

OCENA STABILNOŚCI WYNIKÓW KLASYFIKACJI WOJEWÓDZTW POLSKI POD WZGLĘDEM POZIOMU ŻYCIA LUDNOŚCI

Małgorzata Machowska-Szewczyk

Katedra Metod Sztucznej Inteligencji i Matematyki Stosowanej
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
e-mail: mmachowska@wi.zut.edu.pl

Agnieszka Sompolska-Rzechuła

Katedra Zastosowań Matematyki w Ekonomii
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
e-mail: asompolska@zut.edu.pl

Streszczenie: W pracy przedstawiono ocenę stabilności klasyfikacji województw Polski pod względem poziomu życia w latach 2002-2009. Dobór cech diagnostycznych został przeprowadzony w sposób dynamiczny, dzięki czemu otrzymany zestaw cech końcowych jest aktualny nie tylko w okresie badanym, ale także w przyszłości. Do klasyfikacji województw zastosowano metodę Warda oraz sprawdzono dopasowanie dendrogramów do macierzy odległości D za pomocą współczynnika korelacji kofenetycznej, wskaźnika całościowego dopasowania i miary STRESS. Wykorzystując wskaźniki podobieństwa wyników klasyfikacji stwierdzono, że klasyfikacje charakteryzują się dość dobrą stabilnością w czasie.

Słowa kluczowe: stabilność wyników klasyfikacji, metoda Warda, miary dopasowania

WPROWADZENIE

W ostatnich latach można zaobserwować wzrost zainteresowania takimi pojęciami jak: dobrobyt ekonomiczny, poziom życia, jakość życia, warunki życia. Po okresie fascynacji tempem postępu technicznego i ekonomicznego oraz intensywnych badań nad rozwojem gospodarczym, nastąpił etap analizy korzyści i zagrożeń, wynikających z postępu cywilizacyjnego dla człowieka. Zatem, poza badaniem poziomu życia pod względem materialnym, pojawiły się propozycje

badania poziomu życia pod względem zaspokajania potrzeb duchowych, moralnych, emocjonalnych, psychicznych. Jednak w pełni obiektywna ocena tej kategorii jest zadaniem niezmiernie trudnym, gdyż z poziomem życia związane są takie cechy jak: ukryte bezrobocie, liczba samobójstw, ilość oraz sposób wykorzystania czasu wolnego czy możliwości korzystania z dóbr kultury, o których trudno jest pozyskać informacje i dokonać ich prawidłowej kwantyfikacji [Zeliaś 2004]. Z uwagi na wieloaspektowość do badania poziomu życia można wykorzystać metody wielowymiarowej analizy danych.

W artykule przedstawiono wyniki taksonomicznego badania poziomu życia w województwach Polski w ujęciu dynamicznym. Celem pracy jest ocena stabilności wyników klasyfikacji województw Polski ze względu na wyróżnioną kategorię w latach 2002-2009. W okresie trwającego w Polsce rozwarstwienia społecznego i pauperyzacji znacznej części społeczeństwa skierowanie uwagi na problem oceny poziomu życia ma szczególne znaczenie.

DOBÓR CECH DIAGNOSTYCZNYCH W UJĘCIU DYNAMICZNYM

Do doboru zmiennych, które charakteryzują się relatywnie wysokimi wahaniami wartości oraz gdy celem jest przewidywanie przyszłego rozwoju badanego zjawiska, wykorzystać można współczynniki zmienności oraz korelacji liniowej, obliczane na podstawie danych dotyczących nie tylko ostatniego okresu, ale również kilku poprzednich okresów [Zeliaś 2004].

Ocenę stabilności klasyfikacji przeprowadzono na podstawie podziałów województw Polski w poszczególnych latach okresu 2002-2009 pod względem poziomu życia mieszkańców. Wykorzystano dane statystyczne udostępnione przez Główny Urząd Statystyczny w Banku Danych Lokalnych na stronie internetowej <http://www.stat.gov.pl/bdl/app/portret.dims> do wyznaczenia wartości wskaźników, które charakteryzują analizowane zjawisko. W badaniu przyjęto następujący zestaw potencjalnych cech diagnostycznych:

X_1 – liczba ludności na 1 km²,

X_2 – udział ludności w wieku przedprodukcyjnym w ogólnej liczbie ludności,

X_3 – udział ludności w wieku produkcyjnym w ogólnej liczbie ludności,

X_4 – udział ludności w wieku poprodukcyjnym w ogólnej liczbie ludności,

X_5 – ludność w wieku nieprodukcyjnym na 100 osób w wieku produkcyjnym,

X_6 – ludność w wieku poprodukcyjnym na 100 osób w wieku przedprodukcyjnym,

X_7 – ludność w wieku poprodukcyjnym na 100 osób w wieku produkcyjnym,

X_8 – liczba kobiet na 100 mężczyzn,

X_9 – zgony na 1000 ludności,

X_{10} – przyrost naturalny na 1000 ludności,

X_{11} – urodzenia żywe na 1000 ludności,

X_{12} – liczba małżeństw zawartych w ciągu roku na 1000 ludności,

X_{13} – liczba rozwodów na 1000 ludności,

X_{14} – zgony niemowląt na 1000 urodzeń żywych,

- X_{15} – przeciętne miesięczne wydatki na 1 osobę w zł,
 X_{16} – stopa bezrobocia w %,
 X_{17} – liczba ofert pracy ogółem na 1 bezrobotnego,
 X_{18} – wskaźnik zatrudnienia ogółem w %,
 X_{19} – przeciętne miesięczne wynagrodzenie brutto w relacji do średniej krajowej (Polska=100),
 X_{20} – przeciętna powierzchnia użytkowa mieszkania w m² na 1 osobę,
 X_{21} – liczba mieszkań na 10 tys. ludności,
 X_{22} – liczba studentów na 10 tys. ludności,
 X_{23} – liczba praktyk lekarskich w miastach na 10 tys. ludności,
 X_{24} – liczba praktyk lekarskich na wsi na 10 tys. ludności,
 X_{25} – liczba ludności na 1 aptekę ogólnodostępną,
 X_{26} – liczba osób korzystających ze świadczeń pomocy społecznej na 10 tys. ludności,
 X_{27} – stopień wykorzystania miejsc noclegowych w %,
 X_{28} – czytelnicy bibliotek publicznych na 1000 ludności,
 X_{29} – wypożyczenia księgozbioru na 1 czytelnika,
 X_{30} – liczba ludności na 1 miejsce w kinach stałych,
 X_{31} – liczba widzów i słuchaczy w teatrach i instytucjach muzycznych na 1000 ludności,
 X_{32} – PKB na 1 mieszkańca,
 X_{33} – nakłady inwestycyjne na 1 mieszkańca w zł,
 X_{34} – drogi publiczne o twardej nawierzchni w km na 100 km² powierzchni,
 X_{35} – liczba samochodów osobowych zarejestrowanych na 1000 ludności,
 X_{36} – liczba ofiar śmiertelnych na 100 wypadków drogowych,
 X_{37} – emisja przemysłowych zanieczyszczeń pyłowych powietrza w t na 100 km²,
 X_{38} – emisja przemysłowych zanieczyszczeń gazowych powietrza w t na 100 km²,
 X_{39} – udział parków narodowych w ogólnej powierzchni w %,
 X_{40} – plony z 1 ha zbóż ogółem w dt,
 X_{41} – dochody budżetu województwa ogółem na 1 mieszkańca w zł,
 X_{42} – wydatki z budżetu województwa ogółem na 1 mieszkańca w zł,
 X_{43} – liczba podmiotów gospodarczych na tys. mieszkańców.

Metoda doboru cech diagnostycznych w ujęciu dynamicznym oraz procedura doboru zostały przedstawione w pracy [Machowska-Szewczyk, Sompolska-Rzechuła 2012]. Po selekcji cechy: X_1 , X_{22} , X_{26} , X_{32} , X_{36} , X_{39} tworzą nową listę zmiennych, które są słabo skorelowane między sobą i równocześnie silnie skorelowane z pozostałymi cechami, wyeliminowanymi ze zbioru cech diagnostycznych.

METODA WARDA I OCENA JAKOŚCI DENDROGRAMU

Spośród bardzo wielu metod hierarchicznych do badań wybrano metodę Warda. Na każdym etapie tej metody optymalizuje się otrzymany podział, przez

połączenie dwóch elementów, stosując kryterium minimalnego wzrostu łącznej wewnątrzgrupowej sumy kwadratów odchyłeń wszystkich wartości zmiennych dla każdego obiektu od ich średnich grupowych. Zapewnia ona zatem homogeniczność wewnątrz skupień i heterogeniczność między skupieniami, przez co uznawana jest za bardzo efektywną [Ward 1963]. Na podstawie przeprowadzonych badań można stwierdzić, że metoda Warda jest prawie o 40% bardziej efektywna¹ w porównaniu z drugą z kolei pod względem efektywności – metodą najdalszego sąsiedztwa [Malina 2004]. Ponadto metoda Warda może być wykorzystana do określenia wstępnej liczby klas na podstawie wzrokowej analizy dendrogramu, jednak jest to sposób subiektywny, ponieważ trudno jednoznacznie odpowiedzieć na pytanie, w którym miejscu należy przerwać proces aglomeracji, tzn. na ile klas podzielić daną zbiorowość.

Zaletą metod hierarchicznych (w tym także metody Warda) jest graficzna prezentacja klasyfikacji za pomocą dendrogramu, który ilustruje kolejność łączenia się obiektów w klasy i poziomy, na których obiekty po raz pierwszy się połączyły. W literaturze przedmiotu można znaleźć propozycje mierzenia dopasowania dendrogramu do macierzy odległości **D**, takie jak np.: współczynnik korelacji kofenetycznej, wskaźnik całościowego dopasowania, statystyka Gowera, miara STRESS [Gordon 1999]). Porównuje się wyjściową macierz odległości **D** z macierzą odległości **C**, otrzymaną na podstawie dendrogramu - zwaną macierzą dendrogramu lub macierzą kofenetyczną. Elementami macierzy **C** są odległości odpowiadające poziomom łączenia, na których pary obiektów łączą się w tym samym skupieniu pierwszy raz.

Dopasowanie dendrogramu można ocenić za pomocą tzw. współczynnika korelacji kofenetycznej (cophenetic correlation coefficient) [Balicki 2009]:

$$r_{kof.} = \frac{\left[\frac{n(n-1)}{2} \sum_{r<s} d_{rs} \cdot c_{rs} - \sum_{r<s} d_{rs} \cdot \sum_{r<s} c_{rs} \right]}{\sqrt{\left[n(n-1)/2 \sum_{r<s} d_{rs}^2 - \left(\sum_{r<s} d_{rs} \right)^2 \right] \left[n(n-1)/2 \sum_{r<s} c_{rs}^2 - \left(\sum_{r<s} c_{rs} \right)^2 \right]}}, \quad (1)$$

gdzie: n - liczba obiektów, $r, s \in \{1, \dots, n\}$, d_{rs} - elementy macierzy odległości między obiektami **D**, c_{rs} - elementy macierzy kofenetycznej **C**, n - liczba obiektów.

Dendrogram dobrze odzwierciedla różnice między obiektami lub podobieństwa obiektów, jeżeli wartość współczynnika korelacji kofenetycznej jest bliska 1. Wartość współczynnika przeciętnie maleje wraz ze wzrostem liczby obiektów i prawie nie zależy od liczby zmiennych.

¹ Przez efektywność metod taksonomicznych należy rozumieć zdolność prawidłowego rozpoznawania rzeczywistej struktury obiektów w wielowymiarowych przestrzeniach cech, którą bada się na przykładach empirycznych lub na podstawie zbiorów sztucznie wygenerowanych [Malina 2004].

Inne sposoby oceny grupowania oparte są na różnicach odległości w dwóch macierzach: wyjściowej macierzy odległości i macierzy kofenetycznej. Ocenia się zatem poziom niezgodności między d_{rs} a c_{rs} w macierzach **D** i **C**. Wykorzystują one niektóre relacje, jakie zachodzą między sumami odległości w tych macierzach: $\sum d_{rs}$ oraz $\sum c_{rs}$, różnicami między odpowiadającymi sobie odległościami: $d_{rs} - c_{rs}$ oraz ich sumami: $\sum (d_{rs} - c_{rs})$. Dodatnia suma różnic będzie świadczyła o łącznym skróceniu odległości na dendrogramie w stosunku do odległości wyjściowych, i odwrotnie – ujemna suma różnic będzie informowała o rozciągnięciu odległości. Im większa jest bezwzględna wartość tej wielkości, tym gorzej dendrogram oddaje rzeczywiste odległości. Badanie różnic stanowi podstawę wyznaczania wskaźnika całościowego dopasowania [Balicki 2009]:

$$\delta = \sum_{\substack{r,s \\ r < s}} |d_{rs} - c_{rs}|. \quad (2)$$

Inną miarą, wprowadzoną w 1964 roku przez Kruskala, stosowaną w niemetrycznym skalowaniu wielowymiarowym jest miara o nazwie STRESS (STandardized Residual Sum of Squares) [Balicki 2009]:

$$\text{STRESS} = \left[\left(\sum_{\substack{r,s \\ r < s}} (d_{rs} - c_{rs})^2 \right) / \left(\sum_{\substack{r,s \\ r < s}} d_{rs}^2 \right) \right]^{\frac{1}{2}}, \quad (3)$$

gdzie c_{rs} są elementami macierzy kofenetycznej, $r, s \in \{1, \dots, n\}$.

Tabela 1. Wybrane odległości d_{rs} z wyjściowej macierzy odległości **D** i odległości dendrogramu c_{rs}

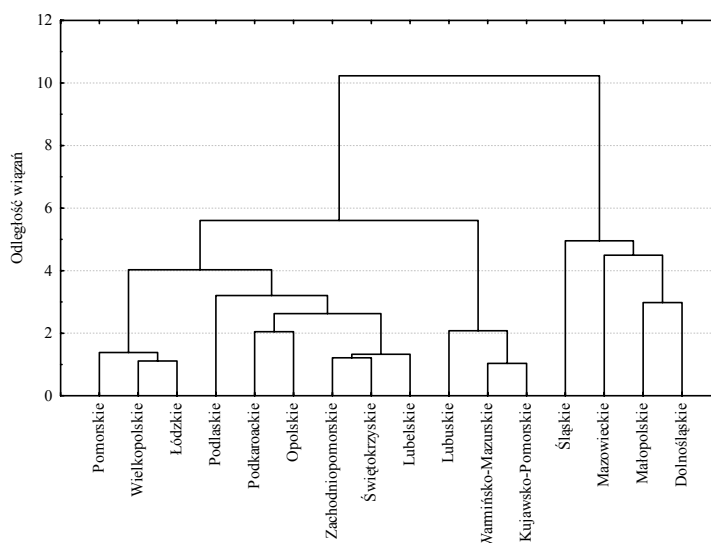
Pary obiektów	2002		2003		...	2008		2009	
	d_{rs}	c_{rs}	d_{rs}	c_{rs}	...	d_{rs}	c_{rs}	d_{rs}	c_{rs}
1-2	4,53	8,58	4,43	7,81	...	3,18	10,95	3,08	10,59
1-3	4,33	8,58	4,01	7,81	...	2,65	10,95	2,78	10,59
1-4	5,33	8,58	5,05	7,81	...	4,25	10,95	3,91	10,59
1-5	4,04	6,53	3,80	6,36	...	1,35	1,33	1,70	1,76
...
13-15	1,94	8,58	1,77	2,31	...	2,20	10,95	2,17	10,59
13-16	1,77	2,60	2,58	7,81	...	0,99	0,99	1,08	1,08
14-15	3,27	8,58	2,92	7,81	...	3,61	10,95	3,62	10,59
14-16	2,62	2,60	2,07	1,95	...	2,38	3,34	2,33	2,46
15-16	1,98	8,58	1,55	7,81	...	1,70	10,95	1,63	10,59
$r_{kof.}$	0,5259		0,4762		...	0,5402		0,5220	
δ	1,1016		1,0019		...	1,5450		1,4708	
STRESS	361,9405		345,0522		...	509,4228		475,2159	

Źródło: opracowanie własne

W każdym roku badanego okresu 2002-2009 przeprowadzono klasyfikację województw metodą Warda i sporządzono dendrogramy. Na tej podstawie wyznaczono macierze kofenetyczne i zbadano dopasowanie dendrogramów do macierzy odległości **D** oraz policzono współczynniki korelacji kofenetycznej, wskaźniki całościowego dopasowania i miary STRESS. Ponieważ macierze kwadratów odległości euklidesowych oraz macierze kofenetyczne mają wymiar 16×16 , ponadto wszystkich kombinacji pomiędzy różnymi województwami jest $n(n-1)/2 = 120$, to w tabeli 3 przedstawiono jedynie fragmenty tych macierzy oraz wartości miar dopasowania.

Najwyższa wartość współczynnika korelacji kofenetycznej została zaobserwowana dla roku 2005, co świadczy o dobrym dopasowaniu dendrogramu do wyjściowej macierzy odległości. Współczynniki korelacji kofenetycznej dla dendrogramów w pozostałych latach przyjmują wartości bliskie 0,5, co oznacza umiarkowane dopasowanie macierzy **C** i **D**. Najmniejsze różnice między odległościami w macierzach **C** i **D** a tym samym najniższe wartości miar: całościowego dopasowania oraz STRESS, występują dla roku 2003. Zatem jedynie w roku 2005 województwa wykazują wyraźną hierarchiczną strukturę grupową (rys. 1).

Rysunek 1. Dendrogram podziału województw Polski w roku 2005 za pomocą metody Warda



Źródło: opracowanie własne

W celu oceny stabilności klasyfikacji województw Polski ustalono jednakową liczbę skupień oraz wcześniej wyodrębniony taki sam zbiór zmiennych diagnostycznych w każdym roku badanego okresu 2002-2009. Po analizie

dendrogramów przyjęto trzy klasy typologiczne i otrzymano wyniki klasyfikacji, które prezentuje tabela 2.

Tabela 2. Wyniki dynamicznej klasyfikacji województw Polski w latach 2002-2009

Rok	Skupienie	Województwa
2002	G1	lubuskie, podlaskie, zachodniopomorskie, pomorskie, podkarpackie, warmińsko-mazurskie, świętokrzyskie, lubelskie, kujawsko-pomorskie
	G2	śląskie, małopolskie, opolskie, wielkopolskie, łódzkie
	G3	mazowieckie, dolnośląskie
2003	G1	lubuskie, warmińsko-mazurskie, zachodniopomorskie, lubelskie, podkarpackie, pomorskie, kujawsko-pomorskie
	G2	śląskie, małopolskie, świętokrzyskie, opolskie, wielkopolskie, łódzkie
	G3	mazowieckie, podlaskie, dolnośląskie
2004	G1	lubuskie, warmińsko-mazurskie, zachodniopomorskie, lubelskie, podkarpackie, kujawsko-pomorskie, podlaskie
	G2	śląskie, małopolskie, świętokrzyskie, pomorskie, opolskie, wielkopolskie, łódzkie
	G3	mazowieckie, dolnośląskie
2005	G1	podlaskie, zachodniopomorskie, świętokrzyskie, pomorskie, opolskie, wielkopolskie, łódzkie, lubelskie, podkarpackie
	G2	lubuskie, warmińsko-mazurskie, kujawsko-pomorskie
	G3	śląskie, mazowieckie, małopolskie, dolnośląskie
2006	G1	podlaskie, podkarpackie, opolskie, zachodniopomorskie, świętokrzyskie, lubelskie, lubuskie, warmińsko-mazurskie, kujawsko-pomorskie
	G2	śląskie
	G3	mazowieckie, małopolskie, dolnośląskie, pomorskie, łódzkie, wielkopolskie
2007	G1	podlaskie, podkarpackie, opolskie, zachodniopomorskie, świętokrzyskie, lubelskie, lubuskie, warmińsko-mazurskie, kujawsko-pomorskie
	G2	śląskie
	G3	mazowieckie, małopolskie, dolnośląskie, pomorskie, łódzkie, wielkopolskie
2008	G1	podlaskie, podkarpackie
	G2	opolskie, zachodniopomorskie, świętokrzyskie, lubelskie, lubuskie, warmińsko-mazurskie, kujawsko-pomorskie
	G3	śląskie, mazowieckie, małopolskie, pomorskie, łódzkie, dolnośląskie, wielkopolskie
2009	G1	podlaskie
	G2	opolskie, lubelskie, lubuskie, podkarpackie, warmińsko-mazurskie, zachodniopomorskie, świętokrzyskie, kujawsko-pomorskie
	G3	śląskie, mazowieckie, małopolskie, pomorskie, łódzkie, dolnośląskie, wielkopolskie

Źródło: opracowanie własne

OCENA STABILNOŚCI KLASYFIKACJI

Do oceny stopnia podobieństwa wyników klasyfikacji obiektów w dwóch momentach można wykorzystać wskaźniki zgodności wyników podziałów, oparte na tablicy kontyngencji. Rozpatruje się wszystkie pary obiektów $O_i, O_j, i, j \in \{1, \dots, n\}, i \neq j$ ze względu na ich przynależność do grup typologicznych z obu podziałów. Każdej parze obiektów badania przyporządkowuje się wartość 1, jeśli obiekty te należą do tej samej grupy typologicznej lub wartość 0, jeśli obiekty te należą do dwóch różnych grup typologicznych. W ten sposób otrzymuje się dwie macierze przyporządkowania: \mathbf{P}_t dla klasyfikacji w momencie t oraz $\mathbf{P}_{t'}$ dla klasyfikacji w momencie t' . Na tej podstawie buduje się czteropolową tablicę kontyngencji (tabela 5).

Tabela 3. Czteropolowa tablica kontyngencji wyników klasyfikacji w dwóch momentach

Podział w momencie t	Podział w momencie t'	
	1	0
1	a	b
0	c	d

Źródło: opracowanie własne na podstawie [Nowak 1990, s. 139].

Wskaźnik podobieństwa wyników klasyfikacji, oparty na tablicy kontyngencji w wymiarach 2×2 ma następującą postać [Nowak 1990]:

$$w_{tt'} = \frac{a + d}{a + b + c + d}. \quad (4)$$

Miara (4) przyjmuje wartości z przedziału $\langle 0, 1 \rangle$, wartości bliskie 1 wskazują na dużą zgodność wyników obu klasyfikacji. Jeżeli w jednej klasyfikacji otrzyma się jedną grupę n elementową a w drugiej n grup jednoelementowych, to wskaźnik (4) przyjmie wartość 0. Natomiast $w_{tt'} = 1$, gdy oba podziały dają identyczne wyniki. Jeżeli porównuje się wyniki klasyfikacji z wielu momentów, to wskaźniki zgodności podziałów można przedstawić w macierzy zgodności o postaci [Nowak 1990]:

$$\mathbf{W} = \begin{bmatrix} 1 & w_{12} & \dots & w_{1T} \\ w_{21} & 1 & \dots & w_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ w_{T1} & w_{T2} & \dots & 1 \end{bmatrix}. \quad (5)$$

Dla klasyfikacji otrzymanych metodą Warda w kolejnych latach okresu 2002-2009 wyznaczono macierz wskaźników podobieństwa wyników klasyfikacji (tabela 6).

Tabela 4. Macierz wskaźników podobieństwa wyników klasyfikacji w latach 2002-2009

Lata	2002	2003	2004	2005	2006	2007	2008	2009
2002	1,0000	0,8167	0,8000	0,6167	0,7167	0,7167	0,6333	0,6667
2003	0,8167	1,0000	0,8333	0,5833	0,5833	0,5833	0,6167	0,6500
2004	0,8000	0,8333	1,0000	0,6167	0,7000	0,7000	0,6667	0,6833
2005	0,6167	0,5833	0,6167	1,0000	0,6000	0,6000	0,5833	0,5833
2006	0,7167	0,5833	0,7000	0,6000	1,0000	1,0000	0,8333	0,8833
2007	0,7167	0,5833	0,7000	0,6000	1,0000	1,0000	0,8333	0,8833
2008	0,6333	0,6167	0,6667	0,5833	0,8333	0,8333	1,0000	0,9333
2009	0,6667	0,6500	0,6833	0,5833	0,8833	0,8833	0,9333	1,0000

Źródło: opracowanie własne

Analizując elementy macierzy wskaźników podobieństwa wyników klasyfikacji województw Polski pod względem poziomu życia mieszkańców w latach 2002-2009 można uznać, że podziały te są dość stabilne w czasie. Największą zgodność wykazują podziały w sąsiednich latach, np. klasyfikacje w latach 2006 i 2007 dają jednakowe wyniki ($w_{2006,2007}=1$). Klasyfikacja otrzymana w roku 2003 wykazuje najniższą zgodność z klasyfikacjami w latach 2005, 2006 i 2007 oraz podział w roku 2005 z podziałami w latach 2008 i 2009 (wskaźniki zgodności podziałów w tych przypadkach wynoszą 0,5833).

Najniższa zgodność klasyfikacji w roku 2003 w porównaniu z latami 2005, 2006 i 2007 wynika z dość dużego tempa wzrostu PKB na 1 mieszkańca, liczby studentów na 10 tys. ludności, liczby osób korzystających ze świadczeń pomocy społecznej na 10 tys. ludności.

PODSUMOWANIE

W pracy przedstawiono ocenę stabilności klasyfikacji województw Polski pod względem poziomu życia mieszkańców. Badany okres obejmował lata 2002-2009. Dynamiczne ujęcie zastosowano na etapie doboru cech diagnostycznych, wykorzystując funkcje trendów liniowych dla współczynników zmienności przyjętych cech oraz funkcje trendów liniowych współczynników korelacji. Dynamiczny dobór zmiennych diagnostycznych ma tę zaletę, że otrzymany zestaw cech końcowych jest nie tylko aktualny w okresie badanym, ale także w przyszłości. Na podstawie wyłonionego zbioru cech diagnostycznych dokonano, metodą Warda, klasyfikacji województw Polski w każdym roku badanego okresu oraz zbadano dopasowanie dendrogramów do macierzy odległości **D** za pomocą współczynnika korelacji kofenetycznej, wskaźnika całościowego dopasowania i miary STRESS. Wartości miar wskazują na dość dobrą lub umiarkowaną zgodność macierzy odległości i dendrogramów. W kolejnym kroku badania dokonano oceny stabilności otrzymanych podziałów, za pomocą wskaźników podobieństwa wyników klasyfikacji i otrzymano, że klasyfikacje charakteryzują się dość dobrą stabilnością w czasie.

Przedstawione podejście pozwoliło na wykorzystanie danych przekrojowo-czasowych w dokonaniu klasyfikacji województw Polski pod względem poziomu życia, uwzględniając dynamikę zmian zachodzących w wartościach cech ilustrujących badane obiekty w całym okresie badania. Województwa Polski są zróżnicowane pod względem poziomu życia. Badania o charakterze dynamicznym powinny być prowadzone w sposób ciągły, gdyż systematyczna analiza i ocena przestrzennego zróżnicowania poziomu życia pozwala na bieżącą kontrolę jego stanu oraz stosowanie odpowiedniej polityki społeczno-gospodarczej, w przypadku zaobserwowania negatywnych zjawisk.

BIBLIOGRAFIA

- Balicki A. (2009) Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne, Wydawnictwo Uniwersytetu Gdańskiego, str. 283-291.
- Gordon A. D. (1999) Classification, Chapman & Hall/CRC, Boca Raton.
- Machowska-Szewczyk M., Sompolska-Rzechuła A. (2012), Dynamiczny dobór cech w taksonomicznej analizie obiektów, Wiadomości Statystyczne, nr 9, str. 34-49.
- Malina A. (2004) Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw, Wydawnictwo Akademii Ekonomicznej w Krakowie.
- Nowak E. (1990) Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych, PWE, Warszawa.
- Zeliaś A. (red) (2004) Poziom życia w Polsce i krajach Unii Europejskiej, PWE, Warszawa.
- Ward J. H. (1963) Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association 58.

EVALUATION OF THE RESULTS STABILITY OF THE POLISH VOIVODSHIPS' CLASSIFICATION

Abstract: The work presents evaluation of the results stability of the Polish regions' classification in terms of a living standard. The research covered the period 2002-2009. Selection of diagnostic features was carried out in a dynamic way, so that the final set of features is valid not only in the period considered, but also in the future. For the regions' classification Ward's method was used and the fit of dendrograms distance to matrix D was checked by using the cophenetic correlation coefficient, the indicator of the overall fit and the STRESS measure. Using the similarity indices of the classification results, it was found that the classifications have a fairly good stability in time.

Keywords: stability of results of classification, Ward's method, measures of fit