

Teorema do limite central aplicado em notas finais de alunos na disciplina de Estatística.

Lucas Glauber da Silva do Carmo, Elizangela Cabral dos Santos, Genevile Carife Bergamo

Resumo: O Teorema do Limite Central (TLC) é um princípio fundamental da estatística que afirma que, ao coletar uma variável aleatória repetidamente, a distribuição das médias amostrais se aproxima de uma distribuição normal à medida que o tamanho da amostra aumenta. Este estudo visa demonstrar a aplicação prática do TLC utilizando notas finais de alunos na disciplina de Estatística, abrangendo 546 estudantes entre os semestres de 2020.1 a 2023.1. Para a análise, foram realizadas 4000 amostragens aleatórias com reposição, em cada amostra com $n = 5, 10, 20, 30$ e 40 notas. As médias amostrais foram calculadas e plotadas em histogramas. À medida que o tamanho da amostra foi aumentando foi possível notar uma aproximação maior da distribuição das médias para uma normal, corroborando as propriedades do TLC. A pesquisa reforça a importância do TLC na estatística inferencial, permitindo generalizações sobre a população com base em amostras representativas. A metodologia utilizada, incluindo o processamento de dados no Excel, mostrou a viabilidade de técnicas estatísticas em contextos acadêmicos. Os resultados destacam a relevância da amostragem e da análise estatística na compreensão do desempenho educacional, sublinhando a robustez do TLC em aplicações práticas.

Palavras-chave: teorema do limite central; distribuição normal; amostragem.

1. INTRODUÇÃO

O Teorema do Limite Central é um importante conceito na estatística que descreve o comportamento das médias amostrais retiradas de uma população. Ele afirma que, mesmo que a distribuição original da população não seja normal, a distribuição das médias amostrais tende a uma distribuição normal à medida que o tamanho das amostras aumenta [1]. Essa convergência de outras distribuições para a distribuição normal ocorre pois ao coletarmos um tamanho 'n' amostral maior conseguimos encontrar todos os tipos de variação dos dados incluindo os scores mais extremos, ou seja, se pegarmos uma variável com distribuição mesmo que assimétrica, e repetirmos esse estudo muitas vezes, todos os picos irão se concentrarem em um certo intervalo, tanto maior quanto menor que a média central, por conta da variabilidade entre as amostras. Assim como os scores mais extremos tanto superiores quanto inferiores continuarão com baixa frequência por conta de serem mais raros [1].

O TLC é um pilar da estatística inferencial, um ramo fundamental da Estatística que se concentra na realização de inferências e conclusões sobre uma população a partir de uma amostra representativa. Essa abordagem é crucial, pois permite que possamos generalizar informações coletadas de um grupo menor para um universo maior, oferecendo uma compreensão mais ampla e fundamentada de um fenômeno. Vários matemáticos fizeram contribuições significativas para o desenvolvimento deste teorema, que levou quase 200 anos desde sua primeira concepção até seu completo aperfeiçoamento. A primeira formulação do teorema foi proposta pelo matemático francês Abraham Moivre em 1733. Posteriormente, o renomado matemático alemão Carl Friedrich Gauss, conhecido como o "príncipe da matemática", associou seu nome à curva normal, ou gaussiana, que fundamenta o teorema. Em 1812, Pierre Simon de Laplace, outro destacado matemático francês, se inspirou nos trabalhos de Gauss para suas próprias contribuições. Em 1901, o matemático russo Aleksandr Lyapunov elucidou a base matemática do teorema do limite central. Somente nos anos 1930, o matemático francês Paul Lévy estabeleceu as condições gerais de aplicabilidade do teorema, entretanto, muito antes dessas contribuições, outros matemáticos, como Blaise Pascal, que precedeu Moivre, e Daniel Bernoulli, que veio antes dos demais, desempenharam papéis importantes na construção desse teorema [2].

O intuito desse trabalho é mostrar na prática a aplicação do TLC utilizando as notas finais de alunos na disciplina de estatística, com o auxílio da ferramenta EXCEL.

2. DESENVOLVIMENTO

A distribuição normal representa a distribuição de probabilidades mais importante e mais utilizada dentre todas as distribuições de probabilidade. Muitos fenômenos no mundo real são exatamente ou aproximadamente distribuídos de forma normal. Como por exemplo: variáveis aleatórias contínuas que representam peso e altura de pessoas; resultados em um exame; vida útil de um item; tempo gasto para terminar determinada tarefa [3]. Uma variável seguindo uma distribuição Normal é representada por $X \sim N(\mu; \sigma^2)$ em que: μ é a média e σ^2 a variância da variável e sua função densidade de probabilidade é dada por [1].

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad (1)$$

em que $\mu, \sigma \in R, \sigma > 0$. Quando $\mu = 0$ e $\sigma = 1$ a distribuição normal padrão está representada na Figura 1.

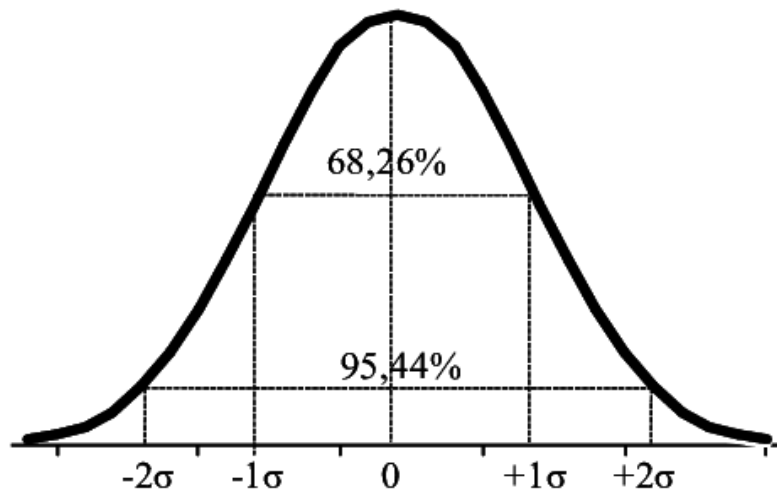


Figura 1. Distribuição Normal Padronizada. [3]

O benefício de conseguir aproximar uma distribuição qualquer para uma normal é que conhecemos as propriedades da curva normal e como ela se comporta. Por exemplo, sabemos que a área total sob uma curva da distribuição normal é igual a 1 ou 100%. Sabemos também que essa curva é simétrica em torno da média aritmética e que as caudas da distribuição normal se estendem infinitamente em ambas as direções, ou seja, assintótica em relação ao eixo horizontal. A média, a moda e a mediana da distribuição gaussiana são iguais. É sabido também que 95% das observações vão estar contidas entre a média e aproximadamente ± 2 desvios padrão.

Para a aplicação do teorema do limite central é necessário satisfazer algumas condições, como a independência entre as variáveis, ou seja, a ocorrência ou não ocorrência de uma variável não interfere em nada na ocorrência ou não ocorrência de outra variável [4]. Outra condição é a aleatoriedade, a escolha de cada elemento da amostra deve ser aleatória, o que significa que cada elemento da população tem a mesma probabilidade de ser selecionado. Na análise de dados, a aleatoriedade é fundamental para a interpretação dos resultados, pois garante que a amostra seja representativa da população em estudo. Ao coletar dados de maneira aleatória, reduz-se o risco de viés, permitindo que as análises estatísticas sejam aplicadas com maior confiança. Dessa forma, a aplicação rigorosa de métodos aleatórios contribui para conclusões mais robustas e generalizáveis. Outra condição é para amostras suficientemente grandes, quando o número de amostras é de pelo menos $n = 30$ elementos, assim, a distribuição das médias amostrais aproxima-se satisfatoriamente da distribuição normal [2].

Ao se fazer infinitas amostragens e calcular a média de cada amostra, na distribuição destas médias amostrais serão observadas três propriedades:

1. A média na distribuição das médias amostrais é igual a média da população.
2. O desvio padrão da distribuição das médias amostrais será igual ao desvio padrão da população dividido por raiz de 'n'.
3. Independe da distribuição da variável na população, se 'n' for maior do que 30 elementos, a distribuição das médias amostrais vai tender para uma distribuição normal [2].

2.1. Metodologia

Para verificar o Teorema do Limite Central, foi realizado um estudo com as notas de várias turmas de alunos na disciplina de Estatística, utilizando dados coletados de semestres que abrangem de 2020.1 a 2023.1. O processo de análise envolveu as seguintes etapas:

Coleta de Dados: As notas finais foram obtidas a partir de registros acadêmicos, calculando-se a média final de cada aluno na disciplina abrangendo um total de 546 alunos.

Amostragem: Foram realizadas 4000 amostragens aleatórias com reposição, em cada amostra com $n = 5, 10, 20, 30$ e 40 notas e para cada amostra calculou-se a média das notas.

Análise Estatística: As médias amostrais foram plotadas em histogramas e avaliadas quanto à sua distribuição comparando-se com a distribuição normal.

Para fazer o sorteio aleatório das amostras foi usada a função ALEATÓRIOENTRE, como mostra a Figura 2

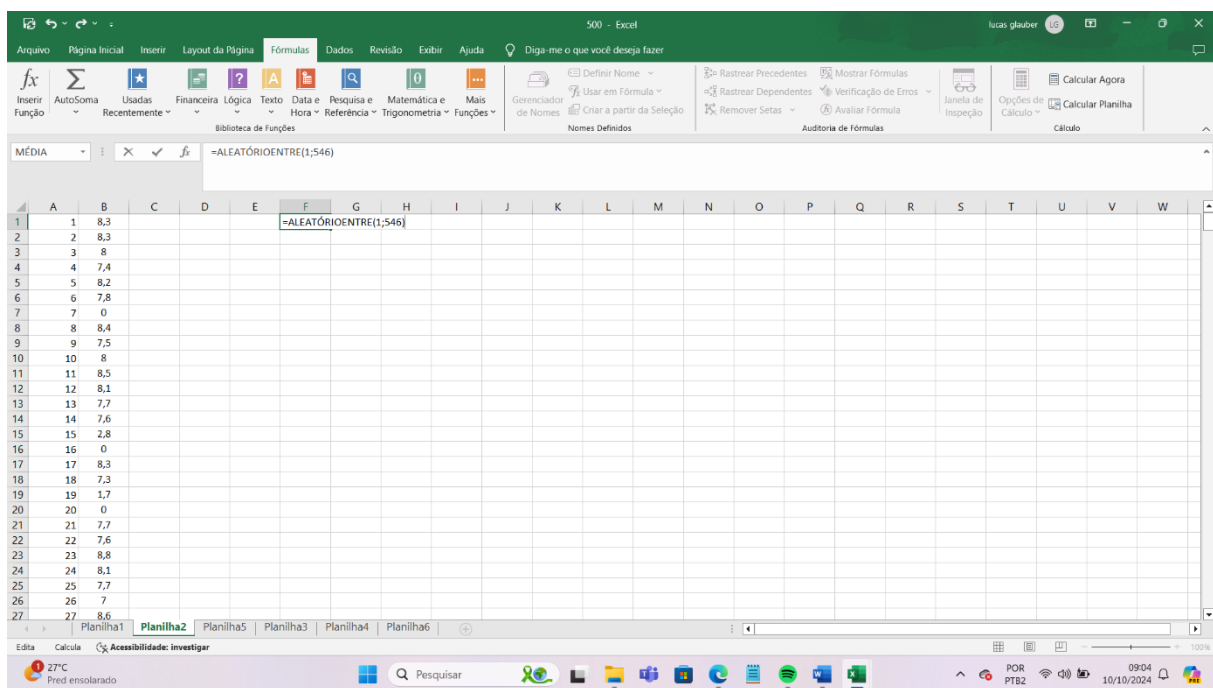


Figura 2. Função ALEATÓRIOENTRE no EXCEL

Em seguida foi usada a função PROCV para retornar os valores correspondentes aos números aleatórios, como mostra a Figura 3

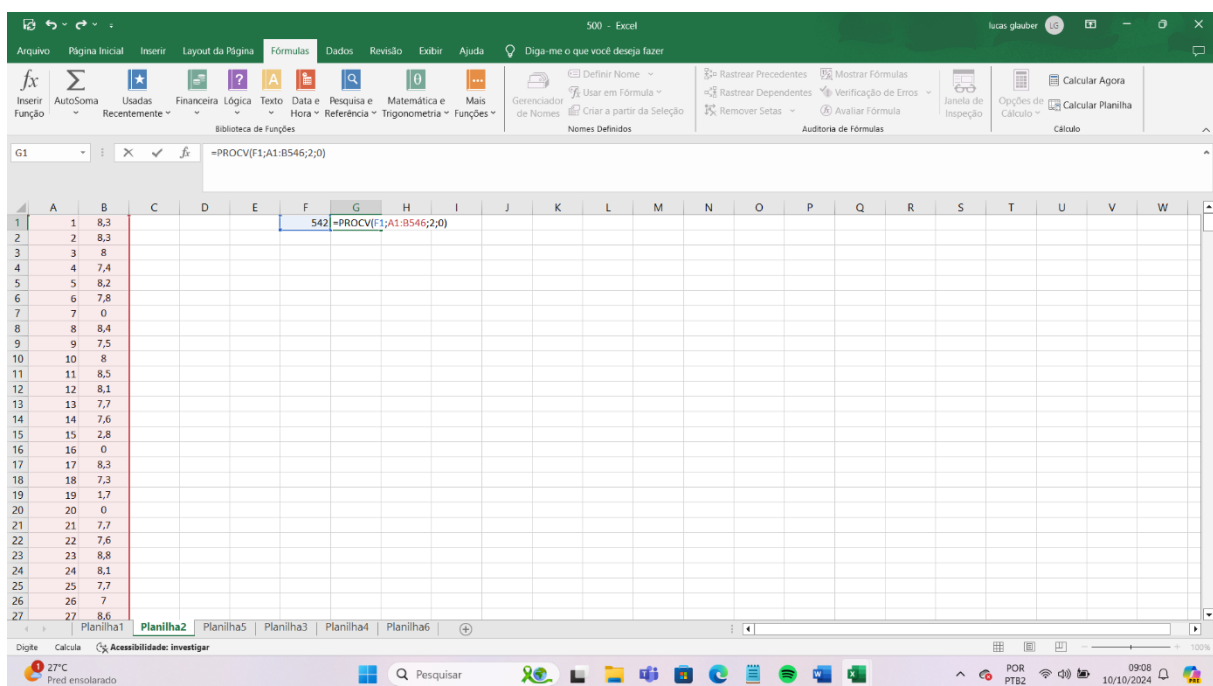


Figura 3. Função PROCV no EXCEL

Em seguida foi utilizada a função média para calcular a média da amostra, como mostra a Figura 4.

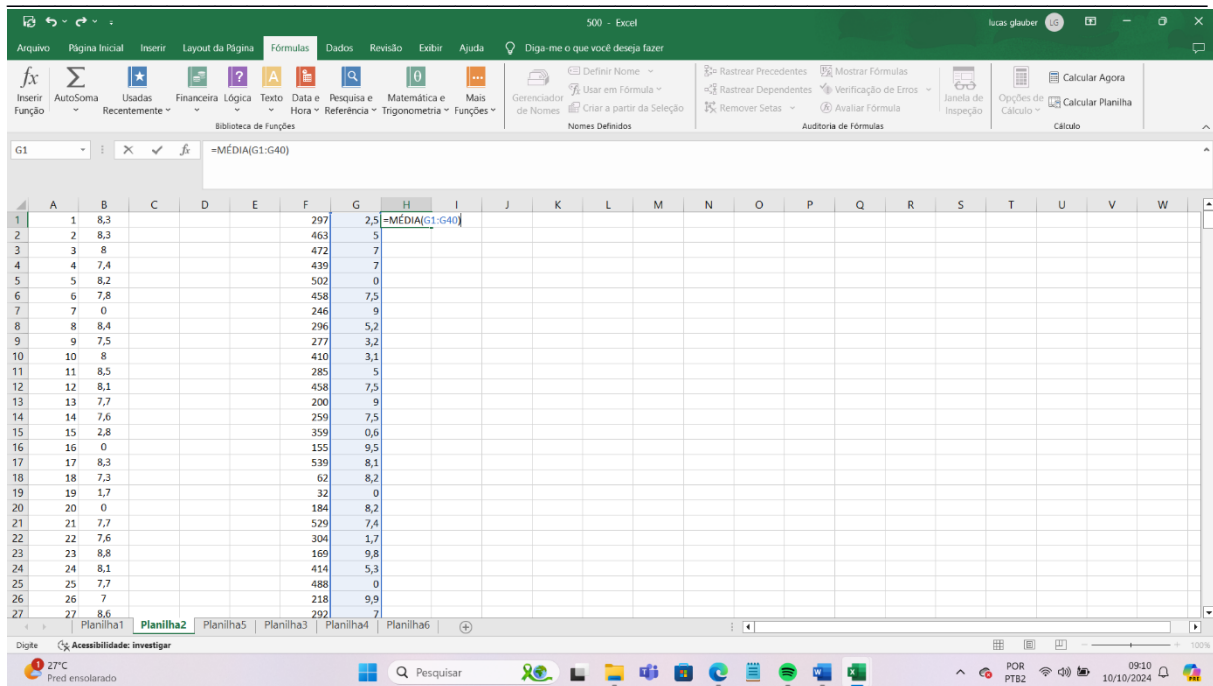


Figura 4. Função MÉDIA no EXCEL

Esse processo foi repetido 4.000 vezes para cada tamanho amostral, e ao final foi calculada a média das médias amostrais e o desvio padrão das médias amostrais e as distribuições das médias amostrais foram representadas em histogramas, permitindo a visualização e análise da tendência das médias em relação à uma distribuição normal.

Para calcular a média das médias amostrais foi usada novamente a função MÉDIA, como mostra a Figura 5

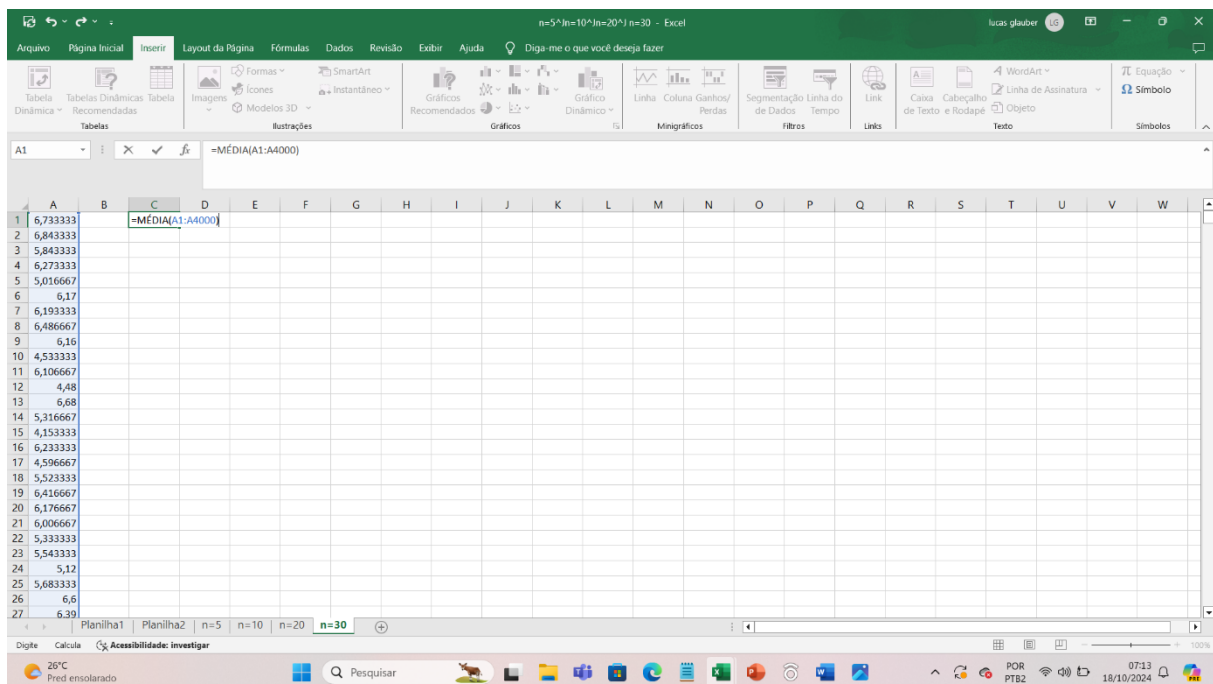


Figura 5. Função MÉDIA no EXCEL

Para calcular o desvio padrão das médias amostrais foi usada a função DESVPAD.P, como mostra a Figura 6

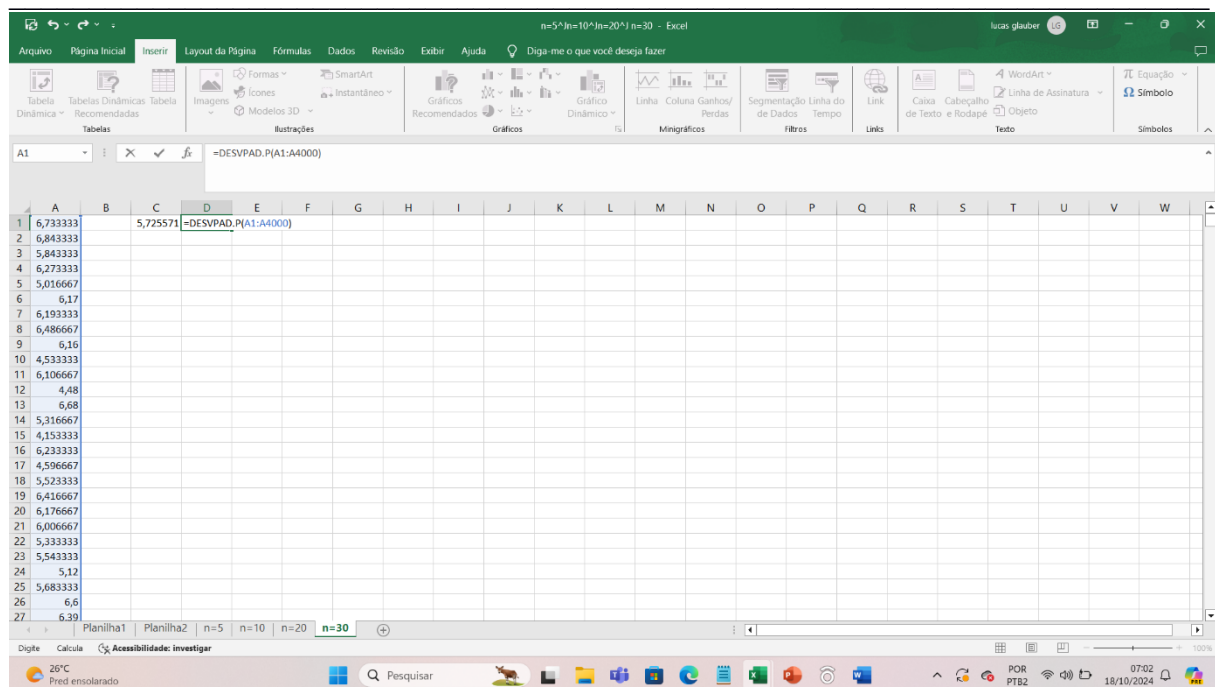


Figura 6. Função DESVPAD.P do Excel

2.2 Resultados e Discussão

Na Figura 6, observa-se a distribuição das notas finais, esta possui uma distribuição assimétrica em forma de “U”, portanto em nada semelhante à distribuição simétrica de uma Normal.

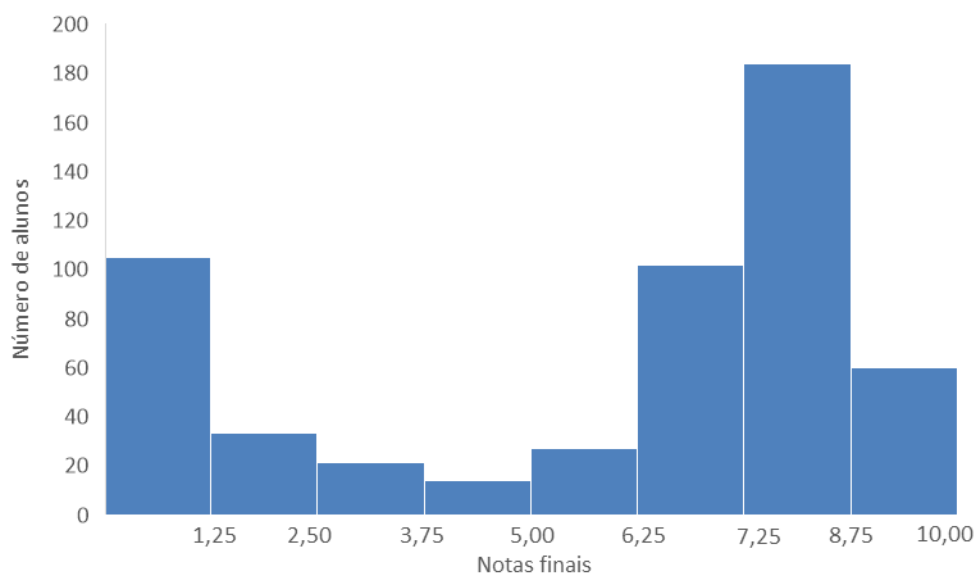


Figura 7. Distribuição da Notas

As Figuras com distribuição amostral das médias para $n = 5$, $n = 10$, $n = 20$, $n = 30$ e $n = 40$ estão representados abaixo.

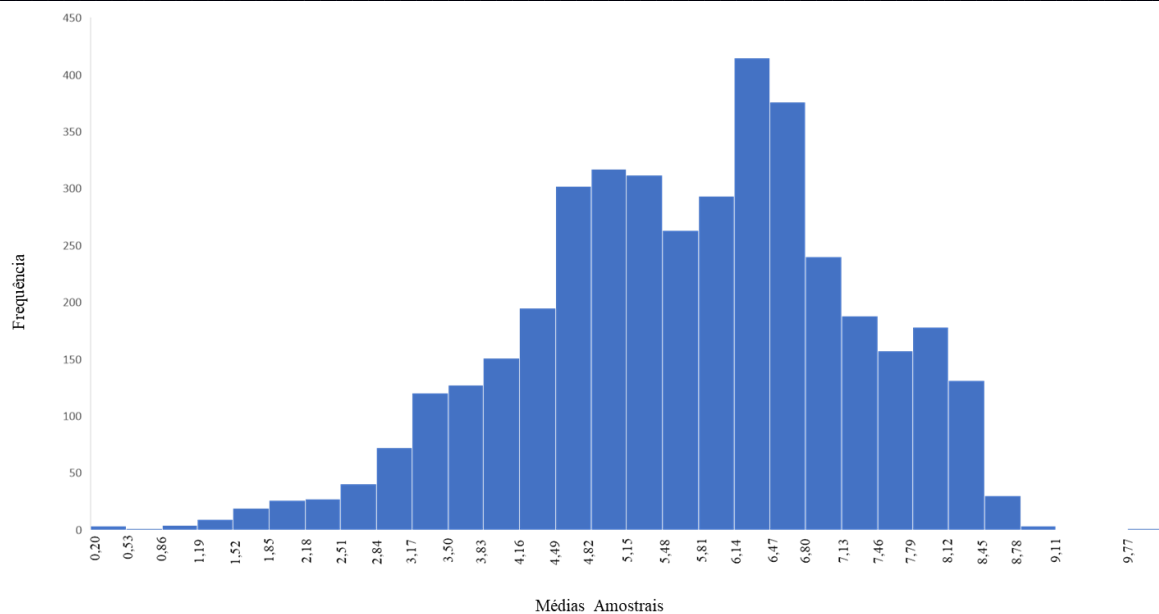


Figura 8. Distribuição amostral das médias com tamanho amostral $n = 5$

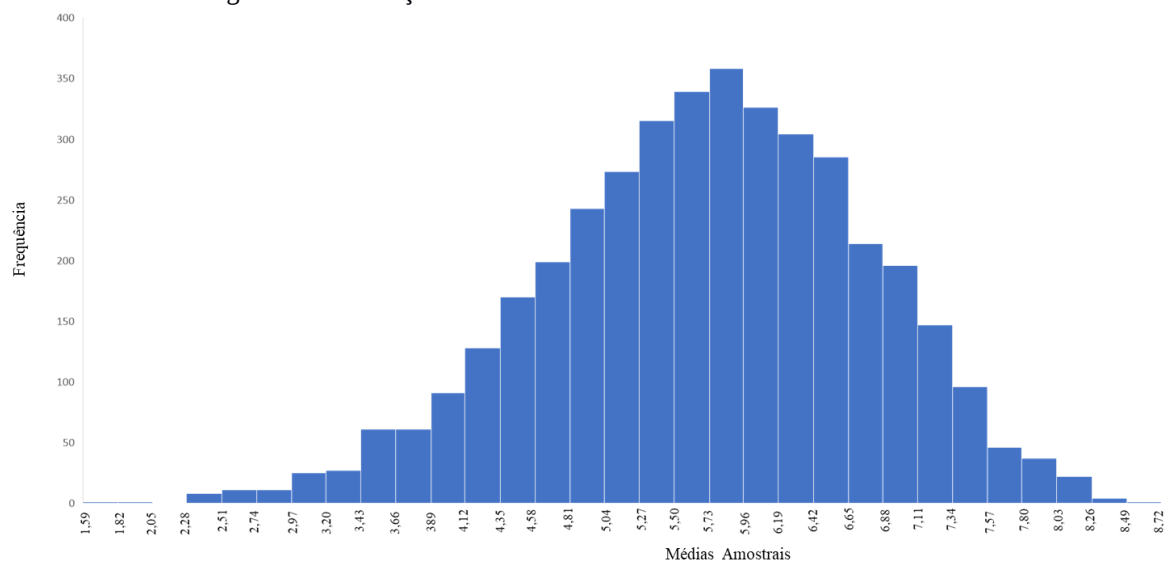


Figura 9. Distribuição amostral das médias com tamanho amostral $n = 10$

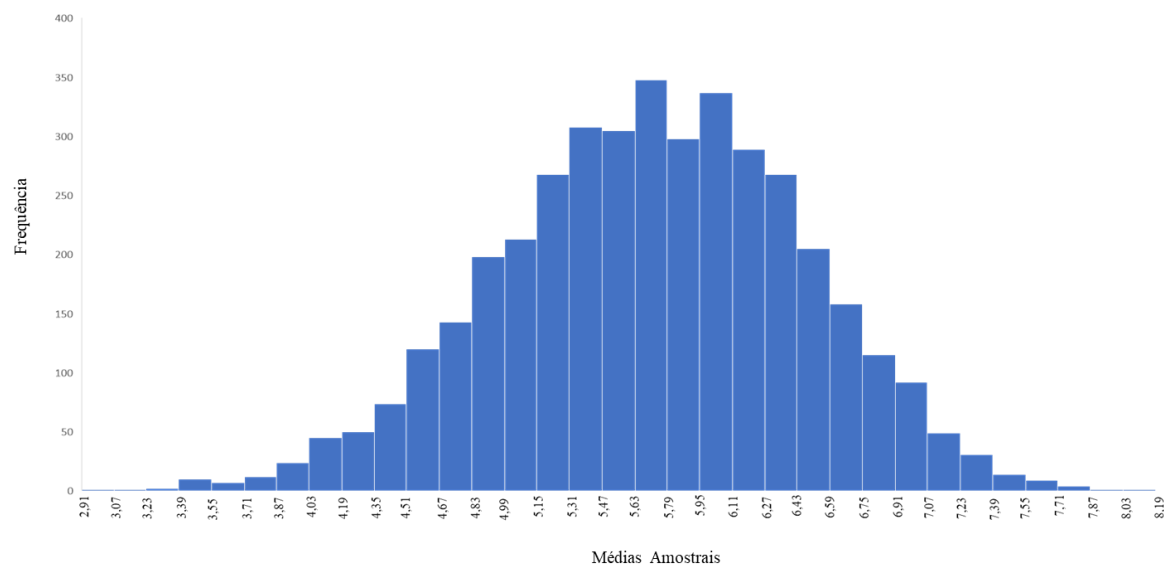


Figura 10. Distribuição amostral das médias com tamanho amostral $n = 20$

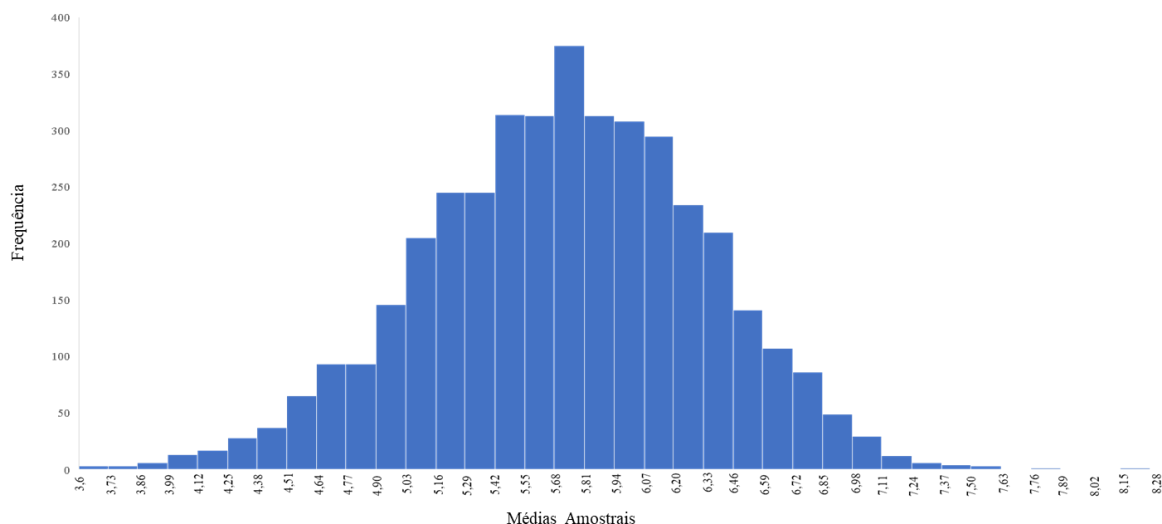


Figura 11. Distribuição amostral das médias com tamanho amostral $n = 30$

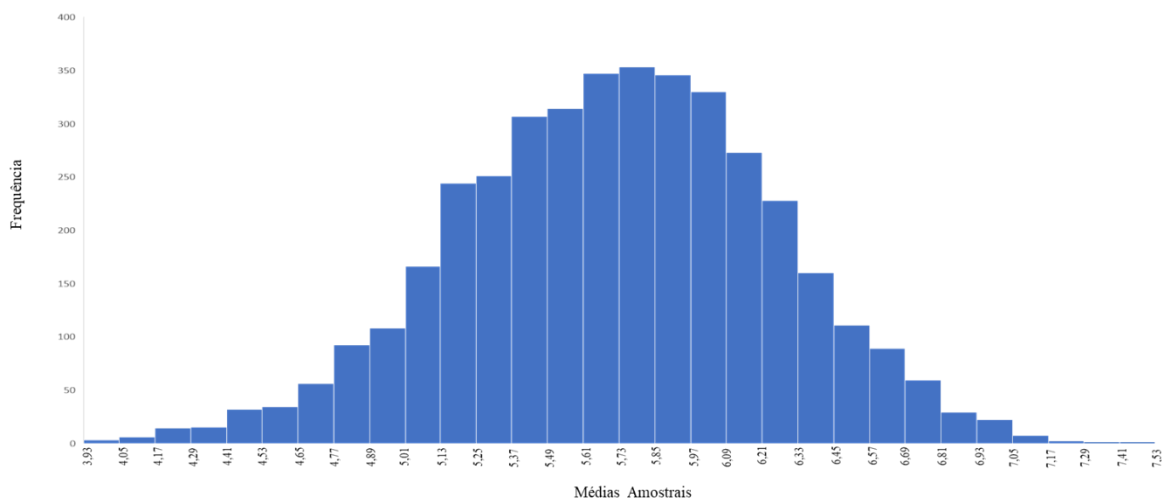


Figura 12. Distribuição amostral das médias com tamanho amostral $n = 40$

A Tabela 1 mostra os valores da média e do desvio padrão da população, o valor $\sigma/\sqrt{n}$, médias amostras e seu erro padrão.

Tabela 1. Média (μ) e desvio padrão (σ) populacional, médias amostras ($\mu_{\bar{x}}$) e seu erro padrão ($\sigma_{\bar{x}}$)

População	μ	Σ	
	5,724	3,324	
Amostras	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}}$	$\sigma_{\bar{x}} = \sigma/\sqrt{n}$
$n = 5$	5,713	1,479	1,487
$n = 10$	5,707	1,058	1,051
$n = 20$	5,719	0,748	0,743
$n = 30$	5,726	0,610	0,607
$n = 40$	5,725	0,530	0,530

Na Tabela 1, verifica-se que ao aumentar o tamanho da amostra, mais próximo o valor da média das médias amostrais se aproximava do valor da média da população. Assim como se aproximava o valor do desvio padrão da distribuição das médias amostrais do valor do desvio padrão da população dividido por raiz de 'n'.

3. CONCLUSÕES

A análise realizada com as notas dos alunos na disciplina de Estatística, através da aplicação do Teorema do Limite Central, demonstrou resultados significativos que reforçam a teoria estatística.

Os histogramas gerados revelaram que, à medida que o tamanho das amostras aumentou, as médias amostrais apresentaram uma forma mais simétrica e campanular, corroborando a previsão do Teorema do Limite Central. Foi possível observar também que conforme o tamanho da amostra aumentou a média das médias amostrais ficou muito próxima da média populacional, assim como o desvio padrão das médias amostrais ficou muito próximo ao desvio padrão populacional dividido por raiz de 'n'. Em suma, os resultados reforçam a robustez do Teorema do Limite Central e destacam a relevância da amostragem e análise estatística.

3. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] DATA SCIENCE. Escola Dnc. Teorema do Limite Central: Fundamentos e Aplicações na Inferência Estatística. 2024. Disponível em: <https://www.escoladnc.com.br/blog/teorema-do-limite-central-fundamentos-e-aplicacoes-na-inferencia-estatistica/>. Acesso em: 24 ago. 2024.
- [2] DE ALMEIDA, C. O. **Ensaio**: da amostra ao teorema do limite central um pouco dos fundamentos e uma aplicação prática. 2019.
- [3] BUSSAB, Wilton de O.; MORETTIN, Pedro A. **ESTATÍSTICA BÁSICA**. 5. ed. São Paulo: Saraiva, 2003. 526 p.
- [4] MEYER, Paul L. **Probabilidade: aplicações à estatística**. Livros Técnicos e Científicos, 1982.