CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open access journal indexed by

Volume 11 Number 2
December 2020

ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society

SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Contents

Chilean Journal of Statistics
Vol. 11, No. 2, December 2020, 117–136

Multivariate Statistical Inference
Research Paper

On new robust tests for the multivariate normal mean vector with high-dimensional data and applications

Henrique José de Paula Alves[1,*] and Daniel Furtado Ferreira[1]

[1]Department of Statistics, Federal University of Lavras, Lavras, Brazil

(Received: 04 August 2020 · Accepted in final form: 10 December 2020)

Abstract

New alternative tests to the Hotelling T2 and the likelihood ratio tests for the multivariate normal and non-normal population mean vector are proposed here. These new tests are based on the ordinary and robust comedian covariance matrix estimator. The new adapted likelihood ratio test overcomes the high dimensional issue that occurs with both T2 and likelihood ratio tests. The asymptotic and parametric bootstrap distributions for test statistics are used and the performance of these new tests based on normal and non-normal distributions is evaluated through Monte Carlo simulations. Contaminated normal multivariate populations are also considered to evaluate the effects of outliers on test performances. Type I error probabilities and power in all simulations are computed using the R software. The non-robust parametric bootstrap version of the likelihood ratio test performs better and is recommended since it is easy to implement and computationally fast. An application of the proposed new and T2 tests to a real data set is provided. We use an R package of our authorship to perform the tests described here.

Keywords: Bootstrapping · Hotelling and likelihood ratio tests · Types I-II errors.

Mathematics Subject Classification: Primary 62G10 · Secondary 62F05.

1. Introduction

A big challenge in statistics is based on verifying if a p-variate normal mean vector $\boldsymbol{\mu}$ is equal to a known vector $\boldsymbol{\mu}_0$ when the dimensionality p is greater than the sample size n. The Hotelling T^2 test is widely used to test the null hypothesis. However, under non-normal distributions or in the presence of outliers, the use of the Hotelling T^2 test is not recommended. First, this statistic is built under multivariate normality. Second, even under normality, this statistic considers the average sample vector $\overline{\boldsymbol{X}}$ and the covariance matrix $\boldsymbol{S}$ that are strongly influenced by outliers, as in the univariate case (Willems et al., 2002). Third, T^2 cannot be calculated when the number of variables p is greater than or equal to the number of observations n, since the sample covariance matrix, that is present in this statistic, is a singular matrix (Bai and Saranadasa, 1996). In addition, Bai and Saranadasa (1996) noted that the test power based on the T^2 statistic has low power when under these same conditions, as also shown in Pan and Zhou (2011).

*Corresponding author. Email: jpahenrique@gmail.com

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Another widely used statistical method is the likelihood ratio (LR) test (Ferreira, 2018; Wagala, A., 2020). In addition, testing the null hypothesis on a vector of population means becomes a challenge under non-normal asymmetric distributions or in the presence of outliers. The Hotelling or likelihood ratio tests consider the sample estimators of the vector of means and the covariance matrix in their expressions, which are highly influenced by outliers. Some robust testing proposals can be found by Tiku (1982); Mudholkar and Srivastava (2000); Willems et al. (2002). In contrast, Srivastava and Du (2008); Srivastava (2009); Chen et al. (2010); Lee et al. (2012); Srivastava et al. (2013); Wang et al. (2013); Marozzi (2015) proposed alternative nonparametric tests for the LR test in non-normal populations, the number of variables p is greater than or equal to the number of observations n.

In this article, new statistical tests are proposed for the null hypothesis involving tests on the vector of multivariate population means. The idea is to obtain robust adaptations of the T^2 and LR statistics using robust comedian estimators (Sajesh and Srinivasan, 2012) for the vector of averages and population covariance matrix. The fundamental concept is to replace ordinary estimators with the mean vector and covariance matrix with their respective robust comedian estimators and provide accurate tests for the mean vector considering the parametric bootstrap distribution under the null hypothesis. In addition, for the LR test statistic with the original and robust comedian estimator of the covariance matrix, the determinants (generalized variances) are replaced by the trace operator, which represents the total sample variance. These new tests are potentially more advantageous than the adapted tests mentioned above, since they can perform better under non-normality and in the presence of outliers. In addition, they are computationally easy to implement and apply.

The performance of these proposed new tests is evaluated by Monte Carlo simulations calculating the type I error probabilities and power of the tests. In Section 2, the new proposed tests are introduced. The results regarding the type I error probability and power of the tests are shown in Section 3. The exact binomial test proposed by Cardoso de Oliveira and Ferreira (2010) is evaluated by Monte Carlo simulations. Section 4 applies the results obtained in this work to real data. In Section 5, the conclusions are presented.

2. Methods

2.1 General context

Consider the problem of testing the hypotheses given by

$$\mathrm{H}_0\colon \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad \mathrm{H}_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu}_0. \tag{1}$$

In order to do this, let $\boldsymbol{X}_j = [X_{\mathrm{j}1}, \ldots, X_{jp}]^\top$, with $j = 1, \ldots, n$, be a random sample of size n from a p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Here, n refers to the number of observations, and p refers to the number of variables (dimensionality) in each random vector. In general, the p components in the random vectors are correlated variables, where its $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ is positive definite.

Under the null hypothesis H_0 as in Equation (1) the test statistic Hotelling T^2 is given by

$$T_{\mathrm{c}}^2 = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0). \tag{2}$$

where $\overline{\boldsymbol{X}} = \sum_{\mathrm{j}=1}^{n} \boldsymbol{X}_{\mathrm{j}}/n$ is the sample mean vector, $\boldsymbol{S} = (1/(n-1)) \sum_{\mathrm{j}=1}^{n} (\boldsymbol{X}_{\mathrm{j}} - \overline{\boldsymbol{X}})(\boldsymbol{X}_{\mathrm{j}} - \overline{\boldsymbol{X}})^\top$ is the sample covariance matrix, and n is the sample size. Under H_0 and with the assumption of normality and homoscedastic covariance matrix, the T_c^2 given in Equation (2) follows a Hotelling T^2 distribution given by $(n-1)pF_{p,n-p}/(n-p)$, where $F_{p,n-p}$ is the F distribution with p and $n-p$ degrees of freedom.

Considering the hypotheses given in Equation (1), another statistical method used is the LR test. Let $\boldsymbol{X} \sim \mathrm{N}_{\mathrm{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is unknown. Therefore, the LR statistic is given by the expression stated as

$$-2\log(\Lambda) = n[\log(|\boldsymbol{S} + \boldsymbol{H}|) - \log(|\boldsymbol{S}|)], \tag{3}$$

where $\boldsymbol{H} = (\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top$ and log is the natural logarithm. Consider $\Omega \in \mathrm{R}^s$ the unrestricted parametric space and $\Omega_0 \subseteq \mathrm{R}^r$ the restricted parametric space, with $\Omega_0 \subset \Omega$. In general, under certain conditions of regularity, Equation (3) follows an asymptotic chi-square distribution with $r - s$ degrees of freedom (Ferreira, 2018) under the null hypothesis H_0. Thus, the rejection region of H_0 is given by $\mathrm{R} = \{\boldsymbol{x} | -2\log(\Lambda(\boldsymbol{x})) > \chi^2_{1-\alpha}(r-s)\}$, where α is the nominal significance level and $\chi^2_{1-\alpha}(r-s)$ is the $100(1-\alpha)\%$ percentile of a chi-square distribution with $r - s$ degrees of freedom. In this case, for testing hypothesis about normal mean vector, the degrees of freedom $r - s$ are equal to p.

Here, the proposed new tests based on the modifications of the T^2 statistic defined in Equation (2) and LR statistic stated in Equation (3) are shown. Also, the adopted Monte Carlo simulation procedure to assess their performance is described. For this, consider a sample of size n from the normal p-variate distribution with a mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ to test the null hypothesis H_0 given in Equation (1). In all cases, the original and transformed LR expressed in Equation (3) are applied, consider the trace operator replacing the determinant operator and the robust estimator replacing the traditional estimator of the covariance matrix. Only in the cases where $p < n$, the T^2 test and its modifications that use the comedian estimators. The theoretical justification can be seen in Section 1. To evaluate test performance, first, type I error probabilities are calculated by generating sample sizes from populations under H_0, with a mean vector of $\boldsymbol{\mu}_0$. Second, random samples are generated under H_1, with $\boldsymbol{\mu} \neq \boldsymbol{\mu_0}$. In both cases, samples from normal and non-normal populations are generated. The p-variate Student-t distributions with 5 degrees of freedom for the non-normal distribution case. We also consider contaminated normal (CN) populations for generating outliers. Some factorial combinations of the number of variables p and sample size n are considered.

Without loss of generality, the population covariance matrix $\boldsymbol{\Sigma}$ with the compound symmetry structure given by

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} = \sigma^2[(1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}] \tag{4}$$

is considered, where $\boldsymbol{J}$ is a $p \times p$ matrix with all entries equal to 1 and $\boldsymbol{I}$ is an identity array of the same order as $\boldsymbol{J}$. Also, without loss of generality, $\sigma^2 = 1$ and $\rho = 0.9$ since the test statistics are invariant under the true covariance structure. The p-variate Student-t distribution is used to pick a distribution that has heavier tails than the multivariate normal distribution and violate the assumptions of the T^2 and LR statistics. The p-variate CN distribution is $\omega\mathrm{N}_1(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1) + (1-\omega)\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$. Again, without loss of generality, $\omega = 0.9$, $\boldsymbol{\Sigma}_1$ is defined by Equation (4) and $\boldsymbol{\Sigma}_2$ is constructed using the constraint: $|\boldsymbol{\Sigma}_2|/|\boldsymbol{\Sigma}_1| = \Delta$, and thus, $\boldsymbol{\Sigma}_2 = \Delta^{1/p}\boldsymbol{\Sigma}_1$, where $\Delta = 2$. Sample sizes are $n = 10, 50, 70, 100$ and 200 and the nominal significance level α is $\alpha = 5\%$. The number of variables is $p = 2, 5$ and 200 and $2\,000$ Monte Carlo simulations to evaluate the empirical estimates of the type I error probabilities and power of each test are considered. The parametric bootstrap null distribution are generated with $2\,000$ resamples from a $\mathrm{N}(\boldsymbol{0}, \boldsymbol{S}^\bullet)$ distribution, where the null hypothesis H_0 is imposed by considering $\boldsymbol{\mu} = \boldsymbol{0}$ to generate the null distribution of the statistic. Also, the covariance matrix used to generate the bootstrap null distribution, given $\boldsymbol{S}^{\bullet}$, is the sample covariance matrix computed by using the traditional or robust comedian estimator in the original sample for traditional and robust bootstrap tests, respectively.

Without loss of generality, to evaluate the type I error probabilities the $\boldsymbol{\mu}_0$ vector under the null hypothesis given in Equation (1) is the p-dimensional $\mathbf{0}$ null vector and the true vector of population mean is also $\boldsymbol{\mu} = \mathbf{0}$. Under the alternative hypothesis H_1 stated in Equation (1), for the power study, the true population mean vector $\boldsymbol{\mu}$ is chosen considering a fixed generalized Mahalanobis distance $\delta(\boldsymbol{\mu}, \boldsymbol{\mu}_0)$ between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_0$ given by

$$\delta(\boldsymbol{\mu}, \boldsymbol{\mu}_0) = n(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0). \tag{5}$$

In this case, $\boldsymbol{\mu}_0 = \mathbf{0}$, and the true population mean vector is calculated by trial and error in Equation (5) by taken a fixed δ value and the final value is used as a parameter in each of the population distributions considered under H_1. The chosen values from δ are $0, 0.5, 1, 1.5, 3, 5, 10$. Therefore, since the values of the mean vector change as n changes, keeping fixed the value of the distance of Mahalanobis $\delta(\boldsymbol{\mu}, \boldsymbol{\mu}_0) = n(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$, the power does not change as n increases.

Thus, five new tests based on the traditional T^2 statistic defined in Equation (2) and in the LR statistic defined in Equation (3) are proposed, including the traditional and robust versions that use the comedian mean vector and covariance matrix estimators (Falk, 1997; Maronna and Zamar, 2002; Sajesh and Srinivasan, 2012). Some tests are based on parametric bootstrap versions as well as the asymptotic chi-square distribution. However, some asymptotic chi-square tests have not been shown since they did not control the type I error probabilities. The performance of these new tests is evaluated by Monte Carlo simulations. Below, each of the proposed new tests for testing the null hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is described. The new LR test has a chi-square asymptotic distribution with p degrees of freedom, as the original LR test.

One special case is considered regarding the distributions and for some values of n, p and δ. A shifted zero mean exponential distribution with parameter $\boldsymbol{\lambda} = \mathbf{1}_\mathrm{p}$, where $\mathbf{1}_\mathrm{p}$ is a p-dimensional vector with 1 in all entries. The latter is a case of a skewed distribution. For the exponential distribution, a p-dimension random vector $\boldsymbol{Z}$ is generated from a $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. A p-dimensional random vector $\boldsymbol{Y}$ from this distribution is obtained considering for the ith entry the random variable stated as $Y_i = F^{-1}(\Phi(\mathrm{Z}_i); \lambda_i) - 1/\lambda_i$, for $i = 1, \ldots, p$, where $F^{-1}(\mathrm{x}; \lambda_i)$ is the quantile function of the exponential distribution of parameter λ_i evaluated at x and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution evaluated at x.

2.2 The parametric bootstrap T^2 test

We construct the parametric bootstrap T^2 test, called T^2_{PB} (T2PB), where PB stands for parametric bootstrap, adopting the following steps:

(1) From the original sample, the parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are estimated, respectively, by $\boldsymbol{S}^*$ and $\overline{\boldsymbol{X}}^*$, where $\boldsymbol{S}^*$ and $\overline{\boldsymbol{X}}^*$ are the traditional sample covariance matrix and vector mean, respectively. The test statistic is computed by

$$T^{*2} = n\left(\overline{\boldsymbol{X}}^* - \boldsymbol{\mu}_0\right)^\top \boldsymbol{S}^{*-1}\left(\overline{\boldsymbol{X}}^* - \boldsymbol{\mu}_0\right). \tag{6}$$

(2) By using the original covariance estimates $\boldsymbol{S}^*$, a random sample of size n is generated from a p-variate normal distribution imposing H_0, that is, by setting $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. Also $\boldsymbol{\Sigma} = \boldsymbol{S}^*$. Therefore, a sample of size n is generated from a $\mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{S}^*)$ distribution.

(3) In each parametric bootstrap sample, the sample mean $\overline{\boldsymbol{X}}_{\rm PB}$ and the sample covariance matrix $\boldsymbol{S}_{\rm PB}$ are estimated.
(4) In each parametric bootstrap sample, compute the test statistic by means of

$$T^2_{\rm PB} = n(\overline{\boldsymbol{X}}_{\rm PB} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}_{\rm PB}(\overline{\boldsymbol{X}}_{\rm PB} - \boldsymbol{\mu}_0). \tag{7}$$

(5) Steps (2) to (4) are repeated B times and a set of size $B+1$ is constructed with the test statistic values computed in Equation (7) and the original value calculated in Equation (6). The null distribution of the parametric bootstrap test is constituted by this set. Therefore, if the ith member of this set is represented by T^2_i, for $i = 1, \ldots, B+1$, then the p-value is computed by

$$p\text{-value} = \frac{\sum_{i=1}^{B+1} I(T^2_i \geq T^{*2})}{B+1}, \tag{8}$$

where $I(T^2_i \geq T^{*2})$ is the indicator function.
(6) The null hypothesis given in Equation (1) is rejected at the significance level α if the p-value defined in Equation (8) is less than α.

Note that the traditional T^2 test is also considered, by computing the p-value direct from the Hotelling T^2 distribution of the test statistic value obtained in Equation (6). This is named by T2 and is considered the benchmark test.

2.3 The robust parametric bootstrap T^2 test

The robust parametric bootstrap T^2 test, called $T^2_{\rm RPB}$ (T2RPB), in which RPB stands for robust parametric bootstrap, are performed by adopting the same steps described for the previous test. However, some of them are modified as in the following sequence.

In Step 1, the estimators $\boldsymbol{S}^*$ and $\overline{\boldsymbol{X}}^*$ are replaced by comedian estimators $\boldsymbol{S}_{\rm R}$ and $\overline{\boldsymbol{X}}_{\rm R}$ the test statistic in the original sample is computed by $T^{*2} = n(\overline{\boldsymbol{X}}_{\rm R} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}_{\rm R}(\overline{\boldsymbol{X}}_{\rm R} - \boldsymbol{\mu}_0)$.

In Step 2, the sample of size n is generated from a $\mathrm{N}(\boldsymbol{\mu}_0,\, \boldsymbol{S}_{\rm R})$ distribution, where again, the null hypothesis is imposed by considering the multivariate normal mean equal to $\boldsymbol{\mu}_0$, the null value of the population mean.

In Step 3, the mean and the sample covariance in each parametric bootstrap sample are denoted, respectively, by $\overline{\boldsymbol{X}}_{\rm RPB}$ and $\boldsymbol{S}_{\rm RPB}$.

In Step 4, the test statistic is computed by $T^2_{\rm RPB} = n(\overline{\boldsymbol{X}}_{\rm RPB} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}_{\rm RPB}(\overline{\boldsymbol{X}}_{\rm RPB} - \boldsymbol{\mu}_0)$.

Steps 5 and 6 are identically as described in the previous test, with T^2_i replaced now by the ith value from the bootstrap null distribution of $T^2_{\rm RPB}$. In this case, the asymptotic chi-square distribution with p degrees of freedom is not considered as an alternative test, since the corresponding robust T^2 test did not control the type I error probability (results omitted here). More details on the performance of the above tests can be seen in Alves and Ferreira (2019).

We have shown in Section 1, in addition to the problems presented for their use in data following non-normal distributions, that the traditional T^2 test is not valid for high dimensional data ($p > n$) due to the singularity of the sample covariance matrix $\boldsymbol{S}$. The LR test has the same limitations of the T^2 to be implemented in high dimensional data sets and non-normal circumstances. Considering to Ledoit and Wolf (2002) as reference, we propose an alternative test to the LR test that is based on replacing the determinants of the matrices $\boldsymbol{S}$ and $\boldsymbol{S} + \boldsymbol{H}$ for their respective traces. Here, $\boldsymbol{H} = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top$. In this way, we have obtained a new test that applies to high-dimensional ($p > n$) data sets and that maintains the same distributional properties of LR test. The validity of the asymptotic null distribution of the new test with traditional sample estimators of the mean vector and covariance matrix is evaluated by Monte Carlo simulations. Also, the parametric bootstrap and robust parametric bootstrap versions for this latest proposal are built, as we have done for the T^2 test. In the following subsections, we present the new procedures.

2.4 The asymptotic LR trace test

The asymptotic trace version of the LR test, named asymptotic trace LR (ATLR) test, is obtained by directly replacing the determinant given in the expression of the LR by the trace operator tr. Let the null hypothesis be H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, then the ATLR test statistic is $T^2_{\mathrm{ATLR}} = n\{\log[\mathrm{tr}(\boldsymbol{S}^* + \boldsymbol{H})] - \mathrm{tr}(\boldsymbol{S}^*)\}$, where $\boldsymbol{H} = (\overline{\boldsymbol{X}}^* - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}}^* - \boldsymbol{\mu}_0)^\top$, that under the null hypothesis H_0 and normality has a chi-square distribution with p degrees of freedom, since the plim (probability limit) of $\boldsymbol{H}$ is $\boldsymbol{0}$ as $n \to \infty$, still under the null H_0. The null hypothesis should be rejected if the $T^2_{\mathrm{ATLR}} \geq \chi^2_{1-\alpha}(p)$.

2.5 The TLR parametric bootstrap test

There is no guarantee the T^2_{ATLR} has an asymptotic chi-square distribution with p degrees of freedom under H_0 and multivariate normality. To overcome this issue we proposed the TLR parametric bootstrap test, named trace likelihood ratio parametric bootstrap (TLRPB) test. The steps to apply for this test are the same as described previously for T^2_{PB}, except for some details explained as follows.

In Step 1, the estimators $\boldsymbol{S}^*$ and $\overline{\boldsymbol{X}}^*$ are computed in the original sample and the test statistic is $T^{*2} = n\{\log[\mathrm{tr}(\boldsymbol{S}^* + \boldsymbol{H}^*)] - \mathrm{tr}(\boldsymbol{S}^*)\}$, where $\boldsymbol{H}^* = (\overline{\boldsymbol{X}}^* - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}}^* - \boldsymbol{\mu}_0)^\top$.

In Step 2, the sample of size n is generated from a $\mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{S}^*)$ distribution, where the null hypothesis is imposed by considering the multivariate normal mean equal to $\boldsymbol{\mu}_0$, the null value of the population mean.

In Step 3, the mean and the sample covariance in each parametric bootstrap sample are denoted respectively by $\overline{\boldsymbol{X}}_{\mathrm{TLRPB}}$ and $\boldsymbol{S}_{\mathrm{TLRPB}}$. In Step 4, the test statistic is $T^2_{\mathrm{TLRPB}} = n\{\log[\mathrm{tr}(\boldsymbol{S}_{\mathrm{TLRPB}} + \boldsymbol{H}_{\mathrm{TLRPB}})] - \mathrm{tr}(\boldsymbol{S}_{\mathrm{TLRPB}})\}$, with $\boldsymbol{H}_{\mathrm{TLRPB}} = (\overline{\boldsymbol{X}}_{\mathrm{TLRPB}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}}_{\mathrm{TLRPB}} - \boldsymbol{\mu}_0)^\top$.

Steps 5 and 6 are identically as described in the previous test, with T^2_i replaced now by the ith value from the bootstrap null distribution of T^2_{TLRPB}.

2.6 The robust TLR parametric bootstrap test

For overcoming problems with outliers the robust parametric bootstrap version of the previous TLRPB, called robust trace likelihood ratio parametric bootstrap (RTLRPB) is constructed. The steps necessary for this test to be applied are the same as the previous steps described for the T^2_{PB}, except for some details explained below.

In Step 1, the comedian estimators $\boldsymbol{S}_{\mathrm{R}}$ and $\overline{\boldsymbol{X}}_{\mathrm{R}}$ are computed in the original sample and the test statistic is $T^{*2} = n\{\log[\mathrm{tr}(\boldsymbol{S}_{\mathrm{R}} + \boldsymbol{H}_{\mathrm{R}})] - \mathrm{tr}(\boldsymbol{S}_{\mathrm{R}})\}$, where $\boldsymbol{H}_{\mathrm{R}} = (\overline{\boldsymbol{X}}_{\mathrm{R}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}}_{\mathrm{R}} - \boldsymbol{\mu}_0)^\top$.

In Step 2, the sample of size n is generated from the $\mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{S}_{\mathrm{R}})$ null distribution.

In Step 3, the comedian sample mean and sample covariance in each parametric bootstrap sample are denoted respectively by $\overline{\boldsymbol{X}}_{\mathrm{RTLRPB}}$ and $\boldsymbol{S}_{\mathrm{RTLRPB}}$.

In Step 4, the test statistic is $T^2_{\mathrm{RTLRPB}} = n\{\log[\mathrm{tr}(\boldsymbol{S}_{\mathrm{RTLRPB}} + \boldsymbol{H}_{\mathrm{RTLRPB}})] - \mathrm{tr}(\boldsymbol{S}_{\mathrm{RTLRPB}})\}$, where $\boldsymbol{H}_{\mathrm{RTLRPB}} = (\overline{\boldsymbol{X}}_{\mathrm{RTLRPB}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}}_{\mathrm{RTLRPB}} - \boldsymbol{\mu}_0)^\top$.

Steps 5 and 6 are identically as described in the previous test, with T^2_i replaced now by the ith value from the bootstrap null distribution of T^2_{RTLRPB}.

2.7 The exact binomial test

The test type I error probabilities are evaluated by Monte Carlo simulations, and according to Cardoso de Oliveira and Ferreira (2010), these estimates are not error-free. Therefore, an exact binomial test is used to decide whether each of the modified or the original statistical test is considered accurate, liberal or conservative. In this sense, considering a nominal level of significance of 1%, the hypotheses to be tested are defined as

$$\mathrm{H}_0\colon \alpha = 5\% \quad \text{versus} \quad \mathrm{H}_1\colon \alpha \neq 5\%. \tag{9}$$

The statistic of the exact binomial test is given by

$$F_{\mathrm{c}} = \left(\frac{z+1}{N-z}\right)\left(\frac{1-\alpha}{\alpha}\right), \tag{10}$$

where z is the number of rejection of the null hypothesis accounted by one of the tests considering the nominal significance level of α and N is the number of Monte Carlo simulations performed. Under the null hypothesis defined in Equation (9), the F_{c} statistic defined in (10) follows a F distribution with $\nu_1 = 2(N-z)$ and $\nu_2 = 2(z+1)$ degrees of freedom. If the null hypothesis is rejected and the type I error probability is considered significantly less than the nominal level adopted of the 5%, the test can be considered conservative; if the null hypothesis is rejected and the type I error is considered significantly higher than the nominal level adopted of the 5%, the test can be considered liberal; and if the null hypothesis is not rejected the test can be considered accurate.

A computer with a Core-I7 processor with 4 cores and 8 GB of RAM is used. The simulations are performed in `R` with functions developed by the authors, except for the `mvrnomr`, `var`, `pf`, `pchisq`, `ginv` and `covComed` functions of the `MASS`, `statistics` and `robustbase` packages. The execution time of each simulation is 12 hours on average considering small sample sizes (10, 50, and 70) and, in the case of larger sample sizes (100, and 200), the average duration of the simulations is 2 days, regardless of the dimension considered.

3. Monte Carlo simulations

3.1 General context

The performance results for the proposed new tests are presented in two stages. First, the results regarding type I error probability control and power for cases where $p = 2$ and $p = 5$ where shown. Second, the results for the special case of high dimension ($p = 200$) are shown. The performance of these new tests is evaluated considering the multivariate normal, Student-t with 5 degrees of freedom and CN distributions.

3.2 Type I error probabilities

The type I error probabilities for the five new tests proposed via Monte Carlo simulations are shown in Table 1, considering the dimension $p = 2$ at the significance level of $\alpha = 0.05$. The exact binomial test is used to classify these tests as exact, liberal, or conservative (see Section 2). The traditional and ordinary Hotelling T^2 test is also applied in each circumstance and it is invoked as a benchmark test. We note that for all n sample sizes considered, as well as for all evaluated multivariate distributions, the proposed tests T2, T2PB, and T2RPB are exact since they showed test size equal to the nominal significance level α. The traditional T^2 test constituted an exception when considering the bivariate Normal distribution with a sample size of 50. In this case, this test is conservative but still acceptable. The same is not the case for LR adaptations. The ATLR is conservative on all distributions and sample sizes considered. For this test, a substantial loss of power is expected to occur, which is an important fact to be taken into account. The TLRPB test and RTLRPB test (see Section 2), in all the evaluated scenarios, are all accurate. A single exception occurs for the TLRPB test, considering the multivariate normal distribution with $n = 50$, where it shows a liberal performance in the control of the type I error probability. In practice, a test is considered reliable if it has an exact size. Otherwise, if it is conservative, it can be considered acceptable. However, if it is a liberal test, then it must be discarded. It does not appear the case for the TLRPB test, once it showed a unique exception. Therefore, since in circumstances where the normality assumption is violated, the performance regarding the type I error probability control of the proposed new tests is acceptable.

Table 1. Type I error probabilities of the six tests with $\alpha = 5\%$ and p = 2, considering the multivariate normal (N), Student-t with 5 degrees of freedom (t_5) and CN distributions.

Test	Model	n 10	50	70	100	200
T2	N	0.0525^-	0.0380^-	0.0505	0.0485	0.0460
	t_5	0.0450	0.0495	0.0420	0.0485	0.0510
	CN	0.0520	0.0450	0.0420	0.0490	0.0570
T2PB	N	0.0510	0.0475	0.0455	0.0485	0.0475
	t_5	0.0510	0.0430	0.0405	0.0485	0.0560
	CN	0.0535	0.0435	0.0440	0.0430	0.0515
T2RPB	N	0.0430	0.0480	0.0525	0.0450	0.0535
	t_5	0.0430	0.0390	0.0520	0.0500	0.0485
	CN	0.0530	0.0455	0.0525	0.0490	0.0615
ATLR	N	0.0195^-	0.0085^-	0.0120^-	0.0110^-	0.0130^-
	t_5	0.0275^-	0.0085^-	0.0105^-	0.0140^-	0.0145^-
	CN	0.0200^-	0.0125^-	0.0120^-	0.0140^-	0.0120^-
TLRPB test	N	0.0535	0.0940^+	0.0555	0.0445	0.0535
	t_5	0.0500	0.0680	0.0460	0.0520	0.0580
	CN	0.0515	0.0520	0.0485	0.0445	0.0470
RTLPBT	N	0.0560	0.0490	0.0445	0.0430	0.0530
	t_5	0.0415	0.0415	0.0555	0.0510	0.0555
	CN	0.0550	0.0505	0.0480	0.0435	0.0570

$^-$: significantly ($p < 0.01$) less than the nominal significance level of 5%.

To evaluate if the patterns presented in Table 1 are maintained we decided to increase the dimensionality p and keep the sample sizes n fixed. Table 2 presents the results for the empirical type I error probabilities considering the dimension $p = 5$. Similar behavioral pattern to the control of the type I error probabilities showed for the case with $p = 2$ holds. The atypical situation (conservative) for the T^2 test no longer occurs. We note that the T2RPB for the multivariate Student-t distribution with 5 degrees of freedom and $n = 10$ and 200 is conservative in this case. The TLRPB test and RTLRPB test are considered exact tests. The ATLR remains conservative and therefore acceptable. Likewise, it expects that this test substantially loses power for this dimension. Next, we present the results of the power of the tests for these same cases.

3.3 Power

The powerful performance of the proposed new tests is evaluated in the same cases used to evaluate type I error probabilities: distributions, sample sizes, and dimensions. In all circumstances, the power curves are plotted against Mahalanobis distances (δ), given in 5, between the true population vector $\boldsymbol{\mu}$ and the hypothetical mean vector $\boldsymbol{\mu_0}$. These distances have been fixed (see Section 2) to establish the true value of the population mean vector.

Table 2. Type I error probabilities of the six tests with $\alpha = 5\%$ and p = 5, considering the multivariate Normal (N), Student-t with 5 degrees of freedom (t_5) and CN distributions.

Test	Model	n = 10	50	70	100	200
T2	N	0.0525	0.0545	0.0430	0.0500	0.0565
	t_5	0.0420	0.0545	0.0435	0.0440	0.0445
	CN	0.0475	0.0535	0.0465	0.0475	0.0520
T2PB	N	0.0520	0.0545	0.0430	0.0500	0.0560
	t_5	0.0545	0.0505	0.0440	0.0440	0.0465
	CN	0.0475	0.0540	0.0555	0.0470	0.0515
T2RPB	N	0.0505	0.0595	0.0465	0.0460	0.0605
	t_5	0.0315^-	0.0385	0.0430	0.0405	0.0365^-
	CN	0.0430	0.0525	0.0460	0.0475	0.0500
ATLR	N	0.0050^-	0.0000^-	0.0005^-	0.0005^-	0.0005^-
	t_5	0.0000^-	0.0005^-	0.0000^-	0.0005^-	0.0010^-
	CN	0.0005^-	0.0005^-	0.0005^-	0.0000^-	0.0105^-
TLRPB test	N	0.0555	0.0550	0.0490	0.0585	0.0570
	t_5	0.0445	0.0455	0.0610	0.0495	0.0530
	CN	0.0510	0.0575	0.0465	0.0545	0.0525
RTLRPB test	N	0.0555	0.0550	0.0490	0.0585	0.0570
	t_5	0.0490	0.0460	0.0535	0.0455	0.0405
	CN	0.0495	0.0550	0.0425	0.0450	0.0520

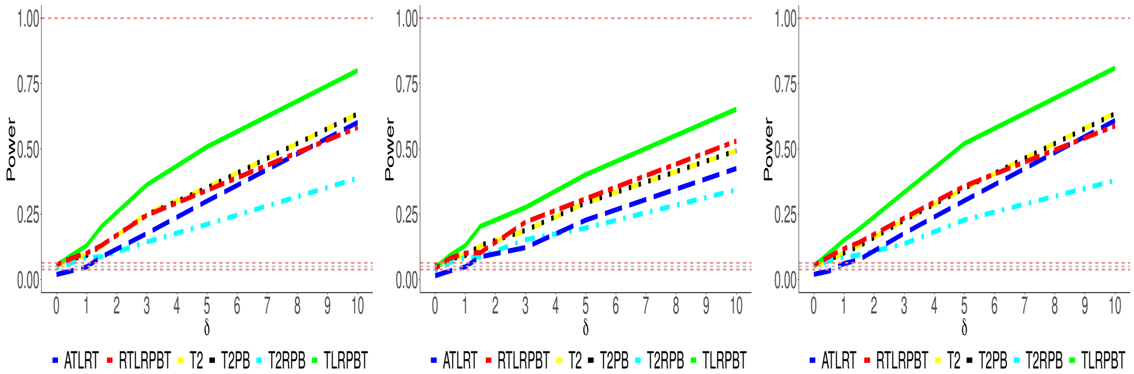
$^-$: significantly ($p < 0.01$) less than the nominal significance level of 5%.

Note in Figure 1 that the TLRPB test showed the best performance among all evaluated tests with $p = 2$. Under non-normality or in the presence of outliers, the performance of this test showed a loss of power only when the multivariate Student-t distribution with 5 degrees of freedom is considered (Figure 1 (b)). We also noticed that the asymptotic version ATLR has lower power as expected; see Tables 1 and 2.

In order to verify if the patterns observed in Figure 1 remain the dimensionality is fixed in $p = 2$ and the multivariate distributions are the same, but the number of observations varied in $n = 50$, 70, 100 and 200 for generating the power curves by Monte Carlo simulations. Under these circumstances, the behavioral power patterns shown in Figure 1 remain the same (Figures 2, 3, 4 and 5). The only exceptions are for the multivariate Student-t distribution with 5 degrees of freedom (Figures 2(b), 3(b), 4(b) and 5(b)). In this case, the performance of the TLRPB test and RTLRPB test is equivalent and higher than the performance of the other tests.

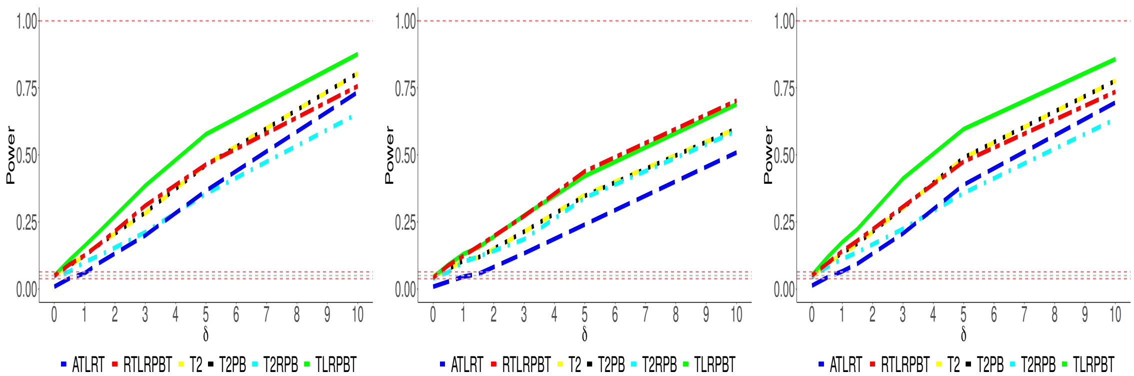
In general, for this dimension ($p = 2$), the parametric bootstrap version (TLRPB test) performed better when compared to the other tests. Also, all tests show a very robust behavioral pattern, since they control the type I error probabilities and show higher power when compared with the multivariate normal case.

We decide to increase the dimension to $p = 5$ and to maintain the same sample sizes n and distributions. Similar performance of all tests that presented at $p = 2$ remains. Considering the dimension $p = 5$, Figures 6, 7, 8, 9 and 10 show the results obtained for the power of the tests. Figures 6(b), 7(b), 8(b), 9(b) and 10(b) show that power of the parametric bootstrap version (TLRPB test) and the robust parametric bootstrap version (RTLRPB test) are equivalent and higher than the other tests in the multivariate Student-t distribution with 5 degrees of freedom. In general, the TLRPB test performs better. The ATLR continues to show substantial low power, once it has shown to be a conservative test, this pattern of behavior is expected; see Tables 1 and 2.



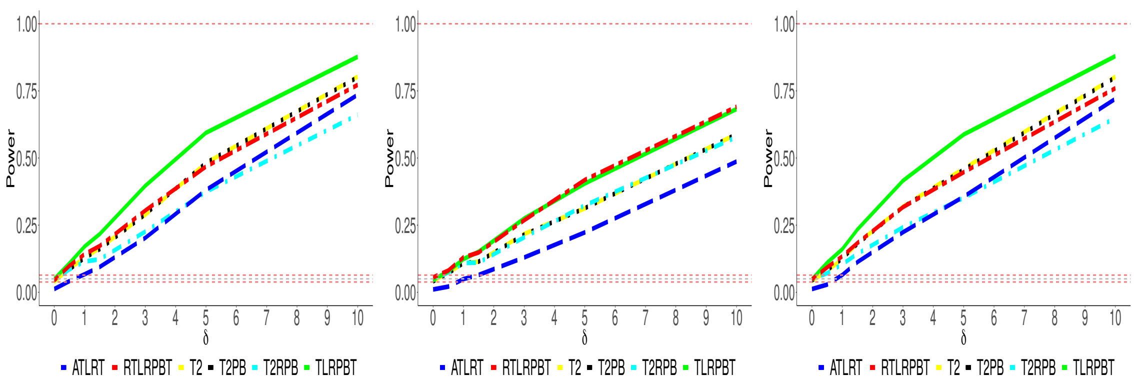
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 1. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 10$ and $p = 2$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



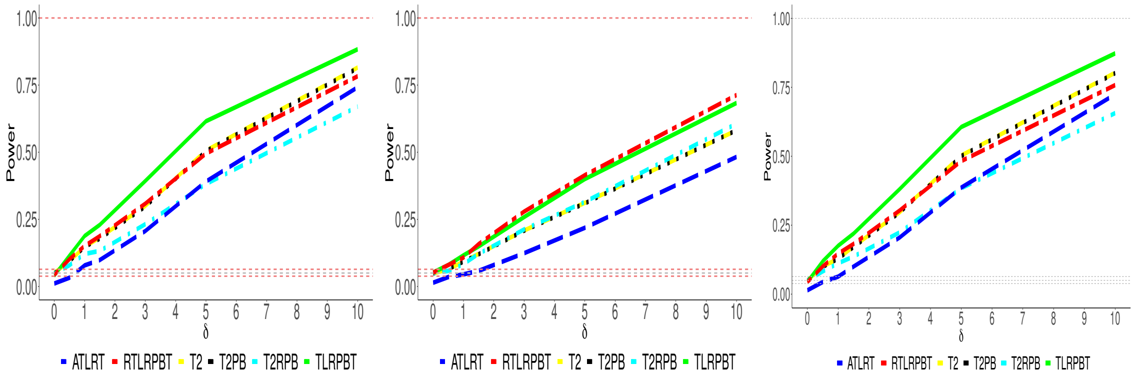
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 2. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 50$ and $p = 2$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



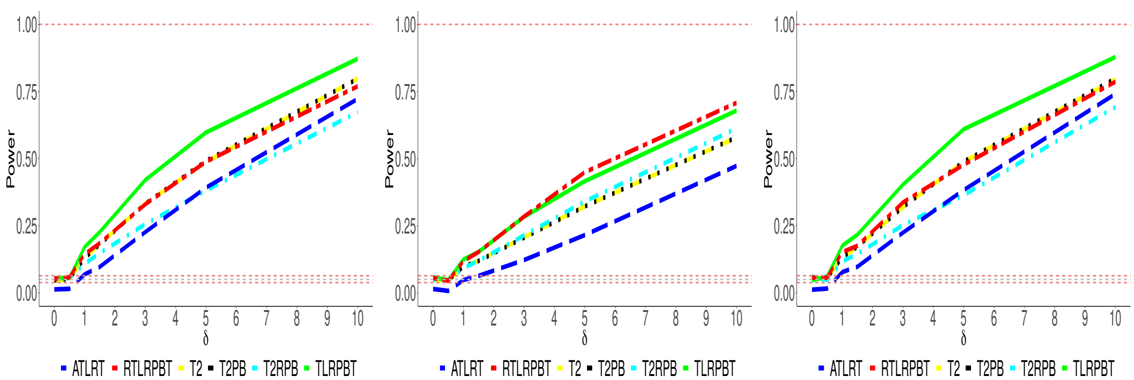
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 3. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 70$ and $p = 2$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



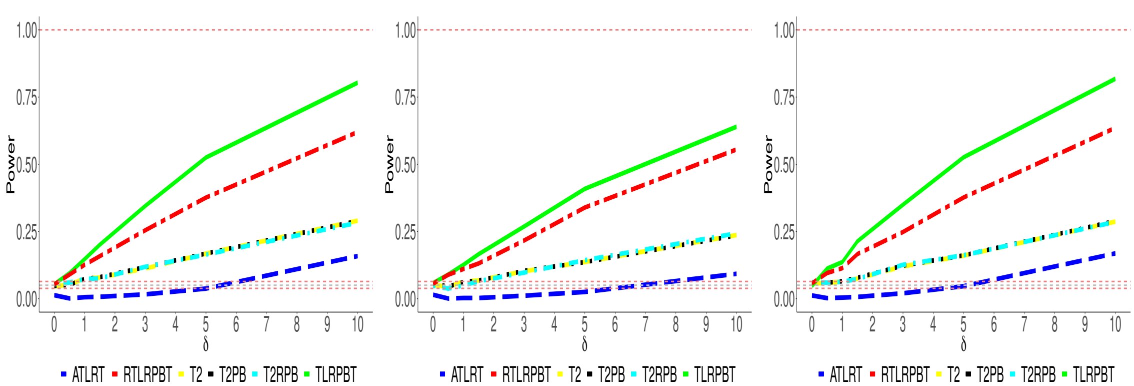
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 4. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 100$ and $p = 2$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



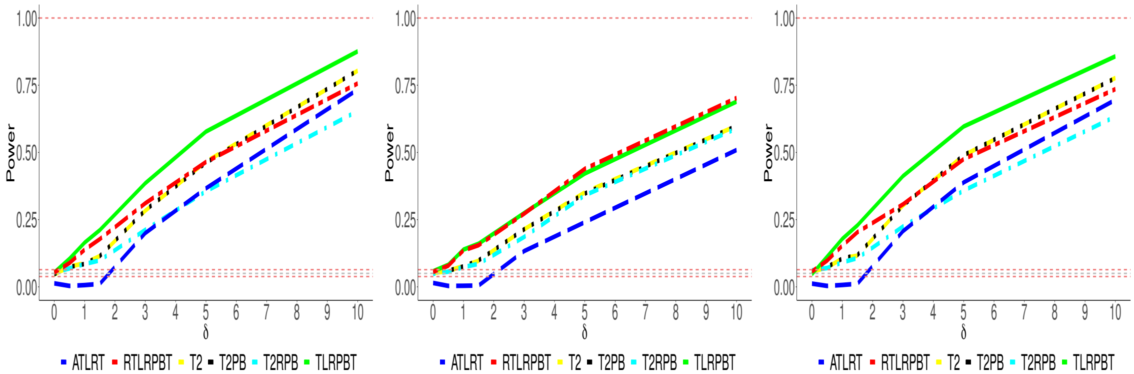
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 5. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 200$ and $p = 2$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



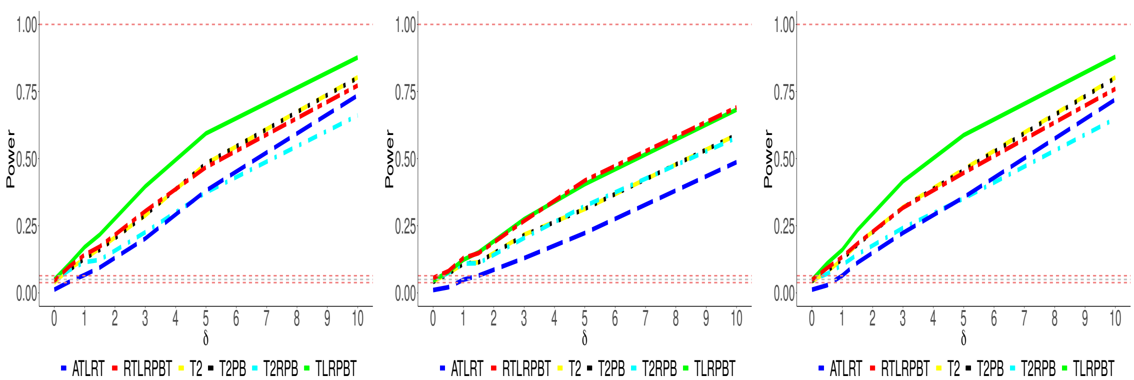
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 6. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 10$ and $p = 5$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



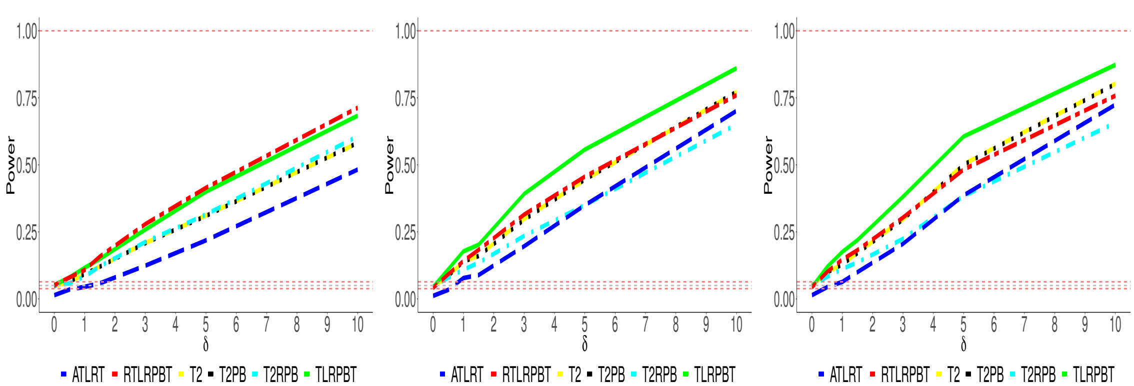
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 7. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 50$ and $p = 5$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



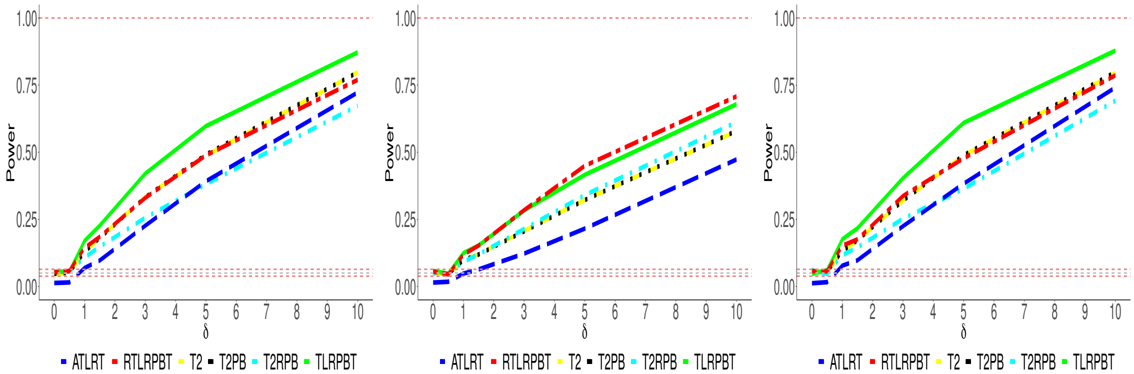
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 8. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 70$ and $p = 5$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 9. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 100$ and $p = 5$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 10. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 200$ and $p = 5$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.

3.4 Special case of $p = 200$: type I error probabilities and power

Table 3 shows the results considering this case of high dimensionality ($p = 200$). It is considering sample sizes of $n = 10, 50, 70, 100$, and 200 and therefore deal with the cases where $p \geq n$. For this circumstance, the traditional T^2 test and the adapted T2PB, T2RPB can not be applied due to the high dimension issue (see Section 1). It is noticed that the ATLR is conservative in all scenarios considered. It is expected that this test shows substantial low power. It is also noticed that the other adapted tests, TLRPB test and RTLRPB test, are exact, according to the binomial test (see Section 2).

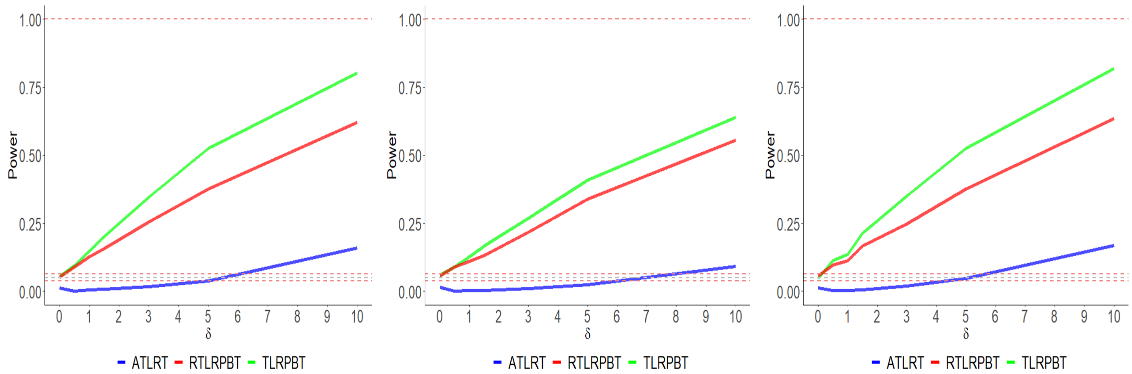
Table 3. Type I error probabilities of the three tests with $\alpha = 5\%$ and $p = 200$, considering the multivariate Normal (N), Student-t with 5 degrees of freedom (t_5) and CN distributions.

Test	Model	n				
		10	50	70	100	200
ATLR	N	0.0130^-	0.0130^-	0.0120^-	0.0110^-	0.0130^-
	t_5	0.0145^-	0.0145^-	0.0105^-	0.0140^-	0.0145^-
	CN	0.0120^-	0.0120^-	0.0120^-	0.0140^-	0.0120^-
TLRPB test	N	0.0535	0.0535	0.0495	0.0450	0.0535
	t_5	0.0580	0.0580	0.0460	0.0520	0.0580
	CN	0.0470	0.0470	0.0485	0.0445	0.0470
RTLRPB test	N	0.0530	0.0530	0.0445	0.0430	0.0530
	t_5	0.0555	0.0555	0.0555	0.0510	0.0555
	CN	0.0570	0.0570	0.0480	0.0435	0.0570

$^-$: significantly ($p < 0.01$) less than the nominal significance level of 5%.

The power results for $p = 200$ are shown in Figures 11 to 15. Considering $n = 10$, Figure 11 shows that the TLRPB test outperformed the other proposed tests. It is also noticed that, unlike the other dimensions considered ($p = 2$ and $p = 5$), the TLRBP and RTLRPB test perform similarly with the multivariate Student-t distribution with 5 degrees of freedom (Figure 11 (b)). It is also noticed for this scenario that the asymptotic version (ATLR) shows substantial low power. This fact is already expected, as it did control the type I error probabilities in a conservative way (Table 3).

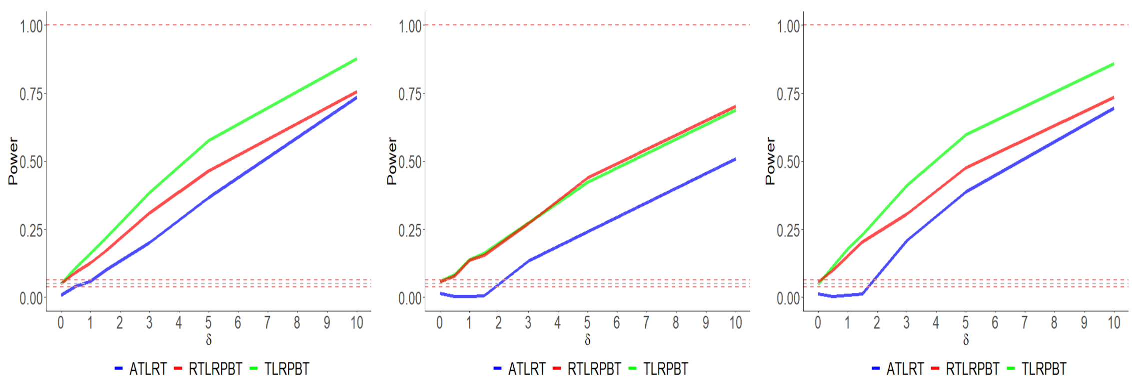
When the sample size n is increased, it is noticed that the TLRPB test continues to perform better (Figures 12, 13, 14 and 15). For the multivariate Student-t distribution with 5 degrees of freedom (Figures 12(b), 13(b), 14(b) and 15(b)), the TLRPB test and RTLRPB test have similar performance. This pattern is identical to those for the dimensions of $p = 2$ and $p = 5$. In contrast, the ATLR has substantial power gain for this dimension (p = 200).



(a) Normal distribution (b) t_5 distribution (c) CN distribution

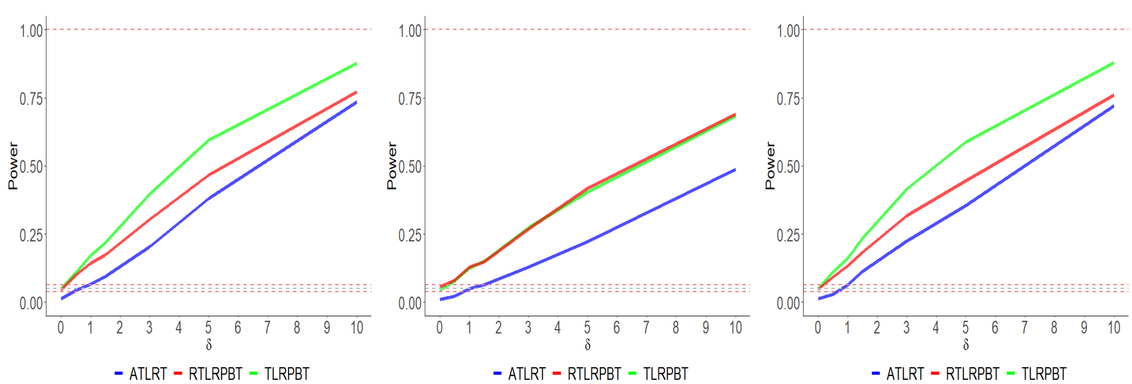
Figure 11. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 10$ and $p = 200$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.

This fact can be noticed in Figures 13, 14 and 15. In general, the TLRPB test (parametric bootstrap version test) outperformed in power when confronted with other tests.



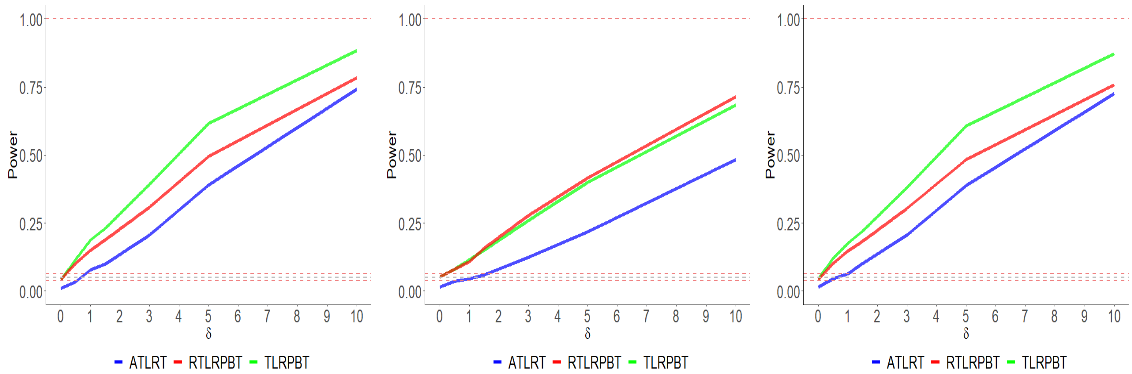
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 12. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 50$ and $p = 200$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



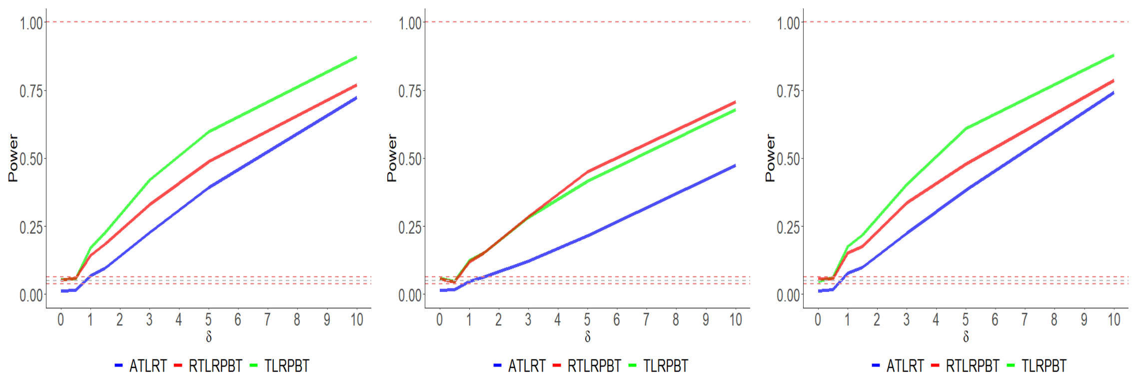
(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 13. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 70$ and $p = 200$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 14. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 100$ and $p = 200$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.



(a) Normal distribution (b) t_5 distribution (c) CN distribution

Figure 15. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with $n = 200$ and $p = 200$, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.

In a more general analysis, considering all scenarios evaluated ($n = 10, 50, 70, 100$, and 200 and $p = 2, 5, 200$), the Monte Carlo simulations for type I error probability and power showed that the TLRPB test performed better. We recommend using this test as it is not hard to implement and computationally fast.

Willems et al. (2002) concluded that his proposed new test T^2_{R} showed power losses when compared to the traditional Hotelling T^2 test for several configurations of n and p ($n > p$). Under p-variate CN populations, with 10% of contamination, the T^2_{R} has also less power than the traditional Hotelling T^2 test. Dong et al. (2016) indicated that his proposed new test for high dimensional data, based on a shrinkage process of the traditional Hotelling T^2 statistic test, showed high power under p-variate normal and Student-t with 4 degrees of freedom distributions, considering different values of $\boldsymbol{\mu}$.

In both tests (Willems et al., 2002; Dong et al., 2016), the effect of the sample size n influenced the power, which is an expected fact. In the results of the present work it does not occur, as presented before, because the values of the population means change when the sample sizes change, keeping fixed the value of the distance of Mahalanobis $\delta(\boldsymbol{\mu}, \boldsymbol{\mu}_0) = n(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$. To clarify, note that Willems et al. (2002) and Dong et al. (2016) fixed the population mean vector $\boldsymbol{\mu}$, and therefore, the value of Mahalanobis distance δ increases with increasing sample size n. Thus the power grew with an increase of n. On the contrary, in the present work $\boldsymbol{\mu}$ changed in each simulation when n varied, keeping the value of δ fixed, which keeps the power practically constant, as can be seen from Figures 1 to 15. Differences in the power values found under identical configuration, but at different values of n, are attributed to the Monte Carlo error.

4. Special cases and real data analysis

4.1 General context

Marozzi (2015) proposed an alternative multivariate test class for case-control studies for high dimensional data, considering heavy tails or skewed distributions. The proposed tests are based on the combination of tests on inter point distances. The Euclidean distance is utilized. These tests are exact, unbiased and consistent. The results showed that the proposed tests are very powerful under normality, heavy tails, and skewed distributions. Marozzi (2015) applied these same tests to magnetic resonance data which are usually with few observations and many variables, that is, high-dimensional data.

We decided to verify the behavior of our proposed tests regarding the type I error control and power considering heavy tails and skewed distributions. For this, we consider the multivariate exponential distributions with parameter $\boldsymbol{\lambda} = \mathbf{1}$. In the latter case, data are shifted to zero mean by subtracting the population exponential mean $\boldsymbol{\mu} = \mathbf{1}$. We consider in our simulations only the dimensions $p = 2$ and 5 and the same sample sizes n of Section 2.

4.2 Multivariate exponential distribution

For the multivariate exponential distribution, in general, the tests are very liberal. The exception occurred for ATLR and TLRBPT. The ATLR is exact for $n = 10$ and conservative for $n \geq 50$ with $p = 2$ or $p = 5$ (see Tables 4 and 5). The TLRPB test did not control the type I error for $n = 10$ and $n = 100$ with $p = 2$ and $p = 5$, showing a liberal behavior (see Tables 4 and 5, though with no expressive difference from the nominal significance level of 5%. In the other cases, it is exact. For large enough n, say $n = 200$, the T2 and T2BP tests showed either type I error probabilities control or inexpressive liberal behavior, although significant (see Tables 4 and 5).

Table 4. Type I error probabilities of the six tests with $\alpha = 5\%$ and $p = 2$, considering the multivariate exponential distribution.

Test	Model	n = 10	50	70	100	200
T2		0.1296^+	0.0778^+	0.0648^+	0.0658^+	0.0558
T2BP		0.5414^+	0.9651^+	0.0648^+	0.0658^+	0.0558
T2RBP		0.3470^+	0.9412^+	0.9821^+	1.0000^+	1.0000^+
ATLR	exponential	0.0459	0.0269^-	0.0160^-	0.0199^-	0.0070^-
TLRPB test		0.0927^+	0.0608	0.0439	0.0658^+	0.0489
RTLRPB test		0.2393^+	0.6112^+	0.7149^+	0.8335^+	0.9811^+

$^-$: significantly ($p < 0.01$) less than the nominal significance level of 5%.
$^+$: significantly ($p < 0.01$) greater than the nominal significance level of 5%.

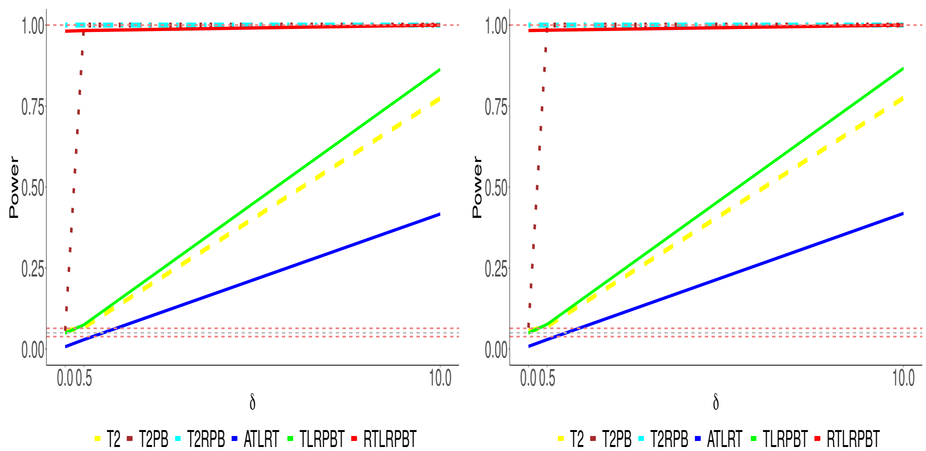
The power of the tests for exponential distribution with $n = 200$ and $p = 2$ and $p = 5$ can be seen in Figures 16(a) and (b). Only the T2, ATLR, and TLRPB test tests should be considered in the comparison, as they are those who controlled the Type I error probabilities. Again, the TLRPB test test is the most powerful, especially with $\delta = 10$, followed by the T2 and ATLR tests, in this order.

Table 6 shows the results of type I error probabilities for multivariate exponential distribution. We consider the dimension $p = 200$ and the sample size $n = 200$, with the nominal

Table 5. Type I error probabilities of the six tests with $\alpha = 5\%$ and $p = 5$, considering the multivariate exponential distribution.

Test	Model	n 10	50	70	100	200
T2	exponential	0.1296^+	0.0778^+	0.0648^+	0.0658^+	0.0558
T2BP		0.5517^+	0.9753^+	0.0676^+	0.0759^+	0.0658^+
T2RBP		0.3490^+	0.9423^+	0.9827^+	1.0000^+	1.0000^+
ATLR		0.0449	0.0267^-	0.0180^-	0.0299^-	0.0080^-
TLRPB test		0.0929^+	0.0618	0.0459	0.0755^+	0.0495
RTLRPB test		0.2397^+	0.6125^+	0.7193^+	0.8435^+	0.9831^+

$^-$: significantly ($p < 0.01$) less than the nominal significance level of 5%.
$^+$: significantly ($p < 0.01$) greater than the nominal significance level of 5%.



(a) $n = 200$, $p = 2$ (b) $n = 200$, $p = 5$

Figure 16. Power of the tests as a function of the generalized Mahalanobis distance δ between the parametric and hypothetical vector means, with multivariate exponential distribution, where the gray lines represent the lower and upper limits of the exact binomial interval and the maximum power.

significance level of 5%, that is, an extreme case. We realized that only the ATLR and TLRPB test tests controlled the type I error, being the first conservative and the second, exact. The margin of error of the Monte Carlo simulations is also presented (see Table 6). The RTLRPB test test showed all type I error probabilities close to 1 and is extremely liberal in this high-dimensional case.

Table 6. Type I error probabilities ($\delta = 0$) and power ($\delta = 0.5$ and 10) of the three tests with $\alpha = 5\%$, $p = 200$ and $n = 200$, considering the multivariate exponential distribution, where the values in parenthesis express the margin of error of Monte Carlo simulations.

Test	Model	δ 0	0.5	10
ATLR	exponential	0.0069^-(0.0051)	0.0000(0.0036)	0.0000(0.0036)
TLRPB test		0.0488 (0.0133)	0.0473(0.0131)	0.9909(0.0058)
RTLRPB test		0.9810^+(0.0084)	0.9673(0.0110)	1.0000(0.0036)

$^-$: significantly ($p < 0.01$) less than the nominal significance level of 5%.
$^+$: significantly ($p < 0.01$) greater than the nominal significance level of 5%.

We develop a package that is available in the R software (R Core Team , 2020) to assist the user in executing the mentioned methodology called multivariate tests for the vector of means (Alves and Ferreira, 2020). Then, we introduce the use of this package to one real data set.

4.3 Application to real data

In this section, the proposed methodology is applied to one real data set, that deals with the contents of sand and clay from capoeira nova, in the Amazon, Brazil, available in Ferreira (2018). The data set has two variables (sand and clay) and 30 observations ($p = 2$, $n = 30$). We want to verify that the new capoeira soil has an average sand and clay content equal to that of a forest population, at a level of 5% of significance. An exploratory analysis is previously carried out and we verified that the variables sand and clay are correlated and the data do not show normal p-variable according to the Royston test. There is also the presence of outliers in the data. Table 7 presents the data set of sand and clay contents in a new capoeira soil in the Amazon to be analyzed.

The vector of sample averages for the sand and clay contents takes on the values of 22 and 36.1, respectively, that is, $\overline{\boldsymbol{X}} = [22, 36.1]^\top$. According to Ferreira (2018), it is known that in a forest soil the average levels of sand and clay content have values equal to 14 and 42, respectively, that is, $\boldsymbol{\mu}_0 = [14, 42]^\top$. So, in possession of the samples collected of sand and clay contents in a new capoeira soil, in the Amazon, the hypotheses to be tested are H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The T2, T2PB, T2R, T2RPB, ATLR, TLRPB and RTLRPB tests have been applied; see Section 2.

Table 8 shows that all tests took the same decision to reject the null hypothesis H_0. However, since the assumption of p-variate normality is not met, we suggest choosing the result of the TLRPB test because this is the most powerful among all tests evaluated in Alves and Ferreira (2019). All tests provided the same decision to reject the null hypothesis.

Table 7. Sand and clay contents in a new capoeira soil in the Amazon.

sand	clay	sand	clay	sand	clay
11	38	20	32	13	47
24	25	18	34	28	32
16	49	17	39	11	45
18	34	30	32	27	36
5	64	45	24	7	59
11	40	11	50	42	23
17	38	41	21	21	35
9	40	22	36	48	21
13	40	14	32	12	36
53	21	25	28	31	32

Table 8. Tests for the vector of population means for the levels of sand and clay in a new capoeira soil, in the Amazon.

Test	Statistics	p-value	Decision
T2	11.93406	0.00802	Reject H_0
T2PB	11.93406	0.00899	Reject H_0
T2RPB	45.19158	0.00049	Reject H_0
ATLR	9.21556	0.00997	Reject H_0
TLRPB test	9.21556	0.00299	Reject H_0
RTLRPB test	7.11871	0.02848	Reject H_0

5. Conclusions

The trace likelihood ratio parametric bootstrap test is recommended for testing hypothesis about a multivariate population mean vector of normal and non-normal populations, including the presence of outliers. For the case of the contaminated multivariate normal distribution, the robust average and comedian covariance matrix estimators performed below tests that do not use these estimators. This fact occurred in all scenarios evaluated considering this distribution. It is possible to conclude that the use of robust comedian mean and covariance estimators is not helpful for testing hypotheses on a population mean vector.

These tests have some limitations, as in the multivariate lognormal distribution, where they did not perform well in controlling type I error probabilities, being considered liberal. As a future work, we will intend to adapt these tests to data from two or more populations.

References

Alves, H.J. and Ferreira, D.F., 2019. Proposition of new alternative tests adapted to the traditional T^2 test. Communications in Statistics: Simulation and Computation, 1-14.

Alves, H.J.P. and Ferreira, D.F., 2020. TVMM: Multivariate Tests for the Vector of Means. R package version 3.1. https://CRAN.R-project.org/package=TVMM.

Bai, Z. and Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statistica Sinica, 311–329.

Chen, S.X. and Qin, Y.L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics, 38, 808–835.

Dong, K., Pang, H., Tong, T., and Genton, M.G., 2016. Shrinkage-based diagonal Hotelling's tests for high-dimensional small sample size data. Journal of Multivariate Analysis, 143, 127–142.

Falk, M., 1997. On MAD and comedians. Annals of the Institute of Statistical Mathematics, 49s, 615–644.

Ferreira, D.F., 2018. Estatística Multivariada. Ufla, Brazil.

Ledoit, O. and Wolf, M., 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. The Annals of Statistics, 30, 1081–1102.

Lee, S.H., Lim, J., Li, E., Vannucci, M., and Petkova, E., 2012. Order test for high-dimensional two-sample means. Journal of Statistical Planning and Inference, 142, 2719–2725.

Maronna, R.A. and Zamar, R.H., 2002. Robust estimates of location and dispersion for high-dimensional datasets. Technometrics, 44, 307–317.

Marozzi, M., 2015. Multivariate multidistance tests for high-dimensional low sample size case-control studies. Statistics in Medicine, 34, 1511–1526.

Mudholkar, G.S. and Srivastava, D.K., 2000. A class of robust stepwise alternatives to Hotelling's T^2 tests. Journal of Applied Statistics, 27, 599-619.

Cardoso de Oliveira, I.R. and Ferreira, D.F., 2010. Multivariate extension of chi-squared univariate normality test. Journal of Statistical Computation and Simulation, 80, 513–526.

Pan, G.M. and Zhou, W., 2011. Central limit theorem for Hotelling's T^2 statistic under large dimension. The Annals of Applied Probability, 1860–1910.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sajesh, T.A. and Srinivasan, M.R., 2012. Outlier detection for high dimensional data using the Comedian approach. Journal of Statistical Computation and Simulation, 82, 745–757.

Srivastava, M.S., 2009. A test for the mean vector with fewer observations than the dimension under non-normality. Journal of Multivariate Analysis, 100, 518–532.

Srivastava, M.S. and Du, M., 2008. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis, 99, 386–402.

Srivastava, M.S., Katayama, S., and Kano, Y., 2013. A two sample test in high dimensional data. Journal of Multivariate Analysis, 114, 349–358.

Tiku, M.L., 1982. Robust statistics for testing equality of means or variances. Communications in Statistics: Theory and Methods, 11, 2543–2558.

Wang, R., Peng, L., and Qi, Y., 2013. Jackknife empirical likelihood test for equality of two high dimensional means. Statistica Sinica, 23, 667–690.

Wagala, A., (2020). A likelihood ratio test for correlated paired multivariate samples. Chilean Journal of Statistics, 11, 41–51.

Willems, G., Pison, G., Rousseeuw, P.J., and Van Aelst, S., 2002. A robust Hotelling test. Metrika, 55, 125–138.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://chjs.mat.utfsm.cl/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright