

Przykład

Dla zilustrowania metody analizy kanonicznej przeprowadzono badanie powiązania zbiorów wybranych zmiennych charakteryzujących 242 gminy miejskie w Polsce w 2005 r. Ze zbioru gmin wyłączono gminy będące jednocześnie powiatami grodzkimi.

Do pierwszego zbioru zmiennych, charakteryzujących potencjał gmin należą:

C_{11.1} – liczba mieszkań ogółem na 1 mieszkańca,

D_{1.1} – liczba aptek ogółem na 1 tys. mieszkańców,

E_{13.1} – liczba gimnazjów dla dzieci i młodzieży na 100 osób w wieku 13-15 lat,

N_{1.1} – liczba jednostek (firm) zarejestrowanych w systemie REGON.

Drugi zbiór tworzą zmienne dotyczące podstawowych źródeł dochodów gmin, a mianowicie:

O_{1.1A} – dochody własne gminy w tys. zł na osobę,

O_{1.10} – subwencje ogólne w tys. zł na osobę,

O_{1.12} – dotacje celowe z budżetu państwa w tys. zł na osobę.

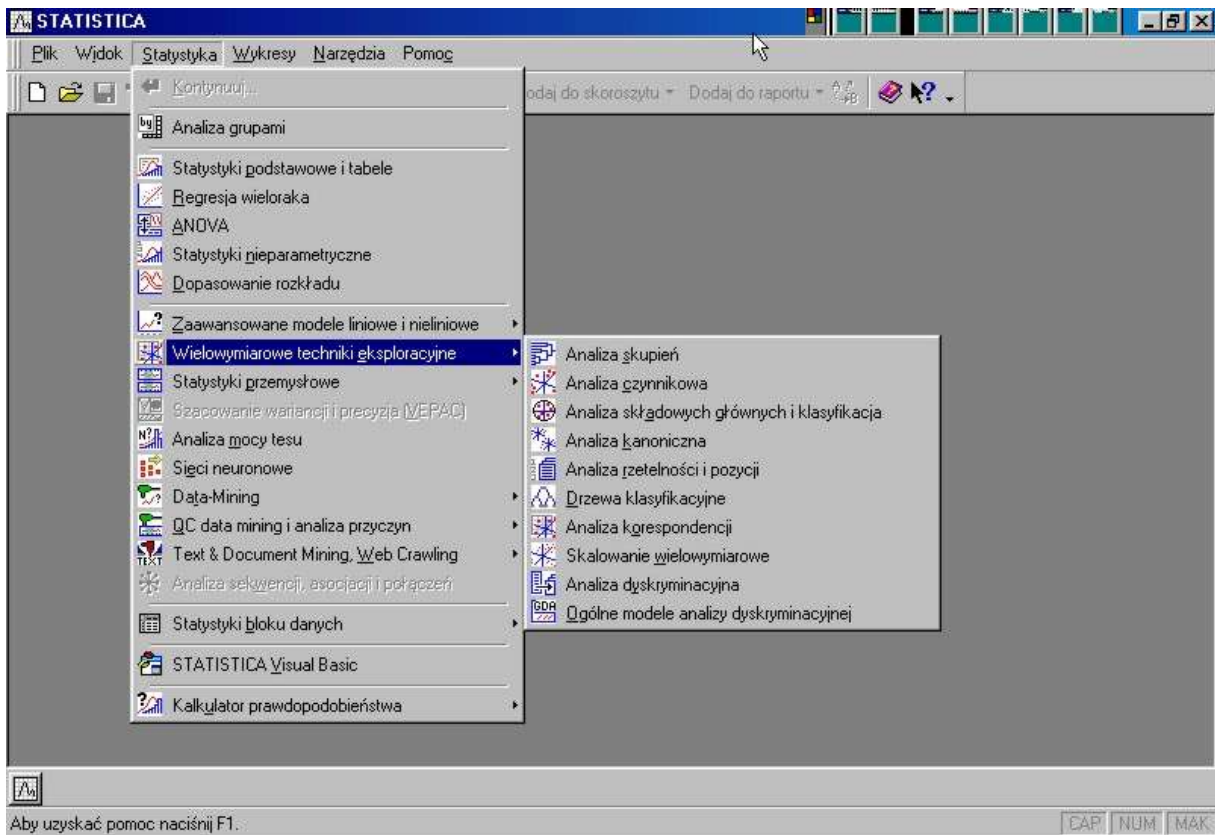
Wartości zmiennych zostały następnie poddane standaryzacji. Fragment arkusza danych znajduje się na rysunku 7.1.

Rys. 7.1. Fragment tablicy z danymi do przykładu 7.1.

	1 GMINA	2 C 11 1	3 D 1 1	4 E 13 1	5 K 1 1	6 N 1 1	7 O 1 1	8 O 1 1A	9 O 1 10	10 O 1 12	11 O 1 16
1	02 01 01	0,430	0,365	0,217	-0,648	0,857	0,182	-0,018	-0,526	-0,432	0,824
2	02 01 02	0,973	-0,923	0,210	0,307	-0,026	-0,786	-0,744	-0,022	0,151	-0,039
3	02 01 03	1,120	-0,201	-0,176	1,473	0,865	-0,562	-0,221	-0,564	-0,105	-0,235
4	02 01 04	0,034	-2,139	-0,367	0,269	-0,911	-1,264	-1,387	-0,206	0,976	0,460
5	02 01 05	-0,345	-0,612	-0,297	2,049	-0,982	-0,366	-1,103	0,669	1,121	-0,437
6	02 01 06	-0,103	0,323	-0,711	1,260	-0,056	-0,304	-0,170	-0,154	0,181	-0,123
7	02 01 07	2,181	0,209	3,134	-1,764	-0,803	4,789	3,992	-0,688	0,884	-0,003
8	02 01 08	0,434	-0,266	-0,569	-1,167	-0,640	-0,892	-0,816	-0,491	0,949	-0,152
9	02 01 09	0,104	-0,505	0,279	-1,148	-0,910	-0,699	-0,240	-0,955	0,150	-0,531
10	02 01 10	1,181	-0,712	1,544	-2,015	-0,758	1,240	1,688	-0,801	0,572	-0,497
11	02 01 111	0,566	-0,735	0,111	0,489	-0,294	-0,642	-0,678	0,067	0,625	-0,531
12	02 01 112	1,563	1,746	3,476	-0,849	-1,000	0,983	0,902	0,154	1,196	-0,351
13	02 01 113	1,009	-0,137	1,062	1,682	1,181	-0,465	-0,089	-0,892	0,283	0,104
14	02 01 114	0,480	-0,616	-0,446	-0,229	-0,762	0,537	0,528	-0,380	0,368	1,571
15	02 01 115	0,887	-0,668	-0,097	0,873	-0,062	-1,073	-1,006	-0,204	0,357	-0,518
16	02 01 116	0,923	-0,654	1,381	-0,767	-0,791	1,782	2,090	-0,395	0,103	1,050
17	02 01 10	0,360	-0,446	-0,036	0,156	-0,224	-0,423	-0,629	-0,277	0,661	-0,474
18	02 01 11	0,818	0,546	0,693	-0,229	-1,005	0,974	0,382	0,851	2,564	-0,481

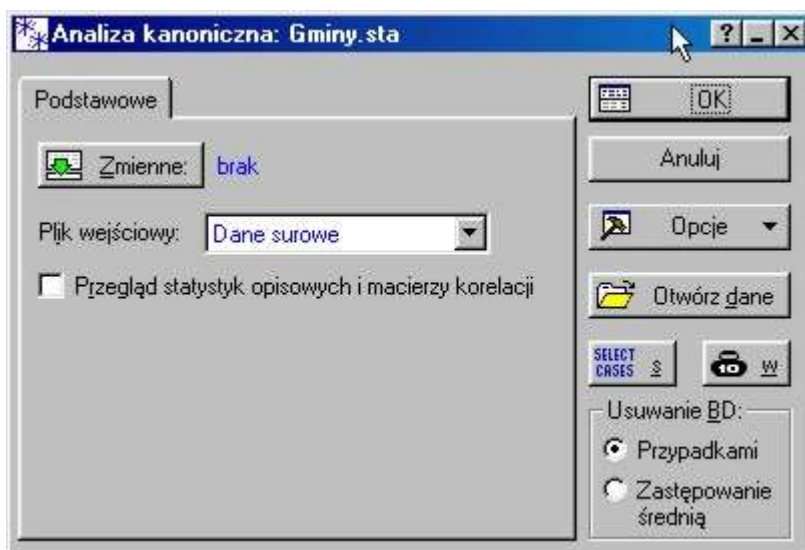
Do przeprowadzenia analizy kanonicznej zastosowano pakiet *STATISTICA*. Analizę wykonujemy w module **Analiza kanoniczna**. Moduł ten uruchamiamy wybierając z menu **Statystyka** opcję **Wielowymiarowe techniki eksploracyjne/Analiza kanoniczna** (rys. 7.2).

Rys. 7.2. Opcje modułu Wielowymiarowe techniki eksploracyjne.



Wybranie tego modułu wywołuje okno **Analiza kanoniczna** (rys. 7.3).

Rys. 7.3. Wstępne okno modułu analizy kanonicznej.



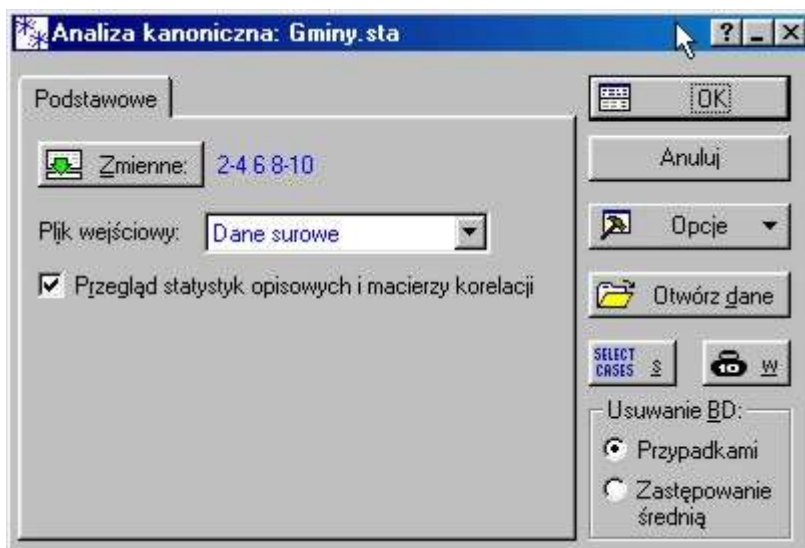
W module tym wybieramy zmienne do analizy, określamy postać pliku wejściowego oraz deklarujemy, czy chcemy wykonać obliczenia szczegółowych statystyk opisowych zmiennych wykorzystywanych w analizie. W celu wyboru zmiennych do analizy klikamy klawisz **Zmienne** otrzymując okno **Wybierz zmienne do analizy**. Z listy wybieramy kursorem, przy wciśniętym na klawiaturze klawiszem **Ctrl**, zmienne do analizy i akceptujemy wybór klawiszem **OK**. (rys. 7.4).

Rys. 7.4. Okno wyboru zmiennych do analizy.



Następnie ustalamy postać pliku wejściowego danych. Możemy wybrać jedną z dwóch opcji: **Dane surowe** i **Macierz korelacji**. Ponieważ dane mają postać zwykłego pliku danych *STATISTICA*, pozostawiamy wybraną domyślnie na liście opcję **Dane surowe**. Wreszcie zaznaczamy pole **Przegląd statystyk opisowych i macierzy korelacji** (rys. 7.5).

Rys. 7.5. Okno analizy kanonicznej z przyjętymi założeniami.



Nasze wybory akceptujemy klawiszem **OK**. Powoduje to otwarcie okna **Przegląd statystyk opisowych** (rys. 7.6).

Rys. 7.6. Okno Przegląd statystyk opisowych.



Najbardziej interesująca dla nas jest macierz korelacji pomiędzy zmiennymi. Czym silniejsze korelacje pomiędzy dwoma wyróżnionymi zbiorami zmiennych tym lepszy efekt da analiza kanoniczna. Aby uzyskać macierz współczynników korelacji między zmiennymi klikamy klawisz **Korelacje** (rys. 7.7).

Rys. 7.7. Macierz współczynników korelacji pomiędzy zmiennymi charakteryzującymi gminy.

		Korelacje (Gminy.sta)						
zmienna		C_11_1	D_1_1	E_13_1	N_1_1	O_1_1A	O_1_10	O_1_12
C_11_1		1,000	-0,061	0,360	0,197	0,496	-0,319	0,055
D_1_1		-0,061	1,000	0,267	-0,141	0,028	0,375	0,230
E_13_1		0,360	0,267	1,000	-0,307	0,434	0,360	0,169
N_1_1		0,197	-0,141	-0,307	1,000	0,037	-0,456	-0,269
O_1_1A		0,496	0,028	0,434	0,037	1,000	-0,253	-0,242
O_1_10		-0,319	0,375	0,360	-0,456	-0,253	1,000	0,463
O_1_12		0,055	0,230	0,169	-0,269	-0,242	0,463	1,000

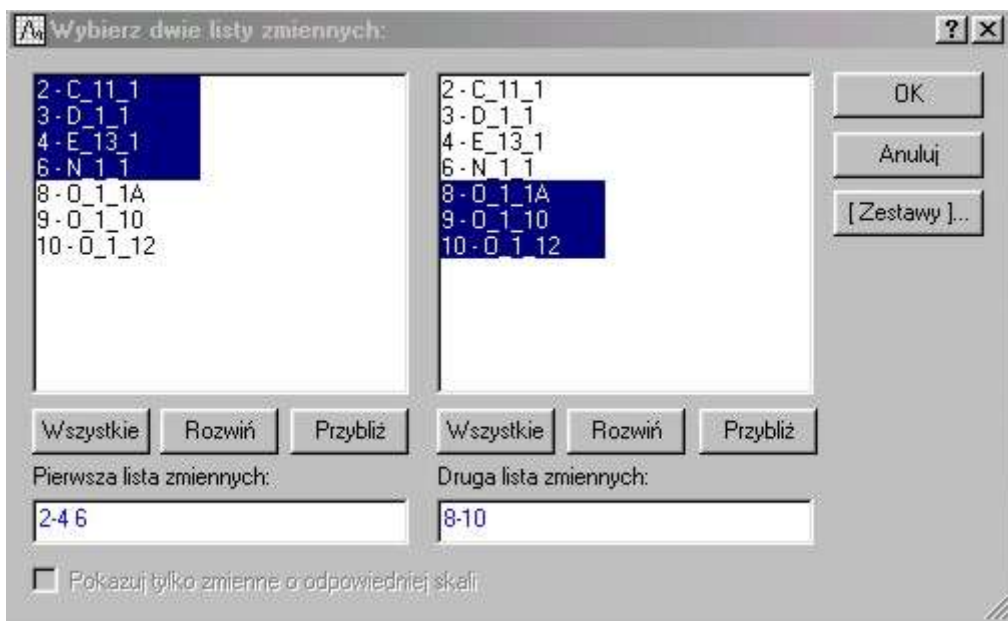
Aby przejść do właściwej analizy kanonicznej klikamy w oknie **Przegląd statystyk opisowych** klawisz **OK**. Pojawia się okno **Definicja modelu** (rys 7.8), w którym klikamy klawisz **Zmienne do analizy kanonicznej**.

Rys. 7.8. Okno Definicja modelu.



Otwiera się okno **Wybierz dwie listy zmiennych** (rys. 7.9).

Rys.7.9. Okno wyboru list zmiennych do analizy.



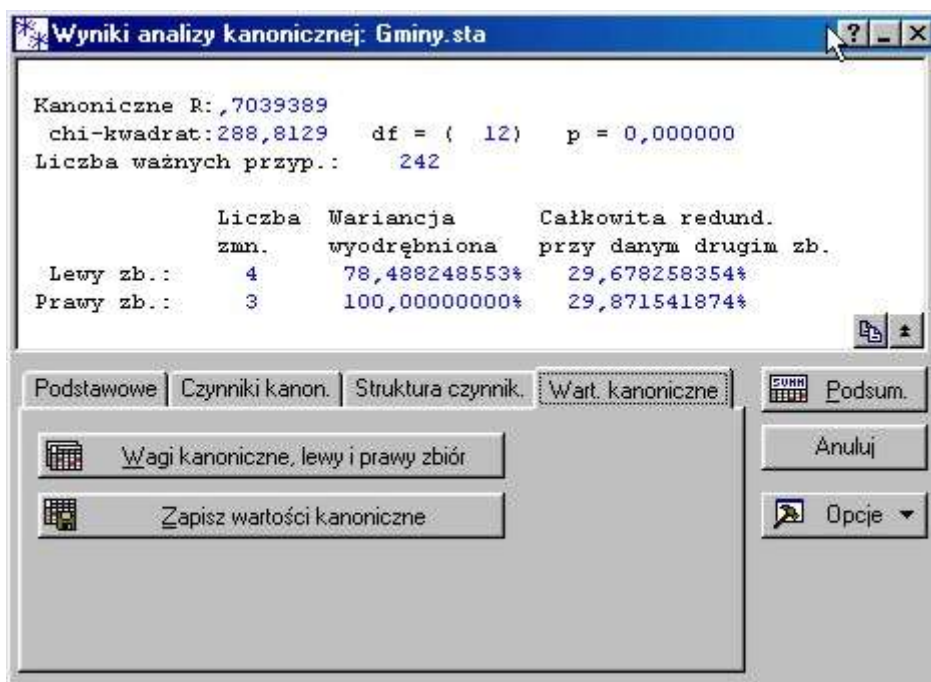
W oknie tym na pierwszej liście zaznaczamy zmienne należące do pierwszego zbioru, a na drugiej z list zmienne należące do drugiego zbioru. Następnie akceptujemy wybory klawiszem **OK**. Na ekranie pojawia się okno **Definicja modelu** z wstępnymi założeniami analizy kanonicznej (rys. 7.10).

Rys. 7.10. Okno analizy kanonicznej z przyjętymi założeniami.



Akceptujemy powyższe założenia klawiszem **OK** otrzymując okno **Wyniki analizy kanonicznej** (rys. 7.11).

Rys. 7.11. Okno Wyniki analizy kanonicznej.



W górnej części okna znajdują się wybrane wyniki analizy. Są one, wraz z wynikami szczegółowymi, przedstawione także na kartach **Czynniki kanoniczne**, **Struktura czynnikowa** oraz **Wartości kanoniczne** otwieranymi poprzez kliknięcie odpowiednich klawiszy.

Analizę wyników rozpoczynamy od wyliczenia wag kanonicznych. Dla otrzymania wag kanonicznych klikamy klawisz **Wartości kanoniczne, Wagi kanoniczne, lewy i prawy zbiór**. Uruchomienie klawisza powoduje pojawienie się tablic z wartościami wag kanonicznych (rys. 7.12).

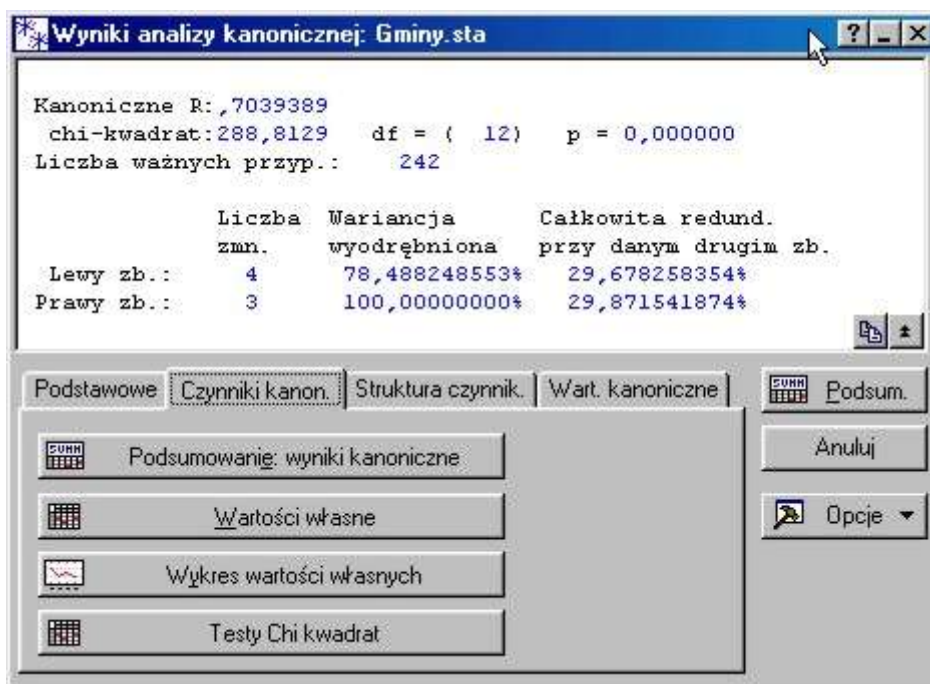
Rys. 7.12. Tablica z wagami kanonicznymi.

Zmienna	Wagi kanoniczne, prawy zbiór (Gm)			Wagi kanoniczne, lewy zbiór (Gm)		
	Pierw 1	Pierw 2	Pierw 3	Pierw 1	Pierw 2	Pierw 3
O_1_1A	-0,565774	-0,830108	-0,286469	0,228903	-0,937459	0,639035
O_1_10	-0,951617	0,532523	-0,345453	-0,292431	0,060213	0,558872
O_1_12	-0,068426	-0,512959	1,016494	-0,767507	-0,114559	-0,948592
				0,290922	-0,062860	-0,845078

Wagi te określają wkład poszczególnych zmiennych wejściowych w tworzenie kolejnych zmiennych kanonicznych. Ze względu na operowanie wystandaryzowanymi wartościami zmiennych wejściowych odpowiadają one współczynnikom beta w regresji wielorakiej. Czym

większa wartość bezwzględna wagi danej zmiennej wejściowej tym większy jej wkład do danej zmiennej kanonicznej. W naszym przykładzie wyznaczyliśmy trzy pary zmiennych kanonicznych. Zdecydowanie największe udziały w tworzeniu pierwszej pary zmiennych kanonicznych mają zmienne $D_{1.1}$ oraz $O_{1.10}$. Oznacza to, że przede wszystkim korelacja pomiędzy tymi zmiennymi wpłynęła na powstanie pierwszej korelacji kanonicznej pomiędzy dwoma wyróżnionymi zbiorami danych wejściowych. Z kolei zmienne $C_{11.1}$ i $O_{1.11}$ mają największy wkład do drugiej pary zmiennych kanonicznych, a zmienne $E_{13.1}$, $N_{1.1}$ oraz $O_{1.12}$ do trzeciej pary zmiennych kanonicznych. Aby zapisać obliczone wagi kanoniczne i wartości zmiennych kanonicznych dla badanych gmin klikamy przycisk **Zapisz wartości kanoniczne**. Do dalszej analizy brane są pod uwagę wyłącznie istotne pary zmiennych kanonicznych. Dla zbadania istotności par zmiennych kanonicznych wykorzystamy test istotności Barletta. W tym celu w oknie **Wyniki analizy kanonicznej** klikamy przycisk **Czynniki kanoniczne** otwierając kartę przedstawioną na rysunku 7.13.

Rys. 7.13. Karta Czynniki kanoniczne.



Następnie uruchamiamy klawisz **Testy Chi kwadrat** otwierając okno z wynikami testowania istotności kolejnych par zmiennych kanonicznych (rys. 7.14).

Rys. 7.14. Tablica z wynikami testowania istotności par zmiennych kanonicznych.

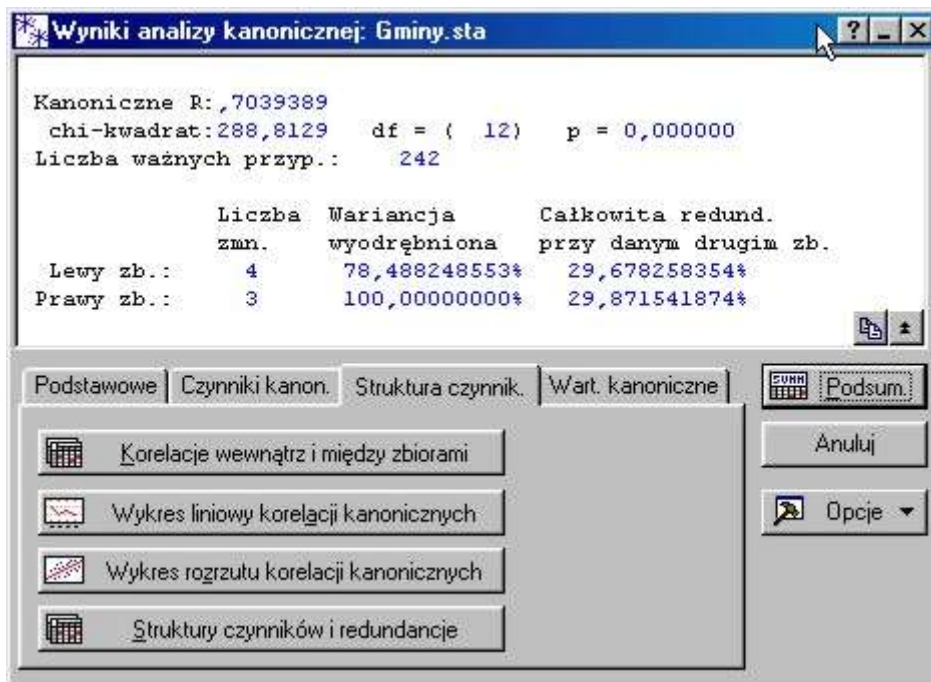
Pierw. Usunięty	Testy chi-kwadrat z usuwaniem kolejnych pierw. (Gminy)					
	Kanon. R	Kanon. R ²	Chi ²	df	p	Lambda pierwot.
0	0,703939	0,495530	288,8129	12	0,000000	0,295638
1	0,612559	0,375228	126,6464	6	0,000000	0,586036
2	0,248997	0,062000	15,1692	2	0,000509	0,938000

Weryfikacja istotności kolejnych par zmiennych kanonicznych odbywa się w sposób iteracyjny. Wartość krytyczna poziomu istotności (p) w pierwszym wierszu tablicy wskazuje, że przy każdym poziomie istotności $\alpha > 0,000000$ przynajmniej pierwsza para zmiennych kanonicznych jest istotna. Na podstawie krytycznej wartości poziomu istotności w drugim wierszu tabeli stwierdzamy, że przy każdym poziomie istotności $\alpha > 0,000000$ przynajmniej druga para zmiennych kanonicznych (trzeci wiersz tablicy) jest istotna. Trzecia para zmiennych kanonicznych (trzeci wiersz tablicy) jest istotna przy każdym poziomie istotności $\alpha > 0,000509$. Tym samym w dalszej analizie powinniśmy uwzględnić wszystkie trzy pary zmiennych kanonicznych.

W kolumnach tablicy przedstawionej na rysunku 7.14 znajdują się także inne informacje dotyczące przeprowadzonej analizy kanonicznej. W pierwszej jej kolumnie podane są wartości współczynnika korelacji kanonicznej pomiędzy zmiennymi z kolejnych par zmiennych kanonicznych (7.4). W drugiej kolumnie znajdują się wartości kwadratów tych korelacji kanonicznych. Stanowią one miary stopnia wyjaśniania zmienności (poprzez związki liniowe) jednego zbioru zmiennych wejściowych przez drugi ze zbiorów zmiennych wejściowych przez kolejne pary zmiennych kanonicznych. W sumie, poprzez wszystkie pary zmiennych kanonicznych, ponad 93% ($R_k^2 \cdot 100\%$) zmienności zmiennych jednego zbioru jest wyjaśniana zmiennością drugiego ze zbiorów. W kolejnych kolumnach podawane są wartości statystyki chi-kwadrat (7.13), testującej istotność par zmiennych kanonicznych, oraz liczbę jej stopni swobody.

W celu interpretacji zmiennych kanonicznych otwieramy w oknie **Wyniki analizy kanonicznej** kartę **Struktura czynnikowa** klikając odpowiedni klawisz (rys. 7.15).

Rys. 7.15. Karta Struktura czynnikowa.



W ramach tej karty dysponujemy czterema opcjami: **Korelacje wewnątrz i między zbiorami**, **Wykres liniowy korelacji kanonicznych**, **Wykres rozrzutu korelacji kanonicznych** oraz **Struktury czynników i redundancje**. Klikając ostatni z tych klawiszy otrzymujemy okna z wartościami kanonicznych ładunków czynnikowych oraz wariancji wyodrębnionych i redundancji (rys. 7.16).

Rys. 7.16. Tablice z wartościami kanonicznych ładunków czynnikowych oraz wariancji wyodrębnionych i redundancji.

Pierw. Zmienna	Struk.czynnik., lewy zbiór (Gminy.sta)		
	Pierw 1	Pierw 2	Pierw 3
C_11_1	0,027621	-0,994795	0,096675
D_1_1	-0,551974	0,095954	0,385629
E_13_1	-0,852220	-0,416937	-0,310251
N_1_1	0,612461	-0,220784	-0,506938

Pierw. Czynniki	Wariancja wyodrębniona (udziały), lewy zb.	
	Warianc. wyodrębn	Redund.
Pierw 1	0,351706	0,174281
Pierw 2	0,305352	0,114576
Pierw 3	0,127824	0,007925

Pierw. Zmienna	Struk.czynnik., prawy zbiór (Gminy.sta)		
	Pierw 1	Pierw 2	Pierw 3
O_1_1A	-0,308858	-0,840593	-0,444984
O_1_10	-0,840419	0,504472	0,198001
O_1_12	-0,372656	-0,065453	0,925658

Pierw. Zmienna	Wariancja wyodrębniona (udziały), prawy zb.	
	Warianc. wyodrębn	Redund.
Pierw 1	0,313523	0,155360
Pierw 2	0,321791	0,120745
Pierw 3	0,364686	0,022610

Kanoniczne ładunki czynnikowe są współczynnikami korelacji liniowej pomiędzy zmiennymi pierwotnymi, a zmiennymi kanonicznymi (7.16) i (7.17). Tym samym przy interpretacji zmiennych kanonicznych bierzemy pod uwagę te zmienne wejściowe, którym odpowiadają wysokie wartości odpowiednich kanonicznych ładunków czynnikowych (zwyczajowo ich

kwadraty powinny być większe niż 0,5). Najpierw przeanalizujemy kanoniczne ładunki czynnikowej dla pierwszego (lewego) zbioru zmiennych. W przypadku pierwszej zmiennej kanonicznej, decydujący kanoniczny ładunek czynnikowy wnosi zmienna $E_{13,1}(c_{3,1}=-0,852)$. Zmienność drugiej zmiennej kanonicznej w decydującym stopniu jest wyjaśniana zmiennością zmiennej $C_{11,1}(c_{12}=-0,995)$. Natomiast w przypadku trzeciej ze zmiennych kanonicznych żadna ze zmiennych wejściowych nie posiada wysokiego kanonicznego ładunku czynnikowego.

Podobnie analizujemy wartości kanonicznych ładunków czynnikowych dla drugiego (prawego) ze zbiorów zmiennych. Decydujący kanoniczny ładunek czynnikowy w przypadku pierwszej zmiennej kanonicznej wnosi zmienna $O_{1,10}(d_{2,1}=-0,840)$. Dla drugiej zmiennej kanonicznej decydujący kanoniczny ładunek czynnikowy wnosi zmienna $O_{1,11}(d_{1,2}=-0,841)$, a dla trzeciej zmiennej kanonicznej zmienna $O_{1,12}(d_{3,3}=-0,926)$.

Dla każdej ze zmiennych kanonicznych podane są wartości wariancji wyodrębnionych ((7.18) albo (7.19)) określające jaki procent wariancji zmiennych wejściowych wyjaśniają te zmienne kanoniczne. Przykładowo pierwsza ze zmiennych kanonicznych pierwszego typu wyjaśnia przeciętnie 35,2% wariancji pierwszego (lewego) zbioru zmiennych wejściowych. Natomiast pierwsza ze zmiennych kanonicznych drugiego typu wyjaśnia przeciętnie 31,4% wariancji drugiego (prawego) zbioru zmiennych wejściowych.

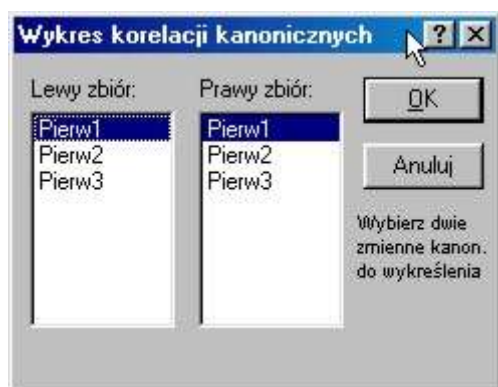
Przemnażając wariancję wyodrębnioną przez kwadrat korelacji kanonicznej otrzymujemy współczynniki redundancji (7.20) i (7.21). Miary te mówią jaki odsetek przeciętnej wariancji zmiennych wejściowych danego typu jest wyjaśniane przez zmienne kanoniczne drugiego typu.

Przy danym drugim (prawym) zbiorze zmiennych wejściowych możemy wyjaśnić odpowiednio 17,4%, 11,5% i 0,8% wariancji pierwszego (lewego) zbioru zmiennych wejściowych w oparciu o kolejną pierwszą, drugą i trzecią zmienną kanoniczną (rys. 7.16). Natomiast przy pierwszym (lewym) zbiorze zmiennych wejściowych wyjaśniamy odpowiednio 15,5% 12,1% i 0,2% wariancji zmiennych drugiego (prawego) zbioru zmiennych wejściowych na podstawie kolejno pierwszej, drugiej i trzeciej zmiennej kanonicznej. Użycie klawisza **Korelacje wewnątrz i między zbiorami** z karty **Struktura czynnikowa** otwiera macierze korelacji pomiędzy zbiorami zmiennych wejściowych

stanowiące bloki macierzy korelacji uzyskanej w ramach opcji okna **Przegląd statystyk opisowych** (rys. 7.7).

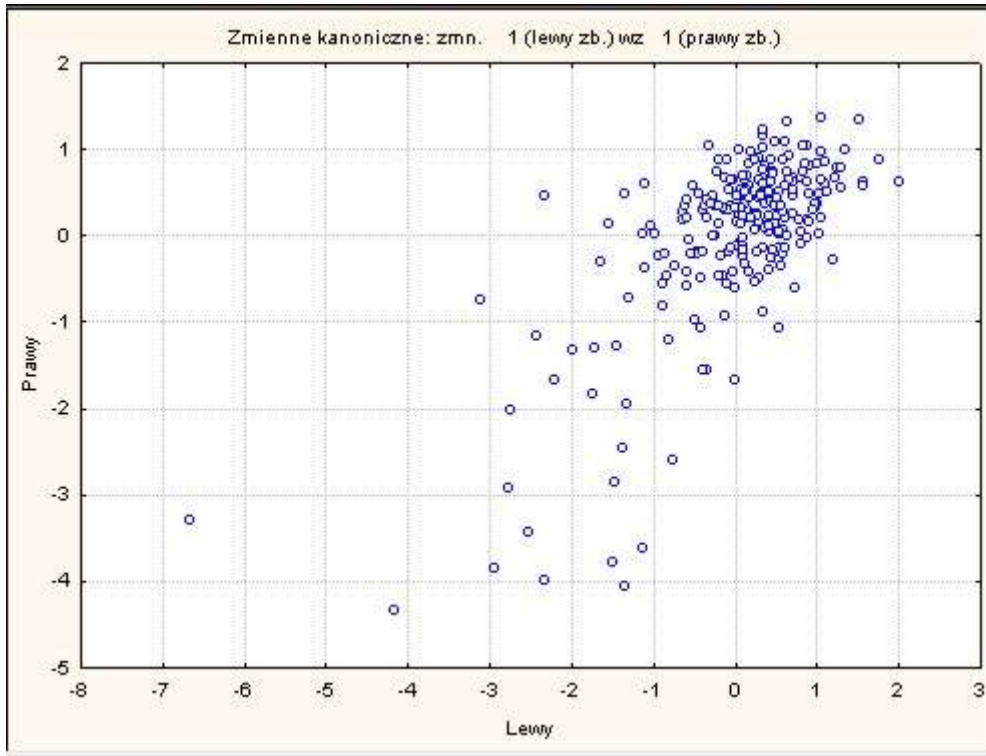
Dwa pozostałe klawisze na karcie **Struktura czynnikowa** wywołują okna z graficzną prezentacją wyników analizy kanonicznej. Kliknięcie klawisza **Wykres rozrzutu korelacji kanonicznej** powoduje otwarcie okna **Wykres korelacji kanonicznych** (rys. 7.17).

Rys. 7.17. Okno wyboru osi czynnikowych do wykresu korelacji kanonicznych.



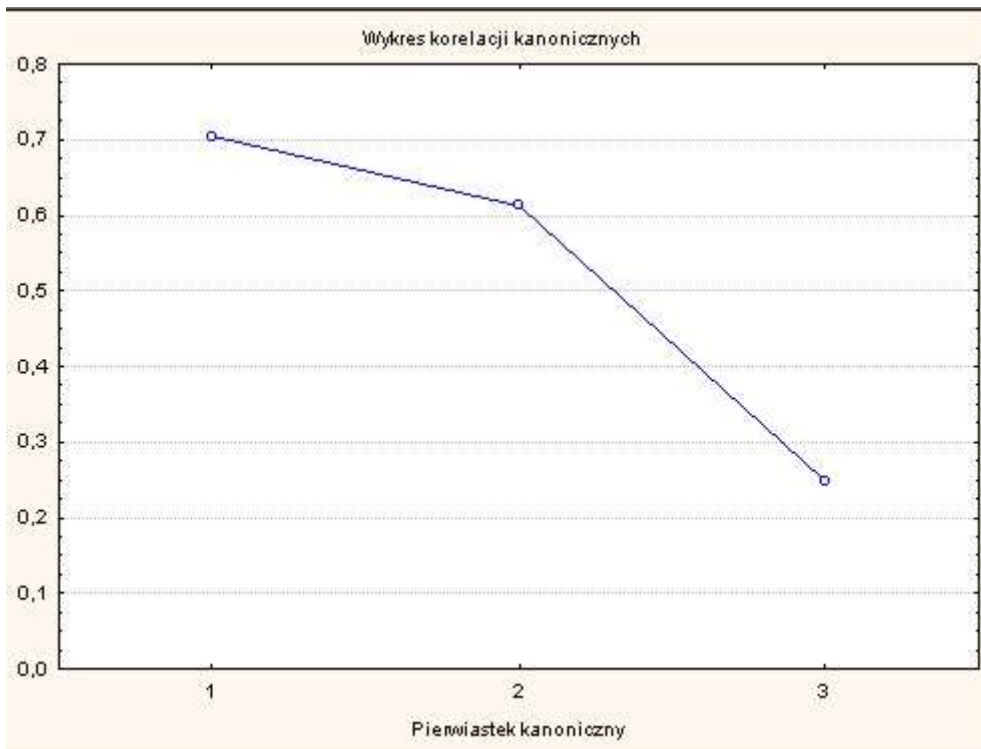
W oknie tym definiujemy parę zmiennych kanonicznych, które określają płaszczyznę (przestrzeń dwuwymiarową), na której zostaną umieszczone punkty reprezentujące gminy miejskie. Współrzędnymi tych punktów będą wartości współrzędnych kanonicznych, wybranej pary zmiennych kanonicznych, dla badanych gmin. W oknie zostawiamy wybraną domyślnie pierwszą zmienną kanoniczną dla pierwszego zbioru zmiennych wejściowych i pierwszą zmienną kanoniczną dla prawego zbioru zmiennych wejściowych. Wybór ten wynika z faktu, że właśnie pierwsza para zmiennych kanonicznych ma największy wkład w wyjaśnienie powiązań między zbiorami zmiennych wejściowych. Akceptujemy nasz wybór klawiszem **OK**. Pojawia nam się wykres punktów reprezentujących gminy na płaszczyźnie wyznaczonej przez pierwszą parę zmiennych kanonicznych (rys. 7.18).

Rys. 7.18. Wykres konfiguracji punktów reprezentujące gminy w układzie pierwszej pary zmiennych kanonicznych.



Na wykresie znajdują się punkty odstające, reprezentujące gminy o nietypowych wartościach zmiennych wejściowych. Większość punktów skupiona jest w prawym, górnym rogu wykresu. Reprezentują one gminy o zbliżonym układzie wartości zmiennych wejściowych. Jednocześnie punkty reprezentujące gminy nie są rozłożone wokół jakiejś linii regresji liniowej, co sugeruje, że istnieją jeszcze inne istotne pary zmiennych kanonicznych wyjaśniające powiązania między zbiorami zmiennych wejściowych. Kliknięcie na karcie **Struktura czynnikowa** klawisza **Wykres liniowy korelacji kanonicznych** powoduje otwarcie okna przedstawionego na rysunku 7.19.

Rys. 7.19. Wykres wartości korelacji kanonicznych.



Przedstawione na wykresie korelacje kanoniczne są pierwiastkami kwadratowymi z wartości własnych.

Podsumowanie wyników dotychczasowych analiz uzyskujemy klikając klawisz **Podsumowanie: wyniki kanoniczne** na karcie **Czynniki kanoniczne** (rys. 7.13). Otwieramy w ten sposób okno przedstawione na rysunku 7.20.

Rys. 7.20. Tablica z podsumowaniem analizy kanonicznej.

		Podsumow. analizy kanonicznej	
		Kanoniczne R: ,70394	
		Chi2(12)=288,81 p=0,00	
N=242		Lewy zb.	Prawy zb.
Liczba zmiennych		4	3
Wariancja wyodręb.		78,4883%	100,000%
Całkowita redund		29,6783%	29,8715%
Zmienne:	1	C_11_1	O_1_1A
	2	D_1_1	O_1_10
	3	E_13_1	O_1_12
	4	N_1_1	

W pierwszym wierszu tablicy mamy informacje o liczebnościach zbiorów zmiennych wejściowych. W drugim wierszu podane są wartości wariancji wyodrębnionych z pierwszego i drugiego zbioru zmiennych wejściowych. Wskazują one przeciętną ilość wariancji

wyodrębnionej ze zbiorów tych zmiennych przez wszystkie zmienne kanoniczne łącznie (ilość informacji przenoszonych przez wszystkie zmienne kanoniczne ze zbiorów zmiennych wejściowych). Wszystkie trzy zmienne kanoniczne wyodrębniają 78,49% wariancji pierwszego (lewego) zbioru zmiennych wejściowych i 100% wariancji z drugiego (prawego) ze zbiorów. Innymi słowy operując zmiennymi kanonicznymi tracimy ponad 20% informacji zawartych w pierwszym zbiorze zmiennych wejściowych. Nie występują natomiast straty informacji zawartych w drugim zbiorze zmiennych wejściowych, gdyż liczebność tego zbioru jest równa liczebności par zmiennych kanonicznych. Całkowite redundancje, których wartości są podane w kolejnym wierszu tabeli, są sumą redundancji kolejnych zmiennych kanonicznych. Przy znajomości wartości zmiennych z drugiego zbioru możemy wyjaśnić prawie 30% wariancji zmiennych z pierwszego zbioru. Podobnie, przy znajomości wartości zmiennych z pierwszego zbioru możemy wyjaśnić prawie 30% wariancji zmiennych z drugiego zbioru.

Na karcie **Czynniki kanoniczne** znajduje się jeszcze klawisz **Wartości własne**. Jego kliknięcie otwiera tablicę z wartościami własnymi (kwadratami korelacji kanonicznych) kolejnych par zmiennych kanonicznych (rys. 7.21).

Rys. 7.21. Tablica z wartościami własnymi par zmiennych kanonicznych.

	Wart. własne (Gminy.sta)		
Pierw.	Pierw 1	Pierw 2	Pierw 3
Wart.	0,495530	0,375228	0,062000

Wielkości te były już podane w tablicy na rysunku 7.14.