

AGRUPAMENTO DE MÉDIAS VIA
BOOTSTRAP PARA POPULAÇÕES
NORMAIS E NÃO-NORMAIS

PATRÍCIA DE SIQUEIRA RAMOS

2007

PATRÍCIA DE SIQUEIRA RAMOS

**AGRUPAMENTO DE MÉDIAS VIA
BOOTSTRAP PARA POPULAÇÕES
NORMAIS E NÃO-NORMAIS**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso 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

2007

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

Ramos, Patrícia de Siqueira.

Agrupamento de médias via *bootstrap* para populações normais e não-normais/ Patrícia de Siqueira Ramos. - Lavras: UFLA, 2007.
76p. : il.

Orientador: Daniel Furtado Ferreira.
Dissertação (Mestrado) - UFLA.
Bibliografia.

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

CDD-519.5

PATRÍCIA DE SIQUEIRA RAMOS

**AGRUPAMENTO DE MÉDIAS VIA
BOOTSTRAP PARA POPULAÇÕES
NORMAIS E NÃO-NORMAIS**

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

APROVADA em 14 de fevereiro de 2007

Prof. Dr. Décio Barbin	ESALQ/USP
Prof. Dr. Júlio Silvio de Sousa Bueno Filho	UFLA
Prof. Dr. Marcelo Ângelo Cirillo	UFLA

Prof. Dr. Daniel Furtado Ferreira
(Orientador)

LAVRAS
MINAS GERAIS-BRASIL

Aos meus pais, Darlan e Eneida
e minhas irmãs Priscila e Paula

Dedico.

AGRADECIMENTOS

A Deus, por tudo.

A minha mãe, pela doçura, incondicional amor e apoio, sem os quais eu não estaria aqui e por me acompanhar em cada momento de minha vida, sempre torcendo por mim.

Ao meu pai, pelo amor e confiança, por acreditar em minha capacidade e me conceder tudo o que eu precisava para chegar onde cheguei. Levarei pra vida toda seus valiosos conselhos.

A minha irmã Paulinha, pelo carinho, atenção e coração tão bondoso, por vibrar comigo em cada vitória e estar disposta a me ouvir a qualquer hora.

A minha irmã Priscila, pelo exemplo de determinação, capacidade e preocupação, mesmo à distância, e por sempre me ajudar com suas opiniões.

Ao meu namorado, Lincoln, por me inspirar tanto encantamento e admiração e pelo amor e incentivo, tão importantes para a conclusão dessa etapa.

Amo cada um de vocês.

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

À FAPEMIG, pela concessão da bolsa de estudos.

Ao meu orientador, Daniel, pela imensa competência, solicitude e amizade com que me orientou, sempre com muito boa vontade. Guardarei eternamente seus ensinamentos e o carinho dispensados a mim.

Aos professores do DEX, pela atenção e capacidade que contribuíram muito para a minha formação.

Ao Leonardo, por ter despertado em mim o interesse pelo Mestrado em Estatística.

À amiga Elisângela, pelo carinho e apoio.

Às amigas e colegas Lourdinha, Elisa, Nádia, Roberta e tantos outros que me ajudaram a seguir e não me deixaram desanimar, pelos estudos, troca de conhecimentos e momentos que jamais esquecerei.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	iv
RESUMO	vi
ABSTRACT	viii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Comparações entre médias	3
2.1.1 Erros envolvidos e poder	7
2.2 Procedimentos de comparações múltiplas (PCM)	10
2.2.1 A amplitude estudentizada	11
2.2.2 Descrição de alguns testes	13
2.3 Análise de agrupamento	17
2.4 Comparação entre os métodos	20
2.5 Monte Carlo	27
2.6 <i>Bootstrap</i>	28
2.7 Distribuições	29
2.7.1 Normal	29
2.7.2 Lognormal	30
2.7.3 Exponencial	30
3 METODOLOGIA	32
3.1 Teste de agrupamento original	32
3.2 Teste proposto	33
3.3 Simulações	34
3.4 Exemplo real	37
4 RESULTADOS E DISCUSSÃO	38
4.1 Erro tipo I sob H_0 completa	38

4.1.1	Distribuição normal	38
4.1.2	Distribuições não-normais	39
4.2	Erro tipo I sob H_0 parcial	44
4.3	Poder sob H_1	48
4.3.1	Distribuição normal	48
4.3.2	Distribuições não-normais	51
4.4	Poder sob H_0 parcial e normalidade	56
4.5	Exemplo real	57
4.6	Considerações finais	59
5	CONCLUSÕES	61
6	REFERÊNCIAS BIBLIOGRÁFICAS	62
	ANEXOS	67

LISTA DE TABELAS

1	Relação entre os valores K do teste t -bayesiano e o nível de significância	17
2	Poder dos testes, em 2.000 experimentos simulados, com 4 repetições para cada número de níveis do fator, ao nível nominal de significância $\alpha = 0,05$, considerando-se a distribuição normal dos resíduos.	21
3	Poder do teste de Scott-Knott, em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, considerando-se a distribuição normal dos resíduos.	22
4	Poder do teste de Scott-Knott, em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,05$, considerando-se a distribuição normal dos resíduos.	23
5	Taxas de erro por comparação (TPC) e por experimento (TPE) do teste de Scott-Knott, em função do número de repetições r , número de níveis k do fator, coeficientes de variação (CV) e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa.	24
6	Taxas de erro tipo I do teste de Scott-Knott, em função do número de níveis k do fator, número de repetições r para os níveis nominais de significância $\alpha = 0,01$ e $\alpha = 0,05$, considerando-se a distribuição normal dos resíduos.	26
7	Teor de nitrogênio, em mg, de plantas de trevo-vermelho inoculadas com combinações de culturas de <i>R. trifolii</i> e <i>R. meliloti</i>	37
8	Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C), sua versão <i>bootstrap</i> (CB) em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição normal(10, 1).	39

9	Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição lognormal(0,1).	40
10	Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição lognormal(0,5).	43
11	Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição exponencial(0,1).	44
12	Taxas de erro tipo I dos testes Caliński & Corsten (C) e <i>bootstrap</i> (CB), em função do número de níveis k do fator, número de repetições r , diferenças δ entre as médias, para os níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob a distribuição normal e H_0 parcial.	46
13	Poder dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, sob H_1 completa, considerando-se a distribuição normal (10,1).	50
14	Poder dos testes de Caliński & Corsten e sua versão <i>bootstrap</i> (C e CB), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, sob H_1 completa, considerando-se a distribuição lognormal (0,1) na escala logarítmica.	54

15	Poder para os testes de Caliński & Corsten e sua versão <i>bootstrap</i> (C e CB), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, sob H_1 completa, considerando-se a distribuição exponencial (0,1).	55
16	Valores das amplitudes R_ℓ e dos valores- p obtidos em cada passo de fusão, pelos procedimentos de Caliński & Corsten(C) e pela sua versão <i>bootstrap</i> (CB), para $B = 1000$ reamostragens.	58

LISTA DE FIGURAS

1	Taxas de erro tipo I por experimento, dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), e do teste de Scott-knott (SK), em função do número de níveis k do fator, considerando-se as distribuições normal (N) (10, 1) e lognormal (L) (0, 1), sob H_0 completa, $r = 20$, para $\alpha = 0,05$ (a) e $\alpha = 0,01$ (b). As linhas pontilhadas representam os limites superior e inferior do IC exato para proporções com 99% de confiança.	42
2	Taxas de erro tipo I por experimento dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), em função do número de níveis k do fator, considerando-se a distribuição exponencial (E) (0,1), sob H_0 completa, $r = 20$, para $\alpha = 0,05$ (a) e $\alpha = 0,01$ (b). As linhas pontilhadas representam os limites superior e inferior do IC exato para proporções com 99% de confiança.	45
3	Poder dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), para diferentes números de repetições r , considerando-se a distribuição normal (10, 1), em função das diferenças entre médias δ , para diferentes números de níveis k do fator, sob H_1 e $\alpha = 0,05$	49
4	Poder dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), para diferentes números de repetições r , considerando-se a distribuição lognormal (0, 1) em função das diferenças entre médias δ , para diferentes números de níveis k do fator, sob H_1 e $\alpha = 0,05$	52
5	Poder dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), para diferentes números de repetições r , considerando-se a distribuição exponencial (0,1), em função das diferenças entre médias δ , para diferentes números de níveis k do fator, sob H_1 e $\alpha = 0,05$	53

6	Poder dos testes de Caliński & Corsten (C) e sua versão <i>bootstrap</i> (CB), para $k = 5$ e $k = 80$ níveis do fator, considerando-se a distribuição normal (N) $(10, 1)$, sob H_0 parcial, para $\alpha = 0,01$ e $0,05$	56
7	Dendrograma com os pontos de corte obtidos pelos testes C e CB para $\alpha = 0,05$	58

RESUMO

RAMOS, Patrícia de Siqueira. **Agrupamento de médias via *bootstrap* para populações normais e não-normais.** 2007. 76 p. Dissertação (Mestrado em Agronomia / Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.*

Procedimentos de Comparações Múltiplas são utilizados para comparar médias de níveis de um fator, porém, os testes mais populares apresentam problemas de ambigüidade dos resultados e de controle do erro tipo I, sendo alguns considerados conservativos e outros liberais. Métodos baseados em análise de agrupamento têm sido propostos para contornar o problema da ambigüidade. Este trabalho tem por objetivos propor uma alternativa *bootstrap* para um dos procedimentos de comparações múltiplas de Caliński & Corsten (1985), baseada em análise de agrupamento e avaliar as versões original e *bootstrap* desse procedimento por meio de simulação Monte Carlo, considerando modelos probabilísticos normais e não-normais. A metodologia de Caliński & Corsten (1985) utiliza a extensão de um teste simultâneo baseado na amplitude estudentizada. Foram consideradas $N = 1000$ simulações de k populações não-estruturadas e qualitativas. Em cada simulação foram geradas amostras de tamanho r (4, 10 e 20) de cada uma das k (5, 10, 20 e 80) populações (níveis do fator). Foram considerados os modelos probabilísticos normal, lognormal e exponencial para as populações. Sob H_0 completa, os parâmetros dos modelos probabilísticos foram considerados iguais para todos os níveis do fator. Assim, os níveis do fator apresentavam médias e variâncias comuns. Sob H_0 parcial, foram considerados dois grupos, dentro dos quais o procedimento utilizado para H_0 completa é aplicado. As densidades entre os grupos tiveram valores diferentes para os parâmetros. Sob H_1 , as densidades foram consideradas todas diferentes. No caso normal, elas se diferenciaram apenas pela média (μ), mantendo constante a variância (σ^2). Os dois métodos foram aplicados em todas as configurações simuladas e nos N experimentos gerados de cada uma delas foi avaliado o desempenho em relação às taxas de erro tipo I por experimento (sob H_0 completa e parcial) e poder (sob H_0 parcial e H_1). A situação de distribuição normal foi utilizada como ambiente favorável, pois o teste de Caliński & Corsten (1985) foi idealizado sob a pressuposição de normalidade. Os programas para simulação foram implementados em R . Assim, pôde-se concluir que os

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

dois testes são exatos sob H_0 completa e normalidade; a versão *bootstrap* sob não-normalidade e H_0 completa controla o erro tipo I por experimento e é considerado robusto; o teste original sob não-normalidade e H_0 completa é conservativo para valores pequenos de k e liberal para grandes valores de k ; sob H_0 parcial os dois testes são, em geral, liberais para menores diferenças entre os grupos e conservativos para maiores diferenças; o poder do teste *bootstrap* é um pouco mais elevado do que o do teste original sob H_0 parcial ou sob H_1 ; o desempenho do teste *bootstrap* é, em geral, considerado superior ao do teste original e robusto, sendo, portanto, recomendada a sua utilização rotineiramente.

ABSTRACT

RAMOS, Patrícia de Siqueira. **Grouping means via *bootstrap* for normal and non-normal populations.** 2007. 76 p. Dissertation (Master in Agronomy / Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, Minas Gerais, Brazil.*

Multiple comparison procedures are used to compare factor levels means. Nevertheless the most popular tests show problems concerning the ambiguity of results and the control of type I error, some of which were considered conservative and other liberal. Methods based on cluster analysis have been proposed to avoid ambiguity. The present work aims to propose a bootstrap alternative to one of Caliński & Corsten (1985) multiple comparison procedures based on cluster analysis and evaluate the original and bootstrap tests by Monte Carlo simulation considering normal and non-normal probabilistic models. The methodology by Caliński & Corsten (1985) uses the extension of a simultaneous test based on studentized range. $N = 1000$ simulations of k unstructured and qualitative populations were considered. In each simulation r sample sizes (4, 10 e 20) of each of the k populations (5, 10, 20 e 80) (factor levels) were generated. The probabilistic models exponential, lognormal and normal were considered for the populations. Under complete H_0 there were no factor effects. Thus the factor's levels presented common mean and common variance. Under partial H_0 situation it was considered two groups for which the procedure used for H_0 complete is applied. The densities between the groups had different values for the parameters. Under H_1 the densities were considered all different. In the normal case they differentiate themselves only for the mean (μ), maintaining variance (σ^2) constant. The two methods were applied in all the simulated configurations. In the N experiments generated from each one of them the performance was evaluated in relation to the type I error rates per experiment (under complete and partial H_0 situations) and in relation to the power (under partial H_0 and H_1 situations). The normal distribution situation was used as suitable environment due to the fact that the Caliński & Corsten test was idealized under the assumption of normality. The simulation software was implemented in R. Thus, it could be concluded that the two tests are exact under complete H_0 and normality; under non-normality and complete H_0 the bootstrap test controls the type I error per experiment and it is robust; under non-normality and complete H_0 the original test is conservative for

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

small values of k and it is liberal for great values of k ; in general, under partial H_0 situation, the two tests are liberal for smaller differences between groups and they are conservative for greater differences; the power of bootstrap test is considered higher than the original test under partial H_0 and H_1 ; in general, the bootstrap test performance is considered robust and superior to the original test being, therefore, recommended.

1 INTRODUÇÃO

Na experimentação, o pesquisador, geralmente, depara-se com o problema de comparar médias de diferentes níveis do fator. Essas comparações são normalmente realizadas utilizando-se os procedimentos de comparações múltiplas. Os testes corriqueiramente aplicados são os testes Tukey, Duncan, Student-Newman-Keuls (SNK), Scheffé e t de Student, entre outros.

O grande problema desses testes é a ambigüidade dos resultados, ou falta de transitividade da não-significância, ou seja, quando dois níveis do fator tidos como diferentes entre si não diferem de um terceiro (Machado *et al*, 2005). Isso dificulta a interpretação dos resultados. Um segundo problema não menos importante é o do controle do erro tipo I. Alguns testes, como t de Student e Duncan, possuem elevadas taxas de erro tipo I por experimento, sendo muito liberais, enquanto outros, como Scheffé e Tukey, possuem taxas de erro tipo I por experimento, em geral, menores do que o nível nominal de significância e são considerados bastante conservativos. O teste SNK, apesar de controlar adequadamente o erro tipo I por experimento sob a hipótese nula (H_0) completa, é bastante liberal sob H_0 parcial. Muitos dos Procedimentos de Comparações Múltiplas (PCM) se relacionam à distribuição da amplitude estudentizada (teste t , Tukey, SNK) e necessitam da obtenção dos quantis superiores $q_\alpha(k, \nu)$, em que k é o número de médias incluídas na amplitude e ν são os graus de liberdade do resíduo da análise de variância. Esta tarefa demanda grandes esforços computacionais.

O primeiro problema, o da ambigüidade, é contornado com métodos alternativos de agrupamento. A análise de agrupamento de médias de níveis do fator tem sido sugerida para criar grupos de níveis do fator distintos e não sobrepostos, simplificando a interpretação dos resultados. Esses procedimentos separam as médias dos níveis do fator em grupos homogêneos, pela minimização da variação dentro e maximização da variação entre grupos. Um exemplo típico desse procedimento é o de Scott-Knott. Dois outros procedimentos de comparações múltiplas baseados em técnicas de agrupamento foram apresentados em Caliński & Corsten (1985). O primeiro é baseado

na distribuição da amplitude estudentizada e o segundo na distribuição F . O principal problema destes testes é que são válidos apenas sob normalidade. Ademais, nenhuma validação teórica ou computacional foi realizada no trabalho original.

Muitos estudos têm sido focados nas comparações múltiplas das médias dos níveis do fator em situações de heterogeneidade de variâncias sob modelos probabilísticos normais ou não-normais. Métodos de reamostragem *bootstrap* têm sido utilizados. A idéia básica de *bootstrap*, na ausência de qualquer conhecimento sobre a população, é realizar reamostragem com reposição de tamanho n da amostra original. A distribuição *bootstrap* de algum estimador de interesse é utilizada no lugar da “distribuição teórica” deste mesmo estimador, em função da dificuldade de desenvolvê-la ou do desconhecimento da distribuição da população de onde foi obtida a amostra aleatória.

Assim, este trabalho tem por objetivos propor uma alternativa *bootstrap* para um dos procedimentos de comparações múltiplas de Caliński & Corsten e avaliar as versões original e *bootstrap* deste procedimento por meio de simulação Monte Carlo, considerando modelos probabilísticos normais e não-normais.

2 REFERENCIAL TEÓRICO

2.1 Comparações entre médias

Na experimentação, o pesquisador geralmente se depara com o problema de comparar médias de diferentes níveis do fator em estudo. Assim, admite-se a hipótese inicial, chamada de hipótese de nulidade (H_0) como verdadeira, que supõe que não existem diferenças entre os níveis de um fator qualquer, de efeitos fixos (Machado et al., 2005):

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k \quad (1)$$

em que k é o número de níveis do fator.

O teste F é utilizado para testar se existem diferenças reais entre níveis do fator. Quando a hipótese nula não é rejeitada, não existem mais questionamentos a serem feitos. Entretanto, quando há rejeição da hipótese H_0 , ou seja, a hipótese de nulidade é rejeitada, a hipótese alternativa (H_1 ou H_A) de que há pelo menos uma diferença entre os níveis do fator é considerada verdadeira. O passo seguinte é a investigação sobre onde se encontram essas diferenças entre os níveis do fator (Banzatto & Kronka, 1989). Em situações simuladas, é comum gerar dados experimentais sob a situação de nulidade parcial ou H_0 parcial, em que há diferenças entre alguns níveis do fator em estudo e outros são considerados iguais. Neste caso, ao serem aplicados os testes, a hipótese nula da equação (1) é que é testada. Isso é feito para retratar situações reais que os pesquisadores julgam comuns (Borges & Ferreira, 2003; Silva et al., 1999).

Estudos comparativos são comumente empregados na pesquisa científica. Uma ferramenta popular para analisar dados desses estudos comparativos é um teste de homogeneidade dos grupos (níveis do fator) sob investigação. Apenas a utilização do teste não provê inferências sobre as várias comparações entre os pares dos grupos. Algumas das comparações podem ser pré-especificadas (antes de se ter os dados) e outras podem ser selecionadas após a coleta dos dados (Hochberg & Tamhane, 1987).

Há vários tipos de comparações entre os níveis do fator que podem ser agrupados em categorias baseadas na metodologia estatística utilizada para analisá-las. Hinkelmann & Kempthorne (1987) realizaram a categorização destes tipos de comparações:

- 1) comparações pré-planejadas para níveis do fator qualitativo:
 - contrastes de níveis do fator (tratamentos);
 - contrastes ortogonais;
 - partição da soma de quadrados de níveis do fator (SQ Tratamentos);
- 2) comparações para níveis do fator quantitativo:
 - comparações de níveis do fator com níveis equidistantes;
 - polinômios ortogonais;
 - soma de quadrados dos contrastes e análise de variância;
- 3) níveis do fator qualitativo e não-estruturado:
 - procedimentos de comparações múltiplas (PCM).

Alguns desses procedimentos são apropriados para alguns tipos de níveis do fator e inapropriados para outros. Em alguns tipos de experimentos, os níveis do fator em estudo apresentam relações entre si que sugerem comparações de maior interesse. Essas comparações são determinadas pela intenção do pesquisador, sendo denominadas pré-planejadas ou *a priori* (Hinkelmann & Kempthorne, 1987).

Considerando uma função linear $Y = f(x) = c_1x_1 + c_2x_2 + \dots + c_kx_k$ e se ainda for verificado que $\sum_{i=1}^k c_i = c_1 + c_2 + \dots + c_k = 0$, então Y é um contraste nas variáveis x . De forma geral, não se conhecem as médias verdadeiras dos níveis do fator e, por isso, não se conhece o valor do contraste dessas médias. Conhecendo-se as estimativas de médias, calculam-se as estimativas dos contrastes (Banzatto & Kronka, 1989). O valor paramétrico do contraste é definido por:

$$Y = c_1\mu_1 + c_2\mu_2 + \dots + c_k\mu_k, \quad \sum_{i=1}^k c_i = 0. \quad (2)$$

Cujo estimador é:

$$\widehat{Y} = c_1\bar{X}_1 + c_2\bar{X}_2 + \dots + c_k\bar{X}_k \quad (3)$$

em que $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ são as médias amostrais dos k níveis do fator.

Sejam dois estimadores de contrastes:

$$\widehat{Y}_1 = a_1\bar{X}_1 + a_2\bar{X}_2 + \dots + a_k\bar{X}_k \quad (4)$$

$$\widehat{Y}_2 = b_1\bar{X}_1 + b_2\bar{X}_2 + \dots + b_k\bar{X}_k \quad (5)$$

em que $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ são as médias amostrais dos k níveis do fator. Os dois contrastes \widehat{Y}_1 e \widehat{Y}_2 são considerados ortogonais se há independência entre suas comparações, ou seja, a variação de um contraste é inteiramente independente da variação do outro. A condição para que dois contrastes sejam ortogonais entre si é que a covariância entre eles seja nula (Banzatto & Kronka, 1989). Se, além das variâncias, os números de repetições também forem iguais, a condição de ortogonalidade é:

$$\sum_{i=1}^k a_i b_i = 0. \quad (6)$$

Para k níveis do fator podem ser formulados vários grupos de $k - 1$ contrastes ortogonais, porém apenas alguns desses grupos são úteis para interpretar os resultados de um experimento. A soma de quadrados de tratamentos (níveis do fator) pode ser particionada em $k - 1$ somas de quadrados de contrastes individuais com 1 grau de liberdade cada (Hinkelmann & Kempthorne, 1987).

Se o fator em estudo é do tipo quantitativo, comparações entre os níveis também são consideradas pré-planejadas, como, por exemplo, doses crescentes de um produto. Em vez de comparar os níveis do fator uns com os outros, é mais interessante verificar se existe alguma tendência na resposta, linear ou curva, em relação à aplicação das doses. Para descobrir qual modelo

melhor se ajusta à relação entre níveis do fator e resposta, utiliza-se a análise de regressão, pelo método dos mínimos quadrados. Se o objetivo é descobrir se há curvatura e qual o seu tipo, pode-se utilizar o método dos polinômios ortogonais (Hinkelmann & Kempthorne, 1987).

Os polinômios ortogonais representam os coeficientes de um conjunto de contrastes ortogonais, podendo ser utilizados para particionar a soma de quadrados de tratamentos (níveis do fator) em $k - 1$ somas de quadrados associadas aos coeficientes de regressão. Dessa forma, pode ser testado, pelo teste F , se um dado modelo (linear, quadrático ou de ordem superior) se ajusta aos dados (Hinkelmann & Kempthorne, 1987).

Se após a realização do ensaio o pesquisador desejar comparar níveis do fator cujas médias aparentam ser diferentes, ou quiser comparar quaisquer outras de seu interesse, surgem as comparações não planejadas ou *a posteriori*. Essas ainda podem ser denominadas comparações sugeridas pelos dados, *data snooping* (Montgomery, 1991) ou *data dredging* (Steel & Torrie, 1980).

Se o fator for qualitativo e estruturado de forma que seus níveis sugiram comparações, aconselha-se a utilização de contrastes e posteriormente um teste específico (Ramalho *et al*, 2000). Entretanto, se o fator for qualitativo e não-estruturado, procedimentos de comparações múltiplas (PCM) devem ser aplicados. Os PCM são procedimentos estatísticos que comparam mais de duas médias e possuem uma vasta bibliografia sobre eles (Hochberg & Tamhane, 1987; Machado *et al.*, 2005; Ramalho *et al*, 2000; entre outros). Quando os PCM são utilizados, todas as comparações pareadas possíveis serão realizadas. É preciso ter alguns cuidados para se utilizar procedimentos de inferência corretos quando se realizam testes de hipóteses ou estimação intervalar. O maior problema é o efeito de multiplicidade que pode levar a muitos testes com resultados significativos se procedimentos incorretos forem escolhidos (Hinkelmann & Kempthorne, 1987). Todos os testes desse tipo permitem analisar diferenças entre médias após a conclusão do experimento, para detectar possíveis grupos entre um grupo de níveis do fator não-estruturados.

Os PCM são populares pela automatização da construção dos parâmetros e pela popularidade dos testes envolvidos (Machado et al., 2005). Estes autores ressaltam alguns problemas envolvidos na utilização dos PCM:

1. número elevado de parâmetros envolvidos: todas as comparações duas a duas são realizadas e, por isso, o número de parâmetros cresce com o aumento do número de níveis k do fator em estudo;
2. falta de transitividade da não-significância: é a chamada ambigüidade dos resultados. Considerando-se três médias, a maior pode diferir da menor, mas ambas não diferirem da média intermediária;
3. dificuldade de interpretação: devido à ambigüidade.

Apesar de constituírem uma metodologia bastante utilizada, os PCM são, muitas vezes, mal empregados. 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 houver significância da interação, então, toda a informação se encontra nos efeitos principais e as médias de cada nível podem ser comparadas recorrendo 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 em estudo é do tipo quantitativo, deve-se utilizar a regressão e não os PCM. 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.

2.1.1 Erros envolvidos e poder

Ao testar a hipótese de nulidade, o pesquisador corre um risco global de tomar decisões erradas que deverá ser considerado. O pesquisador está

sujeito a incorrer em erros de decisão (Banzatto & Kronka, 1989; Gomes, 1984; Machado et al., 2005; Mood et al., 1974):

- Erro Tipo I: erro cometido ao se rejeitar a hipótese nula verdadeira, quando deveria ser aceita. É diretamente controlável pelo pesquisador. A probabilidade de se cometer esse erro é dada por: $P[\text{Erro Tipo I}] = P[\text{rejeitar } H_0 | H_0 \text{ verdadeira}] = \alpha$;
- Erro Tipo II: erro cometido ao não se rejeitar a hipótese nula falsa. Não é controlável diretamente pelo pesquisador. A probabilidade de se cometer esse erro é dada por: $P[\text{Erro Tipo II}] = P[\text{não rejeitar } H_0 | H_0 \text{ falsa}] = \beta$;
- Erro Tipo III: erro cometido ao se declarar uma média maior do que a outra quando, na verdade, ocorre o contrário (Carmer & Swanson, 1973). A probabilidade de se incorrer nesse tipo de erro é: $P[\text{Erro Tipo III}] = P[\text{rejeitar } H_0 | H_0 \text{ falsa, mas a favor de } H_1 \text{ errada}]$

O poder do teste é a probabilidade de se rejeitar H_0 , sendo H_0 falsa e é igual a $1 - \beta$. Assim, poder é a capacidade do teste de detectar todas as reais diferenças entre os efeitos dos níveis do fator.

Esses erros devem ser levados em conta ao se escolher o teste de comparações a aplicar. As probabilidades de se cometerem os erros tipo I e II são inversamente proporcionais, sendo impossível controlá-las ao mesmo tempo em um único experimento. Por isso, cabe ao pesquisador avaliar os objetivos propostos para fazer a sua opção. Segundo Carmer & Swanson (1973), as taxas de erro tipo III são baixas e, normalmente, desprezíveis e são medidas considerando-se todas as comparações (*comparisonwise*) ou por experimento (*experimentwise*).

Segundo Sampaio (1998), a teoria estatística aconselha que, quando se têm k níveis do fator, os $k-1$ graus de liberdade poderiam ser decompostos e testados, quando, na verdade, há $N = k(k-1)/2$ comparações possíveis. O interesse maior está nestas N comparações entre as médias, já que os $k-1$ contrastes compreendem grupos de níveis do fator e nem sempre a

comparação entre dois deles. Com o aumento do número k de níveis do fator e o conseqüente aumento da amplitude da maior para a menor média, percebe-se que a probabilidade de se encontrarem diferenças significativas entre os N contrastes também aumenta.

A preocupação maior 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, sendo o teste, 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 poderoso ou liberal (Machado et al., 2005).

Existem três formas muito utilizadas para medir as taxas de erro tipo I nos procedimentos de comparações múltiplas (Steel & Torrie, 1980). A primeira forma mede a taxa de erro tipo I por comparação, TPC (*comparisonwise* ou *per-comparison error rate*), que é a probabilidade de se rejeitar uma hipótese verdadeira em todas as possíveis combinações de médias de níveis do fator, tomadas duas a duas:

$$\text{TPC} = \frac{\text{Número de inferências erradas}}{\text{Número total de inferências}}.$$

Na segunda forma, calcula-se a taxa de erro tipo I por experimento, TPE (*experimentwise error rate*), que é a probabilidade de se realizar pelo menos uma inferência errada por experimento:

$$\text{TPE} = \frac{\text{Número de experimentos com, pelo menos, 1 inferência errada}}{\text{Número total de experimentos}}.$$

Uma TPE real deve permitir o teste de qualquer hipótese. É freqüente testar apenas um subconjunto, ou família, da hipótese nula. Todas as comparações pareadas possíveis são uma família, assim como uma comparação

de todos os níveis do fator com a testemunha também o é. Se houver a restrição da família de hipóteses conter todas as hipóteses possíveis, um valor crítico menor seria mais adequado. Assim, a taxa de erro por família (TPF) é definida por:

$$TPF = \frac{\text{Número de famílias com, pelo menos, 1 inferência errada}}{\text{Número total de famílias testadas}}.$$

A taxa de erro por experimento (TPE) é um caso da TPF. Hinkelmann & Kempthorne (1987) mostram que a relação entre TPC e TPF é:

$$1 - TPF = (1 - TPC)^N$$

ou

$$TPF = 1 - (1 - TPC)^N.$$

Vários trabalhos comparam os diferentes PCM 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 da análise de variância, ele mantiver, pelo menos aproximadamente, o desempenho delineado originalmente na elaboração de sua teoria (Borges & Ferreira, 2003).

2.2 Procedimentos de comparações múltiplas (PCM)

Os PCM são utilizados para testar contrastes envolvendo duas médias e todas as combinações entre elas. Para se aplicar os PCM, deve ser realizada uma padronização das estatísticas utilizadas por meio do estimador do erro padrão da diferença das médias entre dois níveis do fator (O'Neill

& Wetherill, 1971). Esse estimador envolve o quadrado médio do erro obtido pela análise de variância aplicada aos dados e denomina-se Diferença Mínima Significativa (DMS). A DMS varia para cada teste na sua teoria. Sua forma geral é:

$$DMS = \gamma S_d \quad (7)$$

em que: $S_d = \sqrt{2QME/r}$ é o estimador do erro padrão da diferença de duas médias em um delineamento balanceado; QME é o quadrado médio do resíduo da análise de variância associado a ν graus de liberdade; r é o número de repetições e γ depende do método, dos graus de liberdade do erro e do número de comparações simultâneas.

Em qualquer PCM, a diferença observada entre duas médias (consecutivas ou não) é comparada com um valor crítico apropriado. Se a estimativa do contraste $\hat{Y} = |\bar{X}_i - \bar{X}_j|$ excede o valor crítico, a diferença é dita significativa e as médias são consideradas diferentes. Se o contrário ocorrer, a diferença é considerada não-significativa. Os valores críticos variam de um procedimento para outro, sendo que, para um mesmo grupo de dados, os diferentes PCM aplicados podem apresentar resultados diferentes (Carmer & Swanson, 1973).

2.2.1 A amplitude estudentizada

Para experimentos que envolvam k níveis do fator que sejam igualmente repetidos (r) e com ν graus de liberdade do erro, resultando em médias não correlacionadas com variâncias homogêneas, a maioria dos PCM é função da distribuição da amplitude estudentizada, que está relacionada à distribuição t de Student quando há duas médias na determinação da amplitude (Machado et al., 2005). Seja Q a variável aleatória (Pachares, 1959) definida por:

$$Q = \frac{X_{(k)} - X_{(1)}}{S_d} \quad (8)$$

em que $X_{(k)}$ e $X_{(1)}$ são a maior e a menor observação em uma amostra aleatória de tamanho k , S é o estimador independente do desvio padrão, com ν graus de liberdade.

A função de distribuição da variável Q , amplitude estudentizada, é definida como:

$$F(q) = C \int_0^\infty x^{\nu-1} e^{-\frac{\nu x^2}{2}} \times \left\{ k \int_{-\infty}^\infty \phi(u) [\Phi(u + qx) - \Phi(u)]^{k-1} du \right\} dx \quad (9)$$

em que:

$$C = \frac{\nu^{\nu/2}}{2^{\nu/2-1} \Gamma(\frac{\nu}{2})}$$

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

$$\Phi(u) = \int_0^u \phi(t) dt.$$

A obtenção dos quantis superiores $q_\alpha(k, \nu)$, sendo k o número de médias incluídas na amplitude, é uma dificuldade computacional para a realização dos testes. Existem algoritmos como o de Lund & Lund (1983), que são limitados por permitirem a obtenção de quantis para valores de α entre 0,01 e 0,10. Outro algoritmo, de Copenhaver & Holland (1988), não possui as mesmas limitações do anterior, porém, os códigos em Fortran não foram publicados e Gleason (1999) propõe um método não iterativo para aproximar quantis da amplitude estudentizada. Dado um quantil da

distribuição $t_{\alpha/2}(\nu)$, o método utiliza operações numéricas e uma tabela de constantes, além de possuir implementação simples. O autor afirma que os quantis aproximados apresentam acurácia muito próxima da tabela utilizada desde 1960 (havendo diferenças na quarta casa decimal).

2.2.2 Descrição de alguns testes

Há muitas dúvidas que envolvem a escolha de qual procedimento de comparações múltiplas utilizar. O pesquisador, mesmo da área estatística, fica indeciso ao ter que tomar uma decisão. Isso ocorre porque cada teste possui uma teoria e princípios próprios que poderão ser adequados ou não, dependendo da situação.

Fisher, em 1924, apresentou a fundamentação do teste F , que foi elaborado para avaliar a variação média de uma determinada fonte em relação à variação individual:

$$F = \frac{\text{Variância da fonte testada}}{\text{Variância do resíduo}}.$$

O valor obtido é, então, comparado ao valor da tabela apropriada, em relação aos graus de liberdade da fonte testada e do resíduo. Se a fonte testada não for uma interação e se referir a um grau de liberdade, o teste F é adequadamente aplicado e equivale ao teste t de Student ($t^2 = F$) (Sampaio, 1998).

O primeiro teste baseado na amplitude estudentizada a ser tratado é o t de Student ou teste da diferença mínima significativa (LSD - *Least Square Difference*), cujo valor crítico é:

$$LSD = \frac{q_{\alpha}(2, \nu)S_d}{\sqrt{2}} = t_{\alpha/2}(\nu)S_d.$$

em que $t_{\alpha/2}(\nu)$ é o quantil superior $100(\alpha/2)\%$ da distribuição t de Student.

O número de pares possíveis (contrastos pareados) cresce rapida-

mente com o aumento do número de níveis do fator, fazendo com que a probabilidade de que, pelo menos, um par apresente diferença significativa, aumente (Gomez & Gomez, 1984). O teste t de Student controla o erro tipo I no nível nominal α em um contraste pareado testado individualmente, porém, não controla esse tipo de erro para todos os testes pareados possíveis, nesse mesmo nível α . É um método recomendado para realizar comparações planejadas *a priori* porque controla apenas a taxa de erro por comparação em um nível nominal máximo igual a α (Machado et al., 2005).

O teste LSD pode ser precedido por um teste F preliminar, razão entre os quadrados médios de níveis do fator e do erro, para testar a hipótese nula global. Esse teste é chamado LSD protegido de Fisher, por ser atribuído a R. A. Fisher (Hinkelmann & Kempthorne, 1987). Se o valor observado de F for significativo, em um nível de significância α , o teste LSD é aplicado. Porém, se o valor de F for não significativo, nenhuma comparação entre médias é realizada, eliminando a possibilidade de se cometer o erro tipo I. Sua estatística é calculada por:

$$FSD = LSD = \frac{q_{\alpha}(2, \nu)S_d}{\sqrt{2}} = t_{\alpha/2}(\nu)S_d.$$

Esta proteção não garante que os resultados sejam satisfatórios, pois a suposição da veracidade da hipótese nula (1) é improvável, na maioria das situações. O caso mais comum é de uma nulidade parcial, em que grupos de níveis do fator são homogêneos internamente e heterogêneos entre si. Essa situação leva à significância do teste F , sendo as comparações dos níveis do fator dentro dos grupos homogêneos pelo teste t sujeitas a altas taxas de erro tipo I por experimento. O teste controla o erro tipo I por comparação em todos os casos e o erro tipo I por experimento sob H_0 completa, mas não sob H_0 parcial (Machado *et al.*, 2005).

A diferença significativa de Fisher, ou correção de Bonferroni, também é uma alternativa ao teste t . É realizado um ajuste do nível de significância α para o número de comparações realizadas $p = \binom{k}{2}$ e divide-se a significância nominal pelo número p de comparações pareadas para se de-

terminar o quantil da distribuição da amplitude estudentizada. Seu valor crítico é calculado por meio de:

$$BLSD = \frac{q_{\alpha/p}(2, \nu)S_d}{\sqrt{2}} = t_{\alpha/(2p)}(\nu)S_d.$$

O *BLSD* protege contra o erro tipo I por comparação e por experimento com limites máximos iguais ao valor nominal (Machado et al., 2005).

Outro teste dependente da amplitude estudentizada é o de Tukey. Esse teste foi elaborado para controlar o erro tipo I por experimento e requer que todos os níveis do fator tenham o mesmo número de repetições e que as inferências sejam feitas para todo o conjunto de comparações duas a duas (Ramalho *et al*, 2000). Utiliza-se apenas um valor de DMS, independente do número de médias testadas, caracterizando-o como um teste muito conservativo. Controla muito bem o erro tipo I, porém, permite o aparecimento do erro tipo II. Sua DMS é calculada por:

$$TSD = \frac{q_{\alpha}(k, \nu)S_d}{\sqrt{2}}.$$

Se k for maior do que 2, o valor *TSD* 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).

O teste de Student-Newman-Keuls exige o cálculo de $k - 1$ valores críticos e não apenas de um valor, como o *TSD* e o *LSD*. Seus valores críticos são calculados por:

$$SNK_p = \frac{q_{\alpha}(p, \nu)S_d}{\sqrt{2}}.$$

em que $p = 2, 3, \dots, k$ é o número de médias abrangidas pelo contraste.

SNK_2 é igual ao valor crítico *LSD* e SNK_k é igual ao valor crítico *TSD*. Para valores intermediários de p , o valor SNK_p é um valor interme-

diário entre os valores críticos LSD e TSD (Machado et al., 2005).

Duncan (1955) apresentou amplitudes estudentizadas especiais, em que o nível de significância α varia de acordo com o número de médias abrangidas, da seguinte forma: $\alpha_p = [1 - (1 - \alpha)^{p-1}]$ para $p = 2, 3, \dots, k$. Assemelha-se ao teste SNK por utilizar múltiplos valores críticos, porém, é menos conservativo do que este. O teste de Duncan necessita que sejam calculados $k - 1$ valores críticos, por:

$$MRT_p = \frac{q_\alpha(p, \nu) S_d}{\sqrt{2}}.$$

Os valores críticos MRT_p são maiores do que LSD para $p = 2$, mas inferiores ao valor crítico TSD e ao SNK_p . Assim, o teste de Duncan é considerado mais conservativo do que o teste LSD , porém, mais liberal do que os testes de Tukey e SNK .

O método de Scheffé (1953) é baseado na distribuição F e é muito geral, sendo aplicável a qualquer contraste linear das médias dos níveis do fator. Pode também ser usado como um procedimento de comparações múltiplas. Seu valor crítico é dado por:

$$SSD = [mF_\alpha(p, \nu)]^{1/2} S_d.$$

em que: $S_d = \sqrt{\sum \frac{c_i^2 QME}{r}}$; c_i é o coeficiente da média do nível i do fator; $F_\alpha(p, \nu)$ é o quantil superior $100\alpha\%$ da distribuição F com $m = k - 1$ e ν graus de liberdade.

Exceto para o caso de dois níveis do fator do fator, o valor de SSD é maior e, portanto, este método é mais conservativo do que todos os outros anteriormente apresentados e seu poder é baixo (Carmer & Swanson, 1973). É um teste mais apropriado para contrastes sugeridos pelos dados (Steel & Torrie, 1980).

Uma solução bayesiana para procedimentos de comparações múltiplas foi apresentada por Waller & Duncan (1972). Neste teste, não há nível

de significância envolvido, mas considera o peso do erro ou a grandeza dos erros tipo I e II, expressa pela constante K (Steel & Torrie, 1980). Na Tabela 1 são apresentadas as relações de equivalência entre os valores de K e o nível de significância α dos testes mais utilizados, em que $K = \frac{K_1}{K_2}$ e as constantes K_1 e K_2 medem a grandeza dos erros tipo I e tipo II, respectivamente.

TABELA 1: Relação entre os valores K do teste t -bayesiano e o nível de significância

α	10%	5%	1%
K	50	100	500

Fonte: Perecin & Malheiros (1989) citados por Borges & Ferreira (2003)

O método TEB , ou teste exato de Bayes, tem como valor crítico:

$$TEB = t(K, F, \nu, q) \sqrt{\widehat{V}(D)} = t(K, F, \nu, q) \sqrt{\frac{2QME}{r}},$$

em que t é um valor tabelado igual ao risco médio mínimo para os valores de $K = K_1/K_2$, assim, t minimiza o risco global porque as constantes K_1 e K_2 medem a grandeza dos erros tipo I e II, F é o valor calculado na análise de variância, ν e q são os graus de liberdade do resíduo e dos níveis do fator, respectivamente.

Esse teste é denominado método da diferença mínima significativa de Bayes (BET), por Carmer & Swanson (1971), sendo semelhante ao teste LSD quando F é grande, porém, quando esse é um valor pequeno, menor do que 2, 5, o TEB se aproxima do teste de Tukey, sendo assim considerado conservativo. Espera-se que o teste TEB evite o erro tipo I quando F é pequeno e evite o erro tipo II quando F é grande.

2.3 Análise de agrupamento

O objetivo dos PCM é dividir os níveis do fator em grupos “homogêneos” para facilitar a interpretação dos resultados e auxiliar na definição das recomendações a serem feitas. Porém, alguns dos PCM resultam em

grupos sobrepostos. Para contornar esse problema, utilizam-se idéias de análise de agrupamento combinadas aos testes de hipóteses (Hinkelmann & Kempthorne, 1987). Nos métodos baseados em análise de agrupamento particionam-se grupos de médias e não se trabalha com comparações duas a duas, como é comum nos PCM. Assim, esses procedimentos eliminam o problema da ambigüidade, separando-se as médias dos níveis do fator em grupos homogêneos, pela minimização da variação dentro e maximização da variação entre grupos.

A sugestão de se particionar médias utilizando análise de agrupamento foi feita por Plackett em sua discussão de O'Neill & Wetherill (1971). Métodos deste tipo foram sugeridos por vários autores (Caliński & Corsten, 1985; Scott & Knott, 1974). O método de Caliński & Corsten, baseado em uma extensão da amplitude estudentizada, é citado também por Hinkelmann & Kempthorne (1987) e por Hochberg & Tamhane (1987). Estes últimos autores apresentam uma série de procedimentos de comparações múltiplas, baseados em diversas técnicas, além da análise de agrupamento.

O procedimento divisivo 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 de valor 3,141593; B_0 é o valor máximo da soma de quadrados máxima entre 2 grupos de médias, sobre todas as $(p - 1)$ partições e $\hat{\sigma}_0^2$ é a estimativa de máxima verossimilhança de $\hat{\sigma}^2$.

É um teste que utiliza a razão de verossimilhança para testar a significância de que os k níveis do fator podem ser divididos em dois grupos que maximizem a soma de quadrados entre grupos. Scott & Knott (1974) mostram que a distribuição da estatística λ pode ser aproximada por uma χ^2 com $\nu_0 = \frac{k}{(\pi-2)}$ 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 separadamente. Se o teste não for significativo, conclui-se que as médias dos níveis do fator 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.

Caliński & Corsten (1985) propuseram dois procedimentos: o primeiro é uma extensão da amplitude estudentizada e o segundo é baseado na distribuição F . Sua proposta seria agrupar os níveis do fator em um pequeno número de grupos distintos, mas internamente homogêneos. O primeiro método é hierárquico, aglomerativo e com regra de parada baseada na amplitude estudentizada estendida. É denominado hierárquico porque, quando uma média é incluída em um grupo homogêneo, ela não será retirada dele e aglomerativo porque, a cada passo, dois grupos são unidos para formar um novo grupo. O algoritmo inicia-se com os k níveis ordenados em relação às suas médias, formando k grupos. No primeiro passo, as duas médias mais próximas, em relação à sua amplitude $R_0 = |\bar{X}_i - \bar{X}_j|$, são unidas e a sua amplitude é comparada com o valor crítico $R_\alpha = q_\alpha(k; \nu) \sqrt{\frac{QME}{r}}$ em que $q_\alpha(k; \nu)$ é o quantil superior $100\alpha\%$ da distribuição estudentizada para k médias e ν graus de liberdade. A homogeneidade dentro de cada grupo é testada comparando-se a amplitude com R_α novamente.

O segundo procedimento de Caliński & Corsten (1985) é não-hierárquico e com regra de parada baseada na razão F estendida. O processo se inicia com os k níveis do fator formando um grupo único. Em cada passo n ($n = 1, 2, \dots$), a melhor partição em r grupos (P_1, P_2, \dots, P_n) é encontrada pela minimização da soma de quadrados dentro dos grupos:

$$r \sum_{j=1}^n \sum_{i \in P_j} (\bar{X}_i - \bar{\bar{X}}_{P_j})^2$$

em que r é o número de repetições de cada nível do fator; \bar{X}_i é a média amostral do nível i , pertencente ao j -ésimo subgrupo P e $\bar{\bar{X}}_{P_j}$ é a média das médias amostrais do j -ésimo subgrupo P . Se esse valor não ultrapassa $(k-1)F_\alpha(k-1, \nu)QME$, o valor crítico comum, o agrupamento é aceito.

2.4 Comparação entre os métodos

Carmer & Swanson (1973) avaliaram as taxas de erro tipos I e III e o poder de dez testes de comparações múltiplas, por meio de simulação Monte Carlo. Os autores concluíram que o procedimento de Scheffé não deve ser utilizado como um teste de comparações múltiplas por ser muito conservativo. Para diferenças entre as médias de $0,5\sigma_{\bar{X}}$ a $2,5\sigma_{\bar{X}}$, os testes de Tukey e *SNK* possuem a menor capacidade de detectar diferenças reais. Os testes que apresentam menor proteção ao erro tipo I, sendo considerados liberais, são o *LSD* e *FSD* para $\alpha = 0,10$. De acordo com os autores, a decisão mais razoável seria optar entre o *t*-Bayesiano ou o *FSD* para $\alpha = 0,05$, levando-se em consideração o poder e o controle do erro tipo I dos dois procedimentos.

Perecin & Barbosa (1988) compararam alguns PCM quanto ao seu poder, com número de níveis do fator variando de 5 a 100 e valores pré-estabelecidos de diferenças entre estes níveis, representados pelos erros padrões da média. Quando $2\sigma_{\bar{X}}$ de diferença foram considerados, os testes de Duncan, *SNK* e *LSD* apresentaram poder semelhante e por volta de 22%. O maior poder para esse caso foi obtido com o teste *t*-Bayesiano e o menor poder com o teste de Tukey, por volta de 1%. O último foi muito afetado pelo número de níveis do fator considerado; quanto maior o número de níveis, menor o poder. Isso pode ser verificado pela análise da Tabela 2, em que se encontram os valores para poder, obtidos para quatro repetições.

TABELA 2: Poder dos testes, em 2.000 experimentos simulados, com 4 repetições para cada número de níveis do fator, ao nível nominal de significância $\alpha = 0,05$, considerando-se a distribuição normal dos resíduos.

Teste	Níveis do fator	Diferença real entre as médias (δ)			
		$2\sigma_{\bar{x}}$	$4\sigma_{\bar{x}}$	$6\sigma_{\bar{x}}$	$8\sigma_{\bar{x}}$
Newman-Keuls	5	0,219	0,604	0,916	0,994
	10	0,222	0,652	0,946	0,998
	20	0,236	0,674	0,960	1,000
	40	0,233	0,678	0,963	1,000
	100	0,238	0,692	0,966	1,000
Newman-Keuls modificado	5	0,241	0,608	0,916	0,994
	10	0,242	0,656	0,946	0,998
	20	0,257	0,677	0,960	1,000
	40	0,254	0,681	0,963	1,000
	100	0,259	0,696	0,966	1,000
Duncan	5	0,239	0,736	0,976	1,000
	10	0,240	0,775	0,985	1,000
	20	0,265	0,791	0,989	1,000
	40	0,254	0,790	0,991	1,000
	100	0,259	0,799	0,991	1,000
t	5	0,239	0,769	0,988	1,000
	10	0,242	0,814	0,991	1,000
	20	0,257	0,824	0,994	1,000
	40	0,254	0,821	0,995	1,000
	100	0,259	0,829	0,995	1,000
Tukey	5	0,66	0,426	0,877	0,994
	10	0,25	0,288	0,814	0,992
	20	0,09	0,180	0,718	0,981
	40	0,02	0,106	0,592	0,961
	100	0,00	0,47	0,452	0,925
t -bayesiano	5	0,266	0,781	0,981	1,000
	10	0,325	0,871	0,996	1,000
	20	0,346	0,883	0,998	1,000
	40	0,342	0,883	0,998	1,000
	100	0,345	0,889	0,998	1,000

Fonte: Perecin & Barbosa (1988).

Nas Tabelas 3 e 4 se encontram os valores do poder do teste de Scott-Knott obtidos com as simulações realizadas por Silva et al. (1999), para os níveis de significância $\alpha = 0,01$ e $\alpha = 0,05$, respectivamente. Pela Tabela 3, pode-se observar que o poder do teste aumentou ao se elevar o número de níveis do fator de 5 para 10, especialmente quando a diferença entre médias foi $2\sigma_{\bar{x}}$. Outra influência importante para o aumento do poder foi o maior número de repetições, especialmente de 4 para 10. Porém, esta influência é menor do que a do número de níveis do fator.

TABELA 3: Poder do teste de Scott-Knott, em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, considerando-se a distribuição normal dos resíduos.

r	k	Diferença real entre as médias (δ)				
		$2\sigma_{\bar{x}}$	$4\sigma_{\bar{x}}$	$6\sigma_{\bar{x}}$	$8\sigma_{\bar{x}}$	$10\sigma_{\bar{x}}$
4	5	0,2373	0,5703	0,6853	-	-
	10	0,3632	0,7320	0,9392	0,9945	0,9999
	20	0,3946	0,7636	0,9585	0,9977	0,9999
	40	0,4082	0,7722	0,9625	0,9985	1,0000
	96	0,4127	0,7673	0,9599	0,9987	1,0000
10	5	0,3266	0,7640	0,9218	-	-
	10	0,3791	0,7658	0,9633	0,9995	1,0000
	20	0,3996	0,7732	0,9668	0,9991	1,0000
	40	0,4122	0,7787	0,9666	0,9993	1,0000
	96	0,4139	0,7713	0,9619	0,9987	1,0000
20	5	0,3425	0,7837	0,9555	-	-
	10	0,3809	0,7711	0,9664	0,9992	1,0000
	20	0,4012	0,7768	0,9672	0,9988	1,0000
	40	0,4123	0,7779	0,9655	0,9989	1,0000
	96	0,4159	0,7715	0,9619	0,9989	1,0000

Fonte: Silva et al. (1999).

De maneira geral, com o aumento da diferença entre médias consecutivas, há aumento do poder. Os testes de comparações múltiplas mais populares apresentam poder elevado quando as diferenças entre as médias dos níveis do fator diferem de 6 ou mais erros padrões (Perecin & Barbosa, 1988; Silva et al., 1999). Perecin & Barbosa (1988) obtiveram resultados semelhantes para diferenças de 6 ou mais $\sigma_{\bar{x}}$. Pela observação da Tabela 4, pode-se perceber que o teste de Scott-Knott apresenta melhor desempenho,

pelo fato do nível de significância ser menos rigoroso. O aumento do número de níveis do fator continuou a causar aumento do poder e a influência do número de repetições foi menor do que ao nível $\alpha = 0,01$. A porcentagem de decisões corretas do teste de Scott-Knott, expressa na Tabela 4, foi maior do que a porcentagem de decisões corretas do teste t -bayesiano, contida na Tabela 2, sendo este considerado o teste de maior poder por Perecin & Barbosa (1988). Os outros testes analisados pelos autores apresentaram poderes baixos e Tukey foi o teste com o poder mais baixo de todos.

TABELA 4: Poder do teste de Scott-Knott, em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,05$, considerando-se a distribuição normal dos resíduos.

r	k	Diferença real entre as médias (δ)				
		$2\sigma_{\bar{x}}$	$4\sigma_{\bar{x}}$	$6\sigma_{\bar{x}}$	$8\sigma_{\bar{x}}$	$10\sigma_{\bar{x}}$
4	5	0,3945	0,8142	0,9578	-	-
	10	0,4434	0,8236	0,9767	0,9993	0,9999
	20	0,4639	0,8346	0,9824	0,9998	0,0000
	40	0,4720	0,8361	0,9827	0,9997	1,0000
	96	0,4845	0,8429	0,9835	0,9998	1,0000
10	5	0,4054	0,8402	0,9840	-	-
	10	0,4503	0,8379	0,9829	0,9996	1,0000
	20	0,4651	0,8361	0,9844	0,9998	1,0000
	40	0,4740	0,8392	0,9840	0,9997	1,0000
	96	0,4856	0,8442	0,9838	0,9998	1,0000
20	5	0,4124	0,8467	0,9878	-	-
	10	0,4498	0,8360	0,9848	0,9998	1,0000
	20	0,4645	0,8386	0,9842	0,9998	1,0000
	40	0,4743	0,8382	0,9842	0,9998	1,0000
	96	0,4862	0,8446	0,9841	0,9997	1,0000

Fonte: Silva et al. (1999).

Outra análise realizada por Silva et al. (1999) foi em relação às taxas de erro tipo I do teste de Scott-Knott. Na Tabela 5 se encontram os resultados obtidos pelos autores para a situação de nulidade completa. Para as taxas de erro por experimento (TPE), verifica-se que há poucos valores que ultrapassaram os limites do intervalo de confiança (superior e inferior). Valores acima do limite superior só ocorreram para número de níveis do fator igual a 5. Não houve efeito do coeficiente de variação nas taxas de

erro. Os autores concluíram que o teste de Scott-Knott controlou os dois tipos de erro tipo I. Para as taxas de erro por comparação (TPC), nenhum valor excedeu o nível de confiança adotado.

TABELA 5: Taxas de erro por comparação (TPC) e por experimento (TPE) do teste de Scott-Knott, em função do número de repetições r , número de níveis k do fator, coeficientes de variação (CV) e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa.

r	k	CV	$\alpha = 0,01$		$\alpha = 0,05$	
			TPC	TPE	TPC	TPE
4	5	1	0,0013++	0,0025++	0,0347++	0,0660++
4	5	10	0,0042++	0,0070	0,0278++	0,0505
4	5	20	0,0032++	0,0060	0,0335++	0,0615
4	5	30	0,0022++	0,0045++	0,0285++	0,0540
10	5	1	0,0067	0,0125	0,0311++	0,0590
10	5	10	0,0047++	0,0085	0,0354++	0,0665**
10	5	20	0,0057	0,0115	0,0342++	0,0630
10	5	30	0,0054	0,0100	0,0331++	0,0635
20	5	1	0,0054	0,0105	0,0341++	0,0625
20	5	10	0,0109	0,0210**	0,0314++	0,0585
20	5	20	0,0073	0,0135	0,0349++	0,0665**
20	5	30	0,0069	0,0125	0,0341++	0,0640**
4	10	1	0,0030++	0,0065	0,0212++	0,0420
4	10	10	0,0024++	0,0045++	0,0205++	0,0425
4	10	20	0,0021++	0,0040++	0,0211++	0,0435
4	10	30	0,0025++	0,0055	0,0248++	0,0490
10	10	1	0,0055	0,0110	0,0261++	0,0525
10	10	10	0,0053	0,0105	0,0307++	0,0615
10	10	20	0,0041++	0,0085	0,0281++	0,0570
10	10	30	0,0041++	0,0080	0,0268++	0,0545
20	10	1	0,0066	0,0135	0,0245++	0,0505
20	10	10	0,0059	0,0120	0,0247++	0,0510
20	10	20	0,0040++	0,0080	0,0276++	0,0565
20	10	30	0,0077	0,0155	0,0242++	0,0515

...continua...

Tabela 5 - Cont.

r	k	CV	$\alpha = 0,01$		$\alpha = 0,05$	
			TPC	TPE	TPC	TPE
4	20	1	0,0017++	0,0035++	0,0216++	0,0440
4	20	10	0,0022++	0,0045++	0,0186++	0,0380++
4	20	20	0,0025++	0,0055	0,0182++	0,0375++
4	20	30	0,0033++	0,0065	0,0191++	0,0400
10	20	1	0,0031++	0,0065	0,0228++	0,0480
10	20	10	0,0041++	0,0085	0,0251++	0,0515
10	20	20	0,0041++	0,0080	0,0242++	0,0500
10	20	30	0,0044++	0,0090	0,0249++	0,0505
20	20	1	0,0042++	0,0085	0,0298++	0,0610
20	20	10	0,0055	0,0110	0,0274++	0,0555
20	20	20	0,0054	0,0115	0,0221++	0,0455
20	20	30	0,0050++	0,0105	0,0256++	0,0535
4	80	1	0,0027++	0,0055	0,0174++	0,0355++
4	80	10	0,0017++	0,0035++	0,0174++	0,0355++
4	80	20	0,0037++	0,0075	0,0169++	0,0345++
4	80	30	0,0033++	0,0065	0,0181++	0,0370++
10	80	1	0,0047++	0,0095	0,0227++	0,0460
10	80	10	0,0037++	0,0075	0,0255++	0,0515
10	80	20	0,0037++	0,0075	0,0241++	0,0490
10	80	30	0,0047++	0,0095	0,0175++	0,0355++
20	80	1	0,0050++	0,0100	0,0231++	0,0470
20	80	10	0,0055	0,0110	0,0245++	0,0500
20	80	20	0,0047++	0,0095	0,0243++	0,0495
20	80	30	0,0055	0,0110	0,0270++	0,0550

Fonte: Silva et al. (1999).

**Ultrapassou o LS do IC exato, com 99% de confiança para $\alpha = 0,01$ (0,017270) e $\alpha = 0,05$ (0,063914).

++Não atingiu o LI do IC exato, com 99% de confiança para $\alpha = 0,01$ (0,005188) e $\alpha = 0,05$ (0,038282).

Silva et al. (1999) ainda mediram as taxas de erro tipo I na situação de nulidade parcial em que a diferença entre níveis do fator adjacentes, quando existia, foi igual a dois erros padrões da média ($2\sigma_{\bar{x}}$). Assim, foi possível computar as taxas de erro tipo I e poder. Pela observação da Tabela 6, percebe-se que a maioria dos valores ultrapassou os níveis de confiança estabelecidos, mas sem se afastarem muito destes. Os resultados não foram tão satisfatórios como sob H_0 completa, com taxas de erro maiores à medida que se aumenta o número de níveis do fator.

TABELA 6: Taxas de erro tipo I do teste de Scott-Knott, em função do número de níveis k do fator, número de repetições r para os níveis nominais de significância $\alpha = 0,01$ e $\alpha = 0,05$, considerando-se a distribuição normal dos resíduos.

r	k	Nível Nominal de Significância	
		$\alpha = 0,01$	$\alpha = 0,05$
4	5	0,0065	0,0620
	10	0,0230**	0,0715**
	20	0,0355**	0,0790**
	40	0,0320**	0,0800**
	96	0,0295**	0,0820**
10	5	0,0285**	0,0665**
	10	0,0310**	0,0795**
	20	0,0275**	0,0815**
	40	0,0350**	0,0845**
	96	0,0375**	0,0830**
20	5	0,0305**	0,0725**
	10	0,0380**	0,0775**
	20	0,0350**	0,0735**
	40	0,0430**	0,0705**
	96	0,0440**	0,0830**

Fonte: Silva et al. (1999).

**Ultrapassou o LS do IC exato, com 99% de confiança para $\alpha = 0,01$ (0,017270) e $\alpha = 0,05$ (0,063914).

++Não atingiu o LI do IC exato, com 99% de confiança para $\alpha = 0,01$ (0,005188) e $\alpha = 0,05$ (0,038282).

Santos (2000) modificou o procedimento de Scott-Knott original e apresentou novas alternativas de agrupamento. Borges & Ferreira (2003) avaliaram o poder e as taxas de erro tipo I dos testes de comparações múltiplas Scott-knott, Tukey e *SNK* sob distribuições normais e não-normais dos resíduos. As taxas de erro tipo I foram medidas por comparação (TPC) e por experimento (TPE) sob H_0 completa e parcial. Os autores constataram que o teste de Scott-Knott controlou a TPC sob H_0 completa e não controlou a TPE para todas as situações das distribuições consideradas (normal, lognormal, exponencial e Weibull). Sob H_0 parcial, esse teste não controlou a TPC e a TPE, mesmo considerando a normalidade dos resíduos. É um procedimento mais poderoso do que os outros considerados e é robusto, sendo recomendado.

2.5 Monte Carlo

A aplicação dos PCM depende da pressuposição de normalidade, que nem sempre é verificada e há fortes indicativos de que a maioria desses procedimentos não é robusta à violação dessa pressuposição. Há alternativas para esta situação, como utilizar procedimentos não-paramétricos (Hochberg & Tamhane, 1987). Porém, o problema da dificuldade de interpretação de resultados da maioria dos PCM pode ser mais crítico com a utilização de PCM não-paramétricos. Para estes casos, a escolha do nível de significância é importante, como $\alpha = 0,10$ ou $0,20$, bem maiores do que $\alpha = 0,05$ convencional (Hinkelmann & Kempthorne, 1987). Outra alternativa para situações não-normais é utilizarem-se estimadores robustos para os efeitos do níveis do fator, como os estimadores M . Porém, dificuldades existem, pois as distribuições das estatísticas dos testes são difíceis ou impossíveis de se obter (Hinkelmann & Kempthorne, 1987).

Uma outra saída é utilizar, em muitos casos, simulação Monte Carlo, tornando a obtenção de resultados uma tarefa mais simples do que a obtenção analítica. Em estudos de desempenho de testes estatísticos, devido à complicação de se obter analiticamente informações sobre as taxas de erro tipo I e poder, esta é uma alternativa que está se tornando muito popular

e é utilizada em vários estudos de avaliação de PCM, como Borges & Ferreira (2003), Carmer & Swanson (1973), Perecin & Barbosa (1988), Santos (2000), Silva et al. (1999), entre outros.

A simulação é um processo que tenta reproduzir o comportamento de um sistema real, em geral por meio de programas de computadores, para estudar seu funcionamento sob condições alternativas. O método de Monte Carlo é uma técnica que consiste na simulação de dados por meio da 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).

Como as bases teóricas dos procedimentos de comparações múltiplas não são as mesmas, há muitas dificuldades teóricas em compará-los analiticamente, levando à necessidade de realizar estas comparações com o auxílio computacional (Carmer & Swanson, 1973).

2.6 *Bootstrap*

Muitos estudos têm focado as comparações múltiplas das médias dos níveis do fator em situações de heterogeneidade de variâncias sob modelos probabilísticos normais ou não-normais. Métodos de reamostragem *bootstrap* têm sido utilizados. A idéia básica de *bootstrap*, na ausência de qualquer conhecimento sobre a população, é realizar reamostragem com reposição de tamanho n da amostra original. A distribuição *bootstrap* de algum estimador de interesse é utilizada no lugar da “distribuição teórica” deste mesmo estimador, em função da dificuldade de desenvolvê-la ou do desconhecimento da distribuição da população de onde foi obtida a amostra aleatória. A diferença entre teste de reamostragem e aleatorização está no fato de que a primeira é efetuada com reposição (Manly, 1998).

O *bootstrap* é um método computacional desenvolvido recentemente e é usado principalmente na obtenção de estimativas de parâmetros. O fato de que as medidas de precisão são obtidas diretamente dos dados, não

dependendo completamente do Teorema do Limite Central (TLC), favorece o método do *bootstrap* em suas aplicações (Efron & Tibshirani, 1993).

A técnica consiste em se retirar uma amostra de tamanho n da população e reamostrá-la com reposição, obtendo uma nova amostra de tamanho n da amostra original. Cada uma das amostras obtidas pelas reamostragens é uma amostra *bootstrap*. Esse procedimento é executado milhares de vezes, obtendo-se, assim, as estimativas dos parâmetros que serão usadas para gerar a distribuição denominada distribuição *bootstrap*. Algumas vezes, a estimativa do parâmetro de interesse obtida na amostra original é utilizada como parâmetro da função de densidade da variável aleatória correspondente. Essa densidade é uma estimativa da verdadeira densidade populacional. Por meio desta, são realizadas amostragens e, em cada etapa, é obtida uma estimativa do parâmetro de interesse. Esse processo é repetido milhares de vezes, gerando-se a distribuição de *bootstrap* que, nesse caso, é denominada *bootstrap* infinito (Conlon & Thomas, 1990).

2.7 Distribuições

2.7.1 Normal

A distribuição normal é a distribuição de uma variável aleatória contínua mais importante, amplamente utilizada em pesquisa científica e tecnológica. Realizam-se inferências assumindo normalidade dos resíduos dos modelos adotados para as variáveis consideradas (Ramalho *et al*, 2000).

A maioria das técnicas empregadas na estatística é baseada nesta distribuição. Vários fenômenos aleatórios apresentam comportamentos que podem ser descritos de forma exata ou aproximada, pelo modelo probabilístico normal; é a forma limitante para outras distribuições de probabilidade, como consequência do teorema do limite central. Muitas distribuições de estatísticas se aproximam da normal com o aumento do tamanho amostral (Ferreira, 2005).

Uma variável aleatória X é definida como normalmente distribuída $X \sim N(\mu, \sigma^2)$, com parâmetros μ e σ^2 , se a sua função densidade de probabilidade é dada por:

$$f_X(x) = f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2},$$

$-\infty < X < \infty$, em que os parâmetros μ e σ são a média e a variância da distribuição normal e satisfazem $-\infty < \mu < \infty$ e $\sigma > 0$ (Mood *et al*, 1974).

É uma distribuição que apresenta forma de sino e é simétrica em relação à média.

2.7.2 Lognormal

É uma distribuição muito utilizada em estudos econômicos (Ferreira, 2005). Seja X uma variável aleatória positiva e uma variável Y definida por $Y = \ln X$. Se Y apresenta distribuição normal, então X tem uma distribuição lognormal. A função de distribuição de probabilidade da lognormal é dada por:

$$f_X(x) = f_X(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2},$$

$0 < X < \infty$, em que os parâmetros μ e σ satisfazem $-\infty < \mu < \infty$ e $\sigma > 0$ (Mood *et al*, 1974).

A função de densidade de probabilidade acumulada da lognormal não possui forma explícita e seus valores devem ser obtidos por meio de métodos numéricos (Ferreira, 2005). A média e a variância da variável aleatória lognormal X são dadas por:

$$\mu_X = e^{\mu + \frac{1}{2}\sigma^2} \quad \text{e} \quad \sigma_X^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

2.7.3 Exponencial

A distribuição exponencial modela, em geral, o tempo de vida de insetos, de componentes eletrônicos, etc. São variáveis aleatórias que surgem no tempo entre ocorrências consecutivas de variáveis Poisson e esse tempo para que ocorra um sucesso é modelado pela distribuição exponencial. A variável aleatória exponencial é conhecida como variável aleatória contínua

de tempo de espera, ou seja, refere-se ao tempo necessário para que ocorra o primeiro sucesso (Ferreira, 2005).

A função distribuição de probabilidade da exponencial é dada por:

$$f_X(x) = f_X(x; \lambda) = \lambda e^{-\lambda x},$$

$0 < X < \infty$, em que o parâmetro λ satisfaz $\lambda > 0$ (Mood *et al*, 1974).

A média e a variância da variável aleatória exponencial X são dadas por:

$$\mu_X = \frac{1}{\lambda} \quad \text{e} \quad \sigma_X^2 = \frac{1}{\lambda^2}$$

A distribuição exponencial é um caso especial da gama e a soma de variáveis aleatórias exponenciais, independentes e identicamente distribuídas, têm distribuição gama.

3 METODOLOGIA

Foram comparados, por simulação Monte Carlo, os desempenhos de duas metodologias de agrupamento de médias na análise de variância. A primeira metodologia, de Caliński & Corsten (1985), utiliza um teste simultâneo, baseado em uma extensão da amplitude estudentizada. A segunda é uma proposta do presente trabalho, na qual o procedimento de Caliński & Corsten (1985) é estendido para situações não-normais por meio do uso de *bootstrap*. O desempenho é mensurado em relação às taxas de erro tipo I por experimento e ao poder. Para isso, foram realizadas simulações Monte Carlo sob a hipótese nula (H_0) completa e parcial e sob a hipótese alternativa (H_1), em que foram mensurados os erros tipo I e poder de acordo com a situação.

3.1 Teste de agrupamento original

Seja a amostra aleatória $X_{11}, X_{12}, \dots, X_{1r}, X_{21}, \dots, X_{2r}, \dots, X_{kr}$, em que X_{ij} representa a observação do i -ésimo nível do fator na j -ésima unidade amostral ou experimental ($i = 1, 2, \dots, k$) e ($j = 1, 2, \dots, r$). O estimador da média de cada i -ésima população é dado por:

$$\bar{X}_i = \frac{\sum_{j=1}^r X_{ij}}{r} \quad (10)$$

As médias ordenadas dos níveis do fator foram representadas por $\bar{X}_{(1)}, \bar{X}_{(2)}, \dots, \bar{X}_{(k)}$. Foram calculadas todas as $m = k(k-1)/2$ distâncias entre médias definidas por: $d_{ii'} = \bar{X}_{(i')} - \bar{X}_{(i)}$, para $i' > i = 1, 2, \dots, k$.

Assim, a matriz de distâncias $D = \{d_{ii'}\}$ com dimensão $(k \times k)$ foi criada, sendo $d_{ii'} = d_{i'i}$. Na seqüência, o método de agrupamento do vizinho mais distante foi aplicado (Johnson & Wichern, 1998). Em cada passo de fusão dos grupos, a distância R_ℓ ($\ell = 1, 2, \dots, k-1$) entre eles é registrada. O próximo passo consistiu em verificar a significância para o teste de hipótese da igualdade dos dois grupos que são candidatos a serem unidos.

O teste de Caliński & Corsten (1985) utiliza a distribuição da amplitude estudentizada ($q_\alpha(k, \nu)$) para calcular o valor- p por:

$$\text{valor-}p = Pr \left(q_\alpha(k, \nu) > \frac{R_\ell}{\sqrt{\frac{QME}{r}}} \right) \quad \ell = 1, 2, \dots, k - 1 \quad (11)$$

em que k é o número de níveis do fator, ν são os graus de liberdade associados ao QME e QME é o quadrado médio do resíduo da análise de variância.

O valor- p foi confrontado com o nível nominal de significância α . O primeiro passo de fusão i para o qual o valor- p é menor do que o nível de significância α foi considerado o ponto de corte do dendrograma e os grupos de níveis do fator foram obtidos.

3.2 Teste proposto

O teste proposto neste trabalho (*bootstrap*) difere do teste de Caliński & Corsten (1985) no cálculo do valor- p , usado para verificar a significância para o teste de hipótese da igualdade dos dois grupos que são candidatos a serem unidos. Este valor foi obtido de acordo com o procedimento seguinte. As rk observações amostrais foram combinadas em uma única amostra de onde foram realizadas r reamostragens com reposição para compor novas amostras de cada população. Assim, a hipótese nula de igualdade das médias dos níveis do fator foi imposta. Novas médias foram computadas nestas amostras e a amplitude estudentizada foi calculada por:

$$q_b = \frac{\bar{X}_{(k)}^b - \bar{X}_{(1)}^b}{\sqrt{\frac{QME_b}{r}}} \quad (12)$$

em que QME_b é o quadrado médio do resíduo da b -ésima amostra *bootstrap* e $\bar{X}_{(1)}^b$ e $\bar{X}_{(k)}^b$ são as médias amostrais ordenadas da primeira e k -ésima reamostragem de *bootstrap*, tendo o sobrescrito b utilizado para indicar a b -ésima amostra de *bootstrap*. Este processo foi repetido $B = 1000$ vezes e o conjunto com todos os valores q_b obtidos ($b = 1, 2, \dots, B$) foi utilizado para

computar o valor- p por:

$$\text{valor-}p = \frac{\sum_{b=1}^B I\left(\frac{R_\ell}{\sqrt{\frac{QME}{r}}} < q_b\right)}{B} \quad \ell = 1, 2, \dots, k-1 \quad (13)$$

em que $I(x)$ é a função indicadora de x .

Da mesma forma, o valor- p foi confrontado com α em cada ponto de fusão e o mesmo critério utilizado no teste de Caliński & Corsten (1985) foi aplicado para a obtenção dos grupos de níveis homogêneos do fator.

3.3 Simulações

Foram consideradas $N = 1000$ simulações de k populações não-estruturadas e qualitativas. Em cada simulação, foram geradas amostras de tamanho r de cada uma das k populações (níveis do fator). Foram considerados os modelos probabilísticos normal, lognormal e exponencial para as populações, sendo as duas últimas distribuições comuns na experimentação em casos reais. A lognormal, muitas vezes, é considerada como uma distribuição mais apropriada do que a normal para modelar dados de produção de plantas ou de animais, por considerar somente valores positivos da variável aleatória. O modelo exponencial é considerado adequado quando são feitas mensurações de tempo de vida de animais, plantas ou componentes eletrônicos. Estas distribuições foram utilizadas para avaliar a robustez dos dois testes considerados neste trabalho.

Assim, a variável aleatória X_{ij} pode ser originária de um desses três modelos para o i -ésimo nível do fator ($i = 1, 2, \dots, k$) e j -ésima unidade amostral ou experimental ($j = 1, 2, \dots, r$) e a sua densidade representada por $f_{X_i}(x_{ij})$.

Foram consideradas diversas configurações envolvendo o número k de níveis do fator, número de repetições r e número (δ) de erros padrões de diferença entre médias consecutivas. Assim, foram utilizados valores de k iguais a 5, 10, 20 e 80 e valores de r iguais a 4, 10 e 20. Esses valores foram combinados fatorialmente, totalizando 12 situações para cada distribuição,

tanto para o poder quanto para o erro tipo I. Sob H_1 , diferentes valores de δ foram considerados. Em todos os casos, os testes foram aplicados considerando os níveis nominais de significância α iguais a 0,01 e 0,05.

Sob H_0 completa, os parâmetros dos modelos $f_{X_i}(x_{ij}), i = 1, 2, \dots, k$ foram considerados iguais para todos os valores de i . Assim, os níveis do fator apresentavam médias e variâncias comuns. Sob H_0 parcial, foram considerados dois grupos dentro dos quais o procedimento utilizado para H_0 completa é aplicado. As densidades entre os grupos tiveram valores diferentes para os parâmetros. Sob H_1 , as densidades foram consideradas todas diferentes. No caso normal, elas se diferenciaram apenas pela média (μ), mantendo constante a variância (σ^2).

Os dois testes foram aplicados em todas as configurações simuladas e nos N experimentos gerados de cada uma delas foram estimadas as taxas de erro tipo I por experimento e o poder. As taxas de erro tipo I foram avaliadas sob H_0 completa e parcial e o poder foi avaliado sob H_0 parcial e H_1 . A situação de distribuição normal foi utilizada como ambiente favorável, pois o teste de Caliński & Corsten (1985) foi idealizado sob a pressuposição de normalidade. Os modelos lognormal e exponencial propiciaram que fossem realizadas comparações entre os dois testes em ambientes não-favoráveis.

Sob a hipótese nula completa e modelo normal, foram adotados os valores para média e variância iguais a 10 e 1, respectivamente, sem perda de generalidade. Para o modelo lognormal, a média foi igual a 0 e as variâncias iguais a 1 e 5, na escala logarítmica. Para o modelo exponencial, adotou-se o valor de $\lambda = 0,1$, o que corresponde à média igual a 10 e variância igual a 100.

Sob H_1 , no modelo normal, fixou-se a variância em 1 e variaram-se as médias dos níveis dos fatores. Entre dois níveis consecutivos do fator, foi fixada a diferença de um erro padrão da média ($\sigma_{\bar{x}} = \sigma/\sqrt{r}$). Para o modelo lognormal, foram adotados os valores 0 para a média (μ) e 1 para a variância (σ^2), na escala logarítmica. A média do primeiro nível do fator e a variância na escala exponencial dependeram desses valores originais, estipulados da seguinte forma:

$$\mu_\ell = e^{\mu + \frac{\sigma^2}{2}} \quad (14)$$

$$\sigma_\ell^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} \quad (15)$$

O i -ésimo nível ($i \geq 2$) é especificado da forma seguinte. Inicialmente, é obtida a sua média na escala exponencial, somando-se um erro padrão da média:

$$\mu_i = \mu_{i-1} + \frac{\sigma_{i-1}}{\sqrt{r}} \quad (16)$$

$$\mu_i^* = \ln(\mu_i) - \frac{\sigma^2}{2} \quad (17)$$

e a nova variância na escala exponencial fica determinada por:

$$\sigma_i^2 = e^{2\mu_i^* + 2\sigma^2} - e^{2\mu_i^* + \sigma^2}. \quad (18)$$

As equações (16), (17) e (18) são utilizadas até que os parâmetros do k -ésimo nível tenham sido especificados. É conveniente chamar a atenção para o fato de que a variância é diferente em cada nível do fator. Isso contribui ainda mais para propiciar um ambiente desfavorável à aplicação do teste original.

Para a distribuição exponencial foi utilizado um procedimento semelhante ao da lognormal, considerando $\lambda = 0, 1$. A média e a variância do primeiro nível é igual a $\mu_1 = 1/\lambda$ e $\sigma_1^2 = 1/\lambda^2$. A média do segundo nível i é obtida por $\mu_2 = \mu_1 + \sigma_1/\sqrt{r}$ de onde se determina o parâmetro λ_2 (parâmetro correspondente ao segundo nível do fator). Assim, especificou-se $\sigma_2^2 = 1/\lambda_2^2$ e repetiu-se o processo para se obter μ_3 , λ_3 e σ_3^2 . Isso foi feito até se obterem os parâmetros do k -ésimo nível.

Sob H_0 parcial, foram estipulados dois grupos de níveis do fator e, em cada um deles, as médias eram consideradas iguais. Entre os grupos foram consideradas diferenças de 1, 2, 4, 8, 16 e 32 erros padrões e modelo normal. O mesmo critério utilizado sob H_1 foi adotado neste caso.

3.4 Exemplo real

A fim de se aplicar o teste Caliński & Corsten e sua versão *bootstrap* foi utilizado um exemplo real. O objetivo do estudo original foi testar diferentes tipos de combinações de culturas de bactérias fixadoras de nitrogênio em plantas de trevo-vermelho. Havia cinco culturas de *Rhizobium trifolii*, que é própria do trevo-vermelho, mais um composto com cinco culturas de *Rhizobium meliloti*, típica da alfafa. Cada uma das cinco culturas de *R. trifolii* foi testada individualmente, com um composto das *R. meliloti*, totalizando cinco níveis do fator. O sexto nível do fator foi uma composição das cinco culturas de *Rhizobium trifolii* com as cinco culturas de *Rhizobium meliloti*. O experimento foi conduzido em casa de vegetação, utilizando delineamento inteiramente casualizado, com cinco repetições, sendo cada uma constituída de um vaso. Na Tabela 7 se encontram os teores de nitrogênio em miligramas obtidos.

TABELA 7: Teor de nitrogênio, em mg, de plantas de trevo-vermelho inoculadas com combinações de culturas de *R. trifolii* e *R. meliloti*.

	Repetições				
	1	2	3	4	5
3Dok1	19,4	32,6	27,0	32,1	33,0
3Dok5	17,7	24,8	27,9	25,2	24,3
3Dok4	17,0	19,4	9,1	11,9	15,8
3Dok7	20,7	21,0	20,5	18,8	18,6
3Dok13	14,3	14,4	11,8	11,6	14,2
Composto	17,3	19,4	19,1	16,9	20,8

FONTE: Steel & Torrie, p. 141, 1980.

A esses dados foram aplicadas as duas versões do teste de Caliński & Corsten, original e *bootstrap*, sendo adotado o nível nominal de 0,05 de significância. O objetivo foi ilustrar a aplicação de ambos os testes.

4 RESULTADOS E DISCUSSÃO

Na seqüência são apresentados os resultados das taxas de erro tipo I por experimento e dos valores de poder dos testes em diferentes situações de distribuições normais e não-normais, sob H_0 completa e parcial e sob H_1 . Também foram considerados níveis nominais de significância de 0,01 e 0,05, diferentes números de repetições e níveis do fator, entre outros parâmetros de simulação.

4.1 Erro tipo I sob H_0 completa

4.1.1 Distribuição normal

Na Tabela 8 são apresentadas as taxas de erro tipo I por experimento (TPE) sob normalidade em função de r , k e α dos testes Caliński & Cors-ten (C) e sua versão *bootstrap*(CB) para H_0 completa. As taxas observadas foram confrontadas com os intervalos de 99% de confiança para proporções considerando os níveis nominais de 0,01 e 0,05. Em todas as situações (diferentes r , k e α), ambos os testes apresentaram taxas não significativamente diferentes dos níveis nominais correspondentes.

TABELA 8: Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C), sua versão *bootstrap* (CB) em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição normal(10, 1).

k	r	C		CB	
		0,05	0,01	0,05	0,01
5	4	0,049	0,009	0,053	0,010
5	10	0,047	0,009	0,046	0,012
5	20	0,057	0,018	0,058	0,016
10	4	0,055	0,009	0,057	0,011
10	10	0,042	0,007	0,042	0,007
10	20	0,044	0,010	0,046	0,012
20	4	0,047	0,006	0,050	0,008
20	10	0,048	0,010	0,054	0,013
20	20	0,054	0,009	0,054	0,010
80	4	0,044	0,010	0,051	0,015
80	10	0,048	0,008	0,050	0,009
80	20	0,041	0,007	0,042	0,011

Assim, sob normalidade, não houve nenhuma tendência de o teste de Caliński & Corsten (1985) ser liberal ou conservativo. Este resultado confirma a afirmativa dos autores em apontar que o teste proposto controla a taxa global de erro tipo I de falsas partições. O mesmo ocorreu com a versão *bootstrap*.

A princípio, isso poderia ser um indicativo de que o teste *bootstrap* não deveria ser aplicado, devido ao maior esforço computacional exigido. No entanto, outras situações devem ser consideradas, envolvendo distribuições não-normais e H_0 parcial, que serão abordadas na seqüência.

4.1.2 Distribuições não-normais

Na Tabela 9 são apresentadas as TPE sob H_0 completa, considerando uma distribuição lognormal com parâmetros de posição e escala iguais a 0 e 1, respectivamente, na escala logarítmica. Foram considerados diferentes níveis k do fator, número de repetições r e níveis de significância α .

TABELA 9: Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição lognormal(0, 1).

k	r	0,05		0,01	
		C	CB	C	CB
5	4	0,029++	0,045	0,007	0,007
5	10	0,026++	0,041	0,006	0,006
10	4	0,039	0,048	0,008	0,009
10	10	0,036	0,041	0,006	0,010
20	4	0,067	0,054	0,022**	0,007
20	10	0,066	0,053	0,017	0,013
80	4	0,307**	0,052	0,174**	0,010
80	10	0,236**	0,058	0,131**	0,010

**Ultrapassou o LS do IC exato, com 99% de confiança para $\alpha = 0,01(0,021276)$ e $\alpha = 0,05(0,070504)$.

++Não atingiu o LI do IC exato, com 99% de confiança para $\alpha = 0,01(0,003727)$ e $\alpha = 0,05(0,033927)$.

O teste C para $\alpha = 0,05$ se mostrou conservativo ($P < 0,01$) para um menor número de níveis do fator ($k = 5$) associado a um menor número de repetições ($r \leq 10$). Para este mesmo valor de α , este teste foi significativamente ($P < 0,01$) considerado liberal para um maior número de níveis do fator ($k = 80$) considerando todos os valores de repetições. As taxas de erro tipo I por experimento, neste último caso, para $r = 4$, foram de 0,307, para $r = 10$, de 0,236. Estes valores, além de serem significativamente ($P < 0,01$) diferentes de 0,05, são expressivamente elevados. Nota-se que há uma redução de 7,1 pontos percentuais com o aumento do número de repetições de 4 para 10. Como este teste é recomendado, principalmente, em situações em que o número de níveis do fator é elevado, corre-se um risco considerável de se cometer o erro tipo I, se a distribuição for a lognormal.

Para o nível de significância α igual a 0,01, o teste C apresentou a mesma tendência de se tornar liberal, à medida que o número k de níveis do fator aumenta. Isso aconteceu para um menor valor de k ($k = 20$), quando o número de repetições foi pequeno ($r = 4$). Não houve, no entanto, nenhuma

ocorrência de casos conservativos para $\alpha = 0,01$.

Para o teste *bootstrap* (CB), não ocorreram taxas de erro tipo I por experimento significativamente diferentes dos valores nominais de significância de 0,01 e 0,05. Esse fato confirma uma importante propriedade dos testes *bootstrap*, que é a sua robustez, como é preconizado pela literatura (Manly, 1998). Neste caso particular, justifica-se o esforço computacional adicional demandado pelo teste CB e, portanto, sua utilização.

Na Figura 1 são apresentadas as TPE para a situação de H_0 completa, sob a distribuição lognormal (0, 1) para $r = 20$, em função de k . Para fins de comparação, foram adicionados os resultados sob normalidade dos testes C e de Scott-Knott (SK) (Scott & Knott, 1974) obtidos por Silva et al. (1999). Verifica-se o mesmo comportamento relatado anteriormente para os testes C e CB sob a lognormal, tanto para $\alpha = 0,05$ quanto para $\alpha = 0,01$. Sob normalidade, o mesmo padrão de comportamento também foi seguido, ou seja, houve controle do erro tipo I por experimento, para estes dois testes. O teste SK apresentou controle do erro tipo I para $k \leq 20$, mas foi conservativo ($P < 0,01$) para $k = 80$. Assim, nesta situação de H_0 completa, conclui-se que houve um melhor desempenho do teste CB no controle do erro tipo I por experimento sob normalidade ou sob distribuição lognormal (0, 1).

Na Tabela 10 são apresentadas as taxas de erro tipo I por experimento sob H_0 completa, considerando a lognormal com parâmetros de posição e escala iguais a 0 e 5, respectivamente, na escala logarítmica. Nesta situação, o parâmetro de escala foi considerado quatro vezes superior ao da situação anterior para refletir uma distribuição de maior variabilidade e de maior assimetria à direita.

O teste C apresentou taxas de erro tipo I por experimento significativamente ($P < 0,01$) inferiores aos valores nominais de 0,05 e 0,01, em geral, para pequenos valores de k ($k \leq 10$) associados a pequenos valores de r , sendo considerado conservativo. Se forem comparadas as taxas de erro tipo I por experimento deste teste sob a lognormal (0, 1) e lognormal (0, 5) nessas circunstâncias, verifica-se que houve acentuação da sua característica conservativa.

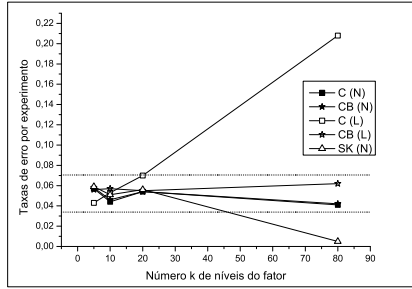
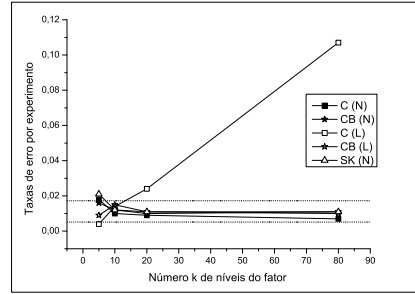
(a) $\alpha = 0,05$ (b) $\alpha = 0,01$

FIGURA 1: Taxas de erro tipo I por experimento, dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), e do teste de Scott-knott (SK), em função do número de níveis k do fator, considerando-se as distribuições normal (N) (10, 1) e lognormal (L) (0, 1), sob H_0 completa, $r = 20$, para $\alpha = 0,05$ (a) e $\alpha = 0,01$ (b). As linhas pontilhadas representam os limites superior e inferior do IC exato para proporções com 99% de confiança.

Por outro lado, o aumento do número de níveis k do fator para 80 acentuou o aspecto liberal do teste C independente do número de repetições r e do nível de significância α . Tal aspecto foi extremamente negativo neste teste, uma vez que as taxas de erro foram superiores a 0,80 e 0,90, para os níveis nominais de 0,01 e 0,05, respectivamente. Assim, este teste não pode ser considerado robusto à violação de normalidade quando se considera a distribuição lognormal. O fato de o teste ser liberal nas situações de maior número de níveis agrava ainda mais o problema, pois estas representam as situações ideais para a aplicação de testes de agrupamento de médias.

O teste CB também apresentou taxas de erro significativamente ($P < 0,01$) inferiores aos valores nominais de 0,05 e 0,01, em geral, para valores de k pequenos ($k \leq 20$) associados a valores menores de r ($r \leq 10$). Aparentemente, houve um maior efeito do número de níveis k do fator, comparado ao efeito do número de repetições r .

TABELA 10: Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição lognormal(0, 5).

k	r	0,05		0,01	
		C	CB	C	CB
5	4	0,005++	0,017++	0,000++	0,002++
5	10	0,002++	0,023++	0,000++	0,001++
5	20	0,005++	0,023++	0,000++	0,005
10	4	0,010++	0,025++	0,001++	0,001++
10	10	0,016++	0,038	0,003++	0,006
10	20	0,009++	0,032++	0,000++	0,002++
20	4	0,028++	0,029++	0,013	0,003++
20	10	0,036	0,033++	0,015	0,005
20	20	0,042	0,043	0,013	0,006
80	4	0,940**	0,040	0,847**	0,009
80	10	0,921**	0,045	0,829**	0,013
80	20	0,910**	0,055	0,825**	0,008

**Ultrapassou o LS do IC exato, com 99% de confiança para $\alpha = 0,01(0,021276)$ e $\alpha = 0,05(0,070504)$.

++Não atingiu o LI do IC exato, com 99% de confiança para $\alpha = 0,01(0,003727)$ e $\alpha = 0,05(0,033927)$.

A característica conservativa deste teste desapareceu à medida que o número de níveis k do fator aumentou de 20 para 80. Isso ocorreu tanto para $\alpha = 0,01$ quanto para $\alpha = 0,05$ e, diferente do teste C, o teste CB não se tornou liberal com $k = 80$, o que o classifica como um teste robusto.

Na Tabela 11 estão apresentadas as taxas de erro tipo I por experimento dos dois testes, em função de k , r e α , sob H_0 completa e distribuição exponencial. O teste C apresentou comportamento conservativo para $k = 5$ e $r = 4$ e liberal para $k = 80$, associado a um pequeno número de repetições ($r = 4$). O teste CB apresentou tamanho do teste sempre igual ao nível α (0,01 ou 0,05), independentemente de r e k .

TABELA 11: Taxas de erro por experimento (TPE), dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), em função do número de repetições r , número de níveis k do fator e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob H_0 completa, considerando-se a distribuição exponencial(0,1).

k	r	0,05		0,01	
		C	CB	C	CB
5	4	0,033++	0,040	0,008	0,009
5	10	0,049	0,051	0,011	0,014
10	4	0,038	0,048	0,005	0,007
10	10	0,046	0,053	0,008	0,011
20	4	0,050	0,048	0,011	0,008
20	10	0,044	0,045	0,013	0,012
80	4	0,076**	0,050	0,029**	0,007
80	10	0,070	0,051	0,019	0,010

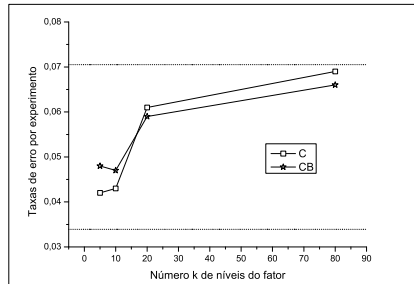
**Ultrapassou o LS do IC exato, com 99% de confiança para $\alpha = 0,01(0,021276)$ e $\alpha = 0,05(0,070504)$.

++Não atingiu o LI do IC exato, com 99% de confiança para $\alpha = 0,01(0,003727)$ e $\alpha = 0,05(0,033927)$.

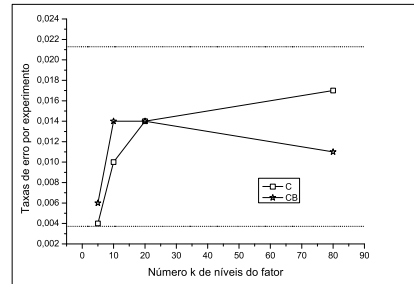
Na Figura 2 são apresentadas as taxas de erro tipo I por experimento para os dois testes, fixando $r = 20$ em função de k sob distribuição exponencial para $\alpha = 0,05$ (Figura 2a) e $\alpha = 0,01$ (Figura 2b). Nenhum dos dois testes apresentou taxas de erro tipo I por experimento significativamente ($P > 0,01$) diferentes dos níveis nominais de 0,05 ou de 0,01.

4.2 Erro tipo I sob H_0 parcial

Na Tabela 12 estão apresentadas as taxas de erro tipo I por experimento em função de k , r , δ e α . É conveniente salientar que estas taxas foram mensuradas nas comparações entre níveis do fator de um mesmo grupo, embora entre os dois grupos haja uma diferença de δ erros padrões da média do nível do fator. Obviamente, as comparações entre níveis do fator de dois grupos diferentes estão relacionadas ao poder. Com poucas exceções pode-se observar que ambos os testes apresentaram o mesmo desempenho, tanto para $\alpha = 0,01$ quanto para $\alpha = 0,05$.



(a) $\alpha = 0,05$



(b) $\alpha = 0,01$

FIGURA 2: Taxas de erro tipo I por experimento dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), em função do número de níveis k do fator, considerando-se a distribuição exponencial (E) (0,1), sob H_0 completa, $r = 20$, para $\alpha = 0,05$ (a) e $\alpha = 0,01$ (b). As linhas pontilhadas representam os limites superior e inferior do IC exato para proporções com 99% de confiança.

Pode-se observar, fixados α e r , que as TPE aumentam com o incremento do número k de níveis do fator de 5 para 80, sendo os testes liberais com δ pequenos ($\delta \leq 4$). Com $\delta \geq 8$, os testes tenderam a ser conservativos, exceto o teste CB, principalmente com maiores valores de k , associados a menores valores de r .

Borges & Ferreira (2003) avaliaram as taxas de erro tipo I por experimento em situações similares às consideradas neste trabalho sob H_0 parcial, para o teste de Scott-Knott (SK). Considerando $\delta = 4$ para $r = 10$ e $\alpha = 0,05$, esses autores encontraram TPE de 10%, 42%, 61%, 98% e 100% para $k = 5, 10, 20, 40$ e 96 níveis do fator, respectivamente. Exceto para $k = 5$, o desempenho dos testes C e CB foi superior ao do teste SK, pois as taxas de erro tipo I por experimento foram menores nos dois primeiros. O mesmo efeito de aumento da TPE com o aumento de níveis k do fator observado no teste SK foi constatado para os testes C e CB.

TABELA 12: Taxas de erro tipo I dos testes Caliński & Corsten (C) e *bootstrap* (CB), em função do número de níveis k do fator, número de repetições r , diferenças δ entre as médias, para os níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$, sob a distribuição normal e H_0 parcial.

δ	r	k	Nível Nominal de Significância (α)			
			0,05		0,01	
			C	CB	C	CB
2	4	5	0,164**	0,169**	0,061**	0,059**
		10	0,307**	0,307**	0,112**	0,123**
		20	0,407**	0,419**	0,170**	0,187**
		80	0,634**	0,666**	0,308**	0,352**
	10	5	0,204**	0,199**	0,075**	0,073**
		10	0,358**	0,364**	0,142**	0,153**
		20	0,474**	0,475**	0,202**	0,224**
		80	0,683**	0,685**	0,328**	0,341**
	20	5	0,175**	0,178**	0,077**	0,077**
		10	0,358**	0,366**	0,160**	0,166**
		20	0,478**	0,478**	0,234**	0,234**
4	4	5	0,156**	0,157**	0,095**	0,095**
		10	0,302**	0,308**	0,224**	0,233**
		20	0,547**	0,549**	0,468**	0,488**
		80	0,914**	0,914**	0,909**	0,912**
	10	5	0,158**	0,160**	0,117**	0,117**
		10	0,311**	0,311**	0,264**	0,267**
		20	0,521**	0,520**	0,493**	0,495**
		80	0,914**	0,914**	0,908**	0,910**
	20	5	0,168**	0,168**	0,129**	0,129**
		10	0,309**	0,310**	0,268**	0,268**
		20	0,500**	0,502**	0,476**	0,479**

...continua...

Tabela 12 - Cont.

δ	r	k	Nível Nominal de Significância (α)			
			0,05		0,01	
			C	CB	C	CB
8	4	5	0,016++	0,016++	0,003++	0,003+
		10	0,028++	0,032++	0,008++	0,011
		20	0,023++	0,057	0,006	0,016
		80	0,037	0,156**	0,017	0,067**
	10	5	0,027++	0,026++	0,004	0,005
		10	0,022++	0,025++	0,007	0,007
		20	0,037	0,040	0,007	0,011
		80	0,040	0,048	0,013	0,022**
	20	5	0,022++	0,021++	0,007	0,008
		10	0,030++	0,032++	0,011	0,011
		20	0,038	0,036	0,009	0,011
		80	0,032++	0,028++	0,002++	0,002++
32	4	10	0,023++	0,033++	0,005	0,014
		20	0,028++	0,103**	0,009	0,059**
		80	0,024++	0,582**	0,006	0,497**
		5	0,033++	0,036	0,004	0,005
	10	10	0,024++	0,029++	0,003++	0,005
		20	0,026++	0,033++	0,005	0,005
		80	0,040	0,048	0,013	0,022**
		5	0,024++	0,027++	0,005	0,005
	20	10	0,027++	0,031++	0,007	0,006
		20	0,024++	0,030++	0,007	0,009

**Ultrapassou o LS do IC exato, com 99% de confiança para $\alpha = 0,01$ (0,021276) e $\alpha = 0,05$ (0,070504).

++Não atingiu o LI do IC exato, com 99% de confiança para $\alpha = 0,01$ (0,003727) e $\alpha = 0,05$ (0,033927).

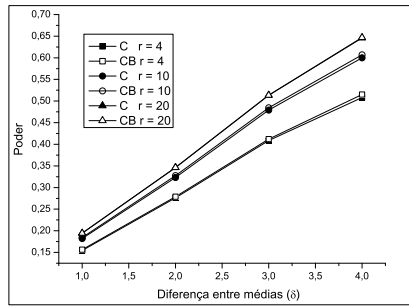
4.3 Poder sob H_1

4.3.1 Distribuição normal

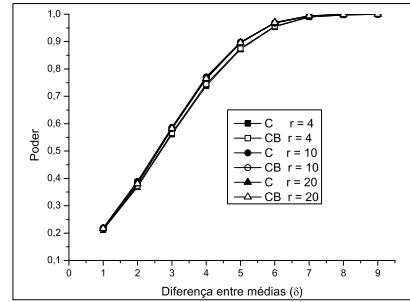
Na Figura 3 estão apresentados os gráficos para poder dos testes Caliński & Corsten (C) e sua versão *bootstrap* (CB), em função do número de repetições r , número k de níveis do fator e da diferença δ em erros padrões entre médias, considerando a distribuição normal sob H_1 e $\alpha = 0,05$. Pode-se observar, comparando-se as Figuras 3a, 3b, 3c e 3d, que há um pequeno aumento no valor do poder à medida que o valor de k cresce, fixadas as outras variáveis do estudo. Este fato é mais evidente quando k cresce de 5 para 10, com valores maiores de δ ($\delta = 4$). Como já é preconizado pela teoria, o poder cresce com o aumento de δ . Em quase todos os casos praticamente não houve diferenças entre os valores de poder dos testes C e CB. A exceção ocorre, principalmente, para $k = 5$, em que o teste CB apresenta uma pequena vantagem em relação ao teste C.

O efeito de repetição só é evidente no caso de $k = 5$. Já era esperado que não houvesse efeito do número de repetições, uma vez que as diferenças paramétricas entre médias dos níveis dos fatores foram fixadas em um valor constante (δ) em erros padrões. Assim, as diferenças na escala das variáveis alteravam-se em função do valor de r para manter constante a diferença entre os níveis do fator em número de erros padrões.

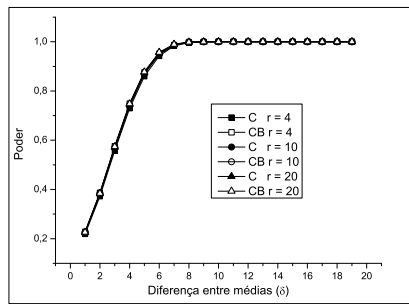
Em todos os casos (Figuras 3 a, 3b, 3c e 3d), o valor de 0,80 de poder foi atingido ou ultrapassado somente com $\delta > 4$. Quando $\delta \geq 8$, o poder se aproximou consideravelmente ou atingiu o valor de 1,00. Para valores de δ menores ($\delta \leq 4$), houve uma maior oscilação dos valores de poder quando o número de níveis do fator k aumentou de 5 para 80. Com $k = 5$ e $r = 4$, por exemplo, o poder foi de 0,15425, para o teste C e 0,15625, para o teste CB e passou para 0,21180 e 0,23073 com $k = 80$ e $r = 4$, para os testes C e CB, respectivamente. Borges & Ferreira (2003) avaliaram o poder sob H_0 parcial e verificaram o mesmo padrão de comportamento, ou seja, houve um aumento do poder com o aumento de δ (como era esperado) e de k . Os autores observaram comportamento diferente dos testes Tukey e SNK, que tenderam a apresentar menores valores de poder com o aumento de k .



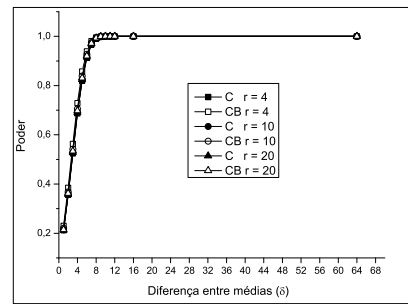
(a) $k = 5$



(b) $k = 10$



(c) $k = 20$



(d) $k = 80$

FIGURA 3: Poder dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), para diferentes números de repetições r , considerando-se a distribuição normal $(10, 1)$, em função das diferenças entre médias δ , para diferentes números de níveis k do fator, sob H_1 e $\alpha = 0,05$.

Os resultados para $\alpha = 0,01$ seguem o mesmo padrão de resposta e são apresentados na Tabela 13. Do mesmo modo, o teste CB apresentou pequena vantagem relativa para quase todas as situações, principalmente quando $\delta \leq 4$.

TABELA 13: Poder dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, sob H_1 completa, considerando-se a distribuição normal $(10, 1)$.

k	r	Teste	Diferença real entre as médias (δ)					
			$1 \sigma_{\bar{x}}$	$2 \sigma_{\bar{x}}$	$4 \sigma_{\bar{x}}$	$8 \sigma_{\bar{x}}$	$16 \sigma_{\bar{x}}$	$32 \sigma_{\bar{x}}$
5	4	C	0,07025	0,12967	0,24300	-	-	-
		CB	0,07025	0,13000	0,24300	-	-	-
5	20	C	0,10925	0,19733	0,38600	-	-	-
		CB	0,11175	0,20300	0,39600	-	-	-
10	4	C	0,17444	0,30638	0,65450	0,99100	-	-
		CB	0,17833	0,31400	0,66483	0,99000	-	-
10	20	C	0,17967	0,31800	0,68100	0,99750	-	-
		CB	0,18122	0,32088	0,68467	0,99750	-	-
20	4	C	0,19079	0,32822	0,66325	0,99042	1,00000	-
		CB	0,20147	0,34494	0,68925	0,99350	1,00000	-
20	20	C	0,20137	0,34478	0,69169	0,99400	1,00000	-
		CB	0,20363	0,34800	0,69619	0,99433	1,00000	-
80	4	C	0,19463	0,32953	0,64464	0,98194	1,00000	1,00000
		CB	0,21844	0,36622	0,70191	0,99265	1,00000	1,00000
80	20	C	0,19671	0,33277	0,65066	0,98461	1,00000	1,00000
		CB	0,20042	0,33876	0,66039	0,98679	1,00000	1,00000

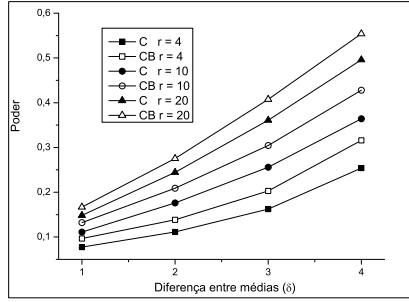
Cox & Cowpertwait (1992) compararam quatro métodos de agrupamento de médias. Duas metodologias foram propostas pelos autores, uma delas baseada na razão de verossimilhança e outra sendo uma extensão da estatística de Welch. Os dois métodos foram comparados com um teste baseado na distribuição F e com o teste de Caliński & Corsten, extensão da amplitude estudentizada. O objetivo foi comparar o desempenho dos métodos em obter os agrupamentos de médias corretos, sem assumir homogeneidade de variâncias. Sob H_1 , verificou-se que o teste de Caliński & Corsten (1985) apresenta vantagem sobre os outros métodos em obter o agrupamento correto. Nas outras simulações realizadas, o seu desempenho foi equivalente ao dos outros.

4.3.2 Distribuições não-normais

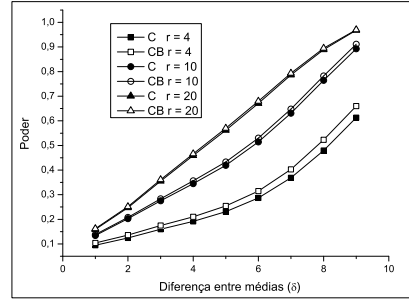
Na Figura 4 estão apresentados os gráficos para o poder dos testes C e CB, em função do número de erros padrões ($\sigma_{\bar{x}}$), repetições r e número k de níveis do fator, sob distribuição lognormal $(0, 1)$ e $\alpha = 0,05$. Pode-se observar, de maneira geral, que os valores de poder dos testes aumentam com o aumento de δ e r e com o aumento de k , exceto quando k passa de 5 para 10. A taxa de aumento em função de δ para o valor de poder é menor para situações com valores de k maiores. Exceto por este último fato, o padrão de resposta dos testes quanto ao poder foi semelhante ao observado sob normalidade (Figura 3).

É conveniente salientar que, sob esta distribuição, houve um efeito pronunciado do número de repetições para todos os valores de k , mas, principalmente para $k \leq 20$. Acredita-se que esse efeito do número de repetições tenha sido mais pronunciado na distribuição lognormal do que na normal, como consequência da heterogeneidade das variâncias na primeira situação. No sistema de simulação empregado, adotou-se o procedimento de gerar amostras aleatórias de cada nível do fator. Assim, como a média e a variância nas distribuições não-normais não são independentes, ao se determinar os parâmetros da distribuição para gerar outro nível do fator, automaticamente alteravam-se a média e a variância. Em consequência disso, espera-se que quanto maior for o valor de r , menor será a heterogeneidade e maior será o poder, haja vista que as diferenças entre médias dos níveis de fator estão fixadas em função do número de erros padrões $\sigma_{\bar{x}}$.

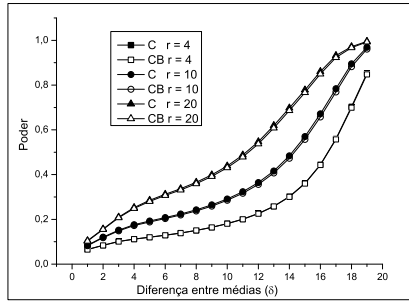
Comparando-se os testes C e CB, pode-se verificar que o desempenho do teste CB foi superior ao teste C nas situações adversas, ou seja, para $k = 5$, com todos os valores de r e $k = 10$ com $r = 4$. Para $k = 80$, o teste C superou o teste CB, mas somente para grandes valores de δ ($\delta \geq 40$). No entanto, a diferença de desempenho pode ser considerada desprezível. Esses fatos demonstram a robustez dos testes, com destaque para o teste baseado no método de *bootstrap*.



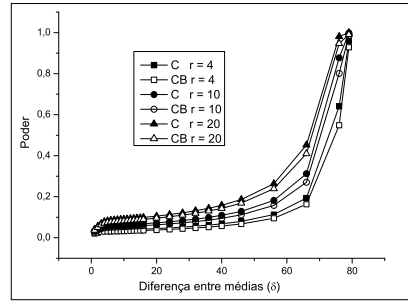
(a) $k = 5$



(b) $k = 10$



(c) $k = 20$

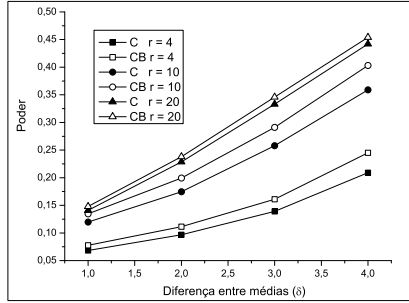


(d) $k = 80$

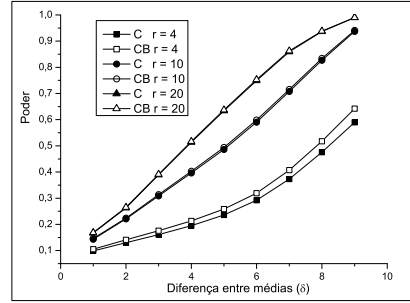
FIGURA 4: Poder dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), para diferentes números de repetições r , considerando-se a distribuição lognormal $(0, 1)$ em função das diferenças entre médias δ , para diferentes números de níveis k do fator, sob H_1 e $\alpha = 0,05$.

Na Figura 5 estão apresentados os gráficos para o poder dos testes C e CB, em função do número de erros padrões $(\sigma_{\bar{x}})$, repetições r e número k de níveis do fator, sob distribuição exponencial com parâmetro $\lambda = 0,1$ e $\alpha = 0,05$. O padrão de resposta em relação ao poder de ambos os testes nesta distribuição foi semelhante ao relatado para a distribuição lognormal e $\alpha = 0,05$. É interessante observar que, para $k = 5$, os valores de poder dos testes são menores na distribuição exponencial e, para $k \geq 5$, são menores na lognormal (Figuras 4 e 5). O efeito do teste CB no poder permaneceu superior ao do teste C para $k = 5$ e para $k = 10$ e $r = 4$. No entanto, para

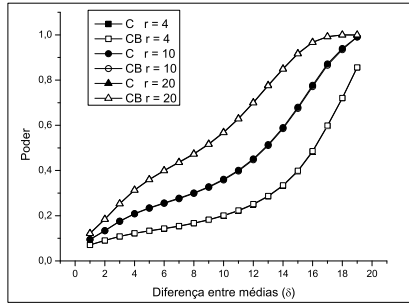
$k = 5$ e $r = 20$, houve uma redução de sua vantagem em relação ao teste C.



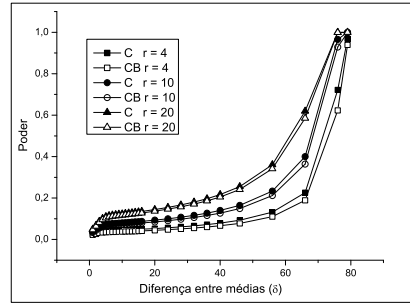
(a) $k = 5$



(b) $k = 10$



(c) $k = 20$



(d) $k = 80$

FIGURA 5: Poder dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), para diferentes números de repetições r , considerando-se a distribuição exponencial $(0,1)$, em função das diferenças entre médias δ , para diferentes números de níveis k do fator, sob H_1 e $\alpha = 0,05$.

Os resultados apresentados anteriormente referem-se ao valor nominal de 0,05 para a significância. Os resultados para $\alpha = 0,01$ estão apresentados nas Tabelas 14 e 15, para as distribuições lognormal e exponencial, respectivamente. De forma geral, o padrão de resposta dos testes quanto ao poder foi semelhante ao observado para $\alpha = 0,05$, exceto pelas menores magnitudes dos valores de poder, como já era esperado. Em função disso, nenhuma outra discussão se faz necessária.

TABELA 14: Poder dos testes de Caliński & Corsten e sua versão *bootstrap* (C e CB), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, sob H_1 completa, considerando-se a distribuição lognormal (0, 1) na escala logarítmica.

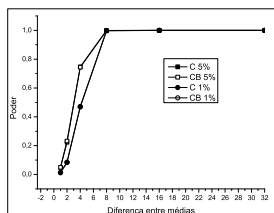
k	r	Teste	Diferença real entre as médias (δ)					
			$1 \sigma_{\bar{x}}$	$2 \sigma_{\bar{x}}$	$4 \sigma_{\bar{x}}$	$8 \sigma_{\bar{x}}$	$16 \sigma_{\bar{x}}$	$32 \sigma_{\bar{x}}$
5	4	C	0,02825	0,03867	0,09500	-	-	-
		CB	0,03400	0,04667	0,11600	-	-	-
5	20	C	0,08100	0,12567	0,28000	-	-	-
		CB	0,09475	0,14667	0,31900	-	-	-
10	4	C	0,06000	0,07575	0,11367	0,29550	-	-
		CB	0,06167	0,07850	0,11817	0,30500	-	-
10	20	C	0,13644	0,21550	0,40317	0,82600	-	-
		CB	0,13722	0,21650	0,40383	0,82800	-	-
20	4	C	0,05395	0,06917	0,09025	0,12158	0,36025	-
		CB	0,04611	0,05944	0,07713	0,10400	0,30850	-
20	20	C	0,09222	0,14067	0,22706	0,32858	0,81050	-
		CB	0,08705	0,13228	0,21350	0,30875	0,77925	-
80	4	C	0,02359	0,02756	0,03297	0,03544	0,03988	0,05317
		CB	0,01691	0,02045	0,02458	0,02638	0,02967	0,03956
80	20	C	0,03899	0,05318	0,07438	0,08374	0,09423	0,12565
		CB	0,03114	0,04368	0,06236	0,07040	0,07923	0,10565

TABELA 15: Poder para os testes de Caliński & Corsten e sua versão *bootstrap* (C e CB), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,01$, sob H_1 completa, considerando-se a distribuição exponencial $(0,1)$.

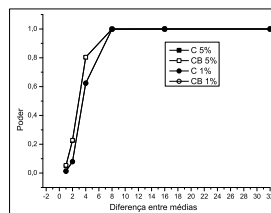
k	r	Teste	Diferença real entre as médias (δ)					
			$1 \sigma_{\bar{x}}$	$2 \sigma_{\bar{x}}$	$4 \sigma_{\bar{x}}$	$8 \sigma_{\bar{x}}$	$16 \sigma_{\bar{x}}$	$32 \sigma_{\bar{x}}$
5	4	C	0,02075	0,03033	0,06700	-	-	-
		CB	0,02600	0,03700	0,08600	-	-	-
5	20	C	0,06575	0,10667	0,21900	-	-	-
		CB	0,07375	0,12033	0,24300	-	-	-
10	4	C	0,05878	0,07525	0,10800	0,27500	-	-
		CB	0,05900	0,07425	0,10600	0,27200	-	-
10	20	C	0,14867	0,23400	0,45450	0,87900	-	-
		CB	0,14989	0,23675	0,46117	0,88750	-	-
20	4	C	0,06042	0,07717	0,10175	0,13833	0,40525	-
		CB	0,05311	0,06722	0,08900	0,12117	0,35425	-
20	20	C	0,10737	0,16461	0,28363	0,43192	0,93975	-
		CB	0,10368	0,15894	0,27531	0,41967	0,92725	-
80	4	C	0,02551	0,03022	0,03788	0,04122	0,04638	0,06183
		CB	0,01872	0,02232	0,02838	0,03088	0,03473	0,04631
80	20	C	0,04652	0,06392	0,09597	0,11421	0,12864	0,17152
		CB	0,04080	0,05709	0,08645	0,10338	0,11641	0,15521

4.4 Poder sob H_0 parcial e normalidade

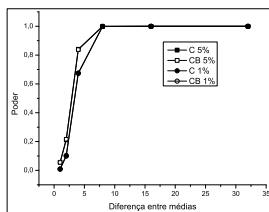
Na Figura 6 estão apresentados os valores de poder dos testes C e CB em função da diferença de δ erros padrões, número de repetições r e de níveis k do fator, sob normalidade e H_0 parcial para $\alpha = 0,05$ e $0,01$.



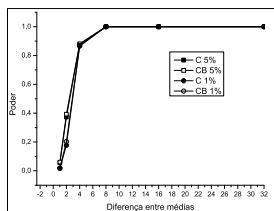
(a) $r = 4, k = 5$



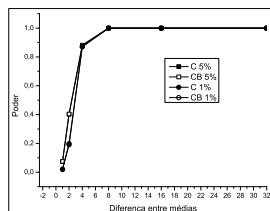
(b) $r = 10, k = 5$



(c) $r = 20, k = 5$



(d) $r = 4, k = 80$



(e) $r = 10, k = 80$

FIGURA 6: Poder dos testes de Caliński & Corsten (C) e sua versão *bootstrap* (CB), para $k = 5$ e $k = 80$ níveis do fator, considerando-se a distribuição normal (N) (10, 1), sob H_0 parcial, para $\alpha = 0,01$ e $0,05$.

Verifica-se um aumento de poder à medida que δ aumenta, como é esperado. A taxa de aumento do poder é grande, sendo que, com $\delta \geq 4$, o poder aproxima-se ou ultrapassa 0,80, principalmente para maiores valores

de k e para $\alpha = 0,05$. Os valores de poder são maiores para $\alpha = 0,05$ em relação a $\alpha = 0,01$, como também é esperado pela teoria. Esse resultado permite que se infira sobre a qualidade da simulação. Para valores de δ iguais ou superiores a 8, os valores de poder alcançam 1,00, tanto para $\alpha = 0,01$ quanto para $\alpha = 0,05$. Praticamente, não há diferenças entre os dois testes estudados.

Como as taxas de erro tipo I foram elevadas nesta situação de H_0 parcial, esperava-se que o poder fosse realmente maior do que sob H_1 , o que ocorreu. Devido à característica liberal dos testes C e CB nestas circunstâncias, não foram avaliados os seus desempenhos sob situações não-normais.

4.5 Exemplo real

Para os dados do exemplo descrito na seção 3.4, foram obtidas as médias ordenadas dos níveis do fator:

$$\begin{array}{ll} \bar{X}_{(1).} = 13,26 \text{ (3Dok13)} & \bar{X}_{(2).} = 14,64 \text{ (3Dok4)} \\ \bar{X}_{(3).} = 18,70 \text{ (Composto)} & \bar{X}_{(4).} = 19,92 \text{ (3Dok7)} \\ \bar{X}_{(5).} = 23,98 \text{ (3Dok5)} & \bar{X}_{(6).} = 28,82 \text{ (3Dok1)} \end{array}$$

A matriz de distâncias $D = \{d_{ii'}\}$ com dimensão (6 x 6) foi criada, sendo $d_{ii'} = d_{i'i}$:

$$D = \begin{pmatrix} 0,00 & 1,38 & 5,44 & 6,66 & 10,72 & 15,56 \\ 1,38 & 0,00 & 4,06 & 5,28 & 9,34 & 14,18 \\ 5,44 & 4,06 & 0,00 & 1,22 & 5,28 & 10,12 \\ 6,66 & 5,28 & 1,22 & 0,00 & 4,06 & 8,90 \\ 10,72 & 9,34 & 5,28 & 4,06 & 0,00 & 4,84 \\ 15,56 & 14,18 & 10,12 & 8,90 & 4,84 & 0,00 \end{pmatrix}$$

Na Tabela 16 são apresentados os valores obtidos ao se aplicarem os testes C e CB aos dados do exemplo. Os valores apresentados se referem às amplitudes R_ℓ entre as médias e aos valores- p obtidos em cada passo de

fusão pelos dois procedimentos.

TABELA 16: Valores das amplitudes R_ℓ e dos valores- p obtidos em cada passo de fusão, pelos procedimentos de Caliński & Corsten(C) e pela sua versão *bootstrap* (CB), para $B = 1000$ reamostragens.

Passo	Médias unidas	R_ℓ	Valor- p (C)	Valor- p (CB)
1	3 - 4	1,22	0,993	0,991
2	1 - 2	1,38	0,987	0,989
3	5 - 6	4,84	0,262	0,254
4	1 - 4	6,66	0,053	0,046
5	1 - 6	15,56	0,000	0,000

Na Figura 7 se encontra o dendrograma obtido com a utilização da função *agnes* do programa *R*, aplicada à matriz de distâncias obtida com as médias dos dados. Os pontos de corte considerados foram os pontos em que $\alpha = 0,05$ e resultaram em diferentes agrupamentos para os dois testes: para o teste C foram obtidos dois grupos de médias (1-4)(5-6) e, para o teste CB, foram obtidos três grupos (1-2)(3-4)(5-6), com $B = 1000$ reamostragens.

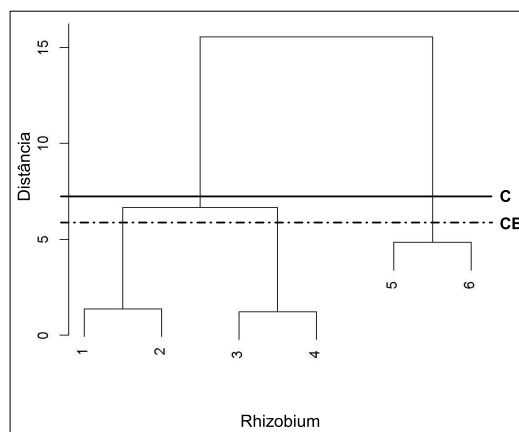


FIGURA 7: Dendrograma com os pontos de corte obtidos pelos testes C e CB para $\alpha = 0,05$.

4.6 Considerações finais

A versão *bootstrap* (CB) do teste de agrupamento de Caliński & Cors-ten (1985) foi proposta com sucesso neste trabalho e seu desempenho foi comparado com o do original, por meio de simulação Monte Carlo. Para realizar esta avaliação, foram consideradas situações simuladas sob normalidade e não-normalidade e sob H_0 completa e parcial e sob H_1 . Como critério de comparação do desempenho do teste, usaram-se as taxas de erro tipo I e o poder. Foram consideradas somente as taxas de erro tipo I por experimento (TPE).

As taxas de erro tipo I por experimento de ambos os testes sob normalidade e H_0 completa foram sempre iguais ao valor nominal de significância. Esse fato indica que o teste C é exato nesta situação, na qual seu desenvolvimento foi baseado. O teste CB apresentou desenvolvimento equivalente. A princípio, pode-se aventar que a utilização do teste CB seria não recomendada devido ao maior esforço computacional requerido. No entanto, é necessário que se considere a possibilidade de sua utilização em situações não favoráveis à aplicação do teste C, como situações de não-normalidade. É interessante salientar que o teste SK é conservativo para valores grandes de k ($k = 80$) sob normalidade, o que confirma o melhor desempenho dos testes C e CB.

A avaliação do desempenho dos testes na distribuição lognormal, que é bastante assimétrica à direita, mostrou que o teste CB foi ou conservativo ou exato, no controle do erro tipo I por experimento sob H_0 completa. O teste C, por sua vez, foi conservativo para valores pequenos de k e liberal para grandes valores de k . Na lognormal $(0, 5)$ e para $k = 80$, os valores das taxas de erro tipo I por experimento foram superiores a 0,80. Na exponencial, embora em menor intensidade, houve o mesmo desempenho relativo dos testes. Assim, o teste *bootstrap* foi considerado robusto, o que não ocorreu com o teste original.

Infelizmente, sob H_0 parcial e normalidade, nenhum dos testes apresentou controle do erro tipo I por experimento, principalmente para $\delta \leq 4$. Nesta circunstância, os testes apresentaram TPE consideradas de pouco li-

beral ($k = 5$) a muito liberal ($k = 80$). Para $\delta \geq 8$, os testes passaram para conservativos, exceto CB com $k = 80$ e $r \leq 10$. Borges & Ferreira (2003) relatam o mesmo padrão de desempenho para o teste SK.

O poder sob H_1 e H_0 , em situações normais no teste CB, é, em geral, superior, embora de forma inexpressiva, ao poder do teste C. Essa superioridade é mais evidente para menores valores de k e com $\delta \leq 4$. O poder sob H_0 parcial é maior do que sob H_1 . Isso é esperado, pois, sob H_0 parcial, não houve controle do erro tipo I.

O desempenho do poder dos testes sob H_1 e distribuições não-normais indica um pronunciado efeito do número de repetições r , principalmente para $k \leq 20$. Esse resultado foi atribuído ao sistema de simulação Monte Carlo adotado neste trabalho, que considerou heterogeneidade de variâncias, além das diferenças entre médias. Nesta situação, o teste CB apresentou melhor desempenho em relação ao teste C para $k \leq 10$, sendo superado apenas com $k = 80$ e $\delta \geq 40$. Entretanto, as diferenças de desempenho foram quase sempre desprezíveis.

5 CONCLUSÕES

O teste *bootstrap* de agrupamento de médias foi proposto com sucesso. Seu desempenho é, em geral, considerado superior ao do teste original e robusto, sendo, portanto, recomendada a sua utilização rotineiramente.

Os testes C e CB são exatos sob H_0 completa e normalidade. O teste CB sob não-normalidade e H_0 completa controla o erro tipo I por experimento e é considerado robusto. O teste C sob não-normalidade e H_0 completa é conservativo para valores pequenos de k e liberal para grandes valores de k .

Sob H_0 parcial, os testes C e CB são, em geral, liberais para $\delta \leq 4$ e conservativos para $\delta \geq 8$.

O poder do teste CB supera, de forma inexpressiva, o poder do teste C sob H_0 parcial ou sob H_1 .

6 REFERÊNCIAS BIBLIOGRÁFICAS

BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. Jaboticabal: FUNEP, 1989. 247 p.

BOARDMAN, T. J.; MOFFITT, D. R. Graphical Monte Carlo Type I error rates for multiple comparison procedures. **Biometrics**, Washington, v. 27, p. 738-744, Sep. 1971.

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

CARMER, S. G.; SWANSON, M. R. Detection of differences between means: a Monte Carlo study of five pairwise multiple comparison procedures. **Agronomy Journal**, Madison, v. 63, n. 6, p. 940-945, Nov./Dec. 1971.

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.

CALIŃSKI, T.; CORSTEN, L. C. A. Clustering means in ANOVA by Simultaneous Testing. **Biometrics**, Washington, v. 41, n. 1, p. 39-48, Mar. 1985.

CONLON, M.; THOMAS, R. G. A new confidence interval for the difference of two binomial proportions. **Computational Statistics & Data Analysis**, Amsterdam, v. 9, n. 2, p. 237-241, Mar. 1990.

COPENHAVER, M. D.; HOLLAND, B. S. Computation of the distribution of the maximum studentized range statistics with application to multiple significant testing of simple effects. **Journal of Statistical Computation**

and Simulation, v. 30, p. 1-15, 1988.

COX, T. F.; COWPERTWAIT, P. S. P. Clustering population means under heterogeneity of variance. **The Statistician**, Oxford, v. 41, n. 4, p. 591-598, 1992.

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

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

EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York: Chapman-Hall, 1993. 436 p.

FERREIRA, D. F. **Estatística Básica**. Lavras: UFLA, 2005. 664 p.

GLEASON, J. R. An accurate, non-iterative approximation for studentized range quantiles. **Computational Statistics and data analysis**, Amsterdam, v. 31, n. 2, p. 147-158, 1999.

GOMEZ, K. A.; GOMEZ, A. A. **Statistical procedures for agricultural research**. 2. ed. New York: J. Wiley & Sons, 1984. 680 p.

HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments**. New York: J. Wiley & Sons, 1987. v. 1, 495 p.

HOCHBERG, Y.; TAMHANE, A. C. **Multiple comparison procedures**. New York: J. Wiley & Sons, 1987. 450 p.

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

LUND, R. E.; LUND, J. R. Algorithm AS 190: probabilities and upper quantiles for the studentized range. **Applied Statistics**, Oxford, v. 32, n. 2, p. 204-210, 1983.

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. **Curso...** Londrina: ISBN, 2005. 290 p.

MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology**. 2. ed. London: Chapman-Hall, 1998. 399 p.

MONTGOMERY, D. C. Design and analysis of experiments. 3. ed. New York: Wiley & Sons, 1991. 649 p.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. New York: J. Wiley & Sons, 1974. 564 p.

PACHARES, J. Table of the upper 10% points of the studentized range. **Biometrika**, London, v. 46, n. 3/4, p. 461-466, 1959.

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

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

PETERSEN, R. G. Use and misuse of multiple comparison procedures. **Agronomy Journal**, Madison, v.69, n. 2, p. 205-208, Mar./Apr. 1977.

R Development Core Team. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2005. Disponível em: <http://www.R-project.org>. Acesso em 2006.

RAMALHO, M. A. P.; FERREIRA D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2000. 303 p.

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

SANTOS, C. **Novas alternativas de testes de agrupamento avaliadas por meio de simulação Monte Carlo**. Lavras: UFLA. 2000. 54 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras.

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

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

SILVA, E. C. da; 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. 3, p. 687-696, jul./set. 1999.

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

WALLER, R. A.; DUNCAN, D. B. A Bayes rule for the symmetric multiple comparison problem. **Journal American Statistical Association**, Washington, v. 67, p. 253-255, Mar, 1972.

ANEXOS

ANEXO 1: Programa de aplicação dos testes C e CB utilizando dados do exemplo proposto na seção 3.4	63
--	----

ANEXO 1: Programa de aplicação dos testes C e CB utilizando dados do exemplo proposto na seção 3.4.

```
# simular dados experimentais com média mu e variância comum e arma-
#zenar em uma matriz de 2 colunas com a primeira identificando os k níveis
# do fator e a segunda os valores da variável resposta
simula.dados = function(mu, sig2,r,k){
  x = matrix(0,r*k,2)
  i_ini = 1
  for (jj in 1:k){
    i_fin=i_ini+r-1
    x[i_ini:i_fin,2]=rnorm(r, mu, sig2^0.5)
    x[i_ini:i_fin,1]=jj
    i_ini=jj*r+1
  }
  list(trt=x[,1],y=x[,2]) }

# função para transformar uma matriz com os níveis do fator e médias
# já fornecidos em um objeto igual ao resultante de “simula.dados”
trans.fator = function(matriz){
  trt = matriz[,1]
  y = matriz[,2]
  list(trt=trt,y=y)}

# retorna um objeto com a análise de variância dos dados simulados
anova.dados=function(dados){
  dados$trt=as.factor(dados$trt)
  anovadados=lm(dados$y ~ dados$trt)
  anovadad=anova(anovadados)
  anovadad }

# calcula médias dos níveis do fator
anova.medias=function(dados){
  medias=tapply(dados$y,list(dados$trt),mean)}

# função que retorna os valores necessários para
#o cálculo do valor-p da amplitude estudentizada
valores.anava = function(dados,r){
  anava = anova.dados(dados)
```

```

S2 = anava$“Mean Sq”[2]
f = anava$“Df”[2]
raiz = sqrt(S2/r)
list(S2=S2,f=f,raiz=raiz)}

# reamostra os dados
reamostra.dados = function(dados,r,k){
  x = matrix(0,r*k,2)
  x[,1] = dados$trt
  numdados=r*k
  unif = trunc(runif(numdados)*numdados)+1
  x[,2] = dados$y[unif]
  list(trt=x[,1],y=x[,2]) }

# obtenção do vetor qihb
qihb = function(dados,r,k,B){
  for(b in 1:B) {
    dadosb = reamostra.dados(dados,r,k)
    anavab=anova.dados(dadosb)
    mediasb=anova.medias(dadosb)
    med.ordb=sort(mediasb)
    if(b == 1){
      qihb = qih.f(med.ordb,r,k,anavab)} # primeira chamada
    else{
      qihb = c(qihb,qih.f(med.ordb,r,k,anavab))} # nas chamadas con-
secutivas
  }
  qihb }

# cálculo dos valores de qih para as amostras bootstrap
qih.f = function(dad,r,k,anava){
  q = (dad[k]-dad[1])/(anava$“Mean Sq”[2]/r)**0.5
}

# cálculo da matriz de distâncias (amplitudes)
dih.f = function(dad,r,k){
  D = matrix(0,k,k)
  for (i in 1:(k-1))
  {

```

```

    seq = (i+1):k
    for (j in seq)
    {
        D[i,j] = dad[j]-dad[i]
    }
}
D }

```

agrupamento de médias usando vizinho mais distante

```

hclstr = function(D,k){
  v = matrix(1:k,k,1)
  res = matrix(1:(k-1),k-1,6)
  ct = 1
  fmin = function(D,k){ # achar mínimo
    min = D[1,2];II=1;JJ=2
    for (ii in 1:(k-1))
    for (jj in (ii+1):k){
      if (min>D[ii,jj]){
        min = D[ii,jj];II=ii;JJ=jj }
    }
  }
  list(min=min,II=II,JJ=JJ) } # fim da função fmin
  arma.res = function(ct,min,II,JJ) { # armazena resultados do passo ct
em res
    if ((v[II] > 0) & (v[JJ] > 0)) {
      res[ct,2] = min(v[II],v[JJ])
      res[ct,3] = max(v[II],v[JJ])
      res[ct,4] = min }
    if ((v[II] < 0) & (v[JJ] > 0)) {
      L = res[abs(v[II]),2]
      U = res[abs(v[II]),3]
      min1 = min(L,U);max1=max(L,U)
      res[ct,2] = min(min1,v[JJ])
      res[ct,3] = max(max1,v[JJ])
      res[ct,4] = min }
    if ((v[II] > 0) & (v[JJ] < 0)) {
      L = res[abs(v[JJ]),2]
      U = res[abs(v[JJ]),3]
      min1 = min(L,U);max1=max(L,U)
      res[ct,2] = min(min1,v[II])

```

```

    res[ct,3] = max(max1,v[II])
    res[ct,4] = min }
if ((v[II] < 0) & (v[JJ] < 0)) {
  L = res[abs(v[II]),2]
  U = res[abs(v[II]),3]
  min1 = min(L,U);max1=max(L,U)
  L = res[abs(v[JJ]),2]
  U = res[abs(v[JJ]),3]
  min2 = min(L,U);max2=max(L,U)
  res[ct,2] = min(min1,min2)
  res[ct,3] = max(max1,max2)
  res[ct,4] = min }
  #cálculo do valor-p das amostras simuladas
R = res[ct,4]
Q = R/valores$raiz
P = 1 - ptukey(Q,k,valores$f)
res[ct,5] = P
  #cálculo do valor-p das amostras bootstrap
pvalorb = (length(qihb[Q<qihb]))/B #dividir pelo n° de amostras
bootstrap
res[ct,6] = pvalorb #incluir a nova coluna de valor-p
res }

  #montar nova matriz D
monta.res = function(D,k,II,JJ) {
  D1 = matrix(0,k-1,k-1);iii=2;jjj=iii+1
  for (ii in 1:(k-1))
  for (jj in (ii+1):k) {
    if ((abs(ii-II) > 1e-6) & (abs(jj - JJ)>1e-6) &
        (abs(ii-JJ) > 1e-6) & (abs(jj - II)>1e-6)) {
      D1[iii,jjj] = D[ii,jj]
      jjj = jjj + 1
      if (jjj >= k) {
        iii = iii + 1; jjj = iii + 1 }
    }
  }
  jj = 2;
  for (ii in 1:k) {
    if ((abs(ii-II) > 1e-6) & (abs(ii-JJ) > 1e-6)){

```



```

        if ((ii>II) & (ii>JJ)) aux = max(D[II,ii],D[JJ,ii])
        if ((ii>II) & (ii<JJ)) aux = max(D[II,ii],D[jj,JJ])
        if ((ii<II) & (ii>JJ)) aux=max(D[ii,II],D[JJ,ii])
        if ((ii<II) & (ii<JJ)) aux=max(D[ii,II],D[ii,JJ])
        D1[1,jj] = aux
        jj = jj + 1 }
} D1 } # fim de monta.res

        # cria o vetor auxiliar
monta.vet = function(v,II,JJ) {
  v1=v[v != v[II]]
  v1=v1[v1 != v[JJ]]
  v1=c(-ct,v1)
  v1 }

        # realiza o agrupamento
Dct = D;kt = k
for (ct in 1:(k-1)){
  estmin = fmin(Dct,kt)
  res = arma.res(ct,estmin$min,estmin$II,estmin$JJ)
  Dct = monta.res(Dct,kt,estmin$II,estmin$JJ)
  v = monta.vet(v,estmin$II,estmin$JJ)
  kt = kt - 1 }
res }

#função para desmascarar o cluster
desmascara = function(cluster,k,nomes){
  for(i in 1:(k-1)){
    cluster[i,2] = nomes[cluster[i,2]]
    cluster[i,3] = nomes[cluster[i,3]]}
  cluster }

#função para desmascarar o grupo
desmascara.grupo = function(grupo,k,nomes){
  for(i in 1:k) {
    grupo[i,1] = nomes[grupo[i,1]] }
  grupo }

# Segunda Parte

```

```

# Criar outra matriz de resultados com os grupos formados a 5% e 1%
agrupar = function(cluster,k) {
  grupos = matrix(1:k,k,14)
  grupos[,1] = seq(1:k) #coluna das médias mascaradas
  grupos[,2] = grupos[,3] = grupos[,6] = grupos[,7] =
  grupos[,10] = grupos[,11] = grupos[,12] = grupos[,13] = 0
  #cols 2-3,6-7 guardarão os passos para q e bootstrap a 5 e 1% e
10-13 são as indicadoras
  grupos[,4] = grupos[,5] = grupos[,8] = grupos[,9] = grupos[,14] = 1
  #cols 4-5 e 8-9 são os grupos formados e 14 é o grupo original
  i = 1
  while((cluster[i,5] > 0.05) & (i < k)){
for(j in cluster[i,2]:cluster[i,3]){
  grupos[j,2] = i
  grupos[j,10] = 1 }
  if(i != (k-1)) i = i + 1
  else break
} # fim do while
# quando o valor-p for significativo, devem-se formar os grupos
  aux = 1
  for(q in 1:k){
    if(q == 1){
      grupos[q,4] = aux }
    else {
      if((grupos[q,2] == grupos[q-1,2]) & (grupos[q,10] == 1)) {
        grupos[q,4] = aux }
      if((grupos[q,2] == grupos[q-1,2]) & (grupos[q,10] == 0)) {
        aux = aux + 1
        grupos[q,4] = aux }
      if((grupos[q,2] != grupos[q-1,2]) & (grupos[q,10] == 1)) {
        aux = aux + 1
        grupos[q,4] = aux }
      if((grupos[q,2] != grupos[q-1,2]) & (grupos[q,10] == 0)) {
        aux = aux + 1
        grupos[q,4] = aux }
    }
  }
  i = 1
  while((cluster[i,5] > 0.01) & (i < k)){

```

```

for(j in cluster[i,2]:cluster[i,3]){
  grupos[j,6] = i
  grupos[j,12] = 1 }
if(i != (k-1)) i = i + 1
else break
}
aux = 1
for(q in 1:k){
  if(q == 1){
    grupos[q,8] = aux }
  else{
    if((grupos[q,6] == grupos[q-1,6]) & (grupos[q,12] == 1)) {
      grupos[q,8] = aux }
    if((grupos[q,6] == grupos[q-1,6]) & (grupos[q,12] == 0)) {
      aux = aux + 1
      grupos[q,8] = aux }
    if((grupos[q,6] != grupos[q-1,6]) & (grupos[q,12] == 1)) {
      aux = aux + 1
      grupos[q,8] = aux }
    if((grupos[q,6] != grupos[q-1,6]) & (grupos[q,12] == 0)) {
      aux = aux + 1
      grupos[q,8] = aux }
  }
}
i = 1
while((cluster[i,6] > 0.05) & (i < k)) {
for(j in cluster[i,2]:cluster[i,3]){
  grupos[j,3] = i
  grupos[j,11] = 1 }
if(i != (k-1)) i = i + 1
else break }
aux = 1
for(q in 1:k){
  if(q == 1) {
    grupos[q,5] = aux }
  else{
    if((grupos[q,3] == grupos[q-1,3]) & (grupos[q,11] == 1)) {
      grupos[q,5] = aux }
    if((grupos[q,3] == grupos[q-1,3]) & (grupos[q,11] == 0)) {

```

```

        aux = aux + 1
        grupos[q,5] = aux }
    if((grupos[q,3] != grupos[q-1,3]) & (grupos[q,11] == 1)) {
        aux = aux + 1
        grupos[q,5] = aux }
    if((grupos[q,3] != grupos[q-1,3]) & (grupos[q,11] == 0)) {
        aux = aux + 1
        grupos[q,5] = aux }
    }
}
i = 1
while((cluster[i,6] > 0.01) & (i < k)) {
for(j in cluster[i,2]:cluster[i,3]) {
    grupos[j,7] = i
    grupos[j,13] = 1 }
    if(i != (k-1)) i = i + 1
    else break
}
aux = 1
for(q in 1:k){
    if(q == 1) {
        grupos[q,9] = aux }
    else {
        if((grupos[q,7] == grupos[q-1,7]) & (grupos[q,13] == 1)) {
            grupos[q,9] = aux }
        if((grupos[q,7] == grupos[q-1,7]) & (grupos[q,13] == 0)) {
            aux = aux + 1
            grupos[q,9] = aux }
        if((grupos[q,7] != grupos[q-1,7]) & (grupos[q,13] == 1)) {
            aux = aux + 1
            grupos[q,9] = aux }
        if((grupos[q,7] != grupos[q-1,7]) & (grupos[q,13] == 0)) {
            aux = aux + 1
            grupos[q,9] = aux }
    }
}
grupos }

calinski = function(dados){

```

```

dados=trans.fator(dados)
medias = anova.medias(dados)
med.ord = sort(medias)
nomes = rownames(med.ord)
nomes = as.numeric(nomes)
valores = valores.anava(dados,r)
qihb1 = qihb(dados,r,k,B)
D = dih.f(med.ord,r,k)
for(i in 1:(k-1)) {
  for(j in (i+1):k) {
    D[j,i] = D[i,j] }
}
library(cluster)
agn <- agnes(D, diss = TRUE, method = "complete")
plot(agn)
cluster = hclstr(D,k,valores,qihb1)
grupos = agrupar(cluster,k)
grupo.desm = desmascara.grupo(grupos,k,nomes)
agrupamento = matrix(0,k,5)
agrupamento[,1] = grupo.desm[,1]
agrupamento[,2] = grupo.desm[,4]
agrupamento[,3] = grupo.desm[,5]
agrupamento[,4] = grupo.desm[,8]
agrupamento[,5] = grupo.desm[,9]
list(Passo_Medias_R_pC_pCB = cluster,
Medias_GC5_GCB5_GC1_GCB1 = agrupamento) }

# Exemplo de aplicação dos testes C e CB - dados de Steel & Torrie (1980)
k = 6 # número de níveis do fator
r = 5 # número de repetições
B = 1000 # número de reamostragens bootstrap
# dados - inserir valores da variável resposta colocando na ordem:
#todas as repetições de cada nível do fator
Rhizobium = matrix(c(rep(seq(1:k),each=r),19.4, 32.6, 27, 32.1, 33, 17.7,
24.8, 27.9, 25.2, 24.3, 17, 19.4, 9.1, 11.9, 15.8, 20.7, 21, 20.5, 18.8, 18.6, 14.3,
14.4, 11.8, 11.6, 14.2, 17.3, 19.4, 19.1, 16.9, 20.8),r*k,2)
C = calinski(Rhizobium)
C

```
