

# Wielowymiarowa analiza porównawcza

Dr hab. Inż. Piotr Zapotoczny, prof. UWM

- Eksploracja danych odnosi się do poszukiwania lub odkrywania związków, zależności, schematów w danych. Jest analizą zbiorów danych obserwacyjnych w celu znalezienia nieoczekiwanych związków i podsumowania danych w oryginalnych sposób, tak aby były zarówno zrozumiałe, jak i przydatne dla ich właściciela.

# Wielowymiarowa analiza porównawcza

- Dyscyplina naukowa zajmująca się porównywaniem obiektów określonych za pomocą wielu cech
- Jeden z jej działów to **metody grupowania**, służące do badania podobieństw obiektów pod względem poziomu zjawiska złożonego
- Metody grupowania można podzielić na **dyskryminacyjne i klasyfikacyjne**

# Dyskryminacja a klasyfikacja

- **Dyskryminacja** – przydział obiektów do znanych wcześniej klas
- **Klasyfikacja** – podział obiektów na nieznane wcześniej klasy
- Często proces podziału obiektów na klasy jest nazywany klasyfikacją, niezależnie od tego, czy klasy te zostały wcześniej ustalone, czy nie



Terminem dane określać będziemy zbiór wektorów cech  
(atrybutów, parametrów, obserwacji) ilościowo  
opisujących analizowane obiekty, którymi mogą być  
wszelkie przedmioty i zjawiska fizyczne, rzeczy,  
organizmy żywe, a także ludzie. Cechy odwzorowują  
różnego rodzaju właściwości obiektów (np. wiek,  
rozmiar, skład chemiczny, parametry elektryczne,  
porowatość powierzchni itp.), które można zmierzyć i  
wyrazić liczbowo.

# Czym jest analiza dyskryminacyjna?

- **Zbiór** metod pozwalający na znalezienie kombinacji cech, które najlepiej rozróżniają dwie lub więcej klas obiektów lub zdarzeń
- **Zbiór**, bo istnieje kilka wariantów analizy dyskryminacyjnej zależnych od posiadanych informacji na temat badanej populacji i przyjętych założeń

O grupie obiektów podobnych pod względem określonych właściwości mówi się, że tworzą one odrębną klasę, lub kategorię w zbiorze danych, przy czym określanie klas obiektów odbywa się na podstawie etykiet kategorii wektorów cech. Obiekty nie należące do tej samej klasy wykazują duże różnice w wartościach (przynajmniej części) atrybutów.

Próba wektorów cech, których przynależność do odpowiednich klas jest znana, nazywana jest zbiorem danych treningowych. Na jego podstawie algorytm klasyfikacji (klasyfikator) konstruuje regułę decyzyjną, pozwalającą przypisywać etykiety kategorii nowym wektorom cech, które nie należą do zbioru danych treningowych.

Program komputerowy, który realizuje zadanie klasyfikacji w oparciu o tak wyuczoną regułę, określa się mianem systemu uczącego się pod nadzorem.

Spośród algorytmów klasyfikacji nadzorowanej wymienić należy m.in.:

- wielowarstwowe sieci perceptronowe,
- sieci neuronowe z radialną funkcją bazową,
- sieci bayesowskie,
- maszyny wektorów podpierających (ang. support vector machines)
- systemy działające w oparciu o regułę „k” najbliższych sąsiadów (ang. k nearest neighbor)

Z drugiej strony, gdy żadna wiedza *a priori* o obiektach nie jest dostępna, klasyfikator dokonuje ich podziału jedynie na podstawie wykrytych statystycznych zależności między wektorami cech — w tym przypadku mowa jest o uczeniu *nienadzorowanym*.

*Klasyfikację nienadzorowaną można realizować za pomocą: algorytmów analizy skupień (ang. *Cluster analysis*), zwanych także metodami *klasteryzacji* lub *grupowania*. Wśród nich można wyróżnić grupowanie metodą *średnich*, algorytm *Expectation Maximization* z wykorzystaniem *gaussowskiego modelu mieszaniny* (ang. *Gaussian mixture model*) lub np. algorytmy *klasteryzacji hierarchicznej*.*

Z praktycznego punktu widzenia, kwestią o zasadniczym znaczeniu jest dokładność klasyfikatora, czyli jego zdolność **przypisywania** obiektom poprawnych etykiet kategorii. Dokładność systemów wyuczonych pod nadzorem można oszacować licząc dla zbioru treningowego *błąd klasyfikacji (lub inaczej ryzyko empiryczne)*, wyrażający się liczbą niewłaściwie sklasyfikowanych wektorów cech w stosunku do wszystkich dostępnych wektorów.

Poprawność klasyfikacji, zarówno nadzorowanej jak i nienadzorowanej, zależy w dużym stopniu od atrybutów opisujących obiekty, a ściślej rzecz ujmując, od tego na ile zawarta w nich informacja koresponduje z prawdziwym podziałem zbioru danych na klasy. O cechach, które posiadają takie właściwości, mówi się, że są znaczące lub informatywne, zaś o całym ich zbiorze (bądź przestrzeni), że ma dużą zdolność dyskryminacyjną.



Podejmując decyzję o opisie analizowanych obiektów za pomocą określonych cech, trudno najczęściej przewidzieć, które z nich będą najlepsze do dyskryminacji klas. Dlatego dokonuje się pomiaru wielu parametrów w nadziei, że znajdą się wśród nich również takie, które umożliwią poprawną klasyfikację.

- Założenia analizy wielowymiarowych

1. Minimalna ilość obiektów. Stwierdzono, że aby błąd klasyfikacji na zbiorze uczącym nie odbiegał znacząco (tzn. więcej niż 15%) od błędu optymalnego klasyfikatora Bayesa, spełniony musi być warunek  $Q/N > 3$  dla każdej klasy, gdzie  $Q$  — liczba obiektów,  $N$  — liczba cech.

2. Problemem przewagi atrybutów nieistotnych nad znaczącymi — gdy jest ona zbyt duża, wektory cech reprezentujące różne klasy są nierozróżnialne. Wówczas, nawet jeśli cechy znaczące mają bardzo dużą zdolność dyskryminacyjną, informacja w nich zawarta staje się nieczytelna dla systemu uczącego.

- **Redukcja wymiarowości przestrzeni cech**

Metody redukcji wymiarowości przestrzeni obserwacji można podzielić na dwie główne podgrupy: ekstrakcję oraz selekcję cech.

- **ekstrakcja**

Polega na liniowej, bądź nieliniowej transformacji wektorów cech, prowadzącej do opisu obiektów w przestrzeni  $n$  nowych parametrów ( $n < N$ ), będących agregatami obserwacji pochodzących z pomiaru. Spośród metod należących do tej podgrupy warto wymienić przede wszystkim analizę składowych głównych PCA (ang. *principal component analysis*), liniową analizę dyskryminacyjną LDA (ang. *linear discriminant analysis*), a także nieliniową analizę dyskryminacyjną (ang. *nonlinear discriminant analysis*)

## • selekcja cech

Prowadzi do wyznaczenia, spośród wszystkich dostępnych, podzbioru niewielu cech zapewniających maksymalną poprawność klasyfikacji

Najczęściej wymieniane metody tworzące ranking cech wykorzystują:

współczynnik	Fishera,
prawdopodobieństwo błędu klasyfikacji	wraz z
zakumulowanym współczynnikiem	korelacji,
współczynnik informacji wzajemnej.	

## • Założenia analizy dyskryminacyjnej

### • **Wielowymiarowa normalność rozkładu normalnego**

Zakłada się, że dane (ujęte w postaci zmiennych) reprezentują próbę z wielowymiarowego rozkładu normalnego.

### • **Jednorodność wariancji/kowariancji**

Zakłada się, że macierze wariancji/kowariancji zmiennych są jednorodne (homogeniczne) w grupach, nieznaczące odchylenia nie są aż tak ważne zwłaszcza przy dużej liczebności grup.

### • **Korelacje między średnimi i wariancjami**

Rzeczywiste zagrożenie dla trafności testów istotności pojawia się wówczas, gdy średnie zmiennych w grupach są skorelowane z wariancjami (lub odchyleniami standardowymi). Intuicyjnie, jeśli w grupie występuje duża zmienność przy szczególnie wysokich średnich niektórych zmiennych, to te wysokie średnie nie są rzetelne. W praktyce taki układ może się pojawić wtedy, gdy jedna z badanych grup zawiera kilka odstających obserwacji, które mają duży wpływ na średnie, a także zwiększają zmienność.

- Założenia analizy dyskryminacyjnej

- **Problem złego uwarunkowania macierzy.**

Inne założenie analizy funkcji dyskryminacyjnej wymaga, by zmienne wykorzystane do dyskryminacji grup nie były całkowicie redundantne. W ramach obliczeń związanych z analizą dyskryminacyjną występuje operacja odwracania macierzy wariancji/kowariancji zmiennych w modelu. Jeśli któraś ze zmiennych jest całkowicie redundantna wobec innych zmiennych, to o macierzy mówi się, że jest *źle uwarunkowana* i nie może być odwrócona. Na przykład, jeśli zmienna jest sumą trzech innych zmiennych, które także znajdują się w modelu, to macierz jest źle uwarunkowana.

- **Rozmiar próby**

Regułą jest, aby liczebność grupy miała 4 do 5 razy więcej przypadków niż liczba zmiennych dyskryminujących. Liczebność grupy nie jest ważna.

- **Wartości odstające**

Analiza dyskryminacyjna wrażliwa jest na wartości dostające. Przed analizą należy takie wartości usunąć.

- Program statistica

StatSoft, Inc. (2011). STATISTICA  
(data analysis software system),  
version 10. [www.statsoft.com](http://www.statsoft.com).



- Plik danych: Dane\_doktoranci\_1.xls

F:\Dydaktyka\KAO\_cwiczenia\Dane\_doktoranci\_1.xls : Arkusz1

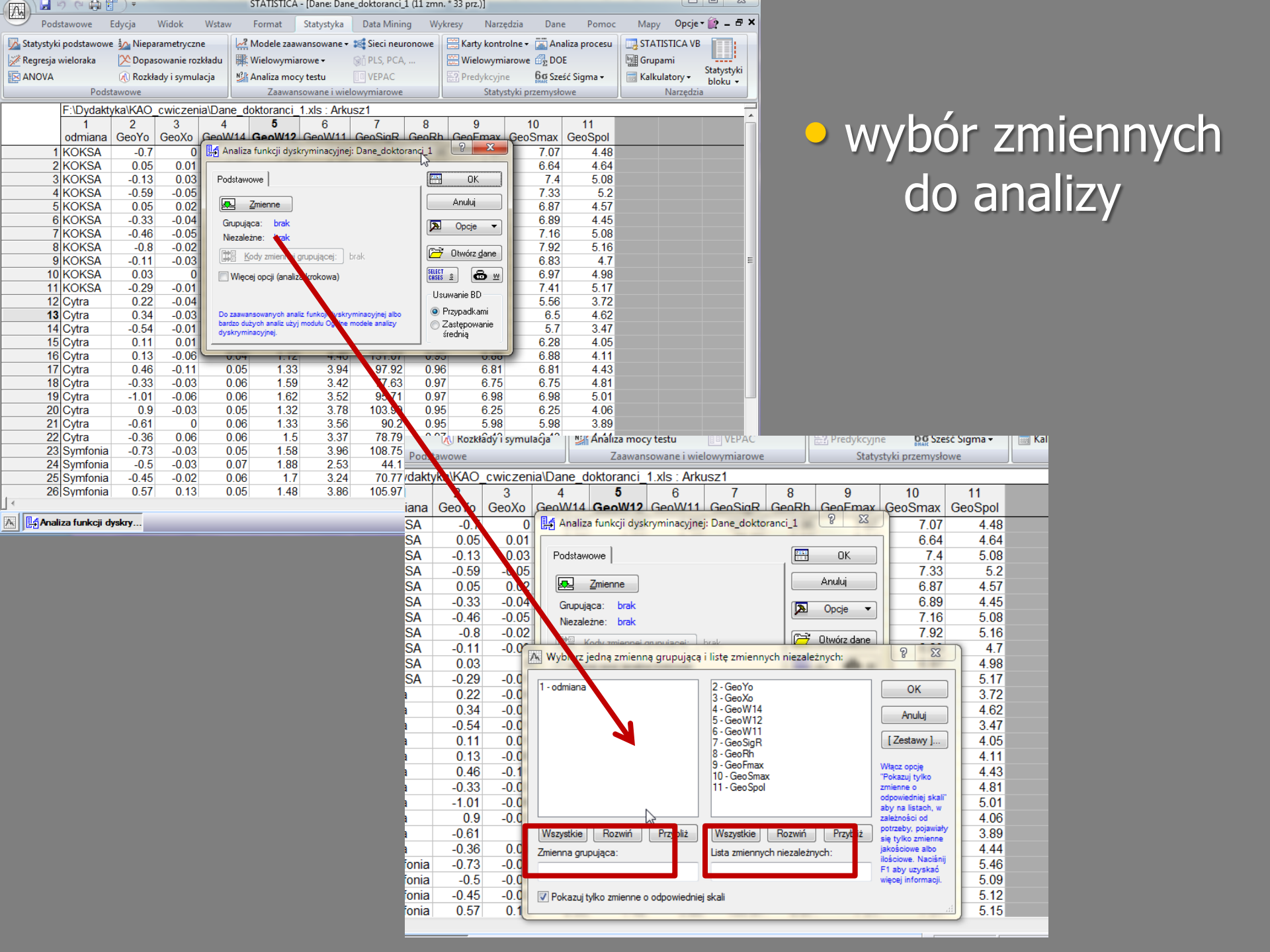
	1	2	3	4	5	6	7	8	9	10	11
	odmiana	GeoYo	GeoXo	GeoW14	GeoW12	GeoW11	GeoSigR	GeoRh	GeoFmax	GeoSmax	GeoSpol
1	KOKSA	-0.7	0	0.04	1.26	4.26	121.13	0.96	7.07	7.07	4.48
2	KOKSA	0.05	0.01	0.06	1.53	3.43	79.89	0.97	6.64	6.64	4.64
3	KOKSA	-0.13	0.03	0.05	1.48	4.06	109.8	0.97	7.4	7.4	5.08
4	KOKSA	-0.59	-0.05	0.05	1.58	3.86	106.77	0.97	7.33	7.33	5.2
5	KOKSA	0.05	0.02	0.05	1.39	3.81	98.6	0.96	6.87	6.87	4.57
6	KOKSA	-0.33	-0.04	0.05	1.31	4.21	122.34	0.95	6.89	6.89	4.45
7	KOKSA	-0.46	-0.05	0.06	1.58	3.77	101.47	0.97	7.16	7.16	5.08
8	KOKSA	-0.8	-0.02	0.04	1.33	4.71	144.95	0.96	7.92	7.92	5.16
9	KOKSA	-0.11	-0.03	0.05	1.49	3.64	90.51	0.97	6.83	6.83	4.7
10	KOKSA	0.03	0	0.06	1.6	3.63	92.62	0.97	6.97	6.97	4.98
11	KOKSA	-0.29	-0.01	0.05	1.53	3.68	87.73	0.97	7.41	7.41	5.17
12	Cytra	0.22	-0.04	0.06	1.4	3.11	68.11	0.96	5.56	5.56	3.72
13	Cytra	0.34	-0.03	0.06	1.59	3.42	83.93	0.97	6.5	6.5	4.62
14	Cytra	-0.54	-0.01	0.05	1.16	3.67	83.1	0.95	5.7	5.7	3.47
15	Cytra	0.11	0.01	0.05	1.31	3.8	96.23	0.96	6.28	6.28	4.05
16	Cytra	0.13	-0.06	0.04	1.12	4.46	131.07	0.95	6.88	6.88	4.11
17	Cytra	0.46	-0.11	0.05	1.33	3.94	97.92	0.96	6.81	6.81	4.43
18	Cytra	-0.33	-0.03	0.06	1.59	3.42	77.63	0.97	6.75	6.75	4.81
19	Cytra	-1.01	-0.06	0.06	1.62	3.52	95.71	0.97	6.98	6.98	5.01
20	Cytra	0.9	-0.03	0.05	1.32	3.78	103.99	0.95	6.25	6.25	4.06
21	Cytra	-0.61	0	0.06	1.33	3.56	90.2	0.95	5.98	5.98	3.89
22	Cytra	-0.36	0.06	0.06	1.5	3.37	78.79	0.97	6.43	6.43	4.44
23	Symfonia	-0.73	-0.03	0.05	1.58	3.96	108.75	0.97	7.69	7.69	5.46
24	Symfonia	-0.5	-0.03	0.07	1.88	2.53	44.1	0.99	6.58	6.58	5.09
25	Symfonia	-0.45	-0.02	0.06	1.7	3.24	70.77	0.98	6.96	6.96	5.12
26	Symfonia	0.57	0.13	0.05	1.48	3.86	105.97	0.97	7.51	7.51	5.15
27	Symfonia	-0.51	0.02	0.07	1.8	2.98	61.98	0.98	6.71	6.71	5.08
28	Symfonia	-0.69	-0.05	0.06	1.67	3.42	83.68	0.97	7.01	7.01	5.11
29	Symfonia	-0.63	0.02	0.06	1.6	3.35	73.98	0.98	7.01	7.01	4.99
30	Symfonia	-0.61	-0.03	0.06	1.64	3.15	64.92	0.98	6.42	6.42	4.64
31	Symfonia	0.22	0.01	0.05	1.58	3.6	84.93	0.97	7.2	7.2	5.11
32	Symfonia	-0.41	0.05	0.07	1.74	2.78	51.87	0.98	6.05	6.05	4.5
33	Symfonia	0.71	0	0.05	1.6	3.95	108.61	0.97	7.61	7.61	5.43

Podstawowe Edycja Widok Wstaw Format Statystyka Data Mining Wykresy Narzędzia Dane Pomoc Mapy Opcje

Statystyki podstawowe Nieparametryczne Modele zaawansowane Sieci neuronowe Karty kontrolne Analiza procesu  
 Regresja wieloraka Dopasowanie rozkładu Wielowymiarowe PLS, PCA... Wielowymiarowe DOE  
 ANOVA Rozkłady i symulacja Predykcyjne Sześć Sigma  
 Podstawowe Statystyki przemysłowe STATISTICA VB  
 Grupami  
 Kalkulatory Statystyki bloku  
 Narzędzia

- Wielowymiarowe
- Analiza skupień
- Analiza czynnikowa
- Analiza składowych głównych i klasyfikacja
- Analiza kanoniczna
- Analiza rzetelności i pozycji
- Drzewa klasyfikacyjne
- Analiza korespondencji
- Skalowanie wielowymiarowe
- Analiza dyskryminacyjna
- Ogólne modele analizy dyskryminacyjnej

	1	2	3	4		9	10	11
	odmiana	GeoYo	GeoXo	GeoW		GeoFmax	GeoSmax	GeoSpol
1	KOKSA	-0.7	0	0		7.07	7.07	4.48
2	KOKSA	0.05	0.01	0		6.64	6.64	4.64
3	KOKSA	-0.13	0.03	0		7.4	7.4	5.08
4	KOKSA	-0.59	-0.05	0		7.33	7.33	5.2
5	KOKSA	0.05	0.02	0		6.87	6.87	4.57
6	KOKSA	-0.33	-0.04	0		6.89	6.89	4.45
7	KOKSA	-0.46	-0.05	0		7.16	7.16	5.08
8	KOKSA	-0.8	-0.02	0.01	1.33	4.71	144.95	0.96
9	KOKSA	-0.11	-0.03	0.05	1.49	3.64	90.51	0.97
10	KOKSA	0.03	0	0.06	1.6	3.63	92.62	0.97
11	KOKSA	-0.29	-0.01	0.05	1.53	3.68	87.73	0.97
12	Cytra	0.22	-0.04	0.06	1.4	3.11	68.11	0.96
13	Cytra	0.34	-0.03	0.06	1.59	3.42	83.93	0.97
14	Cytra	-0.54	-0.01	0.05	1.16	3.67	83.1	0.95
15	Cytra	0.11	0.01	0.05	1.31	3.8	96.23	0.96
16	Cytra	0.13	-0.06	0.04	1.12	4.46	131.07	0.95
17	Cytra	0.46	-0.11	0.05	1.33	3.94	97.92	0.96
18	Cytra	-0.33	-0.03	0.06	1.59	3.42	77.63	0.97
19	Cytra	-1.01	-0.06	0.06	1.62	3.52	95.71	0.97
20	Cytra	0.9	-0.03	0.05	1.32	3.78	103.99	0.95
21	Cytra	-0.61	0	0.06	1.33	3.56	90.2	0.95
22	Cytra	-0.36	0.06	0.06	1.5	3.37	78.79	0.97
23	Symfonia	-0.73	-0.03	0.05	1.58	3.96	108.75	0.97
24	Symfonia	-0.5	-0.03	0.07	1.88	2.53	44.1	0.99
25	Symfonia	-0.45	-0.02	0.06	1.7	3.24	70.77	0.98
26	Symfonia	0.57	0.13	0.05	1.48	3.86	105.97	0.97



• wybór zmiennych do analizy

# Wybór obiektów do analizy

**Podstawowe**

Zmienne

Grupująca: odmiana

Niezależne: GeoYo-GeoSpol

Kody zmiennej grupującej: brak

Więcej opcji (analiza krokowa)

Do zaawansowanych analiz funkcji dyskryminacyjnej, bardzo dużych analiz użyj modułu Ogólne metody dyskryminacyjnej.

**Wybierz kody zmiennej grupującej:**

odmiana

Wszystko Przybliż OK Anuluj

**Wartości, statyst.: Zmienna 1**

Nazwa: odmiana

Długa nazwa: (Brak długiej nazwy)

RD: -999999998

Typ: ...

Format: General

Wyślij do arkusza

101 KOKSA

102 Cytra

103 Symfonia

Statystyki opisowe

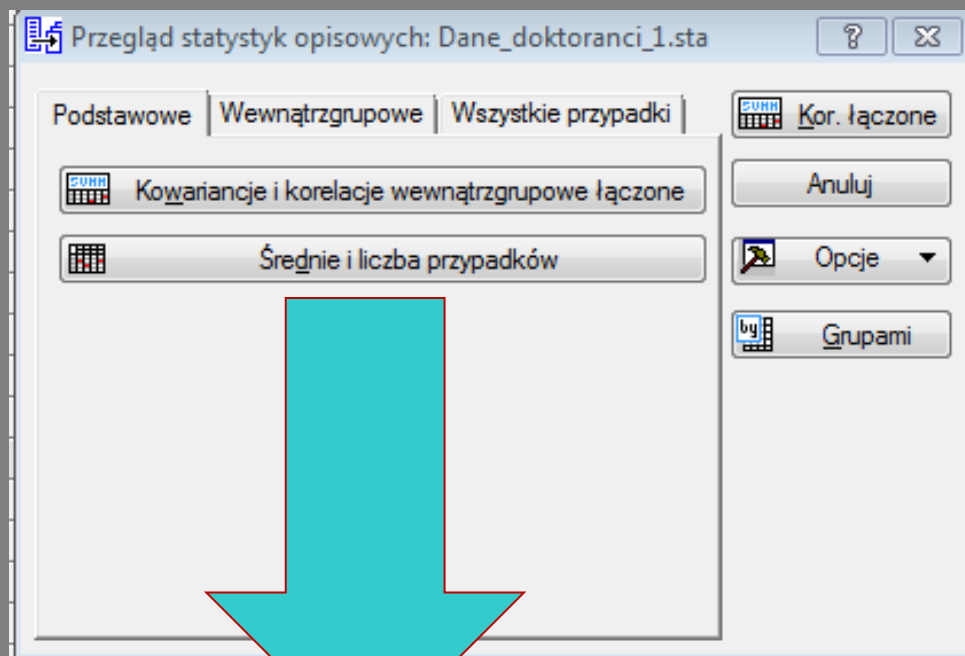
N = 33

Średnia = 102.000000

Odch.std. = 0.829156

	7	8	9	10	11
	GeoSigR	GeoRh	GeoFmax	GeoSmax	GeoSpol
1	121.13	0.96	7.07	7.07	4.48
2	79.89	0.97	6.64	6.64	4.64
3	109.8	0.97	7.4	7.4	5.08
4	106.77	0.97	7.33	7.33	5.2
5	6.8	6.8	6.8	6.8	6.8
6	7.1	7.1	7.1	7.1	7.1
7	7.9	7.9	7.9	7.9	7.9
8	6.8	6.8	6.8	6.8	6.8
9	6.8	6.8	6.8	6.8	6.8
10	7.4	7.4	7.4	7.4	7.4
11	5.5	5.5	5.5	5.5	5.5
12	6.2	6.2	6.2	6.2	6.2
13	6.2	6.2	6.2	6.2	6.2
14	6.8	6.8	6.8	6.8	6.8
15	6.8	6.8	6.8	6.8	6.8
16	6.7	6.7	6.7	6.7	6.7
17	6.9	6.9	6.9	6.9	6.9
18	6.2	6.2	6.2	6.2	6.2
19	5.9	5.9	5.9	5.9	5.9
20	6.4	6.4	6.4	6.4	6.4
21	7.6	7.6	7.6	7.6	7.6
22	6.5	6.5	6.5	6.5	6.5
23	6.9	6.9	6.9	6.9	6.9
24	7.5	7.5	7.5	7.5	7.5
25	7.5	7.5	7.5	7.5	7.5
26	7.5	7.5	7.5	7.5	7.5

# Przeglądanie statystyk opisowych.



odmiana	Średnie (Dane_doktoranci_1.sta)										
	GeoYo	GeoXo	GeoW14	GeoW12	GeoW11	GeoSigR	GeoRh	GeoFmax	GeoSmax	GeoSpol	N ważn.
KOKSA	-0.298182	-0.012727	0.050909	1.461818	3.914546	105.0736	0.965454	7.135455	7.135455	4.864545	11
Cytra	-0.062727	-0.027273	0.054545	1.388182	3.640909	91.5164	0.960000	6.374546	6.374546	4.237273	11
Symfonia	-0.275455	0.006364	0.059091	1.660909	3.347273	78.1418	0.976364	6.977273	6.977273	5.061818	11
Ogół grp	-0.212121	-0.011212	0.054848	1.503636	3.634243	91.5773	0.967273	6.829091	6.829091	4.721212	33



# • Tworzenie histogramu z arkusza.

Średnie (Dane_doktoranci_1.sta)								
odmiana	GeoYo	GeoXo	GeoW14	GeoW12	GeoW11	GeoSigR	GeoRh	GeoFm
KOKSA	-0.298182	-0.012727	0.050909	1.461				7.1354
Cytra	-0.062727	-0.027273	0.054545	1.388				6.3749
Symfonia	-0.275455	0.006364	0.059091	1.660				6.0777
Ogół grp	-0.212121	-0.011212	0.054848	1.503				

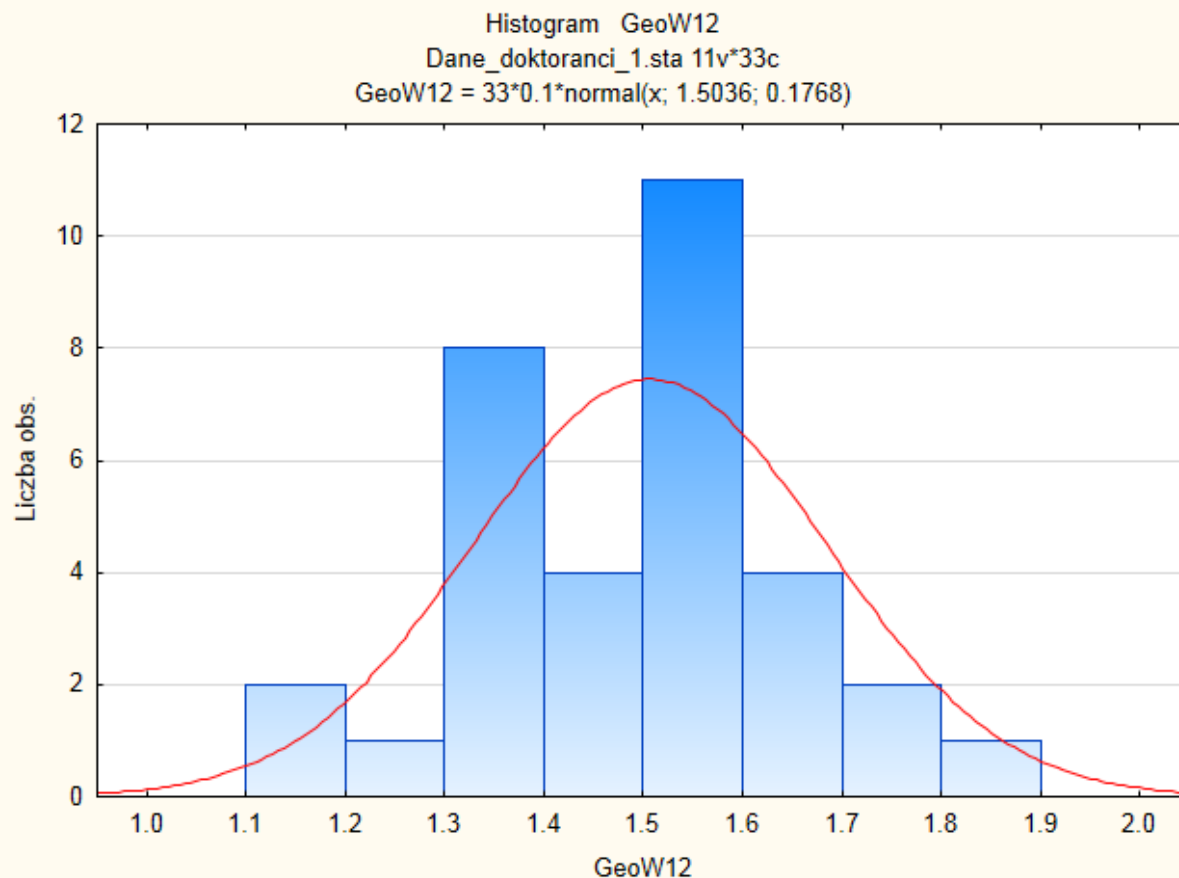
  

Odmiana	Dł działki	Szer działki	Dł płatka
SETOSA	5,006000	3,428000	1,462000
VERSICOL	5,936000	2,770000	4,260000
VIRGINIC	6,588000	2,974000	5,552000

Wykresy danych wejściowych

- Wartości, statystyki GeoW12...
- Histogram GeoW12**
- Ramka-wąsy... GeoW12
- Wykres prawdopodobieństwa GeoW12

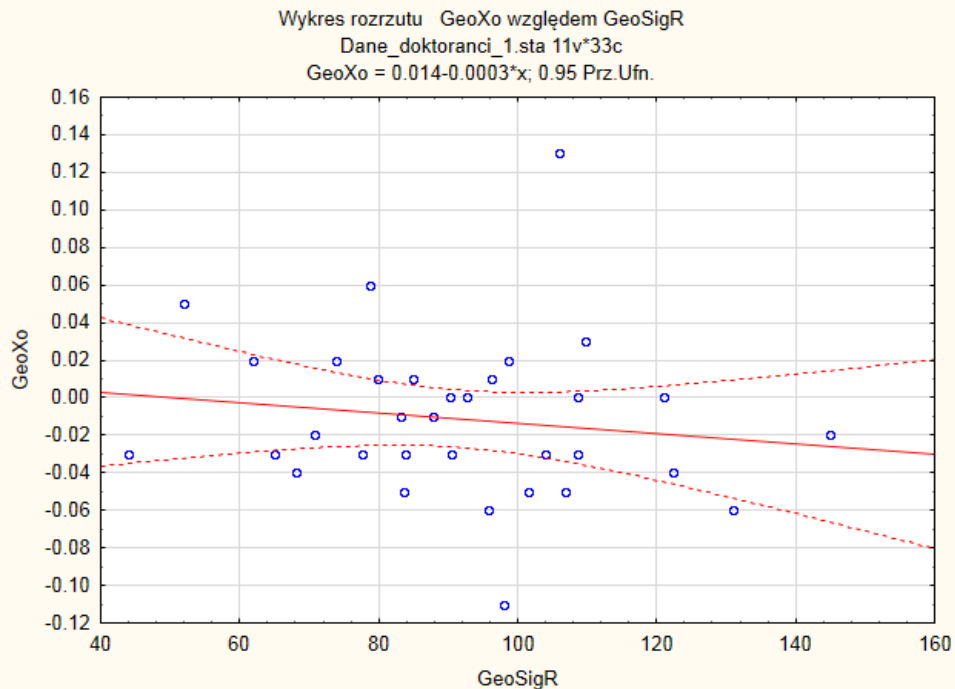
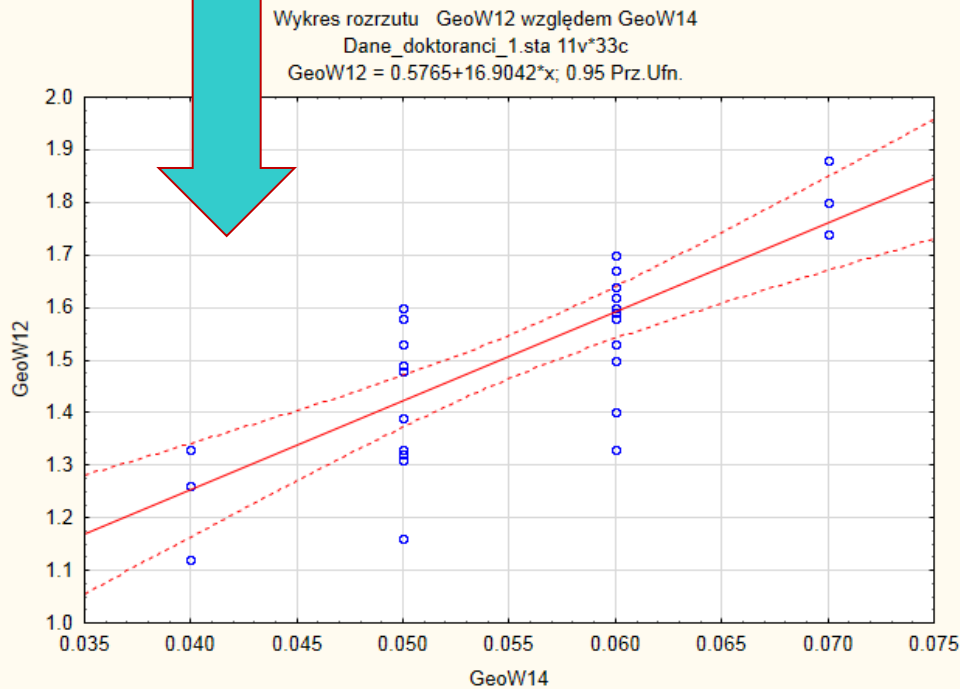
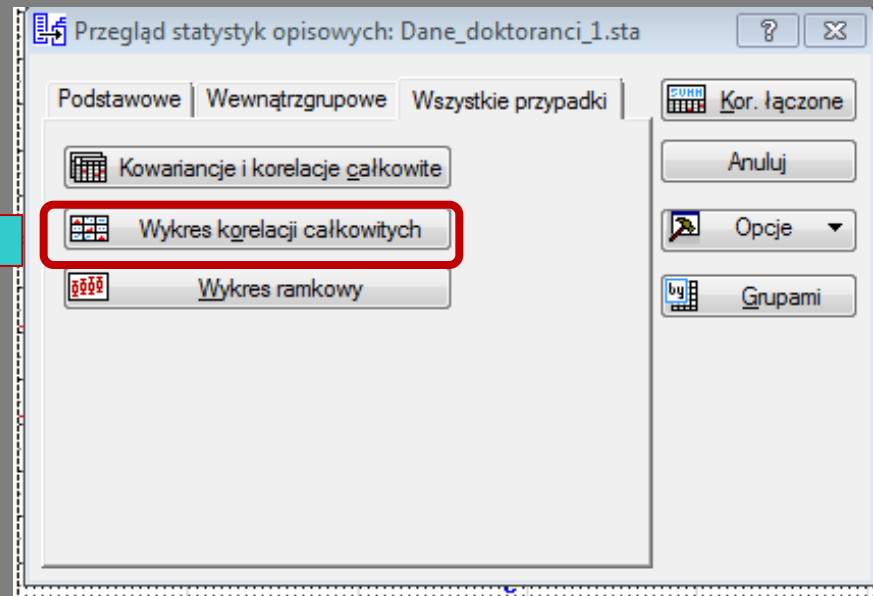
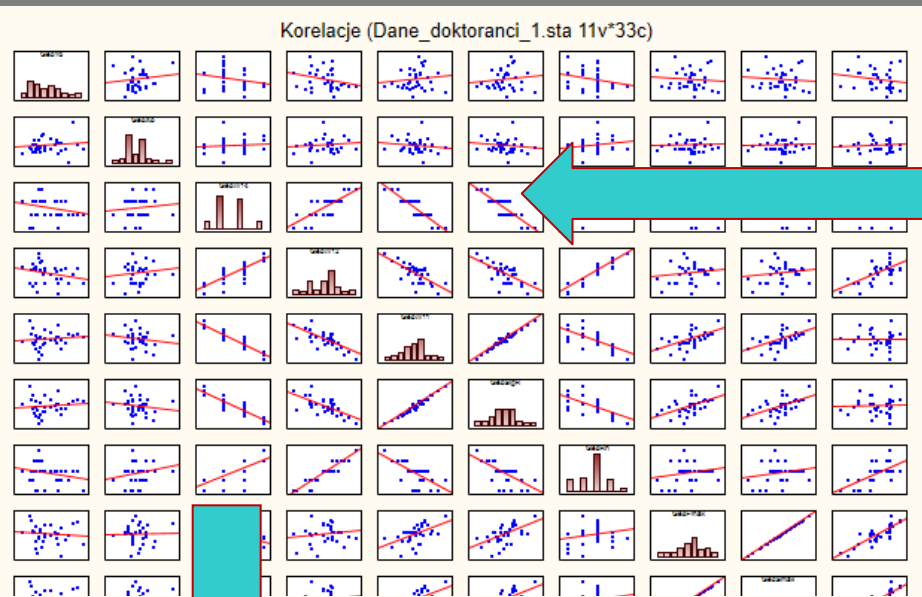
Wytnij Ctrl+X  
Kopiuj Ctrl+C  
Kopiuj z nagłówkami



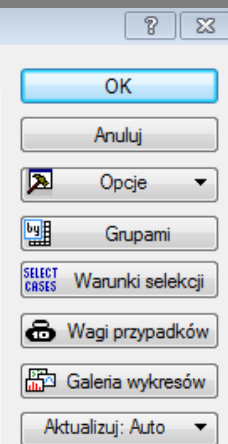
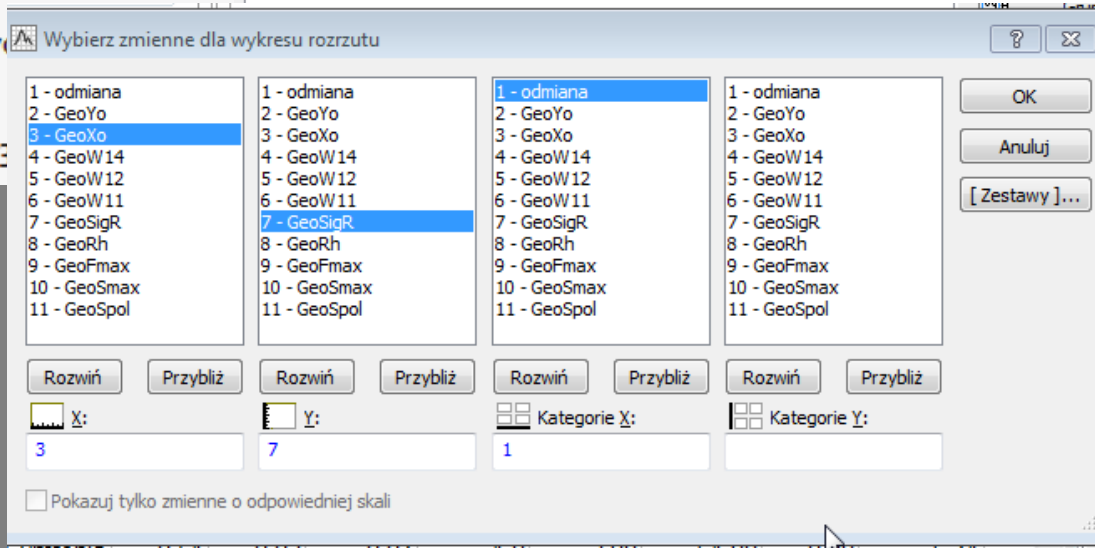
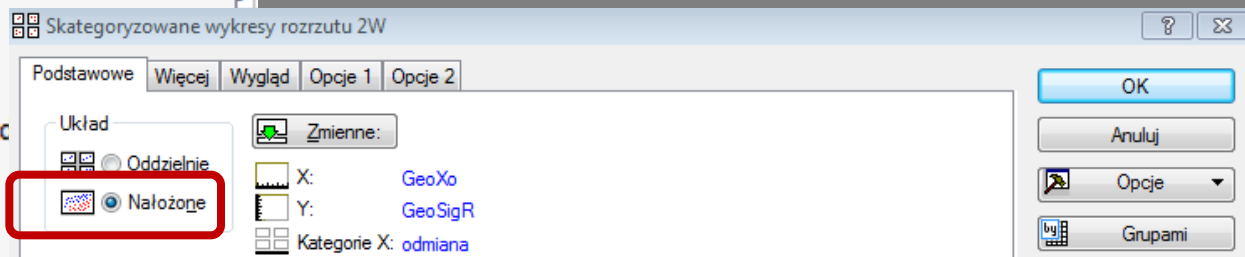
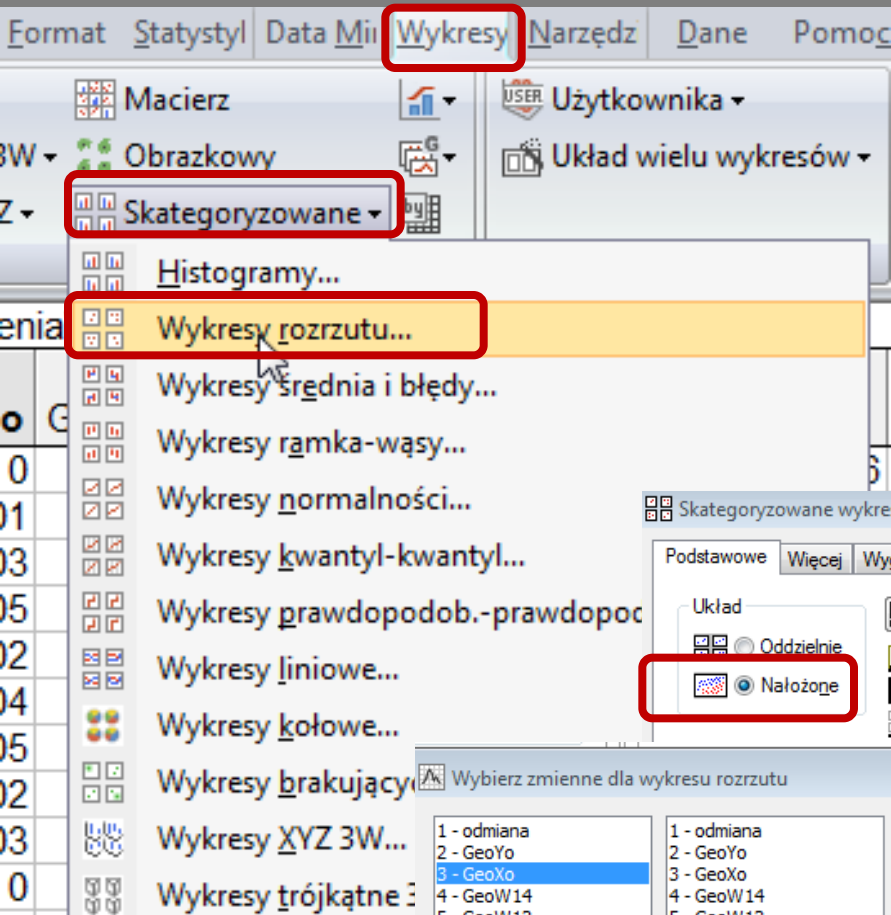
Względem...  
Względem...  
Względem...  
bieństwa wz.  
wykres rozrzutu...

Histogram (Irisdat.sta  
Szer działki = 150\*0,2\*normal

# Wykresy rozrzutu.



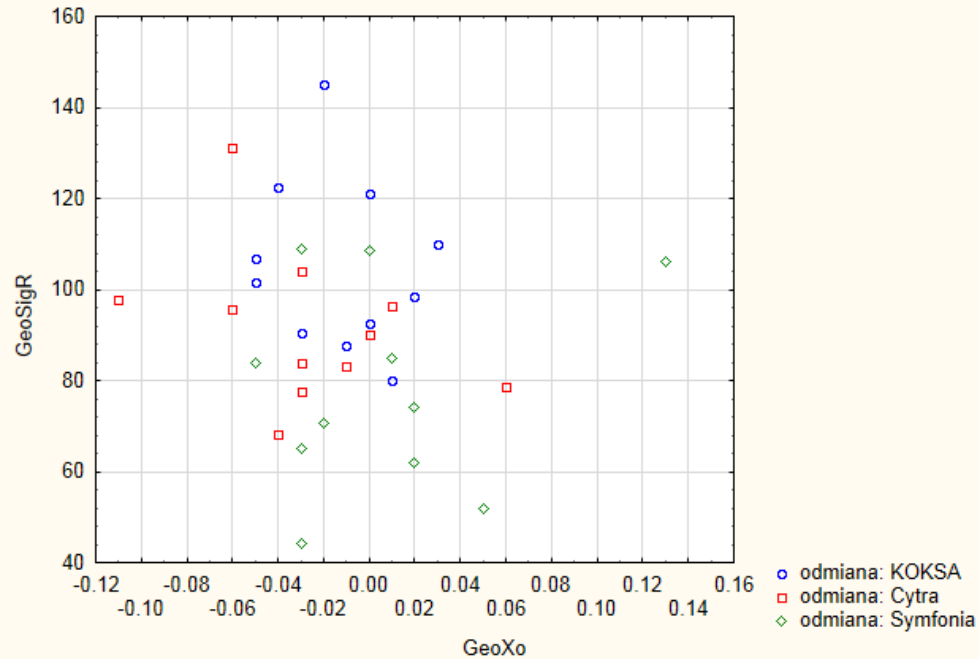
# Wykresy rozrzutu- skateryzowane



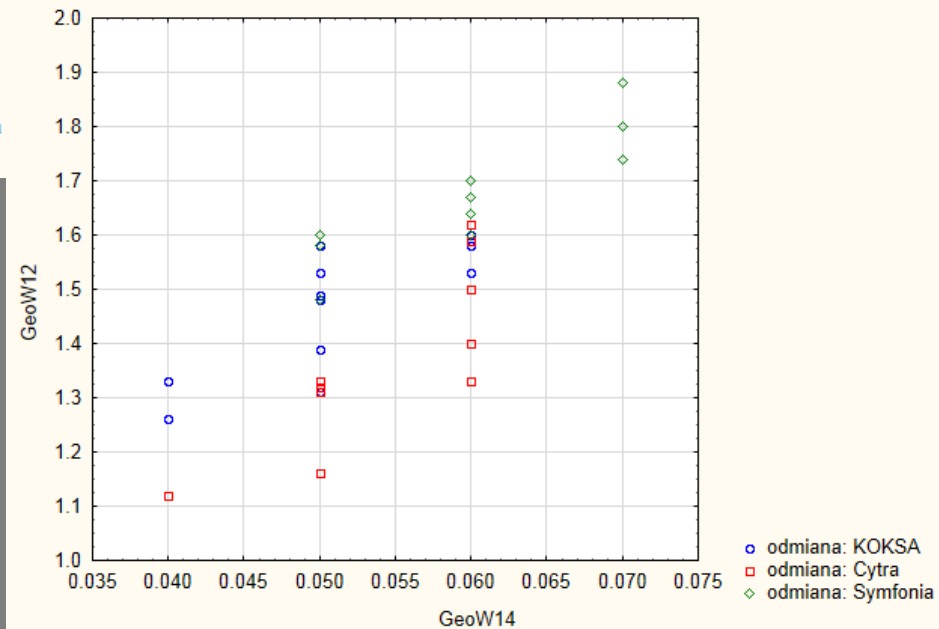


# Wykresy rozrzutu- skategoryzowane

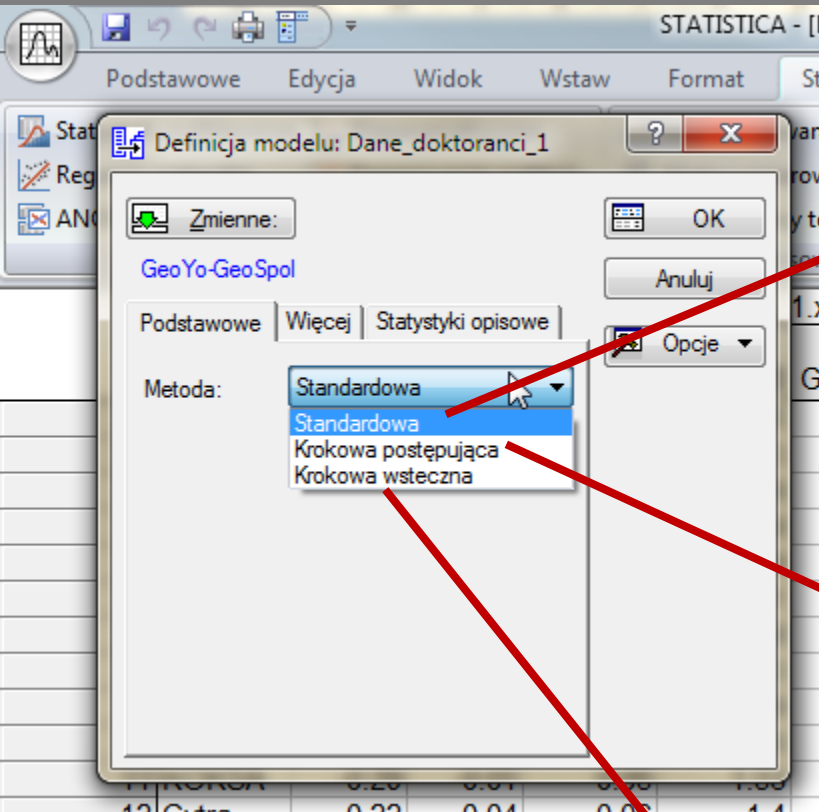
Wykres rozrzutu GeoSigR względem GeoXo; kategorie względem odmiana  
Dane\_doktoranci\_1.sta 11v\*33c



Wykres rozrzutu GeoW12 względem GeoW14; kategorie względem odmiana  
Dane\_doktoranci\_1.sta 11v\*33c

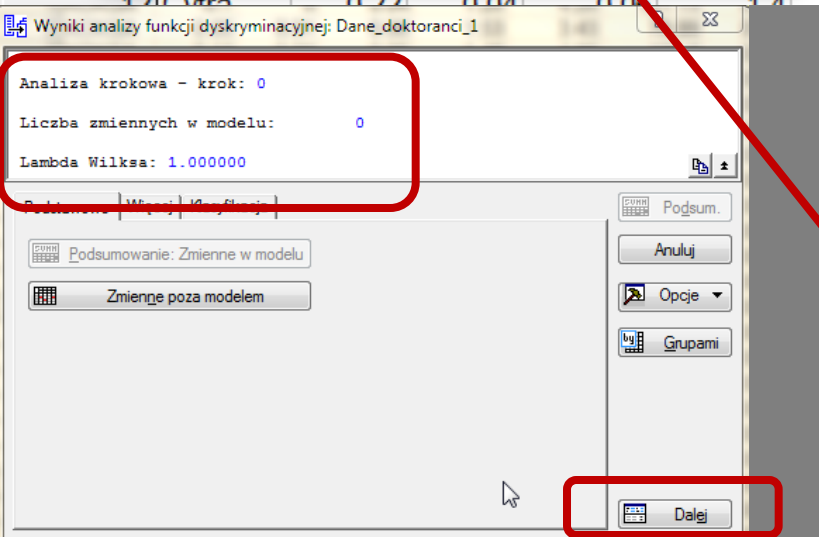


# • Wybór metody analizy



• Opcja ta powoduje jednoczesne wprowadzenie do modelu wszystkich wybranych zmiennych

• W tej wersji zmienne są kolejno dodawane do modelu. Przykładowo może ona polegać w pierwszym kroku na wyborze do modelu tej zmiennej objaśniającej, która jest najsilniej skorelowana ze zmienną objaśnianą i wyznacza model o istotnych parametrach. W drugim kroku wybierana jest kolejna zmienna objaśniająca, której wartości są najsilniej skorelowane z resztami kroku pierwszego, a rozszerzony model charakteryzuje się istotnością wszystkich parametrów.



• Polega w pierwszym kroku na skonstruowaniu modelu zawierającego wszystkie potencjalne zmienne objaśniające, a następnie na stopniowym eliminowaniu zmiennych tak, aby utrzymać model z najwyższą wartością współczynnika determinacji przy zachowaniu istotności parametrów.

• **Lambda Wilksa.** Mówiąc ogólnie, *lambda Wilksa* jest standardową statystyką stosowaną do wyznaczenia istotności statystycznej mocy dyskryminacyjnej aktualnego modelu. Jej wartość mieści się w zakresie od 1,0 (brak mocy dyskryminacyjnej) do 0,0 (doskonała moc dyskryminacyjna).

Wyniki analizy funkcji dyskryminacyjnej: Dane\_doktoranci\_1

Analiza krokowa - krok: 1

Liczba zmiennych w modelu: 1  
Ostatnia wprov. zm.: GeoSpr F (2,31) = 15.67589 p < .0000  
Lambda Wilksa: .4889833 przybl. 2,30) = 15.67589 p < .0000

Podstawowe | Więcej | Klasyfikacja

Podsumowanie: Zmienne w modelu

Zmienne poza modelem

Anuluj

Wyniki analizy funkcji dyskryminacyjnej: Dane\_doktoranci\_1

Analiza krokowa - krok: 2

Liczba zmiennych w modelu: 2  
Ostatnia wprov. zm.: GeoSigR F (2,30) = 6.135797 p < .0058  
Lambda Wilksa: .3435902 przybl. 4,58) = 10.23703 p < .0000

Podstawowe | Więcej | Klasyfikacja

Podsumowanie: Zmienne w modelu

Zmienne poza modelem

Anuluj

Opcje

Wyniki analizy funkcji dyskryminacyjnej: Dane\_doktoranci\_1

Analiza krokowa - krok: 3 Końcowy krok

Liczba zmiennych w modelu: 3  
Ostatnia wprov. zm.: GeoXo F (2,28) = .8578532 p < .4349  
Lambda Wilksa: .3062023 przybl. 6,56) = 7.533462 p < .0000

Podstawowe | Więcej | Klasyfikacja

Podsumowanie: Zmienne w modelu

Zmienne poza modelem

Anuluj

Opcje

Grupami

Analiza krokowa - krok: 3 Końcowy krok

Liczba zmiennych w modelu: 3  
 Ostatnia wprow. zmn.: GeoXo F (2, 28) = .8578532 p < .4349  
 Lambda Wilksa: .3062023 przybl. 6, 56) = 7.533462 p < .0000

Podstawowe | Więcej | Klasyfikacja

Podsum.

Podsumowanie: Zmienne w modelu

Anuluj

Zmienne poza modelem

Opcje

Grupami

Podsumowanie analizy funkcji dyskryminacyjnej. (Dane\_doktoranci\_1)  
 Krok 3, N zmn. w modelu: 3;Grupująca: odmiana (3 grup)  
 Lambda Wilksa: .30620 przybl. F (6,56)=7.5335 p< .0000

N=33	Lambda Wilksa	Cząstk. Wilksa	F usun. (2,28)	p	Toler.	1-Toler. (R-kwad)
GeoSpol	0.673278	0.454793	16.78322	0.000016	0.916025	0.083975
GeoSigR	0.426301	0.718278	5.49106	0.009730	0.944109	0.055891
GeoXo	0.343590	0.891184	1.70943	0.199318	0.966609	0.033391

Analiza krokowa - krok: 3 Końcowy krok

Liczba zmiennych w modelu: 3  
 Ostatnia wprow. zm.: GeoXo F (2,28) = .8578532 p < .4349  
 Lambda Wilksa: .3062023 przybl. 6,56) = 7.533462 p < .0000

Podstawowe Więcej Klasyfikacja

Podsumowanie: Zmienne w modelu  
 Zmienne poza modelem  
 Odległości między grupami  
**Wykonaj analizę kanoniczną**  
 Podsumowanie analizy krokowej

Podsum. Anuluj Opcje Grupami

- Pierwszy wiersz zawiera zawsze test istotności dla wszystkich pierwiastków; drugi wiersz podaje istotność pozostałych pierwiastków, po usunięciu pierwszego pierwiastka i tak dalej. Zatem arkusz ten mówi nam, ile pierwiastków kanonicznych (funkcji dyskryminacyjnych) interpretować.

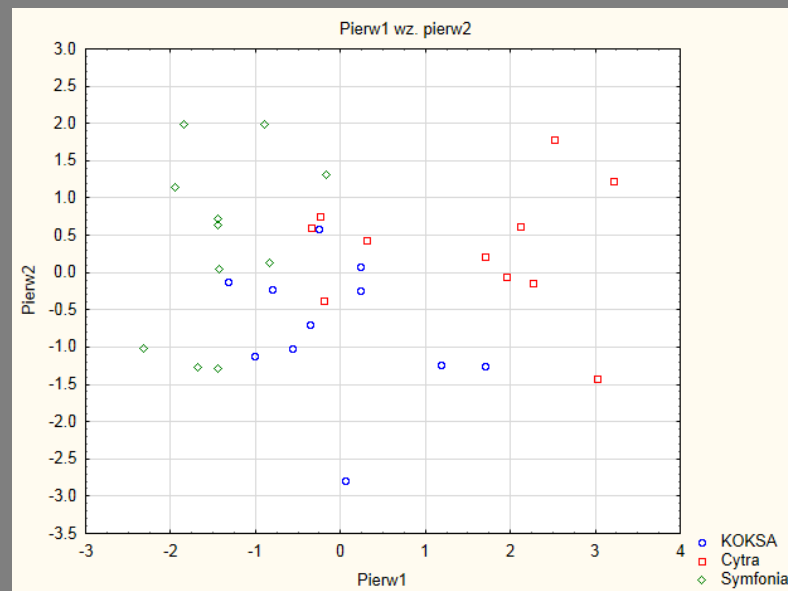
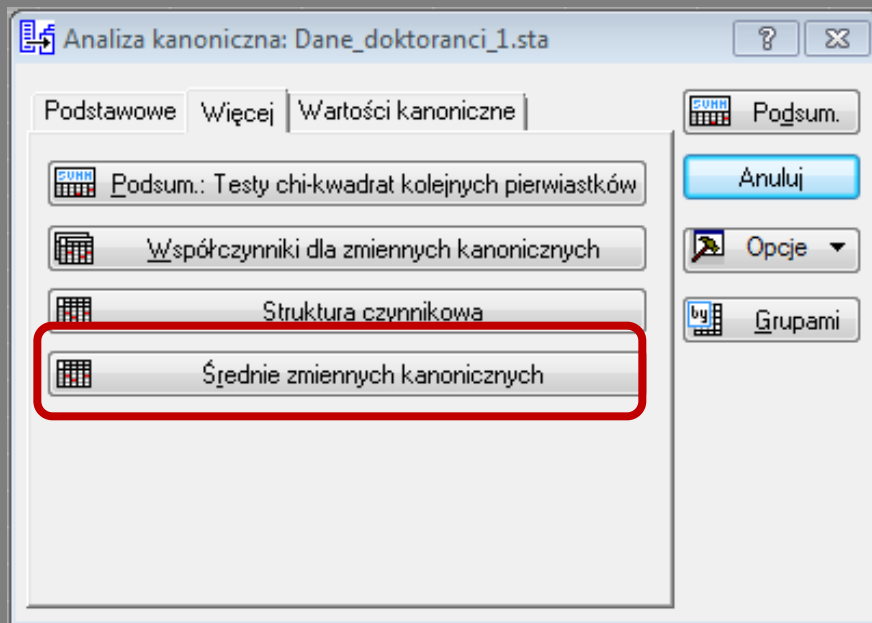
Pierw. Usunięte	Testy chi-kwadrat kolejnych pierwiastków (Dane_doktoranci_1.sta)					
	Wartość własna	Kanonicz R	Lambda Wilksa	chi-kwad	df	p
0	1.525697	0.777219	0.306202	34.32177	6	0.000006
1	0.293035	0.476052	0.773374	7.45278	2	0.024080

Zmienna	Surowe wsp. (Dane_doktoranci_1) dla zmiennych kanonicz.	
	Pierw1	Pierw2
GeoSpol	-2.7215	-0.669458
GeoSigR	0.0217	-0.046953
GeoXo	-10.3371	-0.404473
Stała	10.7497	7.455940
Wart.wł.	1.5257	0.293035
Skum.pro	0.8389	1.000000

Zmienna	Współczynniki standaryzow (Dane_doktoranci_1) dla zmiennych kanonicz.	
	Pierw1	Pierw2
GeoSpol	-0.981542	-0.241452
GeoSigR	0.422758	-0.916690
GeoXo	-0.431571	-0.016887
Wart.wł.	1.525697	0.293035
Skum.pro	0.838879	1.000000



- Po określeniu i sprawdzeniu istotności dyskryminacyjnej, przystępujemy do określenia dyskryminacji dla każdej funkcji dyskryminacyjnej.



Grupa	Średnie zmien. kanonicznych (Dane_doktoranci_1.sta)	
	Pierw1	Pierw2
KOKSA	-0.08217	-0.729037
Cytra	1.48172	0.333333
Symfonia	-1.39955	0.395704

- Funkcje klasyfikacyjne do analizy nowych przypadków

Zmienna	Funkcje klasyfikacyjne; grupująca: odmi:		
	KOKSA p=.33333	Cytra p=.33333	Symfonia p=.33333
GeoSpol	36.9886	32.0214	39.821
GeoSigR	0.1213	0.1053	0.040
GeoXo	54.1949	37.5991	67.358
Stała	-97.0947	-73.2467	-103.659

Wyniki analizy funkcji dyskryminacyjnej: Dane\_doktoranci\_1.sta

Analiza krokowa - krok: 3 Końcowy krok

Liczba zmiennych w modelu: 3  
 Ostatnia wprowadz. zm.: GeoXo F (2, 28) = .8578532 p < .4349  
 Lambda Wilksa: .3062023 przybl. 6, 56) = 7.533462 p < .0000

Podstawowe | Więcej | Klasyfikacja

**Funkcje klasyfikacyjne**

Wybór przypadków do klasyfikacji SELECT CASES Selekcja

Macierz klasyfikacji

Klasyfikacja przypadków

Kwadraty odległ. Mahalanobisa

Prawdopodobieństwa a posteriori

Zapisz wartości

Prawdopod. klasyfikacyjne a priori

Proporcjonalne do wielkości grup  
 Jednokowe dla wszystkich grup  
 Zdefiniowane przez użytkownika

Do zapisania dla każdego przypadku

Zapisz klasyfikację przypadku  
 Zapisz odległość przypadku  
 Zapisz prawdop. a posteriori przyp.

Maks. liczba przypadk. w jednym arkuszu wyników: 100000

Podsum. Anuluj Opcje Grupami

- $D_1 = -97,095 + 36,989 * \text{GeoSpol} + 0,123 * \text{GeoSigR} + 54,195 * \text{GeoXo}$

# • Funkcje klasyfikacyjne do analizy nowych przypadków

Zmienna	Funkcje klasyfikacyjne; grupująca: odmi:		
	KOKSA p=.33333	Cytra p=.33333	Symfonia p=.33333
GeoSpol	36.9886	32.0214	39.821
GeoSigR	0.1213	0.1053	0.040
GeoXo	54.1949	37.5991	67.358
Stała	-97.0947	-73.2467	-103.659

Wyniki analizy funkcji dyskryminacyjnej: Dane\_doktoranci\_1.sta

Analiza krokowa - krok: 3 Końcowy krok

Liczba zmiennych w modelu: 3  
 Ostatnia wprowadz. zm.: GeoXo F (2,28) = .8578532 p < .4349  
 Lambda Wilksa: .3062023 przybl. 6,56 = 7.533462 p < .0000

Podstawowe | Więcej | **Klasyfikacja**

Funkcje klasyfikacyjne

Wybór przypadków do klasyfikacji SELECT CASES Selekcja

**Macierz klasyfikacji**

Klasyfikacja przypadków

Kwadraty odległ. Mahalanobisa

Prawdopodobieństwa a posteriori

Zapisz wartości

Prawdopod. klasyfikacyjne a priori

Proporcjonalne do wielkości grup  
 Jednokowe dla wszystkich grup  
 Zdefiniowane przez użytkownika

Do zapisania dla każdego przypadku

Zapisz klasyfikację przypadku  
 Zapisz odległość przypadku  
 Zapisz prawdop. a posteriori przyp.

Maks. liczba przypadk. w jednym arkuszu wyników: 100000

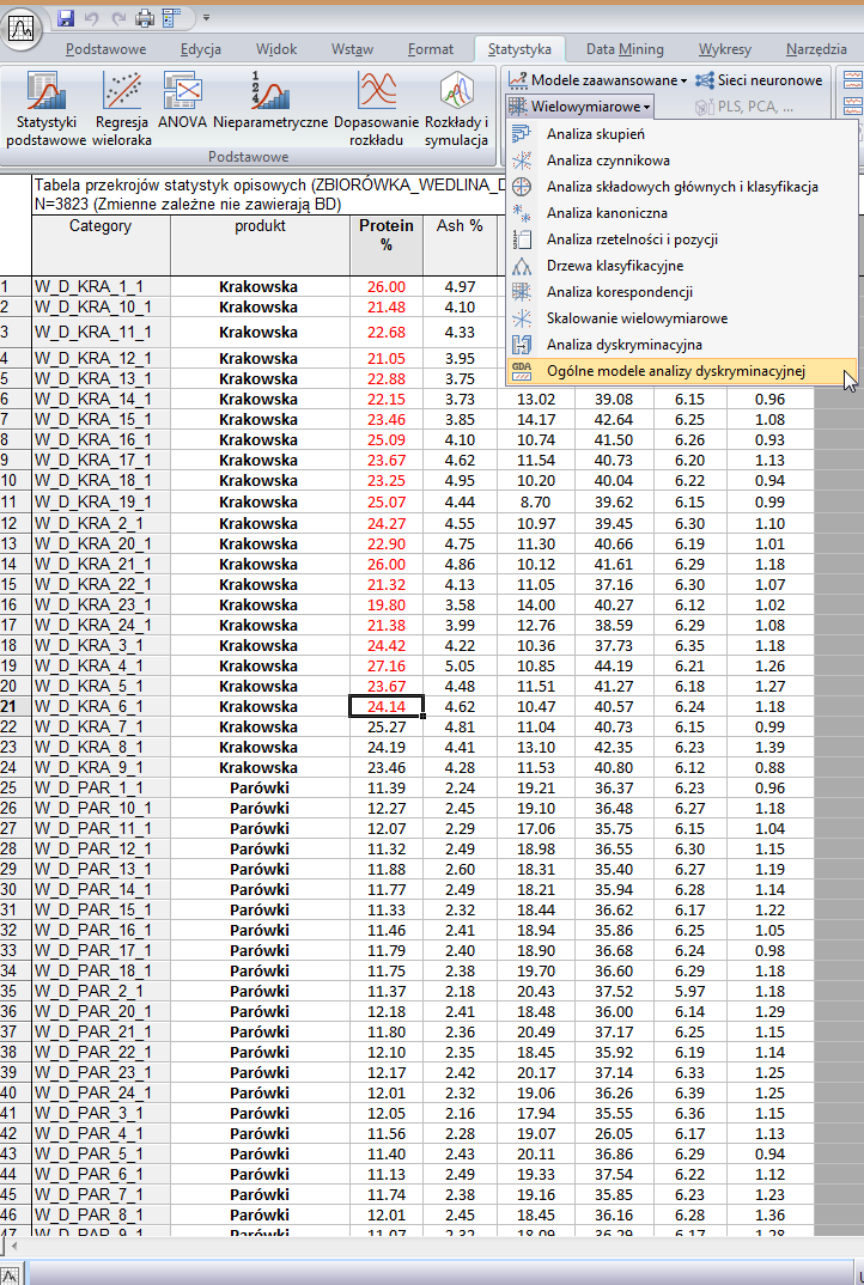
Podsum. Anuluj Opcje Grupami

$$• D_1 = -97,095 + 36,989 * \text{GeoSpol} + 0,123 * \text{GeoSigR} + 54,195 * \text{GeoXo}$$

Grupa	Macierz klasyfikacji (Dane_doktoranci_1.sta)			
	Procent Poprawne	KOKSA p=.33333	Cytra p=.33333	Symfonia p=.33333
KOKSA	63.63636	7	1	3
Cytra	72.72727	1	8	2
Symfonia	81.81818	2	0	9
Razem	72.72727	10	9	14



# GDA – tradycyjna metoda dyskryminacyjna



The screenshot shows the Statistica software interface. The main window displays a data table titled "Tabela przekrojów statystyk opisowych (ZBIORÓWKA\_WEDLINA, N=3823 (Zmienne zależne nie zawierają BD))". The table has columns for "Category", "produkt", "Protein %", "Ash %", and several unlabeled columns. The data is grouped by "produkt" into "Krakowska" and "Parówki". A menu is open over the table, showing various statistical analysis options, with "Ogólne modele analizy dyskryminacyjnej" (General discriminant analysis models) selected.

	Category	produkt	Protein %	Ash %				
1	W_D_KRA_1_1	Krakowska	26.00	4.97				
2	W_D_KRA_10_1	Krakowska	21.48	4.10				
3	W_D_KRA_11_1	Krakowska	22.68	4.33				
4	W_D_KRA_12_1	Krakowska	21.05	3.95				
5	W_D_KRA_13_1	Krakowska	22.88	3.75				
6	W_D_KRA_14_1	Krakowska	22.15	3.73	13.02	39.08	6.15	0.96
7	W_D_KRA_15_1	Krakowska	23.46	3.85	14.17	42.64	6.25	1.08
8	W_D_KRA_16_1	Krakowska	25.09	4.10	10.74	41.50	6.26	0.93
9	W_D_KRA_17_1	Krakowska	23.67	4.62	11.54	40.73	6.20	1.13
10	W_D_KRA_18_1	Krakowska	23.25	4.95	10.20	40.04	6.22	0.94
11	W_D_KRA_19_1	Krakowska	25.07	4.44	8.70	39.62	6.15	0.99
12	W_D_KRA_2_1	Krakowska	24.27	4.55	10.97	39.45	6.30	1.10
13	W_D_KRA_20_1	Krakowska	22.90	4.75	11.30	40.66	6.19	1.01
14	W_D_KRA_21_1	Krakowska	26.00	4.86	10.12	41.61	6.29	1.18
15	W_D_KRA_22_1	Krakowska	21.32	4.13	11.05	37.16	6.30	1.07
16	W_D_KRA_23_1	Krakowska	19.80	3.58	14.00	40.27	6.12	1.02
17	W_D_KRA_24_1	Krakowska	21.38	3.99	12.76	38.59	6.29	1.08
18	W_D_KRA_3_1	Krakowska	24.42	4.22	10.36	37.73	6.35	1.18
19	W_D_KRA_4_1	Krakowska	27.16	5.05	10.85	44.19	6.21	1.26
20	W_D_KRA_5_1	Krakowska	23.67	4.48	11.51	41.27	6.18	1.27
21	W_D_KRA_6_1	Krakowska	24.14	4.62	10.47	40.57	6.24	1.18
22	W_D_KRA_7_1	Krakowska	25.27	4.81	11.04	40.73	6.15	0.99
23	W_D_KRA_8_1	Krakowska	24.19	4.41	13.10	42.35	6.23	1.39
24	W_D_KRA_9_1	Krakowska	23.46	4.28	11.53	40.80	6.12	0.88
25	W_D_PAR_1_1	Parówki	11.39	2.24	19.21	36.37	6.23	0.96
26	W_D_PAR_10_1	Parówki	12.27	2.45	19.10	36.48	6.27	1.18
27	W_D_PAR_11_1	Parówki	12.07	2.29	17.06	35.75	6.15	1.04
28	W_D_PAR_12_1	Parówki	11.32	2.49	18.98	36.55	6.30	1.15
29	W_D_PAR_13_1	Parówki	11.88	2.60	18.31	35.40	6.27	1.19
30	W_D_PAR_14_1	Parówki	11.77	2.49	18.21	35.94	6.28	1.14
31	W_D_PAR_15_1	Parówki	11.33	2.32	18.44	36.62	6.17	1.22
32	W_D_PAR_16_1	Parówki	11.46	2.41	18.94	35.86	6.25	1.05
33	W_D_PAR_17_1	Parówki	11.79	2.40	18.90	36.68	6.24	0.98
34	W_D_PAR_18_1	Parówki	11.75	2.38	19.70	36.60	6.29	1.18
35	W_D_PAR_2_1	Parówki	11.37	2.18	20.43	37.52	5.97	1.18
36	W_D_PAR_20_1	Parówki	12.18	2.41	18.48	36.00	6.14	1.29
37	W_D_PAR_21_1	Parówki	11.80	2.36	20.49	37.17	6.25	1.15
38	W_D_PAR_22_1	Parówki	12.10	2.35	18.45	35.92	6.19	1.14
39	W_D_PAR_23_1	Parówki	12.17	2.42	20.17	37.14	6.33	1.25
40	W_D_PAR_24_1	Parówki	12.01	2.32	19.06	36.26	6.39	1.25
41	W_D_PAR_3_1	Parówki	12.05	2.16	17.94	35.55	6.36	1.15
42	W_D_PAR_4_1	Parówki	11.56	2.28	19.07	26.05	6.17	1.13
43	W_D_PAR_5_1	Parówki	11.40	2.43	20.11	36.86	6.29	0.94
44	W_D_PAR_6_1	Parówki	11.13	2.49	19.33	37.54	6.22	1.12
45	W_D_PAR_7_1	Parówki	11.74	2.38	19.16	35.85	6.23	1.23
46	W_D_PAR_8_1	Parówki	12.01	2.45	18.45	36.16	6.28	1.36
47	W_D_PAR_9_1	Parówki	11.07	2.22	19.00	36.20	6.17	1.20

- Do analizy dyskryminacyjnej może wykorzystywać wielomiany kwadratowe jako funkcje dyskryminacyjne
- Nie nakłada ograniczeń na zmienne dyskryminujące (jakościowe i ilościowe)
- Umożliwia zastosowanie więcej analiz (np. najlepszego podzbioru)



# Wybór zmiennych do analizy

Tabela przekrojów statystyk opisowych (ZBIORÓWKA\_WEDLINA\_D)  
N=3823 (Zmienne zależne nie zawierają BD)

Category	produkt	podzial	Protein %	Ash %	Fat %	D.M %	ph	Collagen %
W_D_KRA_1_1	Krakowska	uczacy	26.00	4.97	9.42	42.03	6.34	1.34
W_D_KRA_10_1	Krakowska	testowy	21.48	4.10	9.81	37.96	6.10	1.06
W_D_KRA_11_1	Krakowska	testowy	22.68	4.33	11.44	39.01	6.18	1.07
W_D_KRA_12_1	Krakowska	uczacy	21.05	3.95	12.07	38.52	6.19	1.17
W_D_KRA_13_1	Krakowska	uczacy	22.88	3.75	11.65	38.13	6.28	0.88
W_D_KRA_14_1	Krakowska	uczacy	22.15	3.73	13.02	39.08	6.15	0.96
W_D_KRA_15_1	Krakowska	uczacy	23.46	3.85	14.17	42.64	6.25	1.08
W_D_KRA_16_1	Krakowska	testowy	25.09	4.10	10.74	41.50	6.26	0.93
W_D_KRA_17_1	Krakowska	uczacy	23.67	4.62	11.54	40.73	6.20	1.13
W_D_KRA_18_1	Krakowska	testowy	23.25	4.95	10.20	40.04	6.22	0.94
W_D_KRA_19_1	Krakowska	uczacy	25.07	4.44	8.70	39.62	6.15	0.94
W_D_KRA_2_1	Krakowska	testowy	24.27	4.55	10.97	39.45	6.30	0.94
W_D_KRA_20_1	Krakowska	testowy	22.90	4.75	11.30	40.66	6.19	0.94
W_D_KRA_21_1	Krakowska	uczacy	26.00	4.86	10.12	41.61	6.29	0.94
W_D_KRA_22_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_23_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_24_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_3_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_4_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_5_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_6_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_7_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_8_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_9_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_PAR_1_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_10_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_11_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_12_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_13_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_14_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_15_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_16_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_17_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_18_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15

Wybierz zmienną zależną i predyktory ciągłe:

- 1 - Category
- 2 - produkt
- 3 - podzial

- 4 - Protein %
- 5 - Ash %
- 6 - Fat %
- 7 - D.M %
- 8 - ph
- 9 - Collagen %

Zmienna zależna: 2  
Predyktory: 4-9

Pokazuj tylko zmienne o odpowiedniej skali

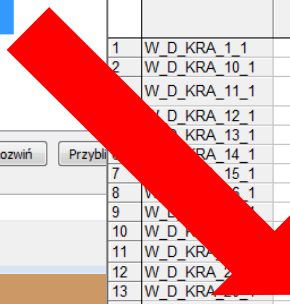


Tabela przekrojów statystyk opisowych (ZBIORÓWKA\_WEDLINA\_D)  
N=3823 (Zmienne zależne nie zawierają BD)

Category	produkt	podzial	Protein %	Ash %	Fat %	D.M %	ph	Collagen %
W_D_KRA_1_1	Krakowska	uczacy	26.00	4.97	9.42	42.03	6.34	1.34
W_D_KRA_10_1	Krakowska	testowy	21.48	4.10	9.81	37.96	6.10	1.06
W_D_KRA_11_1	Krakowska	testowy	22.68	4.33	11.44	39.01	6.18	1.07
W_D_KRA_12_1	Krakowska	uczacy	21.05	3.95	12.07	38.52	6.19	1.17
W_D_KRA_13_1	Krakowska	uczacy	22.88	3.75	11.65	38.13	6.28	0.88
W_D_KRA_14_1	Krakowska	uczacy	22.15	3.73	13.02	39.08	6.15	0.96
W_D_KRA_15_1	Krakowska	uczacy	23.46	3.85	14.17	42.64	6.25	1.08
W_D_KRA_16_1	Krakowska	testowy	25.09	4.10	10.74	41.50	6.26	0.93
W_D_KRA_17_1	Krakowska	uczacy	23.67	4.62	11.54	40.73	6.20	1.13
W_D_KRA_18_1	Krakowska	testowy	23.25	4.95	10.20	40.04	6.22	0.94
W_D_KRA_19_1	Krakowska	uczacy	25.07	4.44	8.70	39.62	6.15	0.94
W_D_KRA_2_1	Krakowska	testowy	24.27	4.55	10.97	39.45	6.30	0.94
W_D_KRA_20_1	Krakowska	testowy	22.90	4.75	11.30	40.66	6.19	0.94
W_D_KRA_21_1	Krakowska	uczacy	26.00	4.86	10.12	41.61	6.29	0.94
W_D_KRA_22_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_23_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_24_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_3_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_4_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_5_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_6_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_7_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_8_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_KRA_9_1	Krakowska	uczacy	21.32	4.13	11.05	37.16	6.30	0.94
W_D_PAR_1_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_10_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_11_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_12_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_13_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_14_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_15_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_16_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15
W_D_PAR_17_1	Parówki	uczacy	12.07	2.29	17.06	35.75	6.15	1.04
W_D_PAR_18_1	Parówki	testowy	11.32	2.49	18.98	36.55	6.30	1.15

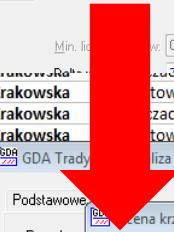
GDA Tradycyjna analiza dyskryminacyjna: Dane\_doktoranci\_2.sta

Podstawowe Więcej

Prawdopodobieństwa a priori klasyfikacji:  
 Szacowane  Równe  Podane: 1 2

Opcje budowania modelu:  
 Wszystkie efekty  
 Krokowa postępująca  
 Krokowa wsteczna  
 Wprowadz. postępow.  
 Eliminacja wstecz  
 Najlepszy podzbiór

Opcja krzyżowa: nie



GDA Tradycyjna analiza dyskryminacyjna: Dane\_doktoranci\_2.sta

Podstawowe Więcej

Prawdopodobieństwa a priori klasyfikacji:  
 Szacowane  Równe  Podane: 1 2

Opcje budowania modelu:  
 Wszystkie efekty  
 Krokowa postępująca  
 Krokowa wsteczna  
 Wprowadz. postępow.  
 Eliminacja wstecz  
 Najlepszy podzbiór

Zmienna identyfikująca próby: podzial

Kod próby do analizy: uczacy

Stan:  
 Włączona  
 Wyłączona

Wybierz zmienną identyfikującą próby i kod, który jednoznacznie określa przypadki dla których ma być wykonana analiza. Pozostałe ważne przypady (o dodatkich wagach przypadków, spełniające warunki selekcji) tworzą próbe do oceny krzyżowej. Dla wszystkich ważnych przypadków wartości przewidywane i statystyki reszt można obliczać, zapisywać i wykorzystywać w dalszych analizach.

# Wybór metody klasyfikacji

Dane: Dane\_doktoranci\_2.sta (9 zmn. \* 354 prz.)

Tabela przekrojów statystyk opisowych (ZBIORÓWKA\_WEDLINA\_D)  
N=3823 (Zmienne zależne nie zawierają BD)

	Category	produkt	podzial	Protein %	Ash %	Fat %	D.M %	ph	Collagen %
1	W_D_KRA_1_1	Krakowska	uczacy	26.00	4.97	9.42	42.03	6.34	1.34
2	W_D_KRA_10_1	Krakowska	testowy	21.48	4.10	9.81	37.96	6.10	1.06
3	W_D_KRA_11_1	Krakowska	testowy	22.68	4.33	11.44	39.01	6.18	1.07
4	W_D_KRA_12_1	Krakowska	uczacy	21.05	3.95	12.07	38.52	6.19	1.17
5	W_D_KRA_13_1	Krakowska	uczacy	22.88	3.75	11.65	38.13	6.28	0.88
6	W_D_KRA_14_1	Krakowska	uczacy	22.15	3.73	13.02	39.08	6.15	0.96
7	W_D_KRA_15_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
8	W_D_KRA_16_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
9	W_D_KRA_17_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
10	W_D_KRA_18_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
11	W_D_KRA_19_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
12	W_D_KRA_2_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
13	W_D_KRA_20_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
14	W_D_KRA_21_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
15	W_D_KRA_22_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
16	W_D_KRA_23_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
17	W_D_KRA_24_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
18	W_D_KRA_3_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
19	W_D_KRA_4_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
20	W_D_KRA_5_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
21	W_D_KRA_6_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
22	W_D_KRA_7_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
23	W_D_KRA_8_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
24	W_D_KRA_9_1	Krakowska	uczacy	23.46	4.28	11.53	40.80	6.12	0.88
25	W_D_PAR_1_1	Parówki	uczacy	11.39	2.24	19.21	36.37	6.23	0.96

GDA Tradycyjna analiza dyskryminacyjna: Dane\_doktoranci\_2.sta

Podstawowe Więcej

Prawdopodobieństwa a priori klasyfikacji  
 Szacowane  Równe  Podane:

Ocena krzyżowa: tak

Opcje budowania modelu:

Wszystkie efekty  Lambda Wilksa  Błęd. klas. (analiz.)  
 Krokowa postępująca  Błęd. klas. (walid.)  
 Krokowa wsteczna  Błęd. klas. (walid.)  
 Wprowadz. postęp.  
 Eliminacja wstecz  
 Najlepszy podzbiór

Start: 1 Stop: 6 Wyświetl podzb. 10

Min. liczba efektów: 2  
 Delta wymiat.: 1.E-7  
 Delta odwrac.: 1.E-12

OK Anuluj Opcje Edytor składni



# Przeglądanie wyników

GDA - Wyniki 1: Dane\_doktoranci\_2....

Profile | Reszty regresji | Macierze | Raport  
 Podstawowe | Funkcje | Przypadki | Efekty

Średnie predyktorów w klasach

Odchylenia standard. predyktorów w klasach

Testy chi-kwadrat kolejnych pierwiastków

**Współczynniki standaryzowane**

Surowe współczynniki

Współczynniki struktury czynnikowej

Średnie zmiennych kanonicznych w klasach

Kwadrat. odległości Mahalanobisa w klasach

Testy istotności odległości

Współczynniki funkcji klasyfikacyjnej

Macierz klasyfikacji dla próby do analizy

Więcej wyników   Zmień   Zamknij

Grupami   Opcje

Surowe współczynniki kanonicznej funkcji dyskryminacyjnej (Dane\_doktoranci\_2....)  
 Parametryzacja z sigma-ograniczeniami

Efekt	Funkcja 1	Funkcja 2	Funkcja 3	Funkcja 4	Funkcja 5	Funkcja 6
<b>Wyraz wolny</b>	-7.62757	-19.5116	7.97741	25.57403	51.59732	-9.89377
Protein %	-0.39313	0.6495	0.13934	0.40303	-0.18665	0.49765
Ash %	-0.17113	0.4609	2.69494	-2.39545	1.27878	0.82565
Fat %	0.07254	0.0530	0.06613	0.11528	0.14308	0.66426
D.M %	0.39752	0.1848	-0.35283	-0.24660	-0.04361	-0.65615
ph	-0.24234	0.2733	-2.04495	-3.07707	-8.14853	2.07741
Collagen %	2.10262	-1.4824	4.34042	1.87825	-1.99173	0.34123
Wart. własna	22.24818	14.1893	2.70928	1.40693	0.37626	0.14293
Skum.prop.	0.54168	0.8871	0.95310	0.98736	0.99652	1.00000

Standaryzowane współcz. kanonicznej funkcji dyskryminacyjnej (Dane\_doktoranci\_2....)  
 Parametryzacja z sigma-ograniczeniami

Efekt	Funkcja 1	Funkcja 2	Funkcja 3	Funkcja 4	Funkcja 5	Funkcja 6
<b>Wyraz wolny</b>	0.00000	0.00000	0.000000	0.000000	0.000000	0.00000
Protein %	-0.49504	0.81785	0.175459	0.507500	-0.235029	0.62666
Ash %	-0.04679	0.12601	0.736803	-0.654921	0.349623	0.22574
Fat %	0.15582	0.11387	0.142054	0.247632	0.307347	1.42691
D.M %	0.72632	0.33774	-0.644680	-0.450572	-0.079682	-1.19889
ph	-0.02736	0.03085	-0.230840	-0.347350	-0.919833	0.23451
Collagen %	0.39536	-0.27874	0.816149	0.353176	-0.374514	0.06416
Wart. własna	22.24818	14.18933	2.709280	1.406932	0.376257	0.14293
Skum.prop.	0.54168	0.88714	0.953105	0.987359	0.996520	1.00000

# Przeglądanie wyników

GDA - Wyniki 1: Dane\_doktoranci\_2....

Profile | Reszty regresji | Macierze | Raport  
Podstawowe | Funkcje | Przypadki | Efekty

Średnie predyktorów w klasach  
Odchylenia standard. predyktorów w klasach  
Testy chi-kwadrat kolejnych pierwiastków  
Współczynniki standaryzowane  
Surowe współczynniki  
Współczynniki struktury czynnikowej  
**Średnie zmiennych kanonicznych w klasach**

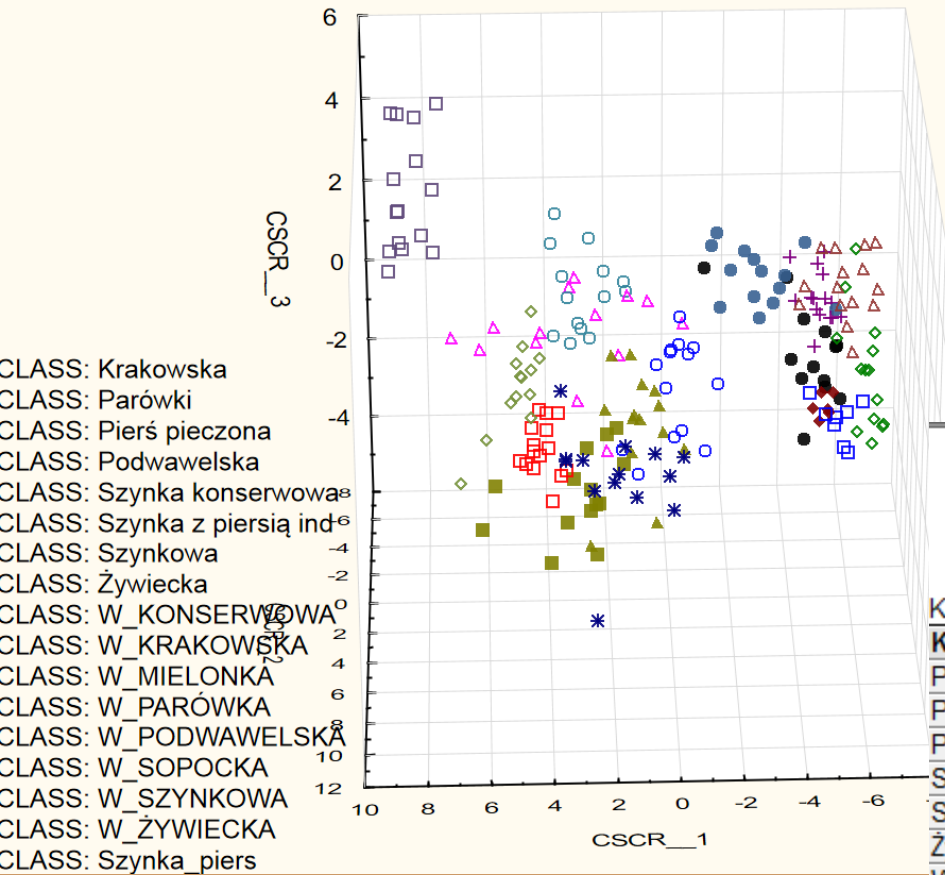
Średnie zm. kanonicznych w klasach (Dane\_doktoranci\_2.sta)  
Parametryzacja z sigma-ograniczeniami

Pierwiastek	Krakowska	Parówki	Pierś pieczona	Podwawelska	Szynka konserwowa	Szynkowa	Żywiecka	W_KONSERWOWA	W_KRAKOWSKA	W_MIELONKA	W_PARÓWKA	W_PODWAWELSK A	W_SOPOCKA	W_SZYNKOWA	W_ŻYWIECKA	Szynka_piers
1	0.069732	4.19053	-6.35495	3.16419	-4.72660	-5.09229	1.129545	-4.96472	1.904295	2.80369	8.50285	4.97430	-5.68933	-2.97405	3.16445	-5.49414
2	5.798854	-3.11074	1.14382	-2.73179	-4.41198	-0.03486	3.605368	-2.38910	7.526508	-3.70737	-2.60502	-1.77141	-0.05024	-2.13226	4.85120	0.24490
3	1.293732	-3.19534	0.02824	-0.26190	-1.34258	-1.20057	-0.335591	0.71631	-0.194697	0.53478	3.34781	-1.21354	1.75235	1.35133	-1.52633	-1.41724
4	-0.945424	0.93010	-0.06433	-1.59031	-0.59362	2.61424	-0.531324	-0.29279	-0.281599	-1.07381	1.83415	-1.52026	-0.39959	0.77461	1.06381	2.22428
5	-0.291283	0.48960	-0.82469	-0.30882	-0.45133	-0.87617	0.422432	0.51342	-0.538745	-0.52529	-0.30245	0.08117	1.09999	0.47638	0.80941	-0.75137
6	-0.198074	-0.07672	0.34036	0.04609	-0.78361	0.53196	-0.063254	0.10142	-0.305499	0.29447	-0.00623	0.39416	0.43434	-0.73107	0.07302	0.33680

Macierz klasyfikacji dla próby do analizy

Więcej wyników | Zmień | Zamknij  
Grupami | Opcje

# Przeglądanie wyników



Macierz klasyfikacji (Dane_doktoranci_2.sta)					
Wiersze: obserwowana klasyfik.					
Kolumny: Przewidywana klasyfikacja					
	Procent Poprawne	Krakowska p=.0660	Parówki p=.0708	Pierś pieczona p=.0660	Pd
Klasa					
<b>Krakowska</b>	100.0000	10.00000	0.000000	0.000000	
Parówki	87.5000	0.00000	7.000000	0.000000	
Pierś pieczona	80.0000	0.00000	0.000000	8.000000	
Podwawelska	50.0000	0.00000	0.000000	0.000000	
Szynka konserwowa	87.5000	0.00000	0.000000	0.000000	
Szynkowa	0.0000	0.00000	0.000000	0.000000	
Żywiecka	80.0000	0.00000	0.000000	0.000000	
W_KONSERWOWA	100.0000	0.00000	0.000000	0.000000	
W_KRAKOWSKA	60.0000	2.00000	0.000000	0.000000	
W_MIELONKA	66.6667	0.00000	0.000000	0.000000	
W_PARÓWKA	88.8889	0.00000	0.000000	0.000000	
W_PODWAWELSKA	100.0000	0.00000	0.000000	0.000000	
W_SOPOCKA	88.8889	0.00000	0.000000	0.000000	
W_SZYNKOWA	100.0000	0.00000	0.000000	0.000000	
W_ŻYWIECKA	80.0000	0.00000	0.000000	0.000000	
Szynka_piers	83.3333	0.00000	0.000000	0.000000	
<b>Ogół</b>	<b>80.9859</b>	<b>12.00000</b>	<b>7.000000</b>	<b>8.000000</b>	



A teraz do pracy

