

Anna Turczak

Zachodniopomorska Szkoła Biznesu w Szczecinie
e-mail: aturczak@zpsb.szczecin.pl

Patrycja Zwiech

Uniwersytet Szczeciński
e-mail: patrycjazwiech@tlen.pl

PODOBIENSTWO WOJEWÓDZTW W POLSCE POD WZGLĘDEM ROZKŁADU WYDATKÓW ICH MIESZKAŃCÓW

SIMILARITY OF VOIVODESHIPS IN POLAND IN TERMS OF THEIR RESIDENTS' DISTRIBUTION OF EXPENDITURES

DOI: 10.15611/e21.2015.3.06

JEL Classification: C12, D12, R29

Streszczenie: Celem artykułu było określenie stopnia podobieństwa rozkładu całkowitych miesięcznych wydatków na osobę w poszczególnych województwach oraz wyodrębnienie województw najbardziej do siebie pod względem podobnych. Wszystkie obliczenia przeprowadzono w oparciu o nieidentyfikowalne dane jednostkowe z badania budżetów gospodarstw domowych przez Główny Urząd Statystyczny. Co niezmiernie istotne, badanie budżetów przez GUS prowadzone jest metodą reprezentacyjną, która daje możliwość uogólnienia uzyskanych wyników na wszystkie gospodarstwa w Polsce. W artykule objęto analizą 37 427 gospodarstw domowych, które pogrupowano w szesnaście zbiorowości statystycznych dotyczących województw. Zrealizowano dwa zadania badawcze. Pierwsze z nich dotyczyło sprawdzenia, czy rozkład wydatków na mieszkańca w poszczególnych województwach jest taki sam. W celu realizacji tego zadania postawiono odpowiednie hipotezy statystyczne i zweryfikowano je za pomocą testu Kołmogorowa–Smirnowa. Procedurę weryfikacyjną przeprowadzono dla każdej pary województw, czyli sto dwadzieścia razy. Przyjęto poziom istotności 0,01, a zatem ryzyko odrzucenia hipotezy prawdziwej wynosiło jedynie 1 na 100 przypadków. Sfinalizowanie zadania pierwszego pozwoliło wyciągnąć wniosek, że zaobserwowane różnice między rozkładami całkowitych miesięcznych wydatków *per capita* w poszczególnych województwach Polski są statystycznie istotne, a zatem zmienna będąca przedmiotem badania ma w każdym z województw inny rozkład. Drugie zadanie badawcze polegało na podzieleniu województw na grupy o jak najbardziej podobnych rozkładach. Do wykonania tego zadania posłużyła taksonomia wrocławska. Jako miernik stopnia podobieństwa rozkładów wykorzystano statystykę λ (lambda) opartą na maksymalnej bezwzględnej wartości różnicy między dwiema dystrybucjami empirycznymi. Na podstawie wartości statystyki λ obliczonej dla każdej z par rozkładów szesnaście województw podzielono na trzy jednolite klasy. Podział ten skutkowało utworzeniem jednej grupy jednoelementowej, jednej grupy ośmioelementowej oraz jednej grupy siedmio-

elementowej. W grupie jednoelementowej znalazło się województwo mazowieckie, w grupie ośmioelementowej – województwa dolnośląskie, śląskie, pomorskie, opolskie, łódzkie, zachodniopomorskie, małopolskie i lubuskie, a w grupie siedmioelementowej – województwa kujawsko-pomorskie, podlaskie, lubelskie, wielkopolskie, świętokrzyskie, warmińsko-mazurskie i podkarpackie. W 2012 roku średnie miesięczne wydatki w woj. mazowieckim wynosiły 1330 zł/osobę, w dolnośląskim: 1139 zł/osobę, w śląskim: 1123 zł/osobę, w pomorskim: 1081 zł/osobę, w opolskim: 1079 zł/osobę, w łódzkim: 1075 zł/osobę, w zachodniopomorskim: 1057 zł/osobę, w małopolskim: 998 zł/osobę, w lubuskim: 996 zł/osobę, w kujawsko-pomorskim: 947 zł/osobę, w podlaskim: 939 zł/osobę, w lubelskim: 938 zł/osobę, w wielkopolskim: 931 zł/osobę, w świętokrzyskim: 884 zł/osobę, w warmińsko-mazurskim: 865 zł/osobę i w podkarpackim: 848 zł/osobę. Z kolei przeciętne różnice w miesięcznych wydatkach mieszkańców woj. mazowieckiego w 2012 r. opiewały na 1171 zł/osobę, dolnośląskiego: 854 zł/osobę, śląskiego: 783 zł/osobę, pomorskiego: 843 zł/osobę, opolskiego: 606 zł/osobę, łódzkiego: 755 zł/osobę, zachodniopomorskiego: 731 zł/osobę, małopolskiego: 649 zł/osobę, lubuskiego: 581 zł/osobę, kujawsko-pomorskiego: 635 zł/osobę, podlaskiego: 712 zł/osobę, lubelskiego: 793 zł/osobę, wielkopolskiego: 675 zł/osobę, świętokrzyskiego: 578 zł/osobę, warmińsko-mazurskiego: 700 zł/osobę, podkarpackiego: 478 zł/osobę.

Słowa kluczowe: rozkład wydatków, wydatki ogółem na osobę, test Kołmogorowa–Smirnowa, gospodarstwa domowe, taksonomia.

Summary: The purpose of the paper was to determine the degree of similarity between the distribution of total monthly expenditures *per capita* in the individual voivodeships and to single out voivodeships most similar in that respect. All calculations were carried out based on non-identifiable individual data from the household budget survey by the Central Statistical Office. What is really important is the fact that surveying budgets by the CSO is conducted using the representative method, which allows for the generalization of the obtained results over all households in Poland. The paper included 37,427 analyzed households which were grouped into sixteen statistical populations with respect to the voivodeships. Two research tasks were implemented. The first concerned verification whether the distribution of spending *per capita* in individual voivodeships was the same. In order to complete this task relevant statistical hypotheses were set and verified using the Kolmogorov-Smirnov test. The verification procedure was performed for each pair of voivodeships, i.e. one hundred and twenty times. The significance level of 0.01 was adopted, and therefore the risk of rejection of a true hypothesis was only 1 in 100 cases. The finalization of the first task has allowed the conclusion that the observed differences between the distributions of total monthly spending *per capita* in individual voivodeships of Poland are statistically significant, and thus the variable which is the subject of research has a different distribution in each of the voivodeships. The second research task was to divide the voivodeships into groups of most similar distributions. In order to accomplish this task the Wrocław taxonomy was employed. As a measure of the degree of similarity of distributions λ (lambda) statistic was used, which is based on the maximum absolute value of the difference between two empirical cumulative distribution functions. On the basis of the value of the λ statistic calculated for each of the pairs of distributions the sixteen voivodeships were divided into three uniform classes. This division resulted in the creation of a single-element group, one eight-element group and one seven-element group. The single-element group comprised the Mazovian voivodeship, the eight-element group included the Lower Silesian, Silesian, Pomeranian, Opole, Łódź, West Pomeranian, Lesser Poland and Lubuskie voivodeships, and the seven-element group comprised the Kujawsko-Pomorskie, Podlaskie, Lublin, Greater Poland, Świętokrzyskie, Warmia-Mazury and Podkarpackie voivodeships. In 2012, the average monthly spending in

the individual voivodeships amounted to: in the Mazovian voiv. PLN 1,330 per person, in the Lower Silesian PLN 1,139 per person, in the Silesian voiv. PLN 1,123 per person, in the Pomeranian voiv. PLN 1,081 per person, in the Opole voiv. PLN 1,079 per person, in the Łódź voiv. PLN 1,075 per person, in the West Pomeranian voiv. PLN 1,057 per person, in the Lesser Poland voiv. PLN 998 per person, in the Lubuskie voiv. PLN 996 per person, in the Kujawsko-Pomorskie voiv. PLN 947 per person, in the Podlaskie voiv. PLN 939 per person, in the Lublin voiv. PLN 938 per person, in the Greater Poland voiv. PLN 931 per person, in the Świętokrzyskie voiv. PLN 884 per person, in the Warmia-Mazury voiv. PLN 865 per person and in the Podkarpackie voiv. PLN 848 per person. In turn, the average difference in monthly spending of residents of the Mazovian voivodeship in 2012 amounted to PLN 1,171 per person, the Lower Silesian voiv. PLN 854 per person, the Silesian voiv. PLN 783 per person, the Pomeranian voiv. PLN 843 per person, the Opole voiv. PLN 606 per person, the Łódź voiv. PLN 755 per person, the West Pomeranian voiv. PLN 731 per person, the Lesser Poland voiv. PLN 649 per person, the Lubuskie voiv. PLN 581 per person, the Kujawsko-Pomorskie voiv. PLN 635 per person, the Podlaskie voiv. PLN 712 per person, the Lublin voiv. PLN 793 per person, the Greater Poland voiv. PLN 675 per person, the Świętokrzyskie voiv. PLN 578 per person, the Warmia-Mazury voiv. PLN 700 per person, the Podkarpackie voiv. PLN 478 per person.

Keywords: distribution of expenditures, total expenditures *per capita*, Kolmogorov-Smirnov test, households, taxonomy.

1. Wstęp

Poziom wydatków jest jednym z ważniejszych i obiektywnych mierników określających materialny standard życia ludności [Grzywińska-Rapca 2014]. Poziom ten uzależniony jest między innymi od geograficznego umiejscowienia gospodarstwa domowego. Badania empiryczne potwierdzają, iż istnieje statystycznie istotna zależność między położeniem danego gospodarstwa na mapie Polski a wielkością i strukturą realizowanych przez nie wydatków. Zależności takie można zaobserwować w przypadku przeprowadzania analiz porównawczych dla różnych jednostek podziału terytorialnego kraju, takich jak gminy czy powiaty [Kasprzyk, Leszczyńska 2012].

Pozytywny wpływ wielkości realizowanych wydatków na poziom zaspokojenia potrzeb ludzi jest bezsprzeczny. Interesujące jest natomiast, czy rozkład całkowitych wydatków¹ na osobę w poszczególnych województwach Polski jest identyczny,

¹ Za Głównym Urzędem Statystycznym przyjęto, że na całkowite wydatki składają się wydatki na towary i usługi konsumpcyjne oraz pozostałe wydatki. Wydatki na towary i usługi konsumpcyjne przeznaczane są na zaspokojenie potrzeb gospodarstwa domowego i obejmują towary zakupione za gotówkę (w tym przy użyciu karty płatniczej bądź kredytowej), na kredyt, otrzymane bezpłatnie, jak również spożycie naturalne (tj. towary i usługi konsumpcyjne pobrane na potrzeby gospodarstwa domowego z działalności rolniczej albo działalności gospodarczej na własny rachunek). Natomiast pozostałe wydatki składają się między innymi z: 1) darów przekazanych innym gospodarstwom domowym i instytucjom niekomercyjnym, 2) niektórych podatków (w tym podatku od spadków i darowizn, podatku od nieruchomości, opłaty za wieczyste użytkowanie gruntu), 3) zaliczek na podatek od dochodów osobistych oraz składek na ubezpieczenia społeczne płaconych samodzielnie przez podatnika [Budżety... 2014].

a jeśli nie, to które województwa można uznać za te o podobnym rozkładzie. Określenie stopnia podobieństwa rozkładów wydatków *per capita* w poszczególnych województwach oraz wyodrębnienie województw najbardziej do siebie pod tym względem zbliżonych stało się celem niniejszego artykułu. Celowi temu służyć będzie realizacja następujących zadań badawczych:

1) sprawdzenie, czy rozkład wydatków na mieszkańca w poszczególnych województwach jest taki sam,

2) podzielenie województw na grupy o podobnym rozkładzie badanej zmiennej.

Zadanie pierwsze zrealizowano z wykorzystaniem testu Kołmogorowa–Smirnowa, natomiast do wykonania zadania drugiego posłużyła taksonomia wrocławska.

Niniejszy artykuł ma charakter badawczy. Wszystkie zawarte w nim obliczenia przeprowadzono w oparciu o nieidentyfikowalne dane jednostkowe z badania budżetów gospodarstw domowych w Polsce². Wspomniana baza w 2012 r. objęła 37 427 gospodarstw domowych. Co warte podkreślenia, badanie budżetów przez Główny Urząd Statystyczny prowadzone jest metodą reprezentacyjną, która daje możliwość uogólnienia uzyskanych wyników na wszystkie gospodarstwa w Polsce [Budżety... 2014].

2. Metodyka badania

Do przeprowadzenia badań wykorzystano test Kołmogorowa–Smirnowa i taksonomię wrocławską. Test Kołmogorowa–Smirnowa służy do weryfikacji hipotezy stanowiącej, że dwie próby pochodzą z tej samej populacji (albo inaczej – że dwie populacje mają ten sam rozkład).

Rozpatrywaną zmienną (tj. całkowite miesięczne wydatki *per capita*) oznaczono przez X . Dystrybuanta $F(x)$ w pełni opisuje rozkład zmiennej X w populacji [Kot i in. 2007]. Stąd porównanie rozkładu zmiennej w dwóch populacjach można sprowadzić do porównania wartości dystrybuant w tych populacjach i jeśli dwie populacje mają ten sam rozkład, to wartości ich dystrybuant powinny być we wszystkich punktach identyczne. Aby udowodnić, że dwie populacje (oznaczone odpowiednio subskryptami 1 i 2) mają jednakowy rozkład, należy zatem sprawdzić hipotezę zerową [Razali, Wah 2011]:

$$H_0 : F_1(x) = F_2(x)$$

wobec hipotezy alternatywnej:

$$H_1 : F_1(x) \neq F_2(x).$$

Wynika z tego, że jeśli dwie próby pochodzą z jednej populacji (bądź z dwóch identycznych populacji), to wartości dystrybuant empirycznych:

² Bazę nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych za rok 2012 udostępnił GUS na podstawie Umowy nr 20/Z/DI-6-611/632/2013/RM między GUS i US.

$$F_{n_1}(x) \text{ i } F_{n_2}(x),$$

gdzie n_1 to liczebność pierwszej próby, a n_2 to liczebność drugiej próby, powinny być we wszystkich punktach zbliżone.

W celu określenia różnic między wartościami dystrybuant empirycznych wszystkie obserwacje występujące w badanych próbach uporządkowano w kolejności nie-malejącej. Następnie dla każdej i -tej obserwacji obliczono wartości obu dystrybuant odpowiednio według wzorów:

$$F_{n_1}(x_i) = \frac{n_{1sk.i}}{n_1} \quad F_{n_2}(x_i) = \frac{n_{2sk.i}}{n_2},$$

gdzie $n_{1sk.i}$ i $n_{2sk.i}$ oznaczają liczebności skumulowane dla pierwszej i drugiej próbki.

W następnym kroku realizacji omawianego testu dla każdej wartości zmiennej X obliczono wartość bezwzględną z różnicy między dystrybuantami. Znaną największą bezwzględną wartość takiej różnicy oznaczono przez D_{12} i zdefiniowano jako [Taylor, Emerson 2011]:

$$D_{12} = \max_{x_i} |F_{n_1}(x_i) - F_{n_2}(x_i)|.$$

W oparciu o statystykę D wyznaczono statystykę λ wyrażoną wzorem:

$$\lambda_{12} = D_{12} \sqrt{n_{12}},$$

gdzie [Rószkiewicz 2012]:

$$n_{12} = \frac{n_1 n_2}{n_1 + n_2}.$$

Z budowy statystyki λ wynika, że im większa będzie maksymalna różnica D , tym większą wartość będzie miała statystyka λ i tym większe będą podstawy do odrzucenia przypuszczenia o identyczności rozkładów w populacjach, z których wylosowano próby.

Przy założeniu prawdziwości hipotezy H_0 statystyka λ ma asymptotyczny rozkład λ Kołmogorowa [Witkowski 2010]. Z tablicy tego rozkładu dla przyjętego z góry poziomu istotności α odczytano taką wartość krytyczną λ_α , aby spełnione było równanie:

$$P\{\lambda \geq \lambda_\alpha\} = \alpha.$$

Następnie obliczoną wartość λ porównano z wartością krytyczną λ_α . Jeśli zaszła by nierówność $\lambda \geq \lambda_\alpha$, to hipotezę H_0 należałoby odrzucić na rzecz hipotezy alternatywnej. Oznaczałoby to, że obie próby nie pochodzą z tej samej populacji (albo inaczej – dwie populacje, z których pochodzą próby, mają inny rozkład). Natomiast

gdyby spełniona była nierówność $\lambda < \lambda_a$, wówczas nie znaleziono by podstaw do odrzucenia hipotezy zerowej o jednakowych rozkładach.

Jak już wspomniano, postawionym zadaniem jest porównanie rozkładów cechy X (tj. całkowitych miesięcznych wydatków na osobę) w szesnastu województwach Polski. Ustalono też, że podobieństwo rozkładów będzie mierzone za pomocą statystyki λ , przy czym dwa rozkłady będą tym bardziej podobne, im mniejszą wartość będzie miała λ .

Dzięki przeprowadzeniu klasyfikacji zbiór szesnastu województw Polski zostanie podzielony na ustaloną liczbę podzbiorów, które będą jednorodne pod względem przyjętego kryterium. Pożądany będzie taki podział, w którym wartość statystyki λ obliczonej dla dowolnej pary województw należących do tej samej klasy jest mniejsza niż wartość tej statystyki dotyczącej dowolnej pary województw należących do różnych klas [Sompolska-Rzechuła 2002].

W celu porównania rozkładów miesięcznych wydatków na osobę w szesnastu województwach Polski wykorzystano taksonomię wrocławską, zwaną także metodą dendrytową (realizowana jest bowiem w sposób graficzny za pomocą dendrytu). Taksonomia wrocławska przeprowadzana została w następujących etapach [Dziechciarz (red.) 2002]:

Etap 1. Dla każdego województwa poszukiwano województwa najbardziej podobnego. Do j -tego województwa najbardziej podobne jest województwo l -te, dla którego wartość statystyki λ jest najmniejsza, co można matematycznie zapisać jako [Młodak 2006]:

$$\min_l \lambda_j, (l \neq j).$$

W następnej kolejności zbudowano dendryt składający się z wierzchołków i wiązań. Konstrukcję dendrytu rozpoczęto od połączenia każdego województwa (a ściślej – wierzchołka grafu odpowiadającego województwu) z najbardziej do niego podobnym. W wyniku przeprowadzenia procedury łączenia uzyskano graf złożony z tzw. skupień pierwszego rzędu³. Po sfinalizowaniu etapu pierwszego mogły zaistnieć dwie sytuacje. Mogło mianowicie się okazać, że utworzony dendryt był grafem spójnym (czyli otrzymano jedno skupienie pierwszego rzędu, w którym wszystkie wierzchołki zostały połączone nieprzerwanym ciągiem wiązań) [Piszczala 2000]. W przypadku wystąpienia takiej sytuacji po etapie pierwszym przechodzi się bezpośrednio do etapu trzeciego. Jeśli natomiast w etapie pierwszym otrzyma się co najmniej dwa skupienia pierwszego rzędu, należy przeprowadzić etap drugi.

Etap 2. W etapie tym dla każdego skupienia pierwszego rzędu poszukuje się skupienia najbardziej podobnego spośród wszystkich pozostałych skupień.

³ Skupienie takie to grupa województw połączonych ze sobą za pomocą wiązań bezpośrednio albo pośrednio.

Jako wartość statystyki λ odnoszącą się do pary skupień przyjęto minimalną wartość tej statystyki obliczoną dla poszczególnych województw należących do tych dwóch skupień. W rezultacie połączenia każdego skupienia pierwszego rzędu ze skupieniem, które jest do niego najbardziej podobne, tworzą się tzw. skupienia drugiego rzędu. Procedurę łączenia trzeba powtarzać aż do momentu, w którym wszystkie skupienia będą ze sobą połączone i otrzyma się graf spójny.

Etap 3. Należy dokonać podziału grafu spójnego. W przypadku, gdy ostateczna klasyfikacja ma wyodrębnić k rozłącznych grup, konieczne jest usunięcie z otrzymanego dendrytu $k - 1$ najdłuższych wiązań.

3. Zbadanie podobieństwa rozkładu wydatków między poszczególnymi województwami

Dla każdego gospodarstwa domowego ankietowanego przez GUS w ramach badania budżetów gospodarstw domowych za 2012 rok wyznaczono średnie miesięczne wydatki przypadające na osobę. Informacje zawarte w udostępnionej bazie danych pozwoliły także przyporządkować poszczególne gospodarstwa do odpowiednich województw. Dzięki temu możliwe stało się wyodrębnienie szesnastu zbiorowości statystycznych. Postawiono hipotezę zerową głoszącą, że dystrybuanty rozkładów tej samej cechy ilościowej X (tj. całkowitych miesięcznych wydatków na osobę) w dwóch województwach są takie same, wobec hipotezy alternatywnej, że są różne. Weryfikację hipotezy zerowej przeprowadzono oddzielnie dla każdej pary województw. W tab. 1 podano uzyskane wartości statystyki D , a w tab. 2 – statystyki λ .

Niech przyjęty z góry poziom istotności α będzie równy 0,01. Wówczas odczytana z tablicy granicznego rozkładu λ Kołmogorowa wartość krytyczna dla założonego współczynnika $\alpha = 0,01^4$ wynosi $\lambda_\alpha = 1,63$. Dla każdej pary województw otrzymano $\lambda \geq \lambda_\alpha$, toteż wartość statystyki λ znalazła się w obszarze krytycznym i hipotezę H_0 należało w każdym przypadku odrzucić. Nie można zatem twierdzić, że którakolwiek z rozpatrywanych par województw ma taki sam rozkład wydatków miesięcznych ogółem na osobę. Oznacza to, że różnice między wartościami dystrybuant empirycznych w próbach były na tyle duże, iż przypuszczenie o identyczności rozkładów w populacjach generalnych zostało odrzucone.

⁴ Obniżenie wartości współczynnika istotności poniżej poziomu 0,01 nie wydaje się zasadne. Z kolei podwyższenie jego wartości ponad 0,01 wpłynie na zmniejszenie wartości krytycznej (przykładowo dla $\alpha = 0,05$ odszukana z tablic wartość λ_α wynosi już tylko 1,36). Oznacza to, że zwiększenie poziomu istotności spowodowałoby zwiększenie obszaru krytycznego i hipotezę zerową należałoby tym bardziej odrzucić. A więc wzrost α nie wpłynąłby w żaden sposób na wyniki przeprowadzanych weryfikacji.

Tabela 1. Wartości statystyki D dla poszczególnych par województw

D	DŚ	KP	LB	LS	ŁD	MP	MZ	OP	PK	PL	PM	ŚL	ŚK	WM	WP	ZP
DŚ	0,000	0,146	0,178	0,081	0,037	0,104	0,086	0,037	0,198	0,175	0,079	0,023	0,195	0,214	0,158	0,069
KP	0,146	0,000	0,041	0,103	0,125	0,063	0,187	0,154	0,066	0,044	0,101	0,142	0,062	0,077	0,020	0,096
LB	0,178	0,041	0,000	0,120	0,159	0,094	0,211	0,189	0,049	0,032	0,120	0,175	0,036	0,060	0,032	0,126
LS	0,081	0,103	0,120	0,000	0,056	0,046	0,151	0,091	0,136	0,128	0,073	0,077	0,141	0,162	0,111	0,044
ŁD	0,037	0,125	0,159	0,056	0,000	0,082	0,115	0,050	0,171	0,149	0,063	0,037	0,170	0,192	0,139	0,043
MP	0,104	0,063	0,094	0,046	0,082	0,000	0,158	0,115	0,108	0,096	0,071	0,103	0,115	0,134	0,077	0,050
MZ	0,086	0,187	0,211	0,151	0,115	0,158	0,000	0,114	0,242	0,217	0,103	0,088	0,232	0,247	0,189	0,117
OP	0,037	0,154	0,189	0,091	0,050	0,115	0,114	0,000	0,205	0,182	0,102	0,028	0,203	0,219	0,161	0,080
PK	0,198	0,066	0,049	0,136	0,171	0,108	0,242	0,205	0,000	0,057	0,157	0,194	0,035	0,060	0,060	0,141
PL	0,175	0,044	0,032	0,128	0,149	0,096	0,217	0,182	0,057	0,000	0,131	0,169	0,039	0,054	0,045	0,121
PM	0,079	0,101	0,120	0,073	0,063	0,071	0,103	0,102	0,157	0,131	0,000	0,088	0,147	0,156	0,105	0,045
ŚL	0,023	0,142	0,175	0,077	0,037	0,103	0,088	0,028	0,194	0,169	0,088	0,000	0,192	0,207	0,149	0,064
ŚK	0,195	0,062	0,036	0,141	0,170	0,115	0,232	0,203	0,035	0,039	0,147	0,192	0,000	0,041	0,048	0,146
WM	0,214	0,077	0,060	0,162	0,192	0,134	0,247	0,219	0,060	0,054	0,156	0,207	0,041	0,000	0,065	0,161
WP	0,158	0,020	0,032	0,111	0,139	0,077	0,189	0,161	0,060	0,045	0,105	0,149	0,048	0,065	0,000	0,102
ZP	0,069	0,096	0,126	0,044	0,043	0,050	0,117	0,080	0,141	0,121	0,045	0,064	0,146	0,161	0,102	0,000

Źródło: obliczenia własne na podstawie bazy nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych za 2012 rok (Umowa nr 20/Z/DI-6-611/632/2013/RM między GUS i US).

Tabela 2. Wartości statystyki λ dla poszczególnych par województw

λ	DŚ	KP	LB	LS	ŁD	MP	MZ	OP	PK	PL	PM	ŚL	ŚK	WM	WP	ZP
DŚ	0,00	8,32	10,60	3,57	2,26	6,76	6,17	1,71	11,53	8,55	4,63	1,60	9,68	10,97	10,36	3,66
KP	8,32	0,00	2,25	4,36	6,93	3,69	11,86	6,77	3,53	2,04	5,44	8,73	2,88	3,72	1,19	4,77
LB	10,60	2,25	0,00	5,16	9,19	5,79	14,18	8,50	2,75	1,54	6,71	11,32	1,73	2,96	2,00	6,42
LS	3,57	4,36	5,16	0,00	2,43	2,08	7,12	3,39	5,81	4,92	3,12	3,55	5,51	6,45	5,04	1,78
ŁD	2,26	6,93	9,19	2,43	0,00	5,13	7,89	2,25	9,67	7,16	3,57	2,44	8,26	9,64	8,85	2,23
MP	6,76	3,69	5,79	2,08	5,13	0,00	11,92	5,41	6,51	4,77	4,27	7,41	5,83	7,07	5,27	2,70
MZ	6,17	11,86	14,18	7,12	7,89	11,92	0,00	5,67	15,84	11,47	6,73	7,12	12,50	13,84	14,53	6,82
OP	1,71	6,77	8,50	3,39	2,25	5,41	5,67	0,00	9,09	7,24	4,54	1,35	8,17	9,01	7,64	3,35
PK	11,53	3,53	2,75	5,81	9,67	6,51	15,84	9,09	0,00	2,68	8,64	12,26	1,68	2,95	3,63	7,08
PL	8,55	2,04	1,54	4,92	7,16	4,77	11,47	7,24	2,68	0,00	6,13	8,71	1,63	2,30	2,27	5,29
PM	4,63	5,44	6,71	3,12	3,57	4,27	6,73	4,54	8,64	6,13	0,00	5,55	6,97	7,64	6,41	2,26
ŚL	1,60	8,73	11,32	3,55	2,44	7,41	7,12	1,35	12,26	8,71	5,55	0,00	10,06	11,27	10,92	3,65
ŚK	9,68	2,88	1,73	5,51	8,26	5,83	12,50	8,17	1,68	1,63	6,97	10,06	0,00	1,79	2,47	6,48
WM	10,97	3,72	2,96	6,45	9,64	7,07	13,84	9,01	2,95	2,30	7,64	11,27	1,79	0,00	3,46	7,34
WP	10,36	1,19	2,00	5,04	8,85	5,27	14,53	7,64	3,63	2,27	6,41	10,92	2,47	3,46	0,00	5,62
ZP	3,66	4,77	6,42	1,78	2,23	2,70	6,82	3,35	7,08	5,29	2,26	3,65	6,48	7,34	5,62	0,00

Źródło: obliczenia własne na podstawie tab. 1.

4. Podział województw na grupy podobne pod względem rozkładu wydatków

Realizacja zadania pierwszego ukazała, że wydatki na osobę mają w poszczególnych województwach Polski inny rozkład. Warto więc sprawdzić, które województwa mają rozkłady podobne. W tym celu zastosowano metodę dendrytową.

Postawionym zadaniem jest więc podział zbioru szesnastu województw na takie rozłączne i niepuste podzbiory, aby województwa należące do tych samych klas były pod względem rozkładów jak najbardziej podobne, a województwa należące do różnych klas były jak najmniej podobne. Realizowana w tym celu procedura metody dendrytowej przeprowadzona zostanie w oparciu o wartość statystyki λ .

W każdym wierszu tab. 2 pogrubioną czcionką zaznaczono najmniejsze dodatnie wartości λ . W oparciu o te liczby sporządzono graf, na którym województwa (tj. wierzchołki grafu) oznaczono kółkami. Otrzymano cztery skupienia pierwszego rzędu. Składają się one z następujących województw (sposób numerowania poszczególnych skupień jest dowolny):

- skupienie **I**: woj. dolnośląskie (DŚ), śląskie (ŚL), opolskie (OP), mazowieckie (MZ);
- skupienie **II**: woj. łódzkie (ŁD), zachodniopomorskie (ZP), pomorskie (PM), lubuskie (LS), małopolskie (MP);
- skupienie **III**: woj. lubelskie (LB), podlaskie (PL), świętokrzyskie (ŚK), warmińsko-mazurskie (WM), podkarpackie (PK);
- skupienie **IV**: woj. kujawsko-pomorskie (KP), wielkopolskie (WP).

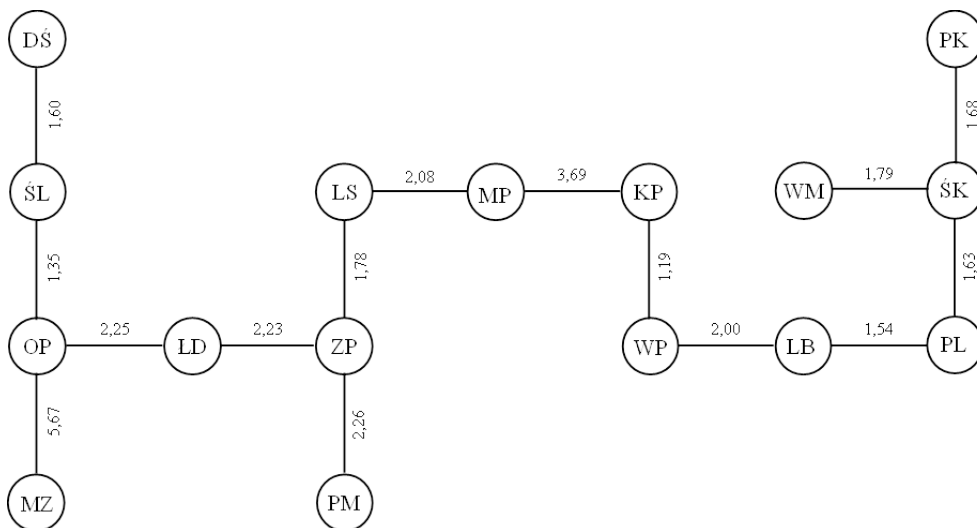
W następnym etapie połączone zostaną ze sobą odpowiednie skupienia pierwszego rzędu. Wartości statystyki λ dla par województw, które można byłoby ze sobą połączyć w celu połączenia skupień, do których te województwa należą, znajdują się w tab. 3. W tabeli tej w każdym wierszu wytłuszczono najmniejszą niezerową wartość λ .

Tabela 3. Wartości statystyki λ dla poszczególnych par skupień

λ	(DŚ, ŚL, OP, MZ)	(ŁD, ZP, PM, LS, MP)	(LB, PL, ŚK, WM, PK)	(KP, WP)
(DŚ, ŚL, OP, MZ)	0,00	2,25	7,24	6,77
(ŁD, ZP, PM, LS, MP)	2,25	0,00	4,77	3,69
(LB, PL, ŚK, WM, PK)	7,24	4,77	0,00	2,00
(KP, WP)	6,77	3,69	2,00	0,00

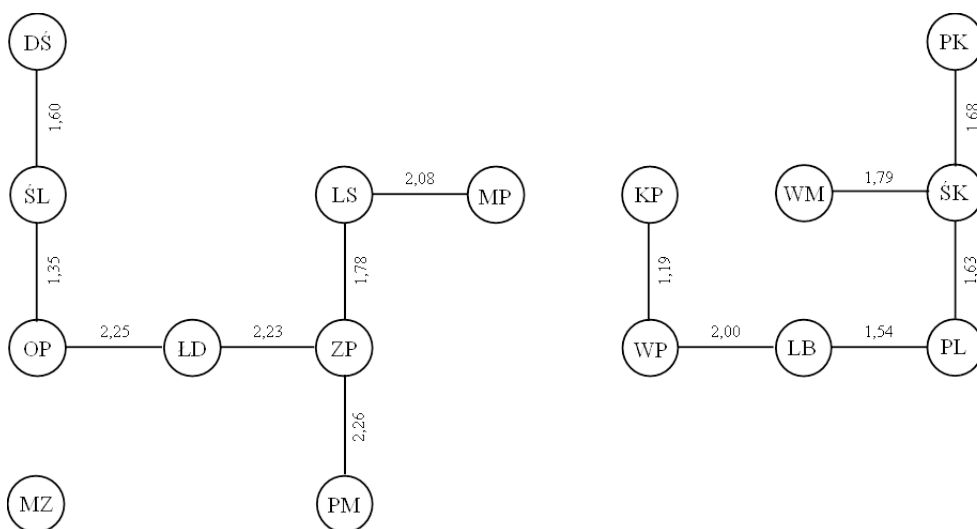
Źródło: obliczenia własne na podstawie tab. 2.

Połączono skupienie **I** ze skupieniem **II**, skupienie **III** ze skupieniem **IV**, a następnie skupienie **II** ze skupieniem **IV**. Powstały w ten sposób graf jest grafem spójnym, zatem można już przejść do jego odpowiedniego podzielenia. Przy szesnastu



Rys. 1. Dendryt spójny

Źródło: opracowanie własne na podstawie tab. 2.



Rys. 2. Podział województw na trzy klasy

Źródło: opracowanie własne na podstawie rys. 1.

wierzchołkach grafu nie jest zasadne dzielenie go na więcej niż trzy klasy. W drodze przeprowadzenia poszczególnych etapów taksonomii wrocławskiej ostatecznie powstały następujące grupy województw (w nawiasach podano średnie miesięczne

wydatki w zł na osobę, odchylenie standardowe w zł na osobę oraz klasyczny współczynnik zmienności):

- {woj. mazowieckie (1330, 1171, 88%)};
- {woj. dolnośląskie (1139, 854, 75%), śląskie (1123, 783, 70%), pomorskie (1081, 843, 78%), opolskie (1079, 606, 56%), łódzkie (1075, 755, 70%), zachodniopomorskie (1057, 731, 69%), małopolskie (998, 649, 65%), lubuskie (996, 581, 58%)};
- {woj. kujawsko-pomorskie (947, 635, 67%), podlaskie (939, 712, 76%), lubelskie (938, 793, 85%), wielkopolskie (931, 675, 73%), świętokrzyskie (884, 578, 65%), warmińsko-mazurskie (865, 700, 81%), podkarpackie (848, 478, 56%)}

Utworzony dendryt spójny przedstawiono na rys. 1, a wynik podziału województw na trzy klasy przedstawiono na rys. 2.

5. Zakończenie

Celem artykułu była odpowiedź na pytanie, czy w poszczególnych województwach Polski jest taki sam rozkład wydatków *per capita*, a jeśli nie, to które województwa mają najbardziej podobne do siebie rozkłady analizowanej zmiennej. Aby osiągnąć postawiony cel, zrealizowano dwa zadania badawcze.

Realizacja pierwszego zadania badawczego polegała na weryfikacji hipotezy statystycznej głoszącej, że rozkłady badanej cechy w dwóch populacjach są takie same, a więc istniejące różnice w wartościach dystrybuant obliczonych na podstawie wyników z prób są statystycznie nieistotne. Sprawdzenia prawdziwości tak sformułowanej hipotezy za pomocą testu Kołmogorowa–Smirnowa dokonano dla każdej pary województw, a zatem test przeprowadzono 120 razy. W przypadku każdej pary województw wartość statystyki empirycznej λ znalazła się w prawostronnym obszarze krytycznym określonym równaniem: $P\{\lambda \geq 1,63\} = 0,01$. W związku z tym we wszystkich ze 120 przypadków przypuszczenie o identyczności rozkładów wydatków na osobę w danych województwach trzeba było odrzucić.

Realizacja drugiego zadania badawczego pozwoliła pogrupować województwa w klasy o najbardziej podobnych do siebie rozkładach całkowitych miesięcznych wydatków na osobę. Grupowania tego dokonano w oparciu o wartości statystyki λ obliczonej w ramach pierwszego zadania badawczego i przy wykorzystaniu metody dendrytowej. W efekcie zastosowanego narzędzia badawczego uzyskano jedną grupę jednoelementową (woj. mazowieckie), jedną grupę ośmioelementową (woj. dolnośląskie, śląskie, pomorskie, opolskie, łódzkie, zachodniopomorskie, małopolskie, lubuskie) oraz jedną grupę siedmioelementową (woj. kujawsko-pomorskie, podlaskie, lubelskie, wielkopolskie, świętokrzyskie, warmińsko-mazurskie, podkarpackie).

Dotychczas powstało wiele opracowań obejmujących badania kształtowania się wybranej zmiennej (w tym wydatków *per capita*) w poszczególnych województwach. W dostępnych analizach województwa zostały nierzadko też pogrupowane

na klasy, a klasyfikacja taka służyła zazwyczaj określeniu zasięgu terytorialnego tzw. Polski A i B, czy nawet Polski A, B i C. Wspomniane opracowania bazowały jednak na pewnych uproszczeniach, poszczególne województwa bowiem najczęściej zostały w nich porównane pod względem wyłącznie jakiejś jednej syntetycznej miary bądź ewentualnie kilku wybranych statystyk (np. średniej arytmetycznej czy mediany jako miar tendencji centralnej, odchylenia standardowego albo współczynnika zmienności jako miar dyspersji, a w przypadku badania nierówności społecznych – współczynników asymetrii i/lub wybranych miar koncentracji). Zaprezentowana w niniejszym artykule metodyka pozwala natomiast na pogrupowanie województw ze względu na kształtowanie się całych rozkładów, a zatem obejmuje wszystkie istniejące różnice w kształtowaniu się badanej zmiennej w poszczególnych województwach – i te w zakresie tendencji centralnej, i te w zakresie dyspersji, i te w zakresie asymetrii. Skoro więc celem grupowania ma być podzielenie województw na jak najbardziej jednorodne klasy, to wydaje się, że proponowany w niniejszym artykule sposób – prezentujący podejście całościowe – pozwala w największym stopniu ten cel osiągnąć. Przedstawiona w artykule metodyka ma zatem bardzo duży walor aplikacyjny.

Literatura

- Budżety gospodarstw domowych w 2013 r.*, GUS, 2014, Warszawa, s. 14, 20.
- Dziechciarz J. (red.), 2002, *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław, s. 273.
- Grzywińska-Rapca M., 2014, *Analiza wydatków konsumpcyjnych rolniczych gospodarstw domowych*, Zeszyty Naukowe Firma i Rynek, Wydawnictwo Zachodniopomorskiej Szkoły Biznesu w Szczecinie, Szczecin, s. 92.
- Kasprzyk B., Leszczyńska M., 2012, *Dochody i wydatki jako determinanty dobrobytu ekonomicznego gospodarstw domowych w Polsce – ujęcie regionalne*, Nierówności społeczne a wzrost gospodarczy nr 28, Katedra Teorii Ekonomii i Stosunków Międzynarodowych Uniwersytetu Rzeszowskiego, Rzeszów, s. 267.
- Kot S.M., Jakubowski J., Sokołowski A., 2007, *Statystyka*, Difin, Warszawa, s. 267.
- Młodak A., 2006, *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa, s. 77.
- Piszczala J., 2000, *Matematyka i jej zastosowanie w naukach ekonomicznych*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, s. 23.
- Razali N.M., Wah Y.B., 2011, *Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests*, Journal of Statistical Modeling and Analytics, vol. 2, no. 1, s. 23.
- Rószkiewicz M., 2012, *Metody ilościowe w badaniach marketingowych*, PWN, Warszawa, s. 304.
- Sompolska-Rzechuła A., 2002, *Zastosowanie taksonomii rozmytej do klasyfikacji spółek na Gieldzie Papierów Wartościowych*, [w:] *Rynek kapitałowy. Skuteczne inwestowanie*. Część I, Konferencja naukowa zorganizowana przez Katedrę Ekonometrii i Statystyki Wydziału Nauk Ekonomicznych i Zarządzania, Międzyzdroje, s. 525.
- Taylor B.A., Emerson J.W., 2011, *Nonparametric goodness-of-fit tests for discrete null distributions*, The R Journal, vol. 2, no. 3, s. 34.
- Witkowski M. (red.), 2010, *Statystyka matematyczna w zarządzaniu*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, s. 92.