EVALUATION OF APPROXIMATED AND EXACT MULTIVARIATE TESTS FOR MEAN VECTORS: A DATA SIMULATION STUDY

Maurício Silva LACERDA[1]
Jhennifer dos Santos NASCIMENTO[2]
Eduardo Campana BARBOSA[2]
Rômulo César MANULI[2]
Moysés NASCIMENTO[2]
Ana Carolina Campana NASCIMENTO[2]
Paulo César EMILIANO[2]

- ABSTRACT: The present study aimed to evaluate, through data simulation, the multivariate statistical tests Likelihood ratio test (LRT) and Hotelling's T^2 test for mean vectors regarding the type I error rate and the power of test. The scenarios were designed to analyze test performance under the influence of $p-$variate normality, correlation, and homogeneity of variance, as well as number of variables and sample size. Our results show that the type I error rate was not affected by the violation of the assumptions of independence and homogeneity of variances, due to the presence of $p-$variate normality, differently from the power of test. In data simulation of $p-$variate distribution with heavier tails than usual (Student$-t$ with 1 degree of freedom), the Hotelling's T^2 showed to be conservative, while the LRT showed better results, especially for small sample sizes.

- KEYWORDS: Likelihood ratio test (LRT); Hotelling's T^2; type I error; most powerful test.

[1]Instituto Federal de Rondônia - IFRO, CEP: 76993-000 - Caixa Postal 51, Colorado do Oeste, RO, Brasil. E-mail: *mauriciolacerda57@gmail.com*

[2]Universidade Federal de Viçosa - UFV, Centro de Ciências Exatas, Departamento de Estatística, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: *jhennifer.nascimento@ufv.br, duducampana@hotmail.com, romulomanuli@ig.com.br, moysesnascim@ufv.br, ana.oliveira8@ufv.br, paulo.emiliano@ufv.br*

1 Introduction

When a hypothesis test is used, either in a univariate or multivariate statistical context, we expect results that are reliable and allow the researcher to correctly infer about the unknown population parameters of interest. An ideal scenario is that where the test does not reject the null hypothesis when it is true and rejects the null hypothesis when it is false. However, this is a hypothetic result, which may not happen in reality. Therefore, previous knowledge about the hypothesis test is essential, especially regarding the different conditions (sample size, number of variables, etc.) and assumptions (normality, independence, homogeneity of variance, etc.) involved (CANTELMO and FERREIRA, 2007). According to Rafter *et al.* (2002), if any condition or assumption is violated, and the test still shows satisfactory performance or a performance close to what was outlined during the elaboration of its theory, we call this test robust to that condition or assumption. Otherwise, the test is called sensitive.

When there is no previous information about the application of a hypothesis test available, we can apply a data simulation technique as an alternative to verify the test's robustness and sensitivity to a given condition or assumption. Through this technique, we can establish scenarios and recreate different real situations without carrying out an experiment, stopping production, wasting raw materials, workforce, and financial resources (DACHS, 1988).

In data simulation studies, the comprehension of some statistical concepts is important. The first concept is the significance level (α) that, according to Steel and Torrie (1980), can be controlled by the researcher and refers to the probability of rejecting the null hypothesis (H_0) when it is true, in other words:

$$\alpha = P\left[\text{reject } H_0 \middle|\, H_0 \text{ is true}\right].$$

In a situation like this, we are dealing with an error known as type I error. It means that by repeating the test several times, we expect the type I error to occur at $\alpha\%$ times. A second important concept is the size of type II error (β), that according to Mood *et al.* (1974), refers to the probability of not rejecting the null hypothesis (H_0) when it is false, in other words:

$$\beta = P\left[\text{not reject } H_0 \middle|\, H_0 \text{ is false}\right].$$

The size of type II error cannot be directly controlled by the researcher, and therefore, we adopt the power of test (Pw) analysis. According to Oliveira *et al.* (2005), the power of the test is defined as $(1-\beta)$ and refers to the probability of rejecting the null hypothesis (H_0) when it is false, in other words:

$$Pw = P\left[\text{reject } H_0 \middle|\, H_0 \text{ is false}\right].$$

In data simulation, the main results focus on type I error and the power. In such procedure, it is estimated a rate for type I error ($\widehat{\alpha}$) that is further compared to the nominal value of α. If $\widehat{\alpha}$ surpasses α significantly, the test is considered liberal

(HOCHBERG and TAMHANE, 1987); otherwise, it is considered conservative (CARMER and SWANSON, 1973). In addition, a test is exact if $\widehat{\alpha}$ is significantly equal to α (MOOD *et al.*, 1974). It is also estimated a rate for the Pw. In general, a test is considered powerful if Pw is close to or surpasses 0.80 (SANTOS and FERREIRA, 2003; RAMOS and FERREIRA, 2009; MINGOTI and SILVA, 2010; BARROSO *et al.*, 2012).

The two error measurements mentioned are inversely proportional. Therefore, liberal statistical tests tend to display low values for $\widehat{\beta}$ and hence high values for $\widehat{Pw}$ that are not actually true and thus should be corrected. For that reason, liberal tests show weak performance and are not recommended. On the other hand, conservative tests tend to display high values for $\widehat{\beta}$, and hence low values for $\widehat{Pw}$ and for that reason, they are also not recommended. The exact test is considered ideal because it reliably reproduces Pw values.

Data simulation is applied in several works. In an econometric context, Lemonte *et al.* (2004) assessed the augmented Dickey and Fuller's test applied to Brazilian inflation tax series, which have long-dependence properties or, in other words, stochastic processes generated by ARFIMA models. Cantelmo and Ferreira (2007) used data simulation to compare the performance of Shapiro-Wilk's multivariate normality tests with asymmetry and kurtosis tests proposed by Mardia (1970, 1974, 1975). Mingoti and Silva (2010) proposed two new multivariate statistical tests to monitor variance and covariance matrices in multivariate processes by comparing them with the traditional generalized variance test. Barroso *et al.* (2012) assessed the efficiency of the Durbin-Watson (DWG) generalized test in detecting serial autocorrelation of up to fourth order in time series. Riboldi *et al.* (2014) compared the performance of parametric (Bartlett, Brow-Forsythe, O'Brien, and Levene's test) and nonparametric tests (Siegel-Tukey, Ansari-Bradley, Klotz and Mood's test) of homogeneity of variances, among others.

In this work, we assessed the performance of two multivariate statistical tests for mean vectors, the Likelihood Ratio Test (LRT) and the Hotelling's T^2 test, through data simulation. In sum, both tests were assessed for the behavior of type I error and power in situations with different sample sizes, number of variables, absence of $p-$variate normality, and independence and homogeneity of variances. We believe that our results will help minimize incorrect analysis and erroneous results since knowing the weakness of a test and the conditions on which the test performance is best or worse beforehand is important for its application.

2 Materials and methods

2.1 Likelihood ratio test (LRT)

The multivariate LRT is used when the researcher wants to test the hypothesis that a parametric vector $\boldsymbol{\theta}$ belongs to any $\mathbb{R}^p$ subspace, either restricted (Ω_0) or unrestricted (Ω). More specifically, in a hypothesis test, the restricted space refers to the null hypothesis (H_0) while the unrestricted space refers to the alternative hypothesis (H_a), in other words:

$$\begin{cases} H_0 : \boldsymbol{\theta} \in \Omega_0 \\ H_a : \boldsymbol{\theta} \in \Omega_1 = \Omega - \Omega_0 \end{cases}$$

Given that $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)$ represents a random sample (independent and equally distributed), where $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \cdots, Y_{ip})'$ for $i = 1, \cdots, n$, the LRT consists of obtaining the maximum of the likelihood function for both subspaces and further calculate its ratio according to shown in (1), in order to obtain a decision criterion for the hypothesis:

$$\Lambda = \frac{\max \left[L_{\Omega_0}(\boldsymbol{Y}; \widehat{\boldsymbol{\theta}}) \right]}{\max \left[L_{\Omega_1}(\boldsymbol{Y}; \widehat{\boldsymbol{\theta}}) \right]}. \tag{1}$$

For (1), we can infer that if Λ big, the restricted space possibly will contain the values of the parameters present in vector $\boldsymbol{\theta}$, resulting in not rejection of the null hypothesis $H_0 : \boldsymbol{\theta} \in \Omega_0$. The opposite will happen if Λ is small and, possibly, the null hypothesis, $H_0 : \boldsymbol{\theta} \in \Omega_0$ will be rejected. To establish a rejection region for H_0, based on probabilities, it would be necessary to know Λ's exact distribution, conditioned to the fact that H_0 is true (Ferreira, 2011). Since this is not a simple task, according to Mood *et al.* (1974), we can use the approximated expression $-2\ln(\Lambda)$, that if $\Omega_0 \subset \Omega$, with $\Omega_0 \subset \mathbb{R}^s$ and $\Omega \subset \mathbb{R}^r$, has asymptotically distribution of χ^2_{r-s}, where r is the number of parameters to be estimated on the unrestricted space and s is the number of parameters to be estimated on the restricted space.

To formulate the LRT in this work, we considered that the $\boldsymbol{Y}$ vector has normal $p-$variate distribution, $\boldsymbol{Y} \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\theta}, \Sigma)$. According to what was shown in Johnson and Wichern's work (2002), in (2) and (3), we have the maximum of the likelihood functions for the restricted and unrestricted spaces, respectively.

$$L_{\Omega_0}\left(\boldsymbol{\theta}_0, \hat{\Sigma}\right) = (2\pi)^{-\frac{np}{2}} \left| \hat{\Sigma} + \left(\overline{\boldsymbol{Y}}_. - \boldsymbol{\theta}_0\right)\left(\overline{\boldsymbol{Y}}, - \boldsymbol{\theta}_0\right)' \right|^{-\frac{n}{2}} \exp\left\{-\frac{np}{2}\right\} \tag{2}$$

$$L_{\Omega}(\widehat{\boldsymbol{\theta}}, \hat{\Sigma}) = (2\pi)^{-\frac{np}{2}} \left| \hat{\Sigma} \right|^{-\frac{n}{2}} \exp\left\{-\frac{np}{2}\right\}, \tag{3}$$

where $\overline{\boldsymbol{Y}}_.$ and $\hat{\Sigma}$, are defined according to the equations (4) and (5)

$$\widehat{\boldsymbol{\theta}} = \overline{\boldsymbol{Y}}. = \frac{\sum\limits_{i=1}^{n} \boldsymbol{Y}_i}{n} \tag{4}$$

$$\hat{\Sigma} = S_n = \frac{\sum\limits_{i=1}^{n} \left(\boldsymbol{Y}_i - \boldsymbol{\theta}_0\right)\left(\boldsymbol{Y}_i - \boldsymbol{\theta}_0\right)'}{n}. \tag{5}$$

Ferreira (2011) demonstrates that the Λ statistics and the approximated statistics described by Mood *et al.* (1974) for hypothesis tests of mean vectors for a population, with unknown variance and covariance matrices (Σ), being the $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \cdots, \theta_{0p})'$, vector pre-specified, are obtained from (6) and (7).

$$\Lambda = \left[1 + \left(\overline{\boldsymbol{Y}}. - \boldsymbol{\theta}_0\right)' S_n^{-1} \left(\overline{\boldsymbol{Y}}, - \boldsymbol{\theta}_0\right)\right] \tag{6}$$

$$-2\ln(\Lambda) = n \ln\left[1 + \left(\overline{\boldsymbol{Y}}. - \boldsymbol{\theta}_0\right)' S_n^{-1} \left(\overline{\boldsymbol{Y}}_0 - \boldsymbol{\theta}_0\right)\right], \tag{7}$$

where $r = p + \frac{p(p+1)}{2}$ and $s = \frac{p(p+1)}{2}$, for $r - s = p$. Therefore

$$-2\ln(\Lambda) = n \ln\left[1 + \left(\overline{\boldsymbol{Y}}. - \boldsymbol{\theta}_0\right)' S_n^{-1} \left(\overline{\boldsymbol{Y}}. - \boldsymbol{\theta}_0\right)\right] \sim \chi^2_{\alpha,p}.$$

2.2 Hotelling's T^2

In multivariate statistics, the Hotelling's T^2 is the most formally used test to draw conclusions about mean vectors and mean differences of multivariate normal populations. According to Giri (2004), given that the $\boldsymbol{Y}$ vector has normal $p-$variate distribution, under independence and homogeneous variance and covariance matrices, $\boldsymbol{Y} \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\theta}, \Sigma)$, the test statistics, under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, is exact and has central $F-$distribution with p and $\nu + 1 - p$ degrees of freedom, according to (8).

$$T^2 = n\left(\overline{\boldsymbol{Y}}. - \boldsymbol{\theta}_0\right)' S^{-1} \left(\overline{\boldsymbol{Y}}. - \boldsymbol{\theta}_0\right) \sim \frac{\nu p}{\nu + 1 - p} F_{a;p;\nu+1-p}, \tag{8}$$

where $\overline{\boldsymbol{Y}}.$ and S, described in (9) and (10), are the sample estimators of the parametric vector $\boldsymbol{\theta}$ and the population variance and covariance matrix Σ, respectively.

$$\overline{\boldsymbol{Y}}. = \widehat{\boldsymbol{\theta}} = \frac{\sum\limits_{i=1}^{n} \boldsymbol{Y}.}{n} \tag{9}$$

$$\hat{\Sigma} = S = \frac{\sum\limits_{i=1}^{n} \left(\boldsymbol{Y}. - \overline{\boldsymbol{Y}}\right)\left(\boldsymbol{Y}. - \overline{\boldsymbol{Y}}\right)'}{n - 1}. \tag{10}$$

2.3 Data simulation

The different scenarios were simulated as follows. As to sample size (n), we assessed the tests' performance for $n = 10, 20, 50, 100$ and 200, in other words, in critical conditions (small sample sizes) and asymptotic conditions (large sample sizes). The number of variables (p) was established as $p = 2, 4$ and 6. The variance and covariance matrix structure (Σ) was defined as:

i) $$\Sigma = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{p \times p} = \sigma^2 I_{p\times p}$$

ii) $$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}_{p \times p} = \sigma^2 \left[(1-\rho)I_{p\times p} + \rho J_{p\times p}\right].$$

ii) is called compound symmetry structure, with equicorrelated variables of same variance, in which ρ is the correlation intensity ($\rho = 0.2, 0.5$ and 0.9) and J is a $p \times p$ matrix of 1s. When $\rho = 0$ the compound symmetry structure equals i) (FERREIRA, 2011).

The two next structures for the variance and covariance matrix were defined to assess test behavior when the assumption(s) iii) homogeneity of variances and iv) homogeneity of variances together with independence, were violated, in other words:

iii) $$\Sigma = \begin{bmatrix} \sigma^2_{Y_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{Y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{Y_p} \end{bmatrix}_{p \times p}, \sigma^2_{Y_1} \neq \sigma^2_{Y_2} \neq \cdots \neq \sigma^2_{Y_p}$$

iv) $$\Sigma = \begin{bmatrix} \sigma^2_{Y_1} & \rho & \cdots & \rho \\ \rho & \sigma^2_{Y_2} & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & \sigma^2_{Y_p} \end{bmatrix}_{p \times p}, \sigma_{Y_1^2} \neq \sigma^2_{Y_2} \neq \cdots \neq \sigma^2_{Y_p}.$$

To make sure the variances are heterogeneous in this study, we assumed that $\sigma^2_{Y_i}/\sigma^2_{Y_j} > 7$, for $i \neq j$ and $i, j \in \{1, 2, \cdots, p\}$. According to Pimentel-Gomes (1990), if the ratio between the highest and lowest variance is higher than 7, we can conclude that the variances are heterogeneous. Estimates of parametric values of n, p and ρ were based in the works of Cantelmo and Ferreira (2007), and Cirilo *et al.* (2006).

To verify test performance in the absence of $p-$variate normality, we simulated data of a $p-$variate Student$-t$ distribution with one degree of freedom, generated as follows:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \frac{\boldsymbol{X}}{\sqrt{\frac{U}{\nu}}} \sim t_\nu(\boldsymbol{\mu}, \Sigma),$$

where $X \sim N_p(\mathbf{0}, \Sigma)$, $U \sim \chi^2_\nu$. We adopted $\boldsymbol{\mu} = \mathbf{0}$ and $\nu = 1$, which resulted in $\boldsymbol{Y} \sim t_1(\mathbf{0}, \Sigma)$. $p-$variate normal distribution data were generated using the *mvrnorm* function from the MASS package in the R software (R CORE TEAM, 2021).

2.4 Type I error and power of test estimations

We developed an R script to calculate test statistics and further repeat the procedure 10,000 times. To evaluate the type I error rate, we assumed H_0 to be true, in other words, $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. $\widehat{\alpha}$ was computed by the proportion of rejection of the null hypothesis, with the calculated statistics and critical values obtained via significance level ($\alpha = 0.05$) and degrees of freedom from the chi-square and F distributions. The significance of population proportions was assessed via exact binomial tests at 1%, significance level, according to Riboldi *et al.* (2014). In sum, we tested the null hypothesis $H_0 : \alpha = 5\%$ against $H_a : \alpha \neq 5\%$. The test was considered exact if the H_0 was not rejected or liberal/rigorous otherwise (liberal if $\widehat{\alpha} > \alpha$ and rigorous if $\widehat{\alpha} \leq \alpha$). As for the Pw, we assumed H_0 to be false, in other words, $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + 0.5$, in order to assess test performance in detecting small changes on mean vectors. $\widehat{Pw}$ was estimated similarly to $\widehat{\alpha}$.

3 Results and discussion

3.1 Type I error

Table 1 shows type I error rate estimates for LRT with simulated data under a $p-$variate normal distribution.

As observed in Table 1, the LRT showed to be liberal in many cases, overestimating the type I error rate. The LRT was exact only in extremely large sample sizes, especially for $n = 100$ (if $p = 2$) and $n \geq 200$. According to Cantelmo and Ferreira (2007), results like this are already expected as the LRT statistics have a chi-square asymptotic distribution. Hence, it has problems when the sample size is small. Therefore, the bigger the sample size, the better the LRT performance. It is important to point out that similar results were found in the study of Mingoti and Silva (2010) when an asymptotic distribution was used for generalized variance and eigenvalue tests. Additionally, in this study, when n values fitted within the 10 to 50 interval, the increase in p resulted in an increase in type I error rates. Therefore, we recommend increasing sample size according to the increase in variable number. No effect of correlation and heterogeneity of variances on type I error was detected, no matter the n and p values used. Therefore, we can consider the LRT robust for independence and homogeneity of variances. In Table 2 are the results of Hotelling's T^2 test.

The Hotelling's T^2 was exact in all cases and robust for the assumptions and conditions tested. According to Giri (2004), the Hotelling's T^2 should have shown approximated behavior since its assumptions were violated, which made the sample not random. However, this was not verified. For Johnson and Wichern (2002),

Table 1 - Type I error rate estimates for the LRT ($p-$variate normality)

Compound symmetry									
$\alpha = 0.05$		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.088^+	0.088^+	0.091^+	0.090^+	0.090^+	0.087^+	0.089^+	0.094^+
	4	0.168^+	0.172^+	0.161^+	0.160^+	0.166^+	0.159^+	0.169^+	0.161^+
	6	0.322^+	0.316^+	0.309^+	0.322^+	0.321^+	0.327^+	0.323^+	0.317^+
20	2	0.065^+	0.064^+	0.067^+	0.070^+	0.064^+	0.072^+	0.063^+	0.068^+
	4	0.091^+	0.090^+	0.094^+	0.085^+	0.090^+	0.094^+	0.087^+	0.088^+
	6	0.120^+	0.127^+	0.125^+	0.126^+	0.131^+	0.128^+	0.126^+	0.123^+
50	2	0.060^+	0.056^+	0.057^+	0.059^+	0.059^+	0.057^+	0.056^+	0.058^+
	4	0.061^+	0.065^+	0.062^+	0.062^+	0.061^+	0.069^+	0.065^+	0.063^+
	6	0.071^+	0.073^+	0.072^+	0.071^+	0.074^+	0.068^+	0.066^+	0.071^+
100	2	0.054	0.054	0.055	0.055	0.052	0.055	0.053	0.054
	4	0.054	0.056^+	0.055^+	0.055^+	0.057^+	0.056^+	0.054	0.053
	6	0.060^+	0.062^+	0.059^+	0.057^+	0.058^+	0.060^+	0.061^+	0.058^+
200	2	0.050	0.053	0.051	0.048	0.049	0.052	0.050	0.052
	4	0.056^+	0.053	0.054	0.053	0.053	0.053	0.052	0.054
	6	0.054^+	0.054	0.054	0.053	0.055	0.055	0.056^+	0.055

$^+$ means that the type I error rate exceeded the nominal value of 5% significance ($p -$ value < 0.01).

Table 2 - Type I error rate estimates for the Hotelling's T^2 ($p-$variate normality)

Compound symmetry									
$\alpha = 0.05$		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.050	0.049	0.050	0.052	0.049	0.049	0.049	0.052
	4	0.050	0.051	0.051	0.049	0.049	0.047	0.052	0.048
	6	0.053	0.049	0.049	0.051	0.051	0.050	0.049	0.053
20	2	0.049	0.049	0.049	0.052	0.051	0.045	0.049	0.050
	4	0.051	0.049	0.049	0.050	0.051	0.049	0.052	0.051
	6	0.049	0.048	0.049	0.051	0.052	0.051	0.050	0.048
50	2	0.049	0.050	0.048	0.049	0.050	0.051	0.051	0.049
	4	0.053	0.051	0.050	0.050	0.053	0.052	0.053	0.052
	6	0.048	0.049	0.050	0.050	0.050	0.047	0.048	0.049
100	2	0.050	0.053	0.048	0.050	0.051	0.052	0.046	0.051
	4	0.047	0.047	0.047	0.052	0.048	0.048	0.051	0.052
	6	0.049	0.050	0.052	0.052	0.049	0.049	0.049	0.053
200	2	0.052	0.047	0.047	0.051	0.048	0.048	0.050	0.053
	4	0.047	0.051	0.052	0.052	0.052	0.053	0.047	0.046
	6	0.049	0.049	0.050	0.048	0.048	0.047	0.047	0.050

the multivariate normality of the random vector is the only requirement for the Hotelling's T^2 to be valid. Since we did not verify any limitations to the use of Hotelling's T^2 test, regarding sample size and variable number, we recommend its use over the LRT if the conditions are the same as those in this study. This conclusion could be foreseen because the Hotelling's T^2 was exact and will correctly reproduce Pw results, unlike the LRT.

In Table 3 are the results of type I error rate for the LRT, with simulated data of a $p-$variate Student$-t$ distribution.

Table 3 - Type I error rate estimates for the LRT ($p-$variate Student$-t$)

Compound symmetry									
$\alpha = 0.05$		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.060^+	0.061^+	0.058^+	0.062^+	0.061^+	0.061^+	0.059^+	0.060^+
	4	0.114^+	0.109^+	0.110^+	0.105^+	0.107^+	0.106^+	0.103^+	0.109^+
	5	0.254^+	0.241^+	0.255^+	0.254^+	0.251^+	0.250^+	0.245^+	0.253^+
20	2	0.054	0.046	0.054	0.053	0.051	0.053	0.050	0.050
	4	0.062^+	0.063^+	0.058^+	0.061^+	0.060^+	0.064^+	0.064^+	0.059^+
	6	0.088^+	0.083^+	0.085^+	0.082^+	0.083^+	0.087^+	0.085^+	0.088^+
50	2	0.050	0.051	0.050	0.049	0.051	0.049	0.049	0.051
	4	0.048	0.050	0.050	0.048	0.053	0.051	0.050	0.050
	6	0.055	0.059^+	0.054	0.054	0.059^+	0.056^+	0.054	0.055
100	2	0.050	0.049	0.048	0.046	0.049	0.051	0.048	0.049
	4	0.048	0.051	0.050	0.051	0.048	0.049	0.049	0.046
	6	0.050	0.052	0.048	0.052	0.047	0.051	0.047	0.049
200	2	0.045	0.050	0.048	0.047	0.048	0.050	0.049	0.047
	4	0.053	0.052	0.048	0.050	0.048	0.050	0.048	0.050
	6	0.049	0.048	0.053	0.049	0.052	0.051	0.048	0.052

$^+$ means that the type I error rate exceeded the nominal value of 5% significance ($p -$ value < 0.01).

Similar results to those in Table 1 were observed for $n = 10$, which means that the LRT was still liberal. However, with simulated data from a $p-$variate Student$-t$ distribution, the LRT became an exact test for smaller sample sizes (starting from $n = 50$), except for ($n = 50$, $p = 6$ and $\rho = 0.2$), which did not occur with simulated data from $p-$variate normal distribution. Especially for $n < 50$, we suggest not using the LRT or use it with a small number of variables (maximum 3). The same not-significant response pattern was also observed for the ρ and/or Σ effects on type I error rate. Curiously, for $n = 20$ and $p = 2$, the LRT was exact. Such result was unexpected, especially due to the LRT's asymptotic statistics. In addition, this was also a great result because $n = 20$ is possible and most likely to be adopted in real situations.

In Table 4 are the type I error rate estimates for the Hotelling's T^2 with simulated data from a $p-$variate Student$-t$ distribution.

Table 4 - Type I error rate estimates for the Hotelling's T^2 ($p-$variate Student$-t$ with 1 degree of freedom)

Compound symmetry									
$\alpha = 0.05$		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.025^-	0.024^-	0.023^-	0.024^-	0.030^-	0.025^-	0.024^-	0.027^-
	4	0.016^-	0.016^-	0.017^-	0.017^-	0.015^-	0.013^-	0.014^-	0.018^-
	6	0.013^-	0.013^-	0.011^-	0.012^-	0.011^-	0.013^-	0.010^-	0.014^-
20	2	0.033^-	0.036^-	0.035^-	0.035^-	0.039^-	0.034^-	0.034^-	0.036^-
	4	0.027^-	0.026^-	0.027^-	0.024^-	0.029^-	0.027^-	0.026^-	0.027^-
	6	0.024^-	0.021^-	0.021^-	0.020^-	0.022^-	0.020^-	0.022^-	0.018^-
50	2	0.042^-	0.043^-	0.044^-	0.044^-	0.042^-	0.046	0.047	0.046
	4	0.037^-	0.039^-	0.039^-	0.036^-	0.038^-	0.036^-	0.040^-	0.039^-
	6	0.039^-	0.035^-	0.032^-	0.033^-	0.032^-	0.036^-	0.035^-	0.032^-
100	2	0.047	0.046	0.049	0.046	0.044	0.048	0.048	0.046
	4	0.045	0.043^-	0.045	0.042^-	0.041^-	0.043^-	0.044^-	0.043^-
	6	0.041^-	0.039^-	0.040^-	0.038^-	0.040^-	0.043^-	0.042^-	0.043^-
200	2	0.046	0.047	0.049	0.048	0.046	0.051	0.047	0.049
	4	0.046	0.048	0.048	0.045	0.047	0.047	0.049	0.049
	6	0.047	0.045	0.047	0.047	0.046	0.047	0.049	0.046

$^-$ means that the type I error rate was lower than the nominal value of 5% significance ($p-$value < 0.01).

Hotelling's T^2 results in Table 4 differed from the previous Hotelling's T^2 results shown in Table 2. The test was conservative in many situations, underestimating the type I error rate. When $n = 10$ and $p = 6$, estimates were close to 0.01, and, in truth, the nominal value of α was 0.05. Only in large sample sizes, $n = 100$ (and $p = 2$) and for $n \geq 200$, the test was exact. Notice that, again, violating the assumptions of independence and homogeneity of variances led to changes in type I error rates. Such results are consistent with Johnson e Wichern's theory (2002) that the $p-$variate normality assumption is indeed the most important for test validation.

The previous result has been explored in other studies such as Arnold (1964), Mardia (1970, 1975), Everitt (1979), and Kariya (1981). Chase and Bulgren (1971) studied the bivariate Hotelling's T^2 with correlated variables. The data was generated through normal, uniform, exponential, lognormal, gamma, and double exponential bivariate distributions. For uniform distributions the Hotelling's T^2 was approximately exact. Mardia (1974) supports this finding by saying that the Hotelling's T^2 is relatively robust to the lack of normality if the distribution is approximately symmetric and liberal for gamma/exponential/lognormal distributions (asymmetric) and conservative for double exponential distributions. The double exponential distribution has denser tails than normal distributions as well as the Student$-t$. Therefore, such results are consistent with those found in Table 4. In addition, the authors noticed that the correlation effect upon the type I error rate was not significant like we also verified here. It is important to point out that the LRT has advantages over the Hotelling's T^2 if data is generated by a Student$-t$ distribution since with $n = 50$ its use is already recommended just like it is for $n = 20$ and $p = 2$.

3.2 Power of test

In Table 5 are Pw results for the LRT with simulated data from a $p-$variate normal distribution.

Table 5 - Power of test for the LRT ($p-$variate normality)

		Compound symmetry							
		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.4905	0.4314	0.3764	0.3132	0.3293	0.3206	0.3070	0.3012
	4	0.7041	0.5361	0.4260	0.3358	0.3560	0.3535	0.3346	0.3245
	6	0.8483	0.6685	0.5476	0.4663	0.4773	0.4704	0.4670	0.4608
20	2	0.7931	0.7132	0.6038	0.5159	0.5455	0.5305	0.5057	0.4982
	4	0.9488	0.7907	0.5971	0.4452	0.4647	0.4551	0.4304	0.4239
	6	0.9845	0.8213	0.5860	0.4365	0.4594	0.4427	0.4221	0.4043
50	2	0.9942	0.9858	0.9558	0.9028	0.9190	0.9116	0.8959	0.8831
	4	1.0000	0.9963	0.9553	0.8433	0.8569	0.8361	0.8215	0.8048
	6	1.0000	0.9998	0.9452	0.7831	0.8050	0.7883	0.7606	0.7435
100	2	1.0000	1.0000	1.0000	0.9969	0.9983	0.9980	0.9963	0.9960
	4	1.0000	1.0000	1.0000	0.9917	0.9940	0.9919	0.9908	0.9859
	6	1.0000	1.0000	1.0000	0.9880	0.9893	0.9846	0.9790	0.9765
200	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The Pw was affected by the violation of the assumptions of independence and homogeneity of variances. The increase in correlation degree led to reductions in Pw, no matter the p values, especially for $n < 100$. Therefore, the more correlated the variables in study, the less powerful the LRT. The addition of a heterogeneous structure in the Σ matrix also led to a decrease in Pw values. The non-homogeneity of variances seems to reduce more sharply the Pw than the independence of variables, except in situations where $\rho = 0.9$ and $p = 6$. In such cases, the fact that the variances are different did not cause reductions in Pw. This result was more notorious for $n \leq 50$, as p increases. We also verified that, for $n = 10$, the Pw increased as p increased, no matter the correlation or covariance structure. However, for $n \geq 20$, this result was not observed, especially for correlated variables ($\rho \geq 0.20$) and with different variances.

Overall, $Pw \geq 0.80$ was observed for $n \geq 50$, except where the homogeneity and independence of variances were violated, with $p = 6$. Such result was expected since the Pw increases as n increases (Mood *et al.*, 1974). However, despite p and n relations with the Pw can be true, according to Cantelmo and Ferreira (2007), the estimates for Pw found, as well as the Pw estimates for $n < 50$, are not real and should be corrected since in these cases the LRT was extremely liberal. Therefore, in these scenarios, by opting to use the LRT, we should be aware of this problem. Only for $n \geq 200$ was the test exact amongst the p values assessed. Noticed that in real situations, a sample size of $n = 200$ is not easily seen for several reasons (time, logistics, cost, etc.). Therefore, under data normality, the Hotelling's T^2 is recommended over the LRT. In Table 6 are the Pw results for the Hotelling's T^2.

Table 6 - Power of test for the Hotelling's T^2 ($p-$variate normality)

Compound symmetry									
		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.3610	0.3060	0.2550	0.2060	0.2230	0.2090	0.2080	0.1980
	4	0.3890	0.2500	0.1710	0.1320	0.1360	0.1340	0.1270	0.1220
	6	0.3050	0.1770	0.1160	0.0910	0.0920	0.0890	0.0820	0.0890
20	2	0.7400	0.6480	0.5550	0.4520	0.4810	0.4750	0.4490	0.4510
	4	0.8900	0.6860	0.4660	0.3270	0.3400	0.3420	0.3100	0.2980
	6	0.9310	0.6410	0.3840	0.2470	0.2640	0.2540	0.2330	0.2240
50	2	0.9950	0.9850	0.9520	0.8910	0.9060	0.9050	0.8860	0.8780
	4	1.0000	0.9950	0.9490	0.8010	0.8330	0.8080	0.7880	0.7790
	6	1.0000	0.9970	0.9190	0.7280	0.7530	0.7330	0.7090	0.6910
100	2	1.0000	1.0000	1.0000	0.9960	0.9980	0.9980	0.9960	0.9970
	4	1.0000	1.0000	1.0000	1.0000	0.9920	0.9910	0.9890	0.9860
	6	1.0000	1.0000	1.0000	0.9810	0.9870	0.9830	0.9780	0.9710
200	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Conclusions similar to that of the LRT were also obtained for the Hotelling's T^2 as for the violation of the assumptions of independence and homogeneity of variances from the variables in study. It means that, when both were neglected, the Pw decreased. These Pw estimates are true and reliably represent the actual values since the Hotelling's T^2 was exact (Table 2). For this reason, we did not verify a completely defined pattern that demonstrates an increase in Pw as the variable number increases, the same as happened to the LRT, which proves that the higher the p values, the lower the Pw. Under data $p-$variate normality, the Hotelling's T^2 is more recommended than the LRT; however, its use is recommended only for $n \geq 50$, conditions where this test showed Pw estimates close to or higher than 0.80.

Table 7 - Power of test for the LRT ($p-$variate Student$-t$ with 1 degree of freedom)

Compound symmetry									
		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.315	0.269	0.229	0.197	0.201	0.187	0.184	0.181
	4	0.420	0.324	0.247	0.205	0.214	0.207	0.202	0.196
	6	0.593	0.470	0.379	0.342	0.347	0.335	0.334	0.337
20	2	0.579	0.510	0.417	0.345	0.369	0.343	0.340	0.335
	4	0.749	0.550	0.389	0.276	0.286	0.278	0.263	0.258
	6	0.825	0.563	0.381	0.270	0.272	0.269	0.254	0.249
50	2	0.945	0.905	0.831	0.726	0.764	0.744	0.723	0.704
	4	0.996	0.949	0.810	0.627	0.659	0.633	0.608	0.579
	6	1.000	0.958	0.770	0.567	0.578	0.567	0.532	0.516
100	2	0.999	0.998	0.988	0.958	0.971	0.963	0.960	0.951
	4	1.000	1.000	0.988	0.922	0.938	0.929	0.914	0.905
	6	1.000	0.999	0.984	0.892	0.905	0.887	0.867	0.856
200	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	4	1.000	1.000	1.000	0.998	0.998	0.999	0.998	0.998
	6	1.000	1.000	1.000	0.998	0.998	0.998	0.995	0.995

In Table 7 are the Pw results for the LRT, with simulated data from a $p-$variate Student$-t$ distribution. Under independence and homogeneity of variances, Pw results for the LRT were affected by the violation of the assumption of $p-$variate normality. Such outcome was verified, for example, by comparing the LRT's Pw for $n = 10$ and $p = 2$ (0.4905 for bivariate normal, and 0.3150 for bivariate Student$-t$, respectively). The decrease in Pw can also be observed when adding correlation and heterogeneity structures into the Σ matrix. Remember that, under the proposed data distribution (Student$-t$), the LRT showed to be liberal for small sample sizes, but exact if $n = 20$ (if $p = 2$) and $n \geq 50$ (except if $p = 6$). Therefore, in situations where $n \geq 50$, with homogeneous variances, only for $p = 6$ and high correlation levels, $Pw \geq 0.80$ was not found. As for heterogeneous variances, where $p = 2$, Pw estimates were close to 0.80, and non-satisfactory results were verified for $p = 4$ and 6. When $n = 20$, Pw estimates were much lower than 0.80; however, in these same conditions the Hotelling's T^2 showed to be conservative and possibly will show low estimates for Pw, as it will be verified in Table 8. Therefore, under these circumstances, the LRT is more recommended.

Table 8 - Power of test estimates for the Hotelling's T^2 ($p-$variate Student$-t$ with 1 degree of freedom)

		Compound symmetry							
		Homogeneity				Heterogeneity			
n	p	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
10	2	0.178	0.144	0.119	0.099	0.095	0.097	0.096	0.096
	4	0.102	0.067	0.053	0.036	0.042	0.041	0.038	0.036
	6	0.052	0.035	0.025	0.022	0.022	0.018	0.019	0.020
20	2	0.509	0.427	0.354	0.271	0.304	0.281	0.280	0.277
	4	0.587	0.385	0.240	0.151	0.166	0.163	0.149	0.149
	6	0.547	0.279	0.161	0.093	0.103	0.098	0.093	0.089
50	2	0.934	0.897	0.815	0.705	0.743	0.718	0.707	0.680
	4	0.995	0.932	0.775	0.578	0.616	0.589	0.556	0.534
	6	0.999	0.934	0.706	0.485	0.498	0.487	0.451	0.433
100	2	0.999	0.997	0.986	0.957	0.970	0.962	0.956	0.950
	4	1.000	1.000	0.986	0.915	0.930	0.922	0.902	0.896
	6	1.000	0.999	0.981	0.876	0.885	0.869	0.844	0.834
200	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	4	1.000	1.000	1.000	0.998	0.999	0.999	0.998	0.998
	6	1.000	1.000	1.000	0.997	0.998	0.997	0.995	0.994

Hotelling's T^2 was conservative for simulated data from a Student$-t$ distribution (Table 4), therefore, its Pw estimates were lower than that under $p-$variate normality. Such difference can be verified by comparing the results from Tables 2 and 4, especially for $n \leq 50$. The effects of correlation and heterogeneity of variables were also significant, indicating a decrease in Pw. In addition, with heterogeneous variances, for $n = 10$ and $p = 2$, 4 and 6, Pw estimates were extremely low ($0.02 < Pw < 0.10$). As we can see, the violation of all three assumptions, in small sample sizes, makes the Hotelling's T^2 test inappropriate.

Conclusions

Our results demonstrated that under $p-$variate normality the Hotelling's T^2 test had better performance compared to the LRT in terms of controlling the type I error rate and Pw. In this specific situation, the Hotelling's T^2 was exact, including for small sample sizes and, therefore, reliably reproduced Pw values. The LRT, because of its asymptotic properties, showed to be liberal for small sample sizes, overestimating Pw values. Therefore, when $p-$variate normality of variables is identified, we suggest using the Hotelling's T^2 test. With original data from a $p-$variate Student$-t$ distribution, with 1 degree of freedom, the Hotelling's T^2 was conservative in many cases and, only for relatively large sample sizes ($n \geq 200$), it was exact. The LRT, unexpectedly, showed satisfactory performance, even with its asymptotic properties, becoming exact for sample sizes starting from $n = 50$. In particular, when $n = 20$ and $p = 2$ the LRT was exact. This result is very important since this scenario is more likely to happen in real situations. Therefore, if the data is simulated from a distribution with tails heavier than normal, we suggest the use of the asymptotic LRT over the Hotelling's T^2.

In addition, the type I error rate was not significantly affected by the structures of correlation and heterogeneity of variances established for the Σ matrix. The negative impact of adding these structures can be verified by looking at the decrease in Pw estimates.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We would like to thank editors and reviewers for their comments and sugestions.

LACERDA, M. S.; NASCIMENTO, J. S.; BARBOSA, E. C.; MANULI, R. C.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; EMILIANO, P. C. Avaliação de testes multivariados aproximados e exatos para vetores de médias: um estudo por simulação de dados. *Braz. J. Biom.,* Lavras, v.40, n.3, p.273-290, 2022.

- *RESUMO: O presente trabalho tem por objetivo avaliar, via simulação de dados, os testes estatísticos multivariados da Razão de Verossimilhanças (TRV) e T^2 de Hotelling (T^2) para vetores de médias, em relação à taxa do erro de tipo I e ao poder do teste. Os cenários propostos foram formados visando analisar o desempenho destes sob a influência de normalidade $p-$variada, correlação e homogeneidade de variâncias das variáveis em estudo, bem como o número de variáveis e o tamanho da amostra. Os resultados demonstraram que a taxa do erro de tipo I não foi afetada pela violação das pressuposições de independência e homogeneidade de variâncias das variáveis, dado a presença de normalidade $p-$variada, diferentemente do poder do teste. Simulando dados de uma distribuição $p-$variada com caldas mais densas que a normal ($t-$Student com 1 grau de liberdade), o T^2 mostrou-se um teste conservador, enquanto o TRV apresentou melhores resultados, inclusive em pequenas amostras.*

- *PALAVRAS-CHAVE: Teste de razão de verossimilhanças (TRV); T^2 de Hotteling; erro de tipo I; teste mais poderoso.*

References

ARNOLD, H. J. Permutation support for multivariate techniques. *Biometrika*, v.51, n.1-2, p. 65-70, 1964.

BARROSO, L. M. A.; NASCIMENTO, M.; SILVA, F. F.; NASCIMENTO, A. C. C.; PETERNELLI, L. A. Avaliação do teste generalizado de Durbin-Watson. *Revista Brasileira de Biometria*, v.30, p.432-441, 2012.

CANTELMO, N. F.; FERREIRA, D. F. Desempenho de testes de normalidade multivariados avaliados por simulação monte carlo. *Ciência e Agrotecnologia*, v.31, n.6, p.1630-1636, 2007.

CARMER, S. G.; SWANSON, M. R. An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association*, v.68, n.341, p.66-74, 1973.

CHASE, G. R.; BULGREN, W. G. A Monte Carlo investigation of the robustness of T^2. *Journal of the American Statistical Association*, v.66, n.335, p.499-502, 1971.

CIRILO, M. A.; FERREIRA, D. F.; SÁFADI, T. Estudo do poder e tamanho do teste de Lavene multivariado via simulação Monte Carlo e Booststrap. *Acta Scientiarum. Technology*, v.28, n.1, p.105-112, 2006.

DACHS, J. N. W. *Estatística computacional: uma introdução em turbo-pascal.* LTC, 1988.

EVERITT, B. S. A Monte Carlo investigation of the robustness of Hotelling's one- and two-sample T^2 tests. *Journal of the American Statistical Association*, v.74, n.365, p.48-51, 1979.

FERREIRA, D. F. *Estatística Multivariada.* 2 ed. Lavras. Ed: UFLA, 2011.

GIRI, N. C. *Multivariate Statistical Analysis.* 2.ed. New York: M. Dekker, 2004. 558 p.

HOCHBERG, Y.; TAMHANE, A. C. *Multiple comparison procedures.* New York: John Wiley & Sons, Inc., 1987.

JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis.* London: Pearson Education Limited, 2002.

KARIYA, T. A robustness property of Hotelling's T^2-test. *The Annals of Statistics*, p.211-214, 1981.

LEMONTE, A. J.; REISEN, V. A.; TOSCANO, E. M. O poder do teste de Dickey - Fuller aumentado contra processos ARFIMA(p, d, q). *Revista Brasileira de Estatística*, v.65, p.33-44, 2004.

MARDIA, K. V. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, v.57, n.3, p.519-530, 1970.

MARDIA, K. V. Applications of some measures of multivariate skewness and Kurtosis for testing normality and robustness studies. *Sankhyã A*, v.36, p.115-128, 1974.

MARDIA, K. V. Assessment of multinormality and the robustness of Hotelling's T^2 test. *Applied Statistics*, p.63-171, 1975.

MINGOTI, S. A.; SILVA, R. A. Comparação de testes estatísticos para matrizes de covariâncias de processos multivariados. *Revista Ingepro: Inovação, Gestão e Produção*, v.2, p.55-66, 2010.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction of the theory of statistics*. New York: McGraw-Hill, 1974. 564 p.

OLIVEIRA, A. C.; FERREIRA, D. F.; RAMALHO, M. A. P. *Experimentação em genética e melhoramento de plantas*. Lavras: Editora Ufla, 2005.

PIMENTEL-GOMES, F. *Curso de estatística experimental*. 13.ed. Piracicaba: Nobel, 1990. 468 p.

R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for statistical computing. 2021.

RAFTER, J. A.; ABELL, M. L.; BRASELTON, J. P. Multiple comparison methods for means. *Siam Review*, v.44, n.2, p.259-278, 2002.

RAMOS, P. S.; FERREIRA, D. F. Agrupamento de médias via bootstrap de populações normais e não-normais. *Revista Ceres*, v.56, n.2, p.140-149, 2009.

RIBOLDI, J.; BARBIAN, M. H.; KOLOWSKI, A. B. S.; SELAU, L. P. R.; TORMAN, V. B. L. Precisão e poder de testes de homocedasticidade paramétricos e não-paramétricos avaliados por simulação. *Revista Brasileira de Biometria*, v.32, p.334-344, 2014.

SANTOS, A. C.; FERREIRA, D. F. Definição do tamanho amostral usando simulação Monte Carlo para o teste de normalidade baseado em assimetria e curtose. i. abordagem univariada. *Ciência e Agrotecnologia*, v.27, n.2, p.432-437, 2003.

STEEL, R. G. D.; TORRIE, J. H. *Principles and procedures of statistics: a biometrical approach*, 2.ed, New York: McGraw-Hill, 1980. 633 p.

Received on 17.06.2021.

Approved after revised on 01.10.2021.