A QUESTÃO DA NÃO NORMALIDADE:
UMA REVISÃO[1]

Francisco Alberto Pino[2]

RESUMO: Apresenta-se, neste estudo, uma revisão da extensa literatura sobre aspectos teóricos da não normalidade em dados estatísticos. O problema é apresentado na seção 1 e suas causas, isto é, as situações nas quais a não normalidade ocorre, são apresentadas na seção 2. Os efeitos da não normalidade aparecem na seção 3. Na seção 4, testes para normalidade são discutidos com o objetivo de detectar o problema no conjunto de dados observados. Finalmente, na seção 5 discute-se como lidar com a questão, principalmente através de procedimentos de transformação dos dados para remover os efeitos da não normalidade.

Palavras-chave: *teste de normalidade, transformação de Box-Cox, assimetria e curtose.*

THE QUESTION OF NON-NORMALITY:
A REVIEW

ABSTRACT: This study presents a review of the extensive literature on the theoretical aspects of non-normality in statistical data. The problem is presented in section 1, and its causes, i.e., the situations in which non-normality occurs, are presented in section 2. The effects of non-normality appear in section 3. In section 4, tests for normality are discussed in order to detect the problem in the observed data set. Finally, section 5 discusses how to address this question mainly through data transformation procedures to remove non-normality effects.

Key-words: *test for normality, Box-Cox transformation, skewness and kurtosis.*

JEL Classification: *C12, C89.*

[1]Registrado no CCTC, REA-16/2015.

[2]Engenheiro Agrônomo, Doutor, Pesquisador Científico Aposentado do Instituto de Economia Agrícola, São Paulo, Estado de São Paulo, Brasil (e-mail: drfapino@gmail.com).

1 - INTRODUÇÃO

No início do século XX, o aparecimento dos testes de significância revolucionou a teoria e a prática estatística. Entretanto, eles se apoiavam na suposição de que os dados observados eram uma amostra aleatória de uma população hipotética com distribuição normal (HOTELLING; PABST, 1936). Logo surgiram estudos a respeito dos erros que poderiam resultar da aplicação desses testes quando a distribuição não fosse normal, como Carlson (1932) e outros[3]. Um século depois, há extensa literatura sobre normalidade e não normalidade (HENZE; WAGNER, 1997), assunto que tem recebido muita atenção na estatística aplicada. O objetivo geral deste artigo é apresentar uma revisão metodológica com base na vasta literatura sobre a questão, tendo em vista, principalmente, o contexto de variáveis agronômicas.

O conhecimento da forma da distribuição de probabilidade de uma variável aleatória é útil e, às vezes, essencial em problemas estatísticos[4]. Uma vez que a forma da distribuição esteja determinada é possível estimar seus parâmetros, construir intervalos de confiança e testar hipóteses. A caracterização das distribuições de probabilidade mais comuns, com a normal, encontra-se em qualquer livro de Estatística Matemática. Uma distribuição de probabilidade pode ser caracterizada de diversas formas: pela sua função densidade, pela sua função característica, pela sua função geradora de momentos, pelo conjunto de seus momentos. Na prática, é usual utilizar os quatro primeiros momentos para caracterizar uma distribuição amostral: a) o primeiro momento dá uma medida de localização ou tendência central (média, mediana e moda); b) o segundo, uma medida de dispersão (variância, desvio padrão, coeficiente de variação e amplitude); c) o terceiro, uma medida de assimetria; e d) o quarto, chamado *kurtosis* (ou curtose), uma medida da proeminência do pico e da cauda da curva de distribuição (FINUCAN, 1964). Também o conjunto M das esperanças das estatísticas de ordem de amostras de uma distribuição pode determiná-la completamente (ARNOLD; MEEDEN, 1975).

Distribuição normal (ou gaussiana, ou de Gauss). Diz-se que uma variável aleatória Y tem distribuição normal, com média μ e variância σ^2, e escreve-se $Y \sim N(\mu, \sigma^2)$, se sua função densidade de probabilidade é dada por:

$$f(y) = (2\pi\sigma^2)^{1/2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

para $-\infty < y < \infty$, e $\sigma > 0$. Ela é chamada de distribuição normal padrão (Figura 1) se tiver média igual a zero e variância igual a um: $Y \sim N(0,1)$

2 - NÃO NORMALIDADE

De maneira tautológica, considere-se que a não normalidade ocorre quando alguma das variáveis que descrevem um fenômeno segue qualquer distribuição de probabilidade que não seja a normal, por razões intrínsecas ao fenômeno. Existem casos em que a não normalidade é evidente, por exemplo: a) quando há restrições sobre os valores das observações; b) quando a distribuição tem caudas pesadas ou deformações em relação à distribuição normal; e c) quando uma variável aleatória é definida pela razão entre outras duas.

Restrições. Uma das restrições mais comuns aos valores que as observações podem assumir é que elas sejam estritamente positivas (ou pelo menos, não negativas). Isso acontece com muitas variáveis que aparecem em estudos com estatísticas agrícolas, por exemplo, área plantada e produção. Um caso ainda mais restritivo é o de dados de contagem, que devem ser estritamente inteiros e não negativos, por exemplo, número de plantas, número de animais e número de trabalhadores.

[3]Shewhart e Winters (1928), Rider (1929), Rietz (1931), Chesire, Oldis e Pearson (1932), Perlo (1933 apud HOTELLING; PABST, 1936).

[4]A normalidade pode ser estudada diretamente sobre uma variável ou sobre os erros de um modelo para essa variável (como num modelo de regressão) e até mesmo sobre os erros de uma série temporal.

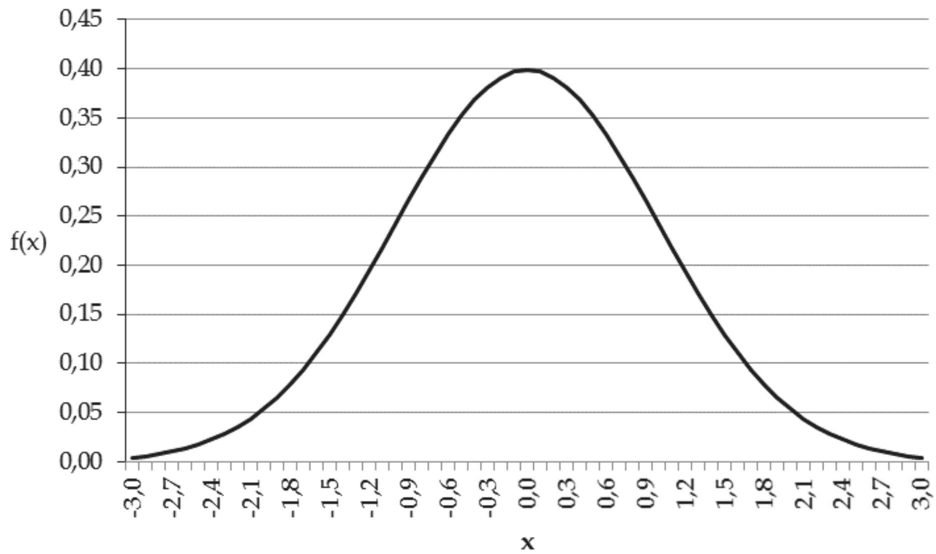

Figura 1 - Distribuição Normal Padrão.
Fonte: Dados da pesquisa.

Caudas pesadas. Diz-se que algumas distribuições possuem caudas pesadas, no sentido de que elas apresentam valores distantes das medidas de localização[5], com probabilidade maior do que a distribuição normal. Caudas pesadas ocorrem, por exemplo, quando a variância[6] é muito grande, até mêsmo infinita, como é o caso da distribuição de Cauchy. Também podem estar associadas à ocorrência de valores discrepantes ou extremos (*outliers*), entendidos como valores muito distantes daqueles das demais observações, em outras palavras, valores excessivamente grandes ou pequenos em relação aos outros. Embora um valor discrepante possa resultar de um erro de medida, também pode ser um resultado genuíno, indicando um comportamento extremo da variável, que merece ser estudado, e não removido.

Caudas pesadas, geralmente, se manifestam com uma proeminência do pico e da cauda da curva de distribuição, podendo ser medida pela curtose. A curtose[7] igual a zero chama-se **mesocúrtica**, e indica uma distribuição com achatamento semelhante ao da distribuição normal. Para valores maiores do que zero ela chama-se **leptocúrtica**, indicando uma distribuição mais afunilada ou centralmente concentrada do que a normal, e com caudas mais pesadas que esta, no sentido de que se podem obter valores muito distantes da média. A curtose negativa, chamada **platicúrtica**, indica uma distribuição mais achatada do que a normal. Portanto, curvas mais ou menos achatadas em relação a uma distribuição normal significam não normalidade.

Assimetria. A distribuição dos dados é simétrica quando eles se distribuem da mesma forma tanto acima quanto abaixo do meio da distribuição, como é o caso da distribuição normal. Distribuições assimétricas, obviamente, afastam-se da normalidade, podendo ser medidas pelo terceiro momento. A assimetria (ou obliquidade[8]) igual a zero indica uma distribuição simétrica, como a normal. Valores posi-

[5]Medidas de tendência central, ou medidas de posição, ou medidas de localização indicam o meio da distribuição dos dados (como a média, a mediana e a moda) ou outros pontos importantes da distribuição (como os quantis).

[6]Medidas de dispersão indicam a variabilidade dos dados (como a variância, o desvio padrão, o desvio médio absoluto, a amplitude total e a distância interquartílica).

[7]Em inglês, *kurtosis*.

[8]Em inglês, *skewness*.

tivos indicam assimetria à direita, isto é, a cauda direita da distribuição, onde estão os valores acima da média, é mais pesada. Valores negativos indicam assimetria à esquerda.

Razão. Considere-se uma variável Z definida como a razão entre outras duas variáveis, X e Y:

$$Z = \frac{X}{Y}$$

Da teoria estatística sabe-se que, se duas dessas variáveis tiverem distribuição normal, a terceira não o terá, mesmo se houver independência entre duas delas (ver, por exemplo, KEENE, 1995). De fato: a) se X e Y forem normais, então, Z terá distribuição de Cauchy (que tem média e variância infinitas); b) se Z e X forem normais, então, Y terá distribuição de Cauchy; e c) se Z e Y forem normais, então, X não terá distribuição normal[9]. Andrews e Mallows (1974) apresentam condições necessárias e suficientes para que uma variável aleatória Z possa ser gerada com a razão X/Y, onde X e Y são independentes e X tem distribuição normal padrão.

Exemplos da razão entre duas variáveis são:

a) A produtividade (ou rendimento) agrícola[10] é calculada pela produção[11] dividida pela área plantada (ou pelo número de plantas);
b) A densidade de cultivo agrícola é calculada pelo número de plantas dividido pela área plantada[12];
c) A produtividade pecuária é calculada pela produção dividida pelo número de animais[13]; e
d) O preço é dado pelo valor da transação dividido pela quantidade vendida (expressa em peso, ou em volume, ou em número de unidades, etc.), embora, na prática comercial, o preço seja estabelecido e o valor da transação, calculado depois.

Em todos esses casos, certamente a suposição de normalidade não é correta, às vezes, por mais de um motivo.

3 - EFEITOS DE DESVIOS DA NORMALIDADE

Comumente espera-se que as observações de amostras de populações sigam uma distribuição normal ou aproximadamente normal. De fato, a suposição de normalidade é uma das mais comuns nos procedimentos estatísticos. Entretanto, essa suposição é, frequentemente, a que menos provavelmente é válida. É mais comum que as observações tenham uma distribuição apenas aproximadamente normal. Felizmente, boa parte da análise de variância pode ser desenvolvida sem tal suposição, que é necessária somente para justificar o uso de certos testes de significância formalmente precisos e certas fórmulas de estimação (JOHNSON; LEONE, 1964). Pode-se mostrar que é possível tolerar um razoável afastamento da normalidade, com pequeno efeito prático na análise de variância convencional. A não normalidade não leva a erros sérios de interpretação de médias simples, que na maioria dos casos são aproximadamente normais, ao contrário do que acontece com a distribuição das estatísticas de segunda ordem (HOTELLING; PABST, 1936).

Num modelo de regressão, o grau de não normalidade das variáveis independentes é decisivo para a não normalidade da variável dependente e para a sensibilidade da análise de variância a esse afastamento da normalidade (BOX; WATSON, 1962). A utilidade da estimação num modelo de regressão depende do grau em que as suposições do modelo, incluindo heterocedasticidade e normalidade dos erros, são satisfeitas. A falta de normalidade não introduz viés na estimação dos parâmetros, mas sim, na dos desvios padrões, afetando a validade dos

---

[9]Pode-se mostrar que a distribuição do produto de duas variáveis normais não é normal (em alguns casos, ela converge para uma normal), podendo ser apenas aproximada (WARE; LAD, 2003; CRAIG, 1936).

[10]Em inglês, *field crop yield*.

[11]Em inglês, *output, production, yield.*

[12]Também chamada densidade de plantio, ou *stand* de plantas, palavra eventualmente aportuguesada para estande (em inglês, *plant stand*), é definida como o número de plantas por unidade de área, sendo determinada pelo espaçamento entre as plantas no campo. A expressão "número de plantas" pode ser dita "número de pés" por alguns autores.

[13]A expressão "número de animais" pode ser dita "número de cabeças" por alguns autores.

intervalos de confiança e dos testes de hipótese (BERNIER; FENG; AZAKAWA, 2011). Kronmal (1993) discute problemas que surgem quando variáveis do tipo razão aparecem como dependentes ou independentes num modelo de regressão. O autor recomenda que razões sejam utilizadas somente no contexto de modelo linear completo em que o intercepto está presente, mostrando que seu uso pode levar a inferências enganosas.

Os efeitos da não normalidade podem ocorrer:

a) Na **estimação de máxima verossimilhança**, que pressupõe uma distribuição de probabilidade para poder deduzir as fórmulas de estimação de seus parâmetros.
b) Na **estimação por intervalo**. Na estimação por ponto não é necessário supor uma distribuição, exceto para estimadores de máxima verossimilhança.
c) Associada à assimetria da distribuição, quando as medidas de localização (média, mediana e moda) deixam de coincidir. De modo geral, a não normalidade não conduz a erros muito sérios na interpretação de médias simples, embora deva ser assinalado que a média é mais sensível a *outliers* do que a mediana.
d) Na aplicação de **testes de significância** baseados na suposição de normalidade, como o teste t de Student ou o teste F, e na análise de variância, esses efeitos podem se mostrar sérios. Entretanto, estudos de simulação têm mostrado que o teste F resiste bastante a afastamentos da normalidade. Simulações mostraram que o teste t para o coeficiente de correlação é robusto a afastamentos da normalidade quando as variáveis são independentes, mas não quando elas são dependentes (EDGELL; NOON, 1984). O teste t de Student também se mostra robusto, a menos que a distribuição seja muito assimétrica ou com caudas muito pesadas (LACHENBRUCH, 2003).
e) Quando se **comparam grupos**. O efeito da não normalidade não é sério quando se comparam médias em experimentos com controle interno, como ocorre com a maioria deles, porém, é mais sério quando se comparam variâncias de grupos independentes de observações.
f) Em casos de **heterocedasticidade**, ou falta de homogeneidade das variâncias, que costuma ser motivo de preocupação maior. Quando se comparam médias de dois grupos de observações pelo teste t, uma suposição básica é a de que a variância em cada grupo de observações é a mesma, caso contrário, as probabilidades calculadas serão diferentes daquelas dadas nas tabelas de significância. O problema é mais sério quando o número de observações nos dois grupos é muito diferente. No estudo dos modelos lineares, quer os de posto completo (regressão), quer os de posto incompleto, a suposição de normalidade dos resíduos é necessária quando se introduzem testes de hipótese e os afastamentos da normalidade costumam estar associados à heterocedasticidade.

4 - TESTANDO A SUPOSIÇÃO DE NORMALIDADE

É bom procedimento verificar se as suposições do modelo ou da análise que se pretende utilizar num trabalho estão satisfeitas, em particular a de normalidade. Duas questões surgem de imediato: a) como testar se um conjunto de observações provém de uma população com distribuição normal; b) em caso negativo, como proceder.

Para verificar se a distribuição é normal, a primeira coisa a fazer é um gráfico de frequências das observações, para examinar se existem assimetrias. Ao se desconfiar da existência de não normalidade, o passo seguinte é testar a hipótese nula de que a distribuição das observações é normal, contra a hipótese alternativa de que não o é.

Chama-se teste de aderência (ou de ajustamento) o problema de testar a hipótese de que uma dada amostra provém de uma população com uma função densidade específica. Um caso clássico é o de testar se as observações provêm de uma população com distribuição normal. Para tanto, utiliza-se uma estatística que pode ser escrita de forma geral como

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

onde O_i é uma frequência observada e E_i é uma frequência esperada. Com essa estatística constrói-se um teste de qui-quadrado de aderência (MOOD; GRAYBILL; BOES, 1974).

Desde que surgiu o interesse pela questão, testes para normalidade/não normalidade e suas variações, para o caso univariado e o caso multivariado, têm sido propostos às catadupas. Ao final do século XX, existiam cerca de 40 testes[14] mais específicos para não normalidade que se fundamentavam em características da distribuição normal (DUFOUR et al., 1998). Na década seguinte, já se contavam mais de 50 métodos para estudar o ajustamento à normal (MECKLIN; MUNDFROM, 2004; DESMOULINS-LEBEAULT, 2004)[15]. Há várias maneiras de categorizar esses testes, como a seguinte: a) testes baseados na função de distribuição empírica[16]; b) testes baseados em regressão e correlação; c) testes baseados em momentos (DUFOUR et al., 1998; SEIER, 2002). Alguns dos mais utilizados são descritos a seguir.

4.1 - Testes Baseados na Função de Distribuição Empírica

Existem importantes testes baseados na função de distribuição empírica[17] (ou na função característica empírica), que consistem em compará-la com a função de distribuição acumulada[18] da normal. Apresentam-se os três principais a seguir[19]. Sejam n observações $\hat{z}_{in}$ ou $\hat{z}_i$, para $1 \le i \le n$, e a transformação $U_i = \Phi(\hat{Z}_{in}/s)$, onde $\Phi(.)$ denota a função de distribuição cumulativa da normal padrão $N(0,1)$.

Teste de Kolmogorov-Smirnov[20]. Esse teste é definido por:

$$D = \max(D^+, D^-)$$

onde $D^+ = \max[(i/n) - U_i]$ e $D^- = \max[U_i - (i-1)/n]$. A tabela para o caso em que a média e a variância precisam ser estimadas a partir da amostra é dada em Lilliefors (1967). Uma revisão dos pontos de significância foi apresentada em D'Agostino e Stephens (1986 apud DUFOUR et al., 1998), enquanto que uma extensão para o caso multivariado foi considerada por Justel, Peña e Zamar (1994).

Teste de Cramer-von Mises[21]. Esse teste é definido por:

$$W^2 = \sum_{i=1}^{n} [U_i - (2i-1)/2n]^2 + 1/12n$$

Chen, Lockhart e Stephens (1993) desenvolveram a teoria para que o teste de Cramer-von Mises possa ser aplicado após o uso da transformação de Box-Cox.

Teste de Anderson-Darling[22]. Esse teste, é definido por:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\ln U_i + (2n+1-2i)\ln(1-U_i)]$$

Chen, Lockhart e Stephens (1993) desenvolveram a teoria para que o teste de Anderson-Darling possa ser aplicado após o uso da transformação de Box-Cox.

---

[14] Geralmente, são testes abrangentes (em inglês, *omnibus tests*) ou globais, no sentido de que compreendem vários itens, ou cuja hipótese nula refere-se globalmente a todos os grupos do estudo.

[15] Um bom trabalho de revisão sobre testes para normalidade, especialmente para o caso multivariado, é o de Mecklin e Mundfrom (2004).

[16] A palavra "empírica" tem aqui o significado de "baseada nos dados ou observações".

[17] Referenciada pela sigla em inglês EDF.

[18] Referenciada pela sigla em inglês CDF.

[19] Outros testes baseados na função de distribuição empírica são apresentados por Csörgő (1986), Henze e Wagner (1997), Rao e Ali (1998), Bogdan (1999), Akbilgiç e Howe (2011), Su e Kang (2015).

[20] Desenvolvido por Kolmogorov (1933 apud DUFOUR et al., 1998) e por Smirnov (1948).

[21] Desenvolvido por Cramér (1928 apud SHAPIRO; WILK, 1965) e por Von Mises (1928 apud ARNOLD; EMERSON, 2011),

[22] Desenvolvido por Anderson e Darling (1954).

4.2 - Testes Baseados em Regressão e Correlação

Os testes de regressão e correlação baseiam-se no fato de que a variável $Y \sim N(\mu, \sigma^2)$ pode ser expressa como $y = \mu + \sigma x$, onde $X \sim N\ (0,1)$, conforme Seier (2002). Esses testes usam a razão de duas estimativas obtidas de estatísticas de ordem: uma estimativa ponderada de mínimos quadrados, dado que a população é normalmente distribuída, e a estimativa não viesada da variância amostral, para qualquer população (DUFOUR et al., 1998). Dois desses testes são apresentados a seguir[23].

Teste de Shapiro-Wilk[24]. Esse é o mais conhecido teste baseado em regressão e correlação, e definido por:

$$W = \frac{1}{(n-k)s^2}\left(\sum_{i=1}^{n} a_i \hat{z}_i\right)^2$$

com

$$a' = (a_1, \ldots, a_n) = \frac{c'V^{-1}}{(c'V^{-2}c)^{1/2}}$$

onde $c' = (c_1, \ldots, c_n)$ e V são o vetor de valores esperados e a matriz de covariâncias das estatísticas de ordem da normal padrão, respectivamente. Uma modificação desse teste, para grandes amostras, é o teste de Shapiro-Francia (*SF*), apresentado por Shapiro e Francia (1972). Uma extensão para o caso multivariado é dada por Srivastava e Hui (1987).

Teste de D'Agostino[25]. Esse teste consiste numa combinação linear das observações ordenadas e é definido por:

$$D = \frac{1}{n^2 s}\sum_{i=1}^{n} \hat{z}_i[i - (n+1)/2]$$

com

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (\hat{z}_i - \bar{z})^2$$

onde $\bar{z}$ é a média amostral.

4.3 - Testes Baseados em Momentos

Há testes baseados na assimetria e na curtose, os quais consistem em comparar a assimetria e a curtose da função de distribuição empírica (baseada nos dados) e aquelas da função de distribuição normal (SEIER, 2002):

$$Sk = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{z}_i^3}{\hat{\sigma}^3}$$

$$Ku = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{z}_i^4}{\hat{\sigma}^4}$$

onde *Sk* representa o terceiro momento (assimetria) e Ku representa a curtose[26]. Os testes podem usar um desses momentos ou ambos conjuntamente. Um teste usando assimetria e curtose é descrito a seguir[27].

Teste de Jarque-Bera[28]. Esse teste é definido por:

$$JB = n\left[\frac{1}{6}(Sk)^2 + \frac{1}{24}(Ku)^2\right]$$

onde $6/n$ e $24/n$ e são as variâncias assintóticas da assimetria e da curtose, respectivamente (ou *Ku -3*, se a curtose for definida com média 3).

4.4 - Comparação de Testes

O poder de cada um desses testes depende da natureza da não normalidade (SEIER, 2002). A falta de soluções exatas para as distribuições amostrais levaram ao desenvolvimento de muitos estudos de

[23]Há outros testes de regressão e correlação como aqueles descritos em Filliben (1975) e em Weisberg e Bingham (1975 apud DUFOUR et al., 1998).

[24]Desenvolvido por Shapiro e Wilk (1965).

[25]Desenvolvido por D'Agostino (1971).

[26]A assimetria também pode ser representada por $\sqrt{b_1}$ e a curtose, por b_2 .

[27]Outros testes desse tipo são os descritos por Bowman e Shenton (1975), Isogai (1982), Doornik e Hansen (1994), Urzúa (2007); Nakagawa, Hashiguchi e Niki (2012).

[28]Descrito em Jarque e Bera (1987) e em Bowman e Shenton (1975).

comparação de testes de normalidade quanto ao seu poder, usando simulações de Monte Carlo (POITRAS, 2006).

Os testes de Kolmogorov-Smirnov, Cramer-von Mises e Anderson-Darling apresentam algumas vantagens sobre o teste de aderência de qui-quadrado tradicional, como maior poder[29].

O poder desses testes também varia de acordo com o número de observações e com o formato da distribuição empírica. Num estudo de simulação, o teste de Shapiro-Wilk mostrou-se o mais poderoso para todos os tipos de distribuição e tamanhos de amostras, enquanto que o de Kolmogorov-Smirnov mostrou-se o menos poderoso. O teste de Anderson-Darling foi comparável ao de Shapiro-Wilk, seguido do teste de Lilliefors (RAZALI; WAH, 2010). O mesmo resultado foi obtido por Mendes e Pala (2003), ao comparar o poder dos testes de Shapiro-Wilk, Lilliefors e Kolmogorov-Smirnov.

Uma desvantagem dos testes baseados em assimetria e curtose é que qualquer distribuição simétrica tem esses momentos iguais aos de uma distribuição normal. Por esse motivo, os testes baseados em EDF e o teste de Shapiro-Wilk têm sido os mais utilizados no caso univariado (MECKLIN; MUNDFROM, 2004). Poitras (2006) argumenta que o poder de testes baseados em momentos pode ser comparado favoravelmente, em relação aos testes baseados em EDF ou em correlação. Além disso, argumenta que em alguns casos, um teste direcional, baseado na assimetria ou na curtose, pode ser preferível a um teste abrangente, mesmo que baseado em ambos os momentos (assimetria e curtose).

Inúmeras outras comparações entre testes podem ser encontradas na literatura[30]. A conclusão geral é que nenhum teste domina os demais sob todas as condições (AKBILGIÇ; HOWE, 2011).

[29]Segundo D'Agostino e Stephens (1986, cap. 2 apud SAS, 2010).

[30]Como em Farrell, Salibian-Barrera e Naczk (2007), Hanusz e Tarasińska (2009), Adefisoye (2015).

5 - ESTRATÉGIAS PARA LIDAR COM A NÃO NORMALIDADE

Quando as observações afastam-se da distribuição normal, pode-se tomar um dos caminhos indicados a seguir.

Métodos sem suposição de normalidade. A maneira mais simples consiste em utilizar métodos estatísticos que dispensem a suposição de uma distribuição de probabilidade específica, como a normal. Embora isso seja possível em alguns casos, na maioria das vezes, significa abdicar de poderosas ferramentas estatísticas.

Métodos para a distribuição correta. Utilizar métodos estatísticos adequados para distribuições de probabilidade diferentes da normal, quando esse for o caso, constitui maneira lógica de tratar o problema. Por exemplo, a estimação de desvios absolutos mínimos, associada à norma L_1, é apropriada para populações com distribuição exponencial dupla, enquanto que a estimação de mínimos quadrados, associada à norma L_2, é apropriada para populações com distribuição normal.

Em muitos casos existem razões teóricas para a escolha da forma da distribuição, mas, às vezes, é necessário procurar o ajustamento de uma distribuição com base somente na amostra de que se dispõe. Muitas tentativas bem sucedidas têm sido feitas para ajustar modelos matemáticos a diferentes populações.

Métodos robustos. Utilizar métodos robustos, no sentido de não serem sensíveis a afastamentos da normalidade, constitui o caminho seguinte, mesmo que não sejam os mais adequados para a real distribuição da variável.

Normalidade assintótica. O teorema do limite central fornece a distribuição limite da média ou da soma de variáveis aleatórias independentes identicamente distribuídas. À medida que aumenta o tamanho da amostra dessa distribuição, ela se aproxima da normal; assim, em grande parte dos casos, para um tamanho conveniente da amostra já se pode admitir que as variáveis tenham distribuição aproximadamente normal e utilizar os métodos usuais de análise. A atenção deve concentrar-se, portanto, nas variáveis cuja convergência não é muito rápida.

A normalidade assintótica também pode ocorrer para outras estatísticas amostrais, como correlações, estatísticas de ordem, estatísticas de máxima verossimilhança, quantis, conforme Hoeffding (1948), entre outros. Uma estimativa da precisão da aproximação normal, com o objetivo de obter limites para amostras razoáveis é apresentada por Reiss (1974). Condições necessárias e suficientes para que sequências de valores subamostrais de uma estatística sejam normalmente assintóticas são apresentadas por Hartigan (1975). Mesmo em algumas situações complexas a normalidade assintótica pode acontecer (Dupačová; Wets, 1987). Condições para estimação com normalidade assintótica de modelos de séries estacionárias com raiz unitária comum foram apresentadas por West (1988).

Transformação dos dados. Finalmente, é possível aplicar uma transformação aos dados, de tal forma que os dados transformados tenham distribuição normal ou aproximadamente normal. Segundo Aitchison (1982), esse tipo de questão começou a ser considerada por McAlister (1879). Às vezes, o tipo de transformação pode ser sugerido pelo próprio problema. Por exemplo, num experimento de fixação de nitrogênio em plantas, como o crescimento das plantas acontece de forma aproximadamente logarítmica, parece natural que tomando os logaritmos da quantidade de nitrogênio nas plantas, as variâncias diferirão pouco entre tratamentos. Entretanto, nem sempre a transformação lógica é tão evidente (Eisenhart; Wilson, 1943). Efron (1982) discute condições sob as quais existe uma simples transformação monótona que torne uma variável aproximadamente normal, mas alerta para o fato de que nem sempre a normalização da variável conduz a uma estabilização da variância. Sprot (1973) assinala que uma análise da função de verossimilhança pode indicar quando a teoria da estimação para grandes amostras pode ser aplicada e quando ela é inadequada podendo levar a erros. O autor analisa transformações que podem melhorar a aproximação à normalidade das funções de verossimilhança e, também, a precisão de níveis de significância e dos intervalos de confiança baseados na teoria para grandes amostras. Shenton (1965) apresenta um método para transformar distribuições de Pearson de tipos I, III, V e VI em distribuições aproximadamente normais, trabalhando com regressões. Quando a população não é normal, a probabilidade de se obter um valor da estatística t de Student (ou de F) maior que um dado valor crítico difere do valor tabulado por certo fator. Bradley (1952) estuda correções para tais testes, permitindo o uso das tabelas comuns. Curtiss (1940) apresenta uma teoria matemática geral para certos tipos de transformação em uso, como a transformação pela raiz quadrada, a transformação logarítmica e a transformação pelo inverso do seno.

Geralmente, os desvios de normalidade são seguidos de heterocedasticidade. Além disso, uma correlação entre a variância e a média frequentemente implica assimetria excessiva. Assim, o que se procura é uma mudança de escala para estabilizar a variância, procedimento este que costuma também diminuir a assimetria e aproximar da normal a distribuição da variável. Deve-se alertar para o fato de que nem sempre a normalização da variável conduz a uma estabilização da variância (Efron, 1982). Entretanto, desvios moderados da normalidade não costumam constituir problema muito sério (Bartlett, 1936, 1947). Num modelo de regressão, uma maneira de tratar a violação da suposição de normalidade consiste em transformar a variável da saída, a fim de tornar simétrica a distribuição dos resíduos, não o das variáveis dependentes (Bernier; Feng; Asakawa, 2011).

5.1 - Transformação Potência

Talvez a família mais geral de transformações para normalidade seja a família paramétrica de transformação potência, também chamada transformação de Box-Cox, dada por:

$$y^{(\lambda)} = \begin{cases} \dfrac{(y+c)^{\lambda}}{\lambda}, & \text{para } \lambda \neq 0 \\ \log(y+c), & \text{para } \lambda = 0 \text{ e } y > -c \end{cases}$$

onde y representa a observação original, $y^{(\lambda)}$ representa a observação transformada, λ e c são parâmetros desconhecidos e log representa o logaritmo natural, proposta no artigo seminal de Box e Cox (1964)[31]. Supõe-se que para algum valor de λ e algum valor de c as observações transformadas sejam independentes e normalmente distribuídas com variância constante σ^2 e esperança $\mathbf{a}\boldsymbol{\theta}$, onde $\mathbf{a}$ é uma matriz conhecida de posto completo e $\boldsymbol{\theta}$ é um vetor de parâmetros desconhecidos. A densidade de probabilidade das observações é obtida por

$$\frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} \exp\left[-\frac{\left(\mathbf{y}^{(\lambda)}\mathbf{a}\boldsymbol{\theta}\right)'\left(\mathbf{y}^{(\lambda)}\mathbf{a}\boldsymbol{\theta}\right)}{2\sigma^2}\right] \times J(\lambda, \mathbf{y})$$

onde o Jacobiano da transformação é dado por

$$J(\lambda, \mathbf{y}) = \prod_{i=1}^{n} \frac{dy_i^{(\lambda)}}{dy_i}$$

Os parâmetros na fórmula acima podem ser estimados de duas maneiras: a) aplicando a teoria de máxima verossimilhança para grandes amostras; ou b) aplicando a teoria de Bayes, na qual se admite que as distribuições *a priori* dos θ´s e $\log(\sigma)$sejam uniformes sobre a região onde a função de verossimilhança está definida, e se obtém a distribuição *a posteriori* de λ. Um algoritmo para estimar o parâmetro da transformação de Box-Cox foi proposto por Asar e Dağ (2014). Entretanto, como o estimador de máxima verossimilhança do parâmetro da transformação de Box-Cox é muito sensível a valores discrepantes (ANDREWS, 1971), Yeo, Jonhson e Dene (2014) propuseram estimar esse parâmetro mediante a minimização da distância quadrática ponderada entre a função característica empírica dos dados transformados e a função característica de uma distribuição normal. Por outro lado, Foudjo (2013), trabalhando no âmbito de séries temporais, tratou de testes robustos para normalidade (em especial, o de Shapiro-Wilk), visando obter um estimador robusto para o parâmetro da transformação Box-Cox.

O uso da transformação de Box-Cox antes da análise ou da modelagem é especialmente recomendado quando a variável assume somente valores positivos (POIRIER, 1978). Draper e Cox (1969) mostram que, mesmo quando não se consegue normalidade exata com a transformação, ela pode ser utilizada, podendo se obter um estimador consistente. A precisão da estimativa de λ depende muito do coeficiente de variação dos dados transformados. Atkinson (1973) compara três testes de hipóteses sobre o parâmetro da transformação: o teste da razão de verossimilhança utilizado por Box e Cox (1964), o teste exato proposto por Andrews (1971), para o parâmetro da transformação de Box-Cox, e um teste estatístico de uma normal assintótica derivada da função de verossimilhança. Conclui que o teste de Andrews é exato e mais fácil de calcular, mas os testes derivados da função de verossimilhança são uniformemente mais poderosos. Entretanto, considerando-se que a estimativa da média é muito afetada por valores extremos, intervalos para a média podem ser muito sensíveis a variações nos valores do parâmetro da transformação de Box-Cox, pois as caudas da distribuição resultante podem diferir bastante. A incorporação de informação *a priori* numa estimação bayesiana pode, eventualmente, diminuir essa sensibilidade da média a valores de λ (RUBIN, 1984).

Como a transformação de Box-Cox é definida somente para variáveis positivas, isto é, em que $y = 0$, a transformação foi estendida para valores negativos, resultando na transformação de Yeo-Johnson[32]:

$$x^{(\lambda)} = \begin{cases} \dfrac{(x+1)^{\lambda}-1}{\lambda}, & \text{para } \lambda \neq 0, x \geq 0 \\ \log(x+1), & \text{para } \lambda = 0, x \geq 0 \\ -\dfrac{(1-x)^{2-\lambda}-1}{2-\lambda}, & \text{para } \lambda \neq 2, x < 0 \\ -\log(1-x), & \text{para } \lambda = 2, x < 0 \end{cases}$$

[31]Pode-se mostrar que $\lim_{\lambda \to 0} y^{(\lambda)} = \log(y + c)$, conforme SAS (2008).

[32]Proposta por Yeo e Johnson (2000 apud WEISBERG, 2001).

Transformação semelhante, abrangendo valores negativos, é apresentada em Ahmad, Naing e Hussein (2007).

No caso multivariado, a transformação de Box-Cox pode ser escrita como:

$$y_j^{(\lambda_j)} = \begin{cases} \dfrac{(y_j + c_j)^{\lambda_j}}{\lambda_j} \times m_j^{\lambda_j - 1}, & \text{para } \lambda_j \neq 0 \\ \log(y_j + c_j) \times m_j, & \text{para } \lambda_j = 0 \text{ e } y_j > -c_j \end{cases}$$

para o j-ésimo elemento do vetor $v \times 1$ de respostas $y'i = (y_{i1}, \ldots, y_{iv})$. O caso de transformação conjunta de dados multivariados discutido em Bozdogan e Ramirez (1986) toma m_j=1, mas em Riani (2004), $m_j = \dot{y}_j$ é a média geométrica da j-ésima resposta.

Em resumo, trata-se de uma transformação bastante estudada sob muitos aspectos (Sakia, 1992). Muitas transformações comuns, como a logarítmica e a raiz quadrada, são apenas casos particulares da transformação potência, como será visto a seguir.

Transformação cúbica. É o caso da transformação potência em que $\lambda = 3$:

$$y^{(\lambda)} = \frac{(y + c)^3}{3}$$

Transformação quadrática. É o caso da transformação potência em que $\lambda = 2$, sendo usada para dados assimétricos à esquerda:

$$y^{(\lambda)} = \frac{(y + c)^2}{2}$$

Transformação linear. É o caso da transformação potência em que $\lambda = 1$, e tem-se apenas uma mudança de origem:

$$y^{(\lambda)} = y + c$$

Neste caso, nenhuma transformação é necessária, produzindo-se um resultado idêntico ao original.

Transformação raiz quadrada. É o caso da transformação potência em que $\lambda = \frac{1}{2}$:

$$y^{(\lambda)} = 2\sqrt{y + c}$$

A transformação raiz quadrada, introduzida por Bartlett (1936), é usada para variáveis com distribuição de Poisson e para dados de contagem de ocorrências (Osborne, 2010).

Transformação logarítmica. É o caso da transformação potência em que $\lambda = 0$:

$$y^{(\lambda)} = \log(y + c)$$

Quando uma variável é restrita a valores não negativos, a transformação logarítmica também estende os valores da variável para a reta real. A transformação logarítmica tem grande apelo e vem sendo utilizada há muito tempo com sucesso, sendo popular em análise de regressão e econometria. Já no trabalho de Bartlett (1947) encontram-se as transformações pela raiz quadrada, a angular e a logarítmica, que são as mais utilizadas. O autor indica a transformação $\log(1 + x)$ no lugar de $\log(x)$, para evitar dificuldades com zeros. As transformações mais comuns para dados assimétricos à direita são a inversa, a logarítmica e a raiz quadrada. Certos aspectos da transformação logarítmica foram vistos também por Bartlett e Kendall (1946).

Aplicando-se logaritmos dos dois lados da igualdade de uma razão entre duas variáveis, $Z = X / Y$, obtém-se:

$$\log Z = \log X - \log Y$$

e, neste caso, as três variáveis transformadas podem ser normais. Ao expressar a razão das variáveis como uma diferença de duas variáveis, as suposições da análise de variância ou da análise de regressão, geralmente, se tornam mais realistas (Keene, 1995).

Apesar de seu apelo, o fato de a transformação logarítmica produzir intervalos de confiança assimétricos em relação à média pode não ser desejável em alguns estudos. A demonstração é evidente, mas seja a seguinte estimativa por intervalo para a variável Z tomada nos logaritmos:

$$\log L_i = \log \hat{Z} - t_\alpha s \text{ e } \log L_s = \log \hat{Z} + t_\alpha s$$

Então,

$$L_i = \exp[\log \hat{Z} - t_\alpha s] = \hat{Z}/\exp[t_\alpha s]$$

e

$$L_s = \exp[\log \hat{Z} + t_\alpha s] = \hat{Z}\exp[t_\alpha s]$$

onde L_i é o limite inferior e L_s é o limite superior do intervalo de confiança para a variável original. O intervalo superior resulta diferente do intervalo inferior ao redor da estimativa:

$$L_s - \hat{Z} = \hat{Z}(e^{t_\alpha s} - 1) \neq \hat{Z}\left(\frac{e^{t_\alpha s} - 1}{e^{t_\alpha s}}\right) = L_i - \hat{Z}$$

Na verdade, é maior se a distribuição for assimétrica à direita, menor se a distribuição for assimétrica à esquerda, e igual se a distribuição for simétrica. Isso é especialmente ruim quando se trabalha com previsões, em modelos de séries temporais[33]. De fato, em que pesem as vantagens no uso da transformação de Box-Cox, da logarítmica em especial, em problemas de econometria (particularmente, em modelos ARIMA), no caso de dados econômicos que se afastam muito da normalidade, o uso dessa transformação parece não reduzir muito o problema, não compensando a inconveniência, esforço e custo extra, nem produzindo previsões melhores (NELSON JUNIOR; GRANGER, 1979).

Muitas vezes, o desvio padrão varia proporcionalmente com a média e, então, uma transformação nas observações pode ser suficiente para tornar constante a variância. Acontece que uma correlação entre o desvio padrão e a média é geralmente acompanhada de um grande desvio de normalidade, o que indica que a forma das observações é inadequada. É frequente, além disso, que a transformação que torna a variância constante também torne a distribuição das observações mais próxima da normal (DAVIES, 1960). Na verdade, "a transformação logarítmica é explicitamente recomendada quando o desvio padrão é proporcional ao valor da média" (KEENE, 1995).

As transformações logarítmicas (ou transformações log) constituem toda uma classe de transformações, não apenas uma transformação (OSBORNE, 2010). Variáveis log-normais (i.e., que se tornam normais após a transformação logarítmica) ocorrem em muitos campos, parecendo ser mais comuns quando os resultados são influenciados por muitos fatores independentes, como em ciências biológicas e também em ciências sociais. Deve-se notar, ainda, que diferentes bases do logaritmo podem produzir diferentes resultados de transformação, embora seja usual usar o logaritmo natural, com base *e* (OSBORNE, 2010).

Transformação raiz quadrada inversa. É o caso da transformação potência em que $\lambda = -\frac{1}{2}$:

$$y^{(\lambda)} = \frac{-2}{\sqrt{y + c}}$$

Transformação inversa (ou hiperbólica de primeira ordem, ou recíproca). É o caso da transformação potência em que $\lambda = -1$

$$y^{(\lambda)} = -\frac{1}{y + c}$$

É usada para razões e para dados fortemente assimétricos à direita. Essa transformação faz números pequenos se tornarem grandes, e vice versa, invertendo sua ordem (OSBORNE, 2010).

Transformação quadrática inversa (ou hiperbólica de segunda ordem). É o caso da transformação potência em que $\lambda = -2$:

$$y^{(\lambda)} = \frac{-2}{(y + c)^2}$$

Transformação cúbica inversa. É o caso da transformação potência em que $\lambda = -3$:

$$y^{(\lambda)} = \frac{-3}{(y + c)^3}$$

[33]O efeito da transformação sobre o intervalo de confiança é ainda pior no caso da transformação raiz quadrada e da transformação recíproca (BLAND; ALTMAN, 1996).

5.2 - Transformações Angulares

Algumas transformações envolvem a função seno ou o respectivo ângulo ou outras funções.

Transformação inverso do seno. É dada por:

$$y^{(\lambda)} = \begin{cases} \sqrt{n}\,\text{sen}^{-1}\sqrt{y + \alpha/n}, & \text{para} -\frac{\alpha}{n} \leq y \leq 1 - \frac{\alpha}{n} \\ 0, \text{caso contrário} \end{cases}$$

onde α é uma constante arbitrária. A transformação inverso do seno é usada para variáveis com distribuição binomial. Sua teoria pode ser vista em Beall (1942)[34]. Entretanto, seu uso vem diminuindo e seus méritos questionados, sendo substituída com algumas vantagens pela análise de regressão logística sobre os dados originais (OSBORNE, 2010; WILSON et al., 2010).

Transformação angular (ou arco seno). É dada por:

$$y^{(\lambda)} = \text{arc sen}\sqrt{y}$$

É usada para proporções (ou, equivalentemente, percentagens), que são transformadas num ângulo. Foi usada por Fischer (1922), no contexto de Genética e por Zubin (1935).

Transformação seno hiperbólico inverso. É dada por:

$$y^{(\lambda)} = \text{senh}^{-1}(y) = \text{arcsenh}(y) =$$
$$= \frac{1}{\lambda}\log\left(\lambda y + \sqrt{\lambda^2 y^2 + 1}\right), \qquad \text{para } \lambda > 0$$

Essa transformação foi proposta por Johnson (1949, apud BURBIDGE; MAGEE; ROBB, 1988), enquanto uma generalização dessa transformação foi proposta por Burbidge, Magee e Robb (1988).

Transformação tangente hiperbólica inversa. É dada por:

$$y^{(\lambda)} = \tanh^{-1}(y) = \text{arctanh}(y) = \frac{1}{2}\log\left(\frac{1+y}{1-y}\right)$$

Essa transformação foi usada por Taylor (1984).

5.3 - Outras Transformações

Transformação exponencial. É dada por:

$$y^{(\lambda)} = e^{y}$$

É usada para dados assimétricos à esquerda.

5.4 - Efeitos da Transformação

Depois de usar uma transformação sobre os dados para garantir normalidade, aplicam-se os procedimentos de análise estatística sobre os dados transformados. Todavia, em alguns casos é preciso fazer a transformação inversa para fornecer resultados sobre a variável original, por exemplo, intervalos de confiança. Então, alguns problemas podem aparecer. Por exemplo, ao se ajustar um modelo linear a uma variável de resposta transformada, as predições obtidas devem passar pela transformação inversa para serem expressas nas unidades originais de observação. Ocorre, segundo Perry e Walker (2015) que essas predições da variável de resposta original contêm um viés, teoricamente[35] originado do fato de que $E(Y^k)$ é uma função não linear de μ e σ^2.

6 - CONSIDERAÇÕES FINAIS

A distribuição normal e a estimação por mínimos quadrados vêm sendo utilizados desde o início do século XIX, mas tornou-se quase onipresente em muitos procedimentos estatísticos desenvolvidos ao longo do século XX na forma de um dos pres-

[34] Também em Tippett (1934 apud BEALL, 1942).

[35] Ver Land (1974 apud PERRY; WALKER, 2015).

supostos básicos. Contudo, na prática encontraram-se muitas situações em que a não normalidade é a regra. Para detectar tais casos desenvolveram-se dezenas de testes para teste de normalidade das observações. Para resolver o problema há diversos caminhos, mas o mais frequentemente usado é o da transformação dos dados, a fim de aproximar sua distribuição amostral da distribuição normal. A transformação mais geral e uma das mais usadas é a transformação potência proposta por Box-Cox. Estudos e simulações têm mostrado que o pesquisador deve analisar cada problema e os respectivos dados antes de decidir como enfrentar o problema, bem como qual teste utilizar e qual transformação ou enfoque considerar para minimizar os efeitos da não normalidade.

LITERATURA CITADA

ADEFISOYE, J. O. **An assessment of the performances of several univariate tests for normality**. 2015. 70 p. Dissertation (Master's degree in Statistics) - Florida International University, Miami, 2015.

AHMAD, A. M. A. W.; NAING, N. N.; HUSSEIN, M. T. A. R. A modification of Box-Cox transformation. **Statistika**, Vol. 7, pp. 45-550, 2007.

AITCHISON, J. The statistical analysis of compositional data. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 44, pp. 139-177, 1982. (Series B - Methodological).

AKBILGIÇ, O.; HOWE, J. A. A novel normality test using an identity transformation of the Gaussian function. **European Journal of Pure and Applied Mathematics**, Vol. 4, pp. 448-454, 2011.

ANDERSON, T. W.; DARLING, D. A. A test of goodness of fit. **Journal of the American Statistical Association**, Vol. 49, pp. 765-769, 1954.

ANDREWS, D. F. A note on the selection of data transformations. **Biometrika**, Vol. 58, pp. 249-254, 1971.

______.; MALLOWS, C. L. Scale mixtures of normal distributions. **Journal of the Royal Statistical Society**, Vol. 36, pp. 99-102, 1974. (Series B - Methodological).

ARNOLD, B. C.; MEEDEN, G. Characterization of distributions by sets of moments of order statistics. **The Annals of Statistics**, Vol. 3, pp. 754-758, 1975.

ARNOLD, T. B.; EMERSON, J. W. Nonparametric goodness-of-fit tests for discrete null distributions. **The R Journal**, Vol. 3, pp. 34-39, 2011.

ASAR, Ö.; ÍLK, Ö.; DAĞ, O. **Estimating Box-Cox power transformation via goodness of fit tests**. Accepted to be published in Communications in Statistics/ Simulation and Computation. 2014. 17 p.

ATKINSON, A. C. Testing transformations to normality. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 35, pp. 473-479, 1973.

BARTLETT, M. S.; KENDALL, D. G. The statistical analysis of variance: heterogeneity and the logarithmic transformation. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 7, Issue 1, pp. 128-138, 1946.

______. The square root transformation in analysis of variance. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 3, Issue 1, pp. 68-78, 1936.

______. The use of transformations. **Biometrics**, Vol. 3, pp. 39-52, 1947.

BEALL, G. The transformation of data from entomological field experiments so that the analysis of variance become applicable. **Biometrika**, Vol. 32, pp. 243-262, 1942.

BERNIER, J.; FENG, Y.; ASAKAWA, K. Strategies for handling normality assumptions in multi-level modeling: a case study estimating trajectories of health utilities index mark 3 scores. **Health Report**, Ottawa, Vol. 22, pp. 45-51, 2011.

BLAND, J. M.; ALTMAN, D. G. The use of transformation when comparing two means. **The British Medical Journal**, Vol. 312, pp. 1152, 1996.

BOGDAN, M. Data driven smooth tests for bivariate normality. **Journal of Multivariate Analysis**, Vol. 68, pp. 26-53, 1999.

BOWMAN, K. O.; SHENTON, L. R. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. **Biometrika**, Vol. 62, pp. 243-250, 1975.

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 26, Issue 2, pp. 211-252, 1964. (Series B - Methodological).

______; WATSON, G. S. Robustness to non-normality of regression tests. **Biometrika**, Vol. 49, pp. 93-106, 1962.

BOZDOGAN, H.; RAMIREZ, D. E. Testing for model fit: assessing and Box-Cox transformations of multivariate data to “near” normality. **Computational Statistics Quarterly**, Vol. 3, pp. 127-150, 1986.

BRADLEY, R. A. Corrections for nonnormality in the use of the two-sample t- and F- tests at high significance levels. **The Annals of Mathematical Statistics**, Durham, Vol. 23, pp. 103-113, 1952.

BURBIDGE, J. B.; MAGEE, L.; ROBB, A. L. Alternative transformation to handle extreme values of the dependet variable. **Journal of the American Statistical Association**, Washington, Vol. 83, pp. 123-127, 1988.

CARLSON, J. L. A study of the distribution of means estimated from small samples by the method of maximum likelihood for Pearson's type II curve. **The Annals of Mathematical Statistics**, Durham, Vol. 3, pp. 86-107, 1932.

CHEN, G.; LOCKHART, R.; STEPHENS, M. A. **EDF tests for normality in linear models after a Box-Cox transformation**. Stanford: Department of Statistics, University of Stanford, 1993. (Technical Report, 472).

CHESIRE, L.; OLDIS, E.; PEARSON, E. S. Further experiments on the sampling distribution of the correlation coefficient. **Journal of the American Statistical Association**, Washington, Vol. 27, pp. 121-128, 1932.

CRAIG, C. C. On the frequency function of *xy*. **The Annals of Mathematical Statistics**, Durham, Vol. 7, pp. 1-15, 1936.

CSÖRGŐ, S. Testing for normality in arbitrary dimension. **The Annals of Statistics**, Philadelphia, Vol. 14, pp. 708-723, 1986.

CURTISS, J. H. On transformations used in the analysis of variance. **The Annals of Mathematical Statistics**, Durham, Vol. 14, pp. 107-122, 1940.

D'AGOSTINO, R. An omnibus test of normality for moderate and large size samples. **Biometrika**, Vol. 58, pp. 341-348, 1971.

______.; STEPHENS, M. Goodness-of-fit techniques. New York: Marcel Dekker, 1986.

DAVIES, O. L. **The design and analysis of industrial experiments**. London: Oliver and Boyd, 1960.

DESMOULINS-LEBEAULT, F. **Semi-moments based tests of normality and the evolution of stock returns towards normality**. Paris: Association Française de Finance, 2004. Disponível em: <http://www.affi.asso.fr/uploads/Externe/36/CTR_FICHIER_108_1226315203.pdf >. Acesso em: set. 2015.

DOORNIK, J. A.; HANSEN, H. **An omnibus test for univariate and multivariate normality**. Oxford: Economics Group/Nuffield College/University of Oxford, 1994. (Economic Papers, W4e91).

DRAPER, N. R.; COX, D. R. On distributions and their transformation to normality. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 31, pp. 472-476, 1969. (Series B - Methodological).

DUFOUR, J. M. et al. Simulation-based finite-sample normality tests in linear regressions. **Econometrics Journal**, Vol. 1, pp. 154-173, 1998.

DUPAČOVÁ, J.; WETS, R. J.-B. **Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems**. Laxenburg: International Institute for Applied Systems Analysis, 1987. 20 p. (Working Paper, 9).

EDGELL, S. E.; NOON, S. M. Effect of violation of normality on the t test of the correlation coefficient. **Psychological Bulletin**, Vol. 95, pp. 576-583, 1984.

EFRON, B. Transformation theory: how normal is a family of distributions? **The Annals of Statistics**, Philadelphia, Vol. 10, pp. 323-339, 1982.

EISENHART, C.; WILSON, P. W. Statistical methods and control in bacteriology. **Bacteriological Reviews**, Vol. 7, pp. 57-137, 1943.

FARRELL, P. J.; SALIBIAN-BARRERA, M.; NACZK, K. On tests for multivariate normality and associated simulation studies. **Journal of Statistical Computation and Simulation**, Vol. 77, pp. 1065-1080, 2007.

FILLIBEN, J. J. The probability plot correlation coefficient test for normality. **Technometrics**, Vol. 17, pp. 111-117, 1975.

FINUCAN, H. M. A note on kurtosis. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 26, Issue 1, pp. 111-112, 1964. (Series B-Methodological).

FISCHER, R. A. On the dominance ratio. **Proceedings of the Royal Statistical Society of Edinburgh**, Vol. 42, pp. 321-341, 1922.

FOUDJO, A. N. **Robust normality test and robust power transformation with application to state change detection in non normal processes**. 2013. 135 p. Dissertation (Doktor der Naturwissenschaften) - Fakultät Statistik, Technische Universität Dortmund, Dortmund, 2013.

HANUSZ, Z.; TARASIŃSKA, J. Simulation study for a test of multivariate normality based on Shapiro-Wilk statistic. **Colloquium Biometricum**, Lublin, Vol. 39, pp. 45-51, 2009.

HARTIGAN, J. A. Necessary and sufficient conditions for asymptotic joint normality of statistic and its subsample values. **The Annals of Statistics**, Philadelphia, Vol. 3, pp. 573-580, 1975.

HENZE, N.; WAGNER, T. A new approach to the BHEP tests for multivariate normality. **Journal of Multivariate Analysis**, Vol. 62, pp. 1-23, 1997.

HOEFFDING, W. A class of statistics with asymptotically normal distribution. **The Annals of Mathematical Statistics**, Durham, Vol. 19, pp. 293-325, 1948.

HOTELLING, H.; PABST, M. R. Rank correlation and tests of significance involving no assumption of normality. **The Annals of Mathematical Statistics**, Durham, Vol. 7, pp. 29-43, 1936.

ISOGAI, T. On a measure of multivariate skewness and a test for multivariate normality. **Annals of the Institute of Statistical Mathematics**, Tokyo, Vol. 34, pp. 531-541, 1982.

JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. **International Statistical Review**, Malden, Vol. 55, pp. 163-172, 1987.

JOHNSON, N. L.; LEONE, F. C. **Statistics and experimental design in engineering and the physical sciences**. New York: Wiley, 1964. v. 2.

JUSTEL, A.; PEÑA, D.; ZAMAR, R. **A multivariate Kolmogorov-Smirnov test of goodness of fit**. Madrid: Departamento de Estadística/UC3M, 1994. 15 p. (Working Paper, 94-32).

KEENE, O. N. The log transformation is special. **Statistics in Medicine**, Vol. 14, pp. 811-819, 1995.

KRONMAL, R. A. Spurious correlation and the fallacy of the ratio standard revisited. **Journal of the Royal Statistical Society**, Edinburgh, Vol. 156, pp. 379-392, 1993. (Series A-Statistics in Society).

LACHENBRUCH, P. A. Proper metrics for clinical trials: transformations and other procedures to remove non-normality effects. **Statistics in Medicine**, Vol. 22, pp. 3823-3842, 2003.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. **Journal of the American Statistical Association**, Washington, Vol. 62, pp. 399-402, 1967.

MCALISTER, D. The law of the geometric mean. **Proceedings of the Royal Society of London**, London, Vol. 29, pp. 367-376, 1879.

MECKLIN, C. J.; MUNDFROM, D. J. An appraisal and bibliography of tests for multivariate normality. **International Statistical Review**, Vol. 72, pp. 123-138, 2004.

MENDES, M.; PALA, A. Type I error rate and power of three normality tests. **Pakistan Journal of Information and Technology**, Vol. 2, pp. 135-139, 2003.

Mood, A. M.; Graybill, F. A.; Boes, D. C. **Introduction to the theory of statistics**. Tokio: Mc-Graw Hill Kogakusha, 1974.

NAKAGAWA, S.; HASHIGUCHI, H.; NIKI, N**. A measure of skewness for testing departures from normality**. Preprint submitted to Computational Statistics and Data Analysis, Tokyo, 2012. Disponível em: <http://arxiv.org/pdf/1202.5093v1.pdf>. Acesso em: set. 2015.

NELSON JUNIOR, H. L.; GRANGER, C. W. J. Experience with using the Box-Cox transformation when forecasting economic time series. **Journal of Econometrics**, Vol. 10, pp. 57-69, 1979.

OSBORNE, J. W. Improving your data transformations: applying the Box-Cox transformation. **Practical Assessment, Research and Evaluation**, Vol. 15, pp. 1-9, 2010.

PERRY, M. B.; WALKER, M. L. A prediction interval estimator for the original response when using Box-Cox transformations. **Journal of Quality Technology**, Vol. 47, pp. 276-295, 2015.

POIRIER, D. J. The use of the Box-Cox transformation in limited dependent variable models. **Journal of the American Statistical Association**, Vol. 73, pp. 284-287, 1978.

POITRAS, G. More on the correct use of omnibus tests for normality. **Economic Letters**, Vol. 90, pp. 304-309, 2006.

RAO, C. R.; ALI, H. An overall test for multivariate normality. **Student**, Vol. 2, pp. 317-324, 1998.

RAZALI, N. M.; WAH. Y. B. Power comparisons of some selected normality tests. In: REGIONAL CONFERENCE ON STATISTICAL SCIENCES, 10., 2010, Kelantan. **Proceedings**… Kelantan: Malaysia Institute of Statistics/Faculty of Computer and Mathematical Sciences/UiTM, June 2010, p. 126-138.

REISS, R.-D. On the accuracy of the normal approximations for quantiles. **Annals of Probability**, Vol. 2, pp. 741-744, 1974.

RIANI, M. Robust multivariate transformations to normality: constructed variables and likelihood ratio tests. **Statistical Methods and Applications**, Vol. 13, pp. 179-196, 2004.

RIDER, P. R. On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes. **Biometrika**, Vol. 21, pp. 124-143, 1929.

RIETZ, H. L. Note on the distribution of the standard deviation of sets of three variates drawn at random from a rectangular distribution. **Biometrika**, Vol. 23, Issue 3/4, pp. 424-426, 1931.

RUBIN, D. B. Bayesianly justifiable and relevant frequency calculations for the applied statistician. **The Annals of Statistics**, Philadelphia, Vol. 12, pp. 1151-1172, 1984.

SAKIA, R. M. The Box-Cox transformation technique: a review. **The Statistician**, Vol. 41, pp. 169-178, 1992.

SAS. **BASE SAS® 9.2 procedures guide**: statistical procedures. 3. ed. Cary, NC: SAS Institute Inc., 2010.

______. **SAS/STAT® 9.2 user's guide**: the Transreg procedure. 2. ed. Cary, NC: SAS Institute Inc., 2008.

SEIER, E. **Comparison of tests for univariate normality**. Tennessee: ETSU, 2002. 17 p. Disponível em: <http://interstat.statjournals.net/YEAR/2002/abstracts/0201001.php>. Acesso em: set. 2015.

SHAPIRO, S. S.; FRANCIA, R. S. An approximate analysis of variance for normality. **Journal of the American Statistical Association**, Vol. 67, pp. 215-216, 1972.

______.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Vol. 52, pp. 591-611, 1965.

SHENTON, L. R. **Transforming non-normal distributions into nearly normal distributions**. Athens: University of Georgia, College of Agriculture, 1965. (Technical Bulletin N.S., 49 and 50).

SHEWHART, W. A.; WINTERS, F. W. Small samples: new experimental results. **Journal of the American Statistical Association**, Vol. 23, pp. 144-153, 1928.

SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. **The Annals of Mathematical Statistics**, Durham, Vol. 19, pp. 279-281, 1948.

SPROT, D. A. Normal likelihoods and their relation to large sample theory estimation. **Biometrika**, Vol. 60, Issue 3, pp. 457-465, 1973.

SRIVASTAVA, M. S.; HUI, T. K. On assessing multivariate normality based on Shapiro-Wilk W statistic. **Statistics and Probability Letters**, Vol. 5, pp. 15-18, 1987.

SU, Y.; KANG, S. -Y. Testing for multivariate normality of disturbances in the multivariate linear regression model. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS RESEARCH AND MECHATRONICS ENGINEERING, 2015, Zhengzhou. **Proceedings**… Zhengzhou: Atlantis Press, 2015. p. 420-425.

TAYLOR, C. R. A flexible method for empirically estimating probability functions. **Western Journal of Agricultural Economics**, Vol. 9, pp. 66-76, 1984.

URZÚA, C. M. **Portable and powerful tests for normality**. México: Tecnológico de Monterrey, Campus Ciudad de México, 2007. 9 p.

WARE, R.; LAD, F. **Approximating the distribution of sums of products of normal variables**. Christchurch: University of Canterbury, 2003. Disponível em: <http://www.math.canterbury.ac.nz/research/ucdms2003n15.pdf>. Acesso em: set. 2015.

WEISBERG, S. **Yeo-Johnson power transformations**. Minneapolis: Department of Applied Statistics/University of Minnesota, 2001. Disponível em: <https://www.stat.umn.edu/arc/yjpower.pdf>. Acesso em: jun. 2015.

WEST, K. D. Asymptotic normality, when regressors have a unit root. **Econometrica**, New York, Vol. 56, pp. 1397-1417, 1988.

WILSON, E. et al. **The arcsine transformation**: has the time come for retirement? Canada: Memorial University of Newfoundland, 2010. 188 p. Disponível em: <http://www.mun.ca/biology/dschneider/b7932/B7932Final10Dec2010.pdf>. Acesso em: set. 2015.

YEO, I.-K.; JOHNSON, R. A.; DENG. X. W. An empirical characteristic function approach to selecting a transformation to normality. **Communications for Statistical Applications and Methods**, Vol. 21, pp. 213-224, 2014.

ZUBIN, J. Note on a transformation function for proportions and percentages. **Journal of Applied Psychology**, Washington, Vol. 19, pp. 213-220, 1935.

Recebido 30/11/2015. Liberado pra publicação em 16/02/2016.