

**Michał Trzęsiok**

Uniwersytet Ekonomiczny w Katowicach

---

**GDZIE JEST TRZECI ŚWIAT?  
– ANALIZA TAKSONOMICZNA**

---

**Streszczenie:** W artykule przedstawiono próbę wielowymiarowego spojrzenia na problem identyfikacji krajów Trzeciego Świata z wykorzystaniem modeli taksonomicznych i porównanie wyników klasyfikacji państw otrzymanych z analizy danych z ONZ za 2008 oraz 1973 rok. Celem artykułu jest weryfikacja hipotezy sformułowanej przez prof. Hansa Roslinga, głoszącej, że używany niegdyś podział państw na kraje rozwinięte i kraje Trzeciego Świata staje się coraz mniej uprawniony wobec znaczących zmian w sytuacji społeczno-gospodarczej i szybkiego rozwoju licznej grupy państw Ameryki Południowej, Azji i Afryki.

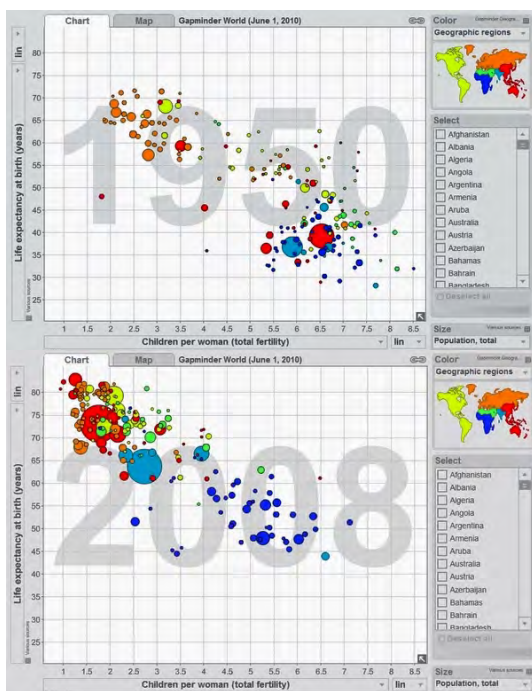
**Słowa kluczowe:** taksonomia, wielowymiarowa analiza porównawcza.

## 1. Wstęp

Termin „kraje Trzeciego Świata” powstał w okresie zimnej wojny i określano nim wszystkie kraje, które nie należały do żadnego z dwóch bloków państw – stron zimnej wojny. Wyróżniano więc układ państw socjalistycznych oraz tzw. „kraje Zachodu”, utożsamiane z sojuszem NATO. Niemal wszystkie pozostałe państwa, tj. około 130 krajów afrykańskich, azjatyckich oraz Ameryki Łacińskiej, określano mianem „Trzeciego Świata”. Szybko jednak nazwa ta przestała odnosić się bezpośrednio do politycznego zaangażowania (braku zaangażowania) kraju w konflikcie zaistniałym w ramach zimnej wojny i zaczęła funkcjonować jako określenie państw znacznie ustępujących pozostałym krajom pod względem rozwoju cywilizacyjnego.

Prof. Hans Rosling jest ekspertem w sprawach zdrowia globalnego i współautorem obszernej pracy na ten temat [Lindstrand i in. 2006]. Spędził wiele lat pracując w Afryce, gdzie miał okazję obserwować zmiany zachodzące w tempie rozwoju niektórych państw afrykańskich. Obecnie prof. Rosling wykłada na najbardziej renomowanej uczelni szwedzkiej – Karolinska Institute, a ponadto jest zaangażowany w niekomercyjne projekty, które mają na celu upowszechnianie wiedzy o zmianach w sytuacji państw i regionów w zakresie zdrowia globalnego. Zdaniem

prof. Roslinga istnieje wiele nieuprawnionych lub zdezaktualizowanych przekonań, które znacznie obniżają skuteczność zarządzania i pomocy państwom, które takiej pomocy mogą potrzebować. Znakomitą formą upowszechniania wiedzy z obszaru zdrowia globalnego i rozwoju cywilizacyjnego są m.in. wykłady prof. Roslinga dostępne na portalu [www.TED.com](http://www.TED.com) oraz niekomercyjny projekt [gapminder.org](http://gapminder.org), umożliwiający dostęp do narzędzi wspomagających wizualizację danych statystycznych zaczerpniętych głównie ze statystyk ONZ.



**Rys. 1.** Zmiany w zakresie dzietności kobiet (oś odciętych) oraz przeciętnego trwania życia (oś rzędnych) – rok 1950 i 2008

Źródło: opracowanie własne na podstawie oprogramowania dostępnego przez [www.gapminder.org](http://www.gapminder.org).

dzia badawczego. Służą jedynie do wizualizacji danych i są bardzo przydatne do stawiania hipotez badawczych. Głównym celem artykułu jest próba wielowymiarowego spojrzenia na problem identyfikacji krajów Trzeciego Świata z wykorzystaniem analizy taksonomicznej i porównanie wyników sytuacji obecnej (w zakresie dostępnych danych statystycznych) z sytuacją sprzed 35 lat. Można więc stwierdzić, że celem jest przeprowadzenie analogicznej analizy z tą przeprowadzoną przez prof. Roslinga – z tym wyjątkiem, że dane pierwotne zostaną przetworzo-

Niniejszy artykuł jest inspirowany jednym z wykładów prof. Roslinga, w którym przedstawił on dynamikę zmian dwóch zmiennych uznawanych za główne cechy dyskryminujące kraje rozwinięte i kraje Trzeciego Świata: przeciętne trwanie życia oraz dzietność kobiet (por. rys. 1). Kraje rozwinięte charakteryzują się niską dzietnością kobiet (mało liczne rodziny) oraz długim przeciętnym trwaniem życia, natomiast w krajach Trzeciego Świata charakterystyki te układają się odwrotnie, tzn. niskie wartości przeciętnego trwania życia oraz wysoka dzietność kobiet. Z rysunku 1 widać, że o ile w roku 1950 można było dosyć wyraźnie wyróżnić dwa skupienia państw mało oraz dobrze rozwiniętych, o tyle w roku 2008 nastąpiła znaczna koncentracja państw w lewym górnym rogu.

Zaznaczyć należy, że wykresy bąbelkowe, którymi posłużył się prof. Rosling, nie stanowią narzędzia

ne metodami wielowymiarowej analizy statystycznej. Dodatkowo, żeby szerzej ująć zagadnienia rozwoju cywilizacyjnego państw, w analizie uwzględniono zestaw czterech zmiennych opisujących badane państwa, dołączając aspekt ekonomiczny w postaci zmiennej PKB *per capita* oraz zmiennej – czynnik opisujący umieralność dzieci przed ukończeniem pierwszego roku życia. Poszczególne etapy analizy taksonomicznej, omówione m in. w pracy [Gatnar, Walesiak (red.) 2004], przedstawiono w kolejnym punkcie artykułu.

## 2. Procedura badawcza

### 2.1. Wybór obiektów i zmiennych

W analizie wykorzystano zbiór danych skonstruowany w oparciu o zmienne i odpowiadające im szeregi czasowe dostępne przez portal [www.gapminder.org](http://www.gapminder.org). Dostępność i kompletność danych w dużym stopniu zdeterminowała okres oraz zbiór obserwacji poddany analizie. Porównanie podziału państw na skupienia wymagało wykorzystania tego samego zestawu obiektów dla dwóch badanych okresów. Ze względu na braki w danych niektóre państwa zostały pominięte w badaniu. Statystyki ONZ zawierają jednak oszacowane wartości zmiennych objaśniających, np. dla byłych republik radzieckich lub państw byłej Jugosławii (zdezagregowane), dla okresów sprzed odzyskania niepodległości. W analizie uwzględniono 182 państwa charakteryzowane czterema zmiennymi objaśniającymi:

- a) średnia dzietność kobiet,
- b) przeciętne trwanie życia,
- c) PKB *per capita* liczone według parytetu siły nabywczej,
- d) zgony niemowląt na 1000 urodzeń żywych.

Niemalże kompletne dane statystyczne dla wybranych cech dostępne są od roku 1973, dlatego ten rok przyjęto za podstawę porównania. Strukturę oraz jakość otrzymanej klasyfikacji porównano z wynikami dla roku 2008.

### 2.2. Wybór formuły normalizacyjnej i miary odległości

Wszystkie zmienne diagnostyczne są mierzone na skali ilorazowej, co uprawnia do zastosowania standaryzacji zerowanej jako formuły normalizacyjnej. Ponadto w przypadku każdej z metod taksonomicznych, która wymaga decyzji o wyborze sposobu pomiaru odległości, wybrano metrykę euklidesową. Ta sama metryka posłużyła również do wyznaczenia wartości indeksu sylwetkowego w etapie oceny jakości klasyfikacji.

### 2.3. Wybór metody klasyfikacji

Do przeprowadzenia analizy wykorzystano pięć różnych metod taksonomicznych: dwie metody aglomeracyjne (Warda i kompletnego połączenia), metody optymalizacyjne (metoda najmniejszych kwadratów i metoda najmniejszych kwadratów z wagami).

zacyjne (metodę  $k$ -medoidów, metodę  $k$ -średnich) [Kauffman, Rousseeuw 1990; Walesiak, Gatnar (red.) 2009] oraz taksonomiczną metodę wektorów nośnych (SVM) [Ben-Hur i in. 2001; Trzęsiok 2009]. Metody te zostały porównane pod względem jakości otrzymanego grupowania, użyteczności interpretacyjnej opisu skupień oraz zgodności wyników. Całość badania empirycznego zrealizowana została w programie statystycznym R.

#### 2.4. Ustalenie liczby klas

Sformułowana przez eksperta (prof. Roslinga) hipoteza badawcza stanowi merytoryczną podstawę do poszukiwania podziału badanych obiektów na dwa skupienia. Dodatkowo, dla porównania, zbudowano także modele realizujące grupowanie badanych państw na trzy skupienia.

W przypadku metody wektorów nośnych nie zadaje się *a priori* liczby klas, które mają być utworzone w procesie modelowania. Liczba zidentyfikowanych skupień zależy jednak bardzo silnie od doboru parametrów metody SVM. Do zbudowania modelu taksonomicznego metodą SVM posłużono się algorytmem zaproponowanym w pracy [Trzęsiok 2009], wykorzystującym indeks Randa jako miarę stabilności otrzymywanych podziałów dla różnych, losowo dobieranych, prób.

#### 2.5. Ocena wyników klasyfikacji

Do oceny jakości podziału zbioru państw na skupienia posłużono się indeksem sylwetkowym (Silhouette). Zaproponowane w pracy [Kauffman, Rousseeuw 1990] wartości progowe dla tego miernika pozwalają uznać, że odkryto poważną strukturę klas, gdy wartość indeksu sylwetkowego przekracza 0,5, silną zaś strukturę klas otrzymuje się, gdy indeks Silhouette jest większy od 0,7.

### 3. Wyniki analizy

Po zbudowaniu wielu modeli taksonomicznych, wykorzystując różne metody oraz dzieląc analizowany zbiór danych na dwa oraz trzy skupienia, obliczono indeksy sylwetkowe dla każdego wariantu (por. tab. 1). Wszystkie modele, w których zadeklarowano identyfikowanie trzech skupień, zostały odrzucone, gdyż wartość indeksu Silhouette w każdym z tych wariantów zawierała się między 0,36 a 0,48 (nie przekroczyła 0,5), co oznacza, że wykryto jedynie słabą strukturę klas i zalecane jest użycie innej metody grupowania. Metoda wektorów nośnych zbudowała model, który zidentyfikował dwa skupienia, ale do drugiego przydzieliła zaledwie trzy obserwacje. Model ten ze względu na bardzo małą wartość interpretacyjną został również odrzucony.

Bardzo zbliżone wartości uzyskano dla modeli aglomeracyjnych oraz optymalizacyjnych w przypadku grupowania na dwie klasy, z niewielką przewagą jakości modeli optymalizacyjnych. Dodatkowo warto zauważyć, że wprawdzie uznać nale-

ży, że modele te odkrywają poważną strukturę klas, to jednak wartość indeksu sylwetkowego jedynie nieznacznie przekroczyła 0,5.

**Tabela 1.** Wartości indeksu sylwetkowego (Silhouette) dla różnych wariantów modeli taksonomicznych

Rok	1973	2008	1973	2008	1973	2008	1973	2008
Indeks Silhouette	Metoda $k$ -medoidów				Metoda $k$ -średnich			
	$k = 2$		$k = 3$		$k = 2$		$k = 3$	
	0,5225	0,5301	0,3588	0,4785	0,5273	0,5254	0,4114	0,4845
Indeks Silhouette	Metoda Warda				Metoda kompletnego połączenia			
	$k = 2$		$k = 3$		$k = 2$		$k = 3$	
	0,4918	0,5272	0,3988	0,4795	0,4918	0,5272	0,3988	0,4795

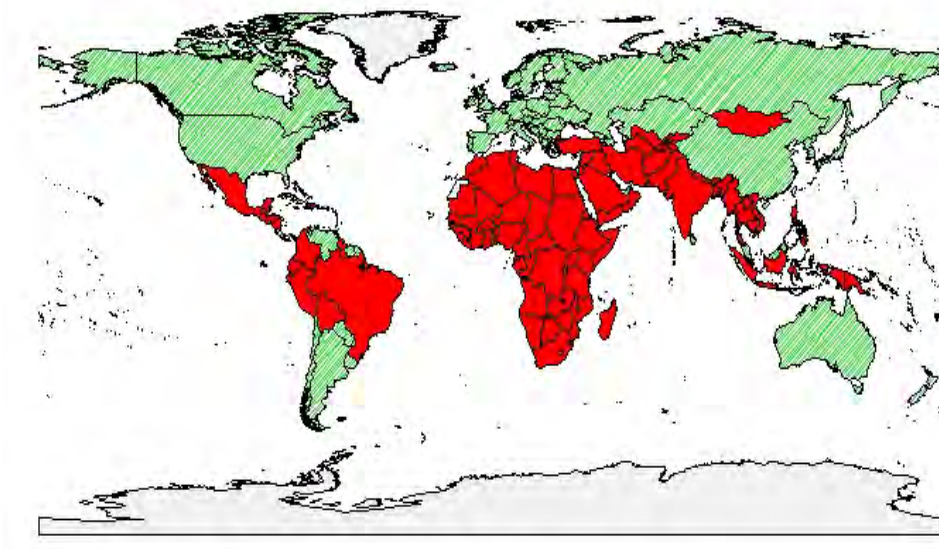
Źródło: opracowanie własne.

Do interpretacji przyjęto model optymalizacyjny zbudowany metodą  $k$ -medoidów (dla  $k = 2$ ). Wybór tej metody jest arbitralny, ale nie ma wielkiego wpływu na sformułowane dalej wnioski, gdyż indeks Randa dla wskazań klasyfikacji metodą  $k$ -medoidów i  $k$ -średnich jest równy 0,96, co oznacza, że modele te w 96% są zgodne we wskazywaniu przynależności do klas dla badanych obiektów. Na poziomie 97% kształtuje się zgodność modelu  $k$ -medoidów z modelami zbudowanymi metodami aglomeracyjnymi (Warda i kompletnego połączenia). Ze względu na relatywnie dużą liczbę obiektów, zamiast wymieniać nazwy państw przynależących do rozpoznanych dwóch skupień, posłużono się mapami świata (rys. 2 i 3). Obszary zakreskowane (jaśniejsze) stanowią pierwsze skupienie, natomiast obszary wypełnione całkowicie (ciemniejsze) – drugie skupienie.

Wykorzystując wykresy pudełkowe, można przejrzeć charakterystyki zmiennej w każdej z dwóch klas (por. rys. 4 i 5).

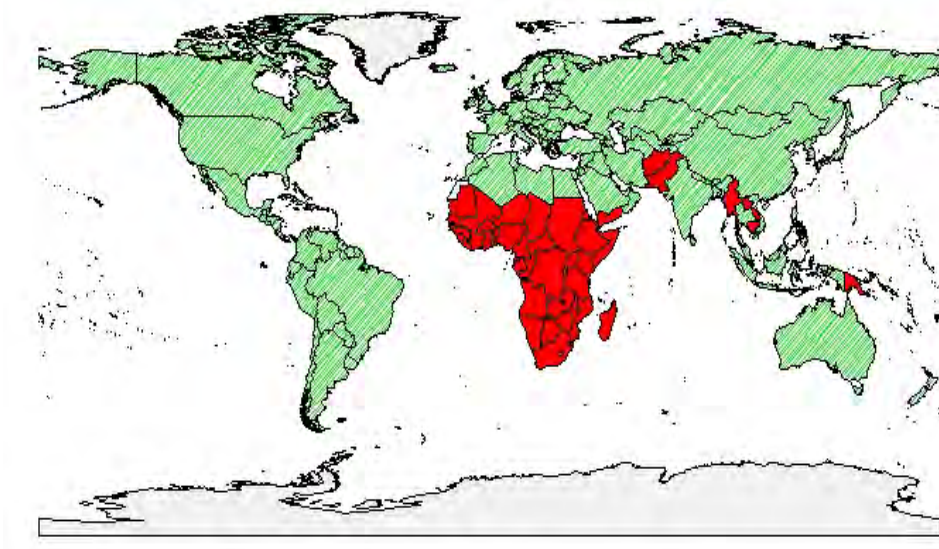
Z rysunku 4 odczytać można, że kraje przynależące w 1973 r. do pierwszej klasy charakteryzują się średnią dzietnością równą niecałe 3, przeciętną długością życia równą około 70 lat, PKB *per capita* równym około 10 000 \$, a około 40 dzieci na 1000 nie dożywa pierwszego roku życia. Drugie skupienie tworzą kraje, w których średnia dzietność kobiet jest równa niecałe 7, przeciętna długość życia to około 50 lat, PKB *per capita* jest równy 2000 \$ i około 120 dzieci na 1000 nie dożywa pierwszego roku życia. Z obserwacji tych wynika, że różnice w badanych charakterystykach między krajami przynależącymi do rozważanych dwóch skupień są znaczące. Drugiemu skupieniu można nadać etykietę „kraje Trzeciego Świata”.

Analizując przynależność obiektów do skupień wyznaczonych w 2008 r. (rys. 3), można odczytać, że zmiana polegała na tym, iż w przypadku 45 państw zaobserwowano migrację ze skupienia drugiego do pierwszego (w porównaniu z rokiem 1973). W grupie drugiej w 2008 r. nie ma już ani jednego państwa Ameryki Łacińskiej



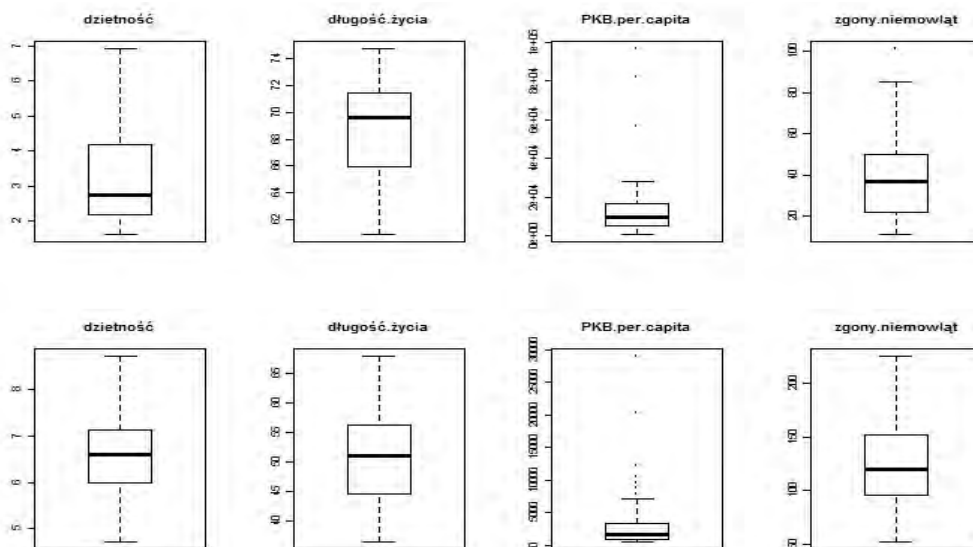
Rys. 2. Wynik grupowania 182 państw metodą  $k$ -medoidów dla  $k = 2$  według charakterystyk z 1973 r.

Źródło: opracowanie własne.



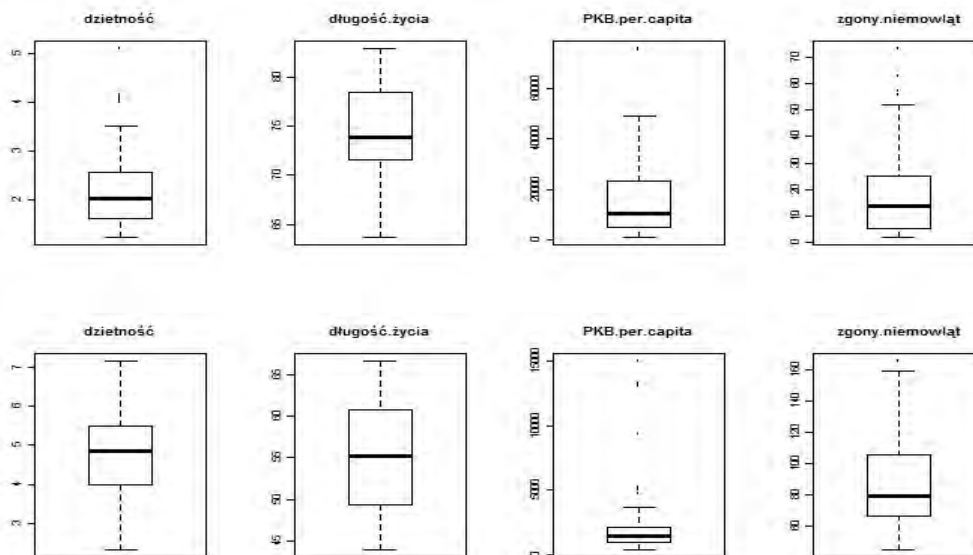
Rys. 3. Wynik grupowania 182 państw metodą  $k$ -medoidów dla  $k = 2$  według charakterystyk z 2008 r.

Źródło: opracowanie własne.



**Rys. 4.** Wykresy pudełkowe dla zmiennych opisujących państwa pierwszego skupienia (w pierwszej linii powyżej) oraz drugiego skupienia (w drugiej linii) dla modelu zbudowanego metodą  $k$ -medoidów dla  $k = 2$  według charakterystyk z 1973 r.

Źródło: opracowanie własne.



**Rys. 5.** Wykresy pudełkowe dla zmiennych opisujących państwa pierwszego skupienia (w pierwszej linii) oraz drugiego skupienia (w drugiej linii) dla modelu zbudowanego metodą  $k$ -medoidów dla  $k = 2$  według charakterystyk z 2008 r.

Źródło: opracowanie własne.

i północnej części Afryki. Z krajów Azji pozostało jedynie kilka państw, np. Afganistan, Pakistan, Laos, Kambodża, Birma. W związku z zaistniałym rozwojem i przesunięciami krajów z klasy drugiej do pierwszej nastąpiło zmniejszenie różnic między średnimi wartościami badanych charakterystyk w dwóch skupieniach.

#### 4. Podsumowanie

Funkcjonujący w świadomości wielu osób podział wykorzystujący kategorię „państwa Trzeciego Świata” staje się coraz mniej uprawniony. Nastąpiły bardzo wyraźne zmiany w sytuacji społeczno-ekonomicznej państw azjatyckich oraz Ameryki Łacińskiej, których nie można ignorować.

Różnice między badanymi charakterystykami w wyróżnionych skupieniach, średnio rzecz biorąc, zmniejszają się. Ponadto zarówno w badanym roku 1973, jak i 2008 ocena zwartości i separowalności zidentyfikowanych skupień jedynie nieznacznie przekroczyła próg, który przyjmuje się za dolną granicę, by można się było posługiwać odkrytą strukturą klas. W roku 1973 odległość między centrami skupień pozwalała wzmocnić wniosek i przyjąć za słuszne nadanie klasie drugiej etykiety „kraje Trzeciego Świata”. W 2008 r. ten argument znacznie osłabł. Przedstawione wnioski potwierdzają hipotezę sformułowaną na podstawie wizualizacji danych (wykresów bąbelkowych) i opinii prof. Roslinga.

Przedstawiona analiza nie prowadzi do wniosku, że w obrazie współczesnego świata nie ma już państw zacofanych lub źle wykorzystujących swój potencjał. Można natomiast odrzucić koncepcję wspólnej, jednakowej polityki pomocy względem tych państw, traktowanych jako jedna kategoria, względem której te same narzędzia pomocy gospodarczej będą jednakowo skuteczne. Wniosek końcowy może więc brzmieć tak: nie istnieją już dwie kategorie państw, między którymi jest przepaść (separowalność) w rozwoju cywilizacyjnym, natomiast istnieje silne zróżnicowanie, które objawia się pełnym spektrum krajów znajdujących się na różnym poziomie rozwoju społeczno-ekonomicznego.

#### Literatura

- Ben-Hur A., Horn D., Siegelman H.T., Vapnik V., *Support Vector Clustering*, „Journal of Machine Learning Research” 2001, 2, s. 125-137.
- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej, Wrocław 2004.
- Kauffman L., Rousseeuw P.J., *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley & Sons, New Jersey 1990.
- Lindstrand A., Bergstrom S., Rosling H., *Global Health: An Introductory Textbook*, Studentlitteratur AB, Lund 2006.
- Trzęsiok M., *Wybór wartości parametrów przez walidację wyników klasyfikacji taksonomicznej metody wektorów nośnych*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teo-*



*ria i zastosowania*, Taksonomia 15, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 7 (1207), Wrocław 2008, s. 354-363.

Trzęsiok M., *On Some Properties of Support Vector Clustering*, C. Domański, J. Białek (red.), Acta Universitatis Lodziensis, Folia Oeconomica 228, Multivariate Statistical Analysis – Statistical Inference, Statistical Models and Applications, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2009, s. 221-228.

Walesiak M., Gatnar E. (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009.

## WHERE IS THE THIRD WORLD? – CLUSTER ANALYSIS

**Summary:** The paper presents a clustering approach to the issue of identifying the countries of the third world. The analysis is based on the UN data sets and compares the classification of countries from 2008 and 35 years earlier. The main goal of the paper is to verify the hypothesis formulated by prof. Hans Rosling that the distinction and the use of the term “the Third World” is no longer justifiable, considering the significant changes in the socio-economic situation in Latin America, Asia and some African countries.