

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

Taksonomia 27

**Klasyfikacja i analiza danych –
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

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

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2016

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp	9
Beata Bal-Domańska: Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....	11
Tomasz Bartłomowicz: Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers’ preferences.....	20
Maciej Beręsewicz, Marcin Szymkowiak: Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....	30
Beata Bieszk-Stolorz: Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....	40
Justyna Brzezińska: Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....	49
Mariola Chrzanowska, Nina Drejerska: Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union.....	58
Sabina Denkowska: Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....	66
Adam Depta: Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jaskających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....	76
Mariusz Doszyń, Sebastian Gnat: Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....	84

Marta Dziechciarz-Duda, Anna Król: Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences	94
Ewa Genge: Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour	103
Joanna Górna, Karolina Górna: Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools	112
Alicja Grześkowiak: Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population	122
Agnieszka Kozera, Feliks Wysocki: Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods	131
Mariusz Kubus: Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....	143
Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski: Analiza wydźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....	153
Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki: Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R	165
Dorota Rozmus: Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method	176
Adam Sagan, Aneta Rybicka, Justyna Brzezińska: <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....	184
Michał Stachura: O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case	195
Michał Stachura, Barbara Wodecka: Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events	205
Iwona Staniec, Jan Żółtowski: Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości	

technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....	215
Marcin Szymkowiak, Wojciech Roszka: Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....	224
Lucyna Wojcieszka: Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy.....	234

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

Agnieszka Kozera, Feliks Wysocki

Uniwersytet Przyrodniczy w Poznaniu
e-mails: {akozera; wysocki}@up.poznan.pl

**PROBLEM USTALANIA WSPÓLRZĘDNYCH
OBIEKTÓW MODELOWYCH
W METODACH PORZĄDKOWANIA
LINIOWEGO OBIEKTÓW**

**THE PROBLEM OF DETERMINING
THE COORDINATES OF MODEL OBJECTS
IN OBJECT LINEAR ORDERING METHODS**

DOI: 10.15611/pn.2016.427.14

Streszczenie: Celem pracy jest analiza porównawcza wyników porządkowania liniowego obiektów i identyfikacji typów rozwojowych zjawisk złożonych przy zastosowaniu metody TOPSIS i trzech podejść do ustalania współrzędnych obiektów modelowych, tj. klasycznego, pozycyjnego z wykorzystaniem mediany przestrzennej Webera oraz proponowanego w artykule podejścia kwartyłowego, polegającego na korygowaniu współrzędnych obiektów modelowych w odniesieniu do cech mających wartości odstające poprzez ich identyfikację i przypisanie im wartości modelowych. Zastosowanie podejścia kwartyłowego lub pozycyjnego umożliwia bardziej jednoznaczną identyfikację typów rozwojowych w przypadkach, gdy w zbiorze rozpatrywanych cech występują takie, które charakteryzują się silną asymetrią lub obserwacjami odstającymi. Proponowane podejście zostało przedstawione na przykładzie dotyczącym budowy miernika syntetycznego i identyfikacji typów poziomu samodzielności finansowej gmin wiejskich województwa wielkopolskiego w 2013 r.

Słowa kluczowe: współrzędne obiektów modelowych, metody wzorcowe porządkowania liniowego obiektów, obserwacje odstające, samodzielność finansowa JST, gminy wiejskie.

Summary: The aim of the study is a comparative analysis of the results of linear ordering of objects and identification of developmental types of complex phenomena based on the TOPSIS method and three approaches to determination of the coordinates of model objects, i.e. the classic approach, the positional approach based on Weber's spatial median, and the quartile approach which is suggested in this article. The article presents the suggested approach with the case study concerning the construction of a synthetic measure and identification of the types of financial self-sufficiency in rural communes of Wielkopolska province in 2013.

Keywords: coordinates of the object model, reference methods of linear ordering of objects, outliers, financial self-sufficiency, rural communes.

1. Wstęp

Procedura budowy cechy syntetycznej jest procesem wieloetapowym, a prowadzący badania na każdym jego etapie podjąć musi wiele decyzji, dotyczących m.in. tego, jakie przyjąć metody doboru cech prostych do badań, jakie zastosować podejście do ustalania systemu wag czy też jaką przyjąć procedurę normalizacji lub agregacji cech prostych. Ważnym zagadnieniem, które nie było dotychczas szeroko poruszane w artykułach naukowych i literaturze przedmiotu, jest również problem ustalania współrzędnych obiektów modelowych w metodach wzorcowych porządkowania liniowego obiektów. Najczęściej jako jednostki modelowe przyjmuje się wartości maksymalną i minimalną dla danej cechy prostej, jako współrzędne odpowiednio wzorca i antywzorca rozwoju, rzadziej są to wartości modelowe ustalone przez ekspertów. Jak wykazały dotychczasowe badania, w pierwszym przypadku, często to postępowanie okazuje się nie najwłaściwsze [Łuczak, Wysocki 2013]. Jeżeli bowiem w zbiorze wartości cech pojawią się cechy mające wartości odstające lub charakteryzujące się silną asymetrią, to przyjęcie wartości maksymalnych i minimalnych ze zbioru wartości cech jako współrzędnych modelowych prowadzi do nadmiernie dużego ich oddalenia od wartości typowych rozważanych cech i w konsekwencji do zmniejszenia zakresu zmienności wartości konstruowanego miernika rozwoju [Wysocki 2010]. W rezultacie uzyskane wartości miernika syntetycznego mogą pokryć tylko niewielką część potencjalnego obszaru jego zmienności (najczęściej określanego przedziałem $[0, 1]$) i wtedy pojawić mogą się problemy związane z identyfikacją typów rozwojowych badanych obiektów¹ [Łuczak, Wysocki 2013].

Propozycja rozwiązania tego problemu po raz pierwszy przedstawiona została m.in. przez Lirę i in. [Lira, Wagner, Wysocki 2002] i polegała na zastosowaniu metody pozycyjnej opartej na medianie przestrzennej Webera do konstrukcji cechy syntetycznej, w sytuacji gdy w zbiorze wartości cech występują wartości odstające. Statystyka ta charakteryzuje się bowiem dużą odpornością na występowanie obserwacji odstających. W niniejszym artykule przedstawiono kolejną propozycję rozwiązania tego problemu – podejście kwartyłowe, polegające na korygowaniu współrzędnych obiektów modelowych (wartości maksymalnych i minimalnych) w odniesieniu do cech mających wartości odstające poprzez ich identyfikację i przypisanie im wszystkim wartości modelowych (wzorca i antywzorca rozwoju).

¹ Problem identyfikacji typów rozwojowych pojawia się w wielu pracach. Na przykład A. Majka [2015] dokonała oceny poziomu życia ludności według województw w latach 2004, 2008 i 2012. Zastosowała metodę Hellwiga i, wykorzystując średnią i odchylenie standardowe z wartości miernika syntetycznego (podejście statystyczne), uzyskała cztery klasy (typy) poziomu życia. Do klasy I reprezentującej wysoki poziom życia zaliczyła województwa (śląskie, dolnośląskie, małopolskie), dla których „syntetyczne mierniki poziomu życia przyjmowały zdecydowanie niskie wartości. Maksymalna wartość miernika była niższa od 0,4” [Majka 2015, s. 33].

Celem artykułu jest więc analiza porównawcza wyników porządkowania liniowego obiektów i identyfikacji typów rozwojowych badanych zjawisk złożonych przy zastosowaniu metody TOPSIS i trzech podejść do ustalania współrzędnych obiektów modelowych, tj. klasycznego, pozycyjnego z wykorzystaniem mediany przestrzennej Webera oraz proponowanego w pracy podejścia kwartylowego. Problem badawczy został przedstawiony na przykładzie dotyczącym budowy miernika syntetycznego i identyfikacji typów poziomu samodzielności finansowej gmin wiejskich województwa wielkopolskiego w 2013 r. Dane empiryczne pochodziły z Głównego Urzędu Statystycznego [GUS 2013] oraz Ministerstwa Finansów [2013].

2. Metodyka badań

Procedura budowy cechy syntetycznej jest procesem wieloetapowym, w którym można wyróżnić 6 etapów [Wysocki 2010]. W artykule szczególną uwagę zwrócono na etap określenia współrzędnych obiektów modelowych i konsekwencje zastosowania wybranych podejść dla uzyskanych wyników, tj. otrzymanego zakresu zmienności wartości miernika syntetycznego i przeprowadzonej na ich podstawie klasyfikacji typologicznej obiektów.

Pierwszy etap konstrukcji cechy syntetycznej obejmuje (w oparciu o kryteria merytoryczne i statystyczne) dobór cech opisujących wybrane obiekty, a także ustalenie kierunku ich preferencji w stosunku do rozpatrywanego kryterium ogólnego. Cechy uznane za destymulanty można przekształcić w stymulanty za pomocą przekształcenia różnicowego [Wysocki 2010]:

$$x_{ik} = a - b \cdot x_{ik}^D, \quad (1)$$

gdzie: x_{ik}^D – wartość k -tej cechy będącej destymulantą ($k \in I_D$) dla i -tego obiektu ($i = 1, \dots, N$); x_{ik} – wartość k -tej cechy ($k = 1, \dots, K$) przekształconej w stymulantę dla i -tego obiektu; a, b – stałe przyjmowane w sposób arbitralny, najczęściej $a = 0$ oraz $b = 1$.

W kolejnym etapie (II) polegającym na normalizacji wartości cech prostych, w podejściu klasycznym zastosowano standaryzację klasyczną:

$$z_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k}, \quad (2)$$

natomiast w podejściu pozycyjnym standaryzację medianową Webera [Lira, Wagner, Wysocki 2002; Młodak 2006]:

$$z_{ik} = \frac{x_{ik} - m\tilde{d}_k}{1,4826 \cdot m\tilde{d}_k}, \quad (3)$$

gdzie: x_{ik} – wartość k -tej cechy w i -tym obiekcie; \bar{x}_k, s_k – odpowiednio średnia arytmetyczna i odchylenie standardowe dla k -tej cechy; $m\tilde{d}_k$ – składowa wektora medianowego Webera (mediana Webera) dla k -tej cechy;

$m\ddot{a}d_k = med_i |x_{ik} - m\ddot{e}d_k|$ – medianowe odchylenie bezwzględne będące medianą z bezwzględnych odchyleń wartości cechy od składowej mediany Webera odpowiadającej k -tej cesze; 1,4826 jest stałym współczynnikiem skalowania, który zależy od rozkładu wartości cech ($\sigma \approx E(1,4826 \cdot m\ddot{a}d_k(X_1, X_2, \dots, X_K))$), σ – odchylenie standardowe.

Rozkład zestandaryzowanych w ten sposób wartości cech „jest zbliżony do rozkładu o zerowej wartości oczekiwanej i odchyleniu standardowym równym 1” [Młodak 2009].

W kolejnym etapie (III) ustalane są współrzędne obiektów modelowych. Najczęściej definiuje się je według następujących zasad – wartości wzorca jako [Wysocki 2010]:

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

oraz antywzorca rozwoju jako:

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

W rzeczywistych zbiorach danych mogą jednak wystąpić nietypowe wartości cech – odstające lub ekstremalne, co wynikać może ze specyfiki badanego zjawiska. Obserwacje te mogą mieć istotny wpływ na wyniki analizy (np. klasyfikacji typologicznej), dlatego wymagają szczególnej uwagi [Trzęsiok 2014]. W takich przypadkach przyjęcie, że wartości maksymalne i minimalne cech w metodach wzorcowych porządkowania liniowego obiektów (np. TOPSIS) stanowią obiekty modelowe, prowadzi do nadmiernie dużego ich oddalenia od wartości typowych rozważanych cech i w konsekwencji do zmniejszenia zakresu zmienności konstruowanego miernika rozwoju oraz trudności przeprowadzenia na jego podstawie prawidłowej identyfikacji typów rozwojowych badanych obiektów [Łuczak, Wysocki 2013].

Identyfikacja obserwacji odstających odbywać może się z wykorzystaniem różnorodnych metod jedno- i wielowymiarowych [Trzęsiok 2014]. Ze względu na to, że wartości modelowe w metodach wzorcowych porządkowania liniowego obiektów ustala się osobno dla każdej z cech, w proponowanym podejściu do identyfikacji wartości odstających zastosowano podejście jednowymiarowe – kryterium kwartylowe². Wartości pojedynczej cechy uznaje się za odstające, jeżeli znajdują się poza przedziałem [Trzęsiok 2014]:

$$[Q_{1k} - 1,5 \cdot IQR_k, Q_{3k} + 1,5 \cdot IQR_k], \quad (6)$$

gdzie: Q_{1k} , Q_{3k} – odpowiednio pierwszy i trzeci kwartył z wartości k -tej cechy; IQR_k – odchylenie ćwiartkowe z wartości k -tej cechy.

² Jest ono wykorzystywane do budowy wykresów *box-plot* wprowadzonych przez Tukeya [Trzęsiok 2014].

Na podstawie przyjętego kryterium kwartylowego (6) współrzędną wzorca rozwoju (A_k^+) dla k -tej cechy (stymulanty) określa się jako:

$$A_k^+ = \begin{cases} \max_{i=1, \dots, N}(z_{ik}), & \text{gdy } z_{ik} \in [Q_{1k} - 1,5 \cdot IQR_k, Q_{3k} + 1,5 \cdot IQR_k] \text{ dla } i \in [1, \dots, N], \\ Q_{3k} + 1,5 \cdot IQR_k, & \text{gdy } \max_{i=1, \dots, N}(z_{ik}) > Q_{3k} + 1,5 \cdot IQR_k, \end{cases} \quad (7)$$

a antywzorca rozwoju (A_k^-) jako:

$$A_k^- = \begin{cases} \min_{i=1, \dots, N}(z_{ik}), & \text{gdy } z_{ik} \in [Q_{1k} - 1,5 \cdot IQR_k, Q_{3k} + 1,5 \cdot IQR_k] \text{ dla } i \in [1, \dots, N], \\ Q_{1k} - 1,5 \cdot IQR_k, & \text{gdy } \min_{i=1, \dots, N}(z_{ik}) < Q_{1k} - 1,5 \cdot IQR_k. \end{cases} \quad (8)$$

Wszystkim obserwacjom odstającym k -tej cechy znajdującym się w przedziałach $[Q_{3k} + 1,5 \cdot IQR_k, \max_{i=1, \dots, N}(z_{ik})]$ oraz $[\min_{i=1, \dots, N}(z_{ik}), Q_{1k} - 1,5 \cdot IQR_k]$ przypisuje się więc odpowiednio współrzędną wzorca równą $A_k^+ = Q_{3k} + 1,5 \cdot IQR_k$, a antywzorca równą $A_k^- = Q_{1k} - 1,5 \cdot IQR_k$.

Współrzędne obiektów modelowych stanowią podstawę do obliczenia odlegania każdego ocenianego obiektu od wzorca (A^+) i antywzorca rozwoju (A^-) (etap IV), tj. w podejściu klasycznym i kwartylowym za pomocą odległości euklidesowych³ [Wysocki 2010]:

$$d_i^+ = \sqrt{\sum_{k=1}^K (z_{ik} - z_k^+)^2}, \quad d_i^- = \sqrt{\sum_{k=1}^K (z_{ik} - z_k^-)^2}, \quad (9)$$

natomiast w podejściu pozycyjnym za pomocą medianowego odchylenia bezwzględnego:

$$d_i^+ = \text{med}_k(|z_{ik} - z_k^+|), \quad d_i^- = \text{med}_k(|z_{ik} - z_k^-|), \quad (i = 1, \dots, N), \quad (10)$$

gdzie: med_k – mediana brzegowa dla k -tej cechy.

Do konstrukcji miernika syntetycznego (etap V) zastosowano metodę TOPSIS [Hwang, Yoon 1981]:

$$S_i = \frac{d_i^-}{d_i^- + d_i^+}, \quad (i = 1, \dots, N), \quad (11)$$

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

Wyznaczone wartości miernika syntetycznego stanowią podstawę do uporządkowania liniowego badanych obiektów według jego wartości nierosnących.

³ Jeżeli wartość k -tej cechy (stymulanty) dla i -tego obiektu zidentyfikowano jako odstającą, wówczas odległość tego obiektu od wzorca albo antywzorca rozwoju wynosi 0.

W etapie VI na ich podstawie wyodrębnić można klasy typologiczne obiektów w sposób arbitralny lub oparty na średniej i odchyleniu standardowym z wartości miernika syntetycznego [Wysocki 2010] (tab. 1).

Tabela 1. Sposoby identyfikacji typów rozwojowych

Arbitralny	Statystyczny*
$\langle 0,80, 1,00 \rangle$ – poziom bardzo wysoki (klasa I)	$S_i \geq \bar{S} + s_s$ – poziom wysoki (klasa I)
$\langle 0,60, 0,80 \rangle$ – wysoki (klasa II)	$\bar{S} \leq S_i < \bar{S} + s_s$ – średni wyższy (klasa II)
$\langle 0,40, 0,60 \rangle$ – przeciętny (klasa III)	$\bar{S} - s_s \leq S_i < \bar{S}$ – średni niższy (klasa III)
$\langle 0,20, 0,40 \rangle$ – niski (klasa IV)	$S_i < \bar{S} - s_s$ – niski (klasa IV)
$\langle 0,00, 0,20 \rangle$ – bardzo niski (klasa V)	

*Oparty na średniej arytmetycznej (\bar{S}) i odchyleniu standardowym (s_s) z wartości miernika syntetycznego.

Źródło: opracowanie własne na podstawie [Wysocki 2010].

3. Wyniki badań

W pierwszym etapie oceny poziomu samodzielności finansowej dokonano wyboru sześciu cech opisujących samorządy gminne, tj.: wskaźnik samodzielności finansowej (wydatkowej) pierwszego stopnia (%) (*WSFW I*), wskaźnik poziomu dochodów własnych w złotych *per capita* (*WDWM*), wskaźnik autonomii podatkowej (%) (*WAP*), wskaźnik finansowej ingerencji państwa (%) (*WFIP*), wskaźnik samofinansowania (%) (*WS*) oraz wskaźnik udziału wydatków inwestycyjnych w wydatkach ogółem (%) (*WIWO*). Jedną z nich, tj. *WFIP* uznano za destymulantę⁴, natomiast pozostałe za stymulanty samodzielności finansowej samorządów gminnych.

Jak wynika z danych przedstawionych w tab. 2, cechy związane z dochodami własnymi w złotych *per capita* (*WDWM*), a także ze zdolnością samorządów do samofinansowania (*WS*) charakteryzowały się relatywnie wysoką asymetrią prawostronną. Współczynnik skośności dla cechy *WDWM* wynosił 2,9, a dla *WS* – 3,7. Cechy te wykazywały również dużą zmienność ich wartości (przekraczającą 50%). Wpływ na to miały wartości odstające, zaobserwowane m.in. dla gmin Przykona oraz Suchy Las. W 2013 r. w gminach tych dochody własne w złotych *per capita* przekraczały ponad trzykrotnie średnią dla ogółu gmin wiejskich województwa wielkopolskiego, która wyniosła 1329,0 złotych. Także w przypadku udziału wydatków inwestycyjnych w wydatkach ogółem samorządów gminnych (*WIWO*) należy zauważyć wysoki stopień rozproszenia wartości tej cechy (współczynnik zmienności klasyczny wyniósł blisko 50%).

⁴ Cechę uznaną za destymulantę przekształcono w stymulantę za pomocą formuły (1).

Tabela 2. Statystyki opisowe cech charakteryzujących poziom samodzielności finansowej gmin wiejskich województwa wielkopolskiego w 2013 roku (statystyki klasyczne)

Charakterystyki	<i>WSFW</i>	<i>WDWM</i>	<i>WAP</i>	<i>WFIP</i>	<i>WS</i>	<i>WIWO</i>
Min	19,5	616,7	5,5	19,9	29,7	2,1
Max	80,1	4662,8	57,1	78,6	773,1	34,7
Średnia	40,2	1329,0	18,5	56,1	139,9	14,9
Odchylenie standardowe	13,4	688,4	8,3	12,3	89,3	7,1
Współczynnik zmienności (%)	33,3	51,8	44,7	22,0	63,8	47,8
Skośność	0,9	2,9	1,5	-0,9	3,7	0,6
Kurtoza	0,6	10,7	3,8	0,8	21,8	-0,1

Źródło: obliczenia własne na podstawie danych pochodzących z Banku Danych Lokalnych [GUS 2013] i Ministerstwa Finansów [2013].

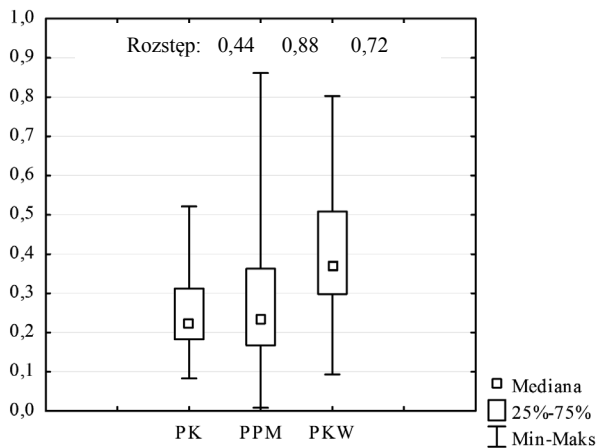
W kolejnym etapie (II) wartości cech poddano normalizacji wykorzystując formułę (2) w przypadku podejścia klasycznego i kwartyłowego oraz (3) w przypadku podejścia pozycyjnego. Następnie w podejściu klasycznym i pozycyjnym współrzędne wzorca rozwoju ustalono przyjmując wartości maksymalne cech, a antywzorca – ich wartości minimalne w zbiorze wszystkich gmin wiejskich województwa wielkopolskiego ($N = 117$). W proponowanym podejściu – kwartyłowym, wartości modelowe określono według formuł (7) i (8) (tab. 3). W przypadku cechy *WSFW* nie dokonano korekty wartości modelowych, bowiem w zbiorze wartości tej cechy nie zidentyfikowano wartości odstających.

Tabela 3. Wartości współrzędnych obiektów modelowych w konstrukcji syntetycznego miernika poziomu samodzielności finansowej gmin wiejskich województwa wielkopolskiego w 2013 roku

Charakterystyki	<i>WSFW</i>	<i>WDWM</i>	<i>WAP</i>	<i>WFIP</i>	<i>WS</i>	<i>WWIWO</i>
Statystyki opisowe						
Min	19,5	616,7	5,5	-78,6	29,7	2,1
Max	80,1	4662,8	57,1	-19,9	773,1	34,7
Mediana	37,5	1176,9	17,2	58,7	116,6	13,3
Wartości odstające (kryterium kwartyłowe)						
Wartości odstające (dolna granica)	×	×	×	×	×	×
Wartości odstające (górna granica)	×	> 1971,2	> 28,1	×	> 217,4	> 26,8
Liczba obiektów odstających	0	4	5	0	3	1
Wartości modelowe – podejście klasyczne						
Antywzorzec	19,5	616,7	5,5	-78,6	29,7	2,1
Wzorzec	80,1	4662,8	57,1	-19,9	773,1	34,7
Wartości modelowe – podejście kwartyłowe						
Antywzorzec	19,5	616,7	5,5	-78,6	29,7	2,1
Wzorzec	80,1	1971,2	28,1	-19,9	217,4	26,8

Źródło: jak w tab. 2.

Na podstawie ustalonych wartości modelowych obliczono odległości każdego ocenianego obiektu od wzorca i antywzorca rozwoju (etap IV), następnie wyznaczono wartości syntetycznego miernika rozwoju metodą TOPSIS według trzech rozważanych podejść (etap V). Zakresy zmienności uzyskanych wartości mierników syntetycznych przedstawiono na rys. 1. Można zauważyć, że zastosowanie podejścia kwartylowego, jak również pozycyjnego jest tutaj uzasadnione, gdyż w zbiorze wartości cech wystąpiły cechy charakteryzujące się relatywnie silną asymetrią, wysoką zmiennością wartości cech i obserwacjami odstającymi. Można zauważyć także, że w wyniku zastosowania metody TOPSIS w ujęciu kwartylowym, jak również pozycyjnym uzyskuje się znacznie większą rozpiętość wartości cechy syntetycznej w stosunku do metody klasycznej. Rozstęp pomiędzy maksymalną a minimalną wartością cechy syntetycznej w podejściu kwartylowym wyniósł 0,72, pozycyjnym 0,88, podczas gdy w podejściu klasycznym tylko 0,44.



* Podejścia: PK – klasyczne, PPM – pozycyjne oparte na medianie Webera, PKW – kwartylowe.

Rys. 1. Zakresy zmienności wartości mierników syntetycznych poziomu samodzielności finansowej gmin wiejskich województwa wielkopolskiego obliczonych z wykorzystaniem metody TOPSIS i różnych podejść* do ustalania wartości modelowych

Źródło: jak w tab. 2.

Na podstawie uzyskanych wartości miernika syntetycznego uporządkowano badane gminy w sposób nierosnący i dokonano oceny zgodności tych uporządkowań. Przyjmując za podstawę rangi gmin obliczono wartości współczynnika korelacji tau Kendalla. Najniższa zgodność uporządkowań badanych gmin wystąpiła pomiędzy podejściem klasycznym i pozycyjnym (współczynnik korelacji tau Kendalla = 0,75), natomiast relatywnie wyższa pomiędzy pozycyjnym i kwartylowym (0,81).

Celem budowy miernika syntetycznego jest nie tylko analiza szczegółowa rankingu badanych obiektów, szczególnie gdy badania obejmują dużą liczbę obiektów ($N > 100$), ale także analiza wewnętrznej struktury badanego zjawiska złożonego na podstawie przeprowadzonej identyfikacji typów rozwojowych. Wielu badaczy wyodrębnia klasy typologiczne (zazwyczaj cztery) w sposób oparty na statystykach opisowych – na podstawie średniej i odchylenia standardowego z wartości miernika syntetycznego. Uzyskana wówczas typologia może być nieuzasadniona w szczególności wtedy, gdy wartości miernika syntetycznego wyróżniają się niską zmiennością i małym rozstępem. W efekcie wysoki poziom rozwoju można nawet przypisać obiektom, które charakteryzuje niska, bądź bardzo niska wartość miernika syntetycznego. Jak wynika z analizy danych przedstawionych w tab. 4 taka sytuacja wystąpiła w przypadku klasyfikacji opartej na metodzie klasycznej, w której gminy wiejskie o wartości miernika z przedziału $[0,29, 0,52]$ zaklasyfikowano do klasy gmin o wysokim poziomie samodzielności finansowej, przy potencjalnym zakresie zmienności wartości miernika z przedziału $[0, 1]$.

Tabela 4. Klasyfikacja typologiczna gmin wiejskich województwa wielkopolskiego według poziomu samodzielności finansowej (identyfikacja typów oparta na statystykach opisowych)

Klasa typologiczna (poziom samodzielności finansowej)	Metoda TOPSIS – podejście								
	klasyczne			pozycyjne			kwartyłowe		
	S_i	N_c^*	%	S_i	N_c	%	S_i	N_c	%
I (wysoki)	$\langle 0,29, 0,52 \rangle$	19	16,2	$\langle 0,43, 0,89 \rangle$	10	8,5	$\langle 0,54, 0,81 \rangle$	23	19,7
II (średni wyższy)	$\langle 0,25, 0,29 \rangle$	29	24,8	$\langle 0,27, 0,43 \rangle$	34	29,1	$\langle 0,40, 0,54 \rangle$	25	21,4
III (średni niższy)	$\langle 0,15, 0,25 \rangle$	54	46,2	$\langle 0,11, 0,27 \rangle$	62	53,0	$\langle 0,26, 0,40 \rangle$	52	44,4
IV (niski)	$\langle 0,08, 0,15 \rangle$	15	12,8	$\langle 0,01, 0,11 \rangle$	11	9,4	$\langle 0,09, 0,26 \rangle$	17	14,5

* N_c – liczebność c -tej klasy ($c = 1, \dots, C$).

Źródło: jak w tab. 2.

Klasyfikację typologiczną badanych gmin wiejskich przeprowadzono również w sposób arbitralny. W rezultacie, na podstawie wartości miernika syntetycznego uzyskanego metodą klasyczną wyodrębniono tylko trzy klasy typologiczne (III, IV, V), w tym większość gmin (prawie 61%) zakwalifikowana została do klasy IV, reprezentującej niski poziom samodzielności finansowej (tab. 5). Nie wyodrębniono natomiast klas gmin charakteryzujących się bardzo wysokim i wysokim poziomem samodzielności finansowej, pomimo, że z przeprowadzonych analiz wynika, że gminy zlokalizowane w pierwszym pierścieniu wokół miasta Poznań, takie jak Suchy Las, Komorniki czy Tarnowo Podgórne, na terenie których rozwija się silnie funkcja rezydencyjna i usługowa (gminy te zatraciły typowy dla obszarów wiejskich charakter rolniczy), powinny reprezentować wysoki bądź bardzo wysoki poziom samodzielności finansowej.

Tabela 5. Klasyfikacja typologiczna gmin wiejskich województwa wielkopolskiego według poziomu samodzielności finansowej (identyfikacja typów w sposób arbitralny)

Wyszczególnienie		Metoda TOPSIS – podejście					
Klasa typologiczna (poziom samodzielności finansowej)	S_i	klasyczne		pozycyjne		kwartylowe	
		N_c	%	N_c	%	N_c	%
I (bardzo wysoki)	(0,80, 1,00)	0	0,0	1	0,9	2	1,7
II (wysoki)	(0,60, 0,80)	0	0,0	4	3,4	7	6,0
III (przeciętny)	(0,40, 0,60)	6	5,1	14	12,0	38	32,5
IV (niski)	(0,20, 0,40)	71	60,7	57	48,7	62	53,0
V (bardzo niski)	(0,00, 0,20)	40	34,2	41	35,0	8	6,8

Źródło: jak w tab. 2.

Na podstawie wartości miernika syntetycznego uzyskanego w podejściu kwartylowym uzyskano znacznie większy zakres zmienności wartości miernika syntetycznego, a zatem było możliwe wyodrębnienie pięciu klas typologicznych, tj. wyróżnienie gmin zarówno o bardzo wysokiej, wysokiej, przeciętnej, niskiej i bardzo niskiej samodzielności finansowej (tab. 5). Wyodrębnione na tej podstawie klasy typologiczne znacznie lepiej z merytorycznego, ale także statystycznego punktu widzenia odzwierciedliły różnice międzyklasowe w poziomie samodzielności finansowej gmin wiejskich w województwie wielkopolskim.

Tabela 6. Wartości wskaźników homogeniczności, heterogeniczności i poprawności skonstruowanych typów*

Wyszczególnienie	Metoda TOPSIS – podejście		
	klasyczne	pozycyjne	kwartylowe
Homogeniczność	544,1	644,0	455,0
Heterogeniczność	5,1	8,5	7,1
Poprawność	106,7	75,8	64,1

*Oparte na koncepcji środka ciężkości klasy i odległości od niej (środek ciężkości danej klasy stanowi mediana Webera jej elementów) [Młodak 2006].

Źródło: jak w tab. 2.

Potwierdza to analiza wartości wskaźników dotyczących jakości klasyfikacji typologicznych⁵ (tab. 6). Można stwierdzić, że klasyfikacja oparta na wartościach miernika syntetycznego uzyskanego za pomocą podejścia kwartylowego wyróżnia

⁵ Wskaźnik homogeniczności mierzy zwartość wyodrębnionej klasy (im niższa wartość wskaźnika tym klasa jest bardziej jednorodna), natomiast wskaźnik heterogeniczności bada jej odrębność od innych klas (im wyższa jego wartość, tym różnice międzyklasowe są większe). Im niższe są wartości wskaźnika poprawności typów (uzyskane poprzez podzielenie wartości wskaźnika homogeniczności przez wartości wskaźnika heterogeniczności), tym uzyskana typologia jest bardziej poprawna pod względem statystycznym.

się najwyższą poprawnością skupień na tle pozostałych. Najniższą jakością klasyfikacji charakteryzuje się typologia uzyskana na podstawie podejścia klasycznego, na podstawie którego uzyskano tylko trzy klasy typologiczne.

4. Zakończenie

W przypadku gdy w zbiorze wartości cech występują obserwacje odstające, zastosowanie klasycznych metod wzorcowych prowadzić może do nadmiernego zmniejszenia zakresu zmienności wartości budowanego syntetycznego miernika rozwoju, a w efekcie do problemów związanych z pełną i prawidłową identyfikacją typów rozwojowych badanego zjawiska złożonego. Problemy te wynikają stąd, że w przypadku zastosowania sposobu identyfikacji opartego na statystykach opisowych, obiektom o względnie niskich wartościach miernika syntetycznego (np. $< 0,5$) może nawet zostać przyporządkowany wysoki poziom rozwoju rozpatrywanego zjawiska. Z kolei w przypadku przyjęcia sposobu arbitralnego identyfikacji typów mogą pojawić się klasy puste.

Przeprowadzone badania empiryczne wykazały, że zastosowanie podejścia kwartyłowego umożliwia, w porównaniu z podejściem klasycznym, poprawniejsze odzwierciedlenie różnic międzyklasowych w poziomie rozwoju analizowanych zjawisk (w sensie merytorycznym i statystycznym), co pociąga za sobą pełniejszą identyfikację typów rozwojowych.

Zastosowanie klasycznej metody TOPSIS do oceny poziomu samodzielności finansowej gmin wiejskich województwa wielkopolskiego doprowadziło do wyodrębnienia tylko trzech klas typologicznych (z pominięciem klasy reprezentującej bardzo wysoki i wysoki poziom rozwoju), bądź przypisania bardzo wysokiego poziomu samodzielności finansowej gminom o stosunkowo niskich wartościach syntetycznego miernika rozwoju ($< 0,5$). Natomiast w proponowanym podejściu kwartyłowym (jak również w podejściu pozycyjnym) uzyskano poprawniejszą identyfikację typów rozwojowych samodzielności finansowej gmin wiejskich, o czym świadczy wyodrębnienie także klas o bardzo wysokim i wysokim poziomie samodzielności finansowej. Na przykład gminy Suchy Las i Tarnowo Podgórne w podejściu kwartyłowym zostały zgodnie z oczekiwaniami, zakwalifikowane do typu rozwojowego o bardzo wysokim poziomie samodzielności finansowej, osiągając wartości miernika syntetycznego powyżej 0,8, podczas gdy w podejściu klasycznym wartości tego miernika wynosiły co najwyżej 0,5.

Literatura

- GUS, 2013, *Bank Danych Lokalnych*, www.stat.gov.pl (12.03.2015).
Hwang C.L., Yoon K., 1981, *Multiple attribute decision-making: Methods and applications*, Springer, Berlin.

- Lira J., Wagner W., Wysocki F., 2002, *Mediana w zagadnieniach porządkowania obiektów wielocechowych*, [w:] J. Paradysz (red.), *Statystyka regionalna w służbie samorządu terytorialnego i biznesu*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, s. 87–99.
- Łuczak A., Wysocki F., 2013, *Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 278, *Taksonomia 20: Klasyfikacja i analiza danych – teoria i zastosowanie*, s. 63–73.
- Majka A., 2015, *Przestrzenne zróżnicowanie poziomu życia ludności w Polsce w ujęciu dynamicznym*, *Wiadomości Statystyczne*, nr 5, Główny Urząd Statystyczny, Warszawa.
- Ministerstwo Finansów, 2013, *Wskaźniki do oceny sytuacji finansowej jednostek samorządu terytorialnego*, www.finanse.mf.gov.pl (12.03.2015).
- Młodak A., 2006, *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa.
- Młodak A., 2009, *Historia problemu Webera*, *Matematyka Stosowana*, t. 10/51: *Matematyka dla społeczeństwa*, s. 3–21.
- Trzęsiok M., 2014, *Wybrane metody identyfikacji obserwacji oddalonych*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 327, *Taksonomia 22: Klasyfikacja i analiza danych – teoria i zastosowanie*, s. 157–166.
- Wysocki F., 2010, *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań.