UNIVERSIDADE FEDERAL DE LAVRAS

PODER E TAXAS DE ERRO TIPO I DE QUATRO CRITÉRIOS MULTIVARIADOS PARA O TESTE DE IGUALDADE DE EFEITOS DE TRATAMENTOS AVALIADOS POR MEIO DO MÉTODO DE MONTE CARLO

DILETA CECCHETTI

1999

DILETA CECCHETTI

PODER E TAXAS DE ERRO TIPO I DE QUATRO CRITÉRIOS MULTIVARIADOS PARA O TESTE DE IGUALDADE DE EFEITOS DE TRATAMENTOS AVALIADOS POR MEIO DO MÉTODO DE MONTE CARLO

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração: Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Orientador
Dr. Daniel Furtado Ferreira

LAVRAS
MINAS GERAIS - BRASIL
1999

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

Cecchetti, Dileta

Poder e taxas de erro tipo I de quatro critérios multivariados para o teste de igualdade de efeitos de tratamentos avaliados por meio do método de Monte Carlo / Dileta Cecchetti. – Lavras : UFLA, 1999.

56 p. il.

Orientador: Daniel Furtado Ferreira.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Simulação. 2. Wilks. 3. Hotelling. 4. Pillai. 5. Roy. 6. Taxa de erro tipo I. 7. Poder. I. Universidade Federal de Lavras. II. Título.

CDD-551

-519.282

DILETA CECCHETTI

PODER E TAXAS DE ERRO TIPO I DE QUATRO CRITÉRIOS MULTIVARIADOS PARA O TESTE DE IGUALDADE DE EFEITOS DE TRATAMENTOS AVALIADOS POR MEIO DO MÉTODO DE MONTE CARLO

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

APROVADA em 10 de fevereiro de 1999.

Prof. Dr. Augusto Ramalho de Morais UFLA

Prof. Dr. Julio Silvio de Sousa Bueno Filho UFLA

Prof. Dr. Daniel Furtado Ferreira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

"É fácil mentir usando a estatística,
difícil é falar a verdade sem ela".

Andrejs Dunkels

"Nós não fomos capazes de responder a todos os nossos problemas. Na verdade, não respondemos completamente a nenhum deles. As respostas que obtivemos serviram apenas para levantar um novo conjunto de questões.
De alguma maneira, nos sentimos tão confusos quanto antes, mas pensamos que estamos confusos em nível muito mais elevado e sobre coisas mais importantes".

Autor desconhecido

OFEREÇO

A DEUS, pela vida

DEDICO,

Aos meus pais, Reinaldo e Ermelinda,
pureza de amor e simplicidade.
Aos meus irmãos, Enedi, Hermes, Marirosa e Gilmar.

AGRADECIMENTOS

À Universidade de Passo Fundo, nas pessoas da Drª Tania Marisa Kuchenbecker Rösing e Dr. Erlei Melo Reis, pela confiança e apoio.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas (DEX), pelos conhecimentos.

Ao CNPq, pela bolsa de estudos.

Ao professor Daniel Furtado Ferreira, pela orientação, pelas críticas, sugestões e, principalmente, pela disponibilidade e ensinamentos durante o curso.

Aos professores Joel, Augusto, Lucas, Eduardo, Luiz Henrique, Delly, Julio, Agostinho, Maria do Carmo e Thema, pelos ensinamentos e companheirismo.

Aos meu colegas de curso, Hugo, Janaína, Marcelo, Telde e Valéria, pelo coleguismo e amizade.

Às funcionárias do DEX, Maria, Ivone, Renata, Odila e Claúdia, pela colaboração e amizade.

Aos queridos sobrinhos Edivandro, Joclei (*in memorian*), Edinei, Maurício, Morgana, Jordana, Juliano, Gemanir (obrigado pelas "cartinhas") e a Ana Carolina, fontes de inspiração para finalizar este trabalho.

As "mães por empréstimo", Dona Deise e Irene, pelo carinho com que sempre me receberam.

A Lu, Flávia, Yoko, Heloíse e Alcivania pela amizade e paciência no convívio diário.

A todos aqueles que direta ou indiretamente colaboraram e entenderam esta dificil, mas, importante etapa de minha vida.

MUITO OBRIGADA !!!

SUMÁRIO

RESUMO

CECCHETTI, Dileta. **Poder e taxas de erro tipo I de quatro critérios multivariados para o teste de igualdade de efeitos de tratamentos avaliados por meio do método de Monte Carlo.** Lavras: UFLA, 1999. 57p. (Dissertação - Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária).*

Na análise de variância multivariada existem quatro critérios considerados principais para testar a hipótese de igualdade de efeitos dos vetores médias de tratamentos, com aproximações por meio da distribuição F, quais sejam: os critérios de Wilks, Lawley-Hotelling, Roy e o de Pillai. O presente trabalho teve como objetivo avaliar, por meio de simulação, o poder e as taxas de erro tipo I destes critérios. Utilizou-se a linguagem do sistema SAS (Statistical Analysis System) para simular 2000 experimentos em diversas situações e em duas etapas. Na primeira etapa foram avaliadas as taxas de erro tipo I. Para isso foram feitas diversas combinações, considerando o número de tratamentos igual a 5, 10 e 20, número de repetições igual a 3 e 5, número de variáveis igual a 2 e 3 e coeficiente de variação de 5, 10 e 20%, fixados, respectivamente, para as variáveis 1, 2 e 3 e ainda considerando a não existência de efeito de tratamentos. A segunda etapa foi realizada nas mesmas condições da primeira, exceto a forma de gerar os efeitos de tratamentos. Para que os vetores de efeito de tratamentos fossem especificados, considerou-se diferenças consecutivas dadas em função do erro padrão da média paramétrica. Em ambas as etapas, foram adotados os níveis nominais de significância de 1% e 5%. Foi gerado um total de 408.000 experimentos. Pode-se concluir que a estrutura de correlação entre as variáveis não afetou as taxas de erro tipo I, porém tiveram influências no poder dos quatro critérios. Os critérios de Pillai, Wilks e Hotelling mostraram-se similares entre si quanto às taxas de erro tipo I, sendo o critério de Pillai o mais rigoroso, por ter apresentado taxas significativamente ($p < 0,05$) inferiores aos níveis nominais, além de ter mostrado menor poder. O critério de Hotelling apresentou tendência, em algumas circunstâncias, de possuir taxas de erro tipo I estatisticamente ($p < 0,05$) superiores ao nível nominal e tendência geral de possuir maiores taxas de erro tipo I do que os critérios de Wilks e Pillai. O critério de Roy tem um alto poder, mas as taxas de erro tipo I foram estatisticamente ($p < 0,05$) superiores às taxas nominais (α=1% ou 5%) em todas as situações, tornando-se inviável sua utilização. O critério de Wilks é recomendado como ideal para os testes de hipótese de igualdade do efeito do vetor média de tratamentos da análise de variância multivariada, por apresentar taxas de erro

---

* Comitê Orientador: Daniel Furtado Ferreira - UFLA (Orientador), Augusto Ramalho de Moraes - UFLA e Thelma Sáfadi - UFLA

tipo I iguais aos níveis nominais de significância e poder superior ao critério de Pillai, embora possua poder ligeiramente inferior ao critério de Hotelling.

ABSTRACT

CECCHETTI, Dileta. **Power and type I error rate of four multivariate criteria for testing equality of treatments evaluated by means of the Monte Carlo's method**. LAVRAS: UFLA, 1999. 57p. (Dissertation-Master in Agronomy, major in Statistics and Agricultural Experimentation).*

In the multivariate analysis of variance, four criteria regarded as chief exist for testing the hypothesis of equality of treatment means vectors with approaches by means of the F distribution namely the, criteria by Wilks, Lawley-Hotelling, Roy and that by Pillai. The present work aimed to evaluate, by means of simulation, the power and type I error rates of these criteria. The language of the SAS system - Statistical Analysis System - was utilized to simulate 2000 experiments in several situations and in two steps. In the first one, the type I error rates were evaluated. For such a number of combinations were done, by taking into account the number of treatments equal to 5,10 and 20, replication number equal to 3 and 5, number of variables equal 2 and 3 and variation coefficient of 5, 10 and 20%, fixed, respectively to variables 1, 2 and 3 and still considering the non-existence of effect of treatments. The second step was proceeded under the same condition of the first one, excepting the manner of generating the effects of treatments. For the treatments effect vectors might be specified, successive differences given in function of the standard error of the parametric means were specifieg. In both steps, the nominal significance levels of 1% and 5% were adopted. A total of 408,000 experiments was generated. It follows that correlation structure among the variables did not affect the type I error rates, but had they influences upon the power of the four criteria. Pillai, Wilks and Hotelling's criteria proved similar among them as to the type I error rates. Pillai's criterium was the most rigorous for having presented type I error rate significantly ($p < 0.05$) inferior to the nominal levels, besides having shown less power. Hotelling's criterion presented trend, in some circumstances, of possessing type I error rates statistically ($p < 0.05$) superior to the nominal level and general trend of possessing greater type I error rates than Wilks' and Pillai's criteria. Roy's criterion has a high power, but the type I error rate were statistically ($p < 0.05$) higher than nominal rates (α=1% or 5%) in every situation, making their utilization unfeasible. Wilks' criterion is advised as ideal to the tests of the equality hypothesis of the effect of the means vector of treatments of the multivariate analysis of variance, for presenting type I error rates equal to the nominal levels of significance and

* Guidance Committee : Daniel Furtado Ferreira - UFLA (Major professor), Augusto Ramalho de Moraes - UFLA and Thelma Sáfadi - UFLA

power superior to Pillai's criterion, although it possesses slightly poorer power than Hotelling's criterion.

1 INTRODUÇÃO

A análise de variância multivariada pode ser feita para qualquer delineamento experimental sem maiores dificuldades. É preciso trabalhar com todas as variáveis simultaneamente e obter as somas de quadrados (SQ) e as somas de produtos (SP). As técnicas de análise multivariada, combinam simultaneamente as múltiplas variáveis provenientes de uma unidade experimental, fornecendo informações nem sempre possíveis de serem obtidas com o uso das análises univariadas. Para muitos tipos de dados nas áreas de biologia, ciências agrárias, medicina, sociologia e outras, a correlação entre variáveis é comum e, assim, as informações fornecidas por análises univariadas isoladas não a considera.

Na análise de variância multivariada são apresentados quatro critérios considerados principais para testar a hipótese de igualdade de efeitos dos vetores médias de tratamentos: o de Wilks (Razão de Verossimilhança), o traço de Lawley-Hotelling, a raiz máxima de Roy e o traço de Pillai. Pouco se conhece com relação ao poder e taxas de erros tipo I e II destes critérios. Segundo Harris (1975), a maioria dos textos e programas de computador empregam o critério de Wilks para os testes de significância em razão de sua precedência histórica, bem como das excelentes aproximações das distribuições desta estatística por meio da distribuição F. O mesmo autor justifica a utilização do critério de Roy pelo seu valor didático. Nenhum destaque é feito na literatura sobre os critérios de Hotelling e Pillai.

O presente trabalho teve o objetivo de avaliar o poder e as taxas de erro tipo I destes quatro critérios para o teste de hipótese de igualdade de vetores média de tratamentos, na análise da variância multivariada (MANAVA), por meio de simulação computacional, usando o método de Monte Carlo em diferentes situações de experimentos de classificação simples, em relação ao número de tratamentos, ao número de repetições, às estruturas de correlações residuais, às diferenças consecutivas entre vetores média e nível de significância, com a finalidade de identificar o melhor entre eles.

2 REVISÃO DE LITERATURA

A oportunidade de resumir sucintamente grandes conjuntos de dados, especialmente em estádios exploratórios de uma investigação, tem contribuído para o aumento do interesse nos métodos multivariados.

Análise de Variância Multivariada é um procedimento para testar diferenças entre tratamentos de acordo com as médias de todas as variáveis, sendo aplicado a dados com distribuição normal multidimensional (James e McMulloch, 1990). Dados com estas características são muito comuns nas áreas biológica, agropecuária, etc, tais como: peso de matéria seca, altura de plantas e produção de grãos. Nesses casos é possível obter resultados não-significativos em testes univariados e resultados significativos em teste multivariado ou vice-versa (Manly, 1986). Segundo Demétrio (1985), a análise de variância multivariada, além de fornecer resultados com base na análise conjunta de todas as variáveis utilizadas, levando-se em consideração um nível de significância conhecido, permite estimar a melhor combinação de variáveis que leva a um valor de F máximo.

Ao se analisar os efeitos de tratamentos para diferentes variáveis x_1, x_2, ..., x_p, observadas nas mesmas unidades experimentais através de análises univariadas, não são levadas em consideração as correlações existentes entre elas (Winer, 1971). Este mesmo autor comentou que as respostas simultâneas nas unidades experimentais para todas as variáveis, consideradas como uma única resposta, geralmente contêm mais informação sobre o efeito total de tratamentos do que uma série de respostas consideradas separadamente; na realidade, isto é o que acontece quando dados multivariados são tratados como uma série de dados univariados independentes.

Para ilustrar esse fato, o autor utilizou um exemplo com três tratamentos e duas variáveis e verificou que, embora o teste F para a análise de variância multivariada tenha sido significativo ao nível de 1% de probabilidade, ao se fazerem as análises univariadas os resultados não foram significativos, nem ao nível α = 0,05. Ele justificou este fato mostrando que a alta correlação intra-classes existente entre as variáveis (r = 0,879) é desprezada ao se fazerem análises univariadas e verificou que o teste multivariado de certa forma incorpora uma série de testes univariados envolvendo um conjunto de elementos ponderados de forma ótima.

Nos experimentos mais comuns, cada parcela fornece o valor de uma só variável a ser analisada pelos métodos tradicionais. Porém, há casos em que cada parcela fornece observações relativas a duas ou mais variáveis a serem analisadas. Se estas variáveis forem independentes entre si, as análises de variância isoladas de cada variável seriam suficientes. Todavia, segundo Harris (1975), a probabilidade de se cometer um erro tipo I (α) aumenta rapidamente com o aumento do número de testes, ou seja, com o aumento do número de variáveis.

A estatística multivariada não é recente. Adrian (1808), Laplace (1811), Plana (1813), Gauss (1823) e Bravais (1846), citados por Anderson (1984), estudaram a densidade da distribuição normal bivariada. Posteriormente, Wilks (1932), Lawley (1938), Hotelling (1951), Roy (1953) e Pillai (1955) propuseram testes para a hipótese de nulidade (Ho) na análise de variância multivariada. Porém, a relativa complexidade de seus cálculos, aliada à escassez de equipamentos eletrônicos naquela época, limitou o uso das técnicas de análise multivariada, além de ter inibido o intercâmbio desses conhecimentos entre os pesquisadores.

De acordo com Dagnelie (1982), pode-se considerar a análise multivariada, em sentido amplo, como formada por um conjunto de métodos estatísticos que têm por objetivo o estudo das relações existentes entre variáveis dependentes ou interdependentes e que, especificamente, têm por objetivo comparar as médias de duas ou mais de duas populações.

Este mesmo autor apresenta a hipótese multivariada da seguinte forma:

$$Ho: \underset{\sim}{\tau_1} = \underset{\sim}{\tau_2} = ... = \underset{\sim}{\tau_I},$$

em que $\underset{\sim}{\tau_I}$ é o vetor de médias das p variáveis associado ao tratamento i (i=1, 2, ..., I) e, para testar esta hipótese Ho, pode-se utilizar os critérios de Wilks, de Lawley-Hotelling, de Roy e de Pillai.

Mardia (1971) apresentou o modelo linear geral multivariado, citando como estatísticas para o teste da hipótese de igualdade de vetores média de tratamentos as estatísticas abaixo, as quais são análogas à estatística F na análise de variância univariada.

$$\Lambda = \prod_{i=1}^{p}(1-\lambda_i) = \frac{|E|}{|E+H|} \qquad \text{(Wilks, 1932)}$$

$$U = \sum_{i=1}^{p}\frac{\lambda_i}{1-\lambda_i} = \mathrm{tr}\,(HE^{-1}) \qquad \text{(Lawley, 1938; Hotelling, 1951)}$$

$$\theta = \lambda_1 = \text{Máx}\left|\frac{\lambda_i}{1-\lambda_i}\right| \qquad \text{(Roy, 1953)}$$

$$V = \sum_{i=1}^{p}\lambda_i = \mathrm{tr}\,[\,H\,(E+H)^{-1}\,] \qquad \text{(Pillai, 1955),}$$

em que λ_i é o maior autovalor da matriz H, satisfazendo à equação $[H-\lambda(E+H)]=0$, H é a matriz de somas de quadrados e de produtos de tratamentos e E é a matriz de somas de quadrados e de produtos do resíduo.

Chatfield e Collins (1980) comentaram que os cálculos da análise de variância multivariada são essencialmente similares aos da univariada, porém, os testes de hipóteses e a interpretação de resultados são mais complicados.

Acrescentaram ainda que estudos feitos sobre os testes sob Ho são inconclusivos e mostram que a estrutura da hipótese alternativa afeta o resultado com relação ao poder. Assim, o critério de Roy foi considerado melhor sob a hipótese alternativa de que as médias são colineares ou aproximadamente colineares, enquanto que os outros três critérios são assintoticamente equivalentes e diferem pouco em poder para pequenas amostras. Os autores justificam o uso do critério de Wilks devido à facilidade de cálculo em situações de dimensão pequena e, principalmente, devido a existência de aproximações que facilitam a determinação de valores críticos.

Muitos autores, entre eles Harris (1975) e Mardia (1971), utilizaram a estatística de Wilks (Λ) por quatro razões: a) precedência histórica; b) embora a distribuição de Λ seja extremamente complexa (Schatzoff, 1966b), existem boas aproximações em termos das distribuições χ^2 e F para consulta de tabelas; c) sob certas circunstâncias, especialmente para dados em casos em que os autovalores de $E^{-1}H$ são aproximadamente iguais, testes estatísticos baseados em Λ são mais poderosos que o do maior autovalor; d) determinantes são mais fáceis de serem computados do que autovalores.

Harris (1975) justificou a utilização do critério de Roy tendo em vista seu valor didático e ainda porque: a) leva direta e naturalmente a procedimentos de comparações múltiplas; b) faz a conexão entre a análise discriminante e MANAVA, enquanto textos empregando a estatística de Hotelling têm que introduzir os procedimentos de raízes características como uma parte separada para a análise discriminante; c) o poder do teste de maior autovalor supera o da razão de verossimilhança quando o maior autovalor excede em muito os outros e os dois são equivalentes para amostras de tamanho grande; d) em relação ao item b, a fonte de qualquer diferença estatisticamente significativa é prontamente discriminada (através das funções discriminantes estatisticamente significativas) quando as técnicas do teste de Roy são empregadas.

Pillai e Jayachandran (1967), comparando os quatro critérios de Wilks, Pillai, Hotelling e Roy, inferiram que os três primeiros são bons para testar as hipóteses: (i) a independência entre um conjunto com p e um conjunto com q variáveis de uma população normal de variáveis (p+q) e (ii) a equivalência de vetores média de populações normais de variáveis que têm matriz de covariância comum. Os autores fizeram as seguintes observações com respeito a hipótese (i): a) para pequenos desvios das hipóteses, o critério de Pillai teve maior poder do que o critério de Wilks e o critério de Wilks mais poder do que o de Hotelling; b) para valores constantes da soma dos quadrados dos autovalores ($\lambda_1^2 + \lambda_2^2$), o poder do critério de Pillai aumentou à medida em que as duas raízes características tenderam a se igualar, mas os poderes de Hotelling e de Wilks decresceram a exceção de pequenos tamanhos de amostra (n); c) para grandes desvios da hipótese mencionada, quando os autovalores se distanciam um do outro, o poder do critério de Hotelling e de Wilks excedem em alguns casos o critério de Pillai para grandes valores de n, mas se λ_1 e λ_2 são próximos, o poder de Pillai sempre é maior que os de Wilks e Hotelling; além disso, o poder do critério de Hotelling é maior do que o de Wilks para grandes valores de n.

Para a hipótese (ii) verificaram que: a) para valores constantes de λ_1 e λ_2, o poder do critério de Pillai e Wilks aumentou à medida em que as raízes características tenderam a se igualar, enquanto o de Hotelling decresceu; b) ocorre da mesma forma como descrito no item c da hipótese (i), exceto que o poder de Hotelling só será maior que o de Wilks para n grande, se os valores de λ_1 e λ_2 tenderem a ser diferentes.

Tatsuoka (1971) apresenta como critérios alternativos de teste da hipótese Ho, o critério de Hotelling, de Roy e o de Pillai, sem, entretanto, chegar a resultados numéricos. Diz que, de acordo com um estudo de Schatzoff (1966), os critérios de Wilks e de Hotelling apresentam sensibilidade semelhante, enquanto que o de Roy é considerado melhor no caso em que a maioria das diferenças concentra-se em uma só dimensão.

Segundo Anderson (1984), os quatros critérios que receberam mais consideração foram os de Wilks, Lawley-Hotelling, Pillai e Roy. Para guiar a escolha de um desses, comparou-se a função do poder dos mesmos. Os três primeiros foram comparados por Rothenberg (1977) baseando-se nas expressões assintóticas de suas distribuições, no caso não nulo. Foi verificado que o poder dos critérios de Lawley-Hotelling e de Pillai diferem do poder de razão de verossimilhança (Wilks) e que o poder de Lawley-Hotelling é maior do que o de Wilks que, por sua vez, é maior que o de Pillai, se a variabilidade relativa dos autovalores for tão grande que:

$$\frac{\sigma_{\lambda}}{\bar{\lambda}} > \sqrt{\frac{(p-1)(p+2)}{pm+2}}$$

em que, σ_{λ}^{2} é a variância da população para os autovalores ($\lambda_1, \lambda_2, ..., \lambda_p$); p é o número de variáveis e m é o grau de liberdade de tratamentos. Se a desigualdade for revertida, o poder dos critérios também é revertido.

Comparações numéricas foram feitas por Schatzoff (1966b) e Olson (1974) usando o método de Monte Carlo. Mikhail (1965), Pillai e Jayachandran (1967) e Lee (1971) usaram as expansões assintóticas de distribuições para comparar os quatro critérios. Todos os resultados concordam com os de Rothenberg (1977). Entre esses procedimentos, o critério de Pillai deve ser preferido se as raízes forem aproximadamente iguais na hipótese alternativa. Já o traço de Lawley-Hotelling é mais poderoso quando as raízes são substancialmente desiguais e o teste de razão de verossimilhança de Wilks parece ser o segundo melhor.

A maior raiz de Roy tem distribuição limitante, que não é a distribuição de χ^2 sob hipótese nula e uma distribuição χ^2 não-central sob sequência da hipótese alternativa. Consequentemente, a comparação de Rothenberg não pode ser ampliada até este caso. As distribuições no caso não nulo são difíceis de serem avaliadas. Olsen (1974), estudando a baixa robustez sob desvios de homogeneidade da matriz de covariância bem como desvios de normalidade dos erros, verificou que os dois critérios traços e o de razão de verossimilhança foram moderadamente robustos, sendo que o critério da raiz máxima (Roy) foi o menos robusto.

2.1 Erros tipo I e II e poder dos testes

Na maioria dos testes de hipóteses realizados, é possível o controle do chamado erro tipo I, mas não do erro tipo II, a menos que as hipóteses sob testes sejam do tipo simples. Por definição, quando Ho é rejeitada sendo de fato verdadeira, comete-se o erro tipo I (α) ou, alternativamente, se Ho for falsa e se não for rejeitada, comete-se o erro tipo II (β). Mayer (1984) comenta que não é possível evitar completamente esses erros, mas que se deve manter relativamente pequena a probabilidade de cometê-los. O poder de um teste é definido por Mood, Graybill e Boes (1974) como uma função do parâmetro em questão, correspondendo à probabilidade de rejeitar a hipótese nula (Ho).

Tonhasca Jr. (1991), discutindo aspectos normalmente negligenciados no uso de metodologias estatísticas, diz que, estabelecida a hipótese alternativa, o maior problema do teste é calcular o valor de $1-\beta$. Apesar disso, pesquisadores deveriam estar cientes das limitações impostas por experimentos com poucas repetições e analisados com os níveis usuais de α (5 e 1%), os quais frequentemente fornecem baixos valores de $1-\beta$, indicando que não haveria alta probabilidade de rejeitar Ho falsa. Acrescenta que o poder de um teste depende do valor de α, da hipótese alternativa e do tamanho da amostra.

Costa (1992) comenta que o poder representa a "sensibilidade da região crítica" para perceber (e rejeitar) uma hipótese falsa. Rodrigues Jr.(1995), trabalhando com simulação, observou que a sensibilidade do teste da interação entre dois fatores diminui à medida em que a magnitude do efeito principal de tratamentos aumenta. Comentou, ainda, que quanto mais evidente for o efeito de tratamentos, maior será a freqüência esperada de erro tipo I e, consequentemente, a probabilidade de erro tipo II se altera com esta situação, só que de maneira inversa. Observou também que o aumento, em magnitude, do efeito de tratamentos, fixando um valor paramétrico para a interação, sugere a diminuição da freqüência esperada do erro tipo II. Mostrou que o aumento do tamanho amostral diminui o "viés" causado pela presença de efeito da interação, pois o tamanho da amostra influiu diretamente na frequência esperada dos erros tipo I e II.

Um bom teste é aquele que rejeita Ho com alta probabilidade quando ela é falsa, e aceita-a igualmente com alta probabilidade quando o parâmetro θ está contido no espaço (ponto ou intervalo) especificado por Ho do espaço paramétrico Θ. A construção de testes de hipóteses terá sempre a dificuldade de estabelecer a região crítica (C), ou seja, os valores amostrais para os quais Ho é rejeitada. Para construir estas regras de decisão que forneçam o menor β, para testes de hipóteses simples, existe o lema de Neyman-Pearson que, para qualquer decisão, com nível de significância $\alpha' \leq \alpha$, tem-se que o erro tipo II associado a esse outro critério de decisão, β' é tal que $\beta' \geq \beta$ (Mood, Graybill e Boes, 1974).

Segundo Guerra e Donaire (1982), em um teste decisório deve-se considerar os erros α e β, e somente uma amostra suficientemente grande pode reduzir o erro β, fixando-se um baixo risco α. Usualmente os pesquisadores fixam o valor de α para analisar os dados e com isso podem chegar a conclusões errôneas, pois não pode ser tratado como um teste decisório inequívoco e, sim, apenas como um teste de significância.

A Figura 1 apresenta o poder do teste ($1-\beta$) em função do parâmetro μ (medida de uma população normal) em um teste bilateral do tipo:

$$Ho : \mu = \mu o$$

$$H_1 : \mu \neq \mu o$$

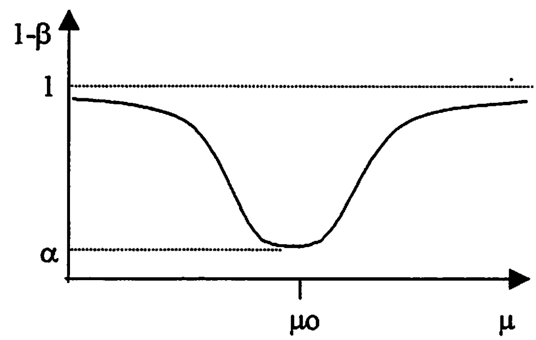

FIGURA 1. Curva do poder de um teste bilateral em que $1-\beta$ está em função de μ.

A análise da curva representada na Figura 1 permite verificar que: a) a altura da curva é a medida de seu poder e quanto mais elevada a curva em um dado ponto, maior é o poder do teste; b) a região de baixo poder está associada ao valor especificado em Ho. Para que o poder aumente rapidamente, para valores de μ próximos a μo deve-se aumentar o tamanho da amostra, que proporcionará um teste com maior poder e taxas de erro tipo I baixas.

2.2 Simulação pelo método de Monte Carlo

A idéia de construir modelos de simulação é muito antiga. As primeiras simulações feitas pelo homem talvez tenham sido na forma de maquetes para edificações e, mais tarde, em combates, usando pedras e galhos de árvores para representar batalhões e obstáculos. Possivelmente, o avanço da ciência não teria alcançado o estágio em que a humanidade se encontra, ou pelo menos teria levado-a por caminhos bastantes diferentes, se o homem não tivesse a habilidade quase inata de criar modelos e simular o comportamento de fenômenos do mundo real através de sua utilização (Dachs, 1988).

Naylor et al. (1971) comentaram que o principal fundamento lógico para utilizar a simulação em qualquer assunto (seja em pesquisa operacional, científica, etc.) é o incessante desejo do homem de conhecer o futuro. Esta busca de conhecimento e o desejo de conhecer o futuro são tão velhos como a história da raça humana. Dizem ainda os autores que existem várias razões para usar a simulação em computadores, sendo que uma delas considera que a simulação pode servir como uma primeira avaliação de um sistema para gerar novas políticas e regras de decisões antes de se correr o risco de experimentá-las no sistema real.

Para estudar ou avaliar um teste estatístico, muitas vezes torna-se bastante difícil obter, analiticamente, informações sobre o poder e taxas de erro tipo I de um teste. Uma maneira de se obter as informações desejadas é através de simulação. Na maioria dos trabalhos que envolvem simulação sempre está associado a ela o termo "método de Monte Carlo", que é muito utilizado para obter resultados de maneira mais simples.

O método de Monte Carlo pode ser definido como a representação da solução de um problema, utilizando para isso uma seqüência de números aleatórios como objetivo de construir uma amostra da população. Assim, estimativas estatísticas de parâmetros representativos do problema podem ser obtidas. Essa definição mostra que o alcance das aplicações é enorme e ~~fascinante~~.

Segundo Dachs (1988), o nome "Monte Carlo" é usado atualmente com freqüência crescente para referir-se ao uso de técnicas computacionais que geram amostras de acordo com determinadas distribuições teóricas conhecidas, visando estudar novos comportamentos de diferentes técnicas estatísticas que poderiam ser empregadas num dado problema. Naylor et al. (1971) dizem que informações sobre valores ou momentos esperados podem não ser suficientes para descrever o processo. Nesses casos, o método de Monte Carlo pode ser a única maneira satisfatória de prover as informações necessárias. Ele baseia-se em simulações de amostras, segundo distribuições previamente estruturadas, que serão analisadas pelos testes estatísticos em estudo, considerando a hipótese nula estabelecida. Com o crescente avanço do computador, é cada vez mais fácil e prática a simulação de variáveis ou de amostras baseadas em modelos estatísticos apropriados com parâmetros conhecidos, com o objetivo de se verificar a adequação de determinada metodologia ou na realização de comparações entre métodos (Dachs, 1988).

O uso da simulação de dados tem uma grande diversidade de áreas de aplicação. Em pesquisas na estatística e melhoramento genético, constantemente se tem interesse em comparar diferentes métodos de estimação de componentes de (co)variância ou diferentes métodos de melhoramento genéticos ou, ainda, comparar diferentes maneiras de se analisar dados de uma determinada população.

Vários autores têm recorrido à simulação de dados em seus estudos. Entre eles citam-se Rodrigues Jr. (1995), que utilizou o método de Monte Carlo para avaliar a probabilidade dos erros tipo I e II dos testes para delineamentos cross-over 2x2 de respostas binárias; Ferreira (1995), que utilizou os recursos de programação em simulação de dados para avaliar a eficiência de métodos de mapeamento de locos quantitativos (QTLs) e da seleção assistida por marcadores moleculares; Durbin e Koopman (1997), que, através do método de Monte Carlo testaram a estimação de máxima verrossimilhança de modelos de espaço de estado não-Gaussianos; Silva (1998), que utilizou a linguagem Pascal, pelo método de Monte Carlo, para simular dados para comparar a taxas de erro tipo I e o poder do teste Scott-Knott, e Veiga (1998), que utilizou os recursos do SAS para simular cruzamento dialélico com 50 pais, visando avaliar a eficiência de dialelos circulantes em relação ao dialelo completo, quanto às estimativas da capacidade geral e específica de combinação.

Para gerar um vetor de observações considerando mais de uma característica, pode-se utilizar a decomposição de Cholesky para a matriz de variância-covariância dos efeitos aleatórios. Seja, para isso, a matriz de covariância residual paramétrica Σ :

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \dots \sigma_{1p} \\ \sigma_{21} & \sigma_{22} \dots \sigma_{2p} \\ \vdots & \vdots \ddots \vdots \\ \sigma_{p1} & \sigma_{p2} \cdots \sigma_{pp} \end{bmatrix}$$

Van Vleck (1994) descreveu como utilizar essa decomposição de Cholesky, que também foi utilizada por Rodda et al. (1977), Walter e Mao (1985) e Huber, White e Hodge (1994). Uma matriz Σ pode ser escrita como o produto de outras duas matrizes provenientes da decomposição de Cholesky (L) de Σ, ou seja,

$$\Sigma = LL'$$

em que, L é uma matriz triangular inferior e L' é uma matriz triangular superior.

Um vetor simples $\underset{\sim}{\varepsilon}$, de variáveis provenientes de uma população com média zero e matriz de variância-covariância Σ, $\underset{\sim}{\varepsilon} \sim N(\phi, \Sigma)$, pode ser simulado pré-multiplicando um vetor $\underset{\sim}{v}$ de desvios normais aleatórios, $\underset{\sim}{v} \sim N(\phi, I)$, por L, ou seja

$$\underset{\sim}{\varepsilon} = L\underset{\sim}{v}.$$

Pode-se verificar que:

$$E(\underset{\sim}{\varepsilon}) = E(L\underset{\sim}{v}) = LE(\underset{\sim}{v}) = \phi$$

$$V(\underset{\sim}{\varepsilon}) = E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}') - E(\underset{\sim}{\varepsilon})E(\underset{\sim}{\varepsilon}') = E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}') = E(L\underset{\sim}{v}\,\underset{\sim}{v}'L') =$$

$$= LE(\underset{\sim}{v}\,\underset{\sim}{v}')L'$$

e como $E(\underset{\sim}{v}\,\underset{\sim}{v}') = I$, tem-se

$$V(\underset{\sim}{\varepsilon}) = LL' = \Sigma.$$

De forma que a obtenção dos vetores por simulação não é tarefa limitante.

3 METODOLOGIA

Para comparar os critérios multivariados para o teste de hipótese de igualdade de efeitos de vetores média de tratamentos foram realizadas simulações computacionais. O modelo estatístico adotado, sem perda de generalidade, foi o seguinte:

$$y_{ijk} = \mu_k + t_{ik} + e_{ijk}$$

sendo, $i = 1,2,...,I$; $j = 1, 2, ..., r$ $k = 1,2,..., p$:

- y_{ijk} representa a resposta simulada obtida da k-ésima variável, do i-ésimo tratamento na j-ésima repetição;
- μ_k é a média geral da k-ésima variável;
- t_{ik} é o efeito do i-ésimo tratamento na k-ésima variável;
- e_{ijk} é o erro aleatório da ijk-ésima combinação de tratamento, repetição e variável. O vetor de erros $\underset{\sim ij}{e}' = (e_{ij1}, ..., e_{ijp})$ tem distribuição normal p-dimensional com vetor de médias nulo e uma matriz de covariâncias Σ comum a todos os tratamentos e os $\underset{\sim ij}{e}$ correspondentes a diferentes unidades experimentais são independentemente distribuídos.

Para gerar o vetor de erros do modelo linear apresentado anteriormente, foi utilizado o procedimento descrito por Van Vleck (1994). Gerou-se um vetor de erros ($\underset{\sim}{v}$) considerando variáveis independentes ($\rho = 0$) com vetor média zero ($\underset{\sim}{\phi}$) e covariância I (identidade), a partir da função RANNOR do SAS (1995), sendo ρ a correlação residual entre as variáveis. A partir da especificação de Σ e do seu fator de Cholesky (L), foi gerado o vetor de erros por :

$$\underset{\sim}{\varepsilon} = \mathrm{L}\,\underset{\sim}{\mathrm{V}}$$

o qual possui distribuição normal multivariada com média igual $\underset{\sim}{\phi}$ (vetor nulo) e covariância Σ.

Em cada simulação foi realizada a análise de variância multivariada de acordo com os procedimentos descritos em Anderson (1984), Johnson e Wichern (1992) e Bock (1975). Os critérios para realizar o teste de hipótese de igualdade de efeitos de vetores média de tratamentos ($\underset{\sim}{\tau}$).:

$$\mathrm{Ho}: \underset{\sim}{\tau_1} = \underset{\sim}{\tau_2} = \ldots = \underset{\sim}{\tau_I}$$

bem como as aproximações de F correspondentes estão apresentadas na Tabela 1.

As simulações foram feitas considerando números de tratamentos (I) iguais a 5, 10 e 20, números de repetições (r) iguais a 3 e 5, coeficiente de variação iguais a 5%, 10% e 20%, número de variáveis (p) iguais a 2 e 3 e diferentes estruturas para a matriz Σ.

Para realizar a comparação dos critérios, as simulações foram realizadas considerando-se duas etapas, sendo que na primeira delas, os critérios foram comparados quanto à taxa de erro tipo I. Para gerar os dados simulados, adotando-se o modelo apresentado anteriormente, combinou-se fatorialmente os valores adotados para I, r, p e ρ, obtendo-se 36 combinações diferentes. Em todos esses casos, o vetor de efeito de tratamentos $\underset{\sim}{\tau_i}$ foi fixado como $\underset{\sim}{\phi}$ (vetor nulo), para todo i = 1,2,...,I. Em cada combinação foram gerados 2000 experimentos, dos quais eram computadas as taxas de erro tipo I dos quatro critérios.

TABELA 1 - Estatísticas multivariadas e sua equivalência aproximada com a distribuição de F

Critério	Estatística	Aproximação F	GL de F
Wilks (1932)	$\Lambda = \Pi \frac{1}{1+\lambda_i} = \frac{\lvert E \rvert}{\lvert E+H \rvert}$	$F = \left(\frac{1-\Lambda^{\frac{1}{t}}}{\Lambda^{\frac{1}{t}}} \right) \left(\frac{rt-2f}{pq} \right)$	$V_1 = pq$ $V_2 = rt\text{-}2f$
Lawley-Hotelling (1938; 1951)	$U = \Sigma \lambda_i = tr(HE^{-1})$	$F = \frac{2(sn+1)U}{s^2(2m+s+1)}$	$V_1 = s(2m+s+1)$ $V_2 = 2(sn+1)$
Roy (1953)	$\theta = \lambda_1$	$F = \frac{\theta(\nu - d + q)}{d}$	$V_1 = d$ $V_2 = \nu\text{-}d\text{+}q$
Pillai (1955)	$V = \sum \frac{\lambda_i}{1+\lambda_i} = tr[H(H+E)^{-1}]$	$F = \left(\frac{V}{s-V} \right) \left(\frac{2n+s+1}{2m+s+1} \right)$	$V_1 = s(2m+s+1)$ $V_2 = s(2n+s+1)$

p: nº de variáveis = posto(H+E); q: GL de tratamentos (ou do contraste); ν : GL do erro; S=min(p,q); $r = \nu\text{-}(p\text{-}q\text{+}1)/2$; f = (pq-2)/4); d=max(p,q); m= (|p-q|-1)/2; n=(ν-p-1)/2 ; e

$$t = \begin{cases} \sqrt{\dfrac{p^2q^2-4}{p^2+q^2-5}} & \text{se } p^2+q^2-5>0 \\ 1 & \text{Caso contrário} \end{cases}$$

Obs.: Critério de Wilks possui aproximação exata de F se min(p,q) ≤ 2

3.1 Taxa de erro tipo I

A taxa de erro tipo I para cada critério foi computada estimando-se a proporção de análises que apresentaram significância para os níveis nominais (α) adotados. Esses níveis nominais foram fixados em 1% e 5%. Como a distribuição das estatísticas dos quatro critérios não era bem conhecida, foram utilizadas as aproximações de F apresentadas na Tabela 1 para avaliar a significância dos testes para a hipótese Ho. Para comparar as taxas observadas de erro tipo I com os valores nominais, o intervalo exato da binomial foi utilizado (Leemis e Trivedi, 1996). Os limites inferiores com 95% de confiança encontrados para o nível de significância (α) dos testes multivariados (α = 1% e 5%) foram de 0,612% e 4,086% e os superiores foram de 1,540% e 6,048%, respectivamente. Assim, os valores que não se encontraram nestes limites foram considerados diferentes das taxas nominais de significância adotados.

O coeficiente de variação de 5, 10 e 20% foi fixado, respectivamente, para as variáveis 1, 2 e 3. Nesse caso, as variáveis tiveram precisões diferentes, quanto ao coeficiente de variação paramétrico. No caso de p = 2 variáveis, o coeficiente de variação foi de 5% para a primeira variável e 10% para a segunda variável. Procedendo assim e adotando a média de cada variável (μ) igual a 100, sem perda de generalidade, foi possível compor a matriz Σ para cada valor de ρ adotado. As correlações residuais entre as variáveis, quando p = 3, foram consideradas as mesmas ($\rho_{12} = \rho_{13} = \rho_{23}$).

3.2 Poder

Na segunda etapa, os quatro critérios foram comparados quanto ao poder. As mesmas situações descritas para a simulação da primeira etapa foram realizadas na segunda, a exceção da forma de gerar os efeitos de tratamentos. Para que os vetores de efeitos de tratamentos fossem especificados, foram consideradas sempre diferenças entre tratamentos consecutivos. Essas diferenças foram específicas para cada variável componente deste vetor.

Os valores das diferenças consecutivas entre tratamentos, para cada variável, foram fixados parametricamente em função do erro padrão da média

($\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{r}}$), uma vez que as variáveis tinham diferentes precisões (CV). Os efeitos dos tratamentos consecutivos foram estipulados considerando-se diferenças dadas por k$\sigma_{\bar{X}}$, sendo que k foi fixado em ½, 1, 2 e 3.

O poder de cada um dos quatro critérios foi computado pela proporção estimada nas 2000 simulações de todas as situações em que as aproximações de F apresentaram significância nos níveis nominais de 1% e 5%, pré-estabelecidos. O efeito do número de repetições ou do número de tratamentos, ou da estrutura de correlação residual, ou da diferença consecutiva entre médias no poder de cada critério foi estudado, individualmente, fixando-se os demais fatores.

3.3 Poder entre dois grupos de tratamentos

Para estudar o poder dos quatro critérios numa situação em que a amplitude total paramétrica entre os tratamentos não se alterasse se o valor de I aumentasse, foi estipulada a estratégia apresentada a seguir. Considerando I = 10 ou 20, dois grupos de tratamento foram gerados, com efeitos iguais dentro dos grupos e cuja diferença entre grupos foi dada por k$\sigma_{\bar{X}}$. Os valores de k foram os mesmos apresentados anteriormente.

4 RESULTADOS E DISCUSSÃO

4.1 Taxas de erro tipo I

Nas Tabelas 2 e 3 estão apresentadas as taxas empíricas de erro tipo I para os quatro critérios multivariados (Wilks, Pillai, Hotelling e Roy) mensurados em 2000 simulações para cada combinação de tratamento (I), número de repetições (r) e correlação residual (ρ). A Tabela 2 refere-se às análises bivariadas e a Tabela 3 refere-se às análises realizadas com três variáveis.

De acordo com os resultados apresentados na Tabela 2 (análises bivariadas), fixando-se o número de tratamentos e a correlação residual, pode-se verificar que três dos critérios estudados apresentaram taxas de erro similares entre si e semelhantes às taxas nominais fixadas (α = 5 ou 1%). Estes três critérios são o de Wilks, Pillai e Hotelling. O aumento do número de repetições de 3 para 5, exceto nas situações de pequeno número de tratamentos (I= 5 ou I = 10), não alterou as taxas de erro tipo I destes diferentes critérios citados anteriormente.

É importante salientar que, no caso bivariado, tanto para valores de α de 5% ou de 1%, o critério de Wilks não apresentou casos de taxas de erro tipo I significativamente diferente dos níveis nominais. O critério de Pillai, em alguns casos, apresentou taxas de erro significativamente menores que os valores nominais. Esses casos referem-se principalmente às situações em que foi utilizado um número menor de tratamentos (I= 5), concomitantemente com número reduzido de repetições (r = 3). Este fato pode ser um indicativo de que este critério, nessas circunstância, seja mais rigoroso que os demais. Foi

[redacted]

encontrada apenas uma exceção para α = 1%, com I= 10, r = 3 e ρ = 0,5 (Tabela 2).

TABELA 2. Taxas de erro tipo I para os quatro critérios adotados para o teste da hipótese de igualdade dos vetores de efeitos de tratamentos, considerando p=2 e diferentes valores de I, r e ρ obtidos por meio de simulação, para os níveis nominais de significância de 1% e 5%

			Critérios							
			Wilks		Pillai		Hotelling		Roy	
I	r	ρ	5%	1%	5%	1%	5%	1%	5%	1%
5	3	0,0	0,0515	0,0095	0,0410	$0,0035^-$	0,0540	$0,0195^+$	$0,2040^+$	$0,0485^+$
5	3	0,5	0,0445	0,0070	$0,0305^-$	$0,0050^-$	0,0505	0,0145	$0,1725^+$	$0,0430^+$
5	3	0,9	0,0530	0,0130	$0,0360^-$	$0,0055^-$	0,0605	$0,0235^+$	$0,1890^+$	$0,0560^+$
5	5	0,0	0,0490	0,0140	0,0445	0,0090	0,0555	0,0150	$0,1760^+$	$0,0470^+$
5	5	0,5	0,0520	0,0110	0,0435	0,0065	0,0550	$0,0155^+$	$0,1730^+$	$0,0445^+$
5	5	0,9	0,0515	0,0130	0,0440	0,0085	0,0555	0,0145	$0,1740^+$	$0,0415^+$
10	3	0,0	0,0560	0,0120	0,0500	0,0105	$0,0615^+$	$0,0175^+$	$0,1835^+$	$0,0465^+$
10	3	0,5	0,0530	0,0085	0,0490	$0,0060^-$	0,0555	0,0105	$0,1880^+$	$0,0505^+$
10	3	0,9	0,0520	0,0140	0,0460	0,0120	0,0565	$0,0180^+$	$0,1870^+$	$0,0485^+$
10	5	0,0	0,0490	0,0085	0,0455	0,0065	0,0545	0,0120	$0,1775^+$	$0,0455^+$
10	5	0,5	0,0435	0,0105	0,0425	0,0085	0,0490	0,0110	$0,1590^+$	$0,0425^+$
10	5	0,9	0,0455	0,0070	0,0420	0,0065	0,0495	0,0115	$0,1690^+$	$0,0425^+$
20	3	0,0	0,0520	0,0120	0,0460	0,0100	0,0540	0,0120	$0,1805^+$	$0,0430^+$
20	3	0,5	0,0555	0,0110	0,0540	0,0090	$0,0610^+$	0,0130	$0,1790^+$	$0,0485^+$
20	3	0,9	0,0505	0,0085	0,0485	0,0075	0,0570	0,0095	$0,1965^+$	$0,0465^+$
20	5	0,0	0,0440	0,0070	0,0425	0,0080	0,0465	0,0075	$0,1750^+$	$0,0405^+$
20	5	0,5	0,0580	0,0115	0,0565	0,0120	$0,0610^+$	0,0110	$0,1830^+$	$0,0510^+$
20	5	0,9	0,0450	0,0115	0,0435	0,0110	0,0475	0,0120	$0,1780^+$	$0,0480^+$
Média			0,0503	0,0105	0,0448	0,0081	0,0547	0,0138	0,1802	0,0463
Amplitude			0,0435	0,0070	0,0305	0,0035	0,0475	0,0075	0,1590	0,0415
			0,0580	0,0140	0,0565	0,0120	0,0615	0,0195	0,2040	0,0560
Nº de +			0	0	0	0	3	5	18	18
Nº de –			0	0	2	4	0	0	0	0

$^-$ significativamente (P < 0,05) menor que α nominal

+ significativamente (P < 0,05) maior que α nominal

TABELA 3. Taxas de erro tipo I para os quatro critérios adotados para o teste da hipótese de igualdade dos vetores de efeitos de tratamentos, considerando p=3 e diferentes valores de I, r e ρ obtidos por meio de simulação, para os níveis nominais de significância de 1% e 5%

			Critérios							
			Wilks		Pillai		Hotelling		Roy	
I	r	ρ	5%	1%	5%	1%	5%	1%	5%	1%
5	3	0,0	0,0520	0,0105	0,0305⁻	0,0025⁻	0,0715⁺	0,0225⁺	0,4255⁺	0,1665⁺
5	3	0,5	0,0460	0,0085	0,0285⁻	0,0030⁻	0,0665⁺	0,0180⁺	0,4420⁺	0,1665⁺
5	3	0,9	0,0485	0,0105	0,0260⁻	0,0060⁻	0,0650⁺	0,0240⁺	0,4065⁺	0,1490⁺
5	5	0,0	0,0395⁻	0,0100	0,0320⁻	0,0045⁻	0,0520	0,0125	0,3480⁺	0,1130⁺
5	5	0,5	0,0390⁻	0,0080	0,0305⁻	0,0060⁻	0,0540	0,0125	0,3555⁺	0,1115⁺
5	5	0,9	0,0515	0,0080	0,0375⁻	0,0020⁻	0,0645⁺	0,0150	0,3785⁺	0,1365⁺
10	3	0,0	0,0510	0,0110	0,0410	0,0075	0,0615⁺	0,0175⁺	0,3805⁺	0,1290⁺
10	3	0,5	0,0430	0,0080	0,0350⁻	0,0070	0,0575	0,0135	0,3975⁺	0,1380⁺
10	3	0,9	0,0605	0,0125	0,0460	0,0080	0,0785⁺	0,0190⁺	0,4220⁺	0,1605⁺
10	5	0,0	0,0475	0,0100	0,0430	0,0055⁻	0,0515	0,0125	0,3715⁺	0,1265⁺
10	5	0,5	0,0535	0,0105	0,0480	0,0080	0,0610⁺	0,0140	0,3675⁺	0,1250⁺
10	5	0,9	0,0545	0,0135	0,0495	0,0085	0,0625⁺	0,0175⁺	0,3680⁺	0,1300⁺
20	3	0,0	0,0540	0,0135	0,0535	0,0125	0,0605	0,0170⁺	0,3860⁺	0,1240⁺
20	3	0,5	0,0480	0,0095	0,0470	0,0060⁻	0,0585	0,0135	0,3825⁺	0,1300⁺
20	3	0,9	0,0470	0,0090	0,0400⁻	0,0085	0,0535	0,0115	0,3985⁺	0,1425⁺
20	5	0,0	0,0465	0,0105	0,0460	0,0105	0,0515	0,0120	0,3735⁺	0,1240⁺
20	5	0,5	0,0510	0,0100	0,0470	0,0075	0,0510	0,0115	0,3860⁺	0,1405⁺
20	5	0,9	0,0450	0,0080	0,0425	0,0065	0,0460	0,0095	0,3910⁺	0,1165⁺
Média			0,0488	0,0101	0,0402	0,0067	0,0593	0,0152	0,3878	0,1350
Amplitude			0,0390	0,0080	0,0260	0,0020	0,0460	0,0095	0,3480	0,1115
			0,0605	0,0135	0,0535	0,0125	0,0785	0,0240	0,4420	0,1665
Nº de +			0	0	0	0	8	7	18	18
Nº de –			2	0	8	8	0	0	0	0

⁻ significativamente (P < 0,05) menor que α nominal

⁺ significativamente (P < 0,05) maior que α nominal

O critério de Hotelling, por sua vez, apresentou alguns casos em que as taxas de erro tipo I obtidas por simulação foram significativamente ($p < 0,05$) superiores às taxas nominais. Isto ocorreu em maior proporção para o nível nominal de 1% e nas situações de menor número de tratamentos e de repetições. Esse fato chama a atenção para um maior risco em se utilizar tal critério nas circunstâncias anteriormente mencionadas.

Para o caso trivariado, a mesma tendência foi observada, porém os resultados foram mais expressivos, ocorrendo em maior intensidade as significâncias comentadas para o caso bivariado (Tabela 3). É curioso notar que para o critério de Wilks com 5 tratamentos e com 5 repetições, as taxas de erro tipo I, em alguns casos, foram significativamente ($p < 0,05$) diferentes do valor nominal. No entanto, apesar destas diferenças, estes valores observados foram significativamente ($p < 0,05$) inferiores ao nível nominal, não trazendo nenhum risco adicional para sua utilização nestas circunstâncias. Para o critério de Pillai verificou-se um maior rigor do teste nas diferentes situações e para o critério de Hotelling verificou-se um maior risco de utilização do mesmo, uma vez que as taxas de erro tipo I foram significativamente maiores do que as taxas nominais em um maior número de situações.

Chama a atenção o fato de que com um número maior de tratamentos ($I = 20$) e um número maior de repetições ($r = 5$), somente uma exceção ocorreu, em que a taxa de erro tipo I diferiu da taxa nominal. Esse caso ocorreu para o critério de Hotelling com $\alpha = 5\%$ no caso bivariado (Tabela 2). A princípio, poderia-se imaginar que esses resultados devem-se ao aumento isolado do número de tratamentos ou do número de repetições. No entanto, pode-se inferir, a partir de uma análise mais aprofundada, que isto provavelmente está associado ao aumento na precisão das estimativas das matrizes de somas de quadrados e produtos de tratamentos e de resíduo, devido ao maior tamanho de amostras. O efeito julgado mais importante é o do aumento dos graus de liberdade residuais, uma vez que a matriz de covariância residual nesta circunstância seria estimada com maior precisão e também devido à sua participação na obtenção das estatísticas de todos os critérios.

O efeito da correlação residual nas taxas de erro tipo I mostrou-se similar nas análises bivariadas ou trivariadas, bem como não apresentaram diferenças com relação aos três critérios comentados (Wilks, Pillai e Hotelling). O aumento da correlação residual, fixados o número de tratamentos e o número de repetições, não provocou alterações consistentes nas taxas de erro tipo I dos diferentes critérios multivariados para os testes de hipótese.

O critério de Roy, por sua vez, apresentou taxas empíricas de erro tipo I superiores significativamente ($p < 0,05$) aos níveis nominais adotados, para todas as situações estudadas neste trabalho. É interessante observar que, na situação trivariada, os resultados das taxas de erro tipo I deste critério foram praticamente o dobro das obtidas no caso bivariado para o nível nominal de 5%, sendo agravada ainda mais tal diferença para o nível nominal de 1%. Os valores da ordem de 15% a 44% no caso do nível nominal de 5%, 4% a 17% para $\alpha = 1\%$, destas taxas de erro, o tornaram praticamente inviável para ser recomendado como um critério adequado nos testes multivariados, por considerar essas taxas de erro de magnitude elevadas, como sendo de alto risco (Tabelas 2 e 3). As Figuras 2 e 3 apresentam os quatro critérios para as situações com $I = 5$ ($r = 3$) e $I = 20$ ($r = 5$), respectivamente, do caso bivariado e $\alpha = 5\%$, em função da correlação residual (ρ). Destaca-se, nas Figuras 2 e 3, a acentuada discrepância do critério de Roy em relação aos demais, considerando as taxas de erro tipo I observadas nas simulações efetuadas. É interessante observar que existe uma tendência do critério de Hotelling apresentar maiores valores de taxas de erro tipo I do que os critérios de Wilks e de Pillai apresentar menores taxas, conforme já comentado anteriormente.

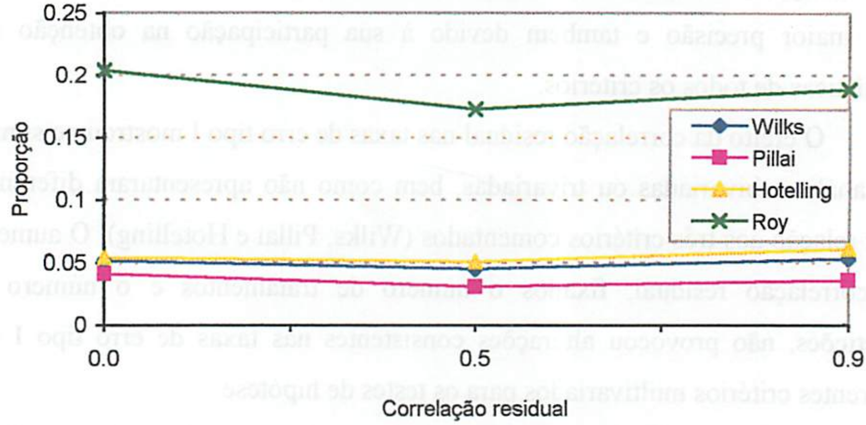

FIGURA 2. Taxa de erro tipo I dos quatro critérios multivariados considerando cinco tratamentos, três repetições e duas variáveis, para o nível de significância de 5%, em função de correlação residual.

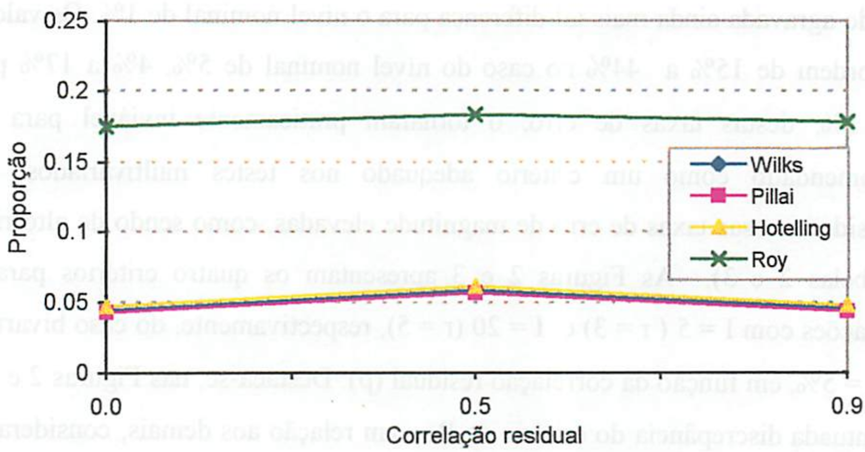

FIGURA 3. Taxa de erro tipo I dos quatro critérios multivariados considerando vinte tratamentos, três repetições e duas variáveis, para o nível de significância de 5%, em função de correlação residual.

4.2 Poder dos critérios multivariados

Nas Tabelas 4, 5 e 6 estão apresentados os poderes dos quatro critérios multivariados estudados no presente trabalho. Cada linha das tabelas corresponde ao poder do teste para diferentes números de repetições, estruturas de correlações residuais e diferenças entre vetores média consecutivas ($\sigma_{\bar{X}}$). A Tabela 4 refere-se a cinco tratamentos, a Tabela 5 a dez tratamentos e a Tabela 6 a vinte tratamentos. O efeito de cada fator no poder dos testes será apresentado isoladamente nas seções seguintes.

Comparando-se o poder dos critérios de Wilks, Pillai, Hotelling e Roy para as mesmas condições (mesmo I, r, ρ e $\sigma_{\bar{X}}$), pode-se inferir que eles apresentaram grandes diferenças. O critério de Pillai apresenta os menores valores de poder, seguido do critério de Wilks, Hotelling e Roy, nesta ordem. Para I = 5 tratamentos e r = 3, as diferenças entre os poderes dos testes são as maiores encontradas (Tabela 4). À medida em que aumenta o número de tratamentos e o número de repetições, verifica-se que há uma tendência dos poderes dos critérios de se aproximarem. Reduzindo-se o nível nominal de significância de 5% para 1%, a mesma tendência foi observada. No entanto, os poderes dos critérios se aproximam mais lentamente quando comparados com os obtidos nos casos de níveis nominais de 5%, em função do aumento de I e r (Tabelas 4, 5 e 6).

Chama a atenção o critério de Roy que, em todas as circunstâncias, apresentou poder superior e na pior das situações, poder semelhante ao dos demais critérios. Essa superioridade foi evidenciada nas situações de pequeno número de tratamentos (I = 5) e de poucas repetições (r = 3) com α = 1% (Tabela 4).

TABELA 4. Poder para os quatro critérios adotados para o teste da hipótese de igualdade dos vetores de efeitos de tratamentos, considerando p=3 variáveis, I=5 tratamentos e diferentes valores de r (repetições), ρ (correlação) e do fator multiplicador (k) do erro padrão da média ($\sigma_{\bar{X}}$), obtidos por meio de simulação, para os níveis nominais de significância de 1% e 5%

			Critérios							
			Wilks		Pillai		Hotelling		Roy	
r	ρ	K	5%	1%	5%	1%	5%	1%	5%	1%
3	0,0	0,5	0,1815	0,0515	0,1110	0,0200	0,2180	0,0910	0,7240	0,4040
3	0,0	1	0,6370	0,2860	0,3090	0,0760	0,7460	0,4880	0,9855	0,8935
3	0,0	2	0,9950	0,9285	0,5455	0,1980	0,9995	0,9915	1,0000	1,0000
3	0,0	3	1,0000	0,9990	0,5895	0,2075	1,0000	1,0000	1,0000	1,0000
3	0,5	0,5	0,1170	0,0280	0,0720	0,0095	0,1510	0,0535	0,6020	0,3020
3	0,5	1	0,3365	0,1145	0,1955	0,0380	0,4185	0,2030	0,8815	0,6295
3	0,5	2	0,9055	0,6365	0,4535	0,1400	0,9640	0,8495	1,0000	0,9965
3	0,5	3	0,9980	0,9510	0,5375	0,1690	1,0000	0,9960	1,0000	1,0000
3	0,9	0,5	0,0850	0,0225	0,0545	0,0075	0,1055	0,0365	0,5265	0,2240
3	0,9	1	0,2490	0,0735	0,1495	0,0220	0,2995	0,1260	0,8010	0,4985
3	0,9	2	0,7910	0,4525	0,3850	0,1040	0,8835	0,6860	1,0000	0,9680
3	0,9	3	0,9865	0,8530	0,5020	0,1540	0,9985	0,9780	1,0000	0,9995
5	0,0	0,5	0,2675	0,0810	0,2035	0,0425	0,3045	0,1300	0,7390	0,4470
5	0,0	1	0,8450	0,5820	0,6970	0,2950	0,8890	0,7215	0,9965	0,9595
5	0,0	2	1,0000	1,0000	0,9975	0,8425	1,0000	1,0000	1,0000	1,0000
5	0,0	3	1,0000	1,0000	1,0000	0,9745	1,0000	1,0000	1,0000	1,0000
5	0,5	0,5	0,1425	0,0390	0,1070	0,0205	0,1630	0,0630	0,5705	0,2690
5	0,5	1	0,5010	0,2400	0,3870	0,1100	0,5710	0,3355	0,9155	0,7200
5	0,5	2	0,9915	0,9465	0,9400	0,5795	0,9975	0,9820	1,0000	1,0000
5	0,5	3	1,0000	1,0000	1,0000	0,8755	1,0000	1,0000	1,0000	1,0000
5	0,9	0,5	0,1155	0,0300	0,0865	0,0175	0,1350	0,0465	0,5155	0,2255
5	0,9	1	0,3655	0,1475	0,2835	0,0695	0,4240	0,2185	0,8405	0,5855
5	0,9	2	0,9605	0,8050	0,8375	0,4230	0,9790	0,9075	1,0000	0,9945
5	0,9	3	1,0000	0,9995	0,9950	0,7730	1,0000	1,0000	1,0000	1,0000
Média			0,6446	0,5111	0,4766	0,2570	0,6770	0,5793	0,8791	0,7549
Amplitude			0,1155	0,0225	0,0545	0,0075	0,1055	0,0365	0,5265	0,2255
			1,0000	1,0000	1,0000	0,9745	1,0000	1,0000	1,0000	1,0000

TABELA 5. Poder para os quatro critérios adotados para o teste da hipótese de igualdade dos vetores de efeitos de tratamentos, considerando p=3 variáveis, I=10 tratamentos e diferentes valores de r (repetições), ρ (correlação) e do fator multiplicador (k) do erro padrão da média ($\sigma_{\bar{X}}$), obtidos por meio de simulação, para os níveis nominais de significância de 1% e 5%

			Critérios							
			Wilks		Pillai		Hotelling		Roy	
r	ρ	k	5%	1%	5%	1%	5%	1%	5%	1%
3	0,0	0,5	0,9225	0,7150	0,6705	0,3125	0,9710	0,8825	0,9995	0,9925
3	0,0	1	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	0,0	2	1,0000	1,0000	0,9510	0,7035	1,0000	1,0000	1,0000	1,0000
3	0,0	3	1,0000	1,0000	0,9710	0,7490	1,0000	1,0000	1,0000	1,0000
3	0,5	0,5	0,6160	0,3300	0,4395	0,1500	0,7110	0,4755	0,9755	0,8730
3	0,5	1	0,9990	0,9765	0,8395	0,5000	1,0000	0,9995	1,0000	1,0000
3	0,5	2	1,0000	1,0000	0,9485	0,6935	1,0000	1,0000	1,0000	1,0000
3	0,5	3	1,0000	1,0000	0,9645	0,7325	1,0000	1,0000	1,0000	1,0000
3	0,9	0,5	0,4605	0,2060	0,3405	0,1075	0,5525	0,2970	0,9265	0,7280
3	0,9	1	0,9885	0,8995	0,7850	0,4055	0,9985	0,9845	1,0000	1,0000
3	0,9	2	1,0000	1,0000	0,9435	0,6630	1,0000	1,0000	1,0000	1,0000
3	0,9	3	1,0000	1,0000	0,9580	0,7160	1,0000	1,0000	1,0000	1,0000
5	0,0	0,5	0,9915	0,9295	0,9475	0,7470	0,9965	0,9755	1,0000	1,0000
5	0,0	1	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,0	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,0	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,5	0,5	0,7800	0,5040	0,6725	0,3420	0,8380	0,6210	0,9920	0,9435
5	0,5	1	1,0000	1,0000	0,9990	0,9815	1,0000	1,0000	1,0000	1,0000
5	0,5	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,5	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9	0,5	0,5850	0,3240	0,5010	0,2225	0,6440	0,4005	0,9650	0,8145
5	0,9	1	0,9990	0,9925	0,9900	0,9055	1,0000	0,9990	1,0000	1,0000
5	0,9	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Média			0,9309	0,8699	0,8717	0,7055	0,9463	0,9015	0,9941	0,9730
Amplitude			0,4605	0,2060	0,3405	0,2225	0,5525	0,2970	0,9265	0,7280
			1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

TABELA 6. Poder para os quatro critérios adotados para o teste da hipótese de igualdade dos vetores de efeitos de tratamentos, considerando p=3 variáveis, I=20 tratamentos e diferentes valores de r (repetições), ρ (correlação) e do fator multiplicador (k) do erro padrão da média ($\sigma_{\bar{X}}$), obtidos por meio de simulação, para os níveis nominais de significância de 1% e 5%

			Critérios							
			Wilks		Pillai		Hotelling		Roy	
r	ρ	K	5%	1%	5%	1%	5%	1%	5%	1%
3	0,0	0,5	1,0000	1,0000	0,9995	0,9760	1,0000	1,0000	1,0000	1,0000
3	0,0	1	1,0000	1,0000	1,0000	0,9970	1,0000	1,0000	1,0000	1,0000
3	0,0	2	1,0000	1,0000	1,0000	0,9970	1,0000	1,0000	1,0000	1,0000
3	0,0	3	1,0000	1,0000	1,0000	0,9994	1,0000	1,0000	1,0000	1,0000
3	0,5	0,5	1,0000	1,0000	0,9925	0,9310	1,0000	1,0000	1,0000	1,0000
3	0,5	1	1,0000	1,0000	1,0000	0,9930	1,0000	1,0000	1,0000	1,0000
3	0,5	2	1,0000	1,0000	1,0000	0,9970	1,0000	1,0000	1,0000	1,0000
3	0,5	3	1,0000	1,0000	1,0000	0,9975	1,0000	1,0000	1,0000	1,0000
3	0,9	0,5	1,0000	0,9995	0,9810	0,8650	1,0000	1,0000	1,0000	1,0000
3	0,9	1	1,0000	1,0000	0,9995	0,9925	1,0000	1,0000	1,0000	1,0000
3	0,9	2	1,0000	1,0000	1,0000	0,9985	1,0000	1,0000	1,0000	1,0000
3	0,9	3	1,0000	1,0000	0,9995	0,9965	1,0000	1,0000	1,0000	1,0000
5	0,0	0,5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,0	1	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,0	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,0	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,5	0,5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,5	1	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,5	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,5	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9	0,5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9	1	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Média			1,0000	0,9999	0,9988	0,9892	1,0000	1,0000	1,0000	1,0000
Amplitude			1,0000	0,9995	0,9810	0,8650	1,0000	1,0000	1,0000	1,0000
			1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

4.2.1 Efeito do número de repetições

Para medir o efeito das repetições no poder dos quatro critérios estudados, deve-se fixar os demais fatores estudados. Procedendo-se assim verificou-se (Tabelas 4, 5 e 6) que a repetição teve um efeito marcante no poder dos testes, efeito que foi mais pronunciado nas situações de menor número de tratamentos (I = 5). A magnitude desse efeito foi menor nas situações de correlação residual nula ou baixa e de menores diferenças de tratamentos. Assim, para ilustrar tal fato, o poder do critério de Wilks para I = 5, r = 3, $\rho = 0$ e $0{,}5\,\sigma_{\bar{X}}$ foi de 18,15% e com I = 5, r = 5, $\rho = 0$ e $0{,}5\,\sigma_{\bar{X}}$ foi de 26,75%. Por outro lado, com I = 5, r = 3, $\rho = 0{,}9$ e $2\,\sigma_{\bar{X}}$, o poder foi de 79,1% e com I = 5, r = 5, $\rho = 0{,}9$ e $2\,\sigma_{\bar{X}}$, o poder foi de 96,05%.

O efeito do número de repetições no poder dos quatro critérios multivariados quando se considerou um maior número de tratamentos (I = 20) (Tabela 6) praticamente não pôde ser detectado. Isso ocorreu devido ao fato de todos os critérios terem apresentado valores de poder, na sua maioria, iguais a 100%, mesmo se forem fixadas as situações com as menores diferenças entre os vetores médios $0{,}5\,\sigma_{\bar{X}}$. As exceções neste caso referem-se aos resultados para o critério de Pillai com $\alpha = 1\%$. O efeito da repetição foi mais marcante quando se considerou I =20, r =3, $\rho = 0{,}9$ e $0{,}5\,\sigma_{\bar{X}}$, cujo poder foi de 86,5% se comparado com a mesma situação com r = 5, cujo poder foi de 100%. Esse resultado uma vez mais reforça a idéia de que o critério de Pillai tende a ser menos poderoso que os demais, sendo mais afetado por este fator (repetição).

Chamou a atenção o fato de que nas situações com pequeno número de tratamentos (I = 5), pequeno número de repetições (r = 3) e pequenas diferenças entre vetores média ($0{,}5\,\sigma_{\bar{X}}$), o poder de todos os critérios, a exceção do critério de Roy, foi baixo. Nessa situação, todos os poderes encontrados para estes 3 critérios foram inferiores a 25%. O critério de Roy, por sua vez, apresentou poderes da ordem de 50% ou mais.

4.2.2 Efeito da correlação residual

O efeito da correlação residual nos quatro critérios alternativos usados para os testes de hipótese multivariados, pode ser visto nas Tabelas 4, 5 e 6. É conveniente observar que este efeito deve ser avaliado tomando-se o cuidado de fixar os demais fatores (I, r e $\sigma_{\bar{X}}$). Nestas tabelas fica evidente que o aumento da correlação residual provocou uma redução no poder de todos os critérios. O critério de Roy, em grande parte das situações, foi o menos influenciado pelo incremento do valor de ρ. Esse fato foi mais evidente nas situações com maior número de tratamentos e maior número de repetições.

Os resultados do poder dos quatro critérios, se for considerada a situação com menor número de tratamentos (I = 5), menor número de repetições (r = 3), 0,5 $\sigma_{\bar{X}}$ de diferença entre vetores média de tratamentos e α = 5%, podem ser observados na Figura 4. Verificou-se que o critério de Roy apresentou sempre maiores valores de poder e os quatro critérios apresentaram tendência de redução do poder com o aumento da correlação residual. Assim, o critério de Roy apresentou poder da ordem de 72% com $\rho = 0$ e reduzindo o valor do poder para 53% com $\rho = 0{,}9$. O critério de Wilks, por sua vez, apresentou poder de 18% com $\rho = 0$ e de 8,5% com $\rho = 0{,}9$.

Na Figura 5, a situação é semelhante a anterior, somente diferindo quanto ao número de repetições que passou a ser de r = 5. A mesma tendência de redução do poder foi observada com o aumento da correlação residual entre as variáveis estudadas. Destaca-se nessas Figuras o fato de que os critérios de Wilks, Hotelling e Pillai apresentaram poderes semelhantes entre si, mas com a tendência do critério de Hotelling ter os maiores poderes, seguido pelos critérios de Wilks e de Pillai, nessa ordem, se for considerada a magnitude dos poderes encontrados.

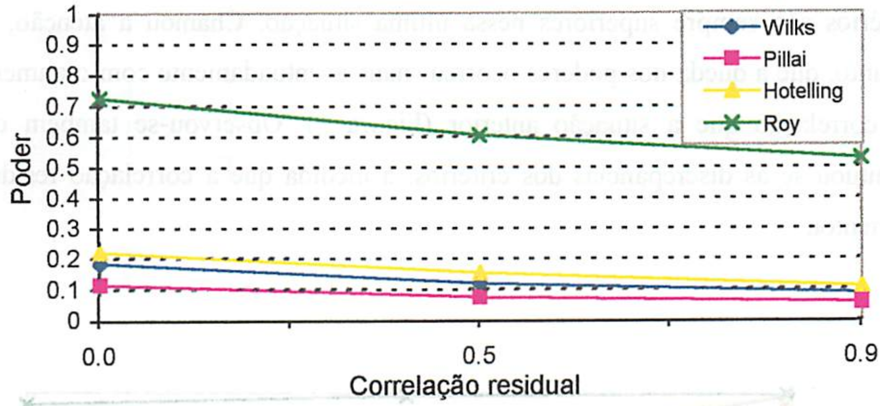

FIGURA 4. Poder dos quatro critérios multivariados considerando cinco tratamentos, três repetições e 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de correlação residual.

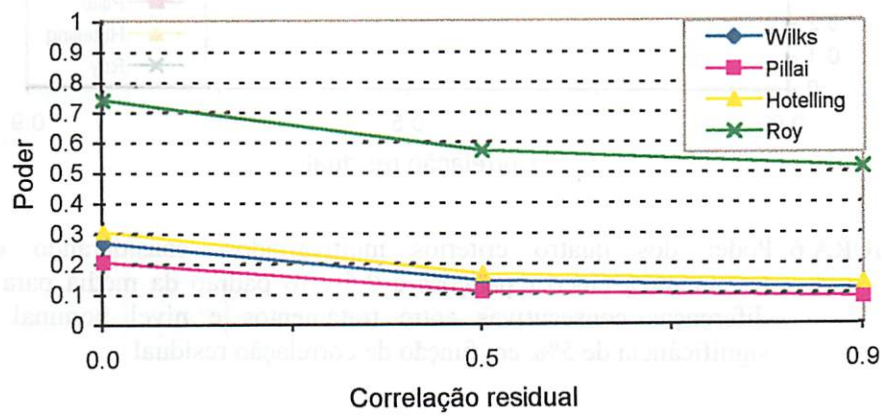

FIGURA 5. Poder dos quatro critérios multivariados considerando cinco tratamentos, cinco repetições e 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de correlação residual.

Para a situação com um número intermediário de tratamentos (I = 10) e um maior número de repetições (r = 5), a Figura 6 apresentou os poderes dos quatro critérios. É interessante observar que, se forem comparados os resultados da Figura 5 com os apresentados na Figura 6, os poderes dos respectivos critérios são sempre superiores nessa última situação. Chamou a atenção, no entanto, que a queda nos poderes ocorreu mais acentuadamente com o aumento da correlação que a situação anterior (Figura 5). Observou-se também que acentuou-se as discrepâncias dos critérios, à medida que a correlação residual aumentou.

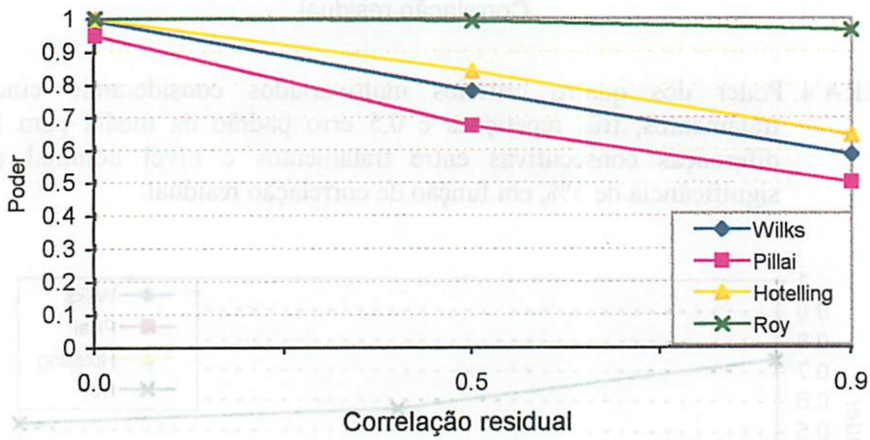

FIGURA 6. Poder dos quatro critérios multivariados, considerando dez tratamentos, cinco repetições e 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de correlação residual.

Considerando uma situação de maior diferença entre os vetores média de tratamentos ($1\sigma_{\bar{X}}$) com I = 10 tratamentos, r = 3 repetições e α = 5%, verificou-se que os critérios de Wilks, Hotelling e Roy apresentaram praticamente 100% de poder para todos os valores de ρ. O critério de Pillai, no entanto, apresentou acentuada taxa de decréscimo de poder com o aumento de ρ. Assim, para este critério, o poder, que foi de 100% com ρ = 0, passou a ser de 78,5% com ρ = 0,9 (Figura 7). A situação com I = 20 (Tabela 6), em quase todos os casos, apresentou praticamente 100% de poder, não permitindo assim verificar o efeito da correlação residual, excetuando a situação em que se considerou o critério de Pillai com α = 1%.

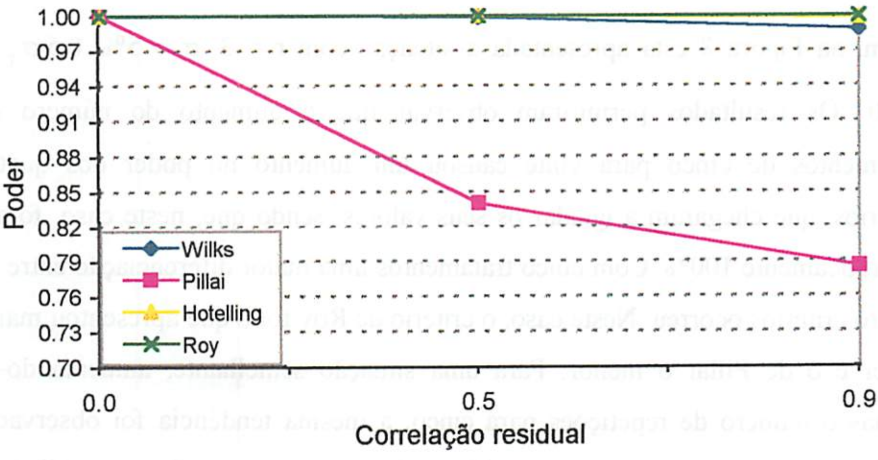

FIGURA 7. Poder dos quatro critérios multivariados considerando dez tratamentos, três repetições e um erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de correlação residual.

4.2.3 Efeito do número de tratamentos

Os efeitos do número de tratamentos para o poder dos quatro critérios multivariados podem ser observados nas Tabelas 4, 5 e 6, fixando-se os demais fatores estudados, como comentado nos casos anteriores.

Verificou-se, em todas as situações, que o aumento do número de tratamentos provocou grandes incrementos no poder dos critérios multivariados, tanto nos níveis nominais de 1%, quanto de 5%. Esse efeito é mais pronunciado se forem consideradas as situações de pequenas diferenças ($0{,}5\,\sigma_{\bar{X}}$) entre os vetores média de tratamentos.

Com a finalidade de salientar alguns pontos importantes e sem perdas de generalidade, serão destacadas a seguir algumas situações. Essas referem-se aos casos em que houve uma maior diferenciação entre os critérios, ou seja, aos casos de $\alpha = 5\%$ e $0{,}5\,\sigma_{\bar{X}}$, com $r = 3$ ou $r = 5$ e $\rho = 0$ ou $\rho = 0{,}9$. Procedendo-se assim, na Figura 8 está apresentada a situação com $r = 3$, $\alpha = 5\%$, $0{,}5\,\sigma_{\bar{X}}$ e $\rho = 0$. Os resultados permitiram observar que o aumento do número de tratamentos de cinco para vinte causou um aumento no poder dos quatro critérios, que chegaram a igualar os seus valores, sendo que, neste caso, foram de praticamente 100%. Com cinco tratamentos uma maior diferenciação entre os quatro critérios ocorreu. Neste caso, o critério de Roy foi o que apresentou maior poder e o de Pillai o menor. Para uma situação semelhante, aumentando-se apenas o número de repetições para cinco, a mesma tendência foi observada, sendo que com dez tratamentos os critérios já apresentavam poderes semelhantes e próximos de 100% (Figura 9).

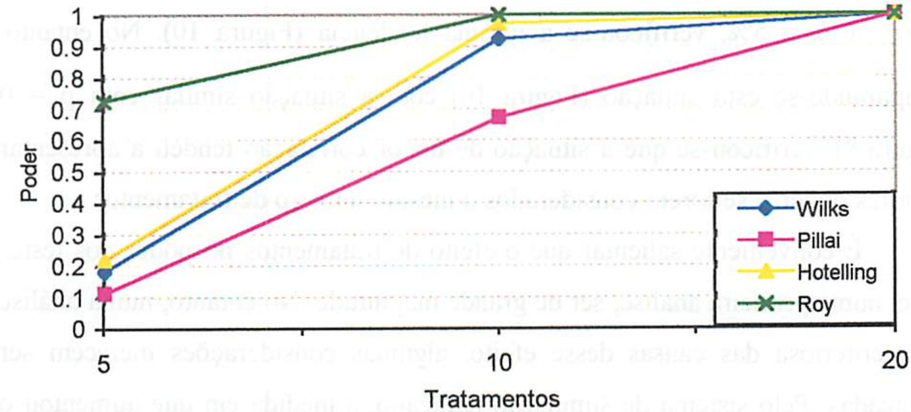

FIGURA 8. Poder dos quatro critérios multivariados considerando três repetições, zero de correlação e 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de número de tratamentos.

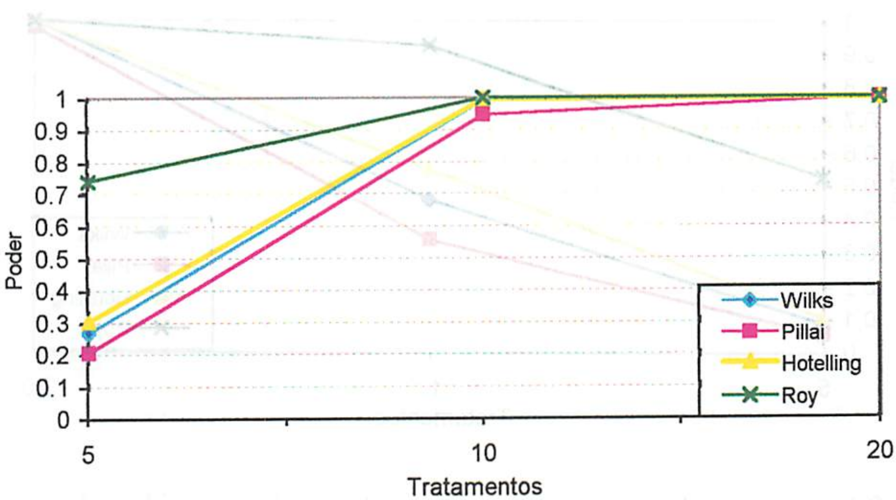

FIGURA 9. Poder dos quatro critérios multivariados considerando cinco repetições, zero de correlação e 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de número de tratamentos.

Com uma situação de maior correlação residual (ρ = 0,9), com r = 3, $0{,}5\sigma_{\bar{X}}$ e α = 5%, verificou-se a mesma tendência (Figura 10). No entanto, comparando-se esta situação (Figura 10) com a situação similar com ρ = 0 (Figura 8), verificou-se que a situação de maior correlação tendeu a apresentar menores poderes se forem considerados o mesmo número de tratamentos.

É conveniente salientar que o efeito de tratamentos no poder dos testes pode, numa primeira análise, ser de grande magnitude. No entanto, numa análise mais criteriosa das causas desse efeito, algumas considerações merecem ser destacadas. Pelo sistema de simulação realizado, à medida em que aumentou o número de tratamentos, a amplitude total paramétrica entre o maior e o menor vetor média de tratamentos aumentou de forma proporcional ao $\sigma_{\bar{X}}$.

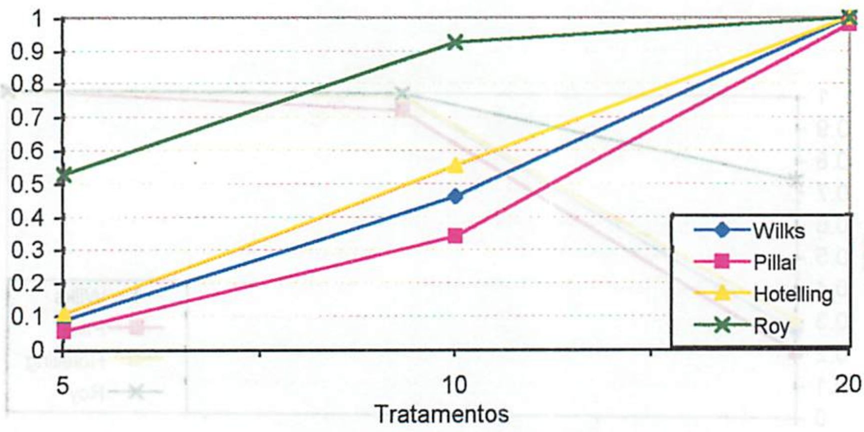

FIGURA 10. Poder dos quatro critérios multivariados considerando três repetições, 0,9 de correlação e 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos e nível nominal de significância de 5%, em função de número de tratamentos.

Essa constante de proporcionalidade foi igual a (I -1)k. Assim, a título de ilustração, em uma situação com diferenças consecutivas de k = 0,5 de $\sigma_{\bar{X}}$, com I = 5, a amplitude total foi de $2\sigma_{\bar{X}}$ e com I = 20 foi de $9,5\sigma_{\bar{X}}$. Portanto, o incremento de poder nestas circustâncias não pode ser atribuído exclusivamente ao incremento no número de tratamentos mas ao incremento nas diferenças a serem detectadas, uma vez que a amplitude total aumenta com o aumento do número de tratamentos. Posteriormente, nesse trabalho, serão discutidas as formas de contornar este problema e medir o efeito isolado do número de tratamentos utilizados.

4.2.4 Efeito da diferença entre médias

O poder dos critérios multivariados, bem como do teste F na análise de variância univariada, é afetado pelas diferenças existentes entre os tratamentos. As diferenças paramétricas entre tratamentos se constituem num dos principais causadores de diferenciação dos testes aplicados na estatística (Graybill, 1974; Searle, 1971; Anderson, 1984). Os resultados do poder dos critérios multivariados apresentados nas Tabelas 4, 5 e 6 permitem a realização de inferências a partir do aumento das diferenças existentes entre vetores média de tratamentos.

Assim, para ilustrar, tomando-se uma situação com I=5 tratamentos, r= 3 repetições, correlação residual entre variáveis igual a zero (ρ = 0) e α igual a 5%, houve uma tendência dos poderes dos quatro critérios de aumentar, à medida em que aumentaram as diferenças entre vetores média, dados em função dos erros padrão das médias. Esta situação pode ser visualizada na Figura 11. Neste caso, verificou-se que as mesmas tendências de classificação dos critérios foram observadas, ou seja, o critério mais poderoso foi o de Roy, seguido do critério de Hotelling, Wilks e Pillai. O aumento das diferenças entre tratamentos consecutivos aumentou a diferenciação entre os critérios de Wilks, Roy e Hotelling com relação ao critério de Pillai. Por outro lado, com pequenas diferenças consecutivas entre tratamentos ($0{,}5\,\sigma_{\bar{x}}$), o critério de Roy teve maior poder se comparado com os demais.

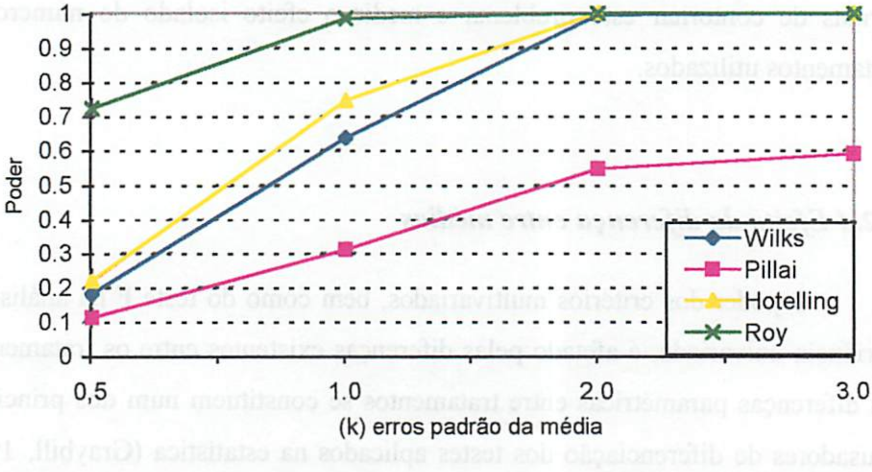

FIGURA 11. Poder dos quatro critérios multivariados considerando cinco tratamentos, três repetições, zero de correlação, com três variáveis e nível nominal de significância de 5%, em função das diferentes quantidades (k) dos erros padrão da média.

Nas Figuras 12 e 13 a mesma situação foi representada, variando apenas a correlação residual. A Figura 12 apresenta o poder dos quatro critérios para $\rho = 0{,}5$ e a Figura 13 para $\rho = 0{,}9$. Os comentários feitos anteriormente se aplicam a estas situações, porém, o que se verifica é que a magnitude do poder é reduzida com o aumento da correlação residual.

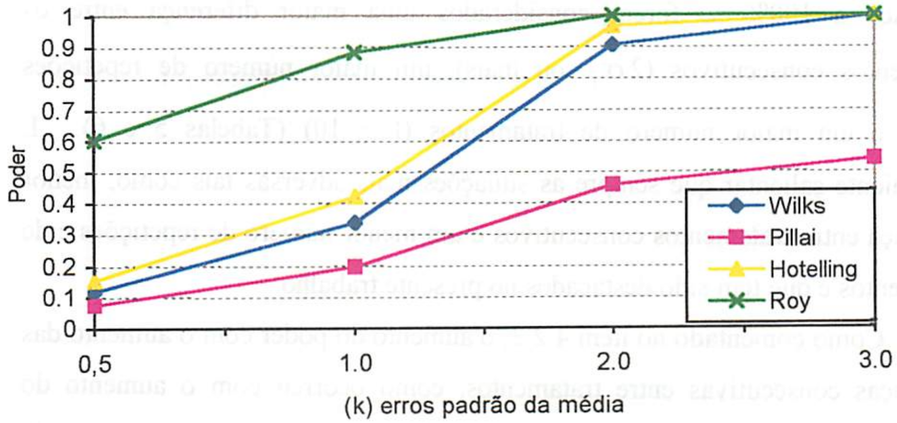

FIGURA 12. Poder dos quatro critérios multivariados considerando cinco tratamentos, três repetições, 0,5 de correlação, com três variáveis e nível nominal de significância de 5%, em função das diferentes quantidades (k) dos erros padrão da média.

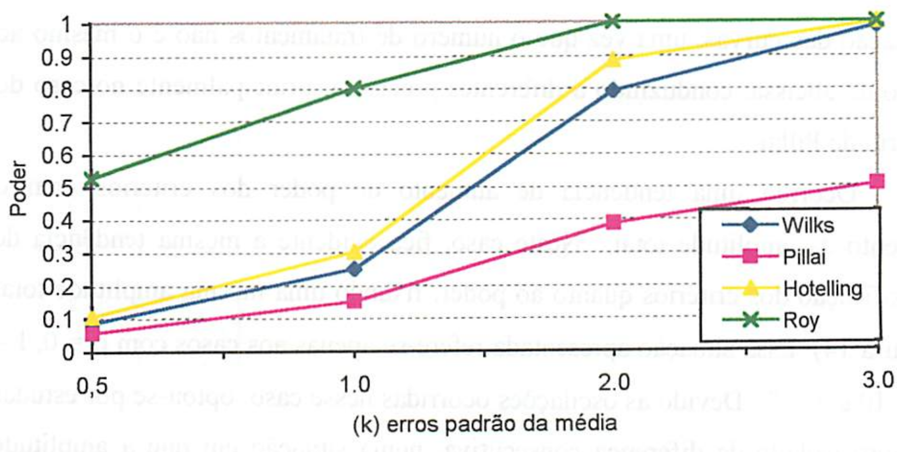

FIGURA 13. Poder dos critérios multivariados considerando cinco tratamentos, três repetições, 0,9 de correlação, com três variáveis e nível nominal de significância de 5%, em função das diferentes quantidades (k) dos erros padrão da média.

De forma geral, os quatro critérios apresentarem poderes elevados próximos a 100% se forem considerados uma maior diferença entre os tratamentos consecutivos ($2\sigma_{\bar{X}}$ ou mais), um maior número de repetições ($r = 5$) e um maior número de tratamentos ($I \geq 10$) (Tabelas 5 e 6). É conveniente salientar que sempre as situações mais adversas tais como, menor diferença entre tratamentos consecutivos e um menor número de repetições e de tratamentos é que têm sido destacados no presente trabalho.

Como comentado no item 4.2.3, o aumento do poder com o aumento das diferenças consecutivas entre tratamentos, como ocorreu com o aumento do número de tratamentos, não deve ser atribuído exclusivamente ao aumento do número de tratamentos. O aumento do poder se deve ao aumento nas diferenças a serem detectadas, uma vez que fixado o número de tratamentos, o aumento do valor de k faz com que a amplitude total seja incrementada nessa mesma razão. Dessa forma, a Figura 14 apresenta o poder dos quatro critérios em função da amplitude total expressa por $(I\text{-}1)k\sigma_{\bar{X}}$. Nessa figura, observa-se uma maior oscilação das curvas, uma vez que o número de tratamentos não é o mesmo ao longo da abcissa, conduzindo a diferentes precisões, principalmente no caso do critério de Pillai.

Ocorreu uma tendência de aumento de poder dos critérios com o aumento da amplitude total. Neste caso, fica evidente a mesma tendência de classificação dos critérios quanto ao poder, fixando uma mesma amplitude total (Figura 14). Essa situação apresentada refere-se apenas aos casos com $\rho = 0$, $I = 5$ ou 10 e $r = 3$. Devido às oscilações ocorridas nesse caso, optou-se por estudar o efeito isolado da diferença consecutiva, numa situação em que a amplitude total se mantivesse constante. Para solucionar o problema resultante, foram elaboradas situações em que os tratamentos foram divididos em dois grupos, sendo que em cada grupo todos possuiam efeitos iguais e entre grupos existia uma diferença de $k\sigma_{\bar{X}}$ (k =0,5; 1; 2 ou 3). Essa situação foi apresentada na Tabela 7.

Os resultados encontrados fixando-se o número de tratamentos e de repetições, permitiram que se inferisse a respeito do poder de cada critério em realmente detectar diferenças entre tratamentos ou entre grupos de tratamentos nas magnitudes pré-estabelecidas. Assim, para ilustrar uma situação dessas, com I =10, r = 3 e ρ = 0 o critério de Wilks apresentou poder de 7,85% com $0,5\sigma_{\bar{X}}$ de diferenças entre grupos e poder de 94,25% com $3\sigma_{\bar{X}}$ de diferença.

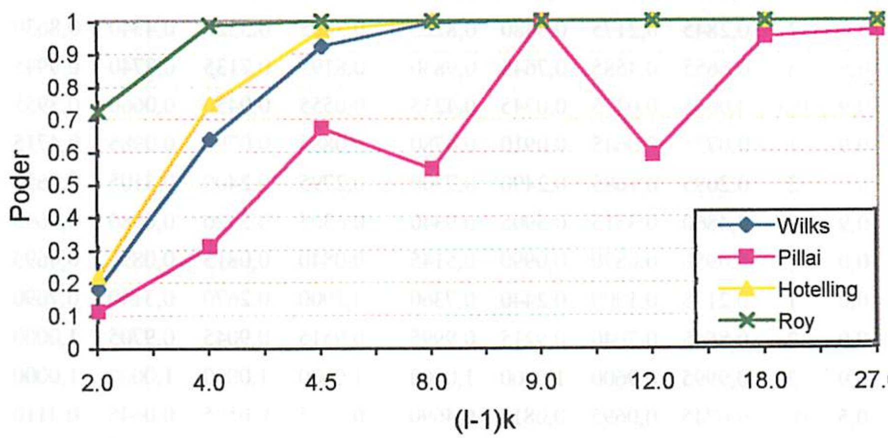

FIGURA 14. Poder dos quatro critérios em função do fator multiplicador da amplitude total (I-1)k, para I = 5 e 10, r=3 e ρ = 0, obtidos por simulação.

[redacted]

TABELA 7. Poder para os quatro critérios adotados para o teste da hipótese de igualdade dos vetores de efeitos de tratamentos, considerando p=3 variáveis e diferentes valores de I (tratamentos) e diferentes valores de r (repetições), ρ (correlação) e do fator multiplicador (k) do erro padrão da média ($\sigma_{\bar{x}}$), obtidos por meio de simulação, para o nível nominal de significância de 5% e situações em que os tratamentos foram divididos em dois grupos

			r = 3				r = 5			
I	ρ	k	Wilks	Pillai	Hotelling	Roy	Wilks	Pillai	Hotelling	Roy
10	0,0	0,5	0,0785	0,0630	0,0935	0,4570	0,0820	0,0710	0,0940	0,4580
10	0,0	1	0,1465	0,1230	0,1650	0,6390	0,1905	0,1625	0,2125	0,6550
10	0,0	2	0,5705	0,3900	0,6785	0,9700	0,7620	0,6575	0,8130	0,9845
10	0,0	3	0,9425	0,6775	0,9830	1,0000	0,9950	0,9680	0,9980	1,0000
10	0,5	0,5	0,0675	0,0535	0,0795	0,4275	0,0685	0,0570	0,0740	0,4270
10	0,5	1	0,0885	0,0800	0,1005	0,5090	0,1070	0,0965	0,1220	0,5090
10	0,5	2	0,2845	0,2175	0,3480	0,8225	0,3865	0,3320	0,4340	0,8630
10	0,5	3	0,6655	0,4585	0,7645	0,9830	0,8195	0,7135	0,8740	0,9945
10	0,9	0,5	0,0645	0,0475	0,0745	0,4235	0,0555	0,0450	0,0660	0,3955
10	0,9	1	0,0735	0,0645	0,0910	0,4780	0,0885	0,0790	0,0985	0,4715
10	0,9	2	0,2095	0,1685	0,2490	0,7300	0,2795	0,2400	0,3105	0,7650
10	0,9	3	0,4860	0,3515	0,5905	0,9340	0,6305	0,5420	0,6980	0,9695
20	0,0	0,5	0,0950	0,0870	0,0990	0,5145	0,0840	0,0815	0,0895	0,4695
20	0,0	1	0,2175	0,1900	0,2440	0,7360	0,2900	0,2670	0,3140	0,7690
20	0,0	2	0,8645	0,7040	0,9215	0,9995	0,9515	0,9045	0,9705	1,0000
20	0,0	3	0,9995	0,9600	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
20	0,5	0,5	0,0745	0,0695	0,0815	0,4690	0,0615	0,0595	0,0645	0,4110
20	0,5	1	0,1230	0,1130	0,1345	0,5510	0,1400	0,1335	0,1485	0,5885
20	0,5	2	0,5145	0,4270	0,5800	0,9350	0,6175	0,5530	0,6565	0,9650
20	0,5	3	0,9070	0,7535	0,9540	1,0000	0,9765	0,9440	0,9875	1,0000
20	0,9	0,5	0,0680	0,0625	0,0755	0,4525	0,0555	0,0540	0,0630	0,3830
20	0,9	1	0,1015	0,0910	0,1105	0,4935	0,1215	0,1120	0,1280	0,5360
20	0,9	2	0,3630	0,3020	0,4200	0,8625	0,4340	0,3865	0,4680	0,8915
20	0,9	3	0,7410	0,5895	0,8315	0,9950	0,8795	0,8225	0,9160	0,9995
Média			0,3644	0,2935	0,4029	0,7243	0,4199	0,3867	0,4417	0,7294
Amplitude			0,0645	0,0475	0,0745	0,4235	0,0555	0,0450	0,0645	0,3830
			0,9995	0,9600	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Com a finalidade de expressar, de uma forma mais clara e de fácil interpretação, a comparação do poder dos critérios sob o efeito da diferença entre os dois grupos, algumas figuras foram geradas. A Figura 15 apresenta o poder dos critérios em função da diferença entre grupos, para I = 10 tratamentos, r = 3 repetições, ρ = 0 e α = 5%. Verificou-se que os critérios de Wilks, Hotelling e Pillai apresentaram praticamente os mesmos poderes da ordem de 10% com $0,5\,\sigma_{\bar{X}}$ de diferenças entre grupos, sendo que a partir daí (mais de $0,5\,\sigma_{\bar{X}_i}$ de diferença) eles passaram a ter uma classificação diferenciada. O critério de Wilks sempre teve poderes inferiores ao de Hotelling, apresentando diferenças pequenas. Este critério foi sempre superior ao de Pillai, diferenciando pouco nas situações de $0,5\,\sigma_{\bar{X}}$ ou de $1\,\sigma_{\bar{X}}$ e diferenciando-se mais nas situações de mais de $1\,\sigma_{\bar{X}}$ de diferença entre tratamentos. O critério de Roy nessa situação sempre apresentou os maiores poderes, sendo que a sua diferenciação dos demais foi mais evidente com pequenos valores para a diferença entre grupos de tratamentos.

Com valores de dois erros padrão de diferença, o poder do critério de Wilks é da ordem de 57,05%. Quando o número de repetições foi aumentado para cinco na mesma circunstância, o poder do critério de Wilks passou a ser da ordem de 76,20% (Figura 16). Porém, se a correlação residual entre as variáveis for aumentada para 0,9, o poder deste critério passou para ordem de 27,95% (Figura 17). Resultados semelhantes podem ser verificados para os demais critérios.

Do exposto, pode-se inferir que a repetição tem um papel importante no poder dos testes, como já comentado, ocorrendo o mesmo com a correlação residual. O aumento de r provoca um aumento no poder do teste, mas um aumento no valor de ρ causas reduções no mesmo, sendo que este último efeito foi mais pronunciado. Isto pode ser explicado porque, uma vez fixada a situação ideal (Figura 16), dentre as três comentadas, o poder foi de 76,20%; a Figura 15 apresentou, para o critério de Wilks, 57,05% de poder e mudou apenas o número de repetições e a Figura 17 apresentou 30%, mudando apenas o valor de ρ, considerando $2\sigma_{\bar{X}}$ de diferença entre grupos. Em todas as circunstâncias, os critérios de Wilks, Hotelling e Roy apresentaram elevado poder, da ordem de 96,9% ou mais, quando se considerou $3\sigma_{\bar{X}}$ de diferença entre os 2 grupos, excetuando-se as situações de maior correlação residual.

Se forem comparadas as situações de maior e menor diferença entre grupos, verificou-se que o efeito desse fator foi relevante, como era esperado, para o poder do teste e para a classificação dos critérios. Como já comentado, o aumento da diferença entre grupos provocou uma maior separação de três dos quatro critérios estudados (Wilks, Hotelling e Pillai) e diminuiu a diferença destes para o critério de Roy, principalmente nas situações com maior número de repetições e ausência de correlação (Figura 16).

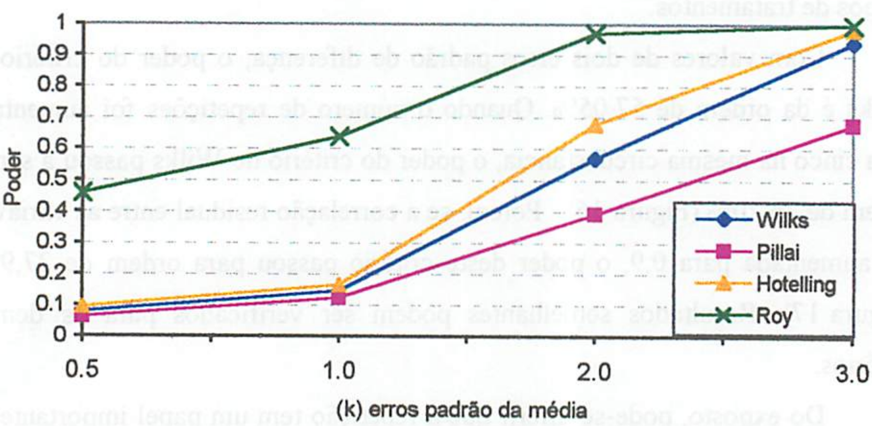

FIGURA 15. Poder dos quatro critérios multivariados considerando dez tratamentos, três repetições, zero de correlação, com três variáveis e nível nominal de significância de 5%, na situação em que os tratamentos foram divididos em 2 grupos, em função das diferentes quantidades (k) dos erros padrão da média.

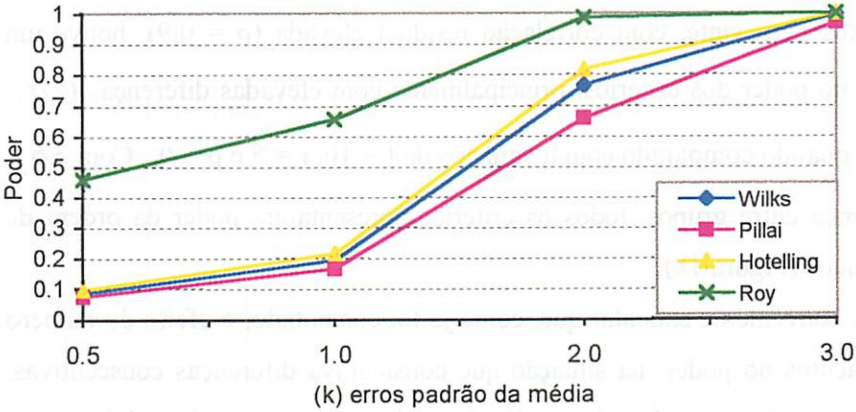

FIGURA 16. Poder dos quatro critérios multivariados considerando dez tratamentos, cinco repetições, zero de correlação, com três variáveis e nível nominal de significância de 5%, na situação em que os tratamentos foram divididos em 2 grupos, em função das diferentes quantidades (k) dos erros padrão da média.

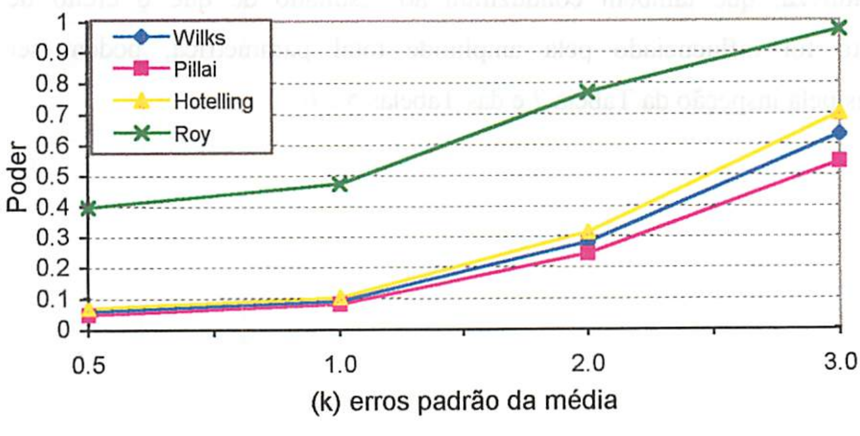

FIGURA 17. Poder dos quatro critérios multivariados considerando dez tratamentos, cinco repetições, 0,9 de correlação, com três variáveis e nível nominal de significância de 5%, na situação em que os tratamentos foram divididos em dois grupos, em função das diferentes quantidades (k) dos erros padrão da média.

Chama atenção o fato de que com r = 5 e um aumento no número de tratamentos para vinte, com correlação residual elevada (ρ = 0,9), houve um aumento no poder dos critérios, principalmente com elevadas diferenças ($2\sigma_{\bar{X}}$ ou mais) quando comparado com a situação de I = 10, r = 5 e ρ = 0. Com $3\sigma_{\bar{X}}$ de diferença entre grupos, todos os critérios apresentaram poder da ordem de 80% ou mais (Figura 18).

É conveniente salientar que, como já foi comentado, o efeito do número de tratamentos no poder, na situação que considerava diferenças consecutivas, estava mascarado em parte pela amplitude total que aumentava à medida que se aumentava o valor de I. Assim como se observa na Figura 8, tomando-se por base o critério de Wilks, verifica-se que com aumento do número de tratamentos de dez para vinte, o poder de 90% passou para 100%. Considerando-se dois grupos e a mesma diferença de $0{,}5\sigma_{\bar{X}}$, o poder de 7,85% com dez tratamentos aumentou para 9,5%, com vinte tratamentos (Figura 19). Outras observações dessa natureza, que também conduziram ao resultado de que o efeito de tratamento foi influenciado pela amplitude total paramétrica, podem ser realizadas pela inspeção da Tabela 7 e das Tabelas 5 e 6.

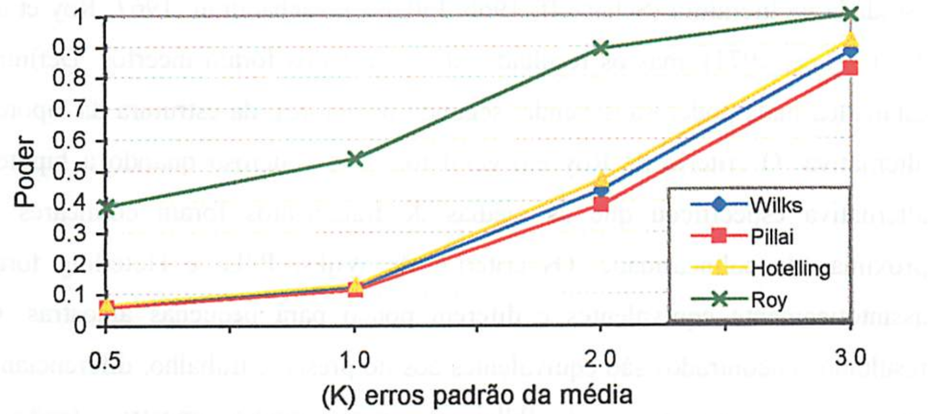

FIGURA 18. Poder dos quatro critérios multivariados considerando vinte tratamentos, cinco repetições, 0,9 de correlação, com três variáveis e nível nominal de significância de 5%, na situação em que os tratamentos foram divididos em dois grupos, em função das diferentes quantidades (k) dos erros padrão da média.

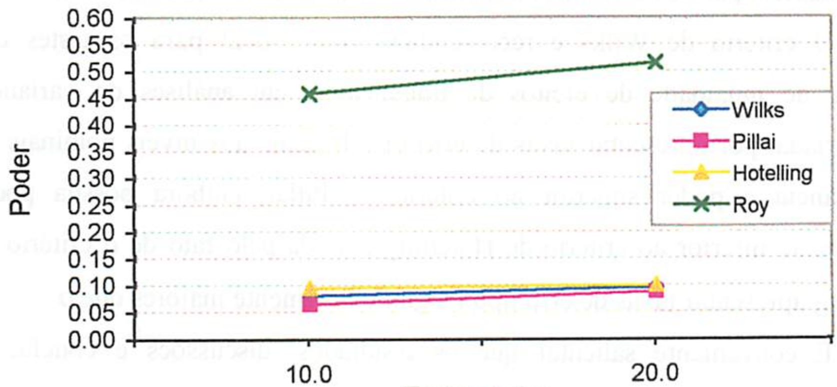

FIGURA 19. Poder dos quatro critérios multivariados considerando três repetições, zero de correlação, 0,5 erro padrão da média para as diferenças consecutivas entre tratamentos, com três variáveis e nível nominal de significância de 5%, na situação em que os tratamentos foram divididos em dois grupos, em função do número de tratamentos.

A comparação entre o poder dos quatro critérios multivariados foi estudada na literatura (Schatzoff, 1966; Pillai e Jayachandran, 1967; Roy et al., 1971 e Lee, 1971), mas os resultados desses estudos foram incertos. Definir a estatística mais poderosa depende, segundo os autores, da estrutura da hipótese alternativa. O critério de Roy em geral foi mais poderoso quando a hipótese alternativa especificou que as médias de tratamentos foram colineares ou próximas da colinearidade. Os critérios de Wilks, Pillai e Hotelling foram assintoticamente equivalentes e diferem pouco para pequenas amostras. Os resultados encontrados são equivalentes aos do presente trabalho, diferenciando apenas quanto ao critério de Pillai que, para grandes amostras, tendeu a diferenciar dos demais.

Olsen (1974) comparou essas estatísticas e suas propriedades de robustez e favoreceu ligeiramente o critério de Pillai, embora novamente tenham ocorrido pequenas diferenças com relação aos demais critérios. O autor se posicionou contra o uso do critério de Wilks quando a igualdade de estrutura de covariância não se mantém para os diferentes tratamentos, que foi a situação estudada.

O critério de Wilks é recomendado como ideal para os testes das hipótese de igualdade de efeitos de tratamentos, em análises de variância multivariada, por apresentar taxas de erro tipo I iguais aos níveis nominais de significância e poder superior ao critério de Pillai, embora possua poder ligeiramente inferior ao critério de Hotelling e ainda pelo fato de o critério de Hotelling apresentar taxas de erro tipo I significativamente maiores que α.

É conveniente salientar que os resultados, discussões e conclusões obtidos no presente trabalho referem-se única e exclusivamente aos casos em que os resíduos possuem distribuição normal multivariada e os tratamentos possuem a mesma estrutura de correlação.

5 CONCLUSÕES

- As taxas de erro tipo I do critério de Wilks são iguais ou inferiores às taxas nominais de significância adotadas em todas as situações estudadas.
- O critério de Pillai apresenta taxas de erro tipo I que tendem a ser menores que os níveis nominais, principalmente para situações de menor número de tratamentos e de repetições, e taxas de erro tipo I iguais aos níveis nominais adotados nas outras situações.
- O critério de Hotelling apresenta tendência, em algumas circunstâncias, de apresentar taxas de erro tipo I superiores aos níveis nominais e tendência geral de apresentar maiores taxas de erro tipo I do que os critérios de Wilks e Pillai.
- O critério de Roy, em todas as situações, apresenta taxas de erro tipo I superiores aos níveis nominais, não podendo ser recomendado para avaliar a hipótese de igualdade de efeitos de vetores média de tratamentos da análise de variância multivariada.
- A estrutura de correlação entre as variáveis não afeta as taxas de erro tipo I.
- O número de repetições, o número de tratamentos, a correlação residual e as diferenças a serem detectadas influenciam no poder dos quatro critérios multivariados.
- O critério de Roy apresenta maior poder, seguido pelo critério de Hotelling, Wilks e Pillai. O critério de Pillai, que apresenta menor poder, diferencia-se dos critérios de Wilks e Hotelling nas situações de maior diferença entre tratamentos e de maior correlação residual entre as variáveis.

- O critério de Wilks é considerado ideal para os testes de hipótese de igualdade de efeitos de vetores média de tratamentos da análise de variância multivariada.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, T.W. **An introduction to multivariate statistical analysis.** 2ª ed. New York : John Wiley, 1984. 675p.

BOCK, R.D. **Multivariate statistical methods in behavioral research.** New York : McGraw Hill. 1975. 623p.

CHATFIELD, C.; COLLINS, A.J. **Introduction to Multivariate Analysis.** Londres: Chapman e Hall, 1980. 246p.

COSTA, S.F. **Introdução ilustrada à estatística (com muito humor).** 2ª ed. São Paulo : Harbra, 1992. 303p.

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

DAGNELIE, P. **Analyse statistique à plusieurs variables.** 2ª Reimpressão, Gembloux: Les Presses Agronomiques, 1986. 362 p.

DEMÉTRIO, C.G.B. **Análise multidimensional para dados de cana-de-açúcar**. Piracicaba : ESALQ/USP, 1985. 144p. (Tese - doutorado em Agronomia).

DURBIN, J.; KOOPMAN, S.J. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. **Biometrika,** v. 84, n.3, p. 669-684, 1997.

FERREIRA, D.F. **Eficiência de métodos de mapeamento de locos quantitativos (QLTs) e da seleção assistida por marcadores através de simulação.** Piracicaba : ESALQ/USP, 1995. 210p. (Tese - doutorado em Genética).

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

GRAYBILL, F.A. **Introduction to matrices with aplications in statistics.** Belmont: Wadsworth Publishing Company, 1974. 371p.

HARRIS, R.J. **A Primer of Multivariate Statistics.** New York: Academic Press, 1975. 331p.

HOTELLING, H. A generalized T test and measure of multivariate dispersion. **Proceedings of the Second Berkeley Simposium on Mathematical Statistics and Probability.** University of California press, p.23-41. 1951.

HUBER, D.A.; WHITE,G.R.; HODGE,G.R. Variance component estimation techniques compared for two mating designs wih forest genetic architecture through computer simulation. **Theoretical Applied Genetics,** Berlin, v.88, p.236-42. 1994.

JAMES, F.C.; McCULLOCH, C.E. Multivariate analysis in ecology and systematics: Panacea or pandora's box? **Annual Review Ecology Systematic,** v.21, p.129-66, 1990.

JOHNSON, R. A., WICHERN, D.W. **Applied multivariate statistical analysis**. New Jersey : Prentice-Hall. Inc, 1992. 642p.

LAWLEY, D.N. A generalization of Fisher's z-Test. **Biometrika,** Cambridge, v.30, p.180-187, 1938.

LEE, Y.S. Asymptotic formulae for the distribuiton of a multivariate test statistic: power comparisons of certain multivariate tests. **Biometrika**, v.58 n.3, p.647-651, 1971.

LEEMIS, L.M.; TRIVEDI, K.S. A comparison of approximate interval estimators for the bernoulli parameter. **The American Statistician,** v.50, n.1, p.63-68. February, 1996.

MANLY, B.F.J. **Multivariate statistical methods: A primer.** London : Chapman and Hall, 1986. 159p.

MARDIA, K.V. The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. **Biometrika**, v.58, n.1, p.105-121. 1971.

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

MIKHAIL, N.N. A comparison of tests of the Wilks-Lawley hypothesis in multivariate analysis. **Biometrika,** v.52, n.1,2. p.149-156. 1965.

MOOD, A .M.; GRAYBILL, F. A.; BOES, D.C. **Introduction to the theory of statistics**. 3ª ed.Tokyo: McGraw-Hill, 1974. 564p.

NAYLOR, T.H.; BALINTFY, J.L.; BURDICK, D.S.; CHU, K. **Técnicas de simulação em computadores**. Petrópolis: Vozes, 1971. 402p.

OLSEN, C.L. Comparative robustness of six test in multivariate analysis of variance. **Journal of the American Statistical Association**. v.69, n.348, p.894-908. 1974.

PILLAI, K.C.S. Some new test criteria in multivariate analysis. **Annals of Mathematical Statistics,** v.26, p.117-21, 1955.

PILLAI, K.C.S.; JAYACHANDRAN, K. Power comparisons of tests of two mltivariate hypotheses based on four criteria. **Biometrika,** v.54, n.1, p.195-210. 1967.

RODDA, D.D.; SCHAEFFER,L.R.; MULLEN,K., FRIARS,G.W. Measuring the precision of genetic parameters by a simulation technique. **Theoritical and Applied Genetics**, Berlin, v.51, p.35-9, 1977.

RODRIGUES JR. A.L. **Probabilidade dos erros tipo I e II dos testes para delineamento cross-over 2x2 de respostas binárias : estimação pelo método de Monte Carlo.** Piracicaba : ESALQ/USP. 88p. 1995. (Dissertação- Mestrado em Agronomia)

ROTHENBERG, T.J. Edgeworth expansions for multivariate test statistics, IP-255, Center for Research in Management Science, University of California, **Berkeley**. 1977.

ROY. S.N. On a heuristic method of test construction and its use in multivariate analysis. **Annals of Mathematical Statistics,** v.24, p.220-238. 1953.

ROY, S.N.; GNANADESIKAN, R.; SRIVASTAVA, J.N. Analysis and design of certain quantitative multiresponse experiments. **Oxford: Pergamon Press.** 1971.

SAS Institute. **SAS Language and Procedures : Usage**. Version 6, 1st ed. Cary NC. SAS Institute Inc., 1995. 373p.

SCHATZOFF, M. Sensitivity comparisons among tests of the general linear hypothesis. **Journal of the Amererican Statistical Association,** v.61, p. 415-35. 1966.

SCHATZOFF, M. Exact distributions of wilks's likelihood ratio criterion. **Biometrika,** v.53, n.3 e 4, p. 347-358. 1966b.

SEARLE, E.R. **Linear Models.** New York : John Wiley, 1971. 531p.

SILVA, E.C. da . **Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo.** Lavras : UFLA. 54p. 1998. (Dissertação - Mestrado em Agronomia).

TATSUOKA, M.M. **Multivariate Analysis: Techniques for Educational and Psychological Research.** New York : John Wiley, 1971. 310p.

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

VEIGA, R.D. **Eficiência dos dialelos circulantes na escolha de genitores, avaliada com simulação de dados.** Lavras : UFLA, 95p. 1998. (Tese - doutorado em Genética).

VAN VLECK, L.D. Algorithms for simulation of animal models with multiple traits and nonadditive genetic effects. **Journal Series, Nebraska Agricultural Research Division,** 1994. 15p. (No prelo)

WALTER, J.P.; MAO, I.L. Multiple and single trait analysis for estimating genetic parameters in simulated populations under selection. **Journal of Dairy Science,** Champaign, v.68, p.91-8, 1985.

WILKS, S.S. Certain generalizations in the analysis of variance. **Biometrics.** Raleigh, v.24, p. 471-494. 1932.

WINER, B.J. **Statistical principles in Experimental Design.** 2ª ed. New York : McGraw-Hill Book Co, 1971. 907p.

ANEXOS

```
/******************************************************
 * Programa - para testar as taxas de erro tipo I  *
 * 5 trat. 5 rep. 3 variáveis; correl=0,5; out/98  *
******************************************************/
options ps = 500 nodate nonumber ls=85;
data expm553;
   do exp = 1 to 2000;
    do trat = 1 to 5;
    do rep = 1 to 5;

       ee1=rannor(111212);
       ee2=rannor(132235);
       ee3=rannor(124563);

        e1=ee1*5;
        e2=ee1*5+ee2*8.660254;
        e3=ee1*10+ee2*5.773507+ee3*16.329932;

        p11=100+e1;
        p22=100+e2;
        p33=100+e3;
        output;
      end;
    end;
  end;

Proc anova;
  Class trat;
  model p11 p22 p33 = trat/nouni;
  manova h=trat;
  by exp;
run; quit;
```

$$\Sigma = \begin{bmatrix} 25 & 25 & 50 \\ 25 & 100 & 100 \\ 50 & 100 & 400 \end{bmatrix}$$

os erros e1, e2 e e3 são multiplicados pelo fator de cholesky (L)

$$L = \begin{bmatrix} 5 & 0 & 0 \\ 5 & 8,660254 & 0 \\ 10 & 5,773507 & 16,329932 \end{bmatrix}$$

```
/*********************************************************
*  Programa - para testar as taxas de erro tipo I      *
* 20 trat. 5 rep., 3 variáveis e correl=zero - out/98 *
*********************************************************/
options ps = 500 nodate nonumber ls=85;
data exp2053;
   do exp = 1 to 2000;
     do trat = 1 to 20;
       do rep = 1 to 5;

        ee1=rannor(131212);
        ee2=rannor(112235);
        ee3=rannor(132541);

        e1=ee1*5;
        e2=ee2*10;
        e3=ee3*20;

        p11=100+e1;
        p22=100+e2;
        p33=100+e3;
        output;
       end;
     end;
   end;

Proc anova;
  Class trat;
  model p11 p22 p33 = trat/nouni;
  manova h=trat;
  by exp;
run; quit;
```

```
/**********************************************************
* Programa -Simulação para testar o poder dos 4 critérios *
* 10 trat. 5 rep e 3 variáveis; correl=0,0; erro padrão=2 *
**********************************************************/
options ps = 500 nodate nonumber ls=85;
data t105Z2;
   do exp = 1 to 2000;
     do trat = 1 to 10;
       do rep = 1 to 5;
        ee1=rannor(160212);
        ee2=rannor(112562);
        ee3=rannor(88112);
        e1=ee1*5;
        e2=ee2*10;
        e3=ee3*20;

      if trat =1 then do;
        p11=100-20.16+e1; p22=100-40.23+e2; p33=100-80.46+e3;    end;
     if trat =2 then do;
        p11=100-15.68+e1; p22=100-31.29+e2; p33=100-62.58+e3;    end;
     if trat =3 then do;
        p11=100-11.20+e1; p22=100-22.35+e2; p33=100-44.70+e3;    end;
     if trat =4 then do;
        p11=100-6.72+e1; p22=100-13.41+e2; p33=100-26.82+e3;     end;
     if trat =5 then do;
        p11=100-2.24+e1; p22=100-4.47+e2; p33=100-8.94+e3;       end;
     if trat =6 then do;
        p11=100+2.24+e1; p22=100+4.47+e2; p33=100+8.94+e3;       end;
     if trat =7 then do;
        p11=100+6.72+e1; p22=100+13.41+e2; p33=100+26.82+e3;     end;
     if trat =8 then do;
        p11=100+11.20+e1; p22=100+22.35+e2; p33=100+44.70+e3;    end;
     if trat =9 then do;
        p11=100+15.68+e1; p22=100+31.29+e2; p33=100+62.58+e3;    end;
     if trat =10 then do;
        p11=100+20.16+e1; p22=100+40.23+e2; p33=100+80.46+e3;    end;
        output;
       end;
     end;
   end;

Proc anova;
  Class trat;
  model p11 p22 p33 = trat/nouni;
  manova h=trat;
  by exp;
run; quit;
```

```
/*****************************************************************
*Programa - Simulação para testar o poder dos 4 critérios      *
*20 trat.3 rep., 3 variáveis correl= 0,5; desvio padrão = 0,5 *
*****************************************************************/
options ps = 500 nodate nonumber ls=85;
data t203m05;
   do exp = 1 to 2000;
     do trat = 1 to 20;
       do rep = 1 to 3;

        ee1=rannor(102132);
        ee2=rannor(195692);
        ee3=rannor(135012);

        e1=ee1*5;
        e2=ee1*5+ee2*8.660254;
        e3=ee1*10+ee2*5.773507+ee3*16.329932;

     if trat =1 then do;
p11=100-13.78+e1; p22=100-27.46+e2;  p33=100-54.9+e3; end;
     if trat =2 then do;
p11=100-12.33+e1; p22=100-24.57+e2;  p33=100-49.13+e3;end;
     if trat =3 then do;
p11=100-10.9+e1; p22=100-21.7+e2;  p33=100-43.4+e3;   end;
     if trat =4 then do;
p11=100-9.425+e1; p22=100-18.79+e2; p33=100-37.57+e3; end;
     if trat =5 then do;
p11=100-7.975+e1; p22=100-15.90+e2; p33=100-31.79+e3; end;
     if trat =6 then do;
p11=100-6.5+e1; p22=100-13.01+e2;  p33=100-26.01+e3;  end;
     if trat =7 then do;
p11=100-5.075+e1; p22=100-10.12+e2; p33=100-20.23+e3; end;
     if trat =8 then do;
p11=100-3.625+e1; p22=100-7.23+e2;  p33=100-14.45+e3; end;
     if trat =9 then do;
p11=100-2.175+e1; p22=100-4.34+e2; p33=100-8.67+e3;   end;
     if trat =10 then do;
p11=100-0.725+e1; p22=100-1.45+e2;  p33=100-2.89+e3;  end;
     if trat =11 then do;
p11=100+0.725+e1; p22=100+1.45+e2;  p33=100+2.89+e3;  end;
     if trat =12 then do;
p11=100+2.175+e1; p22=100+4.34+e2;  p33=100+8.67+e3;  end;
     if trat =13 then do;
p11=100+3.625+e1; p22=100+7.23+e2;  p33=100+14.45+e3; end;
```

```
      if trat =14 then do;
 p11=100+5.075+e1; p22=100+10.12+e2; p33=100+20.23+e3; end;
      if trat =15 then do;
p11=100+6.525+e1; p22=100+13.01+e2; p33=100+26.01+e3; end;
      if trat =16 then do;
p11=100+7.975+e1; p22=100+15.90+e2; p33=100+31.79+e3; end;
      if trat =17 then do;
p11=100+9.425+e1; p22=100+18.79+e2; p33=100+37.57+e3; end;
      if trat =18 then do;
p11=100+10.88+e1; p22=100+21.68+e2; p33=100+43.35+e3; end;
      if trat =19 then do;
p11=100+12.325+e1; p22=100+24.57+e2; p33=100+49.13+e3; end;
      if trat =20 then do;
p11=100+13.78+e1; p22=100+27.46+e2; p33=100+54.91+e3; end;

          output;
        end;
      end;
    end;

 Proc anova;
   Class trat;
   model p11 p22 p33 = trat/nouni;
   manova h=trat;
   by exp;
 run; quit;
```