

Artur Zaborski

Akademia Ekonomiczna we Wrocławiu

IDENTYFIKACJA RÓŻNIC INDYWIDUALNYCH W ANALIZIE PROCRUSTESA

1. Wstęp

Punktem wyjścia skalowania wielowymiarowego, które ma na celu graficzną prezentację relacji zachodzących pomiędzy badanymi obiektami, jest macierz $\Delta = [\delta_{ij}]_{n \times n}$. Jej elementy przedstawiają niepodobieństwa między obiektami i oraz j ($i, j = 1, 2, \dots, n$), gdzie n jest liczbą obiektów. Jeżeli korzystamy z opinii wielu respondentów, mamy do czynienia z sytuacją, kiedy danych jest m macierzy $\Delta_k = [\delta_{ijk}]$, gdzie $k = 1, 2, \dots, m$, przy czym każda macierz Δ_k przedstawia niepodobieństwa między obiektami postrzegane przez k -tego respondenta. Na podstawie macierzy Δ_k można wyznaczyć k konfiguracji punktów X_k reprezentujących obiekty, tworzących trójwymiarową kostkę zmiennych o współrzędnych x_{iak} , gdzie:

$i = 1, 2, \dots, n$ – liczba badanych obiektów,

$a = 1, 2, \dots, p$ – wymiar przestrzeni, w której prezentowane są wyniki skalowania wielowymiarowego,

$k = 1, 2, \dots, m$ – liczba konfiguracji punktów.

Poszczególni respondenci zazwyczaj różnią się w ocenie ważności wpływu określonych zmiennych na podobieństwa między obiektami. Wzrokowe porównanie X_k może być jednak utrudnione, ponieważ różnice w układach konfiguracji punktów na mapach percepcyjnych nie zawsze wynikają z różnic w postrzeganiu obiektów przez poszczególnych respondentów (zob. [Zaborski 2004]). Dlatego porównanie wyników skalowania wielowymiarowego powinno być poprzedzone analizą Procrustesa. Dopuszczalnymi przekształceniami poszczególnych konfiguracji

punktów w analizie Procrustes są translacja, rotacja i jednokładność, czyli te, które zachowują proporcje odległości między punktami.

Dla porównywalnych wyników skalowania wielowymiarowego poszczególnych respondentów, identyfikacja różnic możliwa jest poprzez wyznaczenie wag w_{ka} zmiennej a dla respondenta k , tworzących przestrzeń wag $\mathbf{W} = [w_{ka}]$. Geometrycznie punkty przestrzeni wag przedstawiają respondentów, a współrzędne tych punktów – wagi przypisywane poszczególnym wymiarom.

Dla ustalonej „wspólnej” dla wszystkich respondentów konfiguracji $\mathbf{Z}$ (*centroid configuration*) indywidualne konfiguracje punktów dla danego respondenta $\mathbf{X}_k$ otrzymuje się, mnożąc współrzędne $\mathbf{Z}$ przez odpowiednie wagi. Wtedy:

$$\mathbf{X}_k = \mathbf{Z}\mathbf{W}_k,$$

gdzie $\mathbf{W}_k$ jest diagonalną macierzą wag o wymiarach $p \times p$ k -tego respondenta.

W artykule pokazano, jak, poprzez wyznaczenie przestrzeni wag, możliwe jest wskazanie różnic w ocenach respondentów. Sposób dopasowywania konfiguracji indywidualnych, wyznaczania konfiguracji „wspólnej”, przestrzeni wag oraz miar dopasowania przedstawiono na podstawie modelu PINDIS (*procrustean individual differences scaling*).

Algorytm analizy Procrustes zaproponował J.C. Gower [1975]. Rozwinięcie myśli Gowera oraz przykłady zastosowań są przedstawione m.in. w pracach: [Ten Berge 1977; Lingoes, Borg 1987; Digby, Kempton 1978; Peay 1988; Krzanowski 1988; Commandeur 1991; Goodall 1991].

2. Zasady analizy Procrustes dla dwóch wyników skalowania wielowymiarowego

Niech $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n]^T$ i $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_n]^T$ będą macierzami współrzędnych dwóch konfiguracji n punktów w euklidesowej przestrzeni p -wymiarowej, uzyskanych w wyniku skalowania wielowymiarowego. Punkty obydwu konfiguracji reprezentują te same obiekty lub zmienne, co oznacza, że punkt $\mathbf{x}_i$ odpowiada punktowi $\mathbf{y}_i$ dla każdego $i = 1, 2, 3, \dots, n$. Najlepsze dopasowanie konfiguracji uzyskujemy wtedy, gdy suma kwadratów odległości między odpowiadającymi sobie punktami $\mathbf{X}$ i $\mathbf{Y}$ jest najmniejsza, czyli gdy funkcja dopasowania przybierająca postać:

$$R^2 = \text{tr}(\mathbf{X} - \mathbf{Y})^T (\mathbf{X} - \mathbf{Y}) \quad (1)$$

osiąga wartość najmniejszą.

Dopasowanie konfiguracji $\mathbf{X}$ do $\mathbf{Y}$, minimalizujące wartość funkcji dopasowania (1), przebiega przez następujące etapy (zob. [Cox, Cox 2001, s. 128; Zaborski 2004]):

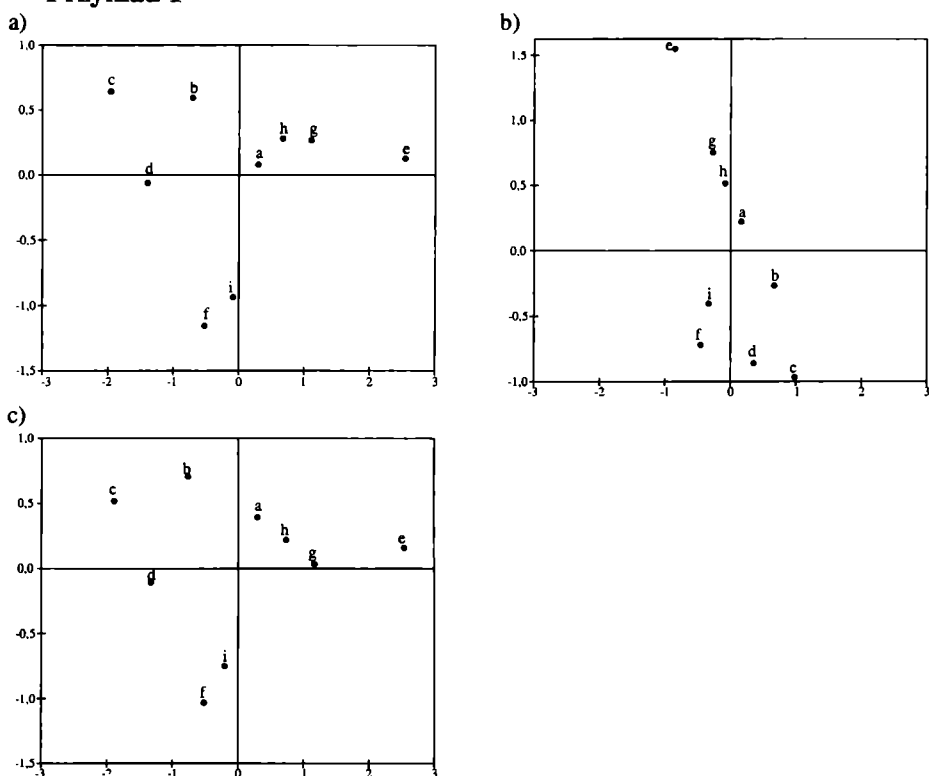
1. Translacja $\mathbf{X}$ i $\mathbf{Y}$ tak, aby ich środki ciężkości znajdowały się w początku układu współrzędnych.

2. Wymnożenie macierzy $\mathbf{X}$ przez macierz rotacji $\mathbf{A} = (\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})^{\frac{1}{2}} (\mathbf{Y}^T \mathbf{X})^{-1}$.
3. Znalezienie obrazu $\mathbf{X}$ przy jednokładności o stosunku $\hat{\phi} = \frac{\text{tr}(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})^{\frac{1}{2}}}{\text{tr}(\mathbf{X}^T \mathbf{X})}$.
4. Obliczenie miary dopasowania $R^2 = 1 - \frac{\{\text{tr}(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})^{\frac{1}{2}}\}^2}{\text{tr}(\mathbf{X}^T \mathbf{X}) \text{tr}(\mathbf{Y}^T \mathbf{Y})}$.

Większość algorytmów skalowania wielowymiarowego dokonuje takiego rozmieszczenia punktów na mapie percepcyjnej, aby środek ciężkości punktów znajdował się w początku układu współrzędnych, a średnia kwadratów odległości punktów od początku układu współrzędnych wynosiła jeden. Dlatego w praktyce etapy pierwszy i trzeci są pomijane, a funkcja dopasowania (1) przybiera postać:

$$R^2 = \text{tr}(\mathbf{Y} - \mathbf{X}\mathbf{A})^T (\mathbf{Y} - \mathbf{X}\mathbf{A}). \quad (2)$$

Przykład 1



Rys. 1. Konfiguracje punktów reprezentujących 9 obiektów otrzymanych: a) z wykorzystaniem modelu ALSCAL; b) z wykorzystaniem MDSCAL-2; c) w wyniku dopasowania konfiguracji (b) do konfiguracji (a) za pomocą analizy Procrustes

Źródło: [Zaborski 2004].

Na podstawie macierzy odległości między dziewięcioma obiektami przeprowadzono skalowanie wielowymiarowe. Rysunek 1a przedstawia konfigurację punktów reprezentujących badane obiekty wyznaczone z zastosowaniem modelu ALSCAL, rys. 1b zaś konfigurację otrzymaną za pomocą modelu MDSCAL-2. Różnice w układach punktów na obydwu mapach percepcyjnych uzasadniały potrzebę przeprowadzenia analizy Procrustes w celu dopasowania obydwu konfiguracji. Efekt dopasowania przy $R^2 = 0,1369$ prezentuje rys. 1c.

3. Dopasowanie konfiguracji indywidualnych i wyznaczanie konfiguracji wspólnej

Dla wielu wyników skalowania wielowymiarowego dokonuje się transformacji indywidualnych konfiguracji $\mathbf{X}_k$ tak, aby każda z nich została jak najlepiej dopasowana do pozostałych. Podobnie jak w analizie dla dwóch wyników skalowania wielowymiarowego, dopuszczalnymi przekształceniami są translacja, rotacja i jednokładność, czyli te, które zachowują proporcje odległości między punktami. Proces dopasowywania wielu konfiguracji indywidualnych oraz wyznaczania konfiguracji wspólnej nazywamy uogólnioną analizą Procrustes.

Uogólniona analiza Procrustes prowadzi do wyznaczenia takich konfiguracji $\tilde{\mathbf{X}}_k$, dla których funkcja dopasowania (por. [Borg, Groenen 1997, s. 362])

$$R^2 = \sum_{k < l}^m \text{tr}(\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l)^T (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l) \quad (3)$$

przyjmuje wartość najmniejszą, przy czym $\tilde{\mathbf{X}}_k = \varphi_k \mathbf{X}_k \mathbf{A}_k^T + \mathbf{1b}_k^T$ ($k, l = 1, 2, \dots, m$).

Algorytm minimalizujący wartość funkcji R^2 ma charakter iteracyjny i polega na cyklicznym dopasowywaniu kolejnych konfiguracji punktów (tak jak w przypadku analizy Procrustes dla dwóch wyników skalowania), przy założeniu, że pozostałe są ustalone.

Konfiguracja wspólna (*centroid configuration*) jest średnią ze wszystkich $\tilde{\mathbf{X}}_k$ i przybiera postać $\mathbf{Z} = \frac{1}{m} \sum_{k=1}^m \tilde{\mathbf{X}}_k$. Wtedy funkcję (3) można zapisać jako:

$$R^2 = m \sum_{k=1}^m \text{tr}(\tilde{\mathbf{X}}_k - \mathbf{Z})^T (\tilde{\mathbf{X}}_k - \mathbf{Z}). \quad (4)$$

Równanie (4) wskazuje, że uogólniona analiza Procrustes może być realizowana poprzez kolejne dopasowywanie konfiguracji $\tilde{\mathbf{X}}_k$ ($k = 1, 2, \dots, m$) do uśrednionej konfiguracji $\mathbf{Z}$, przy jednoczesnym uaktualnianiu $\mathbf{Z}$ w kolejnych cyklach iteracyjnych. Proces dopasowywania $\tilde{\mathbf{X}}_k$ jest taki sam jak dla dwóch wyników skalowania wielowymiarowego.

Wzajemne dopasowywanie konfiguracji indywidualnych, wyznaczanie konfiguracji wspólnej i przestrzeni wag w modelu PINDIS przebiega w dwóch etapach (por. [Lingoes, Borg 1978]). Na etap pierwszy składają się:

a) translacja i wyskalowanie $\mathbf{X}_k$ ($k=1,2,\dots,m$) tak, aby ich środki ciężkości znajdowały się w początku układu współrzędnych, a średni kwadrat odległości punktów od początku układu współrzędnych był równy jeden, tzn. $tr(\mathbf{X}_k^T \mathbf{X}_k) = 1$;

b) szacowanie konfiguracji $\mathbf{Z}$. W tym celu dopasowuje się (wykorzystując analizę Procrustes dla dwóch wyników skalowania wielowymiarowego) $\tilde{\mathbf{X}}_2$ do $\mathbf{X}_1$, a

następnie wyznacza macierz $\mathbf{Z}$ jako $\frac{1}{2}(\mathbf{X}_1 + \tilde{\mathbf{X}}_2)$. Po dopasowaniu $\mathbf{X}_3$ do oszacowanej wcześniej $\mathbf{Z}$ obliczana jest nowa wartość $\mathbf{Z} = \frac{1}{3}(\mathbf{X}_1 + \tilde{\mathbf{X}}_2 + \tilde{\mathbf{X}}_3)$. Proces ten powtarzany jest tak długo, aż wykorzystane zostaną wszystkie konfiguracje $\mathbf{X}_k$.

W etapie drugim, po ponownym dopasowaniu wszystkich konfiguracji $\tilde{\mathbf{X}}_k$ do oszacowanej $\mathbf{Z}$, oblicza się miarę dopasowania:

$$h = \frac{1}{m} \sum_{k=1}^m \left(1 - R_k^2(\tilde{\mathbf{X}}_k, \mathbf{Z})\right)^2, \quad (5)$$

gdzie $R_k^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$ jest miarą Procrustes, gdy $\tilde{\mathbf{X}}_k$ jest dopasowywane do $\mathbf{Z}$, czyli:

$$R_k^2(\tilde{\mathbf{X}}_k, \mathbf{Z}) = \frac{\{tr(\tilde{\mathbf{X}}_k^T \mathbf{Z} \mathbf{Z}^T \tilde{\mathbf{X}}_k)\}^{\frac{1}{2}}}{tr(\tilde{\mathbf{X}}_k^T \tilde{\mathbf{X}}_k) tr(\mathbf{Z}^T \mathbf{Z})}. \quad (6)$$

Po wyznaczeniu nowej wartości $\mathbf{Z}$, jako średniej z otrzymanych w poprzednim etapie $\tilde{\mathbf{X}}_k$, wszystkie $\tilde{\mathbf{X}}_k$ ponownie dopasowywane są kolejno do $\mathbf{Z}$, a następnie oblicza się miarę dopasowania h . Proces powtarzany jest tak długo, aż różnica pomiędzy h wyznaczonymi w kolejnych cyklach iteracyjnych przestanie ulegać zmianie lub będzie mniejsza od ustalonej stałej.

4. Przestrzeń wag

Respondenci zazwyczaj różnie oceniają ważności wpływu określonych zmiennych na podobieństwa między obiektami. Jeżeli zatem przyjmiemy założenie, że przypisują oni różne znaczenie poszczególnym wymiarom konfiguracji wspólnej, to istnieją takie macierze $\mathbf{W}_k$, dla których $\tilde{\mathbf{X}} \approx \mathbf{Z} \mathbf{W}_k$ ($\mathbf{W}_k$ jest diagonalną macierzą o wymiarach $p \times p$ niezerowych wag poszczególnych wymiarów, $\mathbf{Z} \mathbf{W}_k$ zaś ważoną konfiguracją wspólną k -tego respondenta). Dla ustalonych konfiguracji $\mathbf{Z}$ i $\mathbf{X}_k$ należy więc znaleźć taką macierz $\mathbf{W}_k$, aby funkcja dopasowania:

$$R^2 = (\mathbf{Z}\mathbf{W}_k - \tilde{\mathbf{X}}_k)^T (\mathbf{Z}\mathbf{W}_k - \tilde{\mathbf{X}}_k) \quad (7)$$

przyjmowała wartość najmniejszą.

Jeżeli $\mathbf{z}_a$ i $\mathbf{x}_{ka}$ są odpowiednio kolumnami macierzy $\mathbf{Z}$ i $\mathbf{X}_k$, a z_{ia} i x_{iak} i -tymi elementami tych wektorów, wtedy znalezienie w_{ka} , który jest a -tym elementem diagonalnym $\mathbf{W}_k$, sprowadza się do minimalizacji wyrażenia:

$$\sum_{i=1}^n (w_{ka} z_{ia} - x_{iak})^2. \quad (8)$$

Po zróżniczkowaniu (8) i przyrównaniu do zera otrzymujemy:

$$w_{ka} = \frac{\sum_{i=1}^n x_{iak} z_{ia}}{\sum_{i=1}^n z_{ia}^2}. \quad (9)$$

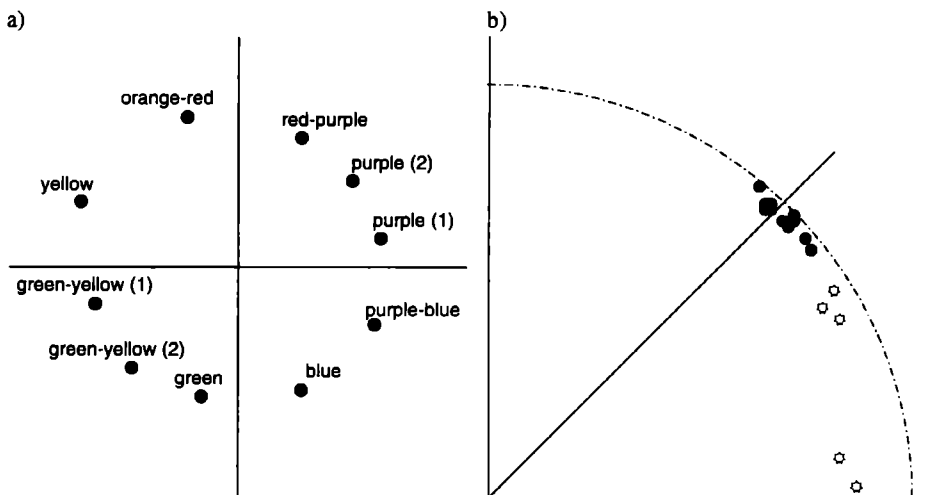
Wyznaczone w ten sam sposób dla każdego $a=1,2,\dots,p$ i dla każdego $k=1,2,\dots,m$ wagi tworzą przestrzeń wag $\mathbf{W}=[w_{ka}]$. Punkty przestrzeni wag (wiersze) przedstawiają respondentów, a współrzędne tych punktów – wagi przypisywane przez nich poszczególnym wymiarom.

5. Przykład wykorzystania modelu PINDIS do identyfikacji różnic indywidualnych¹

14 respondentów, w tym 4 z wadą wzroku (2 respondentów powtórzyło eksperyment po upływie czterech tygodni), poproszono o określenie podobieństwa między 10 kolorami. W tym celu każdemu z nich przedstawiono 120 trójek różnych kolorów z prośbą, aby w pierwszej kolejności dla każdej trójki wskazali dwa kolory ich zdaniem najbardziej różniące się, a następnie umieścili trzeci kolor w takim miejscu na płaszczyźnie, żeby odległości między kolorami odpowiadały ocenom podobieństwa. Uśrednione oceny każdego respondenta utworzyły 16 indywidualnych macierzy niepodobieństw między kolorami (zob. [Borg, Groenen 1997, s 360]). Na ich podstawie przeprowadzono klasyczne skalowanie wielowymiarowe, a dla wyników skalowania – analizę Procrustes z wykorzystaniem modelu PINDIS.

Wartości miar dopasowania zawarte w tab. 1 wskazują na bardzo dobre dopasowanie konfiguracji indywidualnych $\tilde{\mathbf{X}}_k$ respondentów bez wady wzroku do konfiguracji wspólnej $\mathbf{Z}$ (rys. 2a). Świadczy o tym również rozkład odpowiednich punktów w przestrzeni wag. Dla tej grupy badanych znajdują się one w bliskim sąsiedztwie dwusiecznej kąta. Oznacza to, że ci respondenci przypisują obydwu wymiarom $\mathbf{Z}$ podobne znaczenie. Zdecydowanie gorsze dopasowanie (zwłaszcza

¹ Zob. [Lingoes, Borg 1978].



Rys. 2. a) konfiguracja wspólna dla respondentów bez wady wzroku; b) przestrzeń wag respondentów bez wady wzroku (●) i z wadą wzroku (☆)

Źródło: opracowano na podstawie [Lingoes, Borg 1978].

Tabela 1. Miary dopasowania konfiguracji dla prostej i ważonej analizy Procrustes

Respondenci bez wady wzroku	$r^2(\bar{X}_k, Z)$	$r^2(\bar{X}_k, ZW_k)$	Respondenci z wadą wzroku	$r^2(\bar{X}_k, Z)$	$r^2(\bar{X}_k, ZW_k)$
k_1	0,98	0,98	k_{13}	0,89	0,92
k_2	0,98	0,98	k_{14}	0,92	0,92
k_3	0,99	0,99	k_{15}	0,94	0,97
k_4	0,99	0,99	k_{16}	0,50	0,75
k_5	0,99	0,99	k_{17}	0,44	0,82
k_6	0,97	0,97			
k_7	0,98	0,99			
k_8	0,98	0,98			
k_9	0,97	0,97			
k_{10}	0,98	0,98			
k_{11}	0,96	0,97			

Źródło: opracowano na podstawie [Lingoes, Borg 1978].

dla nieważonej analizy Procrustes) uzyskano dla respondentów z wadą wzroku. Rozkład punktów przestrzeni wag wskazuje, że większe znaczenie przypisują oni wymiarowi pierwszemu, a dwóch z nich wręcz nie rozróżnia kolorów na osi „czerwony-zielony”.

6. Podsumowanie

Niezwykle ważnym elementem skalowania wielowymiarowego jest interpretacja wyników polegająca zarówno na określeniu treści wymiarów uzyskanej prze-

strzeni, jak i na identyfikacji zależności występujących między badanymi obiektami. Ponieważ takie przekształcenia jak translacja, jednokładność czy rotacja zbioru punktów nie powodują zmiany proporcji odległości między nimi, dlatego dla danego zbioru obiektów można otrzymać (pozornie) różne konfiguracje końcowe. Zaprezentowana analiza Procrustes pozwala na wzajemne dopasowanie indywidualnych dla poszczególnych respondentów konfiguracji punktów oraz wyznaczenie tzw. konfiguracji wspólnej (średniej). Z kolei wyznaczenie przestrzeni wag oraz miar dopasowania konfiguracji indywidualnych do konfiguracji wspólnej umożliwia wskazanie różnic w ocenach respondentów.

Literatura

- Borg I., Groenen P. (1997), *Modern Multidimensional Scaling. Theory and Applications*, Springer-Verlag, New York.
- Commandeur J.J.F. (1991), *Matching Configurations*, DSWO Press, Leiden.
- Cox T.F., Cox M.A.A. (2001), *Multidimensional Scaling. Second Edition*, Chapman and Hall, London.
- Digby P.G.N., Kempton R.A. (1987), *Multivariate Analysis of Ecological Communities*, Chapman and Hall, London.
- Goodal C. (1991), *Procrustes Methods in the Statistical Analysis of Shape*, „Journal of the Royal Statistical Society” vol. 53, nr 2, s. 285-339.
- Gower J.C. (1975), *Generalized Procrustes Analysis*, „Psychometrika” nr 40, s. 33-51.
- Krzanowski W.J. (1988), *Principles of Multivariate Analysis: a User's Perspective*, Clarendon, Oxford.
- Lingoes J.C., Borg I. (1978), *A Direct Approach to Individual Differences Scaling Using Increasingly Complex Transformations*, „Psychometrika” nr 4, s. 491-519.
- Peay E.R. (1988), *Multidimensional Rotation and Scaling of Configurations to Optimal Agreement*, „Psychometrika” nr 53, s. 199-208.
- Ten Berge J.M.F. (1977), *Orthogonal Procrustes Rotation for Two or More Matrices*, „Psychometrika” nr 42, s. 267-276.
- Zaborski A. (2004), *Porównywanie wyników skalowania wielowymiarowego metodą Procrustes*, *Ekonometria* 14, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1021, s. 51-59.

IDENTIFICATION OF INDIVIDUAL DIFFERENCES IN PROCRUSTES ANALYSIS

Summary

The paper presents Procrustes analysis in which several given coordinate matrices are fit to each other in such a way that uninformative differences are eliminated and an optimal average configuration (group space) is formed. In Procrustes analysis individuals are represented as points in the space of weights. The dimensional weights may be interpreted as expressions of the different importance that individual k attaches to the dimensions of the group space.