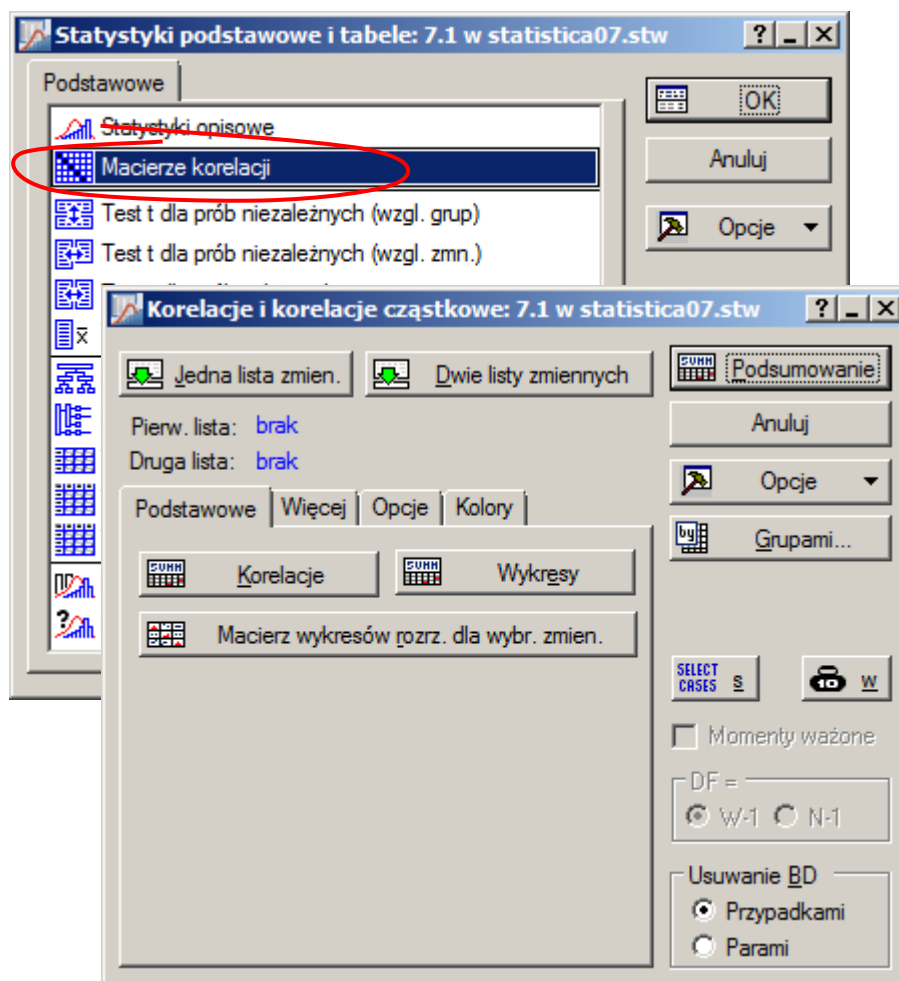


8. ANALIZA KORELACJI I REGRESJI

8.1. Analiza korelacji

Analizę korelacji można przeprowadzić w programie z poziomu okna **Korelacje i korelacje cząstkowe** (dostępne z menu **Statystyka/Statystyki podstawowe** po wskazaniu opcji **Macierz korelacji**).



Sposób przeprowadzania analizy zostanie omówiony w oparciu o przykłady 1. i 2. z części teoretycznej.

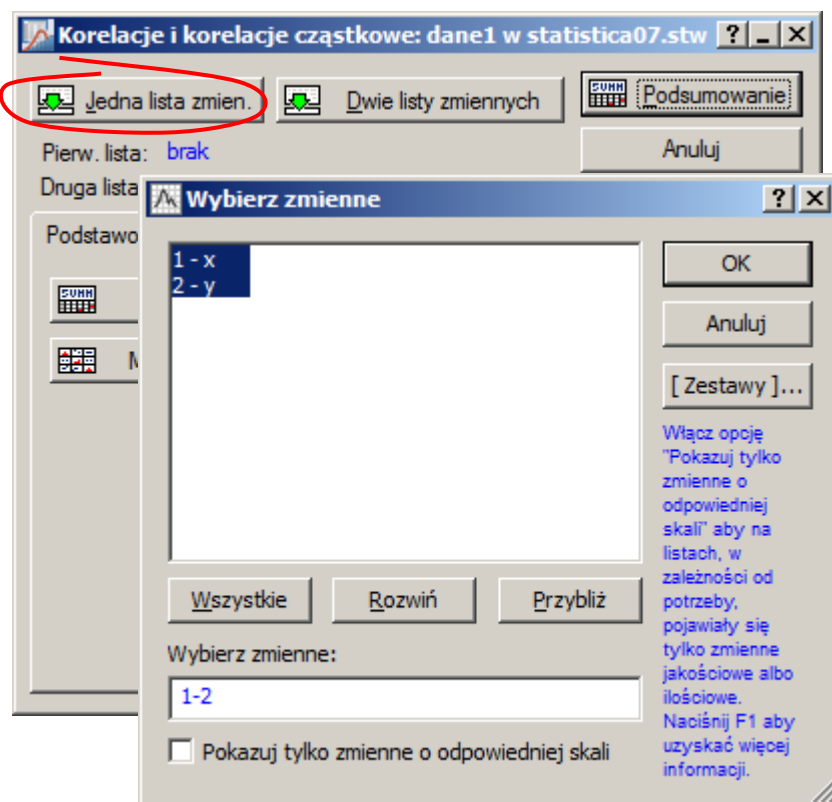
Przykład 1.i 2.

Arkusz *dane1* zawiera wyniki pomiarów dwóch wymiarów pewnego detalu otrzymane z 20-elementowej próby.

	1	2
	x	y
1	3,3	15,1
2	2,4	38
3	3,2	26
4	10,5	31,6
5	5,1	45,6

W oknie **Korelacje i korelacje cząstkowe** przede wszystkim należy wskazać zmienne dla których przeprowadzona zostanie analiza. Zmienne te mogą być wskazane na dwa sposoby: poprzez jedną lub dwie listy. Wybór pojedynczej listy zmiennych pozwala na obliczenie korelacji pomiędzy wszystkimi zmiennymi tej listy – w wyniku otrzymywana jest macierz korelacji. Wybór zmiennych na dwóch listach prowadzi do wyznaczenia współczynników korelacji pomiędzy zmiennymi listy pierwszej a zmiennymi listy drugiej, opcja ta pozwala więc na wyznaczenie macierzy korelacji w przypadku gdy w obydwu listach zostaną wybrane te same zmienne.

W przykładzie 1. w części teoretycznej obliczona została macierz korelacji dla zmiennych x i y . Macierz tą można otrzymać w programie wskazując zmienne np. poprzez jedną listę zmiennych.

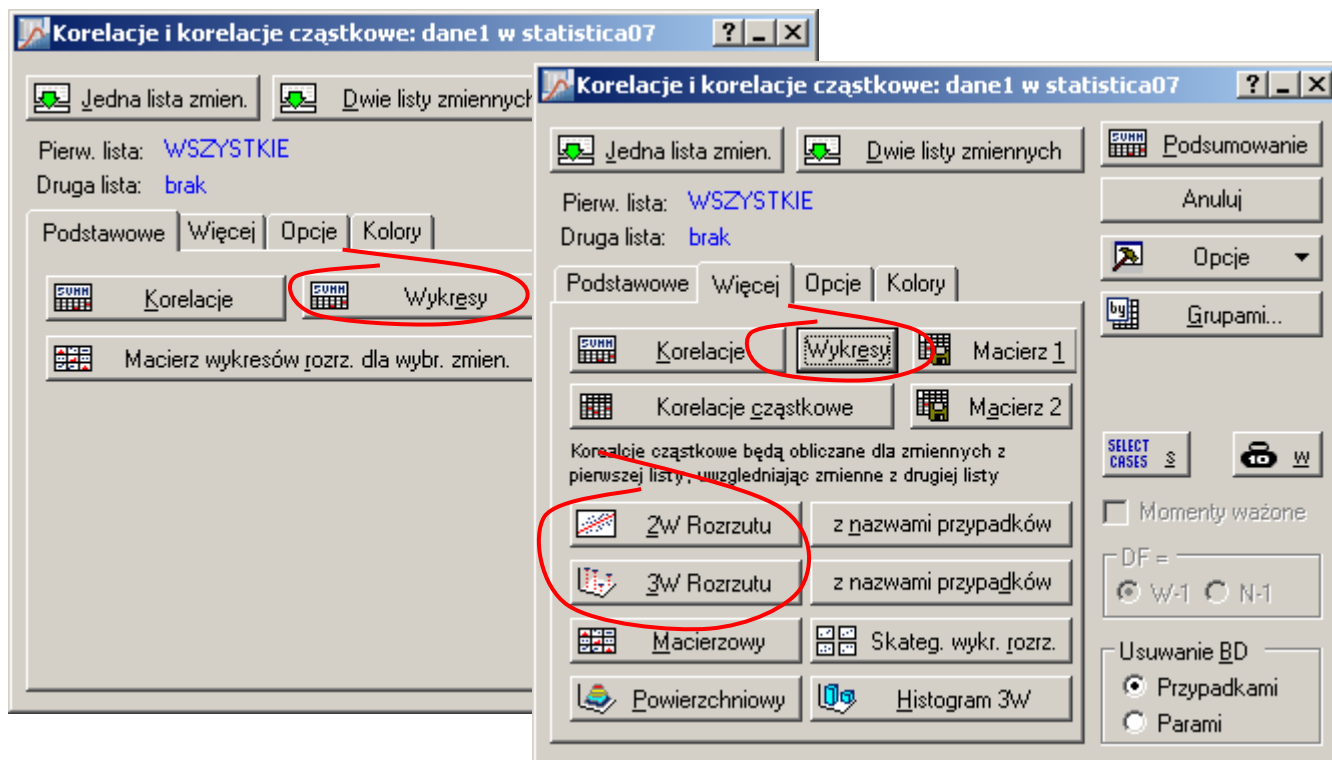


Wyniki analizy dla zmiennych wskazanych w liście są wyświetlane po naciśnięciu przycisku **Podsumowanie** lub **Korelacje** na zakładce **Podstawowe** lub **Więcej**. Zawartość wynikowego arkusza zależy od ustawień zakładki **Opcje**, domyślnie wyświetlane są: średnia i analizowanych zmiennych oraz macierz korelacji.

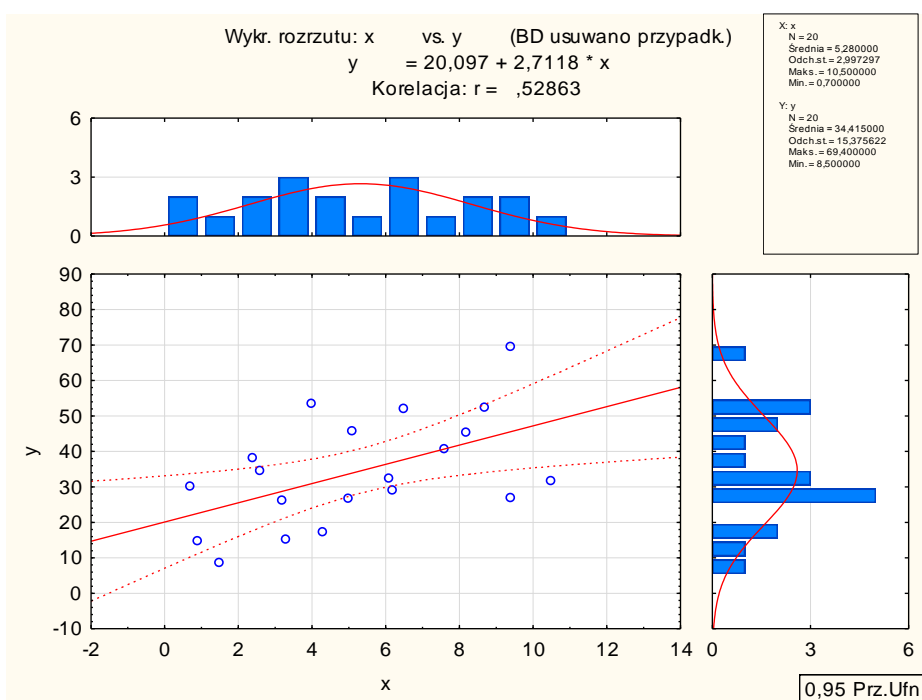
Zmienna	Średnia	Odch.std	x	y
x	5,28000	2,99730	1,000000	0,528626
y	34,41500	15,37562	0,528626	1,000000

Korelacje (dane1 w statistica07.stw)
 Oznaczone wsp. korelacji są istotne z $p < ,05000$
 N=20 (Braki danych usuwano przypadkami)

Otrzymana wartość współczynnika korelacji $\rho = 0,528626$ dla zmiennych x i y wskazuje, że zmienne te nie są zbyt mocno skorelowane. Wniosek ten można potwierdzić rysując wykres rozrzutu. Wykresy dla potrzeb przeprowadzanej analizy można wygenerować przy pomocy przycisków zaznaczonych na poniższych rysunkach.

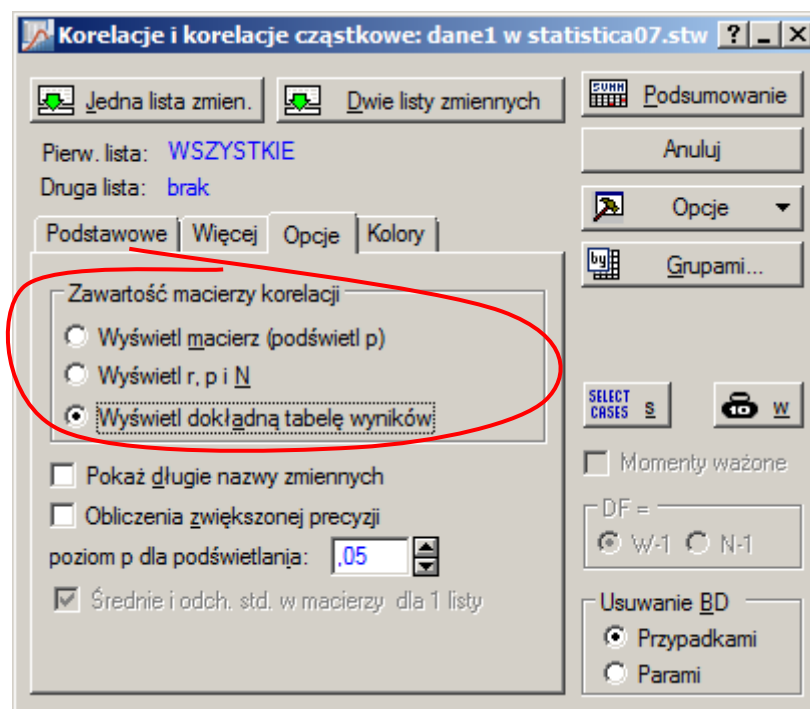


Przyciski **Wykresy** generują wykresy rozrzutu dla wszystkich par zmiennych wskazanych wcześniej w pojedynczej liście lub dwóch listach zmiennych. Przycisk **2WRozrzutu** generuje tylko jeden wykres dla wybranych dwóch zmiennych. Na wykresach rozrzutu oprócz punktów odpowiadających zmiennym nanoszona jest dodatkowo linia przedstawiająca najlepiej dopasowaną do danych zależność liniową wraz z tzw. krzywymi ufności. Dodatkowo, na wykresach wykreślonych przy pomocy przycisków **Wykresy**



kreślone są histogramy przedstawiające rozkład analizowanych zmiennych.

Wygenerowany przez program arkusz analizy, w zależności od ustawień na zakładce **Opcje**, może zawierać dodatkowe informacje.



Po wskazaniu opcji **Wyświetl dokładną tabelę wyników** wyświetlane są dla wszystkich par zmiennych:

- $r(X, Y)$ – współczynnik korelacji liniowej Paersona,
- r^2 – kwadrat współczynnika korelacji,
- t – wartość statystyki testowej wykorzystywanej w teście istotności współczynnika korelacji,
- p – wartość granicznego poziomu istotności p -value otrzymana z testu hipotezy o braku istotności korelacji,
- Ważnych – liczba poprawnych wartości zmiennych,
- Stała zal: Y, Nachylenie zal: Y, Stała zal: X, Nachylenie zal: X – parametry zależności liniowej zmiennych (patrz wykres rozrzutu na poprzedniej stronie).

Zmn. X & Zmn. Y	Średnia	Odch.st.	$r(X,Y)$	r^2	t	p	Ważnych	Stała zal: Y	Nachyle zal: Y
x	5,28000	2,99730							
x	5,28000	2,99730	1,000000	1,000000			20	0,00000	1,000000
x	5,28000	2,99730							
y	34,41500	15,37562	0,528626	0,279446	2,642113	0,016563	20	20,09690	2,711762
y	34,41500	15,37562							
x	5,28000	2,99730	0,528626	0,279446	2,642113	0,016563	20	1,73355	0,103049
y	34,41500	15,37562							
y	34,41500	15,37562	1,000000	1,000000			20	0,00000	1,000000

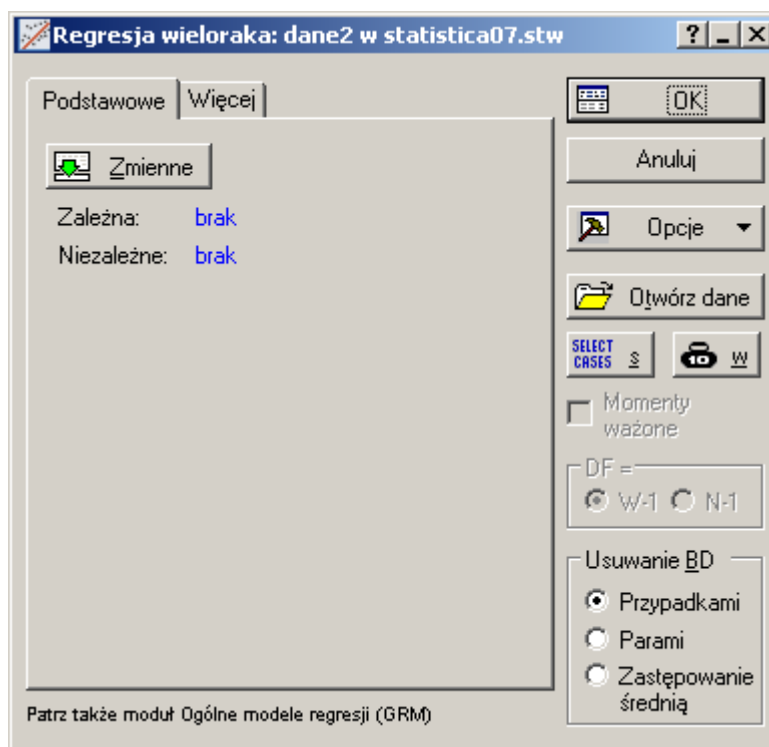
Z analizy powyższej tabeli wynika, że przyjmując (tak jak w części teoretycznej) poziom istotności testu $\alpha = 0,05$, hipotezę o braku istotności korelacji należy odrzucić ($\alpha > p\text{-value}$) – korelację zmiennych $\rho = 0,528626$ należy więc uznać za istotną.

8.2. Analiza regresji

Analiza regresji jest dostępna w programie w kilku różnych miejscach. W materiale omówione zostanie okno **Regresja wieloraka** (dostępne z menu **Statystyka/Regresja wieloraka**) oraz Estymacja nieliniowa (dostępne z menu **Statystyka/Modele zaawansowane**). Sposób przeprowadzania analizy zostanie omówiony w oparciu o przykłady 3.–7. z części teoretycznej.

8.2.1. Okno Regresja wieloraka

W oknie regresji należy przede wszystkim wskazać zmienne dla których przeprowadzona zostanie analiza.



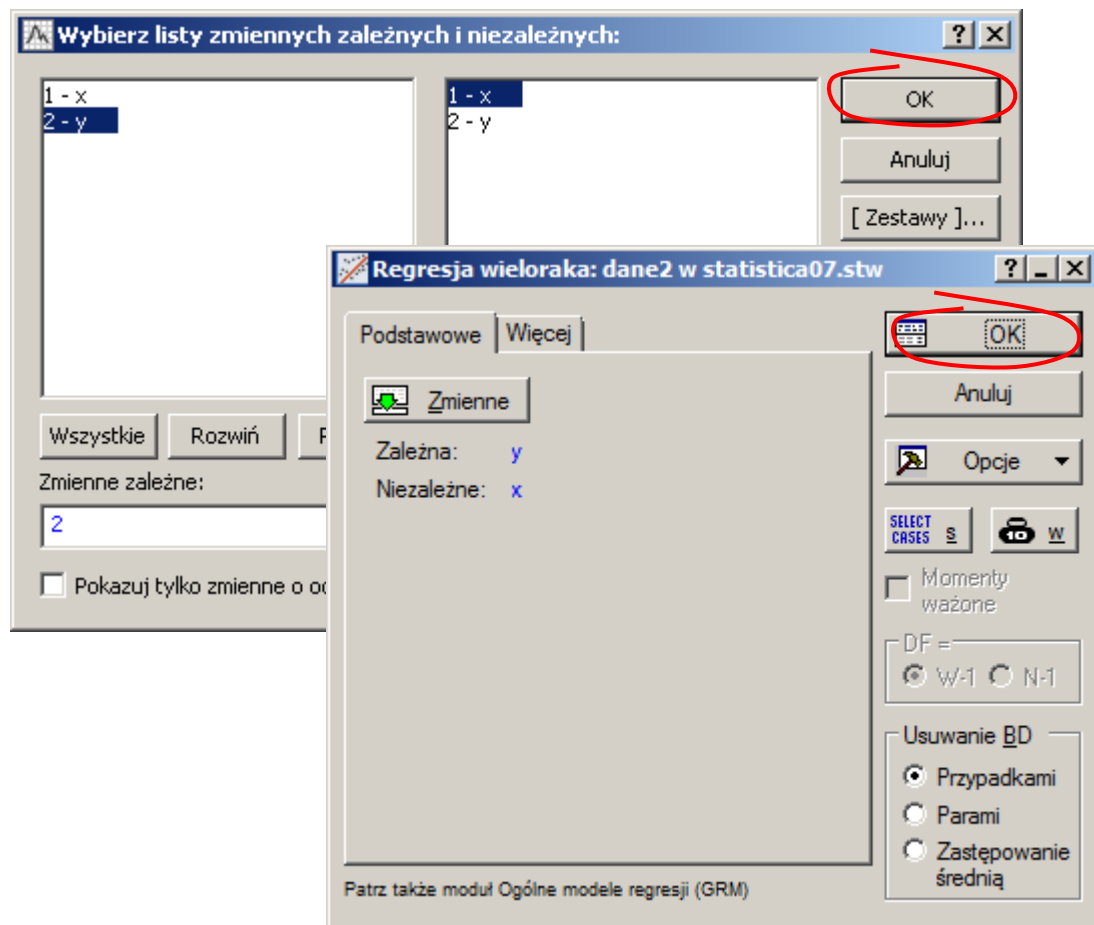
Wybierając zmienne należy wskazać zmienne zależne i niezależne. Na początek zostanie omówiony najprostszy przykład z części teoretycznej – z jedną zmienną niezależną i jedną zmienną zależną.

Przykład 3.

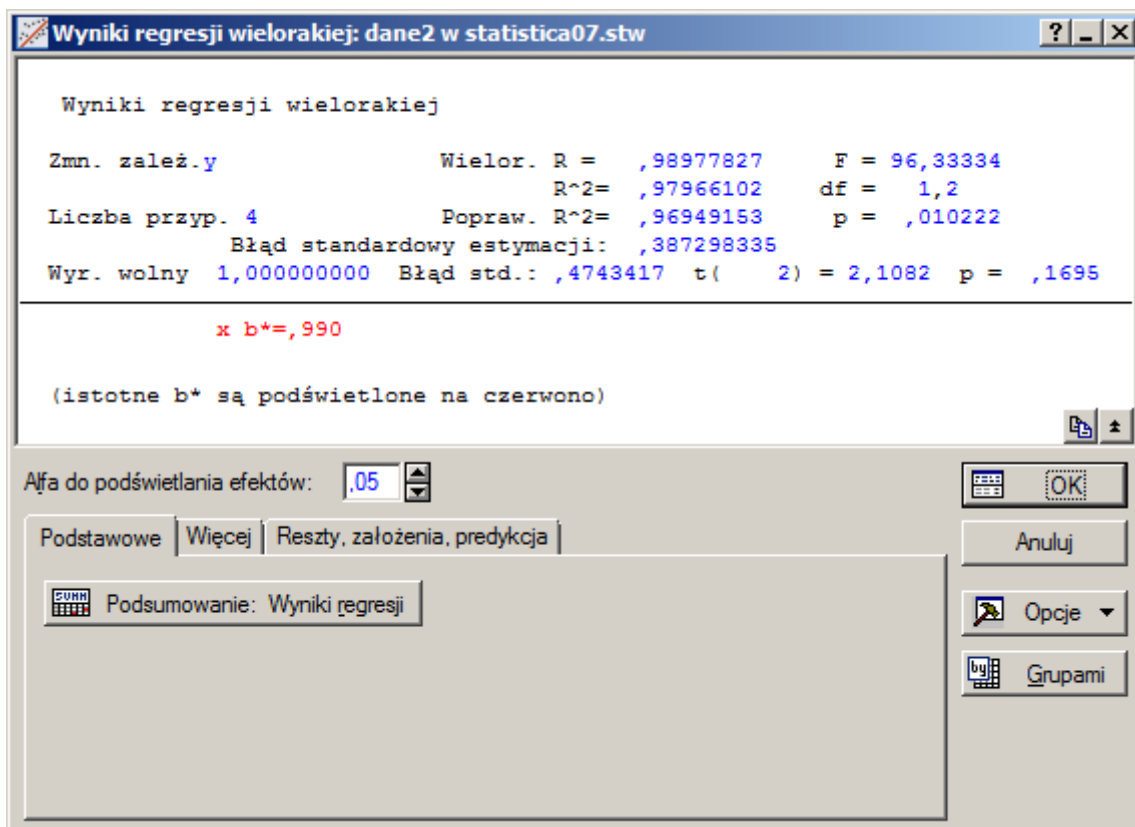
Arkusz *dane2* zawiera wyniki zebrane podczas badania zależności kosztów produkcji y od ilości produkowanych sztuk x .

	1 x	2 y
1	1	3
2	2	4
3	3	6
4	4	8

Zmienną niezależną jest w tym przypadku zmienna x a zmienną zależną zmienna y .



Po wskazaniu zmiennych i zaakceptowaniu okna regresji przyciskiem OK wyświetlane są wyniki analizy.



Górne pole okna wyników zawiera:

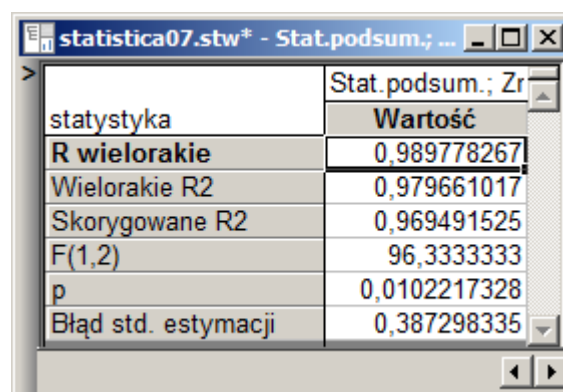
- Wielo. R** – wartość współczynnika korelacji (pierwiastek ze współczynnika determinacji R^2),
- R²** – wartość współczynnika determinacji R^2 ,
- Popraw. R²** – wartość skorygowanego współczynnika determinacji \bar{R}^2 ,
- Błąd standardowy estymacji** – wartość błędu standardowego estymacji s_e ,
- F** – wartość statystyki F wykorzystywanej do oceny istotności funkcji regresji,
- df** – ilość stopni swobody statystyki F ,
- p** – wartość granicznego poziomu istotności p -value testu dla hipotezy zerowej o braku wpływu zmiennych niezależnych na zmienną zależną (hipoteza jest odrzucana jeśli graniczny poziom istotności jest mniejszy od założonego α),

Jeśli równanie regresji zawiera stałą w polu wyświetlane są dodatkowo:

- Wyr. wolny** – wartość stałej,
- błąd std.** – błąd oceny stałej (patrz część teoretyczna: *Istotność współczynników funkcji regresji*: s_i estymator odchylenia standardowego parametru β_i),
- t(df)** – wartość statystyki t wykorzystywanej do oceny istotności współczynnika funkcji regresji (w tym przypadku do oceny istotności stałej),
- df** – ilość stopni swobody statystyki t ,
- p** – wartość granicznego poziomu istotności p -value testu dla hipotezy zerowej o braku istotności współczynnika,

Pod kreską w górnej części pola wyników wyświetlane są standaryzowane wartości współczynników równania regresji. Współczynniki takie otrzymano by w sytuacji gdyby przed wykonaniem analizy regresji wszystkie zmienne zostały przekształcone na zmienne standaryzowane (o średniej zero i odchyleniu standardowym 1). Standaryzacja współczynników umożliwia porównanie wpływu zmiennych niezależnych na zmienną zależną (im większa wartość współczynnika standaryzowanego tym większy wpływ związanej z nim zmiennej). Istotne współczynniki równania regresji wyróżniane są na czerwono (hipoteza o braku ich istotności – dla α wskazanego w oknie wyników – powinna zostać odrzucona).

Pełne wyniki analizy wyświetlane są w dwóch dodatkowych arkuszach po naciśnięciu przycisku **Podsumowanie: Wyniki regresji**. W pierwszym arkuszu znajdują się współczynniki oceniające jakość dopasowania regresji: współczynnik korelacji R , współczynniki determinacji R^2 i \bar{R}^2 i błąd standardowy estymacji s_e



statystyka	Stat. podsum.; Zr
R wielorakie	0,989778267
Wielorakie R2	0,979661017
Skorygowane R2	0,969491525
F(1,2)	96,3333333
p	0,0102217328
Błąd std. estymacji	0,387298335

oraz wyniki testu istotności funkcji regresji: wartość statystyki F wraz z otrzymaną wartością p -value.

W górnej części okna drugiego arkusza wyniki te są dodatkowo powtórzone a w samym arkuszu dla każdego składnika funkcji regresji wyświetlane są:

- b*** – wartość standaryzowanego współczynnika,
- błąd std. z b*** – błąd oceny standaryzowanego współczynnika,
- b** – wartość właściwego współczynnika,
- błąd std. z b** – błąd oceny współczynnika,
- t(df)** – wartość statystyki t wykorzystywanej do oceny istotności współczynnika,
- df** – ilość stopni swobody statystyki t ,
- p** – wartość granicznego poziomu istotności p -value testu dla hipotezy zerowej o braku istotności współczynnika,

Dodatkowo, wiersze odpowiadające istotnym (uwzględniając poziom istotności α wskazany w oknie wyników) składnikom funkcji regresji wyróżniane są na czerwono.

Podsumowanie regresji zmiennej zależnej: y (dane2 w statistica07.stw)
 R= ,98977827 R^2= ,97966102 Popraw. R2= ,96949153
 F(1,2)=96,333 p<,01022 Błąd std. estymacji: ,38730

N=4	b*	Bł. std. z b*	b	Bł. std. z b	t(2)	p
W. wolny			1,000000	0,474342	2,108185	0,169545
x	0,989778	0,100844	1,700000	0,173205	9,814955	0,010222

Z przedstawionych arkuszy analizy wynika, że znaleziona funkcja regresji opisująca zależność kosztów produkcji od ilości produkowanych sztuk x ma postać:

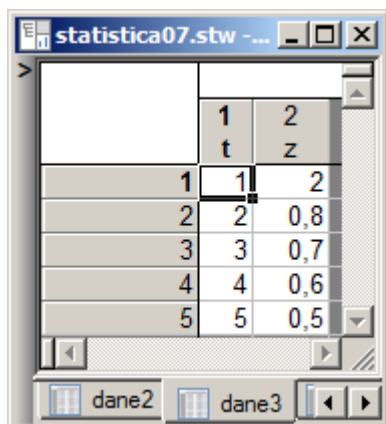
$$\hat{y} = 1 + 1,7x.$$

Dla domyślnego poziomu istotności $\alpha = 0,05$ funkcja ta jest funkcją istotną (hipotezę o jej braku istotności należy odrzucić ponieważ $\alpha = 0,05 > p\text{-value} = 0,01022$). Stała w równaniu regresji jest nieistotna ($p\text{-value} = 0,169545$) a współczynnik związany z ilością produkowanych sztuk jest istotny ($p\text{-value} = 0,010222$).

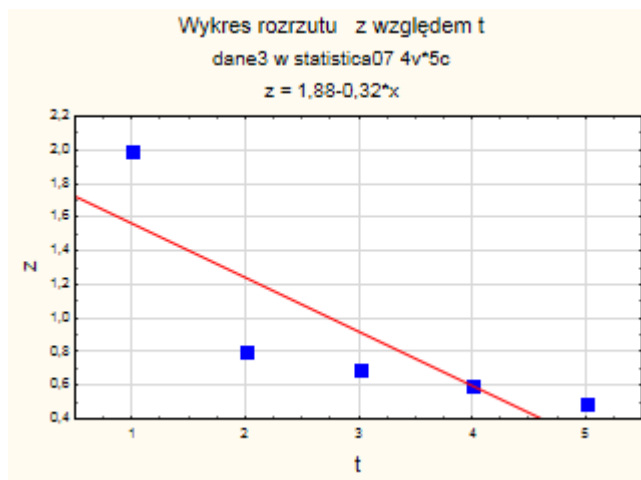
STATISTICA pozwala na dalszą analizę problemu udostępniając (po naciśnięciu przycisku OK w oknie wyników regresji) analizę reszt. Analiza ta umożliwi ocenę prawdziwości przyjmowanych założeń oraz pozwala na ocenę poprawności znalezionej funkcji regresji, temat ten nie zostanie jednak omówiony w tym materiale.

Przykład 4.

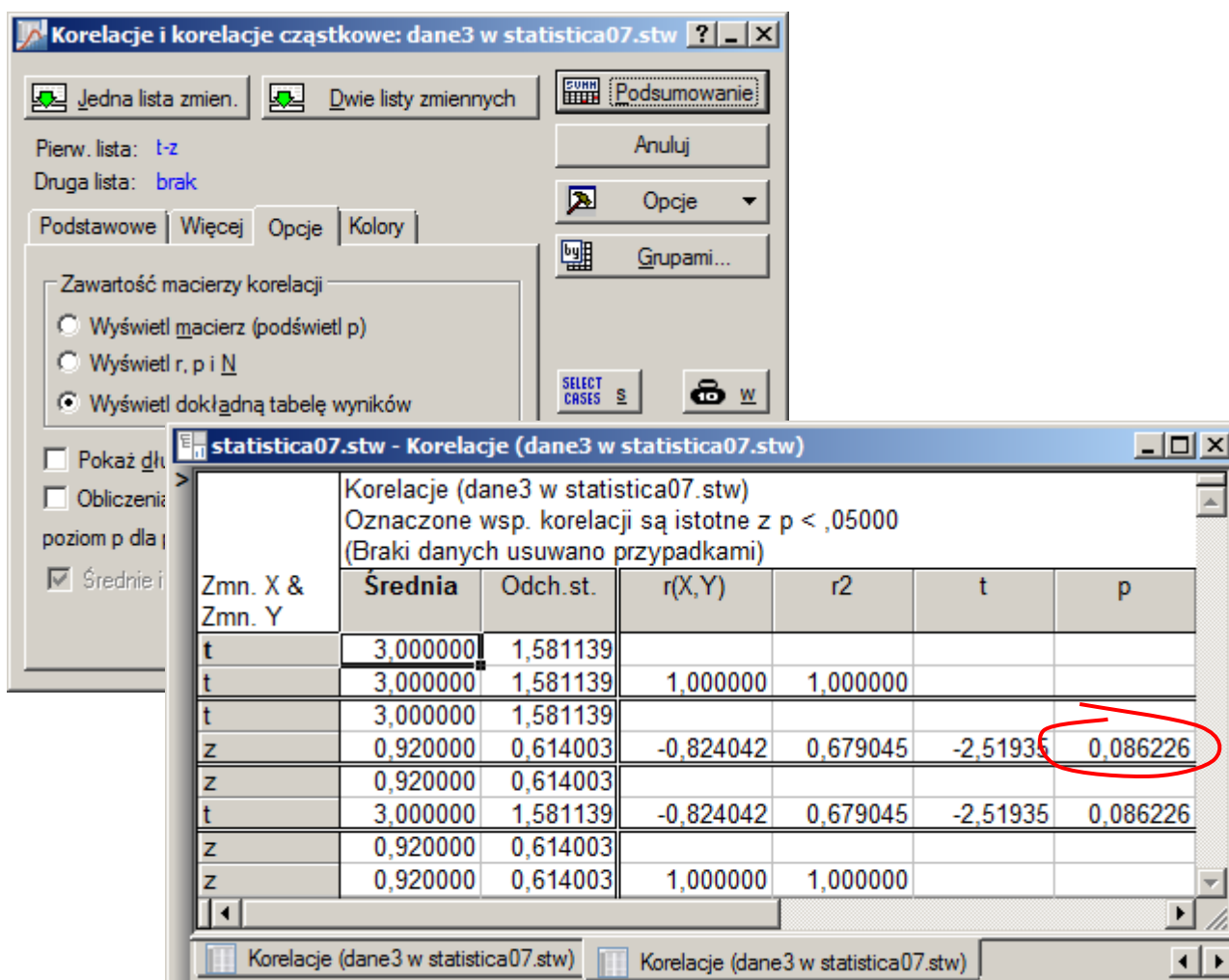
Arkusz *dane3* zawiera wyniki zebrane podczas badania zależności zmian parametru z od czasu t . W części teoretycznej pokazano, że zależność $z(t)$ nie jest zależnością liniową.



	1	2
t	1	2
z	0,8	0,7
	4	0,6
	5	0,5



Analiza korelacji wykazała, że nie można odrzucić hipotezy o braku korelacji liniowej parametrów z i t ($\alpha = 0,05 < p - value = 0,086226$).



Korelacje i korelacje cząstkowe: dane3 w statistica07.stw

Pierw. lista: t-z
Druga lista: brak

Zawartość macierzy korelacji:
 Wyświetl macierz (podświetl p)
 Wyświetl r, p i N
 Wyświetl dokładną tabelę wyników

statistica07.stw - Korelacje (dane3 w statistica07.stw)

Korelacje (dane3 w statistica07.stw)
Oznaczone wsp. korelacji są istotne z $p < ,05000$
(Braki danych usuwano przypadkami)

Zmn. X & Zmn. Y	Średnia	Odch.st.	r(X,Y)	r2	t	p
t	3,000000	1,581139				
t	3,000000	1,581139	1,000000	1,000000		
t	3,000000	1,581139				
z	0,920000	0,614003	-0,824042	0,679045	-2,51935	0,086226
z	0,920000	0,614003				
t	3,000000	1,581139	-0,824042	0,679045	-2,51935	0,086226
z	0,920000	0,614003				
z	0,920000	0,614003	1,000000	1,000000		

Zmienne z i t mogą być jednak przekształcone w taki sposób, żeby wiązała je zależność liniowa:

$$\ln(z) = b_0 + b_1 \frac{1}{t}.$$

W celu znalezienia parametrów funkcji regresji należy uzupełnić arkusz *dane3* o zmienne $\ln(z)$ i $1/t$.

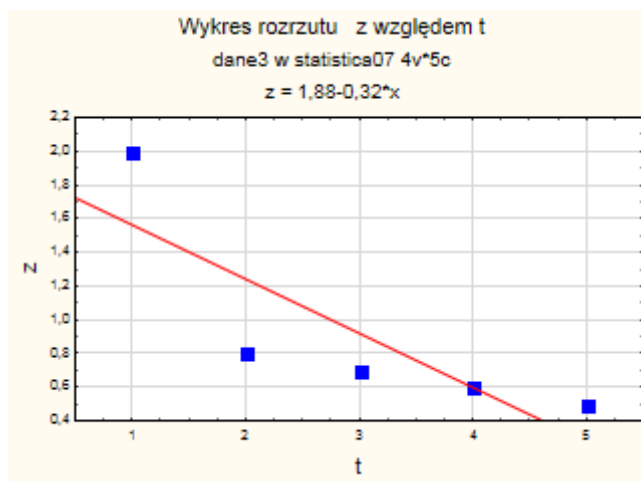
The screenshot shows the Statistica 7 interface. In the background, a spreadsheet window titled "statistica07.stw - dane3" displays a table with columns labeled 1, 2, 3, 4 and rows labeled t, z, 1/t, ln(z). The first row of data contains the values 1, 1, 2, and 0,69315.

Two dialog boxes are overlaid on the spreadsheet:

- Zmienna 3**: The "Nazwa" (Name) field contains "1/t". The "Długa nazwa (etykieta lub funkcja)" (Long name (label or formula)) field contains "=1/t".
- Zmienna 4**: The "Nazwa" field contains "ln(z)". The "Typ" (Type) is set to "Podw. precyzji" (Double precision). The "Długa nazwa (etykieta lub funkcja)" field contains "=log(z)".

Both dialog boxes have "Format wyświetlania" (Display format) sections with a list of options: Ogólny, Liczby, Data, Czas, Naukowy, Waluta, Procent, Ułamek, and Użytkownika. The "Przewodnik po funkcjach" (Use function separator) checkbox is checked in the Zmienna 4 dialog.

Wykres rozrzutu i test istotności korelacji potwierdzają trafność zaproponowanego podejścia.



statistica07.stw* - Korelacje (dane3 w statistica07.stw)

Korelacje (dane3 w statistica07.stw)
Oznaczone wsp. korelacji są istotne z $p < ,05000$
(Braki danych usuwano przypadkami)

Zmn. X & Zmn. Y	Średnia	Odch.st.	r(X,Y)	r ²	t	p
1/t	0,456667	0,324380				
1/t	0,456667	0,324380	1,000000	1,000000		
1/t	0,456667	0,324380				
ln(z)	-0,218129	0,538739	0,993398	0,986839	14,99830	0,000643
ln(z)	-0,218129	0,538739				
1/t	0,456667	0,324380	0,993398	0,986839	14,99830	0,000643
ln(z)	-0,218129	0,538739				
ln(z)	-0,218129	0,538739	1,000000	1,000000		

Korelacje (dane3 w statistica07.stw)

Po wprowadzeniu nowych zmiennych można już przeprowadzić analizę regresji.

Regresja wieloraka: dane3 w statistica07.stw

Podstawowe Więcej

Zmienne

Zależna: brak
Niezarne: brak

Anuluj

Wybierz listy zmiennych zależnych i niezależnych:

1 - t
2 - z
3 - 1/t
4 - ln(z)

1 - t
2 - z
3 - 1/t
4 - ln(z)

OK
Anuluj
[Zestawy]...

Włącz opcję "Pokazuj tylko zmienne o odpowiedniej skali"

Wyniki regresji wielorakiej: dane3 w statistica07.stw

Wyniki regresji wielorakiej

Zmn. zależ. ln(z) Wielor. R = ,99339779 F = 224,9491
R² = ,98683917 df = 1,3
Liczba przyp. 5 Popraw. R² = ,98245222 p = ,000643
Błąd standardowy estymacji: ,071365661
Wyr. wolny -,971565961 Błąd std.: ,0595160 t(3) = -16,32 p = ,0005

1/t b* = ,993
(istotne b* są podświetlone na czerwono)

Alfa do podświetlania efektów: ,05

Podstawowe Więcej Reszty, założenia, predykcja

Podsumowanie: Wyniki regresji

OK
Anuluj
Opcje
Grupami



statistica07.stw* - Podsumowanie regresji zmiennej zależnej: ln(z) (dane3 w statisti... - [] [] [X]

Podsumowanie regresji zmiennej zależnej: ln(z) (dane3 w statistica07.stw)
 R= ,99339779 R^2= ,98683917 Popraw. R2= ,98245222
 F(1,3)=224,95 p<,00064 Błąd std. estymacji: ,07137

	b*	Bł. std. z b*	b	Bł. std. z b	t(3)	p
N=5						
W. wolny			-0,971566	0,059516	-16,3245	0,000500
1/t	0,993398	0,066234	1,649862	0,110003	14,9983	0,000643

Stat.podsum.: Zmn. zal. ln(z) (dane3 w statistica07.stw) Podsumowanie regresji zmiennej zależnej: [] [] [] []

Wyniki analizy wskazują że znaleziona funkcja:

$$\ln(z) = -0,971566 + 1,649862 \frac{1}{t}$$

jest funkcją istotną (p -value = 0,000643), istotne są też obydwa współczynniki funkcji (p -value = 0,0005 dla wyrazu wolnego i p -value = 0,000643 dla współczynnika związanego z czasem).

Ostateczną postać funkcji otrzymuje się eliminując logarytm:

$$z = e^{-0,971566} e^{1,649862 \frac{1}{t}} = 0,37849 e^{1,649862 \frac{1}{t}}$$

Przykład 5.

Arkusz *dane4* zawiera wyniki zebrane podczas badania wpływu dwóch parametrów procesu obróbki x_1 i x_2 na wysokość nierówności y . W części teoretycznej założono, że zależność $y(x_1, x_2)$ ma postać:

$$\hat{y}(x) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$$

dane w arkuszu musiały więc zostać uzupełnione o dodatkową zmienną $x_1 x_2$ zawierającą iloczyn wartości parametrów x_1 i x_2 .

statistica07.stw* - dane4 [] [] [X]

	1 x1	2 x2	3 y	4 x1x2
1	2,1	5,8	24,4	12,18
2	1,1	4,6	18,4	5,06
3	3,1	2,4	16	7,44
4	1,1	5,6	22,2	6,16
5	2,4	2,2	14	5,28
6	4,4	2,3	18,6	10,12
7	4,1	4,1	24,2	16,81
8	1,7	3,7	17,7	6,29
9	1,1	3,7	16,4	4,07
10	1,1	2,1	11,4	2,31

[] dane1 [] dane2 [] dane3 [] dane4

Znaleziona w wyniku analizy funkcja regresji:

$$\hat{y} = 3,217286 + 1,696326 x_1 + 2,848357 x_2 + 0,125511 x_1 x_2$$

jest funkcją istotną (p -value = 0,000000), istotne są też trzy pierwsze współczynniki funkcji (p -value = 0,00826 dla wyrazu wolnego, p -value = 0,002326 dla współczynnika związanego ze zmienną x_1 i p -value = 0,000011 dla współczynnika związanego ze zmienną x_2), współczynnik czwarty nie jest istotny (p -value = 0,244856) i powinien zostać wyeliminowany z modelu.

The screenshot displays the results of a multiple regression analysis in Statistica 7.0. It includes three windows: 'Wybierz listy zmiennych zależnych i niezależnych', 'Wyniki regresji wielorakiej', and 'Podsumowanie regresji zmiennej zależnej: y (dane4 w statistica07.stw)'. The regression equation is $\hat{y} = 3,217286 + 1,696326 x_1 + 2,848357 x_2 + 0,125511 x_1 x_2$. The results show a highly significant model with $R^2 = 0,99485858$ and $p < 0,000000$. The coefficients for x_1 and x_2 are significant, while the coefficient for $x_1 x_2$ is not.

Wyniki regresji wielorakiej:

Zmn. zależ. y: Wielor. R = ,99742598 F = 386,9976
 R²= ,99485858 df = 3,6
 Liczba przyp. 10 Popraw. R²= ,99228787 p = ,000000
 Błąd standardowy estymacji: ,373365616
 Wyr. wolny 3,217285619 Błąd std.: ,8312066 t(6) = 3,8706 p = ,0083

Ważne współczynniki (podświetlone na czerwono):
 x1 b* = ,505 x2 b* = ,932 x1x2 b* = ,128
 (istotne b* są podświetlone na czerwono)

Podsumowanie regresji zmiennej zależnej: y (dane4 w statistica07.stw)

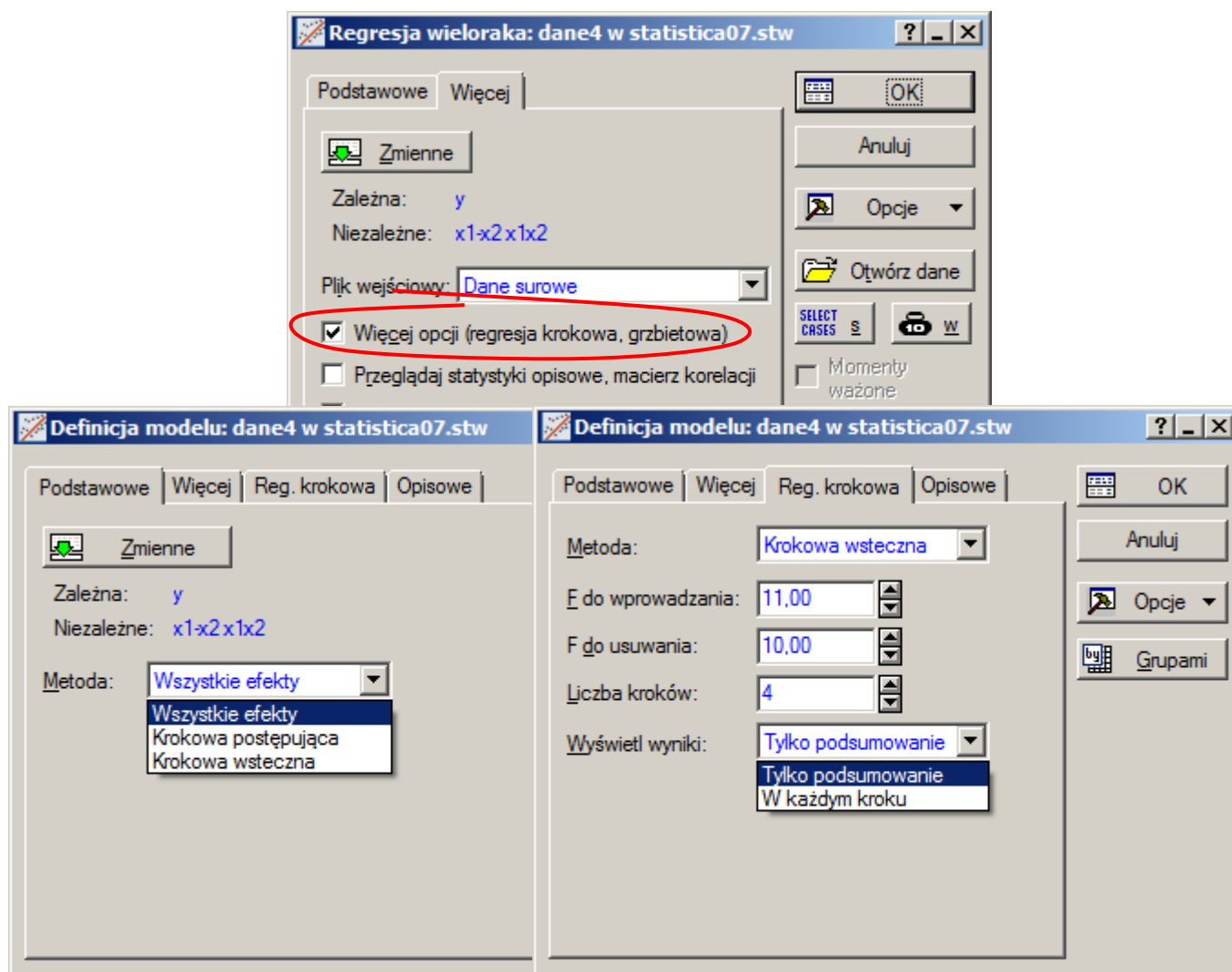
R = ,99742598 R² = ,99485858 Popraw. R² = ,99228787
 F(3,6) = 387,00 p < ,00000 Błąd std. estymacji: ,37337

	b*	Bł. std. z b*	b	Bł. std. z b	t(6)	p
N=10						
W. wolny			3,217286	0,831207	3,87062	0,008260
x1	0,504617	0,099857	1,696326	0,335679	5,05341	0,002326
x2	0,932207	0,069981	2,848357	0,213827	13,32085	0,000011
x1x2	0,127636	0,099019	0,125511	0,097370	1,28902	0,244856



8.2.2. Okno Regresja wieloraka – regresja krokowa

STATISTICA pozwala na krokowe konstruowanie modelu z wykorzystaniem selekcji postępującej i eliminacji wstecznej. Włączenie takiego trybu analizy wymaga ustawienia opcji **Więcej opcji (regresja krokowa, grzbietowa)** na zakładce **Więcej** w oknie regresji. Po włączeniu regresji krokowej przed wykonaniem właściwej analizy wyświetlane jest okno **Definicji modelu**, w którym można wskazać dokładnie sposób konstruowania modelu. Domyślnie wybrana metoda **Wszystkie efekty** poszukuje funkcji regresji zawierającej wszystkie zmienne niezależne. Dostępne są również metody krokowe: postępująca i wsteczna – ich działanie zależy od ustawień zakładki **Reg. krokowa**.



Załóżmy, że w zadaniu z przykładu poprzedniego należy znaleźć najprostszą postać funkcji regresji stosując eliminację wsteczną. W tym celu na zakładce **Podstawowe** lub **Reg. krokowa** należy wskazać metodę: **Krokowa wsteczna**.

Zakładka **Reg. krokowa** zawiera dodatkowe parametry wpływające na sposób przeprowadzania obliczeń. Pole **Liczba kroków** domyślnie jest ustawiane na maksymalną liczbę kroków analizy, liczba ta odpowiada maksymalnej liczbie składników funkcji regresji (w przykładzie 5. funkcja regresji miała 4 składniki). Pole **Wyświetl wyniki** pozwala wyświetlić wyniki obliczeń z każdego kroku regresji (opcja **W każdym kroku**) lub tylko ostateczny rezultat (opcja **Tylko podsumowanie**).

Na poniższych rysunkach przedstawione zostały wyniki kolejnych kroków analizy zadania z przykładu 5 z ustawioną eliminacją wsteczną. Wyniki z kroku 0 pokrywają się z wynikami z przykładu 5, po ich przeanalizowaniu przejście do kolejnego kroku analizy umożliwia przycisk **Następny**.

Wyniki regresji wielorakiej: dane4 w statistica07.stw

Wyniki regresji wielorakiej (Krok 0)

Zmn. zależ.y Wielor. R = ,99742598 F = 386,9976
R²= ,99485858 df = 3,6
Liczba przyp. 10 Popraw. R²= ,99228787 p = ,000000
Błąd standardowy estymacji: ,373365616
Wyr. wolny 3,217285619 Błąd std.: ,8312066 t(6) = 3,8706 p = ,0083

x1 b* = ,505 x2 b* = ,932 x1x2 b* = ,128
(istotne b* są podświetlone na czerwono)

Alfa do podświetlania efektów: ,05

statistica07.stw* - Podsumowanie regresji zmiennej zależnej: y (dane4 w statistic...

	b*	Bł. std. z b*	b	Bł. std. z b	t(6)	p
N=10						
W. wolny			3,217286	0,831207	3,87062	0,008260
x1	0,504617	0,099857	1,696326	0,335679	5,05341	0,002326
x2	0,932207	0,069981	2,848357	0,213827	13,32085	0,000011
x1x2	0,127636	0,099019	0,125511	0,097370	1,28902	0,244856

Podsumowanie regresji zmiennej zależnej: y (dane4 w statistica07.stw)
R= ,99742598 R²= ,99485858 Popraw. R²= ,99228787
F(3,6)=387,00 p<,00000 Błąd std. estymacji: ,37337

Wyniki regresji wielorakiej: dane4 w statistica07.stw

Wyniki regresji wielorakiej (krok 1 rozwiązanie)
żadne inne F do usuwania nie jest mniejsze od pro

Zmn. zależ.y Wielor. R = ,99671199 F = 529,6127
R²= ,99343478 df = 2,7
Liczba przyp. 10 Popraw. R²= ,99155901 p = ,000000
Błąd standardowy estymacji: ,390610497
Wyr. wolny 2,353317359 Błąd std.: ,5142982 t(7) = 4,5758 p = ,0026

x1 b* = ,627 x2 b* = 1,01
(istotne b* są podświetlone na czerwono)

Alfa do podświetlania efektów: ,05

statistica07.stw* - Podsumowanie regresji zmiennej zależnej: y (dane4 w statistic...

	b*	Bł. std. z b*	b	Bł. std. z b	t(7)	p
N=10						
W. wolny			2,353317	0,514298	4,57578	0,002557
x1	0,626944	0,032506	2,107542	0,109274	19,28686	0,000000
x2	1,013035	0,032506	3,095326	0,099323	31,16427	0,000000

Podsumowanie regresji zmiennej zależnej: y (dane4 w statistica07.stw)
R= ,99671199 R²= ,99343478 Popraw. R²= ,99155901
F(2,7)=529,61 p<,00000 Błąd std. estymacji: ,39061

W kroku 1 składnik zawierający iloczyn zmiennych x_1 i x_2 został wyeliminowany a otrzymana w tym kroku funkcja regresji:

$$\hat{y} = 2,353317 + 2,107542x_1 + 3,095326x_2$$

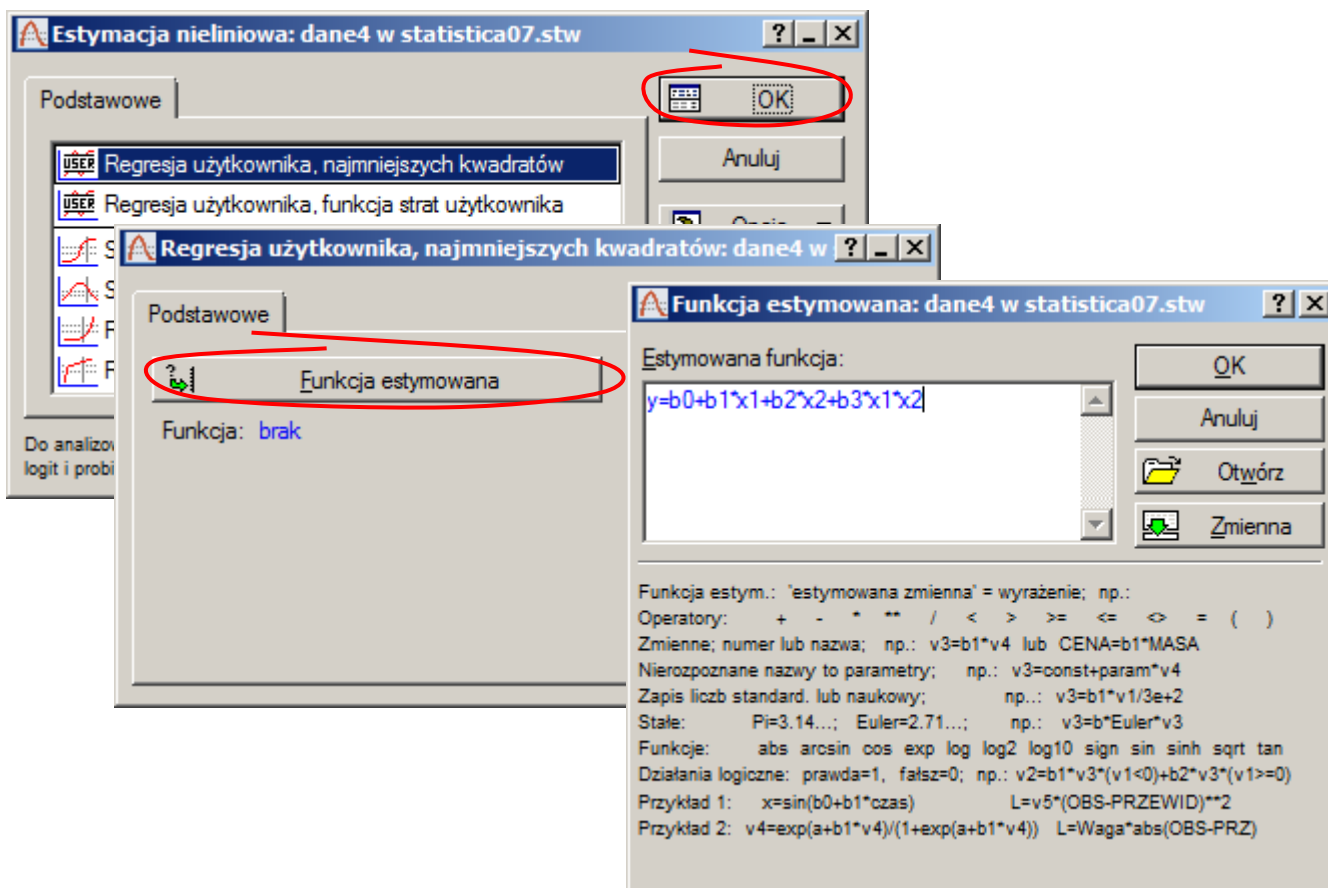
jest funkcją istotną (p -value = 0,000000), istotne są też jej wszystkie współczynniki (p -value = 0,002557 dla wyrazu wolnego, p -value = 0,000000 dla współczynników związanych ze zmienną x_1 i x_2). Uproszczenie postaci funkcji regresji nie spowodowało znaczącego pogorszenia wskaźników dopasowania:

krok	wsp. determinacji R^2	błąd std. estymacji s_e
0	$R^2 = 0,99485858$	$s_e = 0,373365616$
1	$R^2 = 0,99343478$	$s_e = 0,390610497$

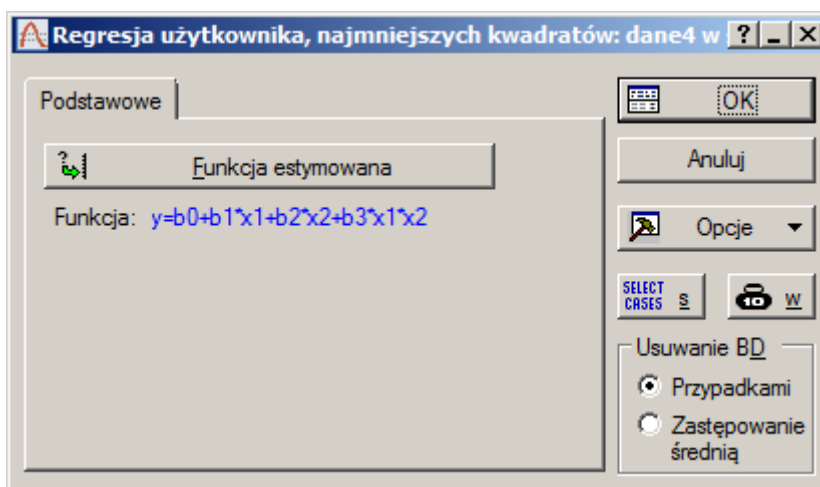
Po usunięciu z funkcji regresji składnika zawierającego iloczyn zmiennych dalsza eliminacja jest już niedostępna (przycisk **Następny** nie jest dostępny), analiza kończy się gdy w modelu występują wyłącznie zmienne, które w istotny sposób wyjaśniają zmienność zmiennej zależnej.

8.2.2. Okno Estymacja nieliniowa

W oknie estymacji nieliniowej należy wskazać model regresji. Opcja **Regresja użytkownika, najmniejszych kwadratów** – umożliwi użytkownikowi podanie ogólnej postaci funkcji regresji, przyjmuje jednocześnie sposób oceny jakości dopasowania danych do modelu w postaci kwadratu różnicy pomiędzy wartościami obserwowanymi a estymowanymi.

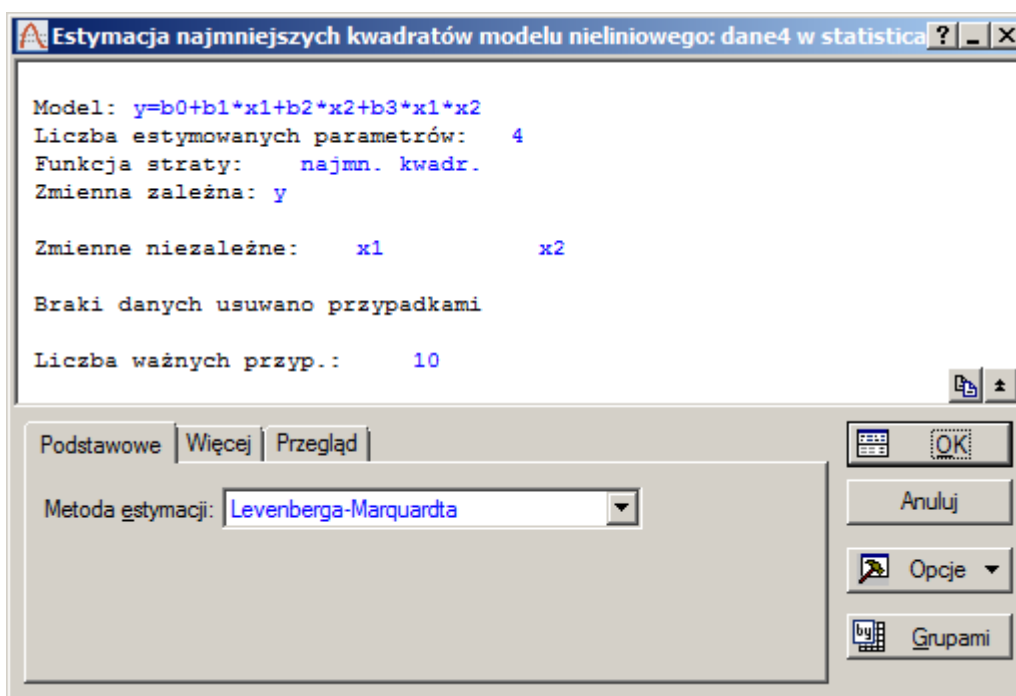


Zakładaną postać funkcji regresji należy wskazać w oknie **Funkcja estymowana** podając wyrażenie opisujące wartość zmiennej zależnej w funkcji zmiennych niezależnych. Wszystkie symbole występujące w wyrażeniu, które nie odpowiadają zmiennym aktywnego arkusza są interpretowane jako parametry funkcji regresji.

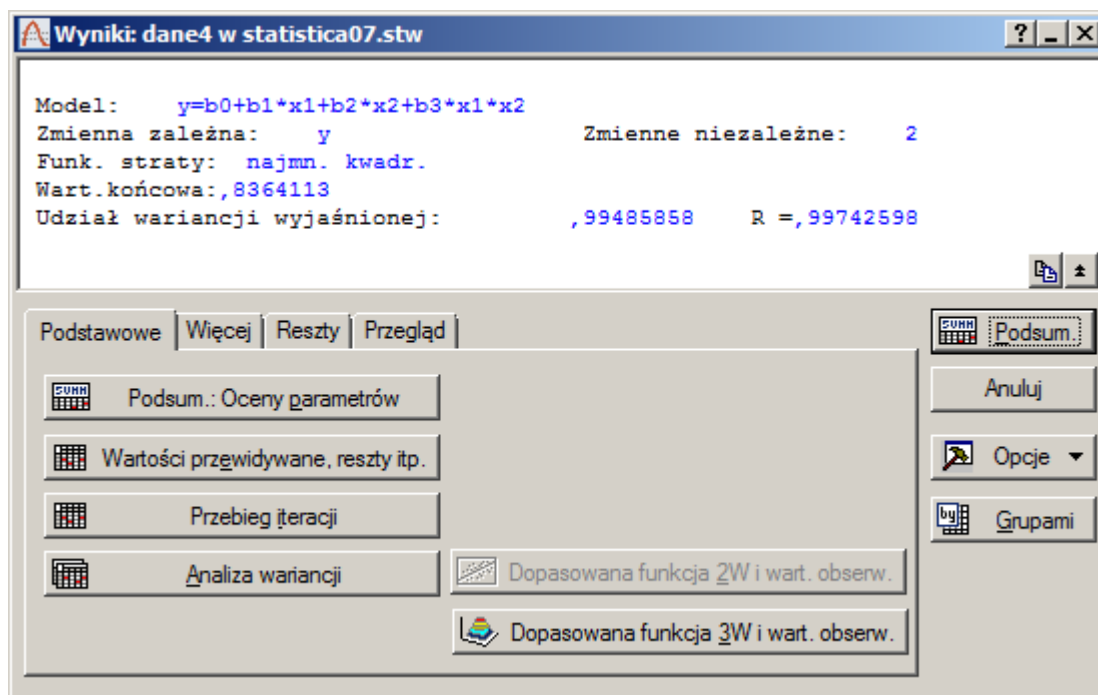


Funkcja zdefiniowana na powyższym rysunku odpowiada funkcji regresji z przykładu 5. Warto zauważyć, że wykorzystanie okna **Estymacja nieliniowa** nie wymaga (w odróżnieniu od rozwiązania przedstawionego w poprzednim punkcie) uzupełniania arkusza o dodatkową zmienną odpowiadającą iloczynowi zmiennych x_1 i x_2 .

Po wprowadzeniu funkcji i zaakceptowaniu przyciskiem OK okna **Regresja użytkownika** w kolejnym oknie należy wskazać metodę estymacji – domyślnie wybrany jest algorytm *Levenberga-Marquardta*.



Wyniki analizy wyświetlane są w następnym oknie. W górnym polu okna wyników wyświetlane są: **Wart. końcowa** – zmienność zmiennej zależnej niewyjaśniona równaniem regresji, tzn. *sse*, **Udział wariancji wyjaśnionej** – wartość współczynnika determinacji R^2 , **R** – wartość współczynnika korelacji.



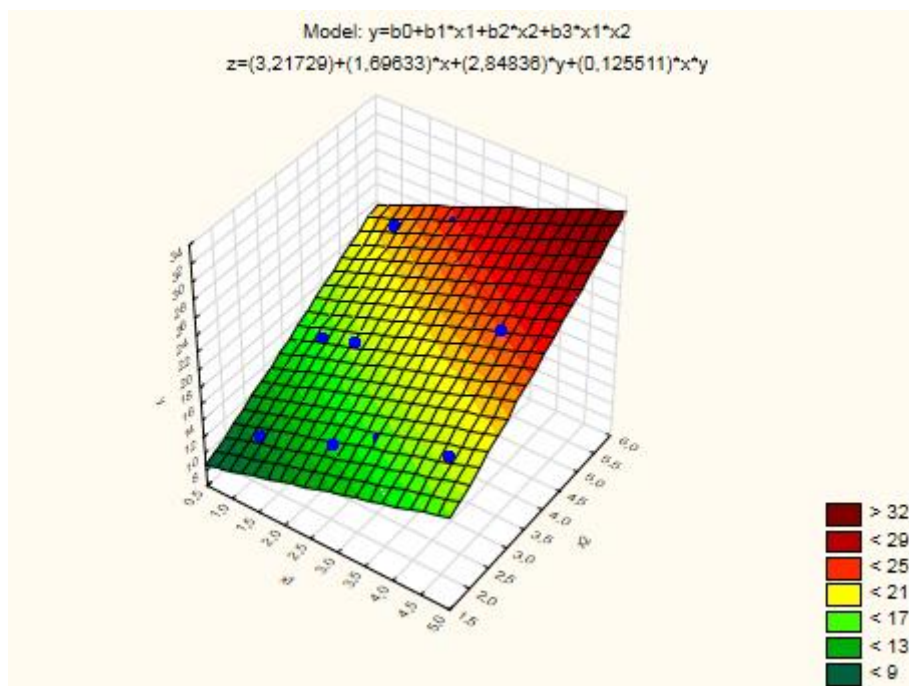
Pełne wyniki analizy wyświetlane są w dodatkowym arkuszu po naciśnięciu przycisku **Podsum: Oceny parametrów**. Dla każdego składnika funkcji regresji wyświetlane są:

- Ocena** – wartość właściwego współczynnika,
- Błąd stand.** – błąd oceny współczynnika,
- Wart. t** – wartość statystyki t wykorzystywanej do oceny istotności współczynnika,
- p** – wartość granicznego poziomu istotności p -value testu dla hipotezy zerowej o braku istotności współczynnika,

Doln. uf Granica, Górn. uf Granica – przedział ufności dla współczynnika.

	Ocena	Błąd stand.	Wart. t df = 6	p	Doln. uf Granica	Górn. uf Granica
b0	3,217286	0,831207	3,87062	0,008260	1,183397	5,251175
b1	1,696326	0,335679	5,05341	0,002326	0,874948	2,517704
b2	2,848356	0,213827	13,32085	0,000011	2,325141	3,371572
b3	0,125511	0,097370	1,28902	0,244856	-0,112744	0,363766

Wyniki analizy pokrywają się z wynikami otrzymanymi przy pomocy okna **Regresja wieloraka**. Okno estymacji udostępnia dla funkcji regresji jednej i dwu zmiennych niezależnych wygodny sposób przygotowania wykresów rozrzutu z nałożoną funkcją regresji (przyciski: **Dopasowana funkcja 2W i wart. obserw.**, **Dopasowana funkcja 3W i wart. obserw.**). W oknie estymacji nieliniowej nie jest natomiast możliwe przeprowadzenie regresji krokowej.



Z otrzymanej analizy wynika, że współczynnik b_3 jest nieistotny. Podobnie jak poprzednio funkcję regresji można uprościć eliminując z równania regresji składnik zawierający iloczyn zmiennych x_1 i x_2 . Wyniki analizy dla uproszczonej funkcji zostały pokazane poniżej (wyniki te są identyczne z otrzymanymi w punkcie Regresja krokowa).

Regresja użytkownika, najmniejszych kwadratów: dane4 w

Podstawowe

Funkcja estymowana

Funkcja: $y=b_0+b_1*x_1+b_2*x_2$

OK

Anuluj

Opcje

SELECT CASES

Wyniki: dane4 w statistica07.stw

Model: $y=b_0+b_1*x_1+b_2*x_2$

Zmienna zależna: y Zmienne niezależne: 2

Funk. straty: najmn. kwadr.

Wart. końcowa: 1,06803592

Udział wariancji wyjaśnionej: ,99343478 $R = ,99671199$

Podstawowe Więcej Reszty Przegląd

Podsum.: Oceny parametrów

Wartości przewidywane, reszty itp.

Przebieg iteracji

Analiza wariancji

Dopasowana funkcja 2W i wart. obserw.

Dopasowana funkcja 3W i wart. obserw.

Podsum

Anuluj

Opcje

Grupami

statistica07.stw* - Model: $y=b_0+b_1*x_1+b_2*x_2$ (dane4 w statistica07.stw)

Model: $y=b_0+b_1*x_1+b_2*x_2$ (dane4 w statistica07.stw)
 Zmn. zal. : y
 Poziom ufności: 95.0% (alfa=0.050)

	Ocena	Błąd stand.	Wart. t df = 7	p	Doln. uf Granica	Górn. uf Granica
b0	2,353317	0,514298	4,57578	0,002557	1,137195	3,569439
b1	2,107542	0,109273	19,28686	0,000000	1,849152	2,365933
b2	3,095326	0,099323	31,16427	0,000000	2,860464	3,330187

Model: $y=b_0+b_1*x_1+b_2*x_2$ (dane4 w statistica07.stw)

