

Testy nieparametryczne

Hipotezy nieparametryczne:

Nie mają żadnych wymagań co do rozkładu cechy w populacji (nie wymagają rozkładu normalnego), liczebności próby, czy jednorodności wariancji. Mogą być stosowane do niedużych prób

Ich moc wnioskowania jest słabsza niż testów parametrycznych.

Na ogół, po zastosowaniu hipotezy nieparametrycznej, weryfikacja na korzyść hipotezy alternatywnej następuje przy znacznie łatwiejszych do zaobserwowania wynikach, niż miało by to miejsce w przypadku stosowania hipotezy parametrycznej.

Stosujemy je w sytuacji, gdy nasze dane nie spełniają wymagań ich parametrycznych odpowiedników

Test weryfikujący hipotezę o niezależności cech

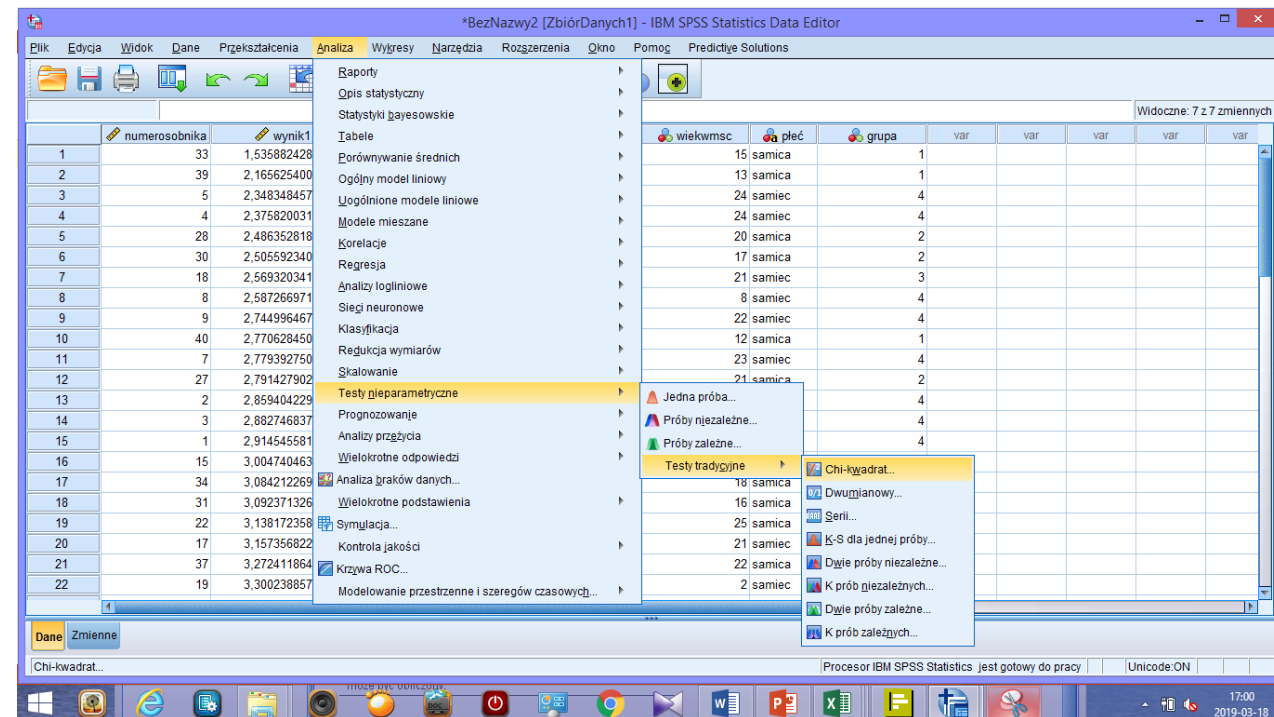
test χ^2 niezależności- dla prób dwuwymiarowych

H_0 : cechy są NIEZALEŻNE od siebie

Bardzo ważne: wszystkie zmienne powinny być skategoryzowane! Nie przeprowadza się analizy, gdy jedna zmienna jest skategoryzowana (np. płeć) a druga określona niepowtarzającymi się wartościami. Ten test bada zależność stochastyczną !

Siłę zależności badanych cech TYLKO po odrzuceniu H_0 najlepiej mierzyć współczynnikiem V Cramera (biorącym pod uwagę liczbę wierszy i kolumn w tabeli krzyżowej).

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - p_{ij}N)^2}{p_{ij}N} \quad v = (k - 1)(l - 1)$$



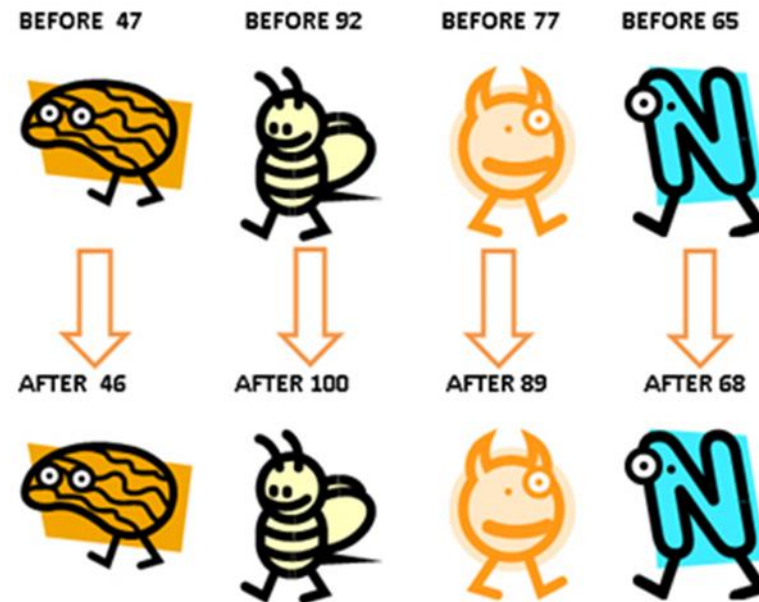
Testy weryfikujące równość/ identyczność danych dla zmiennych EWIDENTNIE zależnych

Zmienne ewidentnie zależne, to takie, które wynikają z następstwa działań i mogą być porównywane parami. Ten sam osobnik znajduje się w jednej i drugiej grupie

Odpowiednikiem tych testów w przypadku hipotez parametrycznych jest test t- Studenta dotyczący równości wartości oczekiwanych.

W przypadku hipotez nieparametrycznych poszczególnym wartościom nadaje się rangi lub znaki, które sumuje się parami.

Porównywane zmienne powinny być zawsze równoliczne.



Testy weryfikujące równość/ identyczność danych dla zmiennych EWIDENTNIE zależnych

RANGI

Rangowanie obserwacji polega na **uszeregowaniu wartości od najmniejszych do największych** oraz przypisaniu im kolejnych rang zaczynając od 1

Oceny	Po uporządkowaniu	Rangi
3,5	2,0	1
4,5	3,0	2
2,0	3,5	3
4,0	4,0	4,5
5,0	4,0	4,5
3,5	4,5	6
3,0	5,0	7

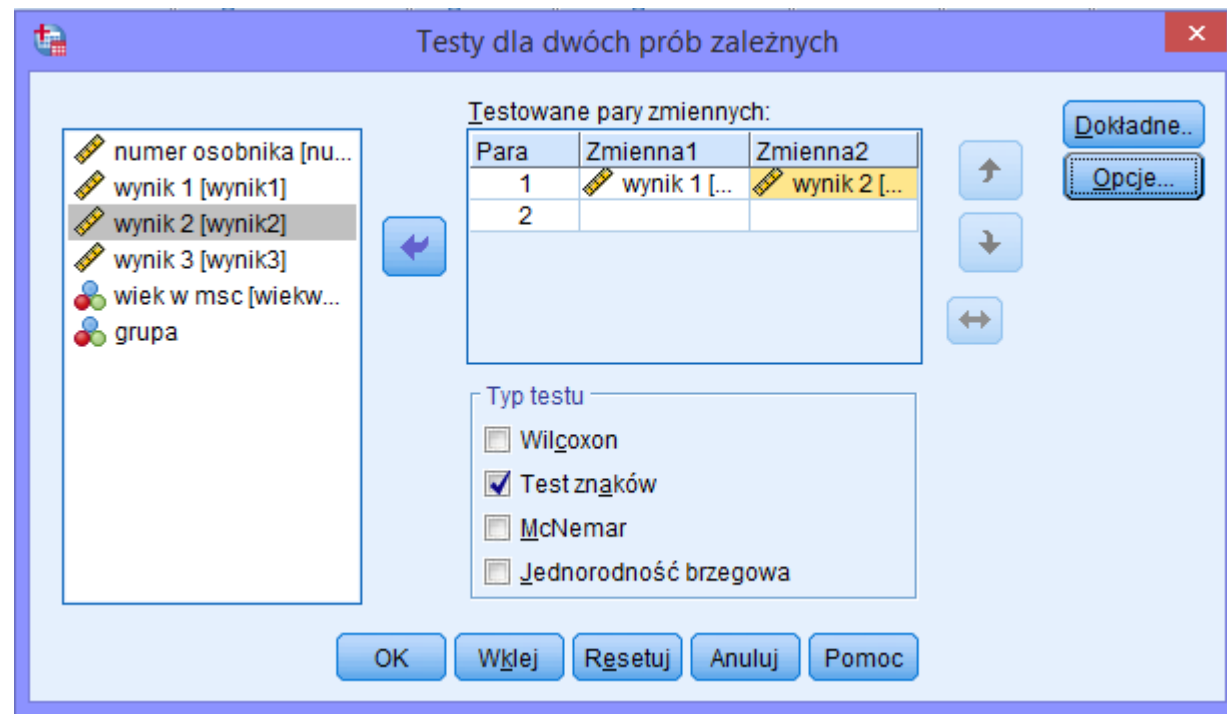
Rangi wiązane:
Dla wartości powtórzonych n razy wyciągana jest średnia arytmetyczna z n sąsiednich rang

Testy weryfikujące równość/ identyczność danych dla zmiennych EWIDENTNIE zależnych

Test znaków

H_0 : wyniki obu zmiennych SĄ takie same

Polega na otrzymaniu znaku +/- (lub 0) różnicy pomiędzy parami wyników porównywanych zmiennych. Liczba plusów i minusów jest zliczana i porównywana z wartością teoretyczną, gdzie liczba znaków + podlega rozkładowi dwumianowemu z parametrem (n - liczba „0”) oraz prawdopodobieństwem równym 0.5.

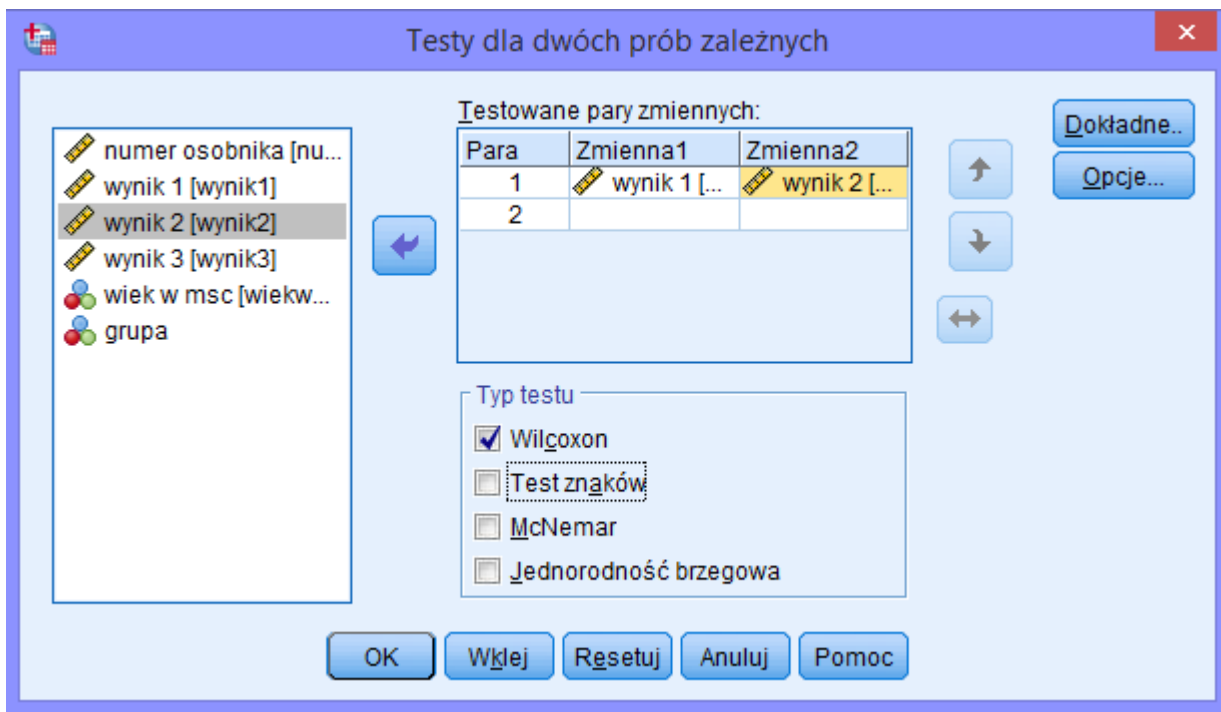


Testy weryfikujące równość/ identyczność danych dla zmiennych EWIDENTNIE zależnych

Test rangowanych znaków (kolejności par) Wilcoxona

Test uwzględnia zarówno znak różnic, ich wielkość, jak i kolejność.

Po uporządkowaniu wartości bezwzględnej różnic w sposób rosnący są im przypisywane rangi, a następnie sumowane osobno rangi różnic dodatnich i ujemnych. Ich suma po porównaniu z tabelą wartości teoretycznych decyduje o przyjęciu lub nie hipotezy zerowej.

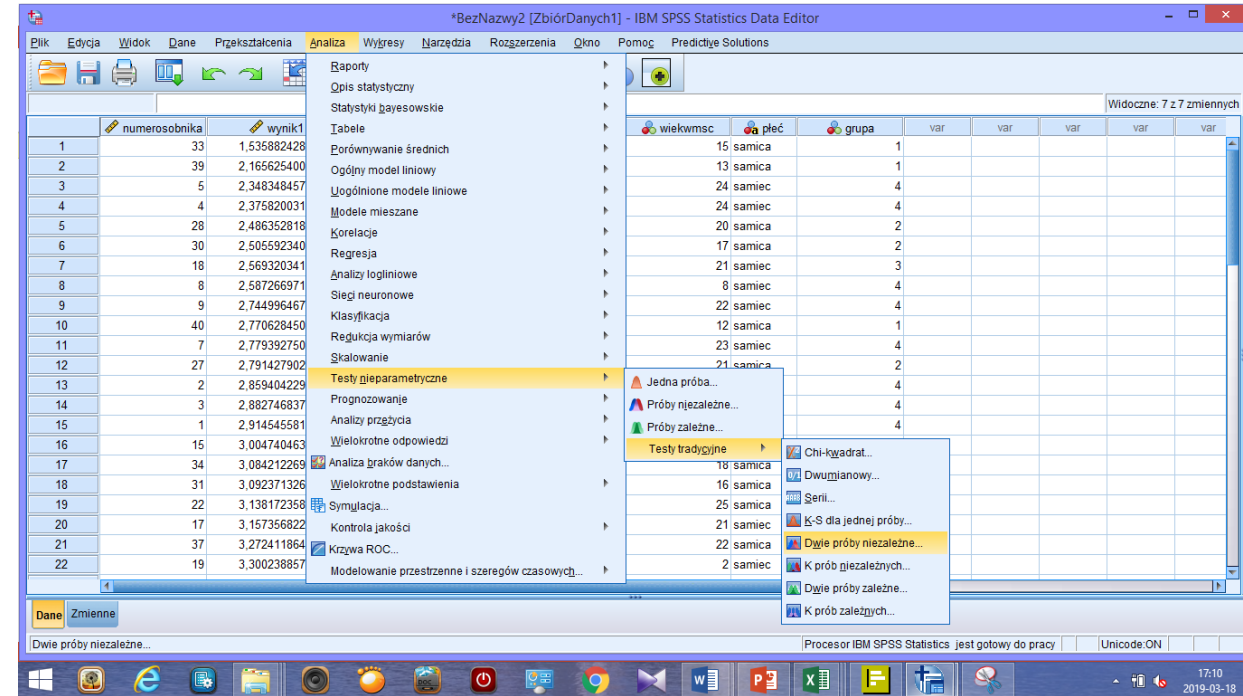


Testy weryfikujące równość/ identyczność danych dla zmiennych niezależnych

Zmienne niezależne, to takie, które otrzymywane są w sposób niezależny, czyli dane zbierane niezależnie w dwóch populacjach - osobnik reprezentujący jedna populację nie powinien znaleźć się w drugiej

Odpowiednikiem tych testów w przypadku hipotez parametrycznych jest test t- Studenta dotyczący równości wartości oczekiwanych.

W przypadku hipotez nieparametrycznych poszczególnym wartościom nadaje się rangi które się sumuje.



Testy weryfikujące równość/ identyczność danych dla zmiennych niezależnych

Test U Manna- Whithneya

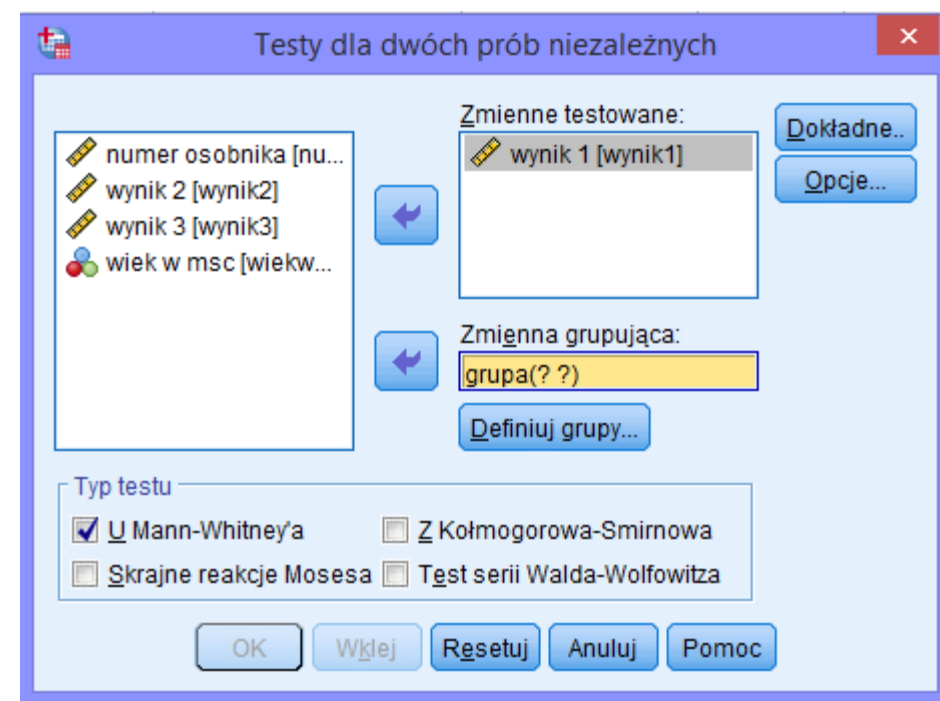
Ho: Próby pochodzą z populacji o tej samej wartości oczekiwanej

Zmienna zależna musi być mierzona na skali co najmniej porządkowej

Nie wymaga równoliczności grup.

Miarą tendencji centralnej dla tego testu jest mediana.

Obszary krytyczne dla ustalonego α odczytuje się z tablic rozkładu normalnego standardowego.



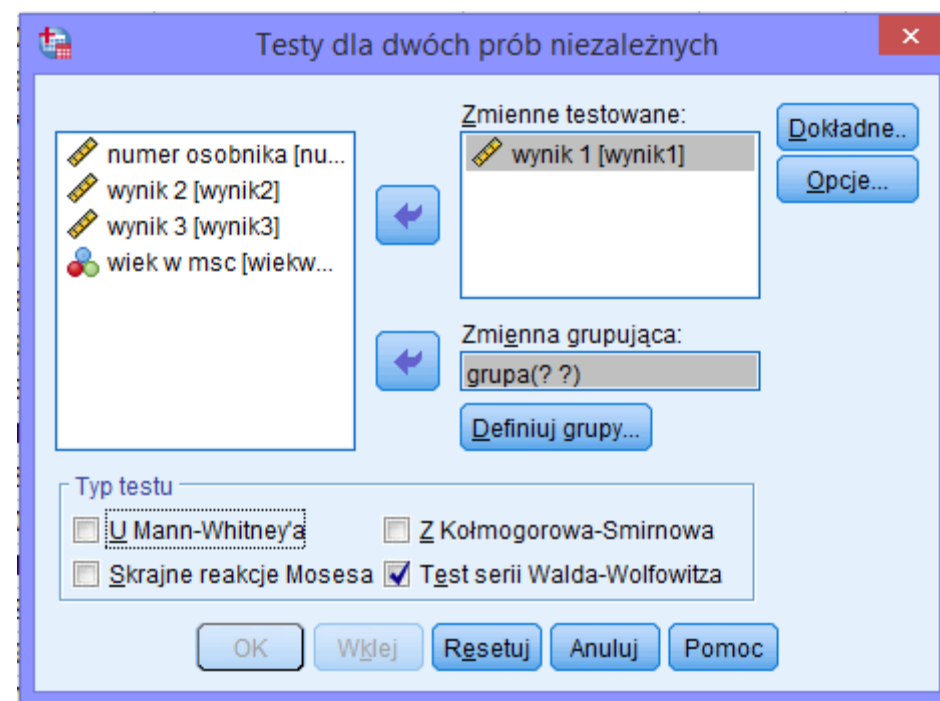
Testy weryfikujące równość/ identyczność danych dla zmiennych niezależnych

Test serii Walda-Wolfowitza

Wartością empiryczną tego testu jest stwierdzona liczba serii. Serię stanowią identyczne znaki (np. 0, 1) otrzymane po posortowaniu wartości z obu porównywanych prób łącznie rosnąco.

Jednakowe wartości występujące w obu próbach są usuwane.

Testem można porównać TYLKO dwie populacje.



Testy weryfikujące równość/ identyczność danych dla zmiennych niezależnych

Test Kruskala-Wallisa

Test jest alternatywą dla jednoczynnikowej analizy wariancji ANOVA,

Zmienna zależna powinna być mierzona na skali co najmniej porządkowej

obserwacje w analizowanych grupach powinny być niezależne wobec siebie.

Test ten wymaga rangowania obserwacji

Istotny wynik testu Kruskala-Wallisa wymaga przeprowadzenia porównań wielokrotnych.

Testy nieparametryczne: Dwie lub więcej prób niezależnych

Cele Zmienne Ustawienia

Wybierz element:

- Wybierz testy
- Opcje testów
- Braki danych

☐ Automatycznie wybierz testy odpowiednie do danych

☒ Pozwól użytkownikowi wybrać testy

Porównanie rozkładów w grupach

- ☒ U Manna-Whitneya (2 próby)
- ☒ Jednoczynnikowa analiza wariancji Kruskala-Wallisa (k prób)
- ☐ Test Kolmogorowa-Smirnowa (2 próby)
- ☐ Test na uporządkowane alternatywy (Jonckheere-Tespstra dla k prób)
- ☐ Test sekwencji na losowość (Walda-Wolfowitza dla 2 prób)

Wielokrotne porównania: Wszystkie parami

Kolejność hipotez: Od najmniejszych

Wielokrotne porównania: Wszystkie parami

Porównanie przedziałów w grupach

- ☒ Test Mosesa skrajnych reakcji (2 próby)
- ☐ Test mediany (k prób)

☒ Oblicz wartości odstające w próbie

☒ Niestandardowa liczba wartości odstających

Wartości odstające: 1

Porównanie median w grupach

- ☒ Mediana z łączonych prób
- ☒ Użytkownika

Mediana: 0

Wielokrotne porównania: Wszystkie parami

Oszacowanie przedziału ufności w grupach

- ☒ Estymator Hodgesa-Lehmana (2 próby)

Uruchom Wklej Resetuj Anuluj Pomoc

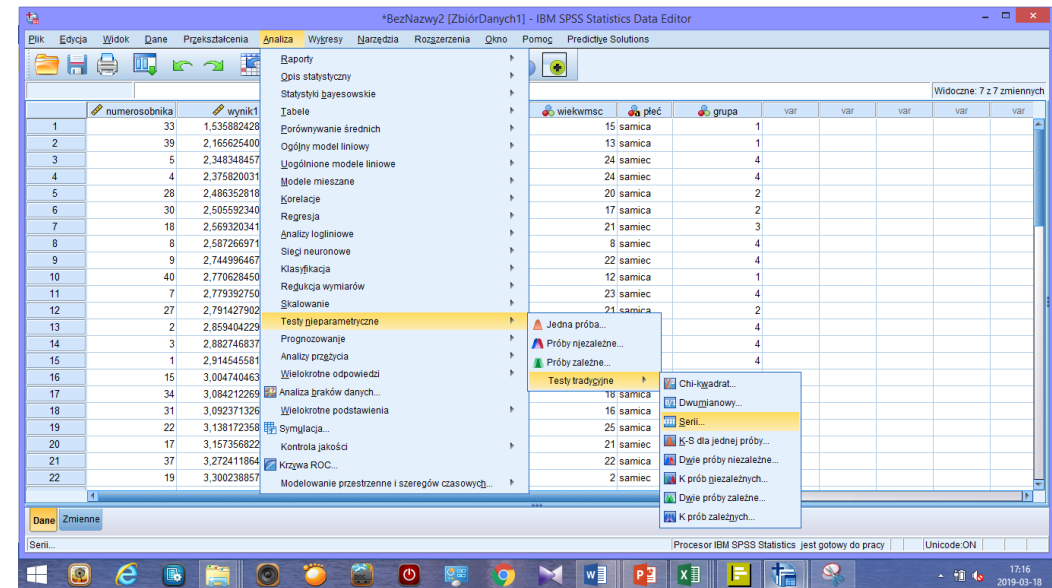
Inne testy nieparametryczne czasami przydatne

Test serii na losowość/ test mediany

H_0 : elementy stanowiące próbę zostały wybrane losowo

Z uporządkowanego wg. kolejności pobierania elementów do próby ciągu obliczamy medianę, me . Każdemu elementowi próby x_i w tym ciągu przypisujemy symbol 'a' jeśli $x_i < me$, bądź symbol 'b' jeżeli $x_i > me$. Wyniki $x_i = me$ odrzucamy.

Otrzymujemy w ten sposób ciąg złożony z symboli 'a' i 'b' np. a b b a a a a b b b b a b a a b. W ciągu tym otrzymujemy pewną liczbę serii – ciągu takich samych elementów (tutaj 8). Statystyką 'k' jest liczba serii. Przy założeniu prawdziwości hipotezy o losowości próby, liczba serii k ma znany i tablicowany rozkład zależny tylko od n_1 i n_2 liczebności elementów 'a' i 'b'



Testy dla jednej proporcji

Sprawdzają czy proporcja obserwacji w populacji jest zgodna z założoną

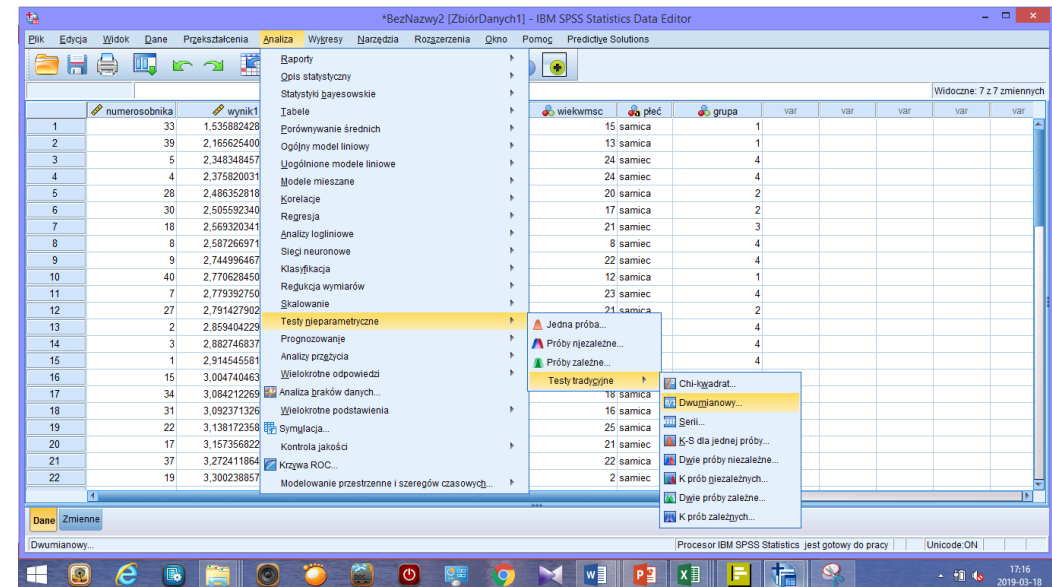
W zależności od wielkości próby mamy do wyboru

test Z dla jednej proporcji - dla dużych prób
oraz

test dwumianowy - dla prób o małej licznosci.

Analiza > Testy nieparametryczne > Testy tradycyjne > Dwumianowy

Analiza > Testy nieparametryczne > Jedna Próba



Inne testy nieparametryczne czasami przydatne

Korelacja rangowa Spearmana

nie jest testem lecz współczynnikiem obliczanym dla stwierdzenia siły zależności pomiędzy zmiennymi

Współczynnik ten jest wykorzystywany do opisu siły i kierunku zależności dwóch cech, wtedy gdy są one mierzalne, badana zbiorowość jest nieliczna oraz mają charakter jakościowy i istnieje możliwość ich uporządkowania.

Miarę tę stosuje się również do badania zależności między cechami ilościowymi w przypadku niewielkiej liczby obserwacji.

Współczynnik jest wrażliwy na obserwacje odstające.

Metoda obliczeniowa różnicuje rangi, dlatego pozwala na otrzymanie współczynnika ze znakiem + lub – dookreślając kierunek zależności.