

Amanda Ramalho Ferreira

**Testes de normalidade: Um estudo
comparativo**

Niterói - RJ, Brasil

26 de Julho de 2022

Amanda Ramalho Ferreira

Testes de normalidade: Um estudo comparativo

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em Estatística pela Universidade Federal Fluminense.

Orientador(a): Profa. Dra. Karina Yuriko Yaginuma

Co-Orientador(a): Prof. Dr. Douglas Rodrigues Pinto

Niterói - RJ, Brasil

26 de Julho de 2022

Amanda Ramalho Ferreira

**Testes de normalidade: Um estudo
comparativo**

Monografia de Projeto Final de Graduação sob o título “*Testes de normalidade: Um estudo comparativo*”, defendida por Amanda Ramalho Ferreira e aprovada em 26 de Julho de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Douglas Rodrigues Pinto
Departamento de Estatística – UFF

Profa. Dra. Estelina Serrano de Marins Capistrano
Departamento de Estatística – UFF

Prof. Dr. Hugo Henrique Kegler dos Santos
Departamento de Estatística – UFF

Niterói, 26 de Julho de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

F383t Ferreira, Amanda Ramalho
Testes de normalidade: Um estudo comparativo / Amanda Ramalho Ferreira ; Karina Yuriko Yaginuma, orientadora ; Douglas Rodrigues Pinto, coorientador. Niterói, 2022.
36 f.

Trabalho de Conclusão de Curso (Graduação em Estatística)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Análise multivariada. 2. Simulação. 3. Produção intelectual. I. Yaginuma, Karina Yuriko, orientadora. II. Pinto, Douglas Rodrigues, coorientador. III. Universidade Federal Fluminense. Instituto de Matemática e Estatística. IV. Título.

CDD -

Resumo

A suposição de normalidade dos dados é muito utilizada para realizar inferências sobre os parâmetros populacionais, diversos métodos estatísticos e fenômenos aleatórios podem ser descritos de forma aproximada pela distribuição normal. Por isso é de extrema importância que sejam realizados testes para avaliar o atendimento a esta suposição para um determinado conjunto de dados. O objetivo deste trabalho foi investigar quais testes de normalidade apresentam melhores resultados para diferentes tamanhos de amostras e distribuições. A metodologia utilizada foi um estudo de simulação, realizado no R (R Core Team, 2014), que utilizou amostras geradas aleatoriamente, para realizar sucessivos testes de normalidade e comparar as taxas de acerto. Os resultados desta pesquisa podem auxiliar em estudos futuros, na escolha dos testes que serão utilizados para as análises. Para amostras grandes, maiores que 1000, o teste Jarque Bera Robusto apresentou o melhor desempenho e no caso de amostras pequenas, de uma maneira geral, o teste Shapiro Francia obteve os melhores resultados.

Palavras-chave: Distribuição normal. Testes de normalidade. Simulação.

Conteúdo

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 9
1.1	Motivação	p. 9
1.2	Revisão Bibliográfica	p. 9
1.3	Objetivos	p. 10
1.4	Organização	p. 11
2	Materiais e Métodos	p. 12
2.1	A Distribuição Normal	p. 12
2.2	Testes de Normalidade	p. 15
2.2.1	Shapiro Wilk	p. 16
2.2.2	Shapiro Francia	p. 17
2.2.3	Lilliefors	p. 18
2.2.4	Cramér-Von Mises	p. 19
2.2.5	Anderson Darling	p. 19
2.2.6	Jarque-Bera Robusto	p. 20
2.3	Acurácia, Precisão, Sensibilidade e Especificidade	p. 20
3	Análise dos Resultados	p. 23
3.1	Distribuições utilizadas no estudo de simulação	p. 23

3.2	Metodologia da Simulação	p. 25
3.3	Resultados	p. 25
3.3.1	Resultados para a distribuição normal	p. 26
3.3.2	Resultados para as distribuições não normais	p. 27
4	Conclusões	p. 31
	Referências	p. 32
	Anexo 1 – Tabela para o valor-p do teste Lilliefors	p. 34
	Anexo 2 – Tabela da distribuição acumulada da normal padrão	p. 35

Lista de Figuras

1	Exemplos da curva de densidade Normal para diferentes valores dos parâmetros μ e σ	p. 14
2	Distribuição amostral da Exponencial e de sua média para amostras de tamanho 50	p. 15
3	Comparação das curvas das distribuições simétricas com a curva da Normal	p. 24
4	Comparação das curvas das distribuições simétricas com a curva da Normal	p. 24
5	Comparação do resultado da distribuição Normal para os diferentes testes de normalidade	p. 26
6	Comparações das taxas de acertos das distribuições para cada um dos testes	p. 29

Lista de Tabelas

1	Matriz confusão	p. 22
2	Distribuições com parâmetros e simetria da curva utilizadas neste estudo	p. 25
3	Taxa de acerto para a sensibilidade dos testes	p. 27
4	Taxa de acerto para a especificidade dos testes	p. 28

1 Introdução

1.1 Motivação

Segundo Dufour et al. (1998), existem cerca de 40 testes de normalidade. Estes diferentes testes, muitas vezes, geram resultados opostos que podem rejeitar ou não a hipótese de normalidade para o mesmo conjunto de dados. Isto acontece, pois, cada teste utiliza um método diferente de verificação, além de ser possível selecionar infinitos tipos de amostra. Logo, dependendo da amostra analisada, um determinado método pode ser mais eficiente do que outro. Diante deste cenário, deseja-se comparar diferentes testes, em situações diversas, para determinar qual teste performa melhor em cada caso.

1.2 Revisão Bibliográfica

Torman, Birck e Riboldi (2005), comparam os testes Kolmogorov-Smirnov, Anderson-Darling, Cramér-Von Mises e Shapiro-Wilk através de simulação Monte Carlo. Para verificar a eficiência desses testes, foram realizadas simulações no software SAS, com 1000 replicações para 5 distribuições e 4 tamanhos de amostras diferentes. Os resultados mostraram que para dados normais, apenas o Kolmogorov-Smirnov se mostrou inferior comparado aos demais. Para os dados não-normais o teste Shapiro-Wilk sempre se mostra superior, concluindo-se então que este é aparentemente o melhor teste de aderência à normalidade.

Em Torman, Coster e Riboldi (2012), os autores comparam os testes Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, Lilliefors, Qui-quadrado, Shapiro-Wilk, Shapiro-Francia e Jarque-Bera através de simulação. Para analisar a eficiência dos testes foi averiguado o percentual de vezes que a decisão correta foi tomada através do software R, com 5000 replicações para 5 distribuições e 6 tamanhos de amostras diferentes. Os resultados mostraram que para amostras pequenas os procedimentos não paramétricos são mais indicados, em função da baixa performance dos testes de aderência à normalidade.

Para amostras maiores, recomenda-se o uso dos testes Shapiro-Francia ou Shapiro-Wilk. Os testes Kolmogorov-Smirnov e Qui-Quadrado, apresentaram os piores desempenhos.

Razali e Wah (2011) realiza uma comparação entre os testes Kolmogorov-Smirnov, Anderson-Darling, Lilliefors e Shapiro-Wilk via simulação Monte Carlo. A eficiência dos testes foi verificada utilizando o software FORTRAN, com 10000 replicações para 14 distribuições e 15 tamanhos de amostras diferentes. Os resultados indicam que o teste Shapiro-Wilk é o mais poderoso, seguido do Anderson-Darling, Lilliefors e Kolmogorov-Smirnov, respectivamente. Entretanto, o poder de todos os testes são baixos para tamanhos pequenos de amostras. Concluiu-se que o Lilliefors sempre performa melhor do que o KS o que justifica a escolha deste teste para este trabalho.

O artigo Bayoud (2019) compara os testes Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, Shapiro-Wilk, Shapiro-Francia, Watson, Kuiper, Vasicek's, Park-Park teste baseado no estimador Vasicek's, Jarque-Bera, Jarque-Bera Robusto e o teste proposto pelo autor através de simulação Monte Carlo. A eficiência dos testes foi verificada utilizando o software R, com 10000 replicações para 12 distribuições e 4 tamanhos de amostras diferentes. Para amostras simétricas variando de $(-\infty, \infty)$, o teste mais poderoso foi o Jarque-Bera Robusto. Já para o caso de amostras assimétricas variando de $(0, \infty)$ ou de $(-\infty, \infty)$ o teste mais poderoso foi o teste proposto pelo autor para amostras pequenas e moderadas, para amostras grandes o teste Shapiro-Wilk é o mais indicado.

Em Yap e Sim (2011) os autores comparam os testes Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, Cramer-Von Mises, Anderson-Darling, D'Agostino-Pearson, Jarque-Bera e Qui-quadrado através de simulação via Monte Carlo. A eficiência dos testes foi testada utilizando diferentes softwares, pois não havia um único software que realizasse todos os testes escolhidos, com 10000 replicações para 24 distribuições e 15 tamanhos de amostras diferentes. A simulação indica que para amostras simétricas de cauda curta os testes D'Agostino e Shapiro-Wilk são mais poderosos. Para os casos em que as amostras são simétricas e de cauda longa, o poder dos testes Jarque-Bera e D'Agostino se compara ao Shapiro-Wilk teste. Para amostras assimétricas, o teste Shapiro-Wilk é o mais poderoso seguido do Anderson-Darling.

1.3 **Objetivos**

O objetivo é comparar os testes de normalidade Shapiro-Wilk, Shapiro-Francia, Lilliefors, Anderson-Darling, Cramér-Von Mises e Jarque-Bera Robusto, através de um estudo

de simulação implementado no programa R (R Core Team, 2014) para decidir quais apresentam o melhor desempenho em diferentes situações, como por exemplo, para os casos de amostras pequenas, amostras grandes, com ruído e sem ruído.

Os objetivos específicos são:

- Realizar um estudo para comparar os diferentes tipos de testes de normalidade, em situações diversas.
- Identificar quais testes apresentam o melhor desempenho para distribuições normais e não normais.
- Identificar qual o melhor teste para os diferentes tamanhos de amostras.

1.4 Organização

No Capítulo 2 é apresentado mais detalhes sobre a distribuição normal, os testes de normalidade utilizados na simulação e as definições de acurácia, precisão, sensibilidade e especificidade. No Capítulo 3 é possível encontrar as distribuições utilizadas neste estudo, a metodologia da simulação e os resultados obtidos. Por fim, no Capítulo 4 estão registradas as principais conclusões desta pesquisa.

2 Materiais e Métodos

2.1 A Distribuição Normal

A distribuição normal, também conhecida como curva Gaussiana ou curva de Gauss, possui grande importância na estatística possuindo aplicações práticas e teóricas. Atualmente, ela é muito utilizada na prática, em diversas áreas de aplicação, como por exemplo, descreve bem diversos fenômenos naturais, em particular, as variáveis biométricas.

Na parte teórica, esta distribuição possui um papel fundamental na inferência estatística, uma vez que muitos procedimentos utilizados nesta área possuem como suposição a normalidade dos dados, ou seja, outros modelos probabilísticos tanto discretos como contínuos podem ser representados pela normal, desde que os critérios de convergência sejam satisfeitos. Alguns exemplos de distribuições que se aproximam da curva normal quando as condições são respeitadas são: Binomial, Poisson, Hipergeométrica, T-Student, Qui-Quadrada e a distribuição F. Em Viali (2014), o autor explica mais detalhadamente sobre a importância e aplicações da normal.

Sabe-se que ao realizar o experimento de lançar uma moeda não viciada e observar sua face, a probabilidade de ocorrência dos eventos A: obter cara ou B: obter coroa, é de 50%. Observe que estes eventos possuem dois resultados possíveis, sair a face desejada (sucesso) ou não (fracasso). Diz-se, então, que estes experimentos seguem uma distribuição de Bernoulli. Ao replicar este experimento n vezes e observar o número de sucessos para as tentativas realizadas, tem-se uma distribuição binomial. Estas distribuições descritas por Bernoulli, foram fundamentais nos estudos de Moivre para descrever pela primeira vez, em 1738, a curva que mais tarde ficou conhecida como curva de Gauss, dando razão a lei da eponímia de Stigler que afirma “nenhuma descoberta científica é batizada com o nome do seu descobridor original”, Viali (2014).

Viali (2014) conta um pouco da história da normal. O matemático francês Abraham de Moivre, em seu livro “The doctrine of the chances”, percebeu que conforme aumentava

o número de lançamentos de uma moeda, a distribuição binomial aproximava-se de uma curva suave, ou seja, quanto maior o número de tentativas, mais exata era a aproximação para a probabilidade de ocorrência de um evento, pois os erros ou desvios, diminuem a medida que aumentam os números de lançamentos.

Segundo Viali (2014), os resultados do trabalho de Moivre foram estendidos por Pierre Simon Laplace e Carl Friedrich Gauss. Laplace foi um matemático, astrônomo e físico francês que, em 1810, apresentou e provou o teorema central do limite (TCL) e, em 1812, em seu livro “Théorie Analytique des probabilités” provou através do TCL que, sob certas circunstâncias, a binomial e outras distribuições podem ser aproximadas por uma Normal. Gauss, um alemão que assim como Laplace era matemático, astrônomo e físico, em seus trabalhos “Theoria Motus Corporum Coelestium” de 1809 e “Bestimmung der Genauigkeit der Beobachtungen, Zeitschrift Astronomi” de 1816, associou a curva normal ao método dos mínimos quadrados e estabeleceu métodos padrões amplamente utilizados no século XIX. Por este motivo, a distribuição Normal também é conhecida como curva de Gauss.

Para este estudo é necessário definir a distribuição normal, a normal padrão e o teorema do limite central.

Definição 2.1 *A variável aleatória X tem distribuição normal com média μ e variância σ^2 , se a sua função de probabilidade é dada por:*

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], -\infty < x < \infty \quad (2.1)$$

em que $-\infty < \mu < \infty$ e $\sigma^2 > 0$. Denotamos esta distribuição por $X \sim N(\mu, \sigma^2)$ em que μ e σ^2 são os parâmetros de posição e escala, respectivamente.

A distribuição Normal padrão é um caso particular da Normal, em que a média μ é zero e a variância σ^2 é 1. Em muitos casos é interessante padronizar, isto é, transformar uma X que segue uma distribuição Normal em uma variável aleatória Z que segue uma distribuição Normal padrão, conforme mostra a proposição 2.1. Este procedimento é realizado, pois, facilita a leitura e interpretação dos dados. Além disso, diversos testes de normalidade utilizam a Normal padrão, que pode ser definida como:

Definição 2.2 *A variável aleatória Z tem distribuição Normal padrão, então a função*

de densidade de probabilidade da Normal reduz-se a:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right], -\infty < z < \infty \quad (2.2)$$

Denotamos esta distribuição por $Z \sim N(0, 1)$.

O próximo resultado mostra como transformar uma distribuição Normal com média μ e variância σ^2 em uma distribuição Normal padrão.

Proposição 2.1 *Seja $X \sim N(\mu, \sigma^2)$. A padronização da variável aleatória X é realizada através da seguinte transformação:*

$$Z = \frac{X - \mu}{\sigma} \quad (2.3)$$

em que $-\infty < \mu < \infty$ e $\sigma > 0$. Então, $Z \sim N(0, 1)$.

A função densidade da distribuição Normal possui forma de sino, sendo uma curva simétrica em torno da média e duplamente assintótica, ou seja, o limite quando x tende para ∞ é igual ao limite quando x tende para $-\infty$, sendo ambos os limites iguais a zero. O intervalo entre $\mu - 3\sigma$ e $\mu + 3\sigma$ representa 99,73% da área total sob a curva. Exemplos da densidade para diferentes valores dos parâmetros μ e σ podem ser encontrados na Figura 1.

Como a distribuição Normal padrão é um caso particular da Normal, o seu gráfico possui as mesmas propriedades de simetria da Normal.

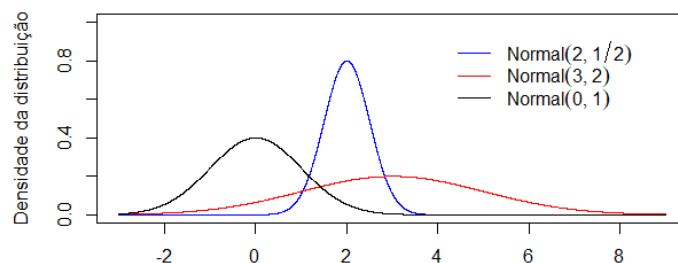


Figura 1: Exemplos da curva de densidade Normal para diferentes valores dos parâmetros μ e σ

Segundo Ross (2010), pelo TCL, mesmo que os dados não sigam uma distribuição

normal, quando a dimensão da amostra é suficientemente grande a distribuição da média amostral de uma população com variância finita aproxima-se de uma distribuição Normal.

Teorema 2.1 *Para amostras aleatórias simples (X_1, X_2, \dots, X_n) retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral da média \bar{X} aproxima-se para n suficientemente grande, de uma distribuição Normal com média μ e variância $\frac{\sigma^2}{n}$.*

Na Figura 2, é apresentado uma ilustração referente ao TCL. Observe que apesar da curva de distribuição da Exponencial não se parecer com a normal, a curva da média para uma quantidade grande de repetições da amostra é bem similar a normal. Para estes gráficos da exponencial o seu parâmetro é $\lambda = 1$, a amostra utilizada foi de tamanho 50 e o número de replicações realizadas foi de 10000.

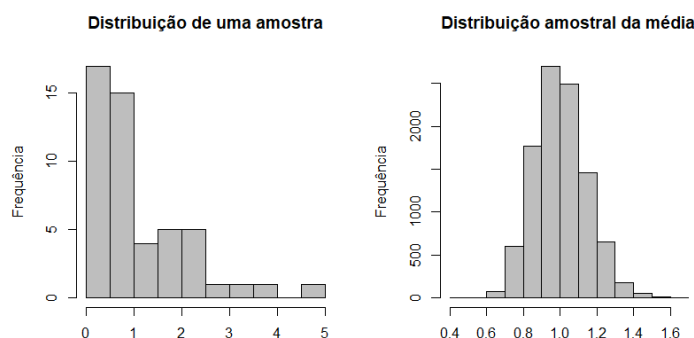


Figura 2: Distribuição amostral da Exponencial e de sua média para amostras de tamanho 50

A seguir serão apresentados os testes de normalidade utilizados nesta simulação.

2.2 Testes de Normalidade

Seier (2002) classificou os testes de normalidade em quatro subcategorias, sendo elas: testes de assimetria e curtose, testes de distribuição empírica, testes de regressão e correlação e outros testes especiais. Neste trabalho serão utilizados testes de normalidade das três primeiras categorias.

Para a categoria dos testes de assimetria e curtose, compara-se os resultados de assimetria e curtose dos dados com os resultados destas medidas da distribuição Normal. O teste Jarque-Bera Robusto é um exemplo dessa categoria de testes.

A ideia dos testes de distribuição empírica é comparar a função de distribuição empírica (FDE) estimada através dos dados com a função de distribuição Normal, ou seja, compara-se a distribuição empírica e a teórica. Dentro desta categoria, ainda é possível classificar os testes como pertencentes à classe suprema de estatísticas baseadas na FDE ou à classe quadrática. Da classe suprema, será utilizado o teste Lilliefors e da classe quadrática os testes Anderson Darling e Cramér-Von Mises.

Na categoria de correlação e regressão deseja-se comparar os valores esperados das estatísticas de ordem da distribuição Normal com as estatísticas de ordem observadas para a distribuição da qual se deseja verificar a suposição de normalidade. Nesta categoria temos os testes Shapiro-Wilk e sua modificação, o Shapiro-Francia.

Antes de definir estes testes, é preciso enunciar a definição de estatística de ordem, já que a mesma é utilizada nos testes de correlação e regressão.

Dado uma amostra x_1, x_2, \dots, x_n , os dados ordenados de forma crescente, formam uma sequência denotada por $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, onde os parênteses no subscrito indicam a ordem. Dessa forma, $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ são as estatísticas de ordem da amostra e $x_{(i)}$ indica a i -ésima estatística de ordem, ou seja, a i -ésima observação ordenada. A estatística de ordem pode ser enunciada conforme Definição 2.3, Ross (2010).

Definição 2.3 *Seja (X_1, X_2, \dots, X_n) um vetor aleatório n -dimensional e (x_1, x_2, \dots, x_n) uma n -upla assumida por (X_1, X_2, \dots, X_n) . Vamos organizar x_1, x_2, \dots, x_n em ordem crescente de magnitude, para que*

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}, \quad (2.4)$$

onde $x_{(1)} = \min(x_1, x_2, \dots, x_n)$, $x_{(2)}$ é o segundo menor valor em x_1, \dots, x_n e assim por diante, $x_{(n)} = \max(x_1, \dots, x_n)$. Se quaisquer dois x_i, x_j forem iguais, a ordem não importa.

Em seguida, será apresentado cada um dos testes de normalidade considerados nesse trabalho.

2.2.1 Shapiro Wilk

O teste proposto por Shapiro e Wilk (1965) é, até os dias de hoje, um dos mais utilizados devido ao fato dele ser o mais sensível para diversas distribuições, de acordo com Althouse, Ware e Ferron (1998). Para verificar se uma amostra segue uma distribuição

normal, utiliza-se a seguinte estatística de teste:

$$SW = \frac{\sum_{i=1}^n (a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.5)$$

em que y_i é o i -ésimo valor da amostra observada, \bar{y} é a média amostral e a_1, a_2, \dots, a_n são constantes calculadas da seguinte forma:

$$(a_1, a_2, \dots, a_n)^T = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}, \quad (2.6)$$

sendo $m = (m_1, m_2, \dots, m_n)^T$ o vetor de valores esperados das estatísticas de ordem da amostra e V é a matriz das covariâncias destas estatísticas.

No R, a função `shapiro.test()` do pacote *nortest* fornece a estatística de teste e o valor-p para o teste Shapiro-Wilk e será a função utilizada neste trabalho.

2.2.2 Shapiro Francia

Shapiro e Francia (1972) apresentaram uma modificação para o teste acima, em que mantiveram a mesma estrutura e critérios utilizados no teste Shapiro-Wilk, sua diferença está na definição dos coeficientes a_n 's. Logo, a estatística de teste considerada neste teste é a mesma apresentada na Equação 2.5 e o vetor a é definido da seguinte forma:

$$(a_1, a_2, \dots, a_n)^T = \frac{m^T}{(m^T m)^{1/2}}, \quad (2.7)$$

sendo $m = (m_1, m_2, \dots, m_n)^T$ o vetor de valores esperados das estatísticas de ordem da amostra.

A grande vantagem deste teste é que não é preciso calcular a matriz de covariâncias V , ou seja, se torna mais simples obter os coeficientes associados ao vetor a , segundo Silva (2014). Atente-se ao fato de que isto não significa que este teste é mais assertivo do que o Shapiro-Wilk.

No R, a função que realiza este teste é a `sf.test()`, também do pacote *nortest*, ela fornece as informações sobre estatística de teste e o valor-p.

2.2.3 Lilliefors

De acordo com Seier (2002), o teste Kolmogorov-Smirnov (KS) proposto inicialmente por Kolmogorov, em 1933, utiliza a FDE para verificar a normalidade dos dados e pertence a classe suprema de estatísticas baseadas na FDE. Este teste é muito utilizado até os dias de hoje, porém, ele possui a limitação de que os parâmetros μ e σ^2 devem ser totalmente conhecidos para que seja possível realizar os cálculos.

Em seus estudos, Lilliefors (1967) notou que com uma modificação nos parâmetros do teste Kolmogorov-Smirnov era possível simplificar os cálculos deste teste e resolver a limitação citada utilizando as estimativas dos parâmetros μ e σ^2 , enquanto que no KS utiliza-se as estimativas dos parâmetros amostrais \bar{x} e S. A estatística de teste é:

$$D = \max|D^+, D^-| \quad (2.8)$$

sendo

$$D^+ = \max_{(i=1, \dots, n)} \left\{ \frac{i}{n} - p_{(i)} \right\} \text{ e } D^- = \max_{(i=1, \dots, n)} \left\{ p_{(i)} - \left(\frac{i-1}{n} \right) \right\}$$

$p_{(i)} = \phi\left(\frac{x_{(i)} - \bar{x}}{s}\right)$, sendo ϕ a função de distribuição acumulada normal padrão, com \bar{x} e s, respectivamente, a média e o desvio padrão amostral.

O valor-p para este teste pode ser obtido através da tabela do teste Lilliefors (ANEXO 1), calculada por Dallal e Wilkinson (1986). Entretanto, esta fórmula possui uma limitação, ela só é confiável quando valor-p menor que 0,1. Criada por Stephens (1979), para os casos em que o valor-p da fórmula acima fossem maiores que 0,1, utiliza-se então a estatística de teste modificada:

$$Z = D \left(\sqrt{n} - 0,01 + \frac{0,85}{\sqrt{n}} \right). \quad (2.9)$$

Para o teste Lilliefors a função do R é a `lillie.test()` do pacote `nortest`, que nos fornece as mesmas informações dos outros testes, a estatística do teste e o valor-p.

2.2.4 Cramér-Von Mises

Este teste foi proposto inicialmente Harald Cramér e Richard Von Mises, em 1928, porém sugestões de modificações e melhorias foram criadas por diversos pesquisadores, acompanhando o avanço dos estudos sobre testes de normalidade. Darling (1957) definiu a estatística deste teste da seguinte forma:

$$W = \frac{1}{12n} + \sum_{i=1}^n \left[\frac{2i-1}{2n} - p_{(i)} \right]^2 \quad (2.10)$$

onde n é o tamanho amostral, $p_{(i)} = \phi\left(\frac{x_{(i)} - \bar{x}}{s}\right)$, ϕ é a função de distribuição acumulada normal padrão, com \bar{x} e s , respectivamente, a média e o desvio padrão amostral.

O cálculo do valor-p foi proposto por Stephens (1979), e pode ser encontrado através da estatística modificada,

$$Z = W(1 + 0,5n) \quad (2.11)$$

No R a função `cvm.test()` do pacote *nortest* fornece a estatística de teste e o valor-p para o teste acima.

2.2.5 Anderson Darling

Anderson e Darling (1952) propuseram um teste de normalidade baseado na função de distribuição empírica, para distribuições com parâmetros conhecidos, definido por:

$$A = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1)(\ln p_{(i)} + \ln(1 - p_{(n-1+i)}))] \quad (2.12)$$

em que $p_{(i)} = \phi\left(\frac{x_{(i)} - \bar{x}}{s}\right)$ são os percentis ordenados da distribuição normal padrão, ϕ é a função de distribuição acumulada normal padrão, \bar{x} é a média amostral e s é o desvio padrão amostral. D'Agostino e Stephens (1986) sugeriram uma modificação no teste Anderson-Darling para que fosse possível usá-lo com algumas distribuições com parâmetros desconhecidos, incluindo a Normal, exponencial, gama, weibull e entre outras. O teste modificado é definido como:

$$A^* = A \left(1 + \frac{0,75}{n} + \frac{2,25}{n^2} \right) \quad (2.13)$$

Para um nível de significância de 0,05, a hipótese de normalidade é rejeitada se $A^* > 0,752$ para qualquer tamanho amostral n . No R, a função `ad.test()` do pacote *nortest* fornece a estatística de teste e o valor-p para o teste acima.

2.2.6 Jarque-Bera Robusto

Este teste foi proposto por Jarque e Bera (1980) e baseia-se na diferença entre a simetria e a curtose dos dados amostrais e teóricos da distribuição normal. Apesar deste teste ser conhecido por ter boas propriedades para verificar a normalidade, ele possui algumas limitações, como por exemplo, valores discrepantes podem afetar os coeficientes utilizados e nos levar a decisão errada. Gel e Gastwirth (2008) apresentaram uma modificação neste teste, com o objetivo de tentar reduzir o impacto destes valores discrepantes na decisão do teste. A estatística de teste do Jarque-Bera robusto é definida como:

$$JBR = \frac{n}{C_1} \left(\frac{\hat{\mu}_3}{J_n^3} \right)^2 + \frac{n}{C_2} \left(\frac{\hat{\mu}_4}{J_n^4} - 3 \right)^2 \quad (2.14)$$

em que $\hat{\mu}_k = \sum_{i=1}^n (x_i - \bar{x})^k$, $J_n = \sqrt{\frac{\pi}{2}} \frac{\sum_{i=1}^n |x_i - M|}{n}$, M é a mediana dos valores x_i da amostra e C_1 e C_2 são constantes positivas que podem ser obtidas através de simulação via Monte Carlo. Em particular, é possível preservar um nível de significância de 0,05 para $C_1 = 6$ e $C_2 = 64$. Segundo Bayoud (2019), a hipótese de normalidade é rejeitada para um $\alpha = 0,05$, se a estatística de teste JBR for maior que 10,658, 9,278, 8,403 e 8,793 para $n = 10, 20, 30$ e 50 respectivamente. Estes valores críticos foram obtidos via simulação com 10.000 replicações. No R, a função `rjb.test()` do pacote *lawstat* fornece a estatística de teste e o p-valor para o teste acima.

2.3 Acurácia, Precisão, Sensibilidade e Especificidade

As medidas de acurácia, precisão, sensibilidade e especificidade serão utilizadas na análise de eficiência para este estudo de simulação. O site de aprendizado de máquinas

criado pelos professores e alunos da UFF, Rodrigues et al. (2022) fala um pouco sobre esses conceitos.

Acurácia é a capacidade de medir corretamente aquilo que se propõe medir, ou seja, é a proximidade entre o valor verdadeiro e o experimental ao realizar a medição de uma grandeza física qualquer. Neste estudo é a capacidade de identificar corretamente se uma distribuição é normal ou não.

A sensibilidade do teste é capacidade de identificar corretamente os elementos da amostra que possuem a característica de interesse do estudo. Pode-se dizer também que é a probabilidade de não rejeitar a hipótese nula quando ela é realmente verdadeira. Pode ser representada como:

$$\text{Sensibilidade} = \frac{VP}{VP + FN}, \quad (2.15)$$

sendo VP = verdadeiros positivos e FN = Falsos negativos. Na Tabela 1 é mostrada a matriz confusão, esta matriz é determinada de acordo com a decisão que o teste toma e a realidade.

Já a especificidade é capacidade de identificar corretamente os elementos que não possuem as características de interesse, ou seja, é a probabilidade de rejeitar a hipótese nula quando a mesma é realmente falsa. Pode ser apresentada através da seguinte equação:

$$\text{Especificidade} = \frac{VN}{VN + FP}, \quad (2.16)$$

sendo VN = verdadeiros negativos e FP = falsos positivos.

Neste estudo, a sensibilidade do teste se refere a proporção de não rejeições da hipótese de nulidade para as distribuições normais e a especificidade é a proporção de rejeições da hipótese de nulidade para distribuições que não são normais.

Já a precisão, está relacionada ao grau de consistência das medidas obtidas, ou seja, é a capacidade de fornecer os mesmos resultados quando repetido, neste caso é a capacidade dos testes fornecerem os mesmos resultados ao ser repetido para diferentes amostras e distribuições. A medida de acurácia está relacionada a erros sistemáticos e aleatórios, enquanto a precisão está relacionada apenas a erros aleatórios.

$$Precisão = \frac{VP + VN}{VP + VN + FN + FP}. \quad (2.17)$$

Tabela 1: Matriz confusão

Decisão	Realidade	
	H_0 é verdadeiro	H_0 é falso
Aceitar H_0	Verdadeiro positivo	Falso negativo (Erro tipo II)
Rejeitar H_0	Falso positivo (Erro tipo I)	Verdadeiro negativo

3 Análise dos Resultados

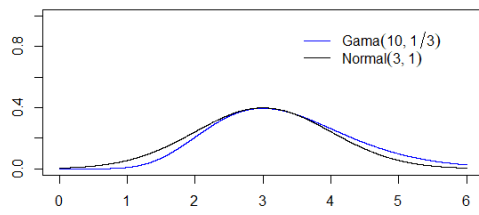
Neste capítulo serão apresentados os resultados obtidos através das análises de simulações realizadas. Para este trabalho o software R 4.0.4 (R Core Team, 2014) foi utilizado, através da interface R Studio.

3.1 Distribuições utilizadas no estudo de simulação

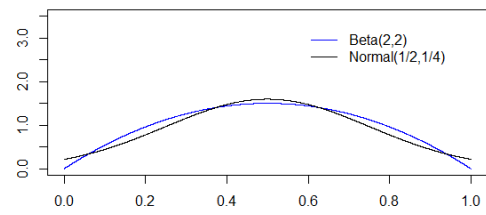
Para testar a eficiência dos testes de normalidade utilizou-se neste trabalho tanto distribuições normais como as seguintes distribuições não normais: beta, gama, t-student, qui-quadrado e cauchy. Estas distribuições foram escolhidas de forma a existir amostras que se aproximem da normal e amostras diferentes da normal. Dessa forma, é possível avaliar como os testes se comportam para as diferentes distribuições. Espera-se que nos casos em que a curva da densidade se parece com a curva da normal, a taxa de acerto seja menor.

As distribuições beta e gama podem possuir um gráfico similar ao da normal ou totalmente diferente, o que as tornam distribuições interessantes de avaliar se o comportamento varia conforme os parâmetros mudam, por este motivo foram escolhidas duas distribuições betas e duas gamas. A distribuição beta $(2, 2)$ e a gama $(10, 1/3)$ possuem um gráfico simétrico similar a curva de uma normal, Figuras 3(b) e 3(a), enquanto que a beta $(2, 1)$ e a gama $(5, 5)$ possuem um gráfico assimétrico, sendo a beta uma assimetria à esquerda (ou negativa) e a gama uma assimetria à direita (ou positiva), conforme mostra as Figuras 4(b) e 4(a).

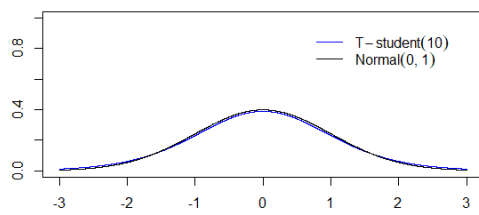
A distribuição t-student com 10 graus de liberdade (g.l.) é uma distribuição simétrica muito similar a normal padrão, pois sua média também é zero. Esta é a distribuição que mais se aproxima graficamente da normal, por este motivo, espera-se que seja uma distribuição com baixa taxa de acerto para amostras pequenas, conforme indica a Figura 3(c). A distribuição qui-quadrado com 4 g.l. é assimétrica a direita e a cauchy $(1, 1)$ é uma distribuição simétrica, respectivamente, Figuras 4(c) e 3(d).



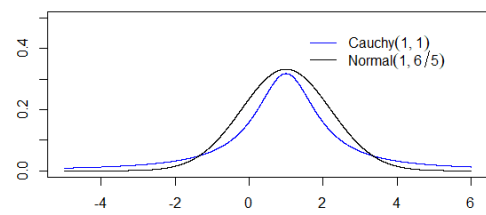
(a) Distribuição Gama



(b) Distribuição Beta

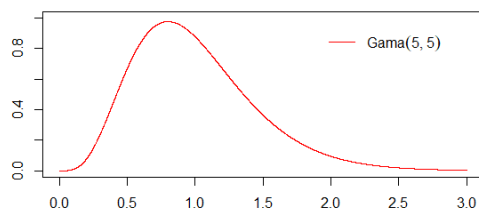


(c) Distribuição T-Student

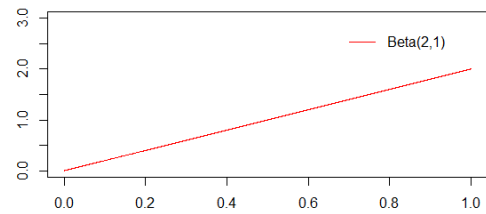


(d) Distribuição Cauchy

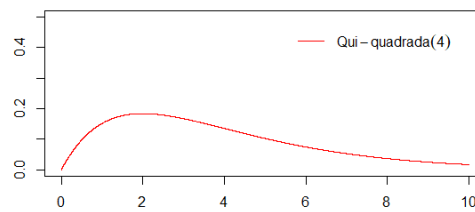
Figura 3: Comparação das curvas das distribuições simétricas com a curva da Normal



(a) Distribuição Gama



(b) Distribuição Beta



(c) Distribuição Qui-quadrada

Figura 4: Comparação das curvas das distribuições simétricas com a curva da Normal

3.2 Metodologia da Simulação

Para realizar o processo de simulação foram definidas as distribuições a serem utilizadas, os tamanhos amostrais e o número de replicações realizadas. Foram simuladas variáveis de 10 distribuições diferentes. A Tabela 2 apresenta estas distribuições e seus parâmetros, que foram escolhidos de modo a gerar curvas muito parecidas com as curvas de distribuições normais, ou totalmente diferentes.

Distribuição	Parâmetros	Simetria da curva
Normal	$(\mu = 0, \sigma^2 = 1)$	Simétrica
Normal	$(\mu = 2, \sigma^2 = 0,5)$	Simétrica
Normal	$(\mu = 3, \sigma^2 = 2)$	Simétrica
T-student	$(\nu = 10)$	Simétrica
Gama	$(\alpha = 10, \beta = 1/3)$	Simétrica
Gama	$(\alpha = 5, \beta = 5)$	Assimétrica
Beta	$(\alpha = 2, \beta = 1)$	Assimétrica
Beta	$(\alpha = 2, \beta = 2)$	Simétrica
Cauchy	$(\gamma = 1, \delta = 1)$	Assimétrica

Tabela 2: Distribuições com parâmetros e simetria da curva utilizadas neste estudo

Para cada distribuição citada acima, vamos considerar 5 tamanhos de amostras distintos, sendo 10, 30, 100, 1000 e 5000.

Para cada distribuição, tamanho amostral e teste de normalidade, 1000 replicações foram feitas e os p-valores armazenados. A eficiência foi analisada através do percentual de vezes que o teste tomou a decisão correta, ou seja, para cada replicação verificou-se se o teste rejeitou a normalidade para os casos em que não era normal e aceitou para os casos em que as distribuições eram normais. O nível de significância utilizado foi de 0,05.

3.3 Resultados

Alguns resultados já esperados se confirmaram, como por exemplo, para os casos das distribuições não normais com tamanho de amostras pequenas a assertividade de todos os testes foi baixa, e as distribuições assimétricas, que possuem curvas muito diferentes da normal, foram as que obtiveram maior taxa de acerto, visto que fica mais fácil para os testes identificar que a curva não é uma normal.

3.3.1 Resultados para a distribuição normal

Independente do tipo de normal escolhida, a taxa de acerto dos testes é sempre a mesma. Diferentes normais foram utilizadas para analisar se os valores de média e variância influenciam no resultado do teste. Logo, os resultados indicam que os testes não são influenciados por estas variáveis, e por este motivo apenas uma das normais utilizadas no estudo é apresentada nas Figuras 5 e 6.

Para as distribuições normais, o teste Shapiro Wilk mostrou-se mais apropriado em casos de amostras muito pequenas, de apenas 10 elementos. Para amostras de tamanho 30 e 100, o teste Lilliefors, apresentou melhor desempenho e para amostras grandes, de 1000 e 5000, o teste Jarque Bera Robusto foi o melhor nesta simulação. É possível verificar estes resultados na Figura 5, ela apresenta as taxas de acerto de cada teste, para cada tamanho de amostra, considerando os resultados encontrados para a distribuição normal.

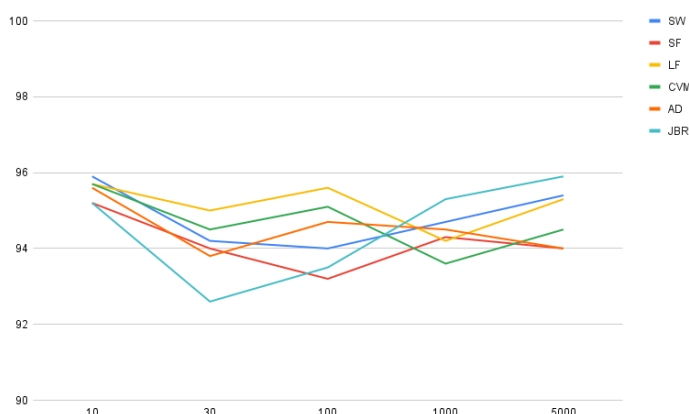


Figura 5: Comparação do resultado da distribuição Normal para os diferentes testes de normalidade

A Tabela 3 apresenta os resultados obtidos da sensibilidade para cada teste e tamanho amostral. Observe que existe pouca diferença entre os resultados dos testes para os casos de amostras pequenas, a taxa de acerto varia entre 95,2% e 95,9%. Logo, qualquer um dos testes apresentaria um resultado com um nível de confiança superior a 95%. Conforme os tamanhos de amostras vão crescendo estes testes vão apresentando maiores diferenças entre os resultados encontrados, observe que para amostras de tamanho 30 apenas o Lilliefors conseguiu obter este nível de confiança.

Tabela 3: Taxa de acerto para a sensibilidade dos testes

Testes	N	Normal (0,1)	Normal (2,1/2)	Normal (3,2)
SW	10	95,9	95,9	95,9
	30	94,2	94,2	94,2
	100	94	94	94
	1000	94,7	94,7	94,7
	5000	95,4	95,4	95,4
SF	10	95,2	95,2	95,2
	30	94	94	94
	100	93,2	93,2	93,2
	1000	94,3	94,3	94,3
	5000	94	94	94
LF	10	95,7	95,7	95,7
	30	95	95	95
	100	95,6	95,6	95,6
	1000	94,2	94,2	94,2
	5000	95,3	95,3	95,3
CVM	10	95,7	95,7	95,7
	30	94,5	94,5	94,5
	100	95,1	95,1	95,1
	1000	93,6	93,6	93,6
	5000	94,5	94,5	94,5
AD	10	95,6	95,6	95,6
	30	93,8	93,8	93,8
	100	94,7	94,7	94,7
	1000	94,5	94,5	94,5
	5000	94	94	94
JBR	10	95,2	95,2	95,2
	30	92,6	92,6	92,6
	100	93,5	93,5	93,5
	1000	95,3	95,3	95,3
	5000	95,9	95,9	95,9

3.3.2 Resultados para as distribuições não normais

Para tamanhos de amostras pequenos, as distribuições não normais apresentaram baixo desempenho em todos os testes. Para as distribuições gama e a qui-quadrada, o teste Shapiro-Francia apresenta o melhor desempenho, para as distribuições betas o Shapiro-Wilk possui as melhores taxas e para a T-Student e a Cauchy o Jarque-Bera Robusto é o melhor. Entretanto, todas essas taxas são inferiores a 70%, conforme a Tabela 4, e por este motivo, não é recomendado apenas o uso destes testes para verificar a normalidade. Nestes casos, recomenda-se o uso de métodos gráficos para a avaliação se a distribuição é uma normal. Em Razali e Wah (2011), o resultado obtido se assemelha

Tabela 4: Taxa de acerto para a especificidade dos testes

Testes	N	T- student (10)	Gama (10,1/3)	Gama (5,5)	Beta (2,1)	Beta (2,2)	Qui- quadrado (4)	Cauchy (1,1)
SW	10	5,7	7,9	11,8	4,5	11,9	21,4	63,7
	30	12,1	20	35,8	6,9	52,8	73,8	96,4
	100	22,8	60,3	89,3	44,9	99,7	100	100
	1000	90,1	100	100	100	100	100	100
	5000	100	100	100	100	100	100	100
SF	10	6,9	9	12,3	3,1	9,6	22,3	68,4
	30	15,7	20,3	33,3	2,2	38,5	69,6	97,2
	100	28	57,2	86,8	19	98,9	99,9	100
	1000	92,2	100	100	100	100	100	100
	5000	100	100	100	100	100	100	100
LF	10	6,4	5,7	7,7	3,8	9,3	14,7	62,3
	30	8,1	12,5	22	5,4	25,6	44,7	94,7
	100	11,7	32,3	59,4	13,5	79,3	95,7	100
	1000	50,9	100	100	100	100	100	100
	5000	100	100	100	100	100	100	100
CVM	10	6,1	6,7	10,3	4	10,1	17,1	66,5
	30	10,4	14,4	26,4	7,3	36,4	58,4	96,7
	100	14,9	42,6	75	22,5	92,9	99	100
	1000	71,6	100	100	100	100	100	100
	5000	100	100	100	100	100	100	100
AD	10	5,8	7,1	10,9	4,6	10,7	19,5	65,7
	30	11,4	16,2	30,5	7,2	43	64,2	96,5
	100	16	47,6	80,9	28,1	98	99,5	100
	1000	78	100	100	100	100	100	100
	5000	100	100	100	100	100	100	100
JBR	10	8,5	7,9	10,9	1,8	5,9	17,8	68,8
	30	18,2	20,6	29,5	0	89	54,4	97,9
	100	33,2	49,2	77,3	0	77,3	98,6	100
	1000	93,8	100	100	100	100	100	100
	5000	100	100	100	100	100	100	100

ao encontrado neste estudo.

O uso de recursos gráficos para decidir se uma distribuição é normal possuem a desvantagem de serem subjetivos, pois dependem da interpretação visual de quem está realizando a análise. Entretanto, para os casos em que as amostras são muito pequenas, ou ainda em conjunto com os resultados obtidos através dos testes de normalidade, esta pode ser uma poderosa ferramenta para garantir que a decisão tomada está correta. Os resultados encontrados neste trabalho convergem com o trabalho de Torman, Coster e Riboldi (2012), em que também recomenda-se o uso de procedimentos não paramétricos para o caso de amostras pequenas.

Alguns exemplos de gráficos que podem ser utilizados para esta análise são o histograma e o quantil-quantil plot, ou simplesmente Q-Q plot. O histograma é um método para verificação do formato da distribuição de uma variável contínua, em que busca-se

verificar se a forma de sino da distribuição Normal está presente. Já o Q-Q plot busca comparar os valores dos quantis observados e esperados, caso a variável tenha distribuição Normal, existirá uma boa aderência dos valores observados com os esperados, ou seja, esses valores encontrados serão muito próximos e o gráfico será aproximadamente uma reta crescente na diagonal.

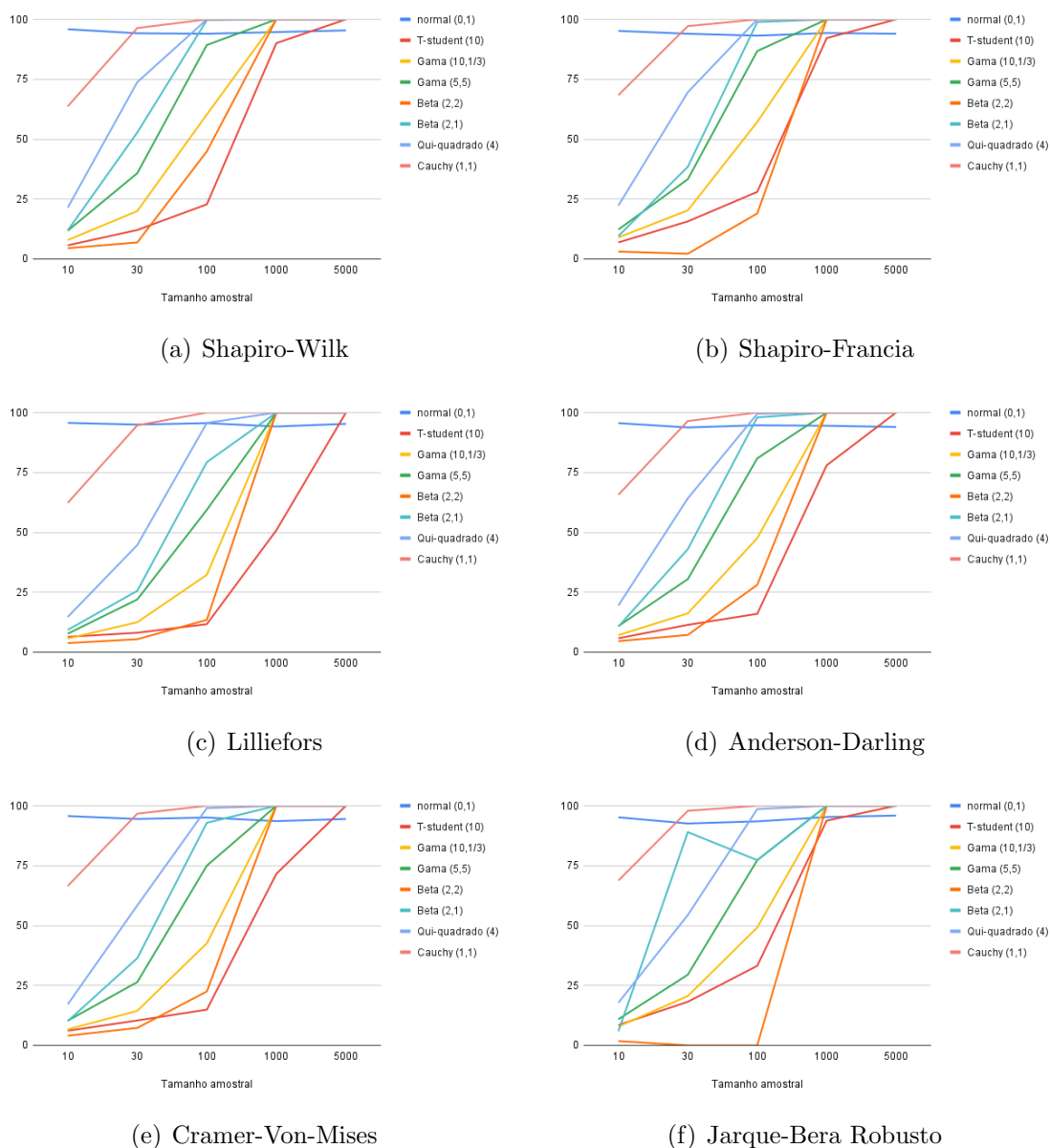


Figura 6: Comparações das taxas de acertos das distribuições para cada um dos testes

Para amostras de tamanho 30, os testes começam a melhorar sua assertividade nos casos em que as distribuições não são normais, porém ainda com taxa de acerto muito baixa para algumas distribuições, como por exemplo, a Beta (2,1) e a T-Student (10). Estas são distribuições que possuem uma curva muito similar a normal, por este motivo é preciso possuir amostras maiores, para que os testes possam ser mais precisos. A única

distribuição não normal que apresenta uma taxa de acerto superior a 90% para este tamanho de amostra é a Cauchy, sendo o teste Jarque Bera Robusto o que possui o melhor desempenho, a Figura 6 explicita esses resultados. Observe que esta Figura mostra os resultados de cada teste, para cada tamanho de amostra escolhido para este estudo.

Para amostras de tamanho 100, em que as distribuições não são normais, o Shapiro Wilk é o mais eficiente, exceto pela T-student, para esta distribuição recomenda-se o teste Shapiro Francia.

Para amostras de 1000 e 5000 o teste Jarque Bera Robusto apresentou o melhor resultado para todas as distribuições. Mantendo a taxa de acerto superior a 95% para as distribuições normais e de 100% para as distribuições não normais. Os resultados aqui encontrados corroboram com Bayoud (2019), que recomenda o teste Jarque Bera Robusto no caso de amostras simétricas.

O teste Jarque Bera Robusto foi o que apresentou melhor desempenho para alguns casos em que a distribuição é muito similar a normal, como por exemplo, para a T-student e a Cauchy, este teste apresentou as maiores taxas para todos os tamanhos de amostra. Logo, este teste pode ser considerado a melhor escolha, ao lidar com distribuições simétricas muito parecidas com a normal, principalmente para as distribuições T-Student e Cauchy.

Para as distribuições assimétricas Gama (5, 5), Beta (2, 1) e Qui-quadrada (4) o teste Shapiro Wilk é o que apresenta os melhores resultados. Yap e Sim (2011), o teste Shapiro Wilk é o mais indicado para distribuições assimétricas. Bayoud (2019), também recomenda o uso deste teste para distribuições assimétricas, no caso de amostras grandes. O teste recomendado para amostras pequenas e moderadas, não foi utilizado neste estudo.

4 Conclusões

O objetivo deste trabalho foi verificar dentre os testes de normalidade escolhidos quais são mais eficientes para os diferentes tamanhos de amostras e distribuições selecionadas. Esta pesquisa pode auxiliar pesquisadores na decisão de qual teste de normalidade utilizar em futuros estudos.

Para realizar a simulação foram definidas as distribuições a serem utilizadas, os tamanhos amostrais e o número de replicações a serem feitas. Foram simuladas variáveis de 10 distribuições diferentes, para 5 tamanhos amostrais distintos. Para cada distribuição e tamanho de amostra, foram realizadas 1000 replicações e o nível de significância utilizado foi de 0,05. A eficiência foi analisada através do percentual de vezes que a decisão correta foi tomada.

Os resultados encontrados indicam que para as distribuições normais os testes Shapiro-Wilk, Lilliefors e Jarque-Bera Robusto apresentaram os melhores desempenhos, o primeiro para amostras de tamanho 10, o segundo para as amostras de tamanho 30 e 100 e o último teste citado para as amostras de tamanho 1000 e 5000.

Para as distribuições que não são normais, nenhum teste apresentou um resultado aceitável para amostras pequenas, todos os testes obtiveram taxas de acerto inferiores a 70%. Logo, recomenda-se o uso de recursos não paramétricos. Já para amostras grandes o teste que apresentou o melhor desempenho foi o Jarque-Bera Robusto.

Dentre as dificuldades encontradas destacam-se a grande quantidade de testes, tamanhos de amostras e distribuições possíveis, não sendo possível testar todos. Além disso, vale ressaltar as limitações dos softwares disponíveis para estas análises, no caso do R, software escolhido para realizar esta simulação, os testes possuem um limite de tamanho de amostra.

Referências

- ALTHOUSE, L. A.; WARE, W. B.; FERRON, W. B. Detecting departures from normality: A monte carlo simulation of a new omnibus test based on moments. *Annual Meeting of the American Educational Research Association*, p. 3–6, 1998.
- ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, JSTOR, p. 193–212, 1952.
- BAYOUD, H. A. Tests of normality: new test and comparative study. *Communications in Statistics - Simulation and Computation*, 2019. Disponível em: <https://doi.org/10.1080/03610918.2019.1643883>.
- D'AGOSTINO, R. B.; STEPHENS, M. A. *Goodness-of-fit-techniques*. [S.l.]: CRC press, 1986. v. 68.
- DALLAL, G. E.; WILKINSON, L. An analytic approximation to the distribution of lilliefors's test statistic for normality. *The American Statistician*, Taylor & Francis, v. 40, n. 4, p. 294–296, 1986.
- DARLING, D. A. The kolmogorov-smirnov, cramer-von mises tests. *The Annals of Mathematical Statistics*, JSTOR, v. 28, n. 4, p. 823–838, 1957.
- DUFOUR, J.-M. et al. Simulation-based finite sample normality tests in linear regressions. *The Econometrics Journal*, Wiley Online Library, v. 1, n. 1, p. 154–173, 1998.
- GEL, Y. R.; GASTWIRTH, J. L. A robust modification of the jarque–bera test of normality. *Economics Letters*, Elsevier, v. 99, n. 1, p. 30–32, 2008.
- JARQUE, C. M.; BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, Elsevier, v. 6, n. 3, p. 255–259, 1980.
- LILLIEFORS, H. W. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, Taylor & Francis, v. 62, n. 318, p. 399–402, 1967.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>.
- RAZALI, N. M.; WAH, Y. B. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2011. Disponível em: <https://www.researchgate.net/publication/267205556>.

- RODRIGUES, D. et al. Aprendizado de máquinas com r. 2022. Disponível em: <https://cienciadedados.uff.br/aprendizado-de-maquinas-com-r/#link9>.
- ROSS, S. M. *A First Course in Probability*. 8. ed. [S.l.]: Prentice Hall, 2010.
- SEIER, E. Comparison of tests for univariate normality. *InterStat Statistical Journal*, v. 1, p. 1–17, 2002.
- SHAPIRO, S. S.; FRANCIA, R. An approximate analysis of variance test for normality. *Journal of the American statistical Association*, Taylor & Francis, v. 67, n. 337, p. 215–216, 1972.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.
- SILVA, R. B. V. Extensão do teste de normalidade de shapiro-francia para o caso multivariado. In: . [S.l.: s.n.], 2014.
- STEPHENS, M. A. The anderson-darling statistic. *Universidade de Stanford*, 1979.
- TORMAN, V. B. L.; BIRCK, A. R.; RIBOLDI, J. Comparação dos testes de aderência à normalidade kolmogorov-smirnov, anderson-darling, cramer–von mises e shapiro-wilk por simulação. *Anais do 11^o Simpósio de Estatística Aplicada à Experimentação Agrônômica*, 2005. Disponível em: <https://pt.scribd.com/document/134412762>.
- TORMAN, V. B. L.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. *Revista HCPA*, 2012. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/29874>.
- VIALI, L. Algumas considerações sobre a denominada curva normal. *Vidya (Santa Maria. Impreso)*, p. 99–116, 2014.
- YAP, B. W.; SIM, C. H. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 81, n. 12, p. 2141–2155, 2011.

ANEXO 1 – Tabela para o valor-p do teste Lilliefors

Em Dallal e Wilkinson (1986) é possível encontrar mais detalhes sobre como esta tabela foi construída.

**Table 2. The Analytic Approximation and Lilliefors's Simulation
Compared With the Dallal–Wilkinson Simulation**

<i>Sample size</i>	<i>Nominal level</i>					
	<i>.100</i>		<i>.050</i>		<i>.010</i>	
	<i>L</i>	<i>DW</i>	<i>L</i>	<i>DW</i>	<i>L</i>	<i>DW</i>
4	.089	.110	.044	.050	.008	.017
5	.112	.107	.061	.055	.007	.010
6	.110	.107	.057	.050	.013	.010
7	.114	.104	.058	.050	.011	.009
8	.114	.103	.056	.050	.011	.009
9	.112	.104	.056	.050	.013	.010
10	.108	.103	.058	.050	.015	.009
11	.106	.104	.056	.051	.014	.010
12	.100	.105	.051	.051	.013	.010
13	.104	.102	.050	.050	.012	.010
14	.104	.102	.049	.051	.011	.010
15	.103	.105	.049	.051	.009	.010
16	.103	.105	.050	.051	.009	.010
17	.106	.104	.053	.051	.008	.010
18	.107	.104	.055	.050	.008	.010
19	.109	.102	.055	.051	.007	.010
20	.113	.105	.056	.052	.007	.010
25	.077	.103	.036	.052	.009	.010
30	.113	.102	.046	.051	.009	.010
40	.102	.097	.047	.050	.009	.010
100	.113	.094	.053	.049	.011	.010
400	.123	.104	.059	.050	.012	.011
900	.130	.093	.061	.054	.014	.011

NOTE: L represents the actual significance levels of Lilliefors's critical values. DW represents the analytic approximation for percentiles in Table 1.

**ANEXO 2 - Tabela da distribuição
acumulada da normal padrão**

