UNIVERSIDADE FEDERAL DE LAVRAS

RENATA APARECIDA CINTRA

TESTES DE NORMALIDADE MULTIVARIADA BASEADOS EM AMOSTRAS BETAS INDEPENDENTES

LAVRAS - MG

2016

RENATA APARECIDA CINTRA

TESTES DE NORMALIDADE MULTIVARIADA BASEADOS EM AMOSTRAS BETAS INDEPENDENTES

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

Orientador
Dr. Daniel Furtado Ferreira

LAVRAS - MG
2016

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Cintra, Renata Aparecida.
Testes de normalidade multivariada baseados em amostras betas independentes / Renata Aparecida Cintra. – Lavras : UFLA, 2016.
102 p.

Dissertação (mestrado acadêmico) – Universidade Federal de Lavras, 2016.
Orientador(a): Daniel Furtado Ferreira.
Bibliografia.

1. Distribuição beta 2. Bootstrap paramétrico 3. Kolmogorov-Smirnov . I. Universidade Federal de Lavras. II. Título.

RENATA APARECIDA CINTRA

TESTES DE NORMALIDADE MULTIVARIADA BASEADOS EM AMOSTRAS BETAS INDEPENDENTES

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

APROVADA em 08 de abril de 2016.

Dra. Carla Regina Guimarães Brighenti	UFSJ
Dr. Renato Ribeiro de Lima	UFLA

Dr. Daniel Furtado Ferreira
Orientador

LAVRAS - MG
2016

AGRADECIMENTOS

Ao meu orientador, Daniel Furtado Ferreira, pelo incentivo, pela atenção, pela paciência, pela inspiração, pela simpatia, sempre presentes em suas orientações e ensinamentos, aos quais serei eternamente grata.

Aos meus pais, pelo apoio incondicional, por me ensinarem todos os valores e, sobretudo, por acreditarem em mim e me amarem em todos os momentos.

Ao meu noivo, Marcel, pelo amor, compreensão, carinho, paciência e pela ajuda concedida para a realização do trabalho. Sobretudo, por me incentivar a seguir na busca pelo conhecimento. E aos meus sogros e cunhados, pelo respeito, pela gentileza, pelo apoio e pela hospitalidade em todos os momentos.

Aos meus irmãos, Renan e Rosana, pelo carinho, companheirismo, apoio e alegria que sempre me fortalecem.

Aos membros da banca examinadora, Renato Ribeiro de Lima e Carla Regina Guimarães Brighenti, por aceitarem contribuir com este trabalho, por meio de suas importantes avaliações e sugestões. Também, aos membros da banca de qualificação, Izabela Regina Cardoso de Oliveira e Patrícia Siqueira, por toda a contribuição.

Aos funcionários do DEX que estão sempre dispostos a ajudar e colaborar para que o desenvolvimento acadêmico ocorra, em especial, à Nádia e à Josi.

Aos meus colegas do mestrado e demais estudantes do DEX, pela convivência, que possibilitou troca de experiências favoráveis ao crescimento social e profissional, pela amizade e pela diversão. Especialmente, aos amigos: Guilherme, Janaína, Kelly, Lílian, Sidcleide, Henrique e Carlos, pela proximidade, por todo o apoio e pelas risadas compartilhadas.

A todas às pessoas que de algum modo contribuíram para que eu pudesse concluir o mestrado.

À CAPES, pelo apoio financeiro e por contribuir com o desenvolvimento da pesquisa e do conhecimento.

RESUMO

Em inferência, a verificação de normalidade multivariada é muito importante, pois muitos métodos são baseados em hipóteses de que os dados provêm de uma distribuição normal multivariada. Gnanadesikan e Kettenring (1972) provaram que é possível obter amostras betas a partir de amostras normais utilizando uma transformação na distância quadrática de Mahalanobis. Verificando a aderência da amostra obtida pela transformação com a distribuição beta, teria-se um indício de que a amostra original seria proveniente de uma distribuição normal multivariada. Embrechts, Frey e McNeil (2005) propuseram um teste baseado na estrutura do teste de Kolmogorov-Smirnov utilizando esses conceitos. Contudo, esse teste é afetado pela dependência amostral presente na distância quadrática utilizada. Liang, Pan e Yang (2004) apresentaram uma forma de obter amostras betas univariadas, cada uma independente e identicamente distribuída, por meio de transformações em uma amostra normal p-variada. Este trabalho teve como principal objetivo propor dois testes para normalidade multivariada com base no que foi proposto por Liang, Pan e Yang (2004): um teste de aderência a partir do teste de Kolmogorov-Smirnov e um teste intensivo fundamentado em *bootstrap* paramétrico. O programa R (R CORE TEAM, 2015) foi utilizado para implementar os algoritmos dos dois testes de normalidade multivariada propostos e para realizar simulações Monte Carlo com o propósito de estimar as taxas de erro tipo I e o poder dos testes. Realizou-se comparações dos testes propostos com o teste de normalidade multivariada que foi apresentado por Embrechts, Frey e McNeil (2005) e com o teste de Shapiro-Wilk de normalidade multivariada proposto por Royston (1983). Embora os testes propostos tenham obtido um bom controle das taxas de erro tipo *I*, o uso desses testes não foi recomendado devido ao fraco desempenho de poder apresentado por eles.

Palavras-chave: Distribuição beta, *Bootstrap* paramétrico, Kolmogorov-Smirnov, Programa R.

ABSTRACT

In inference, multivariate normality tests are very important, since many methods are based on assumptions that the data come from a multivariate normal distribution. Gnanadesikan and Kettenring (1972) proven that it is possible to obtain beta samples from normal samples using a transformation in the Mahalanobis quadratic distance. Checking the fit of the sample obtained by transformation to the beta distribution is an indication that the original sample is from a multivariate normal distribution. Embrechts, Frey and McNeil (2005) proposed a test based on Kolmogorov-Smirnov test using these concepts. However, this test is influenced by the sample dependence present in the quadratic distance. Liang, Pan and Yang (2004) presented a way to obtain univariate beta samples, each independent and identically distributed, through transformations in a p-variate normal sample. This work aimed to propose two tests for multivariate normality: a goodness-of-fit test based on Kolmogorov-Smirnov test and an intensive test based on parametric *bootstrap*. The R program (R CORE TEAM, 2015) was used to implement the algorithms of both proposed tests and Monte Carlo simulations were used in order to estimate type *I* error rates and the power of the tests. Comparisons were conducted between the proposed tests and the multivariate normality test that was presented by Embrechts, Frey and McNeil (2005) and the Shapiro-Wilk multivariate normality test proposed by Royston (1983). Although the proposed tests have obtained good control of the type *I* error rates, the use of these tests was not recommended due to the poor performance of power presented by them.

Keywords: Beta distribution, Parametric bootstrap, Kolmogorov-Smirnov, R Software.

Lista de Tabelas

SUMÁRIO

1 INTRODUÇÃO

A verificação de normalidade multivariada é muito importante para garantir que a inferência esteja correta em uma grande parte dos métodos que existem, pois esses, diversas vezes, são baseados em hipóteses de que os dados provém de uma distribuição normal multivariada. Nesse sentido, os testes de normalidade multivariada se mostram extremamente relevantes. Contudo, os testes existentes quase sempre possuem limitações. Mecklin e Mundfrom (2005) mostraram que não há um único teste que seja o mais poderoso em todas as situações.

O uso de um teste de hipótese formal é indispensável para presumir a normalidade. No entanto, comumente, faz-se uso de testes visuais para sugerir a normalidade no caso univariado, os quais constituem uma ferramenta de apoio. Embora esse tipo de teste esteja sujeito à interpretação subjetiva e, por isso, imprecisa, algumas adaptações foram feitas ao longo dos anos, possibilitando o uso desses testes visuais para detectar possíveis desvios de normalidade no caso multivariado.

Gnanadesikan e Kettenring (1972) provaram que é possível obter amostras betas a partir de amostras normais utilizando uma transformação na distância quadrática de Mahalanobis. Partindo desse resultado, construíram gráficos quantil-quantil (Q-Q *plots*) utilizando as amostras obtidas a partir das transformações nos dados iniciais, para verificar se essas provinham da distribuição beta. Em caso de as amostras serem provenientes de uma beta, poder-se-ia indicar que a amostra multivariada seria de uma distribuição normal multivariada.

Embrechts, Frey e McNeil (2005) propuseram um teste baseado na estrutura do teste de Kolmogorov-Smirnov utilizando esses conceitos. Contudo, além de serem formas de verificação puramente visuais, os Q-Q *plots* propostos são afetados pela dependência amostral presente na distância quadrática utilizada, que afeta também o teste de Embrechts, Frey e McNeil (2005).

Uma forma de contornar o problema da dependência amostral nesses Q-Q *plots* e no teste de Embrechts, Frey e McNeil (2005) foi apresentada por Liang, Pan e Yang (2004). Tratam-se de transformações em uma amostra p-variada, que sob a hipótese nula de normalidade multivariada, conduzem a $p - 1$ amostras betas univariadas. Cada uma dessas amostras betas é independente e identicamente distribuída. Utilizando uma dessas amostras betas, Liang, Pan e Yang (2004) apresentaram Q-Q plots como ferramentas complementares para detectar um possível desvio de normalidade em análise de dados de alta dimensão.

Tendo em vista o problema da dependência amostral presente nos testes construídos com base na obtenção de amostras betas, este trabalho tem como principal objetivo propor dois testes para normalidade multivariada baseados na obtenção de amostras betas que não sejam afetados pela dependência amostral. A partir do que foi proposto por Liang, Pan e Yang (2004), os objetivos específicos deste trabalho são descritos a seguir:

1. propor um teste de normalidade multivariada baseado na obtenção de amostras betas utilizando o teste de aderência de Kolmogorov-Smirnov (TKS);
2. propor um teste *bootstrap* paramétrico para a normalidade multivariada baseado na obtenção de amostras betas (TBPb);
3. realizar a avaliação do desempenho dos testes, utilizando simulação Monte Carlo, para verificar o controle da taxa de erro tipo I e o poder de cada um dos testes propostos;
4. comparar os testes propostos com o teste de normalidade multivariada de Royston (1983) (TSW) e com o teste de normalidade multivariada proposto por Embrechts, Frey e McNeil (2005) (TEM).

2 REFERENCIAL TEÓRICO

2.1 Distâncias quadráticas

Antes de se abordar as distâncias quadráticas, será apresentada a definição de forma quadrática, pois toda distância quadrática é uma forma quadrática.

Dada uma matriz $\boldsymbol{A}_{n\times n}$, sua forma quadrática é definida por (FERREIRA, 2011):

$$Q(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i=1}^{n-1}\sum_{k=i+1}^{n} a_{ik}x_i x_k = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}x_i x_k, \qquad (1)$$

em que $\boldsymbol{x} \neq \boldsymbol{0}$ é um vetor definido em $\mathbb{R}^n$.

O nome forma quadrática se deve ao fato de a expressão (1) apresentar somente termos quadráticos ou de duplos produtos dos elementos do vetor $\boldsymbol{x}$.

No contexto de inferência, quando se lida com análise multivariada é muito comum a utilização do conceito de distância quadrática. Dados os vetores $\boldsymbol{x}$ e $\boldsymbol{y} \in \mathbb{R}^p$ e a matriz $\boldsymbol{\psi}$ positiva definida, denominada métrica, pode-se expressar a distância quadrática entre os pontos $\boldsymbol{x}$ e $\boldsymbol{y}$ como em Ferreira (2011):

$$d^2(\boldsymbol{x},\boldsymbol{y}) = ||\boldsymbol{x}-\boldsymbol{y}||^2{}_{\boldsymbol{\psi}} = (\boldsymbol{x}-\boldsymbol{y})^\top \boldsymbol{\psi}(\boldsymbol{x}-\boldsymbol{y}). \qquad (2)$$

A distância euclidiana quadrática é obtida quando a métrica em (2) é $\boldsymbol{\psi} = \boldsymbol{I}$. Nesse caso, a expressão (2) se reduz a:

$$d^2(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})^\top(\boldsymbol{x}-\boldsymbol{y}) = \sum_{k=1}^{p}(x_k - y_k)^2.$$

Se a métrica em (2) for definida como $\boldsymbol{\psi} = \boldsymbol{S^{-1}}$, em que $\boldsymbol{S}$ é a matriz de covariâncias amostral, uma vez que os vetores $\boldsymbol{x}$ e $\boldsymbol{y}$ sejam realizações das variáveis aleatórias $\boldsymbol{X}$ e $\boldsymbol{Y}$, tem-se a distância generalizada de Mahalanobis, dada por:

$$d^2(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})^\top\boldsymbol{S^{-1}}(\boldsymbol{x}-\boldsymbol{y}).$$

Caso a métrica em (2) seja $\boldsymbol{\psi} = \boldsymbol{D^{-1}} = diag(1/S_{kk}), k = 1,2,...,p$, então é definida a distância euclidiana padronizada quadrática ou, como também é conhecida, a distância estatística de Karl-Pearson, como segue:

$$d^2(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})^\top\boldsymbol{D^{-1}}(\boldsymbol{x}-\boldsymbol{y}) = \sum_{k=1}^{p}\frac{(x_k - y_k)^2}{S_{kk}}.$$

2.2 Distribuição normal multivariada

Sejam p variáveis normais independentes tais como $X_1, X_2, ..., X_p$ cujas funções densidades se definem por:

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_{ii}}}exp\left\{-\frac{(x_i-\mu_i)^2}{2\sigma_{ii}}\right\},$$

em que μ_i é a média e σ_{ii} é a variância para a i-ésima variável.

Ao se agruparem as p variáveis independentes no vetor $\boldsymbol{X}$, é possível escrever a função densidade conjunta como:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} f(x_i) = (2\pi)^{-\frac{p}{2}} \left(\prod_{i=1}^{p} \sigma_{ii} \right)^{-\frac{1}{2}} exp \left\{ - \sum_{i=1}^{p} \frac{(x_i - \mu_i)^2}{2\sigma_{ii}} \right\}. \tag{3}$$

Rearranjando a expressão (3) com matrizes e vetores e utilizando uma matriz de covariâncias positiva definida mais geral, $\boldsymbol{\Sigma}$, é possível obter a função densidade da distribuição normal multivariada, que tem a seguinte forma (FERREIRA, 2011):

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp \left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}, \tag{4}$$

em que $\boldsymbol{\mu}$ é o vetor de médias e $\boldsymbol{\Sigma}$ é a matriz de covariâncias, expressos, matricialmente, da seguinte forma:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}.$$

2.3 Distribuições de formas quadráticas

O teorema enunciado a seguir pode ser encontrado em Ferreira (2011):

Teorema 1. *Seja $\boldsymbol{X} \in \mathbb{R}^p$ um vetor que segue distribuição normal multivariada com função densidade $f_{\boldsymbol{X}}(\boldsymbol{x})$ dada em (4), com média $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, então*

$$Q = \boldsymbol{X}^\top \boldsymbol{A} \boldsymbol{X}$$

tem distribuição qui-quadrado não-central com $\nu = k$ graus de liberdade e parâmetro de não-centralidade $\sigma^2 = \boldsymbol{\mu}^\top \boldsymbol{A} \boldsymbol{\mu}$, se e somente se, $\boldsymbol{\Sigma A}$ for uma matriz idempotente de posto k, ou seja, se $\boldsymbol{\Sigma A \Sigma A} = \boldsymbol{\Sigma A}$, ou ainda, $\boldsymbol{A \Sigma A} = \boldsymbol{A}$.

Do Teorema 1, provêm os seguintes resultados:

1. $\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X} \sim \chi^2_p(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})$;
2. $(\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_p$; e
3. $(\boldsymbol{X} - \boldsymbol{\alpha})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\alpha}) \sim \chi^2_p(\boldsymbol{\mu} - \boldsymbol{\alpha})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\alpha})$ para qualquer vetor $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_p]^\top \in \mathbb{R}^p$.

Ferreira (2011) afirma que, dada uma amostra aleatória $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$, do vetor aleatório $\boldsymbol{X}$ que segue uma distribuição normal multivariada $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a distribuição assintótica de

$$D_j^2 = (\boldsymbol{X}_j - \bar{\boldsymbol{X}}_.)^\top \boldsymbol{S}^{\mathbf{-1}} (\boldsymbol{X}_j - \bar{\boldsymbol{X}}_.), \qquad j = 1,\ 2,\ \ldots,\ n$$

é uma qui-quadrado com $\nu = p$ graus de liberdade. Nesse caso, se n não for grande, a aproximação não é boa. Então, pode-se utilizar a seguinte transformação, de acordo com Gnanadesikan e Kettenring (1972):

$$B_j = \frac{n D_j^2}{(n-1)^2},$$

que segue distribuição beta exata com parâmetros $\alpha = p/2$ e $\beta = (n - p - 1)/2$.

2.4 Erros tipo I e II

Nos testes de hipóteses, duas suposições são discutidas: a hipótese que está sendo testada, chamada de hipótese nula, denotada por H_0, e a outra, chamada de hipótese alternativa, denotada por H_1. A ideia é que se a hipótese nula for falsa, a alternativa é verdadeira e vice-versa (MOOD; GRAYBILL; BOES, 2001).

Um teste de hipótese clássico é baseado em uma determinada estatística de teste T. A distribuição nula é a distribuição da estatística T, quando H_0 é verdadeira. Para testar alguma hipótese nula, deve-se dividir a área abaixo da curva da função densidade da estatística T em duas regiões. Se o valor observado de T pertence a uma região R, deve-se rejeitar a hipótese nula, se T pertence à região complementar R^c, não se deve rejeitar. A região R é chamada de região crítica do teste, cujo tamanho é α e R^c, de região de aceitação, de tamanho $(1-\alpha)$ (KENDALL; STUART, 1961; MOOD; GRAYBILL; BOES, 2001; BARON, 2013). Por conseguinte, pode-se ter dois tipos de erros quando se realiza um teste de hipótese:

- Erro tipo I: decisão errada ao se rejeitar a hipótese nula quando esta é verdadeira. A probabilidade do erro tipo I é igual ao tamanho da região crítica usada, definida pelo nível de significância α assumido previamente, ou seja, $P[T \in R \mid H_0] = \alpha$.
- Erro tipo II: decisão errada ao não se rejeitar a hipótese nula quando esta é falsa. A probabilidade do erro tipo II é uma função da hipótese alternativa considerada e é, usualmente, denotada por β. Assim, $P[T \in R^c \mid H_1] = \beta$.

A probabilidade complementar do erro tipo II, ou seja, $P[T \in R \mid H_1] = 1-\beta$, representa a decisão correta de se rejeitar a hipótese nula quando esta é falsa e é denominada poder do teste de hipótese.

Para uma amostra de tamanho conhecido e fixado, o valor α é inversamente proporcional ao valor β, portanto, a única maneira de se fazer com que as probabilidades do erro tipo I e tipo II se reduzam, simultaneamente, é aumentar o tamanho da amostra (FERREIRA, 2009). A escolha da probabilidade do erro tipo I, que é um valor α pré-definido, deve ocorrer de modo que a probabilidade do erro tipo II, β, seja minimizada (KENDALL; STUART, 1961). De acordo com Baron (2013), geralmente, β pode ser minimizado utilizando valores entre α 1% e 10%.

2.5 Simulação Monte Carlo

Alguns dos precursores do método Monte Carlo foram Metropolis e Ulam (1949), que o apresentaram como uma abordagem estatística para o estudo de equações diferenciais e integrais, as quais ocorrem em vários ramos das ciências naturais. Trata-se de um método computacionalmente intensivo que, como os autores evidenciaram, requer computadores mais modernos, pois um desempenho elevado da máquina implica em redução do tempo de execução das simulações.

Em Estatística, geralmente, define-se o método Monte Carlo como a representação da solução de um problema, como, por exemplo, obtenção de um parâmetro de uma população hipotética. Para isso, utiliza-se uma sequência de números aleatórios para construir uma amostra da população, da qual se pode obter as estimativas do parâmetro (HALTON, 2006).

Segundo Gentle (2003), o uso de números aleatórios na Estatística tem se expandido para além da amostragem aleatória ou aleatorização de tratamentos em unidades experimentais. O autor relata que os usos mais comuns são em processos estocásticos, expressões matemáticas analiticamente intratáveis ou obtenção de amostra aleatória de uma população obtida por meio de reamostragem de dada amostra dessa população. Nessas três áreas, os métodos Monte Carlo são, respectivamente, conhecidos como “simulação estocástica”, “Monte Carlo” e “reamostragem”. Exemplos de obras que abrangem essas aplicações, na devida ordem, são Ripley (1987), Robert e Casella (2009) e Gentle (2003).

Ferreira (2013) descreve o uso dos métodos Monte Carlo para a realização de testes de normalidade, bem como para a validação desses testes. Com o uso de simulações de realizações de variáveis aleatórias que seguem distribuições normais e, também, distribuições não normais é possível avaliar o desempenho do poder e o controle das taxas de erro tipo I de testes estatísticos recém-propostos. Outras utilidades do método descritas pelo autor são: avaliar propriedades de um estimador ou de um método de estimação intervalar, determinar tamanhos de amostras e em Cadeias de Markov via Métodos Monte Carlo na inferência bayesiana.

2.6 *Bootstrap*

O *bootstrap* é um método computacional que foi introduzido por Efron (1979), mas passou a ser amplamente utilizado a partir da década de 90 por requerer o uso de computadores mais potentes. Para entender a ideia do método, suponha que se deseje obter informações sobre a distribuição amostral de uma estatística. Então, considere a possibilidade de realizar amostras repetidas um grande número de vezes, do mesmo tamanho, da população de interesse. Dessa forma, poder-se-ia ter alguma noção da distribuição amostral com base nos valores provenientes das amostras repetidas. Contudo, na prática, esse procedimento seria caro e se tornaria inviável, pois o objetivo dos estudos amostrais é conseguir informações sem custo elevado e em um tempo hábil (SINGH; XIE, 2008).

Finalmente, como Singh e Xie (2008) mencionam, o auxílio computacional viabilizou esse processo, consolidando a ideia que levou à criação do *bootstrap*: utilizar os dados da amostra que se tem para gerar um grande número de amostras (com reposição), conhecidas como amostras *bootstrap*, permitindo obter informações sobre a distribuição da amostra da população de interesse.

Certamente, esse é apenas o princípio básico do método *bootstrap*, que possui inúmeras aplicações, tais como em: estimativas de erro padrão, modelos de regressão, estimativas de viés, intervalos de confiança, testes de permutação, testes de hipóteses, estimativas de predição de erro, calibração, aproximação da verossimilhança, entre outros (EFRON; TIBSHIRANI, 1993).

De modo geral, pode-se definir, como em Ferreira (2013), que o método *bootstrap* se constitui na reamostragem com reposição da amostra original, fornecendo uma nova amostra de tamanho igual. Com a repetição do processo um grande número de vezes, computa-se o valor da estatística ou da quantidade pivotal de interesse cuja distribuição é chamada de distribuição de *bootstrap* e é usada para a realização da inferência. Pode-se classificar o método em dois tipos com base na reamostragem a ser realizada: *bootstrap* paramétrico e não paramétrico.

2.6.1 *Bootstrap* não paramétrico

Davison e Hinkley (1997) relataram que, nos problemas não paramétricos mais simples, obtêm-se, literalmente, amostras a partir da amostra, o que parece não fazer muito sentido a princípio. No entanto, é razoável se for considerado que a amostra contém toda a informação sobre a população. Ao fazer isso, a única suposição que se assume é a de que os valores da amostra são permutáveis, o que, de acordo com Ferreira (2013, p. 583), significa que "a distribuição de probabilidade de quaisquer k observações consecutivas $(k = 1,\ 2,\ \ldots,\ n)$ não se altera quando a ordem das observações é trocada por meio de alguma permutação", sendo a forma mais fraca de se supor que as variáveis aleatórias são independentes e identicamente distribuídas. Segundo esse autor, há a substituição da distribuição desconhecida da população pela distribuição empírica, que é obtida.

Davison e Hinkley (1997) mencionaram também que uma variedade de problemas estatísticos pode ser resolvida com base no método *bootstrap* não paramétrico. Alguns exemplos de aplicação citados por eles são: checar a adequação de medidas de incerteza padrão, relaxar hipóteses e fornecer rápidas soluções aproximadas, dentre outros.

Dada uma amostra aleatória $X_1,\ X_2,\ \ldots,\ X_n$ de uma população com distribuição desconhecida em que o interesse é o parâmetro θ, realiza-se um grande número de reamostragens com reposição de mesmo tamanho n da amostra original, obtendo B amostras *bootstrap*, cada uma como $X_1^*,\ X_2^*,\ \ldots,\ X_n^*$. Então, estima-se o parâmetro θ por uma função $\hat{\theta}$ de cada amostra obtida, de modo que o i-ésimo estimador *bootstrap* seja dado por $G(X_{i1}^*,\ X_{i2}^*,\ \ldots,\ X_{in}^*) = \hat{\theta}_i^*$ em que X_{ij}^* é o j-ésimo elemento da i-ésima amostra *bootstrap*, com $i = 1,\ 2,\ \ldots,\ B$ e $j = 1,\ 2,\ \ldots,\ n$. Então, obtêm-se B estimativas que constituem uma representação de uma amostra da distribuição do estimador, chamada distribuição *bootstrap*, que permite que a inferência sobre θ possa ser realizada (FERREIRA, 2013).

Assim, nos casos em que a distribuição da variável aleatória é desconhecida ou não pode ser tratada de forma analítica, atribui-se a cada elemento da amostra a mesma probabilidade $1/n$ de ser selecionado e se usa a distribuição de probabilidade empírica para a reamostragem (GIVENS; HOETING, 2012).

2.6.2 *Bootstrap* paramétrico

O *bootstrap* paramétrico se aplica aos casos em que se assume uma distribuição de probabilidade para a variável aleatória amostrada, porém, os parâmetros são desconhecidos. A distribuição de amostragem do estimador também é desconhecida ou não pode ser obtida analiticamente. Desse modo, utilizam-se as estimativas dos parâmetros de interesse, obtidas a partir da amostra que se tem, como parâmetros da distribuição de probabilidade assumida. Geram-se, então, novas amostras. Em cada uma dessas, pode-se realizar as estimativas que permitam obter uma amostra aleatória da distribuição do estimador (FERREIRA, 2013).

Suponha uma amostra aleatória $X_1,\ X_2,\ \ldots,\ X_n$ de uma população com função densidade probabilidade (f.d.p.) $f_X(x|\theta)$. Seja $\hat{\theta}$ o estimador de θ — frequentemente o estimador de máxima verossimilhança (EMV) — que será usado na f.d.p. $f_X(x|\hat{\theta})$ para obter B amostras de tamanho n, de modo que cada amostra aleatória $X_1^*,\ X_2^*,\ \ldots,\ X_n^* \sim f_X(x|\hat{\theta})$. Com o uso das B amostras obtidas, pode-se calcular estimativas de alguma função dos dados de interesse e de sua respectiva distribuição (CASELLA; BERGER, 2002).

2.7 Teste de aderência de Kolmogorov-Smirnov

Comumente, na Estatística, tem-se que testar a hipótese sobre a distribuição de uma população. Se o teste busca relacionar a distribuição de um conjunto de valores amostrados com uma distribuição teórica, ele é chamado de teste de aderência. Massey Júnior (1951) apresentou evidências de que, caso a distribuição acumulada da população seja contínua, o teste de aderência de Kolmogorov-Smirnov é o melhor a ser utilizado. Neste estudo, lidar-se-á com distribuições contínuas, portanto se optou por utilizá-lo. Esse teste faz parte dos testes não paramétricos e também pode ser empregado em comparações de amostras quando os pressupostos inerentes aos testes paramétricos não são atendidos (MELLO; PETERNELLI, 2013). O teste será descrito como apresentado em Mood, Graybill e Boes (2001).

Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com função de distribuição acumulada (f.d.a.) contínua $F_X(.) = F_0(.)$. Defina:

$$D_n = d_n(X_1, \ldots, X_n) = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|,$$

em que $F_n(x)$ é a função de distribuição acumulada empírica.

Suponha que a intenção seja testar se a distribuição que está sendo amostrada é alguma distribuição contínua especificada, tal que: $H_0 : X_i \sim F_0(.)$, em que $F_0(.)$ é alguma f.d.a. completamente especificada. Se H_0 é falsa, então $F_n(.)$ irá se aproximar da verdadeira f.d.a. $F(.)$ e não de $F_0(.)$ e, consequentemente, D_n tenderá a ser grande. Desse modo, um critério de teste razoável é rejeitar H_0 se o valor de D_n for grande. D_n segue uma distribuição assintótica, denomine de $H(.)$ e, quando H_0 é verdadeira, $d_{1-\alpha}$ pode ser determinado de modo que $1 - H(d_{1-\alpha}) = \alpha$ e então $P[D_n > d_{1-\alpha}] \approx \alpha$. Isto é, rejeita-se H_0 se, e somente se, $D_n > d_{1-\alpha}$, dado um nível de significância α.

2.8 Testes de normalidade multivariada

Quando a necessidade de se testar a normalidade multivariada se tornou evidente, a primeira ideia que surgiu foi a de testar a normalidade das distribuições marginais para verificar a normalidade conjunta. Isso faria com que o problema de dimensão multivariada pudesse ser tratado em dimensões univariadas. Contudo, essa ideia é errônea. Pois, se por um lado, distribuição conjunta normal multivariada implica em marginais normais univariadas, o contrário não é sempre verdadeiro. Há casos em que as distribuições marginais são normais univariadas e a conjunta não é normal multivariada (FERREIRA, 2011).

Ademais, testar a normalidade multivariada com base nas distribuições marginais univariadas está sujeito ao problema da multiplicidade, isto é, o nível de significância nominal α estabelecido para cada teste de normalidade na distribuição marginal não é garantido simultaneamente. Se α é a probabilidade do erro tipo *I* em cada um dos p testes individualmente, têm-se um nível de confiança $1-\alpha$ para cada teste. Se as estatísticas de teste forem independentes para os p testes simultaneamente e se a hipótese nula de cada teste for verdadeira, o nível de confiança global é de $(1-\alpha)^p$. Desse modo, com o aumento de p, a probabilidade do erro tipo *I* global das inferências simultâneas aumenta e é de $1-(1-\alpha)^p$ (FERREIRA, 2011).

Um dos testes para normalidade multivariada mais utilizados é a generalização do teste de Shapiro e Wilk (1965) para o caso multivariado, feita por Royston (1983). Para o procedimento do teste, considere $X_{(1)k},\ X_{(2)k},\ ..., X_{(n)k}$ a amostra da k-ésima variável, em ordem crescente. Royston (1983) sugere a seguinte estatística:

$$W_k = \frac{\left[\sum\limits_{j=1}^{n} \tilde{a}_j X_{(j)k}\right]^2}{\sum\limits_{j=1}^{n} \left(X_{jk} - \bar{X}_{.k}\right)^2},$$

em que $\bar{X}_{.k} = \sum_{j=1}^{n} \frac{X_{jk}}{n}$ e $\tilde{a}_j$ é o estimador de a do j-ésimo componente de $\boldsymbol{a} = [a_1,\ a_2,\ ...,\ a_n]^\top$. Esse vetor $\boldsymbol{a}$ é definido da seguinte forma:

$$\boldsymbol{a} = \frac{\boldsymbol{V^{-1}m}}{\sqrt{\boldsymbol{m}^\top \boldsymbol{V^{-2}m}}},$$

em que $\boldsymbol{m} = [m_1,\ m_2,\ ...,\ m_n]^\top$ define o vetor de médias e $\boldsymbol{V} = [v_{ij}]$ define a matriz de covariâncias de dimensão $(n \times n)$ das estatísticas de ordem da normal padrão.

A estatística W_k não segue distribuição normal, portanto, o próximo passo é a transformação da estatística W_k em Y_k. Uma transformação da família Box-Cox, proposta por Royston (1993) é dada por:

$$Y_k = \begin{cases} -ln[\gamma - ln(1 - W_k)] & \text{se} \quad 4 \leq n \leq 11, \\ ln(1 - W_k) & \text{se} \quad 12 \leq n \leq 5000, \end{cases}$$

em que $\gamma = -2{,}273 + 0{,}459n$. A média e o desvio padrão de Y_k são apresentados em Royston (1993). A partir de Y_k, obtém-se um escore normal padrão Z_k e, a seguir, realiza-se o cálculo da quantidade expressa como (ROYSTON, 1983):

$$\kappa_k = \left\{ \Phi^{-1} \left[\frac{1}{2} \Phi\left(-Z_k\right) \right] \right\}^2, \quad k = 1,\ 2,\ ..., p,$$

em que $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} exp \frac{-t^2}{2} dt.$

A estatística do teste, definida por Royston (1983), é dada por:

$$H = \frac{\nu}{p} \sum_{k=1}^{p} \kappa_k,$$

que, sob a hipótese nula de normalidade, possui distribuição aproximada de qui-quadrado com ν graus de liberdade, cujos cálculos também são apresentados por Royston (1983). O valor-p é calculado por:

$$\text{valor-}p = 1 - F_{\chi^2}(H;\nu),$$

em que $F_{\chi^2}(H;\nu)$ é a função de distribuição qui-quadrado com ν graus de liberdade avaliada no ponto H.

Mardia (1970) apresentou um teste de normalidade multivariada baseado no coeficiente de assimetria e outro, fundamentado no coeficiente de curtose. Seja a amostra aleatória $\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ ...,\ \boldsymbol{X}_n$ de uma distribuição normal multivariada $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$, considere os coeficientes de assimetria e curtose, respectivamente, dados por β_{1p} e β_{2p}. O estimador não viesado de β_{1p} é $\hat{\beta}_{1p}$ e o estimador viesado de β_{2p} é $\hat{\beta}_{2p}$, uma vez que $E(\hat{\beta}_{1p}) = 0$ e $E(\hat{\beta}_{2p}) = p(p+2)(n-1)/(n+1)$. Mardia (1970) mostrou as variâncias dos estimadores como $Var(\hat{\beta}_{1p}) = 6/n$ e $Var(\hat{\beta}_{2p}) = 8p(p+2)/n$. E utilizou as propriedades assintóticas da distribuição desses estimadores para propor os testes. O autor propôs um teste para a hipótese nula $H_0\ :\ \beta_{1p} = 0$, cuja estatística é:

$$\chi^2_c = \frac{n\hat{\beta}_{1p}}{6}$$

e tem distribuição assintótica de qui-quadrado com $\nu = p(p+1)(p+2)/6$ graus de liberdade. Rejeita-se H_0 se o valor de $\chi^2_c \geq \chi^2_{\alpha,\nu}$. Também foi proposto um teste para $H_0\ :\ \beta_{2p} = p(p+2)$, cuja estatística é:

$$Z_c = \frac{\hat{\beta}_{2p} - p(p+2)(n-1)/(n+1)}{\sqrt{8p(p+2)/n}}$$

e tem distribuição assintótica normal padrão $N(0{,}1)$, em que se rejeita H_0 se $|Z_c| \geq Z_{\alpha/2}$. Em ambos, α é o nível de significância nominal. Mardia (1970) relata que o tamanho da amostra, no caso da distribuição T^2 de Hotelling, influencia mais no coeficiente de assimetria do que no coeficiente de curtose. Contudo, deve-se ressaltar que tanto os testes com base no coeficiente de assimetria quanto aqueles com base no coeficiente de curtose são limitados à amostras de tamanhos relativamente grandes, em razão das propriedades assintóticas.

Doornik e Hansen (2008) propuseram um teste conjunto de assimetria e curtose. Primeiramente, realiza-se uma transformação linear na amostra aleatória multivariada para obter uma amostra aleatória univariada. Em seguida, são calculados os coeficientes de assimetria e curtose de cada marginal univariada. Com os coeficientes, faz-se outra transformação na amostra aleatória univariada, conseguindo amostras independentes e identicamente distribuídas da normal padrão. Então, é possível calcular a estatística do teste, expressa como:

$$E_p = Z_1^\top Z_1 + Z_2^\top Z_2 \dot{\sim} \chi^2_{2p},$$

em que $Z_1^\top = [z_{11},\ ...,\ z_{2p}]$ são os escores normais padrão obtidos com base no coeficiente de assimetria e $Z_2^\top = [z_{21},\ ...,\ z_{2p}]$, os escores normais padrão obtidos baseados no coeficiente de curtose. O teste tem a mesma limitação do de Mardia (1970), devido à distribuição assintótica da estatística. Nos casos de amostras de tamanho grande, o teste proposto por Doornik e Hansen (2008) apresenta bom desempenho de poder, conforme Farrell, Salibian-Barrera e Naczk (2007) avaliaram.

Um generalização do teste de Shapiro-Francia univariado para o caso multivariado foi proposto por Silva (2009). O método é similar ao que foi criado por Royston (1983), sendo que as principais diferenças se referem à estatística do teste, mais especificamente, aos coeficientes e ao cálculo dos graus de liberdade da distribuição da estatística. O desempenho do teste foi testado por Silva (2009) e comparado ao do teste de Royston (1983). Os resultados mostraram que os dois testes têm desempenho equivalente.

Székely e Rizzo (2005) propuseram um teste para normalidade multivariada fundamentado na distância euclidiana entre os elementos da amostra. O teste parte da padronização dos dados para que se tenha vetor de médias nulo e matriz de covariâncias igual à identidade. A estatística do teste é dada por:

$$\varepsilon_{n,p} = n\left(\frac{2}{n}\sum_{j=1}^{n} E\|z_j - \boldsymbol{Z}\| - E\|\boldsymbol{Z} - \boldsymbol{Z'}\| - \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\|z_j - z_k\|\right),$$

em que $\boldsymbol{Z}$ e $\boldsymbol{Z'}$ são vetores de variáveis aleatórias independentes e identicamente distribuídas $N_p(\mathbf{0}, \mathbf{1})$, e

$$E\|\boldsymbol{Z} - \boldsymbol{Z'}\| = \sqrt{2}E\|\boldsymbol{Z}\| = 2\frac{\Gamma\left(\dfrac{p+1}{2}\right)}{\Gamma\left(\dfrac{p}{2}\right)},$$

$$E\|z_j - \boldsymbol{Z}\| = \frac{\sqrt{2}\Gamma\left(\dfrac{p+1}{2}\right)}{\Gamma\left(\dfrac{p}{2}\right)} + \sqrt{\frac{2}{\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!2^k}\frac{\|z_j\|^{2k+2}}{(2k+1)(2k+2)}\frac{\Gamma\left(\dfrac{p+1}{2}\right)\Gamma\left(k+\dfrac{3}{2}\right)}{\left(k+\dfrac{p}{2}+1\right)}$$

e

$$z_j = \boldsymbol{S}^{-1/2}(x_j - \bar{\boldsymbol{X}}).$$

A distribuição nula da estatística do teste é obtida por *bootstrap* paramétrico. As avaliações de desempenho realizadas por Székely e Rizzo (2005) apontaram excelentes resultados em relação ao controle da taxa de erro tipo I e um poder elevado.

Um outro teste de generalização do teste de Shapiro e Wilk (1965) univariado para o caso multivariado foi proposto por Alva e Estrada (2009). Trata-se de um teste Monte Carlo em que, primeiramente, realiza-se uma transformação nos dados da seguinte forma:

$$\boldsymbol{Z}_j = \hat{\boldsymbol{\Sigma}}^{-1/2}\left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right), \quad j = 1, 2, ..., n.$$

A estatística do teste é dada por:

$$W^* = \frac{1}{p}\sum_{i=1}^{p} W_{Z_i},$$

em que W_{Z_i} é a estatística do teste de Shapiro-Wilk, avaliada na i-ésima variável transformada Z_i.

Para se obter a distribuição nula da estatística do teste, computa-se milhares de vezes a estatística W_m^*, $m = 1, 2, ..., B$. Então, o valor-p do teste é calculado como segue:

$$\text{valor - } p = \frac{\sum\limits_{m=1}^{B} I(W_m^* \leq W^*)}{B}.$$

Para um determinado nível de significância α, rejeita-se H_0 se, e somente se, valor-$p \leq \alpha$.

Alva e Estrada (2009) avaliaram o desempenho do teste por meio de simulação Monte Carlo e concluíram que o teste proposto por eles foi mais poderoso do que os testes de Mardia (1970), de Henze e Zirkler (1990) e de Srivastava e Hui (1987).

A partir do resultado 2 proveniente do Teorema 1 apresentado na Seção 2.3, foram desenvolvidos outros testes. Trata-se de que a distância quadrática, dada por $(\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ segue distribuição qui-quadrado com p graus de liberdade. Como $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são desconhecidos, usa-se os seus respectivos estimadores, $\bar{\boldsymbol{X}}$ e $\boldsymbol{S}$. Desse modo, tem-se que:

$$D_j^2 = (\boldsymbol{X} - \bar{\boldsymbol{X}})^\top \boldsymbol{S}^{-1}(\boldsymbol{X} - \bar{\boldsymbol{X}}) \dot{\sim} \chi_p^2\,, \qquad j = 1,\ 2,\ ...,\ n, \tag{5}$$

ou seja, essa distância quadrática de Mahalanobis segue, assintoticamente, uma distribuição qui-quadrado com p graus de liberdade (FERREIRA, 2011).

Paulson, Roohan e Sullo (1987) propuseram uma extensão do teste de normalidade univariado de Anderson e Darling (1952) para o caso multivariado. A estatística do teste é dada por:

$$\begin{aligned} A_{n,p}^2 &= n\int_0^\infty \frac{F_n(D^2) - F_p(D^2)}{F_p(D^2)(1 - F_p(D^2))} dF_p(D^2) \\ &= -\left(\frac{1}{n}\right)\sum_{j=1}^{n}(2j-1)\{\log F_p(D_{(j)}^2) + \log[1 - F_p(D_{(n+1-j)}^2)]\} - n, \end{aligned}$$

em que $F_n(D^2)$ é a função da distribuição empírica com correção de continuidade k/n, $F_p(\cdot)$ representa a função de distribuição acumulada empírica da distribuição $\chi_{(p)}^2$ e D_j^2 é a distância de Mahalanobis, dada pela expressão (5).

Outros testes para normalidade multivariada foram sugeridos com base nessa distância. Gráficos quantil-quantil (Q-Q plots) foram construídos utilizando as estatísticas de ordem, $D_{(j)}^2$, e os quantis da distribuição qui-quadrado com p graus de liberdade, χ_j^2. Um teste formal foi proposto por Cirillo e Ferreira (2003), baseado no coeficiente de correlação quantil r_q, dado por:

$$r_q = \frac{\sum_{j=1}^{n} D_{(j)}\chi_j - \frac{\sum_{j=1}^{n} D_{(j)} \sum_{j=1}^{n} \chi_j}{n}}{\sqrt{\sum_{j=1}^{n} D_j^2 - \frac{\left(\sum_{j=1}^{n} D_j\right)^2}{n}}\sqrt{\sum_{j=1}^{n} \chi_j^2 - \frac{\left(\sum_{j=1}^{n} \chi_j\right)^2}{n}}}.$$

A hipótese de normalidade multivariada deve ser rejeitada, se, e somente se, $r_q \leq r_{q,\alpha}$, para um nível de significância α. Os valores críticos para α, n e p estão disponíveis em Cirillo e Ferreira (2003). De acordo com os autores, o teste foi eficaz para controle do erro tipo I, mas o poder se reduziu com o aumento do número de variáveis.

Uma limitação dos testes que se baseiam na distância D_j^2 é a de que a distribuição qui-quadrado não é adequada para quando o tamanho da amostra é pequeno, por se tratar de uma distribuição assintótica (FERREIRA, 2011).

Como já foi apresentado na Seção 2.3, Gnanadesikan e Kettenring (1972) apresentaram a transformação $B_j = \frac{nD_j^2}{(n-1)^2}$, cuja distribuição é uma beta exata com parâmetros $\alpha = p/2$ e $\beta = (n - p - 1)/2$, com $j = 1,\ 2,\ ...,\ n$. A próxima seção relata um teste visual e um teste formal propostos com base nessa transformação.

2.8.1 Teste de normalidade multivariada utilizando amostra dependente de uma distribuição beta

Considere uma amostra aleatória $\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ ...,\ \boldsymbol{X}_n$ de uma distribuição normal multivariada $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$. Sejam $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ desconhecidos e $\bar{\boldsymbol{X}}$ e $\boldsymbol{S}$ os seus estimadores, respectivamente. Dada a distância de Mahalanobis expressa na equação (5), Gnanadesikan e Kettenring (1972) mostraram a transformação apresentada na

Seção 2.3, tal que:

$$B_j = \frac{nD_j^2}{(n-1)^2}$$

segue distribuição beta exata com parâmetros $\alpha = p/2$ e $\beta = (n-p-1)/2$.

Para se construir uma análise gráfica Q-Q plot com base nos B_j, obtêm-se as estatísticas de ordem $B_{(1)}, B_{(2)}, ..., B_{(n)}$, ordenando os valores de B_j. A próxima etapa é estimar os quantis da distribuição beta (b_j^*), como segue (FERREIRA, 2011):

$$b_j^* = F_X^{-1}\left(\frac{j-a}{n-a-b+1}\right),$$

em que $F_X^{-1}(.)$ é a função de distribuição inversa da beta, $a = (p-2)/(2p)$ e $b = (n-p-3)/[2(n-p-1)]$.

Finalmente, deve-se plotar os valores de $B_{(j)}$ e b_j^* e observar se há algum padrão de não-linearidade, se houver é um indício de desvio de normalidade multivariada. O problema desse teste visual é que os B_j não são variáveis aleatórias independentes e identicamente distribuídas.

Embrechts, Frey e McNeil (2005) propuseram um teste formal utilizando as transformações na distância de Mahalanobis que conduzem à amostra beta exata. Para isso, Embrechts, Frey e McNeil (2005) basearam-se no teste de falta de ajuste de Kolmogorov-Smirnov ao conjunto de pares $(B_j,\ b_j^*)$ para calcular o valor-p. Desse modo, sob a hipótese nula de normalidade multivariada, pode-se testar a aderência de B_j à distribuição beta com parâmetros $\alpha = p/2$ e $\beta = (n-p-1)/2$. Os autores apresentaram o mesmo teste para a aderência de D_j^2 a uma distribuição χ_p^2. Ambos tiveram resultados equivalentes na rejeição da hipótese de normalidade multivariada.

Quanto ao desempenho dos testes, não foram encontradas avaliações que o descrevesse. Contudo, em virtude das variáveis B_j, com $j = 1,\ 2,\ ...,\ n$ não serem variáveis aleatórias independentes, a aplicabilidade do teste de aderência de Kolmogorov-Smirnov é comprometida.

2.8.2 Q-Q plot utilizando amostra independente de uma distribuição beta

As definições, teorema e corolário desta seção são encontrados em Liang, Pan e Yang (2004).

Seja $\boldsymbol{X} \in \mathbb{R}^p$ uma variável aleatória que segue distribuição normal multivariada $N_p(\mathbf{0},\mathbf{\Sigma})$, com média $\mathbf{0}$ e matriz de covariâncias $\mathbf{\Sigma}$. Dada uma amostra aleatória $\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ ...,\ \boldsymbol{X}_n$ independente e identicamente distribuída (i.i.d.), da variável aleatória $\boldsymbol{X}$, temos:

$$\boldsymbol{S}_k = \sum_{i=1}^{k} \boldsymbol{X}_i \boldsymbol{X}_i^{\top},$$

$$\boldsymbol{Y}_k = \boldsymbol{S}_k^{-1/2} \boldsymbol{X}_k, \tag{6}$$

$$\boldsymbol{Z}_k = \frac{\boldsymbol{Y}_k}{\sqrt{1 - \mathbf{Y}_k^{\top} \mathbf{Y}_k}}, \tag{7}$$

para $k = p+1, ..., n$, em que $\boldsymbol{S}_k^{-1/2} = (\boldsymbol{S}_k^{1/2})^{-1}$ e $\boldsymbol{S}_k^{1/2}$ é a raiz quadrada positiva definida de $\boldsymbol{S}_k$.

Então, é possível fazer uma caracterização para a distribuição normal multivariada $N_p(\mathbf{0},\mathbf{\Sigma})$, como descreve o teorema enunciado a seguir, apresentado em Yang, Fang e Liang (1996):

Teorema 2. *Seja $\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ ...,\ \boldsymbol{X}_n$ uma amostra i.i.d. da $N_p(\mathbf{0},\mathbf{\Sigma})$. Defina os vetores aleatórios $\mathbf{Y}_k$ e $\boldsymbol{Z}_k$ como em (6) e (7), respectivamente. Então,*

1. *$\mathbf{Y}_{p+1}, ..., \mathbf{Y}_n$ são mutuamente independentes e $\mathbf{Y}_k (k \geq p+1)$ segue uma*

distribuição multivariada simétrica Pearson tipo II com a função densidade probabilidade (f.d.p.)

$$f_{\boldsymbol{Y}}(\boldsymbol{y};k) = \frac{\Gamma(k/2)}{\pi^{p/2}\Gamma(k-p)/2})(1-\boldsymbol{y}^\top\boldsymbol{y})^{(k-p-2)/2},\ \boldsymbol{y}\in\mathbb{R}^p,\ \boldsymbol{y}^\top\boldsymbol{y}<1;$$

2. *$\boldsymbol{Z}_{p+1},...,\boldsymbol{Z}_n$ são mutuamente independentes e $\boldsymbol{Z}_k(k \geq p+1)$ tem uma distribuição multivariada simétrica Pearson tipo VII com a f.d.p.*

$$h_{\boldsymbol{Z}}(\boldsymbol{z};k) = \frac{\Gamma(k/2)}{\pi^{p/2}\Gamma(k-p)/2})(1+\boldsymbol{z}^\top\boldsymbol{z})^{-k/2},\ \boldsymbol{z}\in\mathbb{R}^p; \tag{8}$$

3. *Seja $\boldsymbol{X}_1,\ ...,\ \boldsymbol{X}_n \in \mathbb{R}^p$ i.i.d. com uma f.d.p. $\nu(\boldsymbol{x})$ que é contínua em $\boldsymbol{x}\in\mathbb{R}^p$ e $\nu(\boldsymbol{0})>0$. Defina os vetores aleatórios $\boldsymbol{Z}_k$ como em (7). Se $\boldsymbol{Z}_k$ e $\boldsymbol{Z}_{k-1}(k\geq p+2)$ tem f.d.p.'s $h_{\boldsymbol{Z}}(\boldsymbol{z};k)$ e $h_{\boldsymbol{Z}}(\boldsymbol{z};k-1)$ definidas em (8). Então, $\boldsymbol{X}_i$ $(i=1,...,n)$ tem uma distribuição normal multivariada.*

Corolário 1. *Assuma que $\boldsymbol{X}_1,\ \ldots,\ \boldsymbol{X}_n$ são i.i.d. $\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Defina os vetores aleatórios*

$$\boldsymbol{U}_i = \frac{\boldsymbol{X}_1+\ldots+\boldsymbol{X}_i - i\boldsymbol{X}_{i+1}}{\sqrt{i(i+1)}},\ i=1,\ldots,n-1,$$

$$\boldsymbol{S}_k = \sum_{i=1}^{k}\boldsymbol{U}_i\boldsymbol{U}_i^\top \text{ e } \mathbf{Y}_k = \boldsymbol{S}_k^{-1/2}\boldsymbol{U}_k,\ k=p+1,\ldots,n-1.$$

$$\textit{Seja } \mathbf{Y}_k = (Y_{k1},\ldots,Y_{kp})^\top : p\times 1(k=p+1,\ldots n-1),$$

$$\bar{Y}_k = \frac{1}{p}\sum_{i=1}^{p}Y_{ki},\ T_k = \frac{\sqrt{p}\bar{Y}_k}{S_k},\ S_k^2 = \frac{1}{p-1}\sum_{i=1}^{p}(Y_{ki}-\bar{Y}_k)^2;$$

$$F_{kq} = \frac{(p-q)\sum_{i=1}^{q} Y_{ki}^2}{q\sum_{i=q+1}^{p} Y_{ki}^2}, \; B_{kq} = \frac{\sum_{i=1}^{q} Y_{ki}^2}{\sum_{i=1}^{p} Y_{ki}^2}, \; q = 1, ..., p-1;$$

para $k = 1, ..., n-1$. *Então,*

1. $\{T_k \; : \; k = p+1, ..., n-1\}$ *são i.i.d.* $\sim t(p-1)$*, a distribuição t-Student com* $p-1$ *graus de liberdade;*

2. *Para cada* $q(q = 1, ..., p-1)$, $\{F_{kq} \; : \; k = p+1, ..., n-1\}$ *são i.i.d.* $\sim$ $F(q, p-q)$*, a distribuição F com* $(q, p-q)$ *graus de liberdade;*

3. *Para cada* $q(q = 1, ..., p-1)$, $\{B_{kq} \; : \; k = p+1, ..., n-1\}$ *são i.i.d.* $\sim$ $B(q/2, (p-q)/2)$*, a distribuição beta com parâmetros* $q/2$ *e* $(p-q)/2$*.*

Se agruparmos os B_{kq} em uma matriz $\boldsymbol{B}$, temos que:

$$\boldsymbol{B} = \begin{bmatrix} B_{p+1,1} & B_{p+1,2} & \cdots & B_{p+1,p-1} \\ B_{p+2,1} & B_{p+2,2} & \cdots & B_{p+2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n-1,1} & B_{n-1,2} & \cdots & B_{n-1,p-1} \end{bmatrix},$$

$$\underbrace{\qquad}_{beta\left(\frac{1}{2},\frac{p-1}{2}\right)} \quad \underbrace{\qquad}_{beta\left(1,\frac{p-2}{2}\right)} \quad \underbrace{\qquad}_{beta\left(\frac{p-1}{2},\frac{1}{2}\right)}$$

ou seja, cada coluna dessa matriz $\boldsymbol{B}$ é amostra aleatória de uma distribuição beta com parâmetros $q/2$ e $(p-q)/2$.

Conforme foi apresentado, as variáveis aleatórias B_{kq}, $q = 1, ..., p-1$, provenientes das transformações são independentes. Portanto, o Teorema 2 e Corolário 1 se propõem como procedimentos, sugestivamente, melhores do que aqueles apresentados na Seção 2.8.1.

Obtidas as variáveis aleatórias B_{kq}, Li, Fang e Zhu (1997) sugerem que seja usado o valor $q = \lfloor p/2 \rfloor$, ou seja, a parte inteira da metade de p, na construção do Q-Q plot. A escolha está fundamentada no bom desempenho desse valor em experiências com dados. Utilizando esse valor de q recomendado por Li, Fang e Zhu (1997), basta seguir os passos descritos na seção anterior para construir os Q-Q plots: obter as estatísticas de ordem $B_{(k)q}$, estimar os quantis da distribuição beta (b^*_{kq}) e plotar os valores $B_{(k)q}$ e b^*_{kq}, observando se há algum padrão linear. Caso não haja, interpreta-se como um desvio de normalidade multivariada.

2.9 Testes de normalidade multivariada implementados no programa R

O programa R, ambiente R, ou apenas R, como é conhecido pelos seus usuários, foi criado em 1996 pelos professores de estatística Ross Ihaka e Robert Gentleman. Por ser um programa de código fonte aberto, é livre para qualquer pessoa usar e modificar. O R é semelhante a outras linguagens de programação, como C, Java e Perl, que dá acesso a vários comandos, ajudando a executar uma grande variedade de tarefas de computação. O programa pode ser considerado uma implementação da linguagem S, que foi desenvolvida nos Laboratórios Bell (atual Lucent Technologies), por Rick Becker, John Chambers e Allan Wilks (R CORE TEAM, 2015; VENABLES; SMITH; TEAM, 2015).

A equipe R CORE TEAM (2015) que mantém o programa R o define como uma suíte integrada para manipulação de dados, cálculos e exibição gráfica. Algumas das vantagens do R relatadas pelo grupo são:

- manipulação de dados eficaz e facilidade de armazenamento;
- conjunto de operadores para cálculos com matrizes;
- grande coleção de ferramentas intermediárias para análise de dados;
- facilidades para análise gráfica de dados e visualização;

- uma linguagem de programação bem desenvolvida, simples e eficaz, que inclui condicionais, *loops*, funções recursivas definidas pelo usuário e recursos de entrada e saída.

Apesar de muitas pessoas utilizarem o R como um sistema de ferramentas estatísticas, o R CORE TEAM (2015) prefere defini-lo como um ambiente em que muitas técnicas estatísticas clássicas e modernas foram implementadas. Algumas dessas técnicas estão incluídas na base do programa R, como pacotes padrões, que são instalados com o programa. Além disso, há muitas outras fornecidas como pacotes, que são criados pelos colaboradores e disponibilizados no site Comprehensive R Archive Network (CRAN) do programa (https://cran.r-project.org).

O R pode ser obtido gratuitamente no site CRAN, onde é apresentado em versões de acordo com o sistema operacional UNIX, Windows ou Macintosh. A facilidade que o usuário encontra para criar suas próprias funções e rotinas na análise de dados e a característica de ser um programa livre contribuem para aumentar a diversidade de pacotes disponíveis. Assim, a maioria das metodologias estatísticas clássicas e mais recentes está disponível para uso no R e, com uma pesquisa mais avançada, é possível encontrá-las (VENABLES; SMITH; TEAM, 2015). Sobretudo, a facilidade de uso permite que qualquer estatístico, engenheiro ou cientista seja capaz de utilizar o ambiente R, fazendo com que o número de usuários seja crescente.

Como já foi mencionado, há vários pacotes para o programa R criados por contribuidores que podem ser utilizados. Sendo assim, existem muitas implementações dos principais testes de normalidade multivariada que são disponibilizadas como pacotes no site CRAN (https://cran.r-project.org/web/packages/).

Os pacotes e as funções de alguns dos principais testes de normalidade multivariada disponíveis para o programa R estão dispostos na Tabela 1.

Mecklin e Mundfrom (2005) fizeram uma comparação das taxas de erro tipo I e tipo II de treze testes de normalidade multivariada que eles consideraram promissores, com o uso de simulação Monte Carlo. Os autores recomendaram o uso do teste de Henze e Zirkler (1990) como teste formal.

Tabela 1 Alguns testes de normalidade multivariada disponíveis no programa R

Teste de normalidade multivariada	Pacote	Função
Henze-Zirkler	*MVN*	*HZ.test*
Szekely e Rizzo	*energy*	*mvnorm.etest*
Domanski (Shapiro-Wilk Royston)	*mvnormtest*	*mshapiro.test*
Mardia	*dprep*	*mardia*
Shapiro-Francia	*mvsf*	mvsf
Doornik e Hansen	*asbio*	*DH.test*
Anderson-Darling	*robCompositions*	*adtest*
Alva e Estrada	*mvShapiroTest*	*mvShapiro.Test*
McNeil, Frey e Embrechts	*QRMlib*	*jointnormalTest*
Três momentos	*dprep*	*mo3*
Quatro momentos	*dprep*	*mo4*
Baseado em curtose	*ICS*	*mvnorm.kur.test*
Baseado em assimetria	*ICS*	*mvnorm.skew.test*
Ruído branco	*normwhn.test*	*normwhn.test-package*
Conjunta com independência	*normwhn.test*	*normality.test1*
Conjunta sob fraca dependência	*normwhn.test*	*normality.test2*
Baseado em simetria e curtose	*vars*	*normality*

Outro estudo de comparação de desempenho realizado por Cantelmo e Ferreira (2007) mostrou que o teste de Shapiro-Wilk multivariado, função *mshapiro.test*, teve taxas de erro tipo I superiores aos valores nominais. Devido ao desempenho fraco, os autores não recomendaram o uso rotineiro dessa função. Em contrapartida, os testes de assimetria e curtose para a normalidade multivariada foram recomendados em casos de $n \geq 50$ e $n \geq 100$.

Nos estudos, Mecklin e Mundfrom (2005) afirmaram que não encontraram um teste que fosse o mais poderoso em todas as situações. Nesse sentido, Cantelmo e Ferreira (2007) relataram a importância de se fazerem avaliações frequentes de testes de hipóteses implementados como pacotes do programa R.

3 MÉTODOS

Os métodos para realização do trabalho abrangeram a definição de dois testes envolvendo transformações multivariadas para marginais beta e simulação de dados, para verificação da qualidade e desempenho dos testes. Como auxílio computacional foi utilizado o programa R (R CORE TEAM, 2015) para implementar os algoritmos dos dois testes de normalidade multivariada propostos: o teste de aderência baseado no teste de Kolmogorov-Smirnov (TKS) e o teste intensivo baseado em *bootstrap* paramétrico (TBPb). O programa R também foi usado para realizar simulações Monte Carlo com o propósito de estimar as taxas de erro tipo I e o poder dos testes.

Foram realizadas comparações dos testes propostos com dois testes de normalidade multivariada: o teste de normalidade multivariada que foi apresentado por Embrechts, Frey e McNeil (2005) (TEM) e o teste de Shapiro-Wilk de normalidade multivariada proposto por Royston (1983) (TSW), cuja rotina foi implementada por Silva (2009) quando ele comparou esse teste com o teste de normalidade de Shapiro-Francia para o caso multivariado. O primeiro foi escolhido porque, assim como os testes propostos nesse trabalho, também é baseado em amostra beta, no entanto, é afetado pela dependência amostral presente na distância quadrática de Mahalanobis. O segundo foi selecionado para a comparação por ser um teste de normalidade multivariada bastante conhecido e com uso muito difundido. Ambos os testes comparados foram submetidos à validação por simulação Monte Carlo da mesma forma que os dois testes propostos.

3.1 Testes de normalidade multivariada propostos

Considere uma amostra aleatória representada por $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_j, \ldots, \boldsymbol{X}_n$, em que $\boldsymbol{X}_j \in \mathbb{R}^p$ provém de uma população com vetor de médias $\boldsymbol{\mu}$, cuja dimensão é $p \times 1$, e matriz de covariâncias $\boldsymbol{\Sigma}$, de dimensão $p \times p$. Pretende-se testar a hipótese nula seguinte:

H_0 : a amostra é de uma variável aleatória que segue distribuição normal multivariada $f_{\boldsymbol{X}}(\boldsymbol{x})$, dada pela expressão da equação (4), com vetor de médias $\boldsymbol{\mu} \in \mathbb{R}^p$ e matriz de covariância positiva definida $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$.

Os testes para normalidade multivariada que se fundamentam na distância de Mahalanobis sofrem influência da dependência entre os D_j^2, com $j = 1, \ldots, n$, pois cada D_j^2 possui em sua fórmula $\bar{\boldsymbol{X}}$ e $\boldsymbol{S}$. Portanto, a dependência amostral também está presente nos testes que são construídos utilizando amostra de uma distribuição beta exata, obtida por transformação na distância de Mahalanobis. Assim, o intuito deste trabalho foi propor dois testes alternativos para normalidade multivariada que se fundamentassem na obtenção de amostras independentes de distribuições betas exatas, de modo que pudesse eliminar o problema da dependência amostral.

Com base nas definições apresentadas na Seção 2.8.2, embasadas no teorema demonstrado por Yang, Fang e Liang (1996) e utilizadas por Liang, Pan e Yang (2004) para a construção de Q-Q plots, foram propostos dois testes para a normalidade multivariada. A primeira proposta foi de um teste de aderência baseado no teste de Kolmogorov-Smirnov e a segunda se referiu a um teste intensivo baseado em *bootstrap* paramétrico. As seções seguintes se dedicam à apresentação de ambas as propostas.

3.1.1 Teste de normalidade multivariada utilizando o teste de Kolmogorov-Smirnov baseado na obtenção de amostras betas

Por questões de organização, optou-se por descrever os passos dos testes propostos na forma de algoritmo. Inicialmente, considere a amostra aleatória e a hipótese nula apresentadas na seção anterior. Os passos para a realização do teste de normalidade multivariada utilizando Kolmogorov-Smirnov baseado na obtenção de amostras betas são:

1. considere uma matriz de dados $\boldsymbol{X}_{n\times p}$:

$$\boldsymbol{X}_{n\times p} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix};$$

2. calculam-se os vetores aleatórios $\boldsymbol{U}_i$:

$$\boldsymbol{U}_i = \frac{\boldsymbol{X}_1 + \ldots + \boldsymbol{X}_i - i\boldsymbol{X}_{i+1}}{\sqrt{i(i+1)}},\ i = 1, \ldots, n-1;$$

3. realizam-se as transformações para se obter $\boldsymbol{S}_k$:

$$\boldsymbol{S}_k = \sum_{i=1}^{k} \boldsymbol{U}_i \boldsymbol{U}_i^\top,\ k = p+1, \ldots, n-1;$$

4. obtém-se os fatores de Cholesky, $\boldsymbol{S}_k^{1/2}$, das matrizes $\boldsymbol{S}_k$;

5. invertem-se os fatores de Cholesky, de modo que se tenha $\boldsymbol{S}_k^{-1/2}$;

6. determinam-se os vetores $\boldsymbol{Y}_k$:

$$\boldsymbol{Y}_k = \boldsymbol{S}_k^{-1/2} \boldsymbol{U}_k,\ k = p+1, \ldots, n-1,$$

em que $\boldsymbol{Y}_k = (Y_{k1}, \ldots, Y_{kp})^\top : p \times 1 (k = p+1, \ldots n-1)$;

7. obtém-se os B_{kq}:

$$B_{kq} = \frac{\sum_{i=1}^{q} Y_{ki}^2}{\sum_{i=1}^{p} Y_{ki}^2},\ q = 1, ..., p-1,$$

que podem ser arranjados em uma matriz da seguinte forma:

$$\boldsymbol{B} = \begin{bmatrix} B_{p+1,1} & B_{p+1,2} & \dots & B_{p+1,p-1} \\ B_{p+2,1} & B_{p+2,2} & \dots & B_{p+2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n-1,1} & B_{n-1,2} & \dots & B_{n-1,p-1} \end{bmatrix}.$$

Caso os dados da matriz $\boldsymbol{X}_{n\times p}$ sejam uma amostra aleatória de uma distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então, cada coluna da matriz $\boldsymbol{B}$ será uma amostra aleatória da distribuição beta$(q/2,\ (p-q)/2)$, com $q = 1,\ 2,\ ...,\ p-1$ (LIANG; PAN; YANG, 2004).

8. seleciona-se a coluna cujo $q^* = \lfloor p/2 \rfloor$, como sugerido por Li, Fang e Zhu (1997);

9. aplica-se o teste de aderência de Kolmogorov-Smirnov na coluna selecionada para verificar se essa é uma amostra proveniente de uma variável aleatória que segue distribuição beta com parâmetros $q^*/2$ e $(p-q^*)/2$;

10. retorna-se o valor-p e o valor da estatística de Kolmogorov-Smirnov.

De acordo com o valor-p, toma-se a decisão de rejeitar ou não a hipótese nula de que a amostra provém da distribuição beta, dado um nível de significância nominal α. A não rejeição da hipótese nula do teste de Kolmogorov-Smirnov implica na não rejeição da hipótese nula de que a matriz de dados $\boldsymbol{X}_{n\times p}$ é uma amostra de uma normal p-variada, com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Do mesmo modo, a rejeição da hipótese nula do teste de Kolmogorov-Smirnov resulta em rejeição da normalidade multivariada. Essa conclusão é possível devido ao Corolário 1 do Teorema 2 da Seção 2.8.2.

3.1.2 Teste *bootstrap* paramétrico de normalidade multivariada baseado na obtenção de amostras betas

O segundo teste proposto se fundamenta na escolha do q cuja estatística do teste de Kolmogorov-Smirnov é máxima. Sendo, portanto, q^* igual ao máximo da estatística de Kolmogorov-Smirnov e não mais $q^* = \lfloor p/2 \rfloor$. Como a distribuição do máximo da estatística de Kolmogorov-Smirnov é desconhecida, recorreu-se ao método *bootstrap* paramétrico para se obter uma amostra aleatória da distribuição do máximo da estatística de Kolmogorov-Smirnov.

Novamente, considere a amostra aleatória e a hipótese nula apresentadas na Seção 3.1. Os passos para a realização do teste *bootstrap* paramétrico de normalidade multivariada baseado na obtenção de amostras betas são:

1. considere uma matriz de dados $\boldsymbol{X}_{n\times p}$:

$$\boldsymbol{X}_{n\times p} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \\ \vdots \\ \boldsymbol{X}_j^\top \\ \vdots \\ \boldsymbol{X}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jp} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix};$$

2. calculam-se os $\boldsymbol{U}_i$, $\boldsymbol{S}_k$, $\boldsymbol{Y}_k$ e B_{kq}, conforme passos de 2 a 7 do algoritmo da seção 3.1.1, a partir dos dados originais $\boldsymbol{X}_{n\times p}$. A matriz dos B_{kq}

$$\boldsymbol{B} = \begin{bmatrix} B_{p+1,1} & B_{p+1,2} & \dots & B_{p+1,p-1} \\ B_{p+2,1} & B_{p+2,2} & \dots & B_{p+2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n,1} & B_{n,2} & \dots & B_{n,p-1} \end{bmatrix};$$

3. calculam-se $D_q, \quad q = 1, ..., p-1$, ou seja, as estatísticas do teste de Kolmogorov-Smirnov para verificar a aderência da q-ésima coluna da matriz

$\boldsymbol{B}$ com uma distribuição beta com parâmetros $q/2$ e $(p-q)/2$, utilizando essa matriz B obtida da amostra original;

4. obtém-se $D_0^+ = max(D_q), \quad q = 1, ..., p-1$, isto é, o máximo das estatísticas de Kolmogorov-Smirnov obtidas com a matriz $\boldsymbol{B}$ proveniente da amostra original;

5. para a obtenção da distribuição nula *bootstrap*, utilizou-se o fato de que sob H_0, os $\boldsymbol{U}_i \sim N_p(\mathbf{0},\ \boldsymbol{\Sigma})$; Como $\boldsymbol{\Sigma}$ é desconhecida, utilizou-se a matriz de covariância $\boldsymbol{S}$ da amostra original $\boldsymbol{X}_{n\times p}$, para se utilizar como estimador da matriz $\boldsymbol{\Sigma}$ na reamostragem *bootstrap*;

6. geram-se N_B amostras *bootstrap* de tamanho $n-1$, $\boldsymbol{U}_1^*$, $\boldsymbol{U}_2^*$, ..., $\boldsymbol{U}_{n-1}^*$, da distribuição $N_p(\mathbf{0},\ \boldsymbol{S})$;

7. repetem-se os passos seguintes para cada uma dessas N_B amostras *bootstrap* geradas:

    (a) realizam-se as transformações para se obter $\boldsymbol{S}_k$:

    $$\boldsymbol{S}_k = \sum_{i=1}^{k} \boldsymbol{U}_i^* \boldsymbol{U}_i^{*\top},\ k = p+1, \ldots, n-1;$$

    (b) obtém-se os fatores de Cholesky, para se utilizar como $\boldsymbol{S}_k^{1/2}$, das matrizes $\boldsymbol{S}_k$;

    (c) invertem-se os fatores de Cholesky, de modo que se tenha $\boldsymbol{S}_k^{-1/2}$;

    (d) determinam-se os vetores $\boldsymbol{Y}_k$:

    $$\boldsymbol{Y}_k^* = \boldsymbol{S}_k^{-1/2} \boldsymbol{U}_k^*,\ k = p+1, \ldots, n-1,$$

    em que $\boldsymbol{Y}_k^* = (Y_{k1}^*, \ldots, Y_{kp}^*)^\top : p \times 1 (k = p+1, \ldots n-1)$;

(e) obtém-se os B^*_{kq}:

$$B^*_{kq} = \frac{\sum\limits_{i=1}^{q} Y^{*2}_{ki}}{\sum\limits_{i=1}^{p} Y^{*2}_{ki}},\ q = 1, ..., p-1;$$

(f) calculam-se D^*_q, $q = 1,\ ...,\ p-1$, ou seja, as estatísticas do teste de Kolmogorov-Smirnov de cada coluna da matriz $\boldsymbol{B}^*$, que é a matriz $\boldsymbol{B}$ obtida na amostra *bootstrap*;

(g) obtém-se $D^+_i = max(D_q), \quad q = 1,\ ...,\ p-1$, isto é, o máximo das estatísticas de Kolmogorov-Smirnov obtidas com a matriz $\boldsymbol{B}^*$, da i-ésima amostra de *bootstrap*, para $i = 1, 2, \ldots, N_B$;

8. obtida a amostra da distribuição nula *bootstrap*, dada por D^+_0, D^+_1, D^+_2,..., $D^+_{N_B}$, do máximo da estatística de Kolmogorov-Smirnov, calcula-se o valor-p da seguinte forma:

$$valor - p = \frac{\sum\limits_{i=0}^{N_B} I(D^+_0 \leq D^+_i)}{N_B + 1},$$

   em que $I(\cdot)$ é uma função indicadora;

9. retorna-se o valor-p e o valor da estatística D^+_0 da amostra original.

Dado um nível de significância α, rejeita-se H_0, a hipótese nula de que os dados seguem uma distribuição normal multivariada, se, e somente se, o valor-$p \leq \alpha$. Do mesmo modo que no teste proposto anteriormente, a conclusão só é possível devido ao Teorema 2 e Corolário 1 apresentados na Seção 2.8.2.

3.2 Validação do desempenho

Para avaliar as taxas de erro tipo I e o poder dos testes foi utilizada simulação Monte Carlo. Foram geradas amostras de diferentes tamanhos da distribuição normal multivariada e de outras distribuições não normais, utilizando funções geradoras do programa R.

Dado um nível de significância pré-estabelecido, aplicaram-se os testes propostos em cada amostra simulada, verificando se houve ou não rejeição da hipótese nula. Se a hipótese nula fosse rejeitada e a amostra fosse proveniente da distribuição normal multivariada, tinha-se um caso de erro tipo I. Por outro lado, se a hipótese nula fosse rejeitada e a amostra fosse de uma distribuição não normal, tinha-se uma decisão correta. Para que o teste fosse razoável, a taxa de erro tipo I deveria ser idêntica ao valor nominal, exceto pelo erro de Monte Carlo, e a proporção de rejeições sob H_1 teria de ser alta, indicando poder elevado.

3.2.1 Erro tipo I

As taxas de erro tipo I foram avaliadas utilizando simulação de amostras aleatórias normais p-variadas com vetor de médias $\boldsymbol{\mu} = \mathbf{0}$ e matriz de covariâncias $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\rho}$, sem perda de generalidade, em que $\boldsymbol{\rho}$ é dada por:

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}.$$

Para o teste Kolmogorov-Smirnov para normalidade multivariada, baseado na obtenção de amostras betas, pode-se utilizar essa estrutura de covariâncias na simulação para a validação, já que os parâmetros das distribuições betas, obtidas pelas transformações nos dados, são invariantes em relação à matriz $\boldsymbol{\Sigma}$. Deve-se ressaltar que esse teste está restrito a $n > p + 2$.

Quanto ao teste *bootstrap* paramétrico de normalidade multivariada baseado na obtenção de amostras betas, pode-se utilizar também a estrutura de covariâncias $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\rho}$ na simulação, conforme os passos de aplicação do teste descritos na Seção 3.1.2. A restrição desse teste é $n > p + 2$.

Para ambos os testes propostos e para os testes utilizados nas comparações, foram simuladas amostras normais de acordo com os diferentes valores de n e p, conforme apresentado na Tabela 2. Para os testes TBPb e TSW não foram apresentados resultados para a combinação $n = 10000$ e $p = 5$. No caso do TBPb, o tempo de execução das simulações de validação é demasiadamente longo. E no caso do TSW, a aplicação do teste se restringe às amostras de tamanho entre $n = 3$ e $n = 5000$.

Tabela 2 Tamanhos amostrais e número de variáveis utilizados nas simulações Monte Carlo

Testes de normalidade multivariada	n	p
TKS, TBPb, TEM e TSW	10	2, 5, 7
TKS, TBPb, TEM e TSW	20	2, 5, 10, 17
TKS, TBPb, TEM e TSW	50	2, 10, 25, 47
TKS, TBPb, TEM e TSW	100	2, 10, 50, 97
TKS, TBPb, TEM e TSW	200	2, 10, 50, 100
TKS e TEM	10000	5

Foram realizadas 2000 simulações Monte Carlo. Os dois testes propostos bem como os dois testes utilizados para comparação foram aplicados em cada amostra normal multivariada simulada aos níveis de significância nominais de 1%, 5% e 10%, de tal forma que foi computado o número de vezes que a hipótese nula foi rejeitada e dividido pelo número de simulações, fornecendo as taxas de erro tipo I empíricas.

3.2.2 Poder do teste

Para estimar o poder dos testes propostos, ou seja, a probabilidade de se rejeitar a hipótese nula dado que a mesma é falsa, foram simuladas amostras aleatórias de distribuições multivariadas não normais, utilizando as funções geradoras de variáveis aleatórias do programa R ou implementações próprias também em linguagem R. As distribuições utilizadas foram: normal contaminada, uniforme multivariada, log-normal multivariada e *t-Student* multivariada com grau de liberdade $\nu = 1$ e, também, com graus de liberdade $\nu = 30$.

O processo de simulação Monte Carlo foi repetido 2000 vezes para cada uma das distribuições não normais citadas, para cada tamanho da amostra n e os números de variáveis p correspondentes, apresentados na Tabela 2. Os testes de normalidade multivariada propostos e os testes comparados foram aplicados em cada amostra gerada e os níveis de significância nominais α fixados foram 1%, 5% e 10%.

Computou-se, então, o número de vezes que a hipótese nula foi rejeitada para cada teste, dividindo-o pelo total de amostras simuladas. Como as amostras geradas são não normais, o quociente representa a proporção de rejeições da hipótese nula falsa, estimando o poder de cada um dos dois testes propostos e dos testes utilizados para comparação.

3.2.3 Teste binomial exato

As taxas de erro tipo I estimadas utilizando simulações Monte Carlo estão sujeitas a erros e, para verificar o controle do erro tipo *I*, pode-se utilizar o teste binomial exato. Oliveira e Ferreira (2009) descreve esse teste do seguinte modo: seja y o número de H_0 rejeitadas em N simulações Monte Carlo, para um nível nominal de significância α^*, a estatística do teste usando a relação entre a distribuição F e a binomial, com probabilidade de sucesso $p = \alpha^*$, é dada por:

$$F = \left(\frac{y+1}{N-y}\right)\left(\frac{1-\alpha}{\alpha}\right),$$

que, sob H_0, tem distribuição F com $\nu_1 = 2(N-y)$ e $\nu_2 = 2(y+1)$ graus de liberdade.

A hipótese nula deve ser rejeitada ao nível nominal de significância α^*, Se $F \leq F_\alpha$ ou $F \geq F_{1-\alpha}$, em que F_α e $F_{1-\alpha}$ são os quantis da distribuição F com

ν_1 e ν_2 graus de liberdade.

Nesse trabalho o teste binomial exato foi aplicado nas taxas de erro tipo I considerando o tamanho da amostra da simulação Monte Carlo, $N = 2000$, no nível de significância $\alpha^* = 1\%$, para testar as seguintes hipóteses: $H_0 : \alpha = 1\%$ versus $H_1 : \alpha \neq 1\%$; $H_0 : \alpha = 5\%$ versus $H_1 : \alpha \neq 5\%$ e $H_0 : \alpha = 10\%$ versus $H_1 : \alpha \neq 10\%$.

Caso as taxas de erro tipo I sejam inferiores ao nível nominal α (valor-$p <$ 0,01), o teste proposto deve ser considerado conservativo. Caso, sejam superiores ao nível nominal α (valor-$p <$ 0,01), considera-se que o teste é liberal. Por fim, caso as taxas de erro tipo *I* não sejam significativamente diferentes (valor-$p >$ 0,01), o teste proposto deve ser classificado como exato.

4 RESULTADOS E DISCUSSÃO

Esta seção se dedica à apresentação dos resultados das simulações para validação dos dois testes propostos neste trabalho e, também, da comparação com dois outros testes já existentes.

O teste de normalidade multivariada utilizando o teste de Kolmogorov-Smirnov baseado na obtenção de amostra betas (TKS) e o teste *bootstrap* paramétrico de normalidade multivariada baseado na obtenção de amostras betas (TBPb), que foram propostos nesta dissertação, foram comparados com o teste de normalidade multivariada de Embrechts, Frey e McNeil (2005) (TEM) e com o teste de Shapiro-Wilk de normalidade multivariada proposto por Royston (1983), cuja rotina foi implementada por Silva (2009) (TSW).

Os valores das taxas de erro tipo I e do poder dos testes são dispostos de acordo com os níveis de significância nominais fixados de 1%, 5% e 10% e com os valores de n e p que foram definidos na Tabela 2. Desse modo, a presente seção se divide em duas etapas: uma que se refere ao erro tipo *I* e outra, ao poder dos testes.

Nota-se no decorrer desta seção que os resultados referentes ao erro tipo *I* e ao poder dos testes TBPb e TSW não foram apresentados para a combinação $n = 10000$ e $p = 5$. Recorda-se que no caso do TBPb, o tempo de execução das simulações de validação é demasiadamente longo. E no caso do TSW, a aplicação do teste é restrita às amostras de tamanho $n = 3$ até $n = 5000$.

4.1 Erro tipo *I*

Os erros tipo *I* e tipo *II* têm probabilidades associadas inversamente proporcionais e a probabilidade complementar do erro tipo *II* é o poder do teste de hipótese. Desse modo, se a probabilidade de se incorrer no erro tipo *I* for diminuída, a probabilidade de se cometer o erro tipo *II* aumentará. Quando as taxas de erro tipo *I* observadas não são significativamente diferentes do nível nominal de significância fixado, o teste é considerado exato. Em contrapartida, se essas taxas são inferiores ao nível de significância nominal fixado, o teste é tido como conservativo. Finalmente, se as taxas são superiores ao nível nominal de significância, o teste é liberal. Para classificar os testes dessa forma, foi utilizado o teste binomial exato com nível de significância de 1%.

As Tabelas 3, 4 e 5 se referem ao erro tipo *I* dos testes TKS, TBPb, TEM e TSW para os níveis de significância nominais de 10%, 5% e 1%, respectivamente. As combinações de n e p utilizadas foram as apresentadas na Tabela 2, da Subseção 3.2.1.

Na Tabela 3 as taxas de erro tipo *I* para o nível de significância nominal de 10%, considerando a distribuição normal multivariada, são mostradas. Foi verificado se os valores estimados diferem significativamente (valor-$p < 0{,}01$) do nível nominal de 10%.

Os testes TKS e TBPb apresentaram taxas de erro tipo *I* não significativamente (valor-$p > 0{,}01$) diferentes do valor nominal $\alpha = 10\%$ para todas as combinações de n e p. Desse modo, os dois testes propostos neste trabalho podem ser considerados exatos ao nível de significância de 10%. O TSW se mostrou exato para a maioria das combinações de n e p, exceto, para as situações $n = 100$ com $p = 50$ e $p = 97$, e $n = 200$ com $p = 100$, em que o teste foi conservativo, sendo as taxas de errro tipo *I* significativamente (valor-$p < 0{,}01$) menores do que o valor nominal $\alpha = 10\%$.

No teste TEM, as taxas de erro tipo *I* para todas as combinações de n e p foram consideradas abaixo do nível de significância nominal, portanto, o teste foi conservativo para $\alpha = 10\%$.

Comparando os resultados dos testes com os de outros testes de normalidade multivariada existentes na literatura, observa-se que ambos os testes propostos apresentaram um bom controle da taxa de erro tipo *I*. Cirillo e Ferreira (2003) mostraram que o teste de normalidade multivariada baseado no coeficiente de correlação Quantil-Quantil apresentou um bom controle da taxa de erro tipo *I* para

Tabela 3 Erro tipo *I* dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal multivariada em função de n e p para $\alpha = 10\%$

		10%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,1075	0,1055	0,0225 $^{-}$	0,0965
	5	0,0955	0,1085	0,0015 $^{-}$	0,0945
	7	0,1085	0,1060	0,0085 $^{-}$	0,0865
20	2	0,1020	0,0970	0,0275 $^{-}$	0,0990
	5	0,1055	0,1040	0,0125 $^{-}$	0,0970
	10	0,0925	0,1030	0,0035 $^{-}$	0,0870
	17	0.0995	0,0985	0,0175 $^{-}$	0,0855
50	2	0,1070	0,1140	0,0310 $^{-}$	0,1020
	10	0,0945	0,1025	0,0090 $^{-}$	0,0890
	25	0,1060	0,1015	0,0090 $^{-}$	0,0945
	47	0,1030	0,1015	0,0195 $^{-}$	0,0855
100	2	0,1045	0,1030	0,0220 $^{-}$	0,0980
	10	0,0990	0,0960	0,0110 $^{-}$	0,0970
	50	0,0985	0,1080	0,0090 $^{-}$	0,0680 $^{-}$
	97	0,0955	0,1010	0,0240 $^{-}$	0,0715 $^{-}$
200	2	0,0905	0,0940	0,0360 $^{-}$	0,0935
	10	0,0995	0,1005	0,0110 $^{-}$	0,0975
	50	0,1045	0,0940	0,0045 $^{-}$	0,0865
	100	0,1050	0,1085	0,0050 $^{-}$	0,0690 $^{-}$
10000	5	0,1045	-	0,0170 $^{-}$	-

$^{+}$ Taxas de erro tipo *I* superiores ao valor nominal de 10% de significância (valor-$p < 0{,}01$).

$^{-}$ Taxas de erro tipo *I* inferiores ao valor nominal de 10% de significância (valor-$p < 0{,}01$).

amostra de tamanho $n = 100$ com $p = 2$ e $p = 10$, que foram de $0{,}1038$ e $0{,}0965$, respectivamente. O teste de normalidade multivariada de simetria, de acordo com Cirillo e Ferreira (2003), apresentou o valor da taxa de erro tipo *I* de $0{,}0665$, inferior ao nível de significância nominal, para $n = 100$ com $p = 10$. E no teste de normalidade multivariada de curtose, a taxa de erro tipo *I* foi de $0{,}0680$ para

$n = 100$ com $p = 2$, abaixo dos outros dois testes que Cirillo e Ferreira (2003) avaliaram. Deve-se ressaltar que Cirillo e Ferreira (2003) realizaram 10.000 simulações Monte Carlo para a validação, enquanto que neste trabalho foram feitas 2.000 simulações Monte Carlo.

Na Tabela 4, têm-se as taxas de erro tipo *I* dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal multivariada em função das combinações de n e p para $\alpha = 5\%$. O teste TKS apresentou taxas de erro tipo *I* não significativamente (valor-$p > 0{,}01$) diferentes do nível de significância nominal de 5% para a maior parte das combinações de n e p, com exceção da situação $n = 10000$ com $p = 5$, em que o teste foi conservativo. O teste TBPb obteve controle da taxa de erro tipo *I* para todas as combinações de n e p em que foi submetido, mostrando-se um teste exato, excluindo-se da afirmação apenas a situação de $n = 10000$ e $p = 5$, para a qual o resultado não foi obtido.

Ainda na Tabela 4, o TEM se mostrou conservativo em todas as combinações de n e p para $\alpha = 5\%$. Enquanto que o teste TSW mostrou controle da taxa de erro tipo *I* na maioria das situações, a não ser nas combinações $n = 100$ com $p = 50$ e $n = 200$ com $p = 100$, em que a taxa de erro tipo *I* foi significativamente (valor-$p < 0{,}01$) inferior que o nível nominal de 5%, sendo conservativo nesses casos.

Mecklin e Mundfrom (2005) realizaram a comparação das taxas de erro tipo *I* por simulações Monte Carlo para os seguintes testes de normalidade multivariada: teste de assimetria de Mardia (1970), teste de curtose de Mardia (1970), teste de ajuste de Hawkins (1981), teste de Koziol (1982), teste de assimetria e curtose de Mardia e Foster (1983), teste de Royston (ROYSTON, 1983), teste de Paulson, Roohan e Sullo (1987), teste de Henze e Zirkler (1990), teste de assimetria e curtose de MARDIA e Kent (1991), teste de Romeu e Ozturk (1993), teste de correlação clássico de Singh (1993), teste de correlação robusto de Singh (1993) e teste de Mudholkar, Srivastava e Lin (1995).

Segundo Mecklin e Mundfrom (2005), no nível nominal de 5% de significância, o teste de Mardia-Foster foi o que teve maior variação das taxas de erro

Tabela 4 Erro tipo *I* dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal multivariada em função de n e p para $\alpha = 5\%$

		5%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0530	0,0570	0,0080$^-$	0,0490
	5	0,0450	0,0500	0,0000$^-$	0,0470
	7	0,0555	0,0555	0,0010$^-$	0,0400
20	2	0,0525	0,0530	0,0095$^-$	0,0500
	5	0,0530	0,0515	0,0045$^-$	0,0530
	10	0,0465	0,0485	0,0000$^-$	0,0445
	17	0,0445	0,0520	0,0040$^-$	0,0390
50	2	0,0635	0,0560	0,0075$^-$	0,0505
	10	0,0510	0,0495	0,0020$^-$	0,0410
	25	0,0530	0,0550	0,0015$^-$	0,0550
	47	0,0530	0,0510	0,0080$^-$	0,0425
100	2	0,0555	0,0560	0,0110$^-$	0,0500
	10	0,0515	0,0505	0,0020$^-$	0,0490
	50	0,0445	0,0515	0,0025$^-$	0,0365$^-$
	97	0,0510	0,0545	0,0080$^-$	0,0410
200	2	0,0460	0,0455	0,0115$^-$	0,0490
	10	0,0495	0,0450	0,0010$^-$	0,0490
	50	0,0465	0,0530	0,0005$^-$	0,0475
	100	0,0490	0,0540	0,0005$^-$	0,0380$^-$
10000	5	0,0005$^-$	-	0,0040$^-$	-

$^+$ Taxas de erro tipo *I* superiores ao valor nominal de 5% de significância (valor-$p < 0{,}01$).

$^-$ Taxas de erro tipo *I* inferiores ao valor nominal de 5% de significância (valor-$p < 0{,}01$).

tipo *I*, que oscilaram de 0 a 1. O teste de Hawkins apresentou taxa de erro tipo *I* de 1 para $n = 25$ com $p = 5$. Os testes de Singh robusto, Mudholkar–Srivastava–Lin, Romeu-Ozturl e Mardia-Kent apresentaram valores de taxas de erro tipo *I* superiores ao nível de significância nominal de 5%, caracterizando-se liberais. Os testes de Koziol, de assimetria de Mardia e de Singh clássico foram um pouco liberais em todas as situações. Por outro lado, os testes de Henze-Zirkler, de curtose de Mardia e de Paulson–Roohan–Sullo foram um pouco conservativos na maioria das situações. O teste de Royston (1983) foi o que apresentou o melhor controle das taxas de erro tipo *I*, que variaram de 0,0470 a 0,0530.

O teste de normalidade multivariado de curtose avaliado por Cirillo e Ferreira (2003) foi conservativo ao nível nominal de 5%. Já o teste de normalidade multivariada baseado no coeficiente de correlação Quantil-Quantil e o teste de normalidade multivariado de simetria comparados por Cirillo e Ferreira (2003) foram exatos. O teste de normalidade Shapiro-Francia multivariado que Silva (2009) propôs e avaliou também foi exato.

Ressalta-se que os autores Mecklin e Mundfrom (2005), Cirillo e Ferreira (2003) e Silva (2009) realizaram 10.000 simulações Monte Carlo, enquanto a validação realizada neste trabalho utilizou 2.000 simulações. Diante da Tabela 4, é notável que os dois testes propostos foram exatos ao nível nominal de 5% de significância e, comparando com as avaliações que esses autores fizeram, ambos apresentaram controle da taxa de erro tipo *I* melhor do que alguns dos testes de normalidade multivariada avaliados por esses autores.

Na Tabela 5, são apresentadas as taxas de erro tipo *I* dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal multivariada em função das combinações de n e p para $\alpha = 1\%$.

O teste TBPb obteve o controle da taxa de erro tipo *I* em todas as situações em que os resultados foram obtidos, independente do tamanho da amostra (n) e do número de variáveis (p), salvo $n = 10000$ com $p = 5$, em que não se obteve o resultado. Esse controle das taxas de erro tipo *I* pode ser explicado em virtude delas não terem sido significativamente (valor-$p > 0{,}01$) diferentes do nível nominal de significância de 1%, sendo, por isso, um teste considerado exato. O mesmo vale para o teste TKS, que também apresentou controle da taxa de erro tipo *I*, com exclusão da combinação $n = 10000$ com $p = 5$, na qual a taxa observada foi significativamente inferior ao nível nominal, caracterizando o teste como conservativo nessa situação.

Tabela 5 Erro tipo *I* dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal multivariada em função de n e p para $\alpha = 1\%$

		1%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0120	0,0130	0,0005$^-$	0,0110
	5	0,0065	0,0105	0,0000$^-$	0,0130
	7	0,0065	0,0120	0,0000$^-$	0,0070
20	2	0,0100	0,0100	0,0015$^-$	0,0085
	5	0,0110	0,0085	0,0000$^-$	0,0115
	10	0,0075	0,0120	0,0000$^-$	0,0085
	17	0,0095	0,0125	0,0000$^-$	0,0040$^-$
50	2	0,0115	0,0125	0,0005$^-$	0,0095
	10	0,0145	0,0130	0,0000$^-$	0,0070
	25	0,0080	0,0070	0,0000$^-$	0,0140
	47	0,0080	0,0090	0,0005$^-$	0,0120
100	2	0,0105	0,0085	0,0005$^-$	0,0090
	10	0,0130	0,0085	0,0000$^-$	0,0080
	50	0,0075	0,0100	0,0000$^-$	0,0150
	97	0,0120	0,0150	0,0000$^-$	0,0160
200	2	0,0080	0,0090	0,0005$^-$	0,0115
	10	0,0090	0,0085	0,0000$^-$	0,0125
	50	0,0100	0,0165	0,0000$^-$	0,0155
	100	0,0120	0,011	0,0000$^-$	0,0130
10000	5	0,0000$^-$	-	0,0000$^-$	-

$^+$ Taxas de erro tipo *I* superiores ao valor nominal de 1% de significância (valor-$p < 0{,}01$).

$^-$ Taxas de erro tipo *I* inferiores ao valor nominal de 1% de significância (valor-$p < 0{,}01$).

O TSW foi considerado exato para a maioria das combinações de n e p, exceto para $n = 20$ com $p = 17$, em que a taxa de erro tipo *I* foi significativamente (valor-$p < 0{,}01$) inferior ao nível de significância nominal de 1%, portanto, conservativo nesse caso. Já o teste TEM não obteve controle da taxa de erro tipo *I* em nenhuma situação, pois os valores observados foram significativamente inferior ao nível nominal de 1%, independentemente do tamanho da amostra e ao número de variáveis, sendo considerado conservativo.

O teste de coeficiente de correlação Quantil-Quantil multivariado, o teste de normalidade multivariada de simetria e o teste de normalidade multivariada de curtose, avaliados por Cirillo e Ferreira (2003), foram exatos, somente para $n = 100$ com $p = 2$ e $p = 10$, ao nível de significância nominal de 1%. Já o teste de Shapiro-Francia multivariado proposto por Silva (2009) foi exato em todas as combinações de n e p avaliadas, bem como o TKS e o TBPb.

É importante destacar que os dois testes propostos neste trabalho, o TNM-KS e o TBPb, obtiveram controle da taxa de erro tipo *I* em todas as combinações de n e p para $\alpha = 10\%$, $\alpha = 5\%$ e $\alpha = 1\%$, com exceção somente do caso em que $n = 10000$ com $p = 5$, ao nível nominal de 5% e de 1%, em que o TKS foi conservativo. Isso permite considerar que, predominantemente, os novos testes foram exatos. Bem como o TSW, que também pode ser considerado exato. Por outro lado, o TEM foi um teste conservativo na maioria das combinações.

4.2 Poder dos testes

Para as estimativas do poder dos testes TKS, TBPb, TEM e TSW foram consideradas as distribuições não normais multivariadas: *t-Student* multivariada com $\nu = 1$ e $\nu = 30$ graus de liberdade, a distribuição log-normal multivariada, a distribuição uniforme multivariada e a distribuição normal contaminada multivariada. Os níveis de significância nominais foram os mesmos estabelecidos para as estimativas das taxas de erro tipo *I*, ou seja, $\alpha = 10\%$, $\alpha = 5\%$ e $\alpha = 1\%$. E as combinações de n e p foram aquelas apresentadas na Subseção 3.2.1.

Os valores de poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade em função de n e p para $\alpha = 10\%$, são apresentados na Tabela 6.

Os testes TKS e TBPb apresentaram os piores valores de poder nas situa-

Tabela 6 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade em função de n e p para $\alpha = 10\%$

n	p	10%			
		TKS	TBPb	TEM	TSW
10	2	0,1760	0,1495	0,5485	0,7455
	5	0,2610	0,2445	0,1415	0,8620
	7	0,2400	0,2830	0,0005	0,8720
20	2	0,2535	0,2495	0,9490	0,9570
	5	0,4580	0,4610	0,9755	0,9895
	10	0,4935	0,5885	0,8340	0,9965
	17	0,2550	0,5035	0,0005	0,9980
50	2	0,4130	0,4505	1,0000	0,9995
	10	0,9380	0,9605	1,0000	1,0000
	25	0,9775	0,9955	1,0000	1,0000
	47	0,2665	0,0960	0,0005	1,0000
100	2	0,5440	0,6245	1,0000	1,0000
	10	0,9935	0,9985	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	97	0,2775	-	0,0015	1,0000
200	2	0,6155	0,7505	1,0000	1,0000
	10	0,9985	1,0000	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	100	1,0000	-	1,0000	1,0000
10000	5	0,9885	-	1,0000	-

ções em que a diferença entre o tamanho da amostra, n, e o número de variáveis, p, foi maior e quando p se aproximou de n. Deve-se ressaltar a restrição do TKS a $n > p + 2$ e, do TBPb a $n > p + 1$. O TKS apresentou melhor desempenho nas combinações $n = 50$ com $p = 10$ e $p = 25$, $n = 100$ com $p = 10$ e $p = 50$, $n = 200$ com $p = 10$, $p = 50$ e $p = 100$ e $n = 10000$ com $p = 5$, mas o desempenho ainda foi inferior aos testes TEM e TSW. O mesmo se verificou para o TBPb, com a exceção das combinações $n = 100$ com $p = 97$ e $n = 200$ com $p = 100$ em que o programa R retornou um erro de inversão da matriz $\boldsymbol{S}_k$, pois em uma das N simulações Monte Carlo essa matriz foi singular. Também, com exceção de $n = 10000$ e $p = 5$, cujo resultado demanda tempo demasiado.

O TEM apresentou valores de poder inferiores aos do TSW na maioria das situações, tendo o pior desempenho quando o número de variáveis se aproximou do tamanho da amostra, casos em que o poder foi inferior ao nível de significância nominal. O TSW foi o que mostrou melhor desempenho de poder em relação aos demais testes.

Nas Tabelas 7 e 8 são mostrados os desempenhos para o poder dos TNM-KS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade para $\alpha = 5\%$ e $\alpha = 1\%$, respectivamente, em função de n e p. Notou-se comportamentos semelhantes aos descritos para $\alpha = 10\%$.

O teste qui-quadrado de Pearson multivariado e o teste conjunto de assimetria e curtose avaliados por Oliveira e Ferreira (2009) apresentaram bom desempenho de poder considerando a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade para $\alpha = 5\%$ e $\alpha = 1\%$. O teste de normalidade multivariada baseado em *bootstrap* paramétrico e o teste Monte Carlo de normalidade multivariada baseado em distâncias, propostos e avaliados por Biase (2011), foram superiores, em desempenho de poder, aos testes avaliados por Oliveira e Ferreira (2009) e aos testes apresentados nas Tabelas 7 e 8.

Deve-se salientar que a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade é uma distribuição bem distinta da normal multivariada, o que faz com que os valores de poder sejam elevados de modo geral em relação a outras distribuições não normais multivariadas não tão distintas da normal multivariada. Entretanto, o desempenho dos testes propostos, mesmo em uma situação tão favorável, foi muito fraco, em relação ao desempenho de seus concorrentes.

A semelhança com a normal multivariada é o caso, por exemplo, da distribuição *t-Student* com $\nu = 30$ graus de liberdade, cujos resultados dos valores de poder dos testes são exibidos nas Tabelas 9, 10 e 11.

Tabela 7 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade em função de n e p para $\alpha = 5\%$

		5%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,1055	0,0860	0,4095	0,6935
	5	0,1740	0,1560	0,0420	0,8165
	7	0,1715	0,1940	0,0000	0,8485
20	2	0,1815	0,1660	0,9130	0,9415
	5	0,3700	0,3610	0,9425	0,9845
	10	0,3845	0,4865	0,6640	0,9935
	17	0,2265	0,4090	0,0000	0,9965
50	2	0,3385	0,3620	1,0000	0,9995
	10	0,9015	0,9405	1,0000	1,0000
	25	0,9465	0,9865	1,0000	1,0000
	47	0,2410	0,0445	0,0000	1,0000
100	2	0,4810	0,5485	1,0000	1,0000
	10	0,9875	0,9980	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	97	0,2630	-	0,0000	1,0000
200	2	0,5820	0,6860	1,0000	1,0000
	10	0,9985	1,0000	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	100	1,0000	-	1,0000	1,0000
10000	5	0,9880	-	1,0000	-

Os desempenhos de poder para os TKS, TBPb, TEM e TSWNM, considerando a distribuição *t-Student* multivariada com $\nu = 30$ graus de liberdade, para $\alpha = 10\%$ em função de n e p, são indicados na Tabela 9. O TKS e o TBPb apresentaram valores de poder próximos ao do nível de significância nominal de 10% na maioria das combinações. O que significa que ambos os testes propostos não conseguiram distinguir a distribuição normal multivariada de uma distribuição

Tabela 8 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 1$ grau de liberdade em função de n e p para $\alpha = 1\%$

		1%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0365	0,0265	0,2030	0,5905
	5	0,0635	0,0460	0,0000	0,7380
	7	0,0735	0,0770	0,0000	0,7795
20	2	0,0780	0,0715	0,8080	0,9095
	5	0,2075	0,2085	0,8070	0,9720
	10	0,2070	0,3120	0,2445	0,9860
	17	0,1415	0,2510	0,0000	0,9930
50	2	0,2295	0,2400	0,9995	0,9990
	10	0,7910	0,8870	1,0000	1,0000
	25	0,8695	0,9690	1,0000	1,0000
	47	0,1975	0,0125	0,0000	1,0000
100	2	0,3735	0,4120	1,0000	1,0000
	10	0,9695	0,9895	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	97	0,2300	-	0,0000	1,0000
200	2	0,5025	0,5795	1,0000	1,0000
	10	0,9975	0,9995	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	100	1,0000	-	1,0000	1,0000
10000	5	0,9875	-	1,0000	-

muito semelhante.

O TEM apresentou valores de poder inferiores ao nível de significância nominal na maioria das situações quando $n < 200$. O teste mostrou um bom desempenho para o poder somente em tamanhos grandes de amostra, nas combinações $n = 200$ com $p = 50$ e $p = 100$ e $n = 10000$ com $p = 5$.

Tabela 9 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 30$ graus de liberdade em função de n e p para $\alpha = 10\%$

		10%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0925	0,1005	0,0220	0,1105
	5	0,1020	0,1065	0,0070	0,1040
	7	0,0975	0,1070	0,0115	0,0960
20	2	0,1035	0,1040	0,0415	0,1300
	5	0,1020	0,1020	0,0235	0,1275
	10	0,1135	0,1100	0,0070	0,1295
	17	0,1085	0,1165	0,0160	0,1320
50	2	0,1085	0,0955	0,0530	0,1435
	10	0,1100	0,1000	0,0860	0,1715
	25	0,1340	0,1200	0,0465	0,2045
	47	0,1250	0,1075	0,0155	0,2245
100	2	0,1070	0,1150	0,0680	0,1455
	10	0,1155	0,1145	0,2040	0,2275
	50	0,1650	0,1200	0,4415	0,3245
	97	0,1285	0,1615	0,0115	0,3750
200	2	0,0980	0,1025	0,0935	0,1930
	10	0,1200	0,1060	0,4715	0,3040
	50	0,1810	0,1520	0,9975	0,4435
	100	0,3390	0,2325	1,0000	0,4870
10000	5	0,1070	-	1,0000	-

Para a distribuição *t-Student* multivariada com $\nu = 30$ graus de liberdade, o TSW não apresentou bom desempenho de poder. Inclusive, para as situações em que o tamanho da amostra foi pequeno, os valores de poder se aproximaram do nível nominal de significância de 10%.

O teste de normalidade multivariada baseado em *bootstrap* paramétrico e o teste Monte Carlo de normalidade multivariada baseado em distâncias, propostos e avaliados por Biase (2011), apresentaram valores de poder superiores ao nível nominal de 10% para a maioria das combinações de n e p. No entanto, o desempenho de poder ainda não foi satisfatório, embora o teste de normalidade multivariada baseado em *bootstrap* paramétrico de Biase (2011) tenha apresentado valores de poder superiores ao TSW.

Nas Tabelas 10 e 11 são mostrados os desempenhos para o poder dos TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 30$ graus de liberdade para $\alpha = 5\%$ e $\alpha = 1\%$, respectivamente, em função de n e p. O comportamento dos resultados se assemelhou ao que foi discutido perante o nível de significância de 10%.

Destaca-se que o único teste que atingiu poder de 100% foi o TEM, paras as situações em que o tamanho da amostra foi grande, quando $n = 10000$ e $p = 5$ nos níveis de significância de 10%, 5% e 1% e quando $n = 200$ e $p = 100$ no nível de significância de 10%. Nessas ocasiões, o TEM obteve superioridade relevante em relação aos demais.

Os testes propostos por Biase (2011) apresentaram melhores valores de poder do que o teste conjunto de simetria e curtose multivariado e o teste qui-quadrado de Pearson multivariado, avaliados por Oliveira e Ferreira (2009), para as situações de $n = 10$ e $n = 20$. O teste de normalidade multivariada baseado em *bootstrap* paramétrico foi superior em desempenho de poder na maioria das combinações de n e p em relação aos testes avaliados por Oliveira e Ferreira (2009) e aos testes TKS, TBPb e TSW.

Na Tabela 12 são apresentados os poderes dos testes TKS, TBPb, TEM, TSW, considerando a distribuição lognormal multivariada em função das combinações de n e p para um nível de significância nominal $\alpha = 10\%$.

O teste de TEM apresentou poder nulo para as combinações em que n e p foram próximos. Por outro lado, o teste apresentou poder elevado para as combinações em que $n \geq 50$ e poder baixo para $n \leq 20$. Os testes TKS e TBPb mostraram poder baixo para quase todas as situações, exceto para as combinações

Tabela 10 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 30$ graus de liberdade em função de n e p para $\alpha = 5\%$

		5%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0545	0,0470	0,0075	0,0625
	5	0,0470	0,0585	0,0010	0,0580
	7	0,0495	0,0525	0,0020	0,0555
20	2	0,0535	0,0485	0,0110	0,0775
	5	0,0525	0,0510	0,0080	0,0685
	10	0,0540	0,0545	0,0005	0,0705
	17	0,0595	0,0585	0,0035	0,0680
50	2	0,0545	0,0525	0,0160	0,0875
	10	0,0605	0,0450	0,0290	0,1115
	25	0,0735	0,0600	0,0145	0,1425
	47	0,0765	0,0505	0,0055	0,1630
100	2	0,0550	0,0640	0,0280	0,0865
	10	0,0650	0,0580	0,0840	0,1565
	50	0,0940	0,0620	0,2185	0,2300
	97	0,0775	0,1085	0,0025	0,2885
200	2	0,0530	0,0485	0,0410	0,1280
	10	0,0660	0,0560	0,2820	0,2090
	50	0,1060	0,0900	0,9915	0,3505
	100	0,2025	0,1545	0,9990	0,3945
10000	5	0,0590	-	1,0000	-

de $n = 100$ com $p = 50$ e de $n = 200$ com $p = 50$ e $p = 100$. Inclusive, foi somente nos casos em que n e p são próximos, que ambos os testes apresentaram um desempenho melhor que o teste de TEM. Já o teste TSW apresentou poder elevado para todas as combinações de n e p, sendo, portanto, o teste que teve o melhor desempenho, considerando a distribuição lognormal multivariada e $\alpha = 10\%$.

Tabela 11 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição *t-Student* multivariada com $\nu = 30$ graus de liberdade em função de n e p para $\alpha = 1\%$

		1%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0145	0,0065	0,0005	0,0205
	5	0,0110	0,0105	0,0000	0,0165
	7	0,0085	0,0120	0,0000	0,0125
20	2	0,0095	0,0080	0,0015	0,0155
	5	0,0135	0,0120	0,0005	0,0195
	10	0,0160	0,0120	0,0000	0,0225
	17	0,0120	0,0090	0,0000	0,0205
50	2	0,0130	0,0090	0,0025	0,0275
	10	0,0150	0,0120	0,0015	0,0420
	25	0,0160	0,0140	0,0000	0,0585
	47	0,0215	0,0140	0,0005	0,0830
100	2	0,0130	0,0185	0,0030	0,0345
	10	0,0160	0,0110	0,0105	0,0620
	50	0,0245	0,0150	0,0250	0,1265
	97	0,0275	0,0340	0,0000	0,1715
200	2	0,0075	0,0110	0,0050	0,0445
	10	0,0165	0,0160	0,0605	0,0915
	50	0,0265	0,0220	0,8980	0,2040
	100	0,0730	0,0415	0,9415	0,2535
10000	5	0,0175	-	1,0000	-

Nas Tabelas 13 e 14 são apresentados os desempenhos para o poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição lognormal multivariada para $\alpha = 5\%$ e $\alpha = 1\%$, respectivamente, em função de n e p. O comportamento dos resultados foi parecido ao que foi discutido no nível de significância de 10%.

Tabela 12 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição log-normal multivariada em função de n e p para $\alpha = 10\%$

		10%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,1055	0,1070	0,3245	0,8635
	5	0,1545	0,1515	0,1105	0,9830
	7	0,2020	0,2030	0,0000	0,9935
20	2	0,1180	0,1980	0,7795	0,9990
	5	0,2085	0,2110	0,9095	1,0000
	10	0,2995	0,3260	0,7530	1,0000
	17	0,2405	0,3770	0,0005	1,0000
50	2	0,1605	0,3595	0,9975	1,0000
	10	0,4520	0,5015	1,0000	1,0000
	25	0,8245	0,8875	1,0000	1,0000
	47	0,2060	0,3255	0,0000	1,0000
100	2	0,1830	0,4900	1,0000	1,0000
	10	0,5105	0,7235	1,0000	1,0000
	50	0,9975	1,0000	1,0000	1,0000
	97	0,2345	0,7325	0,0005	1,0000
200	2	0,2495	0,6975	1,0000	1,0000
	10	0,5745	0,9610	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	100	1,0000	1,0000	1,0000	1,0000
10000	5	0,4640	-	1,0000	-

Os testes TKS e TBPb obtiveram valores de poder reduzidos para a maioria das combinações de n e p, exceto para as situações de $n = 100$ com $p = 50$ e de $n = 200$ com $p = 50$ e $p = 100$, sendo que, de uma maneira geral, o teste TBPb teve desempenho minimamente melhor que o TKS. No entanto, os testes só foram mais poderosos que o TEM para as combinações em que n e p são próximos, circunstância em que TEM tem poder igual a zero. Excetuando os casos que n e p são próximos, o teste TEM apresentou baixo poder para $n \leq 20$ e alto poder

Tabela 13 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição log-normal multivariada em função de n e p para $\alpha = 5\%$

		5%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0550	0,0575	0,2165	0,7980
	5	0,0840	0,0860	0,0295	0,9685
	7	0,1215	0,1175	0,0000	0,9870
20	2	0,0715	0,1270	0,6805	0,9920
	5	0,1400	0,1300	0,8395	1,0000
	10	0,2115	0,2275	0,5505	1,0000
	17	0,1965	0,2900	0,0000	1,0000
50	2	0,1060	0,2645	0,9920	1,0000
	10	0,3495	0,3800	1,0000	1,0000
	25	0,7415	0,8185	1,0000	1,0000
	47	0,1880	0,2330	0,0000	1,0000
100	2	0,1330	0,3955	1,0000	1,0000
	10	0,4035	0,6135	1,0000	1,0000
	50	0,9970	1,0000	1,0000	1,0000
	97	0,2070	0,6810	0,0005	1,0000
200	2	0,2045	0,5915	1,0000	1,0000
	10	0,4720	0,9210	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	100	1,0000	1,0000	1,0000	1,0000
10000	5	0,3385	-	1,0000	-

(valores próximos ou iguais a 100%) para $n \geq 50$. Enquanto que o teste TSW apresentou um poder elevado para todas as combinações de n e p, exceto para a combinação $n = 10$ e $p = 2$.

Nas comparações de poder realizadas por Mecklin e Mundfrom (2005), para o nível de significância nominal de 5%, considerando a distribuição log-normal multivariada, o teste de curtose de Mardia (1970) apresentou o menor

Tabela 14 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição log-normal multivariada em função de n e p para $\alpha = 1\%$

		1%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0130	0,0100	0,0790	0,6290
	5	0,0275	0,0250	0,0010	0,9090
	7	0,0320	0,0370	0,0000	0,9595
20	2	0,0220	0,0460	0,4705	0,9710
	5	0,0555	0,0460	0,5935	1,0000
	10	0,0855	0,0965	0,1675	1,0000
	17	0,1095	0,1545	0,0000	1,0000
50	2	0,0455	0,1360	0,9730	1,0000
	10	0,1900	0,2160	1,0000	1,0000
	25	0,5420	0,6350	0,9995	1,0000
	47	0,1505	0,1085	0,0000	1,0000
100	2	0,0690	0,2305	1,0000	1,0000
	10	0,2260	0,3940	1,0000	1,0000
	50	0,9835	0,9995	1,0000	1,0000
	97	0,1700	0,5660	0,0000	1,0000
200	2	0,1210	0,4215	1,0000	1,0000
	10	0,2725	0,7810	1,0000	1,0000
	50	1,0000	1,0000	1,0000	1,0000
	100	1,0000	1,0000	1,0000	1,0000
10000	5	0,1645	-	1,0000	-

valor de poder, de 0,8250. No teste de Koziol (1982), o menor valor de poder foi de 0,8260. O teste de assimetria de Mardia (1970), o teste de Royston (1983) e o teste de Henze e Zirkler (1990) obtiveram valor de poder de pelo menos 0,9800 em todas as situações avaliadas.

O teste de Shapiro-Wilk generalizado para o caso multivariado por Silva (2009) mostrou resultados de poder similares ao teste de normalidade multivariado baseado em *bootstrap* paramétrico proposto por Biase (2011) para $n = 20$ com $p = 2$, $n = 100$ e $n = 200$ com $p = 2$ e $p = 10$. Esse teste de Biase (2011) teve melhor desempenho do que o teste qui-quadrado de Pearson multivariado e o teste conjunto de assimetria e curtose multivariado avaliados por Oliveira e Ferreira (2009), para amostras pequenas.

Na Tabela 15 são representados os valores de poder dos testes TNMK, TBPb, TEM e TSW, para a distribuição uniforme multivariada em função das combinações de n e p para o nível de significância $\alpha = 10\%$. De uma maneira geral, pode-se observar que os três testes baseados em amostras betas apresentaram um poder muito baixo, com destaque negativo para o teste TEM, que para a maioria das combinações teve poder muito próximo de zero, salvo para as combinações de $n = 200$ com $p = 50$ e $p = 100$ e de $n = 10000$ e $p = 5$ em que o poder foi satisfatório.

Não muito diferente, os testes TKS e TBPb também apresentaram poder muito aquém do esperado, com uma ligeira vantagem para TBPb que mostrou, na maioria das combinações, poder maior que o de TKS. Diferentemente, o teste TSW foi muito superior que os demais testes e apresentou maior poder em todas as combinações de n e p, com poder igual a 1 a partir das combinações de $n = 50$ com $p = 10$. No entanto, seu desempenho não foi satisfatório para as combinações em que n é pequeno.

Nas Tabelas 16 e 17 são dispostos os valores de poder dos testes TNMK, TBPb, TEM e TSW, considerando a distribuição uniforme multivariada em função das combinações de n e p, para o nível de significância $\alpha = 5\%$ e $\alpha = 1\%$, respectivamente. Os resultados para os casos de $\alpha = 5\%$ e $\alpha = 1\%$ se assemelham aos discutidos sobre o poder dos testes para $\alpha = 10\%$, apresentados na Tabela 15

Os testes TKS, TBPb e TEM se mostraram com poderes muito baixo para a maioria das combinações de n e p e o TSW, superior aos demais, com poder de 100% na maioria das situações, a partir da combinação de $n = 50$ com $p = 10$ e valores de poder menores, ainda que superior aos demais, quando $n \leq 20$. O poder dos testes apresentou pior desempenho nos níveis de 5% e 1% em comparação com

Tabela 15 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição uniforme multivariada em função de n e p para $\alpha = 10\%$

		10%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0680	0,0860	0,0060	0,1845
	5	0,0720	0,0895	0,0045	0,2615
	7	0,0670	0,0690	0,0160	0,2715
20	2	0,0565	0,0935	0,0010	0,4625
	5	0,0520	0,0920	0,0085	0,6780
	10	0,0830	0,0970	0,0070	0,8465
	17	0,0905	0,0770	0,0155	0,9430
50	2	0,0520	0,0985	0,0125	0,9745
	10	0,0610	0,1675	0,0095	1,0000
	25	0,0945	0,1590	0,0190	1,0000
	47	0,0940	0,0600	0,0130	1,0000
100	2	0,0510	0,1455	0,0990	1,0000
	10	0,0370	0,3410	0,0075	1,0000
	50	0,1050	0,2310	0,3465	1,0000
	97	0,0970	0,1535	0,0025	1,0000
200	2	0,0560	0,2365	0,6740	1,0000
	10	0,0215	0,7185	0,0050	1,0000
	50	0,0890	0,6955	0,9735	1,0000
	100	0,1675	0,5155	0,9995	1,0000
10000	5	0,9990	-	1,0000	-

os resultados do nível de significância nominal de 10%, devido à relação inversa entre erro tipo *I* e erro tipo *II*.

O teste Monte Carlo de normalidade multivariada proposto e avaliado por Biase (2011) mostrou melhor desempenho que o teste qui-quadrado de Pearson multivariado e o teste conjunto de assimetria e curtose multivariado avaliados por Oliveira e Ferreira (2009) para todos os números de variáveis quando $n = 10$ e

Tabela 16 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição uniforme multivariada em função de n e p para $\alpha = 5\%$

		5%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0340	0,0425	0,0010	0,0920
	5	0,0290	0,0410	0,0005	0,1270
	7	0,0255	0,0305	0,0030	0,1480
20	2	0,0290	0,0455	0,0000	0,2840
	5	0,0200	0,0475	0,0015	0,5110
	10	0,0420	0,0425	0,0005	0,7320
	17	0,0410	0,0340	0,0055	0,8730
50	2	0,0185	0,0455	0,0005	0,9285
	10	0,0265	0,0905	0,0015	1,0000
	25	0,0475	0,0850	0,0020	1,0000
	47	0,0540	0,0265	0,030	1,0000
100	2	0,0215	0,0685	0,0145	1,0000
	10	0,0130	0,2240	0,0015	1,0000
	50	0,0510	0,1385	0,1555	1,0000
	97	0,0655	0,0920	0,0005	1,0000
200	2	0,0235	0,1265	0,3190	1,0000
	10	0,0060	0,5845	0,0005	1,0000
	50	0,0375	0,5650	0,9095	1,0000
	100	0,0860	0,3670	0,9900	1,0000
10000	5	0,9835	-	1,0000	-

$n = 20$ e, para a maioria dos números de variáveis, quando $n = 100$.

Nas avaliações de poder realizadas por Mecklin e Mundfrom (2005), o teste clássico de Singh (1993) teve valor de poder mais baixo de 0,1350, considerando a distribuição uniforme multivariada, no nível nominal de significância de 5%. O teste de Henze e Zirkler (1990) e o teste de Royston (1983) foram relativamente poderosos, mas tiveram valor de poder de 0,4000 quando $n = 25$ com

Tabela 17 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição uniforme multivariada em função de n e p para $\alpha = 1\%$

		1%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0050	0,0080	0,0000	0,0155
	5	0,0050	0,0055	0,0000	0,0255
	7	0,0025	0,0070	0,0000	0,0355
20	2	0,0025	0,0085	0,0000	0,0610
	5	0,0005	0,0070	0,0005	0,2065
	10	0,0060	0,0100	0,0000	0,4200
	17	0,0050	0,0060	0,0005	0,6250
50	2	0,0020	0,0090	0,0000	0,6995
	10	0,0030	0,0230	0,0000	0,9995
	25	0,0110	0,0225	0,0000	1,0000
	47	0,0170	0,0030	0,0000	1,0000
100	2	0,0030	0,0110	0,0000	0,9995
	10	0,0015	0,0675	0,0000	1,0000
	50	0,0070	0,0290	0,0110	1,0000
	97	0,0345	0,0330	0,0000	1,0000
200	2	0,0035	0,0305	0,0190	1,0000
	10	0,0005	0,3030	0,0000	1,0000
	50	0,0045	0,3000	0,5610	1,0000
	100	0,0160	0,1580	0,8360	1,0000
10000	5	0,7785	-	1,0000	-

$p = 5$, enquanto no teste de curtose de Mardia (1970) os valores do poder variaram de 0,8900 a 1,0000 e os testes de Koziol (1982) e Paulson, Roohan e Sullo (1987) apresentaram valor de poder mínimo de 0,9980.

Na Tabela 18 são apresentados os valores de poder dos testes TNMKS, TNMBPb, TEM e TSW no nível de significância nominal $\alpha = 10\%$ para as combinações de n e p, descritas na Tabela 2, considerando a distribuição normal contaminada multivariada. O teste TKS foi o que apresentou pior desempenho dentre os testes, com poder inferior em todas as combinações de n e p, exceto o caso em que $n = 200$ e $p = 100$. Inclusive, para $n = 10000$ com $p = 5$, o poder do teste foi igual a zero.

Tabela 18 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal contaminada multivariada em função de n e p para $\alpha = 10\%$

		10%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0590	0,1055	0,0215	0,4425
	5	0,0965	0,1155	0,0335	0,5120
	7	0,1065	0,1125	0,0020	0,5450
20	2	0,0335	0,1605	0,0255	0,8140
	5	0,0445	0,1650	0,2155	0,8885
	10	0,1085	0,1300	0,3530	0,9380
	17	0,1510	0,1875	0,0010	0,9530
50	2	0,0175	0,3275	0,0540	0,9990
	10	0,0325	0,4005	0,9205	1,0000
	25	0,2905	0,2530	0,9900	1,0000
	47	0,1150	0,0960	0,0000	1,0000
100	2	0,0130	0,5490	0,1775	1,0000
	10	0,0080	0,8870	0,9960	1,0000
	50	0,7800	0,6475	1,0000	1,0000
	97	0,1055	0,5585	0,0035	1,0000
200	2	0,0040	0,8225	0,5530	1,0000
	10	0,0010	1,0000	1,0000	1,0000
	50	0,5575	0,9600	1,0000	1,0000
	100	0,9995	0,9975	1,0000	1,0000
10000	5	0,0000	-	1,0000	-

O teste TBPb apresentou desempenho um pouco superior em relação ao TKS, obtendo maiores valores de poder na maioria das combinações de n e p. No entanto, o melhor desempenho de poder ocorreu nas situações em que o tamanho da amostra n foi igual a 200.

Já o teste TEM apresentou algumas peculiaridades. Para $n \leq 20$, o poder do teste apresentou valores muito baixos. Para $n > 50$ o teste apresenta menor poder quando $p = 2$ e quando p se aproximou de n. Para os tamanhos da amostra, p, "intermediários", o teste apresentou valores de poder elevados, sendo de 100% ou próximo desse valor na maioria das combinações.

Novamente, o teste TSW foi superior em relação aos demais testes baseados em amostras betas, apresentando um poder elevado para combinações de $n \geq 20$. Mesmo que, para $n = 10$, o poder tenha sido inferior, o desempenho desse teste ainda superou o dos demais, considerando a distribuição uniforme multivariada.

Nas Tabelas 19 e 20 são exibidos os valores de poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal contaminada em função das combinações de n e p, para o nível de significância $\alpha = 5\%$ e $\alpha = 1\%$, respectivamente. Os resultados para os casos de $\alpha = 5\%$ e $\alpha = 1\%$ são muito semelhantes aos discutidos sobre o poder dos testes para $\alpha = 10\%$, apresentados na Tabela 18.

O teste TKS foi o que apresentou pior desempenho dentre os testes, com valores de poderes inferiores para todas as combinações de n e p, exceto para o caso em que $n = 200$ e $p = 100$. O teste TBPb apresentou taxas de poderes maiores do que o TKS, mas os valores foram elevados somente quando $n = 200$. O teste TEM apresentou as mesmas peculiaridades de quando $\alpha = 10\%$. Mais uma vez, o teste TSW obteve melhor desempenho que os demais em todas as combinações de n e p.

Mecklin e Mundfrom (2005), em suas comparações de poder de testes, utilizaram 15 configurações diferentes para a distribuição normal contaminada multivariada e todos os testes avaliados tiveram baixo poder em todas as situações. Isso, de acordo com os autores, se deve ao fato de os vetores de média das distribuições normais contaminadas utilizadas serem iguais ou pouco distintos.

Tabela 19 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal contaminada multivariada em função de n e p para $\alpha = 5\%$

n	p	5%			
		TKS	TBPb	TEM	TSW
10	2	0,0255	0,0570	0,0055	0,2645
	5	0,0455	0,0635	0,0065	0,3330
	7	0,0475	0,0615	0,0005	0,3395
20	2	0,0130	0,0910	0,0075	0,6665
	5	0,0205	0,0955	0,1115	0,7970
	10	0,0565	0,0700	0,1955	0,8770
	17	0,1070	0,1255	0,0005	0,8985
50	2	0,0070	0,2240	0,0160	0,9945
	10	0,0145	0,2835	0,8505	1,0000
	25	0,1805	0,1625	0,9735	1,0000
	47	0,1045	0,0505	0,0000	1,0000
100	2	0,0040	0,4165	0,0590	1,0000
	10	0,0035	0,8155	0,9905	1,0000
	50	0,6865	0,5135	1,0000	1,0000
	97	0,1030	0,5165	0,0005	1,0000
200	2	0,0015	0,7255	0,2860	1,0000
	10	0,0000	1,0000	1,0000	1,0000
	50	0,4200	0,9225	1,0000	1,0000
	100	0,9995	0,9960	1,0000	1,0000
10000	5	0,0000	-	1,0000	-

O teste conjunto de simetria e curtose multivariado avaliado por Oliveira e Ferreira (2009) apresentou baixos valores de poder para amostras de tamanho pequeno e para quando p se aproxima de n o poder foi nulo. O teste qui-quadrado de Pearson multivariado proposto por Oliveira e Ferreira (2009) também apresentou valores de poder reduzidos para amostras pequenas e, quando p se aproximou de n o desempenho de poder também foi afetado, embora tenha sido melhor que do

Tabela 20 Poder dos testes TKS, TBPb, TEM e TSW, considerando a distribuição normal contaminada multivariada em função de n e p para $\alpha = 1\%$

		1%			
n	p	TKS	TBPb	TEM	TSW
10	2	0,0025	0,0145	0,0000	0,0565
	5	0,0060	0,0185	0,0000	0,0765
	7	0,0085	0,0125	0,0000	0,0800
20	2	0,0030	0,0285	0,0010	0,3080
	5	0,0025	0,0235	0,0265	0,4595
	10	0,0165	0,0140	0,0340	0,5690
	17	0,0440	0,0505	0,0000	0,6280
50	2	0,0020	0,0930	0,0010	0,9530
	10	0,0025	0,1205	0,6560	0,9980
	25	0,0695	0,0485	0,8905	1,0000
	47	0,0875	0,0080	0,0000	1,0000
100	2	0,0000	0,2110	0,0040	1,0000
	10	0,0005	0,6225	0,9580	1,0000
	50	0,4270	0,2755	1,0000	1,0000
	97	0,0985	0,4185	0,0000	1,0000
200	2	0,0000	0,5040	0,0365	1,0000
	10	0,0000	0,9970	1,0000	1,0000
	50	0,2240	0,7960	1,0000	1,0000
	100	0,9910	0,9700	1,0000	1,0000
10000	5	0,0000	-	1,0000	-

teste conjunto de simetria e curtose multivariado.

O teste Monte Carlo de normalidade multivariada baseado em distâncias, proposto por Biase (2011), apresentou melhor desempenho de poder do que o teste conjunto de simetria e curtose multivariado avaliado por Oliveira e Ferreira (2009), para amostras $n = 10$ e $n = 20$. E, também, em relação ao teste qui-quadrado de

Pearson multivariado de Oliveira e Ferreira (2009), exceto para $p = 8$ e $p = 18$. Para $n = 200$ com $p = 2$ e $p = 10$ o teste de Biase (2011) apresentou menor valor de poder do que ambos os testes avaliados por Oliveira e Ferreira (2009). Já o teste de normalidade multivariada baseado em *bootstrap* paramétrico proposto por Biase (2011) superou os testes avaliados por Oliveira e Ferreira (2009) e os testes avaliados nas Tabelas 18, 19 e 20, na maioria das situações.

De modo geral, o TKS e o TBPb apresentaram valores de poder baixos, considerando as distribuições não normais utilizadas, sendo que o TBPb apresentou um desempenho minimamente superior ao TKS. Ambos os testes obtiveram o pior desempenho quando p se aproxima de n e o melhor desempenho quando o tamanho da amostra é grande, exceto para as distribuições: normal contaminada multivariada e uniforme multivariada.

Um fator que interferiu nos valores de poder foi o tamanho da amostra efetiva $(n-p-1)$, ou seja, o tamanho da amostra depois de serem feitas as devidas transformações envolvidas no TKS e no TBPb. Nota-se que o pior desempenho de poder do TKS ocorre quando o tamanho da amostra efetiva é muito reduzido (casos em que n é pequeno ou que p se aproxima de n). Embora o TBPb tenha sido menos afetado pelo tamanho da amostra efetiva do que o TKS, o seu desempenho de poder também reduziu em alguns casos de amostra efetiva pequena, como, por exemplo, em $n = 50$ com $p = 47$, cuja amostra efetiva é de tamanho 2.

Outro fator que pode ter prejudicado o desempenho do poder do TBPb é o fato de que a construção desse teste não levou em consideração os parâmetros da distribuição beta. Após as transformações são obtidas $p - 1$ amostras betas, cada uma com sua parametrização. Quando se utiliza o máximo da estatística de Kolmogorov-Smirnov, não se pode garantir que ele será proveniente sempre da amostra cuja distribuição beta tenha os mesmos parâmetros para cada amostra *bootstrap*.

O TEM apresentou melhor desempenho para tamanhos grandes de amostra, exceto quando o p se aproxima de n. Inclusive, para a distribuição *t-Student* com $\nu = 30$ graus de liberdade, quando $n > 200$ o poder foi próximo ou igual a

100%. No entanto, o teste foi muito conservativo e apresentou valores de poder muito pequenos para amostra de tamanho reduzido.

O TSW superou os demais testes em desempenho de poder e foi o teste que mais obteve valores de poder de 100%. Apesar de não ter mostrado poder elevado no caso da distribuição *t-Student* com $\nu = 30$ graus de liberdade, foi o único que apresentou melhores resultados de poder para a distribuição uniforme multivariada. Sendo assim, pode ser considerado o teste mais poderoso nas comparações com o TKS, o TBPb e o TEM.

Embora o TKS e o TBPb tenham obtido um bom controle das taxas de erro tipo *I*, mostrando-se exatos, não é recomendado, em hipótese alguma, o uso desses testes, devido ao fraco desempenho de poder apresentado por eles. Do mesmo modo, não é recomendado, em nenhuma situação, o TEM, pois, além de ter apresentado um fraco desempenho de poder, foi considerado um teste conservativo.

Um teste que apresentou resultados similares ao TKS e ao TNMBPb na validação foi o teste de normalidade multivariada baseado na distribuição *t-Student* proposto por Marques e Melo (2016). Assim como os testes propostos neste trabalho, o teste de Marques e Melo (2016) também se baseou no artigo de Liang, Pan e Yang (2004). Apesar de ter apresentado bom controle da taxa de erro tipo *I*, esse teste mostrou um desempenho de poder muito fraco tal como o TKS e o TBPb.

Comparando com outros testes existentes na literatura, o teste de normalidade multivariada baseado em *bootstrap* paramétrico proposto por Biase (2011) é uma alternativa ao uso do TSW, uma vez que apresentou um bom controle da taxa de erro tipo *I*, mostrando-se um teste exato e teve desempenho de poder equivalente e, em alguns casos, superior ao TSW. Uma importante vantagem dos testes propostos por Biase (2011) é a de que o TSW tem a limitação de poder ser aplicado somente em amostras de tamanho $n = 3$ a $n = 5000$, o que não ocorre com ambos os testes de Biase (2011). Além disso, os dois testes propostos e avaliados por Biase (2011) podem ser aplicados tanto em casos que $n > p$, quanto nas situações em que $n \leq p$. Por fim, os testes propostos por Biase (2011) apresentaram, de modo geral, melhor desempenho de poder do que o teste qui-quadrado de Pearson multivariado e o teste conjunto de simetria e curtose multivariado avaliados por Oliveira e Ferreira (2009).

Por outro lado, temos o estudo realizado por Mecklin e Mundfrom (2005), que recomendou o uso do teste de Henze e Zirkler (1990), de acordo com as avaliações de erro tipo *I* e de poder que realizaram. A recomendação dos autores se baseou no fato de o teste de Henze e Zirkler (1990) possuir uma facilidade de uso comparado aos outros testes e bons resultados nas simulações Monte Carlo, obtendo o melhor desempenho de poder considerando as distribuições não normais testadas. A ressalva foi de que o teste de Henze e Zirkler (1990) não ajuda a diagnosticar a razão da não normalidade, por isso, em caso de rejeição da hipótese de normalidade multivariada, os autores sugerem que a análise seja complementada com procedimentos gráficos e medidas de assimetria e curtose tais como as de Mardia (1970). Quanto ao teste de Royston (1983), Mecklin e Mundfrom (2005) apontaram que, apesar de ter apresentado bom desempenho de poder e controle da taxa de erro tipo *I*, sujeita-se a ter desempenho de poder ruim quando as variáveis são altamente correlacionadas ($\rho > 0{,}9$).

Outro estudo que se pode destacar foi realizado por Farrell, Salibian-Barrera e Naczk (2007), que avaliaram empiricamente o desempenho dos testes de Royston (1983), Royston (1992), Henze e Zirkler (1990) e Doornik e Hansen (2008), considerando 10.000 simulações Monte Carlo, ao nível de significância $\alpha = 0{,}05$ e combinações de $n = 25$, $n = 50$, $n = 75$, $n = 100$ e $n = 250$, com $p = 2$, $p = 3$, $p = 4$, $p = 5$ e $p = 10$. As distribuições *t-Student* multivariada com 10 graus de liberdade, uniforme multivariada, Pearson Tipo II com m = 10 e Knintchine foram utilizadas para a avaliação do poder.

Os autores constataram que o teste de Royston (1983) não apresentou bom controle da taxa de erro tipo *I*, sendo muito conservativo para todas as combinações de n e p. Portanto, não recomendaram seu uso. O teste de Royston (1992) e o teste Doornik e Hansen (2008) obtiveram bom controle das taxas de erro tipo *I* para todas as combinações de n e p. Em contrapartida, o teste de Henze e Zirkler (1990) se mostrou conservativo para os casos em que $n = 25$. Não houve um teste uniformemente mais poderoso. O teste de Royston (1992) obteve melhor desempenho de poder para amostras de tamanho $n = 25$, o teste de Doornik e Hansen (2008) se destacou para a situação em que a distribuição amostrada foi *t-Student* multivariada com 10 graus de liberdade e, de maneira geral, o teste de Henze e Zirkler (1990) apresentou bom desempenho de poder, especialmente, quando $n \geq 75$.

Diante das avaliações mencionadas, destacam-se o teste de normalidade multivariada baseado em *bootstrap* paramétrico proposto por Biase (2011), o teste de Royston (1992) e o teste de Henze e Zirkler (1990) como as melhores alternativas para testar a normalidade multivariada. O teste de Biase (2011) e o teste de Henze e Zirkler (1990) ainda não foram comparados sobre as mesmas condições, isto é, considerando o mesmo número de simulações Monte Carlo e as mesmas combinações de n e p. Portanto, não se pode recomendar o uso de um que tenha melhor desempenho de poder e melhor controle da taxa de erro tipo *I* em relação ao outro. Ressalta-se que o TKS e o TBPb propostos neste trabalho não devem ser utilizados, em nenhuma situação, em virtude do fraco desempenho de poder que apresentaram.

5 CONCLUSÕES

- Os dois testes de normalidade multivariada foram propostos e implementados com sucesso utilizando o programa R.

- A avaliação do desempenho dos testes propostos, utilizando simulação Monte Carlo, foi realizada conforme estabelecido, exceto para o TBPb na combinação $n = 10000$ com $p = 5$, cujo tempo de execução das simulações se mostrou demasiadamente longo.

- O teste de normalidade multivariada baseado na obtenção de amostras betas utilizando o teste de aderência de Kolmogorov-Smirnov apresentou bom controle das taxas de erro tipo *I*, sendo um teste exato. Porém, em relação ao poder, esse teste não obteve bom desempenho. Portanto, não é recomendado o seu uso.

- O teste *bootstrap* paramétrico para a normalidade multivariada baseado na obtenção de amostras betas teve sucesso no controle das taxas de erro tipo *I*. No entanto, não apresentou desempenho de poder satisfatório, se mostrando fraco para detectar a não normalidade multivariada na maioria das situações. Assim, não recomenda-se que esse teste seja utilizado.

Referências

ALVA, J. A. V.; ESTRADA, E. G. A generalization of Shapiro–Wilk's test for multivariate normality. **Communications in Statistics – Theory and Methods**, Taylor & Francis, Abingdon, v. 38, n. 11, p. 1870–1883, May 2009.

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. **The Annals of Mathematical Statistics**, JSTOR, Ann Arbor, v. 23, n. 2, p. 193–212, Jun. 1952.

BARON, M. **Probability and statistics for computer scientists**. 2. ed. New York, NY: CRC Press, 2013. 445 p.

BIASE, A. G. **Proposição de testes computacionalmente intensivos de normalidade multivariada**. 2011. 124 p. Tese (Doutorado em Estatística e Experimentação Agropecuária), Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, 2011.

CANTELMO, N. F.; FERREIRA, D. F. Monte Carlo evaluation of the performance of multivariate normality tests. **Ciência e Agrotecnologia**, Lavras, v. 31, n. 6, p. 1630–1636, nov./dez. 2007.

CASELLA, G.; BERGER, R. L. **Statistical inference**. 2. ed. Austrália, Pacific Grove, CA: Thomson Learning, 2002. 660 p.

CIRILLO, M. A.; FERREIRA, D. F. Extensão do teste para normalidade univariado baseado no coeficiente de correlação quantil-quantil para o caso multivariado. **Revista de Matemática e Estatística**, São Paulo, v. 21, n. 3, p. 57–75, Out. 2003.

DAVISON, A. C.; HINKLEY, D. V. **Bootstrap methods and their application**. New York, NY: Cambridge University Press, 1997. 582 p.

DOORNIK, J. A.; HANSEN, H. An Omnibus test for univariate and multivariate normality. **Oxford Bulletin of Economics and Statistics**, New York, v. 70, n. 1, p. 927–939, Dec. 2008.

EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, JSTOR, New York, v. 7, n. 1, p. 1–26, Jan. 1979.

EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. New York, NY: CRC press, 1993. 436 p.

EMBRECHTS, P.; FREY, R.; MCNEIL, A. **Quantitative risk management**. Princeton, NJ: Princeton University Press, 2005. 538 p.

FARRELL, P. J.; SALIBIAN-BARRERA, M.; NACZK, K. On tests for multivariate normality and associated simulation studies. **Journal of Statistical Computation and Simulation**, Taylor & Francis, New York, v. 77, n. 12, p. 1065–1080, Fev. 2007.

FERREIRA, D. F. **Estatística básica**. 2. ed. Lavras, MG: Editora UFLA, 2009. 664 p.

FERREIRA, D. F. **Estatística multivariada**. 2. ed. Lavras, MG: Editora UFLA, 2011. 675 p.

FERREIRA, D. F. **Estatística computacional em Java**. 1. ed. Lavras, MG: Editora UFLA, 2013. 695 p.

GENTLE, J. E. **Random number generation and Monte Carlo methods**. 2. ed. New York, NY: Springer Science & Business Media, 2003. 381 p.

GIVENS, G. H.; HOETING, J. A. **Computational statistics**. 2. ed. Hoboken, NJ: John Wiley & Sons, 2012. 496 p.

GNANADESIKAN, R.; KETTENRING, J. R. Robust estimates, residuals, and outlier detection with multiresponse data. **Biometrics**, JSTOR, Washington, v. 28, n. 1, p. 81–124, Mar. 1972.

HALTON, J. H. A retrospective and prospective survey of the monte carlo method. **Siam Review**, SIAM, Philadelphia, v. 12, n. 1, p. 1–63, Jul. 2006.

HAWKINS, D. M. A new test for multivariate normality and homoscedasticity. **Technometrics**, Taylor & Francis Group, Abingdon, v. 23, n. 1, p. 105–110, Feb. 1981.

HENZE, N.; ZIRKLER, B. A class of invariant consistent tests for multivariate normality. **Communications in Statistics-Theory and Methods**, Taylor & Francis, Abingdon, v. 19, n. 10, p. 3595–3617, 1990.

KENDALL, M. G.; STUART, A. **The advanced theory of statistics - Inference and Relationship.** London: Charles Griffin and Co. Ltd., 1961. 676 p.

KOZIOL, J. A. A class of invariant procedures for assessing multivariate normality. **Biometrika**, Biometrika Trust, Oxford, v. 69, n. 2, p. 423–427, Aug. 1982.

LI, R.-Z.; FANG, K.-T.; ZHU, L.-X. Some qq probability plots to test spherical and elliptical symmetry. **Journal of Computational and Graphical Statistics**, Taylor & Francis, Abingdon, v. 6, n. 4, p. 435–450, Dec. 1997.

LIANG, J.; PAN, W. S.; YANG, Z.-H. Characterization-based q–q plots for testing multinormality. **Statistics & Probability Letters**, Elsevier, Amsterdam, v. 70, n. 3, p. 183–190, Dec. 2004.

MARDIA, K. V. Measures of multivariate skewness and kurtosis with applications. **Biometrika**, Biometrika Trust, Oxford, v. 57, n. 3, p. 519–530, Dec. 1970.

MARDIA, K. V.; FOSTER, K. Omnibus tests of multinormality based on skewness and kurtosis. **Communications in Statistics-theory and methods**, Taylor & Francis, Abingdon, v. 12, n. 2, p. 207–221, 1983.

MARDIA, K. V.; KENT, J. Rao score tests for goodness of fit and independence. **Biometrika**, Biometrika Trust, Oxford, v. 78, n. 2, p. 355–363, Jun. 1991.

MARQUES E MELO, J. **Proposta de um teste exato para avaliar a normalidade multivariada baseado em uma transformação t-Student**. 2016. 86 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária), Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, 2016.

MASSEY JÚNIOR, F. J. The Kolmogorov-Smirnov test for goodness of fit. **Journal of the American Statistical Association**, Taylor & Francis Group, Abingdon, v. 46, n. 253, p. 68–78, Mar. 1951.

MECKLIN, C. J.; MUNDFROM, D. J. A Monte Carlo comparison of the type I and type II error rates of tests of multivariate normality. **Journal of Statistical Computation and Simulation**, Taylor & Francis, Abingdon, v. 75, n. 2, p. 93–107, Feb. 2005.

MELLO, M.; PETERNELLI, L. **Conhecendo o R - Uma visão mais que estatística**. 1. ed. Viçosa, MG: Editora UFV, 2013. 222 p.

METROPOLIS, N.; ULAM, S. The monte carlo method. **Journal of the American Statistical Association**, Taylor & Francis Group, Abingdon, v. 44, n. 247, p. 335–341, Sep. 1949.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the Theory of Statistics.** 3. ed. New York, NY: McGraw-hill, 2001. 564 p.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; LIN, C. T. Some p-variate adaptations of the shapiro-wilk test of normality. **Communications in Statistics-Theory and Methods**, Taylor & Francis, Abingdon, v. 24, n. 4, p. 953–985, 1995.

OLIVEIRA, I. R. C.; FERREIRA, D. F. Multivariate extension of chi-squared univariate normality test. **Journal of Statistical Computation and Simulation**, Taylor & Francis, Abingdon, v. 80, n. 5, p. 513–526, Mar. 2009.

PAULSON, A.; ROOHAN, P.; SULLO, P. Some empirical distribution function tests for multivariate normality. **Journal of Statistical Computation and Simulation**, Taylor & Francis, Abingdon, v. 28, n. 1, p. 15–30, 1987.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2015. Disponível em: http://www.R-project.org/. Acesso em: 21 jul. 2015.

RIPLEY, B. D. **Stochastic simulation**. New York, NY: John Wiley & Sons, 1987. 237 p.

ROBERT, C.; CASELLA, G. **Introducing Monte Carlo Methods with R**. New York, NY: Springer Science & Business Media, 2009. 283 p.

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

ROYSTON, J. Some techniques for assessing multivarate normality based on the Shapiro-Wilk W. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, JSTOR, London, v. 32, n. 2, p. 121–133, 1983.

ROYSTON, P. Approximating the Shapiro-Wilk W-test for non-normality. **Statistics and Computing**, Springer, London, v. 2, n. 3, p. 117–119, 1992.

ROYSTON, P. A toolkit for testing for non-normality in complete and censored samples. **Journal of the Royal Statistical Society. Series D (The Statistician)**, JSTOR, London, v. 42, n. 1, p. 37–43, 1993.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, JSTOR, Oxford, v. 52, n. 3-4, p. 591–611, Dec. 1965.

SILVA, R. B. V. **Extensão do teste de normalidade de Shapiro-Francia para o caso multivariado**. 2009. 59 p. Tese (Doutorado em Estatística e Experimentação Agropecuária), Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, 2009.

SINGH, A. Omnibus robust procedures for assessment of multivariate normality and detection of multivariate outliers. **Multivariate Environmental Statistics**, G.P. Patil and C.R. Rao, North Holland, p. 445–488, 1993.

SINGH, K.; XIE, M. **Bootstrap: a statistical method**. Rutgers University, USA, 2008. Disponível em: <http://www.stat.rutgers.edu/home/mxie/RCPapers/bootstrap.pdf>. Acesso em: 21 jul. 2015.

SRIVASTAVA, M.; HUI, T. On assessing multivariate normality based on Shapiro–Wilk W statistic. **Statistics & Probability Letters**, Elsevier, Amsterdam, v. 5, n. 1, p. 15–18, Jan. 1987.

SZÉKELY, G. J.; RIZZO, M. L. A new test for multivariate normality. **Journal of Multivariate Analysis**, Elsevier, Amsterdam, v. 93, n. 1, p. 58–80, Mar. 2005.

VENABLES, W.; SMITH, D.; TEAM, R. C. **An Introduction to R: Notes on R, A Programming Environment for Data Analysis and Graphics**. Vienna, Austria, 2015. Disponível em:<http://www.R-project.org/>. Acesso em: 21 jul. 2015.

YANG, Z.-H.; FANG, K.-T.; LIANG, J.-J. A characterization of multivariate normal distribution and its application. **Statistics & Probability Letters**, Elsevier, Amsterdam, v. 30, n. 4, p. 347–352, Nov. 1996.

ANEXO A - Rotinas usadas no R

```
#_____________________________________________________#
#                                                     #
# Implementação da rotina realizada por Silva (2009) #
# para o teste Shapiro-Wilk de normalidade  multiva- #
# riado  proposto por Royston (1983, 1993). Em que x #
# é a matriz de dados (n x p) e as funções auxilia-  #
# res ki, fz, fe, grhon foram baseadas no artigo:    #
# ROYSTON,J. P. A tolkit for testing for             #
# non-normality in complete and censored samples.    #
# The Statistician, London, v. 42, n. 1, p. 37-43,   #
# 1993.                                              #
#_____________________________________________________#

Shapiro.Royston = function(x){
#auxiliar functions
ki = function(zj) {
ki = (qnorm(0.5*pnorm(-zj)))^2
return(ki)
}

fz = function(wj,n){
if (n <= 11) {
lamb = -2.273+0.459*n
u    = log(1-wj)
y    = -log(lamb-u)
muy  = 0.544-0.39978*n+0.025054*n^2-0.0006714*n^3
sigy = exp(1.3822-0.77857*n+0.062767*n^2-0.0020322*
n^3)
} else {
y=log(1-wj)
xx=log(n)
muy=-1.5861-0.31082*xx-0.083751*xx^2+0.0038915*xx^3
sigy=exp(-0.4803-0.082676*xx+0.0030302*xx^2)
}
zc = (y-muy)/sigy
return(zc)
}
```

```
fe  = function(cbar,p){
e = p/(1+(p-1)*cbar)
return(e)
}

grhon = function(rho,n){
mu = 0.715
lamb = 5
xx = log(n)
nu = 0.21364+0.015124*xx^2-0.0018034*xx^3
#nu=0.35
#print(nu)
grijn = rho^lamb*(1-mu/nu*rho*(1-rho)^mu)
grijn
}
# main procedures
if (!is.matrix(x))
stop("x[] is not a matrix with number of rows
(sample size) between 3 and 5000")
p=ncol(x)
p
n=nrow(x)
n
if (n < 3 || n > 5000)
stop("sample size must be between 3 and 5000")
rx = cor(x)
#Calculating the univariate W statistics using R
function
Wis = matrix(0,p,1)
for (i in 1:p){
xj=x[1:n,i]
wi = shapiro.test(xj)
Wis[i,1]=wi$statistic
}
#calculating Cij matrix
Cij = matrix(1,p,p)
if (p>1){
pp=1:p-1
```

```
#print(pp)
for (i in pp){
ii=i+1
seq=ii:p
for (j in seq){
#print(i);print(j);print(seq)
cij = grhon(abs(rx[i,j]),n)
Cij[i,j]=cij
#print(cij)
Cij[j,i]=cij
}
}
}
# test statistics and p-value
if(p>1) cbar=sum(Cij-diag(p))/(p*(p-1)) else
cbar=0
e=fe(cbar,p)
Kis=ki(fz(Wis,n))
G=sum(Kis)/p
W.mean = mean(Wis)
H=e*G
prH =1 - pchisq(H,e)
return(list(W = W.mean, G=G, H = H, dfeq = e,
p.value = prH))
}


#____________________________________________________#
#                                                    #
# Função para gerar números aleatórios da distribui- #
# ção normal contaminada                             #
#____________________________________________________#

mvrnormCT <- function(n, mu1, Sigma1, mu2, Sigma2,
delta)
{
u <- runif(n)
if(u[ 1 ] <= delta) X <- mvrnorm(1, mu1, Sigma1) else
X <- mvrnorm (1, mu2, Sigma2)
for (i in 2:n)
```

```
{
if(u[ i ] <= delta) y <- mvrnorm(1, mu1, Sigma1 ) else
y <- mvrnorm (1, mu2, Sigma2)
X <- rbind(X, y)
}
return(X)
}
```

APÊNDICE A - Rotinas usadas no R

```
#_____________________________________________________________#
#                                                             #
# Implementação da rotina do teste de normalidade            #
# multivariada baseado na obtenção de amostras betas #
# utilizando o teste de Kolmogorov-Smirnov (TNMKS),  #
# em que x é a matriz de dados (n x p)                #
#_____________________________________________________________#

# Função que obtem os Ui, com i = 1, ..., n - 1.

U <- function(x)
{
n <- nrow(x)
i <- 1:(n-1)
U <- (apply(x,2,cumsum)[i,] - i * x[2:n,]) /
 sqrt(i * (i+1))
return(U)
}

# Função que obtem os Sk, com k = p + 1, ..., n - 1.

Sk <- function(x)
{
n      <- nrow(x)
p      <- ncol(x)
u      <- U(x)
S      <- array(NA, dim=c(p,p,(n-1-p)))
S[,,1] <- u[1,] %*% t(u[1,])

for(i in 2:(p+1))
{
S[ , ,1] <- S[, ,1] + (u[i,] %*% t(u[i,]))
}

for(j in 2:(n-1-p))
{
S[,,j] <- S[,,(j-1)] +  (u[(p + j),] %*%
```

```
t(u[(p + j),]))
}
return(S)
}

# Função que obtem os Yk, com k = p + 1, ..., n - 1.

Yk <- function(x)
{
n      <- nrow(x)
p      <- ncol(x)
u      <- U(x)
sk     <- Sk(x)
yk <- matrix(NA,n-1-p,p)
isk <- array(NA, dim=c(p,p,n-1-p))
for(i in 1:(n-1-p))
{
isk[, , i] <- solve(t(chol(sk[, , i])))
yk[i,] <- isk[, , i] %*% u[p+i,]
}
return(yk)
}

# Função que obtem as amostras beta

bkq <- function(x)
{
n  <- nrow(x)
p  <- ncol(x)
yk2 <- (Yk(x))^2
if(p==2) B <- as.matrix(yk2[,1] /
apply(yk2, 1, sum)) else
B  <- t(apply(yk2[,1:(p-1)],1,cumsum)) /
apply(yk2, 1, sum)
return(B)
}

# Teste de normalidade multivariada via
# Kolmogorov Smirnov que testa beta na
```

```
# cauda superior e seleciona a amostra
# de acordo com [p/2].

TNMKS <- function(x)
{
b          <- bkq(x)
p          <- ncol(x)
qescolhido <- p %/% 2
KS         <- ks.test(b[,qescolhido],
              "pbeta", qescolhido/2,
              (p - qescolhido)/2,
              alternative = "greater")
return(KS)
}


#_____________________________________________________#
#                                                     #
# Implementação da rotina do teste bootstrap paramé- #
# trico de normalidade multivariada baseado na obten-#
# ção de amostras betas (TNMBPb), em que x é a ma-   #
# triz de dados (n x p)                               #
#_____________________________________________________#

# Função que obtem os Ui, com i = 1, ..., n - 1.

U <- function(x)
{
n <- nrow(x)
i <- 1:(n-1)
U <- (apply(x,2,cumsum)[i,] - i * x[2:n,]) /
     sqrt(i * (i+1))
return(U)
}

# Função que obtem os Sk, com k = p + 1, ..., n - 1.

Sk <- function(x)
{
n      <- nrow(x)
```

```
p      <- ncol(x)
S      <- array(NA, dim=c(p,p,(n-p)))
S[,,1] <- x[1,] %*% t(x[1,])

for(i in 2:(p+1))
{
S[ , ,1] <- S[, ,1] + (x[i,] %*% t(x[i,]))
}

for(j in 2:(n-p))
{
S[,,j] <- S[,,(j-1)] +  (x[(p + j),] %*%
          t(x[(p + j),]))
}
return(S)
}

# Função que obtem os Yk, com k = p + 1, ..., n - 1.

Yk <- function(x)
{
n      <- nrow(x)
p      <- ncol(x)
sk     <- Sk(x)
yk <- matrix(NA,n-p,p)
#isk <- array(NA, dim=c(p,p,n-1-p))
for(i in 1:(n-p))
{
yk[i,] <- solve(t(chol(sk[, , i]))) %*% x[p+i,]
}
return(yk)
}

# Função que obtem as amostras beta

bkq <- function(x)
{
n  <- nrow(x)
p  <- ncol(x)
```

```
yk2 <- (Yk(x))^2
if(p==2) B <- as.matrix(yk2[,1] /
              apply(yk2, 1, sum)) else
B  <- t(apply(yk2[,1:(p-1)],1,cumsum)) /
apply(yk2, 1, sum)
return(B)
}

# Função que obtem o máximo das estatísticas de
# Kolmogorov-Smirnov

Dmax <- function(x)
{
p  <- ncol(x)
b  <- bkq(x)
q  <- ncol(b)
D  <- c()
for(i in 1:q)
{
D[i] <- ks.test(b[,i],"pbeta", i/2, (p-i)/2)$statistic
}
return(max(D))
}

# Teste bootstrap paramétrico de normalidade
# multivariada baseado nas amostras betas

TNMBPb <- function(x, Nb = 2000)
{
library(mvtnorm)
u  <- U(x)
D0 <- Dmax(u)
S  <- var(u)
n  <- nrow(u)
p  <- ncol(u)
mu <- rep(0,p)
D  <- NULL
for(i in 1:Nb)
{
```

```
y    <- rmvnorm(n, mu, S)
D[i] <- Dmax(y)
}
D <- c(D,D0)
PVAL <- length(D[D>=D0]) /  (Nb + 1)
return(list(p.value = PVAL, D0 = D0))
}



#____________________________________________________#
#                                                    #
# Implementação da rotina do teste de Embrechts, Frey#
# e McNeil (2005) (TNMEM), Quantitative risk manage  #
# ment. Princeton Series in Finance, Princeton,      #
# v. 10, 2005. Em que X é a matriz de dados (n x p). #
#____________________________________________________#

TNMEM <- function(X)
{
n  <- nrow(X)
p  <- ncol(X)
j  <- rep(1,n)
Xb <- 1/n * t(X) %*% j
iS  <- solve(var(X))
D2 <- c()
for(k in 1:n)
{
D2[k] <- t(X[k,] - Xb) %*% iS %*% (X[k,] - Xb)
}
b <- n/(n-1)^2 * D2

# parâmetros da B(alfa,beta)

alfa <- p / 2
beta <- (n - p - 1) / 2
resultado <- ks.test(b, "pbeta", alfa, beta,
             alternative = "greater")
return(list( D2 = resultado$statistic, p.value =
       resultado$p.value))}
```

```
#________________________________________________________#
#                                                        #
# Implementação da rotina para validação dos testes      #
# dist = 1: normal multivariada                          #
# dist = 2: t-Student multivariada com 1 GL              #
# dist = 3: t-Student multivariada com 30 GL             #
# dist = 4: log-normal multivariada                      #
# dist = 5: uniforme multivariada                        #
# dist = 6: normal contaminada multivariada              #
# Nb: número de simulações bootstrap (somente no         #
# TNMBPb)                                                #
# Nmc: número de simulações Monte Carlo                  #
#________________________________________________________#

# Função de validação

SMCbootstrap <- function(n, p, Nb=2000, Nmc=2000,
                dist = 1)
{
library(MASS)
library(mvtnorm)
resultado           <- matrix(0,1,3)
rownames(resultado) <- c("TNMBPb")
colnames(resultado) <- c("0.10", "0.05", "0.01")
mu                  <- rep(0, times = p)
rho                 <- 0.5
sigma2              <- 1
Sigma               <- sigma2 * ( (1 - rho ) *
                       diag(p) + rho *
                       matrix(1, p, p) )

# valores referentes à normal contaminada:

mu2                 <- mu + 10
Sigma2              <- 10 * Sigma
delta               <- 0.3

for(i in 1:Nmc)
{
```

```
if(dist == 1) X <- rmvnorm(n, mu, Sigma) else
if(dist == 2) X <- rmvt(n, Sigma, 1, mu) else
if(dist == 3) X <- rmvt(n, Sigma, 30, mu) else
if(dist == 4) X <- exp(rmvnorm(n, mu, Sigma)) else
if(dist == 5) X <- pnorm(rmvnorm(n, mu, Sigma)) else
if(dist == 6) X <- mvrnormCT(n, mu, Sigma, mu2,
                   Sigma2, delta)
res <- TNMBPb(X, Nb = Nb) # trocar TNMBPb pela
                          # função do teste para
                          # o qual se deseja fazer
                          # a simulação Monte Carlo
if(res$p.value <= 0.10) resultado[1,1] <- resultado[1,
                                          1] + 1 / Nmc
if(res$p.value <= 0.05) resultado[1,2] <- resultado[1,
                                          2] + 1 / Nmc
if(res$p.value <= 0.01) resultado[1,3] <- resultado[1,
                                          3] + 1 / Nmc
}
return(resultado)
}
```