



DEYSE MÁRCIA PACHECO GEBERT

**UMA SOLUÇÃO VIA *BOOTSTRAP*
PARAMÉTRICO PARA O PROBLEMA DE
BEHRENS-FISHER MULTIVARIADO**

LAVRAS-MG

2014

DEYSE MÁRCIA PACHECO GEBERT

**UMA SOLUÇÃO VIA *BOOTSTRAP* PARAMÉTRICO PARA O
PROBLEMA DE BEHRENS-FISHER MULTIVARIADO**

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

Orientador
Dr. Daniel Furtado Ferreira

**LAVRAS-MG
2014**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Gebert, Deyse Márcia Pacheco.

Uma solução via *bootstrap* paramétrico para o problema de Behrens-Fisher multivariado / Deyse Márcia Pacheco Gebert. - Lavras : UFLA, 2014.

120 p. : il.

Tese (Doutorado) - Universidade Federal de Lavras, 2014.

Orientador: Daniel Furtado Ferreira.

Bibliografia.

1.Heterocedasticidade. 2. Teste de vetores de médias. 3. *Bootstrap* paramétrico. 4. Problema de Behrens-Fisher multivariado.

I. Universidade Federal de Lavras. II. Título.

CDD-519.535

DEYSE MÁRCIA PACHECO GEBERT

**UMA SOLUÇÃO VIA *BOOTSTRAP* PARAMÉTRICO PARA O
PROBLEMA DE BEHRENS-FISHER MULTIVARIADO**

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

APROVADA em 29 de janeiro de 2014.

Dr. Denismar Alves Nogueira UNIFAL-MG

Dr. Eric Batista Ferreira UNIFAL-MG

Dr. Júlio Sílvio de Sousa Bueno Filho UFLA

Dr. Marcelo Tavares UFU

Dr. Daniel Furtado Ferreira
Orientador

**LAVRAS-MG
2014**

À Minha Família, o alicerce da minha vida.

DEDICO

AGRADECIMENTOS

A Deus.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade concedida para realização do doutorado.

À Fundação de Amparo a Pesquisa do Estado de Minas Gerais (Fapemig), pela concessão de bolsa de estudos.

Aos membros da banca, pela disponibilidade em ajudar a finalizar este trabalho.

Ào meu orientador Daniel, pela sua grande ajuda, atenção e confiança depositada em mim.

À secretária de pós-graduação, Josi, por mostrar-se sempre simpática e prestativa nas ocasiões em que precisei dela.

À minha família pelo apoio incondicional e compreensão.

À todos os meus amigos, que de alguma forma contribuíram para essa conquista.

Muito Obrigada!

"Por vezes sentimos que aquilo que fazemos
não é senão uma gota de água no mar. Mas
o mar seria menor se lhe faltasse uma gota".

Madre Teresa de Calcuta

RESUMO

Na estatística aplicada é muito comum a necessidade de comparação entre médias. Em situações multivariadas, quando a necessidade está em testar vetores de médias de duas populações normais p -variadas, em que as suas matrizes de covariâncias são diferentes entre si e desconhecidas, está caracterizado o chamado problema de Behrens-Fisher multivariado. Nessa situação, o teste T^2 de Hotelling, utilizado quando as covariâncias entre as populações são homogêneas, torna-se inapropriado, especialmente quando os tamanhos amostrais são diferentes. Por essa razão, várias soluções aproximadas foram propostas, como as de Krishnamoorthy e Lu (2010), Krishnamoorthy e Yu (2004) e Nel e Merwe (1986), entre outras. Krishnamoorthy e Yu (2004) reforçam que uma solução exata com propriedades naturais não existe, havendo uma abertura para desenvolver soluções mais eficientes. Já Krishnamoorthy e Lu (2010) apresentaram uma solução para a análise de variância multivariada (MANOVA), sob heterogeneidade de variâncias, mas não avaliaram o desempenho do teste quando a MANOVA se restringe a apenas duas populações. Para tanto, o objetivo deste trabalho é propor um teste, para a solução do problema de Behrens-Fisher multivariado, baseado em bootstrap paramétrico e avaliar o seu desempenho, bem como a sua comparação com o teste de Nel e Merwe modificado e o teste de Krishnamoorthy e Lu (2010). O método via bootstrap paramétrico (*TBP*) foi proposto e a avaliação do seu desempenho e comparação com os demais testes foram realizadas. As conclusões alcançadas sobre o desempenho dos testes em questão foram divididas em dois casos. O primeiro caso, em que as matrizes de covariâncias, das duas populações envolvidas, possuem estrutura equicorrelacionada, concluiu-se que o *TBP* é superior aos seus competidores em todas as situações estudadas, inclusive sob homogeneidade de covariâncias. O teste de Krishnamoorthy e Lu (2010) foi considerado intermediário e o teste de Nel e Merwe modificado o de pior desempenho, em relação ao teste proposto. No segundo caso, as matrizes de covariâncias das populações envolvidas são não estruturadas e concluiu-se que o *TBP* deve ser utilizado em duas ocasiões: amostras pequenas, de mesmo tamanho, com alto número de variáveis e amostras com tamanhos diferentes e também com grande número de variáveis.

Palavras-chave: Heterocedasticidade. Teste de vetores de médias. *Bootstrap* paramétrico. Problema de Behrens-Fisher multivariado.

ABSTRACT

In applied statistics, the need to compare means is very common. In the multivariate cases when there is a need for testing mean vectors of two p -varied normal populations with unknown and different covariance matrices the Behrens-Fisher multivariate problem is characterized. In this case the Hotelling T^2 test applied when the covariances population matrices are homogeneous is inappropriate, especially when the sample sizes are different. For this reason, many approximate solutions were proposed, such as Nel and Merwe (1986), Krishnamoorthy and Yu (2004) and Krishnamoorthy and Lu (2010), among others. Krishnamoorthy and Yu (2004) reinforce that an exact solution with natural properties does not exist and that efforts are needed to develop more efficient solutions. Krishnamoorthy and Lu (2010) presented a solution for the multivariate variance analysis (MANOVA), under heteroscedasticity. However, they did not evaluate the performance of this test when MANOVA is restricted to only two populations. Thus, the objective of this work is to propose a test, for solving the Behrens-Fisher multivariate problem, based on parametric bootstrap, and evaluate its performance, as well as its comparison to the modified Nel and Merwe test and the Krishnamoorthy and Lu (2010) test. A parametric bootstrap test (PBT) was proposed and the evaluation of its performance and comparison to the other tests was conducted by Monte Carlo simulations. The conclusions reached on the test performance were divided into two cases. The first case, in which the covariance matrices of both populations have equicorrelated structure, the PBT is superior to its competitors in all studied situations, including under covariance homogeneity. The Krishnamoorthy and Lu (2010) test was considered intermediate and the modified Nel Merwe test presented the worst performance. In the second case, the covariance matrices of the populations involved are non-structured and the PBT should only be used in two circumstances: with small sample size of same size in both samples associated with large number of variables, and in samples with different sizes, also with a large number of variables.

Keywords: Heteroscedasticity. Mean vector test. Parametric bootstrap. Multivariate Behrens-Fisher problem.

LISTA DE FIGURAS

Figura 1	Curva da função poder em relação a um parâmetro θ	33
Figura 2	Taxas de erro tipo I do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99	49
Figura 3	Taxas de erro tipo I do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$, $n_2 = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99	54
Figura 4	Taxas de erro tipo I do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99	56
Figura 5	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	58

Figura 6	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	59
Figura 7	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	61
Figura 8	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	64
Figura 9	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	65
Figura 10	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	66
Figura 11	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	68

Figura 12 Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	70
Figura 13 Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	71
Figura 14 Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 8$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	73
Figura 15 Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 8$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	74
Figura 16 Taxas de erro tipo I do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função de tamanhos amostrais iguais, $n_1 = n_2 = 8$; $n_1 = n_2 = 30$ e $n_1 = n_2 = 100$, número de variáveis p e nível nominal de significância α , sob H_0 . Cada par de linhas horizontais indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, respectivamente, com coeficiente de confiança de 0,99	79

Figura 17 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função de tamanhos amostrais diferentes, número de variáveis p e nível nominal de significância α , sob H_0 . Cada par de linhas horizontais indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, respectivamente, com coeficiente de confiança de 0,99 80

LISTA DE TABELAS

Tabela 1	Resultados possíveis de um teste de hipótese e suas probabilidades de ocorrência	32
Tabela 2	Taxas de erro tipo I do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0	52
Tabela 3	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$ e $n_2 = 8$, $k = 4$ e 8 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α	76
Tabela 4	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função de tamanhos amostrais iguais, $k = 1$ e 2 , número de variáveis (p) e nível nominal de significância α	83
Tabela 5	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$; $n_2 = 30$ e $n_1 = 30$; $n_2 = 8$, $k = 1$ e 2 , número de variáveis (p) e nível nominal de significância α	85
Tabela 6	Poder do teste <i>bootstrap</i> paramétrico proposto (TBP), do teste <i>bootstrap</i> paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$; $n_2 = 100$ e $n_1 = 100$; $n_2 = 8$, $k = 1$ e 2 , número de variáveis (p) e nível nominal de significância α	87

SUMÁRIO

1	INTRODUÇÃO	16
2	REFERENCIAL TEÓRICO	18
2.1	O problema de Behrens-Fisher univariado	18
2.2	O Problema de Behrens-Fisher multivariado	21
2.3	Simulação <i>bootstrap</i>	27
2.3.1	<i>Bootstrap</i> paramétrico	28
2.3.2	<i>Bootstrap</i> não-paramétrico	28
2.4	Simulação Monte Carlo	29
2.5	Erro tipo I, erro tipo II, poder e robustez	31
2.6	Distribuição normal multivariada	34
3	MÉTODOS	37
3.1	Testes	38
3.1.1	Teste de Nel e Merwe Modificado - TNVMM	38
3.1.2	Teste de Krishnamoorthy e Lu - TBPCL	38
3.1.3	Teste <i>bootstrap</i> paramétrico proposto - TBP	40
3.2	Validação Monte Carlo	41
3.2.1	Erro Tipo I	42
3.2.2	Poder	45
4	RESULTADOS E DISCUSSÃO	47
4.1	Caso de matrizes de covariâncias equicorrelacionadas	47
4.1.1	Taxas de erro tipo I	47
4.1.2	Poder	55
4.2	Caso de matrizes de covariâncias não estruturadas	78
4.2.1	Erro tipo I	78
4.2.2	Poder	81
5	CONCLUSÃO	89
	REFERÊNCIAS	92
	ANEXOS	96

1 INTRODUÇÃO

A necessidade da comparação de médias de duas populações é uma das situações mais comuns dentro da estatística aplicada. Se estas populações são normais e possuem variâncias iguais, a solução para esse problema é muito conhecida na literatura. No caso unidimensional, o teste de hipótese clássico indicado é o teste t . Entretanto, quando as variâncias apresentarem heterogeneidade e forem desconhecidas, o teste t não deve ser utilizado pois nessa situação a distribuição da estatística do teste não segue exatamente a distribuição t de Student. O primeiro a apresentar uma aproximação foi Behrens (1929) e seis anos depois Fisher (1935) mostrou que tal situação poderia ser justificada utilizando-se a teoria da inferência sob o aspecto frequentista. Por serem os primeiros a estudar essa situação, tal problema ficou conhecido como “Problema de Behrens-Fisher”.

Para o caso multidimensional com matrizes de covariâncias consideradas iguais, para se testar a igualdade entre vetores de médias pode-se utilizar o teste T^2 de Hotelling (FERREIRA, 2011). Ocasões em que as matrizes de covariâncias não puderem ser consideradas homogêneas, o teste mencionado anteriormente torna-se inapropriado, especialmente quando os tamanhos das amostras forem diferentes (LIX; KESELMAN; HINDS, 2005). Com isso, várias alternativas para a solução do problema de Behrens-Fisher multivariado foram propostas na literatura, como as de Krishnamoorthy e Lu (2010), Krishnamoorthy e Yu (2004) e Nel e Merwe (1986), entre outras.

A solução apresentada por Krishnamoorthy e Yu (2004) é uma modificação da proposta de Nel e Merwe (1986). Os autores reforçam que uma solução exata com propriedades naturais ainda não existe, havendo uma abertura para desenvolver soluções mais eficientes. Já Krishnamoorthy e Lu (2010) apresentaram uma

solução para a análise de variâncias multivariada (MANOVA), sob heterogeneidade de variâncias, mas não avaliaram o desempenho do teste quando a MANOVA se restringe a apenas duas populações.

Neste trabalho, objetivou-se propor um método para a solução do problema de Behrens-Fisher multivariado, baseado em *bootstrap* paramétrico, e avaliar o seu desempenho. Além disso, objetivou-se também comparar o teste proposto com o teste de Nel e Merwe (1986) modificado por Krishnamoorthy e Yu (2004) e o teste apresentado por Krishnamoorthy e Lu (2010). O desempenho dos três testes em estudo foram avaliados por meio do erro tipo I e poder, sob diferentes graus de heterogeneidade de covariâncias, tamanhos amostrais, níveis de significância e diferenças entre as médias das duas populações.

2 REFERENCIAL TEÓRICO

Nesta seção serão apresentados o problema de Behrens Fisher univariado e multivariado, algumas soluções já propostas, na literatura, para este problema e alguns conceitos importantes utilizados para realização deste trabalho.

2.1 O problema de Behrens-Fisher univariado

Seja Y_{11}, \dots, Y_{1n_1} uma amostra aleatória, de tamanho n_1 , com distribuição normal, média μ_1 e variância σ_1^2 e também Y_{21}, \dots, Y_{2n_2} uma amostra aleatória, de tamanho n_2 , com distribuição normal, média μ_2 e variância σ_2^2 . O interesse está na quantidade $\mu_1 - \mu_2$ e se quer testar a hipótese:

$$H_0 : \mu_1 - \mu_2 = \delta_0,$$

em que δ_0 é um valor fixado de interesse.

O caso mais simples surge quando as variâncias de ambas as populações, σ_1^2 e σ_2^2 , são conhecidas. Neste caso a estatística Z ,

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

é usada. Sob H_0 , Z tem distribuição normal padrão e o teste pode ser conduzido baseado nesse fato (FERREIRA, 2009).

A situação mais complicada ocorre quando as variâncias das populações são desconhecidas. Nesse caso, o teste Z não pode ser aplicado, pois as variâncias precisam ser estimadas. Se for assumido que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, é razoável combinar as duas amostras e estimar σ_c^2 , ao invés de usar os estimadores S_1^2 e S_2^2 separadamente. O estimador mais usual para a variância comum é dado por:

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

sendo S_1^2 e S_2^2 as variâncias amostrais de Y_1 e Y_2 respectivamente.

Então, a estatística do teste

$$T_1 = \frac{\bar{Y}_1 - \bar{Y}_2 - \delta_0}{\sqrt{S_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

é usada, a qual tem distribuição t de Student com $\nu_1 = n_1 + n_2 - 2$ graus de liberdade, sob H_0 .

Um caso mais realista surge quando as duas variâncias são diferentes ($\sigma_1^2 \neq \sigma_2^2$). Nesta situação, que é conhecida como o problema de Behrens-Fisher, a estatística do teste

$$T_2 = \frac{\bar{Y}_1 - \bar{Y}_2 - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (1)$$

não segue distribuição t de Student exata. A dificuldade está no fato de que

$$\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{-1} \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)$$

não segue distribuição χ^2 . Entretanto, Satterthwaite (1946) propôs um método o qual se aproxima da distribuição exata pela distribuição χ^2 . Logo, usando a aproximação de Satterthwaite

$$\nu_2 \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{-1} \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)$$

apresenta distribuição aproximadamente Qui-quadrado com ν_2 graus de liberdade. Então T_2 , dado em (1), segue aproximadamente uma distribuição t de Student com

ν_2 graus de liberdade, que é estimado por:

$$\hat{\nu}_2 = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2 \left\{ \frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1} \right\}^{-1}.$$

Cochran e Cox (1957) também propuseram uma aproximação para a estatística T_2 (1), a qual é comparada ao valor da distribuição t de Student dado por

$$t^* = \frac{W_1 t_1 + W_2 t_2}{W_1 + W_2}$$

em que, t^* refere-se ao valor determinante da região de rejeição da hipótese de igualdade das médias populacionais; $W_1 = S_1^2/n_1$ e $W_2 = S_2^2/n_1$; t_1 e t_2 são os valores tabelados da distribuição t (unilateral) com $n_1 - 1$ e $n_2 - 1$ graus de liberdade, respectivamente, no nível de significância estipulado previamente.

Borges e Ferreira (1999), utilizando simulações Monte Carlo, avaliaram o desempenho das aproximações de Cochran e Cox (1957) e Satterthwaite (1946). Os autores concluíram que os dois métodos são bem semelhantes em relação ao controle dos erros tipo I e II.

Schechtman e Sherman (2007) apresentaram uma alternativa para o problema de Behrens-Fisher considerando conhecida a razão entre as variâncias das populações em que se quer comparar as médias. Relatam que existem situações em que as variâncias das populações não são conhecidas, mas a razão entre elas sim. A solução proposta por eles, apresentou melhor controle de erro tipo I e poder que a aproximação de Satterthwaite (1946) na maioria das situações em estudo.

Segundo Steland, Padmanabhan e Akran (2011), nas áreas de engenharia e finanças são muito frequentes as situações em que não se pode utilizar métodos paramétricos para a comparação de médias com heterogeneidade de variâncias, decorrente da falta de grandes amostras. Com isso, os autores propuseram um

teste *bootstrap* baseado em estatísticas de ordem. O teste proposto apresentou um excelente desempenho, em relação as taxas de erro tipo I e poder, comparado ao teste de Wilcoxon (1945), que também contempla o problema de Behrens-Fisher não paramétrico.

Assim, como o problema de Behrens-Fisher se estende para situações não paramétricas, ele também é abordado em situações p -dimensionais, ou seja, situações onde se utilizam mais de uma variável. A seção 2.2 a seguir trata deste problema.

2.2 O Problema de Behrens-Fisher multivariado

No caso multivariado, em que duas populações normais multivariadas possuem vetores de médias $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ e matrizes de covariâncias $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$, respectivamente, a hipótese que se deseja testar é dada por

$$H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0, \quad (2)$$

em que $\boldsymbol{\delta}_0$ é um vetor fixado de interesse.

Considerando que as matrizes de covariâncias das duas populações sejam iguais e desconhecidas, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, para se testar a hipótese nula (2) deve-se utilizar a estatística T^2 dada por

$$T^2 = \frac{n_1 n_2}{n} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 - \boldsymbol{\delta}_0)^\top \mathbf{S}_c^{-1} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 - \boldsymbol{\delta}_0) \quad (3)$$

em que n_1 é tamanho da amostra da população 1, n_2 o tamanho da amostra da população 2, $n = n_1 + n_2$, $\bar{\mathbf{Y}}_1$ é o estimador do vetor de médias da população 1, $\bar{\mathbf{Y}}_2$ o estimador do vetor de médias da população 2, \mathbf{S}_c é

$$\mathbf{S}_c = \frac{[(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2]}{(n - 2)},$$

sendo \mathbf{S}_1 o estimador da matriz de covariâncias populacional obtido na amostra 1 e \mathbf{S}_2 o estimador da matriz de covariâncias populacional obtido na amostra 2.

A distribuição da estatística (3), sob normalidade p -variada, é a T^2 de Hotelling, que sob a hipótese nula possui distribuição proporcional a F central. Assim deve-se rejeitar H_0 se o valor de T^2 for maior ou igual ao valor crítico

$$\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - 1 - p)} F_{\alpha, p, n_1 + n_2 - 1 - p}$$

sendo $F_{\alpha, p, n_1 + n_2 - 1 - p}$ o quantil superior $100\alpha\%$ da distribuição F com $f_1 = p$ e $f_2 = n_1 + n_2 - 1 - p$ graus de liberdade (FERREIRA, 2011).

As situações em que as covariâncias populacionais são heterogêneas e desconhecidas, ou seja, $\Sigma_1 \neq \Sigma_2$, ainda sob normalidade, são caracterizadas como o problema de Behrens-Fisher multivariado. Nesse caso, a comparação de vetores de médias não pode ser realizada através do teste T^2 de Hotelling. Nessa situação, o teste em questão torna-se inapropriado, especialmente quando os tamanhos das amostras forem diferentes (LIX; KESELMAN; HINDS, 2005). Por essa razão, várias soluções aproximadas foram propostas, como as de James (1954), Krishnamoorthy e Lu (2010), Krishnamoorthy e Yu (2004) e Nel e Merwe (1986), entre outras.

A solução proposta por James (1954) usa uma expansão assintótica de primeira ordem de formas quadráticas, permitindo utilizar uma aproximação para a distribuição da estatística

$$T^{*2} = \delta_0^\top \mathbf{S}_e^{-1} \delta_0, \quad (4)$$

como uma soma de variáveis que seguem uma distribuição qui-quadrado, em que

$$\mathbf{S}_e = \frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2}. \quad (5)$$

A estatística T^{*2} foi delineada para testar a hipótese nula dada em 2, quando $\delta_0 = 0$. Esta estatística deve ser comparada com o valor crítico $\chi_{\alpha,p}^2(A + B\chi_{\alpha,p}^2)$, em que

$$A = 1 + \frac{1}{2p} \sum_{i=1}^2 \frac{1}{n_i - 1} \left[\text{tr} \left(\frac{\mathbf{S}_e^{-1} \mathbf{S}_i}{n_i} \right) \right]^2,$$

$$B = \frac{1}{2p(p+2)} \sum_{i=1}^2 \frac{1}{n_i - 1} \left\{ 2 \text{tr} \left[\left(\frac{\mathbf{S}_e^{-1} \mathbf{S}_i}{n_i} \right)^2 \right] + \left[\text{tr} \left(\frac{\mathbf{S}_e^{-1} \mathbf{S}_i}{n_i} \right) \right]^2 \right\},$$

e $\chi_{\alpha,p}^2$ é o quantil superior 100 α % da distribuição qui-quadrado com p graus de liberdade.

A sugestão de Nel e Merwe (1986) utiliza a mesma estatística que James (1954), porém considera que esta seja aproximadamente distribuída como uma T^2 de Hotelling com dimensão p e ν graus de liberdade, em que

$$\nu = \frac{\text{tr}[(\mathbf{S}_e)^2] + [\text{tr}(\mathbf{S}_e)]^2}{\sum_{i=1}^2 \frac{1}{n_i - 1} \left\{ \text{tr} \left[\left(\frac{\mathbf{S}_i}{n_i} \right)^2 \right] + \left[\text{tr} \left(\frac{\mathbf{S}_i}{n_i} \right) \right]^2 \right\}}. \quad (6)$$

Rejeita-se a hipótese nula quando $T^{*2} > \nu p F_{\alpha,p,\nu+1-p} / (\nu + 1 - p)$.

Uma outra correção nos graus de liberdade foi proposta por Yao (1965). O autor também considera que a estatística do teste, dada em (4), tem aproximadamente distribuição T^2 de Hotelling, mas com graus de liberdade dados por:

$$\frac{1}{\nu} = \frac{1}{(T^{*2})^2} \sum_{i=1}^2 \frac{1}{(n_i - 1)} \left[(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^\top \mathbf{S}_e^{-1} \frac{\mathbf{S}_i}{n_i} \mathbf{S}_e^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) \right]^2.$$

Yao (1965) avaliou as taxas de erro tipo I e poder da sua proposta e comparou com o desempenho do procedimento de James (1954), concluindo que a sua metodologia foi mais conservativa.

Johansen (1980) também sugeriu um procedimento baseado na estatística de James (1954). Esse método consiste em dividir a estatística T^{*2} (4), pela quantidade C que é dada por:

$$C = p - 2D + \frac{6D}{p(p-1) + 2},$$

sendo D definido como:

$$D = \sum_{i=1}^2 \frac{1}{2(n_i - 1)} \left\{ \text{tr}(I - V^{-1}V_i)^2 + [\text{tr}(I - V^{-1}V_i)]^2 \right\}$$

em que

$$V_i = \left(\frac{\mathbf{S}_i}{n_i} \right)^{-1} \quad \text{para } i = 1, 2 \quad \text{e} \quad V = V_1 + V_2.$$

Com essa divisão a estatística do teste passa a se distribuir aproximadamente como uma $F_{p,\nu}$, em que F_α é o quantil superior da distribuição F com p e ν graus de liberdade. Sendo ν calculado por:

$$\nu = \frac{p(p+2)}{3D}.$$

Kim (1992) apresentou um método que equivale a maximização de razão de formas quadráticas para determinar a correção dos graus de liberdade. Com isso, a estatística do teste proposta pelo autor, que é determinada através de algumas expressões que envolvem transformações lineares e outras quantidades calculadas, possui distribuição F aproximada, sob H_0 .

Em um estudo feito por Christensen e Rencher (1997) foram avaliadas algumas soluções para o problema de Behrens-Fisher, entre elas as soluções de James (1954), Johansen (1980), Kim (1992), Nel e Merwe (1986) e Yao (1965). Segundo os autores, os testes que apresentaram menores taxas de erro tipo I e maior poder foram as soluções de Kim (1992) e Nel e Merwe (1986), sendo esses recomendados pelos autores.

Krishnamoorthy e Yu (2004) propuseram uma modificação do teste de Nel e Merwe (1986) em que o ajuste nos graus de liberdade é dado por

$$\nu = \frac{p + p^2}{\sum_{i=1}^2 \frac{1}{n_i - 1} \left\{ \text{tr} \left[\left(\frac{\mathbf{S}_i \mathbf{S}_e^{-1}}{n_i} \right)^2 \right] + \left[\text{tr} \left(\frac{\mathbf{S}_i \mathbf{S}_e^{-1}}{n_i} \right) \right]^2 \right\}}.$$

A rejeição para H_0 se dá de forma semelhante a do teste de Nel e Merwe (1986). Além de propor essa modificação, os autores compararam seu desempenho com o teste original e observaram que a modificação teve igual poder para pequenas dimensões e foi mais poderosa para maiores dimensões. Krishnamoorthy e Yu (2004) dizem que uma solução exata com propriedades naturais não existe, havendo uma abertura para desenvolver soluções mais eficientes.

No trabalho de Cirillo e Ferreira (2004) foram comparados quatro métodos em suas versões originais e *bootstrap* com a estatística T^2 de Hotelling. Os autores concluíram que em situações de homogeneidade de covariâncias a proposta de

Kim (1992), via convencional, foi mais adequada quando as amostras possuem tamanhos iguais. Os métodos de Nel e Merwe (1986) e Yao (1965) foram os mais apropriados para situações com tamanhos de amostras diferentes. Para amostras com mesmo tamanho, os testes mais adequados variaram de acordo com o número de variáveis, grau de correlação entre as matrizes de covariâncias e nível nominal de significância. Apenas a proposta de Johansen (1980) apresentou taxas de erro tipo I aceitáveis, na versão *bootstrap*.

Krishnamoorthy e Lu (2010) propuseram uma solução para a MANOVA sob heterocedasticidade utilizando *bootstrap* paramétrico. Quando a MANOVA se restringe à comparação de vetores de médias de apenas duas populações sob heterocedasticidade, temos a caracterização do problema de Behrens-Fisher multivariado.

O teste proposto por Krishnamoorthy e Lu (2010) apresentou excelentes resultados para situações utilizando várias populações, mas os autores não avaliaram o desempenho do teste utilizando apenas duas populações. Além disso, apenas o erro tipo I foi estudado, não sendo feito nenhum tipo de avaliação em relação ao poder.

Existem muitas soluções para o problema de Behrens-Fisher na literatura mas, em geral, são muito complexas ou possuem limitações. Como por exemplo, a solução bayesiana apresentada por Johnson e Weerahandi (1988). Nessa solução os autores tinham como objetivo obter regiões de credibilidade elipsoidais *a posteriori* para a diferença entre duas médias oriundas de populações normais multivariadas. Os autores dessa solução não avaliaram seu desempenho por meio de simulações.

Ramos e Ferreira (2012) propuseram uma solução bayesiana baseada na proposta de Johnson e Weerahandi (1988). Os autores utilizaram uma distribuição

a priori conjugada para o vetor de médias populacionais (μ_i) e para a matriz de covariâncias (Σ_i), obtendo-se uma distribuição *a posteriori* *t* multivariada para μ_i . Ramos e Ferreira (2012) compararam o desempenho de sua proposta com o teste de Nel e Merwe modificado (KRISHNAMOORTHY; YU, 2004). O teste bayesiano proposto superou seu competidor apenas em algumas situações.

2.3 Simulação *bootstrap*

Dois dos problemas mais importantes em estatística são a determinação do estimador de um parâmetro de interesse e a avaliação de sua precisão. Efron (1979), ao introduzir sua versão de *bootstrap*, foi particularmente motivado por esses dois problemas. Por causa de sua generalidade, técnicas de *bootstrap* têm sido aplicadas a uma classe de problemas muito maior do que apenas obter estimativas de erros padrões e intervalos de confiança. Em várias disciplinas, como psicologia, geologia, econometria, biologia, engenharia, química, etc., empregam-se essas técnicas (CHERNICK, 2008).

Como já é sabido, o método de simulação *bootstrap* é baseado em reamostragens. Segundo Chernick (2008), o número de reamostragens necessárias para se obter boas estimativas, em intervalos de confiança e problemas de testes de hipóteses, seria de pelo menos 1000 repetições *bootstrap*. Para se fazer estas repetições, existem duas maneiras distintas: paramétrico e não-paramétrico.

Quando existe um modelo matemático particular, com constantes ajustáveis ou parâmetros que determinam totalmente a função de densidade ou de probabilidade, tal modelo é chamado de paramétrico. Métodos estatísticos baseados nesse modelo são métodos paramétricos. Quando nenhum modelo matemático é usado, a análise estatística é não-paramétrica, e usa apenas o fato de que as variáveis aleatórias Y_j são independentes e identicamente distribuídas. Mesmo

que exista um modelo paramétrico plausível para a situação, uma análise não paramétrica pode ser usada para avaliar a robustez das conclusões de uma análise paramétrica (DAVISON; HINKLEY, 2008).

2.3.1 *Bootstrap* paramétrico

Neste método as reamostragens são feitas a partir de uma função de distribuição conhecida F . Quando os dados são originários de uma função de distribuição paramétrica, então $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \sim i.i.d. F(x, \theta)$. Supondo que os dados observados são usados para estimar θ por $\hat{\theta}$, que frequentemente, mas nem sempre, é o estimador de máxima verossimilhança, então sua substituição no modelo fornece o modelo ajustado, $\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_n^* \sim i.i.d. F(x, \hat{\theta})$, o qual pode ser usado para calcular estimativas de parâmetros (DAVISON; HINKLEY, 2008). Quando o modelo é conhecido ou acredita-se ser uma boa representação da realidade, o *bootstrap* paramétrico pode ser uma ferramenta poderosa, permitindo fazer inferências em situações que de outra forma são intratáveis (GIVENS; HOETING, 2005).

Em um procedimento de *bootstrap* paramétrico, o primeiro passo é obter estimativas dos parâmetros que caracterizam a distribuição em questão. A partir desses parâmetros geram-se B amostras aleatórias, cada uma de tamanho n , e para cada amostra calcula-se uma estimativa da estatística de interesse, que tem a mesma forma funcional da estatística original. Com essa distribuição de *bootstrap* é possível fazer inferências sobre o parâmetro de interesse (GENTLE, 2003).

2.3.2 *Bootstrap* não-paramétrico

Uma outra maneira de se obter amostras *bootstrap* é o procedimento de reamostragem não-paramétrico. Ao contrário do método paramétrico, nesse caso

não são feitas pressuposições sobre a distribuição dos dados, ou seja, o conjunto de dados tem função de distribuição desconhecida.

As B pseudoamostras, de tamanho n , são obtidas através de reamostragem com reposição da amostra original. A probabilidade de selecionar qualquer valor do conjunto de dados é igual a $1/n$, pois considera-se que a amostra original é independente e identicamente distribuída com uma função de distribuição desconhecida. O passo seguinte é calcular a estimativa do parâmetro de interesse para cada uma das B pseudoamostras. Assim, tem-se a distribuição de *bootstrap* desse parâmetro, a partir da qual pode-se fazer inferências.

Nos casos em que o dados originais são correlacionados, ou seja, não são independentes e identicamente distribuídos, a aproximação da função de distribuição *bootstrap* ($\hat{F}^*(x)$) em relação a distribuição verdadeira ($F(x)$) não é segura. Métodos *bootstrap* para dados dependentes são muito bem descritos em Efron e Tibshirani (1993) e Kunsch (1989).

2.4 Simulação Monte Carlo

O método Monte Carlo herdou seu nome do mais famoso cassino da Europa, localizado em Mônaco. Distribuições de probabilidade envolvendo jogos de azar são frequentemente muito complicadas, mas podem ser avaliadas via simulações. Simulações computacionais referem-se a reprodução de processos, com um programa computacional adequado, e a observação desses resultados. Métodos Monte Carlo são aqueles baseados em simulações computacionais envolvendo números aleatórios (BARON, 2007).

O método Monte Carlo pode ser definido como a representação da solução de um problema, sendo utilizado sequência de números aleatórios para se fazer essa representação, com o objetivo de criar uma amostra da população. Sendo as-

sim possível o cálculo de estimativas de parâmetros representativos do problema. Segundo Dachs (1988), os métodos Monte Carlo são frequentemente utilizados para produzir amostras de acordo com distribuições teóricas conhecidas, objetivando avaliar novos comportamentos de diferentes técnicas estatísticas que podem ser empregadas em determinado problema. Nesses casos, esse método pode ser a única maneira satisfatória de prover as informações necessárias.

O primeiro passo para aplicação desse método seria a simulação de amostras, de tamanho n , seguindo distribuições previamente estruturadas, que serão avaliadas pelas novas técnicas estatísticas em estudo, considerando a hipótese nula estabelecida. Com as técnicas computacionais disponíveis, torna-se fácil a simulação de realizações de variáveis ou de amostras baseadas em modelos estatísticos apropriados de parâmetros conhecidos, com a finalidade de se verificar a adequação de determinada metodologia ou na realização de comparações entre métodos (DACHS, 1988).

No estudo do desempenho ou avaliação de testes estatísticos, frequentemente torna-se muito complicado obter analiticamente informações sobre taxas de erro tipo I e poder. Uma maneira de obter as informações desejadas de forma eficiente e simples, é o uso do método de simulação Monte Carlo, evitando assim eventuais dificuldades analíticas.

Vários autores já avaliaram o desempenho de testes estatísticos através de simulação, como por exemplo Cecchetti (1999) que estudou quatro critérios multivariados para o teste de igualdade de efeitos de tratamentos, Silva (1998) que estudou o desempenho do teste de Scott-Knott, Cirillo e Ferreira (2004) e Ramos e Ferreira (2012) que analisaram o problema de Behrens-Fisher no espaço p -dimensional, Amorim et al. (2010) na avaliação de testes para a unidimensionalidade de painéis sensoriais, Gebert e Ferreira (2010) na avaliação de teste para

retenção de componentes principais, entre outros.

2.5 Erro tipo I, erro tipo II, poder e robustez

De forma geral, o pesquisador depara-se muitas vezes com a necessidade de testar hipóteses sobre parâmetros. Quando uma hipótese nula é testada, tem-se apenas duas alternativas de decisão, rejeitar ou não rejeitar H_0 . Observa-se que ao tomar essa decisão, de rejeição ou não da hipótese nula, é possível estar incorrendo em erros, mesmo quando tudo é feito corretamente. Logo, é necessário considerá-los. O primeiro deles chama-se erro tipo I, e consiste em rejeitar a hipótese nula quando de fato ela é verdadeira. O segundo erro é aquele que é cometido quando H_0 não é rejeitada e na verdade ela é falsa, esse é o chamado erro tipo II (MOOD; GRAYBILL; BOES, 1974).

Em geral, a probabilidade de cometer o erro tipo I é representada pela letra grega α , sendo habitualmente chamado de nível de significância. A probabilidade de que se cometa o erro tipo II é simbolizada pela letra β e não tem nome específico. No entanto, quando não se rejeita H_0 e ela é verdadeira, ou quando se rejeita H_0 e ela é falsa são decisões corretas. Essa última decisão tem probabilidade de ocorrência $1 - \beta$, pois é o complemento do erro tipo II (β), e normalmente é chamado de poder do teste. O complemento da probabilidade de ocorrência do erro tipo I (α), que consiste em aceitar H_0 sendo ela ser verdadeira, é representada por $1 - \alpha$ e comumente chamado de coeficiente de confiança (OLIVEIRA et al., 2009). Esses conceitos estão sintetizados na Tabela 1.

Sabe-se que não é possível evitar esses erros quando se aplica um teste de hipótese, mas deve-se manter relativamente pequena a probabilidade de cometê-los. O pesquisador pode controlar diretamente apenas o erro tipo I, pois ele deve ser previamente fixado. Como α e β são inversamente proporcionais, a única ma-

Tabela 1 Resultados possíveis de um teste de hipótese e suas probabilidades de ocorrência

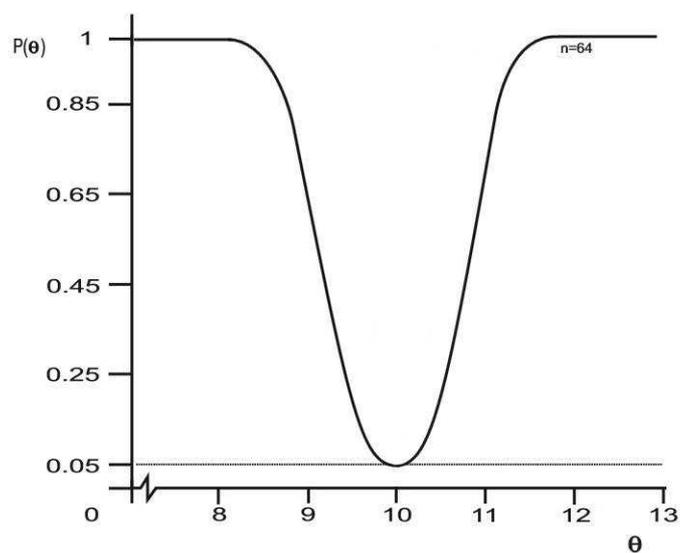
DECISÃO TOMADA	REALIDADE	
	H_0 é verdadeira	H_0 é falsa
H_0 é aceita	Decisão correta Probabilidade : $1 - \alpha$ (Coeficiente de confiança)	Decisão errada Probabilidade : β (Erro Tipo II)
H_0 é rejeitada	Decisão errada Probabilidade : α (Erro Tipo I)	Decisão correta Probabilidade : $1 - \beta$ (Poder do teste)

FONTE: Elaborado pela autora.

neira de causar redução simultânea dos dois erros é aumentando o tamanho da amostra. Devido a essa relação inversa, a fixação de α entre 0,10 e 0,01 é, sempre que possível, uma boa medida (FERREIRA, 2009). Testes que apresentam um valor real de taxas de erro tipo I menor que o nominal α são ditos conservativos ou rigorosos, ou seja, sob H_0 a probabilidade de se encontrar um resultado significativo é inferior ao valor α estabelecido. No entanto, quando testes trazem um valor real de taxa de erro tipo I maior que o nominal, estes são ditos liberais, pois a probabilidade de detectar resultado significativo é maior que o nível de significância estabelecido.

No desenvolvimento de novos testes, a avaliação do poder é um critério largamente empregado para a escolha, ou não, desse teste. Entretanto, o poder de um teste não é simples de ser calculado em situações reais, pois depende do valor desconhecido do parâmetro em questão.

De acordo com Ramalho, Ferreira e Oliveira (2000), o poder de um teste é a habilidade que ele tem em detectar todas as reais diferenças entre tratamentos, ou seja, é a probabilidade $(1 - \beta)$ de rejeitar H_0 quando ela de fato, é falsa. Costa



FONTE: Elaborado pela autora.

Figura 1 Curva da função poder em relação a um parâmetro θ

(1992) define poder como sendo “a sensibilidade da região crítica” para perceber e rejeitar uma hipótese falsa.

Na Figura 1 tem-se um exemplo de curva da função poder em relação a um parâmetro θ com valor hipotético igual a 10. No eixo das ordenadas estão os valores da função poder e na abscissa os valores do parâmetro. Observa-se, que a medida em que os valores de θ se afastam do valor hipotético, o poder aumenta, pois quanto mais distante o parâmetro estiver do valor hipotético mais fácil é detectar essa diferença.

Um bom teste é aquele que não rejeita H_0 com alta probabilidade quando realmente ela é verdadeira, e que rejeita a hipótese nula com grande probabilidade quando de fato ela é falsa. Para qualificar um bom teste é também necessário avaliar a sua robustez.

Quando se aplica um teste estatístico, normalmente algumas condições

devem ser satisfeitas. Um exemplo, muito comum, é o teste F, o qual tem as seguintes exigências (BANZATTO; KRONKA, 2006):

- a) Os efeitos dos fatores que se encontram no modelo estatístico de um dado experimento devem ser aditivos;
- b) Os erros experimentais devem ser aleatórios, independentes e normalmente distribuídos com média zero e variância comum.

Um procedimento deve ser considerado robusto, se ao violar uma das pressuposições estabelecidas, este procedimento, pelo menos de forma aproximada, mantenha as características delineadas originalmente na elaboração de sua teoria.

Um exemplo para a avaliação da robustez seria aplicar um teste baseado na distribuição normal (exige que os dados sigam esta distribuição), em dados com distribuição t de Student, por exemplo. Deve-se repetir esse processo várias vezes e avaliar se o comportamento é similar ou não quando os dados são normais. Quanto mais o desempenho se aproximar dos resultados obtidos quando as pressuposições são atendidas, mais robusto será o teste.

2.6 Distribuição normal multivariada

A distribuição normal é a distribuição de uma variável aleatória contínua mais importante no âmbito univariado. Na estatística multivariada não poderia ser diferente, sendo ela muito utilizada em pesquisas científicas.

Muitas técnicas empregadas na estatística multivariada são baseadas nessa distribuição (RENCHEER, 2002). Inferências são realizadas, assumindo normalidade dos resíduos dos modelos adotados, como, por exemplo a MANOVA.

Um vetor de dados p -dimensional é dito ter distribuição normal multivariada $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, se sua função densidade de probabi-

lidade é dada por:

$$f_x(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

sendo $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]^\top$ o vetor de médias e $\Sigma_{(p \times p)}$ a matriz de covariâncias positiva definida (ANDERSON, 1971), dada por:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Em muitos trabalhos científicos, o pesquisador se depara com a necessidade de simular dados de alguma distribuição, geralmente com o intuito de validar modelos, comparar testes de hipóteses ou métodos de estimação. Para gerar realizações normais p -dimensionais, $N_p(\boldsymbol{\mu}, \Sigma)$, usa-se transformações lineares de variáveis aleatórias normais padrão univariadas e independentes. Uma variável aleatória normal unidimensional, $N(\mu = 0, \Sigma = 1)$, pode ser gerada utilizando o teorema fundamental da transformação de probabilidade (MAGALHÃES, 2006), em que a partir de variáveis uniformes pode-se obter variáveis de qualquer outra distribuição, basta inverter a função de distribuição da variável que se deseja gerar.

Como exemplo, suponha que se deseja gerar uma realização de uma variável aleatória normal-padrão. Primeiramente deve-se gerar uma realização u de uma variável aleatória uniforme $U(0, 1)$, em seguida obter z de forma que:

$$u = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2}{2} \right\} dy,$$

ou seja, $z = \Phi^{-1}(u)$. Deve-se repetir esse processo até se obter p elementos do vetor \mathbf{Z} .

A partir disso, para gerar uma realização de uma variável aleatória com distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, deve-se aplicar a seguinte transformação linear no vetor \mathbf{Z} , $\mathbf{X} = \boldsymbol{\Sigma}^{1/2} \mathbf{Z} + \boldsymbol{\mu}$, em que $\boldsymbol{\Sigma}^{1/2}$ é a matriz raiz quadrada de $\boldsymbol{\Sigma}$. Segundo Ferreira (2011) \mathbf{X} possui distribuição normal multivariada com média $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Esse processo deve ser repetido até se obter o tamanho de amostra n desejado.

3 MÉTODOS

O problema é realizar o teste da hipótese particular

$$H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0} \quad (7)$$

em que $\boldsymbol{\mu}_i$ é o vetor de médias p -dimensional da i -ésima população, considerada normal multivariada com covariância $\boldsymbol{\Sigma}_i$, $i = 1, 2$ e $\mathbf{0}$ é o vetor nulo p -dimensional.

Para isso, da i -ésima população normal $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2$, é gerada uma amostra aleatória de tamanho n_i dada por $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$, considerando que $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. Da i -ésima amostra obtém-se o estimador de $\boldsymbol{\Sigma}_i$ por

$$\mathbf{S}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)^\top \quad (8)$$

em que $\bar{\mathbf{Y}}_i$ é a média amostral referente a i -ésima população.

Os três testes mencionados anteriormente são aplicados, utilizando-se os dados amostrais de ambas as populações. No presente estudo duas situações para avaliação do desempenho foram consideradas. Na primeira, amostras normais de populações com médias iguais ($\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$) foram geradas via simulação Monte Carlo. Nesse caso, a proporção de rejeições da hipótese nula verdadeira para cada teste, refere-se a uma estimativa do erro tipo I. Numa segunda etapa, populações com médias diferentes foram amostradas, permitindo-se avaliar o poder dos testes pelas proporções de vezes, em um número grande de simulações, em que a hipótese nula (7) foi rejeitada. Em ambos os cenários, diferentes tamanhos de amostras, número de variáveis, grau de heterogeneidade das covariâncias populacionais, magnitudes de diferenças entre os vetores de médias e níveis nominais de significância foram considerados.

3.1 Testes

Nesta subseção são apresentados os dois testes da literatura e o desenvolvimento do teste proposto.

3.1.1 Teste de Nel e Merwe Modificado - TNVMM

A partir das duas amostras originais, foram determinados os vetores de médias \bar{Y}_1 e \bar{Y}_2 e as matrizes de covariâncias S_1 e S_2 (8). Assim, a estatística do teste é computada por:

$$T^{*2} = (\bar{Y}_1 - \bar{Y}_2)^\top S_e^{-1} (\bar{Y}_1 - \bar{Y}_2) \quad (9)$$

em que S_e é dado em (5) e S_e^{-1} é a inversa de S_e .

A rejeição da hipótese nula (7) deve ocorrer para um nível nominal de significância α se

$$T^{*2} \geq \frac{\nu p}{\nu + 1 - p} F_{\alpha, p, \nu + 1 - p}, \quad (10)$$

em que $F_{\alpha, p, \nu + 1 - p}$ é o quantil $100\alpha\%$ superior da distribuição F com graus de liberdade p e $\nu + 1 - p$, sendo os graus de liberdade ajustados ν , dados em (6).

A rotina com os comandos do software R DEVELOPMENT CORE TEAM (2013) para a aplicação do teste de Nel e Merwe modificado está apresentada no Anexo Q.

3.1.2 Teste de Krishnamoorthy e Lu - TBPCL

Para aplicação desse teste, a partir dos dados das amostras originais, inicialmente deve ser computado o valor da estatística T^{*2} , dado na expressão (8). A distribuição nula dessa estatística é obtida por *bootstrap* paramétrico. A imposição

da hipótese nula é feita considerando-se que ambas as populações construídas para a realização das reamostragens *bootstrap*, possuem vetor de médias nulo. Formalmente considerou-se que $\bar{\mathbf{Y}}_i^* \sim N_p(\mathbf{0}, \mathbf{S}_i/n_i)$, $i = 1, 2$, sob a hipótese nula, em que $\bar{\mathbf{Y}}_i^*$ é a média da reamostragem *bootstrap* referente a i -ésima população.

Da mesma forma, da teoria normal, sabe-se que $(n_i - 1)\mathbf{S}_i^* \sim W_p(n_i - 1, \mathbf{S}_i)$, ou seja, $(n_i - 1)\mathbf{S}_i^*$ possui distribuição Wishart com dimensão p , graus de liberdade $n_i - 1$ e parâmetro de covariação \mathbf{S}_i , da i -ésima amostra original. Nesse caso, $(n_i - 1)\mathbf{S}_i^*$ é a matriz de soma de quadrados e produtos da reamostragem *bootstrap* referente a i -ésima população.

Dessa forma, Krishnamoorthy e Lu (2010) propuseram que fosse computada a distribuição nula de T^{*2} por $T_1^{*2}, T_2^{*2}, \dots, T_b^{*2}, \dots, T_B^{*2}$, em que $b = 1, 2, \dots, B$, com B sendo considerado o maior possível, e

$$T_b^{*2} = (\bar{\mathbf{Y}}_{1b}^* - \bar{\mathbf{Y}}_{2b}^*)^\top \left(\frac{\mathbf{S}_{1b}^*}{n_1} + \frac{\mathbf{S}_{2b}^*}{n_2} \right)^{-1} (\bar{\mathbf{Y}}_{1b}^* - \bar{\mathbf{Y}}_{2b}^*) \quad (11)$$

sendo $\bar{\mathbf{Y}}_{ib}^*$ uma variável aleatória da distribuição nula da $\bar{\mathbf{Y}}_i^*$ que é $N_p(\mathbf{0}, \mathbf{S}_i/n_i)$ e $(n_i - 1)\mathbf{S}_{ib}^*$ uma variável aleatória $W_p(n_i - 1, \mathbf{S}_i)$ referente a b -ésima reamostragem *bootstrap*, com $i = 1, 2$.

Considerando a distribuição nula de T_b^{*2} , aumentada pelo valor original T^{*2} , aqui denotado por T_{B+1}^{*2} , o valor-p é computado por

$$\text{Valor-p} = \frac{\sum_{b=1}^{B+1} I(T_b^{*2} \geq T^{*2})}{B + 1}$$

em que $I(T_b^{*2} \geq T^{*2})$ é a função indicadora que retorna 1, se $T_b^{*2} \geq T^{*2}$ e 0, caso contrário.

A rejeição da hipótese nula (7) deve ocorrer se o valor-p $\leq \alpha$, sendo α o

nível nominal de significância adotado.

A rotina com os comandos do software R DEVELOPMENT CORE TEAM (2013) para a aplicação do teste de Krishnamoorthy e Lu está apresentada no Anexo R.

3.1.3 Teste *bootstrap* paramétrico proposto - TBP

Considerando $\bar{\mathbf{Y}}_i$ e \mathbf{S}_i , $i = 1, 2$, os vetores de médias e as matrizes de covariâncias das amostras originais, a estatística do teste proposto é computada da seguinte forma:

$$T_c^2 = (\bar{\mathbf{Y}}_{1\cdot} - \bar{\mathbf{Y}}_{2\cdot})^\top (\bar{\mathbf{Y}}_{1\cdot} - \bar{\mathbf{Y}}_{2\cdot}). \quad (12)$$

A distribuição nula de T_c^2 foi construída a partir do método *bootstrap* paramétrico. Para isso, considerou-se que a média amostral de *bootstrap*, $\bar{\mathbf{Y}}_i^*$, da i -ésima população tem distribuição dada por $\bar{\mathbf{Y}}_i^* \sim N_p(\mathbf{0}, \mathbf{S}_i/n_i)$, em que \mathbf{S}_i é a covariância amostral original da i -ésima população e é utilizada como parâmetro da população a ser reamostrada, $i = 1, 2$.

Os seguintes passos devem ser executados para a aplicação desse teste:

1. Gerar uma realização da variável aleatória $\bar{\mathbf{Y}}_{ib}^*$ da distribuição $N_p(\mathbf{0}, \mathbf{S}_i/n_i)$, para $i = 1, 2$, em que b refere-se a b -ésima reamostragem *bootstrap*;
2. Calcular o valor da estatística na b -ésima reamostragem *bootstrap* por

$$T_b^2 = (\bar{\mathbf{Y}}_{1b}^* - \bar{\mathbf{Y}}_{2b}^*)^\top (\bar{\mathbf{Y}}_{1b}^* - \bar{\mathbf{Y}}_{2b}^*); \quad (13)$$

3. Repetir os passos (1) e (2) B vezes e obter a distribuição nula $T_1^2, T_2^2, \dots, T_B^2$. Agrupar o valor original T_c^2 , obtido em (12), aos valores T_b^2 's e denotá-

lo por T_{B+1}^2 . Para este trabalho foi utilizado $B = 2000$;

4. Computar o valor-p por:

$$\text{Valor-p} = \frac{\sum_{b=1}^{B+1} I(T_b^2 \geq T_c^2)}{B + 1}$$

em que $I(T_b^2 \geq T_c^2)$ é a função indicadora que retorna 1, se $T_b^2 \geq T_c^2$ e 0, caso contrário.

5. A rejeição da hipótese nula (7) deve ocorrer se o valor-p $\leq \alpha$, sendo α o nível nominal de significância adotado.

A rotina com os comandos do software R DEVELOPMENT CORE TEAM (2013) para a aplicação do teste de *bootstrap* paramétrico está apresentada no Anexo S.

3.2 Validação Monte Carlo

A validação do teste de *bootstrap* paramétrico proposto e dos demais foi realizada em duas etapas mediante uso de simulação Monte Carlo. Na primeira, simulações Monte Carlo foram realizadas sob a hipótese de nulidade (7). Portanto, a proporção de rejeições da hipótese nula corresponde a estimativa das taxas de erro tipo I. Na segunda etapa, as simulações foram realizadas sob H_1 e a proporção de rejeições refere-se ao poder.

No caso da avaliação do erro tipo I, essas taxas empíricas foram comparadas com os valores nominais de significância por meio do teste binomial exato.

3.2.1 Erro Tipo I

Para a obtenção das taxas de erro tipo I, amostras aleatórias de duas populações normais multivariadas foram simuladas sob H_0 . Assim, se $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ são as médias dessas populações, sem perda de generalidade, considerou-se $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$, em que $\mathbf{0}$ é um vetor p -dimensional de zeros. Assim, gerou-se da i -ésima população normal p -variada $N_p(\mathbf{0}, \boldsymbol{\Sigma}_i)$ a amostra: $\mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \dots, \mathbf{Y}_{ij}, \dots, \mathbf{Y}_{in_i}$, para $i = 1, 2$, em que $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ são as matrizes de covariâncias populacionais, $p \times p$, positivas definidas. Dessa forma, o problema de Behrens-Fisher ficou caracterizado. As amostras de cada população foram simuladas por meio da função `mvrnorm()` da biblioteca *MASS* do programa R DEVELOPMENT CORE TEAM (2013).

Para a especificação de $\boldsymbol{\Sigma}_i$, optou-se pela construção dessas matrizes em dois formatos: estrutura de simetria composta (equicorrelacionada) e matrizes sem nenhuma estrutura.

Para o caso equicorrelacionado, determinou-se que $\boldsymbol{\Sigma}_1$ teria a seguinte estrutura

$$\boldsymbol{\Sigma}_1 = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix},$$

em que $0 \leq \rho \leq 1$ e $\sigma^2 \geq 0$. Sem perda de generalidade, ρ foi fixado em 0,5 e σ^2 em 1.

A matrix $\boldsymbol{\Sigma}_2$ foi estipulada como um múltiplo de $\boldsymbol{\Sigma}_1$, de tal forma que a razão das variâncias generalizadas de ambas as populações fosse igual a uma constante positiva $\delta \geq 1$, que refletiria a heterogeneidade das covariâncias. Para $\delta \geq 1$ fixado, $\boldsymbol{\Sigma}_2$ foi especificada por

$$\Sigma_2 = \delta_p \Sigma_1,$$

em que $\delta_p = \sqrt[p]{\delta}$, com $\delta = |\Sigma_2|/|\Sigma_1|$. Os valores de δ foram fixados em 1, 2, 8, 16 e 32. O valor 1, foi usado como referência, pois indica que as matrizes são homogêneas. Os demais valores indicam graus medianos e elevados de heterogeneidade das covariâncias.

Para a especificação das matrizes não estruturadas, Σ_1 foi determinada da seguinte forma:

$$\Sigma_1 = TT^\top$$

sendo T uma matriz triangular inferior, gerada a partir de variáveis aleatórias independentes e com distribuição Gama.

A matriz Σ_2 foi estipulada da mesma forma descrita anteriormente, no caso equicorrelacionado. Nessa situação utilizou-se apenas $\delta = 32$.

Obtido um par de amostras, de ambas as populações, para alguma configuração de seus parâmetros, conforme descrição anterior, o *TBP*, o *TNVMM* e o *TBPKL* foram aplicados. O valor-p de cada teste foi confrontado com o nível nominal de significância α , em que nesse trabalho foram usados os valores de 0,10; 0,05 e 0,01. Se o valor-p for menor igual a α , a hipótese nula deve ser rejeitada.

O processo de geração das amostras, aplicação dos testes e comparação dos valores-p com os níveis nominais de significância foi repetido N vezes, sendo N fixado em 2000. As probabilidades empíricas do erro tipo I ($\hat{\alpha}_k$) do k -ésimo teste ($k = 1, 2, 3$), sendo $k = 1$ para *TBP*, $k = 2$ para *TNVMM* e $k = 3$ para *TBPKL*, foram estimadas por

$$\hat{\alpha}_k = \frac{\sum_{j=1}^N I(\text{valor-p}_{kj} \leq \alpha)}{N}, \quad (14)$$

sendo $\alpha = 0,10; 0,05$ ou $0,01$, $I(\text{valor-p}_{kj} \leq \alpha)$ uma função indicadora e valor-p_{kj} , o valor-p do k -ésimo teste na j -ésima simulação Monte Carlo, $j = 1, 2, \dots, N$.

Também foram considerados diferentes tamanhos para as amostras de ambas populações. Escolheu-se os tamanhos $n_i = 8, 30$ e 100 , para $i = 1$ e 2 , sendo consideradas as nove configurações possíveis dadas pelo produto cartesiano $|n_1| \times |n_2|$, em que $|n_i|$ é a cardinalidade de n_i . O número de variáveis p , dependeu dos tamanhos amostrais. Quando uma das amostras era de tamanho 8, os valores para p foram 2 e 7. Quando o tamanho mínimo das amostras foi 30, adotaram-se valores de p iguais a 7 e 29. Finalmente, no caso $n_1 = 100$ e $n_2 = 100$, o número de variáveis utilizado foi 7 e 15. Valores de p iguais a $n_i - 1$ foram escolhidos em virtude de Ramos e Ferreira (2012) observarem que o desempenho de alguns testes decaiu abruptamente nessas condições. As rotinas com os comandos do software R DEVELOPMENT CORE TEAM (2013) para a avaliação das taxas de erro tipo I, dos três testes em estudo, estão apresentadas no Anexo T.

As taxas de erro tipo I foram estimadas utilizando simulações Monte Carlo e por isso, não estão livres de erro. Assim, um teste binomial considerando o nível nominal de significância de 0,01, foi aplicado nas taxas estimadas para testar as seguintes hipóteses: $H_0 : \alpha = 0,10$ vs $H_1 : \alpha \neq 0,10$; $H_0 : \alpha = 0,05$ vs $H_1 : \alpha \neq 0,05$ e $H_0 : \alpha = 0,01$ vs $H_1 : \alpha \neq 0,01$. Se H_0 for rejeitada e a taxa de erro tipo I for considerada significativamente inferior ao α correspondente, o teste em questão deve ser considerado conservativo; se a taxa de erro tipo I for considerada significativamente superior a α correspondente, o teste deverá ser

considerado liberal; e por fim, se a taxa não for significativamente diferente de α , ou seja, aceitar H_0 , o teste deverá ser considerado exato (OLIVEIRA; FERREIRA, 2010).

Se y representa o número de rejeições de H_0 em $N = 2000$ simulações Monte Carlo para o nível nominal de significância, então pode-se obter a estatística do teste utilizando a relação exata existente entre a distribuição F e a binomial.

A estatística do teste é

$$F_c = \left(\frac{y - 1}{n - y} \right) \left(\frac{1 - \alpha}{\alpha} \right),$$

que sob a hipótese nula segue uma distribuição F com $\nu_1 = 2(N - y)$ e $\nu_2 = 2(y + 1)$ graus de liberdade. Se $F_c \leq F_{0,005}$ ou $F_c \geq F_{0,995}$, a hipótese nula deve ser rejeitada com 0,01 de significância, onde $F_{0,005}$ e $F_{0,995}$ são quantis da distribuição F com ν_1 e ν_2 graus de liberdade (OLIVEIRA; FERREIRA, 2010).

3.2.2 Poder

Para avaliar o poder dos três testes, os mesmos procedimentos usados na seção 3.2.1 foram adotados, exceto pelo fato das simulações Monte Carlo terem sido realizadas sob H_1 . Nesse caso, $\mu_1 \neq \mu_2$. Assim, amostras de tamanho n_1 foram geradas de um $N_p(\mu_1, \Sigma_1)$ e amostras de tamanho n_2 de uma $N_p(\mu_2, \Sigma_2)$.

Sem perda de generalidade, μ_1 foi fixada em $\mathbf{0}$. A média da população 2 foi estipulada por $\mu_2 = [\mu_{21}, \mu_{22}, \dots, \mu_{2p}]^\top$, sendo $\mu_{2i} = k\sigma_{\bar{Y}_{1i} - \bar{Y}_{2i}}$, em que

$$\sigma_{\bar{Y}_{1i} - \bar{Y}_{2i}} = \sqrt{\frac{\sigma_{ii}^{(1)}}{n_1} + \frac{\sigma_{ii}^{(2)}}{n_2}}$$

é o erro padrão da diferença entre as médias \bar{Y}_{1i} e \bar{Y}_{2i} , para cada variável, $\sigma_{ii}^{(1)}$ e $\sigma_{ii}^{(2)}$ as variâncias da i -ésima variável das populações 1 e 2 e k um número real. Os

valores utilizados para k foram 1, 2, 4 e 8, para o caso de matrizes estruturadas, e $k = 1$ e 2, para o caso de matrizes não estruturadas. Esses valores de k refletem a diferença entre as médias em termos de erros padrões, ou seja, μ_2 foi definido para estabelecer uma diferença de k erros padrões entre as médias das populações para cada variável.

As proporções de rejeições da hipótese nula foram obtidas com fórmula semelhante àquela apresentada na expressão (14). Os valores de poder empíricos (\hat{P}_k) são dados por:

$$\hat{P}_k = \frac{\sum_{j=1}^N I(\text{valor-p}_{kj} \leq \alpha)}{N},$$

em que $I(\text{valor-p}_{kj} \leq \alpha)$, valor-p_{kj} , α e N , são os mesmos descritos em (14). As rotinas com os comandos do software R DEVELOPMENT CORE TEAM (2013) para a avaliação do poder, para os três testes em estudo, estão apresentadas no Anexo U.

4 RESULTADOS E DISCUSSÃO

Nas próximas subseções serão apresentados os resultados das taxas de erro tipo I e poder, para os três testes em estudo, nas duas situações de matrizes de covariâncias.

4.1 Caso de matrizes de covariâncias equicorrelacionadas

4.1.1 Taxas de erro tipo I

De início foram estudadas situações com amostras pequenas e de mesmo tamanho. Na Figura 2 apresenta-se taxas de erro tipo I, dos três testes, utilizando tamanhos amostrais, $n_1 = n_2 = 8$, número de variáveis, $p = 2$ e 7 , nível de significância, $\alpha = 0,10; 0,05$ e $0,01$ e grau de heterogeneidade de covariâncias, $\delta = 1, 2, 8, 16$ e 32 .

Nessa figura pode-se observar que quando $p = 2$ (Figura 2a, 2c e 2e) o teste *bootstrap* paramétrico proposto (*TBP*) apresenta-se com comportamento liberal, exceto quando $\alpha = 0,05$, $\delta = 2$ e quando $\alpha = 0,10$, para $\delta = 2, 8$ e 16 . Já o teste *bootstrap* paramétrico de Krishnamoorthy e Lu, (*TBPKL*), e o teste de Nel e Merwe modificado, (*TNVMM*), mantêm suas taxas dentro do intervalo de linhas pontilhadas, indicando que as mesmas não são consideradas diferentes do nível de significância (α) correspondente, ou seja, os dois testes podem ser considerados exatos, nessa situação. A situação com tamanho amostral $n_1 = n_2 = 8$, $p = 2$ e $\alpha = 0,10; 0,05$ e $0,01$ (Figura 2a), também foi estudada por Ramos e Ferreira (2012). Os resultados encontrados pelos autores foram idênticos para o *TNVMM* e o teste bayesiano proposto por eles também apresentou controle adequado das taxas de erro tipo I, com exceção das situações em que $\delta = 2$ para $\alpha = 0,05$ e $0,01$, em que o teste se mostrou conservativo.

Considerando o valor de $p = 7$, o *TBP* deixa de ser liberal quando $\alpha = 0,10$ (Figura 2b), passando a controlar adequadamente o erro tipo I, já o *TNVMM* tem comportamento conservativo, exceto quando $\delta = 32$. O *TBPKL*, nessa mesma situação, é conservativo apenas para $\delta = 8$, nos demais valores de heterogeneidade de covariâncias, a taxa é controlada adequadamente.

Ainda na Figura 2, mas observando $\alpha = 0,05$ (Figura 2d), nota-se que o *TBP* é levemente liberal para a menor heterogeneidade das matrizes de covariâncias, e a medida que δ aumenta o teste passa a controlar melhor o erro tipo I. Já o *TBPKL* apresenta taxas bem próximas do limite de rejeição, sendo consideradas iguais ao nível de significância apenas para $\delta = 2$ e 32. O *TNVMM* pode ser considerado conservativo, pois não apresentou nenhum valor de taxa que pudesse ser considerado igual a 0,05. Quando $p = 7$ e $\alpha = 0,01$ (Figura 2f), o *TBP* volta a apresentar o comportamento liberal. Já o *TBPKL* e o *TNVMM* somente apresentaram taxas que podem ser consideradas menores que $\alpha = 0,01$ quando $\delta = 2$ e 32, respectivamente.

As situações descritas até o momento são consideradas de pequenas amostras, mas já é possível notar, de uma maneira geral, que quando o número de variáveis aumenta, o controle do erro tipo I do *TBP* melhora, enquanto que os *TBPKL* e *TNVMM* pioram.

Analisando os resultados da Figura 2, observando apenas $\delta = 1$, quando as covariâncias são homogêneas, o que pode-se observar é que o *TBP* é levemente liberal quando $p = 2$ e $\alpha = 0,10$ e 0,05, já para $p = 7$ essa característica aparece apenas para $\alpha = 0,01$. O *TBPKL* apresenta-se exato em todas as situação de $\delta = 1$. O *TNVMM* é exato apenas quando o número de variáveis é 2, quando o valor de p passa para 7 ele torna-se mais conservativo.

Nos anexos A e B, estão apresentadas as taxas de erro tipo I para amostras

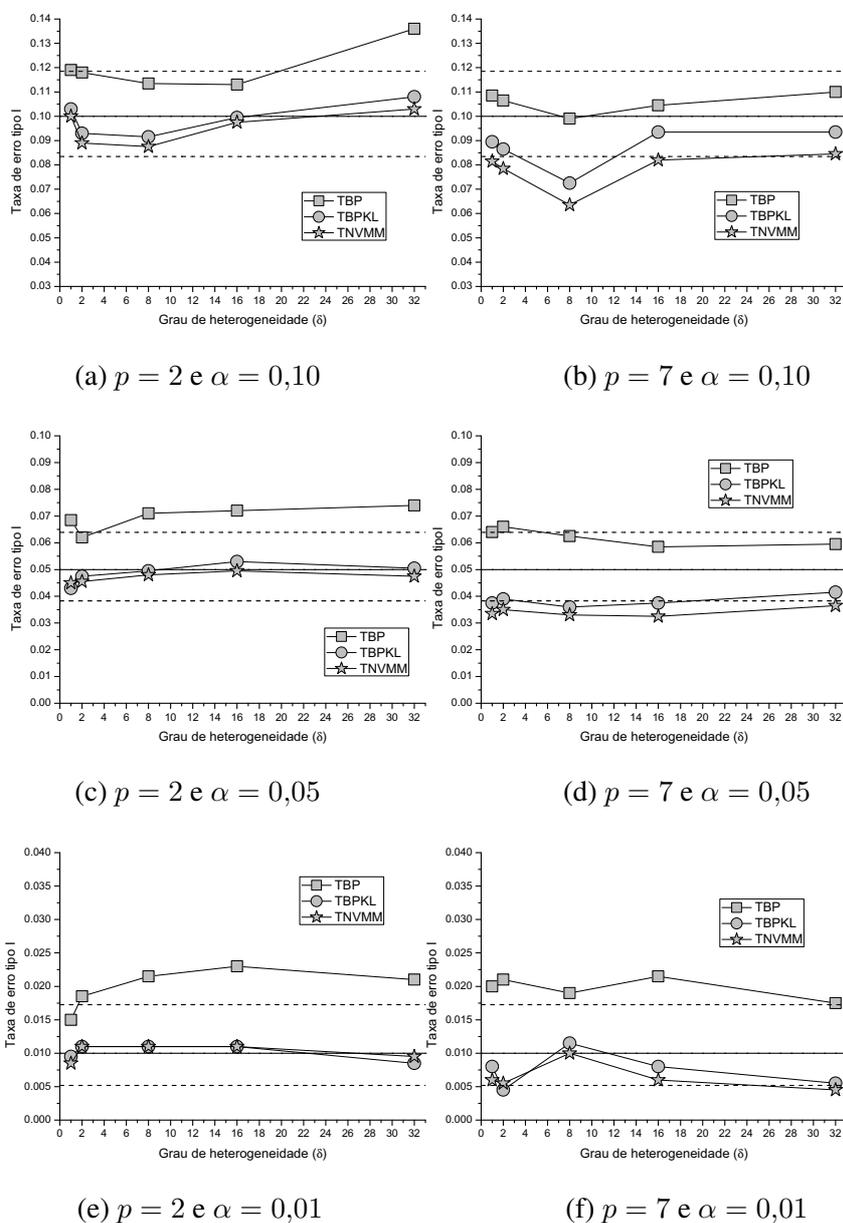


Figura 2 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99

maiores, mas ainda iguais, $n_1 = n_2 = 30$ e $n_1 = n_2 = 100$. Com o aumento do tamanho amostral os três testes em estudo apresentaram melhora no controle do erro tipo I, sendo que em praticamente todas as situações estudadas os valores das taxas podem ser considerados iguais aos valores de α correspondentes. O mesmo resultado foi encontrado nas configurações estudadas por Ramos e Ferreira (2012).

Observando a situação de homogeneidade de matrizes de covariâncias, para os tamanhos amostrais dos Anexos A e B, os três testes apresentaram-se exatos em todas as configurações. Para tamanhos amostrais iguais não foi detectado nenhum tipo de padrão de comportamento para o erro tipo I, em relação à variação do grau de heterogeneidade das matrizes de covariâncias.

Na Tabela 1, a seguir, apresenta-se a primeira situação estudada para tamanhos amostrais diferentes, $n_1 = 8$ e $n_2 = 30$. Pode-se observar que quando $p = 2$ os testes *TBPKL* e *TNVMM* controlam adequadamente o erro tipo I, com exceção da configuração $\delta = 32$ e $\alpha = 0,05$, em que o *TNVMM* mostrou-se conservativo, já o *TBP* apresenta alguns valores das taxas de erro tipo I que podem ser considerados maiores que o α correspondente, principalmente quando $\alpha = 0,05$. No entanto, quando o número de variáveis aumenta para $p = 7$ os resultados se alteram. Quando $\alpha = 0,10$, o *TBPKL* e o *TNVMM* tornam-se liberais, enquanto o *TBP* controla adequadamente o erro tipo I, exceto quando as variâncias são homogêneas ($\delta = 1$), em que apresentou-se levemente liberal. Quando $\alpha = 0,05$, o comportamento liberal do *TBPKL* e do *TNVMM* se mantém, e o *TBP* apresenta caráter liberal apenas quando $\delta = 1$ e 32 , mas com valores menores que os outros dois testes em estudo. Avaliando a situação para $p = 7$ e $\alpha = 0,01$, nessa mesma tabela, observa-se que o *TBPKL* e o *TNVMM* apresentam melhora nas taxas de erro tipo I, enquanto o *TBP* apresentou piora para valores mais altos de δ . De uma maneira geral, esses resultados indicam que

quando o número de variáveis é mais alto, o *TBPKL* e o *TNVMM* apresentam caráter liberal e o *TBP* controla melhor o erro tipo I, mas quando $\alpha = 0,01$, a situação se inverte.

Em relação ao caso em que há homogeneidade das covariâncias, para o tamanho amostral $n_1 = 8$ e $n_2 = 30$, o que pode-se notar é que para $p = 2$, o *TBP* apresenta caráter levemente liberal apenas quando $\alpha = 0,05$, nos demais níveis mostrou-se exato. O *TBPKL* e o *TNVMM* foram considerados exatos nesta situação. Alterando o valor de p para 7, os três testes apresentaram-se liberais, sendo o *TBPKL* com as maiores taxas, em segundo lugar o *TNVMM* e o *TBP* com menores valores para as taxas de erro tipo I. De maneira geral, até mesmo quando as covariâncias são homogêneas e $p = 7$, o *TBP* mostrou-se mais adequado.

Na Figura 3, os tamanhos amostrais são os mesmos apresentados anteriormente, embora invertidos, sendo $n_1 = 30$ e $n_2 = 8$. Esta situação é considerada a mais crítica porque a maior covariância está relacionada ao menor tamanho amostral. O comportamento dos três testes é semelhante ao descrito anteriormente ($n_1 = 8$ e $n_2 = 30$), apenas ficando mais evidente a superioridade do *TBP* para a situação mais crítica, que se caracteriza pela maior covariância relacionada a amostra de menor tamanho e para $p = 7$. Ramos e Ferreira (2012) avaliaram uma situação semelhante, mas usando $p = 5$, e já perceberam esse comportamento liberal do *TNVMM*. Nessas situações, o teste bayesiano, proposto por eles, apresentou-se conservativo.

Quando $p = 2$ (Figura 3a, 3c e 3e), nota-se que o *TBPKL* e o *TNVMM* apresentam taxas de erro tipo I que podem ser consideradas iguais ao α correspondente, em praticamente todas as situações, já o *TBP* apresenta caráter liberal. Para valores de $p = 7$ (Figura 3b, 3d e 3f), o *TBP* apresenta grande melhora no

Tabela 2 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0

		$n_1 = 8$ e $n_2 = 30$			
		δ	TBP	TBPKL	TNVMM
$\alpha = 0,10$	p=2	1	0,1175	0,1050	0,0990
		2	0,1250 ⁺	0,1160	0,1095
		8	0,1090	0,1040	0,1030
		16	0,1120	0,1160	0,1145
		32	0,1160	0,1030	0,0975
	p=7	1	0,1200 ⁺	0,1800 ⁺	0,1460 ⁺
		2	0,1035	0,1830 ⁺	0,1550 ⁺
		8	0,1030	0,1635 ⁺	0,1455 ⁺
		16	0,1080	0,1510 ⁺	0,1245 ⁺
		32	0,1135	0,1600 ⁺	0,1400 ⁺
$\alpha = 0,05$	p=2	1	0,0705 ⁺	0,0635	0,0560
		2	0,0665 ⁺	0,0605	0,0575
		8	0,0620	0,0490	0,0460
		16	0,0675 ⁺	0,0595	0,0570
		32	0,0445	0,0390	0,0380*
	p=7	1	0,0660 ⁺	0,0990 ⁺	0,0765 ⁺
		2	0,0630	0,0910 ⁺	0,0740 ⁺
		8	0,0635	0,0845 ⁺	0,0690 ⁺
		16	0,0565	0,0835 ⁺	0,0655 ⁺
		32	0,0715 ⁺	0,0905 ⁺	0,0785 ⁺
$\alpha = 0,01$	p=2	1	0,0170	0,0125	0,0095
		2	0,0125	0,0115	0,0100
		8	0,0135	0,0105	0,0095
		16	0,0125	0,0115	0,0110
		32	0,0150	0,008	0,0085
	p=7	1	0,0175 ⁺	0,0280 ⁺	0,0180 ⁺
		2	0,0140	0,0205 ⁺	0,0140
		8	0,0165	0,0115	0,0080
		16	0,0190 ⁺	0,0125	0,0065
		32	0,0220 ⁺	0,0140	0,0100

+ O valor pode ser considerado estatisticamente diferente e maior que o α correspondente.

* O valor pode ser considerado estatisticamente diferente e menor que o α correspondente.

Coefficiente de confiança de 0,99.

controle do erro tipo I, podendo ser considerado exato em praticamente todas as situações. Ainda para $p = 7$, mas observando apenas $\alpha = 0,10$ e $0,05$ o *TBPKL* e o *TNVMM* apresentam-se liberais, com taxas bem mais altas que o *TBP*. Para $\alpha = 0,01$, o *TBPKL* se mantém liberal, o *TNVMM* é liberal apenas quando $\delta = 2$ e 8 e o *TBP* controla adequadamente o erro tipo I.

Observando apenas $\delta = 1$, essa inversão dos tamanhos amostrais provocou, mais uma vez, melhora das taxas do *TBP* quando $p = 7$ e piora para o *TBPKL* e o *TNVMM*, sendo o *TBPKL* considerado o mais liberal dos três.

Aumentado a diferença entre o tamanho das amostras, $n_1 = 8$ e $n_2 = 100$ (Anexo C), $n_1 = 100$ e $n_2 = 8$ (Figura 4), o que se observa é um padrão de comportamento semelhante às situações apresentadas na Tabela 1 ($n_1 = 8, n_2 = 30$) e Figura 3 ($n_1 = 30, n_2 = 8$). Apesar de o padrão de comportamento ser o mesmo, esse aumento da diferença entre os tamanhos amostrais, causou grande aumento das taxas de erro tipo I para o *TBPKL* e o *TNVMM*, quando $p = 7$.

Analisando as Figuras 4a, 4c e 4e, observa-se que para $p = 2$ os testes *TBPKL* e *TNVMM* controlam muito bem o erro tipo I, já o *TBP* continua apresentando um leve comportamento liberal. Quando p passa a ser igual a 7 (Figura 4b, 4d e 4f), as taxas do *TBP*, para todos os valores de α , melhoram bastante, apenas em $\alpha = 0,01$ apresenta taxas que podem ser consideradas maiores que o α , mas ainda assim, são bem menores que as taxas dos demais testes em estudo. Mais uma vez o *TBP* mostrou-se superior aos demais testes, no controle do erro tipo I, nas situações consideradas mais críticas. O desempenho do *TBP* quando $\delta = 1$, nos tamanhos amostrais, $n_1 = 8, n_2 = 30$ e $n_1 = 30, n_2 = 8$ e $p = 2$ é o mais liberal dos três, enquanto o *TNVMM* é exato e o *TBPKL* é levemente liberal em algumas ocasiões. Quando $p = 7$, o *TBPKL* torna-se extremamente liberal, em segundo lugar ficou o *TNVMM* e o menos liberal foi

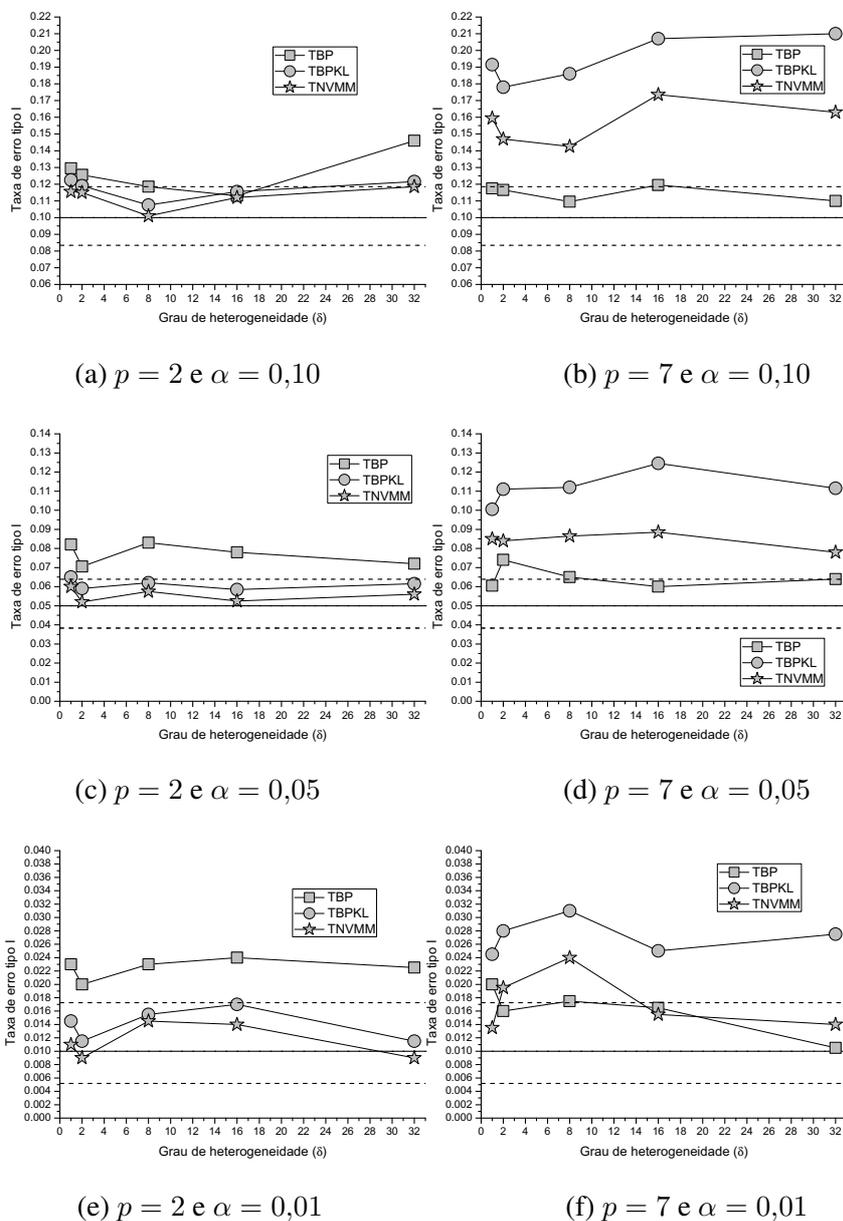


Figura 3 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$, $n_2 = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99

o *TBP*. Novamente a superioridade do *TBP* fica evidente quando $p = 7$ e $\delta = 1$.

As taxas de erro tipo I para tamanhos amostrais $n_1 = 30$, $n_2 = 100$ e $n_1 = 100$, $n_2 = 30$, estão apresentadas no Anexo D. Nesta situação os três testes continuaram apresentando resultados semelhantes aos observados na Figura 4 e no Anexo C. O *TBP* apresentou taxas de erro tipo I que podem ser consideradas iguais ao α em praticamente todas as situações, já o *TBPKL* e o *TNVMM* apresentaram taxas de erro tipo I adequadas apenas para $p = 7$. Quando $p = 29$ tais testes apresentaram altos valores de taxas de erro tipo I, sendo considerados liberais. Avaliando as taxas de erro tipo I quando $\delta = 1$ e $p = 7$, todos os testes mostraram-se exatos, mas quando $p = 29$ o *TBP* torna-se exato enquanto os demais tornam-se muito liberais.

Em nenhum dos casos estudados foi detectado algum padrão de comportamento das taxas de erro tipo I, em relação à variação da heterogeneidade das matrizes de covariâncias.

4.1.2 Poder

Nesta subsecção estão apresentados os resultados do desempenho dos três testes para amostras simuladas sob H_1 , ou seja, quando $\mu_1 \neq \mu_2$.

A princípio foi avaliado o poder em situações de amostras de mesmos tamanhos e pequenas. Na Figura 5 apresenta-se resultados para $n_1 = n_2 = 8$, $p = 2$ e 7 , $\alpha = 0,10$, $0,05$ e $0,01$, para diferença entre médias de 1 erro padrão ($k = 1$). Pode-se observar que nesta situação de diferença pequena entre médias, o poder dos três testes é baixo, pois a dificuldade de detectá-la é maior, além disso tratam-se de amostras consideradas pequenas. Também é possível dizer que a heterogeneidade de covariâncias não afeta o poder dos testes porque em todos os gráficos as curvas de poder são praticamente paralelas ao eixo horizontal do

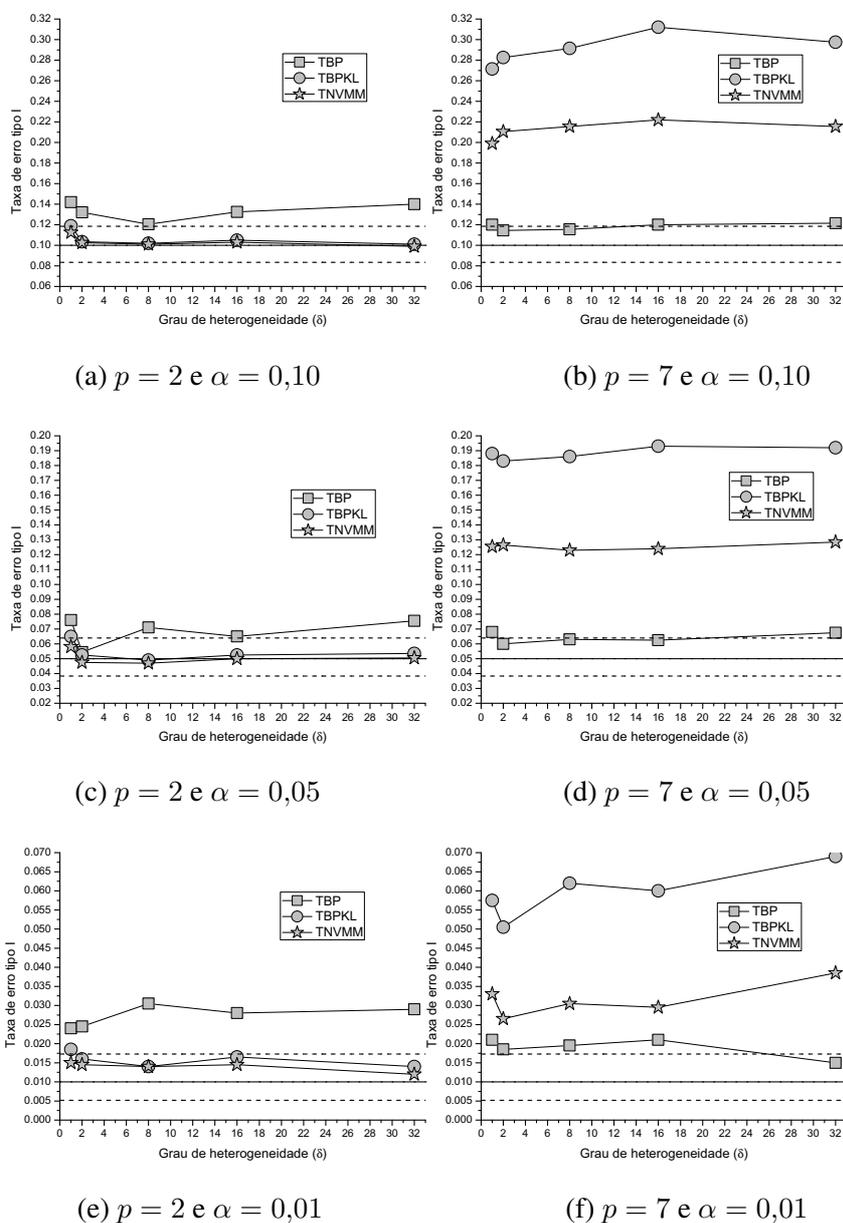


Figura 4 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99

gráfico.

É conveniente chamar atenção para a relação entre nível de significância e poder. Como já é esperado pela teoria, a medida que o α diminuiu o poder também diminuiu, esse fato apenas reforça a confiabilidade dos resultados das simulações realizadas, conforme pode-se verificar pelas comparações dos três testes na Figura 5.

Referindo-se de uma maneira geral, sobre a situação apresentada na Figura 5, o poder do *TBP* aumentou quando o número de variáveis passa de 2 para 7, já o poder do *TBPKL* e do *TNVMM*, que apresentaram valores muito próximos, diminuiu. Em todas as situações o poder do *TBP* foi superior as dos demais testes.

Na Figura 6 apresenta-se a mesma situação da Figura 5, exceto pela diferença entre médias que passa a ser $k = 2$. Esse aumento da diferença entre as médias causou, como já era esperado, um aumento expressivo de poder para todos os testes. Não houve grandes alterações no padrão de comportamento do poder, apenas ficou mais evidente a superioridade do *TBP* em relação aos demais testes, pois o teste aqui proposto teve um aumento com maior magnitude que os demais, comparando com a situação de $k = 1$. A alteração do número de variáveis de 2 para 7 continuou provocando um acréscimo de poder para o *TBP* e decréscimo para o *TBPKL* e o *TNVMM*.

Na Figura 7, apresentam-se os valores de poder quando $k = 4$, nas mesmas configurações de n_1 , n_2 , p e α , das figuras anteriores. Nesse caso, o poder do *TBP* é praticamente 1 em todas as situações, já o *TBPKL* e o *TNVMM* continuam apresentando valores abaixo do *TBP*. Vale ressaltar que o *TBPKL* e o *TNVMM*, para esse mesmo tamanho amostral, foram considerados exatos, em relação ao erro tipo I, na maioria das situações, já o *TBP* apresentou melhores taxas de erro tipo I quando $p = 7$ (ver Figura 2), situação em que seu poder é

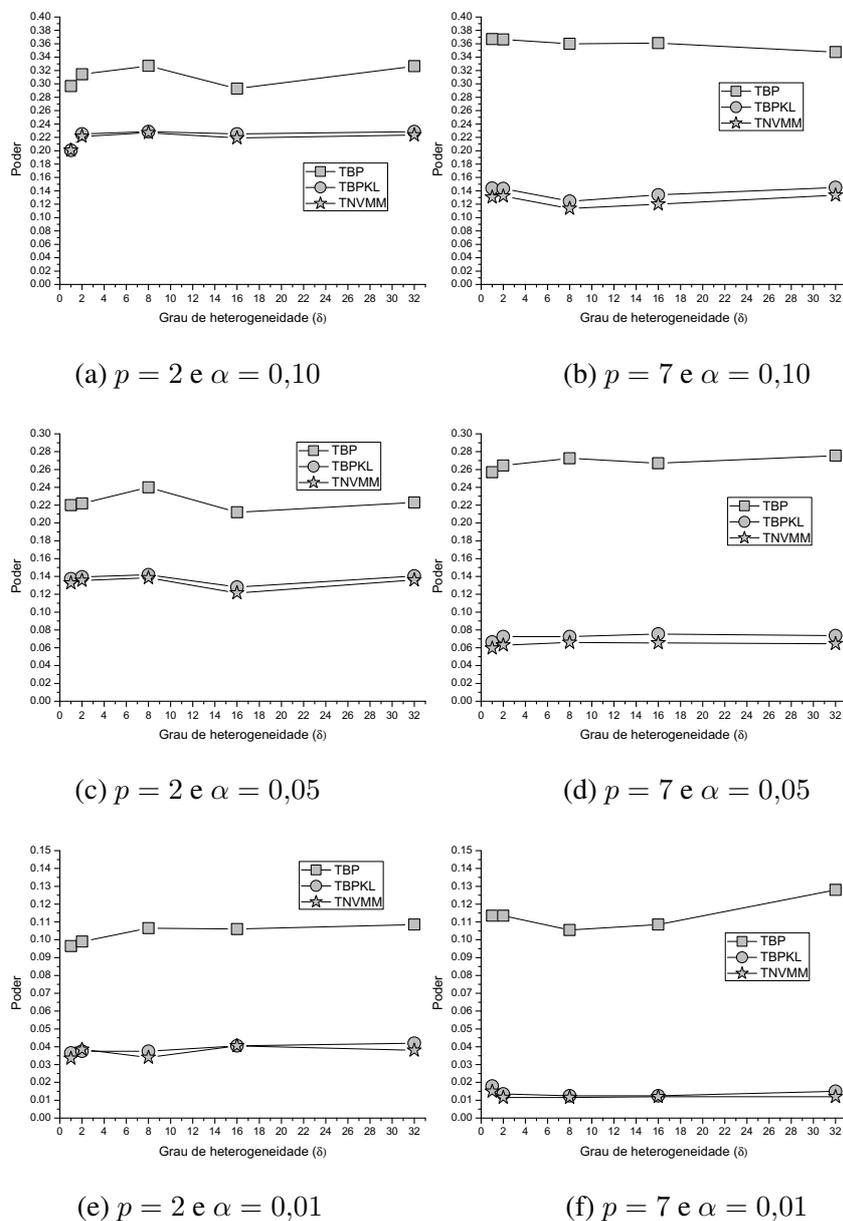


Figura 5 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

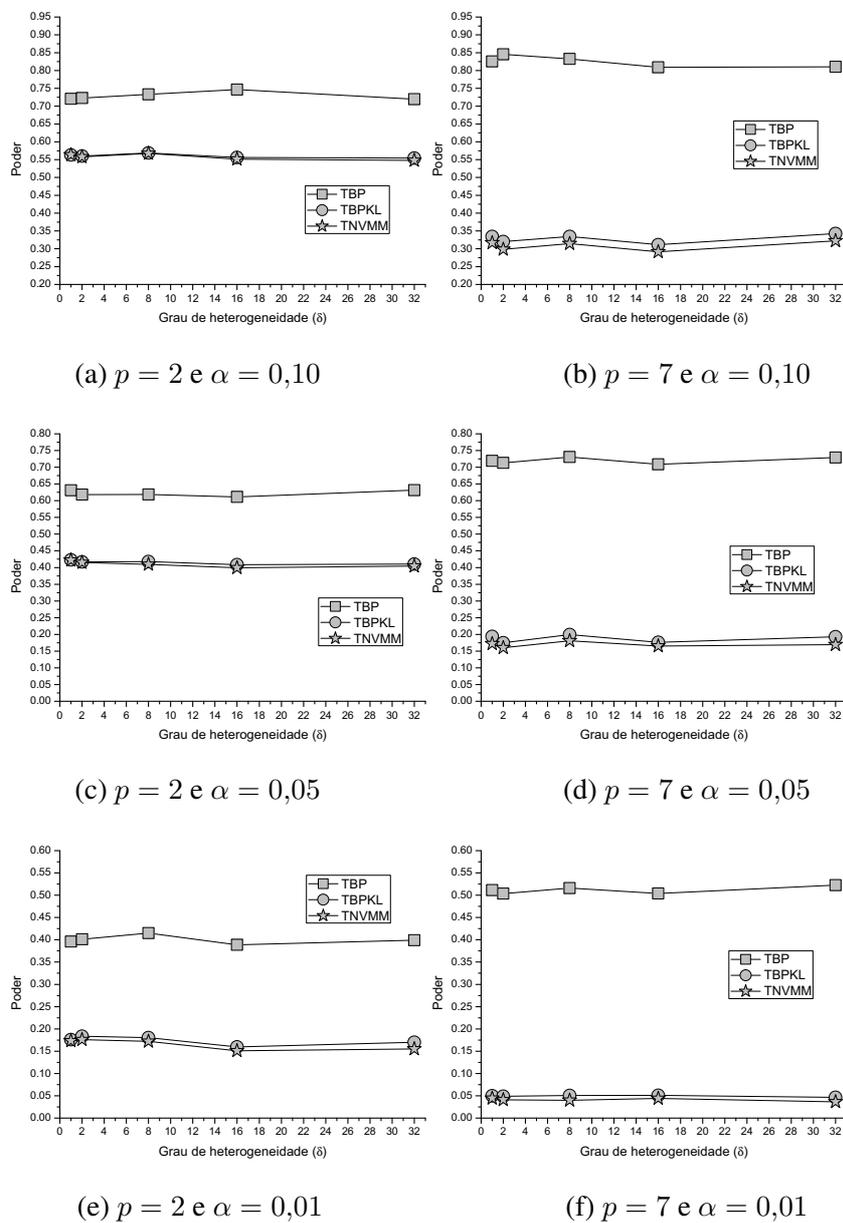


Figura 6 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

muito superior ao dos outros testes.

Avaliando o poder para $k = 8$, ainda com $n_1 = n_2 = 8$ (Anexo E), os três testes em estudo apresentaram poder de igual a 1 em praticamente todas as situações, exceto quando $\alpha = 0,01$ e $p = 7$ o *TBPKL* e o *TNVMM* obtiveram valores de poder um pouco mais baixos, em torno de 0,9758 e 0,9653 respectivamente.

O poder para amostras um pouco maiores, mas ainda iguais, $n_1 = n_2 = 30$, utilizando número de variáveis $p = 7$ e 29 e distância entre médias de $k = 1, 2, 4$ e 8 erros padrões, estão apresentados nos Anexos F e G. Nessas tabelas, observou-se o mesmo padrão de comportamento ocorrido para amostras pequenas (Figuras 5, 6, 7 e Anexo E). O *TBP* continuou apresentando valores superiores aos outros testes e a partir de $k = 2$ (Anexo F), já obteve valores expressivos de poder. Da mesma forma, o *TBPKL* e o *TNVMM* continuaram apresentando valores de poder menores que os de *TBP*.

Ramos e Ferreira (2012) estudaram para esse tamanho amostral ($n_1 = n_2 = 30$) a situação com $p = 2$ e $k = 2$, o poder do *TNVMM* foi um pouco maior que os encontrados neste trabalho. Nota-se que o aumento do número de variáveis de 2 para 7, provocou queda no poder do *TNVMM* e para $p = 29$, a queda foi ainda mais expressiva.

Quando $k = 4$ (Anexo G), o *TBP* apresentou valores superiores a 0,99 em todas as situações. Já os outros dois testes apresentaram melhora de poder, mas ainda continuaram com poder abaixo do *TBP*. Os testes *TBPKL* e *TNVMM* vêm sendo referidos conjuntamente, pois seus valores foram muito próximos, em relação ao poder.

Para $k = 8$, Anexo G, os três testes apresentaram grandes valores de poder, sendo superiores a 0,9960, em todas as situações. Relembrando que para esse

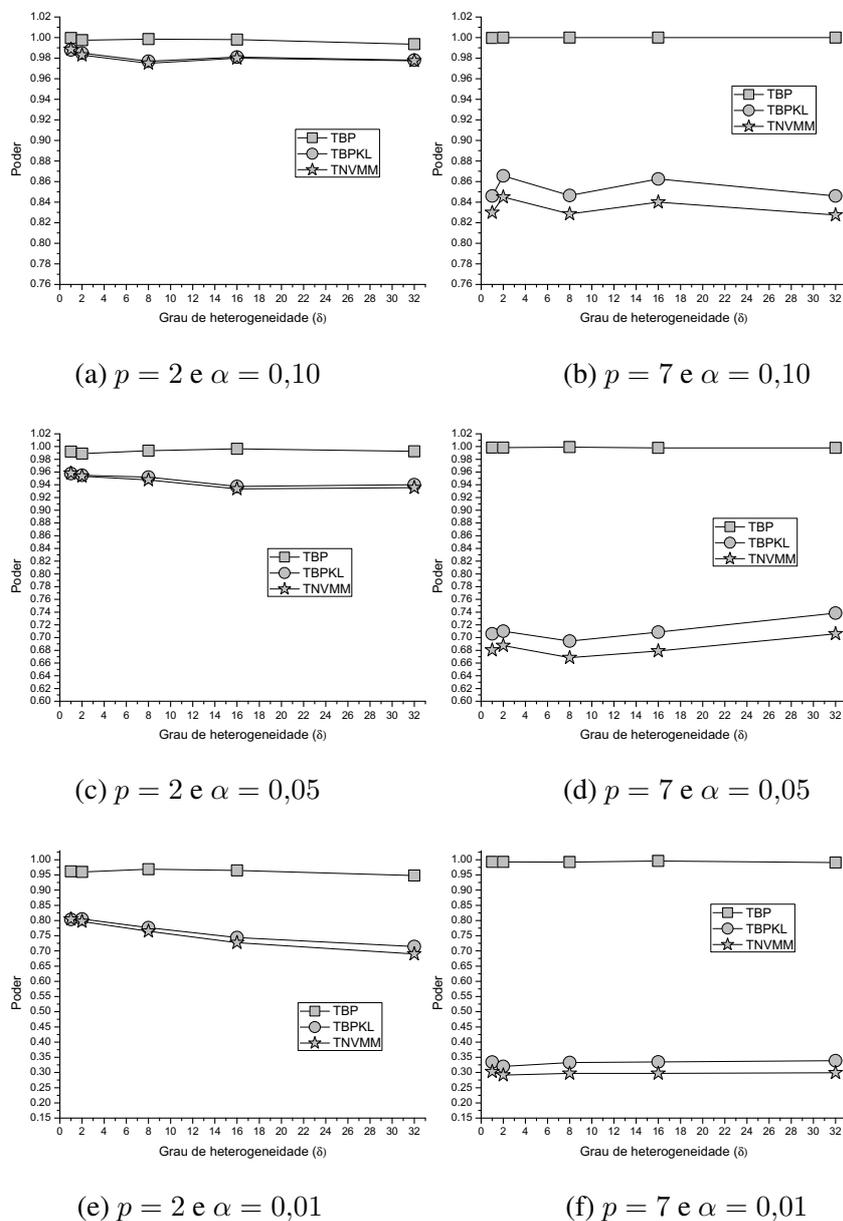


Figura 7 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 8$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

mesmo tamanho amostral, $n_1 = n_2 = 30$, todos os testes foram considerados exatos em relação ao erro tipo I, em praticamente todas as situações estudadas.

Os resultados para amostras grandes e iguais, $n_1 = n_2 = 100$, com $p = 7$ e 15, $k = 1, 2$ e 4, estão apresentados nos Anexos H e I. Observando e comparando estes resultados com os de amostras de tamanho $n_1 = n_2 = 30$, nota-se novamente que o aumento do número de variáveis causou queda no poder dos testes *TBPKL* e *TNVMM*, para todos os valores de α , mas com amostras maiores essa queda foi menos expressiva. Entretanto, isso pode estar relacionado ao fato de terem sido estudadas apenas as situações com $p = 7$ e 15 e não para número de variáveis maiores, $p = 99$, já que a situação é de amostras de tamanho 100. Esse fato implica em dizer, que os testes *TBPKL* e *TNVMM* tem seu poder afetado não somente nas situações em que $p = \min(n_1 - 1, n_2 - 1)$, mas na medida em que p aumenta o poder de cada um desses testes diminui, inclusive em amostras grandes.

Os resultados para $k = 8$ não foram apresentados pois os três testes obtiveram poder igual a 1 em todas as configurações. Em nenhuma das situações estudadas até o momento, observou-se algum efeito da heterogeneidade das covariâncias sobre o poder.

A partir de agora serão apresentados os resultados para amostras com tamanhos diferentes. No primeiro caso estudado foi utilizando $n_1 = 8$, $n_2 = 30$, $p = 2$ e 7, $\alpha = 0,10, 0,05$ e 0,01, para $k = 1, 2, 4$ e 8. Na Figura 8 apresenta-se esses resultados quando $k = 1$. Pode-se observar que o padrão de desempenho dos três testes continuou o mesmo. Em nenhuma situação o poder do *TBP* foi inferior ao *TBPKL* e *TNVMM*. Quando o valor de p passou de 2 para 7 o poder do *TBP* apresentou um leve aumento. Quando $p = 7$, o *TBPKL* e o *TNVMM* não apresentaram curvas de poder tão próximas como anteriormente, o *TBPKL* mostrou-se levemente superior ao *TNVMM*. Como nesta situação, a diferença

entre médias é pequena, os valores de poder obtidos pelos três testes foram baixos.

Nas Figuras 9 e 10 estão apresentados os resultados para $k = 2$ e $k = 4$, respectivamente. Quando k passa de 1 para 2, a melhora do poder dos três testes é clara, sendo o aumento do *TBP* mais expressivo que os demais. Novamente quando o valor de p aumentou para 7 (Figuras 9b, 9d e 9f), o poder do *TBPKL* e do *TNVMM* diminuiu, enquanto que o do *TBP* aumentou. Em $p = 2$, Figuras 9a, 9c e 9e, o poder obtido pelos testes *TBPKL* e *TNVMM* são praticamente indistinguíveis. O mesmo não ocorreu quando $p = 7$, pois o *TBPKL* mostrou-se levemente superior ao *TNVMM*. Os resultados encontrados por Ramos e Ferreira (2012) para $p = 2$ e $n_1 = 8$, $n_2 = 30$, foram idênticos aos encontrados neste trabalho. O teste bayesiano proposto por eles apresentou poder inferior ao *TBP*. Novamente o aumento do número de variáveis feito por Ramos e Ferreira (2012) também provocou queda de poder para *TNVMM*.

Na Figura 10, onde $k = 4$, o *TBP* atinge valores de poder bem próximos ou iguais a 1. Também observa-se que a medida que se reduz α o poder do *TBPKL* e *TNVMM* também diminui, mas o mesmo não ocorreu para o *TBP* que manteve poder bem próximo ou igual a 1 em todas as situações. Mais uma vez pode-se observar que, quando $n_1 = 8$, $n_2 = 30$ e $p = 7$, os testes *TBPKL* e *TNMM* deixaram de apresentar valores semelhantes de poder, como em $p = 2$, sendo o *TBPKL* superior ao *TNVMM*. Já o *TBP* manteve seu comportamento superior aos demais testes, principalmente quando $p = 7$ (ver Figuras 8, 9 e 10). Os resultados de $k = 8$, $n_1 = 8$ e $n_2 = 30$ também não foram apresentados pois todos os testes atingiram valores de poder próximos ou iguais a 1 para todos os valores de p e α .

É bom lembrar que para o tamanho de amostras, $n_1 = 8$, $n_2 = 30$ e $p = 2$, os três testes foram considerados exatos, em relação ao erro tipo I, exceto o *TBP*

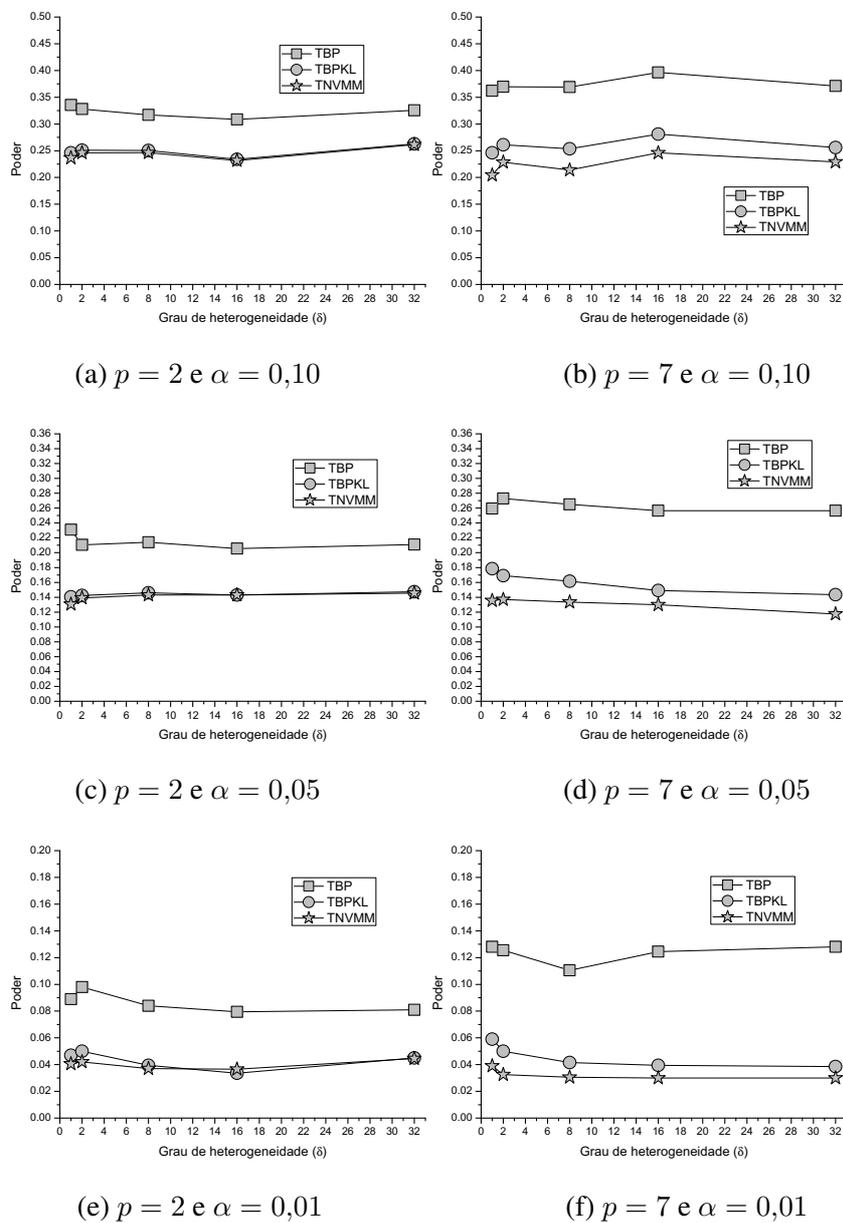


Figura 8 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

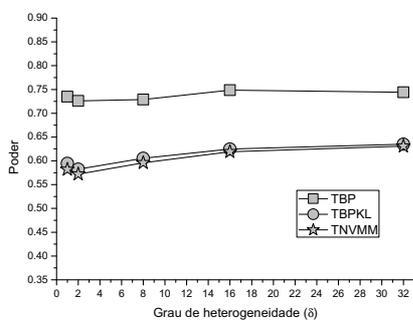
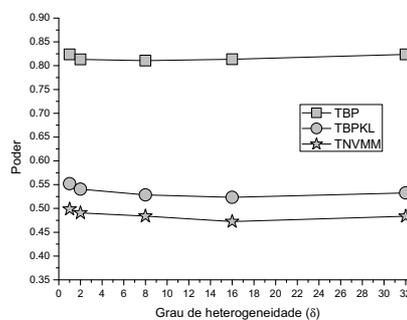
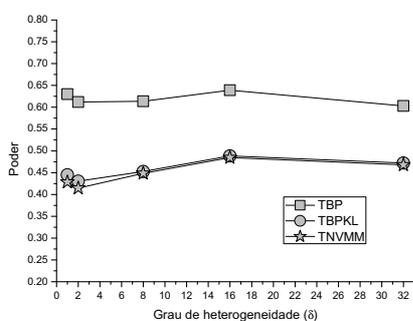
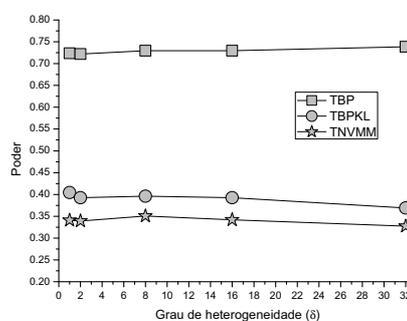
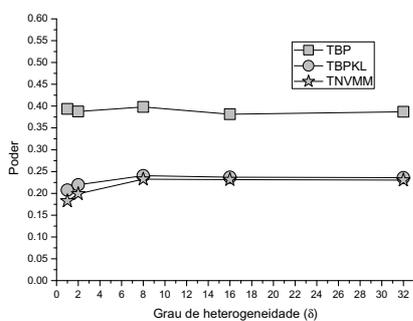
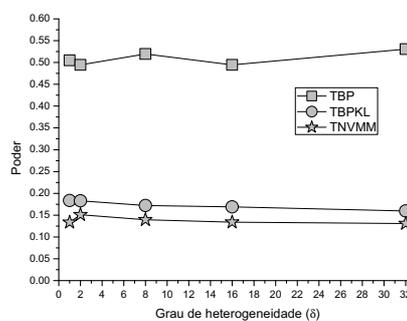
(a) $p = 2$ e $\alpha = 0,10$ (b) $p = 7$ e $\alpha = 0,10$ (c) $p = 2$ e $\alpha = 0,05$ (d) $p = 7$ e $\alpha = 0,05$ (e) $p = 2$ e $\alpha = 0,01$ (f) $p = 7$ e $\alpha = 0,01$

Figura 9 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

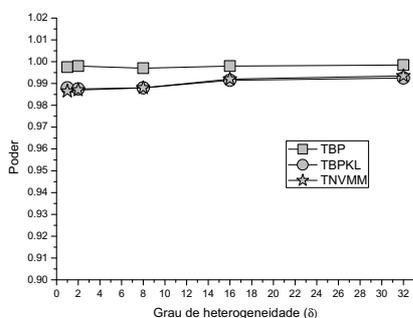
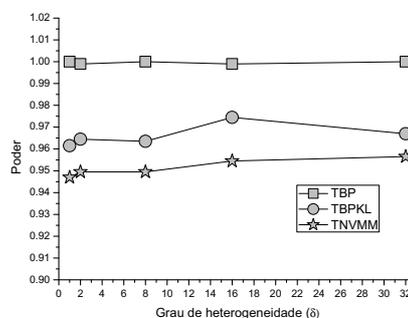
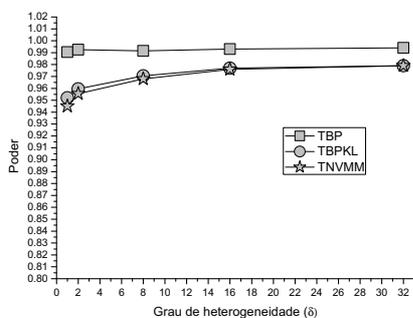
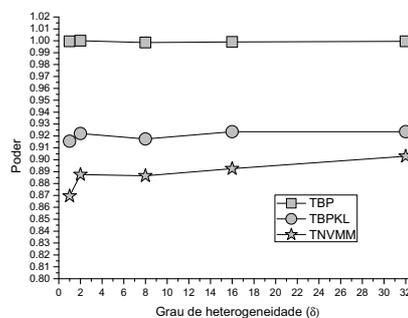
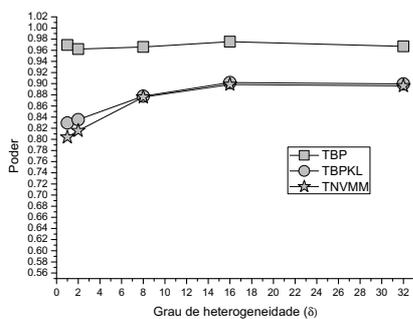
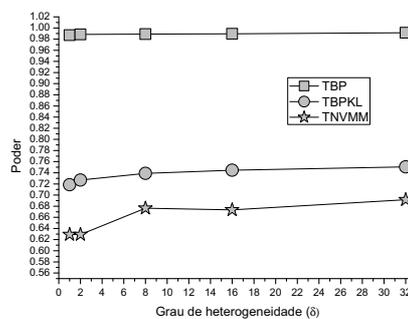
(a) $p = 2$ e $\alpha = 0,10$ (b) $p = 7$ e $\alpha = 0,10$ (c) $p = 2$ e $\alpha = 0,05$ (d) $p = 7$ e $\alpha = 0,05$ (e) $p = 2$ e $\alpha = 0,01$ (f) $p = 7$ e $\alpha = 0,01$

Figura 10 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 30$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

que foi considerado liberal em algumas situações, mas com valores bem próximos ao limite de rejeição (ver Tabela 1). Mas quando $p = 7$, a situação se inverte, o *TBPKL* e o *TNVMM* tornam-se liberais, exceto em $\alpha = 0,01$, enquanto o *TBP* manteve-se exato, também com exceção para $\alpha = 0,01$.

Utilizando as mesmas configurações das Figuras 8, 9 e 10, mas agora com os tamanhos amostrais invertidos, sendo $n_1 = 30$ e $n_2 = 8$, onde a maior covariância está relacionada a menor amostra, estão apresentados nos Anexos J e K. O que pode-se observar nestas situações é um comportamento muito parecido ao obtido quando $n_1 = 8$ e $n_2 = 30$. O *TBP* quando $p = 7$, apresenta um excelente desempenho em relação ao poder e o erro tipo I, enquanto o *TBPKL* e o *TNVMM* apresentaram valores de poder mais baixos e caráter liberal (Figura 3).

Na Figura 11, apresentam-se os resultados de poder para uma diferença maior entre os tamanhos amostrais, $n_1 = 8$ e $n_2 = 100$, utilizando $k = 1$. Mais uma vez a heterocedasticidade não proporciona alterações expressivas no poder para nenhum dos testes. Um fato importante nesta situação é que com $p = 7$ (Figuras 11b, 11d e 11f), o poder do *TBPKL* ficou muito próximo e até superior ao *TBP* quando $\alpha = 0,10$ e $0,05$, isso não se repetiu em nenhuma outra situação apresentada até o momento. Observa-se que esta proximidade é maior para baixos valores de δ , sendo que com $\alpha = 0,10$ e $\delta = 1$ e com $\alpha = 0,05$ e $\delta = 2$, o poder do *TBPKL* é levemente superior, mas isso não deve ser levado em consideração, pois o *TBPKL* foi considerado extremamente liberal nesta situação (Anexo C). Quando $p = 2$ o desempenho dos três testes foram bem similares aos da situação de $n_1 = 8$ e $n_2 = 30$ apresentada na Figura 8. Outro fato a salientar foi que quando o p passa de 2 para 7, o poder aumentou, mas o *TBPKL* apresentou aumento mais expressivo, mas ainda assim com valores menores que o *TBP*.

Para $n_1 = 8$, $n_2 = 100$ e utilizando $k = 2$, resultados apresentados na

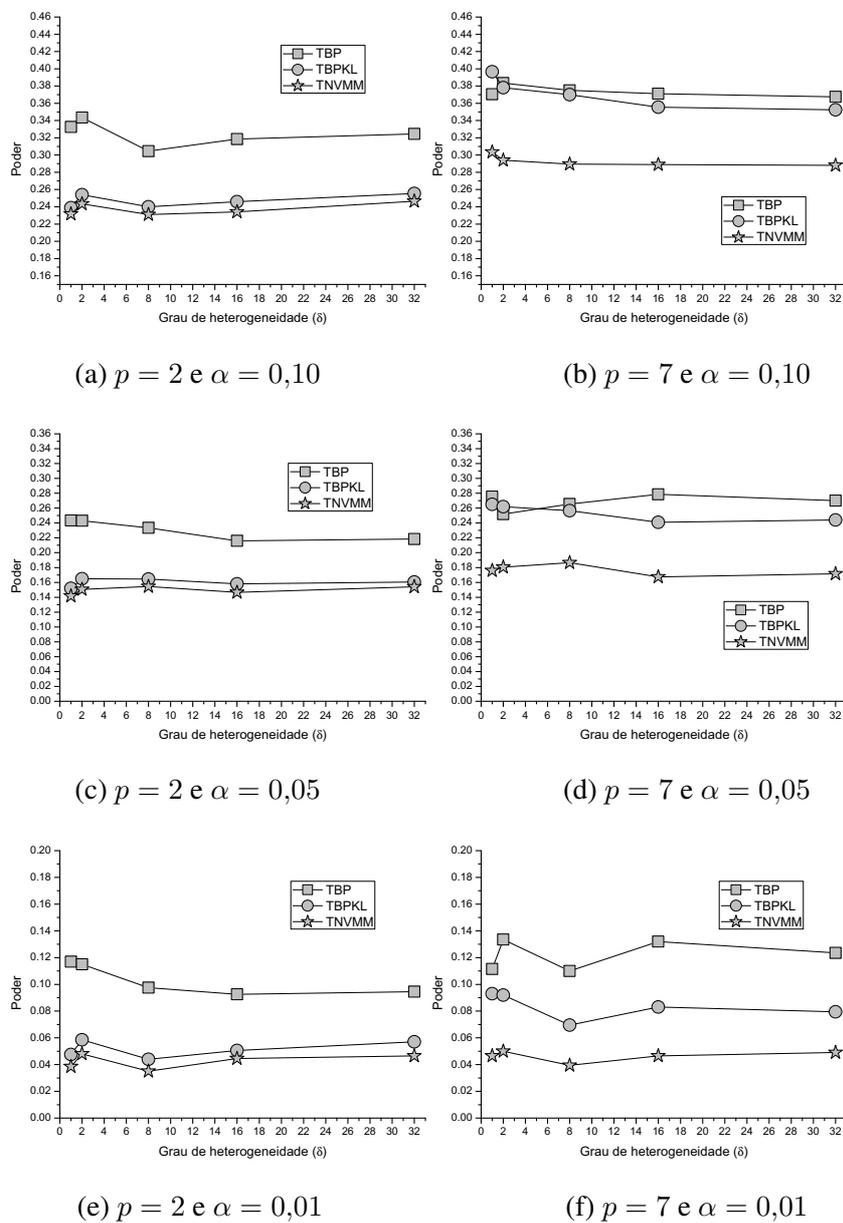


Figura 11 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

Figura 12, o *TBP* voltou a apresentar valores de poder bem mais altos que os dos demais testes. O que pode-se observar é que para $p = 2$, Figuras 12a, 12c e 12e, o *TBPKL* e o *TNVMM* apresentaram valores de poder bem próximos, mas quando o valor de p mudou para 7 o *TBPKL* teve um leve aumento no poder enquanto que o *TNVMM* apresentou queda. Em relação ao *TBP*, o aumento de p provocou aumento expressivo de poder, sendo sempre superior aos demais. De forma geral, amostras com tamanhos bem diferentes alterou apenas o padrão de comportamento do *TBPKL*, pois quando o valor de p aumentou o seu poder teve um leve aumento, fato que não aconteceu para outros tamanhos amostrais.

Ainda estudando o mesmo tamanho de amostras, $n_1 = 8$ e $n_2 = 100$, mas agora para $k = 4$, resultados apresentados na Figura 13, observou-se que a melhora de poder do *TBP* quando o valor de p aumenta, já não ficou tão evidente quanto em $k = 2$, pois quando $k = 4$, o *TBP* já alcança valores bem altos, próximos a 1, em todas as situações. Já o *TBPKL* e o *TNVMM*, novamente apresentaram queda no poder quando $p = 7$. Comparando as Figuras 13b, 13d e 13f com as Figuras 10b, 10d e 10f, $n_1 = 8$ e $n_2 = 30$, pode-se notar que a queda do poder do *TNVMM* é maior quando a diferença entre o tamanho das amostras é maior, já os demais testes praticamente não apresentaram alterações expressivas no poder com esse aumento de diferença entre o tamanho das amostras.

No Anexo L apresentam-se os resultados para $k = 8$, $n_1 = 8$ e $n_2 = 100$. Neste caso observou-se que em praticamente todas as situações, o poder dos três testes foi muito próximo de 1, exceto para $p = 7$ e $\alpha = 0,01$ (Anexo L (f)), em que o *TNVMM* apresentou valores de poder abaixo dos demais, chegando a 0,938 para $\delta = 2$, mas com melhora com o aumento de δ , atingindo 0,975 quando $\delta = 32$.

Os resultados nas mesmas situações apresentados nas Figuras 11, 12, 13

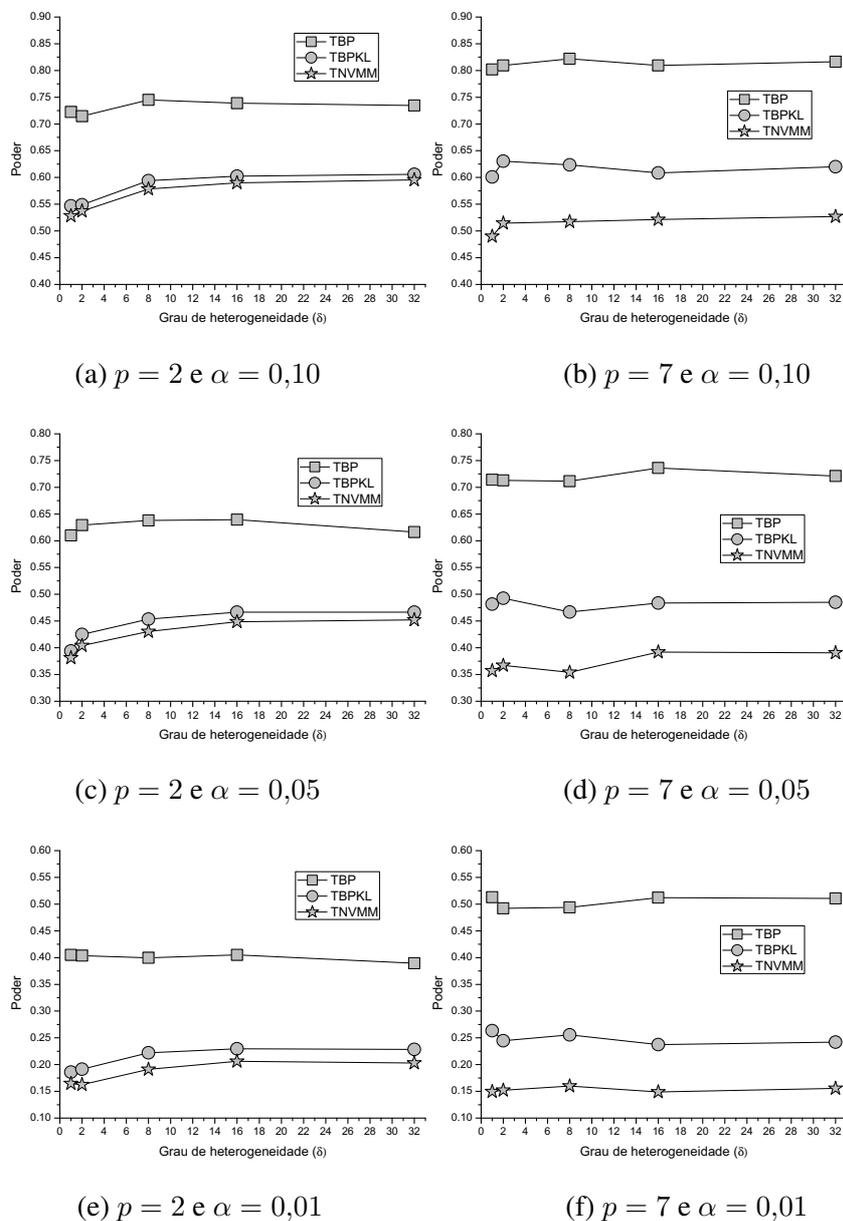


Figura 12 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

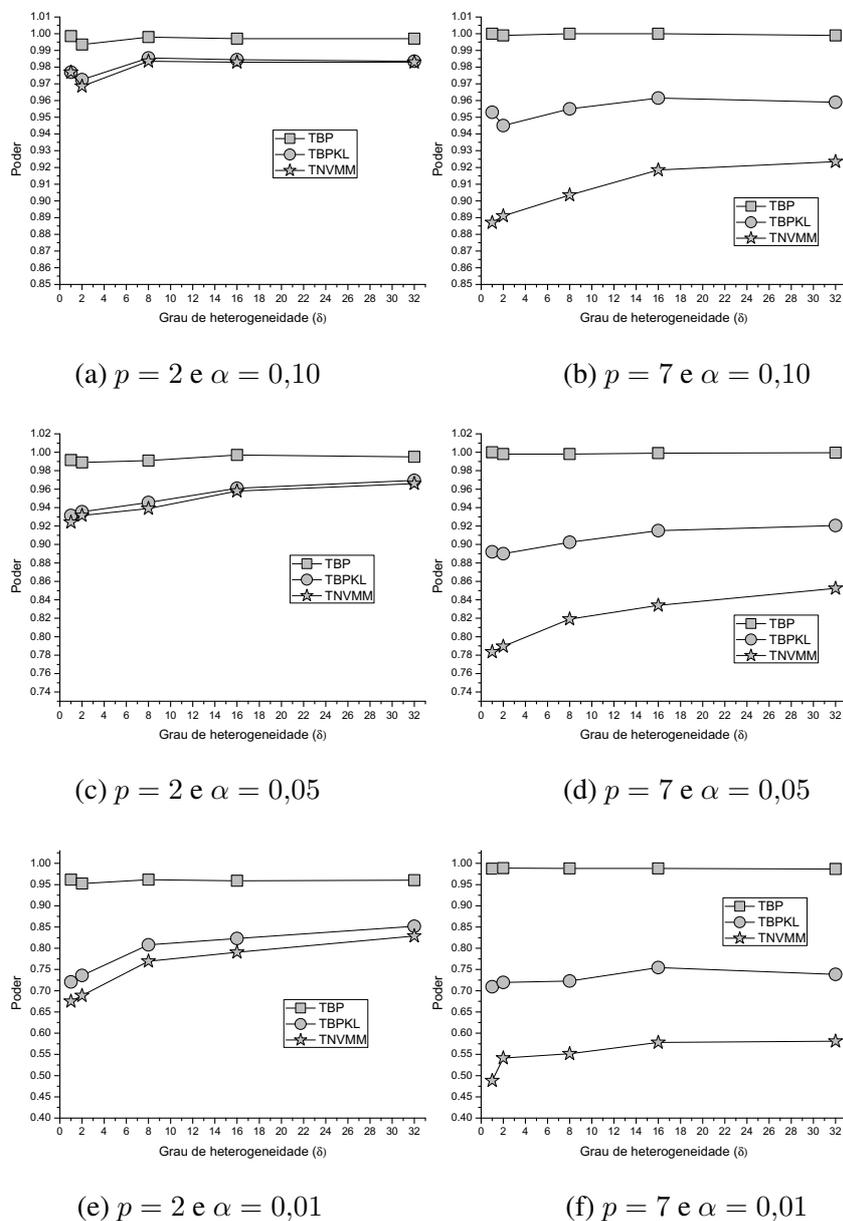


Figura 13 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

e Anexo L, mas agora com os tamanhos das amostras invertidos, $n_1 = 100$ e $n_2 = 8$, cuja a maior covariância está relacionada à amostra de tamanho menor, estão apresentados nas Figuras 14 e 15 e Tabela 2.

Na Figura 14 apresentam-se os resultados obtidos para $k = 1$. A inversão do tamanho amostral não trouxe grandes alterações no poder do *TBP*, em nenhuma das situações. Já para o *TBPKL*, quando $p = 7$, os valores de poder se apresentaram bem próximos e até superiores ao *TBP*, quando $\alpha = 0,10$ e $0,05$. Esse mesmo comportamento foi detectado na Figura 11, onde $n_1 = 8$ e $n_2 = 100$, mas com a inversão do tamanho das amostras isto ficou mais evidente (Figuras 14b e 14d), onde o *TBPKL* mostrou-se levemente superior ao *TBP* para todos os valores de δ . A medida que o valor de α diminuiu a superioridade do *TBPKL* diminuiu, sendo que em $\alpha = 0,01$ o *TBP* mostrou-se levemente superior ao *TBPKL* (Figura 14f).

Um fato interessante observado, quando $k = 1$, $n_1 = 100$ e $n_2 = 8$; $n_1 = 8$ e $n_2 = 100$, foi que o *TBPKL* apresentou boa melhora no poder com o aumento do valor de p . No entanto, o *TBP* e o *TNVMM* também apresentaram melhora, mas em menor magnitude. É importante lembrar que as taxas de erro tipo I, do *TBPKL*, com esses mesmos tamanhos amostrais e $p = 7$, foram bem mais altas que os demais testes (ver Anexo C e Figura 4), sendo assim, essa superioridade nos valores de poder não pode ser considerada, pois o teste é extremamente liberal.

Continuando a avaliar o poder com amostras de tamanhos $n_1 = 100$ e $n_2 = 8$, mas agora para $k = 2$ (Figura 15), o poder dos três testes aumentou, como já era esperado. O *TBP* manteve o mesmo padrão de comportamento encontrado na Figura 12, ou seja, a inversão do tamanho das amostras não afetou o poder do *TBP*, para nenhum p .

Comparando as Figuras 15a, 15c e 15e com as Figuras 12a, 12c e 12e,

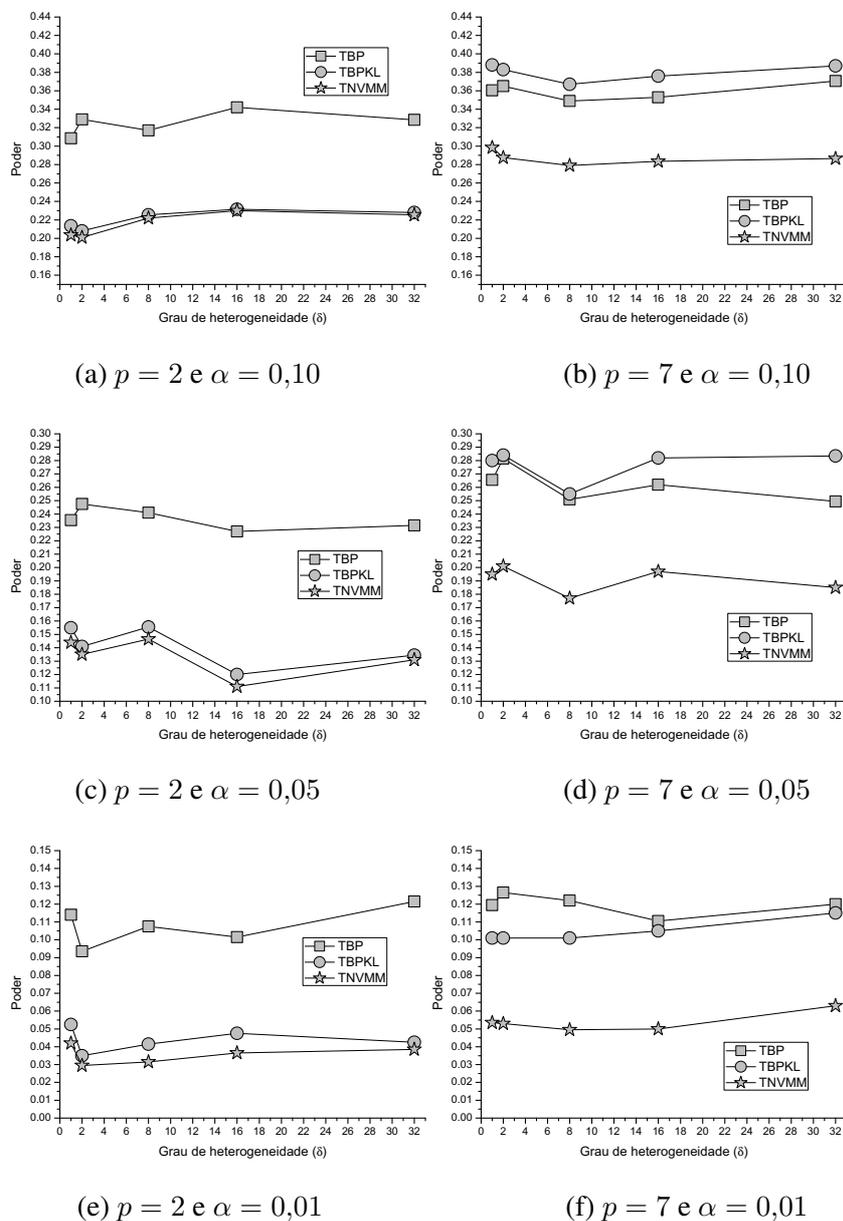


Figura 14 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 8$, $k = 1$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

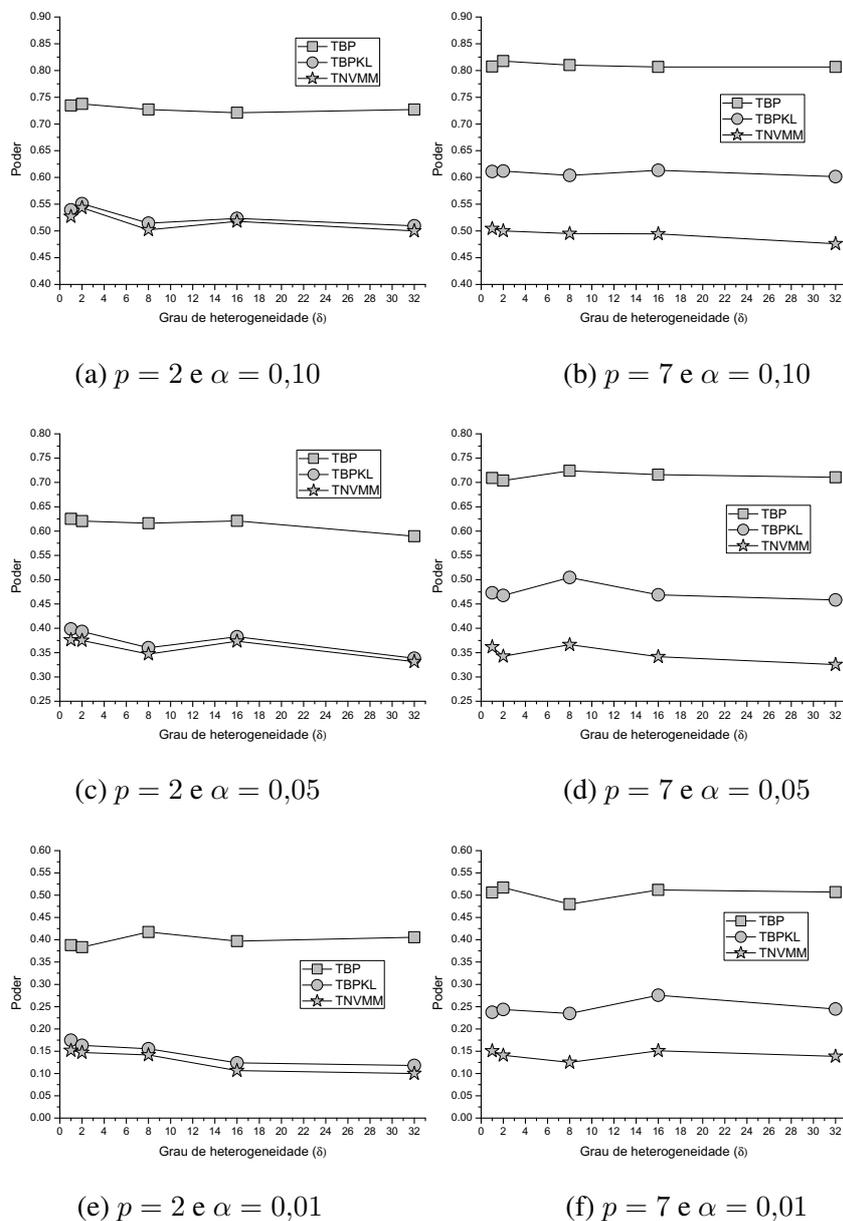


Figura 15 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 8$, $k = 2$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

onde $p = 2$, observou-se que com a inversão das amostras para uma situação mais crítica, os valores de poder do *TBPKL* e *TNVMM* apresentaram uma leve queda, enquanto o *TBP* não alterou seu comportamento. Nessa situação, o *TBPKL* e *TNVMM* foram considerados exatos e o *TBP* liberal, em relação ao erro tipo I, embora sem apresentar expressivas diferenças para os valores nominais de significância. Quando $p = 7$, a inversão do tamanho das amostras não causou grandes alterações no poder do *TBPKL* e *TNVMM*. Comparando a mudança de $p = 2$ para $p = 7$, nas amostras de $n_1 = 100$, $n_2 = 8$ e $k = 2$ (Figura 15), observou-se que *TNVMM* se mantém praticamente constante e o *TBPKL* e *TBP* apresentaram melhora no poder. Nesta ocasião, em relação ao erro tipo I, o *TBP* foi considerado exato na maioria das situações, já o *TBPKL* e o *TNVMM* apresentaram caráter liberal, sendo o *TBPKL* bem mais liberal que o *TNVMM*.

Na Tabela 2 a seguir, apresentam-se os resultados de poder quando $k = 4$ e $k = 8$ para $n_1 = 100$ e $n_2 = 8$. Nestas situações, novamente o *TBP* já apresenta valores de poder bem próximos ou iguais a 1. Quando $k = 4$ e $p = 2$ o *TBPKL* e o *TNVMM* continuaram apresentando valores de poder bem próximos entre si, inferiores ao do *TBP* e à medida em que α diminuiu, a diferença de poder entre o *TBP* e os demais testes aumentou. Alterando o valor de p para 7, o *TBP* continua com valores bem próximos de 1, mas o *TBPKL* e o *TNVMM* passaram a não mais apresentar valores de poder tão próximos entre si, sendo o *TBPKL* superior ao *TNVMM* para todos os valores de α .

Observando os resultados da Tabela 2, mas agora focando em $k = 8$, nota-se que com $p = 2$, os três testes têm um ótimo desempenho, mas quando $p = 7$, o *TNVMM* apresentou-se menos poderoso que o *TBP* e o *TBPKL*, principalmente para $\alpha = 0,01$.

Tabela 3 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$ e $n_2 = 8$, $k = 4$ e 8 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α

		$n_1 = 100$ e $n_2 = 8$						
		$k = 4$			$k = 8$			
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM	
$\alpha = 0,10$	p=2	1	0,9970	0,9700	0,9670	1,0000	1,0000	1,0000
		2	0,9975	0,9765	0,9740	1,0000	1,0000	1,0000
		8	0,9990	0,9665	0,9650	1,0000	1,0000	1,0000
		16	0,9975	0,9695	0,9700	1,0000	1,0000	1,0000
		32	0,9975	0,9675	0,9675	1,0000	1,0000	1,0000
	p=7	1	1,0000	0,9490	0,8930	1,0000	1,0000	0,9990
		2	0,9995	0,9445	0,8820	1,0000	1,0000	0,9970
		8	0,9995	0,9305	0,8540	1,0000	1,0000	0,9985
		16	0,9995	0,9270	0,8430	1,0000	0,9995	0,9935
		32	0,9995	0,9225	0,8310	1,0000	0,9995	0,9940
$\alpha = 0,05$	p=2	1	0,9890	0,9290	0,9205	1,0000	1,0000	1,0000
		2	0,9860	0,9095	0,9040	1,0000	1,0000	1,0000
		8	0,9915	0,8930	0,8905	1,0000	1,0000	1,0000
		16	0,9900	0,9135	0,9120	1,0000	1,0000	1,0000
		32	0,9930	0,9115	0,9070	1,0000	1,0000	1,0000
	p=7	1	0,9980	0,8895	0,7845	1,0000	0,9995	0,9885
		2	0,9990	0,8840	0,7590	1,0000	1,0000	0,9870
		8	1,0000	0,8775	0,7480	1,0000	0,9990	0,9815
		16	0,9975	0,8750	0,7475	1,0000	0,9985	0,9755
		32	0,9980	0,8625	0,7060	1,0000	0,9980	0,9760
$\alpha = 0,01$	p=2	1	0,9605	0,7190	0,6750	1,0000	0,9995	0,9995
		2	0,9490	0,6950	0,6540	1,0000	0,9995	0,9995
		8	0,9515	0,6345	0,6030	1,0000	0,9965	0,9960
		16	0,9480	0,6285	0,6010	1,0000	0,9980	0,9985
		32	0,9575	0,6125	0,5965	1,0000	0,9980	0,9995
	p=7	1	0,9865	0,7230	0,5195	1,0000	0,9900	0,9285
		2	0,9885	0,6940	0,4840	1,0000	0,9950	0,9165
		8	0,9855	0,6870	0,4595	1,0000	0,9905	0,8890
		16	0,9885	0,6650	0,4345	1,0000	0,9855	0,8750
		32	0,9890	0,6600	0,4175	1,0000	0,9810	0,8435

No Anexo M apresentam-se os valores de poder obtidos pelos três testes em estudo, utilizando $n_1 = 30$, $n_2 = 100$, $p = 7$ e 29 , com valores de $k = 1$ e 2 . Com este tamanho amostral não houve grandes alterações no padrão de comportamento dos testes. A heterogeneidade de covariâncias não teve efeito expressivo sobre o poder, pois os valores não apresentaram grandes alterações na medida em que o δ aumenta. Novamente o *TBP* mostrou-se mais poderoso em todas as situações. Quando o valor de p é aumentado, o poder do *TBP* também aumenta, enquanto o *TBPKL* e o *TNVMM* apresentaram queda de poder. Esse comportamento é verificado tanto para $k = 1$ quanto para $k = 2$.

Continuando o estudo com $n_1 = 30$ e $n_2 = 100$, mas para $k = 4$ (Anexo N), o *TBP* apresentou valores bem próximos de 1. Mais uma vez, o aumento do número de variáveis provocou queda no poder do *TBPKL* e do *TNVMM*. O mesmo já não ocorreu para o *TBP*, que manteve-se praticamente constante com o aumento de p . Como já era esperado, para grande amostras e grande distância entre médias, $k = 8$, todos os testes em todas as situações apresentaram poder igual a 1 (resultados não apresentados).

Por fim, o tamanho das amostras foi invertido para $n_1 = 100$ e $n_2 = 30$, os resultados estão apresentados no Anexos O e P. Com essa inversão o comportamento dos três testes, em relação ao poder, não se alterou, mantendo as mesmas características de quando $n_1 = 30$ e $n_2 = 100$.

De uma maneira geral, o que pode-se dizer sobre o poder dos testes em estudo é que o *TBP* apresentou os maiores valores de poder, exceto quando o tamanho das amostras foi muito diferente um do outro, $n_1 = 100$, $n_2 = 8$ e $n_1 = 8$, $n_2 = 100$, mas apenas quando $k = 1$, $p = 7$ e $\alpha = 0,10$ e $0,05$. Nestas situações, apenas o *TBPKL* apresentou uma leve superioridade sobre o *TBP*, mas o *TBPKL* foi considerado bem liberal, enquanto o *TBP* foi mais exato

(Figuras 11 e 14).

Observando apenas as situações em que existe homogeneidade das covariâncias, $\delta = 1$, o *TBP* também apresenta superioridade em relação ao poder, exceto nas situações mencionadas acima. Com isso, pode-se dizer que o *TBP* é mais poderoso que os demais testes mesmo quando as covariâncias são homogêneas.

4.2 Caso de matrizes de covariâncias não estruturadas

4.2.1 Erro tipo I

Na Figura 16 a seguir apresentam-se os resultados de erro tipo I dos três testes em estudo, quando as amostras das duas populações possuem tamanhos iguais. Para amostras pequenas, $n_1 = n_2 = 8$ (Figura 16(a)), os três testes apresentaram taxas adequadas para erro tipo I, exceto o *TBP* que apresentou caráter liberal, quando $p = 2$, $\alpha = 5\%$ e 1% .

Aumentando o tamanho das amostras para, $n_1 = n_2 = 30$ (Figura 16(b)), os testes *TBPKL* e *TNVMM* continuaram apresentando taxas exatas de erro tipo I, já o *TBP* foi levemente conservador apenas quando $p = 7$, $\alpha = 5\%$ e $p = 29$, $\alpha = 1\%$. Para tamanhos amostrais grandes (Figura 16(c)), todos os testes, em todas as situações, apresentaram taxas de erro tipo I que podem ser consideradas iguais ao α correspondente.

Na Figura 17 apresentam-se os resultados de erro tipo I utilizando tamanhos amostrais diferentes. Para $n_1 = 8$, $n_2 = 30$ (Figura 17(a)), os três testes apresentaram taxas exatas de erro tipo I, quando $p = 2$. Com número de variáveis maior, $p = 7$, o *TBP* manteve-se com taxas exatas, mas o *TBPKL* e o *TNVMM* apresentaram caráter liberal para $\alpha = 10\%$ e 5% . Utilizando os mesmos tamanhos amostrais, mas invertidos, $n_1 = 30$ e $n_2 = 8$ (Figura 17(b)), para

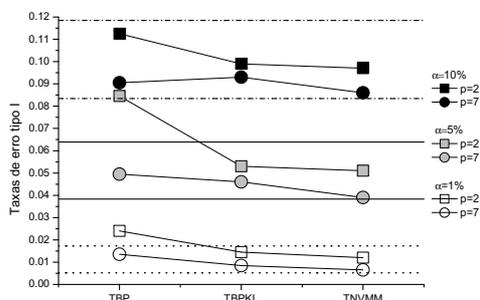
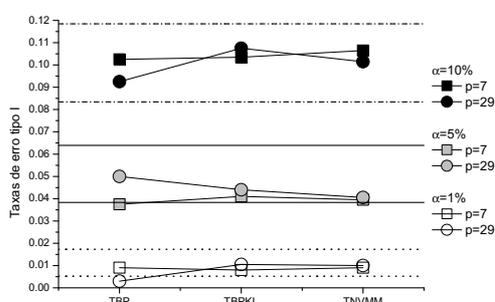
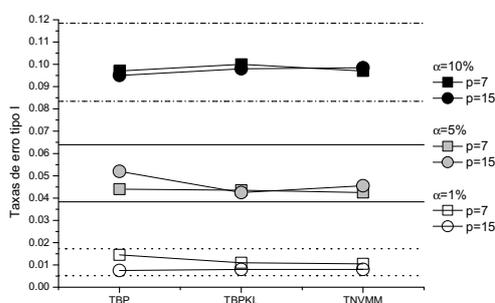
(a) $n_1 = n_2 = 8$ (b) $n_1 = n_2 = 30$ (c) $n_1 = n_2 = 100$

Figura 16 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função de tamanhos amostrais iguais, $n_1 = n_2 = 8$; $n_1 = n_2 = 30$ e $n_1 = n_2 = 100$, número de variáveis p e nível nominal de significância α , sob H_0 . Cada par de linhas horizontais indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, respectivamente, com coeficiente de confiança de 0,99

$p = 2$, o *TBPKL* e o *TNVMM* mantiveram taxas exatas de erro tipo I para todos os valores de α . Já o *TBP* apresentou leve caráter liberal nesta situação. Quando o valor de p foi aumentado para 7, o *TBP* pode ser considerado exato, mas o *TBPKL* e o *TNVMM* tornaram-se extremamente liberais, para $\alpha = 10\%$ e 5%. Já para $\alpha = 1\%$ e $p = 7$, os três testes podem ser considerados exatos.

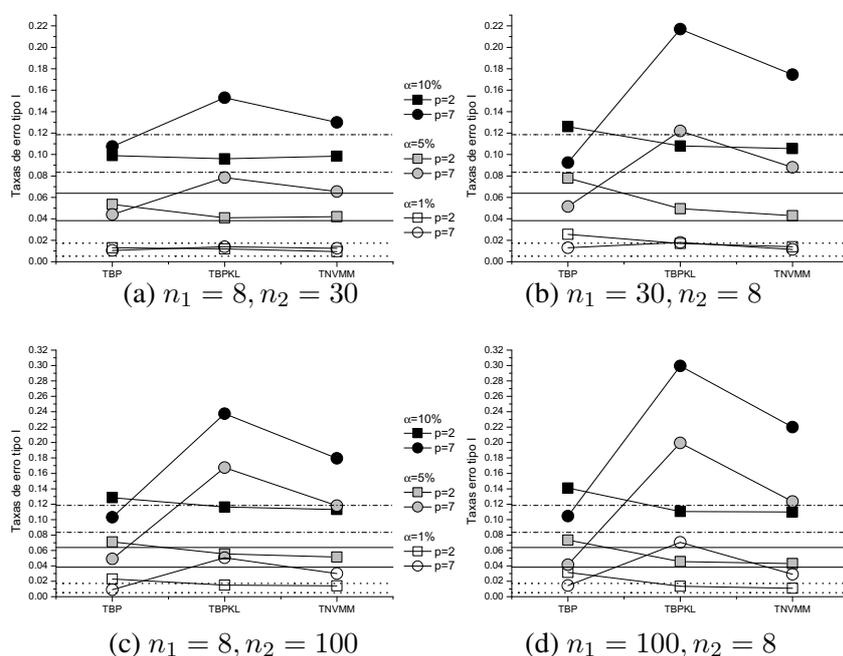


Figura 17 Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função de tamanhos amostrais diferentes, número de variáveis p e nível nominal de significância α , sob H_0 . Cada par de linhas horizontais indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, respectivamente, com coeficiente de confiança de 0,99

Os resultados para maiores diferenças entre os tamanhos amostrais, $n_1 = 8$ e $n_2 = 100$ (Figura 17(c)), o *TBP*, quando $p = 7$, apresentou-se exato, mas

para $p = 2$ apresentou caráter levemente liberal. Já o *TBPKL* e o *TNVMM* mostraram um comportamento inverso ao do *TBP*: taxas exatas de erro tipo I, para $p = 2$, e comportamento extremamente liberal para $p = 7$, principalmente quando $\alpha = 10\%$ e 5% .

Invertendo os tamanhos amostrais, $n_1 = 100$ e $n_2 = 8$ (Figura 17(d)), os três testes mantiveram o mesmo padrão de comportamento de quando $n_1 = 8$ e $n_2 = 100$, exceto que as taxas de erro tipo I do *TBPKL* e do *TNVMM* foram mais altas, quando $p = 7$.

De uma maneira geral, o que pode-se dizer sobre as taxas de erro tipo I, nos casos em que as matrizes de covariâncias são não estruturadas, é que o aumento do número de variáveis provocou queda das taxas de erro tipo I para o *TBP*. No entanto, para o *TBPKL* e o *TNVMM*, esse aumento do número de variáveis provocou expressivo aumento nos valores das taxas de erro tipo I, principalmente quando a população com covariâncias maiores estava relacionada ao menor tamanho amostral, ou seja, para $n_1 = 30$; $n_2 = 8$ e $n_1 = 100$; $n_2 = 8$.

4.2.2 Poder

Na Tabela 3, a seguir, apresentam-se valores de poder para os três testes em estudo, utilizando tamanhos amostrais iguais. Observando os resultados desta tabela, para amostras pequenas, $n_1 = n_2 = 8$ e $k = 1$, nota-se que quando $p = 2$ e $\alpha = 10\%$, o *TBP* apresentou poder levemente inferior aos demais testes. Já para $\alpha = 5$ e 1% , o *TBP* mostrou-se mais poderoso que os demais. No entanto, nessa mesma situação o teste proposto apresentou caráter liberal, em relação ao erro tipo I, enquanto o *TBPKL* e o *TNVMM* foram considerados exatos. (Figura 16 (a))

Ainda para amostras pequenas, mas aumentando o valor de p para 7, provocou queda de poder para o *TBPKL* e o *TNVMM*, já o *TBP* apresentou maior

poder que os demais, mas isso ocorreu apenas para $\alpha = 10$ e 5% . Para $\alpha = 1\%$, os três testes apresentaram aumento de poder, sendo que o mais poderoso foi o *TBPKL*, seguido pelo *TNVMM* e o de menor poder foi o *TBP*. Vale ressaltar que as diferenças de poder, entre os três testes quando $n_1 = n_2 = 8$, $p = 7$ e $\alpha = 1\%$, são bem pequenas. O aumento da diferença entre médias para $k = 2$ provocou, como já era esperado, melhora de poder nos três testes. Observando os resultados quando $p = 2$, nota-se um comportamento semelhante ao encontrado quando $k = 1$, porém com maiores valores de poder. Para $p = 7$, o *TBP* foi mais poderoso que seus concorrentes para todos os valores de α .

Amostras de tamanhos, $n_1 = n_2 = 30$, $k = 1$ e $p = 7$ (Tabela 3), o *TBPKL* e o *TNVMM* apresentaram valores de poder bem semelhantes e mais altos que os do *TBP*, para todos os níveis de significância. O mesmo ocorreu para $p = 29$. Continuando com o mesmo tamanho amostral $n_1 = n_2 = 30$, mas para $k = 2$, pode-se dizer que os três testes apresentaram valores de poder bem próximos, em todas as situações, e mais altos do que quando $k = 1$. Para este tamanho amostral, o *TBPKL* e o *TNVMM* foram considerados exatos, em relação ao erro tipo I, e o *TBP* levemente conservador em algumas situações. (Figura 16(b))

Estudando amostras grandes e iguais, $n_1 = n_2 = 100$, para $k = 1$ (Tabela 3), o que observou-se foi que os testes *TBPKL* e *TNVMM* apresentaram valores de poder bem próximos e maiores que o *TBP* proposto. Isso ocorreu para todos os valores de α , p e k . Sendo que para $k = 2$, as diferenças de poder entre o *TBP* e os seus concorrentes foi menor. Nesta situação os três testes foram considerados exatos, em relação ao erro tipo I, para todas as configurações estudadas. (Figura 16(c))

Analisando os resultados para tamanhos amostrais diferentes, $n_1 = 8$ e

Tabela 4 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função de tamanhos amostrais iguais, $k = 1$ e 2 , número de variáveis (p) e nível nominal de significância α

		k=1					
		$n_1 = 8 \quad n_2 = 8$					
		p=2			p=7		
α		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,3220	0,4020	0,3935	0,4095	0,2990	0,2805
	5	0,2305	0,2070	0,2015	0,2715	0,1960	0,1805
	1	0,0895	0,0400	0,0390	0,0980	0,1280	0,1155
		$n_1 = 30 \quad n_2 = 30$					
		p=7			p=29		
α		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,4365	0,7010	0,6995	0,7955	1,0000	1,0000
	5	0,2645	0,9955	0,9955	0,5890	1,0000	1,0000
	1	0,0935	0,2900	0,2930	0,2975	0,9955	0,9940
		$n_1 = 100 \quad n_2 = 100$					
		p=7			p=15		
α		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,4410	0,5530	0,5560	0,6010	1,0000	1,0000
	5	0,2995	0,7945	0,7945	0,4640	0,8660	0,8665
	1	0,0785	0,4580	0,4610	0,1900	0,9985	0,9990
		k=2					
		$n_1 = 8 \quad n_2 = 8$					
		p=2			p=7		
α		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,8175	0,8700	0,8660	0,8920	0,7720	0,7500
	5	0,6580	0,5545	0,5445	0,9005	0,7935	0,7570
	1	0,4425	0,2485	0,2285	0,7015	0,4580	0,4175
		$n_1 = 30 \quad n_2 = 30$					
		p=7			p=29		
α		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,9915	1,0000	1,0000	1,0000	1,0000	1,0000
	5	0,9610	1,0000	0,9995	1,0000	1,0000	1,0000
	1	0,7430	0,7520	0,7585	0,9995	1,0000	1,0000
		$n_1 = 100 \quad n_2 = 100$					
		p=7			p=15		
α		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,9800	0,9745	0,9745	0,9995	1,0000	1,0000
	5	0,9425	0,9345	0,9345	1,0000	1,0000	1,0000
	1	0,8190	0,9330	0,9360	0,9815	1,0000	1,0000

$n_2 = 30$ (Tabela 4), observa-se que para $k = 1$ e $p = 2$, os três testes apresentaram valores de poder semelhantes, em todos os valores de α . O mesmo ocorreu quando $k = 2$, porém com valores de poder mais altos. Nesta mesma situação os três testes foram considerados exatos, em relação ao erro tipo I. Aumentando o número de variáveis para 7 e com $k = 1$, observou-se que o poder dos três testes melhoraram, em relação a situação de $p = 2$, sendo em primeiro lugar o *TBPKL*, em seguida o *TNVMM* e por último o *TBP*. No entanto, apenas o *TBP* foi considerado exato, em relação ao erro tipo I, os demais testes foram considerados liberais para $\alpha = 10$ e 5%. Para uma diferença entre médias maior, $k = 2$, nota-se que os valores de poder dos três testes se aproximaram, exceto o *TBP* quando $\alpha = 1\%$, que foi inferior aos demais.

A Tabela 4, também aparecem os valores de poder para a mesma situação descrita anteriormente, mas com os tamanhos amostrais invertidos, $n_1 = 30$ e $n_2 = 8$. Para este caso, com $k = 1$ e $p = 2$, a inversão dos tamanhos amostrais causou queda de poder para o *TBPKL* e o *TNVMM*. Já para o *TBP* o poder aumentou, mas nesta situação o teste proposto foi considerado liberal em relação ao erro tipo I. Quando k aumenta para 2, o *TBPKL* e o *TNVMM*, continuaram apresentando valores de poder bem próximos, enquanto o *TBP* foi superior aos seus concorrentes apenas para $\alpha = 5\%$ e 1%. Quando o número de variáveis aumenta para $p = 7$ e com $k = 1$, o poder de todos os testes aumentou, exceto o *TBP* com $\alpha = 1\%$, comparando com os resultados de $p = 2$. Neste caso o *TBP* apresentou valores de poder menores que seus concorrentes, no entanto foi o único a ser considerado exato, para todos os níveis de significância, em relação ao erro tipo I. Quando o valor de k aumentou para 2, o comportamento dos três testes foi semelhante ao caso de $k = 1$, porém com valores de poder mais altos. A única exceção para esse comportamento foi com o *TBP* que apresentou poder

Tabela 5 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$; $n_2 = 30$ e $n_1 = 30$; $n_2 = 8$, $k = 1$ e 2 , número de variáveis (p) e nível nominal de significância α

		k=1					
		$n_1 = 8 \ n_2 = 30$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,3535	0,3605	0,3605	0,4510	0,6615	0,6210
	5	0,2005	0,2095	0,2070	0,2860	0,5080	0,4575
	1	0,0645	0,0725	0,0670	0,1165	0,1515	0,1270
		$n_1 = 30 \ n_2 = 8$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,3290	0,3430	0,3335	0,4300	0,6120	0,5440
	5	0,2575	0,1855	0,1720	0,2975	0,6135	0,5235
	1	0,0940	0,0590	0,0495	0,0925	0,2155	0,1475
		k=2					
		$n_1 = 8 \ n_2 = 30$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,8070	0,8015	0,7995	0,9765	1,0000	1,0000
	5	0,6290	0,5530	0,5465	0,9035	0,9460	0,9235
	1	0,2925	0,3200	0,3200	0,7845	0,9870	0,9775
		$n_1 = 30 \ n_2 = 8$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,7030	0,8275	0,8230	0,9600	0,9925	0,9780
	5	0,6230	0,4330	0,4170	0,9350	1,0000	1,0000
	1	0,4510	0,2415	0,2185	0,6035	0,4960	0,3655

maior que os concorrentes, quando $\alpha = 1\%$, fato que não ocorreu quando $k = 1$.

De maneira geral, os resultados apresentados na Tabela 4 indicam que para os tamanhos amostrais $n_1 = 8$ e $n_2 = 30$, com $p = 2$, o *TBP* apresentou desempenho semelhante aos seus competidores, sendo superior nas situações de $k = 2$, $\alpha = 10\%$ e 5% . No caso de $p = 7$, ainda com $n_1 = 8$ e $n_2 = 30$, as taxas de poder do *TBP* podem ser consideradas reais, no entanto o mesmo não pode ser dito aos seus concorrentes, para $\alpha = 10\%$ e 5% , pois esses testes foram bastante liberais. Com o tamanho de amostras invertido, $n_1 = 30$, $n_2 = 8$ e $p = 7$, também pode-se dizer que os valores de poder são reais para o *TBP*, inclusive quando $k = 2$ e $\alpha = 1\%$. O *TBP* apresentou poder maior que os concorrentes, situação a qual foi considerado exato.

Na Tabela 5, a seguir, observam-se os resultados para diferenças maiores entre os tamanhos amostrais, $n_1 = 8$ e $n_2 = 100$. Quando $k = 1$ e $p = 2$, o *TBP_{KL}* e o *TNVMM* apresentaram poder com valores bem semelhantes e maiores que os do *TBP*, exceto com $\alpha = 10\%$. Para $k = 2$, o poder dos dois testes concorrentes continuaram bem próximos e superiores ao do *TBP*, exceto em $\alpha = 1\%$ o *TNVMM* foi levemente inferior ao *TBP*.

Continuando a estudar esse mesmo tamanho amostral, $n_1 = 8$ e $n_2 = 100$, mas agora com $p = 7$ (Tabela 4), nota-se que quando $k = 1$ o *TBP* continuou apresentando valores de poder inferiores aos seus competidores, no entanto, nesta situação o *TBP* foi considerado exato em relação ao erro tipo I, enquanto o *TBP_{KL}* e o *TNVMM* apresentaram forte caráter liberal. Com o aumento da diferença entre médias para $k = 2$, o comportamento dos três testes foi semelhante ao encontrado quando $k = 1$, porém com valores de poder maiores e as diferenças entre o poder do *TBP* e seus competidores, menores.

Invertendo os tamanhos amostrais para $n_1 = 100$ e $n_2 = 8$, observou-se

Tabela 6 Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8; n_2 = 100$ e $n_1 = 100; n_2 = 8$, $k = 1$ e 2 , número de variáveis (p) e nível nominal de significância α

		k=1					
		$n_1 = 8 \quad n_2 = 100$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,3000	0,2610	0,2560	0,4490	0,6570	0,5585
	5	0,2165	0,2755	0,2635	0,2750	0,4285	0,3395
	1	0,0775	0,1200	0,1085	0,0780	0,5130	0,3575
		$n_1 = 100 \quad n_2 = 8$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,3480	0,2280	0,2260	0,4225	0,6885	0,5685
	5	0,2435	0,1755	0,1705	0,2820	0,4150	0,2930
	1	0,1050	0,0925	0,0835	0,0665	0,3525	0,2130
		k=2					
		$n_1 = 8 \quad n_2 = 100$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,8210	0,8520	0,8425	0,9655	0,9730	0,9480
	5	0,6830	0,7445	0,7320	0,9430	0,9995	0,9960
	1	0,4690	0,4745	0,4420	0,6920	0,8310	0,6875
		$n_1 = 100 \quad n_2 = 8$					
		p=2			p=7		
	α	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM
	10	0,8085	0,7185	0,7145	0,9145	0,9135	0,8155
	5	0,6085	0,3790	0,3760	0,8935	0,8725	0,7280
	1	0,3575	0,2445	0,2330	0,6925	0,8595	0,6040

que os valores de poder do *TBP* foram superiores aos do *TBPKL* e o *TNVMM*, quando $p = 2$. Isso ocorreu tanto para $k = 1$ quanto para $k = 2$. Porém o *TBP* apresentou taxas de erro tipo I consideradas levemente liberais. Já o *TBPKL* e *TNVMM* foram considerados exatos. O aumento de p para 7, provocou aumento de poder em todos os testes, exceto para o *TBP* quando $k = 1$ e $\alpha = 1\%$. Observando os três testes, apenas em $k = 1$, o poder do *TBPKL* e do *TNVMM* foi maior que o do *TBP*, no entanto nesta mesma situação, o dois testes concorrentes, foram considerados extremamente liberais em relação ao erro tipo I. (Figura 17(d))

Ainda estudando os tamanhos amostrais, $n_1 = 100$ e $n_2 = 8$, para $p = 7$ mas agora com $k = 2$, observou-se que o *TBP* apresentou maior poder que seus concorrentes, quando $\alpha = 10\%$ e 5% . Nesta situação o *TBPKL* e o *TNVMM* foram considerados fortemente liberais e o *TBP* exato, no que diz respeito ao erro tipo I. R

De forma geral, os resultados da Tabela 5 indicam que para $p = 7$, o *TBP* apresentou um bom desempenho, pois seus valores de poder podem ser considerados reais, já para os concorrentes não pode-se fazer essa afirmação. Para $p = 2$, o *TBP* apresentou desempenho semelhante aos seus concorrentes.

5 CONCLUSÃO

O método via bootstrap paramétrico foi proposto e a avaliação do seu desempenho e comparação com os demais testes foram realizadas. As conclusões encontradas estão apresentadas a seguir.

1) Caso de matrizes de covariâncias estruturadas

Em relação ao poder, o teste proposto (*TBP*) apresentou os maiores valores em praticamente todas as situações, exceto em algumas em que o *TBPKL* foi levemente superior, mas isso não pode ser levado em consideração, pois tal teste foi extremamente liberal nessas ocasiões.

De forma definitiva, em relação ao poder, o *TBP* apresentou o melhor desempenho para todos os tamanhos amostrais. O *TNVMM* e o *TBPKL* apresentaram valores de poder muito semelhantes quando os tamanhos amostrais são iguais. Para as situações desbalanceadas e com menor dimensão o *TNVMM* e o *TBPKL* mantiveram-se muito próximos, mas quando o número de variáveis aumenta o *TBPKL* torna-se um pouco mais poderoso que o *TNVMM*.

As taxas de erro tipo I, de maneira geral, foram melhores no *TBP* quando em maiores dimensões, pois os demais testes mostraram-se extremamente liberais nessas ocasiões, enquanto o *TBP* é exato ou levemente liberal. Quando o número de variáveis é baixo, o *TBP* proposto neste trabalho, mostrou-se mais uma vez levemente liberal, mas os altos valores de poder compensam essa pequena liberalidade. O *TBP* mantém taxas de erro tipo I praticamente estáveis com a oscilação do número de variáveis, já os outros testes apresentaram grande variabilidade, mostrando um forte caráter liberal com o aumento de p . A superioridade do *TBP* tanto para o poder como nas taxas de erro tipo I, se mantém mesmo sob homogeneidade de covariâncias.

Sendo assim, o teste proposto neste trabalho é recomendado tanto para maiores quanto menores dimensões, pois seu caráter levemente liberal, quando em poucas variáveis, é compensado pelos seus altos valores de poder. Uma alternativa nessa situação seria a aplicação do *TBP* com um nível de significância menor, 0,01. O *TBP* também é recomendado em situações de homogeneidade das covariâncias, ou seja, o teste proposto pode ser aplicado sem a preocupação de homogeneidade ou heterogeneidade das matrizes de covariâncias.

2) Caso de matrizes de covariâncias não estruturadas

Em relação ao erro tipo I, para amostras com tamanhos iguais, o *TBPKL* e o *TNVMM* apresentaram ótimo desempenho em todas as configurações estudadas. Já o *TBP* apresentou bom desempenho em amostras grandes, já para amostras pequenas e com baixo número de variáveis, o *TBP* foi levemente liberal. Em amostras com tamanhos diferentes, o *TBPKL* e o *TNVMM* apresentaram altas taxas de erro tipo I, quando o número de variáveis foi maior. No entanto, o *TBP* proposto, apresentou taxas consideradas iguais ao α correspondente e ou bem próximas a ele. No que diz respeito ao erro tipo I, o *TBP* apresentou desempenho superior aos seus concorrentes.

Em relação ao poder, para amostras de mesmo tamanho, o *TBP* superou seus concorrentes apenas em amostras pequenas, $n_1 = n_2 = 8$, e com maior número de variáveis. Para amostras com tamanhos diferentes e também com maior número de variáveis, o *TBP* mostrou-se menos poderoso, mas foi considerado exato em relação ao erro tipo I, já seus concorrentes apresentaram-se fortemente liberais.

Nas situações em que as matrizes de covariâncias, das duas populações, forem não estruturadas e Σ_1 for múltipla de Σ_2 , recomenda-se o uso do teste proposto, *TBP*, nas seguintes situações: amostras pequenas, de mesmo tamanho e

com alto número de variáveis; e quando as amostras possuírem tamanhos diferentes e alto número de variáveis.

REFERÊNCIAS

- AMORIM, I. S. et al. Monte carlo based test for inferring about the unidimensionality of a Brazilian coffee sensory panel. **Food Quality and Preference**, Barking, v. 21, n. 3, p. 319–323, 2010.
- ANDERSON, T. W. **An introduction to multivariate statistical analysis**. New York: J. Wiley, 1971. 675 p.
- BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. 4. ed. Jaboticabal: FEALQ, 2006. 237 p.
- BARON, M. **Probability and statistics for computer scientists**. Boca Raton: Chapman & Hall; CRC, 2007. 413 p.
- BEHRENS, B. M. A contribution to error estimation with few observations. **Landwirtschaftliche Jahrbücher**, Champaign, v. 68, p. 807–837, 1929.
- 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, mar./abr. 1999.
- CECCHETTI, D. **Poder e taxas de erro tipo I de quatro critérios multivariados para o teste de igualdade de efeitos de tratamentos avaliados por meio do método de Monte Carlo**. 1999. 56 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, Lavras, 1999.
- CHERNICK, M. R. **Bootstrap methods: a guide for practitioners and researchers**. 2nd ed. Newtown: Wiley-Interscience, 2008. 369 p.
- CHRISTENSEN, W. F.; RENCHER, A. C. A comparison of type I error rates and power levels for seven solutions to the multivariate behrens-fisher problem. **Communications in Statistics - Simulation and Computation**, New York, v. 26, n. 4, p. 1251–1273, Apr. 1997.
- CIRILLO, M. A.; FERREIRA, D. F. Avaliação do problema behrens-fisher para p-dimensões via simulação Monte Carlo e Bootstrap. **Revista Matemática Estatística**, São Paulo, v. 22, n. 1, p. 71–94, 2004.
- COCHRAN, W. G.; COX, G. M. **Experimental designs**. 2nd ed. Singapore: J. Wiley, 1957. 640 p.

COSTA, S. F. **Introdução ilustrada à estatística (com muito humor)**. São Paulo: Harbra, 1992. 303 p.

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

DAVISON, A. C.; HINKLEY, D. V. **Bootstrap methods and their application**. Cambridge: Cambridge University, 2008. 582 p.

EFRON, B. Bootstrap methods: another look at the jackknife. **Annals of Statistics**, Hayward, v. 7, n. 1, p. 1–26, Jan. 1979.

EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. Boca Raton: Chapman & Hall, 1993. 436 p.

FERREIRA, D. F. **Estatística básica**. 2. ed. Lavras: UFLA, 2009. 664 p.

FERREIRA, D. F. **Estatística multivariada**. 2. ed. Lavras: UFLA, 2011. 662 p.

FISHER, R. A. The fiducial argument in statistical inference. **Annals of Eugenics**, London, v. 6, n. 3, p. 391–398, 1935.

GEBERT, D. M. P.; FERREIRA, D. F. Proposta de teste bootstrap não-paramétrico de retenção do número de componentes principais. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 2, p. 116–136, abr./jun. 2010.

GENTLE, J. E. **Random number generation and Monte Carlo methods**. 2nd ed. New York: Springer, 2003. 381 p. (Statistics and Computing).

GIVENS, G. H.; HOETING, J. A. **Computational statistics**. Hoboken: Wiley-Interscience, 2005. 418 p.

JAMES, G. S. Testes of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. **Biometrika**, London, v. 41, n. 1/2, p. 19–43, June 1954.

JOHANSEN, S. The welch-james approximation to the distribution of the residual sum of squares in a weighted linear regression. **Biometrika**, London, v. 67, p. 85–92, 1980.

JOHNSON, R. A.; WEERAHANDI, S. A bayesian solution to the multivariate behrens-fisher problem. **Journal of the American Statistical Association**, New York, v. 83, n. 401, p. 145–149, Mar. 1988.

- KIM, S. A practical solution to the multivariate behrens-fisher problem. **Biometrika**, London, v. 79, n. 1, p. 171–176, Mar. 1992.
- KRISHNAMOORTHY, K.; LU, F. A. A parametric bootstrap solution to the manova under heteroscedasticity. **Journal of Statistical Computation and Simulation**, Lafayette, v. 80, n. 8, p. 873–887, Aug. 2010.
- KRISHNAMOORTHY, K.; YU, J. Modified nel and van der merwe test for the multivariate behrens-fisher problem. **Statistics & Probability Letters**, Amsterdam, v. 66, n. 2, p. 161–169, Aug. 2004.
- KUNSCH, H. R. The jackknife and the bootstrap for general stationary observations. **Annals of Statistics**, Hayward, v. 17, n. 3, p. 1217–1241, Mar. 1989.
- LIX, L. M.; KESELMAN, H. J.; HINDS, A. M. Robust tests for the multivariate behrens-fisher problem. **Computer Methods and programs in Biomedicine**, Amsterdam, v. 77, n. 2, p. 129–139, Feb. 2005.
- MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. 2. ed. São Paulo: EDUSP, 2006. 428 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3rd ed. Singapore: McGraw-Hill, 1974. 564 p.
- NEL, D. G.; MERWE, C. A. V. D. A solution to the multivariate behrens-fisher problem. **Communication in Statistics: Theory and Methods**, New York, v. 15, n. 1, p. 3719–3735, 1986.
- OLIVEIRA, I. R. C.; FERREIRA, D. F. Multivariate extension of chi-squared univariate normality test. **Journal of Statistical Computation and Simulation**, New York, v. 80, n. 5, p. 513–526, 2010.
- OLIVEIRA, M. S. et al. **Introdução à estatística**. Lavras: UFLA, 2009. 334 p.
- R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna, 2013. Acesso em: 10 dez. 2013. Disponível em: <<http://www.R-project.org/>>.
- RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2000. 303 p.
- RAMOS, P. S.; FERREIRA, D. F. A bayesian solution to the multivariate Behrens fisher problem. **Computational Statistics & Data Analysis**, New York, v. 54, n. 6, p. 1622–1633, June 2012.

RENCHEER, A. C. **Methods of multivariate analysis**. 2nd ed. New York: J. Wiley, 2002. 708 p.

SATTHEERTHWAITTE, F. E. An approximate distribution of estimates of variance components. **Biometrics Bulletin**, Washington, v. 2, n. 6, p. 110–114, Dec. 1946.

SCHECHTMAN, E.; SHERMAN, M. The two-sample t -test with a known ratio of variances. **Statistical Methodology**, Oxford, v. 4, p. 508–514, Oct. 2007.

SILVA, E. C. **Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo**. 1998. 54 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, Lavras, 1998.

SMITH, W. B.; HOCKING, R. R. Algorithm AS 53: wishart variate generator. **Applied Statistics: Journal of the Royal Statistical Society - Série C**, London, v. 21, n. 3, p. 341–345, 1972.

STELAND, A.; PADMANABHAN, A. R.; AKRAN, M. Resampling methods for the nonparametric and generalized behrens-fisher problems. **Sankhya: The Indian Journal of Statistics**, Calcutta, v. 73, p. 267–302, Aug. 2011.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics Bulletin**, Washington, v. 1, n. 6, p. 80–83, Dec. 1945.

YAO, Y. An approximate degrees of freedom solution to the multivariate Behrens fisher problem. **Biometrika**, London, v. 52, p. 139–147, 1965.

ANEXOS

ANEXO A: Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = n_2 = 30$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 .

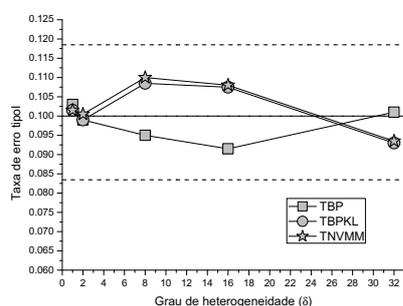
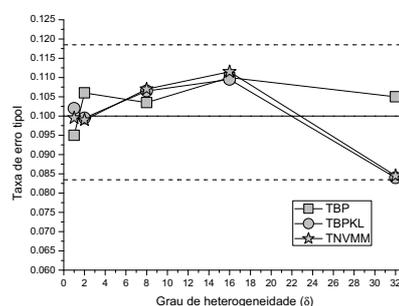
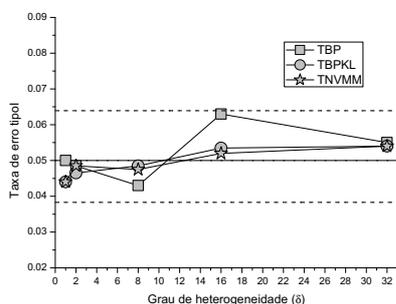
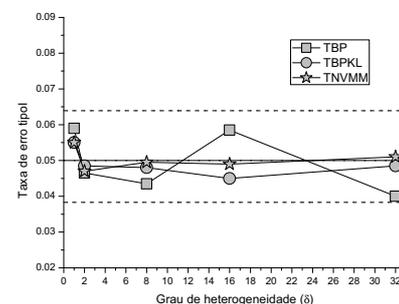
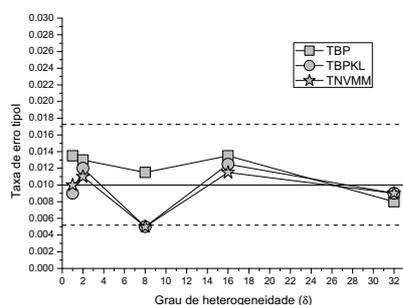
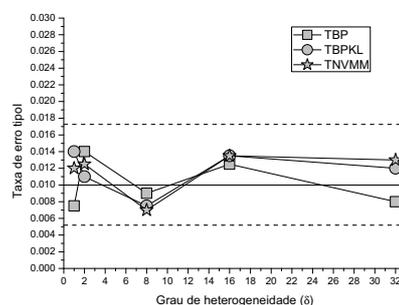
		$n_1 = 30$ e $n_2 = 30$			
		δ	TBP	TBPKL	TNVMM
$\alpha = 0,10$	p=7	1	0,0995	0,0995	0,0995
		2	0,1005	0,1090	0,1040
		8	0,1180	0,1115	0,1110
		16	0,1015	0,0985	0,0975
		32	0,0980	0,1025	0,1010
	p=29	1	0,0975	0,0935	0,0895
		2	0,1050	0,0845	0,0815*
		8	0,1115	0,0920	0,0910
		16	0,1025	0,0805*	0,0775*
		32	0,0975	0,0915	0,0885
$\alpha = 0,05$	p=7	1	0,0565	0,0520	0,0530
		2	0,0490	0,0495	0,0485
		8	0,0540	0,0485	0,0480
		16	0,0535	0,0520	0,0520
		32	0,0565	0,0490	0,0485
	p=29	1	0,0545	0,0475	0,0455
		2	0,0490	0,0465	0,0430
		8	0,0530	0,0445	0,0430
		16	0,0395	0,0415	0,0390
		32	0,0480	0,0470	0,0455
$\alpha = 0,01$	p=7	1	0,0135	0,0070	0,0070
		2	0,0125	0,0150	0,0145
		8	0,0130	0,0060	0,0060
		16	0,0135	0,0115	0,0140
		32	0,0130	0,0125	0,0110
	p=29	1	0,0095	0,0085	0,0070
		2	0,0145	0,0065	0,0070
		8	0,0130	0,0105	0,0115
		16	0,0145	0,0080	0,0065
		32	0,0135	0,0075	0,0070

+ O valor pode ser considerado estatisticamente diferente e maior que o (α) correspondente.

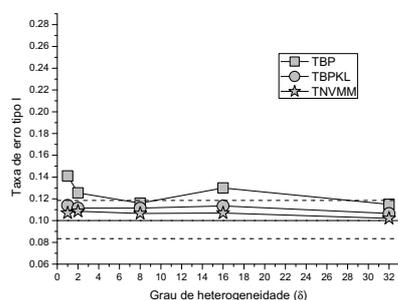
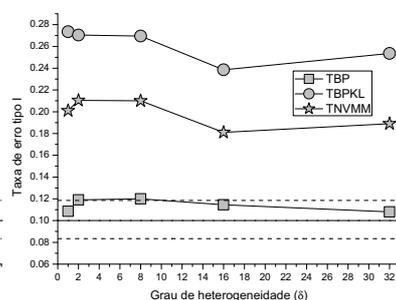
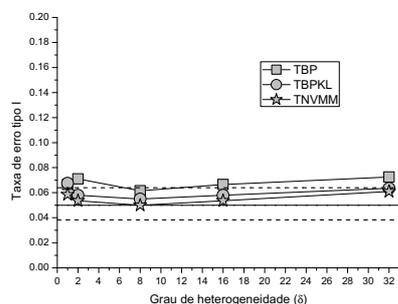
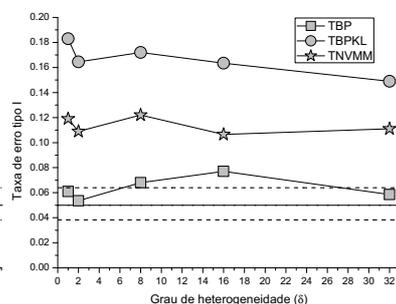
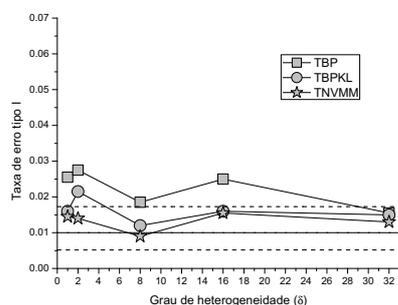
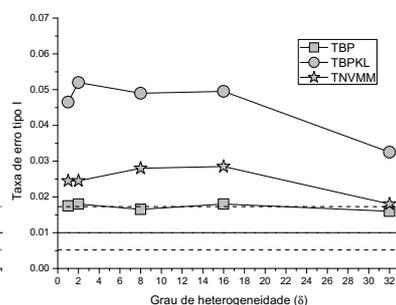
* O valor pode ser considerado estatisticamente diferente e menor que o (α) correspondente.

Coefficiente de confiança de 0,99.

ANEXO B: Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$ e $n_2 = 100$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99.

(a) $p = 7$ e $\alpha = 0,10$ (b) $p = 15$ e $\alpha = 0,10$ (c) $p = 7$ e $\alpha = 0,05$ (d) $p = 15$ e $\alpha = 0,05$ (e) $p = 7$ e $\alpha = 0,01$ (f) $p = 15$ e $\alpha = 0,01$

ANEXO C: Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$ e $n_2 = 100$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99.

(a) $p = 2$ e $\alpha = 0,10$ (b) $p = 7$ e $\alpha = 0,10$ (c) $p = 2$ e $\alpha = 0,05$ (d) $p = 7$ e $\alpha = 0,05$ (e) $p = 2$ e $\alpha = 0,01$ (f) $p = 7$ e $\alpha = 0,01$

ANEXO D: Taxas de erro tipo I do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$, $n_2 = 100$ e $n_1 = 100$, $n_2 = 30$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α , sob H_0 .

p	δ	$n_1 = 30$ e $n_2 = 100$			$n_1 = 100$ e $n_2 = 30$				
		TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM		
$\alpha=0,10$	p=7	1	0,0990	0,1090	0,1045	0,1045	0,1010	0,0965	
		2	0,0915	0,1025	0,1015	0,1105	0,1080	0,1005	
		8	0,1005	0,0970	0,0955	0,0965	0,0960	0,0945	
		16	0,1085	0,1105	0,1070	0,0955	0,1010	0,0995	
	32	0,0940	0,1020	0,1005	0,1050	0,1040	0,1000		
	p=29	1	0,1075	0,2415 ⁺	0,1960 ⁺	0,1175	0,2315 ⁺	0,1975 ⁺	
		2	0,0975	0,2290 ⁺	0,1945 ⁺	0,1020	0,2605 ⁺	0,2185 ⁺	
		8	0,1095	0,2395 ⁺	0,2055 ⁺	0,1030	0,2490 ⁺	0,1935 ⁺	
		16	0,1020	0,2230 ⁺	0,1800 ⁺	0,1045	0,2435 ⁺	0,2025 ⁺	
	32	0,1015	0,2265 ⁺	0,1925 ⁺	0,1160	0,2715 ⁺	0,2300 ⁺		
	$\alpha=0,05$	p=7	1	0,0565	0,0535	0,0505	0,0605	0,0525	0,0500
			2	0,0495	0,0690 ⁺	0,0645 ⁺	0,0465	0,0485	0,0445
8			0,0545	0,0530	0,0505	0,0500	0,0575	0,0570	
16			0,0620	0,0630	0,0610	0,0575	0,0525	0,0505	
32		0,0500	0,0570	0,0560	0,0610	0,0565	0,0520		
p=29		1	0,0635	0,1375 ⁺	0,1090 ⁺	0,0720 ⁺	0,1385 ⁺	0,1080 ⁺	
		2	0,0580	0,1380 ⁺	0,1110 ⁺	0,0460	0,1550 ⁺	0,1210 ⁺	
		8	0,0545	0,1380 ⁺	0,1075 ⁺	0,0425	0,1610 ⁺	0,1255 ⁺	
		16	0,0480	0,1325 ⁺	0,1040 ⁺	0,0515	0,1525 ⁺	0,1145 ⁺	
32		0,0475	0,1265 ⁺	0,0910 ⁺	0,0455	0,1560 ⁺	0,1140 ⁺		
$\alpha=0,01$		p=7	1	0,0145	0,0145	0,0160	0,0125	0,0170	0,0140
			2	0,0095	0,0105	0,0095	0,0170	0,0135	0,0135
	8		0,0165	0,0115	0,0105	0,0140	0,0150	0,0135	
	16		0,0140	0,0110	0,0100	0,0140	0,0105	0,0100	
	32	0,0135	0,0120	0,0100	0,0145	0,0160	0,0130		
	p=29	1	0,0150	0,0425 ⁺	0,0250 ⁺	0,0135	0,0355 ⁺	0,0245 ⁺	
		2	0,0135	0,0390 ⁺	0,0240 ⁺	0,0170	0,0450 ⁺	0,0335 ⁺	
		8	0,0095	0,0350 ⁺	0,0230 ⁺	0,0095	0,0425 ⁺	0,0295 ⁺	
		16	0,0105	0,0425 ⁺	0,0285 ⁺	0,0115	0,0415 ⁺	0,0285 ⁺	
	32	0,0155	0,0320 ⁺	0,0250 ⁺	0,0125	0,0445 ⁺	0,0270 ⁺		

+ O valor pode ser considerado estatisticamente diferente e maior que o (α) correspondente.

* O valor pode ser considerado estatisticamente diferente e menor que o (α) correspondente.

Coefficiente de confiança de 0,99.

ANEXO E: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = n_2 = 8$, $k = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 8$ e $n_2 = 8$			
		$k = 8$			
	δ	TBP	TBPKL	TNVMM	
	$\alpha = 0,10$	p=2	1	1	1
2			1	1	1
8			1	1	1
16			1	1	1
32			1	1	1
p=7		1	1	1	1
		2	1	1	1
		8	1	1	1
		16	1	1	1
		32	1	1	1
$\alpha = 0,05$	p=2	1	1	1	1
		2	1	1	1
		8	1	1	1
		16	1	1	1
		32	1	1	1
	p=7	1	1	0,9995	0,9995
		2	1	1	1
		8	1	0,9990	0,9980
		16	1	1	0,9995
		32	1	1	1
$\alpha = 0,01$	p=2	1	1	1	1
		2	1	1	1
		8	1	1	1
		16	1	1	0,9995
		32	1	1	1
	p=7	1	1	0,9850	0,9740
		2	1	0,9770	0,9705
		8	1	0,9700	0,9640
		16	1	0,9740	0,9625
		32	1	0,9730	0,9555

ANEXO F: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = n_2 = 30$, $k = 1$ e 2, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 30$ e $n_2 = 30$						
		$k = 1$			$k = 2$			
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM	
$\alpha = 0,10$	p=7	1	0,3650	0,1960	0,1960	0,8330	0,5140	0,5150
		2	0,3775	0,2060	0,2095	0,8340	0,5360	0,5320
		8	0,3885	0,2125	0,2130	0,8330	0,5200	0,5200
		16	0,3540	0,1780	0,1775	0,8275	0,5205	0,5225
		32	0,3665	0,1975	0,1970	0,8265	0,5020	0,5020
	p=29	1	0,3970	0,1160	0,1130	0,8660	0,2575	0,2455
		2	0,4010	0,1215	0,1205	0,8640	0,2550	0,2490
		8	0,3925	0,1320	0,1270	0,8565	0,2460	0,2420
		16	0,3945	0,1215	0,1195	0,8615	0,2465	0,2405
		32	0,4015	0,1345	0,1325	0,8770	0,2440	0,2405
$\alpha = 0,05$	p=7	1	0,2575	0,1025	0,1050	0,7410	0,3740	0,3735
		2	0,2575	0,1100	0,1090	0,7320	0,3945	0,3940
		8	0,2540	0,1090	0,1110	0,7405	0,3660	0,3705
		16	0,2540	0,1095	0,1090	0,7365	0,3930	0,3920
		32	0,2520	0,1135	0,1145	0,7370	0,3785	0,3750
	p=29	1	0,2970	0,0665	0,0670	0,7845	0,1400	0,1375
		2	0,2885	0,0605	0,0605	0,7855	0,1645	0,1540
		8	0,2825	0,0635	0,0620	0,7715	0,1380	0,1320
		16	0,2735	0,0615	0,0565	0,7925	0,1420	0,1360
		32	0,2645	0,0705	0,0685	0,7790	0,1380	0,1345
$\alpha = 0,01$	p=7	1	0,1120	0,0390	0,0390	0,5105	0,1595	0,1565
		2	0,1190	0,0375	0,0360	0,5070	0,1585	0,1590
		8	0,1090	0,0305	0,0300	0,5210	0,1700	0,1700
		16	0,1065	0,0280	0,0295	0,5265	0,1630	0,1630
		32	0,1040	0,0295	0,0325	0,5115	0,1695	0,1705
	p=29	1	0,1270	0,0150	0,0145	0,5625	0,0295	0,0295
		2	0,1110	0,0180	0,0170	0,5780	0,0350	0,0335
		8	0,1280	0,0125	0,0110	0,5770	0,0305	0,0265
		16	0,1150	0,0140	0,0140	0,5615	0,0425	0,0400
		32	0,1215	0,0160	0,0150	0,5735	0,0445	0,0450

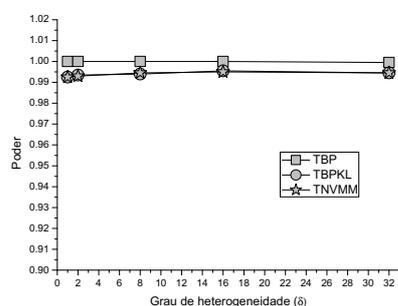
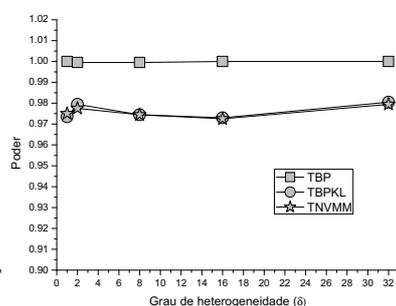
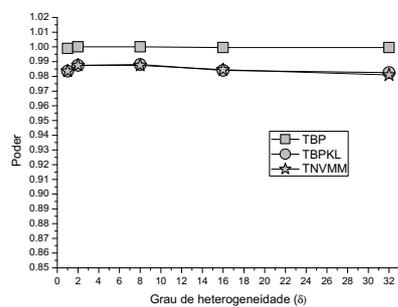
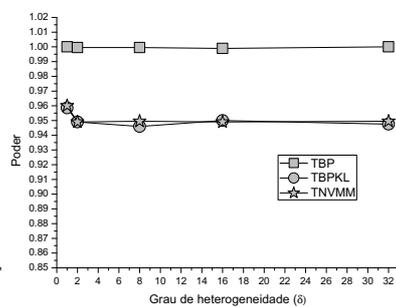
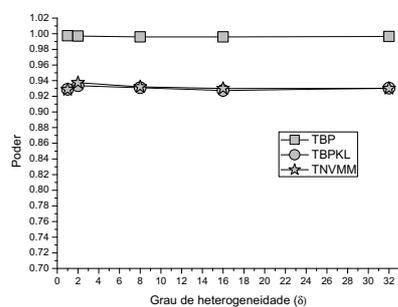
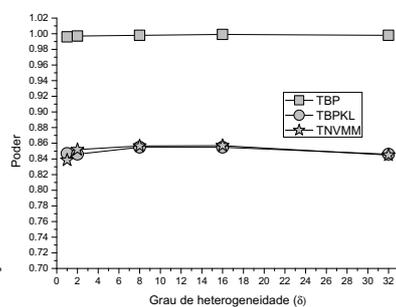
ANEXO G: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = n_2 = 30$, $k = 4$ e 8, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 30$ e $n_2 = 30$						
		$k = 4$			$k = 8$			
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM	
$\alpha = 0,10$	p=7	1	1,0000	0,9905	0,9910	1,0000	1,0000	1,0000
		2	0,9995	0,9880	0,9895	1,0000	1,0000	1,0000
		8	0,9995	0,9850	0,9860	1,0000	1,0000	1,0000
		16	1,0000	0,9880	0,9865	1,0000	1,0000	1,0000
		32	1,0000	0,9845	0,9845	1,0000	1,0000	1,0000
	p=29	1	0,9995	0,7630	0,7560	1,0000	1,0000	1,0000
		2	1,0000	0,7515	0,7475	1,0000	1,0000	1,0000
		8	0,9995	0,7560	0,7525	1,0000	1,0000	1,0000
		16	0,9995	0,7630	0,7545	1,0000	1,0000	1,0000
		32	1,0000	0,7410	0,7350	1,0000	1,0000	1,0000
$\alpha = 0,05$	p=7	1	1,0000	0,9735	0,9740	1,0000	1,0000	1,0000
		2	1,0000	0,9660	0,9650	1,0000	1,0000	1,0000
		8	0,9995	0,9725	0,9750	1,0000	1,0000	1,0000
		16	1,0000	0,9790	0,9790	1,0000	1,0000	1,0000
		32	1,0000	0,9730	0,9735	1,0000	1,0000	1,0000
	p=29	1	0,9995	0,6055	0,6030	1,0000	1,0000	1,0000
		2	0,9995	0,6340	0,6270	1,0000	0,9995	0,9995
		8	0,9995	0,6150	0,6080	1,0000	1,0000	1,0000
		16	1,0000	0,6090	0,5990	1,0000	0,9995	1,0000
		32	0,9995	0,6005	0,5905	1,0000	0,9995	0,9995
$\alpha = 0,01$	p=7	1	0,9960	0,8925	0,8960	1,0000	1,0000	1,0000
		2	0,9960	0,8810	0,8840	1,0000	1,0000	1,0000
		8	0,9950	0,8705	0,8785	1,0000	1,0000	1,0000
		16	0,9925	0,8705	0,8730	1,0000	1,0000	1,0000
		32	0,9945	0,8755	0,8790	1,0000	1,0000	1,0000
	p=29	1	0,9965	0,3305	0,3180	1,0000	0,9975	0,9970
		2	0,9980	0,3120	0,3060	1,0000	0,9965	0,9965
		8	0,9985	0,3160	0,3055	1,0000	0,9970	0,9970
		16	0,9980	0,3370	0,3295	1,0000	0,9975	0,9975
		32	0,9960	0,3255	0,3170	1,0000	0,9980	0,9980

ANEXO H: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = n_2 = 100$, $k = 1$ e 2 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 100$ e $n_2 = 100$						
		$k = 1$			$k = 2$			
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM	
$\alpha = 0,10$	p=7	1	0,3560	0,1975	0,1945	0,8315	0,5580	0,5605
		2	0,3725	0,2095	0,2095	0,8340	0,5455	0,5425
		8	0,3810	0,1940	0,1935	0,8275	0,5490	0,5505
		16	0,3830	0,2140	0,2105	0,8290	0,5500	0,5495
		32	0,3520	0,2020	0,1995	0,8235	0,5485	0,5445
	p=15	1	0,3870	0,1865	0,1870	0,8600	0,4425	0,4445
		2	0,3975	0,1875	0,1865	0,8595	0,4505	0,4500
		8	0,3860	0,1670	0,1680	0,8690	0,4545	0,4565
		16	0,3980	0,1605	0,1605	0,8630	0,4455	0,4440
		32	0,3735	0,1855	0,1855	0,8490	0,4235	0,4215
$\alpha = 0,05$	p=7	1	0,2365	0,1230	0,1225	0,7420	0,4175	0,4200
		2	0,2490	0,1045	0,1060	0,7335	0,4270	0,4300
		8	0,2650	0,1135	0,1160	0,7500	0,4125	0,4135
		16	0,2495	0,1150	0,1170	0,7540	0,4290	0,4305
		32	0,2460	0,1140	0,1150	0,7400	0,4280	0,4230
	p=15	1	0,2785	0,0930	0,0940	0,7570	0,2990	0,3010
		2	0,2700	0,0845	0,0865	0,7905	0,3090	0,3065
		8	0,2865	0,1065	0,1035	0,7835	0,3225	0,3255
		16	0,2840	0,0945	0,0960	0,7830	0,3085	0,3095
		32	0,2995	0,0945	0,0965	0,7540	0,3155	0,3175
$\alpha = 0,01$	p=7	1	0,1075	0,0410	0,0415	0,5270	0,2125	0,2125
		2	0,1020	0,0305	0,0310	0,5105	0,2030	0,2015
		8	0,1100	0,0370	0,0375	0,5195	0,2100	0,2170
		16	0,1040	0,0345	0,0320	0,5020	0,1940	0,1975
		32	0,0960	0,0325	0,0335	0,5175	0,2000	0,2040
	p=15	1	0,1175	0,0255	0,0250	0,5490	0,1415	0,1435
		2	0,1165	0,0240	0,0240	0,5600	0,1390	0,1360
		8	0,1250	0,0250	0,0275	0,5570	0,1250	0,1270
		16	0,1100	0,0190	0,0180	0,5700	0,1225	0,1275
		32	0,1205	0,0220	0,0230	0,5775	0,1290	0,1250

ANEXO I: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 100$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99.

(a) $p = 7$ e $\alpha = 0,10$ (b) $p = 15$ e $\alpha = 0,10$ (c) $p = 7$ e $\alpha = 0,05$ (d) $p = 15$ e $\alpha = 0,05$ (e) $p = 7$ e $\alpha = 0,01$ (f) $p = 15$ e $\alpha = 0,01$

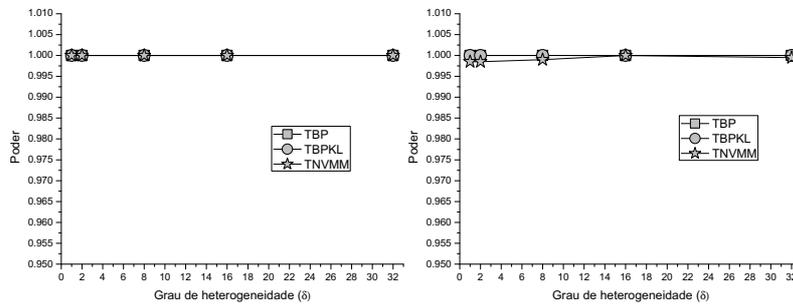
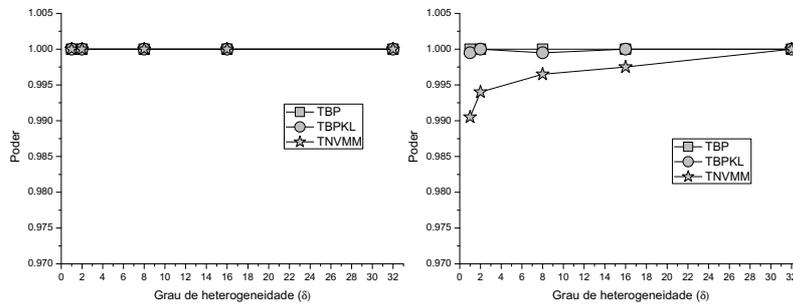
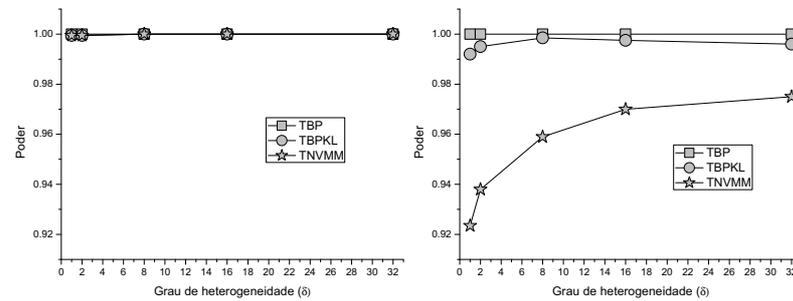
ANEXO J: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$ e $n_2 = 8$, $k = 1$ e 2 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 30$ e $n_2 = 8$							
		$k = 1$			$k = 2$				
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM		
$\alpha = 0,10$	p=2	1	0,3245	0,2510	0,2400	0,7355	0,5890	0,5795	
		2	0,3040	0,2270	0,2180	0,7160	0,5615	0,5530	
		8	0,3230	0,2265	0,2215	0,7215	0,5430	0,5295	
		16	0,3325	0,2360	0,2285	0,7180	0,5345	0,5235	
		32	0,3225	0,2190	0,2130	0,7140	0,5240	0,5175	
	p=7	1	0,3815	0,2775	0,2390	0,8105	0,5560	0,5035	
		2	0,3590	0,2840	0,2400	0,8140	0,5680	0,5050	
		8	0,3770	0,3035	0,2525	0,8125	0,5480	0,4875	
		16	0,3660	0,3215	0,2655	0,8060	0,5755	0,5160	
		32	0,3825	0,3430	0,2885	0,7835	0,5270	0,4510	
	$\alpha = 0,05$	p=2	1	0,2085	0,1330	0,1245	0,6120	0,4350	0,4195
			2	0,2160	0,1385	0,1315	0,6025	0,4245	0,4045
8			0,2185	0,1375	0,1270	0,6220	0,4225	0,4040	
16			0,2445	0,1490	0,1405	0,6240	0,3965	0,3825	
32			0,2105	0,1275	0,1205	0,6285	0,3900	0,3770	
p=7		1	0,2875	0,1860	0,1585	0,7280	0,3785	0,3210	
		2	0,2565	0,1855	0,1540	0,7255	0,3980	0,3395	
		8	0,2650	0,1790	0,1430	0,7080	0,4150	0,3395	
		16	0,2755	0,1970	0,1510	0,7335	0,4090	0,3345	
		32	0,2505	0,1945	0,1495	0,7160	0,4410	0,3565	
$\alpha = 0,01$		p=2	1	0,1065	0,0515	0,0465	0,3920	0,2060	0,1895
			2	0,0980	0,0510	0,0455	0,3870	0,1950	0,1710
	8		0,1045	0,0410	0,0335	0,4115	0,1925	0,1635	
	16		0,1035	0,0450	0,0375	0,4100	0,1725	0,1470	
	32		0,1265	0,0390	0,0355	0,4170	0,1745	0,1480	
	p=7	1	0,1190	0,0470	0,0305	0,5120	0,1805	0,1340	
		2	0,1210	0,0545	0,0395	0,5010	0,1795	0,1360	
		8	0,1195	0,0555	0,0420	0,5075	0,1805	0,1315	
		16	0,1350	0,0655	0,0430	0,5185	0,2105	0,1460	
		32	0,1195	0,0535	0,0330	0,5075	0,1880	0,1330	

ANEXO K: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$ e $n_2 = 8$, $k = 4$ e 8 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 30$ e $n_2 = 8$							
		$k = 4$			$k = 8$				
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM		
$\alpha = 0,10$	p=2	1	0,9970	0,9860	0,9855	1,0000	1,0000	1,0000	
		2	0,9955	0,9830	0,9810	1,0000	1,0000	1,0000	
		8	0,9965	0,9765	0,9745	1,0000	1,0000	1,0000	
		16	0,9955	0,9700	0,9690	1,0000	1,0000	1,0000	
		32	0,9975	0,9720	0,9715	1,0000	1,0000	1,0000	
	p=7	1	0,9995	0,9505	0,9290	1,0000	1,0000	1,0000	
		2	1,0000	0,9590	0,9385	1,0000	1,0000	1,0000	
		8	1,0000	0,9530	0,9185	1,0000	1,0000	1,0000	
		16	0,9990	0,9540	0,9195	1,0000	1,0000	1,0000	
		32	0,9985	0,9545	0,9230	1,0000	1,0000	1,0000	
	$\alpha = 0,05$	p=2	1	0,9920	0,9530	0,9465	1,0000	1,0000	1,0000
			2	0,9940	0,9440	0,9380	1,0000	1,0000	1,0000
			8	0,9890	0,9225	0,9160	1,0000	1,0000	1,0000
			16	0,9890	0,9125	0,9025	1,0000	1,0000	1,0000
32			0,9940	0,9205	0,9120	1,0000	1,0000	1,0000	
p=7		1	0,9990	0,9135	0,8755	1,0000	1,0000	0,9995	
		2	0,9990	0,9045	0,8635	1,0000	1,0000	1,0000	
		8	0,9980	0,9050	0,8585	1,0000	1,0000	0,9995	
		16	0,9975	0,9005	0,8455	1,0000	1,0000	0,9985	
		32	0,9995	0,8980	0,8350	1,0000	1,0000	0,9985	
$\alpha = 0,01$		p=2	1	0,9630	0,8130	0,7835	1,0000	1,0000	1,0000
			2	0,9570	0,7910	0,7560	1,0000	1,0000	1,0000
			8	0,9575	0,7335	0,6970	1,0000	0,9995	1,0000
			16	0,9605	0,6965	0,6560	1,0000	0,9980	0,9985
	32		0,9555	0,6570	0,6270	1,0000	0,9965	0,9955	
	p=7	1	0,9890	0,7160	0,6220	1,0000	0,9995	0,9970	
		2	0,9895	0,7075	0,6085	1,0000	0,9985	0,9925	
		8	0,9905	0,7055	0,5875	1,0000	0,9995	0,9840	
		16	0,9870	0,7065	0,5800	1,0000	0,9960	0,9845	
		32	0,9835	0,7115	0,5795	1,0000	0,9965	0,9795	

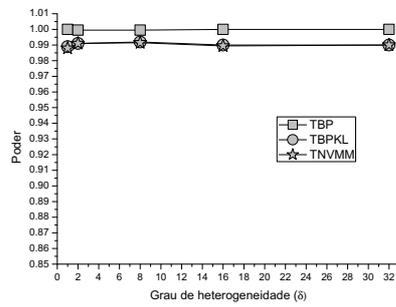
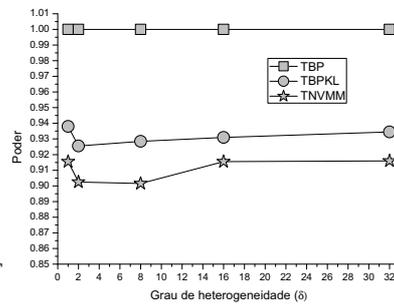
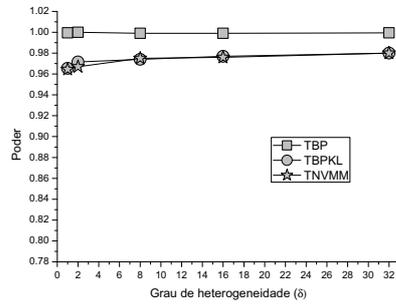
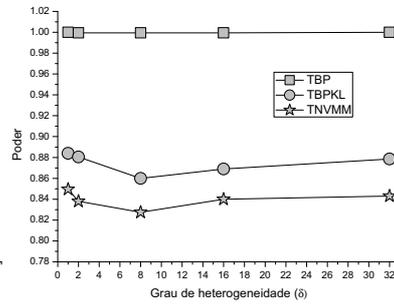
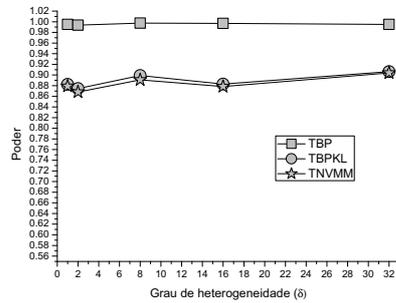
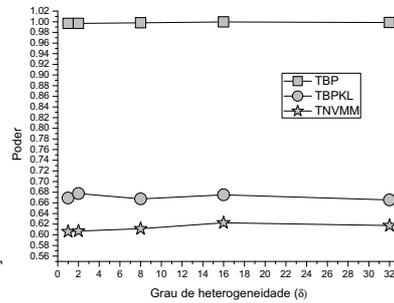
ANEXO L: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 8$, $n_2 = 100$, $k = 8$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99.

(a) $p = 2$ e $\alpha = 0,10$ (b) $p = 7$ e $\alpha = 0,10$ (c) $p = 2$ e $\alpha = 0,05$ (d) $p = 7$ e $\alpha = 0,05$ (e) $p = 2$ e $\alpha = 0,01$ (f) $p = 7$ e $\alpha = 0,01$

ANEXO M: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$ e $n_2 = 100$, $k = 1$ e 2 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 30$ e $n_2 = 100$						
		$k = 1$			$k = 2$			
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM	
$\alpha = 0,10$	p=7	1	0,3630	0,1885	0,1860	0,8235	0,5075	0,5020
		2	0,3585	0,2035	0,1960	0,8330	0,5395	0,5340
		8	0,3705	0,2070	0,2030	0,8310	0,5295	0,5175
		16	0,3580	0,2005	0,1960	0,8285	0,5220	0,5165
		32	0,3680	0,2045	0,2030	0,8320	0,5355	0,5300
	p=29	1	0,4030	0,3040	0,2515	0,8720	0,5015	0,4550
		2	0,3885	0,2930	0,2525	0,8690	0,5010	0,4390
		8	0,4290	0,2895	0,2550	0,8615	0,4910	0,4410
		16	0,3895	0,2890	0,2405	0,8650	0,4950	0,4360
		32	0,3825	0,2780	0,2285	0,8565	0,4865	0,4465
$\alpha = 0,05$	p=7	1	0,2595	0,1205	0,1145	0,7525	0,3975	0,3905
		2	0,2640	0,1190	0,1140	0,7440	0,4195	0,4090
		8	0,2500	0,1045	0,1000	0,7345	0,3955	0,3890
		16	0,2695	0,1225	0,1185	0,7455	0,3910	0,3875
		32	0,2625	0,1290	0,1265	0,7395	0,3950	0,3915
	p=29	1	0,2885	0,1905	0,1535	0,7945	0,3525	0,2950
		2	0,2770	0,1905	0,1485	0,7865	0,3435	0,2960
		8	0,2750	0,1855	0,1540	0,7875	0,3455	0,2910
		16	0,2870	0,1805	0,1510	0,7745	0,3260	0,2780
		32	0,2780	0,1845	0,1540	0,7860	0,3720	0,3225
$\alpha = 0,01$	p=7	1	0,1020	0,0300	0,0250	0,5220	0,1755	0,1700
		2	0,1200	0,0380	0,0365	0,5310	0,1995	0,1920
		8	0,0985	0,0365	0,0305	0,5075	0,1750	0,1620
		16	0,0900	0,0285	0,0255	0,5190	0,1810	0,1765
		32	0,1030	0,0335	0,0290	0,5265	0,1860	0,1765
	p=29	1	0,1260	0,0575	0,0400	0,5745	0,1455	0,1060
		2	0,1225	0,0560	0,0410	0,5705	0,1410	0,1115
		8	0,1215	0,0530	0,0370	0,5555	0,1465	0,1140
		16	0,1350	0,0595	0,0450	0,5760	0,1350	0,1130
		32	0,1170	0,0515	0,0370	0,5610	0,1275	0,1055

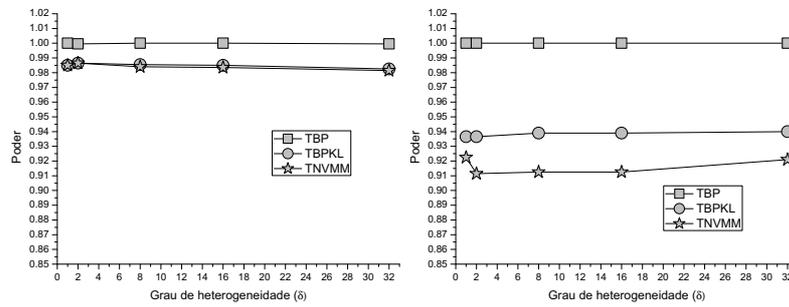
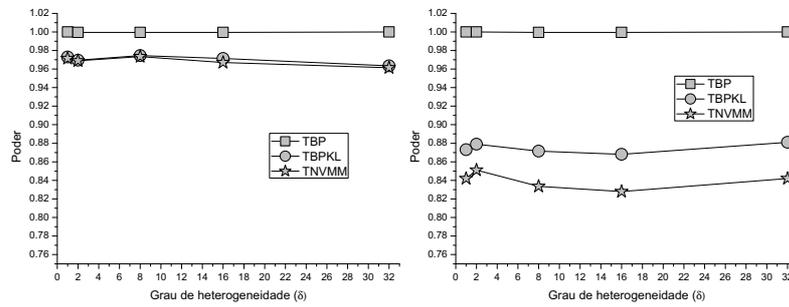
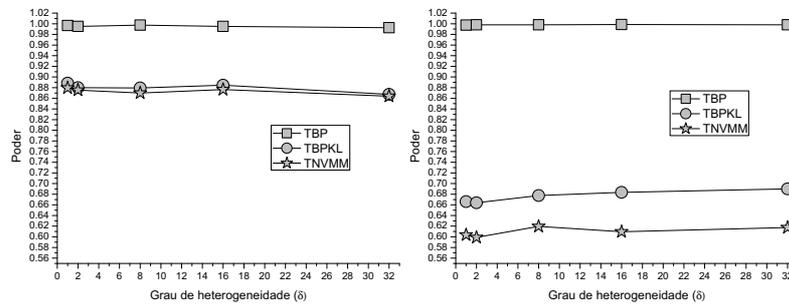
ANEXO N: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 30$, $n_2 = 100$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99.

(a) $p = 7$ e $\alpha = 0,10$ (b) $p = 29$ e $\alpha = 0,10$ (c) $p = 7$ e $\alpha = 0,05$ (d) $p = 29$ e $\alpha = 0,05$ (e) $p = 7$ e $\alpha = 0,01$ (f) $p = 29$ e $\alpha = 0,01$

ANEXO O: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$ e $n_2 = 30$, $k = 1$ e 2 , grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α .

		$n_1 = 100$ e $n_2 = 30$							
		$k = 1$			$k = 2$				
	δ	TBP	TBPKL	TNVMM	TBP	TBPKL	TNVMM		
$\alpha = 0,10$	p=7	1	0,3420	0,2005	0,1945	0,8395	0,5330	0,5275	
		2	0,3830	0,2200	0,2115	0,8445	0,5385	0,5325	
		8	0,3575	0,1890	0,1835	0,8240	0,5035	0,4980	
		16	0,3540	0,1910	0,1840	0,8395	0,5130	0,5060	
		32	0,3480	0,1935	0,1870	0,8210	0,5160	0,5075	
	p=29	1	0,3980	0,3135	0,2690	0,8770	0,4945	0,4415	
		2	0,3915	0,3145	0,2685	0,8785	0,4845	0,4245	
		8	0,3940	0,3085	0,2560	0,8640	0,5055	0,4440	
		16	0,4015	0,3085	0,2530	0,8730	0,5380	0,4675	
		32	0,3915	0,3170	0,2580	0,8650	0,5205	0,4575	
	$\alpha = 0,05$	p=7	1	0,2780	0,1165	0,1120	0,7285	0,3825	0,3720
			2	0,2425	0,1115	0,1070	0,7345	0,3910	0,3865
8			0,2490	0,1235	0,1160	0,7505	0,3795	0,3680	
16			0,2810	0,1280	0,1230	0,7405	0,3955	0,3885	
32			0,2635	0,1360	0,1305	0,7355	0,3740	0,3635	
p=29		1	0,2960	0,1795	0,1465	0,7925	0,3750	0,3150	
		2	0,2735	0,1815	0,1510	0,8085	0,3625	0,3045	
		8	0,2845	0,2185	0,1700	0,7900	0,3725	0,3090	
		16	0,2760	0,1910	0,1530	0,7815	0,3775	0,3210	
		32	0,2830	0,2165	0,1720	0,7935	0,3685	0,3080	
$\alpha = 0,01$		p=7	1	0,1125	0,0320	0,0290	0,5010	0,1760	0,1670
			2	0,0985	0,0330	0,0300	0,5140	0,1770	0,1695
	8		0,1015	0,0280	0,0235	0,5195	0,1890	0,1760	
	16		0,0945	0,0355	0,0350	0,5255	0,1700	0,1615	
	32		0,1140	0,0370	0,0360	0,5220	0,1735	0,1605	
	p=29	1	0,1320	0,0580	0,0400	0,5795	0,1515	0,1195	
		2	0,1100	0,0555	0,0395	0,5690	0,1305	0,0930	
		8	0,1225	0,0680	0,0505	0,5710	0,1605	0,1195	
		16	0,1210	0,0685	0,0505	0,5795	0,1545	0,1135	
		32	0,1095	0,0645	0,0495	0,5730	0,1435	0,1080	

ANEXO P: Poder do teste *bootstrap* paramétrico proposto (TBP), do teste *bootstrap* paramétrico de Krishnamoorthy e Lu (TBPKL) e do teste de Nel e Merwe modificado (TNVMM), em função dos tamanhos amostrais, $n_1 = 100$, $n_2 = 30$, $k = 4$, grau de heterogeneidade das covariâncias (δ), número de variáveis (p) e nível nominal de significância α . As linhas pontilhadas indicam os limites em que os testes das hipóteses $H_0 : \alpha = 0,10$; $H_0 : \alpha = 0,05$ e $H_0 : \alpha = 0,01$ devem ser rejeitadas, considerando um coeficiente de confiança de 0,99.

(a) $p = 7$ e $\alpha = 0,10$ (b) $p = 29$ e $\alpha = 0,10$ (c) $p = 7$ e $\alpha = 0,05$ (d) $p = 29$ e $\alpha = 0,05$ (e) $p = 7$ e $\alpha = 0,01$ (f) $p = 29$ e $\alpha = 0,01$

ANEXO Q: Rotina com os comandos do software R para aplicação do teste de Nel e Merwe modificado - TNVMM.

```

# Função BFNVM: recebe a amostra da população 1 em "X"
# e da população 2 em "Y".
# Aplica o teste de Nel e Merwe (1986) modificado por
# Krishnamoorthy e Yu (2004).

BFNVMM <- function(X,Y,delta0=matrix(0,ncol(X),1))
{
  n1 <- nrow(X); n2 = nrow(Y)
  p <- ncol(X)
  if (n1>n2)
  {
    aux1 <- X
    X <- Y
    Y <- aux1
    delta <- -1*delta
    n1 <- nrow(X); n2 <- nrow(Y)
  } # evitar trocas de dimensões n1>n2
  S1<-var(X); S2<-var(Y)
  Xb <- apply(X,2,mean); Yb <- apply(Y,2,mean)
  SeI<-solve(S1/n1+S2/n2)
  aux<-Xb - Yb - delta0
  T2 <- t(aux)%*%SeI%*%aux
  V1 <- ((S1/n1)%*%SeI); V2 <- ((S2/n2)%*%SeI)
  v2 <- p + p^2
  aux <- (sum(diag(V1%*%V1)) +
          (sum(diag(V1)))^2)/(n1-1)
  aux <- aux + (sum(diag(V2%*%V2)) +
               (sum(diag(V2)))^2)/(n2-1)
  v2<-v2/aux
  v1 <- p
  F<-(v2+1-p)*T2/(v2*p)
  pr.F<-1 - pf(F,v1,v2+1-p)
  return(list(f1=v1,f2=v2+1-p,F = F,pr.F=pr.F,T2=T2,
             nu=v2))
}

```

ANEXO R: Rotina com os comandos do software R para aplicação do teste de Krishnamoorthy e Lu - TBPCL.

```

# Função rWISH: recebe o vetor de médias em "nu" e o
# parâmetro de covariação em "Sigma".
# Esta função gera uma Wishart (matriz de soma de
# quadrados e produtos), baseada no algoritmo de Smith
# e Hocking (1972).
# Matriz w_i, em que  $W_i \sim W_p(\text{nu}, \text{Sigma})$ .

rWISH <- function (nu, Sigma)
{
  p <- nrow(Sigma)
  df <- (nu + nu - p + 1) - (nu - p + 1):nu
  if (p > 1)
  {
    T <- diag(sqrt(rgamma(c(rep(1, p)), df/2, 1/2)))
    T[lower.tri(T)] <- rnorm((p * (p - 1)/2))
  } else T <- sqrt(rgamma(1, df/2, 1/2))
  S <- chol(Sigma) # fator de Cholesky
  w <- t(S) %*% T %*% t(T) %*% S
  return(w)
}

# Variáveis com distribuição normal multivariada foram
# geradas utilizando o pacote MASS do software R.
# A função
# mvrnorm(n = 1, mu, Sigma, tol = 1e-6, empirical = FALSE)
# foi usada. Para isso o pacote deve estar carregado.

library(MASS)

# Função postBootMean: recebe o tamanho da amostra em
# "ni" e a matriz de covariâncias em "si".
# Esta função gera um vetor ybar_{bi}, em que
#  $Ybar_{\{bi\}} \sim N_p(0, s_i/n_i)$ .

postBootMean <- function(ni, si)
{
  p <- nrow(si)
  ybi <- mvrnorm(1, rep(0, p), si/ni)
  return(ybi)
}

# Função T2: recebe o tamanho da amostra da população 1
# em "n1", o tamanho da amostra da população 2 em "n2",
# o vetor de médias da população 1 em "ybar1", o vetor
# de médias da população 2 em "ybar2", a matriz de

```

```

# covariâncias da população 1 em "s1" e a matriz de
# covariâncias da população 2 em "s2".
# Esta função computa a estatística do teste de
# Krishnamoorthy e Lu (2010).

T2 <- function(n1, n2, ybar1, ybar2, s1, s2)
{
  sp <- s1/n1 + s2/n2
  t2 <- t(ybar1-ybar2) %*% ginv(sp) %*% (ybar1-ybar2)
  return(t2)
}

# Função bootNormDist: recebe amostra multivariada da
# população 1 (n_1 x p) em "Y1", amostra multivariada da
# população 2 (n_2 x p) em "Y2" e número de simulações
# bootstrap em "NB".
# Esta função gera a distribuição nula para o teste
# bootstrap paramétrico de Krishnamoorthy e Lu (2010),
# e retorna, além da distribuição nula de bootstrap, a
# estatística original T2* e o p-valor.

bootNormDist <- function(Y1, Y2, NB = 1000)
{
  ybar1 <- apply(Y1, 2, mean)
  ybar2 <- apply(Y2, 2, mean)
  n1 <- nrow(Y1)
  n2 <- nrow(Y2)
  s1 <- var(Y1)
  s2 <- var(Y2)
  p <- ncol(Y1)
  if (p != ncol(Y2)) stop("Diferentes dimensões não são
    permitidas!")
  Q <- T2(n1, n2, ybar1, ybar2, s1, s2)
  for (i in 1:NB)
  {
    sb1 <- (rWISH(n1-1, s1)) / (n1 - 1)
    sb2 <- (rWISH(n2-1, s2)) / (n2 - 1)
    yb1 <- postBootMean(n1, s1)
    yb2 <- postBootMean(n2, s2)
    Q <- c(Q, T2(n1, n2, yb1, yb2, sb1, sb2))
  }
  qc <- Q[1]
  p.val <- length(Q[Q >= qc]) / (NB + 1)
  if (is.null(p.val)) p.val <- 0
  return(list(T2 = qc, p.val = p.val, Q=Q))
}

```

ANEXO S: Rotina com os comandos do software R para aplicação do teste *bootstrap* paramétrico proposto - TBP.

```

# Variáveis com distribuição normal multivariada foram
# geradas utilizando o pacote MASS do software R.
# A função
# mvrnorm(n = 1, mu, Sigma, tol = 1e-6, empirical = FALSE)
# foi usada. Para isso o pacote deve estar carregado.

library(MASS)

# Função postBootMean: recebe o tamanho da amostra em
# "ni" e a matriz de covariâncias em "si".
# Esta função gera um vetor ybar_{bi}, em que
# Ybar_{bi} ~ N_p(0, s_i/n_i).

postBootMean <- function(ni, si)
{
  p <- nrow(si)
  ybi <- mvrnorm(1, rep(0, p), si/ni)
  return(ybi)
}

# Função T2c: recebe o vetor de médias da população 1
# em "ybar1" e o vetor de médias da população 2 em
# "ybar2".
# Esta função computa a estatística do teste de
# bootstrap paramétrico proposto.

T2c<- function(ybar1, ybar2)
{
  t2 <- t(ybar1-ybar2) %*% (ybar1-ybar2)
  return(t2)
}

# Função bootNpDist: recebe amostra multivariada da
# população 1 (n_1 x p) em "Y1", amostra multivariada da
# população 2 (n_2 x p) em "Y2" e número de simulações
# bootstrap em "NB".
# Esta função gera a distribuição nula para o teste
# bootstrap paramétrico proposto e retorna, além da
# distribuição nula de bootstrap, a estatística original
# T2c e o p-valor.

bootNpDist <- function(Y1, Y2, NB = 1000)
{
  ybar1 <- apply(Y1, 2, mean)
  ybar2 <- apply(Y2, 2, mean)

```

```
n1 <- nrow(Y1)
n2 <- nrow(Y2)
s1 <- var(Y1)
s2 <- var(Y2)
p <- ncol(Y1)
if (p != ncol(Y2)) stop("Diferentes dimensões não são
    permitidas!")
Q <- T2c(ybar1, ybar2)
for (i in 1:NB)
{
  yb1 <- postBootMean(n1, s1)
  yb2 <- postBootMean(n2, s2)
  Q <- c(Q, T2c(yb1, yb2))
}
qc <- Q[1]
p.val <- length(Q[Q >= qc]) / (NB + 1)
if (is.null(p.val)) p.val <- 0
return(list(T2 = qc, p.val = p.val, Q=Q))
}
```

ANEXO T: Rotina com os comandos do software R para aplicação da função de avaliação das taxas de erro tipo I para os três testes em estudo.

```

# Função evalTypeIError: recebe o tamanho da amostra
# da população 1 em "n1", o tamanho da amostra da
# população 2 em "n2", o número de variáveis em "p",
# o nível de significância em "alpha", o grau de
# heterogeneidade das matrizes de covariâncias em
# "delta", o coeficiente de correlação em "rho", o
# número de simulações bootstrap em "NB" e número
# de simulações Monte Carlo em NSM.
# Esta função avalia o erro tipo I dos três testes em
# estudo e retorna as proporções de rejeição da
# hipótese nula, na configuração desejada.

evalTypeIError <- function(n1, n2, p, alpha=0.05,
delta = 1, rho=0.9, op = 1, df=10, NB=1000, NSM = 1000)
{
  if (op == 1)
  {
    Sigma1 <- diag(p) * (1-rho) + matrix(rho, p, p)
  } else Sigma1 <- RandomSigma(p, df) #unstructured
  Sigma2 <- delta * Sigma1
  res <- c(0,0,0)
  for (i in 1:NSM)
  {
    Y1 <- mvrnorm(n1, rep(0, p), Sigma1)
    Y2 <- mvrnorm(n2, rep(0, p), Sigma2)
    tc <- bootNpDist(Y1, Y2, NB)
    zc <- bootNormDist(Y1, Y2, NB)
    nvm<- BFNVM(Y1, Y2)
    if (tc$p.val <= alpha) res[1] <- res[1] + 1 / NSM
    if (zc$p.val <= alpha) res[2] <- res[2] + 1 / NSM
    if (nvm$pr.F <= alpha) res[3] <- res[3] + 1 / NSM
  }
  return(list(res=res, Sigma1 = Sigma1))
}

# Função para criar uma matriz de covariância
# não-estruturada para ser usada em simulação
# Monte Carlo com variáveis aleatórias multivariadas

RandomSigma <- function(p = 5, df = 10)
{
  T <- diag(sqrt(rgamma(c(rep(1, p)), df/2, 1/2)))
  T[lower.tri(T)] <- 0.4 * rnorm((p * (p - 1)/2))
  Sigma <- T %*% t(T)
}

```

```
    return(Sigma)
}
```

ANEXO U: Rotina com os comandos do software R para aplicação da função de avaliação do poder para os três testes em estudo.

```

# Função evalPower: recebe o tamanho da amostra
# da população 1 em "n1", o tamanho da amostra da
# população 2 em "n2", o número de variáveis em "p",
# a diferença entre médias em "k", o nível de
# significância em "alpha", o grau de heterogeneidade
# das matrizes de covariâncias em "delta",
# o coeficiente de correlação em "rho", o número de
# simulações bootstrap em "NB" e número de
# simulações Monte Carlo em NSM.
# Esta função avalia o poder dos três testes em
# estudo e retorna as proporções de rejeição da
# hipótese nula, na configuração desejada.

evalPower <- function(n1, n2, p, k = 0, alpha=0.05,
delta = 1, rho=0.9, op=1, df=10, NB=1000, NSM = 1000)
{
  if (op == 1) {
    Sigma1 <- diag(p) * (1-rho) + matrix(rho, p, p)
  } else Sigma1 <- RandomSigma(p, df) #unstructured
  Sigma2 <- delta * Sigma1
  mu1 <- rep(0, p)
  mu2 <- k * (diag(Sigma1 / n1) + diag(Sigma2 / n2))^0.5
  res <- c(0,0,0)
  for (i in 1:NSM)
  {
    Y1 <- mvrnorm(n1, mu1, Sigma1)
    Y2 <- mvrnorm(n2, mu2, Sigma2)
    tc <- bootNpDist(Y1, Y2, NB)
    zc <- bootNormDist(Y1, Y2, NB)
    nvm<- BFNVMM(Y1, Y2)
    if (tc$p.val <= alpha) res[1] <- res[1] + 1 / NSM
    if (zc$p.val <= alpha) res[2] <- res[2] + 1 / NSM
    if (nvm$pr.F <= alpha) res[3] <- res[3] + 1 / NSM
  }
  return(list(res=res, Sigma1 = Sigma1))
}

# Função para criar uma matriz de covariância
# não-estruturada para ser usada em simulação
# Monte Carlo com variáveis aleatórias multivariadas

RandomSigma <- function(p = 5, df = 10)
{

```

```
T <- diag(sqrt(rgamma(c(rep(1, p)), df/2, 1/2)))
T[lower.tri(T)] <- 0.4 * rnorm((p * (p - 1)/2))
Sigma <- T %*% t(T)
return(Sigma)
}
```
