UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Presidente Prudente
Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT)

LEONARDO OLIVEIRA BUTURI

APLICAÇÕES BÁSICAS DE ESTATÍSTICA E DA DISTRIBUIÇÃO NORMAL PARA O ENSINO MÉDIO

PRESIDENTE PRUDENTE

2018

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Presidente Prudente
Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT)

LEONARDO OLIVEIRA BUTURI

APLICAÇÕES BÁSICAS DE ESTATÍSTICA E DA DISTRIBUIÇÃO NORMAL PARA O ENSINO MÉDIO

Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre, junto ao Programa de Mestrado Profissional em Matemática em Rede Nacional da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de Presidente Prudente.
Financiadora: Capes – Proc: 5566715

Orientador: Prof. Dr. José Gilberto Spasiani Rinaldi

PRESIDENTE PRUDENTE
2018

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação - Diretoria Técnica de Biblioteca e Documentação - UNESP, Campus de Presidente Prudente

B993a

Buturi, Leonardo Oliveira.

Aplicações básicas de estatística e da distribuição normal para o ensino médio / Leonardo Oliveira Buturi. - 2018

67 f. : il.

Orientador: José Gilberto Spasiani Rinaldi

Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Ciências e Tecnologia, Presidente Prudente, 2018

Inclui bibliografia

1. Estatística. 2. Distribuição normal. 3. Ensino médio. I. Rinaldi, José Gilberto Spasiani. II. Universidade Estadual Paulista. Faculdade de Ciências e Tecnologia. III. Título.

Alessandra Kuba Oshiro Assunção

CRB-8/9013

TERMO DE APROVAÇÃO

LEONARDO OLIVEIRA BUTURI

APLICAÇÕES BÁSICAS DE ESTATÍSTICA E DA DISTRIBUIÇÃO NORMAL PARA O ENSINO MÉDIO

Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre, junto ao Programa de Mestrado Profissional em Matemática em Rede Nacional da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de Presidente Prudente.

Financiadora: Capes – Proc: 5566715

Comissão examinadora:

Prof. Dr. José Gilberto Spasiani Rinaldi
FCT/UNESP – Campus de Presidente Prudente
Orientador

Prof. Dr. Márcio Luis Lanfredi Viola
UFSCar – Campus São Carlos

Prof. Dr. Luis Carlos Benini
FCT/UNESP – Campus de Presidente Prudente

Presidente Prudente, 15 de Junho de 2018

Aos meus familiares e
Principalmente à minha Esposa e filhos,
pela compreensão nos momentos
que estive ausente......
dedico.

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus por me ajudar em mais esta etapa da minha vida. Agradeço à minha esposa e aos meus familiares pela paciência e pelo apoio, aos meus amigos da minha turma de pós-graduação pela amizade e companheirismo. Á Capes, pelo incentivo financeiro.

“Os que plantam sementes entre lágrimas recolherão com Alegria”

Salmo 125.

RESUMO

Tendo por finalidade apresentar um estudo conciso em alguns tópicos de Estatística, com ênfase em aplicações e na distribuição normal, este trabalho destina-se a alunos e professores do Ensino Médio, que tenham interesse em se aprofundar no estudo destes tópicos. Técnicas estatísticas, principalmente nos problemas que podem ser modelados utilizando-se a distribuição normal, são de interesse geral e muito utilizado nas áreas de biologia, ecologia, administração de empresas e negócios. Inicialmente, são apresentados vários conceitos de Estatística tais como gráficos, medidas de posição e dispersão, conceitos de probabilidade, a distribuição normal, intervalos de confiança para a posterior aplicação em dados envolvendo os preço de combustíveis e notas de alunos dos Ensinos Fundamental e Médio. Os tópicos são expostos de forma que seja acessível aos alunos do ensino médio e as aplicações são motivadoras para a utilização posterior das técnicas apresentadas.
Palavras chave: Estatística. Distribuição normal. Ensino Médio.

ABSTRACT

Having as purpose present a concise study on Statistics and Probability, with emphasis on applications and the normal distribution, this work is intended for students and High School teachers, who have an interest in deepening on these topics. Statistical techniques, mainly in problems that can be modeled using the normal distribution are of general interest and widely used in the areas of biology, ecology, business administration and business. Initially, are presented several concepts of Statistics such as position and dispersion measurements, probability concepts, normal distribution, simple confidence intervals and applications. Topics are exposed in a way that is accessible to high school students and the applications are motivating for the later use of the techniques presented.
Keywords: Statistics. Normal distribution. High school.

SUMÁRIO

1 INTRODUÇÃO

A Estatística surgiu há muitos séculos, quando seus antigos habitantes avistaram a necessidade de contabilizar não somente a população, mas também residências, bens, e também com a finalidade de cobrança de impostos.

Neste contexto, Larson e Farber (2015, p. 3) apontam que "o uso de dados estatísticos remonta aos censos realizados na Antiga Babilônia, Egito e, mais tarde, no Império Romano, quando os dados coletados eram sobre assuntos relacionados ao Estado".

Atualmente a Estatística é uma ferramenta importante para muitas áreas do conhecimento, tais como Medicina, Agronomia e Computação, com intuito de constatar fatos e tendências utilizando informações de pesquisa para prever necessidades futuras.

Partindo deste pressuposto, Larson e Farber (2015, p. 3) destacam a Estatística como a "ciência que trata da coleta, organização e interpretação dos dados para a tomada de decisões".

Com base em Bonafini (2015, p. 2), "a Estatística nada mais é que a ciência de trabalhar com os dados de que dispomos e fazer interpretações sobre eles".

Na fase inicial da fase Estatística, a apresentação gráfica revela-se, como um processo que permite "condensar grandes conjuntos de dados e apresentá-los, numa forma fácil de digeri-los" (FREUND, 2007, p. 42). Com isso, analisando todo o conteúdo e elaborando da melhor maneira, o usuário obterá uma resposta mais visível e eficaz de todos os dados e assim podendo entender as informações para a posterior aplicação de métodos estatísticos inferenciais.

Uma das aplicações da Estatística está nos Seguros. O seguro surgiu há muito tempo. Mais precisamente há 5000 anos entre os comerciantes marítimos, onde aplicavam técnicas para assegurar-se contra naufrágios e roubos. Não se sabe muito sobre essas técnicas utilizadas naquela época, mas pode se afirmar que se baseavam em suas experiências adquiridas com o passar do tempo. Sendo assim, nessa mesma época houve um grande aumento nos negócios, pois os seguros marítimos e os seguradores continuavam a utilizar a antiga técnica baseada em conhecimentos, para que achassem uma melhor solução que os protegessem de acidentes e a perda de suas mercadorias.

Como um ramo ligado à Matemática e que conquistou o status de Ciência, a Estatística tem grande importância também na Política, nas Comunicações e nos Esportes.

A partir dos estudos realizados por Pasquali (2014) observa-se, que a curva normal tem origem no século XVII e está ligada ao estudo da probabilidade, que surgiu com a proposta de se estudar jogos de azar. Um dos primeiros matemáticos a estudá-la foi o Francês Abraham de Moivre, que publicou suas anotações em sua obra *The doctrine of chances*. A descoberta logo foi utilizada por nomes importantes da Matemática, como Laplace, para representar a distribuição de erros, e Gauss, na análise de dados astronômicos. Hoje em dia a curva normal tornou-se uma ferramenta muito importante, pois a normalidade ocorre em muitas medidas físicas, biológicas ou sociais.

1.1 OBJETIVOS

Tem-se, por objetivo geral, apresentar ferramentas básicas de Estatística e noções de distribuição normal, com enfoque em aplicações que podem ser utilizadas por alunos do ensino básico. Pouco explorada no Ensino Médio, a Estatística constitui uma ementa muito importante para a prática da pesquisa e da inferência de dados, que hoje é muito utilizada por empresas públicas e privadas, poder público e diversos setores para melhorar os serviços e dar indicadores. Neste trabalho são apresentadas aplicações reais que podem ser replicadas para o ensino de Estatística.

1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

A finalidade desse trabalho é propor uma maneira de mostrar a curva normal de Gauss e uma essencial ferramenta para o cálculo de probabilidade que é a Distribuição normal.

Assim, nos Capítulos 2 e 3, temos uma abordagem dos conceitos iniciais de Estatística, como o que são variáveis, população e amostra, bem como suas representações gráficas. As medidas de posição, envolvendo a média, moda e mediana bem como as medidas de dispersão, tratando-se do desvio médio, variância e desvio padrão. No Capítulo 4, estudamos o cálculo de probabilidade e a distribuição normal para variáveis aleatórias. Finalmente, no Capítulo 5, os intervalos de confiança com abordagem voltada para aplicação em sala de aula do ensino médio.

2 PRINCIPAIS CONCEITOS DE ESTATÍSTICA

A Estatística é uma ciência fundamental em diferentes campos do conhecimento humano. Neste sentido:

> Qualquer abordagem que diga respeito, por menor que seja, de coleta, processamento, interpretação e apresentação de dados pertence ao domínio da Estatística, assim como o planejamento detalhado que precede todas essas atividades. De fato, a Estatística inclui tarefas tão diversificadas como calcular a média de acertos em arremessos de jogadores de basquete, coletar e registrar dados sobre nascimentos, casamentos e mortes, avaliar a operacionalidade de produtos comerciais e prever o tempo. (FREUND, 2007, p. 13).

Observa-se que a Estatística aplicada envolve desde a coleta, processamento, interpretação de dados sobre um determinado fenômeno até a mensuração de dados em práticas desportivas, ou até mesmo para a previsão da temperatura, na medida em que se determina a probabilidade de ocorrência de mudanças climáticas ou não.

Frente a tal abrangência é que na atualidade tem-se ampliado o número de usuários de métodos estatísticos.

Partindo deste pressuposto:

> Há inúmeros motivos por meio dos quais a amplitude da Estatística e a necessidade de estudar a Estatística têm crescido consideravelmente ao longo dos últimos, aproximadamente, cinquenta anos. Um deles consiste na abordagem crescentemente quantitativa utilizada em todas as ciências, bem como na Administração e em muitas outras atividades que afetam diretamente nossas vidas. Isso compreende a utilização de técnicas matemáticas na avaliação de controles de poluição, no planejamento de inventários, na análise de padrões do trânsito de veículos, no estudo dos efeitos de vários tipos de medicamentos, na avaliação de técnicas de ensino, na análise do comportamento competitivo de administradores e governos, no estudo da dieta e da longevidade, e assim por diante. Também a disponibilidade de computadores aumentou enormemente nossa capacidade de lidar com informação numérica, a tal ponto que um trabalho estatístico sofisticado pode ser realizado mesmo por pequenas empresas e por alunos de escola e de faculdade. (FREUND, 2007, p. 14).

A seguir são apresentadas algumas definições básicas envolvendo Estatística, fundamentais para o desenvolvimento de conceitos que serão analisados no futuro.

2.1 POPULAÇÃO

Conforme Crespo (1997) o termo população diz respeito ao conjunto de portadores de, pelo menos, uma característica comum. Assim, os eleitores de uma cidade, por exemplo, constituem uma população, pois apresentam pelo menos uma característica comum: são os que votam.

Como em qualquer estudo estatístico queremos pesquisar características em comum dos elementos de uma população, faz-se necessário definir bem essas características. Por exemplo, quando queremos fazer uma pesquisa entre os funcionários de uma empresa precisamos definir quais os funcionários que fazem parte da população: se somente os que estão trabalhando ou os que passaram pela empresa. Na maioria das vezes, por impossibilidade ou inviabilidade econômica ou temporal, limitamos as observações referentes a uma determinada pesquisa a apenas uma parte da população. A essa parte proveniente da população em estudo denominamos **amostra**. (Crespo, 1997).

2.2 AMOSTRA

Conforme Dowing e Clark (2010) uma amostra é o conjunto de elementos que são selecionados e retirados da população, ou seja, uma parcela representativa que é tomada por referência, para que, a partir dela, se possa realizar conclusões a respeito da população.

Tem-se, de tal modo que uma amostra estatística se traduz em um subconjunto representativo, ou seja, um determinado conjunto que é extraído de uma população, com o intuito de fazer com que análises estatísticas sobre este determinado grupo apenas, possa oferecer informações essenciais sobre a população de origem.

Segundo destaca Bonafini (2015, p. 5), "para que uma amostra seja válida, é preciso que ela seja extraída com base em critérios rígidos e com credibilidade".

2.3 FASES DA ESTATÍSTICA

A Estatística se divide basicamente em Descritiva e Inferencial:

> A origem da Estatística moderna remonta a duas áreas de interesse que, na aparência, pouco tem em comum: governo (ciência política) e jogos de azar. Os governos vêm, de longa data, utilizando recenseamentos para contar indivíduos e propriedades, e o problema de descrever, resumir e analisar dados de censos levou ao desenvolvimento de métodos que, até recentemente, constituíam quase tudo o que havia na área de Estatística.

> Esses métodos que, no início, consistiam principalmente na apresentação de dados em forma de tabelas e gráficos, constituem o que hoje denominamos estatística descritiva. Esta inclui tudo relacionado com dados que seja projetado para resumir ou descrever dados, mas sem ir além, ou seja, sem procurar inferir qualquer coisa que vá além dos próprios dados. Por exemplo, se os testes feitos com seis carros pequenos importados em 1999 mostraram que eles são capazes de acelerar de 0 a 60 milhas por hora (mph) em 12,9; 16,5; 11,3; 15,2; 18,2; e 17,7 segundos, e se afirmamos que metade deles acelera de 0 a 60 mph em menos de 16,0 segundos, então nosso trabalho pertence ao domínio da estatística descritiva. (FREUND, 2007, p. 16).

De tal modo, a estatística descritiva se preocupa em aplicar inúmeras técnicas para apresentação e sistematização de uma quantidade de dados enquanto a inferencial é a aquela que, partindo de uma amostra, estabelece hipóteses tirando conclusões sobre a população de origem e que formula previsões.

2.3.1 Amostragem

Existem algumas técnicas para obter amostras que são empregadas com a finalidade de assegurar ao máximo, em termos de escolha, a ocorrência do acaso. Consequentemente, cada elemento da população passa a desfrutar da mesma chance de ser selecionado, conferindo assim mais credibilidade aos resultados que dependem diretamente dos resultados obtidos nas amostras de uma população.

Seguem três técnicas de amostragem expressivamente importantes:

. Amostragem casual ou aleatória simples: todos os elementos da população têm a mesma probabilidade não nula de pertencer à amostra, assim como afirmam Baptista e Campos (2016).

. Amostragem sistemática: as populações têm os elementos ordenados, sem a necessidade de construir um sistema de referência. Assim, os elementos que farão parte da amostra são selecionados por um sistema criado pelo próprio pesquisador.

. Amostragem aleatória estratificada: "divide a população em grupos distintos, intitulados de estratos e, então, seleciona uma amostra aleatória simples em cada estrato" (AGRESTI; FINLY, 2012, p. 39).

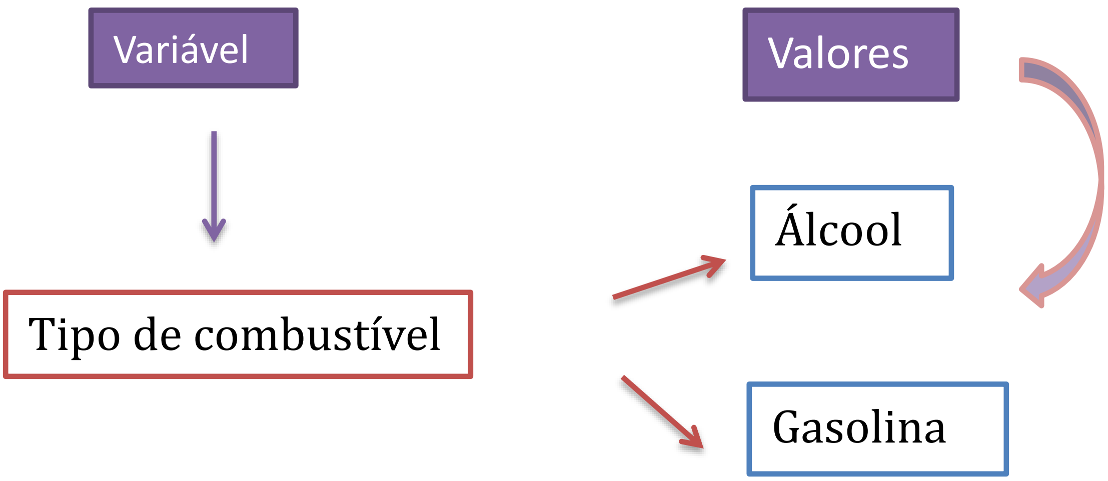
"A amostragem aleatória estratificada é denominada de proporcional se as proporções do estrato amostrado são as mesmas da população" (AGRESTI; FINLY, 2012, p. 39).

“A amostragem aleatória estratificada é chamada de desproporcional se as proporções dos estratos amostrados diferem das proporções da população” (AGRESTI; FINLY, 2012, p. 39).

2.4 VARIÁVEL

A variável é medida em cada elemento da amostra e é reconhecida como sendo a característica de interesse. Os valores das variáveis mudam de acordo com cada elemento e podem ser numéricos ou não numéricos, conforme apresentado na Figura 1.

Figura 1 – Exemplo de variável com valores não numéricos.



Fonte: Elaborado pelo autor

As variáveis, por sua vez, classificam-se em qualitativas ou quantitativas:

> Algumas variáveis, como sexo, educação, estado civil, apresentam como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado, ao passo que outras, como número de filhos, salário, idade, apresentam como possíveis realizações números resultantes de uma contagem ou mensuração. As variáveis do primeiro tipo são chamadas qualitativas, e as do segundo tipo, quantitativas. (MORETTIN, 2013, p. 11).

Observa-se, que há certas particularidades que diferenciam as variáveis qualitativas, das quantitativas, ao passo em que as primeiras revelam uma qualidade ou atributo do indivíduo pesquisado, enquanto as segundas dizem respeito às características que podem ser projetadas em números, ou seja, passíveis de mensuração.

2.4.1 Variáveis qualitativas

“As variáveis qualitativas ou variáveis categorizadas são medidas em escala nominal, como sexo, setor de trabalho e faixa etária” (BRACARENSE, 2012, p. 237), ou seja, quando seus valores são expressos por atributos. E podem ser nominais ou ordinais.

– Variáveis nominais: não existe ordenação dentre as categorias. Exemplos: sexo, cor dos olhos, fumante/não fumante, doente/sadio.

– Variáveis ordinais: existe uma ordenação entre as categorias. Exemplos: escolaridade (1º, 2º, 3º graus), estágio da doença (inicial, intermediário, terminal), mês de observação (janeiro, fevereiro,…, dezembro).

2.4.2 Variáveis quantitativas

As variáveis quantitativas são características que podem ser descritas por números, sendo estas classificadas entre contínuas e discretas.

Variáveis discretas: a variável é avaliada em números que são resultados de contagens e, por isso, somente fazem sentido números inteiros. Exemplos: número de filhos, número de bactérias por litro de leite, número de cigarros fumados por dia.

Variáveis contínuas: a variável é avaliada em números que são resultados de medições e, por isso, podem assumir valores com casas decimais e devem ser medidas por meio de algum instrumento. Exemplos: massa (balança), altura (régua), tempo (relógio), pressão arterial, idade.

2.5 GRÁFICOS NA ESTATÍSTICA

A utilização de gráficos na Estatística viabiliza o acompanhamento de forma mais clara e objetiva dos dados estatísticos sobre um determinado fenômeno.

Assim sendo, a apresentação gráfica refere-se à “apresentação geométrica dos dados, permitindo uma visão rápida, fácil e clara de um determinado fenômeno e sua variação” (CORREA, 2003, p. 15).

A estatística utiliza de vários tipos de gráficos, dependendo do tipo de variável. Alguns deles são: gráfico de barras, gráfico de colunas, de linhas, de setores ou de pizzas, histogramas e polígonos de frequência.

Diante de mapa gráfico pode-se visualizar informações obtidas de uma amostra, respondendo de pronto às indagações do usuário sobre um determinado fenômeno e/ou objeto do conhecimento.

De acordo com Rumsey (2009), os gráficos permitem visualizar de forma fácil os dados que norteiam um determinado fenômeno, possibilitando a quem estiver visualizando analisar uma dada realidade e a partir desta extrair conclusões particulares da amostra selecionada.

2.5.1 Gráfico de pizza

Com base em Rumsey (2009), o gráfico de pizza considera os dados categorizados, apresentando-os em grupos, permitindo a visualização da porcentagem de indivíduos, ou outro signo, que se encontra na mesma situação. Um exemplo de gráfico de pizza pode ser identificado na Figura 2.

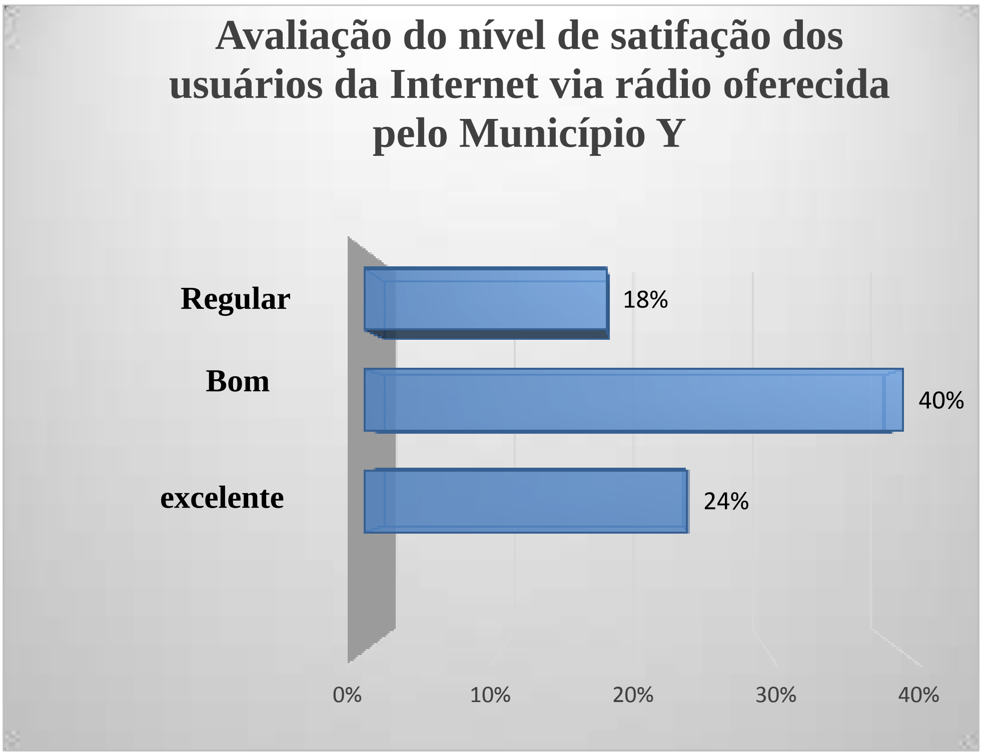
Figura 2 – Representação de dados por meio de um gráfico de pizza.

Fonte: Elaborado pelo autor

2.5.2 Gráfico de barras

Com base em Correia (2003), o gráfico de barras envolve a representação de uma série de dados através de retângulos organizados horizontalmente. A Figura 3 apresenta um exemplo de gráfico de barras.

Figura 3 – Representação de dados por meio de um gráfico de barras.

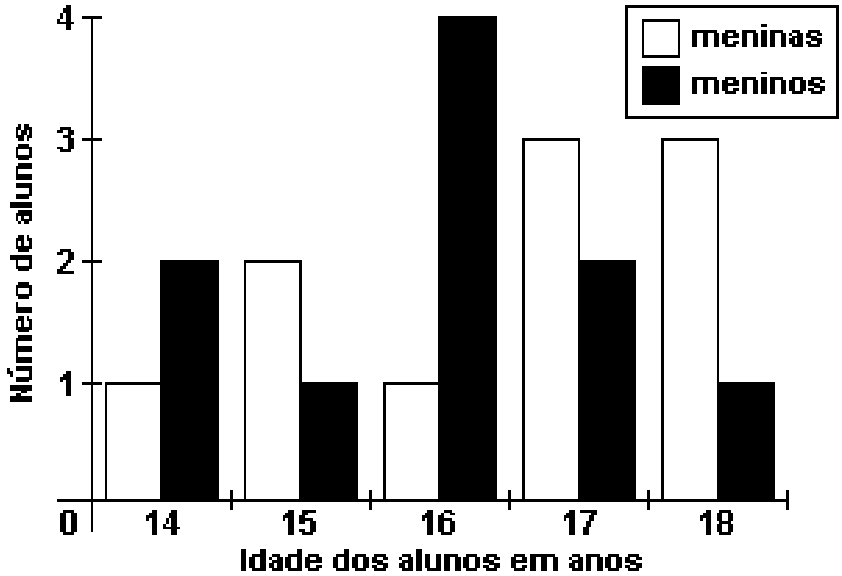


Fonte: Elaborado pelo autor

2.5.3 Gráfico de colunas

Com base em Kühn (2017), o gráfico de colunas permite a comparação entre itens de um determinado fenômeno no que envolve uma dada variável. Todas as colunas do gráfico devem ter a mesma largura, e suas alturas é que variam. Um gráfico deste tipo segue representado na Figura 4.

Figura 4 – Representação de dados por meio de um gráfico de colunas.



Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios 2005

3 MEDIDAS DE TENDÊNCIA CENTRAL E MEDIDAS DE DISPERSÃO

Pensando estatisticamente, evidencia-se, que as informações e dados representativos de um determinado fenômeno são transportados para gráficos e tabelas, devendo a totalidade ser resumidas através de medidas que o caracterizam.

Neste contexto, as medidas de tendência central referem-se aos valores que representam a totalidade, ou seja, o conjunto de dados existentes.

“As medidas de tendência central são valores únicos cuja função é representar o conjunto de dados como um todo” (BONAFINI, 2015, p. 39).

Já para algumas amostras, fazer o cálculo de medidas de tendência central não é suficiente para analisar uma situação, assim faz-se o uso das medidas de dispersão, que analisam o quanto os dados se encontram dispersos ao redor de uma medida de posição central, portanto caracterizam o grau de variação existente em um conjunto de valores.

Tratando-se sobre as medidas de centralidade, as de maior destaque são a média aritmética, a moda e a mediana, e sobre as medidas de dispersão, as mais importantes são o desvio médio, a variância e o desvio padrão, cada qual com particularidades a serem analisadas.

3.1 MÉDIAS

A Média diz respeito a determinado valor que representa um conjunto de dados. Tomando-se, por base o fato de que tais valores passam a inserir-se em um ponto central do conjunto de dados, as medidas passam a ser intituladas medidas de tendência central.

Se a característica desse conjunto for a soma dos seus elementos, atinge-se a média aritmética, se a característica diz respeito ao produto dos elementos, chega-se à média geométrica. As principais médias utilizadas são: Média aritmética, média geométrica e a média harmônica.

3.1.1 Médias – definição e exemplos

A mais importante das médias é a aritmética, “trata-se do quociente da divisão da somatória dos valores da variável pelo número deles, constituindo, via de regra, a de maior relevância entre todas as medidas descritivas” (CORREA, 2003, p. 49).

Considerando um conjunto composto por n números $x_1, x_2, \dots, x_n$, a média aritmética é definida por:

$$M_A = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Exemplo: Afere-se a média aritmética dos números 4; 6; 2 e 5 fazendo:

$$M_A = \frac{4+6+5+2}{4} = \frac{17}{4} = 4{,}25.$$

Propriedade da média aritmética: A somatória dos "desvios" de um conjunto de números em relação a média aritmética é zero.

Exemplo: Os desvios dos números 9; 6; 6 e 3 em relação a sua média 6 são: 3, 0, 0, -3 com soma algébrica zero.

Em algumas situações, os números $x_1, x_2, \dots, x_n$ são relacionados a determinados pesos ou fatores de ponderação $p_1, p_2, \dots, p_n$ que são dependentes do significado ou da relevância correlacionada aos números. Assim, definimos Média Ponderada como:

$$M_P = \frac{p_1 . x_1 + p_2 x_2 + \cdots + p_n x_n}{p_1 + p_2 + \cdots + p_n} = \frac{\sum_{i=1}^{n} p_i . x_i}{\sum_{i=1}^{n} p_i}.$$

Exemplo: Uma prova é feita em três fases com pesos 1, 2 e 3, respectivamente. Um estudante teve notas 6, 6 e 4 nestas fases. Qual foi seu desempenho médio?

$$M_P = \frac{6.1 + 6.2 + 3.4}{6} = \frac{30}{6} = 5.$$

Alcança-se média geométrica dos n números positivos $x_1, x_2, \dots, x_n$ através da seguinte fórmula:

$$M_g = \sqrt[n]{x_1 . x_2 \dots . x_n}.$$

Exemplo: A média geométrica dos números 2, 4 e 8 é:

$$M_g = \sqrt[3]{2.4.8} = \sqrt[3]{64} = 4.$$

Exemplo: Uma indústria ampliou seu processo produtivo ao longo do segundo bimestre do ano passado. Em março e abril, as taxas de ampliação foram de 21% e 8%, respectivamente. Diante deste contexto, como determinar a taxa média de aumento mensal nesse bimestre?

A média procurada é a geométrica. Queremos uma taxa i equivalente ao aumento mês a mês.

$$100 \;\rightarrow\; 100.1{,}21 \;\rightarrow\; 100.1{,}21.1{,}08 = 130{,}68.$$

Se em todos os meses tivéssemos um aumento de taxa, teríamos:

$$100(1+i)^2 = 130{,}68,$$
$$1 + i = \sqrt{1{,}3068,}$$
$$i = 0{,}1432 \text{ ou } 14{,}32\%.$$

Logo, a taxa média procurada é $14{,}32\%$.

A média harmônica dos n números positivos $x_1, x_2, \ldots, x_n$ é definida por:

$$h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots \frac{1}{x_n}}.$$

Exemplo: Um sorteio mensal distribui igualmente entre os vencedores um prêmio total de R$1800,00. Nos últimos três meses houve 2, 1 e 3 premiados, respectivamente. Qual foi o preço médio desses ganhadores? (PASQUALI, 2014).

Busca-se uma média tal que, se todos os prêmios fossem iguais a essa média, o total distribuído seria o mesmo. A média aritmética dos prêmios foram $1800 \div 2 = 900, 1800 \div$

$1 = 1800$ e $1800 \div 3 = 600$. O prêmio médio foi de $(900 + 1800 + 600) \div 3 = 1100$ reais.

Evidencia-se que a média aritmética dos sorteios é igual a:

$$\frac{1800 \times \frac{1}{2} + 1800 \times 1 + 1800 \times \frac{1}{3}}{3} = 1800 \times \frac{\frac{1}{2} + \frac{1}{1} + \frac{1}{3}}{3} =$$
$$= 1800 \div \frac{3}{\frac{1}{2} + \frac{1}{1} + \frac{1}{3}},$$

e que $\frac{3}{\frac{1}{2}+\frac{1}{1}+\frac{1}{3}}$ é a media harmônica do número de ganhadores.

3.2 MEDIDAS DE DISPERSÃO

Segundo Dowing e Clark (2010), utiliza-se o termo dispersão para indicar o grau de afastamento de um conjunto de números no que diz respeito, geralmente, à sua média.

Desta forma, o nível apresentado pelos dados numéricos sobre a tendência de se dispersarem em torno de um valor médio é denominado variação ou dispersão dos dados.

Apresentam-se a seguir algumas medidas de variação, observando-se que, rotineiramente, dentre as mais utilizadas estão a amplitude total, o desvio médio, a variância e o desvio padrão.

3.2.1 Amplitude total

A partir das proposições de Dowing e Clark (2010), a amplitude nos dá uma ideia do afastamento entre o maior e o menor valor, mas não é, na realidade, uma boa medida da dispersão dos dados.

A amplitude viabiliza uma perspectiva da possibilidade de variação dos elementos. Mais precisamente, ela fornece a maior variação possível dos dados.

Neste sentido, a amplitude total de um conjunto de números é resultante da diferença entre o valor mais alto e o valor mais baixo.

Considerando o conjunto de dados ordenados $X_1 \leq X_2 \leq X_3 \leq \cdots \leq X_{n-1} \leq X_n$ a amplitude dos dados é dada pela expressão $X_n - X_1$.

A amplitude total do conjunto $5, 5, 8, 9, 10, 12, 15, 23$ é $23 - 5 = 18$.

3.2.2 O desvio médio

Dado um conjunto $x_1, x_2, \dots, x_n$, contendo n números e cuja média aritmética é $\bar{x}$, o desvio médio é definido por:

$$desvio\ médio = \frac{\sum_1^n |x_i - \bar{x}|}{n}.$$

Exemplo: Calcule o desvio médio dos números $2, 3, 6, 8$ e 11.

Solução:

$$\bar{x} = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6,$$

$$desvio\ médio = \frac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5} = \frac{4+3+0+2+5}{5} =$$

$$= 2{,}8.$$

3.2.3 A variância

Segundo o referencial teórico de Correa (2003), a variância toma por base os valores extremos e os valores intermediários, quer seja, demonstra de forma mais ampla os resultados obtidos.

A variância é uma medida de dispersão muito utilizada. Neste sentido:

> O desvio médio absoluto é uma boa medida de dispersão porque dá a distância média de cada número em relação à média. Todavia, para muitos propósitos, é mais conveniente elevar ao quadrado cada desvio e tomar a média de todos esses quadrados. Essa grandeza é chamada variância. (DOWING; CLARK, 2010, p. 11).

Trata-se do quociente entre a soma dos quadrados dos desvios e o número de elementos.

Exemplo: A variância do exemplo acima é:

$$S^2 = \frac{4^2 + 3^2 + 0^2 + 2^2 + 5^2}{5} = \frac{54}{5} = 10{,}8.$$

A variância é uma medida de dispersão bastante utilizada, porém seu resultado não está na mesma unidade da variação dos dados.

3.2.4 O desvio padrão

Ressaltam Dowing e Clark (2010) que a variância é uma boa medida de dispersão, mas tem uma desvantagem: é difícil interpretar o valor numérico da variância, sendo que em geral, é melhor calcular a raiz quadrada da variância, chamada desvio padrão.

De acordo com o referencial teórico de Correa (2003), o desvio padrão é a medida mais utilizada no comparativo de diferenças entre conjuntos de dados, apresentando, geralmente, maior fidelidade na análise.

O desvio padrão é representado por S e é definido por:

$$S = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n}}.$$

Exemplo: Dois profissionais disputam uma única vaga em uma indústria. São executados diversos testes com esses dois profissionais, Henrique e Mateus. Por meio da Tabela 1 é possível visualizar os desempenhos dos dois candidatos.

Tabela 1 – Desempenho apresentado pelos profissionais.

	Henrique	Mateus
Português	8,0	8,5
Matemática	10,0	10,0
Informática	7,5	8,0
Ciências	7,5	8,5
Espanhol	7,0	5,0

Fonte: Elaborado pelo autor

Note que as médias de Henrique e Mateus são iguais:

Paulo: $\overline{X_E} = \frac{8,0+10+7,5+7,5+7}{5} = 8,0,$

Pedro: $\overline{X_V} = \frac{8,5+10+8,5+8+5}{5} = 8,0.$

Uma pergunta que fica é: Como verificar matematicamente qual dos candidatos merece assumir a vaga?

Podemos fazer uma comparação usando medidas da estatística chamadas de medidas de dispersão, e que envolvem o desvio médio, variância e desvio padrão.

Vamos calcular os desvios de cada um:

Henrique:	Mateus:
$\lvert 8-8\rvert = 0$	$\lvert 8,5-8\rvert = 0,5$
$\lvert 10-8\rvert = 2$	$\lvert 10-8\rvert = 2$
$\lvert 7,5-8\rvert = 0,5$	$\lvert 8-8\rvert = 0$
$\lvert 7,5-8\rvert = 0,5$	$\lvert 8,5-8\rvert = 0,5$
$\lvert 7-8\rvert = 1$	$\lvert 5-8\rvert = 3$

Desvio médio de Henrique:

$$\overline{D_H} = \frac{0+2+0,5+0,5+1}{5} = 0,8.$$

Desvio médio de Mateus:

$$\overline{D_M} = \frac{0,5+2+0+0,5+3}{5} = 1,2.$$

Os cálculos acima mostram que as notas de Henrique estão em média 0,8 unidades afastadas da média, enquanto as notas de Mateus estão em média 1,2 afastadas da média, onde essas medidas apenas quantificam esse afastamento.

Isso demonstra que as notas de Henrique são menos dispersas que as notas de Mateus.

Vamos calcular a variância S^2 de cada candidato:

Variância de Henrique:

$$\delta^2 = \frac{0^2 + 2^2 + 0{,}5^2 + 0{,}5^2 + 1^2}{5} = \frac{5{,}5}{5} = 1{,}1.$$

Variância de Mateus:

$$\delta^2 = \frac{0{,}5^2 + 2^2 + 0^2 + 0{,}5^2 + 3^2}{5} = \frac{13{,}5}{5} = 2{,}7.$$

Diante deste contexto, as notas de Henrique são menos dispersas que as notas de Mateus. Quanto menor a variância, menos dispersas são as notas. Logo, Henrique teve um desempenho mais regular.

Ao aferir-se o Desvio Padrão dos Candidatos:

Desvio padrão de Henrique:

$$S = \sqrt{1{,}1} \approx 1{,}0488.$$

Desvio padrão de Mateus:

$$S = \sqrt{2{,}7} \approx 1{,}6431.$$

Tem-se que as notas de Henrique são menos dispersas que as notas de Mateus. Quanto menor for o desvio padrão, menos dispersas são as notas. Conclusão: Henrique teve melhor desempenho em relação a Mateus e possivelmente, baseado neste critério, merece a vaga.

4 PROBABILIDADE E DISTRIBUIÇÃO NORMAL

O surgimento da teoria das probabilidades deu-se no momento em que o homem se interessou em estudar fenômenos que não se podia prever o resultado, mas que tinham certas possibilidades de ocorrer.

“O termo probabilidade refere-se ao estudo da incerteza e da aleatoriedade” (BONAFINI, 2015, p. 62).

Há alguns indícios de que a teoria das probabilidades iniciou-se nos jogos de azar, muito praticados na Idade Média, com os jogos de apostas. Muitos matemáticos como Pacioli, Cardano e Tartaglia (séc. XVI) trabalharam nesse tema buscando aumentar as suas chances de ganhar e também de prever acontecimentos que nem ao menos haviam ocorrido ainda. As teorias da probabilidade desenvolveram-se a partir da contribuição de vários matemáticos, mas quem deu um maior aprofundamento para essa área foi Blaise Pascal, Pierre de Fermat e Jacob Bernoulli. Atribui-se este feito a Pacioli, Cardano e Tartaglia (séc. XVI).

A teoria das probabilidades e também da análise combinatória sustentam-se nos fundamentos estabelecidos por Pascal e Fermat. Seu início como ciência foi marcado por situações envolvendo apostas em jogos de dados, onde se começou a levantar hipóteses sobre os resultados possíveis.

A partir do referencial teórico de Crespo (1997), Bernoulli contribuiu no sentido de enfatizar os grandes números, através de uma abordagem envolvendo combinações, permutações e a classificação binomial. Laplace formulou a regra da sucessão, enquanto Gauss foi responsável pelo desenvolvimento da lei das distribuições das probabilidades.

Nos dias de hoje, os estudos referentes à teoria das probabilidades têm aplicações em diversos contextos e situações, fato possibilitado em função da existência de axiomas, teoremas e definições. Uma das principais aplicações relaciona-se ao estudo envolvendo jogos e, consequentemente, dos fatores relacionados às vitórias, cuja ênfase é dedicada à estatística indutiva.

4.1 CONCEITOS DE PROBABILIDADE

4.1.1 Experimento aleatório

São aqueles que, mesmo repetido várias vezes sob condições semelhantes, apresentam resultados imprevisíveis (Crespo, 1997). Como exemplo, podemos citar o experimento "observar a face de cima após o lançamento de um dado". Sabemos quais são as possibilidades de sair na face de cima, mas não podemos prever qual número vai aparecer.

Em quase tudo vislumbramos o acaso. Em uma partida de futebol, existem três possibilidades para um time: ganhar a partida, empatar ou perder a partida. Mas não podemos afirmar qual das três vai acontecer.

4.1.2 Espaço amostral

Com base em Correa (2003) o espaço amostral (denotado por Ω) de um experimento aleatório é constituído pelo conjunto de todos os resultados possíveis do experimento em questão. Desta forma, no lançamento de duas moedas, o espaço amostral será:

$$\Omega = \{(cara, cara), (cara, coroa), (coroa, cara), (coroa, coroa)\}.$$

4.1.3 Evento

De acordo com Crespo (1997) um subconjunto do espaço amostral denomina-se evento, ou seja, um evento E é um conjunto, tal que $E \subset \Omega$. No caso em que ocorre $E = \Omega$, então E é um evento certo. Temos o que chamamos de evento elementar quando o conjunto E é unitário. Além disso, existe ainda a possibilidade de $E = \emptyset$, o que caracteriza o evento E como sendo um evento impossível.

Exemplo: No lançamento de um dado, identificamos o espaço amostral como sendo $\Omega= \{1, 2, 3, 4, 5, 6\}$. Dado $A = \{2, 4, 6\}$ um subconjunto de Ω, podemos defini-lo como o evento "obter um número par na face de cima", e o conjunto $B = \emptyset$, também subconjunto de Ω, pode ser o evento "obter um número maior que 6 na face de cima".

4.1.4 Probabilidade

A distribuição de probabilidades indica a percentagem de vezes que, em grande quantidade de observações, podemos esperar a ocorrência de vários resultados de uma variável aleatória. Vamos considerar o espaço amostral Ω de um experimento aleatório. Além disso, S é um conjunto equiprovável, isto é, todos os elementos de Ω têm a mesma chance de ocorrer. A probabilidade de um evento A, denotada por $P(A)$, é calculada como segue:

$$P(A) = \frac{n(A)}{n(\Omega)},$$

sendo $n(A)$ e $n(\Omega)$, o número de elementos do conjunto A e do conjunto Ω, respectivamente.

Exemplos:

a) Ao lançar um dado não viciado, a probabilidade de ocorrer o evento A "número par na face de cima" é:

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$
$$A = \{2, 4, 6\}.$$

Logo, $P(A) = \frac{3}{6} = \frac{1}{2}$.

O resultado nos diz que ao lançarmos um dado não viciado ou em perfeitas condições, há "0,5 de chance" de que o número na face de cima seja par.

b) Um casal deseja ter dois filhos. Vamos determinar o espaço amostral Ω e calcular a probabilidade do evento B "nascer dois meninos".

Denotando H por menino e M por menina, temos:

$$\Omega = \{(H, H), (H, M), (M, M), (M, H)\},$$
$$B = \{(H, H)\}.$$

Logo, $P(B) = \frac{1}{4}$.

Ou seja, há 0,25 de chance de o casal ter dois filhos homens.

Com base nos exemplos acima, sendo $n(\Omega) = n$, concluímos que:

- A probabilidade do evento certo equivale a 1: $P(\Omega) = 1$;
- A probabilidade do evento impossível é igual 0: $P(\emptyset) = 0$;
- A probabilidade de um evento E elementar é $P(E) = \frac{1}{n}$.
- A probabilidade de qualquer evento E é o número $P(E)$: $0 \leq P(E) \leq 1$;

4.1.5 Problemas resolvidos

Problema 1 – Ao lançarmos três moedas simultaneamente, qual é a probabilidade de que ambas caiam com a mesma face para cima?

Resolução:

Utilizando o princípio fundamental da contagem pode-se determinar todos os cenários possíveis considerando-se o lançamento de três moedas.

Em cada lançamento podemos ter apenas dois resultados possíveis, então três moedas produzirão 8 resultados distintos, que compõem o espaço amostral.

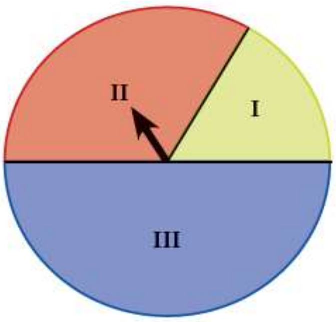
A única situação em que ocorre de todos os lançamentos apresentarem resultados idênticos é aquela na qual ambas será cara ou será coroa, o que significa 2 possibilidades e, então, a probabilidade será dada por:

$$P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{2}{8} = \frac{1}{4} = 0{,}25.$$

Logo, a probabilidade de que se tenha o mesmo resultado nos três lançamentos equivale a 0,25.

Problema 2 – Considere uma roleta circular com um ponteiro central móvel. Ao girarmos livremente esse ponteiro, ele vai parar em uma determinada região da roleta. Na roleta representada pela Figura 5, o ângulo correspondente ao setor I mede $60°$ e o correspondente ao setor III, 180°. Calcule a probabilidade de o ponteiro da roleta, ao ser girado livremente, parar na região II. (SÃO PAULO, 2014a).

Figura 5 – Roleta circular.



Fonte: Caderno do aluno: matemática, ensino fundamental – 8ª série, volume 2. (2014)

Resolução:

A área do setor II corresponde ao ângulo central de $120°$. Não é necessário calcular a área, pois o raio é o mesmo para cada setor. Assim, basta comparar o ângulo correspondente ao setor II com $360°$. $P(II) = \frac{120}{360} = \frac{1}{3} = 0{,}333\ldots$. Logo, a probabilidade de o ponteiro parar na região II é de aproximadamente $0{,}333\ldots$

4.2 A DISTRIBUIÇÃO NORMAL

4.2.1 A história da curva normal

"A curva normal desempenha um papel central em Estatística" (CRILLY, 2014, p. 140). Sua história teve origem há muitos anos e sua história está ligada ao estudo da probabilidade no século XVI, que surgiu inicialmente para resolver problemas de jogos de azar.

Segundo o referencial teórico de Crespo (1997), a curva normal foi muito estudada por Abraham de Moivre, dando sequência aos trabalhos de Jacob Bernoullli e seu sobrinho Nicolau Bernoulli. Seus estudos foram de grande valia, pois posteriormente, homens como Laplace, em 1783, a utilizou para representar a distribuição de erros, e Gauss, em 1809, fez uso da mesma em análises de dados astronômicos.

"A curva normal é determinada por uma fórmula matemática específica que cria uma curva em forma de sino, uma curva com uma bossa que vai diminuindo de cada um dos lados" (CRILLY, 2014, p. 140).

Esse resultado é muito usado hoje na matemática e Estatística, pois a normalidade ocorre em muitas medidas físicas, biológicas e sociais. Vários cientistas do século XVIII perceberam que ao mensurar determinadas medidas, tais como a distância à Lua, esta apresentava uma variação muito semelhante a esta forma gráfica. Tão logo, a mesma foi relacionada aos erros de mensuração, donde surgiu a denotação "Distribuição normal dos erros" e pouco mais tarde "Distribuição normal". Essa mesma curva também é conhecida por "Distribuição Gaussiana", devido ao modelo matemático desenvolvido por Karl F. Gauss, de acordo com Crilly (2014).

4.2.2 Variável aleatória

"Tomando-se, por base um experimento E e um evento e Ω o espaço associado ao experimento, a função F, que associe a cada elemento $s \in \Omega$ um número real $F(s)$, é denominada variável aleatória" (CORREA, 2003, p. 78).

Variável aleatória discreta.

Definição: Uma função X, definida no espaço amostral Ω e com valores num conjunto enumerável de pontos da reta é dita variável aleatória discreta. (Morettin, 2013)

<u>Variável aleatória contínua</u>

<u>Definição</u>: Uma função X, definida no espaço amostral S e com valores num intervalo de números reais é dita variável aleatória contínua. (Morettin, 2013).

Dessa forma, considerando o espaço amostral ao lançarmos simultaneamente duas moedas, o mesmo corresponde a $\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\}$. Como F representa o número da face coroa , então cada ponto de Ω associa-se a um número de F, de acordo com a Tabela 2.

Tabela 2 – Variável aleatória número de coroas no lançamento simultâneo de duas moedas.

Ponto Amostral	F
(Ca, Ca)	0
(Ca, Co)	1
(Co, Ca)	1
(Co, Co)	2

Fonte: Adaptado de Estatística fácil (1997)

Assim podemos escrever o que consta na Tabela 3.

Tabela 3 – Probabilidade de F.

Número de coroas (F)	Probabilidade
2	¼
1	2/4
0	¼
Total	4/4 = 1

Fonte: Adaptado de Estatística fácil (1997)

Vamos agora para outro exemplo de probabilidade.

Neste caso, em particular, vamos considerar a distribuição de frequências referente ao índice de acidentes que ocorrem diariamente na Rodovia Comandante Manoel Antônio, no período equivalente a um mês do ano de 2008, conforme está disposto na Tabela 4.

Tabela 4 – Distribuição da frequência relativa ao número de acidentes diários.

Número de Acidentes	Frequência
0	18
1	7
2	4
3	1

Fonte: Adaptado de Estatística fácil (1997)

Na Tabela 5, temos representada a distribuição de probabilidade.

Tabela 5 – Distribuição de probabilidade.

Número de Acidentes	Probabilidade
0	0,60
1	0,23
2	0,13
3	0,03
Total	1,00

Fonte: Adaptado de Estatística fácil (1997)

Na Tabela 4 temos a frequência da variável aleatória discreta “número de acidentes” e na Tabela 5 a distribuição de probabilidades de ocorrência da mesma variável.

Seja X uma variável aleatória cujo valores são dados por $x_1, x_2, x_3, \dots, x_n$. Cada um dos valores x_i estão relacionados a pontos do espaço amostral. Além disso, vamos associar a cada valor x_i, também, a probabilidade p_i de ocorrências de tais pontos no espaço amostral.

Como consequência, temos:

$$\sum_{i=1}^{n} p_i = 1.$$

Os valores $x_1, x_2, x_3, \dots, x_n$ e seus correspondentes $p_1, p_2, p_3, \dots, p_n$ definem uma distribuição de probabilidades.

Conforme Crespo (1997), a definição de distribuição de probabilidade estabelece uma correspondência unívoca entre os valores da variável aleatória discreta X e entre seus respectivos valores de probabilidade $P(X = x_i)$. Isto define uma função na qual o domínio é composto pelos valores x_i, e os valores p_i definem o conjunto imagem. Consequentemente, no lançamento de um dado, os pontos tirados definem a variável aleatória X, que pode assumir os valores $1, 2, 3, 4, 5$ e 6. Daí segue a distribuição de probabilidade, conforme apresentado na Tabela 6.

Tabela 6 – Distribuição de probabilidade no lançamento de um dado.

X	$P\ (X)$
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$
Total	1

Fonte: Estatística fácil (1997)

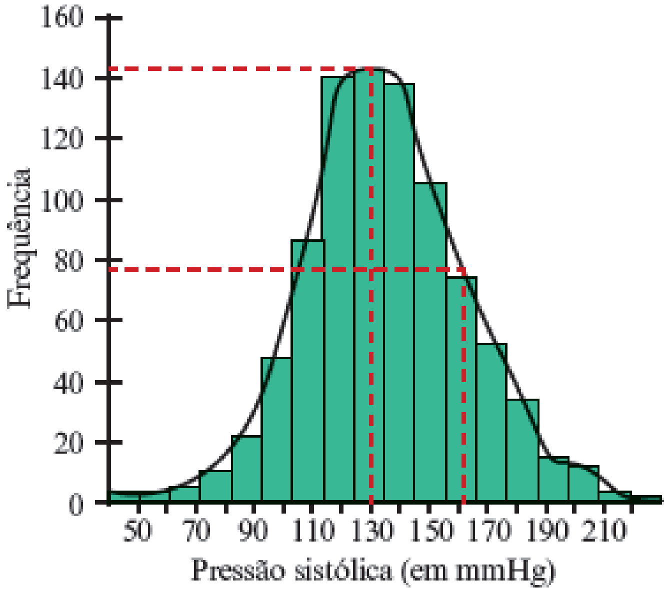
4.2.3 Distribuição normal

Baseando-se, em Correa (2003), "a distribuição normal é a mais importante das distribuições de probabilidade; denominada de "curva em forma de sino", tendo sua origem atrelada aos erros de mensuração".

Como propriedades da distribuição normal, de acordo com Crespo (1997), a primeira delas garante que a variável aleatória X assume um valor real qualquer. Além disso, a representação gráfica da distribuição normal recebe o nome de curva normal ou curva de Gauss, que é caracterizada por sua simetria em torno da média e por seu comportamento assintótico em relação ao eixo das abscissas. A área total da região entre a curva e o eixo das abscissas é igual a 1, representando a probabilidade de a variável aleatória X assumir um valor real no intervalo $(-\infty, \infty)$. Ao mesmo tempo, a simetria da curva implica numa mesma probabilidade de ocorrer valor maior ou menor do que a média, o que significa que cada metade da curva representa 0,5 de probabilidade.

Considerando-se, ao acaso, um grupo composto por 900 pessoas de uma cidade qualquer para medir a pressão arterial de cada uma delas, desenhando um gráfico com os resultados, obter-se-ia, sem dúvida, algo igual ou muito semelhante ao que está representado na Figura 6.

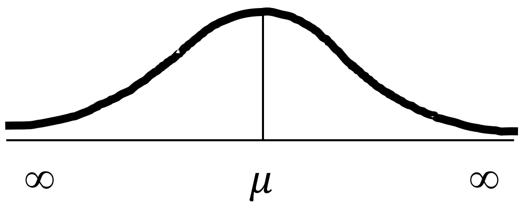
Figura 6 – Exemplo de distribuição normal.



Fonte: Caderno do aluno: matemática, ensino médio – 3ª série, volume 2. (2014)

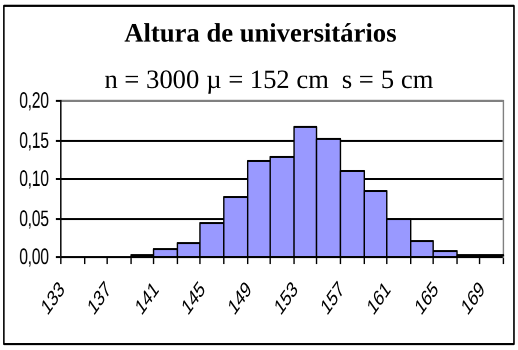
De acordo com Crespo (1997), quando se estuda uma variável aleatória, queremos observar as chances dessa variável aleatória assumir um valor dentro de um intervalo determinado. A distribuição normal possui um aspecto gráfico como o da Figura 7, onde μ é a esperança ou valor médio da distribuição.

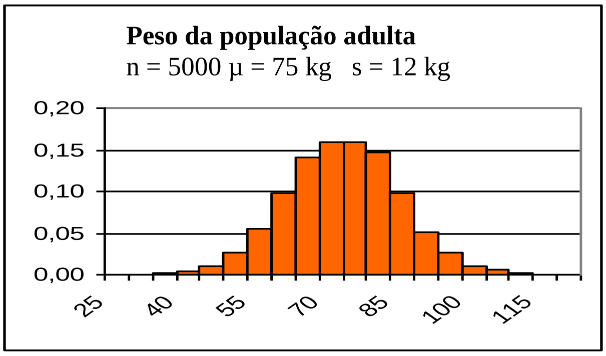
Figura 7 – Aspecto gráfico da distribuição normal.



Fonte: Adaptado de Estatística fácil (1997)

Figura 8 – Exemplos de uma variável aleatória com distribuição normal.





Fonte: http://www.matematiques.com.br/

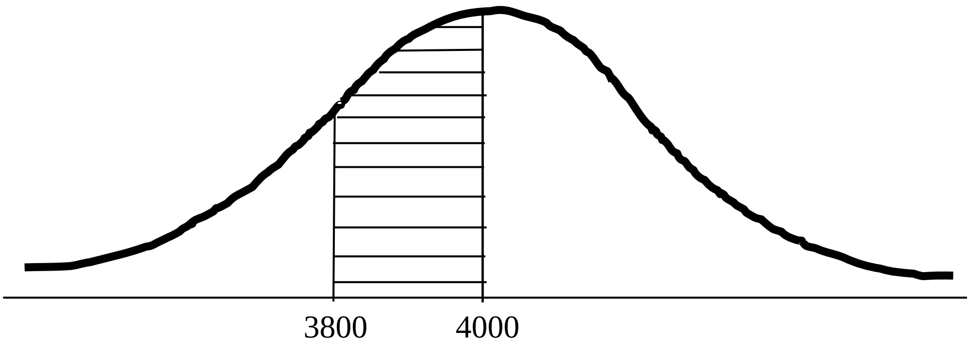
Veja o seguinte exemplo ofertado por Crespo (1997):

Depois de 28 dias de curagem, o cimento apresenta uma resistência compressiva média de $4000\,psi$. Suponha que a resistência obedece a uma distribuição normal com desvio-padrão equivalente a $120\,psi$.

Determine a probabilidade de se comprar um pacote de cimento cuja resistência compressiva de 28 dias esteja entre 3800 e $4000\,psi$.

Percebe-se facilmente que essa probabilidade, indicada por $P(3800 < x < 4000)$, corresponde à área em destaque na Figura 9.

Figura 9 – Representação da probabilidade do exemplo em questão.



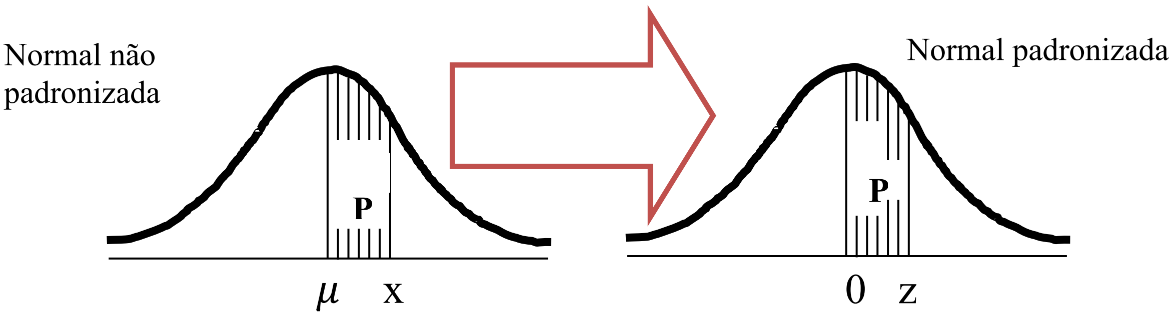
Fonte: Estatística fácil (1997)

Com base em Crespo (1997) o cálculo direto dessa probabilidade exige ferramentas mais avançadas. Mas considerando X uma variável aleatória que possui distribuição normal com média μ e assume um desvio padrão δ, então pode-se obter a variável

$$Z = \frac{X - \mu}{\delta}.$$

Deste modo, segundo as proposições deste autor, esta apresenta distribuição normal reduzida, que é a distribuição normal com média zero e desvio padrão 1. As probabilidades associadas à distribuição normal podem ser identificadas em tabelas que são encontradas em vários textos de estatística (vide Anexo ou Morettin, 2013). Trata-se de uma tabela de distribuição normal reduzida, uma vez que fornece o valor da probabilidade de z' assumir um valor arbitrário entre a média, que é 0, e um valor z determinado, isto é, $P(0 < z' < Z)$, onde z significa a distância entre a média e um ponto qualquer – números de desvios padrões.

Figura 10 – Curvas normais padronizadas e não padronizadas.



Fonte: http://www.matematiques.com.br/

Utilização da Tabela Z: O primeiro passo é procurar na tabela Z o valor de $Z = \frac{3800-4000}{120} = 1{,}666$ ....

Na primeira coluna estão dispostos os valores até uma casa decimal, e com efeito **1,6**. Logo em seguida, procuramos o número 6 na primeira linha, que corresponde ao último algarismo do número **1,66**. Agora buscamos o valor presente na intersecção entre a linha e a coluna correspondentes, o que nos fornece o valor **0,4515**, e então podemos escrever:

$P\ (0\ <\ Z\ <\ 1{,}25\ )\ = 0{,}4515$, que expressa a probabilidade de se comprar um pacote de cimento com resistência de 28 dias entre 3800 e 4000 psi.

Veja um outro exemplo:

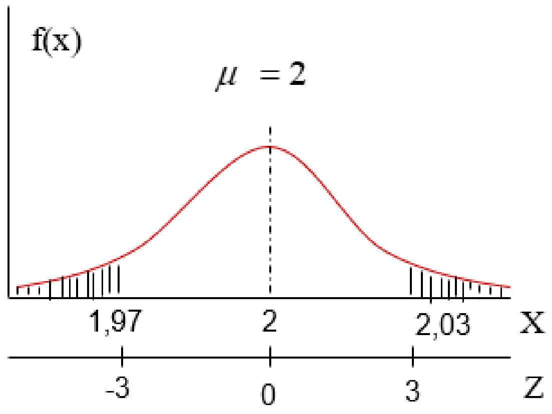
Uma máquina produz peças cujas características descrevem um diâmetro médio de 2,00" e desvio-padrão de 0,01". As peças que estão distantes mais de 0,03" da média são consideradas defeituosas. Qual é a porcentagem de peças defeituosas?

A Figura 11 abaixo contém a curva normal representativa do problema acima.

Em termos matemáticos, resolver este problema significa que devemos determinar o valor de $P([X > 2{,}03]\text{ou } [X < 1{,}97])$. Por isso, calculamos:

$$z_1 = \frac{x-\mu}{\sigma} = \frac{2{,}03-2}{0{,}01} = 3,$$
$$z_2 = \frac{x-\mu}{\sigma} = \frac{1{,}97-2}{0{,}01} = -3,$$

Figura 11 – Curva normal.



Fonte: http://www.matematiques.com.br/

Com base no que está representado acima e no que já foi calculado, a probabilidade que se procura é determinada fazendo:

$$P([X > 2{,}03]\text{ou } [x < 1{,}97]). = 1 - P(1{,}97 \leq x \leq 2{,}03) =$$
$$= 1 - P(-3 \leq z \leq 3) = 1 - \big(P(0 < z < 3) + P(-3 < z < 0)\big) =$$
$$= 1 - (0{,}4986 + 0{,}4986) = 1 - 0{,}9972 = 0{,}0028.$$

Portanto, a porcentagem de peças defeituosas produzidas equivale a 0,28%.

5 INTERVALOS DE CONFIANÇA

Conforme Rumsey (2014, p. 38) “o intervalo de confiança é a forma com que o estatístico abrange seu palpite quando precisa estimar um parâmetro populacional”.

Ampliando-se tais considerações:

> Intervalo baseado em observações de uma amostra e construído de maneira que haja uma probabilidade específica do intervalo conter o verdadeiro valor desconhecido de um parâmetro (em geral calculamos intervalos de confiança que tenham uma chance de 95% de conter o verdadeiro valor). (DOWING; CLARK, 2010, p. 155).

Neste sentido, “a grande sacada do intervalo de confiança é a de apresentar um conjunto de possíveis valores para um parâmetro populacional” (RUMSEY, 2014, p. 38).

Consequentemente, “o nível de confiança representa a probabilidade de se obter um conjunto de possíveis valores que realmente contenha o parâmetro populacional real, caso fosse repetido o processo de amostragem várias e várias vezes” (RUMSEY, 2014, p. 38).

Tem-se o grau de confiança como a “probabilidade de o intervalo conter o verdadeiro valor do parâmetro” (RUMSEY, 2014, p. 155).

Os institutos de estatística costumam acrescentar uma margem de erro aos resultados obtidos em pesquisas estatísticas sobre intenções de voto de uma determinada eleição, por exemplo. Por isso, quando se diz que determinado candidato tem “$35{,}5\%$ das intenções de voto, com 5% de margem de erro para mais ou para menos” significa que o candidato em questão deve obter entre $(35{,}5 - 5)\%$ e $(35{,}5 + 5)\%$ dos votos, isto é, entre $29{,}5\%$ e 41%.

Essa margem de erro, que foi estabelecida, refere-se ao grau de precisão em que o órgão que realizou as pesquisas desejou. Para se definir o grau de precisão de uma pesquisa, precisa-se avaliar o **intervalo de confiança** do resultado. Para entender como são definidos esses intervalos de confiança precisamos retomar a interpretação da curva normal, com a relação entre média aritmética e desvio padrão da amostra.

Considerando uma amostra casual simples com n elementos, uma estimativa da média da população pode ser entendida como a média dos valores observados. Encontrando um intervalo de confiança, podemos ter uma melhor estimativa. Em casos como estes, quanto maior o nível de confiança, maior será o intervalo. Em contrapartida, quanto maior o intervalo, menor será a precisão da estimativa.

Baseando-se nas proposições de Bonafini (2015), o intervalo observado pode subestimar ou superestimar a média da população, contexto diante do qual o intervalo de confiança de 95% constitui-se a faixa provável do verdadeiro parâmetro desconhecido.

Tem-se, com base nas lições deste autor que:

> O intervalo de confiança não reflete a variabilidade no parâmetro desconhecido; pelo contrário, reflete a quantidade de erro aleatório na amostra e fornece um intervalo de valores que são suscetíveis de incluir o parâmetro desconhecido. Outra forma de pensar sobre um intervalo de confiança (definido como a estimativa pontual ± margem de erro) é considerá-lo como um intervalo de valores suscetíveis de conter o parâmetro com um determinado nível de confiança (o qual é semelhante a uma probabilidade). (BONAFINI, 2015, p. 167).

É possível de tal modo, definir um intervalo de confiança de nível c para a média populacional μ, através da probabilidade do evento:

$$\bar{X} - E < \mu < \bar{X} + E.$$

Consequentemente, “a probabilidade de que o intervalo de confiança contenha μ é c (LARSON; FARBER, 2010; BONAFINI, 2015, p. 167)”.

5.1 TEOREMA DO LIMITE CENTRAL

De acordo com Navidi (2012, p. 167), o “Teorema do Limite Central é de longe o mais importante resultado em Estatística”.

“O Teorema do Limite Central diz que se extraímos uma amostra suficientemente grande de uma população, então a distribuição da média amostral é aproximadamente normal, não importando de qual população a amostra foi extraída” (NAVIDI, 2012, p. 167).

Isso quer dizer que, retiradas amostras – com 30 elementos ou mais – de uma população, mesmo que os dados da população não sejam normalmente distribuídos, ainda assim, a distribuição amostral das médias das amostras terá uma distribuição aproximadamente normal.

Além disso, à medida que aumenta-se o tamanho da amostra, a aproximação torna-se cada vez melhor.

Exemplo: Em uma amostra aleatória coletada de forma representativa de uma escola, a média aritmética da altura das alunas foi igual a $1,65\ m$, e o desvio padrão da altura é igual a $0{,}07\,m$. Nessas condições, queremos determinar a faixa de valores da altura em que se encontram 90% das alunas da escola. Quais são os limites dessa faixa?

Reparamos que, neste caso, não há um tamanho amostral revelado. Podemos, inicialmente, dividir pela metade o porcentual fixado, de 90%, imaginando que procuramos uma faixa que compreenda 45% acima e 45% abaixo da média. Consultando a tabela no sentido oposto ao que percorremos anteriormente, vamos considerar o valor 0,4505, o mais próximo de 0,45, ou 45%. A leitura da tabela nos informa que a probabilidade de 0,45 corresponde a 1,65 desvio padrão. Como o desvio padrão, no caso do exemplo, é igual a $0{,}07m$, tem-se:

$$1{,}65 * 0{,}07 = 0{,}1155\,m,$$

e subtraindo este valor de 0,1155 m da média dada de $1{,}65\,m$, temos:

$$1{,}65 + 0{,}1155 = 1{,}7655 \cong 1{,}77\,m,$$
$$1{,}65 - 0{,}1155 = 1{,}5345 \cong 1{,}53\,m.$$

Portanto, a probabilidade de sortearmos uma aluna qualquer da escola com altura entre $1{,}53\,m$ e $1{,}77\,m$ é igual a 0,9, supondo, desde sempre, que a distribuição da altura é normal.

Deveríamos realizar uma discussão menos resumida para o Teorema do Limite Central como a feita acima. Este resultado não é tão simples quanto foi apresentado em Navidi (2012).

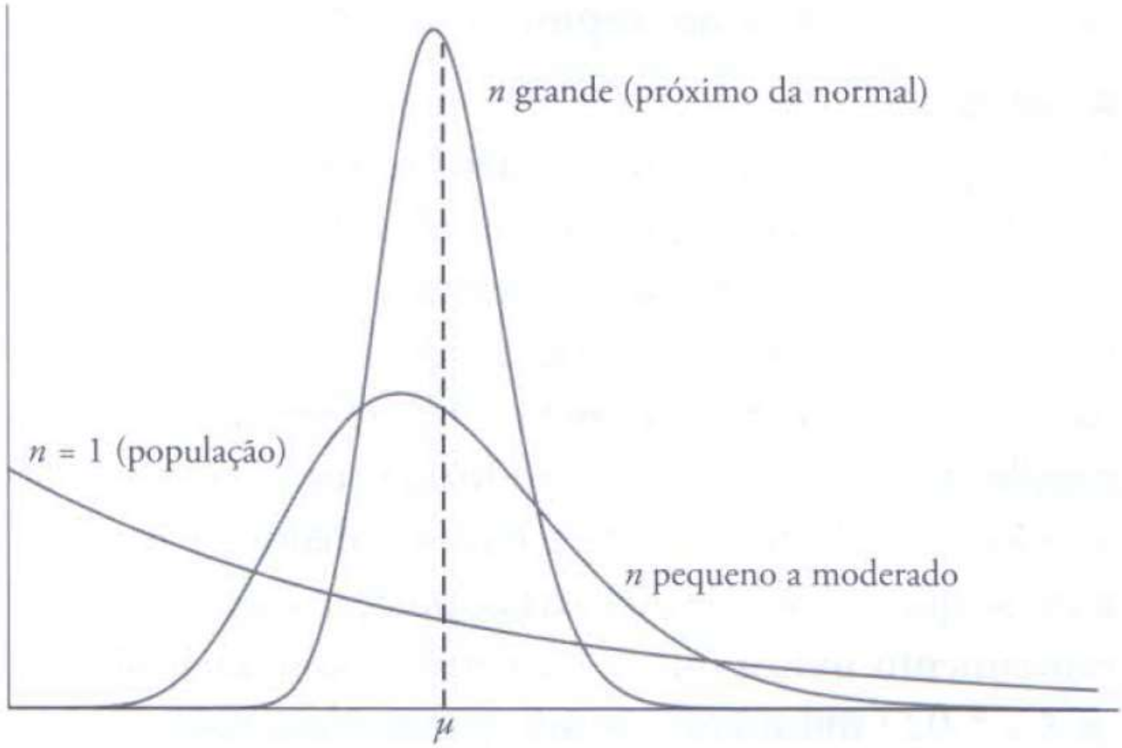
Segundo Walpole (2009) os estatísticos recomendam que, ainda que não se possa assumir uma distribuição normal e, apesar de σ desconhecido, para $n \geq 30$, s pode ser assumido para estimar σ e o intervalo de confiança 1-α de forma que

$$\bar{x} \pm Z_{1-\alpha/2} . \frac{s}{\sqrt{n}}$$

que poderá ser utilizado para μ. Geralmente este é denominado como intervalo de confiança em amostra grande. Uma justificativa aceitável para sua utilização é que para uma amostra maior ou igual a 30, mesmo com uma distribuição de população não muito assimétrica, s estará muito próximo de σ. Devemos ressaltar que esta situação é apenas uma aproximação e que sua qualidade será melhor quanto maior for o tamanho amostral.

Uma figura dada em Walpole (2009) nos revela graficamente como esta aproximação se torna mais eficaz quando o tamanho amostral *n* cresce.

Figura 12 – Ilustração do efeito do aumento de *n*.



Fonte: Walpole, 2009.

O exemplo anterior utilizou dados com distribuição normal e com tamanho amostral desconhecido, contudo, vamos supor que não tenho a hipótese da normalidade, mas conheço o tamanho amostral e estimamos *s*, digamos *n*=40 com *s*=0,07. Então, podemos fazer

$$\bar{x} \pm Z_{1-\alpha/2}.\frac{s}{\sqrt{n}} = 1{,}65 \pm 1{,}65.\frac{0{,}07}{\sqrt{40}} \cong 1{,}65 \pm 0{,}01 => (1{,}64\,;1{,}66)$$

Logo o intervalo com 90% de confiança para a altura das alunas fica (1,64 ; 1,66), portanto, bem menor que o anterior.

Para este trabalho esta situação nos parece válida, um resultado que pode ser aplicado em situações práticas e que não requer conhecimentos tão apurados, como é a situação para o ensino médio. O que se pretende é originar usuários da estatística com aplicações mais simples e motivá-los a estudar os temas relacionados e, consequentemente, obter um aprofundamento futuro no assunto.

6 APLICAÇÕES

As aplicações de estatística são extremamente importantes para motivar o seu estudo. Como já fora discutido anteriormente, as aplicações podem ser realizadas em várias situações, em diferentes campos de conhecimento. São apresentadas neste trabalho duas aplicações com as técnicas expostas.

6.1 PRIMEIRO EXEMPLO

Análise das Notas dos Alunos da Escola Estadual Dom Bosco

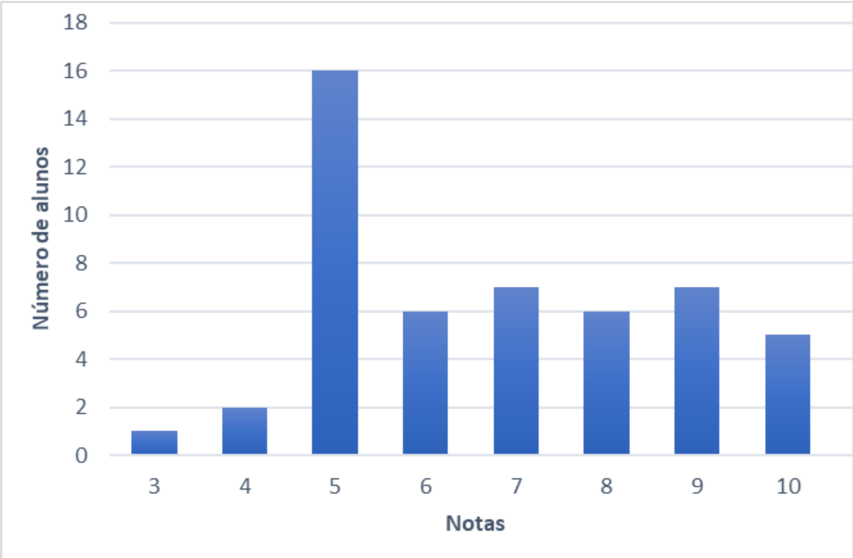
A seguir será feita uma análise do desempenho dos alunos da Escola Estadual Dom Bosco, localizada no município de Osvaldo Cruz, SP. São analisadas as notas de Matemática de alunos do 6º ao 9º ano do Ensino Fundamental, e também de alunos do 1º ao 3º ano do Ensino Médio, em dois bimestres do ano letivo de 2018.

Ensino Fundamental:

Neste nível foram coletadas as notas em dois bimestres dos alunos matriculados no 6º ao 9º ano do Ensino Fundamental. Na escola há duas turmas para cada série, exceto para o 7º ano: 6º ano A, 6º ano B, 7º ano A, 8º ano A, 8º ano B, 9º ano A e 9º ano B. Para facilitar o estudo foram gerados gráficos para a situação de cada série nos bimestres escolhidos.

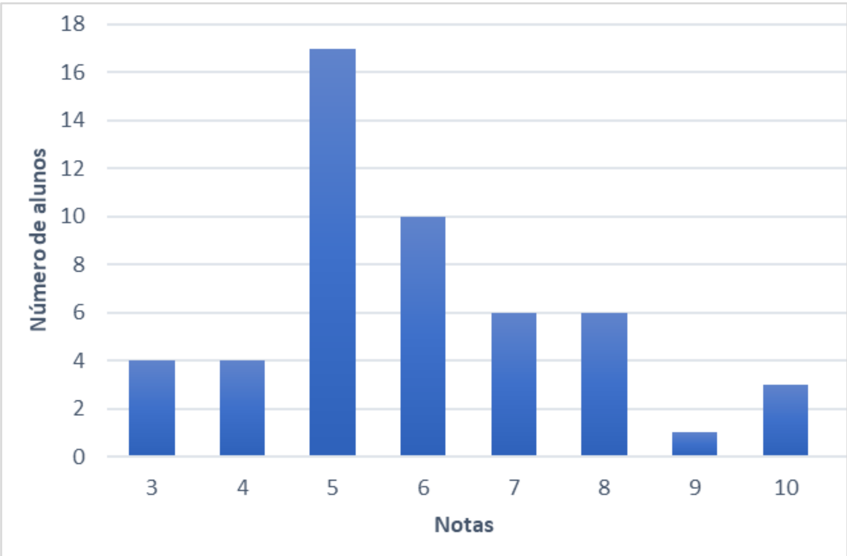
Observando os gráficos relativos ao 6º ano, presentes nas Figuras 13 e 14, verifica-se que os alunos apresentam um desempenho razoável, e que em ambos os bimestres grande parte da turma teve média 5. Além disso, no 2º Bimestre a quantidade de alunos que apresentaram rendimento insuficiente – inferior a 5 – aumentou, enquanto que o número de alunos que tiveram médias mais altas - entre 8 e 10 – diminuiu.

Figura 13 – Distribuição dos alunos do 6º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

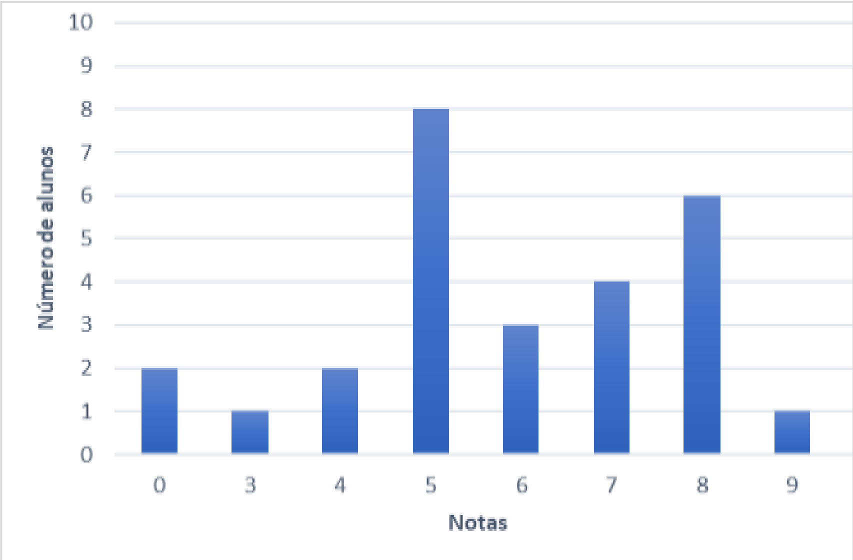
Figura 14 – Distribuição dos alunos do 6º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

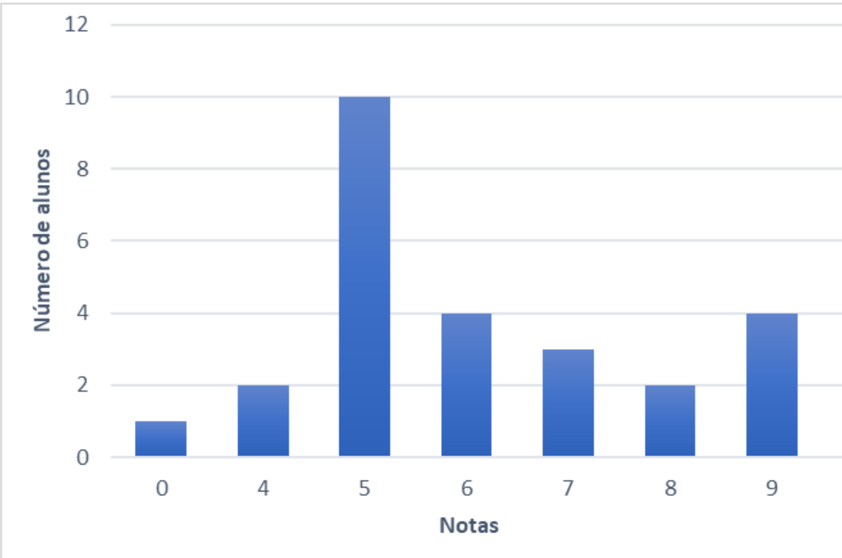
Sobre o 7º ano, uma análise simples das Figuras 15 e 16 permite observar que a maioria dos alunos, no 1º Bimestre, tiveram conceitos entre 5 e 8. Apenas um aluno teve nota 9 e nenhum deles atingiu o conceito máximo, embora 2 alunos tenham tirado zero. No 2º Bimestre a situação ficou um pouco diferente, embora o número de alunos com rendimento insuficiente tenha sido reduzido passando de 5 para 3 alunos, houve redução significativa de alunos que obtiveram nota 8. A quantidade de alunos com conceito 9 também aumentou, mas ainda assim nenhum aluno obteve nota 10.

Figura 15 – Distribuição dos alunos do 7º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

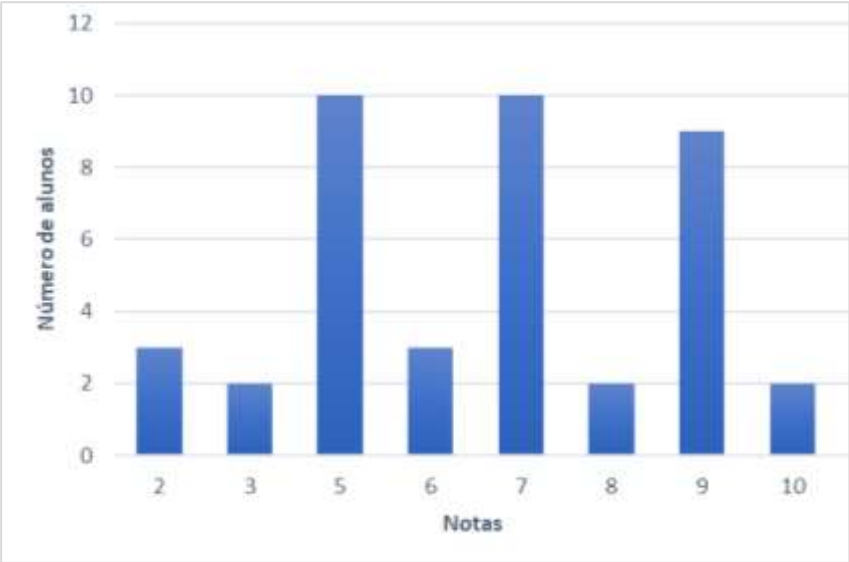
Figura 16 – Distribuição dos alunos do 7º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

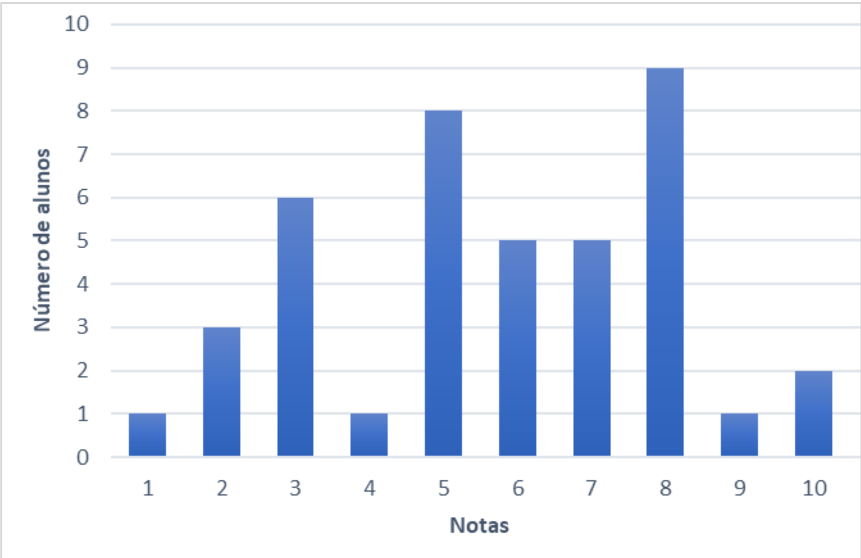
A 8ª turma apresentou um perfil bastante satisfatório no 1º Bimestre, com apenas 5 alunos tendo média inferior a 5. Porém este cenário é transformado no 2º Bimestre, quando 10 alunos passam a ter desempenho inferior ao necessário. Ao mesmo tempo, a quantidade de alunos com nota 9 diminui expressivamente, e grande parte da turma apresenta um desempenho regular, comprovado pelos altos índices de alunos que obtiveram conceitos 5, 6 e 7, conforme consta nas Figuras 17 e 18.

Figura 17 – Distribuição dos alunos do 8º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

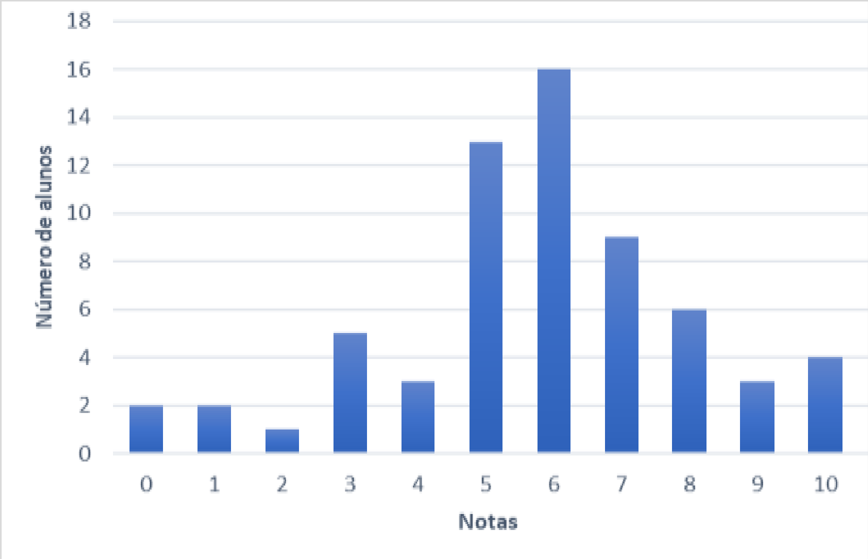
Figura 18 – Distribuição dos alunos do 8º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

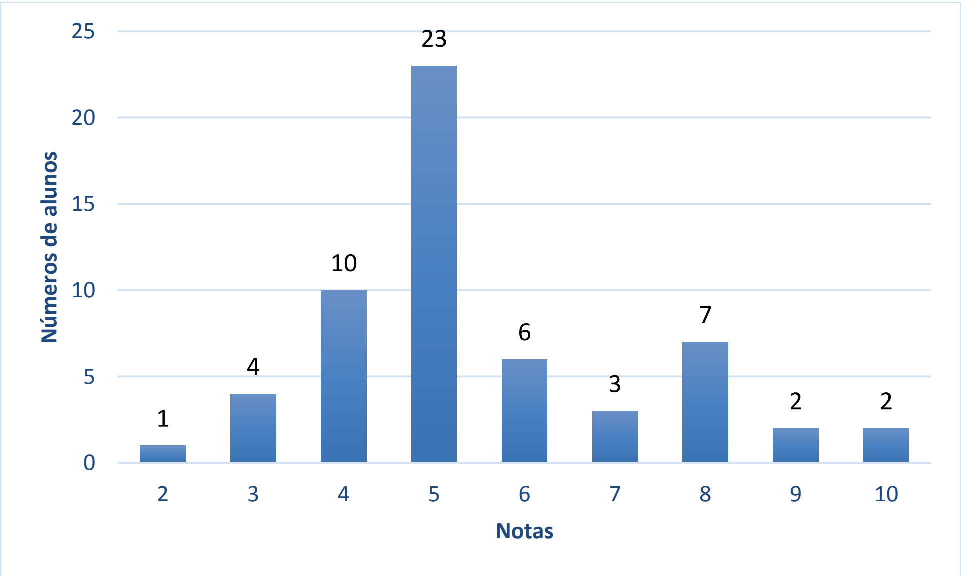
O que se percebe com relação ao 9º ano, a partir das Figuras 19 e 20, é que houve um decaimento do rendimento dos alunos. No 1º Bimestre o desempenho da grande parte dos estudantes se mostrou razoável, mas no 2º Bimestre a distribuição das notas revela que os alunos passaram, em sua maioria, a obter conceito 5, além de que boa parte não tenha nem conseguido atingir esta nota. A quantidade de alunos com notas mais altas também diminuiu entre os bimestres.

Figura 19 – Distribuição dos alunos do 9º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

Figura 20 – Distribuição dos alunos do 9º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

A principal conclusão que se pode tirar das análises apresentadas é que todas as turmas apresentaram um desempenho satisfatório, de um modo geral. Isso pode ser comprovado, ao analisarmos a Tabela 7, onde está presente a média aritmética das notas de cada turma.

Observa-se que as médias estão compreendidas entre 5 e 7. As únicas turmas que apresentaram uma pequena melhora de desempenho foram o 7º ano A e o 9º ano B, observando ainda que, mesmo assim, as suas médias não ultrapassaram o conceito 6. Todas as outras turmas pioraram seu desempenho no 2º Bimestre, mas nota-se que dentro deste cenário, a turma que apresentou melhor desempenho geral foi o 6º ano A.

Tabela 7 – Média aritmética das notas.

	6ºA	6ºB	7ºA	8ºA	8ºB	9ºA	9ºB
1º Bimestre	6,7407	6,7083	5,7037	6,3809	6,7619	6,3437	5,2187
2º Bimestre	6,2963	5,5	5,9615	5,7143	5,8571	5,1613	5,9259

Fonte: Elaborado pelo autor

Nas Tabelas 8 e 9 estão presentes as seguintes medidas de dispersão: desvio médio, variância e desvio padrão, referentes às notas dos alunos matriculados no Ensino Fundamental. Analisando os dados do 1º Bimestre, percebe-se que as notas dos alunos são menos dispersas no 6º ano B, isto é, as notas da maioria dos alunos estiveram de fato, próximas da média que é 6,7083. Ao mesmo tempo, as notas são mais dispersas no 9º ano B, isto é, embora a média seja 5,2187, vários alunos podem ter obtido conceitos ou muito maiores ou muito menores que isto.

Já no 2º Bimestre, informações demonstram que as notas dos estudantes foram mais dispersas no 9º ano A, e menos dispersas no 8º ano A.

Tabela 8 – Medidas de dispersão do 1º Bimestre.

	6ºA	6ºB	7ºA	8ºA	8ºB	9ºA	9ºB
Desvio médio	1,7312	1,5417	1,7147	1,8866	1,7233	1,5547	1,9902
Variância	4,1180	2,8733	4,8752	5,1882	4,3719	3,6631	5,9834
Desvio padrão	2,0293	1,6951	2,2080	2,2778	2,0909	1,9139	2,4461

Fonte: Elaborado pelo autor

Tabela 9 – Medidas de dispersão do 2º Bimestre.

	6ºA	6ºB	7ºA	8ºA	8ºB	9ºA	9ºB
Desvio médio	1,6104	1,125	1,5	2,0136	1,8775	0,7867	1,7860
Variância	3,6900	2	3,8831	5,5374	5,0748	1,8127	4,3649
Desvio padrão	1,9209	1,4142	1,9706	2,3532	2,2527	1,3464	2,0892

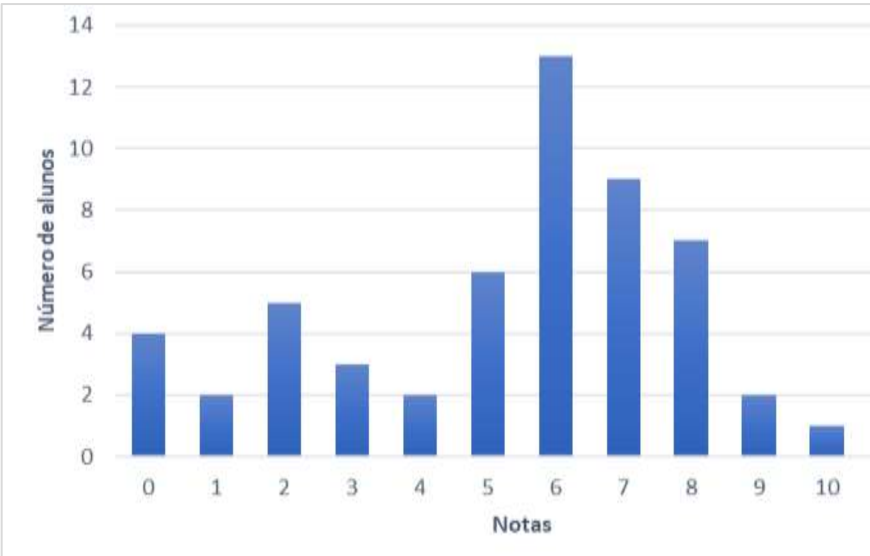
Fonte: Elaborado pelo autor

Ensino Médio:

Neste nível foram coletadas as notas, em dois bimestres, dos alunos matriculados no 1º ao 3º ano do Ensino Médio. Na escola há duas turmas para o 1º e 3º ano e apenas uma turma para o 2º ano: 1º ano A, 1º ano B, 2º ano A, 3º ano A e 3º ano B. Para facilitar o estudo foram gerados gráficos para a situação de cada série nos bimestres escolhidos.

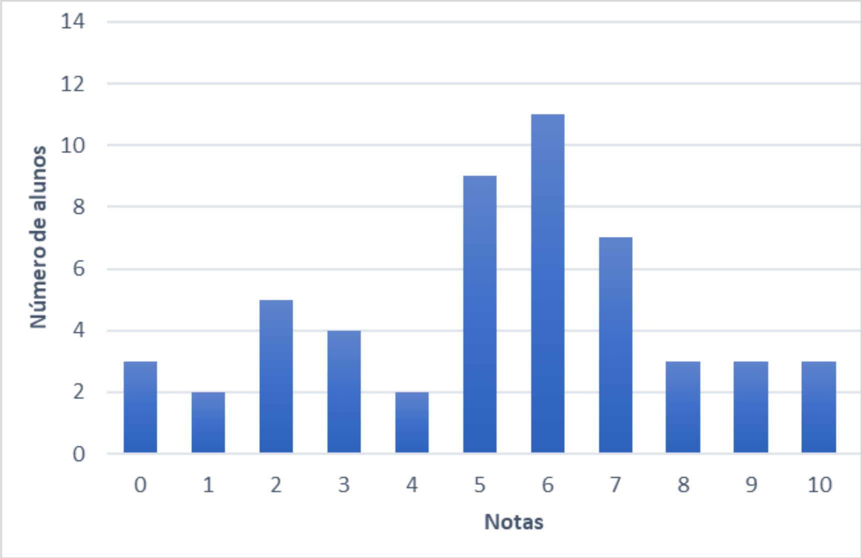
Observando os gráficos relativos ao 1º ano, presentes nas Figuras 21 e 22, verifica-se que os alunos apresentam um desempenho bom, e que em ambos os bimestres grande parte da turma teve conceito compreendido entre 5 e 8. Porém, no 2º Bimestre o desempenho dos alunos foi inferior considerando que mais alunos tiraram 5 e menos alunos obtiveram nota 8.

Figura 21 – Distribuição dos alunos do 1º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

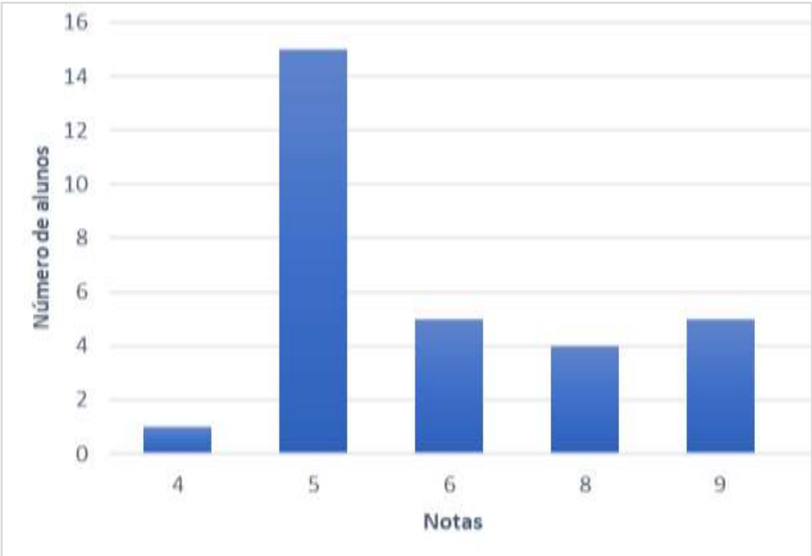
Figura 22 – Distribuição dos alunos do 1º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

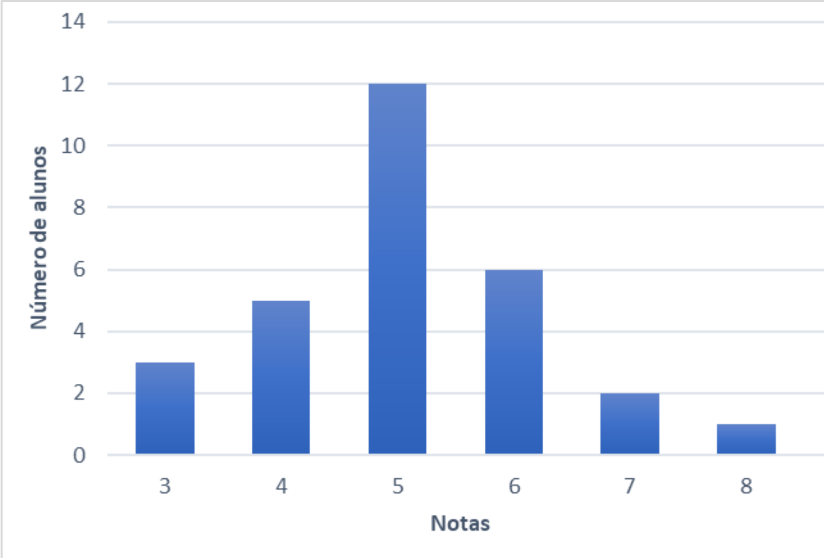
No caso do 2º ano, percebe-se uma piora do desempenho dos alunos, conforme o que está sendo representado nas Figuras 23 e 24. Primeiramente, a nota mais alta no 2º Bimestre foi 8. No 1º Bimestre apenas um dos alunos obteve conceito insuficiente (4). Já no 2º Bimestre esta situação se agravou, pois oito alunos não atingiram o conceito mínimo suficiente.

Figura 23 – Distribuição dos alunos do 2º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

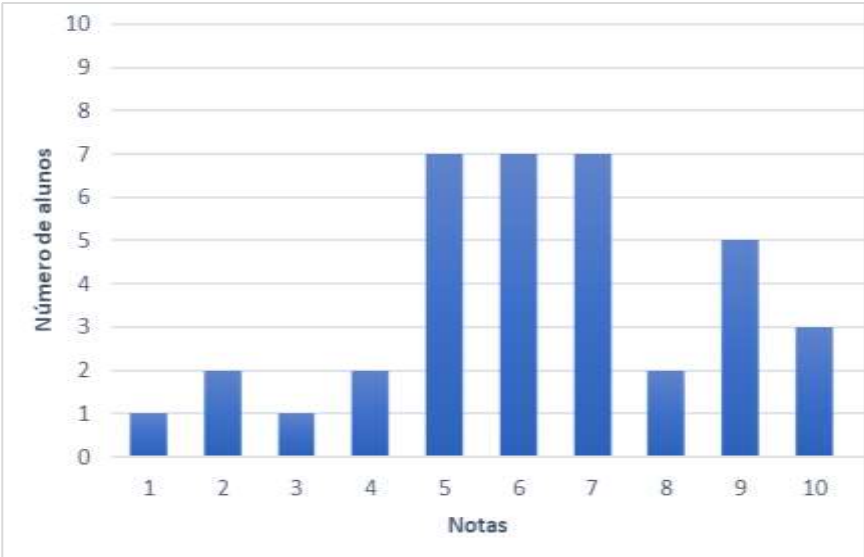
Figura 24 – Distribuição dos alunos do 2º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

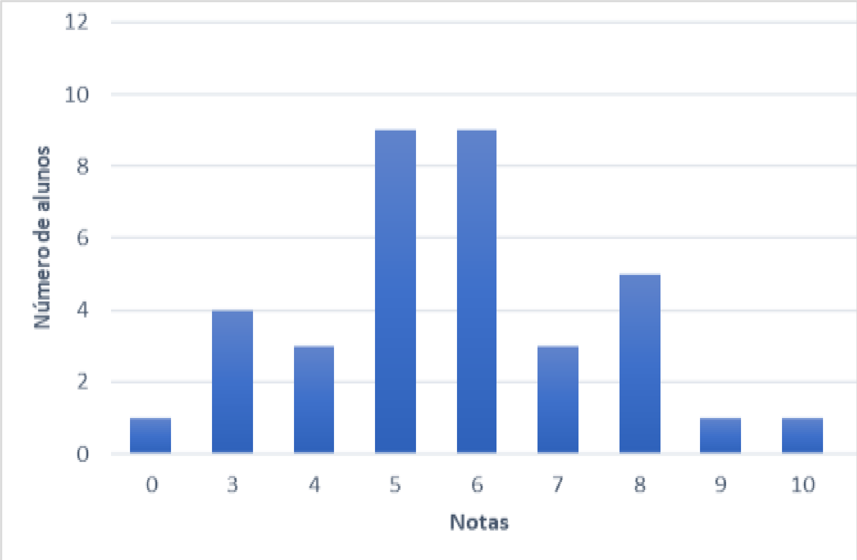
No 3º ano, analisando as Figuras 25 e 26, pode-se notar que os alunos tiveram um bom desempenho no 1º Bimestre. Muito alunos tiveram um conceito regular, entre 5 e 7. Além disso, vários estudantes obtiveram nota 9 e 10. Entretanto, no 2º Bimestre, a quantidade de alunos com notas mais altas diminuiu e, ao mesmo tempo, mais alunos passaram a ter notas mais baixas.

Figura 25 – Distribuição dos alunos do 3º ano por nota no 1º Bimestre.



Fonte: Elaborado pelo autor

Figura 26 – Distribuição dos alunos do 3º ano por nota no 2º Bimestre.



Fonte: Elaborado pelo autor

A principal conclusão que se pode tirar das análises acima se parece muito com o que concluímos na análise das notas dos alunos do Ensino Fundamental. As turmas apresentam um desempenho entre regular e bom. Porém há muitos alunos que não atingiram o conceito 5, que é o mínimo exigido. Isso pode ser comprovado analisando a Tabela 10, onde está presente a média aritmética das notas de cada turma.

Observa-se que as médias estão compreendidas entre 5 e 7. A única turma que melhorou seu desempenho do 1º para o 2º Bimestre foi o 1º ano A, observando ainda que, mesmo assim, as suas médias não ultrapassaram o conceito 6. Quase todas as outras turmas pioraram seu desempenho no 2º Bimestre, mas nota-se que, mesmo assim, o 3º ano A apresentou melhor desempenho, numa visão geral.

Tabela 10 – Média aritmética das notas.

	1ºA	1ºB	2ºA	3ºA	3ºB
1º Bimestre	5,4118	5,2857	6,2000	6,4400	6,1538
2º Bimestre	5,6250	5,0714	5,0690	5,6250	5,6923

Fonte: Elaborado pelo autor

Nas Tabelas 11 e 12 estão presentes as medidas de dispersão, referentes às notas dos alunos matriculados no Ensino Médio. Analisando os dados, percebe-se que, em ambos os bimestres, as notas dos alunos são menos dispersas no 2º ano A, isto é, as notas da maioria dos alunos estiveram de fato, próximas da média aritmética 6,2 no 1º Bimestre e 5,069, no 2º Bimestre. Além disso, as notas são mais dispersas no 1º ano B, isto é, embora a média seja 5,2857 no 1º Bimestre e 5,0714 no 2º Bimestre, vários alunos podem ter obtido conceitos ou muito maiores ou muito menores que isto.

Tabela 11 – Medidas de dispersão do 1º Bimestre.

	1ºA	1ºB	2ºA	3ºA	3ºB
Desvio médio	1,9861	2,1326	1,4133	1,8624	1,5740
Variância	6,3566	6,7041	2,6267	5,5264	3,8225
Desvio padrão	2,5212	2,5892	1,6207	2,3508	1,9551

Fonte: Elaborado pelo autor

Tabela 12 – Medidas de dispersão do 2º Bimestre.

	1ºA	1ºB	2ºA	3ºA	3ºB
Desvio médio	1,7500	2,2806	0,8537	1,6042	1,2071
Variância	5,6096	7,7092	1,3745	4,5677	2,0592
Desvio padrão	2,3684	2,7765	1,1724	2,1372	1,4350

Fonte: Elaborado pelo autor

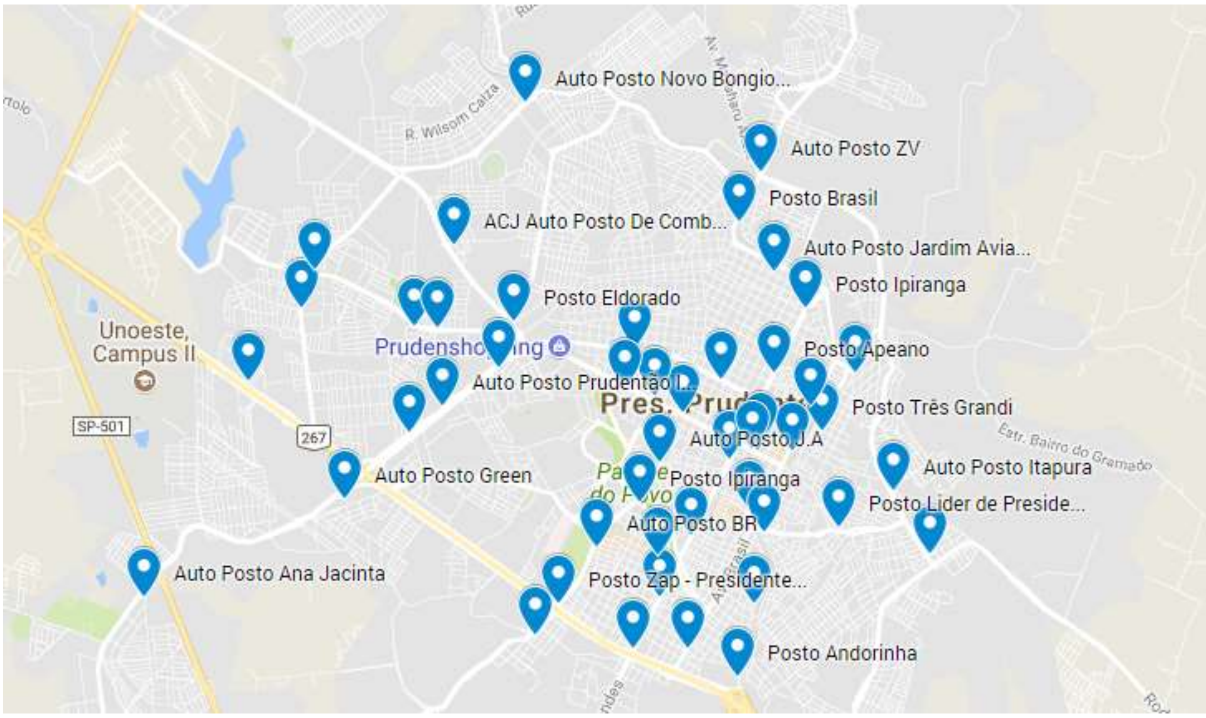
6.2 SEGUNDO EXEMPLO

Análise do preço dos combustíveis em postos de Presidente Prudente

Os preços de combustíveis tem dado relevância a muitas discussões na atualidade, como por exemplo, rendimento e tributos que se pagam por eles e greves por menores preços. Foram realizadas coletas na cidade de Presidente Prudente na época de carnaval, em um curto período (em dois dias) em que parece ter havido relativa estabilidade.

As coletas foram executadas de forma que fossem coletadas a maior quantidade de postos cuja listagem estava em ordem alfabética. Esta estratégia foi interessante pois foi possível distribuir as amostras por toda a cidade (vide figura 27).

Figura 27 – Distribuição espacial dos postos coletados.



Fonte: Google.

Percebe-se, pela figura 27, que esta estratégia apresentou bons resultados. Analisando-se a figura nota-se que os postos estão dispostos nas principais vias arteriais da cidade, onde ocorrem os maiores fluxos de veículos. Este cenário está de acordo com a totalidade dos postos revelando que a coleta foi bem sucedida (estes dados pertencem a um projeto pilôto da FCT/Unesp coordenado pelo prof. Dr. José Gilberto S. Rinaldi e cedidos para este exemplo).

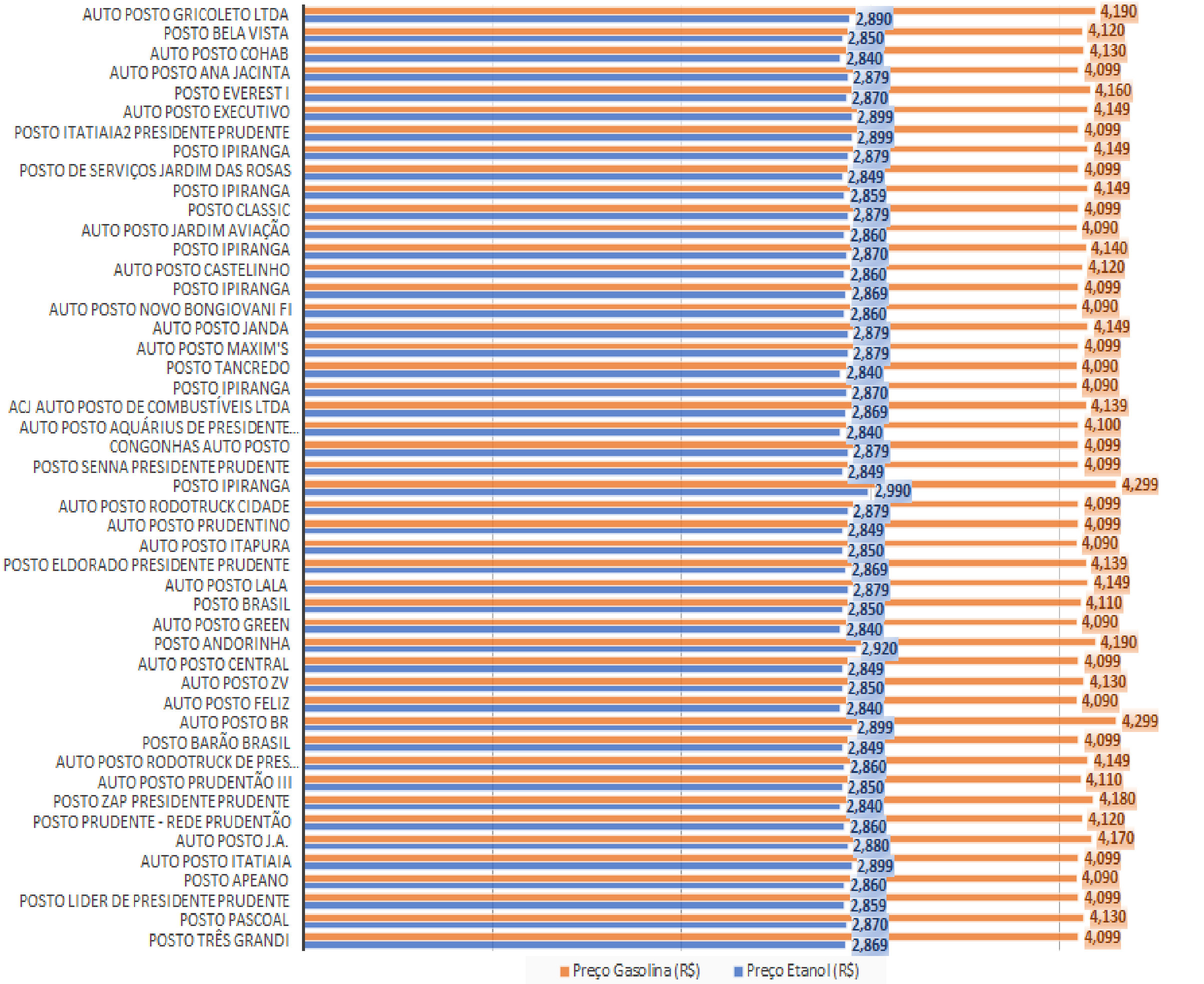
Coletar por meio de ordem alfabética não teve qualquer correlação com as localidades que estarão na amostra, ocasionando uma distribuição espacial abrangente para a coleta.

Pretendendo-se utilizar o teorema do limite central, é interessante coletar o maior número possível de amostras. Também, na realidade, não se sabe a totalidade exata de postos de combustiveis que há em Presidente Prudente por meio de listas telefônicas, há postos que fecharam, outros que abriram e ainda não constam da lista, entre outros casos. Como era de se esperar, houve uma aglomeração da coleta de postos no centro da cidade e no entorno deste.

Foram coletados somente os preços de etanol comum e gasolina comum para que não houvessem tantos produtos a considerar e por serem estes os mais consumidos.

Analisando os preços cobrados por litro de etanol e gasolina em 48 postos de combustíveis em Presidente Prudente, SP, obteve-se os dados apresentados na Figura 27.

Figura 27 – Preço por litro de combustível nos postos de Presidente Prudente, SP.



Fonte: Elaborado pelo autor

A partir desses dados podemos calcular algumas medidas estatísticas que nos ajudam a entender determinados fenômenos. Neste caso em específico, é possível realizar uma análise a fim de determinar quais postos fornecem produtos por preços mais em conta, ou ainda, quando optar por abastecer com gasolina ou com etanol.

A princípio, numa análise superficial identificamos os preços mínimos e máximos pelos quais ambos os combustíveis estão sendo comercializados em Presidente Prudente, conforme segue na Tabela 13.

Tabela 13 – Preços mínimos e máximos.

	Preço mínimo	Preço máximo
Etanol	2,84	2,99
Gasolina	4,09	4,299

Fonte: Elaborado pelo autor

Esses preços foram identificados em alguns dos 48 postos nos quais os dados foram coletados. O etanol pode ser comprado pelo preço mínimo no Posto Zap Presidente Prudente, no Auto Posto Feliz, Auto Posto Green, Auto Posto Aquárius de Presidente Prudente LTDA, Posto Tancredo e no Auto Posto Cohab. Enquanto isso, o Posto Ipiranga comercializa o mesmo produto pelo maior preço do mercado.

Numa mesma análise sobre o preço da gasolina, as melhores opções para abastecer são o Posto Apeano, Auto Posto Feliz, Auto Posto Green, Auto Posto Itapura, Posto Ipiranga, Posto Tancredo, Auto Posto Novo Bongiovani Fi e Auto Posto Jardim Aviação. Em contrapartida, os preços mais altos foram encontrados no Auto Posto BR e no Posto Ipiranga.

A partir das informações da Tabela 13, verifica-se que a amplitude do preço do etanol corresponde a R$0,15, e da gasolina corresponde a R$0,209, que é uma diferença expressiva tendo em vista os valores em questão, embora o preço cobrado pelo litro da gasolina seja mais caro que o preço por litro de etanol.

O conjunto de dados presentes na Figura 27 também nos permite calcular a média dos preços. Após os cálculos, verificou-se que o preço médio por litro de gasolina corresponde a R$4,1265625, e o preço médio por litro de etanol é R$2,868270833. Agora que determinamos a média, podemos calcular as medidas de dispersão para entender se os preços cobrados nos postos estão coerentes. Na Tabela 14 (abaixo) seguem os valores absolutos dos desvios calculados.

Com esses valores vamos calcular o desvio médio, variância e desvio padrão, conforme consta na Tabela 15 (abaixo).

Analisando a Tabela 15, verificamos que tanto o desvio médio, variância, quanto desvio padrão são maiores quando nos referimos à gasolina em comparação com o etanol. Isto significa que os preços da gasolina estão mais dispersos, isto é, os preços de etanol nos postos de combustíveis são mais próximos entre si do que os preços da gasolina, fato confirmado pelo valor das amplitudes calculadas anteriormente, e que acabamos de discutir.

Além disso, mais observações podem ser realizadas. Pesquisas confirmam que é preciso queimar mais etanol para produzir a mesma quantidade de energia que a gasolina. Mais especificamente, a eficiência do etanol corresponde, em média, a 70% da eficiência da gasolina. Consequentemente, se ao abastecer com etanol, verificamos que seu preço ultrapassa 70% do preço da gasolina, não vale a pena preferir por esta opção. Na prática, isso pode ser verificado calculando a razão entre o preço do etanol e da gasolina. É com base neste aspecto que apresentamos a Figura 28.

Tabela 14 – Desvio dos preços dos combustíveis nos postos de Presidente Prudente, SP.

Posto	Desvio Abs. Etanol	Desvio Abs. Gasolina
Posto Três Grandi	0,000729167	0,0275625
Posto Pascoal	0,001729167	0,0034375
Posto Lider de Presidente Prudente	0,009270833	0,0275625
Posto Apeano	0,008270833	0,0365625
Auto Posto Itatiaia	0,030729167	0,0275625
Auto Posto J.A.	0,011729167	0,0434375
Posto Prudente - Rede Prudentão	0,008270833	0,0065625
Posto Zap Presidente Prudente	0,028270833	0,0534375
Auto Posto Prudentão III	0,018270833	0,0165625
Auto Posto Rodotruck de Presidente Prudente	0,008270833	0,0224375
Posto Barão Brasil	0,019270833	0,0275625
Auto Posto BR	0,030729167	0,1724375
Auto Posto Feliz	0,028270833	0,0365625
Auto Posto ZV	0,018270833	0,0034375
Auto Posto Central	0,019270833	0,0275625
Posto Andorinha	0,051729167	0,0634375
Auto Posto Green	0,028270833	0,0365625
Posto Brasil	0,018270833	0,0165625
Auto Posto Lala	0,010729167	0,0224375
Posto Eldorado Presidente Prudente	0,000729167	0,0124375
Auto Posto Itapura	0,018270833	0,0365625
Auto Posto Prudentino	0,019270833	0,0275625
Auto Posto RodoTruck Cidade	0,010729167	0,0275625
Posto Ipiranga	0,121729167	0,1724375
Posto Senna Presidente Prudente	0,019270833	0,0275625
Congonhas Auto Posto	0,010729167	0,0275625
Auto Posto Aquárius de Presidente Prudente LTDA	0,028270833	0,0265625
ACJ Auto Posto De Combustíveis LTDA	0,000729167	0,0124375
Posto Ipiranga	0,001729167	0,0365625
Posto Tancredo	0,028270833	0,0365625
Auto Posto Maxim's	0,010729167	0,0275625
Auto Posto Janda	0,010729167	0,0224375
Auto Posto Novo Bongiovani Fi	0,008270833	0,0365625
Posto Ipiranga	0,000729167	0,0275625
Auto Posto Castelinho	0,008270833	0,0065625
Posto Ipiranga	0,001729167	0,0134375
Auto Posto Jardim Aviação	0,008270833	0,0365625
Posto Classic	0,010729167	0,0275625
Posto Ipiranga	0,009270833	0,0224375
Posto de Serviços Jardim das Rosas	0,019270833	0,0275625
Posto Ipiranga	0,010729167	0,0224375
Posto Itatiaia2 Presidente Prudente	0,030729167	0,0275625
Auto Posto Executivo	0,030729167	0,0224375
Posto Everest I	0,001729167	0,0334375
Auto Posto Ana Jacinta	0,010729167	0,0275625
Auto Posto Cohab	0,028270833	0,0034375
Posto Bela Vista	0,018270833	0,0065625
Auto Posto Gricoleto LTDA	0,021729167	0,0634375

Fonte: Elaborado pelo autor

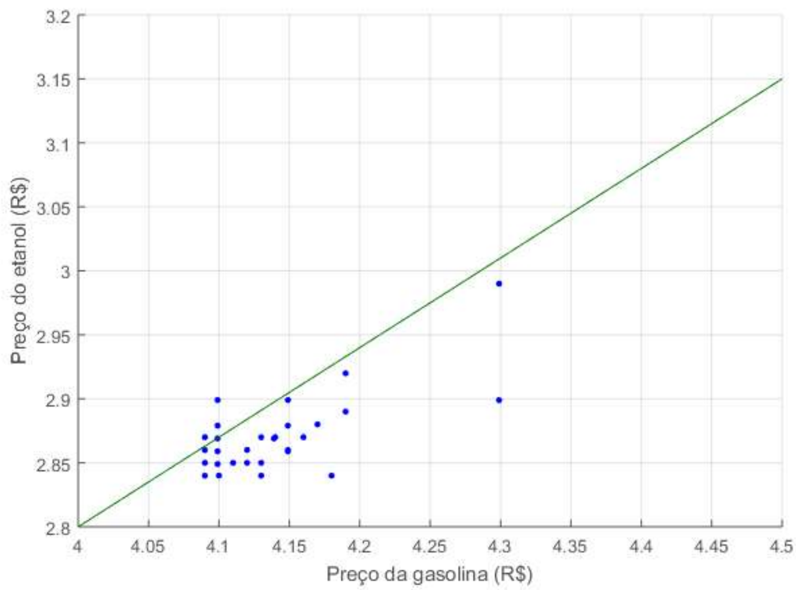
Tabela 15 – Medidas de dispersão.

	Etanol	Gasolina
Desvio médio	0,017729167	0,032721354
Variância	0,000660822	0,002091621
Desvio padrão	0,025706468	0,045734244

Fonte: Elaborado pelo autor

Na Figura 28, cada ponto corresponde a um dos postos de combustíveis. Os pontos localizados abaixo da linha verde referem-se aos postos onde compensa abastecer com etanol, isto é, onde o preço do etanol está abaixo de 70% do preço da gasolina. Em contrapartida, os pontos localizados acima da linha, indicam os postos nos quais é mais aconselhável optar pela gasolina.

Figura 28 – Comparação entre preços de etanol e gasolina.



Fonte: Elaborado pelo autor

Percebemos três pontos acima da linha verde, mas na realidade eles correspondem a oito postos de combustíveis. Isto ocorre devido ao fato de que alguns postos apresentam preços coincidentes para os dois tipos de combustíveis, e consequentemente na Figura 28 há pontos que estão sobrepostos. Mas observando a figura acima, podemos perceber que na maioria dos postos, vale a pena abastecer com etanol.

Para uma análise mais detalhada, apresentamos a Tabela 16 com os valores da razão entre o preço do etanol e da gasolina.

Tabela 16 – Comparação entre o preço por litro de etanol e de gasolina.

Posto	Preço etanol	Preço gasolina	Razão (etanol/gasolina)
Auto Posto BR	2,899	4,299	0,6743428704
Posto Zap Presidente Prudente	2,84	4,18	0,6794258373
Auto Posto Cohab	2,84	4,13	0,6876513317
Posto Ipiranga	2,859	4,149	0,6890817064
Auto Posto Rodotruck de Pres Prudente	2,86	4,149	0,6893227284
Auto Posto Gricoleto LTDA	2,89	4,19	0,6897374702
Posto Everest I	2,87	4,16	0,6899038462
Auto Posto ZV	2,85	4,13	0,6900726392
Auto Posto J.A.	2,88	4,17	0,6906474820
Posto Bela Vista	2,85	4,12	0,6917475728
Auto Posto Aquárius de Pres Prudente LTDA	2,84	4,1	0,6926829268
Posto Eldorado Presidente Prudente	2,869	4,139	0,6931625997
ACJ Auto Posto De Combustíveis LTDA	2,869	4,139	0,6931625997
Posto Ipiranga	2,87	4,14	0,6932367150
Auto Posto Prudentão III	2,85	4,11	0,6934306569
Posto Brasil	2,85	4,11	0,6934306569
Auto Posto Lala	2,879	4,149	0,6939021451
Auto Posto Janda	2,879	4,149	0,6939021451
Posto Ipiranga	2,879	4,149	0,6939021451
Posto Prudente - Rede Prudentão	2,86	4,12	0,6941747573
Auto Posto Castelinho	2,86	4,12	0,6941747573
Auto Posto Feliz	2,84	4,09	0,6943765281
Auto Posto Green	2,84	4,09	0,6943765281
Posto Tancredo	2,84	4,09	0,6943765281
Posto Pascoal	2,87	4,13	0,6949152542
Posto Barão Brasil	2,849	4,099	0,6950475726
Auto Posto Central	2,849	4,099	0,6950475726
Auto Posto Prudentino	2,849	4,099	0,6950475726
Posto Senna Presidente Prudente	2,849	4,099	0,6950475726
Posto de Serviços Jardim das Rosas	2,849	4,099	0,6950475726
Posto Ipiranga	2,99	4,299	0,6955105839
Auto Posto Itapura	2,85	4,09	0,6968215159
Posto Andorinha	2,92	4,19	0,6968973747
Posto Lider de Presidente Prudente	2,859	4,099	0,6974871920
Auto Posto Executivo	2,899	4,149	0,6987225838
Posto Apeano	2,86	4,09	0,6992665037
Auto Posto Novo Bongiovani Fi	2,86	4,09	0,6992665037
Auto Posto Jardim Aviação	2,86	4,09	0,6992665037
Posto Três Grandi	2,869	4,099	0,6999268114
Posto Ipiranga	2,869	4,099	0,6999268114
Posto Ipiranga	2,87	4,09	0,7017114914
Auto Posto RodoTruck Cidade	2,879	4,099	0,7023664308
Congonhas Auto Posto	2,879	4,099	0,7023664308
Auto Posto Maxim's	2,879	4,099	0,7023664308
Posto Classic	2,879	4,099	0,7023664308
Auto Posto Ana Jacinta	2,879	4,099	0,7023664308
Auto Posto Itatiaia	2,899	4,099	0,7072456697
Posto Itatiaia2 Presidente Prudente	2,899	4,099	0,7072456697

Estas considerações a primeira vista parecem óbvias, contudo, estes resultados estão em valores médios, tem uma variabilidade natural. Além disto, os preços estão todos muito próximos da razão de 70%. Então, isto vai depender do tipo de trajeto que se vai andar (longe ou perto), o tipo de veículo que o condutor possui (alguns modelos tem desempenho superior aos 70% e outros abaixo), a forma de condução, entre outros fatores. Resumidamente, haverá pouca diferença se o condutor não se atentar a estas outras características e utilizar como critério de decisão somente a razão de 70%.

Depois desta análise, pode-se também construir um intervalo de confiança para avaliar a média verdadeira dos preços dos dois combustíveis considerados.

A amostra, de 48 elementos, é suficiente para que possamos utilizar o intervalo de confiança para grandes amostras, dado por

$$\bar{x} \pm Z_{1-\alpha/2} . \frac{s}{\sqrt{n}}$$

com 1-α de confiança.

Utilizando os dados dos 48 postos foi possível calcular os intervalos de confiança com 1-α= 0,95, 1-α=0,98 e 1-α= 0,99 tanto para o etanol como para a gasolina comum. Os intervalos podem ser vistos na tabela 17.

Tabela 17 – Intervalos de confiança.

1-α	Etanol	Gasolina
0,95	(2,861 ; 2,876)	(4,113 ; 4,140)
0,98	(2,859 ; 2,877)	(4,111 ; 4,142)
0,99	(2,859 ; 2,878)	(4,109 ; 4,144)

Fonte: Elaborado pelo autor

Pode-se observar que, devido a pouquíssima variabilidade dos dados os intervalos de confiança pouco mudam com 1-α= 0,95, 1-α=0,98 e 1-α=0,99. Existe variabilidade, alguns postos praticam preços mais altos e outros preços mais baixos que a média, mas são pouquíssimos no montante amostrado significantemente distintos. Isto faz com que, juntamente com o tamanho amostral, os intervalos tenham diferenças apenas em casas decimais que praticamente não são praticados como preços de mercado.

Ao que se percebe, a margem de diferença de preços exercitada parece ser bastante apertada, tanto para o etanol como para a gasolina comum.

7 CONCLUSÕES

O processo de ensino aprendizagem constitui algo muito importante para o crescimento e amadurecimento do aluno. O professor é uma peça imprescindível para isso, fazendo da experiência um aliado para atingir os seus objetivos, mesmo quando o meio educacional atual se torna desafiador.

Esse trabalho, de forma simples e aplicada, vem utilizar elementos ensinados na escola básica para solucionar questões de grande interesse nos campos da Ecologia, Biologia e em pesquisas nos mais variados setores.

Foram resolvidos alguns problemas de estatística e probabilidade, principalmente os de distribuição normal a fim de que o aluno veja a importância que a estatística tem nos problemas do dia-a-dia. O tratamento para resolver esses problemas consiste em trabalhar conceitos estudados no ensino médio, de fácil manuseio para eles e de outros mais avançados, como a distribuição normal, mas com resultados com os quais possam ser utilizados de forma mais simples.

Observa-se, por fim, que a proposta deste trabalho foi de promover a utilização dos conceitos de estatística como motivadores, proporcionou executar aplicações em situações reais, algo bem diferente do que realizado nas instituições de ensino e do que é ofertado em manuais escolares da Educação Básica.

Este trabalho, resumidamente, por meio das aplicações reais, procura buscar agentes multiplicadores, ou seja, estudantes que possam descobrir os empregos que a estatística pode proporcionar em diferentes áreas e, desta forma, levar a um maior conhecimento e valorização da mesma.

REFERÊNCIAS

AGRESTI, Alan; FINLAY, Barbara. **Métodos estatísticos para as ciências sociais**. 4.ed. Porto Alegre: Penso, 2012

BAPTISTA, Makilim Nunes; CAMPOS, Dinael Corrêa de. **Metodologias de pesquisa em ciências**: análise quantitativa e qualitativa. 2.ed. Rio de Janeiro: LTC, 2016.

BONAFINI, Fernanda César. **Probabilidade e estatística**. São Paulo: Pearson Education do Brasil, 2015.

BRACARENSE, Paulo Afonso. **Estatística aplicada às ciências sociais**. Curitiba: IESDE Brasil S.A., 2012.

CORREA, Sonia Maria Barros Barbosa. **Probabilidade e estatística**. 2.ed. Belo Horizonte: PUC Minas Virtual, 2003.

CRESPO, Antônio Arnot. **Estatística fácil**. São Paulo: Saraiva, 1997.

CRILLY, Tony. **50 ideias de matemática que precisa mesmo de saber**. Portugal: Editora Leya, 2014.

DOWING, Douglas; CLARK, Jeffrey. **Estatística aplicada**. São Paulo: Saraiva, 2010.

SÃO PAULO (Estado). Secretaria da Educação. Caderno do Aluno: Matemática, ensino fundamental – 8ª série, volume 2. São Paulo: SEE, 2014a.

SÃO PAULO (Estado). Secretaria da Educação. Caderno do Professor: Matemática, ensino médio – 3ª série, volume 2. São Paulo: SEE, 2014b.

SÃO PAULO (Estado). Equipe Técnica de Matemática Área de Matemática Coordenadoria de Estudos e Normas Pedagógicas – CenpSecretaria da Educação do Estado de São Paulo.

FREUND, John. **Economia, administração e contabilidade**: estatística aplicada. 11.ed. São Paulo: Bookman, 2007.

KÜHN, Daniela Dias. **Pesquisa e análise de dados**: problematizando o rural e a agricultura numa perspectiva científica. Porto Alegre: Editora da UFRGS, 2017.

LARSON, Ron; FARBER, Betsy. **Estatística aplicada**. São Paulo: Pearson Education do Brasil, 2010.

LARSON, Ron; FARBER, Betsy. **Estatística aplicada**. São Paulo: Pearson Education do Brasil, 2015.

MORETTIN, Pedro Alberto. **Estatística básica**. 8.ed. São Paulo: Saraiva, 2013.

NAVIDI, Willian. **Probabilidade e estatística para ciências exatas**. São Paulo: AMGH Editora Ltda., 2012.

PASQUALI, Luiz. A Curva Normal. In:______. **Matemática Discreta**. Rio de Janeiro: Coleção PROFMAT, 2014.

PEREIRA, Júlio César Rodrigues. **Análise de dados qualitativos**: estratégias metodológicas para as ciências da saúde, humanas e sociais. São Paulo: Editora da Universidade de São Paulo, 2004.

RUMSEY, Deborah. **Estatística III para leigos**. Rio de Janeiro: Alta Books, 2009.

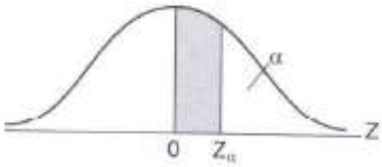
RUMSEY, Deborah. **Estatística III para leigos**. Rio de Janeiro: Alta Books, 2014.

Matematiquês: matemática é fácil. Disponível em: < http://www.matematiques.com.br/ >. Acesso em: 12 mar. 2018.

WALPOLE, Ronald E.. MYERS Raymond H., MYERS Sharon L., YE Keying. Probabilidade e Estatística. São Paulo. Pearson, 2009.

ANEXO

Tabela de valores da Normal Reduzida



DISTRIBUIÇÃO NORMAL: N(0,1)

$P(0 < z < z_\alpha) = \alpha$

zα	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	zα
0,00	-0,000000	0,003989	0,007978	0,011967	0,015953	0,019939	0,023922	0,027903	0,031881	0,035856	0,00
0,10	0,039828	0,043795	0,047758	0,051717	0,055670	0,059618	0,063559	0,067495	0,071424	0,075345	0,10
0,20	0,079260	0,083166	0,087064	0,090954	0,094835	0,098706	0,102568	0,106420	0,110261	0,114092	0,20
0,30	0,117911	0,121719	0,125516	0,129300	0,133072	0,136831	0,140576	0,144309	0,148027	0,151732	0,30
0,40	0,155422	0,159097	0,162757	0,166402	0,170031	0,173645	0,177242	0,180822	0,184386	0,187933	0,40
0,50	0,191462	0,194974	0,198468	0,201944	0,205402	0,208840	0,212260	0,215661	0,219043	0,222405	0,50
0,60	0,225747	0,229069	0,232371	0,235653	0,238914	0,242154	0,245373	0,248571	0,251748	0,254903	0,60
0,70	0,258036	0,261148	0,264238	0,267305	0,270350	0,273373	0,276373	0,279350	0,282305	0,285236	0,70
0,80	0,288145	0,291030	0,293892	0,296731	0,299546	0,302338	0,305106	0,307850	0,310570	0,313267	0,80
0,90	0,315940	0,318589	0,321214	0,323814	0,326391	0,328944	0,331472	0,333977	0,336457	0,338913	0,90
1,00	0,341345	0,343752	0,346136	0,348495	0,350830	0,353141	0,355428	0,357690	0,359929	0,362143	1,00
1,10	0,364334	0,366500	0,368643	0,370762	0,372857	0,374928	0,376976	0,378999	0,381000	0,382977	1,10
1,20	0,384930	0,386860	0,388767	0,390651	0,392512	0,394350	0,396165	0,397958	0,399727	0,401475	1,20
1,30	0,403199	0,404902	0,406582	0,408241	0,409877	0,411492	0,413085	0,414656	0,416207	0,417736	1,30
1,40	0,419243	0,420730	0,422196	0,423641	0,425066	0,426471	0,427855	0,429219	0,430563	0,431888	1,40
1,50	0,433193	0,434478	0,435744	0,436992	0,438220	0,439429	0,440620	0,441792	0,442947	0,444083	1,50
1,60	0,445201	0,446301	0,447384	0,448449	0,449497	0,450529	0,451543	0,452540	0,453521	0,454486	1,60
1,70	0,455435	0,456367	0,457284	0,458185	0,459071	0,459941	0,460796	0,461636	0,462462	0,463273	1,70
1,80	0,464070	0,464852	0,465621	0,466375	0,467116	0,467843	0,468557	0,469258	0,469946	0,470621	1,80
1,90	0,471284	0,471933	0,472571	0,473197	0,473810	0,474412	0,475002	0,475581	0,476148	0,476705	1,90
2,00	0,477250	0,477784	0,478308	0,478822	0,479325	0,479818	0,480301	0,480774	0,481237	0,481691	2,00
2,10	0,482136	0,482571	0,482997	0,483414	0,483823	0,484222	0,484614	0,484997	0,485371	0,485738	2,10
2,20	0,486097	0,486447	0,486791	0,487126	0,487455	0,487776	0,488089	0,488396	0,488696	0,488989	2,20
2,30	0,489276	0,489556	0,489830	0,490097	0,490358	0,490613	0,490863	0,491106	0,491344	0,491576	2,30
2,40	0,491802	0,492024	0,492240	0,492451	0,492656	0,492857	0,493053	0,493244	0,493431	0,493613	2,40
2,50	0,493790	0,493963	0,494132	0,494297	0,494457	0,494614	0,494766	0,494915	0,495060	0,495201	2,50
2,60	0,495339	0,495473	0,495603	0,495731	0,495855	0,495975	0,496093	0,496207	0,496319	0,496427	2,60
2,70	0,496533	0,496636	0,496736	0,496833	0,496928	0,497020	0,497110	0,497197	0,497282	0,497365	2,70
2,80	0,497445	0,497523	0,497599	0,497673	0,497744	0,497814	0,497882	0,497948	0,498012	0,498074	2,80
2,90	0,498134	0,498193	0,498250	0,498305	0,498359	0,498411	0,498462	0,498511	0,498559	0,498605	2,90
3,00	0,498650	0,498694	0,498736	0,498777	0,498817	0,498856	0,498893	0,498930	0,498965	0,498999	3,00
3,10	0,499032	0,499064	0,499096	0,499126	0,499155	0,499184	0,499211	0,499238	0,499264	0,499289	3,10
3,20	0,499313	0,499336	0,499359	0,499381	0,499402	0,499423	0,499443	0,499462	0,499481	0,499499	3,20
3,30	0,499517	0,499533	0,499550	0,499566	0,499581	0,499596	0,499610	0,499624	0,499638	0,499650	3,30
3,40	0,499663	0,499675	0,499687	0,499698	0,499709	0,499720	0,499730	0,499740	0,499749	0,499758	3,40
3,50	0,499767	0,499776	0,499784	0,499792	0,499800	0,499807	0,499815	0,499821	0,499828	0,499835	3,50
3,60	0,499841	0,499847	0,499853	0,499858	0,499864	0,499869	0,499874	0,499879	0,499883	0,499888	3,60
3,70	0,499892	0,499896	0,499900	0,499904	0,499908	0,499912	0,499915	0,499918	0,499922	0,499925	3,70
3,80	0,499928	0,499930	0,499933	0,499936	0,499938	0,499941	0,499943	0,499946	0,499948	0,499950	3,80
3,90	0,499952	0,499954	0,499956	0,499958	0,499959	0,499961	0,499963	0,499964	0,499966	0,499967	3,90
4,00	0,499968	0,499970	0,499971	0,499972	0,499973	0,499974	0,499975	0,499976	0,499977	0,499978	4,00

Fonte: Morettin, 2013.