



**TESTE DE NORMALIDADE E AVALIAÇÃO  
POR SIMULAÇÃO MONTE CARLO DO  
PROBLEMA DE BEHRENS-FISHER NO  
ESPAÇO  $p$ -DIMENSIONAL**

**MARCELO ANGELO CIRILLO**

**2003**

54951

MFN046973

**MARCELO ANGELO CIRILLO**

**TESTE DE NORMALIDADE E AVALIAÇÃO POR  
SIMULAÇÃO MONTE CARLO DO PROBLEMA  
DE BEHRENS-FISHER NO ESPAÇO  $p$ -DIMENSIONAL**

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

Orientador

Prof. Daniel Furtado Ferreira

**LAVRAS  
MINAS GERAIS – BRASIL  
2003**

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

Cirillo, Marcelo Angelo

Teste de normalidade e avaliação por simulação Monte Carlo do Problema de Behrens-Fisher no espaço p-Dimensional / Marcelo Angelo Cirillo. -- Lavras : UFLA, 2003.

128 p. : il.

Orientador: Daniel Furtado Ferreira.

Tese (Mestrado) – UFLA..

Bibliografia.

1. Behrens-Fisher. 2. Heterogeneidade. 3. Poder. 4. Taxa de erro tipo 1.  
5. Normalidade multivariada. 6. Q-Qplot. 7. Quantis. I. Universidade Federal de Lavras. II. Título.

CDD-519.282

-519.5

**MARCELO ANGELO CIRILLO**

**TESTE DE NORMALIDADE E AVALIAÇÃO POR  
SIMULAÇÃO MONTE CARLO DO PROBLEMA  
DE BEHRENS-FISHER NO ESPAÇO p-DIMENSIONAL**

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

Aprovada em sexta feira, 07 de Fevereiro de 2003.

Prof. Dr. Joel Augusto Muniz

UFLA

Prof. Dr. Marcelo Tavares

UFU

Prof. Dr. Paulo César Lima

UFLA



Prof. Dr. Daniel Furtado Ferreira

Dex – UFLA

(Orientador)

**LAVRAS  
MINAS GERAIS – BRASIL  
2003**

## **Dedico**

**Aos  
meus pais**

**Euclides Cirillo e  
Alice Malassise Cirillo**

**Às minhas irmãs**

**Marcia Simone Cirillo  
Eveline Simara Cirillo**

## AGRADECIMENTOS

A **DEUS**, por ter me dado saúde, força e coragem para a realização deste trabalho.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas pela oportunidade e pela confiança no meu trabalho.

Ao CNPq pela concessão da bolsa de estudos.

Ao Prof. Dr. Daniel Furtado Ferreira pela amizade, dedicação, orientação e grande participação na minha formação acadêmica.

Aos professores do DEX/UFLA, Augusto, Delly, Fortunato, Joel, Júlio, Luiz Henrique, Lucas, Mário e Thelma, pelo conhecimento e disponibilidade em ensinar.

Aos funcionários do DEX/UFLA.

Ao grande amigo Paulo Dattori por sempre ter procurado “abrir caminhos” e também por ter acreditado na minha capacidade.

Ao grande amigo Wagner Cavalli pelo apoio e aconselhamento.

Aos novos amigos, Cilinha, Fidélis, Flora, Hedwiges, Lívia, Marinho, Sonia, Simone e Dora.

Aos amigos Daniel, Ítalo, Jaqueline, Edgard, Juliano, Mônica, Paulo José, Tutia, Valesca e Wallace.

Aos amigos de turma, Airton, Alessandra, Ana Lúcia, Elcio, José Ermelindo, João Marcos, José Otaviano, Luís Alberto, Lúcia, Waldemar, Washington.

Em especial a Juliana Vigilato, que passou a fazer parte da minha vida.

Enfim, a todos que incentivaram a realização deste trabalho.

## LISTA DE ABREVIATURAS

<b>ns.</b>	nível de significância
<b>sim.</b>	simetria
<b>cur.</b>	curtose
<b>Dist.</b>	Distribuição
<b>U.</b>	Uniforme
<b>Exp.</b>	Exponencial

## LISTA DE SÍMBOLOS

<b>p</b>	número de variáveis.
<b>n</b>	tamanho amostral.
<b>n<sub>1</sub></b>	tamanho da amostra referente à primeira população.
<b>n<sub>2</sub></b>	tamanho da amostra referente à segunda população.
<b>ρ</b>	correlação.
<b>δ</b>	grau de heterogeneidade.
<b>k</b>	erros padrão.
<b>T<sup>2</sup>H(C)</b>	T <sup>2</sup> de Hotelling convencional.
<b>T<sup>2</sup>H(B)</b>	T <sup>2</sup> de Hotelling bootstrap.
<b>K(C)</b>	Método de Kim via convencional.
<b>K(B)</b>	Método de Kim via bootstrap.
<b>J(C)</b>	Método de Johansen via convencional.
<b>J(B)</b>	Método de Johansen via bootstrap.
<b>N(C)</b>	Método de Nel e Van der Merwe via convencional.
<b>N(B)</b>	Método de Nel e Van der Merwe via bootstrap.
<b>Y(C)</b>	Método de Yao via convencional.
<b>Y(B)</b>	Método de Yao via bootstrap.

## SUMÁRIO

Lista de Abreviaturas .....	i
Lista de Símbolos.....	ii
Lista de Tabelas.....	iii
Lista de Figuras .....	v
Resumo.....	viii
Abstract .....	ix
Introdução Geral.....	x
Referências Bibliográficas .....	xii
<b>CAPÍTULO 1 .....</b>	<b>01</b>
Resumo.....	02
Abstract .....	03
1 - Introdução .....	04
2 - Referencial Teórico.....	07
2.1 – Testes univariados de normalidade .....	07
2.2 – Testes multivariados de normalidade.....	18
3 - Metodologia.....	24
3.1 – Obtenção dos quantis univariados Q-Qplots .....	25
3.2 – Obtenção dos quantis multivariados Q-Qplots.....	26
3.3 – Avaliação do poder do teste de normalidade.....	28
4 - Resultados e Discussão .....	30
4.1 – Quantis do Q-Qplot .....	30
4.2 – Taxa do Erro tipo I .....	43
4.3 – Poder dos testes .....	46
5 - Conclusões .....	54
6- Referências Bibliográficas .....	55

CAPÍTULO 2 .....	59
Resumo.....	60
Abstract .....	61
1 – Introdução .....	62
2 – Referencial Teórico.....	65
2.1 – Comparação de duas médias populacionais .....	65
2.1.1 – Caso univariado.....	65
2.1.2 – Caso multivariado .....	69
3 – Metodologia .....	76
3.1 – Obtenção das matrizes de covariâncias $\Sigma_1$ e $\Sigma_2$ .....	76
3.2 – Simulação das populações.....	78
3.3 – Descrição dos métodos a serem avaliados .....	79
4 – Resultados e Discussão .....	82
4.1 – Taxa do erro tipo I.....	82
4.2 – Poder dos métodos .....	104
5 – Conclusões .....	120
6 – Referências Bibliográficas .....	121
ANEXOS .....	124

## LISTA DE TABELAS

<b>TABELA 2.1.1</b> – Forma geral da tabela de contingência utilizada no teste de aderência qui-quadrado em que, $O_i$ é a freqüência observada da classe $i$ , $E_i$ é a freqüência esperada $i$ , $c$ é o número total de classes para $i=1,2,\dots,c$ .....	09
<b>TABELA 4.1.1</b> – Quantis para o teste coeficiente de correlação $r_q$ definido para as amostras $n=3$ até $n=30$ – caso univariado .....	32
<b>TABELA 4.1.2</b> – Quantis para o teste coeficiente de correlação $r_q$ definido para as amostras $n=35$ até $n=5000$ – caso univariado .....	33
<b>TABELA 4.1.3</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=2$ ) .....	34
<b>TABELA 4.1.4</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=3$ ) .....	35
<b>TABELA 4.1.5</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=4$ ) .....	36
<b>TABELA 4.1.6</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=5$ ) .....	37
<b>TABELA 4.1.7</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=6$ ) .....	38
<b>TABELA 4.1.8</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=7$ ) .....	39
<b>TABELA 4.1.9</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=8$ ) .....	40
<b>TABELA 4.1.10</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=9$ ) .....	41
<b>TABELA 4.1.11</b> – Quantis para o teste coeficiente de correlação $r_q$ para o caso multivariado ( $p=10$ ) .....	42

<b>TABELA 4.2.1</b> – Taxa do erro tipo I ( $p=1$ ) para o caso univariado e bivariado ( $p=2$ ).....	44
<b>TABELA 4.2.2</b> – Taxa do erro tipo I para o caso multivariado .....	46
<b>TABELA 4.3.1</b> – Avaliação do poder do teste coeficiente de correlação $r_q$ , simetria e curtose para os casos univariado ( $p=1$ ) e bivariado ( $p=2$ ) com amostras pequenas.....	50
<b>TABELA 4.3.2</b> – Avaliação do poder do teste coeficiente de correlação $r_q$ , simetria e curtose para os caso univariados ( $p=1$ ) e bivariado ( $p=2$ ) com amostras maiores.....	51
<b>TABELA 4.3.3</b> – Avaliação do poder do teste do coeficiente de correlação $r_q$ simetria e curtose para o caso multivariado ( $p=6$ ) e ( $p=10$ ).....	52
<b>TABELA 4.3.4</b> – Avaliação do poder do teste do coeficiente de correlação $r_q$ simetria e curtose para os casos multivariado ( $p=6$ ) e ( $p=10$ ) para amostras maiores .....	53
<b>TABELA 2.1.2.1</b> – Resumo das combinações dos fatores considerados na avaliação feita por Christensen e Rencher no estudo do problema de Behrens-Fisher .....	75
<b>TABELA 4.2.1</b> - Comparação do poder para todas as amostras em função do aumento do número de variáveis em conjunto com o aumento da heterogeneidade de covariâncias e da diferença entre os vetores de médias em função do erro padrão para o valor nominal de 5% e 1% considerando a correlação $\rho=0,5$ e $p=2$ .....	118
<b>TABELA 4.2.2</b> - Comparação do poder para todas as amostras em função do aumento do número de variáveis em conjunto com o aumento da heterogeneidade de covariâncias e da diferença entre os vetores de médias em função do erro padrão para o valor nominal de 5% e 1% considerando a correlação $\rho=0,5$ e $p=5$ .....	119

## LISTA DE FIGURAS

<b>FIGURA 4.1.1</b> - Taxa de erro tipo I, fixado o valor nominal em 5% e $p=2$ , comparando homogeneidade com máxima heterogeneidade .....	86
<b>FIGURA 4.1.1(a)</b> – Homogeneidade, correlação $\rho=0,2$ .....	86
<b>FIGURA 4.1.1(b)</b> – Homogeneidade, correlação $\rho=0,9$ .....	86
<b>FIGURA 4.1.1(c)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,2$ .....	86
<b>FIGURA 4.1.1(d)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,9$ .....	86
<b>FIGURA 4.1.2</b> - Taxa de erro tipo I, fixado o valor nominal em 1% e $p=2$ comparando homogeneidade com máxima heterogeneidade .....	89
<b>FIGURA 4.1.2(a)</b> – Homogeneidade, correlação $\rho=0,2$ .....	89
<b>FIGURA 4.1.2(b)</b> – Homogeneidade, correlação $\rho=0,9$ .....	89
<b>FIGURA 4.1.2(c)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,2$ .....	89
<b>FIGURA 4.1.2(d)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,9$ .....	89
<b>FIGURA 4.1.3</b> - Taxa de erro tipo I, fixado o valor nominal em 5% e $p=5$ comparando homogeneidade com máxima heterogeneidade .....	94
<b>FIGURA 4.1.3(a)</b> – Homogeneidade, correlação $\rho=0,2$ .....	94
<b>FIGURA 4.1.3(b)</b> – Homogeneidade, correlação $\rho=0,9$ .....	94
<b>FIGURA 4.1.3(c)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,2$ .....	94
<b>FIGURA 4.1.3(d)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,9$ .....	94
<b>FIGURA 4.1.4</b> - Taxa de erro tipo I, fixado o valor nominal em 1% e $p=5$ comparando homogeneidade com máxima heterogeneidade .....	98
<b>FIGURA 4.1.4(a)</b> – Homogeneidade, correlação $\rho=0,2$ .....	98
<b>FIGURA 4.1.4(b)</b> – Homogeneidade, correlação $\rho=0,9$ .....	98
<b>FIGURA 4.1.4(c)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,2$ .....	98
<b>FIGURA 4.1.4(d)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,9$ .....	98

<b>FIGURA 4.1.5</b> - Taxa de erro tipo I, fixado o valor nominal em 5% e $p=5$ comparando graus de heterogeneidade .....	101
<b>FIGURA 4.1.5(a)</b> – Heterogeneidade ( $\delta=2$ ), correlação $\rho=0,2$ .....	101
<b>FIGURA 4.1.5(b)</b> – Heterogeneidade ( $\delta=2$ ), correlação $\rho=0,5$ .....	101
<b>FIGURA 4.1.5(c)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,2$ .....	101
<b>FIGURA 4.1.5(d)</b> – Heterogeneidade ( $\delta=8$ ), correlação $\rho=0,5$ .....	101
<b>FIGURA 4.1.6</b> – Taxa de erro tipo I fixado o valor nominal em 5%, $\delta=8$ e $p=5$ comparando graus de correlações .....	104
<b>FIGURA 4.1.6(a)</b> – $n_1=10$ e $n_2=15$ .....	104
<b>FIGURA 4.2.5(b)</b> – $n_1=15$ e $n_2=10$ .....	104
<b>FIGURA 4.2.1</b> – Poder do teste para amostras de tamanhos iguais, fixado o valor nominal em 5%, $p=2$ , $k=1$ comparando graus de correlações sobre as situações de Homogeneidade e máxima Heterogeneidade.....	107
<b>FIGURA 4.2.1(a)</b> – Homogeneidade, $n_1=10$ e $n_2=10$ .....	107
<b>FIGURA 4.2.1(b)</b> – Homogeneidade, $n_1=30$ e $n_2=30$ .....	107
<b>FIGURA 4.2.1(c)</b> – Heterogeneidade( $\delta=8$ ), $n_1=10$ e $n_2=10$ .....	107
<b>FIGURA 4.2.1(d)</b> – Heterogeneidade( $\delta=8$ ), $n_1=30$ e $n_2=30$ .....	107
<b>FIGURA 4.2.2</b> - Poder do teste para amostras de tamanhos iguais, fixado o valor nominal em 1%, $p=2$ , $k=1$ comparando graus de correlações sobre a situações de Homogeneidade e máxima Heterogeneidade.....	110
<b>FIGURA 4.2.2(a)</b> – Homogeneidade, $n_1=10$ e $n_2=10$ .....	110
<b>FIGURA 4.2.2(b)</b> – Homogeneidade, $n_1=30$ e $n_2=30$ .....	110
<b>FIGURA 4.2.2(c)</b> – Heterogeneidade( $\delta=8$ ), $n_1=10$ e $n_2=10$ .....	110
<b>FIGURA 4.2.2(d)</b> – Heterogeneidade( $\delta=8$ ), $n_1=30$ e $n_2=30$ .....	110
<b>FIGURA 4.2.3</b> - Poder do teste para amostras de tamanhos iguais, fixado o valor nominal em 5%, $p=5$ , $k=2$ comparando graus de correlações sobre a situações de Homogeneidade e máxima Heterogeneidade.....	112

<b>FIGURA 4.2.3(a)</b> – Homogeneidade, $n_1=10$ e $n_2=10$ .....	112
<b>FIGURA 4.2.3(b)</b> – Homogeneidade, $n_1=30$ e $n_2=30$ .....	112
<b>FIGURA 4.2.3(c)</b> – Heterogeneidade( $\delta=8$ ), $n_1=10$ e $n_2=10$ .....	112
<b>FIGURA 4.2.3(d)</b> – Heterogeneidade( $\delta=8$ ), $n_1=30$ e $n_2=30$ .....	112
<b>FIGURA 4.2.4</b> - Poder do teste para amostras de tamanhos iguais, fixado o valor nominal em 1%, $p=5$ , $k=3$ comparando graus de correlações sobre as situações de Homogeneidade e máxima Heterogeneidade.....	114
<b>FIGURA 4.2.4(a)</b> – Homogeneidade, $n_1=10$ e $n_2=10$ .....	114
<b>FIGURA 4.2.4(b)</b> – Homogeneidade, $n_1=30$ e $n_2=30$ .....	114
<b>FIGURA 4.2.4(c)</b> – Heterogeneidade( $\delta=8$ ), $n_1=10$ e $n_2=10$ .....	114
<b>FIGURA 4.2.4(d)</b> – Heterogeneidade( $\delta=8$ ), $n_1=30$ e $n_2=30$ .....	114
<b>FIGURA 4.2.5</b> - Poder do teste fixado o valor nominal em 5%, $k=3$ , $p=5$ e $\delta=8$ comparando as correlações com diferentes tamanhos amostrais .....	115
<b>FIGURA 4.2.5(a)</b> – $n_1=10$ e $n_2=15$ .....	115
<b>FIGURA 4.2.5(b)</b> – $n_1=30$ e $n_2=30$ .....	115

## RESUMO

CIRILLO, Marcelo Angelo. **Teste de Normalidade e avaliação por simulação Monte Carlo do problema de Behrens-Fisher no espaço p-dimensional**. 2003. 128p. (Dissertação – Mestrado em Agronomia / Estatística Experimentação Agropecuária ) – Universidade Federal de Lavras, Lavras-MG.\*

A inferência estatística se resume em testes de hipóteses e estimação de parâmetros populacionais. Neste contexto, um dos testes mais utilizados é o teste conhecido por t-student. A realização deste teste pressupõe normalidade e, preferencialmente, homogeneidade das variâncias. No caso multivariado, o teste  $T^2$  de Hotelling é definido como uma generalização do teste t-student, portanto, as suposições requeridas no caso univariado para o teste t-student são as mesmas para o caso multivariado, ou seja, para a realização do teste  $T^2$  de Hotelling é necessário assumir normalidade multivariada e, preferencialmente, homogeneidade entre as matrizes de covariâncias. No entanto, a inferência sobre estas suposições multivariadas são mais restritas e menos citadas na literatura. Desta forma, neste trabalho objetivou-se apresentar, no capítulo 2, a expansão do teste do coeficiente de correlação  $r_q$  (Q-Qplot) como um teste alternativo para o diagnóstico de normalidade multivariada e, no capítulo 3, avaliar o problema de Behrens-Fisher por meio de simulação Monte Carlo utilizando métodos propostos por diversos autores, visando corrigir o efeito da heterogeneidade das matrizes de covariâncias. Estes métodos foram analisados na abordagem convencional e por meio dos processos de reamostragem conhecidos por bootstrap.

---

\* Comitê Orientador: Daniel Furtado Ferreira (Orientador)

## ABSTRACT

**CIRILLO, Marcelo Angelo. Normality test and Monte Carlo Simulation evaluation of the Behrens- Fisher problem in the p-dimensional space.** Lavras: UFLA, 2003. 128p. (Dissertation – Master Program in Agronomy / Major in Statistics and Agricultural Experimentation) – Federal University of Lavras \*

The statistical inference is limited to hypothesis tests and estimates of population parameters . In this context, one of the most employed test is the test known by t-student. The accomplishment of this test presupposes normality and preferentially homogeneity of variances. In the case multivariate, Hotteling's  $T^2$  test is defined over a generalization of the t-student test therefore, the suppositions required in the univariate case for the t-student test are the same as for the multivariate case, that is, for the accomplishment of Hotelling's  $T^2$  test is necessary to suppose a multivariate normality and, preferentially, homogeneity among the covariance matrixes. Nevertheless, the inference on these multivariate suppositions are more restrict and less quoted in the literature. Thus, in this work presenting in chapter 2 the expansion of the correlation coefficient test  $r_q$  (Q-Qplot) as an alternative test for the diagnosis of multivariate normality and in chapter 3 evaluating the Beheren-Fisher's problem by means of Monte Carlo simulation by utilizing methods proposed by a number of authors, aiming to correct the effect of heterogeneity of the covariance matrices. These methods were assessed in the conventional approach and by means of the re-sampling processes known for bootstrap.

---

\* Guidance Committe: Daniel Furtado Ferreira (Orientador)

## INTRODUÇÃO GERAL

A comparação de duas médias populacionais no caso univariado é feita pelo teste t-student, o qual pressupõe que as médias sejam provenientes de duas populações normais e, preferencialmente, que as variâncias populacionais sejam homogêneas. A importância da suposição de normalidade ser satisfeita implica na estimação precisa de intervalos de confiança. Do grau de homogeneidade das variâncias populacionais depende a qualidade da aproximação entre a distribuição amostral e a verdadeira distribuição t de Student.

Para o caso multivariado segue-se o mesmo princípio, porém estendido para p-dimensões. As médias populacionais são representadas por vetores de médias, sendo que cada componente refere-se a uma variável inerente à população estudada. A suposição de homogeneidade é analisada por meio das matrizes de covariâncias e o teste comumente usado é conhecido por  $T^2$  de Hotelling, sendo este uma extensão do teste t-student utilizado no caso univariado. Desta forma, é perceptível a facilidade da análise multivariada empregada para a comparação de médias populacionais em virtude de que esta análise representa apenas uma expansão do caso univariado. Entretanto, a verificação das suposições é bastante complexa em relação ao caso univariado. Para efeito de comparação entre as duas abordagens (univariado e multivariado), pode-se citar que na literatura existem vários métodos propostos para diagnosticar normalidade univariada, como, por exemplo, análise gráfica Q-Qplots e testes de hipóteses para normalidade: Shapiro Wilk, Lilliefors, Kolmogorov Smirnov. No caso multivariado, as alternativas para diagnosticar normalidade multivariada são mais restritas e menos citadas na literatura, sendo que o procedimento mais utilizado é dado por meio dos coeficientes de simetria e curtose de Mardia. Em relação à homogeneidade, é clara a diferença entre as situações univariadas e multivariada. No caso univariado, quando se tem

heterogeneidade de variâncias, é conveniente aplicar as correções de Satterthwaite (1946) e Cochran & Cox (1957) nos graus de liberdades da estatística  $t$  de student. Em relação ao caso multivariado, em que a heterogeneidade das matrizes de covariâncias está presente devem ser considerados vários fatores na amenização do efeito de heterogeneidade tais como estruturas de correlação, grau de heterogeneidade, tamanhos de amostras associadas a maior e menor variância generalizada. A questão da heterogeneidade das matrizes de covariâncias atualmente é alvo de pesquisa. Recentemente, autores como Christensen & Rencher (1997) e Fouladi & Yockey (2002) estudaram este problema em diversos fatores, mas ainda é um amplo campo de pesquisa.

Em virtude do que foi mencionado, basicamente este trabalho consiste em apresentar a expansão do teste do coeficiente de correlação  $r_q$  (Q-Qplot) para o diagnóstico de normalidade multivariada e avaliar métodos propostos por diversos autores visando corrigir a heterogeneidade das matrizes de covariâncias na abordagem convencional e por meio dos processos de reamostragem conhecidos por bootstrap.

## REFERÊNCIAS BIBLIOGRÁFICAS

CHRISTENSEN, W. F; RENCHER, A. C. A Comparison of type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem. **Communications in Statistics Simulation and Computation**, New York, v. 26, n. 4, p. 1251-1273, 1997.

COCHRAN, W. G.; COX, G. M. **Experimental designs** 2. ed. Singapore: John Wiley & Sons, 1957. 611 p.

FOULADI, R. T.; YOCKEY, R. D. Type I Error Control of Two-Group Multivariate Tests on Means Under Conditions of Heterogeneous Correlation Structure and Varied Multivariate Distributions. **Communications in Statistics**, New York, v. 31, n. 3, p. 375-400, 2002.

# **CAPÍTULO 1**

## **Extensão do Teste para Normalidade, baseado no Coeficiente de Correlação Quantil-Quantil para o caso Multivariado**

## RESUMO

**CIRILLO, Marcelo Angelo. Extensão do Teste para Normalidade, baseado no Coeficiente de Correlação Quantil-Quantil para o caso Multivariado. 2003. 58p. (Dissertação – Mestrado em Agronomia / Estatística Experimentação Agropecuária ) – Universidade Federal de Lavras, Lavras.\***

A inferência estatística consiste em fazer generalizações sobre uma ou mais populações com base nos dados amostrais. Neste contexto, todo o processo de inferência resume-se nos testes de hipóteses de parâmetros populacionais, sendo primordial assumir que os dados são provenientes de uma população normal, pois esse modelo populacional explica muitos fenômenos aleatórios. No caso multivariado, os mesmos princípios do caso univariado são necessários, portanto é de grande importância a normalidade dos dados ou dos resíduos, a qual generalizada para  $p$ -dimensões ( $p > 1$ ) é conhecida por distribuição normal multivariada. Uma alternativa para a identificação de normalidade é dada por meios gráficos (Q-Qplots), no entanto, para o caso multivariado a análise gráfica não é suficiente, de modo que, a utilização de testes estatísticos como meio de inferir a normalidade é necessária. Desta forma, este capítulo objetivou estender o teste do coeficiente de correlação  $r_q$ , apresentado por Johnson & Wichern (1988), para amostras maiores, bem como propor a expansão deste teste para o caso de  $p$ -dimensões, baseado nas distâncias generalizadas de Mahalanobis, apresentando os valores críticos para diferentes tamanhos de amostras e número de variáveis, utilizando a simulação Monte Carlo. Assim, foram obtidas as tabelas para os quantis 0,01; 0,05 e 0,10, variando o número de variáveis ( $p$ ) em  $1 \leq p \leq 10$ . Entre os principais resultados obtidos, constatou-se que o teste do coeficiente  $r_q$  é recomendando para amostras pequenas, corrigindo a limitação assintótica do teste multivariado de simetria e curtose. Em relação ao poder, o teste coeficiente  $r_q$  mostrou-se poderoso principalmente para grandes amostras, porém, seu poder foi reduzido a medida que aumentou o número de variáveis ( $p$ ). O controle da taxa de erro tipo I não apresentou problemas pelo fato de mesmo controlar a taxa de erro tipo I por construção.

---

\* Comitê Orientador: Daniel Furtado Ferreira (Orientador)

## ABSTRACT

CIRILLO, Marcelo Angelo. **Extended of the test for normality based upon the Quantil-Quantil correlation coefficient for hte multivariate case.** Lavras: UFLA, 2003. 58p. (Dissertation – Master Program in Agronomy / Major in Statistics and Agricultural Experimentation)\* – Federal University of Lavras

The statistical inference consists in doing generalizations on a population on the basis of the sample data. In this context, all the inference purpose is limited to the tests of population parameter hypotheses, supposing that the data stem from a normal population being vital, for that population model explain a great deal of random effects. In the multivariate model, the same principles of the univariate case are needed, therefore the normality of the data or of the residues is of great importance, which generalized for  $p$ -dimensions ( $p > 1$ ), is known as multivariate normal distribution. An alternative to the identification of normality is given by graphic means (Q-Qplots); nevertheless, for the multivariate case, the graphic analysis is not enough, so that the utilization of statistical tests as a mean of inferring normality is needed. Thus, this chapter is designed to extend the correlation coefficient test  $r_q$ , presented by Johnson & Wichern (1988), to the larger samples as well as to propose the expansion of this test to the case of  $p$ -dimensions based upon the generalized distances of Mahalanobis, presenting the critical values for different sample sizes and number of variables by utilizing the Monte Carlo simulation. So, the tables for quantils 0.01, 0.05 and 0.10 were obtained, ranging the number of variables ( $p$ ) by  $1 \leq p \leq 10$ . Among the main results obtained, it was found that the coefficient test  $r_q$  is recommended for small samples, correcting the asymptotic limitation of the multivariate test of symmetry and kurtosis. Relative to power, the coefficient  $r$  test proved powerful mainly for large samples, nevertheless, its power was reduced as the number of variables ( $p$ ) increased. The control of the rate of type I error presented no problems by the fact that it controls the rate of type I error by construction.

---

\* Guidance Committee: Daniel Furtado Ferreira (Orientador)

## 1-INTRODUÇÃO

A inferência estatística consiste em fazer generalizações sobre uma ou mais populações com base nos dados amostrais. Neste contexto, todo o processo de inferência resume-se na estimação e testes de hipóteses de parâmetros populacionais. Em ambos os procedimentos presume-se que os dados sejam provenientes de uma população normal pelo fato de muitas distribuições estatísticas serem aproximadamente normais, independentemente da forma da distribuição da população original. Uma outra vantagem da distribuição normal refere-se ao fato de seu modelo populacional apresentar-se coerente a muitos fenômenos aleatórios.

Para a realização de inferências no caso multivariado, seguem-se os mesmos princípios do caso univariado, portanto é de grande importância a normalidade dos dados, a qual, generalizando para muitas dimensões, é conhecida por distribuição normal multivariada.

A identificação de normalidade nos dados, seja no caso univariado ou multivariado, em geral é feita por meio de gráficos, entretanto, a simples constatação via gráficos não é suficiente, principalmente no caso multivariado e, especificamente, nas situações de muitas variáveis. Sendo assim, a utilização de testes estatísticos como meio de inferir sobre a normalidade é necessária. Entre os testes mais utilizados no caso univariado e encontrados em várias literaturas estão o teste de aderência qui-quadrado; Kolmogorov-Smirnov; Lilliefors e Shapiro-Wilk. No caso multivariado, em geral os testes existentes são extensões daqueles propostos para o caso univariado.

A escolha do teste que o pesquisador julga apropriada é dada em vários aspectos como o poder do teste, o controle da taxa do erro tipo I e as suposições que em particular cada teste apresenta. Neste aspecto, Gibbons & Chakraborti (1992) citam os testes Qui-quadrado, Kolmogorov-Smirnov (K-S) e Lilliefors

apresentando suas principais diferenças. Entre estas diferenças pode-se citar que a aplicação do teste de aderência qui-quadrado é feita de forma que as amostras são agrupadas em classes diferenciando-se do teste K-S, que considera cada observação da amostra. Uma outra vantagem do teste K-S é que sua aplicação é irrestrita ao tamanho da amostra, ao passo que o teste de aderência qui-quadrado restringe-se a grandes amostras, pois sem esta condição poderá ocasionar valores pequenos para as freqüências esperadas, isto é, valores inferiores a 5 observados em cada classe, o que põe em risco a validade de sua aplicação, sendo que uma alternativa para esta situação é dada pelo teste exato de Fisher. Para aplicação dos testes K-S e aderência qui-quadrado é necessário conhecer média e variância populacional. Isto faz com que, em muitas situações práticas esses testes sejam impraticáveis. Comparando os testes de aderência de qui-quadrado e o K-S com o teste de Lilliefors, é facilmente verificada uma grande flexibilidade nesse último, pelo fato de que a hipótese nula  $H_0$  refere-se a que a população é pertencente a família de distribuições normais, sem contudo ter a necessidade de especificar a média ou variância da distribuição.

Uma alternativa ainda existente para verificação da normalidade é dada mediante uso dos coeficientes de simetria e curtose. Este procedimento, porém, é limitado a grandes amostras, caracterizando um teste assintótico, além de não garantir resultados referentes à normalidade, pois não são realizadas comparações após o quarto momento. Uma distribuição só pode ser considerada igual à outra se todos os seus momentos forem iguais.

Considerando o poder e o controle da taxa de erro tipo I o teste de Shapiro-Wilk é o melhor. Este fato é comprovado por alguns estudos empíricos mostrando que este teste é poderoso em várias situações de diferentes distribuições e diferentes tamanhos de amostras quando comparado com os resultados de testes retromencionados.

Um procedimento gráfico adotado muitas vezes para verificação da normalidade é o denominado "Q-Qplot". Nesses gráficos são plotados os percentis amostrais versus percentis esperados pelo ajuste de uma distribuição normal. Se os pontos se situarem próximos a uma reta de mínimos quadrados a suposição de normalidade deve ser aceita. Esse processo gráfico é bastante poderoso para verificar desvios de normalidade, entretanto não constitui um teste formal, servindo apenas como uma análise exploratória dos dados e na identificação de outliers. Para contornar essa limitação, Johnson & Wichern (1988) apresentaram um teste complementar este processo gráfico, o qual mede o ajuste dos pontos do Q-Qplot à reta de mínimos quadrados. Esse teste complementar é feito por meio de uma medida do coeficiente de correlação denominado  $r_q$ .

Na verificação da normalidade multivariada torna-se mais complexo o uso dos Q-Qplots, uma vez que os percentis da amostra são substituídos pelas distâncias generalizadas de Mahalanobis e os percentis esperados pelo ajuste de uma distribuição normal são substituídos pelos percentis esperados de uma função inversa da distribuição qui-quadrado. Com a devida substituição dos quantis, a mesma idéia do Q-Qplot do caso univariado é adotada para o caso multivariado. É importante ressaltar que é comum na literatura restringir o Q-Qplot apenas ao caso bivariado, entretanto não há argumentos que limitem a extensão para o caso multivariado qualquer. Nenhuma referência na literatura foi encontrada propondo a obtenção dos valores críticos para o teste de correlação Q-Qplot para o caso multivariado.

Em relação aos testes de hipóteses para o diagnóstico de normalidade no caso multivariado, em geral pode ser afirmado que estes são construídos tomando-se por base a extensão dos testes de normalidade univariados. Pode-se mencionar como exemplo as extensões dos testes de aderência qui-quadrado, Kolmogorov-Smirnov, Shapiro-Wilk e os testes baseados em simetria e curtose,

sendo que estes últimos possuem as mesmas limitações ressaltadas no caso univariado.

Em virtude do que foi mencionado, este capítulo tem como objetivo expandir o teste de normalidade baseado no coeficiente de correlação  $r_q$  do gráfico de probabilidade para amostras maiores, bem como, propor sua extensão para o caso multivariado utilizando simulação Monte Carlo apresentando assim os valores críticos para diferentes tamanhos de amostras e número de variáveis.

## **2 - REFERENCIAL TEÓRICO**

### **2.1 - Testes univariados de normalidade.**

Nas seções a seguir são apresentados detalhadamente os principais testes utilizados para avaliar a veracidade das hipóteses de normalidade univariada. A importância desta revisão é justificada pelo fato de que os testes de normalidade multivariada, em geral, são extensões do caso univariado, desta forma, sendo feita a descrição dos principais testes de normalidade encontrados na literatura, torna-se fácil compreender a extensão dos mesmos para o caso multivariado. Esta seção apresenta uma breve discussão sobre o poder apresentado pelos principais testes estatísticos, o qual é utilizado como critério na comparação dos mesmos.

#### **a) Teste de aderência qui-quadrado.**

As hipóteses estatísticas a serem testadas por meio do teste de aderência qui-quadrado basicamente retratam informações sobre a função de densidade de probabilidade que, em geral, é desconhecida. Desta forma, com o objetivo de verificar se os dados são provenientes de uma distribuição normal, as hipóteses

relativas a este teste deverão ser completamente especificadas conhecendo-se os valores referentes aos parâmetros da distribuição considerada sob a hipótese nula. Neste aspecto, pode-se exemplificar a formalização da hipótese  $H_0$  da seguinte forma:

*“ $H_0$ : A função de densidade de probabilidade da população de onde a amostra foi extraída é uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ .”*

Para a realização deste teste, inicialmente pressupõe-se que a amostra seja aleatória e que a escala de medida seja contínua. Sendo assim, o procedimento a ser adotado é: inicialmente, constrói-se uma tabela de contingência (Tabela 2.1.1), em que a definição do número das classes ( $c$ ) poderá ser feita por vários critérios tais como Sturges (1926), Scott (1979). Em seguida as classes são definidas obtendo-se o limite inferior da primeira classe e a amplitude da classe; as demais classes são obtidas somando-se ao limite inferior a amplitude de classe Lemma (1992), Morettin & Bussab (1987). Posteriormente, torna-se necessário obter as freqüências. Assim, a freqüência observada ( $O_i$ ) é obtida por meio da freqüência absoluta (contagem) relativa ao número de elementos da amostra que pertence à classe  $i$ , para  $i = 1, 2, \dots, c$ . Para se obter a freqüência esperada ( $E_i$ ) especificamente no caso da distribuição normal, é necessário que se calcule a probabilidade das observações serem pertencentes à  $i$ -ésima classe ( $p_i$ ) de modo que  $np_i$  venha a representar o número esperado de observações na classe  $i$  quando  $H_0$  é verdadeira, sendo  $n$  o tamanho da amostra.

Tabela 2.1.1 – Forma geral da tabela de contingência utilizada no teste de aderência qui-quadrado, em que  $O_i$  é a frequência observada da classe  $i$ ,  $E_i$  é a frequência esperada da classe  $i$ ,  $c$  é número total de classes, para  $i = 1, 2, \dots, c$

	Classes				Total
	1	2	...	c	
<b>Frequências Observadas</b>	$O_1$	$O_2$	...	$O_c$	n
<b>Frequências Esperadas</b>	$E_1$	$E_2$	...	$E_c$	n

Definindo a estrutura e o procedimento para a obtenção da tabela de contingência obtém-se, então, a estatística do teste, a qual é dada por:

$$T = \sum_{i=1}^c \frac{(O_i - E_i)^2}{E_i}, \quad i=1,2,\dots,c \quad (2.1.1)$$

Segundo Mood et al. (1974) a regra de decisão é dada por meio da distribuição  $\chi^2_{(1-\alpha)}$  com  $v=c-p-1$  graus de liberdade. Especificamente para o caso da distribuição normal tem-se dois parâmetros  $\mu$  e  $\sigma^2$  desta forma, o valor de  $p$  fica definido por  $p=2$  sendo assim, rejeita-se  $H_0$ , caso o valor da estatística  $T$  seja superior ao quantil  $\chi^2_{(1-\alpha)}$  cujo grau de liberdade pode ser dado por  $c-1$ , ou  $c-3$  conforme mencionado por Mood et. al. (1974). Neste caso, conclui-se que a amostra aleatória não provém de uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ .

Uma consideração importante, feita sobre a validação do teste de aderência qui-quadrado, é dada em função das frequências esperadas. Campos (1983) mencionou que o teste de aderência de qui-quadrado deve ser validado pelos seguintes aspectos:

a) se o número de classes é  $c = 2$ , então, a frequência esperada mínima de uma classe  $i$  qualquer deve ser  $E_i \geq 5$ ;

b) se  $c > 2$ , o teste de aderência qui-quadrado não deve ser usado se mais de 20% das frequências esperadas forem  $< 5$  ou se qualquer uma delas for inferior a 1.

Em algumas situações, estas restrições podem ser contornadas reagrupando as classes adjacentes, conseqüentemente, a frequência esperada em cada nova classe será maior que esses limites restritivos.

### **b) Teste de Kolmogorov-Smirnov para normalidade.**

O teste de Kolmogorov Smirnov (K-S) é utilizado para verificar o grau de concordância entre a distribuição de um conjunto de valores amostrais e alguma distribuição teórica específica. O uso deste teste pressupõe que os dados devem consistir de uma amostra aleatória  $X_1, X_2, \dots, X_n$  de tamanho  $n$ , associada a alguma função de distribuição desconhecida, denotada por  $F(X)$ . A função de distribuição completamente especificada é denotada por  $F_0(X)$ . Desta forma, as hipóteses são descritas por  $H_0: F(X) = F_0(X)$  para todo  $X$  de  $-\infty$  a  $+\infty$  versus  $H_1: F(X) \neq F_0(X)$  para pelo menos um valor de  $X$ .

A estatística do teste é dada por  $T = \text{MAX}|F_0(X) - S_n(X)|$ , em que,  $F_0(X)$  e  $S_n(X)$ , respectivamente, referem-se à função de distribuição da variável aleatória  $X$ , sob a hipótese  $H_0$  e a função de distribuição empírica baseada na amostra aleatória. A obtenção de  $S_n(X)$  é dada por:

$$S_n(X) = \frac{k}{n}, \quad (2.1.2)$$

em que  $k$  é o número de observações menores ou iguais a  $X$ .

A regra de decisão para rejeitar  $H_0$  no valor nominal  $\alpha$  é dada pelo valor da estatística  $T$ , caso este valor for maior que o quantil  $W_{1-\alpha}$  obtido da tabela

apropriada para a distribuição da estatística do teste de Kolmogorov-Smirnov. No caso de a função de distribuição  $F_0(X)$  ser contínua, os quantis desta tabela serão exatos.

A aplicação do teste de Kolmogorov-Smirnov para verificar a normalidade parte do pressuposto de que  $F(X)$  caracteriza a distribuição normal, de forma que a distribuição acumulada é dada por:

$$F(X) = \int_{-\infty}^x f(x) dx = P(x \leq X). \quad (2.1.3)$$

Em que  $f(x)$  é a função de densidade da distribuição normal dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}. \quad (2.1.4)$$

Admitindo  $F_0(X)$  a verdadeira função de distribuição, as hipóteses estatísticas são reescritas da seguinte forma:  $H_0: F(X)=F_0(X)$  versus  $F(X)\neq F_0(X)$ .

Uma vantagem do teste Kolmogorov-Smirnov sobre o teste de aderência qui-quadrado é que o primeiro pode ser aplicado sem restrição, para pequenas amostra. Uma outra característica é que o teste K-S trata os dados individualmente, não perdendo informações devido ao agrupamento de classes, como ocorre no teste de aderência qui-quadrado. Como desvantagem, nota-se que este teste não permite estimar os parâmetros da distribuição teórica, portanto, é necessário conhecer os parâmetros da distribuição sob  $H_0$  para que este teste seja realizado.

### c) Teste de Lilliefors para normalidade.

O teste Lilliefors não requer o conhecimento da média ou variância da distribuição normal a ser testada. Para o uso do teste os dados devem consistir de

uma amostra aleatória, associada a alguma função de distribuição desconhecida, denotada por  $F(X)$ . A hipótese a ser testada  $H_0$  é descrita por:

*“ $H_0$ : A amostra aleatória tem uma distribuição normal com média e variância não especificada.  $H_1$ : A função de distribuição da amostra não é normal”.*

A estatística do teste é calculada sobre os valores dos  $Z_i$ ,  $i = 1, 2, \dots, n$ , definidos como valores amostrais "normalizados", cuja expressão é dada por:

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (2.1.5)$$

Para a determinação da estatística do teste é necessário traçar um gráfico da função de distribuição normal denotada por  $F_0(Z)$  juntamente com o gráfico da função de distribuição empírica da amostra normalizada, denotada por  $S(Z)$ . Desta forma, é determinada a distância vertical máxima entre os dois gráficos. Esta distância é a estatística do teste para normalidade, sendo formalmente denotada por  $T = \text{MAX } |F_0(Z) - S(Z)|$ .

Conforme Campos (1983), a não rejeição da hipótese nula ( $H_0$ ) não implica que a variável aleatória tenha distribuição normal, mas indica simplesmente que a diferença entre a função de distribuição normal e a verdadeira função de distribuição é suficientemente não significativa para ser não detectada.

O critério para rejeição da Hipótese, fixado o valor nominal  $\alpha$ , é dado por: rejeita-se  $H_0$  se a estatística  $T$  apresentar valor superior ao quantil  $(1-\alpha)$  obtidos da tabela apropriada para o teste de normalidade de Lilliefors.

#### **d) Teste de Shapiro-Wilk para normalidade.**

O teste de normalidade proposto por Shapiro & Wilk (1965) basicamente consiste em obter uma estatística referente à razão entre o quadrado de uma combinação linear apropriada dos valores esperados das estatísticas de ordem da

distribuição normal padrão e o quadrado dos desvios em relação à média. A estatística deste teste, representada por  $W$ , apresenta-se flexível à adaptação para hipóteses compostas, pois seus coeficientes lineares ( $a_i$ ) são facilmente computáveis.

A suposição exigida para aplicação deste teste é de que a amostra  $X_1, X_2, X_3, \dots, X_n$  seja aleatória, isto é, independente e identicamente distribuída e, ainda, associada a alguma função de distribuição desconhecida. Com esta suposição sendo satisfeita, a obtenção da estatística  $W$  do teste de Shapiro-Wilk é dada conforme descrição que se segue.

Inicialmente, associa-se um vetor de variáveis aleatórias  $\underline{y}$ , de modo que cada elemento será correspondente às amostras ordenadas do menor ao maior valor, ou seja,  $X^{(1)} \leq X^{(2)} \dots \leq X^{(n)}$ , em que  $X^{(i)}$  denota a  $i$ -ésima estatística de ordem. Associando essa amostra ordenada ao vetor  $\underline{y}$ ; sua representação é dada por  $y_1 \leq y_2 \leq \dots \leq y_n$ . Esta associação é explicada facilmente pela igualdade  $y_i = \mu + \sigma X_i$ , em que cada  $X_i$  ( $i=1, 2, \dots, n$ ) tem distribuição normal padrão com média zero e variância 1.

Em seguida, calculam-se as quantidades:

$$D^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (X_{(i)} - \bar{X})^2 \quad (2.1.6)$$

$$b = \sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i) \quad (2.1.7)$$

Definidas estas quantidades, é necessário obter os valores dos coeficientes  $\{a_{n-i+1}\}$ . Estes valores foram tabulados pelos próprios autores em função do tamanho amostral. Posteriormente define-se o valor de  $k$ , o qual será utilizado na equação 2.1.7 de modo que para  $n$  par usa-se a expressão (2.1.8) e para  $n$  ímpar a expressão (2.1.9), respectivamente.

$$k = \frac{n}{2} \quad (2.1.8)$$

$$k = \frac{n-1}{2} \quad (2.1.9)$$

Definido o valor de  $k$ , a equação (2.1.7) pode ser desenvolvida da seguinte forma:

$$b = a_n(y_n - y_1) + \dots + a_{k+2}(y_{k+2} - y_k) \quad (2.1.10)$$

Sendo assim, a estatística  $W$  do teste é definida por:

$$W = \frac{b^2}{D^2} = \frac{\left( \sum_{i=1}^n a_i y_i \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.1.11)$$

Rejeita-se  $H_0$  caso o valor da estatística  $W$  apresentar valor inferior ao quantil apropriado a este teste. Dada a amostra, calculado o valor de  $W$ , o próximo passo é determinar a distribuição nula de  $W$  ou os valores críticos para determinados valores nominais de significância com a finalidade de se realizar o teste da hipótese nula de normalidade dos dados. Shapiro & Wilk (1965) realizaram tabulações de valores críticos de  $W$  utilizando simulação Monte Carlo, alegando dificuldades de lidar analiticamente com a distribuição conjunta de  $W$  devido à presença de correlações entre valores ordenados. As aproximações de Royston (1982, 1983 e 1995) são usadas para o cálculo de  $p$ -valores, relacionados às significâncias observadas de  $W$ . É importante ressaltar algumas propriedades relacionadas pelos próprios autores sobre a estatística  $W$ , sobre a qual afirmam que:  $W$  mantém-se invariante quanto à escala; no caso de as amostras serem provenientes de uma distribuição normal, fica evidente que a estatística  $W$  depende somente do tamanho da amostra, apresentando-se independente da média e variância; o máximo da estatística  $W$  é 1 e o mínimo é dado por  $na_1^2/(n-1)$ .

Como desvantagem, os autores apontam a aplicação desta estatística em grandes amostras ( $n > 50$ ), sendo que, neste caso, torna-se complexo obter ou aproximar os valores referentes aos coeficientes  $a_i$  presentes no numerador. Royston (1982) mostra que a aproximação dos coeficientes dada por Shapiro & Wilk (1965) é bastante insatisfatória para  $n > 50$ . Assim, para  $n > 50$  Royston (1982) sugere uma aproximação por meio de suavizações polinomiais, ou seja, aproximações numéricas do coeficiente  $a_1$ , que requer o uso de coeficientes específicos de polinômios para realizar a transformação.

Com o p-valor computado a decisão de rejeição ou não de  $H_0$  pode ser tomada, comparando o valor obtido com o valor nominal da significância  $\alpha$  adotado. O teste de normalidade conhecido como teste de Shapiro Wilk é considerado um excelente teste, com poder superior ao de seus concorrentes na maioria das situações e controle das taxas de erro tipo I sempre em consonância com o valor nominal adotado, mesmo para pequenos valores de  $n$ .

#### **e) Teste normalidade baseado nos coeficientes de simetria e curtose.**

Para descrever o teste de normalidade baseado nos coeficientes de simetria e curtose, Santos (2001) apresentou o uso dos momentos não centrados para a média estimados a partir dos dados amostrais, fazendo  $1/n$  como densidade para cada ponto amostral. Desta forma, pode-se definir o r-ésimo momento amostral não centrado para média por:

$$\tilde{m}_r = \frac{1}{n} \sum_{j=1}^n x_j^r \quad (2.1.12)$$

Da mesma forma, em função dos momentos não centrados são definidas as seguintes estatísticas:

$$\text{Média: } \tilde{\mu}_1 = 0 \quad (2.1.13)$$

$$\text{Variância: } \tilde{\mu}_2 = \tilde{m}_2 - \tilde{m}_1^2 \quad (2.1.14)$$

$$\text{Simetria: } \tilde{\mu}_3 = \tilde{m}_3 - 3\tilde{m}_1\tilde{m}_2 + 2\tilde{m}_1^3 \quad (2.1.15)$$

$$\text{Curtose: } \tilde{\mu}_4 = \tilde{m}_4 - 4\tilde{m}_1\tilde{m}_3 + 6\tilde{m}_1^2\tilde{m}_2 - 3\tilde{m}_1^4 \quad (2.1.16)$$

Com base nestas estatísticas, os estimadores dos coeficientes de simetria e curtose são definidos, respectivamente, por:

$$\sqrt{b_1} = \frac{\tilde{\mu}_3}{\tilde{\mu}_2 \sqrt{\tilde{\mu}_2}} \quad (2.1.17)$$

$$b_2 = \frac{\tilde{\mu}_4}{\tilde{\mu}_2^2} \quad (2.1.18)$$

A avaliação do grau de simetria pode ser dada por um teste baseado no coeficiente de simetria, em que os níveis críticos para a estatística (2.1.17) podem ser encontrados em Pearson e Hartley (1966) para  $n > 24$ . No caso de  $n$  variando de 5 a 35, os níveis críticos são encontrados em D'agostinho e Tietjen (1973). A simetria será à esquerda se a estatística apresentar valor negativo; caso o valor da estatística seja positivo, a simetria será à direita. Em grandes amostras, os valores críticos podem ser obtidos com boa aproximação, usando como desvio da normal padrão a estatística dada por:

$$Z_1 = \sqrt{b_1} \sqrt{\frac{(n+1)(n+3)}{6(n-2)}} \quad (2.1.19)$$

Valores críticos para o coeficiente de curtose (2.1.18) podem ser encontrados em Pearson e Hartley (1966) para  $n > 49$  e em D'Agostinho e Tietjen (1973) para  $n$  variando de 7 a 50. Em grandes amostras, os valores críticos para o teste de achatamento da curva podem ser aproximados usando como desvio normal a seguinte estatística:

$$Z_2 = \left( b_2 - 3 + \frac{6}{n+1} \right) \sqrt{\frac{(n+1)^2(n+3)(n+5)}{24n(n-2)(n-3)}} \quad (2.1.20)$$

Com base no estimador beta de curtose, valores de  $b_2$  maiores que 3 indicam que a distribuição é mais afilada, com caldas mais baixas do que a normal; valores menores que 3 indicam uma distribuição achatada no centro e com caudas mais altas do que a distribuição normal.

#### **f) Poder dos testes univariados.**

Comparando os testes retromencionados por meio do poder Gibbons & Subhabrata (1992), apresentaram os resultados obtidos por Kac et al. (1955) e mostraram que o teste Kolmogorov-Smirnov (K-S), quando utilizado para testar a normalidade e sob distribuições não normais, é assintoticamente mais poderoso do que o teste de aderência qui-quadrado. Uma outra vantagem do teste K-S apresentada pelos autores está relacionado ao caso das amostras pequenas, para as quais o teste K-S é melhor do que o teste de aderência qui-quadrado.

Shapiro & Wilk (1965) compararam o seu teste (W) com outros testes em relação ao poder, com base nas seguintes distribuições:  $\chi^2_{(1)}$ ;  $\chi^2_{(2)}$ ;  $\chi^2_{(4)}$ ;  $\chi^2_{(10)}$ ;  $\chi^2$  não central, Log-normal; Cauchy; Uniforme; Logística; Beta (2; 1); Laplace; Poisson (1) e Binomial (4; 0,5). Diante destas distribuições, o teste de aderência qui-quadrado apresentou poder inferior em relação ao teste de Shapiro Wilk. Em relação aos testes de simetria e curtose, os autores mostraram que apenas no caso da distribuição logística o teste de simetria apresentou poder superior ao teste de Shapiro Wilk a mesma ocorrência se deu para o teste de curtose, exceto para as distribuições Cauchy; Uniforme; Logística e Laplace. Devido a estes estudos empíricos, atualmente o teste de Shapiro-Wilk é considerado o mais adequado por apresentar alto poder e controle da taxa de erro tipo I.

## 2.2-Teste normalidade multivariada.

Segundo Mecklin & Mundfrom (2000), os testes de normalidade multivariado são classificados por três aspectos: (a) procedimentos gráficos e coeficientes de correlação; (b) testes de validade de ajuste; e (c) e testes baseados em medidas de simetria e curtose. Desta forma, nas seções a seguir são apresentados breves comentários, comparando vários testes, sendo estes, uma extensão do caso univariado.

Em função de se ter verificado no presente trabalho que especificamente os quantis do teste Q-Qplot para o caso multivariado têm relação com os quantis de uma distribuição beta, são apresentadas, de forma sucinta as características e algumas propriedades dessa distribuição com dois parâmetros.

### a) Procedimentos gráficos e coeficiente de correlação.

Seguindo uma extensão do caso univariado vários pesquisadores apresentaram propostas com base nos gráficos e correlações para verificar normalidade multivariada. Healy (1968) sugeriu um método utilizando a distância generalizada de Mahalanobis, definida em (2.1.21) para analisar a normalidade multivariada.

$$r = (y_i - \bar{y})' S_{p \times p}^{-1} (y_i - \bar{y}) \quad (2.1.21)$$

A distribuição de  $r$  é aproximadamente qui-quadrado com  $p$  graus de liberdade. Plotando a distância generalizada de Mahalanobis versus o quantil esperado de uma  $\chi_p^2$  pode-se obter uma aproximação linear de mínimos quadrados entre as duas variáveis, a qual sugere a presença de normalidade multivariada.

Singh (1993) estudou um procedimento que considera o coeficiente de correlação do Q-Qplot obtido pelos valores observados ordenados da distância

de Mahalanobis versus os quantis esperados da distribuição beta obtidos por simulação. Assim foram obtidas as tabelas e os valores críticos para a estatística do teste. Posteriormente foi observado, por meio de uma simples inspeção, que os outliers multivariados influenciavam grandemente o valor calculado do coeficiente de correlação da distribuição beta. Sendo assim, Singh (1993) recomendou que os outliers deveriam ser removidos utilizando como medida a distância quadrada de Mahalanobis.

Uma outra proposta para a construção de um Q-Qplot foi apresentada por Maronna (1976), pela qual eram obtidos estimadores robustos para o vetor de médias e para a matriz de variância-covariância. Estes estimadores foram denominados como estimadores (M.A). A utilização destes estimadores foi dada no cálculo da distância quadrada de Mahalanobis robusta usada para estimar a estatística do teste de correlação  $r_q$ . Singh (1993) concluiu que a probabilidade calculada por meio da distância generalizada robusta identificou melhor os outliers e que a não remoção destes outliers poderia inflacionar a matriz de variância-covariância, levando a violar a estrutura de correlação. Um outro resultado concluído por Singh (1993) em relação aos estimadores M.A. foi que, a utilização destes estimadores no Q-Qplot é o melhor modo para identificar observações problemáticas.

#### **b) Testes para validade de ajuste.**

A maioria dos testes para validade de ajuste tem como base sobre funções da distribuição empírica. Estes testes, em geral, são derivados do caso univariado. Seguindo essa derivação, Moore & Stubblebine (1981) realizaram a extensão do teste de aderência qui-quadrado para o caso multivariado. Considerando tanto a questão da homocedasticidade quanto a questão da normalidade, Hawkins (1981) propôs um teste baseado na estatística do tipo de Anderson-Darling, a qual, basicamente consistia em converter as distâncias de

Mahalanobis em uma estatística, cuja distribuição seria Uniforme desde que os dados multivariados fossem normais. Em relação à estatística do tipo Kolmogorov-Smirnov, Ward (1988) realizou a extensão deste teste para o caso multivariado considerando as estatísticas de ordem das observações multivariadas.

Entre os testes univariados que foram estudados no caso multivariado, o mais crítico talvez tenha sido a extensão do teste de Shapiro-Wilk. O primeiro autor a apresentar sua extensão para o caso multivariado foi Royston (1983), o qual propôs calcular uma estatística  $W_i$  para cada variável, de modo que, cada  $W_i$  fosse normalizado mediante uma transformação. Srivastava & Hui (1987) criticaram a versão multivariada do teste Shapiro-Wilk proposto por Royston afirmando que a correlação entre os dados originais e transformados não estava justificada, e apresentaram uma extensão alternativa para a estatística do teste Shapiro-Wilk que envolvia a técnica de componentes principais. Seguindo a mesma crítica feita pelos autores Srivastava & Hui (1987), Romeu e Ozturk (1993) notaram problemas relacionados a esta correlação. Mudholkar et al. (1995) apresentaram uma outra versão multivariada para o teste Shapiro-Wilk, o qual consistia em decompor o conjunto de dados em  $p$  variáveis independentes, os  $p$ -valores relacionados a cada variável independente seriam combinados de modo a formalizar a estatística do teste.

### **c) Testes baseados em medidas de simetria e curtose.**

Outra categoria para os testes de normalidade multivariado relaciona-se às medidas de simetria e curtose. Mardia (1970) desenvolveu a extensão multivariada do teste univariado de simetria e curtose. Os coeficientes de simetria e curtose foram, respectivamente, definidos por:

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left[ (y_{ij} - \bar{y})' S_{p \times p}^{-1} (y_{ij} - \bar{y}) \right]^3 \quad (2.1.22)$$

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n \left[ (y_i - \bar{y})' S_{p \times p}^{-1} (y_i - \bar{y}) \right]^2 \quad (2.1.23)$$

A estatística do teste de simetria e curtose é dada por:

$$A = \frac{nb_{1,p}}{6} \quad (2.1.24)$$

$$B = \frac{\{b_{2,p} - p(p+2)\}}{\sqrt{\frac{8p(p+2)}{n}}} \quad (2.1.25)$$

A estatística do teste de simetria, tem uma distribuição assintótica qui-quadrado com  $p(p+1)(p+2)/6$  graus de liberdade, enquanto a estatística do teste de curtose tem distribuição assintótica normal padronizada.

Vários autores intensificaram os estudos em relação aos testes de normalidade multivariada, entre eles Malkovich & Afifi (1973), generalizaram as medidas de simetria e curtose univariada usando o princípio de união-interseção. Srivastava (1984) desenvolveu um teste de simetria e curtose utilizando técnicas de componentes principais para avaliar a normalidade multivariada. Esse teste é muito similar ao teste proposto por Mardia (1970). Bera & John (1983) mostraram que o teste proposto por Mardia (1975), relativo à simetria e curtose multivariada, não são medidas que retratam puramente a simetria e curtose multivariada. Este problema foi dado em função das correlações entre a simetria e curtose. Horswell & Looney (1992) questionaram o termo “diagnóstico” de simetria e curtose. Esta crítica foi feita em função de que os autores verificaram que o teste de simetria é dependente da curtose. DeCarlo (1997) verificou que, de modo geral, os testes envolvendo o vetor de médias são mais sensíveis ao efeito de simetria, enquanto os testes envolvendo

matrizes de covariância são mais sensíveis aos efeitos de curtose. Assim confirmaram-se os estudos feitos por Mardia et.al. (1979).

#### **d) Poder dos testes de normalidade multivariada.**

Em geral, a comparação do poder de testes multivariados via simulação é realizada para avaliar um novo teste com várias alternativas existentes. Infelizmente o processo de simulação acaba sendo uma parte minoritária do trabalho. Muitas vezes, nas simulações, consideram-se apenas as situações de normalidade bivariada. Uma outra dificuldade encontrada é a impossibilidade de testar uma ampla gama de distribuições não normais, sendo assim, os estudos de poder limitam-se a uma categoria particular de testes ou consideram apenas os mais populares e promissores.

Entre os testes estudados, Ward (1988) comparou o poder dos testes de simetria e curtose de Mardia (1970), a extensão do teste de Shapiro-Wilk proposto por Malkovich & Afifi (1973), a extensão do teste de Anderson-Darling proposto por Hawkins (1981) e a extensão do teste Kolmogorov-Smirnov feita por Ward (1988), obtendo os seguintes resultados. O teste de simetria e curtose de Mardia, o teste Hawkins e a extensão do teste de Anderson-Darling são os mais poderosos. Nenhum dos testes estudados apresentou boa performance quando considerou a distribuição multivariada t-student, pois esta apresenta pequenos desvios de um modelo normal multivariado. Em relação ao número de variáveis, verificou-se que o poder do teste de Malkovich & Afifi (1973) diminuiu de acordo com o aumento do número de variáveis, ao contrário dos resultados anteriormente observados por Mardia (1970). Em relação à estrutura de correlação, foi observado que o poder destes testes supostamente foi influenciado pela estrutura de correlação, como apontado por Ward (1988), supostamente por meio do determinante da matriz de covariância.

**e) - Distribuição Beta.**

Em função dos quantis do teste Q-Qplot generalizado para o caso multivariado seguirem uma distribuição beta, é de grande importância apresentar a mesma seguida de algumas propriedades. Desta forma, a variável aleatória X, com parâmetros a e b, cuja notação é dada por  $X \sim \text{beta}(a,b)$  tem função densidade dada por :

$$f_x(x) = \frac{x^{a-1}(1+x)^{-a-b}}{\beta[a,b]} I_{(0,1)}(x) \quad (2.1.26)$$

Em que,  $a > 0$  e  $b > 0$ . Esta condição nos parâmetros é importante para definir a distribuição beta. Para o caso de  $a=b=1$  a distribuição beta se resume na distribuição uniforme.

A função  $\beta(a,b)$  é definida por :

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{(a-1)!(b-1)!}{(a+b-1)!}, \quad (2.1.27)$$

Para  $a$  e  $b > 1$

Em que esta igualdade na integral é simplificada pela propriedade da função gama, de modo que

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx = (a-1)\Gamma(a-1) \quad (2.1.28)$$

Assim,  $\Gamma(a) = (a-1)\Gamma(a-1)$  e para alguns valores especiais tem-se:

$$\Gamma(0) = 0; \quad \Gamma(1) = 1 \quad e \quad \Gamma(1/2) = \sqrt{\pi}.$$

A média e a variância, respectivamente, são dadas por:

$$\mu = \frac{a}{b-1}; \quad b > 1 \quad (2.1.29)$$

$$\sigma^2 = \frac{a(a+b-1)}{(b-1)^2(b-2)}; \quad b > 2 \quad (2.1.30)$$

Entre as propriedades da função beta, as mais comuns e muito utilizadas para simplificar resoluções de integrais são:  $\beta(1,1)=1$ ;  $\beta(1/2,1/2) = \pi$ . A geração de números aleatórios para a distribuição beta, basicamente consiste em inverter a função de distribuição.

### 3 - METODOLOGIA

Para realizar a extensão do teste de normalidade, baseado no coeficiente de correlação do gráfico de probabilidade, para o caso multivariado, foram realizadas simulações Monte Carlo considerando diferentes tamanhos de amostras e diferentes números de variáveis. Inicialmente, foram estendidas as tabelas do teste univariado apresentado por Jonhson & Wichern (1988) para outros tamanhos amostrais não contemplados pelos autores até 5.000. Posteriormente, estendeu-se o teste da correlação Q-Qplot, para o caso multivariado, considerando que o número (p) de variáveis foi definido por  $p = 2(1)10$ . Obedecendo à restrição  $n-p+1$ , esta restrição foi necessária devido às operações matriciais apresentarem singularidade quando a amostra é muito pequena em relação ao número de variáveis. Sendo assim, os tamanhos amostrais considerados foram: i)  $n=2(1)30$ ;  $30(5)50$ ;  $50(25)100$ ;  $100(50)500$ ; para o caso multivariado e ii)  $n=3(1)30$ ;  $30(5)85$ ;  $85(25)500$ ;  $1000$ ;  $1.500$ ;  $2.000$  e  $5.000$  para o caso univariado. Numa segunda etapa, o poder deste teste foi avaliado em simulações Monte Carlo, considerando as distribuições Exponencial, Normal, Uniforme e Gama e comparando-as com os testes de normalidade baseados nos desvios de simetria e curtose.

### 3.1- Obtenção dos Quantis univariados Q-Qplots.

Sem perda de generalidade, foram simuladas amostras de tamanho  $n$  de populações normais com média  $\mu=0$  e variância  $\sigma^2=1$ . Considerando a amostra  $X_1, X_2, \dots, X_n$  uma amostra simulada desta distribuição, foi obtido o coeficiente de correlação Q-Qplot ( $r_q$ ). Os valores  $X_j, j=1, 2, \dots, n$  foram simulados utilizando a função RANNOR do programa SAS, Versão 6.12 (1989).

Os valores das  $X_i$  amostras foram ordenados em ordem crescente, resultando na amostra ordenada  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ . Às probabilidades acumuladas correspondentes a um determinado quantil amostral ordenado  $X_{(j)}$ , foi dado o valor:

$$\left( \frac{j - \frac{1}{2}}{n} \right) \quad (3.1.1)$$

Em seguida foram obtidos os quantis da distribuição normal padrão  $q(j)$ , correspondentes a cada valor  $X_{(j)}, j=1, 2, \dots, n$ . Esses quantis foram obtidos por:

$$\left( \frac{j - \frac{1}{2}}{n} \right) = \int_{-\infty}^{q(j)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \quad (3.1.2)$$

Esse valor foi obtido com o uso da função probit(Argumento) do programa SAS versão 6.12. O argumento utilizado dessa função foi o quantil definido em (3.1.1). O valor retornado é o valor da inversa da função de distribuição normal padrão,  $F(z)$ .

Em seguida, correlação  $r_q$  foi obtida em seguida por:

$$r_q = \frac{\sum_{j=1}^n q_{(j)} X_{(j)} - \frac{\sum_{j=1}^n X_{(j)} \sum_{j=1}^n q_{(j)}}{n}}{\sqrt{\left[ \sum_{j=1}^n q_{(j)}^2 - \frac{\left( \sum_{j=1}^n q_{(j)} \right)^2}{n} \right] \left[ \sum_{j=1}^n X_{(j)}^2 - \frac{\left( \sum_{j=1}^n X_{(j)} \right)^2}{n} \right]}} \quad (3.1.3)$$

Para cada tamanho amostral  $n$  foram repetidos 45.000 vezes os processos descritos anteriormente. Dessa forma, 45.000 valores de  $r_q$  foram obtidos, ordenados e os percentis 1%, 5% e 10% foram gerados e tabulados. Esses percentis correspondem aos valores críticos para o teste de normalidade baseados na correlação Q-Qplot.

### 3.2 - Obtenção dos quantis multivariados Q-Qplots.

Após extensivos testes para populações normais com diferentes médias, estruturas e padrões de covariância, constatou-se, por meio das simulações realizadas, que a escolha desses fatores não influenciavam no teste, então, sem perda de generalidade foram consideradas populações com vetores de médias nulos ( $\underline{\mu} = \underline{0}$ ,  $p \times 1$ ) e matriz de covariância  $\underline{\Sigma} = I$  (Identidade  $p \times p$ ). Nesses casos, foram considerados os mesmos tamanhos de amostras nas  $p$  dimensões mencionadas anteriormente. Para cada valor de  $p$  e de  $n$  foi gerada a amostra  $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ , sendo,  $\underline{X}_j$  um vetor de dimensão  $p \times 1$ ,  $j=1, 2, \dots, n$ . Como as  $p$  variáveis eram não correlacionadas  $\underline{\Sigma} = I$  e com a mesma média  $\underline{\mu} = \underline{0}$ , utilizou-se a função RANNOR para simular cada valor  $X_{ij}$ ,  $i=1, 2, \dots, p$  e  $j=1, 2, \dots, n$ .

Em seguida estimou-se o vetor de médias e a matriz de covariância foi definida por:

$$\bar{X} = \frac{\sum_{j=1}^n X_j}{n} \quad (3.2.1)$$

$$S = \frac{\sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})'}{n-1} \quad (3.2.2)$$

No próximo estágio, calculou-se a distância quadrática, entre cada realização amostral,  $X_j$ , e a média  $\bar{X}$  foi dada por:

$$d_j^2 = (X_j - \bar{X})' S^{-1} (X_j - \bar{X}), \quad j=1,2,\dots,n \quad (3.2.3)$$

De acordo com Johnson & Wichern (1988), a distância quadrática de (3.2.3) tem uma distribuição assintótica de qui-quadrado com  $v=p$  graus de liberdade. Assim, a determinação do coeficiente de correlação Q-Qplot foi feita inicialmente ordenando-se as distâncias quadráticas  $d_{(j)}^2$  e obtendo-se  $d_{(1)}^2, d_{(2)}^2, \dots, d_{(n)}^2$ .

Para cada valor de  $d_{(j)}^2$ , é obtido o quantil da distribuição de qui-quadrado com  $v=p$  graus de liberdade ( $q_{(j)}$ ), considerando as probabilidades acumuladas correspondentes dada por (3.2.4). Então,

$$\frac{j - \frac{1}{2}}{n} = \int_0^{q_{(j)}} \frac{1}{2^{v/2} \Gamma(\frac{v}{2})} y^{v/2-1} e^{-y/2} dy, \quad y > 0 \quad (3.2.4)$$

Esses valores dos quantis da distribuição de  $\chi^2$  com  $v=p$  graus de liberdade foram obtidos com o comando CINV do programa SAS versão 6.12, cujo argumento foi (3.2.4). A correlação Q-Qplot  $r_q$  foi obtida por:

$$r_q = \frac{\sum_{j=1}^n \sqrt{q_{(j)}} d_j^2 - \frac{\sum_{j=1}^n \sqrt{d_j^2} \sum_{j=1}^n \sqrt{q_{(j)}}}{n}}{\sqrt{\left[ \sum_{j=1}^n q_{(j)} - \frac{\left( \sum_{j=1}^n \sqrt{q_{(j)}} \right)^2}{n} \right] \left[ \sum_{j=1}^n d_j^2 - \frac{\left( \sum_{j=1}^n \sqrt{d_j^2} \right)^2}{n} \right]}} \quad (3.2.5)$$

Um total de 25.000 reamostragens foi realizado e um número equivalente de valores  $r_q$  foi obtido para gerar a distribuição de Monte Carlo dessa estatística. Os quantis 1%, 5% e 10% foram obtidos após a ordenação dos valores de  $r_q$ .

### 3.3- Avaliação do poder do teste de normalidade.

Para avaliar o poder do teste foram simuladas 36 situações. Cada uma dessas situações foi resultado de uma configuração de uma combinação fatorial do número de variáveis, tamanho de amostra e distribuição da população. Assim, foram considerados os valores de  $p = 1, 2, 6$  e  $10$ . No caso univariado, os tamanhos das amostras foram iguais a  $n = 3, 20$  e  $100$ . Em relação ao caso multivariado os tamanhos amostrais considerados foram  $n = 15, 25$  e  $100$ . Os tamanhos amostrais fixados para avaliação do poder, seja no caso univariado ou multivariado, caracterizaram-se, respectivamente, por amostras de tamanho pequeno, intermediário e grande. As distribuições multivariadas foram geradas a partir das distribuições univariadas uniforme, gama e exponencial. Para a

geração dessas distribuições, considerou-se a matriz de covariância entre as p variáveis com a seguinte estrutura:

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ \rho & 1 & \rho & \dots & \rho^{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \dots & 1 \end{bmatrix} \sigma^2 \quad (3.3.1)$$

Em que  $\rho$  foi especificado com o valor 0,5

O teorema da probabilidade integral foi aplicado e os valores das distribuições foram gerados. Inicialmente um número aleatório uniforme  $U(0, 1)$  foi gerado com a função RANUNI. Para cada distribuição um vetor  $Z_j$  contendo os valores  $Z_{ij}$  gerado por:  $Z_{ij} = F^{-1}(U)$

Em que  $F^{-1}(\cdot)$  é a inversa da função de distribuição de cada distribuição considerada. Para o caso uniforme a  $F^{-1}$  é a própria função de distribuição. Para a distribuição exponencial com parâmetro  $\alpha=1$ , a  $F^{-1}(\cdot)$  é dada por:

$$F^{-1}(U) = \frac{-\ln(U)}{\alpha} = -\ln(U) \quad (3.3.2)$$

Para a distribuição Gama, foi utilizada a função  $\text{gaminv}(U,a)$ , do programa SAS, versão 6.12, sendo que,  $U$  e  $a$ , respectivamente, correspondem aos parâmetros probabilidade e dispersão utilizados para obter os valores de  $Z_{ij}$ . Para gerar as distribuições multivariadas correspondentes à estrutura de covariância  $\Sigma$ , considerando  $\sigma^2=1$  e  $\mu=0$ , foi realizada a transformação linear singular  $X_j = \Gamma Z_j + \mu = \Gamma Z_j$  em que  $\Gamma$  é o fator de Cholesky da matriz  $\Sigma$  da decomposição  $\Sigma = \Gamma \Gamma'$ .

Gerada uma amostra aleatória, foi aplicado o teste de normalidade proposto considerando os quantis  $r_q$ , obtidos em 3.1 e 3.2 e também os testes de normalidade baseados em simetria e em curtose de Mardia (1970). O processo foi repetido 10.000 vezes e a proporção de vezes que  $H_0$ : distribuição normal p-

variada foi rejeitada foi computada. Essa proporção refere-se ao poder do teste, uma vez que a distribuição real simulada não era normal. Os poderes do teste de normalidade baseados em  $r_q$  e nos desvios de simetria e curtose foram comparados entre si.

## 4 - RESULTADOS E DISCUSSÃO

### 4.1 – Quantis do Q-Qplot.

Os resultados descritos nas tabelas 4.1.1 e 4.1.2, especificamente para o caso univariado, referem-se a uma expansão dos quantis para amostras maiores referente ao teste do coeficiente de correlação  $r_q$  apresentado por Johnson & Wichern (1988). O programa SAS, utilizado para a obtenção destes quantis, é apresentado no Anexo (1). A execução deste programa, para obtenção de novos quantis para amostras maiores que 5.000, consiste em apenas definir um valor para o tamanho da amostra, sendo este definido no programa pela variável  $n$ .

No caso multivariado, os resultados obtidos são apresentados nas tabelas 4.1.3 a 4.1.11, de modo que, cada tabela refere-se ao número de variáveis em que  $p=2, \dots, 10$ . Conforme o aumento do número de variáveis, os quantis obtidos para as amostras, restritos à condição  $n-p+1$ , sofreram uma redução. Esta restrição foi necessária devido às operações matriciais envolvidas no processo, sendo que a matriz de variância e covariância tinha que ser invertida. Se  $n-p+1 \leq 0$ , então essa inversa não será única devido a um problema de estimabilidade.

Em relação ao coeficiente de correlação  $r_q$ , utilizado no programa para obtenção dos quantis no caso multivariado, conforme o Anexo (2), este sofreu uma alteração, de modo que, para cada termo pertinente ao coeficiente de correlação, foi extraída a raiz quadrada. Desta forma, definiu-se uma alternativa

para o cálculo do coeficiente de correlação  $r_q$  para o caso multivariado. Esta alteração na forma original do coeficiente de correlação  $r_q$  foi justificada por extensivas simulações, pois o objetivo é aproximar os quantis em uma reta de mínimos quadrados, e sendo estes quantis provenientes de uma distância quadrática (distância generalizada de Mahalanobis), é conveniente atenuar este efeito não linear extraíndo a raiz quadrada destas distâncias. Um outro fato que levou a fazer esta modificação foi resultante do estudo do poder deste teste, confrontado com o teste de simetria e curtose. Com esta alteração, o teste de correlação  $r_q$  mostrou-se mais poderoso.

A execução deste programa para a obtenção de quantis no caso de amostras maiores que 500 consiste em definir um valor para o número de variáveis e tamanho da amostras, sendo estes representados no programa descrito pelo Anexo (2), pelas variáveis  $p$  e  $n$ . Ressaltadas estas observações, seguem abaixo as tabelas com os respectivos quantis que permitirão aos usuários desse teste de normalidade testar a hipótese nula.

TABELA 4.1.1 – Quantis para o teste coeficiente de correlação  $r_s$  definido para as amostras  $n=3$  até  $n=30$  – caso univariado.

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
<b>3</b>	0,8228	0,8613	0,8907	<b>17</b>	0,9180	0,9448	0,9548
<b>4</b>	0,8255	0,8686	0,8945	<b>18</b>	0,9235	0,9470	0,9570
<b>5</b>	0,8296	0,8808	0,9031	<b>19</b>	0,9253	0,9491	0,9580
<b>6</b>	0,8413	0,8900	0,9116	<b>20</b>	0,9286	0,9511	0,9602
<b>7</b>	0,8521	0,8991	0,9182	<b>21</b>	0,9312	0,9527	0,9618
<b>8</b>	0,8646	0,9063	0,9239	<b>22</b>	0,9342	0,9548	0,9632
<b>9</b>	0,8750	0,9133	0,9299	<b>23</b>	0,9361	0,9561	0,9644
<b>10</b>	0,8813	0,9189	0,9347	<b>24</b>	0,9378	0,9577	0,9653
<b>11</b>	0,8894	0,9248	0,9387	<b>25</b>	0,9407	0,9589	0,9664
<b>12</b>	0,8952	0,9279	0,9419	<b>26</b>	0,9413	0,9600	0,9675
<b>13</b>	0,9022	0,9326	0,9455	<b>27</b>	0,9443	0,9613	0,9685
<b>14</b>	0,9051	0,9359	0,9479	<b>28</b>	0,9457	0,9626	0,9693
<b>15</b>	0,9103	0,9394	0,9506	<b>29</b>	0,9480	0,9638	0,9703
<b>16</b>	0,9166	0,9419	0,9529	<b>30</b>	0,9485	0,9645	0,9710

TABELA 4.1.2 – Quantis para o teste coeficiente de correlação  $r_q$  definido para as amostras  $n=35$  até  $n=5000$  – caso univariado.

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
<b>35</b>	0,9548	0,9688	0,9744	<b>225</b>	0,9916	0,9938	0,9948
<b>40</b>	0,9594	0,9719	0,9770	<b>250</b>	0,9922	0,9944	0,9953
<b>45</b>	0,9638	0,9747	0,9791	<b>275</b>	0,9929	0,9949	0,9957
<b>50</b>	0,9686	0,9768	0,9808	<b>300</b>	0,9935	0,9952	0,9960
<b>55</b>	0,9698	0,9786	0,9823	<b>325</b>	0,9940	0,9956	0,9963
<b>60</b>	0,9723	0,9802	0,9835	<b>350</b>	0,9943	0,9959	0,9965
<b>65</b>	0,9740	0,9815	0,9847	<b>375</b>	0,9947	0,9961	0,9967
<b>70</b>	0,9755	0,9827	0,9856	<b>400</b>	0,9950	0,9964	0,9969
<b>75</b>	0,9767	0,9836	0,9865	<b>425</b>	0,9953	0,9966	0,9971
<b>80</b>	0,9779	0,9846	0,9872	<b>450</b>	0,9956	0,9967	0,9972
<b>85</b>	0,9793	0,9853	0,9878	<b>475</b>	0,9958	0,9969	0,9974
<b>100</b>	0,9822	0,9873	0,9894	<b>500</b>	0,9960	0,9970	0,9975
<b>125</b>	0,9854	0,9895	0,9913	<b>1000</b>	0,9979	0,9984	0,9986
<b>150</b>	0,9876	0,9911	0,9926	<b>1500</b>	0,9985	0,9989	0,9991
<b>175</b>	0,9893	0,9922	0,9935	<b>2000</b>	0,9985	0,9992	0,9993
<b>200</b>	0,9876	0,9911	0,9926	<b>5000</b>	0,9995	0,9996	0,9997

TABELA 4.1.3 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=2$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
4	0,7577	0,7905	0,8175	25	0,9454	0,9617	0,9681
5	0,7774	0,8501	0,8770	26	0,9474	0,9629	0,9695
6	0,8294	0,8808	0,9022	27	0,9477	0,9637	0,9704
7	0,8395	0,8933	0,9140	28	0,9503	0,9646	0,9709
8	0,8577	0,9034	0,9225	29	0,9518	0,9655	0,9717
9	0,8709	0,9116	0,9289	30	0,9528	0,9666	0,9725
10	0,8815	0,9197	0,9351	35	0,9578	0,9705	0,9755
11	0,8928	0,9251	0,9395	40	0,9617	0,9732	0,9780
12	0,8994	0,9311	0,9438	45	0,9654	0,9759	0,9804
13	0,9070	0,9351	0,9477	50	0,9675	0,9779	0,9818
14	0,9127	0,9388	0,9501	75	0,9764	0,9839	0,9868
15	0,9176	0,9428	0,9532	100	0,9817	0,9874	0,9898
16	0,9225	0,9446	0,9547	150	0,9867	0,9912	0,9927
17	0,9262	0,9482	0,9575	200	0,9900	0,9931	0,9943
18	0,9297	0,9503	0,9590	250	0,9918	0,9943	0,9954
19	0,9328	0,9523	0,9607	300	0,9930	0,9952	0,9961
20	0,9345	0,9544	0,9622	350	0,9939	0,9958	0,9965
21	0,9382	0,9556	0,9637	400	0,9947	0,9963	0,9969
22	0,9398	0,9574	0,9648	450	0,9952	0,9967	0,9972
23	0,9416	0,9589	0,9663	500	0,9956	0,9970	0,9975
24	0,9440	0,9599	0,9675				

TABELA 4.1.4 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=3$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
<b>4</b>	-	-	-	<b>25</b>	0,9453	0,9615	0,9680
<b>5</b>	0,7261	0,7590	0,7815	<b>26</b>	0,9465	0,9625	0,9690
<b>6</b>	0,8021	0,8452	0,8669	<b>27</b>	0,9475	0,9639	0,9702
<b>7</b>	0,8260	0,8776	0,8971	<b>28</b>	0,9507	0,9649	0,9711
<b>8</b>	0,8476	0,8933	0,9129	<b>29</b>	0,9521	0,9659	0,9719
<b>9</b>	0,8626	0,9038	0,9214	<b>30</b>	0,9530	0,9665	0,9724
<b>10</b>	0,8732	0,9113	0,9286	<b>35</b>	0,9595	0,9711	0,9759
<b>11</b>	0,8857	0,9206	0,9355	<b>40</b>	0,9638	0,9743	0,9786
<b>12</b>	0,8933	0,9261	0,9402	<b>45</b>	0,9668	0,9764	0,9804
<b>13</b>	0,8999	0,9317	0,9444	<b>50</b>	0,9694	0,9782	0,9819
<b>14</b>	0,9086	0,9369	0,9487	<b>75</b>	0,9786	0,9847	0,9873
<b>15</b>	0,9144	0,9401	0,9513	<b>100</b>	0,9830	0,9880	0,9900
<b>16</b>	0,9180	0,9433	0,9538	<b>150</b>	0,9883	0,9916	0,9930
<b>17</b>	0,9236	0,9465	0,9563	<b>200</b>	0,9908	0,9935	0,9946
<b>18</b>	0,9264	0,9488	0,9580	<b>250</b>	0,9922	0,9945	0,9955
<b>19</b>	0,9303	0,9508	0,9597	<b>300</b>	0,9932	0,9954	0,9962
<b>20</b>	0,9333	0,9533	0,9618	<b>350</b>	0,9942	0,9960	0,9967
<b>21</b>	0,9374	0,9559	0,9638	<b>400</b>	0,9949	0,9964	0,9970
<b>22</b>	0,9392	0,9571	0,9648	<b>450</b>	0,9954	0,9968	0,9973
<b>23</b>	0,9429	0,9590	0,9661	<b>500</b>	0,9958	0,9971	0,9976
<b>24</b>	0,9432	0,9599	0,9670				

TABELA 4.1.5 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=4$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
4	-	-	-	25	0,9438	0,9603	0,9672
5	-	-	-	26	0,9453	0,9615	0,9681
6	0,6992	0,7358	0,7642	27	0,9479	0,9632	0,9698
7	0,7957	0,8361	0,8562	28	0,9491	0,9640	0,9703
8	0,8241	0,8716	0,8924	29	0,9506	0,9648	0,9711
9	0,8468	0,8916	0,9102	30	0,9537	0,9663	0,9722
10	0,8611	0,9028	0,9207	35	0,9583	0,9705	0,9756
11	0,8737	0,9121	0,9287	40	0,9632	0,9737	0,9782
12	0,8846	0,9208	0,9348	45	0,9662	0,9760	0,9801
13	0,8948	0,9270	0,9409	50	0,9691	0,9780	0,9818
14	0,9026	0,9318	0,9440	75	0,9783	0,9848	0,9873
15	0,9087	0,9360	0,9474	100	0,9831	0,9881	0,9900
16	0,9152	0,9400	0,9510	150	0,9881	0,9916	0,9930
17	0,9189	0,9438	0,9535	200	0,9909	0,9935	0,9946
18	0,9231	0,9465	0,9564	250	0,9925	0,9947	0,9955
19	0,9272	0,9489	0,9580	300	0,9937	0,9954	0,9962
20	0,9300	0,9514	0,9602	350	0,9944	0,9960	0,9967
21	0,9327	0,9535	0,9617	400	0,9950	0,9965	0,9970
22	0,9365	0,9557	0,9634	450	0,9955	0,9968	0,9973
23	0,9392	0,9574	0,9649	500	0,9959	0,9971	0,9976
24	0,9414	0,9590	0,9660				

TABELA 4.1.6 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=5$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
<b>4</b>	-	-	-	<b>25</b>	0,9427	0,9586	0,9660
<b>5</b>	-	-	-	<b>26</b>	0,9444	0,9604	0,9674
<b>6</b>	-	-	-	<b>27</b>	0,9468	0,9621	0,9685
<b>7</b>	0,6785	0,7196	0,7497	<b>28</b>	0,9471	0,9633	0,9695
<b>8</b>	0,7830	0,8247	0,8472	<b>29</b>	0,9506	0,9644	0,9705
<b>9</b>	0,8259	0,8667	0,8865	<b>30</b>	0,9508	0,9654	0,9712
<b>10</b>	0,8450	0,8885	0,9069	<b>35</b>	0,9580	0,9702	0,9751
<b>11</b>	0,8618	0,9024	0,9190	<b>40</b>	0,9628	0,9733	0,9779
<b>12</b>	0,8749	0,9118	0,9277	<b>45</b>	0,9657	0,9759	0,9798
<b>13</b>	0,8849	0,9192	0,9334	<b>50</b>	0,9693	0,9779	0,9817
<b>14</b>	0,8970	0,9255	0,9391	<b>75</b>	0,9781	0,9846	0,9870
<b>15</b>	0,9012	0,9311	0,9442	<b>100</b>	0,9836	0,9880	0,9899
<b>16</b>	0,9019	0,9361	0,9476	<b>150</b>	0,9882	0,9916	0,9929
<b>17</b>	0,9140	0,9398	0,9511	<b>200</b>	0,9907	0,9934	0,9945
<b>18</b>	0,9193	0,9430	0,9537	<b>250</b>	0,9925	0,9946	0,9955
<b>19</b>	0,9227	0,9466	0,9563	<b>300</b>	0,9936	0,9954	0,9961
<b>20</b>	0,9266	0,9487	0,9583	<b>350</b>	0,9945	0,9960	0,9967
<b>21</b>	0,9314	0,9516	0,9604	<b>400</b>	0,9951	0,9965	0,9971
<b>22</b>	0,9330	0,9535	0,9620	<b>450</b>	0,9956	0,9968	0,9973
<b>23</b>	0,9368	0,9555	0,9635	<b>500</b>	0,9960	0,9971	0,9976
<b>24</b>	0,9391	0,9572	0,9651				

TABELA 4.1.7 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=6$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
4	-	-	-	25	0,9392	0,9570	0,9649
5	-	-	-	26	0,9415	0,9587	0,9660
6	-	-	-	27	0,9427	0,9598	0,9670
7	-	-	-	28	0,9469	0,9618	0,9686
8	0,6620	0,7098	0,7417	29	0,9480	0,9634	0,9698
9	0,7777	0,8193	0,8425	30	0,9495	0,9641	0,9704
10	0,8233	0,8630	0,8825	35	0,9563	0,9686	0,9742
11	0,8446	0,8857	0,9040	40	0,9617	0,9726	0,9772
12	0,8626	0,9010	0,9175	45	0,9654	0,9752	0,9796
13	0,8757	0,9109	0,9265	50	0,9692	0,9776	0,9814
14	0,8861	0,9189	0,9338	75	0,9788	0,9847	0,9800
15	0,8922	0,9255	0,9391	100	0,9833	0,9880	0,9899
16	0,9026	0,9312	0,9430	150	0,9880	0,9915	0,9929
17	0,9066	0,9353	0,9473	200	0,9906	0,9934	0,9945
18	0,9131	0,9394	0,9508	250	0,9926	0,9946	0,9955
19	0,9175	0,9426	0,9535	300	0,9937	0,9955	0,9962
20	0,9225	0,9457	0,9559	350	0,9945	0,9960	0,9967
21	0,9274	0,9491	0,9582	400	0,9952	0,9965	0,9971
22	0,9304	0,9511	0,9601	450	0,9956	0,9969	0,9974
23	0,9334	0,9535	0,9620	500	0,9960	0,9971	0,9976
24	0,9360	0,9549	0,9632				

TABELA 4.1.8 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=7$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
4	-	-	-	25	0,9357	0,9555	0,9634
5	-	-	-	26	0,9403	0,9578	0,9653
6	-	-	-	27	0,9412	0,9587	0,9663
7	-	-	-	28	0,9435	0,9602	0,9674
8	-	-	-	29	0,9459	0,9618	0,9688
9	0,6509	0,7039	0,7387	30	0,9481	0,9630	0,9694
10	0,7713	0,8127	0,8366	35	0,9558	0,9683	0,9738
11	0,8196	0,8604	0,8798	40	0,9609	0,9721	0,9770
12	0,8450	0,8841	0,9016	45	0,9649	0,9750	0,9791
13	0,8633	0,8997	0,9163	50	0,9690	0,9772	0,9810
14	0,8730	0,9099	0,9254	75	0,9783	0,9843	0,9869
15	0,8864	0,9188	0,9328	100	0,9836	0,9878	0,9898
16	0,8949	0,9253	0,9389	150	0,9883	0,9915	0,9928
17	0,9045	0,9320	0,9436	200	0,9909	0,9934	0,9945
18	0,9071	0,9357	0,9475	250	0,9926	0,9947	0,9955
19	0,9151	0,9395	0,9503	300	0,9938	0,9955	0,9962
20	0,9190	0,9427	0,9532	350	0,9945	0,9960	0,9967
21	0,9218	0,9459	0,9558	400	0,9951	0,9965	0,9970
22	0,9260	0,9488	0,9579	450	0,9957	0,9968	0,9973
23	0,9300	0,9513	0,9604	500	0,9960	0,9971	0,9976
24	0,9340	0,9530	0,9618				

TABELA 4.1.9 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=8$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
4	-	-	-	25	0,9314	0,9531	0,9617
5	-	-	-	26	0,9359	0,9551	0,9632
6	-	-	-	27	0,9390	0,9568	0,9647
7	-	-	-	28	0,9408	0,9584	0,9658
8	-	-	-	29	0,9435	0,9603	0,9675
9	-	-	-	30	0,9458	0,9618	0,9687
10	0,6452	0,7018	0,7385	35	0,9537	0,9673	0,9729
11	0,7630	0,8079	0,8341	40	0,9601	0,9715	0,9764
12	0,8168	0,8569	0,8773	45	0,9640	0,9743	0,9788
13	0,8412	0,8816	0,9007	50	0,9679	0,9771	0,9808
14	0,8606	0,8973	0,9143	75	0,9779	0,9841	0,9867
15	0,8735	0,9083	0,9239	100	0,9830	0,9877	0,9896
16	0,8840	0,9177	0,9320	150	0,9883	0,9914	0,9928
17	0,8932	0,9251	0,9387	200	0,9910	0,9935	0,9945
18	0,9022	0,9299	0,9427	250	0,9924	0,9946	0,9954
19	0,9090	0,9352	0,9466	300	0,9938	0,9954	0,9961
20	0,9136	0,9394	0,9500	350	0,9945	0,9960	0,9966
21	0,9181	0,9424	0,9531	400	0,9951	0,9965	0,9970
22	0,9236	0,9461	0,9556	450	0,9957	0,9968	0,9973
23	0,9264	0,9483	0,9577	500	0,9961	0,9972	0,9976
24	0,9307	0,9511	0,9601				

TABELA 4.1.10 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=9$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
<b>4</b>	-	-	-	<b>25</b>	0,9303	0,9509	0,9600
<b>5</b>	-	-	-	<b>26</b>	0,9333	0,9535	0,9618
<b>6</b>	-	-	-	<b>27</b>	0,9363	0,9556	0,9634
<b>7</b>	-	-	-	<b>28</b>	0,9397	0,9571	0,9649
<b>8</b>	-	-	-	<b>29</b>	0,9418	0,9589	0,9665
<b>9</b>	-	-	-	<b>30</b>	0,9432	0,9603	0,9673
<b>10</b>	-	-	-	<b>35</b>	0,9515	0,9663	0,9723
<b>11</b>	0,6378	0,6960	0,7332	<b>40</b>	0,9586	0,9706	0,9757
<b>12</b>	0,7629	0,8070	0,8327	<b>45</b>	0,9637	0,9739	0,9784
<b>13</b>	0,8128	0,8541	0,8755	<b>50</b>	0,9670	0,9764	0,9805
<b>14</b>	0,8405	0,8798	0,8985	<b>75</b>	0,9769	0,9840	0,9866
<b>15</b>	0,8587	0,8959	0,9128	<b>100</b>	0,9826	0,9875	0,9896
<b>16</b>	0,8735	0,9083	0,9236	<b>150</b>	0,9882	0,9914	0,9928
<b>17</b>	0,8835	0,9175	0,9318	<b>200</b>	0,9909	0,9934	0,9944
<b>18</b>	0,8960	0,9247	0,9377	<b>250</b>	0,9926	0,9946	0,9954
<b>19</b>	0,9012	0,9297	0,9420	<b>300</b>	0,9938	0,9955	0,9962
<b>20</b>	0,9097	0,9349	0,9463	<b>350</b>	0,9946	0,9960	0,9966
<b>21</b>	0,9146	0,9397	0,9508	<b>400</b>	0,9951	0,9964	0,9970
<b>22</b>	0,9193	0,9427	0,9534	<b>450</b>	0,9955	0,9968	0,9973
<b>23</b>	0,9240	0,9457	0,9556	<b>500</b>	0,9960	0,9971	0,9976
<b>24</b>	0,9268	0,9485	0,9579				

TABELA 4.1.11 – Quantis para o teste coeficiente de correlação  $r_q$  para o caso multivariado ( $p=10$ ).

<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>	<b>n</b>	<b>0,01</b>	<b>0,05</b>	<b>0,1</b>
4	-	-	-	25	0,9274	0,9484	0,9580
5	-	-	-	26	0,9310	0,9506	0,9596
6	-	-	-	27	0,9327	0,9528	0,9616
7	-	-	-	28	0,9364	0,9553	0,9632
8	-	-	-	29	0,9377	0,9572	0,9649
9	-	-	-	30	0,9400	0,9581	0,9656
10	-	-	-	35	0,9500	0,9652	0,9715
11	-	-	-	40	0,9578	0,9702	0,9753
12	0,6317	0,6953	0,7329	45	0,9616	0,9730	0,9776
13	0,7539	0,8043	0,8306	50	0,9666	0,9760	0,9800
14	0,8093	0,8531	0,8746	75	0,9770	0,9836	0,9863
15	0,8409	0,8806	0,8907	100	0,9828	0,9876	0,9896
16	0,8586	0,8963	0,9123	150	0,9881	0,9913	0,9927
17	0,8750	0,9087	0,9242	200	0,9910	0,9933	0,9944
18	0,8837	0,9171	0,9314	250	0,9926	0,9946	0,9955
19	0,8925	0,9239	0,9372	300	0,9936	0,9954	0,9961
20	0,9015	0,9296	0,9426	350	0,9944	0,9960	0,9966
21	0,9069	0,9339	0,9460	400	0,9953	0,9965	0,9970
22	0,9124	0,9387	0,9494	450	0,9956	0,9968	0,9973
23	0,9194	0,9425	0,9526	500	0,9960	0,9971	0,9976
24	0,9241	0,9455	0,9555				

## 4.2 – Taxa de erro tipo I

Para avaliar o teste coeficiente de correlação  $r_q$ , inicialmente verificou-se o controle da taxa do erro tipo I, com valores nominais fixados em 1%, 5% e 10%, seja para o caso univariado ou multivariado, condicionados aos tamanhos amostrais mencionados anteriormente em (3.3). Obviamente, a realização deste procedimento foi feita utilizando a distribuição normal. O fato de usar a distribuição normal como referência para o estudo do controle da taxa de erro tipo I é justificado, principalmente pela derivação do teste do coeficiente de correlação  $r_q$  ser baseada nessa distribuição. Também foram aplicados os testes de normalidade baseados nos desvios de simetria e curtose, para fins de comparação das taxas de erro tipo I dessas alternativas com o teste proposto. É oportuno enfatizar que o teste da correlação  $r_q$  possui o tamanho nominal igual ao valor de  $\alpha$ , uma vez que o mesmo foi construído sob a condição de normalidade univariada ou multivariada por meio de simulação Monte Carlo.

No caso univariado (Tabela 4.2.1[p=1]), para amostras pequenas  $n=3$ , os testes de normalidade baseados em desvios de simetria e de curtose apresentaram valores nulos. Este resultado já era esperado devido às limitações assintóticas desses procedimentos. Sendo assim, pode-se apontar uma primeira vantagem do teste coeficiente de correlação  $r_q$ , uma vez que o teste proposto para este tamanho amostral controla, por construção, a taxa de erro tipo I. Esse controle apresenta pequenas distorções dos valores nominais causadas pelo fato de o processo de Monte Carlo ser aleatório. Desta forma, verifica-se que o teste do coeficiente de correlação  $r_q$  é apropriado, mesmo nas situações de pequenos tamanhos amostrais. Para as amostras de tamanho intermediário, ou seja,  $n=20$  os testes de simetria e curtose apresentaram valores referentes ao controle da taxa de erro tipo I menores do que as correspondentes taxas nominais, caracterizando-se conservativos para todos os valores nominais. Para as

amostras de tamanho  $n=100$ , todos os testes em estudo apresentaram valores referentes à taxa de erro tipo I, com valores considerados idênticos aos valores nominais fixados. A diferença entre as probabilidades empíricas referentes ao controle da taxa de erro tipo I com o valor nominal é devida ao erro de Monte Carlo.

TABELA 4.2.1 – Taxa do erro tipo I ( $p=1$ ) para o caso univariado e bivariado ( $p=2$ ).

<b>p=1</b>				<b>p=2</b>			
<b>n=3</b>				<b>n=15</b>			
<b>ns.</b>	<b>sim.</b>	<b>curt.</b>	<b><math>r_q</math></b>	<b>ns.</b>	<b>sim.</b>	<b>curt.</b>	<b><math>r_q</math></b>
0,01	0,0000	0,0000	0,0074	0,01	0,0024	0,0002	0,0095
0,05	0,0000	0,0000	0,0512	0,05	0,0202	0,0013	0,0500
0,10	0,0000	0,0000	0,1000	0,10	0,0425	0,0048	0,0972
<b>n=20</b>				<b>n=25</b>			
0,01	0,0075	0,0063	0,0097	0,01	0,0078	0,0021	0,0099
0,05	0,0314	0,0159	0,0510	0,05	0,0312	0,0092	0,0543
0,10	0,0614	0,0249	0,1019	0,10	0,0604	0,0207	0,1012
<b>n=100</b>				<b>n=100</b>			
0,01	0,0010	0,0143	0,0095	0,01	0,0127	0,0101	0,0103
0,05	0,0441	0,0323	0,0501	0,05	0,0498	0,0316	0,0525
0,10	0,0891	0,0601	0,0991	0,10	0,0922	0,0680	0,1038

Em relação ao caso multivariado, inicialmente foi verificado o controle da taxa de erro tipo I para situação bivariada, isto é, para  $p=2$ . Desta forma, por meio de os resultados apresentados na Tabela 4.2.1 [ $p=2$ ], verificou-se que não houve nenhum desvio das taxas de erro tipo I do teste  $r_q$  em relação aos valores nominais correspondentes em todas as situações de tamanhos amostrais. Entretanto, um importante resultado foi observado para o teste de curtose, o qual apresentou taxas bem inferiores aos valores nominais para as amostras menores,

ou seja, para  $n=15$  e  $n=25$ . Da mesma forma, o teste de assimetria mostrou-se conservador para pequenas amostras. Porém, esse efeito foi de menor magnitude do que o do teste baseado em desvios de mesocurtose. O teste de curtose tendeu a igualar os valores nominais para amostras maiores ( $n=100$ ) e níveis nominais menores ( $\alpha=1\%$ ). Para níveis nominais de 5 ou de 10%, ainda, permaneceu conservador para esse tamanho de amostra. Esses resultados são coerentes com os apresentados por Santos (2001).

Em relação ao número maior de variáveis, o controle da taxa de erro tipo I foi analisado para  $p=6$  e  $p=10$  (Tabela 4.2.2). Para  $n=15$ , com o aumento do número de variáveis de 6 para 10, o teste de simetria apresentou uma leve tendência em reduzir a taxa de erro tipo I para todos os valores nominais e dentro de um mesmo valor nominal de significância. Isso resulta em um maior conservadorismo do teste, o qual cresce à medida que o número de variáveis aumenta. Um maior tamanho de amostra se torna necessário quando o número de variáveis é maior. Em relação ao teste de curtose, o efeito foi o contrário, ou seja, a partir do valor nominal fixado em 5% houve uma “explosão” nas taxas de erro tipo I, resultando em valores bastante superiores aos valores nominais fixados, principalmente para amostras menores e maiores quantidades de variáveis ( $p=10$ ). Em relação ao teste coeficiente de correlação  $r_q$ , o mesmo foi invariante com o aumento do número de variáveis e independente dos tamanhos de amostras utilizados, propiciando praticamente taxas de erro tipo I similares às nominais, salvo erros de Monte Carlo, como já era esperado por construção.

TABELA 4.2.2 – Taxa do erro tipo I para o caso multivariado.

p=6				p=10			
n=15				n=15			
ns.	sim.	curt.	$r_q$	ns.	sim.	curt.	${}^{(6)}r_q$
0,01	0,0000	0,0000	0,0087	0,01	0,0000	0,0000	0,0120
0,05	0,0003	0,0062	0,0504	0,05	0,0000	0,3862	0,0530
0,10	0,0019	0,1155	0,0995	0,10	0,0000	0,8536	0,0949
n=25				n=25			
0,01	0,0025	0,0000	0,0096	0,01	0,0002	0,0018	0,0102
0,05	0,0116	0,0219	0,0489	0,05	0,0007	0,1515	0,0519
0,10	0,0240	0,1026	0,1048	0,10	0,0032	0,3963	0,1018
n=100				n=100			
0,01	0,0116	0,0047	0,0118	0,01	0,0099	0,0064	0,0097
0,05	0,0437	0,0380	0,0511	0,05	0,0374	0,0670	0,0506
0,10	0,0812	0,0963	0,0997	0,10	0,0665	0,1528	0,0965

### 4.3 – Poder do testes

Para mensuração do poder, foi utilizada como referência a comparação dos testes em estudo, aplicados à distribuição uniforme, gama e exponencial. Desta forma, foram obtidos os seguintes resultados referentes às situações univariada e multivariada. Em relação ao caso univariado (Tabela 4.3.1), devido às limitações assintóticas, foi observado que para o tamanho amostral  $n=3$ , os testes de simetria e curtose apresentaram valores nulos para o poder. Sendo assim, nitidamente o teste coeficiente de correlação  $r_q$  mostrou-se vantajoso em relação aos testes de simetria e curtose, pois além de controlar a taxa de erro tipo I, conforme resultados apresentados na tabela 4.2.1, apresentou valores

referentes ao poder superiores aos valores dos testes de simetria e curtose para todos os valores nominais em estudo. No entanto, somente em alguns casos é que os valores de poder foram significativamente superiores às taxas nominais adotadas. Para  $n = 20$ , os testes de simetria e curtose apresentaram valores referentes ao poder inferiores aos do teste do coeficiente de correlação  $r_q$ , esta característica foi verificada para as distribuições uniforme, gama e exponencial. Obviamente, espera-se que desvios de simetria em uma distribuição uniforme, a qual é perfeitamente simétrica, seja inferior a qualquer resultado de poder de seus testes competidores. Essa é uma desvantagem de usar um teste de normalidade baseado em desvio de simetria ou de mesocurtose em distribuições parentes, de que a amostra foi extraída, que não são normais, mas cujos parâmetros de simetria e de curtose se aproximem do valor daqueles da distribuição normal, ou até mesmo algum deles. Os poderes do coeficiente  $r_q$  para amostras de tamanhos relativamente pequenos superaram em alguns casos a taxa de 80 para  $\alpha=5\%$  e se mostra promissor para ser indicado como um teste alternativo aos já existentes e utilizados rotineiramente. Para as amostras maiores, ou seja,  $n=100$  (Tabela 4.3.2) o teste do coeficiente de correlação  $r_q$ , novamente mostrou-se poderoso, entretanto, para as distribuições Gama e Exponencial o teste de simetria apresentou valores de poder iguais ao máximo e idênticos aos valores daqueles do teste do coeficiente de correlação  $r_q$ . Portanto, poder-se-ia se pensar que para grandes amostras o uso do teste do coeficiente de correlação  $r_q$  ou o teste de simetria poderiam ser recomendados indistintamente, em virtude de que ambos controlaram eficientemente a taxa de erro tipo I, conforme resultados apresentados na tabela 4.2.1. É conveniente salientar, entretanto, que para distribuições parentes como a uniforme, o poder do teste de assimetria é praticamente nulo, sendo que os pequenos poderes detectados na Tabela 4.3.1 são meramente resultantes de erros de Monte Carlo. O teste  $r_q$ , considerando todas as situações de distribuições, valores nominais da

significância e de tamanhos de amostras foi o que apresentou maior poder quando comparado com os seus dois competidores.

Na situação multivariada, obedecendo à restrição  $n-p+1$ , devido ao problema de estimabilidade dos estimadores necessários à aplicação dos testes, estudaram-se os tamanhos amostrais  $n=15$ ,  $25$  e  $100$ . Inicialmente, para o caso bivariado, verificou-se que o teste coeficiente de correlação  $r_q$  apresentou valores de poderes superiores aos dos testes de simetria e curtose para as amostras  $n=15$  e  $n=25$  (Tabela 4.3.1), condicionados à distribuição uniforme. Para  $n=100$  (Tabela 4.3.2), este teste apresentou poder similar ao do teste de curtose e ambos apresentaram poder superior àquele do teste de simetria. No caso das distribuições gama e exponencial, o teste do coeficiente de correlação  $r_q$  apresentou poderes inferiores ao teste de simetria para tamanhos amostrais menores, independentemente do valor nominal fixado. No entanto, apresentou poder superior àquele do teste de curtose. Para amostras grandes ( $n=100$ ), nessas distribuições, todos os três testes apresentaram poderes praticamente similares.

Com o aumento do número de variáveis de  $p=2$  para  $p=6$  e  $p=10$ , foi verificado que o teste do coeficiente de correlação  $r_q$  apresentou poder decrescente com o aumento do número de variáveis, independentemente do tamanho amostral. O aumento do número de variáveis refletiu uma redução do poder do teste proposto. Embora isso tenha ocorrido, o teste da correlação  $r_q$  pode ser mais eficiente do que o teste de simetria em relação ao poder em distribuições simétricas, com a uniforme para  $p=6$  ou  $p=10$  ou de eficiência equivalente a ambos os concorrentes, como para distribuições gama e exponencial com  $n=100$  e  $p=6$  ou  $p=10$  (Tabela 4.3.4). Desta forma, concluiu-se que este teste apresenta baixo poder quando se aplica o teste de normalidade para um número maior de variáveis, e pequenos tamanhos amostrais. Nestas condições recomendam-se os testes de simetria e curtose. Devido a esta deficiência, o teste de curtose mostrou-se mais poderoso na distribuição

uniforme para todos os tamanhos amostrais, principalmente para  $p=10$ , pois, com o aumento do número de variáveis o poder deste teste aumentou, ao passo que, no teste coeficiente de correlação  $r_q$ , este efeito foi inverso, ocasionando uma redução no seu poder. Para a distribuição gama, notou-se que o maior poder foi alternado entre o teste de simetria e curtose, de tal forma que, para  $n=15$ , o teste de simetria foi mais poderoso para  $p=6$  (Tabela 4.3.3), aumentando o número de variáveis, para  $p=10$  o teste de curtose foi o que apresentou maior poder. Para as amostras de tamanhos  $n=25$  (Tabela 4.3.3), independentemente do número de variáveis, o teste de simetria foi mais poderoso. Para  $n=100$  (Tabela 4.3.4), o poder do teste de simetria foi idêntico ao de curtose. Para a distribuição exponencial e para  $n=15$  (Tabela 4.3.3) e valor nominal fixado em 1%, o teste de simetria foi mais apropriado, embora com valor de poder extremamente baixo. Já para  $p=10$  (Tabela 4.3.3) verificou-se que o teste coeficiente de correlação  $r_q$  foi o mais apropriado, sujeitos ao mesmo tamanho amostral e valor nominal, embora o valor referente a este poder seja inexpressivo e pouco superior ao valor nominal da significância. Com o aumento do valor nominal para 5%, o teste de simetria continuou mais poderoso para  $p=6$ , porém para  $p=10$  o teste do coeficiente de correlação  $r_q$  perdeu poder para o teste de curtose, ainda para a distribuição exponencial. Esta característica foi também verificada para o valor nominal fixado em 10%.

Para  $n=25$ , o teste de simetria foi o mais indicado, independentemente do valor nominal, tanto para  $p=6$  como  $p=10$  (Tabela 4.3.3), para a distribuição exponencial. Este fato foi também detectado para  $n=100$  (Tabela 4.3.4), sendo que neste último caso o teste de simetria atingiu seu valor máximo de poder, embora os demais testes tenham apresentado poderes relativamente similares.

TABELA 4.3.1 – Avaliação do poder do teste coeficiente de correlação  $r_q$ , simetria e curtose para os casos univariado ( $p=1$ ) e bivariado ( $p=2$ ) com amostras pequenas.

p=1					p=2				
n=3					n=15				
dist.	ns.	sim.	curt.	$r_q$	dist.	ns.	sim.	curt.	$r_q$
U.	0,01	0,0000	0,0000	0,0116	U.	0,01	0,0003	0,0000	0,0546
	0,05	0,0000	0,0000	0,0628		0,05	0,0015	0,0000	0,1677
	0,10	0,0000	0,0000	0,1208		0,10	0,0050	0,0090	0,2598
Gama	0,01	0,0000	0,0000	0,0228	Gama	0,01	0,4628	0,1797	0,2631
	0,05	0,0000	0,0000	0,1070		0,05	0,6722	0,2830	0,4824
	0,10	0,0000	0,0000	0,1813		0,10	0,7730	0,3456	0,5836
Exp.	0,01	0,0000	0,0000	0,0125	Exp.	0,01	0,2675	0,0801	0,1244
	0,05	0,0000	0,0000	0,0736		0,05	0,4635	0,1493	0,2874
	0,10	0,0000	0,0000	0,1393		0,10	0,5832	0,1985	0,3901
n=20					n=25				
U.	0,01	0,0003	0,0000	0,0113	U.	0,01	0,0001	0,0000	0,1382
	0,05	0,0030	0,0001	0,1114		0,05	0,0005	0,0176	0,3181
	0,10	0,0098	0,0004	0,2400		0,10	0,0019	0,2320	0,4397
Gama	0,01	0,5776	0,3341	0,8400	Gama	0,01	0,8909	0,5335	0,6610
	0,05	0,7862	0,4169	0,9543		0,05	0,9632	0,6298	0,8066
	0,10	0,8707	0,4657	0,9782		0,10	0,9820	0,6867	0,8610
Exp.	0,01	0,3956	0,2199	0,5997	Exp.	0,01	0,6904	0,3322	0,4090
	0,05	0,6175	0,2892	0,8062		0,05	0,8424	0,4299	0,5886
	0,10	0,7314	0,3337	0,8818		0,10	0,9029	0,4866	0,6648

TABELA 4.3.2 – Avaliação do poder do teste coeficiente de correlação  $r_q$ , simetria e curtose para os casos univariado ( $p=1$ ) e bivariado ( $p=2$ ) com amostras maiores.

p=1					p=2				
n=100					n=100				
dist.	ns	sim.	curt.	$r_q$	dist.	ns	sim.	curt.	$r_q$
U.	0,01	0,0000	0,2506	0,8537	U.	0,01	0,0000	0,8782	0,8813
	0,05	0,0018	0,9496	0,9833		0,05	0,0000	0,9973	0,9758
	0,10	0,0067	0,9941	0,9466		0,10	0,0009	0,9996	0,9902
Gama	0,01	1,0000	0,9629	1,0000	Gama	0,01	1,0000	0,9955	1,0000
	0,05	1,0000	0,9784	1,0000		0,05	1,0000	0,9980	1,0000
	0,10	1,0000	0,9848	1,0000		0,10	1,0000	0,9985	1,0000
Exp.	0,01	0,9999	0,8462	1,0000	Exp.	0,01	1,0000	0,9582	0,9872
	0,05	1,0000	0,8977	1,0000		0,05	1,0000	0,9772	0,9969
	0,10	1,0000	0,9178	1,0000		0,10	1,0000	0,9847	0,9982

TABELA 4.3.3 – Avaliação do poder do teste coeficiente de correlação  $r_q$ , simetria e curtose para o caso multivariado ( $p=6$ ) e ( $p=10$ ).

p=6					p=10				
n=15					n=15				
dist.	ns.	sim.	curt.	$r_q$	dist.	ns.	sim.	curt.	$r_q$
U.	0,01	0,0000	0,0000	0,0088	U.	0,01	0,0000	0,0000	0,0052
	0,05	0,0000	0,0407	0,0434		0,05	0,0000	0,5459	0,0246
	0,10	0,0000	0,3440	0,0883		0,10	0,0000	0,9442	0,0556
Gama	0,01	0,1214	0,0007	0,0002	Gama	0,01	0,0000	0,0000	0,0029
	0,05	0,2968	0,0046	0,0091		0,05	0,0001	0,0545	0,0438
	0,10	0,4135	0,0150	0,0373		0,10	0,0004	0,3557	0,1044
Exp.	0,01	0,0240	0,0000	0,0003	Exp.	0,01	0,0000	0,0000	0,0030
	0,05	0,0965	0,0005	0,0067		0,05	0,0000	0,1268	0,0363
	0,10	0,1731	0,0133	0,0255		0,10	0,0001	0,5508	0,0953
n=25					n=25				
U.	0,01	0,0000	0,0026	0,0289	U.	0,01	0,0000	0,0000	0,0035
	0,05	0,0000	0,3156	0,0867		0,05	0,0000	0,5522	0,0245
	0,10	0,0001	0,6453	0,1453		0,10	0,0000	0,9504	0,0372
Gama	0,01	0,9270	0,3994	0,0462	Gama	0,01	0,6844	0,0482	0,0003
	0,05	0,9753	0,5518	0,1996		0,05	0,8324	0,1232	0,0089
	0,10	0,9890	0,6325	0,3415		0,10	0,8915	0,1881	0,0356
Exp.	0,01	0,6844	0,1354	0,0150	Exp.	0,01	0,6811	0,0500	0,0001
	0,05	0,8289	0,2462	0,0896		0,05	0,8311	0,1254	0,0084
	0,10	0,8856	0,3186	0,1781		0,10	0,8932	0,2003	0,0358

TABELA 4.3.4 – Avaliação do poder do teste coeficiente de correlação  $r_q$ , simetria e curtose para o caso multivariado ( $p=6$ ) e ( $p=10$ ) para amostras maiores.

<b>p=6</b>					<b>p=10</b>				
<b>n=100</b>					<b>n=100</b>				
<b>dist.</b>	<b>ns.</b>	<b>sim.</b>	<b>curt.</b>	<b><math>r_q</math></b>	<b>dist.</b>	<b>ns.</b>	<b>sim.</b>	<b>curt.</b>	<b><math>r_q</math></b>
U.	0,01	0,0000	0,9916	0,2583	U.	0,01	0,0000	0,9942	0,0997
	0,05	0,0001	0,9997	0,4661		0,05	0,0000	1,0000	0,2319
	0,10	0,0003	0,9999	0,5795		0,10	0,0000	1,0000	0,3373
Gama	0,01	1,0000	1,0000	0,9995	Gama	0,01	1,0000	0,9999	0,9745
	0,05	1,0000	1,0000	1,0000		0,05	1,0000	1,0000	0,9954
	0,10	1,0000	1,0000	1,0000		0,10	1,0000	1,0000	0,9987
Exp.	0,01	1,0000	0,9964	0,9767	Exp.	0,01	1,0000	0,9995	0,8454
	0,05	1,0000	0,9994	0,9945		0,05	1,0000	0,9998	0,9558
	0,10	1,0000	0,9916	0,9981		0,10	1,0000	0,9999	0,9790

## 5- CONCLUSÕES

a) A expansão do teste coeficiente de correlação  $r_q$  para amostras maiores referentes ao caso univariado mostrou-se eficiente na identificação de normalidade.

b) O teste coeficiente de correlação  $r_q$  na versão multivariada foi de grande valia devido à escassez de literatura referente a testes propostos para diagnóstico de normalidade em  $p$ -dimensões.

c) O aumento no número de variáveis ocasionou uma redução no poder independente do tamanho amostral.

d) Apresentação das tabelas de probabilidades referentes aos quantis 0,01 0,05 e 0,10 para  $p=1, \dots, 10$ .

## 6- REFERÊNCIAS BIBLIOGRÁFICAS

BERA, A.; JOHN, S. Tests for multivariate normality with Pearson alternatives. **Communications Statistical Theory Methods**, New York, v. 12, n. 1, p. 103-117, 1983.

CAMPOS H. **Estatística experimental não paramétrica**. 4. ed. Piracicaba: ESALQ, 1983.

D'AGOSTINO, R. B.; TIETJEN, G. L. Approaches to null distribution of  $\sqrt{b_1}$ . **Biometrika**, London, v. 60, n. 1, p. 169-173, Apr. 1973.

DECARLO, L. T. On the meaning and use of kurtosis. **Psychological Methods**, Washington, v. 2, n. 3, p. 292-307, Sept. 1997.

GIBBONS, J. D.; SUBHABRATA, C. **Nonparametric Statistical Inference**. 3. ed. New York, 1992. v. 31. (Statistics: textbooks and monographs).

HAWKINS, D. M. A new Test for Multivariate Normality and Homocedasticity. **Technometrics**, Alexandria, v. 23, n. 1, p. 105-110, Feb. 1981.

HEALY, M. J. R. Multivariate normal plotting. **Journal of the Royal Statistical Society Serie Applied Statistics**, London, v. 17, p. 157-161, Apr. 1968.

HORSWELL, R. L.; LOONEY S. W. A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. **Journal Statistical Computation Simulation**, New York, v. 42, n. 1, p. 21-38, 1992.

IEMMA, A. F. **Estatística Descritiva**. Piracicaba: Pop publicações. 1992.

- JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 4. ed. New Jersey: Prentice Hall, 1988.
- KAC, M.; J. KIEFER; J. WOLFOWITZ On tests of normality and other tests of goodness of fit based on distance methods. **Annals of Mathematical Statistics**, Haywara, v. 26, n. 2, p. 189-211, June 1955.
- MALKOVICH, J. F.; AFIFI, A. A. On tests for multivariate normality. **Journal American Statistics Association**, Alexandria, v. 68, n. 341, p. 176-179, Mar. 1973.
- MARONNA, R. A. Robust M-estimators of multivariate location and scatter. **Annals Statistic**, Hayward, v. 4, n. 1, p. 51-67, Jan. 1976.
- MARDIA, K. V. Measures of multivariate skewness and Kurtosis with applications. **Biometrika**, London, v. 57, n. 3, p. 519-530, Dec. 1970.
- MARDIA, K. V. Assessment of multinormality and the robustness of Hotelling's T<sup>2</sup>. **Journal of the Royal Statistical Society Serie Applied Statistics**, London, v. 24, n. 2, p. 163-171, Apr. 1975.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. New York: Academic Press, 1979.
- MECKLIN, C. J.; MUNDFROM J. D. An appraisal and Bibliography of Tests of Multivariate Normality. **Manuscript Submitted to International Statistical Review**. 2000 Disponível em: <<http://\campus.murraystate.edu/academic>>. Acesso em: 12 de Ago. 2002.
- MOOD A. M.; GRAYBILL F. A. .; BOES D. C. **Introduction to the Theory of Statistics**. 3. ed. London: Macgraw-Hill, 1974. Séries in probability and statistics.
- MOORE, D. S.; STUBBLEBINE, J. B. Chi-square tests for multivariate normality with application to common stock prices. **Communications Statistical Theory Methods**, New York, v. 10, n. 8, p. 713-738, 1981.

MORETTIN, P. A; BUSSAB, W. Q. **Estatística Básica**. São Paulo, Editora Atual, 1987.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; LIN, C. T. Some p-variate adaptations of the Shapiro-Wilk test of normality. **Communications Statistics Theory Methods**, New York, v. 24, n. 4, p. 953-985, 1995.

PEARSON, E. S.; HARTLEY, H. O. **Biometrika tables for statisticians**. 3. ed. New York. Cambridge University, 1966. v. 1.

PURI, M. L.; SEN, P. K. **Nonparametric methods in multivariate analysis**. New York: John Wiley, 1971.

ROMEU, J. L.; OZTURK, A. A comparative study of goodness-of-fit tests for multivariate normality. **Journal Multivariate Analysis**, San Diego, v. 46, n. 2, p. 309-334, Aug. 1993.

ROYSTON, J. P An extension of Shapiro and Wilk's W test for normality in large samples. **Applied Statistics**, London, v. 31, n. 2, p. 115-124, 1982.

ROYSTON, J. P Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. **Journal of the Royal Statistical Society Serie Applied Statistics**, London, v. 32, n. 2, p. 121-133, 1983.

ROYSTON, J. P A remark on Algorithm the W-Test For Normality. **Applied Statistics-Journal of the Royal Statistical Society Series**. v. 44, n. 4, p. 547-551, 1995.

SAS INSTITUTE SAS / STAT. **SAS/IML Software: usage and reference**, version 6. Cary, 1990. 501p.

SANTOS, C. A. **Definição do tamanho amostral usando simulação monte carlo para os testes de normalidade univariado e multivariado baseado em assimetria e curtose**. 2001. p. 65. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

SCOTT, D. W. On optimal and data-based histograms. **Biometrika**, London, v. 66, n. 3, p. 605-610, May 1979.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, London, v. 52, n. 3, p. 591, 1965.

SINGH, A. Omnibus robust procedures for assessment of multivariate normality and detection of multivariate outliers. In: PATIL, G. P.; RAO, C. R. (Ed.). **Multivariate Environmental Statistics**. Amsterdsam: North-Holland, 1993. p. 445-488.

SRIVASTAVA, M. S. A measure of skewness and Kurtosis and a graphical method for assessing multivariate normality. **Statistic Probability Letter**, Amsterdam, v. 2, n. 5, p. 263-267, 1984.

SRIVASTAVA, M. S.; HUI, T. K. On assessing multivariate normality based on Shapiro-Wilk W statistic. **Statistic Probability Letter**, Amsterdam, v. 5, n. 15, p. 15-18, Jan. 1987.

STURGES, W. A. The choice of a class Interval. **Journal of the American Statistical Association**, Alexandria, v. 21, n. 153, p. 65-66, Jan. 1926.

WARD, P. J. **Goodness-of-Fit Tests for multivariate Normality**. 1988. Thesis (Ph. D) – University of Alabama, Alabama.

## **CAPÍTULO 2**

### **Avaliação do problema Behrens-Fisher para p-dimensões via Simulação Monte Carlo e Bootstrap.**

## RESUMO

**CIRILLO, Marcelo Angelo. Avaliação do problema Behrens-Fisher para p-dimensões via Simulação Monte Carlo e Bootstrap.. 2003. 69p. (Dissertação – Mestrado em Agronomia / Estatística Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras.\***

Na comparação de dois vetores de médias populacionais, além da suposição de normalidade multivariada é fundamental a homogeneidade das matrizes de covariâncias, pois com esta suposição violada é caracterizado o problema de Behrens-Multivariado, sendo que a principal consequência é um aumento na probabilidade de se cometerem os erros tipo I e II. Com a alteração das taxas referentes a estes erros, o pesquisador tem grande chance de tomar decisões erradas. Sendo assim, neste capítulo objetivou-se estudar via simulação Monte Carlo e bootstrap, métodos multivariados condicionados a diversos fatores, tais como tamanho amostral, graus de heterogeneidade, diferença nas médias populacionais dada por k erros padrão, graus de correlação e valor nominal, em situações de homogeneidade e heterogeneidade das matrizes de covariâncias. Foram realizadas comparações de vetores p-dimensionais de médias populacionais por meio da estatística  $T^2$  de Hotelling. Utilizando-se como critério de avaliação o controle da taxa de erro tipo I e o poder apresentado por cada método analisado na abordagem convencional e bootstrap, concluiu-se que, de um modo geral, o aumento da correlação ocasionou uma redução no poder. No caso da homogeneidade das matrizes de covariâncias, a qual é a situação ideal do teste  $T^2$  de Hotelling. O método de Nel & Van der Merwe (1986) foi o mais adequado para amostras de tamanhos diferentes. Em relação à abordagem bootstrap, o único método que justificou o esforço computacional foi o Johansen (1980).

---

\* Comitê Orientador: Daniel Furtado Ferreira (Orientador)

## ABSTRACT

**CIRILLO, Marcelo Angelo. Evaluation of Behrens-Fisher problem and p-dimensions via Monte Carlo e bootstrap simulation.** Lavras: UFLA, 2003. 69p. (Dissertation – Master Program in Agronomy / Major in Statistics and Agricultural Experimentation)\* – Federal University of Lavras.

In the comparison of two vectors of population means, in addition to the supposition of multivariate normality homogeneity of the covariance matrices is fundamental, since with this supposition violated, Behrens-Fisher Multivariate problem is characterized, the main consequence being an increase in the probability of making the type I and II errors. With the alteration of the rates concerning these errors, the researcher has a great chance of making wrong decisions. Thus, in this chapter, it was aimed to study via Monte Carlo simulation and bootstrap, multivariate methods conditioned to several factors such as sample size, degrees of heterogeneity, difference in the population means given by k standard errors, degrees of correlation and nominal value in situations of homogeneity and heterogeneity of the covariance matrices. Comparisons of p-dimensional vectors of population means by means of Hotteling's statistic  $T^2$  were accomplished by utilizing as an evaluation criterion, the control of the type I error rate and power presented by each method analyzed in the conventional method and bootstrap, it follows that in general, the increase of correlation brought about a reduction in power. In the case of homogeneity of the covariance matrices, which is the ideal situation of Hotteling test  $T^2$ . The method of Nel and Van der Merwe (1986) was the most adequate for the samples of different sizes. Relative to the bootstrap approach, the only method which justified the computational effort was that of Johansen (1980).

---

\* Guidance Committee: Daniel Furtado Ferreira (Orientador)

# 1- INTRODUÇÃO

Na comparação de duas médias populacionais, a homogeneidade das variâncias é de grande importância na qualidade dos resultados e da inferência, pois uma vez violada esta suposição, a principal consequência é o aumento na probabilidade de se cometerem os erros tipo I e II. Com a alteração das taxas referentes a estes erros, o pesquisador tem grande chance de tomar decisões erradas. Como exemplo clássico desta sensibilidade relativa à heterogeneidade de variâncias cita-se o teste t, uma vez que nesta situação a distribuição da estatística não segue exatamente a distribuição t de Student. Para amenizar esta deficiência é necessário adotar procedimentos estatísticos que proporcionem melhor aproximação da estatística do teste, em relação à verdadeira distribuição. Entre estes procedimentos, o primeiro a proporcionar uma aproximação sob a normalidade das populações amostradas testando a hipótese  $H_0:\mu_1=\mu_2$  contra  $H_1:\mu_1\neq\mu_2$  foi Behrens (1929). Posteriormente, Fisher (1935) justificou o uso da teoria da inferência sob perspectiva frequentista. Por serem estes autores precursores no estudo da heterogeneidade das variâncias, este problema ficou conhecido como "*Problema de Behrens-Fisher*". Desde então, a inferência neste problema tem recebido considerável atenção, principalmente nas últimas décadas.

Para amenizar o efeito do problema retromencionado, lembrando que nesse caso o teste t de student não é exato, Satterthwaite (1946) propôs um ajuste nos graus de liberdade. Uma alternativa foi proposta por Cochran & Cox (1957), sendo que ambas foram avaliadas por Borges & Ferreira (1999) utilizando simulação Monte Carlo, concluíram que estes dois métodos são semelhantes em relação ao controle dos erros tipo I e II.

Ghosh & Kim (2001) estudaram o problema de Behrens-Fisher do ponto de vista da teoria Bayesiana. Inicialmente estes autores mostraram que a priori de Jeffrey, aplicada na estimação dos intervalos de credibilidade proporcionou resultados equivalentes aos intervalos de confiança de Nelman-Pearson, diferenciando-se apenas na interpretação. Posteriormente, estes autores propuseram uma priori alternativa com base nesta priori concluíram, por meio de simulação, que os intervalos de credibilidade assintóticos apresentaram maior probabilidade de cobertura para amostras pequenas do que os intervalos de Nelman-pearson e os intervalos de credibilidade utilizando a priori de Jeffreys.

No caso multivariado, este problema foi estudado sob o aspecto não paramétrico por Brunner et al. (2002). Para a realização deste estudo, os autores consideraram duas amostras, provenientes de populações desconhecidas, de modo que as observações foram independentes e determinadas pela igualdade da distribuição marginal. Foi adotado o modelo proveniente da distribuição marginal pelo fato de ter abrangido as observações empatadas. No entanto, as matrizes de covariâncias foram consideradas desconhecidas e o vetor centrado na mediana dos postos foi calculado. Os critérios utilizados foram baseados nas estatísticas do tipo Wald sugerida por Puri e Sem (1971) e do tipo ANAVA, sendo esta uma extensão do caso univariado aplicada a modelos não paramétricos por Brunner et. al. (1997). Por meio de simulações, os autores concluíram que a estatística do tipo ANAVA manteve a taxa de erro tipo I, com bastante precisão em relação ao valor de significância pré-fixado, até mesmo para amostras pequenas. Em relação ao poder, as estatísticas foram praticamente idênticas.

Em relação ao caso paramétrico, basicamente segue-se uma extensão do caso univariado. Neste caso, o teste apropriado é conhecido como  $T^2$  de Hotelling. Nas situações em que está presente a heterogeneidade das matrizes de covariâncias, fica estabelecido o problema multivariado de Behrens-Fisher. Para

o estudo vários autores, tais como Yao (1965), Johansen (1980), Nel & Van Der Merwe (1986) e Kim (1992), propuseram métodos que visam uma correção nos graus de liberdade da estatística do teste, semelhantemente aos métodos propostos por Satterthwaite (1946) e por Cochran & Cox (1957). No entanto, a formalização destes métodos diferem entre si. Christensen & Rencher (1997) avaliaram estes métodos aplicados à estatística  $T^2$  de Hotelling via simulação, comparando seus poderes e suas taxas de erro tipo I e concluíram que as taxas de erro tipo I, para o teste  $T^2$  de Hotelling em todas as situações analisadas, foram sempre inferiores aos níveis nominais, inclusive para as amostras de tamanhos iguais.

O uso de simulações torna-se primordial na avaliação das performances dos métodos em virtude das complexas operações matriciais envolvidas, o que torna inviável um julgamento analítico comparativo dos mesmos. Neste contexto, destaca-se o método Monte Carlo, o qual consiste em simular experimentos de acordo com modelos probabilísticos reais adotados a priori com a finalidade de determinar propriedades de estimadores ou ainda de comparar métodos e testes estatísticos.

Uma possibilidade para o estudo desse problema é o uso das técnicas de bootstrap (Manly, 1998). A idéia básica do bootstrap é que na ausência de qualquer conhecimento sobre uma população, a distribuição de valores retirados em uma amostra aleatória de tamanho  $n$  é o melhor guia à distribuição na população. Sendo assim, a amostra finita de  $n$  observações é reamostrada com reposição e uma estatística de interesse é calculada. Repetindo-se esse procedimento milhares de vezes é possível obter uma distribuição empírica e os resultados obtidos seriam usados para modelar a real população desconhecida de reamostragem, sendo esta utilizada para testar a hipótese de interesse. A possibilidade de se aproximarem funções de probabilidade por meio do método bootstrap teve como um dos precursores Efron & Tibshirani (1993).

Em virtude do que foi mencionado, este capítulo tem como objetivo avaliar as alternativas mencionadas por Christensen & Rencher (1997), aplicadas à estatística  $T^2$  de Hotelling na comparação de vetores de médias populacionais sobre diferentes graus de heterogeneidade das matrizes de covariâncias, considerando a combinação entre os fatores: tamanhos de amostras, estrutura de correlação, diferenças nas medias definidas por diferentes números de erros padrão e valor nominal, utilizando simulações Monte Carlo e bootstrap.

## 2 - REFERENCIAL TEÓRICO

### 2.1 - Comparação de duas médias populacionais

#### 2.1.1 - Caso univariado.

Na comparação de duas médias populacionais, um teste comumente utilizado é o t-student, cuja hipótese a ser avaliada pode ser escrita por  $H_0: \mu_1 = \mu_2$ , em que  $\mu_1$  e  $\mu_2$  representam, respectivamente, a média populacional da primeira e segunda população da qual as amostras serão retiradas para a realização da inferência estatística. A aplicação do teste t de student difere em dois aspectos, homogeneidade e heterogeneidade de variâncias.

No caso da homogeneidade de variâncias, a qual é a situação ideal para a aplicação do teste t-student, a estatística do teste é dada por

$$t = (\bar{X}_1 - \bar{X}_2) / \left( \sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right), \text{ sendo que} \quad (2.1.1.1)$$

$S_p^2$  representa a variância combinada dada pela expressão (2.1.1.2)

$$S_p^2 = \left[ (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 \right] / (n_1 + n_2 - 2) \quad (2.1.1.2)$$

Os graus de liberdade associados à estatística do teste (2.1.1.1) são  $v = n_1 + n_2 - 2$ . No caso da heterogeneidade de variâncias, a estatística do teste t-student é descrita pela expressão (2.1.1.3)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (2.1.1.3)$$

Em que  $\bar{X}_1$  e  $\bar{X}_2$ , representam, respectivamente, os estimadores das médias populacionais referentes à primeira e à segunda população.  $S_1^2$  e  $S_2^2$  referem-se às variâncias amostrais e  $n_1$  e  $n_2$ , aos tamanhos amostrais.

A aplicação do teste t de student requer as suposições de que as amostras sejam independentes, provenientes de populações normais.

A importância da normalidade é justificada pela construção do teste e intervalos de confiança serem dependentes da distribuição amostral do estimador, assim, temos que o estimador da estatística t provém de uma distribuição t-student, a qual é uma derivação da distribuição normal; sendo assim, justifica-se a validação dos resultados sobre a normalidade dos dados. Em relação à questão de que as amostras sejam independentes, esta suposição implica em que as amostras sejam aleatórias. Esta característica é primordial para qualquer procedimento estatístico, pois com a aleatorização das amostras permite-se mensurar os erros cometidos. Um outro fator importante a ser considerado, para a utilização da estatística do testes t-student, como já mencionado anteriormente, é o fato de que as variâncias populacionais sejam preferencialmente homogêneas, pois nesta situação a comparação entre as médias populacionais poderá ser feita sem nenhuma dificuldade. Entretanto, a violação desta suposição conduz ao conhecido problema de Behrens-Fisher.

Na presença de variâncias heterogêneas a distribuição t de Student não é mais adequada a essa estatística, tornando-se necessária a aplicação de algum método para corrigir esta deficiência.

Uma alternativa para amenizar o efeito desta heterogeneidade é dada pela adaptação de métodos que aproximam a estatística t da verdadeira distribuição t-student. Entre estes métodos, os mais utilizados e conhecidos na literatura são o método de Satterthwaite (1946) e Cochran & Cox (1957). Ambos os métodos são utilizados freqüentemente diante da situação de heterogeneidade. Entretanto, a aplicação dos mesmos difere entre si. O método proposto por Satterthwaite (1946) consiste em aplicar uma correção nos graus de liberdade. Esta correção representada por  $\nu$  e definida por:

$$\nu = \frac{\left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left( \frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left( \frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}. \quad (2.1.1.4)$$

O procedimento de Cochran & Cox (1957) consiste em comparar a estatística t com a quantidade  $t^*$  definida por:

$$t^* = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}, \quad (2.1.1.5)$$

em que  $t_1$  e  $t_2$  são os valores tabelados na distribuição t (unilateral) com  $n_1 - 1$  e  $n_2 - 1$  graus de liberdade, ou seja, esses valores referem-se aos quantis superiores  $\alpha/2$  para valor nominal de significância ( $\alpha$ ) previamente fixada e  $w_1$  e  $w_2$  indicam os pesos dados em (2.1.1.6) e (2.1.1.7).

$$w_1 = \frac{S_1^2}{n_1} \quad (2.1.1.6)$$

$$w_2 = \frac{S_2^2}{n_2} \quad (2.1.1.7)$$

Comparando os procedimentos de Satterthwaite (1946) e Cochran & Cox (1957), Borges & Ferreira (1999) avaliaram a aplicação destes dois métodos em diferentes graus de heterogeneidade de variâncias. Para isso, simularam amostras de duas populações, cujos parâmetros foram definidos sem perda de generalidade como  $\sigma^2=1$ . Para a população 1 e para a população 2, a variância foi restrita a razão:

$$k = \frac{\sigma_2^2}{\sigma_1^2}, \quad (2.1.1.8)$$

em que K foi fixado para os valores 1, 2, 4, 8 e 16, o qual foi denominado como grau de heterogeneidade das variâncias. Para cada uma dessas situações, as populações foram geradas com médias iguais e diferentes. No caso em que as médias foram diferentes, os autores consideraram diferenças padronizadas entre as médias populacionais definidas por 1, 2, 3 e 4 erros padrão.

Para cada uma das diferentes situações foram retiradas 2.000 amostras independentes de cada população, sendo aplicado o teste t para hipótese de igualdade entre as médias populacionais  $H_0: \mu_1 = \mu_2$ . Os tamanhos amostrais considerados foram (a)  $n_1=10, n_2=10$ ; (b)  $n_1=30, n_2=30$  e (c)  $n_1=20, n_2=30$ . O número de erros cometidos nas duas mil repetições foi computado para os níveis de significância de 1% e 5%. Nas situações em que as médias populacionais são iguais, avaliou-se a probabilidade de se cometer o erro tipo I. Nos demais casos, a probabilidade de se cometer o erro tipo II foi avaliada.

Para a geração das amostras foi utilizado o método de Monte Carlo, considerando-se o modelo estatístico  $y_{ij} = \mu_i + e_{ij}$ , para cada população, em que  $y_i$  é o  $i$ -ésimo elemento amostral observado,  $\mu$  é a média paramétrica e  $e_i$  é o valor do erro amostral não observável associado com a distribuição  $N(0, \sigma^2)$ . Este modelo foi adotado para cada população, com os parâmetros  $\mu$  e  $\sigma^2$  estipuladas anteriormente. Para obtenção do valor de  $e_i$ , foi usado um algoritmo baseado no teorema da probabilidade integral. Desta forma, os autores concluíram que para o nível de significância de 5%, o poder do teste não foi influenciado pelas aproximações utilizadas. Em relação ao controle da taxa de erro tipo I, verificou-se que o mesmo não foi influenciado pelo aumento no grau de liberdade, proveniente das aproximações aplicadas. Entretanto, este fato não aconteceu com a taxa de erro tipo II. Com o aumento dos graus de liberdade, em conjunto com o aumento da diferença entre médias populacionais, observou-se uma redução na porcentagem do erro tipo II, aumentando assim o poder do teste. A razão entre as variâncias populacionais teve influência apenas na taxa de erro tipo II, ou seja, as taxas empíricas do erro tipo II foram reduzidas à medida que se reduziu a razão entre as variâncias. Este efeito foi verificado similarmente no caso do aumento do tamanho das amostras e das diferenças entre as médias populacionais.

De um modo geral, concluiu-se que as taxas empíricas de erros foram semelhantes para as duas aproximações, portanto evidenciou-se que estas aproximações não influenciaram o poder do teste.

### **2.1.2 - Caso multivariado**

O processo para comparar duas médias populacionais no caso multivariado segue-se o mesmo princípio do caso univariado. Inicialmente, as médias populacionais são representadas por vetores de médias, provenientes das populações a serem comparadas, portanto percebe-se que as operações

necessárias para a realização desta comparação são feitas em uma linguagem matricial.

As suposições para se efetuar a comparação de duas médias populacionais são uma generalização das suposições referentes ao caso univariado. No caso de se realizar inferência a respeito de diferenças de médias populacionais, isto é,  $\underline{\mu}_1 = \underline{\mu}_2$ , pressupõe-se que as médias populacionais sejam provenientes de uma distribuição normal multivariada, definida pela seguinte expressão:

$$f(\underline{X}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ \frac{1}{2} \left[ (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right] \right\} \quad (2.1.2.1)$$

A suposição das amostras serem aleatórias é fundamental para que se defina a estatística do teste conhecida por  $T^2$  de Hotelling, pois esta condição é suficiente para definir a distribuição amostral de  $\bar{X}$  e  $S$ , no caso multivariado ( $p \geq 2$ ),  $\bar{X} \sim N_p(\underline{\mu}, \Sigma/n)$ , de forma análoga ao caso univariado, em que  $\bar{X} \sim N(\mu, \sigma^2/n)$ . Neste caso, para a variância amostral, sabe-se que a distribuição de  $(n-1)s^2/\sigma^2$  possui uma distribuição  $\chi_{n-1}^2$ . Transpondo este conceito para o caso multivariado temos que, a distribuição da matriz de covariância é definida pela distribuição de Wishart, cuja densidade é definida por:

$$W_{n-1} \left( \frac{S}{\Sigma} \right) = \frac{|\Sigma|^{\frac{(n-p-2)}{2}} e^{-\frac{\alpha(S\Sigma^{-1})}{2}}}{2^{p(n-1)} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{\frac{(n-1)}{2}} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n-i)\right)}, \quad (2.1.2.2)$$

em que  $\Gamma\left(\frac{1}{2}(n-i)\right)$  representa a função gama. Desta forma, com a suposição de normalidade multivariada e a distribuição de Wishart, a estatística  $T^2$  é definida como:

$$T^2 = N_p W_{n-1}^{-1} N_p, \quad (2.1.2.3)$$

com  $n-1$  graus de liberdade.

As observações da amostra associada à primeira população ( $n_1$ ) são obtidas de forma independente das amostras da segunda população ( $n_2$ ). Por se tratar do caso multivariado, associam-se as matrizes de covariâncias  $\Sigma_1$  e  $\Sigma_2$  respectivamente à população 1 com média  $\underline{\mu}_1$  e à população 2 com média  $\underline{\mu}_2$ . Os estimadores de  $\Sigma_1$  e  $\Sigma_2$  são dados por  $S_1$  e  $S_2$ . Desta forma, com base na definição da estatística  $T^2$  de Hotelling mencionada em (2.1.2.3) e com os estimadores definidos, pode-se reescrever a estatística  $T^2$  como uma generalização do teste t-student do caso univariado. Assim tem-se a estatística definida em (2.1.2.4), sob a suposição de homocedastia ( $\Sigma_1 = \Sigma_2 = \Sigma$ ), descrita por:

$$T^2 = [\bar{X}_1 - \bar{X}_2]' \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} [\bar{X}_1 - \bar{X}_2], \quad (2.1.2.4)$$

em que,  $S_p$  indica a matriz de covariância amostral combinada descrita por:

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \quad (2.1.2.5)$$

Uma importante condição para que seja garantida a validade deste teste é feita por meio dos tamanhos amostrais. Quando estes são pequenos, é conveniente assumir a mesma matriz de covariância amostral para ambas as populações, ou seja,  $\Sigma_1 = \Sigma_2 = \Sigma$ . A vantagem desta igualdade das matrizes de covariâncias, sob normalidade, é o fato de que a estatística  $T^2$  de Hotelling pode ser obtida por:

$$\frac{(n_1 + n_2 - 2)}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1(\alpha)} \quad (2.1.2.6)$$

No caso de heterogeneidade das matrizes de covariâncias caracteriza-se o problema de Behrens-Fisher no caso multivariado. Em virtude deste problema,

vários autores, tais como Bennet (1951), James (1954), Yao (1965), Johansen (1980), Nel & Van Der Merwe (1986) e Kim (1992), que apresentaram métodos para correção deste problema.

De um modo geral, com exceção do método proposto por Bennett (1951), os demais métodos consistem em aplicar uma correção nos graus de liberdade da distribuição da estatística  $T^2$  de Hotelling, sendo esta definida por James (1954) diferentemente da estatística convencional, Assim uma “aproximação da estatística  $T^2$  de Hotelling” é dada por:

$$T^{2*} = [\bar{X}_1 - \bar{X}_2]' \left( \frac{S_1}{n_1} + \frac{S_2}{n_2} \right)^{-1} [\bar{X}_1 - \bar{X}_2] \quad (2.1.2.7)$$

A avaliação da performance dos últimos quatro métodos retro mencionados, mediante diferentes situações de tamanhos amostrais, valores nominais, graus de heterogeneidade, erros padrão e estrutura de correlação no aspecto convencional e empírico via processos de reamostragens (bootstrap), é objeto de estudo deste trabalho. Entretanto, vários autores, tais como Yao (1965), Subrahmaniam & Subrahmaniam (1973,1975), Algina e Tang (1988), Algina, Oshima & Tang (1991), Kim (1992), De la Rey & Nel's (1993) e Christensen & Rencher (1997), avaliaram os métodos propostos sobre diversas situações, de modo que, as conclusões obtidas foram diversificadas. Entre as conclusões obtidas por estes autores, Christensen & Rencher (1997) ressaltam as principais:

- Yao (1965) mostrou que as probabilidades empíricas, referentes ao controle da taxa de erro tipo I, apresentaram valor inferior em relação ao método proposto por James (1954) em todos os casos.

- Subrahmaniam & Subrahmaniam (1973,1975) compararam os testes de Bennett (1951), James (1954) e Yao (1965) e concluíram que o método de Bennet (1951) apresentou uma proteção praticamente exata para o nível de significância adotado, isto é, controlou a taxa de erro tipo I com probabilidade

muito próxima àquela do valor nominal, porém, este método apresentou um valor baixo de poder para as amostras de tamanhos diferentes. Em relação ao método de James (1954), este se apresentou mais poderoso entre os três métodos avaliados. No entanto, o controle da taxa de erro tipo I proporcionou uma probabilidade maior do que o método de Yao (1965) apenas na situação em que a amostra menor estava associada a maior matriz de covariância.

- Algina & Tang (1988) ratificaram os resultados obtidos por Yao (1965) e posteriormente Algina, Oshima & Tang (1991) incluíram o método de Johansen nas mesmas condições propostas por Yao (1965). Desta forma, verificaram que o controle da taxa de erro tipo I proporcionado por este método é equivalente ao método de Yao.

- Kim (1992) mostrou que o controle da taxa de erro tipo I, proporcionado por seu método de um modo geral, é mais conservativo do que o método de Yao (1965). Em relação ao poder, praticamente não há diferenças entre estes dois métodos, quando as amostras menores estão associadas com a menor matriz de covariância porém, invertendo a ordem das matrizes de covariâncias, de modo que a menor amostra esteja associada à maior matriz de covariância, o método proposto por Kim geralmente é mais poderoso do que o método de Yao (1965).

- De la Rey & Nel (1993) compararam os métodos de Bennet (1951), James (1954), Yao (1965) e Nel & Van Der Merwe (1986) e concluíram que os últimos dois métodos citados apresentaram melhores resultados. O método de James (1954) mostrou-se poderoso quando submetido a altos valores nominais. Este fato foi mais bem constatado de acordo com o incremento na dimensão  $p$ , ou seja, aumentando o número de variáveis.

- Christensen & Rencher (1997) avaliaram estes métodos mencionados anteriormente por meio de simulação Monte Carlo considerando quatro fatores no processo de simulação: número de variáveis ( $p$ ), grau de intercorrelação

presente entre as variáveis (H), grau de heterogeneidade entre as matrizes de covariâncias (d) e o tamanho das amostras descritos pela razão  $n_1:n_2$ . Combinando estes fatores, totalizaram trinta e seis combinações diferentes. Para cada combinação foram realizadas 10.000 repetições, em que, o grau de intercorrelação (H) foi definido em função do algoritmo proposto por Bryce & Maynes (1979). Este algoritmo consiste em impor os autovalores ( $\lambda_i$ ) de qualquer estrutura da matriz de covariância em função do número de variáveis. Desta forma, H é descrito por:

$$H = \frac{\sum_{i=1}^p \frac{1}{\lambda_i}}{p}, \text{ em que, } \sum_{i=1}^p \lambda_i = p. \quad (2.1.2.8)$$

Quando  $H=1$  significa que as variáveis são independentes, altos valores para H indicam forte correlação entre as variáveis. Desta forma, no processo de simulação, Christensen & Rencher (1997) consideraram as situações extremas, isto é,  $H=1$  e  $H=10.000$ , representando, respectivamente, situação de independência e uma “forte” dependência.

Para a simulação deste método, as observações são geradas de várias estruturas de matrizes de covariâncias, sendo padronizadas pelas correlações, com um valor de H fixado.

No caso do grau de heterogeneidade, foi utilizada no processo de simulação a matriz de covariância da segunda população ( $\Sigma_2$ ), definida em função da matriz de covariância da primeira população ( $\Sigma_1$ ), em conjunto com a razão dos tamanhos amostrais multiplicadas por um escalar (d), representando o grau de heterogeneidade. Desta forma, dada a matriz de covariância da primeira população, a matriz de covariância da segunda população, obedeceu a seguinte expressão:

$$\Sigma_2 = \left(\frac{n_1}{n_2}\right) d \Sigma_1. \quad (2.1.2.9)$$

Resumidamente, as situações avaliadas por Christensen e Rencher (1997) são resumidas pela tabela 2.1.2.1.

TABELA 2.1.2.1 – Resumo das combinações dos fatores considerados na avaliação feita por Christensen e Rencher no estudo do problema de Behrens-Fisher.

	<b>d=3;H = 1</b>			<b>d =9; H=10.000</b>		
	<b>p=2</b>	<b>p=5</b>	<b>p=10</b>	<b>p=2</b>	<b>p=5</b>	<b>p=10</b>
<b>n1 &gt; n2</b>	10:5	20:10	30:20	10:5	20:10	30:20
<b>n1 = n2</b>	10:10	20:20	30:30	10:10	20:20	30:30
<b>n1 &lt; n2</b>	10:20	20:40	30:60	10:20	20:40	30:60

Os resultados obtidos por Christensen & Rencher (1997) foram condizentes com os resultados observados por Holloway & Dunn (1967), a respeito da não robustez, por meio do grau intermediário de heterogeneidade das matrizes de variâncias. Em todas as situações avaliadas, os métodos de James (1954), Yao (1965) e Johansen (1980) apresentaram taxas empíricas inflacionadas em relação ao controle da taxa de erro tipo I. Os métodos de Kim (1992) e Nel & Van Der Merwe (1986), apresentaram alto poder entre os métodos avaliados e controle da taxa de erro tipo I conservativo. Devido a estas características estes métodos foram recomendados por Christensen & Rencher (1997) como os mais adequados.

### 3- METODOLOGIA

#### 3.1 – Obtenção das matrizes de covariâncias $\Sigma_1$ e $\Sigma_2$

Para a realização deste trabalho, inicialmente considerou-se o caso normal bivariado e duas populações, de modo que a matriz  $\Sigma_1$  foi fixada em três estruturas de correlações dadas por: (a)  $\rho=0,2$ ; (b)  $\rho=0,5$  e (c)  $\rho=0,9$ . Desta forma, avaliou-se a influência da estrutura de correlação, sob heterogeneidade das matrizes de covariâncias, em que, para cada correlação foram definidas as matrizes:

$$\text{a) } \Sigma_1 = \begin{bmatrix} 1 & 0,2 \\ 0,2 & 1 \end{bmatrix} \quad \text{b) } \Sigma_1 = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix} \quad \text{e c) } \Sigma_1 = \begin{bmatrix} 1 & 0,9 \\ 0,9 & 1 \end{bmatrix}$$

Em cada situação o vetor média da primeira população foi fixado em

$$\underline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

A matriz de covariância da segunda população foi especificada em função da primeira. O vetor média da segunda população foi inicialmente fixado em  $\underline{\mu}_2 = \underline{\mu}_1$  para estudar as taxas do erro tipo I. Posteriormente, o vetor  $\underline{\mu}_2$  foi especificado por:

$$\underline{\mu}_2 = k \begin{bmatrix} \sqrt{\frac{\sigma_{11(1)}^2}{n_1} + \frac{\sigma_{11(2)}^2}{n_2}} \\ \sqrt{\frac{\sigma_{22(1)}^2}{n_1} + \frac{\sigma_{22(2)}^2}{n_2}} \end{bmatrix}, \quad (3.1.1)$$

em que  $k = 1,2$  ou  $3$  referem-se às diferenças entre as médias das duas populações em função do erro padrão;  $\sigma_{ii(j)}^2$  refere-se à variância da  $i$ -ésima variável ( $i=1,2$ ) para  $j$ -ésima população ( $j=1,2$ );  $n_1$  e  $n_2$  são os tamanhos das

amostras retiradas da população 1 e 2, respectivamente. Nesse caso, mediram-se as taxas do erro tipo II para todas as situações em que  $H_0: \underline{\mu}_1 = \underline{\mu}_2$  não foi rejeitada.

Para especificar a matriz de covariância da população 2 ( $\Sigma_2$ ) para cada situação, (a), (b) e (c) foram considerados diferentes graus de heterogeneidade entre as matrizes de covariâncias das duas populações.

Fazendo  $\delta=1, 2$  e  $8$  referindo-se ao grau de heterogeneidade, obteve-se a matriz  $\Sigma_2$  considerando as variâncias generalizadas de ambas as populações,  $\Sigma_2$ , foi obtida da relação:

$$\delta = \frac{|\Sigma_1|}{|\Sigma_2|} \quad (3.1.2)$$

A situação ideal do teste, ou seja, a situação de homocedastia para cada estrutura de correlação, é descrita por:

(a)  $|\Sigma_{1a}| = 0,96 \rightarrow 0,96\delta = |\Sigma_{2a}|$  fixando  $\delta=1$  obteve-se  $\Sigma_{2a}=\Sigma_{1a}$ .

(b)  $|\Sigma_{1b}| = 0,75 \rightarrow 0,75\delta = |\Sigma_{2b}|$  fixando  $\delta=1$  obteve-se  $\Sigma_{2b}=\Sigma_{1b}$ .

(c)  $|\Sigma_{1c}| = 0,19 \rightarrow 0,19\delta = |\Sigma_{2c}|$  fixando  $\delta=1$ , considerou-se  $\Sigma_{2c}=\Sigma_{1c}$ .

Considerando a estrutura de correlação da segunda população igual a da primeira, foi possível mostrar que  $\sigma_{11(2)}\sigma_{22(2)} = \delta$ . Algumas combinações de  $\sigma_{11(2)}$  e  $\sigma_{22(2)}$ , cujo produto fosse que  $\delta$ , foram considerados.

No caso multivariado, para todas as situações de correlação mencionadas no caso bivariado, a matriz  $\Sigma_1$  foi definida pela estrutura de correlação auto-regressiva de ordem 1. Assim,  $\Sigma_1$  foi representada por:

$$\Sigma_1 = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ \rho & 1 & \rho & \dots & \rho^{p-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \dots & 1 \end{bmatrix}. \quad (3.1.3)$$

em que  $\sigma^2=1$ , sem perda de generalidades.

A matriz  $\Sigma_2$  foi obtida em função de  $\Sigma_1$  e representada pela expressão:

$$\Sigma_2 = \sqrt[p]{\delta} \Sigma_1 \quad (3.1.4)$$

Neste estudo, fixou-se o número de variáveis em  $p=5$  e  $\delta =1, 2$  e  $8$  referindo-se ao grau de heterogeneidade.

### 3.2 - Simulação das populações

As duas populações foram amostradas 2.000 vezes para cada situação, (a), (b) e (c), descritas na seção 3.1. Os tamanhos amostrais considerados foram (1)  $n_1=10, n_2=10$  ; (2)  $n_1=10, n_2=20$  e (3)  $n_1=30, n_2=30$ . O vetor de observações amostrado de cada população foi gerado considerando a distribuição normal bivariada da seguinte forma:

$$\underline{X}_{jl} = \underline{F}_j \underline{Z}_j + \underline{\mu}_j \quad l=1, 2, L, n_j \quad j=1, 2, \text{ em que} \quad (3.1.5)$$

$F_j$  é o fator de Cholesky da matriz  $\Sigma_j$  e  $\underline{Z}_l$  é um vetor  $p \times 1$  de variáveis normais padronizadas, geradas pela inversão da função de distribuição da normal para um ponto aleatório  $U, U \sim U[0,1]$ .

Para cada combinação dos fatores sob estudo, estruturas de médias, covariâncias e tamanhos amostrais das populações, foram geradas 2.000 amostras e calculados o teste  $T^2$  convencional e  $T^2$  bootstrap, para cada método estudado considerando 2.000 reamostragens. Foram computados os erros tipo I e

tipo II, conforme a situação como proporções de erros cometidos nas 2.000 simulações realizadas.

### 3.3 – Descrição dos métodos a serem avaliados.

Todos os métodos a serem avaliados, foram baseados na estatística  $T^{*2}$  obtida por James (1954), pela qual uma correção para os valores críticos é obtida mediante a distribuição de  $\chi^2$ . Esta estatística é definida por:

$$T^{*2} = (\bar{y}_1 - \bar{y}_2)' S_e^{-1} (\bar{y}_1 - \bar{y}_2), \text{ em que,} \quad (3.3.1)$$

$$S_e = \frac{S_1}{n_1} + \frac{S_2}{n_2} \quad (3.3.2)$$

A estatística  $T^{*2}$  possui aproximadamente os valores críticos de uma  $\chi_{\alpha,p}^2 (A+B\chi_{\alpha,p}^2)$ , sendo  $\alpha$  o quantil superior de uma distribuição qui-quadrado com  $p$  graus de liberdade e os valores de  $A$  e  $B$  dados por:

$$A = 1 + \frac{1}{2p} \sum_{i=1}^2 \frac{1}{n_i - 1} \left[ \text{tr} \left( S_e^{-1} \frac{S_i}{n_i} \right) \right]^2 \quad (3.3.3)$$

$$B = \frac{1}{2p(p+2)} \sum_{i=1}^2 \frac{1}{n_i - 1} \left\{ \text{tr} \left[ 2 \left( S_e^{-1} \frac{S_i}{n_i} \right)^2 \right] + \left[ \text{tr} \left( S_e^{-1} \frac{S_i}{n_i} \right) \right]^2 \right\} \quad (3.3.4)$$

#### Método de Yao (1965)

O método de Yao (1965) consiste em uma correção nos graus de liberdade, este teste utiliza  $T^{*2}$ , que é definida em (3.3.1), a qual é aproximadamente distribuída  $T^2$  com dimensão  $p$  e  $v$  graus de liberdade, em que  $S_e$  é definido em (3.3.2). Desta forma, a correção nos graus de liberdade é dada por:

$$\frac{1}{\nu} = \frac{1}{(T^{2*})^2} \sum_{i=1}^2 \frac{1}{(n_i-1)} \left[ (\bar{y}_1 - \bar{y}_2)' S_c^{-1} \frac{S_i}{n_i} S_c^{-1} (\bar{y}_1 - \bar{y}_2) \right]^2 \quad (3.3.5)$$

### Método de Johansen (1980)

O procedimento sugerido por Johansen (1980) consiste em dividir  $T^{2*}$  (3.3.1) por  $C$  (3.3.6). Assim, a estatística do teste possui valores críticos aproximadamente  $F_{\alpha,p,\nu}$ , em que  $\alpha$  é o quantil superior de uma distribuição  $F$  com  $p,\nu$  (3.3.9) graus de liberdade.

$$C = p - \frac{2D+6D}{p(p-1)+2}, \text{ cuja quantidade } D \text{ é definida} \quad (3.3.6)$$

$$D = \sum_{i=1}^2 \frac{1}{2(n_i-1)} \left\{ \text{tr}(I - V^{-1}V_i)^2 + [\text{tr}(I - V^{-1}V_i)]^2 \right\} \text{ e } V_i, \text{ definido} \quad (3.3.7)$$

em (3.3.8)

$$V_i = \left( \frac{S_i}{n_i} \right)^{-1} \text{ para } i=1 \text{ ou } 2 \text{ e } V = V_1 + V_2 \quad (3.3.8)$$

$$\nu = \frac{p(p+2)}{3D} \quad (3.3.9)$$

### Método de Nel & Van der Merwe (1986)

A estatística do teste sugerida por Nel e Van der Merwe (1986) utiliza  $T^{2*}$  (3.3.1), o qual é aproximadamente distribuído  $T^2$  com dimensão  $p$  e  $\nu$  graus de liberdade dados por:

$$\nu = \frac{\text{Tr}(S_e)^2 + [\text{Tr}(S_e)]^2}{\frac{1}{n_1-1} \left\{ \text{tr} \left( \frac{S_1}{n_1} \right)^2 + \left[ \text{tr} \left( \frac{S_1}{n_1} \right) \right]^2 \right\} + \frac{1}{n_2-1} \left\{ \text{tr} \left( \frac{S_2}{n_2} \right)^2 + \left[ \text{tr} \left( \frac{S_2}{n_2} \right) \right]^2 \right\}} \quad (3.3.10)$$

Sendo  $S_e$  definido em (3.3.2)

### Método de Kim (1992)

O método de Kim (1992) consiste em uma maximização de pares de formas quadráticas, cuja solução resulta em uma aproximação dos graus de liberdade semelhante ao procedimento de Yao (1965). A princípio, resolve-se o sistema de equações homogêneas:

$$\frac{S_1}{n_1} q = d \left( \frac{S_2}{n_2} q \right) \quad (3.3.11)$$

para os autovalores  $d_k$  e autovetores  $q_k (k=1, 2, \dots, p)$ .

Seja  $D = \text{diag}(d_1, d_2, \dots, d_p)$ ;  $Q = [q_1, q_2, \dots, q_p]$ ;  $W$ ;  $r$ ;  $l_k$ ;  $c$ ; e  $f$  representadas respectivamente, pelas quantidades (3.3.12), (3.3.13), (3.3.14), (3.3.15) e (3.3.16)

$$W = Q' (\bar{y}_1 - \bar{y}_2) \quad e \quad r = \left( \prod d_k \right)^{1/(2p)} \quad (3.3.12) \text{ e } (3.3.13)$$

$$l_k = (d_k + 1) / (d_k^{1/2} + r)^2 \quad (3.3.14)$$

$$c = \frac{\sum l_k^2}{\sum l_k} \quad e \quad f = \frac{(\sum l_k)^2}{\sum l_k^2} \quad (3.3.15) \text{ e } (3.3.16)$$

A estatística do teste é:

$$G = W' (D^{1/2} + rI)^{-2} W \quad (3.3.17)$$

A regra de decisão é dada por rejeitar  $H_0: \mu_1 = \mu_2$  caso seja satisfeita a seguinte relação:

$$G > cf \nu \frac{F_{\alpha, f, \nu-p+1}}{(\nu-p+1)}, \quad (3.3.18)$$

em que,  $F_{\alpha, f, \nu-p+1}$  é o quantil superior de uma distribuição F com  $f$  e  $\nu-p+1$  graus de liberdade cujo valor de  $\nu$  é dado por:

$$\frac{1}{\nu} = \frac{1}{n_1 - 1} \left[ \frac{w' D (D+I)^{-2} W}{w' D (D+I)^{-1} W} \right]^2 + \frac{1}{n_2 - 1} \left[ \frac{w' D (D+I)^{-2} W}{w' D (D+I)^{-1} W} \right]^2 \quad (3.3.19)$$

## 4. - RESULTADOS E DISCUSSÃO

### 4.1 - Taxa de Erro Tipo I

Para todos os métodos em ambas abordagens considerou-se o controle da taxa de erro tipo I aceitável à medida que as probabilidades foram alocadas dentro do intervalo de 99% de confiança para o nível de significância adotado. Desta forma, estudou-se a performance dos métodos relativos ao controle da taxa de erro tipo I na abordagem convencional e bootstrap, condicionados ao número de variáveis  $p=2$  e  $5$ , grau de correlação definidos por  $\rho=0,2, 0,5$  e  $0,9$ , valor nominal em  $1$  e  $5\%$ . Nestas condições, as situações de homogeneidade e heterogeneidade das matrizes de covariâncias foram comparadas.

Um fator importante e primordial para a discussão dos resultados obtidos neste trabalho é dado pela comparação de outros resultados observados por autores diferentes, sendo que, em muitas situações, o estudo relacionado ao problema multivariado de Behens-Fisher é efetuado sob condições não avaliadas neste trabalho. Neste aspecto, pode-se citar a definição dos tamanhos amostrais como inerente a cada autor. Como exemplo, tem-se Fouladi & Yockey (2002). Esses autores definiram as amostras em função do número de variáveis. Devido a estas considerações características de cada autor, a definição do tamanho das amostras estudadas foi feita arbitrariamente e independente do número de variáveis mencionadas.

Os resultados para o caso bivariado (Figura 4.1.1) considerando o valor nominal fixado em  $5\%$  são discutidos a seguir. Na situação de homogeneidade, condicionados a uma “fraca” correlação ( $\rho = 0,2$ ), a comparação dos métodos com base na Figura 4.1.1(a) permitiu que se verificasse que o método de

Johansen (1980) na abordagem convencional não controlou a taxa de erro tipo I para as amostras  $n_1=10$  e  $n_2=10$  e  $n_1=10$  e  $n_2=20$ , de forma que a taxa de erro tipo I real era maior do que a nominal. Com o aumento da correlação para  $\rho=0,9$  (Figura 4.1.1(b)), este efeito novamente foi detectado. Porém, aplicando este método na versão bootstrap em ambas as correlações, esta deficiência no controle da taxa de erro tipo I foi corrigida apresentando probabilidades aproximadas ao valor nominal. Esta correção, de certa forma, em particular para este método, representa uma vantagem ou eficiência da abordagem bootstrap de forma a compensar o esforço computacional extra requerido. No caso das amostras  $n_1=30$  e  $n_2=30$ , independentemente da correlação, este método não apresentou problemas no controle da taxa de erro tipo I, de modo que suas probabilidades foram consideradas iguais ao valor nominal. Estes resultados ratificam os resultados obtidos por Algina et. al. (1991) que constataram que este método não é apropriado para as amostras pequenas independentemente do número de variáveis.

Outra característica em relação ao método de Johansen (1980) foi observada por Christensen & Rencher (1997), segundo os quais as probabilidades do erro tipo I foram sempre superiores ao valor nominal, independentemente do número de variáveis ou do grau de heterogeneidade das matrizes de covariâncias. Entretanto, os autores ressaltam que esta superioridade não é significativa.

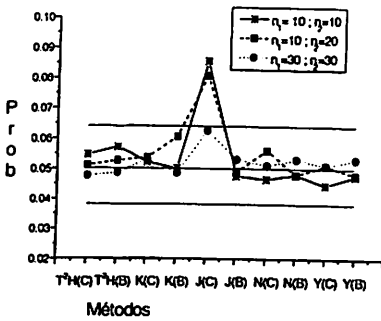
Em relação aos demais métodos, pode-se afirmar que diante das correlações estudadas, o controle da taxa de erro tipo I se deu de maneira semelhante pelo fato de as probabilidades terem sido aproximadas, sendo que na situação de “forte” correlação, ou seja,  $\rho=0,9$  (Figura 4.1.1(b)), os métodos apresentaram um decréscimo nas taxas empíricas. Este efeito foi observado principalmente para as amostras  $n_1=10$  e  $n_2=20$ , sendo mais fortemente constatado no método de Kim (1992) em ambas as abordagens, cujas

probabilidades foram exatamente as mesmas daquelas observadas na estatística  $T^2$  de Hotelling. No caso das amostras de tamanhos iguais,  $n_1=10$  e  $n_2=10$  e  $n_1=30$  e  $n_2=30$ , este efeito mostrou-se ausente, de forma que as probabilidades do erro tipo I, proveniente deste método, permaneceram praticamente imutáveis, concordando com resultados observados por Christensen & Rencher (1997), os quais afirmaram que as taxas empíricas referentes ao controle da taxa de erro tipo I, em particular a este método, não foram inflacionadas.

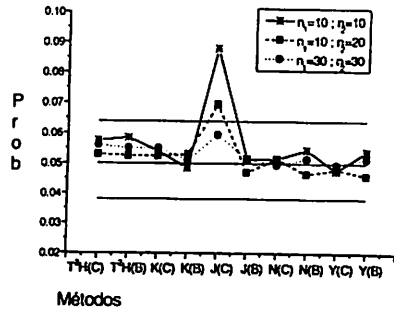
Na situação de máxima heterogeneidade admitida neste trabalho, ou seja,  $\delta = 8$  para  $\rho = 0,2$  (Figura 4.1.1(c)), verificou-se que no caso das amostras  $n_1=30$  e  $n_2=30$  as probabilidades do erro tipo I foram aproximadamente iguais ao valor nominal, concordando com os resultados obtidos por Ito & Schull (1964). O teste  $T^2$  de Hotelling em ambas as abordagens não controlou a taxa de erro tipo I para as amostras  $n_1=10$  e  $n_2=20$ , concordando com as observações feitas por Holloway & Dunn (1967), os quais afirmaram a não robustez da estatística  $T^2$  de Hotelling sobre heterogeneidade de covariâncias. Entretanto, considerando a heterogeneidade das correlações, os resultados relacionados à estatística  $T^2$  de Hotelling contrariou os resultados observados por Fouladi & Yockey (2002). Estes autores afirmaram que para as amostras de tamanhos diferentes, a estatística  $T^2$  de Hotelling é robusta, desde que a magnitude das correlações não seja alta. Diante desta situação, ficou evidenciada a necessidade de se aplicar algum método, seja na versão bootstrap ou convencional, que corrigisse a deficiência apresentada pela estatística  $T^2$  de Hotelling relacionada às amostras de tamanhos diferentes. Com este enfoque, verificou-se uma vantagem nos métodos aplicados via bootstrap em relação aos métodos na abordagem convencional, pois na abordagem bootstrap todos os métodos controlaram a taxa de erro tipo I. Entretanto, para as amostras de tamanhos  $n_1=10$  e  $n_2=10$  ocorreu uma inversão de modo que apenas este método controlou a taxa de erro tipo I (sob heterogeneidade de variâncias), sendo que os demais métodos foram

ineficientes. Em relação ao método de Johansen (1980), novamente a abordagem bootstrap foi satisfatória devido à correção feita no controle da taxa de erro tipo I, cujo problema anteriormente era apresentado na versão convencional. Este mesmo comportamento foi verificado aumentando a correlação para  $\rho=0,9$  (Figura 4.1.1(d)). Para as amostras  $n_1=30$  e  $n_2=30$ , o aumento da correlação apresentou um efeito positivo para o método de Kim (1992) convencional, pois com baixa correlação (Figura 4.1.1(c)) este método não havia controlado a taxa de erro tipo I, aumentando a correlação para  $\rho=0,9$  (Figura 4.1.1(d)). Este método passou a controlar a taxa de erro tipo I. Já a abordagem bootstrap continuou não satisfatória em relação ao controle da taxa de erro tipo I, independentemente da correlação. No caso das amostras pequenas,  $n_1=10$  e  $n_2=10$ , o aumento da correlação influenciou os métodos na abordagem bootstrap, levando ao não controle da taxa do erro tipo I e favorecendo os métodos na abordagem convencional, de forma que as probabilidades provenientes dos métodos sofreram uma redução. Esta redução foi importante em alguns métodos, tais como o método de Nel & Van Der Merwe (1986), que passou a controlar a taxa de erro tipo I, com probabilidade aproximada ao método de Yao (1965).

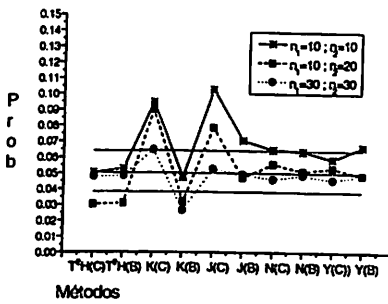
Para as amostras  $n_1=10$  e  $n_2=20$ , com o aumento da correlação todos os métodos de um modo geral, proporcionaram probabilidades inferiores ao valor nominal, exceto o método de Kim (1992) convencional. Uma importante característica verificada com o aumento da correlação, em particular para este tamanho amostral, foi observado na estatística  $T^2$  de Hotelling, a qual passou a controlar eficientemente a taxa de erro tipo I, com probabilidades próximas ao valor nominal, concordando com os resultados feito por Fouladi & Yockey (2002) .



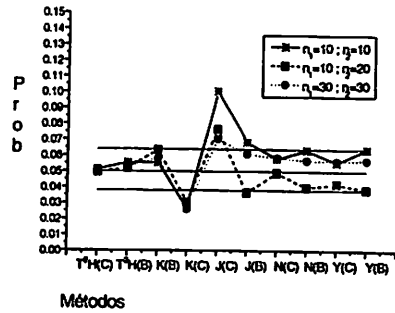
a -  $\delta=1$  e  $\rho=0,2$



b -  $\delta=1$  e  $\rho=0,9$



c -  $\delta=8$  e  $\rho=0,2$



d -  $\delta=8$  e  $\rho=0,9$

**FIGURA 4.1.1** – Taxa de Erro tipo I, fixado o valor nominal em 5% e  $p=2$  comparando homogeneidade com máxima heterogeneidade.

Segundo as mesmas definições anteriores relacionadas ao número de variáveis ( $p=2$ ), correlações e heterogeneidade das matrizes de covariâncias, os resultados apresentados na Figura 4.1.2 diferenciaram apenas no nível de significância, sendo este fixado em 1%. Assim, verificou-se basicamente o mesmo comportamento em relação ao controle da taxa de erro tipo I, sendo que os seguintes resultados merecem destaques.

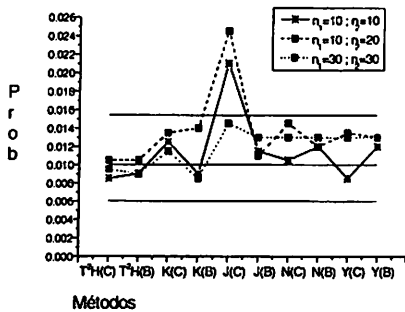
Na situação de homogeneidade, fixada a correlação  $\rho=0,2$  (Figura 4.1.2(a)), verificou-se que a redução no valor nominal não influenciou o controle da taxa de erro tipo I, portanto continuaram válidas as mesmas conclusões sobre a taxa de erro tipo I observadas na Figura 4.1.1(a). Entretanto,

algumas particularidades foram ressaltadas. Entre estas particularidades observou-se que as probabilidades referentes ao controle da taxa de erro tipo I, relacionadas às amostras  $n_1=30$  e  $n_2=30$ , mostraram-se iguais entre os métodos Nel & Van Der Merwe (1986) e Yao (1965) convencional. Esta igualdade nas taxas de erro tipo I foram fatores que determinaram a recomendação destes métodos feita por De La Rey & Nel (1993). A mesma ocorrência foi verificada na abordagem bootstrap. No caso das amostras menores,  $n_1=10$  e  $n_2=10$  e  $n_1=10$  e  $n_2=20$ , notou-se que as diferenças nas taxas empíricas entre estes métodos não foram consideradas como um fator de avaliação entre as abordagens convencional e bootstrap. Seguindo a mesma situação de homogeneidade, porém aumentando o grau de correlação para  $\rho = 0,9$  os resultados na Figura 4.1.2(b) permitiram que se inferisse que a redução no valor nominal prejudicou o método de Johansen (1980) convencional, de modo que para as amostras  $n_1=30$  e  $n_2=30$  passaram a não controlar a taxa de erro tipo I. Em relação aos tamanhos amostrais  $n_1=10$  e  $n_2=10$  e  $n_1=10$  e  $n_2=20$ , foram observados os mesmos comportamentos apresentados na Figura 4.1.1(b).

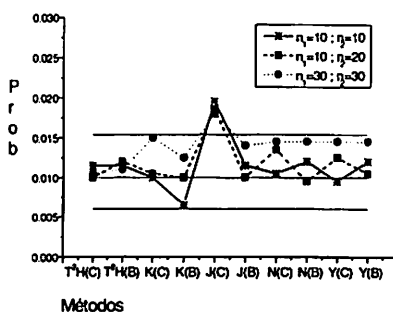
Uma outra característica observada foi em relação ao método de Kim (1992) bootstrap, especificamente para as amostras  $n_1=10$  e  $n_2=10$ , cuja redução no valor nominal ocasionou uma deflação na taxa empírica do erro tipo I. Para os demais métodos, esta deflação foi relativamente pequena; no caso das amostras  $n_1=30$  e  $n_2=30$ , novamente o método de Nel & Van Der Merwe (1986) e de Yao (1965) apresentaram taxas iguais, mostrando-se invariantes quanto ao aumento da correlação.

Na situação de heterogeneidade com  $\rho = 0,2$  (Figura 4.1.2(c)), a redução do valor nominal não apresentou vantagens em relação à estatística  $T^2$  de Hotelling para as amostras  $n_1=10$  e  $n_2=20$ . Na avaliação geral dos métodos a redução do valor nominal favoreceu o método de Kim (1992) bootstrap, sendo este mais adequado no controle da taxa de erro tipo I, principalmente para as

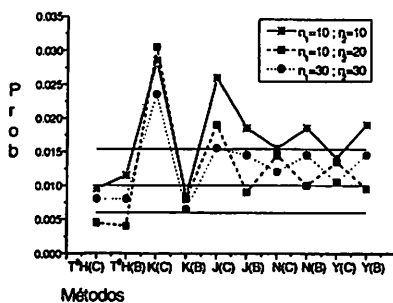
amostras  $n_1=10$ ,  $n_2=20$  e  $n_1=30$ ,  $n_2=30$ . Em relação aos outros métodos, as mesmas características foram mantidas, principalmente para as amostras  $n_1=10$  e  $n_2=10$ , as quais, evidenciaram a ineficiência da abordagem bootstrap quando submetido aos valores nominais pequenos. Aumentando o grau de correlação (Figura 4.1.2(d)), as diferenças entre os valores nominais de 1% e 5% foram maiores, de tal forma que o método de Kim (1992) convencional passou a não controlar a taxa de erro tipo I para as amostras  $n_1=10$ ,  $n_2=20$  e  $n_1=30$ ,  $n_2=30$ . Entretanto, para estes tamanhos amostrais a aplicação deste método via bootstrap corrigiu esta deficiência, o que, para o valor nominal fixado em 5% (Figura 4.1.1(d)) persistia o não controle da taxa de erro tipo I. Em relação ao método de Johansen (1980), as mesmas características, quando fixado o valor nominal em 5% para todos os tamanhos amostrais, foram observadas (Figura 4.1.1(d)). No caso dos demais métodos, a redução no valor nominal refletiu positivamente quando os testes eram aplicados via bootstrap, de forma que estes métodos passaram a apresentar taxas empíricas próximas ao valor nominal.



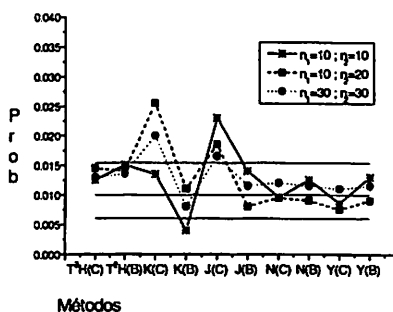
a -  $\delta=1$  e  $\rho=0,2$



b -  $\delta=1$  e  $\rho=0,9$



c -  $\delta=8$  e  $\rho=0,2$



d -  $\delta=8$  e  $\rho=0,9$

**FIGURA 4.1.2** – Taxa de Erro tipo I, fixado o valor nominal em 1% e  $p=2$  comparando homogeneidade com máxima heterogeneidade.

Os resultados apresentados na Figura 4.1.3 são provenientes das mesmas condições relacionada nas Figuras 4.1.1 e 4.1.2. No entanto, o enfoque é dado sobre a influência do aumento do número de variáveis no controle da taxa de erro tipo I. Desta forma, fixando  $p=5$  e valor nominal em 5%, foram obtidos os resultados doravante apresentados e discutidos.

Na situação de homogeneidade inicialmente fixada, a correlação  $\rho=0,2$  (Figura 4.1.3(a)), de um modo geral os métodos continuaram apresentando o mesmo comportamento em relação ao controle da taxa de erro tipo I. O método de Johansen (1980) convencional novamente não controlou a taxa de erro tipo I para as amostras  $n_1=10, n_2=10$  e  $n_1=10, n_2=20$ , o que levou a confirmar os

resultados verificados por Algina et.al (1991), os quais afirmaram que este método era invariante quanto ao número de variáveis no controle do erro tipo I para as amostras de tamanhos iguais e pequenas. Entretanto, a aplicação deste método na abordagem bootstrap corrigiu esta deficiência apresentada pela abordagem convencional para estas amostras. No caso das amostras  $n_1=30$  e  $n_2=30$ , este método não apresentou problemas. Em relação aos demais métodos, verificou-se que, em ambas abordagens, as amostras  $n_1=30$  e  $n_2=30$  praticamente apresentaram taxas empíricas semelhantes. Esta semelhança foi enfatizada por meio dos métodos Nel & Van der Merwe (1986) e Yao (1965), cujas taxas empíricas foram aproximadamente similares àquelas do valor nominal, conforme resultados observados por Christensen & Rencher (1997). No caso das amostras  $n_1=10$  e  $n_2=20$ , verificou-se que estes métodos na abordagem bootstrap apresentaram taxas empíricas inferiores ao valor nominal, sendo contraditórias aos resultados observados por Christensen & Rencher (1997). No caso das amostras  $n_1=10$  e  $n_2=10$ , com exceção do método de Johansen (1980), os demais métodos em ambas as abordagens, apresentaram taxas inferiores à estatística  $T^2$  de Hotelling e superiores ao valor nominal. Aumentando a correlação para  $\rho = 0,9$  (Figura 4.1.3(c)), verificou-se que os resultados apresentados pelo método de Johansen (1980) não foram afetados, pois independentemente do tipo de abordagem, este método continuou apresentando as mesmas características relacionadas à taxa de erro tipo I, quando comparadas com os resultados apresentados na Figura 4.1.1(b). Em relação aos demais métodos, o aumento do número de variáveis praticamente não influenciou as taxas empíricas no controle da taxa de erro tipo I. Porém, algumas considerações foram feitas em relação aos tamanhos amostrais. Entre estas considerações, ressaltou-se que para as amostras  $n_1=10$ ,  $n_2=10$ , com exceção do método de Kim (1992), os demais métodos na abordagem bootstrap apresentaram taxas empíricas maiores em relação às amostras  $n_1=10$ ,  $n_2=20$  e  $n_1=30$ ,  $n_2=30$ . Para estes tamanhos amostrais, ou seja,

em amostras maiores e de tamanhos diferentes, os métodos de Nel & Van der Merwe (1986) e Yao (1965) praticamente foram idênticos.

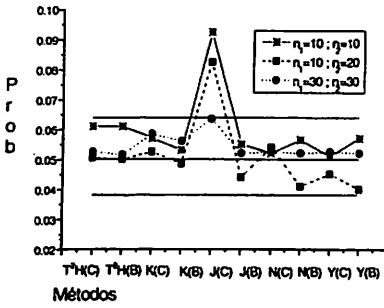
Na situação de heterogeneidade, com baixa correlação ( $\rho = 0,2$ ) conforme resultados apresentados na Figura 4.1.3(c), observou-se que as taxas empíricas da estatística  $T^2$  de Hotelling em ambas abordagens não estavam de acordo com a taxa nominal de erro tipo I para as amostras  $n_1=10$  e  $n_2=20$ . Em relação aos métodos, o mais crítico na abordagem convencional foi o método de Johansen (1980), sendo o que apresentou o não controle da taxa de erro tipo I. Entretanto, quando este teste foi aplicado via bootstrap, este problema foi corrigido. Em relação aos métodos na abordagem bootstrap, apenas o método de Kim (1992) apresentou problemas para todos os tamanhos amostrais. No caso dos demais métodos, as taxas empíricas foram semelhantes e próximas do valor nominal. Aumentando a correlação para  $\rho = 0,9$  conforme resultados apresentados na Figura 4.1.3(d), a estatística  $T^2$  de Hotelling continuou a apresentar o mesmo problema para as amostras de tamanhos diferentes, assim como também os métodos de Kim (1992) bootstrap e Johansen (1980) convencional. Para as amostras de tamanhos iguais, respectivamente as amostras  $n_1=10, n_2=10$  e  $n_1=30, n_2=30$  foram beneficiadas com o aumento da correlação, de tal forma que estes métodos passaram a controlar a taxa de erro tipo I. Em relação aos demais métodos e tamanhos amostrais, as diferenças ocasionadas pelo aumento da correlação não foram expressivas.

Aumentando o grau de heterogeneidade simultaneamente com o número de variáveis, conforme resultados apresentados nas Figuras 4.1.1(a) e 4.1.3(c), sujeitas à “fraca” correlação ( $\rho = 0,2$ ) e valor nominal fixado em 5%, observou-se que apenas o método de Kim (1992) bootstrap foi prejudicado, pois na situação inicial  $p = 2$  e  $\delta = 1$  (Figura 4.1.1(a)), este método controlava a taxa de erro tipo I para todos os tamanhos amostrais, com probabilidade próxima ao valor nominal. Aumentando a heterogeneidade para  $\delta = 8$  e o número de

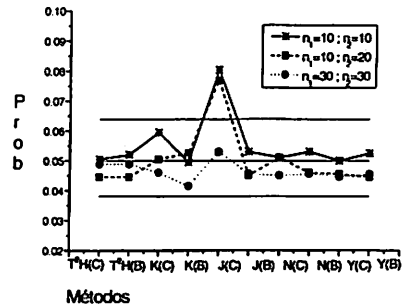
variáveis para  $p = 5$ , este método passou a não controlar a taxa de erro tipo I para todos os tamanhos amostrais. Na abordagem convencional, o método de Johansen (1980) foi bastante prejudicado, para as amostras  $n_1=30$  e  $n_2=30$ , contrariando os resultados observados por Algina et. al. (1991), pois sobre heterogeneidade e condicionado a este tamanho amostral, o método de Johansen (1980) passou a ser sensível em relação ao número de variáveis. Em relação aos demais métodos, o aumento da heterogeneidade em conjunto com o número de variáveis apenas refletiu nos valores das probabilidades, de tal forma que, estes valores provenientes dos métodos para todos os tamanhos amostrais, em ambas as abordagens, passaram a controlar a taxa do erro tipo I, exceto o método de Johansen (1980) convencional quando aplicado às amostras  $n_1=10$  e  $n_2=10$  e  $n_1=10$ ,  $n_2=20$ .

Diante da mesma situação estudada, porém aumentando a correlação para  $\rho = 0,9$  os resultados apresentados nas Figuras 4.1.1(b) e 4.1.3(d) permitiram que se inferisse que, na situação inicial, ou seja,  $\delta = 1$  e  $p = 2$  (Figura 4.1.1(b)), apenas o método de Johansen (1980) convencional, apresentou problemas relativo ao controle da taxa de erro tipo I para as amostras  $n_1=10$ ,  $n_2=10$  e  $n_1=10$ ,  $n_2=20$ , sendo que, nesta última situação amostral a estatística  $T^2$  de Hotelling apresentou os mesmos problemas mencionados anteriormente. Aumentando o grau de heterogeneidade e o número de variáveis, isto é,  $\delta=0,8$  e  $p=5$  (Figura 4.1.3(d)), este método continuou a não controlar a taxa de erro tipo I, para essas amostras, sendo consistente com as observações feitas por Holloway & Dunn (1967) em relação ao incremento do número de variáveis. Em relação à abordagem bootstrap, os métodos de um modo geral foram mais prejudicados em relação ao controle da taxa de erro tipo I para as amostras  $n_1=10$ ,  $n_2=20$ , pois na situação inicial  $\delta = 1$  e  $\rho = 0,9$  todos os métodos na abordagem bootstrap, inclusive a própria estatística  $T^2$  de Hotelling, controlava a taxa de erro tipo I. Aumentando o número de variáveis para  $p=5$  e  $\delta=8$ , foi

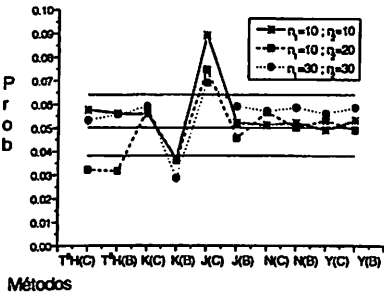
observado que a estatística  $T^2$  de Hotelling e o método de Kim (1992) ambos na abordagem bootstrap, e Johansen (1980) convencional tiveram suas taxas empíricas deflacionadas, passando a não controlar a taxa de erro tipo I. Os métodos de Nel & Van der Merwe (1986) e Yao (1965) continuaram a apresentar taxas empíricas bem aproximadas, porém distantes do valor nominal, de modo que seus respectivos valores ficaram aproximadamente iguais ao limite inferior de controle. Este efeito ocorreu com a mudança na heterogeneidade e o com o número de variáveis e refletiu apenas uma redução nos valores das taxas empíricas, concordando com os resultados observados por Christensen & Rencher (1997). Esses autores afirmaram que em nenhuma situação o método de Yao (1965) mostrou-se conservativo, quando comparado com outros métodos.



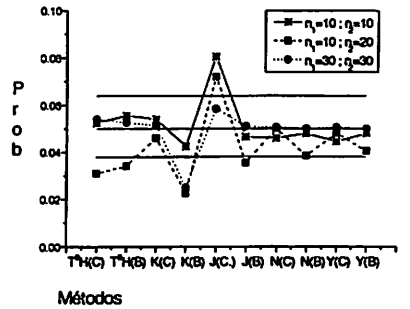
a -  $\delta=1$  e  $\rho=0,2$



b -  $\delta=1$  e  $\rho=0,9$



c -  $\delta=8$  e  $\rho=0,2$



d -  $\delta=8$  e  $\rho=0,9$

**FIGURA 4.1.3 – Taxa de Erro tipo I, fixado o valor nominal em 5% e  $p=5$  comparando homogeneidade com máxima heterogeneidade.**

As discussões apresentadas a seguir foram feitas com base na observação da Figura 4.1.4, visando a complementar os resultados obtidos anteriormente, de modo que a avaliação foi feita nas mesmas situações amostrais, heterogeneidade e grau de correlação, diferenciando apenas no valor nominal, sendo este definido com o valor de 1%. Desta forma, fixado  $p=5\%$  os principais resultados obtidos foram apresentados e discutidos a seguir.

Na situação de homogeneidade, com a correlação  $\rho = 0,2$  (Figura 4.1.4(a)), verificou-se que a redução no valor nominal de um modo geral não afetou os métodos, pois estes basicamente apresentaram o mesmo comportamento no controle da taxa de erro tipo I, diferenciando apenas quanto

ao número de variáveis em relação à Figura 4.1.2(a). Esta diferença repercutiu nas amostras  $n_1=10$  e  $n_2=20$ , em que inicialmente, quando fixado  $p=2$  (Figura 4.1.2(a)), o método de Johansen (1980) convencional apresentou problemas no controle da taxa de erro tipo I, ao passo que, aumentando o número de variáveis para  $p=5$  (Figura 4.1.4(a)), esta deficiência foi corrigida. Em relação aos demais métodos, todos controlaram a taxa de erro tipo I em ambas as abordagens. A redução do valor nominal, juntamente com o aumento do número de variáveis, ocasionou uma redução nos valores das taxas empíricas praticamente em todos os métodos. Este efeito especificamente foi observado nas amostras  $n_1=10$ ,  $n_2=10$  e  $n_1=10$ ,  $n_2=20$  em ambas as abordagens. No caso das amostras  $n_1=30$  e  $n_2=30$ , as taxas empíricas foram exatamente iguais ao valor nominal, com exceção dos métodos de Kim (1992) e Johansen (1980) convencional. Aumentando o grau de correlação para  $\rho = 0,9$  (Figura 4.1.4(b)), estes métodos continuaram a controlar a taxa de erro tipo I. O aumento na correlação levou o método de Johansen (1980) convencional a não controlar a taxa de erro tipo I especificamente para as amostras  $n_1=10$  e  $n_2=20$  (Figura 4.1.4(b)). Os demais métodos praticamente não mostraram diferenças em relação ao aumento do número de variáveis e grau de heterogeneidade, com exceção do método de Kim (1992) bootstrap, que apresentou taxa empírica de erro tipo I próxima do valor nominal, quando condicionada à alta correlação e ao número de variáveis definido por  $p=5$ .

Na situação de heterogeneidade, fixado  $\rho = 0,2$  os resultados da Figura 4.1.4(c) mostraram que as amostras  $n_1=10$ ,  $n_2=10$  e  $n_1=10$ ,  $n_2=20$  relacionadas aos métodos de Kim (1992) bootstrap e Johansen (1980) convencional, não controlaram a taxa de erro tipo I, sendo que esta deficiência relacionada ao método de Johansen (1980) foi corrigida com a versão bootstrap. Comparando esses resultados com aqueles apresentados na Figura 4.1.2(c), referente à mesma situação de heterogeneidade e correlação, diferenciando

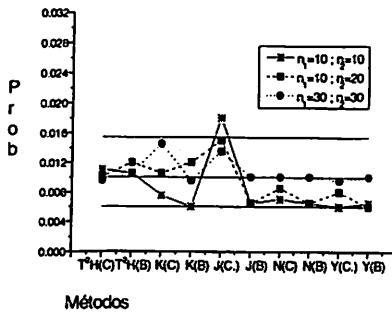
apenas no número de variáveis ( $p=2$ ), observou-se que o aumento no número de variáveis para  $p=5$  prejudicou o método de Kim (1992) na abordagem bootstrap. Este método controlou de maneira conservativa a taxa de erro tipo I para todos os tamanhos amostrais, concordando com os resultados obtidos por Christensen & Rencher (1997).

Em relação ao método de Johansen (1980) convencional, verificou-se que o aumento do número de variáveis não contribuiu no controle da taxa de erro tipo I. Porém, analisando os métodos no âmbito geral, percebeu-se que o aumento no número de variáveis contribuiu de forma notória, principalmente para as amostras  $n_1=10$ ,  $n_2=10$ , de forma que os métodos que apresentaram problema anteriormente passaram a controlar a taxa de erro tipo I com probabilidades próximas ao valor nominal. No caso das amostras de tamanhos  $n_1=10$  e  $n_2=20$ , não apresentaram efeito considerando a mudança no número de variáveis. Para as amostras de tamanhos  $n_1=30$ ,  $n_2=30$ , apenas verificou-se uma redução nas taxas empíricas.

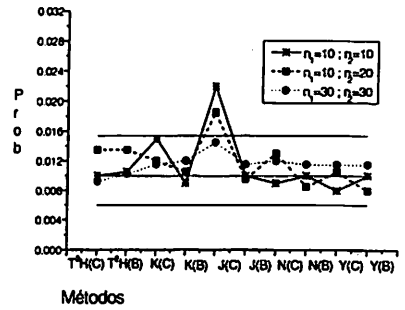
Avaliando a consequência do aumento do grau de correlação, na situação de heterogeneidade, os resultados da Figura 4.1.4(d) permitiram que se inferisse que o método de Johansen (1980) convencional continuou a não controlar a taxa de erro tipo I para as amostras  $n_1=10$ ,  $n_2=10$  e  $n_1=10$ ,  $n_2=20$ . O método de Kim (1992) apresentou-se melhor devido à correção feita no controle da taxa de erro tipo I referente às amostras  $n_1=10$  e  $n_2=10$ . Em particular, foi observado que com o aumento da correlação e para as amostras  $n_1=10$  e  $n_2=20$ , o método de Johansen (1980) bootstrap controlou a taxa de erro tipo I, apresentando-se conservador. Também nesta situação amostral, a própria estatística  $T^2$  de Hotelling mostrou-se não robusta com o aumento do grau de heterogeneidade, conforme retratado nas Figuras 4.1.4(b) e 4.1.4(d). Em relação aos outros métodos, o aumento da heterogeneidade não resultou características que possam

alterar o controle da taxa de erro tipo I, principalmente para as amostras de tamanhos iguais, cujas taxas empíricas foram aproximadas.

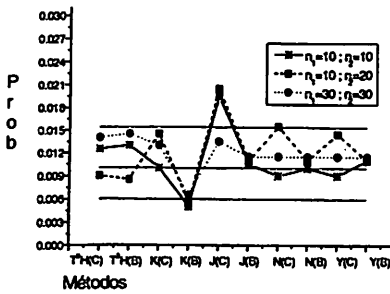
Aumentando o grau de heterogeneidade simultaneamente com o número de variáveis, conforme resultados apresentados nas Figuras 4.1.2(a), 4.1.4(c), sujeitas à fraca correlação ( $\rho=0,2$ ) e valor nominal definido em 1%, verificou-se que apenas o método de Kim (1992) bootstrap foi prejudicado, pois na situação inicial  $p=2$ ,  $\delta=1$  este método controlava a taxa de erro tipo I para todos os tamanhos amostrais. Aumentando a heterogeneidade para  $\delta=8$  e o número de variáveis para  $p=5$ , este método passou a não controlar a taxa de erro tipo I para as amostras  $n_1=10$ ,  $n_2=10$  e  $n_1=10$ ,  $n_2=20$ , sendo que, para as amostras  $n_1=30$ ,  $n_2=30$  mostrou-se conservativo. Os demais métodos apresentaram apenas mudanças nos valores referentes às suas taxas empíricas, não perdendo a propriedade de controlar a taxa de erro tipo I. Aumentando a correlação para  $\rho=0,9$  e analisando o impacto do aumento do grau de heterogeneidade simultaneamente ao número de variáveis, os resultados apresentados nas Figuras 4.1.2(b) e 4.1.4(d) permitiram que se inferisse que na situação inicial, ou seja, para  $\delta=1$  e  $p=2$  (Figura 4.1.2(b)), apenas o método de Johansen (1980) convencional apresentou problemas relativo ao controle da taxa de erro tipo I para todas as amostras. Aumentando o valor da heterogeneidade e o número de variáveis, isto é,  $\delta=0,8$  e  $p=5$  (Figura 4.1.4(d)), apenas as amostras  $n_1=30$  e  $n_2=30$  apresentaram correção na falta de controle da taxa de erro tipo I. Na abordagem bootstrap, os métodos de Kim (1992) e Johansen (1980) foram influenciados, de tal forma que o aumento no grau de heterogeneidade, refletiu nas amostras  $n_1=30$ ,  $n_2=30$  e  $n_1=10$ ,  $n_2=20$  relacionadas ao método de Kim (1992) apresentando o controle da taxa de erro tipo I de forma conservativa. No caso do método de Johansen (1980) para amostras de tamanhos diferentes, apresentou a mesma característica ao método de Kim (1992). Os demais métodos, em ambas as abordagens, não apresentaram problemas com este efeito.



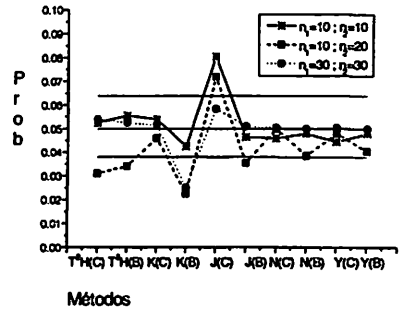
a -  $\delta=1$  e  $\rho=0,2$



b -  $\delta=1$  e  $\rho=0,9$



c -  $\delta=8$  e  $\rho=0,2$




d -  $\delta=8$  e  $\rho=0,9$

**FIGURA 4.1.4** – Taxa de Erro tipo I, fixado o valor nominal em 1% e  $p=5$  comparando homogeneidade com máxima heterogeneidade.

Os resultados discutidos a seguir e apresentados na Figura 4.1.5 são específicos para a situação de heterogeneidade das matrizes de covariância, em que, as Figuras 4.1.5(a); (b) e 4.1.5(c); (d), respectivamente, indicam a situação de baixa heterogeneidade ( $\delta=2$ ) e alta heterogeneidade admitida neste trabalho ( $\delta=8$ ). Para todas as situações, particularizou-se o valor nominal fixado em 5% e o número de variáveis  $p=5$ . Desta forma, compararam-se as correlações  $\rho = 0,2$  e  $\rho = 0,5$  obtendo os seguintes resultados.

Analisando a Figura 4.1.5(a) observou-se que a estatística  $T^2$  de Hotelling praticamente apresentou uma igualdade nas probabilidades de erro



tipo I para todos os tamanhos de amostras em ambas as abordagens. Este fato indicou que o uso do bootstrap poderia ser dispensado devido ao esforço computacional requerido.

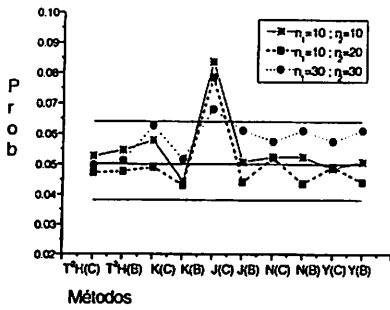
Novamente o método de Johansen (1980) na abordagem convencional apresentou problemas relacionados ao controle da taxa de erro tipo I para todos os tamanhos de amostras. Este efeito no método de Johansen (1980) também foi verificado por Christensen & Rencher (1997), mas somente presente nas situações de amostras de tamanhos iguais e tamanhos diferentes. Uma alternativa eficiente visando reparar esta deficiência apresentada pelo método de Johansen (1980) foi dada por meio da aplicação do bootstrap, independentemente das correlações. Este fato foi comprovado com sucesso mediante aos resultados observados na Figura 4.1.5. Os métodos de Nel & Van Der Merwe (1986) e de Yao (1965) de bootstrap apresentaram diferenças em relação à abordagem convencional. Estes métodos na abordagem convencional apresentaram taxas empíricas iguais para as amostras  $n_1=10$ ,  $n_2=10$  e  $n_1=10$  e  $n_2=20$ , ao passo que na abordagem bootstrap esta igualdade mostrou-se ausente. No caso das amostras  $n_1=30$ ,  $n_2=30$  em ambas as abordagens estes dois métodos apresentaram taxas empíricas de erro tipo I superiores ao valor nominal, confirmando os resultados obtidos por Christensen & Rencher (1997).

Mantendo a mesma heterogeneidade, porém aumentando a correlação para  $\rho = 0,5$  conforme resultados apresentados na Figura 4.1.5(b), constatou-se que, de um modo geral, todos os métodos foram melhorados com o aumento da correlação, de tal forma que a mudança na correlação refletiu uma redução nas taxas empíricas, aproximando-se do valor nominal de modo que a aplicação dos métodos para as amostras  $n_1=10$  e  $n_2=20$  foram mais influenciadas, enquanto a aplicação destes métodos para outros tamanhos amostrais foi menos sensibilizada por este efeito. Sobre o método de Johansen (1980), na abordagem convencional este foi o único que apresentou as mesmas características

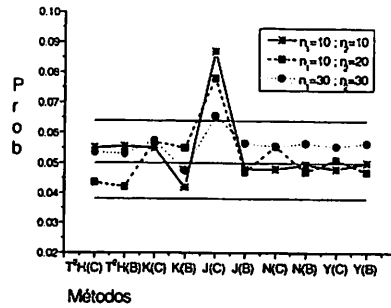
mencionadas quando fixado  $\rho = 0,2$  (Figura 4.1.5(a)), continuando a não controlar o erro tipo I.

Visando analisar o impacto do aumento da heterogeneidade na matriz de covariância diante das correlações, os resultados da Figura 4.1.5(c) permitiram verificar que na situação de fraca correlação ( $\rho = 0,2$ ) e na condição de alta heterogeneidade ( $\delta=8$ ), este aumento refletiu um efeito negativo na estatística  $T^2$  de Hotelling especificamente para as amostras de tamanhos diferentes em ambas as abordagens, o que ocasionou em um não controle da taxa de erro tipo I (Figura 4.1.5(a) e (c)). Em relação aos demais métodos, verificou-se que as diferenças nas probabilidades entre as amostras foram menores e praticamente centralizadas em torno do valor nominal, com exceção do método de Johansen (1980) convencional, que continuou a não controlar a taxa de erro tipo I para todos os tamanhos amostrais, mostrando-se invariante ao aumento do grau de heterogeneidade. Na abordagem bootstrap, o método de Kim (1992) foi o mais prejudicado, passando a controlar a taxa de erro tipo I para todos os tamanhos amostrais de forma conservativa. Aumentando a correlação para  $\rho = 0,5$  (Figura 4.1.5(d)), a única diferença foi verificada no método de Johansen (1980) via convencional, mais especificamente para as amostras  $n_1=30$ ,  $n_2=30$ , pois com estes tamanhos amostrais este método corrigiu a deficiência no controle da taxa de erro tipo I mencionada anteriormente ( $\rho=0,2$ ). Em relação à estatística  $T^2$  de Hotelling, os métodos de Johansen (1980) convencional para outros tamanhos amostrais e Kim (1992) na versão bootstrap apresentaram o mesmo comportamento retratados pela Figura 4.1.5(c).

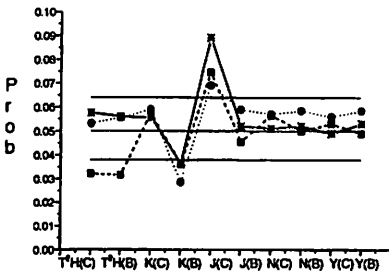
Em relação aos demais métodos, quando comparado com os resultados apresentados na Figura 4.1.5(c), notou-se apenas uma inversão na disposição dos valores em torno do valor nominal. Esta característica ocorreu em ambas as abordagens.



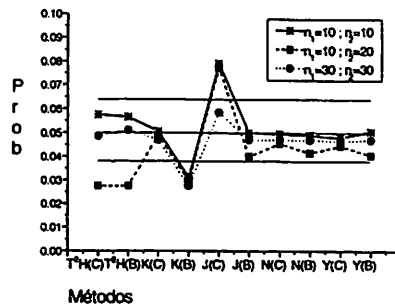
a -  $\delta=2$  e  $\rho=0,2$



b -  $\delta=2$  e  $\rho=0,5$



c -  $\delta=8$  e  $\rho=0,2$



d -  $\delta=8$  e  $\rho=0,5$

**FIGURA 4.1.5 – Taxa de Erro tipo I, fixado o valor nominal em 5% e  $p=5$  comparando graus de heterogeneidade.**

Os resultados da Figura 4.1.6 retrataram uma situação atípica em relação aos resultados apresentados nas figuras anteriores, não apenas pelo tamanho amostral estudado, mas também pela influência das matrizes de covariâncias. Neste aspecto, as amostras  $n_1=10$ ,  $n_2=15$  (Figura 4.1.6(a)) indicam, respectivamente, que a menor amostra está associada à maior matriz de covariância, ao passo que, quando invertidas as amostras, ou seja,  $n_1=15$ ,  $n_2=10$  (Figura 4.1.6(b)), caracterizou-se que a maior amostra passou a se associar à “maior” matriz de covariância. A mensuração da magnitude das matrizes de covariância é dada em função das respectivas variâncias generalizadas.

Justifica-se a escolha destes tamanhos amostrais pela implementação do bootstrap nos métodos, sendo comum o problema de singularidade quando submetido a amostras pequenas, pois a probabilidade de serem amostrados os mesmos elementos novamente é relativamente alta. Com esta preocupação, estes tamanhos amostrais foram definidos de tal forma a evitar a singularidade entre as matrizes envolvidas no processo do bootstrap para obtenção das probabilidades. Um outro aspecto foi de se determinar um tamanho de amostra que não fosse muito grande para que o problema de heterogeneidade de variâncias fosse mascarado pelas propriedades assintóticas do teste. Desta forma, os principais resultados obtidos estão a seguir.

Particularizando o valor nominal em 5%, número de variáveis  $p = 5$  e grau máximo de heterogeneidade ( $\delta = 8$ ), verificou-se que as estatísticas  $T^2$  de Hotelling em ambas as abordagens foram afetadas pelo grau de correlação de tal forma que o controle da taxa de erro tipo I se deu apenas para fraca correlação, ou seja,  $\rho = 0,2$  (Figura 4.1.6(a)). Com a inversão dos tamanhos amostrais, isto é,  $n_1=15$  e  $n_2=10$  (Figura 4.1.6(b)), este efeito causado pelo aumento no grau de correlação em relação à estatística  $T^2$  de Hotelling em ambas as abordagens desapareceu. Este teste controlou a taxa de erro tipo I para as correlações  $\rho = 0,2$  e  $0,9$ . Este fato concordou com os resultados obtidos por Fouladi & Yockey (2002). Estes autores afirmaram que a estatística  $T^2$  de Hotelling, para as amostras de tamanhos diferentes é robusta desde que a magnitude entre as correlações entre as variáveis seja relativamente grande.

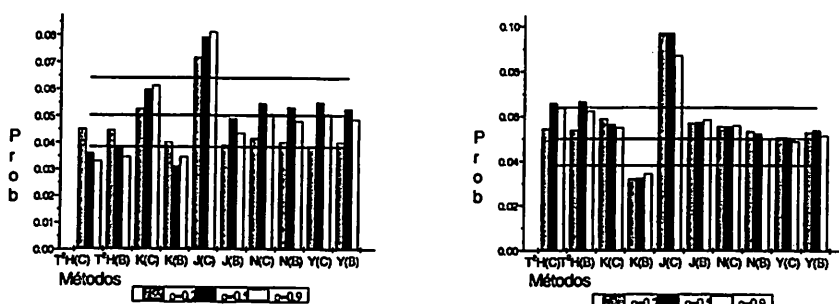
Em relação aos métodos, na situação  $n_1=10$  e  $n_2=15$  (Figura 4.1.6(a)), a influência do aumento das correlações revelou uma igualdade entre os métodos na abordagem convencional, de tal forma que os métodos de Kim (1992) e Johansen (1980) proporcionaram probabilidades praticamente iguais, refletindo o mesmo comportamento. As performances destes dois métodos, quando invertidas as amostras, foram bem diferentes e sensíveis ao aumento da

correlação; esta sensibilidade é facilmente detectada principalmente no método de Kim (1992) convencional, de modo que as probabilidades foram reduzidas gradativamente com o aumento da correlação (Figura 4.1.6(b)), ao passo que, na situação anterior (Figura 4.1.6(a)), esta ocorrência mostrou-se exatamente inversa; na abordagem bootstrap praticamente foi observado a mesma ocorrência.

Analisando o método de Johansen (1980) convencional, é notável a sensibilidade comparando as taxas empíricas, pois os resultados da Figura 4.1.6(a) claramente permitem a percepção de um comportamento crescente à medida que a correlação aumenta. Analisando os resultados da Figura 4.1.6(b), notou-se que esta tendência é inexistente. Na abordagem bootstrap, a sensibilidade relacionada à inversão das matrizes e amostras é mais bem explicada pelas probabilidades demonstradas pela Figura 4.1.6(a), de modo que, para as correlações extremas, o controle da taxa de erro tipo I é feito de forma conservativa, sendo que, na situação posterior (Figura 4.1.6(b)), praticamente todas as correlações apresentaram a mesma probabilidade de erro tipo I.

Analisando os métodos Nel & Van der Merwe (1986) e Yao (1965) na abordagem convencional, conforme resultados apresentados na Figura 4.1.6(a), verifica-se que estes métodos controlaram a taxa de erro tipo I de forma conservativa para a correlação  $\rho = 0,2$ , concordando com os resultados de Fouladi & Yockey (2002) que afirmaram uma tendência conservativa apresentada especificamente para o método de Nel & Van der Merwe (1986). No caso de uma correlação intermediária, ou seja,  $\rho = 0,5$  suas taxas empíricas foram máximas e superiores ao valor nominal (Figura 4.1.6(a)). Invertendo as amostras (Figura 4.1.6(b)), estes métodos apresentaram comportamento semelhante aos resultados da estatística  $T^2$  de Hotelling, com poucas diferenças nos valores das probabilidades, quando submetidas às diferentes correlações. A aplicação dos métodos na versão bootstrap devido à semelhança entre as

probabilidades em relação aos métodos convencionais é recomendada apenas para o método de Johansen (1980), sendo que, na situação retratada pela Figura 4.1.6(a) para as correlações inferiores a  $\rho=0,5$  o controle da taxa de erro tipo I é feito de forma conservativa. Porém, na situação denotada pela Figura 4.1.6(b) as taxa empíricas foram próximas ao valor nominal. Ainda nesta situação, este fato se deu para os demais métodos, além de ter apresentado taxas empíricas similares aos métodos convencionais, exceto o método de Kim (1992).



a -  $n_1=10, n_2=15$

b -  $n_1=15, n_2=10$

FIGURA 4.1.6 – Taxa de erro tipo I, fixado o valor nominal em 5%,  $\delta=8$  e  $p=5$  comparando graus de correlação.

## 4.2 - Poder dos Métodos

Uma outra forma de comparação e avaliação dos métodos é dada por meio da estimação do poder inerente a cada método, associado ao controle da taxa de erro tipo I. Sendo assim, compararam-se os poderes dos testes em situações em que os vetores de médias populacionais diferiram com um determinado número de erros padrão definidos pela constante  $k$ . Desta forma, foi estudado o poder mediante diferentes correlações, inicialmente fixado o número de variáveis ( $p=2$ ), valor nominal em 5% e  $k=1$ , restringindo o estudo apenas

para as amostras de tamanhos iguais (Figura 4.2.1). Assim, os resultados apresentados foram discutidos doravante.

Na situação de homogeneidade, para amostras  $n_1=10$ ,  $n_2=10$  (Figura 4.2.1(a)), os resultados permitiram que se inferisse que o método de Johansen (1980) apresentou-se mais poderoso em todas as estruturas de correlações; entretanto, o uso deste método não foi recomendado devido ao não controle da taxa de erro tipo I verificado anteriormente (Figuras 4.1.1(a) e (b)) representando, respectivamente, as correlações  $\rho = 0,2$  e  $\rho = 0,9$ . Devido a estes valores extremos na correlação, ficou subentendido o mesmo comportamento apresentado por este método, quando fixada uma correlação intermediária, ou seja,  $\rho=0,5$ . Esta suposição foi ratificada com os resultados observados por Christensen & Rencher (1997), referentes ao controle da taxa de erro tipo I, em particular a este método. Com esta característica, evidenciada em todas as correlações, excluiu-se a análise do método de Johansen (1980) convencional, de forma que o método de Kim (1992), nesta mesma abordagem, foi o mais adequado, devido à superioridade do poder, embora esta superioridade tenha sido relativamente pequena em relação à dos demais métodos. Todavia, por controlar a taxa de erro tipo I com probabilidades aproximadas ao valor nominal, comprovado pelos resultados das Figuras 4.1.1(a) e (b) e estendido para a correlação  $\rho = 0,5$  recomendou-se o uso deste método. Segundo Christensen & Rencher (1997), este método também foi recomendado pelos mesmos fatores, ou seja, alto poder associado ao controle da taxa de erro tipo I, com probabilidade acima do valor nominal.

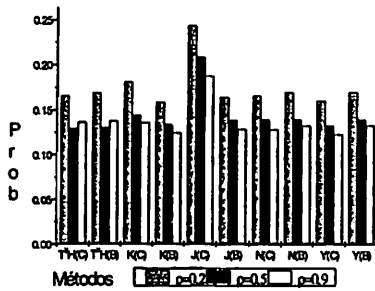
Em relação à abordagem bootstrap, foi verificado que para a recomendação de sua aplicação não foram obtidos valores de poder que pudessem justificar esforço computacional requerido. Este fato é simplesmente comprovado pela igualdade ou proximidade dos valores particularizados a cada método, quando comparados com a abordagem convencional. É importante

ressaltar que os métodos de Nel & Van der Merwe (1986) e Yao (1965) apresentaram valores semelhantes no poder e no controle da taxa de erro tipo I, não havendo, portanto, restrições quanto ao seu uso. Entretanto, Christensen & Rencher (1997) compararam o método de Yao (1965) com o método de Kim (1992) e constataram que este método apresentou maior poder em relação ao método de Yao (1965), quando a amostra menor é associada à maior matriz de covariância.

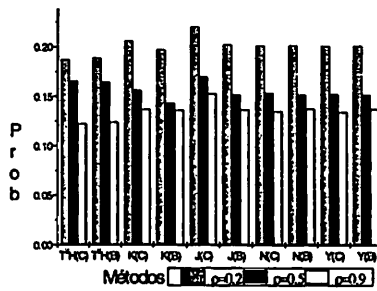
Na situação de homogeneidade, aumentando os tamanhos das amostras para  $n_1=30$ ,  $n_2=30$ , conforme resultados da Figura 4.2.1(b), a indicação de um método adequado foi feita com base no controle da taxa de erro tipo I e alto poder. Com essas suposições, o método de Johansen (1980) convencional praticamente tem o mesmo poder do método de Kim (1992), pois este método, para todas as correlações, controlou a taxa de erro tipo I. Ainda nesta situação, foi observado que nos demais métodos o aumento do tamanho da amostra, especificamente para a correlação  $\rho=0,2$  implicou de um modo geral em um aumento no poder, sendo mais evidente no método de Nel & Van der Merwe (1986). Para as demais correlações, este efeito foi amenizado.

Na situação de heterogeneidade, a principal característica observada foi detectada sobre as amostras  $n_1=10$ ,  $n_2=10$  (Figura 4.2.1(c)), a qual retrata a inexistência de um comportamento padrão nos valores referentes ao poder com o aumento da correlação. Este efeito foi verificado em alguns métodos na abordagem convencional, sendo mais acentuado na abordagem bootstrap, com exceção do método de Kim (1992). A presença deste efeito, quando comparada com a situação de homogeneidade, foi contrastante, pois nesta situação o aumento da correlação implicou, de um modo geral, na redução do poder. Entretanto, a presença deste efeito não influenciou a identificação dos métodos. Baseando-se nos critérios de controle da taxa de erro tipo I e no poder para estas amostras, os métodos de Yao (1965) e Nel & Van der Merwe (1986) foram

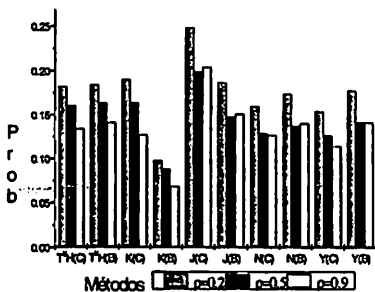
adequados, concordando com as observações feitas por De La Rey e Nel's (1993). Com o aumento das amostras para  $n_1=30$  e  $n_2=30$  (Figura 4.2.1(d)) notou-se um melhoramento em relação ao padrão do comportamento das correlações, mostrando-se semelhante à situação de homogeneidade. Entretanto, não foram indicados os mesmos métodos em virtude do controle da taxa de erro tipo I ter sido diferente em ambas as correlações, conforme resultados apresentados nas Figuras 4.1.1(c) e (d), Sendo assim, a recomendação foi dada em função de ter apresentado simultaneamente nas correlações  $\rho=0,2$  e  $0,9$  o controle da taxa de erro tipo I e alto poder. Assim, a indicação de os métodos de Yao (1965) e de Nel & Van Der Merwe (1986) pode ser feita.



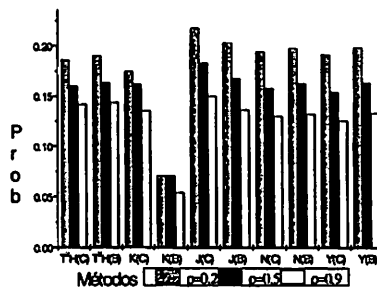
a -  $\delta=1$  e  $n_1=10$  e  $n_2=10$



b -  $\delta=1$  e  $n_1=30$  e  $n_2=30$



c -  $\delta=8$  e  $n_1=10$  e  $n_2=10$



d -  $\delta=8$  e  $n_1=30$  e  $n_2=30$

**FIGURA 4.2.1** – Poder do teste para amostras de tamanhos iguais, fixado o valor nominal 5%,  $p=2$ ,  $k=1$  comparando graus de correlação e sobre situação de homogeneidade e máxima heterogeneidade.

Os resultados apresentados na Figura 4.2.2 basicamente são complementações ao estudo de poder apresentado anteriormente, diferenciando apenas no valor nominal, sendo este fixado em 1%. Desta forma, estudou-se o impacto da redução do valor nominal sobre as mesmas condições amostrais e grau de heterogeneidade relacionados na Figura 4.2.2. Sendo assim, foram obtidos os seguintes resultados.

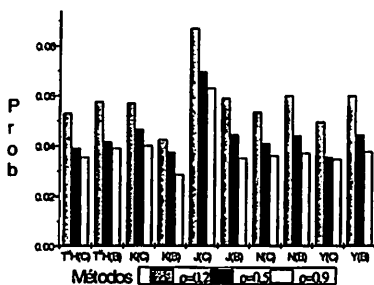
Na situação de homogeneidade, com a redução do valor nominal para as amostras  $n_1=10$ ,  $n_2=10$  conforme resultados apresentados pela Figura 4.2.2(a), novamente o método de Johansen (1980) convencional não foi recomendado pelas mesmas razões mencionadas anteriormente (Figura 4.2.1(a)). Sendo assim, indicou-se o método de Kim (1992) convencional como um método apropriado a este tamanho amostral. Este fato foi comprovado conforme resultados apresentados na Figura 4.2.1(a). Entretanto, reduzindo o valor nominal para 1% (Figura 4.2.2(a)), verificou-se que os métodos Nel & Van der Merwe (1986) e Yao (1965) na abordagem bootstrap apresentaram valores referente ao poder para todas as correlações superiores à abordagem convencional. Este fato é justificado pelas taxas empíricas de erro tipo I, referente a estes métodos via bootstrap, serem superiores à aplicação destes métodos na abordagem convencional (Figura 4.1.2(a)). No caso do método de Johansen (1980), por controlar a taxa de erro tipo I na abordagem bootstrap, mostrou-se vantajoso, embora o poder deste método nesta abordagem seja inferior, quando comparado com a abordagem convencional, como é esperado. Em relação ao método de Kim (1992), quando comparado às versões convencional e bootstrap, necessariamente a recomendação foi justificada também pelo controle da taxa de erro tipo I (Figura 4.1.2(a)). Desta forma, em ambas as abordagens este método controlou a taxa de erro tipo I, porém com taxas empíricas aproximadas ao valor nominal, ocasionando poucas diferenças no poder deste método, quando comparado entre as duas abordagens. Aumentando, o tamanho das amostras para

$n_1=30$  e  $n_2=30$  (Figura 4.2.2 (b)), o poder deste método referente às duas abordagens mostrou-se sensível no aumento da correlação. Entretanto, considerando a máxima correlação admitida ( $\rho=0,9$ ) percebeu-se uma igualdade no poder em ambas as abordagens referentes ao método de Kim (1992); sendo assim, optou-se por recomendar o uso deste método via convencional pela simplicidade na aplicação, quando comparado ao esforço computacional requerido no bootstrap.

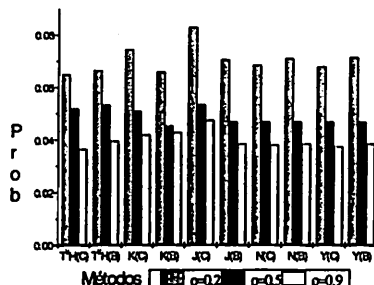
Em relação aos demais métodos, confirmaram-se os mesmos resultados apresentados pela figura 4.2.2(a), diferenciando apenas na correlação  $\rho=0,5$  em que os métodos Nel & Van Der Merwe (1986) e Yao (1965) apresentaram valores de poder equivalentes em ambas abordagens.

Na situação de heterogeneidade (Figura 4.1.2(c)), na avaliação do controle da taxa de erro tipo I, para as amostras  $n_1=10$ ,  $n_2=10$ , fixada a correlação  $\rho=0,2$  de um modo geral em ambas as abordagens os métodos não controlaram a taxa de erro tipo I, exceto o método de Yao (1965) convencional e o de Kim (1992) bootstrap. Desta forma, especificamente para esta correlação, considerando o esforço computacional requerido no bootstrap, avaliado juntamente com o respectivo poder destes métodos, recomendou-se o método de Yao (1965) via convencional como o mais apropriado para esta situação amostral. Aumentando a correlação para  $\rho=0,9$  (Figura 4.1.2(d)), este método passou a controlar a taxa de erro tipo I com probabilidade aproximada ao valor nominal e alto poder, sendo novamente indicado. Aumentando o tamanho das amostras para  $n_1=30$ ,  $n_2=30$ , os problemas relativos ao controle da taxa de erro tipo I continuaram sendo apresentados pelos métodos de Johansen (1980) e Kim (1992) portanto, o estudo do poder foi avaliado em função dos demais métodos, de tal forma que o método de Nel & Van der Merwe (1986) convencional foi o mais apropriado, não descartando a viabilidade da aplicação do método de Yao (1965), devido à aproximação dos resultados entre estes

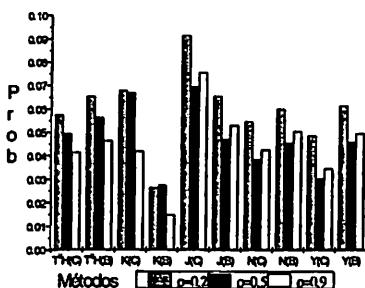
métodos. Na abordagem bootstrap, o principal requisito para a não recomendação foi justamente o esforço computacional exigido e ressaltado em resultados anteriores, além dos valores referentes ao poder, terem sido, de um modo geral, aproximados àqueles da abordagem convencional.



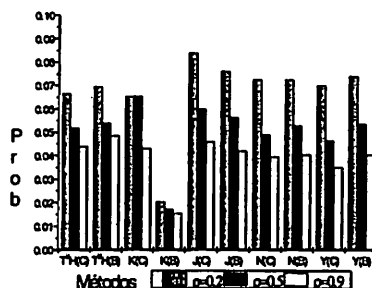
a -  $\delta=1$  e  $n_1=10$  e  $n_2=10$



b -  $\delta=1$  e  $n_1=30$  e  $n_2=30$



c -  $\delta=8$  e  $n_1=10$  e  $n_2=10$



d -  $\delta=8$  e  $n_1=30$  e  $n_2=30$

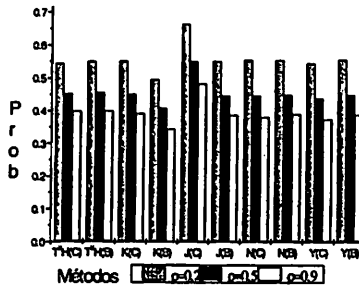
**FIGURA 4.2.2** – Poder do teste para amostras de tamanhos iguais, fixado o valor nominal 1%,  $p=2$ ,  $k=1$  comparando graus de correlação e sobre situação de homogeneidade e máxima heterogeneidade.

Avaliando o aumento do número de variáveis como fator de influência no poder, os resultados apresentados pelos métodos com os vetores de médias populacionais, diferindo em dois erros padrão ( $k=2$ ) e com valor nominal fixado em 5%, com base na Figura 4.2.3, foram obtidos os seguintes resultados.

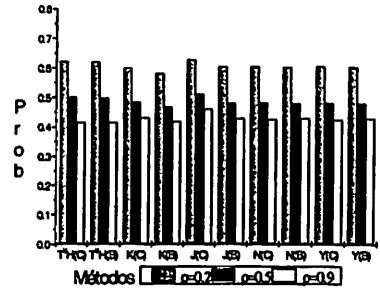
Na situação de homogeneidade, especificamente para as amostras  $n_1=10$ ,  $n_2=10$  o método de Johansen (1980) na abordagem convencional apresentou maior poder, porém não foi recomendado devido ao não controle da taxa de erro tipo I. Sendo assim, comparou-se a influência do aumento do número de variáveis, em conjunto com o aumento da diferença dos erros padrão dado por  $k=2$ . Desta forma, verificou-se que este aumento refletiu um efeito positivo para o método de Kim (1992), de modo que seu poder foi semelhante em relação aos demais métodos. A determinação da utilização da abordagem bootstrap considerou as mesmas características mencionadas anteriormente. Aumentando o tamanho da amostra para  $n_1=30$ ,  $n_2=30$  (Figura 4.2.3(b)), novamente o método de Johansen (1980) convencional foi recomendado, embora os demais métodos tenham mostrado valor de poder aproximado. Em relação à abordagem bootstrap, foram feitas as mesmas ressalvas mencionadas quando  $p$  foi fixado em 2.

Na situação de heterogeneidade, conforme resultados apresentados na Figura 4.2.3(c), as conclusões feitas na situação de homogeneidade continuaram válidas, especificamente para as amostras  $n_1=10$  e  $n_2=10$  e correlação  $\rho=0,2$  (Figura 4.2.3(a)). O aumento do grau de heterogeneidade não influenciou no poder dos métodos, conforme resultados apresentados na Figura 4.2.3(c) porém, quando comparado sobre o número de variáveis, observou-se que na situação de heterogeneidade e com  $p=5$  os métodos apresentaram uma tendência em reduzir o valor referente ao poder, de acordo com o aumento das correlações anteriormente fixado  $p=2$ , este efeito mostrou-se ausente (Figura 4.2.2(c)). Com o aumento das amostras para  $n_1=30$ ,  $n_2=30$  (Figura 4.2.3(d)), não foi detectada nenhuma discrepância em relação ao poder dos métodos, sendo que o método de Johansen (1980) convencional, em particular para este tamanho amostral, controlou a taxa de erro tipo I e apresentou alto poder, sendo que os demais métodos foram também eficientes, com diferença mínima de poder quando

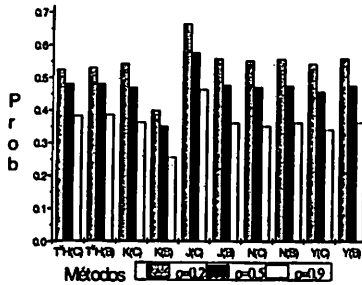
comparados com este método. Em virtude do que foi mencionado, para as amostras maiores e de tamanhos iguais recomendou-se o método de Johansen (1980); para as demais amostras, os métodos apropriados foram Nel & Van der Merwe (1986) e Yao (1965), evidentemente aplicados à abordagem convencional.



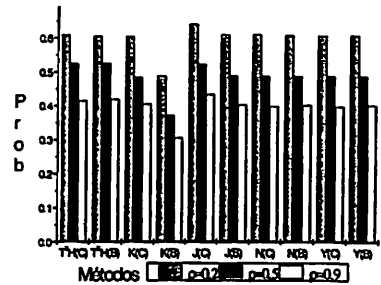
a -  $\delta=1$  e  $n_1=10$  e  $n_2=10$



b -  $\delta=1$  e  $n_1=30$  e  $n_2=30$



c -  $\delta=8$  e  $n_1=10$  e  $n_2=10$



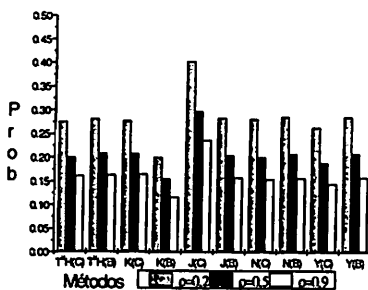
d -  $\delta=8$  e  $n_1=30$  e  $n_2=30$

**FIGURA 4.2.3** – Poder do teste, para amostras de tamanhos iguais fixado o valor nominal 5%,  $p=5$  e  $k=2$ , comparando graus de correlação sobre a situação de homogeneidade e máxima heterogeneidade.

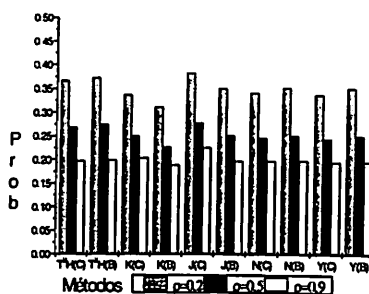
Seguindo as mesmas definições referentes à Figura 4.2.3, os resultados apresentados na Figura 4.2.4 descrevem o comportamento dos valores referentes ao poder inerente a cada método em diferentes graus de correlação sobre a redução do valor nominal para 1%.

Na situação de homogeneidade, conforme resultados apresentados na Figura 4.2.4(a), em relação às amostras  $n_1=10$  e  $n_2=10$  recomendou-se o método de Nel & Van der Merwe (1986). Esta indicação foi dada pelo valor do poder ser coerente com a probabilidade apresentada no controle da taxa de erro tipo I, nas correlações extremas, isto é,  $\rho=0,2$  e  $0,9$  (Figura 4.1.3(a) e (b)). Comparando, com os resultados apresentados pela Figura 4.2.3(a), verificou-se que a redução do valor nominal não influenciou no comportamento dos valores referentes ao poder. Aumentando o tamanho das amostras para  $n_1=30$  e  $n_2=30$  (Figura 4.2.4(b)), os valores referentes ao poder apresentados pelos métodos foram praticamente semelhantes, principalmente na situação de fraca correlação ( $\rho=0,2$ ). Esta semelhança foi justificada pelas probabilidades do controle da taxa de erro tipo I terem sido praticamente iguais, apresentando-se com maior exatidão sobre os métodos de Nel & Van der Merwe (1986) e Yao (1965) em ambas as abordagens. Com o aumento do grau de correlação para  $\rho=0,9$  (Figura 4.2.4(b)), este efeito também se mostrou presente.

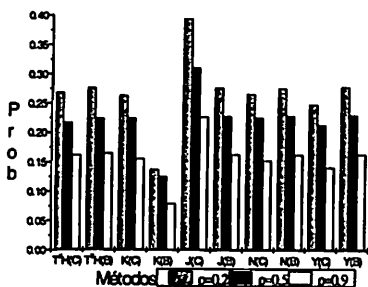
Na situação de heterogeneidade (Figura 4.2.4(c)), as características apresentadas pelos métodos, para as amostras  $n_1=10$  e  $n_2=10$ , foram praticamente as mesmas, quando comparadas com a situação de homogeneidade, sendo que a única diferença detectada foi um pequeno aumento nos valores. Aumentando o tamanho amostral para  $n_1=30$ ,  $n_2=30$  (Figura 4.2.4(d)), notou-se o mesmo fato. Assim, concluiu que a redução no valor nominal não ocasionou diferenças relacionadas ao poder de modo a alterar o comportamento e características apresentadas pelos métodos.



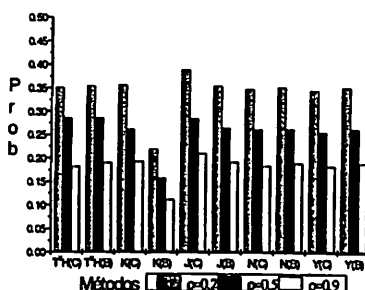
a -  $\delta=1$  e  $n_1=10$  e  $n_2=10$



b -  $\delta=1$  e  $n_1=30$  e  $n_2=30$



c -  $\delta=8$  e  $n_1=10$  e  $n_2=10$



d -  $\delta=8$  e  $n_1=30$  e  $n_2=30$

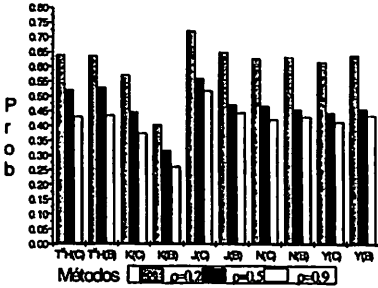
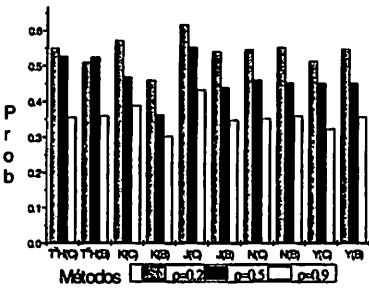
**FIGURA 4.2.4** – Poder do teste, para amostras de tamanhos iguais fixado o valor nominal 1%,  $p=5$  e  $k=3$  comparando graus de correlação sobre a situação de homogeneidade e máxima heterogeneidade.

Avaliando o poder dos métodos (Figura 4.2.5), cujos, vetores de médias populacionais diferiram em 3 erros padrões, na situação amostral mais crítica, isto é,  $n_1=10$ ,  $n_2=15$  e  $n_1=15$ ,  $n_2=10$ , em conjunto com o controle da taxa de erro tipo I (Figura 4.1.6(a) e (b)), foram obtidos os seguintes resultados.

Na primeira situação amostral,  $n_1=10$ ,  $n_2=15$  (Figura 4.2.5(a)), em todas as correlações o método de Johansen (1980) convencional mostrou-se poderoso, entretanto não controlou a taxa de erro tipo I (Figura 4.1.6(a)), sendo considerado um método não adequado, o que novamente confirmou os

resultados verificados por Algina et. al. (1991) e relatados por Christensen e Rencher (1997) sobre a ineficiência deste método no controle da taxa de erro tipo I. No caso dos demais métodos, praticamente todos eles apresentaram o mesmo poder para todas as correlações, entretanto os métodos de Kim (1992) e Yao (1965) mostraram diferenças relacionadas ao poder, de modo que, quando a menor amostra associada à maior matriz de covariância ( $n_1 < n_2$ ), (Figura 4.2.5(a)) o método de Kim (1992) apresentou maior poder. Esta ocorrência foi confirmada por resultados apresentados por Christensen & Rencher (1997) e evidenciada pelo aumento do grau de correlação. Na abordagem bootstrap, este fato foi contrariado, de forma que na mesma situação amostral o método de Yao (1965) foi mais poderoso.

Em relação aos demais métodos, de um modo geral o poder não foi afetado quanto à ordem das amostras, o que contrariou os resultados apresentados por Kim (1992) e relatados por Christensen & Rencher (1997), segundo os quais o poder de alguns métodos não é afetado em relação à ordem das amostras que estão associados as maiores covariâncias.



a -  $n_1=10$  e  $n_2=15$

b -  $n_1=15$  e  $n_2=10$

FIGURA 4.2.5 – Poder do teste, fixado o valor nominal 5%,  $k=3$ ,  $p=5$  e  $\delta=8$  comparando graus de correlação com diferentes tamanhos amostrais.

Em virtude dos resultados mencionados, ficou evidenciado que os métodos, de um modo geral, apresentaram comportamento relacionado ao poder influenciados pelo grau de correlação. Objetivando investigar a influência do número de variáveis, grau de heterogeneidade, valor nominal e erro padrão, os resultados das Tabelas 4.2.1 e 4.2.2 são referentes ao poder dos métodos, mediante a correlação intermediária ( $\rho=0,5$ ). Este valor da correlação foi adotado para se ter uma ampla cobertura de valores de correlações do seu espaço paramétrico para que nenhum tipo de viés pudesse aparecer nos resultados. Os valores de poder descritos na tabela foram provenientes dos métodos com a abordagem convencional, sendo que a justificativa é dada em função de que a abordagem bootstrap, conforme resultados anteriores, apresentou poucas diferenças quando comparada à abordagem convencional. Em alguns métodos, os valores de poder foram iguais, bem como o controle da taxa de erro tipo I foi semelhante, além da inviabilidade da aplicação do bootstrap pelo esforço computacional requerido, o que consolidou ainda mais o uso dos métodos na abordagem convencional. Desta forma, os resultados obtidos e demonstrados pelas tabelas 4.2.1 e 4.2.2 foram discutidos doravante.

A princípio, foi analisado o comportamento dos métodos em função do aumento da heterogeneidade e do valor do erro padrão, definido por  $k$ . Inicialmente, estudou-se este efeito fixando o número de variáveis em  $p=2$  (Tabela 4.2.1). Assim, foi verificado que independentemente do valor nominal e com as médias populacionais diferindo em um erro padrão ( $k=1$ ), a estatística  $T^2$  de Hotelling e o método de Yao (1965) apresentaram uma redução no poder para todos os tamanhos amostrais. Aumentando a diferença entre as médias populacionais para três erros padrão ( $K=3$ ), observou-se a mesma ocorrência no método de Kim (1992). Os demais métodos não apresentaram um comportamento padrão referente ao poder, pois a redução foi dada de forma particular ao valor nominal e tamanho amostral. Ampliando este estudo para o

caso pentavariado ( $p=5$ ) (Tabela 4.2.2), verificou-se que o aumento da heterogeneidade, com as médias populacionais diferenciadas em um erro padrão ( $k=1$ ) e fixado o valor nominal em 1%. Os métodos de Nel & Van der Merwe (1986) e Johansen (1980) tiveram redução no poder para todas as amostras.

Enfatizando o aumento do número de variáveis em conjunto com o aumento da heterogeneidade, cujo vetor de médias diferiu em um erro padrão ( $k=1$ ), observou-se que quando fixado o valor nominal em 1%, os métodos de Johansen (1980), Yao (1965) e Nel & Van der Merwe (1986) apresentaram valores referente ao poder, inflacionado em todos os tamanhos amostrais. Aumentando o valor nominal para 5%, este efeito novamente foi confirmado apenas no método de Yao (1965). Entretanto, com o aumento da diferença entre as médias populacionais com 3 erros padrão, em relação a estes métodos, este efeito mostrou-se ausente, sendo constatado apenas no método de Kim (1992) para ambos os valores nominais.

Em virtude do que foi mencionado, constatou-se o método de Yao (1965) apresentou poder inflacionado devido ao efeito do aumento do número de variáveis em conjunto com o aumento do grau de heterogeneidade independente do valor nominal. Em relação aos métodos de Johansen (1980) e Nel & Van der Merwe (1986), verificou-se que com a redução do valor nominal este efeito na inflação do poder é corrigido. O método de Kim (1992) apresentou sensibilizado com este efeito mediante ao aumento na diferença de erros padrão, independentemente do valor nominal.

**TABELA 4.2.1** – Comparação do poder para todas as amostras em função do aumento do número de variáveis em conjunto com o aumento da heterogeneidade de covariâncias e da diferença entre os vetores médias em função do erro padrão para um valor nominal de 5% e 1% considerando a correlação  $\rho=0,5$  e  $p=2$ .

Métodos	Amostras	$\delta=1$				$\delta=8$			
		k=1		k=3		k=1		k=3	
		1%	5%	1%	5%	1%	5%	1%	5%
T <sup>2</sup> de Hotelling	$n_1=10$ e $n_2=10$	0,0530	0,1660	0,5705	0,8280	0,0495	0,1610	0,5755	0,8270
	$n_1=10$ e $n_2=20$	0,0465	0,1585	0,6270	0,8480	0,0315	0,1090	0,4950	0,7710
	$n_1=30$ e $n_2=30$	0,0650	0,1875	0,6585	0,8690	0,0520	0,1610	0,6880	0,8760
Kim	$n_1=10$ e $n_2=10$	0,0465	0,1445	0,5435	0,8055	0,0670	0,1640	0,4530	0,7095
	$n_1=10$ e $n_2=20$	0,0540	0,1600	0,5520	0,8140	0,0835	0,1900	0,5105	0,7440
	$n_1=30$ e $n_2=30$	0,0510	0,1565	0,6670	0,8625	0,0655	0,1625	0,4450	0,7215
Johansen	$n_1=10$ e $n_2=10$	0,0695	0,2090	0,6725	0,8830	0,0695	0,1990	0,6875	0,8905
	$n_1=10$ e $n_2=20$	0,0800	0,2275	0,6820	0,8885	0,0640	0,2005	0,7160	0,8985
	$n_1=30$ e $n_2=30$	0,0535	0,1705	0,7030	0,8790	0,0600	0,1835	0,7260	0,8895
Yao	$n_1=10$ e $n_2=10$	0,0355	0,1330	0,5300	0,8070	0,0305	0,1265	0,5380	0,8040
	$n_1=10$ e $n_2=20$	0,0495	0,1540	0,5575	0,8145	0,0400	0,1420	0,6495	0,8590
	$n_1=30$ e $n_2=30$	0,0470	0,1530	0,6665	0,8655	0,0465	0,1545	0,6790	0,8655
Nel & Van der Merwe	$n_1=10$ e $n_2=10$	0,0410	0,1395	0,5475	0,8165	0,0385	0,1285	0,5765	0,8145
	$n_1=10$ e $n_2=20$	0,0555	0,1675	0,6110	0,8415	0,0470	0,1500	0,6640	0,8650
	$n_1=30$ e $n_2=30$	0,0470	0,1540	0,6670	0,8655	0,0490	0,1580	0,6875	0,8695

**TABELA 4.2.2** – Comparação do poder para todas as amostras em função do aumento do número de variáveis em conjunto com o aumento da heterogeneidade de covariâncias e da diferença entre os vetores médias em função do erro padrão para um valor nominal de 5% e 1% considerando a correlação  $\rho=0,5$  e  $\rho=5$ .

		$\delta=1$				$\delta=8$			
		k=1		k=3		k=1		k=3	
		1%	5%	1%	5%	1%	5%	1%	5%
T <sup>2</sup> de Hotelling	$n_1=10$ e $n_2=10$	0,0460	0,1440	0,5440	0,8145	0,0405	0,1415	0,5500	0,8095
	$n_1=10$ e $n_2=20$	0,0470	0,1773	0,6200	0,8435	0,0275	0,1145	0,5200	0,7910
	$n_1=30$ e $n_2=30$	0,0500	0,1520	0,6745	0,8675	0,0545	0,1580	0,6800	0,8765
Kim	$n_1=10$ e $n_2=10$	0,0460	0,1455	0,5365	0,8060	0,0470	0,1430	0,5370	0,7905
	$n_1=10$ e $n_2=20$	0,0510	0,1530	0,5660	0,8050	0,0465	0,1460	0,5840	0,8125
	$n_1=30$ e $n_2=30$	0,0590	0,1795	0,6850	0,8765	0,0600	0,1660	0,6745	0,8645
Johansen	$n_1=10$ e $n_2=10$	0,0820	0,2100	0,6810	0,8795	0,0735	0,2135	0,6560	0,8810
	$n_1=10$ e $n_2=20$	0,0730	0,2030	0,6820	0,8670	0,0695	0,1880	0,6690	0,8780
	$n_1=30$ e $n_2=30$	0,0640	0,1965	0,7195	0,8925	0,0680	0,1820	0,7115	0,8830
Yao	$n_1=10$ e $n_2=10$	0,0410	0,1385	0,5295	0,8055	0,0410	0,1360	0,5185	0,7885
	$n_1=10$ e $n_2=20$	0,0470	0,1440	0,5605	0,7965	0,0485	0,1470	0,5685	0,8170
	$n_1=30$ e $n_2=30$	0,0545	0,1740	0,6830	0,8750	0,0555	0,1620	0,6715	0,8625
Nel & Van der Merwe	$n_1=10$ e $n_2=10$	0,0430	0,1415	0,5500	0,8110	0,0440	0,1400	0,5420	0,8010
	$n_1=10$ e $n_2=20$	0,0540	0,1545	0,6210	0,8175	0,0515	0,1495	0,5990	0,8415
	$n_1=30$ e $n_2=30$	0,0545	0,1750	0,6860	0,8750	0,0565	0,1630	0,6770	0,8635

## 5-CONCLUSÕES

(a) O aumento da correlação refletiu uma redução no poder, tanto na situação de homogeneidade como na heterogeneidade.

(b) A eficiência da abordagem bootstrap em todas as situações o método de Johansen (1980), quando submetido à abordagem bootstrap,

(c) A abordagem bootstrap foi eficiente apenas no controle da taxa de erro tipo I para o método de Johansen (1980).

(d) O método de Nel & Van Der Merwe (1986) na abordagem convencional foi o mais apropriado para amostras de tamanhos diferente.

(e) O método de Johansen (1980) na abordagem convencional foi recomendado para amostras maiores.

(f) No caso das amostras menores, a recomendação de um método ficou condicionada ao valor nominal, grau de correlação.

## 6-REFERÊNCIAS BIBLIOGRÁFICAS

ALGINA, J.; OSHIMA, T. C.; TANG, K. L. Robustness of Yao's, Jame's and Johansen's Tests Under Variance-Covariance Heteroscedasticity and Nonnormality. **Journal of Educational Statistics**, washington, v. 16, n. 2, p. 125-139, summer 1991.

ALGINA, J.; TANG, K. L. Type I Error Rates for Yao's and Jame's Tests of Equality of Mean Vectors Under Variance-Covariance Heteroedasticity. **Journal of Educational Statistics**, Washington, v. 13, n. 2, p. 281-290, Fall 1988.

BEHRENS, B. V. Ein Betraig zur Fehlerberechnung bei wenige Beobachtungen. **Landwirtschaftliches Jahrbuch**, Bonn, v. 68, p. 807-837, 1929.

BENNETT, B. M. Note on a Solution of the Generalized Behrens-Fisher Problem. **Annals of the Institute of Statistical Mathematics**, Tokyo, v. 2, n. 1, p. 87-90, 1951.

BORGES, L. C.; FERREIRA, D. F. Comparações de duas aproximações do teste t com variâncias heterogêneas através de simulação. **Ciência e Agrotecnologia**, Lavras, v. 23, n. 2, p. 390-403, abr. /jun. 1999.

BRYCE, G. R.; MAYNES, D. D. **Generation of Multivariate Data Sets**. Technical report no. SD-015-R. Department of Statistics, Brigham Young University, 1979.

BRUNNER, E.; DETTE, H.; MUNK, A. Box type Approximations in Nonparametric Factorial Designs. **Journal American Statistical Association**, Alexandria, v. 92, n. 440, p. 1494-1502, Dec. 1997.

BRUNNER, E.; ULLRICH, M.; PURI, L. M. The Multivariate nonparametric, Behrens-Fisher problem. **Journal of Statistical Planning and Inference**, Amsterdam, v. 108, n. 1/2, p. 37-53, Nov. 2002.

CHRISTENSEN, W. F; RENCHER, A. C. A Comparison of type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem. **Communications in Statistics Simulation and Computation**, New York, v. 26, n. 4, p. 1251-1273, 1997.

COCHRAN, W. G.; COX, G. M. **Experimental designs** 2. ed. Singapore: John Wiley & Sons, 1957. 611 p.

DE LA REY, N.; NEL, D. G. A comparison of the Significance Levels and Power Functions of Several Solutions to the Multivariate Behrens-Fisher Problem. **South African Statistical Journal**, Copetown, v. 27, p. 129-148, 1993.

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

FISHER, R. A. The fiducial argument in statistical inference. **The Annals of Eugenics**, London, v. 11, p. 141, 1935.

FOULADI, R. T.; YOCKEY, R. D. Type I Error Control of Two-Group Multivariate Tests on Means Under Conditions of Heterogeneous Correlation Structure and Varied Multivariate Distributions. **Communications in Statistics**, New York, v. 31, n. 3, p. 375-400, 2002.

GHOSH M.; KIM Y. H. The Behrens-Fisher problem revisited: a Bayes-frequentist synthesis. **The Canadian Journal of Statistics**, Ottawa, v. 29, n. 1, p. 5, Mar. 2001.

HOLLOWAY, L. N.; DUNN, O. J. The Robustness of Hotelling's  $T^2$ . **Journal of the American Statistical Association**, Alexandria, v. 62, n. 317, p. 124-136, Jan. 1967.

ITO, K.; SCHULL, W. J. On the Robustness of the  $T^2$  Test in Multivariate Analysis of Variance When Variance-Covariance Matrices are Not Equal. **Biometrika**, London, v. 51, n. 1/2, p. 71-82, 1964.

JAMES, G. S. Tests of Linear Hypotheses in Univariate and Multivariate Analysis When the Ratios of the Population Variances are Unknown. **Biometrika**, London, v. 41, n. 1/2, p. 19-43, June 1954.

JOHANSEN, S. The Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression. **Biometrika**, London, v. 67, n. 1, p. 85-92, Apr. 1980.

KIM, S. A Practical Solution to the Multivariate Behrens-Fisher Problem. **Biometrika**, London, v. 79, n. 1, p. 171-176, Mar. 1992.

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

NEL, D. G.; VAN DER MERWE, C. A. A solution to the multivariate Behrens-Fisher Problem. **Communications in Statistics Theory and Methods**, New York, v. 15, n. 1, p. 3719-3735, 1986.

SAS INSTITUTE SAS / STAT. **SAS/IML Software: usage and reference**, version 6. Cary, 1990. 501p.

SATTHELTHWAITE, F. E. An approximate distribution of estimates of variance components. **Biometric Bulletin**, London, v. 60, n. 2, p. 110-114, 1946.

SUBRAHMANIAM, K.; SUBRAHMANIAM, K On the Multivariate Behrens-Fisher Problem, **Biometrika**, New York, v. 60, n. 1, p. 107-111, Apr. 1973.

SUBRAHMANIAM, K.; SUBRAHMANIAM, K On the Confidence Region Comparison of Some Solutions for the Multivariate Behrens-Fisher Problem. **Communications in Statistics**, New York, v. 4, n. 1, p. 57-67, 1975.

YAO, Y. An Approximate Degree of Freedom Solution to the Multivariate Behrens-Fisher Problem. **Biometrika**, London, v. 52, n. 1/2, p. 139-147, June 1965.

## ANEXOS

**ANEXO ..... PÁGINA**

**Anexo 1 - Programa utilizado para gerar quantis para o teste de normalidade Q-Qplot – caso univariado. .... 126**

**Anexo 2 - Programa utilizado para gerar quantis para o teste de normalidade Q-Qplot – caso multivariado..... 127**

## Anexo 1

```
proc iml;
    nsim=45000; n=30;
    q=j(n,1,0);
do ii=1 to n;
    p=(ii-0.5)/n;
    q[ii]=probit(p);
end;
    x=j(n,1,0);
    r=j(nsim,1,0);
do j=1 to nsim;
do ii=1 to n;
    x[ii]=rannor(100);
end;
/* *** ordenando o vetor x*** */
    b=x; x[rank(x),]=b;
/* *** fim ordenamento*** */
    sxy=0;sx=0;sy=0;sy2=0;sx2=0;
do ii=1 to n;
    sxy=sxy+x[ii]*q[ii];
    sx=sx+x[ii];
    sy=sy+q[ii];
    sy2=sy2+q[ii]*q[ii];
    sx2=sx2+x[ii]*x[ii];
end;
    rq=(sxy-sx*sy/n)/((sx2-sx*sx/n)*(sy2-sy*sy/n)**0.5;
    r[i]=rq;
end;
    create rqrq from r;
    append from r;
quit;

proc univariate data=rqrq;
run;quit;
```

## Anexo 2

```
proc iml;
    nsim=25000; n=15;
    seed=0;
    p=2;
    sigma=i(p);
    mu =j(p,1,0);
    m=repeat(mu`,n,1);
    g=root(sigma);
    qu=j(n,1,0);

do ii=1 to n;
    pr=(ii-0.5)/n;
    qu[ii]=cinv(pr,p);
end;

    x=j(n,1,0);
    r=j(nsim,1,0);
    dj=j(n,1,0);

do i=1 to nsim;
    z=normal(repeat(seed,n,p));
    x=z*g+m;
    q=i(n) - (1/n)*j(n,n,1);
    S=(1/(n-1))*x`*q*x;
    xb=x`*(1/n)*j(n,1,1);
    si=inv(s);
    xii=j(1,p,0);
do ii=1 to n;
    xii=x[ii,];
    xii=xii-t(xb);
    dj[ii]=xii*si*xii`;
end;

/* *** ordenando o vetor dj *** */
    b=dj;  dj[rank(dj),]=b;
/* *** fim ordenamento *** */
```

```

sxy=0;sx=0;sy=0;sy2=0;sx2=0;

do ii=1 to n;
  sxy=sxy+(dj[ii]**0.5)*(qu[ii]**0.5);
  sx=sx+(dj[ii]**0.5);
  sy=sy+(qu[ii]**0.5);
  sy2=sy2+(qu[ii]**0.5)*(qu[ii]**0.5);
  sx2=sx2+(dj[ii]**0.5)*(dj[ii]**0.5);
end;
rq=(sxy-sx*sy/n)/((sx2-sx*sx/n)*(sy2-sy*sy/n))**0.5;
r[i]=rq;
end;
create rqarq from r;
append from r;
quit;

proc univariate data=rqarq;
  /* histogram /beta; */
run;quit;

/* *** quantis da beta *** */
/*
data teste;
x=betainv(0.01,38.6647,1.800);
run;quit;
proc print data=teste;
run;quit;
*/

```