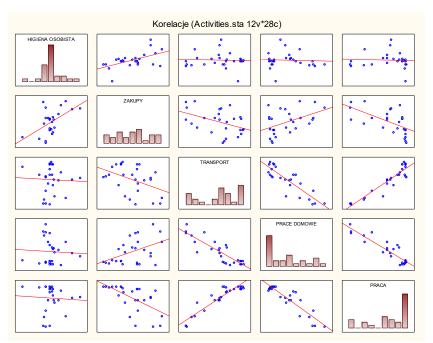
Statystyka i teoria obsługi masowej lab. 7 - Regresja wieloraka Dawid Bitner

Przeprowadź analizę regresji wielorakiej dla danych zawartych w arkuszu Activities.sta(arkuszz przykładów Statistici, na początek wybierz 4 zmienne niezależne). Należy uwzględnić badanie nadmiarowości zmiennych oraz analizę reszt. Jeśli w danych znajdują się punkty odstające, należy je wyeliminować. Po wyznaczeniu ostatecznej wersji równania regresji przeprowadź predykcję dla dwóch wybranych zestawów wartości zmiennych niezależnych.

Wybrana zmienna zależna: PRACA; Wybrane zmienne niezależne: TRANSPORT, PRACE DOMOWE, ZAKUPY, HIGIENA OSOBOSTA

	Dane o cz	asie poświęca	nym na różne czynr	ności dla 2	8 grup po	oulacji. Zmodyfikowan	e dane op	ublikowane	w Explor	atory and Multiva	riate Data	a Analysis
	1	2	3	4	5	6	7	8	9	10	11	12
			PRACE DOMOWE	DZIECI		HIGIENA OSOBISTA		SEN		WYPOCZYNEK	PŁEĆ	REGION
EMU	610		60	10		95		760	175			ZACHÓD
EWU	475	90	250	30		120		775	115			ZACHÓD
UWU	10		495	110		110		785	160			ZACHÓD
MMU	615		65	10		90		765	180			ZACHÓD
MWU	179		421	87	161	112	119	776	143			ZACHÓD
SMU	585		50		150	105		760	150			ZACHÓD
SWU	482		196	18		130		775	132			ZACHÓD
EMW	652		95	7	57	85		807	115			ZACHÓD
EWW	510		307	30		95		815	87			ZACHÓD
UWW	20		567	87	112	90		842	125			ZACHÓD
MMW	655		97	10	52	85		807	122			ZACHÓD
MWW	168		529	69		83		825	119			ZACHÓD
SMW	642		72		62	77	140	812	100			ZACHÓD
SWW	389		262	14	92	97	147	848	84			ZACHÓD
EME	650		122	22	76	94	100	764	96			WSCHÓE
EWE	578		338	42		94		752	64			WSCHÓE
UWE	24		594	72		82		840	86			WSCHÓE
MME	652	133	134	22	68	54	102	762	122	310	MĘŻCZY.	WSCHÓE
MWE	434		431	60		88		770	73			WSCHÓE
SME	627	148	68		88	92		770	58	463	MĘŻCZY.	WSCHÓE
SWE	433		296	21	128	102		798	58			WSCHÓE
EMY	650		120	15		90		760	70		MĘŻCZY.	WSCHÓE
EWY	560		375	45		90		745	60			WSCHÓE
UWY	10		710	55		85		815	60			WSCHÓE
MMY	650		112	15		90		760	80			WSCHÓE
MWY	260	52	576	59	116	85	117	775	65			WSCHÓE
SMY	615	125	95		115	90	85	760	40	475	MĘŻCZY	WSCHÓE
SWY	433	89	318	23	112	96	102	774	45			WSCHÓE

Macierzowy wykres rozrzutu wybranych zmiennych:



```
Wyniki regresji wielorakiej:

Zmn. zależ.PRACA Wielor. R = ,97456511 F = 104,0119
R^2= ,94977715 df = 4,22
Liczba przyp. 27 Popraw. R^2= ,94064572 p = ,000000
Bład standardowy estymacji:52,148697624
Wyr. wolny 360,45821016 Bład std.: 107,6850 t( 22) = 3,3473 p = ,0029

HIGIENA OSOBI b*=,112 ZAKUPY b*=-,29 TRANSPORT b*=,571
PRACE DOMOWE b*=-,28
```

(istotne b* są podświetlone na czerwono)

Sprawdzamy koleracje:

	Korelacje (Activities.sta)				
Zmienna	HIGIENA OSOBISTA	ZAKUPY	TRANSPORT	PRACE DOMOWE	PRACA
HIGIENA OSOBISTA	1,000000	0,542519	-0,043739	-0,052481	-0,053022
ZAKUPY	0,542519	1,000000	-0,428002	0,463065	-0,597700
TRANSPORT	-0,043739	-0,428002	1,000000	-0,869074	0,930138
PRACE DOMOWE	-0,052481	0,463065	-0,869074	1,000000	-0,912716
PRACA	-0,053022	-0,597700	0,930138	-0,912716	1,000000

Możemy zauważyć, że prawie wszystkie zmienne są mocno skolerowane ze zmienną PRACA dodanie lub ujemnie. Bardzo niski współczynnik koleracji występuje w przypadku zmiennych PRACA i HIGIENA OSOBISTA. Poniżej sprawdzimy, czy możemy zredukować nasz badany model o zmienną HIGIENA OSOBISTA.

Koleracje cząstkowe:

	Aktualnie w ró	ktualnie w równaniu są zmienne ; DV: PRACA (Activities.sta)									
	b* w	b* w Cząstk. Semicz. Tolerancja R-kwadr. t(22									
Zmienna		Korelac.	Korelac.	-							
HIGIENA OSOBISTA	0,111929	0,352188	0,084330	0,567647	0,432353	1,76499	0,091444				
ZAKUPY	-0,285028	-0,653464	-0,193464	0,460710	0,539290	-4,04913	0,000535				
TRANSPORT	0,570609	0,777215	0,276808	0,235332	0,764668	5,79348	800000,0				
PRACE DOMOWE	-0,278954	-0,487370	-0,125083	0,201062	0,798938	-2,61793	0,015708				

Współczynnik regresji wielorakiej i istotności modelu:

	Stat.podsum.; Zmn	. zal.:PRACA (Activities.sta)
statystyka	Wartość	
R wielorakie	0,974565108	
Wielorakie R2	0,94977715	
Skorygowane R2	0,940645723	
F(4,22)	104,011906	
р	5,86985769E-14	
Błąd std. estymacji	52,1486976	

Współczynniki równania regresji i ich istotność:

	Podsumowani	Podsumowanie regresji zmiennej zależnej: PRACA (Activities.sta)									
	R= ,97456511 R^2= ,94977715 Popraw. R2= ,94064572										
	F(4,22)=104,0	F(4,22)=104,01 p<,00000 Błąd std. estymacji: 52,149									
	b*	b* Bł. std. b Bł. std. t(22) p									
N=27		z b*		z b							
W. wolny			360,4582	107,6850	3,34734	0,002915					
HIGIENA OSOBISTA	0,111929	0,063416	1,7200	0,9745	1,76499	0,091444					
ZAKUPY	-0,285028	0,070392	-1,9817	0,4894	-4,04913	0,000535					
TRANSPORT	0,570609	0,098492	2,6613	0,4594	5,79348	0,000008					
PRACE DOMOWE	-0,278954	0,106555	-0,3021	0,1154	-2,61793	0,015708					

Dane te potwierdzają moje wcześniejsze przypuszczenie. Wyniki analizy po usunięciu nieistotnej zmiennej prezentują się następująco:

```
Wyniki regresji wielorakiej :
```

```
Zmn. zależ.PRACA Wielor. R = ,97090967 F = 126,0517 R^2 ,94266558 df = 3,23 Liczba przyp. 27 Popraw. R^2= ,93518718 p = ,000000 Błąd standardowy estymacji:54,493903508 Wyr. wolny 494,75154415 Błąd std.: 79,62941 t( 23) = 6,2132 p = ,0000
```

ZAKUPY b*=-,21 TRANSPORT b*=,538 PRACE DOMOWE b*=-,35

(istotne b* są podświetlone na czerwono)

Możemy zauważyć, że poprzez usunięcie zmiennej HIGIENA OSOBISTA zmniejszył nam się

błąd standardowy wyrazu wolnego:

	Aktualnie w ró	ktualnie w równaniu są zmienne ; DV: PRACA (Activities.sta)										
	b* w	t(23)	р									
Zmienna		Korelac.	Korelac.									
ZAKUPY	-0,205325	-0,604443	-0,181675	0,782900	0,217100	-3,63875	0,001373					
TRANSPORT	0,538067	0,742879	0,265719	0,243878	0,756122	5,32205	0,000021					
PRACE DOMOWE	-0,350017	-0,577805	-0,169514	0,234550	0,765450	-3,39518	0,002487					

Koleracje w tej chwili prezentują się następująco:

				-t t							
		Podsumowanie regresji zmiennej zależnej: PRACA (Activities.sta)									
	R = ,97090967	R= ,97090967 R^2= ,94266558 Popraw. R2= ,93518718									
	F(3,23)=126,0	F(3,23)=126,05 p<,00000 Błąd std. estymacji: 54,494									
	b*	b* Bł. std. b Bł. std. t(23) p									
N=27		z b*		z b							
W. wolny			494,7515	79,62941	6,21318	0,000002					
ZAKUPY	-0,205325	0,056427	-1,4275	0,39231	-3,63875	0,001373					
TRANSPORT	0,538067	0,101101	2,5095	0,47154	5,32205	0,000021					
PRACE DOMOWE	-0,350017	0,103092	-0,3791	0,11165	-3,39518	0,002487					

Wzór modelu regresji możemy odczytać z powyższego podsumowania: PRACA = 494,7515 - 1,4275 ZAKUPY + 2,5095 TRANSPORT - 0,3791 PRACE DOMOWE.

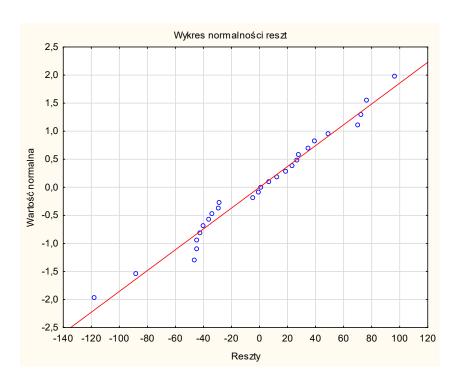
Wartości przewidywane i reszty

	Wartości przew PRACA		У						
	Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard. Przewid.	Standard. Reszta	Bł. std. W.przew.	Mahaln. Odległ.	Usunięte Reszta	Cooka Odległ.
EMU	610,000000	652,038635	-42,038635	0,899450	-0,771437	19,630449	2,410982	-48,307339	0,025494
EWU	475,000000	425,989594	49,010406	-0,188243	0,899374	17,289085	1,654144	54,495850	0,025166
MMU	615,000000	659,790527	-44,790527	0,936750	-0,821936	18,573610	2,057477	-50,677807	0,025118
MWU	179,000000	178,11030€	0,889694	-1,380978	0,016326	25,256727	4,622146	1,133097	0,000023
SMU	585,000000	550,264832	34,735168	0,409739	0,637414	29,072453	6,437192	48,554943	0,056491
SWU	482,000000	455,069489	26,930511	-0,048318	0,494193	19,588842	2,396695	30,926800	0,010405
EMW	652,000000	628,324768	23,675232	0,785344	0,434457	23,656179	3,936706	29,172815	0,013502
EWW	510,000000	439,844482	70,155518	-0,121577	1,287401	16,808662	1,510719	77,532036	0,048148
UWW	20,000000	137,506821	-117,506821	-1,576352	-2,156330	22,928432	3,639881	-142,784271	0,303849
MMW	655,000000	627,175720	27,824280	0,779815	0,510594	25,589138	4,770128	35,695194	0,023653
MWW	168,000000	203,829514	-35,829514	-1,257224	-0,657496	21,138123	2,949141	-42,175468	0,022532
SMW	642,000000	642,453186	-0,453186	0,853327	-0,008316	22,543274	3,486540	-0,546755	0,000004
SWW	389,000000	349,428558	39,571442	-0,556636	0,726163	29,535328	6,674711	56,030891	0,077640
EME	650,000000	696,367432	-46,367432	1,112749	-0,850874	18,480036	2,027119	-52,392765	0,026576
EWE	578,000000	481,320953	96,679047	0,077998	1,774126	18,254845	1,954691	108,899475	0,112036
UWE	24,000000	64,115120	-40,115120	-1,929495	-0,736140	24,214533	4,170728	-49,984550	0,041531
MME	652,000000	680,653137	-28,653137	1,037136	-0,525804	18,431610	2,011469	-32,354534	0,010082
MWE	434,000000	357,588898	76,411102	-0,517371	1,402195	16,699615	1,478726	84,330688	0,056225
SME	627,000000	714,763489	-87,763489	1,201267	-1,610519	17,374287	1,680003	-97,694366	0,081677
SWE	433,000000	420,664124	12,335876	-0,213868	0,226372	13,305758	0,587126	13,117953	0,000864
EMY	650,000000	679,258667	-29,258667	1,030426	-0,536916	16,796217	1,507057	-32,330044	0,008360
EWY	560,000000	487,626709	72,373291	0,108340	1,328099	22,769924	3,576460	87,681969	0,113004
UWY	10,000000	43,721062	-33,721062	-2,027627	-0,618804	26,611948	5,237596	-44,281429	0,039368
MMY	650,000000	694,838806	-44,838806	1,105394	-0,822822	17,903498	1,843461	-50,264301	0,022958
MWY	260,000000	241,314209	18,685791	-1,076856	0,342897	23,202351	3,750515	22,823387	0,007950
SMY	615,000000	608,266113	6,733887	0,688827	0,123571	16,354723	1,378913	7,400462	0,000415
SWY	433,000000	437,674866	-4,674866	-0,132017	-0,085787	11,768654	0,249675	-4,903568	0,000094

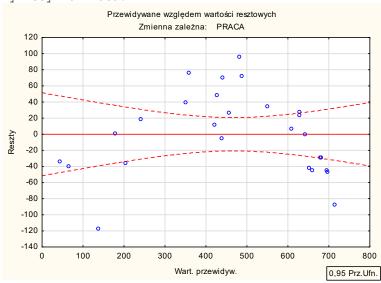
Średnia reszt w tym przypadku wynosi: -0,000000565140336

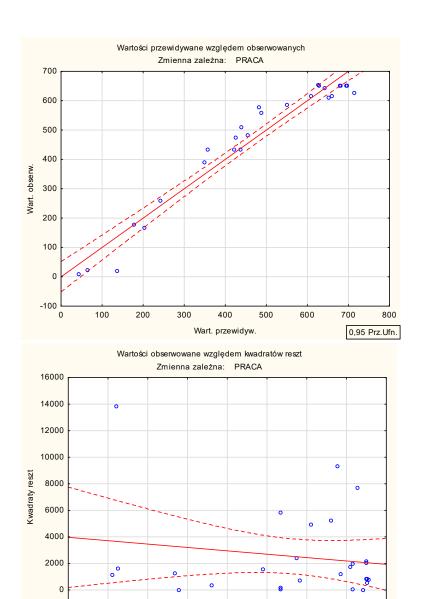
Jest ona bliska zeru, więc reszty są nieobciążone, nasz model regresji wielorakiej jest użyteczny.

WYkres normalności reszt prezentuje się tak:



Wykresy rozrzutu:

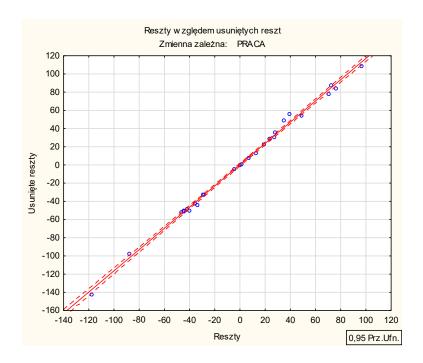




-2000 L -100

Wart. obserw.

0,95 Prz.Ufn.

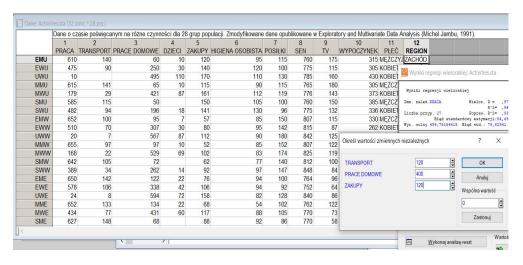


Przeprowadzamy dokładniejszą analizę skolerowaną na poszukiwanie wartości odstających:

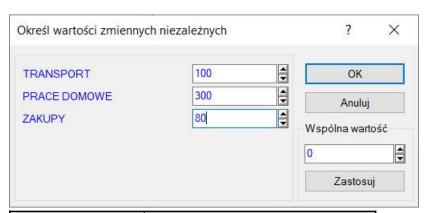
							Standaryz, p	zewidywane (Ad	tivities.sta)					
							Uporządkowa		,					
	St	andar	yz.	prze	ewidywan	ie	Obserw.	Przewidyw.	Reszta	Standard	Standard	Bł. std.	Mahaln.	Usuni
Nazwa przypadku	-2,0 .					1,20	Wartość	Wartość		Przewid.	Reszta	W.przew.	Odległ.	Res
UWY	.*						10,0000	43,7211	-33,721	-2,02763	-0,61880	26,61195	5,23759€	-44
UWE	.*						24,0000	64,1151	-40,115	-1,92950	-0,73614	24,21453	4,170728	-49
UWW		*					20,0000	137,5068	-117,507	-1,57635	-2,15633	22,92843	3,639881	-14:
MWU		.*					179,0000	178,1103	0,890	-1,38098	0,01633	25,25673	4,62214€	1
MWW		. *		- 1			168,0000	203,8295	-35,830	-1,25722	-0,65750	21,13812	2,949141	-4:
SME						. *	627,0000	714,7635	-87,763	1,20127	-1,61052	17,37429	1,680003	-91
EME				- 1		. *.	650,0000	696,3674	-46,367	1,11275	-0,85087	18,48004	2,027119	-51
MMY				- 1		. *.	650,0000	694,8388	-44,839	1,10539	-0,82282	17,90350	1,843461	-50
MWY			*.	- 1			260,0000	241,3142	18,686	-1,07686	0,34290	23,20235	3,750515	-3:
MME				- 1		. * .	652,0000	680,6531	-28,653	1,03714	-0,52580	18,43161	2,011469	-31
EMY				- 1		. * .	650,0000	679,2587	-29,259	1,03043	-0,53692	16,79622	1,507057	-31
MMU						. * .	615,0000	659,7905	-44,791	0,93675	-0,82194	18,57361	2,057477	-50
EMU				- 1		. * .	610,0000	652,038€	-42,039	0,89945	-0,77144	19,63045	2,410982	-48
SMW						.* .	642,0000	642,4532	-0,453	0,85333	-0,00832	22,54327	3,486540	-0
EMW						.* .	652,0000	628,3248	23,675	0,78534	0,43446	23,65618	3,93670€	2
MMW						.* .	655,0000	627,1757	27,824	0,77982	0,51059	25,58914	4,770128	3
SMY				- 1		* .	615,0000	608,2661	6,734	0,68883	0,12357	16,35472	1,378913	- 1
SWW				*			389,0000	349,428€	39,571	-0,55664	0,72616	29,53533	6,674711	50
MWE				*			434,0000	357,5889	76,411	-0,51737	1,40220	16,69962	1,478726	84
SMU					. *		585,0000	550,2648	34,735	0,40974	0,63741	29,07245	6,437192	4
SWE				1	* .		433,0000	420,6641	12,336	-0,21387	0,22637	13,3057€	0,587126	1
EWU				Ì	* .		475,0000	425,989€	49,010	-0,18824	0,89937	17,28909	1,654144	54
SWY				П	* .		433,0000	437,6749	-4,675	-0,13202	-0,08579	11,76865	0,249675	-4
EWW				- 1	* .		510,0000	439,8445	70,156	-0,12158	1,28740	16,80866	1,510719	7
EWY				İ	*		560,0000	487,6267	72,373	0,10834	1,32810	22,76992	3,576460	8
EWE				İ	*.		578,0000	481,3210	96,679	0,07800	1,77413	18,25484	1,954691	10
SWU				Ì	* .		482,0000	455,0695	26,931	-0,04832	0,49419	19,58884	2,396695	30
Minimum	.*			I			10,0000	43,7211	-117,507	-2,02763	-2,15633	11,76865	0,249675	-14
Maksim.				i		. *	655,0000	714,7635	96,679	1,20127	1,77413	29.53533	6.674711	10
Średnia				i	*.		465,1111	465,1111	-0,000	0,00000	-0,00000	20,51031	2,888889	- 1
Mediana				i	*.		560,0000	481,3210	0.890	0.07800	0.01633	19,58884	2,396695	

Tabela przedstawia najbardziej odstające przypadki na podstawie odległości Mahanalobista. W naszym przypadku, zostały uwzględnione tutaj wszystkie przypadki, ponieważ jest ich mniej niż 100.

Predukcja dla dwóch wybranych zestawów wartości zmiennych niezależnych:



	Obliczanie wartości (Activities.sta) zmiennej: PRACA								
	Wagi b Wartość Wagi b								
Zmienna			*Wartość						
TRANSPORT	2,50953	120,0000	301,144						
PRACE DOMOWE	-0,37906	400,0000	-151,624						
ZAKUPY	-1,42753	120,0000	-171,304						
W. wolny			494,752						
Przewidyw.			472,967						
-95,0%GU			411,131						
+95,0%GU			534,804						



	Obliczanie wartości (Activities.sta)							
	zmiennej: PRACA							
	Wagi b Wartość Wagi							
Zmienna			*Wartość					
TRANSPORT	2,50953							
PRACE DOMOWE	-0,37906	300,0000	-113,718					
ZAKUPY	-1,42753	80,0000	-114,203					
W. wolny			494,752					
Przewidyw.			517,784					
-95,0%GU			482,403					
+95,0%GU			553,165					

Rozrzut w 3D badanych zmiennych prezentuje się następująco:

