COMPARAÇÕES MÚLTIPLAS EM PARES DE MÉDIAS

JOSÉ ROBERTO ZORZATTO

UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

CAMPINAS - SÃO PAULO
BRASIL

COMPARAÇÕES MÚLTIPLAS EM PARES DE MÉDIAS

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. José Roberto Zorzatto e aprovada pela comissão julgadora.

Campinas, de 1985.

Adem J. Petenate

Prof. Dr. Ademir José Petenate
Orientador

Dissertação apresentada ao Instituto Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Estatística.

ABRIL / 1985.

À minha esposa Angela e a nossas filhas
Jessica Antonieta e Cristiane pelos
momentos de renúncia e compreensão,
dedicação e amor.

Todos sabem que a produção científica, em geral, não é mais um ato individual, por isso agradeço:

- Ao Prof. Dr. Ademir José Petenate por ter me introduzido neste assunto, por seu estímulo e segura orientação.

- Aos professores do Departamento de Estatística do IMECC - - UNICAMP e aos colegas do curso de Pós-Graduação.

- Aos meus pais João Zorzatto e Sebastiana Suave Zorzatto, a quem muito devo e que sempre me apoiaram.

- Ao PICD/CAPES, pelo apoio financeiro.

- À Fundação Universidade Federal de Mato Grosso do Sul, por investir em minha capacitação docente.

- Aos colegas do Centro Universitário de Corumbá, que contribuíram para o meu afastamento.

- À todos que direta ou indiretamente colaboraram para viabilizar este trabalho.

Meus sinceros agradecimentos.

JOSÉ ROBERTO ZORZATTO

ÍNDICE

INTRODUÇÃO

O procedimento de comparaçõe múltiplas (PCM) para pares de médias comumente recomendado é o método T, de Tukey (1953). Porém, para usar o método de Tukey algumas restrições devem ser consideradas, tais como, normalidade da população, independência das observações, tamanhos de amostras iguais e todas as observações devem ter a mesma variância.

O presente trabalho investiga alternativas quando estas restrições são violadas.

Vários PCM serão apresentados e comparados entre si sendo que avaliaremos os aspectos de conservantismo, optimalidade, conveniência e robustez.

Estaremos, também, pesquisando o inflacionamento do nível de significância α promovido pelo método T em situação de heterogeneidade de variância e qual o comportamento de α se o método T for aplicado após uma estabilização de variância, via transformação nas nas observações.

Finalmente, investigaremos a robustez do método T para comparações múltiplas em pares de médias na presença de uma correlação constante entre as observações de um mesmo tratamento.

CAPÍTULO I

INFERÊNCIA SIMULTÂNEA - O PROBLEMA GERAL

1.1. ANÁLISE DE VARIÂNCIA COM UM FATOR FIXO.

Quando estamos interessados na comparação de médias de vários grupos podemos usar um conjunto de técnicas, desenvolvidas originariamente por Sir Ronald A. Fisher, denominadas Análise de Variância. Essa metodologia nos permite identificar possíveis diferenças entre médias populacionais devidas a vários fatores que atuam sobre os elementos da população.

O modelo linear comumente utilizado para representar a estrutura da análise de variância com um fator fixo é dada por:

$$Y_{ij} = \mu + A_i + e_{ij} \quad , \quad \begin{array}{l} i=1,\ldots,k \\ j=1,\ldots,n_i \end{array} \qquad (1.1.1)$$

onde

Y_{ij} : j-ésima observação no i-ésimo tratamento,

μ : média global,

A_i : efeito do i-ésimo tratamento, com a restrição que $\sum_{i=1}^{k} A_i = 0$,

e_{ij} : erro aleatório associado à observação Y_{ij}.

A notação que vamos adotar para represemtar as observações obtidas no experimento é dada na tabela 1.1.

TABELA 1.1

Observações referentes a um experimento com um fator com k níveis.

	tratamento 1	2	...	k	Total
	Y_{11} . . . Y_{1n_2}	Y_{21} . . . Y_{2n_2}	... ...	Y_{k1} . . . Y_{kn_k}	
média amostral	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$	...	$\bar{Y}_{k.}$	$\bar{Y}_{..}$
tamanho da amostra	n_1	n_2	...	n_k	n

onde

n_i : é o número de unidades experimentais submetidas ao i-ésimo tratamento,

$\bar{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$: média amostral do i-ésimo tratamento ,

$\bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}$: média global das n observações ,

$n = \sum_{i=1}^{k} n_i$.

A partição usual da soma de quadrados correspondente ao modelo descrito em (1.1.1) é expressa por

$$ST = SA + SR$$

sendo

$ST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$ (soma de quadrados total)

$SA = \sum_{i=1}^{k} n_i (y_{i.} - \bar{y}_{..})^2$ (soma de quadrados entre tratamentos)

$SR = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$ (soma de quadrados residual).

As suposições usualmente associadas aos componentes do modelo (1.1.1) são que os e_{ij} são variáveis aleatórias independentes e identicamente distribuídas com distribuição $N(0,\sigma^2)$.

Em resumo, podemos escrever

$$Y_{ij} \sim N(\mu_i,\sigma^2) \quad , \quad \mu_i = \mu + A_i \; .$$

Portanto, estamos supondo que as observações do experimento a ser analisado correspondem a amostras aleatórias de k populações normais com a mesma variância e que podem ou não ter médias iguais.

Assim, em nossa análise de variância estaremos interessados em testar a igualdade de médias, que em termos do modelo (1.1.1) será formulada por:

$$H_o : A_1 = A_2 = \ldots = A_k = 0 \quad , \qquad (1.1.2)$$

Seja o vetor de erros $\underset{\sim}{e} = (e_{11},\ldots,e_{k,r})$.

Por hipótese temos que $\underset{\sim}{e} \sim N(\phi, I\sigma^2)$ e a estatística usual para se testar H_o é

$$F = \frac{SA/_{k-1}}{SR/_{n-k}} \quad ,$$

que sob a hipótese nula (1.1.2) tem distribuição F central, com k-1 e n-k graus de liberdade. Em resumo, indicamos:

$$F \sim F_{k-1,n-k} \quad , \quad \text{sob} \quad H_o \; .$$

Rejeitamos H_o ao nível de significância α se

$$F > F_{k-1,n-k,\alpha}$$

onde $F_{k-1,n-k,\alpha}$ é o quantil de ordem $(1-\alpha)$ da distribuição F-Snedecor com $(k-1)$ e $(n-k)$ graus de liberdade.

1.2. COMPARAÇÕES MÚLTIPLAS.

Através do teste F podemos testar, a um nível de significância α, a hipótese nula H_o de que todas as médias μ_i são iguais. Se H_o for rejeitada, estaremos admitindo que $\mu_i \neq \mu_j$ para pelo menos um par de tratamentos distintos i e j e é interessante prosseguirmos a análise a fim de localizarmos as diferenças entre as médias nos diferentes tratamentos. A continuidade da análise pode ser feita através de técnicas estatísticas denominadas comparações múltiplas as quais permitem testar hipóteses do tipo:

$$H_o : C = 0$$

$$H_1 : C \neq 0 \; ,$$

onde:

$$C = \sum_{i=1}^{k} c_i \mu_i \qquad (1.2.1)$$

com a restrição de que $\sum_{i=1}^{k} c_i = 0$.

A função C definida em (1.2.1) é denominada comparação ou contraste.

Em nossos estudos estaremos interessados em Procedimentos de Comparações Múltiplas (PCM) onde as comparações consideradas serão da forma $C_{ij} = \mu_i - \mu_j$, $i \neq j$. Outros casos são discutidos exaustivamente em Miller (1981).

A discussão abrangerá vários PCM nos quais serão considerados as seguintes situações:

i) Modelo balanceado com homocedasticidade

ii) Modelo não balanceado com homocedasticidade

iii) Modelo balanceado com heterocedasticidade

iv) Modelo não balanceado com heterocedasticidade

v) Estabilização da variância para usar o modelo T.

Além destas cinco situações será considerada também a questão da não independência entre os erros dentro de um mesmo tratamento.

O objetivo principal deste trabalho é investigar qual dos métodos é o procedimento a ser usado em cada situação específica.

Em nossa avaliação consideraremos de maior importância a discussão dos seguintes aspectos:

I - CONSERVANTISMO: Um PCM é conservativo se apresenta um vício negativo no nível de significância. Assim, nossa preocupação é verificar se o nível de confiança simultâneo real é maior ou igual ao nível de confiança simultâneo declarado $1-\alpha$. Lembramos que o nível de significância simultâneo α é da forma $\alpha \leq \sum_{i=1}^{k^*} \alpha_i$, onde α_i é o nível de significância individual para o i-ésimo intervalo de confiança considerado.

II - OPTIMALIDADE: Um procedimento conservativo ou aproximadamente conservativo produz intervalos de confiança estreitos?

III - CONVENIÊNCIA: O procedimento é de fácil utilização? Estão disponíveis tabelas interpretadas? Quais os gastos de processamento computacional associados com o procedimento de comparações múltiplas? O último em questão é de importância crucial quando o procedimento é incorporado a um pacote estatístico.

IV - ROBUSTEZ: Quando um método estatístico é construído, muitas suposições básicas são feitas; algumas delas apenas para facilitar o desenvolvimento matemático, como por exemplo a suposição de normalidade, Scheffé (1959). Usualmente algumas destas suposições são violadas. Métodos estatísticos são ditos robustos se as inferências não são seriamente invalidadas pela violação de tais suposições. Esta é uma questão que dá margem a uma ampla discussão e será parcialmente aprofundada.

Para comparação dos diversos procedimentos de comparações múltiplas alguns comentários e conceitos se fazem necessários:

i) Razão de Erros de Família.

Consideremos uma família de afirmações $F = \{S_f\}$, $f = 1,2,\ldots, N(F)$, onde cada afirmação S_f pode ser correta ou incorreta. Por exemplo, cada afirmação pode ser uma hipótese onde um erro é feito se nós rejeitarmos a hipótese nula que é verdadeira (erro tipo I) ou cada afirmação pode ser um intervalo de confiança onde um erro é feito se o intervalo de confiança não incluir o verdadeiro valor do parâmetro. Nós assumiremos que o número $N(F)$ de afirmações na família é finito. O caso $N(F) = \infty$ é discutido em Miller (1981).

Seja $N_w(F)$ o número de afirmações incorretas na família; definiremos a razão de erros da família por:

$$\text{erro}(F) = \frac{N_w(F)}{N(F)} .$$

A razão de erro é uma variável aleatória bem definida cuja distribuição depende do procedimento utilizado ao construir a família de afirmações e a estrutura probabilística subjacente. Métodos para inferência estatística estão preocupados com a construção S_f de maneira que satisfaça algumas condições colocadas a partir da razão de erros. Uma condição comumente usada e sob a qual nos restringiremos é a probabilidade da razão de erro o qual é definida por:

$$
\begin{aligned}
\alpha = P(F) &= P(\text{erro}(F) > 0) \\
&= P(\frac{N_w(F)}{N(F)} > 0) \\
&= P(N_w(F) > 0) \\
&= \sum_{k=1}^{N(F)} P(N_w(F) = k) \\
&= 1 - P(N_w(F) = 0).
\end{aligned}
$$

Claramente α é a probabilidade que ao menos uma afirmação na família é incorreta. Quando estivermos comparando as médias de tratamentos duas a duas, diremos também que α é o nível de significância conjunto (ou simultâneo) de todos os testes.

Uma pergunta que poderia surgir neste ponto seria o porque não usar vários testes t para comparar as diversas médias de tratamentos. Em um experimento com mais do que dois tratamentos, algum investigador poderia fazer vários testes t ao nível de significância 5% no processo de sumarização dos resultados. Se não existe diferenças entre as médias verdadeiras nos tratamentos a probabilidade de obter um resultado significativo em um destes testes é 0.05. Tem sido reconhecido ao longo do tempo, (Cochran e Cox (1957)), todavia, que se vários testes t forem realizados simultaneamente a probabilidade que ao menos um destes testes seja significativo é maior do que 0.05. Se os testes são independentes esta probabilidade é 0.23 para 5 testes, 0.40 para 10 testes, 0.64 para 20 testes e $1-0.95^n$ para n testes. Destes resultados surge a pergunta se nós devemos tratar de controlar as frequências com que qualquer afirmação errônea seja feita no resumo de um experimento, preferivelmente a uma afirmação errônea seja feita em um teste t individual (Carmer e Walker (1981)).

O mesmo caso se apresenta quando vários intervalos de confiança são determinados. A probabilidade de que ao menos um dos intervalos não inclua a verdadeira diferença dos tratamentos é maior do que o estipulado 0.05.

Recentes trabalhos tem esclarecido esta e questões relacionadas. Um problema em que as pesquisas tem contribuído é esse de ordenar um conjunto de médias de tratamentos em ordem de rendimento. Deste modo, diferenças significativas raramente existem entre todos os pares de tratamentos. O máximo que pode usualmente aparecer é uma separação parcial tal como por exemplo entre 4 tratamentos

T_1 $\quad$ T_2 $\quad$ T_3 $\quad$ T_4

A linha que une T_1 e T_3 indica que não foi demonstrado que estes dois tratamentos sejam significativamente diferentes; T_1 é, porém, significativamente diferente de T_3 e T_4. Similarmente, T_3 difere de T_4 mas não de T_2, e T_2 difere significativamente de T_4.

ii) Razão de Comprimentos de Intervalos.

Ao compararmos métodos A e B cujos intervalos de confiança para $\mu_i - \mu_j$ sejam dados por $\bar{Y}_{i.} - \bar{Y}_{j.} \pm I_{ij}(A)$ e $\bar{Y}_{i.} - \bar{Y}_{j.} \pm I_{ij}(B)$, respectivamente, e que tais intervalos tenham coeficiente de confiança conjunto igual a $\gamma = 1 - \alpha$, podemos definir

$$R_{ij}(A/B) = I_{ij}(A) \ / \ I_{ij}(B)$$

a razão de seus comprimentos de intervalo de confiança. Claramente, se $R_{ij}(A/B) < 1$, então A é preferível a B para estimar $\mu_i - \mu_j$.

1.3. COMPARAÇÕES MÚLTIPLAS PARA PLANEJAMENTOS BALANCEADOS - - O MÉTODO DE TUKEY.

O MÉTODO DE TUKEY.

O método de Tukey é um procedimento de comparações múltiplas que envolve a distribuição da amplitude studentizada e foi proposto por John Tukey em 1953* e, agora, é comumente referido como o método T.

Sejam $\{Y_{ij}, i = 1,\ldots,k, j = 1,\ldots,r\}$ k amostras de tamanhos r de variáveis aleatórias independentes e normalmente distribuídas com variância comum σ^2 e $E(Y_{ij}) = \mu_i$ (esperança de Y_{ij}), $i = 1,\ldots,k$, $j = 1,\ldots,r$. Seja S^2 o estimador usual ANOVA de σ^2 com $v = k(r-1)$ graus de liberdade; $\frac{vS^2}{\sigma^2}$ tem uma distribuição qui-quadrado e independente dos $\{\bar{Y}_{i.}\}$.

No caso do modelo apresentado na secção 1.1 em que comparam-se k amostras, existem $k(k-1)/2$ pares de médias que poderiam ser comparadas. Nós desejamos testar simultaneamente as $\binom{k}{2}$ hipótese nulas $H_o : \mu_i = \mu_j$, $i \neq j$ contra as várias alternativas bilaterais.

Seja R a amplitude dos $\{\bar{Y}_{i.}\}$, isto é,

$$R = \max_i \{\bar{Y}_{i.}\} - \min_i \{\bar{Y}_{i.}\}$$

e consideremos a distribuição da amplitude studentizada

$$SR_{k,v} = \frac{R}{S} \tag{1.3.1}$$

consequentemente teremos $SR_{\alpha,k,v}$ tal que

---

* Conforme Dunnett (1980a), o método T está em um manuscrito não publicado na Princeton University.

$$P(SR_{k,v} > SR_{\alpha,k,v}) = \alpha \tag{1.3.2}$$

A técnica do método T é incorporada na seguinte declaração de probabilidade:

$$P\{|(\overline{Y}_{i.} - \overline{Y}_{j.}) - (\mu_i - \mu_j)| \leq SR_{\alpha,k,v} \frac{S}{\sqrt{r}}\ ,\ i,j=1,\ldots,k\} = 1-\alpha$$

onde

$$\overline{Y}_{i.} = \frac{1}{r} \sum_j Y_{ij} \quad \text{e} \quad S^2 = \frac{1}{k(r-1)} \sum_{i,j} (Y_{ij} - \overline{Y}_{i.})^2$$

e $SR_{\alpha,k,v}$ é o α-ésimo ponto da distribuição da amplitude studentizada como k tratamentos e $k(r-1)$ graus de liberdade.

A probabilidade do erro tipo I, simultaneamente, é α para os seguintes intervalos de confiança das $\binom{k}{2}$ diferenças de médias:

$$\mu_i - \mu_j \in \overline{Y}_{i.} - \overline{Y}_{j.} \pm SR_{\alpha,k,v} \frac{S}{\sqrt{r}} \tag{1.3.3}$$

com $i,j = 1,\ldots,k$ e $i \neq j$.

É importante destacar que as médias populacionais μ_i e μ_j serão declaradas significativamente diferentes se o intervalo de confiança dado em (1.3.3) não incluir o zero e que tais diferenças são julgadas com base nos valores amostrais. Os valores de $SR_{\alpha,k,v}$ são tabelados, entre outros, em Snedecor e Cochran (1969).

O uso do método T nos assegura que a probabilidade de fazer pelo menos um erro, por concluir incorretamente que duas médias populacionais não são iguais quando elas realmente são iguais, não excede ao valor α estipulado.

O uso deste método é restrito a comparações baseadas em médias amostrais onde as amostras devem necessariamente ter o mesmo tamanho. A suposição de iguais variâncias também deve ser satisfeita.

Muitos dos métodos que discutiremos na próxima secção para a estimação simultânea das diferenças em pares de médias em casos não balanceados são aplicáveis para casos balanceados. Seria concebível que um destes métodos poderia produzir intervalos de confiança com amplitude menor do que o obtido pelo método T. Genizi e Hochberg (1978), porém, afasta esta possibilidade em seus estudos de funções lineares $\ell = \sum_{i=1}^{k} \ell_i \mu_i$ de $\mu_1, \ldots, \mu_k$ ao mostrar que o método T e somente o método T chega a estes limites mais baixo para todas as classes de comparações em pares $\mu_1, \ldots, \mu_k$ em casos balanceados com variâncias homogêneas.

Ramseyer e Tcheng (1973) e Brown (1974) dão evidências através de simulação que o método de Tukey compartilha a propriedade de robustez da distribuição F central com respeito a não normalidade sendo que Brown (1974) indica que o método T pode ser não robusto na presença de alguns casos de extrema não-normalidade. Estes pesquisadores consideram ainda que o método de Tukey é razoavelmente robusto em relação a heterogeneidade de variâncias, apesar de obterem em seus resultados um pequeno inflacionamento do nível de significância α, em tal circunstância.

Nos estudos de simulação da secção (3.4) pode-se ter uma idéia do quanto o método T inflaciona o nível de significância α para alguns casos de variâncias heterogêneas e, se o pesquisador estiver interessado em manter um nível de significância conservativo, algumas soluções são propostas em (3.7).

Com respeito ao critério de comparações em pares de média o método de Tukey é exato e óptimo (a menos que suposições adicionais sejam feitas) e é razoavelmente robusto para os casos em que as variâncias não são muito divergentes, o que nos leva a recomendá-lo para estimação simultânea de todas as diferenças em pares de médias $\mu_i - \mu_j$, $i \neq j$, em planejamentos balanceados com um fator.

1.4. PROCEDIMENTOS DE COMPARAÇÕES MÚLTIPLAS PARA PLANEJAMENTOS NÃO BALANCEADOS E VARIÂNCIAS HOMOGÊNEAS.

Uma das limitações do método T é que tenhamos tamanhos de amostras iguais em um experimento. Muitas vezes o pesquisador defronta-se com experimentos em que os tratamentos não podem sofrer um mesmo número de réplicas, por exemplo quando as réplicas em um determinado tratamento representa um custo elevado. Diante disso, muitos métodos foram criados, a maioria deles propondo uma extensão para o método T, para situações onde ocorra não balanceamento experimental. A seguir serão apresentados alguns deste métodos e na secção 2, uma comparação entre eles.

O MÉTODO DE TUKEY-KRAMER.

O método de Tukey-Kramer, também conhecido como método TK e Procedimento de Kramer, aparentemente foi discutido pela primeira vez por John Tukey em 1953 e independentemente por Kramer (1956).

O intervalo de confiança para todas as $k(k-1)/2$ diferentes médias $\mu_i - \mu_j$, $i \neq j$; simultaneamente, de nível $1-\alpha$, do método T é dado por

$$\overline{Y}_{i.} - \overline{Y}_{j.} \pm SR_{\alpha,k,v} \cdot S \cdot \frac{1}{\sqrt{r}} \quad . \qquad (1.4.1)$$

Kramer propôs que se tivéssemos não balanceamento experimental, isto é, se $n_i \neq n_j$ onde n_i e n_j são os números de replicações nas amostras i e j respectivamente, a seguinte modificação poderia ser feita a partir de (1.4.1):

$$\overline{Y}_{i.} - \overline{Y}_{j.} \pm SR_{\alpha,k,v}\, S\sqrt{\frac{1}{2}\left(\frac{1}{r} + \frac{1}{r}\right)}$$

assim,

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm SR_{\alpha,k,v}\, S \sqrt{\frac{1}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

então,

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm SR_{\alpha,k,v} \frac{S}{\sqrt{2}} (n_i^{-1} + n_j^{-1})^{1/2} \tag{1.4.2}$$

onde

$$S^2 = \frac{1}{\sum_{\ell=1}^{k} n_\ell - k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 \ . \tag{1.4.3}$$

Logo os limites propostos em (1.4.2) são os limites do método TK. Desta forma, o método de Tukey-Kramer foi conhecido como uma extensão aproximada do método T para casos não balanceados, onde o tamanho amostral comum r em (1.4.1) é substituído pela média harmônica $2/(n_i^{-1} + n_j^{-1})$ de dois tamanhos amostrais diferentes n_i e n_j.

Note que para casos balanceados o método TK reduz-se ao método T, e para k=2 tratamentos obtém-se o usual intervalo de confiança para $\mu_i - \mu_j$ baseado na distribuição t-student.

Pelas pesquisas de Dunnett (1980a) e Stoline (1981) concluiu-se que método TK é preferível entre todos os métodos aqui descritos para a situação em que tivermos planejamento experimental envolvendo não balanceamento e variâncias homogêneas.

O MÉTODO DE SCHEFFÉ

O método (S) de Scheffé (1953) é aplicável para estimação simultânea de todos os contrastes $C = \sum_{i=1}^{k} C_i \mu_i$, $\sum_{i=1}^{k} C_i = 0$, das médias $\mu_1, \ldots, \mu_k$. O método S fornece intervalos de confiança conjunto de nível $1-\alpha$ para todos os contrastes $\mu_i - \mu_j$,

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm S[(k-1)(F_{\alpha,k-1,v})(n_i^{-1} + n_j^{-1})]^{1/2}$$

onde $F_{\alpha,k-1,v}$ é o quantil de ordem $(1-\alpha)$ da distribuição F-Snedecor com $(k-1)$ e $n-k$ graus de liberdade. O método S é exato mesmo para experimentos não balanceados, no entanto, é quase universalmente admitido que outros métodos produzem intervalos de confiança para pares de médias com menor amplitude, (Scheffé (1959), Miller (1981) e Ury (1976)). Por esta razão o método S é não recomendado para comparação em pares de médias.

O MÉTODO DE BONFERRONI.

Este procedimento de comparações múltiplas é baseado em uma desigualdade de probabilidades conhecida como a desigualdade de Bonferroni (ou desigualdade de Boole) o qual diz que para quaisquer eventos $E_1, E_2, \ldots, E_k$ têm-se

$$P\left(\bigcup_{i=1}^{k} E_i\right) = P(E_1 \text{ ou } E_2 \text{ ou } \ldots \text{ ou } E_k) \leq \sum_{i=1}^{k} P(E_i).$$

Sejam $Z_1, Z_2, \ldots, Z_k$ variáveis aleatórias normalmente distribuídas com médias $\mu_1, \mu_2, \ldots, \mu_k$ e variâncias $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectivamente. Os Z_i's podem (ou não) ser independentes. Sejam $S_1^2, S_2^2, \ldots, S_k^2$ estimadores qui-quadrado de $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ com $v_1, v_2, \ldots, v_k$ graus de liberdade, respectivamente; isto é, $v_i S_i^2 / \sigma_i^2$, tem distribuição qui-quadrado com v_i graus de liberdade, $i=1,\ldots,k$. Os $\{S_i\}$ podem (ou não) ser independentes. Porém, é assunido que Z_i é independente de S_i^2, $i = 1,\ldots,k$, de modo que

$$T_i = \frac{Z_i - \mu_i}{S_i} \quad , \quad i = 1,\ldots,k$$

tenha uma distribuição t-student com v_i graus de liberdade, $i = 1,\ldots,k$.

Suponhamos que intervalos de confiança simultâneos são exigidos para as k médias desconhecidas $\{\mu_i\}$. Seja α_i, o nível do erro tipo I para a i-ésima média μ_i; selecionado de maneira que $\alpha = \sum_{i=1}^{k} \alpha_i$, onde α é a probabilidade da razão de erro; isto é, a razão de erro experimental desejada para a família dos k intervalos de confiança. Seja $t_{\alpha_i/2,\, v_i}$ o quantil de ordem $(1-\alpha_i/2)$ da distribuição t-student com v_i graus de liberdade. Então pela desigualdade de Bonferroni temos:

$$P[z_i - t_{\alpha_i/2, v_i} S_i \leq \mu_i \leq z_i + t_{\alpha_i/2, v_i} S_i, \text{ para todo } i=1,\ldots,k]$$

$$= P[\frac{|z_i - \mu_i|}{S_i} = |t_i| \leq t_{\alpha_i/2, v_i}, \text{ para todo } i = 1,\ldots,k]$$

$$= 1 - P[|t_i| > t_{\alpha_i/2, v_i}, \text{ para pelo menos um } i, i = 1,\ldots,k]$$

$$\geq 1 - \sum_{i=1}^{k} \alpha_i \quad \text{(pela desigualdade de Bonferroni)}$$

$$= 1 - \alpha .$$

Portanto, os limites de confiança simultâneo para os $\{\mu_i\}$, os quais tem pelo menos a probabilidade $1-\alpha$ de conter cada um dos $\{\mu_i\}$, são dados por

$$z_i \pm t_{\alpha/2, v_i} S_i , \quad i = 1,\ldots,k .$$

Equivalentemente, voltando ao nosso modelo dado em (1.1.1), se são k os tratamentos sendo comparados, um intervalo de confiança para todas as $k^* = k(k-1)/2$ diferença entre médias $\mu_i - \mu_j$, $i \neq j$ simultaneamente, de coeficiente de confiança maior ou igual a $1-\alpha$, é dado por

$$\bar{y}_{i.} - \bar{y}_{j.} \pm t_{v,\frac{\alpha}{2k^*}} \; S\left(\frac{1}{n_i} + \frac{1}{n_j}\right)^{1/2}$$

onde $Var(\bar{y}_{i.} - \bar{y}_{j.})$ é estimada por $S^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$ com $v = \sum_{\ell=1}^{k} n_\ell - k$ graus de liberdade.

É importante notar que o método de Bonferroni, para comparações de todos os pares de médias, somente exige variâncias iguais e normalidade, de maneira que a distribuição t-student possa ser usada. Este método não exige tamanhos amostrais iguais.

O MÉTODO GT2.

O método GT2 é um PCM que foi proposto por Hochberg (1974) como uma generalização do método de Tukey em inferência simultânea.

O método GT2 é um procedimento conservativo cujo intervalo de confiança simultâneo, de coeficiente de confiança $1-\alpha$, para as as diferenças de médias $\mu_i - \mu_j$ é dada por:

$$\bar{y}_{i.} - \bar{y}_{j.} \pm SMM_{\alpha,k^*,v}\, S\left(\frac{1}{n_i} + \frac{1}{n_j}\right)^{1/2}$$

onde $SMM_{\alpha,k^*,v}$ é o ponto superior α da distribuição módulo máximo studentizado com parâmetros $k^* = k(k-1)/2$ e $v = \sum_{i=1}^{k} n_i - k$ graus de liberdade, onde o módulo máximo studentizado é definido por

$$SMM = \max_{1 \le i \le k} |\bar{Y}_{i.}|/S \qquad (1.4.4)$$

Tabelas para os pontos $SMM_{\alpha,k^*,v}$, com $\alpha = 0.01, 0.05, 0.1$ e 0.2 e $k = 3 \xrightarrow{(1)} 20$ estão disponíveis em Stoline e Ury (1979).

Para o método GT2 foi mostrado analiticamente a garantia do valor declarado $1-\alpha$ do coeficiente de confiança simultâneo.

O MÉTODO T' .

O método T' foi criado por Spjøtvoll e Stoline (1973). Este procedimento de comparações múltiplas tinha como objetivo estender o método T de Tukey para situações onde o interesse primário era comparar pares de médias simultaneamente em planejamento de experimentos cujos tamanhos amostrais eram não balanceados. O método

T' usa os seguintes limites de confiança no lugar de (1.4.1):

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm SAR_{\alpha,k,v}\ S(\max(n_i^{-1}, n_j^{-1}))^{1/2}$$

onde $SAR_{\alpha,k,v}$ é o ponto α da distribuição de amplitude aumentada studentizada. A estatística SAR é definida por

$$SAR = \max_{1 \leq i < j \leq k} \{|\bar{Y}_{i.}|, |\bar{Y}_{i.} - \bar{Y}_{j.}|\}/S \qquad (1.4.5)$$

Tabelas para $\alpha = 0.01, 0.05, 0.1$ e 0.2 estão disponíveis em Stoline (1978).

Spjøtvoll e Stoline (1973) mostraram analiticamente que o método T' garante o valor declarado $1-\alpha$ do coeficiente de confiança conjunto; isto é, sua razão de erro experimental é igual ou menor que α para qualquer configuração amostral.

O MÉTODO GH.

Um dos mais recentes procedimentos de comparações múltiplas foi criado por Genizi e Hochberg (1978). Tal procedimento, denominado método GH, é conservativo e pode ser usado em situações onde em um planejamento experimental com k tratamentos existem m amostras de tamanho n_1 e as remanescentes k-m são iguais a n_2. Seus limites de confiança de nível de significância α são dados por

$$\overline{Y}_{i.} - \overline{Y}_{j.} \pm SR_{\alpha,k,v}\, gS(n_i^{-1} + n_j^{-1})/2)^{1/2}$$

onde $g = 1$ se $n_i = n_j$ e $g > 1$ se $n_i \neq n_j$ e $SR_{\alpha,k,v}$ e S^2 é dado conforme (1.3.2) e (1.4.3) respectivamente. É importante observar que o método GH dá limites que são idênticos à aqueles do método de Tukey-Kramer se duas médias com o mesmo tamanho amostral são comparadas, entretanto se duas médias com diferentes tamanhos amostrais estão sendo comparadas os limites são mais abertos. Para este método existe provas analíticas da garantia do coeficien de confiança conjunto $1-\alpha$. As tabelas para a execução deste procedimento são limitados para situações onde m ou k-m é igual a 1 e estão disponíveis em Genizi e Hochberg (1978).

O MÉTODO DE GABRIEL.

Este método de comparações múltiplas proposto por Gabriel (1978) fornece intervalos de confiança para as diferenças $\mu_i - \mu_j$, $i \neq j$ da forma

$$\overline{Y}_{i.} - \overline{Y}_{j.} \pm SMM_{\alpha,k^*,v}\, S((2n_i)^{-\frac{1}{2}} + (2n_j)^{-\frac{1}{2}})$$

onde $SMM_{\alpha,k^*,v}$ é dado conforme (1.4.4). O método de Gabriel coincide com o método GT2 quando os tamanhos amostrais n_i's são iguais e apresenta certas vantagens se comparações gráficas são desejadas; Hochberg, Weiss e Hart (1982).

O MÉTODO DS.

Sidák (1967) realizou um melhoramento na desigualdade de Bonferroni. A desigualdade de Sidák foi usada por Dunn (1974) para criar um procedimento de comparações múltiplas denominada método DS. Este método é conservativo e produz intervalos de confiança simultâneos, de nível de significância menor ou igual a α, para as diferenças de médias $\mu_i - \mu_j$, $i \neq j$, dados por

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm t_{\alpha^*,v}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)^{1/2}$$

onde $\alpha^* = \frac{1}{2} - \frac{1}{2}(1-\alpha)^{1/k^*}$ e $t_{\alpha^*,v}$ é o α^*-ésimo ponto da distribuição t-student com v graus de liberdade.

Tabelas especiais para vários procedimentos tipo DS foram tabuladas por Games (1977).

O MÉTODO (H) DE HUNTER.

Este conservativo e tão complicado método foi descrito por

Hunter (1956) e por Gabriel e Begun (1977). Ele é derivado a partir de uma desigualdade de segunda ordem tipo - Bonferroni para união de eventos arbitrários e usa resultado da teoria dos grafos. Para estimar $\mu_i - \mu_j$, o intervalo de confiança, de nível $1-\alpha$, é da forma

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm (H_{\alpha,n_1,n_2,\ldots,n_k})S(n_i^{-1} + n_j^{-1})^{1/2}$$

onde $H_{\alpha,n_1,n_2,\ldots,n_k}$ é o α-ésimo ponto (não tabulado) para este procedimento. $H_{\alpha,n_1,\ldots,n_k}$ pode ser obtido iterativamente pelo uso de rotinas de probabilidades da distribuição t-univariada e t-bivariada. Procedimentos para isso são descritos em Hunter (1976, 1977) e Gabriel e Begun (1977).

CAPÍTULO II

COMPARAÇÕES DE PCM

2.1. UMA AVALIAÇÃO DO MÉTODO TK VERSUS OUTROS PCM.

O método de Tukey-Kramer, dado por (1.4.2) depende da distribuição amostral da seguinte estatística:

$$Z = \max_{i,j} \frac{|\bar{Y}_{i.} - \bar{Y}_{j.}|}{S\,[\frac{1}{2}(n_i^{-1} + n_j^{-1})]^{1/2}} \qquad (2.1.1)$$

Denotemos por $TK_{\alpha,k,v}$ o α-ésimo ponto desta estatística. Se substituíssemos $TK_{\alpha,k,v}$ no lugar de $SR_{\alpha,k,v}$ em (1.4.2) produziríamos um conjunto de intervalos de confiança para todas as diferenças $\mu_i - \mu_j$, obtendo um coeficiente de confiança conjunto exatamente igual a $1-\alpha$.

A distribuição da estatística definida em (2.1.1) não é conhecida exceto para o caso em que tem-se os tamanhos amostrais n_i's iguais, onde obtemos a distribuição da amplitude studentizada. O método TK envolve substituir $TK_{\alpha,k,v}$ por $SR_{\alpha,k,v}$; assim a validade do método depende do efeito desta substituição no coeficiente de confiança.

Em recente trabalho,Dunnett (1980a) estuda através de simulação em computador o efeito de diferentes tamanhos amostrais n_i's sobre a razão de erros de seis procedimentos de comparações múltiplas para k médias de tratamentos com variâncias homogêneas. Os métodos estudados foram o método da média harmônica proposto por Winer (1962), que substitue a média harmônica dos tamanhos amostrais em lugar de r no método T de Tukey,o método GT2 de Hochberg, o método T' de Spjøtvoll e Stoline, o método GH de Genizi e Hochberg, o método de Gabriel e o método de Tukey-Kramer.

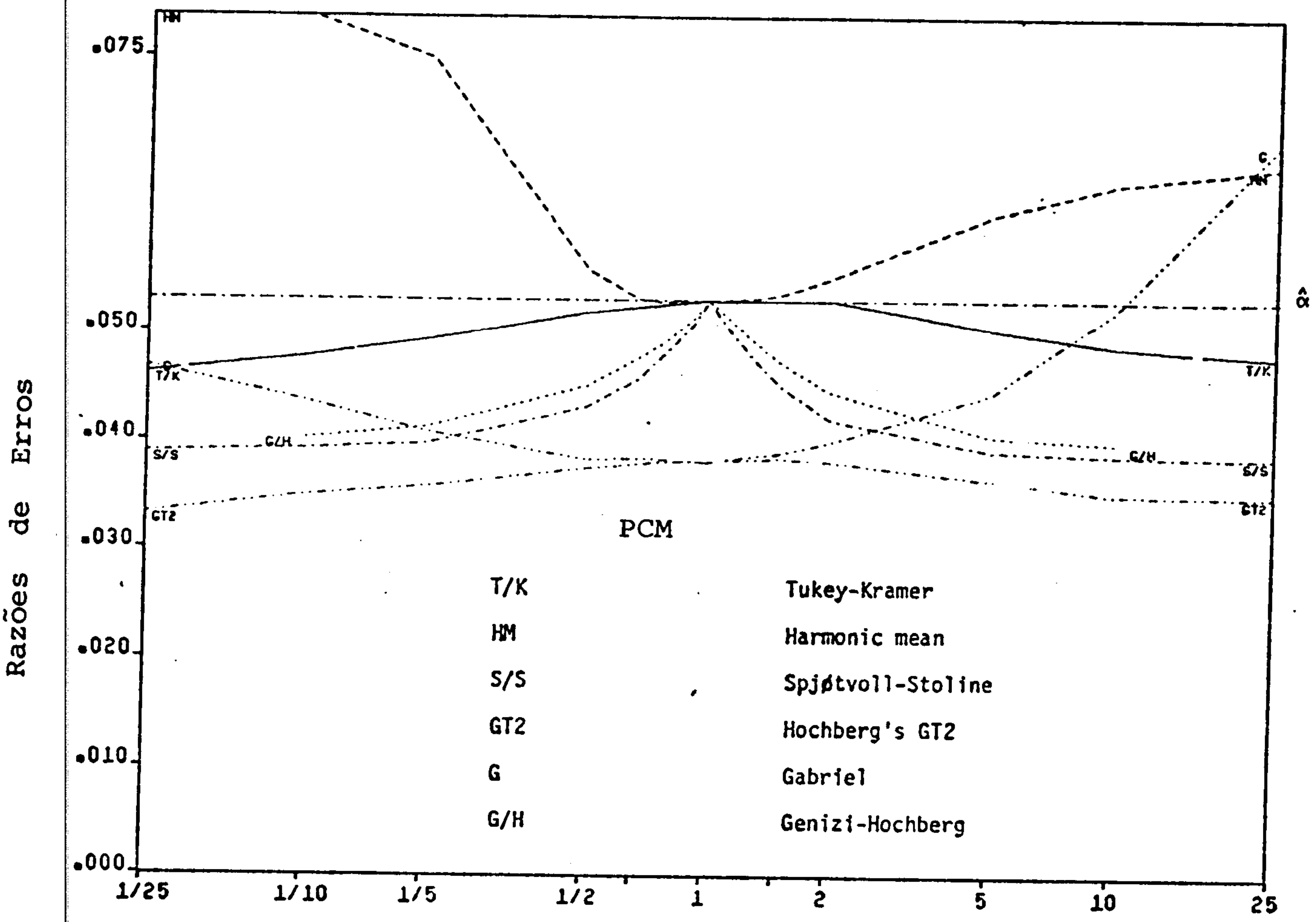
A principal preocupação de Dunnett foi voltada para estimar a razão de erro do procedimento de Tukey-Kramer, denotada por $P(F) = Pr\{Z \geq SR_{\alpha,k,v}\}$ onde Z é definido por (2.1.1). Vários valores de k, v, α e tamanhos amostrais n_i's foram escolhidos e simulações foram realizadas. As simulações também foram aplicadas aos outros cinco procedimentos de comparações múltiplas com o objetivo de comparar seus resultados com aqueles obtidos usando o método de Tukey-Kramer.

Os resultados das simulações mostraram que o método TK é aparentemente conservativo, isto é, os níveis de significância $\hat{\alpha}$ estimados foram sempre menores ou iguais ao nível α estipulado, conforme foi mostrado pelo teste do qui-quadrado de McNemar para proporções. Para pequenas variações entre os tamanhos amostrais o método TK é levemente conservativo e o conservantismo aumenta em baixa escala a medida que as variações nos tamanhos das amostras aumenta.

Os resultados relativos às comparações entre os procedimentos de comparações múltiplas mostram que o método da média harmônica tem razão de erro que torna-se excessivamente alto tão logo exista uma diferença razoável entre os n_i's, indicando assim que este método não deve ser usado nesta situação. Os métodos T' e GH são mais conservativos do que o método TK, exceto para pequenas variações amostrais onde eles são equivalentes. O método GT2 é sempre bem mais conservativo do que o método TK. Finalmente o método de Gabriel é mais conservativo do que o método TK sobre uma ampla faixa, mas sua razão de erro torna-se excessiva para extremas variações nos tamanhos amostrais.

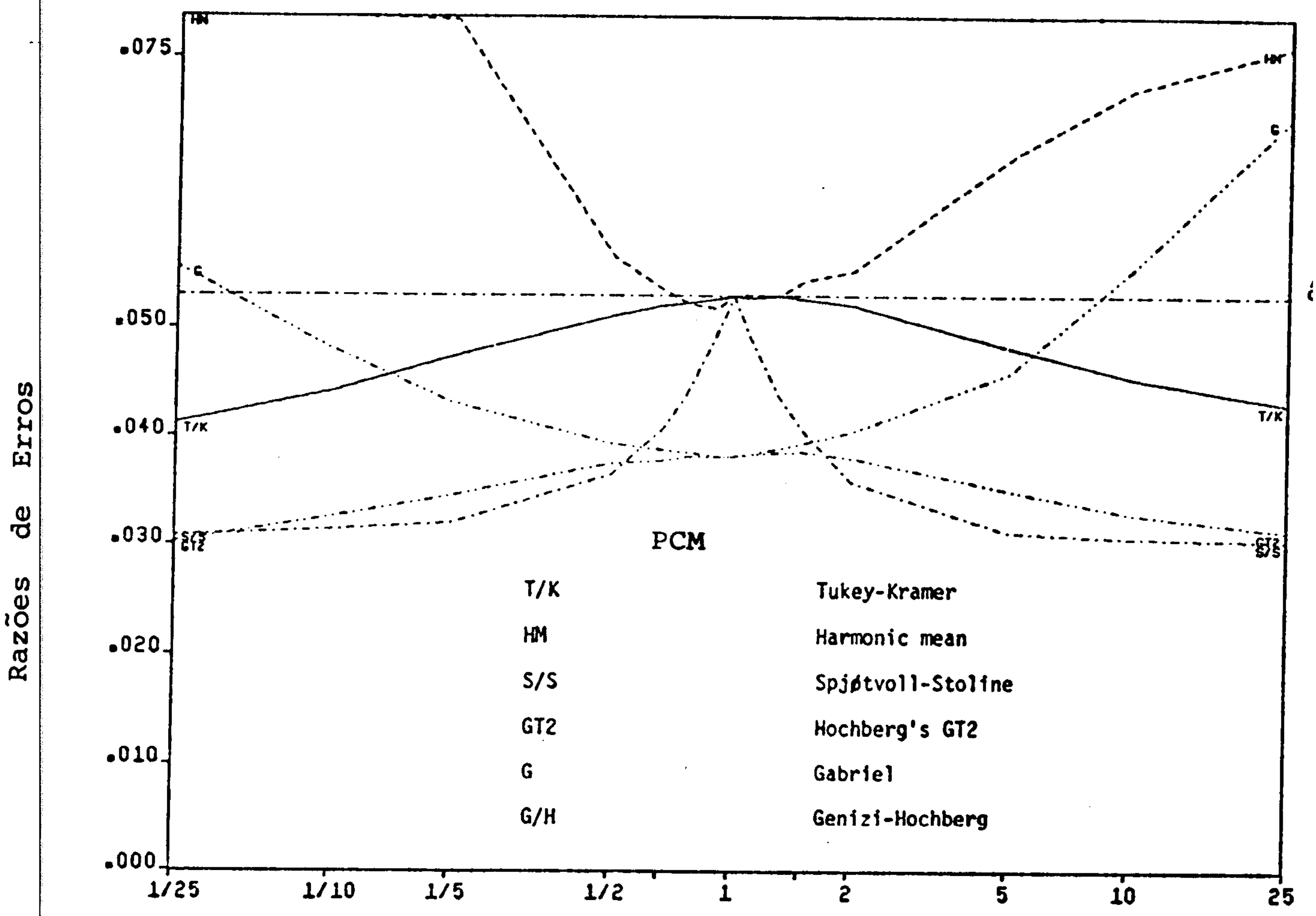
Dunnett comparou os PCM plotando suas razões de erros empíricas versus a razão de tamanho amostral $1/a_i^2$. As figuras A e B apresentam os resultados gráficos para duas situações distintas. É bom lembrar que o valor $\hat{\alpha}$ que aparece em ambas as figuras foi obtido pelo método de Tukey para configurações de amostras iguais.

Note na figura A que $\frac{1}{a_i^2} = \frac{1}{2}$ implica em $a_i^2 = 2$ e desta forma, por exemplo, temos para esse particular valor uma simulação com 5 amostras de tamanhos r e uma amostra de tamanho 2r. Assim a configuração $(a_1,1,1,1,1,1)$ indica que temos 5 amostras de tamanhos iguais a r e uma amostra de tamanho $a_1^2 r$. Para a figura B, a configuração $(a_1,a_2,1,1,1,1)$, $a_1 = a_2$, indica a situação de 4 amostras de tamanhos iguais a r e duas amostras de tamanhos $a_1^2 r$.



Razão do tamanho amostral $1/a_1^2$.

Figura A: Razões de Erros Observados Versus Razão do Tamanho Amostral para $k = 6$, $v = 20$, $\alpha = 0.05$, $(a_1, 1,1,1,1,1)$.



Razão do tamanho amostral $1/a_i^2$, $i = 1,2$.

Figura B: Razões de Erros Observados Versus Razão do Tamanho Amostral $k = 6$, $v = 20$, $\alpha = 0.05$ $(a_1, a_2, 1, 1, 1, 1)$ com $a_1 = a_2$.

Os resultados das figuras A e B indicam que o efeito de pequenas variações nos tamanhos amostrais faz com que o nível de significância da estatística de Tukey-Kramer tenha afastamento próximo a zero,na direção negativa, em relação ao valor $\hat{\alpha}$ do método T. Dunnett obteve em média o afastamento -0.00004, -0,00016, -0,00011 e -0.00039 para $\alpha = 0.01, 0.05, 0.10$ e 0.20, respectivamente. Resaltou ainda que no intervalo $1/1.3 \leq a_i^2 \leq 1.3$ o uso da estatística TK deve ser levemente conservativa. Observe que o ponto 1 no eixo da abcissa corresponde a tamanhos amostrais iguais.

Stoline (1981) desenvolveu um trabalho de comparações entre 9 métodos de comparações múltiplas, acrescentando os métodos de Scheffé, Bonferroni, Dunn-Sidãk e Hunter aos métodos citados no tralho de Dunnett (1980a) e excluindo o método da média harmônica. O critério para comparações dos métodos foram conservantismo, estreiteza dos intervalos de confiança, robustez e facilidade de uso.

Além das mesmas conclusões alcançadas por Dunnett, Stoline destaca uma variedade de comparações entre comprimentos de intervalos de confiança dos métodos TK e Bonferroni. As razões foram determinadas para valores de k (tratamentos) entre 3 e 20 e para v (graus de liberdade) entre 20 e ∞. Os resultados obtidos foram:

I.C. a $100(1-\alpha)\%$	Amplitude dos valores R(TK/B).
80%	0.86 —— 0.94
90%	0.89 —— 0.96
95%	0.91 —— 0.98
99%	0.94 —— 0.99

Claramente, os intervalos de confiança produzidos pelo método TK são significativamente mais estreitos do que aqueles produzidos pelo método de Bonferroni.

Em relação ao estudo entre o método TK e o método de Hunter verificou-se que os valores de R(TK/H) cairam no intervalo 0.93 —— 1.01. O grau de superioridade do método TK é mais notável para os casos de menor valor de $1-\alpha$ e balanceamentos ou proximamente balanceados. Somente para casos de extremo não-balanceamento o método de Hunter tende a ser levemente preferível ao método TK, apesar disso ainda nesse caso, Stoline recomenda o uso do método TK por causa do substancial esforço computacional que o método de Hunter exige.

A tabela a seguir resume os resultados obtidos por Stoline ao comparar os outros procedimentos de comparações múltiplas:

Tabela 1. Comparação de Nove Métodos de Comparações Múltiplas Usadas em Estimação de Pares de Médias - Stoline (1981).

MÉTODOS	CONSERVATIVO	LIMTAÇÕES	COMPRIM.do I.C.
Scheffé (S)	SIM	NÃO HÁ	$B \leq S$ (geralmente)
Bonferroni (B)	SIM	NÃO HÁ	
Dunn-Sidãk (DS)	SIM	NÃO HÁ	
Hochberg (GT2)	SIM	NÃO HÁ	
Gabriel (G)	Provavelmente SIM, exceto para casos de extremo não balanceamento.	NÃO HÁ	$G \leq GT2$ (sempre)
Stoline-Spjøtvoll (T')	SIM	NÃO HÁ	
Genizi-Hochberg (GH)	SIM	Tabelas limitadas p/ situações onde m ou k-m é igual a 1.	$GH \leq T'$
Hunter	SIM	Tabelas não disponíveis.	$H \leq B$ (sempre)
Tukey-Kramer (TK)	(a) SIM, para k=3,4,5 (b) provavelmente SIM para $k \geq 6$	NÃO HÁ	$TK \leq GT2 \leq DS \leq B$ (sempre) $TK \leq T'$ (sempre) $TK \leq GH$ $TK \leq H$ (geralmente)[a] $TK \leq G$ (geralmente)[a]

a : exceto para alguns casos de extremo não balanceamento.

Seria desejável, se possível, a verificação por métodos analíticos dos resultados simulados e registrados para o método de Tukey-Kramer. Kurtz (1956)*, em sua tese de doutorado, desenvolveu uma prova analítica para o caso especial de k=3 grupos de tratamentos. Brown (1979)** desenvolveu uma prova analítica do conservantismo do método TK aplicável a k=3, 4 e 5 tratamentos. Finalmente com o trabalho de Dunnett (1980a), pode ser razoável conjecturar que o método TK deve ser conservativo para todos os casos $k \geq 3$ tratamentos.

Em adição, o método TK sempre produz intervalos de confiança de menor amplitude que os métodos B, DS, GT2 e T' já que são verificadas as seguintes desigualdades:

$$\frac{SR_{\alpha,k,v}}{\sqrt{2}} \leq SMM_{\alpha,k^*,v} \leq t_{\alpha/2k^*,v}$$

e

$$SR_{\alpha,k,v}\,[(1/2((n_i^{-1}+n_j^{-1})]^{1/2} \leq SAR_{\alpha,k,v}/\,[\min(n_i,n_j)]^{1/2}$$

Finalmente podemos concluir que quando comparações multiplas em pares de médias são necessárias e as variâncias podem ser assumidas como homogêneas mas os tamanhos amostrais n_i's são diferentes, a modificação do método T proposta por Tukey e Kramer produz limites

---

(*) Tese de doutorado não publicada; (**) manuscrito não publicado. Ambos citados em Dunnett (1980a).

de confiança conservativos e por ser em geral de menor amplitude que os demais intervalos de confiança dos procedimentos de comparações múltiplas aqui mencionados, e também de maior facilidade de uso, deve ser o preferido entre tais métodos.

Desafortunadamente, existe evidência em simulações indicando que o método TK não é robusto com respeito a heterogeneidade de variâncias (Kelseman e Rogan (1978), Dunnett (1980b)) e para este caso outros métodos serão discutidos na próxima secção.

2.2. COMPARAÇÃO DE PROCEDIMENTOS DE COMPARAÇÕES MÚLTIPLAS EM PARES DE MÉDIAS COM VARIÂNCIAS DIFERENTES.

Nos últimos anos considerável atenção tem sido voltada para o problema de comparações múltiplas sobre as médias μ_i's quando as variâncias σ^2_i's são diferentes; por exemplo, ver Ury e Wiggins (1971), Spjøtvoll (1972), Brown e Forsythe (1974), Games e Howell (1976), Hochberg (1976), Tamhane (1977, 1979), Dalal (1978) e Dunnett (1980b). Destes estudos, apresentaremos quatro procedimentos que mais se adaptam ao caso de comparações múltiplas em pares.

Consideremos o modelo de análise de variância descrito em (1.1.1) onde

$$Y_{ij} = \mu + A_i + e_{ij}$$

e suponha que $e_{ij} \sim N(0,\sigma_i^2)$. Para o caso de σ_i^2's diferentes, usa-se S_i^2 para estimar σ_i^2 , $i=1,\ldots,k$ e devemos considerar intervalos de confiança na forma:

$$\bar{Y}_{i.} - \bar{Y}_{j.} \pm A_{ij,\alpha,k}(S_i^2/n_i + S_j^2/n_j)^{1/2} \qquad (2.2.1)$$

onde $A_{ij,\alpha,k}$ deve ser escolhida convenientemente de modo a obter-se o coeficiente de confiança conjunto $1-\alpha$ estipulado. Os seguintes métodos para escolher $A_{ij,\alpha,k}$ devem ser considerados:

PROCEDIMENTO DE GAMES-HOWELL.

Foi proposto por Games e Howell (1976) e toma-se

$$A_{ij,\alpha,k} = SR_{\alpha,k,\hat{v}_{ij}}/\sqrt{2}$$

onde os graus de liberdade $\hat{v}_{ij}$ são obtidos através da fórmula de Welch (1938) para graus de liberdade dada por

$$\hat{v}_{ij} = \frac{(S_i^2/n_i + S_j^2/n_j)^2}{S_i^4/n_i^2 v_i + S_j^4/n_i^2 v_j} \qquad (2.2.2)$$

onde v_i representa os graus de liberdade associado a S_i^2 .

PROCEDIMENTO C.

O procedimento C é uma extensão da solução aproximada de Cochran (1964) para o problema de Behrens-Fisher. Aqui,

$$A_{ij\alpha,k} = SR_{\alpha,k,v^*_{ij}} / \sqrt{2}$$

onde

$$SR_{\alpha,k,v^*_{ij}} = \frac{SR_{\alpha,k,v_i} S_i^2/n_i + SR_{\alpha,k,v_j} S_j^2/n_j}{S_i^2/n_i + S_j^2/n_j}$$

corresponde a média ponderada de t's de student. O procedimento C, que é uma extensão natural do método de Cochran para $k > 2$, não aparece como sendo previamente sugerido para ser usado em comparações múltiplas em pares.

PROCEDIMENTO T2.

Este procedimento foi proposto por Tamhane (1977, 1979) e propõe

$$A_{ij,\alpha,k} = t_{\gamma,\hat{v}_{ij}}$$

onde $t_{\gamma,\hat{v}_{ij}}$ é o ponto γ bi-lateral da distribuição t-student com $\hat{v}_{ij}$ graus de liberdade e,

$$\gamma = 1 - (1-\alpha)^{1/k^*} \quad , \quad k^* = k(k-1)/2$$

e $\hat{v}_{ij}$ é dado por (1.5.2).

PROCEDIMENTO T3.

Este método não aparece como sendo previamente proposto na literatura e foi introduzido por Dunnett (1980b). O procedimento T3 propõe

$$A_{ij,\alpha,k} = SMM_{\alpha,k^*,\hat{v}_{ij}}$$

onde $SMM_{\alpha,k^*,\hat{v}_{ij}}$ representa o ponto α da distribuição do módulo máximo studentizado de k^* variáveis normais não-correlacionadas com $\hat{v}_{ij}$ graus de liberdade. Tabelas para utilização deste método estão disponíveis em Stoline e Ury (1979).

2.3. COMENTÁRIOS E CONCLUSÕES.

Vários procedimentos de comparações múltiplas foram propostos na literatura, entretanto, eles não foram discutidos aqui em vista dos resultados obtidos por Tamhane (1979) o qual mostrou que na prática a escolha poderia ser restrita aos métodos T2 e Games--Howell e que o procedimento de Games-Howell fornece intervalos de confiança mais estreitos do que o procedimento T2 mas com um coeficiente de confiança conjunto que as vezes excede $1-\alpha$.

Nos trabalhos de Tamhane não foram considerados os procedimentos C e T3 e estes procedimentos foram comparados, juntamente com os procedimentos de Games-Howell e T2, por Dunnet (1980 b) o qual mostrou através de estudos simulados que os procedimentos T2, T3 e C são todos conservativos e que entre estes, T3 tem menor comprimento de intervalo de confiança para pequenos graus de liberdade e C tem menor comprimento para graus de liberdade amplos, enquanto que o procedimento de Games-Howell é levemente liberal quando as variâncias são aproximadamente iguais e tornando-se conservativo em caso contrário. Dunnett (1980b) mostrou ainda que o procedimento T3 sempre produz menor intervalo de confiança do que o procedimento T2, o mesmo ocorrendo com o procedimento de Games-Howell em relação ao procedimento C, e os métodos T2 e T3, Games-Howell e C tornam-se idênticos quando as variâncias são conhecidas.

Poderíamos então concluir que para comparações múltiplas em pares de médias, onde conhecemos de início a existência de heterogeneidade das variâncias entre os tratamentos, o procedimento T3 ou C é recomendado para situações onde os graus de liberdade são pequeno ou grande , respectivamente. O procedimento de Games-Howell é também recomendado, embora tenha-se o risco dele ser as vezes levemente liberal.

Para modelos não-balanceados, onde nada se conhece de início a respeito da natureza das variâncias populacionais, não existe ainda um método óptimo e robusto e trabalho de pesquisa neste sentido se faz necessário.

No capítulo III serão apresentados alguns resultados, obtidos através de simulações, das razões de erros produzidos pelos métodos GH, C e T3 para diferentes configurações das variâncias e estes resultados são confrontados com os resultados obtidos por Dunnett (1980b).

CAPÍTULO III

ESTABILIZANDO A VARIÂNCIA PARA USAR O MÉTODO T.

3.1. INTRODUÇÃO E OBJETIVOS.

O uso do método T para comparações múltiplas em pares de médias é, também, restrito a situações onde todas as diferenças entre pares de médias devem ter a mesma variância. Neste capítulo estaremos estudando a robustez deste PCM quando esta hipótese for violada e qual o efeito da estabilização da variância sobre a sua razão de erros.

Finalmente, alguns PCM serão apresentados como propostas de uso quando ocorrer heterogeneidade entre as variâncias.

3.2. TRANSFORMAÇÕES PARA ESTABILIZAR A VARIÂNCIA.

A necessidade de transformar os dados surge porque a variável original, ou o modelo em termos da variável original, viola uma ou mais das hipóteses padrões. As hipóteses, comumente, mais violadas são aquelas relativas a linearidade do modelo e a homogeneidade da variância dos erros.

Em muitos estudos usa-se transformações sobre as observações para estabilização da variância Scheffé (1959), Box Hunter e Hunter (1978). Ocorre com bastante frequência que σ^2 é, de alguma maneira, uma função do valor de y, ou seja, não constante. Mais ainda, é comum que a variância cresca com y.

Supondo, por exemplo, que o desvio padrão de uma variável aleatória y é proporcional a alguma potência da média, μ de y; por exemplo

$$\sigma_y = c\,\mu^{\theta} ,$$

Box Hunter e Hunter (1978) apresenta uma tabela sugerindo que se faça uma transformação de potência nos dados, do tipo

$$z = y^{\lambda}$$

com o objetivo de estabilizar as variâncias.

TABELA 0: Transformações que Estabilizam a Variância Quando σ é é proporcional a μ

Dependência de σ_y sobre μ	θ	$\lambda = 1 - \theta$	Transf. que Estab. a Variância
$\sigma = c\mu^2$	2	-1	Recíproca
$\sigma = c\mu^{3/2}$	3/2	-1/2	Recíproca de Raiz
$\sigma = c\mu$	1	0	Logaritmo
$\sigma = c\mu^{1/2}$	1/2	1/2	Raiz Quadrada
$\sigma = c$	0	1	não transforma

Existe, ainda, uma série de propostas para estabilização da variância como pode ser visto, por exemplo, em Weisberg (1980).

É comum encontrarmos situações experimentais onde os tratamentos $T_1, \ldots, T_k$ correspondem a um nível de uma variável ordenada como por exemplo dose de medicamentos, adubos, etc. Não é raro também que a variância do erro experimental é função do nível de tratamento.

Investigaremos o efeito da aplicação da transformação de variáveis para estabilizar variância no nível de significância do teste de Tukey para várias estruturas de dependência da variância com relação ao nível do tratamento e utilizando as transformações de acordo com a tabela 0.

3.3. PROCEDIMENTO NAS SIMULAÇÕES.

Nós desejamos estimar as razões de erros promovidas pelo método T antes, $\hat{\alpha}_1 = P_r(Z \geq SR_{\alpha,k,v})$, e após, $\hat{\alpha}_2 = P(T \geq SR_{\alpha,k,v})$, as transformações serem efetuadas; onde

$$Z = \max_{i,j} \frac{|\bar{Y}_{i.} - \bar{Y}_{j.}|}{S} \tag{3.3.1}$$

e

$$T = \max_{i,j} \frac{|\bar{t}_{i.} - \bar{t}_{j.}|}{S_t} \tag{3.3.2}$$

sendo que $\bar{t}_{i.}$ e S_t^2 representam a média e o estimador ANOVA de σ^2 após a transformação das observações, respectivamente.

As simulações foram obtidas pela geração de k grupos (tratamentos), cada qual com r replicações de variáveis aleatórias independentes com distribuição normal de média μ e variância σ_k^2. As variáveis aleatórias foram geradas pelo processo de Box - Muller, utilizando-se o gerador de números aleatórios RND da linguagem BASIC do DEC-SYSTEM 10.

Pela repetição do processo de simulação $N = 10.000$ vezes, foram observados uma contagem de N_1 = número de vezes que $SR_{\alpha,k,v}$ foi excedido por Z e N_2 = número de vezes que $SR_{\alpha,k,v}$ foi excedido por T e N_3 = número de vezes que $SR_{\alpha,k,v}$ foi excedido por Z e T, de tal forma a estimar a razão de erro em cada caso e detectar através do teste do χ^2 de McNemar para proporções, se havia diferença entre α_1 e α_2.

3.4. RESULTADOS DAS SIMULAÇÕES.

As tabelas a seguir, mostram diversas configurações de α, k r e σ_k^2 para efeito de investigar o comportamento do nível de significância do método T.

Razão de Erro Experimental do Método T, antes ($\hat{\alpha}_1$) e após ($\hat{\alpha}_2$) as transformações das observações (Tabelas 1 a 9)

TABELA 1: $\alpha = 0.05$, $r = 6$ (tamanho amostral)

Variâncias ($k = 1,...,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k$ #	Raiz Quadrada	0.0578	0.0575
$\sigma_k^2 = k^2$ #	Logaritmo	0.0723	0.0719
$\sigma_k^2 = k^3$	Rec. da Raiz	0.0846	0.0848
$\sigma_k^2 = k^4$ #	Recíproca	0.0958	0.0466 *

TABELA 2: $\alpha = 0.05$, $r = 10$

Variâncias ($k = 1,...,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k$ #	Raiz Quadrada	0.0577	0.0576
$\sigma_k^2 = k^2$ #	Logaritmo	0.0741	0.0745
$\sigma_k^2 = k^3$	Rec. da Raiz	0.0822	0.0809
$\sigma_k^2 = k^4$ #	Recíproca	0.0888	0.0535 *

NOTA: * $\alpha_1 \neq \alpha_2$ (diferença significante ao nível de 0.01, através do teste χ^2 de MecNemar para pares de proporções).

As observações foram geradas a partir da distribuição $N(20,\sigma_k^2)$; enquanto que as restantes da distribuição $N(100, \sigma_k^2)$.

TABELA 3 : $\alpha = 0.05$, $r = 10$

Variâncias ($k = 1,\ldots,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = 2k-1$	Raiz Quadrada	0.0638	0.0636
$\sigma_k^2 = (2k-1)^2$	Logaritmo	0.0789	0.0804
$\sigma_k^2 = (2k-1)^3$	Rec. da Raiz	0.0841	0.0875
$\sigma_k^2 = (2k-1)^4$	Recíproca	0.0850	0.0660 *

TABELA 4 : $\alpha = 0.05$, $r = 10$

Variâncias ($k = 1,\ldots,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k/2$	Raiz Quadrada	0.0606	0.0605
$\sigma_k^2 = (k/2)^2$	Logaritmo	0.0720	0.0718
$\sigma_k^2 = (k/2)^3$	Rec. da Raiz	0.0836	0.0842
$\sigma_k^2 = (k/2)^4$	Recíproca	0.0810	0.0818

TABELA 5 : $\alpha = 0.05$, $r = 10$

Variâncias $(k = 1,\dots, 6)$	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k$	Raiz Quadrada	0.0702	0.0697
$\sigma_k^2 = k^2$	Logaritmo	0.0919	0.0922
$\sigma_k^2 = k^3$	Rec. da Raiz	0.1114	0.1132
$\sigma_k^2 = k^4$	Recíproca	0.1200	0.1202

TABELA 6 : $\alpha = 0.10$, $r = 10$

Variâncias $(k = 1,\dots, 6)$	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k$	Raiz Quadrada	0.1171	0.1171
$\sigma_k^2 = k^2$	Logaritmo	0.1467	0.1473
$\sigma_k^2 = k^3$	Rec. da Raiz	0.1623	0.1690
$\sigma_k^2 = k^4$	Recíproca	0.1726	0.1927 *

TABELA 7: $\alpha = 0.05$, $r = 10$

Variâncias ($k = 1,...,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = (k/2)^4$	Recíproca	0.0828	0.0836
$\sigma_k^2 = k^4$	Recíproca	0.0869	0.0897
$\sigma_k^2 = k^4$ #	Recíproca	0.0871	0.0597 *

TABELA 8 : $\alpha = 0.10$, $r = 6$

Variâncias ($k = 1,...,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k$ #	Raiz Quadrada	0.1099	0.1108
$\sigma_k^2 = k^2$ #	Logaritmo	0.1218	0.1206
$\sigma_k^2 = k^3$	Rec. da Raiz	0.1357	0.1380
$\sigma_k^2 = k^4$ #	Recíproca	0.1440	0.0924 *

TABELA 9 : $\alpha = 0.10$, $r = 10$

Variâncias ($k = 1,\ldots,4$)	Transformações	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\sigma_k^2 = k$ #	Raiz Quadrada	0.1047	0.1060
$\sigma_k^2 = k^2$	Logaritmo	0.1184	0.1205
$\sigma_k^2 = k^3$ #	Rec. da Raiz	0.1279	0.1301
$\sigma_k^2 = k^4$ #	Recíproca	0.1241	0.1033 *

3.5. DISCUSSÃO DOS RESULTADOS.

Os resultados apresentados nas tabelas 1 a 9 indicam, conclusivamente, que as probabilidades amostrais para a estatística de Tukey (3.3.1) exceder $SR_{\alpha,k,v}$ afasta-se, no sentido positivo, do valor α estipulado se a condição de iguais variâncias não for satisfeita. Tal acréscimo torna-se cada vez maior na medida que aumenta o expoente θ ou o número de tratamentos.

O teste do qui-quadrado de McNemar para proporções, ao nível de significância $\alpha = 0.01$, indicou que, após efetuada as transformações raiz quadrada, logaritmo e recíproca da raiz quadrada para estabilização das variâncias, o nível de significância do método T não foi melhorado, ou seja, as razões de erros, permaneceram praticamente as mesmas em relação às obtidas com as observações originais.

Para alguns casos a transformação recíproca das observações altera a razão de erros do método T. Esta alteração está relacionada com o tamanho amostral, desigualdade das variâncias e magnitude das médias. Um estudo mais detalhado neste sentido não se faz necessário já que esta transformação não torna, na maioria das vezes, o nível de significância conservativo.

É importante lembrar que ao usar a transformação para estabilizar a variância podemos estar alterando as médias e, então, ao continuarmos com a comparação múltipla cometeremos erros fazendo declarações sem sentido, ou seja, estaremos testando algo que não seja a hipótese nula $H_o : \mu_i = \mu_j$, $i \neq j$.

Se a estrutura da variância for da forma $\sigma = c\mu^\theta$ temos, então, uma idéia da natureza do vício de α promovido pelo método T. Observe que se $k = 4$ e $\theta = 1/2$ o vício b de α fica em torno de 0.01, $\theta = 1$ $b \approx 0.02$, $\theta = \frac{3}{2}$ $b \approx 0.03$ e $\theta = 2$ $b \approx 0.04$. Talvez, num estudo mais detalhado, consiga-se com melhor precisão o vício para cada configuração de $c\mu^\theta$ e desta forma poderia surgir, por exemplo, propostas do tipo aplicar o método T a um nível

$\alpha = 0.05$ para garantir-se um nível $\alpha = 0.10$. Fica aqui a sugestão.

Nosso propósito principal, neste capítulo, não era estudar o vício do método T, pois isto já havia sido feito por vários pesquisadores Ramseyer e Tcheng (1973) e Brown (1974). De qualquer forma fica aqui a confirmação da não robustez do método T na presença de variâncias heterogêneas. Quanto a estabilização de variância para aplicação do método T, em situações semelhantes a nossa investigação, é desaconselhável, já que não estaremos melhorando o nível de significância.

3.6. RESULTADOS DE SIMULAÇÕES DE PCM COM VARIÂNCIAS HETEROGENEAS.

Na secção 2.2 foram discutidos os resultados das pesquisas de Dunnett (1980b), referentes a PCM para o caso de variâncias heterogênas. Os resultados obtidos na secção 3.4 indicam, claramente , que o método T inflaciona o nível de significância α se as variâncias não forem constantes. Nesta secção estaremos apresentando as razões de erros, obtidas através de simulações, dos procedimentos GH, C e T3 para algumas configurações de σ_k^2 , $k = 1,...,4$ e $\alpha = 0.05$ e, esses resultados serão confrontados com os resultados obtidos por Dunnett (1980b).

O procedimento nas simulações foi semelhantes ao da secção 3.3, tendo como diferença a inserção dos graus de liberdade no programa. Para calcular os pontos α da amplitude studendizada necessários na aplicação dos métodos GH e C foram usados os valores tabulados em Harter (1960) com imterpolação linear sob o grau de liberdade necessário para aplicar a fórmula de Welch (1.5.2). O mesmo foi feito para calcular os valores α da distribuição módulo máximo studentizado, usando os valores tabulados em Stoline e Ury (1979), indispensáveis para o uso do método T3.

As tabelas 10 e 11 exibem os resultados de nossas simulações.

TABELA 10: Razões de Erros Estimados ao comparar os $\binom{k}{2}$ pares de médias. Nível de Significância $\alpha = 0.05$, $k = 4$, $r = 6$ e $N = 20.000$ simulações.

Variância $(k=1,...,4)$	PCM	$\hat{\alpha}$	I.C. a 95%
	GH	0.0511	[0.0480, 0.0542]
$\sigma_k^2 = 1$	C	0.0441	[0.0413, 0.0469]
	T3	0.0451	[0.0422, 0.0480]
	GH	0.0526	[0.0495, 0.0557]
$\sigma_k^2 = k$	C	0.0463	[0.0434, 0.0492]
	T3	0.0464	[0.0435, 0.0493]
	GH	0.0563 *	[0.0531, 0.0595]
$\sigma_k^2 = k^2$	C	0.0510	[0.0481, 0.0543]
	T3	0.0505	[0.0475, 0.0535]
	GH	0.0591 *	[0.0558, 0.0624]
$\sigma_k^2 = k^3$	C	0.0557	[0.0525, 0.0589]
	T3	0.0531	[0.0500, 0.0562]
	GH	0.0606 *	[0.0573, 0.0639]
$\sigma_k^2 = k^4$	C	0.0582 *	[0.0550, 0.0614]
	T3	0.0544	[0.0513, 0.0575]

* Provavelmente, se as simulações tivessem sido conduzidas de tal forma a testar a igualdade de α contra o nível de significância α_t produzido pelo método T, somente os valores assinalados com * seriam significativas a 5%, pelo teste de McNemar.

TABELA 11. Razões de Erros Estimados ao comparar os $\binom{k}{2}$ pares de médias. Nível de Significância $\alpha = 0.05$, $k=4$, $n = 10$, $N = 10.000$ simulações.

Variância ($k=1,\ldots,4$)	PCM	$\hat{\alpha}$
	GH	0.0426
$\sigma_k^2 = 1$	C	0.0369
	T3	0.0362
	GH	0.0407
$\sigma_k^2 = k$	C	0.0365
	T3	0.0351
	GH	0.0399
$\sigma_k^2 = k^2$	C	0.0362
	T3	0.0346
	GH	0.0399
$\sigma_k^2 = k^3$	C	0.0378
	T3	0.0354
	GH	0.0410
$\sigma_k^2 = k^4$	C	0.0395
	T3	0.0367

3.7. DISCUSSÃO DA RAZÃO DE ERROS PROMOVIDAS PELOS PROCEDIMENTOS GH, C E T3.

Com base nos resultados mostrados nas tabelas 10 e 11 pode-se ter muma idéia do comportamento da razão de erros promovidas pelos procedimentos GH, C e T3.

O procedimento GH é levemente liberal quando os graus de liberdade são pequenos, esta liberalidade aumenta em pequena proporção quando as variâncias tornam-se mais e mais divergentes. Para graus de liberdade grande o procedimento GH torna-se conservativo.

O procedimento C é conservativo. Embora tenhamos observado uma pequena liberalidade (*) deste procedimento, para $\sigma_k^2 = k^4$, $k = 1,\ldots,4$ e $r = 6$, este caso de extrema heterogeneidade, dificilmente ocorre, já que na prática os valores de n_i devem, provavelmente, serem escolhidos de modo que tenhamos os valores de σ_i^2/n_i aproximadamente os mesmos em cada uma das k amostras.

O procedimento T3 é sempre conservativo.

Dunnett (1980b) declara que os procedimentos C e GH tornam-se idênticos para graus de liberdade elevados.

Os resultados aqui obtidos estão em concordância com os resultados obtidos por Dunnett (1980b). Vale a pena ressaltar que a possível liberalidade do procedimento C aqui encontrada não foi destacada por Dunnett (1980b) e que em seus estudos foram tomadas diversas configurações dos tamanhos amostrais, inclusive, amostras de tamanhos diferentes.

Finalmente, poderíamos concluir com base em nossos resultados e nos de Dunnett (1980b) que o procedimento C é preferível quando os graus de liberdade são grande ou moderadamente grande. Porém, para graus de liberdade pequeno o procedimento C torna-se mais conservativo que T3 e desta forma o procedimento T3 deve ser preferido por estar próximo do nível α estipulado.

O procedimento GH é também recomendado embora se tenha o risco dele ser levamente liberal para pequenos graus de liberdade.

CAPÍTULO IV

COMPARAÇÃO MÚLTIPLA EM PARES DE MÉDIAS SOB A PRESENÇA DE CORRELAÇÃO ENTRE OS ERROS.

4.1. INTRODUÇÃO E OBJETIVOS.

Nos últimos tempos, enquanto muitos pesquisadores preocupavam-se com PCM em pares de médias levando em consideração o tamanho amostral, homogeneidade e/ou heterogeneidade das variâncias, surpreendentemente o papel da matriz de covariância foi quase que totalmente ignorado.

Uma das suposições básicas da análise de variância é a não existência de correlação entre os erros, entretanto, na prática, esta suposição pode ser violada.

Scheffé (1959) mostrou que a existência de uma correlação serial entre as observações afeta seriamente o nível de significância do teste F. Neste capítulo estaremos interessados em investigar a robustez do método de Tukey para comparação múltipla em pares de médias na presença de uma correlação serial constante entre as observações de um mesmo tratamento.

4.2. A QUESTÃO DO MODELO.

Conforme foi discutido anteriormente, vamos supor que a hipótese de independência não seja satisfeita devido à existência de uma correlação serial ρ, constante entre as observações de um mesmo tratamento, ou seja, a covariância entre e_{ij} e $e_{ij'}$ é da forma:

$$\operatorname{cov}(e_{ij}, e_{ij'}) = \rho\sigma^2 \quad , \quad j \neq j'$$

enquanto que

$$\operatorname{cov}(e_{ij}, e_{i'j}) = 0 \quad , \quad i \neq i' .$$

Isto pode ser expresso formalmente supondo que a matriz de covariância do erro no modelo (1.1.1) tem a seguinte estrutura:

$$\underset{n\times n}{\Sigma} = \begin{vmatrix} V & \phi & \cdots & \phi \\ \phi & V & \cdots & \phi \\ \cdot & \cdot & \cdots & \cdot \\ \phi & \phi & \cdots & V \end{vmatrix}$$

onde

$$\underset{r\times r}{V} = \begin{vmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \cdot & \cdot & \cdots & \cdot \\ \rho\sigma^2 & \rho\sigma^2 & & \sigma^2 \end{vmatrix}$$

sendo ϕ a matriz nula de ordem r.

4.3. OS PROBLEMAS NA SIMULAÇÃO.

Nas simulações, os componentes dos $Y_i = (y_{i1},\ldots,y_{ir})$, $i = 1,\ldots,k$ são variáveis aleatórias independentes normalmente distribuídas com média zero e variância um e, foram geradas segundo o método de Box-Muller, Newman e Odel (1971); foi usado para isto o gerador de números aleatórios RND (linguagem BASIC) do DEC-SYSTEM-10.

Para introduzir a correlação constante dentro de cada tratamento foi considerado

$$-\frac{1}{r-1} \leq \rho \leq 1 \qquad (4.3.1)$$

onde r é o número de replicações em cada tratamento. O Teorema Espectral nos garante que se A é uma matriz simétrica de dimensão $n \times n$, existe uma matriz ortogonal P tal que

$$P'AP = D = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

onde os λ_i's são os auto-valores de A e as colunas de P são autovetores de A associados aos λ_i's.

Então,

$$PP'VPP' = PDP'$$

logo

$$V = PDP' \quad \text{pois} \quad PP' = I.$$

Como V é positiva definida, tomando-se ρ segundo (4.3.1), então $\lambda_i > 0$, $\forall i$, e existe $D^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r})$ tal que podemos escrever

$$P = PD^{1/2}D^{1/2'}P' = PD^{1/2}(PD^{1/2})' .$$

Assim,

$$V = V^{1/2}V^{1/2'} \text{ , onde } V^{1/2} = PD^{1/2} .$$

Tomando-se $Y_i \sim N(\phi, I)$, para $X_i \sim V^{1/2}Y_i$ tem-se

$$X_i \sim N(\phi, V) \text{ , } i = 1,\dots,k .$$

Para a obtenção das matrizes P e D tal que $V^{1/2} = PD^{1/2}$ algum esforço algébrico se faz necessário. Pode-se observar que a matriz V pode ser colocada na forma $aI + bJ$, onde I é a matriz identidade de ordem r e J é a matriz de ordem r cujos elementos são todos iguais a um. Assim,

$$\underset{r\times r}{V} = \begin{vmatrix} a+b & b & \dots & b \\ b & a+b & \dots & b \\ \cdot & \cdot & \cdot & \cdot \\ b & b & & a+b \end{vmatrix}$$

onde $a = \sigma^2(1-\rho)$ e $b = \rho\sigma^2$.

Determinemos agora a matriz $D = \text{diag}(\lambda_1,\dots,\lambda_r)$ tal que os λ_i's sejam auto-valores da matriz V. Inicialmente, adicionamos à primeira coluna de V todas as demais colunas e nós teremos:

$$\begin{vmatrix} a+rb & b & \dots & b \\ a+rb & a+b & \dots & b \\ \cdot & \cdot & \cdot & \cdot \\ a+rb & b & & a+b \end{vmatrix}$$

Nosso determinante é então $(a+rb)$ vezes o determinante de

$$\begin{vmatrix} 1 & b & \dots & b \\ 1 & a+b & \dots & b \\ \cdot & \cdot & \cdot & \cdot \\ 1 & b & & a+b \end{vmatrix}$$

Desta, subtraímos b vezes a primeira coluna de cada uma das outras colunas. O determinante não muda e nós teremos

$$\det \begin{vmatrix} 1 & 0 & \dots & 0 \\ 1 & a & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \dots & a \end{vmatrix} = a^{r-1}$$

Assim o $\det V = (a + rb)\, a^{r-1}$. Então,

$$\det(V - \lambda I) = \det(aI + bJ + \lambda I)$$

$$= \det[(a-\lambda)I + bJ)]$$

$$= (a+rb-\lambda)(a-\lambda)^{r-1}.$$

Então $a+rb$ é autovalor de V de multiplicidade 1 e a é autovalor de V de multiplicidade $(r-1)$.

Portanto, a matriz D formada pelos autovalores de V será da forma:

$$\underset{r\times r}{D} = \begin{vmatrix} a+rb & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & & a \end{vmatrix}$$

Assim $k\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{r\times 1}$ é o autovetor associado ao autovalor $a+rb$.

A primeira coluna de P é então formada pelo vetor $\frac{1}{\sqrt{r}}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{r\times 1}$

e as outras colunas de P devem formar uma base ortogonal para o subespaço dos autovetores associados ao autovalor a de multiplicidade $(r-1)$.

Pode-se mostrar que P da forma:

$$
\underset{r\times r}{P} = \begin{bmatrix}
\frac{1}{\sqrt{r}} & -\frac{r-1}{\sqrt{(r-1)r}} & 0 & 0 & \cdots & 0 \\
\frac{1}{\sqrt{r}} & \frac{1}{\sqrt{(r-1)r}} & -\frac{r-2}{\sqrt{(r-2)(r-1)}} & 0 & \cdots & 0 \\
\frac{1}{\sqrt{r}} & \frac{1}{\sqrt{(r-1)r}} & \frac{1}{\sqrt{(r-2)(r-1)}} & -\frac{r-3}{\sqrt{(r-3)(r-2)}} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{\sqrt{r}} & \frac{1}{\sqrt{(r-1)r}} & \frac{1}{\sqrt{(r-2)(r-1)}} & \frac{1}{\sqrt{(r-3)(r-2)}} & \cdots & -\frac{1}{\sqrt{2}}
\end{bmatrix}
$$

4.4. RESULTADOS OBSERVADOS NAS SIMULAÇÕES.

Para investigar o efeito da presença de correlação constante ρ dentro de cada tratamento sobre a razão de erro do método de Tukey para comparações múltiplas em pares, foram consideradas várias configurações dos k tratamentos e r replicações para os níveis de significância $\alpha = 0.05$ (tabela 1) e $\alpha = 0.1$ (tabela 2). Para cada configuração foram realizadas $N = 10.000$ simulações.

A coluna 3 de ambas as tabelas mostra as razões de erros estimada $\hat{\alpha}_1$ ao aplicar o método T em tratamentos cujas repetições são independentes enquanto que a coluna 5 mostra as razões de erro $\hat{\alpha}_2$ na presença de correlação constante dentro de cada tratamento.

O teste χ^2 de McNemar para pares de proporções foi aplicado para testar $H_o : \alpha_1 = \alpha_2$, isto é, para verificar se a correlação constante altera o nível de significância.

TABELA 1: Razões de Erros estimadas ao se comparar as $\binom{k}{2}$ pares de média pelo método de Tukey ao nível de significância $\alpha = 0.05$.

Nº de Níveis k	Nº de repetições r	$\hat{\alpha}_1$ para $\rho = 0$	coeficiente de correlação ρ	$\hat{\alpha}_2$ para $\rho \neq 0$
4	6	0.0474	-0.2	0.0542 **
4	6	0.0546	-0.1	0.0535
4	6	0.0487	0.1	0.0492
4	6	0.0465	0.2	0.0467
4	6	0.0478	0.3	0.0495
4	6	0.0471	0.4	0.0513
4	6	0.0465	0.5	0.0489
4	6	0.0525	0.6	0.0508
4	6	0.0516	0.7	0.0507
4	6	0.0453	0.8	0.0527 **
4	6	0.0478	0.9	0.0456
4	10	0.0523	0.1	0.0492
4	10	0.0508	0.2	0.0523
4	10	0.0458	0.3	0.0513
4	10	0.0541	0.4	0.0523
4	10	0.0524	0.5	0.0520
4	10	0.0510	0.6	0.0546
4	10	0.0533	0.7	0.0502
4	10	0.0523	0.8	0.0501
4	10	0.0464	0.9	0.0486
6	10	0.0477	0.1	0.0541 **
6	10	0.0496	0.2	0.0516
6	10	0.0546	0.3	0.0492
6	10	0.0544	0.4	0.0514
6	10	0.0493	0.5	0.0486
6	10	0.0532	0.6	0.0516
6	10	0.0510	0.7	0.0457
6	10	0.0489	0.8	0.0486
6	10	0.0525	0.9	0.0501

* vide rodapé da tabela 2.

TABELA 2: Razões de Erros estimadas ao se comparar as $\binom{k}{2}$ pares de média pelo método de Tukey ao nível de significância $\alpha = 0.1$.

Nº de Níveis k	Nº de repetições r	$\hat{\alpha}_1$ para $\rho = 0$	coeficiente de correlação ρ	$\hat{\alpha}_2$ para $\rho \neq 0$
4	6	0.0995	0.1	0.1015
4	6	0.0944	0.2	0.1065 *
4	6	0.1011	0.3	0.0988
4	6	0.0952	0.4	0.1057 **
4	6	0.0994	0.5	0.1025
4	6	0.0958	0.6	0.1003
4	6	0.0960	0.7	0.1032
4	6	0.0989	0.8	0.1026
4	6	0.0989	0.9	0.0977
4	10	0.0983	0.1	0.0955
4	10	0.0979	0.2	0.1002
4	10	0.1047	0.3	0.1015
4	10	0.1003	0.4	0.1036
4	10	0.0980	0.5	0.1031
4	10	0.0989	0.6	0.0993
4	10	0.0946	0.7	0.0987
4	10	0.0974	0.8	0.1000
4	10	0.1034	0.9	0.0971
6	10	0.1032	0.1	0.1035
6	10	0.1030	0.2	0.1013
6	10	0.1013	0.3	0.0968
6	10	0.1040	0.4	0.1007
6	10	0.1024	0.5	0.1040
6	10	0.1020	0.6	0.0947
6	10	0.1017	0.7	0.0991
6	10	0.1050	0.8	0.1002
6	10	0.1059	0.9	0.0978
6	10	0.0974	-0.1	0.1029

(*) e (**) Indicam que α_1 difere de α_2 ao nível de significância de 1% e 5%, respectivamente. Foi aplicado o teste do χ^2 McNemar para pares de proporções.

4.5. DISCUSSÃO DOS RESULTADOS.

A tabelas 3 e 4 mostram os desvios padrões e médias das razões de erros do método T apresentados nas tabelas 1 e 2.

TABELA 3: Desvios Padrões $(\alpha(1-\alpha)/10.000)^{1/2}$

nível de signif.	Desvio padrão teórico	Desvio padrão observado (coluna 3)	Desvio padrão observado (coluna 5)
0.05	0.00218	0.00294	0.00233
0.10	0.00300	0.00328	0.00294

TABELA 4: Médias das Razões de Erros.

esperança da média	média observada (coluna 3)	média observada (coluna 5)
0.0500	0.05018	0.05050
0.1000	0.09995	0.10067

Podemos observar através das tabelas 3 e 4 uma clara evidência da boa concordância entre os valores teóricos e observados, e com base nestes resultados algumas conclusões são relevantes; como por exemplo, a razão de erro que se comete ao usar o método de Tukey para comparações múltiplas em pares de média quando os tratamentos apresentam uma correlação constante ρ são na maioria das vezes compatíveis com o nível de significância estipulado α. As poucas vezes em que o nível de significância foi alterado (*),(**), verificou-se um pequeno inflacionamento do nível α e que é pouco significante estatisticamente.

A conclusão relativa a presente investigação é que quando comparações múltiplas em pares de médias são necessárias e que tenhamos tratamentos com um mesmo número de replicações e as observações são assumidas como tendo uma correlação constante dentro dos tratamentos, o método de Tukey da limites de confiança compatíveis com o valor α estipulado e pode ser escolhido.

REFERÊNCIAS

BAILEY, B.J.R. (1977). Tables of the Bonferroni t Statistics. Journal of the American Statistical Association, 72, 469-477.

BISHOP, Y.M.M.; FIENBERG, S.E. and HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Mass., The MIT Press, Cambridge.

BOX, G.E.P.; HUNTER, W.G. and HUNTER, J.S. (1978). *Statistics for Experimenters*. Wiley, New York.

BROWN, R.A. (1974). Robustness of the Studentized Range Statistic. Biometrika, 61, 171-175.

BROWN, M.B. and FORYSTHE, A.B. (1974). The ANOVA and Multiple Comparisons for Data With Heterogeneous Variances. Biometrics, 30, 719-724.

CARMER, S.G. and WALKER (1982). Baby Bear's Dilema: A Statistical Tale. Agronomy Journal, 74, 122-124.

COCHRAN, W.G. (1964). Approximate Significance Levels of the Behrens-Fisher Test. Biometrics, 20, 191-195.

COCHRAN, W.G. and COX, G.M. (1957). *Experimental Design.* Wiley, New York.

DALAL, S.R. (1978). Simultaneous Confidence Procedures, For Univariate and Multivariate Behrens-Fisher Type Problems. Biometrika, 65, 221-224.

DUNNETT, C.W. (1980a). Pairwise Multiple Comparisons in the Homogeneous Variance, Unequal Sample Size Case. Journal of the American Statistical Association, 75, 789-795.

DUNNETT, C.W. (1980b). Pairwise Multiple Comparisons in the Unequal Variance Case. Journal of the American Statistical Association, 75, 796-800.

GABRIEL, K.R. (1978). A Simple Method of Multiple Comparisons of Means. Journal of the American Statistical Association, 73, 724-729.

GAMES, P.A. and HOWEL, J.F. (1976). Pairwise Multiple Comparisons Procedures With Unequal N's and/or Variances: A Monte Carlo Study. Journal of Educational Statistics, 1, 113-125.

GENIZI, A. and HOCHBERG, Y. (1978). On Improved Extensions of the T-Method of Multiple Comparisons for Unbalanced Designs. Journal of the American Statistical Association, 73, 879-884.

HARTER, H.L. (1960). Tables of Range and Studentized Range. Annals of Mathematical Statistics, 31, 1122-1147.

HOCHBERG, Y. (1976). A Modification of the T-Method of Multiple Comparisons for a ONE-WAY Layout With Unequal Variances. Journal of the American Statistical Association, 71, 200-203.

HOCHBERG, Y. (1974). Some Generalizations of the T-Method in Simultaneous Inference. Journal of Multivariate Analysis, 4, 224-234.

HOCHBERG, Y., WEISS; G. and HART, S. (1982). On Grafical Procedures for Multiple Comparisons. Journal of the Amercian Statistical Association, 77, 767-772.

KESELMAN, H.J. and ROGAN , J.C. (1978). A Comparison of the Modified-Tukey and Scheffé methods of Multiple Comparisons for Pairwise Contrasts. Journal of the American Statistical Association, 73, 47-52.

KRAMER, C.Y. (1956). Extensions of Multiple Range Tests to Group Means With Unequal Numbers of Replications. Biometrics, 12, 307-310.

MILLER, R.G. (1981). *Simultaneous Statistical Inference*. Mc Graw-Hill, New York.

NEWMANN, T.G. and ODELL, L.P. (1971). *The Generation of Randon Variates*. Charles Griffin & Company Limited, London.

RAMSEYER, G.C. and TCHENG, T.K. (1973). The Robustness of the Studentized Range Statistics to Violations of the Normality and Homogeneity of Variance Assumptions. American Educational Research Journal, 10, 235-240.

SCHEFFÊ, H. (1953). A Method for Judging All Contrasts in the Analysis of Variance. Biometrika, 40, 87-104.

SCHEFFÊ, H. (1965). *The Analysis of Variance*. Wiley, New York.

SIDǍK, ZBYNEK (1967). Rectangular Confidence Regions for Means of Multivariate Normal Distributions. Journal of the American Statistical Association, 62, 626-633.

SPJØTVOLL, E. (1972). Joint Confidence Intervals for All Linear Functions of Means in ANOVA with Unknown Variances. Biometrika, 59, 684-685.

SPJØVOLL, E. and STOLINE, M.R. (1973). An Extension of the T-Method of Multiple Comparison to Include the Cases with Unequal Sample Sizes. Journal of the American Statistical Association, 68, 975-978.

STOLINE, M.R. (1978). Tables of the Studentized Augmented Range and Applications to Problems of Multiple. Comparisons. Journal of the American Statistical Association, 73, 656-660.

STOLINE, M.R. (1981). The Status of Multiple Comparisons: Simultaneous Estimation of All Pairwise Comparisons in One-Way ANOVA Designs. The American Statistician, 35, 134-141.

STOLINE, R.M. and URY, H.K. (1979). Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means. Tecnometrics, 21, 87-93.

TAMHANE, A.C. (1979). A Comparison of Procedures for Multiple Comparisons of Means With Unequal Variances. Journal of the American Statistical Association, 74, 471-480.

TAMHANE, A.C. (1977). Multiple Comparison in Model I Way ANOVA With Unequal Variances. Communications in Statistics, A6(1), 15--32.

URY, H.K. and WIGGINS, A.D. (1971). Large Sample and Other Multiple Comparisons Among Means. British Journal of Mathematical and Statistical Psychology, 24, 174-194.

WEISBERG, S. (1980). *Applied Linear Regression*, Wiley, New York.

WINER, B.J. (1962). *Statistical Principles in Experimental Design.* McGraw-Hill, New York.