STUDIA METODOLOGICZNE

Anna TURCZAK, Patrycja ZWIECH

Zróżnicowanie dochodów ludności według województw

Wpływ wielkości dochodu rozporządzalnego[1] na jakość życia ludzi jest bezsprzeczny. Interesująca zatem może być odpowiedź na pytanie, czy rozkład dochodu rozporządzalnego na osobę w poszczególnych województwach jest identyczny, a jeżeli nie, to które województwa można uznać za te o podobnym rozkładzie. Celem artykułu stało się zatem określenie stopnia podobieństwa rozkładu dochodu rozporządzalnego *per capita* w poszczególnych województwach oraz wyodrębnienie województw najbardziej do siebie pod tym względem zbliżonych. Służyć temu będzie realizacja następujących zadań badawczych:

1) sprawdzenie, czy rozkład dochodu rozporządzalnego na mieszkańca w poszczególnych województwach jest identyczny;
2) podzielenie województw na grupy o podobnym rozkładzie badanej zmiennej.

Zadanie pierwsze zrealizowane zostanie przy wykorzystaniu testu Kołmogorowa-Smirnowa, natomiast do wykonania zadania drugiego posłuży taksonomia wrocławska.

Prezentowane w artykule obliczenia przeprowadzono na podstawie nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych prowadzonego przez GUS[2]. Szczegółowe informacje o budżetach gospodarstw domowych w 2012 r. dotyczą 37427 gospodarstw. Prowadzone przez GUS badanie budżetów metodą reprezentacyjną pozwala na uogólnienie uzyskanych wyników na wszystkie gospodarstwa domowe w Polsce[3].

---

[1] Dochód rozporządzalny zdefiniowano (za GUS) jako sumę bieżących dochodów gospodarstwa domowego z poszczególnych źródeł pomniejszoną: o zaliczki na podatek dochodowy od osób fizycznych płacone przez płatnika w imieniu podatnika, o podatki od dochodów z własności, o podatki płacone przez osoby pracujące na własny rachunek oraz o składki na ubezpieczenia społeczne i zdrowotne. W skład dochodu rozporządzalnego wchodzą dochody pieniężne i niepieniężne, w tym spożycie naturalne (towary i usługi konsumpcyjne pobrane na potrzeby gospodarstwa domowego z gospodarstwa indywidualnego w rolnictwie bądź prowadzonej działalności gospodarczej na własny rachunek) oraz towary i usługi otrzymane nieodpłatnie. Dochód rozporządzalny przeznaczany jest na wydatki oraz przyrost oszczędności (*Budżety*..., 2013, s. 18).

[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 Uniwersytetem Szczecińskim.

[3] *Budżety*... (2013), s. 13.

PODSTAWOWE INFORMACJE O TEŚCIE KOŁMOGOROWA-SMIRNOWA

Test Kołmogorowa-Smirnowa jest nieparametrycznym testem istotności, który służy do weryfikacji hipotezy stanowiącej, że dwie próby pochodzą z tej samej populacji (albo inaczej — że dwie populacje mają ten sam rozkład).

Niech rozpatrywana zmienna oznaczona zostanie przez *X*. Dystrybuanta $F(X)$ w pełni określa rozkład zmiennej *X* w populacji. Stąd porównanie rozkładu zmiennej w dwóch populacjach można sprowadzić do porównania wartości dystrybuant w tych populacjach. Jeżeli dwie populacje mają ten sam rozkład, to wartości ich dystrybuant powinny być we wszystkich punktach identyczne. Gdy więc populacja pierwsza oznaczona zostanie subskryptem 1, a druga — subskryptem 2, to aby udowodnić, że dwie populacje mają jednakowy rozkład, należy sprawdzić hipotezę zerową:

$$H_0 : F_1(x_i) = F_2(x_i) \quad \text{dla każdej wartości zmiennej } X$$

wobec hipotezy alternatywnej:

$$H_1 : F_1(x_i) \neq F_2(x_i) \quad \text{dla przynajmniej jednej wartości zmiennej } X$$

gdzie *i* oznacza numer kolejnej obserwacji zmiennej *X*.

Wynika z tego, że jeśli dwie próby pochodzą z jednej populacji (albo z dwóch identycznych populacji), to wartości dystrybuant empirycznych (inaczej — doświadczalnych) obliczonych dla tych prób powinny być we wszystkich punktach zbliżone.

Niech pierwsza próba pobrana z populacji liczy n_1 elementów, a druga — n_2 elementów. Wówczas przez $F_{n_1}(x_i)$ i $F_{n_2}(x_i)$ oznaczone zostaną wartości dystrybuant empirycznych odpowiednio dla pierwszej i drugiej próby w punkcie x_i. Przedmiotem analizy w rozpatrywanym teście są wielkości różnic pomiędzy wartościami tych dystrybuant. W celu określenia tych różnic wszystkie obserwacje występujące w próbach porządkuje się w kolejności niemalejącej. Następnie dla każdej *i*-tej obserwacji oblicza się wartości obu dystrybuant empirycznych odpowiednio według wzorów:

$$F_{n_1}(x_i) = \frac{n_{1sk.}(x_i)}{n_1} \qquad F_{n_2}(x_i) = \frac{n_{2sk.}(x_i)}{n_2} \tag{1}$$

gdzie:

$F_{n_1}(x_i)$ i $F_{n_2}(x_i)$ — wartości dystrybuant ustalanych na podstawie próbek,

$n_{1sk.}(x_i)$ i $n_{2sk.}(x_i)$ — liczebność skumulowana liczona odpowiednio dla pierwszej i drugiej próbki.

W następnym kroku przeprowadzania testu dla każdej wartości zmiennej X obliczana jest różnica pomiędzy dystrybuantami i szuka się największej wartości bezwzględnej różnicy między $F_{n_1}(x_i)$ i $F_{n_2}(x_i)$. Znaleziona wartość niech oznaczona zostanie przez D_{12}. Wartość statystyki D_{12} będzie zatem zdefiniowana jako[4]:

$$D_{12} = \max_{x_i} \left| F_{n_1}(x_i) - F_{n_2}(x_i) \right| \tag{2}$$

Następnie na podstawie statystyki D wyznacza się statystykę λ wyrażoną wzorem:

$$\lambda_{12} = D_{12}\sqrt{n_{12}} \tag{3}$$

gdzie[5]

$$n_{12} = \frac{n_1 n_2}{n_1 + n_2} \tag{4}$$

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 — bez względu na postać poszczególnych dystrybuant — ma asymptotyczny rozkład λ Kołmogorowa[6]. Z tablicy tego rozkładu, dla przyjętego z góry poziomu istotności α, należy odczytać wartość krytyczną λ_α, tak aby spełnione było równanie:

$$P\{\lambda \geq \lambda_\alpha\} = \alpha \tag{5}$$

Następnie trzeba porównać obliczoną wartość λ z wartością krytyczną λ_α i jeżeli zachodzi nierówność $\lambda \geq \lambda_\alpha$, to hipotezę H_0 odrzuca się na rzecz hipotezy alternatywnej. Oznacza to, że obie próby nie pochodzą z tej samej populacji (albo inaczej — dwie populacje, z których pochodzą próby mają inny rozkład). Z kolei, gdy spełniona jest nierówność $\lambda < \lambda_\alpha$, to stwierdza się brak podstaw do odrzucenia hipotezy zerowej o jednakowych rozkładach zmiennej X w obydwu populacjach.

---

[4] Kot i in. (2007), s. 267.
[5] Krysicki i in. (2003), s. 122.
[6] Jóźwiak, Podgórski (1995), s. 193.

Przedmiotem analizy jest cecha ilościowa X oznaczająca dochód rozporządzalny na osobę. Postawionym zadaniem jest porównanie rozkładów tej cechy w szesnastu województwach. Podobieństwo rozkładów będzie mierzone za pomocą statystyki λ, przy czym dwa rozkłady będą tym bardziej podobne, im λ będzie miała mniejszą wartość.

Dzięki zastosowanej metodzie klasyfikacji zbiór szesnastu województw podzielono na ustaloną liczbę podzbiorów, zwanych klasami, które będą jednorodne pod względem przyjętego kryterium. Pożądany będzie więc taki podział, w którym wartość statystyki λ obliczonej dla dowolnej pary województw należących do tej samej klasy będzie mniejsza niż wartość tej statystyki dotyczącej dowolnej pary województw należących do różnych klas[7].

Jedną z najczęściej stosowanych metod klasyfikacji jest taksonomia wrocławska, zwana także metodą dendrytową (bowiem realizowana jest w sposób graficzny za pomocą dendrytu). Przeprowadzenie tej metody obejmuje trzy etapy[8].

Etap 1. Dla każdego województwa poszukuje się 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[9]:

$$\min_{l} \lambda_{jl} \quad (l \neq j) \tag{6}$$

W następnej kolejności buduje się dendryt (czyli graf). Dendryt taki składa się z wierzchołków i wiązadeł. Konstrukcję dendrytu rozpoczyna się od połączenia każdego województwa (a ściślej — wierzchołka grafu odpowiadającego województwu) z najbardziej do niego podobnym. Niektóre z tak wyznaczonych połączeń mogą być podwójne. W wyniku przeprowadzenia procedury łączenia w odniesieniu do wszystkich rozpatrywanych województw uzyskuje się graf składający się z szesnastu wierzchołków. Graf ten będzie złożony z tzw. skupień pierwszego rzędu, przy czym skupienie to grupa województw połączonych ze sobą za pomocą wiązadeł bezpośrednio albo pośrednio.

Po sfinalizowaniu etapu pierwszego mogą zaistnieć dwie sytuacje. Może się okazać, że zbudowany dendryt będzie grafem spójnym. Otrzyma się wtedy jedno skupienie pierwszego rzędu, w którym wszystkie wierzchołki będą połączone nieprzerwanym ciągiem wiązadeł[10]. 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.

7 Sompolska-Rzechuła (2002), s. 525.
8 Dziechciarz (2002), s. 273.
9 Młodak (2006), s. 77.
10 Piszczała (2000), s. 23.

Etap 2. W etapie tym dla każdego skupienia pierwszego rzędu poszukuje się skupienia najbardziej podobnego spośród wszystkich pozostałych skupień. Jako wartość statystyki λ dotyczącą skupień p i q należałoby przyjąć minimalną wartość tej statystyki obliczoną dla poszczególnych województw należących do tych dwóch skupień, co matematycznie można zapisać następująco:

$$\min_{j,l} \lambda_{jl} \tag{7}$$

gdzie:

$j = 1, 2, ..., n_p$ — województwa należące do skupienia p,

$l = 1, 2, ..., n_q$ — województwa należące do skupienia q,

n_p — liczba województw należących do skupienia p,

n_q — liczba województw należących do skupienia q.

Tak więc na dendrycie każde skupienie pierwszego rzędu łączy się wiązaniem ze skupieniem, które jest do niego najbardziej podobne, przy czym łączenie skupień polega na łączeniu tych obiektów należących do rozpatrywanych skupień, których odległość jest najmniejsza. W rezultacie powstaną tzw. skupienia drugiego rzędu. Procedurę łączenia powtarza się aż do momentu, gdy wszystkie skupienia będą ze sobą połączone i otrzyma się graf spójny. Taki graf będzie składał się z n–1 wiązadeł, gdzie n jest liczbą wierzchołków.

Etap 3. Z kolei w etapie trzecim dokonuje się podziału grafu spójnego. W przypadku gdy ostateczna klasyfikacja ma wyodrębnić k rozłącznych klas, graf spójny należy podzielić na k skupień. W tym celu z otrzymanego dendrytu usuwa się k–1 wiązań odpowiadających największym odległościom.

ZASTOSOWANIE TESTU KOŁMOGOROWA-SMIRNOWA DO SPRAWDZENIA ZGODNOŚCI ROZKŁADÓW W WOJEWÓDZTWACH

Dla każdego gospodarstwa domowego ankietowanego przez GUS w badaniu budżetów gospodarstw domowych za 2012 r. wyznaczono dochód rozporządzalny przypadający na osobę. Informacje zawarte w udostępnionej przez GUS bazie danych pozwoliły także na przyporządkowanie poszczególnych gospodarstw do odpowiednich województw. Dzięki temu możliwe stało się wyodrębnienie szesnastu zbiorowości statystycznych. Następnie postawiono hipotezę zerową — głoszącą, że dystrybuanty rozkładów tej samej cechy X w dwóch województwach są takie same — wobec hipotezy alternatywnej, że są różne. Weryfikację hipotezy zerowej przeprowadzono oddzielnie dla każdej pary województw. W tabl. 1 i 2 podano uzyskane wartości odpowiednio statystyki D i λ.

TABL. 1. WARTOŚCI STATYSTYKI *D* DLA PAR WOJEWÓDZTW

Województwa	DŚ	KP	LB	LS	ŁD	MP	MZ	OP	PK	PL	PM	ŚL	ŚK	WM	WP	ZP
DŚ	0,0000	0,1456	0,2032	0,0790	0,0886	0,1153	0,1143	0,0607	0,2550	0,1156	0,0633	0,0376	0,1500	0,1636	0,1328	0,0769
KP	0,1456	0,0000	0,0784	0,1176	0,0854	0,0569	0,1861	0,1222	0,1285	0,0665	0,1014	0,1359	0,0444	0,0436	0,0368	0,0853
LB	0,2032	0,0784	0,0000	0,1639	0,1318	0,0964	0,2361	0,1683	0,0718	0,1065	0,1621	0,1827	0,0706	0,0469	0,0797	0,1428
LS	0,0790	0,1176	0,1639	0,0000	0,0505	0,0746	0,1732	0,0468	0,2156	0,0733	0,0854	0,0574	0,1090	0,1382	0,0954	0,0585
ŁD	0,0886	0,0854	0,1318	0,0505	0,0000	0,0430	0,1407	0,0526	0,1792	0,0445	0,0567	0,0628	0,0870	0,1070	0,0620	0,0317
MP	0,1153	0,0569	0,0964	0,0746	0,0430	0,0000	0,1777	0,0843	0,1446	0,0381	0,0936	0,0975	0,0480	0,0799	0,0302	0,0699
MZ	0,1143	0,1861	0,2361	0,1732	0,1407	0,1777	0,0000	0,1357	0,2946	0,1492	0,0941	0,1229	0,2155	0,2051	0,1781	0,1262
OP	0,0607	0,1222	0,1683	0,0468	0,0526	0,0843	0,1357	0,0000	0,2194	0,0735	0,0541	0,0458	0,1137	0,1407	0,1067	0,0602
PK	0,2550	0,1285	0,0718	0,2156	0,1792	0,1446	0,2946	0,2194	0,0000	0,1583	0,2180	0,2306	0,1176	0,0965	0,1342	0,1960
PL	0,1156	0,0665	0,1065	0,0733	0,0445	0,0381	0,1492	0,0735	0,1583	0,0000	0,0737	0,0894	0,0702	0,0870	0,0439	0,0531
PM	0,0633	0,1014	0,1621	0,0854	0,0567	0,0936	0,0941	0,0541	0,2180	0,0737	0,0000	0,0716	0,1345	0,1316	0,1022	0,0394
ŚL	0,0376	0,1359	0,1827	0,0574	0,0628	0,0975	0,1229	0,0458	0,2306	0,0894	0,0716	0,0000	0,1306	0,1525	0,1185	0,0686
ŚK	0,1500	0,0444	0,0706	0,1090	0,0870	0,0480	0,2155	0,1137	0,1176	0,0702	0,1345	0,1306	0,0000	0,0536	0,0459	0,1147
WM	0,1636	0,0436	0,0469	0,1382	0,1070	0,0799	0,2051	0,1407	0,0965	0,0870	0,1316	0,1525	0,0536	0,0000	0,0554	0,1081
WP	0,1328	0,0368	0,0797	0,0954	0,0620	0,0302	0,1781	0,1067	0,1342	0,0439	0,1022	0,1185	0,0459	0,0554	0,0000	0,0808
ZP	0,0769	0,0853	0,1428	0,0585	0,0317	0,0699	0,1262	0,0602	0,1960	0,0531	0,0394	0,0686	0,1147	0,1081	0,0808	0,0000

U w a g a. DŚ — dolnośląskie, KP — kujawsko-pomorskie, LB — lubelskie, LS — lubuskie, ŁD — łódzkie, MP — małopolskie, MZ — mazowieckie, OP — opolskie, PK — podkarpackie, PL — podlaskie, PM — pomorskie, ŚL — śląskie, ŚK — świętokrzyskie, WM — warmińsko-mazurskie, WP — wielkopolskie, ZP — zachodniopomorskie.

Ź r ó d ł o: obliczenia własne na podstawie bazy nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych za 2012 r.

TABL. 2. WARTOŚCI STATYSTYKI λ DLA PAR WOJEWÓDZTW

Województwa	DŚ	KP	LB	LS	ŁD	MP	MZ	OP	PK	PL	PM	ŚL	ŚK	WM	WP	ZP
DŚ	0,00	8,28	12,07	3,50	5,36	7,46	8,18	2,80	14,83	5,63	3,69	**2,58**	7,44	8,38	8,73	4,07
KP	8,28	0,00	4,28	4,96	4,74	3,34	11,84	5,35	6,90	3,06	5,45	8,34	**2,07**	2,10	2,19	4,22
LB	12,07	4,28	0,00	7,08	7,62	5,93	15,87	7,56	4,01	5,04	9,06	11,81	3,39	**2,33**	4,97	7,30
LS	3,50	4,96	7,08	0,00	2,20	3,36	8,18	**1,74**	9,21	2,82	3,65	2,66	4,24	5,49	4,33	2,37
ŁD	5,36	4,74	7,62	2,20	0,00	2,70	9,68	2,39	10,16	2,13	3,22	4,15	4,23	5,37	3,94	**1,64**
MP	7,46	3,34	5,93	3,36	2,70	0,00	13,38	3,97	8,70	**1,90**	5,64	6,99	2,44	4,20	2,07	3,81
MZ	8,18	11,84	15,87	8,18	9,68	13,38	0,00	6,73	19,28	7,88	**6,18**	9,99	11,61	11,49	13,68	7,36
OP	2,80	5,35	7,56	**1,74**	2,39	3,97	6,73	0,00	9,74	2,92	2,41	2,22	4,57	5,78	5,07	2,53
PK	14,83	6,90	**4,01**	9,21	10,16	8,70	19,28	9,74	0,00	7,39	11,96	14,55	5,58	4,72	8,17	9,87
PL	5,63	3,06	5,04	2,82	2,13	**1,90**	7,88	2,92	7,39	0,00	3,45	4,61	2,94	3,73	2,21	2,33
PM	3,69	5,45	9,06	3,65	3,22	5,64	6,18	2,41	11,96	3,45	0,00	4,53	6,39	6,44	6,24	**1,99**
ŚL	2,58	8,34	11,81	2,66	4,15	6,99	9,99	**2,22**	14,55	4,61	4,53	0,00	6,86	8,31	8,66	3,88
ŚK	7,44	**2,07**	3,39	4,24	4,23	2,44	11,61	4,57	5,58	2,94	6,39	6,86	0,00	2,33	2,35	5,10
WM	8,38	**2,10**	2,33	5,49	5,37	4,20	11,49	5,78	4,72	3,73	6,44	8,31	2,33	0,00	2,94	4,94
WP	8,73	2,19	4,97	4,33	3,94	**2,07**	13,68	5,07	8,17	2,21	6,24	8,66	2,35	2,94	0,00	4,45
ZP	4,07	4,22	7,30	2,37	**1,64**	3,81	7,36	2,53	9,87	2,33	1,99	3,88	5,10	4,94	4,45	0,00

U w a g a. Jak przy tabl. 1. W każdym wierszu tabl. 2 pogrubioną czcionką zaznaczono najmniejsze dodatnie wartości λ.
Ź r ó d ł o: obliczenia własne na podstawie tabl. 1.

Niech przyjęty z góry poziom istotności α wynosi 0,01. Wówczas odczytana z tablicy granicznego rozkładu λ Kołmogorowa wartość krytyczna dla założonego współczynnika $\alpha = 0{,}01$ 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 trzeba odrzucić. Nie można więc twierdzić, że którakolwiek z rozpatrywanych par województw ma taki sam rozkład dochodu rozporządzalnego na osobę. Oznacza to, że różnice między wartościami dystrybuant empirycznych w próbach były na tyle duże, że przypuszczenie o identyczności rozkładów w obu populacjach zostało odrzucone.

Warto podkreślić, że w stosowanym teście Kołmogorowa-Smirnowa nie trzeba przyjmować jakichkolwiek założeń co do typu rozkładu badanej cechy, ponieważ rozkład statystyki testowej λ nie zależy od postaci rozkładu zmiennej w populacji będącej przedmiotem zainteresowania. Stąd niniejszy test mógł być stosowany bez konieczności określania rodzaju rozkładu dochodu rozporządzalnego *per capita* w poszczególnych województwach.

PODZIELENIE WOJEWÓDZTW NA GRUPY PODOBNE POD WZGLĘDEM ROZKŁADU DOCHODU ROZPORZĄDZALNEGO NA OSOBĘ

Skoro wiadomo, że dochód rozporządzalny na osobę ma w poszczególnych województwach inny rozkład, to warto byłoby sprawdzić, które województwa mają rozkłady podobne. Postawionym zadaniem będzie więc podział zbioru szesnastu województw na takie rozłączne i niepuste podzbiory — zwane klasami — aby województwa należące do tych samych klas były pod względem rozkładów jak najbardziej podobne, a należące do różnych klas były jak najmniej podobne. W tym celu zastosowana zostanie metoda dendrytowa, która przeprowadzona będzie na podstawie wartości statystyki λ.

Wiadomo, że rozkłady w dwóch populacjach są tym bardziej podobne, im mniejszą wartość ma λ. Dodatkowo statystyka λ ma też następujące własności:

1) $\lambda_{jl} \geq 0$,

2) $\lambda_{jj} = 0$,

3) $\lambda_{jl} = \lambda_{lj}$.

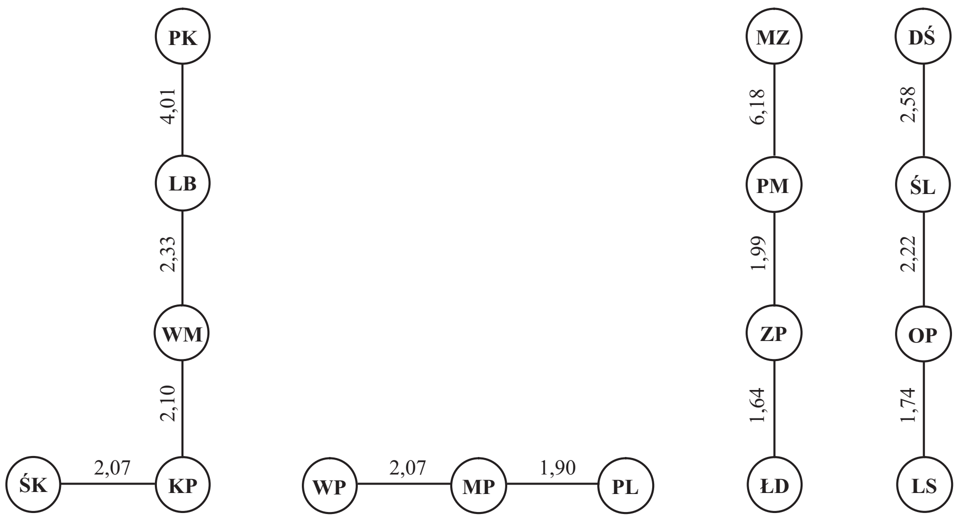
W każdym wierszu tabl. 2 pogrubioną czcionką zaznaczono najmniejsze dodatnie wartości λ. Na podstawie tych liczb sporządzono wykr. 1, na którym województwa (czyli wierzchołki grafu) oznaczono kółkami. Otrzymano cztery skupienia pierwszego rzędu składające się z następujących województw (sposób numerowania poszczególnych skupień jest dowolny):

I — podkarpackie, lubelskie, warmińsko-mazurskie, kujawsko-pomorskie, świętokrzyskie;

II — wielkopolskie, małopolskie, podlaskie;

III — mazowieckie, pomorskie, zachodniopomorskie, łódzkie;
IV — dolnośląskie, śląskie, opolskie, lubuskie.

Wykr. 1. SKUPIENIA PIERWSZEGO RZĘDU



Ź r ó d ł o: opracowanie własne na podstawie tabl. 2.

W następnym etapie należy połączyć ze sobą odpowiednie skupienia pierwszego rzędu. W tym celu dla każdego skupienia poszukuje się skupienia najbardziej podobnego oraz określa się te województwa w poszczególnych skupieniach, które należy ze sobą połączyć. Wartości statystyki λ dla par województw, które należałoby ze sobą połączyć w celu połączenia skupień, do których te województwa należą, znajdują się w tabl. 3.

TABL. 3. WARTOŚCI STATYSTYKI λ DLA POSZCZEGÓLNYCH PAR SKUPIEŃ

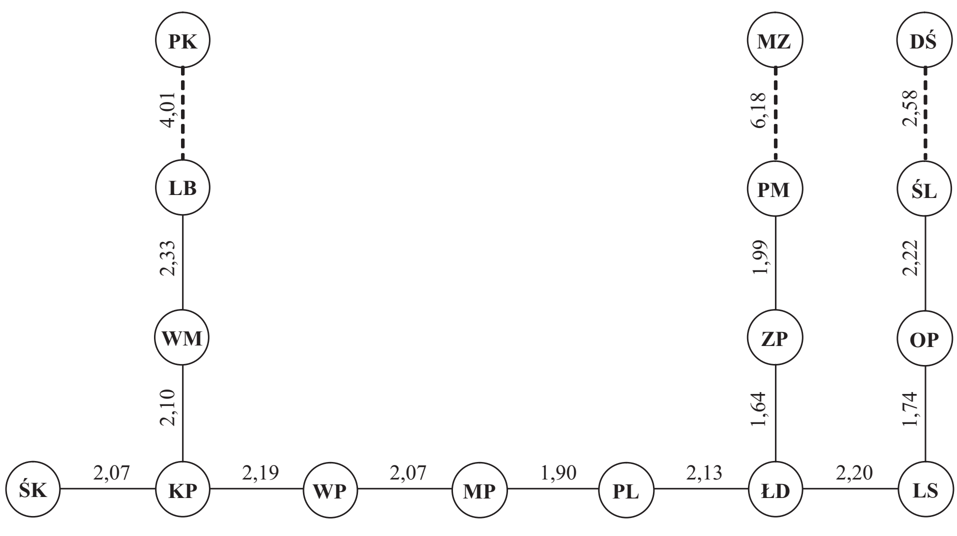
Województwa	PK, LB, WM, KP, ŚK	WP, MP, PL	MZ, PM, ZP, ŁD	DŚ, ŚL, OP, LS
PK, LB, WM, KP, ŚK	0,00	**2,19**	4,22	4,24
WP, MP, PL	2,19	0,00	**2,13**	2,82
MZ, PM, ZP, ŁD	4,22	**2,13**	0,00	2,20
DŚ, ŚL, OP, LS	4,24	2,82	**2,20**	0,00

U w a g a. Jak przy tabl. 1.
Ź r ó d ł o: obliczenia własne na podstawie tabl. 2.

W każdym wierszu tabl. 3 wytłuszczono najniższą niezerową wartość λ, czyli na wykr. 1 należy połączyć ze sobą skupienia:
I z II — woj. kujawsko-pomorskie z woj. wielkopolskim;
II z III (i III z II) — woj. podlaskie z woj. łódzkim;
IV z III — woj. lubuskie z woj. łódzkim.

Powstały w ten sposób graf jest spójny, zatem można już przejść do odpowiedniego podzielenia. W przypadku grupowania województw w dwie jednorodne klasy należałoby z grafu spójnego usunąć jedno najdłuższe wiązanie odpowiadające odległości równej 6,18. Wówczas w jednej grupie znalazłoby się woj. mazowieckie, a w drugiej grupie — pozostałe województwa. W przypadku wyodrębnienia trzech grup należałoby usunąć też wiązanie ilustrujące odległość wynoszącą 4,01 i w ten sposób powstałaby kolejna jednoelementowa grupa z woj. podkarpackim. Z kolei, gdyby celem było wyznaczenie czterech jednorodnych klas, usunięciu uległoby też wiązanie dla dystansu 2,58 i wtedy powstałaby następna jednoelementowa grupa z woj. dolnośląskim. Wydaje się, że przy szesnastu wierzchołkach grafu nie jest zasadne dzielenie go na więcej niż cztery klasy.

Wykr. 2. DENDRYT SPÓJNY[a]



a Linią przerywaną oznaczono trzy wiązania odpowiadające największym odległościom.

Ź r ó d ł o: jak przy wykr. 1.

Ostatecznie po przeprowadzeniu poszczególnych etapów taksonomii wrocławskiej powstały następujące cztery klasy (w nawiasach podano wartości średniego dochodu rozporządzalnego w zł na osobę):

- woj. mazowieckie (1676,14);
- woj. dolnośląskie (1374,23);
- województwa: pomorskie (1359,85), opolskie (1300,14), śląskie (1299,77), zachodniopomorskie (1289,10), podlaskie (1264,26), łódzkie (1260,57), lubu-

skie (1210,77), małopolskie (1187,76), wielkopolskie (1152,77), kujawsko-pomorskie (1131,55), świętokrzyskie (1124,50), warmińsko-mazurskie (1093,56), lubelskie (1068,70);
- woj. podkarpackie (967,77).

Podsumowanie

Celem artykułu była odpowiedź na pytanie, czy w województwach jest taki sam rozkład dochodu rozporządzalnego *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 w każdym ze 120 przypadków przypuszczenie o identyczności rozkładów dochodu rozporządzalnego na osobę w danych województwach trzeba było odrzucić.

Warto też zaznaczyć, że odrzucenie hipotezy zerowej na korzyść hipotezy alternatywnej w teście Kołmogorowa-Smirnowa pozwala jedynie orzec, że różnice w wartościach dystrybuant empirycznych w obu próbach były statystycznie istotne i dlatego rozkłady w populacjach są różne. Nie ma natomiast potrzeby określania charakteru tych różnic. Celem przeprowadzenia testu Kołmogorowa-Smirnowa jest więc wykrycie wszystkich rodzajów różnic łącznie — tak w rodzajach rozkładów, jak i w wartościach parametrów tych rozkładów. Oznacza to, że odrzucenie H_0 może być spowodowane tym, że dwie populacje różnią się typem rozkładu i/lub różnią się parametrami, a rozstrzygnięcie kwestii, czym dokładnie się różnią nie jest potrzebne do realizacji celu tego artykułu. Ważne jest jedynie, aby mieć świadomość istnienia potencjalnej możliwości, że poszczególne populacje mają taki sam typ rozkładu, ale wartości parametrów tych rozkładów na tyle różnią się od siebie, że nie było podstaw do przyjęcia hipotezy zerowej o identyczności rozkładów za prawdziwą.

Realizacja drugiego zadania badawczego pozwoliła na pogrupowanie województw w klasy o rozkładach najbardziej do siebie podobnych. Grupowania tego dokonano na podstawie wartości statystyki λ obliczonej w pierwszym zadaniu badawczym i przy wykorzystaniu metody dendrytowej. W efekcie uzyskano trzy grupy jednoelementowe oraz jedną grupę obejmującą pozostałe trzynaście województw. Okazało się bowiem, że woj. mazowieckie ze średnim dochodem na osobę opiewającym na 1676,14 zł, woj. dolnośląskie z dochodem 1374,23 zł na osobę oraz woj. podkarpackie z dochodem 967,77 zł na osobę mają na tyle różny typ rozkładu i/lub wartość parametrów rozkładu, że nie można ich uznać za podobne ani do siebie nawzajem, ani do jakiegokolwiek innego województwa. Z kolei pozostałe województwa ze średnim dochodem rozporządzalnym *per capita* kształtującym się od 1068,70 do 1359,85 zł można uznać za podobne pod względem rozkładu badanej cechy na tyle, że umieszczone zostają w jednej klasie.

---

dr Anna Turczak — *Zachodniopomorska Szkoła Biznesu w Szczecinie*, **dr Patrycja Zwiech** — *Uniwersytet Szczeciński*

LITERATURA

Budżety gospodarstw domowych w 2012 r. (2013), GUS

Dziechciarz J. (red.) (2002), *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław

Jóźwiak J., Podgórski J. (1995), *Statystyka od podstaw*, Państwowe Wydawnictwo Ekonomiczne, Warszawa

Kot S. M., Jakubowski J., Sokołowski A. (2007), *Statystyka. Podręcznik dla studiów ekonomicznych*, DIFIN, Warszawa

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. (2003), *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Część II: Statystyka matematyczna*, Wydawnictwo Naukowe PWN, Warszawa

Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, DIFIN, Warszawa

Piszczała J. (2000), *Matematyka i jej zastosowanie w naukach ekonomicznych*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań

Sompolska-Rzechuła A. (2002), *Zastosowanie taksonomii rozmytej do klasyfikacji spółek na Giełdzie Papierów Wartościowych*, [w:] *Rynek kapitałowy. Skuteczne inwestowanie. Część I*, Konferencja naukowa zorganizowana przez Katedrę Ekonometrii i Statystyki Wydziału Nauk Ekonomicznych i Zarządzania, Międzyzdroje

SUMMARY

The paper considered the question of whether the distribution of disposable income per capita in the Polish voivodships (provinces) is similar. To study this, the authors put appropriate statistical hypotheses and verified using the Kolmogorov-Smirnov test. Conducting this test allowed them to conclude that there are no grounds to assert that the hypothesis of identical distributions of disposable income per capita is true. The authors determined only the degree of similarity of individual schedules and on this basis, divided the provinces into four homogeneous classes. Thanks to the Wroclaw taxonomy there has been shown that the conduct of the said division results in separation of the three one-piece groups, and one group of thirteen provinces.

РЕЗЮМЕ

В статье рассматривается вопрос, является ли таким самым распределение чистого дохода на душу населения в воеводствах? Для этой цели были поставлены статистические гипотезы и они были проверены с использованием теста Колмогорова-Смирнова. Проведение этого теста позволило прийти к выводу, что нет никаких оснований палагать, что гипотеза по идентичности распределения чистого дохода на душу населения является верной. Была лишь определена степень сходства отдельных распределений и на этой основе воеводства были разделены на четыре равномерных класса. Благодаря использованию вроцлавской таксономии было показано, что проведение обсуждаемого разделения позволило получить результат выделения трех одноэлементных групп и одной группы составленной из тринадцати воеводств.