

COMPARAÇÕES MÚLTIPLAS
MULTIVARIADAS VIA *BOOTSTRAP* E
PERMUTAÇÃO

ELISA NORBERTO FERREIRA SANTOS

2008

ELISA NORBERTO FERREIRA SANTOS

COMPARAÇÕES MÚLTIPLAS MULTIVARIADAS VIA
BOOTSTRAP E PERMUTAÇÃO

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

Orientador

Prof. Dr. Daniel Furtado Ferreira

LAVRAS
MINAS GERAIS-BRASIL
2008

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

Santos, Elisa Norberto Ferreira.

Comparações múltiplas multivariadas via *bootstrap* e permutação/ Elisa Norberto Ferreira Santos. - Lavras: UFLA, 2008.

50p. : il.

Dissertação (Mestrado) - Universidade Federal de Lavras, 2008.

Orientador: Daniel Furtado Ferreira.

Bibliografia.

1. Comparações múltiplas multivariadas. 2. Método Monte Carlo. 3. *Bootstrap*. I. Universidade Federal de Lavras. II. Título.

CDD-519.535

ELISA NORBERTO FERREIRA SANTOS

COMPARAÇÕES MÚLTIPLAS MULTIVARIADAS VIA
BOOTSTRAP E PERMUTAÇÃO

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

APROVADA em 11 de fevereiro de 2008

Prof. Dr. Marcelo Tavares UFU

Prof. Dr. Marcelo Ângelo Cirillo UFLA

Prof. Dr. Renato Ribeiro de Lima UFLA

Prof. Dr. Daniel Furtado Ferreira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS-BRASIL

*A Deus,
pela confiança e coragem.*

*Aos meus pais, Dehon e Neide,
Aos meus irmãos, Alex e Aline,
pelo amor incondicional e
incentivo
Ofereço.*

*Ao meu grande amor, Luciano, pela força,
paciência, com muito amor
Dedico.*

AGRADECIMENTOS

A Deus, pela companhia constante, em cada segundo da minha vida, por toda a ajuda concedida e pelo amor incondicional.

Ao meu esposo, Luciano, pela compreensão, carinho, apoio, paciência e por ter sonhado comigo a conquista deste título.

Aos meus pais, Dehon e Neide, minha eterna gratidão, pelo apoio incondicional, pela confiança, por todo amor e incentivo que sempre me deram em tudo que busquei realizar. Sei que é a vós que devo o fato de ser aquilo que sou hoje.

Aos meus irmãos, Alex e Aline, fiéis companheiros em todos os momentos da minha vida.

O meu mais profundo agradecimento é dirigido ao Prof. Dr. Daniel Furtado Ferreira, não só pela orientação deste trabalho, como pela confiança em mim depositada para a sua concretização e pelo entusiasmo, disponibilidade e amizade que sempre me dispensou.

À UFLA e ao DEX, pela oportunidade de realizar o Mestrado.

Ao CNPq, pela concessão da bolsa de estudos.

Aos professores do DEX, pelas condições para a realização do mestrado e aos funcionários do DEX, pela eficiência e atenção prestada.

Ao Prof. Dr. Marcelo Ângelo Cirillo, pelos conselhos e amizade.

A minha grande amiga Luciana, pelo incentivo e pela amizade.

Aos meus colegas do curso, em especial a Fabricia, companheira de estudos, pela amizade, companheirismo e pela troca de conhecimentos.

Aos meus avós, tias, tios, primos, primas, sogro, sogra, cunhadas, cunhado e, especialmente, para a minha sobrinha Débora e minha priminha Fabiane, que, nos momentos de “stress” foram o meu refúgio.

A todos que colaboraram direta ou indiretamente para a concretização deste sonho.

A todos, os meus sinceros agradecimentos.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	iii
RESUMO	iv
ABSTRACT	v
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Procedimentos de comparações múltiplas univariadas	3
2.2 Tipos de erro e poder do teste	8
2.3 Comparações múltiplas multivariadas.	10
2.4 Simulação Monte Carlo	12
2.5 Método <i>bootstrap</i> e permutação	13
2.6 Distribuição normal multivariada	14
2.7 Distância de Mahalanobis	15
3 METODOLOGIA	17
4 RESULTADOS E DISCUSSÃO	24
4.1 Taxas de erro tipo I	24
4.2 Poder	32
4.3 Exemplo real	35
5 CONCLUSÕES	41
6 REFERÊNCIAS BIBLIOGRÁFICAS	42
ANEXOS	47

LISTA DE TABELAS

4.1	Taxas de erro tipo I, por experimento, em porcentagem em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (<i>bootstrap</i>) e $N = 2.000$ simulações Monte Carlo, $k = 5$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, sem covariância comum.	26
4.2	Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (<i>bootstrap</i>) e $N = 1.000$ simulações Monte Carlo, $k = 5$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.	27
4.3	Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens sem reposição (permutação) e $N = 1.000$ simulações Monte Carlo, $k = 5$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.	29
4.4	Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (<i>bootstrap</i>) e $N = 1.000$ simulações Monte Carlo, $k = 10$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.	31
4.5	Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (<i>bootstrap</i>) e $N = 1.000$ simulações Monte Carlo, $k = 50$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.	32

4.6	Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (<i>bootstrap</i>) e $N = 1.000$ simulações Monte Carlo, $k = 100$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.	33
4.7	Teores de areia e argila referentes ao sistema de uso da terra na Amazônia	37
4.8	Vetores de médias dos sistema de uso da terra (SUTs) na Amazônia.	38
4.9	Resumo dos valores da estatística Γ_m <i>bootstrap</i> e Γ_m^* permutação, com $B=10.000$ reamostragens.	39
4.10	Valores da estatística do teste de comparações múltiplas multivariadas para cada par de vetores de médias dos SUTs. . . .	39
4.11	Valores- p para cada par de vetores de médias dos SUTs obtidos por <i>bootstrap</i> (acima da diagonal) e por permutação (abaixo da diagonal).	40

LISTA DE FIGURAS

4.1	Poder do teste <i>bootstrap</i> de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 5$, $r = 3$, $\rho = 0,5$, $p = 2$ e (a) $\delta = 1$ e (b) $\delta = 16$	34
4.2	Poder do teste <i>bootstrap</i> de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 5$, $r = 10$, $\rho = 0,5$, $p = 5$ e (a) $\delta = 1$ e (b) $\delta = 16$	35
4.3	Poder do teste <i>bootstrap</i> de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 100$, $r = 3$, $\rho = 0,5$, $p = 2$ e (a) $\delta = 1$ e (b) $\delta = 16$	36
4.4	Poder do teste <i>bootstrap</i> de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 100$, $r = 10$, $\rho = 0,5$, $p = 5$ e (a) $\delta = 1$ e (b) $\delta = 16$	36
4.5	Densidade da distribuição de (a) Γ_m via <i>bootstrap</i> e (b) Γ_m^* via permutação.	40

RESUMO

SANTOS, Elisa Norberto Ferreira. **Comparações múltiplas multivariadas via *bootstrap* e permutação**. 2008. 50 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.*

Quando o interesse é identificar os vetores de médias que diferem entre si, os contrastes lineares são o único método existente até o momento. Como não existe nenhum procedimento de comparações múltiplas multivariadas, o presente trabalho foi realizado com o objetivo de propor procedimentos de comparação múltipla para vetores multidimensionais via *bootstrap* e permutação. O procedimento foi desenvolvido e implementado por meio de métodos de computação intensiva e avaliado por meio de simulação Monte Carlo. A idéia foi obter, para todos os pares de médias, a distância T^2 de Hotelling e compará-las com as distâncias obtidas na distribuição *bootstrap* e permutação, originada da reamostragem. As taxas de erro tipo I por experimento e o poder do teste proposto foram computados para avaliar o seu desempenho em $N = 1000$ simulações Monte Carlo. Conclui-se que os testes foram propostos com sucesso e os passos necessários para a sua execução foram implementados em R. Nas situações homocedásticas, os testes tiveram desempenho considerado ótimo, com taxas de erro tipo I por experimento inferiores ou iguais aos valores nominais adotados quando $p \leq k(r - 1)$, sendo p o número de variáveis, k o número de populações e r o número de repetições. Os testes foram liberais nas situações heterocedásticas. O teste de permutação teve melhor desempenho no controle do erro tipo I por experimento quando $p = k(r - 1)$ ou $p \leq k(r - 1)$, mas está na vizinhança dessa desigualdade; o teste de comparações múltiplas multivariadas via *bootstrap* possui valores de poder comparáveis ou superiores aos dos testes univariados. O teste mostrou robustez no desempenho do poder quando a suposição de homocedasticidade for violada.

* **Comitê Orientador:** Daniel Furtado Ferreira - UFLA. (Orientador)

ABSTRACT

SANTOS, Elisa Norberto Ferreira. **Multivariate multiple comparisons using *bootstrapping* and permutation.** 2008. 50 p. Dissertation (Master in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, MG.*

For identifying difference among mean vectors, linear contrasts are the only existent method. For this reason the present work was accomplished for proposing a multiple comparison procedure for multidimensional vectors using *bootstrapping* and permutation. The procedure was developed and implemented using intensive computational methods and s evaluated by Monte Carlo simulations. The idea was to obtain the T^2 distances of Hotelling between pairs of the mean vectors and to compare them with the distances obtained in the bootstrapping and permutation distributions, originated by resampling methods. The experimentwise type I error rates and the power of the proposed test were computed to evaluate its performance in $N = 1000$ Monte Carlo simulations. The tests were successfully proposed and the necessary steps for their application were implemented in R; in the homoscedastic cases the tests achieve optimal performance with experimentwise type I error less or equal to the nominal significance levels when $p \leq k(r - 1)$, where p is the number of variables, k is the number of populations and r is the number of replications; the tests were liberal in heteroscedastic cases; the permutational test had optimal performance in the control of experimentwise type I error rates when $p = k(r - 1)$ or $p < k(r - 1)$ but considering only the neighborhood of this inequality; the multivariate multiple comparisons *bootstrapping* test had power that is comparable or superior to those of the univariate tests; the test was shown robust power performance when the homoscedastic supposition is violated.

***Guidance Committee:** Daniel Furtado Ferreira - UFLA. (Adviser)

1 INTRODUÇÃO

Análise multivariada é o ramo da Estatística que visa analisar dados obtidos quando diversas variáveis são medidas (ou avaliadas) numa mesma unidade experimental. O termo multivariado, ou multidimensional, é usado em virtude de, sob o ponto de vista matemático, cada variável poder ser vista como uma dimensão (Mardia, 1979).

Em muitas situações experimentais, deseja-se testar o efeito ou a eficácia de um tratamento. Para isso, medidas são tomadas nas unidades experimentais antes e após a aplicação do tratamento (Johnson & Wichern, 1998).

No caso univariado, a análise de variância e o teste F são usados para testar a hipótese de igualdade de efeito de tratamento. Se essa hipótese é rejeitada, é necessário identificar onde se encontram as diferenças. Para isso, empregam-se testes apropriados para as comparações das médias dos tratamentos.

Tratando da situação multivariada, avalia-se a hipótese de igualdade de efeitos vetoriais de tratamento por meio da análise de variância multivariada. Se a hipótese for rejeitada, é de interesse do pesquisador identificar os contrastes entre vetores de médias de tratamentos que são responsáveis pela rejeição da hipótese. Nesse caso, quando o interesse é comparar vários vetores de médias, e não os elementos dentro de um vetor, os contrastes lineares são o único método existente até o momento que ajuda a detalhar mais sobre a questão de comparações entre tratamentos. Mas, o uso de contraste ortogonais é limitado às situações em que os tratamentos são qualitativos e estruturados.

No contexto multivariado, há uma grande deficiência quando a hipótese nula é rejeitada, pois, na literatura, não existem testes de comparações múltiplas multivariados. Com isso, a metodologia deste trabalho é inédita e inovadora e tem como um dos objetivos sanar essa deficiência.

Por essas razões, o presente trabalho foi realizado com o objetivo de propor procedimentos de comparação múltipla para vetores multidimensio-

nais via *bootstrap* e permutação. Os procedimentos foram desenvolvidos e implementados por meio de métodos de computação intensiva e avaliados por meio de simulação Monte Carlo. A idéia é obter, para todos os pares de médias, a distância T^2 de Hotelling e compará-las com as distâncias obtidas na distribuição *bootstrap* e permutação, originada da reamostragem. As taxas de erro tipo I por experimento e o poder do teste proposto foram computados para avaliar o seu desempenho.

2 REFERENCIAL TEÓRICO

2.1 Procedimentos de comparações múltiplas univariadas

Os procedimentos de comparações múltiplas servem como um complemento do teste F para detectar possíveis diferenças entre os tratamentos (Banzatto & Kronka, 2006).

Quando os procedimentos de comparações múltiplas são utilizados de maneira incoerente, pode haver perda de informação e redução da eficiência se procedimentos mais apropriados forem avaliados (Petersen, 1977). O uso freqüentemente inapropriado de comparações múltiplas foi considerado, por Pearce (1993), como conseqüência do mau ensinamento e também como a resistência de não estatísticos em se aventurar no território desconhecido da especificação de contrastes.

Um caso freqüente de uso incorreto desses procedimentos diz respeito à análise de experimentos em esquema fatorial, em que cada nível de um fator é combinado com os demais níveis dos outros fatores (Petersen, 1977). Neste tipo de experimento, independente do tipo dos fatores (qualitativos ou quantitativos), é preciso testar primeiro a significância dos efeitos principais e das interações. Se não houve significância da interação, então, toda a informação se encontra nos efeitos principais e as médias de cada nível podem ser comparadas recorrendo-se a métodos mais adequados (Petersen, 1977). Se as médias dos efeitos principais forem comparadas sem se considerar a significância das interações, conclusões incorretas podem ser obtidas.

Sampaio (1998) discorda da afirmação de muitos autores de que, quando o fator é do tipo quantitativo, deve-se utilizar a regressão e não os procedimentos de comparações múltiplas. Esse autor argumenta que apenas uma técnica de análise pode comprometer os objetivos propostos e utilizar mais de uma técnica, concomitantemente, pode prover maiores informações.

Segundo Machado et al. (2005), a aplicação dos procedimentos de comparações múltiplas depende da natureza dos efeitos dos fatores em estudo. Quando os níveis destes fatores são quantitativos, é mais conveniente a utilização de uma metodologia de regressão e, se os fatores são qualitativos com estruturação, é mais apropriado estabelecer comparações entre os

níveis de um dos fatores por meio de contraste, seguidas de um teste específico. No entanto, se os níveis dos fatores são qualitativos e não-estruturados, devem-se aplicar os procedimentos de comparações múltiplas.

Apesar da facilidade de aplicação dos testes de comparações múltiplas, um aspecto que deve ser considerado quando se utilizam esses testes é a ambigüidade dos seus resultados. Esta ambigüidade decorre da possibilidade de que dois tratamentos, tidos como iguais a um terceiro, possam ser considerados diferentes entre si. Este aspecto pode dificultar a interpretação dos resultados, particularmente em presença de muitos tratamentos.

Os testes de comparações múltiplas mais utilizados na literatura são *t* de Student, Tukey, Student-Newman-Keuls (SNK), Duncan, Scheffé e outros que podem ser encontrados em vários livros voltados para a estatística experimental, como os de Pimentel Gomes (1985), Banzatto & Kronka (2006) e Steel & Torrie (1980), entre outros.

O teste *t* de Student, também denominado no contexto das comparações múltiplas de teste da diferença mínima significativa (*LSD* -Least Square Difference), é realizado por meio da estatística

$$LSD = t_{(\nu, \frac{\alpha}{2})} \sqrt{\frac{2QME}{r}}, \quad (2.1)$$

em que *t* é o quantil superior 100($\alpha/2$)% da distribuição *t* de Student, com ν graus de liberdade, QME é o quadrado médio do resíduo e *r* é o número de repetições.

Algumas desvantagens desse teste foram comentadas por Percin & Malheiros (1989). O nível de confiança (1 - α) não é válido para vários contrastes, mas sim para um apenas. Os contrastes devem ser previamente planejados e ortogonais entre si, e não sugeridos pelos dados.

De acordo com O'Neill & Wetherill (1971), uma alternativa ao teste *t* é a utilização da diferença significativa de Fisher. Este procedimento consiste em alterar o nível nominal de significância α para a determinação do valor tabelado de *t*, dividindo-se o nível nominal de significância pelo número de comparações (*p*) que serão realizadas. Este ajuste é também denominado de

correção de Bonferroni da significância nominal. Dessa forma, Ramalho et al. (2000) definiram o valor crítico do teste de Bonferroni da seguinte forma:

$$BLS D = t_{(\alpha/(2p),\nu)} \sqrt{\frac{2QME}{r}}, \quad (2.2)$$

em que t é o quantil superior $100(\alpha/(2p))\%$ da distribuição t de Student, com ν graus de liberdade do resíduo, QME é o quadrado médio do resíduo e r é o número de repetições de cada tratamento e p é o número de comparações.

O teste *BLS D* controla o erro tipo I por comparação e por experimento com limites máximos das taxas iguais ao valor nominal (Machado et al., 2005).

Tukey (1951), segundo Duncan (1955) e Harter (1957), propôs um método de amplitude fixa, baseado na distribuição de amplitude total estudentizada. Esse método requer que todos os níveis de tratamento tenham o mesmo número de repetições e que as inferências de interesse sejam todo o conjunto de comparações duas a duas, ou seja, todos os pares de estimativas (Ramalho et al., 2000). A diferença mínima significativa (Δ) do teste é dada por:

$$\Delta = q_{(\alpha,k,\nu)} \sqrt{\frac{QME}{r}}, \quad (2.3)$$

em que q é o quantil superior $100\alpha\%$ da amplitude estudentizada, no qual k representa o número de populações, ν o número de graus de liberdade do resíduo e α o valor nominal de significância; QME é o quadrado médio do resíduo e r é o número de repetições de cada tratamento.

Se k for maior que 2, o valor de Δ será maior do que o valor *LSD*, o que torna o teste de Tukey mais conservativo do que o teste t em declarar uma diferença significativa (Carmer & Swanson, 1973).

Segundo Perecin & Malheiros (1989), o teste de Tukey é válido para a totalidade dos contrastes de duas médias e que, em princípio, exige balanceamento, sendo exato para testar a maior diferença, porém, nos demais casos é considerado conservativo. O poder deste teste é baixo e apresenta uma redução drástica com o aumento do número de níveis dos tratamentos (Ramalho et al., 2000).

Para o caso em que os tratamentos são não balanceados, é recomendável, no lugar de r , usar a média harmônica, r_h , do número de repetições de cada população, dada por:

$$r_h = \frac{1}{\frac{1}{k} \sum_{i=1}^k \frac{1}{r_i}}, \quad (2.4)$$

em que r_i é o número de repetições de cada população i .

Enquanto os testes t de Student e de Tukey, incluindo suas variações, requerem o cálculo de um simples valor crítico, o teste de Student-Newman-Keuls necessita do cálculo de $(k - 1)$ valores críticos, dados por:

$$SNK_{p'} = q_{(\alpha, p', \nu)} \sqrt{\frac{QME}{r}}, \quad (2.5)$$

para $p' = 2, 3, \dots, k$, que é o número de médias abrangidas no contraste de duas médias ordenadas.

Assim, SNK_2 é igual ao valor crítico LSD e SNK_k é igual ao valor crítico Δ . Para valores intermediários de p' , o valor crítico $SNK_{p'}$ é um valor intermediário entre os valores críticos LSD e Δ (Machado et al., 2005).

Scheffé (1953) propôs um procedimento bem geral, de amplitude fixa, para comparar qualquer contraste entre médias de populações. O seu valor crítico é calculado por:

$$S = \sqrt{(k - 1) \sum_{i=1}^k \frac{c_i^2}{r} F_{(\alpha, k-1, \nu)} QME}, \quad (2.6)$$

em que k é o número de populações do experimento; c_i os coeficientes das médias do contraste em questão, com $i = 1, 2, \dots, k$; $F_{(\alpha, k-1, \nu)}$ o quantil superior $100\alpha\%$ da distribuição F e ν de graus de liberdade do resíduo.

Este teste pode ser aplicado a qualquer contraste de médias e, apesar de ser um teste mais conservativo que o teste t de Student, ele também é mais flexível, pois não exige a condição de ortogonalidade entre os contrastes

e nem que estes sejam estabelecidos antes de examinar os dados (Banzatto & Kronka, 2006).

Amplitudes estudentizadas especiais, apresentadas por Duncan (1955), com $\alpha_{p'} = [1 - (1 - \alpha)^{p'-1}]$, para $p' = 2, 3, \dots, k$, são usadas no teste de Duncan. Para este procedimento, os $(k - 1)$ valores críticos são calculados por:

$$D_{p'} = q_{(\alpha_{p'}, p', \nu)} \sqrt{\frac{QME}{r}}. \quad (2.7)$$

para $p' = 2, 3, \dots, k$ que é o número de médias abrangidas no contraste de duas médias ordenadas.

O teste de Duncan é mais conservativo do que o teste de t de Student, mas mais liberal do que os testes de Tukey e *SNK* (Machado et al., 2005).

O procedimento proposto por Scott & Knott (1974) apresenta como estatística:

$$\lambda = \frac{\pi}{2(\pi - 2)} \times \frac{B_0}{\hat{\sigma}_0^2},$$

em que π é uma constante que equivale a 3,141593; B_0 é o valor máximo da soma de quadrados máxima entre dois grupos de médias, sobre todas as $(k - 1)$ partições possíveis e $\hat{\sigma}_0^2$ é a estimativa de máxima verossimilhança de $\hat{\sigma}^2$.

Este teste utiliza a razão de verossimilhança para testar a hipótese de que k populações podem ser divididas em dois grupos que maximizam a soma de quadrados entre grupos. O número de grupos cresce exponencialmente quando o número (k) de populações é grande, dificultando a aplicação do teste. Existem $2^{k-1} - 1$ partições possíveis das k médias em dois grupos distintos (Ramalho et al., 2000). Mas Fisher (1958) demonstrou em seu trabalho que é necessário considerar apenas $(k - 1)$ partições obtidas pela ordenação das médias e sua divisão entre duas partições sucessivas.

Scott & Knott (1974) mostram que a distribuição da estatística λ pode ser aproximada por uma χ^2 com $\nu_0 = k/(\pi - 2)$ graus de liberdade. O teste é significativo se $\lambda \geq \chi_{\alpha}^2(\nu_0)$ e, assim, rejeita-se a hipótese de que os dois grupos são iguais. O mesmo teste é aplicado para cada grupo se-

paradamente. Se o teste não for significativo, conclui-se que as médias das populações pertencem ao mesmo grupo. O procedimento continua, dividindo cada grupo em dois subgrupos, se o teste para aquela partição for significativo, ou concluindo que o grupo é homogêneo no outro caso. O procedimento termina quando não há grupos remanescentes para serem divididos.

A grande vantagem deste teste é a ausência de ambigüidade e a robustez de normalidade presente nos procedimentos de comparações múltiplas (Silva, 1998).

2.2 Tipos de erro e poder do teste

É preciso ter sempre a consciência de que toda inferência realizada está sujeita a erros. Esses erros podem ser classificados como erro do tipo I, quando rejeita-se a hipótese nula, sendo ela verdadeira. Comete-se o erro do tipo II quando não se rejeita a hipótese nula dado que ela é falsa.

Utilizando a diferença mínima significativa para comparações múltiplas pareadas, Carmer (1976) buscou verificar as freqüências dos erros tipo I, II e III. O erro tipo III é aquele cometido quando se declara que uma média é maior que a outra quando, na verdade, ocorre o contrário. O autor observou que, com o aumento do nível nominal de significância $\alpha = 20\%$ para $\alpha = 40\%$, reduziu-se consideravelmente a proporção de erro tipo II, enquanto aumentou-se razoavelmente os erros tipos I e III. Verificou também que se o pesquisador quiser a redução dos três tipos de erros, ele deve melhorar a precisão do seu experimento.

Segundo Banzatto & Kronka (2006), os erros tipo I e II estão associados de tal forma que, se a probabilidade de ocorrência de um deles diminuir, automaticamente a probabilidade de ocorrência do outro aumenta.

O'Neill & Wetherill (1971) definem duas maneiras básicas para se calcular a taxa de erro do tipo I:

- i) taxa de erro por comparação (*comparisonwise error rate*)

$$\frac{\text{Número de inferências erradas}}{\text{Número total de inferências}}; \quad (2.8)$$

ii) taxa de erro por experimento (*experimentwise error rate*)

$$\frac{\text{Número de experimentos com pelo menos uma inferência errada}}{\text{Número total de experimentos}}. \quad (2.9)$$

A maior preocupação ao se desenvolverem procedimentos para comparar médias centrou-se no controle do erro tipo I para garantir o nível de probabilidade desejado num conjunto de várias comparações. Ao se controlar de forma excessiva esse tipo de erro, aumenta-se a taxa de erro tipo II e diminui-se o poder do teste, que é, nesse caso, denominado conservativo ou conservador (Carmer & Swanson, 1973), dificultando a rejeição da hipótese nula. Porém, ao se aumentar a taxa de erro tipo I, diminui-se a taxa de erro tipo II e o poder é aumentado, sendo considerado um teste liberal (Machado et al., 2005).

O poder de um teste é a capacidade deste de identificar todas as reais diferenças entre os tratamentos (Ramalho et al., 2000), ou seja, a probabilidade de rejeitar a hipótese de igualdade entre médias quando ela realmente é falsa ($1 - \beta$).

Vários trabalhos comparam os diferentes procedimentos de comparações múltiplas quanto às taxas de erro tipo I (Borges & Ferreira, 2003; Carmer & Swanson, 1973; Perecin & Barbosa, 1988). Essa não é uma tarefa fácil de se executar analiticamente, porém, alguns autores já o fizeram, de forma analítica ou por meio de simulação.

Um procedimento é considerado robusto se, ao violar uma das pressuposições básicas para sua aplicação, ele mantiver, pelo menos aproximadamente, o desempenho delineado originalmente na elaboração de sua teoria (Borges & Ferreira, 2003).

Segundo Hocking (1985), os procedimentos de comparações múltiplas têm atraído considerável atenção e é improvável que se chegue a uma

solução final, pois o problema gira em torno de taxas de erro e poder. Ao mesmo tempo em que deseja-se cortar a ocorrência de decisões incorretas (falsas significâncias), deseja-se também detectar diferenças reais entre os tratamentos.

Guerra & Donaire (1982) dizem que se devem estabelecer amostras com tamanhos relativamente grandes para poder reduzir a probabilidade de se cometer o erro tipo II, fixando um baixo risco α para a probabilidade de se cometer o erro tipo I.

2.3 Comparações múltiplas multivariadas

As variáveis possuem relações entre si e qualquer mudança em uma ou algumas delas, afeta as outras. Um tratamento aplicado em um ser, a um solo ou a um sistema, não afeta isoladamente apenas uma variável, e sim todas as variáveis (Mardia et al., 1979).

Segundo Rencher (2002), quando o interesse é analisar todas as variáveis ao mesmo tempo, uma alternativa são os contrastes, embora sejam limitados às situações em que os tratamentos são qualitativos e estruturados. Por meio de contrastes se podem obter respostas mais específicas a respeito de hipóteses de interesse. Isto é, os contrastes lineares ajudam a detalhar mais sobre a questão de comparações entre tratamentos (Negrillo & Perre, 1987). Há dois usos dos contrastes em um ajuste multivariado. Considerando a hipótese $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ com $\mathbf{C}\mathbf{j} = \mathbf{0}$, em que $\boldsymbol{\mu}$ é o vetor de médias, \mathbf{C} é matriz de contraste $(p-1) \times p$, \mathbf{j} é um vetor de 1 e $\mathbf{0}$ é o vetor nulo. A soma de cada linha de \mathbf{C} é zero, pois $\mathbf{C}\boldsymbol{\mu}$ é, conseqüentemente, um confronto de contrastes para comparar os elementos $\mu_1, \mu_2, \dots, \mu_p$ de $\boldsymbol{\mu}$ entre si. Neste trabalho, o interesse está nos contrastes que comparam diversos vetores médias e não os elementos dentro de um vetor.

Um contraste de vetores de médias de uma população é dado por

$$\mathbf{W} = c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \dots + c_k\boldsymbol{\mu}_k, \quad (2.10)$$

em que $\sum_{i=1}^k c_i = 0$.

Assim, se podem efetuar comparações entre vetores de médias das

populações, da mesma forma como se opera no caso univariado.

Para cada contraste é calculado um resíduo adequado. Quando se trabalha com contrastes ortogonais, sabe-se que os $k - 1$ graus de liberdade decompõem-se em $k - 1$ contrastes com um grau de liberdade cada (Rencher, 2002).

Para cada hipótese formulada, deve-se obter o estimador de W_s que é dado pelo correspondente contraste dos vetores de médias amostrais:

$$\widehat{\mathbf{W}} = c_1\bar{\mathbf{y}}_1 + c_2\bar{\mathbf{y}}_2 + \dots + c_k\bar{\mathbf{y}}_k. \quad (2.11)$$

em que $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_k$ são os vetores de médias amostrais.

Para realizar o teste, é obtida a estatística U_s , tal que

$$U_s = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_s|}, \quad (2.12)$$

em que \mathbf{E} é a matriz de soma de quadrados e produtos do resíduo e \mathbf{H}_s é matriz da soma de quadrados e produtos, com $s=1,2,\dots,k-1$.

Se $U_s > U_{(p,1,v_E)}$, rejeita-se H_0 . Cada matriz \mathbf{H}_s tem um grau de liberdade porque o posto de \mathbf{H}_s é igual a um.

Quando a hipótese de efeitos iguais para tratamentos é rejeitada, aqueles efeitos que levaram à rejeição são de interesse. Para comparações simultâneas duas a duas, a aproximação de Bonferroni pode ser usada para construir intervalos de confiança simultâneos para os componentes da diferença $\tau_h - \tau_i$ (diferenças de efeitos das populações h e i , respectivamente)(Mardia et al., 1979). Esses intervalos são mais curtos que os obtidos para todos os contrastes e requerem apenas valores críticos da estatística univariada t , considerando $\tau_{i\ell}$ o ℓ -ésimo componente de τ_i . Desde que τ_i pode ser estimado por $\hat{\tau}_i = \bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}_{..}$, então, $\hat{\tau}_{i\ell} = \bar{Y}_{i.\ell} - \bar{Y}_{.. \ell}$.

Como $\hat{\tau}_{i\ell} = \bar{Y}_{i.\ell} - \bar{Y}_{.. \ell}$ corresponde à diferença entre duas médias amostrais independentes, o teste t de duas amostras é válido, modificando-se adequadamente o nível de significância (Ferreira, 2007). A estimativa da variância do contraste entre duas médias de tratamentos é dada por,

$$\widehat{Var}(\bar{Y}_{h.\ell} - \bar{Y}_{i.\ell}) = \left(\frac{1}{n_h} + \frac{1}{n_i} \right) \frac{E_{\ell\ell}}{\nu}. \quad (2.13)$$

em que n_h é o tamanho da amostra da população h ; n_i é o tamanho da amostra da população i ; $E_{\ell\ell}$ é a soma de quadrados do resíduo e da ℓ -ésima variável e ν é o grau de liberdade do erro.

Segundo Ferreira (2007), a divisão de $E_{\ell\ell}$ pelos seus respectivos graus de liberdade (ν) deve-se ao fato de que o elemento em questão ($E_{\ell\ell}$) refere-se a uma soma de quadrados do resíduo. Dessa forma, desde que p variáveis são consideradas e $k(k-1)/2$ comparações duas a duas serão realizadas, então, o intervalo de confiança protegido por Bonferroni para a diferença de efeitos do tratamento é dado por:

$$\bar{Y}_{h.\ell} - \bar{Y}_{i.\ell} \pm t_{\nu, \alpha/[pk(k-1)]} \sqrt{\left(\frac{1}{n_h} + \frac{1}{n_i} \right) \frac{E_{\ell\ell}}{\nu}}, \quad (2.14)$$

para todos os $\ell = 1, 2, \dots, p$ e todas as diferenças $h < i = 1, 2, \dots, k$.

2.4 Simulação Monte Carlo

Simulação é uma ferramenta poderosa se compreendida e usada corretamente. Naylor et al. (1971) definem simulação como sendo “uma técnica numérica para a condução de um experimento sob a ótica computacional, envolvendo tipos matemáticos seguros e modelos teóricos que descrevem o comportamento de negócios ou sistemas econômicos sobre um grande período, em tempo reduzido”.

Para estudar ou avaliar um teste estatístico, muitas vezes, torna-se bastante difícil obter, analiticamente, informações sobre o poder e as taxas de erro tipo I. Uma maneira de se obter as informações desejadas é por meio de simulação (Cecchetti, 1999).

O método de Monte Carlo é uma técnica que consiste na simulação de dados pela geração de números pseudo-aleatórios, por meio de algum algoritmo em alguma linguagem de programação, de acordo com determinada distribuição de probabilidade. O objetivo é estudar o comportamento de

diferentes técnicas estatísticas que poderiam ser empregadas num problema específico (Dachs, 1988)

O método leva este nome devido à famosa roleta de Monte Carlo, no Principado de Mônaco. Seu nome, bem como o desenvolvimento sistemático do método, ocorreu em 1944, época da Segunda Grande Guerra, tendo sido utilizado como ferramenta de pesquisa para o desenvolvimento da bomba atômica. Porém, existem alguns registros isolados de sua utilização em datas bem anteriores. Por exemplo, pela segunda metade do século XIX várias pessoas executaram experiências nas quais lançavam setas, de maneira fortuita, sobre uma tábua onde havia um conjunto de linhas paralelas e deduziram o valor de $\pi = 3,14\dots$, observando o número de intersecções entre as setas e linhas (Pllana, 2002).

Em síntese, o método de simulação Monte Carlo consiste em simular dados a partir de uma seqüência de números aleatórios, com a finalidade de se obter uma amostra da população. Admite-se também que todo processo simulado que contém um componente aleatório de qualquer distribuição faz parte deste método (Carari, 2004).

2.5 Método *bootstrap* e permutação

O método de *bootstrap* foi originalmente proposto por Bradley Efron em um influente artigo publicado no Annals of Statistics, em 1979. Ele se baseia na construção de distribuições amostrais por meio de reamostragens e é muito utilizado para estimar intervalos de confiança. Também pode ser utilizado, por exemplo, para estimar o viés e a variância de estimadores ou de testes de hipóteses calibrados.

O método de *bootstrap* tem por base a idéia de que o pesquisador pode tratar sua amostra como se ela fosse a população que deu origem aos dados e usar amostragem com reposição da amostra original para gerar pseudo-amostras. A partir dessas pseudo-amostras, é possível estimar características da população, tais como média, variância, percentis, etc.

Vários esquemas diferentes de métodos *bootstrap* têm sido propostos, na literatura e muitos deles apresentam bom desempenho em uma ampla

variedade de situações (Diaconis & Efron, 1983).

A idéia básica do método *bootstrap* é amostrar os dados originais com reposição, obtendo-se valores da estatística de interesse. Repete-se este processo inúmeras vezes, até a obtenção de B valores. Calcula-se o desvio padrão desses valores e, então, tem-se o erro padrão da estatística. Dado o custo alto e a escassez conseqüente de dados em muitas aplicações, combinadas com o custo reduzido e a abundância do poder da computação, o método *bootstrap* se torna uma técnica muito atraente por extrair informações empíricas de dados (Efron & Tibshirani, 1991).

O método *bootstrap* também pode ser visto como um procedimento para tirar conclusões sobre populações, quando não se têm suposições sobre a distribuição das populações de interesse. A única suposição feita é que a amostra original representa bem a população da qual os dados são provenientes.

Os métodos de permutação foram introduzidos na década de 1930 por Fischer e Pitman, mas, por muito tempo, se restringiram ao contexto teórico, devido às dificuldades de cálculos computacionais, somente resolvidos muitos anos depois (Good, 1994). Esse método consiste em amostrar os dados originais sem reposição, no intuito de obter valores da estatística de interesse.

2.6 Distribuição normal multivariada

Em análise multivariada, é importante a generalização da densidade normal univariada para duas ou mais dimensões. De fato, muitos dos métodos estatísticos multivariados baseiam-se na pressuposição de que os dados são obtidos de uma população com distribuição normal multivariada. Apesar de os dados originais não serem quase nunca exatamente normal multivariado, a densidade normal se constitui, muitas vezes, numa aproximação adequada e útil da verdadeira distribuição populacional (Johnson & Wichern, 1998).

Segundo Rencher (2002), a importância da distribuição normal, além de seu fácil manuseio matemático, possui duas razões práticas que justificam a sua utilidade. A primeira diz que a distribuição normal é a mais adequada

para modelos populacionais em várias situações. A segunda refere-se ao fato de a distribuição amostral de muitas estatísticas multivariadas serem aproximadamente normal, independentemente da forma da distribuição da população original, garantida pelo teorema do limite central.

A função de densidade normal multivariada de um vetor aleatório \mathbf{X} é dada por:

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (2.15)$$

em que $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são os vetores de média e a matriz de covariância, simétrica e positiva definida, respectivamente e p é a dimensão do vetor aleatório \mathbf{X} .

Timm (2002) enuncia alguns teoremas importantes. São eles:

Teorema 2.6.1. *Se $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e se $\mathbf{Y} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, então $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ são $N_p(\mathbf{0}, \mathbf{I})$ independentes.*

Teorema 2.6.2. *Se $\mathbf{X}_1, \mathbf{X}_2, \dots$ é uma seqüência infinita de vetores aleatórios independentes e identicamente distribuídos de uma distribuição com média $\boldsymbol{\mu}$ e covariância $\boldsymbol{\Sigma}$, então $\bar{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$, com $n \rightarrow \infty$.*

Teorema 2.6.3. *Se o vetor aleatório $\mathbf{X} \in \mathbb{R}^p$ segue uma distribuição normal multivariada com densidade $f_{\mathbf{X}}(\mathbf{x})$ dada em (2.15), então a sua média $E(\mathbf{X})$ é igual ao parâmetro $\boldsymbol{\mu}$ e sua matriz de covariâncias $E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T$ é igual ao parâmetro $\boldsymbol{\Sigma}$.*

As simulações realizadas neste trabalho foram baseadas nestes teoremas, justificando, assim, sua descrição nesta seção.

2.7 Distância de Mahalanobis

Num caso univariado, a distância entre dois pontos é simplesmente a diferença entre eles. Para o propósito estatístico, esta diferença não é muito informativa (Rencher, 2002).

No caso multivariado, temos que considerar não só as discrepâncias das variáveis como também as covariâncias ou correlações. Assim, para a

distância estatística, utilizou-se a matriz de covariância (\mathbf{S}) e a distância foi dada por

$$D^2 = (\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y}), \quad (2.16)$$

em que D^2 é a distância de Mahalanobis, \mathbf{S}^{-1} é a inversa da matriz de covariância e \mathbf{x} e \mathbf{y} são vetores aleatórios.

Segundo Ferreira (2007), esta distância é conhecida por distância generalizada de Mahalanobis. Em muitas situações experimentais, este tipo de distância possui um apelo intuitivo para o seu uso, justamente por contemplar a estrutura de covariância entre as diferentes variáveis. Ao usarmos esta definição de distância, estamos igualando as condições das escalas e eliminando o efeito da correlação. Isso porque cada variável será padronizada pelas respectivas variabilidades e a correlação entre diferentes variáveis será contemplada.

Deve-se ressaltar que o uso da inversa da matriz covariância numa distância Mahalanobis tem o objetivo de padronizar todas as variáveis para a mesma variância e eliminar correlações (Rencher, 2002).

3 METODOLOGIA

Sejam $\mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \dots, \mathbf{Y}_{ir}$ vetores aleatórios de dimensão $p \times 1$, obtidos na i -ésima população ($i = 1, 2, \dots, k$) com distribuição multivariada qualquer, com vetor de médias $\boldsymbol{\mu}_i$ e matriz de covariância $\boldsymbol{\Sigma}_i$. Então, foi proposto um procedimento de comparações múltiplas multivariado, conforme descrição apresentada a seguir.

Inicialmente, seja família das $m = k(k-1)/2$ hipóteses definidas por:

$$\begin{cases} H_0 : \boldsymbol{\mu}_i = \boldsymbol{\mu}_{i'}, & i \neq i' = 1, 2, \dots, k \\ H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i'}, & \text{para algum } i \neq i'. \end{cases} \quad (3.1)$$

Para testar as hipóteses dessa família de comparações foi proposto método baseado em *bootstrap* e permutação. Inicialmente, foram estimados os vetores de médias e as matrizes de covariância da i -ésima população, dadas por:

$$\bar{\mathbf{Y}}_{i.} = \frac{\sum_{j=1}^r \mathbf{Y}_{ij}}{r}, \quad (3.2)$$

$$\mathbf{S}_i = \frac{1}{r-1} \sum_{j=1}^r (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})^T. \quad (3.3)$$

O estimador da $\text{cov}(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i' .})$ é dado por

$$\widehat{\text{cov}}(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i' .}) = \frac{1}{r}(\mathbf{S}_i + \mathbf{S}_{i'}). \quad (3.4)$$

A estatística do novo procedimento foi definida por:

$$T_{ii'}^2 = r(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i' .})^T(\mathbf{S}_i + \mathbf{S}_{i'})^{-1}(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i' .}), \quad (3.5)$$

para $i \neq i' = 1, 2, \dots, k$.

Uma limitação desse procedimento é que a inversa única $(\mathbf{S}_i + \mathbf{S}_{i'})^{-1}$ só existirá se $p \leq 2(r-1)$. Uma possibilidade para contornar esse problema foi a utilização de inversas generalizadas de $\mathbf{S}_i + \mathbf{S}_{i'}$. Outra possibilidade é utilizar a covariância combinada das k populações, dada por:

$$\mathbf{S}_p = \frac{\sum_{i=1}^k \mathbf{S}_i}{k}. \quad (3.6)$$

Assim, tem-se a seguinte expressão alternativa à expressão (3.5):

$$T_{ii'}^2 = \frac{r}{2}(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i'.})^T \mathbf{S}_p^{-1} (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i'.}), \quad (3.7)$$

para $i \neq i' = 1, 2, \dots, k$.

Com isso, foi possível simular situações em que $p \leq k(r-1)$. Adotando-se esse procedimento, foi possível implementar, no programa de simulação, uma rotina para obter os $k(k-1)/2$ valores necessários sem a utilização de estruturas de repetição, que são ineficientes no programa R (R Development Core Team, 2007). Com isso, houve uma melhoria no desempenho das rotinas de simulação, com redução de 7 vezes do tempo gasto inicialmente. Para a pior configuração, eram gastos 21 dias, havendo uma redução para 3 dias. A rotina de simulação está apresentado no Anexo.

Na ausência de normalidade multivariada ou mesmo sob normalidade, a distribuição conjunta de $T_{ii'}^2$ para os pares de populações não é conhecida. Assim, foi obtida a distribuição nula de $T_{ii'}^2$ por meio de *bootstrap* e permutação. A imposição da hipótese nula de igualdade dos k vetores de médias populacionais ($H_0 = \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k$) foi feita com o agrupamento das rk observações multivariadas em um único conjunto e este reamostrado com reposição e sem reposição, refazendo-se a estrutura amostral das k populações com r observações p -variadas.

Desta amostra de *bootstrap* e permutação foram estimadas as médias e as covariâncias da i -ésima e i' -ésima populações pelas expressões (3.2) e (3.3). A estatística $T_{ii'}^2$ foi calculada utilizando-se as expressões (3.5) ou (3.7) para todos os $k(k-1)/2$ pares de médias e foi denominada por $T_{ii'm}^2$, formando o seguinte conjunto para a m -ésima amostra de *bootstrap* e permutação

$$\Omega = \{T_{12m}^2, T_{13m}^2, T_{14m}^2, \dots, T_{1km}^2, \dots, T_{(k-1)km}^2\}, \quad (3.8)$$

em que $m = 1, 2, \dots, B$.

A partir desse conjunto, foi definido o m -ésimo elemento de um novo conjunto, o qual foi inspirado nos modelos de comparações múltiplas de Tukey, que é baseado na amplitude “estudentizada”, da seguinte forma:

$$\Gamma_m = \max_{i \neq i'} \Omega = \max_{i \neq i'} \{T_{12m}^2, T_{13m}^2, T_{ii'm}^2 \dots, T_{(k-1)km}^2\}. \quad (3.9)$$

Como foram consideradas B amostras de *bootstrap* e permutação incluindo a amostra original e este procedimento foi repetido para todas reamostragens, então, o conjunto Γ foi formado da seguinte forma:

$$\Gamma = \{\Gamma_1, \Gamma_2, \dots, \Gamma_m, \dots, \Gamma_B\}. \quad (3.10)$$

O teste da hipótese (3.1) foi concluído calculando-se o valor- p da seguinte forma:

$$\text{valor} - p_{ii'} = \frac{\sum_{m=1}^B I(\Gamma_m \geq T_{ii'}^2)}{B}, \quad (3.11)$$

em que $I(\Gamma_m \geq T_{ii'}^2)$ é a função indicadora que retorna 1 se a desigualdade for verificada e 0 se não for. O valor- p de cada par de populações foi confrontada com o valor da significância nominal α , para completar o teste da hipótese (3.1).

Para este trabalho, foram considerados $B = 1000$ amostras de *bootstrap* e permutação. Identificado o par populacional que apresenta vetores de médias diferentes, foram determinados os componentes (variáveis) responsáveis pela rejeição da hipótese nula. Para isso, foram utilizados procedimentos semelhantes ao teste conjunto das p -variáveis apresentadas anteriormente, utilizando-se, entretanto, a amplitude studentizada. Assim, para a ℓ -ésima variável, e considerando a i -ésima e i' -ésima populações, para a, amostras *bootstrap* ou permutação e original, foi determinada a quantidade:

$$q_{ii'l} = \frac{\max(\bar{Y}_{i.l}, \bar{Y}_{i'.l}) - \min(\bar{Y}_{i.l}, \bar{Y}_{i'.l})}{\sqrt{\frac{S_{ii'}^2 + S_{i'l}^2}{2r}}}. \quad (3.12)$$

Das np amplitudes estudentizadas, foi formado um conjunto, para o qual foi tomado o seu máximo na m -ésima amostra de *bootstrap* e permutação, da seguinte forma:

$$Q_m = \max_{i \neq i', \ell} \{q_{ii'lm}\}. \quad (3.13)$$

Assim, este conjunto ficou definido por:

$$Q = \{Q_1, Q_2, \dots, Q_m, \dots, Q_B\}. \quad (3.14)$$

O valor- p para identificar quais variáveis possuem efeito significativo no par de populações para o qual anteriormente a hipótese nula (3.1) foi rejeitada, definiu-se por:

$$\text{valor} - p_{ii'\ell} = \frac{\sum_{m=1}^B I(Q_m \geq q_{ii'\ell})}{B}, \quad (3.15)$$

sendo $\ell = 1, 2, \dots, p$.

Para avaliar o desempenho do teste proposto, foram realizadas simulações Monte Carlo em duas etapas. Na primeira, foram consideradas k populações com vetores de médias iguais, para possibilitar que fosse avaliado o erro tipo I. Foi considerada a distribuição normal multivariada.

Na segunda etapa, foram consideradas populações com vetores de médias diferentes, para avaliar o desempenho do teste em relação ao poder. Por não existirem procedimentos equivalentes na literatura, o desempenho do novo teste não pôde ser comparado a outros, a não ser em relação à taxa de erro tipo I, que foi confrontada com os níveis nominais de significância α . Foram considerados valores de α iguais a 1% e 5% e o erro tipo I foi mensurado apenas pela taxa de erro por experimento.

Sob a hipótese nula de vetores de médias iguais ($\boldsymbol{\mu}$), foram realizadas N simulações Monte Carlo para avaliar o desempenho do teste proposto em relação às taxas de erro tipo I por experimento. Foram consideradas r repetições de cada população sob o modelo normal multivariado, obtendo-se, em cada simulação Monte Carlo, a amostra $\mathbf{Y}_{11}, \mathbf{Y}_{12}, \dots, \mathbf{Y}_{1r}, \mathbf{Y}_{21},$

$\dots, \mathbf{Y}_{2r}, \dots, \mathbf{Y}_{ij}, \dots, \mathbf{Y}_{kr}$. Cada vetor foi simulado de uma distribuição normal multivariada por $\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$, em que p é a dimensão (número de variáveis) e $\boldsymbol{\Sigma}_i$ a matriz de covariâncias da i -ésima população ($p \times p$), assumida positiva definida.

A geração de uma realização da variável aleatória \mathbf{Y}_{ij} foi feita da seguinte forma:

- a) gerou-se, inicialmente, um vetor p -dimensional \mathbf{Z}_{ij} de uma distribuição normal $N_p(\mathbf{0}, \mathbf{I})$ por $\mathbf{Z}_{ij} \sim N_p(\mathbf{0}, \mathbf{I})$. Utilizou-se o algoritmo de Box-Muller para a geração desta distribuição (Dachs, 1988) e
- b) realizações do vetor aleatório \mathbf{Y}_{ij} foram obtidas utilizando-se a seguinte transformação linear:

$$\mathbf{Y}_{ij} = \boldsymbol{\Sigma}_i^{1/2} \mathbf{Z}_{ij} + \boldsymbol{\mu}, \quad (3.16)$$

em que $\boldsymbol{\Sigma}_i^{1/2}$ é a raiz quadrada da matriz $\boldsymbol{\Sigma}_i$, ou seja, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i^{1/2} \boldsymbol{\Sigma}_i^{1/2}$.

Estes procedimentos foram executados por meio do programa R, utilizando-se a função *rmvnorm* do pacote *mvtnorm* (R Development Core Team, 2007).

Para realizar as simulações foram consideradas populações homocedásticas e heterocedásticas. No primeiro caso, as covariâncias populacionais foram todas iguais a $\boldsymbol{\Sigma}$, definida por uma estrutura de simetria composta dada por

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}], \quad (3.17)$$

em que \mathbf{J} é uma matriz $p \times p$ formada por valores iguais a 1.

Para as situações, de covariâncias heterogêneas, foi considerada uma matriz de covariância $\boldsymbol{\Sigma}_1$ com estrutura dada pela equação (3.17) para a população 1. As demais populações possuíam matrizes de covariâncias defi-

nidas por:

$$\boldsymbol{\Sigma}_i = \left[1 + \frac{(\sqrt[k]{\delta} - 1)(i - 1)}{k - 1} \right] \boldsymbol{\Sigma}_1, \quad (3.18)$$

em que δ refere-se ao grau de heterogeneidade, pré-fixado e definido como a maior razão entre as covariâncias generalizadas, dado por $\delta = |\boldsymbol{\Sigma}_k|/|\boldsymbol{\Sigma}_1|$, para $i = 2, 3, \dots, k$.

Este critério estabelece que a razão $|\boldsymbol{\Sigma}_i|/|\boldsymbol{\Sigma}_1|$ é sempre um número entre 1 e δ , para as diferentes populações, $i = 2, 3, \dots, k - 1$. O valor desta razão cresce com o aumento de i . Esta situação de heterogeneidade de covariâncias foi considerada para se obter situações desfavoráveis à aplicação do teste, buscando-se, assim, avaliar sua robustez. Outras estruturas de covariâncias, além da estrutura de simetria composta, foram avaliadas, mas não alteraram os resultados.

Nas simulações, foram considerados diferentes valores do número de repetições r , do número de populações k e do número de variáveis p , sendo $r = 3, 10$ e 20 , $p = 2, 5$ e 10 e $k = 5, 10, 50$ e 100 . Diferentes valores de ρ foram considerados, mas, como os resultados não se alteravam expressivamente, deu-se ênfase à situação de $\rho = 0, 5$. O valor de σ^2 foi fixado em $\sigma^2 = 10$ e $\boldsymbol{\mu} = \mathbf{0}$, sem perda de generalidade. Os valores relativos às heterogeneidades de covariâncias δ foram: 1, 2, 8 e 16. O valor $\delta = 1$ determina situações de homocedasticidade. Finalmente, foram consideradas $N = 1000$ simulações Monte Carlo e $B = 1000$ reamostragens *bootstrap* e permutação.

As taxas de erro tipo I por experimento foram obtidas pela proporção de experimentos, em relação às N simulações realizadas, que apresentaram pelo menos uma diferença entre vetores de médias significativamente ($P < \alpha$) diferentes de zero. Os p -valores foram obtidos pela equação (3.11) e os valores nominais de significância α foram iguais a 1% e 5%.

Sob a hipótese alternativa H_1 , as simulações foram realizadas para avaliar o poder do teste. As simulações foram feitas seguindo-se os procedimentos descritos anteriormente para o erro tipo I, exceto pelo fato de uma realização de \mathbf{Y}_{ij} ser obtida de uma distribuição normal multivariada dada por: $\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.

Os vetores de médias populacionais foram especificados como discu-

tido a seguir. Para a população 1,

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (3.19)$$

e para as demais i' , $i' = 2, 3, \dots, k$, considerou-se

$$\boldsymbol{\mu}_{i'} = \boldsymbol{\mu}_{i'-1} + \phi \sqrt{\frac{1}{r} \text{diag}(\boldsymbol{\Sigma}_{i'-1} + \boldsymbol{\Sigma}_{i'})}, \quad (3.20)$$

sendo

$$\text{diag}(\boldsymbol{\Sigma}_{i'-1} + \boldsymbol{\Sigma}_{i'}) = \begin{bmatrix} \sigma_{11}^{i'-1} + \sigma_{11}^{i'} \\ \sigma_{22}^{i'-1} + \sigma_{22}^{i'} \\ \vdots \\ \sigma_{pp}^{i'-1} + \sigma_{pp}^{i'} \end{bmatrix}, \quad (3.21)$$

e $\phi = 1, 2, 4, 8$ e 16 .

A idéia foi criar populações cujo vetor de médias de uma população fosse diferente do vetor de médias da população anterior por uma diferença de ϕ erros padrões da diferença entre médias de cada componente univariado destes vetores.

Os valores de poder foram computados pela proporção de vetores de médias que foram significativamente diferentes (valor- $p < \alpha$) uns dos outros em relação a todos as N simulações e considerando todos os pares cujas diferenças paramétricas foram de ϕ erros padrões. Assim, em cada simulação somente $(k - 1)$ pares de médias foram consideradas, totalizando $N(k - 1)$ comparações, no total das N simulações Monte Carlo realizadas.

4 RESULTADOS E DISCUSSÃO

Inicialmente, é interessante destacar que a metodologia apresentada nesta dissertação é inédita e poderia ter sido apresentada nesta seção de resultados. No entanto, foi descrita na seção anterior, em função de aspectos didáticos apenas. Também é importante notar que esta proposta preenche uma lacuna existente nas análises de variância multivariadas, pois os procedimentos de comparações múltiplas são inexistentes. As únicas alternativas são os contrastes, quando os tratamentos são estruturados.

A apresentação e a discussão dos resultados foram divididas em duas partes distintas. Na primeira, avaliou-se o tamanho dos testes de comparações múltiplas multivariadas. As taxas de erro tipo I foram comparadas com os valores nominais de significância de 1% e 5%. Na segunda, foram apresentados os valores de poder em função de r , k , δ , p e ϕ . Também foi apresentado um exemplo real para ilustrar a metodologia.

4.1 Taxas de erro tipo I

Nas Tabelas 4.1 a 4.6 estão apresentadas as taxas de erro tipo I por experimento dos procedimentos de comparações múltiplas multivariadas via *bootstrap*, em função do número de tratamentos k , do número de repetições r , do número de variáveis p e da heterogeneidade de covariâncias δ para $\alpha = 1\%$ e 5% . As taxas de erro observadas foram comparadas com os valores nominais utilizando-se o intervalo binomial de 99% de confiança.

No caso de $k = 5$, foram utilizadas 2.000 simulações Monte Carlo, e nos demais, 1.000 simulações. Considerando $k = 5$, o intervalo de 99% de confiança para $\alpha = 1\%$ é $[0, 52\%; 1, 72\%]$ e para $\alpha = 5\%$, $[3, 83\%; 6, 39\%]$. Considerando os demais valores de k , o intervalo para $\alpha = 1\%$ é de $[0, 37\%; 2, 13\%]$ e para $\alpha = 5\%$, $[3, 40\%; 7, 05\%]$.

Na Tabela 4.1 estão apresentados os resultados da taxa de erro tipo I por experimento para $k = 5$. É conveniente salientar que todos os resultados para ($k = 5$) foram obtidos utilizando-se a expressão (3.5) para determinar $T_{ii'}^2$. Como, em alguns casos, $p > 2(r - 1)$, como, por exemplo, $r = 3$ e $p = 5$ e $r = 3$ e $p = 10$, os valores de $T_{ii'}^2$ só puderam ser computados substituindo-se $(\mathbf{S}_i + \mathbf{S}_{i'})^{-1}$ por $(\mathbf{S}_i + \mathbf{S}_{i'})^g$, que é uma inversa generalizada

e, no caso específico, foi a Moore-Penrose. Este fato limita a quantidade de informação contida nas p variáveis a uma dimensão menor dada por $2(r-1)$, o que pode alterar as taxas de erro tipo I por experimento. As únicas razões para se utilizar um maior número de simulações Monte Carlo foram a maior velocidade dos programas de simulação para k pequeno ($k = 5$) e a tentativa de obter maior precisão nestes casos, mencionados anteriormente.

Para $k = 5$ (Tabela 4.1), pode-se observar que o teste de comparações múltiplas multivariadas proposto controlou as taxas de erro tipo I por experimento em todos os casos em que $p < 2(r-1)$, de forma exata ou de forma conservativa, para $\alpha = 1\%$. Para $\alpha = 5\%$, houve o mesmo tipo de desempenho, exceto para os casos $r = 10$, $p = 2$ com $\delta = 8$ e 16 , em que houve taxas de erro superiores significativamente (valor- $p < 0,01$) ao valor nominal. Embora, nestes dois casos, as taxas tenham sido significativamente superiores ao valor nominal $\alpha = 5\%$, os resultados não foram expressivamente diferentes.

Não se conseguiu determinar um padrão plausível para explicar as diferenças entre os resultados conservativos ($r = 3$, $p = 2$, qualquer δ , $\alpha = 1\%$ e 5% , $r = 10$, $p = 10$, qualquer δ , $\alpha = 1\%$) e dos exatos para as taxas de erros observadas. Espera-se, no entanto, que os valores de poder nos casos conservativos sejam menores do que os dos casos em que o teste é exato.

Para $k = 5$ (Tabela 4.1), os testes foram liberais quando a condição $p \leq 2(r-1)$ não foi atendida, tanto para $\alpha = 1\%$ quanto para $\alpha = 5\%$. Foi afirmado anteriormente que a substituição de $(\mathbf{S}_i + \mathbf{S}_{i'})^{-1}$ por $(\mathbf{S}_i + \mathbf{S}_{i'})^g$ poderia alterar as taxas de erro tipo I por experimento do teste. Essa alteração foi no sentido de o teste apresentar taxas significativamente (valor- $p < 0,01$) superiores aos valores nominais de 1% e de 5% . Dessa forma, o teste foi considerado liberal e apresentou taxas de erro tipo I por experimento que se situaram entre o dobro e um pouco mais que o triplo do valor nominal α , tanto para 1% quanto para 5% .

Na Tabela 4.2 estão apresentados as taxas de erro tipo I por experimento para $k = 5$, $r = 3$, $p = 5$ e 10 , $\delta = 1, 2, 8$ e 16 , considerando

TABELA 4.1: Taxas de erro tipo I, por experimento, em porcentagem em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (*bootstrap*) e $N = 2.000$ simulações Monte Carlo, $k = 5$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, sem covariância comum.

r	p	δ	α	
			1%	5%
3	2	1	0,00 ⁺⁺	0,70 ⁺⁺
		2	0,05 ⁺⁺	1,40 ⁺⁺
		8	0,00 ⁺⁺	0,70 ⁺⁺
		16	0,00 ⁺⁺	0,85 ⁺⁺
	5	1	2,95 ^{**}	15,10 ^{**}
		2	3,35 ^{**}	14,90 ^{**}
		8	3,60 ^{**}	14,50 ^{**}
		16	2,80 ^{**}	14,30 ^{**}
	10	1	2,45 ^{**}	10,90 ^{**}
		2	2,80 ^{**}	11,45 ^{**}
		8	2,05 ^{**}	11,00 ^{**}
		16	2,20 ^{**}	11,25 ^{**}
10	2	1	0,85	4,60
		2	0,80	4,65
		8	1,45	6,40 ^{**}
		16	1,50	6,85 ^{**}
	5	1	1,15	5,00
		2	1,15	5,15
		8	1,25	5,80
		16	1,20	6,00
	10	1	0,50 ⁺⁺	4,70
		2	0,45 ⁺⁺	4,55
		8	0,30 ⁺⁺	4,65
		16	0,45 ⁺⁺	4,50

⁺⁺ e ^{**} significativamente (valor- $p < 0,01$) inferior e superior, respectivamente, aos valores nominais de significância $\alpha = 5\%$ e $\alpha = 1\%$.

$\alpha = 1\%$ e 5% . Neste caso, foram utilizadas 1.000 simulações e 1.000 reamostragens *bootstrap* e as distâncias entre médias de populações foram calculadas pela expressão (3.7), que leva em conta a matriz \mathbf{S}_p . Em todos

os casos, $p \leq k(r - 1)$, mas \mathbf{S}_p^{-1} foi substituída por \mathbf{S}_p^g , que é uma inversa generalizada de \mathbf{S}_p . Isso foi necessário pois, quando se obtém uma amostra *bootstrap*, uma observação multivariada qualquer \mathbf{Y}_{ij} pode ser reamostrada mais de uma vez na mesma população. Assim, o posto de \mathbf{S}_i , que é dado pelo $\min(r - 1, p)$, fica reduzido de uma unidade e o posto de \mathbf{S}_p , que é $\min(k(r - 1), p)$, pode ser reduzido de 1 unidade se $k(r - 1) \leq p$. A reamostragem de uma observação em uma dada população pode se repetir, no máximo, r vezes e pode acontecer em mais de 1 população. Isso pode reduzir consideravelmente o posto de \mathbf{S}_p , principalmente se k e r forem pequenos. Todas as situações em que o posto de \mathbf{S}_p for inferior a p , a inversa generalizada deve ser utilizada para viabilizar a aplicação do teste.

TABELA 4.2: Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (*bootstrap*) e $N = 1.000$ simulações Monte Carlo, $k = 5$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.

r	p	δ	α		
			1%	5%	
3	5	1	0,20 ⁺⁺	3,40	
		2	0,20 ⁺⁺	3,70	
		8	0,20 ⁺⁺	3,20 ⁺⁺	
		16	0,10 ⁺⁺	3,60	
	10	5	1	35,80 ^{**}	71,50 ^{**}
			2	35,10 ^{**}	70,10 ^{**}
		10	8	35,70 ^{**}	72,40 ^{**}
			16	34,20 ^{**}	69,50 ^{**}

⁺⁺ e ^{**} significativamente (valor- $p < 0,01$) inferior e superior, respectivamente, aos valores nominais de significância $\alpha = 5\%$ e $\alpha = 1\%$.

As taxas de erro tipo I observadas para $k = 5$, usando o método de *bootstrap* com covariância combinada \mathbf{S}_p , foram significativamente (valor- $p < 0,01$) inferiores ao valor nominal de 1% para $r = 3$ e $p = 5$, com qualquer δ . No caso de $r = 3$ e $p = 5$, o teste teve o mesmo tamanho que o

valor nominal $\alpha = 5\%$, exceto para $\delta = 8$. Já no caso de $r = 3$ e $p = 10$, em que $k(r - 1) = p$, as taxas de erro foram muito elevadas e significativamente (valor- $p < 0,01$) superiores aos valores nominais de 1% e 5%. Este foi um exemplo em que a reamostragem de *bootstrap* levou as estimativas da matriz de covariâncias comum com postos inferiores a p . Assim, houve a necessidade de se utilizar inversas generalizadas de \mathbf{S}_p na expressão (3.7), o que explicou a alteração nas taxas de erro tipo I. Então, se $p \geq k(r - 1)$ ou se p for menor que $k(r - 1)$, mas próximo de $k(r - 1)$, não se recomenda utilizar o teste *bootstrap* de comparações múltiplas multivariadas, devido ao caráter liberal do mesmo.

Para solucionar este problema, nas situações em que $p \leq k(r - 1)$, mas nas proximidades da desigualdade, foi proposto o teste de comparações múltiplas multivariadas via permutação, no qual as reamostragens são feitas sem reposição. Assim, em nenhum caso, uma mesma observação vai ser reamostrada mais de uma vez numa mesma população. Na Tabela 4.3 estão apresentados os resultados das taxas de erro tipo I por experimento, para $k = 5$, $r = 3$ e 10 , $p = 2, 5$ e 10 , $\delta = 1, 2, 8$ e 16 , considerando $\alpha = 1\%$ e 5% . Neste caso, foram utilizadas 1.000 simulações e 1.000 reamostragens sem reposição.

Os resultados para as taxas de erro tipo I do teste de comparações múltiplas multivariadas foram excelentes, pois praticamente todos os valores observados não diferiram significativamente dos níveis nominais. Isso indica que o teste possui tamanho exato. Houve duas exceções, sendo uma para $\alpha = 1\%$ e outra para $\alpha = 5\%$. Embora o teste tenha sido considerado liberal nestas duas situações, a diferença para os valores nominais não foi expressiva. O teste baseado em permutação teve melhor desempenho do que o teste *bootstrap* nestas situações estudadas.

Para situações em que $p > k(r - 1)$, não há solução adequada, independente de se estar utilizando reamostragem com ou sem reposição. Uma sugestão é que se faça a seleção de um subconjunto de variáveis de tamanho $s < p$, que obedecerá à relação $s \leq k(r - 1)$. Assim, o problema será paleativamente resolvido. Caso contrário, podem-se utilizar inversas gene-

TABELA 4.3: Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens sem reposição (permutação) e $N = 1.000$ simulações Monte Carlo, $k = 5$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.

r	p	δ	α	
			1%	5%
3	2	1	1,50	5,80
		2	1,20	6,10
		8	2,20**	6,50
		16	1,60	6,70
	5	1	0,80	4,90
		2	1,50	5,90
		8	1,00	5,50
		16	0,60	6,10
10	10	1	1,10	4,80
		2	0,80	5,60
		8	1,00	5,30
		16	1,20	6,20
	2	1	1,00	7,20**
		2	1,20	4,90
		8	1,60	6,00
		16	2,10	7,00
10	5	1	1,20	5,40
		2	1,30	6,10
		8	1,00	5,50
		16	1,60	6,70
	10	1	1,20	4,80
		2	0,90	5,50
		8	1,00	6,00
		16	1,10	5,00

** significativamente (valor- $p < 0,01$) superior aos valores nominais de significância $\alpha = 5\%$ e $\alpha = 1\%$.

realizadas, mas estando ciente de que as taxas de erro tipo I por experimento são muito elevadas e sempre superiores aos valores nominais. Outra alternativa é reduzir a dimensionalidade por meio de componentes principais. Assim, um número $s < p$ de componentes explicando grande parte da variação total contida nas p variáveis originais seria utilizado. A dificuldade

seria no instante de obter as conclusões.

Na Tabela 4.4 estão apresentadas as taxas de erro tipo I por experimento, em função de r , p e δ , para $\alpha = 1\%$ e 5% , amostragem com reposição. Para $r = 3$ o que se observou foram taxas de erros sempre significativamente ($P < 0,01$) inferiores aos valores nominais. Para os demais casos, observou-se uma tendência de taxas que vão da faixa conservativa para a faixa liberal na medida em que a heterogeneidade de covariâncias aumentou. Várias exceções, no entanto, foram observadas. Com $r = 20$ e $p = 5$ e $p = 10$, independentemente de δ e de α , o teste foi sempre exato. Embora as taxas de erro, quando significativamente ($P < 0,01$) foram consideradas superiores aos valores nominais, as diferenças para o real valor α foram pequenos em magnitude. As piores situações ocorreram com $r = 10$ e $p = 2$.

Nas situações de homogeneidade ($\delta = 1$), o teste foi sempre exato, exceto para $r = 3$ e $p = 2$, em que foi conservativo. As piores situações são casos muito atípicos nos exemplos reais, pois a heterogeneidade de covariâncias máxima é de $\delta = 16$. Para situações similares a essa, Machado et al. (2005) mostraram que as taxas de erro dos testes convencionais de Tukey e SNK univariados são muito elevadas, superando 50%, muitas vezes. Assim, os resultados encontrados no presente trabalho, para o caso multivariado, são extremamente promissores e compatíveis com as alternativas univariadas de Machado et al. (2005) para os testes Tukey e SNK via *bootstrap*.

Nas Tabelas 4.5 e 4.6 estão apresentadas as taxas de erro tipo I por experimento para $k = 50$ e $k = 100$ populações, respectivamente, considerando amostragens com reposição para se aplicar o teste. Padrões de respostas similares aos apresentados para o caso de $k = 10$ populações foram observados. Para as situações de homogeneidade de covariâncias ($\delta = 1$), o teste foi exato ou conservativo ($k = 50$, $r = 10$, $p = 2$ e $\alpha = 1\%$). Esta situação ($\delta = 1$) é considerada ideal para a aplicação do teste de comparação múltipla. Na medida em que a heterogeneidade de covariâncias aumentou, houve uma tendência do teste apresentar resultados liberais. Em todas as situações em que $\delta \geq 8$, o teste foi considerado liberal, exceto para $k = 50$,

TABELA 4.4: Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (*bootstrap*) e $N = 1.000$ simulações Monte Carlo, $k = 10$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.

r	p	δ	α	
			1%	5%
3	2	1	0,00 ⁺⁺	0,90 ⁺⁺
		2	0,00 ⁺⁺	0,90 ⁺⁺
		8	0,00 ⁺⁺	0,60 ⁺⁺
		16	0,00 ⁺⁺	1,20 ⁺⁺
10	2	1	1,40	5,10
		2	0,80	6,30
		8	2,70 ^{**}	7,50 ^{**}
		16	2,80 ^{**}	9,50 ^{**}
	5	1	1,00	4,90
		2	0,55	4,50
		8	0,80	4,70
		16	1,65	6,70 ^{**}
20	2	1	1,10	4,40
		2	0,30 ⁺⁺	3,60 ⁺⁺
		8	2,00 ^{**}	6,80 ^{**}
		16	1,20	6,00
	5	1	0,90	4,80
		2	1,20	4,80
		8	1,40	4,50
		16	1,10	4,90
	10	1	1,00	4,90
		2	1,00	5,40
		8	1,00	5,80
		16	1,50	5,20

⁺⁺ e ^{**} significativamente (valor- $p < 0,01$) inferior e superior, respectivamente, aos valores nominais de significância $\alpha = 5\%$ e $\alpha = 1\%$.

$r = 10$, $p = 5$ e $\delta = 8$, independentemente do valor de α .

As taxas de erro tipo I por experimento, quando o teste foi considerado liberal, tenderam a aumentar com o aumento do número de populações (k) (Tabelas 4.4, 4.5 e 4.6). Outro aspecto identificado foi a tendência de diminuição das taxas de erro tipo I com o aumento do número de variáveis nas

TABELA 4.5: Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (*bootstrap*) e $N = 1.000$ simulações Monte Carlo, $k = 50$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.

r	p	δ	α	
			1%	5%
3	2	1	0,90	5,00
		2	1,20	5,80
		8	2,00**	8,40**
		16	3,60**	11,30**
10	2	1	0,50 ⁺⁺	4,20
		2	1,00	4,20
		8	3,00**	11,30**
		16	3,30**	10,00**
	5	1	1,20	4,70
		2	1,10	6,60**
		8	1,20	5,70
		16	2,10**	7,90**

⁺⁺ e ** significativamente (valor- $p < 0,01$) inferior e superior, respectivamente, aos valores nominais de significância $\alpha = 5\%$ e $\alpha = 1\%$.

situações heterocedásticas e liberais. Embora isso tenha sido constatado, é interessante que se ressalte que este padrão de comportamento, em relação ao desempenho do teste, não é muito consistente.

4.2 Poder

Nas Figuras de 4.1 a 4.4 estão apresentados os valores de poder do teste *bootstrap* de comparações múltiplas multivariadas em função da diferença entre duas médias populacionais, dada pelo número de erros padrões ϕ da diferença de duas médias e do nível nominal de significância α para diferentes valores de k , r , p e δ . Em todos os casos, o valor de ρ foi fixado em 0,5, após verificação de que outros valores não afetaram o poder (resultados não apresentados). Os resultados apresentados nas Figuras 4.1 e 4.2

TABELA 4.6: Taxas de erro tipo I, por experimento, em porcentagem, em função de r repetições; p variáveis e δ para distribuição normal, considerando $B = 1.000$ reamostragens com reposição (*bootstrap*) e $N = 1.000$ simulações Monte Carlo, $k = 100$ populações, $\rho = 0,5$ para $\alpha = 5\%$ e $\alpha = 1\%$, com covariância comum.

r	p	δ	α	
			1%	5%
3	2	1	1,10	5,50
		2	1,50	7,20**
		8	2,80**	10,30**
		16	2,80**	12,20**
10	2	1	1,00	5,70
		2	1,20	6,10
		8	4,90**	13,10**
		16	4,20**	14,60**
	5	1	1,10	5,30
		2	1,10	5,50
		8	2,40**	7,30**
		16	2,40**	8,50**

** significativamente (valor- $p < 0,01$) superior aos valores nominais de significância $\alpha = 5\%$ e $\alpha = 1\%$.

foram obtidos utilizando-se o método *bootstrap*, considerando a estatística $T_{ii'}^2$, computada com $(\mathbf{S}_i + \mathbf{S}_{i'})^{-1}$, equação (3.5). Os demais casos diferiram pelo fato de $T_{ii'}^2$ ter sido computada utilizando \mathbf{S}_p^{-1} , equação (3.7).

Nas Figuras 4.1 (a) e (b), pode-se observar o comportamento do teste de comparações múltiplas multivariadas via *bootstrap*, para $k = 5$, $r = 3$ e $p = 2$, considerando $\delta = 1$ e $\delta = 16$, respectivamente, em relação ao poder. Considerando os valores nominais de significância de 5% (linha pontilhada) e de 1% (linha cheia), verifica-se que houve grande superioridade do teste de maior poder, quando adotou-se o valor $\alpha = 5\%$ em relação à situação de $\alpha = 1\%$, na medida que aumentava-se a diferença entre os vetores de médias. Para ambos os casos, sob H_0 (Tabela 4.1), nestas situações, o teste foi extremamente conservativo. Isso, em parte, explica os baixos valores de poder encontrados.

A diferença entre os valores de poder para $\alpha = 5\%$ e $\alpha = 1\%$ foi muito grande sob homogeneidade de variâncias (Figura 4.1 (a)). Para $\alpha = 1\%$, não houve, mesmo com $\phi = 16$, valores de poder que superassem o nível nominal significativamente (valor- $p < 0,01$). Comportamento semelhante foi observado para $\delta = 16$ (Figura 4.1 (b)).

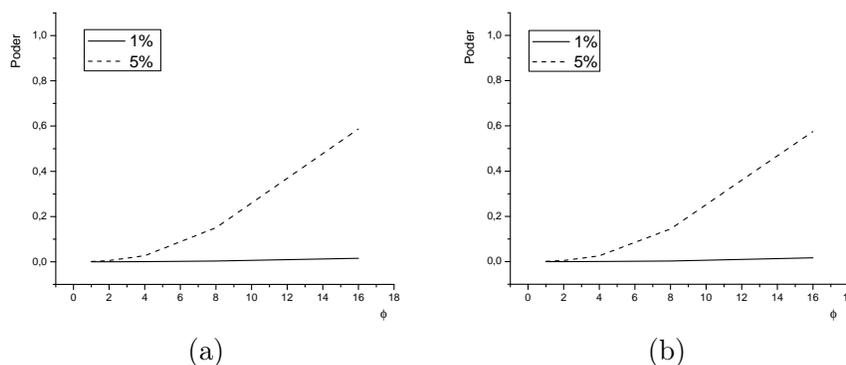


FIGURA 4.1: Poder do teste *bootstrap* de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 5$, $r = 3$, $\rho = 0,5$, $p = 2$ e (a) $\delta = 1$ e (b) $\delta = 16$.

A comparação dos valores de poder entre as situações homocedástica e heterocedástica (Figuras 4.1 (a) e (b)) não mostrou diferenças expressivas no desempenho do teste. Isso confirma, de certa forma, a robustez que é atribuída aos testes realizados via *bootstrap*. Em ambos os casos, o valor máximo de poder foi observado para $\alpha = 5\%$, com $\phi = 16$ erros padrões de diferença entre dois vetores de médias e aproximou-se de 60%.

Na Figura 4.2 estão apresentados os valores de poder em função de ϕ , para uma situação com $k = 5$, $p = 5$ e $r = 10$. Como era esperado, o poder aumentou com o aumento de ϕ e o valor de 100% foi atingido com $\phi = 8$. A heterogeneidade de covariâncias praticamente não afetou o poder do teste. Os valores de poder para $\alpha = 5\%$ foram superiores aos de $\alpha = 1\%$ para ϕ , variando de 1 a 8. Além deste valor, os valores de poder igualaram-se entre si e a 100%.

Nas Figuras 4.3 a 4.4 são apresentados os valores de poder em função

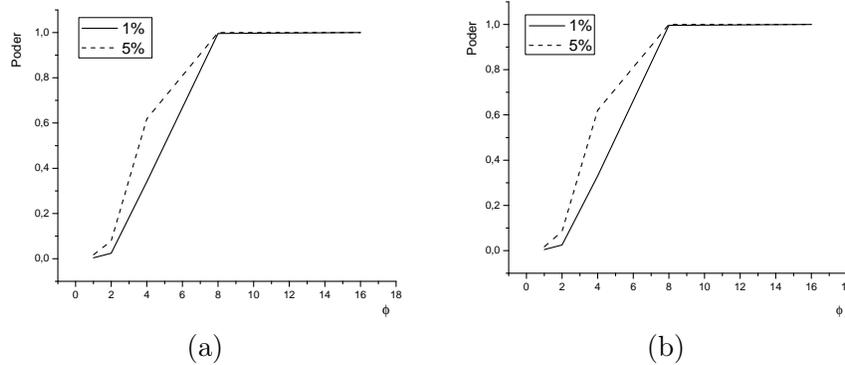


FIGURA 4.2: Poder do teste *bootstrap* de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 5$, $r = 10$, $\rho = 0,5$, $p = 5$ e (a) $\delta = 1$ e (b) $\delta = 16$.

de ϕ e α , para $k = 100$, com $r = 3$ e 10 e $\delta = 1$ e 16 . O padrão de resposta é o mesmo observado anteriormente na Figura 4.2. O número de populações afetou o poder do teste *bootstrap* de comparações múltiplas multivariadas. Os valores de poder encontrados no presente trabalho, no contexto multivariado, são comparáveis aos valores de poder encontrados por Silva (1999), para o teste Scott e Knott, univariado e superiores aos de Ramos (2007), para o teste Calinski & Corsten original e *bootstrap*.

4.3 Exemplo real

Na Tabela 4.7 estão apresentados os dados de areia e argila de seis sistemas de uso da terra (SUTs) da Amazônia, considerando $r = 10$ repetições, caracterizados por Fidalgo et al. (2005), quanto à cobertura vegetal e ao uso atual, como: (Agri) agricultura, (Agro) agrofloresta, (CN) capoeira nova, (CV) capoeira velha, (F) floresta e (P) pastagem. O objetivo da utilização desses dados foi aplicar o teste de comparação múltipla multivariada proposto neste trabalho. Cada ponto de amostragem principal fazia parte de uma malha pré-estabelecida, distribuída em 6 janelas com cerca de 9 ha cada e os pontos distavam entre si, de modo geral, 100 m, com alguns pontos

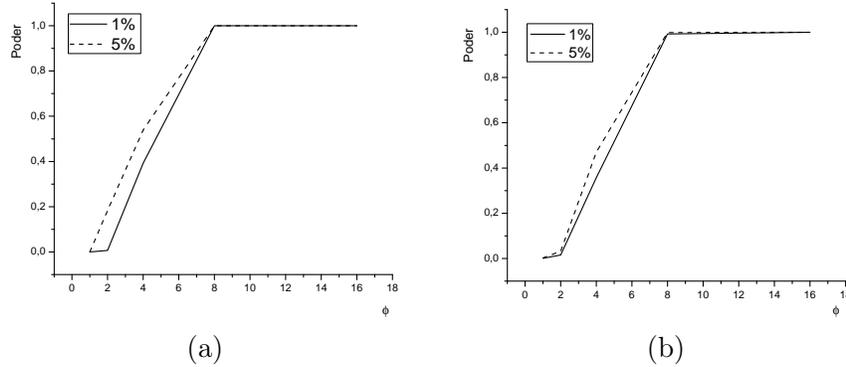


FIGURA 4.3: Poder do teste *bootstrap* de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 100$, $r = 3$, $\rho = 0,5$, $p = 2$ e (a) $\delta = 1$ e (b) $\delta = 16$.

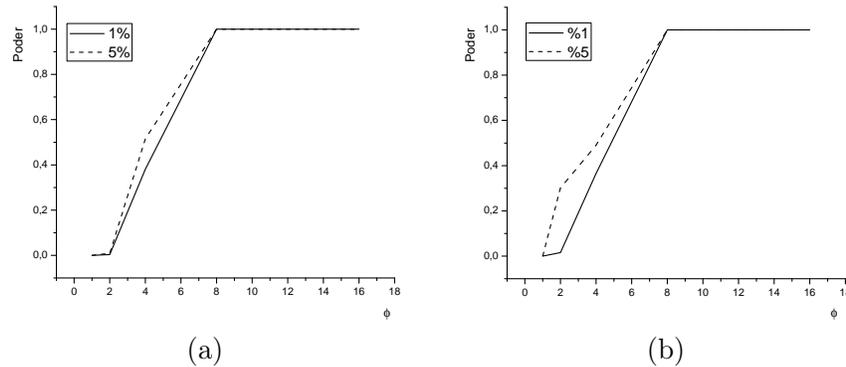


FIGURA 4.4: Poder do teste *bootstrap* de comparações múltiplas multivariadas, em função da diferença entre duas médias ϕ das populações e do nível nominal de significância α , para $k = 100$, $r = 10$, $\rho = 0,5$, $p = 5$ e (a) $\delta = 1$ e (b) $\delta = 16$.

distantes a 50 m.

Antes da realização do teste de comparações múltiplas multivariadas, fixou-se o nível nominal de significância em 5%. Foram estimados os vetores de médias (Tabela 4.8) e as matrizes de covariâncias de cada população.

TABELA 4.7: Teores de areia e argila referentes ao sistema de uso da terra na Amazônia

SUTs	Rep.	Variáveis		SUTs	Rep.	Variáveis	
		Areia	Argila			Areia	Argila
Agri	1	11	42	Agro	1	24	29
	2	10	37		2	29	30
	3	13	35		3	12	55
	4	30	31		4	15	43
	5	7	52		5	22	46
	6	11	46		6	16	50
	7	16	30		7	10	51
	8	11	42		8	19	43
	9	11	43		9	14	42
	10	16	32		10	28	36
CN	1	11	38	CV	1	7	45
	2	13	47		2	27	31
	3	18	34		3	26	34
	4	16	49		4	14	36
	5	11	45		5	8	45
	6	30	32		6	7	42
	7	5	64		7	9	43
	8	7	59		8	17	34
	9	11	50		9	6	47
	10	17	38		10	11	33
F	1	14	33	P	1	26	32
	2	10	41		2	20	32
	3	16	44		3	20	29
	4	17	32		4	12	35
	5	12	36		5	24	29
	6	14	42		6	10	45
	7	27	34		7	46	26
	8	16	35		8	47	43
	9	9	46		9	15	41
	10	5	48		10	50	24

$$S_1 = \begin{bmatrix} 40,4889 & -32,2222 \\ -32,2222 & 51,7778 \end{bmatrix},$$

$$S_2 = \begin{bmatrix} 43,8778 & -47,3889 \\ -47,3889 & 75,3889 \end{bmatrix},$$

TABELA 4.8: Vetores de médias dos sistema de uso da terra (SUTs) na Amazônia.

Variáveis	SUTs					
	Agri	Agro	CN	CV	F	P
Areia	13,6	18,9	13,9	13,2	14,0	27,0
Argila	39,0	42,5	45,6	39,5	39,1	33,6

$$\mathbf{S}_3 = \begin{bmatrix} 49,2111 & -58,8222 \\ -58,8222 & 109,6000 \end{bmatrix}, \quad \mathbf{S}_4 = \begin{bmatrix} 60,8444 & -39,4444 \\ -39,4444 & 31,3889 \end{bmatrix}, \\
 \mathbf{S}_5 = \begin{bmatrix} 34,6667 & -22,4444 \\ -22,4444 & 33,6556 \end{bmatrix} \quad \text{e} \quad \mathbf{S}_6 = \begin{bmatrix} 228,4444 & -46,8889 \\ -46,8889 & 52,4889 \end{bmatrix}.$$

Calculou-se a matriz de covariância combinada por:

$$\mathbf{S}_p = \begin{bmatrix} 76,2556 & -41,2019 \\ -41,2019 & 59,0500 \end{bmatrix}.$$

Na Tabela 4.9 encontra-se um sumário dos valores da estatística dos testes nas 10.000 reamostragens de *bootstrap* e permutação realizadas para o máximo, considerando todos os pares de vetores de médias, de acordo com a equação (3.9). Os valores da estatística do teste de comparações múltiplas multivariadas para o sistema de uso da terra na Amazônia estão dispostos na Tabela 4.10, considerando a expressão (3.7) e os dados originais.

Na Tabela 4.11 estão apresentados os valores- p , obtidos utilizando-se a expressão (3.11) para os testes *bootstrap* (acima da diagonal) e permutação (abaixo da diagonal). Os valores originais da estatística (Tabela 4.10) são comparados com os valores da estatística na distribuição de *bootstrap* ou de permutação (Tabela 4.9) para se determinar os valores- p . Na Figura 4.5 estão apresentados os estimadores *Kernel* das densidades da distribui-

TABELA 4.9: Resumo dos valores da estatística Γ_m *bootstrap* e Γ_m^* permutação, com B=10.000 reamostragens.

Amostragens (B)	Γ_m	Γ_m^*
1	7,38	8,80
2	2,45	7,57
3	5,84	3,14
⋮	⋮	⋮
10000	4,35	4,25

TABELA 4.10: Valores da estatística do teste de comparações múltiplas multivariadas para cada par de vetores de médias dos SUTs.

	Agri	Agro	CN	CV	F	P
Agri	0,00	7,35	6,22	0,02	0,02	12,23
Agro		0,00	1,66	7,15	6,55	7,08
CN			0,00	5,74	5,65	14,54
CV				0,00	0,04	12,82
F					0,00	11,39
P						0,00

ção de reamostragem *bootstrap* e permutação dos máximos (Γ_m e Γ_m^*) da estatística $T_{ii'}^2$, respectivamente. Essas densidades refletem distribuições assimétricas à direita e foram utilizadas para a obtenção dos valores- p associados aos valores da estatística dos testes. Assim, por exemplo, os vetores de médias de P e da CN apresentaram T^2 igual a 14,54. As áreas sob as curvas, nas Figuras 4.5 (a) e (b), entre 14,54 e $+\infty$, são 0,0196 e 0,0176, respectivamente (Tabela 4.11).

Pode-se observar, tanto por *bootstrap* quanto por permutação, quais vetores de médias foram significativamente (valor- $p < 0,05$) diferentes entre si. Considerando o valor nominal α de 5%, mencionado no início desta seção, verificou-se que os SUTs Agri e P, CN e P e CV e P apresentaram vetores de médias significativamente (valor- $p < 0,05$) diferentes entre si, tanto para o teste *bootstrap* quanto para o teste de permutação. O SUT pastagem esteve presente em todos os três pares, cujos vetores de médias diferiram. Este resultado corrobora com aqueles encontrados por Nobrega (2006), que utilizou

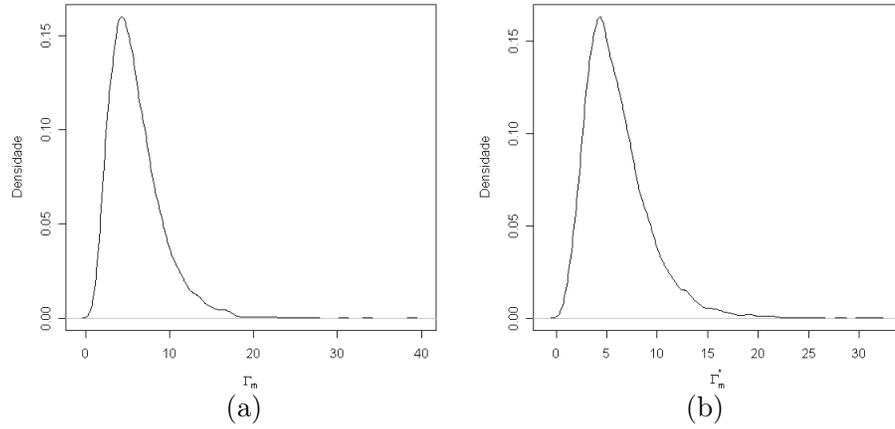


FIGURA 4.5: Densidade da distribuição de (a) Γ_m via *bootstrap* e (b) Γ_m^* via permutação.

TABELA 4.11: Valores- p para cada par de vetores de médias dos SUTs obtidos por *bootstrap* (acima da diagonal) e por permutação (abaixo da diagonal).

	Agri	Agro	CN	CV	F	P
Agri		0,2704	0,3833	1,0000	1,0000	0,0477
Agro	0,2703		0,9797	0,2867	0,3461	0,2934
CN	0,3887	0,9807		0,4470	0,4589	0,0196
CV	1,0000	0,2885	0,4536		1,0000	0,0380
F	1,0000	0,3499	0,4636	1,0000		0,0636
P	0,0433	0,2945	0,0176	0,0349	0,0608	

outras técnicas multivariadas. Uma segunda etapa que deveria ser contemplada direcionaria os esforços na identificação dos componentes do vetor de médias que seriam responsáveis pela diferença significativa encontrada. Este teste já foi planejado e passos para sua execução podem ser vistos a partir da equação (3.12) até a equação (3.15).

5 CONCLUSÕES

Os testes foram propostos com sucesso e os passos necessários para a execução deles foram implementados em R.

Nas situações homocedásticas, os testes tiveram desempenho considerado ótimo, com taxas de erro tipo I, por experimento, inferiores ou iguais aos valores nominais adotados, quando $p \leq k(r - 1)$.

Os testes foram liberais nas situações heterocedásticas.

Recomenda-se a utilização de procedimentos de seleção de variáveis ou de componentes principais nas situações em que $p > k(r - 1)$.

O teste de permutação tem melhor desempenho no controle do erro tipo I por experimento quando $p = k(r - 1)$ ou $p < k(r - 1)$, mas está na vizinhança dessa desigualdade.

O teste de comparações múltiplas multivariadas via *bootstrap* possui valores de poder comparáveis ou superiores aos dos testes univariados.

O teste mostrou robustez no desempenho do poder quando a suposição de homocedasticidade for violada.

6 REFERÊNCIAS BIBLIOGRÁFICAS

BANZATO, D. A.; KRONKA, S. N. **Experimentação Agrícola**. Jaboticabal, SP: FUNEP, 2006. 237 p.

BORGES, L. C.; FERREIRA, D. F. Poder e taxas de erro tipo I dos testes Scott-Knott, Tukey e Student-Newman-Keuls sob distribuição normais e não normais dos resíduos. **Revista de Matemática e Estatística**, São Paulo, v. 21, p. 67-83, 2003.

CARARI, M. L. **Intervalo de confiança pra diferença entre duas proporções binomiais utilizando *Bootstrap* infinito**. 2004 66 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras, MG.

CARMER, S. G. Optimal significance levels for application of the least significant difference in crop performance trials. **Crop Science**, Madison, v. 16, p. 95-99, Jan./Feb. 1976.

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

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 Agronomia) - Universidade Federal de Lavras, Lavras, MG.

DACHS, J. N. W. **Estatística computacional: uma introdução em Turbo Pascal**. Rio de Janeiro, 1988. 236 p.

DIACONIS, P.; EFRON, B. Computer-intensive methods in statistics. **Scientific American**, v. 248, p. 116-130, 1983.

DUNCAN, D. B. Multiple range and multiple F tests. **Biometrics**, Washington, v. 11, p. 1-41, Mar. 1955.

EFRON, B. e TIBSHIRANI, R. Statistical data analysis in the computer age. **Science**, v. 253, p. 390-395, 1991..

FERREIRA, D. F. **Estatística multivariada**. Lavras: UFLA, 2007. 642 p. No prelo.

FIDALGO, E. C. C. et al. **Levantamento do uso e cobertura da terra de seis áreas amostrais relacionadas ao projeto ‘‘Conservation and sustainable management of below-ground biodiversity: Phase 1’’, Município de Benjamin Constant (AM)**. Rio de Janeiro: Embrapa Solos, 2005. Apostila.

FISHER, W. K. On grouping for maximum homogeneity. **Journal American Statistical Association**, Washington, v. 53, p. 789-798, Dec. 1958.

GOOD, P. I. **Permutation tests: a practical guide to resampling methods for testing hypothesis**. New York: Springer-Verlag, 1994.

GUERRA, M. J.; DONAIRE, D. **Estatística indutiva: teoria e exercícios**. 2.ed. São Paulo: Livraria Ciência e Tecnologia, 1982. 311 p.

HARTER, H. L. Errors rates and sample size for range tests in multiple comparisons. **Biometrics**, Washington, v. 13, p. 511-536, Dec. 1957.

HOCKING, R. R. **The analysis of linear models**. Belmont: Brooks/Cole, 1985. 385 p.

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

MACHADO, A. A.; DEMÉTRIO, C. G. B.; FERREIRA, D. F.; SILVA, J. G. C. **Estatística Experimental: uma abordagem fundamentada no planejamento e no uso de recursos computacionais**. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11; Londrina, 2005. **Cursos...** Londrina: Universidade Estadual de Londrina, 2005. p 249-286.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. London: Academic, 1979.

NAYLOR, T.; BALINTFY, J.; BURDICK, D.; CHU, K. **Técnicas de simulação em computadores**. São Paulo: Vozes, 1971.

NEGRILLO, B. G.; PERRE, M. A. Métodos multivariados e aplicações. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 32.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 2; Londrina, 1987. **Cursos...** Londrina: Universidade Estadual de Londrina, 1987. p 78-83.

NOBREGA, R. S. A. **Efeito de sistemas de uso da terra na Amazônia sobre atributos do solo, ocorrência, eficiência e diversidade de bactérias que nodulam caupi [Vigna unguiculata (L.) Walp]**. 2006. 188 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras.

O'NEILL, R.; WETHERILL, G. B. The present state of multiple comparison methods. **Journal of the Royal Statistical Society**, Oxford, v. 33, n. 2, p. 218-250, 1971.

PEARCE, S. C. Data analysis in agricultural experimentation. III. Multiple comparisons. **Experimental Agriculture**, v. 29, p. 1-8, 1993.

PERCIN, D.; BARBOSA, J. C. Uma avaliação de seis procedimentos para comparações múltiplas. **Revista de Matemática e Estatística**, Marília, v. 6, p. 95-103, 1988.

PERCIN, D.; MALHEIROS, E. B. Curso: Procedimentos para comparação múltiplas. IN: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 3; REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMÉTRIA, 34, Lavras, 1989. **Curso...** Lavras: ESAL, 1989. 42 – 65 p.

PETERSEN, R. G. Use and misuse of multiple comparison procedures. **Agronomy Journal**, Madison, v. 69, p. 205-208, Mar./Apr. 1977.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 11.ed. Piracicaba: Nobel, 1985. 468 p.

PLLANA, S. **History of Monte Carlo method**. 2002. Disponível em: <<http://stud2.tuwien.ac.at/~e9527412/index.html>>. Acesso em 10 nov. 2007.

R DEVELOPMENT CORE TEAM . **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2006. Disponível em: <<http://www.R-project.org>>. Acesso em: 05 nov. 2007.

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. **Agrupamento de médias via *bootstrap* para populações normais e não-normais**. 2007. 76 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Lavras, Lavras, MG..

RENCHER, A. C. **Methods of multivariate analysis**. 2002. New York: J. Wiley, 2002. 708 p.

SAMPAIO, I. B. M. **Estatística aplicada à experimentação animal**. Belo Horizonte: FEP-MVZ/UFMG, 1998. 221 p.

SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. **Biometrika**, Langhorne, v. 40, n. 1, p. 87-104, 1953.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, v. 30, n. 3, p. 507-512, Sept. 1974.

SILVA, E. C. ; FERREIRA, D. F. ; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v. 23, n. 2, p. 687-696, 1999.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of**

statistics. 2.ed. New York: McGraw-Hill Book, 1980. 633 p.

TIMM, N. H. **Applied multivariate analysis.** New York: Springer-Verlag, 2002. 693 p.

ANEXOS

ANEXOS A	Páginas
PROGRAMA 1A Programa de comparações múltiplas multivariadas via <i>boots-trap</i> ou permutação.	48

PROGRAMA 1A: Programa de comparações múltiplas multivariadas via *bootstrap* ou permutação.

```

# Esta função retorna um objeto denominado sumar a partir da matriz de
# dados neste objeto a primeira coluna contém os k vetores de médias - um
# a cada p linhas e as demais p colunas com as covariâncias das k populações
# também uma a cada p linhas
dados.sumario = function(dados,r,k)
{
  ybar = apply(dados[1:r,2:(p+1)],2,mean)
  covi = var(dados[1:r,2:(p+1)])
  sumar = cbind(ybar,covi)
  for (i in 2:k)
  {
    ybar = apply(dados[((i-1)*r+1):(i*r),2:(p+1)],2,mean)
    covi = var(dados[((i-1)*r+1):(i*r),2:(p+1)])
    sumari = cbind(ybar,covi)
    sumar = rbind(sumar,sumari)
  }
  return(sumar)
}

# Esta função retorna as T2il de todos os pares entre os k populações
# recebe r, k, p e uma matriz do tipo sumário. O resultado será uma matriz
# T2 k x k com os T2il.
T2il = function(sumar,r,k,p)
{
  T2 = matrix(0,k,k)
  Yb = matrix(sumar[,1],k,p,byrow=T)
  Aux = matrix(t(sumar[,2:(p+1)]),k,p*p,byrow=T)
  Sp = matrix(apply(Aux,2,sum),p,p)/k
  G=r/2*(Yb%*%ginv(Sp)%*%t(Yb))
  for (i in 1:(k-1))
  for (ell in (i+1):k)
  {
    T2[i,ell] = G[i,i]-2*G[i,ell]+G[ell,ell]
    T2[ell,i] = T2[i,ell]
  }
  return(list(T2=T2,Sp=Sp))
}

```

```

}

# Esta função reamostra o conjunto de dados originais, montando a amostra
# bootstrap do experimento, com a mesma estrutura dos dados simulados
reamostragem = function(dados,r,k,boo)
{
  #ii = trunc(runif(r*k)*r*k) + 1
  ii = sample(1:(r*k), r*k, replace = boo)
  return(dados[ii,])
}

# função retorna os valores-p na parte inferior da matriz T2 original
# parte superior com os valores T2ii'
valorp = function(T2,Gamab,k,B)
{
  for (ii in 1:(k-1))
  {
    for (kk in (ii+1):k)
    {
      T2[kk,ii] =length(Gamab[Gamab >= T2[ii,kk]])/B
    }
  }
  return(T2)
}

# função para realizar amostras de bootstrap de tamanhos B
boot = function(dados,r,k,p,B,boo)
{
  dadosb = reamostragem(dados,r,k,boo)
  sumarb = dados.sumario(dadosb,r,k)
  T2b = T2il(sumarb,r,k,p)
  Gama = max(T2b$T2)
  for (b in 2:B)
  {
    dadosb = reamostragem(dados,r,k,boo)
    sumarb = dados.sumario(dadosb,r,k)
    T2b = T2il(sumarb,r,k,p)
    Gammab = max(T2b$T2)
    Gama = rbind(Gama,Gammab)
  }
}

```

```

    }
    return(Gama)
}

# Utilização das funções para análise
library(mvtnorm);library(MASS)
r=10;k=6;p=2;sig2=10;rho=0.5;delta=16;dist=1;boo=TRUE
# FALSE para permutação e TRUE para bootstrap
B=10000
dados=read.table("DadosRafa.txt",header=T) sumar = dados.sumario(dados,r,k)
T2 = T2il(sumar,r,k,p)
Sp=T2$Sp
Gamab = boot(dados,r,k,p,B,boo)
T2 = valorp(T2$T2,Gamab,k,B)
sumar
Sp
T2
Gamab[1];Gamab[2];Gamab[3];Gamab[B]
hist(Gamab)
plot(density(Gamab),main=,xlab=expression(Gamma[m]^"*"),ylab="Densidade")

```