

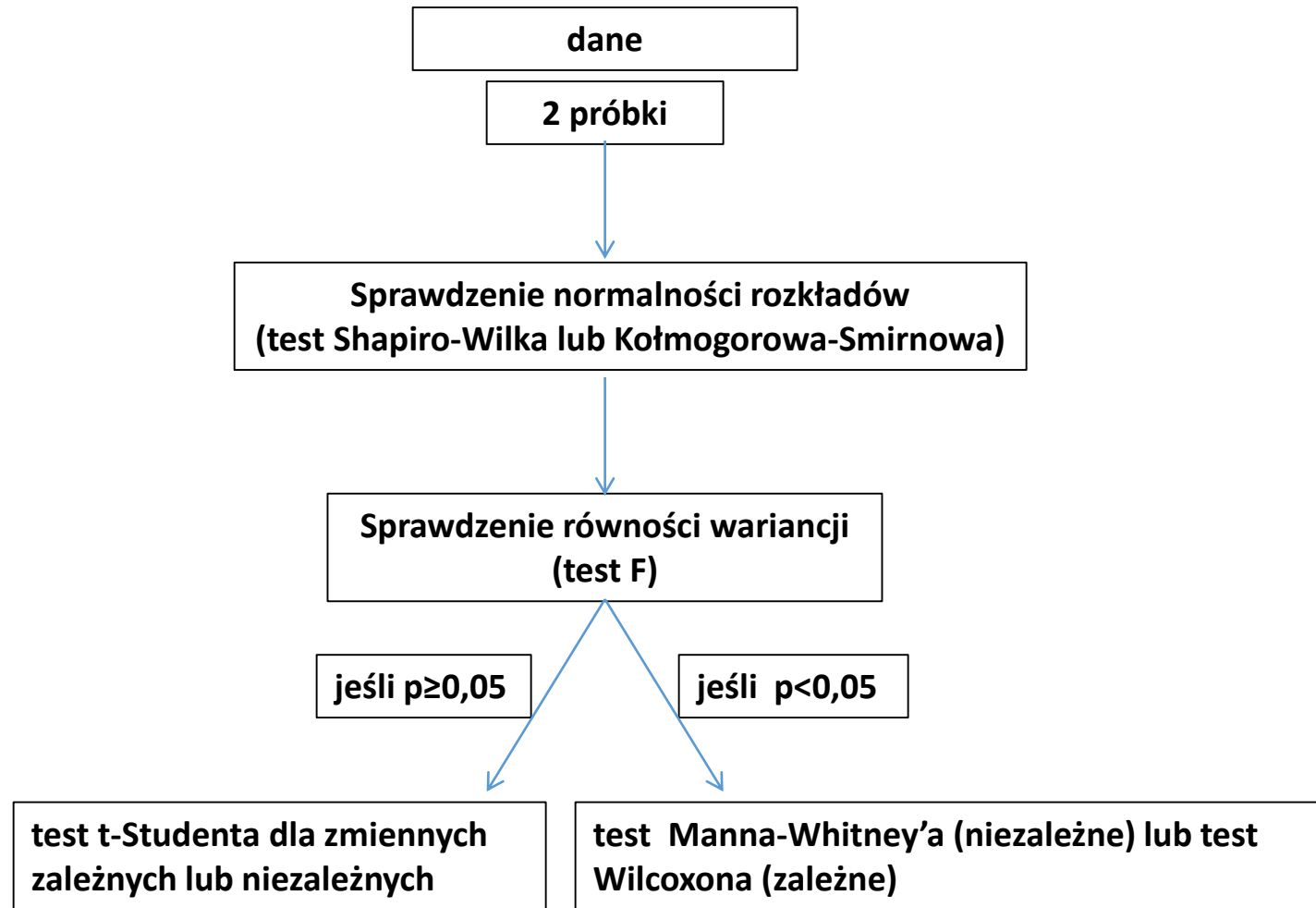
Statystyka i modelowanie w naukach o środowisku

Wykład 5

Testowanie hipotez

Test istotności dla dwóch populacji

Test istotności dla dwóch średnich



Test normalności

test Shapiro-Wilka - standardowy test wykorzystywany do testowania normalności danych

przykład:

sprawdzamy czy próba $x_1, \dots, x_n$ pochodzi z rozkładu normalnego

Hipoteza zerowa i alternatywna w teście Shapiro-Wilka ma następującą postać:

H_0 : Próba pochodzi z populacji o rozkładzie normalnym

H_1 : Próba nie pochodzi z populacji o rozkładzie normalnym

W celu przeprowadzenia testu wykorzystuje się statystykę W .

Test normalności - przykład

Badano zawartość witaminy C w soku pomidorowym przygotowanym według wybranej receptury. Pomiarzy zawartości witaminy C w 17 próbkach dały wyniki (mg/100g): 18,8 18,2 16,5 20,5 20,0 20,2 14,6 26,9 25,0 22,3 20,9 22,1 23,1 13,6 21,4 14,8 17,9.

Sprawdź, czy dane w próbie pochodzą z rozkładu normalnego.

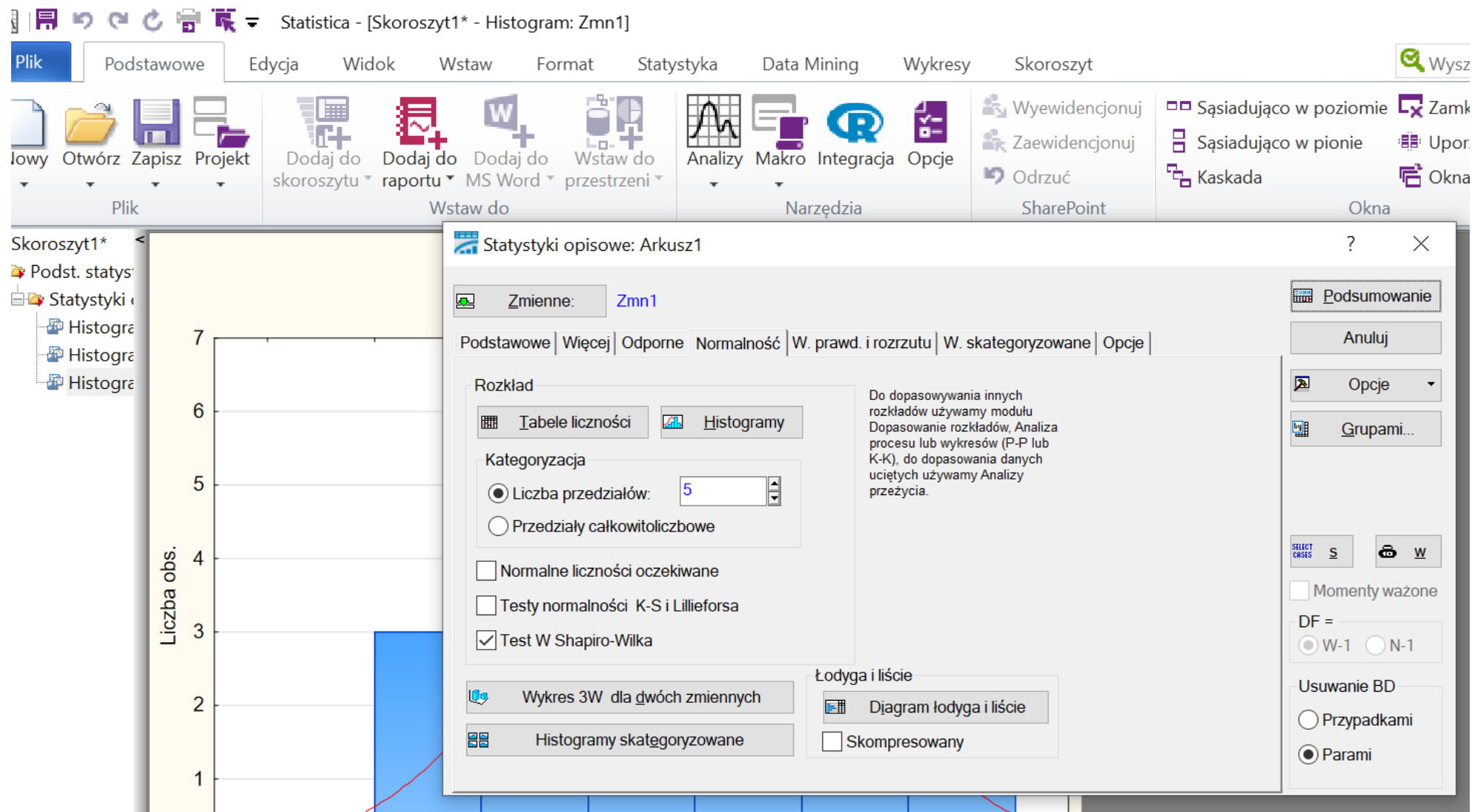
Test normalności - przykład

Statistica:

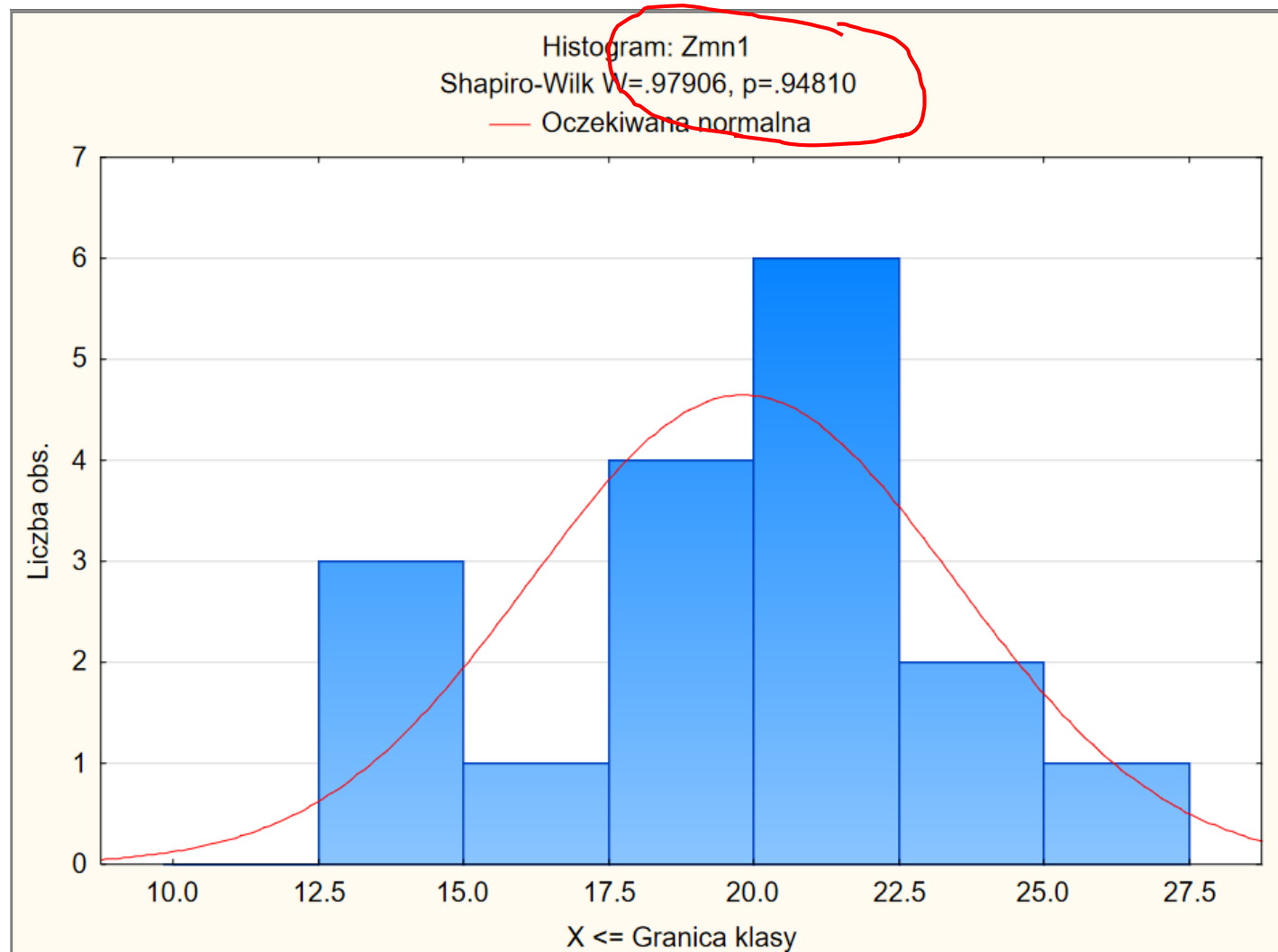
dane w jednej kolumnie

Statystyka → statystyki podstawowe → statystyki opisowe

Test normalności - przykład



Test normalności - przykład



Test normalności - przykład

Statistica:

wartość statystyki testowej (statystyka W)

$W=0,979$

wartość p (p-value)

$p=0,948$

na poziomie istotności $\alpha=0,05$ ($p>\alpha$)

nie ma podstaw do odrzucenia hipotezy zerowej.

odpowiedź:

Dane pochodzą z rozkładu normalnego.

Test normalności - przykład

przykładowy kalkulator on-line dla testu Shapiro-Wilka:

<http://www.statskingdom.com/320ShapiroWilk.html>

Test istotności dla dwóch średnich służący do wnioskowania o równości dwóch średnich w dwóch populacjach normalnych w zmiennych niezależnych

Badamy dwie populacje generalne, mające rozkłady normalne $N(\mu_1, \sigma_1)$ i $N(\mu_2, \sigma_2)$.

Dodatkowo, grupy są niezależne, czyli takie, których jednostki się między sobą nie mieszają.

Warunek normalności rozkładów jest spełniony - korzystamy z testu t na równość dwóch średnich.

Test istotności dla dwóch średnich służący do wnioskowania o równości dwóch średnich w dwóch populacjach normalnych w zmiennych niezależnych

Sprawdzamy, czy $\mu_1 = \mu_2$:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Test istotności dla dwóch średnich

$$t_{emp} = \frac{\overline{x_1} - \overline{x_2}}{s_r}$$

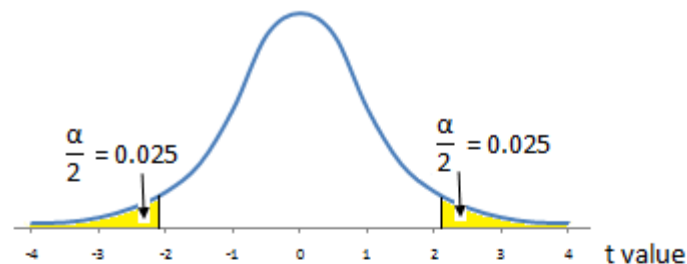
$$s_r = \sqrt{s_e^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$s_e^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$|t_{emp}| > t_{\alpha, n_1 + n_2 - 2}$$

Student's t Distribution Table

For example, the t value for
18 degrees of freedom
is 2.101 for 95% confidence
interval (**2-Tail** $\alpha = 0.05$).



	90%	95%	97.5%	99%	99.5%	99.95%	1-Tail Confidence Level
	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
<i>df</i>	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869	
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370	
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178	
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208	
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405	
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728	
16	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150	
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651	
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216	
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834	

Test na równość dwóch średnich dla dwóch grup zależnych

Grupy zależne to takie, których jednostki się mieszają.

Najczęściej są to dokładnie te same jednostki np. pacjenci, których czas reakcji na jakiś bodziec obserwuje się przed podaniem im leku i po jego podaniu.

Analizowane grupy są więc zależne/powiązane.

Pytanie: czy podanie leku zmienia czas reakcji pacjentów lub czy czas reakcji przed podaniem leku jest średnio równy czasowi reakcji po jego podaniu?

Test na równość dwóch średnich dla dwóch grup zależnych

w teście bada się istotność różnicy średnich dla zmiennych X_1 i X_2

Sprawdzamy, czy $\mu_1 = \mu_2$:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

Test t - przykład

Wykonano badania masy owoców na dwóch odmianach ogórków gruntowych: Przybyszewski (A) i Warszawski Gruntowy (B). Dane pochodzą z rozkładów normalnych. Uzyskano następujące pomiary:

A	150	163	155	164	162	170	174	168	172	171	168
B	152	159	161	164	169	177	173	171	176	176	174

Czy odmiana B istotnie różni się od odmiany A pod względem średniej masy owoców? Zweryfikuj odpowiednią hipotezę na poziomie istotności $\alpha=0,05$.

Test t – przykład (próby niezależne, spełniony warunek o normalności rozkładów)

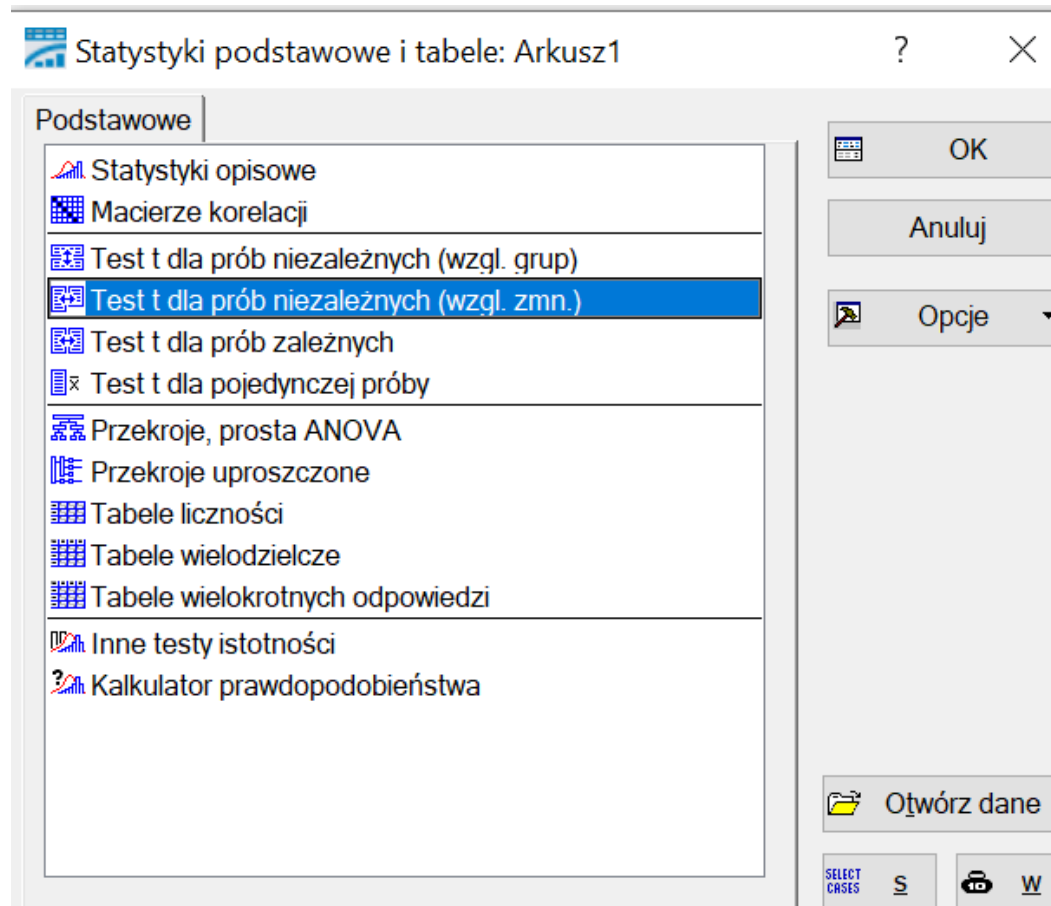
$$H_0: \mu_1 = \mu_2$$

Badane odmiany ogórków nie różnią się średnią masą.

$$H_1: \mu_1 \neq \mu_2$$

Badane odmiany ogórków różnią się średnią masą.

Test t - Statistica



Przybysze	Warszawski	5 Zmn5
150	152	
163	159	
155	161	
164	164	
162	169	
170	177	
174	173	
168	171	
172	176	
171	176	
168	174	

Test t dla prób niezależnych, względem zmiennych: ... ? X

Zmienne (grupy):

Pierwsza lista: brak

Druga lista: brak

Podstawowe | Opcje

Podsumowanie: testy t

Wykres ramka-wasy

Podsumowanie

Anuluj

Opcje

Grupami...

SELECT CASES S W

Wybierz dwie listy zmiennych (listy grup) ? X

1 - Zmn1

2 - Zmn2

3 - Przybyszewski

4 - Warszawski Gruntowy

5 - Zmn5

6 - Zmn6

7 - Zmn7

8 - Zmn8

9 - Zmn9

10 - Zmn10

Wszystkie Rozwiń Przybliż

Pierwsza lista zmiennych (grup):

3

1 - Zmn1

2 - Zmn2

3 - Przybyszewski

4 - Warszawski Gruntowy

5 - Zmn5

6 - Zmn6

7 - Zmn7

8 - Zmn8

9 - Zmn9

10 - Zmn10

Wszystkie Rozwiń Przybliż

Druga lista zmiennych (grup):

4

OK

Anuluj

[Zestawy]...

Włącz opcję "Pokazuj tylko zmienne o odpowiedniej skali" aby na listach, w zależności od potrzeby, pojawiały się tylko zmienne jakościowe albo ilościowe. Naciśnij F1 aby uzyskać więcej informacji.

☐ Pokazuj tylko zmienne o odpowiedniej skali

Testy dla prób niezależnych (Arkusz1)

Uwaga: Zmienne traktowane są jako niezależne próby.

Grupa 1 wz. Grupy 2

Średnia Grupa 1	Średnia Grupa 2	t	df	p	Nważnych Grupa 1	Nważnych Grupa 2	Odch.std Grupa 1	Odch.std Grupa 2	iloraz F Wariancje	p Wariancje
165.1818	168.3636	-0.952227	20	0.352349	11	11	7.400246	8.249518	1.242696	0.737802

Test t - przykład

wartość statystyki testowej

$$t = -0,95$$

wartość p

$$p = 0,35$$

$$p > \alpha$$

nie ma podstaw do odrzucenia hipotezy zerowej

Badane odmiany ogórków nie różnią się średnią masą.

Test t – przykład Excel

Excel – należy włączyć dodatek Analysis ToolPak

Plik → Opcje → Dodatki → Analysis ToolPak → Przejdź

i zaznaczyć Analysis ToolPak

Analysis ToolPak będzie dostępny

Dane → Analiza danych

Test t – przykład Excel

	C	D	E	F	G	H	I
	Zad. 2						
	masa ogórków						
	Przybyszewski	Warszawski Gruntowy					
	150	152					
	163	159					
	155	161					
	164	164					
	162	169					
	170	177					
	174	173					
	168	171					
	172	176					
	171	176					
	168	174					

Test t: z dwiema próbami zakładający równe wariancje

Wejście

Zakres zmiennej 1:

Zakres zmiennej 2:

Różnica średnich wg hipotezy:

☐ Tytuły

Alfa:

Opcje wyjścia

☐ Zakres wyjściowy:

☒ Nowy arkusz:

☐ Nowy skoroszyt

OK

Anuluj

Pomoc

Test t – przykład kalkulatory on-line

<https://www.socscistatistics.com/tests/studentttest/default.aspx>

Proszę wybrać test dwustronny, czyli interesuje nas, czy średnie się różnią, a nie czy Przybyszewski ma średnią masę większą niż Warszawski Gruntowy.

Test t dla prób zależnych

<https://www.socscistatistics.com/tests/ttestdependent/default2.aspx>

Testy nieparametryczne

Jeżeli warunek o normalności rozkładów nie jest spełniony, w szczególności dla prób, których liczność nie przekracza 30, nie jest uzasadnione korzystanie z testu t na równość dwóch średnich.

W przypadku braku spełnienia założenia o normalności równość średnich można weryfikować testem nieparametrycznym.

Testy nieparametryczne

Test Manna-Whitney'a to przykład testu nieparametrycznego dość powszechnie stosowanego w kontekście nieparametrycznego porównywania średnich dla dwóch grup niezależnych.

Testy nieparametryczne

Test par rangowanych znaków Wilcoxona - odpowiednik testu t dla dwóch prób zależnych.

Test U Manna-Whitneya

Służy do sprawdzenia hipotezy, że dwie niezależne próby pochodzą z identycznych populacji.

Zastępuje on test t-Studenta dla dwóch średnich, gdy nie są spełnione wymagane założenia.

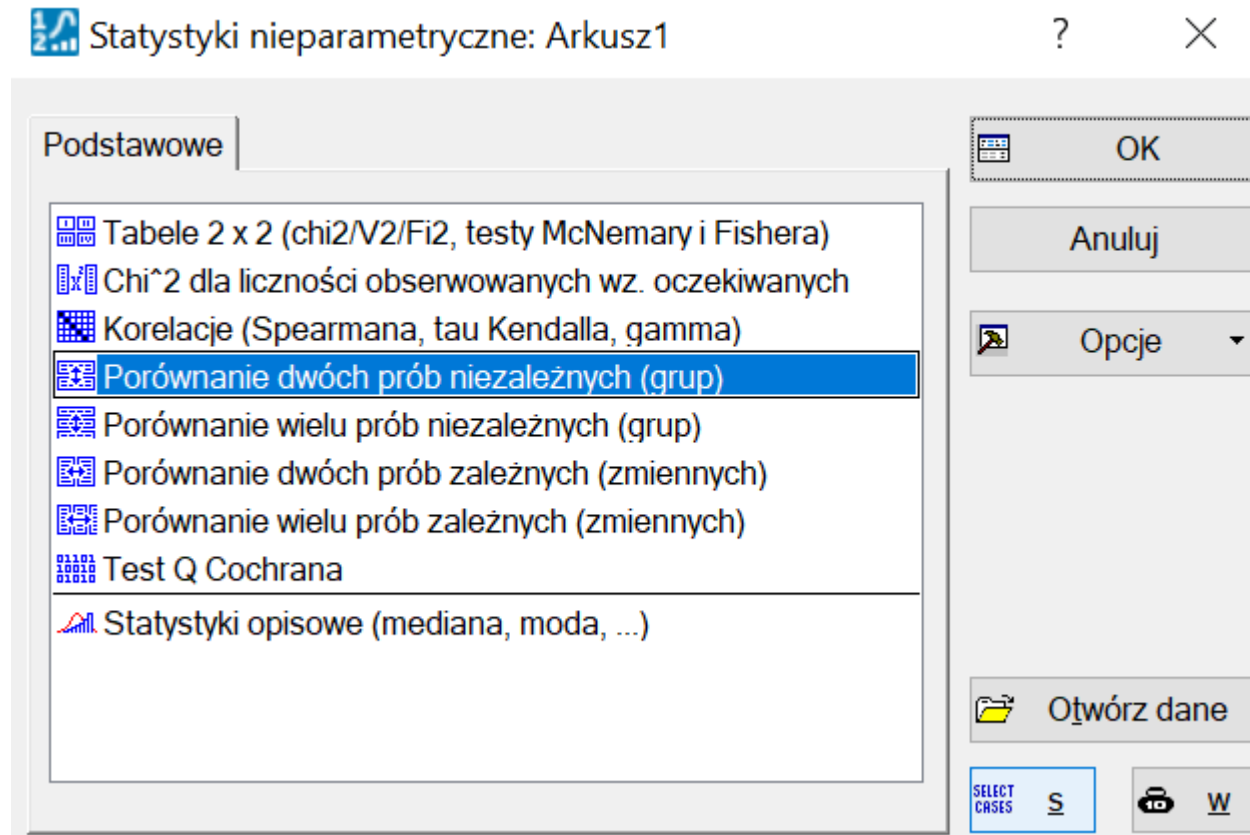
Test U Manna-Whitneya przykład

- Przeprowadzono ocenę porażenia przez choroby wirusowe dwóch odmian ziemniaka (A i B). Ocena była prowadzona w skali 5-stopniowej (1- brak chorób.... 5- bardzo silne porażenie przez choroby wirusowe). Na podstawie losowo wybranych kilkunastu roślin dla każdej z odmian uzyskano następujące wyniki:

Odmiana	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	
Stopień porażenia	3	1	1	2	3	3	4	4	5	1	1	1	1	2	2	1	2	3	3	4	4	5	3	2	1	4	1	5

Test U Manna-Whitneya przykład

Statystyka → Statystyki nieparametryczne



Test U Manna-Whitneya przykład

Statystyka						
Podstawowe						
Zaawansowane i wielowymiarowe						
Test U Manna-Whitneya (z poprawką na ciągłość) (Arkusz1)						
Względem zmiennej: Odmiana						
Zaznaczone wyniki są istotne z $p < .05000$						
Zmienna	Sum.rang A	Sum.rang B	U	Z	p	
Stopień porażenia	179.5000	226.5000	74.50000	-1.05680	0.290605	

Test U Manna-Whitney'a kalkulator on-line

<https://www.socscistatistics.com/tests/mannwhitney/>

Test par rangowanych znaków Wilcoxon

<https://www.socscistatistics.com/tests/signedranks/default.aspx>