PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 384

Taksonomia 24
Klasyfikacja i analiza danych – teoria i zastosowania

Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 384 • 2015
Taksonomia 24 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Andrzej Bąk
Uniwersytet Ekonomiczny we Wrocławiu
e-mail: andrzej.bak@ue.wroc.pl

ZAGADNIENIE WYBORU OPTYMALNEJ PROCEDURY PORZĄDKOWANIA LINIOWEGO W PAKIECIE `PLLORD`

Streszczenie: Metody porządkowania liniowego są wykorzystywane w badaniach zmierzających do ustalenia kolejności lub klasyfikacji obiektów. Przedmiotem porządkowania liniowego mogą być np. takie obiekty, jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe) itp. Na gruncie badań taksonomicznych opracowano wiele procedur porządkowania liniowego. Różnią się one m.in. metodami wyznaczania wag zmiennych, metodami normalizacji zmiennych oraz metodami szacowania wartości zmiennych syntetycznych. W związku z tym pojawia się problem wyboru optymalnej procedury do analizy danych empirycznych o określonych charakterystykach statystycznych. Celem artykułu jest zarys problematyki wyboru optymalnej konfiguracji elementów składowych procedury porządkowania liniowego z wykorzystaniem wybranych mierników oceny jakości i programu R.

Słowa kluczowe: porządkowanie liniowe, optymalna procedura, program R.

DOI: 10.15611/pn.2015.384.03

1. Wstęp

Metody porządkowania liniowego są wykorzystywane w badaniach ekonomicznych w celu ustalenia kolejności lub klasyfikacji obiektów, takich jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe) itp.

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. Przedmiotem porządkowania liniowego mogą być np. takie obiekty, jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory

użytkowe) itp. Takie charakterystyki, jak poziom rozwoju gospodarczego, kondycja finansowa, walory użytkowe są zmiennymi, których realizacje nie są bezpośrednio mierzalne. Zmienne takie są agregatami, których wartości są generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne (funkcje agregujące mogą mieć różną postać analityczną). Uzyskane realizacje zmiennej syntetycznej umożliwiają uporządkowanie obiektów wielowymiarowych w sensie relacji preferencji (dominacji).

Celem artykułu jest wprowadzenie do problematyki wyboru optymalnej konfiguracji elementów składowych procedury porządkowania liniowego, które obejmują wybór współczynników wagowych, normalizację zmiennych, formuły wyznaczania zmiennych syntetycznych, w zależności od danych empirycznych. Optymalna konfiguracja jest wybierana na podstawie mierników oceny jakości metod porządkowania liniowego. W aktualnej wersji pakietu `pllord` wybór optymalnej procedury porządkowania liniowego dotyczy formuł wyznaczania zmiennych syntetycznych.

W pracy przedstawiono implementację wybranych metod porządkowania liniowego oraz wybranych mierników oceny jakości tych metod w pakiecie `pllord` programu R. Aktualna wersja pakietu jest rozwinięciem pierwszej propozycji prezentowanej na Konferencji SKAD w 2012 roku [Bąk 2013]. Program R jest niekomercyjnym projektem o zasięgu światowym w zakresie analizy danych powszechnie wykorzystywanym m.in. w statystycznych i ekonometrycznych badaniach ekonomicznych [R Development Core Team 2014].

2. Metody porządkowania liniowego

Metody porządkowania liniowego, mieszczące się w obrębie wielowymiarowej analizy porównawczej, są w dużej mierze dorobkiem polskiej myśli statystycznej i ekonometrycznej. Pierwszą propozycję przedstawił Z. Hellwig w pracy [Hellwig 1968]. Publikacja ta zainicjowała intensywne badania w tym zakresie, których efektem były kolejne propozycje metod porządkowania liniowego zamieszczone m.in. w pracach [Cieślak 1974; Bartosiewicz 1976; Pluta 1976; Strahl 1978; Borys 1978b; Nowak 1984; Walesiak 1993][1].

Podstawą porządkowania liniowego jest zmienna syntetyczna[2], której wartości są szacowane na podstawie obserwacji zmiennych diagnostycznych opisujących badane obiekty. Zakłada się, że wartości zmiennej syntetycznej, oszacowane za

[1] Wzorcowa metoda porządkowania liniowego wykorzystująca 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]).

[2] W literaturze przedmiotu spotkać można inne określenia zmiennej syntetycznej, takie jak np.: zmienna agregatowa, miara syntetyczna, syntetyczna miara rozwoju, taksonomicznych miernik rozwoju, agregatowa miara rozwoju, miara rozwoju gospodarczego.

pomocą określonej procedury, umożliwiają takie uporządkowanie 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.

Zmienna syntetyczna ma charakter zmiennej ukrytej, ponieważ jej realizacje nie są bezpośrednio obserwowane. Realizacje te są natomiast generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne. Realizacje zmiennej syntetycznej są szacowane za pomocą funkcji agregujących, których postać analityczna może być różna. Rozróżnia się dwie podstawowe grupy metod, które są wykorzystywane do szacowania wartości zmiennej syntetycznej: metody bezwzorcowe i metody wzorcowe.

W procedurze porządkowania liniowego wyróżnia się takie etapy postępowania, jak: 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; 2013].

W aktualnej wersji pakietu `pllord` programu R uwzględniono następujące elementy procedur porządkowania liniowego:

1. Przyjęto, że charakter zmiennych jest identyfikowany na podstawie oceny merytorycznej[4]. Przyjmuje się, że zmienna jest stymulantą, jeżeli jej rosnące wartości wpływają korzystnie na ocenę obiektu. Jeżeli rosnące wartości zmiennej wpływają niekorzystnie na pozycję obiektu, to przyjmuje się, że zmienna ma charakter destymulanty. Zmienna nominata natomiast przyjmuje do pewnego progu wartości wpływające korzystnie na pozycję obiektu, zaś po jego przekroczeniu wartości te wpływają niekorzystnie na ocenę badanego obiektu.

2. Przyjęto, że wagi wszystkich zmiennych są jednakowe. Postulat może być spełniony np. na podstawie przekształcenia:

$$w_j = \frac{1}{m}, \tag{1}$$

gdzie: w_j – waga j-tej zmiennej; m – liczba zmiennych.

W literaturze przedmiotu prezentowane są zróżnicowane stanowiska dotyczące ważenia zmiennych i zagadnienie to nie jest jednoznacznie rozstrzygnięta. Przytaczane są argumenty zarówno za ważeniem, jak i przeciw ważeniu zmiennych. Inny

[3] Pojęcia zmiennej stymulanty i destymulanty zostały wprowadzone do literatury przedmiotu przez Z. Hellwiga [1968], a pojęcie zmiennej nominanty przez T. Borysa [1978a].

[4] W literaturze przedmiotu proponowane są także statystyczne metody identyfikacji charakteru zmiennych, np.: [Grabiński 1984; Bąk 1999].

problem dotyczy sposobu ustalenia wag zmiennych. Wagi mogą być szacowane na podstawie merytorycznych ocen ekspertów lub za pomocą metod statystycznych.

3. Normalizację zmiennych przeprowadzono metodą standaryzacji:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \tag{2}$$

gdzie: z_{ij} – znormalizowana (standaryzowana) wartość *j*-tej zmiennej dla *i*-tego obiektu; $\bar{x}_j$ – średnia arytmetyczna wartości *j*-tej zmiennej; s_j – odchylenie standardowe *j*-tej zmiennej.

4. Uwzględniono dwie bezwzorcowe metody szacowania wartości zmiennej syntetycznej [Grabiński 1984; Bąk 1999]:

a) metodę bezwzorcową wykorzystującą średnią arytmetyczną:

$$q_i = \sum_{j=1}^{m} w_j z_{ij}, \tag{3}$$

gdzie: q_i – wartość zmiennej syntetycznej dla *i*-tego obiektu.

b) metodę bezwzorcową wykorzystującą średnią harmoniczną:

$$q_i = 1 / \sum_{j=1}^{m} \frac{w_j}{z_{ij}}. \tag{4}$$

5. Uwzględniono siedem wzorcowych metod szacowania wartości zmiennej syntetycznej bazujących na miarach odległości [Grabiński 1984; Bąk 1999]:

a) metodę opartą na odległości Hamminga:

$$q_i = w_j \left| z_{ij} - z_{0j} \right|, \tag{5}$$

gdzie: $z_{0j} = \max\{z_{ij}\}$ – współrzędne wzorca rozwoju (górny biegun znormalizowanych wartości zmiennych diagnostycznych).

b) metodę opartą na odległości Euklidesa:

$$q_i = \left[\sum_{j=1}^{m} w_j \left(z_{ij} - z_{0j} \right)^2 \right]^{1/2}, \tag{6}$$

c) metodę opartą na odległości Jeffreysa-Matusita:

$$q_i = \sum_{j=1}^{m} w_j \left(\sqrt{z_{ij}} - \sqrt{z_{0j}} \right)^2, \tag{7}$$

d) metodę opartą na odległości Braya-Curtisa:

$$q_i = \frac{\sum_{j=1}^{m} w_j \left| z_{ij} - z_{0j} \right|}{\sum_{j=1}^{m} w_j \left(z_{ij} + z_{0j} \right)}, \tag{8}$$

e) metodę opartą na odległości Clarka:

$$q_i = \left[\sum_{j=1}^{m} w_j \left(\frac{z_{ij} - z_{0j}}{z_{ij} + z_{0j}} \right)^2 \right]^{1/2}, \tag{9}$$

f) metodę opartą na odległości „Canberra”:

$$q_i = \sum_{j=1}^{m} w_j \frac{|z_{ij} - z_{0j}|}{(z_{ij} + z_{0j})}, \text{ przy czym gdy } z_{ij} = z_{0j}, \text{ to } q_i = 0, \tag{10}$$

g) metodę opartą na odległości kątowej:

$$q_i = 1 - \frac{\sum_{j=1}^{m} w_j z_{ij} z_{0j}}{\left[\sum_{j=1}^{m} w_j (z_{ij})^2 \sum_{j=1}^{m} w_j (z_{0j})^2\right]^{1/2}}. \tag{11}$$

3. Mierniki oceny jakości procedur porządkowania liniowego

Konfiguracje metod porządkowania liniowego uwzględniające różne wagi zmiennych, metody normalizacji i metody szacowania wartości zmiennej syntetycznej prowadzą na ogół do różnego uporządkowania badanych obiektów. W związku z tym powstaje problem wyboru najlepszej konfiguracji w odniesieniu do empirycznego (lub symulacyjnego) zbioru danych o określonych charakterystykach statystycznych (wektor średnich, macierz kowariancji). Ocenę jakości zmiennej syntetycznej można przeprowadzić na podstawie mierników proponowanych w literaturze przedmiotu z obszaru taksonomii [Grabiński 1984; Grabiński, Wydymus, Zeliaś 1989].

Mierniki jakości wykorzystywane do wyboru optymalnej konfiguracji dotyczą w szczególności pomiaru i oceny takich własności procedur porządkowania liniowego, jak [Grabiński 1984; Grabiński, Wydymus, Zeliaś 1989; Bąk 1999]:

- zgodność odwzorowania, mierzona wskaźnikiem zróżnicowania odległości między obiektami w przestrzeni zmiennych diagnostycznych oraz w przestrzeni zmiennej syntetycznej,
- korelacja liniowa pomiędzy zmienną syntetyczną a zmiennymi diagnostycznymi, mierzona przeciętnym współczynnikiem „nieokreśloności” oraz współczynnikiem „jednoznaczności” zmiennej syntetycznej,
- korelacja rangowa zmiennej syntetycznej ze zmiennymi diagnostycznymi, mierzona współczynnikiem „nieokreśloności”, współczynnikiem „jednoznaczności” zmiennej syntetycznej oraz uogólnionym rangowym współczynnikiem rozbieżności,
- zmienność i koncentracja zmiennej syntetycznej, mierzone współczynnikiem obliczonym dla realizacji zmiennej syntetycznej oraz dla pierwszych różnic uporządkowanych niemalejąco wartości zmiennej syntetycznej,
- przeciętna odległość taksonomiczna zmiennej syntetycznej od zmiennych diagnostycznych mierzona na podstawie mierników Hamminga oraz Euklidesa.

W aktualnej wersji pakietu `pllord` programu R uwzględniono następujące mierniki oceny jakości procedur porządkowania liniowego: miernik zgodności odwzorowania, miernik korelacji liniowej zmiennej syntetycznej ze zmiennymi

diagnostycznymi, miernik korelacji rangowej zmiennej syntetycznej ze zmiennymi diagnostycznymi, miernik zmienności i koncentracji zmiennej syntetycznej[5].

Mierniki te mają charakter cząstkowy o jednoznacznym kierunku preferencji – mniejsze wartości liczbowe każdego miernika wskazują na lepszą procedurę porządkowania liniowego [Grabiński, Wydymus, Zeliaś 1989, s. 125]. W związku z tym można przeprowadzić agregację mierników cząstkowych na podstawie wzoru [Seidler i in. 1980]:

$$Q_k = \sqrt{\sum_{l=1}^{g} g_l{}^2}, \tag{12}$$

gdzie: Q_k – miernik agregatowy k-tej konfiguracji elementów procedury porządkowania liniowego, g_l – miernik cząstkowy ($l = 1, \ldots, 7$), g – liczba mierników cząstkowych.

4. Wyniki badań

W badaniach wykorzystano zbiory danych empirycznych (z prac [Hellwig 1968] i [Nowak 1984]) oraz dane symulacyjne.

W przypadku danych z 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:

```
library(pllord)
> 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
```

Wyniki uporządkowane według rosnących wartości miernika agregatowego Q_k:

```
library(pllord)
> print(loqo)
  weigth                 normalize                 agrregate aggrgauge
1    1 standardization ((x-mean)/sd)          arithmetic mean  1.608215
2    1 standardization ((x-mean)/sd)            harmonic mean  1.608215
```

[5] Formuły analityczne tych mierników są zamieszczone w pracach: [Grabiński 1984; Grabiński, Wydymus i Zeliaś 1989; Bąk 1999].

```
4    1 standardization ((x-mean)/sd)       Euclidean distance  4.674878
5    1 standardization ((x-mean)/sd) Jeffreys-Matusita distance 5.085701
3    1 standardization ((x-mean)/sd)       Hamming distance  5.916667
8    1 standardization ((x-mean)/sd)      Canberra distance  5.996884
9    1 standardization ((x-mean)/sd)       angular distance  6.915035
7    1 standardization ((x-mean)/sd)        Clark distance  7.228170
6    1 standardization ((x-mean)/sd)   Bray-Curtis distance  7.745446
```

W przypadku danych z pracy [Nowak 1984] porządkowanymi obiektami jest 15 krajów charakteryzowanych przez 6 zmiennych: X1 – plony pszenicy w q z 1 ha, X2 – plony ziemniaków w q z 1 ha, X3 – plony buraków cukrowych w q z 1 ha, X4 – produkcja mięsa wołowego w kg na 1 ha użytków rolnych, X5 – produkcja mięsa wieprzowego w kg na 1 ha użytków rolnych, X6 – produkcja mleka w litrach na 1 ha użytków rolnych.

Fragment zbioru danych:

```
library(pllord)
> head(ndane84)
            Kraj   X1  X2  X3  X4  X5   X6
1         Belgia 44.7 323 521 188 509 2575
2       Bulgaria 39.7  89 274  21  50  302
3 Czechoslowacja 45.1 136 331  54 127  857
4          Dania 50.6 266 397  85 357 1802
5        Francja 50.6 266 483  57  69 1066
6      Hiszpania 22.2 175 326  12  29  191
```

Wyniki uporządkowane według rosnących wartości miernika agregatowego Q_k:

```
library(pllord)
> print(loqo)
  weigth                    normalize                agrregate aggrgauge
5      1 standardization ((x-mean)/sd) Jeffreys-Matusita distance  5.164260
4      1 standardization ((x-mean)/sd)       Euclidean distance  5.215094
8      1 standardization ((x-mean)/sd)        Canberra distance  5.690777
1      1 standardization ((x-mean)/sd)          arithmetic mean  6.287777
2      1 standardization ((x-mean)/sd)            harmonic mean  6.287777
3      1 standardization ((x-mean)/sd)         Hamming distance  6.423607
9      1 standardization ((x-mean)/sd)         angular distance  6.957684
7      1 standardization ((x-mean)/sd)           Clark distance  7.104344
6      1 standardization ((x-mean)/sd)     Bray-Curtis distance  7.949305
```

Dane symulacyjne z wielowymiarowego rozkładu normalnego (15 obiektów i 2 zmienne) zostały wygenerowane z wykorzystaniem funkcji `mvrnorm()` z pakietu `MASS` za pomocą skryptu:

```
library(pllord)
library(MASS)
gmvrnorm<-
function(n=10,mu=c(0,0),sigma=matrix(c(1,0.5,0.5,1),2,2),ss=TRUE)
{      if(ss) {set.seed(1234567)}
       GD<-mvrnorm(n,mu,sigma)
       return(GD) }
```

Wyniki uporządkowane według rosnących wartości miernika agregatowego Q_k:

```
> print(loqo)
  weigth                    normalize                 agrregate aggrgauge
1      1 standardization ((x-mean)/sd)          arithmetic mean  21.39433
2      1 standardization ((x-mean)/sd)            harmonic mean  21.39433
4      1 standardization ((x-mean)/sd)       Euclidean distance  26.63206
5      1 standardization ((x-mean)/sd) Jeffreys-Matusita distance  31.76824
3      1 standardization ((x-mean)/sd)         Hamming distance  65.82223
8      1 standardization ((x-mean)/sd)        Canberra distance  73.80747
9      1 standardization ((x-mean)/sd)          angular distance  78.43209
6      1 standardization ((x-mean)/sd)     Bray-Curtis distance  84.94770
7      1 standardization ((x-mean)/sd)           Clark distance  87.06585
```

5. Podsumowanie

W aktualnej wersji pakietu `pllord` uwzględniono wybrane konfiguracje procedur porządkowania liniowego i wybrane mierniki oceny jakości otrzymanych uporządkowań. Z tych wstępnych badań wynika, że nie ma jednoznacznych wskazań, które procedury porządkowania liniowego są najlepsze zarówno w przypadku danych empirycznych, jak i danych stymulacyjnych. Problematyka ta powinna być przedmiotem dalszych badań.

Główne kierunki badań i rozwoju pakietu `pllord` to: włączenie pominiętych dotychczas metod porządkowania liniowego, stworzenie możliwości wyboru metody ważenia i normalizacji zmiennych oraz miary odległości w metodach wzorcowych, opracowanie funkcji umożliwiających ocenę jakości aplikacyjnej metod porządkowania liniowego w określonej konfiguracji na podstawie innych mierników, analiza własności mierników jakości metod porządkowania liniowego na podstawie danych symulacyjnych o różnych rozkładach statystycznych, normalizacja miernika agregatowego Q_k w określonym przedziale zmienności.

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++*, Wrocław, Wydawnictwo Akademii Ekonomicznej we Wrocławiu.

Bąk A., 2013, *Metody porządkowania liniowego w polskiej taksonomii – pakiet* `pllord`, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 278, s. 54-62.

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.

Grabiński T., 1984, *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie. Seria specjalna: Monografie nr 61.

Grabiński T., 1992, *Metody taksonometrii*, Kraków, Akademia Ekonomiczna w Krakowie.

Grabiński T., Wydymus S., Zeliaś A., 1989, *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, Warszawa, PWN.

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.

Nowak E., 1984, *Problemy doboru zmiennych do modelu ekonometrycznego*, Warszawa, PWN.

Pluta W., 1976, *Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego*, „Przegląd Statystyczny", z. 4, s. 511-517.

R Development Core Team [2014], *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, URL: http://cran.r-project.org.

Seidler J., Badach A., Molisz W., 1980, *Metody rozwiązywania zadań optymalizacji*, Warszawa, WNT.

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*, Wrocław, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Walesiak M., Dudek A., 2012, `clusterSim` *package*, URL: http://www.R-project.org.

PROBLEM OF CHOOSING THE OPTIMAL LINEAR ORDERING PROCEDURE IN THE `PLLORD` PACKAGE

Summary: Linear ordering methods are used in studies designed to determine the order or classification of objects. The subject of linear ordering can be such objects as countries (due to the level of economic development), business (due to financial condition), products (due to usability), etc. On the basis of taxonomic research a number of procedures for linear ordering have been developed. They mainly differ in methods of determining the weighting of variables, methods of standardization of variables and methods for estimating the values of the synthetic variable. Therefore, there is a problem of choosing the optimal procedure for the analysis of the empirical data with specified statistical characteristics. The aim of the article is to outline the issues of choosing the optimum configuration of the components of the linear ordering procedure using selected measures of quality evaluation and R program.

Keywords: linear ordering, optimal procedure, R program.