

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 426

Taksonomia 26

**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
Jacek Batóg: Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis	13
Andrzej Bąk: Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
Grażyna Dehnel: <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
Andrzej Dudek: <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
Iwona Foryś: Analiza dyskryminacyjna w wyborze obiektów podobnych w procesie szacowania nieruchomości / The discriminant analysis in selection of similar objects in the real estate valuation process	51
Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz: Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
Iwona Konarzewska: Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria	69
Anna Król, Marta Targaszewska: Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
Marek Lubicz: Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
Aleksandra Łuczak: Wpływ różnych sposobów agregacji opinii ekspertów w FAHP na oceny priorytetowych czynników rozwoju / Influence of different methods of the expert judgments aggregation on assessment of priorities for evaluation of development factors in FAHP.....	99
Iwona Markowicz: Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity	108

Małgorzata Markowska, Danuta Strahl: Filary inteligentnego rozwoju a wrażliwość unijnych regionów szczebla NUTS 2 na kryzys ekonomiczny – analiza wielowymiarowa / Smart development pillars and NUTS 2 European regions vulnerability to economic crisis – a multidimensional analysis.....	118
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis	130
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis	139
Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta: Problem wartości odstających w prognozowaniu zagrożenia upadłością przedsiębiorstw (na przykładzie przetwórstwa przemysłowego w Polsce) / Problem of outliers in corporate bankruptcy prediction (case of manufacturing companies in Poland)	148
Wojciech Roszka: Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
Małgorzata Rószkiewicz: Czynniki różnicujące efektywność pracy ankietera w wywiadach <i>face-to-face</i> w środowisku polskich gospodarstw domowych / Factors affecting the efficiency of face-to-face interviews with Polish households.....	166
Adam Sagan, Marcin Pelka: Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data	174
Marcin Salamaga: Zastosowanie drzew dyskryminacyjnych w identyfikacji czynników wspomagających wybór kraju alokacji bezpośrednich inwestycji zagranicznych na przykładzie polskich firm / The use of classification trees in the identification of factors supporting the choice of FDI destination on the example of Polish companies.....	185
Agnieszka Stanimir: Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
Mirosława Sztemberg-Lewandowska: Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge	206
Tadeusz Trzaskalik: Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature	214

Joanna Trzęsiok: Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions	226
Hanna Wdowicka: Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
Artur Zaborski: Zastosowanie skalowania dynamicznego oraz metody wektorów dryfu do badania zmian w preferencjach / The use of dynamic scaling and the drift vector method for studying changes in the preferences.....	245

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. Przewodniczącymi Komitetu Organizacyjnego konferencji byli prof. dr hab. Mirosław Szreder oraz dr hab. Krzysztof Najman, prof. nadzw. UG, sekretarzami naukowymi dr hab. Kamila Migdał-Najman, prof. nadzw. UG oraz dr hab. Anna Zamojska, prof. nadzw. UG, a sekretarzem organizacyjnym Anna Nowicka z Fundacji Rozwoju Uniwersytetu Gdańskiego.

Konferencja Naukowa została dofinansowana ze środków Narodowego Banku Polskiego.

Zakres tematyczny konferencji obejmował takie zagadnienia, jak:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia, itd., aplikacje komputerowe metod statystycznych).

Zasadniczymi celami konferencji SKAD były prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 81 osób. Byli to pracownicy oraz doktoranci następujących uczelni i instytucji: AGH w Krakowie, Politechniki Łódzkiej, Politechniki Gdańskiej, Politechniki Opolskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Jana Kochanowskiego w Kielcach, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwer-

sytetu w Białymstoku, Wyższej Szkoły Bankowej w Toruniu, a także przedstawiciele NBP i PBS Sp. z o.o.

W trakcie dwóch sesji plenarnych oraz trzynastu 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. Obradom w poszczególnych sesjach konferencji przewodniczyli profesorowie: Józef Pocięcha, Eugeniusz Gatnar, Tadeusz Trzaskalik, Krzysztof Jajuga, Marek Walesiak, Barbara Pawełek, Feliks Wysocki, Ewa Roszkowska, Andrzej Sokołowski, Andrzej Bąk, Tadeusz Kufel, Mirosław Krzyśko, Krzysztof Najman, Małgorzata Rószkiewicz, Mirosław Szreder.

Teksty 25 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii „Taksonomia” nr 26. Pozostałe recenzowane artykuły znajdują się w „Taksonomii” nr 27.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pocięcha. Ustalono plan przebiegu zebrania obejmujący następujące punkty:

- A. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
- B. Informacje dotyczące planowanych konferencji krajowych i zagranicznych.
- C. Organizacja konferencji SKAD PTS w latach 2016 i 2017.
- D. Wybór przedstawiciela Rady Sekcji SKAD PTS do IFCS.
- E. Dyskusja nad kierunkami rozwoju działalności Sekcji.

Prof. dr hab. Józef Pocięcha otworzył posiedzenie Sekcji SKAD PTS. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS przedstawiła sekretarz naukowy Sekcji dr hab. Barbara Pawełek, prof. nadzw. UEK. Poinformowała, że obecnie Sekcja liczy 231 członków. Przypomniała, że na stronie internetowej Sekcji znajdują się regulamin, a także deklaracja członkowska. Poinformowała, że zostały opublikowane zeszyty z serii „Taksonomia” nr 24 i 25 (PN UE we Wrocławiu nr 384 i 385). W „Przeglądzie Statystycznym” (zeszyt 4/2014) ukazało się sprawozdanie z ubiegłorocznej konferencji SKAD, która odbyła się w Międzyzdrojach, w dniach 8–10 września 2014 r. Prof. Barbara Pawełek przedstawiła także informacje dotyczące działalności międzynarodowej oraz udziału w ważnych konferencjach członków i sympatyków SKAD.

W konferencji Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS – International Federation of Classification Societies) w dniach 6–8 lipca 2015 r. w Bolonii, zorganizowanej przez Università di Bologna, udział wzięło 19 osób z Polski (w tym 17 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 79,0%). Ponadto prof. Józef Pocięcha był członkiem Komitetu Naukowego Konferencji z ramienia SKAD, członkiem Międzynarodowego Komitetu Nagród IFCS oraz organizatorem i przewodniczącym sesji nt. „Classification models for forecasting of economic processes”.

W konferencji „European Conference on Data Analysis” (Colchester, 2–4 września 2015 r.) zorganizowanej przez The German Classification Society (GfKI) we współpracy z The British Classification Society (BCS) i Sekcją Klasyfikacji i Analizy Danych PTS (SKAD) udział wzięło 18 osób z Polski (w tym 14 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 66,0%). Ponadto profesorowie Krzysztof Jajuga oraz Józef Pociecha byli członkami Komitetu Naukowego konferencji, prof. Andrzej Dudek został poproszony przez organizatorów o przygotowanie referatu i wygłoszenie na Sesji Plenarnej „Cluster analysis in XXI century, new methods and tendencies”, prof. Krzysztof Jajuga był przewodniczącym sesji plenarnej, przewodniczącym sesji nt. „Finance and economics II” oraz organizatorem i przewodniczącym sesji nt. „Data analysis in finance”, prof. Józef Pociecha był organizatorem i przewodniczącym sesji nt. „Outliers in classification procedures – theory and practice”, prof. Andrzej Dudek był przewodniczącym sesji nt. „Machine learning and knowledge discovery II”.

Kolejny punkt posiedzenia Sekcji obejmował zapowiedzi najbliższych konferencji krajowych i zagranicznych, których tematyka jest zgodna z profilem Sekcji. Prof. dr hab. Józef Pociecha poinformował o dwóch wybranych konferencjach krajowych (były to XXXIV Konferencja Naukowa „Multivariate Statistical Analysis MSA 2015”, Łódź, 16–18 listopada 2015 r. i X Międzynarodowa Konferencja Naukowa im. Profesora Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych”, Zakopane, 10–13 maja 2016 r.) oraz o trzech wybranych konferencjach zagranicznych. Konferencja „European Conference on Data Analysis” odbędzie się na Uniwersytecie Ekonomicznym we Wrocławiu w dniach 26–28 września 2017 r. W przeddzień tej konferencji, tj. 25.09.2017 r., odbędzie się Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2017”. Następną konferencją Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS) odbędzie się w 2017 r. w Tokio. W 2019 r. Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2019” organizuje prof. Andreas Geyer-Schultz w Karlsruhe.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. SKAD 2016 zorganizuje Katedra Metod Statystycznych Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego.

W kolejnej części zebrania dokonano wyboru przedstawiciela Rady Sekcji SKAD PTS do IFCS na kadencję 2016–2019. Powołano Komisję Skrutacyjną, której przewodniczącym został prof. Tadeusz Kufel, a członkami dr hab. Iwona Konarzewska i dr Dominik Rozkrut. Profesor Józef Pociecha poprosił zebranych o proponowanie kandydatur zgłaszając jednocześnie prof. Andrzeja Sokołowskiego. Wobec braku następnych kandydatur listę zamknięto. Komisja Skrutacyjna przeprowadziła głosowanie tajne. W głosowaniu uczestniczyło 41 członków Sekcji. Profesor Andrzej Sokołowski został przedstawicielem Rady Sekcji SKAD PTS do

IFCS na kadencję 2016–2019, uzyskując następujący wynik: 39 głosów na „tak”, 1 głos na „nie”, 1 głos był nieważny.

W ostatnim punkcie zebrania dyskutowano nad kierunkami rozwoju działalności Sekcji obejmującymi następujące problemy: udział w międzynarodowym ruchu naukowym (wspólne granty, publikacje), umiędzynarodowienie konferencji SKAD (uczestnicy zagraniczni, dwujęzyczność konferencji), wydawanie własnego czasopisma.

Profesor Józef Pociecha zamknął posiedzenie Sekcji SKAD.

Krzysztof Jajuga, Marek Walesiak

Andrzej Bąk

Uniwersytet Ekonomiczny we Wrocławiu
e-mail: andrzej.bak@ue.wroc.pl

**PORZĄDKOWANIE LINIOWE OBIEKTÓW
METODĄ HELLWIGA I TOPSIS –
ANALIZA PORÓWNAWCZA**

**LINEAR ORDERING OF OBJECTS USING HELLWIG
AND TOPSIS METHODS –
A COMPARATIVE ANALYSIS**

DOI: 10.15611/pn.2016.426.02

Streszczenie: Przedmiotem porządkowania liniowego mogą być takie obiekty, jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe). Takie charakterystyki, jak poziom rozwoju gospodarczego, kondycja finansowa, walory użytkowe są zmiennymi, których realizacje nie są bezpośrednio mierzalne. Zmienne te są agregatami, których wartości są generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne. Uzyskane realizacje zmiennej syntetycznej umożliwiają uporządkowanie obiektów wielowymiarowych w sensie relacji preferencji (dominacji). W obszarze ekonomii pierwsza metoda porządkowania liniowego została przedstawiona przez Z. Hellwiga w 1968 r., natomiast na gruncie teorii decyzji przez C.L. Hwanga i K. Yoona w 1981 r. Celem artykułu jest porównanie wyników porządkowania liniowego (rankingów) otrzymywanych za pomocą tych metod oraz prezentacja procedur obliczeniowych w języku R umożliwiających prowadzenie analiz porównawczych.

Słowa kluczowe: porządkowanie liniowe, metoda Hellwiga, metoda TOPSIS, program R.

Summary: The subject of linear ordering can be objects such as countries (due to the level of economic development), business (due to financial condition), products (due to usability). Such characteristics as the level of economic development, financial condition, usability are variables whose realizations are not directly measurable. These variables are aggregates whose values are generated by observations of diagnostic variables that are directly measurable. The estimated realisations of synthetic variable allow for ordering multi-dimensional objects in terms of preference relations (domination). In the area of economics first linear ordering method was presented by Z. Hellwig in 1968, while that on the decisions theory by C.L. Hwang and Yoon K. in 1981. The purpose of this article is to compare the results of linear ordering (ranking) obtained by these methods and the presentation of computational procedures in R allow conducting comparative analyzes.

Keywords: linear ordering, Hellwig method, TOPSIS method, R program.

1. Wstęp

Metody porządkowania liniowego są wykorzystywane w badaniach ekonomicznych w celu ustalenia kolejności lub klasyfikacji obiektów, takich jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe) itp.

Idea porządkowania liniowego obiektów wielowymiarowych opiera się na pojęciu porządkującej relacji binarnej (zwrotnej, antysymetrycznej, przechodniej i spójnej). Z aksjomatów tej relacji wynika, że jest możliwe stwierdzenie, który z dwóch dowolnych obiektów zbioru jest pierwszy (lepszy), a który drugi (gorszy), a także, czy są one identyczne. Przedmiotem porządkowania liniowego mogą być np. takie obiekty, jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe) itp. Takie charakterystyki jak poziom rozwoju gospodarczego, kondycja finansowa, walory użytkowe są zmiennymi, których realizacje nie są bezpośrednio mierzalne. Zmienne takie są agregatami, których wartości są generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne (funkcje agregujące mogą mieć różną postać analityczną). Uzyskane realizacje zmiennej syntetycznej umożliwiają uporządkowanie obiektów wielowymiarowych w sensie relacji preferencji (dominacji).

Celem artykułu jest prezentacja dwóch historycznie pierwszych metod porządkowania liniowego opracowanych w dwóch różnych obszarach badawczych oraz porównanie wyników porządkowania obiektów za pomocą tych procedur.

Pierwsza metoda porządkowania liniowego została zaproponowana przez Z. Hellwiga (w latach 1967–1968) na gruncie ekonomii (taksonomii), natomiast druga przez C.L. Hwanga i K. Yoona (w latach 1980–1981) na gruncie teorii decyzji (wielokryterialnego podejmowania decyzji).

W obliczeniach porównawczych wykorzystano pakiety programu R – `pllord` [Bąk 2015] i `topsis` [Yazdi 2015]. Program R [R Development Core Team 2015] jest niekomercyjnym projektem w zakresie analizy danych powszechnie wykorzystywanym m.in. w statystycznych i ekonometrycznych badaniach ekonomicznych.

2. Taksonomia i teoria decyzji

W literaturze przedmiotu prezentowanych jest wiele definicji i interpretacji pojęcia „taksonomia”. Stosowane są także terminy bliskoznaczne, takie jak taksologia, taksonometria, taksonomia numeryczna, taksonomia matematyczna, klasyfikacja, analiza skupień, grupowanie, dyskryminacja, delimitacja, rozpoznawanie obrazów. Istnieje także rozróżnienie między taksonomią jakościową i ilościową oraz taksonomią opisową i stochastyczną [Pociecha 2008; Bąk 2013].

Rzeczywiście rozwój taksonomii ilościowej zainicjował na początku XX w. polski antropolog J. Czekanowski, proponując miarę odległości i diagraficzną metodę porządkowania macierzy odległości¹ [Pociecha i in. 1988, s. 13]. Według definicji podanej w pracy [Grabiński, Wydymus, Zeliaś 1989, s. 9] „przez taksonomię rozumie się dyscyplinę naukową zajmującą się zasadami i procedurami klasyfikacji (porządkowania, grupowania, dyskryminacji, delimitacji, podziału)”. Celem badań taksonomicznych może być podział zbioru elementów, porządkowanie elementów zbioru, wybór elementów ze zbioru [Pociecha i in. 1988, s. 17]. Zgodnie z tą definicją i celem badań taksonomicznych metody porządkowania liniowego zalicza się do metod taksonomicznych [Bąk 2013].

W węższym rozumieniu metody porządkowania liniowego należą do podstawowych metod wielowymiarowej analizy porównawczej. Wielowymiarowa analiza porównawcza jest dyscypliną naukową umożliwiającą analizę obiektów i zjawisk złożonych, tj. takich, na których stan i zachowanie wpływa jednocześnie wiele cech (zmiennych) i czynników. Zwięzła definicja podana przez Z. Hellwiga mówi, iż „metody i technika porównywania obiektów wielocechowych nazywają się wielowymiarową analizą porównawczą” [Hellwig 1981, s. 48]. Podstawowym celem wielowymiarowej analizy porównawczej jest konstrukcja miary syntetycznej umożliwiającej porównywanie obiektów opisanych za pomocą wielu zmiennych. Do osiągnięcia tego celu wykorzystuje się metody porządkowania liniowego [Bąk 2013].

Pierwsza propozycja metody porządkowania liniowego, w obszarze badań taksonomicznych, umożliwiającej porządkowanie liniowe obiektów z wykorzystaniem wzorca została przedstawiona przez Z. Hellwiga w 1968 r. pod nazwą „miara rozwoju gospodarczego” [Hellwig 1968]. Rok wcześniej propozycja ta została zaprezentowana w niepublikowanym raporcie pt. *Procedure of Evaluating high Level Manpower Data and Typology of Countries by Means of the Taxonomic Method*, UNESCO, 1967.

W literaturze przedmiotu teoria decyzji obejmuje dwa nurty: normatywny i opisowy [Tyszka 2010]. W podejściu normatywnym poszukuje się optymalnego i racjonalnego rozwiązania określonej sytuacji decyzyjnej – w warunkach pewności, ryzyka i niepewności. Ten nurt rozwinął się głównie na gruncie ekonomii (np. badania wyborów konsumenckich). Ważnym osiągnięciem w tym obszarze badań jest aksjomatyczna (nowoczesna) teoria użyteczności J. von Neumanna i O. Morgensterna [Neumann, Morgenstern 1944]. Racjonalność decyzji i zachowań ludzkich nie zawsze występuje w rzeczywistości. Badania realnych i często nieracjonalnych decyzji podejmowanych w sytuacjach wyboru prowadzone były na gruncie psychologii i psychometrii. Doprowadziły one do wyodrębnienia się opisowego (deskryptywnego) nurtu teorii decyzji, nazywanego także nurtem behawioralnym.

¹ Na stronie internetowej <http://www.antropologia.uw.edu.pl/MaCzek/maczek.html> dostępny jest program komputerowy MaCzek będący implementacją metody (diagramu) Czekanowskiego.

W obszarze badań teorii decyzji stosowane są metody badań operacyjnych, wielokryterialnej analizy decyzyjnej, analizy systemowej, programowania matematycznego. Rozwój teorii decyzji następuje w różnych obszarach badań i zastosowań, m.in. na gruncie ekonomii i zarządzania, matematyki, statystyki, socjologii, psychologii, informatyki.

Na gruncie teorii decyzji (wielokryterialnego podejmowania decyzji) pierwsza metoda porządkowania liniowego z wykorzystaniem wzorca i antywzorca została zaproponowana przez C.L. Hwanga i K. Yoon w 1981 r. pod nazwą TOPSIS – *Technique for Order Preference by Similarity to Ideal Solution* [Hwang, Yoon 1981]. Rok wcześniej propozycja ta została przedstawiona w pracach: K. Yoon, *System Selection by Multiple Attribute Decision Making* oraz K. Yoon i C.L. Hwang, *Technique for Order Preference by Similarity to Ideal Solution – A Multiattribute Decision Making*.

3. Metody porządkowania liniowego

Metody porządkowania liniowego, mieszczące się w obrębie wielowymiarowej analizy porównawczej i szerszej taksonomii, są w dużej mierze dorobkiem polskiej myśli statystycznej i ekonometrycznej. Pierwszą propozycję przedstawił Z. Hellwig w pracy [Hellwig 1968]. Publikacja ta zainicjowała intensywne badania w tym zakresie, których efektem były kolejne propozycje metod porządkowania liniowego zamieszczone m.in. w pracach [Bartosiewicz 1976; Borys 1978b; Cieślak 1974; Nowak 1984; Pluta 1976; Strahl 1978; Walesiak 1993].

Podstawą porządkowania liniowego jest zmienna syntetyczna², której wartości są szacowane na podstawie obserwacji zmiennych diagnostycznych opisujących badane obiekty. Zakłada się, że wartości zmiennej syntetycznej, oszacowane za pomocą określonej metody, umożliwiają takie uporządkowanie zbioru obiektów, w którym [Grabiński 1992, s. 135]: (1) każdy obiekt ma przynajmniej jednego sąsiada oraz nie więcej niż dwóch sąsiadów, (2) jeżeli obiekt a jest sąsiadem obiektu b , to obiekt b jest sąsiadem obiektu a , (3) istnieją tylko dwa obiekty mające jednego sąsiada.

Zmienna syntetyczna ma charakter zmiennej ukrytej, ponieważ jej realizacje nie są bezpośrednio obserwowane. Realizacje te są natomiast generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne. Realizacje zmiennej syntetycznej są szacowane za pomocą funkcji agregujących, których postaci analityczne mogą być różne. Rozróżnia się dwie podstawowe grupy metod, które są wykorzystywane do szacowania wartości zmiennej syntetycznej: metody bezwzorcowe i metody wzorcowe.

² W literaturze przedmiotu spotkać można inne określenia zmiennej syntetycznej, takie jak np.: zmienna agregatowa, miara syntetyczna, syntetyczna miara rozwoju, taksonomiczny miernik rozwoju, agregatowa miara rozwoju, miara rozwoju gospodarczego.

W procedurze porządkowania liniowego wyróżnia się takie etapy postępowania, jak: określenie charakteru zmiennych (stymulanty, nominanty, destymulanty)³, wyznaczenie wag zmiennych, normalizacja zmiennych, wyznaczenie współrzędnych wzorca w przypadku agregacji wzorcowej, agregacja bezwzorcowa lub wzorcowa, klasyfikacja porangowanych obiektów i rozpoznanie typów rozwojowych [Grabiński 1984; Bąk 1999, 2013; Wysocki 2010, s. 145].

Metody porządkowania liniowego Hellwiga i TOPSIS są metodami wzorcowymi, przy czym w metodzie Hellwiga punktem odniesienia obiektów w przestrzeni wielowymiarowej jest wzorzec, a w metodzie TOPSIS wyznacza się dwa punkty odniesienia – wzorzec i antywzorzec.

Konstrukcja miary syntetycznej Hellwiga jest następująca:

a) normalizacja zmiennych (standaryzacja): $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$, x_{ij} – obserwacja j -tej zmiennej dla obiektu i , \bar{x}_j – średnia arytmetyczna obserwacji j -tej zmiennej, s_j – odchylenie standardowe obserwacji j -tej zmiennej;

b) współrzędne wzorca: $z_{0j} = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych stymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych destymulant} \end{cases}$;

c) odległości obiektów od wzorca: $d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2}$;

d) wartości zmiennej agregatowej: $q_i = 1 - \frac{d_{i0}}{d_0}$, przy czym: na ogół $q_i \in [0; 1]$; $\max_i \{q_i\}$ – najlepszy obiekt; $\min_i \{q_i\}$ – najgorszy obiekt; $d_0 = \bar{d}_0 + 2s_d$; $\bar{d}_0 = \frac{1}{n} \sum_{i=1}^n d_{i0}$; $s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_{i0} - \bar{d}_0)^2}$.

Konstrukcja miary syntetycznej TOPSIS Hwanga i Yoona jest następująca:

a) normalizacja zmiennych (przekształcenie ilorazowe): $z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$;

x_{ij} – obserwacja j -tej zmiennej dla obiektu i ;

b) współrzędne wzorca:

1. $z_{0j}^+ = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych stymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych destymulant} \end{cases}$;

c) współrzędne antywzorca:

2. $z_{0j}^- = \begin{cases} \min_i \{z_{ij}\} & \text{dla zmiennych stymulant} \\ \max_i \{z_{ij}\} & \text{dla zmiennych destymulant} \end{cases}$;

d) odległości obiektów od wzorca: $d_{i0}^+ = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j}^+)^2}$;

³ Pojęcia zmiennej stymulandy i destymulandy zostały wprowadzone do literatury przedmiotu przez Z. Hellwiga [Hellwig 1968], a pojęcie zmiennej nominandy przez T. Borysa [Borys 1978a].

e) odległości obiektów od antywzorca: $d_{i0}^- = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j}^-)^2}$;

f) wartości zmiennej agregatowej: $q_i = \frac{d_{i0}^-}{d_{i0}^+ + d_{i0}^-}$, przy czym: $q_i \in [0; 1]$;
 $\max_i\{q_i\}$ – najlepszy obiekt; $\min_i\{q_i\}$ – najgorszy obiekt.

4. Wyniki badań porównawczych

W badaniach porównawczych wyników porządkowania liniowego metodami Hellwiga i TOPSIS wykorzystano zbiory danych empirycznych z oryginalnych prac: [Hellwig 1968] i [Hwang, Yoon 1981]. Zastosowano także metody porządkowania liniowego zaproponowane w tych publikacjach bez żadnych modyfikacji.

W przypadku danych z pracy [Hellwig 1968] (hdane68) porządkowanymi obiektami jest 15 krajów charakteryzowanych przez 6 zmiennych (X3-X6 w przeliczeniu na 10 000 osób): X1 – przeciętne trwanie życia mężczyzn, X2 – procent ludności zawodowo czynnej w rolnictwie, X3 – kadry inżynieryjno-techniczne, X4 – kadry ekonomiczno-administracyjne, X5 – personel urzędniczy, X6 – personel handlowy. Zmienne stymulanty: X1, X3, X4, X5, X6. Zmienna destymulanta: X2. Wagi zmiennych jednakowe: $w = (1,0; 1,0; 1,0; 1,0; 1,0; 1,0)$.

Fragment zbioru danych (hdane68):

```
> data(hdane68)
> head(hdane68)
      Kraj  X1  X2  X3  X4  X5  X6
1  Belgia 62.0  6.2 306.18 100.58 432.06 418.49
2   Dania 70.4 17.5 358.68  76.58 400.97 448.24
3 Finlandia 64.9 35.5 376.32  75.50 237.28 316.97
4   Grecja 67.5 53.9 148.96  31.67 171.41 264.17
5 Holandia 71.4 10.7 332.62 112.30 448.29 343.49
6   Indie 45.2 72.9  73.70  41.26  72.82 156.62
```

W przypadku danych z pracy [Hwang, Yoon 1981] (hwangyoondane81) porządkowanymi obiektami są 4 oferty zakupu samolotów myśliwskich charakteryzowane przez 6 zmiennych (por. [Chen, Hwang 1992]): X1 – prędkość maksymalna, X2 – zasięg, X3 – ładowność maksymalna, X4 – koszt zakupu, X5 – niezawodność, X6 – manewrowość. Zmienne stymulanty: X1, X2, X3, X5, X6. Zmienna destymulanta: X4. Wagi zmiennych zróżnicowane: $w = (0,2; 0,1; 0,1; 0,1; 0,2; 0,3)$.

Zbiór danych (hwangyoondane81):

```
> dane<-read.csv2("hwangyoondane81.csv",header=TRUE)
> head(dane)
  obiekt  X1  X2  X3  X4  X5  X6
1    A1  2.0 1500 20000 5.5  5  9
2    A2  2.5 2700 18000 6.5  3  5
3    A3  1.8 2000 21000 4.5  7  7
4    A4  2.2 1800 20000 5.0  5  5
```

W analizie zbioru danych z pracy [Hellwig 1968] wykorzystano skrypt:

```
library(pllford)
library(topsis)
source("htfunc68.r")
data(hdane68)
vc<-c(1,-1,1,1,1,1); i<-c("+","-","+","+", "+", "+"); w<-c(1,1,1,1,1,1)
hlo<-fhellwig(hdane68,vc)
tlo<-ftopsis(hdane68,w,i)
D<-data.frame(hlo$objn,hlo$dh,tlo$dt,hlo$rh,tlo$rt)
colnames(D)<-c("Object/Alternative","Hellwig score","TOPSIS score","Hellwig
rank","TOPSIS rank")
print(D)
print(cor(hlo$dh,tlo$dt,method="pearson"))
print(cor(hlo$rh,tlo$rt,method="spearman"))
```

Wyniki porządkowania 15 krajów metodami Hellwiga i TOPSIS są następujące:

```
> print(D)
  Object/Alternative Hellwig score TOPSIS score Hellwig rank TOPSIS rank
1          Belgia      0.51395752   0.57422780           8           5
2          Dania      0.54225684   0.53790471           6           8
3      Finlandia      0.40343640   0.40991816          10          11
4          Grecja      0.20198009   0.21427352          12          13
5      Holandia      0.54274018   0.56943878           5           6
6          Indie     -0.07043908   0.03731583          15          15
7          Japonia     0.54978591   0.58178931           4           4
8      Jugosławia     0.16708904   0.20056027          14          14
9          Kanada     0.59597175   0.71549754           2           2
10      Norwegia     0.46083166   0.50017347           9           9
11      Portugalia     0.17805069   0.25574218          13          12
12          USA       0.64844775   0.75442883           1           1
13      Szwajcaria     0.52413369   0.56025822           7           7
14      Szwecja       0.58302556   0.58756180           3           3
15      Węgry         0.33964491   0.41153548          11          10
> print(cor(hlo$dh,tlo$dt,method="pearson"))
[1] 0.9809814
> print(cor(hlo$rh,tlo$rt,method="spearman"))
[1] 0.9678571
```

W analizie zbioru danych z pracy [Hwang, Yoon 1981] wykorzystano skrypt:

```
library(pllford)
library(topsis)
source("htfunc81.r")
dane<-read.csv2("hwangyoondane81.csv",header=TRUE)
vc<-c(1,1,1,-1,1,1); i<-c("+","+", "+", "-","+", "+")
w<-c(0.2,0.1,0.1,0.1,0.1,0.2,0.3)
hlo<-fhellwig(dane,vc,w)
tlo<-ftopsis(dane,w,i)
D<-data.frame(hlo$objn,hlo$dh,tlo$dt,hlo$rh,tlo$rt)
colnames(D)<-c("Object/Alternative","Hellwig score","TOPSIS score","Hellwig
rank","TOPSIS rank")
print(D)
print(cor(hlo$dh,tlo$dt,method="pearson"))
print(cor(hlo$rh,tlo$rt,method="spearman"))
```


Wyniki porządkowania 4 ofert metodami Hellwiga i TOPSIS są następujące:

```
> print(D)
  Object/Alternative Hellwig score TOPSIS score Hellwig rank TOPSIS rank
1                A1    0.4982412    0.6432770             1             1
2                A2    0.1299144    0.2683836             4             4
3                A3    0.4147202    0.6135423             2             2
4                A4    0.2630782    0.3122911             3             3
> print(cor(hlo$dh,tlo$dt,method="pearson"))
[1] 0.9571978
> print(cor(hlo$rh,tlo$rt,method="spearman"))
[1] 1
```

Wyniki porządkowania liniowego obiektów ze zbioru `hdane68` są zgodne, jeżeli chodzi o pierwszą i ostatnią pozycję w rankingu, a także w przypadku sześciu innych obiektów (łącznie 8 z 15 obiektów znalazło się na tych samych miejscach w wyniku zastosowania metody Hellwiga i metody TOPSIS). Współczynnik korelacji liniowej Pearsona między wartościami zmiennych syntetycznych (Hellwig score i TOPSIS score) wyniósł 0,98. Współczynnik korelacji rang Spearmana między pozycjami w rankingu (Hellwig rank i TOPSIS rank) wyniósł 0,97.

Wyniki porządkowania liniowego obiektów ze zbioru `hwangyoondane81` są zgodne dla wszystkich obiektów. Współczynnik korelacji liniowej Pearsona między wartościami zmiennych syntetycznych (Hellwig score i TOPSIS score) wyniósł 0,96. Współczynnik korelacji rang Spearmana między pozycjami w rankingu (Hellwig rank i TOPSIS rank) wyniósł 1,0.

5. Podsumowanie

Miara rozwoju gospodarczego Hellwiga jest pierwszą historycznie metodą porządkowania liniowego zaproponowaną na gruncie ekonomii (taksonomii), natomiast metoda TOPSIS Hwanga i Yoona jest pierwszą metodą porządkowania liniowego zaproponowaną na gruncie teorii decyzji (wielokryterialnego podejmowania decyzji). Celem badań empirycznych z wykorzystaniem obu metod jest ustalenie kolejności (porządku) obiektów opisywanych przez zbiór zmiennych. Obydwie metody są wzorcowe, przy czym w metodzie Hellwiga punktem odniesienia jest wzorzec, a w metodzie Hwanga i Yoona – wzorzec i antywzorzec. Inna jest w obu metodach normalizacja zmiennych – w metodzie Hellwiga jest to standaryzacja, a w metodzie Hwanga i Yoona – przekształcenie ilorazowe. Miara odległości obiektów od punktów odniesienia jest w obu metodach taka sama – jest to odległość euklidesowa. Inna jest natomiast w obu metodach postać analityczna funkcji agregującej, a uzyskane wartości miary syntetycznej w metodzie Hellwiga na ogół są zawarte w przedziale [0; 1], a w metodzie Hwanga i Yoona są zawarte w przedziale [0; 1].

Główne kierunki dalszych badań:

- porównanie wyników porządkowania liniowego na podstawie większej liczby zbiorów danych empirycznych,

- porównanie wyników porządkowania liniowego na podstawie zbiorów danych symulacyjnych o określonych rozkładach statystycznych,
- porównanie jakości wyników porządkowania liniowego na podstawie mierników jakości metod porządkowania liniowego.

Literatura

- Bartosiewicz S., 1976, *Propozycja metody tworzenia zmiennych syntetycznych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 84.
- Bąk A., 1999, *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Bąk A., 2013, *Metody porządkowania liniowego w polskiej taksonomii – pakiet pllord*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 278, s. 54–62.
- Bąk A., 2015, *Linear ordering methods – package pllord*, <http://keii.ue.wroc.pl/pllord>.
- Borys T., 1978a, *Metody normowania cech w statystycznych badaniach porównawczych*, Przegląd Statystyczny, z. 2, s. 227–239.
- Borys T., 1978b, *Propozycja agregatywnej miary rozwoju obiektów*, Przegląd Statystyczny, z. 3, s. 371–381.
- Chen S.J., Hwang C.L., 1992, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York.
- Cieślak M., 1974, *Taksonomiczna procedura prognozowania rozwoju gospodarczego i określania potrzeb na kadry kwalifikowane*, Przegląd Statystyczny, z. 1, s. 29–39.
- Grabiński T., 1984, *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie. Seria specjalna: Monografie nr 61.
- Grabiński T., 1992, *Metody taksonometrii*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Grabiński T., Wydymus S., Zeliaś A., 1989, *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa.
- Hellwig Z., 1968, *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny, z. 4, s. 307–327.
- Hellwig Z., 1981, *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielo- cechowych obiektów gospodarczych*, [w:] W. Welfe (red.), *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, PWE, Warszawa.
- Hwang C.L., Yoon K., 1981, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York.
- Neumann J. von, Morgenstern O., 1944, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.
- Nowak E., 1984, *Problemy doboru zmiennych do modelu ekonometrycznego*, PWN, Warszawa.
- Pluta W., 1976, *Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego*, Przegląd Statystyczny, z. 4, s. 511–517.
- Pociecha J., 2008, *Rozwój metod taksonomicznych i ich zastosowań w badaniach społeczno-ekonomicznych*, <http://www.stat.gov.pl/cps/rde/xbcr/gus/> (11.10.2015).
- Pociecha J., Podolec B., Sokołowski A., Zajac K., 1988, *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa.

- R Development Core Team, 2015, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, <http://cran.r-project.org>.
- Strahl D., 1978, *Propozycja konstrukcji miary syntetycznej*, Przegląd Statystyczny, z. 2, s. 205–215.
- Tyszka T., 2010, *Decyzje. Perspektywa psychologiczna i ekonomiczna*, Wydawnictwo Naukowe Scholar, Warszawa.
- Walesiak M., 1993, *Statystyczna analiza wielowymiarowa w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 654, Monografie i Opracowania, nr 101.
- Wysocki F., 2010, *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań.
- Yazdi M.M., 2015, *TOPSIS method for multiple-criteria decision making (MCDM) – package topsis*, <https://cran.r-project.org/web/packages/topsis/>.