

Przemysław Rompalski\*, Leokadia Róg\*

## WYKORZYSTANIE PAKIETU KOMPUTEROWEGO STATISTICA DO OBRÓBK I ANALIZY DANYCH

### Streszczenie

Do ważniejszych parametrów służących do oceny jakości węgla kamiennych należy zaliczyć takie, na podstawie których można określić jego typ. Spośród nich, decydujące znaczenie mają: zawartość części lotnych, zdolność spiekania – *RI*, wskaźnik wolnego wydymania – *SI* oraz właściwości dylatometryczne. Na wszystkie te parametry duży wpływ wywierają czynniki, wynikające z obecności w węglu tzw. balastu, do którego zalicza się popiół i wodę.

Przy dużym zróżnicowaniu parametrów i przy równoczesnych wzajemnych ich powiązaniach, do wyciążenia wiarygodnych wniosków dotyczących tych powiązań, jest konieczne użycie odpowiednich technik statystycznych. Jednym z takich współczesnych narzędzi statystycznych, pomocnych w tego typu badaniach, jest pakiet komputerowy STATISTICA. W pakiecie tym są wykorzystywane różnego rodzaju testy statystyczne, które służą do scharakteryzowania właściwości badanego węgla, na podstawie opisu zależności między analizowanymi danymi. W badaniach wykorzystano dane zgromadzone w Banku Danych o Jakości Polskich Węgla Kamiennych, prowadzonym od wielu lat w Zakładzie Oceny Jakości Paliw Stałych Głównego Instytutu Górnicztwa; były to węgle energetyczne produkowane w wybranych pięciu kopalniach węgla kamiennego z rejonu Górnosląskiego Zagłębia Węglowego.

Pakiet komputerowy STATISTICA wykorzystano do przeprowadzenia następujących testów statystycznych: testu *t*-Studenta, regresji wielorakiej, analizy wariancji, analizy składowych głównych. Aby stwierdzić czy średnia zawartość popiołu w węglach wyprodukowanych w wytypowanych kopalniach w 2009 r. różni się od średniej zawartości popiołu w węglach produkowanych w latach 2005–2008, wykonano test *t*-Studenta. Następnie, przeprowadzono regresję wieloraką w celu stwierdzenia, czy istnieją zależności między trzema parametrami: zdolnością spiekania węgla według Rogi, zawartością pierwiastka C oraz zawartością części lotnych. Wykazała ona, że zależności istniejące między każdą z tych zmienionych są niewielkie.

Analiza wariancji posłużyła do oszacowania zależności między zawartością popiołu w węglach z pięciu różnych kopalń. Na tej podstawie stwierdzono brak równości wariancji w poszczególnych grupach, co świadczyło o braku zależności między zawartością popiołu w analizowanych kopalniach. Analiza składowych głównych pozwoliła na stwierdzenie, które z badanych parametrów były istotne statystycznie. W analizowanym modelu PCA tylko trzy składowe główne były istotne ( $W^a$ ,  $A^a$  i  $V^{daf}$ ), z czego trzecia składowa miała istotność nieznaczną.

### Use of the computer package STATISTICA for data handling and analysis

#### Abstract

Among more important parameters used for hard coal quality assessment should be ranked such ones, on the basis of which the type of coal can be determined. From among these parameters the deciding significance have: volatile matter content, caking properties – *RI*, swelling index – *SI*, and dilatometric properties. On all these parameters high impact exert factors, resulting from the presence in coal of the so-called ballast, among which ash and water are counted.

Under great differentiation of parameters and their simultaneous mutual connections for drawing reliable conclusions necessary is the use of suitable statistical techniques. One of such contemporary

\* Główny Instytut Górnicztwa

statistical tools, helpful in investigations of this type, is the computer package STATISTICA. In this package different kinds of statistical tests are used for the characterisation of properties of the tested coal, on the basis of relationships between the analysed data. In investigations have been used data collected in the Data Bank about the Quality of Polish Hard Coals, conducted since many years in the Department of Solid Fuels Quality Assessment of the Central Mining Institute. Those were power coals produced in five selected hard coal mines from the Upper Silesian Coal Basin region.

The computer package STATISTICA was used to carry out the following statistical tests: *t*-Student test, multiple regression, analysis of variance, analysis of main components. In order to ascertain if the average ash content in coals produced in the selected mines in 2009 differs from the average ash content in coals produced in the period 2005–2008, the *t*-Student test was carried out. Then the multiple regression was carried out in order to ascertain if there exist relationships between three parameters: caking properties according to Roga, content of element C and volatile matter content. It has pointed out that the relationships existing between each of the variables are small.

The analysis of variance was used for the estimation of relationships between the ash content in coals from five different mines. On this basis the lack of variance equality in individual groups was ascertained, what argued for the lack of relationships between coal contents in the analysed mines. The analysis of main components allowed to ascertain, which from the investigated parameters were statistically essential. In the analysed PCA model only three main components were essential ( $W^a$ ,  $A^a$ , and  $V^{daf}$ ), of which the third component had inconsiderable significance.

## 1. WPROWADZENIE

Dużą rolę w ocenie węgla kamiennych odgrywają podstawowe pomiary jakościowe. Do ważniejszych parametrów oceny jakości węgla kamiennych należy zaliczyć takie, na podstawie których można określić jego typ. Spośród nich decydujące znaczenie mają: zawartość części lotnych, zdolność spiekania – *RI*, wskaźnik wolnego wydymania – *SI* oraz właściwości dylatometryczne. Parametry te mają istotny wpływ na zachowanie się węgla w różnych procesach technologicznych. Na przykład: mała zawartość części lotnych w węglu, przy równoczesnych dużych wartościach pozostałych wymienionych parametrów, znacznie obniża jego właściwości energetyczne. Na wszystkie te parametry duży wpływ wywierają czynniki, wynikające z obecności w węglu tzw. balastu, do którego zalicza się popiół i wodę (Jasieńko 1995).

Przy dużym zróżnicowaniu parametrów i przy równoczesnych wzajemnych ich powiązaniach, do wyciągnięcia wiarygodnych wniosków dotyczących tych powiązań, jest konieczne użycie odpowiednich technik statystycznych. Techniki te pozwalają na określenie różnic występujących między parametrami jakościowymi lub też szacowanie siły i rodzaju związków występujących między nimi (Zastosowanie statystyki... 2008; Praktyczne wykorzystanie... 2008).

Jednym z takich współczesnych narzędzi statystycznych, pomocnych w tego typu badaniach, jest pakiet komputerowy STATISTICA. W pakiecie tym są wykorzystywane różnego rodzaju testy statystyczne, które służą do scharakteryzowania właściwości badanego węgla, na podstawie opisu zależności między analizowanymi danymi. Mogą to być między innymi:

- test *t*-Studenta – służący do szacowania prawdziwej wartości parametru i ustalania przedziałów ufności, w granicach których leży prawdziwa wartość danej zmiennej,
- regresja wieloraka – pozwalająca na liczbowe określenie wpływu poszczególnych zmiennych objaśniających na zmienną objaśnianą,

- analiza wariancji – umożliwiająca poznawanie zależności między różnymi zjawiskami czy czynnikami,
- analiza składowych głównych – pozwalająca na zmniejszenie ilości wymiarów w wielowymiarowej bazie danych bez utraty „niesionych” przez dany wymiar informacji.

Jest to tylko kilka przykładów możliwości zastosowania pakietu komputerowego STATISTICA.

W artykule omówiono badania i analizy umożliwiające wykorzystanie powyższych testów do obróbki wyników analiz fizyczno-chemicznych, wykonanych dla węgla kamiennych.

## 2. PRZEDMIOT I METODYKA BADAŃ

Przedmiotem badań były węgle energetyczne produkowane w pięciu wybranych kopalniach węgla kamiennego w Górnośląskim Zagłębiu Węglowym, oznaczonych literami: A, B, C, D i E. W badaniach wykorzystano dane zgromadzone w Banku Danych o Jakości Polskich Węgla Kamiennych, prowadzonym od wielu lat w Zakładzie Oceny Jakości Paliw Stałych Głównego Instytutu Górnictwa.

W celu scharakteryzowania właściwości jakościowych i technologicznych węgla, wykorzystano ich analizy fizyczno-chemiczne. Do parametrów, które dobrze charakteryzują właściwości węgla kamiennego zaliczono: zawartość popiołu, wilgoci, części lotnych, pierwiastka C, ciepło spalania oraz takie wskaźniki, jak: zdolność spiekania (*RI*), wolne wydymanie (*SI*), dylatacja. Wszystkie oznaczenia wykonano według polskich norm, a mianowicie:

- zawartość wilgoci, popiołu i części lotnych według normy PN-G-04560:1998,
- zawartość pierwiastka C według normy PN-G-04571:1998,
- ciepło spalania według normy PN-G-04513:1981,
- zdolność spiekania według Rogi według normy PN-G-04518:1981,
- wskaźnik wolnego wydymania według normy PN-ISO 501:2007,
- wskaźniki dylatometryczne według normy PN-G-04517:1981.

Uzyskane wyniki badań zostały poddane analizie statystycznej z wykorzystaniem pakietu komputerowego STATISTICA.

W pierwszej kolejności wykonano test *t*-Studenta, następnie regresję wieloraką, analizę wariancji, analizę składowych głównych.

## 3. CEL BADAŃ I ANALIZ STATYSTYCZNYCH

Celem wykonanych badań i analiz statystycznych było określenie zależności między wybranymi parametrami, charakteryzującymi jakość węgla kamiennych energetycznych, produkowanych w wytypowanych kopalniach. Do parametrów tych, jak już podano, zaliczono: zawartość popiołu, wilgoci, pierwiastka C, części lotnych, ciepło spalania oraz wskaźniki: zdolność spiekania *RI*, wolne wydymanie *SI*, dylatację. Analiza statystyczna natomiast miała wykazać, które z parametrów w istotny sposób

wpływają na kształtowanie się wskaźnika zdolności spiekania badanych węgli i jak wybrane ich parametry jakościowe zmieniają się w czasie, co daje możliwość prognozowania jakości węgla.

Aby stwierdzić czy średnia zawartość popiołu w węglach uzyskanych z wytypowanych kopalń w 2009 r. różniła się od średniej zawartości popiołu w węglach z lat 2005–2008, wykonano test *t*-Studenta. Przeprowadzono następnie regresję wieloraką w celu stwierdzenia, czy istnieją zależności między trzema parametrami: zdolnością spiekania węgla według Rogi, zawartością pierwiastka C oraz zawartością części lotnych. Analiza wariancji posłużyła do oszacowania zależności między zawartością popiołu w węglach, a analiza składowych głównych pozwoliła na stwierdzenie, które z badanych parametrów są istotne statystycznie. Analizie tej poddano wszystkie dane, przedstawione w tabeli 1, dotyczące parametrów jakościowych węgla kamiennych pobranych w pięciu kopalniach w 2009 r.

**Tabela 1.** Parametry jakościowe węgla kamiennych wyprodukowanych w 2009 r. w wybranych kopalniach Górnośląskiego Zagłębia Węglowego

Lp.	Kopalnia	Typ węgla	Parametry jakościowe							
			wilgość $W^A$ , %	popiół $A^a$ %	części lotne $V^{daf}$ %	ciepło spalania $Q_s^a$ , kJ/kg	pierwiastek węgla $C^a$ %	R/I	S/I	dylatacja $b$ %
1	2	3	4	5	6	7	8	9	10	11
1	A	32.1	2,47	6,75	32,86	34 215	77	16	1,5	5
2		32.1	4,44	4,09	33,17	33 754	75	10	1,5	10
3		32.1	4,38	3,64	34,00	33 844	81	14	1,5	7
4		33	2,10	5,85	34,19	34 373	78	41	3,5	8
5		33	2,18	3,22	33,30	34 809	79	47	4,5	12
6		32.1	4,09	6,09	32,83	33 570	75	12	1	-2
7		32.1	2,43	3,36	30,92	34 450	71	18	1,5	-6
8		32.2	1,95	4,04	30,91	34 590	86	31	2,5	5
9		32.1	3,69	13,13	34,32	33 519	81	14	1	-8
10		32.1	4,11	5,35	35,69	33 494	75	13	1	2
11		32.1	4,50	6,12	34,80	33 567	72	29	2	10
12		32.1	4,20	6,50	33,28	34 654	76	21	2	8
13	B	34.1	2,52	10,90	37,83	34 765	79	58	6	40
14		33	1,93	9,40	39,21	34 899	69	53	6,5	10
15		34.2	1,71	32,41	37,11	34 789	86	81	8,5	55
16		34.2	1,81	6,02	37,89	35 489	85	79	8	34
17		34.2	2,12	35,50	38,35	34 987	79	79	8	25
18		34.2	2,38	35,18	37,99	35 892	82	76	7,5	48
19		34.2	2,82	15,84	38,98	34 274	81	69	6,5	30
20		34.2	2,54	20,99	37,99	35 006	79	75	7	50
21		34.1	2,73	17,35	39,20	34 240	82	59	5	29
22		34.2	1,72	34,37	39,55	34 481	75	73	6,5	48
23		34.1	2,30	12,67	38,98	34 789	77	60	6	45
24		34.2	2,80	27,89	37,98	34 567	78	55	6	42
25	C	32.2	2,07	6,61	31,19	34 764	65	39	3,5	-6
26		33	1,66	11,34	30,28	34 809	72	49	4,5	15
27		33	2,10	42,35	33,44	35 269	85	53	4,5	20
28		34.2	1,42	13,96	31,66	34 925	78	61	6	17
29		34.2	1,50	25,41	31,26	35 258	84	70	7	35

30		34.2	1,49	18,30	31,60	34 551	79	73	7,5	28	
31		33	2,13	25,30	29,88	34 895	86	44	4	15	
32		34.2	2,05	3,89	34,35	35 326	78	70	7	28	
33		32.2	2,53	2,20	29,74	34 865	71	23	2	0	
34		34.2	1,80	6,73	32,32	34 803	75	63	6,5	-15	
35		34.2	2,76	10,89	31,56	34 952	77	65	6	12	
36		34.2	3,18	11,89	32,43	35 123	71	44	5	19	
37	D	32.1	5,31	12,20	40,04	32 444	65	8	0,5	-17	
38		31.2	6,72	7,88	39,24	32 804	68	2	1	-14	
39		32.1	5,72	12,45	40,45	32 436	63	8	1	-14	
40		32.1	4,87	8,25	39,25	31 963	70	10	1	-5	
41		32.1	6,25	7,35	38,36	33 258	72	5	1	-2	
42		32.1	8,24	6,54	40,25	30 147	69	12	1,5	2	
43		31.2	5,24	10,25	38,79	31 456	64	5	0,5	1	
44		32.1	6,41	11,45	39,82	31 002	66	6	1	8	
45		32.1	5,78	6,45	37,89	32 589	72	9	1	-14	
46		32.1	6,79	8,21	38,96	31 026	75	8	1	-18	
47		32.1	5,87	11,9	39,99	32 123	66	10	1	-12	
48		32.1	7,01	8,33	39,02	32 456	67	4	0,5	-10	
49		E	32.2	3,15	5,94	34,58	34 128	77	22	1,5	-14
50			33	2,16	16,42	32,10	33 612	68	55	4,5	2
51	34.1		2,44	7,16	33,79	34 477	79	61	6	-5	
52	32.2		3,23	10,56	33,60	33 972	73	28	1	-14	
53	32.2		2,25	8,25	32,56	33 698	66	55	5	3	
54	32.2		3,15	5,26	33,26	34 159	78	36	4	-15	
55	32.2		2,78	14,26	33,01	30 999	75	48	4,5	7	
56	32.2		1,99	3,69	34,58	33 698	70	52	5	5	
57	32.2		2,45	5,26	32,98	32 587	68	49	4,5	-10	
58	32.2		2,11	6,89	32,57	33 589	74	60	5,5	6	
59	32.2		3,12	5,67	34,78	34 543	76	34	4	6	
60	32.2	2,98	12,08	33,56	33 456	72	57	6	2		

#### 4. WYNIKI BADAŃ

Wyniki badań węgla uzyskanych z wytypowanych kopalń w 2009 r. przedstawiono w tabeli 1. W tabeli 2 natomiast, podano średnią zawartość popiołu w węglach, wyprodukowanych w poszczególnych kwartałach od 2005 do 2009 r. oraz średnią zawartość popiołu dla każdej kopalni, uzyskaną z wyników za lata 2005–2008. Zawartość popiołu w węglu jest parametrem, od którego w znacznym stopniu zależą właściwości energetyczne. Każda zmiana wartości tego parametru pociąga za sobą charakterystyczne zmiany wszystkich pozostałych parametrów węgla. Dlatego też w niniejszych badaniach do ich scharakteryzowania została wybrana zawartość popiołu.

**Tabela 2.** Zmienność zawartości popiołu  $A^a$  w węglach kamiennych wyprodukowanych w wybranych kopalniach w latach 2005–2009, %

Kopalnia	Kwartał	Rok				Średnia zawartość popiołu $A^a$ w latach 2005–2008, %	Zawartość popiołu $A^a$ w 2009 r., %
		2005	2006	2007	2008		
1	2	3	4	5	6	7	8
A	I	12,53	6,21	5,98	4,25	7,20	5,10
	II	14,28	6,51	5,99	4,61		4,98
	III	10,31	6,31	6,05	4,12		4,56
	IV	13,61	6,89	6,10	4,59		5,12
	średnia roczna	<b>11,93</b>	<b>6,48</b>	<b>6,03</b>	<b>4,39</b>		<b>5,68</b>
B	I	24,89	23,56	22,65	20,89	22,98	21,95
	II	24,32	23,79	22,67	20,90		21,56
	III	24,96	23,65	22,78	20,56		20,56
	IV	24,56	23,89	22,89	20,69		21,12
	średnia roczna	<b>24,68</b>	<b>23,72</b>	<b>22,75</b>	<b>20,76</b>		<b>21,54</b>
C	I	21,58	21,56	18,98	16,35	19,54	16,21
	II	21,69	20,78	19,00	16,79		15,98
	III	21,79	20,69	19,05	16,94		16,56
	IV	21,42	20,56	18,85	16,50		16,23
	średnia roczna	<b>21,62</b>	<b>20,90</b>	<b>18,97</b>	<b>16,65</b>		<b>14,91</b>
D	I	13,65	13,06	11,65	10,80	12,24	10,25
	II	13,89	13,08	11,32	10,76		10,25
	III	13,78	12,98	11,56	10,56		10,65
	IV	13,55	13,00	11,21	10,89		10,90
	średnia roczna	<b>13,72</b>	<b>13,03</b>	<b>11,44</b>	<b>10,75</b>		<b>9,27</b>
E	I	12,03	12,90	10,59	9,86	11,32	9,56
	II	12,06	13,05	10,61	9,99		9,54
	III	12,10	12,84	10,43	9,56		9,23
	IV	11,98	12,79	10,69	9,64		9,68
	średnia roczna	<b>12,04</b>	<b>12,90</b>	<b>10,58</b>	<b>9,76</b>		<b>8,54</b>

#### 4.1. Test $t$ -Studenta

Test  $t$ -Studenta pozwala na oszacowanie prawdziwej wartości zmiennej na podstawie próbki i ustalenie przedziałów ufności, w granicach których znajduje się prawdziwa wartość danej zmiennej. Zmienna  $t$  jest to różnica między średnią uzyskaną z próbki i prawdziwą średnią z populacji, z której pobrano próbkę, podzielona przez oszacowany błąd standardowy średniej. Rozkład  $t$  nie jest rozkładem normalnym, ale w pewnych warunkach zbliża się do normalnego. Jeśli  $\mu$  oznacza prawdziwą średnią z populacji, to  $t$  można zapisać w postaci wzoru

$$t = \frac{|\bar{x} - \mu|}{S(\bar{x})} \quad (1)$$

gdzie:

- $\bar{x}$  – średnia z próbki,
- $\mu$  – prawdziwa średnia z populacji,
- $S(\bar{x})$  – oszacowany błąd standardowy średniej.

Wykonując test istotności dla wartości oczekiwanej w przypadku pojedynczej próbki, weryfikuje się hipotezę  $H_0: m = m_0$ , gdzie  $m$  jest nieznaną wartością oczekiwaną (przeciętną) pewnej cechy mierzalnej  $X$  zaś  $m_0$  wartością oczekiwaną dla hipotezy  $H_0$  w pojedynczej próbce. Jeśli  $X \sim N(m, \sigma)$ , gdzie  $\sigma$  jest znane, statystyka testowa przyjmuje postać

$$U = \frac{\bar{X} - m_0}{\sigma} \sqrt{n} |_{H_0} \approx N(0,1) \quad (2)$$

gdzie:

- $\bar{X}$  – średnia cechy mierzalnej,
- $m_0$  – wartość przypisana dla hipotezy  $H_0$ ,
- $n$  – liczebność próbki,
- $\sigma$  – odchylenie standardowe.

Celem testu  $t$ -Studenta była ocena średniej zawartości popiołu w węglach z poszczególnych kopalń z lat 2005–2009, a dokładnie weryfikacja **hipotezy  $H_0$** , zgodnie z którą średnia zawartość popiołu w węglach wyprodukowanych w 2009 r. była równa średniej zawartości popiołu w węglach kamiennych wyprodukowanych w latach 2005–2008, do **hipotezy  $H_1$** , że ta średnia zawartość popiołu różni się między sobą.

W analizie wykorzystano dane zawarte w tabeli 2. Średnią zawartość popiołu obliczoną dla węgla wyprodukowanych w 2009 r. w każdej z pięciu wybranych kopalń, zamieszczoną w kolumnie 8, porównano ze średnią zawartością popiołu dla czterech lat, podaną w kolumnie 7. Każdą analizę wykonano przy przyjętym poziomie istotności  $\alpha = 0,05$ .

Testy średnich względem stałej wartości odniesienia dla pięciu grup węgla kamiennych dla każdej z pięciu kopalń przedstawiono w tabelach 3–7, natomiast w tabeli 8 zebrano wyniki dla pięciu kopalń.

**Tabela 3.** Analiza statystyczna zawartości popiołu w kopalni A

Zmienna	Test średnich względem stałej wartości odniesienia (ArkusZ21)							
	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t	df	p
Zmn1	5,678333	2,673652	12	0,771817	7,200000	-1,97154	11	0,074336

**Tabela 4.** Analiza statystyczna zawartości popiołu w kopalni B

Zmienna	Test średnich względem stałej wartości odniesienia (ArkusZ21)						
	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t	
Zmn1	21,54333	11,01447	12	3,179604	22,98000	-0,451838	

Tabela 5. Analiza statystyczna zawartości popiołu w kopalni C

Zmienna	Test średnich względem stałej wartości odniesienia (Arkus21)					
	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t
Zmn1	14,90583	11,44299	12	3,303307	19,54000	-1,40289

Tabela 6. Analiza statystyczna zawartości popiołu w kopalni D

Zmienna	Test średnich względem stałej wartości odniesienia (Arkus21)					
	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t
Zmn1	9,271667	2,244151	12	0,647831	12,24000	-4,58196

Tabela 7. Analiza statystyczna zawartości popiołu w kopalni E

Zmienna	Test średnich względem stałej wartości odniesienia (Arkus21)					
	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t
Zmn1	8,453333	4,000930	12	1,154969	11,32000	-2,48203

Tabela 8. Wyniki zbiorcze testu t-Studenta

Symbol kopalni	Test t-Studenta					
	liczba stopni swobody	średnia wyników	odchylenie standardowe	t	pwartość	hipoteza zerowa
A	11	5,68	2,67	-1,97	0,07	+
B	11	21,54	11,01	-0,45	0,66	+
C	11	14,91	11,44	-1,40	0,19	+
D	11	9,27	2,24	-4,58	0,001	-
E	11	8,45	4,00	-2,48	0,03	-

Uwagi: „+” – nie ma podstaw do odrzucenia hipotezy zerowej, „-” – należy odrzucić hipotezę zerową.

W podsumowaniu stwierdzono, że w przypadku kopalń D i E należy odrzucić hipotezę zerową, według której średnia zawartość popiołu w węglach wyprodukowanych w 2009 r. i średnia zawartość popiołu w węglach wyprodukowanych w latach 2005–2008 była taka sama. Świadczy to o istotnej zmianie jakości węgla wyprodukowanych w 2009 r. W przypadku pozostałych kopalń nie ma podstaw do odrzucenia hipotezy zerowej. Czyli można stwierdzić, że w kopalniach A, B i C w 2009 r. zmiana taka nie wystąpiła.

## 4.2. Regresja wieloraka

Regresja jest to liczbowe określenie wpływu poszczególnych zmiennych objaśniających na zmienną objaśnianą. W przypadku regresji prostej jest to zależność jednej zmiennej od drugiej. Gdy ma się do czynienia z regresją wieloraką, liczba zmiennych objaśniających jest powiększona do dwóch i więcej. Statystyczne procedury analizy regresji dwóch i więcej zmiennych objaśniających bazują na procedurach ana-



lizej regresji jednej zmiennej objaśniającej. Problemem statystycznym jest zawsze wybór właściwych zmiennych. Dokonuje się tego ze zbioru zmiennych potencjalnych, przez ich redukcję do podzbioru zmiennych dopuszczalnych. Ogólną zasadą specyfikacji zmiennych diagnostycznych jest preferowanie takich zmiennych objaśniających, które są statystycznie istotnie powiązane ze zmienną objaśnianą i jednocześnie nie są statystycznie istotnie powiązane między sobą.

W wykonywanych badaniach analiza regresji wielorakiej została wykorzystana do oceny wpływu dwóch parametrów objaśniających: zawartości części lotnych i pierwiastka C na parametr objaśniany: zdolność spiekania  $RI$ , który jest najważniejszym wyróżnikiem typu węgla. Celem bezpośrednim przeprowadzonej analizy regresji wielorakiej było zweryfikowanie hipotezy, czy zawartość części lotnych i pierwiastka C w węglach kamiennych, wyprodukowanych w wybranych kopalniach GZW, ma wpływ na wskaźnik zdolności spiekania tych węgli, a tym samym czy ma wpływ na wskaźnik typu węgla.

W analizie wykorzystano dane przedstawione w tabeli 1, w kolumnach 6, 8 i 9, odnoszące się do 60 różnych próbek węgla, pochodzących z wybranych kopalń. Analiza statystyczna obejmowała analizę regresji wielorakiej oraz analizę reszt. Wyniki analizy regresji zmiennej zależnej  $RI$  przedstawiono w tabeli 9.

**Tabela 9.** Regresja wieloraka dla zależności  $RI$  od zawartości części lotnych  $V^{daf}$  i pierwiastka  $C^a$

Podsumowanie regresji zmiennej zależnej: $RI$ (Arkusz10)						
R= ,52524199 R <sup>2</sup> = ,27587914 Skoryg. R2= ,25047139						
F(2,57)=10,858 p<,00010 Błąd std. estymacji: 21,531						
N=60	BETA	Bł. std. BETA	B	Bł. std. B	t(57)	poziom p
W. wolny			-99,0120	52,89146	-1,87199	0,066342
$V^{daf}$	-0,064950	0,115999	-0,4970	0,88756	-0,55992	0,577726
Ca	0,506084	0,115999	2,0865	0,47824	4,36284	0,000055

Liniowa regresja zależności zdolności spiekania węgla  $RI$  od zawartości części lotnych  $V^{daf}$  i zawartości pierwiastka węgla  $C^a$  w badanych próbkach jest statystycznym modelem, powstałym na podstawie tabeli 1. Przybiera ona postać

$$y = -0,50 V^{daf} + 2,09 C^a - 99,01 \quad (3)$$

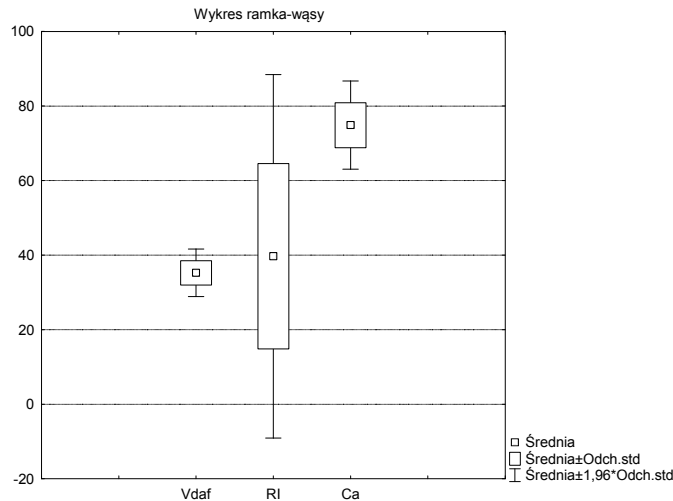
Z tabeli 9 wynika, że istotny statystycznie jest tylko człon związany z pierwiastkiem C. Na jej podstawie można wyciągnąć następujące wnioski:

- Ze wzrostem zawartości części lotnych o jeden procent, przy założeniu niezmienności zawartości pierwiastka C, zdolność spiekania węgla maleje o 0,50% średniej wartości. Wyraz ten występuje ze średnim losowym błędem równym  $\pm 0,89$ . Jest on jednak statystycznie nieistotny na poziomie  $\alpha = 0,05$ , gdyż krytyczny poziom istotności  $\hat{\alpha} = 0,58$ .
- Ze wzrostem zawartości pierwiastka C o jeden procent, przy założeniu niezmienności zawartości części lotnych, zdolność spiekania węgla wzrasta o 2,09% średniej wartości. Wyraz ten występuje ze średnim losowym błędem równym  $\pm 0,48$ .

Jest on statystycznie istotny na poziomie  $\alpha = 0,05$ , gdyż krytyczny poziom istotności  $\hat{\alpha} = 0,000$ .

- Wyraz wolny równy 99,01 występuje ze średnim losowym błędem wynoszącym  $\pm 52,89$  i jest statystycznie nieistotny na poziomie  $\alpha = 0,05$ , gdyż krytyczny poziom istotności  $\hat{\alpha} = 0,07$ .

Średnie wartości oraz odchylenie standardowe badanych zmiennych przedstawiono na rysunku 1.



Rys. 1. Wykres ramka – wąsy dla analizowanych danych

Fig. 1. Box-plot for the analysed data

Z analizy wariancji w rachunku regresji wielorakiej (tab. 10) wynika, że łączny efekt oddziaływania obu zmiennych objaśniających (zawartość części lotnych i pierwiastka C) na zmienną objaśnianą (zdolność spiekania węgla) jest statystycznie istotny. Przeprowadzając następnie test Fischera-Snedecora, który jest testem istotności różnic precyzji, charakteryzujących dwie porównywane serie danych, uzyskano odpowiedź na pytanie: czy wszystkie porównywane serie danych są równorzędne precyzyjnie, to znaczy czy są statystycznie identyczne do oznaczania poziomu istotności. W teście Fischera-Snedecora, określonym przez odchylenie  $s_1 = 2$  (liczbę stopni swobody, czyli liczbę niezależnych wyników obserwacji pomniejszoną o liczbę związków, które łączą te wyniki ze sobą<sup>1</sup>) oraz  $s_2 = 57$ , uzyskano wartość  $F = 10,86$ . Wartość wskaźnika  $F$  jest duża i jest większa od poziomu krytycznego równego 0,0001.

<sup>1</sup> [http://pl.wikipedia.org/wiki/Liczba\\_stopni\\_swobody\\_\(statystyka\)](http://pl.wikipedia.org/wiki/Liczba_stopni_swobody_(statystyka))

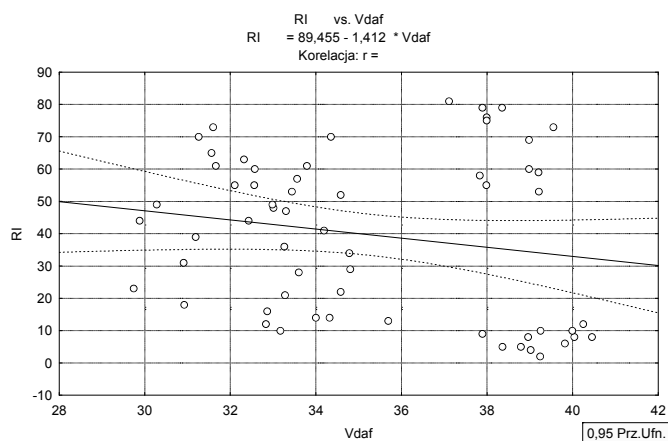
**Tabela 10.** Analiza wariancji

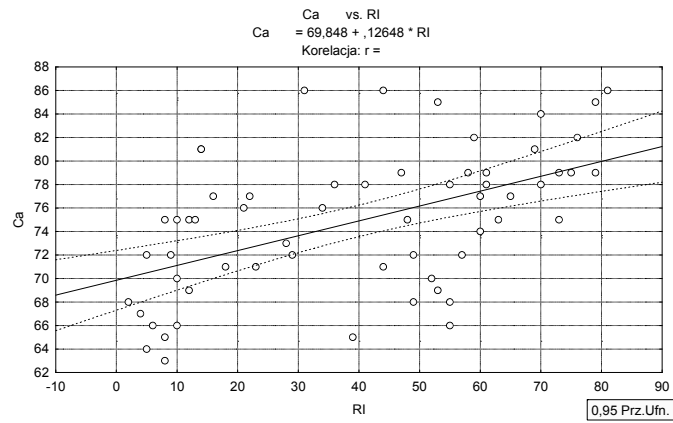
Efekt	Analiza wariancji ; DV: RI (Arkusz10)				
	Suma kwadrat.	df	Średnia kwadrat.	F	poziom p
Regres.	10067,65	2	5033,826	10,85807	0,000101
Reszta	26425,33	57	463,602		
Razem	36492,98				

Z macierzy liniowej współczynników korelacji prostej, przedstawionej w tabeli 11, można wywnioskować, że korelacja między poszczególnymi zmiennymi objaśniającymi a zmienną objaśnianą, jest istotna:  $r_{R|Vdaf} = -0,184579$ ,  $r_{R|Ca} = 0,521437$  oraz  $r_{Ca|Vdaf} = -0,236380$ . Należy jednak zauważyć, że zależności te są małe. Potwierdzeniem tych wniosków są przedstawione wykresy (rys. 2, 3 i 4).

**Tabela 11.** Macierz liniowa współczynników korelacji prostej

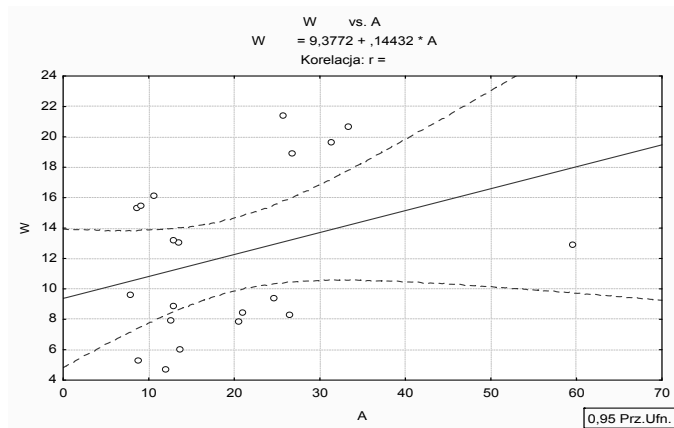
Zmienna	Korelacje (Arkusz10)		
	Vdaf	Ca	RI
Vdaf	1,000000	-0,236380	-0,184579
Ca	-0,236380	1,000000	0,521437
RI	-0,184579	0,521437	1,000000

**Rys. 2.** Korelacja zdolności spiekania węgla  $RI$  i zawartości części lotnych  $Vdaf$ **Fig. 2.** Correlation of coal caking properties  $RI$  and volatile matter  $Vdaf$  content



**Rys. 3.** Korelacja zdolności spiekania węgla  $RI$  i zawartości pierwiastka węgla  $C^a$

**Fig. 3.** Correlation of coal caking properties  $RI$  and carbon element  $C^a$  content



**Rys. 4.** Korelacja zawartości pierwiastka węgla  $C^a$  i zawartości części lotnych  $V^{daf}$

**Fig. 4.** Correlation of carbon element  $C^a$  content and volatile matter  $V^{daf}$  content

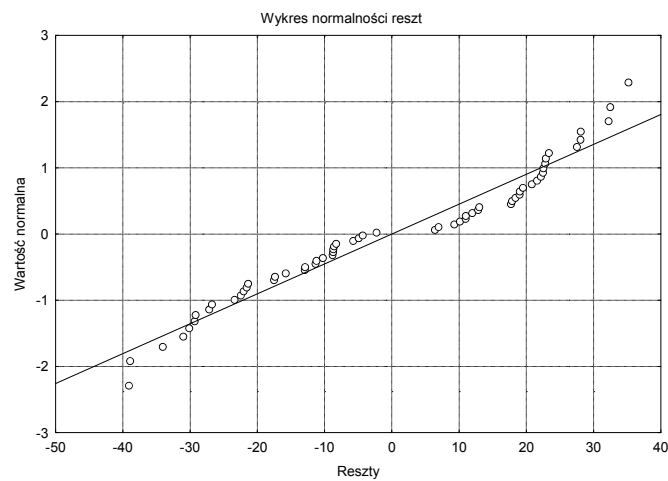
Na podstawie analizy reszt (tab. 12) stwierdzono, że próbki numer 25, 31 i 39 należy wyłączyć z analizowanych reszt, ponieważ wskazują na to duże wartości odległości Mahalanobisa wyznaczone dla tych przypadków, które przyjmują wartości odpowiednio: 54998, 4,9651, 5,2404. Odległości te mierzą odległość przypadku od środka ciężkości wyznaczonego przez zmienne niezależne<sup>2</sup>. Na wykresie normalności reszt (rys. 5) można również zaobserwować, że trzy skrajne reszty różnią się od całości.

<sup>2</sup> <http://zsi.tech.us.edu.pl/~nowak/adb/influence.pdf>

Tabela 12. Wartości przewidywane i reszty analizowanych przypadków (próbek)

Nr przypadku	Wartości przewidywane i reszty (Arkusz10)								
	Zmienna zależna: RI								
	Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard Przewid.	Standard Reszta	Bł. std. W.przew.	Mahaln. Odległ.	Usunięte Reszta	Cooka Odległ.
1	16,00000	45,31789	-29,3179	0,43134	-1,36163	3,494976	0,571181	-30,1112	0,017176
2	10,00000	40,99084	-30,9908	0,10009	-1,43933	3,324857	0,423532	-31,7479	0,017281
3	14,00000	53,09734	-39,0973	1,02688	-1,81583	4,002056	1,054988	-40,4964	0,040737
4	41,00000	46,74342	-5,7434	0,54047	-0,26675	3,190876	0,312432	-5,8724	0,000545
5	47,00000	49,27222	-2,2722	0,73406	-0,10553	3,604514	0,670150	-2,3377	0,000110
6	12,00000	41,15981	-29,1598	0,11303	-1,35429	3,499448	0,575162	-29,9510	0,017038
7	18,00000	33,76302	-15,7630	-0,45322	-0,73209	5,404522	2,733911	-16,8229	0,012821
8	31,00000	65,06545	-34,0654	1,94308	-1,58213	6,416980	4,257108	-37,3861	0,089263
9	14,00000	52,93830	-38,9383	1,01471	-1,80844	3,983115	1,035741	-40,3180	0,039997
10	13,00000	39,73848	-26,7385	0,00422	-1,24184	2,810957	0,022242	-27,2021	0,009068
11	29,00000	33,92129	-4,9213	-0,44110	-0,22856	3,164389	0,291009	-5,0299	0,000393
12	21,00000	43,02267	-22,0227	0,25564	-1,02282	3,256433	0,366222	-22,5382	0,008354
13	58,00000	47,02096	10,9790	0,56172	0,50991	4,365913	1,442476	11,4498	0,003876
14	53,00000	25,47017	27,5298	-1,08806	1,27859	4,830616	1,986359	28,9889	0,030413
15	81,00000	61,98425	19,0158	1,70720	0,88316	6,557667	4,489410	20,9600	0,029300
16	79,00000	59,51012	19,4899	1,51780	0,90518	6,490453	4,377798	21,4379	0,030026
17	79,00000	46,76253	32,2375	0,54193	1,49723	4,671885	1,794401	33,8302	0,038742
18	76,00000	53,20093	22,7991	1,03481	1,05887	5,408527	2,739421	24,3345	0,026865
19	69,00000	50,62244	18,3776	0,83742	0,85352	5,651939	3,082048	19,7376	0,019301
20	75,00000	46,94144	28,0586	0,55563	1,30315	4,457206	1,544986	29,3148	0,026478
21	59,00000	52,59960	6,4004	0,98878	0,29726	6,113756	3,773551	6,9617	0,002810
22	73,00000	37,82018	35,1798	-0,14263	1,63388	4,740448	1,876529	36,9719	0,047640
23	60,00000	42,27645	17,7235	0,19851	0,82315	4,623891	1,737624	18,5804	0,011448
24	55,00000	44,85991	10,1401	0,39628	0,47094	4,195657	1,256968	10,5403	0,003033
25	39,00000	21,10986	17,8901	-1,42185	0,83089	7,137409	5,499839	20,0987	0,031916
26	49,00000	36,16758	12,8324	-0,26914	0,59599	5,643559	3,070003	13,7791	0,009378
27	53,00000	61,72163	-8,7216	1,68710	-0,40507	5,486837	2,848005	-9,3273	0,004062
28	61,00000	48,00075	12,9993	0,63672	0,60373	4,221777	1,284949	13,5190	0,005052
29	70,00000	60,71852	9,2815	1,61031	0,43107	5,656404	3,088474	9,9695	0,004932
30	73,00000	50,11707	22,8829	0,79873	1,06277	4,365589	1,442116	23,8640	0,016833
31	44,00000	65,57733	-21,5773	1,98226	-1,00213	6,836727	4,965105	-23,9967	0,041743
32	70,00000	46,66391	23,3361	0,53438	1,08382	3,168198	0,294079	23,8525	0,008857
33	23,00000	34,34945	-11,3494	-0,40832	-0,52711	6,265381	4,012426	-12,3994	0,009360
34	63,00000	41,41326	21,5867	0,13243	1,00257	3,791415	0,846068	22,2775	0,011064
35	65,00000	45,96395	19,0360	0,48080	0,88411	4,226910	1,290468	19,7991	0,010862
36	44,00000	33,01260	10,9874	-0,51066	0,51030	4,421883	1,505070	11,4712	0,003990
37	8,00000	16,71170	-8,7117	-1,75855	-0,40460	6,215493	3,933184	-9,5036	0,005412
38	2,00000	23,36876	-21,3688	-1,24893	-0,99245	5,059417	2,274339	-22,6176	0,020309
39	8,00000	12,33494	-4,3349	-2,09360	-0,20133	6,993132	5,240384	-4,8461	0,001781
40	10,00000	27,53679	-17,5368	-0,92985	-0,81447	4,678538	1,802318	-18,4058	0,011501
41	5,00000	32,15208	-27,1521	-0,57654	-1,26104	3,933807	0,986060	-28,0897	0,018937
42	12,00000	24,95332	-12,9533	-1,12763	-0,60160	5,428233	2,766598	-13,8325	0,008744
43	5,00000	15,24641	-10,2464	-1,87072	-0,47588	6,075950	3,714902	-11,1329	0,007096
44	6,00000	18,90753	-12,9075	-1,59045	-0,59947	5,837544	3,353442	-13,9316	0,010258
45	9,00000	32,38566	-23,3857	-0,55866	-1,08612	3,689135	0,748697	-24,0929	0,012252
46	8,00000	38,11339	-30,1134	-0,12018	-1,39858	4,327230	1,399680	-31,3809	0,028598
47	10,00000	18,82304	-8,8230	-1,59692	-0,40978	5,918076	3,473921	-9,5441	0,004948
48	4,00000	21,39160	-17,3916	-1,40029	-0,80773	5,211944	2,473718	-18,4741	0,014378

49	22,00000	44,46310	-22,4631	0,36591	-1,04327	2,971353	0,140275	-22,8992	0,007180
50	55,00000	26,91711	28,0829	-0,97729	1,30428	5,532831	2,912508	30,0683	0,042924
51	61,00000	49,02871	11,9713	0,71542	0,55599	3,476994	0,555226	12,2918	0,002833
52	28,00000	36,60415	-8,6041	-0,23572	-0,39961	3,355626	0,449691	-8,8183	0,001358
53	55,00000	22,51551	32,4845	-1,31425	1,50870	6,011782	3,616191	35,2310	0,069574
54	36,00000	47,20560	-11,2056	0,57585	-0,52043	3,437804	0,520739	-11,4987	0,002424
55	48,00000	41,07035	6,9296	0,10618	0,32184	3,404802	0,492000	7,1074	0,000908
56	52,00000	29,85762	22,1424	-0,75219	1,02838	3,759495	0,815395	22,8387	0,011434
57	49,00000	26,47978	22,5202	-1,01077	1,04592	5,065029	2,281569	23,8394	0,022612
58	60,00000	39,20252	20,7975	-0,03681	0,96591	3,739506	0,796318	21,4443	0,009973
59	34,00000	42,27721	-8,2772	0,19857	-0,38442	2,843086	0,045361	-8,4241	0,000890
60	57,00000	34,53753	22,4625	-0,39393	1,04324	3,577688	0,645631	23,1003	0,010593
<b>Minimum</b>	2,00000	12,33494	-39,0973	-2,09360	-1,81583	2,810957	0,022242	-40,4964	0,000110
<b>Maksimum</b>	81,00000	65,57733	35,1798	1,98226	1,63388	7,137409	5,499839	36,9719	0,089263
<b>Średnia</b>	39,68333	39,68333	0,0000	0,00000	0,00000	4,667128	1,966667	0,0772	0,016795
<b>Mediana</b>	44,00000	41,11508	-3,3036	0,10960	-0,15343	4,439545	1,525028	-3,5919	0,010963



Rys. 5. Wykres normalności reszt

Fig. 5. Graph of residue normality

Tabela 13. Regresja wieloraka po odrzuceniu próbek 25, 31 i 39

Podsumowanie regresji zmiennej zależnej: RI (Arkusz10)						
R= ,52871260 R^2= ,27953702 Skoryg. R2= ,25285320						
F(2,54)=10,476 p<,00014 Błąd std. estymacji: 21,750						
N=57	BETA	Bł. std. BETA	B	Bł. std. B	t(54)	poziom p
W. wolny			-115,671	56,44922	-2,04911	0,045323
Vdaf	-0,057678	0,117755	-0,463	0,94484	-0,48981	0,626250
Ca	0,514460	0,117755	2,294	0,52513	4,36888	0,000057

Po odrzuceniu parametrów próbek 25, 31 i 39 (zawartych w tab. 1), z analizy statystycznej uzyskano wyniki analizy regresji przedstawione w tabeli 13.

Z danych zawartych w tej tabeli można wywnioskować, że współczynniki regresji związane z zawartością pierwiastka C<sup>a</sup> oraz wyrazem wolnym są statystycznie istotne.

Wartości przewidywane i reszty po odrzuceniu parametrów próbek 25, 31 i 39 przedstawiono w tabeli 14. Na podstawie tej analizy reszt stwierdzono, że z analizowanych reszt należałoby jeszcze wyłączyć próbkę 8 (wyznaczona wartość Mahalano-bisa dla tego przypadku wynosiła 4,8612 i była największą odległością ze wszystkich 57 analizowanych próbek). Po odrzuceniu tych wszystkich przypadków uzyskano wyniki regresji wielorakiej, przedstawione w tabeli 15.

Tabela 14. Wartości przewidywane i reszty po odrzuceniu przypadków 25, 31 i 39

Nr przypadku	Wartości przewidywane i reszty (Arkusz10)								
	Zmienna zależna: RI								
	Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard Przewid.	Standard Reszta	Bł. std. W.przew.	Mahaln. Odlegl.	Usunięte Reszta	Cooka Odlegl.
1	16,00000	45,77932	-29,7793	0,42123	-1,36918	3,715162	0,651472	-30,6743	0,019345
2	10,00000	41,04734	-31,0473	0,06554	-1,42748	3,525454	0,488865	-31,8851	0,018822
3	14,00000	54,42874	-40,4287	1,07138	-1,85881	4,252964	1,158761	-42,0360	0,047608
4	41,00000	47,45805	-6,4581	0,54741	-0,29692	3,344338	0,341573	-6,6144	0,000729
5	47,00000	50,16419	-3,1642	0,75083	-0,14548	3,832167	0,756010	-3,2656	0,000233
6	12,00000	41,20469	-29,2047	0,07737	-1,34276	3,719885	0,655628	-30,0847	0,018656
7	18,00000	32,91162	-14,9116	-0,54600	-0,68560	5,788667	2,984291	-16,0484	0,012855
8	31,00000	67,33004	-36,3300	2,04113	-1,67036	7,025944	4,861228	-40,5628	0,120983
9	14,00000	54,28065	-40,2807	1,06025	-1,85200	4,218428	1,124127	-41,8551	0,046436
10	13,00000	39,88110	-26,8811	-0,02212	-1,23592	2,902024	0,014508	-27,3683	0,009396
11	29,00000	33,41023	-4,4102	-0,50852	-0,20277	3,377717	0,368135	-4,5192	0,000347
12	21,00000	43,29069	-22,2907	0,23416	-1,02487	3,445473	0,422863	-22,8645	0,009244
13	58,00000	48,06773	9,9323	0,59324	0,45666	4,489889	1,403973	10,3744	0,003232
14	53,00000	24,48655	28,5134	-1,17929	1,31098	5,234125	2,260682	30,2663	0,037382
15	81,00000	64,46072	16,5393	1,82546	0,76043	6,929466	4,701844	18,4078	0,024236
16	79,00000	61,80548	17,1945	1,62587	0,79056	6,815381	4,516213	19,0667	0,025153
17	79,00000	47,82708	31,1729	0,57515	1,43325	4,809542	1,755868	32,7756	0,037014
18	76,00000	54,87645	21,1236	1,10503	0,97121	5,619991	2,756485	22,6348	0,024104
19	69,00000	52,12402	16,8760	0,89814	0,77591	5,852501	3,072260	18,1933	0,016887
20	75,00000	47,99369	27,0063	0,58768	1,24168	4,585011	1,506161	28,2623	0,025012
21	59,00000	54,31646	4,6835	1,06294	0,21534	6,347663	3,787398	5,1196	0,001573
22	73,00000	38,09472	34,9053	-0,15640	1,60485	4,926328	1,890467	36,7928	0,048936
23	60,00000	42,94701	17,0530	0,20833	0,78405	4,765934	1,706435	17,9131	0,010857
24	55,00000	45,70406	9,2959	0,41557	0,42740	4,307130	1,213650	9,6754	0,002587
26	49,00000	35,50206	13,4979	-0,35128	0,62060	6,038691	3,334353	14,6253	0,011619
27	53,00000	63,86492	-10,8649	1,78067	-0,49954	5,926656	3,175663	-11,7364	0,007207
28	61,00000	48,62892	12,3711	0,63542	0,56879	4,532233	1,449198	12,9326	0,005117
29	70,00000	62,57956	7,4204	1,68405	0,34117	6,168671	3,522189	8,0696	0,003691
30	73,00000	50,95094	22,0491	0,80996	1,01376	4,699593	1,632099	23,1289	0,017599
32	70,00000	47,38400	22,6160	0,54185	1,03983	3,313171	0,317010	23,1533	0,008765
33	23,00000	33,45771	-10,4577	-0,50495	-0,48082	6,703270	4,336799	-11,5553	0,008937
34	63,00000	41,44072	21,5593	0,09511	0,99124	4,041981	0,951586	22,3305	0,012135
35	65,00000	46,38095	18,6191	0,46645	0,85606	4,529965	1,446765	19,4633	0,011579
36	44,00000	32,21280	11,7872	-0,59853	0,54195	4,744941	1,682800	12,3762	0,005137
37	8,00000	14,92541	-6,9254	-1,89797	-0,31841	6,850652	4,573274	-7,6882	0,004132
38	2,00000	22,17841	-20,1784	-1,35278	-0,92775	5,516064	2,619480	-21,5655	0,021078
40	10,00000	26,76229	-16,7623	-1,00823	-0,77069	5,032484	2,015616	-17,7105	0,011833
41	5,00000	31,76268	-26,7627	-0,63236	-1,23048	4,165672	1,071766	-27,7818	0,019950
42	12,00000	24,00524	-12,0052	-1,21547	-0,55197	5,863399	3,087373	-12,9461	0,008583
43	5,00000	13,20965	-8,2097	-2,02694	-0,37746	6,734640	4,386701	-9,0802	0,005570
44	6,00000	17,32148	-11,3215	-1,71787	-0,52053	6,414536	3,888427	-12,4000	0,009424
45	9,00000	31,98019	-22,9802	-0,61601	-1,05657	3,907890	0,825393	-23,7468	0,012828
46	8,00000	38,36776	-30,3678	-0,13588	-1,39623	4,485998	1,399839	-31,7170	0,030155
47	10,00000	17,24281	-7,2428	-1,72378	-0,33301	6,499421	4,018197	-7,9530	0,003980
48	4,00000	19,98597	-15,9860	-1,51758	-0,73499	5,716498	2,885998	-17,1722	0,014354
49	22,00000	44,98331	-22,9833	0,36139	-1,05671	3,090458	0,148180	-23,4569	0,007828
50	55,00000	25,48276	29,5172	-1,10441	1,35713	5,968100	3,234020	31,9207	0,054060
51	61,00000	49,93742	11,0626	0,73378	0,50863	3,675642	0,616895	11,3878	0,002610

52	28,00000	36,25983	-8,2598	-0,29432	-0,37977	3,572929	0,528758	-8,4889	0,001370
53	55,00000	20,68137	34,3186	-1,46531	1,57788	6,518169	4,047088	37,7051	0,089972
54	36,00000	47,88845	-11,8885	0,57976	-0,54660	3,646312	0,591472	-12,2323	0,002963
55	48,00000	41,12139	6,8786	0,07110	0,31626	3,614698	0,564298	7,0740	0,000974
56	52,00000	28,92354	23,0765	-0,84577	1,06100	4,066696	0,975310	23,9125	0,014086
57	49,00000	25,07550	23,9245	-1,13502	1,09999	5,482754	2,576109	25,5480	0,029226
58	60,00000	39,03077	20,9692	-0,08604	0,96411	3,986048	0,898430	21,6980	0,011142
59	34,00000	42,59650	-8,5965	0,18198	-0,39524	2,950310	0,047960	-8,7576	0,000994
60	57,00000	33,98409	23,0159	-0,46539	1,05821	3,827806	0,752056	23,7516	0,012312
Minimum	2,00000	13,20965	-40,4287	-2,02694	-1,85881	2,902024	0,014508	-42,0360	0,000233
Maksimum	81,00000	67,33004	34,9053	2,04113	1,60486	7,025944	4,861228	37,7051	0,120983
Średnia	40,17544	40,17544	0,0000	0,00000	0,00000	4,844167	1,964912	0,0841	0,017945
Mediana	47,00000	41,20469	-3,1642	0,07737	-0,14548	4,585011	1,506161	-3,2656	0,011619

Tabela 15. Regresja wieloraka po odrzuceniu próbek 8, 25, 31 i 39

Podsumowanie regresji zmiennej zależnej: RI (Arkuszy10) R= ,56520326 R^2= ,31945472 Skoryg. R2= ,29377377 F(2,53)=12,439 p<,00004 Błąd std. estymacji: 21,311						
N=56	BETA	Bł. std. BETA	B	Bł. std. B	t(53)	poziom p
W. wolny			-122,943	55,45849	-2,21685	0,030946
Vdaf	-0,086918	0,114665	-0,709	0,93586	-0,75802	0,451795
Ca	0,545345	0,114665	2,517	0,52916	4,75600	0,000016

Po odrzuceniu próbek 8, 25, 31 i 39 nadal tylko współczynniki regresji, związane z pierwiastkiem C<sup>a</sup> oraz wyrazem wolnym były statystycznie istotne. Po odrzuceniu kolejnych próbek wyraz wolny stał się statystycznie nieistotny, dlatego analizę przerwano na tym etapie.

Nową liniową regresję zależności zdolności spiekania RI od zawartości części lotnych V<sup>daf</sup> oraz zawartości pierwiastka węgla C<sup>a</sup> w badanych próbkach przedstawiono za pomocą wzoru (4), który jest statystycznym modelem powstałym na podstawie tabeli 15

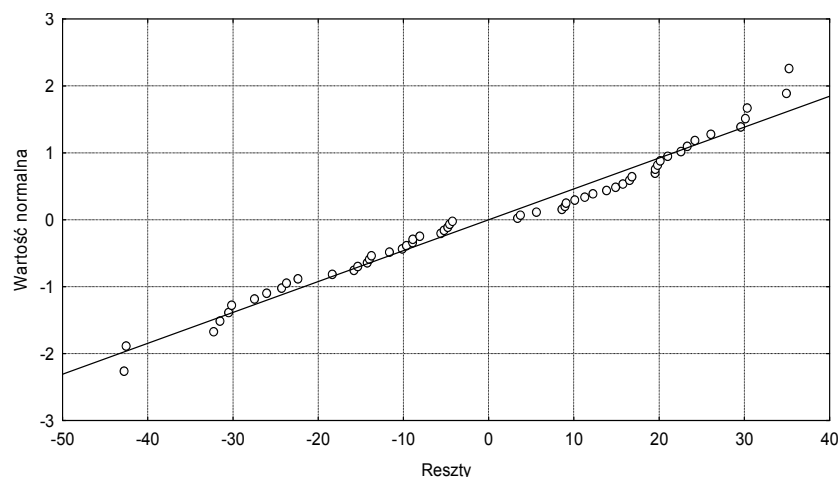
$$y = -0,71V^{daf} + 2,52C^a - 122,94 \quad (4)$$

Z tabeli 15 można wyciągnąć następujące wnioski:

- Ze wzrostem zawartości części lotnych V<sup>daf</sup> o jeden procent, przy założeniu niezmienności zawartości pierwiastka C, zdolność spiekania węgla maleje o 0,71% średniej wartości. Wyraz ten występuje ze średnim losowym błędem równym  $\pm 0,94$ . Jest on jednak statystycznie nieistotny na poziomie  $\alpha = 0,05$ , gdyż krytyczny poziom istotności  $\hat{\alpha} = 0,45$ .
- Ze wzrostem zawartości pierwiastka węgla C<sup>a</sup> o jeden procent, przy założeniu niezmienności zawartości części lotnych, zdolność spiekania węgla wzrasta o 2,52% średniej wartości. Wyraz ten występuje ze średnim losowym błędem równym  $\pm 0,53$ . Jest on statystycznie istotny na poziomie  $\alpha = 0,05$ , gdyż krytyczny poziom istotności  $\hat{\alpha} = 0,000$ .
- Wyraz wolny równy 122,94 występuje ze średnim losowym błędem równym  $\pm 55,46$  i jest statystycznie istotny na poziomie  $\alpha = 0,05$ , gdyż krytyczny poziom istotności  $\hat{\alpha} = 0,03$ .



Wnioski uzyskane po odrzuceniu próbek 8, 25, 31 i 39 potwierdzają wyniki przedstawione na rysunku 6.



**Rys. 6.** Wykres normalności reszt po odrzuceniu próbek 8, 25, 31 i 39

**Fig. 6.** Graph of residue normality after rejection of samples 8, 25, 31, and 39

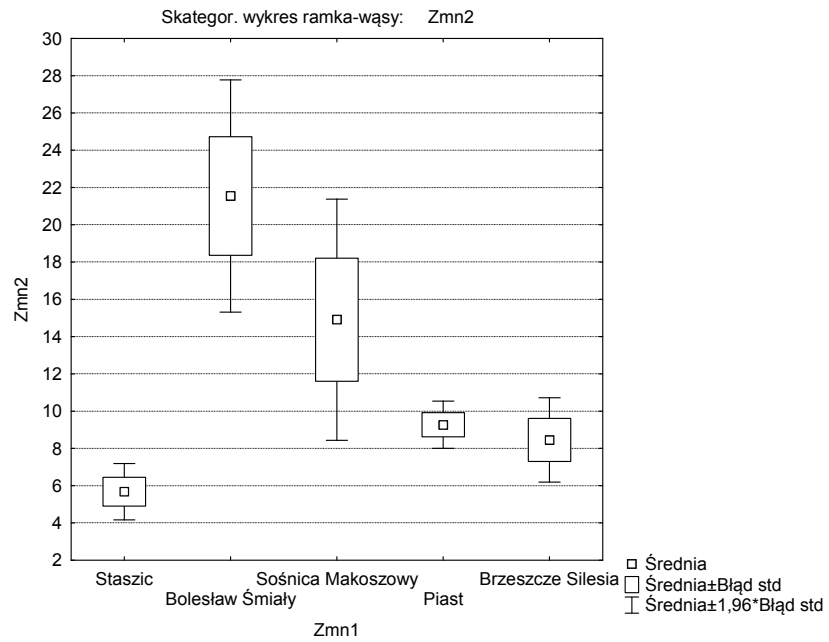
W podsumowaniu stwierdzono, że między każdą ze zmiennych (zdolnością spiekania węgla  $RI$ , zawartością pierwiastka węgla  $C^a$  oraz zawartością części lotnych  $V^{daf}$ ) istnieją niewielkie zależności. Oddziaływanie zawartości pierwiastka węgla  $C^a$  na zdolność spiekania węgla  $RI$ , a tym samym na wskaźnik typu węgla, jest istotne. Po odrzuceniu natomiast danych dla próbek 8, 25, 31 i 39 również wyraz wolny staje się statystycznie istotny.

#### 4.3. Analiza wariancji

Za pomocą analizy wariancji można poznawać zależności między różnymi zjawiskami. Analiza wariancji może być wykorzystywana do obserwacji dwóch zmiennych  $X$  i  $Y$  jednej zbiorowości  $n$ . Może również służyć do analizy przestrzeni wielowymiarowej, w której równoczesnej obserwacji i analizie są poddawane trzy lub więcej zmiennych. Aby określić siłę, kierunek i kształt powiązań zmiennych  $X$  i  $Y$ , można stosować różne metody statystyczne, a mianowicie analiza wariancji, regresji i korelacji. Wszystkie trzy metody są analitycznie komplementarne, ale w analizie statystycznej łączą się one w jedną wzajemnie uzupełniającą się całość.

Podjęto próbę przeprowadzenia analizy wariancji dla zawartości popiołu w węglach należących do pięciu różnych grup z pięciu różnych kopalń. Analizami objęto dane zamieszczone w kolumnie 5 tabeli 1.

Dane w postaci wykresu ramka – wąsy przedstawiono na rysunku 7. Wynika z niego brak normalności między grupami, czyli między zawartością popiołu w węglach pochodzących z pięciu różnych kopalń.



Rys. 7. Wykres ramka – wąsy dla zawartości popiołu w próbkach węgla pochodzących z wybranych kopalń

Fig. 7. Box-plot for ash content in coal samples originating from selected mines

W celu sprawdzenia, czy wariancja w grupach jest jednakowa, zastosowano testy: Levene’a i Browna-Foresythe’a. Test Levene’a jest testem mocnym. Dla każdej zmiennej zależnej przeprowadzono test ANOVA na podstawie bezwzględnych odchyleń wartości od średnich grupowych. Test Browna-Foresythe’a jest podobny do testu Levene’a, ale jest bardziej odporny na odchylenia od normalności zmiennych zależnych w grupach. Zamiast testu ANOVA, na odchyleniach od średniej, można zrobić test na odchyleniach od median grupowych.

Zweryfikowano hipotezy 5 i 6:

$$H_0 : S_1^2 = S_2^2 = \dots = S_n^2 \quad (5)$$

$$H_1 : \neg H_0 \quad (6)$$

Wyniki testu Levene’a przedstawiono w tabeli 16, a wyniki testu Browna-Foresythe’a – w tabeli 17.

Tabela 16. Wyniki testu Levene’a

Test Levene’a jednorodności wariancji (Arkusz18)								
Zaznaczone efekty są istotne z $p < ,05000$								
Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Zmn2	686,3944	4	171,5986	875,7325	55	15,92241	10,77718	0,000002

Obliczony, w teście Levene'a, poziom istotności  $p_{obliczone}$  był bardzo bliski zeru i mniejszy od poziomu istotności  $p$  przyjętego na poziomie 0,05, co wskazywało, że wariancja jest różna w każdej z grup.

Tabela 17. Wyniki testu Browna-Forsythe'a

Zmienna	Test jednorod. wariancji Browna-Forsythe'a (Arkusz18) Zaznaczone efekty są istotne z $p < ,05000$							
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Zmn2	644,8379	4	161,2095	1287,855	55	23,41554	6,884722	0,000144

Obliczony w teście Browna-Forsythe'a poziom istotności  $p_{obliczone}$  był także bardzo bliski zeru i mniejszy od poziomu istotności  $p$  przyjętego na poziomie 0,05, co wskazywało że wariancja jest różna w każdej z grup.

W podsumowaniu można stwierdzić, że oba testy wykazały brak równości wariancji w poszczególnych grupach, co świadczyło, że nie występuje żadna zależność między zawartością popiołu w węglach z analizowanych kopalń.

#### 4.4. Analiza składowych głównych (PCA)

W analizach wielowymiarowych (w analizie danych) często ma się do czynienia z sytuacją, gdy zgromadzone dane dotyczą dużej liczby zmiennych, wzajemnie skorelowanych. Uwzględnienie większej liczby zmiennych pozwala na dokładniejszy i bardziej wiarygodny opis badanego zjawiska, co powoduje jednak większe komplikacje informacji wyjściowej, rosnące trudności obliczeniowe i (przede wszystkim) interpretacyjne. Nierzadko kolejne zmienne okazują się silnie skorelowane z pozostałymi, a ogólne prawidłowości w związkach między zmiennymi pozostają, wobec całkowitego ładunku niesionej informacji, niemożliwe do wykrycia.

W celu stwierdzenia istotnych zależności, jakie zachodzą między zmiennymi opisującymi obiekt wielowymiarowy, stosuje się techniki optymalnej redukcji wymiaru przestrzeni danych, co pozwala na kontrolę stopnia istotności zaniechanianej informacji. Jedną z takich technik jest analiza głównych składowych, polegająca na dokonaniu transformacji początkowych zmiennych w zbiór nowych, wzajemnie niezależnych zmiennych (składowych). Przeobrażone zmienne mają inne znaczenie merytoryczne, ale zachowują jednocześnie znaczną część informacji zawartych w zmiennych pierwotnych.

Analizą składowych zostały objęte wszystkie parametry jakościowe węgla kamiennych, wyprodukowanych w pięciu kopalniach w 2009 r., zawarte w tabeli 1. Po przeprowadzeniu skalowania danych z tabeli 1, uzyskano wyniki PCA, które przedstawiono w tabeli 18.

Tabela 18. Wyniki analizy PCA

Analiza składowych głównych - podsum. (Arkusz7)								
Liczba składowych: 3								
81,7705% sumy kwadratów wyjaśniane jest przez wyodrębnione składowe.								
Składowa	R2X	R2X(skumul.)	Wartości własne	Q2	Granica	Q2(skumul.)	Istotność	Iteracji
1	0,533965	0,533965	4,271716	0,390904	0,139831	0,390904	Tak	4
2	0,192269	0,726234	1,538152	0,173357	0,157635	0,496495	Tak	6
3	0,091471	0,817705	0,731769	-0,100000	0,181287	0,348536	NIEZNANA	13

Z tabeli tej wynika, że tylko trzy składowe wyjaśniają model, z czego składowa trzecia ma nieznaczną istotność. Wskaźnik  $R^2x$ , przedstawiony w tabelicy 18 stanowi procent (frakcję) wyjaśnionej wariancji. Składowa pierwsza wyjaśnia wariancję w 0,53%, składowa druga w 0,19%, a składowa trzecia w 0,09%.

Wartość  $Q^2$  stanowi procent (frakcję) przewidywanej wariancji. Składowa pierwsza wyjaśnia przewidywaną wariancję w 0,39%, a składowa druga w 0,17%. W przypadku, gdy  $Q^2$  jest liczbą ujemną lub gdy  $Q^2$  (skumulowane) maleje, nie należy dodawać więcej składowych, ponieważ będą one nieistotne. Potwierdzeniem tego są dane zawarte w tabeli 19.

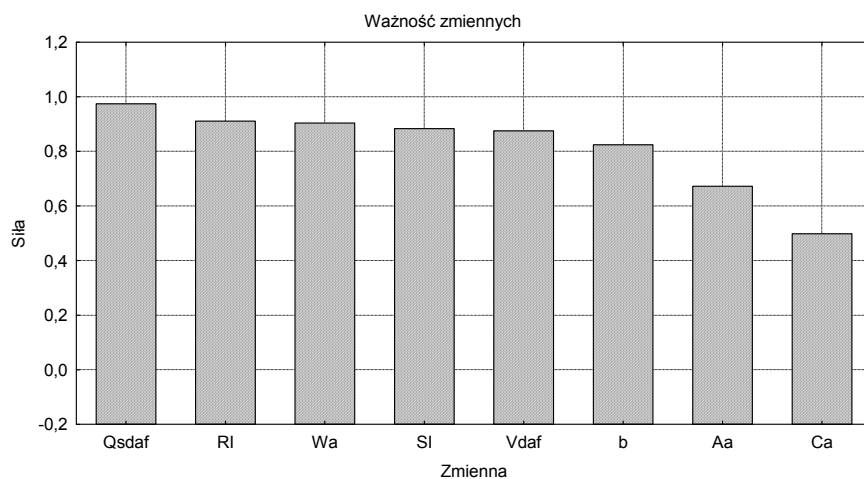
Tabela 19. Analiza PCA po dodaniu kolejnej składowej

Analiza składowych głównych - podsum. (Arkusz7)								
Liczba składowych: 4								
89,8699% sumy kwadratów wyjaśniane jest przez wyodrębnione składowe.								
Składowa	R2X	R2X(skumul.)	Wartości własne	Q2	Granica	Q2(skumul.)	Istotność	Iteracji
1	0,533965	0,533965	4,271716	0,390904	0,139831	0,390904	Tak	4
2	0,192269	0,726234	1,538152	0,173357	0,157635	0,496495	Tak	6
3	0,091471	0,817705	0,731769	-0,100000	0,181287	0,348536	NIEZNANA	13
4	0,080995	0,898699	0,647958	0,057497	0,214286	0,385994	Nie	10

W tabeli 20 i na rysunku 8 pokazano ważność zmiennych. Największą siłą oddziaływania ma zmienna  $Q_s^{daf}$ . Każda następna zmienna ma mniejszą siłę, ale pierwszych sześć zmiennych ma siłę oddziaływania powyżej 0,8.

Tabela 20. Ważność zmiennych

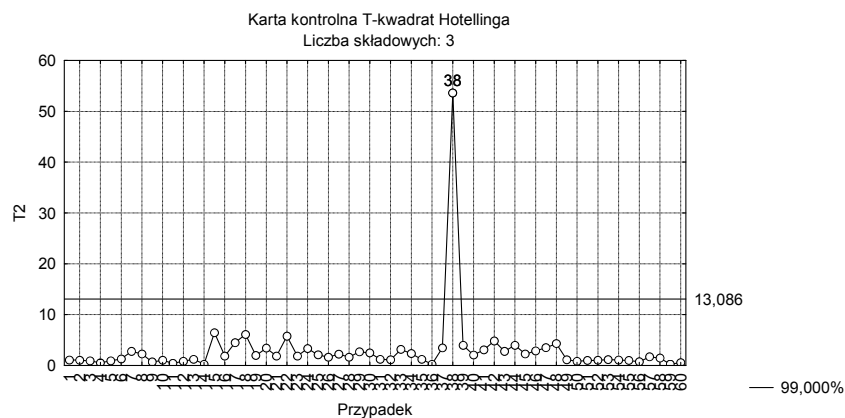
Ważność zmiennych (Arkusz7)			
Liczba składowych: 3			
Zmienna	Numer zmiennej	Siła	Ważność
Qsdaf	4	0,974321	1
RI	6	0,911052	2
Wa	1	0,903625	3
SI	7	0,883463	4
Vdaf	3	0,875108	5
b	8	0,823804	6
Aa	2	0,672086	7
Ca	5	0,498178	8



**Rys. 8.** Graficzne przedstawienie ważności zmiennych

**Fig. 8.** Graphical presentation of variable importance

Na rysunku 9 pokazano kartę T-kwadrat Hottellinga (dane odstające), a w tabeli 21 – kombinację liczbową wyjściowego układu (ładunki niesionych informacji).



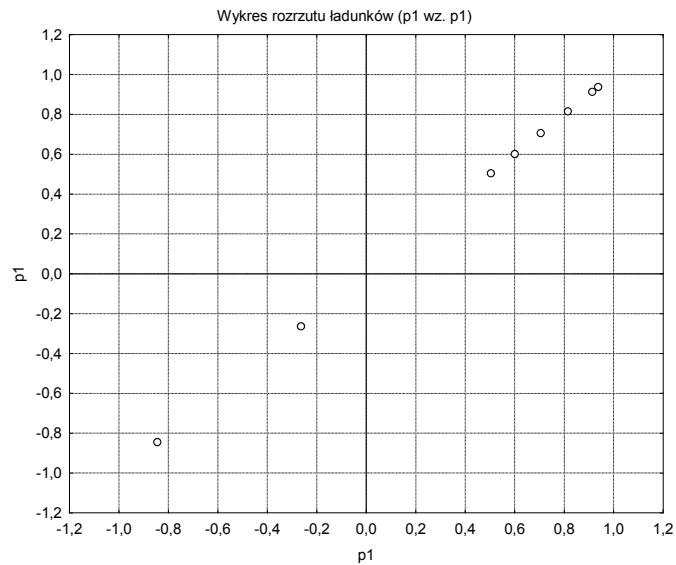
**Rys. 9.** Karta T-kwadrat Hotellinga

**Fig. 9.** Hotelling's T-square chart

Tabela 21. Ładunki niesionych informacji

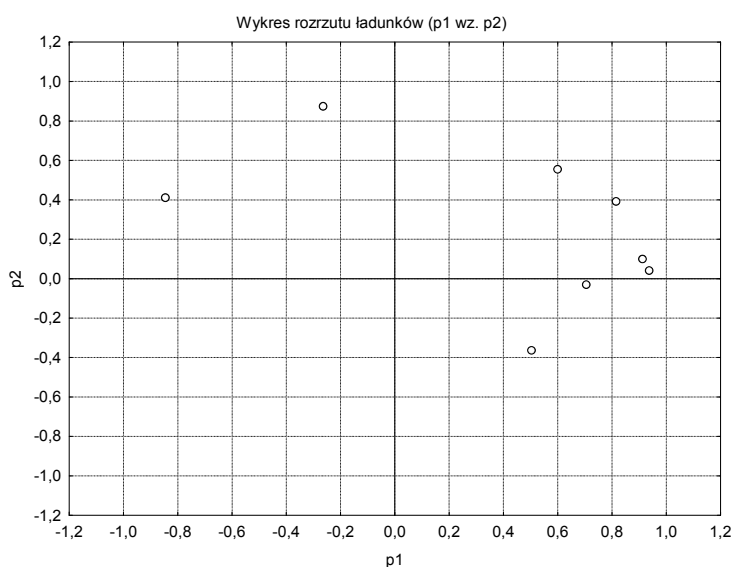
Zmienna	Ładunki (Arkusze7)			
	Numer zmiennej	Składowa 1	Składowa 2	Składowa 3
Wa	1	-0,844811	0,410109	0,148462
Aa	2	0,600184	0,553994	0,063289
Vdaf	3	-0,263945	0,874636	0,208393
Qsdaf	4	0,503908	-0,363212	0,767141
Ca	5	0,705262	-0,030653	0,046749
RI	6	0,937461	0,041449	-0,168895
SI	7	0,912846	0,099887	-0,194591
b	8	0,815282	0,391801	0,072204

Na rysunkach od 10 do 15 zilustrowano rozrzuty między ładunkami niesionych informacji  $p$  dla poszczególnych składowych wyjaśniających model PCA.

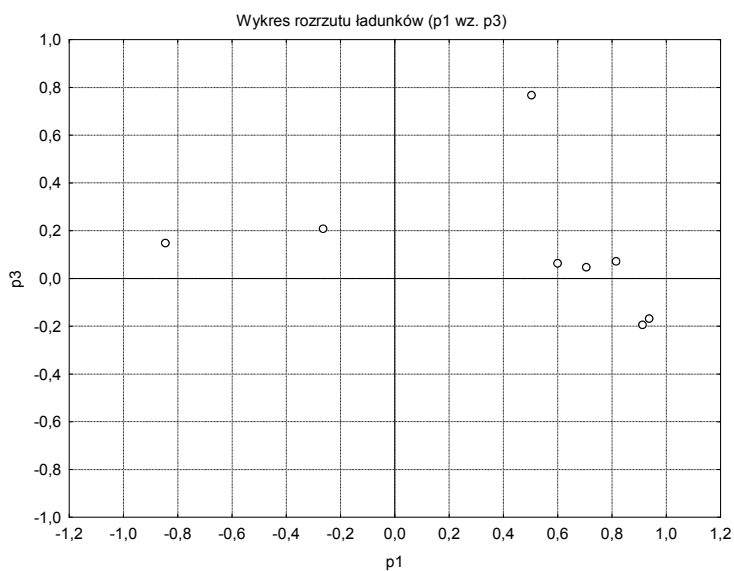


Rys. 10. Wykres rozrzutu ładunków p1 (pierwsza składowa) względem p1

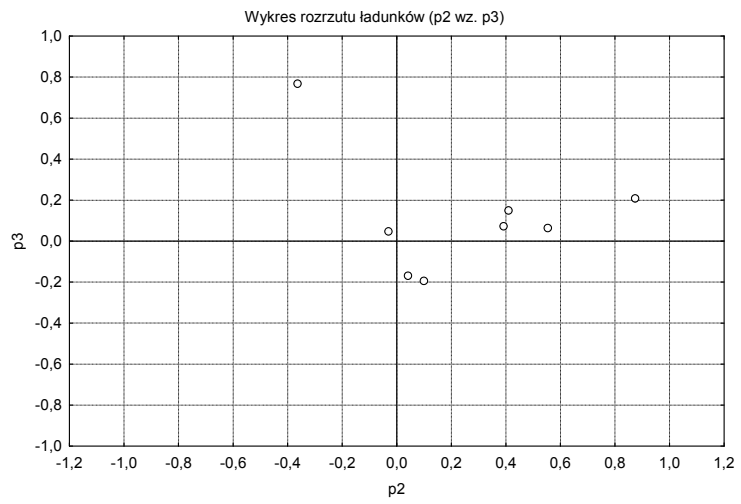
Fig. 10. Graph of charge dispersion p1 (first component) towards p1



**Rys. 11.** Wykres rozrzutu ładunków p1 względem p2 (druga składowa)  
**Fig. 11.** Graph of charge dispersion p1 towards p2 (second component)

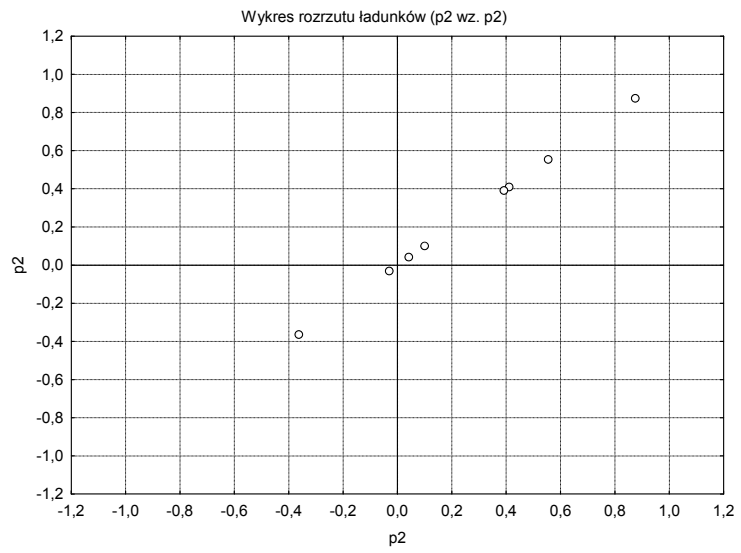


**Rys. 12.** Wykres rozrzutu ładunków p1 względem p3 (trzecia składowa)  
**Fig. 12.** Graph of charge dispersion p1 towards p3 (third component)



**Rys. 13.** Wykres rozrzutu ładunków p2 względem p3

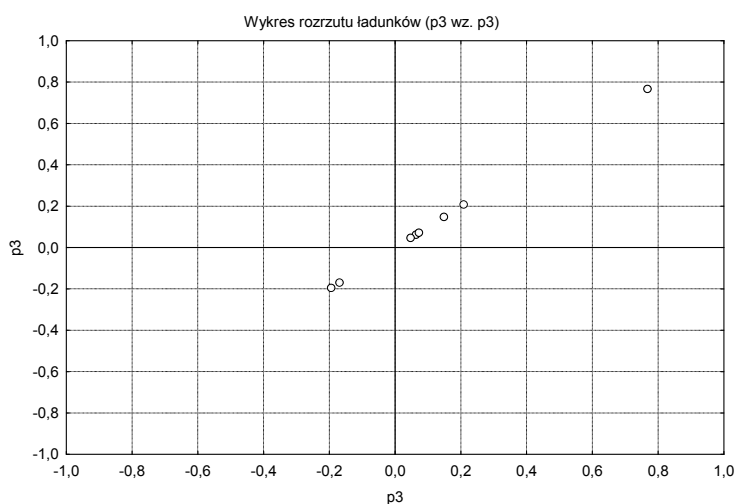
**Fig. 13.** Graph of charge dispersion p2 towards p3



**Rys. 14.** Wykres rozrzutu ładunków p2 względem p2

**Fig. 14.** Graph of charge dispersion p2 towards p2

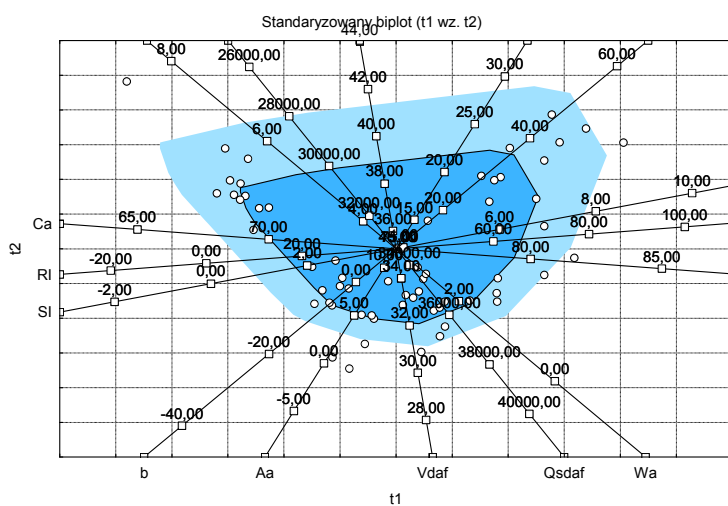




Rys. 15. Wykres rozrzutu ładunków p3 względem p3

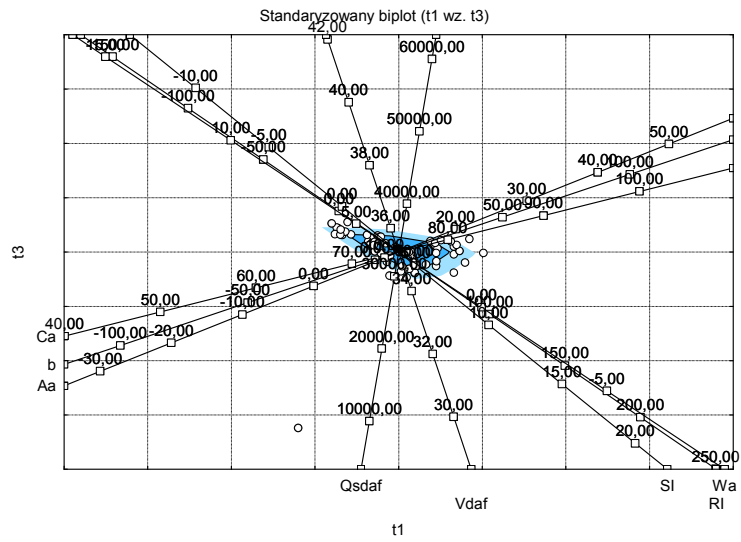
Fig. 15. Graph of charge dispersion p3 towards p3

Na rysunkach od 16 do 18 natomiast przedstawiono rozrzuty ładunków niesionych informacji wszystkich analizowanych zmiennych w postaci grafów Biplot, względem trzech składowych głównych (t1, t2 i t3) wyjaśniających model PCA. Grafy Biplot umożliwiają umieszczenie wszystkich zmiennych na jednym wykresie.

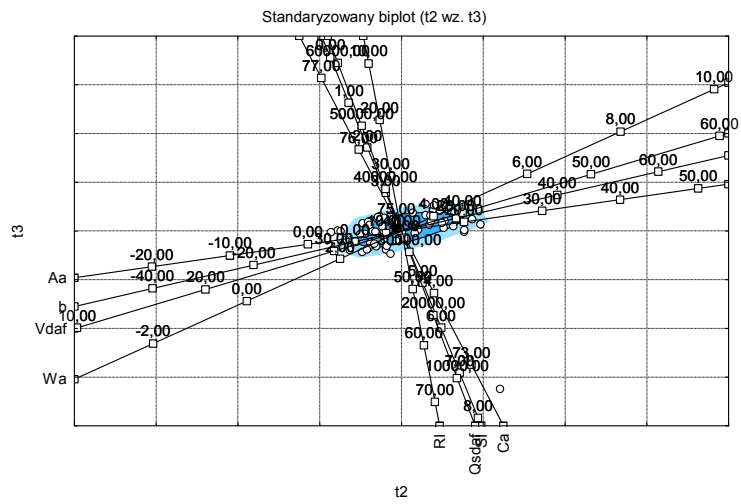


Rys. 16. Biplot t1 względem t2

Fig. 16. Biplot t1 towards t2



Rys. 17. Biplot t1 względem t3  
Fig. 17. Biplot t1 towards t3



Rys. 18. Biplot t2 względem t3  
Fig. 18. Biplot t2 towards t3

W posumowaniu należy stwierdzić, że w analizowanym modelu PCA tylko trzy składowe główne są istotne ( $Q_s^{daf}$ ,  $RI$  i  $W^a$ ), z czego trzecia składowa ma istotność nieznaczną.

## 5. WNIOSKI

- Na podstawie testu  $t$ -Studenta stwierdzono, że jedynie w przypadku kopalni D i kopalni E należałoby odrzucić hipotezę zerową, zgodnie z którą średnia zawartość popiołu w węglach wyprodukowanych w latach 2005–2008 i średnia zawartość popiołu w węglach wyprodukowanych w 2009 r. są jednakowe. Świadczy to o tym, że w kopalniach tych nastąpiła istotna zmiana jakości węgla wyprodukowanych w 2009 r. W przypadku pozostałych kopalń nie ma podstaw do odrzucenia hipotezy zerowej.
- W regresji wielorakiej stwierdzono, że między każdą ze zmiennych (zdolnością spiekania węgla  $RI$ , zawartością pierwiastka węgla  $C^a$  oraz zawartością części lotnych  $V^{daf}$ ) istnieją niewielkie zależności. Oddziaływanie zawartości pierwiastka węgla  $C^a$  na zdolność spiekania węgla  $RI$ , a tym samym na wskaźnik typu węgla, jest istotne. Po odrzuceniu natomiast parametrów próbek 8, 25, 31 i 39 również wyraz wolny staje się statystycznie istotny.
- Analiza wariancji pozwala stwierdzić, że oba testy wykazały brak równości wariancji w poszczególnych grupach, co świadczy, że nie występuje żadna zależność między zawartością popiołu w analizowanych kopalniach.
- Analiza składowych głównych pozwala stwierdzić, że w analizowanym modelu PCA tylko trzy główne są istotne ( $W^a$ ,  $A^a$  i  $V^{daf}$ ), z czego trzecia składowa ma istotność nieznaczną.
- Pakiet komputerowy STATISTICA jest jednym ze współczesnych narzędzi statystycznych, pomocnych w analizie dużej liczby danych i ich przetwarzaniu. W pakiecie tym są wykorzystywane różnego rodzaju testy statystyczne, które pozwalają na scharakteryzowanie właściwości badanych węgla, na podstawie opisu zależności między analizowanymi danymi.

### Literatura

1. Jasieńko S. (1995): Chemia i fizyka węgla. Wrocław, Oficyna Wydaw. Politechniki Wrocławskiej.
2. Luszniwicz A., Słaby T. (2008): Statystyka z pakietem komputerowym Statistica pl. Teoria i zastosowanie. Warszawa, Wydaw. C.H. Beck.
3. Praktyczne wykorzystanie analiz danych i data mining. Materiały na seminarium. Kraków, Wydaw. StatSoft 2008.
4. Zastosowanie statystyki i data mining w badaniach naukowych. Materiały na seminarium. Kraków, Wydaw. StatSoft 2008.
5. [http://pl.wikipedia.org/wiki/Liczba\\_stopni\\_swobody\\_\(statystyka\)](http://pl.wikipedia.org/wiki/Liczba_stopni_swobody_(statystyka)).
6. <http://zsi.tech.us.edu.pl/~nowak/adb/influence.pdf>.

**Recenzent:** dr inż. Tomasz Janoszek