



NETO PASCOAL

**TESTE BOOTSTRAP DE NORMALIDADE UNIVARIADA
BASEADO NA ENTROPIA**

LAVRAS – MG

2018

NETO PASCOAL

**TESTE BOOTSTRAP DE NORMALIDADE UNIVARIADA BASEADO NA
ENTROPIA**

Dissertação apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Prof. Doutor Daniel Furtado Ferreira
Orientador

LAVRAS – MG
2018

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Pascoal, Neto.

Teste Bootstrap de Normalidade Univariada Baseado na
Entropia / Neto Pascoal. - 2018.

70 p.

Orientador(a): Daniel Furtado Ferreira.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2018.

Bibliografia.

1. Bootstrap paramétrico. 2. Poder Erro tipo I. 3. Entropia e
teste de Shapiro-Wilk. I. Ferreira, Daniel Furtado. II. Título.

NETO PASCOAL

**TESTE BOOTSTRAP DE NORMALIDADE UNIVARIADA BASEADO NA
ENTROPIA**

Dissertação apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 15 de junho de 2018.

Prof. Doutor Deive Ciro de Oliveira UNIFAL
Prof. Doutor Paulo Henriques Sales Guimarães UFLA



Prof. Doutor Daniel Furtado Ferreira
Orientador

**LAVRAS – MG
2018**

"Aos meus pais, irmãos, noiva, filha, sobrinhos e a toda família que, com muito carinho e apoio, não mediram esforços para que eu chegasse a essa etapa da vida"

AGRADECIMENTOS

A Deus por ter me dado saúde e força para superar as dificuldades. Ao Departamento de Estatística da Universidade Federal de Lavras (UFLA), ao corpo docente, a secretária e ao Instituto de Bolsa de Estudo de Moçambique, que oportunizaram a janela que hoje vislumbro um horizonte superior, eivado pela acendrada confiança no mérito e ética aqui presentes;

Ao meu orientador Prof. Doutor Daniel Furtado Ferreira, pela maestria e suporte no alcance dessa caminhada acadêmica;

Aos meus pais Pascoal Alficha e Nhauta Polena João, pelos ensinamentos, incentivo, apoio incondicional e recomendações desde pequeno "filho estuda para se tornar alguém amanhã" hoje compreendo, que o estudo é o maior e único bem que é realmente meu;

A tia Ester e Losinha, pela contribuição na caminhada desde o dia do meu nascimento, uma espécie de segunda mãe, lado a lado;

A minha noiva Cacilda Isabel Francisco, pelo amor, carinho, atenção, companheirismo, compreensão, você foi e é geradora da energia que alimenta essa caminhada;

Aos meus irmão(a)s: Flora, Chihuca, Domingos, Gilda, Tiago, Fátima, Mano, Jacinto, Marcelina, Fernando, Minija, Samuel, Rúbene, Melisa, muito obrigado por toda alegria, apoio incondicional e estímulo que sempre injetaram em minha vida;

A minha filha Polena e sobrinho(a)s Helena, Eugênio, Osvaldo, José, Flora, Del, Salomão, Zena, Sendi, Graça, Polena, Amílton Acina, Milena e Pascoal

A toda família, amigos, colegas e a todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

RESUMO

O comportamento de muitos fenômenos, em muitas áreas do saber, é descrito pela distribuição normal de probabilidade. Ao se retirar uma amostra aleatória de uma população, no caso univariado, é comum pressupor que os dados ou resíduos de algum modelo sejam normalmente distribuídos. Essa pressuposição de normalidade deve ser verificada por meio da aplicação de testes estatísticos. A entropia pode ser entendida como uma medida da quantidade de aleatoriedade de um sistema de informação, sendo usada para medir a incerteza de uma variável aleatória. Ela tem sido também utilizada para testar a normalidade. Este trabalho teve como objetivo propor testes de normalidade univariada baseados na entropia. Estes testes são essencialmente procedimentos computacionalmente intensivos baseados na técnica *bootstrap* paramétrico. Além disso, comparou-se o poder do teste de normalidade Shapiro-Wilk (TW) com os testes de normalidade univariada propostos (TKB , $TKRB_1$, $TKRB_2$, $TKRB_C$). As alternativas computacionalmente intensivas, baseada em *bootstrap* paramétrico permitiram contornar as limitações de tamanhos amostrais n dos testes baseados em entropia existentes, que permitem no máximo $n = 100$ e também contornar o problema da não invariância de uma das opções. Os testes propostos apresentaram controle adequado das taxas de erro tipo I, se mostrando exatos, com tamanhos iguais ao nível de significância nominal α . Em relação ao desempenho sob as hipóteses alternativas, ao menos um dos testes TKB , $TKRB_1$, $TKRB_2$ e $TKRB_C$, esteve quase sempre como o mais poderoso. Assim, pode-se recomendá-los, pois estes superaram, de maneira geral, o teste sendo concorrente de melhor desempenho, escolhido para ser o teste de referência, o teste de Shapiro-Wilk (TW) mas, prevalecendo a ideia ainda da não existência de um teste uniformemente mais poderoso para todas as hipóteses alternativas estudadas. As propostas podem ser usadas em amostras superiores a 5.000, que é uma das suas principais virtudes em relação ao teste usado como referência.

Palavras-chave: *Bootstrap paramétrico; Poder; Erro tipo I; entropia; teste de Shapiro-Wilk.*

ABSTRACT

The behaviour of many phenomena, in many areas of the knowledge, is described by the normal distribution of probability. When a random sample of population is taken, in the univariate case, it is common to assume that the data or residues of certain model are normally distributed. This assumption of normality must be verified through application of a statistical test. The entropy can be understood as a measure of the amount of randomness of an information system, and used to measure the uncertainty of a random variable. It is also being used for testing normality. This study is aimed to propose univariate normality tests based on entropy. These tests are computationally intensive procedures based on parametric bootstrap technique. In addition, the power of the Shapiro-Wilk normality test (TW) was compared with the proposed univariate normality tests (TKB , $TKRB_1$, $TKRB_2$, $TKRB_C$). Computational intensive alternatives based on parametric bootstrap allowed to circumvent the limitations of sample sizes n of the existing entropy-based tests, which allow a maximum of $n = 100$ and also to overcome the problem of non-invariance of one of the options. The proposed tests showed adequate control of the type I error rates, if they were accurate, with sizes equal to the level of nominal significance α . Regarding performance under alternative hypotheses, at least one of the TKB , $TKRB_1$, $TKRB_2$, $TKRB_C$, tests demonstrated its supremacy. Thus, we can recommend them, since they generally exceeded the concurrent best-performance test, chosen to be the reference test, the Shapiro-Wilk (TW) test, although, still prevailing the idea of non-existence of a uniformly more powerful test for all the alternative hypotheses studied. Proposals may be used on samples over 5,000, which is one of its main virtues in relation to the test used as a reference.

Keywords: Parametric Bootstrap; Power; Type I Error; entropy; Shapiro-Wilk test.

LISTA DE FIGURAS

Figura 3.1 – Fluxograma da estrutura dos métodos.	38
---	----

LISTA DE TABELAS

Tabela 2.1 – Comparação de distribuição de qui-quadrado, Kolmogorov-Smirnov, Lilliefors e de Shapiro-Wilk	27
Tabela 2.2 – Representação tabular das decisões possíveis de um testes estatístico. . . .	28
Tabela 3.1 – Tamanhos amostrais utilizados nas simulações Monte Carlo e <i>bootstrap</i> . .	36
Tabela 4.1 – Erro tipo I dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , em função de $n=3(1)10, 15(5)50$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%	41
Tabela 4.2 – Erro tipo I dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , em função de $n = 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha= 10\%, 5\%$ e 1%	43
Tabela 4.3 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição exponencial ($\lambda = 1$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha= 10\%, 5\%$ e 1%	45
Tabela 4.4 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição gama ($\alpha = 1, \beta = 3$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha= 10\%, 5\%$ e 1%	47
Tabela 4.5 – Poder dos testes univariados de normalidade baseados na entropia TVB , TKR_1 , $TKRB_2$ e TW , relativo à distribuição t -Student com $\nu = 1$, em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha= 10\%, 5\%$ e 1%	48
Tabela 4.6 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição t -Student com $\nu = 30$, em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%	49
Tabela 4.7 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição uniforme ($a = 0, b = 1$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha= 10\%, 5\%$ e 1%	51

Tabela 4.8 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição log-normal ($\mu = 0, \sigma = 1$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%	52
Tabela 4.9 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição Weibull ($\delta = 2, \alpha = 3$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%	53
Tabela 4.10 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição beta ($\alpha = 0,2, \beta = 0,5$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%	54
Tabela 4.11 – Distribuições não-normais, com suas respectivas abreviaturas e parâmetros adotados	57
Tabela 4.12 – Erro tipo I de teste combinado $TKRB_C$ de normalidade univariada baseado na entropia em função de $n (3, 10, 30, 50, 100, 500, 1.000, 5.000)$, para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%	58
Tabela 4.13 – Poder do teste combinado $TKRB_C$ de normalidade, relativo à oito distribuições não normais univariadas, em função de $n=3, 10, 30, 50, 100, 500, 1.000, 5.000$, para o nível nominal de significância $\alpha=10\%$	59
Tabela 4.14 – Poder do teste combinado $TKRB_C$ de normalidade, relativo à oito distribuições não normais univariadas, em função de $n=3, 10, 30, 50, 100, 500, 1.000, 5.000$, para o nível nominal de significância $\alpha=5\%$	61
Tabela 4.15 – Poder do teste combinado $TKRB_C$ de normalidade, relativo à oito distribuições não normais univariadas, em função de $n=3, 10, 30, 50, 100, 500, 1.000, 5.000$, para o nível nominal de significância $\alpha=1\%$	62

SUMÁRIO

1	Introdução	10
2	Referencial Teórico	13
2.1	Entropia	13
2.1.1	Entropia relativa	14
2.2	Testes univariados de normalidade	19
2.3	Erro tipo I e poder de teste	26
2.4	Método Monte Carlo	29
2.5	Método <i>bootstrap</i>	29
3	Métodos	32
3.1	Métodos	32
3.2	Avaliação do desempenho	36
3.3	Avaliação do erro tipo I e poder do teste combinado	37
3.4	Teste binomial exato	39
4	Resultados e Discussão	40
4.1	Erro tipo I	40
4.2	Poder do teste	44
4.3	Desempenho do teste combinado	56
4.3.1	Resultados do erro tipo I do teste combinado	57
4.3.2	Resultados do poder do teste combinado ($TKRB_C$)	58
5	Conclusão	64
	REFERÊNCIAS	66

1 INTRODUÇÃO

Teste de normalidade univariada baseado na entropia é um campo de pesquisa com potencial importância na área de estatística. Os testes de normalidade são usados para verificar se a distribuição de probabilidade associada a um conjunto de dados ou resíduos do modelo podem ser aproximada a uma distribuição normal. Existem vários critérios ou técnicas amplamente discutidas na literatura, que são usadas para verificar a normalidade. Mesmo assim, tem crescido a tendência de novas propostas ou testes, visto que, na literatura, não se encontra um teste uniformemente poderoso em relação aos tamanhos amostrais e, principalmente, às distribuições.

Quase todo processo de análise de dados tem por finalidade realizar inferências, que consiste em se fazer generalizações sobre uma população com base nos dados amostrais. Ao se retirar uma amostra aleatória de uma população, no caso univariado, é comum muitas vezes pressupor-se que os dados ou resíduos do modelo adotado são proveniente de uma população normalmente distribuída. Isso é feito com a finalidade de garantir que as inferências sejam válidas, ou seja, que descrevam situações da vida real de forma adequada, obtendo-se estimativas relativamente próximas dos parâmetros da população e considerando suas margens de erros. A verificação da pressuposição de normalidade dos dados ou dos resíduos do modelo é preponderante, portanto. Fato este que motiva o uso e a proposição de critérios ou técnicas, para verificar a normalidade.

A entropia é uma medida da quantidade de desordem ou aleatoriedade do comportamento de um sistema de energia ou informação, usado para descrever a incerteza do movimento de partículas de um gás ou ainda, usada para medir a incerteza de uma variável aleatória. Como a entropia é uma propriedade da distribuição da variável aleatória, então se pode utilizar tal medida para se realizar testes de aderência. Durante as últimas décadas várias propostas de estimadores de entropia foram apresentados, sendo que alguns deles são acompanhados de testes de normalidade associados. Entretanto, alguns destes estimadores são não invariantes (os estimadores não mudam em transformação de locação e escala) e os testes requerem usos de tabelas específicas e limitadas em relação aos tamanhos amostrais, que, em geral, são menores que 100.

A entropia é uma medida ou propriedade da distribuição da variável aleatória, que pode ser utilizada para realizar testes de aderência e atua no estado de quase todos os processos do dia à dia. É aplicado como um instrumento fundamental na estatística inferencial. A adoção da

entropia como medida de dispersão global, informação e incerteza na análise dos dados torna-se ferramenta de crescente importância.

Quase todo processo de análise de dados tem por finalidade realizar inferências, e é comum, muitas vezes, pressupor que os dados (ou resíduos) são provenientes de uma população normalmente distribuída. Surge, portanto, a necessidade de verificar se essa suposição de normalidade não foi violada. Por outro lado, na literatura, não se encontra um teste que seja uniformemente mais poderoso em relação aos tamanhos amostrais e distribuições alternativas. Motivando assim que sejam propostos novos testes de normalidade, buscando-se entre eles, aqueles mais poderosos para as mais amplas situações sob hipóteses alternativas de não normalidade e de tamanhos amostrais.

O comportamento de muitos fenômenos da vida real em muitas áreas do saber, são descritos, pelo menos aproximadamente, pelas distribuições normais de probabilidade. Por sua vez, muitas técnicas estatísticas utilizadas para estudar estes fenômenos são baseadas no pressuposto de que os dados ou resíduos são provenientes dessa distribuição. Como consequência do atendimento à esta suposição, os resultados da inferência, serão confiáveis, o que ressalta a necessidade de se verificar se a pressuposição de normalidade não foi violada.

Kohansal e Rezakhah (2016) apresentam um estudo de poder dos testes de normalidade baseado nos estimadores de entropia e concluem que eles são mais poderosos que todas as alternativas comparadas. Entretanto, os valores de n apresentados, são muito limitados ($n \leq 100$), além da inexistência, que se tenha informação, de funções para aplicação destes testes implementadas em *softwares* livres ou proprietários. Ademais, uma das estatísticas de seus testes é invariante a transformações de localização e escala.

Assim, a proposição de testes, que visem superar essas limitações em relação aos valores de n , e cujo desempenho venha a ser avaliado quando estes forem aplicados em uma ampla variedade de distribuições não normais sob a hipótese alternativa, além de serem baseados em procedimentos computacionalmente intensivos via *bootstrap* paramétrico, permitiria que se contorne as limitações da invariância das transformações de localização e escala de uma das alternativas e também poderiam ser implementados em pacotes do programa R.

Desta forma, este trabalho teve por objetivo a proposição de um teste de normalidade univariada baseado na entropia e fundamentado em um procedimento computacionalmente intensivo baseado na técnica de *bootstrap* paramétrico. Espera-se que os testes propostos sejam mais poderosos para uma ampla variedade de distribuições não normais sob a hipótese alter-

nativa. Além disso, alguns objetivos específicos do presente trabalho foram considerados, os quais são:

- i.** Apresentar a teoria e os métodos utilizados nos testes de normalidade univariada, baseados na entropia;
- ii.** Melhorar os testes de normalidade baseados em estimadores de entropia para se obter testes mais poderosos;
- iii.** Apresentar uma alternativa computacional intensiva, baseada em *bootstrap* paramétrico na obtenção dos valores críticos usadas na decisão de rejeição ou não da hipótese de normalidade univariada;
- iv.** Comparar o desempenho (taxas de erro tipo I e poder) do teste de normalidade Shapiro-Wilk com os testes de normalidade univariada propostos.

Na sequência, o trabalho está dividido em cinco capítulos dos quais: o capítulo I foi dedicado a aspectos introdutórios como objetivos, motivação e problematização. O Capítulo II considera as ideias básicas das teorias e métodos dos estimadores de entropia e estatísticas de testes baseados por esses estimadores, ainda nesse capítulo apresenta os testes de normalidade, erro tipo I e poder de testes, método Monte Carlo e método *bootstrap*. O Capítulo III é dedicado aos métodos, dentre as quais destacam-se a avaliação do desempenho (erro tipo I e poder de teste), e teste binomial exato. No capítulo IV, apresentam-se os resultados e discussão, com destaque ao erro tipo I e poder dos testes propostos (TVB , $TKRB_1$, $TKRB_2$), por fim, no capítulo V, apresenta a conclusão e as considerações finais.

2 REFERENCIAL TEÓRICO

Neste capítulo são apresentados os conceitos e teorias da entropia e testes de normalidade univariados, mostrando as ideias contidas nos testes baseados em estimadores de entropia e os seus procedimentos teóricos. Também, retrata as definições de erro e o poder de teste, métodos *bootstrap* e Monte Carlo.

2.1 Entropia

A entropia é uma medida da quantidade de desordem ou aleatoriedade do comportamento de um sistema de energia ou informação (SHANNON, 1948; VASICEK, 1976; KOHANSAL; REZAKHAH, 2016). Este conceito é emprestado da termodinâmica, em que é usado para descrever a incerteza do movimento de partículas de um gás. Na estatística, a entropia é usada para medir a incerteza de uma variável aleatória. Sua definição está intrinsicamente ligada ao conceito de esperança matemática e é apresentada a seguir.

Considere-se uma variável aleatória qualquer com função de probabilidade p_X , caso seja discreta, ou função densidade f_X , caso seja contínua, a entropia é definida por

$$H(X) = \begin{cases} E[-\ln(p_X(X))] = - \sum_{x \in S_X} \ln(p_X(x))p_X(x) & \text{(discreto)} \\ E[-\ln(f_X(X))] = - \int_{x \in S_X} \ln(f_X(x))f_X(x)dx & \text{(contínuo)}, \end{cases} \quad (2.1)$$

em que a soma ou a integração são tomadas em relação ao conjunto S_X que é o conjunto suporte da variável aleatória X .

No caso discreto a entropia é sempre positiva, mas no caso contínuo ela pode ser zero e até mesmo negativa. Para contornar esse problema do ponto de vista do conceito físico, muitas vezes é usado o conceito de entropia relativa. Neste caso a entropia é definida como uma relação entre duas distribuições distintas ou diferentes, tanto no caso discreto quanto no caso contínuo, sendo conhecida como entropia relativa de Kullback-Leibler. Para a distribuição normal, tem-se o seguinte resultado:

Suponha que a variável aleatória $X \sim N(\mu, \sigma^2)$, então a entropia X é:

$$\begin{aligned}
 H(X) &= E[-\ln(f_X(X))] = - \int_{x \in S_X} \ln(f_X(x)) f_X(x) dx \\
 &= - \int_{-\infty}^{\infty} f_X(x) \left[-\frac{\ln 2\pi}{2} - \ln(\sigma) - \frac{(x-\mu)^2}{2\sigma^2} \right] dx \\
 &= \int_{-\infty}^{\infty} \frac{\ln 2\pi}{2} f_X(x) dx + \int_{-\infty}^{\infty} \ln(\sigma) f_X(x) dx + \\
 &\quad + \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma^2} f_X(x) dx \\
 &= \frac{\ln(2\pi)}{2} + \frac{\ln(\sigma^2)}{2} + \frac{(\sigma)^2}{2\sigma^2} \\
 &= \ln(\sigma\sqrt{2\pi e}). \tag{2.2}
 \end{aligned}$$

Assim, se σ tende para zero, $H(X)$ tende para $-\infty$ e se σ tende para ∞ , $H(X)$ tende para ∞ . Em outras palavras, quanto maior a entropia, maior é a variabilidade, ou maior é a desordem do sistema. No caso particular da distribuição normal padrão, a entropia é 1,4189.

2.1.1 Entropia relativa

A entropia relativa é sempre tomada considerando duas funções de densidades ou funções de probabilidade da mesma distribuição. Uma medida da entropia relativa muito utilizada é a da Kullback-Leibler é definida por

$$H_{KL}(X) = \begin{cases} E_{p_X} \left[\ln \left(\frac{p_X(X)}{q_X(X)} \right) \right] & \text{(discreto)} \\ E_{f_X} \left[\ln \left(\frac{f_X(X)}{g_X(X)} \right) \right] & \text{(contínua),} \end{cases} \tag{2.3}$$

sendo que q_X e g_X funções de probabilidade e densidade, da mesma família de distribuição de p_X e f_X , nos casos discreto e contínuo, respectivamente.

Como anteriormente, considera-se o caso da normal e em seguida é obtida a entropia relativa à um outro modelo normal específico. Suponha-se que $X \sim N(\mu_1, \sigma_1^2)$, com função densidade $f_X(x; \mu_1, \sigma_1^2)$, então a entropia relativa da distribuição de X relativa a uma variável aleatória também normal com média μ_2 , variância σ_2^2 e função densidade $g_X(x; \mu_2, \sigma_2^2)$ será de forma:

A entropia relativa de uma variável aleatória normal X é:

$$\begin{aligned} H_{KL}(X) &= E_{f_X} \left[\ln \left(\frac{f_X(X; \mu_1, \sigma_1^2)}{g_X(X; \mu_2, \sigma_2^2)} \right) \right] \\ &= E_{f_X} [(\ln f_X(X; \mu_1, \sigma_1^2)) - (\ln g_X(X; \mu_2, \sigma_2^2))] \\ &= E_{f_X} [(\ln f_X(X; \mu_1, \sigma_1^2))] - E_{f_X} [(\ln g_X(X; \mu_2, \sigma_2^2))]. \end{aligned}$$

Tem-se neste caso que

$$\ln \left(\frac{f_X(X; \mu_1, \sigma_1^2)}{g_X(X; \mu_2, \sigma_2^2)} \right) = \ln \frac{\sigma_2}{\sigma_1} + \frac{(x - \mu_1)^2}{2\sigma_1^2} - \frac{(x - \mu_2)^2}{2\sigma_2^2},$$

portanto,

$$\begin{aligned} H_{KL}(X) &= \int_{-\infty}^{\infty} \ln \left(\frac{f_X(x; \mu_1, \sigma_1^2)}{g_X(x; \mu_2, \sigma_2^2)} \right) f_X(x; \mu_1, \sigma_1) dx \\ &= \int_{-\infty}^{\infty} \left[\ln \frac{\sigma_2}{\sigma_1} - \frac{(x - \mu_1)^2}{2\sigma_1^2} + \frac{(x - \mu_2)^2}{2\sigma_2^2} \right] f_X(x; \mu_1, \sigma_1) dx \\ &= \ln \left(\frac{\sigma_2}{\sigma_1} \right) - \frac{\sigma_1^2}{2\sigma_1^2} + \int_{-\infty}^{\infty} \frac{(x - \mu_2)^2}{2\sigma_2^2} f_X(x; \mu_1, \sigma_1) dx \\ &= \frac{1}{2} \ln \left(\frac{\sigma_2^2}{\sigma_1^2} \right) - \frac{1}{2} + \int_{-\infty}^{\infty} \frac{(x - \mu_1 + \mu_1 - \mu_2)^2}{2\sigma_2^2} f_X(x; \mu_1, \sigma_1) dx \\ &= \frac{1}{2} \ln \left(\frac{\sigma_2^2}{\sigma_1^2} \right) - \frac{1}{2} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} \\ &= \left[\frac{\sigma_1^2}{\sigma_2^2} - \ln \left(\frac{\sigma_1^2}{\sigma_2^2} \right) - 1 \right] + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}. \end{aligned}$$

Considerando-se a desigualdade $\ln(x) \leq x - 1$ na expressão final de $H_{(KL)}(X)$, tem-se

$$H_{(KL)}(X) \geq \left(\frac{\sigma_1^2}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^2} + 1 - 1 \right) + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2} \geq \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}.$$

Logo, a entropia da distribuição de X , relativa a uma outra distribuição normal com média μ_2 e a variância σ_2^2 , $H_{(KL)}(X)$ é não negativa e maior ou igual a metade de distância estatística entre as médias das duas distribuições. Se as variâncias são iguais, então $H_{(KL)}(X)$ é exatamente metade da distância estatística entre as duas médias populacionais.

Vasicek (1976) mostrou que a expressão 2.1 para o caso contínuo pode ser expressa por:

$$H(X) = - \int_{-\infty}^{\infty} \ln(f_X(x)) f_X(x) dx = \int_0^1 \ln \left(\frac{d}{dp} F_X^{-1}(p) \right) dp, \quad (2.4)$$

em que $F_X^{-1}(p)$ é inversa da função de distribuição de probabilidade de X , avaliada em p , que é a probabilidade acumulada. Para provar esta identidade, considere que

$$H(X) = - \int_{-\infty}^{\infty} f_X(x) dt = F_X(t).$$

Assim,

$$F_X^{-1}(p) = F_X^{-1}(F_X(x)) = x.$$

Portanto,

$$\frac{d}{dx} F_X^{-1}(F_X(x)) = \frac{d}{dx} x = 1$$

$$\frac{d}{dp} F_X^{-1}(p) \frac{d}{dx} F_X(x) = 1$$

$$\frac{d}{dp} F_X^{-1}(p) f_X(x) = 1$$

$$\frac{d}{dp} F_X^{-1}(p) = \frac{1}{f_X(x)}.$$

Como,

$$\frac{dp}{dx} = \frac{d}{dx} F_X(x) = f_X(x),$$

então

$$dp = f_X(x) dx,$$

Logo,

$$\begin{aligned} H(X) &= - \int_{-\infty}^{\infty} \ln(f_X(x)) f_X(x) dx = \int_0^1 \ln \left[\frac{1}{dF_X^{-1}(p)/dp} \right] dp \\ &= \int_0^1 \ln \left(\frac{d}{dp} F_X^{-1}(p) \right) dp, \end{aligned}$$

como se queria mostrar.

Vasicek (1976) apresentou um estimador de 2.4, aproveitando-se essa forma de representação, substituindo o operador diferencial pelo operador de diferença. Assim, o estimador da derivada de $F_X^{-1}(p)$ é $(X_{(i+m)} - X_{(i-m)})n/(2m)$, para $(i-1)n < p \leq i/n$, $i = m+1, m+2, \dots, n-m$, em que $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$ são as estatísticas de ordem e m é inteiro positivo menor que $n/2$. Diferenças como $X_{(i+m)} - X_{(1)}$ e $X_{(n)} - X_{(i-m)}$ devem ser usadas no lugar de $X_{(i+m)} - X_{(i-m)}$ quando $p \leq m/n$ e $p > (n-m)/n$, respectivamente (VASICEK, 1976). Assim o estimador não paramétrico de Vasicek é definido por:

$$HV_{mn} = n^{-1} \sum_{i=1}^n \ln \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}, \quad (2.5)$$

em que $X_{(i)} = X_{(1)}$ se $i < 1$ $X_{(i)} = X_{(n)}$, se $i > n$.

Park e Park (2003) derivam a função de distribuição não paramétrica de HV_{mn} por

$$g_v(x) = \begin{cases} 0, & x < \xi_1 \quad \text{ou} \quad x > \xi_{n+1} \\ n^{-1} \frac{2m}{x_{(i+m)} - x_{(i-m)}}, & \xi_i < x \leq \xi_{i+1}, \quad i = 1, 2, \dots, n, \end{cases} \quad (2.6)$$

em que $\xi_i = (x_{(i-m)} + x_{(i-m+1)} + \dots + x_{(i+1)})/(2m)$ e $x_{(i)} = x_{(1)}$ se $i < 1$ e $x_{(i)} = x_{(n)}$ se $i > n$.

Vasicek (1976) propôs um teste de normalidade baseado neste estimador. Como foi mostrado na equação 2.2, o máximo da entropia para o modelo normal com variância σ^2 é $\ln(\sigma\sqrt{2\pi e})$. Assim, para qualquer modelo probabilístico não normal f_X , $\exp(H(X))/\sigma \leq \sqrt{2\pi e}$. Portanto, a hipótese nula de normalidade univariada será rejeitada se $K_{mn} \leq K_\alpha^*$, em que K_α^* é um valor crítico apropriado para o teste e

$$K_{mn} = \frac{n}{2m\hat{\sigma}} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n}, \quad (2.7)$$

é a estatística de teste, sendo $\hat{\sigma}$ o estimador de máxima verossimilhança de σ , para o qual tem-se

$$\hat{\sigma} = n^{-1} \sum_{j=1}^n (X_j - \bar{X})^2.$$

Os valores críticos K_α^* (quantis superiores 100 $\alpha\%$) não foram avaliados analiticamente, uma vez que Vasicek (1976) não derivou a distribuição nula K_{mn} . O autor sugeriu que se utilizasse simulação Monte Carlo para isso. Eles apresentam alguns valores críticos considerando

diversos tamanhos amostrais de n e de m , fixando-se $\alpha = 0,05$. Não há soluções quando as estatísticas de ordem apresentam empates. O autor percebeu que a partir de simulações, o valor de poder se alterava com os diferentes valores de n e m , ou seja, quanto maior era n maior deve ser o valor de m para se alcançar um maior poder do teste. Nos casos simulados para $n = 10, 20, e 50$, Vasicek (1976) recomendou $m = 2, 3$ e 4 , respectivamente. A melhor escolha de m cresce com o aumento de n , mais com a razão m/n decrescendo para zero (0). Crzcgorzewski e Wirczorkowski (1999), mencionam que a melhor escolha de m é dada por $m = \lfloor \sqrt{n+0,5} \rfloor$, em que $\lfloor x \rfloor$ é o maior inteiro não menor que x . Embora utilizem esta recomendação na estimação de entropia, na aplicação de teste de normalidade Crzcgorzewski e Wirczorkowski (1999) recomendam os valores: $m = 1$, se $n \leq 8$; $m = 2$ se $9 \leq n \leq 15$; $m = 3$ se $16 \leq n \leq 35$; $m = 4$ se $36 \leq n \leq 60$; $m = 5$ se $61 \leq n \leq 80$; $m = 6$ se $81 \leq n \leq 100$.

Existem vários outros estimadores melhores do que o estimador de Vasicek (1976), que são na verdade variações ou ajustes do mesmo. Recentemente, Kohansal e Rezakhah (2016) propuseram dois novos estimadores baseados em médias móveis das estatísticas de ordem, que levaram a testes mais poderosos. A ideia é que os estimadores de $F^{-1}(p)$, baseados em estatísticas de ordem, não são suavizados. Portanto, para uma janela de tamanho $w < n$, são determinadas as estatísticas de ordem suavizadas $Y_1, Y_2, Y_3, \dots, Y_n$, por

$$Y_i = \begin{cases} \frac{\sum_{j=1}^i X_j}{i}, & 1 \leq i \leq (w-1)/2 \\ \frac{\sum_{j=i-(w-1)/2}^i X_j}{w}, & (w+1)/2 \leq i \leq n - (w-1)/2 \\ \frac{\sum_{j=i}^n X_j}{n-i+1}, & n - (w-3)/2 \leq i \leq n, \end{cases} \quad (2.8)$$

para $i = 1, 2, 3, \dots, n$ em que $X_{(j)}$'s são estatísticas de ordem amostrais, com $j = 1, 2, 3, \dots, n$.

Os estimadores da entropia modificados, a partir dos apresentados pelos autores são:

$$HKR_{mn}^1 = \frac{1}{n} \sum_{i=1}^n \ln(Y_{i+m-1} - Y_{i-m-1}) + c \quad (2.9)$$

$$HKR_{mn}^2 = \frac{1}{n-m} \sum_{i=2}^{n-m} \ln(Y_{i+m-1} - Y_{i-1}) + \sum_{k=m}^n \frac{1}{k}, \quad (2.10)$$

em que $Y_i = Y_1$, se $i < 1$, $Y_i = Y_n$, para $i > n$ e

$$c = - \left(1 - \frac{2m}{n} \right) \psi(2m) + \psi(n+1) - \frac{2}{n} \sum_{i=1}^m \psi(i+m-1), \quad (2.11)$$

sendo $\psi(x)$ a função digama, que corresponde a primeira derivada do logaritmo neperiano da função gama $\Gamma(x)$, ou seja, $\Psi(x) = d \ln(\Gamma(x))/dx = \Gamma'(x)/\Gamma(x)$ (PRESS et al., 1992; FERREIRA, 2013).

Assim, a partir dos dois testes estatísticos propostos por Kohansal e Rezakhah (2016) tem-se as seguintes estatísticas

$$TKR_{mn}^1 = \frac{\exp(HKR_{mn}^1)}{\hat{\sigma}} = \frac{1}{\hat{\sigma}} \exp(c) \left[\prod_{i=1}^n (Y_{i+m-1} - Y_{i-m-1}) \right]^{1/n} \quad (2.12)$$

$$TKR_{mn}^2 = \frac{\exp(HKR_{mn}^2)}{\hat{\sigma}} = \frac{1}{\hat{\sigma}} \exp \left(\sum_{k=m}^n \frac{1}{k} \right) \left[\prod_{i=2}^{n-m} (Y_{i+m-1} - Y_{i-1}) \right]^{1/(n-m)}, \quad (2.13)$$

em que HKR_{mn}^1 e HKR_{mn}^2 foram definidas em 2.9 e 2.10, respectivamente.

Valores pequenos de TKR_{mn}^1 e TKR_{mn}^2 indicam não normalidade. Valores críticos são apresentados em Kohansal e Rezakhah (2016) em função de $n \leq 100$ e m , para $\alpha = 0,05$. Além disso, apenas o TKR_{mn}^1 é invariante a transformações de locação e escala. Os autores apresentam ainda um estudo de poder dos testes e concluíram que eles são mais poderosos que todas as alternativas comparadas. Entretanto, os valores de n são muito limitados e não existem funções dos testes implementadas em softwares livres ou proprietários, que se tenha informação.

Fazendo a menção das alternativas comparadas, pelos autores, foram sete estatísticas de testes incluindo as suas duas propostas de estatísticas de testes de normalidade baseados na entropia nomeadamente: o de Kohansal e Rezakhah TKR_{mn}^1 e TKR_{mn}^2 , Vasicek (TV_{mn}), Van Es (TVE_{mn}), Correa (TC_{mn}) e dois de Zamanzadeh e Arghami ($TZ1_{mn}$) e ($TZ2_{mn}$), (KOHANSAL; REZAKHAH, 2016).

2.2 Testes univariados de normalidade

Existem várias literaturas especializadas para avaliar a normalidade dos testes univariados, mas serão mencionadas aqui apenas algumas delas. É importante reiterar que marginais não-normais implicam em não-normalidade conjunta, mas marginais univariadas normais não

implicam em normalidade conjunta, embora sejam indicativo disso em interações reais. Assim, testes univariados marginais não são uma estratégia adequada. Ao se retirar uma amostra aleatória de uma população, no caso univariado, é comum muitas vezes pressupor que os dados ou resíduos do modelo adotado são provenientes de uma população normalmente distribuída. Isso é feito com a finalidade de garantir que as inferências sejam válidas, mas essa presunção deve ser testada para verificar se foi atendida ou não.

Sendo assim, a utilização de testes estatísticos como meio de inferir sobre a normalidade é necessária, visto que, a utilidade dos testes de normalidade na sua generalidade são para verificar se a distribuição de probabilidade associada a um conjunto de dados ou resíduos do modelo pode ser aproximada a uma distribuição normal. Existem vários critérios ou técnicas já discutidas na literatura, dentre elas se faz a seguinte menção: os procedimentos gráficos Q-Q plots e P-P plots, o teste de aderência qui-quadrado de Pearson (QQ), o teste de Kolmogorov-Smirnov (KS), Lilliefors (LF) e Shapiro-Wilk (W) (SCOTT, 1979; FERREIRA, 2011; FERREIRA, 2013).

Inicialmente serão abordados os critérios mais simples e empíricos e em seguida serão tratados os testes estatísticos como caminho formal para inferir sobre a normalidade. O critério gráfico (Q-Q plots e P-P plots) utiliza quantis observados e quantis esperados da distribuição hipotética, como também, utiliza as probabilidades acumuladas para avaliar a hipótese de normalidade. Para a tomada de decisão baseia-se em uma reta de mínimos quadrados, em que se muitos pontos estiverem relativamente distante dela, aventa-se a possibilidade de não-normalidade e ou presença de *outliers*. Por não ser um teste formal, acaba servindo somente como uma análise exploratória dos dados.

O teste de normalidade de qui-quadrado possui valor histórico, embora seja repleto de limitações. Ferreira (2011) discute esse assunto e mostra que para uma amostra de tamanho n aleatória, considerando cada um dos componentes dos vetores de cada vez. Assim $X_1, X_2, X_3, \dots, X_j, \dots, X_n$ é essa amostra aleatória de tamanho n para a variável X . Considera-se que a amostra aleatória, relativa a esta variável, seja proveniente de uma população normal com média (μ) e variância (σ^2) desconhecidas. Para se aplicar o teste de qui-quadrado a amostra é subdividida em classes de valores e estima-se a frequência esperada em cada uma delas. Por sua vez, as frequências esperadas são comparadas com as frequências observadas, utilizando uma regra de decisão baseada na distribuição qui-quadrado.

Uma das decisões primárias que deve ser tomada é especificar o número de classes k . Se um número grande de classes é usado, pode-se obter classes com baixas frequências observadas. Por outro lado, se o número de classes for pequeno demais, grandes desvios de normalidade podem ser obscurecidos, tanto no histograma quanto na estatística do teste. Para isso, podem ser utilizados alguns critérios para determinar k . Segundo Ferreira (2005), um critério empírico que pode ser aplicado é dado por

$$k \cong \begin{cases} \sqrt{n} & \text{se } n \leq 100 \quad (\text{inteiro mais próximo}) \\ 5 \log(n) & \text{se } n > 100 \quad (\text{inteiro mais próximo}). \end{cases} \quad (2.14)$$

Outro critério para o número de classes k atribuído a Scott (1979) é baseado em uma população normal, sendo dado por

$$k \cong 1 + \frac{A \sqrt[3]{n}}{3,49S} \quad (2.15)$$

em que S é o desvio padrão, $A = X_{(n)} - X_{(1)}$ é amplitude total, $X_{(n)}$ e $X_{(1)}$ são o máximo e o mínimo (estatística de ordem).

A estatística do teste qui-quadrado é dada por

$$\chi_c^2 = \sum_{\ell=1}^k \frac{(F_{\ell} - FE_{\ell})^2}{FE_{\ell}}, \quad (2.16)$$

em que F_{ℓ} e FE_{ℓ} são as frequências observadas e esperadas, respectivamente, na ℓ -ésima classe, segue aproximadamente uma distribuição qui-quadrado, com ν graus de liberdade sob a hipótese nula de normalidade dos dados. Os graus de liberdades da distribuição qui-quadrado, de uma forma simples e conservadora, são dados por $\nu = k - 1$. A hipótese de normalidade deve ser rejeitada para um nível nominal de significância α se $\chi_c^2 \geq \chi_{\alpha, \nu}^2$, em que $\chi_{\alpha, \nu}^2$ é quantil $100\alpha\%$ superior da distribuição qui-quadrado com $\nu = k - 1$ graus de liberdade.

Ao se trabalhar com a distribuição normal como referência, o grau de coeficiente de curtose caracteriza o achatamento que possibilita a classificação das distribuições de três maneiras: quando $\beta_2 > 3$, a distribuição é leptocúrtica; quando $\beta_2 = 3$, a distribuição é mesocúrtica; e quando $\beta_2 < 3$, é platicúrtica. Uma distribuição é leptocúrtica quando possui a curva da função densidade mais afunilada com um pico mais alto do que a distribuição normal e possui caudas pesadas, enquanto que a função de distribuição mesocúrtica tem o mesmo achatamento da dis-

tribuição normal, e por último, a distribuição platicúrtica é mais achatada do que a distribuição normal.

Os testes para os coeficientes de simetria e curtose são alternativas indiretas para a verificação da normalidade. Por sua vez, esse procedimento é limitado a grandes amostras, caracterizando assim um teste assintótico, além de não garantir resultados referentes à normalidade, pois não são realizadas comparações além do quarto momento. A distribuição só pode ser considerada igual à outra se todos os seus momentos forem iguais. Esse fato se torna impraticável em muitas situações.

O teste de Komogorov-Smirnov (*KS*) é um teste de aderência que visa verificar o grau de concordância entre distribuições em um conjunto de dados observados, buscando identificar se esses dados seguem uma distribuição almejada. O teste *KS* utiliza a distribuição de frequência acumulada teórica e a compara com a distribuição de frequência acumulada observada. A descrição da hipótese para o caso em particular do teste de normalidade é: H_0 assume que a amostra provém de uma distribuição normal (teórica). Para o uso deste teste há restrição quanto ao tamanho da amostra, além de assumir que a distribuição é conhecida (sua forma seja conhecida), não se aplica à dados qualitativos nem a variáveis discretas, a sua aplicação, precisa conhecer média (μ) e variância (σ^2) populacional, tornando assim impraticáveis, em muitas situações práticas. No caso de se pretender, por exemplo, efetuar um ajustamento de uma distribuição normal, sem especificar média (μ) e variância (σ^2), pode-se recorrer a outro teste.

O teste de Lilliefors (*LF*) é uma adaptação do teste de *KS*, usado para verificação de normalidade de um conjunto de dados. A estatística deste teste é dada por

$$D_n = \text{Max} |F(x_i) - F_n(x_i)|,$$

em que, F_n a função de distribuição $N(0,1)$ e F é a função de distribuição da amostra depois de estandardizada, i.e., definida para $x_1, x_2, x_3, \dots, x_n$ as variáveis aleatórias.

Comparando com outros testes, o teste *LF* apresenta uma flexibilidade na sua aplicação por não necessitar de especificação da média ou variância da distribuição, como também, a sua hipótese nula H_0 especifica que a população é pertencente à família de distribuições normais. Em termos de interpretação da estatística, o teste verifica se a distância máxima entre a distribuição de probabilidades acumuladas observadas padronizadas, $F(x_i)$, e a distribuição acumulada teórica (distribuição normal), $F_n(x_i)$, é estatisticamente significativa ou não, a um dado nível

nominal de probabilidade. Assim, se a distância máxima for menor do que o valor tabelado, aceita-se a hipótese H_0 , ou seja, os dados seguem a distribuição normal.

A elevada gama de proposta de novos testes de normalidade é motivada pelo fato de não existência de teste uniformemente poderoso em relação aos tamanhos amostrais e distribuições. Por sua vez, o controle da taxa do erro tipo I e o poder de cada teste depende das suposições e propriedades básicas.

Com isso, são apresentadas algumas diferenças entre alguns testes estatísticos para a normalidade, já discutidos anteriormente. (LEOTTI; COSTER; RIBOLDI, 2012) apresentam comparação de três testes (Qui-Quadrado - QQ , Kolmogorov-Smirnov - KS e Lilliefors - LF). O teste QQ , em que os dados são agrupado em classes, não depende do conhecimento dos parâmetros populacionais, como média e variância, mas necessita de amostras grandes ($n > 30$), pois, em caso contrário apresenta baixo poder e baixo controle do erro tipo I.

O teste de Lilliefors é uma adaptação do teste de KS , usado para verificação de normalidade de um conjunto de dados. Comparando os testes QQ e KS com o LF , verifica-se que, no teste LF apresenta uma flexibilidade na sua aplicação por não necessitar de especificação da média ou variância da distribuição, como também, a sua hipótese nula H_0 especifica que a população é pertencente à família de distribuições normais. De acordo com vários autores, que são unânimes em afirmar que o teste de normalidade de Shapiro-Wilk (W) tem sido o preferido por se mostrar mais poderoso na resolução de problemas de normalidade (LEOTTI; COSTER; RIBOLDI, 2012).

Na literatura, não se encontrou um teste uniformemente mais poderoso em relação aos tamanhos amostrais e distribuições. Isso tem motivado a criação de muitas propostas ou apresentações de novos testes de normalidade. A estatística do teste de Shapiro-Wilk, denotada por W , é um dos testes que tem sido amplamente utilizado como referência em trabalhos de comparação de desempenho de outros testes propostos, como também, em aplicações científicas. Pode ser utilizado para amostras de tamanho maior ou igual a 3. Sob normalidade, espera-se que o valor da estatística do teste aproxime-se de 1.

A estatística W de teste para normalidade é definida por

$$W = \frac{\left(\sum_{i=1}^n a_i X_{(i)} \right)^2}{\sum_{i=1}^n (X_i - \bar{X}_i)^2} \quad (2.17)$$

em que $X_{(i)}$ é i -ésima estatística de ordem da variável aleatória observada e (a_1, a_2, \dots, a_n) são coeficientes ou constantes tabelados.

Para a aplicação do teste é procedido o cálculo da estatística W , de uma mostra aleatória de tamanho n , dada por $X_1, X_2, X_3, \dots, X_n$, procedendo-se primeiro a ordenação das observações em ordem decrescente: $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n)}$, \bar{X} é a média amostral, \mathbf{m} é o vetor $n \times 1$ dos valores esperados da estatística de ordem da normal padrão e \mathbf{V} é a matriz $n \times n$ das variâncias e covariâncias das estatísticas de ordem da normal padrão. As constantes $a_1, a_2, a_3, \dots, a_n$, elementos do vetor \mathbf{a} , são calculadas como solução de

$$\mathbf{a} = \frac{\mathbf{m}^\top \mathbf{V}^{-1}}{(\mathbf{m}^\top \mathbf{V}^{-2} \mathbf{m})^{\frac{1}{2}}}. \quad (2.18)$$

Muitas vezes para o cálculo do vetor de constantes \mathbf{a} , pode-se utilizar aproximações precisas apresentados por Royston (1992) e por Royston (1993) para $n \geq 4$. Para isso é necessário também estimar as médias das estatísticas de ordem da normal padrão. Todavia, o vetor de constantes \mathbf{a} é obtida de forma exata para $n = 3$.

A estimativa do valor esperado da i -ésima estatística de ordem da normal padrão m_i , do i -ésimo componente de \mathbf{m} é dada por:

$$\tilde{m}_i = \Phi^{-1} \left[\frac{j - 3/8}{n + 1/4} \right]. \quad (2.19)$$

O argumento ou razão $(j - 3/8)/(n + 1/4)$ da equação 2.19, em que $j = 1, 2, 3, \dots, n$ representa a função de distribuição empírica com correções de continuidade, para se obter uma aproximação. Assim, o vetor \mathbf{m} estimado é dado por $(n \times 1)\tilde{\mathbf{m}} = (\tilde{m}_i)$. (ROYSTON, 1993) apresenta uma descrição ou roteiro para estimar o vetor de constantes \mathbf{a} :

- Primeiro, \tilde{a}_n e \tilde{a}_{n-1} são determinados por :

$$\begin{aligned} \tilde{a}_n = & c_n + 0,221157u - 0,14798u^2 - 2,071190u^3 \\ & + 4,434685u^4 - 2,706056u^5 \end{aligned}$$

e

$$\begin{aligned} \tilde{a}_{n-1} = & c_{n-1} + 0,042981u - 0,293762u^2 - 1,752461u^3 \\ & + 5,682633u^4 - 3,582633u^5, \end{aligned}$$

em que, c_n e c_{n-1} são o n -ésimo e $(n-1)$ -ésimo elementos do vetor $c = (\tilde{\mathbf{m}}^\top \tilde{\mathbf{m}})^{-1/2} \tilde{\mathbf{m}}$ e $u = 1/\sqrt{x}$.

- Determina-se ϕ de seguinte maneira:

$$\phi = \begin{cases} (\tilde{\mathbf{m}}^\top \tilde{\mathbf{m}} - 2\tilde{m}_n^2)/(1 - 2\tilde{a}_n^2) & \text{se } n \leq 5 \\ (\tilde{\mathbf{m}}^\top \tilde{\mathbf{m}} - 2\tilde{m}_n^2 - 2\tilde{m}_{n-1}^2)/(1 - 2\tilde{a}_n^2 - 2\tilde{a}_{n-1}^2) & \text{se } n > 5 \end{cases} \quad (2.20)$$

- Por último é determinado \tilde{a}_j , para $j = 2, \dots, n-1$ ($n \leq 5$) por:

$$\tilde{a}_j = \tilde{m}_j / \sqrt{\phi}. \quad (2.21)$$

Com o roteiro, mostrado anteriormente, proposto por (ROYSTON, 1993), a estatística W deve ser obtida, utilizando-se o vetor $\tilde{\mathbf{a}}$, $(n \times 1)$, e a estatística de ordem $X_{(j)}$, da seguinte forma:

$$W = \frac{\left[\sum_{j=1}^n \tilde{a}_j X_{(j)} \right]^2}{\sum_{j=1}^n (X_{(j)} - \bar{X})^2}. \quad (2.22)$$

Ao se estudar a normalidade da estatística deste teste, (ROYSTON, 1993) verificou que a estatística W , não segue uma distribuição normal. Sendo assim, propôs que se utilizasse a transformação da família Box-Cox, para se obter a normalidade, em que o valor transformado de W , será denotado por Y e dado por

$$Y = \begin{cases} -\ln[\gamma - \ln(1 - W)], & \text{se } 4 \leq n \leq 11, \\ \ln(1 - W), & \text{se } 12 \leq n \leq 5.000, \end{cases} \quad (2.23)$$

com $\gamma = -2,273 + 0,459n$.

O valor transformado de W , a estatística Y , tem média (μ_Y) e desvio padrão (σ_Y), dados, respectivamente por:

$$\mu_Y = \begin{cases} 0,5440 - 0,39978n + 0,025054n^2 - 0,0006714n^3 & \text{se } 4 \leq n \leq 11, \\ -1,5861 - 0,31082u - 0,083751u^2 + 0,0038915u^3, & \text{se } 12 \leq n \leq 5.000 \end{cases} \quad (2.24)$$

e

$$\sigma_Y = \begin{cases} \exp \{ 1,3822 - 0,77857n + 0,062767n^2 - 0,0020322n^3 \} & \text{se } 4 \leq n \leq 11 \\ \exp \{ -0,4803 - 0,082676u - 0,0030302u^2 \} & \text{se } 12 \leq n \leq 5.000, \end{cases} \quad (2.25)$$

em que $u = \ln(n)$.

Assim, com os parâmetros média (μ_Y) e desvio padrão (σ_Y), da variável Y , pode-se obter uma aproximação pela normal padrão. A estatística, denotada por Z e o seu valor- p são

$$Z = \frac{Y - \mu_Y}{\sigma_Y}, \quad \text{para } n > 3 \quad (2.26)$$

e

$$\text{valor} - p = 1 - \Phi(Z), \quad (2.27)$$

em que valor- p corresponde a área da distribuição normal padrão à direita de Z . No caso particular de $n = 3$, o vetor \mathbf{a} assume o valor $(-\sqrt{2}/2, 0, \sqrt{2}/2)^\top$. A seguir deve-se calcular Z de forma exata associado a W , em que a função arcsen deve ser tomada em radiano, como a seguir:

$$\text{valor} - p = 1 - F(W) = 1 - \frac{6}{\pi} [\arcsen(\sqrt{W}) - \arcsen(\sqrt{3/4})], \quad \text{para } n = 3. \quad (2.28)$$

O teste de Shapiro-Wilk será utilizado como referência nesse trabalho, para efeitos de comparação de desempenho dos testes de normalidade univariada propostos, baseados em entropia, verificando-se o poder em relação aos tamanhos amostrais diferentes, além de se esperar que os testes propostos tenham controle adequado da taxa de erro tipo I. As comparações dos testes qui-quadrado, Kolmogorov-Smirnov, Lilliefors e Shapiro-Wilk estão resumidas na Tabela 2.1.

2.3 Erro tipo I e poder de teste

Em testes de hipóteses, são contempladas duas suposições: a hipótese nula, denotada por H_0 e a outra definida por H_1 , chamada de alternativa. Se a hipótese nula for falsa, a alternativa

Tabela 2.1 – Comparação de distribuição de qui-quadrado, Kolmogorov-Smirnov, Lilliefors e de Shapiro-Wilk

Testes	Suas aplicações			
	Estatísticas de testes	Descrição do uso	Conhecer μ e σ^2	Amostras ($n > 30$)
QQ	$\sum_{j=1}^m \frac{(F_j - FE_j)^2}{FE_j}$	Dados agrupados em classe.	Não depende	Necessita
KS	$Max F_n(x) - F_0(x) $	Compara dist. acumulado teórico com a observada.	Depende	Não necessita
LF	$Max F(x_i) - F_n(x_i) $	Dados originais são estandardizados.	Não depende	Não necessita
W	$\frac{\left[\sum_{j=1}^n \tilde{a}_j X_{(j)} \right]^2}{\sum_{j=1}^n (X_{(j)} - \bar{X})^2}$	Ordenar as n observações da amostra.	Não depende	Não necessita

é verdadeira e vice-versa Krull e MacKinnon (2001). Afirma ainda o autor que, o poder de um teste estatístico é a probabilidade do teste rejeitar H_0 quando H_0 é realmente falsa, ou seja, o poder de um teste é igual a $1 - \beta$, (KRULL; MACKINNON, 2001).

Em estatística, mencionar erros está intrinsicamente ligado aos testes de hipóteses. As probabilidades dos erros são a probabilidade em não rejeitar a hipótese nula quando ela deveria ser rejeitada por ser falsa ou a probabilidade de se rejeitar a hipótese nula quando ela não deveria ser rejeitada, por ser verdadeira. Em geral, a hipótese nula é assumida como verdadeira para a construção dos testes de hipóteses.

O poder de um teste de hipótese é afetado por três fatores como, o tamanho da amostra, o nível de significância e o verdadeiro valor do parâmetro a ser testado. O poder do teste, em geral, refere-se a probabilidade de se rejeitar a hipótese nula quando a mesma é realmente falsa. Na prática, é importante que se tenham testes com nível de significância próximos do nível de significância nominal e que o poder seja alto, mesmo em situações de amostras pequenas (KRULL; MACKINNON, 2001). Deste modo, para se comparar o poder dos testes, calcula-se o percentual de rejeições da hipótese nula, quando amostras de distribuições não-normais foram simuladas. Entre os testes, aquele que detectar uma maior quantidade de amostras não-normais é o que tem maior poder.

A realização de testes de hipóteses, remetem o pesquisador a se ater ao risco global de tomar uma decisão errada. Ao tomar a decisão de rejeitar a hipótese H_0 , o pesquisador estará sob a possibilidade de cometer erros.

Sob H_0 verdadeira, o erro tipo I é cometido quando H_0 é rejeitada. A probabilidade do erro tipo I é também conhecida por tamanho do teste, sendo dada por:

$$\alpha = P(\text{Erro do tipo I}) = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}). \quad (2.29)$$

Em contrapartida, se H_0 é falsa e esta hipótese não é rejeitada, então se tem a ocorrência do erro tipo II. De acordo (TANI et al., 2017), a probabilidade do erro tipo II não é conhecida e nem pode ser especificada *a priori*. Ainda segundo este autor, pode-se concluir que a informação que se tem, é que para um tamanho de amostra conhecido, a probabilidade do erro tipo I é relacionado inversamente ao valor da probabilidade do erro tipo II. Neste contexto, a probabilidade de se cometer o erro tipo II, é denotada por β e dada por

$$\beta = P(\text{Erro do tipo II}) = P(\text{não se rejeitar } H_0 \mid H_0 \text{ falsa}). \quad (2.30)$$

Adicionalmente, pode-se atribuir as probabilidades, $(1 - \alpha)$ e $(1 - \beta)$ as situações em que H_0 não é rejeitada quando a mesma hipótese é verdadeira e H_0 é falsa e se rejeita H_0 , respectivamente. A probabilidade $(1 - \beta)$ é definida como poder do teste. A função poder desempenha o mesmo papel nos testes de hipóteses, sendo útil em dizer o quão bom é o teste particular (MOOD; GRAYBILL; BOES, 1974). Tais informações, podem ser resumidas na Tabela (2.2).

Tabela 2.2 – Representação tabular das decisões possíveis de um testes estatístico.

Decisão tomada	H_0 verdadeira	H_0 falsa
Não se rejeita H_0	Decisão correta	Erro tipo II
Rejeita-se H_0	Erro tipo I	Decisão correta

Um teste é classificado como conservativo, quando seu tamanho real é menor que o nível nominal α (CARMER; SWANSON, 1973). Testes que possuem tamanho igual a α e maior que α , são considerados exatos e liberais, nesta ordem.

Deve-se salientar que no estudo de avaliação de um teste estatístico, em que informações analíticas sobre taxas de erro tipo I e poder são inviáveis pode-se obtê-las indiretamente via simulação Monte Carlo, tópico que será abordado posteriormente.

Segundo (EFRON, 1982), o método *bootstrap* é basicamente uma técnica de reamostragem, que permite aproximar a distribuição de uma função das observações pela distribuição empírica dos dados baseada em uma amostra de tamanho finita. A amostragem é feita, com reposição, da distribuição da qual os dados são obtidos, se esta é conhecida (*bootstrap* Paramétrico) ou da amostra original (*bootstrap* não-paramétrico).

O *bootstrap* aborda o cálculo do intervalo de confiança de parâmetros e cálculos de valores- p , em circunstâncias em que outras técnicas não são aplicáveis, em particular no caso em que o número de amostras é reduzido. Esta técnica foi extrapolada para resolução de muitos problemas de difícil resolução por meio de técnicas de análise estatística tradicionais, baseadas na hipótese de um elevado número de amostras.

O método *bootstrap*, segundo Chernick et al. (2011), pode ser entendido como sendo uma das várias técnicas não paramétricas disponíveis atualmente para fazer inferência estatística. Por sua vez, estas técnicas são também conhecidas como métodos de reamostragem, com objetivo de determinar as propriedades da distribuição do estimador de certo parâmetro e sem fazer suposições sobre a forma da distribuição dos dados. Essa técnica de reamostragem é bastante utilizada em diferentes situações estatísticas, tendo como base a obtenção de uma amostra por reamostragem com reposição do conjunto de dados original (EFRON; TIBSHIRANI, 1993b). Existem duas maneiras de se aplicar *bootstrap*: i) *bootstrap* não-paramétrico e ii) *bootstrap* paramétrico.

No método *bootstrap* não paramétrico, supõe-se que as observações são obtidas da função de distribuição empírica $F_{(x)}$, que designa uma massa de probabilidade igual $(1/n)$ para cada ponto amostral. A amostra *bootstrap* é obtida por meio da reamostragem aleatória com reposição dos dados originais e a distribuição é desconhecida. Ao passo que no método *bootstrap* paramétrico, a amostra *bootstrap* é obtida por meio de reamostragem aleatória da distribuição geradora dos dados, com os parâmetros estimados por meio da amostra original, em que se estima os parâmetros de uma distribuição de probabilidade f_X (TSANG et al., 2001). Para o presente trabalho tem-se o interesse somente na aplicação do método *bootstrap* paramétrico e na estimação dos parâmetros da distribuição de interesse.

No método *bootstrap* paramétrico utiliza-se o conjunto de dados diretamente para se estimar os parâmetros de uma distribuição subjacente que se assume como verdadeira, como μ e σ^2 , por exemplo. A partir desses parâmetros gera-se B amostras aleatórias, cada uma de tamanho n original com reposição, e para cada amostra calcula-se uma estimativa dos parâme-

2.4 Método Monte Carlo

O método Monte Carlo, segundo Dias, Takeda e Koskinas (2006), permite resolver problemas de natureza numérica, física ou matemática, por simulação de forma aleatória, e é uma metodologia regidas por leis de probabilidade conhecidas, tornando simples os problemas que *a priori* seriam tão complexas que não poderiam ser expressos por meio de uma solução analítica. O método Monte Carlo é um mecanismo simples para gerar números aleatórios (MORETTIN; BUSSAB, 2003). A estrutura do algoritmo Monte Carlo também é simples e não precisa que se utilizem as leis de formações que o regem, sendo sua utilização modelada via função densidade de distribuição de probabilidade. Uma vez conhecidas essas distribuições, a simulação Monte Carlo pode seguir com as amostragens a partir das mesmas.

A validação de testes e de procedimentos de estimação nos mais diferentes modelos estatísticos, quando não pode ser realizada diretamente por meios analíticos, é feita por meios aproximados utilizando simulação Monte Carlo. Avaliações de robustez de testes, que são as taxas de erros tipo I e poder sob violação das pressuposições do mesmo, podem ser feitas eficientemente com a utilização de procedimentos Monte Carlo. A base de todo o processo consiste em gerar amostras aleatórias da distribuição subjacente. Assim, para o sucesso dos procedimentos Monte Carlo apesar de serem simples, é necessário ter eficientes algoritmos de geração de dados aleatórios de inúmeras distribuições de probabilidades. Essa eficiência é requerida tanto na acurácia quanto no tempo de processamento para gerar cada número aleatório.

2.5 Método *bootstrap*

O método *bootstrap* foi introduzido na década de 1980 como um método computacional para estimar o erro estandardizado de um estimador $\hat{\theta}$. Por sua vez, esse método possui a vantagem de ser completamente automatizado. Sendo um método totalmente numérico, a estimativa *bootstrap* do erro estandardizado, que também, não requer cálculos teóricos adicionais, é facilmente obtida, independentemente de quão matematicamente complicado possa ser o estimador $\hat{\theta}(x)$ a ser analisado (EFRON; TIBSHIRANI, 1993a). Dois grandes problemas (GEBERT; FERREIRA, 2013) da estimação estatística são a determinação de estimativas de um parâmetro de interesse e a avaliação de sua precisão. De forma a solucionar tais questões, (EFRON, 1982), introduziu o método *bootstrap*.

tros de interesse na inferência e por sua vez, com a distribuição *bootstrap* faz-se as inferências, (FERREIRA, 2013). O algoritmo *bootstrap*, na prática, utiliza o algoritmo Monte Carlo, com um dispositivo gerador de números aleatórios uniforme selecionado inteiros e os relaciona com as posições dos elementos do conjunto original X_1, X_2, \dots, X_n (EFRON, 1982).

A construção da distribuição *bootstrap* \hat{F} é realizada por algoritmo Monte Carlo com um número de repetições B muito grande. O custo computacional e a qualidade da convergência da estimativa do parâmetro para a estimativa natural do parâmetro $\hat{\theta}_B(B \rightarrow \infty) \rightarrow \Theta(F)$ devem ser considerados, sendo que a construção do algoritmo é geralmente simples (EFRON; TIBSHIRANI, 1993a). Os autores afirmam ainda que a sua convergência esteja garantida pela Lei dos *Grandes Números*, pois, os valores $X_1^*, X_2^*, \dots, X_n^*$ nada mais são do que uma amostra de variáveis aleatórias independentes e uniformemente distribuídas com distribuição condicional $\hat{\theta}_B$. Assim, quando B tende a infinito, o estimador $\hat{\theta}_B$ aproxima-se do parâmetro (EFRON; TIBSHIRANI, 1993a).

Silva (2006) sugere a construção do algoritmo *bootstrap* utilizando o método Monte Carlo para estimar os parâmetros por:

- i) A partir da amostra experimental, obtém-se as estimativas dos parâmetros do modelo probabilístico considerado. Utilizando-se estas estimativas, o modelo probabilístico e um gerador de números aleatórios desta distribuição, uma amostra de n valores é gerada para formar as amostras *bootstrap* de mesmo tamanho da amostra original.
- ii) Faz-se a computação da estatística desejada $\hat{\theta}_B$ em cada procedimento de reamostragem (geração de números aleatórios da distribuição alvo).
- iii) Utilizando-se os procedimentos dos passos (i) e (ii), repete-se B vezes o processo, obtendo-se, dessa maneira, B valores de $\hat{\theta}_i$.
- iv) A partir das B estimativas, forma-se a distribuição \hat{F} .
- v) Finalmente, realizam-se as inferências de interesse a partir da distribuição *bootstrap* \hat{F} .

3 MÉTODOS

Nesse capítulo são apresentadas as ideias principais da teoria e dos métodos utilizados para desenvolver um teste de normalidade univariada, baseado na entropia. Neste caso, buscou-se desenvolver um teste que fosse mais poderoso para uma ampla variedade de distribuições não normais sob a hipótese alternativa. A ideia foi aproveitar a sugestão de Vasicek (1976), porém com estimadores mais eficientes, como os de Kohansal e Rezakhah (2016), e desenvolver um teste computacionalmente intensivo via *bootstrap* paramétrico.

Na sequência desta seção são apresentados os estimadores de entropia utilizados, as estatísticas e outros detalhes dos testes, os procedimentos *bootstrap* paramétrico para gerar as distribuições nulas das estatísticas e as expressões para obtenção dos valores- p . Em seguida, é apresentada uma seção de como se dá a avaliação do desempenho dos testes propostos, além do teste de Shapiro-Wilk, usado como referência. Além das propostas de três testes, um quarto teste combinando as melhores alternativas foi apresentada. Finalmente é descrito o procedimento para se aplicar o teste binomial exato na avaliação do tamanho dos testes utilizados no presente trabalho.

3.1 Métodos

Considere uma amostra aleatória $X_1, X_2, X_3, \dots, X_n$ de tamanho n proveniente de uma função densidade comum f_X , para a qual se almeja testar a hipótese

$$H_0 : f_X(x) = f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (3.1)$$

contra a hipótese alternativa

$$H_1 : f_X(x) \neq f_0(x) \quad (3.2)$$

para $\mu \in \mathbb{R}^1$ e $\sigma^2 > 0$, significando que a amostra é originada da distribuição normal com média μ e variância σ^2 .

Assim, o teste para a hipótese 3.1 consistirá em aplicar um procedimento baseado na técnica de *bootstrap* paramétrico. Para isso os estimadores de μ e σ^2 na amostra original são computados por

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad e \quad \hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (3.3)$$

respectivamente. Também são considerados os estimadores de entropia, como o de Vasicek (1976), dado por

$$HV_{mn} = n^{-1} \sum_{i=1}^n \ln \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}, \quad (3.4)$$

e os estimadores de entropia baseados em média móvel de Kohansal e Rezakhah (2016), sendo um deles dado por

$$\begin{cases} HKR_{mn}^1 = \frac{1}{n} \sum_{i=2}^n (Y_{i+m-1} - Y_{i-m-1}) + c, & se \quad n \leq 8, \\ HKR_{mn}^1 = \frac{1}{n} \sum_{i=1}^n (Y_{i+m-1} - Y_{i-m-1}) + c, & se \quad n > 8, \end{cases} \quad (3.5)$$

e o outro, dado por

$$HKR_{mn}^2 = \frac{1}{n-m} \sum_{i=2}^{n-m} \ln (Y_{i+m-1} - Y_{i-1}) + \sum_{k=m}^n \frac{1}{k}, \quad (3.6)$$

em que $Y_i = Y_1$, se $i < 1$, $Y_i = Y_n$, para $i > n$ e

$$c = - \left(1 - \frac{2m}{n} \right) \psi(2m) + \psi(n+1) - \frac{2}{n} \sum_{i=1}^m \psi(i+m-1),$$

sendo $\psi(x)$ a função digama, que corresponde a primeira derivada do logaritmo neperiano da função gama $\Gamma(x)$, ou seja, $\Psi(x) = d \ln(\Gamma(x)) / dx = \Gamma'(x) / \Gamma(x)$ (PRESS et al., 1992; FERREIRA, 2013). Convém salientar que os valores Y_i 's são as estatísticas de ordem suavizadas (médias móveis) dadas para uma janela de tamanho $w < n$, por:

$$Y_i = \begin{cases} \frac{\sum_{j=1}^i X_{(j)}}{i}, & 1 \leq i \leq (w-1)/2, \\ \frac{\sum_{j=i-(w-1)/2}^i X_{(j)}}{w}, & (w+1)/2 \leq i \leq n - (w-1)/2, \\ \frac{\sum_{j=i}^n X_{(j)}}{n-i+1}, & n - (w-3)/2 \leq i \leq n, \end{cases} \quad (3.7)$$

para $i = 1, 2, 3, \dots, n$, em que $X_{(j)}$'s são estatísticas de ordem amostrais, com $j = 1, 2, 3, \dots, n$. Foi usada a recomendação dos autores de usar uma janela w de tamanho 3, ou seja, $w = 3$.

Convém salientar que um dos estimadores de entropia de Kohansal e Rezakhah (2016) e que foi apresentado anteriormente, trata-se na realidade de uma modificação. O estimador modificado no presente trabalho é aquele dado em (3.5).

Com isso, o presente trabalho propôs três diferentes alternativas, baseadas nestes três estimadores, com todas fazendo uso da técnica de *bootstrap* paramétrico para se aplicar o teste de normalidade. Assim, para a amostra original, obtiveram-se os valores das estatísticas modificadas das três alternativas, dadas por

$$KB_{mn} = \frac{\exp(HV_{mn})}{\hat{\sigma}} = \frac{n}{2m\hat{\sigma}} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n}, \quad (3.8)$$

$$\begin{cases} TKRB_{mn}^1 = \frac{\exp(HKR_{mn}^1)}{\hat{\sigma}} = \frac{1}{\hat{\sigma}} \exp(c) \left[\prod_{i=2}^n (Y_{i+m-1} - Y_{i-m-1}) \right]^{1/n}, & se \quad n \leq 8, \\ TKRB_{mn}^1 = \frac{\exp(HKR_{mn}^1)}{\hat{\sigma}} = \frac{1}{\hat{\sigma}} \exp(c) \left[\prod_{i=1}^n (Y_{i+m-1} - Y_{i-m-1}) \right]^{1/n}, & se \quad n > 8, \end{cases} \quad (3.9)$$

e

$$TKRB_{mn}^2 = \frac{\exp(HKR_{mn}^2)}{\hat{\sigma}} = \frac{1}{\hat{\sigma}} \exp\left(\sum_{k=m}^n \frac{1}{k}\right) \left[\prod_{i=2}^{n-m} (Y_{i+m-1} - Y_{i-1}) \right]^{1/(n-m)}. \quad (3.10)$$

Os valores de m usados seguiram aqueles recomendados por Kohansal e Rezakhah (2016), sendo dados por: $m = 1$, se $n \leq 8$; $m = 2$ se $9 \leq n \leq 15$; $m = 3$ se $16 \leq n \leq 35$; $m = 4$ se $36 \leq n \leq 60$; $m = 5$ se $61 \leq n \leq 80$; $m = 6$ se $81 \leq n \leq 100$. Para $n > 100$, o valor de m é aquele apresentado pela recomendação de Crzregorzewski e Wirczorkowski (1999), dada por $m = \lfloor \sqrt{n+0,5} \rfloor$, em que $\lfloor x \rfloor$ é o maior inteiro menor ou igual a x .

Em vez de se comparar os valores das estatísticas dos testes obtidos na amostra com os valores críticos fornecidos pelos autores dos testes originais, foi apresentado uma alternativa computacionalmente intensiva, baseada em *bootstrap* paramétrico. Assim a decisão de rejeição ou não da hipótese de normalidade univariada foi tomada com base na distribuição nula, cujos passos para obtê-las é descrito na sequência.

Com as estimativas de $\hat{\mu}$ e $\hat{\sigma}^2$ de máxima verossimilhança apresentada em (3.3) foram gerados B amostras aleatórias *bootstrap* paramétrico sob normalidade, ou seja, foram gerados n valores independentes $X_j^* \sim N(\hat{\mu}, \hat{\sigma}^2)$, para $j = 1, 2, 3, \dots, n$. Em cada uma destas amostras *bootstrap*, as estatísticas 3.8, 3.9 e 3.10 respectivamente são computadas. Como o processo foi repetido B vezes, então, separadamente para cada estatística, os B valores foram armazenados em três vetores diferentes de dimensão $(B+1) \times 1$, pois cada um deles continha também o valor da respectiva estatística computado na amostra original. Os dados destes vetores se referem à distribuição nula *bootstrap* paramétrico da correspondente estatística considerada. É conveniente reiterar que são três testes independentes de normalidade sendo considerados, mas todos baseados em estimadores de entropia. Como um valor pequeno de qualquer uma das estatísticas indica não normalidade, então os valores- p , considerando que as estatísticas *bootstrap* dos três procedimentos foram representadas por KB_b , $TKRB_b^1$ e $TKRB_b^2$, respectivamente, foram calculados por

$$\text{valor-}p^{KB} = \frac{\sum_{b=1}^{B+1} I(KB_{mn} \leq KB_b)}{B+1},$$

$$\text{valor-}p^{TKRB^1} = \frac{\sum_{b=1}^{B+1} I(TKRB_{mn}^1 \leq TKRB_b^1)}{B+1}$$

e

$$\text{valor-}p^{TKRB^2} = \frac{\sum_{b=1}^{B+1} I(TKRB_{mn}^2 \leq TKRB_b^2)}{B+1},$$

em que $I(\text{condição})$ é uma função indicadora que retorna 1, se a condição for verdadeira e retorna 0, em caso contrário.

Assim, para um nível nominal de significância α , sendo “valor- p ” representante do valor- p de uma das alternativas anteriores, então a hipótese nula de normalidade univariada deve ser rejeitada se $\text{valor-}p \leq \alpha$. Uma das diferenças dos procedimentos propostos e os respectivos testes originais é a não dependência de tabelas específicas e limitadas para os valores de n , em geral, com $n \leq 100$. A outra, é o uso do *bootstrap* paramétrico, que potencialmente contornou o problema da não invariância de um dos testes. Além disso, realizou-se, ainda, combinação linear dada pela média das estatísticas TKR_1 e TKR_2 , resultando em um outro teste combinado $TKRB_C$, e sendo assim, um teste potencialmente mais poderoso. O valor B

considerado foi de 2.000 reamostragens *bootstrap*. As configurações em relação aos tamanhos amostrais consideradas estão apresentadas na Tabela 3.1.

Tabela 3.1 – Tamanhos amostrais utilizados nas simulações Monte Carlo e *bootstrap*

Teste de normalidade baseado na entropia				SW	n
<i>TVB</i>	<i>TKRB₁</i>	<i>TKRB₂</i>	<i>TKRB_C</i>	<i>TW</i>	3(1)10
<i>TVB</i>	<i>TKRB₁</i>	<i>TKRB₂</i>	<i>TKRB_C</i>	<i>TW</i>	15(5)50
<i>TVB</i>	<i>TKRB₁</i>	<i>TKRB₂</i>	<i>TKRB_C</i>	<i>TW</i>	100,500,1.000,5.000

Foram realizadas 1.000 e 2.000 simulações Monte Carlo e *bootstrap*, respectivamente, utilizando as funções geradoras de variáveis aleatórias do programa R ou implementações próprias, também em linguagem R. Os testes de normalidade univariados, baseados na entropia, propostos (*TVB*, *TKRB₁*, *TKRB₂*, *TKRB_C*) neste trabalho e o de comparação, teste de Shapiro-Wilk (*TW*), foram aplicados em cada amostra normal univariada simulada aos níveis de significância nominais de 10%, 5% e 1%. Finalmente, foi computado o número de vezes que a hipótese nula é rejeitada, sendo este resultado dividido pelo número de simulações, fornecendo as taxas de erro tipo I empíricas, sob H_0 e poder, sob H_1 .

3.2 Avaliação do desempenho

Os desempenhos dos testes propostos foram avaliados por simulação Monte Carlo. Foram consideradas duas etapas para isso. Na primeira avaliou-se o desempenho em relação ao erro tipo I e na segunda, em relação ao poder. Para o erro tipo I, foi necessário simular amostras de tamanho n sob a hipótese nula, que no caso, é a hipótese da normalidade. Assim, as amostras de tamanho n da distribuição normal devem ser geradas. Para cada amostra simulada da distribuição normal, os testes propostos foram aplicados e o resultado final, se a amostra for rejeitada, no caso indevidamente, anotado. As taxas de erro tipo I empíricas são computadas ao final de N simulações Monte Carlo pela proporção de rejeições indevidas. Os valores foram comparados com o valor nominal α de significância adotado pelo teste binomial exato (OLIVEIRA; FERREIRA, 2010). Foram considerados valores de α de 1%, 5% e 10% e diferentes valores de n , quais sejam, 3(1)10, 15(5)50, 100, 500, 1.000, 5.000, conforme apresentado na Tabela 3.1. Foram consideradas $N=1.000$ simulações Monte Carlo. O teste de Shapiro-Wilk (*TW*) (ROYSTON, 1993), foi aplicado para fins de comparação, por se tratar do teste de normalidade reconhecidamente como sendo um dos mais poderosos para uma ampla gama de distribuições alternativas.

Numa segunda etapa foi avaliado o poder em amostras de tamanho n , conforme apresentado na Tabela 3.1, sob a hipótese alternativa, que nesse caso é de não normalidade dos dados, logo amostras não normais foram simuladas. Assim, foram consideradas amostras de algumas distribuições não-normais como a exponencial ($\lambda = 1$), gama ($\alpha = 1, \beta = 3$), t de student com ($\nu = 1$) grau de liberdade, t de student com ($\nu = 30$) grau de liberdade, uniforme ($a = 0, b = 1$), log-normal ($\mu = 0, \sigma = 1$), Weibull ($\delta = 2, \alpha = 3$) e a distribuição beta ($\alpha = 0,2; \beta = 0,5$). As mesmas configurações em relação aos valores de α , n , B e N foram consideradas nessa etapa de avaliação de desempenho. Novamente o teste de Shapiro-Wilk (TW) foi aplicado. Tudo isso foi realizado com o objetivo de se encontrar uma proposta de teste mais poderosa para uma ampla gama de distribuições não normais e que, sob normalidade, haja controle adequado do erro tipo I, ou seja, que o teste seja exato. Deste modo, para se comparar o poder dos testes, TVB , $TKRB_1$, $TKRB_2$, $TKRB_C$ e TW , calculou-se o percentual de rejeições da hipótese nula, quando amostras de distribuições não-normais foram simuladas. Entre os testes, aquele que detectar uma maior quantidade de amostras não-normais é o que será considerado mais poderoso.

3.3 Avaliação do erro tipo I e poder do teste combinado

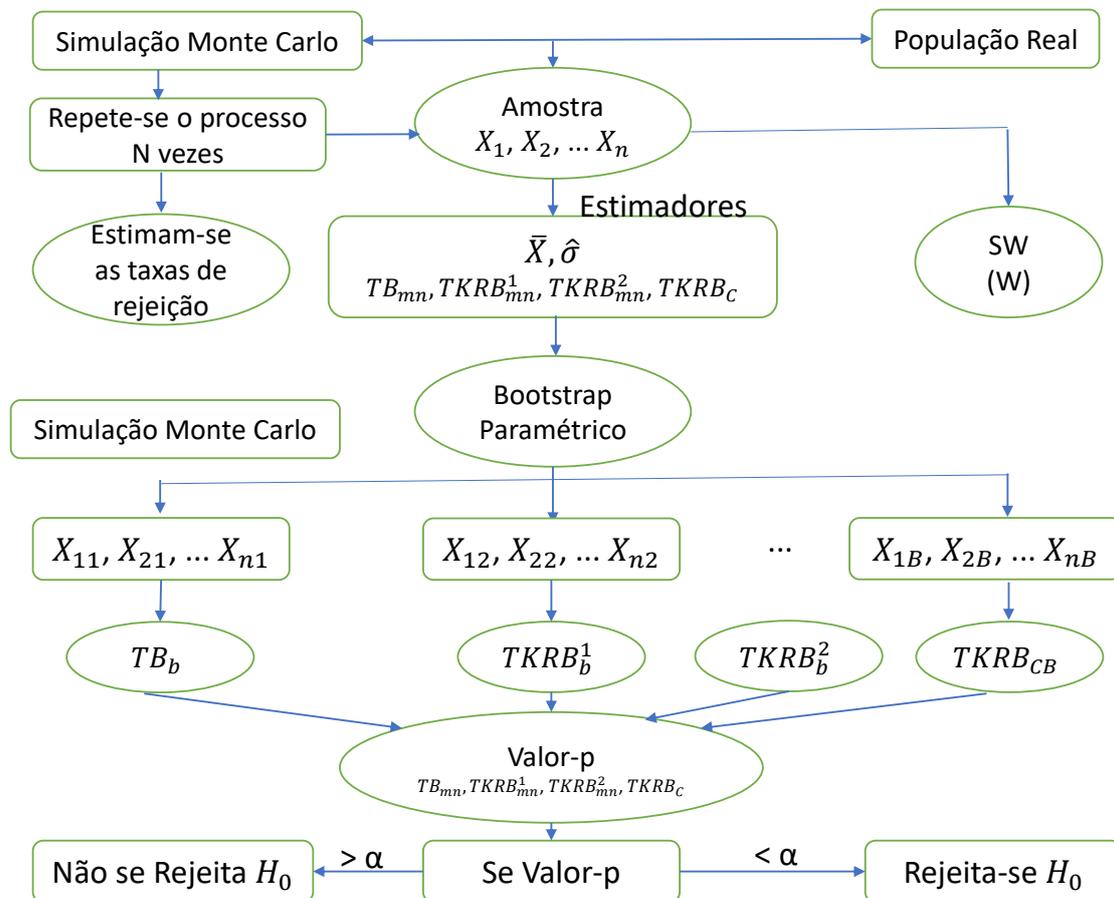
A avaliação do desempenho para o teste combinado ($TKRB_C$), configurando assim, a quarta proposta, foi também realizada via simulação Monte Carlo, consideradas duas etapas, do mesmo modo que foi realizado na Seção (3.2). Na primeira etapa, avaliou-se o desempenho em relação ao erro tipo I e na segunda, em relação ao poder. As configurações em relação aos tamanhos amostrais para o teste combinado ($TKRB_C$) foram simplificadas, sendo dadas por $n = 3, 10, 30, 50, 100, 500, 1.000$ e 5.000 .

A estatística do teste combinado ($TKRB_C$) proposto foi obtida pela média aritmética das estatísticas dos testes $TKRB_1$ e $TKRB_2$, que foram consideradas de melhor desempenho na primeira parte desta proposta. Para o erro tipo I, foram simuladas amostras de tamanho n , dadas pelas configurações apresentadas anteriormente, sob a hipótese nula, que no caso, foi sob normalidade. Assim, amostras de tamanho n da distribuição normal foram geradas e para cada amostra simulada da distribuição normal, o teste proposto foi aplicado e o resultado final, se a amostra for rejeitada, no caso indevidamente, anotado. As taxas de erro tipo I empíricas foram computadas ao final de N simulações Monte Carlo pela proporção de rejeições indevidas. Os valores foram comparados com o valor nominal α de significância adotado pelo teste binomial exato (OLIVEIRA; FERREIRA, 2010). Foram considerados, as mesmas configurações dos

valores de $\alpha = 1\%$, 5% e 10% e diferentes valores de n . Neste caso também foi considerado um valor de simulações Monte Carlo dado por $N = 1.000$.

Na segunda etapa, foi avaliada o poder utilizando amostras de tamanho n , nas mesmas configurações já referenciadas, sob a hipótese alternativa não normais. Assim, foram consideradas amostras de oito 8 distribuições não-normais, como a exponencial ($\lambda = 1$), gama ($\alpha = 1, \beta = 3$), t-Student com ($\nu = 1$) grau de liberdade, t-Student com ($\nu = 30$) grau de liberdade, uniforme ($a = 0, b = 1$), log-normal ($\mu = 0, \sigma = 1$), Weibull ($\delta = 2, \alpha = 3$) e a distribuição beta ($\alpha = 0,2; \beta = 0,5$). As mesmas configurações em relação aos valores de α , n , B e N foram consideradas nessa segunda etapa de avaliação de desempenho. Deste modo, para se comparar o poder do teste, $TKRB_C$, calculou-se o percentual de rejeições da hipótese nula e aquele que apresentou uma maior quantidade de rejeições foi considerado o mais poderoso. Na Figura 3.1 está apresentado o fluxograma, que resume os métodos apresentados nas Seções 3.1, 3.2 e 3.3.

Figura 3.1 – Fluxograma da estrutura dos métodos.



Fonte: O autor

3.4 Teste binomial exato

Para as simulações sob H_0 espera-se que as taxas de erro empíricas sejam iguais ao nível de significância nominal α . Como existe aleatoriedade nas medidas destas taxas tornou-se necessário testar se algum desvio do valor observado em relação ao nível nominal de significância α adotado foi devido ao mero acaso ou não (OLIVEIRA; FERREIRA, 2010). Para isso foi aplicado o teste binomial exato para a hipótese $H_0: \alpha_k = \alpha$, em que α_k é o tamanho real do teste k , sendo que, $k = 1$, refere-se ao teste TVB , $k = 2$, refere-se ao teste $TKRB_1$, $k = 3$, refere-se ao teste $TKRB_2$, $k = 4$, refere-se ao teste $TKRB_C$ e $k = 5$, refere-se ao teste TW , sendo $\alpha = 10\%$, 5% ou 1% .

Assim, considerando que y é o número de hipóteses nulas H_0 rejeitadas, em N simulações Monte Carlo, para um nível nominal de significância α e para algum dos 5 testes, a estatística do teste binomial exato, usando a relação exata entre a distribuição F e a binomial, em um nível de significância α^* , é

$$F = \binom{y+1}{N-y} \left(\frac{1-\alpha}{\alpha} \right)$$

que, sob H_0 , tem distribuição F com $\nu_1 = 2(N - y)$ e $\nu_2 = 2(y + 1)$ graus de liberdade.

A hipótese nula (H_0), deve ser rejeitada ao nível nominal de significância α^* , se $F \leq F_\alpha$ ou $F \geq F_{1-\alpha}$, em que F_α e $F_{1-\alpha}$ são os quantis da distribuição F com ν_1 e ν_2 graus de liberdade, apresentados anteriormente.

No presente trabalho o teste binomial exato foi aplicado para as taxas de erro tipo I considerando o tamanho da amostra da simulação Monte Carlo, $N = 1.000$, no nível de significância $\alpha^* = 1\%$. Portanto, se as taxas de erro tipo I forem significativamente inferiores ($P < 0,01$), no nível nominal α , o teste proposto foi considerado conservativo. Caso, seja significativamente superior ($P < 0,01$) ao nível nominal α , considerou-se que o teste é liberal. Por fim, caso as taxas de erro tipo I não sejam significativamente diferentes ($P > 0,01$), o teste proposto foi considerado exato.

4 RESULTADOS E DISCUSSÃO

Alguns resultados obtidos por via das simulações Monte Carlo para validação dos testes de normalidade univariados baseado na entropia, propostos neste trabalho e, também, da comparação com os resultados teste de Shapiro-Wilk (TW) são apresentados nesta seção. O capítulo foi dividida em três seções, sendo que a primeira se refere ao erro tipo I, a segunda, ao poder dos testes para os níveis de significância fixados de 10%, 5% e 1% e diferentes valores de n , quais sejam, 3(1)10, 15(5)50, 100, 500, 1.000, 5.000, conforme apresentado na seção 3.2. As comparações foram feitas considerando fixos os valores do nível nominal de significância e de n . E por último, na terceira seção é apresentado o desempenho do teste combinado.

Inicialmente, são apresentados os resultados para o erro tipo I na seção (4.1) e, posteriormente, são apresentados os resultados do poder dos testes. Convém salientar que o teste binomial exato foi aplicado a cada resultado das taxas de erro tipo I, conforme descrição apresentada na seção (3.4), considerando um nível nominal de significância de 1%. Assim, os resultados foram classificados em conservativos, exatos e liberais, sendo que os dois primeiros casos configuram uma situação de controle das taxas de erro tipo I do teste. O caso conservativo, como dito, embora represente uma situação de controle do erro tipo I, possui taxas de erro reais inferiores aos valores nominais adotados e, por isso, espera-se que apresente menor poder. Os testes liberais não são desejáveis, uma vez que as rejeições das hipóteses nulas se dão acima dos níveis nominais. Estes tipos de testes estão, em geral, associados à valores de poder elevados.

4.1 Erro tipo I

O erro tipo I, cujas definições e propriedades foram apresentadas na seção (2.3), página 26, tem probabilidade de ocorrência inversamente proporcional ao poder do teste. Além disso, a probabilidade complementar a do erro tipo II é denominada de poder do teste de hipótese. A probabilidade de se cometer o erro tipo II é denotado por β . Assim, se a probabilidade de se incorrer no erro tipo I for reduzida, a probabilidade de se cometer o erro tipo II aumentará. Quando as taxas de erro tipo I observadas não forem significativamente diferentes do nível nominal de significância fixado, o teste é considerado exato. Se essas taxas forem inferiores ao nível de significância nominal fixado, o teste é tido como conservativo. Por fim, se as taxas são superiores ao nível nominal de significância, o teste é liberal.

Nas Tabelas 4.1 e 4.2 são apresentados os erros tipo I dos testes TVB , $TKRB_1$, $TKRB_2$ e TW , nos níveis de significância nominais de 10%, 5% e 1%, respectivamente, e diferentes valores de n , conforme apresentado na Seção (3.1), página 36. Especificamente na Tabela 4.1 são apresentadas as taxas de erro tipo I para os níveis de significância nominais de 10%, 5% e 1%, e diferentes valores de n , 3(1)10, 15(5)50. Em todos estes casos, os dados foram simulados considerando a distribuição normal univariada. As taxas de erros empíricas foram testados pelo teste binomial exato para verificar se os valores estimados diferem significativamente ($P < 0,01$) dos níveis nominais correspondentes fixados.

Tabela 4.1 – Erro tipo I dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , em função de $n=3(1)10, 15(5)50$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

n	α	Teste Normal				n	α	Teste Normal			
		TVB	$TKRB_1$	$TKRB_2$	TW			TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,09	0,094	0,133 ⁺	0,089	15	10	0,108	0,095	0,100	0,091
	5	0,043	0,051	0,066	0,042		5	0,057	0,057	0,057	0,050
	1	0,005	0,006	0,009	0,006		1	0,013	0,016	0,012	0,012
4	10	0,091	0,093	0,114	0,097	20	10	0,111	0,104	0,104	0,107
	5	0,048	0,04	0,052	0,051		5	0,052	0,061	0,051	0,051
	1	0,008	0,006	0,007	0,013		1	0,007	0,012	0,007	0,006
5	10	0,09	0,096	0,115	0,094	25	10	0,096	0,093	0,118	0,110
	5	0,043	0,041	0,049	0,049		5	0,045	0,035	0,046	0,051
	1	0,009	0,009	0,009	0,004		1	0,01	0,007	0,010	0,011
6	10	0,106	0,09	0,105	0,089	30	10	0,100	0,099	0,093	0,089
	5	0,057	0,045	0,054	0,046		5	0,048	0,058	0,043	0,050
	1	0,012	0,012	0,014	0,009		1	0,014	0,011	0,005	0,010
7	10	0,093	0,101	0,117	0,092	35	10	0,109	0,107	0,096	0,095
	5	0,050	0,051	0,060	0,040		5	0,061	0,054	0,048	0,048
	1	0,009	0,008	0,009	0,007		1	0,017	0,017	0,005	0,009
8	10	0,103	0,114	0,107	0,097	40	10	0,092	0,096	0,094	0,090
	5	0,056	0,062	0,059	0,048		5	0,051	0,054	0,041	0,042
	1	0,012	0,012	0,014	0,009		1	0,013	0,015	0,011	0,010
9	10	0,094	0,112	0,108	0,107	45	10	0,114	0,112	0,087	0,098
	5	0,050	0,058	0,051	0,053		5	0,062	0,054	0,036	0,055
	1	0,016	0,017	0,011	0,006		1	0,010	0,011	0,005	0,018
10	10	0,112	0,102	0,106	0,105	50	10	0,091	0,082	0,102	0,107
	5	0,064	0,053	0,056	0,056		5	0,050	0,051	0,047	0,052
	1	0,012	0,008	0,012	0,009		1	0,011	0,007	0,006	0,011

⁺ Taxas de erro tipo I superiores ao valor nominal de significância α respectivo valor ($P < 0,01$)

Para o nível de significância de 10%, em diferentes valores de n , 3(1)10, 15(5)50, verificou-se que os testes em estudo TVB , $TKRB_1$ e $TKRB_2$ apresentaram taxas de erro tipo I não significativamente diferentes ($P > 0,01$) do valor nominal $\alpha = 10\%$ para todos os n 's, exceto em apenas um caso. Desse modo, os testes propostos (TVB , $TKRB_1$, $TKRB_2$) podem ser considerados exatos ao nível de significância de 10%, dado que as suas taxas de erro tipo I não foram significativamente diferentes do nível nominal de 10% ($P > 0,01$). A exceção mencionada ocorreu no teste $TKRB_2$ com $n = 3$ e valor nominal de significância $\alpha = 10\%$. Neste caso, o teste mostrou-se liberal, por ser sua taxa de erro empírica significativamente superior ($P < 0,01$) ao nível nominal α adotado. Embora isso tenha ocorrido, o valor encontrado (0,133) não foi expressivamente diferente de valor nominal 0,10. O teste Shapiro-Wilk (TW), usado para fins de comparação, se mostrou exato para todos os valores de n , 3(1)10, 15(5)50, considerando-se o nível de significância de 10%. Pode-se dizer que os testes apresentaram controle adequado das taxas de erro tipo I.

Verificou-se que as taxas de erro tipo I empíricas foram não significativamente ($P > 0,01$) diferentes dos níveis nominais em todos os casos sem exceção, considerando, ainda, as mesmas situações descritas anteriormente para $\alpha = 10\%$, mas considerando os níveis nominais de significância α de 5% e 1%. O mesmo ocorreu para o TW , usado como teste referência. Comparando os resultados dos testes propostos TVB , $TKRB_1$, $TKRB_2$, com o TW , verificou-se que todos apresentaram taxas de erro tipo I observadas não significativamente diferentes ($P > 0,01$) do nível nominal α de 5% e 1%, para todos os n , 4(1)10, 15(5)50. Assim, pode-se considerar que os testes são considerados exatos, mesmo com a exceção apresentada para $n = 3$, do teste $TKRB_2$ com $\alpha = 10\%$, em que foi liberal. Credita-se este resultado a um possível erro tipo I, uma vez que vários testes simultâneos são realizados e nenhuma correção é feita no nível nominal de significância adotado de 1% para o teste binomial exato.

Na Tabela 4.2 são apresentadas as taxas de erro tipo I para os níveis de significância nominais de 10%, 5% e 1% e valores de n , 100, 500, 1.000 e 5.000, sob distribuição normal univariada dos dados. Novamente foi verificado se os valores estimados das taxas de erro tipo I empíricas diferiram significativamente ($P < 0,01$) ou não dos níveis nominais fixados pelo teste binomial exato. Os resultados apresentados na Tabela 4.2, refletem os resultados de grandes amostras, ($n \geq 100$).

A partir da Tabela 4.2, considerando-se os três níveis de nominais de significância de 10%, 5% e 1% e os valores de n , 100, 500, 1.000 e 5.000, verificou-se que todos os testes em

Tabela 4.2 – Erro tipo I dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , em função de $n = 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

n	α	<i>Teste Normal</i>			
		TVB	$TKRB_1$	$TKRB_2$	TW
100	10	0,113	0,113	0,104	0,093
	5	0,053	0,053	0,052	0,045
	1	0,010	0,008	0,008	0,008
500	10	0,102	0,106	0,096	0,103
	5	0,056	0,051	0,043	0,054
	1	0,015	0,014	0,008	0,014
1000	10	0,107	0,105	0,095	0,104
	5	0,051	0,052	0,045	0,052
	1	0,009	0,009	0,006	0,008
5000	10	0,088	0,089	0,099	0,079
	5	0,043	0,044	0,045	0,041
	1	0,011	0,011	0,011	0,008

estudo, TVB , $TKRB_1$, $TKRB_2$ e TW , apresentaram controle adequado do erro tipo I. Verificou-se que os testes propostos e o teste Shapiro-Wilk (TW), apresentaram taxas de erro tipo I não significativamente diferentes ($P > 0,01$) do valor nominal. Novamente, deste ponto de vista estes testes são exatos, ou seja, os tamanhos são iguais ao nível nominal de significância, como o esperado, verificou-se o melhor deles ou os melhores deles, avaliando-se o poder dos mesmos. Esta etapa foi feita posteriormente neste trabalho, na Seção (4.2).

Ademais, é conveniente salientar que os testes propostos não são limitados a amostras de tamanho 5.000, como acontece com o TW . Assim, esta é uma vantagem relativa dos testes propostos que não deve ser ignorada, principalmente quando se lida com dados ou resíduos de modelos com dezenas ou centenas de milhares de observações, como ocorre em estudos das áreas florestal, melhoramento de animal e vegetal, entre outras possibilidades. Do mesmo modo, em relação ao aspecto conservativo dos testes, não houve nenhum resultado do teste binomial neste sentido. Isso, a princípio, em termo teóricos esperados, reflete que os novos testes apresentaram um potencial ao menos poder idêntico e muitos casos, superior ao do teste de Shapiro-Wilk, o TW .

4.2 Poder do teste

O poder do teste, cujas definições e propriedades foram apresentadas na Seção (2.3), página 26, refere-se em geral a probabilidade de se rejeitar a hipótese nula quando a mesma é realmente falsa, e o erro tipo II é cometido em um teste quando a hipótese nula não é rejeitada, sendo ela falsa. Na prática, é importante que se tenham testes com nível de significância iguais ao nível de significância nominal e que o poder seja o maior possível, mesmo em situações de amostras pequenas (KRULL; MACKINNON, 2001). As mesmas configurações em relação aos valores de α , n , B e N foram consideradas nessa Seção de avaliação do poder. Assim, amostras de tamanho n , nas configurações apresentadas na Tabela 3.1, página 36, sob a hipótese alternativa, foram geradas. O teste de Shapiro-Wilk (TW) foi novamente aplicado, juntamente com os testes propostos. A hipótese alternativa nesse caso é de não normalidade dos dados. Logo amostras foram consideradas, sendo elas oito distribuições não-normais: exponencial ($\lambda = 1$), gama ($\alpha = 1, \beta = 3$), t-Student ($\nu = 1$), t-Student ($\nu = 30$), uniforme ($a = 0, b = 1$), log-normal ($\mu = 0, \sigma = 1$), Weibull ($\delta = 2, \alpha = 3$) e beta ($\alpha = 0,2; \beta = 0,5$).

Para se comparar o poder dos testes, TVB , $TKRB_1$, $TKRB_2$ e TW , calculou-se o percentual de rejeições da hipótese nula, quando amostras de distribuições não-normais foram simuladas. Entre os testes, nas mesmas configurações em relação aos valores de α , n e N , aquele que detectar maior quantidade de amostras não-normais é o que tem maior poder. Os valores obtidos nas simulações foram apresentados em tabelas para cada distribuição considerada, a fim de facilitar a comparação.

Nas Tabelas 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 e 4.10, são apresentados os resultados de poder para as distribuições não-normais exponencial ($\lambda = 1$), gama ($\alpha = 1, \beta = 3$), t-Student com ($\nu = 1$) grau de liberdade, t-Student com ($\nu = 30$) grau de liberdade, uniforme ($a = 0, b = 1$), log-normal ($\mu = 0, \sigma = 1$), Weibull ($\delta = 2, \alpha = 3$) e beta ($\alpha = 0,2; \beta = 0,5$), respectivamente, para os testes TVB , $TKRB_1$, $TKRB_2$ e TW , nos níveis de significância nominais α de 10%, 5% e 1% e diferentes valores de n , quais sejam, 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000.

Inicialmente, é apresentado na Tabela 4.3, o poder dos testes para os níveis de significância nominais de 10%, 5% e 1%, considerando diferentes valores de n e a distribuição exponencial ($\lambda = 1$) univariada. Observou-se para esta distribuição, que o teste $TKRB_1$ apresentou maior poder para qualquer valor de n nos três níveis de significância nominais adotados, quando comparado aos resultados dos outros testes aplicados (TVB , $TKRB_2$ e TW). Uma única exceção ocorreu no caso de $n = 3$, em que o teste $TKRB_2$ apresentou maior poder nos três níveis

de significância nominais adotados. À medida que o tamanho das amostras cresce, o desempenho dos testes tenderam a se igualar, sendo que para $n \geq 50$, eles apresentaram desempenho praticamente iguais.

Tabela 4.3 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição exponencial ($\lambda = 1$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

		<i>Exponencial</i> (λ)						<i>Exponencial</i> (λ)			
<i>n</i>	α	<i>TVB</i>	<i>TKRB</i> ₁	<i>TKRB</i> ₂	<i>TW</i>	<i>n</i>	α	<i>TVB</i>	<i>TKRB</i> ₁	<i>TKRB</i> ₂	<i>TW</i>
3	10	0,143	0,209	0,268	0,141	35	10	0,999	0,999	0,957	0,998
	5	0,078	0,114	0,142	0,078		5	0,996	0,998	0,913	0,994
	1	0,019	0,035	0,043	0,020		1	0,946	0,978	0,802	0,931
5	10	0,217	0,346	0,310	0,261	40	10	1,000	1,000	0,968	0,997
	5	0,122	0,219	0,205	0,160		5	0,996	1,000	0,949	0,996
	1	0,020	0,064	0,053	0,057		1	0,981	0,992	0,858	0,966
8	10	0,333	0,490	0,349	0,416	45	10	1,000	1,000	0,988	1,000
	5	0,199	0,355	0,229	0,306		5	0,999	1,000	0,975	1,000
	1	0,064	0,141	0,078	0,138		1	0,990	0,997	0,909	0,986
10	10	0,577	0,700	0,559	0,595	50	10	1,000	1,000	0,995	1,000
	5	0,457	0,545	0,445	0,477		5	0,999	1,000	0,984	1,000
	1	0,205	0,286	0,229	0,264		1	0,994	0,998	0,956	0,992
15	10	0,774	0,866	0,662	0,798	100	10	1,000	1,000	1,000	1,000
	5	0,671	0,778	0,542	0,682		5	1,000	1,000	1,000	1,000
	1	0,396	0,480	0,318	0,414		1	1,000	1,000	1,000	1,000
20	10	0,913	0,970	0,799	0,908	500	10	1,000	1,000	1,000	1,000
	5	0,845	0,923	0,709	0,844		5	1,000	1,000	1,000	1,000
	1	0,653	0,774	0,489	0,657		1	1,000	1,000	1,000	1,000
25	10	0,955	0,981	0,864	0,958	1.000	10	1,000	1,000	1,000	1,000
	5	0,921	0,965	0,786	0,922		5	1,000	1,000	1,000	1,000
	1	0,801	0,885	0,592	0,782		1	1,000	1,000	1,000	1,000
30	10	0,990	0,998	0,923	0,993	5.000	10	1,000	1,000	1,000	1,000
	5	0,972	0,990	0,858	0,966		5	1,000	1,000	1,000	1,000
	1	0,891	0,949	0,718	0,877		1	1,000	1,000	1,000	1,000

O teste TW mostrou ser o segundo teste mais poderoso para o $n \leq 15$ nos três níveis de significância nominais adotados. Já para $n > 15$, o teste TVB , deixa de ser um dos testes menos poderosos e passa a ser o segundo teste mais poderoso, seguido pelos testes TW , $TKRB_2$ respectivamente, nessa ordem, mostrando que dos quatro testes, o TW sempre se mantém em uma posição intermediária.

Os testes TVB e $TKRB_1$, para $n \geq 40$, considerando os níveis de significância de 10% e 5%, atingiram praticamente 100% de poder, mostrando que os mesmos detectam quase a totalidade das amostras não-normais simuladas. O teste TW , para $n \geq 45$ e níveis de significância de 10% e 5%, atingiu 100% de poder. A partir do valor $n = 100$, tanto os testes propostos TVB , $TKRB_1$ e $TKRB_2$ quanto o teste da comparação TW atingiram 100% de poder, para todos os níveis de significância nominais adotados e mostram que detectam a totalidade das amostras não-normais simuladas.

Observou-se também na Tabela 4.3, que quanto maior for o nível nominal de significância α , maior é o desempenho do poder do teste para todos casos em estudo, com a exceção em grandes amostras $n \geq 100$, que na maioria das vezes atingiram 100% de poder, igualando assim seus valores de poder. Na medida que o tamanho das amostras cresce, o desempenho dos testes também tende a crescer, como esperado pela teoria dos testes de hipóteses.

Na Tabela 4.4 apresentam-se os desempenhos dos testes avaliados em relação ao poder para os níveis de significância nominais de 10%, 5% e 1%, considerando diferentes valores de n , 3, 5, 8, 10, 15(5)50, 100, 500, 1.000 e 5.000, para a distribuição gama ($\alpha = 1$, $\beta = 3$) univariada. Verificou-se que o teste $TKRB_1$ apresentou maior poder para todos os n 's nos três níveis de significância nominais adotados, em relação a todas as alternativas comparadas (TVB , $TKRB_2$ e TW). Houve uma exceção, para o $TKRB_2$, no valor $n = 3$, mostrou-se com maior poder nos três níveis de significância nominais adotados.

O teste $TKRB_1$, foi o primeiro a atingir 100% de poder a partir do valor $n \geq 50$ no nível de significância de 10%. A partir do valor $n \geq 100$, todos os testes propostos TVB , $TKRB_1$ e $TKRB_2$ e o teste de comparação TW atingiram praticamente 100% de poder, para todos os níveis de significância nominais adotados, mostrando que detectam a totalidade das amostras não-normais simuladas. O teste $TKRB_2$ foi o mais poderoso para $n = 3$. Também foi o segundo teste mais poderoso para $n = 5$, nos três níveis de significância nominais adotados, seguido pelos testes TW , TVB respectivamente, nessa ordem. Com o $n \geq 8$, o teste TW passou a ser o segundo teste mais poderoso, sendo seguido pelos testes $TKRB_2$ e TVB , respectivamente, nessa ordem. A partir do valor $n \geq 40$, TW e TVB , apresentaram, de modo geral, desempenho em poder equivalente.

Tabela 4.4 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição gama ($\alpha = 1$, $\beta = 3$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%$, 5% e 1%

n	α	Gama (α, β)				n	α	Gama (α, β)			
		TVB	$TKRB_1$	$TKRB_2$	TW			TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,109	0,174	0,229	0,111	35	10	0,971	0,989	0,859	0,975
	5	0,061	0,078	0,119	0,060		5	0,942	0,970	0,785	0,951
	1	0,012	0,019	0,024	0,013		1	0,799	0,882	0,591	0,842
5	10	0,200	0,309	0,284	0,233	40	10	0,991	0,999	0,909	0,995
	5	0,108	0,197	0,178	0,155		5	0,981	0,994	0,858	0,982
	1	0,019	0,059	0,055	0,051		1	0,909	0,962	0,704	0,901
8	10	0,287	0,432	0,325	0,385	45	10	0,992	0,999	0,946	0,994
	5	0,183	0,304	0,220	0,280		5	0,980	0,994	0,900	0,987
	1	0,047	0,128	0,085	0,129		1	0,936	0,971	0,748	0,954
10	10	0,453	0,561	0,448	0,493	50	10	0,995	1,000	0,955	0,997
	5	0,303	0,397	0,309	0,360		5	0,987	0,998	0,915	0,990
	1	0,111	0,183	0,114	0,170		1	0,959	0,983	0,809	0,967
15	10	0,640	0,730	0,533	0,659	100	10	1,000	1,000	0,999	1,000
	5	0,509	0,607	0,428	0,541		5	1,000	1,000	0,998	1,000
	1	0,251	0,325	0,212	0,310		1	1,000	1,000	0,995	1,000
20	10	0,832	0,907	0,686	0,850	500	10	1,000	1,000	1,000	1,000
	5	0,730	0,828	0,568	0,750		5	1,000	1,000	1,000	1,000
	1	0,457	0,605	0,355	0,519		1	1,000	1,000	1,000	1,000
25	10	0,892	0,955	0,763	0,928	1.000	10	1,000	1,000	1,000	1,000
	5	0,815	0,897	0,662	0,854		5	1,000	1,000	1,000	1,000
	1	0,599	0,723	0,426	0,640		1	1,000	1,000	1,000	1,000
30	10	0,943	0,978	0,821	0,952	5.000	10	1,000	1,000	1,000	1,000
	5	0,882	0,949	0,726	0,909		5	1,000	1,000	1,000	1,000
	1	0,711	0,818	0,524	0,757		1	1,000	1,000	1,000	1,000

Na Tabela 4.5 apresentam-se os desempenhos em relação ao poder de testes TVB , TKR_1 , $TKRB_2$ e TW para os níveis de significância nominais de 10% , 5% e 1% , para diferentes valores de n , considerando a distribuição t -Student com $\nu = 1$ univariada. Observou-se que o teste $TKRB_2$ apresentou maior poder para todos os n 's nos três níveis de significância nominais adotados, em relação a todos os demais teste aplicados (TVB , $TKRB_1$ e TW), exceto em relação ao teste TW , em três casos, quais sejam, $n = 5$ e 8 , que mostrou maior poder nos três níveis de significância nominais adotados. Novamente, os valores de poder dos três testes tenderam a se igualar entre si a partir de $n = 50$, se igualando a 100% para $n \geq 100$, em todos os

níveis nominais de significância, mostrando que detectam a totalidade das amostras não-normais simuladas.

Tabela 4.5 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição t -Student com $\nu = 1$, em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

n	α	t -Student ($\nu = 1$)				n	α	t -Student ($\nu = 1$)			
		TVB	$TKRB_1$	$TKRB_2$	TW			TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,169	0,165	0,206	0,164	35	10	0,962	0,955	0,984	0,987
	5	0,103	0,100	0,126	0,105		5	0,946	0,938	0,978	0,980
	1	0,026	0,027	0,033	0,022		1	0,906	0,890	0,957	0,961
5	10	0,271	0,285	0,349	0,384	40	10	0,969	0,965	0,989	0,988
	5	0,193	0,224	0,253	0,310		5	0,960	0,956	0,984	0,983
	1	0,102	0,116	0,130	0,209		1	0,938	0,923	0,975	0,973
8	10	0,406	0,457	0,476	0,558	45	10	0,984	0,982	0,997	0,997
	5	0,337	0,375	0,400	0,498		5	0,975	0,972	0,994	0,994
	1	0,228	0,265	0,264	0,389		1	0,957	0,949	0,985	0,982
10	10	0,522	0,419	0,703	0,677	50	10	0,994	0,993	0,998	0,998
	5	0,453	0,346	0,626	0,618		5	0,992	0,986	0,997	0,998
	1	0,338	0,220	0,486	0,496		1	0,969	0,965	0,993	0,994
15	10	0,668	0,607	0,812	0,798	100	10	1,000	1,000	1,000	1,000
	5	0,604	0,524	0,746	0,748		5	1,000	1,000	1,000	1,000
	1	0,505	0,418	0,610	0,656		1	1,000	1,000	1,000	1,000
20	10	0,795	0,755	0,913	0,895	500	10	1,000	1,000	1,000	1,000
	5	0,753	0,696	0,882	0,880		5	1,000	1,000	1,000	1,000
	1	0,644	0,579	0,814	0,804		1	1,000	1,000	1,000	1,000
25	10	0,886	0,856	0,953	0,941	1.000	10	1,000	1,000	1,000	1,000
	5	0,854	0,820	0,938	0,916		5	1,000	1,000	1,000	1,000
	1	0,778	0,733	0,880	0,876		1	1,000	1,000	1,000	1,000
30	10	0,940	0,919	0,970	0,969	5.000	10	1,000	1,000	1,000	1,000
	5	0,912	0,896	0,962	0,960		5	1,000	1,000	1,000	1,000
	1	0,854	0,830	0,933	0,937		1	1,000	1,001	1,000	1,000

Na Tabela 4.6 mostram-se os desempenhos em relação ao poder dos testes para os níveis de significância nominais de 10%, 5% e 1%, para diferentes valores de n , 3, 5, 8, 10, 15(5)50, 100, 500, 1.000 e 5.000, considerando a distribuição não-normal t -Student com $\nu = 30$ grau de liberdade univariada. Notou-se que o $TKRB_2$ apresentou maior poder para todo n nos três níveis de significância nominais adotados, em relação a todas as alternativas comparadas (TVB , $TKRB_1$ e TW), exceto em alguns casos, em que perdeu em desempenho para o TVB e o TW . Convém salientar que esta distribuição se assemelha muito à distribuição normal e, portanto, os

valores de poder dos testes são muito baixos, sendo praticamente idênticos aos valores nominais de significância para $n < 30$, exceto pelo erro Monte Carlo. Somente o $TKRB_2$ e o TW é que superaram significativamente os níveis nominais de significância a partir de $n = 30$. Os valores de poder para estes dois testes só mostram resultados expressivos a partir de $n = 5.000$, para $\alpha \geq 5\%$.

Tabela 4.6 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição t -Student com $\nu = 30$, em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

n	α	t -Student ($\nu = 30$)				n	α	t -Student ($\nu = 30$)			
		TVB	$TKRB_1$	$TKRB_2$	TW			TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,102	0,099	0,147	0,102	35	10	0,101	0,099	0,140	0,139
	5	0,050	0,045	0,063	0,051		5	0,053	0,050	0,080	0,085
	1	0,012	0,008	0,011	0,010		1	0,009	0,007	0,021	0,020
5	10	0,115	0,114	0,120	0,103	40	10	0,088	0,093	0,141	0,122
	5	0,065	0,055	0,062	0,056		5	0,044	0,048	0,077	0,069
	1	0,020	0,011	0,010	0,010		1	0,008	0,011	0,022	0,022
8	10	0,105	0,098	0,109	0,107	45	10	0,074	0,073	0,122	0,121
	5	0,047	0,055	0,061	0,056		5	0,032	0,034	0,075	0,070
	1	0,013	0,010	0,012	0,009		1	0,006	0,006	0,015	0,014
10	10	0,090	0,090	0,110	0,105	50	10	0,099	0,088	0,126	0,136
	5	0,041	0,043	0,058	0,057		5	0,046	0,041	0,080	0,077
	1	0,012	0,014	0,016	0,011		1	0,009	0,010	0,016	0,015
15	10	0,092	0,085	0,120	0,118	100	10	0,086	0,080	0,149	0,129
	5	0,042	0,047	0,059	0,064		5	0,041	0,039	0,087	0,070
	1	0,011	0,006	0,012	0,011		1	0,007	0,008	0,021	0,020
20	10	0,092	0,086	0,108	0,105	500	10	0,065	0,066	0,258	0,224
	5	0,046	0,046	0,054	0,050		5	0,031	0,032	0,162	0,143
	1	0,012	0,013	0,011	0,012		1	0,006	0,006	0,066	0,055
25	10	0,102	0,095	0,127	0,104	1.000	10	0,055	0,057	0,358	0,300
	5	0,050	0,050	0,066	0,058		5	0,026	0,027	0,253	0,211
	1	0,011	0,011	0,019	0,015		1	0,006	0,006	0,097	0,086
30	10	0,101	0,097	0,134	0,125	5.000	10	0,060	0,059	0,733	0,747
	5	0,047	0,061	0,079	0,070		5	0,029	0,029	0,618	0,650
	1	0,011	0,008	0,023	0,013		1	0,005	0,005	0,358	0,425

O teste de Shapiro-Wilk (TW) da Tabela 4.6, para $n = 5.000$, apresentou maior poder nos três níveis de significância nominais adotados. Nessa configuração o teste $TKRB_2$ apresentou um poder um pouco inferior ao do Shapiro-Wilk (TW) para todos os níveis de significância. Nenhum teste atingiu 100% de poder que amostras muito grandes seriam necessárias para se

obter rejeições da totalidade das amostras simuladas, em todos os níveis de significância nominais adotados. A razão disso, como já mencionado, é atribuída à semelhança entre a distribuição considerada e a distribuição normal.

Na Tabela 4.7 os desempenhos em relação ao poder do TVB , $TKRB_1$, $TKRB_2$ e TW são apresentados, para os níveis de significância nominais de 10%, 5% e 1%, considerando diferentes valores de n , 3, 5, 8, 10, 15(5)50, 100, 500, 1.000 e 5.000 sob a distribuição não-normal uniforme ($a = 0, b = 1$) univariada. Verificou-se que o TVB apresenta maior poder para todos os n 's nos três níveis de significância nominais adotados, em relação aos demais testes ($TKRB_1$, $TKRB_2$ e TW). Houve uma exceção que ocorreu para o $TKRB_2$ com $n = 3$, em que este se mostrou mais poderoso nos três níveis de significância nominais adotados. O $TKRB_1$ mostrou-se o segundo mais poderoso para o $n \geq 5$ nos três níveis de significância nominais adotados, sendo seguido pelo TW e $TKRB_2$, respectivamente. Assim, os testes TVB e $TKRB_1$ em relação a todos os n 's apresentaram maior poder comparado com o poder do teste de Shapiro-Wilk (TW).

Os testes TVB e $TKRB_1$ foram os primeiros a atingir 100% de poder a partir do valor $n \geq 100$ para os níveis de significância nominais $\alpha = 10\%$ e 5% . À medida que o tamanho das amostras cresce, o desempenho de ambos os testes tende a se igualar, sendo que, a partir do valor $n \geq 500$, todos os testes propostos TVB , $TKRB_1$ e $TKRB_2$ e o teste de comparação TW atingiram 100% de poder, para todos os níveis de significância nominais adotados, mostrando que detectam a totalidade das amostras não-normais simuladas.

Na Tabela 4.8 apresentam-se os desempenhos em relação ao poder dos testes TVB , $TKRB_1$, $TKRB_2$ e TW , para os níveis de significância nominais de 10%, 5% e 1%, com diferentes valores de n , 3, 5, 8, 10, 15(5)50, 100, 500, 1.000 e 5.000, considerando a distribuição não-normal log-normal ($\mu = 0, \sigma = 1$) univariada. Verificou-se que o teste $TKRB_1$ apresenta maior poder para todo n nos três níveis de significância nominais adotados, em relação a todas os testes alternativos ($TKRB_1$, $TKRB_2$ e TW), exceto em relação ao TW para $\alpha = 1\%$ e $n \geq 15$.

O teste TW mostrou-se como o segundo teste mais poderoso em geral, mas em alguns casos, permutou o desempenho com o TVB , para tamanhos de amostras pequenas, sendo seguidos pelo teste $TKRB_2$. A partir do $n \geq 35$ todos os testes em estudos tendem a igualarem os seus valores de poder, aproximando de 100%. À medida que o tamanho da amostra cresce, o desempenho de todos os testes tende a crescer e a se igualar, sendo que, a partir do valor $n \geq 50$ atingiram praticamente 100% de poder, para todos os níveis de significância nominais adotados.

Tabela 4.7 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição uniforme ($a = 0, b = 1$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

n	α	Uniforme (a, b)				n	α	Uniforme (a, b)			
		TVB	$TKRB_1$	$TKRB_2$	TW			TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,127	0,116	0,157	0,126	35	10	0,832	0,791	0,173	0,667
	5	0,065	0,052	0,085	0,066		5	0,725	0,660	0,112	0,488
	1	0,013	0,015	0,019	0,014		1	0,462	0,393	0,026	0,131
5	10	0,151	0,146	0,116	0,134	40	10	0,905	0,884	0,140	0,753
	5	0,088	0,065	0,055	0,073		5	0,826	0,791	0,072	0,584
	1	0,025	0,016	0,010	0,008		1	0,612	0,542	0,017	0,213
8	10	0,220	0,167	0,109	0,146	45	10	0,965	0,952	0,176	0,844
	5	0,126	0,088	0,055	0,061		5	0,916	0,891	0,097	0,686
	1	0,035	0,018	0,014	0,009		1	0,707	0,644	0,025	0,264
10	10	0,281	0,215	0,072	0,176	50	10	0,965	0,957	0,197	0,885
	5	0,177	0,119	0,033	0,073		5	0,931	0,915	0,132	0,743
	1	0,043	0,026	0,005	0,004		1	0,754	0,707	0,037	0,319
15	10	0,435	0,355	0,095	0,264	100	10	1,000	1,000	0,581	0,998
	5	0,294	0,232	0,050	0,123		5	1,000	1,000	0,466	0,994
	1	0,096	0,072	0,010	0,015		1	0,995	0,992	0,230	0,936
20	10	0,601	0,537	0,071	0,368	500	10	1,000	1,000	1,000	1,000
	5	0,446	0,381	0,028	0,222		5	1,000	1,000	1,000	1,000
	1	0,199	0,138	0,009	0,031		1	1,000	1,000	1,000	1,000
25	10	0,696	0,650	0,085	0,477	1.000	10	1,000	1,000	1,000	1,000
	5	0,577	0,496	0,049	0,298		5	1,000	1,000	1,000	1,000
	1	0,282	0,227	0,009	0,052		1	1,000	1,000	1,000	1,000
30	10	0,778	0,741	0,143	0,597	5.000	10	1,000	1,000	1,000	1,000
	5	0,657	0,602	0,077	0,403		5	1,000	1,000	1,000	1,000
	1	0,402	0,331	0,018	0,099		1	1,000	1,000	1,000	1,000

Na Tabela 4.9, os desempenhos em relação ao poder de testes TVB , TKR_1 , $TKRB_2$ e TW para os níveis de significância nominais de 10%, 5% e 1% e diferentes valores de n , considerando a distribuição Weibull ($\delta = 2, \alpha = 3$) univariada, são apresentados. Observou-se que o teste $TKRB_1$ apresentou maior poder nos três níveis de significância nominais adotados para $n \leq 20$, em relação a todas os testes comparados (TVB , $TKRB_2$ e TW). Para $n = 3$, os testes TVB e $TKRB_1$ apresentaram valores de poder iguais e para $n = 5$ as três propostas TVB , $TKRB_1$ e $TKRB_2$ apresentaram maior poder em relação ao teste do Shapiro-Wilk (TW) nos três níveis de significância nominais adotados.

Tabela 4.8 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição log-normal ($\mu = 0, \sigma = 1$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

		<i>Log – Normal (μ, σ)</i>						<i>Log – Normal (μ, σ)</i>			
<i>n</i>	α	<i>TVB</i>	<i>TKRB₁</i>	<i>TKRB₂</i>	<i>TW</i>	<i>n</i>	α	<i>TVB</i>	<i>TKRB₁</i>	<i>TKRB₂</i>	<i>TW</i>
3	10	0,181	0,184	0,141	0,183	35	10	0,998	0,998	0,982	0,998
	5	0,097	0,097	0,074	0,094		5	0,997	0,997	0,961	0,997
	1	0,013	0,013	0,012	0,017		1	0,971	0,978	0,907	0,988
5	10	0,373	0,479	0,474	0,348	40	10	0,999	0,999	0,992	1,000
	5	0,244	0,359	0,341	0,258		5	0,997	0,998	0,984	0,999
	1	0,115	0,148	0,165	0,121		1	0,984	0,991	0,939	0,996
8	10	0,573	0,710	0,591	0,608	45	10	1,000	1,000	0,998	1,000
	5	0,446	0,609	0,501	0,509		5	0,998	1,000	0,994	1,000
	1	0,237	0,368	0,287	0,296		1	0,995	0,998	0,971	0,999
10	10	0,669	0,752	0,655	0,678	50	10	1,000	1,000	0,999	1,000
	5	0,554	0,626	0,558	0,590		5	1,000	1,000	0,997	1,000
	1	0,349	0,417	0,367	0,413		1	1,000	1,000	0,983	1,000
15	10	0,876	0,919	0,814	0,900	100	10	1,000	1,000	1,000	1,000
	5	0,795	0,855	0,736	0,831		5	1,000	1,000	1,000	1,000
	1	0,579	0,636	0,543	0,654		1	1,000	1,000	1,000	1,000
20	10	0,944	0,966	0,883	0,954	500	10	1,000	1,000	1,000	1,000
	5	0,893	0,927	0,839	0,925		5	1,000	1,000	1,000	1,000
	1	0,757	0,795	0,685	0,831		1	1,000	1,000	1,000	1,000
25	10	0,980	0,992	0,947	0,990	1.000	10	1,000	1,000	1,000	1,000
	5	0,961	0,978	0,907	0,983		5	1,000	1,000	1,000	1,000
	1	0,876	0,902	0,793	0,929		1	1,000	1,000	1,000	1,000
30	10	0,993	0,992	0,972	0,996	5.000	10	1,000	1,000	1,000	1,000
	5	0,985	0,990	0,951	0,995		5	1,000	1,000	1,000	1,000
	1	0,940	0,955	0,869	0,966		1	1,000	1,000	1,000	1,000

Ainda considerando a Tabela 4.9, o TW , para o n entre 30 e 500, apresentou maior poder nos três níveis de significância. À medida que o tamanho das amostras cresce, o desempenho de ambos os testes tende a se igualar, sendo que o teste (TW) atingiu primeiro o valor de poder de 100% a partir do $n \geq 500$. Nessa distribuição de Weibull ($\delta = 2, \alpha = 3$), os testes propostos mostraram não apresentar diferenças expressiva para tamanhos de amostras muito grande $n \geq 500$. Deve-se salientar que à medida que os tamanhos de amostras aumentam, os valores de poder tendem a aumentar para todos os testes, e partir do valor $n = 5.000$, todos os testes proposto TVB , $TKRB_1$, $TKRB_2$ e o teste de comparação TW atingiram 100% de poder, para todos os níveis de significância nominais adotados.

Tabela 4.9 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição Weibull ($\delta = 2$, $\alpha = 3$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

		Weibull (δ, α)						Weibull (δ, α)			
n	α	TVB	$TKRB_1$	$TKRB_2$	TW	n	α	TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,110	0,110	0,056	0,107	35	10	0,306	0,387	0,236	0,421
	5	0,062	0,062	0,034	0,060		5	0,176	0,235	0,130	0,280
	1	0,013	0,013	0,007	0,013		1	0,049	0,065	0,028	0,098
5	10	0,123	0,164	0,184	0,117	40	10	0,322	0,398	0,233	0,471
	5	0,062	0,097	0,093	0,056		5	0,198	0,256	0,147	0,320
	1	0,014	0,017	0,017	0,008		1	0,069	0,081	0,040	0,131
8	10	0,103	0,173	0,146	0,105	45	10	0,353	0,459	0,258	0,504
	5	0,055	0,083	0,065	0,051		5	0,227	0,303	0,153	0,388
	1	0,004	0,018	0,013	0,013		1	0,074	0,097	0,039	0,154
10	10	0,165	0,218	0,159	0,154	50	10	0,347	0,468	0,240	0,560
	5	0,091	0,116	0,090	0,095		5	0,223	0,280	0,140	0,412
	1	0,021	0,039	0,019	0,024		1	0,080	0,094	0,049	0,157
15	10	0,198	0,250	0,181	0,219	100	10	0,546	0,642	0,355	0,870
	5	0,112	0,146	0,101	0,131		5	0,410	0,492	0,228	0,769
	1	0,023	0,037	0,027	0,032		1	0,189	0,221	0,091	0,531
20	10	0,205	0,284	0,193	0,243	500	10	0,983	0,990	0,830	1,000
	5	0,125	0,154	0,103	0,148		5	0,960	0,976	0,737	1,000
	1	0,030	0,049	0,021	0,039		1	0,862	0,894	0,499	1,000
25	10	0,259	0,329	0,190	0,293	1.000	10	1,000	1,000	0,977	1,000
	5	0,156	0,192	0,124	0,202		5	0,999	0,999	0,951	1,000
	1	0,040	0,061	0,028	0,064		1	0,989	0,995	0,839	1,000
30	10	0,281	0,358	0,196	0,383	5.000	10	1,000	1,000	1,000	1,000
	5	0,172	0,214	0,115	0,254		5	1,000	1,000	1,000	1,000
	1	0,044	0,065	0,037	0,084		1	1,000	1,000	1,000	1,000

Na Tabela 4.10 apresentam-se os desempenhos dos testes avaliados em relação ao poder para os níveis de significância nominais de 10%, 5% e 1%, com diferentes valores de n , 3, 5, 8, 10, 15(5)50, 100, 500, 1.000 e 5.000, considerando a distribuição Beta ($\alpha = 0,2$, $\beta = 0,5$) univariada. Verificou-se que o teste de normalidade $TKRB_1$ apresentou maior poder para todos os n 's nos três níveis de significância nominais adotados, em relação aos demais testes TVB , $TKRB_2$ e TW . Ocorre uma exceção apenas no caso $n = 3$, em que os testes TVB , $TKRB_1$ e TW apresentaram valores de poder praticamente iguais nos três níveis de significância nominais.

Os testes $TKRB_1$ e TVB foram mais poderosos que o teste TW . A partir do valor $n \geq 20$, todos os testes propostos TVB , $TKRB_1$ e $TKRB_2$ e o teste de comparação TW atingiram 100%

de poder, para todos os níveis de significância nominais adotados, mostrando que detectam o total das amostras não-normais simuladas em todas as alternativas. Exceto o teste $TKRB_2$ atingiu 100% de poder, a partir de $n \geq 30$.

O teste TVB foi o segundo mais poderoso para todos os n 's nos três níveis de significância nominais, seguido pelos testes TW , $TKRB_2$ respectivamente, nessa ordem. Os últimos dois testes apresentaram, de modo geral, valores de poder quase equivalentes.

Tabela 4.10 – Poder dos testes univariados de normalidade baseados na entropia TVB , $TKRB_1$, $TKRB_2$ e TW , relativo à distribuição beta ($\alpha = 0,2$, $\beta = 0,5$), em função de $n = 3, 5, 8, 10, 15(5)50, 100, 500, 1.000, 5.000$ para os níveis nominais de significância $\alpha = 10\%, 5\%$ e 1%

n	α	$Beta(\alpha, \beta)$				n	α	$Beta(\alpha, \beta)$			
		TVB	$TKRB_1$	$TKRB_2$	TW			TVB	$TKRB_1$	$TKRB_2$	TW
3	10	0,352	0,352	0,296	0,354	35	10	1,000	1,000	1,000	1,000
	5	0,266	0,266	0,232	0,263		5	1,000	1,000	1,000	1,000
	1	0,128	0,128	0,119	0,129		1	1,000	1,000	1,000	1,000
5	10	0,566	0,606	0,557	0,552	40	10	1,000	1,000	1,000	1,000
	5	0,468	0,527	0,468	0,427		5	1,000	1,000	1,000	1,000
	1	0,328	0,368	0,331	0,216		1	1,000	1,000	1,000	1,000
8	10	0,871	0,888	0,732	0,798	45	10	1,000	1,000	1,000	1,000
	5	0,818	0,845	0,664	0,681		5	1,000	1,000	1,000	1,000
	1	0,675	0,696	0,540	0,417		1	1,000	1,000	1,000	1,000
10	10	0,962	0,976	0,860	0,888	50	10	1,000	1,000	1,000	1,000
	5	0,930	0,955	0,803	0,800		5	1,000	1,000	1,000	1,000
	1	0,836	0,894	0,681	0,572		1	1,000	1,000	1,000	1,000
15	10	0,997	0,996	0,955	0,979	100	10	1,000	1,000	1,000	1,000
	5	0,993	0,994	0,930	0,949		5	1,000	1,000	1,000	1,000
	1	0,977	0,977	0,879	0,799		1	1,000	1,000	1,000	1,000
20	10	1,000	1,000	0,993	1,000	500	10	1,000	1,000	1,000	1,000
	5	1,000	1,000	0,990	0,993		5	1,000	1,000	1,000	1,000
	1	0,998	0,999	0,962	0,937		1	1,000	1,000	1,000	1,000
25	10	1,000	1,000	0,998	1,000	1.000	10	1,000	1,000	1,000	1,000
	5	1,000	1,000	0,996	1,000		5	1,000	1,000	1,000	1,000
	1	1,000	0,999	0,991	0,989		1	1,000	1,000	1,000	1,000
30	10	1,000	1,000	1,000	1,000	5.000	10	1,000	1,000	1,000	1,000
	5	1,000	1,000	1,000	1,000		5	1,000	1,000	1,000	1,000
	1	1,000	1,000	1,000	0,999		1	1,000	1,000	1,000	1,000

Se forem consideradas todas as situações, de forma geral percebe-se que houve bastante oscilações nos desempenhos dos testes em relação ao poder. Pode-se dizer que não houve um teste uniformemente mais poderoso para todas as hipóteses alternativas estudadas, como

já vem sendo destacado na literatura. Nos estudos comparativos de testes de normalidade univariada, os autores (VASICEK, 1976; LIN; MUDHOLKAR, 1980; JARQUE; BERA, 1987; ARIZONO; OHTA, 1989; GUO; ALEMAYEHU; SHAO, 2010; FILHO; FERREIRA, 2012), verificaram que nenhum dos testes estudados foi uniformemente mais poderoso nas diferentes hipóteses alternativas (distribuições) e tamanhos amostrais utilizados. Além disso, tais testes foram avaliados em situações com limitações dos tamanhos amostrais, que, em geral, apresentaram n igual a 10(10)50, exceto em (JARQUE; BERA, 1987), cujos tamanhos amostrais n foram 20, 35, 50 100, 200, 300 e em (FILHO; FERREIRA, 2012), cujos valores foram 5, 10, 30, 100, 500, 5.000, 10.000. Neste trabalho, isso, correu tanto ao longo dos diferentes tamanhos amostrais quanto em relação aos diferentes níveis nominais de significância adotados. Por outro lado, houve controle perfeito do erro tipo I de todos os testes, sendo que a única exceção foi atribuída a um erro tipo I em virtude da multiplicidade de testes aplicados na avaliação do desempenho, sem que houvesse nenhuma medida de controle adotada. Por essa razão o desempenho dos testes deve focar no poder.

Nas oito hipóteses alternativas, algum dos testes *bootstrap* paramétricos baseados na entropia esteve entre o de melhor desempenho em todos os casos. O TW apresentou melhor desempenho somente para a distribuição Weibull com $n \geq 30$. O teste TVB foi o mais poderoso para a distribuição uniforme. Já o $TKRB_1$ foi mais poderoso em quatro distribuições, sendo que elas são essencialmente distribuições assimétricas à direita. Já o teste $TKRB_2$, foi mais poderoso em 2 casos, sendo representados essencialmente por distribuições simétricas, como é o caso da t , que se assemelha consideravelmente com a normal, principalmente com o aumento dos graus de liberdade.

Jarque e Bera (1987), propuseram o teste (LM), comparando com seis alternativas: skewness ($\sqrt{b_1}$ e b_2), D'Agostino (D^*), Pearson(R), Shapiro-Francia (W'), Shapiro-Wilk (W) e score (LM), com tamanhos amostrais $n = 20, 50, 100, 200, 300$, $\alpha=10\%$ e 5% , em 4 distribuições, log-normal, gama(2,1), beta(3,2) e $t(5)$. Observaram que os testes W , LM e W' respectivamente são preferenciais. Constataram que, para amostras grandes, $n \geq 100$, o teste LM teve o melhor desempenho em relação ao poder para todas as distribuições, com diferenças não expressivos em relação aos testes W e W' . Os autores avançaram ainda, afirmando que, a partir de tamanhos amostrais $n \geq 200$ todos os testes em estudos tenderam a se igualarem em seus valores de poder, aproximando-se de 100%. Por fim concluíram, não haver um teste uniformemente mais poderoso em relação a diferentes tamanhos amostrais e distribuições.

Confrontado os resultados obtidos neste trabalho com os de Jarque e Bera (1987) tem-se que os testes $TKRB_1$ e $TKRB_2$, apresentam maior desempenho do poder, e com tamanhos amostrais $n \geq 20$, tendem a igualarem os seus valores de poder, aproximando de 100%.

O teste TW , ficou sempre entre os melhores, assumindo a segunda posição em desempenho de poder para uma ampla gama de distribuições alternativas estudadas. Como os testes $TKRB_1$ e $TKRB_2$, na maioria dos casos estudados, foram os que apresentaram o melhor desempenho, optou-se por fazer um teste combinado, cujos resultados são apresentados na sequência. A ideia foi realizar uma combinação linear entre as duas estatísticas, seguidas da geração da distribuição nula *bootstrap*. Por tentativa e erro, avaliando-se o poder, verificou-se empiricamente que a média aritmética das estatísticas dos testes $TKRB_1$ e $TKRB_2$ foi a combinação que deu melhor desempenho em poder.

Verificou-se que, em todas hipóteses alternativas ou distribuições não-normais, quanto maior for o nível nominal α de significância, maior é o desempenho do poder do teste para todos casos em estudo, com a exceção, na distribuição t-Student com ($\nu = 1$), $n = 50$, apresentou poder igual (0,998) nos α 10% e 5%. Como também, em grandes amostras $n \geq 100$, que na maioria das vezes atingem 100% de poder, igualando assim seus valores de poder.

4.3 Desempenho do teste combinado

Nesta seção é apresentado o desempenho do teste combinado, sendo que ele foi dividida em duas subseções, uma para o estudo do erro tipo I do teste combinado $TKRB_C$ e outra, para o poder. Foram considerados os níveis de significância nominais de 10%, 5% e 1% e valores de n de 3, 10, 30, 50, 100, 500, 1.000, 5.000. As comparações foram feitas com os testes apresentados na seção anterior para os mesmos valores do nível nominal de significância e de n .

Para análise do erro tipo I do teste combinado $TKRB_C$, foram geradas amostras, via simulação Monte Carlo, de tamanho n sob normalidade. As taxas de erro tipo I empíricas foram computadas ao final de N simulações Monte Carlo pela proporção de rejeições indevidas, cujas definições e propriedades foram apresentadas na Seção (2.3).

Na segunda subseção foi avaliado o poder sob a hipótese alternativa de não normalidade dos dados. Assim, foram consideradas amostras de oito distribuições não-normais já utilizadas na seção anterior, cujas respectivas abreviaturas estão apresentadas na Tabela 4.11.

Tabela 4.11 – Distribuições não-normais, com suas respectivas abreviaturas e parâmetros adotados

Distribuição	Abreviatura	Distribuição	Abreviatura
<i>Exponencial</i> ($\lambda = 1$)	<i>Exp</i> (1)	<i>Uniforme</i> ($a = 0, b = 1$)	<i>Unif</i> (0,1)
<i>Gama</i> ($\alpha = 1, \beta = 3$)	<i>Gama</i> (1,3)	<i>Log – Normal</i> ($\mu = 0, \sigma = 1$)	<i>LogN</i> (0,1)
<i>t-Student</i> ($v = 1$)	<i>t</i> (1)	<i>Weibull</i> ($\delta = 2, \alpha = 3$)	<i>Weib</i> (2,3)
<i>t-Student</i> ($v = 30$)	<i>t</i> (30)	<i>Beta</i> ($\alpha = 0,2; \beta = 0,5$)	<i>Beta</i> (0,2;0,5)

As mesmas configurações, usadas nos testes anteriormente apresentados, em relação aos valores de α , n , B e N foram consideradas nessa etapa de avaliação de desempenho. As comparações foram feitas considerando os correspondentes casos anteriores, sendo o TW , considerado teste de referência. Deste modo, para se comparar o poder do teste combinado, $TKRB_C$, calculou-se o percentual de rejeições da hipótese nula, quando amostras das oito distribuições não-normais: $Exp(1)$, $Gama(1,3)$, $t(1)$, $t(30)$, $Unif(0,1)$, $LogN(0,1)$, $Weib(2,3)$ e $Beta(0,2;0,5)$, foram simuladas (Tabela 4.11). O teste que detectou uma maior quantidade de amostras não-normais (maior proporção de rejeições) foi considerado o mais poderoso.

4.3.1 Resultados do erro tipo I do teste combinado

Na presente subseção, são apresentados os resultados do erro tipo I do teste combinado $TKRB_C$, conforme Tabela 4.12, considerando os níveis de significância nominais de 10%, 5% e 1% e diferentes valores de n , quais sejam, $n = 3, 10, 30, 50, 100, 500, 1.000, 5.000$. Em todos estes casos, os dados foram simulados considerando a distribuição normal univariada. Os valores das taxas de erros empíricas foram testadas pelo teste binomial exato para verificar se os seus valores diferiam significativamente ($P < 0,01$) dos níveis nominais fixados.

Considerando-se o nível de significância de 10%, as taxas de erro tipo I (Tabela 4.12) do teste combinado $TKRB_C$ não foram significativamente diferentes ($P > 0,01$) do valor nominal $\alpha = 10\%$ em todos os valores de n adotados. Desse modo, o quarto teste proposto ($TKRB_C$) pode ser considerado exato ao nível de significância de 10%, dado que a sua taxa de erro tipo I não foi significativamente diferentes do nível nominal de 10% ($P > 0,01$). Pode-se dizer que esse teste combinado ($TKRB_C$) apresentou controle adequado das taxas de erro tipo I.

Considerando os demais níveis nominais de significância, $\alpha = 5\%$ e 1%, verificou-se que as taxas de erro tipo I empíricas foram também não significativamente ($P > 0,01$) diferentes dos níveis nominais em todos os casos sem exceção, considerando, ainda, as mesmas situações

Tabela 4.12 – Erro tipo I de teste combinado $TKRB_C$ de normalidade univariada baseado na entropia em função de n (3, 10, 30, 50, 100, 500, 1.000, 5.000), para os níveis nominais de significância $\alpha = 10\%$, 5% e 1%

<i>Teste Normal $TKRB_C$</i>			
<i>n</i>	<i>α em %</i>		
	10	5	1
3	0,078	0,041	0,007
10	0,121	0,061	0,009
30	0,101	0,043	0,009
50	0,091	0,052	0,011
100	0,109	0,052	0,011
500	0,093	0,054	0,017
1.000	0,092	0,054	0,013
5.000	0,113	0,056	0,021

descritas anteriormente para $\alpha = 10\%$. O mesmo ocorreu para o TW , usado como teste referência, conforme apresentado na Seção 4.1 e para os demais testes propostos. Assim, como o $TKRB_C$ apresentou taxas de erro tipo I observadas não significativamente diferentes ($P > 0,01$) do nível nominal α de 10% , 5% e 1% , para todos os n adotados, ele pode ser considerado um teste exato, como aconteceu para todos os outros testes.

4.3.2 Resultados do poder do teste combinado ($TKRB_C$)

Nessa subseção, são apresentados os resultados referente a poder do teste combinado $TKRB_C$ nas Tabelas 4.13, 4.14 e 4.15, para as oito distribuições não-normais consideradas: $\text{Exp}(1)$, $\text{Gama}(1,3)$, $t(1)$, $t(30)$, $\text{Unif}(0,1)$, $\text{LogN}(0,1)$, $\text{Weib}(2,3)$ e $\text{Beta}(0,2;0,5)$. Também foram considerados os níveis de significância nominais de 10% , 5% e 1% e diferentes valores de n , quais sejam, 3, 10, 30, 50, 100, 500, 1.000, 5.000. Discutiu-se os resultados para cada nível de significância separadamente. Na primeira etapa avaliou-se o desempenho em relação ao $\alpha = 10\%$, na segunda, para $\alpha = 5\%$ e na terceira, para $\alpha = 1\%$, considerando todas as oito distribuições não-normais conjuntamente.

Na Tabela 4.13 apresenta-se o poder de teste combinado $TKRB_C$ para o nível de significância nominal de $\alpha = 10\%$, nas oito distribuições não-normais univariadas. Observou-se para a distribuição $\text{Exp}(1)$, o $TKRB_C$ apresenta maior poder do que teste de Shapiro-Wilk (TW) para o valor de $n \leq 30$. Em amostras de tamanho n entre [30 e 50], o TW foi mais poderoso, sendo

que esta superioridade foi de um valor em torno de 1 ponto percentual. À medida que o tamanho das amostras cresce, o desempenho de ambos os testes tende a se igualar, sendo que ambos os testes, atingiram 100% de poder em amostras a partir de $n \geq 100$. É conveniente lembrar que o teste $TKRB_1$, em relação a todas alternativas, nessa distribuição, foi o mais poderoso, inclusive sendo mais poderoso que o teste combinado.

Tabela 4.13 – Poder do teste combinado $TKRB_C$ de normalidade, relativo à oito distribuições não normais univariadas, em função de $n=3, 10, 30, 50, 100, 500, 1.000, 5.000$, para o nível nominal de significância $\alpha=10\%$

<i>Teste combinado $TKRB_C$ $\alpha = 10\%$</i>								
n	<i>Exp(1)</i>	<i>Gama(1,3)</i>	<i>t(1)</i>	<i>t(30)</i>	<i>Unif(0,1)</i>	<i>LogN(0,1)</i>	<i>Weib(2,3)</i>	<i>Beta(0,2;0,5)</i>
3	0,143	0,113	0,096	0,104	0,072	0,177	0,069	0,355
10	0,683	0,569	0,600	0,126	0,147	0,753	0,239	0,985
30	0,992	0,954	0,937	0,094	0,458	0,997	0,319	1,000
50	0,999	0,985	0,996	0,117	0,682	1,000	0,369	1,000
100	1,000	1,000	1,000	0,104	0,960	1,000	0,522	1,000
500	1,000	1,000	1,000	0,115	1,000	1,000	0,946	1,000
1.000	1,000	1,000	1,000	0,094	1,000	1,000	0,996	1,000
5.000	1,000	1,000	1,000	0,710	1,000	1,000	1,000	1,000

Considerando a distribuição $Gama(1,3)$, o $TKRB_C$ apresentou maior poder em relação ao teste TW para o tamanho amostral n adotado. Deve-se salientar que o $TKRB_C$, para amostra $n = 10$, foi mais poderoso em relação a todas as alternativas em estudo. Na amostra $n = 50$, o teste TW apresentou maior poder, sendo que esta superioridade foi de um valor em torno de 2 pontos percentuais. Ambos os testes, atingiram 100% de poder em amostras a partir de $n \geq 100$. Novamente, o $TKRB_1$ foi o mais poderoso em relação a todas alternativas, incluindo o teste combinado.

Os desempenhos em relação ao poder, considerando a distribuição $t(1)$, $\alpha = 10\%$, mostram que o $TKRB_C$, apresentou menor poder em relação ao teste TW , para amostras $n < 100$, mas com uma diferença, em média, não superior a 3 pontos percentuais. Todos os testes propostos e o teste de comparação atingiram 100% de poder em amostras a partir de $n \geq 100$. Neste caso, o $TKRB_2$ foi o mais poderoso entre todas os testes estudados. Para a distribuição $t(30)$, nenhum dos testes em estudo atingiu poder de 100%, haja vista que isso já era esperado, por esta distribuição ser simétrica e se aproxima consideravelmente da distribuição normal. Constatou-se que o $TKRB_C$ apresentou como praticamente todas os demais testes baixo poder para amostras pequenas de $n \leq 50$. Este teste só foi competitivo com amostras de tamanho

5.000, em que o valor de poder aproximou-se dos valores de poder de TW e $TKRB_2$. Sendo que o $TKRB_2$ mostrou ser mais poderoso para este caso também.

Filho e Ferreira (2012) analisaram o desempenho do poder do teste para as distribuições $t(1)$ e $t(30)$ com nível de significância $\alpha = 10\%$, com tamanhos amostrais $n = 5, 10, 30, 100, 500, 5.000, 10.000$ e observaram que, na $t(1)$, o teste Monte Carlo de normalidade univariado baseado em distâncias proposto por eles, denotado $TMCBD$, apresentou menor poder em amostras pequenas com relação ao teste de Shapiro-Wilk ($TNUSW$), ao passo que, na $t(30)$, ($TMCBD$) apresentou maior desempenho do poder em relação ao $TNUSW$. Confrontando com os resultados do presente trabalho, em $t(1)$ e $t(30)$, o $TKRB_2$ apresentou maior desempenho nas duas distribuições, em relação ao teste de Shapiro-Wilk (TW)

Observa-se, para a distribuição $Unif(0,1)$ com $\alpha = 10\%$, que o teste combinado $TKRB_C$, apresenta menor poder em relação ao TW e ao $TKRB_1$, considerado melhor teste neste caso, para amostras com $n < 500$. À medida que o tamanho das amostras cresce, o desempenho dos três testes mencionados tende a se igualar. Os testes TVB e $TKRB_1$ foram os mais poderosos em relação a todas alternativas, nessa distribuição. Na distribuição $LogN(0,1)$, o $TKRB_C$, apresenta poder semelhante ao poder do teste $TKRB_1$ para todo n adotado. À medida que o tamanho das amostras cresce, o desempenho de ambos os testes tende a se igualar, sendo que ambos os testes, atingiram 100% de poder em amostras $n \geq 50$. Assim, o $TKRB_C$ e o $TKRB_1$ apresentaram maior poder em relação ao TW .

O teste combinado $TKRB_C$ para a distribuição $Weib(2,3)$, com $\alpha = 10\%$, para amostra $n = 10$ apresentou valores de poder elevado, ganhando em desempenho do TW para $n < 30$ e perdendo para ele, para $n \geq 30$. O $TKRB_C$ alcançou 100% de poder em amostras $n = 5.000$ e o TW , com amostra $n \geq 500$. Na distribuição $Beta(0,2;0,5)$ com $\alpha = 10\%$, o $TKRB_C$ apresentou poder maior ou igual ao poder do TW , para todo n . Ainda, o $TKRB_C$, para amostras $n \leq 10$, apresentou maior poder em relação a todos testes em estudo. Os testes, atingiram 100% de poder em amostras $n \geq 30$. Os $TKRB_C$, $TKRB_1$ e TVB respectivamente, apresentam maior poder em relação ao TW , nessa distribuição.

Na Tabela 4.14 apresenta-se o poder de teste combinado $TKRB_C$ para o nível de significância nominal de $\alpha = 5\%$, nas oito distribuições não-normais univariadas e para diferentes tamanhos n de amostras adotados. Verificou-se um padrão bem similar ao encontrado para $\alpha = 10\%$, ocorrendo alguns poucos casos (distribuições e tamanhos amostrais) em que o teste combinado $TKRB_C$ tem desempenho superior aos demais concorrentes. Assim, para a distribui-

ção $Exp(1)$, $t(v = 1)$ e $Unif(0,1)$ não houve nenhum caso em que o poder de $TKRB_C$ superasse o poder dos concorrentes. Na $Exp(1)$, na $Gama(1,3)$, na $t(v = 30)$ e na $Weib(2,3)$, somente no caso de $n = 10$ é que houve superioridade do teste combinado em relação aos demais, excetuando os casos em que o poder é de 100%, que é claro há igualdade de desempenho com os demais testes que também atingiram este patamar de poder. Já nos casos log-normal e Beta, o $TKRB_C$ foi mais poderoso que seus concorrentes para todos os n 's, exceto para a log-normal com $n = 3$.

Tabela 4.14 – Poder do teste combinado $TKRB_C$ de normalidade, relativo à oito distribuições não normais univariadas, em função de $n=3, 10, 30, 50, 100, 500, 1.000, 5.000$, para o nível nominal de significância $\alpha=5\%$

<i>Teste combinado $TKRB_C$ $\alpha=5\%$</i>								
<i>n</i>	<i>Exp(1)</i>	<i>Gama(1,3)</i>	<i>t(1)</i>	<i>t(30)</i>	<i>Unif(0,1)</i>	<i>LogN(0,1)</i>	<i>Weib(2,3)</i>	<i>Beta(0,2,0,5)</i>
3	0,088	0,063	0,059	0,027	0,033	0,095	0,041	0,269
10	0,558	0,438	0,517	0,059	0,067	0,657	0,124	0,963
30	0,967	0,881	0,917	0,048	0,323	0,996	0,197	1,000
50	0,995	0,996	0,993	0,058	0,563	1,000	0,230	1,000
100	1,000	1,000	1,000	0,050	0,916	1,000	0,383	1,000
500	1,000	1,000	1,000	0,056	1,000	1,000	0,895	1,000
1.000	1,000	1,000	1,000	0,055	1,000	1,000	0,990	1,000
5.000	1,000	1,000	1,000	0,595	1,000	1,000	1,000	1,000

Os desempenhos em relação ao poder de teste combinado $TKRB_C$, para o nível de significância nominal $\alpha = 1\%$, são apresentados na Tabela 4.15. Novamente o padrão de desempenho foi parecido com os demais níveis de significância já comentados anteriormente. Assim, nas distribuições t , com $v = 1$ ou $v = 30$, na $Unif(0,1)$ e na $Weib(2,3)$ não houve um caso sequer, em relação a n , que o teste combinado $TKRB_C$ superasse seus concorrentes. Obviamente, quando, nestas distribuições o valor de poder é de 100%, os valores de poder deste teste iguala ao menos ao desempenho de um dos testes estudados.

Observou-se que o $TKRB_C$, na distribuição $Exp(1)$ e $Gama(1,3)$, com $\alpha = 1\%$, apresenta maior poder que os demais concorrentes apenas para $n = 10$. Considerando $\alpha = 1\%$, para as distribuições $LogN(0,1)$ e $Beta(0,2;0,5)$, verifica-se que o teste combinado foi superior ou igual aos demais para todos os tamanhos amostrais. Na distribuição log-normal os testes atingiram o poder máximo com $n \geq 50$ e na beta, o valor máximo de poder foi atingido com $n \geq 30$. Assim, embora haja em algumas situações de hipóteses alternativas, valores de significâncias e tamanhos amostrais vantajosas para o teste combinado, não se pode recomendá-lo,

Tabela 4.15 – Poder do teste combinado $TKRB_C$ de normalidade, relativo à oito distribuições não normais univariadas, em função de $n=3, 10, 30, 50, 100, 500, 1.000, 5.000$, para o nível nominal de significância $\alpha=1\%$

Teste combinado $TKRB_C$ $\alpha=1\%$								
n	$Exp(1)$	$Gama(1,3)$	$t(1)$	$t(30)$	$Unif(0,1)$	$LogN(0,1)$	$Weib(2,3)$	$Beta(0,2;0,5)$
3	0,024	0,015	0,026	0,011	0,014	0,020	0,007	0,149
10	0,289	0,220	0,373	0,012	0,015	0,482	0,035	0,901
30	0,880	0,784	0,873	0,016	0,147	0,972	0,068	1,000
50	0,980	0,886	0,983	0,012	0,290	1,000	0,087	1,000
100	1,000	1,000	1,000	0,014	0,729	1,000	0,172	1,000
500	1,000	1,000	1,000	0,015	1,000	1,000	0,747	1,000
1.000	1,000	1,000	1,000	0,010	1,000	1,000	0,971	1,000
5.000	1,000	1,000	1,000	0,345	1,000	1,000	1,000	1,000

pois nas demais situações em que seu desempenho é pior, muitas vezes ele foi muito aquém dos concorrentes.

No estudo de proposição do teste de normalidade baseado em entropia amostral, com $\alpha = 5\%$ e tamanho de amostra $n = 10, 20, 50$, Vasicek (1976) comparou essa abordagem com os testes de Kolmogorov-Smirnov (D), Cramer-von Mises (W^2), Kuiper (V), Watson (U^2), Anderson-Darling (A^2) e Shapiro-Wilk (W) em cinco distribuições não-normais: $Exp(1)$, $Gama(2)$; $Unif$; $Beta(2,1)$; $Cauchy$, e concluiu que, nenhum dos testes considerados obteve um desempenho de poder uniformemente superior aos demais. No entanto, constatou que a sua proposta (K_{mm}) apresentou maior poder em pelo menos três das cinco distribuições não-normais em relação às demais alternativas. Estes resultados corroboram com a ideia de não existência de um teste uniformemente mais poderoso em relação aos tamanhos amostrais e distribuições sob a hipótese alternativa. Tal constatação também é verificada no trabalho Filho e Ferreira (2012), ao analisarem testes de normalidade baseados em distâncias.

Como não se sabe *a priori* qual é a distribuição alternativa, não há como recomendar um teste específico. Assim, busca-se um teste que seja uniformemente mais poderoso para amplas classes de distribuições alternativas. O que se pode claramente visualizar destes resultados e discussões apresentadas é que os testes de normalidade baseados na entropia via *bootstrap* paramétrico estiveram quase sempre entre a melhor opção, ou seja, ao menos um deles foi o mais poderoso. Assim, pode-se recomendá-los como testes de normalidade, pois estes superaram, de maneira geral, o teste concorrente de melhor desempenho, escolhido para ser o teste de referência, que é o teste de normalidade de Shapiro-Wilk TW .

Ademais, é conveniente salientar que os testes propostos não são limitados a amostras de tamanho 5.000, como acontece com o *TW*. Assim, esta é uma vantagem relativa dos testes propostos que não deve ser ignorada, principalmente quando se lida com dados ou resíduos de modelos com dezenas ou centenas de milhares de observações

5 CONCLUSÃO

O desenvolvimento do presente estudo possibilitou a proposição de testes de normalidade univariada baseados na entropia e fundamentados em um procedimento computacionalmente intensivo, a técnica de *bootstrap* paramétrico (TKB , $TKRB_1$, $TKRB_2$). Além disso, realizou-se uma combinação linear entre duas estatísticas, originando o $TKRB_C$. Essa pesquisa, oferece uma ferramenta para a verificação de normalidade, sem limitações de tamanhos amostrais n .

Os testes TKB , $TKRB_1$, $TKRB_2$, são exatos, pois, controlaram adequadamente os erros tipo I, nos níveis nominais α de 10% 5% e 1%, para todos os tamanhos amostrais n adotados, conforme apresentado na Tabela 3.1. O $TKRB_C$ apresentou taxas de erro tipo I observadas não significativamente diferentes ($P > 0,01$) nos três níveis de significância nominais e tamanhos amostrais n adotados, ele pode ser considerado um teste exato, como aconteceu para todos os outros testes. A comparação com o teste Shapiro-Wilk (TW) mostrou que eles apresentam, controle perfeito do erro tipo I de todos os testes e com desempenhos equivalentes.

Em relação ao desempenho de poder, nas oito distribuições não-normais, ou hipóteses alternativas, os testes oscilaram bastante, porém os testes propostos, considerando todos os cenários estudados, estiveram entre os de melhor desempenho em todos os casos.

- O $TKRB_1$ foi mais poderoso em seis distribuições não-normais em relação ao TW , sendo que elas são essencialmente distribuições assimétricas à direita.
- O $TKRB_C$ foi mais poderoso em quatro distribuições não-normais em relação ao TW , sendo que elas são essencialmente distribuições assimétricas à direita.
- O teste $TKRB_2$, foi mais poderoso em 2 casos, sendo representados essencialmente por distribuições simétricas, como é o caso da t , que se assemelha consideravelmente com a normal, principalmente com o aumento dos graus de liberdade.
- O TVB foi mais poderoso em duas hipóteses alternativas em relação ao TW , sendo que elas são essencialmente distribuições assimétricas à direita.
- Já o teste de referência Shapiro-Wilk (TW) apresentou melhor desempenho somente para a distribuição Weibull ($\delta = 2$, $\alpha = 3$) com tamanho amostral n entre 30 à 100. Para valor $n \geq 100$, todos os testes propostos e o teste de comparação TW atingiram 100% de poder, para todos os níveis de significância nominais adotados. Assim sendo, o TW ficou sempre entre os melhores, se posicionando quase sempre entre a segunda e a terceira colocação em desempenho de poder para uma ampla gama de distribuições alternativas estudadas.

Como não se sabe *a priori* qual é a distribuição alternativa, não há como recomendar um teste específico. Pode-se dizer que não houve um teste uniformemente mais poderoso para todas as hipóteses alternativas estudadas. Assim, a busca de um teste que seja uniformemente mais poderoso para amplas classes de distribuições alternativas deve continuar. O que se pode claramente visualizar destes resultados e discussões apresentadas é que, os testes de normalidade baseados na entropia via *bootstrap* paramétrico estiveram quase sempre na melhor opção, ou seja, ao menos um deles foi o mais poderoso. Assim, pode-se recomendá-los como testes de normalidade, pois estes superaram, de maneira geral, o teste concorrente de melhor desempenho, escolhido para ser o teste de referência, que é o teste de normalidade de Shapiro-Wilk TW .

Os testes propostos TKB , $TKRB_1$, $TKRB_2$ e o $TKRB_C$ podem ser usados em amostras superiores a 5.000, o que é uma das suas principais virtudes. Nesse sentido, para a aplicação e utilização dessas propostas, foram desenvolvidas rotinas R , permitindo que a comunidade acadêmica e científica possam se beneficiar de tais procedimentos.

REFERÊNCIAS

- ARIZONO, I.; OHTA, H. A test for normality based on kullback—leibler information. **The American Statistician**, Taylor & Francis Group, v. 43, n. 1, p. 20–22, 1989.
- CARMER, S. G.; SWANSON, M. R. An evaluation of ten pairwise multiple comparison procedures by monte carlo methods. **Journal of the American Statistical Association**, Taylor & Francis, v. 68, n. 341, p. 66–74, 1973.
- CHERNICK, M. R. et al. Bootstrap methods. In: **International encyclopedia of statistical science**. [S.l.]: Springer, 2011. p. 169–174.
- CRZCGORZEWSKI, P.; WIRCZORKOWSKI, R. Entropy-based goodness-of-fit test for exponentiality. **Communications in Statistics-Theory and Methods**, Taylor and Francis, v. 28, n. 5, p. 1183–1202, 1999.
- DIAS, M. S.; TAKEDA, M. N.; KOSKINAS, M. F. Application of monte carlo simulation to the prediction of extrapolation curves in the coincidence technique. **Applied radiation and isotopes**, Elsevier, v. 64, n. 10-11, p. 1186–1192, 2006.
- EFRON, B. **The jackknife, the bootstrap, and other resampling plans**. New York: Siam, 1982. v. 38.
- EFRON, B.; TIBSHIRANI, R. J. An introduction to the bootstrap. **Monographs on statistics and applied probability**, New York, v. 57, 1993.
- EFRON, B.; TIBSHIRANI, R. J. The jackknife. In: **An introduction to the bootstrap**. New York: Springer, 1993. p. 141–152.
- FERREIRA, D. F. **Estatística básica**. Lavras: Editora UFLA, 2005. 674 p.
- FERREIRA, D. F. **Estatística multivariada**. 2. ed. Lavras: Editora UFLA, 2011. 676 p.
- FERREIRA, D. F. **Estatística computacional em java**. Editora UFLA, Lavras, p. 695, 2013.
- FILHO, N. A. P.; FERREIRA, D. F. Teste Monte Carlo de normalidade univariado baseado em distâncias. **Rev. Bras. Biom**, v. 30, n. 3, p. 401–416, 2012.
- GEBERT, D. M. P.; FERREIRA, D. F. Parametric bootstrap tests for determining the number of principal components. **Journal of Statistical Theory and Practice**, Taylor & Francis, v. 8, n. 4, p. 674–691, 2013.
- GUO, J.; ALEMAYEHU, D.; SHAO, Y. Tests for normality based on entropy divergences. **Statistics in Biopharmaceutical Research**, Taylor & Francis, v. 2, n. 3, p. 408–418, 2010.
- JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. **International Statistical Review/Revue Internationale de Statistique**, JSTOR, p. 163–172, 1987.
- KOHANSAL, A.; REZAKHAH, S. Modified entropy estimators for testing normality. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 86, n. 3, p. 553–568, 2016.
- KRULL, J. L.; MACKINNON, D. P. Multilevel modeling of individual and group level mediated effects. **Multivariate behavioral research**, Taylor & Francis, v. 36, n. 2, p. 249–277, 2001.

- LEOTTI, V. B.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **Revista HCPA. Porto Alegre. Vol. 32, no. 2 (2012), p. 227-234**, 2012.
- LIN, C.-C.; MUDHOLKAR, G. S. A simple test for normality against asymmetric alternatives. **Biometrika**, Oxford University Press, v. 67, n. 2, p. 455–461, 1980.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. Tests of hypotheses. **Introduction to the theory of statistics. Tokio: McGraw-Hill**, p. 401–470, 1974.
- MORETTIN, P. A.; BUSSAB, W. O. Inferência para várias populações. **Estatística básica**, Saraiva São Paulo, v. 5, p. 410–435, 2003.
- OLIVEIRA, I. R. C.; FERREIRA, D. F. Multivariate extension of chi-squared univariate normality test. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 80, n. 5, p. 513–526, 2010.
- PARK, S.; PARK, D. Correcting moments for goodness of fit tests based on two entropy estimates. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 73, n. 9, p. 685–694, 2003.
- PRESS, W. H. et al. Numerical recipes in C++. **The art of scientific computing**, 1992.
- ROYSTON, P. Approximating the shapiro-wilk W-test for non-normality. **Statistics and Computing**, Springer, v. 2, n. 3, p. 117–119, 1992.
- ROYSTON, P. A pocket-calculator algorithm for the shapiro-francia test for non-normality: An application to medicine. **Statistics in medicine**, Wiley Online Library, v. 12, n. 2, p. 181–184, 1993.
- SCOTT, D. W. On optimal and data-based histograms. **Biometrika**, Oxford University Press, v. 66, n. 3, p. 605–610, 1979.
- SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, p. 379–423, 1948.
- SILVA, C. Miguel da. Ra-226 ra-228 na dieta de bovinos leiteiros do agrste semi-árido de pernambuco e avaliação de risco decorrente de consumo de leite por uma população potencialmente exposta. Universidade Federal de Pernambuco, 2006.
- TANI, G. et al. Variabilidade de resposta e processo adaptativo em aprendizagem motora. **Revista Paulista de Educação Física**, v. 6, n. 1, p. 16–25, 2017.
- TSANG, T. S. et al. Left atrial volume: important risk marker of incident atrial fibrillation in 1655 older men and women. In: ELSEVIER. **Mayo Clinic Proceedings**. New York, 2001. v. 76, n. 5, p. 467–475.
- VASICEK, O. A test for normality based on sample entropy. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 54–59, 1976.