

Graficzna prezentacja struktury empirycznych danych wielowymiarowych: opis i zastosowanie metody

Mikołaj Rybaczuk*, Joanicjusz Nazarko
Wydział Zarządzania Politechniki Białostockiej

Mirosława Czerniawska
Wydział Pedagogiki i Psychologii Uniwersytetu w Białymstoku

GRAPHIC PRESENTATION
OF MULTIDIMENSIONAL EMPIRICAL DATA STRUCTURE:
METHOD DESCRIPTION AND APPLICATION

Abstract. This survey paper introduces a new method of graphic presentation of multidimensional data – RGM. The method enables the observation of objects' structure as well as object-feature and feature-feature relations. The algorithm assumes that images of features and objects are presented as points on the plane. The features are placed on a circle with a radius of value 1, and objects inside it, on the surface delimited by the circle. Standardized values of the features (x_{ij} – i -th object, j -th feature) in the data table correspond with the distances between the points of placing the features and objects (d_{ij}) on the plane. The minimum of the function is sought considering the limits that guarantee placing the features on the circle and the objects within it. The algorithm allows organizing points of placing features and objects in the way that minimize the divergence between the values of features describing objects and object-feature distances on the plane. As a result, the observation of the full data structure (object-feature, object-object and feature-feature relations) is possible.

Cechą charakterystyczną badań z dziedziny nauk społecznych jest wieloaspektowość, która na poziomie metodologicznym odzwierciedla się w mnogo-

* Adres do korespondencji: Politechnika Białostocka, Wydział Zarządzania, Katedra Informatyki Gospodarczej i Logistyki, ul. Ojca Stefana Tarasiuka 2, 16-001 Kleosin; e-mail: rym@onet.pl

Praca częściowo została wykonana w ramach badań własnych W/WZ/3/02, finansowanych przez Komitet Badań Naukowych.

ści zmiennych, a w konsekwencji – również w mnogości wskaźników będących operacjonalizacją tych zmiennych. Metody analiz statystycznych, takie jak analiza wariancji i analiza regresji, pozwalają na weryfikację wieloczynnikowo warunkowanych zależności, niemniej jednak znane sposoby graficznej prezentacji danych nie są w stanie odzwierciedlić pełnej złożoności relacji zachodzących między poszczególnymi zmiennymi. Dostępne w programach statystycznych i opisane w literaturze metody wizualne (metody aglomeracyjne, skalowanie wielowymiarowe, analiza skupień, metoda składowych głównych) ukazują jedynie relacje między obiektami lub między cechami, ale jednocześnie powodują utratę informacji o wpływie cech na grupowanie obiektów (Zeliaś, 1989; Grabiński, 1992; Jajuga, 1993; Gordon, 1999). W konsekwencji uniemożliwia to ocenę siły poszczególnych cech w różnicowaniu stanów badanych obiektów. Gabriel (1971) zaproponował metodę (zwaną *biplot*) umożliwiającą graficzną ocenę relacji obiekt–cecha, która stanowi pewne rozszerzenie metody składowych głównych. Jest ona opisana także przez Gordona (1999). Jej realizację zawiera program statystyczny NTSYSpc (Numerical Taxonomy and Multivariate Analysis System). Bliższe informacje na temat programu można znaleźć na stronie internetowej <http://www.exetersoftware.com/cat/statistics-software.html>. Jednak – zdaniem autorów niniejszego artykułu – interpretacja i czytelność uzyskiwanej prezentacji graficznej jest niezadowolająca.

„Niedoskonałość” aktualnie stosowanych metod prezentacji graficznej nie pomniejsza ich znaczenia w analizach statystycznych; wskazuje raczej na konieczność poszukiwania nowych sposobów rozwiązań. Należy bowiem mieć na względzie, że opisywanie przez autorów i odbiór przez czytelników wieloaspektowych i cechujących się dużą złożonością informacji jest szczególnie utrudnione przy wyłącznym wykorzystaniu kodu werbalnego. Kod obrazowy, który aktywizowany jest przy różnego rodzaju metodach graficznych, staje się swoistego rodzaju „narzędziem” pozwalającym zwiększyć efektywność przetwarzania informacji. Informacje graficzne przyczyniają się między innymi do zwiększonej efektywności procesów pamięci, myślenia i rozwiązywania problemów. Graficzne metody prezentacji danych mają więc znaczenie o tyle, o ile przyczyniają się do dokonania względnie łatwej, ale jednocześnie trafnej reprezentacji w ludzkim umyśle dostarczonych wizualnie informacji. Wystarczająco adekwatna reprezentacja daje podstawy do lepszego zrozumienia bogactwa informacji zawartych w danych empirycznych.

METODA GRAFICZNEJ PREZENTACJI STRUKTURY DANYCH WIELOWYMIAROWYCH (RGM)

W niniejszym artykule opisana została metoda graficzna RGM, która pozwala – zdaniem jej autora (Rybaczuk, 2002) – na odwzorowanie umożliwiające obejrzenie obrazu struktury danych wielowymiarowych na płaszczyźnie. Uzyskany za pomocą proponowanej metody obraz struktury danych ułatwia: „uświadomienie” istnienia skupień i określenie ich wyrazistości; ocenę współzależno-

ści między cechami; ocenę udziału poszczególnych cech w tworzeniu się skupień; rozróżnianie stanów badanych obiektów.

Oznaczenia i założenia metody RGM

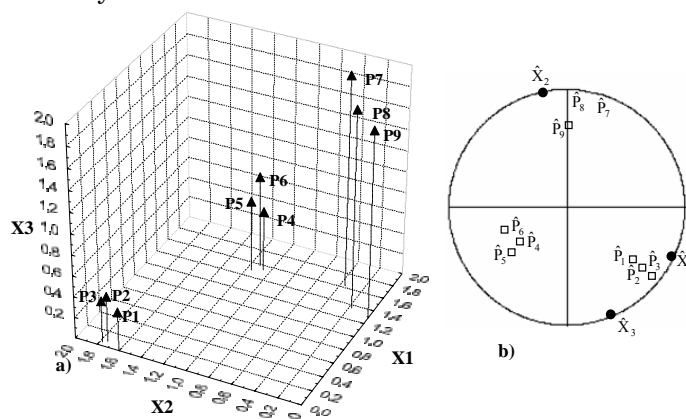
Przez P_i ($i = 1, 2, \dots, n$) oznaczono i -ty obiekt badania spośród zbioru n obiektów $\{P_1, P_2, \dots, P_n\}$, zaś przez \hat{P}_i – jego obraz na płaszczyźnie ($i = 1, 2, \dots, n$). Przez X_j oznaczono j -tą cechę spośród zbioru k cech $\{X_1, X_2, \dots, X_k\}$ opisujących obiekty, a przez \hat{X}_j – jej obraz na płaszczyźnie ($j = 1, 2, \dots, k$). Dane uzyskiwane w trakcie realizacji badań zapisywane są z reguły w postaci prostokątnej tablicy, w której cechy stanowią kolumny, zaś badane obiekty (przypadki) – wiersze. Dane dotyczące n obiektów opisanych k cechami zapisano jako:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

x_{ij} oznacza wartość j -tej cechy dla i -tego obiektu.

Na etapie analizy wyników badań, stawiania hipotez i wnioskowania ważne jest poznanie struktury wielowymiarowych danych, rozumianej jako relacje: obiekt–obiekt (podobieństwo obiektów – obiekty podobne w przestrzeni cech są blisko siebie, zaś niepodobne są odległe); cecha–cecha (cechy podobne – silnie skorelowane, niepodobne – słabo skorelowane; obiekt–cecha (poziom cechy dla poszczególnych obiektów, ich wpływ na grupowanie się obiektów w skupienia).

Przez S oznaczono strukturę danych w k -wymiarowej przestrzeni cech (rysunek 1). Celem metody jest uzyskanie na płaszczyźnie, w kole o promieniu 1, obrazu S^* struktury S .



Rysunek 1. Przykładowa ilustracja struktur: a) S i b) S^* danych wielowymiarowych; P_1, \dots, P_9 – obiekty w przestrzeni cech; X_1, \dots, X_3 – cechy opisujące obiekty; $\hat{P}_1, \dots, \hat{P}_9$ – obrazy obiektów; $\hat{X}_1, \dots, \hat{X}_3$ – obrazy cech. Źródło: opracowanie własne.

Obrazy obiektów mają być optymalnie, według przyjętego kryterium, rozmieszczone jako punkty wewnątrz koła, zaś obrazy cech – jako punkty na okręgu. Przy takim założeniu co do rozmieszczenia obrazów obiektów i obrazów cech widać, iż odległość między obrazem cechy i obrazem obiektu nie może przekroczyć średnicy koła równej 2. W celu uzyskania porównywalności struktur S i S^* dokonuje się normalizacji danych zawartych w macierzy X tak, aby znormalizowane wartości z_{ij} (j -tej cechy dla i -tego obiektu) zawierały się w przedziale $[0;2]$. Przykładową metodą normalizacji spełniającą ten wymóg może być:

$$(1) \quad z_{ij} = \frac{x_{ij} - \bar{x}_j}{\max_i |x_{ij} - \bar{x}_j|} + 1 \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k),$$

gdzie \bar{x}_j jest średnią arytmetyczną j -tej cechy. Ten sposób normalizacji przekształca wartość minimalną cechy w 0, zaś maksymalną w 2.

W x_j daniach z zakresu nauk społecznych często występuje taka sytuacja, że pomiaru wielu cech dokonuje się w skali porządkowej. Wszystkie uwzględniane cechy osiągają minimalną wartość równą a , zaś maksymalną równą b . Środek zakresu zmienności każdej z takich cech oznaczono przez M . Jeżeli celem jest ocena osiąganego poziomu każdej z cech opisujących obiekty, to można zastosować inny sposób normalizacji.

$$(2) \quad z_{ij} = \frac{x_{ij} - M}{\max_i |x_{ij} - M|} + 1 \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k).$$

Przykładowo, gdy cechami jest 18 wartości ostatecznych ze Skali Wartości M. Rokeacha (Brzozowski, 1989) i dąży się do oceny preferencji poszczególnych wartości dla badanych obiektów (według odległości obrazów cecha–obiekt), to zgodnie z przyjętymi wyżej oznaczeniami minimalna ranga $a = 1$, maksymalna ranga $b = 18$, zaś środek zakresu zmienności $M = 9,5$.

Wartości cech po normalizacji zapisano w macierzy \mathbf{Z} , zaś odległości pomiędzy obiektami w przestrzeni znormalizowanych cech w symetrycznej macierzy \mathbf{D} :

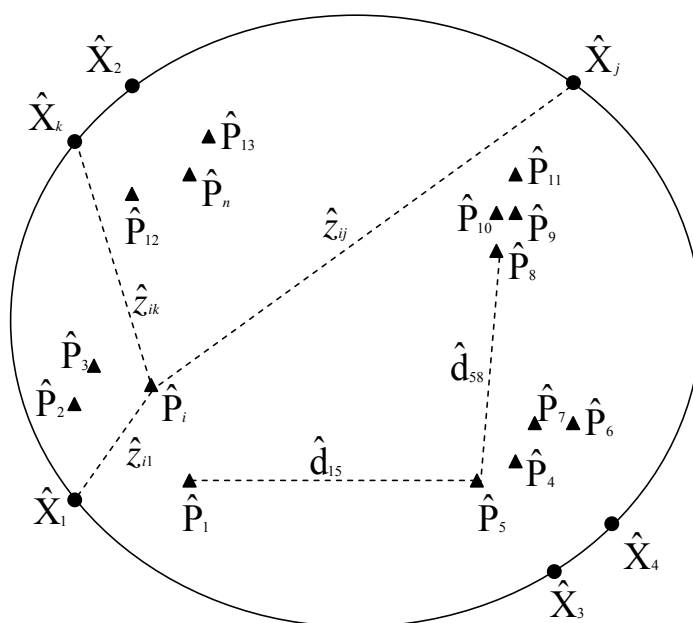
$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1k} \\ z_{21} & z_{22} & \cdots & z_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ z_{n1} & z_{n2} & \cdots & z_{nk} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}.$$

Interpretację obrazu struktury danych na płaszczyźnie przedstawiono na rysunku 1.

Oznaczono przez \hat{z}_{ij} odległości między obrazami i -tego obiektu \hat{p}_i i j -tej cechy \hat{x}_j oraz zapisano w macierzy \hat{Z} ($i = 1, 2, \dots, n; j = 1, 2, \dots, k$), zaś przez \hat{d}_{il} odległości między obrazami i -tego obiektu \hat{p}_i a l -tego obiektu \hat{p}_l i zapisano w symetrycznej macierzy \hat{D} ($i = 1, 2, \dots, n; l = 1, 2, \dots, n$):

$$\hat{Z} = \begin{bmatrix} \hat{z}_{11} & \hat{z}_{12} & \dots & \hat{z}_{1k} \\ \hat{z}_{21} & \hat{z}_{22} & \dots & \hat{z}_{2k} \\ \dots & \dots & \dots & \dots \\ \hat{z}_{n1} & \hat{z}_{n2} & \dots & \hat{z}_{nk} \end{bmatrix}, \quad \hat{D} = \begin{bmatrix} 0 & \hat{d}_{12} & \dots & \hat{d}_{1n} \\ \hat{d}_{21} & 0 & \dots & \hat{d}_{2n} \\ \dots & \dots & \dots & \dots \\ \hat{d}_{n1} & \hat{d}_{n2} & \dots & 0 \end{bmatrix}.$$

Na rysunku 2 przedstawiono istotę konstrukcji obrazu S^* struktury danych S .



Rysunek 2. Interpretacja obrazu struktury danych (relacji obiekt–obiekt, cecha–cecha, obiekt–cecha)

Opis algorytmu optymalizacji

Z przyjętych założeń i oznaczeń wynika, że pełna zgodność struktur S i S^* zostanie osiągnięta, gdy macierze Z i \hat{Z} będą równe, tzn. $z_{ij} = \hat{z}_{ij}$ dla $i = 1, 2, \dots, n$ oraz $j = 1, 2, \dots, k$. Ze względu na znaczną redukcję wymiaru przestrzeni przy przejściu z k -wymiarowej przestrzeni cech na płaszczyznę, z reguły uzyskanie tych równości nie jest możliwe. Przedstawiony algorytm ma na celu zminimalizowanie różnic między elementami macierzy Z i \hat{Z} , przez co uzyskuje się możliwie najtrafniejsze wyobrażenie o strukturze S , na podstawie wizualnej oceny jej obrazu S^* . W algorytmie przyjęto, że środek okręgu o promieniu 1 znajduje się w punkcie $(0,0)$ kartezjańskiego układu współrzędnych.

Przyjęto, że pary liczb (y_{i1}, y_{i2}) , $i = 1, 2, \dots, n$ są współrzędnymi obrazu i -tego obiektu, pary liczb $(\hat{y}_{j1}, \hat{y}_{j2})$, $j = 1, 2, \dots, k$ – współrzędnymi obrazu \hat{X}_j j -tej cechy, elementy \hat{Z}_{ij} macierzy są odległościami euklidesowymi zdefiniowanymi jako:

$$\hat{z}_{ij} = \sqrt{(\hat{y}_{j1} - y_{i1})^2 + (\hat{y}_{j2} - y_{i2})^2} \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k), \quad (3)$$

Zastosowano następujące kryterium minimalizacji rozbieżności struktur S i S^* :

$$F(Z, \hat{Z}) = \sum_{i=1}^n \sum_{j=1}^k (z_{ij} - \hat{z}_{ij})^2 = \text{minimum} \quad (4)$$

przy ograniczeniach:

$$\begin{aligned} y_{i1}^2 + y_{i2}^2 &\leq 1, \quad i = 1, 2, \dots, n, \\ \hat{y}_{j1}^2 + \hat{y}_{j2}^2 &= 1, \quad j = 1, 2, \dots, k. \end{aligned} \quad (5)$$

Ograniczenia gwarantują, że obrazy cech znajdują się na okręgu, a obiektów – w jego wnętrzu.

Ocena jakości odwzorowania (zgodności struktur S i S^*)

Wartość osiągniętego minimum funkcji $F(D)$ nie jest dobrym miernikiem stopnia zgodności struktur, gdyż zależy ona od liczby obiektów i cech.

A. Ocena zgodności struktury obiekt–obiekt. Przez δ_{ij} oznaczono euklidesowe odległości między i -tym a j -tym obiektami opisanymi znormalizowanymi cechami w k -wymiarowej przestrzeni cech:

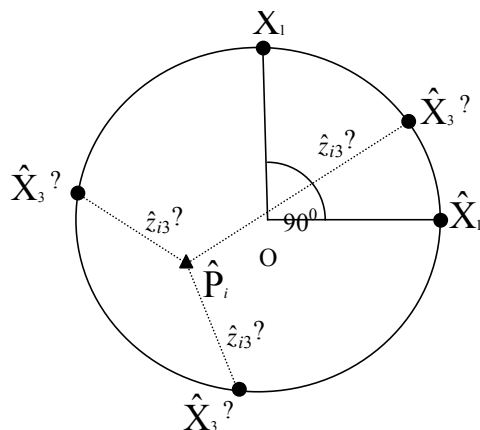
$$\delta_{ij} = \sqrt{\sum_{l=1}^k (z_{jl} - z_{il})^2} \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, n). \quad (6)$$

δ_{ij}^* oznacza odległości między obrazami i -tego i j -tego obiektów na płaszczyźnie:

$$\delta_{ij}^* = \sqrt{\sum_{l=1}^2 (y_{jl} - y_{il})^2} \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, n). \quad (7)$$

Współczynnik korelacji liniowej r_{dd} między odległościami δ_{ij} a δ_{ij}^* może być wykorzystany do oceny zgodności struktur. Im ten współczynnik korelacji jest bliższy jedności, tym struktury obiektów i ich obrazów są bardziej zgodne. Ta prosta miara zgodności może być wykorzystana do analizy porównawczej wyników graficznej prezentacji struktur uzyskanych różnymi metodami (np. skalowania wielowymiarowego, składowych głównych, prezentowaną metodą RGM).

B. Ocena zgodności struktury obiekt–cecha. Występowanie w strukturze S cech słabo skorelowanych z pozostałymi cechami może powodować, że ich obrazy w strukturze S* będą rozmieszczone dosyć przypadkowo. Przykładowo, przy ortogonalnych (zupełnie między sobą nieskorelowanych) cechach X_1 , X_2 , X_3 , obrazy dwóch z nich (np. X_1 i X_2) będą rozmieszczone na okręgu tak, że między odcinkami łączącymi środek okręgu z punktami zaczepienia obrazów cech będzie kąt prosty. Położenie na okręgu obrazu trzeciej z tych cech (np. X_3) będzie dość przypadkowe, a w konsekwencji odległość między obiektem P_i a cechą X_3 będzie również przypadkowa (rys. 3).



Rysunek 3. Wpływ przypadkowego umieszczenia obrazu cechy \hat{X}_3 na odległość \hat{Z}_{13} od obrazu \hat{P}_i

Wskazanie cech, których obrazy na okręgu umiejscowione są przypadkowo, uchroni przed wyciąganiem fałszywych wniosków z obrazu struktury S*. Proponowana metoda graficzna RGM pozwala na łatwe wskazanie takich

cech. Jako miarę jakości wpasowania obrazów każdej z k cech opisujących obrazy obiektów można przyjąć współczynnik korelacji liniowej r_{zz} między wielkościami z_{*j} a \hat{z}_{*j} dla każdej z cech ($j = 1, 2, \dots, k$). Cechy, dla których uzyskano współczynniki korelacji znacznie poniżej +1 (np. poniżej 0,7), są cechami zakłócającymi obraz struktury. Eliminacja wybranych cech spośród zestawu aktualnie analizowanych prowadzi do „oglądania” struktury danych w różnych rzutach (podobnie jak oglądanie struktury obiektów w układzie różnych czynników w metodzie składowych głównych).

C. Ocena zgodności struktury cecha-cecha. Zgodność struktury cecha-cecha w k -wymiarowej przestrzeni cech i jej obrazu na płaszczyźnie jest równoważna zgodności współczynników korelacji liniowej co do znaku i wartości bezwzględnej między każdą parą cech (z_{*j}, z_{*l}) i odpowiadającą jej parą obrazów cech ($\hat{z}_{*j}, \hat{z}_{*l}$) ($j, l = 1, 2, \dots, k$). Jeśli siła wzajemnego skorelowania cech i ich obrazów między sobą jest dla odpowiadających par zbliżona, wtedy na podstawie uzyskanego na płaszczyźnie obrazu struktury danych można wnioskować prawidłowo o skorelowaniu cech. Obrazy cech położone blisko siebie (np. na rys. 2 cechy \hat{X}_2, \hat{X}_4) wskazują na silną korelację dodatnią. Jeśli obrazy cech położone są po przeciwległych stronach na okręgu względem środka koła (np. cechy \hat{X}_2, \hat{X}_3), to są one silnie skorelowane ujemnie. Gdy obrazy cech położone są pod kątem prostym na okręgu względem środka koła (np. cechy \hat{X}_1, \hat{X}_4 na rys. 2), to są one słabo skorelowane. (Autorzy artykułu dysponują autorskim programem odwzorowującym strukturę S w S^* w oparciu o przedstawiony algorytm).

ZASTOSOWANIE METODY GRAFICZNEJ RGM W ANALIZIE PORÓWNAWCZEJ SYSTEMU WARTOŚCI

Charakterystyka psychologicznego problemu badawczego

Zamierzeniem badawczym była diagnoza i charakterystyka porównawcza systemów wartości studentów polskich i rosyjskich. Przyjęto hipotetycznie, że narodowość jest czynnikiem różnicującym preferencje wartości. Dodatkowo analizowanymi zmiennymi była płeć i kierunek studiów. Badaniami objęto 644 osoby, w tym 325 Polaków i 319 Rosjan. W grupie polskiej było 165 studentów pedagogiki i 160 studentów zarządzania (73 mężczyzn i 252 kobiety), w grupie rosyjskiej – 188 studentów pedagogiki/psychologii i 131 studentów zarządzania (63 mężczyzn i 256 kobiet). Zrezygnowano z bliższej charakterystyki grup badawczych, jako że głównym celem pracy jest opis nowej metody graficznej prezentacji struktury danych wielowymiarowych (RGM).

W zamieszczonej poniżej analizie graficznej przyjęto następujące oznaczenia grup: PK1 – polscy studenci pedagogiki, płci żeńskiej; RK1 – rosyjscy studenci pedagogiki, płci żeńskiej; PM1 – polscy studenci pedagogiki, płci męskiej; RM1 – rosyjscy studenci pedagogiki, płci męskiej; PK2 – polscy studenci zarządzania, płci żeńskiej; RK2 – rosyjscy studenci zarządzania, płci żeńskiej;

PM2 – polscy studenci zarządzania, płci męskiej; RM2 – rosyjscy studenci zarządzania, płci męskiej.

W badaniu wykorzystana została Skala Wartości M. Rokeacha (Brzozowski, 1989; Czerniawska, 1995), w wersji wymagającej rangowania. Wartość najwyżej preferowana uzyskuje rangę 1, wartość najniżej preferowana – rangę 18. W przedstawionej poniżej prezentacji wykorzystane zostały dane dotyczące wartości ostatecznych.

W analizie, w pierwszej kolejności skoncentrowano się na określeniu różnic w preferencji wartości, przy wyłącznym uwzględnieniu kryterium narodowości (studenci polscy i rosyjscy). W tabeli 1 zestawione zostały średnie arytmetyczne rang preferencji wartości ostatecznych w poszczególnych grupach badawczych.

Tabela 1.
Wskaźniki preferencji wartości ostatecznych (średnie rang) z uwzględnieniem czynników: narodowość, płeć, kierunek kształcenia

Wartości ostateczne	Polacy	Rosjanie	z	Polacy				Rosjanie			
				kobiety		mężczyźni		kobiety		mężczyźni	
	N= 325	N= 319		N=147	N=105	N=18	N=55	N=167	N=89	N=21	N=42
				PK1	PK2	PM1	PM2	RK1	RK2	RM1	RM2
wo1	10,72	13,34	6,62***	11,61	9,88	11,22	9,78	13,72	12,96	13,43	12,54
wo2	3,41	4,49	4,29***	3,61	2,88	4,28	3,58	4,93	3,53	6,10	3,93
wo3	6,38	4,29	6,53***	5,79	7,02	5,50	7,00	4,28	3,88	4,05	5,33
wo4	11,24	8,29	7,61***	11,97	10,69	10,56	10,55	9,29	6,58	10,33	6,93
wo5	6,26	8,78	7,58***	6,47	6,27	5,00	6,11	8,46	10,40	5,86	8,09
wo6	11,57	11,74	1,01	11,86	11,85	9,94	10,76	11,66	11,86	11,33	11,98
wo7	7,15	7,50	1,27	6,88	6,83	8,56	8,04	7,04	7,75	8,29	8,38
wo8	9,37	11,03	3,71***	9,35	9,41	10,28	9,04	10,85	11,44	10,48	11,19
wo9	7,75	6,72	3,61***	7,52	7,58	7,89	8,62	6,95	6,48	7,62	5,88
wo10	13,27	13,65	1,63	13,00	14,02	12,94	12,67	13,95	13,52	14,52	12,29
wo11	7,51	5,21	7,34***	6,52	8,28	7,11	8,82	3,96	6,09	6,95	7,43
wo12	10,55	13,97	10,47***	10,78	10,67	11,33	9,47	14,13	14,26	12,81	13,29
wo13	7,20	6,34	2,78**	7,01	7,31	7,56	7,38	6,39	5,89	5,62	7,45
wo14	14,05	12,64	4,99***	12,98	15,58	12,56	14,45	12,47	13,28	11,29	12,69
wo15	13,10	11,57	5,04***	13,92	12,16	14,11	12,36	11,88	11,15	12,43	10,81
wo16	7,47	7,40	0,15	7,54	7,62	7,11	7,11	7,10	8,26	6,71	7,12
wo17	9,39	14,54	10,38***	9,53	8,29	10,33	10,78	14,36	14,75	13,76	15,17
wo18	14,50	9,34	14,21***	14,48	14,64	14,72	14,24	9,29	8,93	9,10	10,50

z = statystyka testu Wilcozona dla dwóch prób niezależnych (porównanie grupy polskiej i rosyjskiej); **- $p < 0,01$; ***- $p < 0,001$.

Wyniki analizy statystycznej (test Wilcoxon) wykazały, że studenci rosyjscy cenią w wyższym stopniu wartości hedonistyczne: „dostatnie życie” (wo4), „szczęście” (wo13), „życie pełne wrażeń” (wo18), „świat piękna” (wo14) oraz wartości związane z dobrą kondycją psychiczną: „równowaga wewnętrzna” (wo11), „uznanie społeczne” (wo15). Studenci polscy cechują się silniejszą akceptacją wartości związanych z bezpieczeństwem: „bezpieczeństwo narodowe” (wo1), „bezpieczeństwo rodziny” (wo2), „pokój na świecie” (wo8) oraz wartości: „mądrość” (wo5), „równość” (wo12) i „zbawienie” (wo17). W przypadku wartości „przyjemność” (wo10), „poczucie własnej godności” (wo7), „poczucie dokonania” (wo6), „wolność” (wo16) nie zaobserwowano istotnych statystycznie różnic w preferencji.

Przedstawiony sposób analizy umożliwia zdiagnozowanie różnic w preferowaniu poszczególnych wartości, ale nie pozwala na opis ich strukturalnego układu. Ponadto charakterystyka porównawcza systemów wartości wydaje się problemem jeszcze bardziej złożonym, gdy uwzględnia się większą liczbę czynników warunkujących preferencję, np. narodowość, płeć, kierunek kształcenia. Porównanie międzygrupowe wskaźników akceptacji każdej z osiemnastu wartości w ośmiu grupach badawczych dostarcza wielu szczegółowych informacji, które jednak trudno ująć w bardziej ogólne prawidłowości. Zdaniem autorów, zaprezentowana powyżej metoda RGM stanowi znaczne ułatwienie w analizie porównawczej systemów wartości. Umożliwia ona dokonanie całościowego „oglądu” systemu wartości, jak również ujęcie preferencji wartości w bardziej ogólne dymensje. Pozwala obserwować relacje między wyodrębnionymi grupami studentów (z uwzględnieniem czynników: narodowość, płeć, kierunek kształcenia), wpływ wskaźników preferencji wartości na te relacje oraz relacje między wartościami. Uzyskany obraz struktury danych może stanowić podstawę do doprecyzowania szczegółowych hipotez, podlegających dalszej weryfikacji statystycznej. Wyniki analizy statystycznej mogą być następnie konfrontowane z informacjami zawartymi w prezentacji graficznej.

Stosując metodę graficzną RGM, należy przyjąć, że:

- wartości (cechy) zostały odwzorowane na ich obrazy na okręgu;
- grupy badawcze wydzielone ze względu na narodowość, płeć i kierunek kształcenia (obiekty) zostały odwzorowane na ich obrazy wewnątrz okręgu.

Dokonując interpretacji zaprezentowanych graficznie zależności, należy mieć na względzie, że:

- im mniejszy dystans (odległość) między obrazami cech/wartości na okręgu, tym bardziej zbliżoną pozycję zajmują wartości w systemie (tym silniejsza korelacja dodatnia między wskaźnikami preferencji wartości);
- cechy/wartości, których obrazy położone są po przeciwległej stronie względem środka okręgu, zajmują skrajne pozycje w hierarchii;
- im mniejszy dystans między obrazami obiektów/grup, tym bardziej podobne są grupy badawcze pod względem analizowanych cech/wartości;
- im mniejszy dystans między obrazem cechy/wartości a obrazem obiektu/grupy, tym wyższa preferencja wartości w danej grupie badawczej.

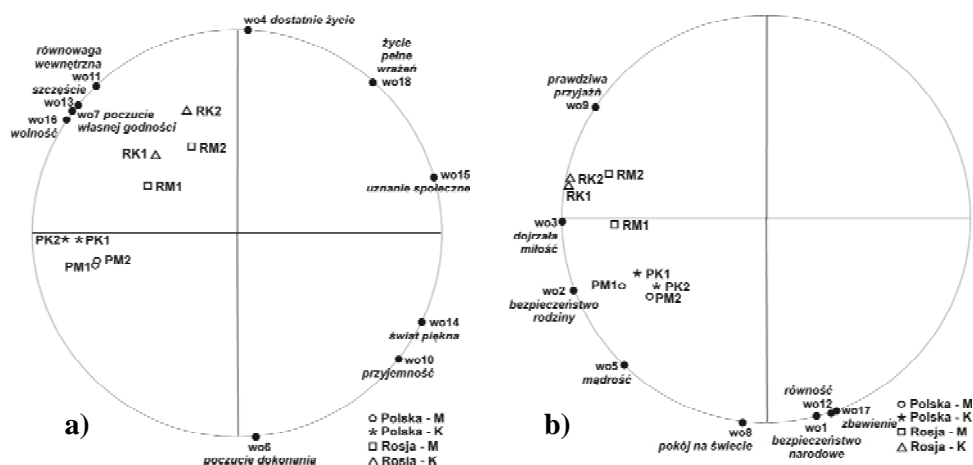
W zależności od zastosowanego sposobu normalizacji uzyskuje się obraz struktury umożliwiający ocenę:

a) **Bezwzględnej pozycji wartości w systemie** (gdy normalizuje się względem 9,5). Za pomocą takiego odwzorowania można więc stwierdzić, jaką pozycję w systemie zajmują poszczególne wartości (opisać w kategoriach preferencji „wysokich”-„niskich”) w wyodrębnionych grupach badawczych (por. rys. 4a, 4b).

b) **Względnej (relatywnej) pozycji wartości w systemie** (gdy normalizuje się względem średniej arytmetycznej rang). W tym przypadku możliwa jest obserwacja preferencji wartości w sposób relatywny (tj. taki, który umożliwia dokonanie opisu w kategoriach „wyżej”-„niżej”, rys. 5). Innymi słowy, dystans między obrazami grup badawczych (obiektów) wskazuje na relatywne zróżnicowanie preferencji wartości, a nie na ich pozycję w systemie. Uzyskany w ten sposób obraz struktury eksponuje zróżnicowanie preferencji wartości w poszczególnych grupach.

Bezwzględne zróżnicowanie pozycji wartości w systemie

Na rysunkach 4a i 4b zaprezentowano strukturę grup i opisujących je wartości ostatecznych. Wartości podzielono na dwie podgrupy. W pierwszej z nich znalazły się przede wszystkim wartości hedonistyczne, skoncentrowane na „ja”, w drugiej – wartości związane z bezpieczeństwem i pozytywnymi relacjami między ludźmi.



Rysunek 4. Struktura grup oraz opisujących je wartości: a) hedonistycznych, skoncentrowanych na „ja”, b) związanych z bezpieczeństwem i relacjami interpersonalnymi

Interpretując zależności na rysunku 4a należy uwzględnić, że dwie wartości, tj. „przyjemność” (wo10) ($r_{zz} = 0,14$) oraz „wolność” (wo16) ($r_{zz} = 0,12$), są źle „wpasowane” w całą strukturę. Miarą jakości wpasowania jest współczynnik korelacji r_{zz} między odległościami cecha-obiekt na płaszczyźnie a znormalizowanymi wartościami cech (por. punkt B. „Ocena zgodności struktury obiekt-cecha”). Oznacza to, że powinno się ostrożnie interpretować wpływ tych wartości na strukturę grup i wnioskować odnośnie do ich preferencji.

Dokonując analizy zależności w wymiarze „obiekt–obiekt” (usytuowanie grup badawczych względem siebie) należy zauważyć, że najsilniejsze zróżnicowanie uwidacznia się przy uwzględnieniu kryterium narodowościowego. Grupy jednolite narodowościowo tworzą dwa wyraźne skupienia: „studenci polscy” i „studenci rosyjscy”. Rozdzielność skupień obiektów zobrazowana jest dużym dystansem (odległością) w usytuowaniu względem siebie grup studentów polskich (PK1, PM1, PK2, PM2) i rosyjskich (RK1, RM1, RK2, RM2). Z interpretacyjnego punktu widzenia oznacza to, że czynnikiem najsilniej różnicującym analizowane wartości była narodowość. Zróżnicowania wewnątrz grupy polskiej i rosyjskiej, oparte na kryterium płci i kierunku kształcenia, są już relatywnie słabsze. Należy jednocześnie stwierdzić większe podobieństwo w zakresie preferencji wartości między czterema grupami polskimi (między symbolami grup badawczych jest niewielki dystans) niż między czterema grupami rosyjskimi (między symbolami grup badawczych jest większy dystans). Analiza dystansu między obiektami/grupami pozwala więc opisać „podobieństwo” vs „zróżnicowanie” w zakresie preferencji wszystkich wartości „wpasowanych” w okrąg.

W dalszej części analizy (rys. 4a) uwzględniono zależności w wymiarze „cecha–cecha” („wartość–wartość”). Obserwuje się między innymi podobną pozycję w systemie następujących wartości: „równowaga wewnętrzna” (wo11), „szczęście” (wo13), „poczucie własnej godności” (wo7), „wolność” (wo16) (niewielki dystans między punktami zaczepienia obrazów cech, por. także średnie rang w tab. 1). Jednocześnie, dokonując analizy relacji „cecha–obiekt” należy zauważyć, że wartości te są wysoko preferowane we wszystkich grupach badawczych, ale wyżej przez studentów rosyjskich (szczególnie kobiety – RK1, RK2), w porównaniu ze studentami polskimi (mniejszy dystans pomiędzy cechami i obiektami w przypadku grup rosyjskich). Wartość „wolność” (wo16) – jak wskazano wyżej – jest źle „wpasowana” w całą strukturę, co ogranicza możliwość trafnej interpretacji.

Wartościami zdecydowanie wyżej preferowanymi przez grupy rosyjskie (szczególnie studentów zarządzania, RK2, RM2) były również: „dostatnie życie” (wo4), „życie pełne wrażeń” (wo18), „uznanie społeczne” (wo15) (por. dystans między cechami a obiektami). Wartości te zajmowały na ogół niższe pozycje w systemie w porównaniu z opisanymi wcześniej (por. różnicowanie dystansu między obiektami a obiema grupami analizowanych cech). Jeszcze niżej (przez wszystkie grupy) preferowane były wartości: „świat piękna” (wo14), „przyjemność” (wo10) (wartość „przyjemność” jest źle „wpasowana”

w całą strukturę), „poczucie dokonania” (wo6). Opisane zależności są zgodne ze wskaźnikami średnich rang i z wynikami analizy statystycznej (por. tab. 1).

W analogiczny sposób można dokonać interpretacji zależności zobrazowanych na rysunku 4b. W tym przypadku widoczne są również dwa skupienia obiektów (PK1, PM1, PK2, PM2 oraz RK1, RM1, RK2, RM2), które są efektem zróżnicowania preferencji wartości w oparciu o kryterium narodowości. Wartościami wysoko preferowanymi były: „bezpieczeństwo rodziny” (wo2) i „dojrzała miłość” (wo3), przy czym pierwszej z nich przypisywała wyższe znaczenie grupa studentów polskich, drugiej – grupa studentów rosyjskich (por. dystans między cechami i obiektami oraz średnie rang i wyniki analizy statystycznej w tab. 1). Niższe pozycje w systemie wartości Polaków, w porównaniu z „bezpieczeństwem rodziny”, zajęły wartości: „mądrość” (wo5), „pokój na świecie” (wo8), „bezpieczeństwo narodowe” (wo1), „równość” (wo12), „zbawienie” (wo17). Każda z tych wartości była jednak wyżej ceniona wśród studentów polskich niż rosyjskich. W grupie Rosjan ujawniła się silniejsza akceptacja wartości „prawdziwa przyjaźń” (wo9).

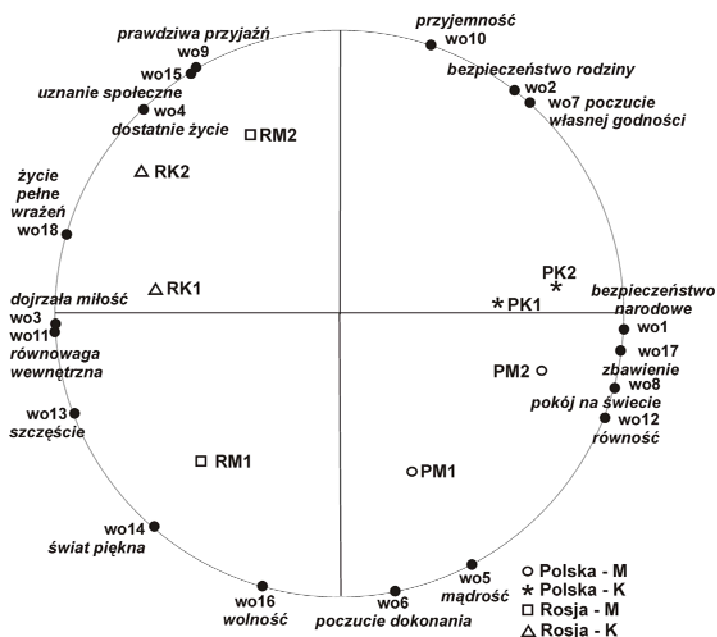
Relatywne zróżnicowanie pozycji wartości w systemie

W dalszej części analizy systemów wartości zastosowano wariant metody RGM, który pozwala na obserwację relatywnego zróżnicowania preferencji wartości. Dzięki takiemu zabiegowi można opisać zależności z uwzględnieniem zarówno najsilniejszego kryterium różnicującego (tj. narodowości), jak i kryteriów słabszych (tj. płci i kierunku kształcenia). Łatwiejsze staje się uchwycenie relacji pomiędzy „obiettami”, „cechami” oraz „obiettami–cechami”. Należy podkreślić, że stwierdzenie „preferowane wyżej” nie oznacza w tym przypadku „preferowane wysoko”. Innymi słowy, opisywane relacje nie wskazują na „bezwzględne” pozycje wartości w systemie (opis preferencji wartości w kategoriach „wysoko”–„nisko” został przeprowadzony na podstawie analizy rys. 4a i 4b).

Na rysunku 5 zaprezentowano odwzorowanie na okręgu 18 cech/wartości. Najważniejszym czynnikiem różnicującym preferencje była narodowość (widoczne są wyraźne skupienia obiektów PK1, PM1, PK2, PM2 oraz RK1, RM1, RK2, RM2). Wartości „bezpieczeństwo narodowe” (wo1), „pokój na świecie” (wo8), „zbawienie” (wo17) oraz „równość” (wo12) preferowane są wyżej przez studentów polskich niż rosyjskich (por. skupienie cech i dystans między cechami i obiektami).

Wartościami preferowanymi wyżej przez Rosjan są: „życie pełne wrażeń” (wo18), „równowaga wewnętrzna” (wo11), „szczęście” (wo13), „świat piękna” (wo14), „dostatnie życie” (wo4), „uznanie społeczne” (wo15) (por. skupienie cech). Uwidacznia się również zróżnicowanie wewnątrz grupy rosyjskiej: dwie wymienione na końcu wartości cechują się wyższą preferencją wśród rosyjskich studentów zarządzania (RK2, RM2), wartość „świat piękna” (wo14) ceniona jest natomiast wyżej przez studentów pedagogiki płci męskiej (RM1). Studenci rosyjscy charakteryzują się też wyższymi wskaźnikami wartości:

„dojrzała miłość” (wo3) (zwłaszcza w grupach RK1, RK2, RM1) i „prawdziwa przyjaźń” (wo9) (zwłaszcza wśród studentów zarządzania RK2, RM2).



Rysunek 5. Struktura grup oraz opisujących je wartości (uwzględniono 18 wartości)

Omawiając preferencje wartości: „przyjemność” (wo10), „poczucie własnej godności” (wo7) i „bezpieczeństwo rodziny” (wo2), należy zauważyć ich zróżnicowanie przy uwzględnieniu kryterium płci i kierunku kształcenia. Zarówno w grupie rosyjskiej, jak i polskiej najniższe wskaźniki akceptacji uzyskali z reguły mężczyźni, studenci pedagogiki (RM1, PM1). Należy jednocześnie podkreślić, że wartości „przyjemność” (wo10) i „poczucie własnej godności” (wo7) nie różniły się preferencją w grupie polskiej i rosyjskiej. Wartość „bezpieczeństwo rodziny” była silniej akceptowana przez Polaków (por. tab. 1, test Wilcoxona).

Wyraźnym zróżnicowaniem preferencji wewnątrz grup narodowościowych cechowały się wartości: „poczucie dokonania” (wo6), „wolność” (wo16) i „mądrość” (wo5). Każda z tych wartości jest silniej akceptowana przez mężczyzn – zwłaszcza studentów pedagogiki, zarówno polskich, jak i rosyjskich (RM1, PM1). Najniższe wskaźniki w tym zakresie uzyskali rosyjscy studenci zarządzania (RK2, RM2). Wartość „mądrość” zajmuje wyższe pozycje w systemie Polaków (relatywnie mniejszy dystans między cechą a obiektami).

WNIOSKI

Analiza porównawcza systemów wartości, przeprowadzona za pomocą metody graficznej RGM, pozwala na sformułowanie następujących wniosków:

(1) Narodowość jest czynnikiem najsilniej różnicującym systemy wartości. Studenci polscy cenili w wyższym stopniu wartości religijne, wartości związane z szeroko rozumianym bezpieczeństwem i odnoszące się do relacji między ludźmi. Studenci rosyjscy wyżej lokowali w systemie wartości hedonistyczne, skoncentrowane na własnej osobie.

(2) Ujawnia się wpływ narodowości, płci i kierunku kształcenia na preferencje wartości, niemniej jednak znaczenie dwóch ostatnich czynników było relatywnie mniejsze. Ponadto płeć i kierunek kształcenia nie różnicowały w analogiczny sposób preferencji wartości w obu grupach narodowościowych. W przypadku niektórych wartości istotniejszym od narodowości kryterium porównań okazała się płeć i kierunek kształcenia.

Podkreślając walory metody graficznej prezentacji struktury danych wielowymiarowych (RGM) w analizie porównawczej systemów wartości należy zauważyć, że pozwala ona na: (1) Całościowy ogłąd systemów wartości: określenie relacji występujących między wartościami (ich wzajemne usytuowanie w systemie), podobieństwo vs zróżnicowanie w wymiarze aksjologicznym wielu grup badawczych, zróżnicowanie w preferencji wartości w poszczególnych grupach badawczych; (2) Opisanie wartości na różnych dymensjach, np. indywidualizm vs kolektywizm. Wartości mogą być dowolnie selekcionowane do prezentacji i przedstawiane na wielu okręgach (na płaszczyźnie mogą być odwzorowane wybrane wartości, takie, które są ważne z merytorycznego punktu widzenia, np. wartości związane z otwartością na zmiany czy konserwatyżmem lub też obie kategorie równocześnie). Interpretacja może być dokonana na różnych poziomach szczegółowości; (3) Dokonanie opisu zależności warunkowanych wieloaspektowo.

Opracowana metoda może być także podstawą do formułowania bardziej szczegółowych hipotez oraz wyboru kategorii zmiennych, które powinny być poddane dalszej weryfikacji empirycznej; można ją stosować również do analizy danych innych niż systemy wartości.

BIBLIOGRAFIA

- Brzozowski, P. (1989). *Skala Wartości (SW). Polska adaptacja Value Survey M. Rokeacha*. Warszawa: Polskie Towarzystwo Psychologiczne–Wydział Psychologii Uniwersytetu Warszawskiego.
- Czerniawska, M. (1995). *Inteligencja a system wartości*. Białystok: Trans Humana.
- Gabriel, K. R. (1971). The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, 58, 453-467.
- Gordon, A. D. (1999). *Classification*. Boca Raton–London–New York–Washington: Chapman & Hall/CRC.

- Grabiński, T. (1992). *Metody taksonometrii*. Kraków: Akademia Ekonomiczna w Krakowie.
- Jajuga, K. (1993). *Statystyczna analiza wielowymiarowa*. Warszawa: PWN.
- Rybaczuk, M. (2002). Graficzna prezentacja struktury danych wielowymiarowych. [W:] K. Jajuga, M. Walesiak (red.), *Taksonomia*, z. 9: *Klasyfikacja i analiza danych – teoria i zastosowania* (s. 146-153). Wrocław: Wydawnictwo Akademii Ekonomicznej. Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 942.
- Zeliaś, A. (red.) (1989). *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*. Warszawa: PWN.