

DEFINIÇÃO DO TAMANHO AMOSTRAL USANDO SIMULAÇÃO MONTE CARLO PARA O TESTE DE NORMALIDADE BASEADO EM ASSIMETRIA E CURTOSE. I. ABORDAGEM UNIVARIADA

ANDRÉA CRISTIANE DOS SANTOS¹
DANIEL FURTADO FERREIRA²

RESUMO – Uma forma alternativa para verificar suposição de normalidade dos dados, refere-se à aplicação de testes baseados nos coeficiente de assimetria e curtose. Realizou-se este trabalho com o objetivo de determinar um tamanho amostral ótimo para as estatísticas univariadas (Z_1 e Z_2) e multivariadas (K_1 e K_2) que, neste caso, foram consideradas como univariadas, com base em simulação. As estatísticas Z_1 e K_1 estão associados às medidas de simetria e Z_2 e K_2 às de curtose. Foram geradas diferentes funções de densidade de probabilidade univariadas, via método de Monte Carlo, com a finalidade de avaliar o erro tipo I e o poder do teste. As simulações foram feitas adotando-se os níveis de probabilidade de 5% e 1%. O critério de avaliação, no caso univariado, foi o da comparação das taxas de poder estimadas com o valor das taxas de poder empírico obtidas pelo teste de Shapiro & Wilk (1965). Pelos

resultados, verificou-se que as estatísticas Z_1 e Z_2 possuem aproximação assintótica normal para $n \geq 25$, com $\alpha = 5\%$ e podem ser recomendadas para uso rotineiro no caso univariado para testar a hipótese de normalidade dos dados; as estatísticas K_1 e K_2 possuem aproximações assintóticas melhores que Z_1 e Z_2 para um menor valor do nível nominal de significância, sendo recomendadas para $n \geq 25$ e $n \geq 100$, respectivamente, garantindo-se o controle da taxa de erro tipo I e um alto poder. No caso de distribuições com simetria próxima de zero e não-normais, as estatísticas baseadas em desvios de simetria apresentam maior poder do que a estatística W de Shapiro-Wilk. Finalmente, pode-se concluir que a estatística de assimetria, em geral, é mais poderosa do que a de curtose, mas os testes da hipótese nula de normalidade devem considerar tanto os testes de desvios de simetria como os de curtose conjuntamente.

TERMOS PARA INDEXAÇÃO: Assimetria, curtose, teste de normalidade univariado, taxa de erro tipo I e poder do teste.

DEFINITION OF THE SAMPLE SIZE BY USING MONTE CARLO SIMULATION FOR THE NORMALITY TEST BASED ON SKEWNESS AND KURTOSIS COEFFICIENTS. I. UNIVARIATE APPROACH

ABSTRACT – An alternative form to verify assumption of data normality is concerned with the application of the tests based on skewness and kurtosis coefficients. The objective of this work was to determine an optimum sample size for the univariate (Z_1 and Z_2) and multivariate (K_1 and K_2) statistics on basis of simulation. The Z_1 and K_1 statistics are related to the skewness and the Z_2 and K_2 are related to the kurtosis. Different univariate probability density functions were generated, by Monte Carlo simulation

method with a view to calculating the type I error rates and the power of the test. The simulations were done by adopting the probability level of 5% and 1%. The evaluation criterion in the univariate case was that of the comparison of the rates obtained through the value of the rates of empirical power obtained by Shapiro & Wilk (1965) test. By considering the univariate case, it was found that the Z_1 e Z_2 statistics possess normal asymptotic approximation for $n \geq 25$ and $\alpha = 5\%$ can be recommended for routine use in the

-
1. Mestre em Estatística e Experimentação Agropecuária, Professora de Estatística da FUOM, Formiga, MG. andrea@ufla.br.
 2. Dr. em Genética e Melhoramento de Plantas, Professor Adjunto III do Departamento de Ciências Exatas da UNIVERSIDADE FEDERAL DE LAVRA/UFLA, Caixa Postal 37 – 372000-000 – Lavras, MG. danielff@ufla.br, Bolsista CNPq.

univariate case. The K_1 and K_2 statistics possess approximation asymptotic better than Z_1 and Z_2 for a lower value of the nominal value of significance, recommended for $n \geq 25$ and $n \geq 100$, respectively, warranting the compromise with the control of the type I error rate and elevated power. In the case of symmetry distributions with efficient of skewness close to zero

and non-normal, the statistics based on skewness deviations present higher power than Shapiro – Wilk's W statistics. It is concluded that the skewness statistic in general, is more powerful than that of kurtosis, but the tests of the null hypothesis of normality must take into account both the tests of skewness deviations and those of kurtosis jointly.

INDEX TERMS: Skewness, kurtosis, test for normality, type I error rates and power of the test.

INTRODUÇÃO

A distribuição normal é importante tanto na estatística teórica como na aplicada por várias razões. Uma delas é que muitas variáveis na natureza, como, por exemplo, variáveis físicas, biológicas ou psicológicas comportam-se de modo aproximadamente simétrico, podendo ser bem representadas por essa distribuição. Verificar a suposição de normalidade em um conjunto de dados é avaliar o comportamento dos mesmos, ou seja, a forma que a distribuição assume, a qual espera-se que seja gaussiana em forma de sino. Na literatura, é possível encontrar vários métodos para testar a normalidade univariada dos dados, entre eles: Kolmogorov (1933), Shapiro & Wilk (1965) e Lilliefors (1967), qui-quadrado (Campos, 1983), Cramér-von Mises (Campos, 1983). Além desses testes consagrados na literatura, essa verificação pode ser feita indiretamente utilizando-se testes baseados nos coeficientes de assimetria e curtose esperados sob a distribuição normal. Essa abordagem tem como fundamento a comparação entre os valores paramétricos desses coeficientes na distribuição normal com os valores estimados na amostra ou no experimento.

Os estimadores dos coeficientes de assimetria, $\sqrt{b_1}$, e curtose, b_2 , são dados por:

$$\sqrt{b_1} = \frac{\tilde{\mu}_3}{\tilde{\mu}_2^{3/2}} \quad b_2 = \frac{\tilde{\mu}_4}{\tilde{\mu}_2^2}$$

em que $\tilde{\mu}$ é o estimador dos momentos centrados na média, referente às respectivas ordens.

As aproximações assintóticas normais, $N(0,1)$, desses estimadores são as estatísticas Z_1 e Z_2 , definidas por:

$$Z_1 = \sqrt{\frac{b_1(n+1)(n+3)}{6(n-2)}}$$

$$Z_2 = \left(b_2 - 3 + \frac{6}{n+1} \right) \sqrt{\frac{(n+1)^2(n+3)(n+5)}{24n(n-2)(n-3)}}$$

em que n é o tamanho da amostra (Bock, 1975).

A alternativa multivariada é baseada na medida de assimetria e curtose multivariada. Para uma amostra de tamanho n , os estimadores dos coeficientes de assimetria e curtose multivariados, $\hat{\beta}_{1,p}$ e $\hat{\beta}_{2,p}$, são respectivamente:

$$\hat{\beta}_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n g_{ij}^3 \quad \hat{\beta}_{2,p} = \frac{1}{n} \sum_{i=1}^n g_{ii}^2 = \frac{1}{n} \sum_{i=1}^n d_i^4$$

em que $g_{ij} = (y_i - \bar{y})' S_n^{-1} (y_j - \bar{y})$ e $d_i = \sqrt{g_{ii}}$; y_i e y_j são dois vetores de realizações multivariadas independentes e provenientes da mesma distribuição; S_n^{-1} é inversa da matriz de variância-covariância amostral viesada; \bar{y} é o vetor de médias amostrais e p é o número de variáveis.

Para grandes amostras Nordia (1970) notou que:

$$K_1 = \frac{n \hat{\beta}_{1,p}}{6} \sim \chi^2 \quad \text{com} \quad \frac{p(p+1)(p+2)}{6} \text{ graus de liberdade e} \quad K_2 = \frac{\hat{\beta}_{2,p} - p(p+2)}{\{8p(p+2)/n\}^{0,5}} \sim N(0,1)$$

Portanto, as estatísticas K_1 e K_2 são utilizadas para testar a hipótese nula de que os dados em estudo pertencem a uma amostra aleatória proveniente de uma distribuição normal multivariada. Embora as estatísticas K_1 e K_2 tenham sido propostas para o caso multivariado, ambas foram usadas no presente trabalho com $p=1$, ou seja, para testar a normalidade univariada.

Na literatura não existem muitos trabalhos referentes a testes de normalidade univariado, baseados nos coeficientes de assimetria e curtose, no que se refere à determinação dos tamanhos amostrais.

Conduziu-se este trabalho com o objetivo de estimar um tamanho amostral, via simulação Monte Carlo, que seja suficiente para satisfazer as propriedades assintóticas do teste de normalidade univariado baseado nos coeficientes de assimetria e curtose.

MATERIAL E MÉTODOS

Para avaliar as taxas de erro tipo I e poder dos testes assintóticos de normalidade baseados nos coeficientes de assimetria e curtose, foram utilizadas simulações computacionais pelo método de Monte Carlo. Em todos os casos, utilizou-se o teorema da probabilidade integral para gerar amostras aleatórias, extraídas da respectiva população (função de densidade de probabilidade).

Para simular dados univariados, foram geradas amostras com diferentes funções de densidade de probabilidade (f.d.p.). Foram consideradas amostras de diferentes tamanhos (n), variando de 5 em 5 e de 50 em 50: $n = 5(5)100(50)500$. Cada situação foi simulada 5.000 vezes. Para gerar dados univariados, as f.d.p.'s consideradas foram: normal (100,100); log-normal com $\sigma^2 = 1$; 0,1; e 0,001; χ^2 com $\nu = 1, 5$ e 30 graus de liberdade; exponencial, com $\alpha = 1, 5$ e 20; uniforme (0,1).

Para medir as taxas de erro tipo I, foram consideradas f.d.p. normais, para os níveis de probabilidade de 5% e 1%. O total de rejeições foi calculado em cada caso e as proporções de erros tipo I empírica foram obtidas. A inadequação do teste aplicado foi avaliada por um afastamento entre o valor nominal adotado e a taxa empírica calculada.

Para avaliar o poder do teste, foram consideradas f.d.p. não-normais. A hipótese nula de que os dados em estudo pertencem a uma amostra aleatória proveniente de uma distribuição normal, quando não rejeitada, refletiu uma situação na qual se cometeu o erro tipo II. A proporção de casos em que as respectivas hipóteses foram rejeitadas foi calculada. Essa proporção mediu o poder empírico dos testes empregados. O critério de avaliação no caso univariado foi o da comparação dessa taxa com o valor da taxa de poder empírico obtida pelo teste W de Shapiro & Wilk (1965).

RESULTADOS E DISCUSSÃO

Sob a hipótese nula de que os dados provêm de uma distribuição normal, avaliou-se a taxa de erro tipo I. Verificou-se pela Figura 1(a) que as estatísticas Z_1 e W mostraram-se adequadas quanto ao controle da taxa de erro tipo I, mesmo para pequenas amostras, ou seja, $n < 30$. A estatística K_1 aproximou-se do valor de 5% de probabilidade somente para um tamanho amostral igual ou superior a 75. É conveniente salientar que para os valores da significância nominal de 5% e de 1%, são esperados, em 99%

dos casos, resultados das taxas empíricas no intervalo de 4,2% a 5,8% e 0,67% a 1,42%, respectivamente.

Para o caso da curtose, verificou-se que as estatísticas Z_2 e K_2 não controlaram adequadamente a taxa de erro tipo I. A taxa de erro tipo I e estatística Z_2 aproximou-se do valor nominal de forma razoável para $n \geq 100$. O pior resultado, quanto ao controle da taxa de erro tipo I, ocorreu com a estatística K_2 , pois os resultados obtidos estavam aquém do valor nominal estabelecido, independentemente do tamanho amostral. Para essa estatística, houve tendência de a taxa de erro tipo I se aproximar do valor nominal com o aumento do valor de n .

Na Figura 1(b) encontram-se os resultados para o valor nominal de significância de 1%. Verificou-se que o teste W (Shapiro & Wilk, 1965) controlou adequadamente a taxa de erro tipo I para amostras de todos os tamanhos, como era esperado. A estatística K_1 , para amostras de tamanho $n < 25$, tendeu a subestimar o valor nominal de significância (1%). Para amostras superiores a 25, aproximou-se consideravelmente do valor nominal, tornando-se bastante razoável sua aproximação para amostras de tamanho $n \geq 200$. Já a estatística Z_1 tendeu a superestimar o valor da significância nominal para $n < 100$, e a partir desse valor, aproximou-se do valor nominal de 1%. As estatísticas Z_2 e K_2 também não controlaram a taxa de erro tipo I, pois mantiveram-se acima do valor nominal para $n > 100$. A estatística K_2 tendeu lentamente ao valor nominal para grandes tamanhos amostrais ($n > 500$), embora com $n > 25$, as taxas empíricas de erro tipo I já pertenciam ao intervalo de confiança para o valor nominal da significância. Nesse caso, o pior desempenho ocorreu com a estatística Z_2 , que gradualmente tendeu ao valor nominal, porém, superestimando-o para todos os valores de n . As estatísticas relacionadas com assimetria apresentaram-se com melhores resultados para o controle da taxa de erro tipo I. No caso em que o valor nominal foi de 1%, verificou-se que todas as estatísticas, exceto a estatística W, não controlaram a taxa de erro tipo I adequadamente, para pequenas amostras.

À medida que o valor da significância nominal diminuiu, os testes tenderam a apresentar maiores taxas de erro tipo I relativas, indicando que a aproximação assintótica dos testes possui "cauda" mais leve do que a normal $N(0, 1)$ ou a qui-quadrado limitantes. É razoável pensar que o teste deveria ser indicado para ser usado com valor nominal de 5%. Nesse caso, o critério Z_1 para assimetria e o Z_2 para curtose deveriam ser recomendados com $n \geq 25$. Para uso a 1%, as estatísticas K_1 e K_2 foram mais adequadas e poderiam ser recomendadas com n a partir de 25.

Para avaliar o poder do teste, foram consideradas f.d.p.'s diferentes da normal. Os resultados obtidos pelas distribuições assimétricas, ou seja, exponencial com os parâmetros $\alpha = 1, 5$ e 20 , qui-quadrado com 1 e 5 graus de liberdade e log-normal com $\sigma^2 = 1$, foram semelhantes e podem ser bem representados (Figura 2) para fins de ilustração, usando-se a distribuição exponencial com $\alpha = 1$. O teste W foi superior em poder às demais estatísticas, pois alcançou um poder elevado a partir de amostras pequenas, ou seja, $n \geq 20$, para os dois níveis de probabilidades especificados. Resultados em que o teste de Shapiro-Wilk é superior a outros testes de normalidade também podem ser encontrados no estudo feito por Shapiro & Wilk (1965). As estatísticas Z_1 e K_1 , considerando-se o valor nominal de 5%, atingiram poder elevado para $n \geq 20$, sendo a estatística Z_1 superior a K_1 . As estatísticas W, Z_1 e K_1 praticamente igualam-se em poder quando $n > 50$. As estatísticas referentes à curtose, Z_2 e K_2 , atingiram poder elevado ($> 80\%$) quando $n \geq 40$. Considerando-se o valor nominal de 1%, verificou-se que as estatísticas Z_1 e K_1 atingiram poder elevado quando $n \geq 25$. As estatísticas Z_2 e K_2 atingiram elevado poder para $n > 40$, considerando-se o valor nominal de 5% e $n > 45$ para 1%.

As distribuições aproximadamente simétricas foram representadas pelas distribuições qui-quadrado com 30 g.l. e log-normal com $\sigma^2 = 0,01$ e $0,001$. Como os resultados foram similares, optou-se por apresentar os resultados da distribuição de qui-quadrado com 30 graus de liberdade (Figura 3). À medida que ocorre o aumento dos

graus de liberdade na distribuição de qui-quadrado ou a diminuição de σ^2 na distribuição log-normal, ambas as distribuições tornam-se simétricas e mesocúrticas, resultando na redução do poder do teste, o qual aumentou lentamente, à medida que o tamanho da amostra aumentava.

Pela Figura 3(a), verificou-se que Z_1 e K_1 possuíam poder semelhante, atingindo um valor elevado para $n \geq 200$. Os critérios de curtose Z_2 e K_2 apresentaram baixo poder ($< 40\%$) para rejeitar a hipótese de nulidade, mesmo considerando $n = 550$. Na Figura 3(b), verificou-se que Z_1 e K_1 atingiram poder razoável para $n \geq 300$. Os critérios Z_2 e K_2 apresentaram baixo poder ($< 35\%$) em rejeitar a hipótese de que a distribuição é normal. É interessante notar que o teste W manteve-se com poder inferior aos critérios de assimetria, tanto para 5% quanto para 1%, resultado não esperado, uma vez que o mesmo é considerado extremamente poderoso (Shapiro & Wilk, 1965). Esse resultado em que o teste de Shapiro-Wilk é inferior quanto ao poder em relação às estatísticas baseadas em assimetria e curtose também foi encontrado por Oja (1981, 1983). Nesse caso, é notório que um teste baseado em desvios de simetria seja mais poderoso que o teste W de Shapiro & Wilk (1965) para detectar uma distribuição não-normal, principalmente em se tratando de uma distribuição com coeficiente de assimetria próximo de zero. O teste W para ser considerado de alto poder ($> 80\%$) tem que ser empregado nessa distribuição para $n \geq 300$, com valor nominal de significância de 5% e para $n \geq 500$, com significância nominal de 1%.

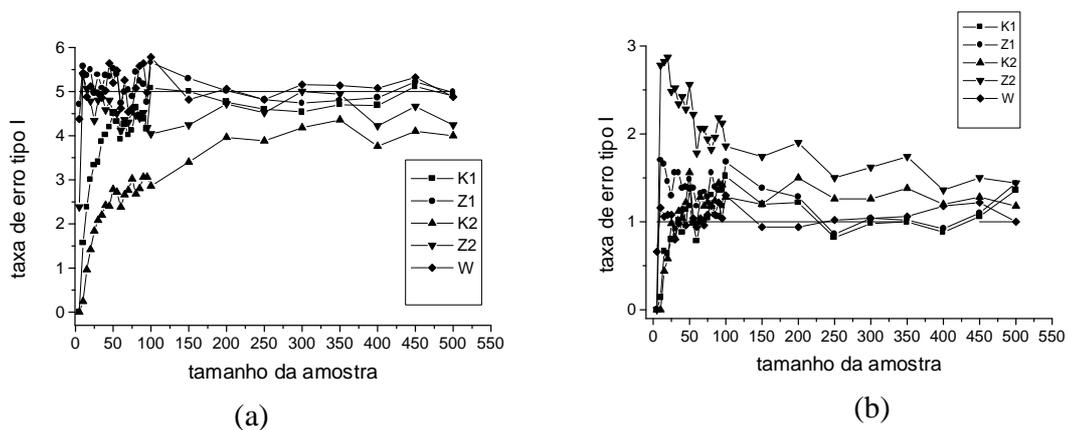


FIGURA 1 – Taxa de erro tipo I considerando o valor nominal de 5% (a) e de 1% (b).

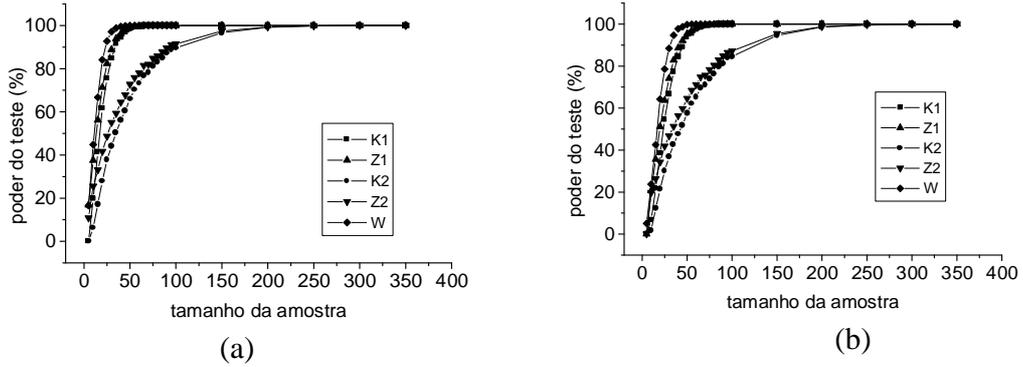


FIGURA 2 – Poder do teste considerando-se a distribuição exponencial com $\alpha = 1$ e valores de significância de 5% (a) e 1% (b).

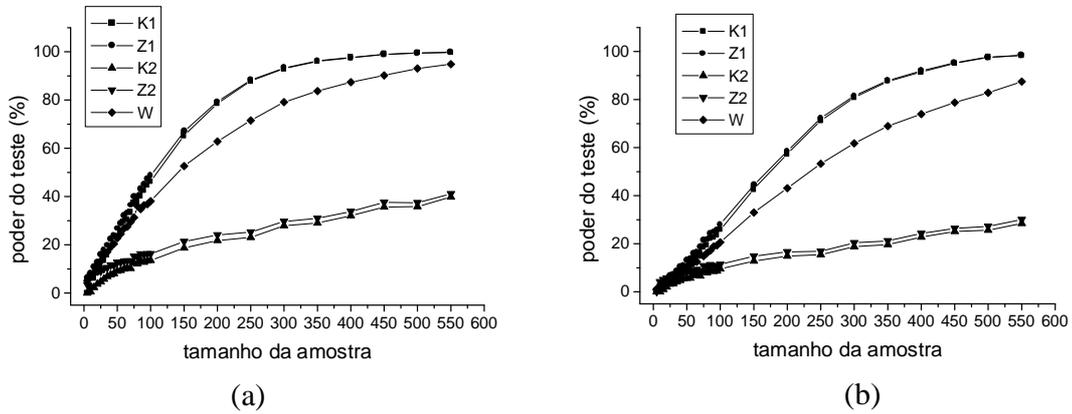


FIGURA 3 – Poder do teste considerando-se a distribuição qui-quadrado com 30 g.l. com os valores de significância de 5% (a) e de 1% (b).

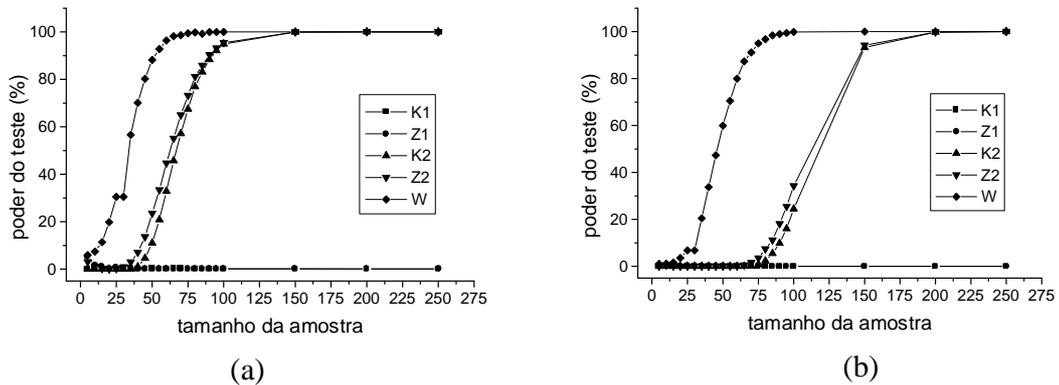


FIGURA 4 – Poder do teste considerando-se a distribuição uniforme (0,1) com os valores de probabilidades de 5% (a) e de 1% (b).

Uma vez que a distribuição uniforme é simétrica e platicúrtica, nas Figuras 4(a) e 4(b), foi possível perceber que as estatísticas Z_1 e K_1 apresentaram taxas de poder próximas a zero para detectar o desvio de normalidade. Sendo assim, a distribuição poderia ser considerada erroneamente como normal. Dessa forma, seria cometido o erro tipo II em 100% dos casos. O que surpreende é o fato de o teste não apresentar resultados que oscilem em torno do valor nominal adotado, como acontece em uma distribuição normal com $\sqrt{\beta_1} = 0$. Nessa situação, Z_2 e K_2 são extremamente sensíveis para detectar os desvios de normalidade. Os resultados para as estatísticas Z_2 e K_2 , conforme a Figura 4(a), mostraram baixo poder em detectar os desvios de normalidade para amostras com $n < 50$. A partir desse tamanho amostral, o poder aumentou conforme aumentou o tamanho da amostra, com uma grande taxa de crescimento. A estatística W foi superior ao critério de curto-se, detectando desvios de normalidade com amostras inferiores a 50. Essas estatísticas tenderam a se igualar em poder quando $n > 150$. No caso em que o valor nominal de significância considerado foi de 1%, conforme apresentado na Figura 4(b), verificou-se que as estatísticas Z_2 e K_2 tiveram baixo poder em detectar os desvios de normalidade quando $n < 100$. O critério W, no caso de 1%, apresentou superioridade em detectar os desvios de normalidade, embora tivesse baixo poder em detectar desvios de normalidade para pequenas amostras, $n < 50$. A partir desse tamanho amostral, o teste atingiu alto poder. Para esse caso, os resultados em que a estatística W superou em poder as demais estatísticas está coerente com os encontrados no estudo de Shapiro & Wilk (1965).

CONCLUSÕES

As estatísticas Z_1 e Z_2 possuem aproximações assintóticas normais para $n \geq 25$, com $\alpha = 5\%$ e podem ser recomendadas para uso rotineiro no caso univariado para testar a hipótese de normalidade dos dados;

As estatísticas K_1 e K_2 possuem aproximações assintóticas melhores que Z_1 e Z_2 para menor valor do nível de significância, sendo recomendadas para $n \geq 25$ e $n \geq 100$, respectivamente, as quais mantêm o controle da taxa de erro tipo I e elevado poder (poder superior a 80%);

No caso de distribuições com simetria próxima de zero e não normais, as estatísticas baseadas em desvios de simetria apresentam maior poder do que a estatística W de Shapiro & Wilk (1965).

REFERÊNCIAS BIBLIOGRÁFICAS

- BOCK, R. D. **Multivariate statistical methods in behavioral research**. Chicago: MacGraw-Hill, 1975. 623 p.
- CAMPOS, H. **Estatística experimental não paramétrica**. Piracicaba: ESALQ, 1983. 349 p.
- KOLMOGOROV, A. N. Sulla determinazione empirica di una legge di distribuzione. **Giornale dell'Istituto degli Attuari**, [S.l.], v. 4, p. 83-91, 1933.
- LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality. **Biometrika**, Great Britain, v. 62, n. 3, p. 399-402, June 1967.
- OJA, H. New tests for normality. **Biometrika**, Great Britain, v. 70, n. 1, p. 297-299, Apr. 1983.
- OJA, H. Two location and scale-free goodness-of-fit tests. **Biometrika**, Great Britain, v. 68, n. 3, p. 637-640, Dec. 1981.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete sample). **Biometrika**, Great Britain, v. 52, n. 3, p. 591-611, 1965.