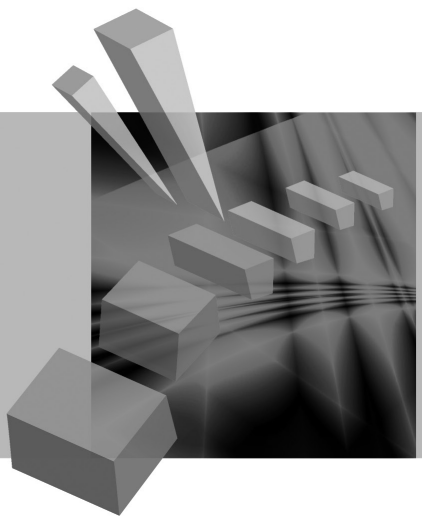
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

279

Taksonomia 21

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka
Redaktor techniczny: Barbara Łopusiewicz
Korektor: Barbara Cibis
Łamanie: Małgorzata Czupryńska
Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon
http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana
Druk: Drukarnia TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 279 • 2013

Taksonomia 21. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

Artur Czech
Politechnika Białostocka

ZASTOSOWANIE WYBRANYCH METOD DOBORU ZMIENNYCH DIAGNOSTYCZNYCH W BADANIACH KONSUMPCJI W UJĘCIU POŚREDNIM

Streszczenie: W artykule podjęto próbę poszukiwania najefektywniejszego narzędzia doboru i analizy zestawu zmiennych diagnostycznych wykorzystywanych do badań konsumpcji w ujęciu pośrednim z użyciem ocen syntetycznych. Analizie poddano różne podejścia w procesie doboru zmiennych diagnostycznych do modelu taksonomicznego w ujęciu województw. Szczególną uwagę zwrócono na postulat oddzielnego traktowania Warszawy na terenie województwa mazowieckiego. Taki sposób prowadzenia analizy może powodować występowanie asymetrii rozkładu empirycznego w rozkładach cech diagnostycznych, co wymaga zastosowania metod mających na celu eliminację niedoskonałości analiz na etapie statystycznego doboru finalnego zestawu cech diagnostycznych.

Słowa kluczowe: konsumpcja w ujęciu pośrednim, dobór cech diagnostycznych, miernik syntetyczny.

1. Wstęp

Podstawą stosowania prawidłowej polityki regionalnej w warunkach występowania ograniczeń środków finansowych pochodzących zarówno z funduszy unijnych, jak i z budżetu państwa, powinna być rzetelna diagnoza zaspokojenia potrzeb społeczeństwa. Potrzeby i stopień ich zaspokojenia można mierzyć nie tylko wprost przez analizę struktury spożycia dóbr materialnych i usług wyłącznie na podstawie danych z Badań Budżetów Gospodarstw Domowych przeprowadzanych przez GUS (konsumpcja w ujęciu bezpośrednim), ale także w sposób pośredni. Konsumpcja w ujęciu pośrednim traktowana jest jako wskaźnik zmian społecznych (np. poziomu, jakości i godności życia); źródłem ocen liczbowych są wyniki empirycznych badań GUS oraz badań według koncepcji różnych autorów; pomiar konsumpcji w tym ujęciu ma ważne znaczenie dla polityki społecznej, która jest postrzegana jako „hamulec wzrostu gospodarczego pasożytujący na bogactwie narodowym" [Słaby 2006b, s. 81].

Oszacowanie szans i zagrożeń w rozwoju społeczno-gospodarczym poszczególnych regionów Polski w ujęciu województw możliwe jest wtedy, gdy wskaza-

ne zostaną mierniki pozwalające w prawidłowy i rzetelny sposób określić rozwój poszczególnych jednostek terytorialnych, na co pozwala m.in. zastosowanie metod taksonomicznych. W każdej analizie taksonomicznej na wstępie konieczne jest określenie obiektów będących przedmiotem badania oraz zestawu cech diagnostycznych, które będą opisywały jednostki podlegające ocenie pod względem rozwoju społeczno-gospodarczego mierzonego stanem konsumpcji w ujęciu pośrednim, np. poziomu życia. Jest to niezwykle istotne, ponieważ to od prawidłowego doboru tzw. finalnego zestawu cech diagnostycznych w znacznym stopniu zależą wyniki dokonywanych ocen, bez względu na stosowane w dalszych etapach analizy taksonomicznej algorytmy postępowania [Panek 2009, s. 16].

Celem pracy jest poddanie weryfikacji wybranych, statystycznych metod doboru finalnego zestawu cech diagnostycznych do badań konsumpcji w ujęciu pośrednim (poziomu życia) z zastosowaniem metod taksonomicznych. W warunkach zaś asymetrii rozkładu empirycznego zastosowanie klasycznych procedur doboru cech nie jest wystarczające.

2. Podstawy teoretyczne zastosowanych metod badawczych

Każde badanie taksonomiczne powinno zostać poprzedzone doborem zestawu cech diagnostycznych. W większości przypadków dobór ten składa się z następujących etapów: doboru merytorycznego, który umożliwia zgromadzenie tzw. potencjalnego zestawu cech diagnostycznych, oraz ich weryfikacji statystycznej. Dobór merytoryczny zawsze niesie za sobą dozę arbitralizmu i może zniekształcać wyniki badań. W wielu przypadkach zestawy potencjalnych zmiennych diagnostycznych mogą przyjmować znaczne rozmiary.

Wówczas pomocne staje się wykorzystanie metod statystycznych, gdzie prawidłowy w sensie statystycznym dobór finalnego zestawu cech powinien obejmować analizę zmienności oraz korelacji. Należałoby podkreślić, iż statystyczny dobór cech diagnostycznych stanowi solidną podstawę prowadzonych badań nie tylko w obszarze konsumpcji, ale może przykładowo dotyczyć również konstrukcji miar syntetycznych do oceny spółek na rynku energii elektrycznej [Halicka 2012, s. 84-88], zastosowań marketingowych [Gatnar, Walesiak 2004, s. 361], czy też w przypadku analiz prowadzonych pod kątem oceny szkół wyższych z zastosowaniem analizy skupień i metody DEA [Nazarko, Chodakowska, Jarocka 2012, s. 163-172].

Analiza zmienności ma na celu wyeliminowanie cech nisko zróżnicowanych, które z merytorycznego punktu widzenia charakteryzują się małą zdolnością dyskryminacyjną, dlatego nie są istotne z analitycznego punktu widzenia. W literaturze można znaleźć propozycje pomiaru zmienności cech diagnostycznych, które odbywają się z wykorzystaniem różnego rodzaju współczynników [Młodak 2005, s. 5-18]. Najczęściej znajduje zastosowanie klasyczny współczynnik zmienności, który w swojej konstrukcji wykorzystuje klasyczne miary rozkładu badanych cech

diagnostycznych, takie jak: średnia arytmetyczna i odchylenie standardowe. Miary te są wrażliwe na asymetrię rozkładu empirycznego.

W przypadku badań konsumpcji w ujęciu województw postuluje się wyodrębnienie Warszawy jako oddzielnego obiektu analizy [Słaby, Czech 2011, s. 7-22]. Wynika to z faktu, iż obszar stolicy jest bardzo wysoko rozwinięty i powoduje to zawyżanie ocen województwa mazowieckiego w stosunku do innych regionów Polski. Znajduje to swoje odzwierciedlenie m.in. w występowaniu dla Warszawy nietypowych (odstających lub ekstremalnych) wartości w rozkładach poszczególnych cech diagnostycznych. Bardzo pomocne wówczas staje się wykorzystanie współczynników zmienności charakteryzujących się odpornością na wpływ asymetrii rozkładu empirycznego. Jedna z takich propozycji wykorzystuje w swojej konstrukcji pojęcie wielowymiarowego wektora medianowego i wyraża się następującym wzorem [Młodak 2006, s. 29]:

$$V_P = \frac{mad\left(X_j\right)}{\theta_j}, \tag{1}$$

gdzie: $mad\left(X_j\right)$ – medianowe odchylenie bezwzględne w rozkładzie *j*-tej cechy [Słaby, Czech 2011, s. 10]:

$$mad\left(X_j\right) = \underset{i=1,2,\ldots,n}{med}\left|x_{ij} - \theta_j\right|, \tag{2}$$

zaś wartości θ_j to składowe poszczególnych wielowymiarowych wektorów medianowych (brzegowego i A. Webera). Wartości poszczególnych elementów wielowymiarowego wektora A. Webera uzyskuje się z następującego wzoru [Słaby, Czech 2011, s. 10]:

$$T\left(\Theta, R^m\right) = \arg\min_{\Theta \in R^m}\left\{\sum_{i=1}^{n}\left[\sum_{j=1}^{m}\left(x_{ij} - \theta_j\right)^2\right]^{1/2}\right\}. \tag{3}$$

Prezentację koncepcji mediany A. Webera można znaleźć w pracy [Młodak 2009, s. 3-21].

Natomiast drugi etap statystycznej weryfikacji zbioru potencjalnych zmiennych diagnostycznych to eliminacja cech wysoko skorelowanych. Pośród szerokiego zestawu metod na uwagę zasługuje metoda odwróconej macierzy współczynników korelacji liniowej Pearsona [Panek 2009, s. 23] oraz parametryczna metoda Z. Hellwiga [Nowak 1990, s. 26-33].

3. Implementacja wybranych metod do badań konsumpcji w ujęciu pośrednim

Jako podstawę analizy zastosowania wybranych metod doboru statystycznego przyjęto zbiór dziesięciu następujących zmiennych diagnostycznych: X_1 – przeciętny

miesięczny dochód rozporządzalny na osobę w gospodarstwie domowym (w zł), X_2 – udział wydatków na żywność i napoje bezalkoholowe w ogólnej ich sumie – w ogólnej sumie wydatków gospodarstwa domowego – (w %), X_3 – udział wydatków na wyposażenie mieszkania i prowadzenie gospodarstwa domowego w ogólnej ich sumie (w %), X_4 – udział wydatków na zdrowie w ogólnej ich sumie (w %), X_5 – udział wydatków na transport w ogólnej ich sumie (w %), X_6 – udział wydatków związanych ze spędzaniem czasu wolnego w ogólnej ich sumie (w %), X_7 – przeciętny wiek głowy gospodarstwa domowego (w latach), X_8 – przeciętny poziom wykształcenia głowy gospodarstwa domowego, X_9 – przeciętny próg miesięcznego dochodu netto uznawany za minimalnie wystarczający (w zł), X_{10} – przeciętna ocena sytuacji materialnej. Wartości poszczególnych cech diagnostycznych uzyskano z rozkładów empirycznych zaczerpniętych z Badań Budżetów Gospodarstw Domowych GUS w 2007 r. Wyselekcjonowany zbiór potencjalnych cech diagnostycznych został poddany analizie zmienności z zastosowaniem dwóch typów współczynników zmienności – klasycznego i jego wersji pozycyjnej. Wyniki przeprowadzonej analizy zawarto w tab. 1.

Tabela 1. Mierniki asymetrii i zróżnicowania zbioru cech diagnostycznych

Miary		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
A_S	**A**	–0,19	0,89	–0,02	–0,46	–0,46	0,07	–0,37	–0,68	0,33	0,65
	B	2,99	–1,20	–0,01	–0,17	–0,51	2,39	–0,32	–2,91	1,59	0,03
V_k	**A**	10,04	6,35	5,44	16,97	25,38	7,10	2,47	5,38	7,67	2,99
	B	19,48	8,85	5,27	17,92	24,53	11,85	2,50	9,66	9,98	3,40
$V_{p.m.b}$	**A**	5,21	3,79	4,82	13,75	16,74	5,26	1,81	3,50	6,56	1,84
	B	4,95	4,03	4,92	12,34	16,75	4,54	1,90	2,57	5,42	2,57
$V_{p.m.W}$	**A**	5,48	3,80	4,19	16,36	16,10	6,30	1,81	3,19	6,53	2,24
	B	4,62	4,01	3,95	15,54	16,55	5,84	2,01	2,84	6,20	2,42

Objaśnienia skrótów: A – 16 województw, B – 16 województw oraz Warszawa (mazowieckie bez Warszawy), A_S – współczynnik asymetrii, V_k – klasyczny współczynnik zmienności, $V_{p.m.b.}$ – pozycyjny współczynnik zmienności z medianą brzegową, $V_{p.m.W}$ – pozycyjny współczynnik zmienności z medianą Webera.

Źródło: opracowanie własne na podstawie [Czech 2010].

Analizując dane z tab. 1, stwierdzono, iż w przypadku oddzielnego uwzględnienia Warszawy w analizie (16 województw oraz Warszawa) wartości klasycznego współczynnika zmienności w przypadku poszczególnych cech mogą wzrastać znacząco. Powodowane jest to silną asymetrią rozkładu empirycznego wynikającą z występowania nietypowych wartości dla stolicy w rozkładach cech.

Stwierdzono, iż zastosowanie pozycyjnego współczynnika zmienności opartego na pojęciu medianowego odchylenia bezwzględnego czyni analizę niewrażliwą na wpływ wartości nietypowych. Dodatkowo zastosowanie mediany A. Webera pozwala uwzględnić interakcje w zbiorze zmiennych diagnostycznych, co powo-

duje w niektórych przypadkach podniesienie wartości współczynnika zmienności w porównaniu do wersji z medianą brzegową, w innych zaś obniża jego wartość. Uwzględniając zawarte w literaturze stwierdzenia na temat wartości progowej pozycyjnego współczynnika zmienności do badań konsumpcji w ujęciu pośrednim [Słaby 2006a, s. 124], w pracy przyjęto jego wartość na poziome około 5%. Stwierdzono, iż zastosowanie mediany Webera do analizy zmienności dla 16 województw i Warszawy (mazowieckie bez Warszawy) przy sztywnym progu dyskryminacyjnym 5% pozwala na uwzględnienie w dalszej analizie cechy X_6.

Zauważono również, że przyjęcie wartości progowej 10% (najczęściej stosowane) powodowałoby pozostawienie do analizy tylko dwóch cech: X_5 i X_6. W związku z tym cechy, dla których wartości współczynników zmienności w ujęciu pozycyjnym były poniżej przyjętego progu dyskryminacyjnego 5%, uznawano za quasi-stałe i eliminowano. W wyniku przeprowadzonej analizy do dalszych badań przyjęto następujące cechy diagnostyczne: X_1, X_4, X_5, X_6 i X_9, które poddano analizie korelacyjnej z zastosowaniem metody odwróconej macierzy współczynników korelacji Pearsona. Wyniki analizy zostały zmieszczone w tab. 2.

Tabela 2. Odwrócone macierze korelacji Pearsona

	16 województw					16 województw oraz Warszawa				
	X_1	X_4	X_5	X_6	X_9	X_1	X_4	X_5	X_6	X_9
X_1	**4,01**	0,58	0,15	–2,61	–1,04	**10,35**	–0,40	0,96	–7,59	–2,50
X_4	0,58	**1,54**	–0,89	–0,46	0,18	–0,40	**1,53**	–0,90	–0,48	0,53
X_5	0,15	–0,89	**1,96**	0,59	–1,25	0,96	–0,90	**1,97**	0,83	–1,86
X_6	–2,61	–0,46	0,59	**3,93**	–1,15	–7,59	–0,48	0,83	**10,36**	–2,47
X_9	–1,04	0,18	–1,25	–1,15	**3,02**	–2,50	0,53	–1,86	–2,47	**5,64**

Źródło: opracowanie własne na podstawie [Czech 2010].

Należałoby zauważyć, iż przypadek oddzielnego ujęcia Warszawy w analizie powoduje znaczący wzrost wartości na głównej przekątnej odwróconej macierzy korelacji. Przekroczenie tej wartości progowej na poziomie 10 może świadczyć o złym uwarunkowaniu numerycznym macierzy i zbyt dużym skorelowaniu odpowiedniej cechy z innymi. Przyczyn zaistniałej sytuacji należy upatrywać jednak w występowaniu nietypowych wartości dochodu rozporządzalnego i odsetka wydatków związanych ze spędzaniem czasu wolnego dla gospodarstw domowych zamieszkujących obszar stolicy, co znajduje swoje odzwierciedlenie w wartościach współczynników asymetrii zamieszczonych w tab. 1. W tym przypadku posłużono się więc zaprezentowaną przez A. Młodaka uwagą o możliwości zastosowania w miejsce współczynników korelacji Pearsona współczynników korelacji Spearmana lub τ-Kendalla [Młodak 2006, s. 31-33]. Wyniki przeprowadzonej powtórnie analizy korelacyjnej dla przypadku 16 województw i Warszawy zestawiono w tab. 3.

Tabela 3. Odwrócone macierze korelacji Spearmana i τ-Kendalla dla 16 województw (mazowieckie bez Warszawy) oraz Warszawa jako siedemnasta jednostka badania

	Spearmana					τ-Kendalla				
	X_1	X_4	X_5	X_6	X_9	X_1	X_4	X_5	X_6	X_9
X_1	**3,21**	0,26	0,04	–1,72	–1,13	**1,77**	0,16	–0,07	–0,88	–0,41
X_4	0,26	**1,19**	–0,47	–0,14	0,12	0,16	**1,10**	–0,29	0,01	0,03
X_5	0,04	–0,47	**1,49**	0,63	–0,94	–0,07	–0,29	**1,15**	0,22	–0,32
X_6	–1,72	–0,14	0,63	**3,09**	–1,06	–0,88	0,01	0,22	**1,79**	–0,49
X_9	–1,13	0,12	–0,94	–1,06	**2,82**	–0,41	0,03	–0,32	–0,49	**1,50**

Źródło: opracowanie własne na podstawie [Czech 2010].

Analiza wartości znajdujących się na diagonalach poszczególnych macierzy pozwala zaobserwować znaczący ich spadek w porównaniu do metody odwróconej macierzy współczynników korelacji Pearsona. Wynika to z faktu, iż współczynnik korelacji liniowej Pearsona w swojej konstrukcji opiera się na średniej arytmetycznej i odchyleniu standardowym, a więc charakterystykach, które nie są odporne na wpływ wartości nietypowych występujących w rozkładach empirycznych cech diagnostycznych.

Analizie poddano również parametryczną metodę Z. Hellwiga w wersji klasycznej oraz wykorzystano wersję zmodyfikowaną zaproponowaną przez A. Młodaka, która polega na zastąpieniu sumy modułów kolumny (wiersza) macierzy współczynników korelacji poprzez ich medianę. Wyniki z zastosowaniem dwóch sposobów konstrukcji wielowymiarowego wektora medianowego zaprezentowana w tab. 4.

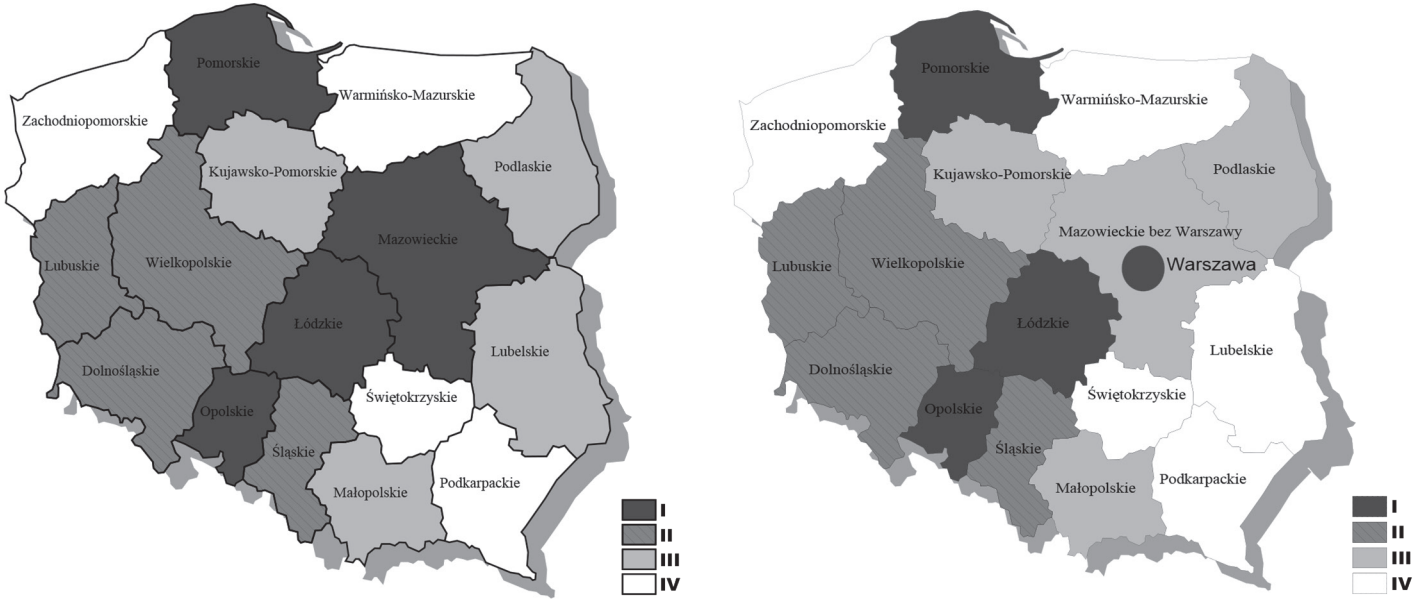
Tabela 4. Wyniki analizy korelacyjnej z zastosowaniem parametrycznej metody Z. Hellwiga w postaci klasycznej i zmodyfikowanej

Rodzaj cechy	Typ analizy	Wariant klasyczny	Wariant z medianą brzegową	Wariant z medianą Webera
Centralne	**A**	X_1, X_9	X_6	X_1
	B	X_9	X_6	X_9
Satelitarne	**A**	X_6	X_1	X_6
	B	X_1, X_6	X_1, X_9	X_1, X_6
Izolowane	**A**	X_4, X_5	X_4, X_5, X_9	X_4, X_5, X_9
	B	X_4, X_5	X_4, X_5	X_4, X_5

Objaśnienia skrótów: A – 16 województw, B – 16 województw oraz Warszawa.

Źródło: opracowanie własne z wykorzystaniem programu Statistica PL i Microsoft Excel.

Zastosowanie wersji zmodyfikowanej z użyciem mediany brzegowej miało na celu uczynić analizę niewrażliwą na asymetrię rozkładu empirycznego cech (przy-



A – 16 województw

B – 16 województw oraz Warszawa

Rys. 1. Przestrzenne zróżnicowanie konsumpcji w ujęciu pośrednim z zastosowaniem mediany A. Webera

Źródło: opracowanie własne z wykorzystaniem programu Statistica PL.

padek analizy dla 16 województw i Warszawy, gdzie województwo mazowieckie rozpatrywane jest bez stolicy). Zastosowanie zaś mediany Webera dodatkowo pozwoliło na uwzględnienie interakcji w zbiorze rozpatrywanych zmiennych diagnostycznych.

Przykład podziału taksonomicznego z zastosowaniem pozycyjnej metody porządkowania liniowego z wykorzystaniem wyników doboru finalnego zestawu cech diagnostycznych przy użyciu metody odwróconej macierzy współczynników korelacji zaprezentowano na rys. 1. W ujęciu pozycyjnym spośród metod normalizacyjnych stosowano standaryzację z zastosowaniem mediany Webera [Słaby 2006].

4. Podsumowanie

Na podstawie przeprowadzonej analizy zastosowania wybranych metod doboru finalnego zestawu cech diagnostycznych do badań konsumpcji w ujęciu pośrednim można stwierdzić, że weryfikacja statystyczna potencjalnego zestawu zmiennych diagnostycznych do badań w ujęciu województw powinna zostać poprzedzona analizą asymetrii rozkładu empirycznego.

Ponadto w procesie statystycznej weryfikacji zgromadzonego zestawu cech diagnostycznych w warunkach występowania silnej asymetrii rozkładu empirycznego należy kierować się pozycyjnymi współczynnikami zmienności będącymi ilorazem medianowego odchylenia bezwzględnego cechy i jej mediany. W przypadku zastosowania pozycyjnych współczynników zmienności w procesie konstrukcji finalnego zestawu cech diagnostycznych jako podstawy konstrukcji miary syntetycznej do badań konsumpcji w ujęciu pośrednim wartość graniczna, poniżej której cechy uznaje się za mało zróżnicowane, powinna się kształtować na poziomie 5%.

Wykorzystanie wyłącznie pozycyjnego współczynnika zmienności opartego na medianie brzegowej może prowadzić do nadmiernej eliminacji zmiennych, a tym samym zmniejszać wartość informacyjną i poznawczą budowanych modeli. Znacznie lepszym rozwiązaniem wydaje się wykorzystanie pojęcia wielowymiarowej mediany A. Webera, której zastosowanie pozwala nie tylko uodpornić analizę na wpływ wartości nietypowych, ale również uwzględnia interakcje w zbiorze zmiennych diagnostycznych na etapie analizy zmienności, co poprzez uwzględnienie w analizie innych komponentów modelu może wpływać na decyzję o zakwalifikowaniu danej cechy do dalszych badań empirycznych.

W warunkach silnej asymetrii rozkładu empirycznego – przypadek analizy dla 16 województw i Warszawy (województwo mazowieckie rozpatrywane bez Warszawy) – wykorzystanie współczynników korelacji liniowej Pearsona nie jest „uprawnione". Wynika to z faktu, iż wartości na głównej przekątnej odwróconej macierzy współczynników korelacji zostają sztucznie zawyżone ze względu na nietypowe wartości cech dla stolicy.

Wykorzystanie metody odwróconej macierzy współczynników korelacji dostarcza bardziej jednoznacznych wyników w zakresie doboru finalnego zestawu cech

diagnostycznych do badań konsumpcji w ujęciu pośrednim niż implementacja parametrycznej metody Z. Hellwiga zarówno w ujęciu klasycznym, jak i zastosowaniem jej wersji zmodyfikowanej wykorzystującej pojęcie wielowymiarowego wektora medianowego.

Literatura

Czech A., *Modelowanie konsumpcji w ujęciu pośrednim. Aspekty metodologiczne* (rozprawa doktorska), 2010.

Gatnar E., Walesiak M., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wyd. Akademii Ekonomicznej, Wrocław 2004.

Halicka K., *Budowa i analiza rankingów spółek dystrybucyjnych z wykorzystaniem metod porządkowania liniowego*, „Rynek Energii" 2012, nr 2.

Młodak A., *Ocena zmienności cech statystycznych w modelu taksonomicznym*, „Wiadomości Statystyczne" 2005, nr 9.

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

Młodak A., *Historia problemu Webera*, „Matematyka Stosowana" 2009, nr 10(51).

Nazarko J., Chodakowska E., Jarocka M., *Segmentacja szkół wyższych metodą analizy skupień versus konkurencja technologiczna ustalona metodą DEA – studium komparatywne*, [w:] Taksonomia 19, *Klasyfikacja i analiza danych – teoria i zastosowania*, Wrocław 2012.

Nowak E., *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa 1990.

Panek T., *Statystyczne metody wielowymiarowej analizy porównawczej*, SGH, Warszawa 2009.

Słaby T., *Konsumpcja. Eseje statystyczne*, Difin, Warszawa 2006a.

Słaby T., *Statystyczny pomiar konsumpcji*, [w:] M. Janoś-Kresło, B. Mróz, *Konsument i konsumpcja we współczesnej gospodarce*, SGH, Warszawa 2006b.

Słaby T., Czech A., *Zróżnicowanie regionalne konsumpcji w ujęciu pośrednim – ujęcie statyczne i przestrzenno-czasowe*, „Studia i Prace Kolegium Zarządzania i Finansów", SGH, Warszawa 2011, nr 111.

APPLICATION OF CHOSEN METHODS FOR THE SELECTION OF DIAGNOSTIC VARIABLES IN INDIRECT CONSUMPTION RESEARCH

Summary: The main aim of the paper is to find the most effective method for the selection and analysis of a set of diagnostic variables for indirect consumption research with the use of synthetic consumption measures. The analysis of potential sets of variables in the process of the assessment of living standard in different voivodeships was carried out using various methods. Special emphasis was put on the fact that within the area of the Mazowieckie Voivodeship, Warsaw ought to be analyzed separately. This way of caring out analysis can cause asymmetry in empirical distribution of diagnostic variables, which requires the implementation of adequate methods that would eliminate imperfections of analyses at the stage of the final selection of a set of diagnostic variables.

Keywords: indirect consumption, selection of diagnostic variables, synthetic measure.