

| Dominika Polko-Zajęc

# A COMPARATIVE STUDY OF THE POWER OF PARAMETRIC AND PERMUTATION TESTS FOR A MULTIDIMENSIONAL TWO-SAMPLE LOCATION PROBLEM

## Abstract

*Objective:* A comparison of multidimensional populations is a very interesting and common statistical problem. It most often involves verifying a hypothesis about the equality of mean vectors in two populations. The classical test for verification of this hypothesis is the Hotelling's  $T^2$  test. Another solution is to use simulation and randomization methods to test the significance of differences between the studied populations. Permutation tests are to enable statistical inference in situations where it is not possible to use classical parametric tests. These tests are supposed to provide comparable power to parametric tests with a simultaneous reduction of assumptions, e.g. regarding the sample size taken or the distribution of the tested variable in the population. The purpose of this study is a comparative analysis of the parametric test, the (usual) permutation test, and the nonparametric permutation procedure using two-stage *ASL* determination.

*Research Design & Methods:* The study considered the analysis of multivariate data. The paper presents theoretical considerations and refers to the Monte Carlo simulation. *Findings:* The article presents a permutational, complex procedure for assessing the overall *ASL* (achieved significance level) value. The applied nonparametric statistical inference procedure uses combining functions. A simulation study was carried out to determine the size and power of the test under normality. A Monte Carlo simulation made it possible to compare the empirical power of this test with that of Hotelling's

| Dominika Polko-Zajęc, University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, Poland, e-mail: dominika.polko@ue.katowice.pl, ORCID: <https://orcid.org/0000-0003-4098-6647>.

| This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License (CC BY-NC-ND 4.0); <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

$T^2$  test. The most powerful test was the permutation test based on a two-stage *ASL* determination method using the Fisher combining function.

*Implications/Recommendations:* The advantage of the proposed method is that it can be used even when samples are taken from any type of continuous distributions in a population.

*Contribution:* The proposed test can be used in the analysis of multidimensional economic phenomena.

**Keywords:** permutation tests, comparing populations, power of test, Monte Carlo simulation, R software.

**JEL Classification:** C30, C150, C880.

## 1. Introduction

Population comparisons most often involve a comparison of characteristics in these populations. If it is assumed that population distributions differ only in a location, there are various parametric and nonparametric tests to verify this hypothesis. Many authors examine both the power and size of tests for the significance of differences between means or medians in two or more populations using for this purpose the simulation methods based on bootstrap or permutation tests (Janssen & Pauls 2005, Chang & Pal 2008, Kończak 2016, Anderson *et al.* 2017).

In a situation where a statistical test for certain measurable variables is conducted in several multidimensional populations, the hypothesis about an equality of mean vectors in these populations may need to be verified. A special case is the study of differences in means of variables  $^1X, ^2X, \dots, ^PX$  in two populations. The problem is to test the hypothesis about an equality of mean vectors of a  $P$ -dimensional random variable in the first and the second population, respectively, in the form of:

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \quad (1)$$

against the alternative hypothesis:

$$H_1: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \quad (2)$$

The parametric test for verification of this hypothesis (1) is Hotelling's  $T^2$  test. The method using the  $T^2$  test was proposed by Hotelling (1931, 1947) and Mahalanobis (1930, 1936) and is a generalisation of the Student's  $t$  test for many variables. To use the test, the assumption that the samples were taken from a population with multidimensional normal distributions is made (Rencher 2002).

In Hotelling's  $T^2$  test, two populations are considered from which two samples are taken independently from the distribution  $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  and from

the distribution  $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ . Assuming that covariance matrices are unknown but the same ( $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ ), in order to verify the null hypothesis (1) on the equality of the mean vectors, this statistic can be used:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \tag{3}$$

where:

$$\mathbf{S} = \frac{1}{n_1 + n_2 - 2} \left( \sum_{i=1}^{n_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)^T + \sum_{i=1}^{n_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)^T \right).$$

If the  $H_0$  hypothesis is true, the statistic (3) has a Hotelling's  $T^2$  distribution with  $P$  and  $n_1 + n_2 - 1$  degrees of freedom, where  $P$  is the number of variables (dimensions) examined and  $n_1, n_2$  are the sizes of samples taken from populations. It is also possible to determine the critical values for this statistic using a statistic of the form (Krzyśko 2009):

$$F = \frac{n_1 + n_2 - P - 1}{(n_1 + n_2 - 2)P} T^2, \tag{4}$$

which has a Snedecor's  $F$  distribution of  $P$  and  $n_1 + n_2 - P - 1$  degrees of freedom.

Hotelling's  $T^2$  test can only be used if variables in each population have a multidimensional normal distribution. The article presents a method for testing a difference between two vectors of mean values that can also be used when the assumption regarding the occurrence of a multidimensional normal distribution in populations is not met. A simulation, randomisation approach was proposed to investigate a significance of differences occurring between the studied populations. The aim of this research is to compare tests for the equality of mean vectors in two populations under multidimensional normality: the parametric test, the (usual) permutation test and the nonparametric permutation procedure using two-stage *ASL* (achieved significance level) determination. A simulation study to determine the size and power of the tests was carried out in the R statistical computing environment (R Core Team 2016).

## 2. Nonparametric Combination Procedures

It is assumed that there are two samples  ${}^1X_1, \dots, {}^pX_1, \dots, {}^pX_1$  and  ${}^1X_2, \dots, {}^pX_2, \dots, {}^pX_2$  independently taken from the population with distribution  $F_1$  and  $F_2$ . These populations have continuous,  $P$ -dimensional

distributions  $F_i$  for  $i = 1, 2$  with unknown parameters. A null hypothesis is verified claiming that two samples were taken from populations with identical distributions in the form of  $H_0: F_1(x) = F_2(x)$ . Data taken from two populations can be noted (Marozzi 2008).

$$\begin{aligned} \underline{\underline{X}} &= \begin{bmatrix} {}^1X_1 & {}^1X_2 \\ \vdots & \vdots \\ {}^pX_1 & {}^pX_2 \\ \vdots & \vdots \\ {}^pX_1 & {}^pX_2 \end{bmatrix} = \begin{bmatrix} {}^1X_{11} & \cdots & {}^1X_{1n_1} & {}^1X_{21} & \cdots & {}^1X_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^pX_{11} & \cdots & {}^pX_{1n_1} & {}^pX_{21} & \cdots & {}^pX_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^pX_{11} & \cdots & {}^pX_{1n_1} & {}^pX_{21} & \cdots & {}^pX_{2n_2} \end{bmatrix} = \\ &= \begin{bmatrix} {}^1X_1 & \cdots & {}^1X_{n_1} & {}^1X_{n_1+1} & \cdots & {}^1X_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^pX_1 & \cdots & {}^pX_{n_1} & {}^pX_{n_1+1} & \cdots & {}^pX_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^pX_1 & \cdots & {}^pX_{n_1} & {}^pX_{n_1+1} & \cdots & {}^pX_n \end{bmatrix} = \begin{bmatrix} \underline{{}^1X} \\ \vdots \\ \underline{{}^pX} \end{bmatrix}, \end{aligned}$$

where  ${}^pX_{ij}$  indicate the  $i$ -th ( $i=1, \dots, n_j$ ) sample for the  $p$ -th variable ( $p=1, \dots, P$ ) in the  $j$ -th ( $j=1, 2$ ) population and  $\underline{{}^pX}$  is the combined sample for the  $p$ -th variable.

The problem of testing equality of means in multidimensional populations can be considered in accordance with the Pesarin (2001) proposal. Pesarin initiated the approach to the nonparametric testing problem. He considered (Pesarin 2001) reducing the scope of the null hypothesis by splitting it into several partial hypotheses. This nonparametric approach is to perform some reasonable tests for each individual partial hypothesis and combine their results with a chosen combining function.

When the study deals with a problem of comparing the  $P$  means in two populations, partial  $P$  hypotheses are taken into account. The null hypothesis about the identity of the means vectors is in the form of:

$$H_0: \bigcap_{p=1}^P {}^p\mu_1 = {}^p\mu_2, \quad (5)$$

against the alternative hypothesis:

$$H_1: \bigcup_{p=1}^P {}^p\mu_1 \neq {}^p\mu_2. \quad (6)$$

The study considered test statistics in the form of:

$${}^pT = {}^p\bar{X}_1 - {}^p\bar{X}_2. \quad (7)$$

The decision was made using the empirical distribution of the test statistic obtained on the basis of permutation of the data set. A nonparametric, complex procedure was used to assess the overall *ASL* values. In the first stage of separate testing of each of the *P* partial hypotheses considered, the *ASL* values are determined in accordance with the usual permutation method used during verification of the hypothesis for one-dimensional data, i.e.:

1. The significance level  $\alpha$  is determined.
2. The statistics values are calculated on the basis of the sample data ( ${}^pT_0$ ).
3. Perform a permutation of data *N*-times, then calculate the statistics test value ( $T_k$ ).
4. Based on the empirical distribution of statistics, the *ASL* value for each of the compared variables is estimated according to the formula:

$$\hat{ASL}_{pT}({}^pT_0) = \frac{0,5 + \sum_{k=1}^N I(|{}^pT_k| \geq |{}^pT_0|)}{N+1} \tag{8}$$

The method of permutation of multidimensional data is shown in Figure 1.

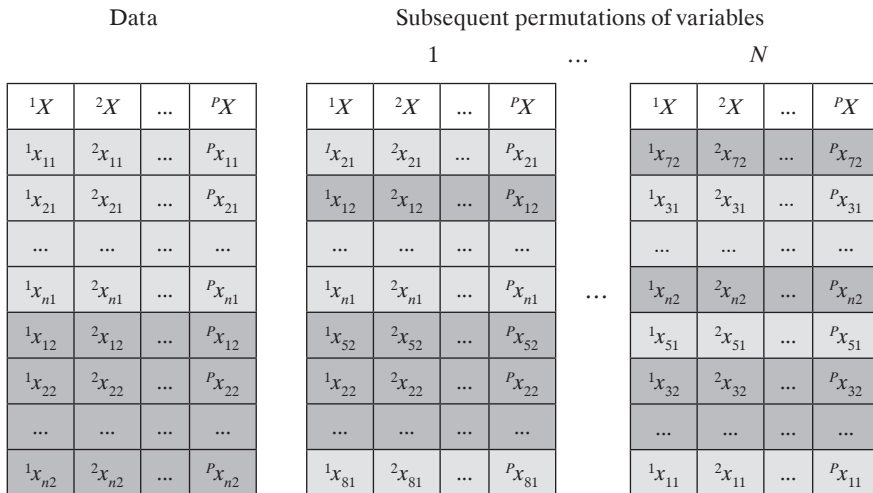


Fig. 1. Scheme of Permutations of Data

Source: author's own work.

The second stage of the nonparametric statistical inference procedure involves the determination of the overall  $ASL$  value using combining functions (Pesarin 2001):

$${}_{\varphi}T = \varphi(ASL_{1T}, \dots, ASL_{pT}).$$

There are many forms of combining functions for determining an overall  $ASL$  value. However, authors most often point to the following functions:

– the Fisher omnibus combining function (Fisher 1932):

$$C^{(F)} = -2 \cdot \sum_{p=1}^P \log(\hat{ASL}({}^pT)),$$

– the Liptak combining function (Liptak 1958):

$$C^{(L)} = \sum_{p=1}^P \Phi^{-1}(1 - \hat{ASL}({}^pT)),$$

where  $\Phi$  denotes the standard normal distribution function,

– the Tippett combining function (Tippett 1931):

$$C^{(T)} = \max\{1 - \hat{ASL}({}^1T), \dots, 1 - \hat{ASL}({}^pT)\}.$$

The observed statistic value for the sample data using Fisher combining functions can be determined as:

$$\underline{T}_0 = -2 \cdot \sum_{p=1}^P \log(\hat{ASL}_{pT}({}^pT_0)), \quad (9)$$

whereas the distribution of this statistic is determined on the basis of the same permutations as in the first step, for example for  $k$ -th permutation:

$$\underline{T}_k = -2 \cdot \sum_{p=1}^P \log(\hat{ASL}_{pT}({}^pT_k)). \quad (10)$$

The overall  $ASL$  value for the test under consideration is estimated using the formula:

$$\hat{ASL}_{\underline{T}} = \frac{\sum_{k=1}^N I(\underline{T}_k \geq \underline{T}_0)}{N}. \quad (11)$$

If  $ASL < \alpha$ , the hypothesis  $H_0$  is rejected, otherwise there is no basis for rejecting the  $H_0$  hypothesis.

### 3. Monte Carlo Simulation

Considering the nonparametric procedure based on the Fisher combining function, the size and power of the test were estimated by a simulation study. A Monte Carlo analysis was carried out allowing comparison of two populations with three-dimensional normal distributions with parameters:

$$\mu_1 = [0, 0, 0], \Sigma_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mu_2 = [x, x, x], \Sigma_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ where } x \in (-1, 1)$$

with the increment 0.2. In the simulations, samples of sizes  $(n_1, n_2) = (10, 10), (20, 20), (30, 30), (50, 50), (100, 100)$  were generated. The results of the simulations carried out to determine the size and power of the tests are presented in Table 1 (small sample sizes) and Table 2 (large sample sizes). For comparative purposes, the tables also include results obtained for the parametric Hotelling's  $T^2$  test and its permutation equivalent. A procedure for conducting each test included 1,000 Monte Carlo simulations and 1,000 permutations of data and the assumed level of significance was  $\alpha = 0.05$ .

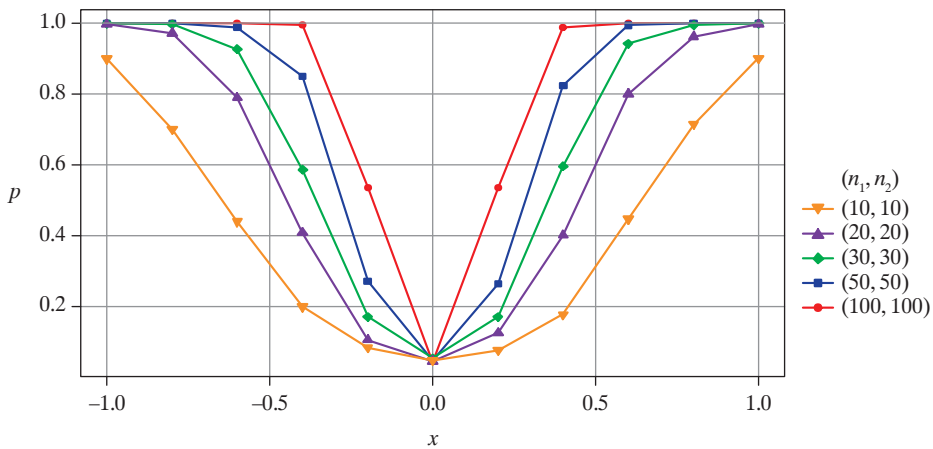


Fig. 2. Graphs of the Empirical Power of the Permutation Test  $\underline{T}$  for Different Sample Sizes

Source: author's own work in the R programme.

In the case of analysis of multidimensional, equinumerous samples, the sizes of the presented tests are close to the assumed level of significance. The values of estimated probabilities of rejecting the hypothesis  $H_0$ , when it was true only slightly differed from  $\alpha = 0.05$ . The three considered tests

Table 1. Hotelling's  $T^2$  Test Power and Estimation of Permutation Tests' Power (Small Sample Sizes)

Test Statistic			
$x$	$T^2$	$T^2 (perm)$	$\underline{T}$
(10, 10)			
-1.0	0.828	0.829	0.900
-0.8	0.632	0.628	0.701
-0.6	0.381	0.387	0.440
-0.4	0.187	0.193	0.199
-0.2	0.079	0.079	0.083
0	0.048	0.046	0.048
0.2	0.075	0.077	0.076
0.4	0.157	0.155	0.177
0.6	0.389	0.385	0.447
0.8	0.629	0.625	0.715
1.0	0.846	0.843	0.902
(20, 20)			
-1.0	0.996	0.995	0.998
-0.8	0.958	0.957	0.972
-0.6	0.747	0.753	0.790
-0.4	0.390	0.394	0.408
-0.2	0.102	0.105	0.105
0	0.045	0.043	0.045
0.2	0.117	0.119	0.125
0.4	0.373	0.380	0.401
0.6	0.760	0.759	0.801
0.8	0.947	0.948	0.963
1.0	0.996	0.996	0.998
(30, 30)			
-1.0	1.000	1.000	1.000
-0.8	0.998	0.997	0.998
-0.6	0.913	0.911	0.927
-0.4	0.548	0.556	0.586
-0.2	0.158	0.160	0.170
0	0.055	0.059	0.054
0.2	0.160	0.160	0.170
0.4	0.552	0.560	0.596



Table 1 cont'd

Test Statistic			
$x$	$T^2$	$T^2 (perm)$	$\underline{T}$
0.6	0.916	0.917	0.943
0.8	0.992	0.992	0.995
1.0	1.000	1.000	1.000

Source: computer simulations in the R programme.

Table 2. Hotelling's  $T^2$  Test Power and Estimation of Permutation Tests' Power (Large Sample Sizes)

Test Statistic			
$x$	$T^2$	$T^2 (perm)$	$\underline{T}$
(50, 50)			
-1.0	1.000	1.000	1.000
-0.8	1.000	1.000	1.000
-0.6	0.987	0.987	0.989
-0.4	0.835	0.836	0.850
-0.2	0.255	0.258	0.271
0	0.045	0.049	0.048
0.2	0.257	0.257	0.264
0.4	0.805	0.805	0.825
0.6	0.992	0.993	0.995
0.8	1.000	1.000	1.000
1.0	1.000	1.000	1.000
(100, 100)			
-1.0	1.000	1.000	1.000
-0.8	1.000	1.000	1.000
-0.6	1.000	1.000	1.000
-0.4	0.997	0.997	0.997
-0.2	0.528	0.528	0.536
0	0.044	0.044	0.041
0.2	0.520	0.528	0.535
0.4	0.986	0.986	0.988
0.6	1.000	1.000	1.000
0.8	1.000	1.000	1.000
1.0	1.000	1.000	1.000

Source: computer simulations in the R programme.

reached comparable assessments of the probabilities of rejecting the  $H_0$  hypothesis when it was false. In the majority of analysed cases, however, the most powerful test was the permutation test based on a two-stage *ASL* determination method using the Fisher combining function.

The probabilities of recognising differences between means vectors increased as the differences between the considered three-dimensional models of the populations increased. Analysing the graphs of the empirical power of the permutation test depending on the sample sizes taken from the populations (Figure 2), it can be seen that for 10 observations the differences in means at level 1 are detected with a probability of around 0.9 by the permutation test. For samples with 50 observations, this probability was obtained for the difference in means of around 0.5.

#### 4. Conclusions

The aim of the simulation research was to determine the ability of the presented permutation test to maintain the nominal probability of committing the type I error and the ability to obtain a high probability of rejecting a false null hypothesis in the conditions of changing distribution parameters in populations from which samples were taken. A simulation study to determine the size and power of the tests was carried out.

The results obtained in the simulation confirm the effectiveness of the permutation procedure and the possibility of its application in order to infer differences between vectors of means in two populations with multidimensional normal distributions. All testing procedures (under normality) ensured control of the type I error at the assumed level of significance. The higher power of the presented tests was achieved thanks to the use of a nonparametric combination procedure that uses Fisher's combining functions to evaluate the overall *ASL* value. The advantage of the presented method is that the method can be used even when samples are taken from any type of continuous distributions in a population. In the further research other forms of combining functions can be considered and a simulation study that takes into account various distributions of the studied variables can be performed.

#### Bibliography

Anderson, M. J., Walsh, D. C. I., Clarke, K. R., Gorley, R. N., Guerra-Castro, E. (2017) "Permutational Multivariate Analysis of Variance (PERMANOVA)". *Statistics Reference Online*: 1–15, <https://doi.org/10.1002/9781118445112.stat07841>.

- Chang, C. H., Pal, N. (2008) "A Revisit to the Behrens-Fisher Problem: Comparison of Five Test Methods". *Communications in Statistics – Simulation and Computation* 37(6): 1064–85, <https://doi.org/10.1080/03610910802049599>.
- Fisher, R. A. (1932) *Statistical Methods for Research Workers*. 4 ed. Edinburgh: Oliver & Boyd.
- Hotelling, H. (1931) "The Generalization of Student's Ratio". *Annals of Mathematical Statistics* 2(3): 360–78, <https://doi.org/10.1214/aoms/1177732979>.
- Hotelling, H. (1947) "Multivariate Quality Control" in C. Eisenhart, M. W. Hastay, W. A. Wallis (eds) *Techniques of Statistical Analysis*. New York: McGraw-Hill.
- Janssen, A., Pauls, T. (2005) "A Monte Carlo Comparison of Studentized Bootstrap and Permutation Tests for Heteroscedastic Two-sample Problems". *Computational Statistics* 20(3): 369–83, <https://doi.org/10.1007/bf02741303>.
- Kończak, G. (2016), *Testy permutacyjne. Teoria i zastosowania*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Krzyżsko, M. (2009) *Podstawy wielowymiarowego wnioskowania statystycznego*. Poznań: Wydawnictwo Naukowe UAM.
- Liptak, I. (1958) "On the Combination of Independent Tests". *Magyar Tudományok Akademia Matematikai Kutató Intézetének Közleményei* 3: 127–41.
- Mahalanobis, P. C. (1930) "On Tests and Measures of Group Divergence". *Journal of the Asiatic Society of Bengal* 26: 541–88.
- Mahalanobis, P. C. (1936) "On the Generalized Distance in Statistics". *Proceedings of the National Institute of Science of India* 12: 49–55, <https://doi.org/10.1007/s13171-019-00164-5>.
- Marozzi, M. (2008) "The Lepage Location-scale Test Revisited". *Far East Journal of Theoretical Statistics* 24: 137–55.
- Pesarin, F. (2001) *Multivariate Permutation Test with Applications in Biostatistics*. Chichester: Wiley.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, <https://www.R-project.org/>.
- Rencher, A. C. (2002) *Methods of Multivariate Analysis*. New York: John Wiley & Sons.
- Tippett, L. H. C. (1931) *The Methods of Statistics*. London: Williams and Norgate.