

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 328

Taksonomia 23

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

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

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

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

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 2014

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	11
Małgorzata Rószkiewicz , Wykorzystanie metaanalizy w budowaniu modelu pomiarowego w przypadku braku niezmienniczości zasad pomiaru na przykładzie pomiaru zadowolenia z życia.....	13
Elżbieta Sobczak , Harmonijność inteligentnego rozwoju regionów Unii Europejskiej	21
Ewa Roszkowska, Renata Karwowska , Analiza porównawcza województw Polski ze względu na poziom zrównoważonego rozwoju w roku 2010.....	30
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Analiza porównawcza wybranych filtrów w analizie synchronizacji cyklu koniunkturalnego.....	41
Marcin Salamaga , Próba konstrukcji tablic „wymierania scenicznego” spektakli operowych na przykładzie Metropolitan Opera.....	51
Iwona Foryś , Wykorzystanie analizy dyskryminacyjnej do typowania rynków podobnych w procesie wyceny nieruchomości niemieszkalnych	59
Jerzy Korzeniewski , Selekcja zmiennych w klasyfikacji – propozycja algorytmu	69
Sabina Denkowska , Testowanie wielokrotne przy weryfikacji wieloczynnikowych modeli proporcjonalnego hazardu Coxa.....	76
Ewa Chodakowska , Teoria równań strukturalnych w klasyfikacji zmiennych jawnych i ukrytych według charakteru ich wzajemnych oddziaływań	85
Iwona Konarzewska , Model PCA dla rynku akcji – studium przypadku	94
Katarzyna Wójcik, Janusz Tuchowski , Dobór optymalnego zestawu słów istotnych w opiniach konsumentów na potrzeby ich automatycznej analizy	106
Aleksandra Łuczak , Zastosowanie metody AHP-LP do oceny ważności determinant rozwoju społeczno-gospodarczego w jednostkach administracyjnych	116
Aleksandra Witkowska, Marek Witkowski , Klasyfikacja pozycyjna banków spółdzielczych według stanu ich kondycji finansowej w ujęciu dynamicznym	126
Adam Depta , Zastosowanie analizy korespondencji do oceny jakości życia ludności na podstawie kwestionariusza SF-36v2	135
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Indukcja reguł dla danych niekompletnych i niezbalansowanych: modele klasyfikatorów i próba ich zastosowania do predykcji ryzyka operacyjnego w torakochirurgii	146

Małgorzata Misztal , Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań.....	156
Anna M. Olszewska , Wykorzystanie wybranych metod taksonomicznych do oceny potencjału innowacyjnego województw	167
Iwona Bąk , Porównanie jakości grupowań powiatów województwa zachodniopomorskiego pod względem atrakcyjności turystycznej.....	177
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Segmentacja gospodarstw domowych według wydatków na turystykę zorganizowaną.....	186
Agnieszka Wałęga , Podejście syntetyczne w analizie spójności ekonomicznej gospodarstw domowych.....	196
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Zastosowanie analizy korespondencji do badania wpływu elektrowni wiatrowych na jakość życia ludności	205
Joanna Banaś, Krzysztof Małecki , Klasyfikacja punktów pomiarów ankietowych kierowców na granicy Szczecina z wykorzystaniem zmiennych symbolicznych.....	214
Aneta Becker , Wykorzystanie informacji granularnej w analizie wymagań rynku pracy.....	222
Katarzyna Cheba, Joanna Holub-Iwan , Wykorzystanie analizy korespondencji w segmentacji rynku usług medycznych.....	230
Adam Depta, Iwona Staniec , Identyfikacja czynników decydujących o jakości życia studentów łódzkich uczelni.....	238
Katarzyna Dębowska, Jarosław Kilon , Reguły asocjacyjne w analizie wyników badań metodą Delphi.....	247
Anna Domagała , O wykorzystaniu analizy głównych składowych w metodzie <i>Data Envelopment Analysis</i>	254
Alicja Grześkowiak , Analiza wykluczenia cyfrowego w Polsce w ujęciu indywidualnym i regionalnym.....	264
Anna M. Olszewska, Anna Gryko-Nikitin , Pomiar postrzegania jakości kształcenia uczelni wyższej na danych porządkowych z wykorzystaniem środowiska R.....	273
Karolina Paradysz , Hierarchiczna metoda grupowania powiatów jako podejście benchmarkowe w ocenie bezrobocia według BAEL-u w wybranych typach małych obszarów	282
Radosław Pietrzyk , Porównanie metod pomiaru efektywności zarządzania portfelami funduszy inwestycyjnych.....	290
Agnieszka Przedborska, Małgorzata Misztal , Wybrane metody statystyki wielowymiarowej w ocenie skuteczności terapeutycznej głębokiej stymulacji elektromagnetycznej u pacjentów z chorobą zwyrodnieniową stawów.....	299

Wojciech Roszka, Marcin Szymkowiak , Podejście kalibracyjne w statystycznej integracji danych	308
Iwona Skrodzka , Zastosowanie wybranych metod klasyfikacji do analizy kapitału ludzkiego krajów Unii Europejskiej	316
Agnieszka Stanimir , Wielowymiarowa analiza czynników sprzyjających włączeniu społecznemu	326
Dorota Strózik, Tomasz Strózik , Przestrzenne zróżnicowanie poziomu życia w województwie wielkopolskim.....	334
Izabela Szamrej-Baran , Identyfikacja przyczyn ubóstwa energetycznego w Polsce przy wykorzystaniu modelowania miękkiego.....	343
Janusz Tuchowski, Katarzyna Wójcik , Klasyfikacja obiektów w systemie Krajowych Ram Kwalifikacji opisanych za pomocą ontologii	353
Aleksandra Matuszewska-Janica , Grupowanie krajów Unii Europejskiej ze względu na poziom feminizacji sektorów gospodarczych	361
Monika Rozkrut, Dominik Rozkrut , Identyfikacja strategii innowacyjnych przedsiębiorstw usługowych w Polsce	369

Summaries

Małgorzata Rószkiewicz , The use of meta-analysis in building the measurement model in case of the absence of measurement invariance on the example of measuring of life satisfaction.....	20
Elżbieta Sobczak , Harmonious smart growth of European Union regions.....	29
Ewa Roszkowska, Renata Karwowska , The comparative analysis of Polish voivodeships with respect to sustainable development in 2010.....	40
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Comparative analysis of chosen filters in business cycles analysis	50
Marcin Salamaga , The attempt of construction of the life tables for opera works on the example of the Metropolitan Opera	58
Iwona Foryś , Using discriminant analysis to select similar markets in non-residential property valuation process.....	68
Jerzy Korzeniewski , Variable selection in classification – algorithm proposal	75
Sabina Denkowska , Multiple testing in the verification process of multifactorial Cox proportional hazards models	84
Ewa Chodakowska , The theory of structural equations modelling in the classification of observed variables and latent constructs according to the character of their relationship.....	93
Iwona Konarzewska , Modelling stock market by PCA factor model – case study	105

Katarzyna Wójcik, Janusz Tuchowski , Selection of the optimal set of relevant words in consumers opinions in the context of the opinion mining ..	115
Aleksandra Łuczak , Application of AHP-LP to the evaluation of importance of determinants of socio-economic development in the administrative units	125
Aleksandra Witkowska, Marek Witkowski , A dynamic approach to the ranking of cooperative banks by their financial condition	134
Adam Depta , Application of correspondence analysis for the measurement of quality of life – questionnaire SF-36v2 based research	145
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Classification rules extraction for missing and imbalance data: models of classifiers and initial results in the rules-based thoracic surgery risk prediction.....	155
Małgorzata Misztal , Selected methods for assessing the performance of classifiers – an overview and examples of applications.....	166
Anna M. Olszewska , The application of selected quantitative methods to the evaluation of voivodeship innovation level potential.....	176
Iwona Bąk , The comparison of the quality of groupings of poviats of West Pomeranian Voivodeship in terms of tourism attractiveness	185
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Household segmentation with respect to the expenditure on organized tourism.....	195
Agnieszka Wałęga , Synthetic approach in the analysis of economic coherence of households	204
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Using the correspondence analysis to examine the impact of wind turbines on the quality of life.....	213
Joanna Banaś, Krzysztof Małecki , Classification of measurement survey points of drivers on the boundary of Szczecin using symbolic variables...	221
Aneta Becker , The use granular information in the analysis of the requirements of the labor market.....	229
Katarzyna Cheba, Joanna Hołub-Iwan , The application of the correspondence analysis of patients segmentation on the medical service market	237
Adam Depta, Iwona Staniec , Identification of the factors that determine the quality of students life at universities in Lodz.....	246
Katarzyna Dębowska, Jarosław Kilon , Association rules in the analysis of research results the Delphi method	253
Anna Domagała , About using Principal Component Analysis in Data Envelopment Analysis	263
Alicja Grześkowiak , Analysis of the digital divide in Poland at the individual and regional level	272

Anna M. Olszewska, Anna Gryko-Nikitin , Assessment of perception of quality of teaching at an institution of higher learning based on the ordinal data with the utilization of R environment.....	281
Karolina Paradysz , The hierarchical method of grouping poviats as a benchmark approach in the assessment of unemployment by BAEL in selected types of small areas	289
Radosław Pietrzyk , Comparison of methods of measuring the performance of investment funds portfolios.....	298
Agnieszka Przedborska, Małgorzata Misztal , Selected multivariate statistical analysis methods in the evaluation of efficacy of deep electromagnetic stimulation in patients with degenerative joint disease	307
Wojciech Roszka, Marcin Szymkowiak , A calibration approach in statistical data integration	315
Iwona Skrodzka , Application of some methods of classification to the analysis of human capital in the European Union.....	325
Agnieszka Stanimir , Multivariate analysis of social inclusion factors.....	333
Dorota Strózik, Tomasz Strózik , Spatial differentiation of the standard of living in Great Poland Voivodeship	342
Izabela Szamrej-Baran , Identification of fuel poverty causes in Poland using soft modelling	352
Janusz Tuchowski, Katarzyna Wójcik , Classification of objects in the National Classification Framework described by the ontology.....	360
Aleksandra Matuszewska-Janica , Clustering of European Union states taking into consideration the levels of feminization of economic sectors..	368
Monika Rozkrut, Dominik Rozkrut , Identification of service sector innovation strategies in Poland.....	379

Anna M. Olszewska

Politechnika Białostocka

WYKORZYSTANIE WYBRANYCH METOD TAKSONOMICZNYCH DO OCENY POTENCJAŁU INNOWACYJNEGO WOJEWÓDZTW

Streszczenie: Innowacyjność staje się coraz częściej wyznacznikiem rozwoju organizacji, ale również całych regionów. Jest pojęciem złożonym, trudnym do jednoznacznego zdefiniowania szczególnie w ujęciu regionalnym, co wynikać może między innymi z tego, że innowacyjność regionalna zależna jest od różnych procesów zachodzących zarówno w obszarze społecznym, jak i gospodarczym. Z powodu złożoności opisywanego zjawiska jego pomiar może sprawiać trudność. Jest też obarczony wieloma ograniczeniami, często wynikającymi z dostępności danych. Dlatego tak ważny w tym kontekście wydaje się właściwy dobór dostępnych zmiennych. Głównym celem artykułu jest dokonanie klasyfikacji województw pod kątem ich potencjału innowacyjnego. Cel ten został osiągnięty poprzez przeprowadzenie analizy wpływu metod doboru zmiennych na wyniki badania. W odniesieniu do dokonanego wyboru przeprowadzono zestawienie klasyfikacji obiektów, uwypuklając zmiany zachodzące w wynikach podziału przy uwzględnieniu różnych form doboru zmiennych.

Słowa kluczowe: innowacyjność, metody doboru zmiennych, klasyfikacja obiektów.

1. Wstęp

Innowacja jest pojęciem wprowadzonym przez J. Schumpetera, który określał ją jako element rozdzielający inwencję i dyfuzję [Mahdjoubi 1997, s. 2]. Definiował ją jako wykonanie nowych bądź już realizowanych rzeczy w sposób unikatowy [Schumpeter 1947, s. 151]. Od tego czasu pojęcie ewaluowało, nabierając coraz szerszego sensu. Obecnie obejmuje zarówno oryginalną ideę czy pomysł, jak też proces, którego celem jest przygotowanie i wprowadzenie nowatorskiego rozwiązania [Nazarko 2013, s. 9]. Natomiast innowacyjność określana jest jako wypadkowa wielu procesów mających zarówno charakter społeczny, jak i przestrzenny czy gospodarczy [Nowakowska (red.) 2009, s. 11]. Związana jest z szeregiem czynników, wśród których wyróżniane są te sprzyjające tworzeniu innowacji, jak też związane z działalnością innowacyjną lokalnych przedsiębiorstw, czy osiąga-

nymi w tym względzie wynikami [Bukowski i in. 2012, s. 15]. Do zmiennych kształtujących poziom innowacyjności regionów – poza tymi, które są bezpośrednio przypisane do przedsiębiorstw – należy zaliczyć również czynniki obejmujące relacje organizacji z otoczeniem instytucjonalnym, w tym m.in. system edukacji, poziom szkolnictwa wyższego, bazę naukowo-badawczą, otoczenie prawne czy uwarunkowania infrastrukturalne [OECD 2005, s. 37].

Podstawowym celem przeprowadzonego badania było dokonanie klasyfikacji województw przy wykorzystaniu miar syntetycznych, umożliwiających zarówno uszeregowanie, jak i pogrupowanie badanych jednostek terytorialnych z punktu widzenia ich potencjału innowacyjnego. Poza tym wskazano, jakie znaczenie dla przeprowadzonych wyników klasyfikacji ma zastosowanie miar statystycznych wykorzystywanych podczas redukcji zgromadzonych pierwotnie zmiennych. Użyte w tytule określenie metod taksonomicznych ograniczono do zakresu klasycznie rozumianej taksonomii numerycznej, obejmującej m.in. porządkowanie liniowe [Pociecha 2008, s. 2], zawiężając zakres zainteresowania do pierwszego etapu postępowania badawczego, a mianowicie doboru zmiennych.

Warto podkreślić, że dobór zmiennych diagnostycznych ma fundamentalne znaczenie w prowadzonym badaniu porównawczym. W znacznym stopniu przyczynia się on do uzyskania wyników końcowych badania, niezależnie od zastosowanych w dalszej kolejności technik i metod [Panek 2009, s. 16].

2. Opis przeprowadzonej procedury badawczej

Realizując postawiony cel, przeprowadzono eksperyment, którego wynikiem miał być ranking województw. Uszeregowanie dotyczyło poziomu innowacyjności badanych obiektów. Pierwotnie uzyskane zmienne, wykorzystane podczas badania, pochodziły z roczników statystycznych GUS i raportów dotyczących innowacyjności regionów publikowanych przez GUS. Obejmowały one przede wszystkim rok 2011, a w szczególnych przypadkach, głównie z powodu problemów z dostępnością danych, odnoszone były do roku 2010 lub 2009.

Zastosowana procedura doboru została podzielona na dwa etapy. Podczas pierwszego etapu, w wyniku analizy merytorycznoformalnej, dokonano redukcji zgromadzonych zmiennych, ograniczając wyjściowy materiał badawczy do 34 wskaźników pozwalających zdefiniować potencjał innowacyjny województw. Podczas doboru merytorycznego autorka starała się zachować reprezentatywność każdej z determinant innowacyjności, a mianowicie [Prystrom 2012, s. 53]: kapitału ludzkiego, transferu technologii, źródeł finansowania, otoczenia instytucjonalnego, polityki innowacyjnej. Wybrany na tym etapie zestaw zmiennych przedstawiono w tabeli 1.

W kolejnym kroku zastosowano do danych ograniczonych w wyniku analizy merytorycznoformalnej wybrane procedury statystyczne. Początkowo wyjściową pulę wskaźników zawężono, wykorzystując kryterium ich zdolności dyskrymina-

Tabela 1. Zestaw zmiennych w porządku alfabetycznym

Z 1	Eksport wyrobów wysokiej i średniowysokiej techniki (w % eksportu ogółem)
Z 2	Gospodarstwa domowe wyposażone w komputer (w % gospodarstw ogółem)
Z 3	Gospodarstwa domowe wyposażone w komputer z dostępem do Internetu (w % gospodarstw ogółem)
Z 4	Liczba absolwentów przypadających na 10 tys. ludności
Z 5	Liczba doktoratów obronionych w 2011 r., przypadających na milion ludności ogółem
Z 6	Liczba habilitacji obronionych w 2011 r., przypadających na milion ludności ogółem
Z 7	Liczba jednostek prowadzących działalność badawczo-rozwojową, przypadających na 10 tys. przedsiębiorstw
Z 8	Liczba słuchaczy studiów podyplomowych przypadających na 10 tys. ludności w wieku produkcyjnym
Z 9	Liczba studentów na 1 tys. ludności w wieku produkcyjnym
Z 10	Liczba uczelni na 100 tys. ludności w wieku produkcyjnym
Z 11	Liczba uczestników studiów doktoranckich przypadających na 10 tys. ludności
Z 12	Nakłady na działalność badawczo-rozwojową ponoszone przez przedsiębiorstwa (w % nakładów)
Z 13	Nakłady na działalność badawczo-rozwojową na 10 tys. przedsiębiorstw
Z 14	Nakłady na działalność badawczo-rozwojową przypadające na 1 osobę (ogółu ludności)
Z 15	Nakłady na działalność innowacyjną na 1 tys. zł nakładów inwestycyjnych
Z 16	Nakłady na działalność innowacyjną na 10 tys. przedsiębiorstw
Z 17	Nakłady ze środków własnych (w % ogółu nakładów) na działalność innowacyjną w zakresie innowacji produktowych i procesowych w przemyśle
Z 18	Osoby z tytułem naukowym profesora, stopniem naukowym doktora lub doktora habilitowanego zatrudnione w B+R (w % zatrudnionych w B+R)
Z 19	Osoby z wykształceniem wyższym (w % ogółu ludności)
Z 20	Patenty i prawa ochronne na wzory użytkowe na milion ludności
Z 21	PKB na jednego mieszkańca w 2010 r. w tys. zł
Z 22	Przedsiębiorstwa mające dostęp do Internetu (w % ogółu przedsiębiorstw)
Z 23	Przedsiębiorstwa przemysłowe, które współpracowały w zakresie działalności innowacyjnej w latach 2009-2011 (w % ogółu przedsiębiorstw)
Z 24	Przedsiębiorstwa wykorzystujące komputery (w % ogółu przedsiębiorstw)
Z 25	Przedsiębiorstwa wykorzystujące zewnętrzną wymianę danych w ogóle przedsiębiorstw
Z 26	Przedsiębiorstwa, które wprowadziły innowacje produktowe lub procesowe (w % ogółu przedsiębiorstw)
Z 27	Przedsiębiorstwa, które wprowadziły nowe lub ulepszone procesy (% ogółu przedsiębiorstw)
Z 28	Przedsiębiorstwa, które wprowadziły nowe lub ulepszone produkty (% ogółu przedsiębiorstw)
Z 29	Przychody netto z eksportu produktów nowych lub istotnie ulepszonych (w % przychodów netto ze sprzedaży ogółem)
Z 30	Przychody netto ze sprzedaży produktów nowych lub istotnie ulepszonych (w % przychodów netto ze sprzedaży ogółem)
Z 31	Relacja nakładów na B+R do produktu krajowego brutto (PKB) (w % w 2010 r.)
Z 32	Zasoby ludzkie dla nauki i techniki (odsetek osób w wieku 15-74 lata)
Z 33	Zatrudnienie w działalności badawczo-rozwojowej na 1 tys. aktywnych zawodowo
Z 34	Zgłoszenia wynalazków i wzorów użytkowych na milion ludności

Źródło: opracowanie własne.

cyjnej. Na tym etapie zastosowano klasyczny współczynnik zmienności [Młodak 2006, s. 28-29], ustalając jego wartość progową na 0,1. Uwzględniając jednak sugestie dotyczące postaci rozkładu [Czech 2013; Młodak 2006], sprawdzono postać rozkładów wybranych w pierwszym kroku zmiennych poprzez określenie poziomu ich asymetrii i koncentracji.

Analizując otrzymane wyniki, zauważono wysoki poziom asymetrii części zmiennych i ich znaczną koncentrację. Dlatego też w dalszym postępowaniu dokonano redukcji otrzymanego zestawu zmiennych za pomocą innej niż klasyczna metody, wykorzystując dwie odmiany pozycyjnych wskaźników zmienności (pierwszy jako iloraz medianowego odchylenia bezwzględnego i mediany [Młodak 2006, s. 28-29], zaś drugi – iloraz odchylenia ćwiartkowego i mediany [Tarka 2013, s. 50]). W tym przypadku również wartość progowa eliminacji zmiennych wynosiła 0,1.

Przeprowadzona redukcja pozwoliła na sporządzenie różnych wynikowo zestawów, stanowiących materiał wejściowy dla kolejnego etapu, obejmującego eliminację zmiennych skorelowanych. W tym kroku zastosowano parametryczną metodę Hellwiga. Metoda ta wykorzystuje macierz współczynników korelacji liniowej Pearsona. Możliwe jest zastosowanie innych miar badających zależność zmiennych, jednak w ramach niniejszego opracowania ograniczono rozważania do wskazanego współczynnika.

Parametryczna metoda Hellwiga polega na ustaleniu w pierwszym kroku wartości progowej współczynnika korelacji r^* , powyżej którego zależność uznaje się za istotną. Wartość ta może być ustalona arbitralnie, zgodnie z metodą minimaksową lub weryfikacyjną istotności współczynnika korelacji [Panek 2009, s. 21]. W przyjętym postępowaniu eksperymentalnym zastosowano dwa podejścia: ustalono arbitralnie najczęściej przyjmowaną wartość jako progową $r^* = 0,7$ [Tarka 2013, s. 236] i wyznaczono ją metodą minimaksową. Wielu autorów wskazuje na pewne niedoskonałości opisywanej procedury, co wynika z jej wrażliwości na wartości odstające [Panek 2009, s. 21-22; Młodak 2006, s. 31; Jarocka 2013, s. 89]. Aby zwiększyć odporność na asymetrię rozkładu A. Młodak zaproponował użycie, zamiast sum bezwzględnych, ich mediany [Młodak 2006, s. 31]. W przeprowadzonym badaniu wykorzystano obie metody. Finalny zestaw użytych metod doboru przedstawiono w tabeli 2.

Otrzymany w wyniku przeprowadzonego badania zestaw zmiennych poddano standaryzacji i przy wykorzystaniu wzorcowej metody Hellwiga [Panek 2009, s. 69] uszeregowano województwa pod kątem ich potencjału innowacyjnego. Porównując otrzymane wyniki rankingów, zbadano ich poziom skorelowania; w tym celu wykorzystano współczynnik korelacji rang Spearmana.

Po zastosowaniu wzorcowej metody Hellwiga możliwe jest przeprowadzenie podziału badanych obiektów na cztery grupy [Młodak 2006, s. 127]. Chcąc określić, jak bardzo wyniki otrzymanych procedur różnią się od siebie, porównano

Tabela 2. Oznaczenie metod doboru zmiennych

Oznaczenie metody	Metody badające zdolność dyskryminacyjną zmiennych	Metoda badająca potencjał informacyjny zmiennych
M1	klasyczny współczynnik zmienności	metoda parametryczna Z. Hellwiga dla $r^* = 0,7$
M2		metoda parametryczna Z. Hellwiga dla r^* wyznaczonego metodą minimaxową
M3		zmodyfikowana metoda parametryczna Z. Hellwiga dla $r^* = 0,7$
M4		zmodyfikowana metoda parametryczna Z. Hellwiga dla r^* wyznaczonego metodą minimaxową
M5	pozycyjny współczynnik zmienności (medianowe odchylenie bezwzględne)	metoda parametryczna Z. Hellwiga dla $r^* = 0,7$
M6		metoda parametryczna Z. Hellwiga dla r^* wyznaczonego metodą minimaxową
M7		zmodyfikowana metoda parametryczna Z. Hellwiga dla $r^* = 0,7$
M8		zmodyfikowana metoda parametryczna Z. Hellwiga dla r^* wyznaczonego metodą minimaxową
M9	pozycyjny współczynnik zmienności (odchylenie ćwiartkowe)	metoda parametryczna Z. Hellwiga dla $r^* = 0,7$
M10		metoda parametryczna Z. Hellwiga dla r^* wyznaczonego metodą minimaxową
M11		zmodyfikowana metoda parametryczna Z. Hellwiga dla $r^* = 0,7$
M12		zmodyfikowana metoda parametryczna Z. Hellwiga dla r^* wyznaczonego metodą minimaxową

Źródło: opracowanie własne.

utworzone grupy, wykorzystując indeks Randa. Zbadano także poprawność użytych metod grupowania poprzez wyznaczenie miar homogeniczności i heterogeniczności grup obiektów oraz miernika poprawności grupowania jako ich ilorazu [Panek 2009, s. 164-167].

3. Wyniki badania poziomu innowacyjności województw

Podczas pierwszego etapu procedury statystycznego doboru zmiennych dokonano redukcji zmiennych cechujących się niewielkim poziomem zmienności. Zestaw zmiennych, które zostały odrzucone na tym etapie, przedstawiono w drugiej kolumnie w tabeli 3.

Po redukcji zmiennych niemających zdolności dyskryminacyjnej przystąpiono do kolejnego etapu – parametrycznej metody Hellwiga i jego wersji zmodyfikowanej. Wyniki z zastosowania tych metod umieszczono w tabeli 3 w kolumnie trzeciej i czwartej.

Wyniki rankingu poszczególnych województw otrzymane po przeprowadzonym doborze zostały przedstawione w tabeli 4. Wartości w nawiasach ukazują przynależność do jednej z czterech utworzonych na podstawie miernika syntetycznego grup województw.

Tabela 3. Zestawienie zmiennych odrzuconych podczas analizy zdolności dyskryminacyjnej oraz zmienne centralne i satelitarne z metod badających potencjał informacyjny

Nazwa metody	Zmienne odrzucone	Zmienne centralne i izolowane	Zmienne satelitarne
M1	Z2, Z3, Z22, Z24	Z1, Z4, Z7, Z12, Z15, Z17, Z18, Z20, Z21, Z23, Z25, Z26, Z29	Z5, Z6, Z8, Z9, Z10, Z11, Z13, Z14, Z16, Z19, Z27, Z28, Z30, Z31, Z32, Z33, Z34
M2		Z1, Z7, Z11, Z12, Z18, Z25, Z28, Z29	Z4, Z5, Z6, Z8, Z9, Z10, Z13, Z14, Z15, Z16, Z17, Z19, Z20, Z21, Z23, Z26, Z27, Z30, Z31, Z32, Z33, Z34
M3		Z1, Z11, Z12, Z15, Z17, Z18, Z20, Z23, Z25, Z26, Z29, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z13, Z14, Z16, Z19, Z21, Z27, Z28, Z30, Z31, Z32, Z34
M4		Z1, Z15, Z17, Z20, Z23, Z30, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z11, Z12, Z13, Z14, Z16, Z18, Z19, Z21, Z25, Z26, Z27, Z28, Z29, Z31, Z32, Z34
M5	Z2, Z3, Z17, Z19,	Z1, Z6, Z7, Z12, Z15, Z18, Z20, Z21, Z23, Z30	Z4, Z5, Z8, Z9, Z10, Z11, Z13, Z14, Z16, Z29, Z31, Z32, Z33, Z34
M6	Z22, Z24, Z25, Z26,	Z1, Z7, Z11, Z12, Z15, Z18, Z29	Z4, Z5, Z6, Z8, Z9, Z10, Z13, Z14, Z16, Z20, Z21, Z23, Z30, Z31, Z32, Z33, Z34
M7	Z27, Z28	Z1, Z11, Z12, Z15, Z18, Z20, Z23, Z30, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z13, Z14, Z16, Z21, Z29, Z31, Z32, Z34
M8		Z1, Z12, Z15, Z18, Z20, Z30, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z11, Z13, Z14, Z16, Z21, Z23, Z29, Z31, Z32, Z34
M9	Z2, Z3, Z17, Z19,	Z1, Z11, Z12, Z15, Z18, Z20, Z25, Z29, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z13, Z14, Z16, Z21, Z30, Z31, Z32, Z34
M10	Z22, Z23, Z24, Z26,	Z1, Z15, Z18, Z30, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z11, Z12, Z13, Z14, Z16, Z20, Z21, Z25, Z29, Z31, Z32, Z34
M11	Z27, Z28	Z1, Z11, Z12, Z15, Z18, Z20, Z25, Z30, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z13, Z14, Z16, Z21, Z29, Z31, Z32, Z34
M12		Z1, Z12, Z16, Z29, Z33	Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z11, Z13, Z14, Z15, Z18, Z20, Z21, Z25, Z30, Z31, Z32, Z34

Źródło: opracowanie własne.

Analizując wyniki rankingów, można zauważyć, że część województw w większości rankingów zajmuje zbliżone miejsca. Przykładem może być województwo dolnośląskie, mające w większości rankingów miejsce pierwsze, czy województwo warmińsko-mazurskie pojawiające się zazwyczaj na ostatniej pozycji. Jest też grupa województw, które w zależności od rankingu przyjmują lokaty o zakresie zmiany aż ośmiu miejsc. Jednak są to województwa plasujące się w każdej klasyfikacji w grupie województw reprezentujących niższy poziom innowacyjności. Mimo rozbieżności pomiędzy wynikami rankingów w odniesieniu do utworzonych grup różnica jest niewielka – zazwyczaj jedynie przeskok na poziomie jednej grupy (tylko w przypadku województwa łódzkiego wystąpił przeskok o dwie grupy).

Tabela 4. Wyniki przeprowadzonych rankingów województw dla danych wybranych analizowanymi metodami

Województwo	Metoda doboru zmiennych											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Dolnośląskie	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	2 (1)	1 (1)	1 (1)	2 (1)	4 (1)	1 (1)	3 (2)
Kujawsko-pomorskie	7 (3)	8 (3)	8 (3)	11 (3)	8 (3)	9 (3)	11 (3)	10 (3)	8 (3)	12 (3)	8 (3)	11 (3)
Lubelskie	9 (3)	11 (3)	11 (3)	10 (3)	11 (3)	12 (3)	13 (3)	12 (3)	14 (3)	10 (3)	13 (3)	12 (3)
Lubuskie	15 (4)	13 (3)	15 (4)	15 (4)	15 (4)	15 (4)	15 (4)	15 (4)	15 (3)	15 (4)	15 (4)	13 (3)
Łódzkie	11 (3)	14 (4)	13 (3)	8 (3)	7 (2)	6 (2)	7 (2)	7 (2)	9 (3)	7 (2)	9 (3)	8 (2)
Małopolskie	4 (2)	6 (2)	3 (1)	3 (1)	5 (2)	7 (2)	4 (2)	4 (1)	5 (2)	2 (1)	6 (2)	7 (2)
Mazowieckie	2 (1)	3 (1)	2 (1)	5 (1)	2 (1)	4 (2)	2 (1)	6 (2)	3 (1)	6 (2)	2 (1)	4 (2)
Opolskie	8 (3)	9 (3)	12 (3)	13 (3)	9 (3)	11 (3)	10 (3)	8 (3)	11 (3)	11 (3)	11 (3)	9 (3)
Podkarpackie	10 (3)	10 (3)	7 (3)	6 (2)	12 (3)	10 (3)	9 (3)	11 (3)	12 (3)	13 (3)	12 (3)	5 (2)
Podlaskie	12 (3)	7 (2)	9 (3)	12 (3)	13 (3)	8 (3)	12 (3)	13 (3)	10 (3)	9 (3)	10 (3)	10 (3)
Pomorskie	6 (2)	2 (1)	6 (2)	7 (2)	6 (2)	1 (1)	5 (2)	2 (1)	1 (1)	3 (1)	3 (1)	1 (1)
Śląskie	3 (1)	5 (2)	4 (1)	2 (1)	3 (1)	5 (2)	3 (1)	5 (1)	6 (2)	5 (2)	5 (2)	6 (2)
Świętokrzyskie	13 (3)	12 (3)	14 (3)	14 (4)	14 (4)	13 (3)	14 (4)	14 (4)	13 (3)	16 (4)	14 (3)	15 (4)
Warmińsko-mazurskie	16 (4)	16 (4)	16 (4)	16 (4)	16 (4)	16 (4)	16 (4)	16 (4)	16 (4)	14 (4)	16 (4)	16 (4)
Wielkopolskie	5 (2)	4 (1)	5 (2)	4 (1)	4 (2)	3 (1)	6 (2)	3 (1)	4 (1)	1 (1)	4 (1)	2 (1)
Zachodnio-pomorskie	14 (3)	15 (4)	10 (3)	9 (3)	10 (3)	14 (3)	8 (3)	9 (3)	7 (3)	8 (3)	7 (3)	14 (3)

Źródło: opracowanie własne.

Tabela 5. Wartości współczynników rang Spearmana i indeksu Randa

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	Współczynnik rang Spearmana											
M1		0,894	0,909	0,844	0,932	0,838	0,865	0,850	0,803	0,741	0,844	0,806		Współczynnik rang Spearmana										
M2	0,700		0,876	0,715	0,788	0,879	0,753	0,779	0,818	0,706	0,824	0,850			Współczynnik rang Spearmana									
M3	0,958	0,708		0,909	0,865	0,809	0,891	0,812	0,856	0,774	0,888	0,806				Współczynnik rang Spearmana								
M4	0,800	0,683	0,825		0,882	0,797	0,929	0,862	0,800	0,826	0,838	0,806					Współczynnik rang Spearmana							
M5	0,850	0,700	0,808	0,817		0,865	0,950	0,932	0,894	0,844	0,935	0,794						Współczynnik rang Spearmana						
M6	0,833	0,706	0,875	0,750	0,833		0,847	0,885	0,885	0,832	0,882	0,924							Współczynnik rang Spearmana					
M7	0,850	0,700	0,808	0,817	1,000	0,833		0,926	0,909	0,844	0,935	0,821								Współczynnik rang Spearmana				
M8	0,800	0,717	0,792	0,833	0,900	0,850	0,900		0,926	0,906	0,921	0,856									Współczynnik rang Spearmana			
M9	0,850	0,800	0,858	0,733	0,717	0,800	0,800	0,817		0,871	0,985	0,803										Współczynnik rang Spearmana		
M10	0,817	0,717	0,792	0,817	0,917	0,833	0,917	0,950	0,733		0,868	0,771											Współczynnik rang Spearmana	
M11	0,925	0,775	0,933	0,792	0,775	0,875	0,775	0,792	0,925	0,792		0,800												Współczynnik rang Spearmana
M12	0,717	0,658	0,733	0,775	0,850	0,758	0,775	0,758	0,692	0,775	0,667													
Indeks Randa													Współczynnik rang Spearmana											

Źródło: opracowanie własne.

Chcąc dokonać wyboru rankingu, przeprowadzono porównanie pod kątem ich wzajemnej zgodności, wyliczając w tym celu współczynnik rang Spearmana (wyniki zaprezentowane w prawym górnym rogu tabeli 5). Porównano także zgodność otrzymanych klas województw. W tym celu wykorzystano indeks Randa (wyliczenia widoczne w lewym dolnym rogu tabeli 5).

Współczynniki rang Spearmana dla utworzonych rankingów były wysokie, przyjmowały wartości z zakresu od 0,706 do 0,985. Podobnie było w przypadku zgodności grup – indeks Randa wahał się od 0,658 do 1. Mimo zgodności widocznej w wyliczeniach obu miar można jednak zauważyć, że część rankingów zdecydowanie mocniej odstaje od pozostałych (np. wyniki doboru zmiennych metodą M2 czy M12), ale część jest w wysokim stopniu zgodna (np. ranking przeprowadzony metodą doboru M7 czy M8).

Oprócz powyżej wymienionych wskaźników do badania zależności pomiędzy otrzymanymi rankingami zastosowano analizę homogeniczności i heterogeniczności utworzonych grup oraz wyznaczono miernik poprawności grupowania (tabela 6), co zgodne jest z zaleceniami dostępnymi w literaturze, dotyczącymi konstrukcji kilku rankingów [Panek 2009, s. 160].

Tabela 6. Wyniki miar poprawności przeprowadzonego grupowania

Miary	Metoda doboru zmiennych											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Miara homogeniczności	2,15	1,92	1,99	1,61	2,06	1,61	1,97	1,68	1,74	1,33	1,78	1,17
Miara heterogeniczności	5,51	3,93	5,20	3,84	4,91	3,87	4,39	3,90	3,48	3,29	4,36	3,44
Miernik poprawności grupowania	0,39	0,49	0,38	0,42	0,42	0,42	0,45	0,43	0,50	0,40	0,41	0,34

Źródło: opracowanie własne.

Wyliczenia dotyczące poprawności wskazują, że dla analizowanych danych najlepszym rozwiązaniem jest wykorzystanie metody M12, gdyż dla niej wartość miernika jest najniższa. Najwyższą wartość, wskazującą na najniższy poziom poprawności przeprowadzonego grupowania, otrzymano dla miary M9. Zdecydowanie lepsze wyniki uzyskano w odniesieniu do pierwotnie zredukowanych danych za pomocą miar klasycznych i w kolejnym kroku (niezależnie od parametrycznej metody Hellwiga klasycznej czy zmodyfikowanej) przy wskaźniku $r = 0,7$. Z kolei w odniesieniu do miar pozycyjnych, badających zdolność dyskryminacyjną, sytuacja była odwrotna. Tu lepiej wypadło zastosowanie metody minimaksowej podczas wyznaczania wartości progowej.

4. Podsumowanie

Przystępując do opisanego w niniejszym artykule eksperymentu, autorka wychodziła z założenia, zgodnego z opiniami innych badaczy [Czech 2013; Młodak 2006], że przy zmiennych wykazujących asymetrię należy zastosować miary pozycyjne. Jednak przeprowadzone badania wykazały, że mimo iż miary pozycyjne dały lepsze wyniki w przypadku grupowania w aspekcie jego poprawności, to różnica pomiędzy nimi, a zastosowanymi miarami klasycznymi nie jest aż tak wyraźna, jakiej oczekiwano. Kolejne zaskoczenie w odniesieniu do uzyskanych wyników przyniosło spostrzeżenie, że stosowanie jednej grupy metod bazujących jedynie na średniej lub jedynie na medianie powinno dawać lepsze wyniki przeprowadzonego grupowania. Co interesujące, podczas porównania wskaźników poprawności grupowania, taka zależność nie była widoczna. Opisane w publikacji badanie może być rozszerzone o dodatkowe aspekty, m. in. uwzględnienie wpływu metod normalizacji zmiennych, jak też zastosowania innych niż euklidesowa odległości czy różnych form tworzenia i grupowania zmiennej syntetycznej. To zadanie dość obszernie, zwłaszcza biorąc pod uwagę, iż już na opisanym etapie wstępnym utworzono aż dwanaście miar. Nie wpływa to jednak na przekonanie autorki o potrzebie i chęci dalszych analiz w tym zakresie.

Literatura

- Bukowski M., Szpor A., Śniegocki A. (2012), *Potencjał i bariery polskiej innowacyjności*, Instytut Badań Strukturalnych, Warszawa.
- Czech A. (2013), *Zastosowanie wybranych metod doboru zmiennych diagnostycznych w badaniach konsumpcji w ujęciu pośrednim*, [w:] *Klasyfikacja i analiza danych – teoria i zastosowania*, K. Jajuga, M. Walesiak (red.), Taksonomia 21, Prace Naukowe UE we Wrocławiu nr 279, Wydawnictwo UE, Wrocław, s. 246-254.
- Jarocka M. (2013), *Wpływ metod doboru cech diagnostycznych na wyniki porządkowania liniowego na przykładzie rankingu polskich uczelni*, [w:] *Klasyfikacja i analiza danych – teoria i zastosowania*, K. Jajuga, M. Walesiak (red.), Taksonomia 21, Prace Naukowe UE we Wrocławiu nr 279, Wydawnictwo UE, Wrocław, s. 85-93.
- Mahdjoubi D. (1997), *Schumpeterian Economics and the Trilog of 'Invention-Innovation-Diffusion'*, www.ischool.utexas.edu/~darius/17-Schumpeter-innovation.pdf.
- Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, „Difin”, Warszawa.
- Nazarko J. (2013), *Regionalny foresight gospodarczy. Scenariusze rozwoju innowacyjności mazowieckich przedsiębiorstw*, Związek Pracodawców Warszawy i Mazowska, Warszawa.
- Nowakowska A. (red.) (2009), *Zdolności innowacyjne polskich regionów*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- OECD (2005), *Guidelines for Collecting and Interpreting Technological Innovation Data*, OECD/European Communities, Oslo Manual, Paris.
- Panek T. (2009), *Statystyczne metody wielowymiarowej analizy porównawczej*, Szkoła Główna Handlowa w Warszawie, Warszawa.

- Pociecha J. (2008), *Rozwój metod taksonomicznych i ich zastosowań w badaniach społeczno-ekonomicznych*, http://www.stat.gov.pl/gus/5840_4193_PLK_HTML.htm.
- Prystrom J. (2012), *Innowacje w procesie rozwoju gospodarczego. Istota i uwarunkowania*, Difin, Warszawa.
- Schumpeter J. (1947), *The Creative Response in Economic History*, „The Journal of Economic History”, t. 7, nr 2, s. 151.
- Tarka D. (2013), *Wpływ metody doboru cech diagnostycznych na wyniki klasyfikacji obiektów na przykładzie danych dotyczących ochrony środowiska*, [w:] *Klasyfikacja i analiza danych – teoria i zastosowania*, K. Jajuga, M. Walesiak (red.), *Taksonomia 21, Prace Naukowe UE we Wrocławiu nr 279*, Wydawnictwo UE, Wrocław, s. 235-245.

THE APPLICATION OF SELECTED QUANTITATIVE METHODS TO THE EVALUATION OF VOIVODESHIP INNOVATION LEVEL POTENTIAL

Summary: Innovation is increasingly becoming a determinant of organizational and regional development. It is a complex notion embracing various processes in the area of social and economic development as well as the activities of enterprises, regional policy makers and the development of R&D sector. The complexity of the described notion makes its measurement potentially difficult. The measurement is also subject to numerous limitations which often arise from data accessibility. Therefore, the right choice of available variables seems to be significant in that context. The main aim of this article is to develop a classification of voivodeships according to their innovation level. The aim has been achieved through conducting an analysis of the influence of variable selection methods on research results. With respect to that choice, object classification has been carried out, emphasizing changes occurring in the results of the classification with consideration of different variable selection methods.

Keywords: innovation, variable selection methods, object classification.