STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 495–506, DOI 10.21307/stattrans-2018-027

# DISCRIMINANT COORDINATES ANALYSIS IN THE CASE OF MULTIVARIATE REPEATED MEASURES DATA

**Mirosław Krzyśko**[1], **Wojciech Łukaszonek**[2], **Waldemar Wołyński**[3]

## ABSTRACT

The main aim of the paper is to adapt the classical discriminant coordinates analysis to multivariate repeated measures data. The proposed solution is based on the relationship between the discriminant coordinates and canonical variables. The quality of these new discriminant coordinates is examined on some real data.

**Key words:** discriminant coordinates analysis, repeated measures data (doubly multivariate data), Kronecker product covariance structure, maximum likelihood estimates.

## 1. Introduction

Let us consider a case where samples originate from $K$ groups (classes). We would often like to present them graphically, to see their configuration. However, it may be difficult to produce such a presentation even only three variables are observed. A different method must therefore be sought for presenting multidimensional data originating from multiple groups. That is the role of the discriminant coordinates (Seber (1984), p. 269). They are also sometimes called canonical variates (Krzanowski (2000), p. 370; Srivastava (2002), p. 257), but this name is misleading, because canonical variates with completely different properties occur in canonical correlation analysis. Another name used is discriminant functions (Rencher (1998), p. 202; Fujikoshi et al. (2010), p. 255) - this is inappropriate because discriminant functions are surfaces that separate $K$ groups from one another.

The aim of the classical discriminant coordinates technique is to replace the input variables by a smaller number of independent coordinates in such a way that the separation among groups (classes) is maximum in the reduced space. In the case of two classes we obtain only one discriminant coordinate, coinciding with the well-known Fisher's linear discriminant function (Fisher (1936)). Generalization on $K > 2$ classes was shown by Rao (1948). The space of discriminant coordinates

[1]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. Faculty of Management, President Stanisław Wojciechowski Higher Vocational State School, Poland. E-mail: mkrzysko@amu.edu.pl

[2]Faculty of Management, President Stanisław Wojciechowski Higher Vocational State School, Poland. E-mail: w.lukaszonek@g.pl

[3]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: wolynski@amu.edu.pl

is a space convenient for the use of various classification methods (methods of discriminant analysis).

In the present paper, we adapt the classical discriminant coordinates analysis to multivariate repeated measures data. Suppose that we have a sample of $n$ objects characterized by $p$-variables measured in $T$ different time points or physical conditions. Such data are referred to in the statistical literature as multivariate repeated data or doubly multivariate data. Analysis of such data is complicated by the existence of correlation among the measurements of different variables as well as correlation among measurements taken at different time points.

The proposed methods are particularly useful when the number of time points is small and the data sets are also small. Such situations concern, from example, the observation of variables in groups of territorial units, the number of which is fixed.

In the case of a large number of variables, a large number of time points and large data sets (e.g. in modern on-line economy), the alternative may be to use discriminant coordinates for functional data (see e.g. Górecki et al. (2018)).

In practice, the use of classical discriminant coordinates described in Section 2 requires the fulfillment of the condition $\max\{n_1,\dots,n_K\} > pT$, where $n_i$ is the sample size derived from the $i$th group, $i = 1,\dots,K$. This condition is very restrictive and requires large samples. If it is not satisfied, then our problem can be partially solved using an existing relationship between the discriminant coordinates and canonical variables (Krzyśko (1979)). The construction of the discriminant coordinates as the canonical variables of the $\boldsymbol{X}$-space is described in Section 3. In this case the condition $n > pT + K - 1$ is required, where $n = n_1 + \cdots + n_K$. Note that the condition $n > pT + K - 1$ is a condition much weaker than the condition $\max\{n_1,\dots,n_K\} > pT$, especially for a small number of groups $K$. If $n \leq pT + K - 1$, then we can construct the discriminant coordinates with the additional condition imposed on the covariance matrix. This construction is presented in Section 4. Section 5 illustrates the approaches presented in the paper on a real data set.

## 2. Classical discriminant coordinates: a review

Let us consider the multivariate discriminant problem with $K$ groups. We observe $(\boldsymbol{X},Y)$, $\boldsymbol{X} = (\boldsymbol{X}_1,\dots,\boldsymbol{X}_p)$, $\boldsymbol{X}_i \in \mathbf{R}^T$, where $\mathrm{vec}\boldsymbol{X} \in \mathbf{R}^{pT}$ is a predictor vector, and $Y \in \{1,\dots,K\}$ is a categorical response variable representing the group membership. We are interested in predicting the class membership $Y$ based on the $p$ variables measured in $T$ different time points or physical conditions. Suppose that group $i$ has group mean vector $\boldsymbol{\mu}_i \in \mathbf{R}^{pT}$, a common (within-group) $pT \times pT$ covariance matrix $\boldsymbol{\Sigma}$ and associated group probability $q_i > 0$, $i = 1,\dots,K$. That is $\mathrm{E}(\mathrm{vec}(\boldsymbol{X})|Y = i) = \boldsymbol{\mu}_i$, $\mathrm{Var}(\mathrm{vec}(\boldsymbol{X})|Y = i) = \boldsymbol{\Sigma} > 0$, for $i = 1,\dots,K$ and $P(Y = i) = q_i > 0$, $q_1 + \cdots + q_K = 1$. Discriminant coordinates are then defined to be the linear combination $U = \boldsymbol{u}' \mathrm{vec}(\boldsymbol{X})$, which maximizes the ratio of the between-group variance to the within-group variance.

Specifically, let $\mathbf{\Delta}$ be the between-group covariance matrix defined by

$$\mathbf{\Delta} = \sum_{i=1}^{K} q_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})',$$

for

$$\boldsymbol{\mu} = \sum_{i=1}^{K} q_i \boldsymbol{\mu}_i.$$

The between-group covariance matrix $\mathbf{\Delta}$ is nonnegative definite matrix. Then, the ratio of the between-group variance to the within-group variance is equal to

$$J(\boldsymbol{u}) = \frac{\boldsymbol{u}'\mathbf{\Delta}\boldsymbol{u}}{\boldsymbol{u}'\mathbf{\Sigma}\boldsymbol{u}}, \tag{1}$$

provided that $\boldsymbol{u} \in \mathbf{R}^{pT} \neq \mathbf{0}$.

If $\boldsymbol{u}_1$ is the vector which maximizes (1), we call the corresponding linear combination $U_1 = \boldsymbol{u}_1' \operatorname{vec}(\boldsymbol{X})$ the first discriminant coordinate. In particular, $\boldsymbol{u}_1$ can be obtained by solving

$$\max_{\boldsymbol{u} \in \mathbf{R}^{pT}} \boldsymbol{u}'\mathbf{\Delta}\boldsymbol{u}$$

subject to

$$\boldsymbol{u}'\mathbf{\Sigma}\boldsymbol{u} = 1.$$

Since $\mathbf{\Sigma}$ is a nonsingular matrix, then $\boldsymbol{u}_1$ is the eigenvector of $\mathbf{\Sigma}^{-1}\mathbf{\Delta}$ corresponding to its largest eigenvalue $\lambda_1$.

The second discriminant coordinate maximizes the measure $J(\boldsymbol{u})$ and satisfies the conditions:

$$\boldsymbol{u}_2'\mathbf{\Sigma}\boldsymbol{u}_2 = 1, \quad \boldsymbol{u}_1'\mathbf{\Sigma}\boldsymbol{u}_2 = 0.$$

Continuing this process, we can define the $k$-th discriminant coordinate as far as maximizing the measure $J(\boldsymbol{u})$, which must also comply with conditions:

$$\boldsymbol{u}_k'\mathbf{\Sigma}\boldsymbol{u}_l = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

$k, l = 1, \ldots, s = \operatorname{rank}\mathbf{\Delta}$.

This means that the discriminant coordinates are uncorrelated and have unit variance.

The vectors $\boldsymbol{u}_k$, which maximize the measure $J(\boldsymbol{u})$, fulfill the equality

$$(\mathbf{\Delta} - \lambda_k\mathbf{\Sigma})\boldsymbol{u}_k = \mathbf{0},$$

where $\lambda_1 \geq \cdots \geq \lambda_s > \lambda_{s+1} = \cdots = \lambda_{pT} = 0$ are the eigenvalues of the matrix $\mathbf{\Sigma}^{-1}\mathbf{\Delta}$, $k = 1, \ldots, s = \operatorname{rank}\mathbf{\Delta}$.

Note that the construction of discriminant coordinates requires knowledge of the vectors $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_K$, prior probabilities $q_1,\ldots,q_K$, and the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$. In practice these parameters are not known, and we need to use their estimates from the sample.

Let $\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{in_i}$ be a sample derived from the $i$th group, where $i=1,\ldots,K$, and let $n=n_1+\cdots+n_K$. Then

$$\hat{q}_i=\frac{n_i}{n},$$

$$\hat{\boldsymbol{\mu}}_i=\bar{\boldsymbol{x}}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}\mathrm{vec}(\boldsymbol{x}_{ij}),$$

$$\hat{\boldsymbol{\mu}}=\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{i=1}^{K}n_i\bar{\boldsymbol{x}}_i,$$

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{n-K}\boldsymbol{W},\quad \text{where}\quad \boldsymbol{W}=\sum_{i=1}^{K}\boldsymbol{A}_i,\ \boldsymbol{A}_i=\sum_{j=1}^{n_i}(\mathrm{vec}(\boldsymbol{x}_{ij})-\bar{\boldsymbol{x}}_i)(\mathrm{vec}(\boldsymbol{x}_{ij})-\bar{\boldsymbol{x}}_i)',$$

$$\hat{\boldsymbol{\Delta}}=\frac{1}{K-1}\boldsymbol{B},\quad \text{where}\quad \boldsymbol{B}=\sum_{i=1}^{K}n_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})',\ i=1,\ldots,K.$$

Note that the matrix $\boldsymbol{W}$ is the sum of the matrices $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K$. The matrix $\boldsymbol{A}_i$ is positive definite with probability 1 if and only if $n_i>pT$, $i=1,\ldots,K$. Then $\boldsymbol{W}$ is also positive definite if and only if $\max\{n_1,\ldots,n_K\}>pT$ (Banerjee and Roy (2004), p. 418; Giri (1996), p. 93). Therefore, if $\max\{n_1,\ldots,n_K\}>pT$, then the estimate $\hat{\boldsymbol{\Sigma}}$ of the positive definite matrix $\boldsymbol{\Sigma}$ is positive definite with probability 1, and we can use the given estimates of unknown parameters. The condition $\max\{n_1,\ldots,n_K\}>pT$ is very restrictive and requires large samples. If it is not satisfied, then the problem of correct estimation of the matrix $\boldsymbol{W}$ can be partially solved using an existing relationship between the discriminant coordinates and canonical variables.

## 3. The relationship between the discriminant coordinates and canonical variables

In the case where $\max\{n_1,\ldots,n_K\}\leq pT$ estimates of the unknown parameters will be calculated using the relationship between discriminant coordinates and canonical variables (Krzyśko (1979)).

Let the $q$-dimensional vector $\boldsymbol{Y}$ be a vector of dummy variables defined as follows:

$$Y_i=\begin{cases}1, & \text{if the matrix } \boldsymbol{X} \text{ is observed in the } i\text{th group},\\ 0, & \text{in other cases,}\end{cases}$$

$i=1,\ldots,q=K-1$.

Let

$$\boldsymbol{Z} = \begin{bmatrix} \operatorname{vec}(\boldsymbol{X}) \\ \boldsymbol{Y} \end{bmatrix}$$

and let

$$\operatorname{Var}(\boldsymbol{Z}) = \begin{bmatrix} \operatorname{Var}(\operatorname{vec}(\boldsymbol{X})) & \operatorname{Cov}(\operatorname{vec}(\boldsymbol{X}),\boldsymbol{Y}) \\ \operatorname{Cov}(\boldsymbol{Y},\operatorname{vec}(\boldsymbol{X})) & \operatorname{Var}(\boldsymbol{Y}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} = \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega}_{21} = \boldsymbol{\Omega}'_{12}$ and $\boldsymbol{\Omega}$ is positive definite.

The estimate $\hat{\boldsymbol{\Omega}}$ of the positive definite matrix $\boldsymbol{\Omega}$ is positive definite with probability 1 if and only if $n > pT + q$, where $n = n_1 + \cdots + n_K$. If the matrix $\hat{\boldsymbol{\Omega}}$ is positive definite, then the matrices $\hat{\boldsymbol{\Omega}}_{11}$ and $\hat{\boldsymbol{\Omega}}_{22}$ are non-singular.

Let

$$\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\Omega}}_{11}^{-1}\hat{\boldsymbol{\Omega}}_{12}\hat{\boldsymbol{\Omega}}_{22}^{-1}\hat{\boldsymbol{\Omega}}_{21}.$$

Consider the equation

$$(\hat{\boldsymbol{\Gamma}} - r^2\boldsymbol{I})\boldsymbol{m} = \boldsymbol{0}. \tag{2}$$

Variables $V_k = \boldsymbol{m}'_k \operatorname{vec}(\boldsymbol{X})$, where $\boldsymbol{m}_k \in \mathbf{R}^{pT}$ are eigenvectors of the matrix $\hat{\boldsymbol{\Gamma}}$ satisfying equation (2), are called sample canonical variables of the $\boldsymbol{X}$-space.

The following relationships are satisfied (Krzyśko (1979)):

$$\boldsymbol{W} = n\hat{\boldsymbol{\Omega}}_{11} - (n\hat{\boldsymbol{\Omega}}_{12})(n\hat{\boldsymbol{\Omega}}_{22})^{-1}(n\hat{\boldsymbol{\Omega}}_{21}),$$
$$\boldsymbol{B} = (n\hat{\boldsymbol{\Omega}}_{12})(n\hat{\boldsymbol{\Omega}}_{22})^{-1}(n\hat{\boldsymbol{\Omega}}_{21}).$$

Thus, equation (2) is equivalent to the equation

$$(\boldsymbol{B} - \lambda\boldsymbol{W})\boldsymbol{m} = \boldsymbol{0},$$

where $\lambda = r^2(1-r^2)^{-1}$.

This means that the discriminant coordinates $U_k = \boldsymbol{u}'_k \operatorname{vec}(\boldsymbol{X})$ are proportional to the canonical variables of the $\boldsymbol{X}$-space $V_k = \boldsymbol{m}'_k \operatorname{vec}(\boldsymbol{X})$, where $\boldsymbol{Y}$ is a vector of dummy variables. Note that the condition $n > pT + q$ is a condition much weaker than the condition $\max\{n_1, \ldots, n_K\} > pT$, especially for a small number of groups $K = q + 1$.

## 4. The special structure of the matrix Ω

If $n \leq pT + q$, then we can construct the discriminant coordinates with the additional condition that assumes that

$$\boldsymbol{\Omega}_{11} = \boldsymbol{U} \otimes \boldsymbol{V},$$

where $\boldsymbol{U} > 0$, $\boldsymbol{V} > 0$.

The matrix $\boldsymbol{U}$ represents the covariance between all $p$-variables on a given object and for a given time point. Likewise, $\boldsymbol{V}$ represents the covariance between repeated measures on a given object and for a given variable. The above covariance structure is subject to an implicit assumption that for all variables the correlation structure between repeated measures remains the same, and that covariance between variables does not depend on time and remains constant for all time points.

Estimates of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, and thus the matrix $\boldsymbol{\Omega}_{11}$ can be obtained using the method given by Srivastava et al. (2008).

Let

$$\bar{\boldsymbol{x}}^{\star} = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j,$$

$$\boldsymbol{x}_{j,c} = \boldsymbol{x}_j - \bar{\boldsymbol{x}}^{\star},\ j = 1,\ldots,n.$$

Then, estimates of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are obtained iteratively with a system of equations

$$\hat{\boldsymbol{U}} = \frac{1}{nT}\sum_{j=1}^{n}\boldsymbol{x}'_{j,c}\hat{\boldsymbol{V}}^{-1}\boldsymbol{x}_{j,c},$$

$$\hat{\boldsymbol{V}} = \frac{1}{np}\sum_{j=1}^{n}\boldsymbol{x}_{j,c}\hat{\boldsymbol{U}}^{-1}\boldsymbol{x}'_{j,c}.$$

In this case, the matrix $\hat{\boldsymbol{\Omega}}_{11} = \hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}$ is positive definite with probability 1 if and only if $n > \max(p,T)$.

Note that the fact that the matrix $\hat{\boldsymbol{\Omega}}_{11}$ is positive definite with probability 1 does not always guarantee that the matrix $\hat{\boldsymbol{\Omega}}$ is positive definite with probability 1. However, the fact that the matrix $\hat{\boldsymbol{\Omega}}_{11}$ is positive definite with probability 1 allows us to determine the discriminant coordinates on the basis of the matrix $\hat{\boldsymbol{\Gamma}}$ because then the matrix $\hat{\boldsymbol{\Omega}}_{11}^{-1}$ exists.

## 5. Example

The described methods were employed here to build the discriminant coordinates based on the annual data on the 38 European countries in the period 2009-2015. These countries were divided into 4 regions purposes by the United Nations Statistics Division: (1) Northern Europe, (2) Western Europe, (3) Eastern Europe, (4) Southern Europe. The list of countries used in the discriminant coordinates analysis is contained in Table 1.

We used the data published by the World Economic Forum (WEF) in its annual reports (`http://www.weforum.org`). Those are comprehensive data, describing exhaustively various socio-economic conditions or spheres of individual states. For statistical analysis, we chose 2 of 12 pillars of variables: technological readiness (consists of 4 variables) and higher education and training (consists of 6 variables). Table 2 describes the pillars used in the analysis.

Table 1: Countries included in analysis

| | Country | Group | | Country | Group |
|---|---|---|---|---|---|
| 1 | Albania (AL) | 4 | 20 | Lithuania(LT) | 1 |
| 2 | Austria (AT) | 3 | 21 | Luxembourg (LU) | 3 |
| 3 | Belgium (BE) | 3 | 22 | Macedonia FYR (MK) | 4 |
| 4 | Bosnia and Herzegovina (BA) | 4 | 23 | Malta (MT) | 4 |
| 5 | Bulgaria (BG) | 2 | 24 | Montenegro (ME) | 4 |
| 6 | Croatia (HR) | 4 | 25 | Netherlands (NL) | 3 |
| 7 | Cyprus (CY) | 4 | 26 | Norway (NO) | 1 |
| 8 | Czech Republic (CZ) | 2 | 27 | Poland (PL) | 2 |
| 9 | Denmark (DK) | 1 | 28 | Portugal (PT) | 4 |
| 10 | Estonia (EE) | 1 | 29 | Romania (RO) | 2 |
| 11 | Finland (FI) | 1 | 30 | Russian Federation (RU) | 2 |
| 12 | France (FR) | 3 | 31 | Serbia (XS) | 4 |
| 13 | Germany (DE) | 3 | 32 | Slovak Republic (SK) | 2 |
| 14 | Greece (GR) | 4 | 33 | Slovenia (SI) | 4 |
| 15 | Hungary (HU) | 2 | 34 | Spain (ES) | 4 |
| 16 | Iceland (IS) | 1 | 35 | Sweden (SE) | 1 |
| 17 | Ireland (IE) | 1 | 36 | Switzerland (CH) | 3 |
| 18 | Italy (IT) | 4 | 37 | Ukraine (UA) | 2 |
| 19 | Latvia (LV) | 1 | 38 | United Kingdom (GB) | 1 |

Table 2: Variables used in analysis

| Pillars | Variables |
|---|---|
| Technological readiness | Availability of latest technologies ($X_1$) |
| | Firm-level technology absorption ($X_2$) |
| | FDI and technology transfer ($X_3$) |
| | Internet users ($X_4$) |
| Higher education and training | Quality of the educational system ($X_1$) |
| | Quality of math and science education ($X_2$) |
| | Quality of management schools ($X_3$) |
| | Internet access in schools ($X_4$) |
| | Local availability of specialized research and training services ($X_5$) |
| | Extent of staff training ($X_6$) |

In both cases $\max\{n_1,\ldots,n_K\} \leq pT$, and we could not use the classical discriminant coordinates algorithm. The unknown parameters were calculated using the relationship between discriminant coordinates and canonical variables described in Section 3. The technological readiness pillar consists of 4 variables ($p = 4$). In this case $n > pT + q$. The discriminant coordinates are uncorrelated. However, they are not orthogonal. In practice, however, the usual procedure is to plot discriminant coordinates on a rectangular coordinate system. The resulting distortion is generally not serious. Projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$ is presented in Figure 1.

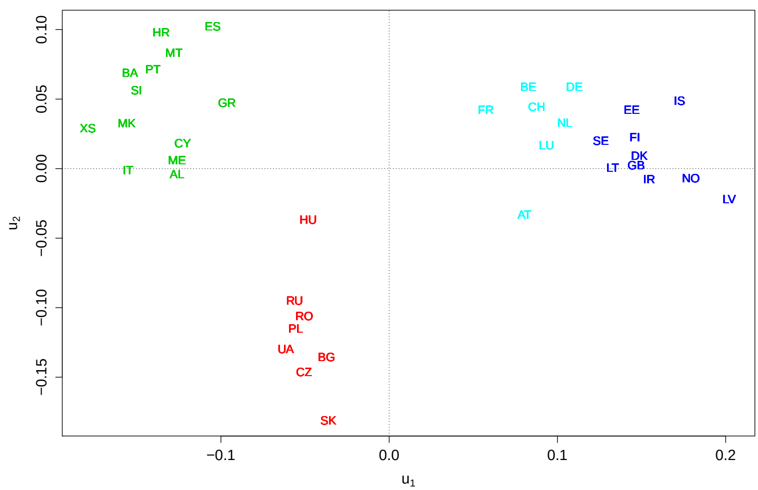


Figure 1: Technological readiness. Projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$. Regions used for statistical processing purposes by the United Nations Statistics Division: ■ – Northern Europe, ■ – Western Europe, ■ – Eastern Europe, ■ – Southern Europe

Figure 1 confirms that the European countries in terms of four characteristics of forming the technological readiness pillar are divided into four groups. However, the difference between the countries of Western Europe and the countries of Northern Europe is small.

The contribution of each variable to the discriminant coordinate is not the same. The correlation between each variable and a discriminant coordinate is widely recommended as a useful measure of variable importance in the discriminant coordinate. These correlations, sometimes called structure coefficients, are provided in many software packages. However, it turns out that these correlations do not show the multivariate contribution of each variable, but rather provide only univariate information, showing how each variable by itself separates the groups, ignoring the presence of the other variables. The better measure are the absolute values of standardized coefficients, because these coefficients show the contribution of the variables in the presence of each other, that is, these coefficients provide a multivariate interpretation that allows for the correlations among the variables (Rencher (1998), p. 214 ). If we denote the $r$th coefficient in the $q$th discriminant coordinate by $m_{qr}$ then the standardized form is $m^{\star}_{qr} = s_r m_{qr}$, where $s_r$ is the within-group standard deviation of the $r$th variable obtained from the diagonal of $(n-K)^{-1}\boldsymbol{W}$.

The absolute values of the standardized coefficients can be used to rank the variables in order of their contribution to separating the groups. Tables 3 and 4 show the standardized discriminant coordinate coefficients of the first and second discriminant coordinate, respectively, for the technological readiness data.

Table 3: Technological readiness. The standardized coefficients of the first discriminant coordinate $\hat{U}_1$.

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|
| $X_1$ | 0.0864 | -0.0537 | -0.0963 | -0.1803 | 0.1950 | -0.1269 | 0.0924 |
| $X_2$ | 0.0133 | 0.1138 | -0.2164 | 0.2248 | -0.1485 | 0.0876 | -0.0004 |
| $X_3$ | 0.0412 | -0.1393 | -0.0500 | 0.1100 | 0.0013 | -0.0387 | 0.0841 |
| $X_4$ | -0.0055 | -0.0324 | 0.0696 | 0.0228 | -0.0400 | -0.0082 | 0.0397 |

Table 4: Technological readiness. The standardized coefficients of the second discriminant coordinate $\hat{U}_2$.

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|
| $X_1$ | -0.0830 | 0.1067 | -0.0487 | -0.0289 | -0.1082 | 0.3113 | -0.0686 |
| $X_2$ | -0.0618 | 0.0751 | -0.0677 | 0.1204 | -0.0575 | 0.0096 | -0.0629 |
| $X_3$ | -0.0829 | 0.1346 | -0.1693 | 0.1705 | -0.0329 | -0.1491 | 0.1036 |
| $X_4$ | -0.0358 | 0.0071 | 0.0471 | -0.0439 | -0.0060 | -0.0108 | 0.0245 |

The higher education and training pillar consists of 6 variables ($p = 6$). In this case $n \leq pT + q$, and we made the calculation with the additional condition $\boldsymbol{\Omega}_{11} = \boldsymbol{U} \otimes \boldsymbol{V}$. A projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$ is presented in Figure 2.

Figure 2 shows three clusters. The countries of Western Europe and Northern Europe in fact form a single group. The outliers countries in this group are Lithuania (LT) and Latvia (LV).

Tables 7 and 8 show the standardized discriminant coordinate coefficients of the first and second discriminant coordinate, respectively, for the higher education and training data.

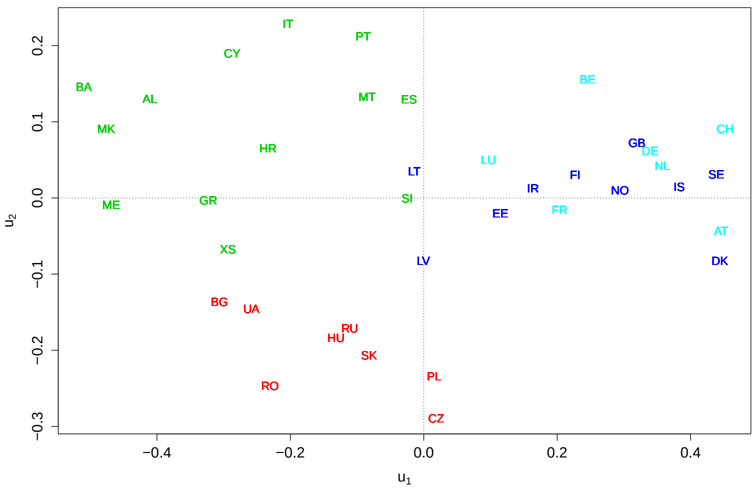


Figure 2: Higher education and training. Projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$. Regions used for statistical processing purposes by the United Nations Statistics Division: ■ – Northern Europe, ■ – Western Europe, ■ – Eastern Europe, ■ – Southern Europe

Table 5: Higher education and training. The standardized coefficients of the first discriminant coordinate $\hat{U}_1$.

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|
| $X_1$ | 0.1232 | -0.2098 | 0.1197 | -0.1592 | 0.2263 | -0.1633 | 0.0831 |
| $X_2$ | -0.0356 | -0.0918 | 0.1149 | 0.1016 | -0.0958 | 0.0126 | -0.0290 |
| $X_3$ | -0.0261 | 0.1900 | -0.1579 | 0.0123 | -0.0119 | 0.0524 | -0.0450 |
| $X_4$ | 0.1023 | -0.0006 | 0.0861 | -0.0974 | -0.0743 | -0.0163 | 0.0537 |
| $X_5$ | 0.0128 | -0.0820 | 0.0797 | -0.0054 | 0.0534 | -0.1114 | 0.1008 |
| $X_6$ | 0.0338 | -0.0280 | -0.0398 | 0.0350 | 0.0367 | 0.0169 | -0.0320 |

Table 6: Higher education and training. The standardized coefficients of the second discriminant coordinate $\hat{U}_2$.

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|
| $X_1$ | -0.0725 | 0.2119 | -0.1886 | 0.1291 | -0.1413 | 0.1061 | -0.0254 |
| $X_2$ | 0.0638 | -0.2120 | 0.1114 | -0.1125 | 0.1617 | -0.0979 | 0.0661 |
| $X_3$ | -0.0615 | 0.0122 | 0.0464 | 0.0027 | 0.0159 | 0.0551 | -0.0473 |
| $X_4$ | -0.0810 | 0.1384 | -0.0576 | 0.0334 | -0.0219 | 0.0517 | -0.0722 |
| $X_5$ | 0.1000 | -0.1485 | 0.0805 | -0.0799 | 0.0571 | -0.0048 | -0.0037 |
| $X_6$ | 0.0296 | -0.0453 | 0.0292 | 0.0333 | -0.1020 | 0.0755 | -0.0132 |

During the numerical calculation process we used R software (R Core Team (2015)).

## REFERENCES

BANERJEE, S., ROY, A., (2014). Linear Algebra and Matrix Analysis for Statistics. CRC Press, Boca Ratop.

FISHER, R. A., (1936). The use of multiple measurements in taxonomic problem. Annals of Eugenics 7, pp. 179–188.

FUJIKOSHI, Y., ULYANOV, V. V., SHIMIZU, R., (2010). Multivariate statistics. High-Dimensional and Large-Sample Approximations. Hoboken, New Jersey.

GIRI, C. G., (1996). Multivariate Statistical Analysis. Marcel Dekker, Inc., New York.

GÓRECKI, T., KRZYŚKO, M., WASZAK, Ł., WOŁYŃSKI, W., (2018). Selected statistical methods of data analysis for multivariate functional data, Statistical Papers 59, pp. 153–182.

KRZANOWSKI, W. J., (2000). Principles of Multivariate Analysis, Revised Edition. Oxford University Press, Oxford.

KRZYŚKO, M., (1979). Discriminant variables. Biometrical Journal 21 (3), pp. 227–241.

RAO, C. R., (1948). The utilization of multiple measurements in problem of biological classification. Journal of the Royal Statistical Society Series B, 10 (2), pp. 159–203.

R CORE TEAM, (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RENCHER, A. C., (1998). Multivariate Statistical Inference and Applications. Wiley, New York.

SEBER, G. A. F., (1984). Multivariate Observations. John Wiley and Sons, New York.

SRIVASTAVA, M. S., (2002). Methods of Multivariate Statistics. Wiley, New York.

SRIVASTAVA, M. S., VON ROSEN, T., VON ROSEN, D., (2008). Models with a Kronecker product covariance structure: estimation and testing. Mathematical Methods of Statistics 17 (4), pp. 357–370.