Universidade Federal de Goiás
Regional de Jataí
Coordenação de Matemática
Programa de Mestrado Profissional em Matemática em Rede Nacional

UMA ABORDAGEM DA DISTRIBUIÇÃO NORMAL ATRAVÉS DA RESOLUÇÃO DE UMA SITUAÇÃO PROBLEMA COM A UTILIZAÇÃO DO *SOFTWARE* GEOGEBRA

Paulo Henrique Rodrigues Gonçalves

Jataí

2014

sistema de bibliotecas ufg

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR ELETRONICAMENTE OS TRABALHOS DE CONCLUSÃO DE CURSO NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **Trabalho de Conclusão de Curso de Mestrado Profissional**

2. Identificação do Trabalho

Autor (a):	Paulo Henrique Rodrigues Gonçalves				
E-mail:	paulor22@bol.com.br				
Seu e-mail pode ser disponibilizado na página? [X]Sim [ ] Não					
Vínculo empregatício do autor		Professor de Ensino Básico, Técnico e Tecnológico			
Agência de fomento:		Coordenação de Aperfeiçoamento de Pessoal de Nível Superior		Sigla:	CAPES
País:	Brasil	UF:	GO	CNPJ:	0889834/0001-08
Título:	Uma abordagem da Distribuição Normal através da resolução de uma situação problema com a utilização do software Geogebra				
Palavras-chave:	Educação Estatística, estimativa, Distribuição Normal, Geogebra				
Título em outra língua:	A Normal Distribution approach through the resolution of a problem situation using the Geogebra software.				
Palavras-chave em outra língua:	Statistics Education, estimate, Normal Distribution, Geogebra				
Área de concentração:	Matemática do Ensino Básico				
Data defesa: (dd/mm/aaaa)	24/10/2014				
Programa de Pós-Graduação:	Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT				
Orientador (a):	Prof. Dr. Gecirlei Francisco da Silva				
E-mail:	gecirlei@yahoo.com				
Co-orientador(a):*					
E-mail:					

*Necessita do CPF quando não constar no SisPG

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC do trabalho de conclusão de curso.

O sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses, dissertações ou trabalhos de conclusão de curso, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.

Paulo H. R. Gonçalves
Assinatura do (a) autor (a)

Data: 19 / 11 / 2014

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Paulo Henrique Rodrigues Gonçalves

UMA ABORDAGEM DA DISTRIBUIÇÃO NORMAL ATRAVÉS DA RESOLUÇÃO DE UMA SITUAÇÃO PROBLEMA COM A UTILIZAÇÃO DO *SOFTWARE* GEOGEBRA

Trabalho de Conclusão de Curso apresentado ao Instituto de Matemática e Estatística da Universidade Federal de Goiás, como parte dos requisitos para obtenção do grau de Mestre em Matemática.
Área de Concentração: Matemática do Ensino Básico

Orientador: Prof. Dr. Gecirlei Francisco da Silva

Jataí

2014

Ficha catalográfica elaborada
automaticamente com os dados fornecidos pelo(a) autor(a).

Gonçalves, Paulo Henrique Rodrigues
Uma Abordagem da Distribuição Normal Através da Resolução de uma Situação Problema com a Utilização do Software Geogebra [manuscrito] / Paulo Henrique Rodrigues Gonçalves. - 2014.
102 f.: il.

Orientador: Prof. Dr. Gecirlei Francisco da Silva.
Dissertação (Mestrado) - Universidade Federal de Goiás, Regional Jataí , Jataí, Programa de Pós-Graduação em Matemática (PROFMAT - Profissional), Jataí, 2014.
Bibliografia.
Inclui siglas, fotografias, símbolos, gráfico, tabelas, lista de figuras, lista de tabelas.

1. Educação Estatística. 2. Estimativa. 3. Distribuição Normal. 4. Geogebra. I. Silva, Dr. Gecirlei Francisco da, orient. II. Título.

Paulo Henrique Rodrigues Gonçalves

Uma abordagem da distribuição normal através da resolução de uma solução problema com a utilização do software Geogebra

Trabalho de Conclusão de Curso defendido no Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT/UFG, Pólo Jataí da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática, área de concentração Matemática do Ensino Básico, aprovado no dia 24 de outubro de 2014, pela Banca Examinadora constituída pelos professores:

Gecirlei Francisco da Silva

Prof. Dr. Gecirlei Francisco da Silva
Coordenação de Matemática-CAJ/UFG
Presidente da Banca

Duelci Aparecido de Freitas Vaz

Prof. Dr. Duelci Aparecido de Freitas Vaz
Membro-IFG/Goiânia

Esdras Teixeira Costa

Prof. Dr. Esdras Teixeira Costa
Membro - Coordenação de Matemática-CAJ/UFG

Paulo Henrique Rodrigues Gonçalves graduou-se em Matemática pela então FESURV, hoje Universidade de Rio Verde (UNIRV), em 2002. Cursou Pós-Graduação “Lato Sensu” em Matemática e Estatística na Universidade Federal de Lavras, no período de setembro de 2003 a maio de 2006. Atualmente é professor efetivo do Instituto Federal Goiano Câmpus Rio Verde.

A

Deus, por tudo que tem me proporcionado.

Pâmella, minha esposa amada, por ter me acompanhado nesta jornada, pelos conselhos dados nos momentos de indecisão e fraqueza, pelas sugestões e correções no texto final, e pelo amor, carinho e paciência que teve comigo nas longas horas dedicadas à produção deste trabalho.

Rosalina e Geraldo, meus pais queridos, por terem me dado condições de aprender, pelo amor, dedicação e paciência que tiveram na minha criação e de meus irmãos.

AGRADECIMENTOS

Agradeço primeiramente ao Senhor Deus por permitir que eu tivesse esta oportunidade, à Coordenação de Matemática da Universidade Federal de Goiás – Campus de Jataí pela condução do programa e pela qualidade e simpatia de seus professores. Agradeço também ao professor Dr. Gecirlei Francisco da Silva, orientador, pela atenção e paciência na elaboração deste trabalho, ao amigo Tarcisio Vieira da Silva, pela ideia, companheirismo, apoio e incentivo. E a todas as pessoas, familiares e amigos que de forma direta ou indireta me auxiliaram e incentivaram na construção deste trabalho. Por último, mas não menos importante, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro através da bolsa de estudos.

"A sorte é lançada no colo, mas

a decisão vem do Senhor".

(Provérbios 16:33)

RESUMO

A Distribuição Normal ou curva de Gauss é um modelo matemático usado para descrever um conjunto de valores e uma das suas principais utilidades é fornecer uma estimação sobre a média desse conjunto. A Distribuição Normal é essencial para a inferência estatística, no entanto, raramente é vista no Ensino Médio, sendo mais comum em currículos de cursos de graduação em disciplinas de probabilidade ou estatística. Os PCNs destacam a relevância do ensino de técnicas de inferência sobre uma população através de evidências fornecidas por uma amostra. Este trabalho propõe uma abordagem da Distribuição Normal através da resolução de uma situação problema com auxílio do *software* Geogebra a ser aplicada numa turma de terceiro ano do Ensino Médio. Na resolução dessa situação problema os alunos serão motivados a estimar o peso médio de massa de quarenta quilos de laranjas pera para, através desta informação, estimar o número de laranjas contidas nessa massa e determinar a quantidade de caixas de um litro de suco que podem ser produzidas. A metodologia usada neste trabalho trata-se de pesquisa bibliográfica e contextualização. O trabalho apresentado é de caráter teórico cujo produto final é uma proposta para ensinar a Função Normal articulando com o *software* Geogebra.

Palavras-chave: Educação Estatística, estimativa, Distribuição Normal, Geogebra.

ABSTRACT

The Normal Distribution, also known as the Gaussian bell curve, is a mathematical model used to describe a compound of values and one of its main uses is providing an estimate about the mean of the compound. The Normal Distribution is essential for the statistical inference; yet, it is rarely studied during high school, being more commonly included in high education course outlines more specifically inside disciplines related to probability and statistics. The PCNs (National Curriculum Parameters) highlight the teaching's relevance of inference techniques about a population through the evidences provided by a sample. This study proposes a Normal Distribution approach through the resolution of a problem situation with the aid of the Geogebra software to be applied to a veterans' high school group. By reaching the resolution of this problem, the students will be motivated to estimate the mean weight of the mass of 40 kilograms of oranges to (considering this information) estimate the number of oranges contained in the mass. The students will also be able to determine the amount of one-liter juice boxes that could be produced from it. The methodology used was bibliographic research and contextualization. The presented work is pure theoretical which final product is a purpose to teach the Normal Function, working with the Geogebra software.

Key-words: Statistics Education; Estimate; Normal Distribution, Geogebra.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A Estatística alcançou grande relevância no cotidiano, pois possibilita extrair informações de uma grande quantidade de dados. Os meios de comunicação, por exemplo, utilizam com frequência a linguagem estatística para explorar as informações transmitidas. Observa-se também a abordagem de conceitos mais complexos, como, tamanho da amostra, margem de erro, nível de confiança (CAZORLA, 2004), desconhecidos para a maioria dos telespectadores (AZEVEDO; LIMA; SILVA, 2011), até mesmo para estudantes da educação básica, que deveriam estar familiarizados com tais conceitos.

Além da importância na pesquisa científica, a Estatística pode ser útil para tomadas de decisões racionais do dia-a-dia. Para Batanero (2001 apud CORRÊA, 2012) a Estatística tem um importante papel na sociedade moderna, pois ela fornece ferramentas que permitem analisar as variáveis sob diversas óticas, verificando as possíveis relações existentes através de estudo e experimentos e, consequentemente, dando caminhos para uma possível tomada de decisão de forma coerente e direcionada.

No Brasil, os conceitos básicos de Estatística eram quase ignorados na educação básica, apenas no final da década de 90 passaram a ser discutidos pela comunidade educacional e acadêmica, tendo sido incorporados oficialmente à estrutura da disciplina de Matemática do Ensino Fundamental e Médio com a publicação dos Parâmetros Curriculares Nacionais – PCNs. (LOPES, et al, 2010 apud SAVIAN; JACOBI; ZANINI, 2012)

Os PCNs de Ensino Fundamental e Médio dão destaque à Estatística no Bloco Tratamento da Informação, devido à relevância que esse tema tem alcançado no cotidiano das pessoas e nas várias áreas das ciências. É necessário que os estudantes saiam da educação básica com, pelo menos, noções de como realizar uma coleta, organizar dados e fazer inferências sobre uma população.

As Tecnologias da Informação e Comunicação - TICs, além de facilitar a vida das pessoas, têm influenciado muito o desenvolvimento e o uso diversificado da Estatística, através da utilização de ferramentas para análise de informações. As tarefas de análises de dados, que antes eram feitas manualmente e tomavam muito tempo, hoje, com o uso do computador, tornaram-se quase automáticas. De acordo com Batanero, Estepa e Godino (1991 apud DUARTE, 2010) o desenvolvimento tecnológico, principalmente no campo da informática, impulsionou o avanço da Estatística, pois a análise de dados e a representação gráfica fornecem um auxílio significativo no alcance dos resultados.

A presença das TICs na educação representa a realidade de várias escolas no Brasil, pois as políticas públicas tem buscado atender à necessidade de equipar as escolas com computadores. A Lei do Programa Nacional de Tecnologia Educacional – ProInfo, por exemplo, promove o uso pedagógico das tecnologias de informação e comunicação nas redes públicas de educação básica. O ProInfo funciona em regime de colaboração entre a União, os Estados, o Distrito Federal e os Municípios, mediante adesão. O programa entrega os computadores prontos para o uso às escolas e o estado ou município participante é responsável por garantir a estrutura adequada para receber os laboratórios e capacitar os educadores para uso das máquinas e tecnologias (BRASIL, 2007).

A utilização das TICs pode ser um aliado do professor na proposição de espaços fecundos para a aprendizagem (CYRINO; BALDINI, 2012). É importante que o professor esteja preparado para fazer a utilização das TICs e reconheça sua importância como recurso didático na aprendizagem de conceitos estatísticos. Dessa forma,

> [...] a formação inicial deve somar-se atualizações, sob pena de cristalização profissional. Para conseguir adequar os recursos educativos a estratégias metodológicas inovadoras é necessário saber sua existência, explorá-los e manejá-los com tempo, com disponibilidade e abertura para recorrer às novas formas de ensinar (RICOY; COUTO, 2011, p. 97 apud CYRINO; BALDIDI, 2012).

De acordo com Rycoy e Couto (2011 apud CYRINO; BALDINI, 2012, p. 44) "a falta de atualizações dos professores já inseridos no processo educativo pode causar-lhes constrangimento ao utilizarem-se das TICs em sala de aula, [...]". Portanto, é necessário que sejam criados espaços para que o professor possa aprender a lidar com tais recursos e que se sinta a vontade, e tenha "confiança" para refletir e discutir a sua utilização (BALDINI; CYRINO, 2012 apud CYRINO; BALDINI, 2012).

Dessa forma, é necessário que o professor faça atualizações e conheça os programas para o tratamento das informações a fim de que possa utilizar o laboratório de forma efetiva e segura na aprendizagem e utilização dos conceitos estatísticos.

O uso do laboratório permitirá ao professor trabalhar conteúdos de Estatística, tais como, a Função de Distribuição Normal, que é um modelo matemático utilizado para modelar um conjunto de dados e que é útil para o cálculo de probabilidades bem como para fazer estimações sobre parâmetros de média e desvio padrão.

Assim, a proposta apresentada neste trabalho é relevante, pois desenvolve uma estratégia didática para abordar a Distribuição Normal, que não tem sido contemplada na educação básica, através do uso da tecnologia como facilitadora da aprendizagem.

O objetivo desse trabalho é propor uma abordagem da Distribuição Normal para uma turma de terceiro ano do Ensino Médio, através da resolução de uma situação problema utilizando o *software* Geogebra, com vistas a contribuir com o ensino e aprendizagem da Estatística Inferencial na educação básica.

Como objetivos específicos destacamos:

- Propor o estudo de uma amostragem estatística, através da apresentação de uma situação problema;
- Elaborar um tutorial do programa, ilustrando as aplicações de algumas ferramentas do *software* no cálculo de medidas estatísticas, histogramas e representações da Distribuição Normal;
- Introduzir os primeiros conceitos relativos à Distribuição Normal no Ensino Médio de forma intuitiva, utilizando como recursos didáticos o computador e o *software* Geogebra;
- O reconhecimento pelos alunos da representação de uma Distribuição Normal, como uma curva em forma de sino, assintótica ao eixo horizontal;
- Utilizar o Modelo Normal para descrever a distribuição dos pesos da amostra e da média amostral;
- Perceber os efeitos da variação dos parâmetros sobre a Curva Normal;
- Explorar funções básicas do Geogebra aplicado à Distribuição Normal;
- O entendimento pelos alunos que a distribuição da variável peso é normal;
- Utilizar a Distribuição Normal para calcular probabilidades e para estimar o peso médio da população de quarenta quilos de laranjas;
- A percepção dos alunos de que a média de várias amostras de uma distribuição tem Distribuição Normal;
- Construir um intervalo de confiança para o peso médio populacional com 95% de confiança;
- Calcular a quantidade aproximada de laranjas em quarenta quilos da fruta utilizando o intervalo de confiança.

O texto será composto pelo primeiro capítulo, no qual será exposta uma discussão do problema de estimar usando amostras de uma população, destacando o uso de novas tecnologias na educação matemática. O capítulo será finalizado com a apresentação do *software* Geogebra.

No segundo capítulo será feita uma síntese histórica com os principais colaboradores da distribuição normal, em seguida, uma apresentação sistemática da Distribuição Normal e finalizado com a distribuição $t - Student$.

No terceiro capítulo será apresentada a sequência didática para a resolução do problema da contagem das laranjas contidas em quarenta quilos, no qual também será exposto um tutorial do *software* Geogebra para a resolução do problema.

Nas considerações finais serão discutidos alguns fatores positivos da realização da proposta e as dificuldades que podem ser encontradas durante a execução da mesma. Também serão expostos os resultados esperados, além de sugestões para futuros trabalhos em sala de aula.

1.1 O PROBLEMA

“Estimar consiste em formar um juízo aproximado a um valor, a um cálculo, a uma quantia ou a uma grandeza” (GIONGO et al, 2013, p. 1).

A abordagem da Estatística no Ensino Médio, atualmente, se limita ao estudo de tabelas, gráficos, medidas de posição e de dispersão. De acordo com Giongo (2013) o estudo de estimativas tem sido pouco contemplado nas aulas de Matemática. No entanto, a habilidade de estimar deve começar a ser desenvolvida na educação básica.

Para Girardi e Giongo (2013, p. 166) “desenvolver atividades sobre estimativa nas aulas de Matemática é uma excelente estratégia para despertar o interesse e a agilidade no desenvolvimento de cálculos aproximados, essenciais para diversas situações no cotidiano”.

Os Parâmetros Curriculares Nacionais (PCNs) também dão destaque ao ensino de estimativas ao afirmar que "[...] uma das finalidades atuais do ensino do cálculo consiste em fazer com que os alunos desenvolvam e sistematizem procedimentos de cálculo por estimativa e estratégias de verificação e controle de resultados" (BRASIL, 1997, p. 77).

Situações que envolvem contagem de grandes quantidades podem tornar-se mais simples se o indivíduo possuir conhecimento de técnicas que lhe permitam obter estimativas dessa contagem. Uma forma de obter-se a contagem de um grande conjunto de valores é calcular o quociente da soma de todos os valores pela média do conjunto, como é mostrado através da expressão (1).

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow n = \frac{\sum_{i=1}^{n} x_i}{\bar{x}} \quad (1)$$

Nem sempre a média de um conjunto de dados é conhecida. A Distribuição Normal é um modelo matemático que permite fazer estimações sobre médias populacionais. Tal modelo raramente é abordado no Ensino Médio, e integra, com mais frequência, o currículo de cursos de graduação em disciplinas que envolvem Probabilidade ou Estatística. O ensino da Distribuição Normal é importante, pois ela é essencial para a Inferência Estatística.

De acordo com Lima et al (2006, p. 147) “os tópicos abordados no Ensino Médio deveriam fornecer resposta às seguintes perguntas: [...] Que inferências podem ser feitas a respeito de uma população a partir dos dados relativos a uma amostra desta população?”.

Para Lopes (2010, p. 01) “a Estatística, com os seus conceitos e métodos para coletar, organizar e analisar informações diversas tem-se revelado um poderoso aliado neste desafio que é transformar a informação bruta em dados que permitem ler e compreender uma realidade”.

Os Parâmetros Curriculares Nacionais (PCNs) do Ensino Médio enfatizam essa perspectiva. Na parte 3 – Ciências da Natureza, Matemática e suas Tecnologias, afirma-se que na área da Matemática, há a necessidade de desenvolver nos alunos capacidades de comunicação, de resolução de problemas, tomada de decisões e possibilitar que os alunos tenham condições de realizar inferências (FERREIRA; CARVALHO; BECKER, 2011).

O Ensino de Estatística e Probabilidade encontra-se no bloco de conteúdos denominados “Tratamento das informações”, no qual são definidas as habilidades a serem desenvolvidas pelos alunos:

> As habilidades de descrever e analisar um grande número de dados, realizar inferências e fazer predições com base numa amostra de população, aplicar as ideias de probabilidade e combinatória a fenômenos naturais e do cotidiano são aplicações da Matemática em questões do mundo real que tiveram um crescimento muito grande e se tornaram bastante complexas. Técnicas e raciocínios estatísticos e probabilísticos são, sem dúvida, instrumentos tanto das Ciências da Natureza quanto

> das Ciências Humanas. Isto mostra como será importante uma cuidadosa abordagem dos conteúdos de contagem, estatística e probabilidade no Ensino Médio, ampliando a *interface* entre o aprendizado da Matemática e das demais ciências e áreas. (BRASIL, 2000, p. 44).

O Exame Nacional do Ensino Médio (ENEM), criado em 1998 pelo Governo Federal, destaca, entre as competências avaliadas, o fato de que o aluno tenha compreensão do caráter aleatório e não determinístico de fenômenos naturais e sociais, devendo este, através de instrumentos adequados de medidas, determinar amostras e calcular probabilidades e saber interpretar informações de variáveis presentes em uma distribuição estatística (FERREIRA, CARVALHO, BECKER, 2011).

A aprendizagem de estatística e probabilidades na educação básica é muito importante, pois fazem parte do cotidiano das pessoas. Noticiários, jornais e revistas frequentemente veiculam notícias acompanhadas de conceitos estatísticos, cada vez mais complexos, tais como: tamanho da amostra, margem de erro, nível de confiança dentre outros (CAZORLA, 2004).

O ensino da Estatística, objetivando o entendimento conceitual de suas ferramentas, deve ser feito através de atividades diferenciadas e não apenas por aulas expositivas (GNANADESIKAN et all, 1997 apud AZEVEDO; LIMA; SILVA, 2011).

Este trabalho propõe o ensino contextualizado da Distribuição Normal para uma turma de terceiro ano do Ensino Médio. Nesse sentido, Doane (2004 apud BITTENCOURT; VIALI, 2006) afirma que o ensino de distribuições probabilísticas, como a Distribuição Normal, pode ser mais eficaz com o uso de simulações.

Dessa forma, propor um problema contextualizado é uma excelente alternativa para o ensino da Estatística, pois motivará o estudante a fazer cálculos, elaborar tabelas e analisar dados reais. As Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais afirmam que

> Aprender Matemática de uma forma contextualizada, integrada e relacionada a outros conhecimentos traz em si o desenvolvimento de competências e habilidades que são essencialmente formadoras, à medida que instrumentalizam e estruturam o pensamento do aluno, capacitando-o para compreender e interpretar situações, para se apropriar de linguagens específicas, argumentar, analisar e avaliar, tirar conclusões próprias, tomar decisões, generalizar e para muitas outras ações necessárias à sua formação (BRASIL, 2002, p. 111).

Para Lopes (2008 apud SILVA; BAYER, 2011) "é necessário o desenvolvimento de práticas pedagógicas envolvendo situações em que os estudantes realizem atividades considerando seus contextos e que estes possam observar e construir os eventos possíveis, por meio de experimentação concreta, de coleta e de organização de dados".

Assim, o professor pode desenvolver os conteúdos de estatística, dentre eles, a Distribuição Normal, utilizando metodologias como a resolução de problemas e o uso das Tecnologias da Informação e Comunicação (SILVA; BAYER, 2011) aplicados a uma situação contextualizada com dados reais.

A proposta deste trabalho é descrever a resolução de uma situação problema, na qual os alunos de uma turma de terceiro ano do Ensino Médio serão motivados a usar a Distribuição Normal como um modelo matemático, a fim de encontrar uma estimativa mais precisa do peso médio de uma massa de quarenta quilos de laranja, e, com auxílio desta, encontrar o número aproximado de laranjas contidas nessa massa. A resolução da situação problema será conduzida de modo que propicie o ensino aprendizagem da Distribuição Normal, utilizando como ferramenta os recursos do *software* Geogebra. Nesse processo, o professor exercerá o papel de mediador, indicando caminhos, dando exemplos, orientando na busca de soluções, apontando padrões, etc.

O processo para obter a estimativa do número de laranjas contidas em quarenta quilos da fruta será composto por: seleção aleatória de uma amostra de quarenta laranjas, construção da tabela de frequência, construção do histograma, verificação de padrões, cálculo de medidas estatísticas, aplicação do Modelo Normal e construção do intervalo de confiança para a média. Em todas as etapas a utilização do programa Geogebra será fundamental para análise dos dados.

1.2 INFORMÁTICA NO ENSINO DE ESTATÍSTICA

O uso de computadores na Estatística tem tornado o trabalho de análise de dados muito mais rápido e eficiente. Na área da educação, a introdução da tecnologia da informação possibilita novas formas de ensinar e de aprender (VASCONCELOS; TOGNI, 2011).

Para Lima, O. (2010, p. 2) a informática passa a ser inserida no contexto educacional como um elemento a mais para contribuir na construção do conhecimento, com o objetivo de promover a autonomia humana. O computador dever ser usado na sala de aula como um instrumento potencializador do desenvolvimento humano.

Com relação à Estatística, Duarte (2010, p. 25) considera que "[...], o desenvolvimento das tecnologias da informação tem surgido como facilitador ao tratamento da informação, o que permite cada vez mais trabalhar com grandes conjuntos numéricos de modo rápido, prático e seguro".

Hoje quase todas as escolas dispõem de um laboratório de informática, o que permite ao professor trabalhar a manipulação de dados com os estudantes através de ferramentas estatísticas, em um intervalo de tempo bem menor. "Nesse contexto, as calculadoras e o computador ganham importância como instrumentos que permitem a abordagem de problemas com dados reais ao mesmo tempo em que o aluno pode ter a oportunidade de se familiarizar com as máquinas e os *softwares"* (BRASIL, 2002, p. 127).

Essas tecnologias permitem rapidez e rigor na obtenção de medidas estatísticas e possibilitam aos alunos, em qualquer estudo estatístico, uma concentração nos aspectos mais elaborados do trabalho, como interpretar, organizar, discutir, argumentar, e não sobre os aspectos mais mecânicos associados à sua realização, como a abertura de fórmulas. Além disso, permite aos alunos lidar com os conceitos estatísticos de uma forma mais rica, explorando seu significado, o que representa e como se manipula (CANAVARRO, 2000).

O uso de computadores é defendido por Batanero (2001 apud DUARTE, 2010, p. 26) devido as suas vantagens como, dinamismo, velocidade e crescente quantidade de novos *softwares* que permitem explorar todos os aspectos do processamento de dados, o que possibilita acrescentar novos tópicos ao ensino de estatística.

O computador, no entanto, deve ser usado com cautela, pois possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema (MAGALHÃES; LIMA, 2004). Nesse ponto, percebe-se a importância do professor, que, além de trabalhar os conceitos básicos da Estatística, orientará as atividades feitas mediante a ferramenta do computador.

Segundo Papert (1985 apud FERREIRA et all, 2011) o computador auxilia na construção do conhecimento e possibilita o desenvolvimento de processos mentais que auxiliam a aprendizagem. Ao professor caberá o papel de mediador, contribuindo no direcionamento das atividades de estudo de forma contextualizada para o aluno. Este trabalho propõe o uso do *software* Geogebra como apoio no ensino da Distribuição Normal.

1.3 O *SOFTWARE* GEOGEBRA

O Geogebra[1] é um *software* de matemática dinâmica que reúne Geometria, Álgebra, Cálculo e Estatística. Foi desenvolvido pelo professor Markus Hohenwarter, na universidade americana Florida Atlantic University, com a ajuda de diversos colaboradores, no intuito de aprender e ensinar matemática nas escolas.

Neste trabalho, optou-se por este *software* por ser um aplicativo de domínio público. Através do Geogebra é possível criar tabelas, gráficos, histogramas, calcular medidas estatísticas, intervalos de confiança, dentre outras funções. Sua *interface* simples não exige do usuário domínio sobre o programa, o conhecimento dos conceitos estatísticos a serem utilizados no desenvolvimento das atividades é suficiente para manipulá-lo.

1.3.1 *Interface*

O Geogebra fornece três diferentes vistas dos objetos matemáticos: *a Janela de Visualização* ou *Zona Gráfica*, *a Janela de Álgebra* ou *Zona Algébrica*, ou numérica, e a *Planillha* ou *Folha de Cálculo*. Elas permitem mostrar os objetos matemáticos em três diferentes representações: graficamente (e.g., pontos, gráficos de funções), algebricamente (e.g., coordenadas de pontos, equações) e nas células da folha de cálculo. Assim, todas as representações do mesmo objeto estão ligadas dinamicamente e adaptam-se automaticamente às mudanças realizadas em qualquer uma delas, independentemente da forma como esses objetos foram inicialmente criados (HOHENWARTER; HOHENWARTER, J., 2009, p. 6).

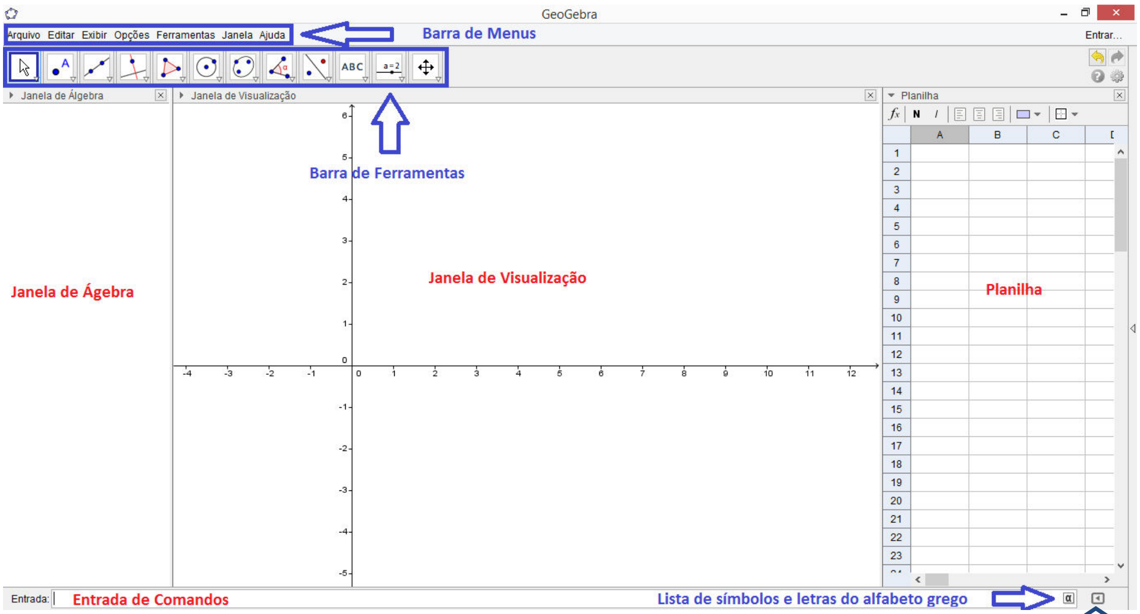
A *interface* do software é composta pelos seguintes itens:

- *Barra de ferramentas*: apresenta os principais recursos para as construções geométricas;
- *Barra de menus*: apresenta as opções de execução do programa;
- *Entrada de comandos*: este campo permite executar praticamente todas as funções do *software* através de comando digitado;

[1] A versão 4.4.43.0 foi utilizada para o desenvolvimento deste trabalho. A versão atual 5.0 está disponível para download em www.geogebra.org

- *Lista de símbolos e letras do alfabeto grego*: apresenta alguns símbolos matemáticos e as letras do alfabeto grego usadas na escrita matemática;
- *Lista geral de comandos*: este campo apresenta a lista de todas as funções do pacote básico do programa. Este campo também apresenta funções criadas pelo usuário.

Figura 1 - *Interface* do Geogebra



Fonte: Geogebra

2 DISTRIBUIÇÃO NORMAL

2.1 HISTÓRICO

A Distribuição Normal é um modelo matemático que descreve fenômenos aleatórios e tem uma importância muito grande para a Inferência Estatística. Esse modelo serve de base na construção de intervalos de confiança e também é útil para calcular probabilidades.

A origem da Distribuição Normal está associada à observação do erro de mensuração por volta do século XVIII. Existem evidências de que tais erros de mensuração eram observados principalmente pelos astrônomos na tentativa de estimar as órbitas dos corpos celestes. Ao fazer essas estimativas havia necessariamente um fator de erro de observação. Esse fator ficou conhecido como lei do erro (STEWART apud DUARTE, 2010, p. 15).

A introdução da Distribuição Normal foi feita inicialmente pelo matemático francês Abraham de Moivre (1667 – 1754) em um artigo que foi reimpresso na segunda edição do seu livro "A doutrina do acaso" de 1738 (BITTENCOURT; VIALI, 2006). Abraham de Moivre foi um exilado político, que foi para Inglaterra onde conheceu Newton e Halley e em 1697 foi eleito membro da Real Academia de Ciências (DUARTE, 2010, p. 15).

De Moivre procurou dar sentido algébrico à teoria das probabilidades, sendo o primeiro matemático a trabalhar com a expressão da "lei do erro" em 1733 (DUARTE, 2010).

Em 1812 o também matemático francês Pierre Simon de Laplace estendeu os trabalhos de Moivre em seu livro "Teoria Analítica da Probabilidade". O resultado de seu trabalho é conhecido hoje como Teorema de Moivre-Laplace. Laplace e também mostrou que mesmo uma distribuição não sendo normal, a média de repetidas amostras dessa distribuição é a aproximadamente normal e que quanto maior for o tamanho da amostra melhor será essa aproximação (BITTENCOURT; VIALI, 2006).

A primeira aplicação da distribuição foi feita pelo belga Adolph Quetelet (1796 – 1874). Ele coletou dados sobre medidas do peito de soldados escoceses e da altura de soldados franceses e verificou que elas podiam ser modeladas pela distribuição (BITTENCOURT; VIALI, 2006).

O matemático alemão Karl F. Gauss (1777 – 1855) fez contribuições importantes e realizou estudos sobre a lei dos Erros. Suas contribuições foram tão importantes que hoje é comum encontrar em livros e artigos do ramo, o enunciado com o título de Distribuição Gaussiana (DUARTE, 2010). Gauss ocupava o cargo de diretor no observatório de Gottingëm, posto que conservou por cerca de quarenta anos (BOYER, 1974, p. 374).

Gauss e o matemático alemão Robert Adrian (1775 – 1843) desenvolveram de forma independente em 1808 a "equação da distribuição" e mostraram que modelava muito bem os erros de observações astronômicas (BITTENCOURT, VIALI, 2006).

No modelo apresentado por Gauss, cada medida ou mensuração é exatamente igual à medida ideal mais o erro de medição. Assim, os resultados obtidos de sucessivas medidas, que se diferem uma da outra, se distribuem em torno de um valor ideal, ou valor nulo (FREEDEMAN apud DUARTE, 2010).

A Distribuição Normal é muito importante pelo fato de que muitos fenômenos naturais apresentam uma distribuição normal ou aproximadamente normal. Além disso, as médias de amostras extraídas de uma distribuição qualquer, mesmo não sendo normal, tendem a apresentar comportamento normal à medida que o tamanho da amostra aumenta (BITTENCOURT; VIALI, 2006).

A denominação de Distribuição Normal ou Curva Normal foi dada de forma independente pelo filósofo americano Charles S. Peirce (1839 – 1914), pelo antropólogo e geneticista britânico Francis Galton (1822 – 1911) e pelo economista alemão Wilhelm Lexis (1837 – 1914) por volta de 1875. Já a terminologia de Curva em Forma de Sino foi cunhada em 1872 pelo francês Esprit Pascal Jouffret (BITTENCOURT e VIALI, 2006).

2.2 O MODELO NORMAL

"Uma distribuição estatística é uma função que descreve o comportamento de uma variável aleatória" (NETO, SCARMINIO e BRUNS, 2003, p. 27). Variável aleatória é uma função que associa um número a cada elemento de um espaço amostral.

A função densidade de probabilidade é uma função sempre não negativa e construída de tal forma que área sob seu gráfico e eixo horizontal é igual a 1.

Assim, uma função matemática é uma função densidade de probabilidade se satisfaz duas condições:

$$i)\ f(x) \geq 0, \text{ para todo } x \in \mathbb{R}; \qquad (2)$$

ii) A área definida por $f(x)$ é igual a 1. (3)

Com o auxílio do cálculo diferencial e integral, podemos caracterizar a condição ii) por

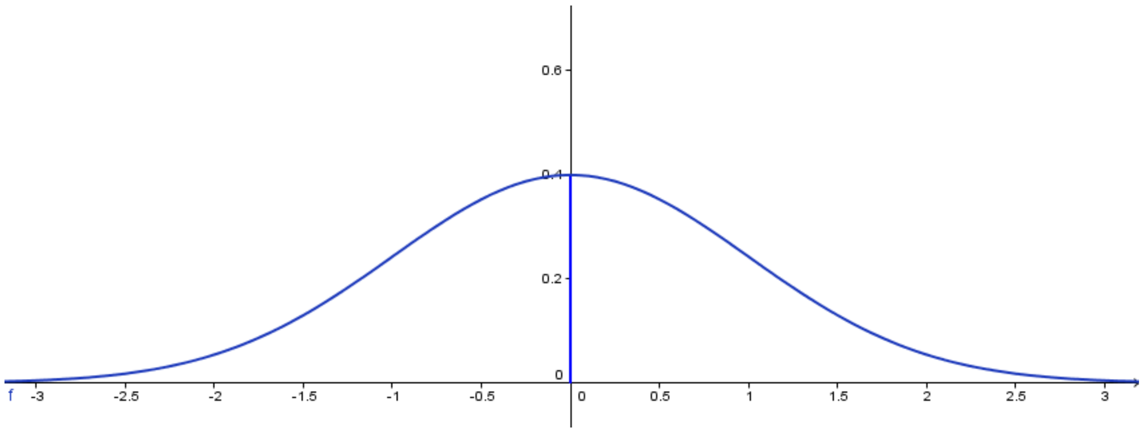
$$\int_{-\infty}^{\infty} f(x)dx = 1. \quad (4)$$

Dizemos que uma variável aleatória contínua X tem distribuição normal com parâmetros μ e σ, se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{, para } x \in R, -\infty < \mu < \infty \text{ e } \sigma > 0 \quad (5)$$

Sua representação gráfica pode ser vista no Gráfico 1.

Gráfico 1 - Gráfico da função de densidade de probabilidade $\boldsymbol{N(0,1)}$



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

A Distribuição Normal é uma distribuição contínua, ou seja, uma distribuição em que a variável pode assumir qualquer valor dentro de um intervalo previamente definido. Para uma variável normalmente distribuída, o intervalo é $(-\infty, +\infty)$, o que significa que ela pode assumir, pelo menos a princípio, qualquer valor real (NETO; SCARMINIO; BRUNS, 2003).

A função $f(x)$ é uma função exponencial de uma variável real, que possui dois parâmetros μ (média) e σ (desvio padrão). Para indicar que X tem distribuição normal com parâmetros μ e σ, usa-se a notação $X \sim N(\mu, \sigma^2)$.

A densidade normal possui as propriedades:

$i)$ $f(x)$ é simétrica em relação à μ, ou seja, ao centro;

ii) $f(x) \to 0$ quando $x \to \pm\infty$;

iii) o valor máximo de $f(x)$ é obtido quando $x = \mu$;

iv) tem dois pontos de inflexões em $x = \mu \pm \sigma$;

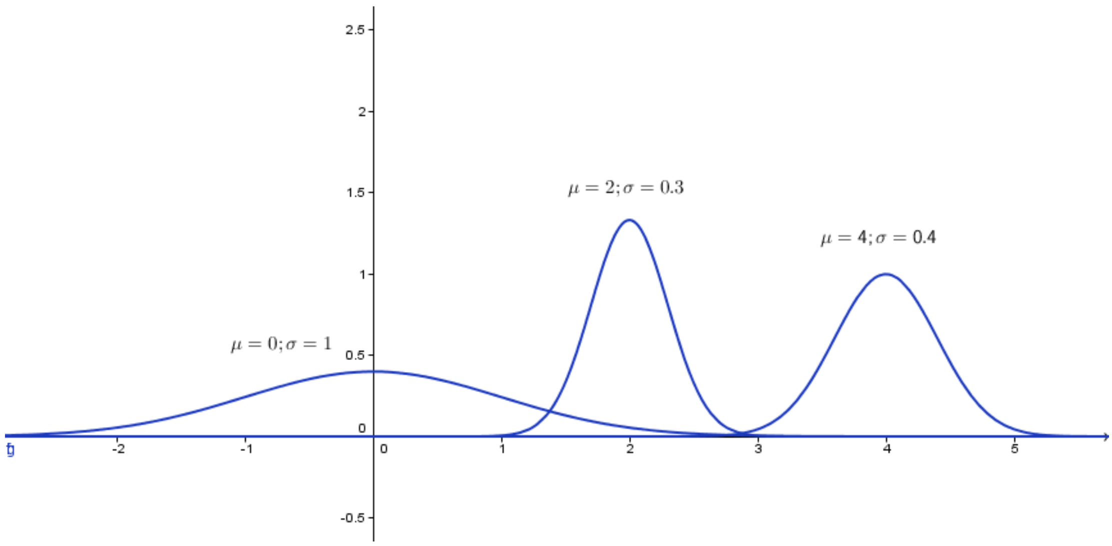
v) A área compreendida entre a curva até o eixo das abscissas é igual a 1.

A Curva Normal é assintótica em relação ao eixo das abscissas devido às propriedades ii) e iii).

O formato da Curva Normal dependerá dos parâmetros μ (média) e σ (desvio padrão). Ao variar esses dois parâmetros temos movimentos de translação e achatamento no gráfico. Mesmo nessas variações, todas as representações normais são simétricas, têm formato semelhante ao contorno de um sino e a área total sob a curva é igual a 1 (DUARTE, 2010).

O Gráfico 2 mostra três variações da Curva Normal de acordo com os respectivos parâmetros μ e σ.

Gráfico 2 - Diferentes curvas normais



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Podemos obsevar pelas representações gráficas que o valor da média indica o centro da distribuição e o desvio padrão mede a dispersão do conjunto, indicando a variabilidade em relação ao centro, ou seja, quanto maior o desvio padrão mais achatado será o gráfico.

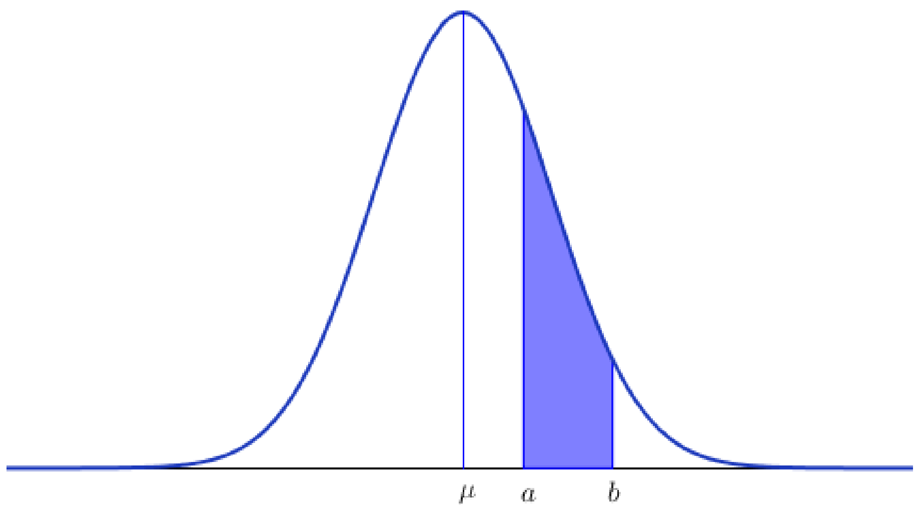
2.3 CÁLCULOS DE PROBABILIDADES

Para calcular probabilidades com o Modelo Normal, devemos resolver a integral da função densidade de probabilidade no intervalo de interesse. Para uma distribuição contínua, portanto, não faz diferença se o intervalo de que estamos falando é aberto ou fechado. A probabilidade de que $a \leq x \leq b$ é igual à probabilidade de que $a < x < b$, isto é

$$P(a < x < b) = P(a \leq x \leq b) = \int_a^b f(x)dx \tag{6}$$

Assim a probabilidade de que um valor da variável aleatória de densidade de probabilidade $f(x)$ seja observado no intervalo $[a, b]$ é obtida pela área sob a curva f e o eixo das abscissas, em relação a esse intervalo, conforme mostrado no Gráfico 3.

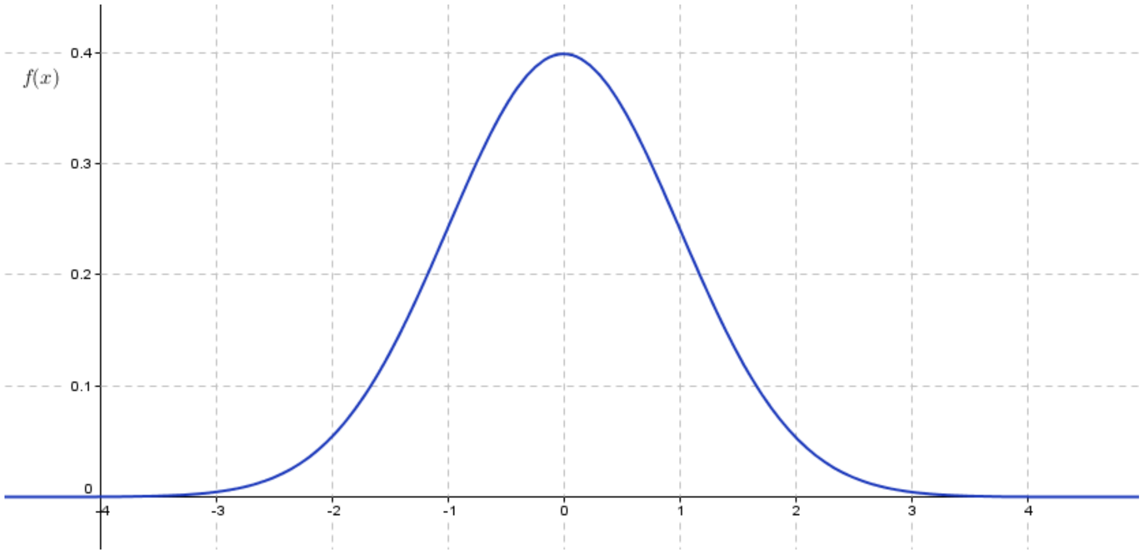
Gráfico 3 - Área sob a Curva Normal entre os pontos $\boldsymbol{a}$ e $\boldsymbol{b}$



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

No Modelo Normal, aproximadamente 68% da área total da curva está localizada entre $\mu \pm \sigma$, aproximadamente 95% da área total está localizada entre $\mu \pm 2\sigma$ e aproximadamente 99,7% da área total da curva está localizada entre $\mu \pm 3\sigma$. Podemos observar esses fatos no gráfico da distribuição de frequência de uma variável aleatória $X \sim N(0,1)$, conforme Gráfico 4.

Gráfico 4 - Curva Normal de uma variável aleatória $\boldsymbol{X}\sim\boldsymbol{N}(\mathbf{0},\mathbf{1})$



Fonte: Construído pelo autor com o auxílio do *software* Geogebra.

Podemos notar pelo gráfico que a maior parte da área sob uma gaussiana está contida no intervalo definido por um desvio padrão em torno da média, e praticamente toda ela está situada entre $\mu + 3\sigma$ e $\mu - 3\sigma$. Para obter os valores numéricos correspondentes a esses fatos, integramos a expressão $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ entre os limites apropriados:

$$P(\mu - \sigma < x < \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0{,}6826 = 68{,}26\% \qquad (7)$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x) = 0{,}9973 = 99{,}73\% \qquad (8)$$

O cálculo dessas integrais é feito pelo método de Simpson, pois a função de densidade normal não possui integral algébrica. Mas na prática não precisamos calcular integral nenhuma, pois podemos utilizar valores tabelados que permitem encontrar facilmente os valores de probabilidades desejados.

A tabela possui valores das integrais para vários intervalos de uma variável z$\sim N(0{,}1)$. Apesar de esses valores corresponderem à distribuição padrão, com média zero e desvio padrão um, esses valores podem ser usados para fazermos inferências a respeito de qualquer Distribuição Normal. Para isso devemos padronizar a variável aleatória x.

2.3.1 Padronização de uma variável aleatória

Padronizar uma variável aleatória x de média μ e desvio padrão σ consiste em construir a partir dela uma nova variável aleatória z, cujos valores são obtidos subtraindo-se de cada valor de x a média populacional e dividindo-se o resultado pelo desvio padrão (NETO; SCARMINIO; BRUNS, 2003):

$$z = \frac{x-\mu}{\sigma} \tag{9}$$

Assim, para determinar a probabilidade de $x \in [a, b]$, temos

$$\begin{aligned} P(a \leq x \leq b) &= P(a - \mu \leq x - \mu \leq b - \mu) \\ &= P(\frac{a-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}) \\ &= P(\frac{a-\mu}{\sigma} \leq z \leq \frac{b-\mu}{\sigma}) \end{aligned} \tag{10}$$

Portanto, quaisquer que sejam os valores de μ e σ, utilizamos a normal padronizada para obter probabilidades com a Distribuição Normal.

Os valores para $P(0 \leq z \leq z_c),\ z \geq 0$ são apresentado na Tabela 1.

Tabela 1 - Área ou probabilidade para Distribuição Normal padrão

Distribuição Normal: Valores de p tais que $P(0 \leq Z \leq z_c) = p$

Segunda decimal de z_c

Parte inteira e primeira decimal de z_c

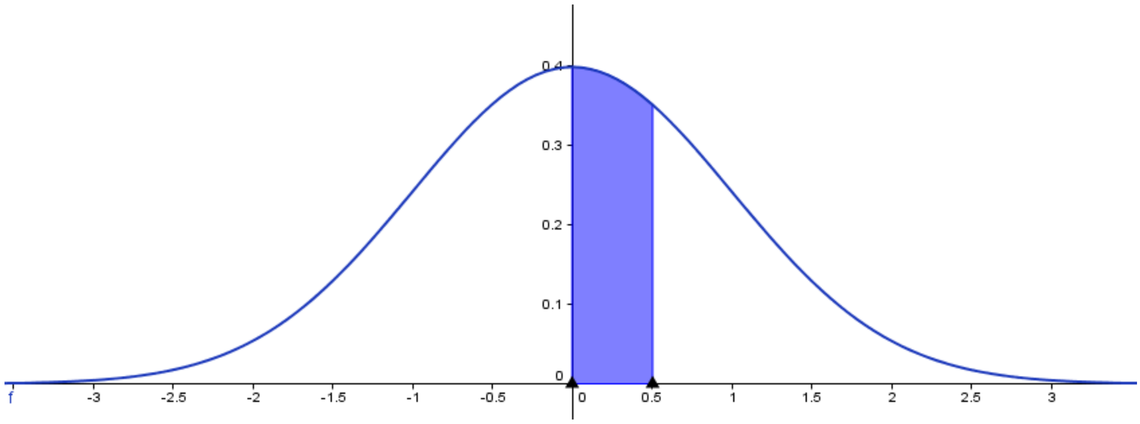
	0	1	2	3	4	5	6	7	8	9	
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359	**0,0**
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753	**0,1**
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141	**0,2**
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517	**0,3**
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879	**0,4**
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224	**0,5**
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549	**0,6**
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852	**0,7**
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133	**0,8**
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389	**0,9**
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621	**1,0**
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830	**1,1**
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015	**1,2**
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177	**1,3**
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319	**1,4**
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441	**1,5**
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545	**1,6**
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633	**1,7**
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706	**1,8**
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767	**1,9**
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817	**2,0**
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857	**2,1**
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890	**2,2**
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916	**2,3**
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936	**2,4**
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952	**2,5**
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964	**2,6**
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974	**2,7**
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981	**2,8**
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986	**2,9**
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990	**3,0**
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993	**3,1**
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995	**3,2**
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997	**3,3**
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998	**3,4**
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	**3,5**
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	**3,6**
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	**3,7**
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	**3,8**
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	**3,9**

Fonte: Magalhães e Lima (2004, p. 355)

A escala padronizada z permite associar os valores da variável x com os valores da escala z, de acordo com os respectivos valores de μ e σ. Neste caso, não será necessário construir tabelas para cada função normal $N(\mu, \sigma^2)$ (DUARTE, 2010).

Na Tabela 1 o valor de z indica a intensidade de afastamento da variável x até a média μ, em relação ao desvio padrão, ou seja, é o comprimento relativo em desvios padrões, entre a média μ até o ponto x (DUARTE, 2010). O Gráfico 5 apresenta o gráfico da distribuição normal padronizada.

Gráfico 5 - Distribuição Normal padronizada



Fonte: Construído pelo autor com o auxílio do *software* Geogebra.

A Distribuição Normal sempre é simétrica em relação à média μ, e devido a esse fato basta obtermos as áreas de apenas um dos lados. Sendo assim, a tabela de valores de z apresenta os valores das áreas compreendidas entre a média e um ponto x à direita da média.

Um exemplo de utilização da tabela:

Seja um conjunto de dados onde estamos estudando o comportamento de uma variável $x \sim N(3, 5^2)$, ou seja, x tem distribuição normal com parâmetro de média igual a 3 e parâmetro de variância igual a 5^2.

Vamos calcular a probabilidade de x variar entre a média até o ponto $x = 10$. Temos:

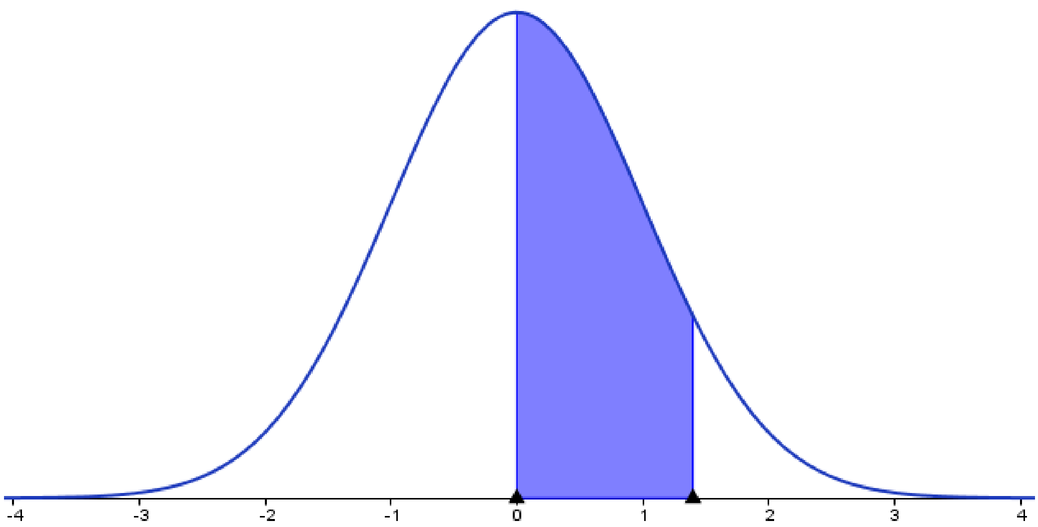
$$P(3 \leq x \leq 10) = P\left(\frac{3-3}{5} \leq \frac{x-3}{5} \leq \frac{10-3}{5}\right) = P(0 \leq z \leq 1{,}4) \qquad (11)$$

A primeira coluna da tabela apresenta a parte inteira e a primeira decimal de z e a primeira linha apresenta a segunda decimal de z. Para saber a probabilidade correspondente a certo valor de z temos que procurar na tabela o valor localizado na interseção da linha e da coluna apropriadas. O valor correspondente a $z = 1{,}4$ está na interseção da linha referente à $z = 1{,}4$ com a coluna encabeçada por 0, que é a segunda decimal de z. Assim temos:

$$P(3 \leq x \leq 10) = P(0 \leq z \leq 1{,}4) = 0{,}4192 \text{ ou } 41{,}92\% \text{ de chance} \qquad (12)$$

que pode ser vista no Gráfico 6.

Gráfico 6 - Área sob a Curva Normal padrão no intervalo $[\mathbf{0}; \mathbf{1,4}]$



Fonte: Construído pelo autor com o auxílio do *software* Geogebra.

2.4 PARÂMETRO, ESTIMADOR E ESTIMATIVA

2.4.1 Parâmetros

Parâmetros são as quantidades desconhecidas da população sobre as quais temos interesse. São representados por letras gregas como μ, σ e θ.

2.4.2 Estimador

É uma função ou combinação dos elementos da amostra, construída com a finalidade de representar, ou estimar, um parâmetro de interesse na população. São denotados pelos símbolos: $\bar{x}, s$ e p.

2.4.3 Estimativa

Estimativas ou estimativas pontuais são os valores numéricos assumidos pelos estimadores.

As notações utilizadas para a média e desvio padrão de uma população são, respectivamente, μ e σ.

Os melhores estimadores para a média e o desvio padrão populacionais são, respectivamente, a média amostral e o desvio padrão amostral com uma pequena alteração na sua expressão. Assim temos:

Estimador para a média populacional μ

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{13}$$

Estimador para o desvio padrão populacional σ

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} \tag{14}$$

A expressão dentro do radical é chamada de variância. O denominador dessa expressão é $n - 1$ e não n. Isso é necessário para que este estimador não seja viciado, isto é, para que seu valor esperado coincida com o parâmetro de interesse.

Esses estimadores têm as propriedades de não serem viciados e terem consistência.

2.5 TEOREMA CENTRAL DO LIMITE

Se a flutuação total numa certa variável aleatória for resultado da soma das flutuações de muitas variáveis independentes e de importância mais ou menos igual, a sua distribuição tenderá para a normalidade, não importando qual seja a natureza das distribuições das variáveis individuais (NETO; SCARMINIO; BRUNS, 2003).

Em símbolos, temos:

$$X_1 + X_2 + X_3 + \cdots + X_n = \sum X_i \sim Normal(\mu_{X,} \sigma_X) \tag{15}$$

Este resultado pode ser utilizado para compreender o comportamento da média amostral (BITTENCOUR; VIALI, 2006). Como a média amostral é uma função de variáveis aleatórias independentes e de igual importância o teorema garante que a média se comporta segundo o Modelo Normal. Utilizando as propriedades da Esperança e da Variância de variáveis aleatórias:

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n} \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{16}$$

Quanto maior for o tamanho da amostra, melhor a aproximação para a distribuição normal. Estudos envolvendo simulações mostram que, em muitos casos, amostras com números de elementos ao redor de 30 fornecem aproximações bastante boas para as aplicações práticas (MAGALHÃES; LIMA, 2004).

2.6 ESTIMAÇÃO POR INTERVALO DE CONFIANÇA

Os estimadores usados para a média e o desvio padrão são estimadores pontuais, pois fornecem como estimativa um único valor numérico para esses parâmetros. Como esses estimadores são variáveis aleatórias, eles têm distribuição de probabilidade. E vimos que a média amostral tem distribuição normal. Devido a este fato, podemos apresentar uma estimativa mais informativa para o parâmetro de interesse que inclua uma medida de precisão do valor obtido. Esse método de estimação é chamado de intervalo de confiança, ele consiste em incorporar à estimativa pontual do parâmetro, informações a respeito de sua variabilidade. Esses intervalos são obtidos através da distribuição amostral de seus estimadores (MAGALHÃES; LIMA, 2004).

2.6.1 Intervalo de confiança para a média de uma população normal com desvio padrão conhecido

Consideremos amostras de n elementos, extraídas aleatoriamente de uma população norma de média (μ) e desvio padrão (σ). Vimos que a média amostral tem distribuição normal com a mesma média μ, mas com desvio padrão igual a $\sigma/\sqrt{n}$.

Padronizando, temos

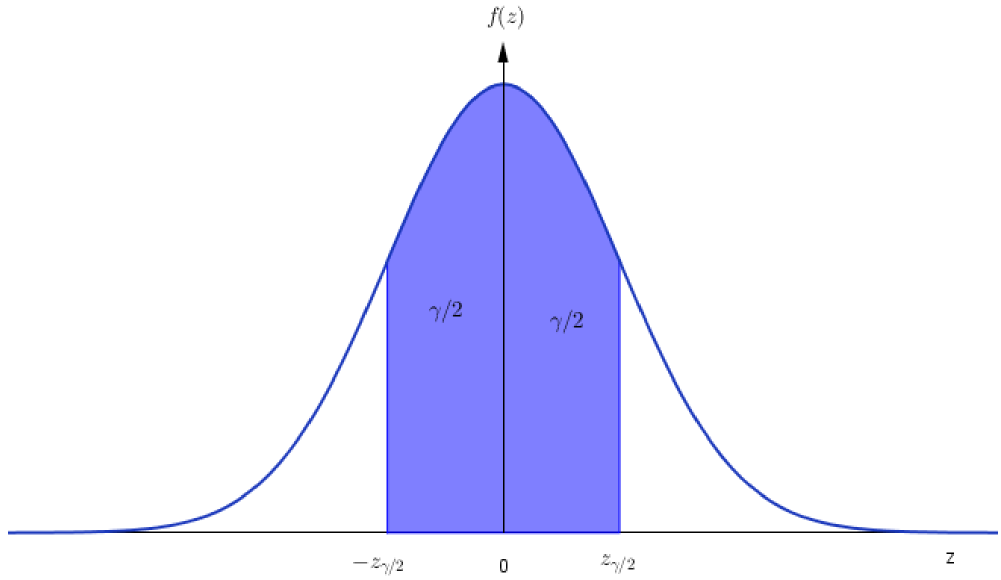
$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \tag{17}$$

Fixando um valor γ tal que $0 < \gamma < 1$, podemos encontrar um valor $z_{\gamma/2}$ tal que

$$P\left(-z_{\gamma/2} < z < z_{\gamma/2}\right) = \gamma \tag{18}$$

Podemos observar que o índice de $z_{\gamma/2}$ apresenta o valor de γ dividido por 2. Isso se deve ao fato da área γ está distribuída igualmente em torno de 0 (Gráfico 7).

Gráfico 7 - Área sobre a Curva Normal padrão no intervalo $[-\boldsymbol{z_{\gamma/2}}, \boldsymbol{z_{\gamma/2}}]$



Fonte: Construído pelo autor com o auxílio do *software* Geogebra

Para obtermos o valor de $z_{\gamma/2}$ através da Tabela 1, devemos fazer o caminho inverso feito para encontrar as probabilidades. Primeiramente localizamos o valor de $\gamma/2$ no corpo da tabela e assim obtém-se o valor $z_{\gamma/2}$ nas margens correspondentes. Assim, temos o intervalo

$$-z_{\gamma/2} < z < z_{\gamma/2} \Rightarrow -z_{\gamma/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\gamma/2} \tag{19}$$

Remanejando os símbolos de modo a isolar a média populacional, μ, chegamos a duas desigualdades,

$$\mu < \bar{x} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}} \qquad \text{e} \qquad \bar{x} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}} < \mu, \tag{20}$$

que podem se combinadas em uma só:

$$\bar{x} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}} \tag{21}$$

Assim, o intervalo de confiança para μ, com coeficiente de confiança γ, é dado por

$$IC(\mu,\gamma) = \left[\bar{x} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right] \tag{22}$$

Devemos ter cuidado ao interpretar um intervalo de confiança. A expressão $IC(\mu,\gamma)$ envolve média amostral $\bar{x}$ que é uma variável aleatória e, portanto, o intervalo obtido também é aleatório. A probabilidade de que ele contenha o verdadeiro valor da média populacional μ é dada por γ. Desta forma, uma interpretação adequada seria: se obtivermos várias amostras de mesmo tamanho e, para cada uma delas, calcularmos os correspondentes intervalos de confiança com coeficiente de confiança γ, esperamos que a proporção de intervalos que contenham o valor de μ seja igual a γ (MAGALHÃES; LIMA, 2004).

2.6.2 Intervalo de confiança para a média populacional com desvio padrão desconhecido

Em geral, não temos informação sobre o desvio padrão σ da variável aleatória em estudo. Neste caso, precisamos contornar essa situação, ou seja, o desvio padrão populacional dever ser estimado.

Mantendo a suposição de que a variável aleatória de interesse tem distribuição normal. Consideremos amostras de n elementos, extraídas aleatoriamente de uma população norma de média μ e desvio padrão σ. Utilizando o melhor estimador que conhecemos para σ que é o desvio padrão amostral:

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} \tag{23}$$

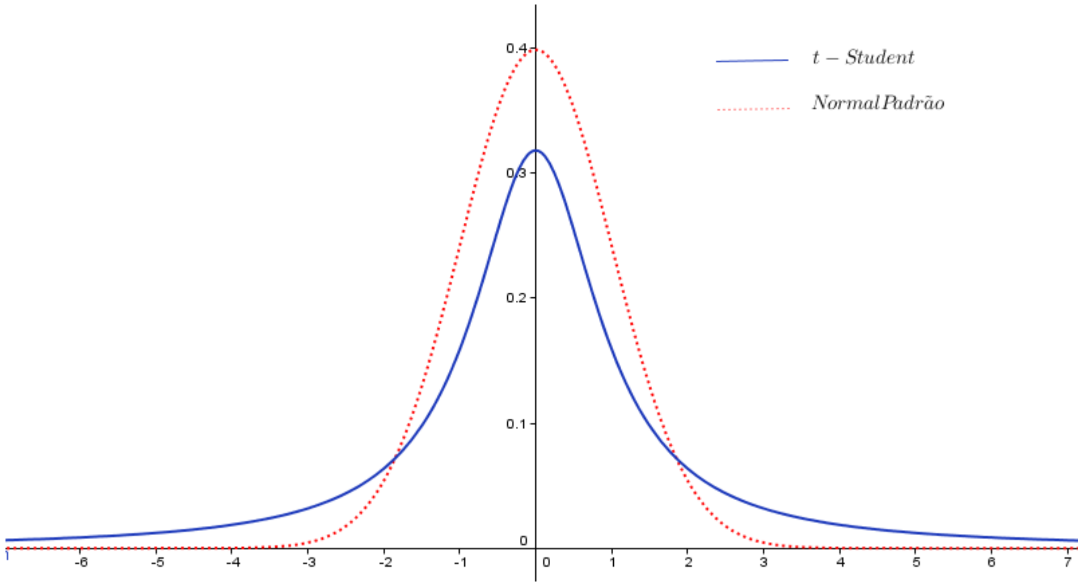
Padronizando, temos uma nova variável aleatória t

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \tag{24}$$

Apesar de $\bar{x}$ ter distribuição normal, o denominador envolve a variável aleatória s, isso fará com que a função densidade de t seja diferente da distribuição normal. Essa nova distribuição é denominada $t\ de\ Student$ e seu parâmetro tem o nome de graus de liberdade, que corresponde ao número de elementos da amostra menos 1 (MAGALHÃES; LIMA, 2004). Usa-se a notação $t_{(n-1)}$ e suas probabilidades são obtidas pela Tabela 2.

A distribuição $t\ de\ Student$ também tem densidade em forma de sino, mas suas caudas têm mais massa que a $N(0,1)$ conforme Gráfico 8.

Gráfico 8 - Gráfico da Distribuição Normal e $\boldsymbol{t - Student}$



Fonte: Construído pelo autor com o auxílio do *software* Geogebra

Quando o tamanho da amostra aumenta, a densidade $t - Student$ converge para normal padrão. Devido a esse fato, as tabelas construídas se limitam a valores de graus de liberdade menores ou iguais a 120. Para os graus superiores a 120, as probabilidades são obtidas da tabela da Distribuição Normal e representadas por ∞ na Tabela 2 (MAGALHÃES; LIMA, 2004). Mas na prática, só se costuma usar a distribuição t quando o número de graus de liberdade na estimativa do desvio padrão é inferior a 30. Para amostras maiores, a equação (22) é satisfatória (NETO; SCARMINIO; BRUNS, 2003).

A tabela $t - Student$ é apresentada na Tabela 2.

Tabela 2 - Área ou probabilidade para distribuição $\boldsymbol{t - Student}$

Distribuição t-Student : Valores t_c tais que $P(-t_c \leq t \leq t_c) = 1 - p$

Graus de liberdade	p->90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%	
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,656	318,289	636,578	1
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,328	31,600	2
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214	12,924	3
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,999	3,747	4,604	7,173	8,610	4
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,757	3,365	4,032	5,894	6,869	5
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	8
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	4,025	4,437	11
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,249	2,602	2,947	3,733	4,073	15
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,235	2,583	2,921	3,686	4,015	16
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	17
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	18
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,861	3,579	3,883	19
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	20
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	21
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	22
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	23
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	24
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,167	2,485	2,787	3,450	3,725	25
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	26
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,689	27
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	28
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,660	29
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	30
35	0,127	0,255	0,388	0,529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	35
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,123	2,423	2,704	3,307	3,551	40
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,109	2,403	2,678	3,261	3,496	50
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,460	60
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,076	2,358	2,617	3,160	3,373	120
∞	0,126	0,253	0,385	0,524	0,675	0,842	1,036	1,282	1,645	1,960	2,054	2,327	2,576	3,091	3,291	∞

Fonte: Magalhães e Lima (2004, p. 356)

Quando a variância é desconhecida, construímos intervalos de confiança para a média populacional utilizando o modelo $t - Student$.

Suponha uma amostra aleatória $X_1, X_2, \dots, X_n$, obtida de uma população com distribuição normal com média e variância desconhecidas, temos que

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{(n-1)} \tag{24}$$

Fixando o coeficiente de confiança γ, $0 < \gamma < 1$, e utilizando a Tabela 2 da distribuição $t - Student$ com $n - 1$ graus de liberdade, obtemos o valor de $t_{\gamma/2}$ tal que

$$P\left(-t_{\gamma/2} < \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{\gamma/2}\right) = \gamma \tag{25}$$

Fazendo as devidas manipulações para isolar μ, teremos o intervalo com coeficiente de confiança γ para μ com desvio padrão desconhecido.

$$IC(\mu, \gamma) = \left[\bar{x} - t_{\gamma/2} \frac{s}{\sqrt{n}}; \bar{x} + t_{\gamma/2} \frac{s}{\sqrt{n}}\right] \tag{26}$$

Exemplo:

Admitindo que o peso da laranja siga o Modelo Normal, 10 laranjas foram selecionadas e tiveram seu peso medido com os seguintes valores: 164,1; 173,3; 182,7; 155,3; 159,5; 155,4; 140,8; 170,3; 165,9; 174,8 (em gramas). Vamos determinar o intervalo de confiança para μ com coeficiente de confiança $\gamma = 95\%$.

Como σ é desconhecido, usaremos o estimador $s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}}$, e sendo o número de valores igual a dez, temos $n - 1 = 9$ graus de liberdade. Assim,

$$T = \frac{\bar{x} - \mu}{s/\sqrt{10}} \sim t_{(9)} \tag{27}$$

Calculando $\bar{x}$ e s, temos

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 164{,}21g \qquad e \qquad s = \sqrt{\frac{\sum_{i=1}^{10} x_i^2 - 10\bar{x}^2}{9}} = 12{,}04g \tag{28}$$

Assim,

$$I(\mu, 95\%) = \left[164{,}21 - t_{\gamma/2}\frac{12{,}04}{\sqrt{10}}; 164{,}21 + t_{\gamma/2}\frac{12{,}04}{\sqrt{10}}\right] \tag{29}$$

A primeira coluna da Tabela 2 apresenta os graus de liberdade $(n - 1)$ e a primeira linha apresenta as probabilidades complementares $(1 - p)$ com que devem ser obtidos os valores de t. Como $1 - p = 5\%$ e $n - 1 = 9$, o valor de t obtido na interseção da linha referente a 9 graus de liberdade com a coluna correspondente a 5% de chance é $t = 2{,}262$. Assim, temos:

$$I(\mu, 95\%) = \left[164{,}21 - 2{,}262\frac{12{,}04}{\sqrt{10}}; 164{,}21 + 2{,}262\frac{12{,}04}{\sqrt{10}}\right] \tag{30}$$

Logo,

$$I(\mu, 95\%) = [155{,}6g; 172{,}82g] \tag{31}$$

3 PROPOSTA DE ATIVIDADE PRÁTICA A SER DESENVOLVIDA

3.1 SITUAÇÃO PROBLEMA

A situação problema proposta é a seguinte:

Um produtor de suco de laranja recebe uma carga com 3 toneladas da fruta. Sabe-se que para a produção de cada caixa de um litro de suco são consumidas 5 laranjas. Diante disso, o produtor gostaria de saber quantas caixas de suco poderiam ser produzidas, aproximadamente, com as 3 toneladas de laranja.

3.1.1 Análise do problema

O problema em questão consiste, na verdade, em descobrir a quantidade de laranjas contidas em 3 toneladas da fruta. Tendo acesso a esta quantidade e dividindo o valor por 5, que é a quantidade de laranjas consumidas na produção de uma caixa de suco de um litro, poderia se saber a quantidade de caixas de suco que seriam produzidas a partir de 3 toneladas de laranjas. Acontece que, além de não terem contato com a carga, numa situação real ninguém contaria o número de laranjas que constituem as 3 toneladas.

O que se faz, na prática, é estimar o número de laranjas que possivelmente constituem as 3 toneladas tendo como base uma quantidade menor, como, por exemplo, um saco de 40 kg. O problema a ser trabalhado em sala de aula pelo professor consiste em descobrir a quantidade aproximada de laranjas contidas em 40kg da fruta que, de agora em diante, será designada como população, tendo como base 40 laranjas selecionadas aleatoriamente dos 40kg de laranjas da Figura 2.

Figura 2 - Sacos de laranja de 20 kg adquiridos no comércio local

Fonte: Imagem do arquivo pessoal do autor.

3.2 DESENVOLVENDO A PROPOSTA

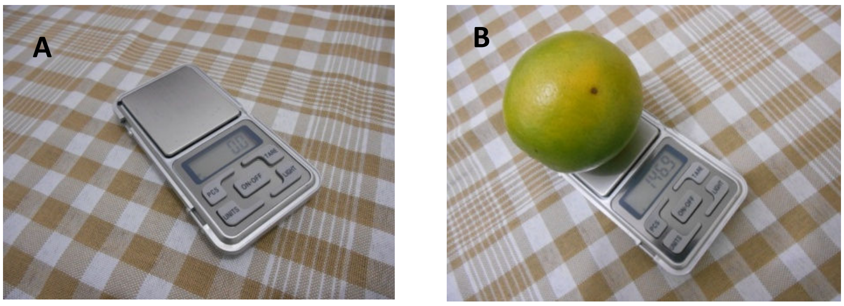
3.2.1 Primeiro contato com as laranjas na sala de aula

Em sala de aula o professor, de posse dos sacos de laranjas pera, poderia iniciar desafiando os estudantes a dizer o número de frutas presentes nos sacos de laranjas. Como certamente surgirão diferentes propostas, o professor novamente desafiaria os estudantes sobre a maneira de resolver o problema em questão. Possivelmente, a primeira solução sugerida seria a contagem de todas as laranjas!

O professor exporia então o problema do produtor de suco de laranja que deseja saber quantas caixas de suco ele conseguiria produzir com uma carga de três toneladas da fruta, levando os estudantes a pensarem se a contagem de todas as frutas seria realmente o modo mais eficiente de resolver o problema em questão.

Dando oportunidade para que os estudantes sugerissem outro método, certamente haverá a sugestão de se dividir os 40 quilos dos sacos pelo peso de uma única laranja. O professor então poderia testar esta possibilidade com os estudantes. Para isto, os estudantes poderiam ser divididos em cinco grupos de, por exemplo, oito integrantes, de modo que cada grupo selecione uma laranja dos sacos e efetue a medida de sua massa ("pesagem") [Figura 3 (B)], fazendo uso de uma balança portátil[2] de baixo custo, facilmente adquirida no comércio local [Figura 3 (A)]. Em seguida, dividindo-se o peso dos sacos de laranjas pelo peso de cada uma das laranjas dos respectivos grupos, obter-se-ia, em tese, o número de laranjas nos sacos, conforme é apresentado na Tabela 3.

Figura 3 - Balança portátil de baixo custo adquirida no comércio local **(A)** e exemplo da medida da massa de uma laranja **(B).**



Fonte: Imagem do arquivo pessoal do autor.

[2] A precisão da balança portátil adquirida é de uma casa decimal.

Tabela 3 - Medidas reais dos pesos[3] de laranjas e as respectivas razões obtidas

Grupo	Peso de uma laranja (gramas)	Razão (peso do saco de laranja/massa da laranja)
1	161,5	247,68
2	182,7	218,94
3	136,3	293,47
4	150,9	265,08
5	126,9	315,21

Fonte: Arquivo pessoal do autor.

Depois de montada a Tabela 3, o professor comentaria que se todas as laranjas dos sacos fossem idênticas à laranja escolhida pelo grupo 1, haveria aproximadamente 248 laranjas nos sacos. No entanto, se todas as laranjas fossem iguais à laranja escolhida pelo grupo 2, haveria aproximadamente 219 laranjas nos sacos. Assim, enfatizando a terceira coluna da Tabela 3, seria evidenciado que, utilizando este método para resolver o problema, para cada grupo haveria um número de laranjas diferentes para os mesmos sacos de laranjas!

Diante disso, o professor poderia indagar, primeiramente, sobre o motivo da obtenção de resultados diferentes e, em seguida, indagar também qual dos resultados poderia ser utilizado para resolver o problema do produtor do suco de laranjas. O professor, então, poderia mostrar que não se deve utilizar o peso de apenas uma fruta, mas sim, usar o peso médio do conjunto de todas as laranjas (NETO; SCARMINIO; BRUNS, 2003), que é uma estatística usada para representar um conjunto de valores. Isso porque, como fica evidente na segunda coluna da Tabela 3, os pesos variam de fruta para fruta. Um problema que surge neste momento é que para obter o peso médio das laranjas contidas nos sacos temos que pesar todas as laranjas dos sacos. Assim, o professor pode comentar que, para contornar essa situação, devemos nos basear não em uma laranja, mas em uma quantidade maior de frutas.

Neste momento, o termo "quantidade maior de frutas" seria utilizado para introduzir a ideia de amostra.

O professor, então, introduziria a ideia de que para estimar a quantidade de frutas em dois sacos de laranjas, o qual seria a população, a partir de uma amostra, pode-se utilizar um conjunto de ferramentas e recursos conhecidos como Estatística Descritiva e Estatística Inferencial.

[3] Estamos utilizando peso e massa como sinônimos.

Em seguida, o professor enfatizaria que estudando estatisticamente o comportamento dos pesos de um pequeno conjunto de laranjas (amostra) é possível fazer inferências sobre a quantidade total de laranjas existente em 40 quilos (população), ou em qualquer outra massa de laranjas, como no caso do produtor que deseja saber a quantidade de caixas de um litro de suco de laranja que poderiam ser produzidas a partir de 3 toneladas de fruta.

3.2.2 Seleção de uma amostra de laranjas

O próximo passo seria orientar e discutir com os estudantes a maneira mais adequada de fazer a seleção das laranjas que irão compor a amostra a ser estudada, de modo que esta amostra seja representativa da população. Neste sentido, o primeiro ponto a ser destacado é que todos os elementos da população (todas as laranjas presentes nos dois sacos da Figura 2) devem ter a mesma chance, isto é, a mesma probabilidade, de fazer parte da amostra. Isto poderia ser feito inicialmente colocando os sacos de laranjas na vertical e mostrando que se as laranjas forem selecionadas com os sacos de laranjas nesta posição, as frutas que estiverem mais acima terão uma chance maior de serem escolhidas do que as frutas que estão mais abaixo. Assim, selecionar as laranjas com o saco nesta disposição não seria correto do ponto de vista estatístico. Então, poderia se indagar aos estudantes sobre como contornar este problema e dar a todas as laranjas do saco a mesma chance de ser escolhida. Talvez uma das sugestões dadas seja retirar todas as laranjas dos sacos e dispô-las sobre uma mesa (Figura 4). Desta forma o professor estaria introduzindo o princípio da amostragem aleatória simples.

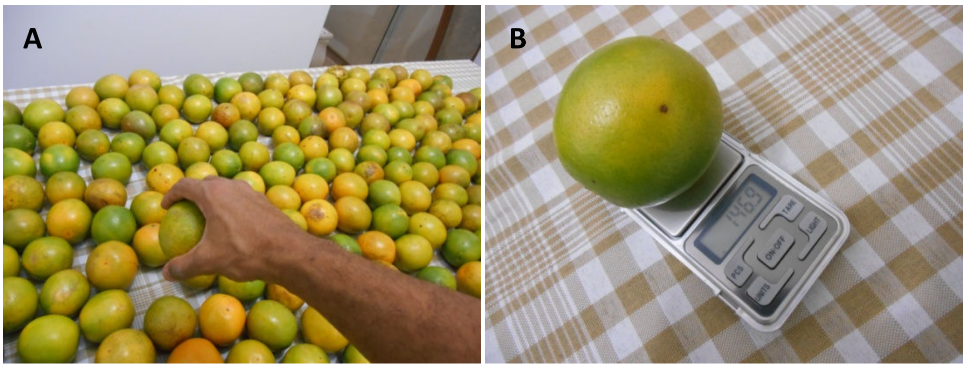
Figura 4 - Disposição de laranjas sobre uma mesa de modo a dar a todas elas a mesma chance de serem selecionadas para constituir uma amostra representativa da população.

Fonte: Imagem do arquivo pessoal do autor.

Outro ponto importante a ser destacado é que depois que uma laranja for selecionada [Figura 5(A)] e tiver seu peso medido [Figura 5(B)], esta deve ser reintegrada ao conjunto [Figura 5(C)] de modo que a próxima laranja a ter o peso medido tenha a mesma chance de ser escolhida. Desse modo, o professor introduziria outro importante conceito estatístico, o conceito de espaço amostral equiprovável.

Ao fazer a coleta da amostra, o professor pode dar a oportunidade a cada aluno da sala de selecionar uma laranja da mesa, pesá-la e devolvê-la à mesa.

Figura 5 - Seleção de uma laranja ao acaso **(A)** para efetuar a medição de seu peso **(B)** e devolvê-la a mesa para que o espaço amostral não se altere **(C)**



Fonte: Imagem do arquivo pessoal do autor.

3.2.3 Construção das tabelas de dados brutos e de distribuição de frequências

Supondo que a turma tenha 40 alunos, então nossa amostra terá 40 pesos. A Tabela 4 apresenta os dados brutos.

Tabela 4 - Pesos das laranjas escolhidas aleatoriamente de um amontoado de 40kg de laranja pera (em gramas)

158,0	164,1	126,9	193,4
161,5	173,3	159,6	165,1
177,3	182,7	158,1	136,3
163,8	155,3	192,7	170,4
171,0	159,5	176,5	150,9
166,7	155,4	141,0	174,0
135,1	140,8	155,4	157,6
193,6	170,3	158,3	174,8
171,6	165,9	155,4	163,9
158,5	174,8	177,3	155,3

Fonte: Arquivo pessoal do autor.

Como a tabela de dados brutos não é prática para fazer análises, o professor deve orientar os alunos para a construção da tabela de frequências, que consiste em listar os valores assumidos pela variável e fazer a contagem do número de ocorrências na tabela de dados brutos. O professor dever lembrar aos alunos que a variável peso da laranja é quantitativa contínua, ou seja, pode assumir qualquer valor dentro de um intervalo real. Neste caso, devemos dividir a faixa total dos pesos em intervalos menores e contarmos as laranjas situadas dentro da cada intervalo (MAGALHÃES; LIMA, 2004). Como o menor peso observado foi 126,9g e o maior 193,6, a faixa 126 – 196g é suficiente para acomodar todos os valores da tabela de frequência. Para determinar o número de classes (ou faixas) o professor pode utilizar uma fórmula que é recomendada pela literatura em geral, e que tem a seguinte formulação: o número de classes é o inteiro mais próximo do resultado da raiz quadrada do valor da amostra.

Na nossa situação, temos 40 observações na Tabela 4, assim, a tabela de frequências dever ter 6 ou 7 classes. O professor dever comentar que é preferível escolher 7, que é um número ímpar, pois assim permitirá enxergar melhor possíveis simetrias (NETO; SCARMINIO; BRUNS, 2003) e ter uma percepção da localização da maior concentração de valores, ou seja, se está localizada no início, no meio, ou na final da distribuição (CRESPO, 2002). Então, pegando-se a amplitude da faixa, 126 – 196g, e dividindo pelo número de classes, 7, obtemos, dessa forma, intervalos de amplitude igual 10g. Agora basta atribuir cada peso medido ao intervalo apropriado.

Seria interessante o professor pedir aos alunos que sentem em dupla para efetuar a contagem do número de ocorrências em cada classe, pois pode haver erros de contagens no trabalho individual e um colega pode alertar o outro. O professor também deve lembrar aos alunos que a notação $126 \vdash 136$ representa o intervalo $[126; 136)$, ou seja o intervalo inclui o limite inferior e exclui o limite superior, que fica incluído no próximo intervalo.

Os alunos chegaram à distribuição de frequência da Tabela 5.

Tabela 5 - Distribuição dos pesos de 40 laranjas extraídas aleatoriamente de um amontoado de 40k de laranja pera.

Intervalo (g)	Nº de laranjas
$126 \vdash 136$	2
$136 \vdash 146$	3
$146 \vdash 156$	6
$156 \vdash 166$	13
$166 \vdash 176$	9
$176 \vdash 186$	4
$186 \vdash 196$	3
Total	40

Fonte: Arquivo pessoal do autor

Percorrendo a coluna de frequências, os alunos podem verificar imediatamente que a maior concentração de laranjas está localizada no meio da distribuição, mais especificamente, ao redor do intervalo $156 \vdash 166$g.

Podemos ainda incluir na tabela frequências uma coluna com as frequências relativas, que consistem em dividir o número de laranjas de cada classe pelo número total de laranjas. Através da frequência relativa os alunos irão perceber a fração do total que representa cada classe. Fazendo as divisões, temos a seguinte Tabela:

Tabela 6 - Distribuição dos pesos de 40 laranjas extraídas aleatoriamente de um amontoado de 40k de laranja pera.

Intervalo (g)	Nº de laranjas	Frequência relativa
$126 \vdash 136$	2	0,050
$136 \vdash 146$	3	0,075
$146 \vdash 156$	6	0,150
$156 \vdash 166$	13	0,325
$166 \vdash 176$	9	0,225
$176 \vdash 186$	4	0,100
$186 \vdash 196$	3	0,075
Total	40	1

Fonte: Arquivo pessoal do autor.

3.2.4 – Cálculo de probabilidades a partir de frequências relativas

Para exemplificar a utilidade da frequência relativa o professor pode dizer aos alunos que no intervalo $156 \vdash 166$g foram observados 13 laranjas de um total de 40. A frequência relativa é, portanto $13 \div 40 = 0{,}325$. Isso significa que $32{,}5\%$ dos pesos ficaram entre 156 e 166g.

Nesse momento o professor pode comentar com os alunos que a frequência relativa também indica a probabilidade de ocorrer certo valor na amostra, pois a probabilidade de um evento ocorrer, no caso de um espaço amostral equiprovável, é calculada pela divisão do número de casos favoráveis que o evento ocorra pelo número de elementos do espaço amostral, que é justamente o que fizemos quando dividimos o número de elementos em cada classe pela frequência total.

E os alunos podem observar também que o somatório da frequência relativa é igual a 1, ou seja, a probabilidade do espaço amostral é 1 ou 100%.

O professor deve alertar que essa probabilidade se refere à amostra e não à população dos 40 quilos de laranjas, pois não temos a distribuição de frequência dessa população.

Assim, o professor pode fazer o seguinte questionamento:

Qual a probabilidade de encontrarmos na amostra uma laranja pesando entre 166g e 176g?

O resultado dessa probabilidade é justamente a frequência relativa da classe $166 \vdash 176$, 0,225, ou seja, 22,5% de chance.

3.2.5 Iniciando a utilização da ferramenta Geogebra

Todo o trabalho proposto com os alunos até agora é feito manualmente. Dessa forma, os alunos ao utilizarem o computador, terão melhor entendimento das técnicas estatísticas utilizadas quando é realizado um comando.

Para tornar as análises dos dados (pesos) mais rápida e eficiente o professor deve agora usar o recurso computacional. Nesse momento o professor pode levar os alunos para o laboratório de informática para efetuar os cálculos com os pesos.

Os alunos podem se sentar em dupla para que haja uma troca de ideias entre eles. O programa utilizado será o Geogebra. É importante que o professor faça uma visão geral do programa antes de iniciar o tratamento dos dados.

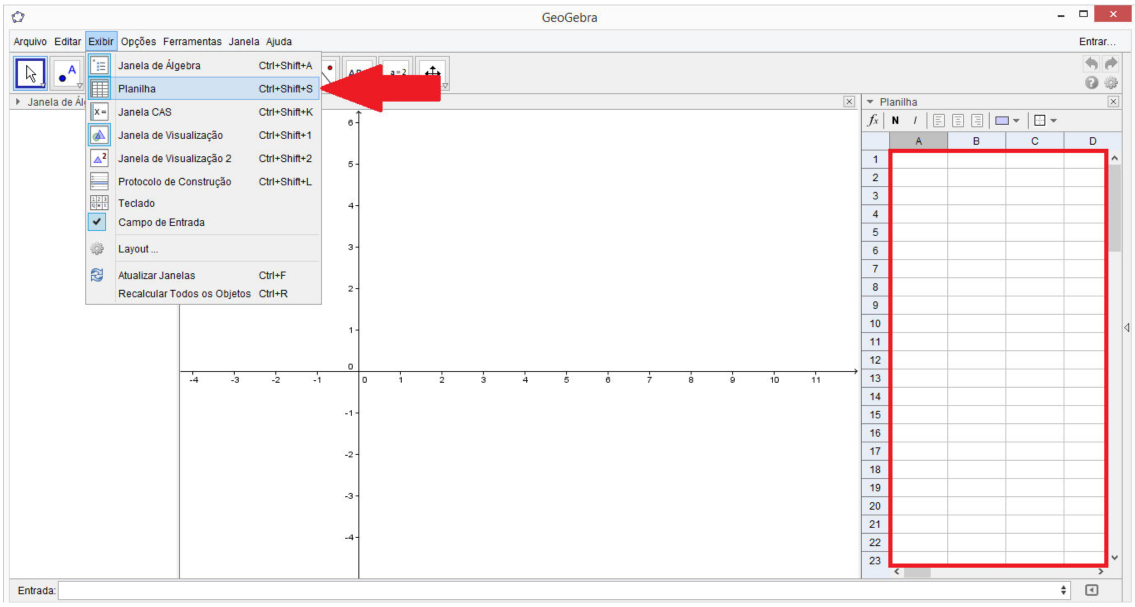
Primeiramente, abrimos o programa com um duplo clique no seu ícone na área de trabalho, como mostra a Figura 6.

Figura 6 - Abrindo o programa Geogebra na área de trabalho.

Fonte: Imagem do arquivo pessoal do autor.

Como estamos trabalhando com dados numéricos devemos habilitar a planilha, pois ela dá agilidade no tratamento de dados (DUARTE, 2010, p 58). Para habilitar a planilha devemos clicar no menu exibir na *Barra de Menus* e selecionar planilha (Figura 7).

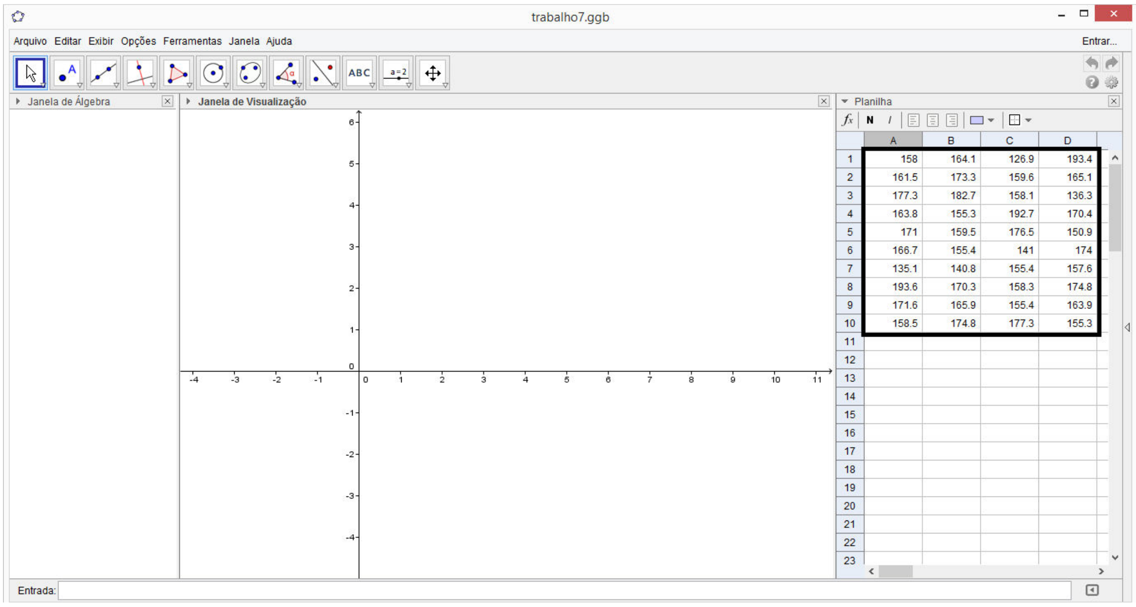
Figura 7 - Exibir planilha



Fonte: Geogebra.

Aberta a planilha no Geogebra o professor orienta aos alunos a inserirem os 40 pesos, cada um em uma célula. A ordem em que os pesos são inseridos na tabela não irá interferir na análise feita pelo Geogebra. Vamos inserir os pesos na mesma disposição da Tabela 4. Neste caso de A1 até D10, conforme a Figura 8.

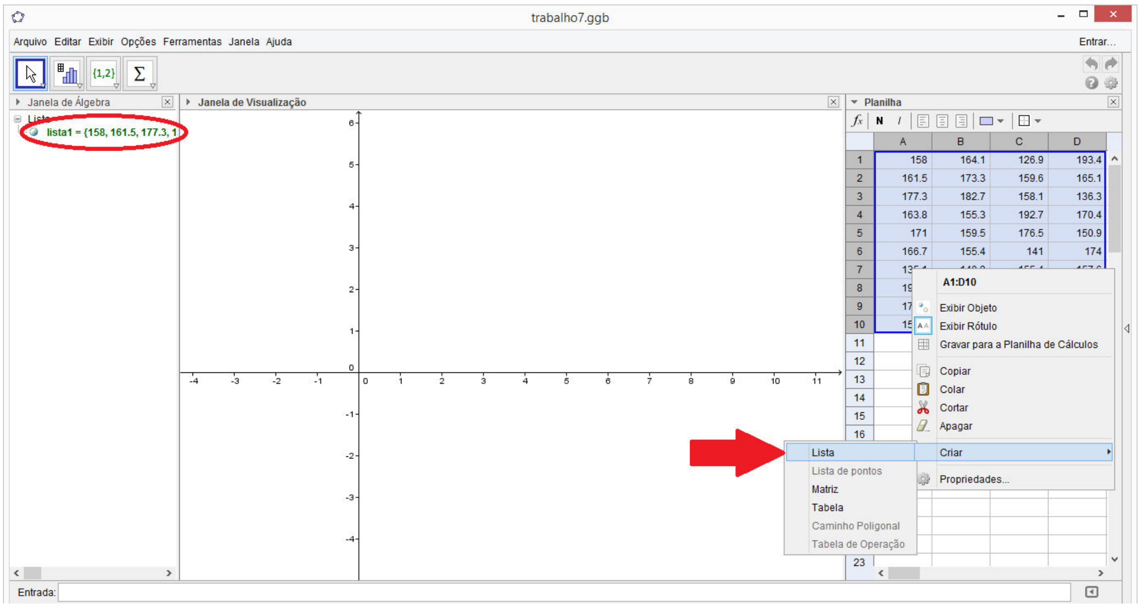
Figura 8 - Inserindo os 40 pesos na planilha



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Depois de lançados os pesos na planilha, o próximo passo é criar uma lista de dados brutos. Selecionamos todas as células preenchidas com os pesos, clicamos no botão direito do *mouse*, aparecerá algumas opções, selecionamos criar e em seguida lista (Figura 9).

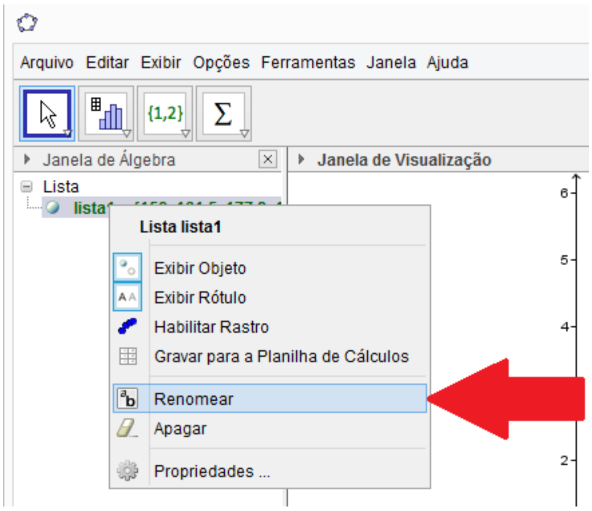
Figura 9 - Criando lista peso



Fonte: Geogebra

Criada a lista de dados brutos, é aconselhável renomear a lista para evitar confusões posteriores. Para isso, clicamos com botão direito sobre a lista e depois em renomear conforme Figura 10.

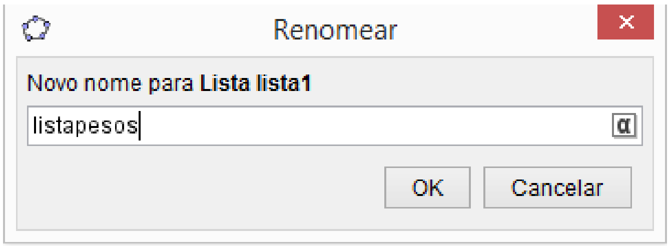
Figura 10 - Renomeando lista de pesos



Fonte: Geogebra.

Aparecerá uma caixa para renomear, e digitamos o nome da lista, por exemplo, "listapesos" e selecionamos *ok* (Figura 11).

Figura 11 - Caixa renomear lista

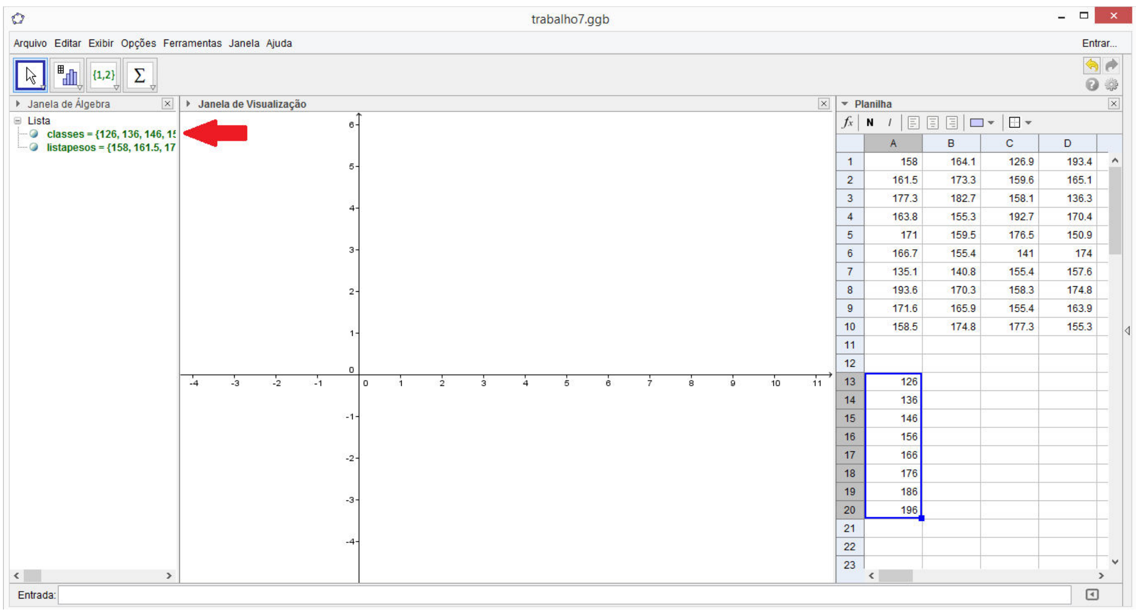


Fonte: Geogebra

3.2.6 Construção da tabela de frequência para variável peso através do Geogebra

Para construir uma tabela de frequência para a variável peso devemos primeiramente criar uma lista com os limites das classes. Repetindo o procedimento que usamos na secção 3.2.3, criaremos limites de classe com amplitude 10g começando por 126g até 196g. Digitamos esses valores nas células da planilha de A13 até A20. Em seguida, renomeamos a lista, criando uma lista chamada "classes" (Figura 12).

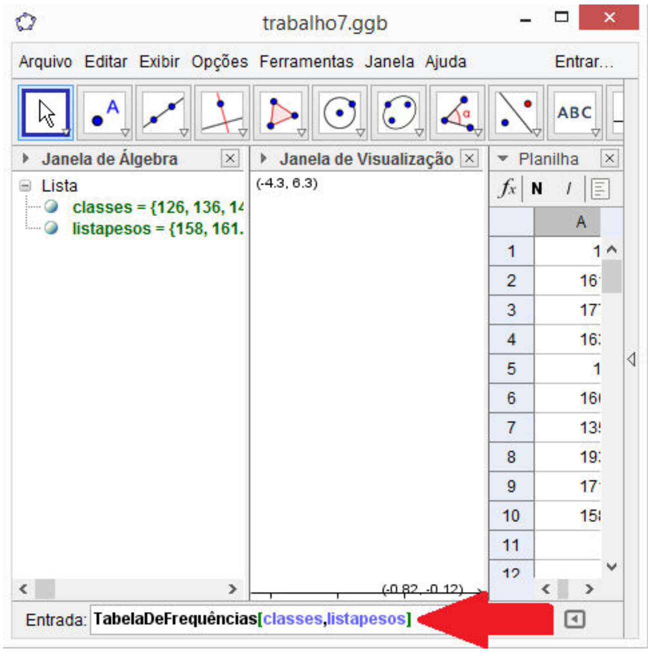
Figura 12 - Criando e renomeando lista de classes



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Para finalizar, digitamos no campo entrada de comandos: $TabelaDeFrequências[< Lista\ dos\ Limites\ das\ Classes >, < Lista\ dos\ Dados\ Brutos >]$. Neste caso $TabelaDeFrequências[classes, listapesos]$ e digitamos *enter* como pode ser visto na Figura 13.

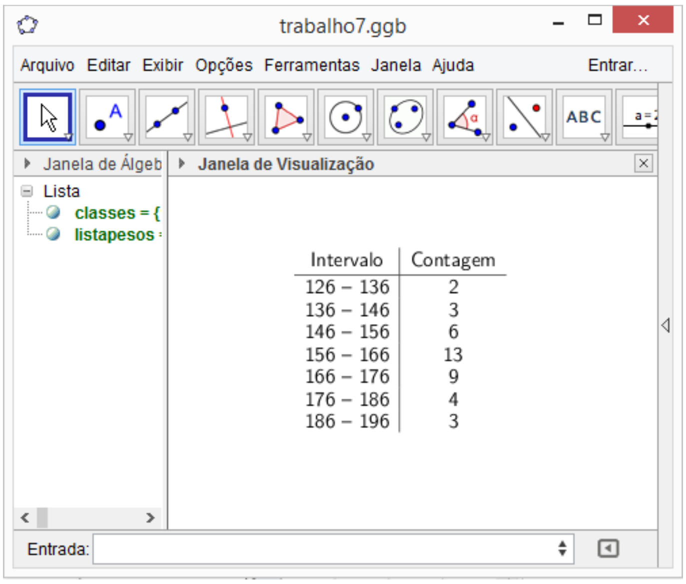
Figura 13 - Criando uma tabela de frequência para a variável peso.



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Assim, criamos através do Geogebra a tabela de frequência para variável peso da laranja conforme Figura 14.

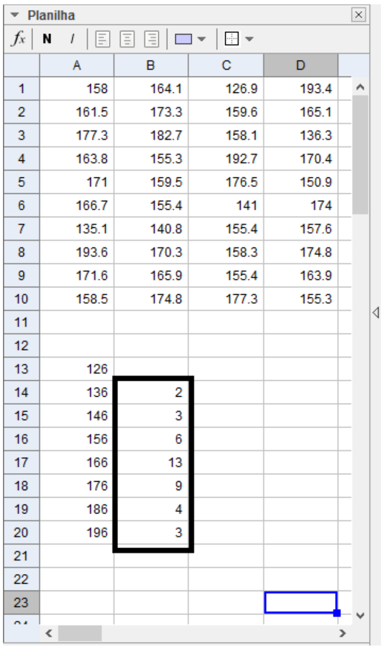
Figura 14 - Tabela de frequência para variável peso



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Também podemos calcular a frequência relativa no Geogebra. Primeiramente digitamos as frequência absolutas em uma coluna da planilha. Para facilitar a identificação da frequência relativa, vamos digitá-las ao lado dos limites de classes, a partir do segundo limite, da célula B14 até B20 (Figura 15).

Figura 15 – Frequências absolutas

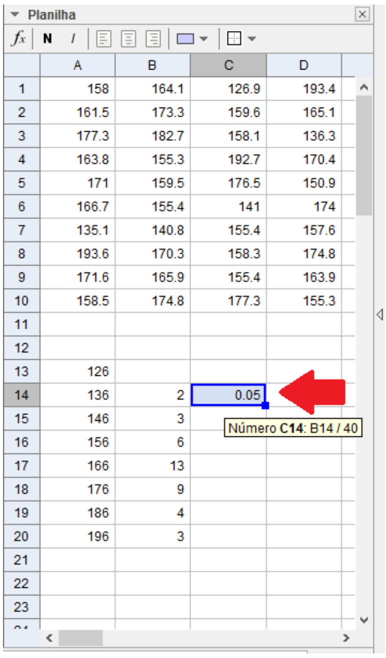


	A	B	C	D
1	158	164.1	126.9	193.4
2	161.5	173.3	159.6	165.1
3	177.3	182.7	158.1	136.3
4	163.8	155.3	192.7	170.4
5	171	159.5	176.5	150.9
6	166.7	155.4	141	174
7	135.1	140.8	155.4	157.6
8	193.6	170.3	158.3	174.8
9	171.6	165.9	155.4	163.9
10	158.5	174.8	177.3	155.3
11				
12				
13	126			
14	136	2		
15	146	3		
16	156	6		
17	166	13		
18	176	9		
19	186	4		
20	196	3		
21				
22				
23				

Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Colocaremos agora um comando na célula C14. Como já é de conhecimento dos alunos, a frequência relativa é o quociente da frequência de cada classe pela frenquência total. Assim, digitamos na célula C14 o comando:= B14/40, onde B14 é celula que contem a frequência da primeira classe, e digitamos *enter* (Figura 16).

Figura 16 - Inserido o comando $= \boldsymbol{B14/40}$ na célula B14

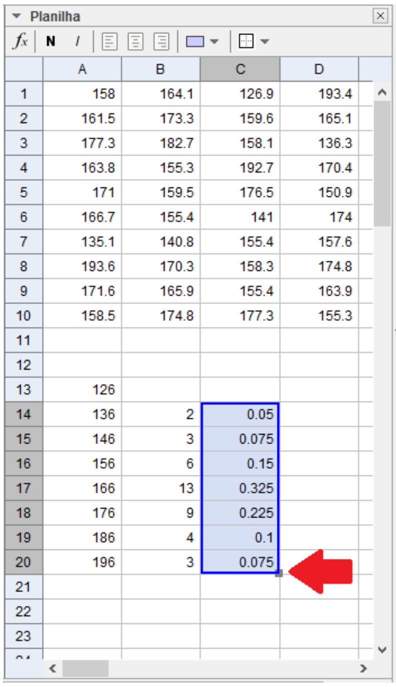


	A	B	C	D
1	158	164.1	126.9	193.4
2	161.5	173.3	159.6	165.1
3	177.3	182.7	158.1	136.3
4	163.8	155.3	192.7	170.4
5	171	159.5	176.5	150.9
6	166.7	155.4	141	174
7	135.1	140.8	155.4	157.6
8	193.6	170.3	158.3	174.8
9	171.6	165.9	155.4	163.9
10	158.5	174.8	177.3	155.3
11				
12				
13	126			
14	136	2	0.05	
15	146	3		
16	156	6		
17	166	13		
18	176	9		
19	186	4		
20	196	3		
21				
22				
23				

Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Obtemos assim, a frequência relativa da primeira classe. Para obtermos a frequência relativa das outras classes, clicamos na borda inferior direita da célula C14 e arastamos até a célula C20 como pode ser visto na Figura 17.

Figura 17 - Frequência relativas



	A	B	C	D
1	158	164.1	126.9	193.4
2	161.5	173.3	159.6	165.1
3	177.3	182.7	158.1	136.3
4	163.8	155.3	192.7	170.4
5	171	159.5	176.5	150.9
6	166.7	155.4	141	174
7	135.1	140.8	155.4	157.6
8	193.6	170.3	158.3	174.8
9	171.6	165.9	155.4	163.9
10	158.5	174.8	177.3	155.3
11				
12				
13	126			
14	136	2	0.05	
15	146	3	0.075	
16	156	6	0.15	
17	166	13	0.325	
18	176	9	0.225	
19	186	4	0.1	
20	196	3	0.075	
21				
22				
23				

Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

3.2.7 Construção do histograma para a variável peso

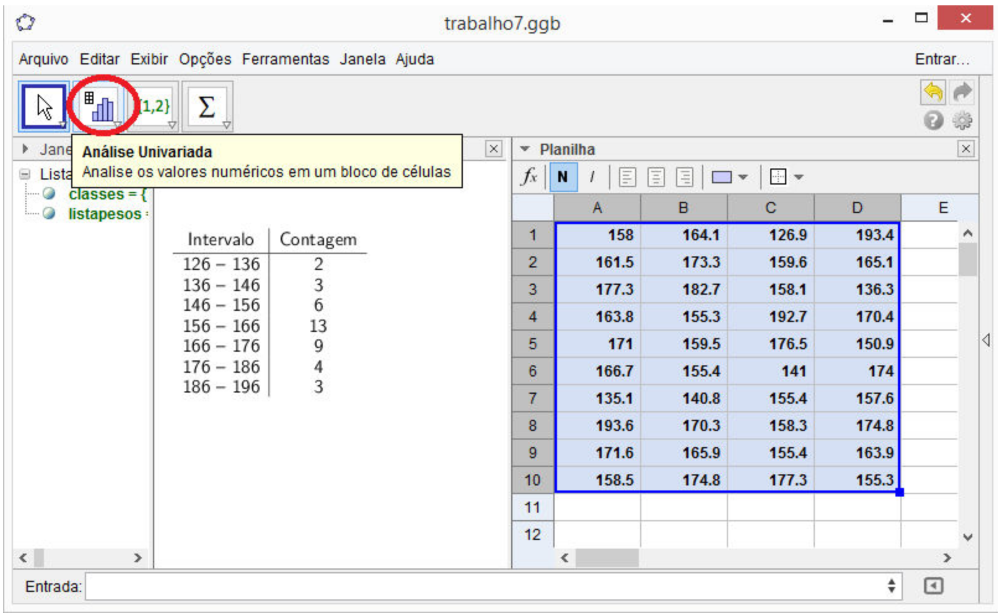
Construída a tabela de frequência, o professor deve questionar os alunos sobre onde está localizada a maior concentração de valores da variável peso da laranja. Os alunos irão perceber que a maior concentração de valores está localizada entre 156 e 166 gramas. Essa concentração pode ser mais bem visualizada através da representação gráfica da distribuição da distribuição de frequência, denominada histograma. O histograma consiste em retângulos justapostos com base na faixa de valores da variável e com área igual à frequência relativa da respectiva faixa. A altura de cada retângulo é denominada densidade de frequência definida pelo quociente da área (frequência relativa da faixa) pela amplitude da faixa. Em símbolos:

$$A_R = frequência\ relativa\ da\ classe = amplitute\ da\ classe\ \times altura$$

$$altura = \frac{frequência\ relativa\ da\ classe}{amplitude\ da\ classe} \qquad (1)$$

Para construir o histograma a partir do Geogebra, selecionamos os dados brutos na planilha e clicamos na janela "análise univariada" na *Barra de Ferramentas* (Figura 18).

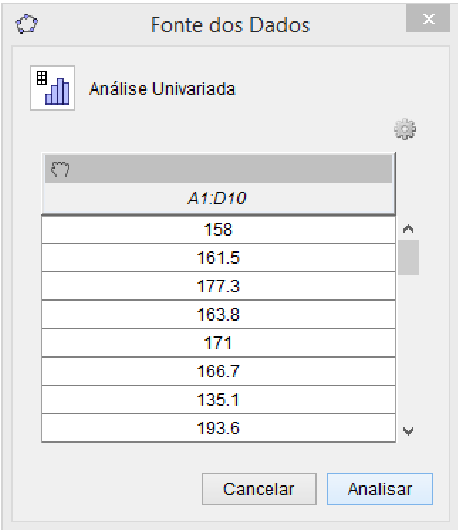
Figura 18 - Construção do histograma para a variável peso



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Aparecerá uma caixa de dialogo e clicamos em "analisar" conforme Figura 19.

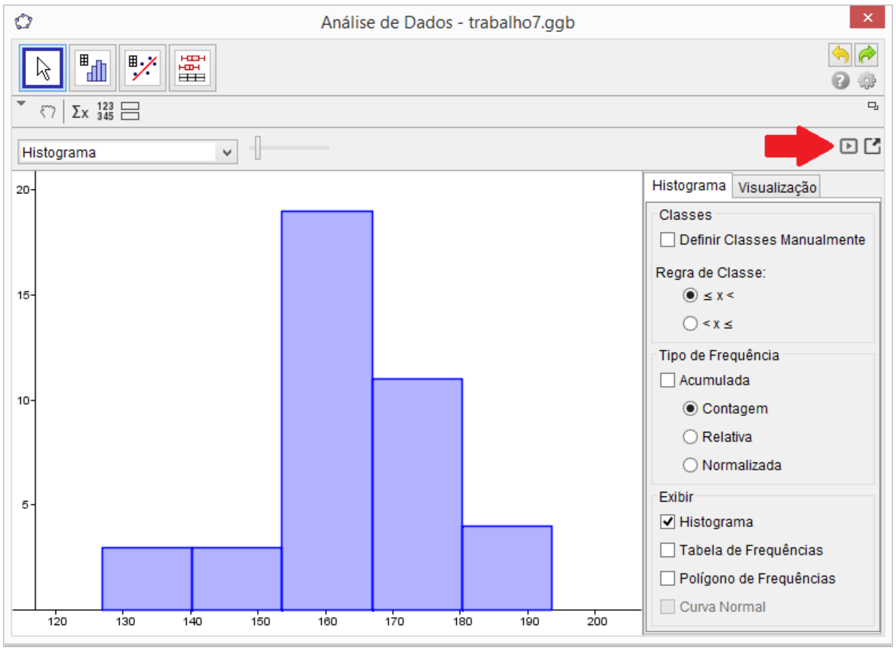
Figura 19 - Caixa de dialogo “analise univariada”



Fonte: Construído no Geogebra

Surgirá um histograma automaticamente. Agora temos que colocar nossas configurações de limites de classes. Clicamos no ícone opções (Figura 20).

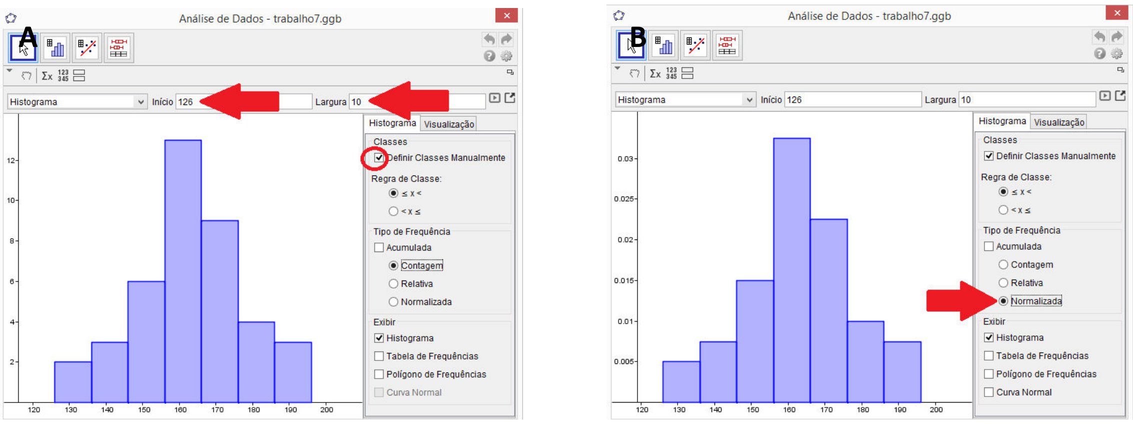
Figura 20 - Opções para o histograma



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

Selecionamos definir classes manualmente e em seguida podemos definir o tipo de frequência que corresponderá à altura dos retângulos. Aqui, o professor pode comentar sobre essa questão e lembrar aos alunos a definição de histograma. Assim, a altura de cada retângulo deve ser obtida de modo que sua área seja igual à frequência relativa da respectiva classe. A opção que dará a ordenada de cada retângulo dessa forma é “normalizada”, que corresponde à densidade de frequência [Figura 21(B)]. Colocando o limite da primeira classe e a amplitude de cada classe, temos o histograma da distribuição como pode ser visto na Figura 21.

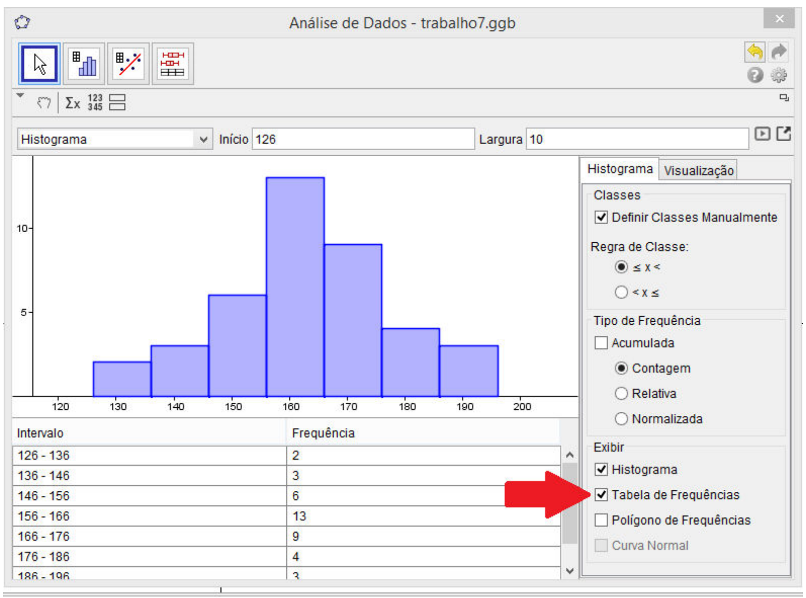
Figura 21 - Selecionando os limites de classes manualmente **(A)** e a opção normalizada (densidade) para a altura dos retângulos **(B)**



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Também podemos selecionar a opção exibir tabela de frequência (Figura 22).

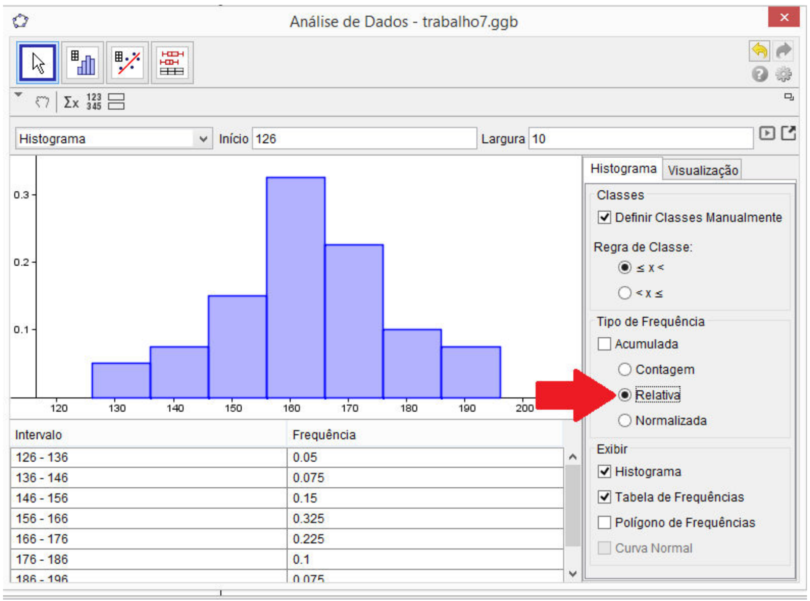
Figura 22 - Exibir tabela de frequência



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

Podemos escolher também o tipo de frequência. Essa opção é uma maneira rápida de se obter a frequência relativa (Figura 23).

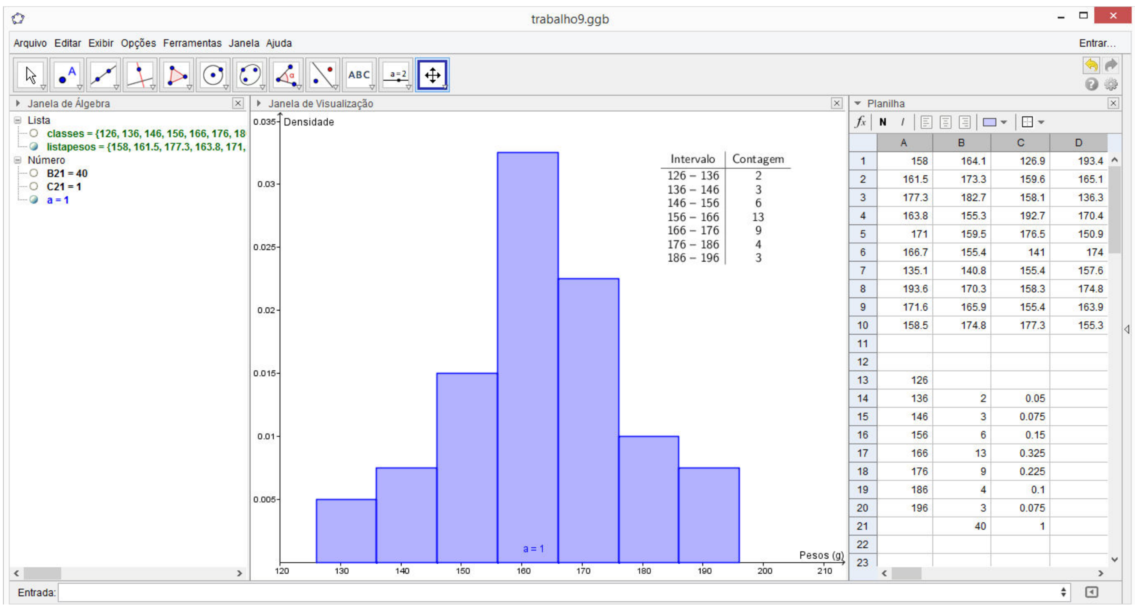
Figura 23 - Exibir tabela de frequência com frequência relativa



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Configurando os eixos, isto é, chamando o eixo x de eixo dos pesos e o eixo y de eixo das densidades, temos o histograma da Figura 24.

Figura 24 - Histograma da variável peso



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Com a construção do histograma os alunos irão perceber melhor onde está localizada a maior concentração de pesos e também perceberão um padrão nesta distribuição, que nela há uma simetria em relação ao retângulo do meio. O professor pode chamar a atenção dos alunos para o fato de que há uma tendência dos pesos em se concentrar em torno do valor 160g.

O professor nesse momento deve lembrar as características do histograma, comentando que a largura da base de cada retângulo corresponde à amplitude da classe correspondente, que a área de cada retângulo é igual à frequência relativa da respectiva classe e que a altura é dos retângulos é chamada de densidade.

3.2.8 Características do histograma construído

Assim, o histograma da Figura 24 é uma representação gráfica dos pesos das 40 laranjas da nossa amostra. E tem como características:

- a localização do conjunto de valores numa certa região do eixo horizontal;

- sua dispersão, ou espalhamento, ao longo dessa região.

Podemos representar numericamente essas características por várias grandezas. As mais usadas são a média aritmética e o desvio padrão (NETO; SCARMINIO; BRUNS, 2003). A primeira nos fornece valor no qual os dados tendem a se aglomerar, e a segunda analisa o quão espalhados estão os dados de uma variável, em torno do valor central.

3.2.9 Média aritmética

A média aritmética de um conjunto de valores é uma medida de localização ou tendência central, e é calculada somando todos os valores e dividindo pelo número total de elementos do conjunto.

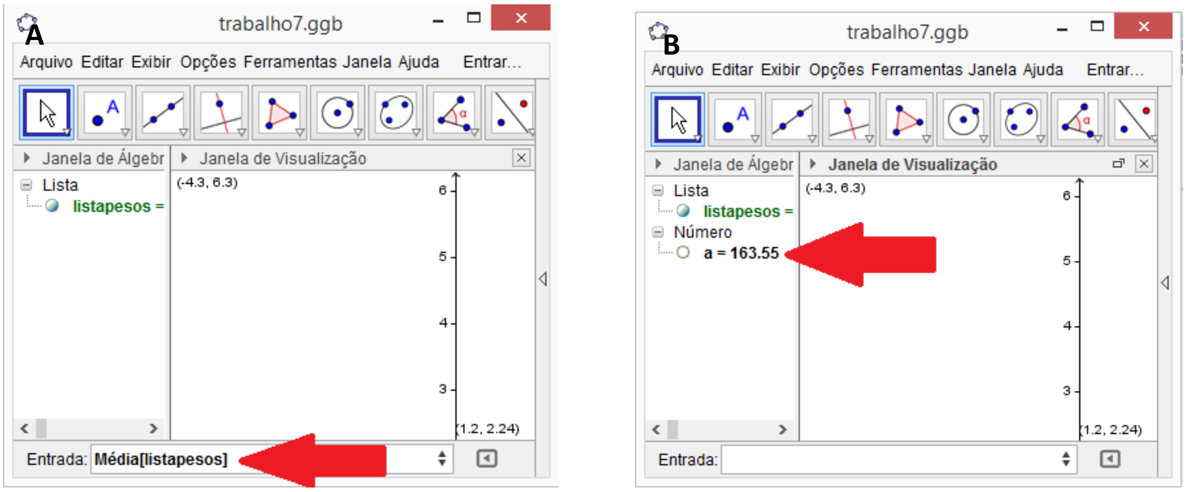
Tendo em mãos os pesos da amostra de 40 laranjas, o professor pode orientar os alunos na busca da média dessas quarenta laranjas utilizando a fórmula:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (2)$$

O cálculo da média desses quarenta valores pode ser um pouco trabalhoso e o professor neste caso pode orientar os alunos a usar o recurso computacional.

Para calcular a média, devemos entrar com o seguinte comando: $Média[< Lista\ dos\ Dados\ Brutos >]$. Digitamos “listapesos” e *enter* (Figura 25).

Figura 25 - Comando média **(A)** para obter o peso médio da amostra de 40 laranjas **(B)**



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

3.2.10 Estimativa do número de laranjas usando a média aritmética

Tendo em mãos o valor[4] do peso médio, o professor orienta os alunos a dividir o peso dos sacos de laranjas, 40000g, pelo peso médio da amostra de 40 laranjas para se obter uma estimativa da quantidade de laranjas contidas em quarenta quilos da fruta. Fazendo os cálculos, temos:

$$\frac{peso\ da\ população\ de\ laranjas}{peso\ médio\ amostral} = \frac{40000g}{163{,}55g} = 244{,}57 \qquad (3)$$

Portando, temos aproximadamente 245 laranjas nos 40kg da fruta.

O professor, neste momento, deve perguntar aos alunos se esta seria a melhor estimativa do número de laranjas que podíamos encontrar.

Para esclarecer os alunos desse fato, o professor pode expor a seguinte problemática:

Se coletarmos outra amostra de 40 laranjas da mesma população teremos exatamente a mesma média que encontramos?

[4] A média será calculada com uma casa decimal a mais que os dados originais.

Certamente haveria uma indecisão por parte dos alunos. O professor então propõe aos alunos fazerem novamente todo o procedimento. Suponha que a nova média encontrada seja $\bar{x} = 186{,}79g$. Fazendo novamente os cálculos, temos:

$$\frac{40000g}{186{,}79g} = 214{,}14 \tag{4}$$

Assim, com a nova média, teríamos aproximadamente 214 laranjas contidas em 40kg da fruta. Um valor diferente do que encontramos com a primeira média.

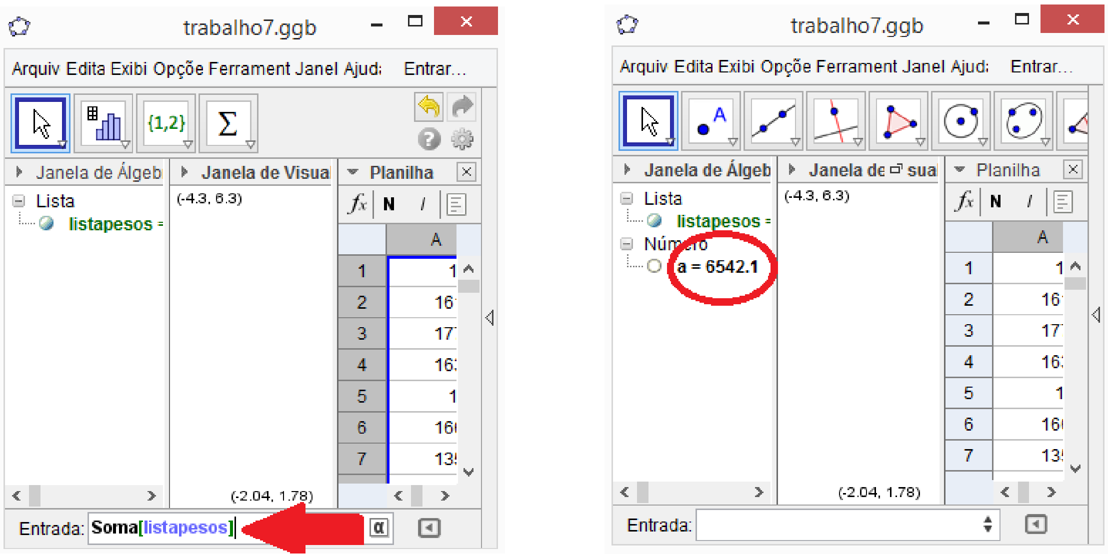
Os alunos podem ficar surpresos com essa situação, pois volta à situação inicial, ou seja, valores diferentes.

Essas estimativas foram obtidas a partir do peso médio de apenas 40 laranjas e supõem que haja em torno de 230 laranjas. Como esses valores são apenas médias calculadas de amostras, ou seja, médias amostrais, não devem corresponder à média exata dos pesos dos 40kg de laranjas (NETO; SCARMINIO; BRUNS, 2003). A média amostral é uma estimativa da verdadeira média. Para obtermos um número de laranjas próximo do verdadeiro número de laranjas contidas em 40kg, devemos obter uma estimativa mais precisa do verdadeiro peso médio de todas as laranjas contidas em 40kg.

Para exemplificar como o peso médio populacional, ou seja, o peso médio de todas as laranjas contidas em 40kg da fruta, é um instrumento eficiente na busca de uma estimativa precisa do número de laranjas contidas nesses 40kg da fruta, considere que a nossa amostra de 40 laranjas seja a população. Suponhamos que não soubéssemos o número de laranjas dessa pequena população, mas que o peso total e o peso médio dessa massa de laranjas fossem conhecidos.

O peso médio das 40 laranjas é $\bar{x} = 163{,}55g$. Para descobrimos o peso dessas 40 laranjas, basta somar seus pesos. Para fazer a soma no Geogebra, digitamos o comando: $Soma[< Lista >]$. Digitamos “listapesos” e teclamos *enter* conforme Figura 26.

Figura 26 - Comando soma **(A).** Soma dos pesos da pequena população de 40 laranjas **(B)**



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Assim, dividindo a soma dos pesos da pequena população de quarenta laranjas pelo peso médio das mesmas, teríamos o número de laranjas dessa população:

$$\frac{Peso\ da\ população\ de\ 40\ laranjas}{Peso\ médio\ populacional} = \frac{6542{,}1g}{163{,}55g} = 40{,}0 \qquad (5)$$

Portanto, o número de laranjas da nossa pequena população é, de fato, 40. Apesar de esse cálculo ser bem lógico, os alunos acharão esse fato interessante e será mais uma motivação na busca de uma estimativa do verdadeiro peso médio das laranjas contidas em 40kg.

Se coletássemos várias amostras de 40 laranjas e calculássemos o respectivo peso médio iríamos obter valores acima e valores abaixo de $\bar{x} = 163{,}55$g. Isso indica que o verdadeiro peso médio pode estar dentro desse intervalo de variação do peso médio amostral.

Assim, para calcular o número de laranjas contidas em 40kg da fruta precisamos encontrar um intervalo de variação que possa conter a verdadeira média.

O professor agora pode iniciar a introdução das técnicas estatísticas para estimar com mais precisão a média de uma população. Os conteúdos vistos na parte de estatística do Ensino Médio são insuficientes para resolver esse problema. Cabe ao professor aqui, esboçar e comentar essas técnicas com os estudantes. Apesar do conteúdo para ser resolver esta situação ser de nível superior e o grau (rigor matemático) de dificuldade ser maior, o professor não deve se preocupar com as demonstrações e definições rigorosas, e sim com os conceitos e utilizações.

O professor dever comentar com os alunos que a técnica estatística utilizada para resolver essa situação se chama estimativa por intervalo de confiança. A questão agora é: como construir esse intervalo de confiança para o peso médio da população de 40k de laranjas?

Primeiramente precisamos do peso médio amostral. Esse valor já foi obtido em cálculos anteriores feitos com os alunos. Precisamos também do desvio padrão amostral.

3.2.11 Desvio padrão

É uma medida de espalhamento dos valores em torno da média.

Para calcularmos o desvio padrão, primeiro devemos calcular a diferença, ou desvio, de cada valor individual em relação à média amostral:

$$d_i = x_i - \bar{x} \tag{6}$$

O próximo passo é somar os quadrados de todos os desvios e dividir o total por $n - 1$, ou seja, o número de elementos da amostra menos 1. O resultado dessas operações é chamado de variância amostral, denotada por s^2.

Temos assim, a seguinte fórmula:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1} = \frac{1}{n - 1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{7}$$

Onde: x_i são os valores da variável

n é o número de valores

$\bar{x}$ é a média amostral

Os estudantes irão notar que a variância é uma espécie de média dos quadrados dos desvios, só que o denominador não é o número total de observações, n, e sim $n - 1$.

Para que os alunos entendam o motivo dessa mudança, o professor deve chamar a atenção dos alunos para o fato de que as observações originais, obtidas por amostragem aleatória, eram todas independentes. Mesmo conhecendo os pesos de todas as 39 primeiras laranjas, não teríamos como prever exatamente qual seria o peso da próxima laranja a ser selecionada, a 40ª. Assim, temos um grupo de 40 valores totalmente independentes, onde um valor individual qualquer não depende dos valores restantes (NETO; SCARMINIO; BRUNS, 2003) .

Na Estatística, dizemos que esse conjunto de valores tem 40 graus de liberdade.

Agora, com os desvios a situação é diferente.

Somando todos os desvios, temos:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x}$$

Como $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, temos

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0 \tag{8}$$

Assim, temos uma propriedade dos desvios em relação à média, e sua soma é 0. Portanto, os 40 desvios não são todos independentes. Se conhecermos 39 deles, o valor que falta estará automaticamente determinado, isto é, é aquele que torna a soma igual a 0. O denominador da expressão da variância é $n - 1$, pois somente $n - 1$ pode flutuar aleatoriamente (NETO; SCARMINIO; BRUNS, 2003).

Para exemplificar, o professor pode propor para os alunos calcular os desvios em relação à média dos 5 primeiros valores da primeira coluna da Tabela 4 e somá-los.

Assim, eles irão perceber que o resultado é nulo. Conhecendo 4 valores, imediatamente saberemos o 5º, aquele que torna a soma igual a 0. Temos então 4 valores independentes, ou 4 graus de liberdade.

Agora, peça para os alunos calcularem a média e a variância desses 5 valores.

Fazendo as contas, temos:

Média

$$\bar{x} = \frac{158 + 161{,}5 + 177{,}3 + 163{,}8 + 171}{5} = \frac{831{,}6}{5} = 166{,}32g \tag{9}$$

Variância

$$\begin{aligned} s^2 &= \frac{1}{4}[(158 - 166{,}32)^2 + (161{,}5 - 166{,}32)^2 + \cdots + (171 - 166{,}32)^2] \\ &= 60{,}32g^2 \end{aligned} \tag{10}$$

O professor deve chamar a atenção dos alunos para o fato de que a média tem a mesma unidade das observações, enquanto a unidade da variância é o quadrado das unidades originais. Assim, para que as medidas de dispersão e de tendência central tenham as mesmas unidades, extraímos a raiz quadrada da variância, obtendo assim uma medida chamada *desvio padrão*.

Assim, extraindo a raiz quadrada da variância, obtemos o desvio padrão[5] dos 5 primeiros pesos da primeira coluna da Tabela 4.

$$s = \sqrt{s^2} = \sqrt{60{,}32} = 7{,}77kg \tag{11}$$

Assim, a fórmula para o desvio padrão amostra será:

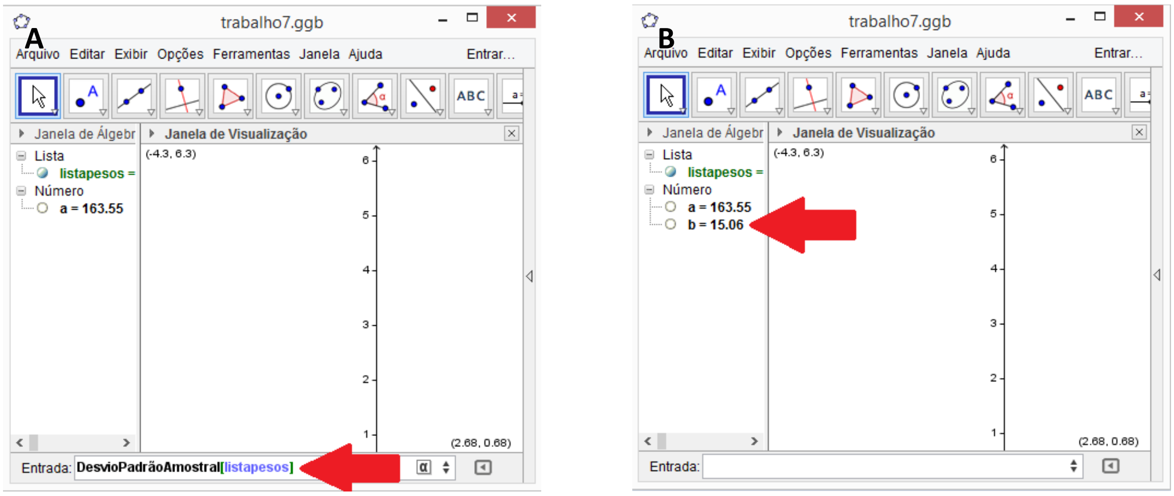
$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}} \tag{12}$$

O cálculo do desvio padrão amostral é mais trabalhoso do que a média. Vamos novamente fazer o uso do Geogebra.

Para calcular o desvio padrão devemos digitar o comando: $DesvioPadrãoAmostral[< Lista\ dos\ Dados\ Brutos >]$. Digitamos "listapesos", que é a lista de pesos das 40 laranjas da amostra, e teclamos *enter* (Figura 27).

[5] O desvio padrão será calculado com uma casa decimal a mais que os dados de partida.

Figura 27 - Comando desvio padrão amostral **(A)** para obter o desvio padrão amostral dos pesos da amostra de 40 laranjas **(B)**



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

3.2.12 Intervalos ao redor da média

Calculada a média e o desvio padrão dos pesos das 40 laranjas da amostra, o professor deve comentar a utilidade dessas medidas na resolução do problema de encontrar o número aproximado de laranjas contidas em 40kg da fruta.

O desvio padrão é usado para definir intervalos em torno da média.

Para exemplificar, o professor pode definir alguns intervalos. Em nossa amostra de 40 laranjas, a média é $\bar{x} = 163{,}55g$ e o desvio padrão é $s = 15{,}06g$, o professor pode definir os seguintes intervalos:

Intervalo definido por um desvio padrão em torno da média: $163{,}55 \pm 15{,}06$, ou 148,49g e 178,61g. Intervalo definido por dois desvios padrão em torno da média: $163{,}55 \pm 30{,}12$, ou 133,43g e 193,67g. Intervalo definido por três desvios padrão em torno da média: $163{,}55 \pm 45{,}18$, ou 118,37g e 208,73g.

Intervalos como estes serão utilizados para estimar o peso médio da população de quarentas quilos de laranja.

O professor deve esclarecer aos alunos que os valores encontrados até agora foram obtidos a partir dos pesos das 40 laranjas, ou seja, são estimativas amostrais.

Para encontrarmos o número aproximado de laranjas contidas em 40kg da fruta, precisamos dos valores populacionais, ou seja, da média populacional e o desvio padrão populacional.

3.2.13 Parâmetros da população

Na Estatística esses valores desconhecidos da população são chamados de parâmetros populacionais e são representados por letras gregas. A média e o desvio padrão populacionais são representados, respectivamente, pelas letras gregas μ e σ.

Como inferir algo a respeito desses valores dispondo apenas dos valores amostrais $\bar{x}$ e s?

É uma pergunta que pode passar na cabeça dos estudantes nessa etapa do problema.

O professor pode expor a seguinte situação para os alunos:

Consideremos as 40 laranjas de nossa amostra como uma minipopulação de 40 elementos. Na tabela de frequência, na coluna das frequências relativas, podemos verificar que 15% desses elementos pesam entre 146g e 156g. Como a frequência relativa corresponde à probabilidade, podemos dizer que a probabilidade de retirarmos aleatoriamente uma laranja com o peso na faixa 146⊢156g é exatamente 0,15.

O Professor deve esclarecer que essa afirmação só pode ser feita porque se conhece a distribuição exata das frequências dos pesos da pequena população de 40 laranjas.

Mas não podemos fazer a mesma afirmação em relação aos 40kg de laranjas, pois não conhecemos a distribuição exata de suas frequências. Para obtermos a distribuição exata da população teríamos que pesar todas as laranjas uma por uma.

Precisamos neste momento de um modelo que descreva o comportamento dos pesos na nossa população de laranjas, ou seja, nos dê uma forma de sua distribuição. Dessa forma poderíamos tirar conclusões do próprio modelo, sem ter que fazer nenhum esforço experimental a mais (NETO; SCARMINIO; BRUNS, 2003, p. 26).

O professor deve comentar com os alunos sobre o uso de modelos na matemática, como por exemplo, as funções, que são usadas para descrever comportamento entre duas ou mais grandezas. Através de um modelo podemos estudar várias características do fenômeno em estudo.

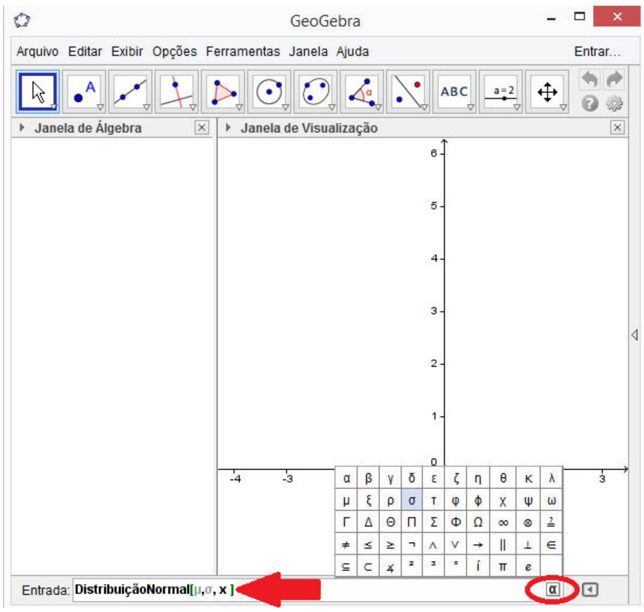
3.2.14 Introdução da Distribuição Normal

O modelo usado para descrever o comportamento da nossa variável é a Distribuição Normal. Distribuição Normal é uma distribuição de probabilidade usada para descrever o comportamento de uma variável aleatória.

O professor pode começar o conteúdo sobre a Distribuição Normal comentando sobre sua história e, em seguida, fazer uma exploração gráfica ressaltando os efeitos dos parâmetros na forma do modelo (BITTENCOURT; VIALI, 2006).

No Geogebra os alunos podem visualizar o gráfico da Distribuição Normal. No campo de entrada de comando digitamos: $DistribuiçãoNormal[< Média >, < Desvio\ Padrão >, x]$. Nos devidos campos colocamos os parâmetros populacionais μ e σ, usando a lista de símbolos e letras gregas do Geogebra. Teclamos *enter* (Figura 28).

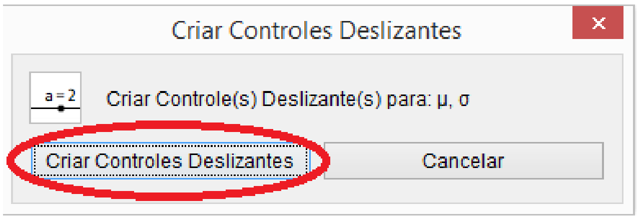
Figura 28 - Comando Distribuição Normal



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Assim, aparecerá uma caixa pedindo para criar controles deslizantes para a média μ e o desvio padrão σ como pode ser visto na Figura 29.

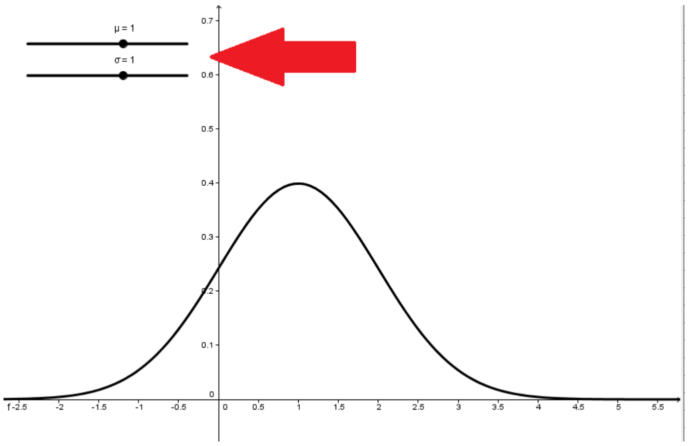
Figura 29 - Caixa de controles deslizantes



Fonte: Geogebra

Com os controles deslizantes podemos variar os valores da média e o desvio padrão populacionais movendo os pontos sobre os segmentos e verificar seus efeitos sobre a Curva Normal (Figura 30).

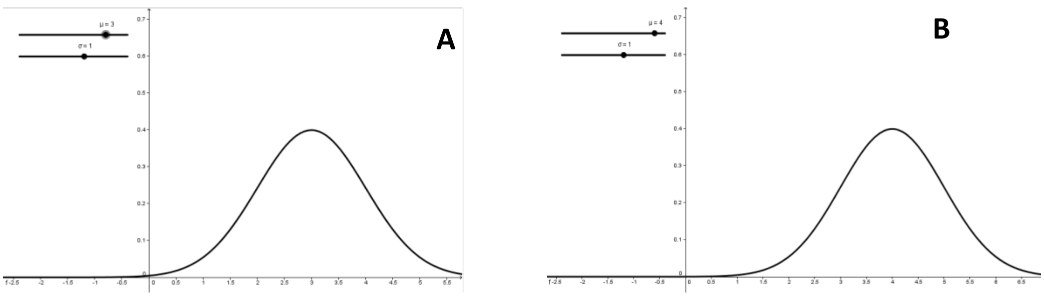
Figura 30 - Controles deslizantes para $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}$



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Movendo o controle deslizante da média μ os alunos perceberão que haverá um movimento de translação da distribuição e que ela indica o centro da distribuição conforme Figura 31.

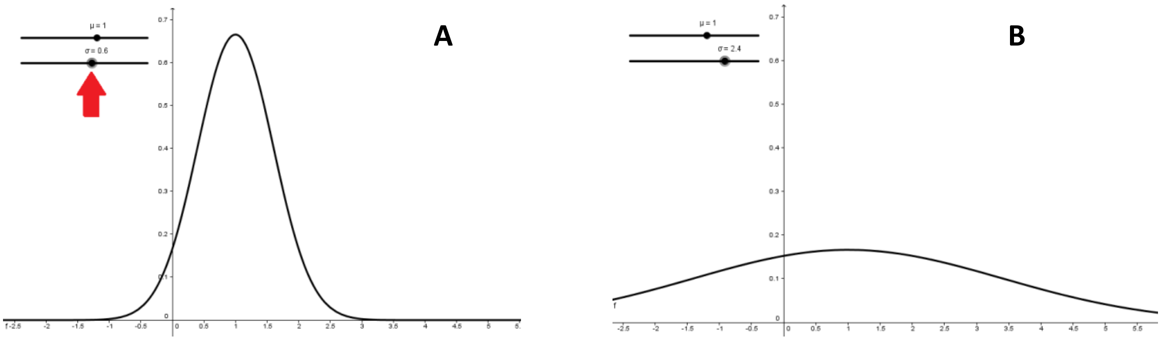
Figura 31 - Movimento de translação da Curva Normal **(A)** e **(B)**



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Movendo, agora, o controle deslizante do desvio padrão σ os alunos perceberão que haverá um alongamento ou achatamento do gráfico indicando que, quanto maior o desvio padrão, mais achatado é o gráfico, como pode ser observado na Figura 32.

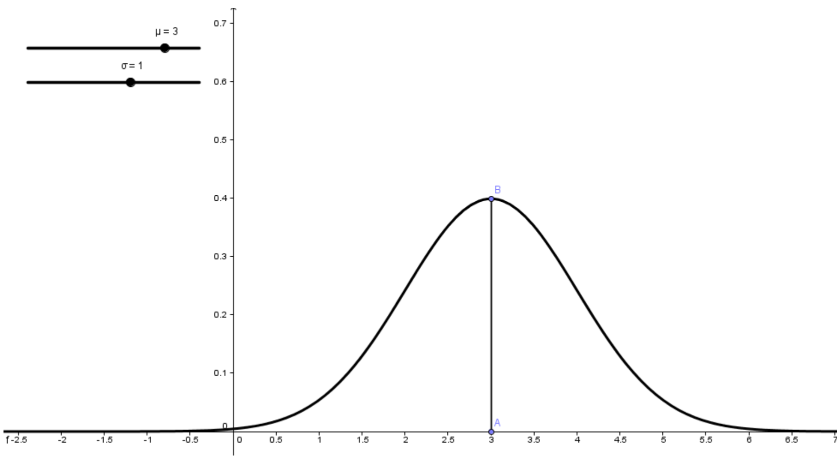
Figura 32 - Alongamento **(A)** e achatamento **(B)** da Curva Normal



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Outro fato que deve ser percebido pelos alunos é que a curva é simétrica em relação ao centro, ou seja, à média μ (Figura 33).

Figura 33 - Simetria da curva normal



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Para indicar que uma variável tem Distribuição Normal com média μ e desvio padrão σ usamos a notação $x \sim N(\mu, \sigma)$.

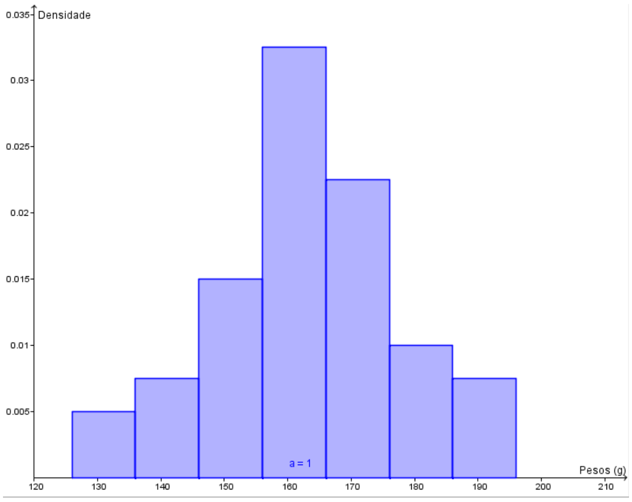
Uma questão que surge agora é:

Como saber se o Modelo Normal descreve o comportamento dos pesos da população de laranjas?

Os testes para verificar se uma Distribuição é Normal não cabem para este nível de ensino. O professor pode fazer essa verificação de modo intuitivo, através da comparação do modelo normal com o histograma da distribuição dos pesos.

Voltemos ao histograma de nossa distribuição no Geogebra:

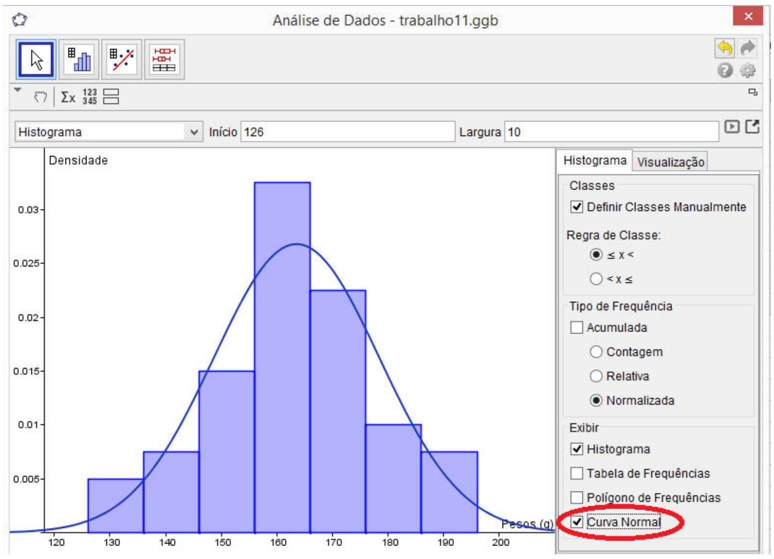
Figura 34 - Histograma da variável peso da laranja



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

Selecionamos novamente os 40 pesos e selecionamos a janela *análise univariada* na *Barra de Ferramentas,* teclamos analisar e abrirá a caixa da análise de dados. Abrimos as opções e selecionamos a caixa Curva Normal conforme é mostrado na Figura 35.

Figura 35 - Curva Norma sobre o histograma da variável peso

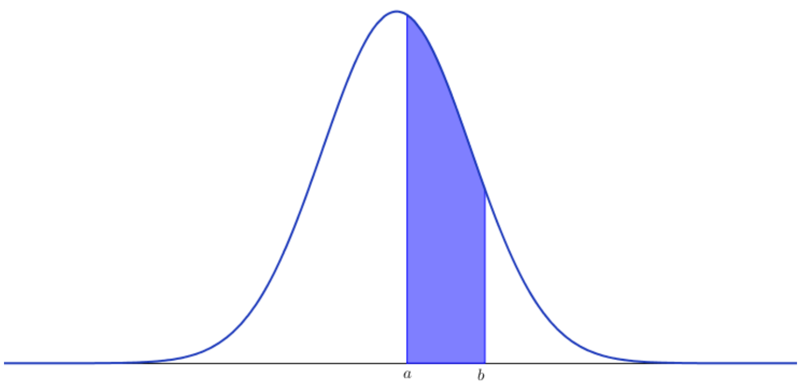


Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Os alunos poderão perceber que há uma semelhança entre o histograma e a Curva Normal. Assim, podemos usar esse modelo para descrever nossa população.

O professor deve lembrar aos alunos que a área de cada retângulo no histograma corresponde à frequência relativa da respectiva classe, que, por sua vez corresponde à probabilidade de ocorrer um valor na referida classe. Os alunos irão perceber no gráfico da Figura 35 que a área sob a curva em um determinado intervalo também é uma probabilidade. Como a Distribuição Normal será o modelo usado para descrever nossa população, podemos calcular probabilidades a partir de áreas sob a curva em relação a um certo intervalo [a,b] (Gráfico 9).

Gráfico 9 - Área sob a Curva Normal no intervalo $[\boldsymbol{a}, \boldsymbol{b}]$



Fonte: Construído pelo autor com o auxílio do *software* Geogebra.

3.2.15 – Padronização da variável peso

O cálculo dessa área usa ferramentas matemáticas vistas no Ensino Superior, e mesmo neste nível de ensino não se calcula diretamente esta área. O valor dessa área é obtido por tabelas. Como existem várias formas de distribuições normais, uma tabela para cada situação seria inviável. Esse problema é resolvido usando uma técnica chamada *padronização*, que consiste em transformar uma variável x, com média μ e desvio padrão σ, em outra variável z com média zero e desvio padrão um. Os valores das probabilidades dessa nova variável z podem ser consultados na tabela da Distribuição Normal padrão.

Para padronizar uma variável x, devemos subtrair de cada valor de x a média populacional μ e dividir o resultado pelo desvio padrão populacional σ, ou seja,

$$z = \frac{x - \mu}{\sigma} \tag{13}$$

Podemos observar que a padronização sugere que a média e o desvio padrão populacionais sejam conhecidos.

Assim, o professor pode comentar as dificuldades de se calcular essa área diretamente e introduzir o uso da tabela normal padrão.

Para isso, o professor pode propor a seguinte situação para os alunos:

Suponhamos que o peso de uma laranja se distribua normalmente, com média μ=163,55 e desvio padrão σ=15,06. Qual a probabilidade de uma laranja retirada aleatoriamente pesar entre 164,5g e 178,4g?

O professor deve deixar bem claro que esta situação supõe que os valores dos parâmetros populacionais são iguais aos valores calculados na amostra.

Para calcular a probabilidade, devemos primeiramente padronizar os valores dos pesos:

$$z_1 = \frac{164{,}5g - 163{,}55g}{15{,}06g} = 0{,}06 \qquad z_2 = \frac{178{,}4g - 163{,}55g}{15{,}06g} = 0{,}99 \tag{14}$$

Como o numerador e o denominador têm a mesma unidade, os alunos perceberão que z é adimensional, ou seja, não tem unidade.

Agora, o pergunta não se refere mais aos pesos das laranjas, e sim à nova variável z. Assim, o problema agora é: qual a probabilidade de z ser encontrado no intervalo [*0,06*; *0,99*]?

Para obter essa probabilidade, o professor deve auxiliar os estudantes a usar a tabela da Distribuição Normal padrão. Com a tabela da Distribuição Normal padrão de posse dos alunos, o professor deve comentar que a primeira coluna da tabela corresponde à parte inteira e a primeira decimal de z, e a primeira linha corresponde a segunda decimal de z. Na intersecção dessa linha e coluna, encontra-se a probabilidade de z (Tabela 7). Fazendo esse procedimento com os valores, $z_1 = 0{,}063\ e\ z_2 = 0{,}986$, temos:

Tabela 7 - Probabilidades de ocorrer $\mathbf{0 \leq Z \leq z_1}$ **(A)** e ocorrer $\mathbf{0 \leq Z \leq z_2}$ **(B)**

Distribuição Normal: Valores de p tais que $P(0 \leq Z \leq z_c) = p$

Segunda decimal de z_c

	0	1	2	3	4	5	6	7	8	9	
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359	0,0
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753	0,1
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141	0,2
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517	0,3
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879	0,4
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224	0,5
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549	0,6

	0	1	2	3	4	5	6	7	8	9	
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359	0,0
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753	0,1
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141	0,2
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517	0,3
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879	0,4
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224	0,5
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549	0,6
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852	0,7
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133	0,8
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389	0,9

Fonte: Magalhães e Lima (2004, p. 355)

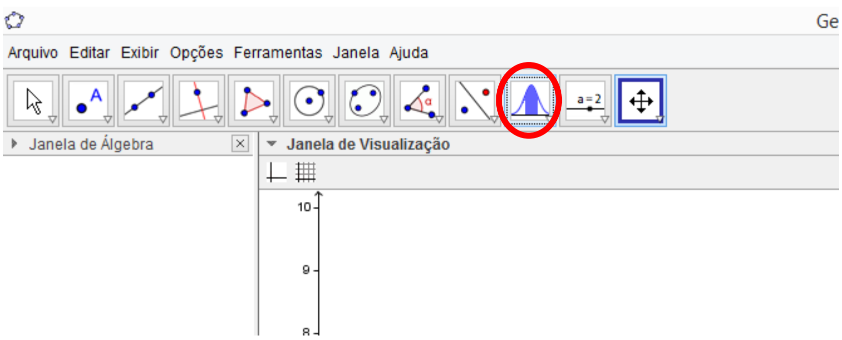
Os valores que encontramos para $z_1 = 0{,}06$ e $z_2 = 0{,}99$, refere-se à probabilidade de z cair nos intervalos, $0 \leq Z \leq z_1$ e $0 \leq Z \leq z_2$, respectivamente.

Assim,

$$P(0{,}06 \leq Z \leq 0{,}99) = P(0 \leq Z \leq 0{,}99) - P(0 \leq Z \leq 0{,}06) = \qquad (15)$$
$$= 0{,}3389 - 0{,}0239 = 0{,}315$$

Usando o Geogebra podemos obter essa probabilidade, na *Barra de Ferramentas* selecionamos calculadora de probabilidades como pode ser visto na Figura 36.

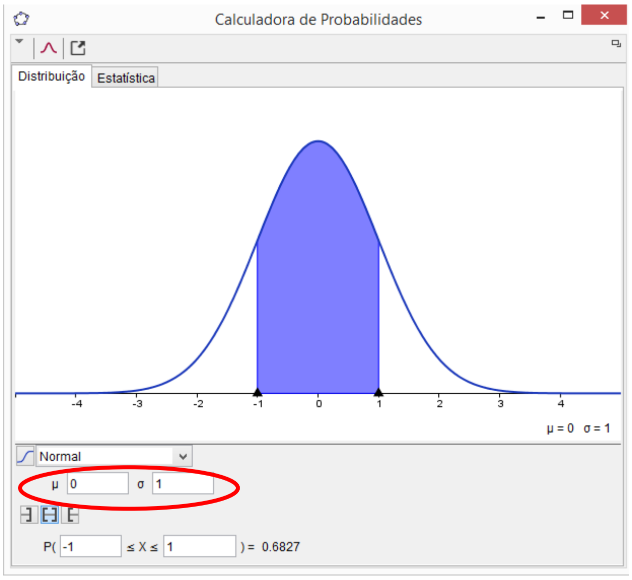
Figura 36 - Selecionando calculadora de probabilidades



Fonte: Geogebra

Abrirá uma nova janela com a calculadora (Figura 37).

Figura 37 - Calculadora de probabilidades



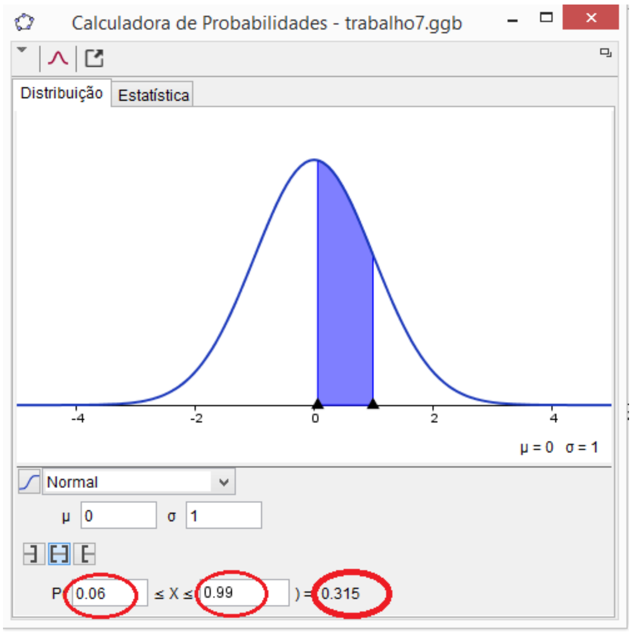
Fonte: Geogebra

Os alunos visualizaram que a média $\mu = 0$ e desvio padrão é $\sigma = 1$, isto é, a Distribuição Normal é padrão.

Um fato importante que o professor pode chamar a atenção é que z, indicado no eixo horizontal, é medido em unidades de desvios padrão.

Colocando os valores padronizados nos locais indicados e teclando *enter*, obtemos a probabilidade conforme Figura 38.

Figura 38 - Probabilidade de z ocorrer no intervalo $[\mathbf{0,06;0,99}]$



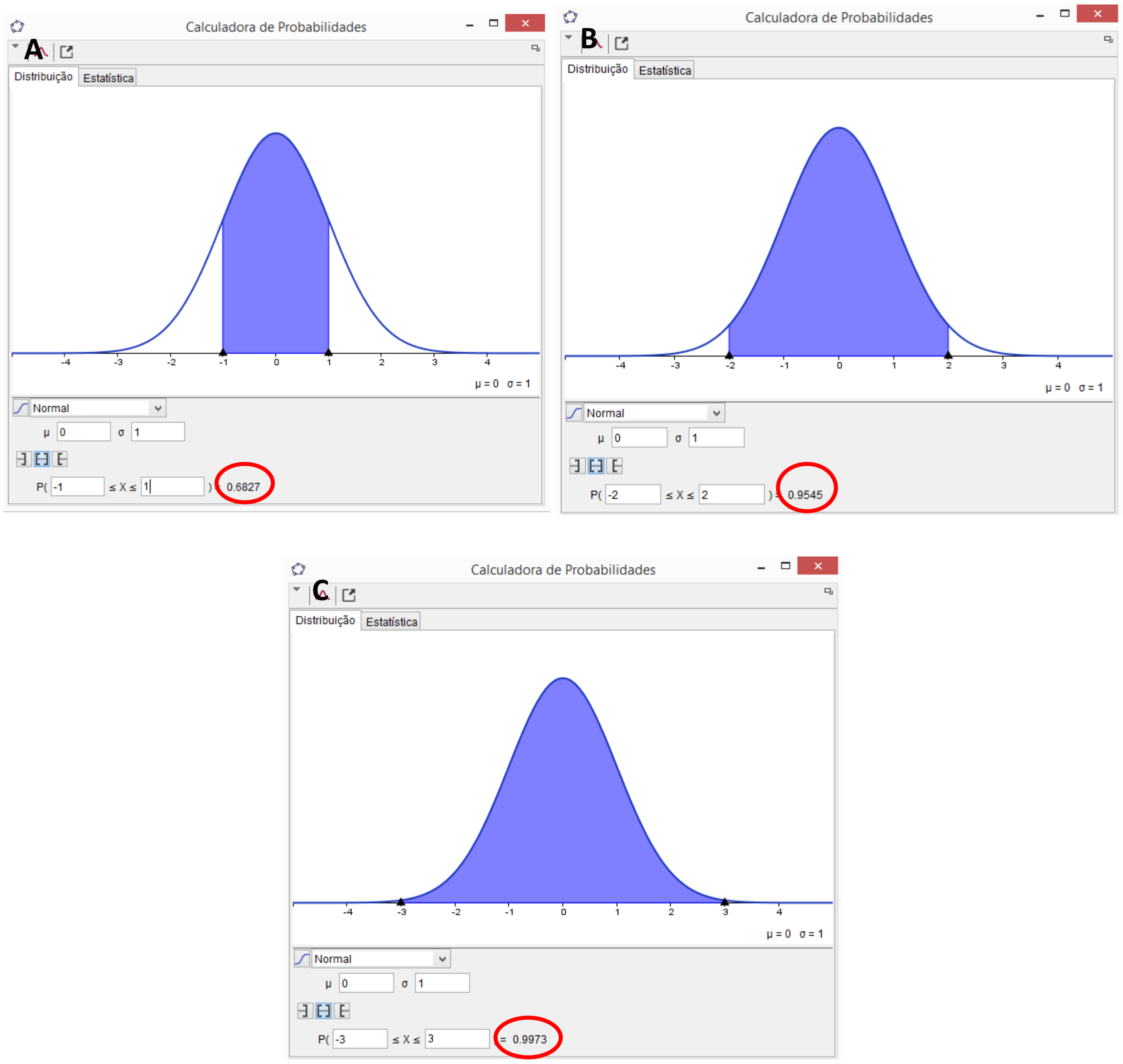
Fonte: Construída pelo autor com o auxílio do *software* Geogebra

Assim, obtemos a resposta do problema, que a probabilidade de uma laranja retirada aleatoriamente pesar entre 164,5g e 178,4g é de 0,3128 ou 31,38% de chance. Os alunos devem observar também que essa probabilidade corresponde à área situada entre os limites indicados.

Essa probabilidade que encontramos se baseia na validade da suposição de que os parâmetros populacionais são iguais aos valores amostrais.

Como foi visto na tabela de probabilidades, o eixo horizontal é medido em desvios padrão. Usando a calculadora, vamos calcular a probabilidades de z nos intervalos definidos por um, dois e três desvios padrão em torno da média (Figura 39).

Figura 39 - Probabilidades de z cair no intervalo de um desvio padrão **(A),** dois desvios padrão **(B)** e três desvios padrão **(C)** ao redor da média 0



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

O professor deve chamar a atenção dos alunos para que percebam que com apenas um desvio padrão em torno da média, temos 68,27% de ocorrer um valor de z neste intervalo e com 3 desvios padrões em torno da média, temos 99,73%, quase 100% de chance.

3.2.16 Intervalo com 95% de confiança de conter z

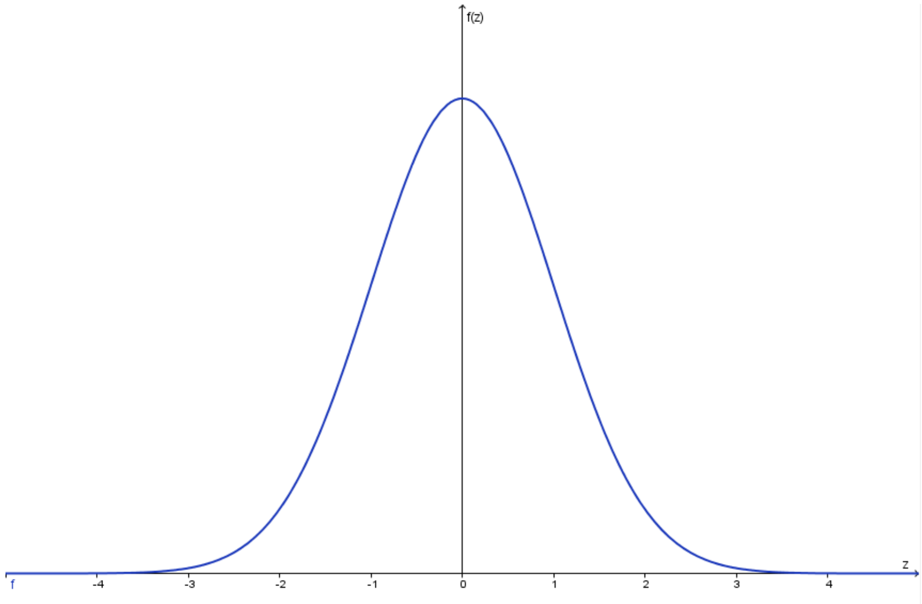
O professor agora pode resolver o seguinte exercício com a turma:

Determinar os limites de um intervalo com 95% de confiança, ou seja, de probabilidade de conter z.

Assim, temos que encontrar $-z_c$ e z_c tal que $P(-z_c < Z < z_c) = 95\%$.

Para obter os limites desse intervalo o professor deve orientar os alunos observar uma propriedade importante da Curva Normal.

Gráfico 10 - Simetria da Curva Normal $\boldsymbol{N(0,1)}$

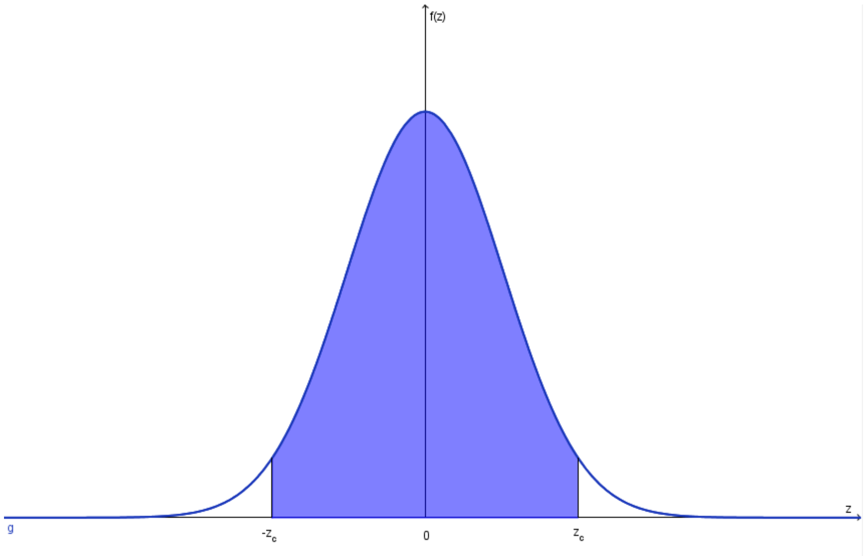


Fonte: Construído pelo autor com o auxílio do *software* Geogebra.

Devido ao fato de a Curva Normal ser simétrica em relação à média 0, como pode ser percebido no Gráfico 10, podemos distribuir a área, que corresponde a 95% de chance, igualmente em torno da média 0, ou seja, metade da área à esquerda e a outra metade à direita.

Assim, obtemos a seguinte situação visualizada no Gráfico 11.

Gráfico 11 - Distribuição da área correspondente a 95% de chance igualmente em torno da média 0



Fonte: Construído pelo autor com o auxílio do *software* Geogebra.

Agora, de volta à tabela, o professor orienta os alunos a fazer o caminho inverso do que foi feito para encontrar probabilidades. Procuramos no corpo da tabela o valor 0,475, que corresponde à metade de 0,95, e em seguida observamos a linha e a coluna correspondentes a esse valor para encontrar z (Tabela 8).

Tabela 8 - Valor de z correspondente a 0,475 de probabilidade

Distribuição Normal: Valores de p tais que $P(0 \leq Z \leq z_c) = p$

Segunda decimal de z_c

	0	1	2	3	4	5	6	7	8	9	
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359	0,0
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753	0,1
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141	0,2
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517	0,3
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879	0,4
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224	0,5
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549	0,6
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852	0,7
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133	0,8
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389	0,9
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621	1,0
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830	1,1
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015	1,2
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177	1,3
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319	1,4
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441	1,5
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545	1,6
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633	1,7
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706	1,8
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767	1,9
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817	2,0
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857	2,1
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890	2,2

Fonte: Magalhães e Lima (2004, p. 356)

Assim, o valor de *z* que corresponde a 0,475 é $z = 1{,}96$

Logo, o intervalo com 95% de confiança de *z* ser encontrado é $-1{,}96 < Z < 1{,}96$

Após essa discussão sobre o Modelo Normal, o professor deve retomar o problema de estimar o peso médio populacional, ou seja, estimar o verdadeiro peso médio do conjunto de todas as laranjas contidas em 40kg.

A média amostral é uma boa aproximação para a média populacional, mas como vimos, ela pode variar de amostra para amostra. Essa variação pode ser tanto para mais ou para menos. A média amostral é uma estimativa pontual, pois fornece um único valor numérico como estimativa para a média populacional.

Dessa forma, o professor pode discutir com os alunos sobre a necessidade de uma estimativa mais precisa para a média populacional, que no caso das laranjas, é o peso médio populacional.

Nesse momento, o professor deve introduzir a ferramenta dada pela Estatística inferencial para se obter uma estimativa mais precisa, que é o intervalo de confiança.

O intervalo de confiança incorpora a estimativa pontual do parâmetro de interesse, informações sobre sua variabilidade e fornece uma medida de precisão do intervalo encontrado (MAGALHÃES, LIMA, 2004). Assim, o intervalo de confiança fornecerá uma estimativa mais informativa sobre o peso médio populacional.

3.2.17 Intervalo de confiança para a média populacional a partir de uma observação

Para exemplificar a construção de um intervalo de confiança para o peso médio populacional, o professor pode fazer o seguinte procedimento com os alunos:

Suponha que tenha sido pesada apenas uma laranja, escolhida aleatoriamente, e que o peso encontrado tenha sido 158,0g, o primeiro valor da Tabela 4 .

O que esse valor nos permite dizer a respeito do peso médio populacional, μ?

Em um exercício feito anteriormente pelos alunos, foi calculado que o intervalo $-1,96 < Z < 1,96$ tem 95% de chance de conter z.

Como z é o resultado da padronização da variável x, que é o peso da laranja, temos:

$$-1,96 < \frac{x-\mu}{\sigma} < 1,96 \Rightarrow -1,96 < \frac{158,0g-\mu}{\sigma} < 1,96 \qquad (16)$$

Multiplicando todos os membros por σ e em seguida somando μ

$$-1{,}96\sigma < 158{,}0g - \mu < 1{,}96\sigma \Rightarrow \mu - 1{,}96\sigma < 158{,}0g < \mu - 1{,}96\sigma \qquad (17)$$

Significa que a observação avulsa, 158,0g, tem 95% de chance de ter sido feita dentro desse intervalo. O professor deve chamar a atenção dos alunos para o fato de que temos 5% de chances de o peso cair fora desse intervalo.

Assim, podemos dizer que temos 95% de confiança na dupla desigualdade:

$$\mu - 1{,}96\sigma < 158{,}0g < \mu - 1{,}96\sigma$$

Adicionando $-\mu$ a todos os membros dessa desigualdade, temos:

$$-1{,}96\sigma < 158{,}0g - \mu < 1{,}96\sigma$$

Subtraindo 158,0g a todos os membros e multiplicando por -1, obtemos:

$$158{,}0g - 1{,}96\sigma < \mu < 158{,}0g + 1{,}96\sigma \qquad (18)$$

Assim, temos um intervalo com 95% de confiança para o peso médio populacional. Deve-se deixar claro para os estudantes que 95% de confiança é a chance de o intervalo conter a verdadeira média populacional e não a probabilidade de a média populacional ser um valor entre os dois limites. O peso médio populacional é um valor bem determinado, para conhecê-lo basta pesar todas as laranjas contidas em 40kg.

Para determinar numericamente os limites desse intervalo, precisamos do valor de σ, que é o desvio padrão populacional. O que temos é o desvio padrão amostral, então a princípio não podemos substituir σ por $s = 15{,}06g$.

Suponhamos que o desvio padrão amostral coincida com o desvio padrão populacional. Fazendo a substituição na dupla desigualdade, temos:

$$158{,}0g - 1{,}96(15{,}06g) < \mu < 158{,}0g + 1{,}96(15{,}06g) \Rightarrow \qquad (18)$$
$$\Rightarrow 128{,}48 < \mu < 187{,}52$$

Podemos dizer com 95% de confiança que o intervalo encontrado contém o verdadeiro peso médio, isto é, de 100 intervalos construídos com a mesma metodologia, 95 deles conteriam a verdadeira média.

A partir dos valores dos limites, e com a suposição feita sobre σ, podemos dizer que o número total de laranjas contidas em 40kg está entre 213 e 311, que são os valores encontrados pela divisão da massa de 40000g de laranjas pelos limites do intervalo (19).

E lembrando aos alunos que ainda há 5% de chance de estarmos errados.

Como esse intervalo de confiança foi obtido a partir de uma única observação, sua precisão não é muito boa. Para obtermos uma estimativa mais precisa para o peso médio populacional, devemos construir o intervalo de confiança a partir de médias.

Isso é natural, pois os valores médios são representações mais adequadas da população como um todo (NETO, SCARMINIO, BRUNS, 2003, p. 37).

Nesse momento os alunos podem sugerir ao professor usar a média amostral na dupla desigualdade (16) para obter o intervalo de confiança. O professor dever fazer uma importante observação:

O intervalo de confiança foi construído aceitando o Modelo Normal para descrever a distribuição da variável peso. E no caso da média amostral? Podemos usar o Modelo Normal para descrevê-la?

Assim como foi feito para a variável peso, devemos construir seu histograma e observar sua características para decidir se a distribuição da média amostral pode ser descrita pelo Modelo Normal. Mas para usar esse procedimento temos que coletar várias amostras de tamanho 40 para obtermos o histograma.

3.2.18 Teorema do Limite Central

Para evitar todo esse trabalho, a Estatística Inferencial conta com um teorema muito importante, o Teorema do Limite Central. O Teorema do Limite Central nos garante que a distribuição da média amostral se aproxima do Modelo Normal à medida que o tamanho da amostra aumenta independente da distribuição da variável em questão.

Assim, para uma variável com distribuição dada por um modelo qualquer com média μ e desvio padrão σ, temos:

$$\bar{x} \sim N(\mu_x, \sigma_x) \tag{20}$$

Pelas propriedades da média e da variância, vem:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{21}$$

Logo, para uma variável com distribuição de probabilidade dada por um modelo qualquer com média (μ) e desvio padrão (σ), a média amostral $\bar{x}$ terá (aproximadamente) Distribuição Normal, com mesma média μ e desvio padrão $\sigma/\sqrt{n}$.

Esse teorema garante que a distribuição do peso médio amostral se aproxime da Distribuição Normal. Quanto maior o tamanho da amostra, melhor é a aproximação.

Para ilustrar o Teorema do Limite Central o professor pode fazer uma simulação usando uma tabela de números aleatórios (Tabela 9). Essa tabela é construída de modo que os dez algarismos, de 0 a 9, são distribuídos ao acaso nas linhas e colunas (CRESPO, 2006, p. 20). Assim, a probabilidade de um algarismo ocorrer numa certa linha e coluna é a mesmo para todos. Dessa forma, podemos usar uma coleção de números dessa tabela como uma amostra, da mesma forma que foi feito com a seleção de uma amostra de laranjas, em ambos os casos, cada elemento tem a mesma probabilidade de está na amostra.

Tabela 9 - Tabela de números aleatórios

57720039848441796771402113975649865408932968745483
28805351590993988758702771771706320278621674696517
92591852873048869748352518887403629838586586424103
90381291743019758907506415597188137495305278301175
80911694675860820666904756184645111235324550411343
22017031329691927540165429727499009597610098243007
56241004302046299053531105844121647919762951626066
79449262029686643000945669302059878735442250977819
53996645088978507753372577412762380223576201416035
18928735885505213651392850146685793019797266643145
53085896630561257022504128966266436306630132798522
03588029287689511824888946474859192987031033996712
27078188656949980028047051300147189733218582454324
05210859010622249891811755446616077307661012317858
40361327843082333639694205586461123389278952667193
54602528858820001059610536613372010119016110512091
71516340767111737352373160458892734371280498090248
61020181739260667358533442682638340327449604466593
82559313463095265506961765917239799612495280632699
89985414217413576819862860894733152628774538480808
00998484146795137758901450794273633106604340125504
62415078204805884352980319939203049725849595036331
94279069246809921186076383193299511555710927026700
44892928843628251582877418972576106326760226745328
97307695332110542695666552049936584803089363581796
39165804448015595983909554668184396085388866333569
60781103266750340961313020769366308351093383647605
03192347628957779133884760593754394877674985384391
41285267562539599665513690322239330522990339979699
77549850392537425297100356049281668670014889558210
28634161916424838137344883279638716973067750256460
74244885401233596750149814264279791352896978804471
00240337964668750532421663332897263647277365383446
05414769694536167118955197220413239658600369487983
62698497974723665156130869115275592686818043009892

Fonte: Crespo (2006, p. 223)

Para fazer a simulação, selecionaremos várias amostras de números aleatórios dessa tabela. Numa primeira etapa selecionaremos 37 amostras de 10 elementos representadas pelas 37 primeiras colunas da tabela entre a primeira e quarta linha e da sexta a décima primeira conforme Tabela 10.

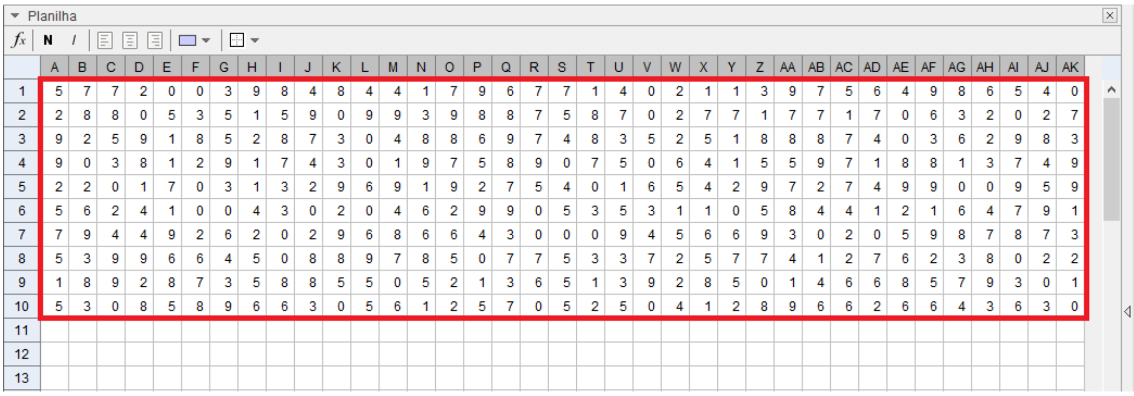
Tabela 10 - Seleção de 37 amostras de 10 números aleatórios

57720039848441796771402113975649865408932968745483
28805351590993988758702771771706320278621674696517
92591852873048869748352518887403629838586586424103
90381291743019758907506415597188137495305278301175
80911694675860820666904756184645111235324550411343
22017031329691927540165429727499009597610098243007
56241004302046299053531105844121647919762951626066
79449262029686643000945669302059878735442250977819
53996645088978507753372577412762380223576201416035
18928735885505213651392850146685793019797266643145
53085896630561257022504128966266436306630132798522
03588029287689511824888946474859192987031033996712

Fonte: Crespo (2006, p. 223)

Na sala de informática, o professor pode orientar aos grupos a transpor esses números para a planilha do Geogebra, de forma que cada coluna represente uma amostra, como pode ser visto na Figura 40.

Figura 40 - 37 amostra de 10 números aleatório inseridas no Geogebra

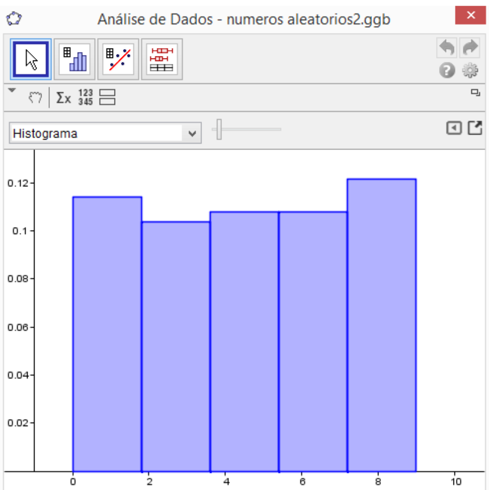


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
1	5	7	7	2	0	0	3	9	8	4	8	4	4	1	7	9	6	7	7	1	4	0	2	1	1	3	9	7	5	6	4	9	8	6	5	4	0
2	2	8	8	0	5	3	5	1	5	9	0	9	9	3	9	8	8	7	5	8	7	0	2	7	7	1	7	7	1	7	0	6	3	2	0	2	7
3	9	2	5	9	1	8	5	2	8	7	3	0	4	8	8	6	9	7	4	8	3	5	2	5	1	8	8	8	7	4	0	3	6	2	9	8	3
4	9	0	3	8	1	2	9	1	7	4	3	0	1	9	7	5	8	9	0	7	5	0	6	4	1	5	5	9	7	1	8	8	1	3	7	4	9
5	2	2	0	1	7	0	3	1	3	2	9	6	9	1	9	2	7	5	4	0	1	6	5	4	2	9	7	2	7	4	9	9	0	0	9	5	9
6	5	6	2	4	1	0	0	4	3	0	2	0	4	6	2	9	9	0	5	3	5	3	1	1	0	5	8	4	4	1	2	1	6	4	7	9	1
7	7	9	4	4	9	2	6	2	0	2	9	6	8	6	6	4	3	0	0	0	9	4	5	6	6	9	3	0	2	0	5	9	8	7	8	7	3
8	5	3	9	9	6	6	4	5	0	8	8	9	7	8	5	0	7	7	5	3	3	7	2	5	7	7	4	1	2	7	6	2	3	8	0	2	2
9	1	8	9	2	8	7	3	5	8	8	5	5	0	5	2	1	3	6	5	1	3	9	2	8	5	0	1	4	6	6	8	5	7	9	3	0	1
10	5	3	0	8	5	8	9	6	6	3	0	5	6	1	2	5	7	0	5	2	5	0	4	1	2	8	9	6	6	2	6	6	4	3	6	3	0
11																																					
12																																					
13																																					

Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

O professor deve lembrar que cada coluna representa uma amostra de 10 elementos. Selecionando esses valores e clicando no comando "análise univariada" na *Barra de Ferramentas* obtemos o histograma desses números aleatórios conforme Figura 41.

Figura 41 - Histograma dos números aleatórios contidos nas 37 amostras de 10 números aleatórios



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

O histograma mostra uma tendência de uniformidade de ocorrência dos valores da variável e confirma a hipótese de que a probabilidade de ocorrência de cada valor é a mesma. O professor deve chamar a atenção dos alunos para este fato.

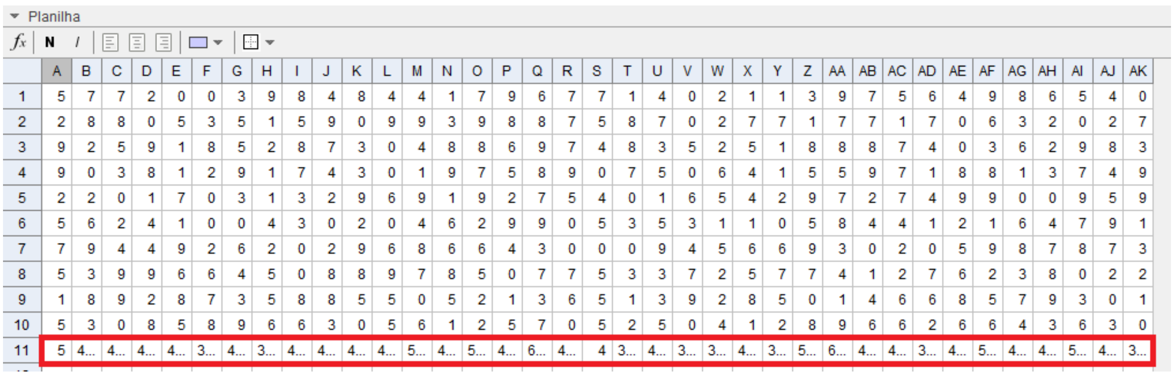
Para calcular a média dessas amostras, selecionamos todas as colunas e abrimos o comando *média* na barra de ferramentas do Geogebra (Figura 42).

Figura 42 - Opção calcular média na *Barra de Ferramentas*

Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Dessa forma, obtemos a média de cada amostra, como pode ser visto na Figura 43.

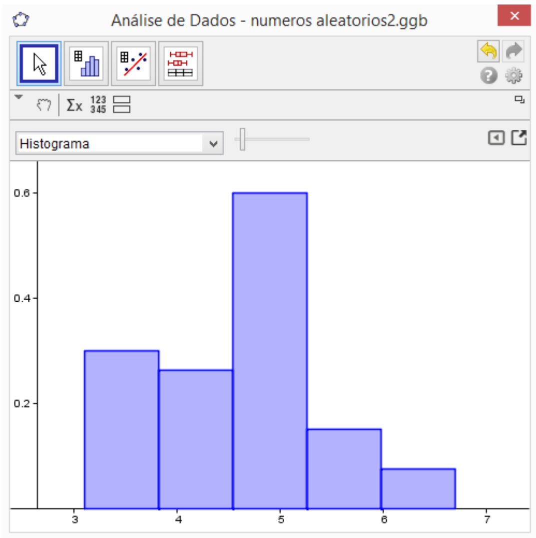
Figura 43 - Médias das 37 amostras de 10 números aleatórios



Fonte: Construída pelo autor com o auxílio do *software* Geogebra

Selecionando a linha de médias e construindo o histograma, temos o histograma da Figura 44.

Figura 44 - Histograma das médias das 37 amostras de 10 números aleatórios



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Na segunda etapa, vamos selecionar 37 amostras de 20 elementos cada (Tabela 11).

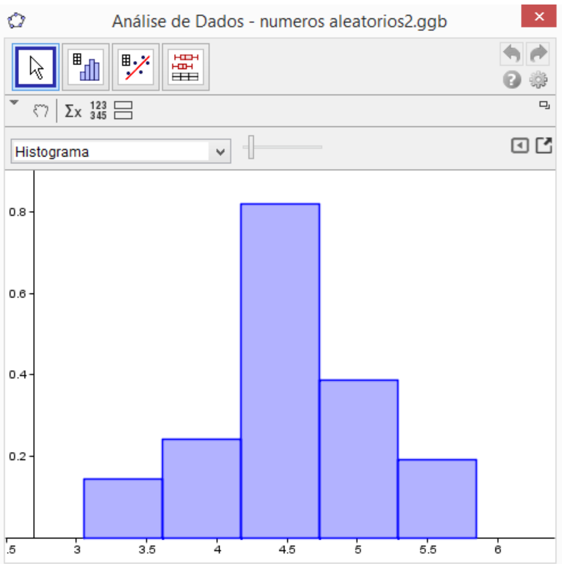
Tabela 11 - Seleção de 37 amostras de 20 números aleatórios

57720039848441796771402113975649865408932968745483
28805351590993988758702771771706320278621674696517
92591852873048869748352518887403629838586586424103
90381291743019758907506415597188137495305278301175
80911694675860820666904756184645111235324550411343
22017031329691927540165429727499009597610098243007
56241004302046299053531105844121647919762951626066
79449262029686643000945669302059878735442250977819
53996645088978507753372577412762380228576201416035
18928735885505213651392850146685793019797266643145
53085896630561257022504128966266436306630132798522
03588029287689511824888946474859192987031033996712
27078188656949980028047051300147189738218582454324
05210859010622249891811755446616077307661012317858
40361327843082333639694205586461123389278952667193
54602528858820001059610536613372010119016110512091
71516340767111737352373160458892734371280498090248
61020181739260667358533442682638340327449604466593
82559313463095265506961765917239799612495280632699
89985414217413576819862860894733152623774538480808
00998484146795137758901450794273633106604340125504
62415078204805884352980319939203049725849595036331
94279069246809921186076383193299511555710927026700

Fonte: Crespo (2006, p. 223)

Fazendo o mesmo procedimento, obtemos o seguinte histograma da Figura 45.

Figura 45 - Histograma das médias de 37 amostras de 20 números aleatórios



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Com a observação dos histogramas das médias amostrais (Figuras 44 e 45), os alunos perceberão o Teorema do Limite Central em ação. À medida que o tamanho da amostra aumenta, a distribuição da média amostral se aproxima cada vez mais do Modelo Normal, independente da distribuição da variável, que neste caso, não tinha Distribuição Normal.

Podemos concluir que a distribuição do peso médio amostral se aproxima do Modelo Normal. Padronizando o peso médio amostral, temos:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \tag{22}$$

3.2.19 Intervalo de confiança para a média populacional a partir da Distribuição Normal

Construindo um intervalo de confiança para a média populacional μ com 95% de confiança, obtemos:

$$-1{,}96 < Z < 1{,}96 \Rightarrow -1{,}96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < -1{,}96$$

Multiplicando todos os membros por $\sigma\ /\ \sqrt{n}$ e em seguida adicionando $-\bar{x}$, temos:

$$\bar{x} - \frac{1{,}96\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{1{,}96\sigma}{\sqrt{n}} \tag{23}$$

Para determinar numericamente os limites desse intervalo, devemos conhecer os valores da média amostral e do desvio padrão populacional. A média amostral da nossa amostra de laranjas é $\bar{x} = 163{,}55$. Mas não conhecemos o desvio padrão do conjunto de todas as laranjas contidas em 40kg, ou seja, o desvio padrão populacional. Neste caso o desvio padrão deve ser estimado pelo desvio padrão amostral. Dessa forma, obtemos uma nova variável com a padronização.

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{24}$$

Não podemos afirmar que o peso médio padronizado dessa forma tem Distribuição Normal, pois trocamos o parâmetro σ, que é uma quantidade bem determinada, por uma variável aleatória s. A nova padronização tem outra distribuição de probabilidade, denominada Distribuição $t - Student$.

Nesse momento, o professor deve comentar a consequência da substituição de σ por s, e chamar atenção que na realidade essa substituição não vai alterar nossos cálculos, pois a Distribuição de $Student$ se aproxima do Modelo Normal à medida que o tamanho da amostra aumenta. A distribuição de $Student$ é usada na prática para uma amostra de tamanho até 30 elementos. Para amostras maiores o intervalo (22) é satisfatório (NETO, SCARMINIO, BRUNS, 2003, p. 53)

Em nosso caso, com uma amostra de 40 laranjas, podemos usar sem problemas o intervalo (22).

Finalmente, temos condições de responder qual o número aproximado de laranjas contidas em 40kg da fruta.

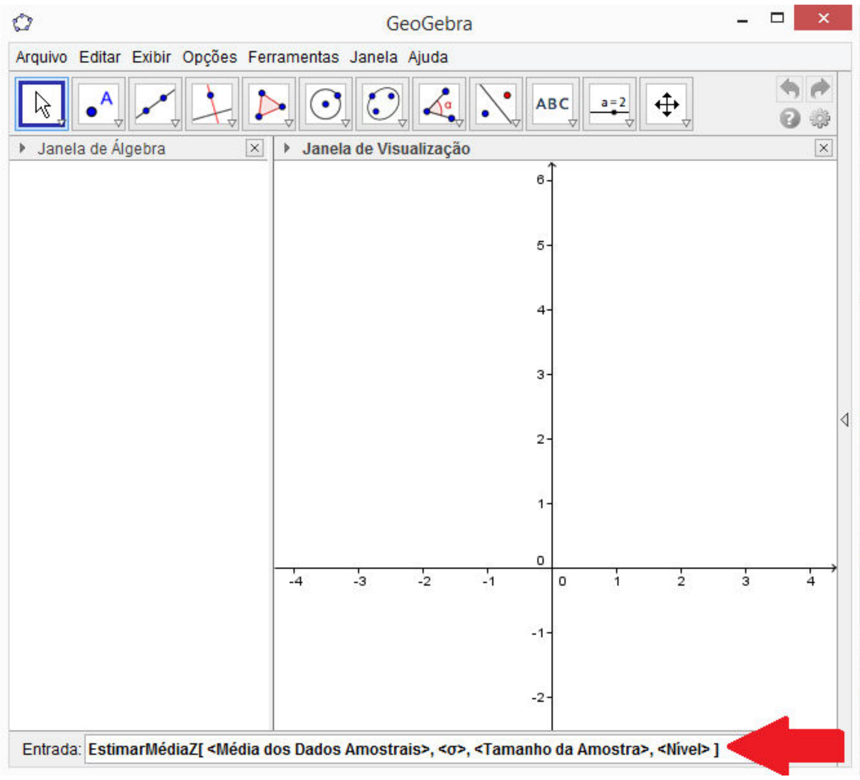
O professor orienta os alunos a calcular um intervalo com 95% de confiança para o peso médio de todas as laranjas contidas em 40kg. Como o peso médio amostral é $\bar{x} = 163{,}55$g, o desvio padrão amostral é $s = 15{,}06g$ e tamanho da amostra é $n = 40$, temos:

$$\bar{x} - \frac{1{,}96\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{1{,}96\sigma}{\sqrt{n}} \Rightarrow 163{,}55 - \frac{1{,}9615{,}06}{\sqrt{40}} < \mu < 163{,}55 + \frac{1{,}9615{,}06}{\sqrt{40}} \Rightarrow 158{,}88 < \mu < 168{,}22 \quad (25)$$

Após os estudantes se familiarizarem com o processo para se obter um intervalo de confiança, o professor deve orientá-los a utilizar o Geogebra para calcular o intervalo de confiança.

Na entrada de comando do Geogebra digitamos $EstimarMédiaZ[< Média\ dos\ Dados\ Amostrais, < \sigma >, < Tamanho\ da\ Amostra >, < Nível >]$ conforme Figura 46.

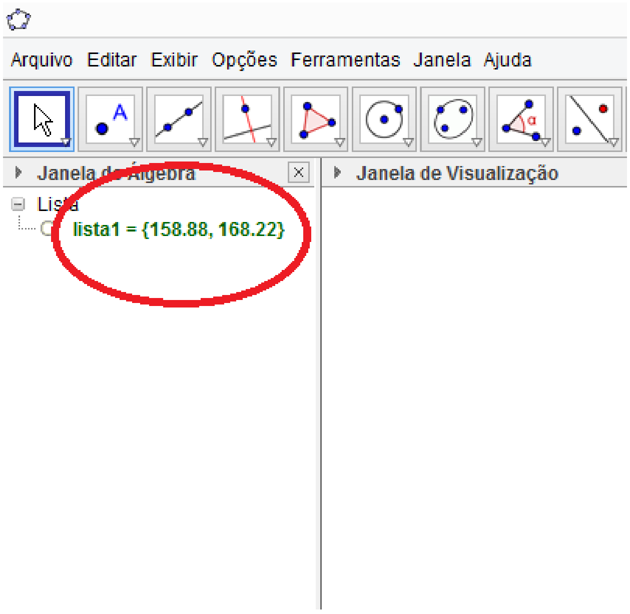
Figura 46 - Comando para calcular intervalo de confiança para média populacional



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Nos campos correspondentes, inserimos o peso médio amostral, o desvio padrão amostral, o tamanho da amostra e o nível de confiança. Teclamos *enter* para obter o intervalo de confiança (Figura 47).

Figura 47 - Intervalo de confiança para o peso médio populacional



Fonte: Construída pelo autor com o auxílio do *software* Geogebra.

Assim, podemos dizer com 95% de confiança que o peso médio de todas as laranjas contidas em 40kg pode ser encontrado no intervalo obtido, ou seja, de 100 intervalos de confiança construídos, 95 deles conteriam o verdadeiro peso médio populacional.

3.3 CONCLUSÃO DA PROPOSTA

3.3.1 Quantidade aproximada de laranjas contidas em 40kg da fruta

Com os limites do intervalo de confiança calculados, o professor faz o questionamento fundamental, que polarizou todo trabalho até aqui:

Qual a quantidade aproximada de laranjas contidas em 40kg da fruta?

Dividindo 40000g pelos limites do intervalo de confiança, temos:

$$\frac{40000g}{158{,}88g} = 251{,}76 \quad \text{e} \quad \frac{40000g}{168{,}22g} = 237{,}78 \tag{19}$$

Assim, chegamos à conclusão que o número de laranjas contidas em 40kg da fruta deve estar entre 238 e 252 laranjas. Podemos dizer isso com 95% de confiança. Lembrando sempre aos alunos que temos 5% de chances de estarmos errados.

3.3.2 Quantidade aproximada de caixas de um litro de suco de laranja que podem ser produzidas com 40kg

Portanto, pode-se produzir entre 48 e 50 caixas de um litro de suco de laranja com as laranjas contidas em 40kg.

Para finalizar, o professor deve discutir com os alunos sobre o uso dessas técnicas da Estatística Inferencial pelo produtor de suco de laranja.

3.3.3 Questões importantes a serem tratadas em aulas posteriores

Algumas questões que devem ser discutidas pelo professor em aulas posteriores:

O que acontece com intervalo de confiança quando aumentamos o tamanho da amostra?

O que aconteceria com o intervalo de confiança se o tamanho da amostra fosse de 100 laranjas? E se fosse de 200 laranjas?

Se aumentarmos o nível de confiança, o que acontece com o intervalo?

O que pode ser feito para diminuirmos a amplitude do intervalo de confiança, tornando o mais informativo?

Calcular um intervalo de confiança com 99,5% de confiança para o peso médio populacional e comparar com o intervalo obtido com 95% de confiança.

Todas essas questões podem ser resolvidas utilizando o Geogebra.

4 CONSIDERAÇÕES FINAIS

Ao final da proposta espera-se que o aluno compreenda que a utilidade principal da Distribuição Normal, neste trabalho, é fornecer uma estimativa mais precisa para o peso médio da população de 40 quilos de laranja, e com auxílio desta estimativa, obter o número aproximado de laranjas contidas nesses quarenta quilos.

A realização da proposta possibilitará aos alunos o contato com uma situação de inferência sobre uma população a partir de uma amostra dessa e também oferecerá através do *software* Geogebra, simulações de construções de tabelas, gráficos e cálculos estatísticos. Propiciará, ainda, o reconhecimento por parte dos alunos da eficiência no tratamento de dados obtida pela utilização do recurso computacional.

A implementação da proposta em sala de aula exigirá um número expressivo de aulas, aproximadamente 10 aulas da disciplina de matemática do Ensino Médio. Por outro lado, seriam contemplados vários conceitos estatísticos, como média, desvio padrão, probabilidades, construções de gráficos e tabelas, estimação, dentre outros. Assim, a abordagem dos conhecimentos estatísticos supracitados superaria fragmentação do conteúdo nessas aulas.

Explorar uma situação problema com dados reais, utilizando as Tecnologias da Informação e Comunicação para fazer o tratamento dos dados, poderá constituir uma alternativa eficaz para o ensino aprendizagem de conceitos estatísticos complexos, como por exemplo, a função de Distribuição Normal. Portanto, o uso do Geogebra aplicado à abordagem da Distribuição Normal oferecerá ao professor uma maneira eficiente de explorar um conteúdo complexo em turmas de Ensino Médio, devido à sua *interface* prática e da fácil manipulação do *software*.

A proposta pode ser desenvolvida na prática para determinar, por exemplo, a quantidade aproximada de laranjas em uma plantação de 10 mil pés, ou ainda, para determinar o número aproximado de peixes de uma determinada espécie em um carregamento de 5 toneladas. A resolução desses problemas pode ser tratada em estudos posteriores.

REFERÊNCIAS

AZEVEDO, I. F.; LIMA, M. F.; SILVA, W. F. A estatística no Ensino Médio: Preparando alunos do 1º e 2º anos para uma melhor visão da estatística. In: VIII EPAEM Encontro Paranaense de Educação Matemática, 2011, Belém: UNAMA. **Anais do VIII EPAEM.** Belém: SBEM-PA, 2011. Disponível em: http://www.sbempa.mat.br/Boletim/Anais/secoes/CC0127.pdf. Acesso em: 27 set. 2014.

BÍBLIA DE ESTUDO: Nova Versão Internacional [traduzida pela comissão da Sociedade Bíblica Internacional]. São Paulo: Editora Vida, 2003.

BITTENCOURT, H. R.; VIALI, L. Contribuições para o ensino da distribuição normal ou curva de Gauss em cursos de graduação. In: III Seminário Internacional de Pesquisa em Educação Matemática, 2006, Águas de Lindóia. **Anais do III Seminário Internacional de Pesquisa em Educação Matemática**. Brasília : SBEM, 2006. Disponível em: http://www.pucrs.br/famat/viali/mestrado/literatura/artigos/planilhas/Sipem_06.pdf. Acesso em: 27 set. 2014.

BOYER, Carl B. **História da Matemática.** São Paulo: Editora Edgard Blucher Ltda, 1974.

BRASIL, Ministério da Educação. Secretaria da Educação Média e Tecnológica**. PCN+ - Ensino Médio, Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais**. Brasília: MEC/SEMTEC. 2002. Disponível em: http://portal.mec.gov.br/index.php?option=com_content&id=12598%3Apublicacoes&Itemid=859. Acesso em: 27 set. 2014.

______, Ministério da Educação. Secretaria da Educação Média e Tecnológica. **Parâmetros Curriculares Nacionais: ensino médio: ciências da natureza, matemáticas e suas tecnologias**. Brasília: MEC/SEMTEC. 2000. Disponível em: http://portal.mec.gov.br/index.php?option=com_content&id=12598%3Apublicacoes&Itemid=859. Acesso em: 27 set. 2014.

______, Ministério da Educação. Secretaria de Educação Fundamental. **Parâmetros curriculares nacionais: matemática: 1º e 2º ciclos**. Brasília, DF: MEC/SEF, 1997. Disponível em: http://portal.mec.gov.br/seb/arquivos/pdf/livro03.pdf. Acesso em: 27 set. 2014.

______, Decreto nº 6.300, de 12 de dezembro de 2007. **Dispõe sobre o Programa Nacional de Tecnologia Educacional – ProInfo**. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6300.htm. Acesso em: 27 set. 2014.

CANAVARRO, Ana Paula. Estatística e Calculadoras Gráficas**.** In: LOUREIRO, Cristina; OLIVEIRA, Fernanda; BRUNHEIRA, LINA (Org.) **Ensino e Aprendizagem da Estatística.** Lisboa: GRAFIS, 2000. p.159- 167. Disponível em: http://www.educ.fc.ul.pt/docentes/jponte/sd/textos/canavarro.pdf. Acesso em: 27 set. 2014.

CAZORLA, I. M. Estatística ao alcance de todos. In: VIII Encontro Nacional de Educação Matemática, 2004, Recife: UFPE. **Anais do VIII ENEM.** Brasília: SBEM, 2004. Disponível em: http://www.sbembrasil.org.br/files/viii/pdf/12/MC11915634806.pdf. Acesso em: 27 set. 2014.

CYRINO, M. C. C. T., BALDINI, L. A. L. O *software* Geogebra na formação de professores de Matemática – Uma visão a partir de dissertações e teses. **RPEM**, Campo Mourão, Pr, v.1, n.1, jul-dez, 2012. Disponível em: http://www.fecilcam.br/rpem/documentos/v1n1/Software%20Geogebra.pdf. Acesso em 19 nov. 2014.

CRESPO, A. A. **Estatística fácil.** 18ª ed. São Paulo: Saraiva, 2002.

CORRÊA, A. A. Saberes docentes e educação estatística: composições analíticas no ensino médio. **EMP**, São Paulo, v. 14, 2012, n. 1. Disponível em: http://revistas.pucsp.br/index.php/emp/article/view/8125. Acesso em: 27 set. 2014.

DUARTE, L. R. **A utilização do *Software* Geogebra no ensino da distribuição normal:** Uma aproximação entre a Geometria Dinâmica e a Educação Estatística. 2010. 129 f. Dissertação (Mestrado) - Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2010.

FERREIRA, I. F.; CARVALHO, K. S.; BECKER, A .J. Applets no Geogebra: Atividades de Estatística e Probabilidade no Ensino Médio. In: XIII CIAEM Conferencia Internacional de Educação Matemática, 2011, Recife: UFPE. **Anais da XIII CIAEM**. Recife: UFPE, 2011. Disponível em: http://www.gente.eti.br/lematec/CDS/XIIICIAEM/artigos/2524.pdf. Acesso em: 27 set. 2014.

GIONGO, I. M.; QUARTIERI, M. T.; REHFELDT, M. J. H. Problematizando o uso da estimativa em aulas de matemática da escola básica. In: XI ENEM Encontro Nacional de Educação Matemática, 2013, Curitiba: PUCPR. **Anais do XI Encontro Nacional de Educação Matemática**. Brasília: SBEM, 2013. Disponível em: http://sbem.esquiro.kinghost.net/anais/XIENEM/pdf/1099_200_ID.pdf. Acesso em: 27 set. 2014.

GIRARDI, F.; GIONGO, I. M. Geometria e estimativa nos anos iniciais do ensino fundamental. **Revista Destaques Acadêmicos,** CETEC/UNIVATES-RS, v. 5, n. 4, 2013. Disponível em: http://www.univates.br/revistas/index.php/destaques/article/viewFile/844/534. Acesso em: 27 set. 2014.

HOHENWARTER, M.; HOHENWARTER, J. **Ajuda Geogebra 3.2**. Hanover: 2009. Disponível em: http://www.Geogebra.org. Aceso em: 27 set 2014.

LIMA, E. L. et al. **Temas e Problemas Elementares.** Coleção do professor de matemática. 2ª ed. Rio de Janeiro: SBM, 2006.

LIMA, O. A. Distribuição Normal: Uma introdução voltada ao ensino médio por simulações via planilha eletrônica e exercícios interativos. In: X Encontro Nacional de Educação Matemática, 2010, Salvador. **Anais do X Encontro Nacional de Educação Matemática**. Recife: UFPE, 2010. Disponível em: http://www.lematec.net/CDS/ENEM10/artigos/MC/T15_MC1136.pdf. Acesso em: 27 set. 2014.

LOPES, C. E. A educação estatística no currículo de matemática: um ensaio teórico. **Anais da 33ª Reunião Anual da ANPEd**. Caxambu, 2010. Disponível em: http://33reuniao.anped.org.br/internas/ver/trabalhos-gt19. Acesso em: 27 set. 2014.

MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de Probabilidade e Estatística**. 6ª ed. São Paulo: Editora da Universidade de São Paulo, 2004.

NETO, B. B.; SCARMINIO, I. S.; BRUNS, R E. **Como fazer experimentos**: Pesquisa e desenvolvimento na ciência e na indústria. 2ª ed. Campinas, SP: Editora da Unicamp, 2003.

SAVIAN, M. C. B.; JOCABI, L. F.; ZANINI, R. R. O ensino e a aprendizagem da estatística no ensino fundamental e médio. In: III EIEMAT Escola de inverno de Educação Matemática e 1ª Encontro Nacional PIBID-Matemática, 2012, Santa Maria: UFSM. **Anais do 3º EIEMAT e 1º Encontro Nacional PIBID Matemática**. Volume 1. Número 1. Santa Maria: UFSM, 2012. Disponível em: http://w3.ufsm.br/ceem/eiemat/Anais/arquivos/PO/PO_Savian_Monica.pdf. Acesso em: 27 set. 2014.

SILVA, A. M.; BAYER, A. Estatística no Ensino Médio: Como Tornar o Desenvolvimento dos Conteúdos mais Interessantes?. In: CUREM 3 Congresso Uruguayo de Educación Matemática, 2011, Montevideo: Instituto Tecnológico Superior del Buceo. **Actas Del 3er CUREM**. Uruguay: SEMUR: 2011. Disponível em: http://www.semur.edu.uy/curem3/actas/105.pdf. Acesso em: 27 set. 2014.

VASCONCELOS, M. H. S.; TOGNI, A. C. Aprendendo Estatística no Ensino Médio e no Curso Técnico Agropecuária Utilizando o Objeto de Aprendizagem Estatísticanet. **Revista Saber Acadêmico**, Faculdade de Presidente Prudente SP – v. 11, jun. 2011. Disponível em: http://www.uniesp.edu.br/revista/revista11/pub_artigos.asp. Acesso em: 27 set. 2014.