DESEMPENHO DE TESTES DE NORMALIDADE MULTIVARIADOS AVALIADO POR SIMULAÇÃO MONTE CARLO

Monte Carlo evaluation of the performance of multivariate normality tests

Narjara Fonseca Cantelmo[1], Daniel Furtado Ferreira[2]

RESUMO

Neste trabalho, objetivou-se avaliar a performance do teste multivariado de normalidade de Shapiro-Wilk implementado no **R** comparando o seu desempenho com os testes de assimetria e curtose de Mardia (1970, 1974, 1975) utilizando simulação Monte Carlo. Foram mensuradas e comparadas as taxas de erro tipo I e poderes dos testes. Pode-se concluir que o teste de Shapiro-Wilk multivariado do programa **R**, função *mshapiro.test* do pacote mvnormtest, tem fraco desempenho (liberal) e não é recomendado para uso rotineiro.

Termos pra indexação: Shapiro-Wilk, assimetria, curtose, poder, erro tipo I.

ABSTRACT

This work aimed to evaluate the performance of the multivariate normality test of Shapiro-Wilk implemented in **R** in the library *mvnormtest* and to compare it with the asymmetry and kurtosis normality test proposed by Mardia (1970, 1974, 1975) using Monte Carlo simulation. The multivariate normality test of Shapiro-Wilk is not recommended for regular use.

Index terms: Shapiro-Wilk, asymmetry, kurtosis, power, type I error rate.

(Recebido em 19 de março de 2007 e aprovado em 10 de julho de 2007)

INTRODUÇÃO

A suposição de normalidade dos dados amostrais ou experimentais é uma condição exigida para a realização de muitas inferências válidas a respeito de parâmetros populacionais. Vários dos diferentes métodos de estimação e testes de hipóteses existentes foram formulados sob a suposição de que a amostra aleatória tenha sido extraída de uma população normal. Da mesma forma que ocorre no caso univariado, têm-se inferências multivariadas sobre parâmetros que são vetores ou matrizes, pois a estatística multivariada é lida com observações simultâneas de várias variáveis. Os testes de hipóteses e os métodos de estimação são, em geral, baseados na suposição de normalidade multivariada da amostra aleatória.

A necessidade de testar a hipótese de normalidade multivariada fica evidenciada quando o pesquisador pretende avaliar se as condições pressupostas para a validade da inferência que irá realizar foram atendidas. A existência de um teste adequado, com propriedades ótimas sempre foi questionada. Um destes testes de normalidade é baseado nos desvios de assimetria e curtose (MARDIA, 1970, 1974, 1975). Alguns problemas podem ser destacados desta abordagem. A princípio, a não rejeição da hipótese de distribuição simétrica e mesocúrtica não garante que a distribuição seja realmente normal. Existem contra-exemplos de distribuições não-normais simétricas e mesocúrticas. A segunda dificuldade é a propriedade assintótica das estatísticas dos testes. As distribuições destas estatísticas são apenas assintoticamente válidas, o que dificulta a aplicação dos testes em conjuntos de dados de tamanhos relativamente pequenos.

Um teste multivariado para normalidade foi proposto por Royston (1983) generalizando o teste univariado de Shapiro-Wilk para o caso multivariado. A justificativa apresentada para a proposição deste teste refere-se ao fato de que muitos testes multivariados possuem distribuições matematicamente intratáveis sob a hipótese nula para a estatística. No entanto, as propriedades do teste de Shapiro-Wilk generalizado para o caso multivariado não foram avaliadas por simulação, principalmente para dimensões dos vetores das variáveis aleatórias maiores do que três.

O programa **R** (R DEVELOPMENT CORE TEAM, 2006) tem tido grande impacto no meio científico e por ter código fonte aberto, tem recebido contribuições de pesquisadores de todo o mundo. O pacote *mvnormtest* e a

[1]Aluna do curso de Agronomia – Departamento de Ciências Exatas/DEX – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – nacantelmo@hotmail.com – Bolsista CNPq
[2]Pós-Doutorado, Professor – Departamento de Ciências Exatas/DEX – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – danielff@ufla.br – Bolsista CNPq

função *mshapiro.test* deste pacote possibilitam ao usuário aplicar o teste de normalidade multivariada de Shapiro-Wilk que não trata da extensão multivariada de Royston (1983). Esta função é baseada na generalização multivariada do teste proposto por Domanski (1998) que se baseia em buscar uma combinação linear das *p* variáveis originais e aplicar o teste de Shapiro-Wilk nesta nova variável. O procedimento de Royston (1983) prevê a estimação da estatística W de Shapiro-Wilk para cada uma das variáveis, sendo a estatística final do teste baseada na soma dos seus valores. É utilizada uma transformação da estatística e a correlação entre as variáveis é utilizada para obter os graus de liberdade da distribuição de qui-quadrado resultante (ROYSTON, 1983).

Uma forma de avaliar o desempenho de um teste é mensurar tanto as taxas de erro tipo I, em diferentes condições da hipótese nula de normalidade, quanto o poder do teste, simulando amostras sob a hipótese alternativa de não-normalidade. Um teste ideal, apesar de não existir, seria aquele que não rejeitasse para nenhuma amostra observada a hipótese nula verdadeira e rejeitasse 100 % das vezes as hipóteses nulas falsas. Como isso não ocorre em situações reais, busca-se um teste que mantenha as taxas de erro tipo I menores ou iguais a um valor nominal de probabilidade escolhido, diga-se, α , e que tenha o maior poder possível. O poder é a probabilidade de rejeitar a hipótese nula quando ela é realmente falsa.

A motivação para este trabalho surgiu a partir do interesse de avaliar o desempenho do teste multivariado de normalidade de Shapiro-Wilk implementado no **R**. Assim, foi proposto este trabalho objetivando-se comparar o desempenho do teste multivariado de normalidade de Shapiro-Wilk com o teste de assimetria e curtose de Mardia utilizando simulação Monte Carlo.

MATERIAL E MÉTODOS

Duas estratégias foram consideradas neste trabalho. A primeira teve o intuito de avaliar as taxas de erro tipo I dos testes de Shapiro-Wilk e de assimetria e curtose de Mardia. A segunda foi delineada para avaliar o poder destes testes. Em ambos os casos foi utilizada simulação Monte Carlo. Em cada simulação foram aplicados os testes de normalidade em um nível nominal pré-estabelecido de significância, sendo verificado se a hipótese nula foi ou não rejeitada. Caso tenha sido rejeitada a hipótese nula e a amostra seja da distribuição normal multivariada, foi cometido um erro do tipo I. Da mesma forma se a hipótese nula for rejeitada e a amostra foi obtida de uma população não-normal, uma decisão correta foi tomada. Este processo, em cada caso, foi repetido 2.000 vezes e a proporção de decisões incorretas no primeiro caso é a taxa de erro tipo I empírica e, no segundo caso, a proporção de decisões corretas é o poder empírico. Os valores da taxa de erro tipo I empírica foram comparados com o valor nominal por meio de um intervalo de confiança para proporções. Também foram comparadas as taxas de erro e o poder dos dois testes aplicados. Na sequência são descritas todas as etapas necessárias para a simulação das amostras normais e não-normais, aplicação dos testes e cômputo das taxas de erro tipo I e poder. Todas as simulações foram realizadas no programa **R**.

Taxas de Erro Tipo I

Foram simuladas amostras aleatórias normais multivariadas de tamanho n no espaço p-dimensional dadas por $X_1, X_2, ..., X_n$. O vetor aleatório $X_j = \left[X_{j1}, X_{j2}, ..., X_{jp}\right]'$ de dimensões $p \times 1$ tem densidade normal multivariada dada por:

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

em que

Σ é matriz de covariância e

μ é o vetor de médias $p \times 1$ populacionais.

Para simular um vetor aleatório X_j da distribuição apresentada em (1) foi inicialmente obtido o fator de Cholesky (Γ) da matriz Σ, tal que $\Sigma = \Gamma\Gamma'$. Posteriormente foi simulado um vetor Z_j, composto de elementos independentes de uma distribuição normal univariada, sendo que Z_j tem distribuição $N_p(0, I)$. Utilizando uma transformação linear do vetor dada por $X_j = \Gamma Z_j + \mu$, obteve-se o vetor desejado com distribuição dada por (1). Repetiu - se este processo inúmeras vezes para formar a amostra aleatória.

O teste de normalidade multivariado de Shapiro-Wilk foi aplicado utilizando-se a função *mshapiro.test* do pacote *mvnormtest* do programa ***R***. Inicialmente foi aplicada uma transformação linear do vetor linear X_j dada por $Y_j = CX_j$, em que o vetor C foi determinado a partir de $C = W^{-1}R_k$, sendo W^{-1} a inversa da matriz de somas de quadrados e produtos amostral e R_k o vetor da unidade amostral k contendo os desvios de cada observação para a sua respectiva média. Sendo R a matriz de desvios de cada observação para a sua média de dimensão n x p, a unidade amostral k, selecionada foi considerada a linha da matriz R correspondente ao valor máximo da diagonal da matriz $RW^{-1}R'$. As realizações da variável Y são ordenadas de forma crescente obtendo-se $Y_{(1)},\ Y_{(2)},\ \cdots,\ Y_{(n)}$ e o teste de Shapiro-Wilk foi aplicado a estes valores. Inicialmente foi obtida a estatística de Shapiro-Wilk por

$$W = \frac{\left(\sum_{j=1}^{n} \tilde{a}_j Y_{(j)}\right)^2}{\sum_{j=1}^{n}\left(Y_j - \bar{Y}_.\right)^2}$$

em que os coeficiente $\tilde{a}_j$ são os estimadores apresentados por Royston (1983) para os coeficientes populacionais a_j relacionadas a j-ésima estatística de ordem da distribuição normal padrão. O vetor n x 1destes coeficientes populacionais são definidos por $a = V^{-1}m/\left(m'V^{-2}m\right)$, sendo m o vetor n x 1 de médias das estatísticas de ordem da normal padrão e V a matriz $n \; x \; n$ de variâncias e covariâncias das n estatísticas de ordem da normal padrão. Os valores-p são obtidos a partir de uma transformação Box-Cox da estatística do teste, utilizando-se assim uma aproximação normal (ROYSTON, 1983).

Os testes de normalidade baseados nos desvios de assimetria e curtose foram aplicados de acordo com os procedimentos descritos na sequência. Foram utilizadas as definições dos coeficientes multivariados de Mardia (1970, 1974) para assimetria $\beta_{1p} = E\left[(X-\mu)'\Sigma^{-1}(Y-\mu)\right]^3$ e para curtose $\beta_{2p} = E\left[(X-\mu)'\Sigma^{-1}(X-\mu)\right]^2$, para uma distribuição p-variada qualquer, para se definir os valores hipotéticos sob normalidade. Assim, sob a hipótese nula de normalidade multivariada estes coeficientes correspondem a $\beta_{1p} = 0$ e $\beta_{2p} = p(p+2)$, respectivamente.

Inicialmente foram estimados os coeficientes de assimetria e curtose utilizando os seguintes estimadores:

$$\hat{\beta}_{1p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g^3_{ij} \qquad \text{e} \qquad \hat{\beta}_{2p} = \frac{1}{n}\sum_{i=1}^{n} g^2_{ii}$$

em que $g_{ij} = (X_i - \bar{X}.)'S_n^{-1}(X_j - \bar{X}.)$, $\bar{X}. = \frac{1}{n}\sum_{j=1}^{n} X_j$ é o vetor de médias amostrais e $S_n = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X}.)(X_j - \bar{X}.)'$ é o estimador viesado de máxima verossimilhança da matriz de covariâncias amostrais.

Em seguida foi aplicado o teste de simetria, cuja estatística, dada por,

$$k_1 = \frac{n\hat{\beta}_{1p}}{6}$$

possui distribuição assintótica de qui-quadrado com $p(p+1)(p+2)/6$, graus de liberdade sob a hipótese nula $H_0: \beta_{1p} = 0$.

Da mesma forma foi aplicado o teste para distribuição mesocúrtica. A estatística do teste dada por

$$k_2 = \frac{\hat{\beta}_{2p} - p(p+2)}{\sqrt{8p(p+2)/n}}$$

possui distribuição assintótica normal padrão sob a hipótese nula .

Se na aplicação dos testes, a hipótese nula for rejeitada a um nível nominal α então, a distribuição dos dados é considerada não-normal.

A proporção de rejeições da hipótese nula foi computada para o teste multivariado de Shapiro-Wilk e para o teste baseado em desvios de simetria e de curtose. As taxas de erro obtidas foram comparadas entre si e com o valor nominal α adotado.

Poder dos testes

Foram simuladas amostras aleatórias de distribuições não-normais para avaliar o poder dos testes em rejeitar a hipótese nula que por construção é falsa. O mesmo procedimento para simulação descrito na seção anterior foi aplicado. A diferença consistiu na determinação de diferentes distribuições para os elementos Z_i do vetor Z. Foram consideradas amostras aleatórias das distribuições t multivariada com ν =1, 2 e 30 graus de liberdade. Foram utilizadas as funções geradoras de variáveis aleatórias do programa **R** para realizar tal tarefa.

Serão aplicados os testes de normalidade descritos anteriormente em 2.000 repetições deste procedimento. A proporção de rejeições da hipótese nula falsa foi computada para o teste multivariado de Shapiro-Wilk e para o teste baseado em desvios de simetria e de curtose. Estas proporções mediram os poderes dos testes, que foram comparados entre si.

RESULTADOS E DISCUSSÃO

Na Tabela 1 são apresentadas as taxas de erro tipo I dos testes de normalidade baseados em desvios de assimetria *(ta)* e curtose *(tc)* e do teste de Shapiro-Wilk *(ts)* em função da correlação *(* ρ *)*, tamanho da amostra *(n)* e número de variáveis *(p)*. Pode-se observar que, praticamente, não há efeito do ρ nas taxas de erro tipo I, embora exceções existam, como, por exemplo, na situação de *n* = 100 e *p* = 2, considerando o teste *tc*, que apresentou significância com $\rho = 0,5$ e $\rho = 0,9$, mas não apresentou diferenças significativas em relação ao nível nominal α com $\rho = 0$.

O tamanho da amostra afeta expressivamente as taxas de erro tipo I de todos os testes. No caso dos testes baseados nos desvios de assimetria e curtose é esperado que quanto maior for a amostra melhor serão os testes, uma vez que suas estatísticas são distribuídas assintoticamente como qui-quadrado e normal,respectivamente. O teste *ta* com amostras pequenas tendeu a ser conservativo e passou, em amostras grandes, a apresentar tamanho do teste igual ao valor nominal de 5%. O efeito do aumento do número de variáveis no teste *ta* foi o de reduzir as taxas de erro tipo I para um mesmo *n*. Assim, para os casos estudados o teste de assimetria apresentou taxas significativamente inferiores ou iguais ao valor nominal. Quanto maior for o número de variáveis recomenda-se aumentar o tamanho da amostra.

O teste de curtose *(tc)* com amostras pequenas tendeu a ser conservativo, considerando-se um menor número de variáveis envolvidas e tendeu a ser liberal com um maior número de variáveis. Para grandes amostras o mesmo efeito foi observado, exceto que para valores intermediários de *p* observaram-se taxas de erro similares ao valor nominal $\alpha = 5\%$. Este teste apresenta o inconveniente de ser liberal quando o número de variáveis é grande em relação ao tamanho da amostra utilizado. Assim, por exemplo, para *n*=50, *p*=20 e $\rho = 0,5$ a taxa de erro tipo I é igual a 60,40%.

O teste de normalidade de Shapiro-Wilk multivariado apresentou o pior desempenho de todos. As taxas de erro tipo I foram sempre superiores ao valor nominal, independentemente de *p* ou de *n*, chegando a atingir 100%. O aumento de *n* provocou uma redução do erro tipo I quando *p* estava fixado. De forma similar, o aumento de número de variáveis, fixado o valor de *n*, provocou um grande aumento na taxa do erro tipo I. A taxa de crescimento do erro tipo I em função de *p* foi muito grande, fazendo com que houvesse casos com taxa de 100%. Assim, o teste de normalidade de Shapiro-Wilk multivariado implementado no programa **R** é extremamente liberal e a sua aplicação é desaconselhada.

Para as taxas de erro tipo I dos três testes, considerando $\alpha = 1\%$ (resultados não apresentados), praticamente as mesmas constatações foram feitas. O teste de assimetria foi conservativo ou de mesmo tamanho do nível nominal. O teste de curtose foi conservativo para pequenas amostras e pequenos valores de *p* e liberal com grandes valores de *n* e *p*. O teste de normalidade de Shapiro-Wilk multivariado foi sempre liberal e com pior desempenho quando *p* era grande com relação a *n*. Não houve efeito da estrutura de correlação, da mesma forma como foi observado para $\alpha = 5\%$ (Tabela 1). Para $\alpha = 1\%$ o teste de normalidade de Shapiro-Wilk multivariado implementado no programa **R** também não deve ser recomendado.

Poder

Na Tabela 2 estão apresentados os valores de poder dos testes *ta*, *tc* e *ts* em função de diferentes valores de *n*, *p* e ρ para $\alpha = 5\%$. Foi considerada a distribuição *t* de *Student* com $\nu = 1$ grau de liberdade, criando uma situação de grande afastamento da normalidade para $\alpha = 5\%$ (Tabela 2). Conforme aconteceu com as taxas de erro tipo I praticamente não houve efeito da correlação r nos valores de poder. Para o teste de Shapiro-Wilk os valores de poder foram quase sempre elevados (>80%) e na maioria das vezes iguais a 100%. Infelizmente, estes valores não são reais, pois o teste não controlou as taxas de erro tipo I, sendo sempre liberal. Para este teste o poder cresceu rapidamente com o aumento do número de variáveis atingindo 100%, mesmo em situações de pequenas amostras como, por exemplo, para *n*=10 e *p*=6. Os testes de assimetria e curtose nesta mesma situação apresentaram valores de poder nulos, independentemente de ρ. Assim, não se aconselha a comparação do poder do teste *ts* com os demais e não se recomenda seu uso para testar a normalidade multivariada.

O teste de desvios de curtose (*tc*) apresentou valores de poder bastante elevados para *n* superior a 20, mas foi severamente afetado pelo número de variáveis. Quanto maior for o valor de *p* para $n \leq 20$, menor é o poder deste teste. Este resultado é surpreendente, pois o erro tipo I nestas condições foi significativamente (*p*<0,01) superior aos valores nominais (Tabela 1) de 5% e também de1% (resultados não apresentados). Ocorreu exatamente o contrário, o poder foi relativamente mais baixo que o esperado. Para $n \geq 50$ este efeito não foi observado, embora o erro tipo I continuasse significativamente superior a α.

O teste baseado em desvios de simetria (*ta*) apresentou valores de poder elevados (>90%) para $n \geq 20$ e para quaisquer valores de *p*. O teste *ta* para $n \geq 50$ foi bastante conservativo e com isso eram esperados valores

Tabela 1 –Taxas de erro tipo I dos três testes de normalidade multivariados: teste de assimetria (*ta*), teste de curtose (*tc*) e teste de Shapiro-Wilk (*ts*) e $\alpha = 5\%$.

		$\rho = 0,0$			$\rho = 0,5$			$\rho = 0,9$		
n	***p***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***
10	2	0,25	0,00	16,80	0,55	0,00	17,85	0,50	0,00	17,40
	4	0,00	0,00	79,85	0,00	0,00	77,70	0,00	0,00	79,40
	6	0,00	0,00	100,00	0,00	0,00	100,00	0,00	0,00	100,00
20	2	2,55	0,70	14,35	2,55	0,40	11,50	3,25	0,85	13,85
	4	1,70	0,30	58,25	1,45	0,35	59,50	1,60	0,15	56,40
	6	0,60	1,75	97,05	0,35	1,75	97,70	0,45	1,60	97,30
	10	0,00	19,40	100,00	0,00	20,45	100,00	0,00	20,15	100,00
50	2	3,35	1,90	10,50	4,20	1,80	11,20	4,30	1,75	11,25
	4	3,25	1,80	43,60	3,95	1,75	43,70	3,95	2,20	47,40
	6	3,15	2,85	89,20	3,55	2,95	87,80	3,30	2,90	88,40
	10	2,15	8,05	100,00	1,35	9,60	100,00	2,35	9,00	100,00
	20	0,30	59,55	100,00	0,10	60,40	100,00	0,00	60,80	100,00
100	2	4,20	3,40	9,85	4,65	3,00	10,35	4,40	2,80	10,00
	4	5,65	2,15	40,20	5,10	3,40	38,80	4,55	2,60	38,60
	6	4,80	4,40	82,55	4,50	4,40	80,50	4,35	3,50	81,80
	10	3,80	6,80	100,00	3,15	8,20	100,00	3,10	5,95	100,00
	20	1,85	26,30	100,00	1,35	25,70	100,00	1,80	26,65	100,00
200	2	4,80	3,85	11,20	5,70	3,70	9,90	4,40	3,50	9,10
	4	5,15	4,65	37,20	5,65	4,55	38,25	5,10	3,50	35,85
	6	4,75	4,45	77,10	5,10	3,85	76,65	5,20	4,35	78,40
	10	5,20	5,85	100,00	4,95	5,85	100,00	4,90	5,70	100,00
	20	3,50	14,60	100,00	3,35	14,05	100,00	2,75	13,15	100,00

baixos de poder, porém isso não ocorreu. A exceção ocorreu para amostras pequenas como *n*=10 e principalmente maior número de variáveis ($p \geq 4$) Com *n*=10, *p*=2 e $\rho = 0$ o poder do teste *ta* foi de 54,45%, o que pode ser considerado relativamente satisfatório. Para *p*=6, com o mesmo tamanho de n, o poder foi nulo. Com amostras de tamanho *n*=200 todos os valores de poder foram iguais a 100%.

Os valores de poder para os testes em função de *n*, *p* e ρ com $\alpha = 1\%$, considerando a distribuição *t* de *Student* com $\nu = 1$ teve o mesmo padrão de resposta apontado anteriormente para $\alpha = 5\%$ (Tabela 2). O que ocorre, no entanto, como já era esperado, é que os valores de poder são menores do que aqueles observados para $\alpha = 5\%$. Em todos os casos em que $n \geq 20$, o poder foi superior a 85%, exceto para *tc* com *n*=20 e *p*=10 em que os valores eram de aproximadamente de 30%. Houve uma tendência do poder do teste *ta* ser superior ao do teste *tc*. Com *n*=200 todos os valores foram iguais a 100%.

Na Tabela 3 estão apresentados os valores de poder dos três testes estudados para $\alpha = 1\%$ e 5% em função de *p, n* e ρ. Apenas valores de *n*=20 e 200 foram considerados, uma vez que os padrões anteriores se repetiram nesta situação na qual foi considerada a distribuição *t* com $\nu = 30$ graus de liberdade. Os valores de poder dos testes *ta* e *tc* foram muito pequenos $(\leq \alpha)$ para *n*=20. Houve algumas exceções para o teste *tc* dados pela situação de *p*=10. Novamente pode-se observar que o teste *ts* foi o mais poderoso e que o poder aumentou com aumento de *p*. Neste caso particular de *n*=20, o poder de 100% foi atingido com *p*=10 mesmo para $\alpha = 1\%$. Novamente deve-se afirmar que os grandes valores de poder encontrados para o teste *ts* não devem ser considerados, pois o seu tamanho é significativamente e expressivamente superior ao valor nominal α.

Para *n*=200 os valores de poder foram classificados na seguinte ordem *ta*<*tc*<*ts* para todas as combinações de *p*, ρ e α. A comparação dos valores de poder correspondentes às distribuições *t* com $\nu = 1$ e 30 graus de liberdade (Tabelas 2 e 3) permite que se observe uma drástica redução com o aumento dos graus de liberdade

Tabela 2 – Poder dos três testes de normalidade multivariados: teste de assimetria (*ta*), teste de curtose (*tc*) e teste de Shapiro-Wilk (*ts*) para a distribuição *t* com $\nu = 1$ grau de liberdade e $\alpha = 5\%$.

		$\rho = 0,0$			$\rho = 0,5$			$\rho = 0,9$		
n	***p***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***
	2	54,45	16,85	82,00	54,90	16,20	82,40	54,30	16,00	82,45
10	4	20,50	0,00	99,00	21,05	0,00	99,20	21,60	0,00	99,50
	6	0,00	0,00	100,00	0,00	0,00	100,00	0,00	0,00	100,00
	2	92,40	91,70	97,50	91,90	91,05	97,20	93,00	93,00	97,00
20	4	98,40	96,20	99,90	98,55	96,80	100,00	98,55	96,75	99,95
	6	99,40	96,15	100,00	99,35	95,80	100,00	99,25	96,05	100,00
	10	97,70	51,45	100,00	97,55	52,30	100,00	97,20	51,20	100,00
	2	99,75	100,00	99,95	99,50	100,00	99,95	99,50	100,00	100,00
	4	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
50	6	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	10	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	20	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	2	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	4	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
200	6	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	10	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	20	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

de 1 para 30. Isso ocorreu por causa da aproximação da distribuição *t* multivariada à normal multivariada quando $\nu \to \infty$. A superioridade do teste *tc* em relação ao teste *ta* provavelmente possa ser explicada pela aproximação mais lenta da mesocurtose na medida que o parâmetro ν aumenta.

Considerações finais

O programa **R** (R DEVELOPMENT CORE TEAM, 2006) tem alcançado grande popularidade entre os pesquisadores das mais diferentes áreas da ciência. Este programa possui código fonte aberto e recebe contribuições de profissionais do mundo todo. Estas contribuições são denominadas de pacotes e muitas vezes são avaliadas e recomendadas pelo grupo de desenvolvimento do programa.

O teste de Shapiro-Wilk generalizado para o caso multivariado é um desses pacotes. Este teste teve uma pobre performance nos estudos de simulação realizados neste trabalho. O erro tipo I deste teste foi elevado e muito superior aos valores nominais adotados. Esse resultado foi avaliado em várias situações e as taxas de erro tipo I deste teste cresceram na medida que o número de variáveis cresceu, atingindo 100% com valores moderados de *p*. O poder, por outro lado, foi quase sempre superior ao dos testes concorrentes. Isso, no entanto, não pode ser considerado, pois o tamanho do teste foi superior ao nível nominal α.

Os testes de desvios de assimetria e de curtose mostraram-se como alternativas eficientes para avaliar a hipótese nula de normalidade. Por se tratarem de testes assintóticos, foi recomendado por Santos & Ferreira (2003) utilizar $n \geq 50$ para assimetria e $n \geq 100$ para curtose.

Em função desses resultados, recomendam-se avaliações científicas rotineiras de procedimentos de estimação e de testes de hipóteses implementados como pacotes do programa **R**. Para o caso particular do teste de Shapiro-Wilk multivariado não se recomenda a utilização da função *mshapiro.test* do pacote *mvnormtest* deste programa.

Tabela 3 – Poder dos três testes de normalidade multivariados: teste de assimetria (*ta*), teste de curtose (*tc*) e teste de Shapiro-Wilk (*ts*) para a distribuição *t* com $\nu = 30$ graus de liberdade e $\alpha = 5\%$ e 1%.

		$\alpha = 5\%$								
		$\rho = 0,0$			$\rho = 0,5$			$\rho = 0,9$		
n	***p***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***
20	2	4,45	0,90	16,15	5,15	1,65	17,55	5,10	1,20	17,50
	4	3,40	0,20	65,70	3,80	0,00	65,30	2,90	0,25	64,70
	6	1,40	1,00	98,45	2,50	0,40	98,35	1,35	0,80	98,50
	10	0,00	13,20	100,00	0,00	11,95	100,00	0,00	1,21	100,00
200	2	12,70	19,10	24,35	11,45	17,30	22,75	12,20	17,00	23,10
	4	21,40	30,80	67,40	22,30	32,15	67,75	20,20	31,35	65,20
	6	31,60	46,80	96,55	30,25	44,35	95,60	31,75	43,60	95,70
	10	58,20	72,00	100,00	56,20	72,70	100,00	57,00	73,80	100,00
	20	95,70	96,75	100,00	89,50	90,60	93,35	95,05	97,00	100,00
		$\alpha = 1\%$								
		$\rho = 0,0$			$\rho = 0,5$			$\rho = 0,9$		
n	***p***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***	***ta***	***tc***	***ts***
20	2	1,10	0,25	4,70	1,50	0,35	6,45	1,30	0,40	6,45
	4	0,70	0,00	35,85	0,90	0,00	35,95	0,60	0,00	34,75
	6	0,20	0,00	86,20	0,70	0,00	84,40	0,15	0,00	83,90
	10	0,00	0,00	100,00	0,00	0,00	100,00	0,00	0,00	100,00
200	2	5,55	10,85	11,35	4,10	9,05	8,95	4,25	8,65	10,15
	4	9,75	19,15	41,10	9,50	20,20	43,65	8,95	17,45	39,60
	6	16,40	30,75	80,95	16,15	28,65	79,65	15,55	28,15	79,50
	10	39,10	55,95	100,00	36,75	56,05	100,00	38,90	56,40	100,00
	20	89,85	91,95	100,00	83,00	85,55	93,35	89,05	91,30	100,00

CONCLUSÕES

O teste de Shapiro-Wilk multivariado do programa **R**, função *mshapiro.test* do pacote mvnormtest, tem fraco desempenho e não é recomendado para uso rotineiro.

O teste de normalidade baseado em desvios de assimetria e curtose é recomendado para situações de $n \geq 50$ e $n \geq 100$, respectivamente.

REFERÊNCIAS BIBLIOGRÁFICAS

DOMANSKI, C. **Wlasnosci testu wielowymiarowej normalnosci Shapiro-Wilka i jego zastosowanie**. Cracow: University of Economics Rector's Lectures, 1998.

MARDIA, K. V. Measures of multivariate skewness and kurtosis with applications. **Biometrika**, London, v. 57, n. 3, p. 519-530, 1970.

MARDIA, K. V. Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. **Sankhyã A**, [S.l.], v. 36, p. 115-128, 1974.

MARDIA, K. V. Assessment of multinormality and the robustness of Hotelling's T^2 test. **Applied Statistics**, London, v. 24, n. 2, p. 163-171, 1975.

R DEVELOPMENT CORE TEAM. **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2006.

ROYSTON, J. B. Some techniques for assessing multivariate based on the Shapiro-Wilk W. **Applied Statistics**, London, v. 32, n. 2, p. 121-133, 1983.

SANTOS, A. C.; FERREIRA, D. F. Definição do tamanho amostral usando simulação Monte Carlo para o teste de normalidade baseado em assimetria e curtose: II. Abordagem multivariada. **Ciência e Agrotecnlogia**, Lavras, v. 24, n. 1, p. 62-69, 2003.