

**COMPARAÇÕES MÚLTIPLAS BAYESIANAS  
EM MODELOS NORMAIS  
HOMOCEDÁSTICOS E  
HETEROCEDÁSTICOS**

**PAULO CÉSAR DE RESENDE ANDRADE**

**2008**

**PAULO CÉSAR DE RESENDE ANDRADE**

**COMPARAÇÕES MÚLTIPLAS BAYESIANAS EM MODELOS  
NORMAIS HOMOCEDÁSTICOS E HETEROCEDÁSTICOS**

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

Orientador

Prof. Dr. Daniel Furtado Ferreira

LAVRAS  
MINAS GERAIS – BRASIL  
2008

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

Andrade, Paulo César de Resende.

Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos / Paulo César de Resende Andrade. -- Lavras : UFLA, 2008.  
96 p. : il.

Tese (Doutorado) – Universidade Federal de Lavras, 2008.

Orientador: Daniel Furtado Ferreira.

Bibliografia.

1. Amplitude padronizada. 2. Erro tipo I. 3. Monte Carlo. 4. Poder. 5. Simulação. I. Universidade Federal de Lavras. II. Título.

CDD – 519.54

**PAULO CÉSAR DE RESENDE ANDRADE**

**COMPARAÇÕES MÚLTIPLAS BAYESIANAS EM MODELOS  
NORMAIS HOMOCEDÁSTICOS E HETEROCEDÁSTICOS**

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

APROVADA em 12 de fevereiro de 2008

Prof. Dr. Ednaldo Carvalho Guimarães	UFU
Prof. Dr. Marcelo Tavares	UFU
Prof. Dr. João Domingos Scalon	UFLA
Prof. Dr. Renato Ribeiro de Lima	UFLA

Prof. Dr. Daniel Furtado Ferreira  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS – BRASIL

## **OFEREÇO**

A DEUS, pela vida e pela saúde.

## **DEDICO**

À minha esposa Valéria

Aos meus filhos Guilherme e Júlia

Aos meus pais, Valter e Ana Maria

## **AGRADECIMENTOS**

A Deus, por tudo.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas (DEX), pela oportunidade de realizar o Doutorado.

Ao meu orientador, Daniel, pelo profissionalismo, pela boa vontade, pela orientação, críticas, sugestões e, principalmente, pela amizade, compreensão e ensinamentos durante o curso. Sempre lembrarei de seus conselhos e de sua imensa competência.

Aos professores do DEX, pela atenção e capacidade que contribuíram muito para a minha formação. Aos funcionários do DEX, pela amizade e ajuda. Aos colegas de turma, pela amizade, pelos estudos e troca de conhecimentos.

Ao irmão Valter, pelo companheirismo, colaboração, amizade e preocupação, mesmo à distância, e por sempre me ajudar com suas opiniões. À irmã Adriana, pelo exemplo de determinação, pelo incentivo e por estar sempre torcendo por mim. Aos sobrinhos, cunhados e sogros, pelo estímulo.

Aos meus pais Valter e Ana Maria, pelo amor, carinho, por acreditar em minha capacidade e por tudo mais que puderam proporcionar. Levarei para vida toda seus valiosos conselhos.

Aos meus filhos Guilherme e Júlia, fontes de inspiração, encantamento e incentivo para superar todas as dificuldades.

À minha esposa Valéria, pelo amor, pelo apoio, pela força, pelo estímulo, pela compreensão, por vibrar comigo em cada vitória, por estar disposta a me ouvir a qualquer hora e por me acompanhar em cada momento de minha vida, me ajudando com suas opiniões, tão importantes para a conclusão dessa etapa.

A todos aqueles que direta ou indiretamente colaboraram e contribuíram para a realização deste trabalho.

## SUMÁRIO

LISTA DE TABELAS. . . . .	i
LISTA DE FIGURAS. . . . .	iii
RESUMO. . . . .	iv
ABSTRACT. . . . .	vi
1 INTRODUÇÃO. . . . .	1
2 REFERENCIAL TEÓRICO. . . . .	3
2.1 Inferência bayesiana. . . . .	3
2.1.1 Introdução. . . . .	3
2.1.2 Teorema de Bayes. . . . .	4
2.1.3 Distribuições <i>a priori</i> . . . . .	6
2.1.4 Distribuição <i>a posteriori</i> . . . . .	9
2.1.5 Estimação. . . . .	10
2.1.5.1 Estimação pontual . . . . .	10
2.1.5.2 Estimação por intervalos. . . . .	11
2.1.5.3 Intervalos de credibilidade e HPD's. . . . .	12
2.2 Estudo do modelo normal. . . . .	13
2.2.1 Distribuição normal com média conhecida e variância desconhecida. . . . .	14
2.2.2 Distribuição normal com média e variância desconhecidas. . . . .	15
2.3 Distribuição t multivariada. . . . .	18
2.4 Teoria bayesiana e freqüentista na tomada de decisão. . . . .	20
2.5 Análise de variância. . . . .	21
2.6 Procedimentos de comparações múltiplas. . . . .	23
2.6.1 Comparações entre médias. . . . .	23
2.6.2 Erros envolvidos e poder. . . . .	25
2.6.3 Testes de comparações múltiplas. . . . .	28
2.6.3.1 Introdução. . . . .	28
2.6.3.2 Teste <i>t</i> de Student. . . . .	30
2.6.3.3 Teste de Tukey. . . . .	32
2.6.3.4 Teste SNK. . . . .	33
2.6.3.5 Teste de Duncan. . . . .	33
2.6.3.6 Teste de Scheffé. . . . .	34
2.6.3.7 Testes de agrupamento. . . . .	35
2.6.3.8 Testes de comparação múltipla via <i>bootstrap</i> . . . . .	36
2.6.3.9 Algumas considerações. . . . .	36
2.6.4 Testes de comparações múltiplas bayesianos. . . . .	38
3 METODOLOGIA. . . . .	44
3.1 Resultados teóricos. . . . .	44
3.3 Distribuição nula da amplitude padronizada . . . . .	48
3.4 Estatísticas descritivas . . . . .	49

3.5 Exemplos . . . . .	52
3.5.1 Exemplos simulados sob $H_0$ . . . . .	52
3.5.2 Exemplos simulados sob $H_1$ . . . . .	52
3.5.3 Exemplo real . . . . .	53
4 RESULTADOS E DISCUSSÃO. . . . .	55
4.1 Resultados sob $H_0$ . . . . .	55
4.2 Resultados sob $H_1$ . . . . .	65
4.3 Resultados para dados reais. . . . .	74
4.4 Considerações finais. . . . .	82
5 CONCLUSÕES. . . . .	84
REFERÊNCIAS BIBLIOGRÁFICAS. . . . .	85
ANEXOS. . . . .	91



## LISTA DE TABELAS

1	Teor de nitrogênio, em mg, de plantas de trevo-vermelho inoculadas com combinações de culturas de <i>R. trifolii</i> e <i>R. metiloti</i> . . . . .	54
2	Médias originais e <i>a posteriori</i> obtidas a partir de dados simulados sob $H_0$ . . . . .	56
3	Quantis superiores da amplitude padronizada da distribuição <i>a posteriori</i> sob $H_0$ . . . . .	57
4	Resultado dos testes de comparações múltiplas utilizando os quantis 95% da distribuição <i>a posteriori</i> da amplitude padronizada $q$ . . . . .	58
5	Intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada <i>a posteriori</i> $q$ . . . . .	59
6	Probabilidades <i>a posteriori</i> $P(\mu_i > \mu_r)$ , considerando todos os pares $(\mu_i, \mu_r)$ da distribuição conjunta <i>a posteriori</i> . . . . .	61
7	Probabilidades do intervalo de confiança bayesiano (56) <i>a posteriori</i> conter o valor zero. . . . .	63
8	HPD's de 95% para cada cadeia $\mu_i - \mu_r$ sob $H_0$ . . . . .	65
9	Médias originais e <i>a posteriori</i> obtidas a partir de dados simulados sob $H_1$ . . . . .	66
10	Quantis superiores da amplitude padronizada da distribuição <i>a posteriori</i> sob $H_1$ . . . . .	67
11	Resultado dos testes de comparações múltiplas utilizando os quantis 95% da distribuição <i>a posteriori</i> da amplitude padronizada $q$ . . . . .	68
12	Intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada <i>a posteriori</i> $q$ . . . . .	69

13	Probabilidades <i>a posteriori</i> $P(\mu_i > \mu_{i'})$ , considerando todos os pares $(\mu_i, \mu_{i'})$ da distribuição conjunta <i>a posteriori</i> . . . . .	70
14	Probabilidades do intervalo de confiança bayesiano (56) <i>a posteriori</i> conter o valor zero. . . . .	73
15	HPD's de 95% para cada cadeia $\mu_i - \mu_{i'}$ sob $H_1$ . . . . .	74
16	Médias originais e <i>a posteriori</i> obtidas a partir dos dados reais. . . . .	74
17	Quantis superiores da amplitude padronizada da distribuição <i>a posteriori</i> . . . . .	75
18	Resultado dos testes de comparações múltiplas utilizando os quantis 95% da distribuição <i>a posteriori</i> da amplitude padronizada $q$ . . . . .	76
19	Intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada <i>a posteriori</i> $q$ . . . . .	77
20	Probabilidades <i>a posteriori</i> $P(\mu_i > \mu_{i'})$ , considerando todos os pares $(\mu_i, \mu_{i'})$ da distribuição conjunta <i>a posteriori</i> . . . . .	78
21	Probabilidades do intervalo de confiança bayesiano (56) <i>a posteriori</i> conter o zero. . . . .	80
22	HPD's de 95% para cada cadeia $\mu_i - \mu_{i'}$ . . . . .	80
23	Legenda dos procedimentos de comparações múltiplas. . . . .	81
24	Resultados dos vários procedimentos quanto às comparações múltiplas realizadas. . . . .	82

## LISTA DE FIGURAS

1	Distribuição da amplitude padronizada $q$ sob $H_0$ (a) comportado e (b) não comportado. . . . .	57
2	Densidade <i>a posteriori</i> de $\mu_i - \mu_{i'}$ para (a) $i = 1$ e $i' = 4$ e (b) $i = 4$ e $i' = 6$ sob $H_0$ comportado. . . . .	61
3	Densidade <i>a posteriori</i> de $\mu_i - \mu_{i'}$ para (a) $i = 1$ e $i' = 4$ e (b) $i = 4$ e $i' = 6$ sob $H_0$ não comportado. . . . .	62
4	Distribuição <i>a posteriori</i> da amplitude padronizada $q$ sob $H_1$ (a) comportado e (b) não comportado. . . . .	67
5	Densidade <i>a posteriori</i> de $\mu_i = \mu_{i'}$ para (a) $i = 3$ e $i' = 5$ e (b) $i = 5$ e $i' = 6$ sob $H_1$ comportado. . . . .	71
6	Densidade <i>a posteriori</i> de $\mu_i = \mu_{i'}$ para (a) $i = 3$ e $i' = 5$ e (b) $i = 5$ e $i' = 6$ sob $H_1$ não comportado. . . . .	72
7	Distribuição da amplitude padronizada $q$ . . . . .	75
8	Densidade <i>a posteriori</i> de $\mu_i = \mu_{i'}$ para (a) $i = 2$ e $i' = 4$ e (b) $i = 1$ e $i' = 6$ . . . . .	79

## RESUMO

ANDRADE, Paulo César de Resende. **Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos**. 2008. 96 p. (Tese Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.<sup>1</sup>

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. Métodos baseados em análise de agrupamento têm sido propostos para contornar o problema da ambigüidade, mas apresentam o problema de serem válidos apenas sob normalidade. Uma alternativa é o uso de procedimentos bayesianos. Este trabalho teve por objetivo propor alternativas bayesianas para comparações múltiplas considerando os casos de homogeneidade e heterogeneidade de variâncias. A metodologia utilizada foi estabelecida a partir da distribuição *a posteriori* t multivariada. Foram geradas  $k$  cadeias de médias, utilizando o método de Monte Carlo assumindo médias constantes, impondo a hipótese nula  $H_0$  no método bayesiano. Foi realizada a geração da amplitude padronizada da *posteriori*, sob  $H_0$ , obtida na distribuição *a posteriori* das médias, contemplando a possibilidade de se analisar tanto o caso de variâncias heterogêneas como o caso de variâncias homogêneas. Para descrever os resultados finais das análises dos métodos propostos foi considerado um exemplo real e quatro exemplos simulados. Nos exemplos foram simulados dados de experimentos sob  $H_0$  e sob  $H_1$ , sendo que a diferença entre as médias dos níveis do fator foi de dois erros padrões. Em ambos foram consideradas situações balanceadas e homocedásticas e não-balanceadas e heterocedásticas. Para sintetizar os resultados, inicialmente, foram obtidas as densidades *a posteriori* da distribuição da amplitude padronizada  $q$ . Para realizar a inferência a respeito da hipótese  $H_0$ , foram obtidos a diferença mínima significativa e o intervalo de confiança bayesiano. Foi calculada a probabilidade *a posteriori*  $P(\mu_i > \mu_j)$  e obtida uma nova cadeia com os limites inferior e superior de um intervalo *a posteriori* para cada par de médias, calculando-se a probabilidade *a posteriori* dos intervalos conterem o valor zero. Obteve-se o HPD nas cadeias *a posteriori* da distribuição das diferenças de duas médias, utilizando-se o pacote BOA. Os procedimentos baseados na amplitude padronizada foram superiores aos demais procedimentos estudados por terem controlado, nos exemplos simulados, o erro tipo I e detectadas a maior parte das diferenças sob  $H_1$ . Os procedimentos de comparações múltiplas bayesianos foram propostos com sucesso.

---

<sup>1</sup> Orientador: Daniel Furtado Ferreira - UFLA.

## ABSTRACT

ANDRADE, Paulo César de Resende. **Bayesian multiple comparisons in homocedastic and heterocedastic normal models**. 2008. 96 p. (Thesis Doctor in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, Minas Gerais, Brazil.<sup>2</sup>

Multiple comparison procedures are used to compare factor's levels means. Nevertheless, the most popular tests show problems concerning the ambiguity of results and the control of type I error. Methods based on cluster analysis have been proposed to avoid ambiguity, but they present the problem of being valid only under normality. An alternative to avoid this problem is the use of bayesian procedures. This work proposed bayesian alternatives for multiple comparisons, considering the cases of homogeneity and heterogeneity of variances. The methodology used was the *a posteriori* multivariate  $t$  distribution. A number of  $k$  chains of averages were generated, using the Monte Carlo method, assuming constant averages, imposing the null hypothesis  $H_0$  in the bayesian method. It was done the standardized range of a *posteriori*, under  $H_0$ , obtained in a *posteriori* distribution of the averages, contemplating the possibility of analyzing both the case of heterogeneous variances and the case of homogeneous variances. To describe the final results of the analyses of the proposed methods, it was considered one real example and four simulated examples. In all of them, data from experiments under  $H_0$  and  $H_1$  were simulated, being the difference between the means of the levels of the factor two standard errors. In both situations, balanced and homocedastics and not-balanced and heterocedastics were considered. To synthesize the results, the *posteriori* densities of the distribution of standardized range  $q$  were initially gotten. To program the inference about the hypothesis  $H_0$ , the minimum significant difference and the bayesian confidence interval were obtained. It was calculated a *posteriori* probability  $P(\mu_i > \mu_j)$  and obtained a new chain with the lower and upper limits of a *posteriori* interval for each pair of averages, calculating the *a posteriori* probability of the intervals to be zero. It was obtained the HPD in chains of *posteriori* distribution of the differences of two averages, using it in the BOA package. The procedures based on the standardized range were higher than the other studied procedures, once they controlled, in the simulated examples, the type I error and detected most of the differences under  $H_1$ . The bayesian procedures of multiple comparisons were considered successful.

---

<sup>2</sup> Adviser: Daniel Furtado Ferreira – UFLA.

# 1 INTRODUÇÃO

Na experimentação, o pesquisador, geralmente, utiliza a análise de variância para avaliar a significância dos efeitos dos tratamentos, deparando-se com o problema de comparar médias de diferentes níveis do fator. Para verificar diferenças entre médias de tratamentos são sugeridos vários procedimentos de comparações múltiplas, entre eles: 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.

Muitos procedimentos de comparações múltiplas controlam um dos dois tipos de taxas de erro tipo I: por comparação ou por experimento. Nenhum desses procedimentos é sempre apropriado. Alcançar alto poder para detectar diferenças significativas faz a taxa de erro do tipo I por experimento padecer. Por outro lado, controlar a taxa de erro para todas inferências simultâneas torna o procedimento muito conservativo, ou seja, declara poucas diferenças significativas, quando elas existem.

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 Calinski & 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.

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, utilizando métodos de reamostragem *bootstrap*, para tentar contornar o problema.

Uma alternativa é o uso de procedimentos bayesianos. Uma quantidade razoável de artigos, a maioria recente, leva em consideração o problema de comparações múltiplas do ponto de vista bayesiano. Porém, nenhum destes procedimentos enfoca o caso normal homocedástico ou heterocedástico.

Assim, este trabalho tem por objetivos propor alternativas bayesianas para comparações múltiplas, considerando os casos de homogeneidade e heterogeneidade de variâncias em modelos com distribuição normal, além de serem ilustradas em exemplos simulados e em dados reais.

## 2 REFERENCIAL TEÓRICO

Em conformidade com a metodologia estatística utilizada neste trabalho, dividiu-se o referencial teórico nos seguintes tópicos: inferência bayesiana (Seção 2.1); estudo do modelo normal (Seção 2.2); distribuição t multivariada (Seção 2.3); teoria bayesiana e freqüentista na tomada de decisão (Seção 2.4); análise de variância (Seção 2.5); procedimentos de comparações múltiplas (Seção 2.6). Posteriormente, procedeu-se a descrição de cada seção.

### 2.1 Inferência bayesiana

#### 2.1.1 Introdução

Segundo Paulino et al. (2003), a análise bayesiana de dados é uma alternativa importante aos procedimentos clássicos de estimação e testes de hipóteses que vem recebendo espaço crescente em aplicações em diversas áreas, sendo que em problemas ambientais e de manejo e conservação de populações animais tem recebido atenção especial. Trata-se de um procedimento geral de inferência a partir de dados, utilizando modelos probabilísticos tanto para descrever as variabilidades em quantidades observadas ( $Y$ ) quanto para descrever incertezas em quantidades sobre as quais queremos aprender (geralmente os parâmetros dos modelos ou funções destes parâmetros  $g(\theta)$ ).

A primeira abordagem bayesiana na solução de problemas da inferência estatística data do ano de 1763, com a publicação do livro póstumo “*An essay towards solving a problem in the doctrine of chances*”, do reverendo Thomas Bayes, por Richard Price, de acordo com Paulino et al. (2003). As idéias bayesianas e suas aplicações à estatística foram ressuscitadas em meados do



século passado, por Jeffreys (1961). A partir daí a lista de bayesianos foi avolumando-se paulatinamente.

Segundo Paulino et al. (2003), no modelo clássico, o parâmetro  $\theta$ ,  $\theta \in \Theta$ , é um escalar ou vetor desconhecido, mas fixo. No modelo bayesiano o parâmetro  $\theta$ , é tomado como um escalar ou vetor aleatório (não observável). A filosofia bayesiana é a seguinte: o que é desconhecido é incerto e toda incerteza deve ser quantificada em termos de probabilidade.

Segundo Gelman et al. (2003), a informação que se tem sobre uma quantidade de interesse é fundamental na Estatística. O verdadeiro valor de  $\theta$  é desconhecido e a idéia é tentar reduzir este desconhecimento. Além disso, a intensidade da incerteza a respeito de  $\theta$  pode assumir diferentes graus. Do ponto de vista bayesiano, estes diferentes graus de incerteza são representados através de modelos probabilísticos para  $\theta$ . Neste contexto, é natural que diferentes pesquisadores possam ter diferentes graus de incerteza sobre  $\theta$  (especificando modelos distintos). Sendo assim, não existe nenhuma distinção entre quantidades observáveis e os parâmetros de um modelo estatístico, sendo que todos são considerados quantidades aleatórias.

### **2.1.2 Teorema de Bayes**

Seja uma quantidade de interesse desconhecida  $\theta$  (tipicamente não observável). A informação de que dispomos sobre  $\theta$ , resumida probabilisticamente através de  $p(\theta)$ , pode ser aumentada observando-se uma quantidade aleatória  $Y$  relacionada com  $\theta$ . A distribuição amostral  $p(y|\theta)$  define esta relação. A idéia de que, após observar  $Y = y$ , a quantidade de informação sobre  $\theta$  aumenta é bastante intuitiva e o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação (Gelman et al., 2003). Assim,

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(\theta,y)d\theta} \cdot \quad (1)$$

Para um valor fixo  $y$ , a função  $l(\theta; y) = p(y|\theta)$  fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de  $\theta$ , enquanto  $p(\theta)$  é chamada distribuição *a priori* de  $\theta$ . Estas duas fontes de informação, *priori* e verossimilhança, são combinadas levando à distribuição *a posteriori* de  $\theta$ ,  $p(\theta|y)$  (Gelman et al., 2003). Assim, como o denominador não depende de  $\theta$ , a forma usual do teorema de Bayes é

$$p(\theta|y) \propto l(\theta; y) p(\theta). \quad (2)$$

Observa-se que na fórmula (1), o termo  $1/p(y)$  não depende de  $\theta$ , funcionando como uma constante normalizadora de  $p(\theta|y)$ . A constante normalizadora da *posteriori* pode ser facilmente recuperada, pois  $p(\theta|y) = k p(y|\theta) p(\theta)$ , em que

$$k^{-1} = \int p(y|\theta) p(\theta) d\theta = E[p(Y|\theta)] = p(y), \quad (3)$$

a qual é chamada de distribuição preditiva. Segundo Ehlers (2005), esta é a distribuição esperada para a observação  $y$  dado  $\theta$ . Assim,

- antes de observar  $Y$  podemos checar a adequação da *priori* fazendo previsões via  $p(y)$ .
- se  $Y$  observado recebe pouca probabilidade preditiva, o modelo deve ser questionado.

Segundo Paulino et al. (2003), o teorema de Bayes propõe caracterizar a aprendizagem com a experiência, isto é, a modificação da atitude inicial em relação aos “antecedentes”, “causas”, “hipóteses” ou “estados” depois de obter a informação adicional de que certo acontecimento ou acontecimentos se realizaram.

A função de verossimilhança representa o meio através do qual os dados,  $y$ , transformam o conhecimento *a priori* sobre  $\theta$ ; quer dizer, a verossimilhança pode ser interpretada como expressão da informação sobre  $\theta$  fornecida pelos dados  $y$ .

Paulino et al. (2003) resumem: para os bayesianos, a distribuição *a posteriori* incorpora, por via do teorema de Bayes, toda a informação disponível sobre o parâmetro (informação inicial + informação da experiência ou da amostra).

O princípio da verossimilhança postula que para fazer inferência sobre uma quantidade de interesse  $\theta$  só importa aquilo que foi realmente observado e não aquilo que “poderia” ter ocorrido, mas efetivamente não ocorreu.

### **2.1.3 Distribuições *a priori***

A utilização de informação *a priori* em inferência bayesiana requer a especificação de uma distribuição *a priori* para a quantidade de interesse  $\theta$ . Esta distribuição deve representar (probabilisticamente) o conhecimento que se tem sobre  $\theta$  antes da realização do experimento (Paulino et al., 2003).

A partir do conhecimento que se tem sobre  $\theta$ , pode-se definir uma família paramétrica de densidades. Neste caso, segundo Ehlers (2005), a distribuição *a priori* é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros

indexadores da família de distribuições *a priori* são chamados de hiperparâmetros para distingui-los dos parâmetros de interesse  $\theta$ .

Gelman et al. (2003) afirmam que esta abordagem, em geral, facilita a análise, e o caso mais importante é o de *prioris* conjugadas. A idéia é que as distribuições *a priori* e *a posteriori* (Seção 2.1.4) pertençam à mesma classe de distribuições. Assim, a atualização do conhecimento que se tem em relação ao parâmetro  $\theta$  envolve apenas uma mudança nos hiperparâmetros.

Segundo Paulino et al. (2003), quando não se tem informação *a priori* palpável ou o conhecimento *a priori* é pouco relevante relativamente à informação cadastral, usa-se distribuições não informativas<sup>3</sup>. A primeira idéia de “não informação” *a priori* que se pode ter é pensar em todos os possíveis valores de  $\theta$  como igualmente prováveis, com uma distribuição *a priori* uniforme. Um grupo especial de *prioris* não informativas, cuja aplicação é mais ampla, em face de diversos problemas, foi construído por Jeffreys (1961).

A classe de *prioris* não informativas proposta por Jeffreys (1961) é invariante a transformações um a um, embora, em geral, seja imprópria e será definida a seguir (Paulino et al., 2003). Antes, porém, precisamos da definição da medida de informação de Fisher.

Definição 1: Considere uma única observação  $Y$  com função de (densidade) de probabilidade  $p(y|\theta)$ . A medida de informação esperada de Fisher de  $\theta$  através de  $Y$  é definida como

$$I(\theta) = E \left[ - \frac{\partial^2 \log p(\theta | y)}{\partial \theta^2} \right]. \quad (4)$$

---

<sup>3</sup> Discussões mais detalhadas são encontradas em Berger (1985), Box & Tiao (1992), Bernardo & Smith (1994) e O’Hagan (1994).

Se  $\theta$  for um vetor paramétrico define-se então a matriz de informação esperada de Fisher de  $\theta$  através de  $Y$  como

$$I(\theta) = E \left[ - \frac{\partial^2 \log p(\theta | y)}{\partial \theta \partial \theta'} \right]. \quad (5)$$

Segundo os autores, o conceito de informação aqui está sendo associado a uma espécie de curvatura média da função de verossimilhança no sentido de que, quanto maior a curvatura, mais precisa é a informação contida na verossimilhança, ou equivalentemente maior o valor de  $I(\theta)$ . Em geral espera-se que a curvatura seja negativa e por isso seu valor é tomado com sinal trocado. Note também que a esperança matemática é tomada em relação à distribuição amostral  $p(y|\theta)$ . Podemos considerar então  $I(\theta)$  uma medida de informação global, enquanto que uma medida de informação local é obtida quando não se toma o valor esperado na definição acima. A medida de informação observada de Fisher  $J(\theta)$  fica então definida como

$$J(\theta) = - \frac{\partial^2 \log p(\theta | y)}{\partial \theta^2}. \quad (6)$$

Definição 2: Seja uma observação  $Y$  com função de (densidade) de probabilidade  $p(y|\theta)$ . A priori não informativa de Jeffreys tem função de densidade dada por

$$p(\theta) \propto [I(\theta)]^{1/2}. \quad (7)$$

Se  $\theta$  for um vetor paramétrico então  $p(\theta) \propto |\det I(\theta)|^{1/2}$ .

Em geral a priori não informativa é obtida fazendo-se o parâmetro de escala da distribuição conjugada tender a zero e fixando-se os demais

parâmetros convenientemente (Gelman et al., 2003). Além disso, *a priori* de Jeffreys assume formas específicas em alguns modelos, sendo dada por  $p(\theta) \propto \text{constante}$ , para o modelo de locação e  $p(\theta) \propto \sigma^{-1}$ , para o modelo de escala. Em modelos de locação e escala, *a priori* não informativa pode ser obtida assumindo-se independência a priori entre  $\theta$  e  $\sigma$  de modo que

$$p(\theta, \sigma) = p(\theta) p(\sigma) \propto \sigma^{-1} \quad (8)$$

Segundo Paulino et al. (2003), *a priori* não informativa de Jeffreys viola o princípio da verossimilhança, já que a informação de Fisher depende da distribuição amostral.

#### 2.1.4 Distribuição *a posteriori*

A análise bayesiana é conduzida combinando uma informação *a priori*  $p(\theta)$  com uma informação amostral  $p(y|\theta)$ . Segundo Paulino et al. (2003), a distribuição resultante dessa combinação  $p(\theta|y)$  é denominada distribuição *a posteriori* de  $\theta$  dado  $y$  e é dada por

$$p(\theta|y) \propto p(\theta) p(y|\theta) . \quad (9)$$

Segundo Box & Tiao (1992), a distribuição *a posteriori* pode ser vista como uma regra de atualização de  $p(\theta|y)$  pelo fato de que a distribuição *a priori*  $p(\theta)$  contém pouca informação sobre o parâmetro. Porém, a informação relativa à situação experimental é observada na amostra. Assim, acredita-se que combinando essas “informações” haverá inferências mais precisas, principalmente se estas informações forem atualizadas a cada realização amostral e/ou experimental. Paulino et al. (2003) salientam que, nesta regra de

atualização, está implícito o princípio da verossimilhança, pois toda informação sobre o parâmetro está contida na distribuição amostral  $p(y|\theta)$ .

Uma vez que a distribuição *a posteriori* para os parâmetros está disponível, pode-se pensar em inferência. Em princípio, esta distribuição é a descrição mais completa possível de toda a informação disponível. Porém, para comparar e discutir resultados, bem como para resumir conclusões, é mais prático apresentar algumas quantidades que sintetizam esta distribuição (Gelman et al., 2003). Examinam-se, então, as alternativas bayesianas para os elementos centrais da inferência estatística clássica: estimação e intervalos de confiança.

### 2.1.5 Estimação

Segundo Box & Tiao (1992), o processo de estimação pode ser realizado de duas maneiras básicas, satisfazendo as propriedades desejáveis: estimação pontual e estimação por intervalo.

#### 2.1.5.1 Estimação pontual

Segundo os autores, é o caso mais simples, em que procura-se determinar um único valor de  $\theta$  que resuma a distribuição como um todo. Este valor é denotado por  $\hat{\theta}$  e é denominado estimador pontual de  $\theta$ . Como estimativa pontual de  $\theta$ , três medidas são comumente utilizadas, a saber:

- i) **estimador de máxima verossimilhança generalizado** - definido como a moda *a posteriori*;
- ii) **média da distribuição a posteriori** - esperança matemática de  $\theta$ , com respeito *a posteriori*;
- iii) **mediana da distribuição a posteriori**.

De acordo com Box & Tiao (1992) a estimação pontual é mais bem caracterizada num contexto de teoria da decisão, em que se define uma regra de decisão e a cada decisão associa-se uma perda, interpretada como a punição que se sofre por tomar a decisão quando o valor do parâmetro é  $\theta$ . O risco, ou perda esperada *a posteriori*, associado a uma determinada regra de decisão, é obtido tomando-se a esperança da função perda com relação à distribuição *a posteriori* de  $\theta$ . Uma regra de decisão é ótima se tem risco mínimo. O estimador pontual é a regra de decisão ótima segundo uma função perda especificada, de acordo com Gelman et al. (2003).

Box & Tiao (1992) afirmam que todo estimador deve ser acompanhado pelo respectivo erro padrão. Para isso, é preciso conhecer a distribuição amostral da estatística em questão, que nem sempre é possível. A utilização de determinada estatística em dado problema dependerá da teoria disponível (Gelman et al., 2003). Em muitos casos, o problema só pode ser contornado recorrendo-se a resultados assintóticos, mesmo quando o tamanho da amostra é pequeno, ou utilizando-se alternativas computacionalmente intensas: os métodos de simulação Monte Carlo e *bootstrap* (Ehlers, 2005).

Segundo Ehlers (2005), na análise bayesiana, em princípio, essa dificuldade não existe. Pois, para qualquer estimador  $\hat{\theta}$  de  $\theta$ , podemos calcular a variância *a posteriori* e, desta forma, ter uma idéia da precisão associada ao estimador.

### **2.1.5.2 Estimação por intervalo**

Sob o ponto de vista bayesiano, a forma mais adequada de avaliar a informação disponível a respeito de uma quantidade desconhecida  $\theta$  é por meio da distribuição *a posteriori*, uma vez que a estimação pontual não nos informa sobre a precisão da estimativa e restringe toda informação presente na



distribuição *a posteriori* num único resultado (Paulino et al., 2003). Segundo os autores, uma maneira de contornar esse problema é calcular intervalos de credibilidade para estes valores. Segundo eles, o ideal é que os intervalos de credibilidade sejam definidos com o menor tamanho e maior nível de credibilidade possível.

A distribuição *a posteriori* de um parâmetro  $\theta$  contém toda a informação probabilística a respeito deste parâmetro e um gráfico da sua função de densidade *a posteriori* é a melhor descrição do processo de inferência (Box & Tiao, 1992).

### 2.1.5.3 Intervalos de credibilidade e HPD's.

Uma alternativa para o intervalo de confiança clássico é o intervalo de credibilidade  $[\theta_1, \theta_2]$  de  $100(1 - \alpha)\%$ , tal que  $P(\theta \leq \theta_1) = P(\theta \geq \theta_2) = \alpha/2$ .

Outra possibilidade é selecionar o subconjunto  $C$  de valores para  $\theta$  com maiores probabilidades *a posteriori* (denominado por Gelman et al. (2003) como “*Highest posterior Density (HPD) Interval*”) tal que,

$$C = \{\theta \in \Theta : p(\theta|y) \geq k(\alpha)\}, \quad (10)$$

em que  $k(\alpha)$  é a maior constante que pode ser obtida, satisfazendo  $P(C|y) \geq 1 - \alpha$ .

Para distribuições unimodais e simétricas, o intervalo de credibilidade e o HPD são idênticos. Havendo assimetria, os dois critérios produzem intervalos diferentes; sendo esta diferença dependente do grau de assimetria. Se, no entanto, a distribuição *a posteriori* é multimodal, o conjunto  $C$  pode ser constituído de vários subintervalos em torno das modas mais destacadas

(Gelman et al., 2003). Neste caso o HPD leva vantagem sobre o intervalo de credibilidade, fornecendo mais informações.

## 2.2 Estudo do modelo normal

Segundo Ehlers (2005), um resultado importante ocorre quando se tem uma única observação da distribuição normal com média desconhecida. Se *a priori*, para a média for uma distribuição normal, os parâmetros *a posteriori* são obtidos de uma forma bastante intuitiva.

**Teorema 1:** Se  $Y|\theta \sim N(\theta, \sigma^2)$  com  $\sigma^2$  conhecido e  $\theta \sim N(\mu_0, \tau_0^{-2})$  então  $\theta|y \sim N(\mu_1, \tau_1^{-2})$  em que

$$\mu_1 = \frac{\tau_0^{-2} \mu_0 + \sigma^{-2} y}{\tau_0^{-2} + \sigma^{-2}} \quad e \quad \tau_1^{-2} = \tau_0^{-2} + \sigma^{-2}. \quad (11)$$

Note que, definindo *precisão* como o inverso da variância, segue do teorema que a *precisão a posteriori* é a soma das *precisões a priori* e da verossimilhança e não depende de  $y$ . Interpretando *precisão* como uma medida

de informação e definindo  $w = \frac{\tau_0^{-2}}{\tau_0^{-2} + \sigma^{-2}} \in (0, 1)$  então  $w$  mede a informação

relativa contida na *priori* com respeito à informação total (Ehlers, 2005).

Para uma amostra de tamanho  $n$ , a função de verossimilhança pode ser escrita como

$$l(\theta; y) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right\} \propto \exp\left\{-\frac{n}{2\sigma^2} (y - \theta)^2\right\}, \quad (12)$$

em que os termos que não dependem de  $\theta$  foram incorporados à constante de proporcionalidade. Segundo Ehlers (2005), a verossimilhança tem a mesma forma daquela baseada em uma única observação bastando substituir  $y$  por  $\bar{y}$  e  $\sigma^2$  por  $\sigma^2/n$ . Logo, vale o Teorema 1 com as devidas substituições, ou seja, a distribuição *a posteriori* de  $\theta$  dado  $y$  é  $N(\mu_1, \tau_1^{-2})$  onde

$$\mu_1 = \frac{\tau_0^{-2} \mu_0 + n\sigma^{-2} \bar{y}}{\tau_0^{-2} + n\sigma^{-2}} \quad e \quad \tau_1^{-2} = \tau_0^{-2} + n\sigma^{-2}. \quad (13)$$

A seguir serão apresentados alguns resultados para membros importantes da família exponencial.

### 2.2.1 Distribuição normal com média conhecida e variância desconhecida

Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória da distribuição  $N(\theta, \sigma^2)$ , com  $\theta$  conhecido e  $\phi = \sigma^2$  desconhecido. Segundo Ehlers (2005), neste caso a função de densidade conjunta é dada por

$$p(y | \theta, \phi) \propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} \sum_{i=1}^n (y_i - \theta)^2\right\}. \quad (14)$$

Note que o núcleo desta verossimilhança tem a forma daquele de uma distribuição Gama. Sabendo que a família Gama é fechada por amostragem, considera-se uma distribuição *a priori* Gama com parâmetros  $n_0/2$  e  $n_0 \sigma_0^2/2$ , isto é,

$$\phi \sim \text{Gama} \left( \frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2} \right). \quad (15)$$

Definindo  $n s_0^2 = \sum_{i=1}^n (y_i - \theta)^2$  e aplicando o Teorema de Bayes

obtém-se *a posteriori* de  $\phi$  (Ehlers, 2005), dada por

$$\begin{aligned} p(\phi | y) &\propto \alpha \phi^{n/2} \exp \left[ -\frac{\phi}{2} n s_0^2 \right] \phi^{n_0/2 - 1} \exp \left[ -\frac{\phi}{2} n_0 \sigma_0^2 \right] \propto \\ &\propto \phi^{(n_0 + n)/2 - 1} \exp \left[ -\frac{\phi}{2} (n_0 \sigma_0^2 + n s_0^2) \right]. \end{aligned} \quad (16)$$

Esta expressão corresponde ao núcleo de uma distribuição Gama, portanto:

$$\phi | y \sim \text{Gama} \left( \frac{n_0 + n}{2}, \frac{n_0 \sigma_0^2 + n s_0^2}{2} \right). \quad (17)$$

### 2.2.2 Distribuição normal com média e variância desconhecidas

Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória da distribuição  $N(\theta, \sigma^2)$ , com ambos  $\theta$  e  $\sigma^2$  desconhecidos. Neste caso a distribuição *a priori* conjugada será especificada em dois estágios segundo Ehlers (2005). No primeiro estágio,

$$\theta | \phi \sim N(\mu_0, (c_0 \phi)^{-1}), \quad \phi = \sigma^{-2} \quad (18)$$

e a distribuição *a priori* marginal de  $\phi$  é a mesma do caso anterior, expressão (15).

Segundo o autor, a distribuição conjunta de  $(\theta, \phi)$  é geralmente chamada de Normal-Gama com parâmetros  $(\mu_0, c_0, n_0, \sigma_0^2)$  e sua função de densidade conjunta é dada por:

$$\begin{aligned}
p(\theta, \phi) &= p(\theta | \phi) p(\phi) \\
&\propto \phi^{1/2} \exp\left[-\frac{c_0 \phi}{2} (\theta - \mu_0)^2\right] \phi^{n_0/2 - 1} \exp\left[-\frac{n_0 \sigma_0^2 \phi}{2}\right] \\
&\propto \phi^{(n_0 + 1)/2 - 1} \exp\left[-\frac{\phi}{2} (n_0 \sigma_0^2 + c_0 (\theta - \mu_0)^2)\right].
\end{aligned} \tag{19}$$

A partir da densidade conjunta pode-se obter a densidade marginal de  $\theta$  por integração (Ehlers, 2005),

$$\begin{aligned}
p(\theta) &= \int p(\theta | \phi) p(\phi) d\phi \\
&\propto \int_0^\infty \phi^{1/2} \exp\left[-\frac{c_0 \phi}{2} (\theta - \mu_0)^2\right] \phi^{n_0/2 - 1} \exp\left[-\frac{n_0 \sigma_0^2 \phi}{2}\right] d\phi \\
&\propto \int_0^\infty \phi^{(n_0 + 1)/2 - 1} \exp\left[-\frac{\phi}{2} (n_0 \sigma_0^2 + c_0 (\theta - \mu_0)^2)\right] d\phi \\
&\propto \left[\frac{n_0 \sigma_0^2 + c_0 (\theta - \mu_0)^2}{2}\right]^{-\frac{n_0 + 1}{2}} \propto \left[1 + \frac{(\theta - \mu_0)^2}{n_0 (\sigma_0^2 / c_0)}\right]^{-\frac{n_0 + 1}{2}},
\end{aligned} \tag{20}$$

que é o núcleo da distribuição  $t$  de Student com  $n_0$  graus de liberdade, parâmetro de locação  $\mu_0$  e parâmetro de escala  $\sigma_0^2/c_0$ . Ou seja,  $\theta \sim_{n_0} (\mu_0, \sigma_0^2/c_0)$ . A distribuição condicional de  $\phi$  dado  $\theta$  é facilmente obtida como:

$$p(\phi | \theta) \propto p(\theta | \phi) p(\phi) \propto \phi^{(n_0 + 1)/2 - 1} \exp\left[-\frac{\phi}{2} (n_0 \sigma_0^2 + c_0 (\theta - \mu_0)^2)\right], \tag{21}$$

e portanto,

$$\phi | \theta \sim \text{Gama} \left( \frac{n_0 + 1}{2}, \frac{n_0 \sigma_0^2 + c_0 (\theta - \mu_0)^2}{2} \right). \quad (22)$$

A *posteriori* conjunta de  $(\theta, \phi)$  é também obtida em 2 etapas como segue, segundo Ehlers (2005). Primeiro, para  $\phi$  fixo pode-se usar o resultado (13) de modo que a distribuição *a posteriori* de  $\theta$  dado  $\phi$  fica

$$\theta | \phi, y \sim N \left( \mu_1, (c_1 \phi)^{-1} \right) \quad (23)$$

em que

$$\mu_1 = \frac{c_0 \phi \mu_0 + n \phi \bar{y}}{c_0 \phi + n} = \frac{c_0 \mu_0 + n \bar{y}}{c_0 + n} \quad e \quad c_1 = c_0 + n. \quad (24)$$

Na segunda etapa, combinando a verossimilhança com *a priori* de  $\phi$  obtém-se que

$$\phi | y \sim \text{Gama} \left( \frac{n_1}{2}, \frac{n_1 \sigma_1^2}{2} \right), \quad (25)$$

em que

$$n_1 = n_0 + n \quad e \quad n_1 \sigma_1^2 = n_0 \sigma_0^2 + \sum (y_i - \bar{y})^2 + c_0 n (\mu_0 - \bar{y})^2 / (c_0 + n). \quad (26)$$

Assim *a posteriori* conjunta é  $(\theta, \phi | y) \sim \text{Normal-Gama} (\mu_1, c_1, n_1, \sigma_1^2)$  e portanto *a posteriori* conjunta de  $\theta$  fica

$$\theta | y \sim t_{n_1} \left( \mu_1, \sigma_1^2 / c_1 \right) \quad (27)$$

Os resultados anteriores são aplicados ao modelo normal para estimação da média e variância em problemas de uma ou mais amostras e em modelos de regressão linear.

### 2.3 Distribuição t multivariada

Segundo Giri (2004), a distribuição elíptica simétrica é uma família de distribuições que engloba a normal multivariada e tem sido considerada em vários estudos para avaliar o efeito da sensibilidade de um teste ou de um processo de estimação por região a desvios de normalidades dos dados multivariados. Suas funções de densidades possuem contornos de iguais probabilidades que têm forma elíptica.

Giri (2004) afirma que esta família tornou-se bastante popular devido ao fato de que alguns resultados ótimos de inferência estatística, obtidos para a distribuição normal multivariada, conservarem suas propriedades para todos os membros da família. Um caso particular desta distribuição são as distribuições esféricas.

Segundo o autor, as distribuições elípticas simétricas são definidas para um vetor aleatório  $Y = [Y_1, Y_2, \dots, Y_k]^T \in \mathfrak{R}^k$  com média  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_k]^T \in \mathfrak{R}^k$  e matriz de parâmetros simétrica e positiva definida  $\boldsymbol{\Sigma}_{(k \times k)}$  como funções da forma quadrática  $(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$  por:

$$f_Y(y) = |\boldsymbol{\Sigma}|^{-1/2} g \left[ (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right], \quad (28)$$

em que  $g$  é uma função  $[0, \infty)$  que satisfaz a condição  $\int_{\mathfrak{R}^k} g(z^T z) dz = 1$  para todo  $z \in \mathfrak{R}^k$ .

**Teorema 1:** (Transformações lineares nas distribuições elípticas). Se  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_k]^T \in \mathfrak{R}^k$  é um vetor aleatório com distribuição elíptica simétrica com parâmetros dados pelo vetor de médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_k]^T \in \mathfrak{R}^p$  e matriz simétrica e positiva  $\boldsymbol{\Sigma}_{(k \times k)}$ , então  $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2} (\mathbf{Y} - \boldsymbol{\mu})$  possui distribuição esférica simétrica, em que  $\boldsymbol{\Sigma}^{-1/2}$  é a inversa da matriz raiz quadrada de  $\boldsymbol{\Sigma}$ , tal que  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{1/2}$ .

Giri (2004) conclui que  $\mathbf{Z}$  possui distribuição elíptica simétrica com parâmetros  $\mathbf{0}$  e  $\mathbf{I}$ . Uma vez que as variâncias dos componentes de  $\boldsymbol{\Sigma}$  são iguais a 1 e as covariâncias entre eles são nulas, os contornos da função de densidade de mesma probabilidade são esféricos e, portanto, a distribuição é chamada de distribuição esférica simétrica, sendo um caso particular da família elíptica simétrica.

Um exemplo a destacar é o da distribuição normal multivariada.

Um segundo e importante exemplo é a  $t$  de Student multivariada com  $\nu$  graus de liberdade (Giri, 2004). A função de densidade do vetor aleatório  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_k]^T \in \mathfrak{R}^k$  com parâmetros dados pelo vetor de médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_k]^T \in \mathfrak{R}^k$  e matriz simétrica e positiva definida  $\boldsymbol{\Sigma}_{(k \times k)}$  é

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{(\pi\nu)^{k/2} \Gamma(\nu) |\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{1}{\nu} (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right]^{-\frac{\nu+k}{2}}, \quad (29)$$

em que  $g$  é dada por  $g(z) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{(\pi\nu)^{k/2} \Gamma(\nu)} \left[1 + \frac{z}{\nu}\right]^{-\frac{\nu+k}{2}}$ . A variável aleatória

$\mathbf{Y}$  tem média  $\boldsymbol{\mu}$  e matriz de covariâncias  $\nu\boldsymbol{\Sigma}/(\nu-2)$  no caso de  $\nu > 2$ .



Efetuada-se a transformação  $X = \Sigma^{-1/2} (Y - \mu)$  obtém-se a distribuição t multivariada esférica simétrica, cuja densidade é dada por (Giri, 2004):

$$f_X(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu + k}{2}\right)}{(\pi\nu)^{k/2} \Gamma(\nu)} \left[ 1 + \frac{1}{\nu} \mathbf{x}^T \mathbf{x} \right]^{-\frac{\nu + k}{2}} \quad (30)$$

A variável aleatória  $X$  terá vetor de médias nulo, matriz de covariâncias  $\nu I/(\nu - 2)$  ( $\nu > 2$ ) e a densidade terá contornos esféricos de mesma probabilidade.

## 2.4 Teoria bayesiana e freqüentista na tomada de decisão

Segundo Lindsey (1996), após a observação dos dados, a função de verossimilhança pode assumir uma determinada forma, tal que *a priori* pode ter pouco efeito sob o parâmetro pertencente a algum espaço paramétrico. Desta forma, o autor ressalta que a distribuição *a priori* só deveria ser considerada após examinar a função de probabilidade.

Berger & Berliner (1986) ressaltaram que, no caso de a distribuição *a posteriori* concentrar-se especificamente sobre a função de verossimilhança, poderão ocorrer estimativas imprecisas, além de as abordagens freqüentista ou bayesiana serem praticamente análogas em relação à hipótese avaliada sob a amostra.

Segundo Paulino et al. (2003), do ponto de vista freqüentista, a tomada de decisão tem como base a média de amostras repetidas, dado que o verdadeiro valor do parâmetro é fixo e desconhecido. Do ponto de vista bayesiano, essa repetição é feita sempre com a mesma amostra, mas o parâmetro é aleatório e geralmente desconhecido.

Recentemente, Agresti & Min (2005) utilizaram argumentos comuns à teoria freqüentista, tais como probabilidades de cobertura e região de confiança, para avaliar a performance da inferência bayesiana realizada em tabelas de contingências. Eles analisaram a probabilidade de cobertura de intervalos de confiança para o risco relativo, diferença entre duas proporções e razão de chances de tabelas de contingência 2 x 2 nas abordagens freqüentista e bayesiana, sendo que nesta última, os autores utilizaram as *prioris* dadas pelas distribuições beta, logit-normal e *prioris* correlacionadas para os dois parâmetros da distribuição binomial. Analisando as possíveis configurações que essas *priori* poderiam assumir através dos valores dos hiperparâmetros, os autores recomendaram os intervalos HPD para o risco relativo e para a razão de chances. Todavia, em situações em que as amostras eram pequenas, verificaram que os intervalos de credibilidade, usando uma *priori* informativa apresentaram uma performance inferior em relação aos intervalos de confiança obtidos na abordagem freqüentista.

## **2.5 Análise de variância**

Um dos métodos de análise de dados proveniente de experimentos é o da análise de variância, atribuído a Fisher (Steel & Torrie, 1980). De acordo com Nogueira (1997), a análise de variância é uma técnica que consiste em decompor a variância total apresentada pelos dados em suas causas de variação associadas ao delineamento experimental (aos fatores em estudo). Por variação total entende-se como a variação entre as unidades experimentais. Para se utilizar corretamente este processo é fundamental conhecer, além da variação inerente às unidades experimentais, quais são as demais fontes de variação previstas.

Segundo Nogueira (1997), as fontes de variação presentes em uma análise da variância deverão ser os efeitos principais, os efeitos de interação, os

efeitos de aninhamento e o resíduo, que estimará a variabilidade inerente às unidades experimentais.

De acordo com a autora, o efeito principal corresponde ao efeito de um fator tomado como uma média dos demais fatores, incluindo as repetições, presentes no experimento. O efeito inerente às interações mede o diferencial de um fator quando o outro muda de nível – quando envolve mais de dois fatores, o seu significado é mais complexo. O efeito aninhado também é um efeito médio em relação aos demais fatores dentro dos quais ele não está aninhado. O resíduo é a fonte de variação que representa toda a variação não provocada pelo pesquisador.

Todo delineamento experimental possui um modelo estatístico que deve ser considerado para se efetuar a análise de variância, assim como a aceitação de algumas hipóteses de validação da mesma, a saber (Banzatto & Kronka, 1989):

i) aditividade: os efeitos dos fatores que ocorrem no modelo matemático devem ser aditivos;

ii) independência: os erros ou desvios, devido ao efeito de fatores não controlados, devem ser independentes, ou seja, não correlacionados;

iii) homocedasticidade ou homogeneidade de variâncias: os erros ou desvios, devidos ao efeito de fatores não controlados, devem possuir variância comum;

iv) normalidade: os erros ou desvios, devidos ao efeito de fatores não controlados, devem possuir uma distribuição normal.

Segundo Steel & Torrie (1980), caso seja constatada a não verificação de uma ou mais dessas hipóteses, pode-se transformar os dados antes de se proceder à análise de variância ou utilizar algum outro tipo de método de análise mais robusto, como testes não-paramétricos, testes de aleatorização ou *bootstrap*, além de modelos lineares generalizados.

## 2.6 Procedimentos de comparações múltiplas (PCM)

### 2.6.1 Comparações entre médias

Um dos objetivos mais comuns dos experimentos é a comparação dos níveis dos tratamentos. Assim, admite-se a hipótese de nulidade ( $H_0$ ) como verdadeira, que supõe que não existem diferenças entre os níveis de um fator qualquer, de efeitos fixos, com  $k$  de níveis do fator (Machado et al., 2005):

$$H_0: \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k. \quad (31)$$

O teste F, aplicado à análise de variância, é utilizado para testar a hipótese de nulidade dos efeitos de tratamentos, ou seja, para verificar se existem diferenças reais entre os níveis do fator. Se a decisão for para a não rejeição da hipótese apresentada em (31), constata-se a não existência do efeito do fator em estudo sobre a variável observada. Caso se rejeite essa hipótese, aceita-se a hipótese alternativa, na qual pelo menos um par de médias de tratamentos ou níveis do fator difere entre si.

O passo seguinte é a investigação sobre a que se devem as diferenças, quais são as médias que diferem entre si e 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 (31) é que é testada. Isso é feito para retratar situações reais que os pesquisadores julgam comuns (Borges & Ferreira, 2003; Silva et al., 1999).

Assim, com a rejeição desta hipótese global de interesse, deve-se continuar a análise estatística dos dados observados, buscando procedimentos para identificar quais são os níveis deste fator que diferem entre si. Assim, de

acordo com o tipo de níveis atribuídos para o fator, a técnica apropriada a ser utilizada será:

i) se os níveis deste fator em estudo são **quantitativos**, o uso de uma metodologia de análise de regressão é mais apropriado.

ii) se os níveis deste fator em estudo são **qualitativos com uma estruturação** que sugerem comparações específicas – pré-planejadas - entre os níveis do fator, utilizam-se contrastes de níveis do fator (tratamentos), contrastes ortogonais, partição da soma de quadrados de níveis do fator (Hinkelmann & Kempthorne, 1987).

iii) quando os níveis do fator são **qualitativos e não estruturados** deve-se aplicar os procedimentos de comparações múltiplas (PCM).

Os PCM são procedimentos estatísticos que comparam duas ou mais médias e possuem uma vasta bibliografia sobre o assunto (Hochberg & Tamhane, 1987; Machado et al., 2005; Ramalho et al, 2000; entre outros). Quando os PCM são utilizados, comparações emparelhadas possíveis serão realizadas. É preciso ter alguns cuidados para se utilizar procedimentos de inferência corretos quando se realiza testes de hipóteses ou estimação intervalar. Todos os testes desse tipo permitem analisar diferenças entre médias após a conclusão do experimento para detectar possíveis grupos para os níveis do fator não-estruturado.

Machado et al. (2005) ressaltam alguns problemas envolvidos na utilização dos procedimentos de comparação de médias duas a duas:

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;

3. dificuldade de interpretação: devido à ambigüidade.

Conforme Machado et al.(2005), a escolha do método adequado de comparações múltiplas que se deve adotar depende das qualidades estatísticas deste procedimento. Esta qualidade é função do tipo de erro que é controlado e da forma como estes erros são controlados. Pode-se, ainda, procurar métodos que sejam robustos a algumas violações básicas dos pressupostos necessários para a aplicação de um procedimento de comparação múltipla, tais como normalidade e homogeneidade de variâncias.

### 2.6.2 Erros envolvidos e poder

Quando se realiza um teste de hipótese, o pesquisador corre um risco global de tomar decisões erradas que deverá ser considerado. Ele está sujeito a incorrer em erros de decisão (Banzatto & Kronka, 1989; Gómez & Gomez, 1984; Machado et al., 2005; Mood et al., 1974):

i) se uma hipótese for rejeitada quando deveria ser aceita, diz-se que foi cometido erro Tipo I. A probabilidade de se cometer esse erro é:  
 $P[\text{Erro Tipo I}] = P[\text{rejeitar } H_0 | H_0 \text{ verdadeira}] = \alpha$

ii) se for aceita uma hipótese que deveria ser rejeitada, diz-se que foi cometido erro Tipo II. A probabilidade de se cometer esse erro é:  
 $P[\text{Erro Tipo II}] = P[\text{não rejeitar } H_0 | H_0 \text{ falsa}] = \beta$

iii) se declarar uma média maior do que a outra quando, na verdade, ocorre o contrário (Carmer & Swanson, 1973), diz-se que cometeu o erro Tipo III. A probabilidade de se incorrer nesse tipo de erro é:  
 $P[\text{Erro Tipo III}] = P[\text{rejeitar } H_0, \text{ mas a favor de } H_1 \text{ errada} | H_0 \text{ falsa}]$ .

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*).

Meyer (1984) comenta que não é possível evitar completamente esses erros, mas que deve-se manter relativamente pequena a probabilidade de cometê-los, ou seja, os testes de hipóteses devem ser planejados de modo que os erros de decisão sejam reduzidos ao mínimo. Isto não é fácil, pois, para um dado tamanho de amostra, baixas probabilidades de se cometer certo tipo de erro são associadas a altas probabilidades de se cometer outro tipo. Na prática, um tipo de erro pode ser mais importante que o outro, de modo que se deve procurar uma acomodação que favoreça a limitação do erro mais sério.

A preocupação maior ao se desenvolverem procedimentos de comparação de 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, mas, ao mesmo tempo, liberal (Machado et al., 2005).

Tonhasca Jr. (1991), discutindo aspectos normalmente negligenciados no uso de metodologias estatísticas, diz que, estabelecida a hipótese alternativa, o maior problema é calcular o poder do teste. Acrescenta-se que o poder de um teste depende do valor de  $\alpha$ , da hipótese alternativa e do tamanho da amostra.

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 executar analiticamente. 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).

Em seu trabalho, Leonel (2000) avaliou um procedimento alternativo para comparação de tratamentos, nas análises em grupos de experimentos. Este procedimento baseou-se na combinação das probabilidades dos contrastes de interesse de cada experimento. Os resultados obtidos por este teste foram comparados com os resultados da análise de variância tradicional. Por meio do processo de simulação Monte Carlo, foi possível avaliar as suas taxas de erro Tipo I e o seu poder. Nas situações avaliadas, observou-se um controle da taxa de erro Tipo I, com os valores nominais de 5% e 1%. Quanto ao poder do teste, verificou-se que apresenta baixo poder quando comparado com a análise de variância tradicional.

Santos (2000) propõe, em seu trabalho, uma modificação para o processo divisório do teste de Scott-Knott e novas alternativas de agrupamento. O comportamento do teste quanto às taxas de erro tipo I e o poder do teste, sob situações de hipótese de nulidade completa e parcial, foi avaliado por meio de simulação Monte Carlo. Entre outras conclusões, o autor verifica que o teste de Scott-Knott original controla adequadamente as taxas de erro tipo I, por comparação e por experimento, sob hipótese de nulidade completa, mantendo o



poder elevado, mas não controla adequadamente sob hipótese de nulidade parcial. O teste de Scott-Knott modificado apresenta os mesmos resultados do original sob hipótese de nulidade completa, mas, sob hipótese de nulidade parcial, apresenta taxas de erro tipo I mais elevadas, embora tenha maior poder.

A *false discovery rate* (FDR) (Benjamin & Hochberg, 1995) é um critério alternativo de proteção do erro. A FDR é a proporção esperada de hipóteses verdadeiras entre todas aquelas que são rejeitadas, ou seja, o valor esperado de  $Q/R$ , em que  $Q$  é o número de hipóteses rejeitadas falsamente (hipóteses verdadeiras que são rejeitadas), e  $R$  é o número total de rejeições. O controle de FDR implica que certo número de erros é permissível com alta probabilidade, se eles representam uma proporção suficientemente pequena de fortes conclusões atingidas.

### **2.6.3 Testes de comparações múltiplas**

#### **2.6.3.1 Introdução**

Em muitas situações experimentais é desejável fazer um grande número de comparações, tais como todas as possíveis combinações entre duas delas ou comparações sugeridas pelos próprios dados. Deve-se tomar cuidado para o uso de procedimentos de inferência corretos, caso contrário, pode ocorrer o efeito da multiplicidade (Tukey, 1977), levando à grande quantidade de resultados significativos. Segundo Hinkelmann & Kempthorne (1987), o problema envolve as noções de taxas de erro por comparação (*comparisonwise error rate* - CWE) e taxa de erro por experimento (*familywise error rate* - FWE). A idéia é controlar a taxa de erro por experimento e não para cada comparação individual. A relação entre CWE e FWE pode ser expressa como

$$1 - FWE = (1 - CWE)^N \quad (32)$$

em que  $N$  é o número de combinações a serem feitas.

Embora todos PCM testem a mesma hipótese de nulidade, eles usam diferentes métodos para controlar a FWE e também diferem quanto à sensibilidade da hipótese alternativa (Hinkelmann & Kempthorne, 1987).

Os PCM são utilizados para testar contrastes envolvendo duas médias e todas as combinações entre os níveis do fator em questão. 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 na 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:

$$DMS = \gamma Sd \quad (33)$$

em que:  $Sd = \sqrt{2 QME / r}$  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  $\nu$  graus de liberdade;  $r$  é o número de repetições e  $\gamma$  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 o valor absoluto da diferença entre duas médias é maior que a DMS, as médias são consideradas estatisticamente diferentes, ao nível de significância estabelecido. 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).

Segundo Machado et al. (2005), para experimentos que envolvem  $n$  níveis do fator que sejam igualmente repetidos ( $r$ ) e  $\nu$  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.

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.

### 2.6.3.2 Teste $t$ de Student

Este teste é conhecido como teste da diferença mínima significativa (LSD - *Least Square Difference*) cujo valor crítico é dado por:

$$LSD = t_{(\alpha/2, \nu)} Sd \quad (34)$$

em que  $t_{(\alpha/2, \nu)}$  é o quantil 100 $\alpha$ % superior da distribuição  $t$  de Student com  $\nu$  graus de liberdade e  $Sd$  é o estimador do erro padrão da diferença de duas médias.

O teste  $t$  de Student controla o erro tipo I no nível nominal  $\alpha$  em um contraste pareado testado individualmente, porém, não controla esse tipo de erro para todos os testes pareados possíveis, neste mesmo nível nominal  $\alpha$ . 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  $\alpha$  (Machado et al., 2005).

Para se preservar as taxas de erro por experimento, alguns autores, entre eles Neter & Wasserman (1974) e Steel & Torrie (1980), recomendam que se utilize o  $t$  protegido de Fisher (1935), que requer que o teste  $F$  seja efetuado antes da realização do mesmo. Se o valor observado de  $F$  for significativo em um nível nominal de significância  $\alpha$ , o teste LSD é aplicado. No entanto, se o valor de  $F$  é não significativo no mesmo nível nominal, nenhuma comparação de médias é feita, eliminando a possibilidade de se cometer o erro tipo I nas comparações múltiplas. O seu valor crítico é calculado por:

$$FSD = LSD = t_{(\alpha/2, v)} Sd \quad (35)$$

se o valor de  $F$  for significativo, ou

$$FSD = \infty \quad (36)$$

se o valor de  $F$  for não significativo.

Outra alternativa ao teste  $t$ , segundo O'Neill & Wetherill (1971), é a utilização da diferença significativa de Fisher, também denominado de correção de Bonferroni da significância nominal. Neste tipo de teste o valor da significância nominal  $\alpha$  é ajustado para o número de comparações realizadas  $k$ . Faz-se simplesmente dividindo a significância nominal pelo número de comparações pareadas que se quer fazer para determinar o quantil da distribuição da amplitude estudentizada. Assim, o valor crítico do teste  $t$  protegido por Bonferroni é dado por:

$$BLSD = t_{(\alpha/2k, v)} Sd \quad (37)$$

em que  $k$  é o número de comparações realizadas.

Segundo Machado et al. (2005), o teste LSD protege apenas contra o erro tipo I por comparação e o teste LSD protegido de Fisher (FSD) controla o erro tipo I por comparação em todos os casos e o erro tipo I por experimento sob  $H_0$  completa, mas não controla este último erro sob  $H_0$  parcial. O teste BLSD, por sua vez, controla o erro tipo I por comparação e por experimento, com limites máximo das taxas iguais ao valor nominal.

### 2.6.3.3 Teste de Tukey

Outro teste dependente da amplitude estudentizada é o de Tukey (Tukey, 1977). Este teste foi delineado para controlar o erro por experimento e o valor crítico do teste (TSD) é dado por:

$$TSD = \frac{q(\alpha, n, v) Sd}{\sqrt{2}} \quad (38)$$

Se a diferença entre um par de médias é maior do que TSD, então a hipótese de nulidade, ou seja, a hipótese de que as médias são iguais, é rejeitada. Este teste foi delineado para controlar o erro tipo I por experimento, sendo que controla muito bem este tipo de erro, porém, possui elevadas taxas de erro tipo II. Assim, o teste Tukey é mais conservador do que o LSD em declarar uma diferença como significativa e tem um baixo poder.

#### 2.6.3.4 Teste SNK

Keuls (1952) sugere um outro método, usando a amplitude estudentizada. Neste método, as médias devem ser ordenadas da maior para a menor. Diferentes  $q$  valores (quantis da amplitude estudentizada) são usados, dependendo se as médias contrastadas são adjacentes ou se separadas por um ou mais intervalos.

O teste de Student-Newman-Keuls (SNK) exige o cálculo de  $k-1$  valores críticos, dados por:

$$SNK_p = \frac{q(\alpha, p, v) Sd}{\sqrt{2}} \quad (39)$$

para  $p = 2, 3, \dots, k$ , que é o número de médias abrangidas pelos contrastes entre duas médias.

Segundo Machado et al. (2005),  $SNK_2$  é igual ao valor crítico LSD e  $SNK_k$  é igual ao valor crítico TSD. Para valores intermediários de  $k$ , o valor crítico  $SNK_p$  é um valor intermediário, entre os valores críticos LSD e TSD.

É uma alternativa para controlar a taxa de erro por comparação, mas não controla a taxa de erro por experimento sob  $H_0$  parcial, embora a controle sob  $H_0$  completa. Segundo Ramalho et al. (2000), estudos de simulação mostram que este é um algoritmo eficiente no controle da taxa de erro tipo I por experimento em relação aos níveis nominais adotados.

#### 2.6.3.5 Teste de Duncan

Amplitudes estudentizadas especiais, em que, para cada contraste o nível de significância  $\alpha$  é alterado em função do número de médias abrangidas, são usadas no teste de Duncan (Duncan, 1955). É um procedimento de comparação

múltipla similar ao SNK, apresentando uma ligeira diferença na estatística  $q$ , cujos valores críticos são calculados por:

$$MRT_p = \frac{q(\alpha_p, p, v) Sd}{\sqrt{2}} \quad (40)$$

com  $\alpha_p = [1 - (1 - \alpha)^{p-1}]$  para  $p = 2, 3, \dots, k$ .

Assim, o teste de Duncan é mais rigoroso do que o teste LSD, porém mais liberal do que os testes de Tukey e SNK. Pode ser considerado um compromisso de se preocupar em controlar as taxas de erro tipo I e tipo II, apresentando uma baixa taxa de erro tipo II e, então, um poder maior que os anteriores (Davis & Gaito, 1984).

#### 2.6.3.6 Teste de Scheffé

O método proposto por Scheffé (1953) é apresentado na forma de intervalos de confiança simultâneos para os contrastes e é baseado na distribuição  $F$ . Segundo Machado et al. (2005), este método é bastante geral e aplicável a qualquer contraste linear das médias de tratamento, podendo ser usado como um procedimento de comparações múltiplas. Quando aplicado neste contexto seu valor crítico simplificado é dado por:

$$SSD = [mF(\alpha, m, v)]^{1/2} Sd \quad (41)$$

em que  $F(\alpha, m, v)$  é o quantil superior  $100\alpha\%$  da distribuição  $F$  com  $m = k - 1$  e  $v$  graus de liberdade.

Ele é muito rigoroso para, por exemplo, comparações 2 a 2 entre todos os tratamentos. É extremamente conservador, estabelecendo que a taxa de erro

por experimento não exceda o nível  $\alpha$  para todas as comparações feitas, se tornando tanto mais conservador quanto maior o número de comparações a se fazer.

### **2.6.3.7 Testes de agrupamento**

Todos os procedimentos apresentados até então consideram válida a pressuposição de normalidade dos resíduos e há fortes indicações de que eles não sejam muito robustos em situações de desvios de normalidade (Ringland, 1983). Hochberg & Tamhane (1987) apresentam procedimentos não paramétricos, livres de pressuposições de normalidade, que podem ser empregados.

Freqüentemente surgem resultados pouco claros nas apresentações resumidas dos testes como os anteriormente apresentados. Uma das mais famosas complicações práticas é a da ambigüidade, que surge da sobreposição de parte dos intervalos de confiança. Para contornar esse problema, utilizam-se idéias de análise de agrupamento combinadas aos testes de hipóteses (Hinkelmann & Kempthorne, 1987).

Segundo os mesmos autores, um dos objetivos dos PCM, além de comparar níveis de tratamento, é chegar a grupos de tratamentos “homogêneos”. Isto facilitaria a interpretação dos resultados e ajudaria a fazer recomendações acerca de ações futuras.

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 no artigo de O'Neill & Wetherill (1971). Métodos deste tipo foram sugeridos por vários autores (Scott & Knott, 1974; Calinski & Corsten, 1985).

Hochberg & Tamhane (1987) apresentam uma série de procedimentos de comparações múltiplas, baseados em diversas técnicas, além da análise de agrupamento.

#### **2.6.3.8 Testes de comparação múltipla via *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. Ramos (2007) propôs uma alternativa *bootstrap* para um dos procedimentos de comparações múltiplas de Calinski & Corsten (1985), baseada em análise de agrupamento para avaliar a versão original e *bootstrap* desse procedimento por meio de simulação Monte Carlo. O desempenho do teste *bootstrap* é, em geral, considerado superior ao do teste original e robusto, sendo, portanto, recomendada a sua utilização rotineiramente.

#### **2.6.3.9 Algumas Considerações**

Segundo Machado et al (2005), a escolha do teste adequado depende do seu controle do erro tipo I e do poder. O teste de Tukey controla adequadamente o erro tipo I por experimento e se torna rigoroso com o controle do erro tipo I por comparação à medida que o número de níveis do fator aumenta. Por controle adequado devemos entender que o valor nominal da significância é o tamanho real do teste. Isto acontece para este teste sob a hipótese nula completa ou parcial. O teste de Scheffé também apresenta comportamento similar ao de

Tukey, mas é mais rigoroso ainda que o teste de Tukey. Um teste que controla a taxa de erro tipo I por experimento sob hipóteses nulas completas, mas se torna razoavelmente liberal quando a hipótese nula é parcial, é o teste SNK. Muitas vezes este fato é um atrativo para a utilização deste procedimento, uma vez que o seu poder é superior aos dos testes de Tukey e de Scheffé.

Machado et al. (2005) afirmam que as melhores alternativas são os testes de computação intensiva de Tukey (comparações pareadas) e Dunnett (comparações com o controle). Isto porque, em situações adversas para a distribuição do erro, estes dois procedimentos controlaram o erro tipo I por experimento, com taxas inferiores ao valor nominal  $\alpha$  adotado ou quando superaram o desvio foi de no máximo alguns poucos pontos percentuais, enquanto os testes originais correspondentes apresentavam taxas superiores a 50%.

De acordo com Machado et al. (2005), o avanço computacional na área de *software* e na área de *hardware* é cada vez maior, possibilitando que testes computacionais mais elaborados e complexos sejam avaliados e propostos, buscando a solução dos problemas que ainda persistem desde a formulação das primeiras teorias sobre o assunto. Mas a opinião de Hocking (1985) ainda persiste, ou seja, ao mesmo tempo em que requer proteger de decisões incorretas (falsas significâncias), deseja-se também detectar diferenças reais entre os tratamentos.

Outros testes de comparações múltiplas foram publicados na literatura especializada como, por exemplo, os testes de Duncan (1965) e de Waller-Duncan (Waller & Duncan, 1969). Estes dois testes envolvem metodologias bayesianas nas suas formulações. Estes métodos são baseam-se no risco de Bayes, ao invés de terem relação com algum tipo particular de taxa de erro.

#### 2.6.4 Testes de comparações múltiplas bayesianos

Uma quantidade razoável de artigos, a maioria recentes, aparece na literatura na qual a aproximação bayesiana tem sido utilizada em alguns problemas que envolvem comparações múltiplas.

Duncan (1965) dá a primeira exposição completa de uma aproximação bayesiana para o problema de comparação múltipla, em que a decisão é feita usando magnitudes relativas de perdas devidas a erros do tipo I e tipo II. O procedimento envolve a especificação de uma distribuição *a priori* conjunta para os parâmetros. Ele assume o modelo usual de efeitos aleatórios para as médias  $\theta_i$  e uma relação conhecida para a razão entre as variâncias entre e dentro  $k = \sigma_{\theta}^2 / \sigma^2$ , obtendo distribuições *a posteriori* para as médias. Mas este é somente o primeiro passo. O segundo passo, parte importante das realizações de Duncan (1965), é a incorporação da teoria da decisão, modelando o fato de duas ou mais médias serem iguais em termos de funções de perda. Assumindo a mesma função de perda entre as comparações par a par, especificou-se somente uma constante  $k$ , o que indica uma importância relativa do erro tipo I em relação ao tipo II para cada comparação. Segundo Berry & Hochberg (1999), o procedimento é adaptativo no que depende da (aparente) heterogeneidade entre as médias, como indicado pelo teste F.

Waller & Duncan (1969) ampliaram o procedimento original de Duncan (1965) usando uma distribuição *hiperpriori* para a desconhecida razão entre as variâncias entre e dentro  $\sigma_{\theta}^2 / \sigma^2$ . Eles forneceram uma conexão entre possíveis valores de  $k$  e o nível de significância implicado na comparação.

Berry (1988) atribuiu uma distribuição Dirichlet *a priori* para a população de onde foram retiradas as médias amostrais. Gopalan & Berry (1998) desenvolveram esta aproximação, mostrando que a aproximação bayesiana difere da freqüentista, mas há maneiras em que as duas podem levar a resultados

similares. Eles aplicaram *prioris* de Dirichlet na forma de combinações básicas *priori*/verossimilhança para obter probabilidades *a posteriori* para várias hipóteses de igualdade entre médias populacionais. As combinações usadas foram beta/binomial e normal/gama inversa, com variâncias iguais. Os autores encontraram distribuições *a posteriori* que requerem métodos numéricos ou procedimentos de simulação para serem conhecidas, usando assim o amostrador de Gibbs. O procedimento proposto mostrou ser mais poderoso do que o de Duncan sob certas hipóteses alternativas.

Westfall et al. (1997) apresentaram uma perspectiva bayesiana para PCM no contexto da análise de variância. Eles afirmam que a FWE depende da credibilidade que os bayesianos atribuem às hipóteses de nulidade que serão testadas.

Shaffer (1999) modificou o procedimento de Waller/Duncan, ajustando  $k$  para encontrar FWE de 0,05 para hipótese de nulidade global, ou seja, a probabilidade de que todas as hipóteses sejam aceitas é 0,95 quando todas são verdadeiras. Assim, evitou a necessidade de especificar uma *priori* para uma razão de variâncias geralmente desconhecida, postulando *prioris* não-informativas e usando os dados para estimar a razão. O objetivo do procedimento de Shaffer (1999) foi comparar o procedimento de comparações múltiplas bayesiano de Duncan (Duncan, 1965; Waller & Duncan, 1969) com procedimentos não bayesianos comumente usados. Os procedimentos foram comparados quanto a FWE, FDR, poder e risco de Bayes.

Shaffer (1999) afirma que os resultados indicaram considerável similaridade, tanto no risco de Bayes quanto no poder entre o procedimento de Duncan modificado e o controle de FDR de Benjamin & Hochberg (1995), com o mesmo controle fraco da taxa de erro por experimento. Risco e poder desses procedimentos foram próximos ao risco e poder dos testes  $t$  individuais de comparações de médias. Essas conclusões são válidas considerando

pressuposição de normalidade<sup>4</sup> das médias dos tratamentos, graus de liberdade para o erro grande e variâncias iguais para as médias amostrais dos tratamentos.

Berry & Hochberg (1999) afirmaram que comparações múltiplas e problemas a elas relacionadas, tais como análise de agrupamentos (Berry, 1990), estavam entre os mais difíceis encontrados por estatísticos e outros pesquisadores. Segundo os autores, a perspectiva bayesiana é similar à freqüentista, que faz inferências com base em comparações, preservando a taxa de erro por experimento. Eles descreveram as diferenças entre assumir distribuições *a priori* independentes e hierárquicas e usaram o processo Dirichlet como distribuição *a priori* no contexto de multiplicidades. Discutiram também alguns procedimentos quase bayesianos, que combinam idéias freqüentistas e bayesianas, mostrando o potencial da metodologia bayesiana na produção de procedimentos que podem ser avaliados usando critérios objetivos. Comentaram o papel da subjetividade em aproximações bayesianas para problemas de comparações múltiplas e complexos na robustez versus *prioris* informativas.

Gelman & Tuerlinckx (2000) focaram seus estudos na taxa de erro tipo III. Eles definiram um procedimento bayesiano para comparações múltiplas como um conjunto de “afirmações com confiança” da forma  $\theta_j > \theta_k$ , tal que a probabilidade *posteriori* é 0,95 de que todas essas afirmações são verdadeiras. O procedimento é um teste de multiestágio: retira-se do vetor de parâmetros os valores de  $\theta$ , consideram-se as afirmações da forma “ $\theta_j > \theta_k$ ” uma de cada vez, começando com a comparação que tem a maior probabilidade *a posteriori* (baseada em simulações), continuando em ordem de probabilidade decrescente. Se nenhuma das comparações tem ao menos 0,95 de probabilidade de ter o correto sinal, nada se pode afirmar com confiança. Por outro lado, eles

---

<sup>4</sup> Para discussão da elucidação de *prioris* subjetivas para hiperparâmetros normais em comparações múltiplas, ver DuMouchel (1988).

encerravam quando muitas das afirmações eram possíveis, parando quando a probabilidade *a posteriori* conjunta de que todas elas fossem verdadeiras era menor que 0,95. Eles avaliaram as taxas de erro tipo III para procedimentos clássicos e bayesianos no contexto de modelo hierárquico normal com hiperparâmetros conhecidos e computaram as taxas de erro tipo III para afirmações de confiança clássicas e bayesianas, encontrando taxas de erro extremamente altas (superiores a 0,50) para as clássicas. Segundo os autores, afirmações de confiança bayesianas são conservativas, no sentido de que, afirmações baseadas em intervalos posteriores de 0,95 apresentam taxas de erro tipo III entre 0 e 0,025. Para comparações múltiplas, as conclusões são similares.

Gelman & Tuerlinckx (2000) demonstraram também que a taxa de erro tipo III difere drasticamente da taxa de erro tipo I, tanto para procedimentos clássicos como para os bayesianos.

Scott & Berger (2003) exploraram vários aspectos de comparações múltiplas em análise de DNA, mostrando alguns exemplos.

Lewis & Thayer (2004) mostraram que, minimizar o risco de Bayes para função de perda de uma comparação “0-1”, também controla uma versão de FDR para efeitos aleatórios, confirmando e estendendo os resultados de Shaffer (1999). Segundo os autores, o procedimento de Shaffer (1999) é essencialmente um desenvolvimento de efeitos aleatórios, pois as médias populacionais são amostradas de uma população normal. Para situações práticas em que as variâncias entre médias são desconhecidas, os autores propuseram uma modificação do procedimento que continua a garantir o controle da FDR. Simulações de comparações duas a duas para efeitos aleatórios demonstravam poderes similares àqueles obtido por Benjamini & Hochberg (1995).

Bratcher & Hamilton (2005) apresentaram um procedimento de comparações múltiplas bayesianas para médias ranqueadas e obtidas de populações distribuídas normalmente. Os autores compararam o procedimento

proposto, via simulação, com os usuais, usando uma função de perda linear. Segundo esses autores, mesmo em situações de *prioris* não-informativas, o desempenho do método é muito mais favorável do que o procedimento de Benjamini & Hochberg (1995) com respeito às duas propriedades de maior interesse: a probabilidade de seleção correta e o número esperado de classificações produzidas.

Müller et al. (2006) discutem aproximações bayesianas para problemas de comparações múltiplas, usando uma perspectiva teórica de decisão para comparar criticamente aproximações competentes. Eles estabeleceram problemas de decisão que levaram ao uso de regras baseadas na FDR e generalizações. Usando uma função de perda que controla a FDR, derivaram um critério ótimo de Bayes que é uma variação do procedimento de Benjamin & Hochberg (1995). O ponto de corte é baseado em incrementos nas probabilidades *a posteriori* ordenadas ao invés de p-valores ordenado. É um procedimento totalmente bayesiano, com foco na FDR condicional esperada, condicionada aos dados. Os autores apresentaram uma ampla variedade de aplicações dos problemas de comparações múltiplas em diferentes áreas de pesquisa.

Cho et al. (2006) consideraram o problema específico de comparações múltiplas para populações com distribuição binomial negativa. Eles aplicaram a família Dirichlet na forma de *prioris* base para computar probabilidades *a posteriori* para várias hipóteses dos parâmetros da distribuição binomial negativa. Utilizaram o amostrador de Gibbs devido à dificuldade de se encontrar formas analíticas e ilustraram com um exemplo contendo dados simulados.

Neath & Cavanaugh (2006) consideraram o problema de selecionar amostras independentes de várias populações para comparar grupos, através do teste de hipóteses ou da estimação de diferenças de médias. O objetivo de determinar quais grupos têm médias iguais requer testar uma coleção de

hipóteses relacionadas. Os autores examinaram esse problema para testar as hipóteses do ponto de vista bayesiano, formulando-as como um problema de seleção de modelo. A informação de todos os modelos é combinada através de métodos bayesianos no intuito de fornecer a realidade da incerteza. As probabilidades foram computadas via critério de informação bayesiano. Os autores ilustraram com um exemplo como a aproximação bayesiana levou a uma apresentação lógica dos resultados de comparações múltiplas.



### 3 METODOLOGIA

A descrição da metodologia utilizada neste trabalho foi feita em várias etapas e é apresentada nas subseções seguintes.

#### 3.1 Resultados teóricos

Inicialmente foi considerado o modelo linear, de acordo com a proposta de Box & Tiao (1992), dado por

$$y = X\theta + \varepsilon \quad (42)$$

em que  $y$  é um vetor de observações  $n \times 1$ ,  $X$  uma matriz de constantes conhecidas  $n \times k$ ,  $\theta$  um vetor de parâmetros  $k \times 1$  e  $\varepsilon$  um vetor dos erros aleatórios  $n \times 1$ , tendo média zero (vetor de erros).

Foi assumido que os erros têm distribuição multivariada esfericamente normal  $N_n(0, \sigma^2 I)$  e que o posto da matriz  $X$  é  $k$ , tal que

$$\begin{aligned} p(y|\theta, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2} (y - X\theta)'(y - X\theta)\right] = \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \left[\nu s^2 + (\theta - \hat{\theta})'(X'X)(\theta - \hat{\theta})\right]\right\} \end{aligned} \quad (43)$$

em que

$$\hat{\theta} = (X'X)^{-1} X'y, \quad \nu = n - k, \quad s^2 = (1/\nu) (y - \hat{y})'(y - \hat{y}) \text{ e } \hat{y} = X\hat{\theta}.$$

Segundo Box & Tiao (1992), com variância conhecida, a distribuição *a posteriori* de  $\theta$  é

$$p(\theta | y) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})' (X' X) (\theta - \hat{\theta}) \right\}. \quad (44)$$

Todas as inferências relevantes sobre  $\theta$  podem ser feitas pelo fato de se saber que a distribuição *a posteriori* de  $\theta$  é a distribuição normal multivariada  $N_k(\hat{\theta}, \sigma^2 (X' X)^{-1})$  (Box & Tiao, 1992).

Em muitos casos, a variância  $\sigma^2$  não é conhecida, e a informação sobre  $\sigma^2$  deve ser obtida de uma amostra. A distribuição *a posteriori* de  $(\theta, \sigma^2)$  é então:

$$p(\theta, \sigma^2 | y) \propto p(\theta, \sigma^2) p(s^2 | \sigma^2) p(\hat{\theta} | \theta, \sigma^2). \quad (45)$$

Foi adotada uma distribuição *a priori* não informativa dada por

$$p(\theta | \sigma^2) \propto \sigma^{-2}. \quad (46)$$

Substituindo (46) em (45), a distribuição conjunta *a posteriori* de  $(\theta, \sigma^2)$  pôde ser fatorada de forma que

$$p(\theta, \sigma^2 | y) \propto p(s^2 | \sigma^2) p(\theta | \hat{\theta}, \sigma^2), \quad (47)$$

em que a distribuição marginal *a posteriori* de  $\sigma^2$  é Gama Invertida e a distribuição condicional *a posteriori* de  $\theta$ , dado  $\sigma^2$ , é a normal multivariada, respectivamente, dadas por

$$\sigma^2 | y \sim GI\left(\frac{k}{2}, \frac{k s^2}{2}\right) \quad \theta | \sigma^2, y \sim N_k\left[\hat{\theta}, \sigma^2 (X' X)^{-1}\right] \quad (48)$$

Segue-se então que a distribuição marginal *a posteriori* de  $\theta$  é

$$p(\theta | y) = \frac{\Gamma\left(\frac{\nu+k}{2}\right) |X' X|^{1/2} s^{-k}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^k \Gamma\left(\frac{\nu}{2}\right) (\sqrt{\nu})^k} \left[1 + \frac{(\theta - \hat{\theta})' (X' X) (\theta - \hat{\theta})}{\nu s^2}\right]^{-\frac{\nu+k}{2}} \quad (49)$$

que é a densidade da distribuição t multivariada.

Foi considerado o caso especial relatado em Box & Tiao (1992) onde os elementos do vetor de observações  $y$  são amostras independentes  $y'_1 = (y_{11}, \dots, y_{1n_1}), \dots, y'_k = (y_{k1}, \dots, y_{kn_k})$ , de tamanho  $n_1, \dots, n_k$  ( $\sum n_i = n$ ) de  $k$  populações com médias  $(\theta_1, \dots, \theta_k)$ , respectivamente, e variância comum  $\sigma^2$ . Neste caso, a distribuição *a posteriori* de  $\theta$  em (49), se reduz a

$$p(\theta | y) \propto \left[1 + \frac{\sum_{i=1}^k n_i (\theta_i - \bar{y}_i)^2}{\nu s^2}\right]^{-\frac{\nu+k}{2}}, \quad -\infty < \theta_i < \infty, \quad i=1, \dots, k, \quad (50)$$

com  $\bar{y}_i = \frac{1}{n_i} \sum y_{ij}$ ,  $\nu = n - k$ , e  $s^2 = \frac{1}{n-k} \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$  que foi usada para se realizar as comparações de médias de  $k$  populações normais.

Box & Tiao (1992) realizam as comparações das médias de  $k$  populações normais através da análise de contrastes lineares. A idéia é, a partir da t multivariada, propor procedimentos bayesianos para se comparar médias.

### 3.2 Geração de uma amostra de tamanho $n$ da $t$ multivariada

Considere  $k$  populações com médias  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^t$ . Observações  $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_k\}$  são obtidas dessas populações, em que  $Y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^t$  é um vetor de observações  $n_i \times 1$  condicionalmente independentes no tratamento  $i$ ,  $i = 1, 2, \dots, k$  e  $\sum_{i=1}^k n_i = n$ . O problema de comparações múltiplas (PCM) é fazer inferências sobre as relações entre os  $\boldsymbol{\theta}$  baseados em  $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_k\}$ .

Pela equação (50) pode-se observar que a distribuição *a posteriori* do parâmetro  $\boldsymbol{\theta}$  foi utilizada para se realizar comparações múltiplas de médias normais. Considerando o modelo na forma  $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$  e faz-se  $\boldsymbol{\theta} = \boldsymbol{\mu}_i = \mu + \tau_i$ . A distribuição  $t$  multivariada,  $t(n, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ , possui parâmetros que podem ser especificados por:

$$\boldsymbol{\mu} = \bar{\mathbf{y}} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix}, \quad \boldsymbol{\Sigma} = S_{\bar{\mathbf{y}}} = \begin{bmatrix} \frac{s_1^2}{n_1} & 0 & \dots & 0 \\ 0 & \frac{s_2^2}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{s_k^2}{n_k} \end{bmatrix}, \quad \nu = n - k$$

em que  $n = \sum_{i=1}^k n_i$ ,  $k$  é o número de médias populacionais e  $\boldsymbol{\Sigma}$  é a matriz de covariâncias das médias.

Os elementos localizados “fora” da diagonal indicaram covariâncias nulas entre as populações, sendo considerados independentes, ao passo que o

$k$ -ésimo elemento da diagonal de  $\Sigma$  representa a variância da  $k$ -ésima população sobre o número de elementos da mesma.

Na seqüência foi considerada a forma geral usada para gerar variáveis aleatórias  $k$ -dimensionais  $t$  multivariada com  $\nu$  graus de liberdade e parâmetros  $\mu$  e  $\Sigma$ . Seja um vetor aleatório  $X$  com distribuição  $N_k(\mathbf{0}, \Sigma)$  e a variável escalar  $U$  com distribuição qui-quadrado com  $\nu$  graus de liberdade, então o vetor aleatório  $Y$  dado pela transformação  $Y = \sqrt{\nu} \frac{X}{\sqrt{U}} + \mu$ , possui distribuição  $t$  multivariada com  $\nu$  graus de liberdade.

Foi aplicada a transformação anterior  $n$  vezes, a  $n$  diferentes vetores aleatórios  $X$  e a  $n$  variáveis escalares  $U$ , tendo ao final do processo uma amostra de tamanho  $n$  da distribuição  $t$  multivariada com  $\nu$  graus de liberdade.

Foi utilizado o programa R (R. Development Core Team, 2006) como ambiente de programação, onde todos os procedimentos foram implementados e estão apresentados em anexo.

### 3.3 Distribuição nula da amplitude padronizada

A partir da distribuição *a posteriori*  $t$  multivariada,  $t(n, \mu, \Sigma, \nu)$ , foram geradas  $k$  cadeias de médias  $\mu_i$ , utilizando o método de Monte Carlo e assumindo médias constantes, vetor  $\mu = (\mu, \mu, \dots, \mu)^t$ , ou seja, todos os componentes iguais. Assim, sem perda de generalidade, foi assumido  $\mu = 0$  (para todos os  $k$  componentes), impondo a hipótese nula  $H_0$  no método bayesiano.

No próximo passo foi realizada a geração da amplitude padronizada da *posteriori*, sob  $H_0$ , e obtida na distribuição *a posteriori* das médias da seguinte forma:

$$q = \frac{\text{máx}(\mu_i) - \text{mín}(\mu_i)}{\sigma_h} \quad (51)$$

em que  $\sigma_h$  representa a média harmônica das variâncias das  $k$  médias, dada por:

$$\sigma_h = \sqrt{\frac{1}{\frac{1}{k} \left( \frac{n_1}{s_1^2} + \frac{n_2}{s_2^2} + \dots + \frac{n_k}{s_k^2} \right)}} \quad (52)$$

para contemplar a possibilidade de se analisar tanto o caso de variâncias heterogêneas como o caso de variâncias homogêneas.

O fato de se gerar a amplitude padronizada da *posteriori* sob  $H_0$ , obtida da distribuição *a posteriori* das médias da forma mostrada na equação (52), tem um significado muito importante, pois possibilita a análise sob heterogeneidade de variâncias e sob condições de não-balanceamento, diferentemente dos testes de comparações múltiplas convencionais.

Calcula-se a seguir a diferença mínima significativa  $\Delta = \sigma_h \cdot q_\alpha$ , em que  $q_\alpha$  é o quantil superior 100 $\alpha$ % da distribuição *a posteriori* de  $q$ . Para qualquer amplitude maior que  $\Delta$ , a diferença é significativamente considerada diferente de zero.

### 3.4 Estatísticas descritivas

Para descrever os resultados finais das análises dos métodos propostos neste trabalho, foi considerado um exemplo real e quatro exemplos simulados. Nos exemplos simulados foram consideradas duas situações. Na primeira situação, foram simulados dados de experimentos sob  $H_0$  com  $k = 6$  níveis de um

fator. Na segunda, foram simulados experimentos sob  $H_1$ , sendo que a diferença entre as médias dos níveis do fator foi de dois erros padrões. Em ambos foram consideradas situações balanceadas e homocedásticas e não-balanceadas e heterocedásticas.

Para sintetizar os resultados, inicialmente foram obtidas as densidades *a posteriori* da distribuição da amplitude padronizada  $q$  em cada um dos cinco exemplos utilizados. Para isso foi utilizado um estimador de densidades Kernel do programa R (R. Development Core Team, 2006). Da mesma forma foram obtidos os quantis superiores 10%, 5% e 1% desta distribuição.

Para realizar a inferência a respeito da hipótese  $H_0: \mu_i = \mu_{i'}$ ,  $i \neq i' = 1, 2, \dots, k$ , considerando todos os pares e todos os exemplos, foi obtido

$$\Delta = \sigma_h \cdot q_\alpha, \quad (53)$$

para  $\alpha = 10\%$ ,  $5\%$  e  $1\%$  sendo  $\sigma_h$  obtido a partir da expressão (52).

Da mesma forma foi obtida a média na distribuição *a posteriori* de  $\mu$  e assim, para cada par  $(\mu_i, \mu_{i'})$ , foi obtido o intervalo de confiança bayesiano

$$\text{ICB}_{1-\alpha}(\mu_i - \mu_{i'}): \bar{\mu}_i - \bar{\mu}_{i'} \pm \Delta, \quad (54)$$

em que  $\bar{\mu}_i$  e  $\bar{\mu}_{i'}$  são as médias *a posteriori* das cadeias de  $\mu_i$  e  $\mu_{i'}$ . Todos os intervalos que contiverem zero indicarão que a hipótese  $H_0: \mu_i = \mu_{i'}$ , não deve ser rejeitada; se os limites do  $\text{ICB}_{1-\alpha}$  forem positivos, indicarão que  $\mu_i > \mu_{i'}$ ; se forem negativos,  $\mu_i < \mu_{i'}$ .

Duas outras alternativas foram avaliadas. Na primeira, considerando todos os pares  $(\mu_i, \mu_{i'})$ , foi calculada a seguinte probabilidade *a posteriori*

$$P(\mu_i > \mu_{i'}). \quad (55)$$

Esta probabilidade serviu de forma descritiva como uma evidência a favor ou contra  $H_0$ .

A segunda alternativa seria obter uma nova cadeia com os limites inferior ( $LI^{ii'}$ ) e superior ( $LS^{ii'}$ ) de um intervalo *a posteriori* para cada par de médias ( $\mu_i, \mu_{i'}$ ), da seguinte forma

$$\begin{cases} LI^{ii'} = \mu_{ij} - \mu_{i'j} - q_j \sigma_h \\ LS^{ii'} = \mu_{ij} - \mu_{i'j} + q_j \sigma_h \end{cases} \quad (56)$$

Como medida de evidência a favor ou contra  $H_0: \mu_i = \mu_{i'}$ , calculou-se a probabilidade *a posteriori* dos intervalos conterem o valor zero. Seja  $I(0 \in [LI_j^{ii'}, LS_j^{ii'}])$  a função indicadora para verificar se o valor 0 pertence ao intervalo na  $j$ -ésima unidade amostral de Monte Carlo da cadeia *a posteriori*, logo,

$$P(ICB \supset 0) = \frac{\sum_{j=1}^n I(0 \in [LI_j^{ii'}, LS_j^{ii'}])}{n}. \quad (57)$$

Uma última análise foi realizada utilizando o HPD nas cadeias *a posteriori* da distribuição das diferenças de duas médias. Para cada cadeia  $\mu_i - \mu_{i'}$ , foi obtido o HPD utilizando um coeficiente de credibilidade de 95%, utilizando-se para isso o pacote BOA, um aplicativo do programa R (R Development Core Team, 2006). Os resultados de todos os procedimentos foram comparados em experimentos simulados e reais.



### 3.5 Exemplos

#### 3.5.1 Exemplos simulados sob $H_0$

Foram simulados dois exemplos sob a hipótese nula, ou seja, todas as médias dos níveis do fator foram consideradas iguais. No primeiro caso, considerou-se uma situação balanceada e homocedástica. Foi considerado o valor  $k = 6$  e  $r = 5$ , com  $\sigma^2 = 1$ . Assim, as observações  $y_{ij}$  foram geradas de uma normal  $N(0, 1)$ , para  $i = 1, 2, \dots, k$  e  $j = 1, 2, \dots, r$ .

No segundo caso, considerou-se que as amostras de cada população seriam geradas de distribuições normais com média 0 e variâncias  $\sigma_1^2 = 1$ ,  $\sigma_2^2 = 4$ ,  $\sigma_3^2 = 7$ ,  $\sigma_4^2 = 10$ ,  $\sigma_5^2 = 13$  e  $\sigma_6^2 = 16$ . Isso foi realizado para que houvesse heterogeneidade da variância máxima de  $\delta = \sigma_6^2 / \sigma_1^2 = 16$ . Assim, os valores  $y_{ij}$  foram gerados de  $N(0, \sigma_i^2)$ , considerando os valores das médias dos  $n_i$ 's dados por 3, 5, 4, 3, 2 e 4, para  $i = 1, 2, 3, 4, 5$  e 6, respectivamente.

#### 3.5.2 Exemplos simulados sob $H_1$

Neste caso foram simulados dois exemplos, como na seção 3.5.1, sendo o primeiro balanceado e homocedástico e o segundo não-balanceado e heterocedástico. Os valores de  $k$ ,  $r$  ou  $n_i$  e  $\sigma^2$  ou  $\sigma_i^2$  foram os mesmos. A diferença é que as amostras das populações consideradas possuíam diferentes médias. A média da população 1 foi considerada igual a zero ( $\mu_1 = 0$ ) sem perda de generalidade. As demais médias  $\mu_i$  foram definidas por:

$$\mu_{i+1} = \mu_i + 2 \sigma_h , \quad (58)$$

sendo  $\sigma_h$  a raiz quadrada da média harmônica das variâncias das médias  $\mu_i$ , dada por:

$$\sigma_h = \sqrt{\frac{1}{\frac{1}{k} \left( \frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2} + \dots + \frac{n_k}{\sigma_k^2} \right)}}. \quad (59)$$

### 3.5.3 Exemplo Real

A fim de se aplicar a metodologia, foi utilizado um exemplo real (Steell & Torrie, 1980). 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 1 se encontram os teores de nitrogênio em miligramas obtidos.

TABELA 1: Teor de nitrogênio, em mg, de plantas de trevo-vermelho inoculadas com combinações de culturas de *R. trifolii* e *R. metiloti*.

	Repetições				
	1	2	3	4	5
3Dok1	19,4	32,6	27,0	32,1	33,0
3Dok4	17,7	24,8	27,9	25,2	24,3
3Dok5	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: Steell & Torrie (1980).

A esses dados foi aplicada a metodologia, sendo adotado o nível nominal de 0,05 de significância. O objetivo foi ilustrar a aplicação do procedimento e comparar com os frequentistas existentes.

## 4 RESULTADOS E DISCUSSÃO

Na seqüência, são apresentadas as médias *a posteriori*, a distribuição *a posteriori* da amplitude padronizada  $q$  e seus quantis a 90%, 95% e 99%, os resultados dos testes de comparações múltiplas utilizando os quantis 95% da distribuição  $q$ , os intervalos de confiança  $ICB_{1-\alpha}(\mu_i - \mu_j)$ , usando a amplitude padronizada  $\Delta$ , as probabilidades *a posteriori*  $P(\mu_i > \mu_j)$ , as probabilidades do IC (56) *a posteriori* conter o zero e os HPD's, em três situações, sob  $H_0$ , sob  $H_1$  e aplicados a dados reais. Também foram considerados  $H_0$  e  $H_1$  em dois casos, experimentos balanceados e homocedásticos (comportados) e não balanceados e heterocedásticos (não comportados).

A discussão é prejudicada pela escassez de artigos sob o problema.

### 4.1 Resultados sob $H_0$

Na Tabela 2 são apresentadas as médias originais e *a posteriori* de experimentos simulados sob  $H_0$  comportado e não comportado com  $k = 6$  níveis de um fator. Em ambas as situações (comportado e não comportado), os resultados das médias originais e *a posteriori* foram similares, indicando que a simulação foi consistente. Este resultado já era esperado, uma vez que a média *a posteriori* teoricamente é igual à média *a priori*. Assim, as diferenças podem ser atribuídas apenas ao erro de Monte Carlo, obtido nas 10.000 simulações realizadas.

TABELA 2: Médias originais e *a posteriori* obtidas a partir de dados simulados sob  $H_0$

	Comportado		Não Comportado	
	Média original	Média <i>a posteriori</i>	Média original	Média <i>a posteriori</i>
1	0,255456	0,254579	0,011598	0,043037
2	0,003093	-0,002943	-0,260981	-0,266368
3	-0,207752	-0,207871	-0,964964	-0,965178
4	0,379949	0,371173	-2,338673	-2,357626
5	0,182213	0,184449	0,605279	0,606313
6	-0,087732	-0,091142	1,201431	1,215459

As densidades *a posteriori* da distribuição da amplitude padronizada  $q$  sob  $H_0$  são apresentadas nas Figuras 1 (a) e (b) para as duas situações: (a) comportado e (b) não comportado. Como esperado, a densidade (b) apresenta uma maior variabilidade, comparado com a densidade (a), devido à heterogeneidade das variâncias. As densidades *a posteriori* apresentadas permitem que se afirme que houve uma grande flexibilidade do método, indicado pelo ajuste ao conjunto de dados e as suas características. Assim, diferentemente do caso clássico, os quantis superiores da amplitude padronizada se modificam em função do não balanceamento e da heterogeneidade de variâncias que porventura possam existir nos conjuntos de dados submetidos à análise.

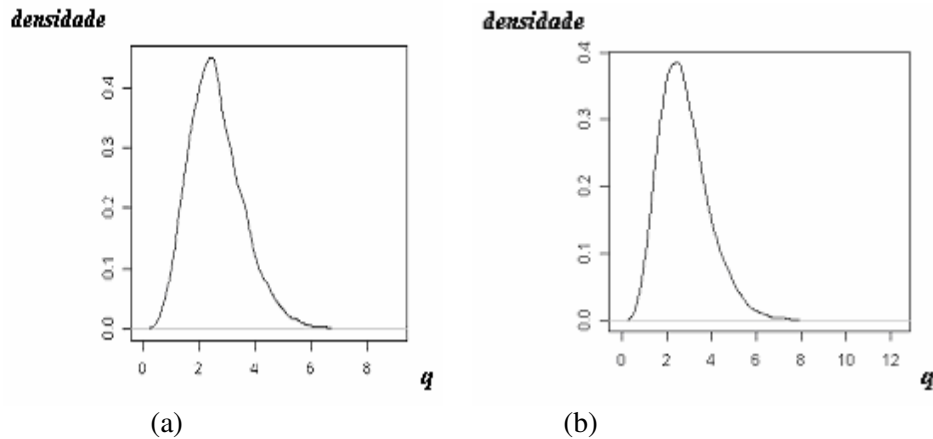


FIGURA 1: Distribuição da amplitude padronizada  $q$  sob  $H_0$  (a) comportado e (b) não comportado.

Na Tabela 3 são apresentados os quantis superiores 10%, 5% e 1% da distribuição da amplitude padronizada  $q$  sob  $H_0$  comportado e não comportado. Os valores dos quantis da amplitude padronizada são muito similares aos da amplitude *estudentizada*  $q$  do teste clássico de comparações múltiplas de Tukey para médias de 6 fatores com 24 graus de liberdade, cujos valores são 3,90, 4,37 e 5,37 para 10%, 5% e 1%, respectivamente, principalmente com os do caso comportado. Isso é esperado, pois a distribuição da amplitude *estudentizada* foi desenvolvida assumindo normalidade, balanceamento e homogeneidade de variâncias. Os resultados encontrados para o caso não comportado reforçam a idéia do desempenho ruim do teste de Tukey, quando não se satisfazem as pressuposições.

TABELA 3: Quantis superiores da amplitude padronizada da distribuição *a posteriori* sob  $H_0$ .

$1 - \alpha$	Comportado	Não Comportado
90%	3,877603	4,278497
95%	4,373322	4,849312
99%	5,389085	6,114813

Os resultados dos testes de comparações múltiplas utilizando os quantis 95% da distribuição *a posteriori* da amplitude padronizada  $q$  sob  $H_0$  comportado e não comportado, são apresentados na Tabela 4. Considerando as diferenças entre todos os pares de médias e comparando com o  $\Delta$ , obtido a partir da equação 53, verifica-se que realmente não há diferença significativa entre as médias dos fatores em ambas as situações. Novamente, este resultado é muito interessante, pois ambas as situações foram simuladas sob a hipótese nula em que as médias dos fatores são iguais e com isso não se esperava detectar diferenças entre as médias.

TABELA 4: Resultado dos testes de comparações múltiplas utilizando os quantis 95% da distribuição *a posteriori* da amplitude padronizada  $q$  sob  $H_0$ .

Comportado		Não Comportado	
Tratamento	Média <i>a posteriori</i>	Tratamento	Média <i>a posteriori</i>
4	0,371173 a	6	1,215459 a
1	0,254579 a	5	0,606313 a
5	0,184449 a	1	0,043037 a
2	-0,002943 a	2	-0,266368 a
6	-0,091142 a	3	-0,965178 a
3	-0,207871 a	4	-2,357626 a
$\Delta = 1,611806$		$\Delta = 4,897561$	

Os intervalos de confiança bayesianos dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada *a posteriori*  $q$  sob  $H_0$  comportado e não comportado, são apresentados na Tabela 5. Todos os intervalos englobaram o zero, que é um indicativo de que a hipótese  $H_0: \mu_i = \mu_r$ , não deve ser rejeitada. Este procedimento é alternativo ao procedimento anterior, considerando a amplitude padronizada *a posteriori*. Portanto, é esperado que os resultados sejam idênticos. Outro fato interessante de ser destacado é que os intervalos de confiança bayesianos para a situação não comportada são mais amplos. Isso é decorrente da maior diferença mínima

bayesiana neste caso, em consequência da maior variabilidade no caso heterocedástico. Assim, o procedimento é adequado tanto para situações comportadas como para situações em que as condições não seriam favoráveis à aplicação de um teste clássico de comparações múltiplas. Em ambos os casos, no entanto, a suposição de normalidade não foi violada.

TABELA 5: Intervalos de confiança bayesianos dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada *a posteriori*  $q$  sob  $H_0$ .

Parâmetros	Comportado			Não Comportado		
	Médias	LI	LS	Médias	LI	LS
$\mu_1 - \mu_2$	0,257522	-1,354284	1,869328	0,309405	-4,588156	5,206966
$\mu_1 - \mu_3$	0,462450	-1,149356	2,074256	1,008215	-3,889346	5,905776
$\mu_1 - \mu_4$	-0,116594	-1,728400	1,495212	2,400663	-2,496898	7,298224
$\mu_1 - \mu_5$	0,070130	-1,541676	1,681936	-0,563276	-5,460837	4,334285
$\mu_1 - \mu_6$	0,345721	-1,266085	1,957527	-1,172422	-6,069983	3,725139
$\mu_2 - \mu_3$	0,204928	-1,406878	1,816734	0,698810	-4,198751	5,596371
$\mu_2 - \mu_4$	-0,374116	-1,985922	1,237690	2,091258	-2,806303	6,988819
$\mu_2 - \mu_5$	-0,187392	-1,799198	1,424414	-0,872681	-5,770242	4,024880
$\mu_2 - \mu_6$	0,088199	-1,523607	1,700005	-1,481827	-6,379388	3,415734
$\mu_3 - \mu_4$	-0,579044	-2,190850	1,032762	1,392448	-3,505113	6,290009
$\mu_3 - \mu_5$	-0,392320	-2,004126	1,219486	-1,571491	-6,469052	3,326070
$\mu_3 - \mu_6$	-0,116730	-1,728536	1,495076	-2,180637	-7,078198	2,716924
$\mu_4 - \mu_5$	0,186724	-1,425082	1,797540	-2,963939	-7,861500	1,933622
$\mu_4 - \mu_6$	0,462315	-1,149491	2,074121	-3,573085	-8,470646	1,324476
$\mu_5 - \mu_6$	0,275591	-1,336215	1,887397	-0,609146	-5,506707	4,288415

Na Tabela 6 são apresentadas probabilidades *a posteriori*  $P(\mu_i > \mu_r)$ , considerando todos os pares  $(\mu_i, \mu_r)$  da distribuição conjunta *a posteriori*  $\leq 2,5$ . Quando a probabilidade  $P(\mu_i > \mu_r)$  ou  $P(\mu_r > \mu_i) \geq 97,5$  tem-se uma evidência contra  $H_0$ , caso  $2,5 < P(\mu_i > \mu_r) < 97,5$ , tem-se uma evidência a favor de  $H_0$ . Esta evidência foi considerada neste estudo por meio da escolha de 95% de confiança, para termos uma situação comparável ao normalmente aplicado aos testes clássicos de comparações múltiplas. Verifica-se que, sob  $H_0$  comportado,



todas as probabilidades evidenciam que a hipótese  $H_0: \mu_i = \mu_{i'}$  não deve ser rejeitada. Contudo, sob  $H_0$  não comportado, há uma situação de evidência contra  $H_0$ , parâmetros  $\mu_4$  e  $\mu_6$ , que indica que o procedimento é mais liberal, com mais chance de cometer erros. Neste caso haveria uma evidência de que  $\mu_6$  supera  $\mu_4$ , o que não era esperado, pois todas as médias por construção foram simuladas sob a hipótese nula de igualdade entre médias. Este procedimento é o que os bayesianos normalmente fazem, ou seja, tomam suas decisões baseadas na distribuição *a posteriori* conjunta dos parâmetros e numa medida probabilística a favor de uma hipótese ou de outra.

Assim, este procedimento, aplicado para fins de comparação, não é equivalente ao da proposta de trabalho que considera a distribuição *a posteriori* da amplitude padronizada para construir o critério de decisão. No caso univariado, essa alternativa seria comparável, diga-se assim, com o teste *t* de *Student* para comparar as médias de tratamentos. O teste de Tukey seria a alternativa equivalente ao procedimento bayesiano, que considera a amplitude *a posteriori* padronizada e que controlaria as taxas de erro por experimento. Portanto, a maior liberalidade desta alternativa seria esperada, como acontece com os seus pares clássicos.

A título de ilustração, na Figura 2 são apresentadas as densidades *a posteriori* da distribuição das diferenças de duas médias  $\mu_i = \mu_{i'}$  sob  $H_0$  comportado para (a)  $i = 1$  e  $i' = 4$  e (b)  $i = 4$  e  $i' = 6$ . Em ambos os casos, a distribuição da diferença englobou o zero em uma região de alta densidade, indicando que a hipótese  $H_0: \mu_i = \mu_{i'}$  não deve ser rejeitada.

TABELA 6: Probabilidades *a posteriori*  $P(\mu_i > \mu_{i'})$ , considerando todos os pares  $(\mu_i, \mu_{i'})$  da distribuição conjunta *a posteriori* sob  $H_0$ .

Parâmetros	Comportado	Não Comportado
$\mu_1$ e $\mu_2$	68,77	58,56
$\mu_1$ e $\mu_3$	80,41	75,15
$\mu_1$ e $\mu_4$	40,65	92,84
$\mu_1$ e $\mu_5$	55,21	37,54
$\mu_1$ e $\mu_6$	74,08	21,90
$\mu_2$ e $\mu_3$	65,21	70,24
$\mu_2$ e $\mu_4$	23,79	92,23
$\mu_2$ e $\mu_5$	36,12	29,55
$\mu_2$ e $\mu_6$	56,75	13,52
$\mu_3$ e $\mu_4$	13,74	82,42
$\mu_3$ e $\mu_5$	23,38	17,95
$\mu_3$ e $\mu_6$	41,31	6,35
$\mu_4$ e $\mu_5$	64,12	5,92
$\mu_4$ e $\mu_6$	81,04	<b>1,45</b>
$\mu_5$ e $\mu_6$	69,74	35,79

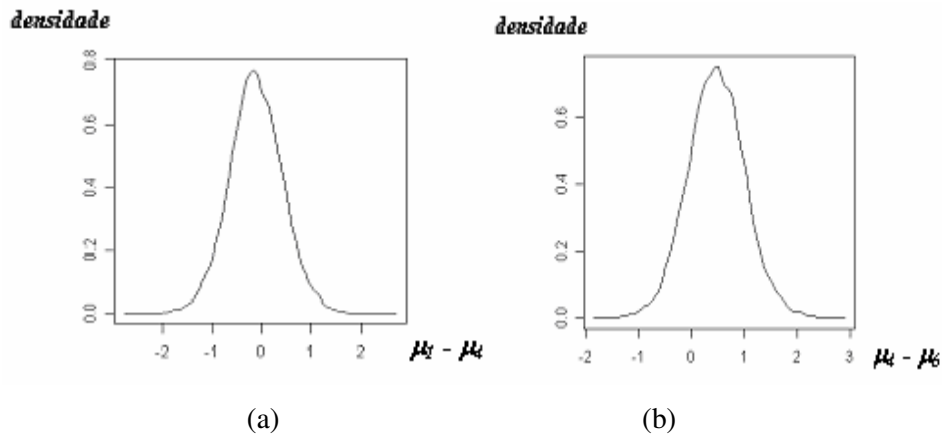


FIGURA 2: Densidade *a posteriori* de  $\mu_i - \mu_{i'}$  para (a)  $i = 1$  e  $i' = 4$  e (b)  $i = 4$  e  $i' = 6$  sob  $H_0$  comportado.

As densidades *a posteriori* da distribuição das diferenças de duas médias  $\mu_i - \mu_{i'}$  sob  $H_0$  não comportado para (a)  $i = 1$  e  $i' = 4$  e (b)  $i = 4$  e  $i' = 6$ , são

apresentadas na Figura 3. Na situação (a), a distribuição *a posteriori* da diferença englobou o zero em uma região de maior densidade, o que é uma evidência forte de que a hipótese  $H_0: \mu_1 = \mu_4$  não deve ser rejeitada. Já na situação (b), o zero está muito próximo do limite superior (resultado similar ao da Tabela 6), que é uma região de baixa densidade, indicando uma situação de forte evidência contrária a hipótese nula  $H_0$ . Estes resultados são equivalentes aos apresentados na Tabela 6 e, por essa razão, apenas alguns casos escolhidos foram plotados para exemplificar as comparações múltiplas entre médias utilizando a distribuição conjunta *a posteriori* das médias por meio de uma função (diferença) entre pares de medidas. As cadeias das diferenças foram obtidas na distribuição *a posteriori* conjunta e os gráficos correspondentes obtidos. É conveniente comentar que, embora os procedimentos gráficos possuam um apelo visual muito grande, eles dificultam a tomada de decisão por não darem um resultado preciso da evidência, como o caso das probabilidades apresentadas na Tabela 6.

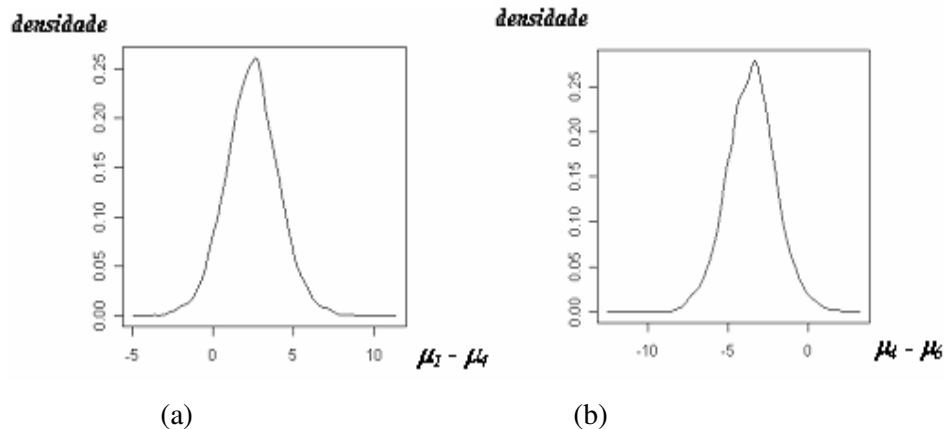


FIGURA 3: Densidade *a posteriori* de  $\mu_i - \mu_{i'}$  para (a)  $i = 1$  e  $i' = 4$  e (b)  $i = 4$  e  $i' = 6$  sob  $H_0$  não comportado.

As probabilidades *a posteriori* dos intervalos de confiança (equação 56) conter o valor zero são apresentadas na Tabela 7. A diferença entre esse método

e o anterior é que as cadeias das diferenças de médias consideram os valores de  $\Delta$  determinados em cada ponto amostral desta cadeia. Assim, espera-se uma maior proteção do erro tipo I quando se utiliza tal procedimento. Quando a probabilidade for menor ou igual a 5% tem-se uma evidência contra  $H_0$ , caso contrário, a favor de  $H_0$ . Verifica-se que, tanto sob  $H_0$  comportado como não comportado, todas as probabilidades constituíram-se em uma forte evidência a favor de  $H_0$ , ou seja, nenhuma das hipóteses  $H_0: \mu_i = \mu_j$  devem ser rejeitadas. Esse procedimento é equivalente ao intervalo de confiança bayesiano usando o  $\Delta$ . A diferença entre eles é que o intervalo bayesiano é obtido como um sumário da cadeia da diferença entre duas médias e a obtenção das probabilidades da Tabela 7 envolve o cálculo do intervalo em cada ponto simulado da cadeia *a posteriori*.

TABELA 7: Probabilidades do intervalo de confiança bayesiano (56) *a posteriori* conter o valor zero sob  $H_0$ .

Parâmetros	Comportado	Não Comportado
$\mu_1$ e $\mu_2$	84,28	87,92
$\mu_1$ e $\mu_3$	77,63	81,19
$\mu_1$ e $\mu_4$	86,40	57,03
$\mu_1$ e $\mu_5$	86,74	80,27
$\mu_1$ e $\mu_6$	81,35	78,52
$\mu_2$ e $\mu_3$	85,61	86,83
$\mu_2$ e $\mu_4$	79,95	63,98
$\mu_2$ e $\mu_5$	85,43	80,44
$\mu_2$ e $\mu_6$	86,53	76,48
$\mu_3$ e $\mu_4$	71,73	76,06
$\mu_3$ e $\mu_5$	79,28	70,60
$\mu_3$ e $\mu_6$	86,39	62,41
$\mu_4$ e $\mu_5$	85,26	45,61
$\mu_4$ e $\mu_6$	76,87	33,07
$\mu_5$ e $\mu_6$	83,76	80,92

Na Tabela 8 são apresentados os HPD's de 95% para cada cadeia  $\mu_i = \mu_j$ . Novamente é interessante que se saliente que este é um procedimento que seria genuinamente adotado pelos estatísticos bayesianos. Neste caso, as cadeias *a posteriori* das diferenças de médias, para todos os pares de médias possíveis, são submetidas a uma análise para obtenção dos intervalos de credibilidade. Este procedimento deve ser melhor do que os procedimentos que utilizam a distribuição *a posteriori* das diferenças de médias apresentados anteriormente, mas não contorna o problema do controle do erro tipo I global nas diferentes análises realizadas simultaneamente.

Verifica-se que, sob  $H_0$  comportado, todos os intervalos contiveram o valor zero sendo um indicativo de que a hipótese  $H_0: \mu_i = \mu_j$  não deve ser rejeitada. Contudo, sob  $H_0$  não comportado, há uma situação de evidência contra  $H_0$  (intervalo  $\mu_4$  e  $\mu_6$  não contém o valor zero). Esse procedimento foi equivalente ao das probabilidades a posteriori  $P(\mu_i > \mu_j)$  neste caso. Espera-se que resultados similares entre estes dois procedimentos sejam observados com muita frequência em situações práticas reais, mas não há garantias de equivalência teórica entre estes dois métodos. O que, no entanto, fica claro, é que ambos os métodos não possuem garantias de controle do erro tipo I simultaneamente nas múltiplas inferências realizadas nas cadeias *a posteriori* das diferenças entre médias. Assim, observando-se o erro tipo I no caso não comportado, tem-se uma evidência de que este procedimento é mais liberal que os que utilizam a distribuição *a posteriori* da amplitude padronizada.

TABELA 8: HPD's de 95% para cada cadeia  $\mu_i - \mu_j$ , sob  $H_0$ .

Parâmetros	Comportado		Não Comportado	
	LIC	LSC	LIC	LSC
$\mu_1$ e $\mu_2$	-0,817692	1,341100	-2,788709	3,213795
$\mu_1$ e $\mu_3$	-0,571935	1,552182	-1,958805	4,092564
$\mu_1$ e $\mu_4$	-1,160501	0,978029	-0,775643	5,801071
$\mu_1$ e $\mu_5$	-0,977498	1,160892	-4,395291	3,002879
$\mu_1$ e $\mu_6$	-0,778737	1,387350	-4,177383	1,954940
$\mu_2$ e $\mu_3$	-0,872901	1,249687	-1,922337	3,497666
$\mu_2$ e $\mu_4$	-1,431171	0,673814	-0,869857	5,139772
$\mu_2$ e $\mu_5$	-1,267582	0,889139	-4,195682	2,559633
$\mu_2$ e $\mu_6$	-0,926989	1,222086	-4,244258	1,176342
$\mu_3$ e $\mu_4$	-1,630423	0,506152	-1,750661	4,415628
$\mu_3$ e $\mu_5$	-1,414343	0,735317	-5,020998	1,980766
$\mu_3$ e $\mu_6$	-1,204674	0,934645	-5,175465	0,627017
$\mu_4$ e $\mu_5$	-0,893001	1,248009	-6,864609	0,627510
$\mu_4$ e $\mu_6$	-0,547104	1,588037	<b>-6,650266</b>	<b>-0,390251</b>
$\mu_5$ e $\mu_6$	-0,808203	1,334888	-4,164519	2,916772

#### 4.2 Resultados sob $H_1$

As médias originais e *a posteriori* de experimentos simulados sob  $H_1$  comportado e não comportado, são apresentadas na Tabela 9, com  $k = 6$  níveis de um fator. Em ambas as situações (comportado e não comportado), os resultados das médias originais e a posteriori foram similares, indicando que a simulação foi consistente, pois estes resultados estão em conformidade com o que é esperado teoricamente. É conveniente salientar que cada média simulada difere da outra por uma diferença de dois erros padrões de uma média. No caso heterocedástico, utilizou-se a média harmônica da variância de uma média para computar o erro padrão. Os resultados que se seguem possuem o mesmo padrão anteriormente relatado para a situação simulada sob a hipótese nula de igualdade

das médias do fator. Assim, as discussões serão apresentadas de forma mais simplificada.

TABELA 9: Médias originais e *a posteriori* obtidas a partir de dados simulados sob  $H_1$ .

	Comportado		Não Comportado	
	Média original	Média <i>a posteriori</i>	Média original	Média <i>a posteriori</i>
1	0,846103	0,850243	0,639328	0,629005
2	1,215129	1,219349	1,465521	1,457039
3	1,592824	1,590603	4,429088	4,431684
4	2,905023	2,905513	9,285582	9,285229
5	3,940113	3,944073	2,892533	2,880282
6	4,591469	4,598216	10,476197	10,485834

Na Figura 4 são apresentadas as densidades *a posteriori* da distribuição da amplitude padronizada  $q$  sob  $H_1$  para as duas situações (a) comportado e (b) não comportado. Como esperado, a densidade (b) apresenta uma maior variabilidade, devido à heterogeneidade das variâncias. O padrão de resposta foi exatamente o mesmo observado sob a hipótese nula. Isso tem uma explicação lógica, pois a distribuição *a posteriori* é gerada sob a imposição da hipótese de igualdade das médias do fator, independentemente da situação ter sido simulada sob  $H_0$  ou sob  $H_1$ . Assim, se o número de repetições e as variâncias forem os mesmos, espera-se obter os mesmos quantis superiores *a posteriori* e o mesmo padrão de resposta das densidades.

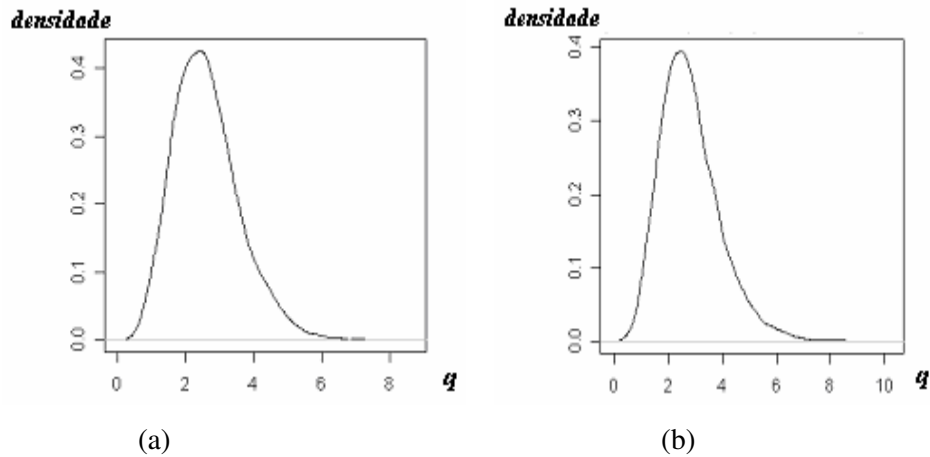


FIGURA 4: Distribuição *a posteriori* da amplitude padronizada  $q$  sob  $H_1$  (a) comportado e (b) não comportado.

Os quantis superiores 10%, 5% e 1% da distribuição *a posteriori* da amplitude padronizada  $q$  sob  $H_1$  comportado e não comportado, são apresentados na Tabela 10. Os quantis encontrados representam os valores que limitam áreas sob a curva de 90%, 95% e 99%. Os valores observados, quando confrontados com os observados sob  $H_0$  comportado e não comportado, mostram uma pequena diferença, sendo um pouco maiores neste último caso.

TABELA 10: Quantis superiores da amplitude padronizada da distribuição *a posteriori* sob  $H_1$ .

$1 - \alpha$	Comportado	Não Comportado
90%	3,933255	4,286779
95%	4,421675	4,873247
99%	5,374668	6,203168

Na Tabela 11 são apresentados os resultados dos testes de comparações múltiplas utilizando os quantis 95% da distribuição *a posteriori* da amplitude padronizada  $q$  sob  $H_1$  comportado e não comportado. Considerando as diferenças entre todos os pares de médias e comparando com o  $\Delta$ , verifica-se que há diferenças significativas entre as médias dos fatores em ambas as situações, mas



que muitas diferenças foram não significativas, indicando que se cometeu o erro tipo II. A heterogeneidade de variâncias afetou consideravelmente o poder do teste. A quantidade de diferenças detectadas foi muito menor no exemplo não comportado em relação às diferenças detectadas no caso comportado, embora muitos erros do tipo II tenham sido cometidos em ambos os casos. As diferenças mínimas significativas foram muito discrepantes, sendo uma delas mais de quatro vezes superior à outra. Isso é um indicativo que o teste, quando aplicado em situações heterocedásticas, deve ser menos poderoso.

TABELA 11: Resultado dos testes de comparações múltiplas utilizando os quantis 95% da distribuição *a posteriori* da amplitude padronizada  $q$  sob  $H_1$ .

Comportado		Não Comportado	
Tratamento	Média <i>a posteriori</i>	Tratamento	Média <i>a posteriori</i>
6	4,598216 a	6	10,485830 a
5	3,944073 ab	4	9,285229 a
4	2,905513 b	3	4,431684 ab
3	1,590603 bc	5	2,880282 b
2	1,219349 c	2	1,457039 b
1	0,850243 c	1	0,629005 b
$\Delta = 1,485137$		$\Delta = 6,222290$	

Os intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada  $q$ , sob  $H_1$  comportado e não comportado, são apresentados na Tabela 12. Alguns intervalos contiveram o valor zero, que é um indicativo de que a hipótese  $H_0: \mu_i = \mu_r$  não deve ser rejeitada, ou seja, representam casos em está se cometendo o erro tipo II. Cada média populacional neste caso difere da sua vizinha por dois erros padrões. Médias vizinhas de um salto igual a 2, possuem valores distantes em quatro erros padrões e assim sucessivamente. Assim, médias mais distantes tendem a ser diferenciadas. Por exemplo,  $\mu_1$  e  $\mu_2$  bem como  $\mu_1$  e  $\mu_3$  não foram identificadas como sendo diferentes no caso comportado, mas  $\mu_1$  e  $\mu_4$ ,  $\mu_1$  e  $\mu_5$  e

$\mu_1$  e  $\mu_6$  foram. As oscilações nos valores das médias experimentais são devidas aos maiores erros associados à situação denominada não comportada.

TABELA 12: Intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada *a posteriori*  $q$  sob  $H_1$ .

Parâmetros	Comportado			Não Comportado		
	Médias	LI	LS	Médias	LI	LS
$\mu_1 - \mu_2$	-0,369106	<b>-1,854243</b>	<b>1,116031</b>	-0,828034	<b>-7,050324</b>	<b>5,394256</b>
$\mu_1 - \mu_3$	-0,740360	<b>-2,225497</b>	<b>0,744777</b>	-3,802679	<b>-10,024969</b>	<b>2,419611</b>
$\mu_1 - \mu_4$	-2,055270	-3,540407	-0,570133	-8,656224	-14,878514	-2,433934
$\mu_1 - \mu_5$	-3,093830	-4,578967	-1,608693	-2,251277	<b>-8,473567</b>	<b>3,971013</b>
$\mu_1 - \mu_6$	-3,747973	-5,233110	-2,262836	-9,856825	-16,079115	-3,634535
$\mu_2 - \mu_3$	-0,371254	<b>-1,856391</b>	<b>1,113883</b>	-2,974645	<b>-9,196935</b>	<b>3,247645</b>
$\mu_2 - \mu_4$	-1,686164	-3,171301	-0,201027	-7,828190	-14,050480	-1,605900
$\mu_2 - \mu_5$	-2,724724	-4,209861	-1,239587	-1,423243	<b>-7,645533</b>	<b>4,799047</b>
$\mu_2 - \mu_6$	-3,378867	-4,864004	-1,893730	-9,028791	-15,251081	-2,806501
$\mu_3 - \mu_4$	-1,314910	<b>-2,800047</b>	<b>0,170227</b>	-4,853545	<b>-11,075835</b>	<b>1,368745</b>
$\mu_3 - \mu_5$	-2,353470	-3,838607	-0,868333	1,551402	<b>-4,670888</b>	<b>7,773692</b>
$\mu_3 - \mu_6$	-3,007613	-4,492750	-1,522476	-6,054146	<b>-12,276436</b>	<b>0,168144</b>
$\mu_4 - \mu_5$	-1,038560	<b>-2,523697</b>	<b>0,446577</b>	6,404947	0,182657	12,627237
$\mu_4 - \mu_6$	-1,692703	-3,177840	-0,207566	-1,200601	<b>-7,422891</b>	<b>5,021689</b>
$\mu_5 - \mu_6$	-0,654143	<b>-2,139280</b>	<b>0,830994</b>	-7,605548	-13,827838	-1,383258

Na Tabela 13 são apresentadas as probabilidades *a posteriori*  $P(\mu_i > \mu_j)$ , considerando todos os pares  $(\mu_i, \mu_j)$ . Quando a probabilidade  $P(\mu_i > \mu_j) \leq 2,5$  ou  $P(\mu_i > \mu_j) \geq 97,5$  tem-se uma evidência considerada significativa contra  $H_0$  ou a favor de  $H_1$ , caso  $2,5 < P(\mu_i > \mu_j) < 97,5$ , tem-se uma evidência a favor de  $H_0$  ou contra  $H_1$ . Verifica-se que, sob  $H_1$  comportado, há 4 (quatro) casos, resultados das probabilidades, evidenciando que a hipótese  $H_0: \mu_i = \mu_j$  não deve ser rejeitada, e que, sob  $H_1$  não comportado, há 5 (cinco) casos. Estes resultados quando comparados com aqueles que utilizaram a distribuição *a posteriori* da amplitude padronizada permitem que infira que este método é mais poderoso do que o anterior. No entanto, sob a hipótese nula, principalmente sob situações não

comportadas, este procedimento foi considerado mais liberal. Assim, a vantagem em relação ao maior poder deve ser olhada com ressalva.

TABELA 13: Probabilidades *a posteriori*  $P(\mu_i > \mu_{i'})$ , considerando todos os pares  $(\mu_i, \mu_{i'})$  da distribuição conjunta *a posteriori* sob  $H_1$ .

Parâmetros	Comportado	Não Comportado
$\mu_1$ e $\mu_2$	<b>22,29</b>	<b>32,50</b>
$\mu_1$ e $\mu_3$	<b>6,41</b>	1,89
$\mu_1$ e $\mu_4$	0,00	0,03
$\mu_1$ e $\mu_5$	0,00	<b>15,73</b>
$\mu_1$ e $\mu_6$	0,00	0,01
$\mu_2$ e $\mu_3$	<b>21,88</b>	2,35
$\mu_2$ e $\mu_4$	0,06	0,00
$\mu_2$ e $\mu_5$	0,00	<b>24,12</b>
$\mu_2$ e $\mu_6$	0,00	0,00
$\mu_3$ e $\mu_4$	0,53	0,86
$\mu_3$ e $\mu_5$	0,00	<b>76,70</b>
$\mu_3$ e $\mu_6$	0,00	0,12
$\mu_4$ e $\mu_5$	1,89	99,29
$\mu_4$ e $\mu_6$	0,13	<b>26,51</b>
$\mu_5$ e $\mu_6$	<b>8,76</b>	0,08

As densidades *a posteriori* da distribuição das diferenças de duas médias  $\mu_i - \mu_{i'}$  sob  $H_1$  comportado para (a)  $i = 3$  e  $i' = 5$  e (b)  $i = 5$  e  $i' = 6$ , são apresentadas na Figura 5. Foram escolhidos dois exemplos ilustrativos para evidenciar alguns aspectos interessantes desta abordagem. Na situação (a), a distribuição da diferença englobou o zero em uma região de baixa densidade, indicando que a hipótese nula  $H_0: \mu_i = \mu_{i'}$  deve ser rejeitada, com uma forte evidência a favor da hipótese alternativa. Já na situação (b), a densidade para a diferença de médias contém o zero em uma região de alta densidade, o que permite que se infira que esta é uma situação de forte evidência a favor de  $H_0$ , levando a um erro tipo II, neste caso. Veja que no primeiro caso as duas médias diferem por quatro erros padrões e no segundo, em dois erros padrões.

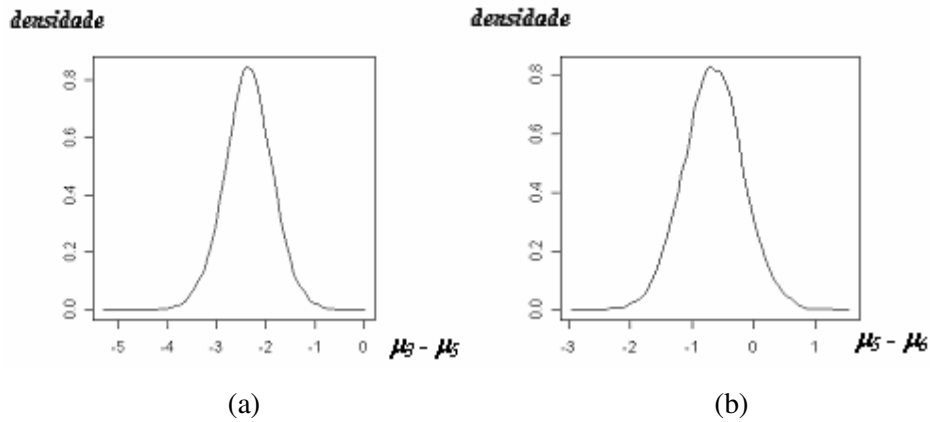


FIGURA 5: Densidade *a posteriori* de  $\mu_i = \mu_{i'}$  para (a)  $i = 3$  e  $i' = 5$  e (b)  $i = 5$  e  $i' = 6$  sob  $H_1$  comportado.

As densidades *a posteriori* da distribuição das diferenças de duas médias  $\mu_i - \mu_{i'}$  sob  $H_1$  para o caso não comportado, considerando os exemplos (a)  $i = 3$  e  $i' = 5$  e (b)  $i = 5$  e  $i' = 6$ , são apresentadas na Figura 6. Na situação (a), o zero na densidade *a posteriori* é um valor típico, ou seja, de alta densidade, sendo que a hipótese  $H_0: \mu_i = \mu_{i'}$  não deve ser rejeitada, o que caracteriza um erro do tipo II. Já na situação (b), o zero está muito próximo do limite superior, indicando uma situação de forte evidência contra  $H_0$ , o que caracterizaria uma decisão correta. Pode-se observar que, no primeiro caso, as diferenças paramétricas eram de quatro erros padrões no primeiro caso, e de dois no segundo. Assim, esperava-se que ambas as situações apresentassem evidências significativas a favor da hipótese alternativa e que no primeiro caso essa evidência fosse mais forte, mas foi o oposto que ocorreu. Isso não é um problema da inferência bayesiana, mas da imprecisão com que as médias experimentais são estimadas devido ao grande erro experimental associado. Pode-se observar que a média do nível 5 do fator (Tabela 9, caso não comportado) é menor do que a média do nível 4, o que não era esperado, uma vez que esta média deveria ser dois erros padrões maior do que a média do nível 4 do fator.

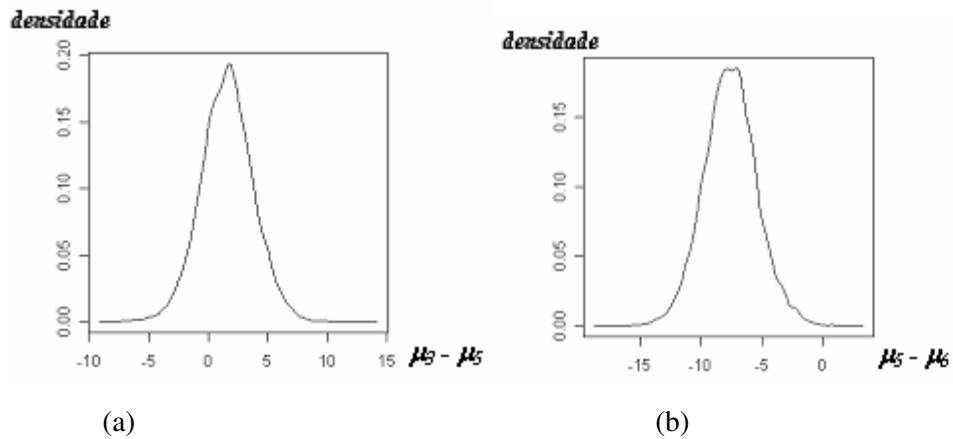


FIGURA 6: Densidade *a posteriori* de  $\mu_i = \mu_{i'}$  para (a)  $i = 3$  e  $i' = 5$  e (b)  $i = 5$  e  $i' = 6$  sob  $H_1$  não comportado.

Na Tabela 14 são apresentadas as probabilidades *a posteriori* dos intervalos IC (56) conter o valor zero, método que deveria ser equivalente ao do intervalo apresentado na Tabela 12. Quando a probabilidade for menor ou igual a 5% tem-se uma evidência a favor de  $H_1$ . Verifica-se que, tanto sob  $H_1$  comportado (8 casos) e não comportado (11 casos), existem probabilidades que evidenciam uma decisão a favor de  $H_0$ , ou seja, a hipótese  $H_0: \mu_i = \mu_{i'}$  não deve ser rejeitada, indicativo de se estar cometendo erro do tipo II. Este procedimento apresentou piores resultados do que os obtidos utilizando o intervalo baseado na média *a posteriori* das cadeias da diferença das médias e com a amplitude padronizada, pois houve mais erros do tipo II no presente caso. Assim, como aconteceu na situação sob a hipótese nula, os métodos não são equivalentes e há um melhor resultado aparente para o caso do intervalo (54).

TABELA 14: Probabilidades do intervalo de confiança bayesiano (56) *a posteriori* conter o valor zero sob  $H_1$ .

Parâmetros	Comportado	Não Comportado
$\mu_1$ e $\mu_2$	<b>78,90</b>	<b>86,29</b>
$\mu_1$ e $\mu_3$	<b>58,13</b>	<b>44,61</b>
$\mu_1$ e $\mu_4$	2,89	2,82
$\mu_1$ e $\mu_5$	0,02	<b>67,23</b>
$\mu_1$ e $\mu_6$	0,00	0,88
$\mu_2$ e $\mu_3$	<b>79,24</b>	<b>59,60</b>
$\mu_2$ e $\mu_4$	<b>8,74</b>	4,33
$\mu_2$ e $\mu_5$	0,16	<b>78,00</b>
$\mu_2$ e $\mu_6$	0,01	1,34
$\mu_3$ e $\mu_4$	<b>22,49</b>	<b>28,68</b>
$\mu_3$ e $\mu_5$	0,92	<b>76,40</b>
$\mu_3$ e $\mu_6$	0,10	<b>13,74</b>
$\mu_4$ e $\mu_5$	<b>38,68</b>	<b>14,62</b>
$\mu_4$ e $\mu_6$	<b>8,77</b>	<b>82,51</b>
$\mu_5$ e $\mu_6$	<b>64,70</b>	<b>6,85</b>

Os HPD's de 95% para cada cadeia  $\mu_i - \mu_r$  são apresentados na Tabela 15. Verifica-se que, tanto sob  $H_1$  comportado (4 casos) e não comportado (7 casos), existem intervalos contendo o valor zero sendo um indicativo de que a hipótese  $H_0: \mu_i = \mu_r$  não deve ser rejeitada, ou seja, há indícios de se estar cometendo erro tipo II. Os resultados deste procedimento, se comparados com os dos demais testes anteriores utilizando a amplitude padronizada, é mais poderoso, por possuir menores taxas de erro do tipo II. No entanto, o erro do tipo I é provavelmente menor neste procedimento de inspiração genuinamente clássica, fator este que deve ser considerado antes de se escolher entre os diversos métodos apresentados neste trabalho. O procedimento que utiliza a amplitude padronizada é um procedimento bayesiano, como os demais propostos no trabalho, mas cuja motivação é baseada em argumentos frequentistas.

TABELA 15: HPD's de 95% para cada cadeia  $\mu_i - \mu_j$  sob  $H_I$ .

Parâmetros	Comportado		Não Comportado	
	LIC	LSC	LIC	LSC
$\mu_1$ e $\mu_2$	<b>-1,366074</b>	<b>0,620659</b>	<b>-4,569094</b>	<b>2,917308</b>
$\mu_1$ e $\mu_3$	<b>-1,720346</b>	<b>0,222036</b>	<b>-7,671973</b>	<b>0,117352</b>
$\mu_1$ e $\mu_4$	-3,011327	-1,058353	-12,756760	-4,462047
$\mu_1$ e $\mu_5$	-4,081105	-2,128489	<b>-7,104385</b>	<b>2,158895</b>
$\mu_1$ e $\mu_6$	-4,756779	-2,805941	-13,802760	-6,055668
$\mu_2$ e $\mu_3$	<b>-1,369324</b>	<b>0,581855</b>	<b>-6,543924</b>	<b>0,330364</b>
$\mu_2$ e $\mu_4$	-2,635127	-0,659352	-11,534620	-4,019615
$\mu_2$ e $\mu_5$	-3,669051	-1,733160	<b>-5,722182</b>	<b>2,929257</b>
$\mu_2$ e $\mu_6$	-4,395516	-2,422127	-12,509800	-5,645992
$\mu_3$ e $\mu_4$	-2,330097	-0,375910	-8,679467	-0,816287
$\mu_3$ e $\mu_5$	-3,354348	-1,383923	<b>-2,694470</b>	<b>6,120848</b>
$\mu_3$ e $\mu_6$	-3,960627	-1,999558	-9,700422	-2,514968
$\mu_4$ e $\mu_5$	-2,021750	-0,092038	1,723547	11,135510
$\mu_4$ e $\mu_6$	-2,703934	-0,745852	<b>-5,118156</b>	<b>2,546080</b>
$\mu_5$ e $\mu_6$	<b>-1,622429</b>	<b>0,298863</b>	-11,899310	-3,094075

### 4.3 Resultados para dados reais

As médias originais e *a posteriori* de experimento real são apresentadas na Tabela 16, com  $k = 6$  níveis de um fator. Da mesma maneira observada anteriormente, as médias originais e *a posteriori* foram similares, indicando que a simulação foi consistente.

TABELA 16: Médias originais e *a posteriori* obtidas a partir dos dados reais.

	Média original	Média <i>a posteriori</i>
1	28,82	28,3822
2	23,98	23,8398
3	14,64	14,4637
4	19,92	19,4399
5	13,26	13,4456
6	18,70	18,2529

A densidade *a posteriori* da distribuição da amplitude padronizada  $q$  é apresentada na Figura 7. Na Tabela 17 são apresentados os quantis superiores 10%, 5% e 1% da distribuição *a posteriori* da amplitude padronizada  $q$  para o conjunto de dados. Os valores encontrados se adequam bem à densidade apresentada na Figura 7. Os valores dos quantis da amplitude padronizada são muito similares aos quantis da distribuição de Tukey para médias de 6 fatores com 24 graus de liberdade, cujos valores são 3,90, 4,37 e 5,37 para 10%, 5% e 1%, respectivamente.

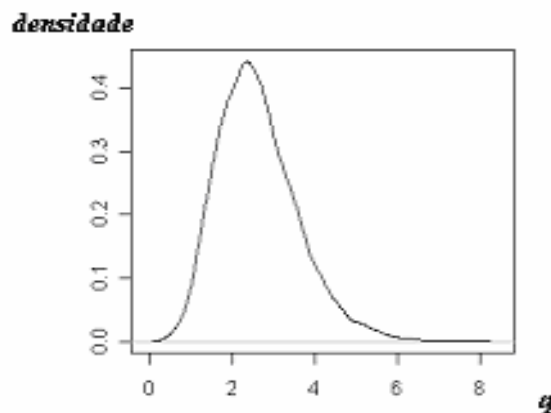


FIGURA 7: Distribuição da amplitude padronizada  $q$ .

TABELA 17: Quantis superiores da amplitude padronizada da distribuição *a posteriori*  $q$ .

$1 - \alpha$	$q$
90%	3,9020
95%	4,3844
99%	5,4722

Na Tabela 18 são apresentados os resultados dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada  $q$  (Tabela 17) e uma comparação com os resultados de Steel & Torrie (1980). Verifica-se que os resultados são muito diferentes, pois pares de tratamentos



indicados como sendo diferentes pelo teste Tukey clássico, foram considerados similares pelo método bayesiano, dado pela equação (54). Além disso, há de se observar que o problema da ambigüidade continua existindo, mas em uma dimensão menor.

TABELA 18: Resultado dos testes de comparações múltiplas utilizando os quantis 95% da distribuição *a posteriori* da amplitude padronizada  $q$ .

Tratamento	Média <i>a posteriori</i>	Steel & Torrie (Teste Tukey)
1	28,3822 a	28,820 a
2	23,8398 ab	23,980 b
4	19,4399 bc	19,920 bc
6	18,2529 bc	18,700 c
3	14,4637 c	14,640 cd
5	13,4456 c	13,260 d
	$\Delta = 6,7321$	dms = 4,482

Na Tabela 19 são apresentados os intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada  $q$ . Os resultados são idênticos aos apresentados na Tabela 18.

TABELA 19: Intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada *a posteriori*  $q$ .

Parâmetros	Médias	LI	LS
$\mu_1 - \mu_2$	4,5424	-2,1897	11,2745
$\mu_1 - \mu_3$	13,9185	7,1864	20,6506
$\mu_1 - \mu_4$	8,9423	2,2102	15,6744
$\mu_1 - \mu_5$	14,9366	8,2045	21,6687
$\mu_1 - \mu_6$	10,1293	3,3972	16,8614
$\mu_2 - \mu_3$	9,3761	2,6440	16,1082
$\mu_2 - \mu_4$	4,3999	-2,3322	11,1320
$\mu_2 - \mu_5$	10,3942	3,6621	17,1263
$\mu_2 - \mu_6$	5,5869	-1,1452	12,3190
$\mu_3 - \mu_4$	-4,9762	-11,7083	1,7559
$\mu_3 - \mu_5$	1,0181	-5,7140	7,7502
$\mu_3 - \mu_6$	-3,7892	-10,5213	2,9429
$\mu_4 - \mu_5$	5,9943	-0,7378	12,7264
$\mu_4 - \mu_6$	1,1870	-5,5451	7,9191
$\mu_5 - \mu_6$	-4,8073	-11,5394	1,9248

As probabilidades *a posteriori*  $P(\mu_i > \mu_r)$ , são apresentadas na Tabela 20, considerando todos os pares  $(\mu_i, \mu_r)$ . Quando a probabilidade  $P(\mu_i > \mu_r) \leq 2,5$  ou  $P(\mu_i > \mu_r) \geq 97,5$  tem-se uma evidência contra  $H_0$ , caso  $2,5 < P(\mu_i > \mu_r) < 97,5$ , tem-se uma evidência a favor de  $H_0$ . Verifica-se que os resultados são similares para os pares de médias que foram identificados como não diferentes por este último método. No entanto, alguns resultados que foram não significativamente diferentes no método anterior, como por exemplo para as médias 1 e 2, foram considerados diferentes neste. Isso indica que este método é mais liberal e que a proposta de considerar a amplitude padronizada, como é feita tradicionalmente nos procedimentos frequentistas, pode ser uma boa estratégia.

TABELA 20: Probabilidades *a posteriori*  $P(\mu_i > \mu_{i'})$ , considerando todos os pares  $(\mu_i, \mu_{i'})$  da distribuição conjunta *a posteriori*.

Parâmetros	Probabilidade
$\mu_1$ e $\mu_2$	97,95
$\mu_1$ e $\mu_3$	100,00
$\mu_1$ e $\mu_4$	99,99
$\mu_1$ e $\mu_5$	100,00
$\mu_1$ e $\mu_6$	99,97
$\mu_2$ e $\mu_3$	99,98
$\mu_2$ e $\mu_4$	96,02
$\mu_2$ e $\mu_5$	100,00
$\mu_2$ e $\mu_6$	98,97
$\mu_3$ e $\mu_4$	1,21
$\mu_3$ e $\mu_5$	73,75
$\mu_3$ e $\mu_6$	3,74
$\mu_4$ e $\mu_5$	99,77
$\mu_4$ e $\mu_6$	70,81
$\mu_5$ e $\mu_6$	0,76

As densidades *a posteriori* da distribuição das diferenças de duas médias  $\mu_i = \mu_{i'}$  para (a)  $i = 2$  e  $i' = 4$  e (b)  $i = 1$  e  $i' = 6$ , são apresentadas na Figura 8. Na situação (a) a distribuição *a posteriori* da diferença entre médias possuía o zero como um valor típico da distribuição, indicando que a hipótese  $H_0: \mu_i = \mu_{i'}$  não deve ser rejeitada. Já na situação (b) a distribuição da diferença não continha o valor zero como um valor típico, sendo um indicativo de que as médias são diferentes. Estes dois resultados são equivalentes aos resultados apresentados na Tabela 20, como era esperado.

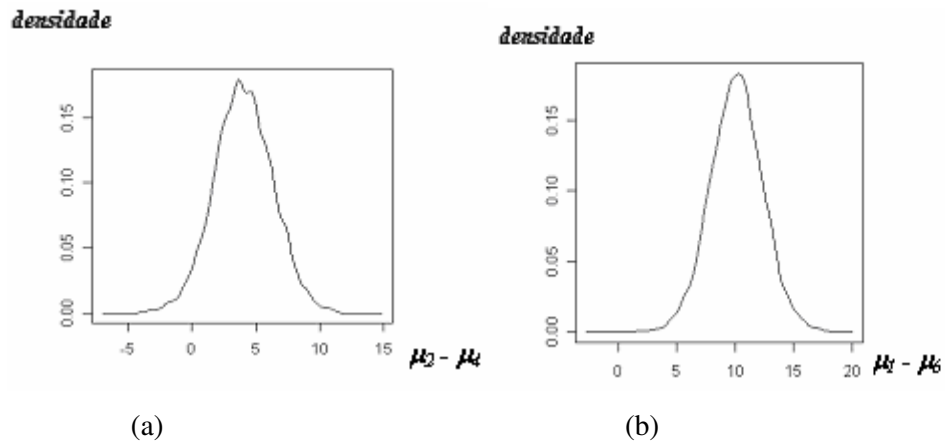


FIGURA 8: Densidade *a posteriori* de  $\mu_i = \mu_{i'}$  para (a)  $i = 2$  e  $i' = 4$  e (b)  $i = 1$  e  $i' = 6$ .

As probabilidades *a posteriori* dos intervalos de confiança (equação 56) conter o valor zero, são apresentadas na Tabela 21. Quando a probabilidade for menor o igual a 5% tem-se uma evidência contra  $H_0$  (6 casos), caso contrário, a favor de  $H_0$  (9 casos). Verifica-se que os resultados são equivalentes aos encontrados nos intervalos de confiança bayesiano dos testes de comparações múltiplas utilizando os quantis 95% da distribuição da amplitude padronizada *a posteriori*  $q$ .

Os HPD's de 95% para cada cadeia  $\mu_i = \mu_{i'}$  são apresentados na Tabela 22. Verifica-se que alguns dos intervalos contém o valor zero (4 casos), o que é um indicativo de que a hipótese  $H_0: \mu_i = \mu_{i'}$  não deve ser rejeitada. Esse procedimento foi equivalente ao das probabilidades *a posteriori*  $P(\mu_i > \mu_{i'})$ , considerando todos os pares  $(\mu_i, \mu_{i'})$  da distribuição conjunta *a posteriori*.

TABELA 21: Probabilidades do intervalo de confiança bayesiano (56) *a posteriori* conter o zero.

Parâmetros	Probabilidade
$\mu_1$ e $\mu_2$	36,63
$\mu_1$ e $\mu_3$	<b>0,05</b>
$\mu_1$ e $\mu_4$	<b>4,33</b>
$\mu_1$ e $\mu_5$	<b>0,00</b>
$\mu_1$ e $\mu_6$	<b>1,54</b>
$\mu_2$ e $\mu_3$	<b>2,89</b>
$\mu_2$ e $\mu_4$	48,93
$\mu_2$ e $\mu_5$	<b>1,02</b>
$\mu_2$ e $\mu_6$	31,49
$\mu_3$ e $\mu_4$	30,33
$\mu_3$ e $\mu_5$	81,43
$\mu_3$ e $\mu_6$	47,73
$\mu_4$ e $\mu_5$	16,16
$\mu_4$ e $\mu_6$	82,67
$\mu_5$ e $\mu_6$	28,35

TABELA 22: HPD's de 95% para cada cadeia  $\mu_i - \mu_r$ .

Parâmetros	LIC	LSC
$\mu_1$ e $\mu_2$	0,3504	9,4803
$\mu_1$ e $\mu_3$	9,6751	18,5621
$\mu_1$ e $\mu_4$	4,3638	13,3429
$\mu_1$ e $\mu_5$	10,9692	19,9846
$\mu_1$ e $\mu_6$	5,5586	14,4815
$\mu_2$ e $\mu_3$	4,8919	13,8872
$\mu_2$ e $\mu_4$	<b>-0,4954</b>	<b>8,5665</b>
$\mu_2$ e $\mu_5$	6,3833	15,3025
$\mu_2$ e $\mu_6$	0,8580	9,7804
$\mu_3$ e $\mu_4$	-9,8140	-0,7551
$\mu_3$ e $\mu_5$	<b>-3,1187</b>	<b>5,9731</b>
$\mu_3$ e $\mu_6$	<b>-8,6322</b>	<b>0,3874</b>
$\mu_4$ e $\mu_5$	2,1095	11,1305
$\mu_4$ e $\mu_6$	<b>-3,3139</b>	<b>5,6222</b>
$\mu_5$ e $\mu_6$	-9,8757	-0,9268

Para simplificar a comparação entre os diversos testes, uma legenda foi montada e apresentada na Tabela 23.

TABELA 23: Legenda dos procedimentos de comparações múltiplas.

Teste	Legenda
Tukey	T
Calinski & Corsten	C
Calinski & Corsten - versão bootstrap	CB
$ICB_{1-\alpha}(\mu_i - \mu_{i'})$ usando a amplitude padronizada $\Delta$	IC $\Delta$
Probabilidades a posteriori $P(\mu_i > \mu_{i'})$	$P(\mu_i > \mu_{i'})$
Probabilidades do IC (56) <i>a posteriori</i> conter o zero	IC 56
HPD's	HPD

Na Tabela 24 são apresentados os resultados comparativos dos testes acima citados para o conjunto de dados reais. Os resultados indicam uma similaridade entre os intervalos de confiança  $ICB_{1-\alpha}(\mu_i - \mu_{i'})$  usando a amplitude padronizada  $\Delta$  e as probabilidades do IC (56) *a posteriori* conter o zero. O mesmo ocorre entre as probabilidades a posteriori  $P(\mu_i > \mu_{i'})$  e os HPD's, sendo, além disso, equivalentes ao Tukey.

Os intervalos de confiança  $ICB_{1-\alpha}(\mu_i - \mu_{i'})$ , usando a amplitude padronizada  $\Delta$  (IC  $\Delta$ ), são mais conservativos, enquanto as probabilidades a posteriori  $P(\mu_i > \mu_{i'})$  são indicadoras de uma maior liberalidade.

Já os testes de Calinski & Corsten (C) e sua versão bootstrap (CB) não controlam o erro tipo I para os dados em questão.

TABELA 24: Resultados dos vários procedimentos quanto às comparações múltiplas realizadas.

Parâmetros	T	C	CB	IC $\Delta$	$P(\mu_i > \mu_i')$	IC 56	HPD
$\mu_1$ e $\mu_2$	*	ns	ns	ns	*	ns	*
$\mu_1$ e $\mu_3$	*	ns	*	*	*	*	*
$\mu_1$ e $\mu_4$	*	ns	*	*	*	*	*
$\mu_1$ e $\mu_5$	*	*	*	*	*	*	*
$\mu_1$ e $\mu_6$	*	*	*	*	*	*	*
$\mu_2$ e $\mu_3$	*	ns	*	*	*	*	*
$\mu_2$ e $\mu_4$	ns	ns	*	ns	ns	ns	ns
$\mu_2$ e $\mu_5$	*	*	*	*	*	*	*
$\mu_2$ e $\mu_6$	*	*	*	ns	*	ns	*
$\mu_3$ e $\mu_4$	*	ns	ns	ns	*	ns	*
$\mu_3$ e $\mu_5$	ns	*	*	ns	ns	ns	ns
$\mu_3$ e $\mu_6$	ns	*	*	ns	ns	ns	ns
$\mu_4$ e $\mu_5$	*	*	*	ns	*	ns	*
$\mu_4$ e $\mu_6$	ns	*	*	ns	ns	ns	ns
$\mu_5$ e $\mu_6$	*	ns	ns	ns	*	ns	*

\* - significativo

ns - não significativo

#### 4.4 Considerações Finais

O desempenho da versão bootstrap (CB) é, em geral, considerado superior ao do teste original de Calinski & Corsten (C), sendo mais conservativos que os demais testes apresentados.

Os testes de comparações múltiplas bayesianos encontrados na literatura apresentam, em geral, resultados similares aos de Tukey, utilizando distribuição de Dirichlet .

O procedimento que encontra os intervalos de confiança  $ICB_{1-\alpha}(\mu_i - \mu_i')$  usando a amplitude padronizada  $\Delta$  (IC  $\Delta$ ) é mais conservativo e flexível a situações de heterocedasticidade. Apresenta resultados praticamente

equivalentes ao método que calcula as probabilidades do IC (56) *a posteriori* conter o zero (IC 56).

O procedimento que calcula as probabilidades *a posteriori*  $P(\mu_i > \mu_j)$  e os HPD's, foram equivalentes entre si e ao de Tukey. São também mais liberais, identificando diferenças em situações em que elas não existem.

Os procedimentos propostos apresentam a desvantagem de ainda não estarem implementados. Em compensação, apresentam vantagens em relação aos testes convencionais, no sentido de não haver necessidade de balanceamento dos dados, o que é muito significativo do ponto de vista prático, e de poderem ser utilizados em situações homo e heterocedásticas.

Um próximo passo seria fazer análises de validação e implementação dos testes.



## 5 CONCLUSÕES

Os procedimentos baseados na amplitude padronizada foram superiores aos demais procedimentos estudados por terem controlado, nos exemplos simulados, o erro tipo I e detectadas a maior parte das diferenças sob  $H_1$ .

Os procedimentos de comparações múltiplas bayesianos foram propostos com sucesso para situações de normalidade, com ou sem homogeneidade de variâncias.

## REFERÊNCIAS BIBLIOGRÁFICAS

AGRESTI, A.; MIN, Y. Frequentist performance of Bayesian Confidence Intervals for Comparing Proportions in 2 x 2 Contingency Tables, **Biometrics**, Oxford, v.61, n.2 June, p.515-523, 2005.

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

BENJAMIN, Y., HOCHBERG, Y. Controlling the false discovery rate. **Journal of the Royal Statistical Society**. 57, p.289-300, 1995.

BERGER, J.O. **Statistical Decision Theory and Bayesian Analysis**. New York: Springer Series in Statistics, 2º ed., 1985. 606p.

BERGER, J.O.; BERLINER, L.M. Robust Byes and empirical Bayes analysis with  $\epsilon$  contaminated priors. **Annals Statistics**, Hayward, v.14, n.3 May, p.461-486, 1986.

BERNARDO, J.M.; SMITH, A.F.M. **Bayesian Theory**. New York: Wiley, 1994.

BERRY, D.A. Multiple comparisons, multiple tests, and data dredging: a Bayesian perspective (with discussion). In: BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V., SMITH, A. F. M., **Bayesian Statistics**, vol. 3. Oxford: Oxford University Press, p.79-94, 1988.

BERRY, D.A. Subgroup analyses. **Biometrics**. 47, p.1227-1230, 1990.

BERRY, D.A.; HOCHBERG, Y. Bayesian perspectives on multiple comparisons. **Journal of Statistical Planning and Inference** 82, p.215-227, 1999.

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

BOX, G.E.P.; TIAO, G.C. **Bayesian Inference in Statistical Analysis**. Toronto: Wiley Classics Library ed. Wiley-Interscience, 1992.

BRATCHER, T., HAMILTON, C. A Bayesian multiple comparison procedure for ranking the means of normally distributed data. **Journal of Statistical Planning and Inference** 133, p.23-32, 2005.

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

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.

CHO, J.S., MASOOM, M.A., BEGUM, M. Bayesian multiple comparisons with nonparametric Dirichlet process prioris for negative binomial populations. **Pak. Journal Statistical**, vol 22(2), p.89-98, 2006.

DAVIS, C.; GAITO, J. Multiple comparison procedures within experimental research. **Canadian Psychology**, 25, 1-13, 1984.

DUMOUCHEL, W. A Bayesian model and graphical elicitation procedure for multiple comparisons. In: BERNARDO, J.M., DEGROOT, M.H., LINDLEY, D.V., SMITH, A.F.M., **Bayesian Statistics**, vol. 3. Oxford: Oxford University Press, 1988.

DUNCAN, D.B. Multiple-range and multiple-F tests. **Biometrics**, 11(1):1-42, 1955.

DUNCAN, D.B. A Bayesian approach to multiple comparisons. **Technometrics**, 7:171-222, 1965.

EHLERS, R.S. **Introdução à Inferência Bayesiana**. Departamento de Estatística, UFPR. Disponível em <http://leg.est.ufpr.br/~ehlers/notas/bayes2006>. 2005. 78p.

FISHER, R.A. **The Design of Experiments**. Edinburgh and London: Oliver and Boyd, Ltda. 1935.

GELMAN, A.; CARLIN, J.B.; STERN, I.S.; RUBIN, D.B. **Bayesian Data Analysis**. London: Chapman and Hall, 2003. 668p.

GELMAN, A., TUERLINCKX, F. Type S error rates for classical and Bayesian single and multiple comparison procedures. **Computational Statistics** 15, p.373-390, 2000.

GIRI, N.C. **Multivariate Statistical Analysis**. New York, STATISTICS: textbook and monographs, Marcel Dekker, vol.149, 2004. 378p.

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

GOPALAN, R.; BERRY, D.A. Bayesian multiple comparisons using Dirichlet process priors. **Journal American Statistical Association**. 93 (443), p.1130-1139, 1998.

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.

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

JEFFREYS, H. **Theory of probability**. Oxford University Press, London, 1961.

KEULS, M. The use of the studentized range in connection with an analysis of variance. **Euphytica**, 1, 112-122, 1952.

LEONEL, A.C. **Combinação de probabilidade na comparação em análise conjunta de experimentos**. Lavras: UFLA-DEX, 2000 p. 81. (Dissertação – Mestrado em Estatística e Experimentação Agropecuária).

LEWIS, C.; THAYER, D.T. A loss function related to the FDR for random effects multiple comparisons. **Journal of Statistical Planning and Inference**. 125, p.49-58, 2004.

LINDSEY, D.V. **Parametric Statistical Inference**, New York, Clarendon Press Oxford, 1996. 487p.

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 AGRÔNOMICA, 11., Londrina, 2005. **Curso...** Londrina: ISBN, 2005. 290p.

MEYER, P. L. **Probabilidade: Aplicações à Estatística**. Tradução do prof. Ruy de C. B. Lourenço Filho. 2.ed. Rio de Janeiro: LTC, 1984. 426 p.

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

MÜLLER, P., PARMIGIANI, G., RICE, K. FDR and Bayesian Multiple Comparisons Rules. In: PROC. VALENCIA / ISBA 8 th WORLD MEETING ON BAYESIAN STATISTICS., Benidorm (Alicante, Spain), June 1<sup>st</sup>-6<sup>th</sup>, 2006.

NEATH, A.A., CAVANAUGH, J.E. A Bayesian Approach to the Multiple Comparisons Problem. **Journal of Data Science** 4, p.131-146, 2006.

NETER, J.; WASSERMAN, W. **Applied linear statistical models – regression, analysis of variance and experimental designs**. Illinois: Richard D. Irwin, 1974. 841p.

NOGUEIRA, M.C.S. **Curso de estatística experimental aplicada à experimentação agrônômica**. ESALQ/USP. Piracicaba, 1997.

O'HAGAN, A. **Bayesian Inference**, Volume 2B. Edward Arnold, Cambridge, 1994.

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.

PAULINO, C.D.; TURKMAN, M.A.A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa, Fundação Calouste Gulbenkian, 2003. 444p.

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

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

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

RAMOS, P. S. **Agrupamento de médias via bootstrap para populações normais e não-normais** Lavras: UFLA-DEX, 2007. 76p. (Dissertação – Mestrado em Estatística e Experimentação Agropecuária).

RINGLAND, J.T. Robust multiple comparisons. **Journal American Statistical Association**, 78, p.145-151, 1983.

SANTOS, C. **Novas alternativas de testes de agrupamento avaliadas por meio de simulação de Monte Carlo**. Lavras: UFLA-DEX, 2000. p.85. (Dissertação-Mestrado em Estatística e Experimentação Agropecuária).

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.

SCOTT, J., BERGER, J. An exploration of aspects of Bayesian multiple testing. **Tech. rep.**, Duke University, ISDS, 2003.

SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. **Biometrika**, 40(1/2): 87-104, 1953.

SHAFFER, P. J. A semi-Bayesian study of Duncan's Bayesian multiple comparison procedure. **Journal of Statistical Planning and Inference**, 82, 197-213. 1999.

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: a biometrical approach**. New York: MacGraw-Hill Book Company, 1980, 633p.

TONHASCA, Jr.A. The three “capital sins” of statistics used in biology. **Ciência e cultura**, v. 43, n.6, p.417-422, 1991.

TUKEY, J.W. Some thoughts on clinical trials, especially problems of multiplicity. **Science**, 198, p.679-684, 1977.

WALLER, R.A.; DUNCAN, D.B. A Bayes rule for the symmetric multiple comparisons problem. **Journal of the American Statistical Association**, 64(328):1484-1503, 1969.

WESTFALL, P. H.; JOHNSON, W. O.; UTTS, J. M. A Bayesian perspective on the Bonferroni adjustment. **Biometrika** **84**, p.419-427, 1997.

## **ANEXOS**

ANEXO 1: Programa .....	92
-------------------------	----



## ANEXO 1: Programa

---

```
# Tese PC - Funções para realizar inferências Bayesianas para realizar compara-  
# ções múltiplas
```

```
# Função para gerar uma amostra N da t multivariada: t(N,Yb,Syb,nu=n-k)  
# Posteriori marginal para o vetor de médias populacionais  
# em que  $n = \sum_i^k n_i$ , k: número de médias populacionais, e Syb é a  
# matriz de Covariâncias das médias (diagonal- independentes)  
# Utiliza o pacote mvtnorm - deve ser carregado antes de utilizar a função
```

```
tpostmult = function(N=100,Yb=c(0,0,0),Syb=diag(3),nu=10)  
{  
  y=rmvnorm(N,sigma=diag(nrow(Syb)))/sqrt(rchisq(N,nu)/nu)  
  Sybroot = diag(diag(Syb)^0.5)  
  x=y%*%Sybroot+matrix(c(rep(t(Yb),times=N)),N,nrow(Syb),byrow =  
TRUE)  
  return(x)  
}
```

```
# Funções para computar a distribuição nula da amplitude padronizada:  
#  $q=(\mu_{(k)}-\mu_{(1)})/\text{Sig}_{(\mu_i)}$ , em que  $\text{Sig}^2_{(\mu_i)} = (S^2/n_i)*\text{nu}/(\text{nu}-2)$   
# utilizar média harmônica das variâncias, pois existe possibilidade  
# de serem diferentes. Utilizada somente com imposição da  $H_0$  via  
# bayesiana.  $P(\text{Muly}) = t(0,\text{Syb},\text{nu})$ 
```

```

qpostbayes = function(N,Yb,Syb,nu)
{
  y=rmvnorm(N,sigma=diag(nrow(Syb)))/sqrt(rchisq(N,nu)/nu)
  Sybroot = diag(diag(Syb)^0.5)
  y=y%%*%%Sybroot
  sigh = (1/(sum(1/diag(Syb))/nrow(Syb)))^0.5
  q = (apply(y,1,max)-apply(y,1,min))/sigh
  q=list(q=q,sigh=sigh)
  return(q)
}

# Exemplo de dados amostrais reais
# Dissert. Patrícia
library(mvtnorm)
ni=c(5,5,5,5,5);N=100000
kk=length(ni)+1
dados=read.table("c:\\Doutorado\\Tese\\DadosReaisPatricia.txt",header=T)
X=as.matrix(dados)
trt=X[,1]
y=X[,2]
dados=list(trt=trt,y=y)
Yb=tapply(dados$y,list(dados$trt),mean)
dados$trt=as.factor(dados$trt)
anovadados=lm(dados$y~dados$trt)
anava= anova(anovadados)
S2=anava$"Mean Sq"[2]
nu=anava$"Df"[2]
Syb = diag(S2/ni)

```

```

Cadeia=tpostmult(N,Yb,Syb,nu)
var(Cadeia); Syb*nu/(nu-2)
apply(Cadeia,2,mean)
q = qpostbayes(N,Yb,Syb,nu)
q$q=sort(q$q)
#q$q
#plot(q$q)
q$q[N*0.90]
q$q[N*0.95]
q$q[N*0.99]
q$sigh
delta=q$sigh*q$q[N*0.95]
delta

# como juntar as coisas
Cadeia1=cbind(Cadeia,as.matrix(q$q))
# para usar o Boa
m1m2=Cadeia1[,1]-Cadeia1[,2]
m1m3=Cadeia1[,1]-Cadeia1[,3]
m1m4=Cadeia1[,1]-Cadeia1[,4]
m1m5=Cadeia1[,1]-Cadeia1[,5]
m1m6=Cadeia1[,1]-Cadeia1[,6]
m2m3=Cadeia1[,2]-Cadeia1[,3]
m2m4=Cadeia1[,2]-Cadeia1[,4]
m2m5=Cadeia1[,2]-Cadeia1[,5]
m2m6=Cadeia1[,2]-Cadeia1[,6]
m3m4=Cadeia1[,3]-Cadeia1[,4]
m3m5=Cadeia1[,3]-Cadeia1[,5]

```

```

m3m6=Cadeia1[,3]-Cadeia1[,6]
m4m5=Cadeia1[,4]-Cadeia1[,5]
m4m6=Cadeia1[,4]-Cadeia1[,6]
m5m6=Cadeia1[,5]-Cadeia1[,6]

# Calcular probabilidades a posteriori de mi>mi'
p12=length(m1m2[m1m2>0])/N*100
p12
p13=length(m1m3[m1m3>0])/N*100
p13
p14=length(m1m4[m1m4>0])/N*100
p14
p15=length(m1m5[m1m5>0])/N*100
p15
p16=length(m1m6[m1m6>0])/N*100
p16
p23=length(m2m3[m2m3>0])/N*100
p23
p24=length(m2m4[m2m4>0])/N*100
p24
p25=length(m2m5[m2m5>0])/N*100
p25
p26=length(m2m6[m2m6>0])/N*100
p26
p34=length(m3m4[m3m4>0])/N*100
p34
p35=length(m3m5[m3m5>0])/N*100
p35

```

```

p36=length(m3m6[m3m6>0])/N*100
p36
p45=length(m4m5[m4m5>0])/N*100
p45
p46=length(m4m6[m4m6>0])/N*100
p46
p56=length(m5m6[m5m6>0])/N*100
p56

# Obtenção das probabilidades do IC 5.6 conter o valor 0
ii=1;jj=2
LIj=Cadeia1[,ii]-Cadeia1[,jj]-Cadeia1[,kk]*q$sigh
LSj=Cadeia1[,ii]-Cadeia1[,jj]+Cadeia1[,kk]*q$sigh
y=c((0>=LIj) & (0<=LSj))
p12=length(y[y==TRUE])/N*100
p12
# repetir o procedimento para todas as 15 combinações

hist(q$q)
plot(density(q$q, kernel="gaussian"))

```

---