UNIWERSYTET SZCZECIŃSKI
WYDZIAŁ NAUK EKONOMICZNYCH I ZARZĄDZANIA

STUDIA I PRACE WYDZIAŁU NAUK EKONOMICZNYCH I ZARZĄDZANIA NR 42, t. 1

DOI: 10.18276/sip.2015.42/1-06

Anna Turczak[*]
Zachodniopomorska Szkoła Biznesu w Szczecinie
Patrycja Zwiech[**]
Uniwersytet Szczeciński

GRUPOWANIE ROZKŁADÓW W JEDNORODNE KLASY NA PRZYKŁADZIE KLASYFIKACJI GOSPODARSTW DOMOWYCH O RÓŻNYM TYPIE BIOLOGICZNYM

Streszczenie

Celem artykułu jest zaprezentowanie nowego sposobu dzielenia rozkładów na jednorodne grupy oraz wykorzystanie go do podziału typów biologicznych gospodarstw domowych na klasy o jak najbardziej podobnych rozkładach. Jako miernik stopnia podobieństwa rozkładów wykorzystano statystykę λ (lambda), która jest oparta na maksymalnej bezwzględnej wartości różnicy między dwiema dystrybuantami empirycznymi. Na podstawie wartości statystyki λ obliczonej dla każdej z par rozkładów trzynaście typów biologicznych gospodarstw domowych podzielono na osiem jednolitych klas. Podział ten skutkował utworzeniem pięciu grup jednoelementowych, dwóch grup dwuelementowych oraz jednej grupy czteroelementowej.

Słowa kluczowe: dochód rozporządzalny, gospodarstwo domowe, test Kołmogorowa-Smirnowa, taksonomia

---

[*] Adres e-mail: aturczak@zpsb.szczecin.pl.
[**] Adres e-mail: patrycjazwiech@tlen.pl.

Wprowadzenie

W debacie publicznej często powraca temat konieczności zintensyfikowania działań państwa mających na celu wsparcie ekonomiczne uboższej części społeczeństwa (Domański et al. 2012, s. 115). Bardzo mocno zakorzenione jest w świadomości społecznej przeświadczenie o tym, że małżeństwa (jak i osoby samotne) wychowujące liczne potomstwo są tą grupą, która w Polsce najbardziej jest dotknięta ubóstwem. Teza taka jest głoszona również przez naukowców zajmujących się problemem ogromnych – jak nierzadko są one określane – różnic w poziomie życia poszczególnych grup społecznych (Kołodko 2014, s. 35). Warto byłoby więc porównać ze sobą gospodarstwa domowe o różnym typie biologicznym i wskazać typy gospodarstw rzeczywiście drastycznie odbiegające od pozostałych, którym powodzi się lepiej. Wydaje się, że analizę taką należałoby oprzeć przede wszystkim na zestawieniu poziomu osiąganego dochodu przez poszczególne gospodarstwa domowe, gdyż dochód jest bezsprzecznie tym czynnikiem, którego wpływ na jakość życia jest kluczowy (Bal 2012, s. 254). Stąd określenie rodzajów[1] gospodarstw domowych, które charakteryzują się bardzo podobnym rozkładem dochodu rozporządzalnego[2] na osobę, oraz tych rodzajów gospodarstw, które pod względem rozkładu rozpatrywanej zmiennej odbiegają od pozostałych, stało się celem niniejszego artykułu. Celowi temu służyć będzie realizacja następujących zadań badawczych:

– wyznaczenie wartości statystyki pozwalającej na stwierdzenie, czy rozkład dochodu rozporządzalnego na osobę w analizowanych typach biologicznych gospodarstw domowych jest taki sam,
– podzielenie rodzajów gospodarstw na grupy[3] o identycznym rozkładzie badanej zmiennej,

[1] Określenia „typy biologiczne”, „rodzaje” i „kategorie” będą używane w niniejszym artykule zamiennie.

[2] 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, 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 (tj. 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 gospodarstw*… 2014, s. 19).

[3] Określenia „grupy” i „klasy” będą używane zamiennie.

– obliczenie wartości klasycznych miar tendencji centralnej, zróżnicowania i asymetrii opisujących rozkłady odnoszące się do grup wyodrębnionych w ramach drugiego zadania.

Artykuł niniejszy ma charakter badawczy. Wszystkie zawarte w nim obliczenia przeprowadzono na podstawie nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych zrealizowanego przez GUS[4]. Wspomniana baza zawiera szczegółowe informacje dotyczące budżetów gospodarstw domowych w Polsce i w 2012 r. objęła 37 427 gospodarstw. Co warte podkreślenia, badanie budżetów przez Główny Urząd Statystyczny prowadzone jest metodą reprezentacyjną, która daje możliwość uogólnienia uzyskanych wyników na wszystkie gospodarstwa domowe w Polsce (*Budżety gospodarstw*... 2014, s. 14).

1. Metodyka badania

W przeprowadzonych badaniach wykorzystano oryginalny i nieznany wcześniej konglomerat narzędzi statystycznych, a mianowicie test Kołmogorowa-Smirnowa[5] 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). Z kolei taksonomia wrocławska jest znaną metodą klasyfikacji.

Rozpatrywaną zmienną (tj. miesięczny dochód rozporządzalny *per capita*) oznaczono przez X. Dystrybuanta $F(x)$ w pełni opisuje rozkład zmiennej X w populacji (Kot et al. 2007, s. 267). Z tego powodu porównanie rozkładu zmiennej w dwóch populacjach można sprowadzić do porównania wartości dystrybuant w tych populacjach i jeżeli 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 sprawdzić hipotezę zerową (Razali, Wah 2011, s. 23):

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

[4] 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.

[5] Autorki wykorzystały test Kołmogorowa-Smirnowa we wcześniejszych badaniach przedstawionych w pracach Turczak, Zwiech 2015a oraz Turczak, Zwiech 2015b.

wobec hipotezy alternatywnej:

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

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:

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

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 niemalejącej. Następnie dla każdej *i*-tej obserwacji obliczono wartości obu dystrybuant odpowiednio według wzorów:

$$F_{n1}(x_i) = \frac{n_{1sk.}(x_i)}{n_1} \qquad F_{n2}(x_i) = \frac{n_{2sk.}(x_i)}{n_2} \tag{1}$$

gdzie $n_{1sk.}(x_i)$ i $n_{2sk.}(x_i)$ oznaczają liczebności skumulowane liczone odpowiednio dla pierwszej i drugiej próbki.

W następnym kroku realizacji omawianej procedury dla każdej wartości zmiennej *X* obliczono wartość bezwzględną z różnicy między dystrybuantami. Znalezioną największą bezwzględną wartość takiej różnicy oznacza się przez *D* i definiuje jako (Arnold, Emerson 2011, s. 34):

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

Na podstawie statystyki D wyznaczono statystykę λ wyrażoną wzorem:

$$\lambda = D\sqrt{n_{12}} \tag{3}$$

gdzie (Rószkiewicz 2012, s. 304):

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

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.

W celu porównania rozkładów cechy X (tj. miesięcznego dochodu rozporządzalnego na osobę) zbudowanych dla poszczególnych typów biologicznych gospodarstw domowych i podzielenia ich na ustaloną liczbę klas, które to klasy będą jednorodne pod względem przyjętego kryterium, wykorzystano taksonomię wrocławską, zwaną także metodą dendrytową. Taksonomię wrocławską zrealizowano w następujących trzech etapach (Dziechciarz 2002, s. 273):

Etap 1. Na podstawie wartości statystyki λ dla każdej kategorii gospodarstw domowych znaleziono kategorię najbardziej podobną. Na tej podstawie zbudowano dendryt składający się z wierzchołków i wiązadeł, przy czym każdy z wierzchołków grafu odpowiada innemu typowi biologicznemu. Konstrukcję dendrytu rozpoczęto od połączenia wszystkich kategorii z najbardziej podobnymi. Uzyskano graf złożony ze skupień pierwszego rzędu[6]. Gdyby okazało się, że utworzony dendryt jest grafem spójnym (czyli otrzymano jedno skupienie pierwszego rzędu, w którym wszystkie wierzchołki połączono nieprzerwanym ciągiem wiązadeł) (Piszczała 2000, s. 23), to po etapie pierwszym należałoby przejść bezpośrednio do etapu trzeciego. Jeśli natomiast w etapie pierwszym otrzymano co najmniej dwa skupienia pierwszego rzędu, należy przeprowadzić etap drugi.

[6] Skupienie takie to grupa kategorii połączonych ze sobą za pomocą wiązadeł bezpośrednio albo pośrednio.

Etap 2. W etapie tym dla każdego skupienia pierwszego rzędu poszukiwano skupienia najbardziej podobnego spośród wszystkich pozostałych skupień. Jako wartość statystyki λ odnoszącą się do pary skupień przyjęto minimalną wartość tej statystyki obliczoną dla poszczególnych kategorii gospodarstw domowych należących do tych dwóch skupień (Młodak 2006, s. 77). W rezultacie połączenia każdego skupienia pierwszego rzędu ze skupieniem, które jest do niego najbardziej podobne, uformowano skupienia drugiego rzędu. Procedurę łączenia powtarzano aż do momentu, w którym wszystkie skupienia były ze sobą połączone i otrzymany graf był spójny.

Etap 3. W tym etapie podzielono graf spójny. W tym celu należało określić te wiązadła, do których przyporządkowano wartość statystyki λ większą lub równą wartości krytycznej (λ_α). Z tablicy rozkładu λ Kołmogorowa dla przyjętego z góry poziomu istotności α odczytano taką wartość krytyczną, aby spełnione było równanie $P\{\lambda \geq \lambda_\alpha\} = \alpha$ (Witkowski 2010, s. 92). Wartości λ przyporządkowane poszczególnym wiązadłom porównano ze znalezioną wartością λ_α. W przypadku zajścia nierówności $\lambda \geq \lambda_\alpha$ hipotezę zerową należało odrzucić na rzecz hipotezy alternatywnej, co było równoznaczne ze stwierdzeniem, że rozpatrywane próby nie pochodzą z tej samej populacji (albo inaczej – populacje, z których pochodzą próby, mają inny rozkład). Natomiast gdy spełniona była nierówność $\lambda < \lambda_\alpha$, wówczas nie było podstaw do odrzucenia H_0 o identyczności rozkładów. Ponieważ celem przeprowadzanego badania było podzielenie rozpatrywanych typów gospodarstw domowych na jednorodne klasy, toteż w ostatnim etapie realizowanej procedury z otrzymanego dendrytu usunięto wyłącznie te wiązadła, które odpowiadały wartościom λ większym bądź równym λ_α.

2. Podział gospodarstw domowych na grupy o jednakowym rozkładzie dochodu rozporządzalnego na osobę

Dla każdego gospodarstwa domowego ankietowanego przez GUS w ramach badania budżetów gospodarstw domowych za 2012 r. wyznaczono średni miesięczny dochód przypadający na osobę. Informacje z bazy danych GUS pozwoliły także na przyporządkowanie poszczególnych gospodarstw do odpowiednich typów biologicznych. Dzięki temu wyodrębniono trzynaście następujących zbiorowości statystycznych: A – gospodarstwa jednoosobowe, B – małżeństwa (albo osoby żyjące w związkach nieformalnych) bez dzieci, C – małżeństwa (albo osoby żyjące w związkach nieformalnych) z 1 dzieckiem, D – małżeństwa (albo osoby żyjące w związkach nieformalnych) z 2 dzieci, E – małżeństwa (albo osoby żyjące w związkach nieformalnych) z 3 dzieci, F – małżeństwa (albo osoby żyjące w związkach nieformalnych) z co najmniej 4 dzieci, G – matki z dziećmi, H – ojcowie z dziećmi, I – małżeństwa (albo osoby żyjące w związkach nieformalnych) z dziećmi i innymi osobami, J – matki z dziećmi i innymi osobami, K – ojcowie z dziećmi i innymi osobami, L – inne osoby z dziećmi na utrzymaniu, M – pozostałe gospodarstwa domowe.

Postawiona hipoteza zerowa głosi, że dystrybuanty rozkładów miesięcznych dochodów rozporządzalnych na osobę w dwóch kategoriach gospodarstw domowych są takie same, a hipoteza alternatywna, że są różne. W tabelach 1 i 2 podano wartości statystyk D i λ, których znajomość jest konieczna do przeprowadzenia weryfikacji hipotezy H_0.

W celu podzielenia zbioru {A, B, C, …, M} na takie rozłączne i niepuste grupy, aby kategorie gospodarstw domowych należące do tych samych grup były jak najbardziej do siebie podobne, a kategorie należące do różnych grup były jak najmniej do siebie podobne, na podstawie liczb z tabeli 2 sporządzono graf, na którym poszczególne kategorie gospodarstw domowych (tj. wierzchołki grafu) oznaczono kółkami. Otrzymano trzy skupienia pierwszego rzędu, które połączono, wykorzystując dane z tabeli 3.

Tabela 1. Wartości statystyki *D* dla poszczególnych par kategorii gospodarstw domowych

D	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0,0000	0,0241	0,1891	0,3975	0,6132	0,7861	0,4866	0,4710	0,5038	0,5754	0,5959	0,6144	0,2317
B	0,0241	0,0000	0,1789	0,3899	0,5977	0,7731	0,4732	0,4621	0,4922	0,5649	0,5872	0,5997	0,2272
C	0,1891	0,1789	0,0000	0,2172	0,4489	0,6700	0,3259	0,2927	0,3261	0,4041	0,4211	0,4501	0,0597
D	0,3975	0,3899	0,2172	0,0000	0,2660	0,5289	0,1523	0,1198	0,1206	0,1963	0,2268	0,2572	0,1751
E	0,6132	0,5977	0,4489	0,2660	0,0000	0,3022	0,1372	0,2511	0,1692	0,1049	0,1171	0,0513	0,4182
F	0,7861	0,7731	0,6700	0,5289	0,3022	0,0000	0,3917	0,4792	0,4415	0,3842	0,4109	0,2860	0,6529
G	0,4866	0,4732	0,3259	0,1523	0,1372	0,3917	0,0000	0,1236	0,0717	0,1094	0,1262	0,1418	0,2975
H	0,4710	0,4621	0,2927	0,1198	0,2511	0,4792	0,1236	0,0000	0,1234	0,1980	0,2099	0,2303	0,2422
I	0,5038	0,4922	0,3261	0,1206	0,1692	0,4415	0,0717	0,1234	0,0000	0,0907	0,1332	0,1670	0,2873
J	0,5754	0,5649	0,4041	0,1963	0,1049	0,3842	0,1094	0,1980	0,0907	0,0000	0,0565	0,1190	0,3664
K	0,5959	0,5872	0,4211	0,2268	0,1171	0,4109	0,1262	0,2099	0,1332	0,0565	0,0000	0,1459	0,3877
L	0,6144	0,5997	0,4501	0,2572	0,0513	0,2860	0,1418	0,2303	0,1670	0,1190	0,1459	0,0000	0,4128
M	0,2317	0,2272	0,0597	0,1751	0,4182	0,6529	0,2975	0,2422	0,2873	0,3664	0,3877	0,4128	0,0000

Źródło: obliczenia własne na podstawie bazy nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych za 2012 r.

Tabela 2. Wartości statystyki λ dla poszczególnych par kategorii gospodarstw domowych

λ	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0,000	**1,713**	12,670	27,949	34,052	33,079	18,658	5,558	35,867	28,227	9,814	26,115	16,604
B	**1,713**	0,000	15,149	35,753	38,505	35,235	19,370	5,484	46,097	30,989	9,782	27,658	21,516
C	12,670	15,149	0,000	18,140	27,556	29,791	13,073	**3,467**	27,714	21,397	6,991	20,237	5,122
D	27,949	35,753	18,140	0,000	16,994	24,004	6,215	**1,422**	11,101	10,705	3,777	11,813	16,290
E	34,052	38,505	27,556	16,994	0,000	12,260	5,102	2,954	10,924	4,899	**1,917**	2,102	27,133
F	33,079	35,235	29,791	24,004	12,260	0,000	12,620	**5,546**	20,144	14,533	6,521	9,887	29,866
G	18,658	19,370	13,073	6,215	5,102	12,620	0,000	**1,420**	2,935	3,833	1,975	4,595	12,212
H	5,558	5,484	3,467	1,422	2,954	5,546	**1,420**	0,000	1,464	2,315	2,040	2,668	2,875
I	35,867	46,097	27,714	11,101	10,924	20,144	2,935	**1,464**	0,000	4,985	2,220	7,711	27,322
J	28,227	30,989	21,397	10,705	4,899	14,533	3,833	2,315	4,985	0,000	**0,914**	4,540	20,204
K	9,814	9,782	6,991	3,777	1,917	6,521	1,975	2,040	2,220	**0,914**	0,000	2,318	6,461
L	26,115	27,658	20,237	11,813	**2,102**	9,887	4,595	2,668	7,711	4,540	2,318	0,000	19,108
M	16,604	21,516	5,122	16,290	27,133	29,866	12,212	**2,875**	27,322	20,204	6,461	19,108	0,000

Źródło: obliczenia własne na podstawie tabeli 1.

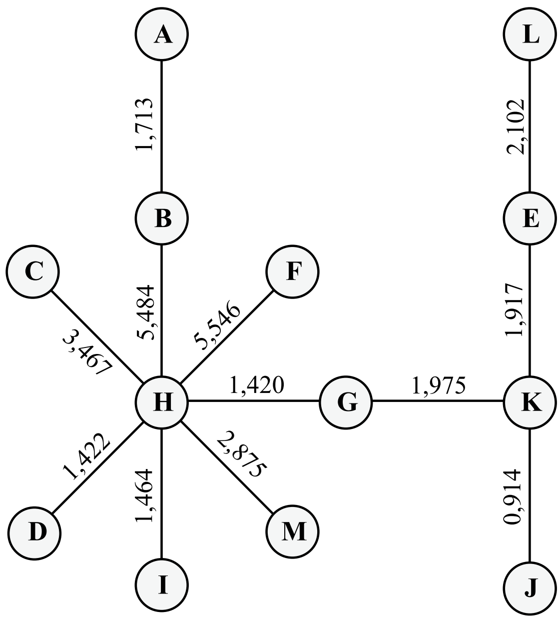
Tabela 3. Wartości statystyki λ dla poszczególnych par skupień pierwszego rzędu

λ	(A, B)	(C, D, F, G, H, I, M)	(E, J, K, L)
(A, B)	0,000	**5,484**	9,782
(C, D, F, G, H, I, M)	5,484	0,000	**1,975**
(E, J, K, L)	9,782	**1,975**	0,000

Źródło: obliczenia własne na podstawie tabeli 2.

Połączono skupienie (A, B) ze skupieniem (C, D, F, G, H, I, M), a potem skupienie (C, D, F, G, H, I, M) ze skupieniem (E, J, K, L). Powstały w ten sposób graf był spójny. Zaprezentowano go na rysunku 1.

Rysunek 1. Dendryt spójny

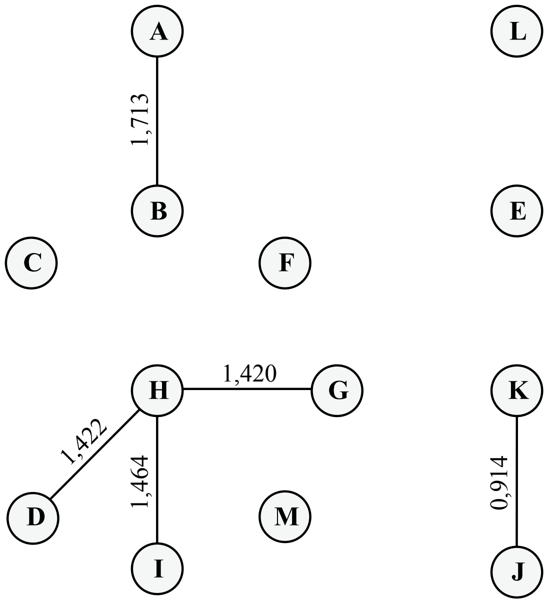


Źródło: opracowanie własne na podstawie tabel 2 i 3.

Kolejnym etapem było odpowiednie podzielenie otrzymanego dendrytu spójnego, dla poziomu istotności α wynoszącego 0,005, $\lambda_{\alpha} = 1,73$. Ponieważ dla siedmiu wiązań dendrytu spójnego otrzymano relację: $\lambda \geq \lambda_{\alpha}$ (wartość statystyki λ znalazła się w obszarze krytycznym), to hipotezę H_0 odrzucono. Nie można więc twierdzić, że w przypadku par: (F, H), (B, H), (C, H), (M, H), (E, L), (G, K) i (E, K), jest taki sam rozkład dochodu rozporządzalnego na osobę – różnice między wartościami dystrybuant empirycznych w próbach były na tyle duże, że przypuszczenie o identyczności odrzucono. Z kolei w przypadku pięciu wiązań spełniona została nierówność: $\lambda < \lambda_\alpha$, czyli wartość statystyki λ nie znalazła się w obszarze krytycznym. W odniesieniu do par: (J, K), (G, H), (D, H), (H, I) i (A, B), brak było zatem podstaw do odrzucenia hipotezy zerowej o identycznym rozkładzie dochodu rozporządzalnego na osobę.

Ostatecznie w drodze przeprowadzenia metody dendrytowej – zrealizowanej na podstawie wartości statystyki λ – powstało osiem klas, które zaprezentowano na rysunku 2.

Rysunek 2. Podział dendrytu na osiem klas



Źródło: opracowanie własne na podstawie rysunku 1.

W tabeli 4 podano wartości klasycznych miar tendencji centralnej, dyspersji i skośności, które to miary syntetycznie opisują rozkłady zbudowane dla wyodrębnionych ośmiu grup (grupy w tabeli 4 – z wyjątkiem ostatniej – uporządkowano według malejącej wartości średniej arytmetycznej).

Tabela 4. Średnia arytmetyczna, odchylenie standardowe, klasyczny współczynnik zmienności oraz klasyczny współczynnik asymetrii dla każdej z grup

Wyszczególnienie	Średnia arytmetyczna [zł/osobę]	Odchylenie standardowe [zł/osobę]	Klasyczny współczynnik zmienności [%]	Klasyczny współczynnik asymetrii
Gospodarstwa jednoosobowe oraz małżeństwa* bez dzieci	1799,20	1428,69	79,41	8,97
Małżeństwa* z 1 dzieckiem	1441,73	1024,91	71,09	1,33
Małżeństwa* z 2 dzieci, matki z dziećmi, ojcowie z dziećmi, mał-żeństwa* z dziećmi i innymi osobami	1028,09	847,28	82,41	1,26
Małżeństwa* z 3 dzieci	854,51	975,31	114,14	2,67
Matki z dziećmi i innymi osobami, ojcowie z dziećmi i innymi osobami	851,36	479,09	56,27	0,99
Inne osoby z dziećmi na utrzymaniu	820,26	1083,85	132,13	3,52
Małżeństwa* z 4 (i więcej) dzieci	566,10	361,42	63,84	0,36
Pozostałe gospodarstwa domowe	1325,00	1212,37	91,50	5,96

* Albo osoby żyjące w związkach nieformalnych.

Źródło: obliczenia własne na podstawie bazy nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych za 2012 r.

Z tabeli 4 wynika, że małżeństwa (albo osoby żyjące w związkach nieformalnych) z co najmniej czwórką dzieci są tą grupą społeczną, w której średni dochód rozporządzalny jest na poziomie najniższym ze wszystkich rozpatrywanych grup. Co zaskakujące, wspomniana grupa charakteryzuje się również względnie małym zróżnicowaniem dochodu *per capita* (odchylenie standardowe stanowi tutaj 63,84% średniej i jest to jedna z najniższych wartości znajdujących się w przedostatniej kolumnie tabeli 4). Dodatkowo jedynie w tej grupie gospodarstw domowych rozkład rozpatrywanej zmiennej jest prawie symetryczny (tj. tylko w przypadku tej grupy wykazana asymetria prawostronna jest asymetrią nieznaczną). Można więc orzec, że około 50% małżeństw (lub osób żyjących w związkach nieformalnych) z czwórką i większą liczbą dzieci ma dochód na osobę niższy niż 566 zł.

Na podstawie przeprowadzonych badań można stwierdzić, że rodziny wielodzietne są rzeczywiście tą grupą społeczną, której dochody w przeliczeniu na osobę są na bardzo niskim poziomie. Typowe małżeństwo (albo osoby żyjące w związkach nieformalnych) z przynajmniej czwórką dzieci ma dochód na osobę zawierający się w przedziale (204,68; 927,52)[7] i rozpiętość tego przedziału jest stosunkowo mała, jeśli porównać ją z rozpiętością typowego przedziału zmienności wyznaczonego dla pozostałych grup społecznych. Grupę rodzin wielodzietnych należałoby więc wskazać jako relatywnie najuboższą w Polsce.

Podsumowanie

Celem artykułu było określenie tych typów biologicznych gospodarstw domowych, które charakteryzują się identycznym rozkładem dochodu rozporządzalnego na osobę. Aby tego dokonać, wyznaczono wartości statystyki λ będącej miarą podobieństwa rozkładów, przy czym dwa rozkłady są tym bardziej podobne, im λ ma mniejszą wartość. Trzynaście typów gospodarstw domowych pogrupowano w osiem klas o takich samych rozkładach badanej zmiennej, przy wykorzystaniu metody dendrytowej. W efekcie uzyskano pięć grup jednoelementowych: (C), (E), (F), (L) i (M), dwie grupy dwuelementowe: (A, B) i (J, K), oraz jedną grupę czteroelementową: (D, G, H, I).

Ostatnim etapem przeprowadzonych badań było scharakteryzowanie każdej z otrzymanych klas za pomocą klasycznych miar struktury. Dzięki temu zidentyfikowano tę kategorię gospodarstw domowych, tj. małżeństwa (albo osoby żyjące w związkach nieformalnych) z przynajmniej czwórką dzieci, która jest szczególnie narażona na ubóstwo (Panek 2011, s. 155) i najpewniej wymagałaby finansowego wsparcia państwa (Pliszka 2004, s. 360)[8].

[7] Końce tego przedziału obliczono jako 566,10 zł/os. (średnia) ±361,42 zł/os. (odchylenie standardowe).

[8] Gospodarstw domowych należących do tej klasy jest 0,97% i stanowią one 2,24% ludności Polski (obliczenia własne na podstawie bazy nieidentyfikowalnych danych jednostkowych z badania budżetów gospodarstw domowych za 2012 r.).

Literatura

Arnold T.B., Emerson J.W. (2011), *Nonparametric goodness-of-fit tests for discrete null distributions*, „The R Journal", vol. 2, nr 3.

Bal I. (2012), *Marginalizacja i wykluczenie społeczne jako bariera rozwoju regionalnego*, „Nierówności Społeczne a Wzrost Gospodarczy", nr 28.

Budżety gospodarstw domowych w 2013 r. (2014), GUS, Warszawa.

Domański H., Karpiński Z., Pokropek A., Przybysz D., Sawiński Z., Słomczyński K.M., Trzciński R. (2012), *Metodologia badań nad stratyfikacją społeczną*, Wydawnictwo Naukowe Scholar, Warszawa.

Dziechciarz J. (red.) (2002), *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna im. Oskara Langego we Wrocławiu, Wrocław.

Kołodko G. (2014), *Społeczne i przestrzenne aspekty zróżnicowania dochodów we współczesnym świecie*, „Nierówności Społeczne a Wzrost Gospodarczy", nr 39.

Kot S.M., Jakubowski J., Sokołowski A. (2007), *Statystyka*, Difin, Warszawa.

Piszczała J. (red.) (2000), *Matematyka i jej zastosowanie w naukach ekonomicznych*, Akademia Ekonomiczna w Poznaniu, Poznań.

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

Panek T. (2011), *Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa.

Pliszka T. (2004), *Skutki nierówności społecznych*, „Nierówności Społeczne a Wzrost Gospodarczy", nr 5.

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, nr 1.

Rószkiewicz M. (2012), *Metody ilościowe w badaniach marketingowych*, Wydawnictwo Naukowe PWN, Warszawa.

Witkowski M. (red.) (2010), *Statystyka matematyczna w zarządzaniu*, Uniwersytet Ekonomiczny w Poznaniu, Poznań.

Turczak A., Zwiech P. (2015a), *Zróżnicowanie dochodów ludności według województw*, „Wiadomości Statystyczne", nr 12.

Turczak A., Zwiech P. (2015b), *Podobieństwo województw w Polsce pod względem rozkładu wydatków ich mieszkańców*, „Ekonomia XXI Wieku", nr 3 (7).

GROUPING DISTRIBUTIONS INTO HOMOGENEOUS CLASSES EXEMPLIFIED BY CLASSIFICATION OF DIFFERENT BIOLOGICAL TYPES OF HOUSEHOLDS

Abstract

The purpose of the paper was to present a new way of dividing distributions into homogeneous groups and to use it to divide biological types of households into classes of most similar distributions. 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 thirteen biological types of households were divided into eight uniform classes. This division resulted in the creation of five single-element groups, two two-element groups and one four-element group.

Translated by Anna Turczak

Keywords: disposable income, household, Kolmogorov-Smirnov test, taxonomy

JEL Codes: D12, C10