

ANALIZA WIELOWYMIAROWA

Temat: Ocena stanu środowiska w województwach za pomocą analizy wielowymiarowej w 2019r.

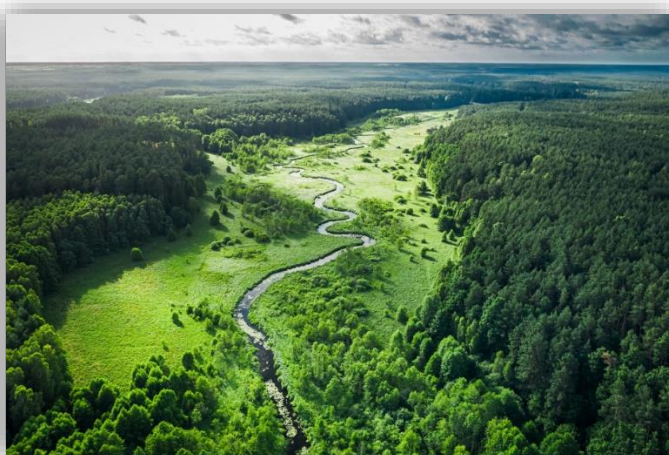
Spis treści

1. Wprowadzenie	3
2. Wstępna analiza	6
3. Porządkowanie liniowe	13
3.1. Metoda wzorca rozwoju Hellwiga	14
3.2. Metoda wzorca rozwoju Hellwiga z wagami	16
3.3. Metoda TOPSIS	18
3.4. Metoda standaryzowanych sum	20
3.5. Metoda sumy rang	21
3.6. Metoda dystansów	22
3.7. Podsumowanie metod porządkowania liniowego	24
4. Analiza głównych składowych	25
5. Analiza skupień	33
6. Podsumowanie, wnioski	39
7. Bibliografia	41

1. WPROWADZENIE

W obecnych czasach temat ochrony środowiska jest bardzo ważnym aspektem życia codziennego. Głównym powodem degradacji natury są działania człowieka, przede wszystkim w obszarze przemysłu. W związku z tym eksploatowane są coraz intensywniej nieodnawialne zasoby naturalne, a także odnawialne, takie jak przede wszystkim woda. Emitowane są ogromne ilości substancji trujących oraz niebezpiecznych do atmosfery. Zanieczyszczona zostaje woda, powietrze oraz podłoże. Obszary terenów zieleni z biegiem czasu ulegają degradacji oraz redukcji co do ich powierzchni.

Obecnie Polska jest krajem odznaczającym się wysokim stopniem zanieczyszczenia środowiska, skąd pojawia się potrzeba stałego monitorowania jego stanu, w celu zapobiegania spotęgowania działalności tego procesu. Ważny jest również aspekt kontroli obszarów zieleni, ich ilości oraz jakości.



Celem analizy jest ocena stanu środowiska w województwach na terenie Polski w 2019r., na podstawie danych statystycznych pochodzących z Głównego Urzędu Statystycznego, z pomocą wybranych metod analizy wielowymiarowej. W analizie danych wykorzystane zostanie porządkowanie liniowe, analiza głównych składowych oraz analiza skupień. W analizie danych empirycznych zastosowano następujące metody porządkowania liniowego obiektów wielowymiarowych:

- metodę wzorca rozwoju Hellwiga,
- metodę wzorca z wagami,
- metodę TOPSIS,
- metodę standaryzowanych sum,
- metodę sumy rang,
- metodę dystansów.

Problemem badawczym jest pytanie:

Które województwa na terenie Polski odznaczają się dobrym, a które złym stanem środowiska naturalnego, na podstawie wybranych zmiennych?

W oparciu o wyniki z analiz zostaną utworzone rankingi, podziały województw, zaczynając od tych, które charakteryzują się najlepszym stanem środowiska do tych, które odznaczają się najgorszym stanem natury.

Do przeprowadzenia badania stanu środowiska naturalnego w Polsce w 2019r. wybrano 16 następujących zmiennych:

X1	Zużycie wody na potrzeby gospodarki narodowej i ludności [hm ³ /100km ²]
X2	Ścieki przemysłowe i komunalne nieoczyszczane odprowadzone z zakładów przemysłowych do wód lub ziemi [dam ³ /100km ²]
X3	Liczba przemysłowych i komunalnych oczyszczalni ścieków przypadająca na 10 000 tys. mieszkańców [sztuki]
X4	Obszary Natura 2000 specjalnej ochrony ptaków [km ² /10 km ²]
X5	Liczba pomników przyrody przypadająca na 1000 tys. mieszkańców
X6	Masa wytworzonych odpadów komunalnych przez 10 mieszkańców [tona]
X7	Liczba czynnych składowisk odpadów przypadająca na 100 000 tys. mieszkańców
X8	Nakłady finansowe na środki trwałe na ochronę środowiska i gospodarki wodnej przypadające na 0,001 mieszkańca [zł]
X9	Nasadzenia drzew w miastach i na wsiach [sztuk/1km ²]
X10	Liczba pożarów lasów przypadająca na 100 km ² powierzchni
X11	Powierzchnia ekologicznych gospodarstw rolnych przypadająca na 1000 km ² powierzchni [km ²]
X12	Grunty rolne i leśne wyłączone z produkcji rolniczej i leśnej na tereny osiedlowe [km ² /10000km ²]
X13	Grunty zdewastowane i zdegradowane zrehabilitowane i zagospodarowane na cele rolnicze i leśne [km ² /10000km ²]
X14	Produkcja energii elektrycznej z odnawialnych nośników energii [GWh/10 km ²]
X15	Emisja tlenków azotu [tona/10 km ²]
X16	Emisja metali ciężkich z zakładów szczególnie uciążliwych przypadających na 10 km ² powierzchni [kg]

Powyższe zmienne zostały podzielone na bloki oraz na stymulanty i destymulanty. Stymulanty są to zmienne pozytywne, których wysokie wartości są korzystne z punktu widzenia wpływu na stan środowiska naturalnego. Niskie wartości tych zmiennych są niepożądane. Destymulanty są to zmienne negatywne, których niskie wartości są korzystne, natomiast wartości wysokie niewskazane.

1. Zasoby wodne

D	X1	Zużycie wody na potrzeby gospodarki narodowej i ludności [hm ³ /100km ²]
D	X2	Ścieki przemysłowe i komunalne nieoczyszczane odprowadzone z zakładów przemysłowych do wód lub ziemi [dam ³ /100km ²]
S	X3	Liczba przemysłowych i komunalnych oczyszczalni ścieków przypadająca na 10 000 tys. mieszkańców [sztuki]

2. Obszary prawnie chronione

S	X4	Obszary Natura 2000 specjalnej ochrony ptaków [km ² /10 km ²]
S	X5	Liczba pomników przyrody przypadająca na 1000 tys. mieszkańców

3. Odpady

D	X6	Masa wytworzonych odpadów komunalnych przez 10 mieszkańców [tona]
D	X7	Liczba czynnych składowisk odpadów przypadająca na 100 000 tys. mieszkańców

4. Nakłady finansowe

S	X8	Nakłady finansowe na środki trwałe na ochronę środowiska i gospodarki wodnej przypadające na 0,001 mieszkańca [zł]
----------	----	--

5. Zadrzewienie

S	X9	Nasadzenia drzew w miastach i na wsiach [sztuk/1km ²]
D	X10	Liczba pożarów lasów przypadająca na 100 km ² powierzchni

6. Rolnictwo

S	X11	Powierzchnia ekologicznych gospodarstw rolnych przypadająca na 1000 km ² powierzchni [km ²]
D	X12	Grunty rolne i leśne wyłączone z produkcji rolniczej i leśnej na tereny osiedlowe [km ² /10000km ²]
S	X13	Grunty zdewastowane i zdegradowane zrekultywowane i zagospodarowane na cele rolnicze i leśne [km ² /10000km ²]

7. Energia

S	X14	Produkcja energii elektrycznej z odnawialnych nośników energii [GWh/10 km ²]
---	-----	--

8. Powietrze

D	X15	Emisja tlenków azotu [tona/10 km ²]
D	X16	Emisja metali ciężkich z zakładów szczególnie uciążliwych przypadających na 10 km ² powierzchni [kg]

W badaniu wzięto pod uwagę 16 następujących województw (n=16):

- Dolnośląskie
- Kujawsko-pomorskie
- Lubelskie
- Lubuskie
- Łódzkie
- Małopolskie
- Mazowieckie
- Opolskie
- Podkarpackie
- Podlaskie
- Pomorskie
- Śląskie
- Świętokrzyskie
- Warmińsko-mazurskie
- Wielkopolskie
- Zachodniopomorskie

2. WSTĘPNA ANALIZA

Poniżej przedstawiona jest uporządkowana tabela 2.1 wartości zmiennych, doprowadzonych do jednego rzędu. Zmienne x11 oraz x15 zostały zlogarytmowane ze względu na zbyt odstające wartości od reszty.

Tabela 2.1 Wartości zmiennych

	D	D	S	S	S	D	D	S	S	D	S	D	S	S	D	D
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16
dolnośląskie	1,87	22,12	1,01	1,47	0,88	4,04	0,86	4,39	0,89	2,67	1,16	1,23	0,69	0,39	0,68	4,09
kujawsko-pomorskie	1,28	9,25	0,89	0,88	1,19	3,32	1,11	2,48	0,66	2,94	0,63	0,33	0,27	1,96	0,71	0,78
lubelskie	1,10	2,34	1,65	1,34	0,68	2,34	1,28	3,80	0,20	1,77	1,06	0,42	0,52	0,23	0,33	1,89
lubuskie	0,54	0,06	1,25	2,10	1,38	3,80	0,99	3,12	0,28	4,17	1,47	0,66	0,33	0,52	0,46	0,10
łódzkie	1,43	27,15	1,10	0,22	0,98	3,33	0,73	3,67	0,69	4,48	0,71	1,13	0,81	0,99	1,23	3,27
małopolskie	2,59	91,07	0,92	0,88	0,64	3,17	0,41	3,52	0,83	1,25	0,81	1,78	1,21	0,36	0,90	7,77
mazowieckie	6,89	16,09	0,81	1,21	0,75	3,21	0,46	5,13	0,66	7,59	1,09	0,57	0,59	0,49	0,85	5,67
opolskie	1,41	0,07	1,10	0,15	0,67	3,39	1,83	4,76	0,96	1,59	0,54	0,32	0,56	0,53	1,17	5,96
podkarpackie	1,35	4,55	1,31	2,85	0,96	2,42	0,56	4,07	0,59	1,82	0,89	0,83	0,77	0,37	0,46	2,84
podlaskie	0,44	0,21	1,25	2,87	1,67	2,83	0,93	3,48	0,54	1,61	1,41	0,34	0,54	0,48	0,07	0,32
pomorskie	1,01	0,16	0,78	1,99	1,19	3,72	0,51	3,62	0,63	2,18	1,06	2,82	0,49	1,33	0,43	1,31
śląskie	2,94	446,87	0,74	0,51	0,32	3,76	0,42	5,74	1,44	4,14	0,46	1,77	1,68	0,77	1,39	38,04
świętokrzyskie	11,74	68,14	1,21	0,19	0,59	2,32	1,05	3,69	0,21	5,20	0,88	0,45	0,26	1,73	1,04	2,13
warmińsko-mazurskie	0,50	14,46	1,89	2,38	1,64	3,06	0,70	2,34	0,29	1,07	1,65	0,29	0,88	0,54	0,16	0,17
wielkopolskie	4,34	1,17	1,19	1,37	1,07	3,55	0,80	3,76	0,61	2,52	0,97	0,43	0,97	0,81	0,62	5,42
zachodniopomorskie	4,19	1,54	1,64	3,03	1,56	3,91	0,77	3,33	0,37	2,24	1,65	0,72	0,38	1,92	0,45	0,27
średnia arytmetyczna	2,73	44,08	1,17	1,47	1,01	3,26	0,84	3,81	0,62	2,95	1,03	0,88	0,68	0,84	0,68	5,00
odchylenie standardowe	2,87	107,05	0,32	0,95	0,39	0,53	0,36	0,85	0,31	1,69	0,36	0,69	0,36	0,56	0,37	8,84
współczynnik zmienności [%]	105,39	242,86	27,39	64,87	38,66	16,38	42,74	22,44	50,59	57,39	34,84	78,15	52,82	66,97	54,61	176,73

Źródło: Opracowanie własne na podstawie danych z Głównego Urzędu Statystycznego oraz obliczeń w programie Excel

W powyższej tabeli zawarte są również obliczenia dotyczące średniej arytmetycznej, odchylenia standardowego oraz współczynnika zmienności dla podanych zmiennych. Najmniejszy współczynnik zmienności można odnotować dla zmiennej opisującej masę wytworzonych odpadów komunalnych przez 10 mieszkańców (x6) równy 16,40%, natomiast największy dla zmiennej dotyczącej objętości nieoczyszczonych ścieków przemysłowych i komunalnych odprowadzonych z zakładów przemysłowych do wód lub ziemi (x2) wynoszący 242,86%. Przyjęto minimalny próg równy 10% dla współczynnika zmienności, tak aby zmienna wzięła udział w badaniu. Im większa wartość współczynnika zmienności tym cecha lepiej różnicuje jednostki i jest bardziej pożądana w badaniu. Taki wynik świadczy o niskiej jednorodności zmiennej. Na tej podstawie nie ma powodów do odrzucenia jakiegokolwiek zmiennej.

W tabeli 2.2 zostały zawarte podstawowe statystyki opisowe dla zmiennych. Na jej podstawie można zauważyć, że zmienna x2 (Ścieki przemysłowe i komunalne nieoczyszczane odprowadzone z zakładów przemysłowych do wód lub ziemi [$\text{dam}^3/100\text{km}^2$]) oraz x16 (Emisja metali ciężkich z zakładów szczególnie uciążliwych przypadających na 10 km^2 powierzchni [kg]) istotnie odznaczają się od reszty.

Ze względu na jednostki indywidualne wartości minimum oraz maksimum są od siebie znacząco oddalone. Również taka sama sytuacja jest w przypadku dolnego i górnego kwartylu. Przykładowo kwartył dolny, który dzieli uporządkowaną niemalejąco zbiorowość na 2 części w ten sposób, że 25% jednostek zbiorowości ma wartości zmiennej mniejsze lub równe kwartyłowi pierwszemu, w przypadku zmiennej x8 (wartość nakładów finansowych

na środki trwałe na ochronę środowiska i gospodarki wodnej przypadające na 0,001 mieszkańca [zł]) stanowi 3,41zł, natomiast górny kwartyl wynosi 4,23 zł i stanowi 25% województw. W przypadku zmiennej x2 wartości te mocno od siebie odbiegają, stanowią kolejno 0,69 i 24,64 dam³/100km², przy czym maksimum wynosi aż 446,87 dam³/100km².

Wariancja:

Wariancja jest miarą zmienności obserwowalnych wyników. Pokazuje ona jak duże jest, zróżnicowanie wyników w badanym zbiorze zmiennych. Ponownie zmienna x2 odznacza się bardzo istotnie różną od pozostałych wariancją równą 12223,56 oraz zmienna x16, której wariancja wynosi 83,35. Zbiorowości tych 2 zmiennych są bardzo zróżnicowane. Kolejno największą wariancją odznacza się zmienna x1 (Zużycie wody na potrzeby gospodarki narodowej i ludności [hm³/100km²]) oraz x10 (Liczba pożarów lasów przypadająca na 100 km² powierzchni).

Odchylenie standardowe:

Jest to wartość, która informuje o tym czy rozrzut wyników wokół średniej jest niewielki czy wielki. Przykładowo wartości zmiennej x2 średnio odchyła się od średniej arytmetycznej o $\pm 110,56$ dam³/100km², zmiennej x16 o $\pm 9,13$ kg, a x4 (Obszary Natura 2000 specjalnej ochrony ptaków) o 0,96 [km²/10 km²].

Skośność:

Wartość skośności masy wytworzonych odpadów komunalnych przez 10 mieszkańców [tona] jako jedyna jest ujemna, co wskazuje na lewoskośny rozkład asymetrii. Dla pozostałych zmiennych wartości skośności są dodatnie co świadczy o prawoskośnym rozkładzie. Zmienna x2 oraz x16 odznaczają się bardzo dużą skośnością, wykraczającą poza wartości $\pm 3,09$ dlatego zmienne te zostały odrzucone w dalszej analizie.

Kurtoza:

Kurtoza jest miarą występowania wartości odstających. Im wyższa wartość kurtozy tym więcej pojawia się wartości bliskich skrajnych lub skrajnych – ponownie zmienna x2 oraz x16. Ujemne wartości kurtozy – np. zmienna x5 (Liczba pomników przyrody przypadająca na 1000 tys. mieszkańców) oraz x14 (Produkcja energii elektrycznej z odnawialnych nośników energii [GWh/10 km²]) wskazują na brak wartości odstających i większe skupienie wyników wokół średniej.

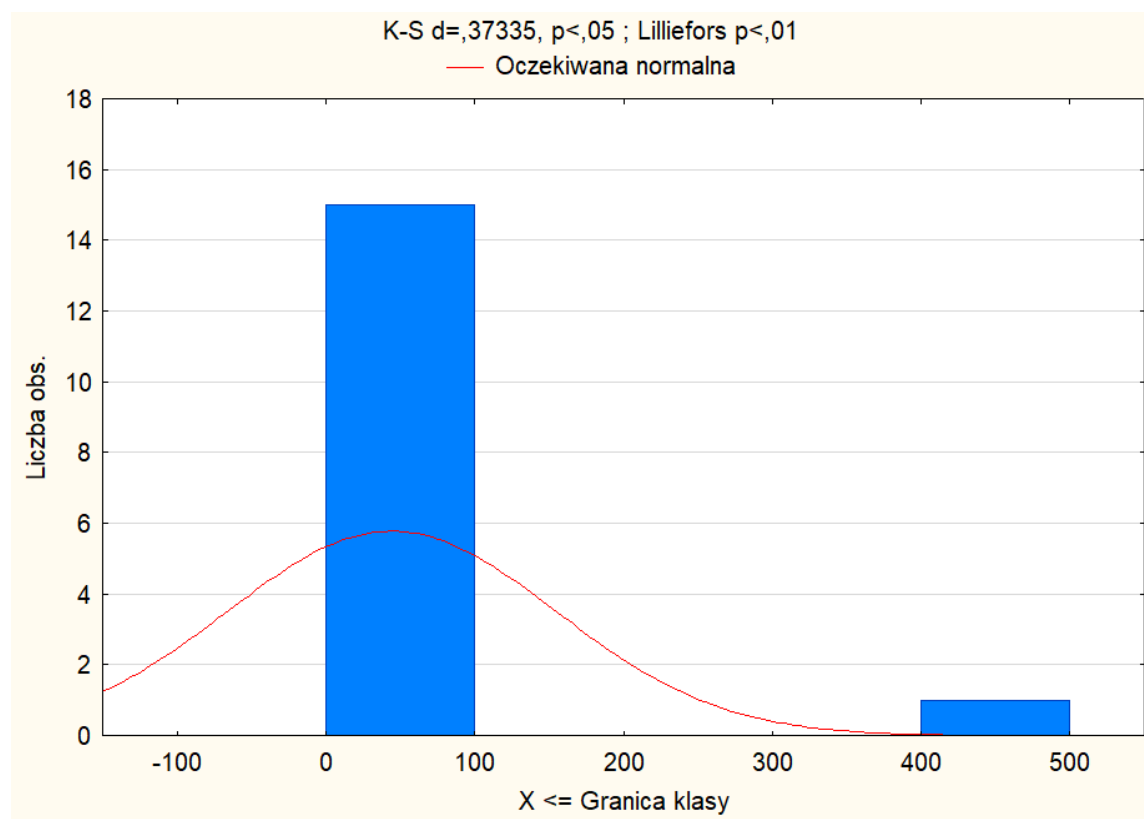
Tabela 2.2 Statystyki opisowe

Zmienna	Statystyki opisowe (AW-ochrona-srodowiska)												
	Nważnych	Średnia	Mediana	Minimum	Maksimum	Dolny Kwartyl.	Górny Kwartyl.	Wariancja	Odch.std	Wsp.zmn.	Standard. Błąd	Skośność	Kurtoza
x1	16	2,73	1,42	0,44	11,74	1,06	3,57	8,81	2,97	108,84	0,74	2,21	5,31
x2	16	44,08	6,90	0,06	446,87	0,69	24,64	12223,56	110,56	250,83	27,64	3,65	13,87
x3	16	1,17	1,15	0,74	1,89	0,91	1,28	0,11	0,33	28,29	0,08	0,75	0,04
x4	16	1,47	1,36	0,15	3,03	0,70	2,24	0,96	0,98	66,99	0,25	0,23	-1,17
x5	16	1,01	0,97	0,32	1,67	0,68	1,29	0,16	0,40	39,93	0,10	0,25	-0,83
x6	16	3,26	3,33	2,32	4,04	2,95	3,74	0,30	0,55	16,92	0,14	-0,52	-0,72
x7	16	0,84	0,79	0,41	1,83	0,54	1,02	0,14	0,37	44,14	0,09	1,25	2,23
x8	16	3,81	3,68	2,34	5,74	3,41	4,23	0,78	0,88	23,18	0,22	0,52	0,57
x9	16	0,62	0,62	0,20	1,44	0,33	0,76	0,10	0,32	52,25	0,08	0,93	1,58
x10	16	2,95	2,38	1,07	7,59	1,69	4,16	3,06	1,75	59,27	0,44	1,39	1,92
x11	16	1,03	1,02	0,46	1,65	0,76	1,29	0,14	0,37	35,98	0,09	0,33	-0,73
x12	16	0,88	0,62	0,29	2,82	0,38	1,18	0,51	0,71	80,71	0,18	1,65	2,53
x13	16	0,68	0,58	0,26	1,68	0,44	0,85	0,14	0,37	54,55	0,09	1,36	2,22
x14	16	0,84	0,54	0,23	1,96	0,44	1,16	0,34	0,58	69,16	0,15	1,09	-0,21
x15	16	0,68	0,65	0,07	1,39	0,44	0,97	0,15	0,39	56,40	0,10	0,28	-0,76
x16	16	5,00	2,49	0,10	38,04	0,55	5,55	83,35	9,13	182,53	2,28	3,55	13,42

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

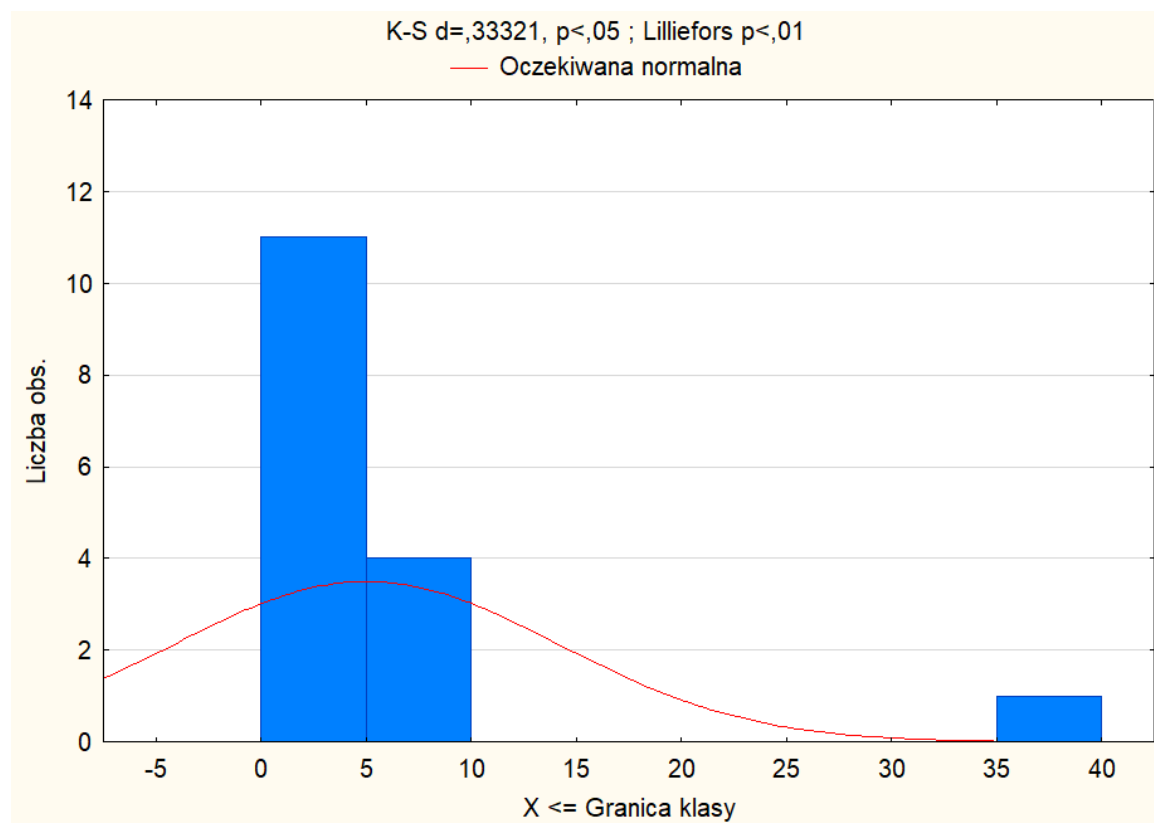
Poniżej zostały przedstawione przykładowe histogramy wybranych zmiennych. Na początku zostały zaprezentowane histogramy zmiennych nietypowych tzn. x2 oraz x16, są one dosyć nieczytelne i prezentują skrajne asymetrie. Kolejno zostały przedstawione histogramy zmiennych x7 oraz x4, które charakteryzują się asymetrią prawostronną. Na końcu zaprezentowany został histogram zmiennej x6, która jako jedyna ma asymetrię lewostronną.

Wykres 1.1 Histogram zmiennej x2



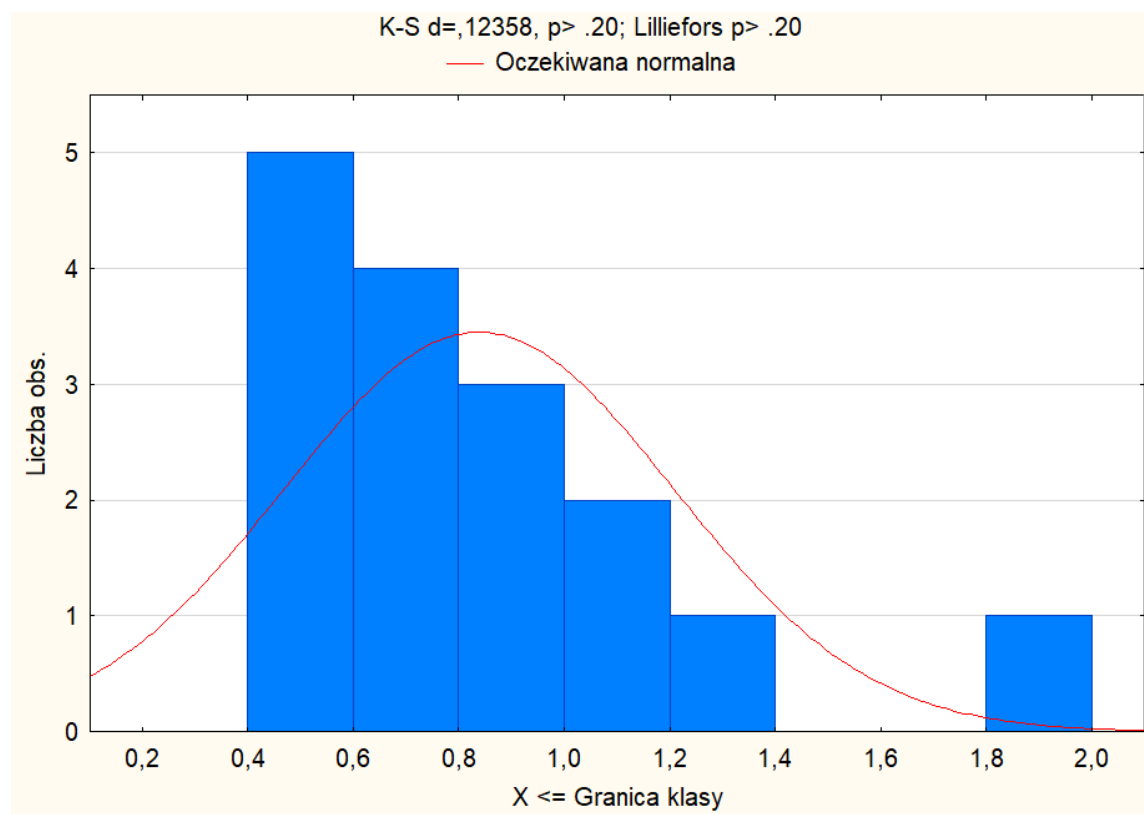
Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Wykres 1.2 Histogram zmiennej x16



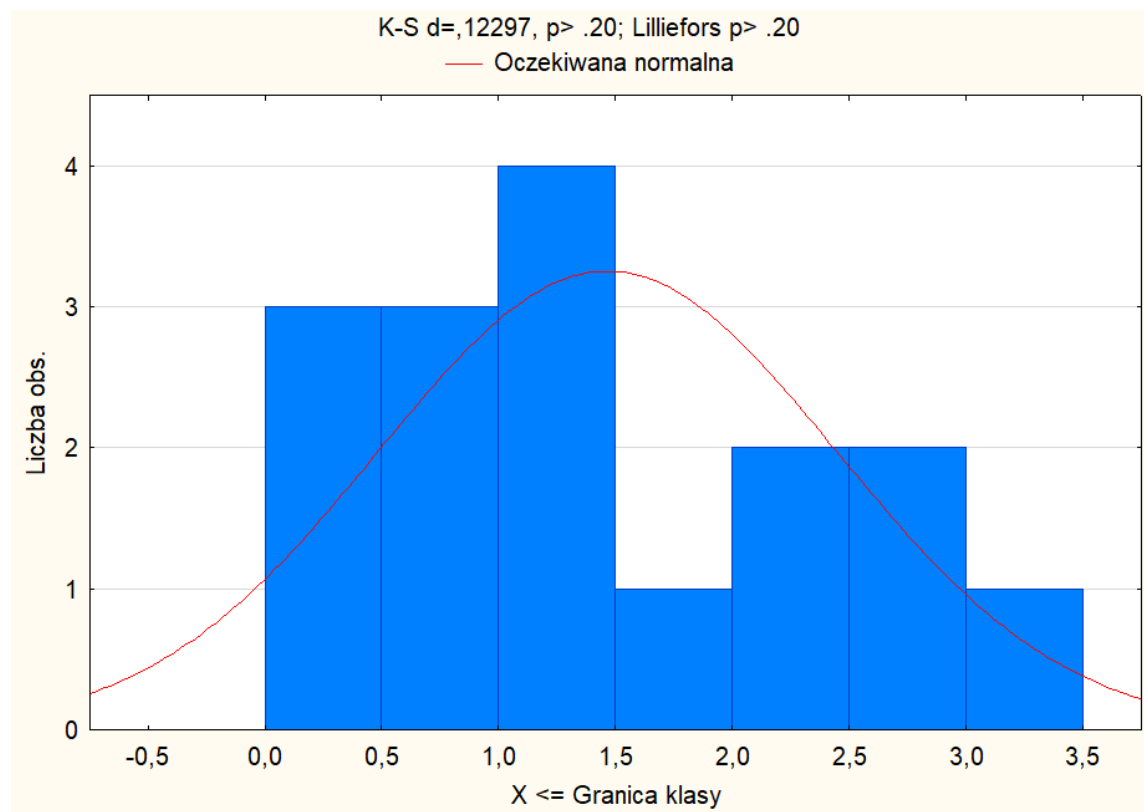
Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Wykres 1.3 Histogram zmiennej x7



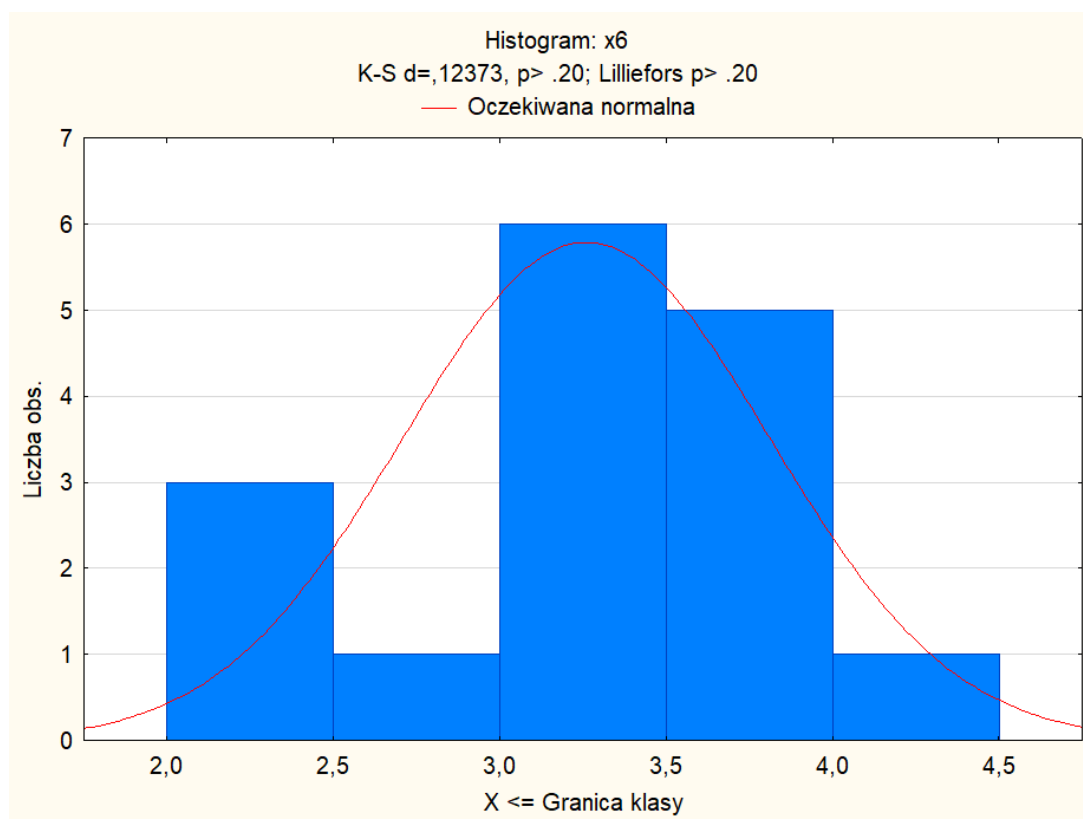
Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Wykres 1.4 Histogram zmiennej x4



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Wykres 1.5 Histogram zmiennej x6



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

W tabeli 2.3 przedstawione są wartości korelacji wszystkich zmiennych, które przygotowano do badania. Kolorem zielonym zostały zaznaczone przykładowe wysokie wartości korelacji (powyżej $\pm 0,5$) jakie zmienne między sobą tworzą. Kolorem żółtym zostały zaznaczone korelacje o bardzo niskich wartościach, co świadczy, że zmienne nie budują między sobą istotnych związków – zależność liniowa. Główną przekątną macierzy korelacji stanowią jedynki – korelacja zmiennej z samą sobą stanowi 1,00.

Tabela 2.3 Macierz korelacji wszystkich zmiennych

Korelacje (AW-ochrona-srodowiska)																
Oznaczone wsp. korelacji są istotne z $p < ,05000$																
N=16 (Braki danych usuwano przypadkami)																
Zmienna	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16
x1	1,00	0,13	-0,14	-0,36	-0,40	-0,25	-0,09	0,25	-0,16	0,61	-0,12	-0,15	-0,14	0,39	0,38	0,08
x2	0,13	1,00	-0,39	-0,36	-0,55	0,18	-0,36	0,57	0,68	0,21	-0,46	0,38	0,75	-0,02	0,57	0,97
x3	-0,14	-0,39	1,00	0,50	0,51	-0,34	0,26	-0,52	-0,70	-0,43	0,66	-0,55	-0,25	-0,08	-0,59	-0,46
x4	-0,36	-0,36	0,50	1,00	0,76	0,03	-0,27	-0,38	-0,41	-0,39	0,79	-0,06	-0,21	-0,09	-0,86	-0,39
x5	-0,40	-0,55	0,51	0,76	1,00	0,17	-0,01	-0,72	-0,52	-0,34	0,78	-0,25	-0,42	0,19	-0,77	-0,61
x6	-0,25	0,18	-0,34	0,03	0,17	1,00	-0,16	0,13	0,45	0,06	0,10	0,39	0,17	0,17	0,18	0,24
x7	-0,09	-0,36	0,26	-0,27	-0,01	-0,16	1,00	-0,10	-0,20	-0,22	-0,13	-0,57	-0,51	0,04	0,02	-0,31
x8	0,25	0,57	-0,52	-0,38	-0,72	0,13	-0,10	1,00	0,68	0,41	-0,50	0,25	0,47	-0,34	-0,60	0,70
x9	-0,16	0,68	-0,70	-0,41	-0,52	0,45	-0,20	0,68	1,00	0,06	-0,65	0,45	0,70	-0,17	0,64	0,79
x10	0,61	0,21	-0,43	-0,39	-0,34	0,06	-0,22	0,41	0,06	1,00	-0,17	-0,05	-0,10	0,17	0,48	0,20
x11	-0,12	-0,46	0,66	0,79	0,78	0,10	-0,13	-0,50	-0,65	-0,17	1,00	-0,23	-0,37	-0,04	-0,80	-0,54
x12	-0,15	0,38	-0,55	-0,06	-0,25	0,39	-0,57	0,25	0,45	-0,05	-0,23	1,00	0,39	0,03	0,22	0,36
x13	-0,14	0,75	-0,25	-0,21	-0,42	0,17	-0,51	0,47	0,70	-0,10	-0,37	0,39	1,00	-0,41	0,40	0,80
x14	0,39	-0,02	-0,08	-0,09	0,19	0,17	0,04	-0,34	-0,17	0,17	-0,04	0,03	-0,41	1,00	0,14	-0,13
x15	0,38	0,57	-0,59	-0,86	-0,77	0,18	0,02	-0,60	0,64	0,48	-0,80	0,22	0,40	0,14	1,00	0,62
x16	0,08	0,97	-0,46	-0,39	-0,61	0,24	-0,31	0,70	0,79	0,20	-0,54	0,36	0,80	-0,13	0,62	1,00

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

W tabeli 2.4 przedstawiona jest korelacja cząstkowa pomiędzy zmiennymi x2 oraz x16. Korelacja cząstkowa używana jest między 2 zmiennymi po uwzględnieniu wpływu innego czynnika (jednego lub wielu) na obie analizowane zmienne, dla cech o wartościach 0,9 i większej. Pomiedzy wspomnianymi zmiennymi zachodzi dodatnia korelacja. W związku z tym należałoby wysunąć wnioski, że wraz ze wzrostem ilości ścieków przemysłowych i komunalnych nieoczyszczanych odprowadzonych z zakładów przemysłowych do wód lub ziemi [$\text{dm}^3/100\text{km}^2$] rośnie emisja metali ciężkich z zakładów szczególnie uciążliwych przypadających na 10 km^2 powierzchni [kg]. W związku pomiędzy zmiennymi trzeba uwzględnić wielkość przemysłu, który może, lecz nie musi powodować jednoczesną emisję obu związków natomiast nie są one ze sobą ściśle zależne i bezpośrednio na siebie nie wpływają.

Tabela 2.4 Korelacja cząstkowa pomiędzy zmiennymi x2 i x16

Zmienna	Aktualnie w równaniu są zmienne ; DV: x2 (AW-ochrona-srodowiska)						
	b* w	Cząstk. Korelac.	Semicz. Korelac.	Tolerancja	R-kwadr.	t(14)	p
x16	0,966	0,966	0,966	1,000	0,0	14,02	0,000

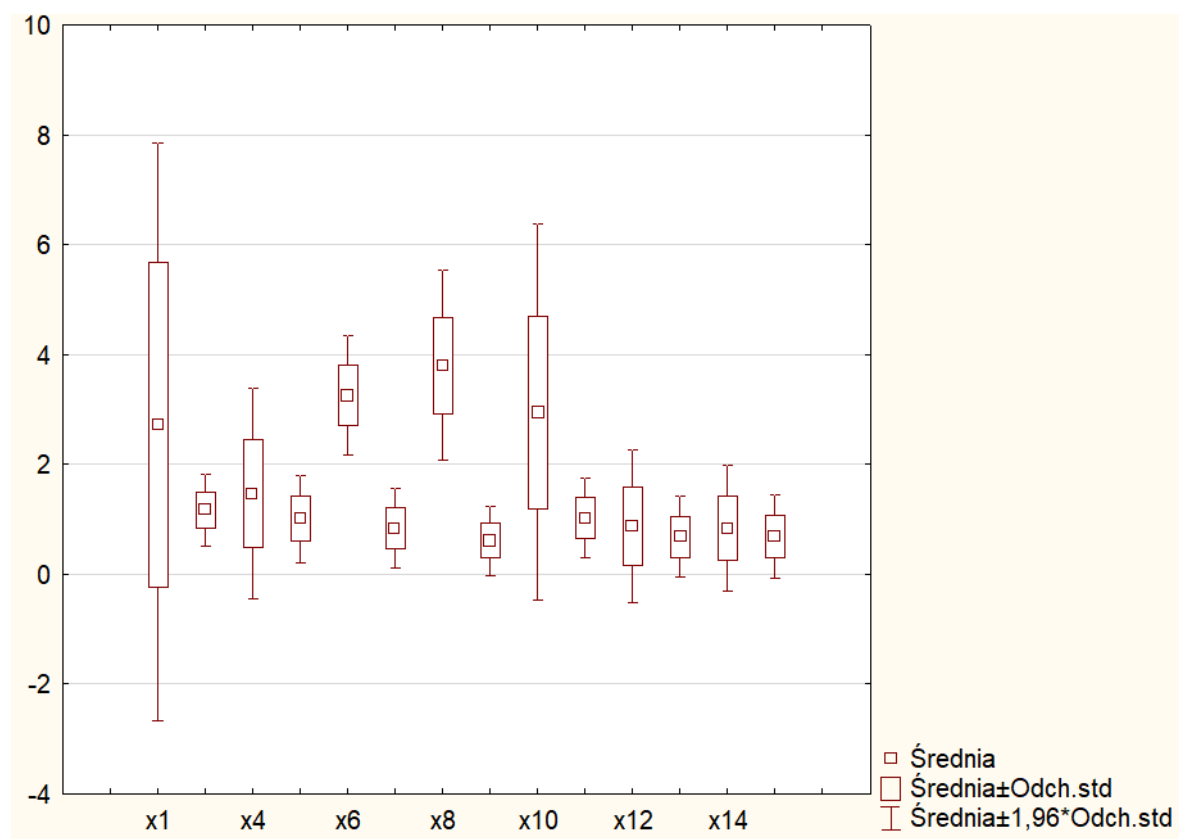
Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Finalnie do porządkowania liniowego zostały wyeliminowane tylko dwie zmienne – x2 oraz x16. Do przeprowadzenia analizy PCA oraz analizy skupień zostały wyeliminowane zmienne x1, x2, x6, x7, x10, x12, x14 oraz x16, co zostało dokładniej opisane w rozdziale 4.

Poniżej przedstawione są 2 wykresy ramka-wąsy pokazujące zakresy rozpiętości wartości zmiennych, które zostały przyjęte do analiz. Wykresy te są przedstawieniem w formie graficznej wyników podstawowych statystyk. Zróznicowanie wynika stąd, że niektóre zmienne mają większy rozrzut w ekstremach, a inne mniejsze.

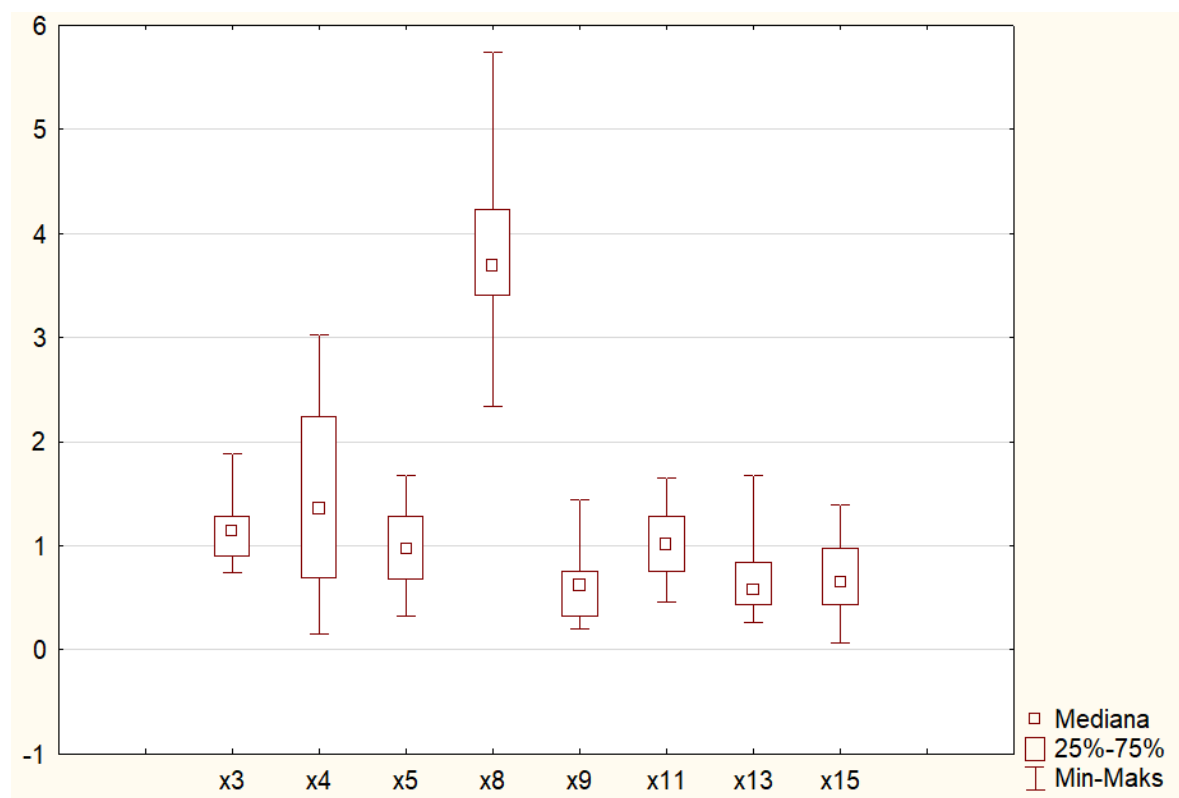
Mediana przeważnie zlokalizowana jest w okolicach średniej każdej ze zmiennych. Jeśli chodzi o długość wąsów, każdej zmiennej dolny i górny jest w podobnej długości. Dla zmiennych x3, x8, x9, x13, dłuższy jest górny wąs, a więc wartość minimalna wynosi zauważalnie mniej od pozostałych. Wykres zmiennej x4 jest prawie symetryczny, występuje w nim minimalna asymetria ujemna. Zmienna x9 oraz x13 wykazuje dominującą liczbę wyników poniżej średniej.

Wykres 1.6 Wykres ramka-wąsy dla zmiennych wybranych do porządkowania liniowego



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Wykres 1.7 Wykres ramka-wąsy dla zmiennych wybranych do analizy PCA oraz skupień



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

3. PORZĄDKOWANIE LINIOWE

Porządkowanie liniowe jest porządkowaniem ze względu na 1 cechę, ale porządkowane obiekty są wielowymiarowe. Pozwala ono ponumerować wszystkie obiekty od 1 do n. liniowe porządkowanie zbioru n elementowego pozwala na uszeregowaniu ich w kolejności od „najlepszego” – numer 1, do „najgorszego” – numer n.

Chcąc zawrzeć wiele wymiarów, należy dla grupy cech charakteryzujących obiekty znaleźć sumaryczną cechę, którą można nazwać funkcją porządkującą. Jest to zwykle jakiś syntetyczny wskaźnik, będący funkcją wartości zmiennych zespołu cech.

Ze względu na to, że podstawą porządkowania liniowego jest zespół zmiennych, to pomimo faktu, że są one przekształcane w jeden wskaźnik syntetyczny, to można to porządkowanie umiejscowić wśród metod analizy wielowymiarowej.

W pierwszym etapie przedstawiona jest macierz zmiennych wybranych do badania, wraz z obliczoną średnią arytmetyczną, odchyleniem standardowym oraz współczynnikiem zmienności dla każdej zmiennej. Na podstawie obliczeń oraz analizy wstępnej z poprzedniego rozdziału nie ma podstaw do odrzucenia innych zmiennych.

Tabela 3.1 Macierz zmiennych wybranych do badania

	D	S	S	S	D	D	S	S	D	S	D	S	S	D
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
dolnośląskie	1,87	1,01	1,47	0,88	4,04	0,86	4,39	0,89	2,67	1,16	1,23	0,69	0,39	0,68
kujawsko-pomorskie	1,28	0,89	0,88	1,19	3,32	1,11	2,48	0,66	2,94	0,63	0,33	0,27	1,96	0,71
lubelskie	1,10	1,65	1,34	0,68	2,34	1,28	3,80	0,20	1,77	1,06	0,42	0,52	0,23	0,33
lubuskie	0,54	1,25	2,10	1,38	3,80	0,99	3,12	0,28	4,17	1,47	0,66	0,33	0,52	0,46
łódzkie	1,43	1,10	0,22	0,98	3,33	0,73	3,67	0,69	4,48	0,71	1,13	0,81	0,99	1,23
małopolskie	2,59	0,92	0,88	0,64	3,17	0,41	3,52	0,83	1,25	0,81	1,78	1,21	0,36	0,90
mazowieckie	6,89	0,81	1,21	0,75	3,21	0,46	5,13	0,66	7,59	1,09	0,57	0,59	0,49	0,85
opolskie	1,41	1,10	0,15	0,67	3,39	1,83	4,76	0,96	1,59	0,54	0,32	0,56	0,53	1,17
podkarpackie	1,35	1,31	2,85	0,96	2,42	0,56	4,07	0,59	1,82	0,89	0,83	0,77	0,37	0,46
podlaskie	0,44	1,25	2,87	1,67	2,83	0,93	3,48	0,54	1,61	1,41	0,34	0,54	0,48	0,07
pomorskie	1,01	0,78	1,99	1,19	3,72	0,51	3,62	0,63	2,18	1,06	2,82	0,49	1,33	0,43
śląskie	2,94	0,74	0,51	0,32	3,76	0,42	5,74	1,44	4,14	0,46	1,77	1,68	0,77	1,39
świętokrzyskie	11,74	1,21	0,19	0,59	2,32	1,05	3,69	0,21	5,20	0,88	0,45	0,26	1,73	1,04
warmińsko-mazurskie	0,50	1,89	2,38	1,64	3,06	0,70	2,34	0,29	1,07	1,65	0,29	0,88	0,54	0,16
wielkopolskie	4,34	1,19	1,37	1,07	3,55	0,80	3,76	0,61	2,52	0,97	0,43	0,97	0,81	0,62
zachodniopomorskie	4,19	1,64	3,03	1,56	3,91	0,77	3,33	0,37	2,24	1,65	0,72	0,38	1,92	0,45
średnia arytmetyczna	2,73	1,17	1,47	1,01	3,26	0,84	3,81	0,62	2,95	1,03	0,88	0,68	0,84	0,68
odchylenie standardowe	2,87	0,32	0,95	0,39	0,53	0,36	0,85	0,31	1,69	0,36	0,69	0,36	0,56	0,37
współczynnik zmienności [%]	105,39	27,39	64,87	38,66	16,38	42,74	22,44	50,59	57,39	34,84	78,15	52,82	66,97	54,61

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.1. Metoda wzorca rozwoju Hellwiga

Metoda wzorca polega na skonstruowaniu hipotetycznego wzorca, względem którego określa się odległości rzeczywistych punktów. Odległości są należycie przekształcone, tak aby metoda mogła obliczyć odległości obiektów w przestrzeni wielowymiarowej od konkretnego punktu, który jest odpowiednio zdefiniowany.

Na początku dokonano normalizacji zmiennych – standaryzacji:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

Gdzie:

x_{ij} – obserwacja j-tej zmiennej dla obiektu i

\bar{x}_j – średnia arytmetyczna obserwacji j-tej zmiennej

s_j – odchylenie standardowe obserwacji j-tej zmiennej

	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14
	D	S	S	S	D	D	S	S	D	S	D	S	S	D
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
dolnośląskie	-0,30	-0,50	0,01	-0,33	1,46	0,06	0,68	0,88	-0,17	0,37	0,51	0,02	-0,80	-0,01
kujawsko-pomorskie	-0,50	-0,88	-0,62	0,46	0,11	0,76	-1,55	0,14	-0,01	-1,11	-0,80	-1,15	2,00	0,07
lubelskie	-0,57	1,49	-0,13	-0,85	-1,72	1,23	-0,01	-1,33	-0,70	0,09	-0,67	-0,45	-1,08	-0,95
lubuskie	-0,76	0,25	0,67	0,95	1,01	0,42	-0,80	-1,08	0,72	1,24	-0,32	-0,98	-0,57	-0,60
łódzkie	-0,45	-0,22	-1,31	-0,08	0,13	-0,30	-0,16	0,24	0,90	-0,89	0,36	0,35	0,27	1,46
małopolskie	-0,05	-0,78	-0,62	-0,95	-0,17	-1,20	-0,34	0,69	-1,00	-0,61	1,31	1,45	-0,85	0,58
mazowieckie	1,45	-1,13	-0,27	-0,67	-0,09	-1,06	1,55	0,14	2,74	0,17	-0,45	-0,26	-0,62	0,44
opolskie	-0,46	-0,22	-1,38	-0,87	0,24	2,77	1,12	1,11	-0,80	-1,36	-0,81	-0,34	-0,55	1,30
podkarpackie	-0,48	0,43	1,46	-0,13	-1,57	-0,78	0,31	-0,08	-0,67	-0,38	-0,07	0,24	-0,83	-0,60
podlaskie	-0,80	0,25	1,48	1,69	-0,81	0,26	-0,38	-0,24	-0,79	1,07	-0,79	-0,40	-0,64	-1,64
pomorskie	-0,60	-1,22	0,55	0,46	0,86	-0,92	-0,22	0,05	-0,46	0,09	2,82	-0,54	0,87	-0,68
śląskie	0,07	-1,34	-1,00	-1,77	0,93	-1,17	2,26	2,65	0,70	-1,59	1,29	2,75	-0,12	1,89
świętokrzyskie	3,14	0,12	-1,34	-1,08	-1,76	0,59	-0,14	-1,30	1,33	-0,41	-0,63	-1,17	1,59	0,95
warmińsko-mazurskie	-0,77	2,24	0,96	1,61	-0,38	-0,39	-1,72	-1,05	-1,11	1,74	-0,86	0,54	-0,53	-1,40
wielkopolskie	0,56	0,06	-0,10	0,15	0,54	-0,11	-0,05	-0,02	-0,26	-0,16	-0,65	0,79	-0,05	-0,17
zachodniopomorskie	0,51	1,46	1,65	1,41	1,22	-0,19	-0,56	-0,79	-0,42	1,74	-0,23	-0,84	1,93	-0,63

Obliczono współrzędne wzorca:

$$z_{0j} = \begin{cases} \max_i \{z_{ij}\}, & \text{dla zmiennych stymulant} \\ \min_i \{z_{ij}\}, & \text{dla zmiennych destymulant} \end{cases}$$

Obliczono współrzędne antywzorca:

$$z_{0j} = \begin{cases} \min_i \{z_{ij}\}, & \text{dla zmiennych stymulant} \\ \max_i \{z_{ij}\}, & \text{dla zmiennych destymulant} \end{cases}$$

wzorzec	-0,80	2,24	1,65	1,69	-1,76	-1,20	2,26	2,65	-1,11	1,74	-0,86	2,75	2,00	-1,64
antywzorzec	3,14	-1,34	-1,38	-1,77	1,46	2,77	-1,72	-1,33	2,74	-1,59	2,82	-1,17	-1,08	1,89
(WZ-ANWZ) ²	15,47	12,85	9,19	11,94	10,37	15,72	15,84	15,85	14,81	11,05	13,51	15,43	9,49	12,47

Kolejno obliczono odległość między wzorcem a antywzorcem:

$$\text{Suma } (WZ-ANWZ)^2 = 183,99$$

$$d_0 = \sqrt{183,99} = 13,56$$

Następnie obliczono odległości euklidesowe między każdą obserwacją a wzorcem:

$$d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2}$$

	D	S	S	S	D	D	S	S	D	S	D	S	S	D			
	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	suma	d_{i0}	m_i
dolnośląskie	0,25	7,52	2,69	4,09	10,37	1,58	2,50	3,12	0,89	1,87	1,87	7,50	7,81	2,66	54,73	7,398	0,455
kujawsko-pomorskie	0,09	9,71	5,12	1,51	3,51	3,82	14,56	6,27	1,22	8,12	0,00	15,22	0,00	2,93	72,08	8,490	0,374
lubelskie	0,05	0,56	3,16	6,42	0,00	5,90	5,16	15,85	0,17	2,72	0,04	10,30	9,49	0,48	60,30	7,765	0,428
lubuskie	0,00	3,98	0,96	0,55	7,68	2,62	9,41	13,87	3,35	0,25	0,29	13,95	6,57	1,09	64,57	8,035	0,408
łódzkie	0,12	6,06	8,74	3,12	3,58	0,80	5,87	5,80	4,05	6,89	1,49	5,79	2,98	9,63	64,93	8,058	0,406
małopolskie	0,56	9,14	5,12	6,95	2,53	0,00	6,75	3,84	0,01	5,51	4,69	1,69	8,11	4,93	59,83	7,735	0,430
mazowieckie	5,04	11,33	3,67	5,55	2,78	0,02	0,51	6,27	14,81	2,45	0,17	9,09	6,85	4,36	72,88	8,537	0,371
opolskie	0,11	6,06	9,19	6,55	4,01	15,72	1,32	2,38	0,09	9,61	0,00	9,60	6,48	8,66	79,79	8,932	0,341
podkarpackie	0,10	3,27	0,04	3,30	0,04	0,18	3,82	7,45	0,20	4,51	0,62	6,34	8,01	1,09	38,95	6,241	0,540
podlaskie	0,00	3,98	0,03	0,00	0,91	2,11	7,00	8,35	0,10	0,45	0,01	9,95	6,94	0,00	39,82	6,311	0,535
pomorskie	0,04	11,97	1,20	1,51	6,87	0,08	6,16	6,76	0,43	2,72	13,51	10,84	1,26	0,93	64,27	8,017	0,409
śląskie	0,76	12,85	7,03	11,94	7,27	0,00	0,00	0,00	3,28	11,05	4,62	0,00	4,49	12,47	75,77	8,704	0,358
świętokrzyskie	15,47	4,49	8,93	7,64	0,00	3,19	5,76	15,60	5,94	4,63	0,05	15,43	0,17	6,74	94,04	9,697	0,285
warmińsko-mazurskie	0,00	0,00	0,47	0,01	1,92	0,66	15,84	13,64	0,00	0,00	0,00	4,90	6,39	0,06	43,87	6,624	0,512
wielkopolskie	1,84	4,76	3,05	2,36	5,30	1,19	5,37	7,10	0,73	3,61	0,04	3,86	4,19	2,17	45,57	6,751	0,502
zachodniopomorskie	1,70	0,61	0,00	0,08	8,86	1,01	7,96	11,80	0,48	0,00	0,39	12,93	0,01	1,03	46,87	6,846	0,495

Finalnie obliczono dla każdej jednostki tzw. miarę taksonomiczną m_i według wzoru:

$$m_i = 1 - \frac{d_{i0}}{d_0}$$

Tabela 3.2 Ranking województw metodą wzorca rozwoju Hellwiga

Województwa	m_i	RANKING
podkarpackie	0,540	1
podlaskie	0,535	2
warmińsko-mazurskie	0,512	3
wielkopolskie	0,502	4
zachodniopomorskie	0,495	5
dolnośląskie	0,455	6
małopolskie	0,430	7
lubelskie	0,428	8
pomorskie	0,409	9
lubuskie	0,408	10
łódzkie	0,406	11
kujawsko-pomorskie	0,374	12
mazowieckie	0,371	13
śląskie	0,358	14
opolskie	0,341	15
świętokrzyskie	0,285	16

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.2. Metoda wzorca rozwoju Hellwiga z wagami

Metoda wzorca z wagami opiera się na tym samym co zwykła metoda z dodatkiem wag każdej zmiennej. Jest ona obliczana poprzez podzielenie wartości współczynnika zmienności danej zmiennej przez sumę wszystkich wartości współczynników zmienności dla wszystkich cech, która wynosi 713,23% lub w wartościach liczbowych 7,13.

	D	S	S	S	D	D	S	S	D	S	D	S	S	D
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
dolnośląskie	1,87	1,01	1,47	0,88	4,04	0,86	4,39	0,89	2,67	1,16	1,23	0,69	0,39	0,68
kujawsko-pomorskie	1,28	0,89	0,88	1,19	3,32	1,11	2,48	0,66	2,94	0,63	0,33	0,27	1,96	0,71
lubelskie	1,10	1,65	1,34	0,68	2,34	1,28	3,80	0,20	1,77	1,06	0,42	0,52	0,23	0,33
lubuskie	0,54	1,25	2,10	1,38	3,80	0,99	3,12	0,28	4,17	1,47	0,66	0,33	0,52	0,46
łódzkie	1,43	1,10	0,22	0,98	3,33	0,73	3,67	0,69	4,48	0,71	1,13	0,81	0,99	1,23
małopolskie	2,59	0,92	0,88	0,64	3,17	0,41	3,52	0,83	1,25	0,81	1,78	1,21	0,36	0,90
mazowieckie	6,89	0,81	1,21	0,75	3,21	0,46	5,13	0,66	7,59	1,09	0,57	0,59	0,49	0,85
opolskie	1,41	1,10	0,15	0,67	3,39	1,83	4,76	0,96	1,59	0,54	0,32	0,56	0,53	1,17
podkarpackie	1,35	1,31	2,85	0,96	2,42	0,56	4,07	0,59	1,82	0,89	0,83	0,77	0,37	0,46
podlaskie	0,44	1,25	2,87	1,67	2,83	0,93	3,48	0,54	1,61	1,41	0,34	0,54	0,48	0,07
pomorskie	1,01	0,78	1,99	1,19	3,72	0,51	3,62	0,63	2,18	1,06	2,82	0,49	1,33	0,43
śląskie	2,94	0,74	0,51	0,32	3,76	0,42	5,74	1,44	4,14	0,46	1,77	1,68	0,77	1,39
świętokrzyskie	11,74	1,21	0,19	0,59	2,32	1,05	3,69	0,21	5,20	0,88	0,45	0,26	1,73	1,04
warmińsko-mazurskie	0,50	1,89	2,38	1,64	3,06	0,70	2,34	0,29	1,07	1,65	0,29	0,88	0,54	0,16
wielkopolskie	4,34	1,19	1,37	1,07	3,55	0,80	3,76	0,61	2,52	0,97	0,43	0,97	0,81	0,62
zachodniopomorskie	4,19	1,64	3,03	1,56	3,91	0,77	3,33	0,37	2,24	1,65	0,72	0,38	1,92	0,45
Wagi	0,15	0,04	0,09	0,05	0,02	0,06	0,03	0,07	0,08	0,05	0,11	0,07	0,09	0,08

Następnie dokonano normalizacji zmiennych tak jak w poprzedniej metodzie, ale dodatkowo pomnożono je przez wagi:

	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14
	D	S	S	S	D	D	S	S	D	S	D	S	S	D
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
dolnośląskie	-0,04	-0,02	0,00	-0,02	0,03	0,00	0,02	0,06	-0,01	0,02	0,06	0,00	-0,08	0,00
kujawsko-pomorskie	-0,07	-0,03	-0,06	0,02	0,00	0,05	-0,05	0,01	0,00	-0,05	-0,09	-0,08	0,19	0,01
lubelskie	-0,08	0,06	-0,01	-0,05	-0,04	0,07	0,00	-0,09	-0,06	0,00	-0,07	-0,03	-0,10	-0,07
lubuskie	-0,11	0,01	0,06	0,05	0,02	0,03	-0,03	-0,08	0,06	0,06	-0,04	-0,07	-0,05	-0,05
łódzkie	-0,07	-0,01	-0,12	0,00	0,00	-0,02	-0,01	0,02	0,07	-0,04	0,04	0,03	0,03	0,11
małopolskie	-0,01	-0,03	-0,06	-0,05	0,00	-0,07	-0,01	0,05	-0,08	-0,03	0,14	0,11	-0,08	0,04
mazowieckie	0,21	-0,04	-0,02	-0,04	0,00	-0,06	0,05	0,01	0,22	0,01	-0,05	-0,02	-0,06	0,03
opolskie	-0,07	-0,01	-0,13	-0,05	0,01	0,17	0,04	0,08	-0,06	-0,07	-0,09	-0,03	-0,05	0,10
podkarpackie	-0,07	0,02	0,13	-0,01	-0,04	-0,05	0,01	-0,01	-0,05	-0,02	-0,01	0,02	-0,08	-0,05
podlaskie	-0,12	0,01	0,13	0,09	-0,02	0,02	-0,01	-0,02	-0,06	0,05	-0,09	-0,03	-0,06	-0,13
pomorskie	-0,09	-0,05	0,05	0,02	0,02	-0,05	-0,01	0,00	-0,04	0,00	0,31	-0,04	0,08	-0,05
śląskie	0,01	-0,05	-0,09	-0,10	0,02	-0,07	0,07	0,19	0,06	-0,08	0,14	0,20	-0,01	0,14
świętokrzyskie	0,46	0,00	-0,12	-0,06	-0,04	0,04	0,00	-0,09	0,11	-0,02	-0,07	-0,09	0,15	0,07
warmińsko-mazurskie	-0,11	0,09	0,09	0,09	-0,01	-0,02	-0,05	-0,07	-0,09	0,08	-0,09	0,04	-0,05	-0,11
wielkopolskie	0,08	0,00	-0,01	0,01	0,01	-0,01	0,00	0,00	-0,02	-0,01	-0,07	0,06	0,00	-0,01
zachodniopomorskie	0,08	0,06	0,15	0,08	0,03	-0,01	-0,02	-0,06	-0,03	0,08	-0,03	-0,06	0,18	-0,05

Obliczono współrzędne wzorca i antywzorca, zgodnie z wzorami podanymi w poprzedniej metodzie:

wzorzec	-0,12	0,09	0,15	0,09	-0,04	-0,07	0,07	0,19	-0,09	0,08	-0,09	0,20	0,19	-0,13
antywzorzec	0,46	-0,05	-0,13	-0,10	0,03	0,17	-0,05	-0,09	0,22	-0,08	0,31	-0,09	-0,10	0,14
(WZ-ANWZ) ²	0,34	0,02	0,08	0,04	0,01	0,06	0,02	0,08	0,10	0,03	0,16	0,08	0,08	0,07

Kolejno obliczono odległość między wzorcem a antywzorcem:

$$\text{Suma } (WZ-ANWZ)^2 = 1,15$$

$$d_0 = \sqrt{1,15} = 1,07$$

Kolejno obliczono odległości euklidesowe między każdą obserwacją a wzorcem, tak jak w poprzedniej metodzie:

	D	S	S	S	D	D	S	S	D	S	D	S	S	D		
	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	suma	d ₀
dolnośląskie	0,005	0,011	0,022	0,012	0,005	0,006	0,002	0,016	0,006	0,004	0,022	0,041	0,069	0,016	0,24	0,488
kujawsko-pomorskie	0,002	0,014	0,042	0,004	0,002	0,014	0,014	0,032	0,008	0,019	0,000	0,083	0,000	0,017	0,25	0,502
lubelskie	0,001	0,001	0,026	0,019	0,000	0,021	0,005	0,080	0,001	0,006	0,000	0,056	0,084	0,003	0,30	0,551
lubuskie	0,000	0,006	0,008	0,002	0,004	0,009	0,009	0,070	0,022	0,001	0,003	0,076	0,058	0,006	0,27	0,524
łódzkie	0,003	0,009	0,072	0,009	0,002	0,003	0,006	0,029	0,026	0,016	0,018	0,032	0,026	0,056	0,31	0,555
małopolskie	0,012	0,013	0,042	0,020	0,001	0,000	0,007	0,019	0,000	0,013	0,056	0,009	0,072	0,029	0,30	0,543
mazowieckie	0,110	0,017	0,030	0,016	0,001	0,000	0,001	0,032	0,096	0,006	0,002	0,050	0,060	0,026	0,45	0,668
opolskie	0,002	0,009	0,076	0,019	0,002	0,056	0,001	0,012	0,001	0,023	0,000	0,053	0,057	0,051	0,36	0,602
podkarpackie	0,002	0,005	0,000	0,010	0,000	0,001	0,004	0,037	0,001	0,011	0,007	0,035	0,071	0,006	0,19	0,436
podlaskie	0,000	0,006	0,000	0,000	0,000	0,008	0,007	0,042	0,001	0,001	0,000	0,055	0,061	0,000	0,18	0,425
pomorskie	0,001	0,018	0,010	0,004	0,004	0,000	0,006	0,034	0,003	0,006	0,162	0,059	0,011	0,005	0,32	0,570
śląskie	0,017	0,019	0,058	0,035	0,004	0,000	0,000	0,000	0,021	0,026	0,056	0,000	0,040	0,073	0,35	0,590
świętokrzyskie	0,338	0,007	0,074	0,022	0,000	0,011	0,006	0,078	0,038	0,011	0,001	0,085	0,001	0,039	0,71	0,844
warmińsko-mazurskie	0,000	0,000	0,004	0,000	0,001	0,002	0,016	0,069	0,000	0,000	0,000	0,027	0,056	0,000	0,18	0,418
wielkopolskie	0,040	0,007	0,025	0,007	0,003	0,004	0,005	0,036	0,005	0,009	0,000	0,021	0,037	0,013	0,21	0,461
zachodniopomorskie	0,037	0,001	0,000	0,000	0,005	0,004	0,008	0,059	0,003	0,000	0,005	0,071	0,000	0,006	0,20	0,446

Finalnie obliczono dla każdej jednostki tzw. miarę taksonomiczną m_i zgodnie z wzorem podanym w poprzedniej metodzie:

Tabela 3.3 Ranking województw metodą wzorca rozwoju Hellwiga z wagami

Województwa	m_i		Województwa	m_i	RANKING
dolnośląskie	0,545		warmińsko-mazurskie	0,610	1
kujawsko-pomorskie	0,532		podlaskie	0,604	2
lubelskie	0,486		podkarpackie	0,594	3
lubuskie	0,512		zachodniopomorskie	0,584	4
łódzkie	0,483		wielkopolskie	0,571	5
małopolskie	0,494		dolnośląskie	0,545	6
mazowieckie	0,377		kujawsko-pomorskie	0,532	7
opolskie	0,439		lubuskie	0,512	8
podkarpackie	0,594		małopolskie	0,494	9
podlaskie	0,604		lubelskie	0,486	10
pomorskie	0,469		łódzkie	0,483	11
śląskie	0,450		pomorskie	0,469	12
świętokrzyskie	0,213		śląskie	0,450	13
warmińsko-mazurskie	0,610		opolskie	0,439	14
wielkopolskie	0,571		mazowieckie	0,377	15
zachodniopomorskie	0,584		świętokrzyskie	0,213	16

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.3. Metoda TOPSIS

Metoda porządkowania liniowego TOPSIS jest to metoda wzorcowa, w której wyznacza się dwa punkty odniesienia obiektów w przestrzeni wielowymiarowej – wzorzec i antywzorzec.

Na początku dokonano normalizacji zmiennych – standaryzacji:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$$

Gdzie:

x_{ij} – obserwacja j-tej zmiennej dla obiektu i

Tabela ze zmiennymi ustandaryzowanymi jest przedstawiona poniżej:

	D	S	S	S	D	D	S	S	D	S	D	S	S	D
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
dolnośląskie	0,118	0,208	0,210	0,203	0,306	0,236	0,281	0,323	0,196	0,267	0,275	0,223	0,097	0,218
kujawsko-pomorskie	0,081	0,183	0,126	0,275	0,251	0,304	0,159	0,239	0,216	0,145	0,074	0,087	0,485	0,228
lubelskie	0,069	0,340	0,192	0,157	0,177	0,351	0,244	0,072	0,130	0,244	0,094	0,168	0,057	0,106
lubuskie	0,034	0,257	0,301	0,318	0,288	0,272	0,200	0,101	0,306	0,338	0,148	0,107	0,129	0,147
łódzkie	0,090	0,226	0,031	0,226	0,252	0,200	0,235	0,250	0,329	0,163	0,253	0,262	0,245	0,394
małopolskie	0,163	0,189	0,126	0,148	0,240	0,112	0,226	0,301	0,092	0,186	0,398	0,391	0,089	0,289
mazowieckie	0,435	0,167	0,173	0,173	0,243	0,126	0,329	0,239	0,557	0,250	0,128	0,191	0,121	0,273
opolskie	0,089	0,226	0,021	0,155	0,257	0,502	0,305	0,348	0,117	0,124	0,072	0,181	0,131	0,375
podkarpackie	0,085	0,270	0,408	0,222	0,183	0,154	0,261	0,214	0,134	0,204	0,186	0,249	0,092	0,147
podlaskie	0,028	0,257	0,411	0,385	0,214	0,255	0,223	0,196	0,118	0,324	0,076	0,174	0,119	0,022
pomorskie	0,064	0,161	0,285	0,275	0,281	0,140	0,232	0,228	0,160	0,244	0,631	0,158	0,329	0,138
śląskie	0,186	0,152	0,073	0,074	0,284	0,115	0,368	0,522	0,304	0,106	0,396	0,543	0,191	0,446
świętokrzyskie	0,741	0,249	0,027	0,136	0,176	0,288	0,236	0,076	0,382	0,202	0,101	0,084	0,428	0,333
warmińsko-mazurskie	0,032	0,389	0,341	0,378	0,232	0,192	0,150	0,105	0,079	0,379	0,065	0,284	0,134	0,051
wielkopolskie	0,274	0,245	0,196	0,247	0,269	0,219	0,241	0,221	0,185	0,223	0,096	0,313	0,201	0,199
zachodniopomorskie	0,264	0,338	0,434	0,360	0,296	0,211	0,213	0,134	0,165	0,379	0,161	0,123	0,476	0,144

Obliczono współrzędne wzorca:

$$z_{0j}^{+} = \begin{cases} \max_i \{z_{ij}\}, & \text{dla zmiennych stymulant} \\ \min_i \{z_{ij}\}, & \text{dla zmiennych destymulant} \end{cases}$$

Obliczono współrzędne antywzorca:

$$z_{0j}^{-} = \begin{cases} \min_i \{z_{ij}\}, & \text{dla zmiennych stymulant} \\ \max_i \{z_{ij}\}, & \text{dla zmiennych destymulant} \end{cases}$$

wzorzec z_{0j}^{+}	0,03	0,39	0,43	0,39	0,18	0,11	0,37	0,52	0,08	0,38	0,06	0,54	0,49	0,02
antywzorzec z_{0j}^{-}	0,74	0,15	0,02	0,07	0,18	0,11	0,15	0,07	0,08	0,11	0,06	0,08	0,06	0,02
$(WZ-ANWZ)^2$	0,51	0,06	0,17	0,10	0,00	0,00	0,05	0,20	0,00	0,07	0,00	0,21	0,18	0,00

Kolejno obliczono odległość między wzorcem a antywzorcem:

$$\text{Suma } (WZ-ANWZ)^2 = 1,55$$

$$d_0 = \sqrt{1,55} = 1,24$$

Następnie obliczono odległości euklidesowe między każdą obserwacją a wzorcem:

$$d_{i0+} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j+})^2}$$

oraz między każdą obserwacją a antywzorcem, zgodnie ze wzorem:

$$d_{i0-} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j-})^2}$$

Finalnie obliczono wartości zmiennej agregowanej q_i według wzoru:

$$q_i = \frac{d_{i0-}}{d_{0+} + d_{i0-}}$$

Poniżej przedstawiony jest ranking województw wykonany metodą TOPSIS. Ponownie tak jak w dwóch poprzednich metodach na pierwszych trzech miejscach znajduje się województwo warmińsko-mazurskie oraz podlaskie. Województwo z najgorszym stanem środowiska naturalnego jest województwo świętokrzyskie, który trzeci raz z rzędu zajmuje ostatnie miejsce.

Tabela 3.4 Ranking województw metodą TOPSIS

Województwa	q_i	RANKING
warmińsko-mazurskie	0,589	1
podlaskie	0,584	2
zachodniopomorskie	0,571	3
podkarpackie	0,556	4
pomorskie	0,544	5
śląskie	0,541	6
kujawsko-pomorskie	0,528	7
dolnośląskie	0,526	8
lubuskie	0,525	9
łódzkie	0,511	10
małopolskie	0,503	11
wielkopolskie	0,495	12
opolskie	0,492	13
lubelskie	0,478	14
mazowieckie	0,422	15
świętokrzyskie	0,342	16

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.4. Metoda standaryzowanych sum

Metoda standaryzowanych sum polega na wyznaczeniu liniowej funkcji porządkującej w_i – będącą średnią arytmetyczną standaryzowanych wartości zmiennych. Jednostki lokuje się od najwyższej wartości do najmniejszej – im wyższa wartość tym województwo odznacza się lepszym stanem środowiska naturalnego.

W tej metodzie należy dokonać normalizacji zmiennych – standaryzacji. Dla zmiennych pozytywnych oblicza się wartość znormalizowaną według pierwszego wzoru, natomiast dla negatywnych z drugiego:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad z_{ij} = \frac{\bar{x}_j - x_{ij}}{s_j}$$

Poniżej przedstawiona jest tabela z obliczonymi ustandaryzowanymi wartościami

	D	S	S	S	D	D	S	S	D	S	D	S	S	D		
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	Suma	w_i
dolnośląskie	0,298	-0,503	0,005	-0,334	-1,459	-0,061	0,683	0,881	0,167	0,370	-0,508	0,016	-0,799	0,012	-1,232	-0,088
kujawsko-pomorskie	0,503	-0,877	-0,616	0,459	-0,111	-0,759	-1,553	0,142	0,007	-1,110	0,800	-1,146	1,996	-0,069	-2,332	-0,167
lubelskie	0,566	1,492	-0,132	-0,846	1,724	-1,234	-0,007	-1,335	0,698	0,091	0,669	-0,455	-1,084	0,948	1,096	0,078
lubuskie	0,761	0,245	0,668	0,945	-1,010	-0,424	-0,803	-1,078	-0,719	1,236	0,321	-0,980	-0,567	0,600	-0,804	-0,057
łódzkie	0,451	-0,222	-1,310	-0,078	-0,130	0,302	-0,159	0,239	-0,901	-0,887	-0,362	0,348	0,269	-1,460	-3,902	-0,279
małopolskie	0,047	-0,783	-0,616	-0,949	0,170	1,195	-0,335	0,688	1,005	-0,608	-1,307	1,454	-0,852	-0,577	-1,467	-0,105
mazowieckie	-1,449	-1,126	-0,268	-0,667	0,095	1,056	1,550	0,142	-2,737	0,175	0,451	-0,261	-0,621	-0,443	-4,104	-0,293
opolskie	0,458	-0,222	-1,384	-0,872	-0,242	-2,769	1,116	1,106	0,804	-1,362	0,815	-0,344	-0,550	-1,299	-4,745	-0,339
podkarpackie	0,479	0,432	1,457	-0,130	1,574	0,776	0,309	-0,082	0,668	-0,384	0,074	0,237	-0,835	0,600	5,177	0,370
podlaskie	0,796	0,245	1,479	1,688	0,806	-0,256	-0,382	-0,243	0,792	1,068	0,786	-0,399	-0,639	1,644	7,384	0,527
pomorskie	0,597	-1,219	0,552	0,459	-0,860	0,916	-0,218	0,046	0,456	0,091	-2,818	-0,538	0,875	0,681	-0,980	-0,070
śląskie	-0,074	-1,344	-1,005	-1,768	-0,935	1,167	2,264	2,647	-0,701	-1,585	-1,292	2,754	-0,122	-1,888	-1,882	-0,134
świętokrzyskie	-3,137	0,121	-1,342	-1,077	1,761	-0,592	-0,136	-1,302	-1,326	-0,412	0,626	-1,174	1,587	-0,952	-7,355	-0,525
warmińsko-mazurskie	0,775	2,240	0,963	1,611	0,376	0,386	-1,716	-1,046	1,111	1,739	0,858	0,541	-0,532	1,403	8,709	0,622
wielkopolskie	-0,562	0,058	-0,100	0,152	-0,542	0,106	-0,054	-0,018	0,255	-0,161	0,655	0,790	-0,051	0,172	0,702	0,050
zachodniopomorskie	-0,509	1,461	1,647	1,406	-1,216	0,190	-0,558	-0,789	0,420	1,739	0,233	-0,842	1,925	0,627	5,736	0,410

Tabela 3.5 Ranking województw metodą standaryzowanych sum

Województwa	w_i	RANKING
warmińsko-mazurskie	0,622	1
podlaskie	0,527	2
zachodniopomorskie	0,410	3
podkarpackie	0,370	4
lubelskie	0,078	5
wielkopolskie	0,050	6
lubuskie	-0,057	7
pomorskie	-0,070	8
dolnośląskie	-0,088	9
małopolskie	-0,105	10
śląskie	-0,134	11
kujawsko-pomorskie	-0,167	12
łódzkie	-0,279	13
mazowieckie	-0,293	14
opolskie	-0,339	15
świętokrzyskie	-0,525	16

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.5. Metoda sumy rang

Metoda sumy rang polega na ustaleniu rangi dla zmiennych, od 1 do 16 w przypadku badania województw. Dla stymulant o najwyższych wartościach przypisuje się najwyższe rangi, natomiast dla destymulant najwyższą rangę otrzymują wartości najniższe. W ten sposób powstaje tabela z przypisanymi rangami poniżej:

	D	S	S	S	D	D	S	S	D	S	D	S	S	D	
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	Suma rang
dolnośląskie	10	11	7	10	16	10	4	3	10	5	13	7	13	9	128
kujawsko-pomorskie	6	13	11,5	5,5	8	14	15	6,5	11	14	3	15	1	10	134
lubelskie	5	2	9	12	2	15	6	16	5	7,5	5	11	16	3	115
lubuskie	3	5,5	5	4	14	12	14	14	13	3	9	14	10	6,5	127
łódzkie	9	9,5	14	8	9	7	9	5	14	13	12	5	5	15	135
małopolskie	11	12	11,5	14	6	1	11	4	2	12	15	2	15	12	129
mazowieckie	15	14	10	11	7	3	2	6,5	16	6	8	8	11	11	129
opolskie	8	9,5	16	13	10	16	3	2	3	15	2	9	9	14	130
podkarpackie	7	4	3	9	3	5	5	10	6	10	11	6	14	6,5	100
podlaskie	1	5,5	2	1	4	11	12	11	4	4	4	10	12	1	83
pomorskie	4	15	6	5,5	12	4	10	8	7	7,5	16	12	4	4	115
śląskie	12	16	13	16	13	2	1	1	12	16	14	1	7	16	140
świętokrzyskie	16	7	15	15	1	13	8	15	15	11	7	16	3	13	155
warmińsko-mazurskie	2	1	4	2	5	6	16	13	1	1,5	1	4	8	2	67
wielkopolskie	14	8	8	7	11	9	7	9	9	9	6	3	6	8	114
zachodniopomorskie	13	3	1	3	15	8	13	12	8	1,5	10	13	2	5	108

Przydzielone rangi zostały zsumowane, a następnie posegregowane od najmniejszej wartości do największej. Im mniejsza suma rang tym wyższa pozycja w rankingu.

Tabela 3.6 Ranking województw metodą sumy rang

Województwa	Suma rang	RANKING
warmińsko-mazurskie	66,5	1
podlaskie	82,5	2
podkarpackie	99,5	3
zachodniopomorskie	107,5	4
wielkopolskie	114	5
lubelskie	114,5	6
pomorskie	115	7
lubuskie	127	8
dolnośląskie	128	9
małopolskie	128,5	10
mazowieckie	128,5	11
opolskie	129,5	12
kujawsko-pomorskie	133,5	13
łódzkie	134,5	14
śląskie	140	15
świętokrzyskie	155	16

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.6. Metoda dystansów

Wśród metod porządkowania liniowego można także wyróżnić metodę opartą na dystansach. Jest ona dosyć podobna do metody wzorca. Polega ona na obliczeniu odległości rzeczywistych jednostek w stosunku do obiektu najlepszego pod względem danej cechy, przy czym zależnie od charakteru zmiennej będzie to obiekt o najwyższej lub najniższej wartości.

Poniżej przedstawiona jest macierz danych wraz z obliczonymi wartościami max i min dla danej zmiennej.

	D	S	S	S	D	D	S	S	D	S	D	S	S	D
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
dolnośląskie	1,87	1,01	1,47	0,88	4,04	0,86	4,39	0,89	2,67	1,16	1,23	0,69	0,39	0,68
kujawsko-pomorskie	1,28	0,89	0,88	1,19	3,32	1,11	2,48	0,66	2,94	0,63	0,33	0,27	1,96	0,71
lubelskie	1,10	1,65	1,34	0,68	2,34	1,28	3,80	0,20	1,77	1,06	0,42	0,52	0,23	0,33
lubuskie	0,54	1,25	2,10	1,38	3,80	0,99	3,12	0,28	4,17	1,47	0,66	0,33	0,52	0,46
łódzkie	1,43	1,10	0,22	0,98	3,33	0,73	3,67	0,69	4,48	0,71	1,13	0,81	0,99	1,23
małopolskie	2,59	0,92	0,88	0,64	3,17	0,41	3,52	0,83	1,25	0,81	1,78	1,21	0,36	0,90
mazowieckie	6,89	0,81	1,21	0,75	3,21	0,46	5,13	0,66	7,59	1,09	0,57	0,59	0,49	0,85
opolskie	1,41	1,10	0,15	0,67	3,39	1,83	4,76	0,96	1,59	0,54	0,32	0,56	0,53	1,17
podkarpackie	1,35	1,31	2,85	0,96	2,42	0,56	4,07	0,59	1,82	0,89	0,83	0,77	0,37	0,46
podlaskie	0,44	1,25	2,87	1,67	2,83	0,93	3,48	0,54	1,61	1,41	0,34	0,54	0,48	0,07
pomorskie	1,01	0,78	1,99	1,19	3,72	0,51	3,62	0,63	2,18	1,06	2,82	0,49	1,33	0,43
śląskie	2,94	0,74	0,51	0,32	3,76	0,42	5,74	1,44	4,14	0,46	1,77	1,68	0,77	1,39
świętokrzyskie	11,74	1,21	0,19	0,59	2,32	1,05	3,69	0,21	5,20	0,88	0,45	0,26	1,73	1,04
warmińsko-mazurskie	0,50	1,89	2,38	1,64	3,06	0,70	2,34	0,29	1,07	1,65	0,29	0,88	0,54	0,16
wielkopolskie	4,34	1,19	1,37	1,07	3,55	0,80	3,76	0,61	2,52	0,97	0,43	0,97	0,81	0,62
zachodniopomorskie	4,19	1,64	3,03	1,56	3,91	0,77	3,33	0,37	2,24	1,65	0,72	0,38	1,92	0,45
MAX	11,74	1,89	3,03	1,67	4,04	1,83	5,74	1,44	7,59	1,65	2,82	1,68	1,96	1,39
MIN	0,44	0,74	0,15	0,32	2,32	0,41	2,34	0,20	1,07	0,46	0,29	0,26	0,23	0,07

W przypadku stymulant należy podzielić wartości cechy jednostek przez najkorzystniejszą zaobserwowaną wartość cechy wśród obiektów, czyli przez wartość **największą** zgodnie ze wzorem:

$$w_{ij} = \frac{x_{ij}}{\max_i(x_{ij})}$$

W przypadku destymulant należy podzielić wartości cechy jednostek przez najkorzystniejszą zaobserwowaną wartość cechy wśród obiektów, czyli przez wartość **najmniejszą** zgodnie ze wzorem:

$$w_{ij} = \frac{\min_i(x_{ij})}{x_{ij}}$$

Sumaryczny wskaźnik uzyskuje się poprzez uśrednienie mierników dystansów zgodnie ze wzorem:

$$W_i = \frac{1}{p} \sum_{j=1}^p w_{ij} \text{ dla } p=14 \text{ w tym przypadku}$$

Poniżej znajduje się tabela wraz z obliczonymi wartościami w_{ij} , ich summi oraz obliczoną wartością W_i

	D	S	S	S	D	D	S	S	D	S	D	S	S	D		
	x1	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	Suma	W_i
dolnośląskie	0,24	0,53	0,49	0,53	0,57	0,48	0,76	0,62	0,40	0,70	0,24	0,41	0,20	0,10	6,27	0,448
kujawsko-pomorskie	0,34	0,47	0,29	0,71	0,70	0,37	0,43	0,46	0,36	0,38	0,88	0,16	1,00	0,10	6,66	0,476
lubelskie	0,40	0,87	0,44	0,41	0,99	0,32	0,66	0,14	0,60	0,64	0,69	0,31	0,12	0,21	6,81	0,487
lubuskie	0,81	0,66	0,69	0,83	0,61	0,41	0,54	0,19	0,26	0,89	0,44	0,20	0,27	0,15	6,96	0,497
łódzkie	0,31	0,58	0,07	0,59	0,70	0,56	0,64	0,48	0,24	0,43	0,26	0,48	0,51	0,06	5,90	0,421
małopolskie	0,17	0,49	0,29	0,38	0,73	1,00	0,61	0,58	0,86	0,49	0,16	0,72	0,18	0,08	6,74	0,482
mazowieckie	0,06	0,43	0,40	0,45	0,72	0,89	0,89	0,46	0,14	0,66	0,51	0,35	0,25	0,08	6,30	0,450
opolskie	0,31	0,58	0,05	0,40	0,68	0,22	0,83	0,67	0,67	0,33	0,91	0,33	0,27	0,06	6,32	0,451
podkarpackie	0,33	0,69	0,94	0,57	0,96	0,73	0,71	0,41	0,59	0,54	0,35	0,46	0,19	0,15	7,62	0,544
podlaskie	1,00	0,66	0,95	1,00	0,82	0,44	0,61	0,38	0,66	0,85	0,85	0,32	0,24	1,00	9,79	0,699
pomorskie	0,44	0,41	0,66	0,71	0,62	0,80	0,63	0,44	0,49	0,64	0,10	0,29	0,68	0,16	7,08	0,506
śląskie	0,15	0,39	0,17	0,19	0,62	0,98	1,00	1,00	0,26	0,28	0,16	1,00	0,39	0,05	6,64	0,474
świętokrzyskie	0,04	0,64	0,06	0,35	1,00	0,39	0,64	0,15	0,21	0,53	0,64	0,15	0,88	0,07	5,76	0,412
warmińsko-mazurskie	0,88	1,00	0,79	0,98	0,76	0,59	0,41	0,20	1,00	1,00	1,00	0,52	0,28	0,44	9,84	0,703
wielkopolskie	0,10	0,63	0,45	0,64	0,65	0,51	0,66	0,42	0,42	0,59	0,67	0,58	0,41	0,11	6,86	0,490
zachodniopomorskie	0,11	0,87	1,00	0,93	0,59	0,53	0,58	0,26	0,48	1,00	0,40	0,23	0,98	0,16	8,11	0,579

Finalnie uporządkowano obliczone wartości W_i i otrzymano następujący ranking województw:

Tabela 3.7 Ranking województw metodą dystansów

Województwa	W_i	RANKING
warmińsko-mazurskie	0,703	1
podlaskie	0,699	2
zachodniopomorskie	0,579	3
podkarpackie	0,544	4
pomorskie	0,506	5
lubuskie	0,497	6
wielkopolskie	0,490	7
lubelskie	0,487	8
małopolskie	0,482	9
kujawsko-pomorskie	0,476	10
śląskie	0,474	11
opolskie	0,451	12
mazowieckie	0,450	13
dolnośląskie	0,448	14
łódzkie	0,421	15
świętokrzyskie	0,412	16

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

3.7. Podsumowanie metod porządkowania liniowego

W poniższej tabeli zostały zaprezentowane wyniki wszystkich metod porządkowania liniowego. Województwa zostały uszeregowane od tych, które uzyskały najmniej punktów, do tych które uzyskały ich najwięcej. W tym przypadku najmniejsza liczba punktów oznacza województwo, które zajęło najlepszą pozycję względem stanu środowiska. Kolorem zielonym (od najjaśniejszego do ciemnego) zostały zaznaczone województwa, które zajęły w poszczególnych metodach pierwsze 3 miejsca, natomiast kolorem żółtym, pomarańczowym i czerwonym trzy ostatnie miejsca (14,15 i 16).

Sumarycznie według wszystkich metod porządkowania liniowego, województwem dysponującym najlepszym stanem środowiska naturalnego jest województwo warmińsko-mazurskie, a następnie podlaskie i podkarpackie. Najgorsze okazały się województwo łódzkie, ex aequo mazowieckie i opolskie oraz świętokrzyskie. Można zauważyć, że wyniki są dosyć podobne pomiędzy metodami. Województwo warmińsko-mazurskie 4 na 5 razy zajęło pierwsze miejsce, a świętokrzyskie za każdym razem ostatnie.

Tabela 3.8 Porównanie metod porządkowania liniowego

	M. wzorca rozwoju Hellwiga	M. wzorca z wagami	M. TOPSIS	M. standaryzowanych sum	M. sumy rang	M. dystansów	SUMA PUNKTÓW
warmińsko-mazurskie	3	1	1	1	1	1	8
podlaskie	2	2	2	2	2	2	12
podkarpackie	1	3	4	4	3	4	19
zachodniopomorskie	5	4	3	3	4	3	22
wielkopolskie	4	5	12	6	5	7	39
pomorskie	9	12	5	8	7	5	46
lubuskie	10	8	9	7	8	6	48
lubelskie	8	10	14	5	6	8	51
dolnośląskie	6	6	8	9	9	14	52
małopolskie	7	9	11	10	10	9	56
kujawsko-pomorskie	12	7	7	12	13	10	61
śląskie	14	13	6	11	15	11	70
łódzkie	11	11	10	13	14	15	74
mazowieckie	13	15	15	14	11	13	81
opolskie	15	14	13	15	12	12	81
świętokrzyskie	16	16	16	16	16	16	96

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

W poniższej tabeli została przedstawiona macierz korelacji pomiędzy metodami porządkowania liniowego. Współczynnik liniowej korelacji Pearsona ma wysokie wartości dla wszystkich metod, jednakże w metodzie TOPSIS są one najniższe, co można też zauważyć po zestawieniu wyników w poprzedniej tabeli 3.8.

Tabela 3.9 Macierz korelacji dla metod porządkowania liniowego

Zmienna	M. wzorca rozwoju Hellwiga	M. wzorca z wagami	M. TOPSIS	M. standaryzowanych sum	M. sumy rang	M. dystansów
M. wzorca rozwoju Hellwiga	1,00	0,91	0,63	0,89	0,91	0,76
M. wzorca z wagami	0,91	1,00	0,74	0,87	0,82	0,76
M. TOPSIS	0,63	0,74	1,00	0,71	0,60	0,74
M. standaryzowanych sum	0,89	0,87	0,71	1,00	0,94	0,91
M. sumy rang	0,91	0,82	0,60	0,94	1,00	0,89
M. dystansów	0,76	0,76	0,74	0,91	0,89	1,00

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

4. ANALIZA GŁÓWNYCH SKŁADOWYCH

Analiza głównych składowych PCA (ang. *principal components analysis*) polega na ortogonalnej transformacji układu badanych zmiennych X_j w zbiór nowych zmiennych Y_l , które są liniowymi kombinacjami badanych zmiennych, które można zapisać w postaci równania:

$$Y_1 = w_{11}X_1 + w_{21}X_2 + \dots + w_{p1}X_p$$

$$Y_2 = w_{12}X_1 + w_{22}X_2 + \dots + w_{p2}X_p$$

...

$$Y_m = w_{1m}X_1 + w_{2m}X_2 + \dots + w_{pm}X_p$$

Nowe zmienne Y_l to główne składowe zmiennych X_j . Podział przebiega w taki sposób, aby pierwsza składowa wyjaśniała możliwie jak największą część zmienności danych.

4.1. Analiza wyboru zmiennych – analiza korelacji zmiennych i wykresy rozrzutu

Poniżej, w Tabeli 4.1 przedstawiona została macierz korelacji wszystkich zmiennych. Przedstawiono wykresy rozrzutu dla zmiennych charakteryzujących się najwyższym współczynnikiem korelacji.

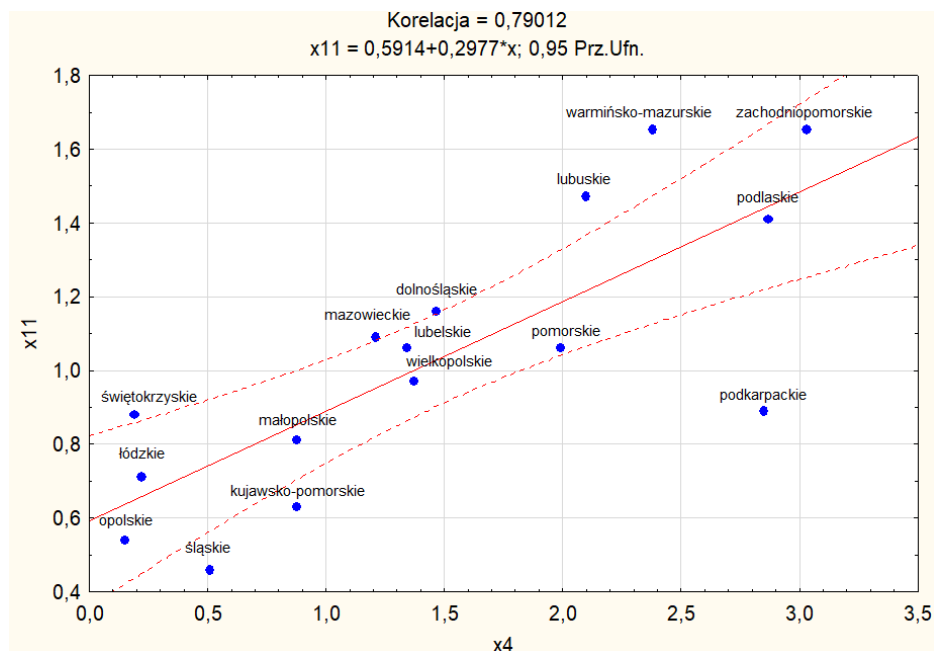
Tabela 4.1 Macierz korelacji wszystkich zmiennych

Zmienna	Korelacje (AW-ochrona-srodowiska)															
	Oznaczone wsp. korelacji są istotne z p < ,05000 N=16 (Braki danych usuwano przypadkami)															
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16
x1	1,00	0,13	-0,14	-0,36	-0,40	-0,25	-0,09	0,25	-0,16	0,61	-0,12	-0,15	-0,14	0,39	0,38	0,08
x2	0,13	1,00	-0,39	-0,36	-0,55	0,18	-0,36	0,57	0,68	0,21	-0,46	0,38	0,75	-0,02	0,57	0,97
x3	-0,14	-0,39	1,00	0,50	0,51	-0,34	0,26	-0,52	-0,70	-0,43	0,66	-0,55	-0,25	-0,08	-0,59	-0,46
x4	-0,36	-0,36	0,50	1,00	0,76	0,03	-0,27	-0,38	-0,41	-0,39	0,79	-0,06	-0,21	-0,09	-0,86	-0,39
x5	-0,40	-0,55	0,51	0,76	1,00	0,17	-0,01	-0,72	-0,52	-0,34	0,78	-0,25	-0,42	0,19	-0,77	-0,61
x6	-0,25	0,18	-0,34	0,03	0,17	1,00	-0,16	0,13	0,45	0,06	0,10	0,39	0,17	0,17	0,18	0,24
x7	-0,09	-0,36	0,26	-0,27	-0,01	-0,16	1,00	-0,10	-0,20	-0,22	-0,13	-0,57	-0,51	0,04	0,02	-0,31
x8	0,25	0,57	-0,52	-0,38	-0,72	0,13	-0,10	1,00	0,68	0,41	-0,50	0,25	0,47	-0,34	0,60	0,70
x9	-0,16	0,68	-0,70	-0,41	-0,52	0,45	-0,20	0,68	1,00	0,06	-0,65	0,45	0,70	-0,17	0,64	0,79
x10	0,61	0,21	-0,43	-0,39	-0,34	0,06	-0,22	0,41	0,06	1,00	-0,17	-0,05	-0,10	0,17	0,48	0,20
x11	-0,12	-0,46	0,66	0,79	0,78	0,10	-0,13	-0,50	-0,65	-0,17	1,00	-0,23	-0,37	-0,04	-0,80	-0,54
x12	-0,15	0,38	-0,55	-0,06	-0,25	0,39	-0,57	0,25	0,45	-0,05	-0,23	1,00	0,39	0,03	0,22	0,36
x13	-0,14	0,75	-0,25	-0,21	-0,42	0,17	-0,51	0,47	0,70	-0,10	-0,37	0,39	1,00	-0,41	0,40	0,80
x14	0,39	-0,02	-0,08	-0,09	0,19	0,17	0,04	-0,34	-0,17	0,17	-0,04	0,03	-0,41	1,00	0,14	-0,13
x15	0,38	0,57	-0,59	-0,86	-0,77	0,18	0,02	0,60	0,64	0,48	-0,80	0,22	0,40	0,14	1,00	0,62
x16	0,08	0,97	-0,46	-0,39	-0,61	0,24	-0,31	0,70	0,79	0,20	-0,54	0,36	0,80	-0,13	0,62	1,00

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Poniżej zaprezentowane są wykresy rozrzutu zmiennych wygenerowane w programie Statistica.

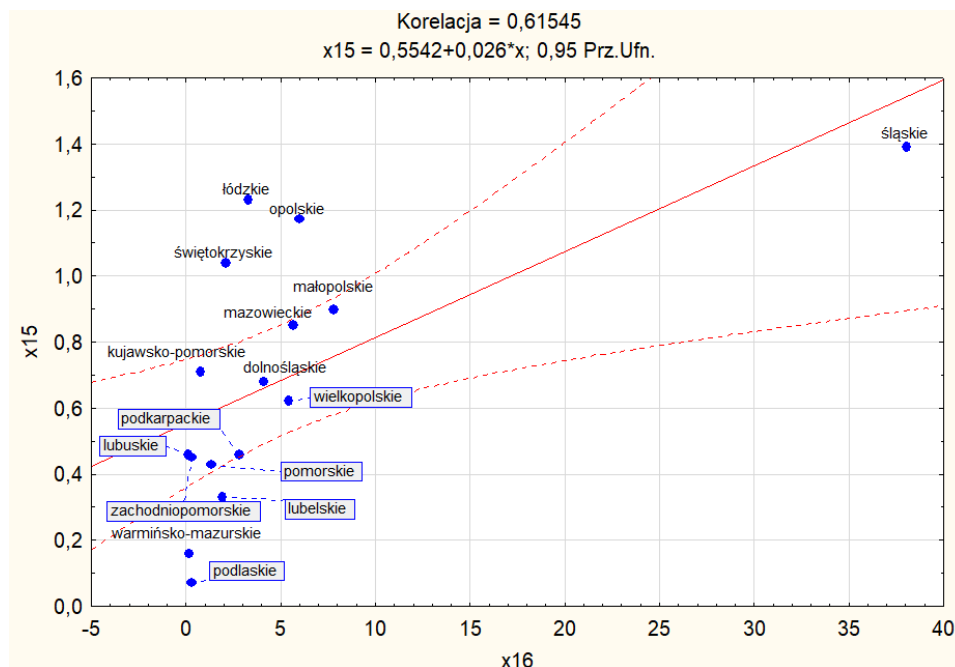
Wykres 4.1 Wykres rozrzutu zmiennej **x11** względem **x4**



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie 4.1 zaprezentowany jest wykres rozrzutu zmiennej **x11** względem **x4**. Współczynnik korelacji między zmiennymi wynosi 0,79012 i jest on dodatni. Oznacza to, że województwa dysponujące i posiadające większą powierzchnię gospodarstw ekologicznych charakteryzują się większą powierzchnią obszarów Natura 2000. Ogólnie rzecz ujmując województwa posiadające większą powierzchnię obszarów prawnie chronionych kładą również większy nacisk na ekologiczne rolnictwo.

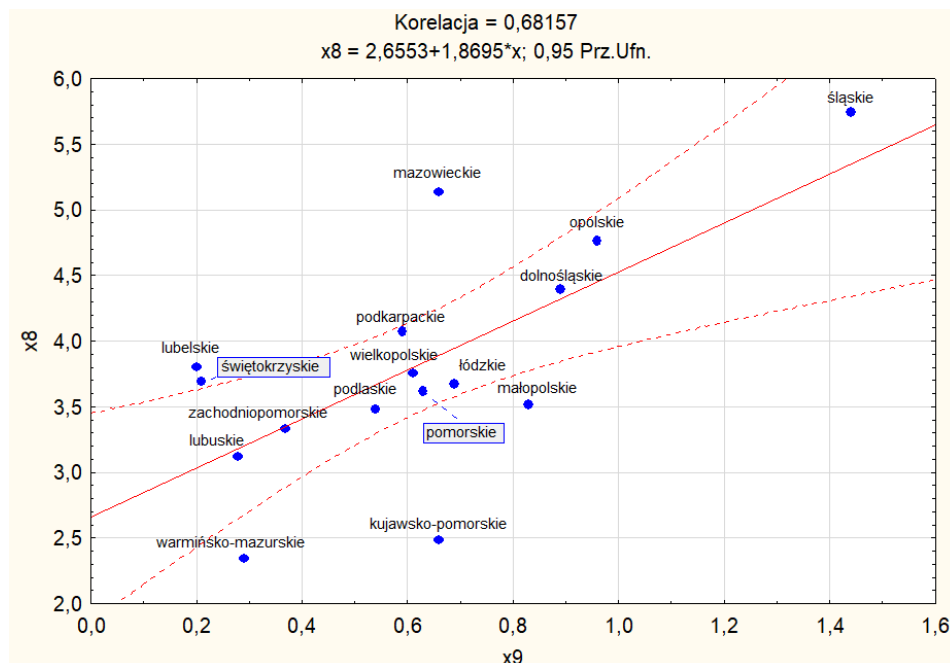
Wykres 4.2 Wykres rozrzutu zmiennej **x15** względem **x16**



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie 4.2 zaprezentowany jest wykres rozrzutu zmiennej **x15** względem **x16**. Współczynnik korelacji między zmiennymi wynosi 0,61545 i jest on dodatni. Oznacza to, że wraz ze wzrostem emisji tlenków azotu rośnie emisja metali ciężkich z zakładów szczególnie uciążliwych. Wszystkie województwa mniej więcej znajdują się obok siebie. Województwo śląskie zdecydowanie odstaje od pozostałych i cechuje się bardzo dużą emisją obydwu rodzaju substancji.

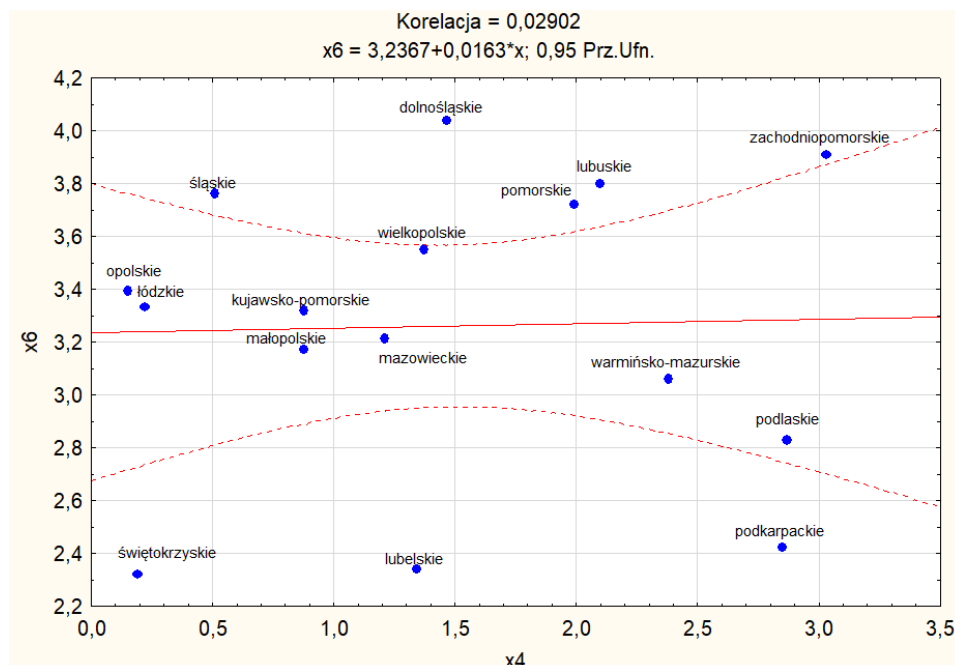
Wykres 4.3 Wykres rozrzutu zmiennej **x8** względem **x9**



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie 4.3 zaprezentowany jest wykres rozrzutu zmiennej **x8** względem **x9**. Współczynnik korelacji między zmiennymi wynosi 0,68157 i jest on dodatni. Oznacza to, że wraz ze wzrostem nakładów finansowych na środki trwałe służące ochronie środowiska i gospodarki wodnej rośnie ilość nasadzonych drzew w miastach i na wsiach. Od grupy województw zdecydowanie odbiega województwo śląskie, następnie mazowieckie oraz kujawsko-pomorskie i warmińsko-mazurskie.

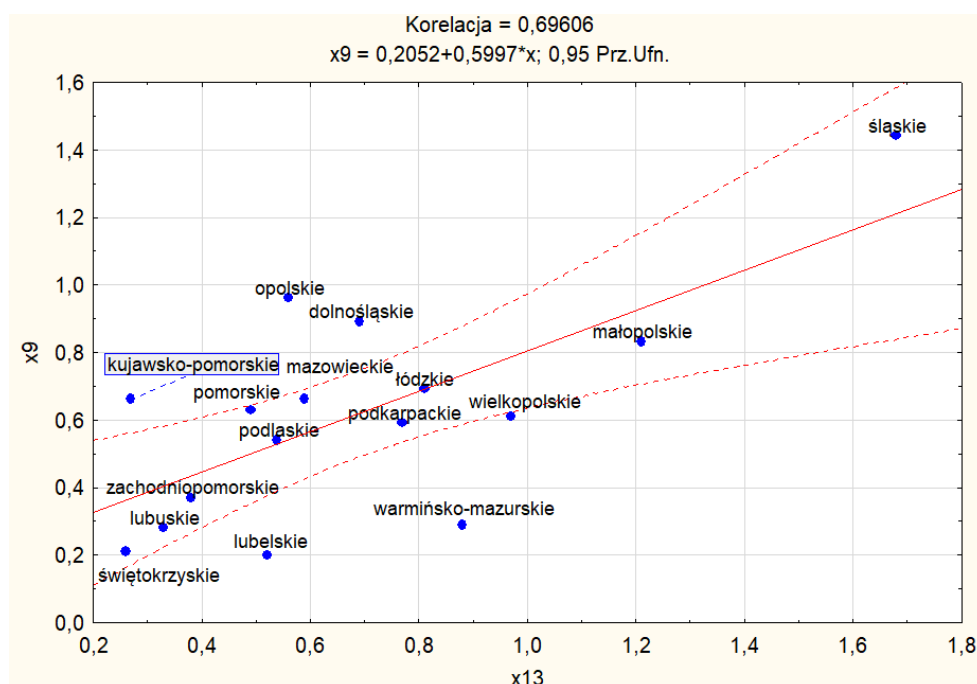
Wykres 4.4 Wykres rozrzutu zmiennej **x6** względem **x4**



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie 4.4 zaprezentowany jest wykres rozrzutu zmiennej **x6** względem **x4**. Współczynnik korelacji między zmiennymi wynosi 0,02902 i jest on dodatni. Jest to bardzo mała wartość i związek między zmiennymi jest liniowy. Oznacza to, że masa wytworzonych odpadów komunalnych przez mieszkańców w niemalże całkowitym stopniu nie wpływa na powierzchnię obszarów Natura 2000.

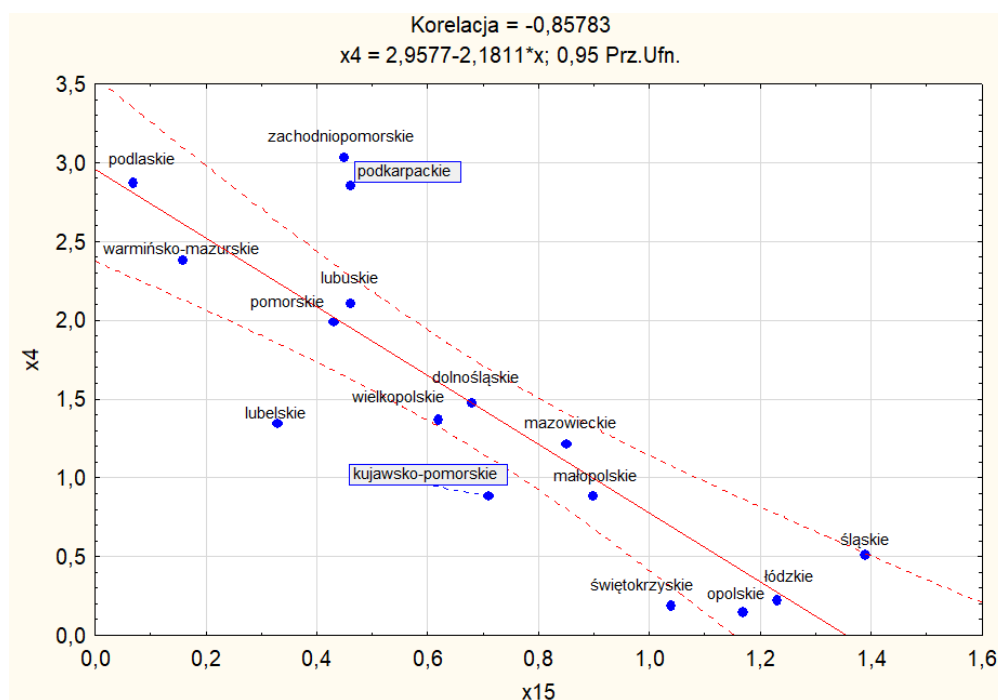
Wykres 4.5 Wykres rozrzutu zmiennej **x9** względem **x13**



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie 4.5 zaprezentowany jest wykres rozrzutu zmiennej **x9** względem **x13**. Współczynnik korelacji między zmiennymi wynosi 0,69606 i jest on dodatni. Oznacza to, że wraz ze wzrostem ilości nasadzonych drzew w miastach i na wsiach rośnie powierzchnia obszarów gruntów zdewastowanych i zdegradowanych, które zostały zrekultywowane i zagospodarowane na cele rolnicze i leśne. Województwo śląskie również w tym przypadku odstaje od pozostałych.

Wykres 4.6 Wykres rozrzutu zmiennej **x4** względem **x15**



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie 4.6 zaprezentowany jest wykres rozrzutu zmiennej **x4** względem **x15**. Współczynnik korelacji między zmiennymi -0,85783 i jest on ujemny. Oznacza to, że wraz ze wzrostem powierzchni obszarów Natura 2000 maleje emisja tlenków azotu, a zmienne są mocno skorelowane.

Z dalszej analizy wykluczamy zmienne x1, x7, x10, x12 ponieważ słabo korelują z innymi zmiennymi. Zmienne x6 i x14 nie posiadają znaczących korelacji z żadną zmienną, więc także należy je wyeliminować, ponieważ wykazują nieistotność statystyczną. Zmienne x2 i x16 charakteryzują się skośnością wykraczającą poza obszar -od 3,09 do 3,09, więc też należy je wyeliminować. Zmienna x13 pomimo tworzenia tylko jednej pary korelacji powyżej 0,50, również została użyta w badaniu ze względu na stosunkowo wysokie korelacje z innymi zmiennymi oraz istotne znaczenie dla środowiska. **Poniżej przedstawiona jest macierz wybranych zmiennych do analizy.**

Tabela 4.2 Macierz wybranych zmiennych do analizy

Zmienna	Korelacje (AW-ochrona-srodowiska) Oznaczone wsp. korelacji są istotne z $p < ,05000$ N=16 (Braki danych usuwano przypadkami)							
	x3	x4	x5	x8	x9	x11	x13	x15
x3	1,00	0,50	0,51	-0,52	-0,70	0,66	-0,25	-0,59
x4	0,50	1,00	0,76	-0,38	-0,41	0,79	-0,21	-0,86
x5	0,51	0,76	1,00	-0,72	-0,52	0,78	-0,42	-0,77
x8	-0,52	-0,38	-0,72	1,00	0,68	-0,50	0,47	-0,60
x9	-0,70	-0,41	-0,52	0,68	1,00	-0,65	0,70	0,64
x11	0,66	0,79	0,78	-0,50	-0,65	1,00	-0,37	-0,80
x13	-0,25	-0,21	-0,42	0,47	0,70	-0,37	1,00	0,40
x15	-0,59	-0,86	-0,77	-0,60	0,64	-0,80	0,40	1,00

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

4.2. Wybór głównych składowych oraz ich równania

W etapie tym należy wybrać główne składowe. Liczba głównych składowych powinna być mniejsza od liczby zmiennych wyjściowych. Aby ograniczyć liczbę głównych składowych w dalszej analizie skorzystano z następujących kryteriów:

- wybranie takiej liczby głównych składowych, których skumulowany procent wyjaśnia 70-90% całkowitej zmienności
- kryterium Kaisera – wybranie tych głównych składowych, których wartości własne są większe od jedności. Należy odrzucić składowe o wartościach własnych mniejszych od 1,00.
- Reguła Catetella – polega na analizie wykresu Osypiska. Zgodnie z kryterium należy „odciąć” składowe, w miejscu w którym ma miejsce łagodny spadek wartości własnych. Odrzucane są składowe za tzw. kolankiem.

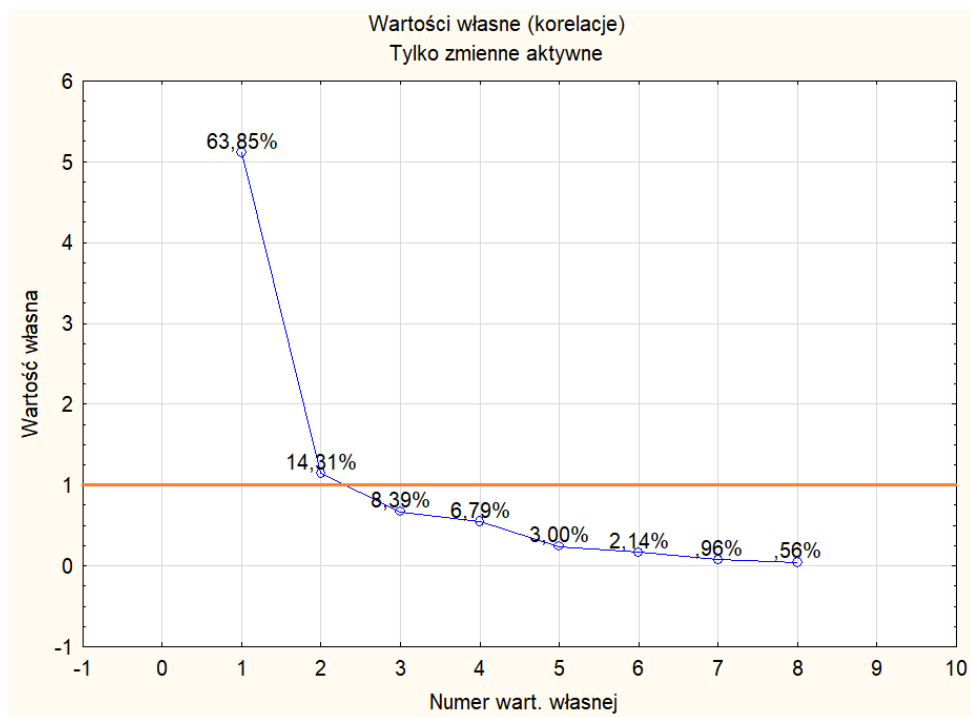
Tabela 4.3 Wartości własne składowych

Nr wartości	Wartość własna	% ogółu Warianc.	Skumulowana Wartość wł	Skumul. %
PCA 1	5,11	63,85	5,11	63,85
PCA 2	1,14	14,31	6,25	78,16
PCA 3	0,67	8,39	6,92	86,54
PCA 4	0,54	6,79	7,47	93,33
PCA 5	0,24	3,00	7,71	96,33
PCA 6	0,17	2,14	7,88	98,48
PCA 7	0,08	0,96	7,96	99,44
PCA 8	0,04	0,56	8,00	100,00

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

W tabeli 4.3 zostały przedstawione wartości własne wszystkich składowych. W celu ograniczenia liczby głównych składowych w dalszej części analizy, zastosowałam kryterium Kaisera. Zgodnie z nim do analizy zostały przyjęte 2 główne składowe, **PCA 1** i **PCA 2**, których wartość jest większa od 1,00. Sumarycznie tłumaczą one 78,16% zmienności.

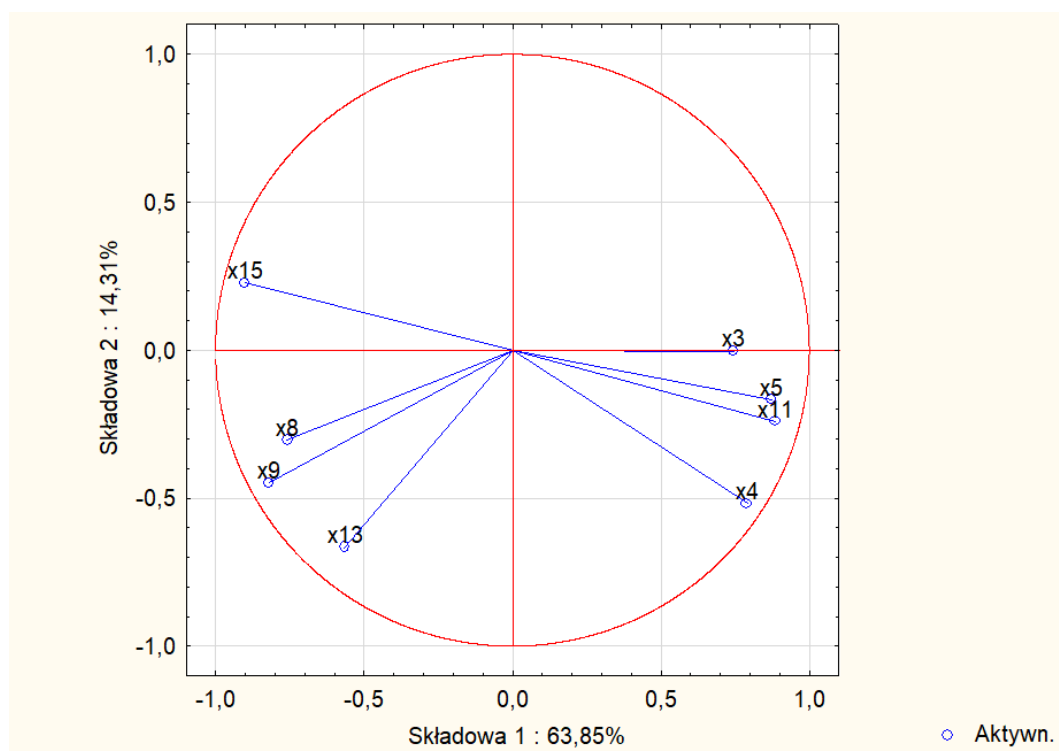
Wykres Osypiska



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na wykresie Osypiska „kolanko” tworzy się pomiędzy drugą a trzecią składową. Od trzeciej składowej wartość własne nie różnią się znacząco. Wniosek z kryterium Cattella pokrywa się z wnioskiem z kryterium Kaisera, co potwierdza słuszność wybrania do dalszej analizy 2 pierwszych głównych składowych.

4.3. Projekcja zmiennych na płaszczyznę składowych (1 x 2)



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Długość wektora zmiennej oznacza moc dyskryminacyjną tzn. zdolność zmiennej do różnicowania elementów w wierszach. Z projekcji wynika, że zmienne mają zróżnicowane znaczenie, ponieważ mają różne długości wektorów.

Z projekcji zmiennych na płaszczyznę składowych **PCA 1** oraz **PCA 2** można wywnioskować, że najmniejszy wpływ na budowanie pierwszej oraz drugiej składowej ma zmienna x3, której wektor jest najkrótszy, a następnie zmienna x8. Największy wpływ na pierwszą i drugą składową ma zmienna x15, a następnie x4 których wektory są najdłuższe. Para zmiennych x5 i x11 jest ze sobą dosyć silnie skorelowana oraz x8 i x9 jest ze sobą również dosyć skorelowana, ponieważ kąty pomiędzy nimi są najmniejsze.

Tabela 4.4 Wektory własne macierzy korelacji

Zmienna	Składowa 1	Składowa 2	Składowa 3	Składowa 4	Składowa 5	Składowa 6	Składowa 7	Składowa 8
x3	0,330	-0,003	0,766	-0,031	-0,137	0,513	-0,108	-0,102
x4	0,350	-0,483	-0,205	-0,207	0,229	0,226	0,648	-0,190
x5	0,386	-0,158	-0,329	0,338	-0,458	0,239	-0,075	0,575
x8	-0,336	-0,285	0,033	-0,758	-0,158	0,131	-0,132	0,413
x9	-0,363	-0,420	-0,259	0,202	-0,199	0,401	-0,353	-0,509
x11	0,392	-0,226	0,044	-0,240	-0,510	-0,573	-0,165	-0,348
x13	-0,250	-0,623	0,423	0,402	0,120	-0,349	0,086	0,255
x15	-0,399	0,212	0,120	0,086	-0,616	0,035	0,622	-0,079

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Równania głównych składowych:

- Bilans jakości powietrza**

$$Y_1 = 0,330x_3 + 0,350x_4 + 0,386x_5 - 0,336x_8 - 0,363x_9 + 0,392x_{11} - 0,250x_{13} - 0,399x_{15}$$

- Obszary podnoszące walory stanu środowiska naturalnego**

$$Y_2 = -0,003x_3 - 0,483x_4 - 0,158x_5 - 0,285x_8 - 0,420x_9 - 0,226x_{11} - 0,623x_{13} + 0,212x_{15}$$

Tabela 4.5 Macierz korelacji składowych

Zmienna	Składowa 1	Składowa 2
x3	0,745	-0,004
x4	0,791	-0,517
x5	0,873	-0,169
x8	-0,759	-0,305
x9	-0,820	-0,449
x11	0,885	-0,241
x13	-0,566	-0,667
x15	-0,902	0,227

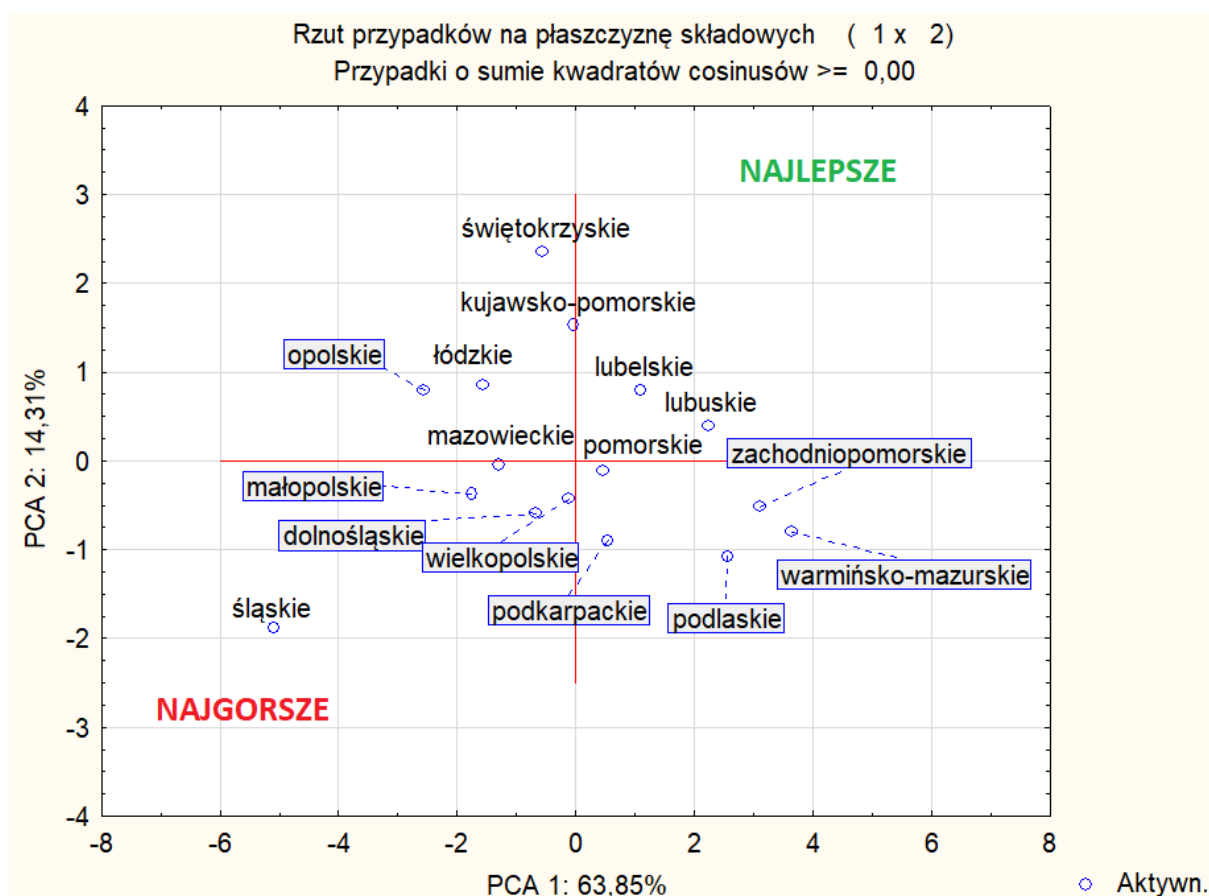
Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Analizując macierz korelacji składowych można zauważyć, że duży wkład w budowanie pierwszej składowej mają wszystkie zmienne, a przede wszystkim zmienna x5, x11 oraz x15. Spośród zmiennych budujących pierwszą składową x8, x9, x13 oraz x15 wykazują korelację ujemną. Największy wkład w budowanie drugiej składowej mają zmienne x4, x9 i x13, gdzie każda zmienna wykazuje korelację ujemną.

W równaniu pierwszej głównej składowej najsilniejsze korelacje występują dla zmiennych mówiących o obszarach Natura 2000, liczbie pomników przyrody, nasadzeniach drzew, powierzchni ekologicznych gospodarstw rolnych oraz emisji tlenków azotu. Wzrostowi niektórych zmiennych towarzyszy spadek innych i na odwrót lub jednocześnie maleją albo rosną. W związku z tym, że zmienne działają korzystnie lub nie na stan jakości powietrza pierwszą składową nazwano: **bilans jakości powietrza**.

W równaniu drugiej głównej składowej najsilniejsze korelacje występują dla zmiennych mówiących o obszarach Natura 2000, nasadzeniach drzew oraz gruntach zdewastowanych i zdegradowanych zrehabilitowanych oraz zagospodarowanych na cele rolnicze i leśne. W związku z tym, że zmienne powodują wzrost powierzchni obszarów korzystnie wpływających na stan jakości środowiska drugą składową nazwano: **obszary podnoszące walory stanu środowiska naturalnego**.

4.4. Rzut obserwacji na płaszczyznę składowych (1 x 2)



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na powyższym rysunku widnieje projekcja rzutu obserwacji – województw na płaszczyznę dwóch głównych składowych PCA 1 i PCA 2. Wyodrębnione zostały dwie główne składowe stanowiące osie główne układu, które wyjaśniają 78,16% całkowitej zmienności zmiennych aktywnych wziętych do analizy.

Na podstawie projekcji można zauważyć, że w grupie województw charakteryzujących się najlepszym stanem środowiska naturalnego są województwo kujawsko-pomorskie, lubelskie oraz lubuskie. Niewiele pod linią znajduje się województwo pomorskie, zachodniopomorskie oraz warmińsko mazurskie. W ćwiartce z województwami odznaczającymi się najgorszym stanem środowiska są województwo mazowieckie, małopolskie, dolnośląskie, wielkopolskie oraz śląskie. Zdecydowanie najgorszym stanem środowiska naturalnego odznacza się województwo śląskie.

5. ANALIZA SKUPIEŃ

Analiza skupień (*ang. cluster analysis*) to dział wielowymiarowej analizy statystycznej obejmujący zbiór technik znajdujących zastosowanie w badaniu danych o obiektach wielowymiarowych (jednostek lub zmiennych) mający za zadanie podzielenie zbioru niejednorodnych obiektów na grupy – **skupienia**, które są do siebie podobne, jednorodne.

Celem analizy skupień jest wykrycie homogenicznych, grup obiektów w zbiorze danych. Na jego podstawie bada się podobieństwa lub odrębności obiektów i ich zbiorów.

Wyodrębnienie klas powinno spełniać dwa kryteria:

- **Wewnętrzną spójność obiektów**
- **Zewnętrzną izolację z innymi obiektami poza grupą**

Do analizy zostały użyte te same zmienne co w analizie PCA.

Tabela 5.1 Macierz odległości

Nr przypadku	DLS	K-P	LUBE	LBU	ŁDZ	MŁP	MAZ	OPO	PDKR	PODL	POM	ŚLĄ	SWI	W-M	WLKP	Z-POM
DLS	0,00	2,15	1,21	1,72	1,63	1,27	0,88	1,61	1,52	2,01	1,09	2,33	1,76	2,70	0,82	2,25
K-P	2,15	0,00	1,84	1,73	1,58	1,54	2,77	2,54	2,65	2,53	1,69	3,83	1,67	2,31	1,61	2,69
LUBE	1,21	1,84	0,00	1,38	1,71	1,44	1,73	2,05	1,68	1,97	1,29	3,13	1,46	2,17	0,91	2,06
LBU	1,72	1,73	1,38	0,00	2,36	2,00	2,44	2,98	1,51	1,04	0,91	3,98	2,30	1,26	1,36	1,07
ŁDZ	1,63	1,58	1,71	2,36	0,00	0,95	1,89	1,21	2,79	3,08	2,04	2,52	0,87	3,11	1,35	3,23
MŁP	1,27	1,54	1,44	2,00	0,95	0,00	1,79	1,64	2,21	2,59	1,55	2,48	1,38	2,69	0,87	2,80
MAZ	0,88	2,77	1,73	2,44	1,89	1,79	0,00	1,36	2,08	2,69	1,81	1,88	1,92	3,48	1,54	2,92
OPO	1,61	2,54	2,05	2,98	1,21	1,64	1,36	0,00	2,95	3,49	2,45	1,70	1,40	3,90	1,85	3,68
PDKR	1,52	2,65	1,68	1,51	2,79	2,21	2,08	2,95	0,00	1,16	1,18	3,41	2,85	2,19	1,54	1,35
PODL	2,01	2,53	1,97	1,04	3,08	2,59	2,69	3,49	1,16	0,00	1,23	4,19	3,13	1,48	1,84	0,68
POM	1,09	1,69	1,29	0,91	2,04	1,55	1,81	2,45	1,18	1,23	0,00	3,29	2,10	1,98	0,93	1,57
ŚLĄ	2,33	3,83	3,13	3,98	2,52	2,48	1,88	1,70	3,41	4,19	3,29	0,00	2,90	4,80	2,73	4,43
SWI	1,76	1,67	1,46	2,30	0,87	1,38	1,92	1,40	2,85	3,13	2,10	2,90	0,00	3,15	1,57	3,21
W-M	2,70	2,31	2,17	1,26	3,11	2,69	3,48	3,90	2,19	1,48	1,98	4,80	3,15	0,00	2,15	1,35
WLKP	0,82	1,61	0,91	1,36	1,35	0,87	1,54	1,85	1,54	1,84	0,93	2,73	1,57	2,15	0,00	2,07
Z-POM	2,25	2,69	2,06	1,07	3,23	2,80	2,92	3,68	1,35	0,68	1,57	4,43	3,21	1,35	2,07	0,00

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

W tabeli 5.1 została przedstawiona macierz odległości pomiędzy wszystkimi województwami. Na jej podstawie można wskazać jednostki najbardziej do siebie zbliżone, województwo podlaskie oraz zachodnio-pomorskie. Od tych 2 województw rozpoczyna się przebieg aglomeracji przedstawiony w tabeli 5.2. Najdalej oddalonymi od siebie jednostkami są województwa warmińsko-mazurskie oraz śląskie.

Tabela 5.2 Przebieg aglomeracji

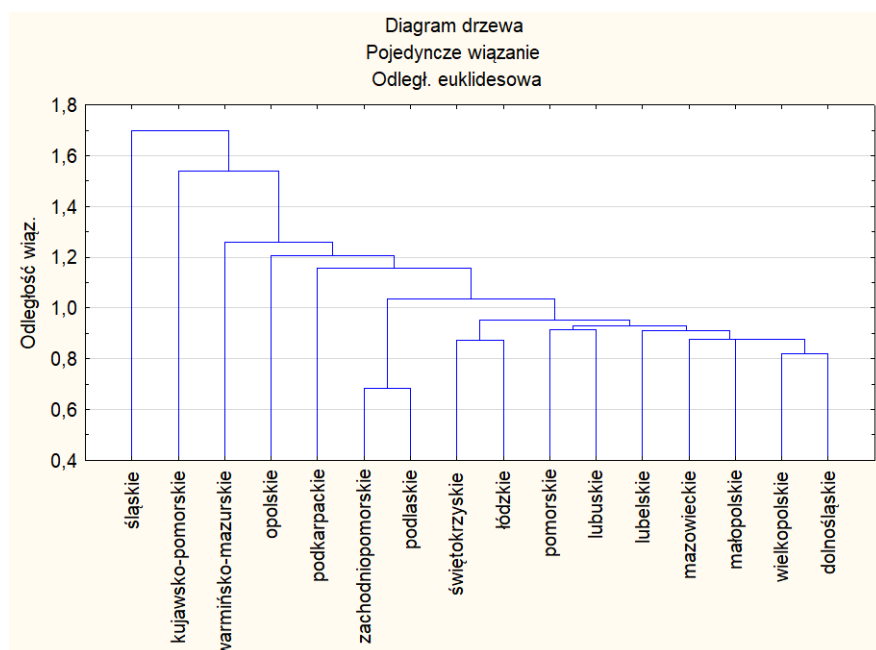
Odległość	Przebieg aglomeracji (AW-ochrona-srodowiska)															
0,6846897	PODL	Z-POM														
0,8196951	DLS	WLKP														
0,8737276	ŁDZ	SWI														
0,9144397	LBU	POM														
1,205031	DLS	WLKP	LUBE													
1,351703	PDKR	PODL	Z-POM													
1,358970	MAZ	OPO														
1,379493	ŁDZ	SWI	MŁP													
1,571751	LBU	POM	PDKR	PODL	Z-POM											
1,666313	K-P	ŁDZ	SWI	MŁP												
1,880665	MAZ	OPO	ŚLĄ													
2,148907	DLS	WLKP	LUBE	K-P	ŁDZ	SWI	MŁP									
2,186847	LBU	POM	PDKR	PODL	Z-POM	W-M										
3,227848	DLS	WLKP	LUBE	K-P	ŁDZ	SWI	MŁP	LBU	POM	PDKR	PODL	Z-POM	W-M			
4,797218	DLS	WLKP	LUBE	K-P	ŁDZ	SWI	MŁP	LBU	POM	PDKR	PODL	Z-POM	W-M	MAZ	OPO	ŚLĄ

Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

W tabeli 5.2 został przedstawiony przebieg aglomeracji wraz z dokładnymi odległościami pomiędzy skupieniami. Pierwsze słupienie powstało na wysokości 0,68 dla województwa podlaskiego oraz zachodnio-pomorskiego tak jak zaobserwowano w macierzy odległości. Najdalej oddaloną jednostką od pozostałych jest województwo śląskie.

Poniżej został przedstawiony graficzny przebieg aglomeracji w postaci dendrogramów wykonanych różnymi metodami. Ilustruje on kolejność oraz poziomy na jakich wystąpiło łączenie jednostek w skupienia. W zależności od konfiguracji dendrogramu ich wizualizacje różnią się między sobą, jednakże finalnie wyniki są podobne.

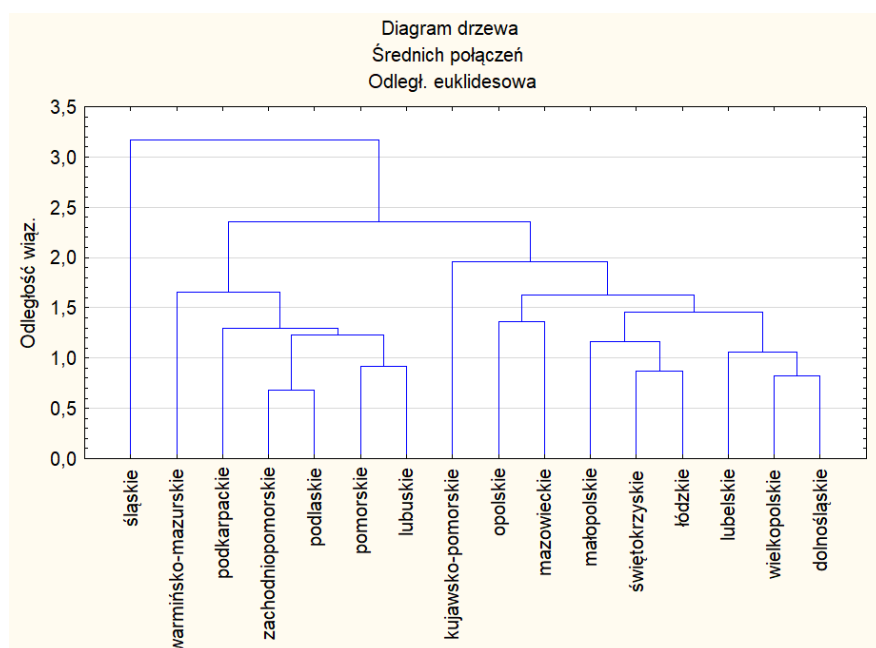
Rysunek 5.1 Dendrogram metodą najbliższego sąsiada



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Rysunek 5.1 przedstawia metodę najbliższego sąsiada – pojedyncze wiązanie, odległość euklidesowa. Dendrogram powstał poprzez łączenie się w skupienia jednostek znajdujących się najbliżej siebie. Ze względu na powstanie „efektu schodków” nie należy on do udanych i nie przedstawia dobrego podziału aglomeracji. Trudno jest na jego podstawie pogrupować jednostki w skupienia.

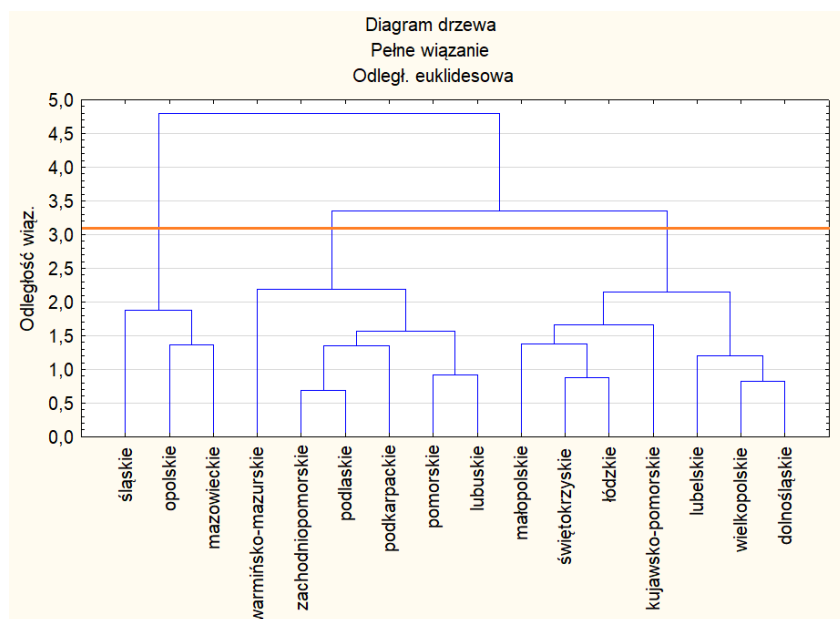
Rysunek 5.2 Dendrogram metodą średnich połączeń



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Rysunek 5.2 przedstawia metodę średnich połączeń – odległość euklidesowa. Dendrogram powstał poprzez łączenie się w skupienia jednostek znajdujących się w określonej (średniej) odległości. Odległość między dwoma skupieniami obliczona jest jako średnia odległość między wszystkimi jednostkami znajdującymi się w dwóch różnych skupieniach. Ta metoda również dosyć nieczytelnie dzieli jednostki, w bardzo nierównoliczne skupienia. Powstają 3 skupienia gdzie jedno z nich zawiera tylko jedno województwo. Jednakże na podstawie dwóch powyższych dendrogramów można już zauważyć, że województwo śląskie odstaje od pozostałych.

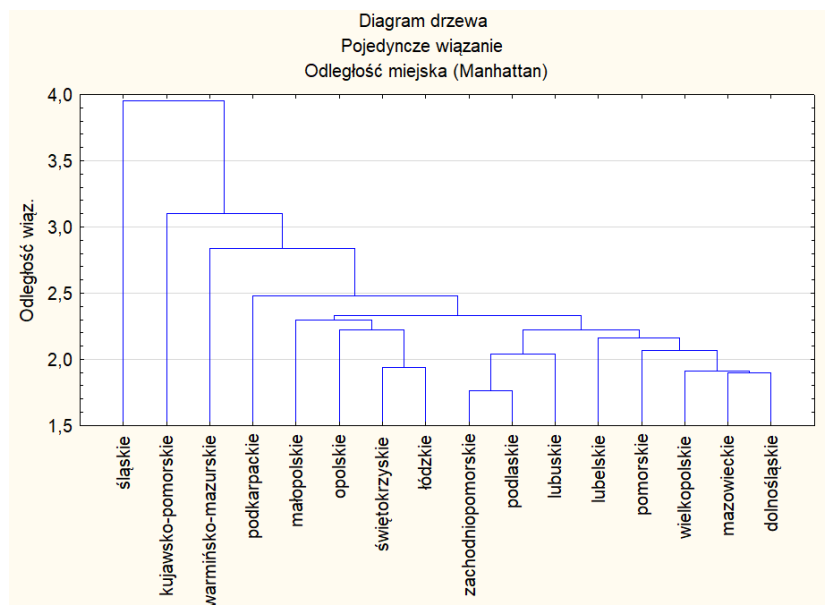
Rysunek 5.3 Dendrogram metodą najdalszego sąsiada



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Rysunek 5.3 przedstawia metodę najdalszego sąsiada – pełne wiązanie, odległość euklidesowa. Dendrogram powstał poprzez łączenie się w skupienia jednostek znajdujących się w najdalszej odległości od siebie. Odległość między skupieniami jest określona przez największą z odległości pomiędzy 2 dowolnymi jednostkami przynależącymi do innych skupień. Na powyższym dendrogramie można wyróżnić 3 skupienia dokonując cięcia na wysokości zaznaczonej pomarańczową linią. W skład skupień wchodzi odpowiednio następujące województwa, skupienie nr 1: śląskie, opolskie i mazowieckie, skupienie nr 2: dolnośląskie, kujawsko-pomorskie, lubelskie, łódzkie, małopolskie, świętokrzyskie i wielkopolskie oraz skupienie nr 3: lubuskie, podkarpackie, podlaskie, pomorskie, warmińsko-mazurskie i zachodniopomorskie.

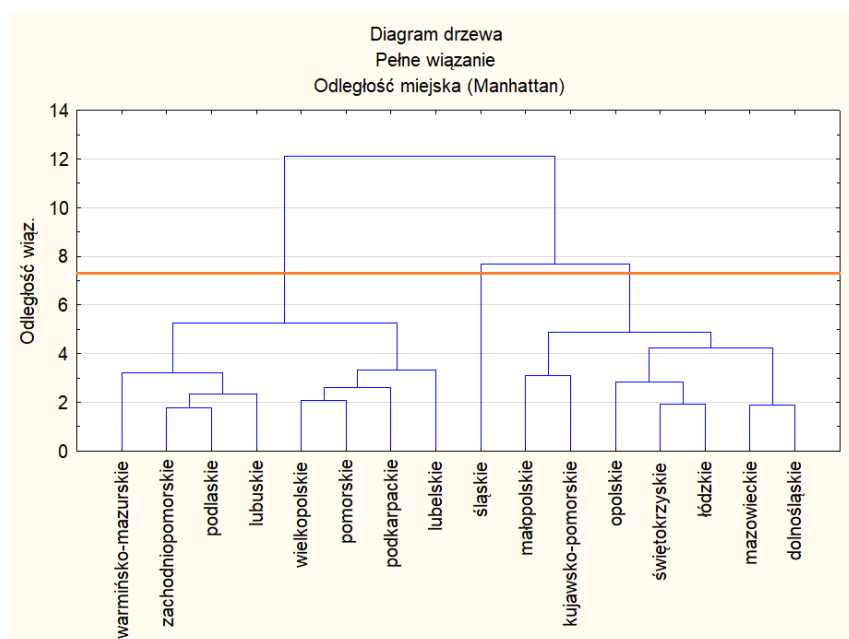
Rysunek 5.4 Dendrogram metodą najbliższego sąsiada



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Rysunek 5.4 przedstawia metodę najbliższego sąsiada – pojedyncze wiązanie, odległość miejska. Dendrogram powstał poprzez łączenie się w skupienia jednostek znajdujących się najbliżej siebie. W tej metodzie tak samo jak z wykorzystaniem odległości euklidesowej powstał „efekt schodków”, więc dendrogram nie przedstawia dobrego podziału aglomeracji. Trudno jest na jego podstawie pogrupować jednostki w skupienia.

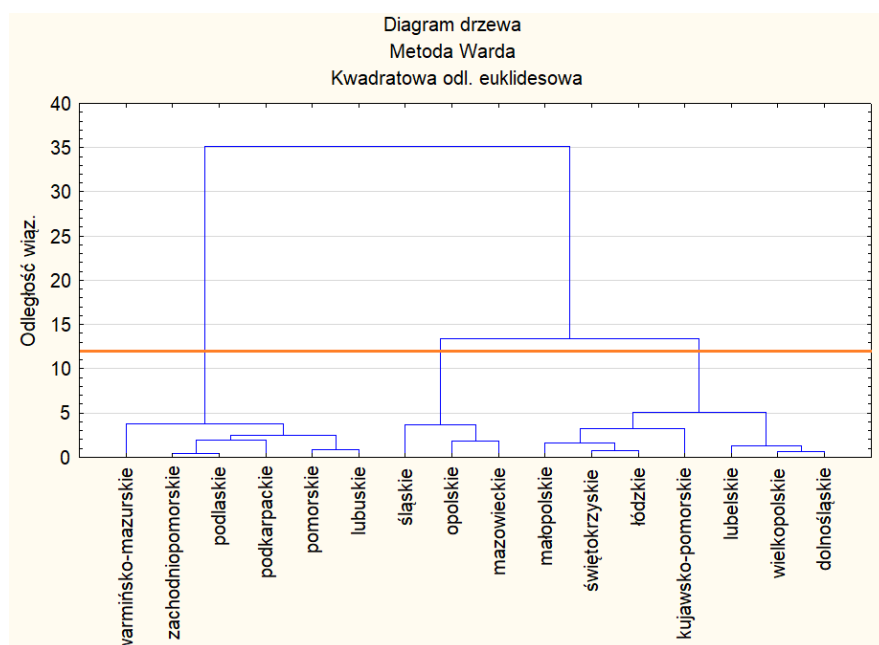
Rysunek 5.5 Dendrogram metodą najdalszego sąsiada



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Rysunek 5.5 przedstawia metodę najdalszego sąsiada – pełne wiązanie, odległość miejska. Dendrogram powstał poprzez łączenie się w skupienia jednostek znajdujących się w najdalszej odległości od siebie. W tym przypadku podział na skupienia jest inny niż w zastosowaniu odległości euklidesowej. Ta metoda zdecydowanie mniej różnicuje skupienia i dzieli jednostki w taki sposób, że powstaje skupienie w którym znajduje się jedno województwo.

Rysunek 5.6 Dendrogram metodą Warda



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Rysunek 5.6 przedstawia metodę Warda z kwadratową odległością euklidesową, która zapewnia spójność wewnętrzną klas, dlatego, że uwzględnia wariancję. Dendrogram w tej metodzie charakteryzuje się spłaszczeniem w porównaniu do innych metod. Wiązania występują na dużych odległościach. Można zauważyć, że podział na skupienia jest tak samo rozdzielony jak w przypadku metody najdalszego sąsiada z odległością euklidesową, tyle, że w tym przypadku województwa charakteryzujące się najlepszym stanem środowiska naturalnego są w najdalszej odległości od pozostałych. Uzyskujemy podział na 3 skupienia.

Rysunek 5.7 Wykres przebiegu aglomeracji w grupowaniu metodą najdalszych sąsiadów



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na rysunku 5.7 został przedstawiony wykres przebiegu aglomeracji w grupowaniu metodą najdalszych sąsiadów. W miejscu pierwszego znaczącego skoku odległości aglomeracyjnej zostało wyznaczone optymalne przecięcie dendrogramu (rysunek 5.3), w wyniku którego 16 województw zostało podzielonych na 3 skupienia.

Następnie dla skupień otrzymanych metodą najdalszego sąsiada zostały obliczone średnie przedstawione w tabeli 5.3. kolorem zielonym zostały zaznaczone wartości powyżej średniej dla skupień dla danej zmiennej.

Tabela 5.3 Średnie dla uzyskanych skupień

Zmienna	Skupienie 1	Skupienie 2	Skupienie 3	Średnia
x3	0,88	1,14	1,35	1,17
x4	0,62	0,91	2,54	1,47
x5	0,58	0,86	1,40	1,01
x8	5,21	3,62	3,33	3,81
x9	1,02	0,58	0,45	0,62
x11	0,70	0,89	1,36	1,03
x13	0,94	0,68	0,57	0,68
x15	1,14	0,79	0,34	0,68

Źródło: Opracowanie własne na podstawie obliczeń w programie Excel

Poniżej przedstawiono elementy wchodzące w skład poszczególnych skupień wraz z odległościami oraz nadano im nazwy.

Skupienie 1 – zły stan środowiska

Mazowieckie	0,3309
Opolskie	0,2867
Śląskie	0,3907

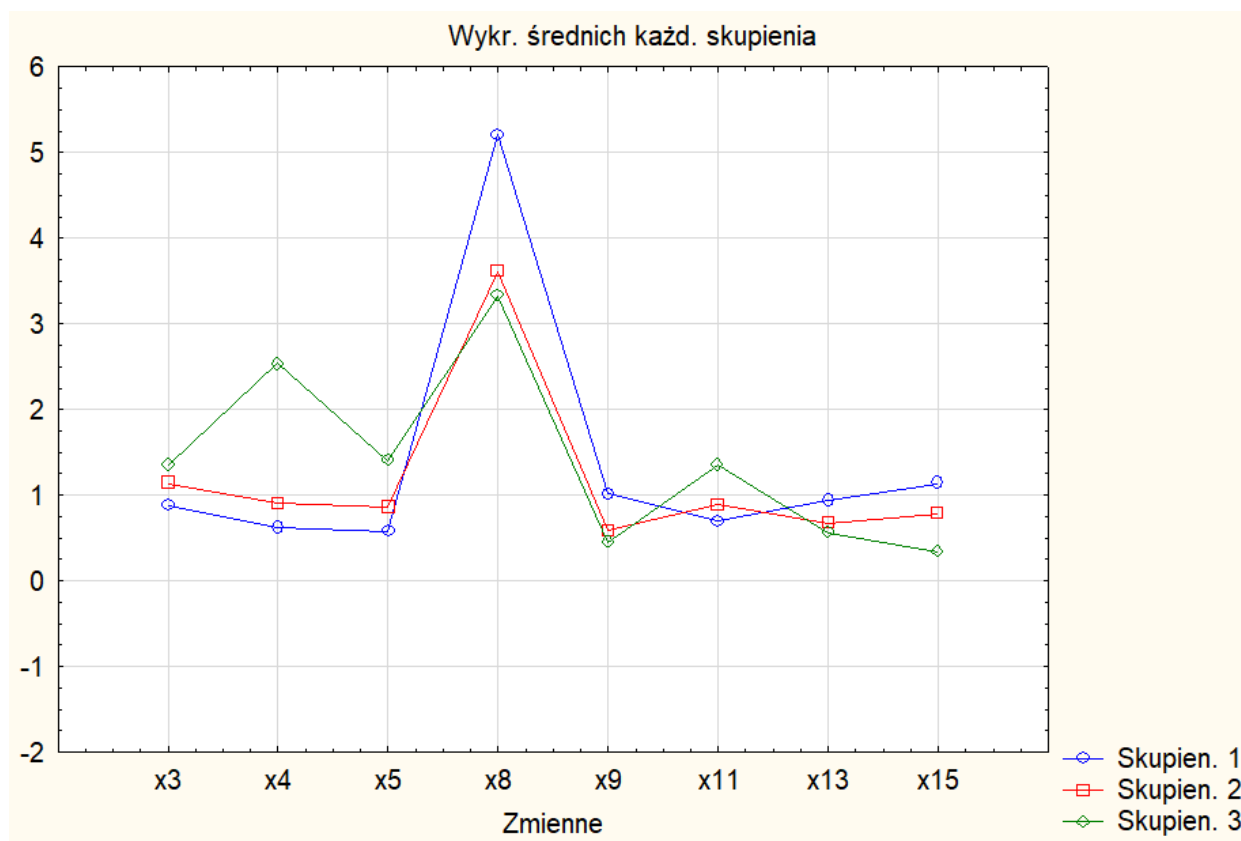
Skupienie 2 – umiarkowany stan środowiska

Dolnośląskie	0,3728
Kujawsko-pomorskie	0,4614
Lubelskie	0,3402
Łódzkie	0,3058
Małopolskie	0,2428
Świętokrzyskie	0,3492
Wielkopolskie	0,2245

Skupienie 3 – dobry stan środowiska

Lubuskie	0,2111
Podkarpackie	0,3773
Podlaskie	0,1943
Pomorskie	0,3338
Warmińsko-mazurskie	0,4454
Zachodniopomorskie	0,2478

Rysunek 5.8 Średnie dla trzech skupień



Źródło: Opracowanie własne na podstawie obliczeń w programie Statistica

Na podstawie wykresu dla średnich skupień (rysunek 5.7) można zauważyć, że każda zmienna odpowiednio różnicuje profile danych skupień, odległości są znaczące, co świadczy o tym, że zostały one poprawnie dobrane. Zmienna x4 oraz x11 spowodowała „rozjechanie się” wartości z zupełnie inną stroną co znacznie różnicuje dane skupienia. Można zauważyć, że wartości skupień dla zmiennych mające charakter zdecydowanie pozytywny dla środowiska, takie jak obszary Natura 2000, liczba pomników przyrody czy powierzchnia ekologicznych gospodarstw rolnych są zdecydowanie wyższe dla trzeciego skupienia, czyli tego gdzie znajdują się województwa o dobrym stanie środowiska. W przypadku skupienia 1 sytuacja jest odwrotna. Dla tego skupienia wartości te są najniższe.

6. PODSUMOWANIE, WNIOSKI

W poniższej tabeli 6.1 zostały zestawione ze sobą metody porządkowania liniowego oraz analiza skupień. W tym przypadku kolory zostały nadane na podstawie wyników z analizy skupień. Kolor zielony to grupa województw charakteryzująca się dobrym stanem środowiska naturalnego, kolor żółty to województwa odznaczające się umiarkowanym stanem środowiska naturalnego, natomiast kolor czerwony to województwa w których stan środowiska jest zły.

Na podstawie poniższej tabeli można wyróżnić grupę województw: warmińsko-mazurskie, podlaskie, podkarpackie oraz zachodniopomorskie, które niezależnie od wybranej metody znajdują się na szczycie każdego rankingu.

Województwa z grupy pośredniej – żółtej, mniej więcej znajdują się na środku tabeli, jednakże można zauważyć, że świętokrzyskie zdecydowanie inaczej zostało sklasyfikowane w metodach porządkowania liniowego niż w analizie skupień.

Grupa województw z najgorszym stanem środowiska rozlokowana jest w dolnej części tabeli, z jednym wyjątkiem w metodzie TOPSIS, gdzie województwo śląskie jest zdecydowanie wyżej niż w pozostałych.

Wyniki z powyższych analiz przekładają się na rzeczywistość. W województwach zaznaczonych na czerwono zdecydowanie bardziej rozwinięty jest przemysł co przejawia się złym stanem środowiska. Województwa z grupy żółtej odznaczają się mniej lub bardziej rozwiniętym stanem sektora przemysłowego, więc znajdują się w klasie pośredniej. Województwa zaznaczone na kolor zielony zdecydowanie są znane ze swoich walorów przyrodniczych oraz odpowiedniego dbania o nie.

Jeśli chodzi o porównanie metod porządkowania liniowego oraz analizy skupień z analizą głównych składowych to wyniki są dosyć odmienne. Jedynie województwo lubuskie i ewentualnie pomorskie znajduje się w tej samej grupie – charakteryzującej się dobrym stanem środowiska. Województwo mazowieckie oraz śląskie, w obu przypadkach są w grupie z najgorszym stanem środowiska. Województwo warmińsko-mazurskie, podlaskie oraz podkarpackie, które znajdowało się na podium, w metodzie PCA znalazło się w grupie pośredniej.

Tabela 6.1 Podsumowanie metod porządkowania liniowego i analizy skupień

Porządkowanie liniowe						Analiza skupień
M. wzorca rozwoju Hellwiga	M. wzorca z wagami	M. TOPSIS	M. standaryzowanych sum	M. sumy rang	M. dystansów	
podkarpackie	warmińsko-mazurskie	warmińsko-mazurskie	warmińsko-mazurskie	warmińsko-mazurskie	warmińsko-mazurskie	lubuskie
podlaskie	podlaskie	podlaskie	podlaskie	podlaskie	podlaskie	podkarpackie
warmińsko-mazurskie	podkarpackie	zachodniopomorskie	zachodniopomorskie	podkarpackie	zachodniopomorskie	podlaskie
wielkopolskie	zachodniopomorskie	podkarpackie	podkarpackie	zachodniopomorskie	podkarpackie	pomorskie
zachodniopomorskie	wielkopolskie	pomorskie	lubelskie	wielkopolskie	pomorskie	warmińsko-mazurskie
dolnośląskie	dolnośląskie	śląskie	wielkopolskie	lubelskie	lubuskie	zachodniopomorskie
małopolskie	kujawsko-pomorskie	kujawsko-pomorskie	lubuskie	pomorskie	wielkopolskie	dolnośląskie
lubelskie	lubuskie	dolnośląskie	pomorskie	lubuskie	lubelskie	kujawsko-pomorskie
pomorskie	małopolskie	lubuskie	dolnośląskie	dolnośląskie	małopolskie	lubelskie
lubuskie	lubelskie	łódzkie	małopolskie	małopolskie	kujawsko-pomorskie	łódzkie
łódzkie	łódzkie	małopolskie	śląskie	mazowieckie	śląskie	małopolskie
kujawsko-pomorskie	pomorskie	wielkopolskie	kujawsko-pomorskie	opolskie	opolskie	świętokrzyskie
mazowieckie	śląskie	opolskie	łódzkie	kujawsko-pomorskie	mazowieckie	wielkopolskie
śląskie	opolskie	lubelskie	mazowieckie	łódzkie	dolnośląskie	mazowieckie
opolskie	mazowieckie	mazowieckie	opolskie	śląskie	łódzkie	opolskie
świętokrzyskie	świętokrzyskie	świętokrzyskie	świętokrzyskie	świętokrzyskie	świętokrzyskie	śląskie

Źródło: Opracowanie własne na podstawie programu Excel

Na podstawie przeprowadzonej analizy został omówiony cel badania, czyli ocena stanu środowiska w województwach na terenie Polski w 2019r., poprzez wpływ wybranych zmiennych. Zostały wykonane rankingi województw od tych, które charakteryzują się dobrym stanem środowiska naturalnego do tych w których stan środowiska jest zły, wykorzystując do tego metody analizy wielowymiarowej. Zdecydowanie województwa odznaczające się wyższym stopniem „zazielenienia” oraz wysokimi współczynnikami wartości pozytywnych dla ekologii, wyróżniają się lepszym stanem środowiska naturalnego od reszty.

7. BIBLIOGRAFIA

- Balicki A., *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2013
- Bąk A., Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza, Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu
- Główny Urząd Statystyczny – Ochrona środowiska 2020
<https://stat.gov.pl/obszary-tematyczne/srodowisko-energia/srodowisko/ochrona-srodowiska-2020,1,21.html>
- Główny Urząd Statystyczny – STAN I OCHRONA ŚRODOWISKA
<https://bdl.stat.gov.pl/BDL/dane/podgrup/temat>