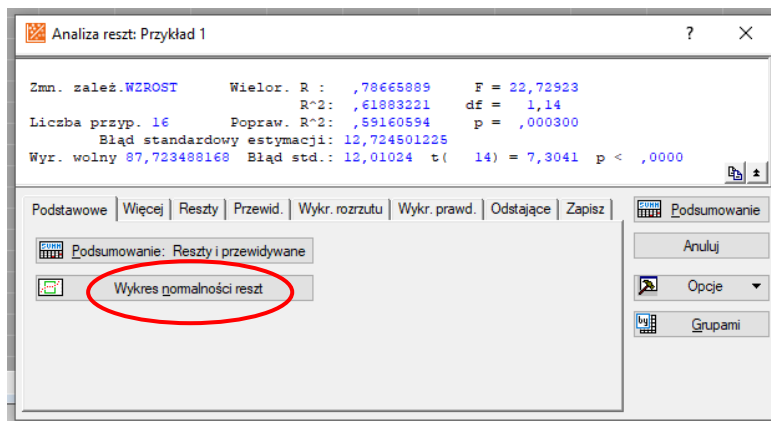
- Liczba obserwacji musi być odpowiednio duża. Im większa liczba obserwacji, tym większa precyzja oszacowań.
- Czynniki nieuwzględnione w modelu nie wpływają istotnie na zmienną Y, co oznacza, że składnik losowy ma wartość oczekiwaną równą zero.
- Wariancja składnika losowego jest taka sama dla wszystkich obserwacji.
- Reszty są nieskorelowane – dotyczy głównie szeregów czasowych;
- Reszty modelu mają rozkład normalny.



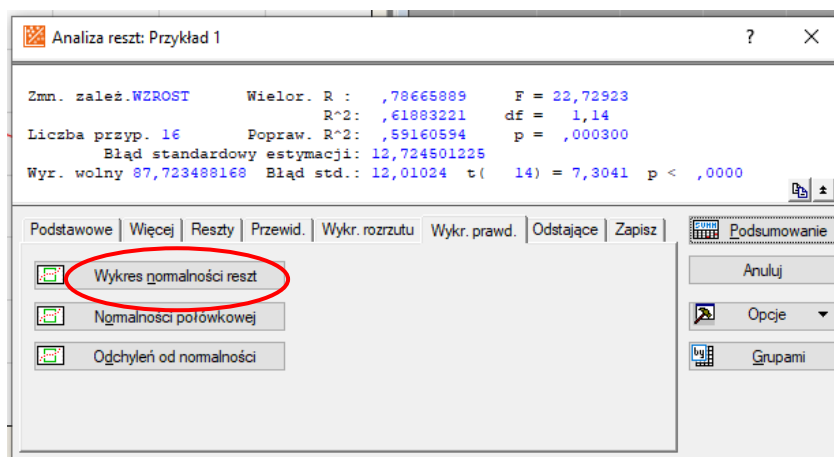
Sprawdzanie założeń



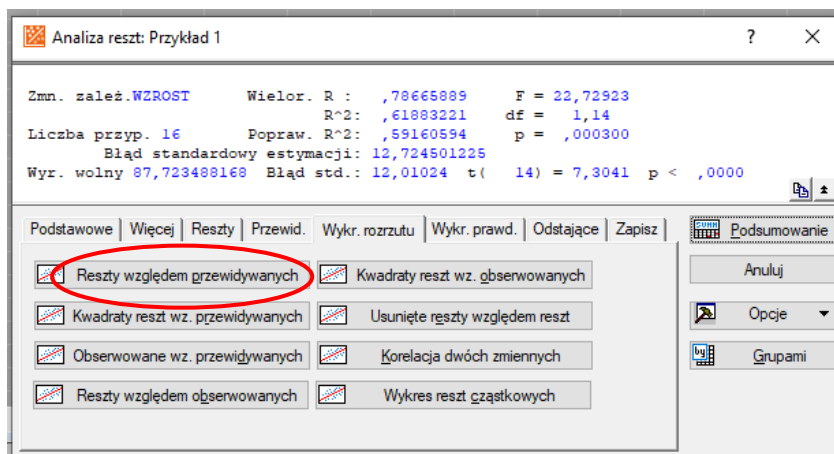
Wybieramy kartę **Reszty, założenia, predykcja**, a dalej **Wykonaj analizę reszt**.



Zakładka **Podstawowe** pozwala zweryfikować normalność reszt. Możemy to zrobić wzrokowo wybierając **wykres normalności reszt** lub, w razie wątpliwości, testem Shapiro-Wilka. W tym celu należy wybrać **Podsumowanie: Reszty i przewidywane**, w kolumnie reszty znajdziemy interesujące nas dane. (Test Shapiro-Wilka – patrz seminarium 1).

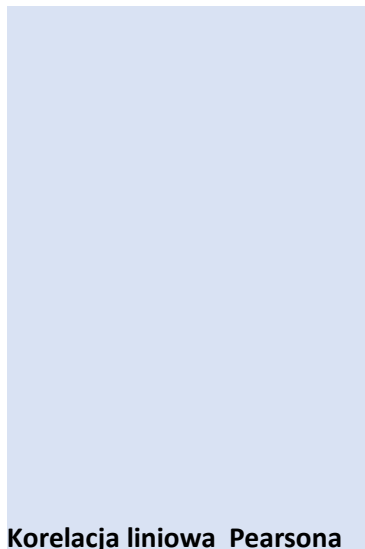


Wykresy pozwalające zweryfikować normalność reszt znajdziemy również w zakładce **wykr.prawd**.

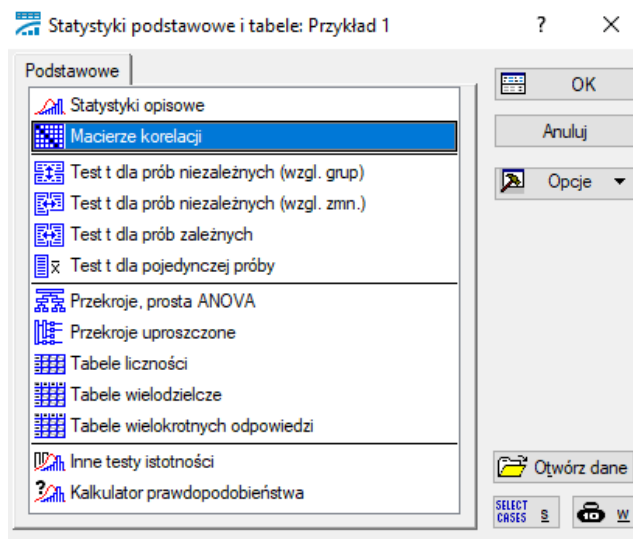


W celu sprawdzenie założenia, że wariancja składnika losowego jest taka sama dla wszystkich obserwacji wybieramy zakładkę **wykr. rozrzutu**. Sporządzenie wykresu **reszty względem przewidywanych (kwadraty reszt wz. przewidywanych)** pozwala graficznie sprawdzić czy nie występuje jakaś tendencja/regularność, tzn. czy reszty nie są bardziej zróżnicowane dla jednych wartości, a mniej dla innych.

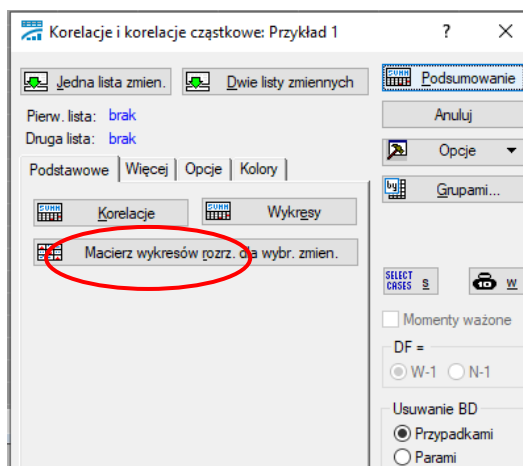
Korelacja Pearsona i Spearmana.



Korelacja liniowa Pearsona



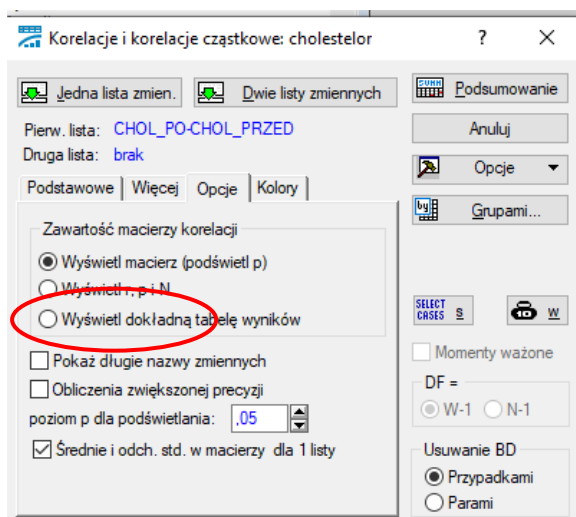
W zakładce **Statystyka** wybieramy **statystyki podstawowe i tabele**. W kolejnym kroku wybieramy **macierze korelacji**.



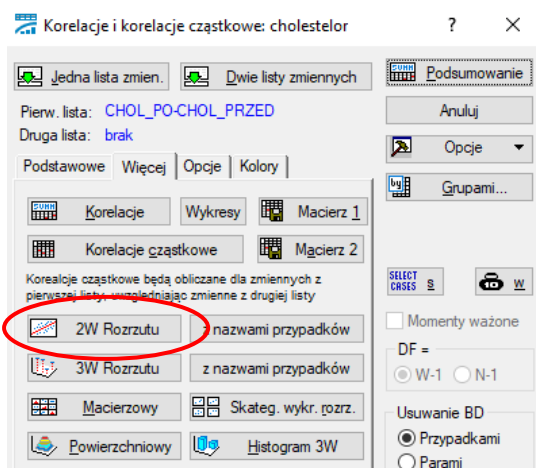
Do wyboru mamy dwie metody wyboru zmiennej do analizy: **jedna lista zmiennych** lub **dwie listy zmiennych**. W pierwszym przypadku pakiet wyliczy współczynniki korelacji dla wszystkich możliwych par zmiennych utworzonych z naszej listy zmiennych. W drugim przypadku otrzymamy tylko korelacje pomiędzy zmiennymi z pierwszej listy, a zmiennymi z drugiej listy.

Po wskazaniu zmiennych do analizy wybieramy przycisk **korelacje**. W efekcie otrzymujemy tabelę, która oprócz średniej i odchylenia każdej ze zmiennych podaje współczynniki korelacji. Korelacje istotne statystycznie (na poziomie 0,05) zaznaczone są na czerwono.

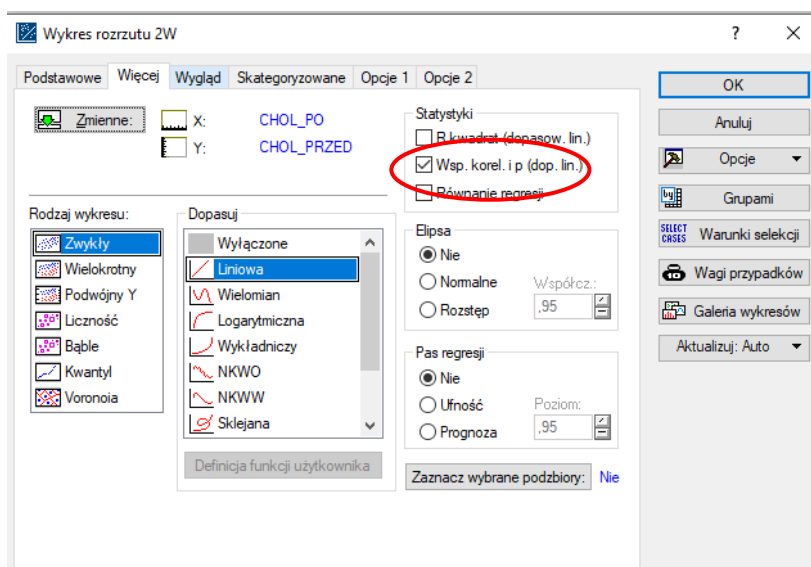
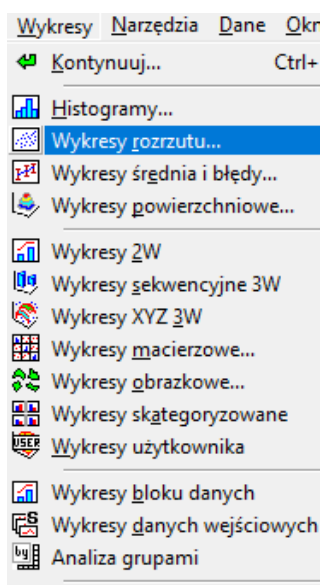
Jeśli chcemy zobaczyć dodatkowe informacje (wartość p, licznosc próby) należy na karcie **opcje** wybrać **wyświetl r, p, N**. Tam również można zmienić poziom istotności.



Dodatkowo należy się upewnić, że punkty wykazują liniowy trend. W tym celu na karcie **więcej** należy wybrać opcję **2W Rozrzutu** lub **wykresy**.

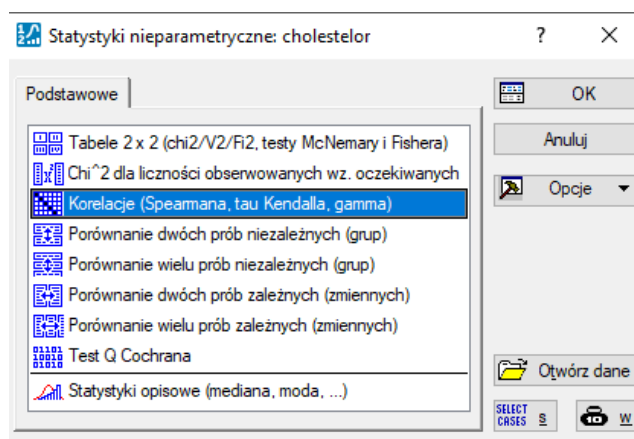


Podobne efekty możemy otrzymać wybierając na karcie **wykresy** opcję **wykresy rozrzutu**.



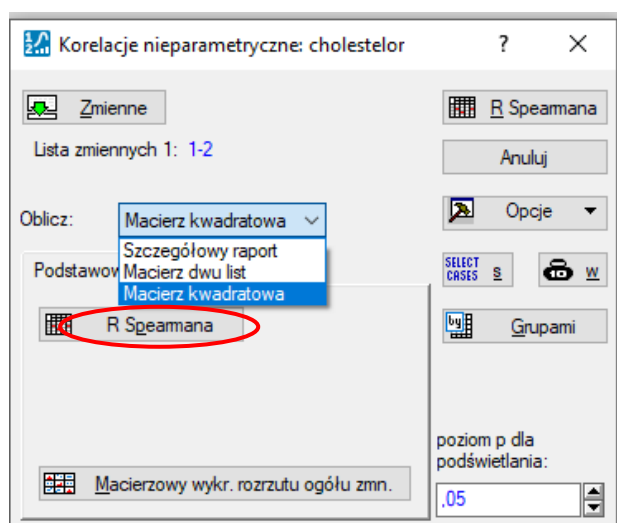
Na karcie **więcej** należy zaznaczyć opcję **wsp. korel. i p (dop. lin.)**. W tej wersji można od razu ocenić czy dane wykazują liniowy trend.

Korelacja Spearmana



Wybieramy kartę **statystyka**, następnie **statystyki nieparametryczne**. Po wybraniu opcji **korelacje** należy wskazać zmienne do analizy. Domyślnie ustawiona jest jedna lista zmiennych (**macierz kwadratowa**). Można to zmienić wybierając **macierz dwu list**.

Wybranie przycisku **R Spearmana** daje współczynniki korelacji rangowej Spearmana. Jeśli współczynniki są istotne statystycznie wyświetlane są w kolorze czerwonym. Aby otrzymać wartości p należy wybrać opcję **szczegółowy raport** (podaje licznosc, wartość współczynnika, wartość statystyki testowej oraz p).



PRZYKŁADY I ZADANIA

Przykład 1.1 Do danych z pliku AUC.sta zastosuj test Shapiro-Wilka. Następnie wprowadź nową zmienną, której wartościami będą zlogarytmowane wartości AUC. Dane → przekształcanie wielu zmiennych. Zastosuj ponownie test Shapiro Wilka do przekształconych danych. W następnej kolejności wprowadź nową zmienną, uzyskaną przez przekształcenie Boxa-Coxa (Dane → przekształcenie Boxa-Coxa) zmiennej AUC i dla niej ponownie zweryfikuj hipotezę o normalności rozkładu.

Zadanie 1.1 Przeprowadź kolejno analizy wymienione w powyższym przykładzie dla danych z pliku zadanie1.sta uwzględniając zmienną grupującą – lek.

Przykład 1.2. Korzystając z pliku fish.xlsx:

- Włączyć opcję selekcja przypadków, wybierając do analizy dane dotyczące rybek hodowanych w preparacie Phenol;
- Sprawdzić czy zmienna liczba uderzeń serca na minutę zawiera dane odstające. Analizę przeprowadź graficznie z wykorzystaniem odstępu kwartylowego, zasady trzech sigma oraz za pomocą testu Grubbsa na poziomie istotności 0,05 (wcześniej należy zweryfikować normalność rozkładu).

Zadanie 1.2 Korzystając z pliku CPK.sta sprawdzić dane odstające włączając selekcję przypadków – wybierz dane z grupy BUDOWA.

Zadanie 1.3. W losowej próbie 10 pacjentów leczonych w szpitalu na pewną chorobę zakaźną, badano czas trwania tej choroby [dni] otrzymując wyniki: 30, 15, 25, 20, 30, 25, 55, 35, 30, 30. Zweryfikować przypuszczenie, że średni czas choroby nie przekracza 30 dni. Przyjąć poziom istotności 0,05. Sprawdzić moc testu. Dobrać n tak, by moc przekraczała 80%. (Sprawdzić wszystkie potrzebne założenia).

Zadanie 1.4. W pewnym eksperymencie roztwór substancji został infuzyjnie podany 10 znieczulonym świnkom morskim. Wielkości dawki śmiertelnej prezentują poniższe wyniki: 29, 23, 26, 19, 25, 17, 21, 25, 20, 19. Na poziomie istotności 0,05 zweryfikuj przypuszczenie, że wartość przeciętna dawki śmiertelnej jest większa niż 20. Czy wynik testu zmieniłby się, gdyby hipoteza alternatywna była obustronna? Sprawdzić moc testu. Dobrać n tak, by moc przekraczała 80%.

Przykład 1.3 Korzystając z pliku heartdisease.xlsx:

- Sporządź na jednym rysunku histogramy dla systolic blood pressure w zależności od tabaco intake level.
- Sporządź na jednym rysunku wykresy ramka-wąsy z zaznaczoną średnią odchyleniem standardowym i zakresem min-max dla zmiennej systolic blood pressure w zależności od tabaco intake level.

- Sporządź histogram 3-wymiarowy dla zmiennych niezależnych: *tabaco intake level*, *alcohol intake level*
- Sporządź wykres powierzchniowy 3-wymiarowy dla zmiennych niezależnych: *tabaco intake*, *alcohol intake* i zmiennej zmiennej zależnej *systolic blood pressure*.
- Dla zmiennych z poprzedniego podpunktu sporządź wykres macierzowy.

Zadanie 1.5. Korzystając z pliku *I_rok_stat_do_analizy.xlsx*

- zweryfikuj przypuszczenie, że studenci, którzy w szkole średniej mieli matematykę na poziomie rozszerzonym uzyskują wyniki powyżej 85%. Oblicz moc tego testu.
- Sporządź na jednym rysunku dwa czytelne histogramy dla sumy uzyskanych punktów w semestrze I i semestrze II dla osób po matematyce podstawowej i rozszerzonej.
- Sporządź na jednym rysunku wykresy ramka-wąsy z zaznaczoną średnią odchyleniem standardowym i zakresem min-max dla zmiennej wynik.
- Sporządź histogram 3-wymiarowy dla zmiennych niezależnych: ocena i poziom matury.
- Sporządź wykres powierzchniowy 3-wymiarowe dla zmiennych niezależnych: suma kartkówek z I semestru, suma kartkówek z II semestru i zmiennej zmiennej zależnej wynik końcowy.
- Dla zmiennych z poprzedniego podpunktu sporządź wykres macierzowy.

Przykład 2.1. Przeprowadzono badanie poziomu beta-endorfiny w warunkach stresu związanego ze zbliżającym się zabiegiem chirurgicznym. W grupie 8 pacjentów dokonano pomiaru poziomu beta-endorfiny (w fmol/ml) na 12-14 godzin przed zabiegiem (A) oraz na 10 minut przed zabiegiem (B). Otrzymano następujące dane:

A	10	6,5	8	12	5	11,5	4,7	8
B	6,5	14	13,5	18	14,5	9	25	12

- Wprowadzić dodatkową zmienną będącą różnicą tych dwóch zmiennych. (Dane->przekształcanie wielu zmiennych)
- Sprawdzić czy ta nowa zmienna nie ma danych odstających.
- Sprawdzić czy ta nowa zmienna ma rozkład normalny.
- Zweryfikować hipotezę, że czas ma istotny wpływ na wzrost poziomu beta-endorfin. Zastosować test dla dwóch zmiennych zależnych.
- Zweryfikować hipotezę, że czas ma istotny wpływ na wzrost poziomu beta-endorfin. Zastosować test dla pojedynczej próby.
- Sprawdzić moc ostatniego testu.
- Sprawdzić moc testu dla dwóch zmiennych zależnych

Zadanie 2.1. W celu zbadania czy zastosowany lek powoduje istotny spadek ciśnienia tętniczego w grupie osób chorujących na nadciśnienie wybrano losowo 10 pacjentów. Tabela przedstawia poziom ciśnienia przed i po zastosowaniu leku.

pacjent	1	2	3	4	5	6	7	8	9	10
przed leczeniem	220	185	270	285	200	295	255	190	225	230
po leczeniu	190	175	215	260	215	195	260	150	155	175

Przyjmując poziom istotności 0,05 zweryfikuj odpowiednią hipotezę wykorzystując najpierw test

dla dwóch zmiennych zależnych, a następnie obliczając różnicę ciśnień dla każdego pacjenta testem dla jednej zmiennej. Sprawdź potrzebne założenia. Porównaj otrzymane wartości p. Sprawdź moc testów. Dobierz n tak, by moc przekraczała 80%.

Zadanie 2.2. Badano poziom CD4 przed leczeniem i po leczeniu pewnym lekiem przeciwwirusowym (plik **przed_po_test_cw_2.sta**). Przyjmując poziom istotności 0,05 zweryfikuj hipotezę, że występują istotne różnice badanego parametru wykorzystując najpierw test dla dwóch zmiennych zależnych, a następnie obliczając odpowiednie różnice, dla jednej zmiennej. Sprawdź potrzebne założenia. Porównaj otrzymane wartości p. Sprawdź moc testów. Dobierz n tak, by moc przekraczała 80%.

Przykład 2.2. Oceniano wpływ dodatków do pożywienia na masę zwierząt. Na poziomie istotności 0,05 sprawdź czy stosowanie dodatków zwiększa masę zwierząt. Przeanalizuj moc testu. Plik *Animal Weights.sta*

Zadanie 2.3. Badano poziom agresji 12-tu chłopców i 12-tu dziewczynek podczas 15-to minutowej sesji. Sprawdzić na poziomie istotności 0,05 czy różnice pomiędzy wartościami oczekiwanymi poziomu agresji pomiędzy chłopcami i dziewczynkami są istotne statystycznie. Przeanalizuj moc testu. Plik *aggressn.sta*.

Zadanie 2.4. Badano wpływu rodzaju podręcznika na umiejętności matematyczne. Sprawdzić na poziomie istotności 0,05 czy podręcznik ma wpływ na umiejętności matematyczne. Przeanalizuj moc testu. Plik *textbook.sta*.

Zadanie 2.5. Porównujemy poziom wskaźnika ścieków wiosną i zimą (wskaźnik sciekow.sta)

z	104,2	67,2	63	67,2	43,6	48,8	59,1	45,3					
w	77	66,8	72,7	338,1	68,5	138,7	116,5	94,8	78,2	63,2	119,8	106,5	96,5

Na poziomie istotności 0,05 sprawdzić, czy istnieją różnice pomiędzy grupami.

Zadanie 2.6. Badano zależność długości snu od palenia papierosów w pewnej populacji. Wybrano losowo grupę 15 osób palących i 12 niepalących. Wyznaczono średnie i odchylenia standardowe długości snu w obu grupach, otrzymując

	niepalący	palący
średnia z próby	8,2	7,0
odchylenie standardowe z próby	2,2	3,8

Zakładając, że dane w obydwu grupach podlegają rozkładowi normalnym, zweryfikować przypuszczenie, że osoby palące śpią krócej. Przyjąć poziom istotności 0,05.

Zadanie 2.7 Czy piątek 13. Jest istotni dniem pechowym? Plik *piatek13.sta*. Analizę przeprowadź na poziomie istotności 0,05.

Przykład 3.1. Celem badania było sprawdzenie czy dieta bogatsza w wapń jako niefarmakologiczne leczenie podniesionego ciśnienia krwi ma dobroczynny wpływ na funkcję śródbłonna. W tym celu badacze podzielili 7-tygodniowe szczury płci męskiej z jednakowym ciśnieniem na cztery podgrupy: nieleczone szczury na normalnej diecie, nieleczone szczury z dietą wzbogaconą w Ca, leczone na normalnej diecie oraz leczone na diecie wzbogaconej w Ca. Plik wapn.sta zawiera masy 61 szczurów. Na poziomie istotności 0,05 sprawdź, czy istnieją statystycznie istotne różnice w średnich wagach szczurów. Oblicz moc testu.

Zadanie 3.1. W eksperymencie badano na świnkach morskich dawki śmiertelne dla czterech różnych substancji kardiologicznych. Plik dawka smiertelna.sta. Na poziomie istotności 0,05 porównaj wielkości dawek śmiertelnych. Oblicz moc testu.

Zadanie 3.2. Na poziomie istotności 0,05 porównaj średnie masy ciała dla trzech ras szczurów. Plik szczury.sta → anova dwuczynnikowa?

Zadanie 3.3. Badano czterema różnymi metodami czas krzepnięcia osocza krwi dziesięciu losowo wybranym pacjentom. Plik czas krzepniecia.sta. Na poziomie istotności 0,05 porównaj czas krzepnięcia dla każdej metody. Dla jakiej liczności grup moc przekroczy 90%?

Zadanie 3.4. Na poziomie istotności 0,05 porównaj tempo wzrostu ryb w różnych środowiskach z grupą kontrolną. Plik ryby.sta.

Zadanie 3.5. Leki A, B oraz C były stosowane w niezależnych trzech próbach pacjentów cierpiących na bezsenność. Liczby godzin snu pacjentów podczas drugiej nocy po rozpoczęciu leczenia były następujące:

A	3,5	5,7	3,5	11	12,5	3,8
B	6,2	7,1	6,4	4,5	5,1	
C	4,5	11,7	10,8	4,5	6,2	3,8

Na poziomie istotności 0,05 zweryfikować hipotezę czy lek ma wpływ na długość snu.

Zadanie 3.6 Badano wybór nagradzanej odpowiedzi jako funkcji rodzaju bodźca. Plik Kruskal.sta. Na poziomie istotności 0,05 przeprowadź odpowiednią analizę.

Zadanie 3.7.! PROBLEM Z DANYMI ODSTAJĄCYMI – przykład 8.7 sta

Zadanie 4.1. Z populacji dzieci i młodzieży wybrano losowo 16 osób i określono ich wiek i wzrost (plik przykład 1.sta)

- Znajdź prostą regresji zmiennej y względem zmiennej x. Jakim błędem obarczone są współczynniki prostej dla całej populacji?
- O ile wartości przewidywane przez otrzymane równanie różnią się od wartości empirycznych?
- Czy wszystkie parametry modelu są istotne? (test F)

- d. Jaki procent rozproszenia zmiennej wzrost daje się wyjaśnić jej liniową zależnością względem wieku?
- e. Na poziomie ufności 95% znajdź przedział ufności dla predykcji wagi dziecka 11-letniego.
- f. Zweryfikuj hipotezy (poziom istotności 0,05), że współczynniki prostej są różne od 0.
- g. Sprawdź założenie normalności reszt;
- h. Przeanalizuj wykres rozrzutu reszt względem wartości przewidywanych.

Zadanie 4.2. W pewnej klinice u chorych z ostrą i przewlekłą niewydolnością wątroby analizowano powiązania między aktywnością enzymów aminotransferazy (x) a stężeniem amoniaku (y) we krwi. Zebrano wyniki dla 10 losowo wybranych pacjentów. Plik przykład 2.sta.

- a. Znajdź prostą regresji zmiennej y względem zmiennej x. Jakim błędem obarczone są współczynniki prostej dla całej populacji?
- b. O ile wartości przewidywane przez otrzymane równanie różnią się od wartości empirycznych?
- c. Czy wszystkie parametry modelu są istotne? (test F)
- d. Jaki procent rozproszenia stężenia amoniaku daje się wyjaśnić jej liniową zależnością względem aktywności enzymu aminotransferazy.
- e. Zweryfikuj hipotezy (poziom istotności 0,05), że współczynniki prostej są różne od 0.
- f. Sprawdź założenie normalności reszt;
- g. Przeanalizuj wykres rozrzutu reszt względem wartości przewidywanych.

Zadanie 4.3. Badano oddziaływanie leku na parametr biochemiczny krwi – plik przykład3.sta
Zastosuj analizę regresji.

Zadanie 4.4. Wysunięto hipotezę, że związek między czasem leczenia chorych (X) na chorobę układu krążenia a aktywnością pewnego enzymu (Y) w ich organizmie zależy od płci. Losowa próba dała wyniki przedstawione w tabeli 2 (czas leczenia w dniach) (zadania4.xls - czas).

czas leczenia [dni]		1	2	3	4	5	7	10
aktywność enzymu	kobiety	41	44	35	43	35	43	36
	mężczyźni	42	40	37	39	36	35	30

Obliczając współczynniki korelacji dla każdej z grup i badając ich istotność statystyczną na poziomie 0,05 sprawdź tę tezę.

Zadanie 4.5. Badano zależność pomiędzy ocenami uzyskiwanymi na egzaminie przez studentów, a ilorazem inteligencji. Otrzymano następujące dane (zadania4.xls - oceny)

ocena	ndst	dst	db	dst+	dst	db+	ndst	bdb	dst	db
IQ	101	96	111	107	130	114	120	147	130	120

(a) Wyznaczyć rangowy współczynnik Spearmana.

(b) Zweryfikować hipotezę, że istnieje korelacja pomiędzy ilorazem inteligencji IQ, a ocenami z egzaminu. Przyjąć poziom istotności $\alpha = 0,05$.

Zadanie 4.6. Na podstawie kontroli całokształtu pracy zawodowej i kwalifikacji nauczycieli dyrektor szkoły i wizytator wydali opinię o każdym z nauczycieli. Wyniki ujęto w punktach (zadania4.xls - wyniki)

Nauczyciele	A	B	C	D	E	F	G	H	I	J	K
Dyrektor	41	27	35	33	25	47	38	53	43	35	36
Wizytator	38	24	34	29	27	47	43	52	39	31	29

(a) Wyznaczyć rangowy współczynnik Spearmana.

(b) Zweryfikować hipotezę, że istnieje korelacja pomiędzy oceną wizytatora i dyrektora. Przyjąć poziom istotności $\alpha = 0,05$.

Zadanie 4.7 Na podstawie pliku cholesterol.sta znajdź współczynnik korelacji, oraz sprawdź jego istotność statystyczną, pomiędzy zmiennymi z podziałem na diety. Analizę przeprowadź na poziomie istotności 0,05.