

# AVALIAÇÃO DE TESTES DE NORMALIDADE IMPLEMENTADOS NO PROGRAMA R POR SIMULAÇÃO MONTE CARLO

Renata Guidi Carradori<sup>1</sup>, Patrícia de Siqueira Ramos<sup>2</sup>

**Resumo:** *A distribuição normal é o modelo probabilístico contínuo mais importante porque a maioria dos métodos estatísticos é baseado nesse modelo e a maioria dos fenômenos aleatórios pode ser descrita de forma aproximada por ele. Há vários testes de normalidade que, muitas vezes, produzem resultados diferentes podendo rejeitar ou não a hipótese nula de normalidade. O objetivo do trabalho foi avaliar as taxas de erro tipo I e poder dos testes de normalidade Shapiro-Wilk, Anderson-Darling, D'Agostino-Pearson e Lilliefors implementados no programa R para decidir quais apresentam o melhor desempenho. As simulações de diferentes tamanhos amostrais e algumas distribuições simétricas e assimétricas foram feitas no programa R para avaliar as taxas de erro tipo I e poder dos testes. Sob normalidade, os quatro testes foram exatos e, em relação ao poder, o desempenho foi afetado pelo tamanho da amostra e pelo tipo de distribuição. Em distribuições simétricas, o teste D'Agostino-Pearson (DG) foi um pouco mais poderoso do que o teste Shapiro-Wilk (SW) enquanto que, nas distribuições assimétricas, ocorreu o contrário, porém, os dois foram superiores aos outros. Entretanto, o teste SW pode ser mais indicado pois o teste DG só pode ser aplicado para amostras maiores do que 30.*

**Palavras-chave:** *Normalidade Univariada, Programa Estatístico Livre, Teste de Shapiro-Wilk.*

**Abstract:** *We aimed to evaluate the Type I error rates and power of some normality tests implemented on R software and conclude about the one with the best performance. Under normality all tests were exact and the power was affected by sample size and type of distribution.*

**Keywords:** *Univariate Normality, Free Statistical Software, Shapiro-Wilk Test.*

## 1 Introdução

Grande parte dos procedimentos estatísticos de estimação e inferência de parâmetros supõem distribuição normal dos dados amostrais. Porém, tal suposição é, muitas vezes, deixada de lado pelos pesquisadores, seja por esquecimento, por falta de conhecimento ou por desconhecerem qual o melhor método a ser aplicado para se verificar se os dados seguem ou não a distribuição normal.

A distribuição normal é o modelo probabilístico contínuo mais importante porque a maioria dos métodos estatísticos é baseado nesse modelo e a maioria dos fenômenos aleatórios pode ser

<sup>1</sup>ICB - UNIFAL-MG. e-mail: [renatinha\\_guidi@hotmail.com](mailto:renatinha_guidi@hotmail.com), agradecimento ao CNPq pelo apoio financeiro.

<sup>2</sup>ICEX - UNIFAL-MG. e-mail: [patricia.ramos@unifal-mg.edu.br](mailto:patricia.ramos@unifal-mg.edu.br), agradecimento à FAPEMIG pelo apoio financeiro.

descrita de forma aproximada por ele. A distribuição normal de uma variável aleatória contínua  $X$  tem a função densidade de probabilidade dada por:

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

em que  $\mu$ , média, e  $\sigma$ , desvio padrão, são os parâmetros da distribuição (Bussab; Morettin, 2003).

Avaliar o pressuposto de normalidade é exigido por inúmeros procedimentos estatísticos, tais como construção de alguns intervalos de confiança, testes de hipóteses e análise de variância, pois a maioria deles foi desenvolvido supondo que os dados são normais. Assim, torna-se importante verificar esse pressuposto antes de se prosseguir com os procedimentos estatísticos que o exijam (RAZALI; WAH, 2011).

A maior parte dos programas estatísticos disponibilizam maneiras de verificar a normalidade dos dados, sendo os mais populares: métodos gráficos, métodos numéricos e alguns testes, tais como: testes de Shapiro-Wilk, Anderson-Darling, Lilliefors, qui-quadrado etc. Métodos gráficos, como o Q-Q *plot*, são muito usados e úteis para avaliar a normalidade dos dados, mas são testes subjetivos e é importante efetuar um teste formal quanto à normalidade dos dados para tirar uma conclusão (FERREIRA, 2009).

Métodos numéricos incluem os baseados em coeficientes de assimetria e curtose enquanto um teste de normalidade é um teste formal da hipótese de normalidade dos dados (RAZALI; WAH, 2011). Os procedimentos baseados em desvios de assimetria e curtose devem ser usados com cautela pois a não rejeição da hipótese nula não implica que a distribuição seja normal, já que há distribuições que não são normais e são simétricas e mesocúrticas. Assim, utilizar um teste formal de normalidade é importante, porém, os pesquisadores ficam confusos ao terem que decidir qual escolha fazer.

Há quantidade significativa de testes de normalidade disponíveis na literatura. No entanto, os procedimentos mais comuns de normalidade disponíveis em programas de estatística são os testes Shapiro-Wilk, Anderson-Darling, Lilliefors, Jarque-Bera, Cramer-von Mises, D'Agostino Pearson, Kolmogorov-Smirnov, entre outros. Alguns destes testes só podem ser aplicados em uma determinada condição ou pressuposto. Além disso, diferentes testes de normalidade muitas vezes produzem resultados diferentes e algum teste pode rejeitar enquanto outros não rejeitam a hipótese nula de normalidade. Os resultados contraditórios são enganosos e muitas vezes confundem os profissionais. Assim, a escolha do teste de normalidade a ser utilizado é uma tarefa difícil (RAZALI; WAH, 2011).

Duas formas de avaliar o desempenho de um teste é por meio de suas taxas de erro tipo I, sob a hipótese nula ( $H_0$ ), e pelo poder, simulando-se dados sob a hipótese alternativa ( $H_1$ ). Um teste ideal não deveria nunca rejeitar a hipótese de normalidade para uma amostra normal e sempre rejeitar para amostras não normais. Como tal teste não existe, o que se busca é um teste que apresente taxas de erro tipo I próximas do nível nominal de significância adotado e que apresente altos valores de poder para poder ser recomendado.

De acordo com alguns trabalhos que avaliaram o poder de testes de normalidade, como Razali; Wah (2011) e Yap; Sim (2011), observou-se que os testes de Shapiro-Wilk e Anderson-Darling são os mais poderosos e o teste de Kolmogorov-Smirnov é o menos poderoso. Porém, os testes são afetados pelo tamanho da amostra analisada e todos não são poderosos para amostras de tamanho menor do que 30. Outro aspecto que afeta o poder dos testes é o tipo de distribuição utilizada. Para distribuições simétricas de caudas curtas, os testes de Shapiro-Wilk e Anderson-Darling são os mais poderosos. Para distribuições simétricas de caudas longas, em relação ao poder, os testes Jarque-Bera e D'Agostino-Pearson equiparam com o teste de Shapiro-Wilk. Para distribuições assimétricas, Shapiro-Wilk é o mais poderoso teste seguido pelo Anderson-Darling.

Como não é possível avaliar o desempenho de métodos gráficos por meio de taxas de erro tipo I e poder foi avaliado o desempenho de alguns testes formais de normalidade. Assim, o objetivo do trabalho foi avaliar as taxas de erro tipo I e poder dos testes de normalidade Shapiro-Wilk

(SHAPIRO; WILK, 1965), Anderson-Darling (ANDERSON; DARLING, 1954), D’Agostino-Pearson (D’AGOSTINO; PEARSON, 1973) e Lilliefors (LILLIEFORS, 1967) implementados no programa R (R CORE TEAM, 2013) para decidir quais apresentam o melhor desempenho ao responder se os dados podem ser considerados normais.

## 2 Material e métodos

A simulação Monte Carlo foi utilizada para simular amostras normais sob a hipótese nula de normalidade,  $H_0$ , para avaliar as taxas de erro tipo I dos testes e amostras sob a hipótese alternativa  $H_1$ , ou seja, dados de distribuições não normais para avaliar o poder. Em cada situação, os testes de normalidade foram aplicados ao nível nominal de significância pré-definido  $\alpha$  com o valor de 0,05. Se o teste rejeita a hipótese nula e a amostra é da distribuição normal, comete-se um erro do tipo I e, se não for normal, uma decisão correta foi tomada.

### 2.1 Simulação Monte Carlo

Foram efetuadas  $N = 2.000$  simulações sob  $H_0$  e sob  $H_1$ . Cada simulação de cada distribuição produziu amostras de tamanho  $n$  e esses valores foram: 10, 20, 30, 50, 100, 500 e 1.000. Valores de  $n$  menores do que 30 foram escolhidos porque, de acordo com a construção de alguns dos testes e por alguns estudos de simulação, o poder tende a ser menor quando o tamanho amostral é reduzido e isso foi verificado. Porém, o teste D’Agostino-Pearson exige que as amostras tenham tamanho mínimo  $n = 30$ , por isso, os tamanhos menores do que este não foram simulados para esse teste.

A proporção de decisões incorretas sob normalidade representa a taxa de erro tipo I empírica e a proporção de decisões corretas sob não normalidade representa o poder empírico. As taxas de erro tipo I foram comparadas com o nível nominal de significância por meio de um intervalo de confiança para proporções a 0,99 de confiança. O IC exato para  $p$  com confiança de  $1 - \alpha$  é:

$$IC_{1-\alpha}(p) = \left[ \frac{1}{1 + \frac{(n-y+1)F_{\alpha/2; \nu_1=2(n-y+1), \nu_2=2y}}{y}}; \frac{1}{1 + \frac{(n-y)}{(y+1)F_{\alpha/2; \nu_1=2(y+1), \nu_2=2(n-y)}}} \right], \quad (2)$$

em que  $F_{\alpha/2}$  é o quantil superior da distribuição  $F$  com  $\nu_1$  e  $\nu_2$  graus de liberdade (Ferreira, 2009). As taxas de erro tipo I e poder foram comparadas entre os testes de normalidade.

Os desempenhos dos seguintes testes implementados no programa R foram avaliados. O teste de Shapiro-Wilk foi aplicado utilizando-se a função `shapiro.test()` do pacote `stats`. Os valores- $p$  desse teste são obtidos a partir de uma aproximação realizada por Royston (1982). A função `ad.test()` do pacote `nortest` foi usada para o teste de Anderson-Darling. Para o teste Lilliefors, a função `lillie.test()` do pacote `nortest` foi aplicada. Tais funções foram escolhidas por permitirem a recuperação do valor- $p$  facilmente para comparar com o nível  $\alpha = 0,05$  a ser utilizado. Já para o teste de D’Agostino-Pearson foi usada a função `dagostino()` do pacote `fBasics`.

É importante ressaltar que a função `agostino.test()` do pacote `moments` do programa R também realiza o teste D’Agostino. Porém, sua implementação difere da função `dagostino()` e é baseada no trabalho de D’Agostino (1970), sendo mais indicada para detectar assimetria em dados normalmente distribuídos. O desempenho dessa função também foi avaliado em relação às taxas de erro tipo I e poder e comparado com o desempenho da função `dagostino()`.

### 2.2 Erro tipo I dos testes

Duas mil amostras da distribuição normal de tamanho  $n$  foram simuladas, ou seja, cada amostra  $X_1, X_2, \dots, X_n$  teve densidade normal. Assim, diz-se que  $X$  segue uma distribuição normal com os parâmetros  $\mu$  e  $\sigma$ , ou seja,  $X \sim N(\mu, \sigma)$ . Para gerar tais amostras foi utilizada a função `rnorm` com média  $\mu = 10$  e desvio padrão  $\sigma = 1$ , sem perda de generalidade.

Se, ao aplicar os testes, a hipótese nula de normalidade é rejeitada a um nível nominal de significância de 0,05, então a distribuição dos dados é considerada erroneamente como não normal. A proporção de rejeições incorretas foi calculada para cada teste e representam as taxas de erro tipo I. Se o valor obtido for considerado estatisticamente diferente e menor do que o nível  $\alpha$ , de acordo com o intervalo de confiança para proporções construído, o teste é considerado conservador. Se, ao contrário, o valor for considerado diferente e maior do que o nível nominal de significância, o teste é liberal.

### 2.3 Poder dos testes

Diferentes distribuições foram simuladas para contemplar diferentes situações de assimetria e curtose. Duas mil amostras de tamanho  $n$  de algumas distribuições não normais, simétricas e assimétricas, foram simuladas. As distribuições simétricas foram a Uniforme(0,1), Beta(2,2) e  $t(10)$ . Já as distribuições assimétricas foram Beta(6,2),  $\chi^2(20)$ , Lognormal(0,1), Exponencial(1) e Weibull(3,1).

A distribuição  $t$  de Student teve um caso,  $\nu = 10$  graus de liberdade e foi simulada por meio da função `rt` do programa R. Com poucos graus de liberdade, a distribuição  $t$  fica mais distante da distribuição normal e com maior número de graus de liberdade, ela se aproxima mais da normal. Para o modelo lognormal, foram adotados os valores 0 para a média ( $\mu$ ) e 1 para o desvio padrão ( $\sigma$ ), na escala logarítmica, com o uso da função `rlnorm`. Para o modelo exponencial, foi adotado o valor de  $\lambda = 1$ , o que corresponde à média igual a 1 e desvio padrão igual a 1, pois  $\mu = \sigma = 1/\lambda$ , e foi utilizada a função `rexp`. Para as outras distribuições, procedimentos parecidos foram adotados com o uso das funções `runif`, `rbeta`, `rchisq` e `rweibull` para gerar amostras da uniforme, beta,  $\chi^2$  e Weibull, respectivamente.

Os testes de normalidade foram aplicadas em cada uma das  $N = 2.000$  amostras de cada distribuição simulada e a proporção de rejeições corretas da hipótese nula foi computada. Os valores obtidos representam o poder dos testes, que foram comparados entre si.

## 3 Resultados e discussões

Os resultados serão apresentados de acordo com as taxas de erro tipo I, na situação de distribuição normal, e poder, na situação de outras distribuições.

Na Figura 1 estão representadas as taxas de erro tipo I dos testes de normalidade Shapiro-Wilk (SW), Anderson-Darling (AD), D'Agostino-Pearson (DG) e Lilliefors (LL) para diferentes tamanhos amostrais  $n$  e dados simulados da distribuição normal com média 10 e desvio padrão 1. Todos os testes apresentaram taxas de erro próximas de 0,05 sendo, por isso, considerados exatos. Seus valores ficaram dentro do IC para proporções ( $p$ ) construído a 99% de confiança com  $\alpha = 0,05$ ,  $IC_{99\%}(p)$ :  $[0,033927; 0,070504]$ .

Keskin (2006) avaliou quatro testes de normalidade, Kolmogorov-Smirnov, qui-quadrado, Shapiro-Wilk e D'agostino Pearson, quanto às taxas de erro tipo I e poder para amostras de tamanhos  $n = 10$  a  $n = 150$ . Os testes DG e SW foram exatos para todos os tamanhos amostrais simulados a partir de uma normal com média 0 e desvio padrão 1, sendo estes testes a apresentar os melhores resultados de erro tipo I. SW também mostrou ser o teste mais poderoso, principalmente para amostras maiores do que  $n = 60$ .

Santos; Ferreira (2003) avaliaram testes de normalidade baseados nos coeficientes de assimetria e curtose, além do teste de Shapiro-Wilk. De acordo com os autores, o teste SW controla adequadamente o valor do erro tipo I para amostras de todos os tamanhos e, ainda, os testes relacionados com assimetria apresentaram melhores resultados para o controle do erro tipo I.

Na Figura 2 estão apresentados os valores de poder para os quatro testes em função do tamanho amostral  $n$  para as três distribuições simétricas simuladas:  $U(0,1)$ ;  $B(2,2)$  e  $t(10)$ .

Os testes foram mais poderosos para dados simulados da uniforme e menos poderosos para a  $t$ . Para a uniforme, os valores de poder dos testes SW e DG atingiram aproximadamente 0,80

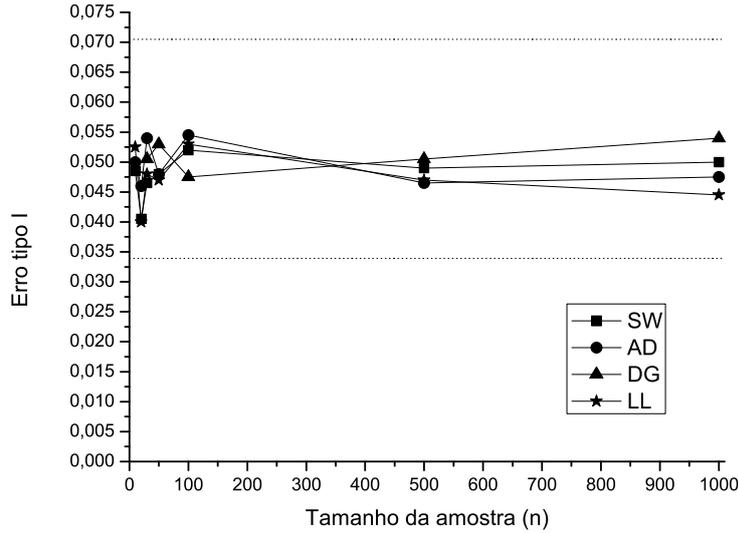


Figura 1: Erro tipo I dos testes SW, AD, DG e LL em função do tamanho amostral  $n$  para a distribuição  $N(10,1)$ .

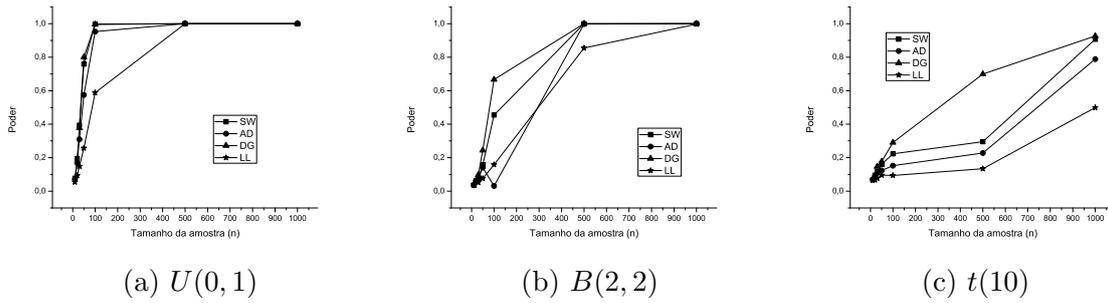


Figura 2: Poder dos testes SW, AD, DG e LL, em função do tamanho amostral  $n$  para as distribuições simétricas.

para  $n = 50$ , já o LL só atingiu poder de 1,00 para  $n = 500$ , ou seja, um tamanho amostral muito grande. Em relação à distribuição uniforme os resultados obtidos eram esperados pois, de acordo com Yap; Sim (2011), os testes SW, AD e DG possuem os maiores valores de poder para essa distribuição e o teste LL apresenta os piores resultados, principalmente para amostras menores do que  $n = 50$ . O teste DG apresentou poder superior ao SW para as distribuições  $B(2,2)$  e  $t(10)$ . Na  $B(2,2)$ , para  $n = 100$ , o teste DG apresentou poder de 0,67 e o SW de 0,45. Da mesma forma, na  $t$ , para  $n = 500$ , DG apresentou poder de 0,70 enquanto o SW foi de 0,30 e todos os testes tiveram seu poder alto somente para amostras maiores do que  $n = 500$ . Ainda assim, DG teve melhores resultados de poder do que SW em todos os tamanhos amostrais mesmo não sendo esse o resultado esperado, pois, de acordo com a literatura, o teste SW seria o mais poderoso. Na maioria das situações, o teste LL foi o menos poderoso.

Na Figura 3 estão representados os valores de poder dos testes para as distribuições assimétricas  $B(6,2)$ ,  $\chi^2(20)$ ,  $ln(0,1)$ ,  $exp(1)$  e  $W(3,1)$ .

As distribuições em que os testes apresentaram maior poder foram a lognormal e a exponencial. Os valores ficaram em torno de 0,80 a partir de  $n = 30$  para todos os testes, o que foi semelhante aos resultados obtidos por Santos; Ferreira (2003) para os testes analisados pelos autores.

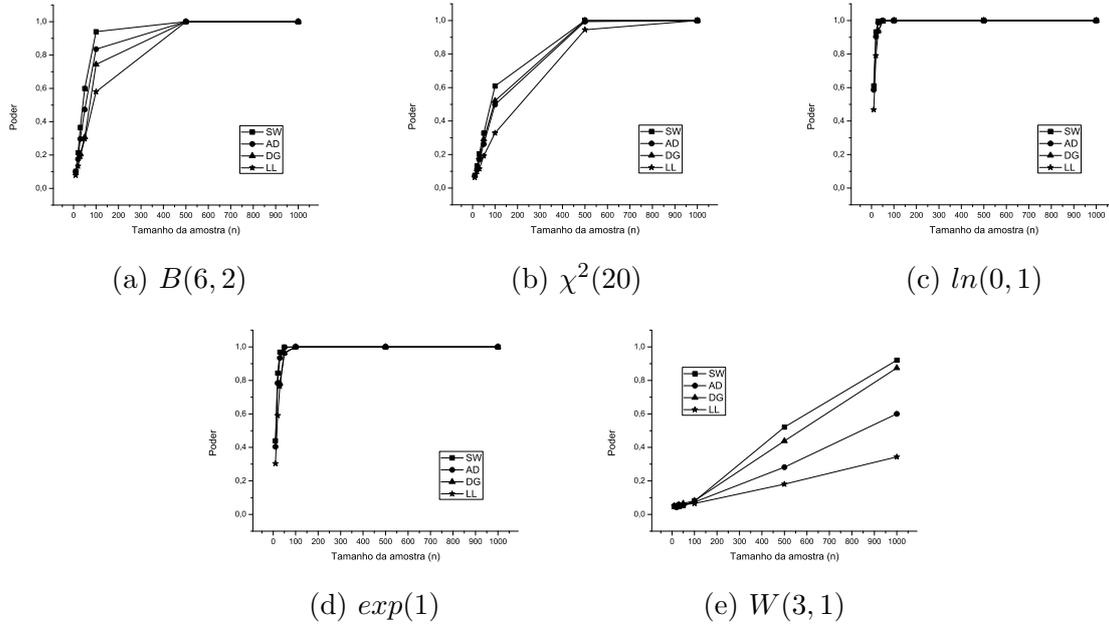


Figura 3: Poder dos testes SW, AD, DG, e LL, em função do tamanho amostral  $n$  para as distribuições assimétricas.

O teste SW foi o mais poderoso, seguido do DG, para as distribuições assimétricas  $B(6, 2)$ ,  $\chi^2(20)$  e  $W(3, 1)$ . Na Weibull, os valores de poder dos testes SW e DG só se destacaram em relação aos outros a partir de  $n = 100$ . Os valores de poder de SW só foram altos (em torno de 0,90) em  $n = 1000$ , seguido de DG, confirmando o que se esperava baseando-se em Yap; Sim (2011), mostrando que, para esta distribuição, os testes apresentam baixo poder. Em oposição a isso, tem-se que a distribuição lognormal apresentou ótimos resultados de poder para todos os testes (a partir de  $n = 50$ ), inclusive para o teste LL, condizendo com o que foi postulado por Yap; Sim (2011).

Os resultados obtidos com a distribuição  $\chi^2(20)$  condizem com Keskin (2006), o qual informa que o teste de SW é adequado para amostras maiores do que  $n = 40$ . De fato, para amostras de tamanho  $n = 50$ , SW teve o melhor poder, porém, para essa distribuição, o poder só é alto a partir de  $n = 500$ . Sendo assim, o teste SW se mostrou o mais poderoso e tem seu poder aumentado de acordo com o aumento do tamanho da amostra.

De acordo com Ahad et al. (2011), o melhor teste é SW, seguido de AD mas, de acordo com as simulações deste trabalho, o segundo melhor é o teste DG. Razali; Wah (2011) tiveram a mesma conclusão sobre o teste SW e afirmam que ele pode ser utilizado tanto para distribuições simétricas não normais quanto para distribuições assimétricas. O poder dos testes, ainda de acordo com esses autores, pode variar de acordo com o tamanho da amostra e tipo de distribuição. Assim como Ahad et al. (2011), Razali; Wah (2011) também afirmam que o segundo melhor teste é o AD, e ainda apontam o desempenho de testes como KS, Kolmogorov-Smirnov, que não foi testado neste estudo e esse teste se mostrou muito pouco poderoso, assim como o LL.

Na Figura 4 estão ilustrados os valores obtidos para o erro tipo I dos testes SW e das duas implementações do teste DG, as funções `dagostino` e `agostino.test`, sendo a primeira a utilizada neste trabalho.

As taxas de erro tipo I do SW e da função `dagostino` já foram apresentadas na Figura 1 e os testes se mostraram exatos. Já a função `agostino.test` apresentou valores de erro muito baixos sendo considerado conservador para todos os valores de  $n$ .

Na Figura 5 encontram-se os valores de poder do teste SW para todas as distribuições

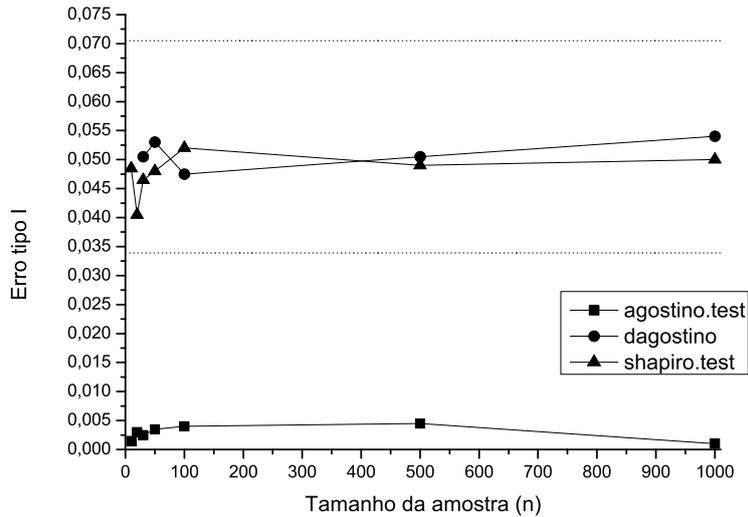


Figura 4: Erro tipo I dos testes SW e as duas implementações de DG função do tamanho amostral  $n$  para a distribuição  $N(10,1)$ .

simuladas. Este teste foi analisado em particular por ter sido o mais poderoso na maioria das situações. Os valores de poder do SW foram menores para a  $t(10)$  e  $W(3,1)$ ; intermediários para a  $\chi^2(20)$  e  $B(2,2)$ ; e superiores para as outras distribuições,  $U(0,1)$ ,  $exp(1)$ ,  $ln(0,1)$  e  $B(2,2)$ .

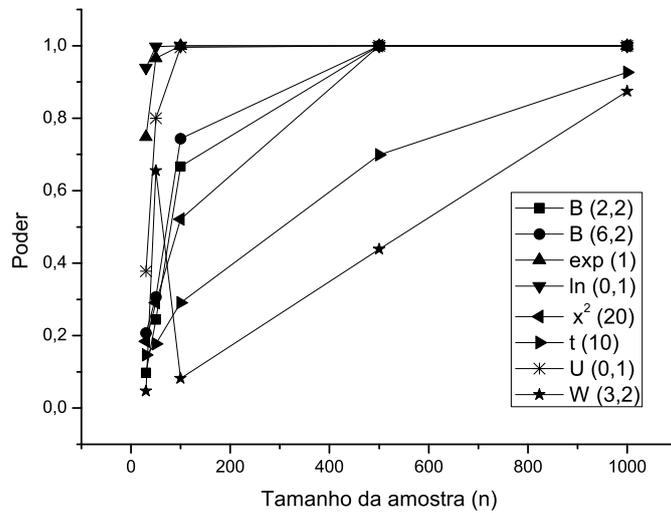


Figura 5: Poder do teste SW em função do tamanho amostral  $n$  para as distribuições assimétricas.

## 4 Conclusões

Para dados de distribuição normal, os quatro testes avaliados foram exatos, controlando a taxa de erro tipo I. Em relação ao poder, o desempenho dos testes foi afetado pelo tamanho

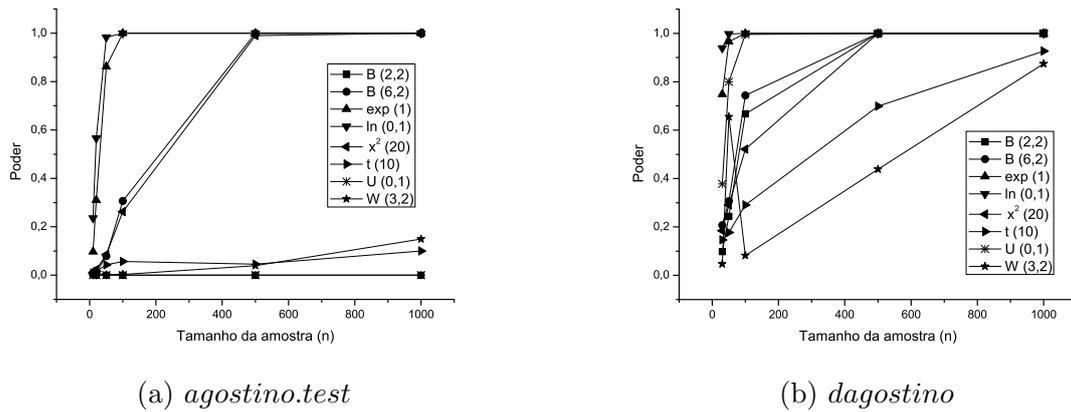


Figura 6: Poder do teste DG para as funções *agostino.test* e *dagostino*.

da amostra e pelo tipo de distribuição. Todos se tornaram mais poderosos à medida que se aumentou o tamanho amostral. Na maior parte das situações, o teste Lilliefors foi o menos poderoso. Em distribuições simétricas, o teste D’Agostino-Pearson (DG) foi um pouco mais poderoso do que o teste Shapiro-Wilk (SW) enquanto que, nas distribuições assimétricas, ocorreu o contrário, porém, os dois foram superiores aos outros. Entretanto, o teste SW pode ser mais indicado pois o teste DG só pode ser aplicado para amostras maiores do que 30.

## Referências

- [1] AHAD, N. A.; YIN, T. S.; OTHMAN, A. R.; YAACOB, C. R.; Sensitivity of Normality Tests to Non-normal Data. *Sains Malaysiana*, v.40, p.637-641, 2011.
- [2] ANDERSON, T. W.; DARLING, D. A. A test of goodness of fit. *Journal of American Statistical Association*, v.49, p. 765-769, 1954.
- [3] BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 5.ed., São Paulo: Atual, 2003.
- [4] D’AGOSTINO, R. B. Transformation to Nomality of the Null Distribution of G1. *Biometrika*, v.57, n.3, p.679-681, 1970.
- [5] D’AGOSTINO, R. B.; PEARSON, E. S. Testing for departures from normality. I. Fuller empirical results for the distribution of  $b_2$  and  $\sqrt{b_1}$ . *Biometrika*, v.60, p. 613-622, 1973.
- [6] FERREIRA, D. F. *Estatística Básica*. 2.ed., Lavras: Editora UFLA, 2009.
- [7] KESKIN, S. Comparison of Several Univariate Normality Tests Regarding Type I Error Rate and Power of the Test in Simulation based Small Samples. *Journal of Applied Science Research* v.2, n.5, p.296-300, 2006.
- [8] LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of American Statistical Association*, v.62, p.534-544, 1967.
- [9] R CORE TEAM. R: A Language and Environment for Statistical Computing. Vienna, Austria: *R Foundation for Statistical Computing*, 2013. Disponível em: <http://www.R-project.org>. Acesso em 15 Abr. 2013.
- [10] RAZALI, N. M.; WAH, Y. B. Power Comparisons of Shapiro-Wilk , Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, v.2, n.1, p.21-33, 2011.

- [11] ROYSTON, J. P. Remark AS R94: A remark on Algorithm AS 181: The W-test for normality. *Applied Statistics*, v.44, p.547-551, 1995.
- [12] SANTOS, A. C. dos e FERREIRA, D. F. Definição do tamanho amostral usando simulação Monte Carlo para o teste de normalidade baseado em assimetria e curtose: I. Abordagem univariada. *Ciência e Agrotecnologia*, vol.27, n.2, pp. 432-437, 2003.
- [13] SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v.52, p. 591-611, 1965.
- [14] YAP, B. W.; SIM, C. H. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, v.81, p.2141-2155, 2011.