

Statystyka i teoria obsługi masowej

Laboratorium 5 - Korelacja i regresja
Damian Kwaśniok

W zadaniu szukając pary zmiennych dla których będzie można utworzyć model regresji przyjmiemy zmienna **GEOMETRY** jako zmienną objaśnianą i dla niej spróbujemy stworzyć model regresji. Za poziom istotności alfa przyjmiemy wartość 0,05.

Przed przystąpieniem do wyznaczania równania regresji należy sprawdzić podstawowe charakterystyki danych.

Arkusz zawierający wyznaczone współczynniki korelacji liniowej Pearsona pomiędzy wybranymi zmiennymi:

Zmienna	Korelacje (School performance.sta) Oznaczone wsp. korelacji są istotne z p < ,05000 N=80 (Braki danych usuwano przypadkami)							
	GEOMETRY	READING	GRAMMAR	DRAWING	CALCULUS	HISTORY	WRITING	SPELLING
GEOMETRY	1,000000	-0,035582	-0,014342	0,903247	0,809318	-0,020458	-0,031098	0,018067
READING	-0,035582	1,000000	0,943131	-0,034940	-0,009269	0,950956	0,942327	0,860010
GRAMMAR	-0,014342	0,943131	1,000000	-0,025917	0,005717	0,924746	0,901965	0,837787
DRAWING	0,903247	-0,034940	-0,025917	1,000000	0,814418	-0,032686	-0,033889	0,016422
CALCULUS	0,809318	-0,009269	0,005717	0,814418	1,000000	-0,017948	0,022055	0,040587
HISTORY	-0,020458	0,950956	0,924746	-0,032686	-0,017948	1,000000	0,909612	0,848156
WRITING	-0,031098	0,942327	0,901965	-0,033889	0,022055	0,909612	1,000000	0,882936
SPELLING	0,018067	0,860010	0,837787	0,016422	0,040587	0,848156	0,882936	1,000000

W powstałym raporcie widać, że zmienna **DRAWING** jest silnie skorelowana ze zmienną **CALCULUS** zatem nie włączymy ich obu do jednego modelu.

Wyniki regresji wielorakiej dla wszystkich zmiennych z arkusza prezentują się następująco:

Zmn. zależ.	GEOMETRY	Wielor. R =	,91391552	F =	52,14335
		R^2=	,83524157	df =	7,72
Liczba przyp.	80	Popraw. R^2=	,81922339	p =	0,000000
		Błąd standardowy estymacji:	,764733743		
Wyr. wolny	20,769339474	Błąd std.:	5,022569	t(72) =	4,1352 p = ,0001

READING b*=-,16	GRAMMAR b*=-,068	DRAWING b*=-,716
CALCULUS b*=-,228	HISTORY b*=-,144	WRITING b*=-,07
SPELLING b*=-,012		

(istotne b* są podświetlone na czerwono)

N=80	Podsumowanie regresji zmiennej zależnej: GEOMETRY (School performance.s R= ,91391552 R^2= ,83524157 Popraw. R2= ,81922339 F(7,72)=52,143 p<0,0000 Błąd std. estymacji: ,76473					
	b*	Bł. std. z b*	b	Bł. std. z b	t(72)	p
W. wolny			20,76934	5,022569	4,135203	0,000095
READING	-0,155143	0,213959	-0,08467	0,116762	-0,725108	0,470736
GRAMMAR	0,067514	0,150800	0,03457	0,077220	0,447708	0,655708
DRAWING	0,716388	0,083775	0,62174	0,072706	8,551316	0,000000
CALCULUS	0,227645	0,083974	0,17515	0,064610	2,710906	0,008384
HISTORY	0,143645	0,163394	0,07401	0,084190	0,879128	0,382257
WRITING	-0,067609	0,160631	-0,03600	0,085537	-0,420896	0,675085
SPELLING	0,011787	0,105724	0,00564	0,050610	0,111491	0,911537

Z otrzymanych danych wynika, że uwzględnienie zmiennych **READING**, **GRAMMAR**, **HISTORY**, **WRITING** oraz **SPELLING** jest zbyt duże, ponieważ ich udział w wariancji zmiennej **GEOMETRY** jest niewielki, a wyniki testu dla odpowiadających im współczynników równania regresji świadczą o ich nieistotności.

Korelacje cząstkowe prezentują się następująco:

Zmienna	Aktualnie w równaniu są zmienne ; DV: GEOMETRY (School performance.sta)						
	b* w	Cząstk. Korelac.	Semicz. Korelac.	Tolerancja	R-kwadr.	t(72)	p
READING	-0,155143	-0,085144	-0,034686	0,049987	0,950013	-0,725108	0,470736
GRAMMAR	0,067514	0,052690	0,021417	0,100626	0,899374	0,447708	0,655708
DRAWING	0,716388	0,709842	0,409063	0,326050	0,673950	8,551316	0,000000
CALCULUS	0,227645	0,304329	0,129680	0,324509	0,675491	2,710906	0,008384
HISTORY	0,143645	0,103055	0,042054	0,085712	0,914288	0,879128	0,382257
WRITING	-0,067609	-0,049542	-0,020134	0,088687	0,911313	-0,420896	0,675085
SPELLING	0,011787	0,013138	0,005333	0,204722	0,795278	0,111491	0,911537

Na podstawie otrzymanych wyników wybieramy te zmienne, których współczynniki regresji są istotne, a ich wartość tolerancji jest w miarę duża, są to zmienne: **DRAWING** oraz **CALCULUS**.

Wyniki regresji wielorakiej dla wybranych zmiennych prezentują się następująco:

Zmn. zależ.	GEOMETRY	Wielor. R =	,91213216	F =	190,6463
		R^2=	,83198507	df =	2,77
Liczba przyp.	80	Popraw. R^2=	,82762104	p =	0,000000
		Błąd standardowy estymacji:	,746760359		
Wyr. wolny	20,213513414	Błąd std.:	4,079597	t(77) =	4,9548 p = ,0000

DRAWING b*=-,725	CALCULUS b*=-,219
------------------	-------------------

(istotne b* są podświetlone na czerwono)

Eliminacja zmiennych poprawiła model, wartość błędu standardowego dla wyrazu wolnego się zmniejszyła.

Pamiętamy, że zmienne **DRAWING** oraz **CALCULUS** były ze sobą bardzo silnie skorelowane (wartość korelacji liniowej Pearsona = 0,8144), więc jedną z nich należy usunąć z modelu. Przeanalizujemy, która zmienna jest istotniejsza.

Wyniki bez zmiennej **DRAWING**:

Zmn. zależ.	GEOMETRY	Wielor. R =	,80931803	F =	148,0841
		R^2=	,65499567	df =	1,78
Liczba przyp.	80	Popraw. R^2=	,65057254	p =	0,000000
		Błąd standardowy estymacji:	1,063206559		
Wyr. wolny	37,542254650	Błąd std.:	5,121782	t(78) =	7,3299 p = ,0000

CALCULUS b*=-,809

(istotne b* są podświetlone na czerwono)

Wyniki bez zmiennej **CALCULUS**:

Zmn. zależ.	GEOMETRY	Wielor. R =	,90324713	F =	345,5801
		R^2=	,81585538	df =	1,78
Liczba przyp.	80	Popraw. R^2=	,81349455	p =	0,000000
		Błąd standardowy estymacji:	,776756496		
Wyr. wolny	21,627063266	Błąd std.:	4,208867	t(78) =	5,1385 p = ,0000

DRAWING b*=-,903

(istotne b* są podświetlone na czerwono)

W obu przypadkach mamy wartość **p** dla testu istotności równania regresji bliska 0, a więc równanie jest istotne statystycznie. Włączymy do modelu zmienną **DRAWING**, ponieważ wartość **R** jest w tym przypadku większa niż z zmienną **CALCULUS**.

Podsumowanie regresji jest następujące:

N=80	Podsumowanie regresji zmiennej zależnej: GEOMETRY (School performance.s R= ,90324713 R^2= ,81585538 Popraw. R2= ,81349455 F(1,78)=345,58 p<0,0000 Błąd std. estymacji: ,77676					
	b*	Bł. std. z b*	b	Bł. std. z b	t(78)	p
W. wolny			21,62706	4,208867	5,13845	0,000002
DRAWING	0,903247	0,048588	0,78391	0,042169	18,58979	0,000000

Korelacje między współczynnikami regresji:

Zmienna	Korelacje (School performance.sta) Oznaczone wsp. korelacji są istotne z p < ,05000 N=80 (Braki danych usuwano przypadkami)	
	GEOMETRY	DRAWING
DRAWING	0,903247	1,000000
GEOMETRY	1,000000	0,903247

Korelacje cząstkowe i tolerancje:

Zmienna	Aktualnie w równaniu są zmienne ; DV: GEOMETRY (School performance.sta)						
	b* w	Cząstk. Korelac.	Semicz. Korelac.	Tolerancja	R-kwadr.	t(78)	p
DRAWING	0,903247	0,903247	0,903247	1,000000	0,00	18,58979	0,000000

Modelem regresji jest więc: **GEOMETRY = 21,627063266 + 0,78391 DRAWING**

Następnie wykonujemy weryfikację modelu.

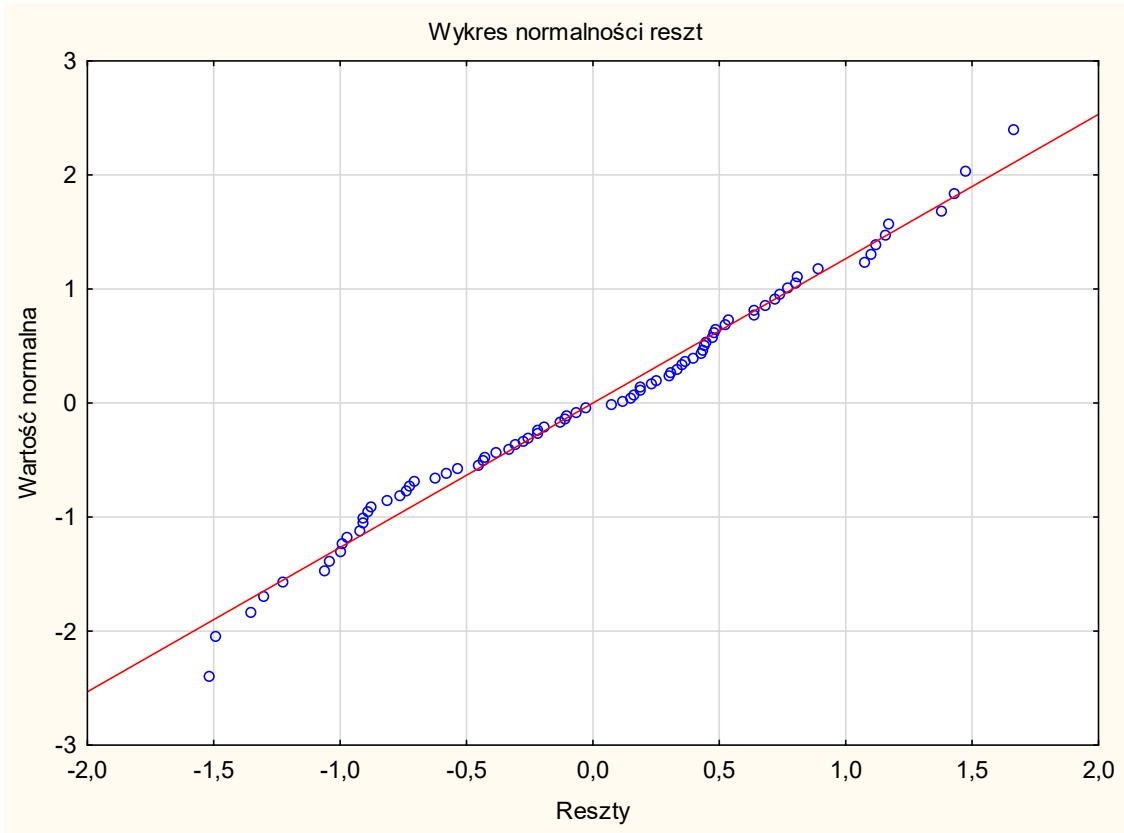
ANALIZA RESZT:

	Wartości przewidywane i reszty GEOMETRY								
	Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard. Przewid.	Standard. Reszta	Bl. std. W. przew.	Mahaln. Odlegl.	Usunięta Reszta	Cooka Odlegl.
1	98.654907	99.361320	-0.706413	-0.302235	-0.909440	0.090772	0.09134E	-0.716194	0.005805
2	98.700600	98.349281	0.351319	-0.925187	0.452285	0.119655	0.855971	0.359712	0.002502
3	98.399396	97.626350	0.773046	-1.370181	0.995226	0.147920	1.877395	0.802137	0.019337
4	98.032120	98.568375	-0.536255	-0.790325	-0.690377	0.110961	0.624615	-0.547425	0.005068
5	97.962402	99.482773	-1.520370	-0.227481	-1.957332	0.089090	0.05174E	-1.540635	0.025876
6	98.981255	98.540947	0.440308	-0.807212	0.566854	0.111885	0.651590	0.449637	0.003476
7	94.023918	94.455406	-0.431488	-3.322015	-0.555500	0.303025	11.03578E	-0.508945	0.032669
8	99.409615	98.960350	0.449265	-0.549054	0.578385	0.099215	0.301460	0.456715	0.002820
9	100.327110	101.299950	-0.972835	0.891056	-1.252438	0.116644	0.793985	-0.995283	0.018512
10	99.013870	98.715866	0.298004	-0.699541	0.383652	0.106204	0.489355	0.303681	0.001429
11	102.357625	102.740257	-0.382632	1.777615	-0.492595	0.177975	3.15992E	-0.403830	0.007095
12	98.469650	98.222214	0.247437	-1.003404	0.318551	0.123415	1.006820	0.253845	0.001348
13	97.688625	97.292603	0.396027	-1.575615	0.599847	0.162795	2.482566	0.414221	0.008246
14	102.656545	102.326057	0.330490	1.522663	0.425475	0.158900	2.318504	0.344925	0.004126
15	101.586311	100.695747	0.890564	0.519145	1.148516	0.097981	0.269514	0.904963	0.010799
16	102.202347	101.401863	0.800484	0.953790	-1.030547	0.120373	0.909715	0.820181	0.013388
17	101.535896	100.064903	1.470993	0.130835	1.893763	0.087594	0.017115	1.489940	0.023394
18	98.464444	97.989155	0.480286	-1.146860	0.618322	0.132617	1.315287	0.494705	0.005912
19	102.979915	101.603535	1.376381	1.077925	1.771955	0.128125	1.161925	1.414877	0.045137
20	99.449928	100.679645	-1.229721	0.509237	-1.583145	0.097583	0.259322	-1.249440	0.020418
21	100.606705	101.062813	-0.456106	0.745091	-0.587196	0.108544	0.555160	-0.465192	0.003502
22	99.961823	99.239635	0.722185	-0.377135	0.929742	0.092885	0.142233	0.732661	0.006361
23	100.382904	100.492411	-0.109512	0.393992	-0.140987	0.093421	0.155225	-0.111120	0.000148
24	100.528023	100.860100	-0.332077	0.620313	-0.427516	0.102375	0.384785	-0.337947	0.001644
25	95.838936	97.332001	-1.493065	-1.551363	-1.992175	0.161005	2.406725	-1.560094	0.086655
26	101.483414	100.407593	1.075821	0.341777	1.385017	0.091837	0.116811	1.091073	0.013790
27	100.228233	99.545517	0.682716	-0.188863	0.878932	0.088395	0.035665	0.691675	0.005135
28	100.361580	99.241600	1.119980	-0.375934	1.441867	0.092851	0.141327	1.136215	0.015287
29	100.580643	99.774117	0.806526	-0.048145	1.038326	0.086945	0.002315	0.816760	0.006927
30	97.530182	98.346153	-0.815971	-0.927115	-1.050485	0.118771	0.859543	-0.835505	0.013525
31	98.970093	100.322445	-1.352355	0.289370	-1.741030	0.090451	0.083735	-1.370945	0.021120
32	100.295517	99.139641	1.155875	-0.438693	1.488081	0.094930	0.192451	1.173402	0.017042
33	97.965736	97.815788	0.149948	-1.253575	0.193044	0.139795	1.571451	0.154965	0.000645
34	97.782440	98.500653	-0.724213	-0.826323	-0.932355	0.113055	0.686120	-0.739887	0.009611
35	101.963915	101.890395	0.073517	1.254495	0.094945	0.139952	1.573764	0.075980	0.000155
36	99.911305	99.680977	0.230331	-0.105475	0.296530	0.087532	0.011125	0.233280	0.000570
37	101.348545	100.810225	0.538322	0.589612	0.693035	0.100980	0.347643	0.547577	0.004199
38	99.652252	99.873940	-0.221687	0.013293	-0.285401	0.086852	0.000177	-0.224494	0.000522
39	97.931564	98.188904	-0.257335	-1.023907	-0.331300	0.124695	1.048385	-0.264147	0.001490
40	100.306035	100.615395	-0.309357	0.469686	-0.398267	0.096055	0.220605	-0.314161	0.001251
41	100.264153	99.830395	0.433754	-0.013505	0.558417	0.086852	0.000182	0.439245	0.001999
42	100.621785	99.522583	1.099205	-0.202977	1.415122	0.088637	0.041200	1.113707	0.013385
43	100.728531	101.718055	-0.989525	1.148415	-1.273915	0.132720	1.318855	-1.019282	0.025135
44	98.993256	99.901585	-0.908333	0.030312	-1.169392	0.086884	0.000915	-0.919842	0.008773
45	99.494598	99.714965	-0.220367	-0.084557	-0.283702	0.087155	0.007150	-0.223177	0.000520
46	100.306503	99.669884	0.636620	-0.112305	0.819587	0.087397	0.012613	0.644782	0.004362
47	97.901291	97.783345	0.117945	-1.273540	0.151940	0.141170	1.621904	0.121972	0.000407
48	100.129143	99.491095	0.638054	-0.222362	0.821434	0.088992	0.049445	0.646540	0.004547
49	99.771614	100.351788	-0.580154	0.307415	-0.746894	0.090905	0.094504	-0.588211	0.003927
50	100.226227	99.744115	0.482105	-0.066615	0.620665	0.087035	0.004435	0.488235	0.002480
51	102.914673	103.107803	-0.193130	2.003860	-0.248637	0.195472	4.015454	-0.206185	0.002231
52	99.820675	99.886612	-0.065933	0.021094	-0.084883	0.086864	0.000445	-0.066765	0.000046
53	99.301956	99.929265	-0.627312	0.047351	-0.807604	0.086943	0.002242	-0.635271	0.004190
54	103.304207	101.877265	1.426941	1.246417	1.837051	0.139305	1.553555	1.474364	0.057942
55	99.452942	98.927425	0.525513	-0.569315	0.676548	0.100087	0.324123	0.534385	0.003929
56	101.572731	102.310852	-0.738121	1.513305	-0.950261	0.155215	2.290094	-0.770070	0.020389
57	104.196355	103.831105	0.365242	2.449075	0.470214	0.230975	5.997987	0.400671	0.011764
58	98.307050	99.195053	-0.888023	-0.405951	-1.144532	0.093745	0.183175	-0.902155	0.008824
59	101.867813	101.899391	-0.031578	1.260034	-0.040654	0.140241	1.587685	-0.032642	0.000029
60	99.993195	100.915465	-0.922272	0.654393	-1.187337	0.103983	0.428230	-0.939101	0.013097
61	98.162033	99.157825	-0.995795	-0.427495	-1.281993	0.094535	0.182754	-1.010765	0.012542

Średnia otrzymanych reszt wynosi: **-9,53674315073982E-8**

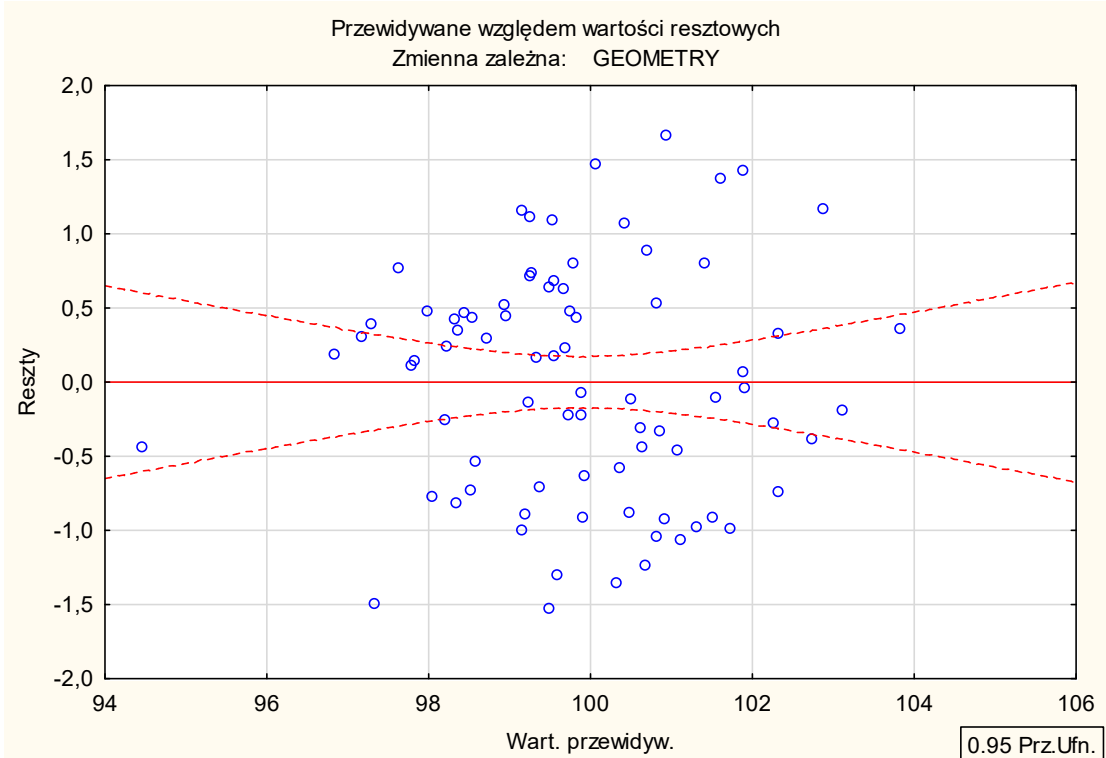
Średnia bardzo bliska zeru świadczy o nieobciążoności reszt, dowodzi to temu, że model regresji jest użyteczny.

Wykres normalności reszt:

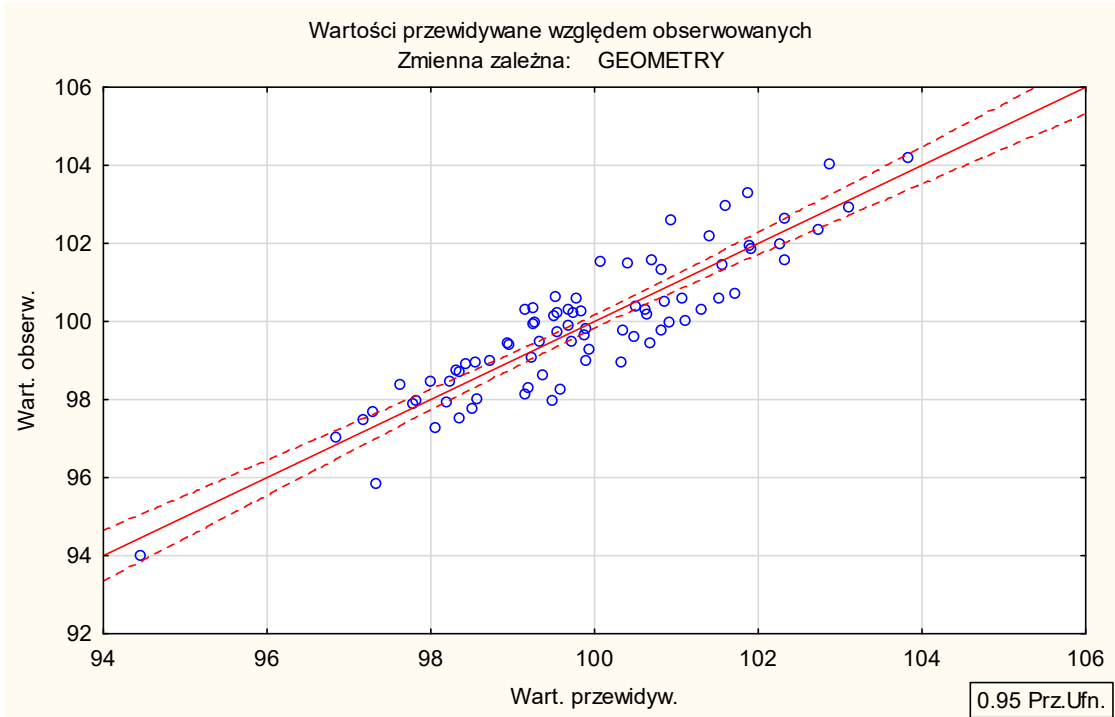


Otrzymany wykres potwierdza warunek normalności reszt.

Wykres rozrzutu reszt względem wartości przewidywanych:



Wykres rozrzutu reszt wartości przewidywanych względem obserwowanych:



Punkty empiryczne są losowo rozmieszczone wokół prostej regresji, co świadczy o poprawności zbudowanego modelu regresji.

Wartości odstające prezentują się następująco:

Reszty standaryz.							Reszty standaryz.: GEOMETRY (School performance.sta)				
Przyp. -5. -4. -3. ±2. 3. 4. 5.							Odstające				
							Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard Przewid.	Standard Reszta
63 . . . *							102,6003	100,9389	1,661385	0,668829	2,138874
Minimum . . . *							102,6003	100,9389	1,661385	0,668829	2,138874
Maksim. . . . *							102,6003	100,9389	1,661385	0,668829	2,138874
Średnia . . . *							102,6003	100,9389	1,661385	0,668829	2,138874
Mediana . . . *							102,6003	100,9389	1,661385	0,668829	2,138874

Obliczmy jeszcze współczynniki korelacji:

Współczynnik R Spearmana:

Zmienna	Korelacja porządku rang Spearmana (School performance.sta BD usuwane parami	
	Oznaczone wsp. korelacji są istotne z p <,05000	
GEOMETRY	GEOMETRY	DRAWING
GEOMETRY	1,000000	0,872269
DRAWING	0,872269	1,000000

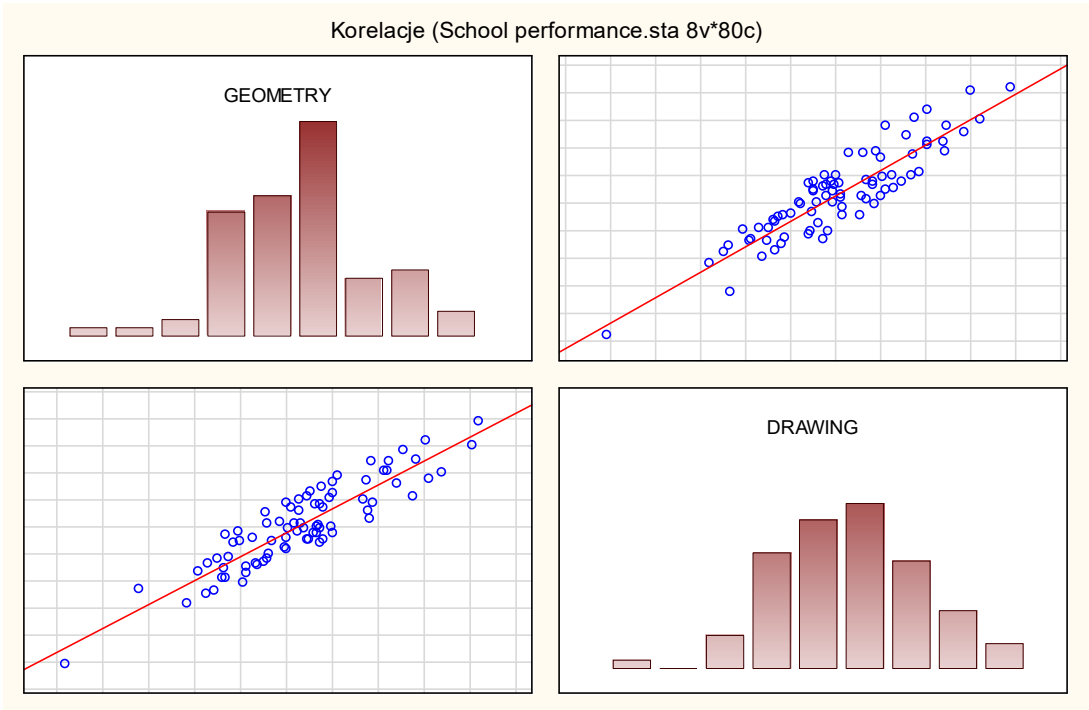
Uzyskana wartość **|R| = 0,8722** świadczy o bardzo silnej korelacji między zmiennymi.

Współczynnik Tau Kendalla:

	Korelacja tau Kendalla (School performance.sta) BD usuwane parami Oznaczone wsp. korelacji są istotne z p <,05000	
Zmienna	GEOMETRY	DRAWING
GEOMETRY	1,000000	0,698101
DRAWING	0,698101	1,000000

Otrzymana wartość **0,698** potwierdza poprawność wykonanej regresji.

Macierzowy wykres rozrzutu zmiennych:



Powyższy wykres potwierdza otrzymane wnioski, że zbudowany model regresji jest istotny statystycznie. Ma to również uzasadnienie w rzeczywistości: umiejętności geometryczne wiążą się silnie z umiejętnościami rysowania.

W taki sam sposób spróbujemy znaleźć model regresji dla zmiennej objaśnianej **HISTORY**.

	Korelacje (School performance.sta) Oznaczone wsp. korelacji są istotne z p <,05000 N=80 (Braki danych usuwano przypadkami)							
Zmienna	GEOMETRY	READING	GRAMMAR	DRAWING	CALCULUS	HISTORY	WRITING	SPELLING
GEOMETRY	1,000000	-0,035582	-0,014342	0,903247	0,809318	-0,020458	-0,031098	0,018067
READING	-0,035582	1,000000	0,943131	-0,034940	-0,009269	0,950956	0,942327	0,860010
GRAMMAR	-0,014342	0,943131	1,000000	-0,025917	0,005717	0,924746	0,901965	0,837787
DRAWING	0,903247	-0,034940	-0,025917	1,000000	0,814418	-0,032686	-0,033889	0,016422
CALCULUS	0,809318	-0,009269	0,005717	0,814418	1,000000	-0,017948	0,022055	0,040587
HISTORY	-0,020458	0,950956	0,924746	-0,032686	-0,017948	1,000000	0,909612	0,848156
WRITING	-0,031098	0,942327	0,901965	-0,033889	0,022055	0,909612	1,000000	0,882936
SPELLING	0,018067	0,860010	0,837787	0,016422	0,040587	0,848156	0,882936	1,000000

Badając korelacji widzimy, że zmienne **READING**, **GRAMMAR**, **WRITING** oraz **SPELLING** są silnie z sobą skorelowane, więc nie umieścimy ich wszystkich w modelu regresji.

Wyniki regresji wielorakiej dla wszystkich zmiennych z arkusza prezentują się następująco:

```
Zmn. zależ.HISTORY      Wielor. R =      ,95665999      F = 111,0057
                        R^2=      ,91519835      df =    7,72
Liczba przyp. 80        Popraw. R^2=      ,90695374      p = 0,000000
                        Błąd standardowy estymacji: 1,064792989
Wyr. wolny - ,522846567  Błąd std.: 7,779283  t( 72) = - ,0672  p =
,9466
```

GEOMETRY b*=-,074	READING b*=-,620	GRAMMAR b*=-,223
DRAWING b*=-,02	CALCULUS b*=-,06	WRITING b*=-,057
SPELLING b*=-,079		

(istotne b* są podświetlone na czerwono)

	Podsumowanie regresji zmiennej zależnej: HISTORY (School performance.sta) R= ,95665999 R^2= ,91519835 Popraw. R2= ,90695374 F(7,72)=111,01 p<0,0000 Błąd std. estymacji: 1,0648					
N=80	b*	Bł. std. z b*	b	Bł. std. z b	t(72)	p
W. wolny			-0,522847	7,779283	-0,067210	0,946601
GEOMETRY	0,073934	0,084100	0,143490	0,163219	0,879128	0,382257
READING	0,620353	0,135608	0,657033	0,143626	4,574621	0,000019
GRAMMAR	0,222566	0,105116	0,221188	0,104465	2,117346	0,037685
DRAWING	-0,023863	0,085283	-0,040193	0,143647	-0,279805	0,780430
CALCULUS	-0,058343	0,062870	-0,087121	0,093881	-0,927997	0,356510
WRITING	0,057305	0,115185	0,059224	0,119041	0,497506	0,620348
SPELLING	0,079011	0,075283	0,073404	0,069940	1,049519	0,297449

Z otrzymanych danych wynika, że uwzględnienie zmiennych **GEOMETRY**, **DRAWING**, **CALCULUS**, **WRITING** oraz **SPELLING** jest zbyteczne jako, że ich udział w wariancji zmiennej **HISTORY** jest niewielki, a wyniki testu dla odpowiadających im współczynników równania regresji świadczą o ich nieistotności.

Korelacje cząstkowe prezentują się następująco:

	Aktualnie w równaniu są zmienne ; DV: HISTORY (School performance.sta)						
Zmienna	b* w	Cząstk. Korelac.	Semicz. Korelac.	Tolerancja	R-kwadr.	t(72)	p
GEOMETRY	0,073934	0,103055	0,030171	0,166527	0,833473	0,879128	0,382257
READING	0,620353	0,474552	0,156997	0,064048	0,935952	4,574621	0,000019
GRAMMAR	0,222566	0,242108	0,072665	0,106595	0,893405	2,117346	0,037685
DRAWING	-0,023863	-0,032957	-0,009603	0,161937	0,838063	-0,279805	0,780430
CALCULUS	-0,058343	-0,108717	-0,031848	0,297976	0,702024	-0,927997	0,356510
WRITING	0,057305	0,058531	0,017074	0,088773	0,911227	0,497506	0,620348
SPELLING	0,079011	0,122752	0,036015	0,207818	0,792182	1,049519	0,297449

Na podstawie otrzymanych wyników wybieramy te zmienne, których współczynniki regresji są istotne, a ich wartość tolerancji jest w miarę duża, są to zmienne: **READING** oraz **GRAMMAR**.

Wyniki regresji wielorakiej dla wybranych zmiennych prezentują się następująco:

```
Zmn. zależ.HISTORY      Wielor. R =      ,95464432      F = 395,7715
                        R^2=      ,91134578      df =    2,77
Liczba przyp. 80        Popraw. R^2=      ,90904307      p = 0,000000
                        Błąd standardowy estymacji: 1,052770299
Wyr. wolny 1,591653527  Błąd std.: 3,587851  t( 77) = ,44362  p =
,6586
```

READING b*=-,713	GRAMMAR b*=-,252
------------------	------------------

(istotne b* są podświetlone na czerwono)

Eliminacja zmiennych poprawiła model, wartość błędu standardowego dla wyrazu wolnego się zmniejszyła.

Pamiętamy, że zmienne **READING** oraz **GRAMMAR** były ze sobą bardzo silnie skorelowane (wartość korelacji linionej Pearsona = 0,9431), więc jedną z nich należy usunąć z modelu. Przeanalizujmy, która zmienna jest istotniejsza.

Wyniki bez zmiennej **READING**:

```
Zmn. zależ.HISTORY      Wielor. R =      ,92474596      F = 460,5070
                        R^2=      ,85515508      df =    1,78
Liczba przyp. 80        Popraw. R^2=      ,85329810      p = 0,000000
                        Błąd standardowy estymacji: 1,337006744
Wyr. wolny 10,313760282  Błąd std.: 4,271750  t( 78) = 2,4144  p =
,0181
```

GRAMMAR b*=-,925

(istotne b* są podświetlone na czerwono)

Wyniki bez zmiennej **GRAMMAR**:

```
Zmn. zależ.HISTORY      Wielor. R =      ,95095561      F = 737,1882
                        R^2=      ,90431658      df =    1,78
Liczba przyp. 80        Popraw. R^2=      ,90308986      p = 0,000000
                        Błąd standardowy estymacji: 1,086676619
Wyr. wolny 1,443253062  Błąd std.: 3,702885  t( 78) = ,38976  p =
,6978
```

READING b*=-,951

(istotne b* są podświetlone na czerwono)

W obu przypadkach mamy wartość **p** dla testu istotności równania regresji bliska 0, a więc równanie jest istotne statystycznie. Włączymy do modelu zmienną **READING**, ponieważ wartość **R** jest w tym przypadku większa niż z zmienną **GRAMMAR**.

Podsumowanie regresji jest następujące:

Podsumowanie regresji zmiennej zależnej: HISTORY (School performance.sta) R= ,95095561 R^2= ,90431658 Popraw. R2= ,90308986 F(1,78)=737,19 p<0,0000 Błąd std. estymacji: 1,0867						
N=80	b*	Bł. std. z b*	b	Bł. std. z b	t(78)	p
W. wolny			1,443253	3,702885	0,38976	0,697773
READING	0,950956	0,035024	1,007184	0,037095	27,15121	0,000000

Korelacje między współczynnikami regresji:

Korelacje (School performance.sta) Oznaczone wsp. korelacji są istotne z p < ,05000 N=80 (Braki danych usuwano przypadkami)		
Zmienna	READING	HISTORY
HISTORY	0,950956	1,000000
READING	1,000000	0,950956

Korelacje cząstkowe i tolerancje:

Zmienna	Aktualnie w równaniu są zmienne ; DV: HISTORY (School performance.sta)						
	b* w	Cząstk. Korelac.	Semicz. Korelac.	Tolerancja	R-kwadr.	t(78)	p
READING	0,950956	0,950956	0,950956	1,000000	0,00	27,15121	0,00

Modelem regresji jest więc: **HISTORY = 1,4432 + 1,0071 READING.**

Następnie wykonujemy weryfikację modelu.

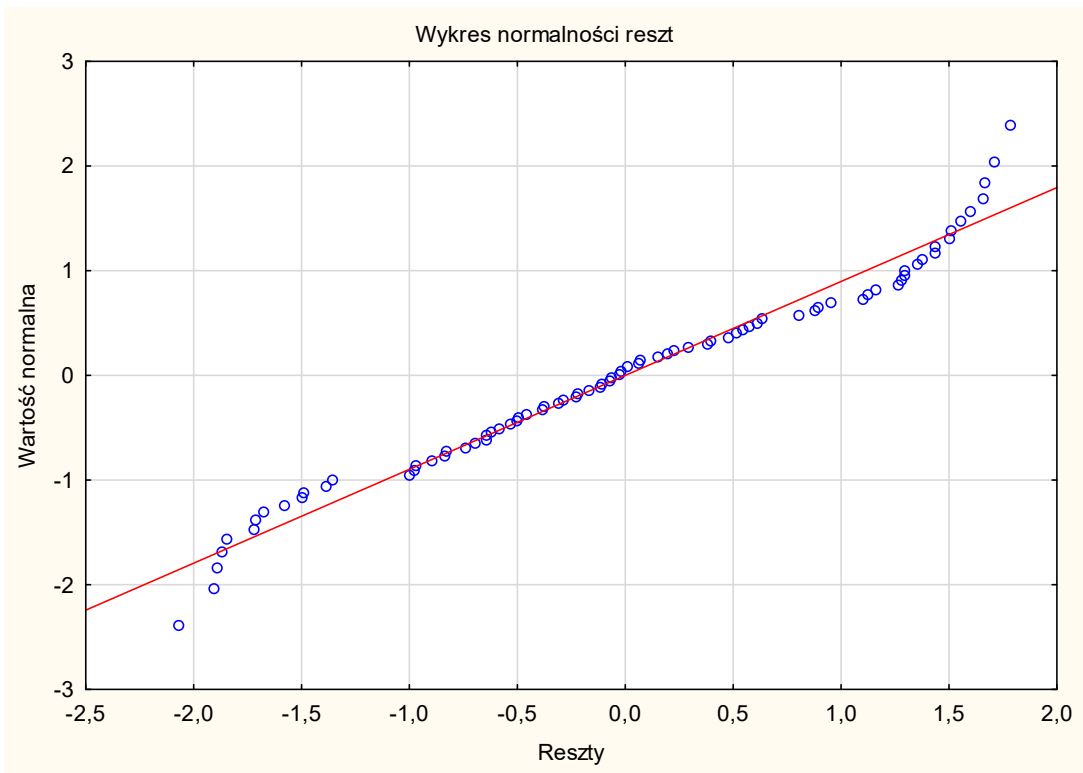
ANALIZA RESZT:

	Wartości przewidywane i reszty HISTORY									
	Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard. Przewid.	Standard. Reszta	Bł. std. W. przew.	Mahaln. Odlegl.	Usunięte Reszta	Cooka Odlegl.	
1	99,986923	100,633614	-0,646690	-0,389607	-0,595108	0,130498	0,151794	-0,656155	0,002625	
2	103,135422	102,558968	0,576553	0,190372	0,530566	0,123703	0,036241	0,584123	0,001872	
3	101,657375	99,944836	1,712540	-0,597100	1,575942	0,141740	0,356525	1,742175	0,021864	
4	102,063446	102,369972	-0,306526	0,133467	-0,282077	0,122585	0,017813	-0,310477	0,000515	
5	101,457085	101,302986	0,154098	-0,187961	0,141807	0,123646	0,035325	0,156120	0,000134	
6	102,773781	104,842495	-2,068718	0,878311	-1,903711	0,162146	0,771430	-2,115827	0,042204	
7	101,825745	100,271912	1,553833	-0,498565	1,429895	0,135926	0,248571	1,578531	0,016506	
8	107,434486	109,152222	-1,717735	2,176600	-1,580724	0,292536	4,737612	-1,851945	0,105240	
9	100,003996	100,376755	-0,372757	-0,466986	-0,343025	0,134241	0,218076	-0,378534	0,000926	
10	101,214226	101,441002	-0,226776	-0,146384	-0,208688	0,122805	0,021428	-0,229710	0,000285	
11	102,090721	101,706093	0,384626	-0,066526	0,353945	0,121766	0,004426	0,389515	0,000807	
12	99,464696	101,368011	-1,903313	-0,168372	-1,751495	0,123226	0,028345	-1,928106	0,020241	
13	105,933914	105,961853	-0,027935	1,215513	-0,025710	0,191952	1,477471	-0,028836	0,000011	
14	97,203323	99,075142	-1,871815	-0,859095	-1,722517	0,160602	0,738044	-1,913616	0,033867	
15	95,989605	96,488190	-0,498581	-1,638407	-0,458813	0,234275	2,684375	-0,522884	0,005381	
16	100,470947	99,593075	0,877865	-0,703065	0,807847	0,148827	0,494304	0,894655	0,006357	
17	103,421387	102,620041	0,801346	0,208795	0,737428	0,124147	0,043597	0,811943	0,003643	
18	103,924744	102,971355	0,953384	0,314633	0,877340	0,127435	0,098994	0,966675	0,005442	
19	102,579483	101,282990	1,296494	-0,193984	1,193081	0,123787	0,037630	1,313536	0,009480	
20	106,889740	105,290161	1,599575	1,013167	1,471992	0,173507	1,026507	1,641425	0,029083	
21	105,343323	105,844437	-0,501114	1,180141	-0,461144	0,188624	1,392733	-0,516681	0,003406	
22	100,277195	101,772614	-1,495415	-0,046486	-1,376136	0,121627	0,002161	-1,514386	0,012165	
23	102,190582	100,811501	1,379082	-0,336015	1,269082	0,128252	0,112905	1,398563	0,011536	
24	105,939163	104,152260	1,786903	0,670376	1,644375	0,146555	0,449404	1,820007	0,025510	
25	101,324181	103,211586	-1,887405	0,387001	-1,736860	0,130382	0,149770	-1,914973	0,022353	
26	98,814804	98,618725	0,196075	-0,996598	0,180436	0,172066	0,993185	0,201116	0,000425	
27	106,088310	106,674202	-0,585892	1,430105	-0,539155	0,212913	2,045205	-0,609281	0,006034	
28	108,778542	108,842163	-0,063622	2,083200	-0,058547	0,282187	4,339721	-0,068222	0,000133	
29	111,305426	110,022690	1,282736	2,438831	1,180423	0,321975	5,947897	1,406187	0,073502	
30	96,893890	96,665443	0,228447	-1,585005	0,210225	0,228721	2,512255	0,239036	0,001072	
31	103,457970	102,018646	1,439323	0,027631	1,324515	0,121541	0,000763	1,457557	0,011253	
32	102,096382	103,451302	-1,354915	0,459215	-1,246847	0,133835	0,210875	-1,375785	0,012157	
33	100,357185	101,330383	-0,973196	-0,179707	-0,895573	0,123465	0,032294	-0,985925	0,005313	
34	99,949341	100,775391	-0,826055	-0,346897	-0,760162	0,128684	0,120337	-0,837795	0,004168	
35	106,495483	105,373932	1,121552	1,038403	1,032093	0,175723	1,078280	1,151067	0,014685	
36	104,195526	104,655075	-0,459545	0,821846	-0,422894	0,157661	0,675436	-0,469430	0,001964	
37	99,094551	100,582703	-1,488152	-0,404945	-1,369452	0,131194	0,163981	-1,510163	0,014075	
38	101,574081	102,197372	-0,623291	0,081472	-0,573575	0,121902	0,006636	-0,631234	0,002123	
39	100,118370	99,227394	0,890976	-0,813228	0,819905	0,156991	0,661340	0,909965	0,007315	
40	100,969246	102,644844	-1,675596	0,216271	-1,541947	0,124335	0,046773	-1,697826	0,015980	
41	97,954536	98,694824	-0,740288	-0,973665	-0,681240	0,170093	0,948023	-0,758881	0,005974	
42	109,676743	108,324636	1,352104	1,927297	1,244256	0,265110	3,714473	1,437672	0,052085	
43	105,775047	105,708282	0,066765	1,139125	0,061435	0,184816	1,297606	0,068754	0,000055	
44	96,198524	97,175171	-0,976646	-1,431455	-0,898746	0,213046	2,049067	-1,015687	0,016790	
45	103,192802	103,416702	-0,223900	0,448791	-0,206041	0,133310	0,201414	-0,227321	0,000325	
46	103,937363	104,319313	-0,381950	0,720702	-0,351485	0,150085	0,519412	-0,389376	0,001225	
47	99,470917	100,189350	-0,698433	-0,529467	-0,642724	0,137663	0,280335	-0,709825	0,003424	
48	96,773735	97,773567	-0,999832	-1,251191	-0,920083	0,195345	1,565475	-1,033222	0,014605	
49	100,641255	99,347313	1,293945	-0,777102	1,190736	0,154232	0,603887	1,320547	0,014874	
50	101,561523	101,081490	0,480034	-0,254685	0,441745	0,125421	0,064865	0,486515	0,001335	
51	100,917326	100,857834	0,059494	-0,322063	0,054745	0,127716	0,103724	0,060327	0,000021	
52	102,061447	100,795906	1,265541	-0,340715	1,164598	0,128437	0,116085	1,283471	0,009744	
53	104,735855	103,298950	1,436905	0,413320	1,322293	0,131584	0,170833	1,458287	0,013203	
54	100,171085	99,569454	0,601631	0,713202	0,560416	0,140482	0,602734	0,600405	0,003081	

Średnia otrzymanych reszt wynosi: **-2,86102295188329E-7.**

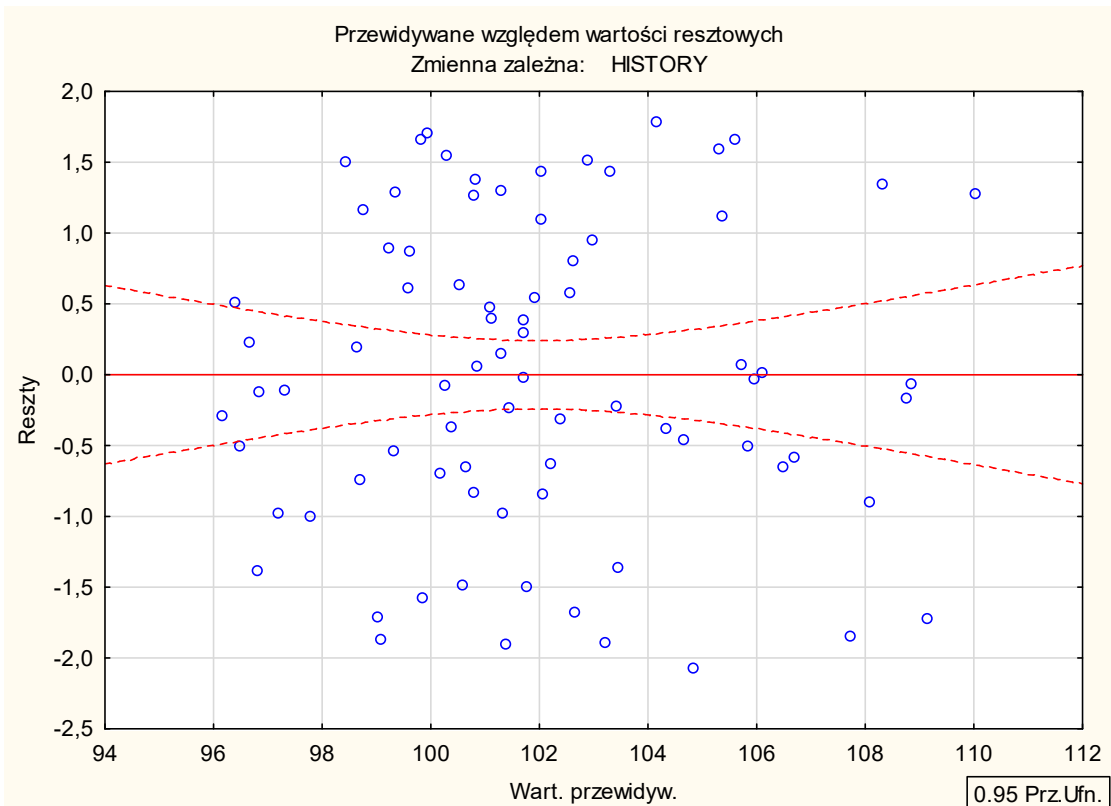
Średnia bardzo bliska zeru świadczy o nieobciążoności reszt, dowodzi to temu, że model regresji jest użyteczny.

Wykres normalności reszt:

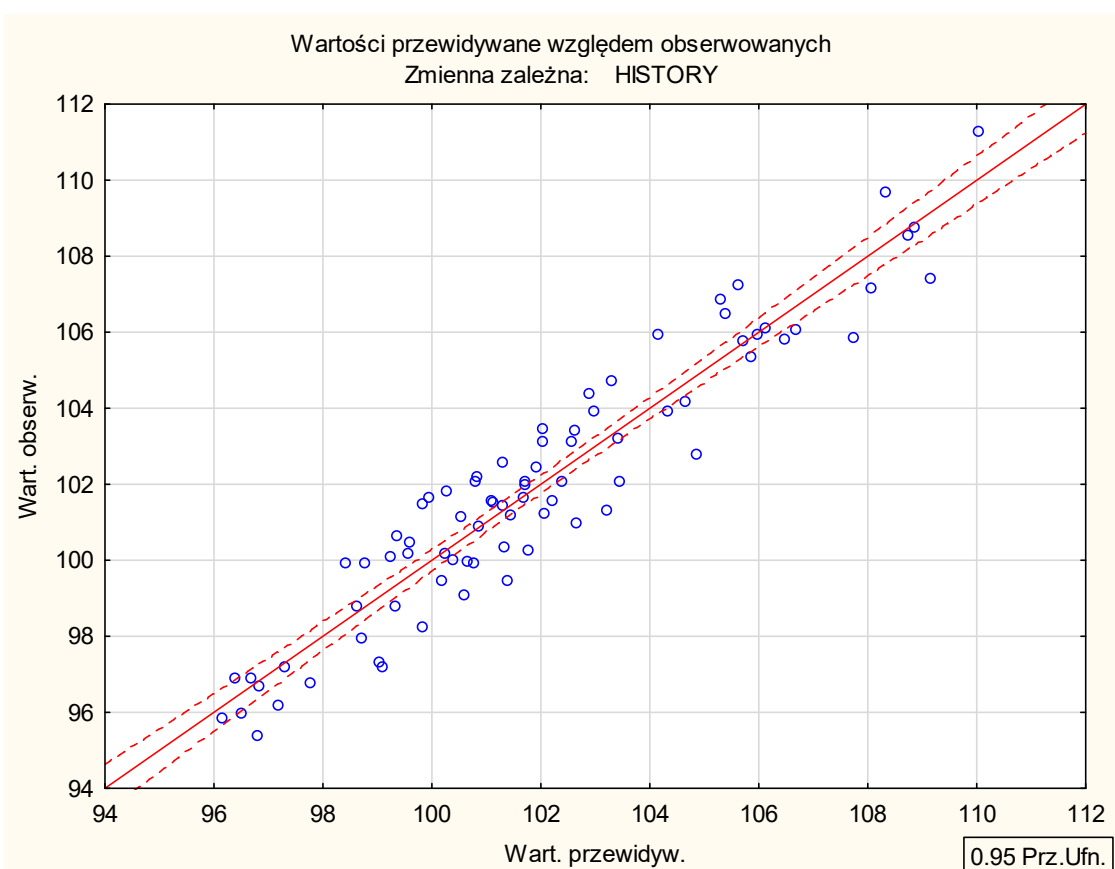


Otrzymany wykres potwierdza warunek normalności reszt.

Wykres rozrzutu reszt względem wartości przewidywanych:

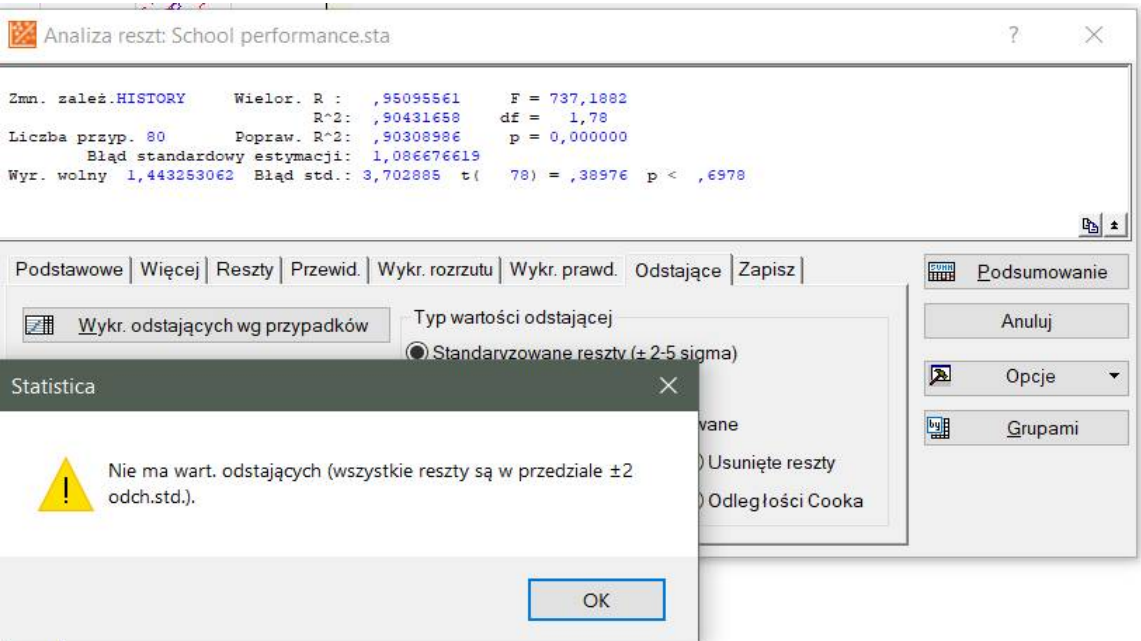


Wykres rozrzutu reszt wartości przewidywanych względem obserwowanych:



Punkty empiryczne są losowo rozmieszczone wokół prostej regresji, co świadczy o poprawności zbudowanego modelu regresji.

W analizowanym przypadku nie wykryto wartości odstających:



Obliczmy jeszcze współczynniki korelacji:

Współczynnik R Spearmana:

	Korelacja porządku rang Spearmana (School performance.sta BD usuwane parami Oznaczone wsp. korelacji są istotne z p <,05000	
Zmienna	READING	HISTORY
READING	1,000000	0,929606
HISTORY	0,929606	1,000000

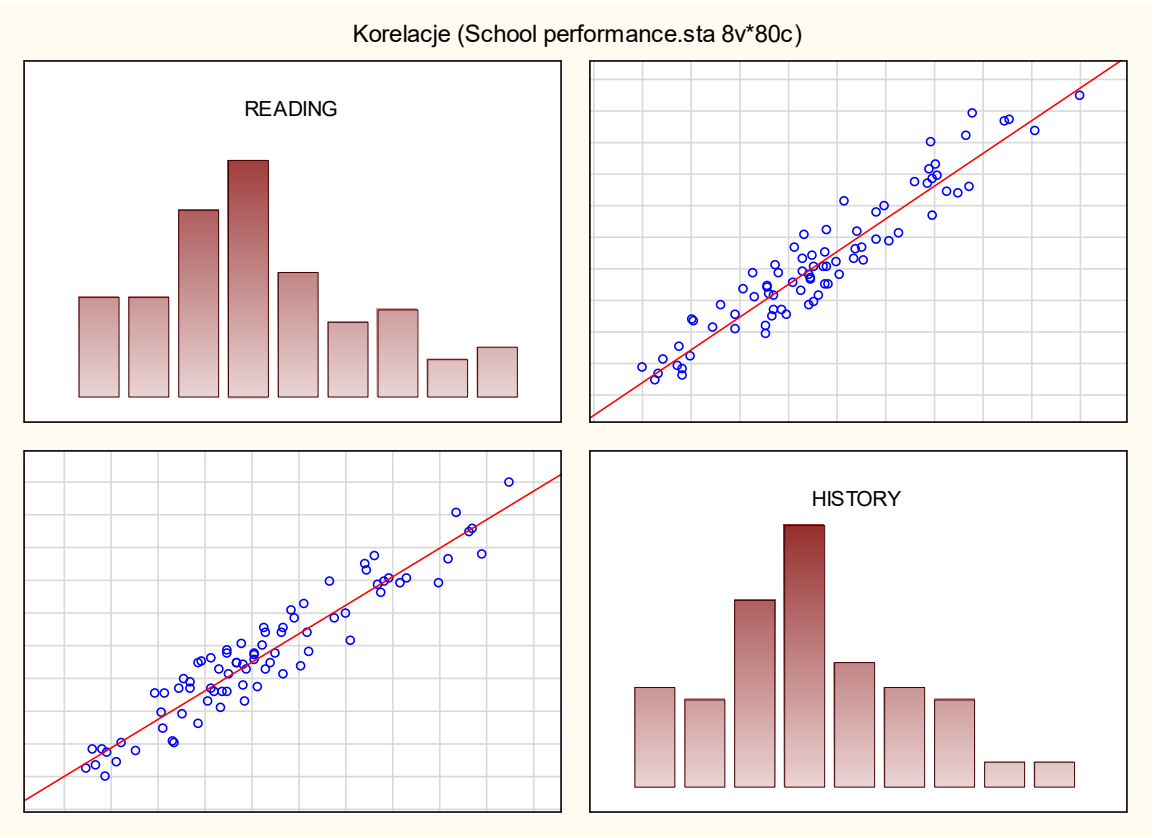
Uzyskana wartość $|R| = 0,9296$ świadczy o korelacji prawie pełnej między zmiennymi.

Współczynnik Tau Kendalla:

	Korelacja tau Kendalla (School performance.sta BD usuwane parami Oznaczone wsp. korelacji są istotne z p <,05000	
Zmienna	READING	HISTORY
READING	1,000000	0,781013
HISTORY	0,781013	1,000000

Otrzymana wartość $0,781$ potwierdza poprawność wykonanej regresji.

Macierzowy wykres rozrzutu zmiennych:



Powyższy wykres potwierdza otrzymane wnioski, że zbudowany model regresji jest istotny statystycznie. Ma to również uzasadnienie w rzeczywistości: zdolności historyczne wiążą się silnie z zdolnościami czytania (czytaniem).