

MAREK WALESIAK¹

WIZUALIZACJA WYNIKÓW PORZĄDKOWANIA LINIOWEGO DLA DANYCH PORZĄDKOWYCH Z WYKORZYSTANIEM SKALOWANIA WIELOWYMIAROWEGO

1. WPROWADZENIE

W artykule przedstawiono propozycję zastosowania skalowania wielowymiarowego (Borg, Groenen, 2005; Borg i inni, 2013) w porządkowaniu liniowym zbioru obiektów, opisanych zmiennymi porządkowymi, bazującym na wzorcu rozwoju (Hellwig, 1968, 1972). Propozycję rozwiązania dla danych metrycznych zaprezentowano w pracy (Walesiak, 2016b).

Zaproponowano dwukrokową procedurę badawczą pozwalającą na wizualizację wyników porządkowania liniowego. Najpierw w wyniku zastosowania skalowania wielowymiarowego otrzymuje się wizualizację rozmieszczenia obiektów w przestrzeni dwuwymiarowej. Następnie przeprowadza się porządkowanie liniowe obiektów na podstawie odległości Euklidesa od wzorca rozwoju. Zaproponowane podejście zilustrowano przykładem empirycznym.

W artykule wykorzystano koncepcję izokwant i ścieżki rozwoju (osi zbioru – najkrótszej drogi łączącej wzorzec i antywzorzec rozwoju²) zaproponowaną w pracy (Hellwig, 1981). Graficzna prezentacja wyników porządkowania liniowego w tej koncepcji możliwa była dla dwóch zmiennych. Zastosowanie skalowania wielowymiarowego rozszerzyło możliwości zastosowania wizualizacji wyników porządkowania liniowego dla m zmiennych.

2. DANE PORZĄDKOWE

W teorii pomiaru rozróżnia się cztery podstawowe skale pomiaru wprowadzone przez Stevensa (1946): nominalną, porządkową, interwałową (przedziałową), ilorazową (stosunkową). Skale przedziałową i ilorazową zalicza się do skal metrycznych, natomiast nominalną i porządkową do niemetrycznych. Skale pomiaru są uporządko-

¹ Uniwersytet Ekonomiczny we Wrocławiu, Wydział Ekonomii, Zarządzania i Turystyki, Katedra Ekonometrii i Informatyki, ul. Nowowiejska 3, 58-500 Jelenia Góra, Polska, e-mail: marek.walesiak@ue.wroc.pl.

² Wzorzec (górną biegun) obejmuje najkorzystniejsze wartości (kategorie) zmiennych, antywzorzec (dolny biegun) zaś najmniej korzystne wartości (kategorie) zmiennych preferencyjnych.

wane od najslabszej (nominalna) do najmocniejszej (ilorazowa). Z typem skali wiąże się grupa przekształceń, ze względu na które skala zachowuje swe własności. Na skali porządkowej dozwolonym przekształceniem matematycznym dla obserwacji jest dowolna ściśle monotonicznie rosnąca funkcja, która nie zmienia dopuszczalnych relacji, tj. równości, różności, większości i mniejszości.

Zasób informacji skali porządkowej jest nieporównanie mniejszy niż skal metrycznych. Jedyną dopuszczalną operacją empiryczną na skali porządkowej jest zliczanie zdarzeń (tzn. wyznaczanie liczby relacji większości, mniejszości i równości). Szczegółową charakterystykę skal pomiaru zawierają m.in. prace (Walesiak 1996, s. 19–24; 2016a, s. 15–17).

3. GENEZA KONCEPCJI WZORCA ROZWOJU I MIARY ROZWOJU

Pierwszy referat poświęcony koncepcji wzorca rozwoju i miary rozwoju w języku angielskim został zaprezentowany przez Profesora Zdzisława Hellwiga na konferencji UNESCO w Warszawie w 1967 roku (Hellwig, 1967). Praca ukazała się drukiem w języku angielskim w monografii pod redakcją Z. Gostkowskiego (Hellwig, 1972).

Pierwszy artykuł poświęcony koncepcji wzorca rozwoju i miary rozwoju w języku polskim ukazał się w czasopiśmie „Przegląd Statystyczny” w 1968 roku (Hellwig, 1968).

W pracach tych wprowadzono pojęcia:

- stymulant i destymulant (ang. *stimulants and dis-stimulants*),
- wzorca rozwoju (ang. *pattern of development*),
- miary rozwoju (ang. *measure of development*) jako odległości od wzorca rozwoju (ang. *distance from the pattern of development*).

Idea Hellwiga zapoczątkowała, można bez przesady powiedzieć, lawinę propozycji tworzenia metod porządkowania liniowego. Modyfikacje te zmierzały do (zob. Borys i inni, 1990; Pocięcha, Zajac, 1990):

- a) różnicowania sposobu normalizacji wartości zmiennych,
- b) wprowadzenia do zbioru zmiennych nominant,
- c) odmiennego ustalania wzorca rozwoju (bazy porównawczej),
- d) wykorzystania różnych konstrukcji miary agregatowej (tzw. miary rozwoju),
- e) wykorzystania zbiorów rozmytych w konstrukcji miary agregatowej.

W ostatnim okresie w porządkowaniu liniowym bazującym na wzorcu rozwoju powstały koncepcje wykorzystujące liczby rozmyte (zob. np. Wysocki, 2010; Jefmański, Dudek, 2016), dane symboliczne interwałowe (Młodak, 2014) oraz uwzględniające zależności przestrzenne (Antczak, 2013; Pietrzak, 2014).

4. TYPOWA PROCEDURA POSTĘPOWANIA W PORZĄDKOWANIU LINIOWYM BAZUJĄCYM NA WZORCU DLA DANYCH PORZĄDKOWYCH

Typowa procedura postępowania w porządkowaniu liniowym bazującym na wzorcu (górnym biegun) lub antywzorcu (dolnym biegun) dla danych porządkowych obejmuje następujące kroki:

$$P \rightarrow A \rightarrow X \rightarrow SDN \rightarrow T_w \rightarrow d_i \rightarrow R, \quad (1)$$

gdzie:

P – wybór zjawiska złożonego (nadrzędne kryterium porządkowania elementów zbioru A , które nie podlega pomiarowi bezpośredniemu),

A – wybór obiektów,

X – dobór zmiennych porządkowych. Zgromadzenie danych i konstrukcja macierzy danych w przestrzeni m -wymiarowej $[x_{ij}]_{n \times m}$ (x_{ij} – kategoria j -tej zmiennej porządkowej dla i -tego obiektu, $i = 1, \dots, n$ – numer obiektu, $j = 1, \dots, m$ – numer zmiennej),

SDN – identyfikacja zmiennych preferencyjnych (stymulanty, destymulanty, nominanty). Zmienna M_j jest stymulantą (Hellwig, 1981, s. 48), gdy dla każdych dwóch jej obserwacji x_{ij}^S, x_{kj}^S odnoszących się do obiektów A_i, A_k jest $x_{ij}^S > x_{kj}^S \Rightarrow A_i \succ A_k$ (\succ oznacza dominację obiektu A_i nad obiektem A_k). Zmienna M_j jest destymulantą (Hellwig, 1981, s. 48), gdy dla każdych dwóch jej obserwacji x_{ij}^D, x_{kj}^D odnoszących się do obiektów A_i, A_k jest $x_{ij}^D > x_{kj}^D \Rightarrow A_i \prec A_k$ (\prec oznacza dominację obiektu A_k nad obiektem A_i). Zmienna M_j jest nominantą jednodalną (Borys, 1984, s. 118), gdy dla każdych dwóch jej obserwacji x_{ij}^N, x_{kj}^N odnoszących się do obiektów A_i, A_k (nom_i oznacza nominalny poziom j -tej zmiennej): jeżeli $x_{ij}^N, x_{kj}^N \leq nom_j$, to $x_{ij}^N > x_{kj}^N \Rightarrow A_i \succ A_k$; jeżeli $x_{ij}^N, x_{kj}^N > nom_j$, to $x_{ij}^N > x_{kj}^N \Rightarrow A_i \prec A_k$.

T_w – transformacja nominant w stymulanty lub destymulanty (w przypadku zastosowania miar odległości od antywzorca rozwoju). Metody transformacji nominant dla zmiennych mierzonych na skali porządkowej w destymulanty przedstawiono w pracy Walesiak (2011),

d_i – obliczenie dla i -tego obiektu wartości miary agregatywnej – zastosowanie miar odległości od wzorca lub antywzorca z udziałem wag,

R – uporządkowanie obiektów według wartości d_i .

W tabeli 1 przedstawiono miary odległości od wzorca rozwoju dla danych porządkowych.

Tabela 1.

Miary odległości od wzorca rozwoju dla danych porządkowych

Nazwa	Odległość d_i	Przedział zmienności
Odległość GDM2 (Walesiak, 1993, 1999)	$1 - GDM2_i^+ = \frac{1}{2} + \frac{\sum_{j=1}^m \alpha_j a_{iwj} b_{wj} + \sum_{j=1}^m \sum_{l=1}^n \alpha_j a_{ilj} b_{wlj}}{2 \left[\sum_{j=1}^m \sum_{l=1}^n \alpha_j a_{ilj}^2 \cdot \sum_{j=1}^m \sum_{l=1}^n \alpha_j b_{wlj}^2 \right]^{\frac{1}{2}}},$ $a_{ijp}(b_{wj}) = \begin{cases} 1 & \text{jeżeli } x_{ij} > x_{pj} \text{ (} x_{wj} > x_{rj} \text{)} \\ 0 & \text{jeżeli } x_{ij} = x_{pj} \text{ (} x_{wj} = x_{rj} \text{)} \\ -1 & \text{jeżeli } x_{ij} < x_{pj} \text{ (} x_{wj} < x_{rj} \text{)} \end{cases}$ <p>dla $p = w, l$; $r = i, l$</p>	[0;1]
GDM2_TOPSIS (miara TOPSIS (Hwang, Yoon, 1981) z odległością GDM2)	$\frac{GDM2_i^-}{GDM2_i^- + GDM2_i^+}$	[0;1]

$GDM2_i^+$ ($GDM2_i^-$) – odległość GDM2 obiektu i -tego od wzorca (antywzorca), $x_{wj} = x_{+j}$ ($x_{wj} = x_{-j}$) dla $GDM2_i^+$ ($GDM2_i^-$); x_{+j} (x_{-j}) – j -ta współrzędna obiektu wzorca (antywzorca), $i, l = 1, \dots, n$ – numer obiektu, $j = 1, \dots, m$ – numer zmiennej, α_j – waga j -tej zmiennej ($\alpha_j \in [0; 1]$ i $\sum_{j=1}^m \alpha_j = 1$ lub $\alpha_j \in [0; m]$ i $\sum_{j=1}^m \alpha_j = m$).

Źródło: opracowanie własne.

5. PROCEDURA BADAWCZA POZWALAJĄCA NA WIZUALIZACJĘ WYNIKÓW PORZĄDKOWANIA LINIOWEGO ZBIORU OBIEKTÓW

Procedura badawcza pozwalająca na wizualizację wyników porządkowania liniowego zbioru obiektów opisanych zmiennymi porządkowymi obejmuje następujące kroki:

1. Wybór zjawiska złożonego w porządkowaniu liniowym, które nie podlega pomiarowi bezpośredniemu (np. poziom atrakcyjności nieruchomości lokalowych).
2. Ustalenie zbioru obiektów oraz zbioru zmiennych porządkowych merytorycznie związanych z badanym zjawiskiem złożonym. Po zgromadzeniu danych konstruuje się macierz danych $[x_{ij}]_{n' \times m}$ (x_{ij} – kategoria j -tej zmiennej dla i -tego obiektu; $i = 1, \dots, n'$ – numer obiektu, $j = 1, \dots, m$ – numer zmiennej).
3. Wśród zmiennych wyróżnia się zmienne preferencyjne (stymulanty, destymulanty, nominanty).
4. Do zbioru obiektów dodaje się wzorzec (górny biegun rozwoju) oraz antywzorzec (dolny biegun rozwoju) otrzymując macierz danych $[x_{ij}]_{n \times m}$ ($n = n' + 2$).

5. Wzmacnia się skalę pomiaru zmiennych wykorzystując metodykę zaproponowaną w pracy Walesiak (2014). Propozycja wzmacniania skali pomiaru zmiennych porządkowych bazuje na odległości GDM2 właściwej do zastosowania dla danych porządkowych. W wyniku przekształcenia zmiennych porządkowych na zmienne metryczne dla destymulant i nominant nastąpi dodatkowo przekształcenie w stymulanty.

6. Oblicza się odległości między obiektami i zestawia w macierz odległości $[\delta_{ik}]$. Zastosowano tutaj kwadrat odległości euklidesowej (uwzględniając wagi zmiennych):

$$\delta_{ik} = \sum_{j=1}^m \alpha_j^2 (x_{ij} - x_{kj})^2, \quad (2)$$

gdzie: $i, k = 1, \dots, n$ – numery obiektów, $j = 1, \dots, m$ – numer zmiennej,

x_{ij} (x_{kj}) – wartość j -tej zmiennej dla i -tego (k -tego) obiektu,

α_j – waga j -tej zmiennej ($\alpha_j \in [0; 1]$ i $\sum_{j=1}^m \alpha_j = 1$).

Można wykorzystać inne miary odległości dla danych metrycznych (odległość miejska, odległość Euklidesa, odległość Czebyszewa, odległość GDM1).

7. Przeprowadza się skalowanie wielowymiarowe: $f: \delta_{ik} \rightarrow d_{ik}$. Skalowanie wielowymiarowe jest metodą reprezentacji macierzy odległości między obiektami w przestrzeni m -wymiarowej $[\delta_{ik}]$ w macierz odległości między obiektami w przestrzeni q -wymiarowej $[d_{ik}]$ ($q < m$) w celu graficznej wizualizacji relacji zachodzących między badanymi obiektami oraz interpretacji wyników. Wymiary q nie są bezpośrednio obserwowalne. Mają one charakter zmiennych ukrytych, które pozwalają na wyjaśnienie podobieństw i różnic między badanymi obiektami. Ze względu na możliwość graficznej prezentacji wyników porządkowania liniowego q wynosi 2. Iteracyjny schemat postępowania w algorytmie **smacof** przedstawiono w pracy (Borg, Groenen, 2005, s. 204–205). Ostatecznie otrzymuje się macierz danych w przestrzeni dwuwymiarowej $[v_{ij}]_{n \times 2}$.

8. Prezentacja graficzna oraz interpretacja wyników w przestrzeni dwuwymiarowej (wyniki skalowania wielowymiarowego) oraz jednowymiarowej (rezultaty porządkowania liniowego):

- na rysunku w przestrzeni dwuwymiarowej (wyniki skalowania wielowymiarowego) łączy się linią prostą punkty oznaczające antywzorzec i wzorzec w tzw. oś zbioru. Wyznacza się od punktu wzorca izokwanty rozwoju (krzywe jednokowego rozwoju)³. Np. podzielenie osi zbioru na 4 części pozwala wyznaczyć 4 izokwanty. Podział osi zbioru na 4 części ma charakter umowny. Można tutaj wykorzystać statystyczne kryteria podziału uwzględniające dwa kryteria: średnią arytmetyczną i odchylenie standardowe lub medianę i medianowe odchylenie bezwzględne (zob. np. Wysocki, 2010, s. 167–169). Obiekty znajdujące się pomiędzy izokwantami prezentują zbliżony poziom rozwoju. Ten sam poziom rozwoju mogą osiągnąć obiekty znajdujące się w różnych punktach na tej samej izokwancie

³ Przebieg izokwant rozwoju zobrazowano z wykorzystaniem funkcji **draw.circle** pakietu **plotrix** (Lemon i inni, 2016).

rozwoju (z uwagi na inną konfigurację obserwacji na zmiennych). Dzięki takiej prezentacji wyników wzbogaca się interpretację wyników porządkowania liniowego;

- oblicza się unormowane odległości d_i^+ obiektu i -tego od wzorca rozwoju zgodnie ze wzorem (Hellwig, 1981, s. 62):

$$d_i^+ = \frac{\sqrt{\sum_{j=1}^2 (v_{ij} - v_{+j})^2}}{\sqrt{\sum_{j=1}^2 (v_{+j} - v_{-j})^2}}, \quad (3)$$

gdzie: $d_i^+ \in [0; 1]$,

$\sqrt{\sum_{j=1}^2 (v_{ij} - v_{+j})^2}$ – odległość Euklidesa obiektu i -tego od obiektu wzorca (górnego bieguna rozwoju),

$\sqrt{\sum_{j=1}^2 (v_{+j} - v_{-j})^2}$ – odległość Euklidesa obiektu wzorca (górnego bieguna rozwoju) od obiektu antywzorca (dolnego bieguna rozwoju).

Porządkuje się obiekty badania według rosnących wartości miary odległości (3). Wyniki porządkowania liniowego przedstawione zostają graficznie na rysunku.

6. WYNIKI BADANIA EMPIRYCZNEGO

W tabeli 2 zaprezentowano dane dotyczące 27 nieruchomości lokalowych na jeleniogórskim rynku nieruchomości opisanych 6 zmiennymi. Celem badania jest przeprowadzenie porządkowania liniowego 27 nieruchomości lokalowych na jeleniogórskim rynku nieruchomości ze względu na ich atrakcyjność.

Mieszkalne nieruchomości lokalowe zostały opisane następującymi zmiennymi:

- x1. Lokalizacja środowiskowa nieruchomości gruntowej, z którą związany jest lokal mieszkalny (1 – zła, 2 – nieodpowiednia, 3 – dostateczna, 4 – dobra, 5 – bardzo dobra).
- x2. Standard użytkowy lokalu mieszkalnego (1 – zły, 2 – niski, 3 – średni, 4 – wysoki).
- x3. Warunki bytowe występujące na nieruchomości gruntowej, z którą związany jest lokal mieszkalny (1 – złe, 2 – przeciętne, 3 – dobre).
- x4. Położenie nieruchomości gruntowej, z którą związany jest lokal mieszkalny, w strefie miasta (1 – centralna, 2 – śródmiejska, 3 – pośrednia, 4 – peryferyjna).
- x5. Typ wspólnoty mieszkaniowej (1 – mała, 2 – duża).
- x6. Powierzchnia gruntu, z którą związany jest lokal mieszkalny (1 – poniżej obrysu budynku, 2 – obrys budynku, 3 – obrys budynku z otoczeniem akceptowalnym, np. na parking, plac zabaw, 4 – obrys budynku z otoczeniem zbyt dużym).

Tabela 2.

Macierz danych

Nr nieruchomości	Zmienne porządkowe					
	x1	x2	x3	x4	x5	x6
1	5	3	1	3	1	3
2	3	3	3	3	2	2
3	5	4	3	4	1	2
4	2	3	1	3	2	3
5	5	4	2	4	1	2
6	4	3	2	3	1	3
7	3	4	3	3	2	2
8	4	4	3	4	1	1
9	5	3	2	4	1	2
10	4	2	1	3	1	3
11	5	4	3	4	1	4
12	4	3	1	4	1	2
13	4	4	3	3	1	1
14	4	4	3	3	2	3
15	5	4	2	3	2	4
16	3	3	2	3	1	1
17	4	2	1	3	2	3
18	4	1	2	4	1	2
19	3	3	2	3	2	4
20	3	2	1	3	1	3
21	4	3	2	3	1	1
22	5	3	2	4	1	2
23	5	4	3	4	1	2
24	4	2	2	3	1	2
25	3	2	1	2	2	3
26	3	3	1	1	2	3
27	2	3	1	1	2	3
Wzorzec	5	4	3	1	1	3
Antywzorzec	1	1	1	4	2	1

Źródło: opracowano na podstawie pracy Pawlukowicz (2006, s. 238).

W artykule zastosowano skrypt programu R przygotowany zgodnie z procedurą badawczą z sekcji 5, w której zastosowano następującą metodykę postępowania:

- zmienne x_1 , x_2 i x_3 są stymulantami, zmienne x_4 i x_5 – destymulantami, a zmienna x_6 jest nominantą o kategorii nominalnej (najkorzystniejszej) wynoszącej 3.
- do zbioru 27 nieruchomości lokalowych dodano wzorzec i antywzorzec. Macierz danych obejmuje zatem 29 obiektów opisanych 6 zmiennymi (zob. tabela 2).
- wzmacnia się skalę pomiaru zmiennych x_1 – x_6 wykorzystując metodykę zaproponowaną w pracy (Walesiak, 2014). Propozycja wzmacniania skali pomiaru zmiennych porządkowych bazuje na odległości GDM2 właściwej do zastosowania dla danych porządkowych. Ze względu na to, że metoda wzmacniania skali pomiaru zmiennych porządkowych z wykorzystaniem odległości GDM2 dotyczy każdej zmiennej z osobna, wzór na odległość GDM2 w tej sytuacji jest następujący:

$$d_{iw} = \frac{1}{2} - \frac{a_{iwj}b_{wij} + \sum_{l=1}^n a_{ilj}b_{wlj}}{2 \left[\sum_{l=1}^n a_{ilj}^2 \cdot \sum_{l=1}^n b_{wlj}^2 \right]^{\frac{1}{2}}} \quad \text{dla } j = 1, \dots, m, \quad (4)$$

gdzie: wyjaśnienie symboli a_{ipj} , b_{wrj} (dla $p = w$, $l; r = i, l$) znajduje się w tabeli 1.

Do przekształcenia zmiennej porządkowej w zmienną metryczną zastosowany zostanie wzór:

$$s_{iw} = 1 - d_{iw} \quad \text{dla } j = 1, \dots, m. \quad (5)$$

W wyniku zastosowania wzoru (5) nastąpi wzmocnienie skali porządkowej w skalę metryczną zgodnie ze schematem:

$$\begin{array}{ccc} \begin{array}{c} \left[\begin{array}{c} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{nj} \end{array} \right] & \Rightarrow & \begin{array}{c} \text{obliczenie podobieństw (5)} \\ \text{bazujących na odległości} \\ \text{GDM2 od obiektu wzorca} \end{array} & \begin{array}{c} \left[\begin{array}{c} s_{1j} \\ \vdots \\ s_{ij} \\ \vdots \\ s_{nj} \end{array} \right] & \Rightarrow & \text{dane metryczne.} \end{array} \end{array}$$

W tej sytuacji we wzorze (4) x_{wj} ($j = 1, \dots, m$) oznaczać będzie kategorię najbardziej korzystną spośród wszystkich kategorii danej zmiennej. Dla stymulanty i destymulanty jest to kategoria odpowiednio maksymalna i minimalna. Z kolei dla nominanty jednodomodalnej jest to kategoria nominalna zmiennej. W wyniku takiego przekształcenia

zmiennej porządkowej na zmienną metryczną dla destymulanty i nominanty nastąpi dodatkowo przekształcenie w stymulantę.

W tabeli 3 zaprezentowano dane dotyczące 27 nieruchomości lokalowych na jeleniogórskim rynku nieruchomości opisanych 6 zmiennymi po wzmocnieniu skali pomiaru zmiennych $x_1 - x_6$.

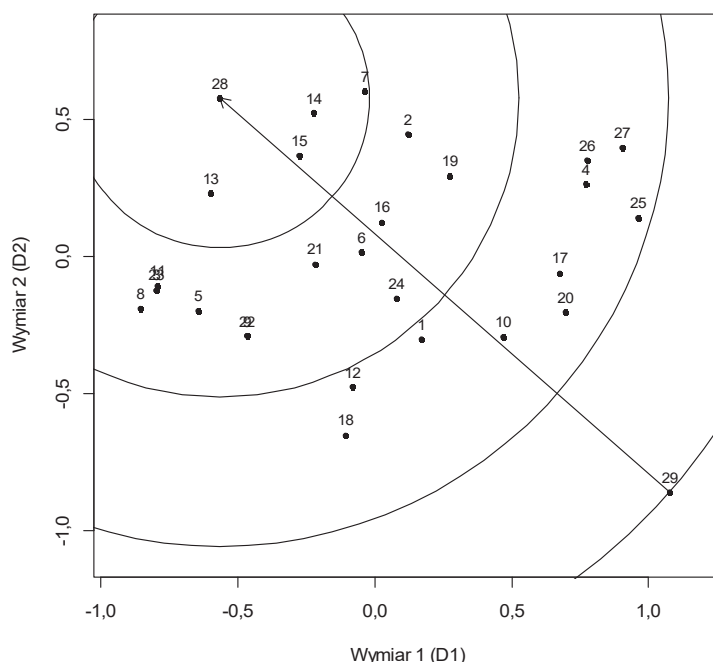
Tabela 3.

Macierz danych po wzmocnieniu skali pomiaru zmiennych

Nr nieruchomości	Zmienne metryczne					
	s1	s2	s3	s4	s5	s6
1	1	0,662221	0,225	0,702548	1	1
2	0,313499	0,662221	1	0,702548	0,465414	0,684466
3	1	1	1	0,12725	1	0,684466
4	0,140809	0,662221	0,225	0,702548	0,465414	1
5	1	1	0,725	0,12725	1	0,684466
6	0,725	0,662221	0,725	0,702548	1	1
7	0,313499	1	1	0,702548	0,465414	0,684466
8	0,725	1	1	0,12725	1	0,311438
9	1	0,662221	0,725	0,12725	1	0,684466
10	0,725	0,247643	0,225	0,702548	1	1
11	1	1	1	0,12725	1	0,817526
12	0,725	0,662221	0,225	0,12725	1	0,684466
13	0,725	1	1	0,702548	1	0,311438
14	0,725	1	1	0,702548	0,465414	1
15	1	1	0,725	0,702548	0,465414	0,817526
16	0,313499	0,662221	0,725	0,702548	1	0,311438
17	0,725	0,247643	0,225	0,702548	0,465414	1
18	0,725	0,109801	0,725	0,12725	1	0,684466
19	0,313499	0,662221	0,725	0,702548	0,465414	0,817526
20	0,313499	0,247643	0,225	0,702548	1	1
21	0,725	0,662221	0,725	0,702548	1	0,311438
22	1	0,662221	0,725	0,12725	1	0,684466
23	1	1	1	0,12725	1	0,684466
24	0,725	0,247643	0,725	0,702548	1	0,684466
25	0,313499	0,247643	0,225	0,937014	0,465414	1
26	0,313499	0,662221	0,225	1	0,465414	1
27	0,140809	0,662221	0,225	1	0,465414	1
Worzec	1	1	1	1	1	1
Antyworzec	0,084773	0,109801	0,225	0,12725	0,465414	0,311438

Źródło: obliczenia własne z wykorzystaniem pakietu clusterSim (Walesiak, Dudek, 2016a) programu R.

- macierz odległości $[\delta_{ik}]$ między obiektami obliczono z wykorzystaniem kwadratu odległości euklidesowej (2), dla której zastosowano wagi jednakowe.
- przeprowadzono skalowanie wielowymiarowe 29 obiektów (27 nieruchomości plus wzorzec i antywzorzec) ze względu na poziom atrakcyjności nieruchomości lokalowych z wykorzystaniem funkcji **smacofSym** pakietu **smacof** (Mair i inni, 2016) otrzymując konfigurację 29 obiektów (punktów) w przestrzeni dwuwymiarowej $[v_{ij}]_{29 \times 2}$.
- na rysunku 1 przedstawiono graficzną prezentację wyników skalowania wielowymiarowego 29 obiektów. Antywzorzec (obiekt 29) i wzorzec (obiekt 28) połączono linią prostą otrzymując tzw. oś zbioru. Wyznaczono 4 izokwanty rozwoju dzieląc oś zbioru na 4 równe części.



Rysunek 1. Graficzna prezentacja wyników skalowania wielowymiarowego w przestrzeni dwuwymiarowej 29 obiektów obejmujących 27 nieruchomości, wzorzec (obiekt 28) i antywzorzec (obiekt 29) ze względu na poziom atrakcyjności

Źródło: opracowanie własne z wykorzystaniem programu R.

- obliczono odległości każdego obiektu (nieruchomości) od wzorca rozwoju zgodnie ze wzorem (3). Uporządkowano nieruchomości według rosnących wartości miary (3), a następnie wyodrębniono 4 klasy nieruchomości podobnych pod względem poziomu atrakcyjności. Uporządkowanie 29 obiektów obejmujących 27 nieruchomości, wzorzec (obiekt 28) i antywzorzec (obiekt 29) ze względu na poziom atrakcyjności według rosnących wartości miary (3) prezentuje tabela 4.

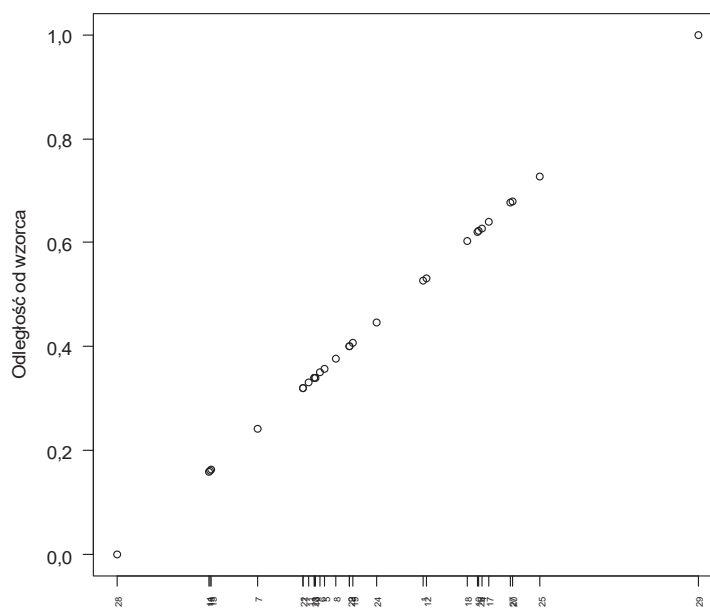
Tabela 4.

Uporządkowanie nieruchomości według rosnących wartości miary (3)

Nr nieruchomości	Odległość
28	0,0000000
14	0,1583928
13	0,1603719
15	0,1632594
7	0,2420548
2	0,3195323
21	0,3198637
11	0,3306159
3	0,3392882
23	0,3392882
16	0,3409411
6	0,3499601
5	0,3579128
8	0,3759838
9	0,4007961
22	0,4007961
19	0,4064514
24	0,4469689
1	0,5260473
12	0,5322337
18	0,6023688
10	0,6201668
26	0,6230692
4	0,6277490
17	0,6398830
27	0,6771382
20	0,6800868
25	0,7277380
29	1,0000000

Źródło: obliczenia własne z wykorzystaniem programu R.

Graficznie wyniki uporządkowania liniowego 29 obiektów obejmujących 27 nieruchomości, wzorzec (obiekt 28) i antywzorzec (obiekt 29) ze względu na poziom atrakcyjności według rosnących wartości miary (3) przedstawia rysunek 2.



Rysunek 2. Graficzna prezentacja wyników porządkowania liniowego 29 obiektów obejmujących 27 nieruchomości, wzorzec (obiekt 28) i antywzorzec (obiekt 29) ze względu na poziom atrakcyjności według rosnących wartości miary (3)

Źródło: opracowanie własne z wykorzystaniem programu R.

Taka forma prezentacji wyników pozwala na:

- przedstawienie uporządkowania nieruchomości ze względu na poziom atrakcyjności według wartości miary (3) oraz w formie prezentacji graficznej na rysunku 2,
- wyodrębnienie klas nieruchomości lokalowych (nieruchomości znajdujące się pomiędzy izokwantami) o zbliżonym poziomie atrakcyjności (zob. rysunek 1),
- zidentyfikowanie nieruchomości o zbliżonym poziomie atrakcyjności, ale różniących się położeniem na izokwancie rozwoju (zob. rysunek 1). Np. nieruchomości 2 i 21, 13 i 14, 19 i 22 oraz 19 i 9 mają zbliżony poziom atrakcyjności, ale różnią się położeniem na izokwancie rozwoju. Zatem nieruchomości te osiągają zbliżony poziom rozwoju, ale mają istotnie różniące się konfiguracje obserwacji na zmiennych.

7. PODSUMOWANIE

W artykule przedstawiono propozycję procedury badawczej pozwalającą na wizualizację wyników porządkowania liniowego zbioru obiektów dla danych porządkowych wykorzystując do realizacji tego celu skalowanie wielowymiarowe. Przy ocenie poprawności wyników skalowania wielowymiarowego w kroku 7 zaprezentowanej procedury badawczej należy wziąć pod uwagę wartość funkcji dopasowania *STRESS*, udziały procentowe obiektów w wartości miary dopasowania *STRESS* (ang. *stress per point*) oraz interpretowalność wyników. W funkcji **smacofSym** pakietu **smacof** stosowana jest miara dopasowania *STRESS-1* Kruskala (Borg, Groenen, 2005, s. 250–254). Z punktu widzenia skalowania wielowymiarowego pożądane jest jak najmniejsze odchylenie rozkładu błędów dla poszczególnych obiektów od rozkładu równomiernego. Dodatkowymi kryteriami akceptowalności wyników skalowania wielowymiarowego są wykresy „Residual plot” oraz „Shepard diagram”, które pozwalają ocenić dopasowanie wybranego modelu skalowania oraz zidentyfikować obiekty odosobnione (De Leeuw, Mair, 2015). Szerzej o tym zagadnieniu traktuje praca Walesiak, Dudek (2016b).

Koncepcja izokwant i ścieżki rozwoju zaproponowana w pracy (Hellwig, 1981) pozwala na graficzną prezentację wyników porządkowania liniowego tylko dla dwóch zmiennych. Zastosowanie skalowania wielowymiarowego rozszerza możliwości zastosowania wizualizacji wyników porządkowania liniowego dla m zmiennych. Dzięki takiemu rozwiązaniu wzbogacono interpretację wyników porządkowania liniowego.

Zaproponowane podejście zilustrowano przykładem empirycznym z zastosowaniem skryptu przygotowanego w środowisku R (R Development Core Team, 2016).

LITERATURA

- Antczak E., (2013), Przestrzenny taksonomiczny miernik rozwoju, *Wiadomości Statystyczne*, 7, 37–53.
- Borg I., Groenen P. J. F., (2005), *Modern Multidimensional Scaling. Theory and Applications*, 2nd Edition, Springer Science+Business Media, New York.
- Borg I., Groenen P. J. F., Mair P., (2013), *Applied Multidimensional Scaling*, Springer, Heidelberg, New York, Dordrecht, London.
- Borys T., (1984), *Kategoria jakości w statystycznej analizie porównawczej*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, 284, Seria: Monografie i Opracowania, 23.
- Borys T., Strahl D., Walesiak M., (1990), *Wkład ośrodka wrocławskiego w rozwój teorii i zastosowań metod taksonomicznych*, w: Pocięcha J., (red.), *Taksonomia – teoria i zastosowania*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków, 12–23.
- De Leeuw J., Mair P., (2015), Shepard Diagram, Wiley StatsRef: Statistics Reference Online, John Wiley & Sons Ltd.
- Hellwig Z., (1967), *Procedure of Evaluating High-Level Manpower Data and Typology of Countries by Means of the Taxonomic Method*, COM/WS/91, Warsaw, 9 December, 1967, UNESCO working paper.
- Hellwig Z., (1968), Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju i strukturę wykwalifikowanych kadr, *Przegląd Statystyczny*, 15 (4), 307–327.
- Hellwig Z., (1972), *Procedure of Evaluating High-Level Manpower Data and Typology of Countries by Means of the Taxonomic Method*, w: Gostkowski Z., (red.), *Towards a System of Human Resources*

- Indicators for Less Developed Countries*, Papers Prepared for UNESCO Research Project, Ossolineum, The Polish Academy of Sciences Press, Wrocław, 115–134.
- Hellwig Z., (1981), *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych*, w: Welfe W., (red.), *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, PWE, Warszawa, 46–68.
- Hwang C. L., Yoon K., (1981), *Multiple Attribute Decision Making – Methods and Applications. A State-of-the-Art Survey*, New York, Springer-Verlag.
- Jefmański B., Dudek A., (2016), *Syntetyczna miara rozwoju Hellwiga dla trójkątnych liczb rozmytych*, w: Appenzeller D. (red.), *Matematyka i informatyka na usługach ekonomii. Wybrane problemy modelowania i prognozowania zjawisk gospodarczych*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, 29–40.
- Lemon J. et al., (2016), plotrix: Various Plotting Functions. R package version 3.6-3, URL <http://CRAN.R-project.org/package=plotrix>.
- Mair P., De Leeuw J., Borg I., Groenen P. J. F., (2016), smacof: Multidimensional Scaling. R package version 1.8-13, URL <http://CRAN.R-project.org/package=smacof>.
- Młodak A., (2014), On the Construction of an Aggregated Measure of the Development of Interval Data, *Computational Statistics*, 29 (5), 895–929.
- Pawlukowicz R., (2006), Klasyfikacja w wyborze nieruchomości podobnych dla potrzeb wyceny rynkowej nieruchomości, *Ekonometria* 16, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1100, 232–240.
- Pietrzak M. B., (2014), Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych, *Przegląd Statystyczny*, 61 (2), 181–201.
- Pociecha J., Zajac K., (1990), *Wkład ośrodka krakowskiego w rozwój teorii i zastosowań metod taksonomicznych*, w: Pociecha J., (red.), *Taksonomia – teoria i zastosowania*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków, 24–32.
- R Development Core Team, (2016), *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, URL <http://www.R-project.org>.
- Stevens S. S., (1946), On the Theory of Scales of Measurement, *Science*, 103 (2684), 677–680.
- Walesiak M., (1993), *Statystyczna analiza wielowymiarowa w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 654, Seria: Monografie i Opracowania, 101.
- Walesiak M., (1996), *Metody analizy danych marketingowych*, PWN, Warszawa.
- Walesiak M., (1999), Distance Measure for Ordinal Data, *Argumenta Oeconomica*, 2 (8), 167–173.
- Walesiak M., (2011), Porządkowanie liniowe z wykorzystaniem uogólnionej miary odległości GDM2 dla danych porządkowych, *Ekonometria*, 30, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 163, 9–18.
- Walesiak M., (2014), Wzmacnianie skali pomiaru w statystycznej analizie wielowymiarowej, *Taksonomia* 22, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 327, 60–68.
- Walesiak M., (2016a), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*. Wydanie drugie poprawione i rozszerzone, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław.
- Walesiak M., (2016b), Visualization of Linear Ordering Results for Metric Data with the Application of Multidimensional Scaling, *Ekonometria*, 2 (52), 9–21.
- Walesiak M., Dudek A., (2016a), clusterSim: Searching for Optimal Clustering Procedure for a Data Set. R package version 0.45-1, URL <http://CRAN.R-project.org/package=clusterSim>.
- Walesiak M., Dudek A., (2016b), Wybór optymalnej procedury skalowania wielowymiarowego dla danych metrycznych z wykorzystaniem programu R, Referat na XXXV Konferencję Naukową nt. „Multivariate Statistical Analysis. MSA 2016”, Łódź, 7–9 listopada 2016 r.
- Wysocki F., (2010), *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań.

WIZUALIZACJA WYNIKÓW PORZĄDKOWANIA LINIOWEGO
DLA DANYCH PORZĄDKOWYCH
Z WYKORZYSTANIEM SKALOWANIA WIELOWYMIAROWEGO

Streszczenie

W artykule zaproponowano dwukrokową procedurę badawczą pozwalającą na wizualizację wyników porządkowania liniowego dla danych porządkowych. W pierwszym kroku w wyniku zastosowania skalowania wielowymiarowego (zob. Borg, Groenen, 2005; Mair i inni, 2016) otrzymuje się wizualizację obiektów w przestrzeni dwuwymiarowej. W następnym kroku przeprowadza się porządkowanie liniowe zbioru obiektów na podstawie odległości Euklidesa od wzorca rozwoju. Zaproponowane podejście zilustrowano przykładem empirycznym z zastosowaniem skryptu przygotowanego w środowisku R.

W artykule wykorzystano koncepcję izokwant i ścieżki rozwoju (osi zbioru – najkrótszej drogi łączącej wzorzec i antywzorzec rozwoju) zaproponowaną w pracy Hellwig (1981). Zaproponowane podejście rozszerzyło możliwości interpretacyjne wyników porządkowania liniowego zbioru obiektów.

Słowa kluczowe: porządkowanie liniowe, skalowanie wielowymiarowe, dane porządkowe, miary agregatowe, program R

VISUALIZATION OF LINEAR ORDERING RESULTS FOR ORDINAL DATA
WITH APPLICATION OF MULTIDIMENSIONAL SCALING

Abstract

A two-step procedure was proposed to visualization of linear ordering results for ordinal data. In the first step as a result of the application of multidimensional scaling (see Borg, Groenen, 2005; Mair et al., 2016) is to visualize objects in two-dimensional space. In the next step, a linear ordering is carried out with the use of the Euclidean distance from the pattern (ideal) object. The proposed approach expanded the possibilities of interpretation of the results of the linear ordering of set of objects.

The article uses the concept of isoquant and path of development (the shortest way connecting ideal and anti-ideal object) proposed by Hellwig (1981). The proposed approach is illustrated by an empirical example with application of script of R environment.

Keywords: linear ordering, multidimensional scaling, ordinal data, composite measures, R environment

