

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

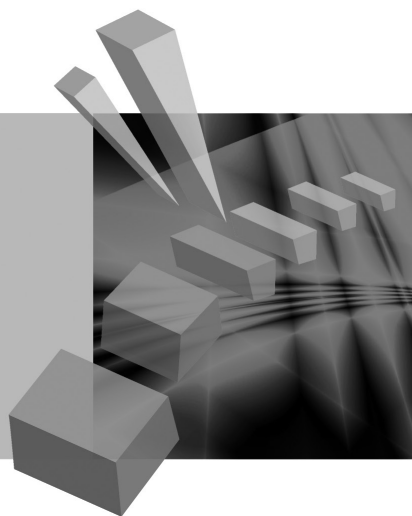
RESEARCH PAPERS

of Wrocław University of Economics

279

Taksonomia 21

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

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
Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: Sejm VI kadencji – maszynka do głosowania	11
Barbara Pawelek, Adam Sagan: Zmienne ukryte w modelach ekonomicznych – respecyfikacja modelu Kleina I	19
Jan Paradysz: Nowe możliwości badania koniunktury na rynku pracy	29
Krzysztof Najman: Samouczące się sieci GNG w grupowaniu dynamicznym zbiorów o wysokim wymiarze	41
Kamila Migdał-Najman: Zastosowanie jednowymiarowej sieci SOM do wyboru cech zmiennych w grupowaniu dynamicznym	48
Aleksandra Matuszewska-Janica, Dorota Witkowska: Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych	58
Iwona Foryś, Ewa Putek-Szeląg: Przestrzenna klasyfikacja gmin ze względu na sprzedaż użytków gruntowych zbywanych przez ANR w województwie zachodniopomorskim	67
Joanna Banaś, Małgorzata Machowska-Szewczyk: Klasyfikacja internetowych rachunków bankowych z uwzględnieniem zmiennych symbolicznych.....	77
Marta Jarocka: Wpływ metody doboru cech diagnostycznych na wynik porządkowania liniowego na przykładzie rankingu polskich uczelni	85
Anna Zamojska: Badanie zgodności rankingów wyznaczonych według różnych wskaźników efektywności zarządzania portfelem na przykładzie funduszy inwestycyjnych.....	95
Dorota Rozmus: Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	106
Ewa Wędrowska: Wrażliwość miar dywergencji jako mierników niepodobieństwa struktur.....	115
Katarzyna Wójcik, Janusz Tuchowski: Wpływ automatycznego tłumaczenia na wyniki automatycznej identyfikacji charakteru opinii konsumenckich ...	124
Małgorzata Misztal: Ocena wpływu wybranych metod imputacji na wyniki klasyfikacji obiektów w modelach drzew klasyfikacyjnych.....	135
Anna Czapkiewicz, Beata Basiura: Badanie wpływu wyboru współczynnika zależności na grupowanie szeregów czasowych	146
Tomasz Szubert: Czynniki różnicujące poziom zadowolenia z życia oraz wartości życiowe osób sprawnych i niepełnosprawnych w świetle badań „Diagnozy społecznej”	154

Marcin Szymkowiak: Konstrukcja estymatorów kalibracyjnych wartości globalnej dla różnych funkcji odległości	164
Wojciech Roszka: Szacowanie łącznych charakterystyk cech nieobserwowanych łącznie	174
Justyna Brzezińska: Metody wizualizacji danych jakościowych w programie R	182
Agata Sielska: Regionalne zróżnicowanie potencjału konkurencyjnego polskich gospodarstw rolnych w województwach po akcesji do Unii Europejskiej	191
Mariusz Kubus: Liniowy model prawdopodobieństwa z regularyzacją jako metoda doboru zmiennych	201
Beata Basiura: Metoda Warda w zastosowaniu klasyfikacji województw Polski z różnymi miarami odległości	209
Katarzyna Wardzińska: Wykorzystanie metody obwiedni danych w procesie klasyfikacji przedsiębiorstw	217
Katarzyna Dębowska: Modelowanie upadłości przedsiębiorstw oparte na próbach niezbilansowanych	226
Danuta Tarka: Wpływ metody doboru cech diagnostycznych na wyniki klasyfikacji obiektów na przykładzie danych dotyczących ochrony środowiska ..	235
Artur Czech: Zastosowanie wybranych metod doboru zmiennych diagnostycznych w badaniach konsumpcji w ujęciu pośrednim	246
Beata Bal-Domańska: Ocena relacji zachodzących między inteligentnym rozwojem a spójnością ekonomiczną w wymiarze regionalnym z wykorzystaniem modeli panelowych	255
Mariola Chrzanowska: <i>Ordinary kriging</i> i <i>inverse distance weighting</i> jako metody szacowania cen nieruchomości na przykładzie warszawskiego rynku	264
Adam Depta: Zastosowanie analizy wariancji w badaniu jakości życia na podstawie kwestionariusza SF-36v2	272
Maciej Beręsewicz, Tomasz Klimanek: Wykorzystanie estymacji pośredniej uwzględniającej korelację przestrzenną w badaniach cen mieszkań	281
Karolina Paradysz: Benchmarkowa analiza estymacji dla małych obszarów na lokalnych rynkach pracy	291
Anna Gryko-Nikitin: Dobór parametrów w równoległych algorytmach genetycznych dla problemu plecakowego	301
Tomasz Ząbkowski, Piotr Jałowicki: Zastosowanie reguł asocjacyjnych do analizy danych ankietowych w wybranych obszarach logistyki przedsiębiorstw przetwórstwa rolno-spożywczego	311
Agnieszka Przedborska, Małgorzata Misztal: Zastosowanie metod statystyki wielowymiarowej do oceny wydolności stawów kolanowych u pacjentów z chorobą zwyrodnieniową leczonych operacyjnie	321
Dorota Perło: Rozwój zrównoważony w wymiarze gospodarczym, społecznym i środowiskowym – analiza przestrzenna	331

Ewa Putek-Szeląg, Urszula Gieraltowska, Analiza i diagnoza wielkości produkcji energii odnawialnej w Polsce na tle krajów Unii Europejskiej..	342
--	-----

Summaries

Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: VIth-term Sejm – a voting machine	18
Barbara Pawelek, Adam Sagan: Latent variables in econometric models – respecification of Klein I model	28
Jan Paradysz: New possibilities for studying the situation on the labour market	40
Krzysztof Najman: Self-learning neural network of GNG type in the dynamic clustering of high-dimensional data.....	47
Kamila Migdał-Najman: Applying the one-dimensional SOM network to select variables in dynamic clustering	57
Aleksandra Matuszewska-Janica, Dorota Witkowska: Gender wage gap: application of classification trees.....	66
Iwona Foryś, Ewa Putek-Szeląg: Spatial classification of communes by usable land traded by the APA in the Zachodniopomorskie voivodeship...	76
Joanna Banaś, Małgorzata Machowska-Szewczyk: Classification of Internet banking accounts including symbolic variables	84
Marta Jarocka: The impact of the method of the selection of diagnostic variables on the result of linear ordering on the example of ranking of universities in Poland.....	94
Anna Zamojska: Empirical analysis of the consistency of mutual fund ranking for different portfolio performance measures.....	105
Dorota Rozmus: Comparison of accuracy of affinity propagation clustering and cluster ensembles based on bagging idea.....	114
Ewa Wędrowska: Sensitivity of divergence measures as structure dissimilarity measurements	123
Katarzyna Wójcik, Janusz Tuchowski: Machine translation impact on the results of the sentiment analysis	134
Małgorzata Misztal: Assessment of the influence of selected imputation methods on the results of object classification using classification trees ...	145
Anna Czapkiewicz, Beata Basiura: Simulation study of the selection of coefficient depending on the clustering time series.....	153
Tomasz Szubert: Factors differentiating the level of satisfaction with life and the life's values of people with and without disabilities in the light of the "Social Diagnosis" survey	162
Marcin Szymkowiak: Construction of calibration estimators of totals for different distance measures	173

Wojciech Roszka: Joint characteristics' estimation of variables not jointly observed.....	181
Justyna Brzezińska: Visualizing categorical data in \mathbf{R}	190
Agata Sielska: Regional diversity of competitiveness potential of Polish farms after the accession to the European Union	200
Mariusz Kubus: Regularized linear probability model as a filter	208
Beata Basiura: The Ward method in the application for classification of Polish voivodeships with different distances.....	216
Katarzyna Wardzińska: Application of Data Envelopment Analysis in company classification process.....	225
Katarzyna Dębowska: Modeling corporate bankruptcy based on unbalanced samples	234
Danuta Tarka: Influence of the features selection method on the results of objects classification using environmental data.....	245
Artur Czech: Application of chosen methods for the selection of diagnostic variables in indirect consumption research.....	254
Beata Bal-Domańska: Assessment of relations occurring between smart growth and economic cohesion in regional dimension using panel models	263
Mariola Chrzanowska: Ordinary kriging and inverse distance weighting as methods of estimating prices based on Warsaw real estate market	271
Adam Depta: Application of analysis of variance in the study of the quality of life based on questionnaire SF-36v2	280
Maciej Beręsewicz, Tomasz Klimanek: Using indirect estimation with spatial autocorrelation in dwelling price surveys.....	290
Karolina Paradysz: Benchmark analysis of small area estimation on local labor markets	300
Anna Gryko-Nikitin: Selection of various parameters of parallel evolutionary algorithm for knapsack problems	310
Tomasz Ząbkowski, Piotr Jałowiecki: Application of association rules for the survey of data analysis in the selected areas of logistics in food processing companies	320
Agnieszka Przedborska, Małgorzata Misztal: Using multivariate statistical methods to assess the capacity of the knee joint among the patients treated surgically for osteoarthritis	330
Dorota Perło: Sustainable development in the economic, social and environmental dimensions – spatial analysis.....	341
Ewa Putek-Szeląg, Urszula Gieraltowska: Analysis and diagnosis of the volume of renewable energy production in Poland compared to EU countries	352

Marta Jarocka

Politechnika Białostocka

WPLYW METODY DOBORU CECH DIAGNOSTYCZNYCH NA WYNIK PORZĄDKOWANIA LINIOWEGO NA PRZYKŁADZIE RANKINGU POLSKICH UCZELNI¹

Streszczenie: W artykule dokonano analizy wpływu metod doboru cech diagnostycznych na wynik hierarchizacji liniowej. W tym celu na podstawie opublikowanych w 2012 r. danych użytych do konstrukcji Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej” skonstruowano rankingi z różną kombinacją cech kryterialnych. W eksperymencie badawczym wykorzystano między innymi następujące narzędzia i metody doboru cech: klasyczny i pozycyjny współczynnik zmienności, metodę odwróconej macierzy korelacji, parametryczną metodę Hellwiga, analizę czynnikową, analizę skupień oraz metodę środka ciężkości.

Słowa kluczowe: dobór cech diagnostycznych, ranking, szkoły wyższe.

1. Wstęp

Dobór cech diagnostycznych jest ważnym etapem wielowymiarowej analizy porównawczej. Jakość zestawu cech analizowanych obiektów determinuje bowiem wiarygodność oczekiwanych rezultatów. E. Nowak podkreśla, że od doboru katalogu cech zależą ostateczne wyniki badania, do których autor zalicza: trafność ocen i analiz, dokładność przewidywań, a w konsekwencji trafność podejmowanych na ich podstawie decyzji [Nowak 1990, s. 23-33]. Proces doboru charakterystyk, ze względu na złożoną strukturę analizowanych zjawisk, nie jest też procesem łatwym. Pierwszy jego etap, tak zwany merytoryczny lub merytoryczno-formalny, polega na sporządzeniu tak zwanych potencjalnych cech diagnostycznych, które zdaniem T. Panka „w świetle posiadanej wiedzy merytorycznej o badanym zjawisku są najważniejsze dla dokonania analizy porównawczej badanych obiektów” [Panek 2009, s. 17]. Idea drugiego etapu – statystycznego – doboru cech diagnostycznych sprowadza się głównie do eliminacji cech o małym stopniu diagnostyczności, czyli charakteryzujących się niskim stopniem zmienności oraz wysokim stopniem skorelowania. Wybór cha-

¹ Artykuł finansowany ze środków Narodowego Centrum Nauki, nr N N111 530140.

rakterystyk może odbywać się zatem na podstawie analizy współczynników korelacji i zmienności. Metody i techniki badawcze, stosowane w ramach statystycznego kryterium zostały szeroko opisane m. in. w pracach [Hellwig 1981; Pluta 1977; Pocięcha i in. 1988; Grabiński, Wydymus, Zeliaś 1989; Nowak 1990; Kurkiewicz, Pocięcha, Zajac 1991; Młodak 2006; Panek 2009]. W tabeli 1 zaprezentowano katalog metod i technik badawczych najczęściej – według powyżej wymienionych badaczy – wykorzystywanych w badaniach empirycznych do weryfikacji zbioru potencjalnych cech diagnostycznych.

Tabela 1. Metody i procedury doboru cech diagnostycznych

Procedury i narzędzia oparte na analizie zmienności (badanie zdolności dyskryminacyjnej zmiennych)
klasyczny i pozycyjny współczynnik zmienności współczynnik względnej amplitudy wahań miara zdolności do hierarchizacji procedury oparte na analizie asymetrii
Procedury i narzędzia oparte na analizie korelacji (badanie potencjału informacyjnego zmiennych)
współczynniki korelacji liniowej metoda parametryczna Z. Hellwiga modyfikacja parametrycznej metody Z. Hellwiga (Młodak) metoda odwróconej macierzy korelacji modyfikacja metody wskaźników pojemności informacyjnej Z. Hellwiga (Nowak) metoda A.W. Bakker'a i M.L. Lukackiej metoda E.M. Brawermana metody B. Kinga procedury zaproponowane przez Plutę procedura taksonomii cech zaproponowana przez S. Bartosiewicza analiza czynnikowa
Procedury oparte na redukcji cech poprzez ich pogrupowanie i wyborze reprezentantek grup
<u>Metody grupowania cech</u>
np. metoda <i>k</i> -średnich, metoda Czekanowskiego, taksonomia wrocławska, metoda Prima, analiza wiązek J. C. Gowera-C.J.S. Rossa, procedury aglomeracyjne z grupy Lance'a-Williamsa-Warda, metody obszarowe – metoda kul i metoda katowicka
<u>Metody wyboru reprezentantek zmiennych:</u> metoda potencjału, metoda środka ciężkości

Źródło: opracowanie własne na podstawie [Hellwig 1981; Pluta 1977; Pocięcha i in. 1988; Grabiński, Wydymus, Zeliaś 1989; Nowak 1990; Kurkiewicz, Pocięcha, Zajac 1991; Młodak 2006; Panek 2009].

W badaniach empirycznych wykorzystuje się różne procedury doboru cech diagnostycznych, w wyniku których badacz może uzyskać inne zbiory charakterystyk

analizowanych obiektów. W związku z tym w pracy podjęto próbę odpowiedzi na pytanie: czy zmiana metody doboru cech diagnostycznych ma istotny wpływ na wynik porządkowania liniowego?

2. Metodyka eksperymentu badawczego

Do eksperymentu badawczego wykorzystano dane dotyczące 88 polskich uczelni zaprezentowane w Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej” 2012, w którym wykorzystano 33 cechy podzielone na sześć kryteriów podrzędnych: prestiż, innowacyjność, potencjał naukowy, efektywność naukowa, warunki studiowania oraz umiędzynarodowienie studiów (tab. 2). Ze względu na ograniczenia stron artykułu autorka pracy nie podaje obszernego opisu jednostek miar poszczególnych charakterystyk uczelni, odsyłając czytelnika do strony internetowej „Perspektyw”².

Tabela 2. Kryteria i cechy wykorzystane do oceny polskich uczelni w Rankingu Szkół Wyższych 2012 „Perspektyw” i „Rzeczpospolitej”

Kryterium	Ozn.	Cecha kryterialna	Waga
1	2	3	4
Prestiż	P1	Preferencje pracodawców	10%
	P2	Ocena przez kadrę akademicką	10%
	P3	Uznanie międzynarodowe	3%
	P4	Wybór olimpijczyków	2%
Innowacyjność	I1	Patenty, prawa ochronne i licencje	2%
	I2	Pozyskane środki z UE	2%
	I3	Zaplecze innowacyjne uczelni	1%
Potencjał naukowy	PN1	Ocena parametryczna	6%
	PN2	Uprawnienia habilitacyjne	2%
	PN3	Uprawnienia doktorskie	2%
	PN4	Nasylenie kadry osobami o najwyższych kwalifikacjach	3%
	PN5	Akredytacje	2%
Efektywność naukowa	EN1	Rozwój kadry własnej	8%
	EN2	Nadane stopnie i tytuły naukowe	7%
	EN3	Efektywność pozyskiwania zewnętrznych środków finansowych na badania	4%
	EN4	Publikacje	2%
	EN5	Cytowania	3%
	EN6	H-index	3%
	EN7	Udział uczelni w 7. Programie Ramowym UE	2%
	EN8	Studia doktoranckie	1%

² http://www.perspektywy.pl/index.php?option=com_content&task=view&id=5053&Itemid=906, stan z dnia 19. 05. 2012 r.

Tabela 2, cd.

1	2	3	4
Warunki studiowania	WS1	Dostępność dla studentów kadr wysokokwalifikowanych	5%
	WS2	Zbiory elektroniczne	1%
	WS3	Zbiory drukowane	1%
	WS4	Warunki korzystania z biblioteki	1%
	WS5	Możliwość rozwijania zainteresowań naukowych	1%
	WS6	Osiągnięcia sportowe	1%
Umieździarnodowienie	U1	Programy studiów prowadzone w j. obcych	4%
	U2	Studiujący w językach obcych	3%
	U3	Wymiana studencka (wyjazdy)	2%
	U4	Wymiana studencka (przyjazdy)	2%
	U5	Studenci cudzoziemcy	2%
	U6	Nauczyciele akademicy z zagranicy	1%
	U7	Wielokulturowość środowiska studenckiego	1%

Źródło: opracowanie własne na podstawie strony internetowej „Perspektywy” http://www.perspektywy.pl/index.php?option=com_content&task=view&id=5053&Itemid=906, stan z dnia 19.05.2012 r.

Pierwotne dane dotyczące polskich szkół wyższych zostały przekształcone zgodnie z formułą $z_{ij} = x_{ij} / \max \{x_{ij}\}$. W omawianym zbiorze cech kryterialnych uczelni występują wyłącznie cechy mierzone na skali ilorazowej.

Eksperyment badawczy polegał na utworzeniu sześciu list rankingowych uczelni z wykorzystaniem różnych, najczęściej używanych w badaniach empirycznych procedur doboru cech diagnostycznych, a następnie na komparacji uzyskanych wyników. Jako formułę agregacji cech obiektów zastosowano średnią arytmetyczną. W tabeli 3 przedstawiono oznaczenia utworzonych rankingów oraz użyte do ich konstrukcji metody doboru charakterystyk szkół wyższych.

Tabela 3. Oznaczenia konstruowanych rankingów oraz użyte do ich budowy procedury doboru cech diagnostycznych

Oznaczenie rankingu	Metody/narzędzia
R1	klasyczny i pozycyjny współczynnik zmienności
R2	metoda parametryczna Z. Hellwiga
R3	zmodyfikowana metoda parametryczna Z. Hellwiga (Młodak)
R4	metoda odwróconej macierzy korelacji
R5	analiza czynnikowa
R6	metoda Warda metoda środka ciężkości

Źródło: opracowanie własne.

W procesie konstrukcji rankingu R1 wykorzystano klasyczny współczynnik zmienności wyrażony wzorem: $V^k(x_j) = \frac{S(x_j)}{\bar{x}}$, $j = 1, 2, \dots, m$ oraz pozycyjny współczynnik zmienności liczony w sposób następujący: $V^p(x_j) = \frac{MOB(x_j)}{M(x_j)}$, $j = 1, 2, \dots, m$, gdzie: $MOB(x_j)$ – medianowe odchylenie bezwzględne j -tej cechy, $M(x_j)$ – mediana j -tej cechy [Młodak 2006, s. 28-29].

Do rankingu R2, jako metodę weryfikacji cech, zaproponowano metodę parametryczną Hellwiga [Hellwig 1981]. Algorytm eliminacji najsilniej skorelowanych ze sobą charakterystyk uczelni przebiegał według następujących etapów: 1) ustalenie w sposób arbitralny wartości progowej współczynnika korelacji na poziomie 0,7; 2) wyliczenie dla każdej kolumny macierzy korelacji \mathbf{R} sum ich bezwzględnych wartości; 3) wyznaczenie kolumny, dla której suma \mathbf{R}_j przyjęła najwyższą wartość; 4) zakwalifikowanie cechy odpowiadającej wyznaczonej kolumnie do cech centralnych; 5) w wyznaczonej kolumnie wybranie elementów o wartościach bezwzględnych większych od przyjętej wartości progowej współczynnika korelacji, na podstawie których stwierdza się istnienie silnej korelacji pomiędzy analizowanymi cechami; 6) eliminacja z macierzy korelacji cechy centralnej oraz cech z nią skorelowanych (satelitarnych); 7) powtarzanie powyższych czynności aż do momentu uzyskania tak zwanych cech izolowanych, dla których stopień skorelowania z innymi jest niższy od ustalonej wartości progowej współczynnika korelacji; 8) ustalenie ostatecznego zbioru charakterystyk uczelni zawierającego cechy centralne oraz izolowane. W literaturze przedmiotu wskazuje się na pewne niedoskonałości metody parametrycznej [Panek 2009, s. 22; Młodak 2006, s. 31]. Badacze wskazują na jej wrażliwość na wartości odstające oraz na fakt, iż uwzględnia ona wyłącznie bezpośrednie powiązania danej cechy z innymi cechami, nie uwzględniając pośrednich powiązań pomiędzy nimi. Proponują oni pewne możliwości zniwelowania tych niedoskonałości. A. Młodak, w celu zwiększenia odporności wyników metody na asymetrię rozkładu cech, zaproponował zastąpienie – w 2 etapie parametrycznej metody Hellwiga – sumy bezwzględnych wartości każdej z kolumn macierzy korelacji ich medianą [Młodak 2006, s. 31]. Odnośnie do drugiej wady A. Malina i A. Zeliaś oraz J. Lira i in. [Panek 2009, s. 22] zaproponowali metodę odwróconej macierzy korelacji, uwzględniającą zarówno bezpośrednie, jak i pośrednie powiązania cech. Algorytm tej metody przebiega w następujących krokach: 1) wyznaczenie macierzy odwrotnej do macierzy korelacji; 2) ustalenie wartości krytycznej elementów diagonalnych macierzy odwrotnej na poziomie 10; 3) wyszukanie elementów diagonalnych macierzy odwrotnej co do modułu mniejszych od przyjętej wartości krytycznej; 4) redukcja pierwotnego zbioru cech diagnostycznych o cechy spełniające warunek sformułowany w kroku 3.

W związku z powyższym w procesie konstrukcji rankingu R3 zastosowano modyfikację metody parametrycznej Hellwiga, zaproponowaną przez Młodaka. Zaś

dobór cech kryterialnych do rankingu R4 przebiegał zgodnie z algorytmem metody odwróconej macierzy korelacji.

W procesie konstrukcji rankingu R5 wykorzystano cechy uczelni uzyskanych w wyniku zastosowania analizy czynnikowej. Ze względu na obszerny opis procedury analizy czynnikowej autorka odsyła czytelnika do literatury przedmiotu [Waleśiak, Gatnar (red.) 2004, s. 186-245; Pluta 1977, s. 51-76].

Dobór cech szkół wyższych do ostatniego rankingu (R6) nastąpił w wyniku pogrupowania wszystkich pierwotnych cech diagnostycznych, a następnie wyboru reprezentantek uzyskanych grup. Klasyfikacji pierwotnych charakterystyk uczelni dokonano, wykorzystując jedną z aglomeracyjnych metod grupowania – metodę Warda, zaś wyboru przedstawicielek uzyskanych skupień dokonano za pomocą metody środka ciężkości. W aglomeracyjnych metodach grupowania klasyfikacji obiektów dokonuje się na podstawie macierzy odległości międzygrupowych, przy czym w metodzie Warda odległości pomiędzy skupieniami definiowane są jako moduł różnicy między sumami kwadratów odległości punktów od środków uzyskanych klastrów, do których te punkty należą [Pociecha i in. 1988, s. 83]. Metoda środka ciężkości polega zaś na wyborze reprezentantek uzyskanych grup, które znajdują się blisko środka ciężkości tychże skupień i jednocześnie nie pozostają blisko siebie w wielowymiarowej przestrzeni obiektów [Pluta 1977, s. 41-42; Pociecha i in. 1988, s. 106].

3. Wyniki i wnioski przeprowadzonego eksperymentu badawczego

W wyniku zastosowania sześciu procedur doboru cech diagnostycznych uzyskano różne zbiory charakterystyk uczelni, na podstawie których zbudowano rankingi polskich szkół wyższych. W tabeli 4 przedstawiono analizowane cechy uczelni, z podziałem na charakterystyki przyjęte do dalszego etapu badania oraz cechy odrzucone.

Z powyższego zestawienia wynika, że w procesie doboru cech diagnostycznych ograniczenie się tylko do analizy współczynników zmienności cech nie doprowadziło do eliminacji żadnej z nich (R1). Okazało się bowiem, że wszystkie 33 charakterystyki uczelni wykazują skuteczną dyskryminację obiektów. W przypadku zastosowania metody parametrycznej Hellwiga oraz zmodyfikowanej metody parametrycznej otrzymano identyczne zbiory cech kryterialnych do rankingu. Wynika to z faktu, że w zbiorze obserwacji nie ma odstających wartości cech. Podobne wyniki dały dwie kolejne metody: metoda odwróconej macierzy korelacji oraz analiza czynnikowa. W tabeli 5 zaprezentowano zestawienie odrzuconych cech uczelni w zależności od metody weryfikacji mającej na celu wyeliminowanie cech najsilniej skorelowanych ze sobą.

W tabeli 5 wyróżniono te cechy, które zostały wyeliminowane w wyniku zastosowania wszystkich czterech zaproponowanych metod. Należą do nich P3, PN5, EN5, EN6, EN7, U1, U2. Można zatem przypuszczać, że cechy te są faktycznie

Tabela 4. Zbiory cech diagnostycznych odrzuconych oraz przyjętych do budowy rankingów uczelni

	Metody/narzędzia	Cechy diagnostyczne do rankingu	Cechy diagnostyczne odrzucone
R1	klasyczny i pozycyjny współczynnik zmienności	wszystkie 33 cechy	–
R2	metoda parametryczna Z. Hellwiga	PN3, U7, EN4, WS2, EN2, WS6, PN2, WS1, P4, EN8, WS3, U4, I1, I3, PN1, PN4, EN1, EN3, WS3, WS4, U5, U6	P1, P2, P3, I2, PN5, EN5, EN6, EN7, U1, U2, WS5, U3
R3	zmodyfikowana metoda parametryczna Z. Hellwiga (Młodak)	PN3, U7, EN4, WS2, EN2, WS6, PN2, WS1, P4, EN8, WS3, U4, I1, I3, PN1, PN4, EN1, EN3, WS3, WS4, U5, U6	P1, P2, P3, I2, PN5, EN5, EN6, EN7, U1, U2, WS5, U3
R4	metoda odwróconej macierzy korelacji	P1, P4, I1, I3, PN1, PN2, PN4, EN1, EN2, EN3, EN8, WS1-WS6, U3-U7	P2, P3, PN3, PN5, EN4, EN5, EN6, EN7, U1, U2
R5	analiza czynnikowa	P2, WS1, P4, I1, I3, PN1, PN2, PN4, EN1-EN4, EN8, WS2-WS6, U3-U6	P1, P3, I2, PN3, PN5, EN5, EN6, EN7, U1, U2, U7
R6	metoda Warda metoda środka ciężkości	P2, WS4, EN6, EN1	P1, P3, P4, I1, I2, I3, PN1, PN2, PN3, PN4, PN5, EN2, EN3, EN4, EN5, EN7, EN8, WS1, WS2, WS3, WS5, WS6, U1, U2, U3, U4, U5, U6, U7

Źródło: opracowanie własne.

Tabela 5. Zestawienie odrzuconych cech diagnostycznych w zależności od metody ich doboru

	P1	P2	P3	I2	PN3	PN5	EN4	EN5	EN6	EN7	U1	U2	U3	U7	WS5
R2	–	–	–	–		–		–	–	–	–	–	–		–
R3	–	–	–	–		–		–	–	–	–	–	–		–
R4		–	–		–	–	–	–	–	–	–	–			
R5	–		–	–	–	–		–	–	–	–	–		–	

Źródło: opracowanie własne.

zbędne w dalszym etapie badania. Kierując się wiedzą na temat analizowanego zjawiska, należałoby zastanowić się nad losem pozostałych odrzuconych charakterystyk obiektów. Niektóre z nich, na przykład EN4 czy U7, zostały wyeliminowane w wyniku zastosowania tylko jednej procedury doboru cech. W pozostałych przypadkach stały się one elementami ostatecznego zbioru cech kryterialnych.

Nawiązując do wyników ostatniej z zaproponowanych do eksperymentu badawczego procedur doboru cech diagnostycznych – metody klasyfikacyjnej oraz metody wyboru reprezentantów grup (R6) – można stwierdzić, że zbiór przyjętych do konstrukcji rankingu charakterystyk uczelni jest znacznie mniej liczny niż w przypadku pozostałych uzyskanych zbiorów. Rezultatem zastosowania metody Warda były cztery grupy podobnych do siebie cech, których reprezentantkami zostały P2, WS4, EN6, EN1. Warto zwrócić uwagę na fakt, iż są one przedstawicielami tylko trzech z sześciu grup kryterialnych Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”, a mianowicie prestiżu, warunków studiowania oraz efektywności naukowej. Zatem ranking uczelni R6 bazuje tylko na ocenie przez kadrę akademicką (P2), warunkach korzystania z biblioteki (WS4), H-indeksie (EN6) oraz rozwoju kadry własnej (EN1). Ocena szkół wyższych nie uwzględnia zatem charakterystyk związanych z ich innowacyjnością, potencjałem naukowym oraz poziomem umiędzynarodowienia.

Na podstawie uzyskanych zbiorów cech diagnostycznych dokonano hierarchizacji uczelni. Zbudowano 6 rankingów polskich szkół wyższych, przy czym listy rankingowe bazujące na cechach uzyskanych w wyniku zastosowania metody parametrycznej Hellwiga i jej modyfikacji (R2, R3) są identyczne. W celu porównania otrzymanych klasyfikacji obliczono współczynniki korelacji rang Spearmana (tab. 6).

Tabela 6. Wartości współczynników korelacji rang Spearmana

	R2, R3	R4	R5	R6
R1	0,979	0,987	0,977	0,842
R2, R3		0,985	0,988	0,847
R4			0,985	0,827
R5				0,848

Źródło: opracowanie własne.

Uzyskane wartości współczynników korelacji świadczą o tym, iż najwyższym stopniem skorelowania wyróżniają się rankingi R2, R3, R4 i R5. Do ich budowy wykorzystano bowiem charakterystyki uczelni wyłonione w wyniku zastosowania procedur doboru cech diagnostycznych mających na celu wyeliminowanie cech najsilniej skorelowanych ze sobą. Najniższe wartości współczynników korelacji na poziomie 0,827-0,848 otrzymano dla rankingu R6 i wszystkich pozostałych R1, R2, R3, R4 i R5. Związane jest to z faktem, iż w procesie konstrukcji listy rankingowej R6 ze zbioru 33 pierwotnych cech kryterialnych odrzucono aż 29 z nich, przez co hierarchizacja uczelni odbyła się na podstawie tylko 4 charakterystyk analizowanych szkół. W pozostałych przypadkach bazowano na co najmniej 21 cechach.

4. Podsumowanie

Wyniki każdego rankingu w dużym stopniu zależą od prawidłowego określenia jego kryteriów. Ich merytoryczny dobór powinien być uzupełniony o odpowiednie procedury statystyczne. Zbiór potencjalnych cech diagnostycznych opracowany przez grupę ekspertów powinien być zweryfikowany ze względu na ich wartość informacyjną. Ze zbioru wielkości kryterialnych powinny być wykluczone te, które charakteryzują się małą zdolnością dyskryminacyjną lub powielają informację niesioną przez inne zmienne. W literaturze przedstawionych jest wiele metod doboru cech, od których zależą zarówno wyniki badań, jak i słuszność podejmowanych na ich podstawie decyzji. Wiele z nich, jak wskazują opisane w pracy wyniki przeprowadzonego eksperymentu badawczego, prowadzi również do różnych rezultatów. Ograniczenie się tylko do jednej procedury weryfikacyjnej cech – zdaniem autorki – może przyczynić się do podważenia wiarygodności przeprowadzonych ocen i analiz, a w konsekwencji trafności podejmowanych na ich podstawie decyzji.

Literatura

- Grabiński T., Wydymus S., Zeliaś A., *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa 1989.
- Hellwig Z., *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielowymiarowych obiektów gospodarczych*, [w:] W. Welfe (red.), *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1981, s. 46-68.
- Kurkiewicz J., Pocięcha J., Zajac K., *Metody wielowymiarowej analizy porównawczej w badaniach rozwoju demograficznego*, Szkoła Główna Handlowa, Instytut Statystyki i Demografii, Monografie i Opracowania nr 336, Warszawa 1991.
- Młodak A., *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa 2006.
- Nowak E., *Metody taksonomiczne w klasyfikacji obiektów społeczno-ekonomicznych*, Wydawnictwo Naukowe PWN, Warszawa 1990.
- Panek T., *Statystyczne metody wielowymiarowej analizy porównawczej*, Szkoła Główna Handlowa w Warszawie, Warszawa 2009.
- Pluta W., *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1977.
- Pocięcha J., Podolec B., Sokołowski A., Zajac K., *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, Wydawnictwo Naukowe PWN, Warszawa 1988.
- Walesiak M., Gatnar E (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, 2004.

THE IMPACT OF THE METHOD OF THE SELECTION OF DIAGNOSTIC VARIABLES ON THE RESULT OF LINEAR ORDERING ON THE EXAMPLE OF RANKING OF UNIVERSITIES IN POLAND

Summary: This article presents the analysis of the impact of the method of the selection of diagnostic variables on the result of linear ordering. Based on data from the ranking 2012 by the „Perspektywy” and „Rzeczpospolita”, using different combinations of criterion variables, the university rankings were constructed. The following method and tools were used in this research: the coefficient of variation, method of inverse correlation matrix, Hellwig method and its modification, factor analysis, cluster analysis and method of center of gravity. Finally, the Spearman’s rank correlation coefficient was calculated.

Keywords: selection of data, ranking, universities.