

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

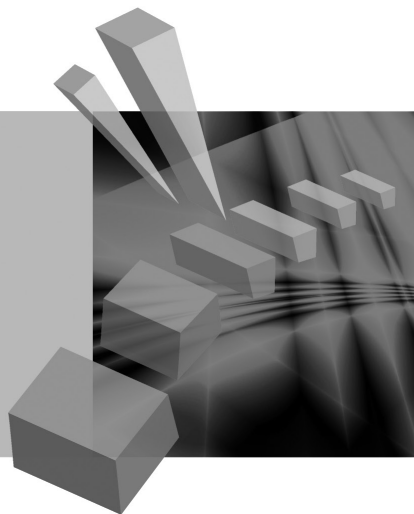
RESEARCH PAPERS

of Wrocław University of Economics

278

Taksonomia 20

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

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

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 2013

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Józef Pocięcha: Wskaźniki finansowe a klasyfikacyjne modele predykcji upadłości firm	15
Eugeniusz Gatnar: Analiza miar adekwatności rezerw walutowych	23
Marek Walesiak: Zagadnienie doboru liczby klas w klasyfikacji spektralnej	33
Joanicjusz Nazarko, Joanna Ejdyś, Anna Kononiuk, Anna M. Olszewska: Analiza strukturalna jako metoda klasyfikacji danych w badaniach foresight	44
Andrzej Bąk: Metody porządkowania liniowego w polskiej taksonomii – pakiet <code>pllord</code>	54
Aleksandra Łuczak, Feliks Wysocki: Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia	63
Ewa Roszkowska: Zastosowanie rozmytej metody TOPSIS do oceny ofert negocjacyjnych	74
Jacek Batóg: Analiza wrażliwości metody ELECTRE III na obserwacje nietypowe i zmianę wartości progowych	85
Jerzy Korzeniewski: Modyfikacja metody HINoV selekcji zmiennych w analizie skupień	93
Małgorzata Markowska, Danuta Strahl: Wykorzystanie referencyjnego systemu granicznego do klasyfikacji europejskiej przestrzeni regionalnej ze względu na filar inteligentnego rozwoju – kreatywne regiony	101
Elżbieta Sobczak: Inteligentne struktury pracujących a efekty strukturalne zmian zatrudnienia w państwach Unii Europejskiej.....	111
Elżbieta Gołata, Grażyna Dehnel: Rozbieżności szacunków NSP 2011 i BAEL.....	120
Iwona Foryś: Wykorzystanie analizy historii zdarzeń do badania powtórnego sprzedaży na lokalnym rynku mieszkaniowym	131
Hanna Dudek, Joanna Landmesser: Wpływ relatywnej deprivacji na subiektywne postrzeganie dochodów.....	142
Grażyna Łaska: Syntaksonomia numeryczna w klasyfikacji, identyfikacji i analizie przemian zbiorowisk roślinnych	151
Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analiza zależności między procesami fundamentalnymi a rynkiem kapitałowym w Chinach	161

Andrzej Bąk, Tomasz Bartłomowicz: Mikroekonometryczne modele wielomianowe i ich zastosowanie w analizie preferencji z wykorzystaniem programu R	169
Andrzej Dudek, Bartosz Kwaśniewski: Przetwarzanie równoległe algorytmów analizy skupień w technologii CUDA	180
Michał Trzęsiok: Wycena rynkowej wartości nieruchomości z wykorzystaniem wybranych metod wielowymiarowej analizy statystycznej	188
Joanna Trzęsiok: Wybrane symulacyjne techniki porównywania nieparametrycznych metod regresji.....	197
Artur Mikulec: Kryterium Mojeny i Wisharta w analizie skupień – przypadek skupień o różnych macierzach kowariancji	206
Artur Zaborski: Analiza <i>unfolding</i> z wykorzystaniem modelu grawitacji	216
Justyna Wilk: Identyfikacja obszarów problemowych i wzrostowych w województwie dolnośląskim w zakresie kapitału ludzkiego	225
Karolina Bartos: Analiza ryzyka odejścia studenta z uczelni po uzyskaniu dyplomu licencjata – zastosowanie sieci MLP	236
Ewa Genge: Segmentacja uczestników Industriady z wykorzystaniem analizy klas ukrytych	246
Izabela Kurzawa: Wielomianowy model logitowy jako narzędzie identyfikacji czynników wpływających na sytuację mieszkaniową polskich gospodarstw domowych	254
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modele eksploracji danych niezbilansowanych – procedury klasyfikacji dla zadania analizy ryzyka operacyjnego.....	262
Aleksandra Łuczak: Zastosowanie rozmytej hierarchicznej analizy w tworzeniu strategii rozwoju jednostek administracyjnych	271
Marcin Pelka: Rozmyta klasyfikacja spektralna <i>c</i> -średnich dla danych symbolicznych interwałowych.....	282
Małgorzata Machowska-Szewczyk: Klasyfikacja obiektów reprezentowanych przez różnego rodzaju cechy symboliczne	290
Ewa Chodakowska: Indeks Malmquista w klasyfikacji podmiotów gospodarczych według zmian ich względnej produktywności działania	300
Beata Bieszk-Stolorz, Iwona Markowicz: Wykorzystanie modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa do badania szansy podjęcia pracy w zależności od rodzaju bezrobocia	311
Marcin Salamaga: Weryfikacja teorii poziomu rozwoju gospodarczego J.H. Dunninga w ujęciu sektorowym w wybranych krajach Unii Europejskiej	321
Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Sytuacja społeczno-gospodarcza jako determinanta migracji wewnętrznych w Polsce.	330
Hanna Gruchociak: Delimitacja lokalnych rynków pracy w Polsce na podstawie danych z badania przepływów ludności związanych z zatrudnieniem	343

Radosław Pietrzyk: Efektywność inwestycji polskich funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych	351
Sabina Denkowska: Procedury testowań wielokrotnych	362

Summaries

Józef Pocięcha: Financial ratios and classification models of bankruptcy prediction	22
Eugeniusz Gatnar: Analysis of FX reserve adequacy measures	32
Marek Walesiak: Automatic determination of the number of clusters using spectral clustering	43
Joanicjusz Nazarko, Joanna Ejdys, Anna Kononiuk, Anna M. Olszewska: Structural analysis as a method of data classification in foresight research	53
Andrzej Bąk: Linear ordering methods in Polish taxonomy – pllord package	62
Aleksandra Łuczak, Feliks Wysocki: The application of spatial median of Weber and the method TOPSIS in positional formulation for the construction of synthetic measure of standard of living	73
Ewa Roszkowska: Application of the fuzzy TOPSIS method to the estimation of negotiation offers.....	84
Jacek Batóg: Sensitivity analysis of ELECTRE III method for outliers and change of thresholds	92
Jerzy Korzeniewski: Modification of the HINoV method of selecting variables in cluster analysis	100
Małgorzata Markowska, Danuta Strahl: Implementation of reference limit system for the European regional space classification regarding smart growth pillar – creative regions	110
Elżbieta Sobczak: Smart workforce structures versus structural effects of employment changes in the European Union countries	119
Elżbieta Gołata, Grażyna Dehnel: Divergence in National Census 2011 and LFS estimates.....	130
Iwona Foryś: Event history analysis in the resale study on the local housing market	141
Hanna Dudek, Joanna Landmesser: Impact of the relative deprivation on subjective income satisfaction	150
Grażyna Łaska: Numerical syntaxonomy in classification, identification and analysis of changes of secondary communities	160
Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analysis of relations between fundamental processes and capital market in China.....	166
Andrzej Bąk, Tomasz Bartłomowicz: Microeconomic polynomial models and their application in the analysis of preferences using R program.....	179

Andrzej Dudek, Bartosz Kwaśniewski: Parallel processing of clustering algorithms in CUDA technology	187
Michał Trzęsiok: Real estate market value estimation based on multivariate statistical analysis	196
Joanna Trzęsiok: On some simulative procedures for comparing nonparametric methods of regression.....	205
Artur Mikulec: Mojena and Wishart criterion in cluster analysis – the case of clusters with different covariance matrices	215
Artur Zaborski: Unfolding analysis by using gravity model	224
Justyna Wilk: Determination of problem and growth areas in Dolnośląskie Voivodship as regards human capital.....	235
Karolina Bartos: Risk analysis of bachelor students' university abandonment – the use of MLP networks	245
Ewa Genge: Clustering of industrial holiday participants with the use of latent class analysis.....	253
Izabela Kurzawa: Multinomial logit model as a tool to identify the factors affecting the housing situation of Polish households.....	261
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modelling class imbalance problems: comparing classification approaches for surgical risk analysis	270
Aleksandra Łuczak: The application of fuzzy hierarchical analysis to the evaluation of validity of strategic factors in administrative districts.....	281
Marcin Pełka: A spectral fuzzy c-means clustering algorithm for interval-valued symbolic data	289
Małgorzata Machowska-Szewczyk: Clustering algorithms for mixed-feature symbolic objects	299
Ewa Chodakowska: Malmquist index in enterprises classification on the basis of relative productivity changes	310
Beata Bieszk-Stolorz, Iwona Markowicz: Using proportional and non proportional Cox hazard models to research the chances for taking up a job according to the type of unemployment	320
Marcin Salamaga: Verification J.H. Dunning's theory of economic development by economic sectors in some EU countries	329
Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Socio-economic situation as a determinant of internal migration in Poland	342
Hanna Gruchociak: Delimitation of local labor markets in Poland on the basis of the employment-related population flows research.....	350
Radosław Pietrzyk: Selectivity and timing in Polish mutual funds performance measurement	361
Sabina Denkowska: Multiple testing procedures.....	369

Andrzej Bąk

Uniwersytet Ekonomiczny we Wrocławiu

METODY PORZĄDKOWANIA LINIOWEGO W POLSKIEJ TAKSONOMII – PAKIET PLLORD

Streszczenie: Celem artykułu jest prezentacja dorobku polskiej taksonomii w zakresie metod porządkowania liniowego. W szczególności w artykule przedstawiono następujące zagadnienia: charakterystyka wybranych metod porządkowania liniowego, implementacja wybranych metod porządkowania liniowego w pakiecie `pllord` programu R, przykłady zastosowań metod porządkowania liniowego z wykorzystaniem pakietu `pllord`.

Słowa kluczowe: taksonomia, metody porządkowania liniowego, program R.

1. Wstęp

Idea porządkowania liniowego obiektów wielowymiarowych opiera się na pojęciu porządkującej relacji binarnej (zwrotnej, antysymetrycznej, przechodniej i spójnej). Z aksjomatów tej relacji wynika, że jest możliwe stwierdzenie, który z dwóch dowolnych obiektów zbioru jest pierwszy (lepszy), a który drugi (gorszy), a także, czy są one identyczne [Bąk 1999].

Celem artykułu jest prezentacja dorobku polskiej taksonomii w zakresie metod porządkowania liniowego i propozycja pakietu programu R zawierającego implementację wybranych metod. W szczególności w artykule przedstawiono następujące zagadnienia: charakterystykę wybranych metod porządkowania liniowego, implementację wybranych metod porządkowania liniowego w pakiecie `pllord` programu R, przykłady zastosowań metod porządkowania liniowego z wykorzystaniem pakietu `pllord`.

Implementacja wybranych metod porządkowania liniowego w pakiecie `pllord` programu R może przyczynić się do popularyzacji tych metod na świecie. Program R jest niekomercyjnym projektem w zakresie analizy danych powszechnie wykorzystywanym m.in. w statystycznych i ekonometrycznych badaniach ekonomicznych.

2. Taksonomia, wielowymiarowa analiza porównawcza i porządkowanie liniowe

W literaturze przedmiotu jest wiele definicji i interpretacji pojęcia „taksonomia”. Stosowane są także terminy bliskoznaczne, takie jak taksologia, taksonometria, tak-

sonomia numeryczna, taksonomia matematyczna, klasyfikacja, analiza skupień, grupowanie, dyskryminacja, delimitacja, rozpoznawanie obrazów. Istnieje także rozróżnienie między taksonomią jakościową i ilościową oraz taksonomią opisową i stochastyczną [Pociecha 2008].

Rozwój taksonomii ilościowej zainicjował na początku XX w. polski antropolog Jan Czekanowski, proponując miarę odległości i diagraficzną metodę porządkowania macierzy odległości¹ [Pociecha i in. 1988, s. 13]. Zgodnie z definicją podaną w pracy [Grabiński, Wydymus, Zeliaś 1989, s. 9] „przez taksonomię rozumie się dyscyplinę naukową zajmującą się zasadami i procedurami klasyfikacji (porządkowania, grupowania, dyskryminacji, delimitacji, podziału)”. Celem badań taksonomicznych może być podział zbioru elementów, porządkowanie elementów zbioru, wybór elementów ze zbioru [Pociecha i in. 1988, s. 17].

Wielowymiarowa analiza porównawcza jest dyscypliną naukową umożliwiającą analizę obiektów i zjawisk złożonych, tj. takich, na których stan i zachowanie wpływa jednocześnie wiele cech (zmiennych) i czynników. Zwięzła definicja podana przez Z. Hellwiga mówi, iż „(...) metody i technika porównywania obiektów wielocechowych nazywają się wielowymiarową analizą porównawczą (...)” [Hellwig 1981, s. 48].

W literaturze przedmiotu przeważa pogląd, iż taksonomia i wielowymiarowa analiza porównawcza stanowią odrębne dyscypliny badawcze [Pociecha i in. 1988, s. 16-17]. Podstawowym celem wielowymiarowej analizy porównawczej jest konstrukcja miary syntetycznej umożliwiającej porównywanie elementów zbioru (obiektów) opisanych za pomocą wielu zmiennych (cech). Do osiągnięcia tego celu wykorzystuje się często metody porządkowania liniowego. W metodach porządkowania liniowego zakłada się, że rezultatem uzyskanym na skutek zastosowania odpowiedniego algorytmu będzie takie uszeregowanie zbioru obiektów, w którym [Grabiński 1992], s. 135]:

- każdy obiekt ma przynajmniej jednego sąsiada oraz nie więcej niż dwóch sąsiadów,
- jeżeli obiekt a jest sąsiadem obiektu b , to obiekt b jest sąsiadem obiektu a ,
- istnieją tylko dwa obiekty mające jednego sąsiada.

Metody porządkowania liniowego prowadzą do ustalenia kolejności obiektów ze względu na jedną agregatową cechę, która jest syntetycznym reprezentantem wielu cech opisujących porządkowane obiekty. W ujęciu geometrycznym metody porządkowania liniowego prowadzą do rzutowania punktów reprezentujących obiekty umieszczone w wielowymiarowej przestrzeni cech na prostą, która reprezentuje cechę agregatową.

¹ Na stronie internetowej <http://eskimo73.republika.pl/maczek.html> dostępny jest program komputerowy MaCzek będący implementacją metody Czekanowskiego.

3. Wybrane metody porządkowania liniowego

Podstawowe kategorie w porządkowaniu liniowym to obiekty i cechy. Zarówno obiekty, jak i cechy posiadają swoje liczbowe reprezentacje (obrazy), co umożliwia ich wszechstronne analizowanie za pomocą algorytmów numerycznych. Porządkowanie liniowe ma charakter wartościujący (oceniający) elementy (obiekty) i znajduje zastosowanie w różnych obszarach badań, w tym również w ekonomii.

Przedmiotem porządkowania liniowego mogą być takie obiekty, jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe). Takie charakterystyki, jak poziom rozwoju gospodarczego, kondycja finansowa, walory użytkowe, są zmiennymi, których realizacje nie są bezpośrednio mierzalne. Zmienne te są agregatami, których wartości są generowane przez obserwacje cech, które są bezpośrednio mierzalne. Zmienne agregatowe w sposób syntetyczny charakteryzują obiekty wielowymiarowe. Realizacje zmiennych agregatowych są generowane przez określone funkcje agregujące, których postać analityczna może być różna (np. metody bezwzorcowe i wzorcowe). Zmienne agregatowe należą do najczęściej wykorzystywanych metod porządkowania liniowego w zastosowaniach ekonomicznych. Zmienne agregatowe są nazywane także zmiennymi syntetycznymi, miarami syntetycznymi, syntetycznymi miarami rozwoju, taksonomicznymi miernikami rozwoju.

W literaturze przedmiotu poświęconej metodom porządkowania liniowego, która w dużej mierze jest dorobkiem polskiej myśli statystycznej, spotkać można wiele propozycji konstrukcji zmiennych syntetycznych. Pierwsza propozycja miary syntetycznej umożliwiającej porządkowanie liniowe obiektów z wykorzystaniem wzorca została przedstawiona przez Z. Hellwiga w 1968 r. pod nazwą „miara rozwoju gospodarczego” [Hellwig 1968].

Propozycja Z. Hellwiga stała się inspiracją do dalszych badań w zakresie metod porządkowania liniowego, które zaowocowały zarówno modyfikacjami miary rozwoju gospodarczego, jak i nowymi oryginalnymi propozycjami konstrukcji zmiennych syntetycznych. Jako przykłady wymienić można m.in.: absolutny miernik rozwoju M. Cieślak [1974], zmienną syntetyczną S. Bartosiewicz [1976], zmodyfikowaną miarę rozwoju gospodarczego W. Pluty [1976], miarę syntetyczną D. Strahl [1978], agregatową miarę rozwoju T. Borysa [1978b], zmienną syntetyczną (taksonomiczny miernik rozwoju) E. Nowaka [1984], syntetyczny miernik rozwoju (wykorzystujący uogólnioną miarę odległości) M. Walesiaka² [1993].

² Wzorcową metodą porządkowania liniowego (syntetyczny miernik rozwoju) wykorzystującą uogólnioną miarę odległości (GDM – *Generalized Distance Measure*) zaproponowaną przez M. Walesiaka w 1993 r. jest oprogramowana w pakiecie `clusterSim` programu R [Walesiak 2011; Walesiak, Dudek 2012]).

Podstawowe etapy w procedurze porządkowania liniowego to: określenie charakteru zmiennych (stymulanty, nominanty, destymulanty)³, wyznaczenie wag zmiennych, normalizacja zmiennych, wyznaczenie współrzędnych wzorca w przypadku agregacji wzorcowej, agregacja bezwzorcowa lub wzorcowa [Grabiński 1984; Bąk 1999].

4. Pakiet `pllord`

W aktualnej wersji pakietu `pllord`⁴ oprogramowane są następujące miary syntetyczne:

1) miara rozwoju gospodarczego Z. Hellwiga (1968 r.):

a) normalizacja zmiennych (standaryzacja): $z_{ij} = \frac{\bar{x}_j - x_{ij}}{s_j}$, x_{ij} – obserwacja j -tej zmiennej dla obiektu i , \bar{x}_j – średnia arytmetyczna obserwacji j -tej zmiennej, s_j – odchylenie standardowe obserwacji j -tej zmiennej;

b) współrzędne wzorca: $z_{0j} = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych stymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych destymulant} \end{cases}$;

c) odległości obiektów od wzorca: $d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2}$;

d) miara rozwoju gospodarczego q_i (na ogół $q_i \in [0;1]$) – $q_i = 1 - \frac{d_{i0}}{d_0}$,

$$d_0 = \bar{d}_0 + 2s_d, \quad \bar{d}_0 = \frac{1}{n} \sum_{i=1}^n d_{i0}, \quad s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_{i0} - \bar{d}_0)^2}.$$

2) zmienna syntetyczna S. Bartosiewicz (1976 r.):

a) współrzędne punktu odniesienia – $x_{0j} = \min_i \{x_{ij}\}$;

b) normalizacja zmiennych – $z_{ij} = \frac{x_{ij}}{x_{0j}}$;

c) zmienna syntetyczna – $q_i = \sum_{j=1}^m z_{ij}$.

³ Do literatury przedmiotu pojęcia zmiennej stymulandy i destymulandy zostały wprowadzone przez Z. Hellwiga [1968], a pojęcie zmiennej nominandy przez T. Borysa [1978a].

⁴ Pakiet `pllord` jest dostępny na stronie internetowej <http://keii.ue.wroc.pl/pllord>.

3) miara syntetyczna D. Strahl (1978 r.):

a) współrzędne punktu odniesienia –

$$x_{0j} = \begin{cases} \max_i \{x_{ij}\} & \text{dla zmiennych stymulant} \\ \min_i \{x_{ij}\} & \text{dla zmiennych destymulant} \end{cases};$$

b) normalizacja zmiennych – $z_{ij} = \begin{cases} \frac{x_{ij}}{x_{0j}} & \text{dla zmiennych stymulant} \\ \frac{x_{0j}}{x_{ij}} & \text{dla zmiennych destymulant} \end{cases};$

c) miara syntetyczna – $q_i = \sum_{j=1}^m z_{ij}$.

4) zmienna syntetyczna E. Nowaka (1984 r.):

a) współrzędne punktu odniesienia – $x_{0j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$, $x_{0j} = \min_i \{x_{ij}\}$,

$$x_{0j} = \max_i \{x_{ij}\}, x_{0j} = \max \{x_{ij}\} - \min \{x_{ij}\}, x_{0j} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, x_{0j} = \sqrt{\sum_{i=1}^n x_{ij}^2};$$

b) normalizacja zmiennych – $z_{ij} = \frac{x_{ij}}{x_{0j}};$

c) zmienna syntetyczna – $q_i = \sum_{j=1}^m z_{ij}$ lub $q_i = \frac{1}{n} \sum_{j=1}^m z_{ij}$.

W tabeli 1 zestawiono funkcje pakietu `pllord` i ich przeznaczenie.

Tabela 1. Funkcje pakietu `pllord`

Lp.	Funkcja	Przeznaczenie
1	<code>hellwig()</code>	Oblicza wartości miary rozwoju gospodarczego Z. Hellwiga [1968]
2	<code>bartosiewicz()</code>	Oblicza wartości zmiennej syntetycznej S. Bartosiewicz [1976]
3	<code>strahl()</code>	Oblicza wartości miary syntetycznej D. Strahl [1978]
4	<code>nowak()</code>	Oblicza wartości zmiennej syntetycznej E. Nowaka [1984]
5	<code>normalization()</code>	Oblicza macierz zmiennych znormalizowanych z wykorzystaniem wskazanej metody normalizacji
6	<code>upperpole()</code>	Oblicza wartości wektora zawierającego współrzędne wzorca w metodzie Hellwiga
7	<code>euclideanist()</code>	Oblicza wartości wektora zawierającego odległości od wzorca w metodzie Hellwiga
8	<code>distnorm()</code>	Oblicza wartości wektora zawierającego znormalizowane odległości od wzorca w metodzie Hellwiga
9	<code>svord()</code>	Sortuje malejąco wartości zmiennej syntetycznej i opcjonalnie nazwy obiektów

Źródło: opracowanie własne.

Dla zilustrowania sposobu użycia wybranych funkcji pakietu `pllord` wykorzystano metodę porządkowania liniowego zaproponowaną przez Z. Hellwiga oraz dane zamieszczone w pracy [Hellwig 1968]. Porządkowanymi obiektami jest 15 krajów charakteryzowanych przez 6 zmiennych (X3-X6 w przeliczeniu na 10 000 osób): X1 – przeciętne trwanie życia mężczyzn, X2 – procent ludności zawodowo czynnej w rolnictwie, X3 – kadry inżynieryjno-techniczne, X4 – kadry ekonomiczno-administracyjne, X5 – personel urzędniczy, X6 – personel handlowy.

Fragment zbioru danych:

```
> head(hdane68)
      Kraj  X1  X2  X3  X4  X5  X6
1   Belgia 62.0 6.2 306.18 100.58 432.06 418.49
2    Dania 70.4 17.5 358.68 76.58 400.97 448.24
3 Finlandia 64.9 35.5 376.32 75.50 237.28 316.97
4   Grecja 67.5 53.9 148.96 31.67 171.41 264.17
5 Holandia 71.4 10.7 332.62 112.30 448.29 343.49
6    Indie 45.2 72.9 73.70 41.26 72.82 156.62
```

Za pomocą skryptu 1 otrzymano wartości miary rozwoju gospodarczego Z. Hellwiga [1968] z wykorzystaniem funkcji `hellwig(X, vc, objn)`. Znaczenie parametrów tej funkcji zawiera tab. 2.

Skrypt 1.

```
> library(pllord)
> data(hdane68)
> X<-hdane68[2:7]
> objn<-hdane68[,1]
> vc<-c(1,-1,1,1,1,1)
> hellwig(X,vc,objn)
```

Otrzymane wartości miary rozwoju gospodarczego Z. Hellwiga:

	Object	Distance
1	Belgia	0.51395752
2	Dania	0.54225684
3	Finlandia	0.40343640
4	Grecja	0.20198009
5	Holandia	0.54274018
6	Indie	-0.07043908
7	Japonia	0.54978591
8	Jugosławia	0.16708904
9	Kanada	0.59597175
10	Norwegia	0.46083166
11	Portugalia	0.17805069
12	USA	0.64844775
13	Szwajcaria	0.52413369
14	Szwecja	0.58302556
15	Węgry	0.33964491

Za pomocą skryptu 2 otrzymano wyniki uporządkowania liniowego obiektów w kolejności od najbardziej do najmniej rozwiniętego oraz ich ilustrację graficzną (rys. 1) na podstawie miary rozwoju gospodarczego Z. Hellwiga [1968]. Wykorzystano funkcje 1 i 5-9 z tab. 1. Przeznaczenie parametrów tych funkcji zawiera tab. 2.

Tabela 2. Wybrane funkcje pakietu `pllord` i ich parametry

Funkcja	Parametry
<code>hellwig(X, vc, objn)</code>	X – zbiór danych; vc – wektor reprezentujący charakter zmiennych (1 – zmienna stymulanta, -1 – zmienna destymulanta), objn – wektor z nazwami obiektów
<code>normalization(X, "n1")</code>	X – zbiór danych; n1 – standaryzacja
<code>upperpole(Z, vc)</code>	Z – macierz danych standaryzowanych, vc – jw.
<code>euclidean(Z, z0)</code>	Z – jw.; z0 – współrzędne wzorca rozwoju
<code>distnorm(d0)</code>	d0 – miara rozwoju gospodarczego
<code>distord(d, objn, ord=TRUE)</code>	d – wektor odległości od wzorca; objn – wektor z nazwami obiektów; ord – zmienna logiczna wskazująca, czy obiekty mają być uporządkowane malejąco

Źródło: opracowanie własne.

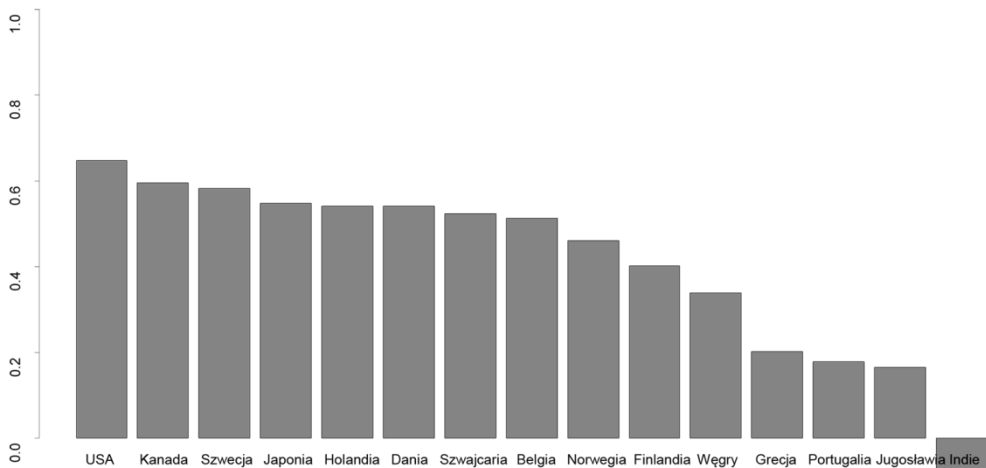
Skrypt 2.

```
library(pllord)
data(hdane68)
X<-hdane68[2:7]
objn<-hdane68[,1]
vc<-c(1,-1,1,1,1,1)
Z<-normalization(X,"n1")
z0<-upperpole(Z,vc)
d0<-euclidean(Z,z0)
d<-distnorm(d0)
distord(d,objn,ord=TRUE)
q<-distord(d,objn,ord=TRUE)
windows(width=14,height=7,pointsize=7)
barplot(q[,2],ylim=c(0,1),names.arg=q[,1])
```

Otrzymane wyniki porządkowania liniowego metodą Z. Hellwiga:

	Object	Distance
1	USA	0.64844775
2	Kanada	0.59597175
3	Szwecja	0.58302556
4	Japonia	0.54978591
5	Holandia	0.54274018
6	Dania	0.54225684
7	Szwajcaria	0.52413369
8	Belgia	0.51395752
9	Norwegia	0.46083166

10	Finlandia	0.40343640
11	Węgry	0.33964491
12	Grecja	0.20198009
13	Portugalia	0.17805069
14	Jugosławia	0.16708904
15	Indie	-0.07043908



Rys. 1. Graficzna prezentacja wyników porządkowania liniowego

Źródło: opracowanie własne.

5. Podsumowanie

Metody porządkowania liniowego, zajmujące ważne miejsce w dorobku polskiej taksonomii, znajdują praktyczne zastosowania w wielu dziedzinach badań, w tym w ekonomii. Ich wykorzystanie w badaniach empirycznych nie jest możliwe bez odpowiedniego oprogramowania komputerowego, a w komercyjnych pakietach statystyczno-ekonometrycznych metody te nie są uwzględniane. Implementacja metod porządkowania liniowego w pakiecie programu R może przyczynić się do ich popularyzacji na świecie, ponieważ program ten jest niekomercyjnym projektem w zakresie analizy danych powszechnie stosowanym m.in. w statystycznych i ekonometrycznych badaniach ekonomicznych. W aktualnej wersji pakietu `pllord` opracowane zostały wybrane metody, w tym historycznie pierwsza metoda Z. Hellwiga [1968]. W opracowywaniu znajdują się kolejne metody porządkowania liniowego zaproponowane w polskojęzycznej literaturze przedmiotu.

Literatura

- Bartosiewicz S. [1976], *Propozycja metody tworzenia zmiennych syntetycznych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 84.
- Bąk A. [1999], *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Borys T. [1978a], *Metody normowania cech w statystycznych badaniach porównawczych*, „Przegląd Statystyczny”, z. 2, s. 227-239.
- Borys T. [1978b], *Propozycja agregatowej miary rozwoju obiektów*, „Przegląd Statystyczny” z. 3, s. 371-381.
- Cieślak M. [1974], *Taksonomiczna procedura prognozowania rozwoju gospodarczego i określania potrzeb na kadry kwalifikowane*, „Przegląd Statystyczny” z. 1, s. 29-39.
- Gabiński T. [1984], *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie. Seria specjalna: Monografie nr 61.
- Gabiński T. [1992], *Metody taksonometrii*, Akademia Ekonomiczna w Krakowie, Kraków.
- Gabiński T., Wydymus S., Zeliaś A. [1989], *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa.
- Hellwig Z. [1968], *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, „Przegląd Statystyczny”, z. 4, s. 307-327.
- Hellwig Z. [1981], *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych*, [w:] W. Welfe (red.), *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, PWE, Warszawa.
- Nowak E. [1984], *Problemy doboru zmiennych do modelu ekonometrycznego*, PWN, Warszawa.
- Pluta W. [1976], *Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego*, „Przegląd Statystyczny” z. 4, s. 511-517.
- Pociecha J. [2008], *Rozwój metod taksonomicznych i ich zastosowań w badaniach społeczno-ekonomicznych*, <http://www.stat.gov.pl/cps/rde/xbcr/gus/>.
- Pociecha J., Podolec B., Sokołowski A., Zając K. [1988], *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa.
- R Development Core Team [2012], *R: A Language and Environment for Statistical Computing*, *R Foundation for Statistical Computing*, <http://cran.r-project.org>.
- Strahl D. [1978], *Propozycja konstrukcji miary syntetycznej*, „Przegląd Statystyczny”, z. 2, s. 205-215.
- Walesiak M. [1993], *Statystyczna analiza wielowymiarowa w badaniach marketingowych*. Wrocław, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 654, Seria: Monografie i Opracowania nr 101.
- Walesiak M. [2011], *Uogólniona miara odległości w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Walesiak M., Dudek A. [2012], *clusterSim package*, <http://www.R-project.org>.

LINEAR ORDERING METHODS IN POLISH TAXONOMY – PLLORD PACKAGE

Summary: The main aim of this article is the presentation achievements of Polish taxonomy. The article especially presents the following issues: characteristics of chosen linear ordering methods, implementation of chosen linear ordering methods in `pllord` R program package, examples of applications of linear ordering methods using `pllord` package.

Keywords: taxonomy, linear ordering methods, R program.