NELSON DE ALMEIDA PEREIRA FILHO

TESTE MONTE CARLO DE NORMALIDADE UNIVARIADO

LAVRAS - MG
2013

NELSON DE ALMEIDA PEREIRA FILHO

TESTE MONTE CARLO DE NORMALIDADE UNIVARIADO

Tese 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 Doutor.

Dr. Daniel Furtado Ferreira
Orientador

LAVRAS - MG
2013

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

Pereira Filho, Nelson De Almeida.
Teste Monte Carlo de normalidade univariado / Nelson de Almeida Pereira Filho. – Lavras : UFLA, 2013.
55 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2013.
Orientador: Daniel Furtado Ferreira.
Bibliografia.

1. Normalidade. 2. Simulação. 3. Shapiro-Wilk. 4. R. 5. Poder. I. Universidade Federal de Lavras. II. Título.

CDD – 519.24

NELSON DE ALMEIDA PEREIRA FILHO

TESTE MONTE CARLO DE NORMALIDADE UNIVARIADO

Tese 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 Doutor.

APROVADA em 13 de junho de 2013.

Profa. Dra. Giovana Oliveira Silva	Universidade Federal da Bahia
Prof. Dr. José Airton Rodrigues Nunes	UFLA
Prof. Dr. Paulo César Lima	UFLA
Profa. Dra. Telma Safadi	UFLA

Prof. Dr. Daniel Furtado Ferreira
Orientador

LAVRAS - MG
2013

AGRADECIMENTOS

A Deus, Senhor de todos nós, pela minha vida;

Ao meu orientador, Professor Daniel Furtado Ferreira, por toda a atenção, dedicação, cuidado e inspiração. Conduziu-me com sabedoria e rigor. Seus ensinamentos e seu compromisso com a pesquisa me servirão como principal referência daqui pra frente;

Aos meus pais, Nelson e Júlia, que construíram o meu caráter; aos meus irmãos, Marilena, Arlindo, Ana, Pedro e Nélia, referências e apoio por toda a minha vida;

Ao meu filho Luiz, por ser a principal razão de todo esforço desprendido;

A minha esposa, Maristela, pelo amor, pelo cuidado e pela paciência que me dedicou durante esse período difícil;

Aos professores do Departamento de Ciências Exatas da UFLA, pelos ensinamentos, pela amizade e excelente convivência;

À minha amiga e colega, Adriele Biase, pelo apoio, companheirismo, amizade e confiança que tivemos mutuamente;

Aos professores membros da banca, principalmente ao Professor José Airton Rodrigues Nunes, pelas contribuições dadas para a conclusão deste trabalho;

Aos meus colegas do DINTER, principalmente ao colega José Otaviano, pela companhia e pelo apoio que sempre dedicaram;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro, ao Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA) e à Universidade Federal de Lavras;

Ao professor, Lurimar Smera Batista, pela sua dedicação, companheirismo, compromisso e profissionalismo, desprendido durante o curso;

Aos colegas e amigos Luiz Vasques, Jaime Filho e Jailson Rodrigues, pelo excelente convívio, grandes companheiros durante a jornada;

A todos os funcionários do Departamento de Ciências Exatas, em especial a Josi (secretária da Pós-Graduação), pela prontidão e suporte;

Muito obrigado a todos!

RESUMO

As distribuições normais de probabilidade descrevem o comportamento de muitos fenômenos da vida real em vários campos da ciência. Ao se retirar uma amostra aleatória de uma população, no caso univariado, é comum se pressupor que os dados ou resíduos do modelo adotado são proveniente de uma população normalmente distribuída. Os gráficos como histogramas e Q-Qplots são maneiras bastante eficientes, porém subjetivas de se verificar a normalidade da distribuição dos dados ou dos resíduos do modelo considerado. No entanto isso não é suficiente para se fazer inferência sobre a normalidade dos dados coletados ou dos resíduos do modelo utilizado para modelar os dados. Objetivou-se com estas pesquisas comparar o teste de normalidade Shapiro-Wilk com um teste de normalidade univariado proposto. Este novo teste é baseado no cálculo do coeficiente de determinação entre os observados das estatísticas de ordem amostrais e os valores esperados das estatísticas de ordem da distribuição normal padrão. A distribuição nula da estatística foi obtida via simulação Monte Carlo. Os resultados obtidos de poder e do erro tipo I, permitem que se conclua que a proposta é, em geral, mais eficiente que o teste Shapiro-Wilk e não possui a limitação prática de ser restrito a tamanho de amostra máximo de 5.000 unidades, que é uma limitação de seu principal concorrente.

Palavras-chave: Normalidade. Simulação. Shapiro-Wilk. R. Poder.

ABSTRACT

The normal probability distributions describe the behavior of many real-life phenomena in various fields of science. When a random sample was drawn from a population, in the univariate case, it is common to assume that the data or the model residuals are normally distributed. Graphs like histograms and Q-Qplots are ways quite efficient, but subjective, to verify the assumption of normality of data or model residuals. However, this procedures are not enough to verify the normality of the data or the residuals of the adopted model. This research aims to compare the Shapiro-Wilk normality test with a proposed univariate normality test. This new test is based on the coefficient of determination between the observed values of order statistics and the expected values of standard normal order statistics. The null distribution of the statistic was obtained via Monte Carlo simulation. The results for power and of type I error rates, allow the conclusion that the proposal method is generally more efficient than the Shapiro-Wilk and has no practical limitation of being restricted to a maximum sample size of $5{,}000$ units, which is the main competitor limitation.

Keywords: Normality. Simulation. Shapiro-Wilk. R. Power.

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A inferência é um ramo da estatística que consiste em fazer afirmativas válidas sobre parâmetros de alguma população, baseadas em dados amostrais. Por exemplo, ao se obter uma estimativa pontual acerca de um certo parâmetro, é quase certo que ela, a estimativa, não será igual ao valor do parâmetro. Deste modo, inferências por meio de estimação intervalar e testes de hipóteses devem ser realizadas. Esses procedimentos inferenciais incorporam as incertezas existentes nos estudos de todos os fenômenos naturais e permitem que sejam obtidas conclusões sobre parâmetros contemplando margem de erro ou confiança de que as decisões estejam corretas nos testes de hipóteses.

Ao se retirar uma amostra de uma população, em grande parte dos casos, pressupõe-se que os dados sejam provenientes de uma população normalmente distribuída. Tal suposição é feita pelo simples fato de que as distribuições normais podem ser usadas para descrever muitas situações da vida real e são largamente aplicadas em vários campos da ciência. Além disso, a facilidade da obtenção de estimadores por intervalo e testes exatos é grande quando se assume esse modelo. Ademais, a maior parte dos testes e procedimentos de estimação é formulada tomando-se a normal com referência para a distribuição dos dados. A validade da inferência, para a maioria dos procedimentos que supõe normalidade, é garantida quando essa distribuição é de fato a distribuição dos dados ou dos resíduos de um modelo linear que estão sendo considerados.

Uma maneira simples, porém subjetiva, de se verificar a normalidade de um conjunto de dados é por meio de gráficos, como histogramas e Q-Qplots. Tal observação, no entanto, não é suficiente pra se fazer inferência sobre a normalidade. No caso multivariado, principalmente nas situações de muitas variáveis, isso se torna ainda mais complicado, porque nem sempre é possível detectar-se alguma violação da normalidade, haja vista a complexa relação existentes entre as varáveis envolvidas. Apesar disso, os gráficos Q-Qplots, são ferramentas viáveis para a visualização de valores discrepantes da amostra. Inúmeros testes de normalidade univariada existem. Entretanto, apesar de a maioria controlar adequadamente os erros tipo I, o poder desses testes variam consideravelmente em função das diferentes distribuições não-normais consideradas em suas avaliações (OLIVEIRA; FERREIRA, 2010).

A ausência de um teste uniformemente poderoso em relação aos tamanhos amostrais e distribuições tem sido o principal motivo da grande quantidade de proposições de testes de normalidade. Um teste que vem sendo largamente utilizado como referência, tanto nas aplicações científicas, em geral, quanto nos trabalhos de comparação de desempenho de outros testes propostos é o de Shapiro-Wilk. O grande entrave ao uso desse teste é a limitação dos tamanhos amostrais ao máximo de 5.000 observações. Em muitas aplicações em zootecnia, na área de melhoramento animal e vegetal, e em ciências florestais, é muito comum o pesquisador se deparar com amostras superiores a esse tamanho. Existem alternativas de testes que são aplicáveis a grandes amostras, mas possuem limitações.

Um teste que pode ser aplicado em grandes amostras é baseado na verificação dos desvios de assimetria e curtose, de forma independente ou em conjunto. Entretanto, esses testes comparam apenas os quatro primeiros momentos da distribuição normal teórica com os mesmos momentos estimados dos dados. Grandes desvios dos momentos amostrais dos seus valores esperados são considerados como um indicativo de não-normalidade. Esse procedimento não garante a normalidade no caso de não-rejeição, pois a igualdade dos quatro primeiros momentos é uma condição necessária, mas não suficiente. Outros exemplos, como o teste de Kolmogorov-Smirnov pode ser aplicado em grandes amostras. Esse teste é baseado na distância entre a função de distribuição empírica e a distribuição teórica, no caso a normal, que se está testando. Mas o teste de Kolmogorov-Smirnov é pouco poderoso.

O advento de computadores com processadores cada vez mais poderosos e atualmente com vários núcleos de processamento em apenas uma máquina, tem possibilitado cada vez mais que testes computacionalmente intensivos baseados em procedimentos Monte Carlo e *bootstrap* tenham mais espaços. A concepção de testes desse tipo é simples e os requerimentos teóricos das distribuições de suas estatísticas são mínimos, uma vez que essas distribuições são geradas por simulação. Usando propriedades das distribuições de interesse, que nesse trabalho é a normal, é possível formular testes que sejam exatos, no sentido de controlarem adequadamente o erro tipo I, ou seja, que tenham tamanho igual ao nível nominal de significância α. Ademais, é possível construir tais testes sem limitações quanto aos tamanhos amostrais máximos e que sejam mais poderosos que os testes já existentes para uma ampla classe de distribuições probabilísticas consideradas sob a hipótese alternativa.

Sendo assim, o uso de testes estatísticos se faz necessário para se inferir sobre a normalidade da distribuição dos dados que estão sendo analisados. A garantia de normalidade, quando se aplica os procedimentos clássicos de inferência estatística, assegura a validade das conclusões e possibilita aos pesquisadores uma maior confiabilidade na divulgação dos resultados de suas investigações científicas. A existência de um procedimento de normalidade tão poderoso quanto o teste de Shapiro-Wilk e que pudesse ser utilizado em grande amostras, ou seja, para tamanhos de amostras,n, para $n > 5.000$ foram as razões que motivaram o desenvolvimento desse trabalho.

Assim, objetivou-se com o presente trabalho realizar a proposição de um teste de normalidade univariada baseado em métodos Monte Carlo e que possa ser usado para qualquer tamanho de amostra, incluindo os casos de amostras superiores a 5.000. Além disso, objetiva-se validar o procedimento proposto por meio da avaliação das taxas de erro tipo I e poder, utilizando para isso os métodos Monte Carlo, considerando diferentes tamanhos de amostras e níveis nominais de significância, além de comparar os resultados do desempenho do teste proposto com os do teste de Shapiro-Wilk, nos casos em que $n \leq 5.000$.

2 REFERENCIAL TEÓRICO

Muitos procedimentos de inferência são baseados em um modelo normal da população sob amostragem (FERREIRA, 2005), pois a normalidade é a mais usual suposição feita no desenvolvimento e no uso dos procedimentos estatísticos (THODE JUNIOR, 2002). Em geral, os testes são feitos para avaliar a veracidade de alguma hipótese de interesse, que é qualquer conjectura sobre aspectos desconhecidos da população. Tais aspectos podem ser parâmetros, ou mesmo a família de distribuições ou a forma da distribuição. Os testes estatísticos usados para verificar se um conjunto de dados de uma variável aleatória pode ser modelado por uma distribuição normal, são chamados de testes de normalidade. No caso univariado, alguns testes de normalidade mais usados ou citados, são os testes Shapiro-Wilk(1965), Shapiro-Francia, qui-quadrado de Pearson, Lilliefors, Kolmogorov-Smirnov(1932) e Jarque-Bera. Embora isso seja verdade, o assunto segundo Thode Junior (2002) não sofreu por falta de atenção, pois, como é apontado pelos próprios autores, há mais testes de normalidade do que eles jamais haviam imaginado. Em seu livro, o autor abordou 40 testes de normalidade.

Thode Junior (2002) aponta que resultados surpreendentes foram observados em seus levantamentos sobre os testes de normalidade. Ele sugere que testes como o de Kolmogorov-Smirnov e de Qui-Quadrado possuem poder tão pequeno que eles deveriam ser desconsiderados como testes de normalidade. Por outro lado, os testes baseados em momentos, testes de assimetria e curtose e o teste de Shapiro-Wilk são tão impressionantes nos seus desempenhos que eles deveriam ser recomendados nas aplicações práticas do dia-a-dia.

Os procedimentos estatísticos como o teste t, o teste F na análise de variância e na verificação da homogeneidade de variâncias, os testes dos coeficientes de regressão entre outros casos possuem uma pressuposição subjacente de que os dados amostrais ou os resíduos do modelo adotado se distribuem como normais. Essa pressuposição deve ser checada por um teste que possa efetivamente constatar se ela foi atendida ou não. Além disso, pode-se usar argumentos sólidos e cientificamente embasados de que a violação da suposição de normalidade não invalida o procedimento utilizado. Muitas pesquisas são realizadas por estatísticos para avaliar a robustez dos procedimentos quando as suposições em que eles, os procedimentos, foram construídos são violadas. Especificamente, muitos trabalhos buscam medir o efeito específico da violação de normalidade nesses procedimentos. Um dos exemplos mais clássicos da sensibilidade de um teste à violação da normalidade é o caso do teste de Bartlett para homogeneidade de variâncias, que é seriamente comprometido pela não-normalidade.

Os muitos testes de normalidade existentes na literatura, em geral, se beneficiam de propriedades especiais da distribuição normal em suas construções ou fundamentações teóricas. Por exemplo, o teste de Shapiro-Wilk compara uma estimativa do desvio padrão usando combinação linear de estatísticas de ordem com a estimativa usual. Em virtude de haver um número muito grande de testes, mas que os principais são os testes de Shapiro-Wilk e de assimetria e curtose, a ênfase dessa revisão será sobre esses procedimentos. Alguns outros testes também serão considerados nessa revisão, em função de seu uso ter sido difundido pelas implementações existentes nos principais programas de computadores para análises estatísticas.

Supondo que a variável aleatória X possua uma função densidade $f(x)$, o problema dos testes de ajustes consiste na verificação da validade da hipótese nula dada por

$$H_0 : f(x) = f_0(x), \tag{1}$$

contra uma hipótese alternativa H_1. A função densidade $f_0(x)$ sob a hipótese nula (1) tem uma forma distribucional específica. Quando os parâmetros dessa distribuição são completamente especificados, a hipótese nula é chamada de hipótese simples. Se um ou mais dos parâmetros de $f_0(x)$ na hipótese nula não forem especificados, H_0 é chamada de hipótese composta.

A hipótese alternativa H_1 pode ser completamente especificada, incluindo os parâmetros, ou pode ser determinada somente pela classe de modelos probabilísticos ou, ainda, pode ser completamente geral, conforme os três tipos de hipóteses alternativas dados por

$$H_1^{(a)} : f(x) = f_1(x;\theta), \qquad H_1^{(b)} : f(x) = f_1(x) \quad \text{e} \quad H_1^{(c)} : f(x) \neq f_0(x). \tag{2}$$

Em geral, como será também feito nesse trabalho, a hipótese nula será a hipótese composta dada por

$$H_0 : f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{3}$$

em que a média $\mu \in \mathbb{R}$ e a variância $\sigma^2 \in \mathbb{R}_+$ em (3) são desconhecidos. A hipótese alternativa que será considerada nesse trabalho será a geral, ou seja, $H_1^{(c)}$ de (2).

Hipótese nula em que se especifica os parâmetros são menos gerais e de uso muito restritos a certas circunstâncias reais específicas. Da mesma forma, a utilização de distribuições específicas nas hipóteses alternativas tornam os testes mais poderosos, à medida que os parâmetros são definidos. Entretanto, tornam a utilização do teste bem limitada e restrita àquela classe de hipótese alternativa, ou seja, àquele modelo particular definido na hipótese alternativa. Assim, o objetivo, como é considerado nos principais testes já citados nesse texto, é considerar a hipótese nula composta e a hipótese alternativa mais geral, para que a utilidade da proposta seja mais ampla, no sentido de detectar uma vasta gama de modelos não-normais. O preço que se paga com esse tipo de estratégia é a obtenção de testes que não são uniformemente poderosos para todas as classes de hipóteses alternativas (THODE JUNIOR, 2002).

Em todos os casos será considerada uma amostra aleatória de tamanho n da população para a qual se almeja testar a normalidade, dada por X_1, X_2, ..., X_n. Os estimadores da média e da variância populacionais são representado por $\bar{X}$ e S^2, respectivamente. Nas discussões e apresentações seguintes essas quantidades serão mencionadas e utilizadas.

2.1 Estatísticas de Ordem

Inicialmente, sejam consideradas as variáveis aleatórias X_1, X_2, ..., X_n, as quais são consideradas independentes e possuem uma função de distribuição comum $F_X(x)$, assim, essas variáveis aleatórias podem ser consideradas uma amostra aleatória de tamanho n de uma população com distribuição F. O conjunto de valores observados x_1, x_2, ..., x_n das variáveis aleatórias X_1, X_2, ..., X_n é chamado de uma realização da amostra. Alternativamente, pode-se dizer que X_1, X_2, ..., X_n representam n independentes cópias independentes da variável aleatória X, que possui função de distribuição F. Apresentado dessa forma, verifica-se que o interesse não está nos valores ordenados dessas variáveis aleatórias.

Por outro lado, se o interesse está focado nos valores amostrais ordenados, então é necessário examinar as variáveis aleatórias

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(j)} \leq \cdots \leq X_{(n)},$$

que são conhecidas por estatísticas de ordem de uma amostra de tamanho n. A variável aleatória $X_{(j)}$ é chamada de j-ésima estatística de ordem, para $1 \leq j \leq n$. Os valores realizados das estatísticas de ordem são denotados por $x_{(1)}$, $x_{(2)}$, ..., $x_{(n)}$. Pode-se perceber que $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$ e $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$, representam o mínimo e o máximo amostral e referem-se à casos particulares das estatísticas de ordem, em geral, de muito interesse nas pesquisas.

A distribuição conjunta das n estatísticas de ordem é dada por

$$f(x_1, x_2, \cdots, x_n) = n! \prod_{i=1}^{n} f(x_k), \quad -\infty < x_1 < x_2 < \cdots < x_n < \infty, \tag{4}$$

em que $f(x)$ é a função densidade da variável aleatória X_i, $i = 1, 2, \ldots, n$.

A função de distribuição do máximo, sendo $F(X)$ a função de distribuição comum das n variáveis aleatórias amostrais pode ser obtida por

$$P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, \cdots, X_n \leq x).$$

Como X_1, X_2, ..., X_n é uma amostra aleatória e, portanto, têm distribuição independente e idêntica, essa expressão pode ser simplificada em

$$\begin{aligned} P(X_{(n)} \leq x) &= P(X_1 \leq x, X_2 \leq x, \cdots, X_n \leq x) \\ &= P(X_1 \leq x) \times P(X_2 \leq x) \times \cdots \times P(X_n \leq x) \\ &= [F(x)]^n. \end{aligned}$$

Da mesma forma, a função de distribuição do mínimo é

$$\begin{aligned} P(X_{(1)} \leq x) =& 1 - P(X_1 > x, X_2 > x, \cdots, X_n > x) \\ =& 1 - [(1 - P(X_1 \leq x)) \times \cdots \times (1 - P(X_n \leq x))] \\ =& 1 - [1 - F(x)]^n. \end{aligned}$$

A função de distribuição geral para a k-ésima estatística de ordem (MOOD; GRAYBILL; BOES, 1974) é

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}, \tag{5}$$

e a função densidade de probabilidade correspondente é

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x). \tag{6}$$

Assim, dependendo da função de distribuição F ou função densidade f, o cálculo das esperanças e variâncias das estatísticas de ordem são muito complexas e, em alguns casos, impossíveis de serem obtidas analiticamente. A distribuição conjunta de duas estatísticas de ordem é necessária para a obtenção das covariâncias entre duas diferentes estatísticas de ordem. Assim, a função densidade conjunta da i-ésima e j-ésima estatística de ordem é

$$\begin{aligned} f_{X_{(i)}, X_{(j)}}(x_1, x_2) =& \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(x_1)]^{i-1} \times \\ & \times [F(x_2) - F(x_1)]^{j-i-1} [1 - F(x_2)]^{n-j} f(x_1) f(x_2), \end{aligned} \tag{7}$$

para $1 \leq i < j \leq n$ e $x_1 < x_2$. Os testes de Shapiro-Wilk e Shapiro-Francia são baseados nos valores esperados, variâncias e covariâncias das estatísticas de ordem da distribuição normal padrão. Muitas aproximações foram apresentadas por diversos autores para a obtenção dessas quantidades, que muitas vezes são as razões das limitações desses testes à amostras de tamanhos máximos iguais a 5.000. Essas dificuldades foram, entre outras razões, a motivação para a proposição do presente trabalho.

2.2 O Teste de Shapiro-Wilk

O teste Shapiro-Wilk é um dos mais usados para testar a normalidade, como já foi mencionado anteriormente, pelo seu alto desempenho em poder, avaliado por inúmeros pesquisadores em uma ampla classe de distribuições. Este teste será descrito em detalhes nessa subseção por ter sido escolhido nesse trabalho como o teste de referência. A estatística do teste é baseada em uma razão de dois estimadores distintos da variância populacional. O estimador do numerador é baseado em uma combinação linear de quantidades relacionadas às estatísticas de ordem da distribuição normal. O estimador, utilizado no denominador dessa razão, é obtido de forma convencional. Ademais, várias aproximações para a distribuição da estatística do teste foram apresentadas por Royston (1983, 1993).

Considere que $Z_{(1)}$, $Z_{(2)}$, $Z_{(3)}$, $\ldots$, $Z_{(n)}$ são as estatísticas de ordem de uma normal padrão, ou seja, Z_i's são normais independentes e identicamente distribuídos com média 0 e variância unitária (FERREIRA, 2005). Deve-se observar que, embora os Z_i's sejam independentemente distribuídos, as estatísticas de ordem $Z_{(i)}$'s não são, para $i = 1, 2, 3, \ldots, n$. Isso decorre do fato de as estatísticas de ordem serem as variáveis aleatórias originais, idênticas e independentemente distribuídas, expressas em ordem crescente. A ordenação gera dependência. Considere que $E(Z_{(i)}) = m_i$, Var$(Z_{(i)}) = v_{ii}$ e, ainda, Cov$(Z_{(i)}, Z_{(k)}) = v_{ik}$, $i \neq k$ $= 1, 2, \ldots, n$. As variâncias e covariâncias das estatísticas de ordem podem ser agrupadas em uma matriz $n \times n$ denotada por $\boldsymbol{V} = [v_{ik}]$. As médias são agrupadas em um vetor $n \times 1$ denotado por $\boldsymbol{m}$. Os valores esperados, as variâncias e as covariâncias podem ser obtidos a partir das distribuições das estatísticas de ordem, conforme é descrito por exemplo em Mood, Graybill e Boes (1974). O problema é que os valores só podem ser obtidos numericamente e os cálculos são proibitivos para valores de n moderados ou grandes.

A estatística do teste de Shapiro-Wilk é denotada por W, sendo calculada por

$$W = \frac{\left[\sum_{i=1}^{n} a_i X_{(i)}\right]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

sendo as constantes a_1,$a_2, \ldots, a_n$, elementos do vetor $\boldsymbol{a}$, calculadas como solução de

$$\boldsymbol{a} = \frac{\boldsymbol{m}^\top \boldsymbol{V}^{-1}}{(\boldsymbol{m}^\top \boldsymbol{V}^{-2}\boldsymbol{m})^{\frac{1}{2}}}, \tag{8}$$

em que $X_{(i)}$ é a i-ésima estatística de ordem amostral e $\bar{X}$ a média amostral.

O valor esperado da i-ésima estatística de ordem da normal padrão m_i pode ser muito bem estimado por

$$\tilde{m}_i = \Phi^{-1}\left(\frac{j - 3/8}{n + 1/4}\right), \tag{9}$$

em que $\Phi^{-1}(p)$ é a inversa da função de distribuição da normal padrão avaliada no argumento p, entre 0 e 1. O argumento $(j-3/8)/(n+1/4)$ de (9), $j = 1, 2, \ldots, n$, representa a função de distribuição empírica com correções de continuidade para se obter uma melhor aproximação, sendo dadas por $3/8$, no numerador, e $1/4$, no denominador. O vetor $\boldsymbol{m}$ é, portanto, estimado pelo vetor $(n \times 1)$ $\tilde{\boldsymbol{m}} = [\tilde{m}_i]$.

O vetor de coeficientes $\boldsymbol{a}$ pode ser calculado utilizando várias aproximações, desde as de Shapiro e Wilk (1965), até as de Royston (1982, 1992, 1993, 1995). As aproximações apresentadas por Shapiro e Wilk (1965) e Royston (1982) não são precisas. As aproximações de Royston (1992, 1993) para o vetor de coeficientes $\boldsymbol{a}$ devem ser utilizadas para $n \geq 4$. Esta aproximação é baseada nas médias das estatísticas de ordem que são estimadas pela equação (9). Para $n = 3$, o vetor de coeficientes $\boldsymbol{a}$ é obtido de forma exata.

Os testes de Shapiro-Wilk e Shapiro-Francia diferem na definição dos coeficientes $\boldsymbol{a}$, conforme destacado em Ferreira (2008) e na aproximação da distribuição das suas estatísticas. No teste de Shapiro-Wilk, o vetor $\boldsymbol{a}$ é definido na equação (8) e no caso do teste de Shapiro-Francia, esse vetor de coeficientes é definido por $\boldsymbol{a} = (\boldsymbol{m}^\top \boldsymbol{m})^{-1/2}\boldsymbol{m}$, que pode ser estimado substituindo o vetor $\boldsymbol{m}$ pelo seu estimador $\tilde{\boldsymbol{m}}$, composto pelos elementos $\tilde{m}_i$'s definidos em (9).

Para se estimar o vetor de coeficientes $\boldsymbol{a}$ do teste Shapiro-Wilk, Royston (1993) apresenta o seguinte roteiro. Deve-se inicialmente obter

$$\begin{aligned}\tilde{a}_n =& c_n + 0{,}221157u - 0{,}14798u^2 - 2{,}071190u^3 \\ &+ 4{,}434685u^4 - 2{,}706056u^5 \end{aligned} \tag{10}$$

$$\begin{aligned}\tilde{a}_{n-1} =& c_{n-1} + 0{,}042981u - 0{,}293762u^2 - 1{,}752461u^3 \\ &+ 5{,}682633u^4 - 3{,}582633u^5,\end{aligned}$$

em que c_n e c_{n-1} são o n-ésimo e $(n-1)$-ésimo elementos do vetor $\boldsymbol{c} = (\tilde{\boldsymbol{m}}^\top \tilde{\boldsymbol{m}})^{-1/2}\tilde{\boldsymbol{m}}$ e $u = 1/\sqrt{n}$.

Deve-se obter a quantidade normalizadora

$$\phi = \begin{cases} (\tilde{\boldsymbol{m}}^\top \tilde{\boldsymbol{m}} - 2\tilde{m}_n^2)/(1 - 2\tilde{a}_n^2) & \text{se } n \le 5 \\ (\tilde{\boldsymbol{m}}^\top \tilde{\boldsymbol{m}} - 2\tilde{m}_n^2 - 2\tilde{m}_{n-1}^2)/(1 - 2\tilde{a}_n^2 - 2\tilde{a}_{n-1}^2) & \text{se } n > 5 \end{cases}$$

e finalmente

$$\tilde{a}_j = \frac{\tilde{m}_j}{\sqrt{\phi}}, \tag{11}$$

para $j = 2, \ldots, n-1$ $(n \le 5)$ ou $j = 3, 4, \ldots, n-2$ $(n > 5)$. Deve-se observar que $\tilde{a}_1 = -\tilde{a}_n$ e $\tilde{a}_2 = -\tilde{a}_{n-1}$.

A estatística do teste Shapiro-Wilk é redefinida a partir das estatísticas de ordem $X_{(j)}$ e do vetor de coeficientes $\tilde{\boldsymbol{a}}$ da seguinte forma

$$W = \frac{\left(\sum_{j=1}^{n} \tilde{a}_j X_{(j)}\right)^2}{\sum_{j=1}^{n} \left(X_j - \bar{X}_.\right)^2}. \tag{12}$$

A estatística W não segue uma distribuição normal, mas Royston (1993) propõe a utilização de uma transformação da família Box-Cox, para obter normalidade. De acordo com a proposta de Royston (1993), deve-se obter Y (valor transformado de W) por

$$Y = \begin{cases} -\ln\left[\gamma - \ln(1-W)\right] & \text{se } 4 \leq n \leq 11 \\ \ln(1-W) & \text{se } 12 \leq n \leq 5.000 \end{cases} \tag{13}$$

sendo $\gamma = -2{,}273 + 0{,}459n$.

A variável Y possui média dada por

$$\mu_Y = \begin{cases} 0{,}5440 - 0{,}39978n + 0{,}025054n^2 - 0{,}0006714n^3 & \text{se } 4 \leq n \leq 11 \\ \\ -1{,}5861 - 0{,}31082u - 0{,}083751u^2 + 0{,}0038915u^3 & \text{se } 12 \leq n \leq 5.000 \end{cases} \tag{14}$$

em que $u = \ln(n)$ e desvio padrão

$$\sigma_Y = \begin{cases} \exp\left\{1{,}3822 - 0{,}77857n + 0{,}062767n^2 - 0{,}0020322n^3\right\} & \text{se } 4 \leq n \leq 11 \\ \\ \exp\left\{-0{,}4803 - 0{,}082676u + 0{,}0030302u^2\right\} & \text{se } 12 \leq n \leq 5.000 \end{cases} \tag{15}$$

Assim, sob a hipótese nula de normalidade

$$Z = \frac{Y - \mu_Y}{\sigma_Y} \tag{16}$$

possui aproximadamente distribuição normal-padrão e o valor-p é estimado por valor-$p = 1 - \Phi(Z)$, ou seja, o valor-p corresponde a área da distribuição normal-padrão à direita de Z.

Para o caso particular de $n = 3$, o vetor $\boldsymbol{a}$ é conhecido e exato, sendo dado por $\boldsymbol{a} = [-\sqrt{2}/2, 0, \sqrt{2}/2]^\top$. O valor-$p$ associado a W é calculado de forma exata por

$$\text{valor-}p = 1 - F(w) = 1 - \frac{6}{\pi}\left[\text{arcsen}(\sqrt{w}) - \text{arcsen}\left(\sqrt{\frac{3}{4}}\right)\right]$$

em que a função arcsen deve ser tomada em radianos.

Da comparação do valor-p com o nível nominal de significância α adotado, toma-se a decisão de rejeitar ou não a hipótese nula de normalidade. Convém salientar, como pode ser observado nas aproximações de Royston (1993), que só se deve aplicar o teste para $n \leq 5.000$. Isso se deve ao fato de que as aproximações obtidas por este autor foram validadas apenas para esse limite máximo do tamanho amostral. Em muitas situações práticas da vida real das pesquisas científicas, os pesquisadores se deparam com conjuntos de dados com dimensões maiores do que 5.000 e ficam impossibilitados de usufruírem de um teste com tantas propriedades ótimas como é o teste de Shapiro-Wilk.

2.3 Outros Testes de Normalidade

Nesta seção são destacados mais alguns dos inúmeros testes de normalidade encontrados na literatura. A escolha dos testes a serem apresentados se deu em decorrência de suas importâncias e pelo fato de aparecerem implementados nos principais programas de análise estatística. Uma primeira abordagem para se checar a normalidade dos dados é a utilização de procedimentos gráficos. Embora esses procedimentos não sejam testes formais de normalidade, eles possuem grande apelo para a sua utilização e se constituem em uma primeira aproximação ao problema. A subjetividade é a característica que limita o uso dos procedimentos gráficos como testes formais de normalidade. Por outro lado, desvios de assimetria e de curtose, detecção de "outliers", entre outras possibilidades são prontamente visualizados nesses procedimentos gráficos, o que não acontece com os teste formais de normalidade.

Os procedimentos gráficos são baseados em gráficos probabilísticos. Os gráficos probabilísticos são aqueles em que as estatísticas de ordem amostrais são confrontadas com os valores teóricos ou com os valores esperados das estatísticas de ordem. Em geral, os gráficos denotados por Q-Q plots e P-P plots são utilizados. Os P-P plots são procedimentos gráficos que confrontam as probabilidades observadas e teóricas do modelo considerado, no caso o modelo normal. Sob normalidade a dispersão desses pontos no plano cartesiano se aproxima de uma linha reta de mínimos quadrados. Não-linearidade sistemática indica algum tipo específico de desvio de normalidade.

Testes de normalidade formais são derivados desses procedimentos gráficos. Em geral, esses testes são baseados nos coeficientes de correlação ou determinação entre esses quantis ou percentis. Esse tipo de teste de normalidade é denominado de teste baseado em regressão linear, pois são derivados a partir de uma medida de linearidade.

Nos Q-Q plots ou P-P plots a escolha da abscissa é uma estimativa ou é dependente de uma estimativa da função de distribuição acumulada empírica da distribuição sob a hipótese nula. Essas estimativas são denotadas por posições gráficas. Blom (1958) apresenta uma posição geral das posições gráficas, p_i, para qualquer distribuição sob H_0, dada por

$$p_i = \frac{i - \alpha_i}{n - 2\alpha_i + 1},$$

em que $0 \leq \alpha_i < 1$, depende de n, i e da distribuição subjacente.

Estimativas da média das estatísticas de ordem são obtidas, invertendo a função de distribuição $F_X(x)$ por

$$E\left(X_{(i)}\right) \cong F^{-1}(p_i),$$

em que $E\left(X_{(i)}\right)$ é o valor esperado da i-ésima estatística de ordem. Considerando as pequenas variações dos α_i's, Blom (1958) sugeriu utilizar as posições gráficas por

$$p_i = \frac{i - 0{,}375}{n + 0{,}25} \tag{17}$$

ou

$$p_i = \frac{i - 0{,}5}{n}. \tag{18}$$

O valor mais utilizado para as posições gráficas é

$$p_i = \frac{i}{n+1}, \tag{19}$$

que corresponde ao valor esperado da função de distribuição da normal padrão, ou seja,

$$\begin{aligned} p_i &= E[\Phi^{-1}(Z_{(i)})] \\ &= E[U_{(i)}] \\ &= \frac{i}{n+1}, \end{aligned}$$

em que $U_{(i)}$ é a i-ésima estatística de ordem de uma distribuição uniforme $(0, 1)$.

Os Q-Q plots são os gráficos das estatísticas de ordem amostrais contra algum valor "esperado" da distribuição normal padrão. Sob normalidade estes gráficos devem ser lineares, exceto pela aleatoriedade presente naturalmente. Para obter os Q-Q plots, as observações amostrais são ordenadas, logo $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ são as estatísticas de ordem observadas. Esses valores são quantis empíricos usados na ordenada desses gráficos. Os pontos das abscissas são então escolhidos. Primeiro, define-se as posições gráficas p_i's por: $i/(n+1)$, $(i-0{,}5)/n$ ou $(i-3/8)/(n+1/4)$. Em seguida obtém-se, para cada i, $\tilde{m}_i = \Phi^{-1}(p_i)$ e os pares ordenados $\left(\tilde{m}_i, x_{(i)}\right)$ são plotados.

Os P-P plots consistem em plotar os pares (p_i, P_i), sendo que
$P_i = \Phi((X_{(i)} - \bar{x})/\hat{\sigma})$, em que

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{20}$$

e

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{21}$$

são as estimativas de máxima verossimilhança do desvio padrão e da média populacionais. Como nos Q-Q plots, os pontos plotados devem recair em uma reta de

45^o de inclinação entre $(0, 0)$ e $(1, 1)$. Nesse caso a posição gráfica recomendada é $p_i = i/(n+1)$.

Embora úteis, os procedimentos gráficos têm a subjetividade como limitação. Critérios objetivos podem ser gerados a partir dos gráficos de probabilidade. A utilização de uma medida de linearidade é bem natural. Assim, em geral se usa o coeficiente de correlação ou o coeficiente de determinação. Se os valores calculados dessa medida se aproximam do zero, então esse fato indica que o modelo normal não deve ser apropriado para os conjuntos de valores que se apresentam. Por outro lado, se os valores se aproximam de 1, conclui-se a favor da normalidade dos conjuntos de dados analisados. Níveis críticos são estabelecidos. Valores observados da estatística inferiores a esses limites, indicam que a hipótese nula de normalidade deve ser rejeitada no nível nominal de significância α adotado. Esses testes são denominados de testes de regressão para normalidade.

O teste de Shapiro e Wilk, apresentado na seção 2.2, é um exemplo desse tipo de teste. O teste de Shapiro e Francia (1972) também é um exemplo de testes de regressão. Em ambos os casos, as estatísticas dos testes são baseadas no coeficiente de determinação entre as estatísticas de ordem observadas e um determinado coeficiente, relativo ao valor normalizado do vetor dos valores esperados das estatísticas de ordem da normal padrão. Shapiro e Wilk (1965) consideraram que as estatísticas de ordem normais padrão são correlacionadas, com que o vetor de seus valores esperados $\boldsymbol{m}$ e a matriz de suas covariâncias $\boldsymbol{V}$, foram definidos na seção 2.2. Assim, definiram o vetor $\boldsymbol{m}$ normalizado por

$$\boldsymbol{a} = \frac{\boldsymbol{m}^\top \boldsymbol{V}^{-1}}{\sqrt{\boldsymbol{m}^\top \boldsymbol{V}^{-2} \boldsymbol{m}}}, \tag{22}$$

ou seja $\boldsymbol{a}$ é considerado o melhor vetor de combinações lineares não-viciado para estimar σ^2, ou seja, é o vetor que permite estimar σ^2 por meio do método dos mínimos quadrados generalizados para a regressão das estatísticas de ordem amostrais em seus valores esperados.

Para contornar a necessidade de estimar $\boldsymbol{V}$, Shapiro e Francia (1972) propuseram obter um estimador de σ^2 que fosse a regressão das estatísticas de ordem observadas em seus valores esperados, mas ignoram haver correlações entre as estatísticas de ordem. Assim, o coeficiente $\boldsymbol{a}$ proposto pelos autores é definido por

$$\boldsymbol{a}^* = \frac{\boldsymbol{m}}{\sqrt{\boldsymbol{m}^\top \boldsymbol{m}}}. \tag{23}$$

Como, nem o vetor das médias $\boldsymbol{m}$ e nem a matriz de covariâncias $\boldsymbol{V}$, das estatísticas de ordem normais são conhecidas, expressões aproximadas para estimar (22) e (23) foram apresentadas por Royston (1993). Assim, $\tilde{\boldsymbol{a}}$ e $\tilde{\boldsymbol{a}}^*$ são os vetores estimados dos vetores $\boldsymbol{a}$ e $\boldsymbol{a}^*$, respectivamente. A estatística do teste de Shapiro e Francia é, portanto, dada por

$$W' = \frac{\left[\displaystyle\sum_{i=1}^{n} \tilde{a}_i^* X_{(i)}\right]^2}{(n-1)S^2} \tag{24}$$

em que $\tilde{a}_i$ é o i-ésimo elemento de $\tilde{\boldsymbol{a}}^*$, definido por $\tilde{\boldsymbol{a}}^* = (\tilde{\boldsymbol{m}}^\top \tilde{\boldsymbol{m}})^{-1/2} \tilde{\boldsymbol{m}}$.

Como no caso do teste de Shapiro e Wilk, Royston (1993) também propôs transformar a estatística W' de (24) pela família Box-Cox. Com isso, foi obtida uma excelente aproximação normal para a estatística transformada, viabilizando a realização do teste até amostras de tamanho 5.000. Detalhes dessas aproximações podem ser vistas em Ferreira (2008) e Royston (1993). Outros testes baseados em correlação de funções dos valores esperados e observados são apresentados em Thode Junior (2002), como o procedimento de Filliben (1975) e os percentis de Schilittgen.

Na sequência alguns dos testes comumente utilizados são discutidos. O primeiro deles, que tem importância apenas histórica e didática é o teste qui-quadrado de Pearson. Esse teste fundamenta-se na comparação das frequências teóricas esperadas sob o modelo normal (hipótese nula) com as frequências observadas em uma amostra aleatória. Para aplicá-lo, torna-se necessário agrupar os dados em classes de frequências e, em seguida, calcular as frequências sob modelo normal. A estatística do teste é calculada por

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(FO_i - FE_i)^2}{FE_i},$$

sendo FO_i e FE_i as frequências observada e esperada sob o modelo normal na i-ésima classe e k, o número de classes. A determinação do número de classes pode ser feita utilizando vários critérios. O mais importantes deles é devido a Scott (1979), que determina que k é dado por

$$k \cong 1 + \frac{An^{1/3}}{3{,}49S},$$

em que $A = X_{(n)} - X_{(1)}$ é a amplitude total e S o desvio padrão amostral.

O teste de qui-quadrado possui grande utilidade, principalmente para testar hipóteses nulas considerando outras distribuições além da normal. Entretanto, seu uso para a hipótese de normalidade não é recomendado, principalmente devido a sua falta de poder quando comparado com os testes de normalidades da atualidade. Por essa razão não se recomenda o teste qui-quadrado para verificar a normalidade de um conjunto de dados.

Uma boa parte dos testes de normalidade é baseada na função de distribuição de probabilidade empírica. Esses testes se baseiam na comparação das funções de distribuição empírica (FDE) e hipotética, no caso, a distribuição normal. Existe uma divisão clara entre esses testes. Uma parte deles se baseia na distância máxima entre as funções de distribuições empírica e teórica, como o teste de Kolmogorov-Smirnov, Lilliefors e Kuiper. Outra parte incluem funções quadráticas das diferenças entre as funções de distribuições empírica e teórica. A diferença entre os testes de Komogorov-Smirnov e Lilliefors consiste na especificação ou não dos parâmetros, média e variância, da distribuição normal a ser testada.

A função de distribuição empírica de uma amostra aleatória é denotada por $F_n(x)$ e consiste em uma função escada, definida por

$$F_n(x) = \begin{cases} 0 & \text{se } x < x_{(1)} \\ \dfrac{i}{n} & \text{se } x_{(i)} \leq x < x_{(i+1)} \quad i = 1, 2, \ldots, n-1 \\ 1 & \text{se } x_{(n)} \leq x. \end{cases} \tag{25}$$

Essa função representa a probabilidade acumulada de observações menores ou iguais a x, com crescimento dado por passos de $1/n$. Os testes anteriormente mencionados, são baseados na distância entre a função de distribuição empírica $F_n(x)$ e a função de distribuição sob H_0, $p_{(i)}$. Grandes discrepâncias entre as distribuições empíricas e teóricas indicam que a hipótese nula deve ser falseada. No caso de haver concordância, não haverá evidências para rejeitar a hipótese nula. Para a normal, em geral, $p_{(i)}$ é definido por

$$p_{(i)} = \Phi\left(\frac{x_{(i)} - \bar{x}}{s}\right). \tag{26}$$

O teste de Komogorov-Smirnov é definido a partir do máximo da diferença entre a função de distribuição empírica e a função de distribuição teórica, normal no presente trabalho. Para o cálculo da estatística, deve-se obter as quantidades

$$D^+ = \max_{i=1,\dots,n}\left[\frac{i}{n} - p_{(i)}\right] \qquad \text{e} \qquad D^- = \max_{i=1,\dots,n}\left[p_{(i)} - \frac{i-1}{n}\right].$$

Assim, a estatística do teste é dada por

$$D = \max\{D^+, D^-\}. \tag{27}$$

A função de distribuição da estatística do teste sob a hipótese nula, denominada de distribuição de Kolmogorov, é dada por

$$\Pr(K \leq x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{i=1}^{\infty}e^{-(2i-1)^2\pi^2/(8x^2)}.$$

Essa função de distribuição é utilizada para calcular valores-p, haja vista que sua convergência para um número finito de termos da série é relativamente rápida. Uma limitação do teste de Kolmogorov-Smirnov refere-se ao problema de que $p_{(i)}$, definido em (26), deveria ser calculado utilizando μ e σ^2 e não estimativas desses parâmetros. Se for usada da forma como se encontra em (26), o teste aplicado será apenas aproximado. Se n for muito grande, então a aproximação será considerada adequada.

O teste de Lilliefors (LILLIEFORS, 1967), uma modificação do teste Kolmogorov-Smirnov, utiliza a estatística D de Kolmogorov-Smirnov, mas usa as estimativas dos parâmetros e não os parâmetros populacionais em si, como no caso do teste Kolmogorov-Smirnov. Esse autor, Liliefors, obteve valores críticos para o teste via simulação. Assim, o uso do teste de Lilliefors depende da consulta de tabelas especiais, que foram obtidas por simulação Monte Carlo. Entretanto, segundo Thode Junior (2002), a estatística modificada

$$D^* = (\sqrt{n} - 0{,}01 + 0{,}85/\sqrt{n})D,$$

pode ser comparada com o valor crítico de $0{,}895$ para um teste ao nível nominal de significância de $0{,}05$. Se o valor calculado exceder esse limite crítico, a hipótese nula de normalidade deve ser rejeitada. Valores críticos para α de $0{,}10$, $0{,}025$ e $0{,}01$ são $0{,}819$, $0{,}995$ e $1{,}035$, podem ser utilizados para a decisão de rejeição ou não da hipótese nula de normalidade. Valores de D^* superiores a esses limites indicam que a hipótese nula deve ser rejeitada, considerando o nível de significância α escolhido.

O teste de Kuiper é uma variação desses dois testes. Em vez de utilizar o máximo, como em (27), o autor propôs somar D^+ e D^-. Valores críticos, transformação da estatística do teste e mais detalhes podem ser vistos em Thode Junior (2002). Outros testes de normalidade baseados na função de distribuição empírica baseiam-se na quantidade

$$n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, \psi(F(x)) dF(x),$$

em que $\psi(F(x))$ é uma função peso. Se essa função for igual a unidade $\psi(F(x)) = 1$, o teste resultante é o de Crammer-von Mises. Se $\psi(p) = [p(1-p)]^{-1}$, o teste resultante é o de Anderson-Darling.

O teste Jarque-Bera baseia-se na diferença entre os coeficientes de simetria e curtose dos dados amostrais X_1, X_2, $\ldots$, X_n e teóricos, provenientes da distribuição normal. A estatística do teste é

$$JB = n \left(\frac{\left(\sqrt{b_1}\right)^2}{6} + \frac{(b_2 - 3)^2}{24} \right),$$

em que

$$\sqrt{b_1} = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3}{\left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^{3/2}}$$

e

$$b_2 = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^4}{\left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^{2}}$$

são os coeficientes de assimetria e curtose amostrais, respectivamente e $\bar{X}$, a média amostral. A estatística *JB* tem distribuição assintótica χ^2 com $\nu = 2$ graus de liberdade sob hipótese nula de normalidade. O teste de Jarque-Bera é conhecido por ter boas propriedades para verificar a normalidade, sendo simples de calcular e muito utilizado no contexto de regressão em econometria. Uma limitação do mesmo é sua especificidade para testar normalidade, ou seja, não pode ser generalizado, como os testes baseados em regressão para testar outras distribuições, além da distribuição normal.

Muitos outros testes foram elaborados fazendo modificações ou extensões dos testes pré-existentes. Assim, existem à disposição do pesquisador, uma gama enorme de testes de normalidade, cuja escolha é feita de acordo com as características apresentadas por eles, como taxa de erro tipo I e poder. Assim, dentre os testes que possuam tamanho exato, ou seja, cuja P(Rejeitar $H_0|H_0$ é verdadeira) $= \alpha$, aqueles de maior poder nas diversas variações quanto às distribuições alternativas e aos tamanhos amostrais considerados devem ser preferidos.

Um processo muito usado na validação dos testes de normalidade é a simulação de dados, pois a avaliação analítica é por demais complexa ou impossível de ser realizada. Por esse mecanismos, inúmeros trabalhos foram realizados comparando o desempenho dos testes de normalidade. Deve-se entender que a simulação utiliza algum mecanismo aleatório de gerar amostras provenientes de alguma distribuição de probabilidade (MOORE; WEATHERFORD, 2005) e assim, permitem estudar facilmente as propriedades de qualquer testes estatísticos.

O método Monte Carlo é um desses modelos de simulação que usa a geração de números aleatórios para atribuir valores às variáveis que se deseja investigar. Os números aleatórios podem ser obtidos por meio de algum processo aleatório gerado diretamente do computador, usando funções específicas (LUSTOSA; PONTE; DOMINAS, 2004). A operacionalização desse processo requer o auxílio de alguns métodos matemáticos. Dentre os mais conhecidos e utilizados, segundo Evans e Olson (1998) e Vose (2000), está o método da transformada inversa, que faz uso das propriedades dos números aleatórios uniformes e da função de distribuição acumulada de uma variável aleatória.

Recentemente, Yazici e Yolocan (2007) realizaram uma comparação entre 15 testes de normalidade usando simulações de Monte Carlo. O resultado da comparação foi, segundo eles, que o poder, a facilidade de uso e consequentemente a escolha do teste, depende de vários fatores, entre eles o tipo de distribuição sob H_1, o tamanho da amostra e os valores críticos. Embora os autores critiquem o teste de Shapiro-Wilk pela necessidade de estimar coeficientes especiais e valores críticos especiais, eles afirmam que este teste fornece um indicativo geral de não-normalidade sobre várias alternativas simétricas, não-simétricas, caudas pesada ou leves e sobre todos os tamanhos de amostras utilizados. Outro trabalho, apresentado por Romão, Delgado e Costa (2010), compara 33 testes de normalidade e não aponta um único teste com sendo mais poderoso do que os demais em todos os casos estudados. Eles classificam os testes de acordo com características comuns, em geral, grupos de três ou quatro testes e, em cada grupo, apontam o teste de maior poder. Especificamente entre os testes mais poderosos para distribuições assimétricas e distribuições que são misturas de normais ou normais com presença de *outliers*, os autores recomendam o teste W de Shapiro-Wilk. Quando a natureza da não-normalidade não é conhecida *a priori*, o teste de Shapiro-Wilk está entre os três testes recomendado pelos autores.

2.4 Distribuições de Probabilidades

Nesta seção é feita uma descrição resumida de características e propriedades de algumas distribuições de probabilidades referenciadas neste trabalho. O objetivo de tal descrição é facilitar a consulta, principalmente no que se refere à função densidade, função de distribuição acumulada, médias e variâncias. O poder dos testes de normalidade, avaliados neste trabalho, foi obtido usando tais distribuições. Para se avaliar o poder é necessário aplicar o teste sob H_1. Assim, simulações de dados sob distribuições não-normais são necessárias para realizar tal avaliação. Foram escolhidas as distribuições *t-Student*, com graus de liberdade $\nu = 1$ e $\nu = 30$, gama, com parâmetros 0,5 e 1, lognormal padrão e beta $(1, 1)$.

2.4.1 Distribuição t de Student

A distribuição t de Student tem notável aplicação em estatística e recebeu este nome em homenagem ao pesquisador W. S. Gosset que realizou uma importante publicação a seu respeito em 1908 usando o pseudônimo de *Student*. A distribuição t é definida pela razão entre uma variável aleatória normal padronizada, $N(0{,}1)$, por outra originada da obtenção da raiz quadrada de uma variável aleatória independente qui-quadrado dividida pelos seus graus de liberdade. Se Z for uma variável aleatória independente qui-quadrado com ν graus de liberdade, a variável aleatória

$$X = \frac{Z}{\sqrt{\frac{U}{\nu}}}$$

possui distribuição t de Student. A função densidade de probabilidade de X é:

$$f(x) = \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1 + \dfrac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} \tag{28}$$

Sabe-se que se uma amostra aleatória de uma distribuição normal com média μ e variância σ^2, então, a variável aleatória

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

possui distribuição normal padrão e

$$U = \frac{(n-1)S^2}{\sigma^2}$$

possui distribuição qui-quadrado com $\nu = n - 1$ graus de liberdade. Como Z e U são independentes, então, a variável aleatória

$$T = \frac{Z}{\sqrt{\frac{U}{(n-1)}}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{((n-1)\sigma^2)}{(n-1)}}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

possui distribuição t de Student com $\nu = n - 1$ graus de liberdade.

A distribuição t de Student se aproxima da distribuição normal, à medida que o grau de liberdade aumenta, tendendo para infinito. Além disso, nota-se que o gráfico de sua função densidade possui caldas mais leves que a distribuição normal, e esse fato fica mais evidenciado à medida que os graus de liberdade diminuem (FERREIRA, 2009).

A média e a variância da distribuição t de Student dependem de ν e são dadas por $\mu = 0$ para $\nu > 1$ e $\sigma^2 = \dfrac{\nu}{\nu - 2}$ para $\nu > 2$.

2.4.2 Distribuição gama

A distribuição gama é uma generalização da distribuição exponencial e tem diversas aplicações na estatística, como o tempo de vida de equipamentos, tempo de retorno de mercadorias com falhas e testes de confiabilidade. A família de densidades gama é definida por dois parâmetros reais λ e *r*, sendo *r* o parâmetro de forma e λ o parâmetro de escala. Uma variável aleatória X com distribuição gama, para o caso de r inteiro, pode ser interpretada como aquela correspondente ao tempo de espera, necessário para a ocorrência do r-ésimo sucesso. Quando $r = 1$, a distribuição gama se reduz na distribuição exponencial (FERREIRA, 2009). A função densidade da gama é:

$$f(x) = \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x}, \tag{29}$$

em que: $\lambda > 0$, $r > 0$, $x > 0$ e $\Gamma(\bullet)$ é a função gama definida por

$$\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt \tag{30}$$

em que $r > 0$.

A importante propriedade de $\Gamma(r)$, é obtida a partir de (30) por meio de integração por partes.

$$\Gamma(r) = (r-1)\Gamma(r-1). \tag{31}$$

Se r é um número inteiro positivo, então, 31, se reduz a

$$\Gamma(r) = (r-1)!$$

Se r for um número ímpar inteiro então vale a relação

$$\Gamma\left(r + \frac{1}{2}\right) = \frac{1 \times 3 \times 5 \times \ldots \times (2r-1)}{2^r}\sqrt{\pi},$$

pois, particularmente, $\Gamma(1/2) = \sqrt{\pi}$.

Assim como para a distribuição normal, a função de distribuição acumulada gama não tem uma derivação explícita. A forma geral da função de distribuição de probabilidade acumulada gama é

$$F(x) = \frac{\lambda}{\Gamma(r)} \int_0^{x} (\lambda z)^{r-1} e^{-\lambda z} dz. \tag{32}$$

A média e a variância da distribuição gama são, respectivamente,

$$\mu = \frac{r}{\lambda}$$

e

$$\sigma^2 = \frac{r}{\lambda^2}.$$

2.4.3 Distribuição Lognormal

A distribuição lognormal é obtida a partir da distribuição normal, $N(\mu, \sigma^2)$, pela transformação dada por

$$X = e^Y.$$

A variável aleatória X, cuja densidade é a lognormal, possui função densidade dada por

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(ln(x)-\mu)^2},$$

em que $-\infty < x < +\infty$, $\sigma > 0$ e $x > 0$.

A sua função de distribuição de probabilidade acumulada também não tem forma explícita e os seus valores são obtidos através de métodos numéricos. A média e variância de X são dadas, respectivamente, por:

$$\mu_X = e^{\mu + \frac{1}{2}\sigma^2}$$

e

$$\sigma_X^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}.$$

2.4.4 Distribuição beta

A família de distribuições beta tem inúmeras aplicações na probabilidade e estatística e assume várias formas de acordo com as diferentes escolhas de seus parâmetros. A variável aleatória X com distribuição beta com parâmetros α e β possui densidade dada por

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{33}$$

em que $\alpha > 0$, $\beta > 0$ e $0 \leq x \leq 1$.

Uma importante relação da função beta, $B(\alpha, \beta)$, com a função gama é

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx. \tag{34}$$

A distribuição beta equivale à distribuição uniforme se $\alpha = \beta = 1$. A média e a variância da distribuição beta são

$$\mu = \frac{\alpha}{\alpha+\beta}$$

e

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

A função de distribuição acumulada beta não possui uma forma explícita e está apresentada na sequência na sua forma geral para $0 < x < 1$. Essa função é conhecida por função beta incompleta, sendo

$$F(x) = \frac{1}{B(\alpha, \beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt. \tag{35}$$

3 MATERIAL E MÉTODOS

No presente trabalho propôs-se um teste computacionalmente intensivo baseado em Monte Carlo para a hipótese nula

$$H_0 : \ f(x) = \phi_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{36}$$

contra a alternativa

$$H_1 : \ f(x) \neq \phi_0(x).$$

Deve ficar claro, que a hipótese é do tipo composta, ou seja, somente a forma da distribuição é importante em (36), mas os parâmetros são desconhecidos. Inicialmente é considerado que uma amostra aleatória X_1, X_2, ..., X_n, de tamanho n, supostamente obtida sob H_0. Em seguida são obtidas as estatísticas de ordem $X_{(1)})$, $X_{(2)}$, ..., $X_{(n)}$, ou seja, a amostra é ordenada de forma crescente.

O presente teste tem como fundamentação a obtenção do coeficiente de determinação entre as estatísticas de ordem observadas e os valores esperados, assumindo que o modelo normal determinado em (36). Como o coeficiente de determinação é uma quantidade invariante em relação ao efeito de escala, então os parâmetros da normal para obtenção dos valores esperados das estatísticas de ordem são $\mu = 0$ e $\sigma^2 = 1$, ou seja, é considerado a normal padrão. Os valores esperados $E\left(Z_{(i)}\right) = m_i$, $i = 1$, 2, ..., n, possuem cálculo difícil de ser obtido numericamente, por envolver distribuições de estatísticas de ordem. As aproximações apresentadas na expressão (9) foram utilizadas para estimá-los.

No próximo passo fez-se o uso de simulações Monte Carlo, considerando um número grande de repetições N_{MC}, para obtenção da distribuição nula da estatística do teste. Para isso amostras de tamanho n com distribuições normais foram geradas e o coeficiente de determinação entre as estatísticas de ordem observadas e os valores esperados das mesmas, estimados utilizando (9). Repetindo o processo N_{MC} vezes e armazenando os coeficientes de determinação r_i^2's, a distribuição nula é obtida. O teste foi concretizado calculando-se o valor-p, pela proporção de coeficientes de determinação da distribuição nula de Monte Carlo que foram inferiores ao coeficiente de determinação obtido na amostra original. O julgamento, considerando um nível nominal de significância α, deve confrontar o valor-p e α. As simulações para a construção do teste de normalidade foram feitas usando o software estatístico R (R DEVELOPMENT CORE TEAM, 2008). Uma função foi implementada para realizar tal teste e encontra-se apresentada no Apêndice A.

3.1 A Proposta

Pode-se descrever o Teste Monte Carlo de Normalidade Univariado (TNUMC) por meio dos seguintes passos:

1. obter a amostra aleatória X_1, X_2, ..., X_n, a ser submetida ao teste;

2. obter as estatísticas de ordem da amostra original, ou seja, a amostra original é ordenada de forma crescente: $X_{(1)})$, $X_{(2)}$, ..., $X_{(n)}$;

3. estimar os valores esperados das estatísticas de ordem da distribuição normal padrão por

$$E(Z_{(j)}) \cong \widetilde{m}_j = \Phi^{-1}(p_j),$$

   em que

$$p_j = \frac{j - 3/4}{n + 1/4},$$

   correspondente a $X_{(j)}$, $j = 1, 2, \ldots, n$;

4. calcular o coeficiente de determinação amostral r_c^2, entre $X_{(j)}$ e $\tilde{m}_j$, $j = 1$, 2, ..., n;

5. gerar, via simulação Monte Carlo, N_{MC} amostras aleatórias de tamanho n na distribuição normal padrão Z_1, Z_2, ..., Z_n;

6. obter as estatísticas de ordem da distribuição normal padrão, $Z_{(1)}$, $Z_{(2)}$, ..., $Z_{(n)}$, em cada uma das N_{MC} amostras simuladas;

7. calcular o coeficiente de determinação r_i^2, entre $Z_{(j)}$ e $\tilde{m}_j$, $j = 1, 2, \ldots, n$, para a i-ésima amostra simuladas, gerando N_{MC} valores;

8. armazenar os N_{MC} simulados r_i^2's juntamente com o valor obtido na amostra original r_c^2;

9. Cálculo do valor-p

$$\text{valor} - p = \frac{1}{N_{MC}+1} \sum_{i=1}^{N_{MC}+1} I(r_i^2 \leq r_c^2),$$

   em que $I(\bullet)$ é a função indicadora;

10. retornar o valor-p e o valor da estatística r_c^2 da amostra original.

Confrontando o valor-p com α, faz-se o julgamento a respeito da normalidade dos dados, isto é, se o valor-p for menor ou igual ao valor de significância nominal α estabelecido pelo pesquisador, deve-se rejeitar a hipótese de normalidade dos dados analisados. Caso contrário, não haverá evidências significativas ao nível nominal de significância α para rejeitar a hipótese de normalidade. A função R, denominada *normMCTeste.R*, está apresentada no Apêndice A.

3.2 Validação do Teste

O teste Monte Carlo de normalidade univariado (TNUMC) apresentado foi validado em duas etapas. Na primeira etapa fez-se a comparação dos dois testes em relação ao erro tipo I. Na segunda etapa avaliou-se o poder do teste proposto. Na primeira, foram realizadas simulações sob H_0 dada em (36), ou seja, foram simuladas amostras de tamanho n da distribuição normal. Sem perda de generalidade foram simulados dados de uma normal padrão com média 0 e variância unitária. Para fins de comparação, o teste de Shapiro-Wilk (TNUSW) foi aplicado a cada amostra simulada. A descrição do teste de Shapiro-Wilk foi feita na subseção 2.2. Foi fixado um valor para o nível de significância α. O confronto dos valores-p com o nível de significância nominal em cada amostra foi realizado em ambos os testes.

Um número N de simulações Monte Carlo de validação foram realizadas sob H_0, sendo ambos os testes aplicados. Todas as simulações foram realizadas no programa R (R DEVELOPMENT CORE TEAM, 2008). A proporção de rejeições da hipótese nula de normalidade foi computada para o total de N simulações. Estes valores são estimativas dos tamanhos reais dos testes. Para avaliar o efeito da aleatoriedade foi aplicado um teste binomial exato para a hipótese nulas de que o nível de significância do teste é igual ao valor nominal, ou seja, para a hipótese que o teste é exato.

Foram considerados níveis nominais de significância iguais a $0{,}10$, $0{,}05$ e $0{,}01$ e tamanhos amostrais n iguais a 5, 10, 30, 100, 500, 5.000 e 10.000. O número N de simulações Monte Carlo para a validação do teste foi igual 2.000. O número de simulações Monte Carlo N_{MC} para a aplicação do teste foi também igual a 2.000.

Testes binomiais exatos, considerando o nível nominal de significância de 1%, para as hipóteses $H_0 : \alpha = 10\%$ versus $H_1 : \alpha \neq 10\%$, $H_0 : \alpha = 5\%$ versus $H_1 : \alpha \neq 5\%$ e $H_0 : \alpha = 1\%$ versus $H_1 : \alpha \neq 1\%$ foram aplicados. Se a hipótese nula for rejeitada e o valor observado das taxas de erro tipo I forem consideradas significativamente $(p < 0{,}01)$ inferior ao nível nominal, o teste deve ser considerado conservativo; se as taxas de erro tipo I forem consideradas significativamente $(p < 0{,}01)$ superiores ao nível nominal, o teste deve ser considerado liberal; e se os valores observados das taxas de erro tipo I não diferirem significativamente $(p > 0{,}01)$ do nível nominal, o teste deve ser considerado exato. Considerando que y representa o número de hipóteses nula de normalidade rejeitadas nas N simulações Monte Carlo para o nível de significância nominal α, então a estatística do teste é obtida, considerando a relação entre as distribuições F e binomial, com probabilidade de sucesso $p = \alpha$, por

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

que, sob a hipótese nula, segue a distribuição F com $\nu_1 = 2(N-y)$ e $\nu_2 = 2(y+1)$ graus de liberdade. Se $F_c \leq F_{0{,}005}$ ou $F \geq F_{0{,}995}$, a hipótese nula deve ser rejeitada no nível nominal de significância de 1%, em que $F_{0{,}005}$ e $F_{0{,}995}$ são quantis da distribuição F com ν_1 e ν_2 graus de liberdade.

Na segunda etapa de validação, avaliou-se o poder do teste proposto, sendo o seu poder comparado com o do teste de normalidade de Shapiro-Wilk. Para isso foram simuladas N amostras considerando os mesmos tamanhos amostrais anteriormente definidos para a avaliação do erro tipo I. Também foram considerados os mesmos níveis nominais de significância retromencionados. Para avaliar o poder foram simuladas amostras sob H_1, ou seja, sob outra distribuição diferente da normal. Optou-se por escolher algumas distribuições de probabilidades comumente encontradas nas pesquisas.

A primeira distribuição considerada foi a t de *Student* com $\nu = 1$ e 30 graus de liberdade. A escolha da t se deu em decorrência de sua forma semelhante à da normal, ou seja, é uma distribuição simétrica centrada em 0. Se os graus de liberdade são pequenos, a distribuição se afasta da normal, embora com a mesma forma; se os graus de liberdade forem grandes, a distribuição se aproxima consideravelmente da normal, e espera-se que o poder do teste seja pequeno. Também foi escolhida a distribuição gama padrão, ou seja, com um único parâmetro, por ser uma distribuição assimétrica à direita. Escolheu-se uma gama com parâmetro $\alpha = 0{,}5$ e com $\alpha = 1{,}5$. Finalmente, considerou-se uma distribuição beta com parâmetros $\alpha = 1$ e $\beta = 1$, que corresponde a distribuição uniforme $(0, 1)$. Essa escolha se deveu ao fato de a distribuição uniforme ser platicúrtica e simétrica. Assim, espera-se contemplar uma série de possibilidades na avaliação do poder em relação à forma das distribuições consideradas sob H_1.

Para computar o poder, do mesmo modo que realizado para o erro tipo I, as taxas de rejeições de H_0 foram calculadas nas N simulações Monte Carlo de cada configuração considerada. O teste Shapiro-Wilk também foi aplicado, como na primeira etapa de validação, para fins de comparação. A exceção se deu para amostras de tamanho $n = 10.000$, uma vez que o teste concorrente é limitado à amostras de tamanhos máximos de 5.000. Essa limitação do teste concorrente foi o grande motivador para a construção do teste Monte Carlo, cuja limitação é unicamente devida a memória disponível nos computadores para lidarem com grandes amostras. Todas as simulações foram realizadas no R (R DEVELOPMENT CORE TEAM, 2008), utilizando as funções de geração de amostras aleatórias das distribuições anteriormente mencionadas. O teste de normalidade de Shapiro-Wilk foi aplicado utilizando a função *shapiro.test*.

4 RESULTADOS E DISCUSSÃO

Nesta seção foi feita a discussão dos resultados obtidos e a validação do TNUMC, por meio da comparação do seu desempenho com o do teste de normalidade Shapiro-Wilk (TNUSW). A seção foi dividida em duas partes: uma, para o erro tipo I e, outra, para o poder dos testes. Na parte referente ao erro tipo I, a comparação dos resultados dos dois testes de normalidade univariados mencionados anteriormente foi realizada para os níveis nominais de significância α, fixados em 10%, 5% e 1%, considerando diferentes tamanhos de amostra.

Na subseção referente ao poder dos testes, as comparações foram feitas para os mesmos valores de α e n considerados na primeira subseção, embora as simulações tivessem sido feitas sob distribuições não-normais. Foram consideradas diferentes distribuições não-normais univariadas. A escolha do teste de normalidade Shapiro-Wilk foi feita devido ao fato de o mesmo ter apresentado resultados superiores ao de outros testes de normalidade, de acordo com o trabalho Shapiro e Wilk (1965), no qual este teste foi proposto, além de ter apresentado melhores resultados entre seus competidores em Romão, Delgado e Costa (2010) e Yazici e Yolocan (2007).

4.1 Erro tipo I

Em um teste de hipóteses, comete-se um erro tipo I ao se rejeitar a hipótese nula quando esta é verdadeira. A probabilidade de se cometer esse erro, se o teste for exato, é denotada por α e recebe o nome de nível de significância do teste ou tamanho do teste. No trabalho aqui apresentado, a avaliação do erro tipo I foi feita observando-se a proporção dos valores simulados sob normalidade que foram superiores significativamente $(P < 0{,}01)$ ao nível de significância nominal adotado. Essa avaliação é necessária, uma vez que a distribuição da estatística do teste sob H_0 é obtida empiricamente. Considera-se que um teste controla adequadamente o erro tipo I, quando a taxa de erro observada não diferir significativamente do nível nominal de significância. Tal significância foi dada por teste binomial, buscando-se, assim, minimizar o efeito dos erros de Monte Carlo e avaliações subjetivas. A ideia básica da avaliação do erro tipo I, no caso do teste proposto nesta pesquisa, foi, portanto, verificar o percentual de amostras simuladas, em um total de 2.000 simulações, que não diferiram significativamente do nível de significância adotado em cada teste e para cada tamanho de amostra.

As probabilidades de se cometer os erros do tipo I e II são inversamente proporcionais, sendo que o complemento da taxa de erro tipo II é o poder do teste. Isso implica que uma baixa probabilidade de se incorrer no erro tipo I está associada a uma alta probabilidade de se cometer o erro tipo II. Novamente, se as taxas de erro tipo I observadas não forem significativamente diferentes do nível nominal de significância considerado, o teste é *exato*, porém, se elas forem inferiores ao nível de significância nominal considerado, o teste será *conservativo*, e se elas forem superiores ao nível nominal de significância, o teste será *liberal*.

O processo de simulação Monte Carlo, sob normalidade, foi repetido 2.000 vezes e os testes TNUMC e TNUSW, foram aplicados em cada amostra ao nível nominal de significância, α, fixado em 10%, 5% e 1%. As taxas de erro tipo I foram calculadas para cada teste e estas foram comparadas com o nível de significância nominal.

Os resultados obtidos com as simulações, tanto para o teste Monte Carlo de normalidade univariado (TNUMC) quanto para o teste de normalidade Shapiro-Wilk (TNUSW), estão apresentados na Tabela 1. Ambos os testes apresentaram, de maneira geral, controle adequado do erro tipo I, sendo suas taxas de erro tipo I equivalentes. O teste binomial para verificar o rigor do controle do erro tipo I para cada teste, em cada configuração simulada foi aplicado e, nenhum resultado diferiu significativamente, ou seja, $(P > 0{,}01)$, dos correspondentes valores nominais de significância. Tanto o teste de Shapiro-Wilk quanto o teste proposto foram considerados exatos para todos os casos considerados e tiveram desempenho equivalentes quanto ao controle do erro tipo I.

Em todos os níveis de significância nominal os valores estimados do tamanho dos testes podem ser considerados idênticos aos valores nominais α's, em todos os tamanhos de amostras, mesmo no caso de $n = 5$, que é uma amostra muita pequena. Esse fato, para o TNUMC pode ser visto como uma forma de confirmação da validade da simulação realizada. Isso porque o teste possui natureza Monte

Carlo, ou seja, a distribuição da estatística é obtida sob H_0 e de forma exata, exceto pelo tamanho da amostra dessa distribuição nula, que é finito. Destaca-se também que o teste proposto foi aplicado em uma amostra de tamanho 10.000, que é uma situação em que é impossível de se aplicar o TNUSW. Esse fato destaca uma das principais vantagens da presente proposta. Novamente, as taxas de erro tipo I estão estreitamente relacionadas aos valores nominais de significância adotados.

Tabela 1 Taxas de erro tipo I dos testes de normalidade univariada TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% em função do tamanho da amostra n

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC†	TNUSW†	TNUMC†	TNUSW†	TNUMC†	TNUSW†
5	0,0995	0,1060	0,0495	0,0530	0,0135	0,0125
10	0,1050	0,0940	0,0530	0,0495	0,0100	0,0090
30	0,1035	0,0965	0,0485	0,0430	0,0130	0,0100
100	0,0995	0,0960	0,0475	0,0425	0,0065	0,0080
500	0,1015	0,1040	0,0560	0,0610	0,0095	0,0130
5000	0,0900	0,0770	0,0445	0,0405	0,0115	0,0105
10000	0,0915	-	0,0485	-	0,0110	-

† todos os resultados foram não significativamente ($P > 0{,}01$) diferentes dos valores nominais.

4.2 Poder dos testes

Em um teste de hipóteses, comete-se erro tipo II quando a hipótese nula não é rejeitada, dado que ela é falsa. A probabilidade de se cometer este erro é denotada por β. O poder de um teste é o complemento, em relação à unidade, desta probabilidade β, ou seja, é a probabilidade de se rejeitar a hipótese nula quando ela é falsa. Sendo assim, a determinação e a comparação do poder dos dois testes, TNUMC e TNUSW, foram realizadas observando o percentual de rejeições da hipótese nula, quando amostras de distribuições não-normais foram simuladas. O teste que detectar uma maior quantidade de amostras não-normais é o que tem maior poder em relação ao outro. Para facilitar a comparação, os valores obtidos nas simulações foram apresentados em tabelas para cada distribuição considerada.

Nesta seção estão apresentados os desempenhos, em relação aos valores de poder, do novo teste de normalidade proposto, TNUMC, e do teste de normalidade de Shapiro-Wilk ao nível de significância $\alpha = 10\%$, $\alpha = 5\%$ e $\alpha = 1\%$. Foram considerando as seguintes distribuições não-normais: *t-Student* com graus de liberdade $\nu = 1$ e $\nu = 30$, gama com parâmetros 0,5 e 1, lognormal e beta $(1, 1)$.

O poder dos testes TNUMC e TNUSW, considerando a distribuição não-normal *t-Student* com grau de liberdade $\nu = 1$ para $\alpha = 10\%$, $\alpha = 5\%$ e $\alpha = 1\%$ é apresentado na Tabela 2. Observa-se que ambos os testes têm valores de poder bastante próximos, sendo que nenhum deles é expressivamente superior ao outro. A exceção ocorre em amostras pequenas, como $n = 5$ e $n = 10$, em que o poder do teste de Shapiro-Wilk é superior. Isso ocorreu com maior evidência, principalmente, para $\alpha = 1\%$. Em amostras de tamanho $n = 30$, os dois testes têm valores de poder praticamente iguais. Em amostras de tamanhos maiores, isto é, $n = 100$, $n = 500$ e $n = 5.000$, os desempenhos em poder dos teste são exatamente iguais para todos os valores de α. Nesse últimos casos, ambos os testes apresentam 100% de poder, ou seja, detectam o total das amostras não-normais simuladas, no caso, a distribuição *t-Student* com $\nu = 1$.

Tabela 2 Poder dos testes TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% relativo à distribuição t com $\nu = 1$ grau de liberdade, considerando diferentes tamanhos de amostras

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC	TNUSW	TNUMC	TNUSW	TNUMC	TNUSW
5	0,3240	0,3570	0,2360	0,2830	0,1010	0,1740
10	0,6430	0,6535	0,5675	0,5900	0,4245	0,4810
30	0,9680	0,9690	0,9600	0,9595	0,9120	0,9245
100	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
500	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10000	1,0000	-	1,0000	-	1,0000	-

Em relação a amostra de tamanho $n = 10.000$, o TNUMC apresentou 100% de poder, o que é uma grande vantagem em relação ao TNUSW, pois este é restrito a amostras de tamanho menores do que 5.000. Essa característica foi a principal motivação para a realização do presente trabalho. Ademais, esperava-se que a utilização da distribuição Monte Carlo da estatística do teste propiciasse um teste mais poderoso que o seu concorrente. Isso em decorrência de o TNUSW utilizar aproximações para a distribuição nula da estatística do teste.

Os desempenhos em relação ao poder dos TNUMC e TNUSW considerando a distribuição *t-Student* com grau de liberdade $\nu = 30$ para $\alpha = 10\%$, $\alpha = 5\%$ e $\alpha = 1\%$ são mostrados na Tabela 3. Para ambos os testes o poder aumenta à medida que o tamanho da amostra aumenta, como é preconizado pela teoria dos testes de hipóteses, que corrobora a adequação das simulações realizadas. Observa-se novamente que nenhum dos testes apresentou desempenho em relação ao poder expressivamente superior ao do outro. Inclusive, pode-se observar que há algumas oscilações em relação a diferentes valores de n e de níveis de significância, entre a liderança de um teste em relação ao outro quanto ao desempenho. Em amostras de tamanhos $n = 5$, $n = 10$ e $n = 30$, o TNUSW apresentou desempenho superior, embora com pequena diferença a seu favor, exceto para o nível de significância nominal $\alpha = 5\%$, em amostras de tamanhos $n = 5$ e $n = 30$ e para $\alpha = 10\%$ com $n = 30$. Em amostras de tamanho $n = 100$, os testes tiveram desempenhos equivalentes, com pequena superioridade de desempenho do TNUMC. Em amostras de tamanhos $n = 500$ e $n = 5.000$ o TNUMC mostrou-se mais poderoso do que o TNUSW.

Convém salientar que os valores de poder, em alguns casos de pequenas amostras, são praticamente idênticos aos valores nominais de significância e na medida que as amostras aumentam, o poder afasta do valor nominal de significância. Isso ocorre como consequência da semelhança que a distribuição *t-Student* com $\nu = 30$ graus de liberdade apresenta em relação à distribuição normal. O melhor desempenho do teste proposto sobre seu concorrente direto em grandes amostras é uma qualidade que deve ser destacada e valoriza o principal enfoque dado ao trabalho de estender a aplicabilidade dos testes de normalidade para amostras maiores que 5.000. Outro aspecto que merece ser destacado é que a distribuição sob H_1 considerada nesses dois primeiros casos é simétrica, situação teoricamente mais difícil de ser detectada por um teste de normalidade, em função de a normal possuir também essa característica.

Na Tabela 4 estão apresentados os valores de poder dos TNUMC e TNUSW relativos à distribuição gama de parâmetro $0{,}5$ para os três níveis de significância nominais adotados. Nota-se que o TNUMC apresenta um desempenho de poder

Tabela 3 Poder dos testes TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% relativo à distribuição t com $\nu = 30$ graus de liberdade

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC	TNUSW	TNUMC	TNUSW	TNUMC	TNUSW
5	0,0975	0,1025	0,0540	0,0445	0,0095	0,0105
10	0,1055	0,1150	0,0595	0,0610	0,0140	0,0150
30	0,1270	0,1240	0,0770	0,0430	0,0210	0,0220
100	0,1465	0,1425	0,0865	0,0805	0,0310	0,0255
500	0,2395	0,2175	0,1595	0,1420	0,0595	0,0595
5000	0,7750	0,7440	0,6800	0,6445	0,4640	0,4395
10.000	0,9575	-	0,9305	-	0,8120	-

expressivamente superior em amostras pequenas, isto é pra $n = 5$ e $n = 10$. Em amostras de tamanho $n = 30$, os dois testes têm valores de poder iguais para α de 10% e 5% e praticamente equivalentes para $\alpha = 1\%$ e, em amostras superiores a 30, o desempenho em relação ao poder dos dois testes é absolutamente igual. A exceção ocorre para amostras maiores do que 5.000, em que o TNUSW não pode ser aplicado.

A distribuição gama é assimétrica à direita. Assim, há um maior poder de ambos os testes de normalidade em relação às mesmas configurações apresentadas para a distribuição t, simétrica, exceto para os casos em que os valores já tenham atingido 100%. Da mesma forma, ampliaram-se as vantagens relativas em pequenas amostras do teste proposto sobre o principal teste de normalidade existente, o teste de Shapiro-Wilk.

Tabela 4 Poder dos testes TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% relativo à distribuição gama - 0,5

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC	TNUSW	TNUMC	TNUSW	TNUMC	TNUSW
5	0,5615	0,4300	0,4215	0,3045	0,1940	0,1405
10	0,8780	0,8105	0,8020	0,7270	0,5715	0,5135
30	1,0000	1,0000	1,0000	0,9995	0,9975	0,9940
100	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
500	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10000	1,0000	-	1,0000	-	1,0000	-

Na tabela 5, os valores de poder para dois testes são apresentados, considerando a distribuição univariada gama de parâmetro $1{,}5$, relativos ao três níveis de significâncias nominais adotados. O desempenho de poder do TNUMC mostra-se novamente expressivamente maior do que o desempenho do TNUSW em amostras de tamanhos $n = 5$, $n = 10$ e $n = 30$. Para amostras de tamanhos $n = 100$, $n = 500$ e $n = 5.000$ os dois testes comparados apresentam resultados de poder iguais a 100%.

Tabela 5 Poder dos testes TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% relativo à distribuição gama - 1,5

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC	TNUSW	TNUMC	TNUSW	TNUMC	TNUSW
5	0,2790	0,1880	0,1725	0,1105	0,0600	0,0390
10	0,5790	0,4400	0,4470	0,3170	0,2000	0,1500
30	0,9665	0,9320	0,9215	0,8790	0,7430	0,6785
100	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
500	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10000	1,0000	-	1,0000	-	1,0000	-

Os desempenhos quanto ao poder dos testes TNUMC e TNUSW considerando a distribuição lognormal padrão podem ser observados e comparados na Tabela 6. Nota-se que em tamanhos de amostras menores do que $n = 30$, o TNUMC apresentou novamente resultados de desempenho de poder expressivamente superiores ao do TNUSW. Os dois testes têm valores de poder equivalentes em amostras de tamanho $n = 30$, não havendo nenhum com poder excessivamente superior ao outro e os testes possuem poderes iguais entre si e iguais a 100% em amostras de tamanhos maiores do que 30.

Novamente, pode-se observar que a lognormal é uma distribuição assimétrica à direita. Em amostras pequenas, nesse tipo de distribuição, pelo menos pelas distribuições consideradas no presente trabalho, o desempenho do TNUMC foi muito melhor. Isso pode ser observado pelas grandes diferenças de poder encontradas entre os dois testes nesses tamanhos de amostras, para todas as distribuições assimétricas à direita consideradas e para todos os valores dos níveis nominais de significância adotados. Em grandes amostras, os dois testes apresentaram desempenhos equivalentes, sendo um forte indício de equivalência assintótica. Entretanto, não existem extensões para a aplicação do TNUSW em amostras superiores a 5.000.

Tabela 6 Poder dos testes TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% relativo à distribuição lognormal padrão

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC	TNUSW	TNUMC	TNUSW	TNUMC	TNUSW
5	0,4665	0,3440	0,3415	0,2525	0,1610	0,1140
10	0,7910	0,7075	0,7005	0,6130	0,4800	0,4160
30	0,9990	0,9975	0,9980	0,9955	0,9755	0,9690
100	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
500	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10000	1,0000	-	1,0000	-	1,0000	-

Para a distribuição beta$(1, 1)$, os resultados de desempenho de poder estão apresentados na Tabela 7. Pode-se observar um melhor desempenho do TNUSW em amostras de tamanhos $n = 5$, $n = 10$, $n = 30$ e $n = 100$, sendo que a superioridade é bem mais expressiva em amostras de tamanho $n = 30$. Ambos possuem 100% de desempenho em relação ao poder em amostras de tamanhos $n = 500$ e $n = 5.000$. A distribuição beta$(1, 1)$ é na verdade a uniforme $(0, 1)$, que é platicúrtica e simétrica. Assim, espera-se que o teste proposto possua desempenho pior que o seu concorrente em distribuições simétricas e platicúrticas. Novas distribuições simuladas sob H_1 devem ser consideradas para que se faça uma investigação mais ampla sobre os desempenhos relativos do novo teste proposto e do teste de Shapiro-Wilk.

Tabela 7 Poder dos testes TNUMC e TNUSW para os níveis nominais de significância de 10%, 5% e 1% relativo à distribuição beta - $(1,1)$

	$\alpha = 0{,}10$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	TNUMC	TNUSW	TNUMC	TNUSW	TNUMC	TNUSW
5	0,1105	0,1240	0,0535	0,0600	0,0135	0,0115
10	0,1115	0,1595	0,0495	0,0790	0,0060	0,0095
30	0,3310	0,5590	0,1610	0,3635	0,0185	0,0810
100	0,9960	1,0000	0,9720	0,9985	0,7480	0,9510
500	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10000	1,0000	-	1,0000	-	1,0000	-

5 CONCLUSÕES

Em relação ao controle de erro tipo I, os dois testes apresentaram resultados equivalentes, uma vez que tanto para o TNUMC quanto para o TNUSW, os valores das taxas de rejeição da hipótese nula não diferem significativamente $(P > 0{,}01)$ dos valores nominais de significância, em todas as situações envolvendo os diferentes tamanhos de amostras simulados. Assim, o teste proposto é considerado exato.

Em relação ao poder, o TNUMC mostra-se mais poderoso do que o teste Shapiro-Wilk para a maioria das distribuições não-normais estudadas considerando os diferentes níveis de significância nominal fixados.

Outra conclusão importante é a possibilidade do TNUMC poder ser aplicado em amostras maiores do que 5.000, fato que o coloca em vantagem em relação a outros testes de normalidade.

REFERÊNCIAS

BLOM, G. **Statistical estimates and transformed beta variables**. New York: J. Wiley, 1958. 176 p.

EVANS, J.; OLSON, D. **Introduction to simulation and risk analysis**. Washington: Prentice Hall, 1998. 392 p.

FERREIRA, D. F. **Estatística básica**. Lavras: UFLA, 2005. 664 p.

FERREIRA, D.F. **Estatística Básica**. Lavras: UFLA, 2009. 664 p.

FERREIRA, D.F. **Estatística Multivariada**. Lavras: UFLA, 2008. 576 p.

FILLIBEN, J. J. The probability plot correlation coefficient test for normality. **Tecnometrics**, Washington, v. 17, n. 1, p. 11-117, 1975.

LILLIEFORS, H. On the Kolmogorov-Smirnov test for normality witn mean and variance unknown. **Journal of the American Statistical Association**, New York, v. 67, p. 399-402, 1967.

LUSTOSA, P. R. B.; PONTE, V. M. R.; DOMINAS, W. R. Simulação. In: CORRAR, L. J.; THEÓPHILO, C. R. (Coord.). **Pesquisa operacional para decisão em contabilidade e administração**: contabilometria. São Paulo: Atlas, 2004. p. 242-284.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. Singapore: McGraw-Hill International, 1974. 480 p.

MOORE, J. H.; WEATHERFORD, L. R. **Tomada de decisões em administração com planilhas eletrônicas**. Porto Alegre: Bookman, 2005. 633 p.

OLIVEIRA, I. R.; FERREIRA, D. F. Multivariate extension of chi-squared normality test. **Journal of Statistical Computation and Simulation**, New York, v. 80, n. 5, p. 513-526, 2010.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2008. Disponível em: <http://www.R-project.org/>. Acesso em: 10 dez. 2012.

ROMÃO, X.; DELGADO, R.; COSTA, A. An empirical power comparison of univariate goodness-of-fit of normality. **Journal of Statistical Computation and Simulation**, London, v. 80, n. 5, p. 545-591, 2010.

ROYSTON, J. P. Approximating the Shapiro-Wilk's test for non-normality. **Statistics and Computing**, London, v. 2, n. 1, p. 117-119, 1992.

ROYSTON, J. P. An extension of Shapiro and Wilk's test for normality to large samples. **Applied Statistics - Journal of the Royal Statistical Society - Series C**, London, v. 31, n. 2, p. 115-124, 1982.

ROYSTON, J. P. A remark on algorithm: the test for normality. **Applied Statistics - Journal of the Royal Statistical Society - Series C**, London, v. 44, n. 4, p. 547-551, 1995.

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

ROYSTON, J. P. A toolkit for testing for non-normality in complete and censored samples. **The Statistician**, London, v. 42, n. 1, p. 37-43, 1993.

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

SHAPIRO, S. S.; FRANCIA, R. S. Approximate analysis of variance test for normality. **Journal of the American Statistical Association**, New York, v. 67, p. 215-216, 1972.

SHAPIRO, S. S.; WILK, M.B. An Analysis of Variance Test for Normality: complete samples. **Biometrika**, London, v.52, n.3/4, p.591-611, Dec. 1965.

THODE JUNIOR, H. **Testing for Normality**, New York: M. Decker, 2002. 154 p.

VOSE, D. **Risk analysis**: a quantitative guide. 2^{nd} ed. Sussex: J. Wiley, 2000. 752 p.

YAZICI, B.; YOLACAN, S. A comparison of various tests of normality. **Journal of Statistical Computation and Simulation**, London, v. 80, n. 2, p. 175-183, 2007.

APÊNDICES

APÊNDICE A - Rotinas R

```
# função para aplicação do teste de normalidade Monte Carlo
# univariado
normMCTeste  <- function(x, NMC = 2000)
{
  n <- length(x) #tamanho da amostra
  x <- sort(x)   #ordena a amostra, obtendo as Est. de Teste
  pj <- ((1:n) - 3/4) / (n + 1/4) # fator de correção, j=1,...,n
  qj <- qnorm(pj) # estimadores dos valores esperados de x(j)
  r2c <- cor(x, qj)^2 #coef. de det. amostral entre x(j) e qj
  plot(qj, x, xlab = paste("Valores Esperados N(", 0, 1,")",
  sep = ""), ylab = paste("Estatísticas de Ordem Observadas",
  sep = ""))
  reg <- lm(x~qj)
  abline(reg$coefficients[1], reg$coefficients[2])
  Z <- matrix(rnorm(n * NMC), NMC, n)  #gera 2000 a.a. de tam. n
  #da normal padrão

  Z <- t(apply(Z, 1, sort))   # est. de ordem z(1),..., z(n)
  r2MC <- (apply(Z, 1, cor, qj))^2  # coef. de det. z(j) e qj
  r2MC <- c(r2MC, r2c)     # armazena r2MC
  valor.p <- length(r2MC[r2MC <= r2c]) /( NMC + 1 )#valor-p
  return(list(r2 = r2c, valor.p=valor.p))
}
```

```
# função para validar o teste e comparar com o Shapiro Wilk
# resultado em uma matriz com valores significância nas colunas
# testes nas linhas; retorna as taxas de rejeições de cada teste
# em cada nível de significância nominal
# argumento op = 1, distribuição normal, op = 2, distr. t com
#  v = 1 GL,op = 3, dist t com v = 30; op = 4 dist gama(0,5);
# op = 5, gama(1,5), op = 6, lognormal; op = 7, beta(1, 1)
validTNU    <- function(n, N = 2000, NMCT=2000, op = 1)
{
   result <- matrix(0, 2, 3)
   colnames(result) <- c("0,10", "0,05", "0,01")
   rownames(result) <- c("TNUMC", "TNUSW")
```

```
  for (i in 1:N)
  {
    if (op==1) x <- rnorm(n) else
     if (op==2) x <- rt(n, 1) else
       if (op==3) x <- rt(n, 30) else
        if (op==4) x <- rgamma(n, 0.50) else
         if (op==5) x <- rgamma(n, 1.5) else
          if (op==6) x <- rlnorm(n) else
           if (op==7) x <- rbeta(n, 1.0, 1.0)
    res <- normMCTeste(x, NMCT)
    if (res$valor.p <= 0.10) result[1,1] = result[1,1] + 1.0/N
    if (res$valor.p <= 0.05) result[1,2] = result[1,2] + 1.0/N
    if (res$valor.p <= 0.01) result[1,3] = result[1,3] + 1.0/N
    res <- shapiro.test(x)
    if (res$p.value <= 0.10) result[2,1] = result[2,1] + 1.0/N
    if (res$p.value <= 0.05) result[2,2] = result[2,2] + 1.0/N
    if (res$p.value <= 0.01) result[2,3] = result[2,3] + 1.0/N
  }
  return(result)
}
```