

# Estudo do poder e tamanho do teste de Levene multivariado via simulação Monte Carlo e *bootstrap*

Marcelo Angelo Cirillo<sup>1\*</sup>, Daniel Furtado Ferreira<sup>2</sup> e Thelma Sáfy<sup>2</sup>

<sup>1</sup>Programa de Pós-graduação em Estatística e Experimentação Agropecuária, Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil. <sup>2</sup>Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil. \*Autor para correspondência. e-mail: marcelocirillo@hotmail.com

**RESUMO.** Este trabalho tem por objetivo avaliar a taxa de erro tipo I e o poder do teste de Levene multivariado definido em função de diferentes parâmetros de posição para comparar k-matrizes de covariâncias. Para isso foram utilizadas simulação Monte Carlo e *bootstrap*. São contempladas situações com diferentes graus de correlação, heterogeneidade entre as matrizes de variâncias e covariâncias, tamanho amostral e número de variáveis. Concluiu-se que o teste de Levene centrado na média foi mais poderoso para amostras pequenas e com baixa heterogeneidade das matrizes de covariâncias; a abordagem *bootstrap* controlou a taxa de erro tipo I quando a mediana foi utilizada como parâmetro de posição.

**Palavras-chave:** simulação *bootstrap*, taxa de erro tipo I, poder do teste de Levene.

**ABSTRACT.** Evaluating the power and the size of multivariate Levene's test using Monte Carlo simulation and *bootstrap*. This work aimed to evaluate the I-type error rate and the power of Levene's multivariate test using different location parameters to compare k covariance matrices. Monte Carlo's simulation and bootstrap were used for this purpose. Configurations were considered with different degrees of correlations, heterogeneity of covariance matrices, sample sizes and number of variants. Results showed that Levene's multivariate test centered on the mean is more powerful in small sample size and with low degree of heterogeneity among covariance matrices; the bootstrap approach controlled the I-type error when median was used as scale parameter.

**Key words:** simulation *bootstrap*, power Levene's test, type I error rate.

## Introdução

A análise estatística de dados experimentais inicialmente é feita considerando as observações de interesse como uma variável aleatória, sendo denominada de variável dependente. Essa denominação é justificada pelo fato de que a variável a ser analisada é dada em função do delineamento experimental e está associado a um modelo estatístico, no qual, descrevem-se as observações obtidas por meio da realização de um experimento planejado.

Convém salientar que a análise estatística apropriada tem ligação direta com o tipo de planejamento experimental realizado, de tal forma que a escolha das estruturas dos tratamentos a serem contemplada pelo modelo por meio de variáveis independentes ou classificatórias é determinante para a obtenção dos resultados. Dentre essas estruturas, Mead e Curnow (1986) destacam os arranjos fatoriais.

Determinado o plano experimental, a coleta das respostas (variáveis dependentes) é realizada, de modo que cada unidade experimental é selecionada para ser avaliada sob condições impostas pelo planejamento. Muitas vezes são coletadas observações restritas a uma única variável dependente e nesse caso os métodos univariados devem ser aplicados. Quando se tem um conjunto de variáveis dependentes, a metodologia estatística adequada para análise dos dados é dada por técnicas multivariadas (Chatfield e Collins, 1995).

Convém salientar que nos dois casos possíveis de serem empregados os procedimentos estatísticos univariados ou multivariados, o plano experimental é o mesmo. Chatfield e Collins (1995) afirmam que a diferenciação básica é a representação da variável resposta, assim, no caso univariado a resposta de cada unidade experimental, submetida a um determinado tratamento é um escalar. Já no caso multivariado, a resposta da mesma unidade

experimental é representada por um vetor aleatório.

Na análise experimental, comumente utiliza-se a técnica da análise de variância conhecida como ANOVA (Steel et al., 1998), cuja hipótese  $H_0$  é definida por:

“ $H_0$ : As médias dos tratamentos são iguais” versus a hipótese alternativa “ $H_1$ : Pelo menos a média de um tratamento difere das demais”. No caso, multivariado a formalização das hipóteses são análogas, Morrison (1990) apresenta a generalização da ANOVA como MANOVA, cujas hipóteses são: “ $H_0$ : Os vetores de médias dos tratamentos são iguais” versus “ $H_1$ : Pelo menos dois tratamentos têm vetores de médias diferentes”.

O uso da análise de variância (ANOVA) torna-se confiável a partir do momento em que as pressuposições tais como: amostras independentes, normalidade e homogeneidade de variâncias são respeitadas (Steel et al., 1998). Em se tratando de que a MANOVA é uma generalização da ANOVA é apropriado pensar que as suposições são as mesmas, assim Johnson e Wichern (1988) enuncia que cada tratamento tem uma distribuição normal multivariada e que todos os tratamentos apresentam matrizes de covariâncias homogêneas.

Dentre essas suposições mencionadas, a verificação da normalidade multivariada, pode ser feita analisando individualmente cada componente, de modo que, se a componente for normal e todas as combinações lineares forem normais então se pode afirmar que a distribuição conjunta é normal multivariada (Johnson e Wichern, 1988). Evidente, que esse processo é bastante trabalhoso em termos práticos, por isso, utiliza-se testes de normalidade multivariada tais como: Teste de Simetria e Curtose proposto por Mardia (1970) ou procedimentos gráficos como Qqplots (Cirillo e Ferreira, 2003).

No caso da homogeneidade de matrizes de variâncias e covariâncias dos tratamentos, possivelmente essa suposição é a mais crítica, pois a heterogeneidade das matrizes de covariâncias ocasionará valores incoerentes referentes às probabilidades de se cometer o erro tipo I e II, lembrando que sob esta heterogeneidade é definido o problema de “Behrens-Fisher” multivariado (Christensen e Rencher, 1997).

Em relação aos testes multivariados utilizados com o propósito de verificar a homogeneidade das matrizes de variâncias e covariâncias, segundo O’Brien (1992) esses testes são geralmente extensões do caso univariado. Especificamente, para o teste de Levene, O’Brien (1992) propôs sua extensão para o caso multivariado, seguindo a mesma formalização do teste original (Carroll e Schneider, 1985), assim, o autor avaliou o controle da taxa de erro tipo I e o

poder mediante as distribuições não normais, ressaltando que esse teste é de fácil alteração em relação a sua formalização.

A referida facilidade na alteração do teste, conforme mencionada por O’Brien (1992), pode ser exemplificada não apenas pela substituição do parâmetro de posição, mas também pela forma como são calculados os desvios, ou seja, na versão original do teste os desvios são obtidos por meio de uma soma de quadrados; entretanto, nada se restringe a adaptar a somas dos desvios considerando o valor absoluto. Um outro detalhe, a extensão proposta pelo autor considera apenas a média como parâmetro de posição. Em virtude de tais mudanças, supostamente o controle da taxa de erro tipo I e poder é afetado, portanto, avaliar esse teste seja na versão uni ou multivariada torna-se ainda um objeto de pesquisa.

Tendo em vista que o teste de Levene multivariado (O’Brien, 1992) é realizado por meio de uma análise de variância multivariada (MANOVA), Johnson e Wichern (1988) mencionam quatro estatísticas multivariadas para testar hipóteses sobre igualdade de vetores de médias conhecidas por: Wilks, Traço de Pillai, Traço de Hotelling Lawley e Raíz máxima de Roy, sendo que todas elas possuem aproximação assintótica da distribuição F. Dentre esses quatro critérios, o mais usual é o critério de Wilks, justamente por se tratar de um teste baseado na razão de verossimilhança.

Considerando a dificuldade de se obter analiticamente a distribuição assintótica de um teste de razão de verossimilhança, surge como alternativa o método de computação intensiva conhecido como *bootstrap* proposto por Efron (1979). Este método é utilizado para estimar a distribuição da estatística através da reamostragem dos dados, além de ser uma importante ferramenta para estudos de simulação. Chou (2004) mostrou que os testes multivariados baseados em razão de verossimilhança na versão *bootstrap* quando submetidos a dados econômicos tende a apresentar um desempenho superior, justamente por apresentar probabilidades de rejeição mais próximas do nível nominal especificado.

Em decorrência as afirmações de O’Brien (1992) esse, trabalho propõe avaliar por meio de simulação Monte Carlo e *bootstrap*, a performance do teste de Levene multivariado referente ao controle da taxa de erro tipo I e poder, em circunstâncias não contempladas pelo autor tais como: diferentes graus de correlação, tamanhos amostrais, parâmetros de posição, heterogeneidade entre as matrizes de variâncias e covariâncias.

**Metodologia**

Para a realização desse trabalho foram consideradas duas populações normais multivariadas em que as respostas foram simuladas de acordo com a estrutura de correlação representada pela estrutura auto-regressiva de ordem 1 AR(1) (expressão 1). Objetivando analisar a performance do teste de Levene multivariado diante de diferentes graus de correlação ( $\rho$ ), fixaram-se os valores de  $\rho$  em 0,2; 0,5 e 0,9 sem perda de generalidades.

$$R_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ \rho & 1 & \rho & \dots & \rho^{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \dots & 1 \end{bmatrix} \quad (1)$$

Com base na matriz de correlação ( $R_1$ ) foi reproduzida a matriz de covariância  $\Sigma_1 = V_1^{1/2} R_1 V_1^{1/2}$ , sendo  $V_1$  uma matriz diagonal  $\sigma_{kk}^2$  ( $k=1,2,\dots,p$ ) em que,  $p$  representou o número total de variáveis. Os valores assumidos para  $p$  foram 3 e 5.

A matriz de covariância da segunda população ( $\Sigma_2$ ) foi obtida pelo produto  $g\Sigma_1$ . O grau de heterogeneidade entre as matrizes de covariâncias foi determinado por  $g$ . Assim os valores assumidos foram  $g=2$  e  $10$ , representando, respectivamente, as situações de baixa e alta heterogeneidade. No caso do valor de  $g=1$  corresponde à hipótese  $\Sigma_1 = \Sigma_2$ . De acordo com essa hipótese, inicialmente avaliou-se o erro tipo I, o qual é cometido quando se rejeita a hipótese nula, sendo esta verdadeira.

Os tamanhos amostrais avaliados foram  $n_1=n_2=20$  e  $n_1=n_2=50$ , de modo que  $n_1$  e  $n_2$  referiram-se, respectivamente, às amostras simuladas da primeira e segunda população. Convém salientar que a escolha dessas amostras foi feita em virtude dos valores de  $p$  assumidos, pois é sabido que o número de amostras aproximadas ao número de variáveis poderia conduzir singularidade nas operações matriciais, ou seja, inversa não única, o que inviabiliza a execução dos procedimentos doravante apresentados.

Com base nas definições mencionadas anteriormente, realizou-se a expansão do teste de Levene Multivariado para os dois parâmetros de posição usuais, isto é, média e mediana. Preliminarmente, para indicar a observação multivariada, considerou-se a seguinte notação  $Z_{ijk}$ , em que o valor  $Z$  representou a  $i$ -ésima observação da  $k$ -ésima variável na  $j$ -ésima população. Em seguida, procederam-se com as seguintes operações matriciais.

$$M = \begin{bmatrix} m_{k1} \\ m_{k2} \\ \vdots \\ m_{kp} \end{bmatrix}_{p \times 1} \quad (2)$$

em que:  $M$  indicou vetor das estimativas do parâmetro de posição utilizado posteriormente para compor a matriz dos desvios. Cada elemento  $m_{ki}$  correspondeu à estimativa do parâmetro de posição da  $k$ -ésima variável. Primeiramente foi considerado o parâmetro de posição dado pela mediana; após a obtenção dos resultados, adaptou-se novamente o teste considerando o parâmetro de posição dado pela média. Desta forma, têm-se duas possíveis definições na formalização do teste, as quais foram comparadas nas configurações entre tamanho amostral, grau de heterogeneidade, grau de correlação e número de variáveis.

Utilizando a matriz  $M$  definida em (2), constituiu-se a matriz dos desvios conforme a expressão:

$$S_j = \begin{bmatrix} s_{11} = Z_{(1)1j} - m_1 & s_{12} = Z_{(1)2j} - m_2 & \dots & s_{1k} = Z_{(1)jk} - m_k \\ \vdots & \vdots & & \vdots \\ s_{n1} = Z_{(n)1j} - m_1 & s_{n2} = Z_{(n)2j} - m_2 & \dots & s_{nk} = Z_{(n)jk} - m_k \end{bmatrix}_{n \times k} \quad (3)$$

em que  $j = 1, 2$ . Nota-se que, para a representação da observação  $Z_{ijk}$ , adicionou-se um índice entre parênteses. Este índice correspondeu a  $n$ -ésima linha da  $j$ -ésima matriz de covariância dos desvios ( $j = 1, 2$ ) a serem calculadas.

Com base nas matrizes  $S_j$ , construiu-se a matriz de soma de quadrados e produtos denominada por  $SP_j$  ( $j = 1, 2$ ) e dada por:

$$SP_j = \begin{bmatrix} (s_{11})^2 & s_{11}s_{12} & \dots & s_{11}s_{1k} & (s_{12})^2 & s_{12}s_{13} & \dots & s_{12}s_{1k} & \dots & (s_{1n})^2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ (s_{n1})^2 & s_{n1}s_{n2} & \dots & s_{n1}s_{nk} & (s_{n2})^2 & s_{n2}s_{n3} & \dots & s_{n2}s_{nk} & \dots & (s_{nm})^2 \end{bmatrix}_{n, \frac{k(k+1)}{2}} \quad (4)$$

Combinando as matrizes  $SP_1$  e  $SP_2$  tem-se uma nova matriz denominada por  $T$  cujas dimensões são de  $(2n)$  linhas por  $k(k+1)/2$ . A matriz  $T$  é definida abaixo por:

$$T = \begin{bmatrix} SP_1 \\ SP_2 \end{bmatrix}_{2n \times \frac{k(k+1)}{2}} = \begin{bmatrix} T_{11} \\ T_{12} \\ \vdots \\ T_{1n} \\ T_{21} \\ T_{22} \\ \vdots \\ T_{2n} \end{bmatrix} \quad (5)$$

em que cada elemento de  $T_{ji}$  representa a  $i$ -ésima linha da matriz soma de quadrados e produtos referente a matriz de  $Sp_j$  ( $j=1,2$ ).

Os vetores colunas que irão compor a matriz  $T$  são padronizados pela sua respectiva norma. Esta padronização é necessária em virtude de que alguns desvios poderão ser relativamente altos, principalmente quando forem obtidos por meio da mediana.

Definida a matriz  $T$ , realizou-se a análise de variância multivariada (MANOVA) e aplicou-se o critério de Wilks mencionado por Johnson e Wichern (1988) para testar a hipótese nula. Especificamente neste trabalho, em função do objetivo proposto, no qual são comparadas matrizes de covariâncias de duas populações, os quatro critérios utilizados comumente na MANOVA são equivalentes; em virtude disso o grau de liberdade será 1 e quase toda a informação será concentrada em um único autovalor. Devido a este valor no grau de liberdade, arbitrariamente optou-se pelo critério de Wilks.

A notação utilizada para representar o critério de Wilks na abordagem convencional e sua generalização *bootstrap* é definida, respectivamente, por  $W$  e  $W(B)$ . A definição deste critério é apresentada na Tabela 1. Convém salientar que a importância de se realizar um teste *bootstrap* é dada por meio da aplicação da estatística de Wilks na matriz  $T$  (equação 5), cuja distribuição é desconhecida, pois a mesma é composta por uma matriz de desvios cujos elementos são dados por variâncias e covariâncias (equação 4).

**Tabela 1.** Critério de Wilks e sua equivalência aproximada com a distribuição F

Critério	Estatística	Aproximação F	Graus de liberdade
Wilks	$W = \frac{ E }{ H+E }$	$F = \left( \frac{1-W}{W} \right) \left( \frac{rt-2f}{pq} \right)$	$V_1 = pq$ $V_2 = rt-2f$

Segundo a notação de Johnson e Wichern (1988),  $p$  refere-se ao número de variáveis e  $q$  ao grau de liberdade do tratamento. As quantidades  $f$  (equação 6);  $r$  (equação 7) e  $t$  (equação 8) foram utilizadas para definir os graus de liberdade  $v_1$  e  $v_2$ . Assim, tem-se:

$$f = \frac{pq - 2}{4} \tag{6}$$

$$r = v - \frac{(p - q + 1)}{2} \tag{7}$$

$$t = \sqrt{\frac{p^2 q^2 - 4}{p^2 + q^2 - 5}}; \text{ se } p^2 + q^2 - 5 > 0 \tag{8}$$

e 1 caso contrário

em que  $v$  corresponde ao grau de liberdade do erro.

Em relação à estatística  $W$  definida pela razão dos determinantes, tem-se que,  $|E|$  refere-se ao determinante da matriz de soma de quadrados e produtos residual e  $|H+E|$  determinante da soma das matrizes  $E$  (Equação 10) e  $H$  (Equação 11), sendo esta última a soma de quadrados e produtos relativos a tratamentos. As matrizes  $E$  e  $H$  são dadas por:

$$E = E_1 + E_2, \text{ em que,} \tag{9}$$

$$E_j = \sum_{i=1}^n (T_{ji} - \bar{T}_{.j})(T_{ji} - \bar{T}_{.j})^t; (j = 1, 2) \tag{10}$$

em que,  $T_{ji}$  foi denominado como a observação multivariada da  $i$ -ésima linha da matriz  $T$  pertencente a  $j$ -ésima população e  $\bar{T}_{.j}$  o vetor de média global das referidas populações.

$$H = \sum_{j=1}^2 n (\bar{T}_{.j} - \bar{T}_{..})(\bar{T}_{.j} - \bar{T}_{..})^t \tag{11}$$

em que  $\bar{T}_{.j}$  é o vetor de médias da  $j$ -ésima população.

Vale ressaltar que, a adaptação do critério de Wilks comumente utilizado para comparar vetores de médias dos tratamentos é feita no sentido de comparar populações, assim, em outras palavras, neste trabalho o efeito de tratamento é similar ao efeito das populações. Como as populações representam um fator em um modelo de classificação simples fica caracterizado assim um delineamento inteiramente casualizado.

Posteriormente à realização da expansão do teste de Levene para o caso multivariado, foi realizado o processo de simulação das populações implementada no software SAS módulo IML (Interactive Matrix Language) de acordo com a descrição apresentada a seguir.

**Simulação das populações**

As duas populações consideradas foram simuladas utilizando um procedimento Monte Carlo, do qual as amostras foram obtidas. Para cada população o vetor de observações amostrado foi gerado considerando a distribuição normal multivariada da seguinte forma:  $X_{ji} = F_j Z_i + \mu_j$  ( $j=1,2$ ;  $i=1,2,\dots,n_j$ ), em que  $F_j$  é o fator de Cholesky da matriz  $\Sigma_j$  e  $Z_i$  é um vetor  $p \times 1$  de variáveis normais padronizadas, geradas pela inversão da função de distribuição da normal univariada padrão

para um ponto aleatório  $U$ ,  $U \sim U[0,1]$ . O vetor  $\mu_j$  foi considerado igual ao vetor nulo por simplicidade, uma vez que não se perdia generalidade, pois foram feitas várias simulações preliminares com diferentes valores deste vetor onde se verificou que tais mudanças não afetavam os resultados dos testes.

Para cada situação avaliada foram geradas 1000 amostras para cada população por meio do método de Monte Carlo. Para cada amostra gerada, foi feita 1000 reamostragens com o objetivo de realizar o teste *bootstrap*. Com base na distribuição empírica proveniente de sucessivas reamostragens computou-se a estatística de Wilks denominada neste trabalho por  $W(B)$ . No caso da estatística de Wilks na sua forma convencional, esta foi obtida nas amostras geradas via o método Monte Carlo. Em seguida, para cada teste (convencional e *bootstrap*) computou-se a probabilidade de ocorrência do erro tipo I (equação 12), considerando  $\Sigma_1 = \Sigma_2$  ( $g=1$ ) estimando a proporção de vezes que  $H_0: \Sigma_1 = \Sigma_2$  era rejeitada nas 1000 simulações Monte Carlo.

$$P(\text{Ocorrência do erro tipo I}) = \frac{A_{\text{smc}}}{1000} \quad (12)$$

onde  $A_{\text{smc}}$  refere-se as amostras em que os teste de Levene na forma convencional e *bootstrap* controlaram o erro tipo I com probabilidade abaixo ou igual ao nível nominal especificado em 5%.

Quando  $\Sigma_2 = g\Sigma_1$  com  $g > 1$ , computou-se o poder para todas as situações em que  $\Sigma_1 = \Sigma_2$  foi rejeitada.

## Resultados e discussão

### Taxa do erro tipo I

Os resultados apresentados nas Figuras 1 a 4 referem-se às probabilidades do erro tipo I computadas via simulação Monte Carlo referente ao teste de Levene Multivariado na versão convencional e na sua generalização *bootstrap*. Esses valores, por serem representativos de uma proporção obtida em 1000 simulações de 2 populações normais com variâncias homogêneas, representam à probabilidade empírica de controle da taxa de erro tipo I.

A performance do teste de Levene relativo ao controle da taxa de erro tipo I é fortemente influenciada pelo uso dos parâmetros de locação, dos quais os desvios são calculados. Uma exemplificação deste fato é mostrada por Carroll e Schneider (1985). Esses autores verificaram que em distribuições simétricas o teste centrado na média produz probabilidades do erro tipo I mais próximas ao nível nominal. Em distribuições assimétricas, o

teste centrado na mediana é mais adequado, porém, esses autores avaliaram o teste no caso univariado. O'Brien (1992) propôs a expansão desse teste considerando apenas a mediana como parâmetro de posição. No entanto, o autor afirmou que em virtude da flexibilidade do teste para adaptações, seja no parâmetro de posição ou na obtenção dos desvios, julgam-se desejáveis novos estudos em relação a sua expansão para o caso multivariado. Diante desta justificativa, foram obtidos os resultados doravante apresentados de acordo com o seguinte critério:

Para um controle efetivo da taxa de erro tipo I foi considerado um intervalo de confiança de 95% para o nível de significância adotado em 5%, de tal forma que, as probabilidades alocadas dentro deste intervalo, para cada configuração avaliada, permitiram verificar que o teste apresentou um controle da taxa de erro tipo I com probabilidade semelhante ao nível nominal fixado; o intervalo de confiança obtido foi de (0,037;0,065).

Numa situação inicial (Figura 1) foi observado que o teste de Levene centralizado na mediana mostrou uma tendência a ser conservativo à medida que houve um aumento na correlação. Esta característica é facilmente verificada em situação de forte dependência, com um número de variáveis fixado  $p=3$  e amostras menores  $n_1=n_2=20$ , conforme é ilustrado pela Figura 1. Nessa mesma situação, a versão *bootstrap* do teste foi relevante, não apenas por apresentar probabilidades superiores em relação à abordagem convencional, mas também por corrigir o efeito conservativo observado para  $\rho=0,9$ , justificando, assim, seu esforço computacional.

Considerando o mesmo tamanho de amostra, porém aumentando o número de variáveis (Figura 2) observou-se que o controle da taxa de erro tipo I foi fortemente afetado pelo número de variáveis. Inicialmente, detectou-se que o teste de Levene, quando centrado na mediana para todas as correlações, tornou-se mais conservativo. Vale ressaltar que este resultado anteriormente verificado para  $p=3$  (Figura 1) foi de fato confirmado apenas para uma alta correlação  $\rho=0,9$ . Em relação ao teste estar centrado na média, o aumento do número de variáveis praticamente não surtiu efeito considerável, pois em ambas as situações,  $p=3$  (Figura 1) e  $p=5$  (Figura 2), os valores referentes ao controle da taxa de erro tipo I foram aproximados e coerentes com o nível nominal.

Avaliando a generalização *bootstrap*, para situações de baixa dependência observou-se que as taxas empíricas referentes ao controle da taxa de erro tipo I foram levemente superiores às taxas proporcionadas pela forma original do teste. Esta

superioridade foi verificada em relação às duas formas usuais em função dos parâmetros de posição (média / mediana) e ocasionará obviamente maiores valores de poder. Ressalta-se que este fato foi detectado para o teste de Levene no caso multivariado, sendo que Loh e Lim (1995) chegaram à mesma conclusão no caso univariado, isto é, o teste de Levene na versão *bootstrap* apresentou-se mais poderoso.

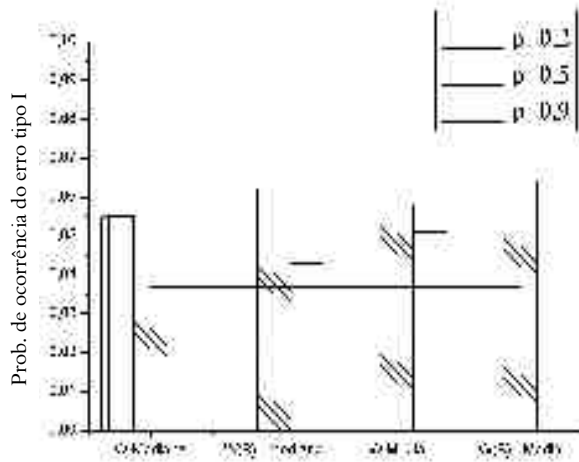


Figura 1. Probabilidade do erro tipo I fixado  $p=3$  e  $n_1=n_2=20$ .

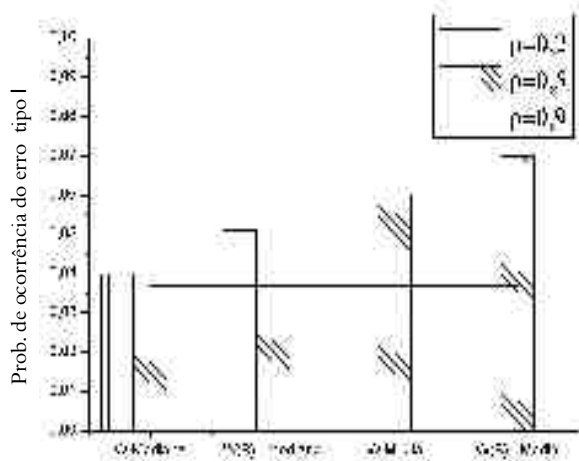


Figura 2. Probabilidade do erro tipo I fixado  $p=5$  e  $n_1=n_2=20$ .

Analisando o controle da taxa de erro tipo I diante de grandes amostras, conforme ilustram as Figuras 3 e 4, para baixas correlações percebe-se que o teste centrado na média apresentou probabilidades superiores em relação ao teste centrado na mediana. Este fato foi notório para ambos os números de variáveis avaliados. O diferencial entre os testes foi observado para um maior grau de correlação  $\rho=0,9$  em que, de fato, verificou-se a sensibilidade do teste em relação ao aumento do número de variáveis. Isto

posto, constatou-se que o teste centrado na mediana, seja na versão original ou *bootstrap*, aumentando o número de variáveis, passou a ser conservativo, (Figura 4). Para as demais situações, os resultados foram semelhantes, em relação às probabilidades do controle do erro tipo I.

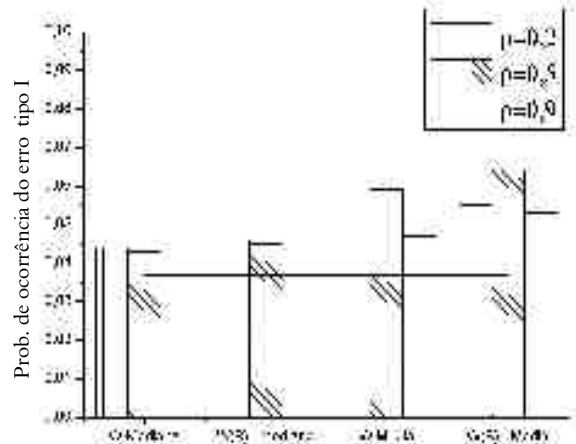


Figura 3. Probabilidade do erro tipo I fixado  $p=3$  e  $n_1=n_2=50$ .

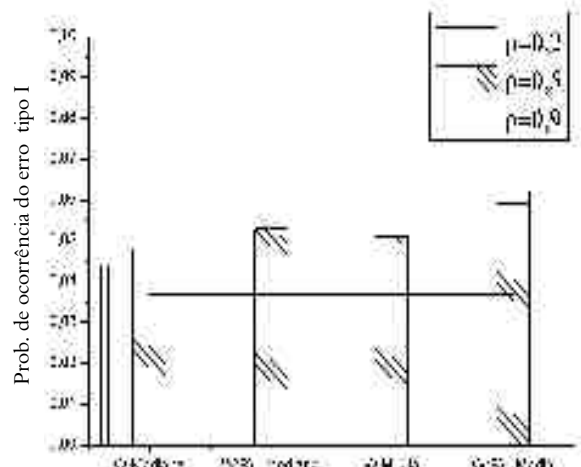


Figura 4. Probabilidade do erro tipo I fixado  $p=5$  e  $n_1=n_2=50$ .

**Poder do teste de Levene multivariado**

Os resultados encontrados na Tabela 2 correspondem aos valores de poder obtidos por meio das mesmas configurações referentes ao controle da taxa de erro tipo I, conforme é mostrado nas Figuras 1 a 4. Isto posto, avaliou-se o teste de Levene em ambos os parâmetros de posição considerando não apenas o valor referente ao poder, mas também o controle da taxa de erro tipo I proporcionado em cada situação avaliada. Desta forma, obtiveram-se os seguintes resultados.

Numa fraca correlação, isto é,  $\rho=0,2$  para um

número de variáveis  $p=3$ , observou-se que para uma baixa heterogeneidade ( $g=2$ ) em todos os tamanhos amostrais, o teste quando centralizado na média mostrou-se mais poderoso comparado ao seu uso dado pela mediana. Confrontando os valores para cada tamanho de amostra nas duas especificações (média / mediana), tem-se que para as amostras pequenas ( $n_1=n_2=20$ ), o teste centralizado na média proporcionou um ganho de aproximadamente 10 pontos percentuais. Aumentando o tamanho da amostra ( $n_1=n_2=50$ ), a diferença entre os dois testes foi apenas de 3 pontos percentuais.

**Tabela 2.** Poder para o teste de Levene multivariado considerando diferentes graus de correlação, cujos valores foram obtidos por meio das configurações entre número de variáveis, tamanho amostral, diferentes parâmetros de posição e grau de heterogeneidade entre matrizes de covariâncias.

		$\rho=0,2$							
		Mediana				Média			
		$p=3$		$p=5$		$p=3$		$p=5$	
		$n_1=n_2=20$	$n_1=n_2=50$	$n_1=n_2=20$	$n_1=n_2=50$	$n_1=n_2=20$	$n_1=n_2=50$	$n_1=n_2=20$	$n_1=n_2=50$
W	$g=2$	0,239	0,770	0,199	0,818	0,331	0,808	0,313	0,852
	$g=10$	1,000	1,000	0,996	1,000	1,000	1,000	1,000	1,000
W(B)	$g=2$	0,259	0,773	0,242	0,831	0,342	0,812	0,351	0,865
	$g=10$	1,000	1,000	0,998	1,000	1,000	1,000	1,000	1,000
		$\rho=0,5$							
W	$g=2$	0,222	0,791	0,191	0,816	0,306	0,797	0,255	0,849
	$g=10$	0,999	1,000	1,000	1,000	0,999	1,000	0,999	1,000
W(B)	$g=2$	0,238	0,799	0,230	0,826	0,331	0,798	0,301	0,857
	$g=10$	0,999	1,000	1,000	1,000	1,000	1,000	0,999	1,000
		$\rho=0,9$							
W	$g=2$	0,197	0,663	0,181	0,689	0,249	0,756	0,209	0,712
	$g=10$	1,000	1,000	0,998	1,000	0,999	1,000	0,996	1,000
W(B)	$g=2$	0,217	0,670	0,238	0,721	0,289	0,769	0,254	0,746
	$g=10$	1,000	1,000	0,998	1,000	1,000	1,000	0,997	1,000

Os valores referentes ao poder, provenientes pela generalização dos testes para abordagem *bootstrap* mostrou-se “levemente” superior em relação aos resultados provenientes da abordagem convencional. Em decorrência da baixa magnitude destes valores, julgou-se a generalização *bootstrap* irrelevante no sentido de justificar seu esforço computacional.

O aumento da correlação refletiu apenas uma redução no poder, pois na situação de extrema dependência ( $\rho=0,9$ ) o teste de Levene centralizado na média manteve-se mais poderoso, independentemente do tamanho amostral. Quando o teste foi submetido a comparar matrizes de covariâncias altamente heterogêneas ( $g=10$ ), independentemente do parâmetro de posição considerando o erro Monte Carlo, os valores referentes ao poder dos testes se igualaram tanto na abordagem convencional como na *bootstrap*.

Aumentando o número de variáveis ( $p=5$ ) no caso das amostras pequenas ( $n_1=n_2=20$ ), o teste de Levene, quando centrado na média, confirmou ser

novamente mais poderoso. Este fato mostrou que o número de variáveis é um fator de baixa importância para diferenciar a performance do teste diante da escolha do parâmetro de posição a ser utilizado. Porém, comparando os resultados da abordagem *bootstrap*, observou-se por meio da Figura 2, que este aumento ocasionou o não controle da taxa de erro tipo I, implicando, assim, em valores não confiáveis para avaliação do poder. Vale ressaltar que esta deficiência no controle da taxa de erro tipo I foi observada para uma situação de baixa e intermediária dependência, isto é,  $\rho=0,2$  e  $\rho=0,5$ .

Avaliando a situação de extrema dependência, isto é, ( $\rho=0,9$ ), conforme ilustrou a Figura 2, essa deficiência no controle da taxa de erro tipo I foi corrigida, de modo que a probabilidade da taxa de erro tipo I foi estatisticamente igual ao nível nominal especificado.

Ficou evidenciado para as amostras pequenas diante de todas as correlações, que o teste de Levene centrado na média apresentou-se mais poderoso em relação ao teste centrado na mediana. Em uma analogia com os resultados referentes à situação univariada, pode-se citar os resultados obtidos por Carroll e Schneider (1985), em que, os autores concluíram que, de fato, o teste de Levene quando centrado na média é mais poderoso.

Embora o teste centrado na média seja mais poderoso, pode-se afirmar que seu poder é mais sensível ao grau de dependência das variáveis quando comparado com o teste centrado na mediana. Para um melhor entendimento, confrontaram-se os resultados relacionados à baixa heterogeneidade ( $g = 2$ ) e constatou-se que, no teste quando centrado na mediana para todas as correlações, os valores de poder apresentaram pouca oscilação em relação ao aumento do número de variáveis. Analisando os resultados provenientes do teste centrado na média, detecta-se facilmente que à medida que há um aumento do grau de correlação, o poder é reduzido em alguns casos ( $\rho = 0,5$  e  $0,9$ ) essa redução foi de aproximadamente 5 pontos percentuais; o mesmo acontece na generalização *bootstrap*.

No caso das amostras maiores, o efeito do aumento no grau de correlação reduziu significativamente o poder do teste definido sobre a média ou mediana, apenas para grau de forte dependência entre as variáveis ( $\rho = 0,9$ ). Este fato já era esperado em virtude de que as probabilidades do controle da taxa de erro tipo I ter sido inferiores em relação às demais correlações avaliadas.

Os resultados relativos ao poder do teste foram bastante relevantes nas situações de baixa

heterogeneidade ( $g = 2$ ). No caso de um grau maior de heterogeneidade, considerando o erro de Monte Carlo, o teste de Levene apresentou valores de poder similares quando foram utilizados a média ou a mediana.

### Conclusão

O teste de Levene multivariado na abordagem *bootstrap* controlou o erro tipo I para o teste centrado na mediana independentemente do número de variáveis e dos tamanhos de amostras;

O teste de Levene multivariado centrado na média é mais poderoso do que o mesmo teste centrado na mediana quando submetido às amostras pequenas condicionadas à baixa heterogeneidade;

O poder do teste de Levene multivariado centrado na média é mais sensível ao aumento do grau de correlação.

### Referências

- CARROLL, J.R.; SCHNEIDER, H.A. Note on Levene's tests for equality of variances, *Stat. Prob. Lett.*, Amsterdam, v. 3, p. 1-194, 1985.
- CHATFIELD, C.; COLLINS, A.J. *Introduction to multivariate analysis*. London: Chapman & Hall, 1995.
- CHOU, P.H. Bootstrap tests for multivariate event studies, *Rev. Quant. Fin. Account.*, Dordrecht, v. 23, p. 275-290, 2004.
- CHRISTENSEN, W.F.; RENCHER, A.C. A comparison of type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem. *Com. Stat. Sim. Comput.*, v. 26, n. 4, p. 1251-1273, 1997.
- CIRILLO, M.A.; FERREIRA, D.F. Extensão do teste para normalidade univariado baseado no coeficiente de correlação quantil-quantil para o caso multivariado. *Rev. Mat. Estat. Marília*, v. 21, n. 3, p. 67-84, 2003.
- EFRON, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.*, Hayward, v. 7, p. 1-26, 1979.
- LOH, W.Y.; LIM, T.S. A comparison of tests of equality of variances. *Comput. Stat. Data Anal.*, Amsterdam, v. 22, p. 287-301, 1995.
- JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*, New Jersey: Prentice Hall, 1988.
- MARDIA, K.V. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, London, v. 57, p. 519-530, 1970.
- MEAD, R.; CURNOW, R.N. *Statistical methods in agriculture and experimental biology*, London: Chapman & Hall, 1986.
- MORRISON, D.F. *Multivariate statistical methods*. 3. ed. New York: McGraw-Hill, 1990.
- O'BRIEN, P.C. Robust procedures for testing equality of covariance matrices, *Biometrics*, London, v. 48, p. 819-827, 1992.
- SAS INSTITUTE. *SAS for windows*, Release 8. Cary, N.C, 2000.
- STEEL, R.G.D. et al. *Principles and procedures of statistics: A biometrical approach*. 3. ed. New York: McGraw-Hill, 1998.

Received on February 22, 2005.

Accepted on April 18, 2006.