# On comparing populations based on two sets of variables

Dominika Polko-Zając[1]

**Abstract**

In an economic and social studies it is often necessary to test the differences between the two sets of variables. Multidimensional comparisons allow researchers to a thorough analysis of the studied phenomenon. The article concerned the problem of comparing multidimensional populations using canonical correlation analysis. In order to identify differences between the analysed sets of variables permutation tests were used. These tests do not require additional assumptions about the form of the distribution in the population; are suitable for small sample sizes and are robust to outliers. The properties of these tests were characterized using a computer simulation in R program.

***Keywords:*** *multidimensional data, canonical correlation, permutation tests, Monte Carlo study*
***JEL Classification:*** C12, C15, C30

## 1 Introduction – testing differences between two correlation coefficients

Significance tests that verify equality of correlation coefficients are important in population studies. From two populations where tested variables have two–dimensional normal distributions of unknown correlation coefficients $\rho_1$ and $\rho_2$ samples of $n_i$ elements $(n_i > 4)$ for $i = 1,2$ are taken. The null hypothesis which says that the correlation coefficients in both compared independent populations are equal

$$H_0 : \rho_1 = \rho_2 \tag{1}$$

against alternative hypothesis

$$H_1 : \rho_1 \neq \rho_2, \tag{2}$$

can be verified using test statistic in form of (Domański, 1990):

$$Z = (z_1 - z_2)\sqrt{\frac{(n_1 - 3)(n_2 - 3)}{n_1 + n_2 - 6}}, \tag{3}$$

where: $z_i = \frac{1}{2}\ln\frac{1+r_i}{1-r_i}$, for $i = 1,2$ and $r_i$ are sample canonical correlation coefficient.

The statistic has asymptotic distribution $N(0,1)$ when the null hypothesis is assumed to be true. Hypothesis *H0* is rejected in favour of *H1* if $|Z| > z_\alpha$.

[1] Corresponding author: University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, Bogucicka 14, 40-226 Katowice, Poland, e-mail: dpolko@gmail.com.

## 2 Correlation coefficients in multidimensional analysis

Multidimensional methods are significant part of statistical methods. This stems mainly from the fact that in many areas of empirical research they relate to phenomena of complex, multidimensional structure. The Pearson's correlation coefficient measures relation between only two random variables *Y* and *X*

$$(Y, X) \longrightarrow \rho = \frac{Cov(Y, X)}{\sqrt{Var(Y)Var(X)}} \tag{4}$$

and is a quantity ranging from -1 to 1.

The multiple correlation coefficient measures, in turn, the relationship between a variable *Y*, and sets of *q* variables $\mathbf{X}=(X_1, X_2, \ldots, X_q)$. In fact this is the maximal correlation coefficient between a variable *Y* and a linear combination of variables *X*

$$(Y, \mathbf{X}) \longrightarrow \rho = \max_{\mathbf{B}} \frac{Cov(Y, \mathbf{B}^{\mathbf{T}}\mathbf{X})}{\sqrt{Var(Y)Var(\mathbf{B}^{\mathbf{T}}\mathbf{X})}} \tag{5}$$

where: $\mathbf{B}=(B_1, B_2, \ldots, B_q)$ and *n* – the number of observations of each variable.

The multiple correlation coefficient has values from the interval $\langle 0,1 \rangle$.

The canonical analysis is a generalization of the concept of the multiple correlation to the case of correlation between the two sets of random variables. The objective of the canonical analysis is to test the strength of the relationship between two sets of variables $\mathbf{Y}=(Y_1, Y_2, \ldots, Y_p)$ and $\mathbf{X}=(X_1, X_2, \ldots, X_q)$

$$(\mathbf{Y}, \mathbf{X}) \longrightarrow \rho = \max_{\mathbf{A},\mathbf{B}} \frac{Cov(\mathbf{A}^{\mathbf{T}}\mathbf{Y}, \mathbf{B}^{\mathbf{T}}\mathbf{X})}{\sqrt{Var(\mathbf{A}^{\mathbf{T}}\mathbf{Y})Var(\mathbf{B}^{\mathbf{T}}\mathbf{X})}} \tag{6}$$

where: $\mathbf{A}=(A_1, A_2, \ldots, A_p)$, $\mathbf{B}=(B_1, B_2, \ldots, B_q)$, *n* – the number of observations of each variable.

The canonical correlation coefficient is a quantity ranging from 0 to 1.

## 3 Canonical correlation analysis

The basic concept of the canonical analysis was developed by Hotelling (1936). Hotelling introduced the idea of canonical variables and canonical correlation and studied the relationship between the sets of variables on the skills of fast comprehension reading, and fast executing arithmetic calculations.

In the canonical correlation analysis both the independent variable and the dependent variable are multidimensional. The population with *p*+*q*–dimensional characteristic is considered to study the dependency between *p*–coordinates of the characteristic which are considered as *p*–dimensional dependent variable and remaining *q*–coordinates of the characteristic which are

considered as *q*–dimensional independent variable (Kosiorowski, 2008). The sample of *n* vectors of observations of the tested characteristic can be specified as follows

$$\left(\mathbf{Y_i}, \mathbf{X_i}\right)^T = \left(Y_{i1}, Y_{i2}, \ldots, Y_{ip}, X_{i1}, X_{i2}, \ldots, X_{iq}\right)^T,\ i = 1,2,\ldots,n. \tag{7}$$

The sample covariance matrix of $p+q$ variables is in the form of

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \tag{8}$$

where: $\mathbf{S}_{11}$ – covariance matrix of *Y*'s of dimension $p \times p$,

$\mathbf{S_{12}} = \mathbf{S_{21}^T}$ – covariance matrix of *Y*'s and *X*'s of dimension $p \times q$,

$\mathbf{S}_{22}$ – covariance matrix of *X*'s of dimension $q \times q$.

The objective of the canonical analysis is to test the strength of the association. Coefficient vectors **A** and **B** are pursued such that linear combinations of dependent variables $\mathbf{U=A^TY}$ and linear combinations of independent variables $\mathbf{V=B^TX}$, called canonical variables (canonical variates) are maximally correlated. These combinations give an insight into the relationship between the two sets of variables.

In the first stage of the canonical analysis coefficient vectors of the first pair of canonical variables are sought. Coefficient vectors are selected to maximize the correlation between the first pair of canonical variables, that is, maximize the expression

$$r_1 = r_{u_1, v_1} = \frac{\left(\mathbf{A_1^T S_{12} B_1}\right)}{\left[\left(\mathbf{A_1^T S_{11} A_1}\right)\left(\mathbf{B_1^T S_{22} B_1}\right)\right]^{\frac{1}{2}}} \tag{9}$$

where $r_{u_1, v_1}$ stands for canonical correlation coefficient.

Eigenvalues are obtained from equations

$$\left|\mathbf{S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}} - \lambda \mathbf{I}\right| = 0$$

$$\left|\mathbf{S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}} - \lambda \mathbf{I}\right| = 0.$$

Number of non–zero eigenvalues of equations is denoted as $k = \min(p, q)$. Once eigenvalue with greatest $\lambda_1$ value is found we search for coefficient vectors for the first pair of canonical variables. They are determined by solving the characteristic equations

$$\left(\mathbf{S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}} - \lambda_1 \mathbf{I}\right)\mathbf{A}_1 = 0$$

$$\left(\mathbf{S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}} - \lambda_1 \mathbf{I}\right)\mathbf{B}_1 = 0$$

where:

$\lambda_1 = r_1^2$ – eigenvalue of matrix $\mathbf{S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}}$ or $\mathbf{S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}}$.

When we obtain the first canonical correlation vectors for *Y* and *X*; further canonical correlation vectors can be found in the uncorrelated directions to the previous ones in the same manner. A few pairs of canonical variates are expected to represent the original sets of variables to explain their relation and variabilities.

In general there are *k* canonical correlations $r_1, r_2, \ldots, r_k$ corresponding to the $k = \min(p, q)$ pairs of canonical variates $u_i = \mathbf{A_i^T Y}$ and $v_i = \mathbf{B_i^T X}$:

$$\begin{array}{cccccc} r_1 & u_1 & = & \mathbf{A_1^T Y} & v_1 & = & \mathbf{B_1^T X} \\ r_2 & u_2 & = & \mathbf{A_2^T Y} & v_2 & = & \mathbf{B_2^T X} \\ \vdots & & \vdots & & & \vdots & \\ r_k & u_k & = & \mathbf{A_k^T Y} & v_k & = & \mathbf{B_k^T X} \end{array}$$

for each *i*=1,2,…,*k*, $r_i$ is the sample correlation between $u_i$ and $v_i$; that is $r_i = r_{u_i, v_1}$.

The pairs ($u_i$,$v_i$), *i*=1,2,…,*k*, provide the *k* dimensions of the relationship.

Canonical correlation analysis is a useful technique for simplifying the correlation structure between two sets of variables (Yamada and Sugiyama, 2006). The most important assumptions of classical canonical analysis are:

- multidimensional normality of distribution of variables in the population,
- suitable size of the sample, at least 20 times larger than number of variables,
- the lack of collinear variables,
- the lack of outliers.

## 4 Testing differences based on canonical correlation coefficients

Comparing populations based on multi–dimensional sets of variables the null hypothesis can be stated as follows: “in the underlying populations all corresponding canonical correlation coefficients are equal”. The alternative hypothesis is formulated: “in the underlying populations corresponding canonical correlation coefficients are not equal for at least one pair”. To test the null hypothesis the permutation test can be used. Formally these hypotheses can be written as follows

$$H_0 : \begin{pmatrix} \rho_{11} \\ \rho_{21} \\ \cdots \\ \rho_{m1} \end{pmatrix} = \begin{pmatrix} \rho_{12} \\ \rho_{22} \\ \cdots \\ \rho_{m2} \end{pmatrix} \tag{10}$$

and the alternative

$$H_1 : \rho_{i1} \neq \rho_{j2} \tag{11}$$

for some *i*, *j* = 1, 2, …, *m*, where *i*=*j*.

It is more interesting for researcher to employ directional alternative hypotheses. The one–sided alternative hypothesis can be stated as follows

$$H_1 : \rho_{i1} > \rho_{j2} \ \ or \ \ H_1 : \rho_{i1} < \rho_{j2} \ . \tag{12}$$

The most important is to test the significance of the difference between the first canonical correlations, which determines to the greatest extent the relationship between the two sets of variables. To test null hypothesis against alternative hypothesis we can use the following test statistic

$$T = r_{i1} - r_{j2} \tag{13}$$

where: $r_{i1}, r_{j2}$ (for *i, j=1*) are first, sample canonical correlation coefficients.

Tests based on permutations of observations were introduced by R.A. Fisher in 1930's (Welch, 1990). Because of the need to perform complex calculations method was widely used only in recent decades, when computing capabilities of computers increased. The basic idea behind permutation methods is to generate a reference distribution by recalculating a statistic for many permutations of the data (Ernst, 2004). Permutation tests in general take a test statistic *T* used for a parametric test, or one derived intuitively (Baker, 1995). Currently, the problem of using the permutation tests in statistical analysis is popular among researchers. The most important references are Good (1994, 2005, 2006), Pesarin (2001), Basso et al. (2009), Pesarin and Salmaso (2010) and Kończak (2016).

These tests do not require additional assumptions about the form of the distribution in the population; are suitable for small sample sizes and are robust to outliers. The goal of the test is to verify hypothesis at certain level of significance to discover a correlation between data sets. After the value of the statistic $T_0$ had been calculated, *N* permutations of variables were performed and values $T_i$ ($i$ = 1, 2, …, $N$) were determined. The decision concerning a verified hypothesis is made on the basis of *ASL* (*achieved significance level*) value (Efron and Tibshirani, 1993):

$$ASL = P_{H_0}\{|T| \geq |T_0|\}. \tag{14}$$

On the basis of the random variable of large size (it is recommended in most cases the number of permutation to be greater than 1000) taken from the set of all possible permutations of the data set the *ASL* is determined using formula (Kończak, 2016)

$$\hat{ASL} = \frac{card\{i : |T_i| \geq |T_0|\}}{N}. \tag{15}$$

Sample data

| $Y_1$ | $Y_2$ | . . | $Y_p$ | $X_1$ | $X_2$ | ... | $X_q$ |
|---|---|---|---|---|---|---|---|
| $y_{111}$ | $y_{121}$ | ... | $y_{1p1}$ | $x_{111}$ | $x_{121}$ | ... | $x_{1q1}$ |
| $y_{211}$ | $y_{221}$ | ... | $y_{2p1}$ | $x_{211}$ | $x_{221}$ | ... | $x_{2q1}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $y_{n11}$ | $y_{n21}$ | ... | $y_{np1}$ | $x_{n11}$ | $x_{n21}$ | ... | $x_{nq1}$ |
| $y_{112}$ | $y_{122}$ | ... | $y_{1p2}$ | $x_{112}$ | $x_{122}$ | ... | $x_{1q2}$ |
| $y_{212}$ | $y_{222}$ | ... | $y_{2p2}$ | $x_{212}$ | $x_{222}$ | ... | $x_{2q2}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $y_{n12}$ | $y_{n22}$ | ... | $y_{np2}$ | $x_{n12}$ | $x_{n22}$ | ... | $x_{nq2}$ |

Subsequent permutations of variables

**1**

| $Y_1$ | $Y_2$ | ... | $Y_p$ | $X_1$ | $X_2$ | ... | $X_q$ |
|---|---|---|---|---|---|---|---|
| $y_{211}$ | $y_{221}$ | ... | $y_{2p1}$ | $x_{211}$ | $x_{221}$ | ... | $x_{2q1}$ |
| $y_{112}$ | $y_{122}$ | ... | $y_{1p2}$ | $x_{112}$ | $x_{122}$ | ... | $x_{1q2}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $y_{n11}$ | $y_{n21}$ | ... | $y_{np1}$ | $x_{n11}$ | $x_{n21}$ | ... | $x_{nq1}$ |
| $y_{512}$ | $y_{522}$ | ... | $y_{5p2}$ | $x_{512}$ | $x_{522}$ | ... | $x_{5q2}$ |
| $y_{212}$ | $y_{222}$ | ... | $y_{2p2}$ | $x_{212}$ | $x_{222}$ | ... | $x_{2q2}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $y_{811}$ | $y_{821}$ | ... | $y_{8p1}$ | $x_{811}$ | $x_{821}$ | ... | $x_{8q1}$ |

**...** **...**

**N**

| $Y_1$ | $Y_2$ | ... | $Y_p$ | $X_1$ | $X_2$ | ... | $X_q$ |
|---|---|---|---|---|---|---|---|
| $y_{712}$ | $y_{722}$ | ... | $y_{7p2}$ | $x_{712}$ | $x_{722}$ | ... | $x_{7q2}$ |
| $y_{311}$ | $y_{321}$ | ... | $y_{3p1}$ | $x_{311}$ | $x_{321}$ | ... | $x_{3q1}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $y_{n12}$ | $y_{n22}$ | ... | $y_{np2}$ | $x_{n12}$ | $x_{n22}$ | ... | $x_{nq2}$ |
| $y_{511}$ | $y_{521}$ | ... | $y_{5p1}$ | $x_{511}$ | $x_{521}$ | ... | $x_{5q1}$ |
| $y_{312}$ | $y_{322}$ | ... | $y_{3p2}$ | $x_{312}$ | $x_{322}$ | ... | $x_{3q2}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $y_{111}$ | $y_{121}$ | ... | $y_{1p1}$ | $x_{111}$ | $x_{121}$ | ... | $x_{1q1}$ |

**Fig. 1.** The scheme of permutation variables (where $x_{ijk}, y_{ijk}$ are respectively the *i*-th observation of the *j*-th variable for the *k*-th population).

The smaller the value of *ASL*, the stronger evidence against $H_0$. Formally we choose a significance level $\alpha$ and reject $H_0$ if *ASL* is less than $\alpha$.

The steps in the permutation test conducted to determine the significance of the difference between two sets of variables are as follows:

1. Assume the level of significance $\alpha$;
2. Calculate the value of the statistics $T_0$ for the sample data;
3. Proceed permutations of data that destroy existing structure of variables (see. Fig. 1) and calculate test statistic values $T_i$ for these permutations.
4. Create empirical distribution of $T_i$, where ($i$ = 1, 2, …, $N$) and locate calculated value of $T_0$ on this distribution and estimate *ASL* value.

The proposed test procedures do not assume underlying distributions of **Y** and **X**. The simulation study was performed using R program (R Core Team, 2016). Package CCA with function cancor() is freely available from the Comprehensive R Archive Network (CRAN, http://CRAN.R-project.org/) (González et al., 2008).

## 5 Empirical example

To illustrate the possibilities of application the method for the analysis of economic data, the data provided by the Central Statistical Office of Poland (GUS) and Social Diagnosis study (Diagnoza Społeczna) were used. Three variables representing a subjective evaluation of life satisfaction of respondents, and three variables determining the level of socio–economic development were determined. In the Table 1 and Table 2 observations of two periods: year 2000 and year 2015 grouped by voivodships were presented.

| **Voivodship** | $Y_1$ | $Y_2$ | $Y_3$ | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|---|---|---|
| Dolnośląskie | 11.39 | 13.41 | 28.38 | 8.90 | 21.30 | 102.90 |
| Kujawsko-Pomorskie | 14.69 | 13.38 | 29.18 | 8.20 | 17.80 | 89.60 |
| Lubelskie | 14.62 | 11.38 | 35.53 | 9.20 | 14.10 | 71.40 |
| Lubuskie | 17.17 | 14.54 | 26.38 | 8.30 | 20.70 | 89.40 |
| Łódzkie | 13.82 | 11.54 | 26.94 | 9.90 | 16.60 | 88.60 |
| Małopolskie | 10.35 | 8.37 | 37.25 | 10.60 | 11.70 | 89.70 |
| Mazowieckie | 15.57 | 15.65 | 31.43 | 11.70 | 13.10 | 152.80 |
| Opolskie | 15.32 | 16.33 | 36.24 | 8.40 | 15.40 | 83.40 |
| Podkarpackie | 14.80 | 10.49 | 33.08 | 8.10 | 15.90 | 72.70 |
| Podlaskie | 17.42 | 11.30 | 26.23 | 10.00 | 15.20 | 73.40 |
| Pomorskie | 16.20 | 15.48 | 34.26 | 10.30 | 16.70 | 98.90 |
| Śląskie | 17.02 | 13.03 | 29.27 | 6.60 | 17.50 | 106.20 |
| Świętokrzyskie | 10.83 | 10.40 | 30.50 | 8.10 | 15.70 | 77.90 |
| Warmińsko-Mazurskie | 15.14 | 14.81 | 24.80 | 8.00 | 23.60 | 77.50 |
| Wielkopolskie | 17.23 | 16.83 | 37.13 | 7.90 | 13.60 | 106.80 |
| Zachodniopomorskie | 13.98 | 15.41 | 29.14 | 9.50 | 19.10 | 99.00 |

**Table 1.** Data determining the subjective evaluation of life satisfaction and the level of socio–economic development in 2000.

Source: Social Diagnosis and Central Statistical Office of Poland (GUS).

| Voivodship | $Y_1$ | $Y_2$ | $Y_3$ | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|---|---|---|
| Dolnośląskie | 34.32 | 22.22 | 43.43 | 24.20 | 7.00 | 111.50 |
| Kujawsko-Pomorskie | 30.13 | 24.62 | 40.54 | 18.80 | 8.00 | 81.60 |
| Lubelskie | 27.56 | 17.00 | 37.79 | 23.70 | 9.30 | 68.60 |
| Lubuskie | 35.14 | 25.68 | 42.19 | 19.50 | 6.30 | 83.50 |
| Łódzkie | 28.67 | 21.78 | 42.82 | 23.80 | 7.70 | 93.50 |
| Małopolskie | 33.27 | 23.48 | 45.31 | 24.90 | 7.20 | 90.10 |
| Mazowieckie | 30.49 | 23.94 | 41.90 | 33.50 | 6.40 | 159.40 |
| Opolskie | 39.70 | 29.97 | 47.05 | 21.90 | 6.50 | 80.80 |
| Podkarpackie | 25.29 | 17.44 | 41.44 | 22.00 | 11.70 | 70.70 |
| Podlaskie | 25.88 | 19.25 | 41.34 | 24.40 | 6.90 | 71.10 |
| Pomorskie | 39.13 | 32.46 | 53.41 | 24.30 | 6.60 | 95.90 |
| Śląskie | 35.17 | 26.77 | 45.94 | 22.80 | 7.20 | 104.10 |
| Świętokrzyskie | 30.93 | 20.74 | 43.36 | 21.80 | 10.10 | 72.40 |
| Warmińsko-Mazurskie | 29.57 | 21.13 | 29.39 | 19.70 | 9.40 | 71.00 |
| Wielkopolskie | 35.73 | 28.08 | 51.05 | 22.90 | 5.80 | 108.80 |
| Zachodniopomorskie | 36.67 | 22.25 | 42.91 | 21.60 | 7.50 | 84.90 |

**Table 2.** Data determining the subjective evaluation of life satisfaction and the level of socio–economic development in 2015.

Source: Social Diagnosis and Central Statistical Office of Poland (GUS).

First set of variables contains:

$Y_1$ – the percentage of people satisfied or very satisfied with the financial situation of their own family (in %),

$Y_2$ – the percentage of people satisfied or very satisfied with the prospects for the future (in %),

$Y_3$ – the percentage of people satisfied or very satisfied with their education (in %).

Second set contains following variables:

$X_1$ – percentage of population with university education (in %),

$X_2$ – unemployment rate (in %),

$X_3$ – gross domestic product per capita (Poland takes ratio = 100; in 2015 the estimated value).

To test null hypothesis (10) against alternative hypothesis $H_1 : \rho_{11} > \rho_{12}$ the permutation test was used. Significance level $\alpha = 0.05$ was assumed and $N$=1000 permutations of

variables were performed. As test statistic (13) was used. Empirical distribution of statistic was presented on Figure 2. *ASL* value calculated with empirical distribution of statistic

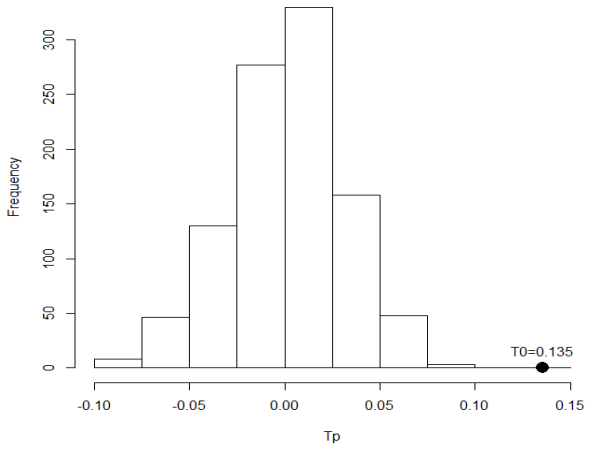


**Fig. 2.** Empirical distribution of statistic.

is lower than assumed significance level and amounts 0. Verified hypothesis $H_0$ should be rejected in favour of alternative hypothesis. There is the significance difference between the first canonical correlations so the dependency between socio-economic situation and satisfaction with life is stronger in 2000 than in 2015.

## Conclusion

The limitation of commonly used classical statistical methods makes the simulation methods being used to an increasing extent in a variety of analyses for both quantitative and qualitative data. Many classical statistical methods not have their counterparts for multidimensional data. The paper presents a method for testing the differences between sets of variables. In order to identify differences between the canonical correlation coefficients permutation test was proposed. The advantage of the proposed method is that the method can be used even when the required assumptions (e.g.: on the distribution of variables in the population) are not fulfilled. The procedure using the permutation test is used to estimate the distribution of the test statistics. The proposed method is illustrated by an empirical example.

## References

Baker, R. D. (1995). Two permutation tests of equality of variances. In: *Statistics and Computing*, 5, 289–296.

Basso, D., Pesarin, F., Salmaso, L., & Solari, A. (2009). *Permutation Tests for Stochastic Ordering and ANOVA*. Heidelberg: Springer Science + Business Media.

Domański, C. (1990). *Testy statystyczne*. Warszawa: Państwowe Wydawnictwo Ekonomiczne.

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. In: *Statist. Sci., 19*(4), 676–685.

González, I., Déjean, S., Martin, P. G., & Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. In: *Journal of Statistical Software*, *23*(12), 1–14.

Good, P. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. New York: Springer Science Business Media.

Good, P. (2006). *Resampling Methods. A Practical Guide to Data Analysis*. Boston–Basel–Berlin: Birkhauser.

Good, P. I. (1994). *Permutation Tests: A Practical Guide for Testing Hypotheses*. New York: Springer–Verlag.

Hotelling, H. (1936). Relations between two sets of variates. In: *Biometrika*, 28, 321–377.

Kończak, G. (2016). *Testy permutacyjne. Teoria i zastosowania*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

Kosiorowski, D. (2008). *Wstęp do wielowymiarowej analizy statystycznej zjawisk ekonomicznych. Kurs z wykorzystaniem środowiska R*. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Chichester: John Wiley & Sons, Ltd.

Pesarin, F., & Salmaso, L. P. (2010). *Permutation Tests for Complex Data. Theory, Applications and Software*. Chichester: John Wiley & Sons, Ltd.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Welch, W. J. (1990). Construction of Permutation Tests. In: *Journal of the American Statistical Association. Theory and Methods, 85*(411), 693–698.

Yamada, T., & Sugiyama T. (2006). On the permutation test in canonical correlation analysis. In: *Computational Statistics & Data Analysis, 50*(8), 2111–2123.