

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 327

Taksonomia 22

**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	9
Eugeniusz Gatnar , Balance of payments statistics and external competitiveness of Poland.....	15
Andrzej Sokolowski, Magdalena Czaja , Efektywność metody k -średnich w zależności od separowalności grup.....	23
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Wielosektorowa analiza ukrytych przejść w modelowaniu zagrożenia upadłością polskich przedsiębiorstw	30
Elżbieta Gołata , Zróżnicowanie procesu starzenia i struktur demograficznych w Poznaniu i aglomeracji poznańskiej na tle wybranych dużych miast Polski w latach 2002-2011.....	39
Aleksandra Łuczak, Feliks Wysocki , Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów	49
Marek Walesiak , Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej	60
Paweł Lula , Identyfikacja słów i fraz kluczowych w tekstach polskojęzycznych za pomocą algorytmu <i>RAKE</i>	69
Mariusz Kubus , Propozycja modyfikacji metody złagodzonego LASSO.....	77
Andrzej Bąk, Tomasz Bartłomowicz , Wielomianowe modele logitowe wyborów dyskretnych i ich implementacja w pakiecie <i>DiscreteChoice</i> programu R.....	85
Justyna Brzezińska , Wykorzystanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004-2012.....	95
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Zastosowanie pakietu <i>dcMNM</i> programu R w badaniach preferencji konsumentów wódki	104
Barbara Batóg, Jacek Batóg , Analiza stabilności klasyfikacji polskich województw według sektorowej wydajności pracy w latach 2002-2010	113
Małgorzata Markowska, Danuta Strahl , Klasyfikacja europejskiej przestrzeni regionalnej ze względu na filary inteligentnego rozwoju z wykorzystaniem referencyjnego systemu granicznego.....	121
Kamila Migdał-Najman, Krzysztof Najman , Formalna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	131
Kamila Migdał-Najman, Krzysztof Najman , Graficzna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	139
Beata Basiura, Anna Czapkiewicz , Badanie jakości klasyfikacji szeregów czasowych	148
Michał Trzęsiok , Wybrane metody identyfikacji obserwacji oddalonych.....	157

Grażyna Dehnel, Tomasz Klimanek , Taksonomiczne aspekty estymacji pośredniej uwzględniającej autokorelację przestrzenną w statystyce gospodarczej.....	167
Michał Bernard Pietrzak, Justyna Wilk , Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji.....	177
Maciej Beręsewicz , Próba zastosowania różnych miar odległości w uogólnionym estymatorze Petersena.....	186
Marcin Szymkowiak, Tomasz Józefowski , Konstrukcja i praktyczne wykorzystanie estymatorów typu SPREE na przykładzie dwuwymiarowych tabel kontyngencji.....	195
Marcin Pelka , Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym.....	202
Małgorzata Machowska-Szewczyk , Ocena klas w rozmytej klasyfikacji obiektów symbolicznych.....	210
Justyna Wilk , Problem wyboru liczby klas w taksonomicznej analizie danych symbolicznych.....	220
Andrzej Dudek , Metody analizy skupień w klasyfikacji markerów map Google.....	229
Ewa Roszkowska , Ocena ofert negocjacyjnych w słabo ustrukturyzowanych problemach negocjacyjnych z wykorzystaniem rozmytej procedury SAW.....	237
Marcin Szymkowiak, Marek Witkowski , Zastosowanie analizy korespondencji do badania kondycji finansowej banków spółdzielczych.....	248
Bartłomiej Jefmański , Budowa rozmytych indeksów satysfakcji klientów z zastosowaniem programu R.....	257
Karolina Bartos , Odkrywanie wzorców zachowań konsumentów za pomocą analizy koszykowej danych transakcyjnych.....	266
Joanna Trzęsiok , Taksonomiczna analiza krajów pod względem dzietności kobiet oraz innych czynników demograficznych.....	275
Beata Bal-Domańska , Próba identyfikacji większych skupisk regionalnych oraz ich konwergencja.....	285
Beata Bieszk-Stolorz, Iwona Markowicz , Wpływ zasiłku na proces poszukiwania pracy.....	294
Marta Dziechciarz-Duda, Klaudia Przybysz , Wykształcenie a potrzeby rynku pracy. Klasyfikacja absolwentów wyższych uczelni.....	303
Tomasz Klimanek , Problem pomiaru procesu dezagrarnizacji wsi polskiej w świetle wielowymiarowych metod statystycznych.....	313
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Wybrane metody analizy danych wzdluznych.....	321
Artur Zaborski , Zastosowanie miar odległości dla danych porządkowych do agregacji preferencji indywidualnych.....	330
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Zastosowanie analizy korespondencji do badania sytuacji mieszkańców strefy podmiejskiej Warszawy na rynku pracy.....	338

Katarzyna Wawrzyniak , Klasyfikacja województw według stopnia realizacji priorytetów Strategii Rozwoju Kraju 2007-2015 z wykorzystaniem wartości centrum wierszowego	346
---	-----

Summaries

Eugeniusz Gatnar , Statystyka bilansu płatniczego a konkurencyjność gospodarki Polski	22
Andrzej Sokółowski, Magdalena Czaja , Cluster separability and the effectiveness of k -means method	29
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Multisectoral analysis of latent transitions in bankruptcy prediction models.....	38
Elżbieta Golata , Differences in the process of aging and demographic structures in Poznań and the agglomeration compared to selected Polish cities in the years 2002-2011	48
Aleksandra Łuczak, Feliks Wysocki , Determination of weights for features in problems of linear ordering of objects	59
Marek Walesiak , Reinforcing measurement scale for ordinal data in multivariate statistical analysis	68
Paweł Lula , Automatic identification of keywords and keyphrases in documents written in Polish.....	76
Mariusz Kubus , The proposition of modification of the relaxed LASSO method.....	84
Andrzej Bąk, Tomasz Bartłomowicz , Microeconomic multinomial logit models and their implementation in the <code>DiscreteChoice</code> R package .	94
Justyna Brzezińska , The analysis of unemployment data in Poland in 2004-2012 with application of log-linear models	103
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Application of the MMLM package of R software for vodka consumers preference analysis.....	112
Barbara Batóg, Jacek Batóg , Analysis of the stability of classification of Polish voivodeships in 2002-2010 according to the sectoral labour productivity	120
Małgorzata Markowska, Danuta Strahl , Classification of the European regional space in terms of smart growth pillars using the reference limit system.....	130
Kamila Migdał Najman, Krzysztof Najman , Formal quality assessment of group structure mapping on the Kohonen's map	138
Kamila Migdał Najman, Krzysztof Najman , Graphical quality assessment of group structure mapping on the Kohonen's map	147
Beata Basiura, Anna Czapkiewicz , Validation of time series clustering	156
Michał Trzęsiok , Selected methods for outlier detection.....	166

Grażyna Dehnel, Tomasz Klimanek , Taxonomic aspects of indirect estimation accounting for spatial correlation in enterprise statistics	176
Michał Bernard Pietrzak, Justyna Wilk , Economic distance in modeling spatial phenomena with the application of gravity model.....	185
Maciej Beręsewicz , An attempt to use different distance measures in the Generalized Petersen estimator	194
Marcin Szymkowiak, Tomasz Józefowski , Construction and practical using of SPREE estimators for two-dimensional contingency tables.....	201
Marcin Pelka , The ensemble conceptual clustering for symbolic data.....	209
Małgorzata Machowska-Szewczyk , Evaluation of clusters obtained by fuzzy classification methods for symbolic objects.....	219
Justyna Wilk , Problem of determining the number of clusters in taxonomic analysis of symbolic data	228
Andrzej Dudek , Clustering techniques for Google maps markers.....	236
Ewa Roszkowska , The evaluation of negotiation offers in ill structure negotiation problems with the application of fuzzy SAW procedure	247
Marcin Szymkowiak, Marek Witkowski , The use of correspondence analysis in analysing the financial situation of cooperative banks.....	256
Bartłomiej Jefmański , The construction of fuzzy customer satisfaction indexes using R program.....	265
Karolina Bartos , Discovering patterns of consumer behaviour by market basket analysis of the transactional data.....	274
Joanna Trzęsiok , Cluster analysis of countries with respect to fertility rate and other demographic factors	284
Beata Bal-Domańska , An attempt to identify major regional clusters and their convergence	293
Beata Bieszk-Stolorz, Iwona Markowicz , The influence of benefit on the job finding process	302
Marta Dziechciarz-Duda, Klaudia Przybysz , Education and labor market needs. Classification of university graduates	312
Tomasz Klimanek , The problem of measuring deagrarianisation process in rural areas in Poland using multivariate statistical methods.....	320
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Selected methods for an analysis of longitudinal data.....	329
Artur Zaborski , The application of distance measures for ordinal data for aggregation individual preferences	337
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Application of correspondence analysis to examine the situation of the inhabitants of Warsaw suburban area in the labour market	345
Katarzyna Wawrzyniak , Classification of voivodeships according to the level of the realization of priorities of <i>the National Development Strategy 2007-2015</i> with using the values of centroid of the rows	355

Barbara Batóg, Jacek Batóg

Uniwersytet Szczeciński

ANALIZA STABILNOŚCI KLASYFIKACJI POLSKICH WOJEWÓDZTW WEDŁUG SEKTOROWEJ WYDAJNOŚCI PRACY W LATACH 2002-2010

Streszczenie: W artykule badaniu poddana została stabilność klasyfikacji polskich województw według wydajności pracy. W ramach grupowania poszczególnych obiektów wykorzystano wydajność pracy w układzie sektorowym, przy czym przedsiębiorstwa w każdym województwie podzielono na 6 makrosektorów. W pierwszej części badania utworzono rankingi województw w poszczególnych latach za pomocą taksonomicznego miernika rozwoju, a w drugiej części województwa pogrupowano metodą *k*-średnich. Przeprowadzone analizy pozwoliły między innymi ocenić, czy obserwowane w ostatnich latach spowolnienie gospodarcze powodowało zmiany podobieństwa badanych obiektów (województw).

Słowa kluczowe: sektorowa wydajność pracy, metoda *k*-średnich, stabilność klasyfikacji.

1. Wstęp

O poziomie rozwoju społeczno-ekonomicznego, jakości życia oraz dobrobycie poszczególnych krajów decydują przede wszystkim długookresowe tendencje w kształtowaniu się wydajności pracy. Zjawisko to odgrywa również dużą rolę w zakresie redukcji ubóstwa oraz wzroście wydatków publicznych [Blinder, Baumol 1993, s.778]. Pośrednim jego efektem jest charakter i natężenie procesów migracyjnych, inwestycyjnych oraz zróżnicowanie dochodów *per capita*.

Jak wskazuje wielu autorów, istotne znaczenie dla kształtowania się ogólnego poziomu wydajności pracy oraz jej przestrzennego i czasowego zróżnicowania, oprócz wielu innych czynników [Batóg, Batóg 2008, s. 60], ma istniejąca struktura sektorowa gospodarki [Benito, Ezcurra 2003], ze szczególnym uwzględnieniem roli sektora usług. Różnice występujące w wydajności pracy w ujęciu międzynarodowym są też wymieniane jako wyznacznik tempa konwergencji dochodowej [Batóg 2010, s.12].

Dotychczasowe analizy wydajności pracy, przeprowadzane zarówno na poziomie przedsiębiorstw, jak i w ujęciu regionalnym, pozwalają zauważyć pewne prawidłowości. Badanie wydajności polskich przedsiębiorstw dokonane w pracy [Ba-

tóg, Batóg 2012] w oparciu o miary stopnia heterogeniczności pozwoliło sformułować wniosek o zwiększaniu się wydajności pracy wraz ze wzrostem wielkości przedsiębiorstw oraz stwierdzić silne zjawisko przynależności przedsiębiorstw do tej samej klasy wydajności pracy w czasie. Natomiast analiza czynników determinujących zróżnicowanie wydajności pracy polskich województw, przedstawiona w pracy [Batóg, Batóg 2008] i przeprowadzona z wykorzystaniem metody *shift-share* [zob. Esteban 2000], wykorzystywanej pierwotnie do badania czynników kształtujących regionalne i międzynarodowe różnice w poziomie zjawisk o charakterze addytywnym, takich jak: zatrudnienie, eksport, inwestycje, pozwoliła pozytywnie zweryfikować hipotezę o utrzymywaniu się różnic w poziomie wydajności pracy pomiędzy poszczególnymi województwami. Podstawową przyczyną tego zróżnicowania było odmienne kształtowanie się czynników regionalnych, określających poziom konkurencyjności poszczególnych województw, a zwłaszcza struktury sektorowej, której wpływ mierzony był tak zwanym efektem strukturalnym¹.

2. Podstawowe cele pracy i metody badawcze

Rozważania zawarte we wstępie jednoznacznie potwierdzają konieczność przeprowadzania wszechstronnych analiz wydajności pracy. Celem badania zaprezentowanego w artykule jest dokonanie klasyfikacji polskich województw według stopnia podobieństwa sektorowej wydajności pracy oraz ocena stopnia stabilności tej klasyfikacji w czasie. Szczególną uwagę zwrócono przy tym na zbadanie wpływu pojawiających się kryzysów gospodarczych w latach 2001-2003 oraz 2008-2010 na przyporządkowanie poszczególnych województw do określonych klas wydajności pracy. Metody i mierniki oceny stabilności i odporności struktur taksonomicznych przedstawione zostały w wielu pracach [zob. m.in. Walesiak, Gatnar (red.) 2009]. Warto też zapoznać się z rozważaniami przeprowadzonymi na ten temat przez Z. Hellwiga [Hellwig 1995]. W pracy do analizy stabilności uzyskiwanych klasyfikacji w badanym okresie wykorzystana została miara zgodności procentowej, wyznaczana jako udział obiektów, które w dwóch odrębnych klasyfikacjach występowały w tej samej klasie wydajności pracy w ogólnej liczbie obiektów.

W wykonanych analizach wykorzystane zostały metody wielowymiarowej analizy statystycznej: taksonomiczny miernik rozwoju [Hellwig 1968] oraz metoda *k*-średnich [Panek 2009]. Obie te metody są powszechnie znane i zostały wyczerpująco przedstawione w literaturze. Warto jedynie wspomnieć, że w pierwszej z nich normowanie danych oparto na unitaryzacji zerowanej², a normowanie odleg-

¹ Zob. również badanie sektorowej wydajności pracy przeprowadzone z wykorzystaniem estymacji jądrowej w pracy [Batóg, Batóg, Mojsiewicz 2009].

² Parametry normalizacyjne były ustalane dla poszczególnych lat osobno ze względu na wykorzystywanie w porównaniach miejsc w rankingach, a nie odległości od wzorca.

łości obiektów od wzorca dokonane zostało, w przeciwieństwie do najczęściej stosowanego podejścia [Tarczyński 1997, s. 268], w oparciu o odległość wzorca od antywzorca. Natomiast wybór początkowych skupień w przypadku metody k -średnich opierał się na maksymalizacji odległości centrów skupień.

3. Wyniki badania

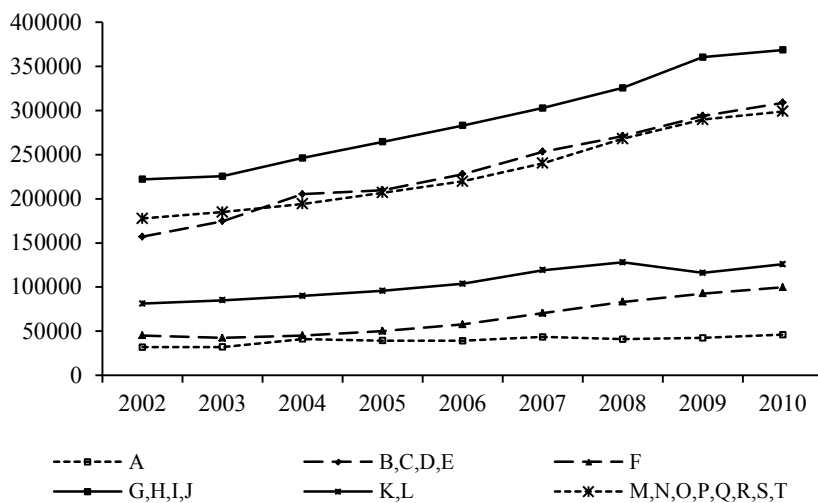
Przeprowadzona analiza wydajności pracy mierzonej wartością dodaną brutto (WDB) na jednego pracującego dotyczyła polskich województw z uwzględnieniem struktury sektorowej w latach 2002-2010. Źródłem danych statystycznych był Bank Danych Lokalnych Głównego Urzędu Statystycznego³. Rozpatrywano sześć następujących sektorów (grup sekcji) według PKD 2007:

1. A – Rolnictwo, leśnictwo, łowiectwo i rybactwo;
2. B+C+D+E – Górnictwo i wydobywanie; Przetwórstwo przemysłowe; Wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną, gorącą wodę i powietrze do układów klimatyzacyjnych; Dostawa wody, gospodarowanie ściekami i odpadami oraz działalność związana z rekultywacją;
3. F – Budownictwo;
4. G+H+I+J – Handel hurtowy i detaliczny, naprawa pojazdów samochodowych, włączając motocykle; Transport i gospodarka magazynowa; Działalność związana z zakwaterowaniem i usługami gastronomicznymi; Informacja i komunikacja;
5. K+L – Działalność finansowa i ubezpieczeniowa; Działalność związana z obsługą rynku nieruchomości;
6. M+...+T – Działalność profesjonalna, naukowa i techniczna; Działalność w zakresie usług administrowania i działalność wspierająca; Administracja publiczna i obrona narodowa, obowiązkowe zabezpieczenia społeczne; Edukacja; Opieka zdrowotna i pomoc społeczna; Działalność związana z kulturą, rozrywką i rekreacją; Pozostała działalność usługowa.

W okresie ostatnich kilkunastu lat wydajność pracy w Polsce znacząco wzrosła. Było to przede wszystkim wynikiem dynamicznego wzrostu wydajności w trzech grupach sektorów (zob. rys. 1): usług rynkowych niefinansowych (G, H, I, J), przemyśle (B, C, D, E) oraz usług nierynkowych (M, N, O, P, Q, R, S, T). Jednocześnie zauważyć można było wzrost regionalnego zróżnicowania tej zmiennej.

W tabeli 1 zaprezentowano rankingi województw w poszczególnych latach badania. Rankingi te wyznaczono za pomocą taksonomicznego miernika rozwoju.

³ www.stat.gov.pl/bdl.



Rys. 1. Wydajność pracy (WDB na jednego pracującego) w grupach sekcji w Polsce w latach 2002-2010 (zł)

Źródło: Bank Danych Lokalnych GUS.

Tabela 1. Ranking województw według TMR w latach 2002-2010

Województwo	2002	2003	2004	2005	2006	2007	2008	2009	2010	Maksymalna różnica pozycji
Dolnośląskie	2	2	2	4	2	2	2	2	2	2
Kujawsko-pomorskie	9	11	11	14	12	11	14	13	12	5
Lubelskie	15	15	15	15	15	15	15	15	15	0
Lubuskie	6	6	5	5	6	10	11	11	11	6
Łódzkie	12	10	10	8	7	9	7	9	8	5
Małopolskie	14	14	13	12	11	14	12	14	14	3
Mazowieckie	1	1	1	1	1	1	1	1	1	0
Opolskie	8	9	7	7	13	6	8	8	9	7
Podkarpackie	16	16	16	16	16	16	16	16	16	0
Podlaskie	7	12	14	11	9	7	10	6	7	8
Pomorskie	3	4	6	3	3	3	5	5	5	3
Śląskie	5	5	3	6	5	5	4	3	4	3
Świętokrzyskie	11	13	12	13	14	12	9	10	10	5
Warmińsko-mazurskie	13	7	8	10	10	13	13	12	13	6
Wielkopolskie	10	8	9	9	8	8	6	7	6	4
Zachodniopomorskie	4	3	4	2	4	4	3	4	3	2

Źródło: obliczenia własne na podstawie danych BDL GUS.

Zauważyć można w całym badanym okresie zajmowanie pierwszej pozycji, charakteryzującej się najwyższą wydajnością pracy, przez województwo mazowieckie. Równie stabilną lokatę (drugą) zajmowało województwo dolnośląskie.

Stabilnością miejsca w rankingach odznaczały się też dwa województwa o najniższym poziomie wydajności: podkarpackie i lubelskie, zajmując odpowiednio 16. i 15. pozycję. Znaczące pogorszenie miejsca odnotowało województwo lubuskie – spadek z pozycji 6. w roku 2002 na pozycję 11. w roku 2010. Odwrotną tendencję zauważyć można w przypadku województw: łódzkiego i wielkopolskiego. Jednocześnie można stwierdzić stosunkowo wysoką stabilność uzyskanych rankingów w całym badanym okresie. Wskazują na to współczynniki τ Kendalla obliczone dla kolejnych par lat – przyjmowały one wartości od 0,72 do 0,93 i były istotne na poziomie 0,05.

W następnej kolejności dokonano grupowania województw według sektorowej wydajności pracy w poszczególnych latach za pomocą metody k -średnich. Województwa podzielono na 5 grup⁴. Wyniki klasyfikacji przedstawiono w tabeli 2, przy czym oznaczenia grup zostały powiązane z poziomem wydajności pracy – grupa I charakteryzowała się najwyższą wydajnością pracy, a grupa V – najniższą wydajnością pracy.

Tabela 2. Przynależność województw do grup wyznaczonych metodą k -średnich w latach 2002-2010

Województwo	2002	2003	2004	2005	2006	2007	2008	2009	2010
Dolnośląskie	II	II	II	II	III	II	II	II	III
Kujawsko-pomorskie	II	II	III	IV	IV	III	III	IV	IV
Lubelskie	V	V	V	V	V	V	V	V	V
Lubuskie	II	II	III	III	IV	III	III	IV	IV
Łódzkie	V	V	V	III	IV	IV	III	III	IV
Małopolskie	V	V	V	V	III	V	IV	V	V
Mazowieckie	I	I	I	I	I	I	I	I	I
Opolskie	III	III	II	II	III	II	II	III	IV
Podkarpackie	V	V	V	V	V	V	V	V	V
Podlaskie	IV	IV	V	III	IV	IV	III	III	IV
Pomorskie	II	II	III	II	III	III	III	IV	IV
Śląskie	III	III	II	II	III	II	II	III	III
Świętokrzyskie	IV	IV	IV	V	IV	IV	IV	III	V
Warmińsko-mazurskie	II	II	III	III	II	III	III	IV	IV
Wielkopolskie	II	II	III	IV	II	III	III	IV	IV
Zachodniopomorskie	II	II	III	III	II	III	III	IV	II

Źródło: obliczenia własne na podstawie danych BDL GUS.

Wyniki klasyfikacji uzyskanych w oparciu o metodę k -średnich w dużym stopniu pokrywają się z tymi zaprezentowanymi w tabeli 1. W celu oceny stabilności tych klasyfikacji obliczono miarę zgodności przynależności województw do poszczególnych grup (zob. tab. 3). Dodatkowo w celu zobrazowania wpływu spadku koniunktury gospodarczej na wydajność pracy w latach 2009-2010 przedstawiono

⁴ Liczba grup została określona na podstawie metody Warda.

szczegółowe tablice zgodności przynależności do grup I-V w latach 2008-2009 oraz 2009-2010 (zob. tab. 4 i 5).

Tabela 3. Współczynnik zgodności klasyfikacji oraz liczby województw zmieniających przynależność do grupy dla par lat

Lata	Zgodność (%)	Wyższa grupa	Niższa grupa
2002-2003	100,0	0	0
2003-2004	43,8	2	7
2004-2005	62,5	3	3
2005-2006	25,0	5	7
2006-2007	43,8	5	4
2007-2008	81,3	3	0
2008-2009	37,5	1	9
2009-2010	62,5	1	5
2002-2010	43,8	1	5

Źródło: obliczenia własne na podstawie danych BDL GUS.

Tabela 4. Tablica zgodności przynależności do grup I-V w latach 2008 i 2009

2008-2009	I	II	III	IV	V
I	1	0	0	0	0
II	0	1	2	0	0
III	0	0	2	6	0
IV	0	0	1	0	1
V	0	0	0	0	2
Zgodność	37,5%				

Źródło: obliczenia własne na podstawie danych BDL GUS.

Tabela 5. Tablica zgodności przynależności do grup I-V w latach 2009 i 2010

2009-2010	I	II	III	IV	V
I	1	0	0	0	0
II	0	0	1	0	0
III	0	0	1	3	1
IV	0	1	0	5	0
V	0	0	0	0	3
Zgodność	62,5%				

Źródło: obliczenia własne na podstawie danych BDL GUS.

Brak jakichkolwiek zmian przynależności województw do poszczególnych klas wydajności pracy charakteryzował wyłącznie lata 2002 i 2003. Mimo że był to okres spowolnienia gospodarczego, to jego wpływ na kształtowanie się wydajności pracy w poszczególnych województwach był bardzo zbliżony. W kolejnych latach odnotowano już jednak znaczący wzrost zróżnicowania w kształtowaniu się wydajności pracy w badanych obiektach. Efektem tego były zmiany (spadek) zgodności

ści kolejnych klasyfikacji. Zjawisko to było również widoczne w latach kolejnego spowolnienia gospodarczego 2008-2010. Wskazuje to na zupełnie odmienny charakter wpływu tego kryzysu na zmiany wydajności pracy w porównaniu z poprzednim okresem kryzysowym w latach 2002-2003. Warto jednocześnie w kolejnych latach odnotować wzrost liczebności grup obiektów, które w danym roku charakteryzowały się niższym poziomem wydajności pracy. O ile w roku 2002 w grupach III-V występowało 8 województw, o tyle w roku 2010 było ich już 14.

4. Wnioski

W latach 2002-2010 w Polsce występował stały wzrost wydajności pracy ogółem oraz w ujęciu sektorowym. Biorąc jednak pod uwagę ujęcie regionalne, zauważyć można zróżnicowane zmiany sektorowej struktury wydajności pracy w poszczególnych województwach. Negatywnym zjawiskiem jest przy tym dominowanie grup zawierających województwa charakteryzujące się niższą wydajnością pracy w odniesieniu do wartości średniej. Pomimo że niektóre województwa zmieniały znacznie swoją pozycję w rankingach, to uzyskane klasyfikacje były stosunkowo stabilne, o czym świadczą obliczone współczynniki Kendalla. Widoczna jest natomiast znacznie niższa stabilność przynależności do grup wydajności pracy uzyskanych metodą k -średnich.

Literatura

- Batóg J. (2010), *Konwergencja dochodowa w krajach Unii Europejskiej. Analiza ekonometryczna*, Uniwersytet Szczeciński, Rozprawy i Studia, t. (DCCCLIV) 780, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
- Batóg J., Batóg B. (2008), *Analiza regionalnych zmian wydajności pracy w Polsce*, „Wiadomości Statystyczne”, Główny Urząd Statystyczny, Polskie Towarzystwo Statystyczne, nr 6, s. 59-69.
- Batóg J., Batóg B. (2012), *Analiza wydajności pracy największych polskich przedsiębiorstw w latach 2004-2008 na podstawie danych panelowych*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 731, Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania nr 26, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin, s. 21-32.
- Batóg J., Batóg B., Mojsiewicz M. (2009), *Application of Kernel Estimation in Analysis of Labour Productivity of the Largest Polish Firms in 2004-2008*, „Folia Oeconomica Stetinensia”, nr 8(16), Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin, s.126-139.
- Benito J.M., Ezcurra R. (2003), *Spatial Disparities in Productivity and Industry Mix. The Case of the European Regions*, Department of Economics, Universidad Pública de Navarra.
- Blinder A., Baumol W. (1993), *Economics: Principles and Policy*, Harcourt Brace Jovanovich, San Diego 1993.
- Esteban J. (2000), *Regional Convergence in Europe and the Industry Mix: a Shift-share Analysis*, „Regional Science and Urban Economics”, 30.
- Hellwig Z. (1968), *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, „Przegląd Statystyczny”, nr 4, s. 307-327.

- Hellwig Z. (1995), *Stabilność i odporność struktur taksonomicznych*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych. Problemy teoretyczne*, Taksonomia, Zeszyt 2, Jelenia Góra – Wrocław – Kraków, s. 7-30.
- Panek T. (2009), *Statystyczne metody wielowymiarowej analizy porównawczej*, Szkoła Główna Handlowa w Warszawie, Warszawa 2009.
- Tarczyński W. (1997), *Rynki kapitałowe. Metody ilościowe*, Vol.1, Agencja Wydawnicza Placet, Warszawa 1997.
- Walesiak M., Gatnar E. (red.) (2009), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa.

ANALYSIS OF THE STABILITY OF CLASSIFICATION OF POLISH VOIVODESHIPS IN 2002-2010 ACCORDING TO THE SECTORAL LABOUR PRODUCTIVITY

Summary: In the paper the authors examined the stability of classification of Polish voivodeships according to labour productivity. The grouping was based on labour productivity in 6 macrosectors. In the first part of analysis taxonomic measures of development were calculated, whereas in the second part the k-means method was applied in order to classify voivodeships into 5 groups. The conducted analysis allowed to evaluate whether worldwide slowdown in economy caused changes in the similarities of objects.

Keywords: sectoral labour productivity, k-means method, stability of classification.