

NORMALIDADE DE VARIÁVEIS: MÉTODOS DE VERIFICAÇÃO E COMPARAÇÃO DE ALGUNS TESTES NÃO-PARAMÉTRICOS POR SIMULAÇÃO

NORMALITY OF VARIABLES: DIAGNOSIS METHODS AND COMPARISON OF SOME NONPARAMETRIC TESTS BY SIMULATION

Vanessa Bielefeldt Leotti Torman^{1,2}, Rodrigo Coster², João Riboldi¹

RESUMO

Introdução: Os principais testes estatísticos têm como suposição a normalidade dos dados, que deve ser verificada antes da realização das análises principais.

Objetivo: Revisar as técnicas de verificação da normalidade dos dados e comparar alguns testes de aderência à normalidade para diferentes distribuições de origem e tamanho amostral.

Metodologia: Através da simulação de cinco distribuições (Normal, t-student, Qui-Quadrado, Gama e Exponencial) e seis tamanhos amostrais (10, 30, 50, 100, 500 e 1000) foram simulados 5000 amostras de cada par distribuição-tamanho amostral e realizados os testes Qui-quadrado, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Shapiro-Francia, Cramer-von Mises, Anderson-Darling e Jarque-Bera.

Resultados: Os resultados obtidos mostram uma clara superioridade dos testes Shapiro-Francia e Shapiro-Wilk, com percentuais de acerto de 72,41% e 72,15%, respectivamente. Entre os piores resultados encontramos o Kolmogorov-Smirnov e Qui-Quadrado, com percentual de acerto de 44,78% e 61,58%, respectivamente.

Conclusões: Para amostras pequenas recomenda-se que sejam utilizados procedimentos não paramétricos diretamente para a análise, em função da baixa performance dos testes de aderência à normalidade, dado o baixo percentual de acertos. Para amostras maiores, recomenda-se o uso dos testes Shapiro-Francia ou Shapiro-Wilk.

Palavras-chave: Distribuição normal; teste de aderência; Qui-quadrado; Kolmogorov-Smirnov; Lilliefors; Shapiro-Wilk; Shapiro-Francia; Cramer-von Mises; Anderson-Darling; Jarque-Bera; poder; erro tipo I

ABSTRACT

Background: The main statistical tests have the normality assumption that must be verified before performing the main analyzes.

Objective: To review the techniques of testing for normality of data and compare some adherence tests for different true distributions and sample size.

Method: Through simulation of five distributions (Normal, t-Student, Chi-Square, Gamma and Exponential) and six sample sizes (10, 30, 50, 100, 500 and 1000) were simulated 5000 samples of each pair sample size-distribution and applied the Chi-square, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Shapiro-Francia, Cramer-von Mises, Anderson-Darling and Jarque-Bera tests.

Results: The results show a clear superiority of the Shapiro-Francia and Shapiro-Wilk tests, with percentages of accuracy of 72.41% and 72.15% respectively. Among the worst results we find the Kolmogorov-Smirnov and Chi-Square, with percentage of accuracy of 44.78% and 61.58% respectively.

Revista HCPA. 2012;32(2):227-234

¹Departamento de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, UFRGS.

²Unidade de Bioestatística, Grupo de Pesquisa e Pós-Graduação, Hospital de Clínicas de Porto Alegre.

Contato:

Vanessa Bielefeldt Leotti Torman
vanessa.leotti@ufrgs.br
Porto Alegre, RS, Brasil

Conclusions: For small samples it is recommended to use non-parametric procedures directly for the analyzes, due to the low performance of the tests of adherence to normality, given the low percentage of accuracy. For larger samples, we recommend the use of the Shapiro-Francia and Shapiro-Wilk tests.

Keywords: Normal distribution; goodness-of-fit tests; chi-square; Kolmogorov-Smirnov; Lilliefors; Shapiro-Wilk; Shapiro-Francia; Cramer-von Mises; Anderson-Darling; Jarque-Bera; power; type I error

Toda variável aleatória, seja idade de um grupo de pessoas ou a ocorrência de um determinado desfecho, assume uma determinada distribuição de frequências na população, que podem ter formas variadas (1). Na literatura estatística encontra-se muitas distribuições teóricas. Essas são modelos que procuram representar o comportamento de determinado evento em função da frequência de sua ocorrência. No caso das variáveis contínuas, esse evento será um intervalo de valores. As distribuições de frequências são, na verdade, distribuições de probabilidade, onde para um evento teremos uma probabilidade de ocorrência associada. Em outras palavras, a partir de uma distribuição de probabilidade completamente especificada, pode-se calcular a probabilidade de uma variável aleatória assumir determinado intervalo de valores.

Nesse artigo o objeto de estudo é a mais conhecida distribuição de probabilidades: a distribuição Normal ou Gaussiana. A suposição de normalidade da(s) variável(is) aleatória(s) é exigida para a realização de muitos métodos de inferência estatística, tais como:

- No caso de uma amostra de uma única população, a suposição de normalidade é exigida quando deseja-se obter intervalo de confiança ou executar teste de hipótese sobre a média dessa população baseados na estatística t.

- No caso de duas amostras de populações independentes, para realizar o teste t de comparação de médias para amostras independentes, é necessário que a variável aleatória assuma distribuição normal em ambas as populações (grupos).

- No caso de duas amostras pareadas (ou relacionadas), para utilizar o teste t de comparação das médias, é necessário que a variável aleatória da diferença entre as duas amostras tenha distribuição normal.

- No caso de três ou mais amostras de populações independentes, para compararmos as médias das populações através da Análise de Variância em classificação simples (One-Way ANOVA), entre outras suposições, tem-se que pressupor que a variável aleatória tenha distribuição normal em cada uma das populações (grupos).

- Na estimação da correlação linear de Pearson, o teste de significância do coeficiente de correlação somente é válido se ambas as variáveis aleatórias tiverem distribuição Normal.

- Nos modelos de regressão linear, uma das suposições é de que os resíduos do modelo tenham distribuição Normal.

Assim, é importante termos métodos para verificar se distribuição dos dados que estamos estudando se ajusta a uma distribuição normal. Há metodologias descritivas, como a análise visual de alguns gráficos, e também testes não-paramétricos de aderência, que testam objetivamente a hipótese de normalidade. Este artigo pretende revisar essas metodologias existentes e comparar o desempenho de alguns testes não-paramétricos por simulação.

Cumpramos ressaltar que os testes t de Student são classificados como testes paramétricos porque exigem uma distribuição de probabilidade específica para a variável aleatória, no caso a distribuição Normal. Entretanto, existem outras técnicas paramétricas que suportam outras distribuições de probabilidade. Assim, quando os dados não são normais, têm-se pelo menos duas soluções: a primeira é procurar outro teste paramétrico cuja distribuição de probabilidade assumida se ajuste melhor aos dados; a segunda é usar testes não-paramétricos, como o teste de Mann-Whitney, que não fazem pressuposição sobre a distribuição dos dados. Como nos programas estatísticos mais conhecidos, como o SPSS, os outros testes paramétricos não estão implementados, usualmente a segunda opção é adotada. Assim, a classificação de paramétrico ou não-paramétrico é referente ao tipo de teste estatístico, e não à variável aleatória. Não é correto dizer que uma específica variável é paramétrica ou não-paramétrica.

Na próxima sessão, os métodos disponíveis para a verificação da hipótese de normalidade serão revisados e o estudo de simulação procedido será descrito. Após, tem-se a sessão que apresentará os resultados da simulação, e na sequência, as conclusões do trabalho.

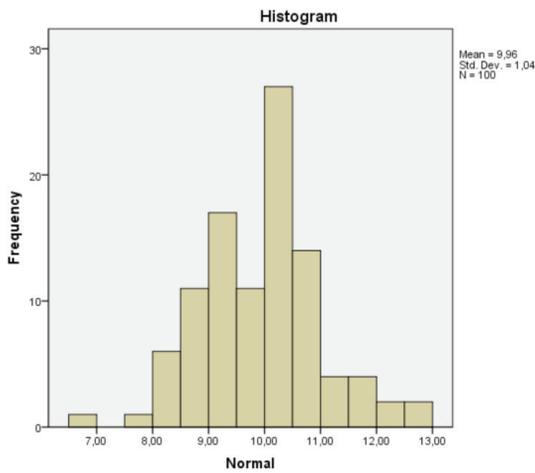
MÉTODOS

O primeiro método para verificação do formato da distribuição de uma variável contínua é a construção do histograma. O histograma é um gráfico de barras justapostas em que no eixo horizontal está a variável de interesse dividida em classes e no eixo vertical a frequência da classe correspondente (2). Este gráfico está disponível na maioria dos programas estatísticos. No SPSS versão 18.0, uma das maneiras de obter esse gráfico é através do menu Analyze -> Descriptive Statistics -> Explore -> Plots marcando-se Histogram. Através do histograma, buscamos verificar se a forma de sino da distribuição Normal está presente.

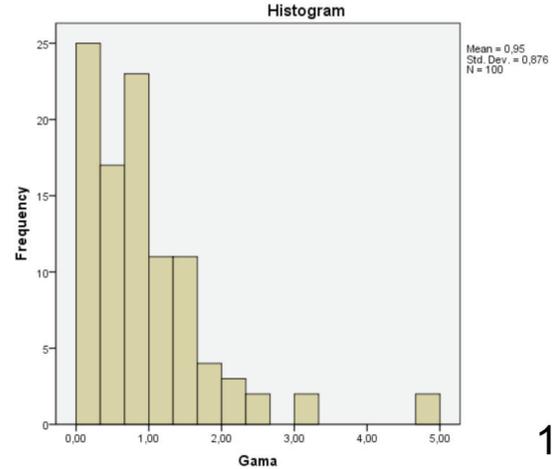
Na Figura 1, tem-se dois exemplos de histogramas provenientes

de amostras de tamanho 100, onde, na Figura 1-a tem-se dados gerados de uma distribuição Normal, e na Figura 1-b, dados gerados de uma distribuição Gama. Verifica-se que a

forma simétrica da Normal está presente na Figura 1-a, mas não na Figura 1-b, como esperado.



1a



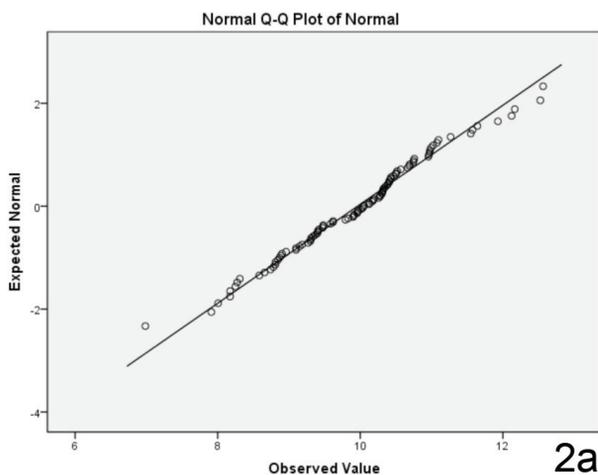
1b

Figura 1 - Exemplos de Histogramas

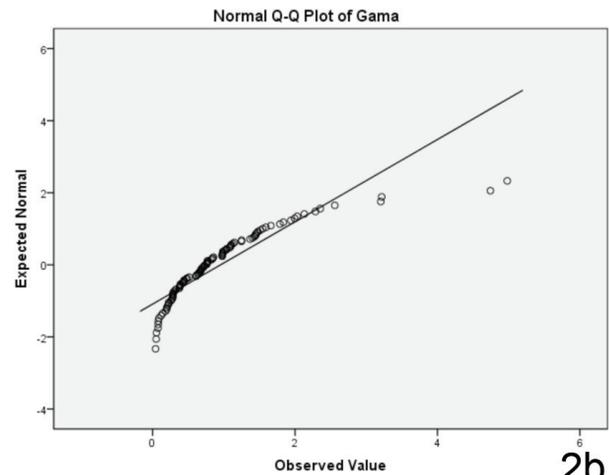
Outro gráfico que pode ser utilizado para avaliar a normalidade de uma variável é o gráfico Quantil-Quantil, ou Q-Q Plot. Neste gráfico, no eixo horizontal tem-se os valores observados da variável, e no eixo vertical, os valores esperados caso a variável tenha distribuição Normal. Se há uma boa aderência dos dados à distribuição Normal os pontos estão próximos a reta de referência apresentada no gráfico. No SPSS versão 18.0, obtém-se o Q-Q Plot marcando-se *Normality Plots with Tests*, na mesma janela onde se

solicitou o Histograma anteriormente.

Na Figura 2, tem-se dois exemplos de Q-Q Plots provenientes de amostras de tamanho 100, novamente de dados gerados de uma distribuição Normal (Figura 2-a), e dados gerados de uma distribuição Gama (Figura 2-b). Verifica-se que os pontos se aproximam bem da reta no caso da variável com distribuição Normal, mas há um grande desvio para a variável com distribuição Gama.



2a



2b

Figura 2 - Exemplos de Q-Q Plots

Os métodos gráficos citados anteriormente têm a desvantagem de serem subjetivos, pois dependem de interpretação visual. Para um resultado mais objetivo, pode-se usar testes não-paramétricos de aderência à distribuição Normal. Alguns deles são: Qui-quadrado de Pearson (QQ), Kolmogorov-Smirnov (KS), Lilliefors (LF) – que é apenas uma correção para o teste KS), Shapiro-Wilk (SW), Shapiro-Francia (SF), Cramer-von Mises (CM), Anderson-Darling (AD) e Jarque-Bera (JB).

Apenas os quatro primeiros testes citados estão disponíveis no SPSS versão 18.0, sendo que o teste qui-quadrado é bastante trabalhoso pois necessita que se classifique a variável manualmente, bem como o cálculo das frequências esperadas. O teste Kolmogorov-Smirnov pode ser obtido em Analyze -> Nonparametric Testes -> Legacy Dialogs -> 1-Sample KS. Os testes de Lilliefors e Shapiro-Wilk são exibidos marcando-se Normality Plots with Tests, da mesma maneira utilizada para obter o Q-Q Plot. Todos os testes citados estão disponíveis no

programa R versão 2.15.0 (3), através das funções `pearson.test`, `ks.test`, `lillie.test`, `shapiro.test`, `sf.test`, `cvm.test`, `ad.test` e `jarque.bera.test`, respectivamente. As funções `ks.test` e `shapiro.test` estão disponíveis na instalação básica do R, já a função `jarque.bera.test` pertence ao pacote `tseries` (4) e as demais funções pertencem ao pacote `nortest` (5).

Estes testes de aderência têm estatísticas de teste e critérios de decisão diferentes, entretanto têm em comum as hipóteses testadas: a hipótese de nulidade é de que a variável aleatória adere à distribuição Normal, contra a hipótese alternativa de que a variável aleatória não adere à distribuição Normal. A maneira mais fácil de tomar a decisão é observar o valor-p dos testes e comparar com o nível de significância adotado. Se o valor-p do teste for menor que o nível de significância escolhido, rejeita-se a hipótese de normalidade. Mais detalhes sobre estes testes podem ser encontrados em (6) e (7).

Procedendo-se os testes de Lilliefors e Shapiro-Wilk para os dados mostrados na Figura 1, tem-se os resultados do Quadro 1.

Quadro 1 - Saída do SPSS para os testes de normalidade.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
Normal	,065	100	,200*	,989	100	,575
Gama	,152	100	,000	,797	100	,000

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Considerando um nível de significância de 5%, através do Quadro 1, percebe-se que tanto os dois testes não rejeitam a hipótese de normalidade para a variável com distribuição Normal e rejeitam esta hipótese para a variável com distribuição Gama. Entretanto, em algumas situações os testes podem discordar na decisão estatística. Zar (8) não recomenda o uso dos testes qui-quadrado e Kolmogorov-Smirnov, pois os mesmos têm poder muito baixo, ou seja, em muitas situações onde os dados realmente não provêm de uma distribuição normal, estes testes não conseguem rejeitar a hipótese de nulidade.

Assim, torna-se interessante verificar o desempenho dos mesmos através de simulação, para tentar identificar algum com melhor desempenho. Para tanto, procedeu-se um estudo de simulação com a seguinte configuração:

1) Foram simuladas variáveis de 6 distribuições diferentes: Normal (0,1), Qui-quadrado (3), t-student (10), Gama (10,1/3) e Exponencial (5).

2) Para todas as variáveis foram utilizados 6 tamanhos amostrais distintos: 10, 30, 50, 100, 500 e 1000.

3) Para cada distribuição e tamanho amostral, simulou-se 5000 amostras.

Para analisar a eficiência dos testes foi averiguado o percentual de vezes que a decisão correta foi tomada, utilizando 5% como nível de significância. Isto significa que para a distribuição normal foi avaliado a proporção de não rejeições da hipótese de nulidade, isto é, o complementar da probabilidade do erro tipo I, ou seja, a especificidade do teste, enquanto para as demais distribuições foi avaliada a proporção de rejeições da hipótese de nulidade, ou seja, o poder ou sensibilidade do teste. As simulações foram procedidas no programa R 2.15.0.

Resultados do estudo de simulação

As funções densidades de probabilidade das distribuições não-normais simuladas, bem como a densidade da distribuição Normal com mesma média e variância de cada uma dessas distribuições encontram-se na Figura 3. Percebe-se que a distribuição mais similar à Normal (representada em vermelho em todos os gráficos) é a t-student (10) e a mais discrepante é a Exponencial (5).

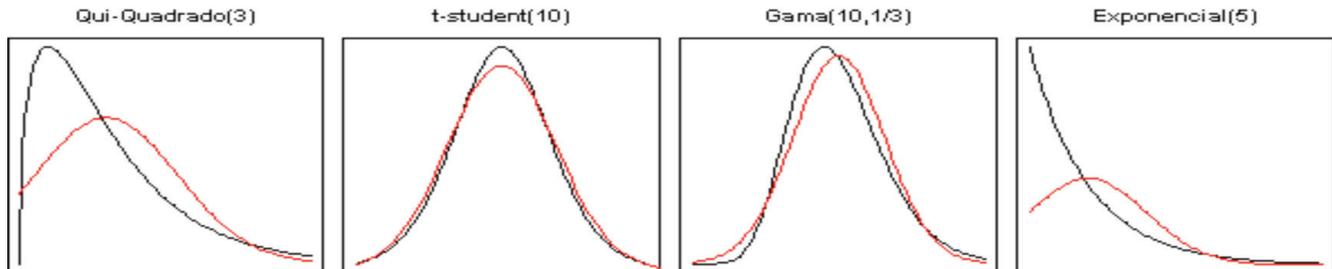


Figura 3 - Funções densidade das distribuições simuladas (em preto) e distribuição normal com mesma média e variância (em vermelho).

Os resultados obtidos nas simulações foram sintetizados e apresentados na Figura 4, Tabela 1 e 2.

Na Figura 4 encontramos o percentual de acertos de cada teste para cada distribuição e tamanho amostral. Em relação às amostras provenientes da distribuição Normal, percebe-se que o teste de Kolmogorov-Smirnov tem um comportamento não esperado em relação ao erro tipo I pois, mesmo realizando-se o teste com 5% de significância, este não rejeitou a hipótese de normalidade em nenhuma amostra independentemente do tamanho amostral. Com exceção desse teste, todos os demais parecem ter taxas de erro tipo I adequadas, próximas do valor nominal de 5%, pois o percentual de acerto foi próximo do esperado (95%) para a maioria dos tamanhos amostrais considerados. Em relação às demais distribuições simuladas, o teste Kolmogorov-Smirnov apresentou o menor poder independentemente da distribuição para amostras menores que 500, sendo portanto não recomendável, concordando com Zar (8).

Para a distribuição t-student os testes diferem bastante mesmo para o maior tamanho amostral, tendo o percentual de acerto variado desde 1,72% (Kolmogorov-Smirnov) até 94,36%

(Jarque-Bera), conforme a Tabela 1. A dificuldade dos testes em identificarem corretamente a falta de normalidade quando os dados seguem a distribuição t-student se deve ao fato das duas distribuições possuírem densidades muito parecidas. Levando em consideração todos os tamanhos amostrais, o teste que melhor identificou os dados provenientes da distribuição t-student foi o Shapiro-Francia, com 39,26% de acertos, e nas distribuições restantes o teste Shapiro-Wilk obteve melhores resultados.

Outro resultado importante que verifica-se na Tabela 1 é o percentual total de acertos de cada teste, onde o teste de Shapiro-Francia obteve melhor percentual de acertos (72,41%), seguido pelo teste Shapiro-Wilk, (72,15%). Já entre os piores resultados, encontram-se o teste Qui-quadrado (61,58%) e o KS (44,78%).

É possível perceber também que o desempenho dos testes pode ser muito ruim para tamanhos de amostrais pequenos. Para as distribuições mais parecidas com a Normal (Gama e t), mesmo os melhores testes só conseguem ter poder acima de 70% a partir de $n=500$.

Tabela 1 - Percentual de acerto dos testes para cada distribuição e tamanho amostral.

Dist	Tamanho Amostral	% Acerto							
		SW	KS	LF	AD	CM	QQ	SF	JB
Normal	10	95,66%	100,00%	95,28%	95,30%	95,26%	93,56%	95,00%	99,36%
	30	95,18%	99,96%	95,60%	95,76%	95,38%	94,88%	94,84%	96,62%
	50	94,10%	100,00%	94,88%	94,42%	94,62%	94,30%	93,88%	95,70%
	100	95,06%	100,00%	94,84%	94,94%	94,96%	94,96%	94,62%	95,54%
	500	94,86%	100,00%	95,36%	95,12%	95,02%	95,48%	94,82%	95,20%
	1000	95,26%	99,96%	95,24%	95,00%	94,86%	95,16%	95,10%	95,08%
	Total	95,02%	99,99%	95,20%	95,09%	95,02%	94,72%	94,71%	96,25%
Qui Quadrado	10	30,66%	0,20%	21,34%	28,26%	25,98%	26,94%	30,58%	9,74%
	30	86,56%	4,40%	58,32%	79,64%	73,76%	59,06%	82,78%	57,78%
	50	98,86%	15,58%	82,50%	96,50%	93,50%	85,58%	98,14%	86,04%
	100	100,00%	56,74%	99,36%	99,96%	99,90%	99,74%	100,00%	99,88%
	500	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
	1000	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
	Total	86,01%	46,15%	76,92%	84,06%	82,19%	78,55%	85,25%	75,57%
t-student	10	7,10%	0,04%	6,52%	7,16%	6,70%	7,52%	8,70%	2,10%
	30	11,70%	0,02%	7,78%	9,74%	8,94%	6,34%	14,46%	11,54%
	50	15,30%	0,04%	8,48%	12,02%	10,64%	6,68%	19,06%	17,16%
	100	22,92%	0,08%	11,62%	16,14%	13,92%	7,96%	29,06%	29,36%
	500	65,96%	0,36%	26,80%	48,41%	41,06%	11,74%	71,64%	75,04%
	1000	90,10%	1,72%	49,48%	78,17%	70,18%	19,94%	92,62%	94,36%
	Total	35,51%	0,38%	18,45%	28,61%	25,24%	10,03%	39,26%	38,26%
Gama	10	8,42%	0,00%	6,92%	8,28%	7,70%	8,08%	9,08%	1,94%
	30	20,26%	0,06%	12,00%	16,36%	14,44%	9,00%	19,60%	13,78%
	50	33,30%	0,26%	18,40%	25,48%	22,96%	11,76%	31,96%	24,26%
	100	61,72%	1,02%	33,58%	47,68%	42,36%	18,72%	58,70%	49,72%
	500	99,98%	34,84%	95,22%	99,60%	98,92%	78,24%	99,98%	99,94%
	1000	100,00%	86,72%	99,96%	100,00%	100,00%	98,90%	100,00%	100,00%
	Total	53,95%	20,48%	44,35%	49,57%	47,73%	37,45%	53,22%	48,27%
Exponencial	10	44,72%	0,48%	29,66%	41,74%	38,54%	39,66%	43,18%	14,86%
	30	96,86%	11,64%	77,92%	93,26%	90,08%	85,14%	94,60%	72,84%
	50	99,88%	37,22%	95,98%	99,62%	99,08%	98,18%	99,78%	95,68%
	100	100,00%	92,08%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
	500	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
	1000	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
	Total	90,24%	56,90%	83,93%	89,10%	87,95%	87,16%	89,59%	80,56%
Total geral		72,15%	44,78%	63,77%	69,29%	67,63%	61,58%	72,41%	67,78%

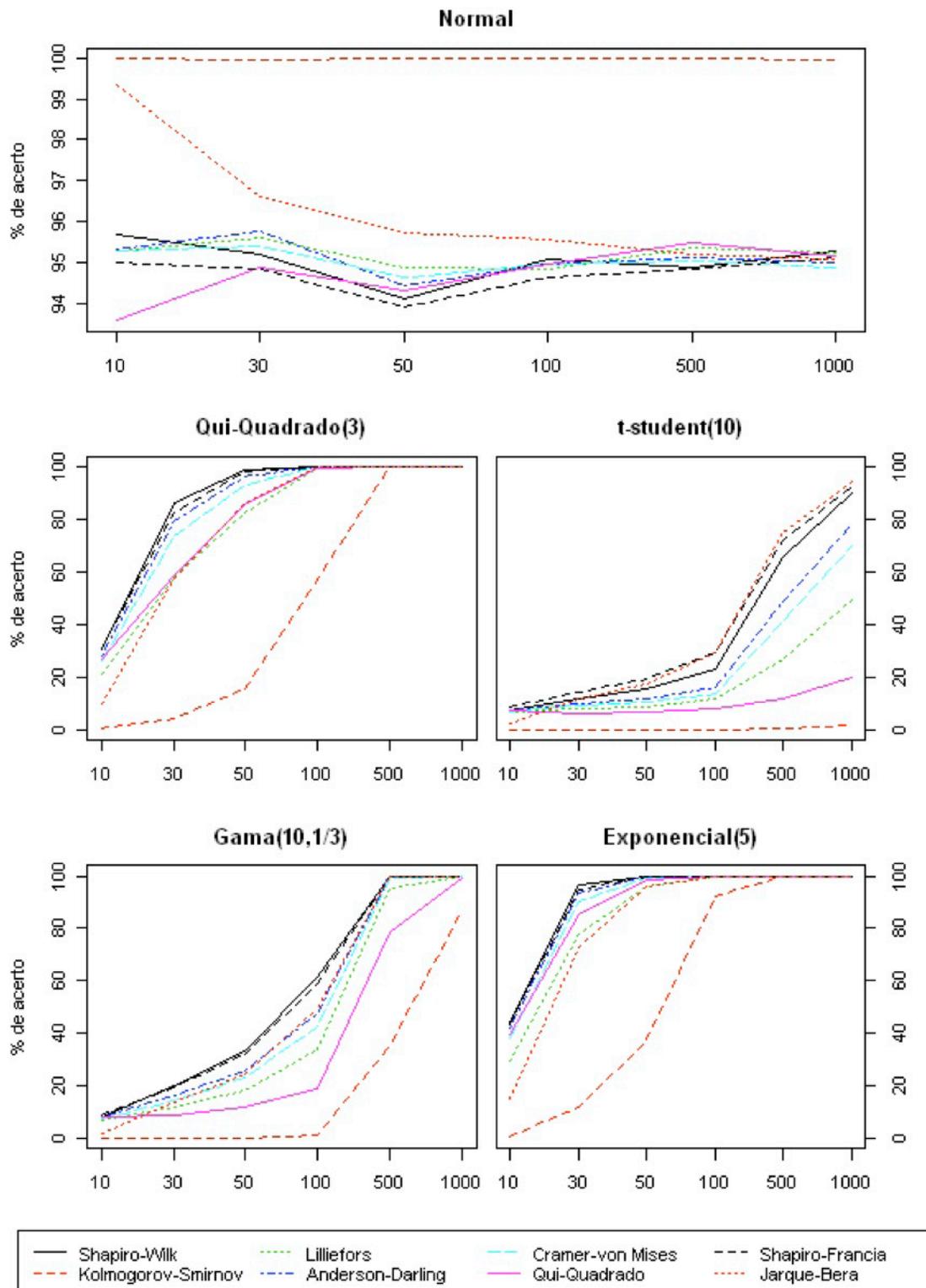


Figura 4 - Percentual de acerto em função dos tamanhos amostrais.

Tabela 2 - Percentual de acerto dos testes para cada tamanho amostral.

Tamanho Amostral	SW	KS	LF	AD	CM	QQ	SF	JB
10	37,31%	20,14%	31,94%	36,15%	34,84%	35,15%	37,31%	25,60%
30	62,11%	23,22%	50,32%	58,95%	56,52%	50,88%	61,26%	50,51%
50	68,29%	30,62%	60,05%	65,61%	64,16%	59,30%	68,56%	63,77%
100	75,94%	49,98%	67,88%	71,74%	70,23%	64,28%	76,48%	74,90%
500	92,16%	67,04%	83,48%	88,63%	87,00%	77,09%	93,29%	94,04%
1000	97,07%	77,68%	88,94%	94,63%	93,01%	82,80%	97,54%	97,89%

Conclusões

Neste trabalho buscou-se apresentar métodos de verificação da normalidade de variáveis, como os gráficos Histograma e Q-Q Plot. A análise visual tem a desvantagem de ser subjetiva e um método mais objetivo é a realização de testes não-paramétricos de aderência à distribuição Normal. Como estes testes podem conduzir a conclusões diferentes,

comparou-se alguns deles através de simulação. Os resultados do estudo de simulação procedidos permitem recomendar o uso dos testes de Shapiro-Wilk e Shapiro-Francia. É importante também salientar que o desempenho dos testes foi em geral muito ruim para tamanhos amostrais pequenos. Para amostras de tamanho menor ou igual a 10, recomenda-se não proceder ao teste de normalidade e partir diretamente para uma estratégia não-paramétrica de análise.

REFERÊNCIAS

1. Callegari-Jacques SM. Bioestatística princípios e aplicações. Porto Alegre: Artmed; 2003.
2. Soares JF. Introdução à estatística médica. Belo Horizonte, MG: COOPMED; 2002.
3. Team RDC. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2011.
4. Trapletti A, Hornik K. tseries: Time Series Analysis and Computational Finance. 2012. Disponível em: <http://CRAN.R-project.org/package=tseries>
5. Gross J. nortest: Tests for Normality. 2012. Disponível em: <http://CRAN.R-project.org/package=nortest>
6. Thode HC. Testing for normality. New York: Marcel Dekker; 2002.
7. D'Agostino RB, Stephens MA. Goodness-of-fit techniques. New York: M. Dekker; 1986.
8. Zar JH. Biostatistical analysis. Upper Saddle River, N.J.: Prentice Hall; 1999.

Recebido: 16/02/2012

Aceito: 16/04/2012