
Universidade Federal de São Paulo

Instituto de Ciência e Tecnologia



**Mestrado Profissional em Matemática
em Rede Nacional - PROFMAT**

Distribuição Amostral no Ensino Médio

Maurício Brito Pereira

Orientadora: Prof^{fa} Dr^a. Juliana Garcia Cespedes

São José dos Campos
Setembro, 2018



PROFMAT

Título: *Distribuição Amostral no Ensino Médio*

Dissertação apresentada ao Instituto de Ciência e Tecnologia da UNIFESP, campus São José dos Campos/SP, como parte dos requisitos exigidos para a obtenção do título de Mestre pelo Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT.

São José dos Campos
Setembro, 2018

Brito Pereira, Maurício

Distribuição Amostral no Ensino Médio, Maurício Brito Pereira – São José dos Campos, 2018.

viii, 40f.

Dissertação (Mestrado) – Universidade Federal de São Paulo. Instituto de Ciência e Tecnologia. Programa de Pós-Graduação em Matemática em Rede Nacional (PROFMAT).

Sampling Distribution in High School

1. Estatística. 2. Distribuição Amostral. 3. Educação. 4. Planilhas Eletrônicas. 5. Excel.

UNIVERSIDADE FEDERAL DE SÃO PAULO
INSTITUTO DE CIÊNCIA E TECNOLOGIA

Mestrado Profissional em Matemática em Rede Nacional
PROFMAT

Chefe de departamento:

Prof. Dr. Eduardo Antonelli

Coordenador do Programa de Pós-Graduação:

Prof. Dr. Angelo Calil Bianchi

MAURÍCIO BRITO PEREIRA
DISTRIBUIÇÃO AMOSTRAL NO ENSINO MÉDIO

A handwritten signature in black ink, appearing to read 'Jespedes', with a large, stylized flourish above the name.

Presidente da banca: Prof^a Dr^a. Juliana Garcia Cespedes

Banca examinadora:

Profa. Dra. Flávia Cristina Martins Queiroz Mariano

Profa. Dra. Denise Beatriz Teixeira Pinto do Areal Ferrari

Prof. Dr. Luiz Leduino de Salles Neto

Data da Defesa: 17 de setembro de 2018

"O que sabemos é uma gota; o que ignoramos é um oceano". Isaac Newton.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus por me dar forças e inspiração para poder superar obstáculos que apareceram durante toda a jornada.

Agradeço a toda minha família, em especial a meus pais, que sempre foram meu porto seguro e pelo apoio incondicional, além do exemplo de luta.

Agradeço a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) por idealizar e implementar o programa PROFMAT (Mestrado Profissional em Matemática em Rede Nacional) em parceria junto a SBM (Sociedade Brasileira de Matemática) com apoio do IMPA (Instituto Nacional de Matemática Pura e Aplicada), e o Instituto de Ciência e Tecnologia da UNIFESP (Universidade Federal de São Paulo) campus São José dos Campos/SP que contribuíram para a realização deste curso.

Agradeço de forma especial, a minha orientadora Prof^a Dr^a. Juliana Garcia Cespedes, pelas suas orientações e paciência durante os meses de trabalho; e a cada um dos professores e tutores que contribuíram para a formação do meu aprendizado.

Agradeço, ainda, a meus colegas de curso por todos os momentos, bons e difíceis, que enfrentamos e que, com a ajuda mútua, conseguimos superar; a amizade e o aprendizado vão além da sala de aula.

Por fim, a todos que, de alguma forma, direta ou indiretamente, colaboraram para que esse curso pudesse ser concluído.

RESUMO

Este trabalho apresenta um breve estudo dos desafios para o ensino da Matemática no Ensino Médio, em especial os tópicos referentes à Estatística, enveredando pela Distribuição Amostral, que tanto podem contribuir para o entendimento do mundo que nos cerca e que por diversos motivos muitas vezes não é capaz de atingir o aluno como deveria. Sugere-se a necessidade de um novo modelo de abordagem, adequando-se às novas tecnologias disponíveis e que se fazem tão presentes e convidativas aos jovens nos tempos atuais. Enfatiza-se a utilização de planilhas eletrônicas, em especial o Excel, como ferramenta para tornar mais atraente ao estudante o aprendizado dos conceitos estatísticos e propõem-se atividades buscando adequar o trabalho docente a esses objetivos que convergem com as ideias que permeiam as sugestões presentes nos Parâmetros Curriculares Nacionais (PCN's) e na nova Base Nacional Comum Curricular (BNCC), que se encontra em etapa final de discussão, mas dá sinais dessas novas necessidades. Pretende-se, com o trabalho, dar uma motivação para aprender, através de atividades simples, elaboradas com o objetivo de aliar a teoria estudada nos livros didáticos à utilização das planilhas eletrônicas, procurando contribuir para uma melhor formação dos estudantes.

Palavras-chave: 1. Estatística. 2. Distribuição Amostral. 3. Educação. 4. Planilhas Eletrônicas. 5. Excel.

ABSTRACT

In that work was realized a short study of challenges for the teaching of Mathematics in High School, in special the topics related to Statistics, going through the Sampling Distribution, that can contribute so much to the understanding of the world around us and that for many reasons many times is not able to reach the student as it should. It is suggested the need for a new model of approach, adapting to the new available technologies that are so present and inviting to the teenagers in the present times. Emphasis is given to the use of electronic spreadsheets, especially the Excel, as a beneficial tool to make the student more attractive to learning of statistical concepts and proposes activities seeking to adapt the teaching work to those objectives that converge with the ideas that permeate the suggestions present in the National Curricular Parameters (NCPs) and in the new National Curricular Commom Base (BNCC), which is in the final stage of discussion but gives signs of these new needs. The purpose of this work is to provide motivation through simple activities designed to combine the theory studied in textbooks and the use of electronic spreadsheets, seeking to contribute to a better education of students.

Keywords: 1. Statistic. 2. Sampling Distribution. 3. Education. 4. Electronics Spreadsheets. 5. Excel.

SUMÁRIO

INTRODUÇÃO	3
1 REVISÃO BIBLIOGRÁFICA	5
1.1 Histórico	5
1.2 Conceito de População e Amostra	6
1.3 Conceito de Amostragem	6
1.3.1 Amostragem Aleatória Simples - AAS	7
1.3.2 Estatísticas e Parâmetros	7
1.4 Medidas de Posição	9
1.5 Medidas de Dispersão ou de Variabilidade	10
1.6 Distribuição de Frequência	11
1.6.1 Tipos de Frequências	11
1.6.2 Sugestão para a determinação do números de classes e cálculo da amplitude de intervalo de classe	12
1.6.3 Histograma	13
1.7 Distribuição Amostral	14
1.7.1 Conceito de Distribuição Amostral	14
1.7.2 Distribuição Amostral da Média	16
1.8 Estimador Não-Viesado	17
1.9 Planilha eletrônica	18
1.10 Sobre o Microsoft Office e o Microsoft Excel	19
2 PROPOSTA DIDÁTICA	21
2.1 Uso da planilha eletrônica	21
2.2 Escolha da variável	21
2.3 Obtenção dos dados	22
2.4 Amostragem	22
2.5 Comparando os resultados.	22
2.6 Construindo o conceito da distribuição amostral	23
2.7 Estimadores viciados	23
3 RESULTADOS E DISCUSSÃO	24
3.1 Exemplo da Construção da TDF no Excel	25
3.2 Cálculo da Média, Variância e Desvio Padrão no Excel	30
3.3 Distribuição Amostral das Médias no Excel	34
3.4 Atividade Investigativa em Estatística	35
3.4.1 Estratégias de Execução da Atividade	35

3.4.2	Simulação da Atividade e Proposta da Discussão dos Resultados Obtidos	36
4	CONCLUSÕES	39
A	ANEXO	43

INTRODUÇÃO

A Estatística tem papel fundamental na formação do cidadão na sua vida cotidiana. Gráficos, tabelas, porcentagens e resumos descritivos são constantemente apresentados à população pela mídia (jornais, revistas, televisão, redes sociais, etc) e podem facilmente induzir leitores a erros de entendimento e interpretação por falta de conhecimento de conceitos básicos de estatística.

Desde 1997 a estatística foi inserida nos Parâmetros Curriculares Nacionais e ainda hoje a formação dos professores de Matemática é deficiente em relação ao conteúdo de Probabilidade e Estatística. Segundo [6] é muito importante que o professor de Matemática esteja preparado para desenvolver em suas aulas os principais conceitos de estatística, e ainda, considera o fato de que existe carência de recursos pedagógicos que auxiliem os professores de matemática em suas aulas, observando que muitos possuem limitadas experiências em Estatística.

Um professor, para ter condições de coordenar um processo eficiente de ensino e aprendizagem de Estatística, deve desenvolver sua competência em três dimensões de forma integrada: a dos conceitos relacionados à Estatística; a da organização de situações didáticas; e a da compreensão do desenvolvimento da aprendizagem dos conceitos estatísticos pelos alunos. [3]

Por isso, deve-se enfatizar o ensino de Estatística desde a Educação Básica e consolidar o seu espaço na sociedade com vistas a uma formação escolar que enriqueça a cultura geral dos cidadãos. [2]

Com a proposta de criação da Base Nacional Comum Curricular [11], a Probabilidade e Estatística é considerada como uma das cinco unidades de conhecimento em que a Matemática está organizada no Ensino Médio, sendo as outras: Geometria, Grandezas e Medidas, Números e Operações, Álgebra e Funções.

Essas unidades de conhecimento, no Ensino Médio, são contempladas em cada uma das cinco Unidades Curriculares em que se organizam os objetivos de aprendizagem e desenvolvimento do componente Matemática e, em cada uma delas, recebe uma ênfase diferente, dependendo do avanço na etapa, buscando garantir que os/as estudante desenvolvam raciocínios cada vez mais sofisticados, em Matemática, ao longo dos anos de escolarização. [11, p. 561]

Considerando as normativas em construção da Base Nacional Comum Curricular, no 9º ano do Ensino Fundamental, o estudante já deve ser capaz de planejar uma pesquisa amostral, definir a técnica de amostragem e a amostra, coletar, organizar e interpretar os dados usando medidas de tendência central e amplitude. O uso intuitivo de histogramas também deve acontecer nesse ano.

No Ensino Médio, outras medidas de dispersão são introduzidas, tais como: desvio padrão e variância e "as medidas de tendência centrais, já conhecidas, devem ser analisadas considerando-se as diferentes medidas de dispersão dos dados". É também no Ensino Médio que se recomenda o uso de tecnologias (planilhas eletrônicas) para que o trabalho não tenha foco em fórmulas e cálculos.[11]

Observando o conteúdo recomendado na Base Nacional Comum Curricular, este trabalho tem por objetivo propor uma prática de aula envolvendo o tema: Distribuição Amostral da Média. Com o uso de histogramas, cálculo de média, desvio-padrão e fazendo o uso de planilhas eletrônicas, o conceito de distribuição amostral da média, bem como o conceito de estimador não viesado e de variância mínima podem ser introduzidos no Ensino Médio através de uma atividade prática.

Como objetivos específicos destacam-se:

- i) Estudar os conceitos estatísticos necessários para o conhecimento das distribuições amostrais;
- ii) Propor uma atividade de aula para o Ensino Médio;
- iii) Fazer uso de planilhas eletrônicas, como apoio pedagógico, para a prática em sala de aula.

O Capítulo 1 é destinado a uma revisão de literatura sobre os conceitos estatísticos necessários para a compreensão do tema: Distribuições Amostrais.

No Capítulo 2, apresenta-se uma proposta de atividade para o Ensino Médio usando planilhas eletrônicas. Resultados e discussões são apresentados no Capítulo 3 e, por fim, as considerações finais são apresentadas no Capítulo 4 abordando sugestões para futuros trabalhos em sala de aula.

REVISÃO BIBLIOGRÁFICA

Apresentam-se nesta seção alguns conceitos estatísticos necessários para o entendimento do tema deste trabalho.

Ao coletar informações referentes às características de um grupo de objetos ou indivíduos, é muitas vezes impossível ou impraticável observar todo o grupo, especialmente se for muito grande. Alguns exemplos são fáceis de apresentar, por exemplo: imagine uma fábrica de palito de fósforos, onde se deseja saber como está a confiabilidade destes, não seria nada útil testar todos os palitos antes que saíssem da fábrica, pois assim todos os palitos seriam inutilizados.

Conforme [13] a utilização de informações de uma amostra com o intuito de utilizar tais informações para tirar conclusões sobre o todo faz parte da atividade diária da maioria das pessoas, sejam estas cientistas ou simples dona de casa. Nos próximos tópicos, espera-se dar uma conceituação mais formal a princípios intuitivos que são utilizados com tanta frequência no cotidiano de várias pessoas para que possam então ser utilizados de uma maneira mais formal e científica. Ao final, o conceito de distribuição amostral da média e seus teoremas serão descritos.

As definições, conceitos e teoremas desta seção foram pesquisados nas referências: [12], [13], [17].

1.1 HISTÓRICO

A Estatística, como toda outra ciência, deve seu surgimento e motivação a alguma necessidade do ser humano ao longo da sua história. Há indícios de que mesmo no período antes de Cristo (a.C.) já se praticava algum tipo de estatística mesmo na sua maneira mais simples e elementar, como, por exemplo, levantamento e controle de população para fins de cobrança de impostos ou fins bélicos. A própria etimologia da palavra estatística está associada a estado, pois era tida anteriormente para lidar em prol de seus negócios.

No entanto, de acordo com [4], a Estatística, tal como a conhecemos hoje, pode ser considerada um ramo novo da Matemática, pois apenas na Idade Média ganha destaque sua importância, quando verificou-se sua necessidade e com isso a elaboração de conceitos e técnicas mais rigorosas para o seu trabalho. Dessa forma, as tabelas tornam-se mais completas, surgem as representações gráficas, o cálculo das probabilidades, e a Estatística deixa de ser uma simples catalogação de dados numéricos coletivos para se tornar uma ciência muito bem estabelecida que tem como objetivo o estudo de como chegar a

conclusões sobre o todo (população), partindo da observação e análise de partes do todo (amostra).

Com esse amadurecimento e desenvolvimento, a Estatística fica muito bem definida como sendo a disciplina que trata do estudo dos métodos científicos de coleta, organização, descrição, resumo, análise, interpretação e apresentação relevante dos dados estatísticos, bem como técnicas que dão sugestões de como podem ser obtidas conclusões válidas e a utilização dessas para a tomada de decisões razoáveis em tais análises.

Por outro lado, a Estatística se divide em duas partes mais ou menos distintas: a *Estatística Descritiva* e a *Estatística Indutiva*.

Definição 1.1. *Estatística Descritiva é a parte da Estatística que procura descrever o comportamento de uma variável em estudo, resumindo dados observados e apresentando-os através de tabelas, gráficos ou medidas que permitam interpretá-los rapidamente.*

Definição 1.2. *Estatística Indutiva é a parte da Estatística que trata das condições sob as quais inferências (deduções lógicas) sobre uma população, a partir de amostras, são válidas.*

Portanto, será discutida com mais atenção, justamente esta última, pois é a *Estatística Indutiva* (ou *Inferência Estatística*) o nome dado ao conjunto de técnicas e procedimentos que permitem dar ao pesquisador um grau de confiabilidade, ou seja, de confiança nas afirmações que faz para a população, baseadas nos resultados das amostras.

1.2 CONCEITO DE POPULAÇÃO E AMOSTRA

Definição 1.3. *População é o conjunto de todos os elementos ou resultados sob investigação em um estudo estatístico. Amostra é qualquer subconjunto finito dessa população.*

Em um estudo estatístico, se faz a análise de uma determinada característica que se tem interesse de uma população, como por exemplo, a altura ou peso de alunos das escolas de uma cidade, ou as peças defeituosas da linha de produção de uma fábrica.

No entanto, por motivos diversos, pode não ser possível verificar todos os elementos do estudo estatístico, seja por não ser viável economicamente ou devido ao tempo gasto para essa atividade, o que naturalmente limita as observações a apenas uma parte da população, sendo essa parte denominada amostra.

É notável a importância de se discutir e estudar Estatística, pois ela tem a ambição de apresentar técnicas de como chegar a conclusões sobre o todo (população), partindo justamente da observação de partes desse todo (amostra).

1.3 CONCEITO DE AMOSTRAGEM

Definição 1.4. *Amostragem é o processo de obtenção (ou extração) de amostras de uma população.*

De acordo com [17], a Teoria da Amostragem é um conjunto de procedimentos especiais que permite estudar as relações existentes entre uma população e as amostras dela extraídas, devendo garantir que as conclusões a serem obtidas com base nas amostras sejam tanto quanto possível as mais próximas se o objeto de estudo fosse toda a população propriamente dita. É importante destacar também que as observações contidas em uma amostra serão tanto mais confiáveis e informativas sobre a população quanto maior for o conhecimento explícito ou implícito dessa mesma população.

1.3.1 Amostragem Aleatória Simples - AAS

Conforme [13] a amostragem aleatória simples (ou simplesmente AAS) é a maneira mais simples de se selecionar uma amostra probabilística de uma população. O procedimento é utilizado quando todos os elementos da população possuem a mesma chance de serem sorteados. Além disso, o conhecimento adquirido do estudo desse procedimento serve como base para o aprendizado e desenvolvimento de outros procedimentos amostrais, planejamento de experimentos, estudos observacionais, etc.

O conceito de AAS, para população finita, consiste em numerar as N unidades elementares da população e realizar um sorteio dessa numeração até que se obtenha uma amostra de tamanho desejado.

Segundo [5], pode-se ter uma AAS *com reposição*, se for permitido que uma unidade possa ser sorteada mais de uma vez, e *sem reposição*, se a unidade sorteada for removida da população.

Conforme [13], do ponto de vista da quantidade de informação contida na amostra, amostrar sem reposição é mais adequado. Contudo, a amostragem com reposição conduz a um tratamento teórico mais simples, pois desta forma tem-se independência entre as unidades sorteadas. Essa independência facilita o desenvolvimento das propriedades dos estimadores que serão considerados.

Portanto, firma-se nesse momento, que o plano amostral a ser considerado ao longo desse trabalho será o de *amostragem aleatória simples com reposição*, que será denotado simplesmente por AAS.

Definição 1.5. *Uma Amostra Aleatória Simples de tamanho n de uma variável aleatória X , com dada distribuição, é o conjunto de n variáveis aleatórias independentes X_1, X_2, \dots, X_n , cada uma com a mesma distribuição de X . Ou seja, a amostra será a n -upla ordenada (X_1, X_2, \dots, X_n) , onde X_i indica a observação do i -ésimo elemento sorteado.*

1.3.2 Estatísticas e Parâmetros

Ao se obter uma amostra deseja-se usá-la para produzir alguma característica específica, por exemplo, calcular a média da amostra (X_1, X_2, \dots, X_n) , dada por:

$$\bar{X} = \frac{1}{n} \cdot \{X_1 + X_2 + \dots + X_n\}$$

É possível verificar que \bar{X} é também uma variável aleatória. Pode-se também estar interessado em qualquer outra característica da amostra, que será sempre uma função do vetor aleatório (X_1, \dots, X_n) .

Definição 1.6. *Uma estatística T é uma função de $\{X_1, X_2, \dots, X_n\}$.*

A média amostral (\bar{X}), a variância amostral (S^2) são exemplos de estatísticas. De acordo com [13] as estatísticas mais comuns são:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} : \text{média da amostra,}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} : \text{variância da amostra,}$$

$$X_{\text{mín}} = X_{(1)} = \min(X_1, X_2, \dots, X_n) : \text{o menor valor da amostra,}$$

$$X_{\text{máx}} = X_{(n)} = \max(X_1, X_2, \dots, X_n) : \text{o maior valor da amostra,}$$

$$W = X_{\text{máx}} - X_{\text{mín}} = X_{(n)} - X_{(1)} : \text{amplitude amostral,}$$

X_i = a i -ésima observação da amostra.

Em geral, pode-se considerar as *estatísticas de ordem*,

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

ou seja, os elementos da amostra ordenados.

Para facilitar a linguagem usada em Inferência Estatística, serão diferenciadas as características da amostra e da população.

Definição 1.7. *Um parâmetro é uma medida usada para descrever uma característica numérica da população. Genericamente representa-se o parâmetro por θ .*

A média (μ), a variância (σ^2) e o coeficiente de correlação (ρ) são alguns exemplos de parâmetros populacionais.

Os símbolos mais comuns, de acordo com [13], são dados na tabela a seguir:

Denominação	População	Amostra
Média	$\mu = E(X)$	$\bar{X} = \sum X_i/n$
Mediana	$Md = Q_2$	$md = q_2$
Variância	$\sigma^2 = Var(X)$	$S^2 = \sum (X_i - \bar{X})^2 / (n - 1)$
Nº de elementos	N	n
Proporção	p	\hat{p}
Quantil	Q(p)	q(p)
Quartis	Q_1, Q_2, Q_3	q_1, q_2, q_3
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função densidade	f(x)	histograma
Função de distribuição	F(x)	$F_e(x)$

1.4 MEDIDAS DE POSIÇÃO

As medidas de posição possibilitam representar um conjunto de dados relativos à observação de determinado fenômeno de forma resumida, traduzindo-os em números. Tais medidas permitem orientar quanto à posição da distribuição em relação ao eixo horizontal (eixo das abscissas), possibilitando assim, comparações de séries de dados entre si pelo confronto desses números. São chamadas medidas de tendência central, pois representam os fenômenos pelos seus valores médios, em torno dos quais tendem a concentrarem-se os dados.

Definição 1.8. *Média Aritmética (\bar{X}) é o quociente da divisão da soma dos valores da variável observados na amostra pelo número deles (tamanho da amostra):*

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

sendo: \bar{X} a média aritmética; X_i os valores da variável e n o número de valores.

Definição 1.9. *Moda (mo) é o valor que ocorre com maior frequência em uma série de valores observados em uma amostra. Alguns conjuntos de dados são multimodais, isto é, possuem mais de uma moda; outros, podem não possuir moda, sendo chamados amodais.*

Definição 1.10. *Mediana (md) é o número que se encontra na posição central em uma série de valores observados em uma amostra, estando estes dispostos em ordem (crescente ou decrescente); ou seja, é o valor que separa o conjunto de valores em dois subconjuntos de mesmo número de elementos. Caso a quantidade de dados seja par, a mediana será a média aritmética dos dois valores centrais.*

1.5 MEDIDAS DE DISPERSÃO OU DE VARIABILIDADE

São estatísticas utilizadas para avaliar o grau de variabilidade ou dispersão, ou seja, o grau em que os dados numéricos tendem a se dispersarem em torno de um valor médio (central). Servem para medir a representatividade da média. Existem várias dessas medidas, no entanto, neste trabalho, serão abordadas somente as mais usadas: Amplitude Total, Desvio Médio, Variância e Desvio Padrão.

Definição 1.11. *Amplitude Total (AT) é a diferença entre o maior e o menor valor observados numa série:*

$$AT = X_{\max} - X_{\min}$$

A utilização da Amplitude Total como medida de dispersão é limitada, pois pode ser afetada por valores discrepantes, considerando apenas os valores extremos.

Definição 1.12. *O Desvio Médio (DM) de um conjunto de N valores observados X_1, X_2, \dots, X_n é dado por:*

$$DM = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{N}$$

As distâncias entre as observações e a média foram consideradas em módulo, evitando-se que a soma de todas as distâncias seja nula.

Definição 1.13. *A Variância (σ^2) de um conjunto de N valores observados X_1, X_2, \dots, X_n é definida por:*

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - E(X))^2}{N}$$

Para o caso do cálculo da Variância Amostral (S^2), utiliza-se o estimador não viesado:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Sendo a variância calculada a partir dos quadrados dos desvios, ela é um número em unidade quadrada em relação à variável em questão, o que, sob o ponto de vista prático, é um inconveniente. Por isso, determinou-se uma nova medida que tem utilidade e interpretação práticas, sendo que essa medida é o Desvio Padrão

Definição 1.14. *O Desvio Padrão (σ) é definido como a raiz quadrada da Variância:*

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - E(X))^2}{N}}$$

Para o caso do cálculo do Desvio Padrão Amostral (S), utiliza-se o estimador não viesado:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

As definições, conceitos e teoremas desta seção foram pesquisados nas referências: [4], [7], [12], [13], [17].

1.6 DISTRIBUIÇÃO DE FREQUÊNCIA

A Tabela de Distribuição de Frequências (TDF) será apresentada como tópico introdutório aos conceitos básicos de Estatística.

Basicamente, pode-se ter duas situações na construção da Tabela de Frequências: com ou sem intervalos de classes. As TDF sem intervalos de classe, são as tabelas em que cada linha corresponde a um único valor da variável que está sendo discutida, sendo esse modelo mais utilizado quando os dados a serem discutidos são poucos e possuem pouca dispersão (baixa variação). As TDF com intervalos de classe, são as tabelas em que cada linha abrange um intervalo de valores correspondente aos dados que se quer estudar, sendo este modelo mais aplicado quando se têm muitos dados ou estes são muito variados. Será dada especial atenção a esse último, pois de acordo com [4], o que se pretende com a construção dessa tabela é realçar o que há de essencial nos dados, bem como tornar possível o uso de técnicas analíticas para sua total descrição.

1.6.1 Tipos de Frequências

Frequência é a quantidade atribuída a uma das variáveis correspondentes a uma das informações coletadas como dados Estatísticos. As frequências mais utilizadas em uma tabela de frequências são: absoluta, relativa e percentual; tanto simples como acumulada.

Definição 1.15. *Frequências absolutas simples (f_i) são os valores que representam o número de vezes que um determinado elemento aparece na amostra ou o número de elementos pertencentes a uma classe. A soma das frequências simples é igual ao número*

$$\text{total dos dados: } \sum_{i=1}^k f_i = \sum f_i = n.$$

Definição 1.16. *Frequência absoluta acumulada (F_i) é o total acumulado (soma) das frequências simples de todas as classes anteriores até a classe que se deseja analisar: $F_i = f_1 + f_2 + \dots + f_i$, onde i é a classe considerada.*

Com o objetivo de facilitar a identificação e diferenciação no trabalho com as frequências, será adotado o "f" minúsculo quando se tratar de frequências simples e o "F" maiúsculo quando se tratar de frequências acumuladas.

Definição 1.17. *Frequências relativas simples (fr_i) são valores que correspondem às razões (quocientes) entre as frequências simples e a frequência total, tendo como propósito permitir análises posteriores ou facilitar comparações: $fr_i = \frac{f_i}{\sum f_i} = \frac{f_i}{n}$. Onde tem-se*

$$\text{que } \sum_{i=1}^k fr_i = 1.$$

Definição 1.18. *Frequência relativa acumulada (Fr_i) é a soma das frequências relativas simples (fr_i) das classes anteriores mais a da classe em questão. Também pode ser calculada fazendo a divisão da frequência acumulada da classe (F_i) pela frequência total da distribuição (n): $Fr_i = \sum fr_i$ ou $Fr_i = \frac{F_i}{n}$, onde i é a classe considerada.*

Definição 1.19. *Frequência percentual simples (fp_i) é obtida diretamente da frequência relativa simples (fr_i) multiplicando-a por 100: $fp_i = fr_i \times 100$ (%), onde i é a classe considerada.*

Definição 1.20. *Frequência percentual acumulada (Fp_i) pode ser obtida facilmente da frequência relativa acumulada (Fr_i) multiplicando-a por 100: $Fp_i = Fr_i \times 100$ (%), onde i é a classe considerada.*

1.6.2 Sugestão para a determinação do número de classes e cálculo da amplitude de intervalo de classe

Uma primeira preocupação que se deve ter na construção de uma Tabela de Distribuição de Frequência (TDF) é saber determinar o número de classes que essa tabela deverá ter, e com isso, poder calcular a amplitude de intervalo de classe e dessa forma escolher adequadamente os limites do intervalo de classe

Existem algumas sugestões para a determinação do número de classes. Será apresentada a Regra de Sturges, que leva em consideração o número de valores correspondente à variável aleatória que se está discutindo:

$$k \cong 1 + 3,22 \cdot \log n$$

onde:

k é a sugestão do número de classes;

n é o número total de dados.

Como informado anteriormente, existem algumas sugestões que têm por finalidade ajudar na escolha do número de classes que a distribuição deve ter, entretanto, de acordo com [4], essas fórmulas não devem levar a uma decisão final, pois esta depende de um julgamento pessoal, ligado à natureza dos dados, do tipo de unidade usada para expressá-los e, ainda do objetivo que se pretende atingir, além disso, deve-se procurar sempre que possível, evitar classes com frequência nula ou com frequência relativa muito exagerada.

Após a decisão do número de classes que a distribuição deve ter, pode-se então calcular a amplitude do intervalo de classe correspondente, que é feito dividindo-se a amplitude amostral (AA) pelo número de classes (k):

$$h \cong \frac{AA}{k}$$

Quando o resultado obtido não for exato, deve-se arredondá-lo sempre para mais, pois dessa forma impede-se que ao construir a Tabela de Distribuição de Frequência, algum valor não seja coberto.

Outra vantagem do fato de arredondar o valor calculado da amplitude de intervalo de classe para cima, é que dessa forma, obtém-se uma folga na escolha para os limites do intervalo de classe, pensando em valores apropriados, que deverão ser escolhidos, na medida do possível, de forma a obter-se pontos médios que facilitem cálculos posteriores (ou seja, números naturais).

1.6.3 *Histograma*

O Histograma é um tipo de Gráfico Estatístico que conforme [4] é uma forma de apresentação dos dados estatísticos, cujo objetivo é de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão que as séries.

De acordo com [13], usar os métodos gráficos têm encontrado um uso cada vez maior devido ao seu forte apelo visual, pois normalmente é mais fácil para qualquer pessoa entender a mensagem de um gráfico do que aquela embutida em tabelas.

É notável a importância e benefícios que os gráficos estatísticos podem adicionar ao trabalho estatístico, mas devem-se, no entanto, tomar alguns cuidados e formalizar alguns processos. Existem diversos métodos estatísticos, ou seja, formas e padrões para a construção desses gráficos, onde se destaca o Histograma, sendo esse escolhido por sua relevância para ser apresentado e utilizado no presente trabalho.

O histograma, de acordo com [13], é um gráfico de barras contíguas (lado a lado), com as bases proporcionais aos intervalos das classes e a área de cada retângulo proporcional à respectiva frequência, podendo-se usar tanto a frequência absoluta (n_i), como a relativa (f_i). Indicando a amplitude do i -ésimo intervalo por Δ_i , para que a área do respectivo retângulo seja proporcional a f_i , a sua altura deve ser proporcional a f_i/Δ_i (ou a n_i/Δ_i), que é chamada densidade de frequência da i -ésima classe. Quanto mais dados se tem em uma classe, mais alto deve ser o retângulo. Com essa convenção, tem-se que a área total do histograma será igual a um (deve-se lembrar que a somatória das frequências relativas é igual a um).

Ainda conforme [13], quando os intervalos das classes forem todos iguais a Δ , a densidade de frequência da i -ésima classe passa a ser f_i/Δ (ou n_i/Δ). É claro que marcar

no eixo das ordenadas os valores n_i , f_i , n_i/Δ ou f_i/Δ leva a obter histogramas com a mesma forma; em que somente as áreas é que serão diferentes.

Na figura 1 pode-se ver um exemplo de um histograma construído com o auxílio do software Excel. Por exemplo, a representação do número de filhos dos funcionários com dependentes de uma grande empresa, em que o eixo das abscissas representa o número de filhos de um a sete, e o eixo das ordenadas a quantidade de funcionários correspondente.

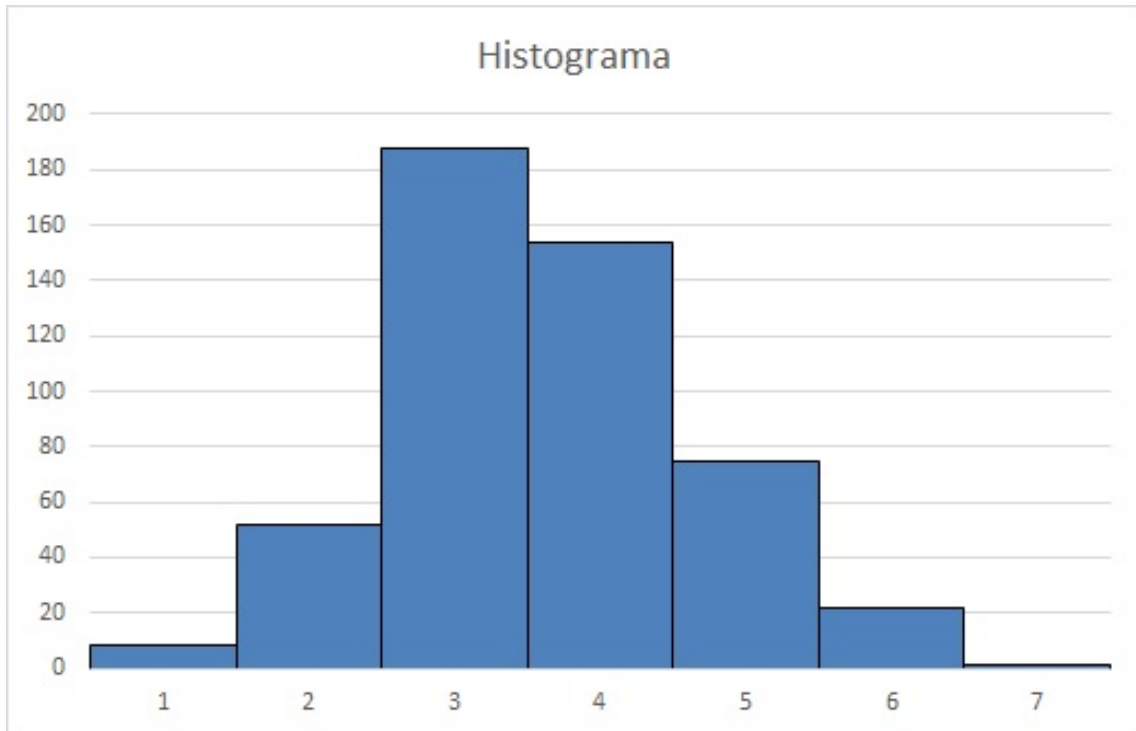


Figura 1: Exemplo de Histograma construído no Excel

1.7 DISTRIBUIÇÃO AMOSTRAL

1.7.1 *Conceito de Distribuição Amostral*

Sabe-se que o problema da Inferência Estatística é fazer uma afirmação sobre os parâmetros da população a partir do que se observa na amostra. É natural que toda conclusão tirada por uma amostragem, quando generalizada para a população, virá acompanhada por um grau de incerteza ou risco que pode ser estimado. É papel fundamental da Inferência Estatística, portanto, medir este grau de incerteza ou risco dessas generalizações; e seus instrumentos têm como objetivo permitir viabilizar as conclusões através de afirmações estatísticas.

Conforme [17], considerando todas as amostras possíveis de tamanho n que podem ser retiradas de uma população de N elementos dada (com ou sem reposição), para cada amostra, pode-se calcular uma estatística, tais como, a média, o desvio padrão, etc., que varia de amostra para amostra.

Com os valores das estatísticas obtidos nas amostras retiradas da população, obtém-se uma distribuição da estatística que é denominada *Distribuição Amostral*.

Se, por exemplo, a estatística particular adotada for a média da amostra, obtém-se a *Distribuição Amostral da Média*. Semelhantemente pode-se ter distribuições amostrais do desvio padrão, da variância, das proporções, etc.

Ainda de acordo com [17], para cada uma das distribuições amostrais obtidas, pode-se calcular a esperança (média), o desvio padrão, etc. Em consequência, pode-se falar da média e do desvio padrão da distribuição amostral das médias e assim sucessivamente.

Também conforme [13], supondo que uma afirmação deva ser feita sobre um parâmetro θ da população (por exemplo, como citado anteriormente, a média, a variância, ou qualquer outra medida). Fica decidido que será utilizada uma AAS de n elementos sorteados dessa população. A decisão será baseada na estatística T , que será uma função da amostra (X_1, X_2, \dots, X_n) , ou seja, $T = f(X_1, X_2, \dots, X_n)$. Colhida essa amostra, teremos observado um particular valor de T , digamos t_0 , e baseados nesse valor é que faremos a afirmação sobre θ , o parâmetro populacional.

A validade da resposta obtida seria melhor compreendida se fosse possível saber o que acontece com a estatística T , quando se retiram todas as amostras da população conhecida segundo o plano amostral adotado. Isto é, qual a distribuição de T quando (X_1, X_2, \dots, X_n) assume todos os valores possíveis. Essa distribuição é chamada *distribuição amostral da estatística T* e desempenha papel fundamental na teoria da inferência estatística, também de acordo com [13].

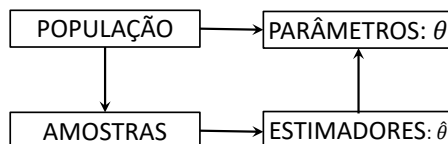
Esquemáticamente, para [13], tem-se o seguinte procedimento:

- (a) uma população X , com determinado parâmetro de interesse θ ;
- (b) todas as amostras retiradas da população, de acordo com certo procedimento;
- (c) para cada amostra, o valor t da estatística T é calculado;
- (d) os valores t formam uma nova população, cuja distribuição recebe o nome de Distribuição Amostral de T .

O objetivo principal, então, é poder identificar um modelo que explique bem a distribuição amostral de T . É evidente que a distribuição de T irá depender da distribuição de X e do plano amostral, no presente caso, reduzido a AAS.

É importante perceber que há a necessidade de determinar propriedades das distribuições amostrais que possam ser aplicadas em situações mais gerais e não em situações muito particulares. Será a seguir, apresentado o estudo de distribuições amostrais de algumas estatísticas importantes que serão utilizadas para fazer inferências sobre populações.

Sendo utilizado nesses casos, conforme [12], o esquema geral a seguir:



Alguns importantes conceitos, que têm papel fundamental nas discussões sobre distribuições, mas que no entanto, são de maior grau de complexidade, serão apresentados na parte anexa como material complementar, sendo estes conceitos introdutórios sobre *Distribuição Normal* e o *Teorema do Limite Central*.

1.7.2 Distribuição Amostral da Média

Será estudada agora a distribuição amostral da estatística \bar{X} , a média da amostra. Considere-se uma população identificada pela variável X , cujos parâmetros média populacional $\mu = E(X)$ e variância populacional $\sigma^2 = Var(X)$ são supostos conhecidos. Admita-se que sejam retiradas todas as possíveis AAS de tamanho n dessa população, e para cada uma sejam calculadas as respectivas médias \bar{X} .

Verificam-se dois fatos: primeiro, a média das médias amostrais coincide com a média populacional; segundo, a variância de \bar{X} é igual à variância de X , dividida por n . Estes dois fatos não são casos isolados. Na realidade, tem-se o seguinte resultado e propriedades.

Definição 1.21. *Distribuição Amostral das Médias é a distribuição de probabilidade da estatística \bar{X} , de todas as amostras de n elementos de uma população.*

Teorema 1.22. *Seja X uma variável aleatória com média μ e variância σ^2 , e seja (X_1, X_2, \dots, X_n) uma AAS de X . Então:*

$$E(\bar{X}) = \mu \text{ e } Var(\bar{X}) = \frac{\sigma^2}{n}.$$

Prova: Pelas propriedades de variáveis aleatórias, tem-se:

$$\begin{aligned} E(\bar{X}) &= \left(\frac{1}{n}\right) \{E(X_1) + \dots + E(X_n)\} \\ &= \left(\frac{1}{n}\right) \{\mu + \mu + \dots + \mu\} = \frac{n\mu}{n} = \mu. \end{aligned}$$

De modo análogo, e pelo fato de X_1, \dots, X_n serem independentes, tem-se:

$$\begin{aligned} \text{Var}(\bar{X}) &= \left(\frac{1}{n^2}\right) \{\text{Var}(X_1) + \dots + \text{Var}(X_n)\} \\ &= \left(\frac{1}{n^2}\right) \{\sigma^2 + \dots + \sigma^2\} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Ficam, então, determinados, a média e a variância da distribuição amostral de \bar{X} .

Teorema 1.23. *Se a população da qual se extraem as amostras têm distribuição normal com média μ e variância σ^2 , então a média amostral é normalmente distribuída com média μ e variância $\frac{\sigma^2}{n}$.*

Teorema 1.24. *Supondo-se que a população da qual se extraem amostras tenha uma distribuição de probabilidade não necessariamente normal, com média μ e variância σ^2 . Então, a variável padronizada associada a \bar{X} , dada por:*

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

é assintoticamente normal, isto é,

$$\lim_{n \rightarrow \infty} P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^z e^{-u^2/2} du$$

É importante conseguir observar, que, conforme n aumenta (ou seja, são tomados cada vez valores maiores da amostra), independente da distribuição da população, a distribuição amostral aproxima-se cada vez mais de uma distribuição normal. Esse resultado, fundamental na teoria da Inferência Estatística, é devido ao *Teorema do Limite Central* (Teorema A.1 - Anexo). A rapidez da convergência (o fato de que a distribuição aproxima-se da normal quando n tende para o infinito) depende da distribuição da população da qual a amostra é retirada. Se a população original tem uma distribuição próxima da normal, a convergência é rápida; se a população original se afasta muito de uma normal, a convergência é mais lenta, ou seja, será necessária uma amostra maior para que a distribuição tenha uma distribuição aproximadamente normal. Para amostras da ordem de 30 ou 50 elementos, a aproximação pode ser considerada boa.

As definições, conceitos e teoremas desta seção foram pesquisados nas referências: [7], [12], [13], [14], [17].

1.8 ESTIMADOR NÃO-VIESADO

Conforme [12], um dos objetivos básicos da experimentação é a estimação de parâmetros. Estuda-se uma população cuja distribuição é considerada conhecida por meio de sua

função de densidade de probabilidade, $f(X, \theta_1, \theta_2, \dots, \theta_p)$, onde X é uma variável aleatória e θ_i , $i = 1, 2, \dots, p$ são os parâmetros da distribuição que se deseja estimar, como por exemplo a média $\mu = E(X)$ ou a variância $\sigma^2 = Var(X)$.

Sabe-se que um dos objetivos da Inferência Estatística é permitir avaliar um ou mais parâmetros de uma distribuição populacional, tomando-se por base uma amostra casual simples x_1, x_2, \dots, x_n ; ao que tem-se que o principal problema é procurar funções de observações que permitam fornecer estimativas verossímeis dos referidos parâmetros. Em que, a distribuição dessas funções devem estar o mais concentradas possível em torno dos verdadeiros valores dos parâmetros de θ ; portanto, de acordo com [13], pode-se então definir o que é um estimador.

Definição 1.25. *Um estimador T do parâmetro θ é qualquer função das observações da amostra, ou seja, $T = g(X_1, \dots, X_n)$.*

Ainda conforme [13], deve-se notar que, segundo essa definição, um estimador é o que definia-se antes por estatística, porém associando-o a um parâmetro populacional. Ao passo que o problema da estimação, conforme posto anteriormente, é poder então determinar uma função $T = g(X_1, \dots, X_n)$ que seja a mais próxima possível de θ , segundo algum critério, o qual nesse momento é de interesse e portanto apenas o qual será abordado, o caso em que esse estimador é dito não-viesado.

Definição 1.26. *O estimador T é não-viesado (ou não viciado ou não tendencioso ou justo) para θ se $E(T) = \theta$, para todo θ .*

Definição 1.27. *Estimativa é o valor assumido pelo estimador em uma particular amostra.*

Ou seja, de acordo com [17], uma estatística é chamada estimador não-viesado de um parâmetro populacional se sua média, ou esperança, for igual ao parâmetro, sendo que o valor correspondente da estatística chama-se estimativa não-viesada do parâmetro.

De maneira análoga, tem-se que, se $E(T) \neq \theta$, T é chamado viesado, em que $V(T) = E(T) - \theta$ é chamado o viés de T . Ainda conforme [13], deve-se notar que a esperança de T em $E(T) = \theta$ é calculada sobre a distribuição amostral de T .

1.9 PLANILHA ELETRÔNICA

A planilha eletrônica utilizada neste trabalho será o Microsoft Excel, um dos programas do Microsoft Office. A escolha deve-se ao fato deste programa ser comum, e de fácil acesso para vários estudantes por estar presente em instituições de ensino, bem como em vários computadores particulares domésticos.

Outras opções de planilhas eletrônicas que podem ser facilmente adaptadas à proposta deste trabalho são as versões gratuitas LibreOffice ou BrOffice. As funcionalidades destas planilhas são semelhantes às do Excel.

Vale ressaltar que a escolha do Excel é interessante, pois além de ser um pacote presente em vários computadores, quando estes são comprados das lojas (a maioria dos computadores vendidos no Brasil, acompanham o sistema operacional Windows, da Microsoft, já com sua devida licença, e normalmente acompanha o pacote Microsoft Office, o qual contém o Excel), a Microsoft, tem parcerias com várias instituições de Ensino Superior, disponibilizando para estas e para seus alunos, o pacote Microsoft Office de maneira gratuita. Mesmo que uma determinada universidade não tenha parceria, um estudante universitário terá direito ao uso gratuito bem como qualquer professor universitário, sendo necessário para isso apenas que possua uma conta de e-mail escolar válido. No próprio site (<https://products.office.com/pt-br/student/office-in-education>) é destacado que *"Não é uma versão de avaliação! Alunos e professores estão qualificados para o Office 365 Education que inclui os aplicativos Word, Excel, PowerPoint e OneNote e o Microsoft Teams, além de ferramentas adicionais para sala de aula. Para começar a usar, você só precisa inserir um endereço de email de estudante válido"*.

Também influenciou a escolha, o fato do Excel ser uma ferramenta poderosa com vários recursos e também ser de fácil utilização, com aplicação de fórmulas e funções de maneira muito intuitiva.

1.10 SOBRE O MICROSOFT OFFICE E O MICROSOFT EXCEL

O Microsoft Excel (ou simplesmente Excel), conforme [15] e [16], é uma planilha de cálculo, do pacote de programas do Microsoft Office para computadores que usam o Microsoft Windows principalmente, e são aplicativos inicialmente concebidos para uso profissional em escritórios, mas que ao longo dos anos se mostraram úteis e com utilização nas mais diversas aplicações, como domésticas e educacionais. O Microsoft Office foi anunciado pela primeira vez, por Bill Gates (presidente do grupo Microsoft) em 01 de agosto de 1988, sendo o Excel tendo sido integrado de maneira implícita ao pacote em 1993. Desde então, foram sendo feitas atualizações e melhorias em todos seus programas, a cada nova versão.

O Excel, como editor de planilhas eletrônicas, possui poderosos recursos que incluem uma interface intuitiva e ferramentas de cálculos e construção de gráficos que o ajudaram a se tornar um dos mais populares aplicativos em todo mundo.

O Excel oferece diversos ajustes na interface ao usuário, mas essencialmente, ele é formado por células que são organizados em linhas e colunas, e cada célula pode conter dados ou fórmulas com referências relativas ou absolutas a outras células, tendo também capacidade avançada de construções de gráficos.

Para o acesso ao recursos da fórmula no Excel, deve ser iniciado com o uso do sinal de igual (=), seguido por alguma combinação entre operações matemáticas ou outras operações de processamento de dados. Para fórmulas simples, os símbolos +, -, *, / e ^

são utilizados em operações de adição, subtração, multiplicação, divisão e exponenciação (um número elevado a uma determinada potência), respectivamente.

Também pode-se utilizar *funções de planilha*, que são funções previamente carregadas no Excel, que funcionam segundo uma determinada sintaxe, algumas destas serão apresentadas no decorrer dos próximos tópicos.

Quando se está utilizando *funções de planilha*, pode-se acessar o assistente de função, para obter maiores orientações para uso destas. Este assistente pode ser acessado a partir do menu *Fórmulas* submenu *Inserir Função* e em seguida faz as entradas e as seleções em uma ou mais caixas de diálogos.

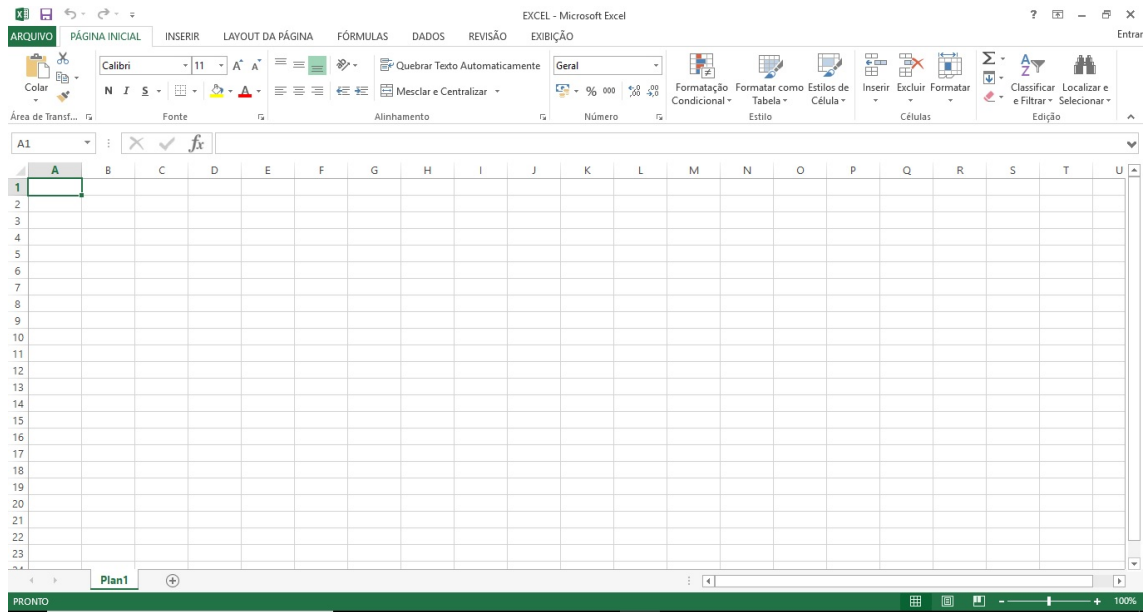


Figura 2: Tela inicial do Microsoft Excel

PROPOSTA DIDÁTICA

A proposta didática aqui apresentada divide-se em algumas etapas. Elas são importantes para o bom desenvolvimento da atividade e para a construção do conhecimento junto ao estudante.

2.1 USO DA PLANILHA ELETRÔNICA

Inicialmente propõe-se aos estudantes uma revisão de conceitos estatísticos já aprendidos com o uso de planilhas eletrônicas. Essa revisão tem por objetivo familiarizá-lo com o uso da planilha eletrônica e suas principais funções.

2.2 ESCOLHA DA VARIÁVEL

Para que o estudante do Ensino Médio seja inserido no contexto da aprendizagem, o segundo momento da prática didática consiste na escolha da variável a ser estudada.

Propõe-se aos estudantes que determinem um tema cuja resposta possa ser representada por uma variável aleatória discreta ou contínua.

Nesse momento, o professor pode introduzir o conceito de variável aleatória de maneira intuitiva, indicando ao estudante que tipo de resposta ele espera, por exemplo, em relação ao seu peso. A resposta obtida ao pesar um estudante é um número pertencente ao conjunto dos reais e que pode ser resumido através da média.

Outros exemplos de assuntos que podem ser trabalhos é o dia do aniversário, o número de irmãos(as), o número de pessoas na família, etc. Cabe ao professor contextualizar o assunto e adaptar para a realidade da classe.

Esse momento pode também ser oportuno para o professor trabalhar outros assuntos, que podem estar envolvidos na pesquisa, e que possuem caráter educativo. Por exemplo, ao usar a variável peso é possível que estudantes sintam-se constrangidos em divulgar o seu peso, abrindo espaço para trabalhar aspectos relacionados ao bullying, auto-estima, empoderamento, etc.

Cabe ressaltar que é importante que o tema escolhido refira-se a uma população finita e com poucas unidades (uma resposta para cada estudante da turma) para que a atividade seja produtiva.

2.3 OBTENÇÃO DOS DADOS

Após a escolha do tema, os estudantes passam para a etapa de coleta das informações.

Nesse momento é possível trabalhar conceitos como erro de medição, aferição de equipamentos, influência do experimentador no resultado de uma pesquisa.

Convém, nesta etapa, explicar aos estudantes que a participação na atividade é seletiva e que aqueles que não se sentirem confortáveis em fornecer as informações sobre o tema proposto não são obrigados a fazê-lo.

Todos os estudantes que aceitarem fornecer informações serão então considerados como sendo a população em estudo.

Os estudantes podem, durante a obtenção dos dados, preencher as informações solicitadas na planilha eletrônica, que foi previamente utilizada para a revisão dos conceitos estatísticos.

O professor, caso seja necessário, deve fazer uma intervenção para que os dados sejam colocados na planilha de maneira correta.

Assim que os dados forem coletados, as descrições populacionais podem ser feitas, tais como, cálculo de média, variância e desvio padrão populacionais.

2.4 AMOSTRAGEM

O conceito de amostragem aleatória simples pode ser trabalhado nesse momento.

Primeiramente enumera-se de 1 até N as respostas dadas pelos estudantes.

Se a atividade estiver sendo feita em um laboratório de informática, cada estudante poderá gerar uma amostra aleatória simples da população.

Na planilha eletrônica, essa amostragem pode ser feita usando a função ALEATÓRIOENTRE.

Inicialmente o estudante deve gerar uma amostra de tamanho 2, depois uma amostra de tamanho 10, e assim sucessivamente até amostrar todos os elementos da população.

Para que o encaminhamento seja proveitoso, todos os estudantes devem gerar os mesmos números de elementos.

2.5 COMPARANDO OS RESULTADOS.

O primeiro passo para a comparação dos resultados é calcular o valor das estatísticas para as amostras selecionadas.

As funções MÉDIA, VAR.A e DESVPAD.A da planilha eletrônica podem ser utilizadas para o cálculo da média, variância e desvio padrão, respectivamente, das amostras de tamanhos diferentes. Com isso, num primeiro momento, os estudantes devem perceber que ao aumentar o tamanho da amostra, as estimativas obtidas nas planilhas eletrônicas tendem a se aproximar.

Num segundo momento o professor deve chamar a atenção para os diferentes resultados obtidos pelos estudantes, levantando os questionamentos: *"Os valores das médias e variâncias que vocês obtiveram são iguais? Se não o são, quais valores vocês escolheriam para representar a média dessa população? De que forma podemos resumir as informações das estimativas de todas as amostras que vocês obtiveram?"*

2.6 CONSTRUINDO O CONCEITO DA DISTRIBUIÇÃO AMOSTRAL

Com os questionamentos levantados pelo professor deve-se encaminhar a discussão para a construção de gráficos, mais especificamente, o histograma (conteúdo abordado no 9º ano do ensino fundamental, segundo [11]). O objetivo do Histograma é permitir a visualização do comportamento das estimativas obtidas nas diferentes amostras.

O histograma pode ser feito no Excel usando o suplemento "Análise de dados", que divide a variável em faixas e faz um gráfico de barras com as frequências de cada faixa. No entanto é preciso corrigir o gráfico de barras para que as barras não tenham espaço entre elas, tal procedimento será detalhado em tópico a seguir.

Com gráfico pronto é interessante perguntar aos estudantes: Qual é o valor mais provável que pode-se encontrar para a média, se fosse feito um sorteio aleatório, com reposição, de amostras de tamanho n ?

Esse mesmo gráfico pode ser construídos para os diferentes tamanhos amostrais gerados pelos estudantes e aqui surgem os questionamentos: *O que acontece com a amplitude desses valores? Diminuem ou aumentam com o tamanho da amostra? Quanto maior o tamanho da amostra, menor é a variabilidade existente na distribuição amostral da média?*

2.7 ESTIMADORES VICIADOS

Nesta proposta também é possível trabalhar o conceito de viés. Para tal, algumas perguntas podem ser feitas aos estudantes, como: *Porque a média é calculada como $\frac{\sum X_i}{n}$? Porque não podemos calcular a média como sendo o maior valor sorteado na amostra menos o menor valor, dividido por 2: $\frac{X_{máx} - X_{mín}}{2}$?*

Em seguida, o professor pode solicitar aos estudantes que calculem a "média" usando a segunda estatística. Refazendo os gráficos observa-se que a média das médias não é mais igual ao valor verdadeiro da média desta população e portanto, essa segunda estatística não representa bem o parâmetro da média.

RESULTADOS E DISCUSSÃO

Nesta seção será feito um exemplo simulado de aplicação desta vivência, com a descrição das funções que podem ser utilizadas na atividade prática.

As atividades propostas não têm a pretensão de servir como um tutorial ou receita pronta, mas como fonte de inspiração para ideias que poderão ser utilizadas e aperfeiçoadas para o ensino de Estatística em sala de aula, com o suporte de planilhas eletrônicas. É responsabilidade de cada professor buscar atividades que melhor se encaixem no seu contexto, que se ajustem aos seus objetivos e melhor sirvam aos seus alunos, visando sempre tornar o ensinamento mais atraente e dinâmico.

Seguindo a estratégia de apresentação da proposta didática, a primeira etapa consiste na revisão de conceitos de estatística para utilização da planilha eletrônica.

Em seguida, usando a função **ALEATÓRIOENTRE** disponível na planilha eletrônica é possível gerar uma amostra aleatória simples com reposição da população em estudo. Esta função gera valores discretos aleatórios entre um intervalo escolhido, com reposição, imitando um sorteio, onde todas as unidades amostrais tem a mesma probabilidade de pertencerem a amostra; para a simulação das atividades é esta função que será utilizada.

Na tela inicial do Excel, figura 3, será inserida na célula A1 a fórmula:

$$=ALEATÓRIOENTRE(X_{\min}; X_{\max}).$$

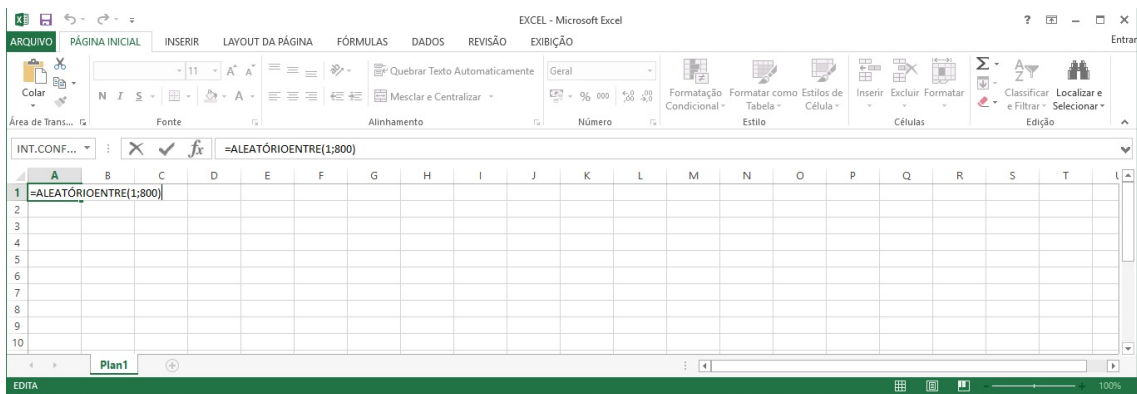


Figura 3: Função ALEATÓRIOENTRE

3.1 EXEMPLO DA CONSTRUÇÃO DA TDF NO EXCEL

Tempo estimado da atividade: acredita-se que após a definição dos conceitos, pode-se utilizar duas aulas para a execução no computador.

Exemplo 3.1. Considerando 100 valores, gerados pelo Excel, no intervalo de 151 a 189, conforme figura 4:

	A	B	C	D	E	F	G	H	I	J
1	154	188	160	152	164	164	180	180	175	170
2	160	151	180	154	177	185	178	155	168	166
3	187	176	170	176	169	163	182	187	156	156
4	182	186	177	170	158	173	155	171	151	187
5	153	164	156	151	160	173	165	177	183	179
6	153	156	159	165	186	167	155	157	159	188
7	153	178	153	174	185	183	161	180	172	158
8	170	168	152	180	175	173	183	167	176	151
9	186	189	154	177	159	171	175	176	181	174
10	168	155	169	170	170	157	184	186	164	165

Figura 4: Números gerados de maneira aleatória no Excel com o uso do comando: "`=ALEATÓRIOENTRE(151;189)`" copiado em 100 células

Para facilitar a visualização e a apresentação, para as próximas imagens, será mostrado apenas a região correspondente as células do programa do Excel, conforme figura 5:

	A	B	C	D	E	F	G	H	I	J
1	154	188	160	152	164	164	180	180	175	170
2	160	151	180	154	177	185	178	155	168	166
3	187	176	170	176	169	163	182	187	156	156
4	182	186	177	170	158	173	155	171	151	187
5	153	164	156	151	160	173	165	177	183	179
6	153	156	159	165	186	167	155	157	159	188
7	153	178	153	174	185	183	161	180	172	158
8	170	168	152	180	175	173	183	167	176	151
9	186	189	154	177	159	171	175	176	181	174
10	168	155	169	170	170	157	184	186	164	165

Figura 5: Conjunto de Dados Gerados no Excel

1. Pede-se então, determinar:

(a) A Amplitude Amostral (AA):

Para apresentar a Amplitude Amostral dos dados, será feita, num primeiro momento, uma verificação do menor e do maior valor presentes na tabela, fazendo uso das seguintes funções:

- i. Função MENOR: Retorna o menor valor k -ésimo de um conjunto de dados.
Sintaxe: = *MENOR*(*Matriz*; k)
- ii. Função MAIOR: Retorna o maior valor k -ésimo de um conjunto de dados.
Sintaxe: = *MAIOR*(*Matriz*; k)

No caso em questão, têm-se como argumento correspondente, a "Matriz": A1 : J10, pois os valores a serem analisados estão nesse intervalo de células e "k": 1 pois se quer o primeiro desses valores. A Amplitude Amostral é apresentada fazendo a diferença entre os valores encontrados, conforme pode-se verificar na figura 6.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	154	188	160	152	164	164	180	180	175	170		MENOR	151
2	160	151	180	154	177	185	178	155	168	166		MAIOR	189
3	187	176	170	176	169	163	182	187	156	156		AA	38
4	182	186	177	170	158	173	155	171	151	187			
5	153	164	156	151	160	173	165	177	183	179			
6	153	156	159	165	186	167	155	157	159	188			
7	153	178	153	174	185	183	161	180	172	158			
8	170	168	152	180	175	173	183	167	176	151			
9	186	189	154	177	159	171	175	176	181	174			
10	168	155	169	170	170	157	184	186	164	165			

Figura 6: Amplitude Amostral do Conjunto de Dados

Dessa forma, sabe-se que o menor valor encontrado é 151 e o maior é 189, portanto a Amplitude Amostral é 38.

(b) *o Rol*:

Sabendo que os valores extremos são 151 e 189, serão então determinadas duas novas colunas, em que uma será dos valores e outra da frequência desses valores sendo que para determiná-la será utilizado a seguinte função:

- i. Função CONT.SE: Calcula o número de células não vazias em um intervalo que corresponde a um determinada condição.
Sintaxe: = *CONT.SE*(*Intervalo*; *Critérios*)

No caso em questão, tem-se como argumento para o "Intervalo": A1 : J10 e como critério cada um dos valores que os dados podem assumir. Pode-se ver na figura 7 uma parte dessa tabela.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	154	188	160	152	164	164	180	180	175	170		MENOR	151		VALORES	FREQÜÊNCIA
2	160	151	180	154	177	185	178	155	168	166		MAIOR	189		151	4
3	187	176	170	176	169	163	182	187	156	156		AA	38		152	2
4	182	186	177	170	158	173	155	171	151	187					153	4
5	153	164	156	151	160	173	165	177	183	179					154	3
6	153	156	159	165	186	167	155	157	159	188					155	4
7	153	178	153	174	185	183	161	180	172	158					156	4
8	170	168	152	180	175	173	183	167	176	151					157	2
9	186	189	154	177	159	171	175	176	181	174					158	2
10	168	155	169	170	170	157	184	186	164	165					159	3
11															160	3
12															161	1
13															162	0
14															163	1
15															164	4
16															165	3
17															166	1
18															167	2
19															168	3
20															169	2

Figura 7: Conjunto de Dados em Rol

(c) *O Número de Classes (k):*

Será feito uso da Regra de Sturges, em que será utilizado a própria tabela do Excel, digitando em uma das células o seguinte comando: $= 1 + 3,22 * \log(100)$; pois o número total de dados (n) é igual a 100; como pode ser visto na figura 8, essa expressão retorna o valor de 7,44, e será usado arredondamento para cima, portanto, $k = 8$, ou seja, a tabela de distribuição de frequências terá 8 classes.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	154	188	160	152	164	164	180	180	175	170		MENOR	151
2	160	151	180	154	177	185	178	155	168	166		MAIOR	189
3	187	176	170	176	169	163	182	187	156	156		AA	38
4	182	186	177	170	158	173	155	171	151	187		CLASSES	7,44
5	153	164	156	151	160	173	165	177	183	179			
6	153	156	159	165	186	167	155	157	159	188			
7	153	178	153	174	185	183	161	180	172	158			
8	170	168	152	180	175	173	183	167	176	151			
9	186	189	154	177	159	171	175	176	181	174			
10	168	155	169	170	170	157	184	186	164	165			

Figura 8: Número de Classes pela Regra de Sturges

(d) *A Amplitude do Intervalo de Classe (h):*

Como foi visto anteriormente, a Amplitude do Intervalo de Classe (h) será calculada dividindo-se a amplitude amostral (AA), que já foi verificado ser igual a 38, pelo número de classes, que também já foi determinado ser $k = 8$, portanto:

$$h = \frac{AA}{K} = \frac{38}{8} = 4,75$$

e como discutido, esse valor será arredondado para cima, tendo-se $h = 5$.

(e) *A Tabela de Distribuição de Frequências:*

Agora podem-se dar os passos para a construção da Tabela de Distribuição de Frequências propriamente dita, feita na mesma planilha, através da inserção de oito colunas correspondentes as classes, dados e as frequências (absolutas, relativas e percentuais, tanto simples quanto acumuladas). Como foi utilizada a amplitude arredondado para cima, pode-se escolher um valor apropriado como limite inferior da primeira classe, nesse caso $l_1 = 150$. Como $h = 5$, a primeira classe será até 155 e como foi discutido, nessa classe serão considerados valores maiores ou iguais a 150 até valores menores que 155 ($150 \leq x < 155$). Deve-se fazer a digitação correspondente até a oitava classe que terá como limite inferior $l_8 = 185$ e limite superior $L_8 = 190$.

Para o preenchimento da coluna da frequência absoluta simples da primeira classe e das subsequentes, será utilizado a função SOMA:

- i. Função SOMA: Soma todos os números em um intervalo de células.

Sintaxe: = *SOMA(células)*

Então deverão ser selecionadas as células da coluna de frequência dos valores que se quer; para a primeira classe, aquelas correspondentes aos valores de 150 a 154; no exemplo, a função será: = *soma(P2 : P5)* pois os valores estão entre as células *P2* até a célula *P5*, como pode-se verificar na figura 7.

A próxima coluna, das Frequências Absolutas Acumuladas (F), como foi definido anteriormente, é a soma da frequência em questão com a somatória das frequências simples anteriores.

A coluna a seguir é a da Frequência Relativa Simples (f_r) que é a frequência absoluta simples dividida pelo número total de dados. No nosso exemplo, para a primeira classe, foi feito: = $T3/100$ e procede-se de maneira análoga para as demais classes (linhas).

A coluna da Frequência Relativa Acumulada (F_r) é a frequência relativa simples da respectiva classe somada com a somatória das frequências relativas simples das classes anteriores.

A coluna da Frequência Percentual Simples (f_p) é o respectivo valor da frequência relativa multiplicado por 100. E de maneira subsequente, a coluna da Frequência Percentual Acumulada (F_p) é a somatória das frequências relativas simples anteriores com a da classe em questão.

Podemos ver na figura 9 a apresentação da respectiva Tabela de Distribuição de Frequências construída no Excel:

	O	P	Q	R	S	T	U	V	W	X	Y
1	VALORES	FREQUÊNCIA	TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS NO EXCEL								
2	151	4		CLASSES	DADOS	f	F	fr	Fr	fp	Fp
3	152	2		1	150 f-155	13	13	0,13	0,13	13	13
4	153	4		2	155 f-160	15	28	0,15	0,28	15	28
5	154	3		3	160 f-165	9	37	0,09	0,37	9	37
6	155	4		4	165 f-170	11	48	0,11	0,48	11	48
7	156	4		5	170 f-175	14	62	0,14	0,62	14	62
8	157	2		6	175 f-180	14	76	0,14	0,76	14	76
9	158	2		7	180 f-185	12	88	0,12	0,88	12	88
10	159	3		8	185 f-190	12	100	0,12	1	12	100
11	160	3			Σ	100	---	1	---	100	---
12	161	1									

Figura 9: Tabela de Distribuição de Frequências construída no Excel

(f) *Histograma:*

Tendo sido feita a Tabela de Distribuição de Frequências, será feito o gráfico tipo Histograma correspondente à frequência absoluta simples. Para isso, deve-se selecionar os valores correspondentes a essa coluna, no caso $T3 : T10$, e no menu superior clicar em "INSERIR", na parte correspondente a "Gráficos", subitem "Inserir Gráfico de Colunas" e selecionar em "Coluna 2D" a opção "Colunas Agrupadas" como pode ser visto na figura 10, então aparecerá o gráfico correspondente na planilha em uso, o usuário poderá posicioná-la adequadamente na planilha para que não fique sobreposta a nenhuma informação importante.

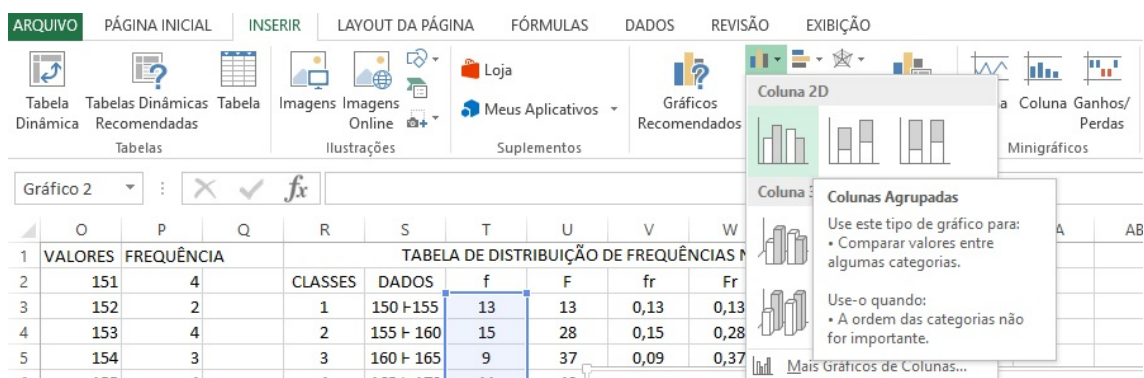


Figura 10: Menu Gráfico do Excel

No entanto, o gráfico apresentado ainda não corresponde ao Histograma de acordo com a sua definição, pois as colunas não são apresentadas de maneira justapostas (uma do lado da outra). É preciso fazer uma correção na sua formatação; para isso, deve-se dar dois cliques rápidos na parte das colunas do gráfico. Assim, aparecerá um menu lateral "Formatar Séries de Dados",

subitem "Opções de Série" e na opção "Largura do Espaçamento" colocar zero. Na figura 11 podemos ver o gráfico resultante e o menu lateral "Formatar Séries de Dados".

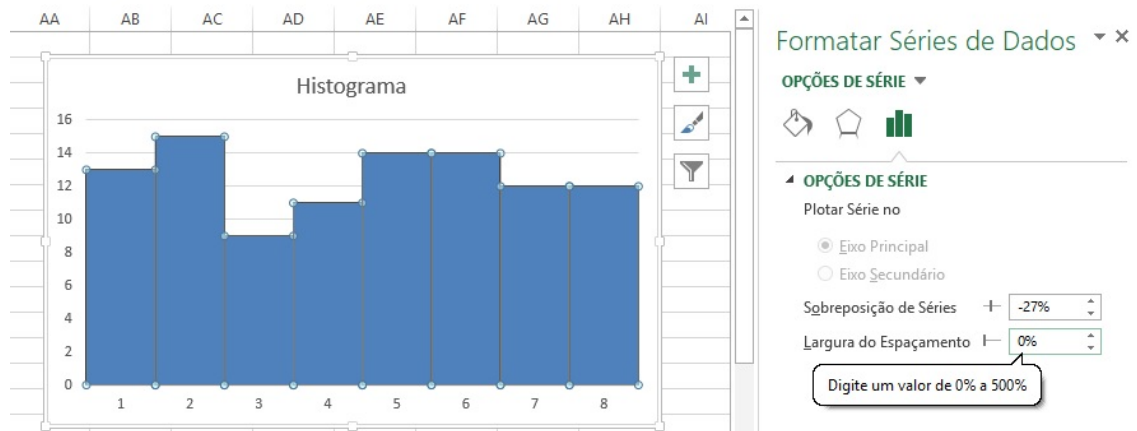


Figura 11: Formatação do Gráfico tipo Histograma

3.2 CÁLCULO DA MÉDIA, VARIÂNCIA E DESVIO PADRÃO NO EXCEL

Tempo estimado da atividade: acredita-se que após a definição dos conceitos, pode-se utilizar duas aulas para a execução no computador.

Exemplo 3.2. Após gerar uma amostra aleatória, calcula-se a média, variância e desvio padrão para amostras de tamanhos: 2, 3, 5, 10, 20, 30 e 50, sendo também apresentado ao fim, para todos os dados de uma população com valores entre 1 e 10. Pode-se assumir, como exemplo, a nota de 100 alunos de uma escola em uma determinada disciplina.

Deve-se conseguir verificar, conforme as figuras 12, 13 e 14; que ao aumentar o tamanho das amostras, a média das médias se aproxima cada vez mais da média de todos os dados, destacado na figura 14, e também têm-se que a variância e o desvio padrão vão diminuindo.

	A	B	C	D	E	F	G	H	I	J
1										
2	DADOS:									
3										
4	10	0	1	7	8	10	5	7	3	3
5	1	9	10	3	9	10	3	5	5	7
6	5	4	1	3	7	5	5	6	0	6
7	9	1	7	5	5	5	4	6	3	2
8	1	2	4	1	2	2	2	1	1	10
9	7	8	3	6	3	3	7	8	6	3
10	9	7	4	9	7	8	6	10	8	6
11	5	9	7	4	8	3	9	0	4	3
12	5	2	2	9	1	3	8	1	6	5
13	2	1	3	3	3	7	3	0	9	10

O17			
K	L	M	N
1	TAMANHO DA AMOSTRA: 02		
2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:
3	=MÉDIA(;	=VAR.A(;	=DESVPAD.A(;
4	5	50	7,071067812
5	5	32	5,656854249
6	4,5	0,5	0,707106781
7	5	32	5,656854249
8	1,5	0,5	0,707106781
9	7,5	0,5	0,707106781
10	8	2	1,414213562
11	7	8	2,828427125
12	3,5	4,5	2,121320344
13	1,5	0,5	0,707106781
14			
15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:
16	4,85	5,113888889	2,261390919

A1			
K	L	M	N
1	TAMANHO DA AMOSTRA: 03		
2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:
3	=MÉDIA(;	=VAR.A(;	=DESVPAD.A(;
4	3,666666667	30,33333333	5,507570547
5	6,666666667	24,33333333	4,932882862
6	3,333333333	4,333333333	2,081665999
7	5,666666667	17,33333333	4,163331999
8	2,333333333	2,333333333	1,527525232
9	6	7	2,645751311
10	6,666666667	6,333333333	2,516611478
11	7	4	2
12	3	3	1,732050808
13	2	1	1
14			
15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:
16	4,633333333	3,813580247	1,952839022

R18			
K	L	M	N
1	TAMANHO DA AMOSTRA: 05		
2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:
3	=MÉDIA(;	=VAR.A(;	=DESVPAD.A(;
4	5,2	19,7	4,438468204
5	6,4	16,8	4,098780306
6	4	5	2,236067977
7	5,4	8,8	2,966479395
8	2	1,5	1,224744871
9	5,4	5,3	2,302172887
10	7,2	4,2	2,049390153
11	6,6	4,3	2,073644135
12	3,8	10,7	3,271085447
13	2,4	0,8	0,894427191
14			
15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:
16	4,84	3,073777778	1,753219261

Figura 12: População e Distribuição Amostral das Médias de Amostras de Tamanhos 02, 03 e 05

S17				AR22			
K	L	M	N	K	L	M	N
1	TAMANHO DA AMOSTRA: 10			1	TAMANHO DA AMOSTRA: 20		
2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:	2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:
3	=MÉDIA(;))	=VAR.A(;))	=DESVPAD.A(;))	3	=MÉDIA(;))	=VAR.A(;))	=DESVPAD.A(;))
4	5,4	12,71111111	3,565264522	4	5,8	11,22105263	3,349783968
5	6,2	10,62222222	3,259175083	5	5,2	8,484210526	2,912766816
6	4,2	5,066666667	2,250925735	6	4,45	5,102631579	2,258900524
7	4,7	5,566666667	2,359378449	7	3,65	7,397368421	2,719810365
8	2,6	7,6	2,75680975	8	4	7,894736842	2,809757435
9	5,4	4,711111111	2,170509413	9	6,4	4,778947368	2,186080366
10	7,4	3,155555556	1,776388346	10	6,3	6,957894737	2,637782163
11	5,2	8,844444444	2,97396107	11	4,7	8,326315789	2,885535616
12	4,2	8,177777778	2,859681412	12	4,15	9,292105263	3,048295468
13	4,1	11,43333333	3,381321241	13	4,75	11,88157895	3,44696663
14				14			
15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:	15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:
16	4,94	1,718222222	1,310809758	16	4,94	0,921555556	0,959976852

AQ21				W13			
K	L	M	N	L	M	N	
1	TAMANHO DA AMOSTRA: 30			1	TAMANHO DA AMOSTRA: 50		
2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:	2	MÉDIA AMOSTRAL:	VARIÂNCIA AMOSTRAL:	DESVIO PADRÃO AMOSTRAL:
3	=MÉDIA(;))	=VAR.A(;))	=DESVPAD.A(;))	3	=MÉDIA(;))	=VAR.A(;))	=DESVPAD.A(;))
4	5,266666667	9,512643678	3,084257395	4	4,62	9,138367347	3,022973263
5	5,033333333	7,343678161	2,709922169	5	4,62	7,668979592	2,769292255
6	3,833333333	6,488505747	2,547254551	6	4,86	7,306530612	2,703059491
7	4,233333333	7,012643678	2,648139664	7	5,06	7,894285714	2,809677155
8	5,133333333	8,809195402	2,968028875	8	4,96	8,488163265	2,913445257
9	6	6,206896552	2,491364396	9	5,26	8,114693878	2,848630176
10	5,6	8,110344828	2,847866715	10	5,26	9,584081633	3,095816796
11	4,5	9,086206897	3,014333574	11	5,02	10,14244898	3,184721178
12	4,566666667	10,39195402	3,223655382	12	4,82	9,538367347	3,088424736
13	5,233333333	11,56436782	3,400642265	13	4,92	8,973061224	2,995506839
14				14			
15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:	15	MÉDIA DAS MÉDIAS:	VARIÂNCIA DA MÉDIA AMOSTRAL:	DESVIO DA MÉDIA AMOSTRAL:
16	4,94	0,428345679	0,654481229	16	4,94	0,050222222	0,224103151

Figura 13: Distribuição Amostral das Médias de Amostras de Tamanhos 10, 20, 30 e 50

A1	:	X	✓	<i>fx</i>	
	K	L	M	N	
1	TODA POPULAÇÃO				
2		MÉDIA POPULAÇÃO:	VARIÂNCIA POPULAÇÃO:	DESVIO PADRÃO POPULAÇÃO:	
3		=MÉDIA(,;)	=VAR.P(,;)	=DESV.PAD.P(,;)	
4		4,94	8,5564	2,939868753	

Figura 14: Estatísticas com os dados populacionais

Na figura 15 pode-se ver também, o gráfico Histograma correspondente as amostras de tamanho 2, 3, 5, 10, 20 e 30.

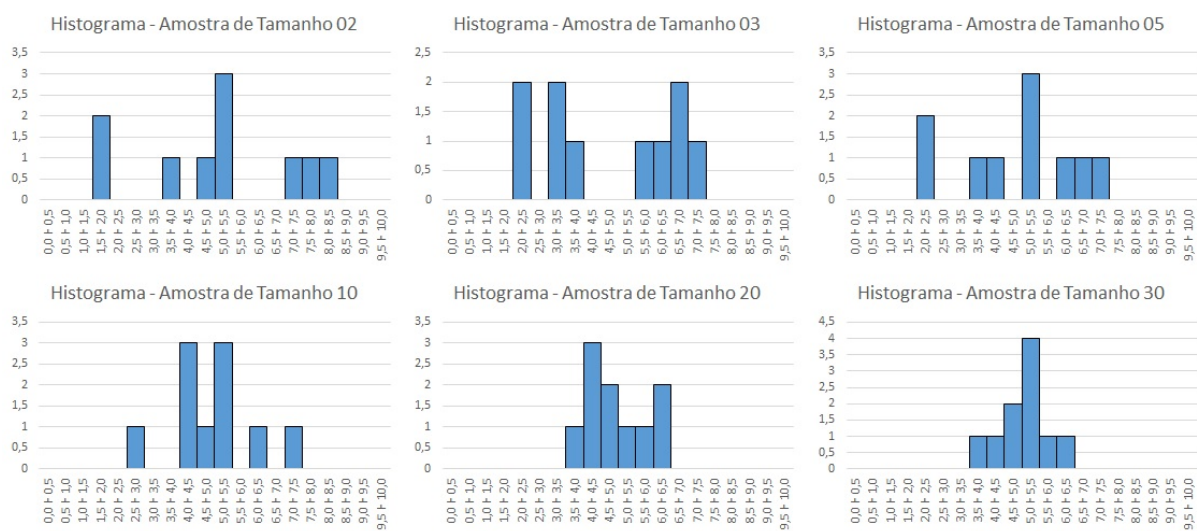


Figura 15: Histogramas

E na figura 16 tem-se em destaque o gráfico Histograma, correspondente a amostra de tamanho 50.

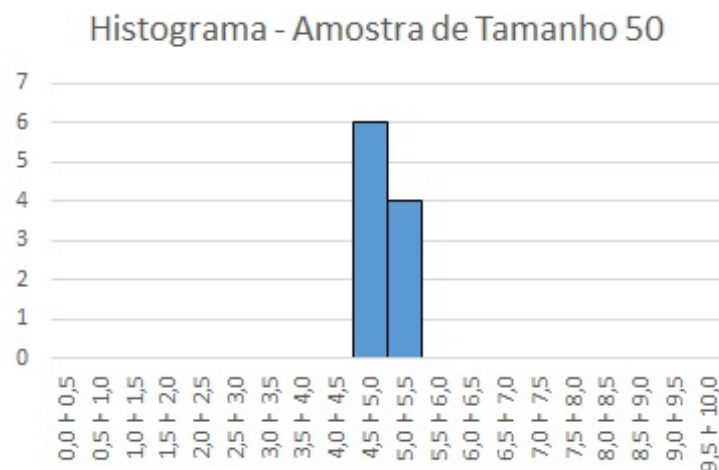


Figura 16: Histograma - Amostra de Tamanho 50

Mesmo o exercício tendo sido feito para apenas dez medições (quando o aconselhável é que seja maior que 30), é possível verificar através do gráfico Histograma que, quanto maior o tamanho da amostra, mais a distribuição converge para o valor da média da população (= 4,94).

3.3 DISTRIBUIÇÃO AMOSTRAL DAS MÉDIAS NO EXCEL

Tempo estimado da atividade: acredita-se que após a definição dos conceitos, pode-se utilizar duas aulas para a execução no computador.

Exemplo 3.3. *Nesse exemplo, será apresentada a simulação do gráfico histograma da distribuição amostral das médias, correspondente as médias de uma população de 0 a 10, tomadas a partir de 50 medições, de amostras de tamanhos 02, 05, 10, 20, 30, 40 e 50. Deve-se destacar a importância de se notar que, quanto maior o tamanho da amostra, mais o gráfico vai tomando a forma de "sino", como pode-se ver nas figuras 17 e 18.*

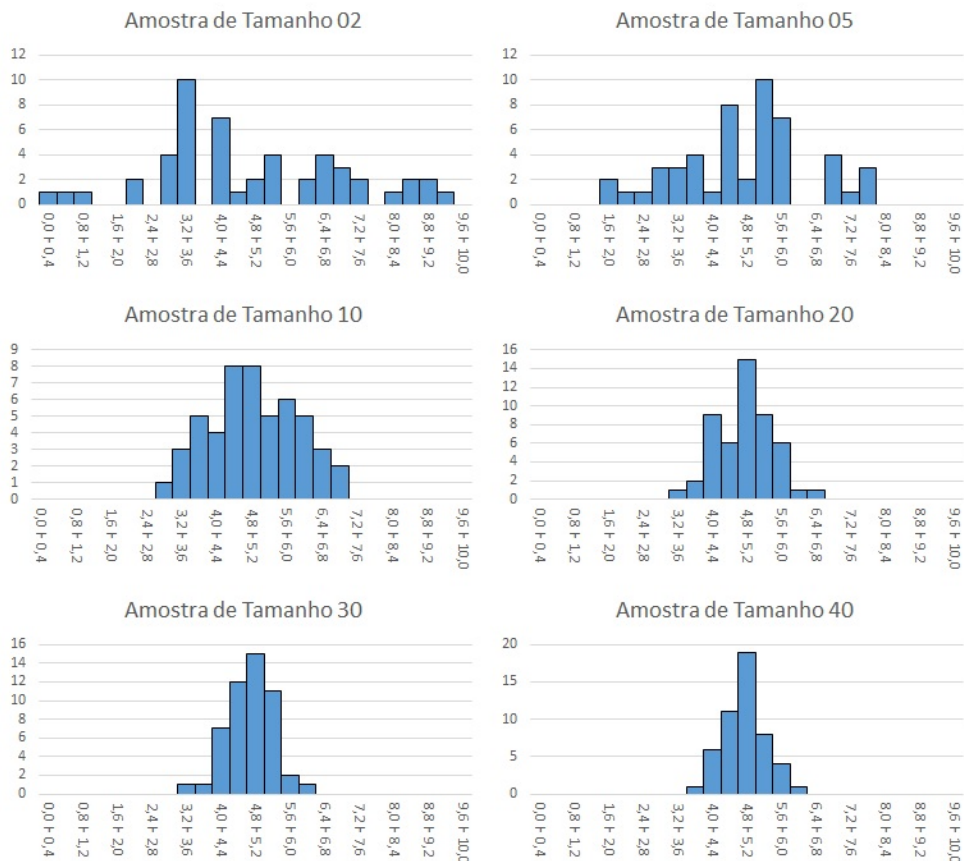


Figura 17: Histogramas para Amostras de Tamanho: 02, 05, 10, 20, 30 e 40

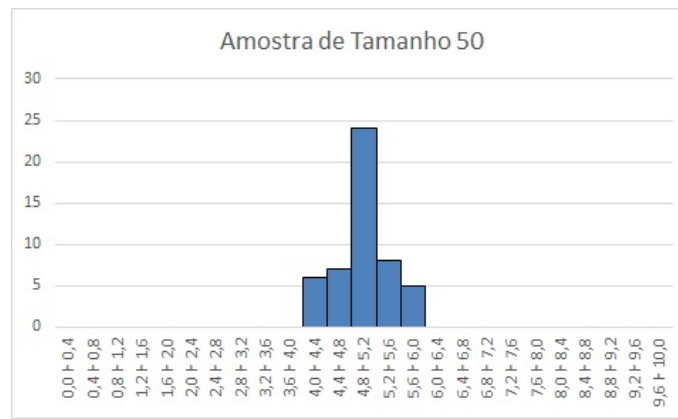


Figura 18: Histograma - Amostra de Tamanho: 50

3.4 ATIVIDADE INVESTIGATIVA EM ESTATÍSTICA

Nesse tópico, será apresentado um problema exemplo que deverá ser explorado como base para atividade investigativa, bem como as etapas que podem ser seguidas para direcionar os conceitos que se deseja discutir.

Tempo estimado da atividade: acredita-se que pode-se utilizar duas aulas para a o experimento em sala e mais duas aulas para a execução no computador.

Exemplo 3.4. *Situação Problema e discussões inspirado em [18]: Um dado honesto é lançado três vezes. Sendo Y_1 , Y_2 e Y_3 o número de pontos obtidos na face superior para os lançamentos 1, 2 e 3 respectivamente. Suponha que estamos interessados em $\bar{Y} = \frac{Y_1 + Y_2 + Y_3}{3}$, o número médio dos pontos observados nos três lançamentos. Questiona-se:*

- (a) *Quais são a média $\mu_{\bar{Y}}$ e o desvio padrão $\sigma_{\bar{Y}}$ de \bar{Y} ?*
- (b) *Como podemos determinar a distribuição amostral de \bar{Y} ?*

3.4.1 Estratégias de Execução da Atividade

1. Introdução com debate sobre a importância dos aspectos da estatística no mundo atual;
2. Leitura direcionada e debate sobre a situação problema;
3. Explicação da atividade a ser desenvolvida;
4. Definição do objetivo da pesquisa - Distribuição Amostral das Médias;
5. Divisão da turma em equipes, e divisão das tarefas e cooperação entres os componentes;

6. Definição da variável que será analisada, bem como a quantidade de medições (os grupos poderão fazer quantidades de medições diferentes);
7. Coleta dos dados - lançamento do dado de seis faces e registro dos valores;
8. Lançamento das informações no Excel;
9. Manipulação no Excel - cálculo das médias, distribuição de frequências, construção do gráfico Histograma;
10. Conclusão e relatório dos resultados obtidos;
11. Socialização das conclusões dos grupos e debate em sala mediado pelo professor.

3.4.2 Simulação da Atividade e Proposta da Discussão dos Resultados Obtidos

A questão a ser motivada pra discussão nesse momento é: "*Como podemos calcular a distribuição da variável aleatória \bar{Y} ?*".

Pode-se então explorar uma possível argumentação, que segue:

Os possíveis valores da variável aleatória $W = Y_1 + Y_2 + Y_3$ são 3, 4, 5, ..., 18 e $\bar{Y} = \frac{W}{3}$. Como o dado é honesto, cada um dos $6^3 = 216$ valores distintos da variável aleatória multivariável (Y_1, Y_2, Y_3) são igualmente prováveis e

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = p(y_1, y_2, y_3) = 1/216$$

em que $y_i = 1, 2, 3, 4, 5, 6$; e $i = 1, 2, 3$.

Portanto,

$$P(\bar{Y} = 1) = P(W = 3) = p(1, 1, 1) = 1/216$$

$$P(\bar{Y} = 4/3) = P(W = 4) = p(1, 1, 2) + p(1, 2, 1) + p(2, 1, 1) = 3/216$$

$$P(\bar{Y} = 5/3) = P(W = 5) =$$

$$p(1, 1, 3) + p(1, 3, 1) + p(3, 1, 1) + p(1, 2, 2) + p(2, 1, 2) + p(2, 2, 1) = 6/216$$

As probabilidades $P(\bar{Y} = i/3)$, $i = 7, 8, \dots, 18$ são obtidos de maneiras análoga.

A dedução da distribuição amostral da variável aleatória \bar{Y} discutida acima, utiliza o cálculo de todos valores possíveis, que tem como desvantagem, ou ser de grande dificuldade em completar os cálculos para a distribuição amostral exata para \bar{Y} ou o processo em si é muito tedioso.

Questão a ser discutida: "*Como podemos ter então uma ideia da forma da distribuição amostral sem o inconveniente de ter que se fazer todos os cálculos?*"

Uma forma é simular a distribuição amostral ao se tomar amostras independentes repetidas, cada uma de tamanho três, calculando o valor observado \bar{y} para cada amostra y e construindo um histograma destes valores observados.

Nessa simulação, serão consideradas 100 observações de tamanho três (ou seja, 300 lançamentos), onde após registrado, as informações são digitadas no Excel e então é feito o cálculo da média dos três lançamentos. Na figura 19 pode-se ver uma parte da tabela.

Para o cálculo da média dos três lançamentos foi utilizado o comando: =MÉDIA(A3:C3), copiado respectivamente para as demais linhas.

	A	B	C	D	E	F
1	LANÇAMENTOS					
2	PRIMEIRO	SEGUNDO	TERCEIRO		MÉDIA	DESVIO PADRÃO
3	3	2	4		3,00	1,00000
4	3	4	2		3,00	1,00000
5	5	1	4		3,33	2,08167
6	1	3	1		1,67	1,15470
7	2	4	3		3,00	1,00000
8	2	4	3		3,00	1,00000
9	3	6	3		4,00	1,73205
10	6	5	3		4,67	1,52753
11	4	6	4		4,67	1,15470
12	2	1	5		2,67	2,08167
13	2	1	2		1,67	0,57735
14	4	4	1		3,00	1,73205
15	5	2	6		4,33	2,08167
16	1	3	4		2,67	1,52753
17	3	4	3		3,33	0,57735
18	2	3	1		2,00	1,00000
19	5	3	6		4,67	1,52753
20	5	4	2		3,67	1,52753
21	4	6	2		4,00	2,00000
22	3	3	6		4,00	1,73205
23	4	5	3		4,00	1,00000
24	4	1	4		3,00	1,73205
25	4	3	6		4,33	1,52753

Figura 19: Lançamentos Registrados no Excel

Então pode-se também calcular a média das médias e o desvio padrão das médias, cujo resultado pode ser visto na figura 20. Para estes cálculos foram utilizados os seguintes comandos:

1. =MÉDIA(E3:E52)
2. =DESVPAD.A(E3:E52)

O resultado do gráfico Histograma obtido desta simulação pode ser visto na figura 21: Questão a ser discutida: "O que se pode observar na figura 21?"

Como já abordado, o máximo valor observado de \bar{Y} é 6 e o valor mínimo é 1. Também tem-se que os valores obtidos na simulação, se acumulam em forma de sino centrado em $10/3 \approx 3,3$ e $11/3 \approx 3,7$ (próximo a 3,5 que é a média teórica de \bar{Y}). Na figura 20 tem-se que a média e o desvio padrão dos cem valores simulados de \bar{Y} são respectivamente $\mu_{\bar{Y}} = 3,49$ e $\sigma_{\bar{Y}} = 0,83344$ (que são muito próximos dos valores teóricos $\mu_{\bar{Y}} = 3,5$ e $\sigma_{\bar{Y}} = 0,9860$ - tais valores podem ser demonstrados mas não serão apresentados pois não faz parte do objetivo da atividade).

96	3	5	4		4,00	1,00000
97	3	6	6		5,00	1,73205
98	4	1	1		2,00	1,73205
99	1	2	2		1,67	0,57735
100	3	3	1		2,33	1,15470
101	5	1	1		2,33	2,30940
102	5	2	4		3,67	1,52753
103						
104					MÉDIA	DESVIO PADRÃO
105					3,49	0,83344

Figura 20: Média e Desvio Padrão das Médias dos Lançamentos

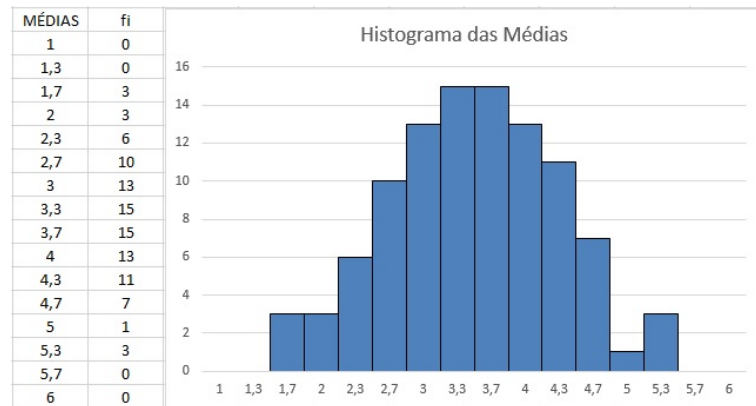


Figura 21: Histograma da Simulação

Questão a ser discutida: *"Qual a relevância desse tipo de ocorrência?"*

Várias situações utilizam a aplicação de uma simulação para explorar a distribuição amostral das médias para diferentes tamanhos de amostras e para situações em que o conjunto de dados são muito grandes. Outras ocorrências demandam o uso da distribuição amostral para o cálculo da média e a variância de amostras de uma distribuição em forma de "sino".

Assim como a distribuição amostral simulada, a forma da distribuição da amostragem teórica de qualquer estatística dependerá da distribuição das variáveis aleatórias observadas na amostra e tais estatísticas podem ser usadas para fazer inferências que bem representam o todo a partir de parte desse todo.

CONCLUSÕES

É notória a necessidade de repensar o atual sistema educacional, as crianças e jovens dos dias atuais são em muitos aspectos diferentes das crianças e jovens de algumas décadas atrás, pois desde muito cedo vivenciam novas tecnologias com um volume de informação enorme. Os desafios para a educação são grandes, uma vez que a sociedade passa por um período de intensas e rápidas transformações, em que as organizações escolares apresentam dificuldades para se prepararem e poder então acompanhá-las; poder público, escolas e professores aparentam ter dificuldade em conseguir convergir na solução dos desafios, o que faz um processo de reformulação andar a passos lentos, enquanto os avanços tecnológicos e a informatização evoluem cada vez mais rápido.

O ensino em nossas escolas ainda tem como base elementar a utilização de livros didáticos e quadro-giz, enquanto nossos alunos são cada vez mais digitais, e as informações estão ao alcance de suas mãos praticamente de forma instantânea. Esse fato aumenta ainda mais a responsabilidade dos professores na tentativa de buscar proporcionar ao seus alunos uma formação sólida e crítica, pois esse excesso de informação disponível para nossos jovens deixa mais difícil a tarefa de fazer com que eles aprendam a analisar fatos criticamente com imparcialidade e de maneira não apenas superficial.

Nesse sentido, reforça-se a importância de se trabalhar unindo os recursos clássicos já existentes no ambiente escolar à utilização dos recursos tecnológicos computacionais, não só na Estatística dentro da Matemática mas em todos os conteúdos curriculares. Acreditar e convencer que o uso dos novos recursos tecnológicos trarão efeitos positivos na formação dos alunos é fator crucial na atualização do atual sistema educacional, proporcionando uma transformação na didática existente, e dessa forma, naturalmente os professores se sentirão confiantes e motivados a buscar a capacitação necessária para conseguir inserir esses novos recursos no desenvolvimento de suas aulas e práticas didáticas, não ficando restritos simplesmente às ferramentas clássicas de livros didáticos e quadro-giz.

Na tentativa de uma abordagem para aproveitar os recursos computacionais e novas tecnologias, o Excel se mostra uma excelente e viável ferramenta, que permite abreviar cálculos repetitivos e enfadonhos, que não contribuem para a compreensão dos conceitos trabalhados, o que muitas vezes desmotiva os alunos, possibilitando direcionar o tempo disponível poupado para que seja aproveitado na realização de experimentos e análise de dados organizados na forma de tabelas e gráficos, sem que esse trabalho se torne cansativo e de desinteresse para eles. Porém, o professor deve ter o cuidado de planejar suas aulas para que o uso de planilhas não fique restrito a um mero repositório de fórmulas, o que poderia ser reducionista no processo ensino-aprendizagem.

As atividades sugeridas nesse trabalho são simples, exigem pouco conhecimento específico na área de informática, para serem desenvolvidos; mas atendem à indicação dos objetivos presentes nos Parâmetros Curriculares Nacionais e na Base Comum Curricular. Dessa forma pretende-se mostrar que a inserção das tecnologias computacionais no trabalho pedagógico é possível, bastando que para isso exista motivação para utilizá-las, bem como disponibilidade de tempo para se dedicar na elaboração e realização das aulas, não devendo ser esquecido que o laboratório de informática deve, além de existir, ter o mínimo de estrutura para que as aulas possam ser efetivamente desenvolvidas com qualidade.

Uma perspectiva futura é a de aplicar essa proposta de ensino em turmas do Ensino Médio (havendo até a possibilidade, de que, com algumas adaptações, possa ser aplicada para os anos finais do Ensino Fundamental, principalmente nono ano) visando analisar os frutos que essa vivência poderá produzir, para que, então, possam ser aprimoradas as atividades apresentadas na busca de uma melhor adequação delas ao cotidiano dos alunos. Incentiva-se que as aplicações possam ser propostas por meio da realização de pesquisas que visem identificar as reais contribuições e benefícios do uso das planilhas, por meio dos quais o trabalho pode ser conduzido de duas formas, uma em que os alunos utilizem as planilhas eletrônicas e outra não, possibilitando que haja parâmetro de comparação e posterior discussão para os resultados obtidos em ambas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BRASIL. Secretaria de Educação Fundamental. **Parâmetros curriculares nacionais: Matemática**/Secretaria de Educação Fundamental. Brasília : MEC /SEF, 1998.148 p. Disponível em: portal.mec.gov.br/seb/arquivos
- [2] CARVALHO A, **A importância do ensino de Estatística na formação inicial do professor de Matemática**. *XIX Encontro Brasileiro de Estudantes de Pós-Graduação em Educação Matemática*, Anais Seção D10, Juiz de Fora - Minas Gerais. Disponível em: <http://www.ufjf.br/ebrapem2015/files/2015/10/Artigo-EBRAPEM-Revisado.pdf>, (2015)
- [3] CAZORLA, I.; MAGINA, S.; GITIRANA, V.; GUIMARÃES, G. **Estatística para os anos iniciais do ensino fundamental**, Sociedade Brasileira de Educação Matemática - SBEM, Brasília (2017). Disponível em: http://www.sbem.com.br/files/ebook_sbem.pdf
- [4] CRESPO, Antônio Arnot. **Estatística Fácil**. 19.ed. São Paulo: Saraiva,2009.218 p.
- [5] DE MELO, R.A.F.P. **Uma Análise Sobre o Estudo de Amostragem no Ensino Médio**. Dissertação - Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, (2017), 80p.
- [6] ECHEVESTE, S.; ÁVILA, M.G. **Estatística no Ensino Fundamental e Médio**, *Acta Scientiae*, v.4, n.1 (2002), p.91-96
- [7] FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade. **Curso de Estatística**. 6. ed. São Paulo: Atlas S.a., 1996. 320 p.
- [8] HOEL, Paul G.; PORT, Sidney C.; STONE, Charles J.. **Introdução à Teoria da Probabilidade**. Rio de Janeiro: Interciência, 1978. 269 p. Tradução: Fernando Yassou Chiyoshi.
- [9] LAPPONI, J. C. **Estatística usando Excel**. 4. ed. Rio de Janeiro: Elsevier, 2005. 476 p.
- [10] LARSON, Ron; FARBER, Betsy. **Estatística Aplicada**. 4. ed. São Paulo: Pearson Prentice Hall, 2010. 527 p. Tradução: Luciane Paulete Viana.
- [11] MINISTÉRIO DA EDUCAÇÃO. **Base Nacional Comum Curricular**, (2016). Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=

40791-bncc-proposta-preliminar-segunda-versao-pdf&category_slug=maio-2016-pdf&Itemid=30192

- [12] MORETTIN, Luiz Gonzaga. **Estatística Básica**: Probabilidade e Inferência, Volume Único. São Paulo: Pearson Prentice Hall, 2010.
- [13] MORETTIN, Pedro Alberto; BUSSAB, Wilton de O.. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010. 540 p. Citação com autor incluído no texto: Morettin e Bussab (2010).
- [14] ROSS, Sheldon. **Probabilidade**: Um curso moderno com aplicações. 8. ed. Porto Alegre: Bookman, 2010. 608 p. Tradutor: Alberto Resende De Conti.
- [15] SILVA, Mário Gomes da. **Informática**: Terminologia Básica - Volume 1. 2. ed. São Paulo: Érica, 1998. 212 p.
- [16] SILVA, Mário Gomes da. **Informática**: Volume 2. 2. ed. São Paulo: Érica, 1998. 260 p.
- [17] SPIEGEL, Murray R.. **Estatística**. 2. ed. [s.l.]: Mcgraw-hill do Brasil, Ltda., 1970. 580 p. Traduzido por Pedro Cosentino.
- [18] WACKERLY, Dennis D.; III, William Mendenhall; SCHEAFFER, Richard L.. **Estatística matemática con aplicaciones**. 7. ed. Santa Fe, México: Cengage Learning, 2010. 911 p. Jorge Humberto Romo Munoz.

ANEXO

DISTRIBUIÇÃO NORMAL

A Distribuição Normal, também conhecida como Distribuição de Gauss, Laplace ou Laplace-Gauss; é um dos mais importantes exemplos de distribuição contínua de probabilidade, sendo aplicada em inúmeros fenômenos e utilizada para o desenvolvimento teórico da estatística. Para sua definição, precisa-se dos conceitos de variável aleatória contínua e de função densidade de probabilidade, que serão apresentados a seguir.

Variável Aleatória Contínua e Função Densidade de Probabilidade

Seja X uma variável aleatória, cujo contradomínio $X(\Omega)$ é infinito não-enumerável, como por exemplo, um intervalo de números reais. Sejam a variável aleatória X e o evento $a \leq X \leq b$ do espaço amostral Ω . Conseqüentemente a probabilidade desse evento, denotado por $P(a \leq X \leq b)$, está bem definida. Considerando que haja uma função contínua $f : R \rightarrow R$, tal que $P(a \leq X \leq b)$ seja igual a área sob o gráfico de $f(x)$ entre $x = a$ e $x = b$.

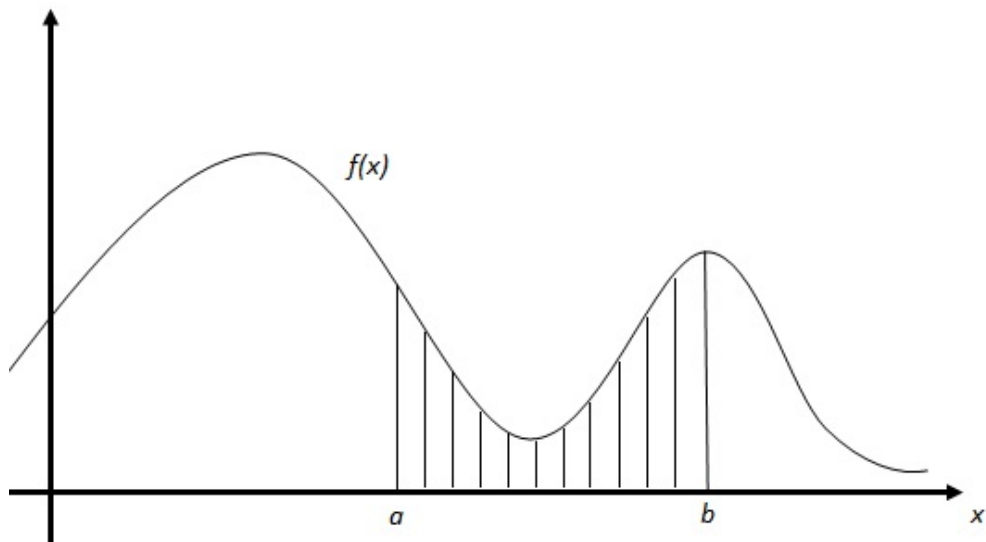


Figura 22: Função Contínua - $P(a \leq X \leq b) = \text{área da região hachurada}$

Em linguagem matemática, tem-se:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Neste caso, diz-se que X é uma variável aleatória contínua e $f(x)$ é denominada função densidade de probabilidade (f.d.p.) ou (função) distribuição de probabilidade contínua, desde que satisfaça as seguintes condições:

- $f(x) \geq 0, \forall x \in R$;
- $\int_{-\infty}^{\infty} f(x)dx = 1$.

Isto é, $f(x)$ é uma função não negativa e a área total sob seu gráfico é igual a 1. A esperança matemática $E(X)$, quando existe, é definida por:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx$$

A variância $Var(X)$ é definida por:

$$Var(X) = E((X - \mu_X)^2) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)dx,$$

quando o lado direito existe. $Var(X)$ existe se, e somente se, $\mu_X = E(X)$ e $E(X^2)$ existem e então:

$$Var(X) = E(X^2) - \mu_X^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu_x^2$$

O desvio padrão é definido por $\sigma_X = \sqrt{Var(X)}$, quando $Var(X)$ existe.

Distribuição Normal

Como dito, a distribuição Normal é a mais importante distribuição de probabilidade, sendo aplicada em inúmeros fenômenos. Uma variável aleatória contínua X terá Distribuição de Probabilidade Normal se sua função densidade de probabilidade (f.d.p.) é dada por:

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

onde:

μ = média da distribuição;

σ = desvio padrão da distribuição;

$\pi = 3,14159... ;$

$e = 2,71828... ;$

Para se registrarem distribuições normais usa-se a seguinte notação:

$X : N(\mu; \sigma^2) \Rightarrow$ Lê-se: a variável aleatória X tem distribuição normal com média μ e variância σ^2 .

Irá nos interessar a área total limitada pela curva correspondente a função densidade de probabilidade (f.d.p.), representada por $f(x)$ e pelo eixo dos X , cujo gráfico é apresentado pela figura 23.

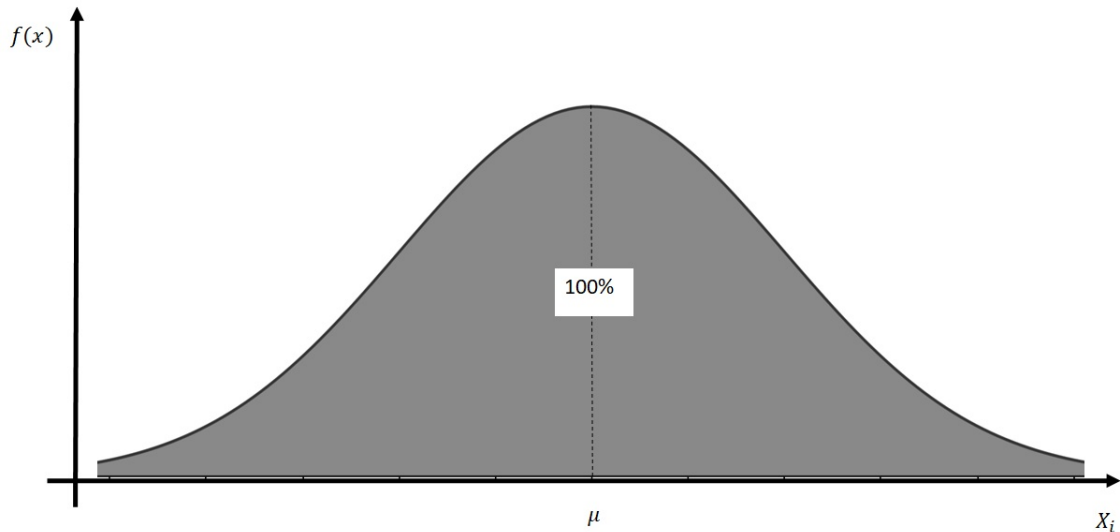


Figura 23: Distribuição Normal

A área total limitada pela curva correspondente a função densidade de probabilidade (f.d.p.) da Distribuição de Probabilidade Normal e pelo eixo dos X é igual a 1; portanto a área sob a curva, compreendida entre as duas abscissas $X = a$ e $X = b$, em que $a < b$, representa a probabilidade de X estar situado entre a e b , representada por $P\{a < X < b\}$.

Para o cálculo das probabilidades, surgem dois grandes problemas: primeiro, para a integração de $f(x)$, pois para o cálculo é necessário o desenvolvimento em séries; segundo, seria a elaboração de uma tabela de probabilidades, pois $f(x)$ depende de dois parâmetros, fato este que acarretaria um grande trabalho para tabelar essas probabilidades considerando-se as várias combinações de μ e σ^2 .

Esses problemas podem ser solucionados por meio de uma mudança de variável, obtendo-se, assim, a distribuição normal padronizada (ou reduzida).

Distribuição Normal Padrão

Seja Z uma variável aleatória contínua tal que:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

em que X é uma variável aleatória com distribuição normal de média μ e variância σ^2 , então Z terá distribuição normal padrão.

A média de z será:

$$E[z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma}\{E[X - \mu]\} = \frac{1}{\sigma}\{E[X] - E[\mu]\} = \frac{1}{\sigma}\{\mu - \mu\} = 0$$

E sua variância:

$$Var[z] = Var\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma^2}\{Var[X - \mu]\} = \frac{1}{\sigma^2}\{Var[X]\} = \frac{\sigma^2}{\sigma^2} = 1$$

Logo, a função densidade de probabilidade da distribuição normal padrão é dada por:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty$$

Sendo o gráfico de $\phi(z)$ igual a:

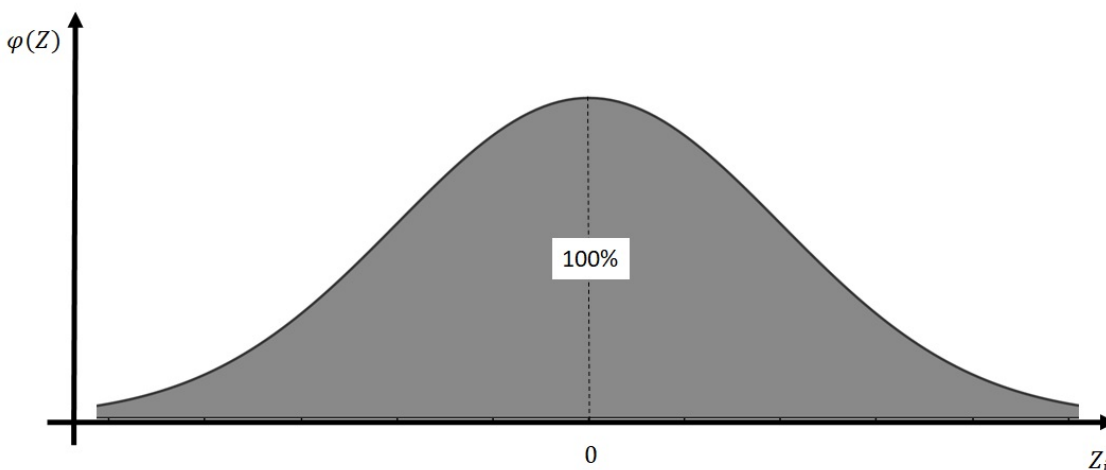


Figura 24: Distribuição Normal Padrão

Como, de acordo com o que foi apresentado, a média de Z é $\mu = 0$ e a variância $\sigma^2 = 1$, as probabilidades (áreas) sob $\phi(z)$ são calculadas e tabeladas.

Para se registrarem distribuições normais padrão usa-se a seguinte notação:

$Z : N(0; 1) \Rightarrow$ Lê-se: a variável aleatória Z tem distribuição normal com média 0 e variância 1. Ou, simplesmente, distribuição normal padrão.

Probabilidade de um Evento

Seja o evento $E = \{a \leq Z \leq b\}$, então a probabilidade do evento E ocorrer é dada por:

$$P(E) = P(a \leq Z \leq b) = \int_a^b \phi(z) dz$$

A integral só tem solução numérica, porém, como ela só depende de Z pode ser calculada uma só vez e então guardada numa tabela.

Relação entre as Distribuições Normal e Normal Padrão

Seja os eventos $E_1 = \{x_1 \leq X \leq x_2\}$ e $E_2 = \{z_1 \leq Z \leq z_2\}$ tal que:

$$z_1 = \frac{(x_1 - \mu)}{\sigma} \text{ e } z_2 = \frac{(x_2 - \mu)}{\sigma},$$

então, as probabilidades dos eventos E_1 e E_2 ocorrerem são iguais, ou seja:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x) dx = P(z_1 \leq Z \leq z_2) = \int_{z_1}^{z_2} \phi(z) dz$$

Como a integral em Z está tabelada, em geral, usa-se esta propriedade para se calcular a probabilidade do evento E_1 .

As definições, conceitos e teoremas desta seção foram pesquisados nas referências: [7], [8], [14].

TEOREMA FUNDAMENTAL DA ESTATÍSTICA – TEOREMA DO LIMITE CENTRAL

O Teorema do Limite Central (TLC) pode ser considerado um dos resultados mais extraordinários na teoria da Probabilidade. Em linhas gerais, ele afirma que a soma de um grande número de variáveis aleatórias independentes tem uma distribuição que é aproximadamente normal. Com isso, ele não somente fornece um método simples para o cálculo de probabilidades aproximadas para soma de variáveis aleatórias independentes, mas também ajuda a explicar o impressionante fato de que frequências empíricas de muitas populações exibem curvas na forma de um sino (isto é, normais).

Para o Teorema a seguir, é necessário saber que Φ representa a função distribuição normal padrão dada por:

$$\Phi(a) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^a e^{-x^2/2} dx, \quad -\infty < a < \infty.$$

De uma forma simples, o Teorema do Limite Central, pode ser enunciado da seguinte maneira:

Teorema A.1. *Seja X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas (isto é, todas têm a mesma função de probabilidade no caso discreto ou a mesma função de densidade no caso contínuo), cada um com média μ e variância σ^2 . Então, a distribuição de:*

$$S_n^* = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

tende a distribuição normal padrão quando $n \rightarrow \infty$.

Isto é, para $-\infty < a < \infty$, temos que:

$$P \left\{ \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^a e^{-x^2/2} dx \text{ quando } n \rightarrow \infty.$$

A chave para a demonstração do Teorema do Limite Central é o Lema a seguir, que será enunciado sem apresentar uma demonstração.

Lema A.2. *Seja Z_1, Z_2, \dots uma sequência de variáveis aleatórias com funções distribuição F_{Z_n} e funções geratrizes de momentos $M_{Z_n}, n \geq 1$; seja também Z uma variável aleatória com função distribuição F_Z e função geratriz de momentos M_Z . Se $M_{Z_n}(t) \rightarrow M_Z(t)$ para todo t , então $F_{Z_n}(t) \rightarrow F_Z(t)$ para todo t no qual $F_Z(t)$ é contínua.*

Se Z é uma variável aleatória normal padrão, então, como $M_Z(t) = e^{t^2/2}$, obtém-se do Lema A.2 que, se $M_{Z_n}(t) \rightarrow e^{t^2/2}$ quando $n \rightarrow \infty$, então $F_{Z_n}(t) \rightarrow \Phi(t)$ (sendo Φ a função de distribuição normal padrão) quando $n \rightarrow \infty$.

Agora pode-se demonstrar o Teorema do Limite Central.

Demonstração. Teorema do Limite Central: Vamos supor primeiro que $\mu = 0$ e $\sigma^2 = 1$. Vamos demonstrar o Teorema sob a hipótese de que a função geratriz de momentos de $X_i, M(t)$, exista e seja finita. Agora, a função geratriz de momentos de X_i/\sqrt{n} é dada por:

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

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

Considere

$$L(t) = \log M(t)$$

e note que

$$\begin{aligned} L(0) &= 0 \\ L'(0) &= \frac{M'(0)}{M(0)} \\ &= \mu \\ &= 0 \\ L''(0) &= \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} \\ &= E[X^2] \\ &= 1 \end{aligned}$$

Agora, para demonstrar o teorema, devemos mostrar que $[M(t/\sqrt{n})]^n \rightarrow e^{t^2/2}$ quando $n \rightarrow \infty$, ou, equivalentemente, que $nL(t/\sqrt{n}) \rightarrow t^2/2$ quando $n \rightarrow \infty$. Para mostrarmos isso, observe que:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n \rightarrow \infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} && \text{Pela Regra de L'Hôpital} \\
&= \lim_{n \rightarrow \infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}} \right] \\
&= \lim_{n \rightarrow \infty} \left[\frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}} \right] && \text{novamente pela Regra de L'Hôpital} \\
&= \lim_{n \rightarrow \infty} \left[L'' \left(\frac{t}{\sqrt{n}} \right) \frac{t^2}{2} \right] \\
&= \frac{t^2}{2}
\end{aligned}$$

Assim, o Teorema do Limite Central está demonstrado quando $\mu = 0$ e $\sigma^2 = 1$. O resultado é agora estendido para o caso geral considerando-se as variáveis aleatórias padronizadas $X_i^* = (X_i - \mu)/\sigma$ e aplicando-se o resultado anterior, já que $E[X_i^*] = 0$ e $Var(X_i^*) = 1$.

Observação: Embora o Teorema A.1 afirme apenas que, para cada a ,

$$P \left\{ \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right\} \rightarrow \Phi(a)$$

ou seja, a função de distribuição de S_n^* se aproxima de Φ a função distribuição normal padrão, quando $n \rightarrow \infty$, pode-se, na verdade, mostrar que a convergência é uniforme em a (dizemos que $f_n(a) \rightarrow f(a)$ uniformemente em a se, para cada $\epsilon > 0$, existir um N tal que $|f_n(a) - f(a)| < \epsilon$ para todo a sempre que $n \geq N$).

□

A primeira versão do Teorema do Limite Central foi demonstrada por De-Moivre em 1733 para o caso especial de variáveis aleatórias de Bernoulli com $p = \frac{1}{2}$. O Teorema foi em seguida estendido por Laplace para o caso de p arbitrário (como uma variável aleatória binomial pode ser tratada como sendo a soma de n variáveis aleatórias de Bernoulli independentes e identicamente distribuídas, isso justifica a aproximação normal para a distribuição binomial). Laplace também descobriu a forma mais geral do Teorema do Limite Central dada no Teorema A.1. Sua demonstração, contudo, não era completamente rigoroso e, de fato, não poderia ter sido feito rigorosamente. Uma demonstração verdadeiramente rigorosa para o Teorema do Limite Central foi apresentada primeiramente pelo matemático russo Liapounoff entre 1901 e 1902.

As definições, conceitos e teoremas desta seção foram pesquisados nas referências: [8], [14].