

**Karol DERĘGOWSKI<sup>1</sup>**  
**Mirosław KRZYŚKO<sup>1</sup>**  
**Waldemar WOŁYŃSKI<sup>2</sup>**

## Podobieństwo struktury wybranych wydatków budżetowych województw a ich zrównoważony rozwój: nowe narzędzia badawcze

### 1. WSTĘP

Inspiracją do przygotowania tego opracowania był artykuł Salamagi (2017). Autor tego artykułu w streszczeniu napisał: „Celem opracowania jest opisanie zależności pomiędzy strukturą wydatków budżetowych województw a ich zrównoważonym rozwojem”. Autorzy tej pracy stawiają sobie ten sam cel, ale proponują odmienne metody statystyczne wykorzystane do osiągnięcia tego celu. Schölkopf i inni (1998) zaproponowali jądrową analizę składowych głównych w przypadku, gdy badane obiekty charakteryzowane są za pomocą wektorów cech obserwowanych w ustalonym momencie czasu. W pracy tej podajemy rozszerzenie jądrowej analizy składowych głównych na przypadek, gdy obiekty charakteryzowane są za pomocą macierzy  $\mathbf{X}$  rozmiaru  $T \times p$ , gdzie  $p$  jest liczbą cech obserwowanych w  $T$  momentach czasowych (patrz sekcja 3). Podział województw na grupy w układzie dwóch pierwszych jądrowych składowych głównych dokonano metodą składowych głównych (patrz sekcja 5). Wyłonienie wskaźników zrównoważonego rozwoju najbardziej charakterystycznych dla wyróżnionych skupień odbywa się przy użyciu odległości Mahalanobisa (patrz sekcja 4). Ponieważ województwa charakteryzowane są za pomocą danych macierzowych, to użyte przez nas techniki statystyczne wydają się być odpowiednie.

---

<sup>1</sup> Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego, Międzywydziałowy Zakład Matematyki i Statystyki, ul. Nowy Świat 4, 62-030 Kalisz, e-mail: kadere@o2.pl; mkrzysko@amu.edu.pl.

<sup>2</sup> Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Matematyki i Informatyki, ul. Umultowska 87, 61-614 Poznań, e-mail: wolynski@amu.edu.pl.

## 2. DANE BADAWCZE

Zgodnie z informacją zawartą w Banku Danych Lokalnych GUS wydatki budżetowe pogrupowano w 33 działach oznaczonych odpowiednimi symbolami. Spośród tych działów wybranych zostało 9 następujących:

1. Dział 801 — Oświata i wychowanie,
2. Dział 851 — Ochrona zdrowia,
3. Dział 852 — Pomoc społeczna,
4. Dział 853 — Pozostałe zadania w zakresie polityki społecznej,
5. Dział 854 — Edukacyjna opieka wychowawcza,
6. Dział 900 — Gospodarka komunalna i ochrona środowiska,
7. Dział 921 — Kultura i ochrona dziedzictwa narodowego,
8. Dział 925 — Ogrody botaniczne i zoologiczne oraz naturalne obszary i obiekty chronionej przyrody,
9. Dział 926 — Kultura fizyczna i sport.

Dane badawcze, to 16 macierzy rozmiaru  $8 \times 9$  zawierających wartości wydatków budżetowych w 9 działach dla 16 województw notowane w latach 2009–2016 (9 lat). Wybór działów oraz lat podyktowany był dostępnością danych.

Za cechy charakteryzujące rozwój zrównoważony województw przyjęte zostały następujące cechy, tożsame z cechami użytymi przez Salamagę (2017):

- $Y_1$  — PKB na mieszkańca w zł,  
 $Y_2$  — nakłady inwestycyjne na mieszkańca w zł,  
 $Y_3$  — udział przedsiębiorstw innowacyjnych w ogóle przedsiębiorstw w %,  
 $Y_4$  — nakłady na działalność badawczo – rozwojową w relacji do PKB,  
 $Y_5$  — przeciętny miesięczny dochód rozporządzalny na osobę w gospodarstwie domowym w zł,  
 $Y_6$  — udział osób w gospodarstwach domowych korzystających ze środowiskowej pomocy społecznej w ludności ogółem w %,  
 $Y_7$  — stopa bezrobocia rejestrowanego w %,  
 $Y_8$  — długość dróg ekspresowych i autostrad na 100 km<sup>2</sup>,  
 $Y_9$  — liczba samochodów osobowych na 1000 ludności,  
 $Y_{10}$  — przeciętne dalsze trwanie życia osób w wieku 65 lat w latach,  
 $Y_{11}$  — emisja zanieczyszczeń pyłowych w t/rok.

## 3. JĄDROWE SKŁADOWE GŁÓWNE

Niech zbiorem danych będzie zbiór  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  macierzy z przestrzeni  $\mathfrak{R}^{T \times p}$ , gdzie  $p$  jest liczbą cech obserwowanych w  $T$  momentach czasu.

Funkcja

$$k: \mathfrak{R}^{T \times p} \times \mathfrak{R}^{T \times p} \rightarrow \mathfrak{R}$$

nazywa się jądrem na przestrzeni  $\mathfrak{R}^{T \times p}$ , jeżeli istnieje przestrzeń Hilberta wyposażona w iloczyn skalarny  $\langle \bullet, \bullet \rangle_{\mathfrak{H}}$  oraz odwzorowanie

$$\varphi : \mathfrak{R}^{T \times p} \rightarrow \mathfrak{Hf}$$

takie, że dla wszystkich  $\mathbf{X}_i, \mathbf{X}_j \in \mathfrak{R}^{T \times p}$  zachodzi

$$k(\mathbf{X}_i, \mathbf{X}_j) = \langle \varphi(\mathbf{X}_i), \varphi(\mathbf{X}_j) \rangle_{\mathfrak{Hf}}.$$

Funkcję  $\varphi$  nazywamy odwzorowaniem cechy, a  $\mathfrak{Hf}$  przestrzenią cech odwzorowanych.

Funkcję jądrową możemy interpretować jako miarę podobieństwa między macierzami  $\mathbf{X}_i$  i  $\mathbf{X}_j$ . Mówimy, że funkcja jądrowa  $k$  jest nieujemnie określona, jeżeli dla każdego zbioru  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  macierzy z przestrzeni  $\mathfrak{R}^{T \times p}$  i dla wszystkich liczb rzeczywistych  $c_1, \dots, c_n$

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j k(\mathbf{X}_i, \mathbf{X}_j) \geq 0.$$

Pojawia się interesujące pytanie: czy dla danej funkcji dwóch argumentów  $k(\mathbf{X}_i, \mathbf{X}_j)$  istnieje funkcja  $\varphi$  taka, że

$$k(\mathbf{X}_i, \mathbf{X}_j) = \langle \varphi(\mathbf{X}_i), \varphi(\mathbf{X}_j) \rangle_{\mathfrak{Hf}}?$$

Odpowiedź daje twierdzenie Mercera (Mercer, 1909), które z grubsza mówiąc orzeka, że jeżeli  $k$  jest jądrem nieujemnie określonym to taka funkcja  $\varphi$  istnieje.

W naszej pracy posługiwać się będziemy gaussowską funkcją jądrową postaci

$$k(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\gamma \|\mathbf{X}_i - \mathbf{X}_j\|_F^2\right), \gamma > 0,$$

gdzie

$$\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}'\mathbf{A})}$$

jest macierzową normą Frobeniusa.

Stała  $\gamma > 0$  jest odpowiednio dobrana do danych  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , gdzie  $\mathbf{X}_i \in \mathfrak{R}^{T \times p}$ ,  $i = 1, \dots, n$ .

Bierzemy pod uwagę dolny trójkąt macierzy symetrycznej, której element  $(i, j)$  jest równy  $\|\mathbf{X}_i - \mathbf{X}_j\|_F^2$ ,  $i, j = 1, \dots, n$ . Za  $\gamma$  wybieramy odwrotność wartości średniej elementów tego dolnego trójkąta.

Dla danego zbioru  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  macierzy z przestrzeni  $\mathfrak{R}^{T \times p}$  i danej funkcji jądrowej określonej na przestrzeni  $\mathfrak{R}^{T \times p}$ , macierz  $\mathbf{K} = (K_{ij})$  stopnia  $n$ , gdzie  $K_{ij} = k(\mathbf{X}_i, \mathbf{X}_j)$ , nazywa się macierzą jądrową lub macierzą Grama funkcji jądrowej  $k$  ze względu na zbiór  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ .

Niech  $\mathbf{X}_i \in \mathfrak{R}^{T \times p}, i = 1, \dots, n$ , będzie zbiorem danych, gdzie  $p$  jest liczbą cech obserwowanych w  $T$  momentach czasu. Nasze dane odwzorowujemy nieliniowo do przestrzeni Hilberta  $\mathcal{H}$  za pomocą funkcji

$$\varphi: \mathfrak{R}^{T \times p} \rightarrow \mathcal{H}.$$

Zauważmy, że elementy przestrzeni  $\mathcal{H}$  mogą być nieskończenie wymiarowe.

W celu zbudowania liniowych składowych głównych przestrzeni  $\mathcal{H}$  musimy znaleźć wartości własne  $\lambda$  i wektory własne  $\mathbf{u} \in \mathcal{H}$  operatora kowariancji z próby w przestrzeni  $\mathcal{H}$ :

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n \langle \Psi(\mathbf{X}_i), \Psi(\mathbf{X}_i) \rangle_{\mathcal{H}} \tag{1}$$

spełniającego równość

$$\mathbf{C}\mathbf{u} = \lambda\mathbf{u}, \tag{2}$$

gdzie  $\langle \mathbf{u}, \mathbf{u} \rangle_{\mathcal{H}} = 1$  oraz

$$\Psi(\mathbf{X}_i) = \varphi(\mathbf{X}_i) - \bar{\varphi}, \bar{\varphi} = \frac{1}{n} \sum_{i=1}^n \varphi(\mathbf{X}_i), i = 1, \dots, n.$$

Zauważmy, że równanie (2) możemy zapisać w postaci

$$\frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{X}_i) \langle \Psi(\mathbf{X}_i), \mathbf{u} \rangle_{\mathcal{H}} = \lambda\mathbf{u}. \tag{3}$$

Oznacza to, że wektory  $\mathbf{u}$  leżą w podprzestrzeni  $\text{span}\{\Psi(\mathbf{X}_1), \dots, \Psi(\mathbf{X}_n)\}$ , tj. leżą w powłoce liniowej przestrzeni generowanej przez zbiór  $\{\Psi(\mathbf{X}_1), \dots, \Psi(\mathbf{X}_n)\}$ , a więc istnieją współczynniki rzeczywiste  $\alpha_1, \dots, \alpha_n$  takie, że

$$\mathbf{u} = \sum_{i=1}^n \alpha_i \Psi(\mathbf{X}_i). \tag{4}$$

Zatem równanie (3) przyjmuje postać

$$\Psi(\mathbf{X}_i) \mathbf{C}\mathbf{u} \rangle_{\mathcal{H}} = \lambda \langle \Psi(\mathbf{X}_i), \mathbf{u} \rangle_{\mathcal{H}}. \tag{5}$$

Łącząc równania (1), (4) oraz (5) otrzymujemy:

$$\tilde{\mathbf{K}} \tilde{\mathbf{K}} \alpha = n\lambda \tilde{\mathbf{K}} \alpha \tag{6}$$

gdzie  $\alpha = (\alpha_1, \dots, \alpha_n)^T, \mathbf{K} = (k(\mathbf{X}_i, \mathbf{X}_j)), \tilde{\mathbf{K}} = \mathbf{H}\mathbf{K}\mathbf{H}, \mathbf{H} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T, \mathbf{1}_n \in \mathfrak{R}^{n \times 1}$ .

Idempotentna macierz  $\mathbf{H}$  ( $\mathbf{H}^2 = \mathbf{H}$ ) nazywa się macierzą centrującą.  
Zauważmy, że każde rozwiązanie równania

$$\tilde{\mathbf{K}} \alpha = n\lambda \alpha, \quad (7)$$

jest zarazem rozwiązaniem równania (6).

Wektory własne  $\mathbf{u}$  w przestrzeni  $\mathcal{H}$  normujemy tak, by miały jednostkową długość:

$$\langle \mathbf{u}, \mathbf{u} \rangle_{\mathcal{H}} = \sum_{i,j=1}^n \alpha_i \alpha_j \langle \Psi(\mathbf{X}_i), \Psi(\mathbf{X}_j) \rangle_{\mathcal{H}} = n\lambda \sum_{i=1}^n \alpha_i \alpha_i = 1.$$

Zatem wektory  $\alpha_i$  należy znormalizować tak, by ich długość była równa  $\frac{1}{\sqrt{n\lambda_i}}$ ,  $i = 1, \dots, n$ .

Stąd rzut odwzorowanej obserwacji  $\varphi(\mathbf{X}_i) \in \mathcal{H}$  na  $j$ -tą jądrową składową główną ma postać:  $\langle \Psi(\mathbf{X}_i), \mathbf{u}_j \rangle_{\mathcal{H}} = \sum_{k=1}^n \alpha_k^j \langle \Psi(\mathbf{X}_i), \Psi(\mathbf{X}_k) \rangle_{\mathcal{H}} = \sum_{k=1}^n \tilde{K}_{ik} \alpha_k^j$ ,  $i, j = 1, \dots, n$ .

Podsumujemy teraz kolejne kroki konstrukcji jądrowych składowych głównych:

1. Danymi są macierze  $\mathbf{X}_i \in \mathbb{R}^{T \times p}$ ,  $i = 1, \dots, n$ .
2. Na bazie danych budujemy macierz jądrową  $\mathbf{K} = (K_{ij})$ , której elementami są wartości  $K_{ij} = k(\mathbf{X}_i, \mathbf{X}_j)$  gaussowskiej funkcji jądrowej  $k$ .
3. Tworzymy macierz jądrową scentrowaną  $\tilde{\mathbf{K}} = \mathbf{H}\mathbf{K}\mathbf{H}$ , gdzie  $\mathbf{H}$  jest idempotentną macierzą centrującą postaci  $\mathbf{H} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$ .
4. Wyliczamy wartości własne  $\lambda_i$  oraz wektory własne  $\alpha_i$  macierzy  $\tilde{\mathbf{K}}$ ,  $i = 1, \dots, n$ .
5. Normalizujemy wektory  $\alpha_i$  tak, aby ich długość była równa  $\frac{1}{\sqrt{n\lambda_i}}$ , co zapewnia, że

$$\langle \mathbf{u}, \mathbf{u} \rangle_{\mathcal{H}} = \sum_{i,j=1}^n \alpha_i \alpha_j \langle \Psi(\mathbf{X}_i), \Psi(\mathbf{X}_j) \rangle_{\mathcal{H}} = n\lambda \sum_{i=1}^n \alpha_i \alpha_i = 1.$$

6. Rzut odwzorowanej obserwacji  $\varphi(\mathbf{X}_i) \in \mathcal{H}$  na  $j$ -tą jądrową składową główną jest równy:

$$U_{ij} = \langle \Psi(\mathbf{X}_i), \mathbf{u}_j \rangle_{\mathcal{H}} = \sum_{k=1}^n \tilde{K}_{ik} \alpha_k^j, i, j = 1, \dots, n.$$

#### 4. CECHY ROZWOJU ZRÓWNOWAŻONEGO NAJBARDZIEJ CHARAKTERYSTYCZNE DLA WYRÓŻNIONYCH SKUPIEŃ

Każda z 11 cech zrównoważonego rozwoju obserwowana jest w każdym z 16 województw. Niech  $\mathbf{y}_{ij}$  będzie 8-wymiarowym wektorem złożonym z wartości  $i$ -tej cechy zrównoważonego rozwoju obserwowanej w  $j$ -tym województwie w latach 2009–2016,  $i = 1, \dots, 11$ ,  $j = 1, \dots, 16$ . Z 16 wektorów  $\mathbf{y}_{i1}, \dots, \mathbf{y}_{i16}$  wyliczamy wektor średni dla  $i$ -tej cechy zrównoważonego rozwoju

$$\bar{y}_i = \frac{1}{16} \sum_{j=1}^{16} y_{ij}$$

oraz nieobciążoną ocenę macierzy kowariancji

$$S_i = \frac{1}{15} \sum_{j=1}^{16} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)'$$

Macierz  $S_i$  jest miarą zmienności wektorów  $y_{ij}$ .

Ponadto w  $k$ -tym skupieniu wyliczamy wewnątrzgrupowy wektor średni oraz kwadrat odległości Mahalanobisa:

$$d_{ik}^2 = (y_{ik} - \bar{y}_i)' S_i^{-1} (y_{ik} - \bar{y}_i), i = 1, \dots, 11, k = 1, \dots, K. \quad (8)$$

## 5. WYNIKI BADAŃ EMPIRYCZNYCH

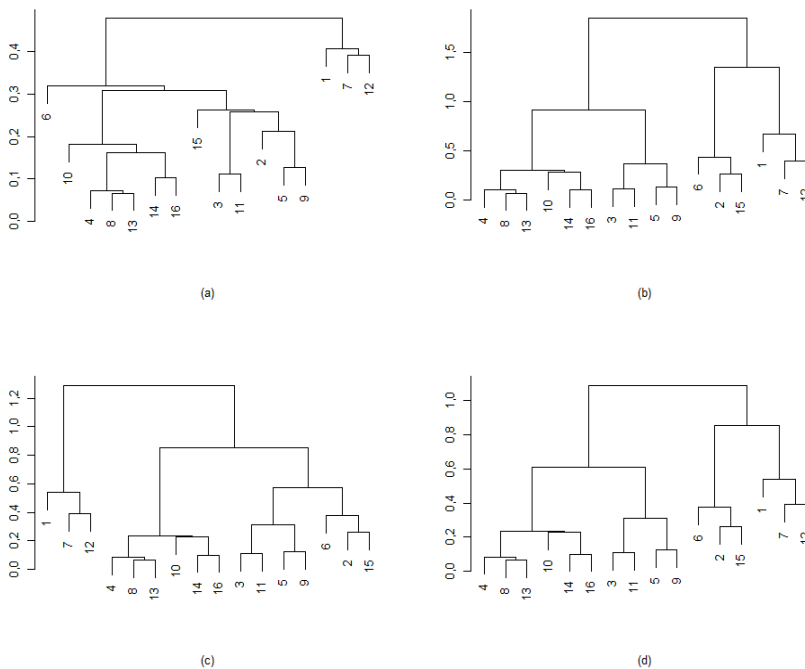
Analizą objęto 16 województw ( $n = 16$ ). Analizowane dane dotyczą lat 2009–2016 ( $T = 8$ ). Każde województwo zostało scharakteryzowane wartościami wydatków budżetowych w  $p = 9$  wybranych działach oraz wartościami 11 cech opisujących zrównoważony rozwój. Dane pochodzą z Lokalnego Banku Danych GUS (<https://bdl.stat.gov.pl>). Najpierw skonstruowane zostały jądrowe składowe główne. Są to składowe nieliniowe, bo chociaż w przestrzeni Hilberta zostały skonstruowane składowe główne liniowe, to dane pierwotne z przestrzeni macierzy  $X$  rozmiaru  $T \times p$  zostały odwzorowane za pomocą nieliniowej funkcji  $\varphi$ . Wartości własne i rzut na dwie pierwsze jądrowe składowe główne zawiera tabela 1.

Tabela 1. WARTOŚCI WŁASNE I RZUT NA DWIE PIERWSZE JĄDROWE SKŁADOWE GŁÓWNE

Lp.	Województwo	Wartość własna	$U_{i1}$	$U_{i2}$
1	Dolnośląskie .....	2,4048	0,7022	-0,0571
2	Kujawsko-pomorskie .....	1,1076	0,3125	0,2224
3	Lubelskie .....	0,9894	-0,1819	0,1938
4	Lubuskie .....	0,7757	-0,6505	-0,2179
5	Łódzkie .....	0,5344	0,0347	0,0304
6	Małopolskie .....	0,4567	0,5938	0,5550
7	Mazowieckie .....	0,3972	1,0648	-0,2437
8	Opolskie .....	0,3610	-0,6703	-0,1489
9	Podkarpackie .....	0,2626	0,1492	0,0860
10	Podlaskie .....	0,2138	-0,4578	-0,1054
11	Pomorskie .....	0,1991	-0,2180	0,0879
12	Śląskie .....	0,1601	1,0450	-0,6350
13	Świętokrzyskie .....	0,1327	-0,7347	-0,1662
14	Warmińsko-mazurskie .....	0,0705	-0,6633	-0,0844
15	Wielkopolskie .....	0,0512	0,2815	0,4840
16	Zachodniopomorskie .....	0,0000	-0,6071	-0,0008

Wykorzystując współrzędne rzutu na dwie pierwsze jądrowe składowe główne, dokonano podziału województw używając metod analizy skupień. Wybrana została hierarchiczna metoda aglomeracyjna. Ponadto, wykorzystano cztery różne metody wiązania skupień: metodę pojedynczego wiązania, metodę pełnego wiązania, metodę średniego wiązania oraz metodę sumy kwadratów odległości. Na rysunku 1 przedstawiono dendrogramy uzyskane przy użyciu tych metod.

Rysunek 1. Metody aglomeracji: (a) – metoda pojedynczego wiązania, (b) – metoda pełnego wiązania, (c) – metoda średniego wiązania, (d) – metoda sumy kwadratów odległości



Źródło: opracowanie własne.

Następnie korzystając ze wskaźnika zarysu, dla każdej z metod, wyznaczono optymalną liczbę skupień. Wyniki zawiera tabela 2.

Tabela 2. OPTYMALNA LICZBA SKUPIEŃ

Metoda aglomeracji	Optymalna liczba skupień
Metoda pojedynczego wiązania .....	2
Metoda pełnego wiązania .....	4
Metoda średniego wiązania .....	3
Metoda sumy kwadratów odległości	4

Źródło: opracowanie własne.

Powstałe w ten sposób skupienia mają następujący skład:

1. Metoda pojedynczego wiązania:

Skupienie I: dolnośląskie, mazowieckie, śląskie.

Skupienie II: pozostałe województwa.

2. Metoda średniego wiązania:

Skupienie I: dolnośląskie, mazowieckie, śląskie.

Skupienie II: kujawsko-pomorskie, małopolskie, wielkopolskie.

Skupienie III: pozostałe województwa.

3. Metoda pełnego wiązania oraz metoda sumy kwadratów odległości:

Skupienie I: dolnośląskie, mazowieckie, śląskie.

Skupienie II: kujawsko-pomorskie, małopolskie, wielkopolskie.

Skupienie III: lubelskie, łódzkie, podkarpackie, pomorskie.

Skupienie IV: pozostałe województwa.

Dalej za podział optymalny przyjmujemy podział na 4 jednorodne skupienia. Podział ten przedstawiony jest na rysunku 2. Następnie oceniono, które spośród 11 wskaźników opisujących najważniejsze aspekty rozwoju zrównoważonego są najbardziej charakterystyczne dla wyróżnionych skupień województw. Jako miarę bliskości wybrano kwadrat odległości Mahalanobisa danej wzorem (8). Odpowiednie wartości zawiera tabela 3.

Rysunek 2. Skupienia wyznaczone zgodnie z metodami pełnego wiązania oraz sumy kwadratów odległości (podział optymalny)

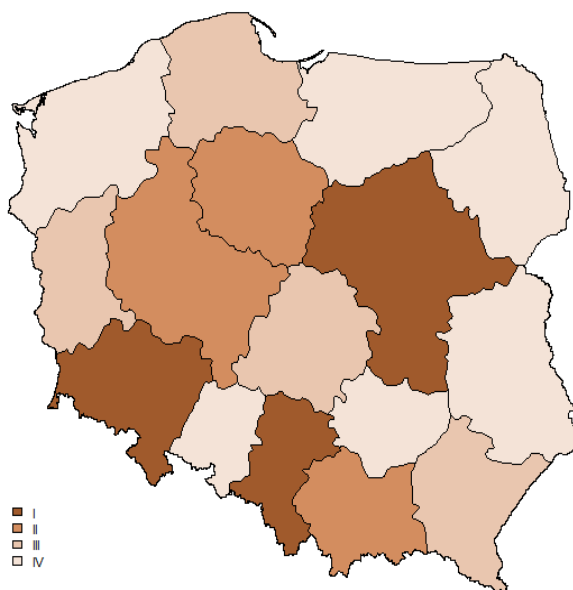




Tabela 3. WARTOŚCI KWADRATÓW ODLEGŁOŚCI MAHALANOBISA DLA 11 WSKAŹNIKÓW ROZWOJU ZRÓWNOWAŻONEGO W CZTERECH OPTYMALNYCH SKUPIENIACH WOJEWÓDZTW

Nr wskaźnika	I	II	III	IV
1 .....	3,7408	1,3836	1,9583	1,0551
2 .....	2,6116	<b>0,7779</b>	1,5462	<b>0,7550</b>
3 .....	2,6836	3,1738	<b>1,5193</b>	<b>0,6170</b>
4 .....	<b>1,5616</b>	1,9270	1,7441	1,2269
5 .....	2,4790	1,5933	1,6222	0,8499
6 .....	2,0197	1,3678	1,6128	1,1425
7 .....	<b>1,6701</b>	<b>0,8771</b>	1,8463	1,1680
8 .....	2,4257	2,3689	1,6410	1,1397
9 .....	<b>0,2789</b>	2,3107	1,4307	<b>0,2955</b>
10 .....	2,7193	<b>1,1381</b>	2,0041	0,8466
11 .....	3,7146	2,6849	<b>1,3009</b>	0,7956

Źródło: opracowanie własne.

Analizując tabelę 3 widzimy, że dla skupienia I złożonego z województw dolnośląskiego, mazowieckiego i śląskiego najbardziej charakterystycznymi wektorowymi wskaźnikami opisującymi zrównoważony rozwój w latach 2009–2016 łącznie są:  $Y_9$  – liczba samochodów osobowych na 1000 ludności,  $Y_4$  – nakłady na działalność badawczo-rozwojową w relacji do PKB oraz  $Y_7$  – stopa bezrobocia rejestrowanego w %. Dla skupienia II najbardziej charakterystycznymi wskaźnikami wektorowymi są:  $Y_2$  – nakłady inwestycyjne na mieszkańca w zł.,  $Y_7$  – stopa bezrobocia rejestrowanego w % oraz  $Y_{10}$  – przeciętne dalsze trwanie życia osób w wieku 65 lat mierzone w latach. Dla skupienia III są to:  $Y_{11}$  – emisja zanieczyszczeń pyłowych w t/rok,  $Y_9$  – liczba samochodów osobowych na 1000 ludności oraz  $Y_3$  – udział przedsiębiorstw innowacyjnych w ogóle przedsiębiorstw w %. Dla skupienia IV są to:  $Y_9$  – liczba samochodów osobowych na 1000 ludności,  $Y_3$  – udział przedsiębiorstw innowacyjnych w ogóle przedsiębiorstw w % oraz  $Y_2$  – nakłady inwestycyjne na mieszkańca w zł.

Wartości danego wskaźnika zrównoważonego rozwoju obserwowane w latach 2009–2016 tworzą ośmiowymiarowy wektor. Tabela 4 zawiera długości tych wektorów. Im wartości danego wskaźnika obserwowane w latach 2009–2016 są większe, tym długość tego wektora jest większa.

Tabela 4. DŁUGOŚCI WEKTORÓW REPREZENTUJĄCYCH WARTOŚCI 11 WSKAŹNIKÓW W LATACH 2009–2016 DLA POSZCZEGÓLNYCH WOJEWÓDZTW

Województwo	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$
Dolnośląskie .....	135821,16	20144,33	43,50	1,88	3845,28	17,45
Kujawsko-pomorskie ...	98504,35	15172,89	37,79	1,00	3311,10	30,64
Lubelskie .....	84521,11	12877,18	42,53	2,40	3134,83	26,71
Lubuskie .....	100776,58	16610,64	35,34	0,58	3595,02	26,44
Łódzkie .....	112523,12	18500,79	34,92	1,92	3608,96	21,91
Małopolskie .....	107055,83	15487,23	42,15	3,61	3448,58	18,54
Mazowieckie .....	192927,14	26761,68	45,14	4,33	4712,73	19,26
Opolskie .....	97768,50	15837,88	45,05	0,73	3473,79	17,28

Tabela 4. DŁUGOŚCI WEKTORÓW REPREZENTUJĄCYCH WARTOŚCI 11 WSKAŹNIKÓW W LATACH 2009–2016 DLA POSZCZEGÓLNYCH WOJEWÓDZTW (dok.)

Województwo	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>	Y <sub>6</sub>
Podkarpackie .....	85127,42	15355,64	43,18	3,42	2807,05	30,82
Podlaskie .....	87443,52	14383,48	42,38	1,45	3473,38	27,61
Pomorskie .....	116027,36	18724,07	37,16	2,78	3830,52	23,43
Śląskie .....	126751,49	17477,24	41,70	1,64	3698,04	15,06
Świętokrzyskie .....	89262,98	13137,97	34,08	1,00	3199,73	30,35
Warmińsko-mazurskie	86392,09	13849,83	35,83	1,09	3330,44	37,56
Wielkopolskie .....	128663,68	17513,54	37,13	1,92	3431,39	20,63
Zachodniopomorskie	101069,37	16422,69	40,37	0,89	3673,73	26,21

Źródło: opracowanie własne.

Tabela 4. DŁUGOŚCI WEKTORÓW REPREZENTUJĄCYCH WARTOŚCI 11 WSKAŹNIKÓW W LATACH 2009–2016 DLA POSZCZEGÓLNYCH WOJEWÓDZTW (KONT.)

Województwo	Y <sub>7</sub>	Y <sub>8</sub>	Y <sub>9</sub>	Y <sub>10</sub>	Y <sub>11</sub>
Dolnośląskie .....	20,28	37,09	1434,54	49,11	11188,59
Kujawsko-pomorskie .....	27,92	24,55	1381,21	48,92	10363,00
Lubelskie .....	24,79	6,04	2802,92	49,86	6460,00
Lubuskie .....	22,85	35,78	1469,00	48,57	3270,01
Łódzkie .....	23,25	46,35	1415,80	48,04	9853,41
Małopolskie .....	18,54	27,16	1342,67	51,10	9956,41
Mazowieckie .....	20,60	16,22	1550,11	50,88	12726,43
Opolskie .....	19,54	26,46	1522,98	49,19	6186,08
Podkarpackie .....	29,12	14,58	1250,76	51,15	4799,10
Podlaskie .....	23,18	4,47	1216,27	51,25	2731,72
Pomorskie .....	19,23	20,08	1399,19	50,01	7260,44
Śląskie .....	16,90	61,71	1374,66	48,58	31524,33
Świętokrzyskie .....	27,86	12,27	1312,43	49,95	7165,45
Warmińsko-mazurskie .....	30,78	13,85	1256,93	48,74	3158,91
Wielkopolskie .....	16,02	29,09	1576,95	49,12	13851,71
Zachodniopomorskie .....	25,42	18,51	1318,15	48,93	8056,73

Źródło: opracowanie własne.

Następnie długości wektorów dla ustalonego wskaźnika nadano rangi poprzez wszystkie województwa. Rangi przypisane województwom znajdującym się w danym skupieniu oraz trzem wskaźnikom najbardziej charakterystycznym dla tego skupienia zostały uśrednione. Średnią wartość rang dla poszczególnych skupień zawiera tabela 5.

Tabela 5. ŚREDNIE WARTOŚCI RANG PRZYPISANE WOJEWÓDZTWOM Z DANEGO SKUPIENIA ORAZ TRZEM WSKAŹNIKOM ROZWOJU ZRÓWNOWAŻONEGO NAJBARDZIEJ CHARAKTERYSTYCZNYM DLA DANEGO SKUPIENIA

I	II	III	IV
17,(3)	22,(3)	23,5	31,8(3)

Źródło: opracowanie własne.

Z tabeli 5 wynika, że pod względem społeczno-gospodarczym najlepiej rozwinięte są województwa tworzące skupienie nr I, a najsłabiej województwa tworzące skupienie nr IV.

W celu stwierdzenia, na ile spójny jest rozwój społeczno-gospodarczy w województwach o podobnej strukturze wydatków budżetowych wyliczono wielowymiarowy współczynnik zmienności  $CV_m$  dla zmiennych  $Y_1$ – $Y_{11}$ , zarówno w każdym wyróżnionym skupieniu województw, jak i dla wszystkich ogółem.

Wielowymiarowy współczynnik zmienności wyliczamy według wzoru (Albert, Zhang, 2010):

$$CV_m = [(\bar{y}'S\bar{y})/(\bar{y}'\bar{y})^2]^{1/2},$$

gdzie  $\bar{y}$  jest 8-wymiarowym wektorem średnich dla lat 2009–2016, a  $S$  jest macierzą kowariancji.

Im mniejsza jest wartość wielowymiarowego współczynnika zmienności, tym bardziej podobny jest rozwój województw mierzony odpowiednim wskaźnikiem. Wartości wielowymiarowego współczynnika zmienności zamieszczone są w tabelach 6–7.

Tabela 6. WARTOŚCI WIELOWYMIAROWEGO WSPÓŁCZYNNIKA ZMIENNOŚCI DLA 11 WSKAŹNIKÓW BEZ PODZIAŁU WOJEWÓDZTW NA SKUPIENIA

Nr wskaźnika	$CV_m$
1 .....	0,2523
2 .....	0,2019
3 .....	0,0960
4 .....	0,5935
5 .....	0,1164
6 .....	0,2553
7 .....	0,1956
8 .....	0,6232
9 .....	0,2501
10 .....	0,0210
11 .....	0,7350

Źródło: opracowanie własne.

Tabela 7. WARTOŚCI WIELOWYMIAROWEGO WSPÓŁCZYNNIKA ZMIENNOŚCI DLA 11 WSKAŹNIKÓW W CZTERECH SKUPIENIACH WOJEWÓDZTW

Nr wskaźnika	I	II	III	IV
1 .....	0,2363	0,1395	0,1714	0,0728
2 .....	0,2230	0,0799	0,1691	0,0936
3 .....	0,0410	0,0697	0,1042	0,1133
4 .....	0,5681	0,6170	0,2441	0,3133
5 .....	0,1341	0,0221	0,1379	0,0500
6 .....	0,1222	0,2781	0,1535	0,2389
7 .....	0,1058	0,3012	0,1700	0,1602
8 .....	0,5968	0,0870	0,8338	0,6085
9 .....	0,0614	0,0876	0,4237	0,0896
10 .....	0,0243	0,0243	0,0259	0,0205
11 .....	0,6145	0,1859	0,2970	0,4558

Źródło: opracowanie własne.

Analizując tabele 6–7 widzimy, że zmienność wskaźników rozwoju społeczno-gospodarczego w skupieniach województw jest na ogół mniejsza niż zmienność dla wszystkich województw ogółem (spośród 44 wartości wielowymiarowego współczynnika zmienności w poszczególnych skupieniach, aż 31 wartości jest mniejszych od wartości tego współczynnika dla wszystkich województw ogółem). Można zatem przyjąć, że większe podobieństwo struktury wybranych grup wydatków budżetowych znajduje odzwierciedlenie w zbliżonym poziomie rozwoju społeczno-gospodarczego województw.

## LITERATURA

- Albert A., Zhang L., (2010), A Novel Definition of the Multivariate Coefficient of Variation, *Biometrical Journal*, 52 (5), 667–675.
- Mercer J., (1909), Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations, *Philosophical Transactions of the Royal Society of London, Series A*, 209, 415–426.
- Salamaga M., (2017), Podobieństwo struktury wydatków budżetowych województw a ich zrównoważony rozwój, *Wiadomości Statystyczne*, 9 (676), 79–93.
- Schölkopf B., Smola A., Müller K. B., (1998), Nonlinear Component Analysis as a Kernel Eigenvalues Problem, *Neural Computation*, 10, 1299–1319.

## PODOBIEŃSTWO STRUKTURY WYBRANYCH WYDATKÓW BUDŻETOWYCH WOJEWÓDZTW A ICH ZRÓWNOWAŻONY ROZWÓJ: NOWE NARZĘDZIA BADAWCZE

### Streszczenie

*Celem tej pracy jest opisanie zależności między strukturą wybranych wydatków budżetowych województw a ich zrównoważonym rozwojem. Do tego opisu użyto rozszerzonej jądrowej analizy składowych głównych na przypadek, gdy województwa charakteryzowane są wieloma cechami obserwowanymi w wielu momentach czasowych. Podział województw na grupy jednorodne proponowany jest metodą analizy skupień. Wybór wskaźników zrównoważonego rozwoju najbardziej charakterystycznych dla wyróżnionych skupień odbywa się przy użyciu odległości Mahalanobisa.*

**Słowa kluczowe:** wydatki budżetowe, zrównoważony rozwój, jądrowe składowe główne, odległość Mahalanobisa

---

**THE SIMILARITY OF THE STRUCTURE OF SELECTED BUDGET  
EXPENDITURES OF POLISH VOIVODSHIPS AND THEIR SUSTAINABLE  
DEVELOPMENT: NEW RESEARCH TOOLS**

**Abstract**

*The purpose of this paper is to describe the relationship between the structure of selected budget expenditure of Polish voivodships and their sustainable development. The extended analysis of the kernel principal components was used for this description when the Polish voivodships are characterised by many features observed in many time points. The division of Polish voivodships into homogeneous groups is proposed by the cluster analysis method. The selection of sustainability indicators most characteristic of the distinguished clusters takes place using the Mahalanonis distance.*

**Keywords:** budget expenditure, sustainable development, kernel principal components, Mahalanobis distance