ROBUSTEZ DE TESTES ASSINTÓTICOS E DE BOOTSTRAP PARA HOMOGENEIDADE DE COVARIÂNCIAS EM POPULAÇÕES MULTIVARIADAS

ROBERTA BESSA VELOSO SILVA

2005

ROBERTA BESSA VELOSO SILVA

ROBUSTEZ DE TESTES ASSINTÓTICOS E DE BOOTSTRAP PARA HOMOGENEIDADE DE COVARIÂNCIAS EM POPULAÇÕES MULTIVARIADAS

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

Orientador
Prof. Dr. Daniel Furtado Ferreira

LAVRAS
MINAS GERAIS – BRASIL
2005

Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da Biblioteca Central da UFLA

Silva, Roberta Bessa Veloso
Robustez de testes de homogeneidade de covariâncias assintóticos e bootstrap para populações multivariadas / Roberta Bessa Veloso Silva. – Lavras : UFLA, 2005.
93 p. : il.

Tese (Doutorado) – Universidade Federal de Lavras, 2009.
Orientador: Ludwig H. Pfenning.
Bibliografia.

1. Inferência bayesiana. 2. Erro tipo. 3. Simulação. 4. Poder. I. Universidade Federal de Lavras. II. Título.

CDD –519.42

ROBERTA BESSA VELOSO SILVA

ROBUSTEZ DE TESTES DE HOMOGENEIDADE DE COVARIÂNCIAS ASSINTÓTICOS E BOOTSTRAP PARA POPULAÇÕES MULTIVARIADAS

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

APROVADA em 28 de fevereiro de 2005

Prof. Dr. Marcelo Tavares FAMAT/UFU

Prof. Dr. Ruben Delly Veiga UFLA

Prof. Dr. Daniel Furtado Ferreira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

*Ao meu pai, que eu amo e me orgulho, **José Oswaldo**, que sempre me ajudou com sua humildade, simplicidade em ver as coisas da vida e o carinho que sempre me dedicou.*

*À minha mãe, **Lenita,** exemplo de vida e que, através da sua enorme fé e amor soube me dar forças pra continuar caminhando e superando os obstáculos que enfrentei.*

*Ao meu querido irmão, **Júnior,** pelo apoio e ajuda nos momentos difíceis que passei durante o mestrado.*

AGRADECIMENTOS

Ao meu **Deus**, força que me impulsiona a viver e que não me deixou fraquejar.

À minha **família**, que é a base e o porto seguro da minha vida, muito obrigada por tudo.

Ao **Daniel,** pela paciência, amizade e atenção sempre incentivando e me instruindo para que eu tomasse sempre as melhores decisões. Seu exemplo de vida e entusiasmo me impulsionou e motivou a chegar até aqui.

A **Lidiane,** pela amizade e carinho, colaborando, sempre que necessário, para a conclusão desta etapa.

Aos meus colegas de mestrado, **Imaculada, Gisele, Nilson, Eric, Taciana, Regilson, Antônio, Luciane, Rômulo, Mônica, Deive e Devanil** por todos os momentos que passamos juntos. Através da ajuda mútua constante conseguimos derrubar barreiras difíceis, muitas vezes rindo e outras chorando também.

Ao amigo **Marcelo Cirilo,** a quem recorri tantas vezes pra que pudesse me ajudar, obrigada pela força.

Aos professores **Lucas, Maria do Carmo, Joel, Eduardo, Júlio, Augusto e Daniel,** que tornaram melhores os dois anos de mestrado transmitindo o conhecimento necessário para a nossa formação.

Ao **CNPq,** por ter propiciado a ajuda financeira necessária pra que eu pudesse ter condições de permanecer estudando.

SUMÁRIO

LISTA DE TABELAS

RESUMO

SILVA, Roberta Bessa Veloso. **Robustez de testes de homogeneidade de covariâncias assintóticos e bootstrap para populações multivariadas.** 2005. 93 p. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG*.

O presente trabalho ressalta a importância da aplicação de testes sobre a hipótese de igualdade de matrizes de covariâncias de k populações. A violação da pressuposição da homogeneidade das covariâncias afeta diretamente a qualidade dos testes e a probabilidade de cobertura das regiões de confiança. Por essa razão, neste trabalho propõe-se aplicar dois testes de homogeneidade de covariâncias e avaliar as suas performances juntamente com a de dois outros testes mediante uso de simulação Monte Carlo em populações normais e a robustez em situações não-normais avaliando-se o erro tipo I e o poder. Os testes utilizados foram: teste de Bartlett multivariado, teste de Samiuddin generalizado para o caso multivariado e as versões bootstrap desses testes. Foram feitas combinações entre os tamanhos amostrais, o número de variáveis, a correlação e o número de populações. Os resultados obtidos permitiram concluir que o teste generalizado para o caso multivariado de Samiuddin (TGS) e a sua versão de bootstrap (TGSB) foram propostos com sucesso; o teste generalizado de Samiuddin (TGS) sob normalidade foi considerado superior ao teste de Bartlett multivariado (TBM) e a sua versão de bootstrap (TBMB); os testes de bootstrap foram considerados superiores aos assintóticos e robustos controlando a taxa de erro tipo I; o (TGSB) foi mais poderoso do que o (TBMB) e é recomendada a sua utilização para testar a hipótese de homogeneidade das matrizes de covariância.

Palavras-chave: simulação; teste de Samiuddin; inferência bayesiana.

* Comitê Orientador: Prof. Dr. Daniel Furtado Ferreira – UFLA (orientador) e Prof. Denismar Alves Nogueira – UFLA.

ABSTRACT

SILVA, Roberta Bessa Veloso. **Robustness of asymptotic and bootstrap tests for multivariate homogeneity of covariance matrices.** 2005. 93 p. Dissertation (Masters in Statistics and Agricultural Experimentation) – Universidade Federal de Lavras, MG[†].

The present work emphasizes the importance of homogeneity covariance matrices tests of multivariate populations, because in general it is presupposed normality, homogeneity of covariance and independence of the samples observations. Then, the assumption of covariance homogeneity violation reduces the quality of the test and the coverage probability of the confidence regions. For this reason this work intends to apply two tests of homogeneity of covariance and to evaluate their performances together with two other tests using Monte Carlo simulation in normal and the robustness in non normal populations evaluating the type I error rates and power. The used tests were: multivariate Bartlett's test, Samiuddin's test generalized for the multivariate case and the bootstrap versions of those tests. It was used samples sizes from 5 to 105, number of variables from 2 to 10, correlation from 0 to 0,9 and number of populations from 2 to 10. The results show that the generalized Samiuddin's test proposed (TGS) and its bootstrap version (TGSB) were successeful; the (TGS) under normality was considered superior than the original and bootstrap Bartlett's test; the bootstrap tests were considered superiors to the asymptotic test and were robust became they control type I error; the generalized bootstrap Samiuddin's test (TGSB) was more powerful than TBMB and its use is recommended for testing of covariance homogeneity.

Key words: Simulation; Samiuddin's test; Bayesian Inference.

[†] Guidance Committee: Prof. DSc. Daniel Furtado Ferreira – UFLA (Adviser),
Prof. Denismar Alves Nogueira.

1 INTRODUÇÃO

Na inferência multivariada, em geral, pressupõe-se normalidade, homogeneidade de covariâncias e independência das observações amostrais. A violação da pressuposição da homogeneidade das covariâncias afeta diretamente a qualidade dos testes e a probabilidade de cobertura das regiões de confiança. Por essa razão, é importante aplicar testes sobre a hipótese de igualdade de covariâncias de k populações. Para o caso específico de amostras de populações normais, existem vários testes na literatura já consagrados. No entanto, o teste de igualdade de k matrizes de covariâncias proposto por Bartlett (1937) é fortemente afetado quando se viola a pressuposição de normalidade multivariada, sendo poderoso quando as populações são normais.

Alternativas existem, na literatura, para o teste de Bartlett. Entre as alternativas destacam-se o teste proposto por Samiuddin (1976) utilizando a teoria Bayesiana e os testes de computação intensiva. Este é um teste de homogeneidade de variâncias em amostras de populações normais.

Atualmente, um assunto que vem se destacando é a inferência bayesiana que ficou um grande período esquecida, devido à necessidade de se calcularem analiticamente integrais complicadas. Tem ganhado cada vez mais espaço, sendo utilizada em inúmeros trabalhos, mas os resultados são muitas vezes questionados, pois muitos deles não utilizam, ou mesmo não abordam com clareza, os critérios implementados para a verificação da convergência. Outro questionamento feito também seria o de que, em casos de modelos complicados, esta exige um esforço computacional grandioso. O problema na inferência bayesiana é a necessidade de conhecer a distribuição, tanto dos dados quanto dos parâmetros.

Os testes de homogeneidade pressupõem normalidade. Se esta pressuposição for violada estes testes mostram-se muitos sensíveis (controle do erro tipo I e poder). Eles não diferenciam entre a falta de normalidade e possível heterogeneidade de variâncias. Os testes bootstrap têm sido sugeridos na literatura como uma alternativa para se realizar inferência nas mais variadas situações. No caso específico de matrizes de covariâncias, eles já foram utilizados com sucesso, no entanto em situações simuladas bastantes restritas. Estudos de simulação sugerem que não existe nenhum teste que seja uniformemente melhor para todas as distribuições e tamanhos amostrais configurados.

Neste trabalho objetivou-se:

i. propor um novo teste a partir da generalização multivariada do teste de Samiuddin para a igualdade de variâncias;
ii. comparar a sua performance com a do teste de Bartlett multivariado utilizando simulação;
iii. propor as versões de bootstrap de ambos os testes e comparar suas performances com as das aproximações assintóticos;
iv. avaliar o poder e a taxa de erro tipo I em situações normais e não-normais, sendo o último caso utilizado para verificar a robustez.

2 REFERENCIAL TEORICO

2.1 Taxas de erros (Tipo I e Tipo II)

O pesquisador, em geral, depara-se, muitas vezes, com a necessidade de testar hipóteses sobre parâmetros populacionais. Ao testar as hipóteses, o pesquisador corre o risco de tomar decisões erradas, ou seja, cometer erros. Estes erros são referenciados como sendo erro tipo I e erro tipo II. O erro tipo I é aquele que se comete ao rejeitar uma hipótese nula, dado que esta hipótese é verdadeira e a probabilidade de incorrer neste tipo de erro é representada pelo valor de significância α (Mood, Graybil & Boes, 1974). O pesquisador não consegue controlar as probabilidades de se cometer os dois tipos de erros ao mesmo tempo, embora consiga controlar a do erro tipo I. O erro tipo II é o erro que se comete ao aceitar a hipótese nula, quando esta hipótese é falsa e a probabilidade associada a esse erro é representada por β. Para que quaisquer testes de hipóteses ou regras de decisão tenham resultados válidos, eles devem ser planejados de modo que os erros de decisão sejam reduzidos ao mínimo. Isto não é tarefa simples, pois, para um dado tamanho de amostra, a tentativa de diminuir certo tipo de erro é acompanhada pelo acréscimo do outro. Deve-se reforçar a importância sobre do controle do erro tipo I e que os erros (tipo I e tipo II) são inversamente proporcionais (Borges e Ferreira, 2002). Portanto, um equilíbrio entre as taxas de erro é essencial, de modo que a taxa de erro tipo II não seja excessivamente aumentada.

Mood, Graybil & Boes (1974) propuseram um teste Υ de uma hipótese nula H_0. A função poder do teste Υ, denotada por $\pi_{\Upsilon}(\theta)$, é definida pela probabilidade de H_0 ser rejeitada quando a distribuição da qual a amostra foi obtida foi parametrizada por θ.

A função poder ideal é aquela que é 0 para aqueles $\theta' s$ correspondentes à hipótese nula e é 1 para aqueles $\theta' s$ correspondentes à hipótese alternativa. A idéia é que não se quer rejeitar a hipótese nula se ela é realmente verdadeira, mas rejeitar a hipótese nula quando ela realmente é falsa.

O poder de um teste, de acordo com Ramalho, Ferreira e Oliveira (2000), citados por Borges e Ferreira (2002), é a probabilidade de detectar diferenças entre os tratamentos quando elas realmente existem, ou seja, $1-\beta$ é a probabilidade de rejeitar a hipótese nula quando ela é realmente falsa. Os testes que possuem o valor real das taxas de erro tipo I menor do que o nominal α são ditos conservadores ou rigorosos. Em um teste conservador, sob H_0, a probabilidade de se encontrar um resultado significativo é inferior ao valor α estabelecido (Snedecor e Cochran, 1980).

Na comparação de duas médias populacionais, Borges e Ferreira (1999) buscaram avaliar o aumento de riscos de serem tomadas decisões erradas (erro tipo I e tipo II). Para avaliar as probabilidades de se cometerem os erros, duas populações foram geradas por meio de simulação Monte Carlo. Foi aplicado o teste t por meio das aproximações de Sattherthwaite (1946) e de Cochran e Cox (1957), para verificar se a diferença entre as médias amostrais era significativa ou não. Quando as variâncias populacionais foram consideradas iguais, as probabilidades de se cometer o erro tipo I foram praticamente as mesmas dos valores nominais estabelecidos. Nesta mesma situação, observou-se que à medida que as diferenças entre as médias populacionais foram aumentadas, a taxa de erro tipo II diminuiu consideravelmente, acarretando um aumento no poder do teste. Ao se aumentar o tamanho da amostra, houve uma redução nas taxas de erro tipo I e do erro tipo II. As duas aproximações do teste t, no caso das variâncias heterogêneas, não diferiram quanto à taxa de erro tipo I e tipo II empíricas.

2.2 Processo de simulação

Dachs (1988) define simulação como o processo para imitar o comportamento de um sistema real, para estudar seu funcionamento sob condições alternativas. Tornou-se cada vez mais freqüente o uso de métodos de simulação para estudar novos procedimentos estatísticos ou para comparar o comportamento de diferentes técnicas estatísticas.

A simulação de dados é, possivelmente, a área em que mais se usa geração de números aleatórios na atualidade, e é também muito antiga. Esta é uma área muito vasta, que cresceu muito e teve sua revolução com o advento do computador. Foi, no entanto, o aparecimento do computador que permitiu a um número crescente de pessoas a possibilidade de realizar simulações, desde simples jogos, até modelos da economia em escala mundial. O importante na simulação é a validação dos modelos, pois, se o modelo reproduzir com alguma fidelidade a realidade, torna-se claro a influência de certas alterações na realidade e, ainda, permite que estas alterações sejam testadas antes de serem levadas para a realidade no modelo, ou seja, com o programa de simulação (Dachs, 1988).

Segundo Naylor (1971), citado por Santos (2001), a simulação é aplicada na construção de modelos de formas extremamente diversas, desde as esculturas e pinturas da Renascença até os modelos em escala de aviões supersônicos e modelos analíticos de processos mentais. Dessa forma, a simulação tornou-se algo quase que indispensável para cientistas e teóricos.

O uso da simulação de dados tem uma grande diversidade de áreas de aplicação, basicamente sob duas linhas de atuação: problemas matemáticos completamente determinísticos, cuja solução é difícil ou em problemas que envolvem o processo estocástico Monte Carlo, cuja técnica de simulação tem base probabilística ou estocástica.

Dachs (1988) descreveu processos de geração de amostras aleatórias a partir de uma distribuição uniforme (0,1). A linguagem computacional Pascal foi utilizada pelo autor na implementação dos algoritmos para a geração de seqüências aleatórias. Este tipo de construção está baseado no teorema da probabilidade integral, que garante que é possível obter, a partir de uma distribuição uniforme (0,1), uma amostra de qualquer outra distribuição. Este teorema afirma que se U tem distribuição uniforme (0,1) e se F é uma função de distribuição qualquer, a variável $X = F^{-1}(U)$ tem função de distribuição F. Analiticamente, tem-se que:

$$P(X \leq x) = P\left(F^{-1}(U) \leq x\right) = P\left(F\left[F^{-1}(U)\right] \leq F(x)\right) = P\left(U \leq F(x)\right)$$

como U é uniforme e F(x) está entre O e 1, então

$$P\left(U \leq F(x)\right) = F(x)$$

de modo que

$$P(X \leq x) = F(x),$$

ou seja, X tem função de distribuição F.

O método Monte Carlo, de uma maneira bastante simplificada, é um método utilizado para que sejam simulados dados (variáveis aleatórias) a partir de uma seqüência pseudo aleatória, baseada na distribuição uniforme (0,1). Todo processo simulado que envolve um componente aleatório de qualquer distribuição é considerado como pertencente ao método Monte Carlo. A única restrição para o uso deste método é a sua impraticabilidade para distribuições cuja função distribuição seja desconhecida ou que a inversão não seja possível pela não existência de algoritmos numéricos. Felizmente, na literatura, existem inúmeros algoritmos eficientes de inversão das funções de distribuições comumente usadas pelos estatísticos (Santos, 2001). Outros métodos podem ser utilizados, como, por exemplo, o método da aceitação\rejeição.

2.3 Usos do método bootstrap e permutação

Segundo Manly (1988), uma opção que vem sendo usada na literatura, além de outras abordagens, é a dos testes que fazem uso de computação intensiva. Esses testes são, por exemplo, os de bootstrap e aleatorização. Esses dois procedimentos diferem entre si apenas, na maioria das circunstâncias, se a reposição é ou não considerada no processo de reamostragem. Para o caso específico de comparações de duas médias populacionais é comum substituir, em cada amostra, as observações pelos resíduos, os quais são obtidos da subtração da observação amostral pela média amostral correspondente. Os resíduos de ambas as amostras são combinadas em uma única amostra de tamanho $n_1 + n_2$, respectivamente, e calculada a estatística t do teste convencional. Repete-se esse processo milhares de vezes e o valor da estatística do teste na amostra original é confrontado com valores da distribuição gerada.

Boos e Brownie (1989) estudaram versões de bootstrap do teste de Bartlett e Layard (1973) para testar a igualdade de k matrizes de covariâncias. O teste de Layard (1973) é uma generalização para k amostras do teste de duas amostras, usando ainda jackknife (Miller, 1968). Seus resultados de simulação mostraram que as versões de bootstrap do teste de Bartlett e o teste jackknife tiveram maior desempenho do que as versões originais. Contudo, quando os dados originavam-se de uma distribuição exponencial, o tamanho do teste de bootstrap tendia a ser bastante grande.

De acordo com Zhang e Boos (1992), a estrutura de covariâncias de populações multivariadas é de interesse em vários campos. Por exemplo, a comparação de matrizes de covariâncias é usada para avaliar se houve transformações no padrão de crescimento de mandíbulas de camundongos causadas por diferenças genéticas. Uma discussão geral das estruturas das covariâncias em estudos genéticos é apresentada em Turelli (1980). Outros exemplos são apresentados em Flury (1988), Flury e Riedwyl (1988), que utilizaram algumas aplicações educacionais, (Lomax, 1985) com aplicações biométricas. Nessas aplicações, testes robustos para comparar matrizes de covariância despertaram o interesse particular dos profissionais que usualmente trabalhavam em contextos nos quais a suposição de normalidade era questionável.

A suposição de igualdade de k matrizes de covariâncias é a base da análise de variância multivariada e da análise discriminante linear. De fato, a motivação para o estudo de Zhang e Boos (1992) foi o procedimento de análise discriminante (proc discrim) do programa estatístico SAS® v.8.2 (SAS, 2000), que usa a função discriminante quadrática em vez da função discriminante linear, baseada na rejeição de um teste preliminar de igualdade de covariância usando o teste de Bartlett. Segundo os autores, a função discriminante linear é relativamente robusta à não normalidade e à heterogeneidade de variância moderada, mas a função discriminante quadrática e a estatística do teste de Bartlett são extremamente sensíveis à não normalidade. Dessa forma, o que pode acontecer para distribuições com caudas pesadas, como a distribuição normal multivariada, é que a hipótese de igualdade de matrizes de covariâncias pelo teste de Bartlett seja falsamente rejeitada em situações na qual a função discriminante linear ainda permanece adequada e melhor que a função discriminante quadrática.

De acordo com Zhang e Boos (1992), quando se deseja comparar a homogeneidade de k matrizes de covariâncias estimadas a partir de vetores aleatórios $p \times 1$ independentes $\left\{ \underset{\sim}{X}_{i1}, \ldots, \underset{\sim}{X}_{in_i}, i = 1, \ldots, k \right\}$, o teste estatístico mais freqüentemente usado é o de razão de verossimilhança de Bartlett modificado:

$$\Lambda = \left(\prod_{i=1}^{k} \left| S_i \right|^{(n_i - 1)/2} \right) \Big/ \left| S_p \right|^{(n-k)/2} \qquad (2.3.1)$$

sendo,

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i \right) \left(\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i \right)^t , \quad \overline{\underset{\sim}{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \underset{\sim}{X}_{ij}$$

$$S_p = \sum_{i=1}^{k} \frac{n_i - 1}{n - k} S_i , \quad n = \sum_{i=1}^{k} n_i$$

Sob a hipótese nula de igualdade de k matrizes de covariâncias e normalidade multivariada (Muirhead, 1982), tem-se que:

$$-2 \log \Lambda \xrightarrow{d} \chi^2_{p(p+1)(k-1)/2} ,$$

com o $min(n_1, \ldots, n_k) \to +\infty$ com $n_i / n \to \lambda_i \in (0,1)$ para $i = 1, \ldots, k$.

Embora a estatística Λ seja obtida sob a suposição de normalidade multivariada, é certamente uma estatística razoável para testar a igualdade de matrizes de covariâncias de distribuições elípticas e outras moderadamente não-normais (Zhang e Boos, 1992). A distribuição assintótica qui-quadrado desta estatística, contudo, é muito sensível à violação da suposição de normalidade. Até mesmo pequenos desvios de normalidade podem perturbar a distribuição assintótica e levar a testes muito liberais. Se a suposição de normalidade é trocada por uma suposição de uma distribuição elíptica comum com parâmetro de curtose κ, então κ pode ser estimado e um teste assintoticamente válido baseado na equação (2.3.1) pode ser construído (Muirhead, 1982). A suposição de distribuição elíptica ainda é bastante forte e difícil de se verificar em pequenas amostras.

Zhang e Boos (1992) investigaram as propriedades dos procedimentos bootstrap sob três diferentes famílias de distribuições: (1) a distribuição normal multivariada, MN (μ,Σ); (2) a distribuição t de Student multivariada com 5 graus de liberdade, MT $(5;\mu,\Sigma)$ que foi gerada como MN (μ,Σ) dividida por $\left(\chi^2_{(5)}/5\right)^{1/2}$ e (3) uma distribuição normal contaminada, CN, que foi gerada com cada componente sendo uma $N(0,1)$, com probabilidade 0,9, e uma $\chi^2_{(2)}$ com probabilidade 0,1, em que os componentes são independentes. Nas três distribuições, observou-se que os quartos momentos são finitos. A distribuição t multivariada é elipticamente distribuída, com o coeficiente de curtose maior do que o da distribuição normal multivariada.

Em todas as simulações, 1.000 conjuntos de amostras aleatórias independentes de Monte Carlo foram gerados e os p-valores foram computados para o procedimento bootstrap e para aproximação χ^2 de (2.3.1). A estatística $L=-2\ln\Lambda$ foi multiplicada por um fator de correção γ, e em seu lugar, foi usado $-2\gamma\ln\Lambda$ para o cálculo do p-valor da aproximação de χ^2 original (Mardia, Kent e Bibby, 1982). A correção adotada foi:

$$\gamma = 1 - \frac{2p^2+3p-1}{6(p+1)(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1}-\frac{1}{n-k}\right)$$

Os resultados obtidos por Zhang e Boos (1992) com relação às taxas de erro tipo I estão apresentados na Tabela 1.

TABELA 1 Simulações sob H_0 para $-2\ln\Lambda$ de Bartlett.

	Distribuições					
	Normal (μ,Σ)		Multi-T $(5;\mu,\Sigma)$		Normal Cont	
$k = 2/6$	χ^2	Bootstrap	χ^2	Bootstrap	χ^2	Boot
$n_1 = n_2 = \ldots = n_k = 20,\quad p = 2$ dimensões						
$k = 2$	0,059	0,046	0,231	0,045	0,315	0,050
$k = 6$	0,042	0,031	0,511	0,031	0,700	0,056
$n_1 = 20,\ n_2 = 40,\ \ p = 2$ dimensões e $k = 2$ $n_1 = n_2 = 20,$ $n_3 = n_4 = 30,\quad n_5 = n_6 = 40,\quad p = 2$ dimensões e $k = 6$						
$k = 2$	0,057	0,55	0,230	0,054	0,335	0,064
$k = 6$	0,046	0,044	0,627	0,036	0,802	0,049
$n_1 = n_2 = \ldots\ n_k = 20,\ \ p = 5$ dimensões						
$k = 2$	0,056	0,012	0,374	0,023	0,390	0,019
$k = 6$	0,061	0,045	0,887	0,061	0,853	0,038
$n_1 = 20,\ \ n_2 = 40,\ \ p = 5$ dimensões e $k = 2$ $n_1 = n_2 = 20,$ $n_3 = n_4 = 30,\ \ n_5 = n_6 = 40,\ \ p = 5$ dimensões e $k = 6$						
$k = 2$	0,045	0,024	0,430	0,030	0,397	0,031
$k = 6$	0,041	0,032	0,905	0,049	0,919	0,044

Pode-se perceber que, embora o procedimento bootstrap seja pouco conservador no caso normal (MN), é muito melhor que nos casos MT e CN, enquanto que a aproximação de χ^2 apresentou baixa performance para estes dois últimos casos. No caso de tamanhos de amostras iguais, o procedimento bootstrap apresentou-se melhor quando o número k de populações amostradas aumentou. A performance do teste de bootstrap não foi adequada quando a dimensão p diminuiu. Para tamanhos de amostras maiores (30,30), (30,50) e $k = 2$, $p = 5$, a distribuição normal convergiu para os valores críticos.

O poder também foi estudado para $k = 2$ (populações) e $p = 2$ (dimensões) sob distribuição normal e distribuição t_5 multivariada. Duas alternativas foram estudadas: (1) uma população com covariância $V = diag\left[2,4\right]$ e a outra, uma matriz identidade; (2) uma população com covariância:

$$c = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}$$

e a outra, uma matriz identidade. Os autores consideraram, no segundo caso, um pequeno desvio da matriz identidade.

Duas estimativas de poder são dadas na Tabela 2. As primeiras linhas em cada caso são as porcentagens de rejeição quando foi usado $\chi^2_{(3)}$ e os valores críticos bootstrap. Pelo fato de $\chi^2_{(3)}$ produzir testes muito liberais quando as amostras são de distribuições não-normais, alguns ajustamentos devem ser feitos para comparar o poder quando as amostras são de uma distribuição t_5 multivariada. As segundas linhas em cada caso foram obtidas usando como valores críticos o quinto percentil da distribuição empírica de p-valores sob H_0 obtidos na construção da Tabela 1.

A Tabela 2 mostra que o procedimento bootstrap perde um pouco de poder para a distribuição normal, mas parece ganhar no poder ajustado para a distribuição t_5 multivariada quando comparada ao procedimento $\chi^2_{(3)}$. Este ganho posterior no poder, contudo, parece improvável e pode ser devido à variabilidade ou viés nas estimativas ajustadas (Zhang e Boos, 1992).

Pelos resultados das Tabelas 1 e 2 são sugeridos que o procedimento bootstrap sustente bem seu nível sob H_0 e perca muito pouco poder sob H_1, enquanto que a performance da aproximação χ^2 é muito ruim para distribuições não-normais. Por essa razão, Zhang e Boos (1992) recomendam o procedimento bootstrap para testar hipótese sob a igualdade de matrizes de covariâncias.

TABELA 2 Estudo do poder para $-2\ln\Lambda$ de Bartlett, $k = 2$ populações, $p = 2$ dimensões e $n_1 = n_2 = 20$ para o nível nominal de 5%, obtidos por Zhang e Boos (1992).

Distribuições							
MN $(0,V)$ e MN $(0,1)$		MT $(5;0,V)$ e MT $(5;0,1)$		MN $(0,C)$ e MN $(0,1)$		MT$(5;0,V)$ e MT$(5;0,1)$	
$\chi^2_{(3)}$	Boot	$\chi^2_{(3)}$	Boot	$\chi^2_{(3)}$	Boot	$\chi^2_{(3)}$	Boot
0,762	0,642	0,763	0,49	0,264	0,233	0,436	0,155
0,735	0,657	0,499	0,52	0,244	0,243	0,164	0,177

Segundo Zhu, Ng e Jing (2002), sob a suposição de multinormalidade, o teste de hipótese de homogeneidade no problema de k amostras pode ser realizado pelo teste de razão de verossimilhança (LRT – *likelihood ratio test*). A distribuição exata do LRT é muito complicada. Quando o tamanho da amostra é suficientemente grande, utiliza-se, usualmente, a distribuição qui-quadrado, que é a distribuição limitante para a distribuição nula do LRT. Box (1949) obteve um fator de correção para o LRT de Bartlett e propôs uma estatística M com a mesma distribuição qui-quadrado para testar a homogeneidade de variâncias para o problema de k amostras. Esta estatística é a mesma que foi apresentada anteriormente, ou seja, $-2\gamma \ln \Lambda$.

Sem considerar a suposição de multinormalidade, as razões de verossimilhança seriam diferentes. Mas, ainda sob a suposição de normalidade multivariada, a distribuição nula assintótica para o teste de homogeneidade de Bartlett não é uma distribuição qui-quadrado, mas uma combinação linear de qui-quadrados como apontado por Zhang e Boos (1992). Segundo Zhu, Ng e Jing (2002), a falta de uma distribuição nula correta para as estatísticas tradicionais impele os pesquisadores a buscarem outras formas de implementação de testes. As técnicas de reamostragem, tais como bootstrap, representam uma solução. Beran e Srivastava (1985) consideraram a implementação de testes bootstrap baseado em função de autovalores de uma matriz de covariância no problema de uma amostra. Para o problema de k amostras, Zhang e Boos (1992) estudaram procedimentos bootstrap para obter valores críticos assintóticos para a estatística homogeneidade de Bartlett sem a suposição de multinormalidade. Entre outras coisas, Zhang e Boos (1993) desenvolveram uma teoria bootstrap para estatística do tipo quadrática e demonstraram a idéia usando o teste de Bartlett como um exemplo.

Uma abordagem alternativa para construir testes multivariados é o princípio da união-interseção de Roy (1953). Esse princípio usa o fato de que um vetor aleatório é normal multivariado, se e somente se, toda função linear não nula de seus elementos é uma normal univariada. Isto permite que se veja a hipótese multivariada como afirmativas conjuntas (interseção) de hipóteses univariadas resultantes de todas as funções lineares de componentes univariados, e que a região de rejeição conjunta resulte da união de todas as regiões de rejeição univariadas correspondentes. O teste para duas amostras de Roy é baseado no maior e no menor autovalor de uma matriz na métrica de outra, ou seja, é definido como sendo:

$$\lambda_i = Max\left[\left(\underset{\sim}{e}_i \Sigma_1 \underset{\sim}{e}_i\right) \Big/ \left(\underset{\sim}{e}_i^t \Sigma_2 \underset{\sim}{e}_i\right)\right]$$

em que λ_i é o critério de Roy para duas matrizes de covariância $\sum_1$ e $\sum_2$, $\underset{\sim}{e}_i$ é o autovetor associado. Mas, não existe nenhum teste de Roy para o problema de mais de duas amostras. Uma razão pode a dificuldade de estender a idéia de utilizar a razão de variâncias para comparar mais de duas amostras.

Sem multinormalidade e sem considerar a razão de verossimilhança ou o princípio da união-interseção, Zhu, Ng e Jing (2002) propuseram testes de homogeneidade para mais de duas amostras, os quais foram baseados nos autovalores das diferenças das matrizes de covariâncias amostrais resultantes de uma alteração na escala dos dados. As distribuições assintóticas dos testes estatísticos foram identificadas. Foi considerada a validade de algumas técnicas de reamostragens, como o bootstrap, simetrização aleatória e procedimentos de permutação, para calcular valores críticos e p-valores.

Existe a teoria que suporta a conclusão de que os procedimentos de permutação e simetrização aleatória apresentam-se melhores que o bootstrap, com p-valores que se aproximam do nível nominal de significância em alguns casos. Os estudos de Monte Carlo indicaram que o teste de permutação geralmente tem maior poder que o teste de bootstrap e a simetrização aleatória é compatível ao bootstrap na performance do poder. A simetrização aleatória, se aplicável, é de fácil implementação. Resultados de simulação sugerem que a proposta de Zhu, Ng e Jing (2002) é melhor que o teste de Bartlett versão bootstrap estudado por Zhang e Boos (1992).

Para uma melhor compreensão, Zhu, Ng e Jing (2002) fizeram a suposição de que $X_{i1}, X_{i2}, ..., X_{in_i}$, $i = 1,...,k$, é uma amostra aleatória de uma distribuição p-dimensional com os quartos momentos finitos, média μ_i e matriz de covariância Σ_i. O interesse dos autores está na hipótese de homogeneidade:

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k \\ \quad\quad\quad \text{vs} \\ H_1 : \Sigma_i \neq \Sigma_j \ \text{ para algum } i \neq j \end{cases} \qquad (2.3.2)$$

A matriz de covariância amostral para a i-ésima amostra é dada por:

$$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(X_{ij} - \hat{\mu}_i \right) \left(X_{ij} - \hat{\mu}_i \right)^T \qquad (2.3.3)$$

sendo $\hat{\mu}_i$ qualquer estimador de μ_i, podendo ser, inclusive, a média amostral. A matriz de covariância ponderada é:

$$S_p = \frac{1}{n} \sum_{i=1}^{k} n_i S_i \,, \text{ sendo } n = \sum_{i=1}^{k} n_i \qquad (2.3.4)$$

Zhu, Ng e Jing (2002) basearam-se na idéia das comparações múltiplas (Dunnett, 1994) e construíram testes por meio das combinações pareadas. As comparações pareadas entre a l-ésima e a i-ésima amostras foram baseadas em:

$$M_{li} = \max\left\{\text{autovalores absolutos de } \sqrt{\frac{n_l n_i}{n}} S_p^{-1/2}\left(S_l - S_i\right) S_p^{-1/2}\right\},$$

$$A_{li} = \text{média } \left\{\text{autovalores absolutos de } \sqrt{\frac{n_l n_i}{n}} S_p^{-1/2}\left(S_l - S_i\right) S_p^{-1/2}\right\}$$

Também utilizaram o procedimento de reamostragem ponderada sugerida por Zhang e Boos (1992). A proposta foi usar a média de $k\left(k-1\right)/2$ comparações pareadas como a estatística do teste:

$$LM = \frac{2}{k\left(k-1\right)} \sum_{i<l} M_{li},$$

$$LA = \frac{2}{k\left(k-1\right)} \sum_{i<l} A_{li}$$

Quando as k amostras tinham o mesmo tamanho, n, Zhu, Ng e Jing (2002) sugeriram outro teste condicional (simetrização aleatória), procedimento que é de mais fácil implementação. O inconveniente desse procedimento é a restrição de tamanhos de amostras iguais. O teste da permutação pode ser aplicado para tamanhos de amostras desiguais. Tem também algumas vantagens com relação ao bootstrap, mas é mais difícil de implementar do que a simetrização aleatória. É fácil observar que, similar à simetrização aleatória, quando todas as variáveis na amostra são i.i.d., a validade exata dos testes de permutação pode ser checada.

De forma análoga à simetrização aleatória, a validade exata dos testes de permutação, para o caso das médias, pode ser obtida. De fato, sob a hipótese nula, a permutação com as partes somadas tem a mesma distribuição do teste estatístico original.

Zhu, Ng e Jing (2002) compararam os estudos de Monte Carlo dos três procedimentos usando três famílias de distribuições multivariadas: multinormal $N(0,I_d)$, distribuição t multivariada $MT(5;0,I_d)$ e a distribuição normal contaminada $NC_2(0,I_d)$ cujos componentes são independentes, $N(0,1)$, com probabilidade 0,9 e uma $\chi^2_{(2)}$ com probabilidade 0,1. Consideraram $k=2$ e $k=6$, e as dimensões do vetor aleatório $p=2$ e $p=5$. O nível nominal adotado foi de 5%. Em cada procedimento, o número de repetições para calcular o valor crítico foi $r=500$. Cada uma das proporções atuais de rejeição de H_0 foi baseada em 1.000 simulações. Como esperado, os testes apresentaram-se melhores quando as médias eram conhecidas. Os autores somente relataram resultados para o caso de uma média desconhecida. Como se pode observar na Tabela 3, a proporção de rejeição pela permutação (PERM) é mais próxima do valor nominal α do que o bootstrap (BOOT) e, neste aspecto, a simetrização aleatória (RAS) é comparável ao bootstrap (BOOT). Para diferentes tamanhos de amostras, em que a RAS não é avaliada, os resultados referentes ao teste BOOT e ao teste PERM constam da Tabela 4. As tabelas mostram que PERM é melhor do que BOOT em 8 de 12 simulações. Comparando LA com LM, os autores encontraram que, para tamanhos de amostras iguais, quando $k=2$ LA é pior que LM na maioria das vezes; quando $k=6$, LA é melhor em todos os casos. Para tamanhos de amostras diferentes, LA é melhor que LM na maioria dos casos.

TABELA 3 Porcentagem de vezes que H_0 (2.3.1) é rejeitada, considerando as situações de $k = 2$ e $n_1 = n_2 = 20$; $k = 6$ e $n_i = 20$, para $p = 2$ e $p = 5$, usando três famílias de distribuições multivariadas: multinormal, distribuição t multivariada e a distribuição contaminada para o nível nominal de 5% de probabilidade.

		$k=2,\ n_1=n_2=20$								
		$N(0,I_d)$			$MT(5;0,I_d)$			$NC_2(0,I_d)$		
		RAS	**BOOT**	**PERM**	**RAS**	**BOOT**	**PERM**	**RAS**	**BOOT**	**PERM**
$p=2$	*LM*	0,053	0,045	0,048	0,053	0,046	0,054	0,055	0,046	0,056
	LA	0,059	0,057	0,054	0,062	0,046	0,057	0,063	0,060	0,061
$p=5$	*LM*	0,055	0,045	0,056	0,050	0,044	0,051	0,057	0,039	0,053
	LA	0,054	0,047	0,054	0,059	0,031	0,060	0,063	0,041	0,059
		$k=6,\ n_i=20,\ i=1,...,6$								
		$N(0,I_d)$			$MT(5;0,I_d)$			$NC_2(0,I_d)$		
		RAS	**BOOT**	**PERM**	**RAS**	**BOOT**	**PERM**	**RAS**	**BOOT**	**PERM**
$p=2$	*LM*	0,033	0,062	0,060	0,032	0,064	0,059	0,033	0,061	0,058
	LA	0,053	0,058	0,057	0,060	0,060	0,056	0,057	0,060	0,055
$p=5$	*LM*	0,040	0,063	0,059	0,043	0,058	0,056	0,043	0,057	0,063
	LA	0,056	0,060	0,054	0,049	0,045	0,053	0,055	0,054	0,055

Como apresentado por Zhang e Boos (1992), o procedimento do teste de homogeneidade para Bartlett apresentou-se pior quando a dimensão p aumentou. O procedimento bootstrap dos testes de Zhu, Ng e Jing (2002) são bastante estáveis. Quando $n_1 = n_2 = 20, \alpha = 0{,}05\ e\ p = 2$, as proporções de rejeição usadas pelos autores para as três distribuições foram 0,045, 0,046 e 0,046, contra 0,046, 0,045 e 0,50, respectivamente. Mas, quando a dimensão aumentou para $p = 5$, os resultados obtidos por Zhu, Ng e Jing (2002) foram 0,045, 0,044 e 0,039, contra 0,012, 0,023 e 0,019 de Zhang e Boos (1992), respectivamente, ou seja, os resultados foram mais favoráveis ao estudo de Zhu, Ng e Jing (2002).

TABELA 4 Porcentagem de vezes que H_0 (2.3.1) é rejeitada, considerando as situações de $k = 2$ e $n_1 = n_2 = 20$; $k = 6$ e $n_i = 20$ para $p = 2$ e $p = 5$, usando três famílias de distribuições multivariadas: multinormal, distribuição t multivariada e a distribuição contaminada para o nível nominal de 5% de probabilidade.

	$k=2,\ n_1=20,\ n_2=40$					
	$N(0,I_d)$		$MT(5;0,I_d)$		$NC_2(0,I_d)$	
	BOOT	**PERM**	**BOOT**	**PERM**	**BOOT**	**PERM**
LM	0,042	0,041	0,046	0,045	0,058	0,055
LA	0,054	0,053	0,058	0,056	0,054	0,053
LM	0,047	0,049	0,053	0,055	0,064	0,060
LA	0,055	0,052	0,049	0,048	0,053	0,053
	$k=6,\ n_1=n_2=20, n_3=n_4=30, n_5=n_6=40$					
	$N(0,I_d)$		$MT(5;0,I_d)$		$NC_2(0,I_d)$	
	BOOT	**PERM**	**BOOT**	**PERM**	**BOOT**	**PERM**
LM	0,060	0,057	0,045	0,043	0,054	0,052
LA	0,056	0,055	0,041	0,045	0,055	0,054
LM	0,066	0,062	0,065	0,063	0,064	0,059
LA	0,060	0,057	0,043	0,045	0,045	0,055

O poder dos testes também foi estudado para $k = 2$ populações em amostras de tamanhos $n_1 = n_2 = 20$, com dimensão $p = 2$. As distribuições multinormal e t multivariada, $N(\mu, \Sigma)$ e $MT(5; 0, \Sigma)$ foram usadas para a geração dos dados. Foram utilizados os pares da matriz identidade I_2 com C_2 e com V_2, respectivamente. As matrizes C_2 e V_2 foram:

$$C_2 = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}, \qquad V_2 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

Os resultados estão apresentados na Tabela 5. Para comparação com o teste de Bartlett estudado por Zhang e Boos (1992), Zhu, Ng e Jing (2002) também calcularam o poder ajustado. A Tabela 5 mostra que o procedimento bootstrap tem poder mais alto quando a distribuição é normal, mas perde poder se comparado com os testes de permutação e simetrização aleatória, no caso da distribuição t multivariada. Além disso, LA tem melhor performance do que LM, na maioria das vezes.

Os resultados dos estudos do poder do procedimento bootstrap são mais favoráveis aos procedimentos adotados por Zhu, Ng e Jing (2002). Em síntese Zhu, Ng e Jing (2002) sugerem as seguintes recomendações: (1) o valor médio do teste é melhor que o teste do valor máximo; (2) os testes LM e LA são preferíveis ao teste de Bartlett, com ou sem bootstrap; (3) se os tamanhos de amostras forem iguais, usar o procedimento de simetrização aleatória, por ser de fácil implementação, embora sua performance de poder seja levemente pior; (4) se os tamanhos de amostras forem diferentes, o procedimento de permutação é uma boa escolha.

TABELA 5 Estudo do poder para $p = 2$ dimensões, $k = 2$ populações, tamanhos amostrais $n_1 = n_2 = 20$, considerando os procedimentos de simetrização aleatória, bootstrap e permutação sob as distribuições normal multivariada e t multivariada e valor nominal de 5% de significância.

	$N(0,C_2) \& N(0,I_2)$			$N(0,V_2) \& N(0,I_2)$		
	RAS	**BOOT**	**PERM**	**RAS**	**BOOT**	**PERM**
LM	0,193	0,227	0,193	0,753	0,781	0,762
$LM(a-p)$	0,229	0,246	0,223	0,815	0,825	0,817
LA	0,256	0,276	0,265	0,784	0,770	0,770
$LA(a-p)$	0,296	0,299	0,297	0,817	0,818	0,814
	$MT(5;0,C_2) \& MT(5;0,I_2)$			$MT(5;0,V_2) \& MT(5;0,I_2)$		
	RAS	**BOOT**	**PERM**	**RAS**	**BOOT**	**PERM**
LM	0,228	0,231	0,233	0,553	0,521	0,561
$LM(a-p)$	0,261	0,262	0,265	0,636	0,604	0,823
LA	0,230	0,229	0,234	0,570	0,538	0,589
$LA(a-p)$	0,274	0,270	0,275	0,640	0,617	0,637

2.4 Homogeneidade de variâncias

Problemas de comparação de variâncias sempre ocorrem em vários campos: minimizar e controlar a variabilidade são importantes na produção e controle de qualidade (em que se enfatiza a variabilidade lote a lote, em função do processo e das técnicas de medição); na indústria farmacêutica, onde a variabilidade de efeitos e de reações é de interesse e, em muitas outras áreas de pesquisa, tais como a da evolução (nas quais as relações genéticas são estudadas). Nessas situações e em muitas outras, a variância populacional é o parâmetro de interesse.

De acordo com Neyman e Pearson (1931), testes para igualdade de variâncias são de interesse em muitas situações, tais como análise de variância ou controle de qualidade. A abordagem clássica para testar hipóteses começa com o teste de razão de verossimilhança sob a suposição de distribuição normal. Entretanto, como esse teste é muito sensível à falta de normalidade, muitos testes alternativos foram propostos desde então. Alguns desses testes são modificações do teste de razão de verossimilhança. Outros são adaptações do teste F para testar variâncias em vez de médias.

Ao contrário do que se esperava, estudos comparativos recentes revelaram que alguns destes testes apresentam falta de robustez e baixo poder (Layard, 1973). Este problema foi também estudado por Conover et al. (1981), que realizaram um interessante estudo de muitos testes paramétricos e não paramétricos existentes. Simulações Monte Carlo de algumas distribuições para vários tamanhos de amostras mostraram que poucos testes são robustos e poderosos.

Muitos dos testes paramétricos e não paramétricos para homogeneidade de variâncias e algumas variações desses testes foram examinados por Conover et al. (1981). O propósito de seu estudo foi proporcionar uma lista de testes que tenham taxas de erro tipo I estáveis quando a suposição de normalidade pode não ser verdadeira e quando os tamanhos amostrais podem ser pequenos e desiguais.

A importância do teste de homogeneidade de variâncias é baseada na premissa de que muitos dos testes de hipóteses sobre médias ou efeitos de tratamentos são realizados pressupondo-se que as variâncias das populações amostradas sejam iguais. A violação dessa hipótese pode afetar a performance do teste de diferentes formas (Johnson e Wichern, 1998). Essa importância dos testes na inferência estatística fica evidenciada nos exemplos destacados anteriormente. Assim, pode-se destacar o teste de Bartlett para comparar variâncias como um dos mais freqüentemente encontrado nos livros textos e ensinado nos cursos introdutórios de estatística, devido à sua facilidade de computação. O teste estatístico proposto, cuja estatística é denominada (B), envolve a comparação das somas de quadrados entre e dentro de amostras de k populações. Os seguintes símbolos são usados nas descrições estatísticas: k é o numero de grupos ou populações amostradas, n_j é o numero total de observações do j-ésimo grupo, s_p^2 é a variância combinada, S_i^2 é o estimador da variância do i-ésimo grupo dos k grupos.

$$B = (n-k) ln\, s_i^2 - \sum_{j=1}^{k} (n_j - 1) ln\, s_p^2$$

em que,

$$s_p^2 = SSW / (n-k)$$

e,

$$SSW = \sum_{i=1}^{K} \sum_{j=1}^{n_j} \left(X_{ij} - \overline{X}_{i.} \right)^2$$

um fator de correção C_B é computado:

$$C_B = 1 + \frac{1}{3(k-1)} \left[\sum_{j=1}^{k} \frac{1}{(n_j - 1)} - \frac{1}{(n-k)} \right]$$

e aplicado em B, para se obter a estatística do teste corrigido por:

$$B_c = B / C_B$$

B_c tem uma distribuição assintótica de $\chi^2_{(k-1)}$ com $k-1$ graus de liberdade. Sabe-se que o teste de Bartlett é poderoso se as populações amostradas são normais, mas é fortemente afetado pela não normalidade (Box, 1949).

No teste de igualdade de k variâncias, Bartlett (1937), primeiro. notou, sob a hipótese nula de variâncias iguais, que a distribuição qui-quadrado com $k-1$ graus de liberdade melhor se ajustou à distribuição da estatística de teste de razão de verossimilhança multiplicada por um fator escalar ($1/C_B$) do que àquela da estatística de teste original. Desde então, correções para uma melhor aproximação da distribuição nula da estatística do teste da razão de verossimilhança por distribuições qui-quadrado receberam considerável atenção na literatura (Bartlett, 1954), entre outros. Para testar a igualdade de variâncias normais em um teste unilateral, sob a hipótese nula a estatística de teste de razão da verossimilhança é assintoticamente distribuída como uma mistura de k distribuições de qui-quadrado (Robertson et al., 1988).

A performance do teste de Bartlett e da modificação do teste de Bartlett proposta por Dixon e Massey (1969) foi avaliada e, no caso de duas amostras, foi comparada com o teste F. O teste de Bartlett foi usado quando mais de duas amostras eram independentes e não foi considerado bom quando as amostras eram de tamanhos diferentes. No caso, a modificação do teste de Bartlett proposta por Dixon e Massey (1969) pode ser considerada boa para amostras de tamanhos iguais e diferentes. Embora alternativas existam, elas são apresentadas, em geral, para o caso univariado.

2.5 Inferência multivariada

De acordo com Nogueira (2004), pelo fato dos fenômenos serem estudados a partir de dados coletados ou mensurados em muitas variáveis, os métodos estatísticos delineados para obter informações a partir destes conjuntos de variáveis são denominados de métodos de análise multivariados. A partir desse conceito, pode-se ter a generalização de densidades multivariadas, como, por exemplo, a distribuição normal, que desempenha um papel fundamental. Segundo Johnson & Wichern (1998), pode-se escrever a distribuição normal multivariada da seguinte forma:

$$f\left(\underset{\sim}{x};\underset{\sim}{\mu},\Sigma\right)=\frac{1}{\left(2\pi\right)^{\frac{p}{2}}\left|\Sigma\right|^{\frac{1}{2}}}exp\left\{-\frac{1}{2}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^{T}\Sigma^{-1}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\right\},$$

sendo p o número de variáveis, $\underset{\sim}{x}$ o vetor de variáveis observadas $\left(1\times p\right)$, $\underset{\sim}{\mu}$ um vetor $\left(p\times 1\right)$ dos valores esperados do vetor $\underset{\sim}{X}$; Σ representa a matriz de covariâncias $\left(p\times p\right)$ do vetor $\underset{\sim}{X}$ e T o símbolo indicador de transposição.

De acordo com Bock (1975), uma propriedade da distribuição normal multivariada é que todos os subconjuntos de $\underset{\sim}{X}$ têm distribuição normal multivariada. Tomando-se uma partição ${}_{p}\underset{\sim}{X}_{1}=\left[\frac{{}_{q}\underset{\sim}{X}_{1}}{{}_{(p-q)}\underset{\sim}{X}_{1}}\right]=\left[\frac{\underset{\sim}{X}_{1}}{\underset{\sim}{X}_{2}}\right]$ e suas correspondentes partições no vetor da média e covariância, tem-se:

$${}_{p}\underset{\sim}{\mu}_{1}=\left[\frac{{}_{q}\underset{\sim}{\mu}_{1}}{{}_{(p-q)}\underset{\sim}{\mu}_{1}}\right]=\left[\frac{\underset{\sim}{\mu}_{1}}{\underset{\sim}{\mu}_{2}}\right]\text{ e }\Sigma=\left(\begin{array}{c|c}{}_{q}\Sigma_{11q} & {}_{q}\Sigma_{12q}\\ \hline {}_{q}\Sigma_{21q} & {}_{(p-q)}\Sigma_{22(p-q)}\end{array}\right),$$

sendo $\underset{\sim}{X}_{1}\sim N_{p}\left(\underset{\sim}{\mu}_{1},\Sigma_{11}\right)$, $\underset{\sim}{X}_{2}\sim N_{p}\left(\underset{\sim}{\mu}_{2},\Sigma_{22}\right)$ para $q<p$, sendo ${}_{q}\Sigma_{11q}$ e ${}_{(p-q)}\Sigma_{22(p-q)}$ matrizes de covariâncias de $\underset{\sim}{X}_{1}$ e $\underset{\sim}{X}_{2}$, respectivamente.

2.6 Inferência bayesiana

Segundo Nogueira (2004), o foco fundamental da estatística é a inferência. Alguns dados são observados e deseja-se fazer inferência sobre uma ou mais características desconhecidas do sistema que teria dado origem a estes dados.

A inferência estatística depende da presença das observações y, cujos valores inicialmente incertos são descritos por meio de uma distribuição de probabilidade $f\left(y/\theta\right)$. O parâmetro θ é o que interessa ao pesquisador e tem um sentido exato dentro da inferência bayesiana. O pesquisador deve ter alguma informação prévia sobre θ e esta informação deve ser incorporada na análise.

Para Gelman et al. (1997), a inferência bayesiana é o processo de ajustar um modelo de probabilidade para um grupo de dados e resumir o resultado por uma distribuição de probabilidade nos parâmetros do modelo e em quantidades não observáveis, como predição para novas observações.

A inferência bayesiana é baseada no conceito de probabilidade subjetiva, que mede o grau de desconhecimento que se tem da ocorrência de um determinado evento do espaço amostral. Se existe uma incerteza, a descrição desta incerteza deve ser feita utilizando-se a probabilidade subjetiva. Então, a análise bayesiana descreve toda a quantidade desconhecida por meio da probabilidade. Ao descrever uma incerteza usando a interpretação subjetivista de probabilidade, deve-se levar em conta o axioma da coerência estabelecido por De Finetti (1964).

Ao incorporar uma opinião sobre θ à análise por meio de uma densidade de probabilidade $P\left(\theta\right)$, o bayesiano está utilizando a densidade a priori. A densidade a priori possui este nome por ser a distribuição de probabilidade de θ antes que se observem os dados. O parâmetro θ pode ser um escalar ou um vetor.

A densidade conjunta de um grupo de observações $y_1,...,y_n$ expressa como uma função do parâmetro, é denominada função de verossimilhança e é representada por $L\left(y_1,...,y_n \mid \theta\right)$, sendo n o número de observações.

Se $P\left(\theta\right)$ é a densidade a priori para θ, então, a densidade a posteriori de θ é dada pelo Teorema de Bayes, ou seja,

$$P\left(\theta \mid Y\right) = \frac{L\left(Y \mid \theta\right) P\left(\theta\right)}{\int L\left(Y \mid \theta\right) P\left(\theta\right) d\theta},$$

sendo $Y = \left\{y_1,...,y_n\right\}$. Como o denominador não depende de θ, pode-se escrever:

$$P\left(\theta \mid Y\right) \propto L\left(Y \mid \theta\right) P\left(\theta\right),$$

em que $\propto$ representa proporcionalidade.

Pode-se pensar no Teorema de Bayes como um mecanismo de atualização da opinião do estatístico sobre θ. O Teorema de Bayes é um elemento essencial para análise bayesiana, pois toda inferência é feita a partir da posteriori.

De acordo com Samiuddin (1976), testar a igualdade de variâncias é de interesse em muitas aplicações estatísticas. Quando a origem da população é normal, o teste F é usado para o caso de duas amostras e o teste de Bartlett é usado quando existem mais de duas amostras independentes. O autor supôs $\mathrm{X}_{ij}\left(j = 1, 2, ..., n_i;\ i = 1, 2, ..., k\right)$ variáveis aleatórias independentes normais com média μ_i e variância σ_i^2. Foram considerados $S_i^2 = \sum_i \left(\mathrm{X}_{ij} - \overline{\mathrm{X}}_i\right)^2$ e

$s_i^2 = S_i^2 \big/ v_i$, sendo $\overline{\mathrm{X}}_i = \sum_i \mathrm{X}_{ij} \big/ n_i$ e $v_i = n_i - 1$. Para o caso de $k = 2$, um teste de igualdade de variâncias exato usa a razão de variâncias $F = s_1^2 \big/ s_2^2$. Quando $k > 2$, usa-se normalmente:

$$\left[v \ln s^2 - \sum_{i=1}^{k} v_i \ln s_i^2 \right] \Bigg/ \left\{ 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^{k} \frac{1}{v_i} - \frac{1}{v} \right) \right\} \qquad (2.6.1)$$

como uma χ^2 com $k-1$ graus de liberdade, sendo $s^2 = \left(\sum_{i=1}^{k} v_i s_i^2 \right) \Big/ v$, e $v = \sum_{i=1}^{k} v_i$. O denominador em (1.1) é um fator de correção do teste de Bartlett e usualmente melhora a aproximação.

Dixon e Massey (1969) deram uma outra versão para o teste de Bartlett. Eles sugeriram tratar $l M \big/ (b - M)$ como uma F com $k-1$ e l graus de liberdade, sendo:

$$M = v \ln s^2 - \sum_{i=1}^{k} v_i \ln s_i^2,$$

$$A = \frac{1}{3(k-1)} \left(\sum_{i=1}^{k} \frac{1}{v_i} - \frac{1}{v} \right),$$

$$l = 3(k-1) \big/ A^2,$$

e,

$$b = l \big/ (1 - A + 2/l)$$

Samiuddin (1976) denominou estes testes de teste de Bartlett 1 e teste de Bartlett 2, respectivamente. Sabe-se que o teste de Bartlett é sensível à não normalidade. Diferentes autores, tais como Layard (1973), Levene (1960), Miller (1968), e Gartside (1972), sugeriram testes para situações não-normais.

O teste bayesiano de homogeneidade de variâncias de Samiuddin (1976) inicia-se com a usual distribuição a priori não informativa de μ_i e:

$$\sigma_i \left(i = 1, 2, ..., k\right), \qquad (2.6.2)$$

$$\pi\left(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k\right)$$
$$\propto d\mu_1 d\mu_2 ... d\mu_k \frac{d\sigma_1 d\sigma_2 ... d\sigma_k}{\sigma_1 \sigma_2 ... \sigma_k}$$

A verossimilhança é proporcional a:

$$\prod_{i=1}^{k} (1/\sigma_i) \exp\left[-\frac{1}{2}\sum_j \left(X_{ij} - \mu_i\right)^2 \Big/ \sigma_i^2\right] \qquad (2.6.3)$$

O produto de (2.6.1) e (2.6.2) gera a distribuição a posteriori conjunta de médias μ_i e variâncias σ_i^2. Integrando-se os $\mu_i's$ chega-se na distribuição a posteriori marginal dos $\sigma_i's$. Isso implica na independência a posteriori marginal dos σ_i^2's e que $S_i^2 \big/ \sigma_i^2 \left(i = 1, 2, ..., k\right)$ tem distribuição χ^2 com v_i graus de liberdade.

Para S_i^2 / σ_i^2, foi aplicada a transformação de Wilson-Hilferty (1931), que é surpreendentemente precisa, mesmo para $v_i's$ pequenos dada por:

$$\hat{\chi}_\alpha^2 = v\left[C_\alpha \left(2/9v\right)^{\frac{1}{2}} + \left(1 - 2/9v\right)\right]^3,$$

sendo C_α o quantil $100\alpha\%$ de uma normal padrão e v são os graus de liberdade da χ^2. O ponto percentual exato da χ^2 é chamado χ^2_α.

Com uma simples manipulação dessa expressão, substituindo $\hat{\chi}^2_\alpha$ por S_i^2 / σ_i^2 e $v = v_i$, obtém-se:

$$C_\alpha = \frac{\left(\frac{s_i^2}{\sigma_i^2}\right)^{1/3} - \left(1 - \frac{2}{9v_i}\right)}{\sqrt{\frac{2}{9v_i}}}$$

Isto sugere que $\left(s_i^2 / \sigma_i^2\right)^{1/3}$ segue aproximadamente uma distribuição normal com média $1 - 2/9v_i$ e variância $2/9v_i$. Neste estágio, Samiuddin (1976) enfatiza que σ_i^2 é a variável aleatória e s_i^2 é fixa. Denominou ϕ_i por $\left(1/\sigma_i^2\right)^{1/3}$ e, após algumas simplificações, obteve:

$$C_\alpha = \frac{\phi_i - \left(1 - \frac{2}{9v_i}\right)\left(\frac{v_i}{S_i^2}\right)^{1/3}}{\sqrt{\frac{2}{9v_i\left(S_i^2\right)^{2/3}}}}$$

de onde concluiu que ϕ_i tem distribuição normal com média $m_i = \left(v_i / S_i^2\right)^{1/3}\left[1 - 2/9v_i\right]$ e variância $a_i^2 = 2 \Big/ \left[9\left(S_i^2\right)^{2/3} v_i^{1/3}\right]$. Então, sob a hipótese nula de igualdade de variância, obteve a estatística do teste por:

$$\chi_s^2 = \sum_{i=1}^{k} \frac{(m_i - m)^2}{a_i^2}$$

em que,

$$m = \frac{\sum \frac{m_i}{a_i^2}}{\sum \frac{1}{a_i^2}}$$

com $k-1$ graus de liberdade.

3 METODOLOGIA

Foram considerados quatro testes neste trabalho: teste de Bartlett proposto para o caso multivariado (TBM), o teste de Samiuddin univariado generalizado para o caso multivariado neste trabalho (TGS) e suas respectivas versões bootstrap (TBMB, TGSB).

Seja $\underset{\sim}{X}_{i1}, \underset{\sim}{X}_{i2}, ..., \underset{\sim}{X}_{ij}, ..., \underset{\sim}{X}_{in_i}$ uma amostra aleatória de tamanho n_i da i-ésima população em que $\underset{\sim}{X}_{ij}^T = \left[X_{ij1}, X_{ij2}, ..., X_{ijp}\right]$ é o vetor de observações amostrais de dimensão p da j-ésima unidade amostral e i-ésima população para $i = 1, 2, ..., k$ e $j = 1, 2, ..., n_i$. Considere ainda que a i-ésima população tenha média $\underset{\sim}{\mu}_i$ $(1 \times p)$ e covariância Σ_i $(p \times p)$ e $n = \sum_{i=1}^{k} n_i$.

A hipótese de interesse é dada por:

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma \\ \qquad\qquad \text{vs} \\ H_1 : \Sigma_i \neq \Sigma \text{ para algum } i = 1, 2, \ldots, k \end{cases} \tag{3.1}$$

3.1 Testes assintóticos

Nas próximas seções serão descritos os testes estudados neste trabalho.

3.1.1 Teste de Bartlett multivariado (TBM)

Sob a hipótese alternativa H_1 de (3.1) e sob normalidade multivariada, a verossimilhança é dada por:

$$L = \left(2\pi\right)^{-\frac{np}{2}} \prod_{i=1}^{k} \left|\Sigma_i\right|^{-\frac{n_i}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\left(\underset{\sim}{X}_{ij} - \underset{\sim}{\mu}_i\right)^T \Sigma_i^{-1}\left(\underset{\sim}{X}_{ij} - \underset{\sim}{\mu}_i\right)\right]\right\} \tag{3.2}$$

A função de verossimilhança (3.2) deve ser maximizada. O usual é tomar o logaritmo e maximizar a função resultante denominada de logaritmo da função de verossimilhança $\left(\ln L\right)$. Obtêm-se as derivadas parciais em relação à Σ_i e $\underset{\sim}{\mu}_i$ e, a partir delas, obtêm-se os estimadores de máxima verossimilhança por:

$$\begin{cases} \hat{\underset{\sim}{\mu}}_i = \overline{\underset{\sim}{X}}_{i.} = \dfrac{\sum_{j=1}^{n_i} \underset{\sim}{X}_{ij}}{n_i} \\ \hat{\Sigma}_i = \dfrac{\sum_{j=1}^{n_i} \left(\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_{i.}\right)\left(\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_{i.}\right)^T}{n_i} \end{cases} \tag{3.3}$$

para $i = 1, 2, ..., k$.

Sob H_0 e normalidade, a função de verossimilhança é:

$$L_0 = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\underset{\sim}{X}_{ij} - \underset{\sim}{\mu}_i\right)^T \Sigma^{-1}\left(\underset{\sim}{X}_{ij} - \underset{\sim}{\mu}_i\right)\right\} \tag{3.4}$$

Procedeu-se da mesma forma anterior e obtiveram-se os estimadores de $\underset{\sim}{\mu}_i$ idênticos aos da equação (3.3) e o estimador da covariância comum (Σ) por:

$$\hat{\Sigma} = \frac{\sum_{i=1}^{k} n_i \hat{\Sigma}_i}{n} \tag{3.5}$$

É conveniente substituir os estimadores de máxima verossimilhança das matrizes de covariâncias por:

$$
\begin{cases}
S_i = \dfrac{n_i \hat{\Sigma}_i}{n_i - 1} \\
\\
S_p = \dfrac{\sum_{i=1}^{k} (n_i - 1) S_i}{n - k}
\end{cases}
\qquad (3.6)
$$

O teste de razão de verossimilhança foi obtido tomando-se o máximo de (3.2) e dividindo-o pelo máximo de (3.4), utilizando ainda os estimadores não viesados. Obteve-se a razão de verossimilhança modificada por:

$$
\Lambda = \frac{\prod_{i=1}^{k} |S_i|^{\frac{n_i - 1}{2}}}{|S_p|^{\frac{n-k}{2}}} \qquad (3.7)
$$

Como $-2 \ln \Lambda$ tem uma distribuição assintoticamente de χ^2 (Mood, Grybil & Boes, 1974), a estatística do teste de TBM é dada por:

$$
\chi_c^2 = (n-k) \ln |S_p| - \sum_{i=1}^{k} (n_i - 1) \ln |S_i| \qquad (3.8)
$$

com $\nu = (k-1)\, p\,(p+1)/2$. Os graus de liberdade são obtidos comparando-se a dimensão da função verossimilhança irrestrita $\left[pk + p(p+1)k/2 \right]$ com a função de verossimilhança restrita $\left[pk + p(p+1)/2 \right]$. Bartlett (1937) propôs uma correção para melhorar as propriedades assintóticas da aproximação. A estatística do teste com a modificação proposta usada nesse trabalho é:

$$\chi_1^2 = \left[1 - \left(\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{n-k}\right)\left(\frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\right)\right] \times \qquad (3.9)$$
$$\left[(n-k)\ln\left|S_p\right| - \sum_{i=1}^{k}(n_i - 1)\ln\left|S_i\right|\right]$$

3.1.2 Teste generalizado de Samiuddin (TGS)

Da função de verossimilhança (3.2) da distribuição normal sabe-se (Johnson e Wichern, 1998) que:

$$\left(\underset{\sim}{\mathrm{X}}_{ij} - \underset{\sim}{\mu}_i\right)^T \Sigma_i^{-1} \left(\underset{\sim}{\mathrm{X}}_{ij} - \underset{\sim}{\mu}_i\right) \qquad (3.10)$$

tem distribuição qui-quadrado com p graus de liberdade. Esse resultado é a base para a generalização multivariada do teste de Samiuddin (Samiuddin, 1976) univariado. Utilizando-se o estimador de máxima verossimilhança de $\underset{\sim}{\mu}_i$, dado por $\bar{\underset{\sim}{\mathrm{X}}}_{i.}$ na expressão 3.10, obtém-se:

$$\left(\underset{\sim}{\mathrm{X}}_{ij} - \bar{\underset{\sim}{\mathrm{X}}}_{i.}\right)^T \Sigma_i^{-1} \left(\underset{\sim}{\mathrm{X}}_{ij} - \bar{\underset{\sim}{\mathrm{X}}}_{i.}\right) \qquad (3.11)$$

que tem distribuição assintótica qui-quadrado com ν graus de liberdade.

Samiuddin (1976) integrou a verossimilhança (3.2) com relação aos $\underset{\sim}{\mu}_i$'s para o caso univariado $(p = 1)$ e obteve que a distribuição posteriori de $\sum_{j=1}^{n_i}\left(\mathrm{X}_{ij} - \bar{\mathrm{X}}_{i.}\right)^2 / \sigma_i^2$ é uma distribuição qui-quadrado com $\nu_i = n_i - 1$ graus de liberdade e independente dos μ_i's. A generalização para o caso multivariado pode ser feita a partir do expoente da função exponencial de (3.2), substituindo-se $\underset{\sim}{\mu}_i$ pelo estimador $\overline{\underset{\sim}{X}}_{i.}$ de máxima verossimilhança da seguinte forma:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T \Sigma_i^{-1}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\right] \qquad (3.12)$$

O argumento das somas de (3.12) são escalares e, portanto, pode-se aplicar a função traço. Utilizando-se as propriedades do traço: i) $tr(AB)=tr(BA)$ e ii) $tr(A+B)=tr(A)+tr(B)$, tem-se o seguinte desenvolvimento:

$$tr\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T \Sigma_i^{-1}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\right]=$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i} tr\left[\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T \Sigma_i^{-1}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\right]=$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i} tr\left[\Sigma_i^{-1}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T\right]=$$

$$=\sum_{i=1}^{k} tr\left\{\sum_{j=1}^{n_i}\left[\Sigma_i^{-1}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T\right]\right\}=$$

$$=\sum_{i}^{k} tr\left\{\Sigma_i^{-1}\left[\sum_{j=1}^{n_i}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T\right]\right\}$$

Como $\sum_{j=1}^{n_i}\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)\left(\underset{\sim}{X}_{ij}-\overline{\underset{\sim}{X}}_{i.}\right)^T$ é igual a $(n_i-1)S_i=\nu_i S_i$, então se obtém:

$$\sum_{i=1}^{k} tr\left(\nu_i \, \Sigma_i^{-1} \, S_i\right) = \sum_{i}^{k} Q_i$$

logo,

$$Q_i = tr\left(\nu_i \, \Sigma_i^{-1} \, S_i\right) \tag{3.13}$$

possui distribuição "assintótica" de qui-quadrado com $\nu_i p$ graus de liberdade. Os graus de liberdade são resultantes da soma de n_i variáveis qui-quadrado com p graus de liberdade, sendo apenas $\nu_i = n_i - 1$ independentes.

Considerando a equação (3.13), pode-se usar o mesmo argumento de Samiuddin (1976) para propor a estatística final do teste. Em um primeiro estágio, foi utilizada a aproximação normal de Wilson e Hilferty (1931) para a distribuição qui-quadrado. Posteriormente, foi imposta a hipótese e utilizado o resultado de que se Z_i são variáveis normais padrão independentes, $\sum_{i=1}^{k} Z_i^2$ tem distribuição qui-quadrado com k graus de liberdade (Mood, Grybil & Boes, 1974). A aproximação de Wilson-Hilferty é dada por:

$$Z \approx \frac{\sqrt[3]{\dfrac{\chi_\nu^2}{\nu}} - \left(1 - \dfrac{2}{9\nu}\right)}{\sqrt{\dfrac{2}{9\nu}}} \tag{3.14}$$

em que Z é uma normal padrão e χ_ν^2 é uma distribuição qui-quadrado com ν graus de liberdade.

Sob a hipótese nula $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma$ e considerando que Σ é desconhecida, mas cujo estimador é S_p, pode-se obter:

$$Z_i = \frac{\sqrt[3]{\dfrac{tr\left(\nu_i S_i S_p^{-1}\right)}{\nu_i p}} - \left(1 - \dfrac{2}{9\nu_i p}\right)}{\sqrt{\dfrac{2}{9\nu_i p}}} \tag{3.15}$$

logo,

$$\chi_2^2 = \sum_{i=1}^{k} Z_i^2 = \sum_{i=1}^{k} \left[\frac{\sqrt[3]{\dfrac{tr\left(\nu_i S_i S_p^{-1}\right)}{\nu_i p}} - \left(1 - \dfrac{2}{9\nu_i p}\right)}{\sqrt{\dfrac{2}{9\nu_i p}}} \right]^2 \tag{3.16}$$

tem distribuição assintótica qui-quadrado com $k-1$ graus de liberdade. Os $k-1$ graus de liberdade são conseqüência direta de não se ter os k Z_i's independentes, por causa de S_p ser uma combinação linear dos S_i's. A estatística χ_2^2 em (3.16) com $k-1$ graus de liberdade foi usada como estatística do teste generalizado de Samiuddin (TGS).

3.2 Testes de bootstrap

Foram propostas as versões bootstrap dos dois testes assintóticos apresentados na seção 3.1. A imposição da hipótese H_0 foi feita combinando-se as k amostras multivariadas em uma única amostra de tamanho $n = \sum_{i=1}^{k} n_i$ e de dimensão p. Tomou-se o cuidado de estimar os desvios $\underset{\sim}{\varepsilon}_{ij} = \underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_{i.}$. Esses desvios não alteram a variabilidade da i-ésima amostra, por terem sofrido apenas uma translação. Esse cuidado foi tomado para evitar que possíveis diferenças nas médias populacionais afetassem os testes de comparação de covariâncias por bootstrap.

As versões bootstrap dos testes de Bartlett multivariado e generalizado de Samiuddin foram denominadas de TBMB e TGSB. A idéia básica é impor H_0 por meio da amostra combinada, reamostrar com reposição esta amostra, formando novas amostras de tamanho n_i, $i = 1, 2, \ldots, k$ preservando a dimensão p e aplicar as estatísticas das equações (3.9) e (3.16), para TBMB e TGSB, respectivamente. Os resultados obtidos em um número B de reamostragens são comparados com os valores obtidos nas amostras originais. Se a proporção de valores da distribuição de bootstrap que superaram os valores originais dos respectivos testes for inferior ao nível de significância nominal α adotado, a hipótese H_0 deve ser rejeitada. O algoritmo para os testes de bootstrap TBMB e TGSB é dado por:

1. calcular χ_1^2 e χ_2^2 nas amostras originais utilizando (3.9) e (3.16);
2. compor o conjunto $\Omega = \left\{ \underset{\sim}{\varepsilon}_{11}, \ldots, \underset{\sim}{\varepsilon}_{1n_1}, \ldots, \underset{\sim}{\varepsilon}_{k1}, \ldots, \underset{\sim}{\varepsilon}_{kn_k} \right\}$ de desvios p-variados, resultante da união de todas as observações amostrais;
3. reamostrar Ω com reposição compondo novas k amostras, cujos elementos são os valores de desvios $\left(\underset{\sim}{\varepsilon}_{ij}^{*} \right)$ de Ω, dados por $\Omega_i^{*} = \left\{ \underset{\sim}{\varepsilon}_{ij}^{*}, \quad j = 1, 2, \ldots, n_i \right\}$ para $i = 1, 2, \ldots, k$;

4. recalcular as matrizes de covariâncias S_i^* de bootstrap e partir de Ω_i;
5. calcular as estatísticas (3.9) e (3.16) e denominá-las χ^2_{1l} e χ^2_{2l} na l-ésima reamostragem de bootstrap, sendo $l = 1, 2, \ldots, B$;
6. repetir B vezes os passos de 3 a 5;
7. calcular os p-valores (p_m) de bootstrap por:

$$p_m = \frac{\sum_{l=1}^{B} \mathrm{I}\left(\chi^2_{ml} \geq \chi^2_m\right)}{B}$$

em que $m = 1$ ou 2 e $\mathrm{I}(\cdot)$ é uma função indicadora. Neste caso, p_1, representa o p-valor do TBMB e p_2, de TGSB.

O valor de $B = 600$ foi escolhido neste trabalho, devido ao grande esforço computacional requerido na avaliação da performance destes testes, realizada por simulação Monte Carlo que será descrita na seção (3.3). Outra razão para a adequação dessa escolha baseou-se no valor $B = 500$ utilizado por Zhang e Boos (1992).

3.3 Simulação Monte Carlo

A performance dos quatro testes foi avaliada por meio de simulação computacional Monte Carlo.

Foram avaliados o erro tipo I e o poder para determinar a qualidade do desempenho dos testes. Também foram consideradas duas situações distintas para a realização dessa avaliação. Na primeira, considerou-se a distribuição normal multivariada e, na segunda, distribuições multivariadas geradas a partir de variáveis uniformes ou gamas independentes. Nesta última situação, pretendeu-se avaliar a robustez dos testes quando a normalidade foi violada.

Para isso, várias situações foram simuladas variando-se o número de populações (k), os tamanhos amostrais $(n_i\text{'s})$ e o número de variáveis (p). Foram consideradas ainda três distribuições p-dimensionais: normal, gama e uniforme. Foi adotado o seguinte método para simular a j-ésima observação da i-ésima amostra populacional:

a. gerou-se p variáveis Z_r de uma das três distribuições consideradas, $r = 1, 2, ..., p$. Para gerar da distribuição nomal univariada utilizou-se o comando rannor (seed) do programa SAS® (SAS, 2000), em que seed é a semente do processo de geração de números aleatórios; da distribuição uniforme, o comando ranuni (seed); e para a gama, rangam (seed, λ). O valor de seed foi estipulado como zero, para que o sistema utilizasse um valor inicial dependente do horário. O parâmetro de forma λ da distribuição gama padrão foi fixado em 1,5, para se obter uma distribuição unimodal e assimétrica à direita;
b. as p variáveis i.i.d Z_r foram consideradas um vetor $\underset{\sim}{Z}^T = \left[Z_1, Z_2, ..., Z_p\right]$;
c. estipulou-se uma única estrutura da matriz de covariância para todas as k populações por:

$$\Sigma_i = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & 1 & \rho & \cdots & \rho^{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & & 1 \end{pmatrix} \quad (3.17)$$

No caso do erro tipo I, simulou-se k populações com a mesma matriz de covariância $\Sigma_i = \Sigma, \ \forall \ i = 1, 2, ..., k$, ou seja, efetuaram-se simulações sob a hipótese H_0 de 3.1. Para o poder, as matrizes $\Sigma_2, \Sigma_3, \cdots, \Sigma_k$ foram estipuladas como um produto de Σ_1, de acordo com um critério de interesse. Para isso Σ_1 foi estipulada de acordo com a expressão (3.17) e as demais covariâncias por:

$$\Sigma_1 = \delta_i^* \Sigma_i, \ \ i = 2, 3, ..., k \quad (3.18)$$

O critério utilizado para determinar δ_i^* foi baseado nas variâncias generalizadas de Σ_i, da seguinte forma:

$$\delta = \frac{|\Sigma_k|}{|\Sigma_1|} \quad (3.19)$$

O valor de δ foi fixado em 2, 4, 8 e 16. Assim, calculou-se uma quantidade auxiliar d dada por:

$$d = \frac{\delta - 1}{k - 1} \quad (3.20)$$

e estipulou-se δ^* por:

$$\delta_i^* = \left[1 + d\left(i-1\right)\right]^{\frac{1}{p}} \qquad (3.21)$$

Dessa forma, a razão entre a maior variância generalizada e a menor é igual a δ e as diferenças consecutivas entre as razões de variâncias generalizadas de cada população em relação à primeira é d. Para tanto, foram fixados valores de ρ em 0, 0,1, 0,5 e 0,9.

d. obteve-se o fator de Cholesky de cada matriz de covariância por:

$$\Sigma_i = F_i F_i^T$$

a matriz F_i é o fator de Cholesky de Σ_i e foi obtida utilizando-se o comando root $\left(F_i\right)$ do sistema SAS (SAS, 2000), $i = 1, 2, \ldots, k$;

e. o vetor $\underset{\sim}{X}_{ij}$ da i-ésima população e j-ésima unidade amostral foi obtido por uma simples transformação, dada por:

$$\underset{\sim}{X}_{ij} = F_i \underset{\sim}{Z} + \underset{\sim}{\mu}_i, \quad i = 1, 2, \ldots, k \text{ e } j = 1, 2, \ldots, n_i$$

para fins de simulação, os $\underset{\sim}{\mu}_i$'s foram considerados nulos $\left(\underset{\sim}{0}\text{'s}\right)$;

f. repetir os passos de (a) a (e), sem recalcular F_i no passo (d) até que os $n = \sum_{i=1}^{k} n_i$ elementos amostrais tenham sido simulados.

Para avaliar a performance dos testes foram simuladas várias combinações fatoriais dos parâmetros k, n_i e p sob as três distribuições retro mencionadas. Os valores de k foram 2, 5 e 10, os de $p = 2$, 3, 5 e 10 e tamanhos amostrais variando de 5 a 105. Foram feitas 600 simulações de cada configuração e de cada distribuição e ainda foram consideradas simulações sob H_0 e H_1. Aplicaram-se os testes TBM, TGS, TBMB e TGSB em cada caso. Adotaram-se dois níveis nominais de significância α $(0,01 \text{ e } 0,05)$. Os p-valores observados foram confrontados com os valores de α nominais. A proporção de resultados em que os p-valores de um teste foram inferiores ou iguais ao valor nominal α foi computada. No caso de simulações sob H_0, essa proporção refere-se à taxa de erro tipo I e sob H_1, ao poder. As situações não-normais foram utilizadas para avaliar a robustez dos testes, tanto no controle do erro tipo I quanto na capacidade de diferenciação das matrizes de covariâncias (poder). A situação normal foi utilizada para propiciar um ambiente ideal e que serviu de referência para as demais situações não-normais. Os programas SAS® (PROC IML) utilizados neste trabalho estão apresentados no apêndice.

4 RESULTADOS E DISCUSSÃO

Na seqüência desta seção serão apresentados os principais resultados para as taxas de erro tipo I e poder dos quatro testes estudados nesse artigo. São abordadas duas situações distintas: considerando amostragem das populações normal e não-normal (uniforme e gama). Também foram considerados dois níveis nominais de significância (5% e 1%).

4.1. Taxas de erro tipo I sob distribuição normal

Nas Tabelas 6, 7, 8 e 9 estão apresentadas as taxas de erro tipo I dos testes TBM, TGS, TBMB e TGSB sob normalidade e considerando $k = 2$ populações em função de ρ, p e n_i. O intervalo de 95% de confiança para o valor nominal de 5% da taxa de erro tipo I é [0,034; 0,071].

Observa-se que o TBM controlou a taxa de erro tipo I em quase todas as situações simuladas, exceto para as situações de $p = 3$, para os tamanhos de amostras $n_1 = 5$ e $n_2 = 5$ para $\rho = 0{,}1$; $n_1 = 5$ e $n_2 = 20$ para $\rho = 0$ e $0{,}9$; $n_1 = 5$ e $n_2 = 60$ para $\rho = 0$ e $0{,}5$; $n_1 = 7$ e $n_2 = 20$ para $\rho = 0{,}1$. Nestas situações, o TBM não controlou o erro tipo I, apresentando taxas superiores ao valor nominal estabelecido. Para $n_1 = n_1 = 60$ e $\rho = 0{,}1$, o TBM foi rigoroso quanto ao controle do erro tipo I.

O TGS foi um teste que tendeu a ser mais rigoroso à medida que os tamanhos amostrais diminuem. Se forem observadas as taxas de erro tipo I não significativamente diferentes da taxa nominal (Tabelas 6, 7, 8 e 9) deste teste, pode-se fazer uma associação, a princípio, com os maiores tamanhos amostrais. Da mesma forma que ocorreu para o teste TBM, considerando, por exemplo,

$n_1 = n_2 = 5$, a correlação teve pouca ou nenhuma influência na taxa de erro tipo I deste teste.

O teste de bootstrap TBMB, em geral, foi rigoroso para amostras pequenas (Tabela 6), inclusive apresentando, em algumas situações, taxas nulas de erro tipo I. Houve uma tendência de melhorar a performance com o aumento dos tamanhos amostrais. Observou-se que o aumento número de variáveis também afetou o teste de bootstrap TBMB no sentido de torná-lo mais rigoroso, ao contrário do efeito causado no teste TBM (Tabelas 7, 8 e 9).

Os resultados do erro tipo I para o TGSB foram similares aos resultados apresentados pelo TGS, exceto por uma sensível melhora do teste TGSB. Este teste deixou de ser rigoroso e passou a ter taxas de erro iguais ao valor nominal, quando pelo menos uma das duas amostras tinha maior tamanho. Neste caso, o maior número de variáveis afetou a taxa de erro tipo I. Por exemplo, para as situações de $n_1 = n_2 = 60$ para $\rho = 0,5$ e $p = 2$; $n_1 = n_2 = 30$ para $\rho = 0$ e $p = 3$, este teste foi liberal. A partir de $p \geq 5$, o TGSB apresentou taxas de erro tipo I iguais a nominal.

Zhang e Boos (1992) avaliaram o erro tipo I do teste de Bartlett e de sua versão bootstrap. Uma situação equivalente à simulada neste trabalho foi $k = 2$, $n_1 = n_2 = 20$ com $p = 5$. Neste caso, os autores encontraram $\hat{\alpha} = 0,056$ e $\hat{\alpha} = 0,012$, para os testes Bartlett assintótico e bootstrap, respectivamente, em uma população normal. No presente trabalho, os valores encontrados foram $\hat{\alpha} = 0,052$ e $\hat{\alpha} = 0,015$ para TBM e TBMB, respectivamente. A comparação destes resultados mostra uma quase perfeita concordância entre eles, ou seja, os resultados de Zhang e Boos (1992) foram reproduzidos neste trabalho.

TABELA 6 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis $(p = 2)$, da correlação (ρ) para o número de populações ($k = 2$).

n_i	ρ	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	0	0,045	0,027*	0,000*	0,025*
(5,5)	0,1	0,052	0,032*	0,000*	0,030*
(5,5)	0,5	0,052	0,020*	0,000*	0,018*
(5,5)	0,9	0,060	0,033*	0,002*	0,025*
(5,20)	0	0,070	0,032*	0,040	0,035
(5,20)	0,1	0,048	0,043	0,028*	0,053
(5,20)	0,5	0,042	0,050	0,018*	0,052
(5,20)	0,9	0,040	0,047	0,017*	0,050
(5,60)	0	0,067	0,040	0,052	0,050
(5,60)	0,1	0,038	0,045	0,027*	0,048
(5,60)	0,5	0,068	0,062	0,062	0,068
(5,60)	0,9	0,040	0,052	0,032*	0,062
(7,20)	0	0,047	0,055	0,025*	0,055
(7,20)	0,1	0,062	0,050	0,043	0,043
(7,20)	0,5	0,048	0,052	0,035	0,055
(7,20)	0,9	0,055	0,033*	0,037	0,050
(15,60)	0	0,052	0,052	0,048	0,063
(15,60)	0,1	0,048	0,058	0,050	0,058
(15,60)	0,5	0,053	0,063	0,055	0,067
(15,60)	0,9	0,053	0,038	0,058	0,045
(20,20)	0	0,037	0,037	0,030*	0,045
(20,20)	0,1	0,057	0,057	0,045	0,055
(20,20)	0,5	0,050	0,053	0,037	0,067
(20,20)	0,9	0,043	0,040	0,042	0,047
(30,30)	0	0,050	0,048	0,047	0,052
(30,30)	0,1	0,040	0,043	0,038	0,037
(30,30)	0,5	0,038	0,043	0,038	0,052
(30,30)	0,9	0,040	0,052	0,035	0,058
(60,60)	0	0,047	0,050	0,042	0,058
(60,60)	0,1	0,053	0,068	0,055	0,077*
(60,60)	0,5	0,060	0,043	0,053	0,052
(60,60)	0,9	0,053	0,037	0,052	0,042

* Significativo a 5% de probabilidade

TABELA 7 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis $(p = 3)$, da correlação (ρ) para o número de populações ($k = 2$).

n_i	ρ	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	0	0,057	0,025*	0,000*	0,023*
(5,5)	0,1	0,080*	0,010*	0,000*	0,013*
(5,5)	0,5	0,067	0,011*	0,000*	0,010*
(5,5)	0,9	0,048	0,010*	0,000*	0,020*
(5,20)	0	0,085*	0,032*	0,003*	0,035
(5,20)	0,1	0,068	0,037	0,002*	0,045
(5,20)	0,5	0,050	0,030*	0,002*	0,040
(5,20)	0,9	0,095*	0,035	0,002*	0,042
(5,60)	0	0,078*	0,042	0,027*	0,055
(5,60)	0,1	0,070	0,035	0,020*	0,047
(5,60)	0,5	0,087*	0,038	0,023*	0,043
(5,60)	0,9	0,067	0,062	0,022*	0,062
(7,20)	0	0,067	0,035	0,018*	0,047
(7,20)	0,1	0,072*	0,045	0,030*	0,048
(7,20)	0,5	0,055	0,030*	0,015*	0,043
(7,20)	0,9	0,040	0,032*	0,012*	0,060
(15,60)	0	0,047	0,042	0,040	0,058
(15,60)	0,1	0,055	0,053	0,043	0,058
(15,60)	0,5	0,047	0,048	0,038	0,065
(15,60)	0,9	0,045	0,047	0,028*	0,040
(20,20)	0	0,053	0,040	0,043	0,063
(20,20)	0,1	0,037	0,042	0,025*	0,057
(20,20)	0,5	0,037	0,038	0,020*	0,055
(20,20)	0,9	0,038	0,052	0,028*	0,063
(30,30)	0	0,050	0,057	0,035	0,075*
(30,30)	0,1	0,053	0,048	0,040	0,063
(30,30)	0,5	0,048	0,043	0,043	0,048
(30,30)	0,9	0,047	0,045	0,033*	0,057
(60,60)	0	0,035	0,060	0,035	0,067
(60,60)	0,1	0,033*	0,057	0,027*	0,062
(60,60)	0,5	0,045	0,045	0,040	0,057
(60,60)	0,9	0,057	0,045	0,052	0,048

* Significativo a 5% de probabilidade.

TABELA 8 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis $(p = 5)$, da correlação (ρ) para o número de populações ($k = 2$).

n_i	ρ	**TBM**	**TGS**	**TBMB**	**TGSB**
(10,10)	0	0,060	0,018*	0,000*	0,022*
(10,10)	0,1	0,063	0,018*	0,000*	0,025*
(10,10)	0,5	0,063	0,018*	0,000*	0,025*
(10,10)	0,9	0,062	0,012*	0,000*	0,020*
(15,60)	0	0,048	0,043	0,028*	0,058
(15,60)	0,1	0,050	0,042	0,025*	0,058
(15,60)	0,5	0,055	0,025*	0,027*	0,037
(15,60)	0,9	0,053	0,042	0,030*	0,055
(20,20)	0	0,055	0,032*	0,020*	0,045
(20,20)	0,1	0,062	0,025*	0,013*	0,047
(20,20)	0,5	0,058	0,040	0,017*	0,045
(20,20)	0,9	0,057	0,035	0,023*	0,052
(30,30)	0	0,042	0,043	0,020*	0,058
(30,30)	0,1	0,055	0,043	0,025*	0,057
(30,30)	0,5	0,048	0,040	0,030*	0,048
(30,30)	0,9	0,057	0,045	0,033*	0,060
(60,60)	0	0,050	0,048	0,045	0,055
(60,60)	0,1	0,040	0,043	0,035	0,050
(60,60)	0,5	0,048	0,045	0,027*	0,050
(60,60)	0,9	0,045	0,037	0,033*	0,052

* Significativo a 5% de probabilidade

TABELA 9 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis $(p=10)$, da correlação (ρ) para o número de populações ($k=2$).

n_i	ρ	**TBM**	**TGS**	**TBMB**	**TGSB**
(20,20)	0	0,065	0,022*	0,000*	0,017*
(20,20)	0,1	0,078*	0,017*	0,000*	0,013*
(20,20)	0,5	0,060	0,025*	0,000*	0,022*
(20,20)	0,9	0,082*	0,018*	0,000*	0,020*
(30,30)	0	0,072*	0,027*	0,003*	0,033*
(30,30)	0,1	0,052	0,032*	0,003*	0,035
(30,30)	0,5	0,043	0,035	0,005*	0,045
(30,30)	0,9	0,062	0,030*	0,008*	0,035
(60,60)	0	0,060	0,052	0,023*	0,052*
(60,60)	0,1	0,055	0,045	0,015*	0,055
(60,60)	0,5	0,058	0,050	0,033*	0,058
(60,60)	0,9	0,048	0,048	0,018*	0,058

* Significativo a 5% de probabilidade

Na Tabela 10 estão apresentadas as taxas de erro tipo I para os quatro testes estudados para $k=5$ populações. O teste TBM não manteve a mesma performance da situação de $k=2$, sendo, em geral, mais liberal. Essa maior liberalidade foi provocada por um maior número de variáveis e por amostras de tamanhos diferentes e intermediários. A pior situação ocorreu para $n_i=20(0)20$ com $p=10$ e $\rho=0,5$, em que a taxa de erro tipo I observada foi de 8,5%.

O TGS tendeu a ser mais rigoroso com amostras intermediárias e com o aumento do número de variáveis. Por exemplo, para $n_i=20$ com $\rho=0,5$ e $p=10$, a taxa de erro tipo I deste teste foi de 3,3%.

O teste de bootstrap TBMB tendeu a ser rigoroso para amostras menores (Tabela 10). A partir de $\min(n_i)$ igual a 30, as taxas de erro tipo I foram semelhantes ao valor nominal de 5%, exceto para $p = 10$. O teste TGSB apresentou também a mesma tendência do teste TBMB de ser rigoroso com amostras menores, só que em menor intensidade. Houve uma melhoria de performance, em geral, na versão bootstrap do teste de Samiuddin.

TABELA 10 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p), da correlação $(\rho = 0,5)$ e para o número de populações ($k = 5$).

n_i	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	2	0,058	0,040	0,012*	0,037
15(0)15	2	0,053	0,045	0,045	0,048
15(5)35	2	0,043	0,037	0,042	0,050
10(10)50	2	0,042	0,050	0,048	0,053
20(0)20	2	0,032*	0,030*	0,032*	0,028*
30(0)30	2	0,048	0,050	0,048	0,062
60(0)60	2	0,040	0,052	0,048	0,060
15(0)15	3	0,033*	0,047	0,018*	0,050
15(5)35	3	0,050	0,042	0,042	0,055
10(10)50	3	0,077*	0,052	0,062	0,065
20(0)20	3	0,073*	0,067	0,055	0,06
30(0)30	3	0,042	0,042	0,043	0,05
60(0)60	3	0,043	0,052	0,037	0,052
10(0)10	5	0,058	0,033*	0,003*	0,045
15(0)15	5	0,038	0,047	0,015*	0,052
15(5)35	5	0,060	0,048	0,042	0,063
10(10)50	5	0,060	0,045	0,038	0,053
20(0)20	5	0,048	0,037	0,027*	0,047
30(0)30	5	0,072*	0,045	0,053	0,053
60(0)60	5	0,070	0,060	0,070	0,073*
20(0)20	10	0,085*	0,033*	0,002*	0,042
30(0)30	10	0,075*	0,028*	0,013*	0,042
60(0)60	10	0,063	0,042	0,037	0,058

* Significativo a 5% de probabilidade.

Na Tabela 11 estão apresentadas as taxas de erro tipo I dos quatro testes para $k = 10$. De maneira geral, percebe-se que houve melhorias em todos os testes, na medida em que k aumentou de 2 para 10. Os dois testes que apresentaram melhor desempenho foram o TBM e TGSB. A única exceção ocorreu com $n_i = 5 \;\; \forall \;\; i = 1, 2, ..., 10$, $p = 2$ e $\rho = 0,5$ para o teste TGSB que se mostrou rigoroso. Estes resultados para $k = 10$ apresentaram as mesmas tendências já apontadas e discutidas para $k = 2$ e $k = 5$, conforme pode ser observado na Tabela 11.

TABELA 11 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) da correlação $(\rho = 0,5)$ e para o número de populações ($k = 10$).

n_i	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	2	0,043	0,015*	0,015*	0,028*
15(0)15	2	0,055	0,033*	0,030*	0,047
10(10)100	2	0,040	0,043	0,048	0,043
20(0)20	2	0,058	0,048	0,057	0,045
30(0)30	2	0,050	0,045	0,052	0,053
60(0)60	2	0,045	0,050	0,045	0,052
15(0)15	3	0,060	0,053	0,038	0,057
10(10)100	3	0,063	0,055	0,062	0,062
20(0)20	3	0,058	0,038	0,045	0,050
30(0)30	3	0,035	0,038	0,038	0,043
60(0)60	3	0,052	0,067	0,052	0,070
15(0)15	5	0,055	0,050	0,015*	0,065
10(10)100	5	0,045	0,042	0,037	0,047
20(0)20	5	0,058	0,035	0,040	0,050
30(0)30	5	0,045	0,053	0,028*	0,063
60(0)60	5	0,070	0,053	0,068	0,062
30(0)30	10	0,058	0,045	0,027*	0,052
60(0)60	10	0,057	0,055	0,045	0,070

* Significativo a 5% de probabilidade.

É conveniente salientar que não há efeito pronunciado da correlação nas taxas de erro tipo I dos diferentes testes. O teste TBMB mostrou-se rigoroso para $p = 10$, enquanto o TBM controlou o erro tipo I.

Na Tabela 12 estão apresentadas as taxas de erro tipo I para os quatro testes, considerando $\alpha = 1\%$, $k = 2$ e somente algumas situações representativas de n_i. O intervalo de 95% de confiança para o valor nominal de 1% da taxa de erro tipo I é [0,004; 0,022].

De maneira geral pode-se observar a mesma tendência nos resultados da situação de $\alpha = 5\%$. O TBM não apresentou resultados de taxas de erro significativamente diferentes da nominal; o TGS foi rigoroso apenas para amostras menores. Em situações de grandes amostras $(n_1 = n_2 = 60)$, se o número de variáveis for aumentado $(p = 10)$, o teste TBMB se tornou rigoroso. O teste TGSB apresentou os mesmos resultados para a taxa de erro tipo I do TGS.

TABELA 12 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 1%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) da correlação (ρ) e para o número de populações ($k = 2$).

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	0,5	2	0,010	0,003*	0,000*	0,003*
(5,20)	0,9	2	0,007	0,010	0,000*	0,005
(5,60)	0,5	2	0,017	0,010	0,013	0,010
(30,60)	0,9	5	0,012	0,012	0,005	0,010
(60,60)	0,9	10	0,010	0,008	0,003*	0,012

* Significativo a 5% de probabilidade.

Na Tabela 13, pelos resultados da taxa de erro tipo I para $\alpha = 1\%$ e $k = 5$ dos quatro testes, pode-se observar o mesmo tipo de performance descrito anteriormente. Destacam-se os efeitos do aumento do número de variáveis, principalmente quando p é expressivo em relação ao $\min(n_i)$. Com o aumento de p os testes TBM e TBMB se tornaram, em geral, rigorosos. O TGS e o TGSB apresentaram resultados praticamente iguais, exceto para $n_i = 60$, $p = 10$ e $\rho = 0,5$. Neste caso, o TGS foi rigoroso e o TGSB apresentou tamanho igual ao valor nominal $\alpha = 1\%$.

TABELA 13 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 1%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p), da correlação (ρ) e para o número de populações ($k = 5$).

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	0,5	2	0,015	0,003*	0,000*	0,003*
5(0)5	0,9	2	0,012	0,002*	0,000*	0,000*
20(0)20	0,9	2	0,012	0,010	0,010	0,010
10(0)10	0	5	0,020	0,007	0,000*	0,010
20(0)20	0	5	0,010	0,008	0,005	0,012
20(0)20	0,5	5	0,005	0,008	0,000*	0,012
30(0)30	0,1	5	0,010	0,012	0,005	0,012
60(0)60	0,9	5	0,012	0,010	0,008	0,020
15(10)55	0,1	10	0,025*	0,008	0,000*	0,010
20(10)60	0,9	10	0,027*	0,012	0,008	0,018
60(0)60	0,5	10	0,010	0,003*	0,002*	0,007

* Significativo a 5% de probabilidade.

Finalmente, na Tabela 14, estão apresentados alguns resultados para $k = 10$. Nenhuma discrepância do padrão de performance dos testes descritos anteriormente foi observada.

TABELA 14 Taxas de erro tipo I para as quatro opções dos testes, considerando normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) da correlação (ρ) e para o número de populações ($k = 10$).

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(5)50	0	2	0,013	0,010	0,005	0,010
10(0)10	0,9	2	0,007	0,007	0,000*	0,012
30(0)30	0,5	5	0,017	0,012	0,008	0,015
60(0)60	0,5	10	0,015	0,007	0,015	0,013

* Significativo a 5% de probabilidade.

4.2 Taxas de erro tipo I para distribuições não-normais (uniforme e gama)

Nesta seção buscou-se a avaliação da robustez dos quatro testes de comparação de covariâncias no controle do erro tipo I. Para isso utilizaram-se amostragens de populações uniformes e gamas multivariadas.

Na Tabela 15 estão apresentadas as taxas de erro tipo I dos quatro testes com $k = 2$, para a distribuição uniforme. Pode-se observar que, de maneira geral, ocorreu uma tendência dos testes apresentarem uma redução da taxa de erro tipo I em relação à amostragem da população normal. Os testes TBM, TGS, TBMB e TGSB tornaram-se mais rigorosos do que na situação normal, principalmente para amostras menores ou para situações com maior número de variáveis. O teste TGS sofreu um maior impacto da mudança de distribuição. Este teste apresentou taxas de erro significativamente $(p < 0,05)$ inferiores ao valor nominal $(\alpha = 5\%)$ em todos os casos. Espera-se que o poder deste teste em distribuições platicúrticas seja inferior ao poder esperado nas situações de normalidade.

O TGSB foi o teste de melhor performance, principalmente com grandes amostras. Notou-se, ainda, a influência do número maior de variáveis no sentido de torná-lo mais rigoroso. O teste TBMB foi rigoroso para pequenas amostras e tendeu a apresentar taxas iguais ao valor nominal, para grandes amostras. As exceções ocorreram com $p > 5$.

Nestes casos de maior número de variáveis, principalmente os testes assintóticos TBM e TGS tenderam a ser tornar rigorosos, com valores de taxas de erro significativamente $(p < 0,05)$ inferiores ao valor nominal. Nas versões bootstrap, as taxas empíricas foram quase sempre não diferentes da nominal.

TABELA 15 Taxas de erro tipo I para as quatro opções dos testes, sob não normalidade (uniforme), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p), da correlação (ρ) e para o número de populações ($k = 2$).

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	0,1	2	0,025*	0,005*	0,002*	0,012*
(5,5)	0,5	2	0,043	0,022*	0,000*	0,018*
(5,20)	0,1	2	0,040	0,007*	0,028*	0,035
(5,20)	0,5	2	0,038	0,013*	0,033*	0,045
(5,20)	0,9	2	0,027*	0,005*	0,025*	0,033*
(7,20)	0	2	0,028*	0,007*	0,037	0,035
(60,60)	0,9	2	0,035	0,003*	0,040	0,053
(60,60)	0,1	5	0,020*	0,000*	0,037*	0,042
(60,60)	0	10	0,022*	0,002*	0,028*	0,030*

* Significativo a 5% de probabilidade.

Na Tabela 16 estão apresentadas as taxas de erro tipo I dos testes TBM, TGS, TBMB e TGSB para simulações realizadas a partir da distribuição gama com parâmetro de forma $\lambda = 1,5$ e $k = 2$ populações. Nota-se que os testes assintóticos TBM e TGS passaram a ser liberais $(p < 0,05)$ para todos os tamanhos amostrais simulados. Quanto maiores os tamanhos amostrais, mais liberais foram os testes. O TGS, nas mesmas situações simuladas, apresentou menores taxas de erro tipo I do que o TBM. Para pequenas amostras, os resultados foram mais parecidos, com uma pequena vantagem a favor do TGS. Para algumas situações de $n_1 = n_2 = 5$, o TGS chegou a apresentar taxas de erro iguais ao valor nominal. Já em grandes amostras, a diferença tendeu a se dilatar, chegando a superar 20 pontos percentuais. Essas maiores diferenças ocorreram principalmente nas maiores dimensões $(p = 5 \text{ e } 10)$.

Isso mostra que esses testes não são robustos à violação de normalidade e que, aparentemente, são mais afetados por desvios de simetria. O TGS mostrou-se mais tolerante a esse desvio de normalidade do que o teste TBM, sendo este último mais sensível à medida que p aumenta. Estes resultados concordam com os resultados de Zhang e Boos (1992) e Zhu, Ng e Zing (2002). Houve um resultado de certa forma surpreendente, que é a tendência de aumentar a liberalidade dos testes com o aumento dos tamanhos amostrais, pois se espera exatamente o contrário.

As versões bootstrap dos testes (TBMB e TGSB) foram rigorosas em pequenas amostras ou tiveram tamanhos iguais ao nominal de $\alpha = 5\%$. Essa passagem de liberal para rigoroso quando se passa dos testes assintóticos para os de bootstrap é desejável, mesmo não sendo ideal. Isso porque há um controle da taxa de erro tipo I, mesmo que em níveis inferiores ao nominal. A comparação dos resultados dos testes TBMB e TGSB permite verificar que o teste TGSB é superior à TBMB, pois apresentou taxas de erro tipo I quase sempre iguais ao valor nominal de 5%. O teste TBMB com grandes amostras e altas dimensões tendeu, inclusive, a ser rigoroso.

TABELA 16 Taxas de erro tipo I para as quatro opções dos testes, sob não normalidade (gama), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) da correlação (ρ) para o número de populações $(k=2)$.

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	0	2	0,102*	0,080*	0,000*	0,035
(5,5)	0,1	2	0,110*	0,085*	0,000*	0,038
(5,5)	0,5	2	0,112*	0,063*	0,002*	0,028*
(5,5)	0,9	2	0,103*	0,067	0,002*	0,030*
(5,20)	0	2	0,147*	0,102*	0,022*	0,037
(5,20)	0,1	2	0,140	0,135*	0,020*	0,062
(5,20)	0,5	2	0,143*	0,143*	0,020*	0,050
(5,20)	0,9	2	0,148*	0,130*	0,043	0,050
(7,20)	0	2	0,207*	0,177*	0,037	0,058
(7,20)	0,1	2	0,177*	0,142*	0,025*	0,057
(7,20)	0,5	2	0,185*	0,138*	0,027*	0,062
(7,20)	0,9	2	0,173*	0,155*	0,035	0,065
(60,60)	0	2	0,282*	0,218*	0,040	0,060
(60,60)	0,1	2	0,313*	0,225*	0,062	0,062
(60,60)	0,5	2	0,263*	0,237*	0,033*	0,058
(60,60)	0,9	2	0,288*	0,228*	0,040	0,065
(60,60)	0,1	5	0,372*	0,190*	0,038	0,050
(60,60)	0,5	5	0,325*	0,212*	0,047	0,070
(60,60)	0,9	5	0,347*	0,200*	0,027*	0,068
(60,60)	0	10	0,410*	0,193*	0,028*	0,067
(60,60)	0,1	10	0,410*	0,203*	0,023*	0,068
(60,60)	0,5	10	0,402*	0,185*	0,018*	0,070
(60,60)	0,9	10	0,407*	0,205*	0,017*	0,038

*Significativo a 5% de probabilidade.

Na Tabela 17 estão apresentadas as taxas de erro tipo I dos testes para $\alpha = 5\%$ e $k = 5$, considerando a distribuição gama p-variada e diferentes tamanhos amostrais. O que se observa é um mesmo padrão de comportamento da situação em que $k = 2$. A diferença que ocorre é uma acentuação da liberalidade dos testes assintóticos TBM e TGS. Por exemplo, com

$n_i = 5 \ (i = 1, 2, ..., k)$, o valor da significância observada do teste TBM foi $\hat{\alpha} = 0,255$ e do TGS foi $\hat{\alpha} = 0,183$ para ρ=0,1 e $k = 5$ (Tabela 8). Para essa mesma situação, com $k = 2$ (Tabela 7) os valores foram $\hat{\alpha} = 0,110$ e $\hat{\alpha} = 0,085$, respectivamente. O teste TGSB apresentou dois resultados significativamente liberais com $n_i = 5$ e $k = 5$ quando se utilizou um maior número de variáveis $p = 5$. O teste TBMB sempre rigoroso ou de tamanho $\alpha = 5\%$.

TABELA 17 Taxas de erro tipo I para as quatro opções dos testes, sob não normalidade (gama), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) da correlação (ρ) para o número de populações $(k = 5)$.

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	0	2	0,237*	0,185*	0,010*	0,032*
5(0)5	0,1	2	0,255*	0,183*	0,010*	0,037
5(0)5	0,5	2	0,247*	0,205*	0,010*	0,045
5(0)5	0,9	2	0,217*	0,177*	0,007*	0,032*
20(0)20	0	2	0,535*	0,385*	0,032*	0,052
20(0)20	0,1	2	0,532*	0,410*	0,042	0,047
20(0)20	0,5	2	0,527*	0,430*	0,025*	0,043
20(0)20	0,9	2	0,587*	0,447*	0,043	0,053
20(0)20	0	5	0,672*	0,398*	0,028*	0,087*
20(0)20	0,1	5	0,708*	0,408*	0,038	0,075*
20(0)20	0,5	5	0,662*	0,405*	0,022*	0,063
20(0)20	0,9	5	0,663*	0,353*	0,030*	0,052

* Significativo a 5% de probabilidade.

Na Tabela 18 apresentam-se as taxas de erro tipo I para $k = 10$ populações e distribuição gama. O que se pode observar é o mesmo padrão de resposta observado para $k = 2$ (Tabela 16) e $k = 5$ (Tabela 17), exceto por um agravamento da liberalidade de TBM e TGS. De maneira geral, verifica-se que os testes assintóticos TBM e TGS não são robustos no controle do erro tipo I em distribuições não-normais assimétricas. Os testes de bootstrap TBMB e TGSB controlaram o erro tipo I, mantendo as taxas empíricas iguais ou inferiores ao valor nominal de 5%. Algumas exceções ocorreram para o teste TGSB, no entanto, as magnitudes das diferenças para o valor nominal não foram expressivas. Assim, pode-se considerar TBMB e TGSB testes robustos no controle do erro tipo I.

TABELA 18 Taxas de erro tipo I para as quatro opções dos testes, sob não normalidade (gama), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) da correlação (ρ) para o número de populações $(k = 10)$.

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	0	2	0,445*	0,417*	0,018*	0,018*
5(0)5	0,1	2	0,458*	0,383*	0,017*	0,018*
5(0)5	0,5	2	0,458*	0,405*	0,010*	0,043
5(0)5	0,9	2	0,460*	0,405*	0,008*	0,038

* Significativo a 5% de probabilidade.

4.3 Poder dos testes sob normalidade

Para o estudo do poder, optou-se por fixar a correlação em $\rho = 0,5$. Este procedimento foi adotado em virtude de os resultados obtidos para as taxas de erro tipo I não terem sido afetados pela estrutura de correlação. De qualquer forma, alguns casos foram simulados para avaliar o poder, considerando diferentes correlações (resultados não apresentados) e foi constatado o mesmo padrão de resposta do erro tipo I. Assim, os resultados de poder dos quatro testes para $\delta = 2$ e 4, considerando diferentes tamanhos amostrais, dimensões, $\alpha = 5\%$, $k = 2$ e amostragem de populações normais estão apresentados na Tabela 19.

TABELA 19 Poder para as quatro opções dos testes, sob normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) e da correlação $(\rho = 0,5)$, para o número de populações ($k = 2$) e diferenças entre as matrizes de covariância (δ).

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	2	2	0,053	0,035	0,000	0,023
(5,20)	2	2	0,048	0,053	0,028	0,052
(7,20)	2	2	0,060	0,107	0,050	0,098
(5,60)	2	2	0,070	0,060	0,053	0,063
(20,20)	2	2	0,138	0,160	0,110	0,163
(30,30)	2	2	0,187	0,267	0,143	0,253
(15,40)	2	2	0,087	0,155	0,070	0,160
(40,15)	2	2	0,120	0,177	0,093	0,187
(15,60)	2	2	0,107	0,147	0,097	0,157
(60,15)	2	2	0,143	0,208	0,127	0,228
(60,60)	2	2	0,303	0,467	0,295	0,462
(5,5)	4	2	0,078	0,107	0,000	0,093
(5,20)	4	2	0,092	0,133	0,030	0,138
(7,20)	4	2	0,143	0,230	0,078	0,220
(5,60)	4	2	0,098	0,155	0,080	0,165
(20,20)	4	2	0,362	0,532	0,285	0,548
(30,30)	4	2	0,543	0,707	0,488	0,708
(15,60)	4	2	0,392	0,577	0,347	0,575
(40,15)	4	2	0,427	0,592	0,387	0,608
(15,40)	4	2	0,327	0,515	0,290	0,523
(60,15)	4	2	0,487	0,655	0,438	0,647
(60,60)	4	2	0,888	0,963	0,870	0,957

Para $\delta = 2$, situação de pequena diferença entre as matrizes de covariância, o poder dos testes foi pequeno, sendo igual ou inferior ao valor nominal, principalmente com pequenas amostras. À medida que os tamanhos amostrais aumentam, o poder dos testes também aumenta. Estes resultados são esperados, haja vista que os testes são, em geral, rigorosos para pequenas amostras. Com amostras intermediárias $(n_1 = n_2 = 30)$, o poder máximo observado foi de 26,7% para o TGS. Para amostras maiores $(n_1 = n_2 = 60)$, ainda com $\delta = 2$, o maior poder observado foi 46,7% do TGS.

É conveniente notar que Zhu, Ng e Jing (2002) e Zhang e Boos (1992) utilizaram uma estrutura de covariâncias para $k = 2$ populações que corresponde a um valor da razão das variâncias generalizadas das matrizes de covariâncias $\delta = 8$. Encontraram poder igual a 73,5% para o teste de Bartlett assintótico e 65,7% para o teste bootstrap, sob normalidade com $n_1 = n_2 = 20$ e $p = 2$. Nesta mesma situação $(n_1 = n_2 = 20 \text{ e } p = 2)$, só que $\delta = 2$, foram encontrados os poderes 13,8% e 11,0% para o teste TBM e TBMB, respectivamente. Percebe-se que o poder de TBMB foi inferior ao de TBM, da mesma forma que foi apontado por Zhang e Boos (1992). Isso ocorreu no presente trabalho para diversos outros tamanhos amostrais. Esse resultado, de certa forma, já era esperado, uma vez que, sob normalidade, TBMB foi, em geral, mais rigoroso que TBM. Os poderes de TBM e TBMB para $\delta = 2$ foram 30,3% e 29,5% com $(n_1 = n_2 = 60)$, respectivamente.

Foi simulada a mesma situação de Zhang e Boos (1992) com $\delta = 8$ e mesma estrutura de covariância das duas populações Os poderes dos testes TBM e TBMB foram 0,705 e 0,587, respectivamente. Estes valores foram levemente superiores aos respectivos valores obtidos por Zhang e Boos (1992). Para os novos testes deste trabalho, TGS e TGSB, os resultados foram 0,857 e 0,833, respectivamente. Estes novos testes propostos são, conforme já destacados anteriormente, superiores ao teste original de Bartlett ou de sua versão de bootstrap. No entanto, para uma outra situação simulada por Zhang e Boos (1992), com $\delta = 0,75$, os novos testes tiveram baixa performance quando comparados com o TBM e TBMB. Para a população 1, Zhang e Boos (1992) consideraram a seguinte matriz de covariância:

$$\sum_1 = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}$$

e para a população 2, $\sum_2 = \mathrm{I}$. Na primeira situação $(\delta = 8)$, a população 1 tinha a seguinte estrutura:

$$\sum_1 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

e a população 2, $\sum_2 = \mathrm{I}$.

Os resultados de Zhang e Boos (1992) para a segunda situação dos testes TBM e TBMB foram 0,244 e 0,243, respectivamente. Os resultados deste trabalho, nesta situação, para os testes TBM, TGS, TBMB e TGSB foram 0,233, 0,032, 0,207 e 0,037, respectivamente. Estes resultados concordam com os de Zhang e Boos (1992), para os testes TBM e TBMB. Conforme já salientado, TGS e TGSB, surpreendentemente, tiveram uma péssima performance. Foi atribuído ao pequeno valor de $\delta = 0,75$ a pequena performance.

A princípio, poderia se pensar que a causa da baixa performance de TGS e TGSB fosse conseqüência da estrutura de covariância, uma vez que, no primeiro caso, a estrutura de covariância era a mesma nas duas populações e, no segundo, as estruturas de covariâncias eram diferentes $\left(\rho_{12}^{(1)} = 0,5 \text{ e } \rho_{12}^{(2)} = 0\right)$ para as populações 1 e 2, respectivamente. Para descaracterizar esse fato e comprovar que o que ocorreu se deveu basicamente ao baixo e alto valor de δ, simulou-se outra situação com estruturas de covariâncias diferentes para as duas populações, mas com δ grande $(\delta = 8)$. A matriz de covariância da população 1 foi:

$$\Sigma_1 = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

e da população 2, $\Sigma_2 = \mathrm{I}$.

Os resultados obtidos para os testes TBM, TGS, TBMB e TGSB foram 0,877, 0,800, 0,820 e 0,775, respectivamente.

Os testes TGS e TGSB apresentaram resultados semelhantes entre si para o poder, com $\delta = 2$, em todos os tamanhos amostrais. Além disso, os resultados destes testes foram superiores aos de TBM e TBMB, exceto para pequenas amostras. Para os maiores tamanhos amostrais simulados $(n_1 = n_2 = 60)$, os poderes para TGS e TGSB foram 46,7% e 46,2%, respectivamente. Estes resultados são aproximadamente 16 pontos percentuais superiores aos dos testes TBM e TBMB na mesma configuração.

O número de variáveis afeta os testes no sentido de reduzir os seus poderes. Pode-se observar que, com $\delta = 2$ e $n_1 = n_2 = 60$, o aumento de p de 2 para 5 provocou uma redução muito grande do poder dos testes. Para TBM, a redução foi de 21,8% para TGS foi de 25,5% para TBMB foi 22,3% e para TGSB, foi 24%. Assim, recomenda-se que os tamanhos amostrais sejam cada vez maiores, na medida em que se pretenda avaliar um maior número de variáveis, para se ter o mesmo poder de uma situação de menor dimensão. A performance relativa aos testes, no entanto, não é alterada, independentemente de p.

TABELA 20 Poder para as quatro opções dos testes, sob normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) e da correlação $(\rho = 0,5)$, para o número de populações ($k = 2$) e diferenças entre as matrizes de covariância (δ).

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	8	2	0,085	0,148	0,000	0,140
(5,20)	8	2	0,163	0,268	0,080	0,273
(7,20)	8	2	0,265	0,465	0,138	0,440
(5,60)	8	2	0,220	0,353	0,180	0,347
(20,20)	8	2	0,705	0,857	0,587	0,833
(30,30)	8	2	0,927	0,967	0,882	0,963
(15,40)	8	2	0,733	0,867	0,655	0,863
(40,15)	8	2	0,808	0,912	0,740	0,905
(15,60)	8	2	0,768	0,902	0,717	0,898
(60,15)	8	2	0,823	0,927	0,793	0,925
(60,60)	8	2	1,000	1,000	1,000	1,000
(5,5)	16	2	0,165	0,297	0,000	0,257
(5,20)	16	2	0,310	0,505	0,153	0,467
(7,20)	16	2	0,518	0,738	0,328	0,698
(5,60)	16	2	0,347	0,590	0,290	0,572
(20,20)	16	2	0,967	0,977	0,890	0,970
(30,30)	16	2	0,992	0,997	0,992	0,995
(15,40)	16	2	0,952	0,988	0,910	0,985
(40,15)	16	2	0,952	0,982	0,933	0,980
(15,60)	16	2	0,975	0,997	0,943	0,995
(60,15)	16	2	0,980	0,988	0,965	0,990
(60,60)	16	2	1,000	1,000	1,000	1,000

Um outro aspecto importante é a associação dos tamanhos amostrais com as populações, quando as amostras possuem tamanhos amostrais diferentes. Pode-se associar sempre a menor amostra com a população de menor variância generalizada, a segunda menor amostra com a população de segunda menor variância generalizada e assim por diante. Pode-se fazer a associação também ao contrário, para verificar-se o efeito que isso provoca no poder. Para o erro tipo I, nenhuma menção a esse fato foi feita, porque as covariâncias populacionais por construção são iguais. Avaliar este efeito é importante, principalmente quando o pesquisador pretende realizar experimentos não balanceados. Nesta situação prática, é impossível para este pesquisador fazer qualquer tipo de associação desta natureza, pois não se conhece as matrizes $\sum_i$. Porém, estudar este efeito é importante, principalmente para checar se há algum tipo de interação testes x associações dos tamanhos amostrais.

Foram avaliadas, para $\delta = 2$, as inversões $n_1 = 5$ e $n_2 = 60$, com $n_1 = 60$ e $n_2 = 5$; $n_1 = 15$ e $n_2 = 60$, com $n_1 = 60$ e $n_2 = 15$ e $n_1 = 15$ e $n_2 = 40$, com $n_1 = 40$ e $n_2 = 15$. O padrão de resposta foi basicamente o mesmo, ou seja, a performance relativa dos testes manteve-se a mesma, independentemente da associação dos tamanhos amostrais com as magnitudes das variâncias generalizadas das populações. Os testes TGS e TGSB apresentaram poderes semelhantes e foram superiores aos testes TBM e TBMB. Entre estes dois últimos, TBM foi levemente superior a TBMB.

Para $\delta = 4$, com $k = 2$, $\alpha = 5\%$ e $p = 2$ sob normalidade multivariada (Tabela 19) o mesmo padrão de resposta foi observado para o comportamento do poder dos quatro testes. O que se pode detectar é o aumento, já esperado, do poder em relação à situação de $\delta = 2$. Por exemplo, com $n_1 = n_1 = 60$ e $p = 2$, os poderes de TBM, TGS, TBMB e TGSB foram iguais a 88,84%, 96,3%, 87,0% e 95,7%, respectivamente, para $\delta = 4$. Nesta mesma situação, porém com $\delta = 2$, os poderes foram 30,3%, 46,7%, 29,5% e 46,2%, respectivamente. Nas associações das amostras menores com as populações de menores variâncias generalizadas, houve um menor poder do que o poder das situações em que as maiores amostras estavam associadas às populações de menores variâncias generalizadas.

Nas Tabelas 21 e 22 estão apresentados os poderes dos testes para $k=5$ e $k=10$, respectivamente, considerando populações normais, com $p=2$ e $\delta=2$ ou 16.

TABELA 21 Poder para as quatro opções dos testes, sob normalidade, valor nominal de 5% de significância, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) para o número de populações ($k=5$) e diferença das matrizes de covariâncias $(\delta = 2 \text{ e } 16)$.

n_i	ρ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	2	2	0,048	0,050	0,005	0,050
5(5)25	2	2	0,057	0,058	0,058	0,065
25(0)5	2	2	0,055	0,098	0,027	0,092
20(0)20	2	2	0,092	0,122	0,078	0,130
15(5)35	2	2	0,097	0,130	0,082	0,123
35(5)15	2	2	0,117	0,162	0,080	0,143
60(0)60	2	2	0,193	0,307	0,173	0,322
5(0)5	16	2	0,158	0,238	0,015	0,167
5(5)25	16	2	0,268	0,415	0,182	0,402
25(5)5	16	2	0,755	0,928	0,770	0,972
20(0)20	16	2	0,885	0,978	0,770	0,972
15(5)35	16	2	0,805	0,963	0,750	0,950
35(5)15	16	2	0,972	0,997	0,940	0,997
60(0)60	16	2	1,000	1,000	1,000	1,000

O padrão de resposta da performance relativa dos testes é exatamente o mesmo das situações de $k=2$. O que se percebe é uma influência do aumento do número de populações amostradas, no sentido de reduzir o poder dos testes. Isso só pode ser verificado fixando-se as amostras das k populações em um valor constante $(n_1 = m)$ e observando-se os poderes nas Tabelas 19 ou 20 e confrontando-os com aqueles correspondentes, para um mesmo δ, nas Tabelas 21 e 22. A magnitude dessa redução, em geral, é pequena e menor do que o efeito provocado pelo aumento do número de variáveis. Também é conveniente salientar que esse efeito ocorre de uma forma mais evidente para amostras de tamanhos intermediários $(5 < n_i < 105)$.

TABELA 22 Poder para as quatro opções dos testes, sob normalidade, valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) para o número de populações ($k = 10$) e diferença das matrizes de covariâncias $(\delta = 2 \text{ e } 16)$.

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	2	2	0,060	0,062	0,025	0,078
5(5)50	2	2	0,077	0,083	0,068	0,083
50(5)5	2	2	0,110	0,175	0,095	0,178
20(0)20	2	2	0,087	0,118	0,062	0,110
15(5)60	2	2	0,115	0,185	0,098	0,180
60(5)15	2	2	0,123	0,247	0,112	0,237
60(0)60	2	2	0,228	0,382	0,222	0,377
5(0)5	16	2	0,152	0,250	0,045	0,177
5(5)50	16	2	0,502	0,763	0,413	0,713
50(5)5	16	2	0,993	1,000	0,983	1,000
20(0)20	16	2	0,888	0,983	0,795	0,973
15(5)60	16	2	0,948	0,998	0,925	0,998
60(5)15	16	2	1,000	1,000	1,000	1,000
60(0)60	16	2	1,000	1,000	1,000	1,000

Observa-se uma maior diferenciação dos testes quando δ é maior $(\delta = 16)$ com amostras menores. Neste caso, os testes TGS e TGSB apresentaram maiores poderes do que TBM e TBMB e, principalmente, a magnitude das diferenças é bem maior do que ocorre com grandes amostras e elevadas razões das variâncias generalizadas $(\delta = 16)$. Se, por outro lado, os δ's forem pequenos $(\delta = 2)$ e as amostras grandes, o TGS e TGSB superaram em poder os testes TBM e TBMB e com uma expressiva diferença.

4.4 Poder dos testes sob não-normalidade

O poder dos quatro testes foi avaliado sob condições atípicas dos pressupostos para a sua aplicação. Duas distribuições foram consideradas, as mesmas da avaliação do erro tipo I (uniforme e gama) nos casos mais relevantes. Na Tabela 23 estão apresentados os poderes dos testes para a uniforme com $k = 2$ e $p = 2$, para os diferentes poderes de δ e n_i's.

O que pode ser observado é, basicamente, uma superioridade dos testes de bootstrap em relação aos seus pares assintóticos em amostras maiores. Conforme já antevisto, o teste TGS apresentou poder inferior na distribuição uniforme com $\delta = 2$ quando comparado com a mesma situação sob normalidade (Tabela 6). A causa disso foi uma maior rigorosidade. Isso também ocorreu, de maneira geral, com os testes TBM e TGSB, exceto para o TBMB com amostras grandes $(n_1 = n_2 = 60)$. Neste caso, o TBMB apresentou poder de 40% sob distribuição uniforme $(\delta = 2)$ e 29,5% sob distribuição normal.

O TGSB nessa distribuição, principalmente para menores valores de δ, em que os testes se diferenciam mais, foi muito superior aos seus concorrentes, especialmente em amostras grandes. A comparação com os poderes deste teste sob a distribuição uniforme em relação ao mesmo teste sob distribuição normal mostrou uma superioridade da primeira situação (uniforme), haja vista que houve um controle da taxa de erro tipo I em amostragens de ambas as distribuições confirmando a robustez deste teste.

Para $\delta = 8$, o TGS tendeu a superar o TBMB, o que não foi evidente para $\delta = 2$. O TBM em pequenas amostras foi superior ao TBMB e, em amostras intermediárias, inferior. Houve uma menor diferenciação dos testes nesta situação de maior δ. O TGSB novamente superou os demais testes em poder em todas as situações de tamanhos amostrais com $\delta = 8$.

TABELA 23 Poder para as quatro opções dos testes, sob não normalidade (uniforme), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) e da correlação $(\rho = 0,5)$, para o número de populações ($k = 2$) e diferenças entre as matrizes de covariâncias $(\delta = 2 \text{ e } 8)$.

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	2	2	0,043	0,025	0,002	0,027
(5,25)	2	2	0,030	0,087	0,028	0,065
(25,5)	2	2	0,035	0,047	0,032	0,115
(20,20)	2	2	0,038	0,083	0,090	0,248
(15,35)	2	2	0,037	0,077	0,095	0,252
(35,15)	2	2	0,032	0,082	0,082	0,283
(60,60)	2	2	0,152	0,428	0,400	0,783
(5,5)	8	2	0,077	0,137	0,000	0,138
(5,25)	8	2	0,133	0,230	0,102	0,452
(25,5)	8	2	0,215	0,542	0,100	0,553
(20,20)	8	2	0,782	0,960	0,840	0,982
(15,35)	8	2	0,708	0,940	0,835	0,987
(35,15)	8	2	0,818	0,963	0,825	0,978
(60,60)	8	2	1,000	1,000	1,000	1,000

Na Tabela 24 estão apresentados os poderes dos testes sob distribuição uniforme com $k = 10$, δ=2 e 8, $p = 2$ e diferentes tamanhos amostrais.

TABELA 24 Poder para as quatro opções dos testes, sob não normalidade (uniforme), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) e da correlação $(\rho = 0,5)$, para o número de populações ($k = 10$) e diferenças entre as matrizes de covariâncias $(\delta = 2 \text{ e } 8)$.

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	2	2	0,012	0,000	0,018	0,050
5(5)50	2	2	0,002	0,002	0,087	0,168
50(5)5	2	2	0,007	0,002	0,092	0,308
20(0)20	2	2	0,003	0,005	0,100	0,220
15(5)60	2	2	0,002	0,002	0,125	0,325
60(5)15	2	2	0,002	0,008	0,128	0,523
60(0)60	2	2	0,015	0,068	0,308	0,793
5(0)5	8	2	0,018	0,018	0,023	0,133
5(5)50	8	2	0,083	0,255	0,463	0,900
50(5)5	8	2	0,543	0,983	0,927	1,000
20(0)20	8	2	0,187	0,765	0,665	0,993
15(5)60	8	2	0,352	0,933	0,922	1,000
60(5)15	8	2	0,918	1,000	0,998	1,000
60(0)60	8	2	1,000	1,000	1,000	1,000

Houve uma sensível redução do poder dos testes TBM e TGS. O teste TBMB sofreu uma redução, porém, de menor magnitude, quando se passou da situação de $k = 2$ para $k = 10$. Basicamente, o TGSB não sofreu redução com o aumento de k de 2 para 10 populações com distribuição uniforme. Sob normalidade, houve uma redução do poder (Tabelas 19 e 20) deste teste.

A comparação entre o poder sob normalidade e sob a distribuição uniforme confirma uma grande superioridade do teste TGSB em diferenciar as matrizes de covariâncias populacionais para a distribuição uniforme. Por exemplo, $\delta = 2$, k=10 e $n_i = 60$, sob normalidade, o poder foi de 37,7% e, sob distribuição uniforme, de 79,3%. Mais uma vez, destaca-se a robustez do teste TGSB.

Nas Tabelas 25 e 26 estão apresentados os poderes para a “distribuição gama” para $k = 2$ e $k = 10$, respectivamente. Para o caso de $k = 2$ (Tabela 25), verificou-se, como já esperado, um maior poder dos testes assintóticos em relação aos de bootstrap. Esse maior poder não é real, pois o tamanho do teste é superior ao valor nominal α (Tabela 24). Assim, serão focadas somente as comparações daqueles que tiveram tamanho máximo igual ao valor nominal α.

Para $\delta = 2$, o poder dos testes bootstrap possuem baixa magnitude se as amostras são pequenas. Para amostras maiores $(n_1 = n_2 = 60)$, os poderes foram 12,2% e 24,2% na gama e 29,5% e 46,2% na normal, para os testes TBMB e TGSB, respectivamente. Houve redução do poder, ao contrário do que ocorreu para a distribuição uniforme. Percebe-se que o teste TGSB superou o TBMB em poder, em todas as situações.

Quando a diferença entre as covariâncias é maior, o padrão de resposta para $\delta = 2$ é mantido. Ocorre uma elevação esperada do poder e, com grandes amostras, o poder do teste TGSB aproxima-se ao dos testes assintóticos. Como as taxas de erro tipo I dos testes assintóticos tendem a aumentar com o aumento do tamanho da amostra, a recomendação de um teste assintótico, que teoricamente seria para amostras maiores, não pode ser feita em distribuições assimétricas. Nessa situação de grandes amostras, o teste bootstrap TGSB apresenta uma melhor performance quanto ao poder e mantém o erro tipo I sob controle. Como seu poder foi maior do que o TBMB para $\delta = 8$ e em tamanhos amostrais considerados, este teste foi considerado melhor.

TABELA 25 Poder para as quatro opções dos testes, sob não normalidade (gama), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) e da correlação $(\rho = 0,5)$, para o número de populações ($k = 2$) e diferenças entre as matrizes de covariâncias $(\delta = 2 \text{ e } 8)$.

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
(5,5)	2	2	0,130	0,087	0,000	0,035
(5,25)	2	2	0,152	0,160	0,032	0,068
(25,5)	2	2	0,170	0,183	0,038	0,083
(20,20)	2	2	0,308	0,257	0,055	0,127
(15,35)	2	2	0,333	0,295	0,052	0,110
(35,15)	2	2	0,293	0,265	0,067	0,132
(60,60)	2	2	0,493	0,460	0,122	0,242
(5,5)	8	2	0,165	0,187	0,000	0,087
(5,25)	8	2	0,318	0,398	0,068	0,182
(25,5)	8	2	0,372	0,428	0,130	0,242
(20,20)	8	2	0,695	0,728	0,303	0,513
(15,35)	8	2	0,722	0,745	0,323	0,520
(35,15)	8	2	0,725	0,777	0,398	0,573
(60,60)	8	2	0,977	0,987	0,872	0,940

Na Tabela 26 apresentam-se os poderes dos testes para $k = 10$. O que se observa é que as diferenças entre os testes assintóticos e bootstrap se dilataram, principalmente para $\delta = 2$. Entretanto, o padrão geral das respostas comparativas dos testes é o mesmo da situação de $k = 2$ populações.

TABELA 26 Poder para as quatro opções dos testes, sob não normalidade (gama), valor nominal de 5%, em função de diferentes combinações dos tamanhos amostrais n_1 e n_2 do número de variáveis (p) e da correlação $(\rho = 0,5)$, para o número de populações ($k = 10$) e diferença entre as matrizes de covariâncias $(\delta = 2 \text{ e } 8)$.

n_i	δ	p	**TBM**	**TGS**	**TBMB**	**TGSB**
5(0)5	2	2	0,483	0,428	0,017	0,028
5(5)50	2	2	0,798	0,685	0,043	0,040
50(5)5	2	2	0,818	0,703	0,062	0,075
20(0)20	2	2	0,850	0,737	0,030	0,070
15(5)60	2	2	0,865	0,775	0,045	0,065
60(5)15	2	2	0,868	0,803	0,070	0,115
5(0)5	2	2	0,522	0,513	0,022	0,067
5(5)50	8	2	0,930	0,878	0,137	0,193
50(5)5	8	2	0,973	0,975	0,318	0,585
20(0)20	8	2	0,947	0,955	0,172	0,388
15(5)60	8	2	0,987	0,973	0,150	0,445
60(5)15	8	2	0,995	0,997	0,522	0,778

4.5 Considerações finais

Uma grande crítica ao teste de Bartlett para realizar comparações de matrizes de covariâncias populacionais é a sua capacidade de diferenciar variâncias realmente heterogêneas ou violação da normalidade. A falta de robustez deste teste foi verificada neste trabalho, principalmente para a gama. Um novo teste foi proposto (TGS), mas pelo fato de ter sido fundamentado teoricamente na multinormalidade das amostras, sofreu em distribuições não-normais. Para distribuições simétricas platicúrticas, foi rigoroso e perdeu poder em relação à situação ideal de multinormalidade. Em distribuições assimétricas, o tamanho do teste foi excessivamente superior ao valor nominal, da mesma forma que ocorreu com o teste TBM.

Os testes bootstrap, conceitualmente, podem ser aplicados a amostras de qualquer distribuição, possuindo, portanto, uma ampla plasticidade. Realmente, esse resultado foi observado para ambos os testes de bootstrap estudados. Nas distribuições assimétricas, eles mantiveram o erro tipo I sob controle, sendo, em algumas situações, de pequenas amostras, até mesmo rigorosos. A comparação do poder dos testes TGSB e TBMB nas diferentes situações estudadas indica uma grande superioridade do teste TGSB em relação ao TBMB. O teste foi robusto, quase sempre apresentou tamanho do teste igual ao valor nominal e, entre os testes que controlaram o erro tipo I, foi quase sempre mais rigoroso.

Ao serem repetidos os estudos de Zhang e Boos (1992), verificou-se que, em uma das situações, sob normalidade, o teste TGSB não manteve sua boa performance relativa. Uma das possíveis causas aventadas foi a influência da razão de variâncias generalizadas. Nesta situação de baixa performance, constatou-se que Zhang e Boos (1992) utilizaram um $\delta = 1,33$ como razão das variâncias generalizadas das $k = 2$ populações. Na outra situação simulada por Zhang e Boos (1992), verificou-se que o teste TGSB foi superior aos concorrentes, conforme demais resultados deste trabalho. A explicação foi atribuída à razão de variâncias generalizadas que, neste caso, foi $\delta = 8$. Esta situação foi simulada para as duas populações, com $\delta = 8$, obtendo resultados semelhantes à situação de $\delta = 8$ com mesma estrutura de covariância, o que descarta ser essa a causa.

Os motivos para o teste de Bartlett e de sua versão de bootstrap terem apresentados maiores poderes para uma situação em que $\delta = 1,33$ não foram identificados, uma vez que se esperava que as variâncias generalizadas tivessem influência marcante no teste. Essa expectativa fundamenta-se na constatação de que a estatística do teste é baseada numa razão de covariâncias generalizadas. Situações muito mais amplas foram estudadas neste trabalho e os resultados observados apontam que o TGSB é um teste robusto, controlando o erro tipo I nos níveis preestabelecidos e apresentando maior poder do que os seus competidores diretos (TBMB). Sob distribuições uniformes, o teste TGSB apresentou poder superior ao mesmo teste aplicado sob normalidade.

A estrutura de covariância das populações não foi um fator relevante, mas a dimensão deve ser considerada. Se o pesquisador almeja estudar um fenômeno utilizando uma grande dimensão, é necessário, para que se tenha um maior poder, utilizar amostras maiores, principalmente se houver suspeitas de assimetria das populações amostradas. Por todas essas razões, recomenda-se a utilização dos testes de bootstrap e, em especial, o teste TGSB.

5 CONCLUSÃO

Por meio deste estudo de proposição de testes de homogeneidade de covariâncias mediante o uso de simulação Monte Carlo, pôde-se concluir que:

1. o teste generalizado de Samiuddin (TGS) e a sua versão bootstrap (TGSB) que foram propostos, tiveram êxito;
2. o teste generalizado de Samiuddin (TGS) sob normalidade foi considerado superior aos testes de Bartlett e de sua versão bootstrap;
3. os testes de bootstrap foram considerados superiores aos assintóticos e robustos, controlando a taxa de erro tipo I;
4. o teste generalizado de Samiuddin de bootstrap (TGSB) foi mais poderoso do que o seu concorrente robusto (TBMB) e é recomendada a sua utilização para testar a hipótese de homogeneidade de covariâncias.

REFERÊNCIAS BIBLIOGRÁFICAS

BARTLETT, M. S. A note on the multiplying factors for various χ^2 approximations. **Journal of Royal Statistical Society B,** London, v. 16, n. 2, p. 296-298, 1954.

BARTLETT, M. S. Properties of sufficiency and Statistical tests. **Proceedings of the Royal Statistical Society -** Serie A, London, v. 60, p. 268-282, 1937.

BERAN, R.; SRIVASTAVA, M. S. Bootstrap tests and confidence regions for functions os a covariance matrix. **Annals of Statistics,** Hayward, v. 13, n. 1, p. 95-115, 1985.

BOCK, R. D. **Multivariate statistical methods in behavior research.** New York: McGrawn Hill, 1975. 623 p.

BOOS, D. D.; BROWNIE, C. Bootstrap methods for testing homogeneity of variances. **Technometrics,** Alexandria, v. 31, n. 1, p. 69-82, Feb. 1989.

BORGES, L. C.; FERREIRA, D. F. Comparações de duas aproximações do teste t com variâncias heterogêneas através de simulação. **Ciência e Agrotecnologia,** Lavras, v.23, n. 2, p. 390-403, abr. 1999.

BORGES, L. C.; FERREIRA, D. F. **Poder e taxas de erro tipo I dos testes Scott-Knott, Tukey e Student-Newman-Keuls sob distribuições normal e não-normais dos resíduos.** 2002. p. 94. Dissertação (Mestrado em Estatística e Experimentação Agorpecuária) – Universidade Federal de Lavras, Lavras, MG.

BOX, G. E. P. A general distribution theory for a class of likelihood criteria. **Biometrika,** London, v. 36, n. 3/4, p. 317-346, 1949.

COCHRAN, W. G.; COX, G. M. **Experimental designs.** 2. ed. Singapure: John Wiley & Sons, 1957. 611 p.

CONOVER, W. J.; JOHNSON, M. E.; JOHNSON, M. M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. **Technometrics,** Alexandria, v. 23, n. 4, p. 351-361, Nov. 1981.

DACHS, J. N. **Estatística computacional:** uma introdução em turbo pascal. Rio de Janeiro: Livros Técnicos e Científicos, 1988. 236 p.

DE FINETTI, B. La prévision:ses lois logiques, ses sources subjectives. **Annales de I´Institut Henri Poincaré,** Paris, v. 7, p. 1-68, 1937. (Translated into English by Henry E. Kyburg Jr., Foresight: Its Logical Laws, its Subjective Sources. In: KYBURG JR., H. E.; SMOKLER, H. E. (Ed). **Studies in Subjective Probability.** New York: Wiley, 1964. p. 53-118).

DIXON, W. J.; MASSEY, F. J. **Introduction to statistical analysis.** 3. ed. New York: McGraw-Hill Book, 1969. p. 308-309.

DUNNETT, C. W. Recent results in multiple testing: Several treatments vc. a specified treatment. Proceedings of the In: INTERNATIONAL CONFERENCE ON LINEAR STATISTICAL INFERENCE LINSTAT 93, Ddordrecht. **Proceedings...** Dordrecht, the Netherlands: Kluwer Acad. Publ., 1994. p. 317-346.

FLURY, B. **Common principle components and related multivariate models.** New York: John Wiley, 1988.

FLURY, B.; RIEDWYL, H. **Multivariate statistics, a practical approach.** London: Chapman & Hall, 1988.

JOHNSON, R. A.;WICHERN, D. W. **Applied multivariate statistical analysis.** 4. ed. New York: Prentice Hall, 1998. 816 p.

GARTSIDE, P. S. A study of methods for comparing several variances. **Journal of the American Statistical Associations,** Alexandria, v. 67, n. 338, June 1972.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analisys.** London: Chapman and Hall, 1997. 526 p.

LAYARD, M. N. J. Robust large-sample tests for homogeneity of variances. **Journal of the American Statistical Association,** Alexandria, v. 68, n. 341, p. 195-198, Mar. 1973.

LEVENE, H. Robust test for equality of variances. In: OLKIN, I. (Ed.). **Contributions to probability and statistics,** Palo Alto.: Stanford University Press, 1960. p. 278-292.

LOMAX, R. G. A statistical model of public and private schools. **Journal of Experimental Education,** Washington, v. 53, n. 4, p. 443-477, 1985.

MANLY, B.F.J. **Randomization bootstrap and Monte Carlo methods in biology**. London: Chapman-Hall, 1988.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analisys.** New York: Academic Press, 1982.

MILLER, R. G., Jr. Jackknifing variances. **Annals of Mathematical Statistics,** Baltimore, v. 39, n. 2, p. 567-582, Apr. 1968.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to theory of statistics.** 3. ed. New York: Wiley & Sons, 1974. 842 p.

MUIRHEAD, R. J. **Aspects of multivariate statistical theory.** New York: John Wiley, 1982.

NEYMAN, J.; PEARSON, E. S. **On the problem of k samples.** Bull. Plish Acad. Sci. A, 460-481 p.; 1931.

NOGUEIRA, D. A. **Proposta e avaliação de critérios de convergência para o método de Monte Carlo via cadeias de Markov: casos uni e multivariados.** 2004. p. 121. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

ROBERTSON, T.; WRIGHT, F. T.; DYKSTRA, R.L. **Other restricted statistical inference.** New York: Wiley, 1988.

ROY, S. N. On a heuristic method of test construction and its use in multivariate analisys. **Annals of Mathematical Statistics,** v. 24, n. 2, p. 220-238, 1953.

SAMIUDDIN, M. Bayesian test of homogeneity of variance. **Journal of the American Statistical Association,** Alexandria, v. 71, n. 354, p. 515-517, June 1976.

SANTOS, A. C. dos. **Definição do tamanho amostral usando simulação Monte Carlo para os testes de normalidade univariado e multivariado baseados em assimetria e curtose.** 2001. p.71. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) Universidade Federal de Lavras, Lavras, MG.

SAS INSTITUTE. **SAS for Windows, Realese 8.** Cary, N. C, 2000.

SATTHERTHWAITE, F. E. Na aproximate distribution of estimates of variance components. **Biometric Bulletin,** London, v. 2, p. 110-114, 1946.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods.** 7. ed. Ames: The lowa State University Press, 1980. 507 p.

TURELLI, M. Phenotypic evolution, constant covariances, and the maintenance of additive variance. **Evolution,** Washington, v. 42, p. 1342-1347, 1980.

ZHANG, J. I.; BOOS, D. D. Bootstrap critical values for testing homogeneity of covariance matrices. **Journal of the American Statistical Association,** Alexandria, v. 87, n. 418, p. 425-429, June 1992.

ZHANG, J. I.; BOOS, D. D. Testing hypothesis about covariance matrices using bootstrap methods. **Communications in Statistics - Theory Methods,** New York, v. 22, n. 3, p. 723-739, 1993.

ZHANG, J. I.; PANTULA, S. G.; BOOS, D. D. Robust methods for testing the pattern of a single covariance matrix. **Biometrika,** London, v. 78, n. 4, p. 787-795, Dec. 1993.

ZHU, L. X.; NG, K. W.; JING, P. Resampling methods for homogeneity tests of covariance matrices. **Statistica Sinica,** Taipei, v. 12, n. 3, p. 769-783, July 2002.

WILSON, E. B.; HILFERTY, M. M. The distribution of chi-square. **Proceedings of the National Academy of Science,** Washington, v. 17, p. 684-688, 1931.

ANEXO

PROGRAMAS UTILIZADOS NOS PROCEDIMENTOS DE SIMULAÇÃO E ANÁLISE

Programa de simulação para a avaliação do erro tipo I do teste de Bartlett multivariado, teste generalizado de Samiuddin e as suas versões bootstrap, considerando as combinações fatoriais entre as p variáveis, k populações e tamanho de amostral, n nível de significância α, sob distribuições normais e não-normais.

```
proc iml;
  alpha=0.05;p=3;k=2;
ni=j(k,1,0);ni[1]=5;ni[2]=5; ni[3]=20;ni[4]=20;ni[5]=20;
ni[6]=20;ni[7]=20;ni[8]=20;ni[9]=20;ni[10]=20;
print ni;
  n=sum(ni);
  nsim=600;nboot=600;
  sigma=j(p,p,0);/*gerar as sigmas e as amostras*/
  pho=0;
  do i=1 to p;
   do j=i to p;
    if j=i then sigma[i,j]=1;
      else do;
         sigma[i,j]=pho**(j-i);
         sigma[j,i]=sigma[i,j];
      end;
   end;
  end;
  F=t(root(Sigma));
  X=j(n,p,0);
  Xb=j(n,p,0);
  scov=j(p*k,p,0);
  print alpha p n k;
  print sigma F;
  z=j(p,1,0);
  *print q;
```

```
  nsig_1=0;
  nsig_2=0;
  nsig_3=0;
  nsig_1b=0;nsig_2b=0;nsig_3b=0;
  ct_simul=0;
  do i=1 to nsim;
      do j=1 to n;
       do jj=1 to p;
              z[jj]=rannor(0);
        *z[jj]=ranuni(0);
        *z[jj]=rangam(0,1.5);
       end;
       X[j,]=t(F*Z);
      end;
      *print X;
      aux=0;
      aux1=0;aux2=0;
      spool=j(p,p,0);
      do j=1 to k;
          xx=x[aux+1:aux+ni[j],1:p];
          aux=aux+ni[j];
          *print xx;
          Q=i(ni[j])-j(ni[j],ni[j],1)/ni[j];
          Sj=(1/(ni[j]-1))*t(XX)*Q*XX;
          *print Sj;
          spool=spool+(ni[j]-1)*Sj/(n-k);
                      auxaux=det(Sj);
          if auxaux<=0 then auxau=log(1e-10);
          else auxau=log(auxaux);
          Scov[(j-1)*p+1:(j-1)*p+p,1:p]=Sj;
          aux1=aux1+(ni[j]-1)*auxau;
          aux2=aux2+1/(ni[j]-1);
      end;
/*teste de Bartlett multivariado*/
      aux2=1-(aux2-1/(n-k))*((2*p**2+3*p-1)/(6*(p+1)*(k-1)));
  quic1=-aux2*(aux1-(n-k)*log(det(spool)));
  if quic1<=0 then quic1=1e-19;
      v1=(k-1)*p*(p+1)/2;
      prqui1=1-probchi(quic1,v1);
      if prqui1<=alpha then nsig_1=nsig_1+1/nsim;
      *print v1 quic1 prqui1;
/*teste de Bartlett univariado adaptado*/
```

```
      aux3=-1*(aux2-1)+1;
    quic2=-aux3*(aux1-(n-k)*log(det(spool)));
    if quic2<=0 then quic2=1e-19;
      v2=(k-1)*p*(p+1)/2;
      prqui2=1-probchi(quic2,v2);
      if prqui2<=alpha then nsig_2=nsig_2+1/nsim;
      *print v2 quic2 prqui2;
/*Samiuddin - adaptado 1-Daniel*/
      Spool_inv=ginv(spool);
      *print spool_inv;
      qui2_rob=0;
      do j=1 to k;
         sj=Scov[(j-1)*p+1:(j-1)*p+p,1:p];
         squi2=trace((ni[j]-1)*Sj*Spool_inv);
         *print squi2;
         If squi2<1e-10 then Zi=0;
              else Zi=(1/((ni[j]-1)*p)*squi2)**(1/3)-(1-2/(9*(ni[j]-1)*p));
         Zi=zi/(2/(9*(ni[j]-1)*p))**0.5;
         qui2_rob=qui2_rob+zi*zi;
      end;
      prqui_rob=1-probchi(qui2_rob,k-1);
      if prqui_rob<=alpha then nsig_3=nsig_3+1/nsim;
      *print qui2_rob prqui_rob;
      *print scov;
      sig_1b=0;sig_2b=0;sig_3b=0;
      do ii=1 to nboot;
           do jj=1 to n;
             kk=int(ranuni(0)*n)+1;
             xb[jj,]=x[kk,];
           end;
           aux=0;
           aux1=0;aux2=0;
           spool=j(p,p,0);
           do j=1 to k;
             xx=xb[aux+1:aux+ni[j],1:p];
             aux=aux+ni[j];
             *print xxb;
             q=i(ni[j])-j(ni[j],ni[j],1)/ni[j];
            Sj=(1/(ni[j]-1))*t(XX)*Q*XX;
             *print Sj;
             spool=spool+(ni[j]-1)*Sj/(n-k);
             Scov[(j-1)*p+1:(j-1)*p+p,1:p]=Sj;
```

```
           auxaux=det(Sj);
            if auxaux<=0 then auxau=log(1e-10);
           else auxau=log(auxaux);
           *print auxaux;
           aux1=aux1+(ni[j]-1)*auxau;
          aux2=aux2+1/(ni[j]-1);
     end;
/*teste de Bartlett multivariado*/
          aux2=1-(aux2-1/(n-k))*((2*p**2+3*p-1)/(6*(p+1)*(k-1)));
              roberta=det(spool);
              if roberta<=0 then roberta=1e-10;
              quic1b=-aux2*(aux1-(n-k)*log(roberta));
          if quic1b>=quic1 then sig_1b=sig_1b+1/nboot;
          *print quic1b quic1;
/*teste de Bartlett univariado adaptado*/
          aux3=-1*(aux2-1)+1;
              quic2b=-aux3*(aux1-(n-k)*log(roberta));
          if quic2b>=quic2 then sig_2b=sig_2b+1/nboot;
*print quic2b quic2;
/*Samiuddin - adaptado 1-Daniel*/
          Spool_inv=ginv(spool);
          *print spool_inv;
          qui2_rob_b=0;
          do j=1 to k;
            sj=Scov[(j-1)*p+1:(j-1)*p+p,1:p];
            squi2=trace((ni[j]-1)*Sj*Spool_inv);
*print squi2;
                        if squi2<1e-10 then Zi=0;
                 else Zi=(1/((ni[j]-1)*p)*squi2)**(1/3)-(1-2/(9*(ni[j]-1)*p));
            Zi=zi/(2/(9*(ni[j]-1)*p))**0.5;
            qui2_rob_b=qui2_rob_b+zi*zi;
          end;
          if qui2_rob_b>=qui2_rob then sig_3b=sig_3b+1/nboot;
*print qui2_rob_b qui2_rob;
*print scov;
        end;*bootstrap;
        if sig_1b<=alpha then nsig_1b=nsig_1b+1/nsim;
        if sig_2b<=alpha then nsig_2b=nsig_2b+1/nsim;
        if sig_3b<=alpha then nsig_3b=nsig_3b+1/nsim;
           ct_simul=ct_simul+1;
  end;
print "numero de simulaçoes:                    " ct_simul;
```

```
print "Teste de Bartlett multivariada:          " nsig_1;
print "Teste de Bartlett univariada:            " nsig_2;
print "Teste de Samiuddin adaptado 1:           " nsig_3;
print "Teste de Bartlett multivariada (boot.):  " nsig_1b;
print "Teste de Bartlett univariada (boot.):    " nsig_2b;
print "Teste de Samiuddin adaptado 1 (boot.):   " nsig_3b;
quit;
```

Programa de simulação de avaliação para o poder dos testes de Bartlett multivariado, teste generalizado de Samiuddin e as suas versões bootstrap, considerando as combinações fatoriais entre as p variáveis, k populações e tamanho de amostral n, nível de significância α, sob distribuições normais e não-normais.

```
proc iml;
  title "daqui ate la em baixo e normal";
  alpha=0.05;p=2;k=5;delta=2;
  ni=j(k,1,0);
  ni[1]=20;
  ni[2]=20;
  ni[3]=20;
  ni[4]=20;
  ni[5]=20;/*
  ni[6]=5;
  ni[7]=5;
  ni[8]=10;
  ni[9]=5;
  ni[10]=5; */
  print ni;
  n=sum(ni);
  nsim=600;nboot=600;
  sigma=j(p,p,0);/*gerar as sigmas e as amostras*/
  pho=0.5;
  do i=1 to p;
   do j=i to p;
    if j=i then sigma[i,j]=1;
      else do;
         sigma[i,j]=pho**(j-i);
         sigma[j,i]=sigma[i,j];
```

```
    end;
 end;
end;
/*Gerando diferentes sigmas*/
diff=(delta-1)/(k-1);
detdelta=det(sigma*((1+diff*(k-1))**(1/p)))/det(sigma);
F=t(root(Sigma));
*sigmacov=j(p*k,p,0);
Fcov=j(p*k,p,0);
*sigmacov[1:p,1:p]=sigma;
Fcov[1:p,1:p]=F;
do i=2 to k;
  *sigmacov[(i-1)*p+1:(i-1)*p+p,1:p]=sigma*((1+diff*(i-1))**(1/p));
  *Fcov[(i-1)*p+1:(i-1)*p+p,1:p]=t(root(sigmacov[(i-1)*p+1:(i-1)*p+p,1:p]));
        Fcov[(i-1)*p+1:(i-1)*p+p,1:p]=t(root(sigma*((1+diff*(i-1))**(1/p))));
end;
X=j(n,p,0);
Xb=j(n,p,0);
scov=j(p*k,p,0);
print alpha p n k delta diff;
*print sigma F;
z=j(p,1,0);
*print q;
nsig_1=0;
nsig_2=0;
nsig_3=0;
nsig_1b=0;nsig_2b=0;nsig_3b=0;
ct_simul=0;
do i=1 to nsim;
    count=1;
    do ipop=1 to k;
            F=Fcov[(ipop-1)*p+1:(ipop-1)*p+p,1:p];
      do j=1 to ni[ipop];
        do jj=1 to p;
          z[jj]=rannor(0);
          *z[jj]=rangam(0,1.5);
*z[jj]=ranuni(0,0.5);
        end;
        X[count,]=t(F*Z);
                  count=count+1;
            end;
    end;
```

```
    *print X;
    aux=0;
    aux1=0;aux2=0;
    spool=j(p,p,0);
    do j=1 to k;
       xx=x[aux+1:aux+ni[j],1:p];
       aux=aux+ni[j];
       *print xx;
       Q=i(ni[j])-j(ni[j],ni[j],1)/ni[j];
       Sj=(1/(ni[j]-1))*t(XX)*Q*XX;
       *print Sj;
       spool=spool+(ni[j]-1)*Sj/(n-k);
                   auxaux=det(Sj);
       if auxaux<=0 then auxau=log(1e-10);
       else auxau=log(auxaux);
       Scov[(j-1)*p+1:(j-1)*p+p,1:p]=Sj;
       aux1=aux1+(ni[j]-1)*auxau;
       aux2=aux2+1/(ni[j]-1);
    end;
    /*teste de Bartlett multivariado*/
    aux2=1-(aux2-1/(n-k))*((2*p**2+3*p-1)/(6*(p+1)*(k-1)));
quic1=-aux2*(aux1-(n-k)*log(det(spool)));
if quic1<=0 then quic1=1e-19;
    v1=(k-1)*p*(p+1)/2;
    prqui1=1-probchi(quic1,v1);
    if prqui1<=alpha then nsig_1=nsig_1+1/nsim;
    *print v1 quic1 prqui1;
    /*teste de Bartlett univariado adaptado*/
    aux3=-1*(aux2-1)+1;
quic2=-aux3*(aux1-(n-k)*log(det(spool)));
if quic2<=0 then quic2=1e-19;
    v2=(k-1)*p*(p+1)/2;
    prqui2=1-probchi(quic2,v2);
    if prqui2<=alpha then nsig_2=nsig_2+1/nsim;
    *print v2 quic2 prqui2;
    /*Samiuddin - adaptado 1-Daniel*/
    Spool_inv=ginv(spool);
    *print spool_inv;
    qui2_rob=0;
    do j=1 to k;
       sj=Scov[(j-1)*p+1:(j-1)*p+p,1:p];
       squi2=trace((ni[j]-1)*Sj*Spool_inv);
```

```
    *print squi2;
        if squi2<1e-10 then Zi=0;
        else Zi=(1/((ni[j]-1)*p)*squi2)**(1/3)-(1-2/(9*(ni[j]-1)*p));
    zi=zi/(2/(9*(ni[j]-1)*p))**0.5;
    qui2_rob=qui2_rob+zi*zi;
end;
prqui_rob=1-probchi(qui2_rob,k-1);
if prqui_rob<=alpha then nsig_3=nsig_3+1/nsim;
*print qui2_rob prqui_rob;
*print scov;
sig_1b=0;sig_2b=0;sig_3b=0;
do ii=1 to nboot;
    do jj=1 to n;
      kk=int(ranuni(0)*n)+1;
      xb[jj,]=x[kk,];
    end;
        *print X xb;
    aux=0;
    aux1=0;aux2=0;
    spool=j(p,p,0);
    do j=1 to k;
      xx=xb[aux+1:aux+ni[j],1:p];
      aux=aux+ni[j];
      *print xx;
      Q=i(ni[j])-j(ni[j],ni[j],1)/ni[j];
      Sj=(1/(ni[j]-1))*t(XX)*Q*XX;
      *print Sj;
      spool=spool+(ni[j]-1)*Sj/(n-k);
      Scov[(j-1)*p+1:(j-1)*p+p,1:p]=Sj;
      auxaux=det(Sj);
                if auxaux<=0 then auxau=log(1e-10);
          else auxau=log(auxaux);
       *print auxaux;
      aux1=aux1+(ni[j]-1)*auxau;
      aux2=aux2+1/(ni[j]-1);
    end;
    /*teste de Bartlett multivariado*/
    aux2=1-(aux2-1/(n-k))*((2*p**2+3*p-1)/(6*(p+1)*(k-1)));
        roberta=det(spool);
        *print roberta;
        if roberta<=0 then roberta=1e-10;
        quic1b=-aux2*(aux1-(n-k)*log(roberta));
```

```
        if quic1b>=quic1 then sig_1b=sig_1b+1/nboot;
        *print roberta quic1b quic1;
            *print sig_1b;
     /*teste de Bartlett univariado adaptado*/
        aux3=-1*(aux2-1)+1;
            quic2b=-aux3*(aux1-(n-k)*log(roberta));
        if quic2b>=quic2 then sig_2b=sig_2b+1/nboot;
        *print quic2b quic2;
        /*Samiuddin - adaptado 1 - Daniel*/
        Spool_inv=ginv(spool);
        *print spool_inv;
        qui2_rob_b=0;
        do j=1 to k;
          sj=Scov[(j-1)*p+1:(j-1)*p+p,1:p];
          squi2=trace((ni[j]-1)*Sj*Spool_inv);
          *print squi2;
                    if squi2<1e-10 then Zi=0;
              else Zi=(1/((ni[j]-1)*p)*squi2)**(1/3)-(1-2/(9*(ni[j]-1)*p));
          Zi=zi/(2/(9*(ni[j]-1)*p))**0.5;
          qui2_rob_b=qui2_rob_b+zi*zi;
        end;
        if qui2_rob_b>=qui2_rob then sig_3b=sig_3b+1/nboot;
        *print qui2_rob_b qui2_rob;
        *print scov;
      end;*bootstrap;
      if sig_1b<=alpha then nsig_1b=nsig_1b+1/nsim;
      if sig_2b<=alpha then nsig_2b=nsig_2b+1/nsim;
      if sig_3b<=alpha then nsig_3b=nsig_3b+1/nsim;
          ct_simul=ct_simul+1;
 end;
 print "Poder - proporcao de analises significativas de cada procedimento";
 print "numero de simulaçoes:                  " ct_simul;
 print "Para conferencia apenas veja se detdelta=delta";
 print "Relacao entre os determinantes orig.:   " detdelta;
 print "Teste de Bartlett multivariada:         " nsig_1;
 print "Teste de Bartlett univariada:           " nsig_2;
 print "Teste de Samiuddin adaptado 1:          " nsig_3;
 print "Teste de Bartlett multivariada (boot.): " nsig_1b;
 print "Teste de Bartlett univariada (boot.):   " nsig_2b;
 print "Teste de Samiuddin adaptado 1 (boot.):  " nsig_3b;
quit;
```