

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 Dębkowska, 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

Monika Rozkrut

Uniwersytet Szczeciński

Dominik Rozkrut

Urząd Statystyczny w Szczecinie, Uniwersytet Szczeciński

IDENTYFIKACJA STRATEGII INNOWACYJNYCH PRZEDSIĘBIORSTW USŁUGOWYCH W POLSCE

Streszczenie: Innowacyjność jest zjawiskiem wielowymiarowym, stąd też wykorzystanie metod klasyfikacji i analizy danych zwiększa możliwości opisu, umożliwiając dokładniejszy wgląd w naturę procesów innowacji. W artykule przetestowano przydatność analizy skupień oraz analizy czynnikowej do poszukiwania prawidłowości w zakresie zachowań innowacyjnych przedsiębiorstw sektora usług. Kształtowanie polityki gospodarczej skierowanej na wzrost konkurencyjności polskich przedsiębiorstw wymaga wykorzystywania metod adekwatnie opisujących skomplikowane aspekty działalności innowacyjnej.

Słowa kluczowe: analiza czynnikowa, analiza skupień, innowacyjność.

1. Wstęp

Nie ulega wątpliwości, że innowacje w sektorze usług są coraz istotniejszym czynnikiem jego rozwoju, niemniej zróżnicowanie podejmowanych działań innowacyjnych jest znaczne, zwłaszcza w zakresie źródeł dla innowacji, przy czym już wcześniejsze badania przewidywały systematyczny wzrost znaczenia nowoczesnych technologii dla innowacji w usługach [Evangelista, Sirilli 1998]. Wziąwszy pod uwagę znaczenie sektora usług dla rozwoju gospodarczego, niniejsze badanie ma na celu dostarczenie wiedzy na temat natury procesów innowacyjnych w przedsiębiorstwach usługowych. Innowacyjność jest zjawiskiem wielowymiarowym, stąd też wykorzystanie metod klasyfikacji i analizy danych zwiększa możliwości opisu, umożliwiając dokładniejszy wgląd w naturę procesów innowacji. W artykule starano się z jednej strony zweryfikować możliwości wykorzystania analizy czynnikowej i metod analizy skupień do badania strategii innowacyjnych przedsiębiorstw z sektora usług w Polsce, z drugiej zaś w przypadku pozytywnej weryfikacji użyteczności proponowanej procedury celem badania była identyfikacja i ocena zróżnicowania strategii innowacyjnych przedsiębiorstw usługowych.

2. Metoda badawcza

Celem eksploracyjnej analizy czynnikowej jest identyfikacja ukrytej struktury analizowanych zmiennych. W ramach analizy redukuje się liczbę zmiennych poprzez wykorzystywanie ich liniowych kombinacji (czynników), które w założeniu reprezentują znaczną część informacji zawartych w zmienności zmiennych pierwotnych. Zmienne te są naturalnie skorelowane i korelacje te znajdują odzwierciedlenie w wartościach ładunków czynnikowych, tak że zmienne, które wykazują wysokie wartości ładunków czynnikowych względem jednego z wyodrębnionych czynników, są w dużej mierze niezależne od pozostałych czynników. Przegląd zagadnień analizy czynnikowej zawarto w pracy [Sztemberg-Lewandowska 2008].

W celu redukcji zmiennych oraz identyfikacji ich ukrytej struktury, tj. w szczególności do znalezienia charakterystycznych zachowań w związku z wewnętrznymi procesami mającymi miejsce w przedsiębiorstwach, związanymi z ich innowacyjnością, proponuje się wykorzystanie analizy czynnikowej [por. Wziątek-Kubiak i in. 2009]. Te charakterystyczne zachowania nazywane są w niniejszym artykule praktykami innowacyjnymi. Zaletą analizy czynnikowej w tym kontekście jest fakt, że w jej wyniku obliczyć można dla każdego obiektu tzw. wartości czynnikowe, co pozwala dokonywać dalszych analiz. W niniejszym badaniu ostateczne czynniki reprezentujące wyodrębnione praktyki innowacyjne wykorzystywane są, poprzez wartości czynnikowe, w analizie skupień metodą k -średnich. Wykorzystanie analizy skupień pozwoliło na identyfikację podejmowanych przez przedsiębiorstwa strategii innowacyjnych.

Celem analizy skupień jest grupowanie obiektów w możliwie jednorodne skupienia (grupy), składające się z możliwie podobnych obiektów, jednocześnie możliwie różnych pomiędzy skupieniami. Podobieństwo obiektów mierzyć można z wykorzystaniem różnych miar [więcej np. Walesiak 1996]. Oczekuje się, że otrzymane skupienia reprezentują możliwe do zinterpretowania naturalne grupy obiektów. Istnieją różne podejścia do problemu wyodrębniania skupisk. Algorytmy aglomeracyjne łączą w kolejnych iteracjach obiekty w większe skupiska, traktując na początku każdy obiekt jako osobną grupę i następnie łącząc najbardziej podobne do siebie grupy. Proces ten jest kontynuowany do momentu połączenia wszystkich obiektów w jednym skupisku lub do otrzymania zadanej liczby skupień. Zaletą metod aglomeracyjnych jest niewątpliwie możliwość prezentacji wyników aglomeracji w postaci dendrogramu.

Inaczej niż algorytmy aglomeracyjne, algorytmy niehierarchiczne przyporządkowują obiekty do zadanej liczby skupień bezpośrednio, nie bazując na iteracyjnym łączeniu obiektów w coraz większe skupiska. Najbardziej popularnym niehierarchicznym algorytmem analizy skupień jest algorytm k -średnich [Hartigan 1975]. W metodzie tej skupienia dobiera się tak, by zminimalizować zmienność wewnątrzgrupową. Główną ideą tego algorytmu jest taka alokacja obiektów, która

minimalizuje zmienność wewnątrz powstałych skupień, a co za tym idzie – maksymalizuje zmienność pomiędzy skupieniami. Poza wymienionymi istnieją także inne metody analizy skupień, jak np. odpowiadająca metodzie k -średnich, metoda k -medoid, sieci samoorganizujące się (SOM), uogólniona analiza skupień w oparciu o mieszaniny rozkładów (algorytm EM).

Warto zaznaczyć, że niezależnie od przyjętej metody, z perspektywy jakości uzyskanych grupowań zawsze istotny jest odpowiedni dobór zmiennych do analizy. Jakość procedury grupowania zależy w oczywisty sposób od ich doboru. Postuluje się, by wyselekcjonowane zmienne były możliwie dobrymi reprezentantami poszukiwanej typologii, patrz np. [Everitt i in. 2011].

Zasadniczym problemem w przypadku analizy skupień jest określenie liczby grup, tak by właściwie reprezentowały one analizowaną zbiorowość. Istnieje wiele sposobów określenia tej liczby, wśród nich indeks Calińskiego-Harabasz, indeks Dudy-Harta, indeks pseudo-F, indeks Krzanowskiego-Lai [1988], indeks Hartigana [1975] (więcej patrz np. [Walesiak, Dudek 2007]). Ciekawą metodę przedstawił także Schonlau [2004], proponując budowę wykresu ilustrującego proces przypisywania obiektów do skupisk dla kolejnych rosnących wariantów liczby skupień. Wykres ten można wykorzystywać w przypadku niehierarchicznej analizy skupień, jak również do analizy aglomeracyjnej, kiedy zbyt duża liczba obiektów czyni korzystanie z dendrogramu niepraktycznym. W literaturze wielokrotnie testowano różne metody określenia liczby skupień (patrz np. [Milligan, Cooper 1985]). Niektóre z tych metod mają zastosowanie jedynie w przypadku hierarchicznej analizy skupień, inne zaś mają charakter uniwersalny. Miarą taką jest m.in. indeks Calińskiego-Harabasz, którego wysokie wartości oznaczają wyraźniejszą odmienność skupień.

Dane wykorzystane w przedstawionych analizach pochodzą z badania działalności innowacyjnej przedsiębiorstw, prowadzonego w ramach Programu Badań Statystycznych Statystyki Publicznej przez służby statystyki publicznej. Badanie to jest elementem szerszych badań, prowadzonych w cyklu dwuletnim w krajach Unii Europejskiej, pod nazwą *Community Innovation Survey*. Wyniki tych badań stanowią nieocenione źródło zharmonizowanej informacji statystycznej odnośnie do innowacyjności przedsiębiorstw w Unii i jej krajach członkowskich. W Polsce badania te, ze względu na swą wagę i duże zapotrzebowanie odbiorców, prowadzone są w cyklu rocznym, w przemyśle i usługach. Wyniki badania wykorzystane w niniejszym artykule odnoszą się do przedsiębiorstw usługowych, badanych w 2011 r. z okresem sprawozdawczym 2008-2010 (pytania dotyczące działalności innowacyjnej odnoszą się do okresów trzyletnich). Badanie niniejsze ograniczono do zbioru podmiotów aktywnych innowacyjnie, tj. takich, które w analizowanym okresie wprowadziły innowacje produktowe lub procesowe, lub realizowały projekty innowacyjne mające w zamierzeniu doprowadzić do innowacji produktowych lub procesowych. Badaniu podlegają podmioty o liczbie pracujących powyżej 9. Przyjmując metodę badawczą opartą na eksploracyjnej analizie czynnikowej,

w badaniach wykorzystano możliwie szeroki, dostępny zakres zmiennych, charakteryzujących działalność innowacyjną przedsiębiorstw. Zestawienie analizowanych zmiennych zawierają tabele 1-5, przy czym dla zwiększenia czytelności tabeli, wykorzystywane są ich skrócone nazwy. Zmienne te można zestawić w jednolite grupy zmiennych o podobnym charakterze. Pierwszą grupę (patrz tab. 1) stanowiły zmienne o charakterze wkładu dla działalności innowacyjnej, w tym wskaźniki wprowadzania nowych technologii (wewnętrzne i zewnętrzne nakłady na B+R, maszyny, urządzenia, oprogramowanie; licencjonowanie), ale także inne wskaźniki, takie jak ponoszenie wydatków na szkolenia, na wprowadzenie innowacji na rynek, podejmowanie działań związanych z zaprojektowaniem lub poprawą designu, wyglądu nowych produktów, *feasibility studies*, testowanie, uzyskanie wsparcie publicznego czy współpraca w działalności innowacyjnej. Drugą grupę stanowiły zmienne o charakterze efektów, świadczące o rodzajach efektów działań innowacyjnych (poszczególne przykłady innowacji produktowych, procesowych, organizacyjnych, marketingowych). Ponadto analizie poddano wskaźniki dotyczące źródeł wiedzy, umiejętności i metod stymulowania innowacji. W pierwszej z wymienionych znalazły się: źródła wewnętrzne, dostawcy, klienci, konkurenci, konsultanci, laboratoria, prywatne i publiczne instytucje, szkolnictwo wyższe, konferencje, targi, wystawy, czasopisma i publikacje naukowe, stowarzyszenia naukowe-techniczne i zawodowe. Zmienne dotyczące umiejętności dla innowacji to: grafika, reklama, projektowanie obiektów lub usług, multimedia, projektowanie stron internetowych; rozwój oprogramowania, badania rynkowe, inżynieria, nauki stosowane; matematyka, statystyka, zarządzanie bazami danych. Wśród wyróżnionych metod stymulowania innowacji znalazły się: burze mózgów, zespoły robocze, rotacja pracowników, zachęty finansowe i niefinansowe, szkolenia ze sposobów wypracowywania nowych pomysłów.

3. Analiza wyników

Ponieważ analizy czynnikowe oparto na zmiennych binarnych, do konstrukcji macierzy korelacji wykorzystano współczynniki korelacji tetrachorycznej, szacowanej z wykorzystaniem metody największej wiarygodności [Olsson 1979]. Analiza czynnikowa przeprowadzana była etapami¹. W pierwszym kroku analizie poddano jednolite grupy zmiennych związanych z działalnością innowacyjną analizowanych przedsiębiorstw. Wyniki otrzymanych analiz zaprezentowano w tab. 1-6. Przyjętą metodą rotacji była metoda *oblimin oblique*. Liczbę czynników dobierano opierając się na kryterium własności własnych.

W pierwszej kolejności analizowano zmienne o charakterze wkładu. Uzyskano cztery czynniki (tab. 1). Pierwszy z nich reprezentuje wewnętrzne prace rozwojo-

¹ Prezentowane w artykule obliczenia wykonane zostały przy wykorzystaniu pakietu Stata (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP).

we, odzwierciedlając tradycyjne rozumienie procesów innowacyjnych poprzez nakłady na B+R, design, testowanie, *feasibility studies*. Drugi reprezentuje tworzenie warunków dla innowacji, obejmując takie działania, jak zakupy maszyn i urządzeń, oprogramowania, licencjonowanie, szkolenia, wydatki na wprowadzanie innowacji na rynek. Zmienne świadczące o otrzymaniu pomocy finansowej z programów wsparcia innowacyjności wyraźnie koncentrują się w czynniku trzecim. Ostatni reprezentuje zakupy B+R ze źródeł zewnętrznych i współpracę w działalności innowacyjnej.

Tabela 1. Ładunki czynnikowe dla zmiennych określających wkład do innowacji

Zmienna	Czynniki związane z nakładami			
	czynnik 1	czynnik 2	czynnik 3	czynnik 4
Ponoszono nakłady na B+R	0,930			
Systematycznie ponoszono nakłady na B+R	0,972			
Zakupowano B+R ze źródeł zewnętrznych				0,548
Zakupowano maszyny, urządzenia i oprogramowanie		0,545		
Zakupowano lub licencjonowano patenty lub inny know-how		0,696		
Ponoszono wydatki na szkolenia		0,804		
Ponoszono wydatki na wprowadzenie innowacji na rynek		0,554		
Podejmowano działania związane z designem nowych produktów	0,693			
Podejmowano inne działania, np. <i>feasibility studies</i> , testowanie	0,734			
Otrzymano wsparcie publiczne z administracji lokalnej			0,743	
Otrzymano wsparcie publiczne z administracji rządowej				
Otrzymano wsparcie publiczne z UE			0,929	
Otrzymano wsparcie publiczne z 6. lub 7. Programu ramowego			0,807	
Współpracowano w działalności innowacyjnej				0,764

Źródło: obliczenia własne.

Analiza zestawu zmiennych dotyczących efektów innowacji (tab. 2) ujawnia trzy łatwe w interpretacji czynniki. Pierwszy związany jest z wprowadzaniem zmian designu, opakowania, metod sprzedaży, dystrybucji, promocji, kalkulacji cen, czyli innowacji marketingowych. Drugi czynnik koncentruje innowacje produktowe i procesowe, tj. wprowadzenie nowego wyrobu, usługi czy też nowej metody produkcji. Trzeci natomiast dotyczy zmian w zakresie systemów, organizacji procesów biznesowych, organizacji miejsc pracy, zmian w relacjach z innymi organizacjami, czyli innowacji organizacyjnych.

W przypadku zmiennych określających pożądane umiejętności dla innowacji wyraźnie wyodrębniły się dwa czynniki (tab. 3), jeden związany z umiejętnością-

Tabela 2. Ładunki czynnikowe dla zmiennych określających efekty innowacji

Zmienna	Czynniki związane z efektami		
	czynnik 1	czynnik 2	czynnik 3
Wprowadzono nowy wyrób		0,535	
Wprowadzono nową usługę		0,734	
Wprowadzony produkt był nowy na rynku		0,852	
Wprowadzono nową metodę produkcji		0,679	
Wprowadzono nową metodę w zakresie logistyki			
Wprowadzono nową metodę w zakresie systemów			0,694
Wprowadzono zmiany w organizacji procesów biznesowych			0,695
Wprowadzono zmiany organizacji miejsc pracy			0,678
Wprowadzono zmiany w relacjach z innymi organizacjami			0,601
Wprowadzono zmiany designu lub opakowania	0,717		
Wprowadzono zmiany w metodach sprzedaży i dystrybucji	0,860		
Wprowadzono zmiany w metodach promocji	0,874		
Wprowadzono zmiany metod kalkulacji cen	0,820		

Źródło: obliczenia własne.

mi, które można by określić mianem zaawansowanych (inżynieria, nauki stosowane, matematyka, statystyka, zarządzanie bazami danych), drugi z pozostałymi wskazanymi umiejętnościami, jeśli nawet nie mniej skomplikowanymi, to z pewnością nieco bardziej powszechnymi (grafika, reklama, projektowanie stron internetowych, multimedia, oprogramowanie, badania rynkowe).

Tabela 3. Ładunki czynnikowe dla zmiennych określających umiejętności dla innowacji

Zmienna	Czynniki związane z umiejętnościami	
	czynnik 1	czynnik 2
Grafika, reklama	0,964	
Projektowanie obiektów lub usług		
Multimedia	0,878	
Projektowanie stron internetowych	0,974	
Rozwój oprogramowania	0,707	
Badania rynkowe	0,611	
Inżynieria, nauki stosowane		0,974
Matematyka, statystyka, zarządzanie bazami danych		0,662

Źródło: obliczenia własne.

Metody stymulowania koncentrowały się w dwóch czynnikach (tab. 4), pierwszym obejmującym burze mózgow i współpracę w ramach grup roboczych, i drugim, łączącym zachęty dla pracowników ze szkoleniami z zakresu wypracowywania nowych pomysłów. Podobnie, w przypadku wykorzystywanych źródeł, wyod-

rębnily się dwa czynniki, jeden związany ze środowiskiem przedsiębiorstwa, łączący źródła wewnętrzne z informacjami od klientów i konkurencji, oraz drugi odnoszący się do środowiska badawczego, tj. konsultantów, prywatnych i państwowych instytucji badawczych, szkolnictwa wyższego (tab. 5).

Tabela 4. Ładunki czynnikowe dla zmiennych określających metody stymulowania innowacyjności

Zmienna	Czynniki związane ze stymulowaniem	
	czynnik 1	czynnik 2
Burze mózgów		0,849
Multidyscyplinarne lub międzyfunkcyjne zespoły robocze		0,937
Wewnętrzna rotacja pracowników		
Finansowe zachęty dla pracowników	0,899	
Niefinansowe zachęty dla pracowników	0,716	
Szkolenia z zakresu sposobów wypracowywania nowych pomysłów	0,885	

Źródło: obliczenia własne.

Tabela 5. Ładunki czynnikowe dla zmiennych określających źródła innowacyjności

Zmienna	Czynniki związane ze źródłami	
	czynnik 1	czynnik 2
Źródła wewnętrzne (przedsiębiorstwo lub grupa)		0,565
Dostawcy materiałów, komponentów, oprogramowania		
Klienci		0,880
Konkurenci		0,849
Konsultanci, laboratoria, prywatne instytucje badawcze	0,764	
Szkolnictwo wyższe	1,001	
Instytucje rządowe, w tym publiczne instytucje badawcze	0,912	
Konferencje, targi, wystawy		
Czasopisma i publikacje naukowe		
Stowarzyszenia naukowe-techniczne i zawodowe	0,855	

Źródło: obliczenia własne.

W następnym kroku dokonano ponownie analizy czynnikowej, tym razem jednak na czynnikach uzyskanych w pierwszym etapie na podstawie zestawów jednorodnych zmiennych. Wyniki tej analizy zestawiono w tabeli 6. Warto zauważyć, że takie podejście zaowocowało procedurą, którą można byłoby określić hierarchiczną analizą czynnikową. Ograniczenie pierwotnych zestawów do zmiennych o jednorodnym charakterze pozwoliło wyodrębnić łatwe w interpretacji czynniki, które ostatecznie posłużyły do identyfikacji strategii innowacyjnych.

Pierwsza z wyodrębnionych praktyk innowacyjnych opiera się na wprowadzaniu innowacji marketingowych i organizacyjnych, które tworzone są na kanwie przede wszystkim umiejętności miękkich, pobudzanych za pomocą całego spek-

Tabela 6. Ładunki czynnikowe dla praktyk innowacyjnych

Zmienna	Czynnik 1	Czynnik 2	Czynnik 3
Nakłady 1 („prace rozwojowe”)		0,627	
Nakłady 2 („warunki dla innowacji”)			
Nakłady 3 („wsparcie finansowe”)			-0,769
Nakłady 4 („zakup B+R i współpraca”)			0,679
Efekty 1 („innowacje marketingowe”)	0,501		
Efekty 2 („innowacje produktowe, procesowe”)		0,704	
Efekty 3 („innowacje organizacyjne”)	0,737		
Umiejętności 1 („pozostałe”)	0,646		
Umiejętności 2 („ściśle”)			
Metody stymulacji 1 („współpraca wewnętrzna”)	0,615		
Metody stymulacji 2 („zachęty i szkolenia”)	0,584		
Źródła dla innowacji 1 („środowisko przedsiębiorcy”)		0,596	
Źródła dla innowacji 2 („środowisko badawcze”)		0,643	

Źródło: obliczenia własne.

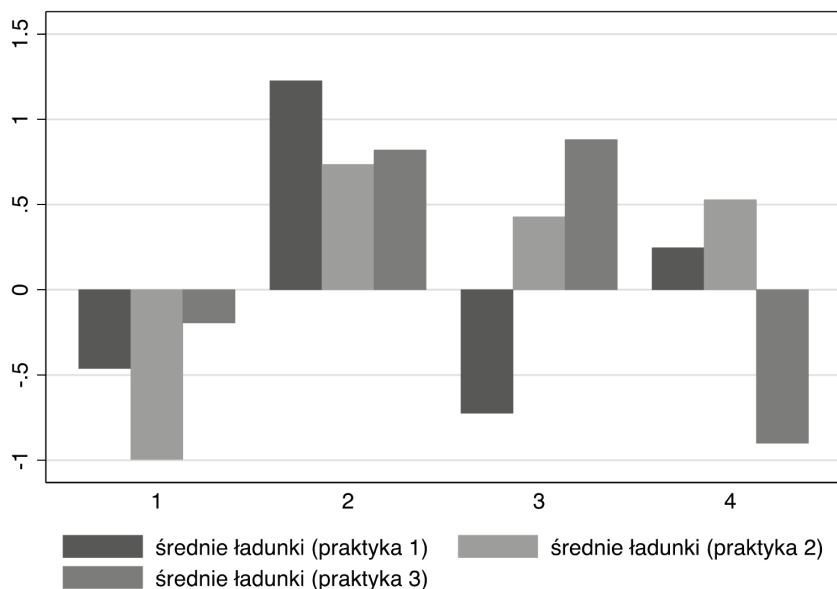
trum metod stymulowania działalności innowacyjnej, obejmujących pobudzanie wewnętrznej współpracy, połączonych z systemami zachęt pracowników natury finansowej i niefinansowej.

Druga wyodrębniona praktyka działań innowacyjnych charakteryzuje innowatorów, których można by określić mianem „twardych”, opierających swoje działania innowacyjne na pracach badawczo-rozwojowych, znajdujących efekt w postaci innowacji produktowych i procesowych. Charakterystyczne dla tej praktyki jest silne wykorzystanie wszelkich dostępnych źródeł wiedzy dla działań innowacyjnych, płynących zarówno ze środowiska przedsiębiorcy, jak i ze środowiska badawczego.

Trzecia wyodrębniona praktyka jest efektem racjonalnych zachowań rynkowych. Badania jednoznacznie bowiem wskazują (tab. 6), że brak dofinansowania jest czynnikiem sprzyjającym nawiązywaniu współpracy w zakresie działalności innowacyjnej. Ładunki czynnikowe wskazują na istotną dodatnią korelację tego czynnika z zaangażowaniem we współpracę w zakresie działalności innowacyjnej i jednocześnie istotną ujemną korelacją z otrzymywaniem zewnętrznego wsparcia finansowego (z programów wsparcia działalności innowacyjnej).

W celu opracowania typologii strategii innowacyjnych przedsiębiorstw w oparciu o rozpoznane praktyki innowacyjne wykorzystano procedury grupowania. Metody analizy skupień umożliwiają grupowanie podobnych obiektów, pozwalając na poznanie struktury danych. Jakkolwiek metody grupowania hierarchicznego są wygodnym narzędziem wizualizacji podobieństwa między obiektami, mają one jednak tendencję do wyodrębniania skupisk o niewielkiej liczebności. Dobrym przykładem jest metoda Warda – relatywnie efektywna w odkrywaniu ukrytej ty-

pologii. Metoda ta oparta jest na koncepcji analizy wariancji, w której dąży się do minimalizacji sumy kwadratów odchyłeń wewnątrz potencjalnych skupień (a więc suma kwadratów odchyłeń wewnątrz skupień jest miarą niepodobieństwa pomiędzy skupieniami). To pożądana własność w sytuacji, w której faktycznie oczekujemy niskiej wartości sumy kwadratów. Okazuje się jednak, że na danym etapie aglomeracji metodą Warda otrzymywana wartość sumy kwadratów dla danej liczby grup może być większa niż uzyskiwana dla tej samej liczby grup metodą k -średnich.



Rys. 1. Porównanie średnich dla czynników końcowych (praktyk innowacyjnych) w otrzymanych grupowaniach

Źródło: obliczenia własne.

W badaniu wykorzystano więc metodę k -średnich, niemniej centroidy otrzymane w efekcie aglomeracji metodą Warda dla określonej liczby grup wykorzystano jako punkt startowy do ostatecznego grupowania z wykorzystaniem metody k -średnich. Celem metody k -średnich jest wyodrębnienie k możliwie odmiennych grup. W tym przypadku należy jednak *a priori* określić liczbę pożądanых grup. Istnieją różne metody doboru liczby klastrów. Jedną z relatywnie skutecznych jest wykorzystanie indeksu Calińskiego i Harabasz [Caliński, Harabasz 1974; Milligan, Cooper 1985]. W badaniu posłużono się tym indeksem. Dodatkowo, aby potwierdzić wybór w oparciu o indeks, przeprowadzono inspekcję clustergramów [Schonlau 2004].

4. Podsumowanie

Analizując wyniki grupowania, sprawdza się zwykle średnie w grupach dla każdej uwzględnionej w badaniu zmiennej w celu oceny odmienności otrzymanych skupisk. Wyniki takiej analizy zaprezentowano na rys. 1. Analiza tego wykresu daje ciekawy obraz innowacyjności przedsiębiorstw sektora usług, szczególnie w przypadku grupy 3. i 4., zważywszy na składowe rozpatrywane praktyki. Pierwsza z ujawnionych strategii innowacyjnych podmiotów cechuje się relatywnie niższym poziomem zaangażowania w działania innowacyjne i w znacznej mierze charakteryzuje przedsiębiorstwa aktywne innowacyjnie, ale nieprzejawiające efektów w tym zakresie (podejmujące działania bez efektu w postaci wprowadzenia innowacji). Druga zidentyfikowana strategia to liderzy innowacji, charakteryzujący się intensywną działalnością innowacyjną zarówno w zakresie innowacji technologicznych, jak i nietechnologicznych, opierający swe działania na współpracy z innymi podmiotami. Trzecia grupa podmiotów to innowatorzy koncentrujący się na produktach i procesach, aktywni we współpracy, ale niekorzystający ze wsparcia, co potencjalnie ogranicza uzyskiwane efekty. Czwarta strategia innowacyjna oparta jest na silnej koncentracji na wykorzystaniu wsparcia dla innowacji, które, co niezmiernie istotne, okazuje się substytutem współpracy w zakresie działalności innowacyjnej. Metody analiz czynnikowych i grupowania zdecydowanie ułatwiają prowadzenie analiz takich wieloaspektowych zjawisk, jak działalność innowacyjna przedsiębiorstw, umożliwiając wgląd w strukturę wielowymiarowych obrazów, niemożliwą do wychwycenia poprzez prostą analizę powszechnie wykorzystywanych wskaźników. Uzyskane wyniki wyraźnie potwierdzają przydatność wykorzystanych metod do identyfikacji strategii innowacyjnych podmiotów gospodarczych z sektora usług. Postulowana w niniejszym badaniu potrzeba identyfikacji strategii innowacyjnych przedsiębiorstw usługowych ujawnia ponadto bardzo istotne, z punktu widzenia potrzeb kształtowania polityki gospodarczej, prawidłowości w zakresie działalności innowacyjnej przedsiębiorstw w Polsce.

Literatura

- Caliński T., Harabasz J.(1974), *A dendrite method for cluster analysis*, „Communications in Statistics” 3 (1).
- Duda R.O., Hart P.E., Stork D.G. (2012), *Pattern Classification*, John Wiley & Sons, New York.
- Evangelista R., Sirilli G. (1998), *Innovation in the Service Sector Results from the Italian Statistical Survey*, „Technological Forecasting and Social Change”, 58.
- Everitt B.S., Landau S., Leese M., Stahl D. (2011), *Cluster Analysis*, John Wiley & Sons, New York.
- Hartigan J.A. (1975), *Clustering algorithms*, Wiley, New York.
- Hollenstein H. (2003), *Innovation modes in the Swiss service sector: A cluster analysis based on firm-level data*, „Res. Policy”, 32.
- Jajuga K. (1993), *Statystyczna analiza wielowymiarowa*, Biblioteka Ekonometryczna. Wyd. Naukowe PWN, Warszawa.

- Krzanowski W.J., Lai Y.T. (1985), *A criterion for determining the number of groups in a data set using sum of squares clustering*. Biometrics, No 44.
- Milligan G.W., Cooper M.C. (1985). *An examination of procedures for determining the number of clusters in a data set*. Psychometrika, 50(2), 159-179.
- Olsson U. (1979), *Maximum Likelihood Estimation of the Polychoric Correlation*. Psychometrika, 44.
- Pociecha J., Podolec B., Sokołowski A., Zając K. (1998), *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa.
- Schonlau M. (2004), *Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams*, Computational Statistics 19, no. 1 (2004): 95-111.
- Sztemberg-Lewandowska M. (2008), *Analiza czynnikowa w badaniach marketingowych*, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław.
- Walesiak M. (1996), *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa.
- Walesiak M., Dudek A. (2007), *clusterSim: Searching for optimal clustering procedure for a data set*. URL <http://www.R-project.org>. R package version 0.30- 7.
- Walesiak M. (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Wziątek-Kubiak A., Balcerowicz E., Pęczkowski M. (2009), *Differentiation of Innovation Behavior of Manufacturing Firms in the New MemberStates. Cluster analysis on Firm-Level Data*, CASE Network Studies & Analyses, No. 394.

IDENTIFICATION OF SERVICE SECTOR INNOVATION STRATEGIES IN POLAND

Summary: Innovation is a multidimensional phenomenon; hence the use of classification and data analysis methods increases the scope of description, enabling a more detailed insight into the nature of innovation processes. The paper tested the usefulness of cluster analysis and factor analysis to look for patterns in the behavior of innovative business services sector. The formation of economic policy aimed at increasing the competitiveness of Polish enterprises requires the use of methods adequately depicting complicated aspects of innovation activities.

Keywords: innovativeness, clustering, factor analysis.