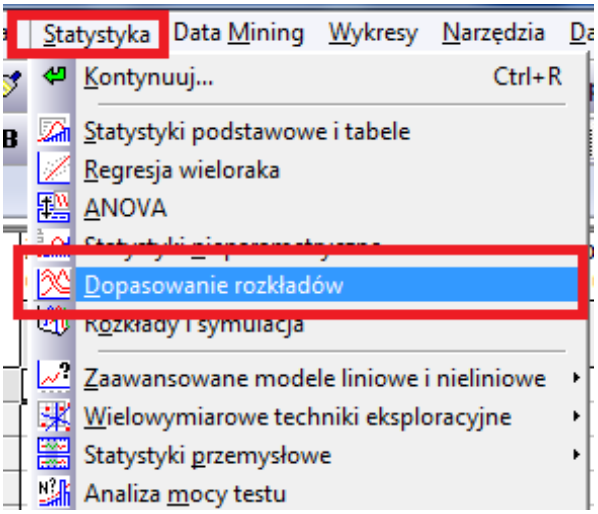
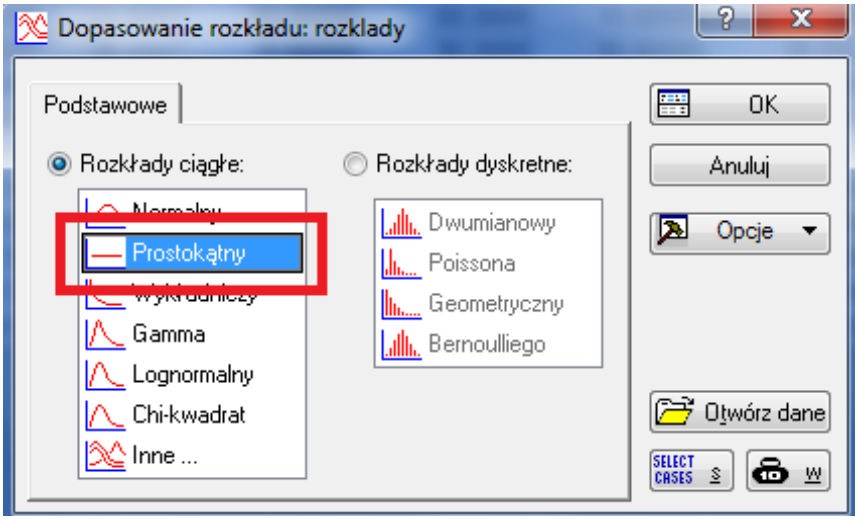
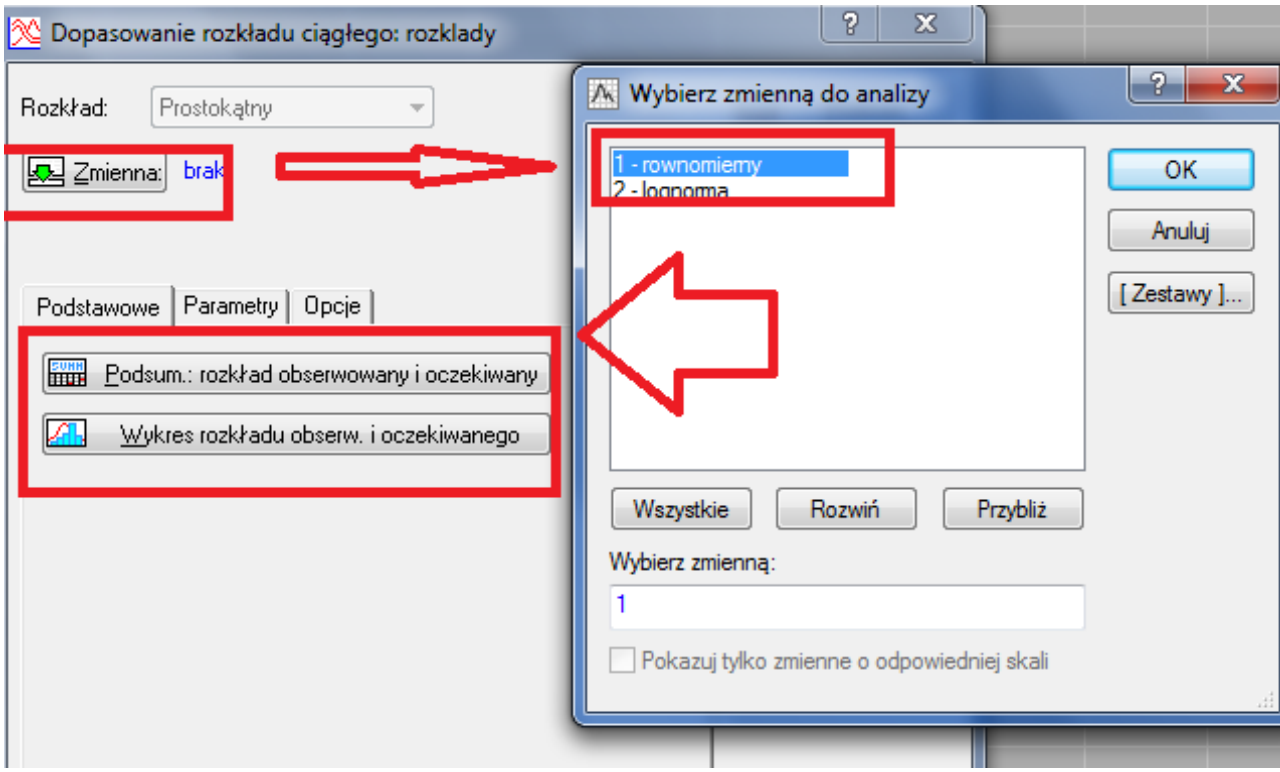
1 Testowanie zgodności rozkładu

Poznaliśmy już sposoby testowania hipotezy o tym, czy próba pochodzi z populacji o rozkładzie normalnym. Dzisiaj zajmiemy się ogólniejszymi metodami weryfikacji zgodności z dowolnym rozkładem. Dla przykładu w pliku rozklady.sav mamy dwie próby pobrane z dwóch różnych rozkładów. Czy można stwierdzić, że pierwsza próba pobrana jest z rozkładu normalnego, a druga z lognormalnego? Dla odmiany zadanie to rozwiążemy najpierw w programie Statistica. Z meny Statystyka wybieramy "Dopasowanie rozkładów"



następnie zaś wybieramy interesujący nas rozkład oraz wybieramy stosowną zmienną





Jako wynik otrzymujemy w zależności od tego co wybraliśmy tabelę z oczekiwanym rozkładem lub odpowiednio wykres(histogram) z nałożoną oczekiwaną

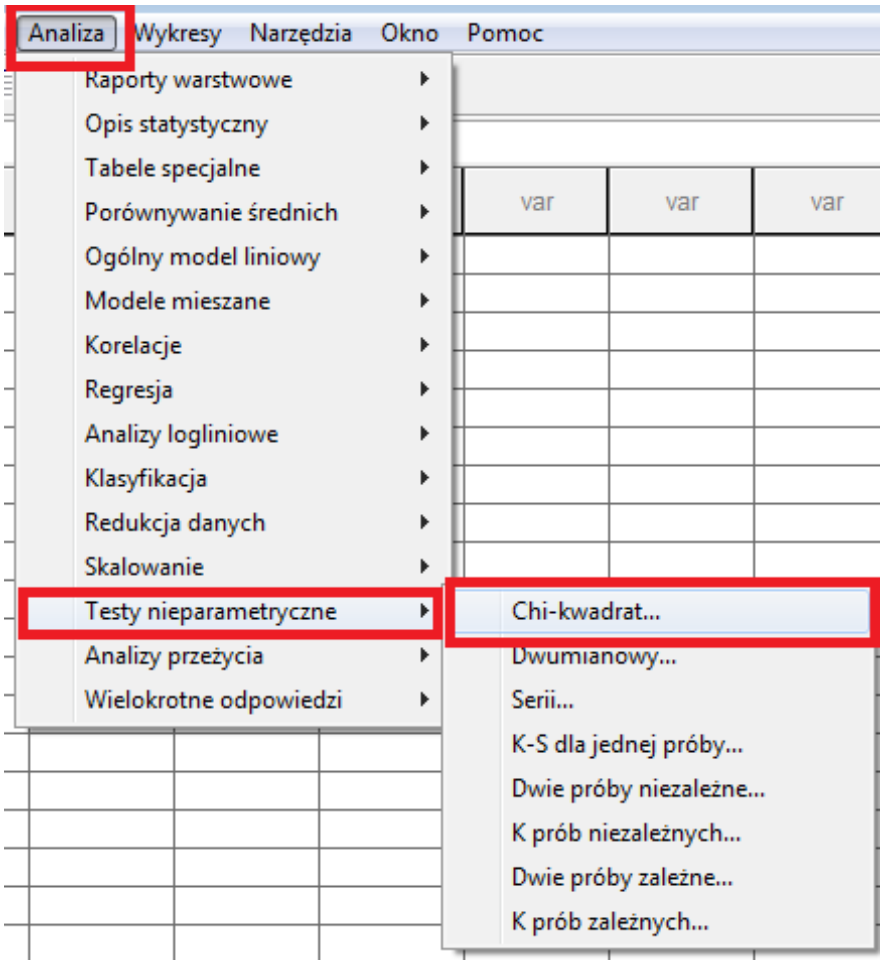
gęstością. Najistotniejsze w obu przypadkach to istotność, która wynosi $p = 0,00042$, zatem odrzucamy hipotezę o równomierności naszego rozkładu. W analogiczny sposób weryfikujemy następujące hipotezy dla drugiej zmiennej

$$H_0: \text{ Y ma rozkład lognormalny}$$

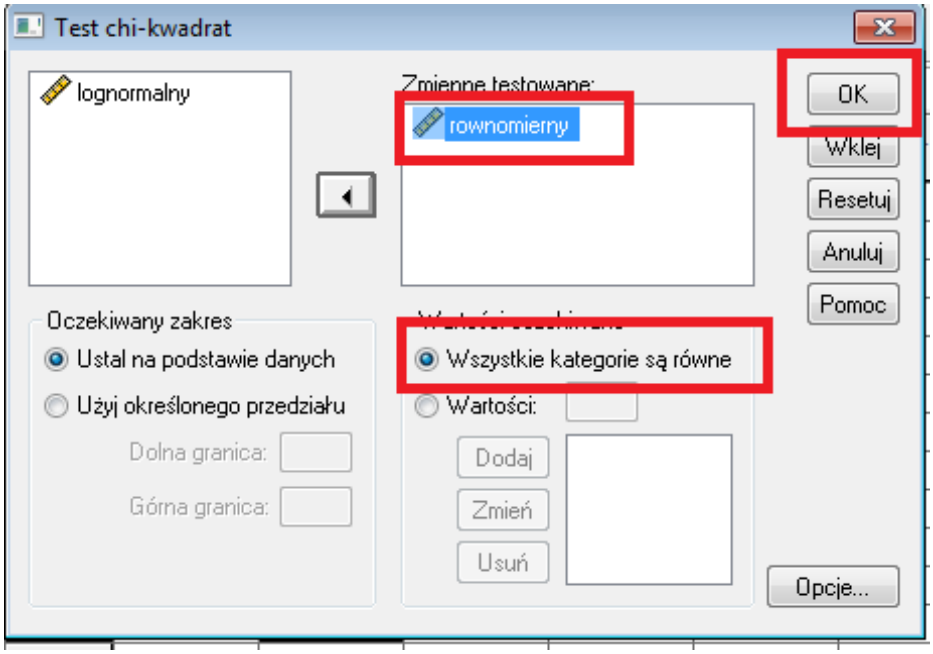
$$H_1: \text{Y nie ma rozkładu lognormalnego}$$

W tym wypadku nie ma podstaw do odrzucenia H_0 ponieważ p=0,4331.

W programie SPSS test χ^2 zgodności znajduje się w sekcji test nieparametryczne



W przypadku rozkładu równomiernego nie ma właściwie żadnego problemu



W raporcie otrzymujemy interesujące nas wyniki

Statystyki testu

	rownomierny
Chi-kwadrat[a]	29,340
df	10
Istotność asymptotyczna	,001

a. 0 komórek (,0%) ma liczebność oczekiwaną mniejszą od 5. Minimalna liczebność oczekiwana w komórce wynosi 27,3.

Jak widzimy istnieją pewne rozbieżności pomiędzy poznawanymi programami. W jednym otrzymaliśmy istotność p=0.00042, zaś w drugim $\alpha = 0.001$. Różnice są powodowane różną dokładnością obliczeń oraz błędami zaokrągleń, nie zmieniają one jednak odpowiedzi podczas weryfikacji hipotez.

2 Weryfikacja hipotez dla jednej średniej

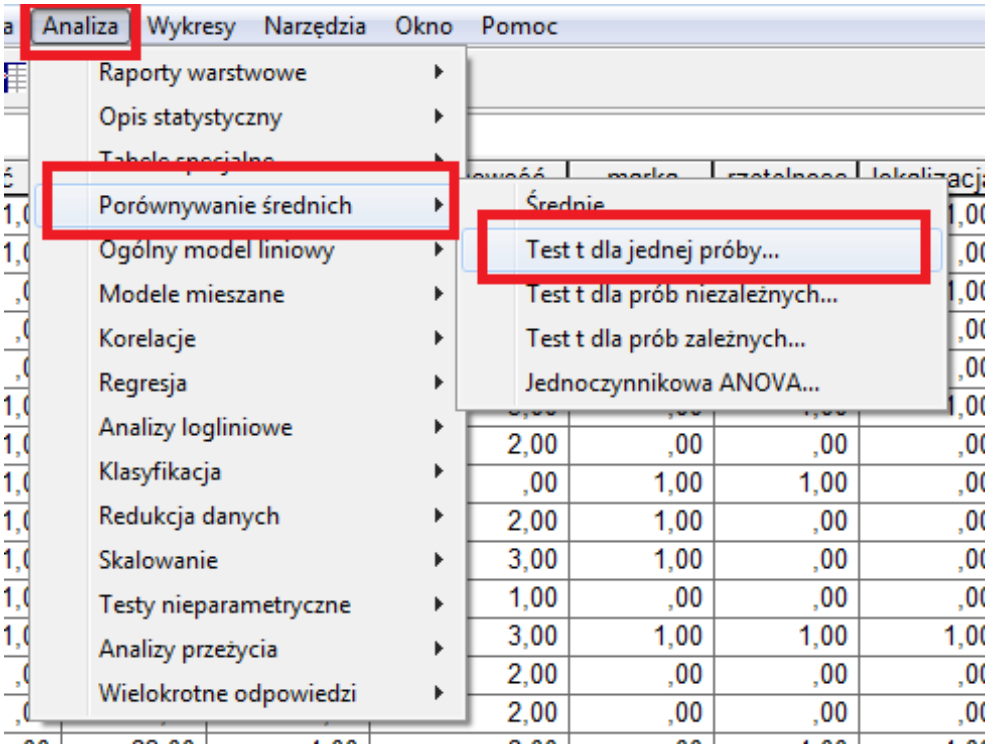
Bardzo często podczas różnych badań w naturalny sposób pojawia się pytanie, czy w analizowanych przez nas danych dla pewnej cechy średnia jest równa pewnej zadanej wartości. Dla przykładu chcemy sprawdzić, czy w rozważanej

przez nas wcześniej ankiecie średnia wieku dla badanej populacji wynosi powiedzmy 35 lat. Oczywiście większość osób stwierdzi, że wystarczy policzyć średnią i będziemy znali odpowiedź. Należy jednak pamiętać, że do wypełnienia ankiety została wybrana pewna próba. Jest oczywiste, że parametry z próby estymują parametry z populacji, to jednak wciąż są to pewne przybliżenia. Fachowej odpowiedzi na takie pytanie dokonujemy posiłkując się stosownym testem do weryfikacji hipotez

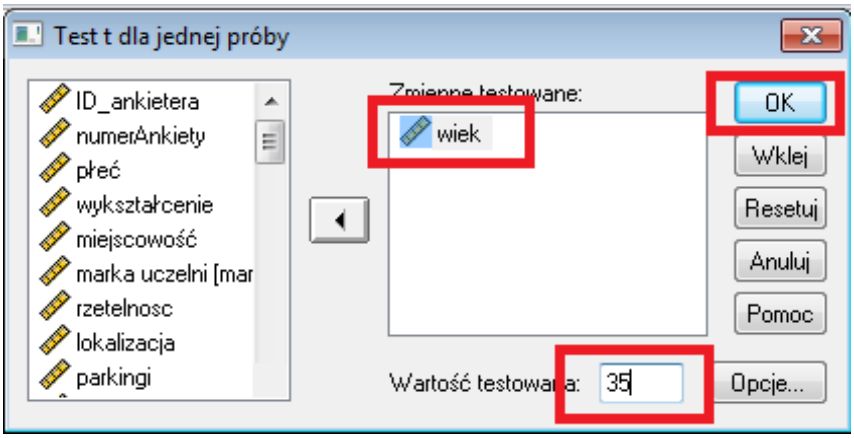
$$H_0 : \mu = \mu_0 \, (35)$$

$$H_1 : \mu = \mu_0 \, (35) \, ,$$

gdzie μ_0 jest zadaną wartością. Oczywiście takiej weryfikacji można dokonać w sposób tradycyjny, jak również z zastosowaniem pakietów statystycznych. My skupimy się raczej na tym drugim sposobie. W programie SPSS odpowiedni test znajdujemy w sekcji porównanie średnich



Następnie wskazujemy interesującą nas zmienną, zadajemy wartość testowaną



Na podstawie otrzymanego raportu

Test dla jednej próby

	Wartość testowana = 35					
					95% przedział ufności dla różnicy średnich	
	t	df	Istotność (dwustronna)	Różnica średnich	Dolna granica	Górna granica
wiek	-4,282	199	,000	-2,25500	-3,2936	-1,2164

łatwo stwierdzamy, że odrzucamy hipotezę H_0 na korzyść H_1.

Sprawdźmy teraz jaki jest poziom istotności jeśli jako wartość testowaną obierzemy 33 lata. Jak widać z poniższego raportu

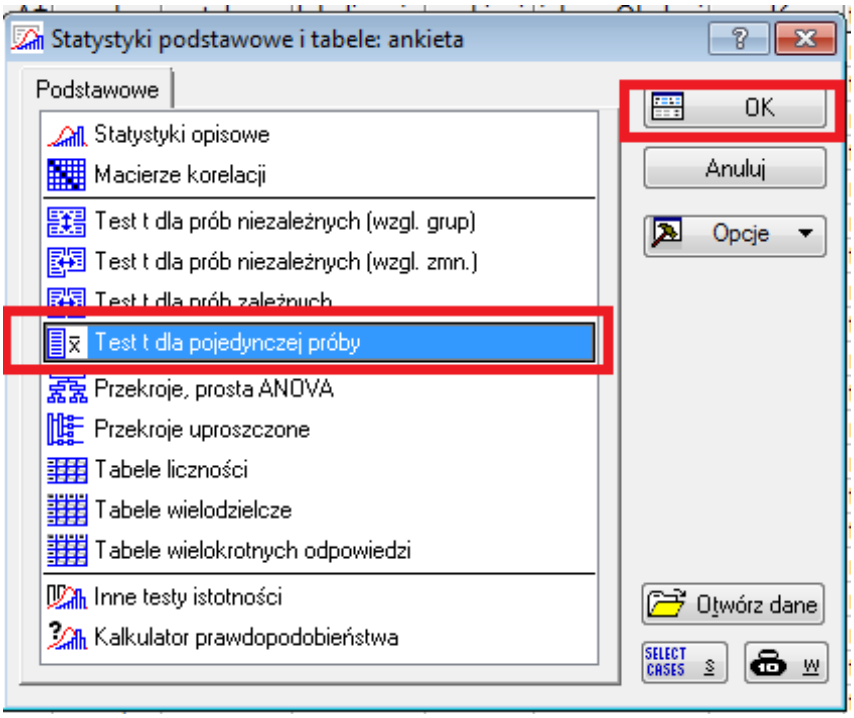
Test dla jednej próby

	Wartość testowana = 33					
					95% przedział ufności dla różnicy średnich	
	t	df	Istotność (dwustronna)	Różnica średnich	Dolna granica	Górna granica
wiek	-,484	199	,629	-,25500	-1,2936	,7836

w takim przypadku nie podstaw do odrzucenia hipotezy zerowej świadczącej o tym, że średnia w populacji wynosi 33 lata.

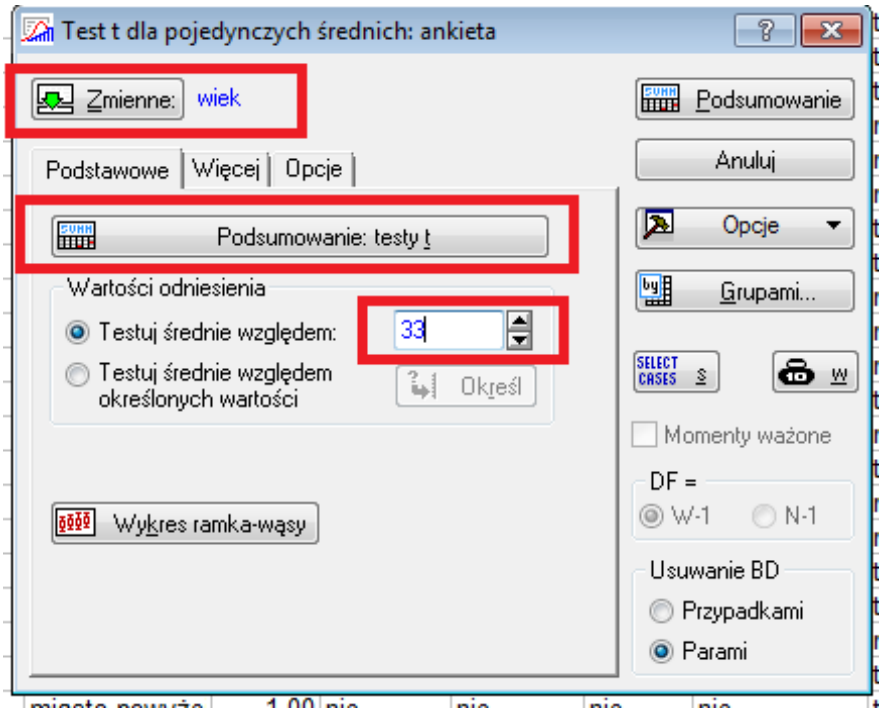
Teraz prześledźmy wykonanie tego samego zadania w programie Statistica. Z menu "Statystyka" wybieramy "Statystyki podstawowe i tabele" a następnie

z okna dialogowego



jak na powyższym rysunku. Po wybraniu odpowiednie zmiennej i testowanej

wartości



otrzymujemy skoroszyt z wynikami testu.

	Test średnich względem stałej wartości odniesienia (ankieta)							
Zmienna	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t	df	p
wiek	32,74500	7,448313	200	0,526675	33,00000	-0,484169	199	0,628798

Tak samo jak wcześniej zauważamy różnice pomiędzy oboma programami wynikające z zaokrągleń.

Dość istotnie z weryfikacją hipotez dla jednej średniej jest wyznaczanie przedziału ufności dla średniej. W statystyce matematycznej przedział ufności dla średniej wyraża się wzorem

$$\mu \in \left[\overline{X} - \frac{S_0}{\sqrt{n}} t_{1-\alpha/2,n-1}; \overline{X} - \frac{S_0}{\sqrt{n}} t_{1-\alpha/2,n-1}\right]$$

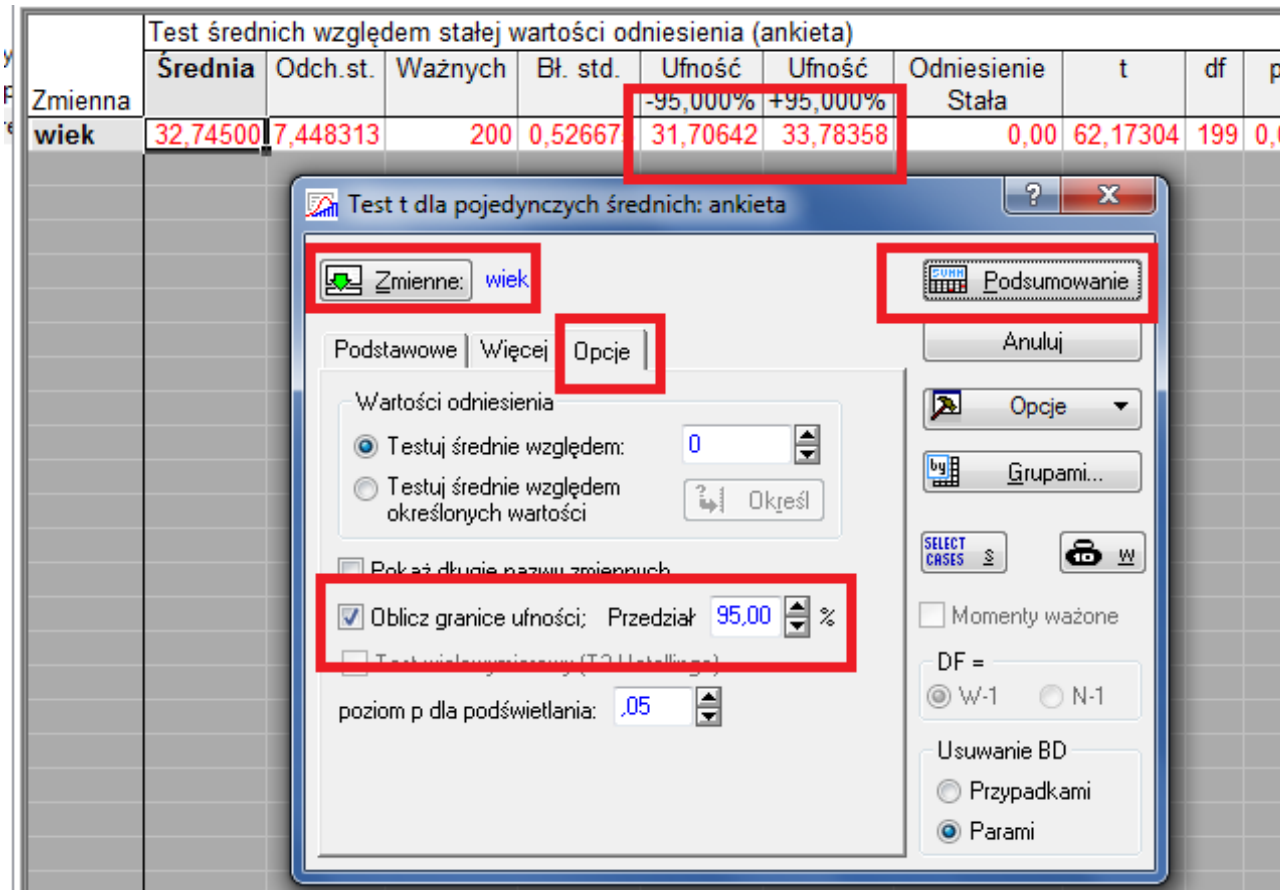
gdzie $\overline{X}$ oznacza średnią z próby;

S_0 - nieobciążone odchylenie standardowe;

n - liczebność próby;

$t_{1-\alpha/2,n-1}$ oznacza kwantyl rzędu $1-\alpha/2$ z rozkładu t-studenta o $n-1$ stopniach swobody.

W programie Statistica musimy wejść do okna dialogowego "test t dla pojedynczych średnich" a następnie przejść na zakładkę Opcje

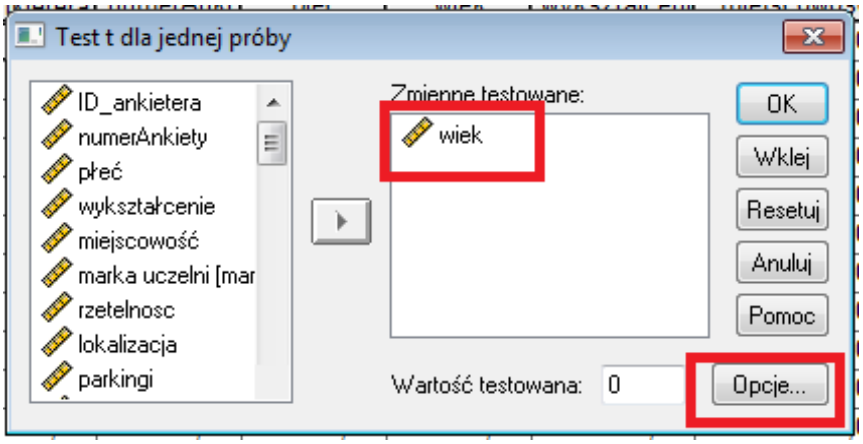


Jako wynik otrzymujemy przedział

$$\mu \in [31.70642; 33.78358]$$

W programie SPSS postępujemy dość podobnie. Wybieramy test dla jednej

próby



wskazujemy zmienną i wchodzimy w opcje, gdzie podajemy poziom ufności i otrzymujemy raport

Test dla jednej próby

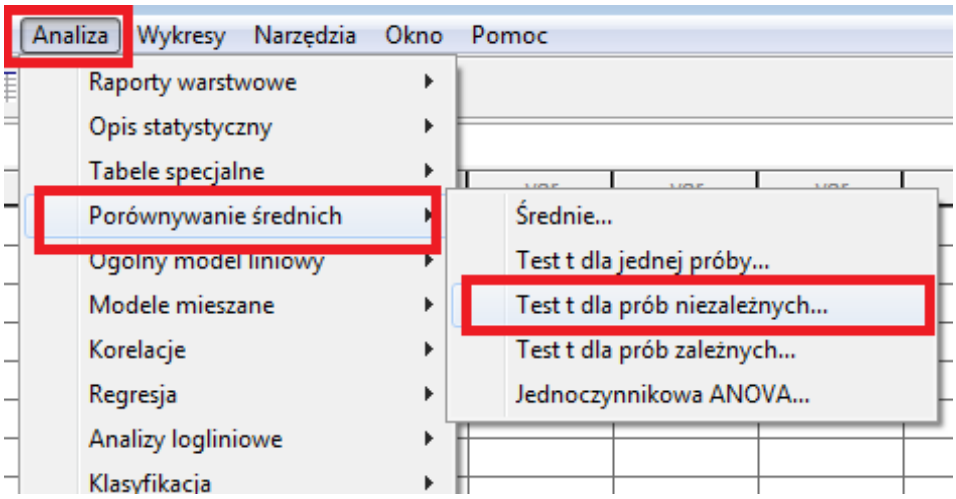
	Wartość testowana = 0					
					95% przedział ufności dla różnicy średnich	
	t	df	Istotność (dwustronna)	Różnica średnich	Dolna granica	Górna granica
wiek	62,173	199	,000	32,74500	31,7064	33,7836

3 Weryfikacja hipotez o równości dwóch średnich

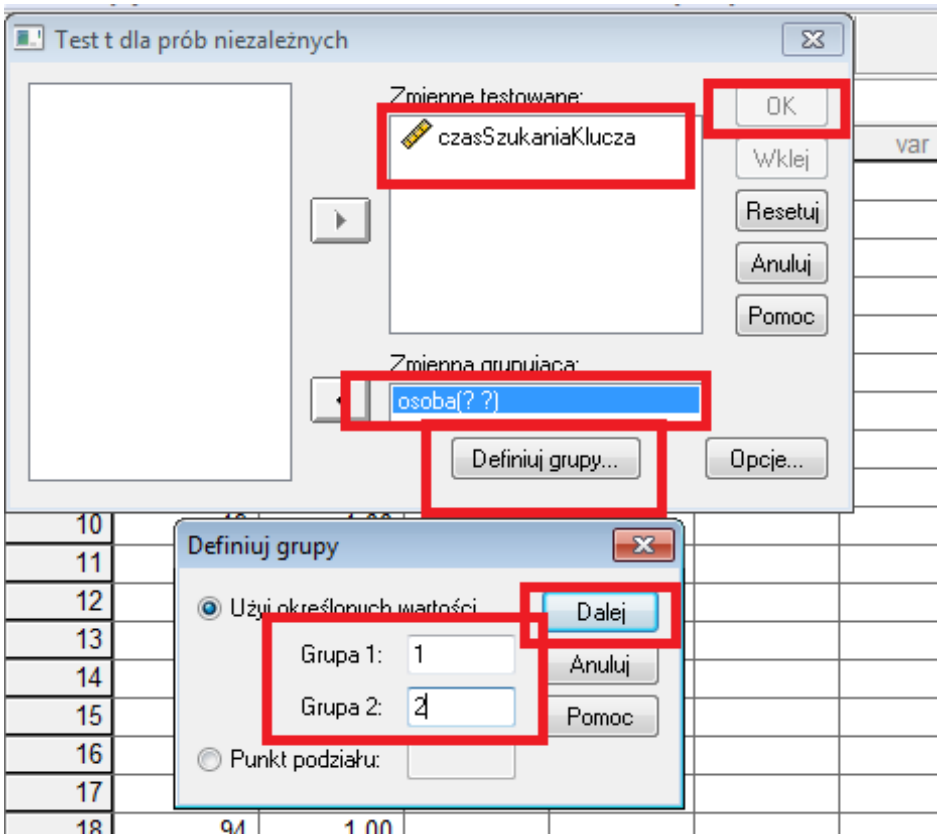
Zajmiemy się teraz problem stwierdzenia faktu, że w dwóch próbach średnia jest taka sama. Musimy tutaj rozważyć dwa przypadki. W przypadku pierwszym zakładamy niezależność dwóch zmiennych(prób), w drugim przypadku nie zakładamy niezależności. W większości przypadków intuicja sama nam podpowiada, który przypadek zastosować w danej sytuacji.

W przypadku zmiennych niezależnych w programie SPSS musimy dysponować

zmienną grupującą dla badanej cechy. Wybieramy następującą metodę



Następnie wskazujemy zmienną testowaną oraz zmienną grupującą, wskazujemy podział dla zmiennej grupującej i następnie otrzymujemy raport, w którym mamy interesującą nas odpowiedź.

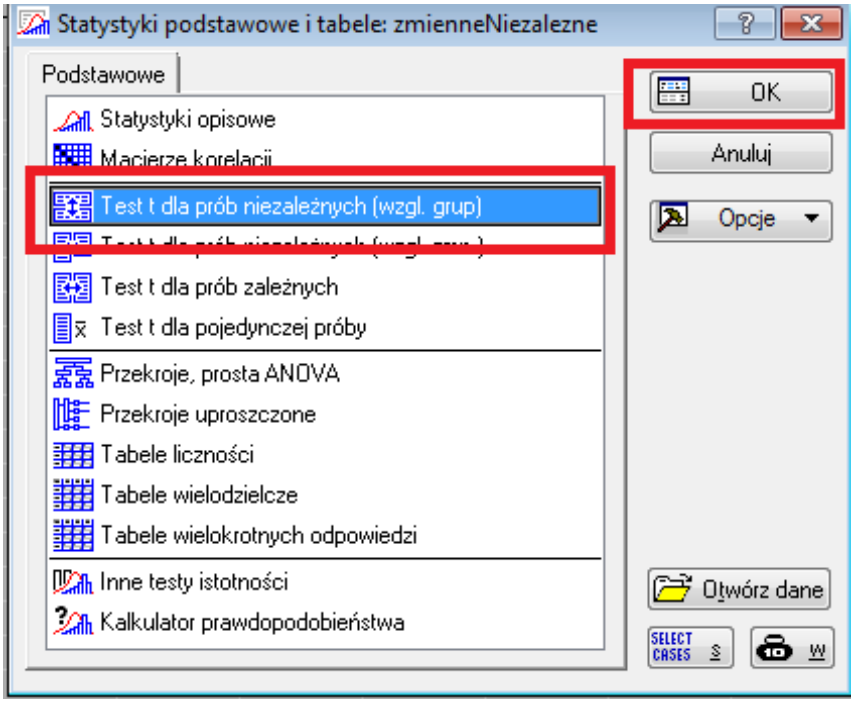


W raporcie oprócz interesującej nas istotności otrzymujemy niejako w gratisie przedział ufności dla różnicy średnich.

Test dla prób niezależnych

	Test Levene'a jednorodności wariancji		Test t równości średnich						
								95% przedział ufności dla różnicy średnich	
	F	Istotność	t	df	Istotność (dwustronna)	Różnica średnich	Błąd standardowy różnicy	Dolna granica	Górna granica
Założono równość wariancji	,074	,786	,552	98	,582	,15320	,27771	-,39790	,70430
Nie założono równości wariancji			,552	97,982	,582	,15320	,27771	-,39790	,70430

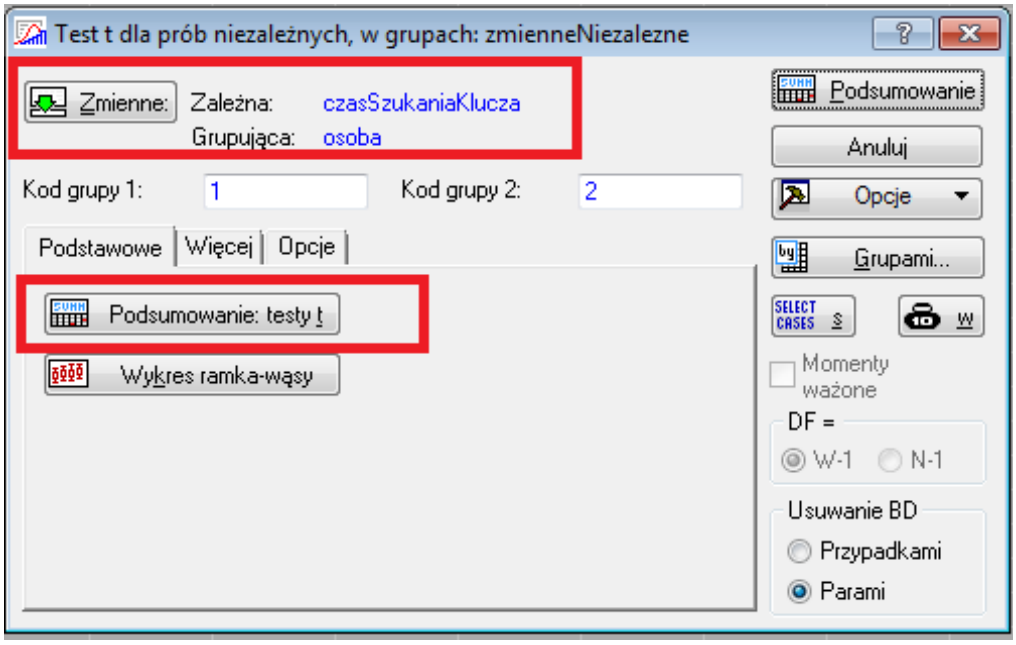
W programie Statistica dla prób niezależnych nie musimy mieć zmiennej grupującej, ale dla danych ze zmienną grupującą jest również stosowna metoda.



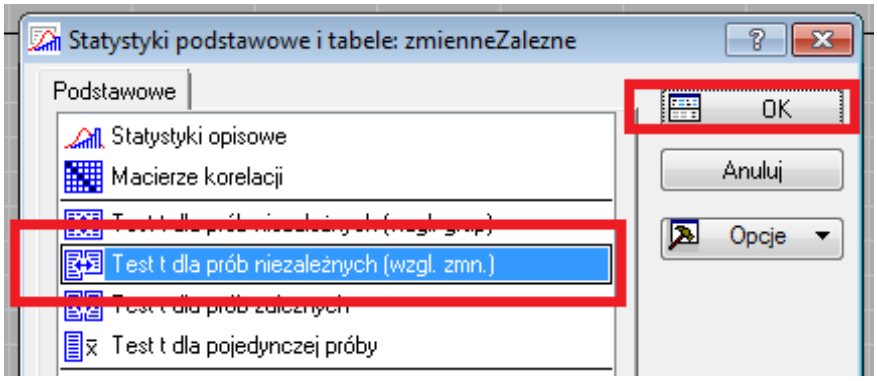
Następnie wskazujemy zmienną grupującą oraz testowaną (jak na rysunku poniżej) i otrzymujemy wynik.

Testy t; Grupująca: osoba (zmienneNiezalezne)
Grupa 1: 1
Grupa 2 2

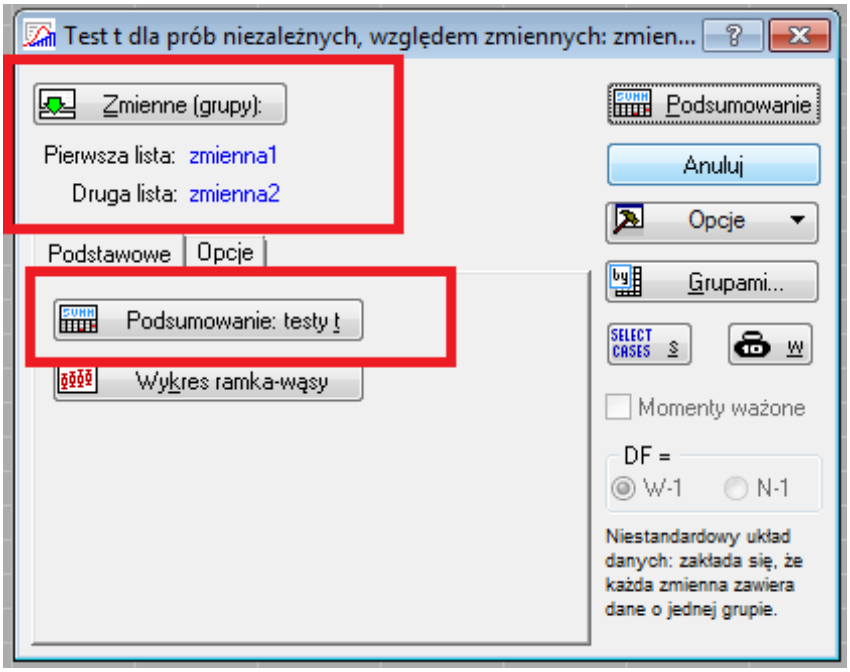
	Średnia 1	Średnia 2	t	df	p	N ważnyc 1	N ważnych 2
a	2,435800	2,282600	0,551664	98	0,582435	50	50



W programie Statistica próby niezależne możemy mieć podane w dwóch zmiennych. W takim przypadku należy zastosować inną metodę.



W odpowiednim oknie dialogowym wskazujemy, które zmienne mają być porównywane. Następnie otrzymujemy interesujące nas wyniki.



Badanie równości średniej dla prób zależnych wykonuje się w bardzo podobny sposób. Musimy jedynie dokładnie czytać wszystkie polecenia.

4 Analiza wariancji

Na wstępie zapoznamy się z metodą pozwalającą porównywać średnie w kilku grupach. Do tego typu analiz służy jednoczynnikowa analiza wariancji, tzw. jednoczynnikowa ANOVA. W pliku czasDojazdu.sav mamy informacje o czasie dojazdu na uczelnie na kolejne zjazdy. Chcemy zweryfikować hipotezę o równości średnich czasów dojazdu w poszczególnych okresach czasu. Postawmy zatem hipotezy

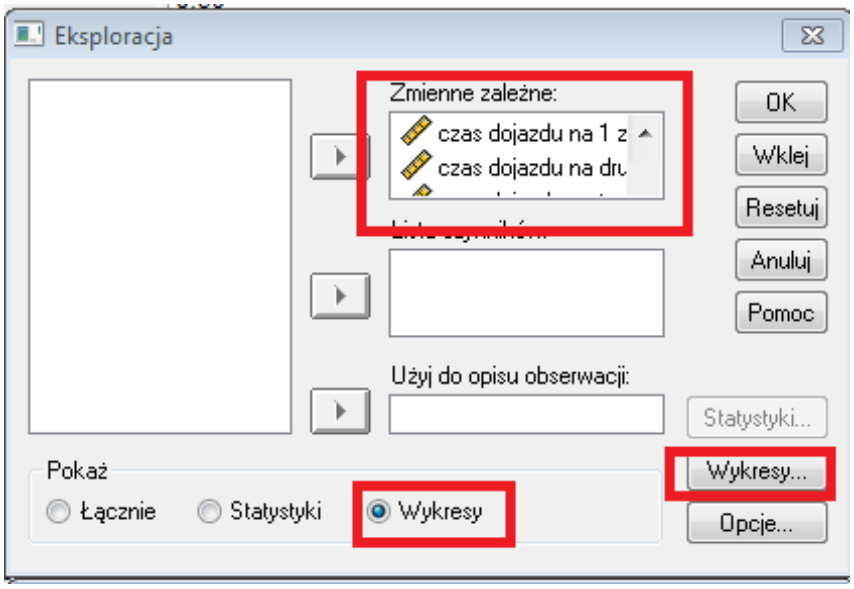
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

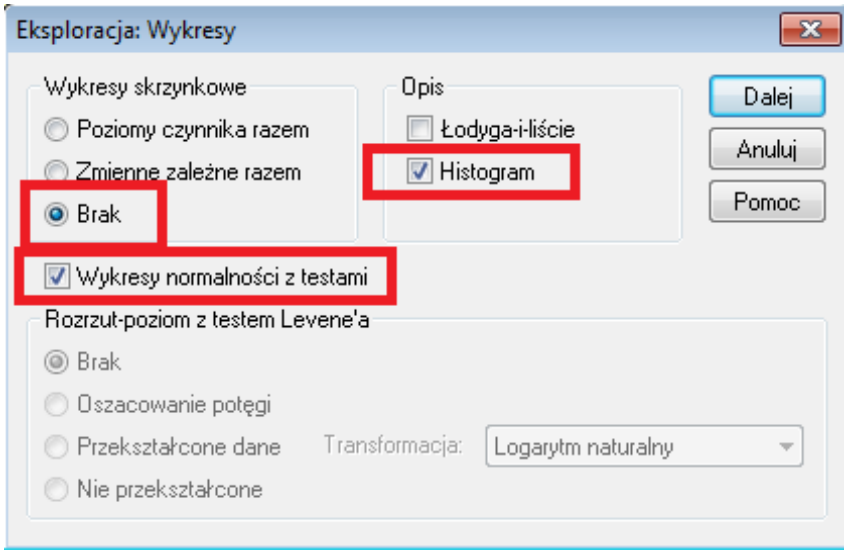
wobec hipotezy alternatywnej, która jest zaprzeczeniem H_0, tzn.

$$H_1 : \mu_1 \neq \mu_2 \vee \mu_1 \neq \mu_3 \vee \mu_1 \neq \mu_4 \vee \mu_2 \neq \mu_3 \vee \mu_2 \neq \mu_4 \vee \mu_3 \neq \mu_4$$

Zgodnie ze stosowną teorią powinniśmy sprawdzić, czy poszczególne próby pochodzą z populacji o rozkładzie normalnym. Przypomnijmy sobie jak sprawdzić,

czy nasz próba ma rozkład normalny. Wybieramy Analiza\Opis statystyczny\Eksploracja danych. Następnie postępujemy jak na poniższych rysunkach





Jako wynik otrzymujemy raport, w którym najistotniejszym punktem jest

tabela

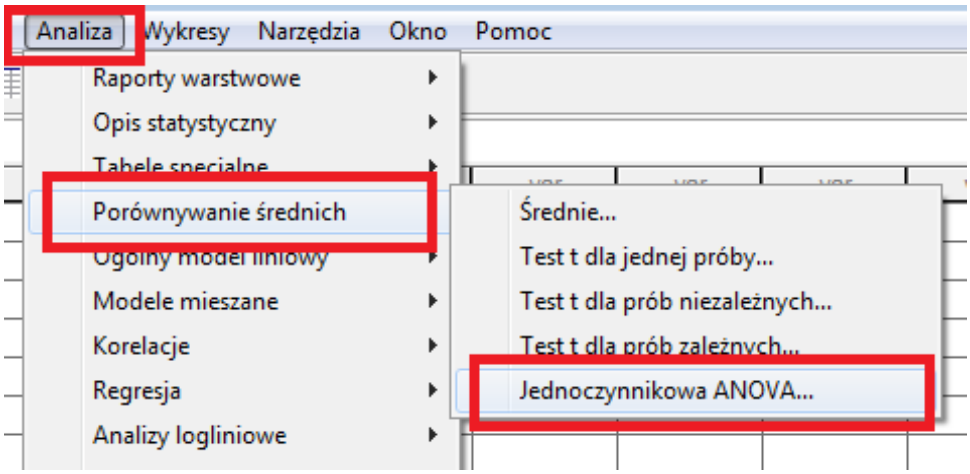
Testy normalności rozkładu

	Kołmogorow-Smirnow[a]			Shapiro-Wilk		
	Statystyka	df	Istotność	Statystyka	df	Istotność
czas dojazdu na 1 zjazd	,041	200	,200*	,986	200	,054
czas dojazdu na drugi zajzd	,036	200	,200*	,993	200	,438
czas dojazdu na trzeci zjazd	,033	200	,200*	,997	200	,988
czas dojazdu na 4 zjazd	,038	200	,200*	,994	200	,637

*. Dolna granica rzeczywistej istotności.

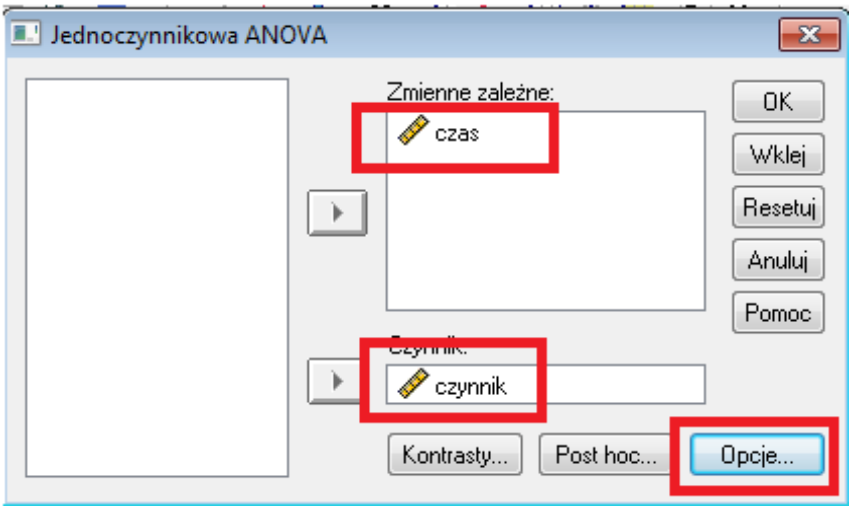
a. Z poprawką istotności Lillieforsa

Łatwo jest stwierdzić, że założenie o normalności rozkładu jest spełnione. W naszym przekonaniu umacnia nas dodatkowo analiza wykresów. W klasycznej analizie wariancji dodatkowym założeniem jest równość wariancji. Warunek ten sprawdzamy za pomocą chociażby testu Levene'a. Sprawdzenia tego dokonamy równolegle z testowaniem naszej głównej hipotezy. Zanim przystąpimy do testowania hipotezy musimy przygotować sobie nasze dane. W programie SPSS w jednej zmiennej powinny znajdować się wartości obserwacji natomiast w innej zmiennej czynnik grupujący. Dlatego dalszej analizy dokonamy na pliku czasDojazdu1.sav, w którym nasze wartości są już odpowiednio przygotowane. Wybieramy z menu jednoczynnikową ANOVA jak na poniższym rysunku

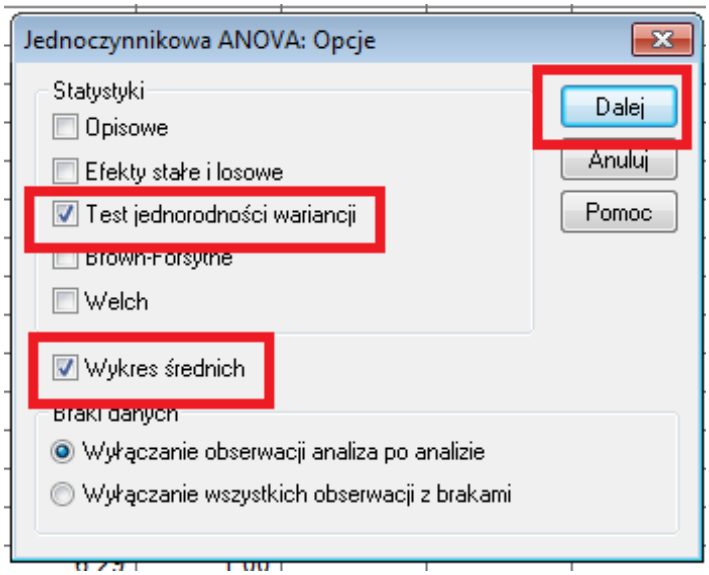


Jako zmienną zależną obieramy czas, natomiast jako czynnik zmienną czyn-

nik jak na rysunku i przechodzimy na kartę opcje



W oknie Opcje wybieramy interesujące nas wskaźniki



Jako wynik otrzymujemy raport, w którym mamy następujące dane

Jednoczynnikowa analiza wariancji (ONEWAY)

[ZbiórDanych2] D:\Users\Adam\Desktop\statistica\sad\czasDojazdu1.sav

Test jednorodności wariancji

czas

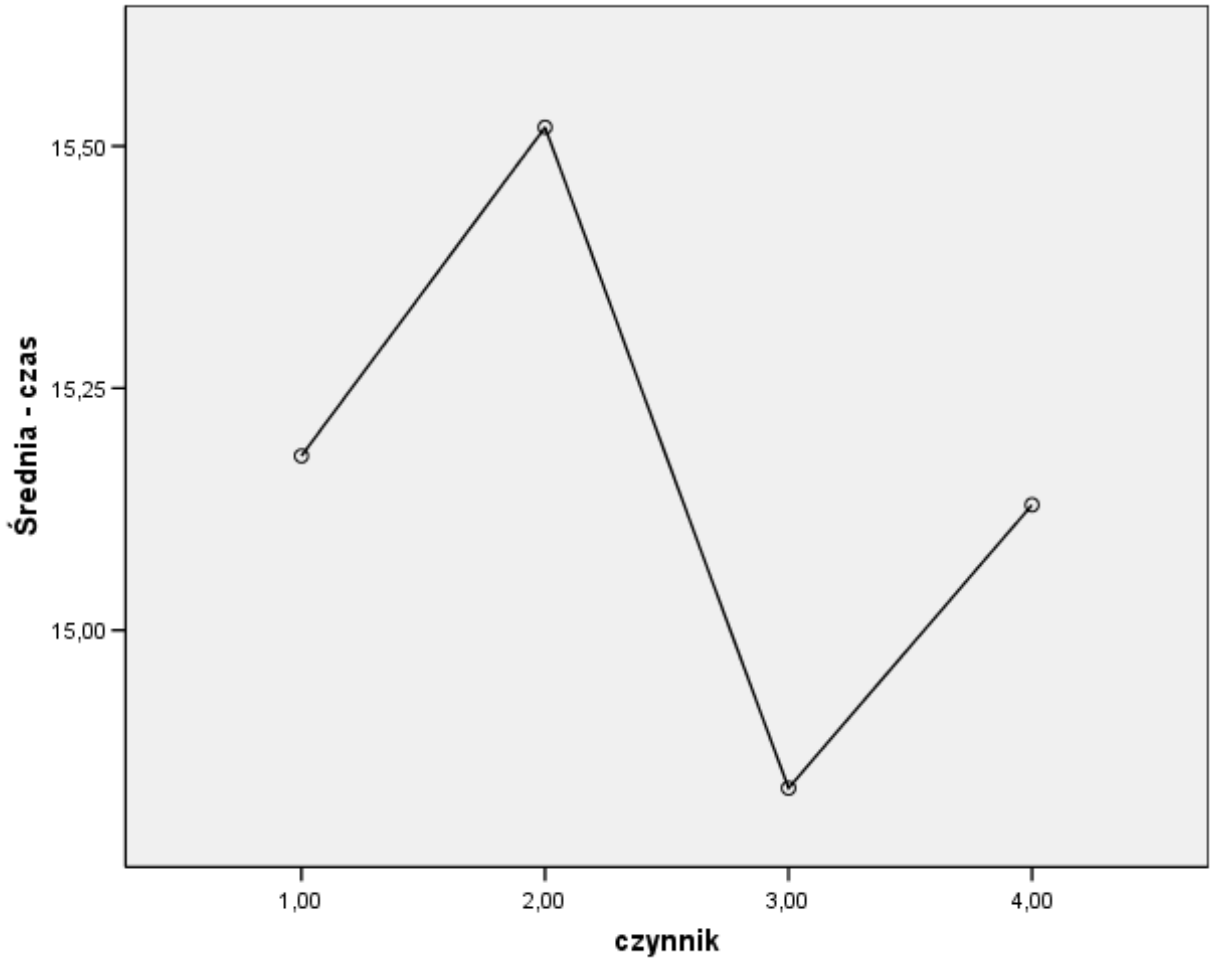
Test Levene'a	df1	df2	Istotność
1,649	3	796	,177

Jednoczynnikowa ANOVA

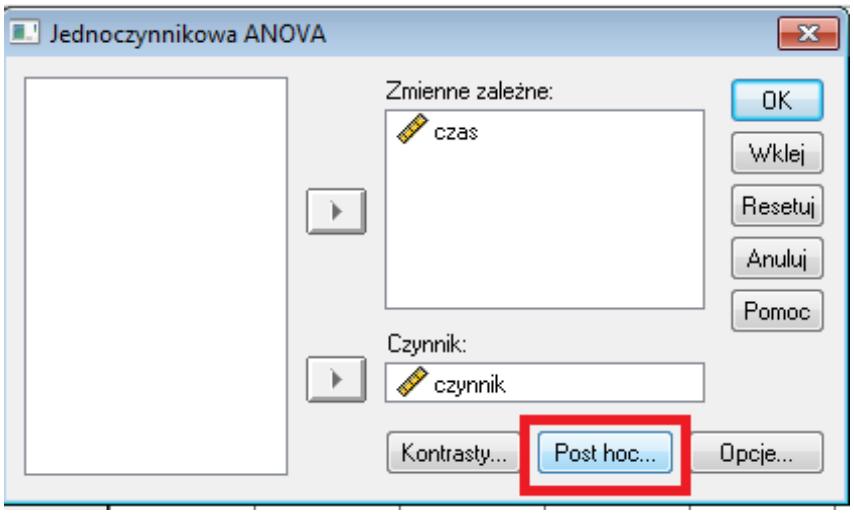
czas

	Suma kwadratów	df	Średni kwadrat	F	Istotność
Między grupami	46,926	3	15,642	,633	,594
Wewnątrz grup	19657,498	796	24,695		
Ogółem	19704,424	799			

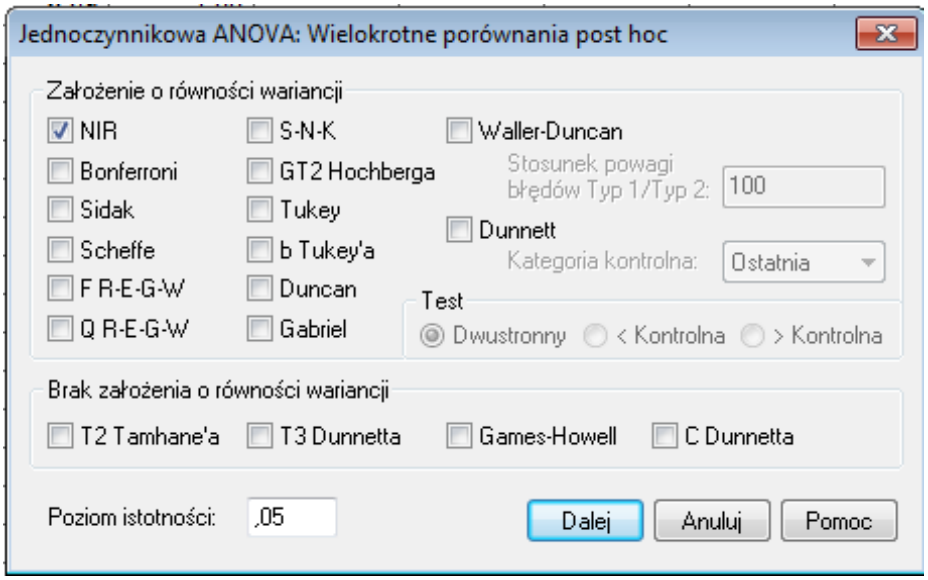
z których wynika, że założenie jednorodności wariancji jest spełnione oraz nie ma podstaw do odrzucenia hipotezy H_0 o równości wariancji. Ponadto zgodnie, z tym co wybraliśmy na karcie opcje otrzymaliśmy wykres reprezentujący poszczególne średnie. Możemy na nim łatwo sprawdzić słuszność naszego osądu.



Teraz spróbujemy stwierdzić, w której grupie jest najwyższa średnia i pomiędzy, którymi parami występuje statystycznie istotna różnica pomiędzy średnimi. W tym celu wykorzystamy test post-hoc



Mamy do dyspozycji wiele testów skorzystamy jednak z testu zaproponowanego przez twórcę analizy wariancji, tj testu NIR. Wybór ten jest jak najbardziej uzasadniony ponieważ już wiemy, że wariancje są równe.



W wynikowym raporcie z łatwością odnajdujemy tabelę

Porównania wielokrotne

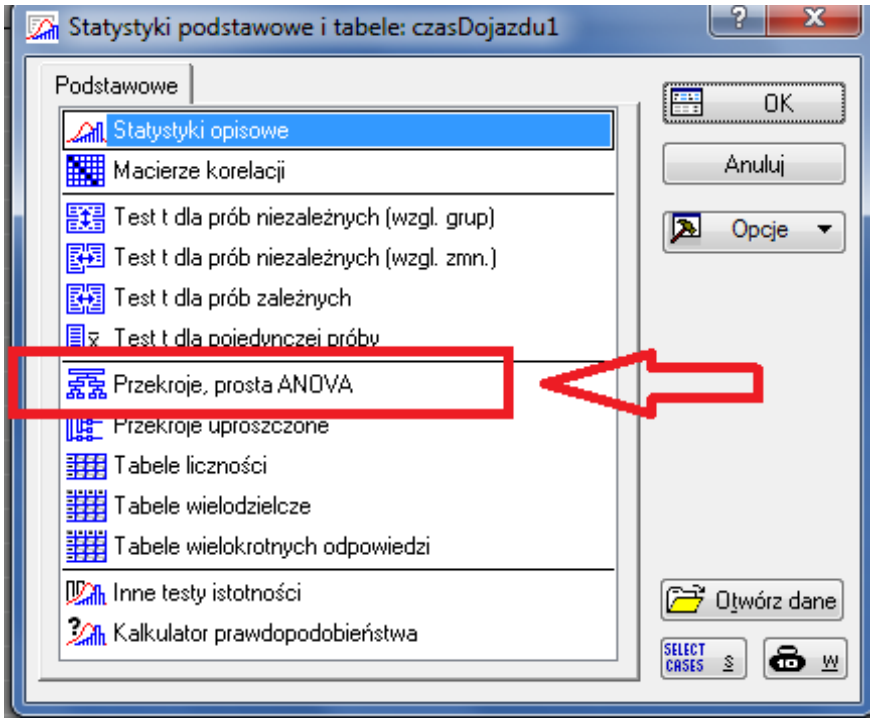
Zmienna zależna: czas

Test NIR

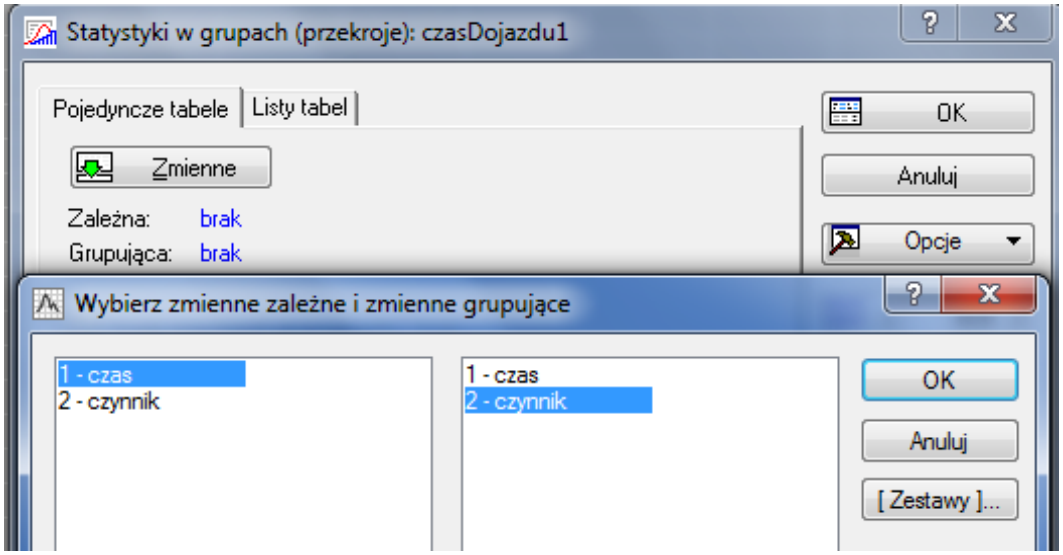
(I) czynnik	(J) czynnik	Różnica średnich (I-J)	Błąd standardowy	Istotność	95% przedział ufności	
					Dolna granica	Górna granica
1,00	2,00	-,33935	,49694	,495	-1,3148	,6361
	3,00	,34300	,49694	,490	-,6325	1,3185
	4,00	,05055	,49694	,919	-,9249	1,0260
2,00	1,00	,33935	,49694	,495	-,6361	1,3148
	3,00	,68235	,49694	,170	-,2931	1,6578
	4,00	,38990	,49694	,433	-,5856	1,3654
3,00	1,00	-,34300	,49694	,490	-1,3185	,6325
	2,00	-,68235	,49694	,170	-1,6578	,2931
	4,00	-,29245	,49694	,556	-1,2679	,6830
4,00	1,00	-,05055	,49694	,919	-1,0260	,9249
	2,00	-,38990	,49694	,433	-1,3654	,5856
	3,00	,29245	,49694	,556	-,6830	1,2679

Analiza powyższej tabeli pozwala nam stwierdzić, że pomiędzy żadną parą nie występuje statystycznie istotna różnica dla średnich. Ponadto najwyższa średnia jest w drugiej grupie a najniższa w 3.

Teraz nadszedł czas na analogiczne rozważania w programie Statistica. Skorzystamy z już przygotowanego pliku czasDojazdu. W programie Statistica testy ANOVA można odszukać w kilku miejscach. Nam w zupełności wystarczy na razie moduł dostępny menu Statystyka\Statystyki podstawowe i tabele.

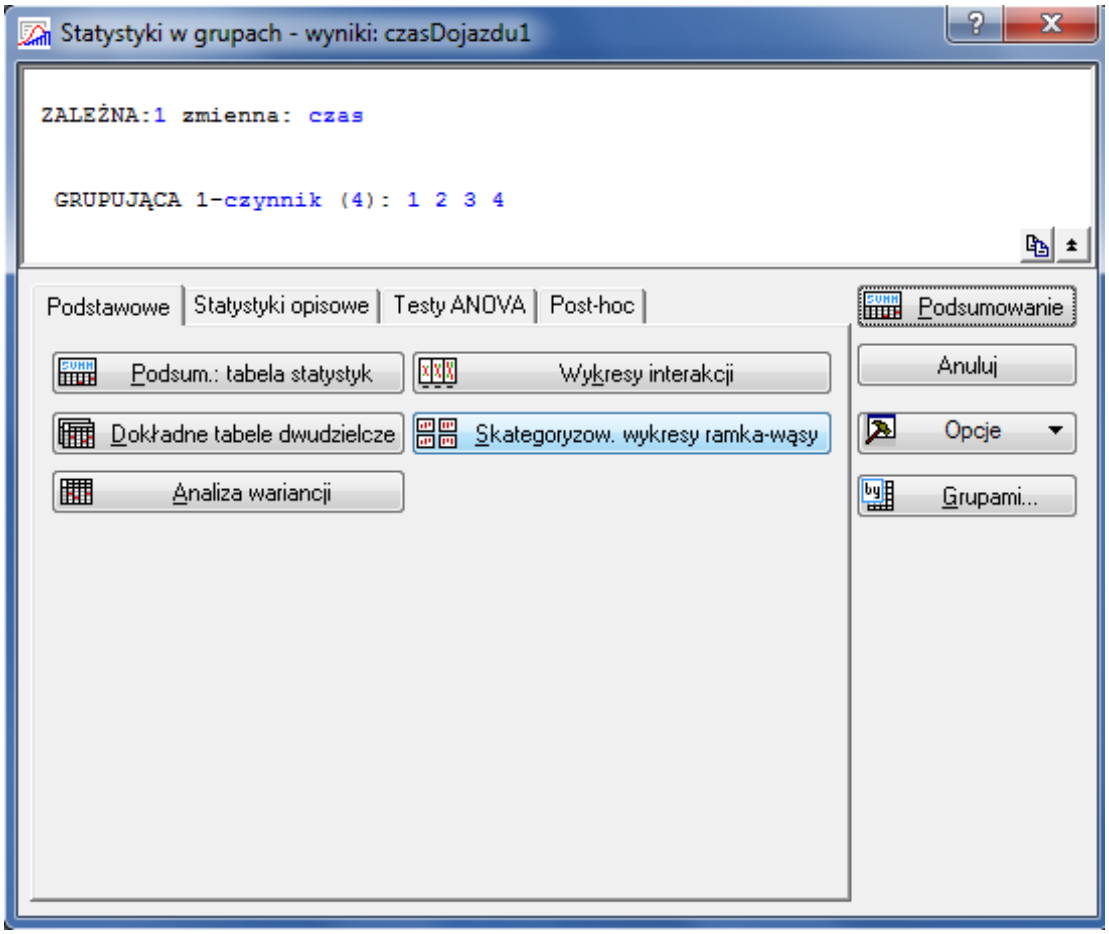


Po wskazaniu odpowiednich zmiennych

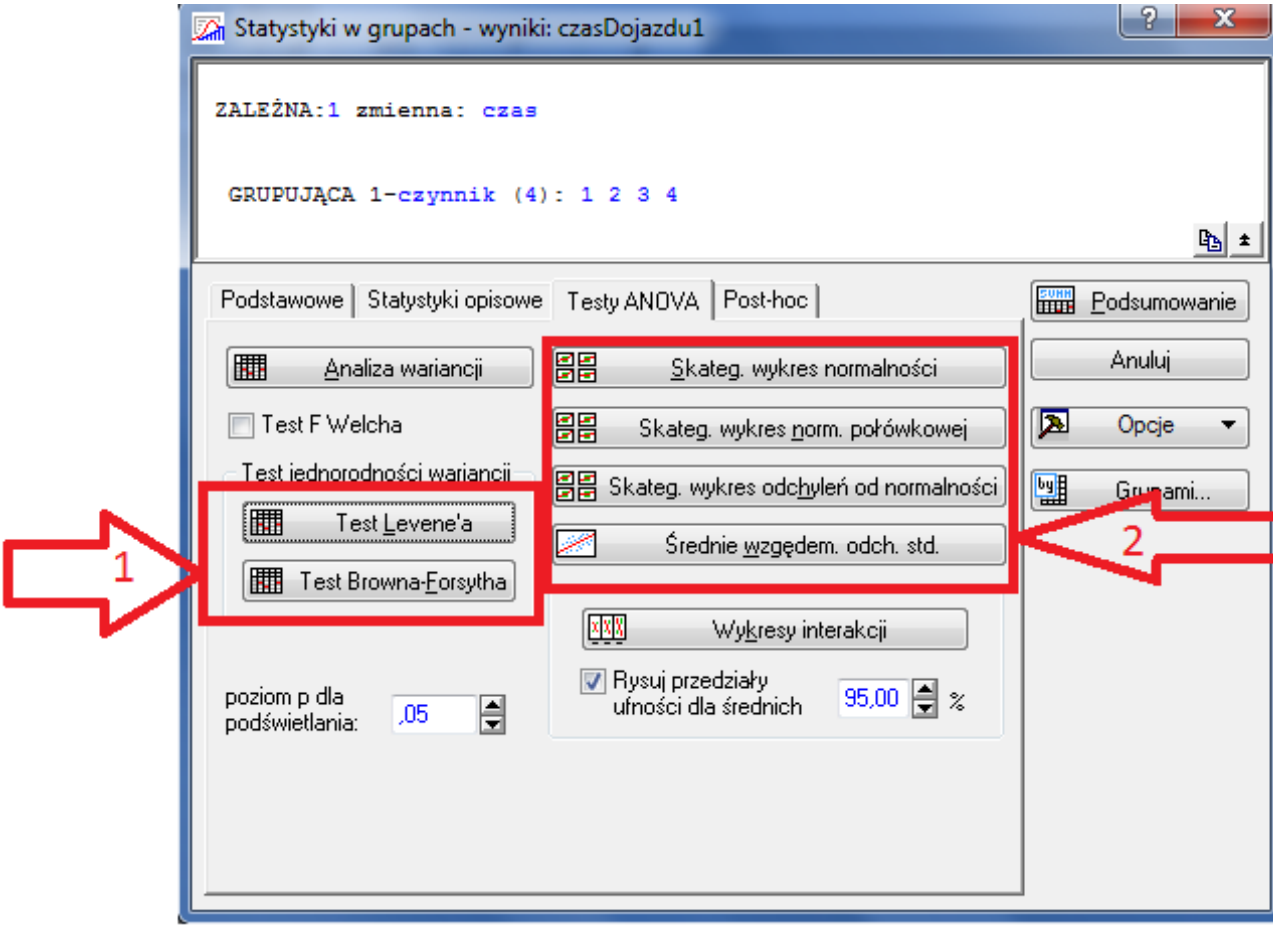


przechodzimy dalej i mamy okienko, w którym mamy kilka testów związanych

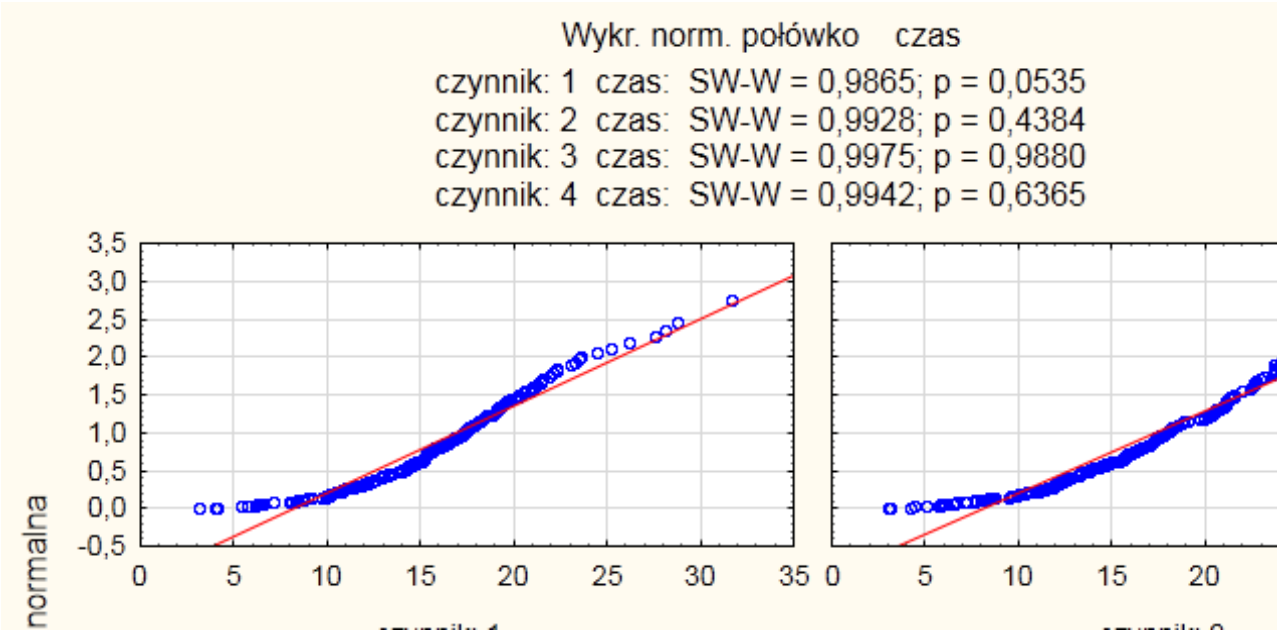
z analizą wariancji



Na zakładce "Testy ANOVA" mamy m. in. do dyspozycji testy jednorodności wariancji (1), jak również skategoryzowane wykresy normalności (2) pozwalające sprawdzić nam założenia klasycznej analizy wariancji.

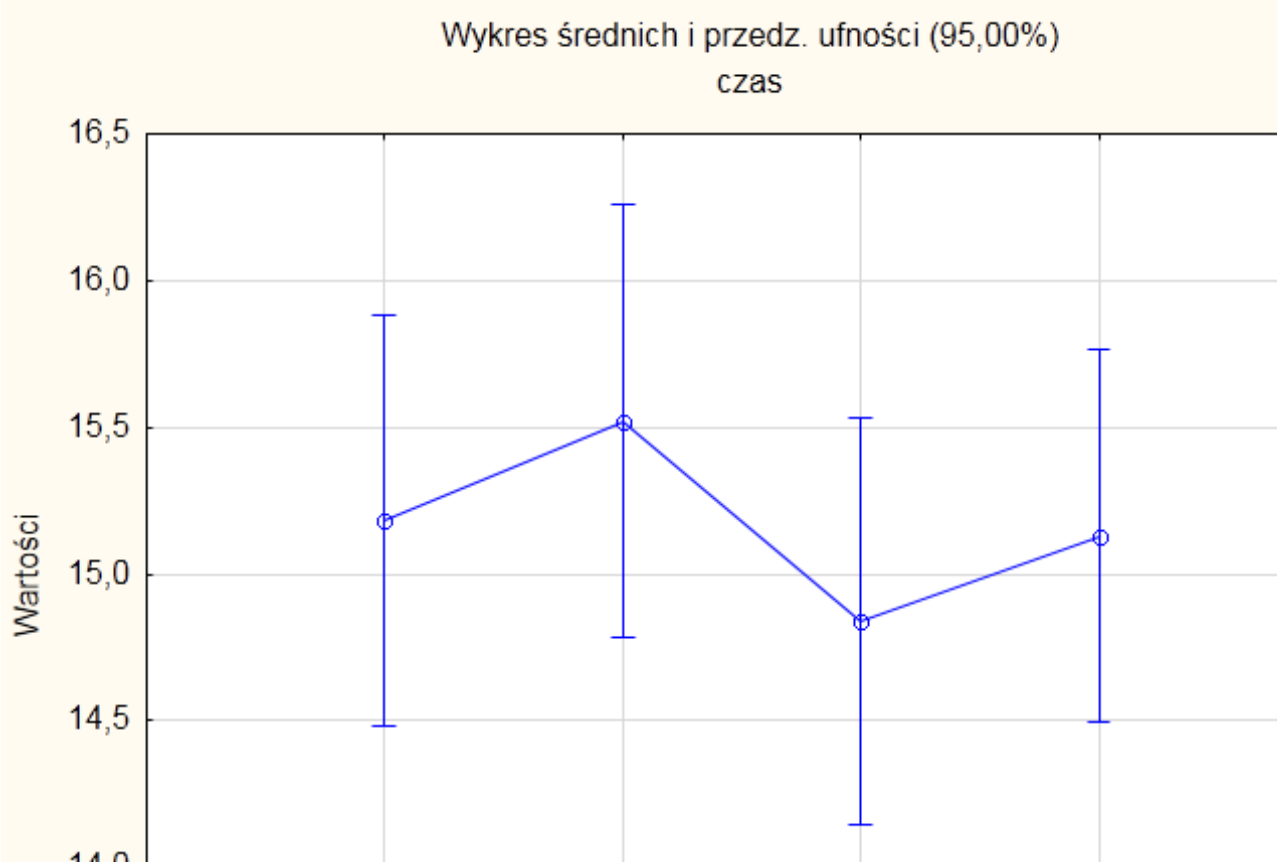


Domyślnie przy wyborze wykresów skategoryzowanych mamy jedynie same wykresy, jednak po dwukrotnym kliknięciu w obszar wykresu możemy wybrać opcję wyświetlającą wyniki testu Shapiro-Wilka.



Oczywiście jeśli ktoś woli mieć wyznaczony histogramy z nałożonymi wynikami testu badającego normalność to można skorzystać z zakładki Statystyki opisowe. Dość ciekawym sposobem wizualizacji danych jest wykres interakcji dostępny na zakładce podstawowe. Jako wynik otrzymujemy wykres, na którym oprócz śred-

nich zaznaczone są 95% przedziały ufności dla średniej.



Jeśli będziemy dysponowali zapasem czasu to powrócimy do tematu i omówimy nieco bardziej skomplikowane zagadnienie jakim jest wieloczynnikowa analiza wariancji.

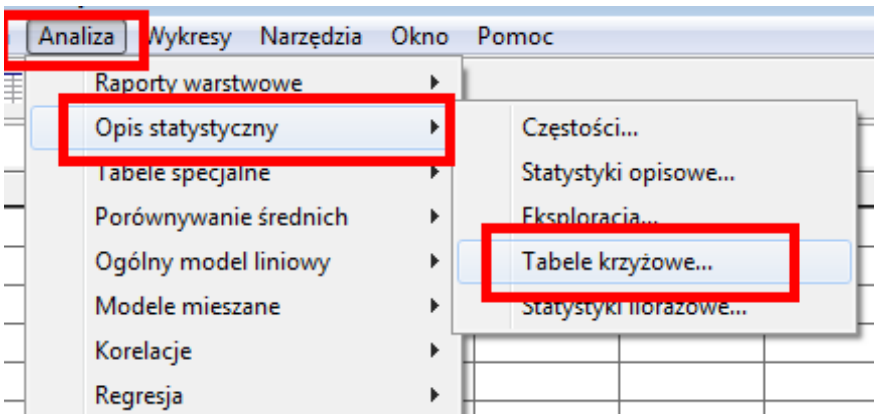
5 Testowanie niezależności

W wielu przypadkach interesuje nas sprawdzenie, czy istnieją zależności pomiędzy pewnymi próbami. Możemy np. wyobrazić sobie sytuację, że chcemy sprawdzić czy istnieje zależność pomiędzy ocenami z "Podstaw statystyki matematycznej" oraz "Podstaw statystyki opisowej". W pliku ocenyPodstawy.sav mamy informacje o ocenach pewnej grupy studentów. Chcemy dokonać weryfikacji następujących hipotez

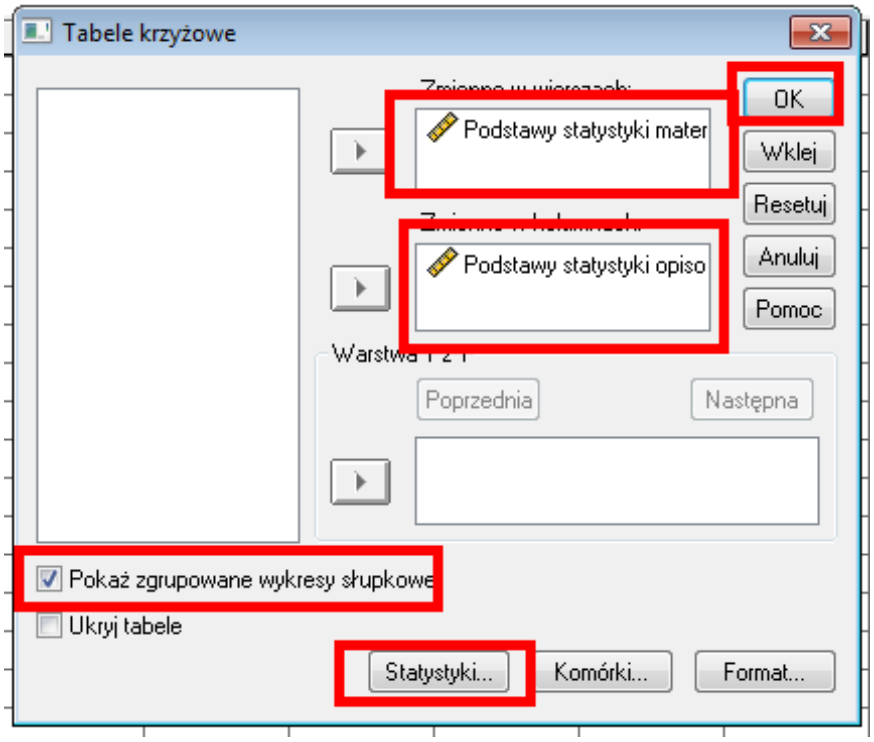
H_0 : oceny z "Podstaw statystyki matematycznej" i oceny z "Podstaw statystyki opisowej" są niezależne

H_1 : istnieje zależność pomiędzy ocenami

W klasycznej statystyce do weryfikacji hipotez o niezależności stosuje się test niezależności $\mathcal{X}^2$. W programie SPSS odnajdujemy go w nieco zaskakującym miejscu



Jedną z naszych zmiennych wskazujemy jako zmienną w wierszu, drugą jako kolumnę (jest to bez znaczenia), w oknie statystyki wskazujemy test chi-kwadrat.



Jako wynik otrzymujemy raport, w którym mamy odpowiedź na interesujące nas pytanie, tabele krzyżowe zależności pomiędzy poszczególnymi ocenami oraz dość interesujący wykres reprezentujący zależności opisane w tabeli krzyżowej.

Testy Chi-kwadrat

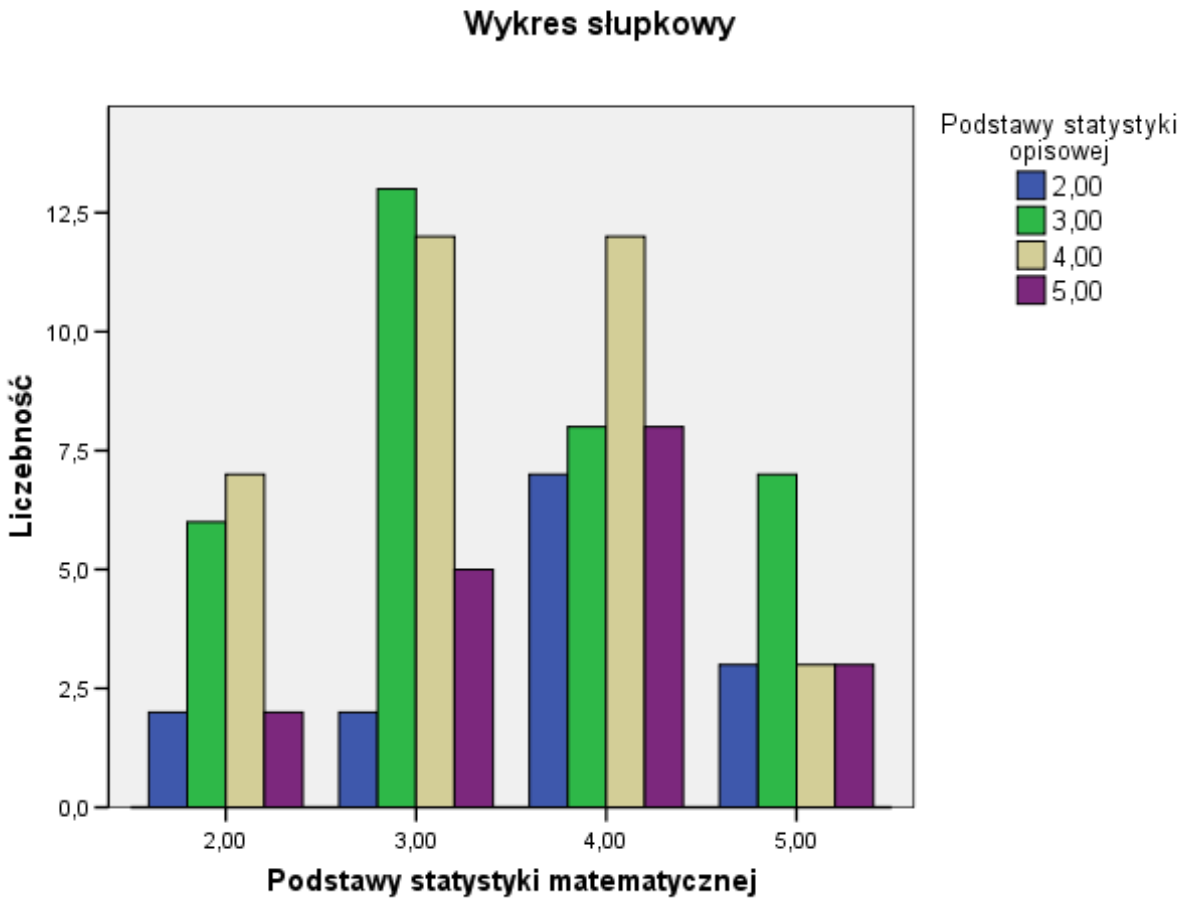
	Wartość	df	Istotność asymptotyczna (dwustronna)
Chi-kwadrat Pearsona	7,138[a]	9	,623
Iloraz wiarygodności	7,691	9	,566
Test związku liniowego	,196	1	,658
N Ważnych obserwacji	100		

a. 37,5% komórek (6) ma liczebność oczekiwaną mniejszą niż 5. Minimalna liczebność oczekiwana wynosi 2,24.

Tabela krzyżowa Podstawy statystyki matematycznej * Podstawy statystyki opisowej

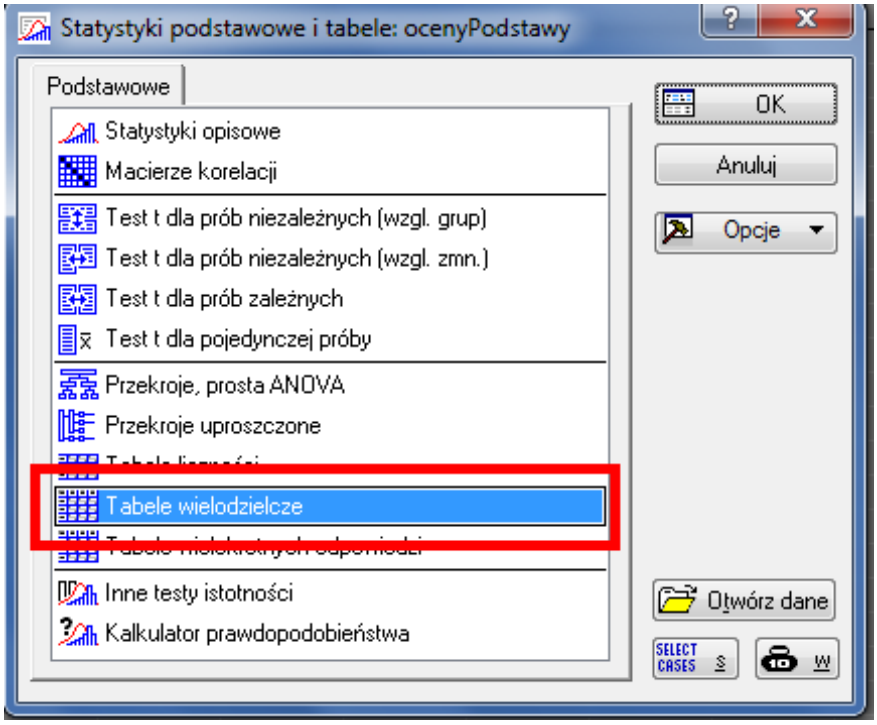
Liczebność

		Podstawy statystyki opisowej				Ogółem
		2,00	3,00	4,00	5,00	
Podstawy statystyki matematycznej	2,00	2	6	7	2	17
	3,00	2	13	12	5	32
	4,00	7	8	12	8	35
	5,00	3	7	3	3	16
Ogółem		14	34	34	18	100

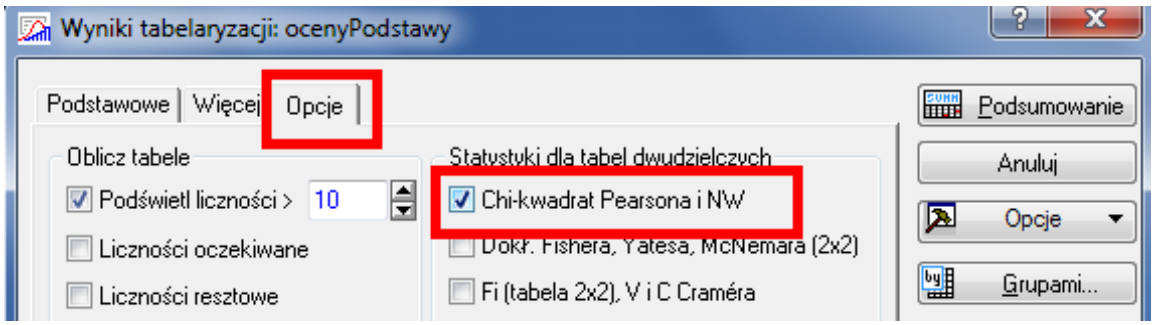


W programie Statistica test ten odnajdujemy w części Statystyki podsta-

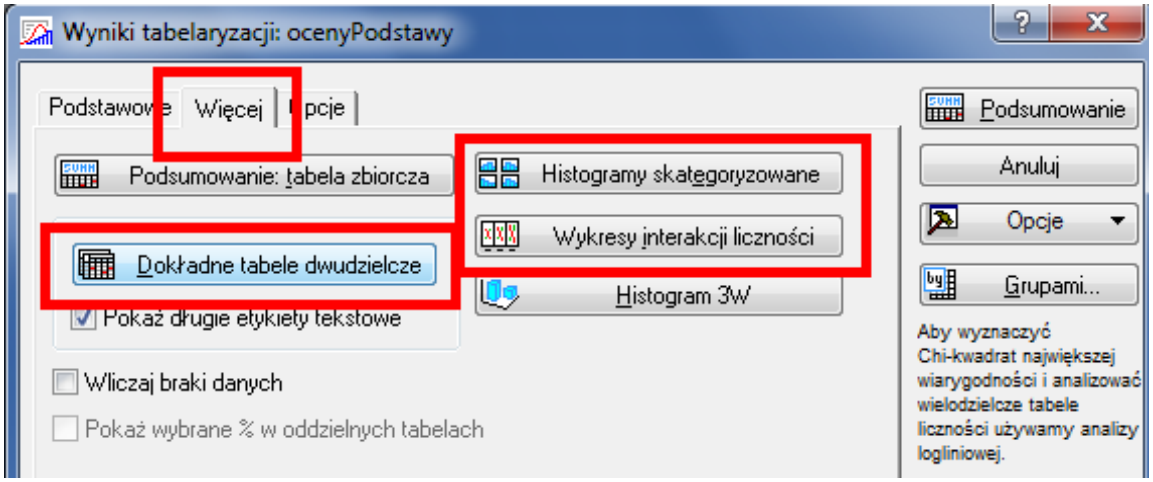
wowe i tabele.



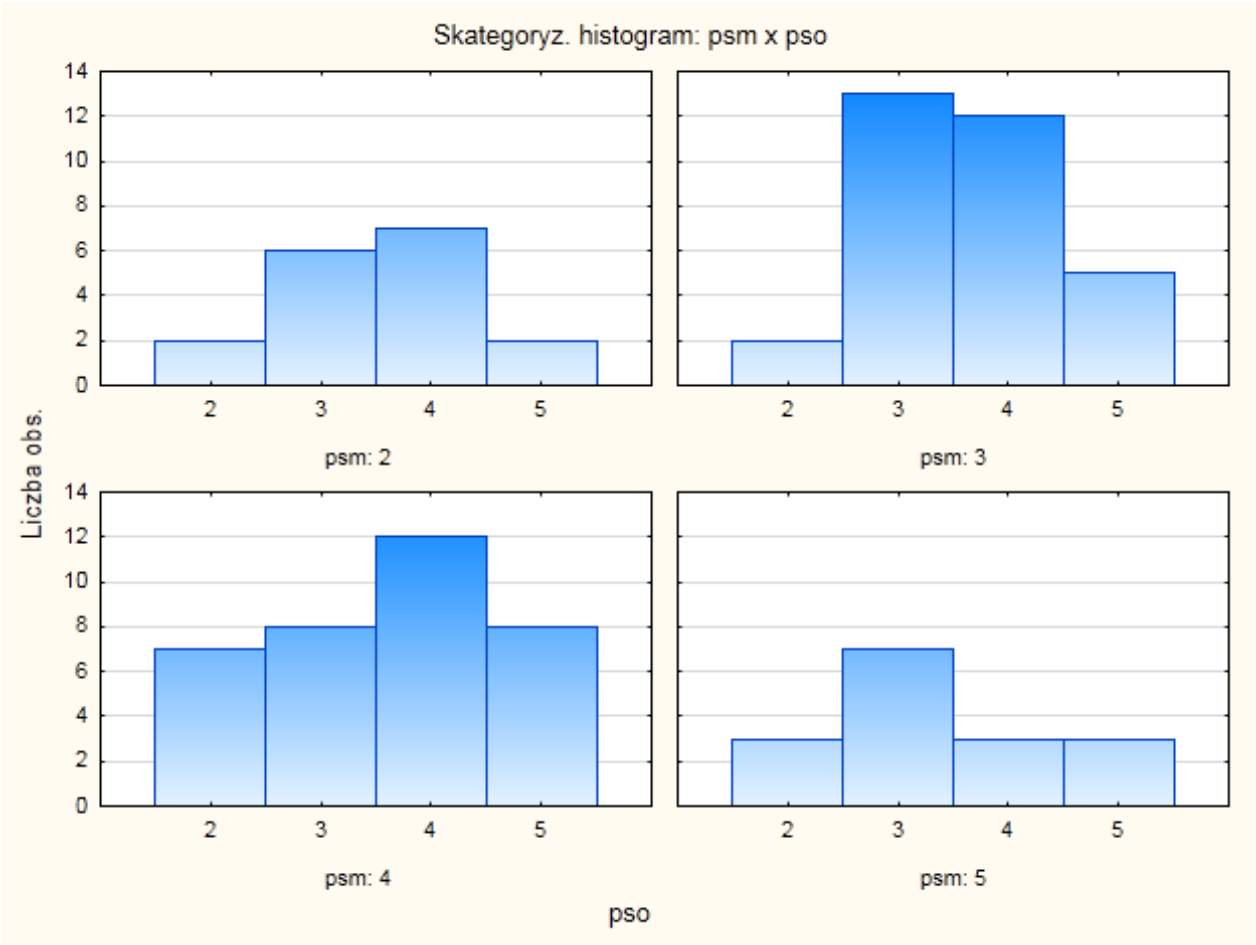
Po określeniu interesujących nas zmiennych w zakładce opcje wybieramy stosowny test



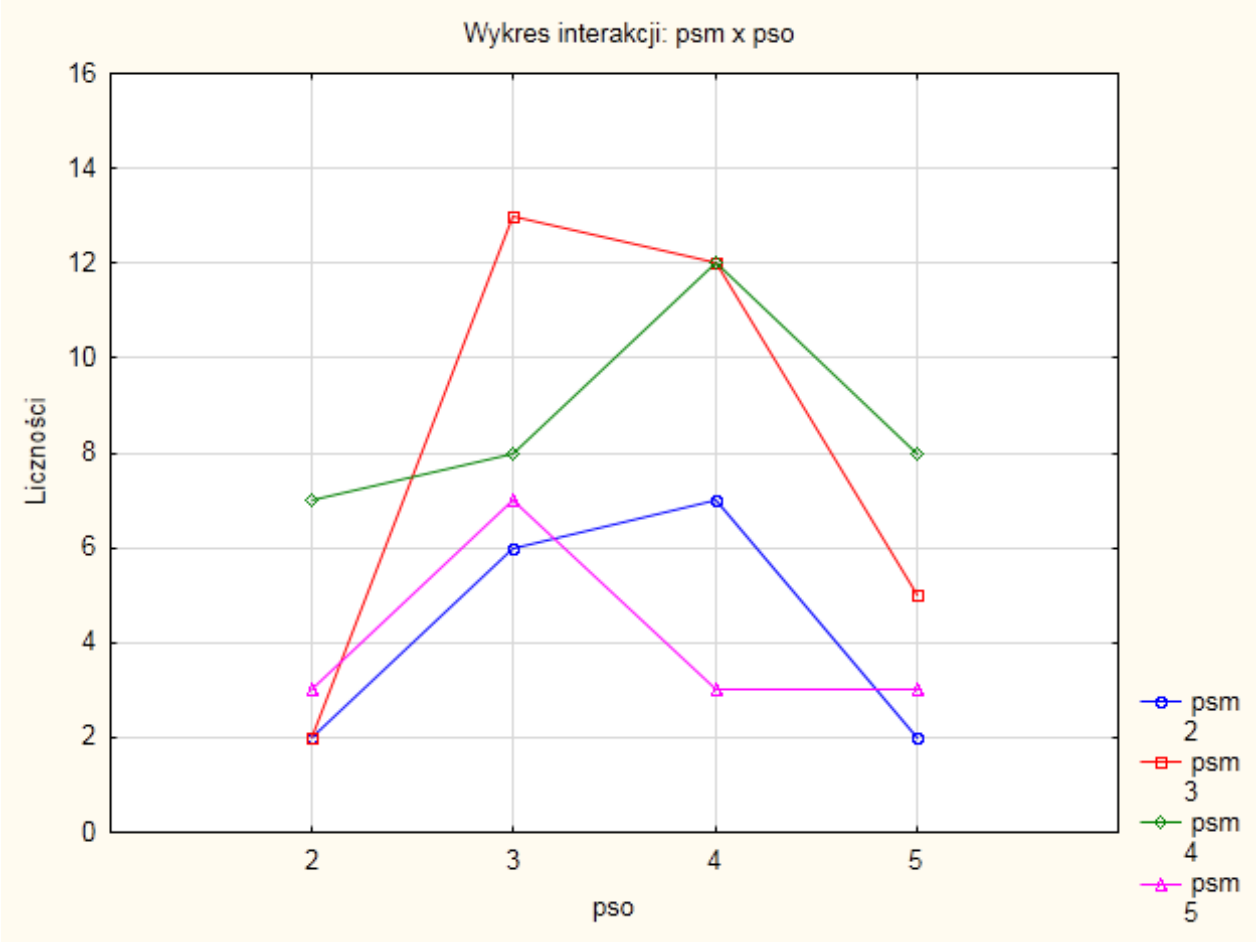
oraz w zakładce Więcej wybieramy dokładne tabele dwudzielcze



Dość interesujące wyniki otrzymujemy wybierając na zakładce Więcej Histogramy skategoryzowane oraz Wykresy interakcji liczności. W pierwszym przypadku otrzymujemy histogramy dla poszczególnych wartości jednej zmiennej.



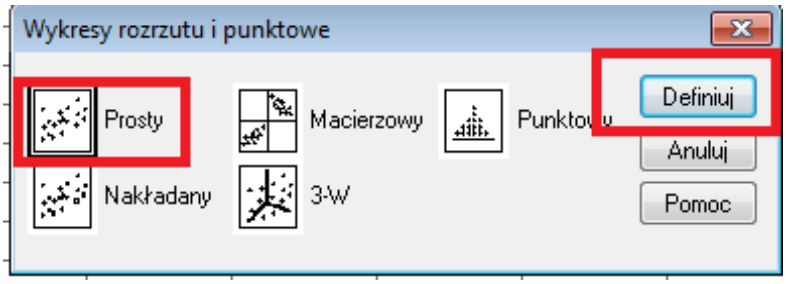
W naszym konkretnym przypadku są to histogramy zmiennej pso dla poszczególnych wartości psm. W drugim przypadku mamy natomiast



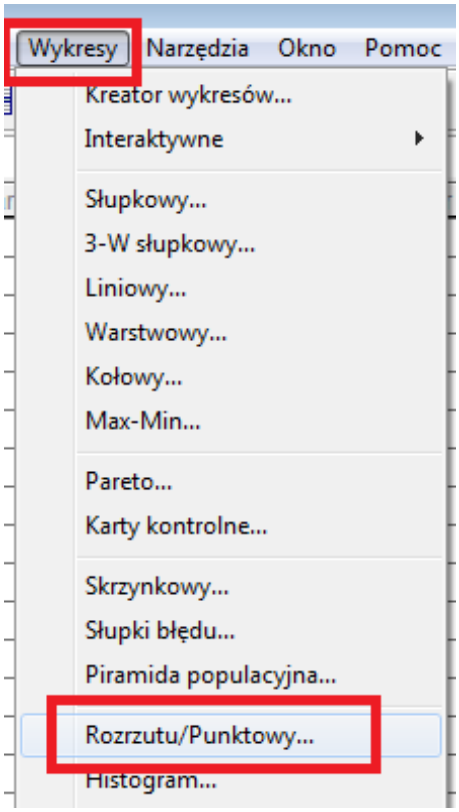
zobrazowane w sposób liniowy analogiczne zależności.

6 Regresja

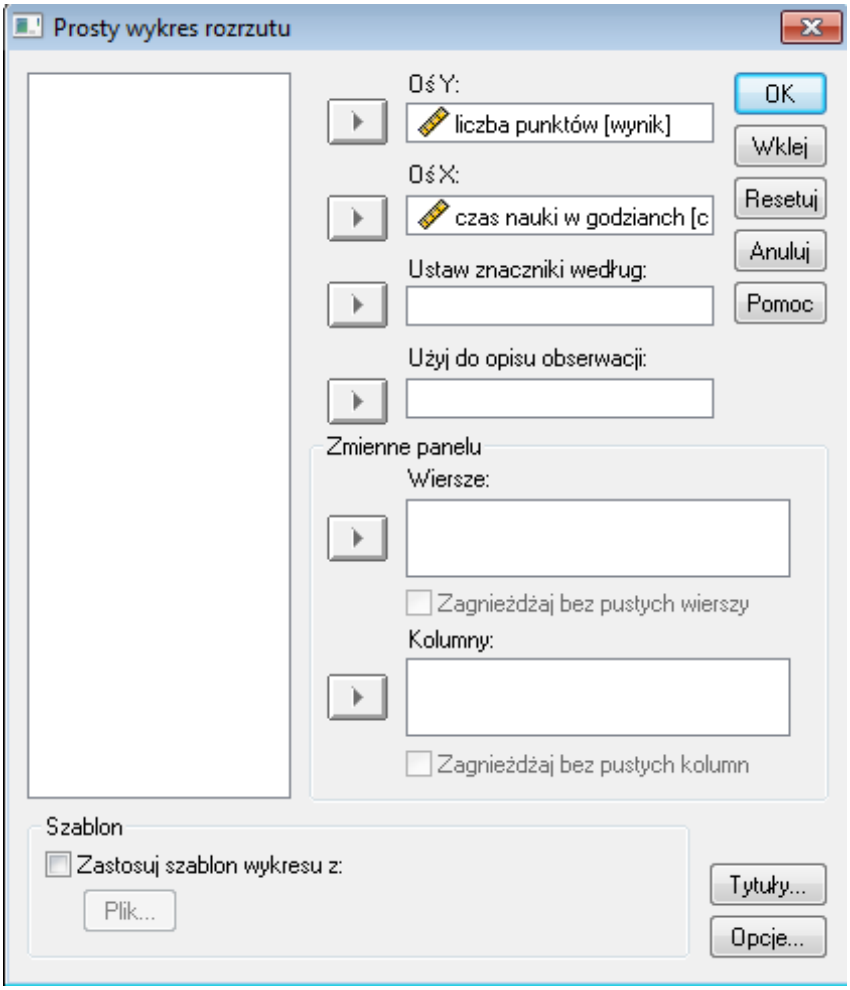
Do tej pory poznaliśmy już metody pozwalające stwierdzić, czy dwie próby są niezależne. W tym miejscu spróbujemy opisać rodzaj zależności pomiędzy badanami cechami. Najprostszym sposobem zależności jest zależność liniowa pomiędzy dwiema cechami. Rozważmy przykładową zależność pomiędzy czasem nauki a wynikiem z egzaminu. Wykres rozrzutu danych z pliku "nauka.sav" pozwala nam oszacować rodzaj zależności. W tym celu wybieramy prosty wykres



rozrzutu

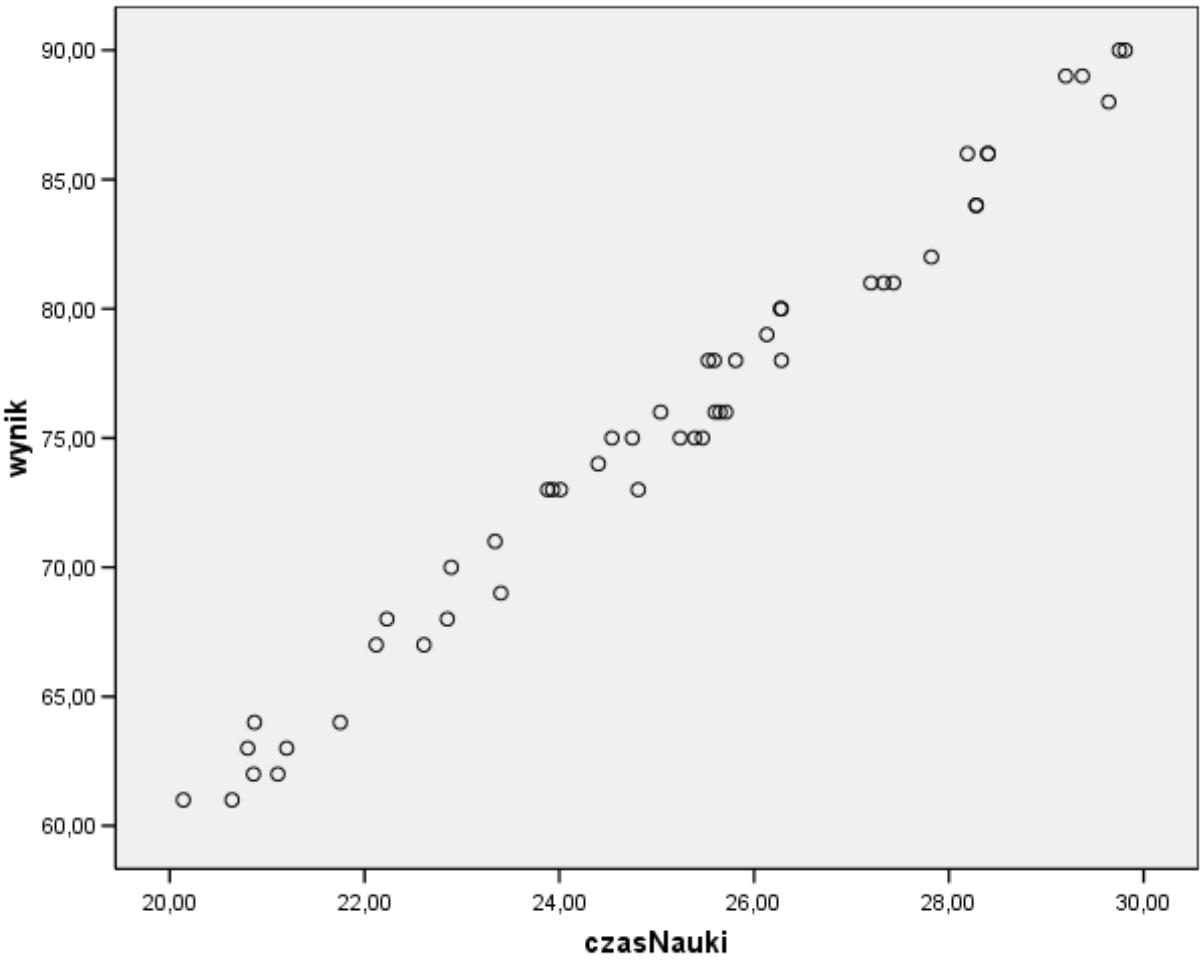


Następnie wskazujemy interesujące nas zmienne, teoretycznie kolejność zmiennych jest bez znaczenia, ale w sposób naturalny odczuwamy, że wynik zależy od czasu nauki, a nie odwrotnie.

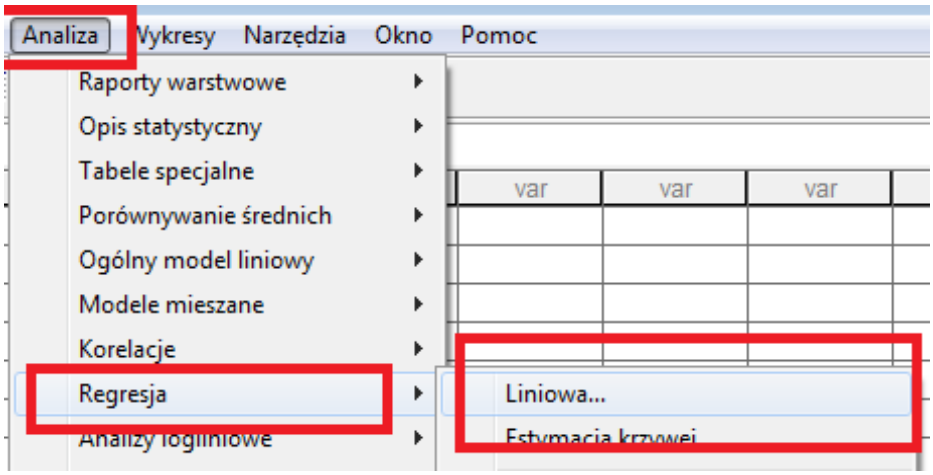


Jako wynik otrzymujemy raport, w którym mamy w sposób graficzny zaprezen-

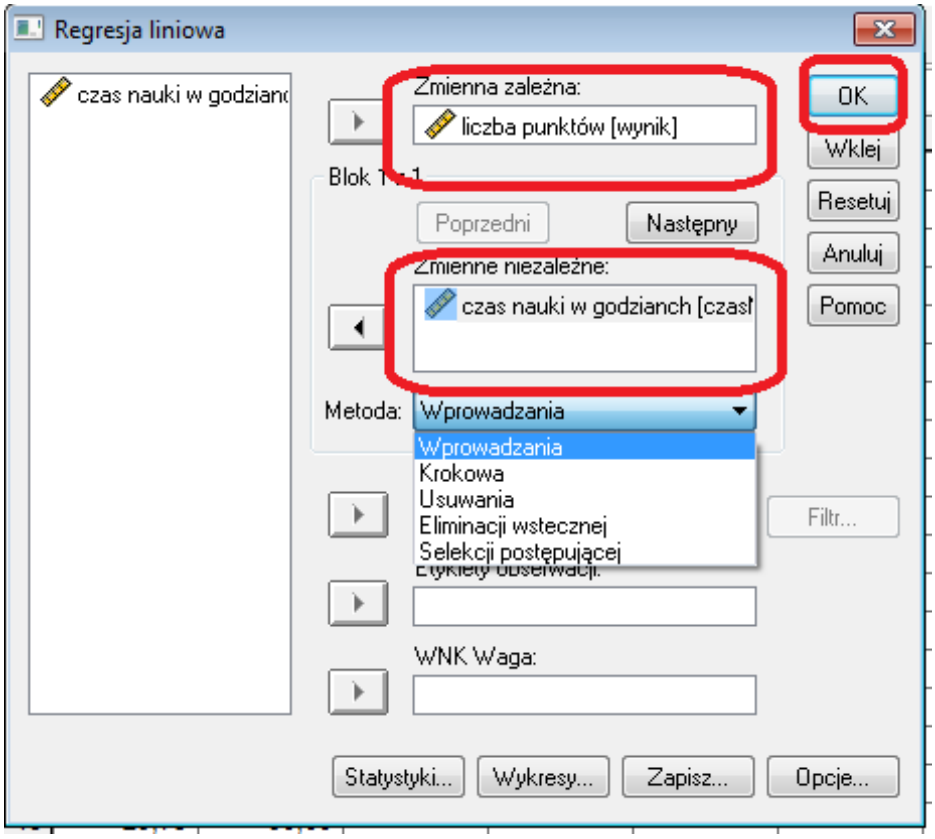
towane występujące zależności



Łatwo zauważyć, że nasze obserwacje rozkładają się wokół pewnej prostej, naszym celem będzie wyznaczenie równania tej prostej. W tym celu skorzystamy z modułu regresja liniowa



Następnie określamy zmienną zależną i niezależną



W wynikowym raporcie najważniejsza jest następująca tabelka.

Współczynniki[a]

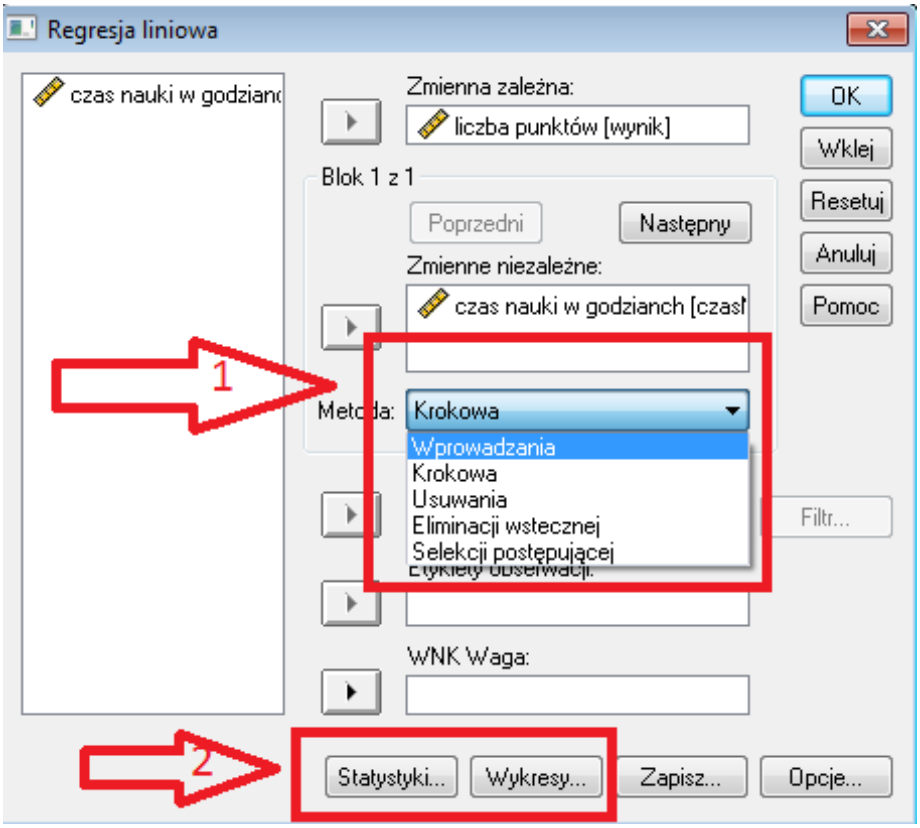
Model		Współczynniki niestandaryzowane		Współczynniki standaryzowane	t	Istotność
		B	Błąd standardowy	Beta		
1	(Stała)	-,153	1,392		-,110	,913
	czas nauki w godzianch	3,010	,055	,992	54,651	,000

a. Zmienna zależna: liczba punktów

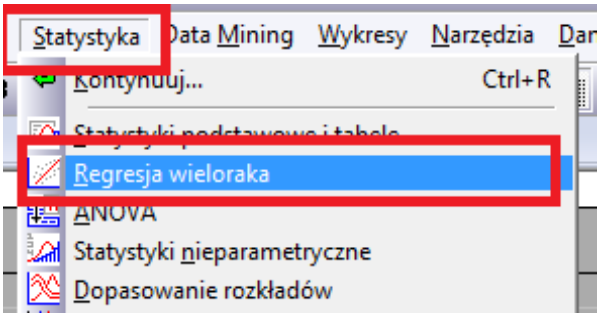
z której możemy odczytać interesujący nas wzór prostej. Okazuje się, że zależność wyniku w zależności od czasu nauki opisuje prosta

$$y = 3.010x - 0.153$$

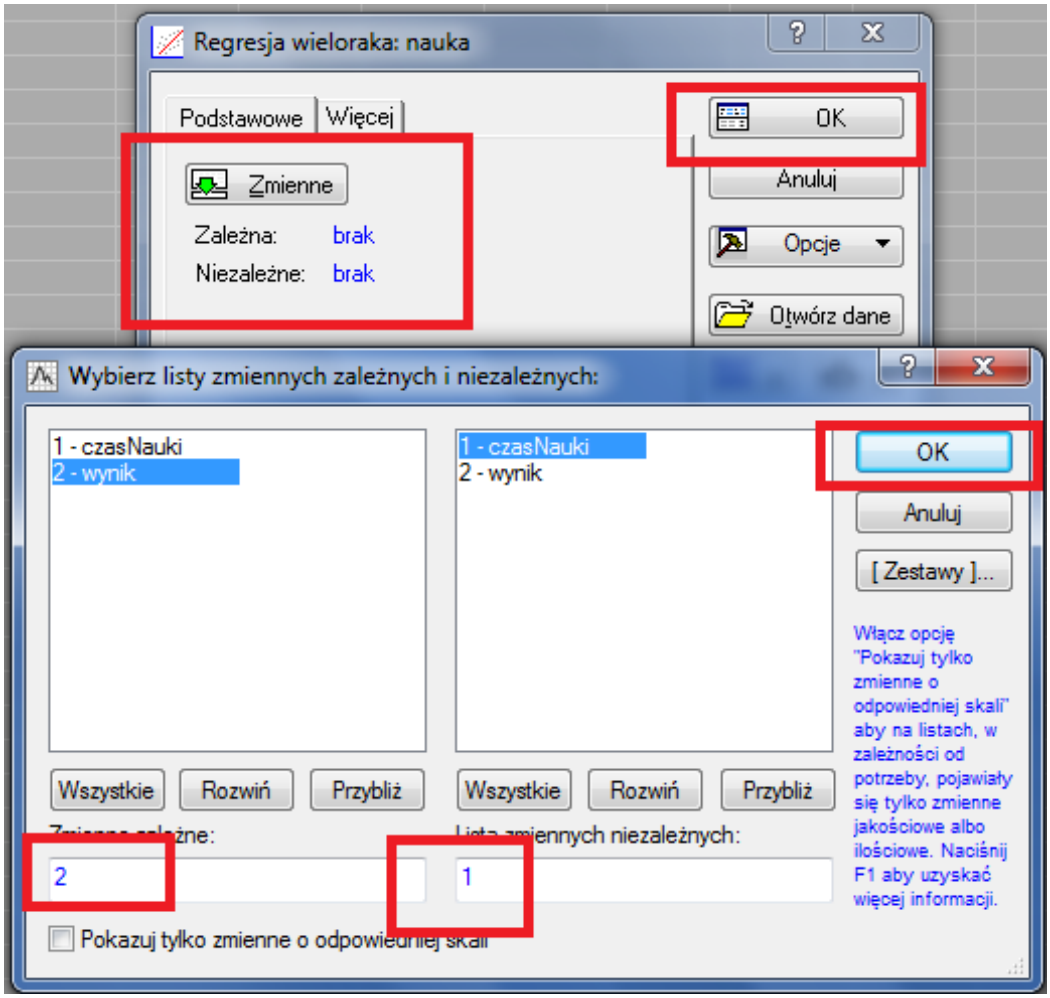
Za pomocą tej prostej możemy szacować wyniki znając czas poświęcony na naukę i dla przykładu osoba ucząca się 50 godzin powinna otrzymać około 150 punktów (na 100 możliwych ;)). Zauważmy jeszcze, że w programie mamy kilka możliwych sposobów wyznaczania równania (1). Mamy również możliwość wyboru z kilku dostępnych statystyk oraz wykresów (2).



Teraz wyznaczymy równanie prostej regresji za pomocą programu Statistica. Wybieramy z menu statystyka regresję wieloraką



Następnie wskazujemy zmienne

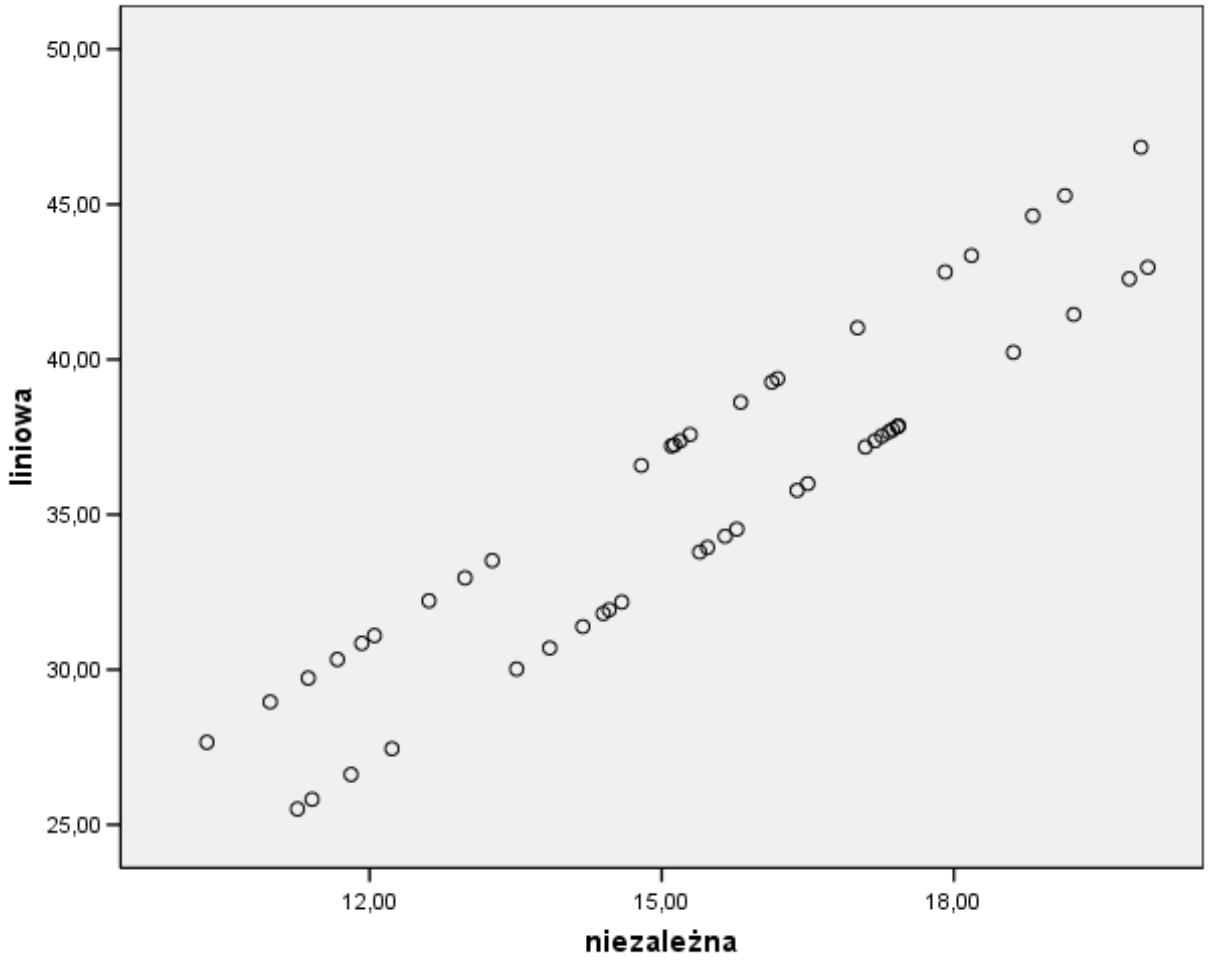


i jako wynik otrzymujemy skoroszyt, w którym mamy interesujący nas wynik.

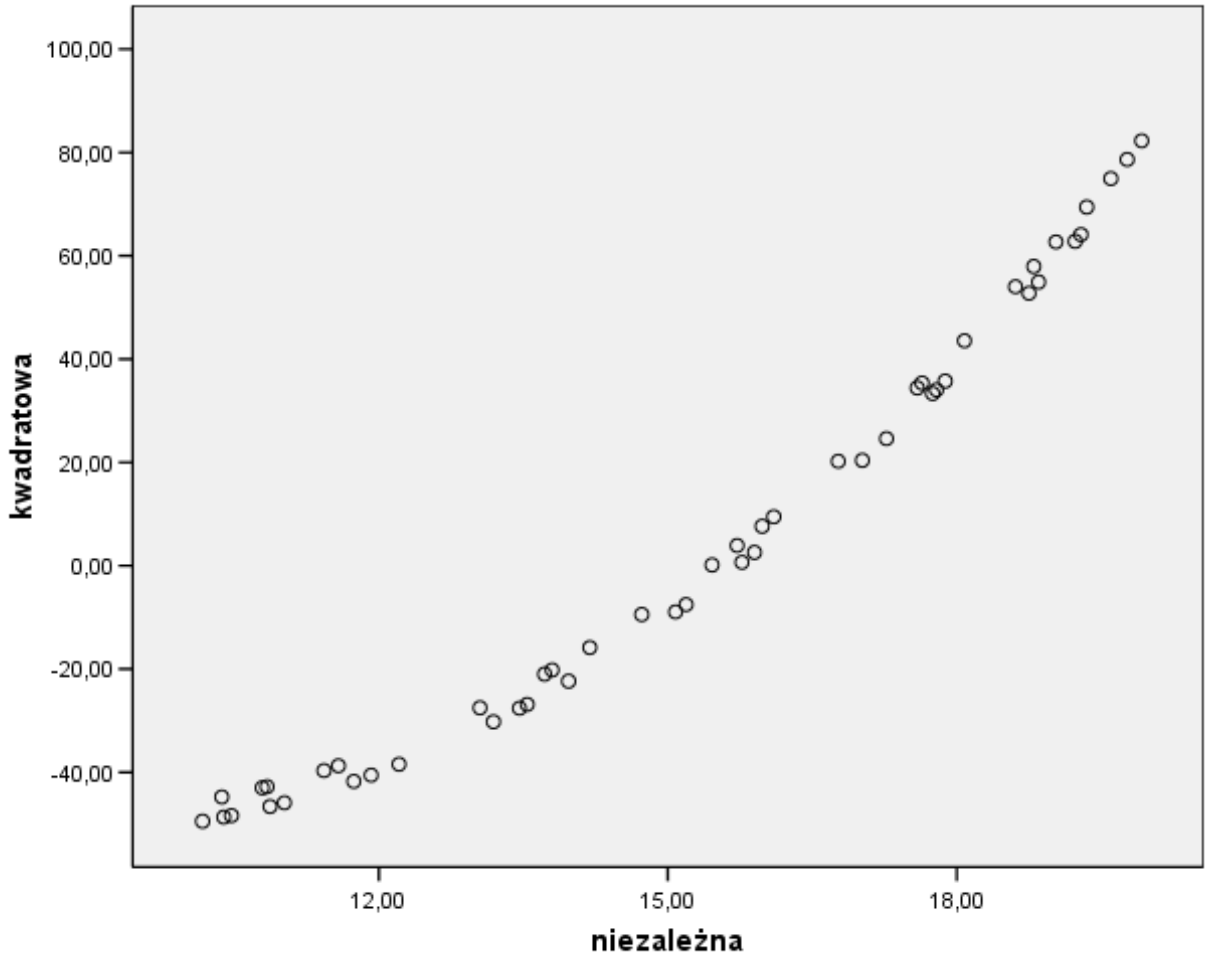
Podsumowanie regresji zmiennej zależnej: wynik (nauka)
R= ,99206004 R^2= ,98418313 Skoryg. R2= ,98385361
F(1,48)=2986,7 p<0,0000 Błąd std. estymacji: 1,0535

N=50	b*	Bł. std. z b*	b	Bł. std. z b	t(48)	p
W. wolny			-0,152589	1,391552	-0,10965	0,913141
czasNauki	0,992060	0,018153	3,009528	0,055068	54,65103	0,000000

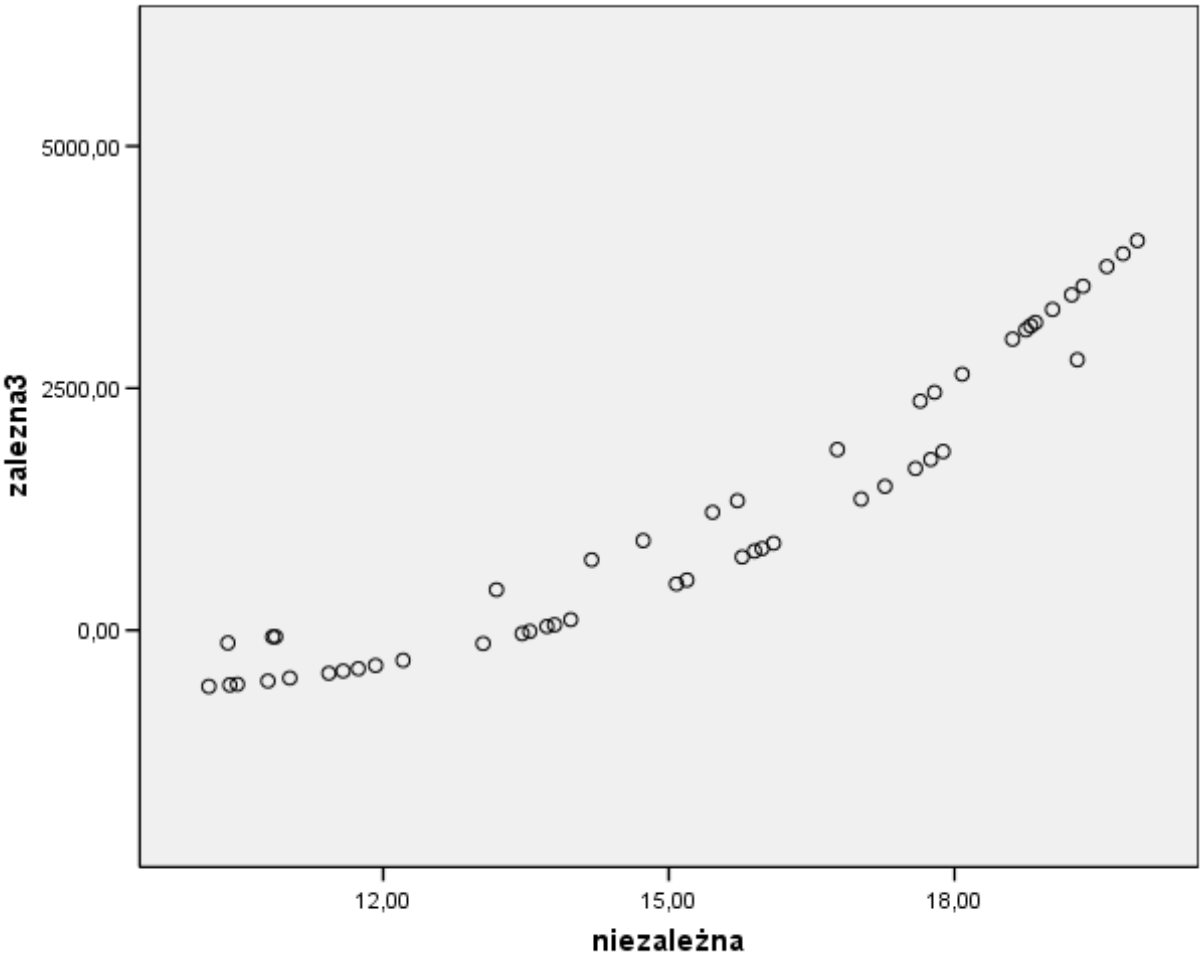
Oczywiście w statystyce można rozważać inne, bardziej złożone typy regresji. W klasyczny sposób (na kartce) wyznaczanie innych krzywych regresji jest dość trudne i pracochłonne. Na szczęście za pomocą programów statystycznych jest stosunkowo proste. Musimy jedynie zasugerować jakiego rodzaju krzywej regresji się spodziewamy. Przy wyborze mogą nam pomóc wykresy rozrzutu. W pliku regresja.sav mamy kilka przykładowych zestawów danych. Pierwsza zmienna jest zmienną niezależną, jak łatwo odczytać z poniższego wykresu rozrzutu druga ze zmiennych zależy od pierwszej w sposób liniowy



Trzecia zmienna jest związana z pierwszą za pomocą trójmianu kwadratowego (chociaż może tego nie widać na rysunku), czyli jak ktoś woli funkcji kwadratowej.

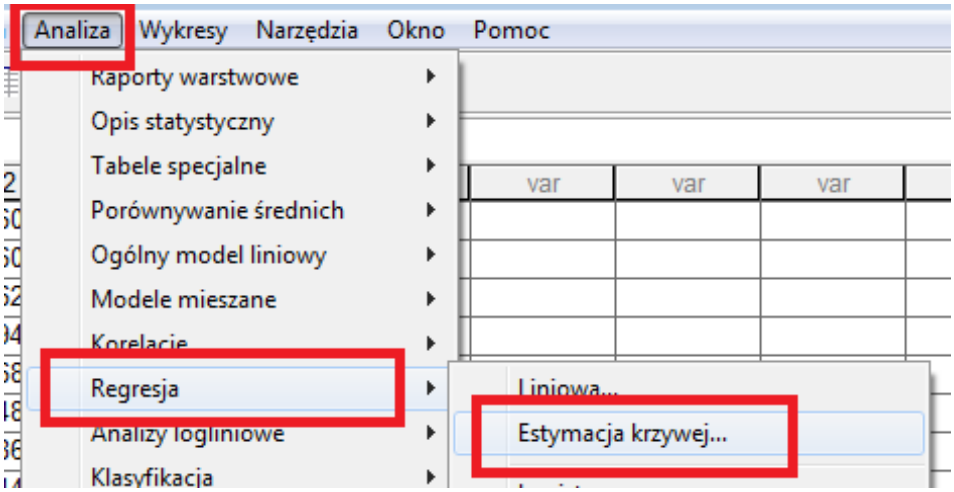


Trzecia ze zmiennych zależnych (u nas jest to zmienna o nazwie zależna3) jest związana wielomianem stopnia trzeciego. Wykres rozrzutu przyjmuje następującą postać.

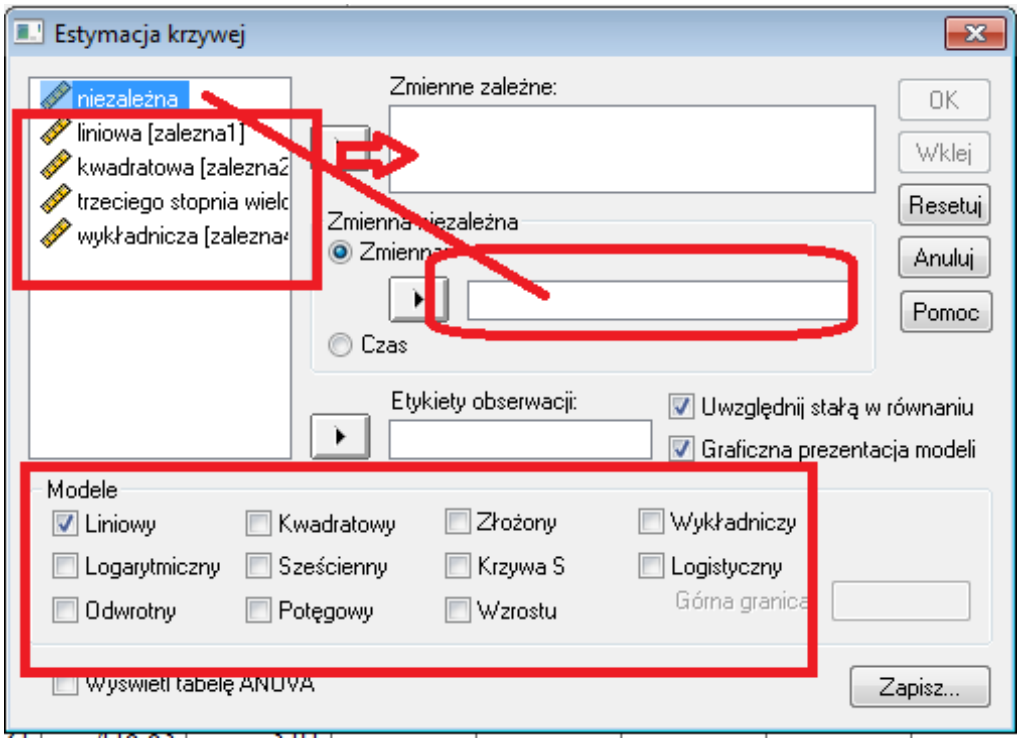


Ostatnia zaś zmienna zależna jest związana za pomocą funkcji wykładniczej. Postaramy się wyznaczyć poszczególne funkcje opisujące nasze zależności.

W programie SPSS skorzystamy z modułu estymacja krzywej...



gdzie musimy wybrać zmienne oraz rodzaj krzywej regresji



Dla pierwszej zmiennej zależnej zgonie z sugestią wykresu rozrzutu wyznaczamy prostą regresji. Otrzymujemy raport w którym mamy wyznaczoną in-

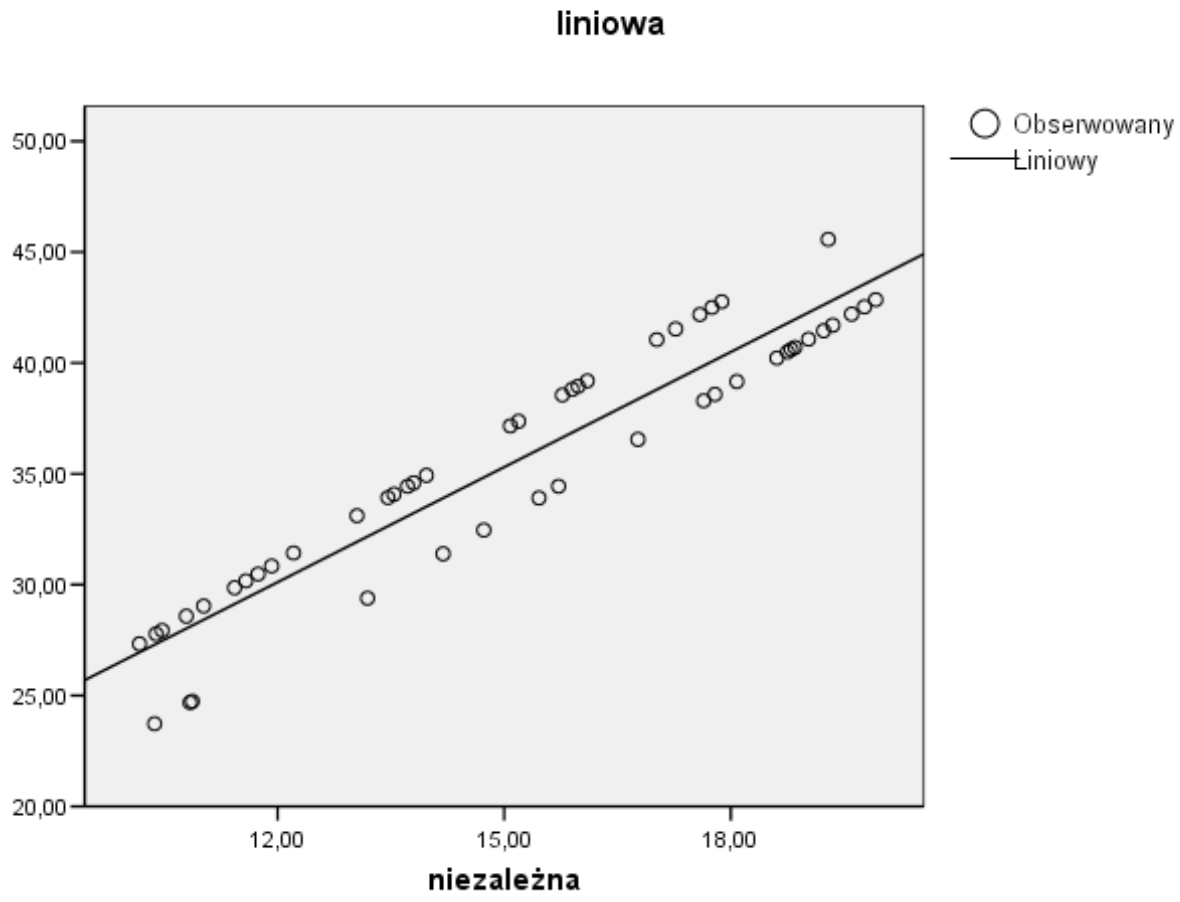
teresującą nas prostą oraz dodatkowo poprzez wybranie stosownego pola mamy na wykres rozrzutu naniesioną naszą prostą.

Podsumowanie modelu i oszacowań parametrów

Zmienna zależna: liniowa

	Podsumowanie modelu					Oszacowania parametrów	
Równanie	R-kwadrat	F	df1	df2	Istotność	Stała	b1
Liniowy	,899	425,710	1	48	,000	9,359	1,730

Zmienną niezależną jest niezależna.



W tym miejscu pojawia się naturalne pytanie z jaką dokładnością nasza wyznaczona prosta pokrywa się ze stanem faktycznym. Ponieważ dane do przykładu zostały dobrane w taki sposób, że zmienna zależna wyraża się za pomocą wzoru

$$y = 2x + 5 + \delta$$

gdzie δ przyjmuje wartości ± 1 z prawdopodobieństwem $\frac{1}{2}$. Natomiast SPSS wyznaczył następujący wzór

$$y = 1.73x + 9.359$$

Może nie jest to idealne przybliżenie, jest ono spowodowane stosunkowo małą liczbą obserwacji. Dla 500 obserwacji nasza estymowana krzywa przyjmuje postać.

Podsumowanie modelu i oszacowań parametrów

Zmienna zależna: zalezna

	Podsumowanie modelu					Oszacowania parametrów	
Równanie	R-kwadrat	F	df1	df2	Istotność	Stała	b1
Liniowy	,896	4313,028	1	498	,000	4,472	2,044

Zmienną niezależną jest niezalezna.

i widzimy, że dopasowanie jest już znacznie lepsze.

Dla zmiennej zależnej2 otrzymujemy następujący wynik

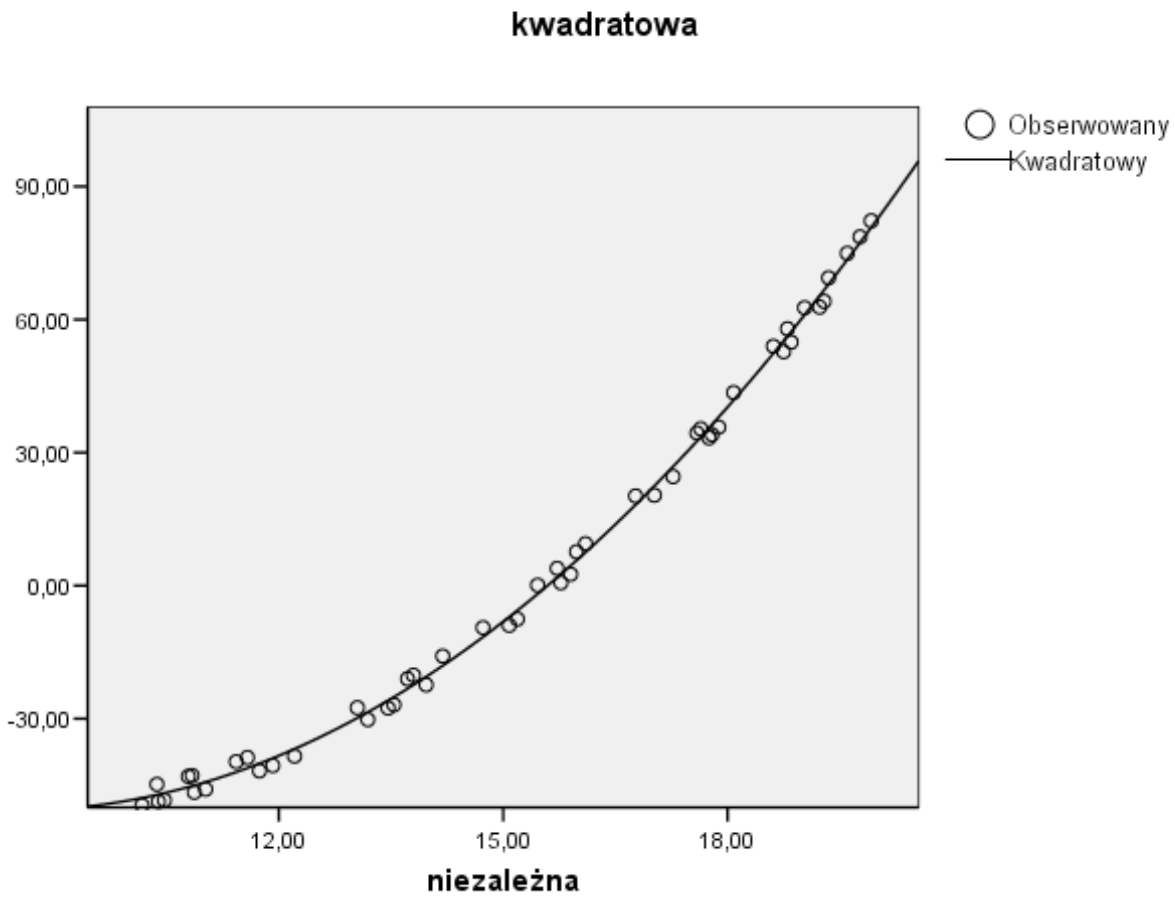
Podsumowanie modelu i oszacowań parametrów

Zmienna zależna: kwadratowa

	Podsumowanie modelu					Oszacowania parametrów		
Równanie	R-kwadrat	F	df1	df2	Istotność	Stała	b1	b2
Kwadratowy	,998	10267,690	2	47	,000	22,257	-17,149	1,008

Zmienną niezależną jest niezależna.

oraz stosowną krzywą



W rzeczywistości wzór ma następującą postać

$$y = x^2 - 17x + 22 + \delta.$$

Dla trzeciej zmienne prawdziwy związek zadany jest za pomocą funkcji

$$y = x^3 - 2x^2 - 5x - 2 + \delta$$

program natomiast proponuje nam następujący wzór oraz jego reprezentację

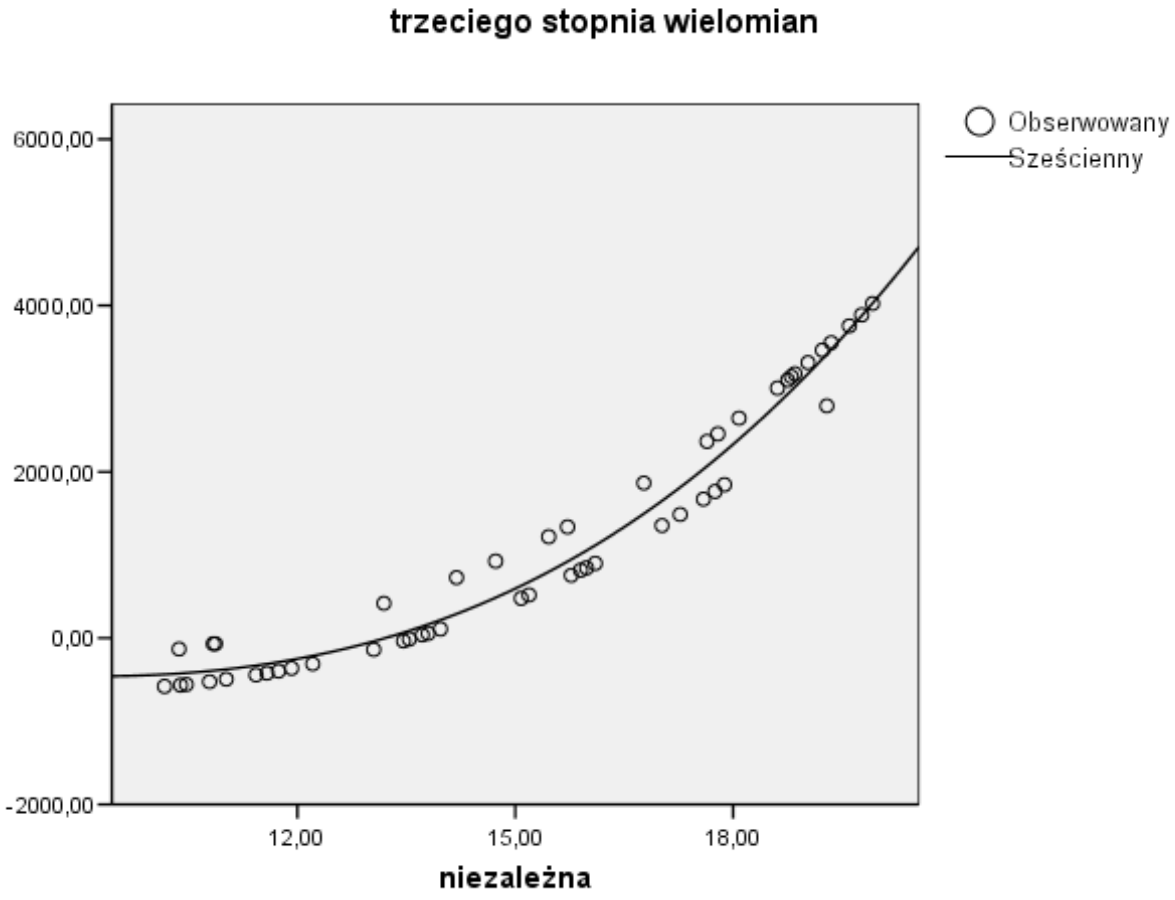
graficzną.

Podsumowanie modelu i oszacowań parametrów

Zmienna zależna: trzeciego stopnia wielomian

Równanie	Podsumowanie modelu					Oszacowania parametrów			
	R-kwadrat	F	df1	df2	Istotność	Stała	b1	b2	b3
Sześcienny	,970	750,114	2	47	,000	149,042	,000	-21,754	1,583

Zmienną niezależną jest niezależna.



Łatwo spostrzec, że sugerowany wzór funkcji nie jest zbytnio zbliżony do stanu faktycznego, jest to spowodowane tym, że nasze wartości nie w pobliżu punktów charakterystycznych jakimi są wierzchołki. W ostatnim przypadku funkcja

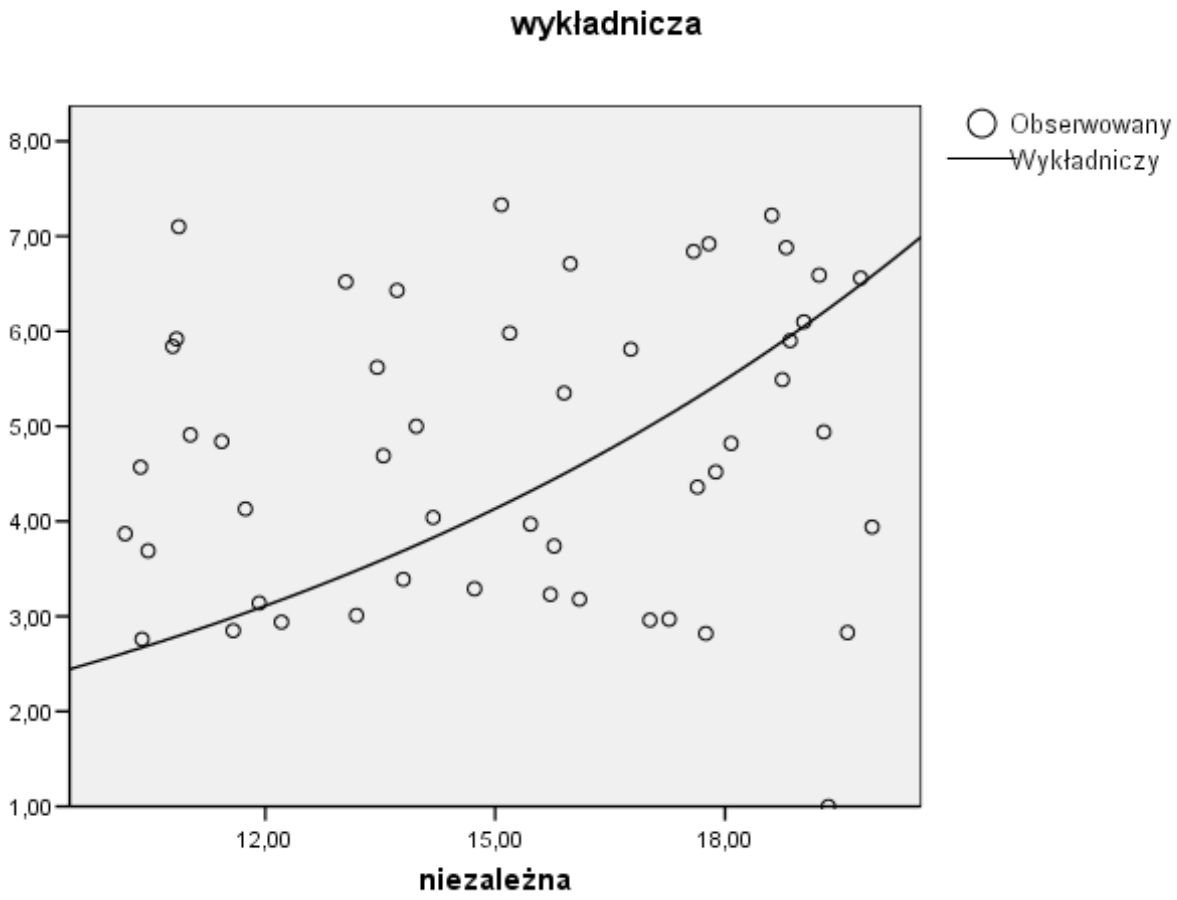
zadane jest wzorem

$$y = \frac{e^x}{100000}$$

natomiast program szacuje nam krzywą w następujący sposób

$$y = e^{0.095x}$$

której reprezentacja graficzna przyjmuje postać



W czasie ćwiczeń prześledzimy również inne przykładowe dane, w tym również inne rodzaje krzywych regresji oraz ich kombinacje.