

---

**Andrzej Bąk**

Uniwersytet Ekonomiczny we Wrocławiu

e-mail: andrzej.bak@ue.wroc.pl

---

## **ANALIZA PORÓWNAWCZA WYBRANYCH METOD PORZĄDKOWANIA LINIOWEGO**

---

### **COMPARATIVE ANALYSIS OF SELECTED LINEAR ORDERING METHODS BASED ON EMPIRICAL AND SIMULATION DATA**

---

DOI: 10.15611/pn.2018.508.02

JEL Classification: C00, C19, C49, C87

**Streszczenie:** Metody porządkowania liniowego są stosowane na gruncie ekonomii w badaniach rankingowych i klasyfikacyjnych dotyczących obiektów i zjawisk opisanych zmiennymi o różnych rozkładach i charakterystykach statystycznych. W literaturze przedmiotu z zakresu wielowymiarowej analizy porównawczej 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 porządkowania liniowego do analizy danych statystycznych. Celem pracy jest prezentacja wyników badań dotyczących oceny jakości wybranych procedur porządkowania liniowego w odniesieniu do danych o rozkładzie normalnym i różnym od rozkładu normalnego. Do oceny jakości rankingów wykorzystane zostały wybrane mierniki jakości metod porządkowania liniowego.

**Słowa kluczowe:** porządkowanie liniowe, analiza porównawcza, program R.

**Summary:** Linear ordering methods are used in economic studies to determine the order or classification of objects described using variables with different distributions and statistical characteristics. In the literature on the subject in the field of multidimensional comparative analysis, many linear ordering procedures have been developed. They differ, among others using the methods for determining variable weights, normalization methods, and methods for estimating the values of synthetic variables. Therefore, there is a problem of choosing the optimal procedure for the analysis of statistical data with specified statistical characteristics. The purpose of this article is to present the results of research concerning the quality evaluation of selected linear ordering procedures in relation to data about normal distribution and different from normal distribution. To assess the quality of rankings, selected measures of the quality of linear ordering methods were used.

**Keywords:** linear ordering, comparative analysis, R program.

## 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.

Celem artykułu jest prezentacja wyników badań dotyczących oceny jakości wybranych procedur porządkowania liniowego (bezwzorcowych i wzorcowych) w odniesieniu do danych (empirycznych i symulacyjnych) o rozkładzie normalnym i różnym od rozkładu normalnego. Do oceny jakości rankingów wykorzystane zostały wybrane mierniki jakości metod porządkowania liniowego. Dodatkowym efektem pracy jest implementacja analizowanych metod porządkowania liniowego, mierników oceny jakości tych metod oraz metod generowania danych symulacyjnych w pakiecie `pllord` [Bąk 2013; 2017] programu R [R Development Core Team 2017].

## 2. Wybrane metody porządkowania liniowego

Podstawą porządkowania liniowego jest zmienna syntetyczna<sup>1</sup>, której wartości są szacowane na podstawie obserwacji zmiennych diagnostycznych opisujących badane obiekty. Zakłada się, że wartości zmiennej syntetycznej, oszacowane za pomocą określonej metody, umożliwiają takie uporządkowanie zbioru obiektów, w którym [Grabiński 1992, s. 135]: (1) każdy obiekt ma przynajmniej jednego sąsiada oraz nie więcej niż dwóch sąsiadów, (2) jeżeli obiekt  $a$  jest sąsiadem obiektu  $b$ , to obiekt  $b$  jest sąsiadem obiektu  $a$ , (3) 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 (addytywne lub multiplikatywne) i metody wzorcowe (wykorzystujące miary odległości od wzorca lub antywzorca).

---

<sup>1</sup> W literaturze przedmiotu spotkać można inne określenia zmiennej syntetycznej, takie jak np.: zmienna agregatowa, miara syntetyczna, syntetyczna miara rozwoju, taksonomiczny miernik rozwoju, agregatowa miara rozwoju, miara rozwoju gospodarczego.

W procedurze porządkowania liniowego wyróżnia się takie etapy postępowania, jak: określenie charakteru zmiennych (stymulanty, nominanty, destymulanty)<sup>2</sup>, wyznaczenie wag zmiennych, normalizacja zmiennych, wyznaczenie współrzędnych wzorca w przypadku agregacji wzorcowej, agregacja bezwzorcowa lub wzorcowa [Grabiński 1984; Pocięcha i in. 1988; Bąk 1999].

Pierwsza propozycja metody porządkowania liniowego w obszarze badań taksonomicznych i ekonomicznych, umożliwiającej porządkowanie liniowe obiektów z wykorzystaniem wzorca, została przedstawiona przez Z. Hellwiga w 1968 r. pod nazwą „miara rozwoju gospodarczego” [Hellwig 1968].

Na gruncie teorii decyzji (wielokryterialnego podejmowania decyzji) pierwsza metoda porządkowania liniowego z wykorzystaniem wzorca i antywzorca została zaproponowana przez C.L. Hwanga i K. Yoon w 1981 r. pod nazwą TOPSIS – *Technique for Order Preference by Similarity to Ideal Solution* [Hwang, Yoon 1981].

Idea zastosowania analizy głównych składowych (PCA – *Principal Component Analysis*) do porządkowania liniowego obiektów (na podstawie wartości pierwszej głównej składowej) została zaproponowana przez J. Perkalą (w latach 1963-1967) na gruncie badań przyrodniczych i rolniczych [Perkal 1967]. Podstawy analizy głównych składowych zostały przedstawione w pracach K. Pearsona [1901] i H. Hotellinga [1933].

W porządkowaniu liniowym stosuje się także metody addytywne i multiplikatywne oparte na średniej arytmetycznej, harmonicznnej lub geometrycznej.

Konstrukcja miary syntetycznej (miary rozwoju gospodarczego) Hellwiga jest następująca [Hellwig 1968]:

a) normalizacja zmiennych (standaryzacja):  $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ ,  $x_{ij}$  – 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:

$$z_{0j} = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych o charakterze stymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych o charakterze destymulant} \end{cases}$$

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

d) wartości zmiennej agregatywnej:  $q_i = 1 - \frac{d_{i0}}{d_0}$ , przy czym: na ogół  $q_i \in [0; 1]$ ;  $\max_i \{q_i\}$  – najlepszy obiekt;  $\min_i \{q_i\}$  – najgorszy obiekt;  $d_0 = \bar{d}_0 + 2s_d$ ;  
 $\bar{d}_0 = \frac{1}{n} \sum_{i=1}^n d_{i0}$ ;  $s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_{i0} - \bar{d}_0)^2}$ .

<sup>2</sup> Pojęcia zmiennej stymulandy i destymulandy zostały wprowadzone do literatury przedmiotu przez Z. Hellwiga [1968], a pojęcie zmiennej nominandy przez T. Borysa [1978].

Konstrukcja miary syntetycznej (TOPSIS) Hwanga i Yoona jest następująca [Hwang, Yoon 1981; Yoon, Hwang 1995]:

- a) normalizacja zmiennych (przekształcenie ilorazowe):  $z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}$ ,  $x_{ij}$  –

obserwacja  $j$ -tej zmiennej dla obiektu  $i$ ;

- 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) współrzędne antywzorca:

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

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

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

- f) wartości zmiennej agregatowej:  $q_i = \frac{d_{i0}^-}{d_{i0}^+ + d_{i0}^-}$ , przy czym:  $q_i \in [0; 1]$ ;

$\max_i \{q_i\}$  – najlepszy obiekt;  $\min_i \{q_i\}$  – najgorszy obiekt.

Konstrukcja miary syntetycznej (PCA) na podstawie wartości pierwszej głównej składowej jest następująca [Perkal 1967; Rusnak 1999; Balicki 2009]:

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

- b) obliczenie macierzy kowariancji **S** lub korelacji **R**;

- c) obliczenie wartości własnych i wektorów własnych macierzy kowariancji **S** lub korelacji **R**;

- d) zmiana znaków współczynników pierwszej głównej składowej (elementów pierwszego wektora własnego), jeżeli więcej jest współczynników ujemnych niż dodatnich;

- e) obliczenie wartości składowych głównych na podstawie danych empirycznych i wektorów własnych:  $\mathbf{Y} = \mathbf{XW}$ : **Y** – macierz głównych składowych, **X** – macierz danych empirycznych, **W** – macierz współczynników głównych składowych (wektory własne);

- f) uporządkowanie obiektów na podstawie wartości pierwszej składowej głównej (malejąco).

W konstrukcji addytywnej miary syntetycznej wykorzystuje się średnią arytmetyczną obliczoną dla znormalizowanych wartości zmiennych diagnostycznych (AV):

a) normalizacja zmiennych (standaryzacja):  $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ ,  $x_{ij}$  – 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) obliczenie wartości zmiennej syntetycznej:  $q_i = \frac{1}{m} \sum_{j=1}^m w_j \cdot z_{ij}$ ,  $w_j$  – waga  $j$ -tej zmiennej diagnostycznej.

Jeżeli metodą normalizacji zmiennych diagnostycznych jest standaryzacja, to wartości miary syntetycznej obliczane są metodą sum standaryzowanych.

### 3. Mierniki oceny jakości procedur porządkowania liniowego

Poszczególne konfiguracje utworzone z wag zmiennych, metod normalizacji i metod 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 i in. 1999; 2015].

Mierniki jakości stosowane do wyboru optymalnej konfiguracji dotyczą w szczególności oceny takich własności procedur porządkowania liniowego, jak: (1) zgodność odwzorowania, mierzona wskaźnikiem zróżnicowania odległości między obiektami w przestrzeni zmiennych diagnostycznych oraz w przestrzeni zmiennej syntetycznej, (2) korelacja liniowa pomiędzy zmienną syntetyczną a zmiennymi diagnostycznymi, mierzona przeciętnym współczynnikiem „nieokreśloności” oraz współczynnikiem „jednoznaczności” zmiennej syntetycznej, (3) korelacja rangowa zmiennej syntetycznej ze zmiennymi diagnostycznymi, mierzona współczynnikiem „nieokreśloności”, współczynnikiem „jednoznaczności” zmiennej syntetycznej oraz uogólnionym rangowym współczynnikiem rozbieżności, (4) zmienność i koncentracja zmiennej syntetycznej, mierzone współczynnikiem obliczonym dla realizacji zmiennej syntetycznej oraz dla pierwszych różnic uporządkowanych niemalejąco wartości zmiennej syntetycznej, (5) 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<sup>3</sup>.

<sup>3</sup> Formuły analityczne tych mierników są zamieszczone w pracach: [Grabiński 1984; Grabiński i in. 1989; Bąk 1999].

Mierniki te mają charakter cząstkowy o jednoznacznym kierunku preferencji – mniejsze wartości liczbowe każdego miernika wskazują na lepszą procedurę porządkowania liniowego. W związku z tym można przeprowadzić agregację mierników cząstkowych na podstawie wzoru [Seidler i in. 1980]:  $Q_k = \sqrt{\sum_{l=1}^g g_l^2}$ , gdzie:  $Q_k$  – miernik agregatowy  $k$ -tej konfiguracji elementów procedury porządkowania liniowego,  $g_l$  – miernik cząstkowy ( $l = 1, \dots, g$ ),  $g$  – liczba mierników cząstkowych.

## 4. Wyniki badań

W badaniach porównawczych czterech metod porządkowania liniowego (Hellwiga, TOPSIS, PCA i AV) wykorzystano:

- dane empiryczne: dane statystyczne dotyczące stanu i ochrony środowiska w powiatach województwa dolnośląskiego w 2015 r.<sup>4</sup> – DE1 (10 zmiennych, 30 obiektów) i dane statystyczne dotyczące rankingu polskich uczelni akademickich w 2016 r.<sup>5</sup> – DE2 (27 zmiennych, 90 obiektów);
- dane symulacyjne otrzymane za pomocą generatora wielowymiarowego rozkładu normalnego dla zadanych wektorów średnich i dodatnio określonych macierzy kowariancji – DS1 (10 zmiennych, 30 obiektów) i DS2 (27 zmiennych, 90 obiektów).

W badaniach porównawczych zastosowano następującą procedurę:

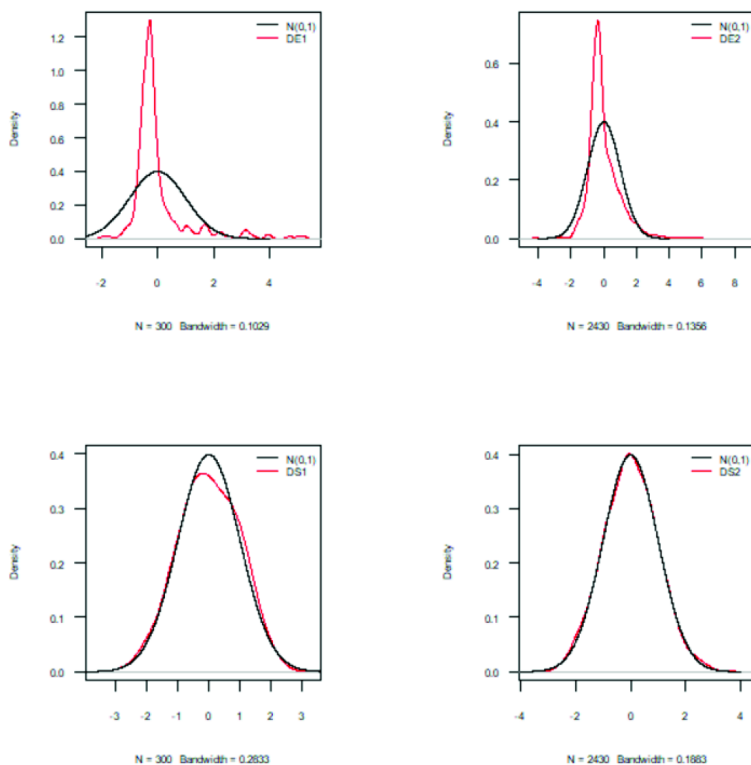
- przyjęto, że wszystkie zmienne mają charakter stymulant,
- przyjęto jednostkowe wagi dla wszystkich zmiennych,
- przeprowadzono testy normalności rozkładu dla danych empirycznych i symulacyjnych – test Shapiro-Wilka,
- przeprowadzono normalizację danych z wykorzystaniem standaryzacji (N1) stosowaną w metodzie Hellwiga oraz za pomocą przekształcenia ilorazowego (N2) stosowanego w metodzie TOPSIS,
- przeprowadzono porządkowanie liniowe metodami: Hellwiga, TOPSIS, PCA i AV) zbiorów danych DE1, DE2, DS1, DS2 przy zastosowaniu normalizacji N1 i N2 (większe wartości zmiennej syntetycznej wskazują wyższą pozycję obiektu w rankingu),
- przeprowadzono ocenę jakości rankingów za pomocą cząstkowych kryteriów oceny jakości i obliczono mierniki agregatowe  $Q$  (mniejsze wartości miernika agregatowego wskazują na lepszą procedurę porządkowania liniowego).

Na podstawie wyników testu Shapiro-Wilka należy odrzucić hipotezę o normalności rozkładów danych empirycznych DE1 i DE2. Dane symulacyjne DS1 i DS2 mają natomiast rozkład normalny. Graficzną ilustrację zgodności rozkładów danych

<sup>4</sup> Źródło: <http://wroclaw.stat.gov.pl>.

<sup>5</sup> Źródło: <http://www.ranking.perspektywy.org>.

empirycznych i symulacyjnych z rozkładem normalnym przedstawiono na rys. 1 za pomocą funkcji gęstości poszczególnych rozkładów.



**Rys. 1.** Funkcje gęstości rozkładu normalnego oraz rozkładów danych empirycznych i symulacyjnych  
Źródło: opracowanie własne.

W tabeli 1 zestawiono wartości agregatowego miernika jakości metod porządkowania liniowego (Hellwiga, TOPSIS, PCA i AV) obliczone dla danych empirycznych (DE1 i DE2), natomiast w tab. 2 dla danych symulacyjnych (DS1 i DS2) przy zastosowaniu metod normalizacji N1 i N2.

Na podstawie otrzymanych wyników analizy danych można sformułować następujące wnioski:

- w przypadku danych empirycznych o rozkładzie różnym od normalnego: najniższe wartości miernika agregatowego otrzymano dla metody PCA,
- w przypadku danych symulacyjnych o rozkładzie normalnym: najniższe wartości miernika agregatowego otrzymano dla metody AV,
- w przypadku danych empirycznych spośród dwóch metod wzorcowych wykorzystujących odległość euklidesową: niższe wartości miernika agregatowego otrzymano dla metody TOPSIS,

**Tabela 1.** Wartości agregatowego miernika jakości metod porządkowania liniowego Q dla danych empirycznych

| Normalizacja          | Metoda Hellwiga | Metoda TOPSIS | Metoda PCA | Metoda AV |
|-----------------------|-----------------|---------------|------------|-----------|
| DE1, $m = 10, n = 30$ |                 |               |            |           |
| N1                    | 9,416211        | 9,170164      | 4,753804   | 5,305854  |
| N2                    | 9,466951        | 9,249728      | 8,799317   | 6,803840  |
| DE2, $m = 27, n = 90$ |                 |               |            |           |
| N1                    | 11,254887       | 11,242065     | 4,517565   | 27,154239 |
| N2                    | 11,309287       | 11,253389     | 10,88955   | 8,851145  |

Źródło: opracowanie własne.

**Tabela 2.** Wartości agregatowego miernika jakości metod porządkowania liniowego Q dla danych symulacyjnych

| Normalizacja          | Metoda Hellwiga | Metoda TOPSIS | Metoda PCA | Metoda AV |
|-----------------------|-----------------|---------------|------------|-----------|
| DS1, $m = 10, n = 30$ |                 |               |            |           |
| N1                    | 9,356263        | 9,590986      | 3,748789   | 3,249898  |
| N2                    | 9,364065        | 9,594696      | 8,355604   | 7,691365  |
| DS2, $m = 27, n = 90$ |                 |               |            |           |
| N1                    | 11,395239       | 11,473243     | 6,583351   | 5,626531  |
| N2                    | 11,366475       | 11,461965     | 11,077245  | 9,812555  |

Źródło: opracowanie własne.

- w przypadku danych symulacyjnych spośród dwóch metod wzorcowych wykrzystujących odległość euklidesową: niższe wartości miernika agregatowego otrzymano dla metody Hellwiga,
- metody bezwzorcowe porządkowania liniowego (PCA i AV) przyjmują niższe wartości miernika agregatowego, jeżeli normalizacja danych zarówno empirycznych, jak i symulacyjnych jest przeprowadzona metodą N1 (z wyjątkiem metody AV dla danych empirycznych DE2),
- metody wzorcowe porządkowania liniowego (Hellwiga i TOPSIS) przyjmują niższe wartości miernika agregatowego, jeżeli normalizacja danych empirycznych (DE1 i DE2) jest przeprowadzona metodą N1,
- w przypadku danych symulacyjnych (DS1 i DS2) metoda TOPSIS przyjmuje niższe wartości miernika agregatowego, jeżeli normalizacja danych jest przeprowadzona metodą N1,
- w przypadku danych symulacyjnych (DS1) metoda Hellwiga przyjmuje niższe wartości miernika agregatowego, jeżeli normalizacja danych jest przeprowadzona metodą N1, natomiast w przypadku danych symulacyjnych (DS), jeżeli normalizacja danych jest przeprowadzona metodą N2.



## 5. Podsumowanie

Metody porządkowania liniowego znajdują praktyczne zastosowania w wielu dziedzinach badań, w tym w ekonomii. Celem badań empirycznych z wykorzystaniem tych metod jest ustalenie kolejności obiektów opisywanych przez zbiór zmiennych. Różne procedury porządkowania liniowego prowadzą do różnych rankingów porównywanych obiektów. W pracy przedstawiono wyniki analizy porównawczej czterech metod porządkowania liniowego (Hellwiga, TOPSIS, PCA i AV) i dwóch metod normalizacji zmiennych (standaryzacji i przekształcenia ilorazowego). Analizę przeprowadzono na podstawie danych empirycznych i symulacyjnych o rozkładzie normalnym i różnym od normalnego. Wyniki analizy wskazują na zależność jakości rankingów od rozkładu danych statystycznych, metody normalizacji zmiennych i metody agregacji.

## Literatura

- Balicki A., 2009, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Sopot.
- Bąk A., 1999, *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Bąk A., 2013, *Metody porządkowania liniowego w polskiej taksonomii – pakiet pllord*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 278, s. 54-62.
- Bąk A., 2015, *Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie pllord*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 384, s. 33-41.
- Borys T., 1978, *Metody normowania cech w statystycznych badaniach porównawczych*, Przegląd Statystyczny, nr 2, s. 227-239
- 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*, Akademia Ekonomiczna w Krakowie, Kraków.
- Grabiński 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.
- Hotelling H., 1933, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology, 24(7), s. 417-441, 498-520.  
<http://wroclaw.stat.gov.pl>  
<http://www.ranking.perspektywy.org>
- Hwang C.L., Yoon K., 1981, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York.
- Pearson K., 1901, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine, nr 6(2), s. 559-572.
- Perkal J., 1967, *Matematyka dla przyrodników i rolników*. Część II, PWN, Warszawa.
- Pociecha J., Podolec B., Sokołowski A., Zajac K., 1988, *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa.

- R Development Core Team, 2017, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, URL: <http://cran.r-project.org>.
- Rusnak J., 1999, *Metoda głównych składowych*, [w:] Ostasiewicz W. (red.), *Statystyczne metody analizy danych*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Seidler J., Badach A., Molisz W., 1980, *Metody rozwiązywania zadań optymalizacji*, WNT, Warszawa.
- Yoon K., Hwang C.L., 1995, *Multiple Attribute Decision Making: An Introduction*, Sage Publications, Thousand Oaks, California.