Brazilian Journal of Development

# Uma aplicação do teorema central do limite

# An application of the central limit theorem

DOI:10.34117/bjdv5n12-293

Recebimento dos originais: 15/11/2019
Aceitação para publicação: 18/12/2019

**Denilson Junio Marques Soares**
Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Viçosa
Doutorando em Educação pela Universidade Federal do Espírito Santo
Professor do Instituto Federal de Minas Gerais
Endereço: Rua Severo Veloso, 1880, Piumhi-MG, Brasil
E-mail: denilson.marques@ifmg.edu.br

**Talita Emidio Andrade Soares**
Mestranda em Educação pela Universidade Federal do Espírito Santo
Endereço: Rua Severo Veloso, 1880, Piumhi-MG, Brasil
E-mail: talitaeandrade@gmail.com

**Paulo César Emiliano**
Doutor em Estatística e Experimentação Agropecuária pela Universidade Federal de Lavras,
Professor Adjunto da Universidade Federal de Viçosa
Endereço: Avenida P.H. Rolfs, s/n., Viçosa-MG, Brasil
E-mail: paulo.emiliano@ufv.br

**RESUMO**

O Teorema Central do Limite é um importante teorema da inferência estatística que diz que quanto maior o tamanho de uma determinada amostra, mais próxima estará de uma distribuição normal, a distribuição amostral de sua média. Neste artigo, propõe-se utilizar este teorema para mostrar que a distribuição amostral do tamanho médio de folhas da árvore Dypsis lutescens, popularmente conhecida como palmeira areca, areca-bambu ou palmeira de jardim, converge para uma distribuição normal. Para isto, recolheu-se aleatoriamente uma certa quantidade de folhas desta árvore e, após uma análise descritiva do tamanho dessas folhas, mensurado com o auxílio de um paquímetro, aplicou-se o teste de Kolmogorov-Smirnov que, a um nível de significância de 5%, confirmou a hipótese de normalidade dos dados. As análises foram realizadas com o auxílio do *software* estatístico R. Espera-se que este trabalho possa ilustrar as aplicabilidades do Teorema Central do Limite trazendo ganhos para o processo de ensino-aprendizagem da estatística, especialmente em cursos da área de Ciências Agrárias.

**Palavras-chave**: Teorema Central do Limite. Palmeira Areca. Distribuição Normal.

**ABSTRACT**

The central limit theorem is an important result from statistical inference that states that the distribution of sample means approximates a normal distribution as the sample size gets larger. In the present paper we propose to make this theorem to show an average leaf size distribution of the Dypsis lutescens tree, popularly known as areca palm, areca bamboo or garden palm, converges to a normal distribution. For this, a certain amount of leaves were randomly collected from this tree and, after a descriptive analysis of the size of these leaves, measured with the aid of a caliper, the Kolmogorov-Smirnov test was applied which, at a significance level of 0.05, confirmed the hypothesis of data normality.The analysis were performed using R Statistical Software. It is hoped that this work may contribute to the understanding of the applicability of the Central Limit Theorem thus providing quality gains to the teaching-learning process in statistics, especially in Agricultural Science.

**Keywords:** Central Limit Theorem. Areca Palm Tree. Normal Distribution.

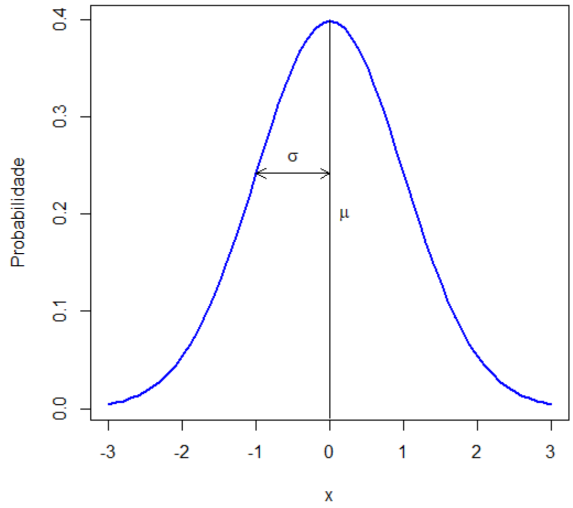
## 1. INTRODUÇÃO

A distribuição normal de probabilidades descreve uma série de fenômenos físicos, financeiros e biológicos nos proporcionando uma ampla variedade de aplicações estatísticas. De acordo com Magalhães (2015, p. 102), uma variável X segue o modelo normal ($\boldsymbol{X\sim N(\mu,\sigma)}$) de parâmetros de média e variância da variável representados por $\boldsymbol{\mu}$ e $\boldsymbol{\sigma^2}$, respectivamente, quando sua função densidade de probabilidade é a que segue:

$$\boldsymbol{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} I_{(-\infty,+\infty)}(x),} \qquad \textbf{(1)}$$

em que $\boldsymbol{\mu, \sigma} \in \mathbb{R}, \boldsymbol{\sigma > 0}$. Quando $\boldsymbol{\mu = 0}$ e $\boldsymbol{\sigma = 1}$ a distribuição normal é dita padrão, cujo gráfico está representado abaixo:

Gráfico 1 - Distribuição Normal Padrão




Para a verificação da normalidade de um conjunto de dados, pode-se utilizar o teste de Kolmogorov-Smirnov (LILLIEFORS, 1967), cuja hipótese de nulidade é a de que os dados seguem a uma distribuição normal de probabilidade e cuja estatística quantifica a distância entre a função distribuição empírica da amostra e a função distribuição acumulada assumida para os dados, no caso a Normal. Esta distância é comparada pautada em um valor crítico, para um dado nível de significância, comumente adotado com o valor de 5%.

O Teorema Central do Limite (TCL) diz que quanto maior o tamanho de uma determinada amostra, mais próxima estará de uma distribuição normal, a distribuição amostral de sua média (BUSSAB; MORETTIN, 2013; FERREIRA, 2009). Em outras palavras, quando o tamanho amostral é suficientemente grande, a distribuição da média é uma distribuição aproximadamente normal.

Neste artigo pretende-se incialmente demonstrar este teorema e, em seguida, apresentar uma aplicação empírica que visa mostrar que a distribuição amostral do tamanho médio de folhas da árvore *Dypsis lutescens*, popularmente conhecida como palmeira areca, areca-bambu ou palmeira de jardim, converge para uma distribuição normal.

## 2. O TEOREMA CENTRAL DO LIMITE

O teorema central do limite é um dos resultados mais extraordinários na teoria da probabilidade, por nos permitir conduzir alguns procedimentos de inferência estatística sem qualquer conhecimento da distribuição da população. Em linhas gerais, trata-se da convergência

Brazilian Journal of Development

em distribuição para o modelo Normal de uma soma de variáveis aleatórias independentes, após uma conveniente padronização.

**Teorema:** Sejam $\{X_n : n \geq 1\}$ variáveis aleatórias independentes, identicamente distribuídas e com esperança $\mu$ e variância $\sigma^2$, com $0 < \sigma^2 < \infty$. Então, para $S_n = X_1 + X_2 + ... + X_n$, temos,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Em outras palavras, para $-\infty < a < \infty$,

$$P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx \text{ quando } n \to \infty.$$

Para demonstrarmos o teorema iremos utilizar o lema abaixo enunciado e cuja demonstração pode ser encontrada em Mood (1974).

**Lema**: Sejam $Z_1,\ Z_2,...$ uma sequência de variáveis aleatórias com funções distribuição $F_{Z_n}$ e funções geradoras de momentos $m_{Z_n}, n \geq 1$; seja também $Z$ uma variável aleatória com função distribuição $F_z$ e função geradora de momentos $m_Z$. Se $m_{Z_n}(t) \to m_Z(t)$ para todo $t$, então $F_{Z_n}(t) \to F_Z(t)$ para todo $t$ no qual $F_Z(t)$ é contínua.

Se $Z$ é uma variável aleatória normal padrão, então, como $m_Z(t) = e^{t^2/2}$, obtemos do Lema acima que, se $m_{Z_n}(t) = e^{t^2/2}$ quando $n \to \infty$, então $F_{Z_n}(t) \to \Phi(t)$ quando $n \to \infty$.

Finalmente, podemos agora demonstrar o teorema central do limite.

Para isto, vamos primeiramente supor que $\mu = 0$ e $\sigma^2 = 1$ e considerar que a função geradora de momentos de $X_i$, $m(t)$, exista e seja finita. Assim, a função geradora de momentos de $X_i / \sqrt{n}$ é dada por

$$E\left[\exp\left\{\frac{tX_i}{\sqrt{n}}\right\}\right] = m\left(\frac{t}{\sqrt{n}}\right)$$

Brazilian Journal of Development

Logo, a função geradora de momentos de $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ é dada por $\left[ m\left( \frac{t}{\sqrt{n}} \right) \right]^n$. Considere

$$L(t) = \ln m(t)$$

e observe que

$$L(0) = \ln m(0) = \ln E\left[ e^{0X} \right] = \ln E[1] = \ln 1 = 0$$

$$L'(0) = \left( \ln m(0) \right)' = \frac{m'(0)}{m(0)} = \mu = 0$$

$$L''(0) = \left( \ln m(0) \right)'' = \left( \frac{m'(0)}{m(0)} \right)' = \frac{m''(0)m(0) - \left[ m'(0) \right]^2}{\left[ m(0) \right]^2} = \frac{E\left[ X^2 \right] - 0^2}{1^2} = E\left[ X^2 \right]$$

$$= E\left[ (X - 0)^2 \right] = E\left[ (X - \mu)^2 \right] = \sigma^2 = 1$$

Finalmente, para demonstrar o teorema, devemos mostrar que $\left[ m\left( \frac{t}{\sqrt{n}} \right) \right]^n \to e^{t^2/2}$ quando $n \to \infty$, ou, equivalentemente, que $nL\left( \frac{t}{\sqrt{n}} \right) \to t^2/2$ quando $n \to \infty$. Para mostrarmos isso, observe que

$$\lim_{n\to\infty} \frac{L\left(t/\sqrt{n}\right)}{n^{-1}} \overset{\substack{\text{Regra de} \\ \text{L´Hôpital}}}{=} \lim_{n\to\infty} \frac{-L'\left(t/\sqrt{n}\right) n^{-3/2} t}{-2n^{-2}} = \lim_{n\to\infty} \frac{L'\left(t/\sqrt{n}\right) t}{2n^{-1/2}} \overset{\substack{\text{Regra de} \\ \text{L´Hôpital}}}{=} \lim_{n\to\infty} \frac{-L''\left(t/\sqrt{n}\right) n^{-3/2} t^2}{-2n^{-3/2}}$$

$$= \lim_{n\to\infty} L''\left( \frac{t}{\sqrt{n}} \right) \frac{t^2}{2} = \frac{t^2}{2}$$

Assim, o teorema central do limite está demostrado quando $\mu = 0$ e $\sigma^2 = 1$. O resultado pode ser estendido para os demais casos. Basta definirmos uma nova variável aleatória $Y = \left(X_i - \mu\right) / \sigma$ que teria média zero e variância 1.

## 3. UMA APLICAÇÃO DO TEOREMA CENTRAL DO LIMITE

Como visto, este artigo se propõe a apresentar uma aplicação empírica do TCL, envolvendo o comprimento de folhas de uma árvore selecionada. Para isto, considerou-se a Dypsis lutescens, uma árvore popularmente conhecida como palmeira areca, areca-bambu ou palmeira de jardim. A escolha desta árvore se deu pela presença expressiva no campus da Universidade Federal de Viçosa, onde o experimento foi realizado. Além disso, trata-se de uma árvore de fácil acesso e com folhas que podem ser medidas de maneira simples e precisa.

Recolheu-se aleatoriamente 1580 folhas, distribuídas em 35 galhos da árvore, atentando-se às diferentes idades, tamanhos e demais características que pudessem afetar a pesquisa. As Figuras 1 e 2 representam a árvore e os galhos coletados para compor a amostra, respectivamente.

Figura 1 – Árvore Dypsis lutescens

Figura 2 – Galhos coletados para compor a amostra

Em seguida, realizou-se as medições das folhas com o auxílio de um paquímetro, instrumento constante de uma escala graduada fixa, duas garras e um cursor com um nônio, utilizado para medir precisamente pequenas distâncias. A Figura 3 ilustra este instrumento e a Figura 4 parte das folhas mensuradas.

Figura 3 - Paquímetro

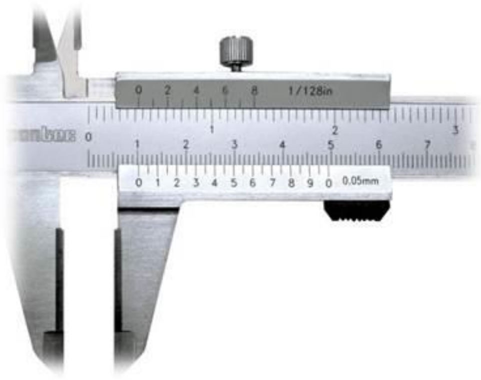
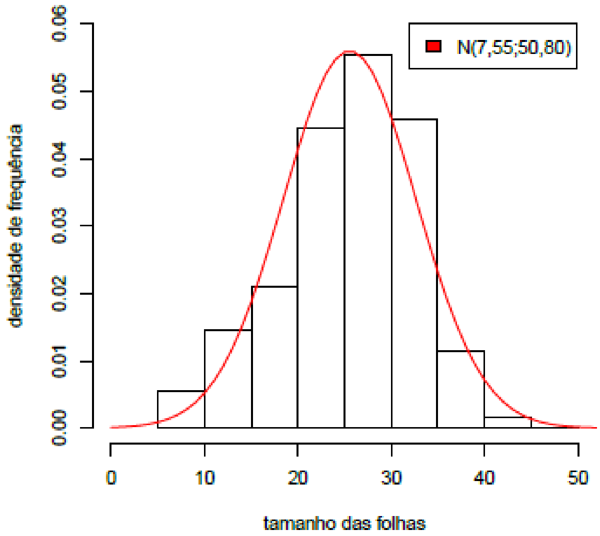



Figura 4 - Folhas mensuradas

Para verificar se a distribuição amostral do tamanho médio de folhas da árvore converge para uma distribuição normal, realizou-se o teste de Kolmogorov-Smirnov. Para isto, utilizou-se a função ks.test, implementada no pacote stats do software estatístico R (R Core Team, 2019). A utilização desse software se fez pela facilitação de seu acesso e por ser um software livre e de código aberto. O software retornou estatística $D = 0{,}5473$ e valor-p $= 0{,}9257$. Dessa forma, não encontrou-se evidências significativas para rejeitar a hipótese de normalidade.

Em seguida, através do método da máxima verossimilhança, procedeu-se à estimação dos parâmetros da distribuição normal, obtendo-se média $\mathbf{25,55}$ e variância $\mathbf{50,80}$ da normal que melhor se ajusta aos dados. O Gráfico 2 representa o histograma e a normal ajustada para os dados deste experimento.

Gráfico 2 - Histograma e normal ajustada para o tamanho das folhas da *Dypsis lutescens*




## 4. CONSIDERAÇÕES FINAIS

Neste artigo propôs-se à verificar, empiricamente, se a distribuição amostral do tamanho médio das folhas da *Dypsis lutescens* converge em distribuição para uma distribuição normal. Com êxito, os resultados obtidos confirmaram esta hipótese.

Este experimento pode ilustrar, didaticamente, o Teorema Central do Limite, pois trata-se de uma aplicação direta do mesmo. Em cursos como os das ciências agrárias que possuem uma parte estatística sólida, experimentos deste tipo podem contribuir significativamente para o processo ensino-aprendizagem, por apresentarem uma ligação mais direta entre a teoria e a prática.

Além de concluirmos que o tamanho médio das folhas da segue assintoticamente a distribuição normal, percebemos o quanto essa distribuição está presente e se aplica nas formas da natureza, ressaltando sua importância.

## REFERÊNCIAS

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica.** 8a ed. São Paulo, Saraiva, 2013. 548 p.

FERREIRA, D. F. Estatística básica. Lavras: Editora UFLA, 2009. 664 p.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. **Journal of the American Statistical Association**, v. 62, n. 318, p. 399-402, 1967.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. 3. ed. São Paulo, SP: Editora da Universidade de São Paulo, 2015. 428 p.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the Theory of Statistics**, Singapura: 1974.

R DEVELOPMENT CORE TEAM. **R:** a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2019. Disponível em http://www.r-project.org. Acesso em: 14 out 2019.