

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

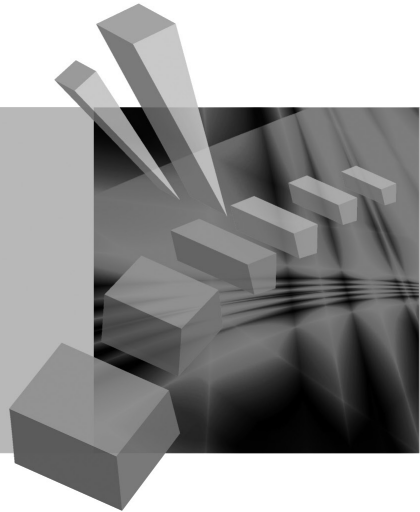
RESEARCH PAPERS

of Wrocław University of Economics

278

Taksonomia 20

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

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Józef Pocięcha: Wskaźniki finansowe a klasyfikacyjne modele predykcji upadłości firm	15
Eugeniusz Gatnar: Analiza miar adekwatności rezerw walutowych	23
Marek Walesiak: Zagadnienie doboru liczby klas w klasyfikacji spektralnej	33
Joanicjusz Nazarko, Joanna Ejdyś, Anna Kononiuk, Anna M. Olszewska: Analiza strukturalna jako metoda klasyfikacji danych w badaniach foresight	44
Andrzej Bąk: Metody porządkowania liniowego w polskiej taksonomii – pakiet <code>pllord</code>	54
Aleksandra Łuczak, Feliks Wysocki: Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia	63
Ewa Roszkowska: Zastosowanie rozmytej metody TOPSIS do oceny ofert negocjacyjnych	74
Jacek Batóg: Analiza wrażliwości metody ELECTRE III na obserwacje nietypowe i zmianę wartości progowych	85
Jerzy Korzeniewski: Modyfikacja metody HINoV selekcji zmiennych w analizie skupień	93
Małgorzata Markowska, Danuta Strahl: Wykorzystanie referencyjnego systemu granicznego do klasyfikacji europejskiej przestrzeni regionalnej ze względu na filar inteligentnego rozwoju – kreatywne regiony	101
Elżbieta Sobczak: Inteligentne struktury pracujących a efekty strukturalne zmian zatrudnienia w państwach Unii Europejskiej.....	111
Elżbieta Gołata, Grażyna Dehnel: Rozbieżności szacunków NSP 2011 i BAEL.....	120
Iwona Foryś: Wykorzystanie analizy historii zdarzeń do badania powtórnych sprzedaży na lokalnym rynku mieszkaniowym	131
Hanna Dudek, Joanna Landmesser: Wpływ relatywnej deprivacji na subiektywne postrzeganie dochodów.....	142
Grażyna Łaska: Syntaksonomia numeryczna w klasyfikacji, identyfikacji i analizie przemian zbiorowisk roślinnych	151
Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analiza zależności między procesami fundamentalnymi a rynkiem kapitałowym w Chinach	161

Andrzej Bąk, Tomasz Bartłomowicz: Mikroekonometryczne modele wielomianowe i ich zastosowanie w analizie preferencji z wykorzystaniem programu R	169
Andrzej Dudek, Bartosz Kwaśniewski: Przetwarzanie równoległe algorytmów analizy skupień w technologii CUDA	180
Michał Trzęsiok: Wycena rynkowej wartości nieruchomości z wykorzystaniem wybranych metod wielowymiarowej analizy statystycznej	188
Joanna Trzęsiok: Wybrane symulacyjne techniki porównywania nieparametrycznych metod regresji.....	197
Artur Mikulec: Kryterium Mojeny i Wisharta w analizie skupień – przypadek skupień o różnych macierzach kowariancji	206
Artur Zaborski: Analiza <i>unfolding</i> z wykorzystaniem modelu grawitacji	216
Justyna Wilk: Identyfikacja obszarów problemowych i wzrostowych w województwie dolnośląskim w zakresie kapitału ludzkiego	225
Karolina Bartos: Analiza ryzyka odejścia studenta z uczelni po uzyskaniu dyplomu licencjata – zastosowanie sieci MLP	236
Ewa Genge: Segmentacja uczestników Industriady z wykorzystaniem analizy klas ukrytych	246
Izabela Kurzawa: Wielomianowy model logitowy jako narzędzie identyfikacji czynników wpływających na sytuację mieszkaniową polskich gospodarstw domowych	254
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modele eksploracji danych niezbilansowanych – procedury klasyfikacji dla zadania analizy ryzyka operacyjnego.....	262
Aleksandra Łuczak: Zastosowanie rozmytej hierarchicznej analizy w tworzeniu strategii rozwoju jednostek administracyjnych	271
Marcin Pelka: Rozmyta klasyfikacja spektralna c -średnich dla danych symbolicznych interwałowych.....	282
Małgorzata Machowska-Szewczyk: Klasyfikacja obiektów reprezentowanych przez różnego rodzaju cechy symboliczne	290
Ewa Chodakowska: Indeks Malmquista w klasyfikacji podmiotów gospodarczych według zmian ich względnej produktywności działania	300
Beata Bieszk-Stolorz, Iwona Markowicz: Wykorzystanie modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa do badania szansy podjęcia pracy w zależności od rodzaju bezrobocia	311
Marcin Salamaga: Weryfikacja teorii poziomu rozwoju gospodarczego J.H. Dunninga w ujęciu sektorowym w wybranych krajach Unii Europejskiej	321
Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Sytuacja społeczno-gospodarcza jako determinanta migracji wewnętrznych w Polsce.	330
Hanna Gruchociak: Delimitacja lokalnych rynków pracy w Polsce na podstawie danych z badania przepływów ludności związanych z zatrudnieniem	343

Radosław Pietrzyk: Efektywność inwestycji polskich funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych	351
Sabina Denkowska: Procedury testowań wielokrotnych	362

Summaries

Józef Pocięcha: Financial ratios and classification models of bankruptcy prediction	22
Eugeniusz Gatnar: Analysis of FX reserve adequacy measures	32
Marek Walesiak: Automatic determination of the number of clusters using spectral clustering	43
Joanicjusz Nazarko, Joanna Ejdys, Anna Kononiuk, Anna M. Olszewska: Structural analysis as a method of data classification in foresight research	53
Andrzej Bąk: Linear ordering methods in Polish taxonomy – pllord package	62
Aleksandra Łuczak, Feliks Wysocki: The application of spatial median of Weber and the method TOPSIS in positional formulation for the construction of synthetic measure of standard of living	73
Ewa Roszkowska: Application of the fuzzy TOPSIS method to the estimation of negotiation offers.....	84
Jacek Batóg: Sensitivity analysis of ELECTRE III method for outliers and change of thresholds	92
Jerzy Korzeniewski: Modification of the HINoV method of selecting variables in cluster analysis	100
Małgorzata Markowska, Danuta Strahl: Implementation of reference limit system for the European regional space classification regarding smart growth pillar – creative regions	110
Elżbieta Sobczak: Smart workforce structures versus structural effects of employment changes in the European Union countries	119
Elżbieta Gołata, Grażyna Dehnel: Divergence in National Census 2011 and LFS estimates.....	130
Iwona Foryś: Event history analysis in the resale study on the local housing market	141
Hanna Dudek, Joanna Landmesser: Impact of the relative deprivation on subjective income satisfaction	150
Grażyna Łaska: Numerical syntaxonomy in classification, identification and analysis of changes of secondary communities	160
Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analysis of relations between fundamental processes and capital market in China.....	166
Andrzej Bąk, Tomasz Bartłomowicz: Microeconomic polynomial models and their application in the analysis of preferences using R program.....	179

Andrzej Dudek, Bartosz Kwaśniewski: Parallel processing of clustering algorithms in CUDA technology	187
Michał Trzęsiok: Real estate market value estimation based on multivariate statistical analysis	196
Joanna Trzęsiok: On some simulative procedures for comparing nonparametric methods of regression.....	205
Artur Mikulec: Mojena and Wishart criterion in cluster analysis – the case of clusters with different covariance matrices	215
Artur Zaborski: Unfolding analysis by using gravity model	224
Justyna Wilk: Determination of problem and growth areas in Dolnośląskie Voivodship as regards human capital.....	235
Karolina Bartos: Risk analysis of bachelor students' university abandonment – the use of MLP networks	245
Ewa Genge: Clustering of industrial holiday participants with the use of latent class analysis.....	253
Izabela Kurzawa: Multinomial logit model as a tool to identify the factors affecting the housing situation of Polish households.....	261
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modelling class imbalance problems: comparing classification approaches for surgical risk analysis	270
Aleksandra Łuczak: The application of fuzzy hierarchical analysis to the evaluation of validity of strategic factors in administrative districts.....	281
Marcin Pełka: A spectral fuzzy c-means clustering algorithm for interval-valued symbolic data	289
Małgorzata Machowska-Szewczyk: Clustering algorithms for mixed-feature symbolic objects	299
Ewa Chodakowska: Malmquist index in enterprises classification on the basis of relative productivity changes	310
Beata Bieszk-Stolorz, Iwona Markowicz: Using proportional and non proportional Cox hazard models to research the chances for taking up a job according to the type of unemployment	320
Marcin Salamaga: Verification J.H. Dunning's theory of economic development by economic sectors in some EU countries	329
Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Socio-economic situation as a determinant of internal migration in Poland	342
Hanna Gruchociak: Delimitation of local labor markets in Poland on the basis of the employment-related population flows research.....	350
Radosław Pietrzyk: Selectivity and timing in Polish mutual funds performance measurement	361
Sabina Denkowska: Multiple testing procedures.....	369

Aleksandra Łuczak, Feliks Wysocki

Uniwersytet Przyrodniczy w Poznaniu

ZASTOSOWANIE MEDIANY PRZESTRZENNEJ WEBERA I METODY TOPSIS W UJĘCIU POZYCYJNYM DO KONSTRUKCJI SYNTETYCZNEGO MIERNIKA POZIOMU ŻYCIA

*Pamięci prof. dra hab. Wiesława Wagnera,
który był prekursorem badań z wykorzystaniem mediany Webera*

Streszczenie: Celem pracy jest przedstawienie możliwości wykorzystania do konstrukcji syntetycznego miernika rozwoju mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym. Mediana Webera zastosowana jest do normalizacji cech, zaś mediana brzegowa i medianowe odchylenie bezwzględne do obliczenia wartości cechy syntetycznej metodą TOPSIS. Statystyki te charakteryzują się dużą odpornością na występowanie obserwacji odstających, dlatego zastosowanie pozycyjnej metody TOPSIS jest korzystne w przypadkach, gdy w zbiorze występują cechy proste charakteryzujące się silną asymetrią lub obserwacjami odstającymi. Proponowane podejście zostało przedstawione na przykładzie dotyczącym budowy syntetycznego miernika poziomu życia ludności w powiatach województwa wielkopolskiego w 2010 r.

Słowa kluczowe: mediana Webera, porządkowanie liniowe, metoda TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution), poziom życia.

1. Wstęp

W rozpoznawaniu typów rozwojowych zjawisk społeczno-gospodarczych często pojawiają się cechy proste, które mają wartości odstające lub charakteryzują się silną asymetrią. W takim przypadku klasyczne metody budowy miernika syntetycznego mogą zawodzić, gdyż występowanie dla danego obiektu nawet jednej wartości odstającej (bardzo dużej, bądź bardzo małej) może wpłynąć w sposób istotny na przypisanie mu nadmiernie wysokiej (lub niskiej) rangi w klasyfikacji końcowej. W związku z tym w badaniach typologicznych dotyczących rankingu obiektów powinno zwracać się więcej uwagi na metody porządkowania liniowego odporne na to zjawisko, wśród nich na korzystające z L_1 -normy [Wysocki 2010]. Do takich można zaliczyć metody wykorzystujące medianę. Statystyka ta charakteryzuje się dużą odpornością na występowanie obserwacji odstających.

Celem pracy jest przedstawienie możliwości wykorzystania do konstrukcji syntetycznego miernika rozwoju mediany przestrzennej Webera i metody TOPSIS (*Technique for Order Preference by Similarity to an Ideal Solution*); [Hwang, Yoon 1981] w ujęciu pozycyjnym [Wysocki 2010]. Mediana Webera pozwala na taką konstrukcję miernika syntetycznego, w której uwzględnia się nie tylko dużą odporność na obserwacje odstające, ale również zależności występujące między badanymi cechami [Lira i in. 2002; Młodak 2006; Wysocki 2010]. Proponowane podejście zostało zastosowane do porządkowania liniowego powiatów ziemskich województwa wielkopolskiego według poziomu życia¹ ludności w 2010 r.

2. Metodyka badań

W procesie tworzenia cechy syntetycznej można wyróżnić sześć etapów postępowania. Pierwszym z nich jest wybór cech opisujących wybrane obiekty (etap I). Istnieją dwa podstawowe podejścia do wyboru cech – statystyczne i merytoryczne [Wysocki 2010].

Następnie należy ustalić kierunek preferencji cech w stosunku do rozpatrywanego kryterium ogólnego – poziomu życia ludności (etap II). Zmienne uznane za destymulanty można przekształcić w stymulanty za pomocą przekształcenia różnicowego:

$$x_{ik} = a - b \cdot x_{ik}^D,$$

gdzie: x_{ik}^D – wartość k -tej cechy będącej destymulantą ($k \in I_D$, gdzie I_D oznacza zbiór numerów cech będących destymulantami) w i -tym obiekcie (powiecie), ($i = 1, \dots, N$),

x_{ik} – wartość k -tej cechy przekształconej na stymulantę w i -tym obiekcie (powiecie),

a, b – stałe przyjmowane w sposób arbitralny, najczęściej $a = 0$ lub $a = \max_i (x_{ik}^D)$ oraz $b = 1$.

Wybrane cechy zazwyczaj są wyrażone w różnych jednostkach, dlatego należy stosować odpowiednie procedury normalizacyjne (etap III) [Wysocki 2010]. Głównym celem normalizacji jest sprowadzenie ich wartości do porównywalności, co polega na pozbawieniu ich mian i ujednoczeniu rzędów wielkości. W pracy zastosowano standaryzację medianową Webera opartą na formule [Lira, Wagner, Wysocki 2002; Młodak 2006; 2009]:

¹ „Poziom życia to ogół warunków, w których żyje społeczeństwo czy grupa społeczno-zawodowa. Obejmują one między innymi warunki pracy, poziom płacy i dochodów, stan opieki zdrowotnej i socjalnej, sytuację mieszkaniową, poziom wyposażenia infrastrukturalnego (w wodę, gaz, energię elektryczną, kanalizację), stan finansów publicznych, poziom ochrony środowiska” [Wysocki 2010].

$$z_{ik} = \frac{x_{ik} - m\tilde{e}d_k}{1,4826 \cdot m\tilde{a}d_k}, \quad (1)$$

- gdzie: x_{ik} – wartość k -tej cechy ($k=1, 2, \dots, K$) w i -tym obiekcie (powiecie) ($i=1, 2, \dots, N$),
- $m\tilde{e}d_k$ – składowa wektora medianowego Webera (mediana Webera) dla k -tej cechy,
- $m\tilde{a}d_k = med_i |x_{ik} - m\tilde{e}d_k|$ – medianowe odchylenie bezwzględne będące medianą z bezwzględnych odchyień wartości cechy od składowej mediany Webera odpowiadającej k -tej cesze,
- 1,4826 jest stałym współczynnikiem skalowania, który zależy od rozkładu wartości cech ($\sigma \approx E(1,4826 \cdot m\tilde{a}d_k (X_1, X_2, \dots, X_K))$,
- σ – odchylenie standardowe) (zob. [Młodak 2006; 2009]).

Wektor medianowy określony według kryterium Webera, tj. kryterium odległości euklidesowej, nazywany jest także „ L_1 medianą”, „medianą przestrzenną”, „centrum medianowym” lub „punktem Webera” i jest zdefiniowany następująco [Lira, Wagner, Wysocki 2002].

Niech $K_N^K = \{X_1, X_2, \dots, X_N\} \in R^K$ będzie zbiorem N wektorów obserwacji obiektów K -cechowych i niech $\hat{\theta} \in R^K$ będzie takim wektorem, iż jest on rozwiązaniem zagadnienia optymalizacyjnego

$$T(\hat{\theta}, \Gamma_N^K) = \min_{\theta \in R^K} T(\theta, \Gamma_N^K), \quad (2)$$

gdzie funkcja celu tego zagadnienia przyjmuje postać

$$T(\theta, \Gamma_N^K) = \sum_{i=1}^N \left[\sum_{k=1}^K (x_{ik} - \theta_k)^2 \right]^{1/2}, \quad (3)$$

przy czym $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iK})^T$, $i=1, 2, \dots, N$ oraz $\theta = (\theta_1, \theta_2, \dots, \theta_K)^T$.

Rozwiązaniem powyższego zagadnienia optymalizacyjnego jest wektor medianowy Webera oznaczony przez $\hat{\theta} = (m\tilde{e}d_1, m\tilde{e}d_2, \dots, m\tilde{e}d_K)^T$ [Wysocki 2010].

W kolejnym etapie (IV) ustalone zostają współrzędne obiektów modelowych – wzorca, np. według zasady:

$$A^+ = \left(\max_i (z_{i1}), \max_i (z_{i2}), \dots, \max_i (z_{iK}) \right) = (z_1^+, z_2^+, \dots, z_K^+) \quad (4)$$

i antywzorca rozwoju:

$$A^- = \left(\min_i(z_{i1}), \min_i(z_{i2}), \dots, \min_i(z_{iK}) \right) = (z_1^-, z_2^-, \dots, z_K^-). \quad (5)$$

Jest to podstawą do obliczenia oddalenia każdego ocenianego obiektu od wzorca A^+ i antywzorca rozwoju A^- (etap IV) [Wysocki 2010]:

$$d_i^+ = \text{med}_k(|z_{ik} - z_k^+|), \quad d_i^- = \text{med}_k(|z_{ik} - z_k^-|) \quad (i = 1, 2, \dots, N), \quad (6)$$

gdzie: d_i^+ i d_i^- – medianowe odchylenie bezwzględne od wzorca A^+ i antywzorca rozwoju A^- dla i -tego obiektu,

med_k – mediana brzegowa dla k -tej cechy.

W przypadku podejścia klasycznego podstawą tworzenia miernika syntetycznego są odległości euklidesowe każdego ocenianego obiektu od wzorca (z^+) (metoda TOPSIS i Hellwiga) [Hwang, Yoon 1981; Wysocki 2010]:

$$d_i^+ = \sqrt{\sum_{k=1}^K (z_{ik} - z_k^+)^2} \quad i = 1, 2, \dots, N \quad (7)$$

i antywzorca rozwoju (z^-) (metoda TOPSIS):

$$d_i^- = \sqrt{\sum_{k=1}^K (z_{ik} - z_k^-)^2}. \quad (8)$$

Do konstrukcji miernika syntetycznego zastosowano metodę TOPSIS za pomocą miernika (etap V) [Hwang, Yoon 1981; Wysocki 2010]:

$$S_i = \frac{d_i^-}{d_i^- + d_i^+}, \quad (i = 1, 2, \dots, N), \quad (9)$$

przy czym $0 \leq S_i \leq 1$.

Im mniejsza jest odległość danego obiektu od obiektu modelowego – wzorca rozwoju, a tym samym większa od drugiego bieguna – antywzorca rozwoju, tym wartość miernika syntetycznego jest bliższa 1.

W przypadku pozycyjnej metody Hellwiga miernik syntetyczny przyjmuje postać [Lira, Wagner, Wysocki 2002]:

$$H_i^{(1)} = 1 - \frac{d_i^+}{d_0}, \quad (10)$$

gdzie: $d_i^+ = \text{med}_k(|z_{ik} - z_k^+|)$, $d_0 = \text{med}(d_i^+) + 2,5 \cdot \text{mad}(d_i^+)$, $\text{mad}(d_i^+) = \text{med}(|d_i^+ - \text{med}(d_i^+)|)$, 2,5 jest tzw. odpornościową wartością progową,

która wyznacza barierę korzystnych wartości pomiarów odległości obiektów od wzorca rozwoju [Rousseuw i Leroy 1987; Młodak 2006; 2009]. W przypadku miernika H_i w skrajnych przypadkach, gdy obiekt jest znacznie oddalony od wzorca, mogą wystąpić wartości poniżej 0.

Wyznaczone wartości syntetycznego miernika wykorzystuje się do liniowego porządkowania powiatów oraz – na tej podstawie – do wyodrębnienia ich klas typologicznych (etap VI). Wyodrębnienie klas dla całego obszaru zmienności cechy syntetycznej może zostać przeprowadzone metodami statystycznymi z zastosowaniem średniej i odchylenia standardowego z wartości cechy syntetycznej lub w sposób arbitralny, przyjmując np. przedziały liczbowe wartości miernika S_i : $\langle 0,00;0,20 \rangle$ – poziom bardzo niski, $\langle 0,20;0,40 \rangle$ – poziom niski, $\langle 0,40;0,50 \rangle$ – poziom średni-niższy, $\langle 0,50;0,60 \rangle$ – poziom średni-wyższy, $\langle 0,60;0,80 \rangle$ – poziom wysoki, $\langle 0,80;1,00 \rangle$ – poziom bardzo wysoki.

3. Wyniki badań

W badaniach poziomu życia ludności wykorzystano dane statystyczne z Głównego Urzędu Statystycznego (*Bank Danych Lokalnych* 2010). W pierwszym etapie oceny poziomu życia ludności dokonano wyboru 10 cech opisujących powiaty województwa wielkopolskiego: 1) udział ludności obsługiwanej przez sieć kanalizacyjną (%) (X_1), 2) udział ludności obsługiwanej przez sieć gazową (%) (X_2), 3) drogi gminne w powiecie o nawierzchni twardej w km na 100 km² (X_3), 4) zanieczyszczenia gazowe w tonach na rok na km² (X_4), 5) zanieczyszczenia pyłowe w tonach na rok na km² (X_5), 6) lekarze według podstawowego miejsca pracy na 1 tys. ludności (X_6), 7) ludność na jedną aptekę (X_7), 8) przeciętne miesięczne wynagrodzenia brutto w zł (X_8), 9) stopa bezrobocia w % (X_9), 10) dochody własne gmin na mieszkańca w zł (średnia z 5 lat 2006-2010) (X_{10}).

W drugim etapie przyjęto, że cztery cechy mają charakter destymulant (X_4 , X_5 , X_7 i X_9), a pozostałe – stymulant (etap II). Cechy o charakterze destymulant zostały przekształcone w stymulanty za pomocą przekształcenia różnicowego: $x_{ik} = a - b \cdot x_{ik}^D$, przyjmując że $a = \max(x_{ik}^D)$ w zbiorze powiatów Polski (z wyjątkiem X_4 , gdzie przyjęto wartości maksymalne dla powiatów województwa) oraz $b = 1$. Opis statystyczny cech przedstawiono w tab. 1.

Jak wynika z obliczeń, cechy X_4 i X_5 związane z emisją zanieczyszczeń gazowych i pyłowych charakteryzują się wysoką asymetrią prawostronną. Współczynnik skośności dla cechy X_4 wynosi 5,51, a dla X_5 5,38. Cechy te wykazują również bardzo dużą zmienność ich wartości (ponad 300%). Taką sytuację w przypadku cech X_4 i X_5 wywołują obserwacje skrajne, zaobserwowane dla powiatu tureckiego, gdzie

emisja zanieczyszczeń gazowych przekracza średnią dla powiatów w województwie wielkopolskim aż ponad 22-krotnie, a zanieczyszczeń pyłowych – ponad 17-krotnie.

Tabela 1. Statystyki opisowe charakteryzujące cechy dotyczące poziomu życia mieszkańców powiatów województwa wielkopolskiego

Charakterystyki	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Statystyki klasyczne										
Max	73,70	70,20	79,25	4 366 436,00	2314,00	1,55	9043,00	3161,97	18,80	1964,25
Min	22,90	0,00	9,37	0,00	0,00	0,14	2783,00	2106,84	3,50	686,79
Średnia	51,09	34,71	34,85	195 559,29	132,52	0,85	4064,42	2743,07	11,61	1066,56
Odchylenie standardowe	12,06	21,28	16,79	764 088,12	402,62	0,36	1106,66	240,59	3,77	214,48
Współczynnik zmienności (%)	23,60	61,31	48,19	390,72	303,83	41,94	27,23	8,77	32,48	20,11
Skośność	-0,35	-0,19	0,28	5,51	5,38	-0,03	2,90	-0,27	0,06	2,24
Statystyki pozycyjne										
Wektor medianowy Webera (LI-mediana) ^{a)}	52,88	38,81	36,11	34353,83	49,50	0,81	4522,11	2720,27	11,70	1032,38
Medianowe odchylenie bezwzględne	6,52	18,29	13,88	30286,83	43,50	0,19	587,11	162,18	2,70	97,03
Współczynnik zmienności pozycyjny (%)	12,34	47,12	38,43	88,16	87,88	23,46	12,98	5,96	23,09	9,40
Wartości modelowe ^{b)}										
Wzorzec	99,70	99,10	186,42	0,00	0,00	3,01	2246,00	4628,24	3,50	3498,34
Antywzorzec	8,60	0,00	2,42	4 366 436,00	2406,00	0,04	11749,00	2106,84	36,00	523,52

a) Obliczenia wykonano z wykorzystaniem programu R; b) Wartości modelowe ustalono, przyjmując wartości maksymalne (dla stymulant) i minimalne (dla destymulant) cech w zbiorze wszystkich powiatów w Polsce (bez wartości skrajnie odstających).

Źródło: obliczenia własne na podstawie danych *Banku Danych Regionalnych* [2010].

W przypadku udziału ludności obsługiwanej przez sieć gazową (X_2) należy zauważyć wysoki stopień rozproszenia wartości cechy. Współczynnik zmienności dla tej cechy wynosi 61,31%. Również cechy X_3 i X_6 charakteryzują się znacznym zróżnicowaniem ich wartości (ponad 40%).

W kolejnym etapie wartości cechy poddano normalizacji, wykorzystując standaryzację medianową Webera² (wzory 1-3, etap III). Zstandaryzowane wartości cech umożliwiły wyznaczenie medianowego odchylenia bezwzględnego każdego ocenianego powiatu od wzorca i antywzorca rozwoju (wzór 6, etap IV). Wartości wzorca ustalono, przyjmując wartości maksymalne cech, a antywzorca – ich wartości minimalne w zbiorze wszystkich powiatów ziemskich w Polsce³.

² Obliczenia wykonano z wykorzystaniem pakietu ClusterSim w programie R.

³ Przyjęcie wartości modelowych na podstawie powiatów ziemskich w Polsce miało na celu określenie pozycji powiatów w województwie wielkopolskim w odniesieniu do sytuacji w kraju.

Następnie obliczono wartości syntetycznego miernika rozwoju metodą TOPSIS ($S^{(1)}$) (wzór 9, etap V). Zastosowanie metody pozycyjnej TOPSIS opartej na medianie Webera i medianowych odchyleniach bezwzględnych od wzorca i antywzorca rozwoju jest tutaj uzasadnione, gdyż w zbiorze wystąpiły cechy charakteryzujące się silną asymetrią i obserwacjami odstającymi. Takie właściwości rozkładu wykazywały zwłaszcza trzy cechy (tab. 1): zanieczyszczenia gazowe i pyłowe w tonach na rok na km² (X_4 , X_5), charakteryzujące się wysoką asymetrią prawostronną (odpowiednio 5,51 i 5,38, na co istotnie wpływał bardzo wysoki poziom zanieczyszczeń w powiecie tureckim), oraz odsetek ludności obsługiwanej przez sieć gazową (X_2), charakteryzujący się dużym zróżnicowaniem w zbiorowości badanych powiatów (61,3%).

W celach porównawczych wyznaczono również wartości syntetycznych mierników, stosując metodę Hellwiga (wzór 10, $H^{(1)}$) w ujęciu pozycyjnym przy wykorzystaniu standaryzacji medianowej opartej na wektorze medianowym Webera i medianowych odchyleniach bezwzględnych od wzorca rozwoju oraz klasyczne metody TOPSIS ($S^{(2)}$) i Hellwiga ($H^{(2)}$) z wykorzystaniem klasycznej formuły standaryzacji cech i odległości euklidesowych obiektów od wzorca i antywzorca rozwoju. Powiaty zostały uporządkowane od najlepszego do najgorszego według wartości cechy syntetycznej $S_i^{(1)}$ (tab. 2).

Należy zauważyć istotne różnice pomiędzy uporządkowaniami powiatów metodami pozycyjnymi i klasycznymi (tab. 3). W przypadku metody TOPSIS zgodność uporządkowania według metody pozycyjnej i klasycznej wynosi 0,838, a dla podejść opartych na metodzie Hellwiga – 0,613. Przykładem znacznych rozbieżności w uporządkowaniu może być powiat śremski. W podejściu $S^{(1)}$ uplasował się on na szesnastym miejscu pod względem poziomu życia, a w podejściu $S^{(2)}$ uzyskał pozycję trzecią. Natomiast powiat gostyński w podejściu $S^{(1)}$ uplasował się na dwudziestym miejscu, a w podejściu $S^{(2)}$ o dziewięć pozycji niżej. Z kolei w przypadku podejść opartych na metodzie Hellwiga rozbieżności są jeszcze bardziej znaczące. Przykładem może być powiat pilski, który w podejściu $H^{(1)}$ zyskał dwudziestą czwartą pozycję, a w podejściu $H^{(2)}$ uplasował się aż o dwadzieścia dwie pozycje wyżej. Różnice te wynikają przede wszystkim z zastosowanej metody pomiaru odległości pomiędzy obiektami a wzorcem rozwoju w przypadku metody Hellwiga, a także antywzorcem rozwoju w metodzie TOPSIS. Zastosowanie medianowego odchylenia bezwzględnego, czyli miary tendencji centralnej, pozwala zlokalizować środek zbioru wartości bezwzględnych różnic pomiędzy każdym obiektem wielocechowym a wzorcem (antywzorcem) rozwoju. Natomiast w przypadku klasycznego podejścia opartego na

W przypadku zanieczyszczeń gazowych (X_4) jako wzorec przyjęto wartości maksymalne dla powiatów województwa wielkopolskiego ze względu na obecność w Polsce powiatu belchatowskiego, gdzie emisja zanieczyszczeń przekraczała maksymalną wartość dla powiatów w województwie wielkopolskim blisko siedmiokrotnie. W przypadku cechy dotyczącej przeciętnego miesięcznego wynagrodzenia brutto wzorec wyznaczono, odrzucając obserwację odstającą dla powiatu lubińskiego, gdzie ze względu na charakter przemysłowo-górnicy powiatu przeciętne wynagrodzenie przekraczało 6 tys. zł (przy średniej dla województwa wielkopolskiego wynoszącej około 2,7 tys. zł).

Tabela 2. Wartości syntetycznych mierników oraz klasyfikacja powiatów województwa wielkopolskiego według poziomu życia ludności

Lp.	Powiaty ^{a)}	Wartości syntetycznego miernika/podejścia				Rangi powiatów/podejścia				Klasa i poziom życia ludności	
		TOPSIS			Hellwig	TOPSIS		Hellwig			
		$S^{(1)}$	$S^{(2)}$	$H^{(1)}$	$H^{(2)}$	$S^{(1)}$	$S^{(2)}$	$H^{(1)}$	$H^{(2)}$		
1	poznański	0,682	0,535	0,472	0,284	1	1	1	1	I wysoki	
2	chodzieski	0,621	0,453	0,421	0,161	2	5	2	6		
3	kościański	0,590	0,447	0,397	0,127	3	9	3	10	II średni- wyższy	
4	średzki	0,570	0,452	0,382	0,163	4	6	4	5		
5	międzychodzki	0,570	0,445	0,367	0,110	5	10	5	14		
6	gnieźnieński	0,555	0,459	0,334	0,167	6	4	6	3		
7	szamotulski	0,543	0,448	0,284	0,157	7	7	12	7		
8	piłski	0,540	0,482	0,061	0,212	8	2	24	2		
9	ostrowski	0,537	0,447	0,306	0,167	9	8	8	4		
10	wolsztyński	0,537	0,433	0,287	0,092	10	14	11	21		
11	obornicki	0,535	0,436	0,297	0,124	11	13	10	11		
12	kępiński	0,533	0,426	0,256	0,079	12	18	14	23		
13	nowotomyski	0,530	0,440	0,233	0,111	13	12	15	13		
14	jarociński	0,524	0,419	0,324	0,102	14	21	7	19		
15	koniński	0,515	0,411	0,149	0,110	15	24	21	15		
16	śremski	0,499	0,463	0,083	0,154	16	3	23	8		III średni- niższy
17	rawicki	0,480	0,417	0,262	0,093	17	22	13	20		
18	grodziski	0,477	0,431	0,232	0,081	18	16	16	22		
19	krotoszyński	0,475	0,432	0,302	0,103	19	15	9	18		
20	gostyński	0,465	0,444	0,162	0,137	20	11	20	9		
21	ostrzeszowski	0,450	0,425	0,145	0,107	21	19	22	16		
22	złotowski	0,439	0,390	0,164	0,049	22	28	19	28		
23	wągrowiecki	0,434	0,401	0,217	0,071	23	27	17	25		
24	leszczyński	0,429	0,404	-0,067	0,044	24	26	29	30		
25	wrzesiński	0,425	0,420	0,166	0,103	25	20	18	17		
26	kolski	0,419	0,429	0,033	0,121	26	17	25	12		
27	pleszewski	0,402	0,410	0,019	0,070	27	25	27	26		
28	czarnkowsko-trzcianecki	0,396	0,389	-0,002	0,054	28	29	28	27	IV niski	
29	słupecki	0,366	0,412	0,024	0,073	29	23	26	24		
30	turecki	0,271	0,355	-0,489	0,047	30	31	31	29		
31	kaliski	0,240	0,358	-0,309	-0,017	31	30	30	31		
	max	0,682	0,535	0,472	0,284						
	min	0,240	0,355	-0,489	-0,017				×		
	rozstęp	0,442	0,179	0,961	0,302						

^{a)} Uporządkowanie liniowe według wartości syntetycznego miernika rozwoju uzyskanego metodą $S^{(1)}$.

Źródło: obliczenia własne na podstawie danych statystycznych Głównego Urzędu Statystycznego z *Banku Danych Lokalnych* (2010), <http://www.stat.gov.pl/bdl>.

odległości euklidesowej uwzględnia się kwadraty odchyleń wartości cech między obiektem badanym i modelowym, a następnie dokonuje się ich sumowania po wszystkich cechach.

Metoda pozycyjna TOPSIS dostarczyła w porównaniu z pozostałymi miernikami większego zakresu zmienności miernika syntetycznego (od 0,240 do 0,689), co pozwoliło poprawniej określić rangi i typy rozwojowe powiatów (obejmujące poziomy „wysoki”, „średni” i „niski”). Natomiast metoda pozycyjna Hellwiga dostarczyła wprawdzie większego zakresu zmienności miernika, obejmującego wartości w granicach od $-0,489$ do $0,472$, lecz nie pozwalają one (zwłaszcza uzyskane wartości ujemne i bliskie zeru) na przeprowadzenie sensownej identyfikacji typów rozwojowych poziomu życia mieszkańców.

Z kolei mierniki klasyczne TOPSIS i Hellwiga przy przyjętych wartościach wzorca i antywzorca rozwoju nie odzwierciedlają zadowalająco różnic międzyklasowych w zakresie poziomu życia. Biorąc pod uwagę otrzymane ich zakresy zmienności (w przypadku metody TOPSIS wartości miernika mieszczą się w przedziale $\langle 0,355; 0,535 \rangle$, a w przypadku metody Hellwiga w przedziale $\langle -0,017; 0,284 \rangle$), większość powiatów można by zakwalifikować do klasy reprezentującej średni (metoda TOPSIS) albo niski i bardzo niski (metoda Hellwiga) poziom życia. Można sądzić, że wpływ na zmniejszenie zakresu zmienności mierników (TOPSIS i Hellwiga), a także przesunięcie wszystkich powiatów do klasy o najniższym poziomie życia (według miernika Hellwiga) miało przyjęcie większych rozpiętości wartości wzorca i antywzorca rozwoju na podstawie danych dla kraju (w porównaniu z wartościami dla województwa). Również nieuwzględnianie wartości antywzorca rozwoju w konstrukcji miernika Hellwiga prowadzi do nadmiernego obniżenia jego wartości.

Na podstawie uporządkowanych wartości cechy syntetycznej uzyskanych pozycyjną metodą TOPSIS wyodrębniono cztery typy poziomu życia mieszkańców województwa wielkopolskiego (tab. 2). Pierwszy typ utworzył powiat poznański i chodzieski. Są to powiaty o najwyższym poziomie życia mieszkańców. Drugi typ utworzyło trzynaście powiatów o średnim-wyższym poziomie jakości życia, głównie z północnej i środkowej części województwa. Kolejny, trzeci typ obejmuje dwanaście powiatów głównie z południowej i północnej części województwa. Powiaty te cechują się średnim-niższym poziomem życia ludności. Czwarty typ obejmuje obszar czterech powiatów głównie ze wschodniej części województwa. Są to tereny o najniższym poziomie życia ludności.

4. Podsumowanie

Na podstawie przeprowadzonych obliczeń i analiz można sformułować następujące stwierdzenia i wnioski. Zaproponowane podejście do porządkowania liniowego obiektów oparte na metodzie TOPSIS i medianie pozycyjnej Webera może być zastosowane do wyznaczenia syntetycznego miernika rozwoju w przypadku, gdy w zbiorze cech opisujących badane obiekty pojawiają się obserwacje odstające lub

silna asymetria. Występowanie dla danego obiektu nawet jednej wartości odstającej (bardzo dużej bądź bardzo małej) może wpłynąć w sposób istotny na przypisanie mu nadmiernie wysokiej (lub niskiej) rangi w klasyfikacji końcowej, co ujawnia się w przypadku stosowania metod klasycznych (TOPSIS i Hellwiga). Istota metod klasycznych polega bowiem na obliczeniu i sumowaniu kwadratów odchyłeń każdego ocenianego obiektu wielocechowego od wzorca (antywzorca) rozwoju. Z kolei zastosowanie tych metod w ujęciu pozycyjnym, tj. z wykorzystaniem medianowego odchylenia bezwzględnego, pozwala zlokalizować środek zbioru wartości różnic bezwzględnych pomiędzy każdym obiektem wielocechowym a wzorcem (antywzorcem) rozwoju, co umożliwia pominięcie obserwacji odstających w konstrukcji miernika syntetycznego.

Metoda TOPSIS w ujęciu pozycyjnym jest odporna na występowanie wartości odstających cech i przyjętych wartości antywzorca i wzorca rozwoju. Utworzona na jej podstawie typologia powiatów dobrze odzwierciedliła różnice międzyklasowe w poziomie życia mieszkańców, obejmując swoim zakresem powiaty zarówno o niskim, średnim, jak i o wysokim poziomie życia w odniesieniu do wszystkich powiatów ziemskich w Polsce.

Metoda Hellwiga w ujęciu pozycyjnym dostarczyła wprawdzie szerokiego zakresu zmienności miernika, gdzie jego rozpiętość wartości waha się w granicach od $-0,489$ do $0,472$, jednak ze względu na wartości ujemne i bliskie zeru nie dała podstawy do poprawnej typologii powiatów.

Przy przyjętych wartościach wzorca i antywzorca rozwoju mierniki uzyskane za pomocą klasycznych metod TOPSIS i Hellwiga słabo odzwierciedliły różnice międzyklasowe w poziomie życia. Biorąc pod uwagę wartości miernika uzyskane metodą TOPSIS, większość powiatów można by zakwalifikować do klasy reprezentującej średni bądź średni-niższy poziom życia. Jeszcze mniej przydatna wydaje się typologia oparta na mierniku Hellwiga, który dał podstawę do wyróżnienia tylko niskiego i bardzo niskiego poziomu życia mieszkańców badanych powiatów.

Można zatem stwierdzić, że przyjęcie większych rozpiętości między wzorcem i antywzorcem rozwoju, a tym samym większego ich oddalenia od wartości cech przyjętych do konstrukcji miernika, prowadzi w przypadku metod klasycznych (Hellwig, TOPSIS) do gorszej jakości klasyfikacji – zmniejszenia zakresu zmienności wartości syntetycznego miernika rozwoju.

Literatura

- Bank Danych Lokalnych* <http://www.stat.gov.pl/bdl> 2010.
- Hwang C.L., Yoon K., *Multiple Attribute Decision Making: Methods and Applications*, Springer, Berlin 1981.
- Lira J., Wagner W., Wysocki F., *Mediana w zagadnieniach porządkowania obiektów wielocechowych*, [w:] *Statystyka regionalna w służbie samorządu terytorialnego i biznesu*, red. J. Paradysz, Akademia Ekonomiczna w Poznaniu, Poznań 2002, s. 87-99.
- Młodak A., *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa 2006.

Młodak A., *Historia problemu Webera*, „Matematyka Stosowana” 2009, nr 10, s. 3-21.

Rousseeuw P.J., Leroy A.M., *Robust Regression and Outlier Detection*, John Wilney and Sons, New York 1987.

Wysocki F., *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wyd. Uniwersytetu Przyrodniczego w Poznaniu, Poznań 2010.

THE APPLICATION OF SPATIAL MEDIAN OF WEBER AND THE METHOD TOPSIS IN POSITIONAL FORMULATION FOR THE CONSTRUCTION OF SYNTHETIC MEASURE OF STANDARD OF LIVING

Summary: The aim of this paper is to investigate the applicability of the spatial median of Weber and the method TOPSIS for the construction of synthetic characteristics. The Weber median is applied to the standardization of characteristics, whereas the marginal median and the median absolute deviation to calculate the values of synthetic characteristic by TOPSIS method. These statistics are characterized by large resistance on the occurrence of outliers, therefore the use of positional TOPSIS method is profitable in cases when there are characteristics with strong asymmetry or outliers in the set of characteristics. The proposed procedure was applied to estimate the standard of living in administrative districts in Wielkopolska Voivodeship.

Keywords: Weber median, linear ordering, technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), standard of living.