

ZASTOSOWANIE METOD STATYSTYCZNEJ ANALIZY WIELOWYMIAROWEJ DO BADANIA STRUKTURY WYDATKÓW GOSPODARSTW DOMOWYCH

Elżbieta Badach

Katedra Statystyki Matematycznej
Uniwersytet Rolniczy im. H. Kołłątaja w Krakowie
rbadach@cyf-kr.edu.pl

Streszczenie: Celem badań było wskazanie analogii w zmianach w strukturze wydatków gospodarstw domowych wybranych krajów UE tj.: Hiszpanii, Portugalii i Grecji oraz Polski na przestrzeni lat 1988-2009. Posługując się zasadą analogii można w ten sposób przewidywać kierunek i tempo zmian w strukturze budżetów polskich gospodarstw domowych. Podziału na klasy podobieństwa dokonano w oparciu o zmodyfikowaną metodę środków ciężkości. Wyodrębniono 3 klasy. Struktura budżetów polskich gospodarstw domowych z lat 2004 – 2008 okazała się najbardziej podobna do struktury wydatków gospodarstw portugalskich z lat 1988-2005.

Słowa kluczowe: wydatki gospodarstw domowych, klasyfikacja, metoda środków ciężkości

WSTĘP

Na przestrzeni ostatnich lat konsumpcja polskich gospodarstw domowych podlegała stałym zmianom i wahaniom. Najistotniejszym czynnikiem determinującym poziom i strukturę tych wydatków są niewątpliwie dochody gospodarstw.

Wzrost dochodów powoduje wzrost spożycia dóbr żywnościowych i nieżywnościowych, ale udział wydatków na żywność w ich strukturze maleje ze względu na niższe tempo spożycia tych dóbr w porównaniu z tempem spożycia towarów nieżywnościowych.

W roku 2010 Polacy przeznaczyci średnio 21,4% swoich dochodów na żywność, podczas gdy przeciętny odsetek wydatków na ten cel liczony dla wszystkich państw Unii wynosi 15,6%.

Malejący udział wydatków na żywność w całości wydatków konsumpcyjnych wskazuje na prawidłowe tendencje w kształtowaniu struktury wydatków gospodarstw domowych i jest efektem oczekiwanym w gospodarce naszego kraju.

Z analizy rozwoju gospodarczego krajów świata wynika, że procesy ekonomiczne występujące w jednym państwie, pojawiają się również po upływie pewnego okresu w innym, co określa się mianem zasady analogii i [Jajuga, 1997] Niniejsze badania stanowią próbę wykorzystania tej zasady przy jednoczesnym zastosowaniu metod statystycznej analizy wielowymiarowej.

Celem było wskazanie podobieństwa w zmianach struktury wydatków gospodarstw domowych Polski i wybranych krajów dawnej Unii. Wytypowano trzy kraje, w których udział wydatków na cele żywnościowe w badanym okresie był najwyższy, tj: Hiszpanię, Portugalię i Grecję. Były to państwa, w których struktura wydatków konsumpcyjnych gospodarstw domowych mogła być "najbardziej podobna" do struktury wydatków polskich gospodarstw, i co do których, można przypuszczać, że przeszły drogą przemian, którą jest podobna do tej, jaka czeka polska gospodarce.

W podziale wydatków wyróżniono cztery składniki, tj: wydatki na żywność, na utrzymanie mieszkania i nośniki energii, na kulturę i relaks oraz pozostałe wydatki. Grupy podobieństwa zbudowano w oparciu o zmodyfikowaną metodę środków ciężkości.

METODYKA

Dane są m -wymiarowe obserwacje $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$, $i=1, 2, \dots, n$

przy czym składowe wektorów x_i sumują się do jedności, czyli $\sum_{j=1}^m x_{ij} = 1$ [Jajuga, 1997].

Klasyfikacja prowadzi do wyznaczenia minimum warunkowego funkcji

$$\sum_{k=1}^K \sum_{i \in \zeta_k} (x_i - v_k)^T (x_i - v_k) \quad (1)$$

przy warunkach opisanych równaniem:

$$\sum_{j=1}^m v_{kj} = 1, \quad k = 1, 2, \dots, K \quad (2)$$

gdzie K – liczba klas

ζ_k – k -ta klasa obserwacji

$v_i = [v_{i1}, v_{i2}, \dots, v_{im}]$ – m -wymiarowa obserwacja stanowiąca charakterystykę k -tej klasy.

Funkcja, której minimum jest poszukiwane, wyraża sumę kwadratów odległości euklidesowej wektorów obserwacji od wektorów położenia klas, do których te obserwacje należą. Warunki dane wzorem (2) dają gwarancję, że wektory wyznaczone jako charakterystyki klas będą mieć również postać danych strukturalnych.

Jeśli zdefiniujemy wektor $J = [1, 1, \dots, 1]$ z przestrzeni R^m , to warunek (2) można zapisać jako

$$v_k^T J = 1, \quad k = 1, 2, \dots, K \quad (3)$$

Do wyznaczenia minimum warunkowego wykorzystujemy metodę mnożników Lagrange'a, [Fichtenholz, 1985] stąd, po ich wprowadzeniu otrzymujemy funkcję:

$$\sum_{k=1}^K \sum_{i \in \zeta_k}^m (x_i - v_k)^T (x_i - v_k) - 2 \sum_{k=1}^K \lambda_k (v_k^T J - 1), \quad (4)$$

gdzie $\lambda_1, \lambda_2, \dots, \lambda_k$ – mnożniki Lagrange'a.

Po obliczeniu pochodnych funkcji (4) po v_k dla $k=1, 2, \dots, K$ i przyrównaniu ich do wektora zerowego z przestrzeni R^m otrzymujemy następujące zależności:

$$\sum_{i \in \zeta_k}^m (-2x_i + 2v_k) - 2\lambda_k J = 0, \quad \text{dla } k = 1, 2, \dots, K \quad (5)$$

A po ich przekształceniu mamy:

$$\lambda_k J = n_k v_k - n_k \bar{x}_k \quad \text{dla } k=1, 2, \dots, K \quad (6)$$

gdzie n_k – liczba obserwacji należących do k -tej klasy,

\bar{x}_k – wektor średnich k -tej klasy (obliczony na podstawie obserwacji należących do k -tej klasy)

Po pomnożeniu lewostronnie przez wektor J^T i przekształceniu otrzymujemy wzór:

$$\lambda_k = (1 - J^T \bar{x}_k) \frac{n_k}{m}, \quad \text{dla } k=1, 2, \dots, K \quad (7)$$

Po podstawieniu (7) do (6) i przekształceniach uzyskujemy postać wektora parametrów położenia

$$v_k = \bar{x}_k + \left[\frac{1}{m} (1 - J^T \bar{x}_k) \right] J \quad (8)$$

Z wzoru wynika, że jest on wektorem średnich arytmetycznych skorygowanym tak, aby jego składowe sumowały się do jedności.

Do zastosowania wzoru potrzebna jest znajomość klasyfikacji, a ta jest wyznaczana przez minimalizację funkcji wyrażającej kwadraty odległości

obserwacji od wektorów położenia klas. Stąd klasyfikacji dokonuje się poprzez zastosowanie algorytmu iteracyjnego, prowadzącego od pewnej (losowej) klasyfikacji, metodą kolejnych poprawek do uzyskania klasyfikacji, której już nie można poprawić.

Algorytm jest następujący:

1. Wyznacza się klasyfikację początkową zbioru obserwacji, np. losowo

$$\zeta^0 = \{\zeta_1^0, \zeta_2^0, \dots, \zeta_K^0\}$$
2. W r -tej iteracji algorytmu oblicza się:
 - wektory położenia klas wyodrębnionych w wyniku klasyfikacji dokonanej w $r-1$ iteracji
 - odległości euklidesowe obserwacji od wektorów położenia klas
 - d_{ik}^r – odległość wektora obserwacji x_k od wektora położenia k -tej klasy uzyskanej w r -tej iteracji
 - ustala się nową klasyfikację zbioru obiektów, przydzielając obiekt do tej klasy, dla której odległość od wektora jej położenia jest najmniejsza możliwych.
3. Procedurę iteracyjną kończy się, gdy klasyfikacje otrzymane w dwóch kolejnych krokach są identyczne.

BADANIA EMPIRYCZNE

Do analizy wykorzystano dane Głównego Urzędu Statystycznego oraz dane EUROSTAT pochodzące z programów badań nad budżetami gospodarstw domowych. Dane dla Polski dotyczą lat 2001-2009, zaś dane dla wybranych krajów unijnych pochodzą z lat 1988 -2005, z tym że są one wykonywane co 5 lub 6 lat. W strukturze wydatków gospodarstw domowych wyróżniono cztery składniki, tj: wydatki na żywność, na utrzymanie mieszkania i nośniki energii, na kulturę i relaks oraz pozostałe wydatki. Zestawienie danych zawiera tabela 1.

Struktura wydatków gospodarstw domowych jest kształtowana przede wszystkim przez wysokość tych dochodów. Według prawa Engla wzrost dochodów powoduje wzrost spożycia dóbr żywnościowych i nieżywnościowych, ale udział wydatków na żywność w ich strukturze maleje ze względu na niższe tempo spożycia tych dóbr w porównaniu z tempem spożycia towarów nieżywnościowych. Pożądanym jest więc taki kierunek zmian w strukturze wydatków gospodarstw domowych, który prowadzi do zmniejszenia udziału wydatków na żywność oraz potrzeby mieszkaniowe na rzecz wydatków na cele nie powiązane bezpośrednio z bytem członków gospodarstwa. Opisaną tendencję można zaobserwować w strukturze budżetów gospodarstw domowych wszystkich wybranych do analizy państw, w tym także Polski, jednak tempo tych zmian jest zróżnicowane. Obliczone średnie tempo zmian wydatków na żywność wskazuje, iż najszybciej zmiany te zachodziły w Hiszpanii (średnie tempo zmian: $i=0,230$) najwolniej Polsce ($i=0,10$).

Wykorzystując zmodyfikowaną metodę środków ciężkości dokonano klasyfikacji badanych obiektów, w wyniku której wyodrębniono (według założeń powziętych z góry) 3 grupy podobieństwa – tabela 2.

Tabela 1 Struktura wydatków gospodarstw domowych w Polsce i wybranych krajach UE w badanym okresie

Kraj	Lata	Żywność	Użytkowanie mieszkania i nośniki energii	Kultura i rekreacja	Pozostałe wydatki
Hiszpania	1988	0,2714	0,1817	0,049	0,4979
	1994	0,2436	0,2363	0,067	0,4531
	1999	0,1948	0,2479	0,062	0,4953
	2005	0,191	0,263	0,064	0,482
Portugalia	1988	0,3125	0,1131	0,04	0,5344
	1994	0,2363	0,1852	0,037	0,5415
	1999	0,2119	0,1853	0,048	0,5548
	2005	0,201	0,1823	0,057	0,5597
Grecja	1988	0,237	0,1804	0,046	0,5366
	1994	0,1905	0,2287	0,039	0,5418
	1999	0,1801	0,2103	0,045	0,5646
	2005	0,163	0,221	0,042	0,574
Polska	2001	0,308	0,179	0,067	0,446
	2002	0,31	0,188	0,065	0,437
	2003	0,295	0,199	0,064	0,441
	2004	0,278	0,21	0,066	0,447
	2005	0,307	0,222	0,074	0,397
	2006	0,281	0,196	0,068	0,454
	2007	0,271	0,196	0,071	0,461
	2008	0,266	0,184	0,076	0,473
	2009	0,256	0,189	0,079	0,476

Źródło: obliczenia własne na podstawie danych GUS i Eurostat

Klasyfikacja wskazuje na podobieństwo struktur budżetów gospodarstw domowych Polski lat 2004-2008 do Portugalii 1988-2005 oraz Hiszpanii z 1988 r.

Przyjmując zatem, że tempo i scenariusz zmian w strukturze budżetów gospodarstw domowych Polski i Portugalii będzie przebiegać w zbliżony sposób, można założyć, iż okres „opóźnienia” Polski względem Portugalii wynosi około 16 lat (biorąc pod uwagę różnicę między najwcześniejszym obiektem Portugalii i najwcześniejszym obiektem Polski zaliczonym do drugiej grupy).

Tabela 2. Wyniki klasyfikacji

grupa 1	grupa 2	grupa 3
Hiszpania 1994	Portugalia 1988	Polska 2000
Hiszpania 1999	Portugalia 1994	Polska 2001
Hiszpania 2005	Portugalia 1999	Polska 2002
Grecja 1988	Portugalia 2005	Polska 2003
Grecja 1994	Hiszpania 1988	
Grecja 1999	Polska 2004	
Grecja 1995	Polska 2005	
	Polska 2006	
	Polska 2007	
	Polska 2008	
	Polska 2009	

Źródło: opracowanie własne

Osobną grupę tworzą obiekty obrazujące strukturę wydatków gospodarstw domowych Polski w pierwszej fazie analizowanego okresu tj w latach 2000-2003, jest to klasa, w której udział wydatków na cele żywnościowe kształtuje się na najmniej korzystnym poziomie tabela 3.

Tabela 3. Wektory położenia klas dla grup uzyskanych w klasyfikacji

Wektory położenia klas	grupa 1	grupa 2	grupa 3
Żywność	0,200	0,262	0,298
Mieszkanie i energia	0,227	0,183	0,194
Kultura i relaks	0,052	0,06	0,066
Pozostałe	0,521	0,495	0,442

Źródło: obliczenia własne

Grupa ta charakteryzuje się także najmniejszym udziałem wydatków na pozostałe cele, czyli niepowiązane bezpośrednio ze sprawami bytowymi członków gospodarstwa domowego.

BIBLIOGRAFIA

- Fichtenholz G. M. (1985) Rachunek różniczkowy i całkowity, PWN, Warszawa
 Jajuga K. (1997) Metody statystycznej analizy wielowymiarowej w projekcji ścieżek rozwojowych, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, 743, str 75 – 80
www.stat.gov.pl
<http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/>

**APPLYING METHODS OF MULTIDIMENSIONAL
STATISTICAL ANALYSIS FOR STUDYING
THE STRUCTURE OF HOUSEHOLD EXPENSES**

Abstract: The aim of the study was to show the analogy in the changes in the structure of household expenses in the chosen EU countries and Poland between 1988 and 2009. The direction and pace of the changes in the budgets of Polish households can be predicted by using the principle of analogy. Classification of the objects has been carried out based on the modified k-means method. Three classes of similarities have been distinguished. The structure of Polish households budgets from 2004 – 2008 has turned out the most similar to the structure of the expenses of Portuguese households from 1988 – 2005.

Key words: households budget, classification, method of average means