

Iwona BAŁ, Katarzyna WAWRZYŃIAK

BADANIE JAKOŚCI ŚRODOWISKA NATURALNEGO W WOJEWÓDZTWIE ZACHODNIOPOMORSKIM W 2006 ROKU Z WYKORZYSTANIEM WIELOWYMIAROWEJ ANALIZY KORESPONDENCJI

THE RESEARCH OF THE QUALITY OF THE NATURAL ENVIRONMENT IN ZACHODNIOPOMORSKIE VOIVODSHIP IN 2006 WITH THE USE OF MULTIPLE CORRESPONDENCE ANALYSIS

Katedra Zastosowań Matematyki w Ekonomii, Zachodniopomorski Uniwersytet Technologiczny
w Szczecinie, ul. K. Janickiego 31, 71-270 Szczecin

Abstract. The article takes the attempt to answer to the question about the condition of the quality of natural environment in Zachodniopomorskie Voivodship with regard to an average level of chosen variables characterizing this environment in 2006. For the analysis were chosen variables describing air, water and soils pollution. Besides, it has taken two variables into considerations which inform properly about: carried expenditures for environmental protection and particular natural values of researched area. As the investigate instrument was used the multiple correspondence analysis based on indicator matrix and Ward method.

Słowa kluczowe: analiza korespondencji, metoda Warda, środowisko naturalne, złożona macierz znaczników.

Key words: indicator matrix, multiple correspondence analysis, natural environment, Ward method.

WSTĘP

Stan środowiska naturalnego jest wypadkową jakości jego trzech podstawowych komponentów: powierzchni ziemi, zasobów wodnych i powietrza atmosferycznego. Ich zanieczyszczenie negatywnie oddziałuje na zasoby przyrody żywej, a także na warunki życia i zdrowie ludności. Rzetelna wiedza o stanie środowiska naturalnego i zmianach w nim zachodzących jest niezbędna do podejmowania racjonalnych decyzji, które mogą dotyczyć np. przeznaczenia gruntów, lokalizacji wszelkiego rodzaju zakładów czy – w szerszym zakresie – restrukturyzacji przemysłu w regionie (województwie, powiecie).

Głównym celem badania było wykazanie, które powiaty odbiegają *in plus* i *in minus* od przeciętnych wartości wybranych zmiennych, charakteryzujących jakość środowiska naturalnego w województwie zachodniopomorskim, oraz jakie są powiązania pomiędzy powiatami w zależności od tych zmiennych. Do analizy wybrano zmienne, które opisują stopień zanieczyszczenia powietrza, wody i gleby oraz szczególne walory przyrodnicze badanego obszaru, a także zmienne informujące o nakładach przeznaczonych w powiatach na ochronę środowiska. Jako narzędzie badawcze wykorzystano wielowymiarową analizę korespondencji na podstawie zbudowanej złożonej macierzy znaczników oraz metodę Warda.

MATERIAŁ I METODY¹

Wielowymiarowa analiza korespondencji jest stosowana wówczas, gdy badane zmienne mierzone są w najniższej skali pomiarowej (skala nominalna) i gdy wśród nich nie można wyróżnić w sposób jednoznaczny zmiennej zależnej (Gatnar i Walesiak 2004). Ze względu na rodzaj pomiaru i liczbę badanych zmiennych zbiorów danych wejściowych w tej metodzie można zapisać w postaci: macierzy Burta, złożonej macierzy znaczników, wielowymiarowej tablicy kontyngencji i łącznej tablicy kontyngencji.

W artykule zastosowano wielowymiarową analizę korespondencji z wykorzystaniem złożonej macierzy znaczników, której budowa rozpoczęła proces badawczy. Następnie wyznaczono wymiar rzeczywistej przestrzeni współwystępowania (K) na podstawie wzoru:

$$K = \sum_{q=1}^Q (J_q - 1) \quad (1)$$

gdzie:

J_q – liczba kategorii zmiennej q ($q = 1, 2, \dots, Q$),

Q – liczba zmiennych.

Ponieważ wynik wielowymiarowej analizy korespondencji jest przedstawiany graficznie zazwyczaj w przestrzeni dwuwymiarowej, sprawdzono najpierw w jakim stopniu wartości własne (λ_k) przestrzeni o niższym wymiarze wyjaśniają inercję całkowitą ($\lambda = \sum_{k=1}^K \lambda_k$). W tym celu zastosowano kryterium Greenacre'a, według którego za istotne dla badania uznaje się te wartości własne (λ_k), dla których spełniona jest nierówność:

$$\lambda_k > \frac{1}{Q} \quad (2)$$

gdzie:

Q – liczba zmiennych.

Następnie jakość odwzorowania w przestrzeni dwuwymiarowej podwyższono poprzez modyfikację wartości własnych według propozycji Greenacre'a na podstawie wzoru:

$$\tilde{\lambda}_k = \left(\frac{Q}{Q-1} \right)^2 \cdot \left(\sqrt{\lambda_k} - \frac{1}{Q} \right)^2 \quad (3)$$

gdzie:

λ_k – k -ta wartość własna ($k = 1, 2, \dots, K$),

Q – liczba zmiennych.

Zmodyfikowane wartości współrzędnych w przestrzeni dwuwymiarowej ($k = 2$) dla kategorii badanych zmiennych wyznaczono na podstawie wzoru:

$$\tilde{\mathbf{F}} = \mathbf{F}^* \cdot \mathbf{\Gamma}^{-1} \cdot \tilde{\mathbf{\Lambda}} \quad (4)$$

gdzie:

$\tilde{\mathbf{F}}$ – macierz zmodyfikowanych wartości współrzędnych dla kategorii badanych zmiennych o wymiarze $K \times k$,

¹ Opis sposobu postępowania w przypadku wielowymiarowej analizy korespondencji opracowano na podstawie pracy Stanimir (2005).

F^* – macierz pierwotnych wartości współrzędnych dla kategorii badanych zmiennych o wymiarze $K \times k$,

Γ^{-1} – diagonalna macierz odwrotna wartości osobliwych (γ_k) o wymiarze $k \times k$,

γ_k – k -ta wartość osobliwa będąca pierwiastkiem kwadratowym z k -tej wartości własnej (λ_k),

$\tilde{\Lambda}$ – diagonalna macierz zmodyfikowanych wartości własnych o wymiarze $k \times k$,

K – wymiar rzeczywistej przestrzeni współwystępowania.

Punkty (kategorie poszczególnych zmiennych), których współrzędne wyznaczono na podstawie wzoru (4), naniesiono w prostokątnym układzie współrzędnych. Przy interpretacji rozrzutu tych punktów uwzględniono ich położenie względem:

- początku układu współrzędnych (centrum rzutowania),
- innych punktów określających kategorie należące do tej samej zmiennej,
- punktu opisującego kategorie innej zmiennej.

Ostateczną interpretację wyników uzupełniono klasyfikacją kategorii zmiennych za pomocą metody Warda².

Badaną zbiorowość stanowiło 21 powiatów województwa zachodniopomorskiego w 2006 roku. Do pomiaru jakości środowiska naturalnego zastosowano następujące zmienne (Ochrona środowiska 2007): X_1 – zanieczyszczenia pyłowe powietrza zatrzymywane w urządzeniach do redukcji zanieczyszczeń, w procentach wytworzonych, X_2 – emisja zanieczyszczeń pyłowych powietrza z zakładów szczególnie uciążliwych, w tys. t na 1 km², X_3 – emisja zanieczyszczeń gazowych powietrza z zakładów szczególnie uciążliwych, w tys. t na 1 km², X_4 – odpady (z wyjątkiem odpadów komunalnych) wytworzone w ciągu roku, w tys. t na 1 km², X_5 – odpady komunalne (stałe) wywiezione w ciągu roku, w dam³ na 1 km², X_6 – ścieki przemysłowe i komunalne oczyszczone, w procentach, X_7 – ludność obsługiwana przez oczyszczalnie ścieków, w procentach, X_8 – nakłady inwestycyjne na ochronę środowiska, w mln zł na 1000 mieszkańców, X_9 – nakłady inwestycyjne na gospodarkę wodną, w mln zł na 1000 mieszkańców, X_{10} – powierzchnia o szczególnych walorach przyrodniczych prawnie chroniona, w procentach, X_{11} – liczba pomników przyrody na 1 km². W zbiorze zmiennych dominują stymulanty. Tylko cztery zmienne uznano za destymulanty (X_2 , X_3 , X_4 , X_5).

WYNIKI I DYSKUSJA

Wielowymiarową analizę korespondencji rozpoczęto od zbudowania złożonej macierzy znaczników. W macierzy tej liczba wierszy była równa liczbie badanych jednostek (powiatów), natomiast liczba kolumn odpowiadała podwojonej liczbie badanych zmiennych. Taka liczba kolumn wynikała z istoty macierzy znaczników, w której elementy przyjmują wyłącznie wartości 1 i 0. Dlatego każdą z badanych zmiennych zamieniono na zmienną zero-jedynkową według następującej zasady:

- dla stymulant

² Metoda Warda jest jedną z aglomeracyjnych metod grupowania. Jest stosowana w badaniach empirycznych zarówno w odniesieniu do klasyfikacji obiektów, jak i cech. W metodzie tej odległość między grupami jest definiowana jako moduł różnicy między sumami kwadratów odległości punktów od środków grup, do których punkty te należą (Malina 2004).

$$xs_i = \begin{cases} 1, & \text{gdy } x_i \geq M, \\ 0, & \text{gdy } x_i < M; \end{cases}$$

– dla destymulant

$$xd_i = \begin{cases} 1, & \text{gdy } x_i \leq M, \\ 0, & \text{gdy } x_i > M. \end{cases}$$

Przyjęcie mediany jako wartości granicznej wynikało z typu rozkładów badanych zmiennych, które w większości charakteryzowały się bardzo dużym zróżnicowaniem i silną asymetrią (tab. 1) – Wawrzyniak (2000).

Tabela 1. Podstawowe statystyki opisowe charakteryzujące rozkład badanych zmiennych

Zmienne	Średnia	Mediana	Współczynnik zmienności	Miara asymetrii
X ₁	68,705	81,700	45,245	-2,058
X ₂	0,001	0,00009	196,195	1,928
X ₃	0,598	0,026	201,064	1,253
X ₄	0,488	0,057	301,374	3,326
X ₅	0,063	0,012	221,904	1,856
X ₆	94,024	98,900	14,233	-4,700
X ₇	70,619	71,100	24,057	-2,318
X ₈	0,242	0,129	173,772	2,694
X ₉	0,061	0,041	119,370	1,221
X ₁₀	19,162	13,700	90,840	-0,240
X ₁₁	0,131	0,087	110,762	2,227

W tabeli 2 przedstawiono niepełną złożoną macierz znaczników. W macierzy tej dla każdego obiektu (powiatu) ujęto tylko wariant badanej zmiennej oznaczony symbolem X_{ip} (i = 1, 2, ..., 11). Wariant X_{in} pominięto, gdyż jest on przeciwieństwem wariantu X_{ip} – jeżeli dla danego powiatu X_{ip} = 1, to X_{in} = 0 (i odwrotnie).

Tabela 2. Niepełna złożona macierz znaczników

Powiaty	Skróty nazw powiatów	X _{1p}	X _{2p}	X _{3p}	X _{4p}	X _{5p}	X _{6p}	X _{7p}	X _{8p}	X _{9p}	X _{10p}	X _{11p}
Białogardzki	B	1	0	0	1	0	0	1	1	1	0	0
Choszczeński	CH	0	1	1	1	1	1	0	0	0	1	0
Drawski	D	0	1	1	1	1	0	0	1	1	1	1
Goleniowski	GL	0	0	1	1	0	1	0	1	1	0	0
Gryficki	GR	0	1	1	0	0	1	1	0	0	0	0
Gryfiński	GF	1	0	0	0	1	1	0	1	0	1	1
Kamieński	KM	0	1	1	1	1	1	0	0	1	0	1
Kołobrzeski	KB	1	0	0	0	0	0	1	1	0	0	1
Koszaliński	KO	0	0	1	1	1	0	0	0	1	1	1
Łobeski	L	1	1	1	1	1	0	0	0	1	0	0
Myśliborski	M	1	1	0	0	0	0	0	1	1	1	1
Policki	PL	1	0	0	0	0	1	1	1	1	0	0
Pyrzycki	PR	0	1	1	1	0	0	1	1	0	0	0
Sławiński	SL	0	1	1	0	1	1	1	0	0	1	1
Stargardzki	ST	1	1	0	0	0	1	1	0	0	0	0
Szczecinecki	SC	1	0	0	0	1	1	1	0	0	1	0
Świdwiński	SW	0	1	1	1	1	0	0	0	1	1	1
Wałecki	W	1	1	1	1	0	1	1	0	0	1	0
Koszalin	K	1	0	0	0	0	1	1	0	0	1	1
Szczecin	SZ	1	0	0	0	0	0	0	1	1	0	1
Świnoujście	S	0	0	0	0	0	0	1	1	1	1	1

Analizę korespondencji przeprowadzono według etapów omówionych w części Materiał i metody³.

W zbiorze analizowanych zmiennych, oprócz 11 zmiennych zero-jedynkowych, uwzględniono zmienną: powiaty, która miała 21 wariantów. W związku z tym wymiar rzeczywistej przestrzeni współwystępowania wyniósł 31 – wzór (1).

Następnie sprawdzono, w jakim stopniu wartości własne przestrzeni o niższym wymiarze wyjaśniają inercję całkowitą ($\lambda = 2,5833$). Zgodnie z kryterium Greenacre'a za istotne dla badania uznano inercje główne większe niż $\frac{1}{Q} = \frac{1}{12} = 0,0833$. Z tabeli 3 wynika, że są to

inercje dla K przyjmującego wartości do 11 włącznie⁴. Dla tych wymiarów przeanalizowano wartości miernika τ_k ⁵ i okazało się, że stopień wyjaśnienia inercji w przestrzeni dwuwymiarowej wynosi 25,1386%. W celu podwyższenia jakości odwzorowania w przestrzeni dwuwymiarowej przeprowadzono modyfikację wartości własnych według wzoru (3). Pierwotne i zmodyfikowane wartości własne wraz ze stopniem wyjaśnienia inercji całkowitej podano w tab. 3.

Tabela 3. Wartości osobliwe oraz wartości własne wraz ze stopniem wyjaśnienia inercji całkowitej w wersjach pierwotnej i zmodyfikowanej

K	Wartości osobliwe γ_k	Wartości własne λ_k	λ_k / λ	τ_k	$\tilde{\lambda}_k$	$\tilde{\lambda}_k / \tilde{\lambda}$	$\tilde{\tau}_k$
1	0,6045	0,3654	14,1458	14,1458	0,3233	0,2299	0,2299
2	0,5329	0,2840	10,9928	25,1386	0,2405	0,1711	0,4009
3	0,4793	0,2297	8,8916	34,0302	0,1866	0,1327	0,5336
4	0,3946	0,1557	6,0289	40,0591	0,1153	0,0820	0,6157
5	0,3750	0,1407	5,4446	45,5037	0,1013	0,0720	0,6877
6	0,3625	0,1314	5,0861	50,5898	0,0927	0,0660	0,7536
7	0,3470	0,1204	4,6609	55,2507	0,0827	0,0588	0,8125
8	0,3370	0,1136	4,3972	59,6479	0,0766	0,0545	0,8669
9	0,3242	0,1051	4,0692	63,7172	0,0691	0,0491	0,9160
10	0,3110	0,0967	3,7436	67,4608	0,0617	0,0439	0,9599
11	0,3010	0,0906	3,5070	70,9677	0,0564	0,0401	1,0000
12	0,2887	0,0833	3,2258	74,1935	$\tilde{\lambda}_k = 1,4061$		

W wyniku przeprowadzonej modyfikacji wyraźnie zwiększył się stopień wyjaśnienia inercji całkowitej. Dwie pierwsze wartości własne stanowią 40,09% zmodyfikowanej inercji całkowitej. Dlatego w graficznej prezentacji wyników wielowymiarowej analizy korespondencji w przestrzeni dwuwymiarowej uwzględniono modyfikację wartości własnych (rys. 1). Nowe wartości współrzędnych w przestrzeni dwuwymiarowej dla kategorii zmiennych zostały wyznaczone na podstawie wzoru (4).

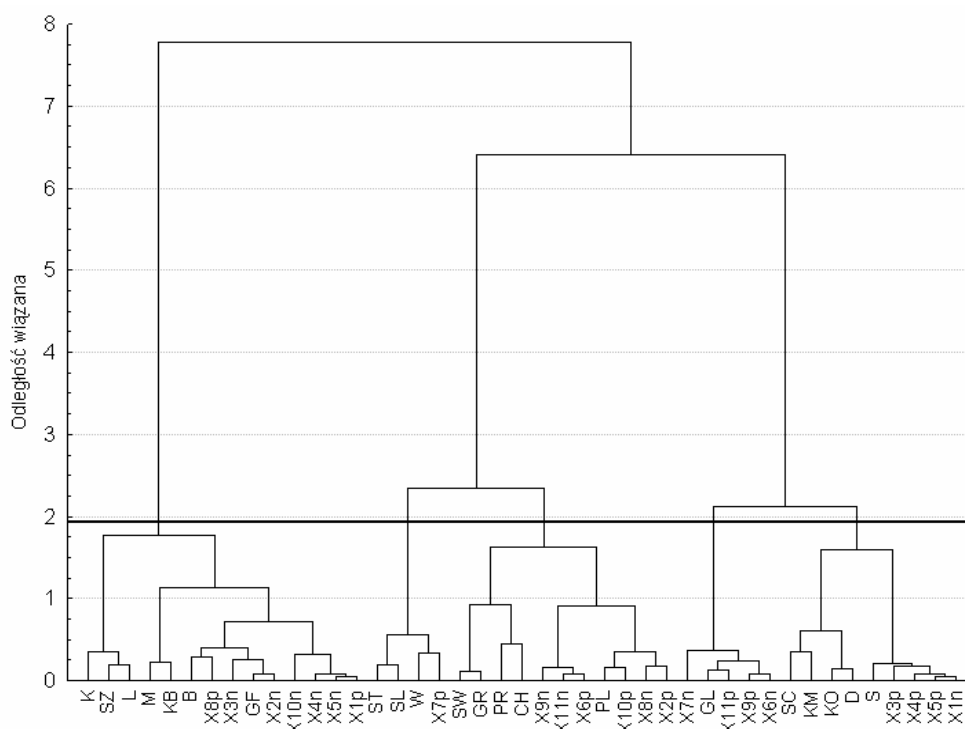
Analiza rozrzutu punktów (rys. 1) pozwala zauważyć, że niewiele kategorii zmiennych znajduje się blisko początku układu współrzędnych, przy czym najdalej od centrum rzutowania rozmieszczone są punkty obrazujące większość powiatów. Taki układ punktów świadczy o występowaniu zależności pomiędzy kategoriami.

³ Do obliczeń i graficznej prezentacji wyników wykorzystano moduł Analiza korespondencji oprogramowany w pakiecie Statistica 8,0.

⁴ W tabeli 3 pominięto wyniki dla $K > 12$, gdyż dla tych wymiarów inercje główne były nie wyższe niż 0,0833, a więc te wymiary były nieistotne w badaniu.

⁵ Miernik ten określa udział inercji wybranego wymiaru (λ_k) w inercji całkowitej (λ).

odchylenia wartości analizowanych zmiennych *in plus* i *in minus* od przeciętnej w województwie. Brak w danej klasie informacji o innych zmiennych oznacza, że ich poziom nie odbiegał w sposób istotny od przeciętnej w województwie.



Rys. 2. Diagram hierarchicznej klasyfikacji kategorii zmiennych wykonanej metodą Warda

Tabela 4. Charakterystyka stanu środowiska naturalnego w powiatach województwa zachodniopomorskiego

Klasy	Powiaty	Charakterystyka stanu środowiska naturalnego
I	Koszalin, Szczecin, łobeski, myśliborski, kołobrzeski, białogardzki, gryfiński	Obiekty z tej klasy charakteryzowały się redukcją zanieczyszczeń pyłowych i nakładami na ochronę środowiska powyżej przeciętnej. Natomiast emisja zanieczyszczeń pyłowych i gazowych powietrza oraz odpady wytworzone i wywiezione w ciągu roku były poniżej przeciętnej w województwie. Atrakcyjność przyrodniczą, mierzoną powierzchnią o szczególnych walorach przyrodniczych, oceniono negatywnie
II	stargardzki, sławieński, wałecki	Powiaty te jedynie ze względu na odsetek ludności obsługiwanej przez oczyszczalnie ścieków wyraźnie odchyłały się <i>in plus</i> od przeciętnej w województwie
III	świdwiński, gryficki, pyrzycki, choszczeński, policki	W tej klasie emisja zanieczyszczeń pyłowych powietrza została oceniona negatywnie. Odsetek oczyszczonych ścieków oceniono pozytywnie. Nakłady inwestycyjne na ochronę środowiska i gospodarkę wodną kształtowały się poniżej przeciętnej. Powiaty charakteryzowały się powierzchnią o szczególnych walorach przyrodniczych powyżej przeciętnej, a liczba pomników przyrody odchyłała się <i>in minus</i> od przeciętnej w województwie
IV	goleniowski	Powiat ten w wyraźny sposób odchyłał się <i>in minus</i> od przeciętnej w województwie pod względem odsetka ludności obsługiwanej przez oczyszczalnie ścieków oraz odsetka oczyszczonych ścieków, pomimo że nakłady inwestycyjne na gospodarkę wodną były powyżej przeciętnej. Liczba pomników przyrody w powiecie była powyżej przeciętnej w województwie
V	szczecinecki, kamieński, koszaliński, drawski, Świnoujście	W tych powiatach emisja zanieczyszczeń gazowych powietrza, odpady wytworzone i wywiezione w ciągu roku kształtowały się poniżej przeciętnej w województwie, czyli zostały ocenione pozytywnie. Natomiast redukcja zanieczyszczeń pyłowych powietrza została oceniona negatywnie

Źródło: opracowanie własne na podstawie rys. 2.

PODSUMOWANIE

Z przeprowadzonych badań wynika, że powiaty województwa zachodniopomorskiego są bardzo zróżnicowane pod względem jakości środowiska naturalnego. Nie wyodrębniono grup powiatów, które odbiegałyby zdecydowanie *in plus* lub *in minus* od przeciętnych wszystkich badanych zmiennych w województwie. Na uwagę zasługują powiaty zakwalifikowane do grupy drugiej, które tylko ze względu na jedną zmienną (odsetek ludności obsługiwanej przez oczyszczalnie ścieków) wyraźnie odbiegały *in plus* od średniej wartości tej zmiennej w województwie. Najwięcej odchyień od wartości przeciętnej w województwie można zaobserwować w powiatach należących do grup pierwszej oraz trzeciej. Powiaty z grupy pierwszej oceniono pozytywnie pod względem aż sześciu zmiennych. Tylko atrakcyjność przyrodnicza, mierzona powierzchnią o szczególnych walorach przyrodniczych, została oceniona negatywnie. W grupie trzeciej sześć zmiennych zdecydowanie odbiegało od mediany (dwie zmienne *in plus*, a cztery zmienne *in minus*). W pozostałych grupach powiatów odchylenia od przeciętnej dotyczyły tylko czterech różnych zmiennych.

Wydzielenie grup powiatów, charakteryzujących się jednakowym zestawem zmiennych odbiegających w sposób wyraźny od przeciętnych wartości w województwie, było możliwe dzięki zastosowaniu do ostatecznej interpretacji wyników zarówno wielowymiarowej analizy korespondencji, jak i metody Warda.

PIŚMIENNICTWO

- Malina A.** 2004. Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw. Kraków, Wydawnictwo AE.
- Metody statystycznej analizy wielowymiarowej w badaniach marketingowych.** 2004. Red. E. Gatnar, M. Walesiak. Wrocław, Wydawnictwo AE.
- Ochrona środowiska.** 2007, Warszawa, GUS.
- Stanimir A.** 2005. Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych, Wrocław Wydawnictwo AE.
- Wawrzyniak K.** 2000. Klasyczne i pozycyjne parametry struktury jako normy w procesie oceny działalności przedsiębiorstwa. Zesz. Nauk. USzczec. 269, 269–282.