



ARGUMENTOS

UNIDADE

3



José Nogueira

Livro Eletrônico

Aula 3 – Estatística Básica

Objetivo da Aula

Compreender conceitos fundamentais de estatística, bem como medidas e técnicas utilizadas, e conhecer o método estatístico.

Apresentação

O ser humano utiliza a matemática e as suas respectivas ramificações para tomada de decisão desde o início da civilização. No princípio da civilização, o homem mapeou a distância das fontes de água ou mesmo a periodicidade de cada estação do ano, de forma a saber quando deveria plantar ou mesmo colher. As técnicas de coleta de dados, interpretação, análise e representação são de uma área da matemática denominada “estatística”.

Nesta aula, vamos conversar sobre temas como conceito de amostra, população e variável ou medidas centrais e de dispersão. Um assunto importante que vamos abordar é a melhor forma de apresentação dos dados por tabelas ou gráficos. Conversaremos, também, sobre o método estatístico e as fases de execução desse método.

1. Por que Estatística?

A estatística é uma área do conhecimento que se dedica a coletar, organizar, analisar e interpretar dados, com o objetivo de extrair informações relevantes e tomar decisões embasadas em evidências. Assim, a estatística é uma ferramenta importante em diversas áreas do conhecimento, como ciências naturais, sociais, medicina, engenharia e finanças. Podemos, também, afirmar que a estatística é um método científico utilizado para estudar fenômenos multicausais, coletivos ou de massa. Esses fenômenos são aqueles que envolvem a ação de múltiplos fatores e/ou a participação de um grande número de indivíduos.

Por exemplo, podemos listar a violência urbana como um fenômeno multicausal, uma vez que envolve diversos fatores, como desigualdade social, falta de oportunidades e educação precária. Por motivos diferentes, também podemos citar as mudanças climáticas como um fenômeno multicausal, pois envolvem a interação de diversos fatores, como atividades humanas, variações naturais do clima e até mesmo eventos externos. Ou seja, fenômenos multicausais são aqueles que possuem diversas variáveis e fontes de causa para serem estudadas simultaneamente.

Já um fenômeno coletivo ou de massa é aquele que envolve a participação de um grande número de indivíduos, os quais estão agindo de forma ordenada ou sendo influenciados por um mesmo estímulo. Considere, por exemplo, que manifestações políticas, compras em massa em grandes promoções (como na **black friday**) e torcidas organizadas de eventos esportivos são todos fenômenos coletivos ou de massa.

Dado o alto volume de dados que podem ser retirados de todos esses fenômenos, torna-se evidente a necessidade de um método sólido, para que seja possível avaliar e produzir estudos em cima deles. Esses estudos e métodos são, então, utilizados para que seja possível tomar uma decisão embasada em observação empírica e correta análise de dados. Esse é exatamente o papel da estatística, uma metodologia que formaliza a coleta, organização, descrição, análise e interpretação de dados que auxiliam em uma tomada de decisões.

2. Conceitos Básicos

Para que seja possível aplicar o método estatístico, é necessário, primeiro, entender alguns conceitos básicos que o formam, como população, amostra, classificação de dados, tipos de representação e frequência. Esses conceitos são peças fundamentais para o correto entendimento dos estudos em cima de estatística.

2.1. População e Amostra

Figura 1: Representação lúdica de uma amostra como um “gráfico fatia de pizza”



Fonte: Mundo Educação. Disponível em: <https://mundoeducacao.uol.com.br/sociologia/minorias.htm>.

Acesso em: 26 mai. 2023.

Chamamos de “**população** estatística” todos os elementos cujas características pretendemos estudar e analisar. É comum, também, chamar esse conjunto de “universo estatístico”. Esses elementos vão, obrigatoriamente, possuir características em comum, as quais podem ter valores distintos, mas todos os elementos possuem algum valor naquela característica, mesmo que nulo.

Já uma **amostra** é quando pegamos apenas parte da população, normalmente uma fração pequena, genuinamente aleatória, para representar a população como um todo. Obviamente, utilizar uma amostra diminui o nível de confiabilidade da pesquisa, porém também possibilita que algumas pesquisas estatísticas mais complexas sejam feitas. Existem técnicas capazes de determinar o nível de confiança de uma amostra, diante de seu tamanho, quando comparado com a população que ela representa. Dentre os motivos para utilizar uma amostra, não a população inteira, destacam-se:

- Custo.
 - Obviamente, coletar dados de menos elementos tornará a pesquisa mais barata e eficiente.
- Tamanho da população.
 - Se a população for grande demais, pode ser inviável fazer a testagem e coleta com todos os elementos.
- Tempo.
 - Quanto menos elementos para coletar dados, menos tempo será gasto na etapa de coleta.

Existem, também, diferentes tipos de amostragem:

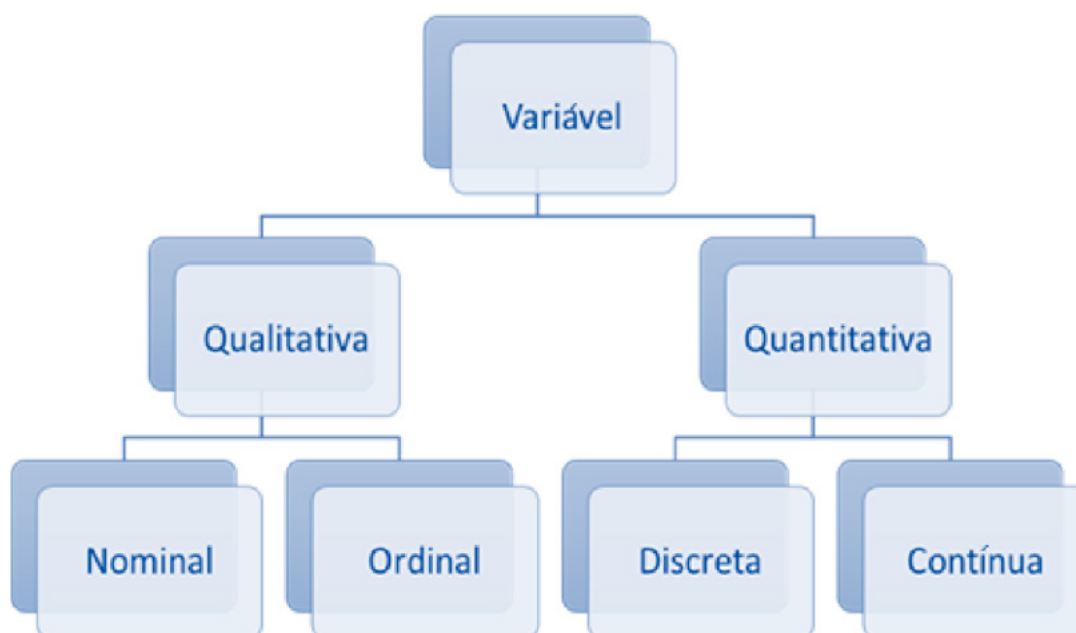
- Aleatória simples.
 - Quando os elementos são escolhidos de forma que todos tenham a mesma chance de serem selecionados para a coleta de dados.
 - Pode ser feita utilizando uma tabela de números aleatórios, um programa de números aleatórios ou qualquer outro método capaz de gerar aleatoriedade real.
- Estratificada.
 - A população é dividida em, no mínimo, dois estratos ou subconjuntos que compartilhem a mesma característica. Então, são feitas amostragens em cada um desses subconjuntos.
 - Um exemplo simples seria, em uma população de pessoas, separá-las de acordo com seu gênero e, então, coletar amostragens de cada gênero por igual.

- Sistemática.
 - É estabelecido um ponto de partida e um passo a ser dado na escolha.
 - Por exemplo, o elemento 2 é escolhido como ponto de partida e, a partir daí, contamos 10 indivíduos e coletamos de novo, de forma a coletar o 12... 22... 32..., e assim por diante.
- Conglomerados.
 - A população é dividida em áreas ou conglomerados; então, são sorteadas as áreas, e todos os elementos daquela área são selecionados.
 - Por exemplo, em uma estatística de uma cidade, os conglomerados poderiam ser bairros ou, em uma estatística de bairro, o conglomerado poderia ser um quarteirão.

2.2. Classificação dos Dados ou Variáveis

Os dados são as peças-chave de todo o método estatístico. O enorme volume de dados é, precisamente, o motivo pelo qual são necessárias toda uma metodologia e uma área de conhecimento específica para que seja possível garantir os melhores resultados. Um dado, na verdade, é uma variável em comum de todos os elementos da população a ser estudada ou inferida. Podemos, então, classificar os dados como quantitativos ou qualitativos, os quais ainda possuem mais duas classificações, como podemos ver na tabela hierárquica a seguir.

Figura 2: Classificação de dados



Fonte: Elaboração própria.

Falamos “variável qualitativa” quando os seus valores são expressos por características ou atributos não expressáveis de forma numérica – por exemplo, sexo, estado civil, escolaridade ou profissão. Podemos, então, classificar essas variáveis de mais duas formas:

- Ordinal.
 - Conseguimos atribuir alguma espécie de ordem aos indivíduos após a definição da variável em si.
 - Como “escolaridade”, em que podemos ordenar os indivíduos.
- Nominal.
 - Quando não é possível atribuir nenhuma classificação extra à variável.
 - Como “profissão” ou “estado civil”, em que não dá para ordenar.

Já as variáveis **quantitativas** são aquelas que podem ter os seus valores expressos por números. Estes são alguns exemplos de variáveis quantitativas: idade, salário e volume. As variáveis quantitativas também são divididas em mais duas categorias.

- Discretas.
 - Quando os valores numéricos podem ser enumerados.
 - Como contagem do número de alunos em classe.
- Contínuas.
 - Quando o valor da variável pertence a um intervalo dos números reais.
 - Como a medição de volume de água em uma caixa d’água ou temperatura de determinado lugar.

2.3. Gráficos e Tabelas

Gráficos e tabelas são elementos essenciais para estatística. Eles permitem que os dados sejam organizados, de forma coerente e sucinta, para que os pesquisadores e tomadores de decisão possam aplicar técnicas estatísticas e, assim, chegar às conclusões e aos resultados necessários. Existem muitos tipos de gráficos e tabelas, alguns com usos bem amplos e outros mais restritos, mas todos com suas vantagens acima do próximo.

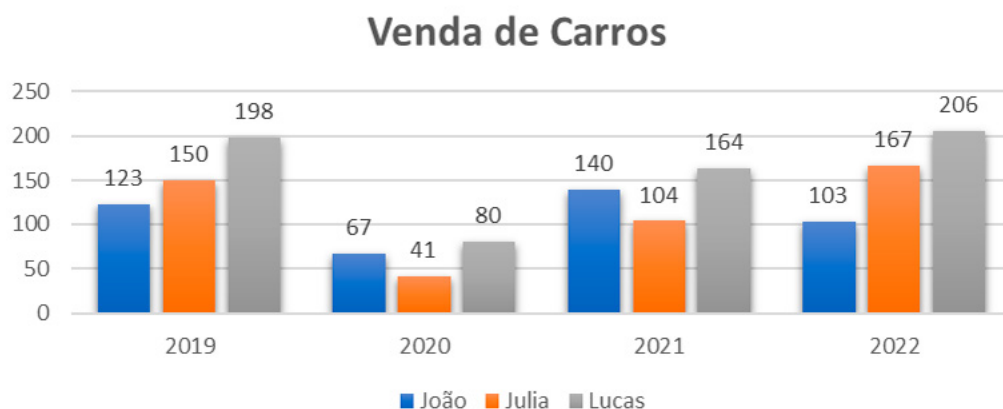
É importante conhecer os principais e mais utilizados, para saber como melhor apresentar os dados em questão, de forma relevante para o problema a ser resolvido.

Por isso, vamos começar com os **gráficos**.

- **Gráfico em colunas.**

Nesse tipo de gráfico, as categorias são organizadas na horizontal, enquanto seus valores são representados pelo tamanho vertical. É muito útil quando se quer mostrar alterações de dados em um período ou simplesmente fazer comparações entre itens distintos. Observe a venda de três vendedores de carros entre os anos 2019 e 2022.

Gráfico 1: Gráfico em colunas

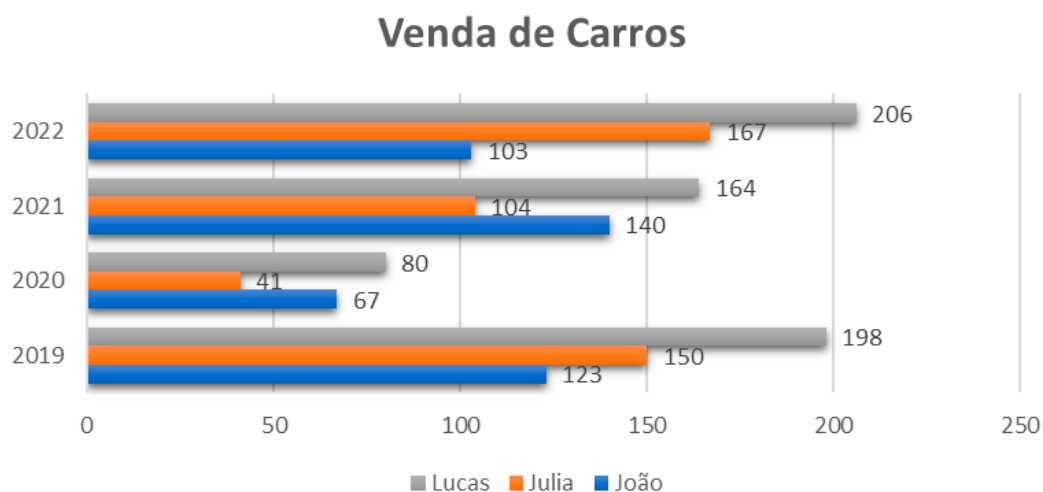


Fonte: Elaboração própria.

- **Gráfico em barras.**

Esse gráfico é semelhante ao de colunas, porém as categorias são dispostas verticalmente, e os valores são retângulos espalhados na horizontal.

Gráfico 2: Gráfico em barras

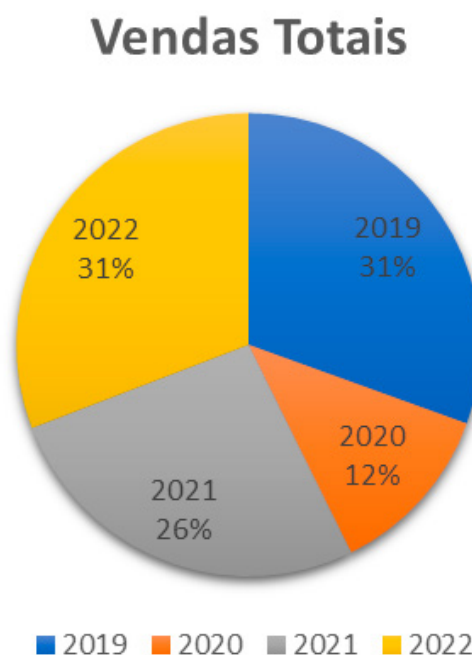


Fonte: Elaboração própria.

- **Gráfico em setores.**

Faz-se um círculo e, por meio de setores relativos, nesse círculo, podemos comparar as partes em relação com o todo. É bom para analisar proporções. Também costuma ser chamado de “gráfico pizza”, em que as fatias são as categorias representadas.

Gráfico 3: Gráfico em setores



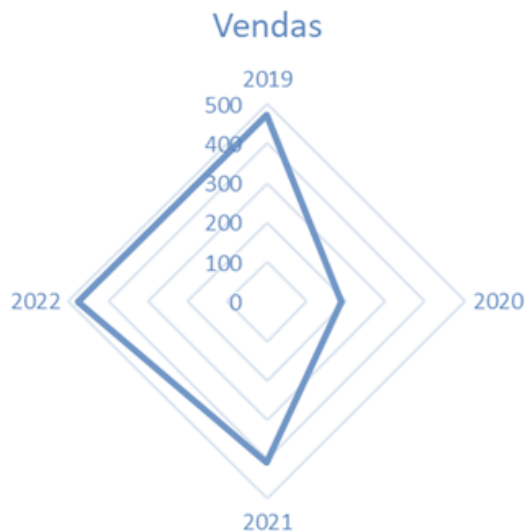
Fonte: Elaboração própria.

- **Gráfico polar.**

Um gráfico polar utiliza polígonos para mostrar os seus valores. Sua construção é feita com uma circunferência que possui tantos arcos quanto dados. Nos pontos de divisão, então, são traçados raios, e cada raio é representado por um valor. Ainda, marcam-se pontos cuja distância até o centro da circunferência é diretamente proporcional a esse valor. Assim, os pontos são unidos.

Esse gráfico costuma ser utilizado em progressões temporais, para avaliar as etapas temporais estabelecidas. Também costuma ser chamado de “gráfico radar”, devido à sua semelhança com imagens geradas por radares.

Gráfico 4: Gráfico polar

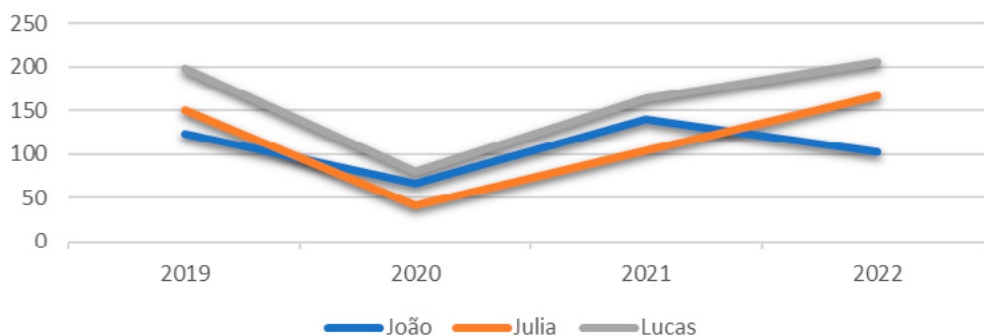


Fonte: Elaboração própria.

- **Gráfico de linhas.**

Um gráfico de linhas é utilizado para mostrar tendências que estão ocorrendo nos dados em intervalos iguais.

Gráfico 5: Gráfico de linhas



Fonte: Elaboração própria.

Vamos, agora, partir para as **tabelas** mais comuns de serem utilizadas no método estatístico.

- **Tabelas cronológicas, evolutivas, históricas ou temporais.**

Nessas tabelas, os dados são ordenados de acordo com a época de ocorrência dos fatos, de forma que a primeira coluna é o período utilizado em questão, enquanto as outras são os dados e valores em si.

- **Tabelas de localização ou geográficas.**

A tabela geográfica permite que a gente ordene os valores de acordo com localidades, como região ou cidade.

- **Tabelas específicas.**

Tabelas específicas servem quando utilizamos um método próprio de ordenar os valores, pois eles são específicos para o método estatístico em questão.

- **Distribuição de frequência.**

Serve para dispor a frequência absoluta de um fenômeno específico, como a frequência de um aluno na sala de aula. Vale ressaltar que existem tipos de tabelas de frequência diferentes.

- Frequência acumulada.
 - Quando vamos acumulando as ocorrências em cada etapa ou categoria da tabela.
- Frequência relativa.
 - Uma divisão da frequência absoluta pelo tamanho da amostra em questão.
 - Ou seja, uma porcentagem do quanto aquela frequência significa em relação ao todo.
- Frequência cumulativa.
 - A soma das frequências absolutas ou relativas desde o valor inicial da variável.

3. Medidas e Técnicas

As medidas e técnicas são utilizadas para formar as tabelas e, por consequência, os gráficos utilizados na metodologia de estatística. Podemos dividir essas medidas em dois principais agrupamentos.

3.1. Medidas de Tendência Central

Também chamadas de “medidas de ordenamento ou posição”, as medidas de tendência central têm como objetivo analisar os valores de uma série de dados. Considere a tabela a seguir para conferir os exemplos das medidas.

Funcionário	Alex	Victoria	Sonia	Matheus	Alice
Salário/h	R\$ 64,00	R\$ 90,00	R\$ 35,00	R\$ 50,00	R\$ 70,00

• **Média.**

- Soma de todos os valores dividida pela quantidade de dados.
 - Por exemplo, a soma de todos os carros vendidos por ano dividida pela quantidade de anos analisados para se decidir a média de carros vendidos ao ano.
- Fácil de calcular e utiliza todos os dados da amostragem em questão. Muito útil para conhecer valores totais.
 - Porém, é necessário possuir todos os dados da amostragem.
 - Além disso, valores extremos podem alterar drasticamente a média, tornando esse dado irrelevante, dependendo da situação.
- No exemplo da tabela acima, ficaria assim:
 - Soma dos valores:

$$64 + 90 + 35 + 50 + 70 = 309$$

- Divisão pelo número de valores:

$$309/5 = 61,8$$

- 61,8 é a média do salário por hora dessa tabela.

• **Mediana.**

- Medida de tendência central de uma série de dados ordenados de forma crescente.
- Se a amostra possuir valores muito extremos na amostragem, a mediana pode ser uma solução melhor do que a média, pois essa primeira não é afetada por inclinações acentuadas.
- Para calcular a mediana, é necessário ordenar todos os valores de forma crescente e, então, ver o valor central.
 - Se for uma quantidade “n” de valores ímpares, a mediana simplesmente é o valor que está na posição $(n + 1)/2$, ou seja, o valor mais central dessa ordenação.
 - Se for uma quantidade “n” de valores pares, a mediana será a média dos valores nas posições $(n/2) + 1$ e $(n/2) - 1$.

- No exemplo de tabela acima:
 - Ordenamos os valores em ordem crescente:

35 – 50 – 64 – 70 – 90

- Como temos um número ímpar de valores, estabelecemos que a mediana é o $(n + 1)/2$, em que n é 5, logo $(5 + 1)/2 = 3$. Então, a mediana é o valor na terceira posição que, nesse caso, é 64.

• **Moda.**

- Moda é o valor que mais se repete na série de dados.
- Não há cálculos complexos a se fazer, apenas somas e contagens simples. Deve-se analisar a série de dados organizados em uma tabela e contar os dados que mais se repetem. Aquele que mais se repete é chamado de “moda”.
- No exemplo da tabela acima, não há nenhuma moda, pois nenhum dado possui valor repetido.

3.2. Medidas de Dispersão

As medidas de dispersão são utilizadas para descrever a variabilidade de uma distribuição de dados. Elas fornecem informações sobre a distância dos dados entre si, da média ou de outros parâmetros relevantes. Vamos utilizar a mesma tabela como exemplo para essa próxima série de medidas.

Funcionário	Alex	Victoria	Sonia	Matheus	Alice
Salário/h	R\$ 64,00	R\$ 90,00	R\$ 35,00	R\$ 50,00	R\$ 70,00

• **Amplitude total.**

- A amplitude geral é a diferença entre o maior e menor valor da série de dados em questão.
- No exemplo acima, o maior valor é 90 e o menor é 35.
- Logo, a amplitude seria $90 - 35 = 55$.

• **Desvio.**

- É a diferença entre o dado em questão e a média da série de dados.
- $x = x - \mu$
 - μ é a média da série de dados, e x é o dado que quer descobrir o desvio.

- A média do exemplo acima é 61,8.
- O desvio de Alex é $x = x - \mu$, logo $64 - 61,8 = 2,2$.
- O desvio de Matheus é $x = x - \mu$, logo $50 - 61,8 = -11,8$.

• **Variância.**

- A variância de uma medida mostra quão distantes estão os valores da média.
- É calculada como a média dos quadrados das diferenças entre cada valor e a média da distribuição.
- Se calcularmos o desvio de toda uma série de dados e resolvermos calcular a média do desvio, rapidamente vamos perceber que isso é impossível, pois o somatório de todos os desvios é sempre 0.
 - No exemplo acima, os desvios seriam, na ordem em que aparecem na tabela, os seguintes: 2,2; 28,2; -26,8; -11,8; 8,2.
 - Somando, ficaria $2,2 + 28,2 + (-26,8) + (-11,8) + 8,2 = 0$.
- Por fim, chamamos de “variância” a média dos quadrados dos desvios.
- Variância = $\Sigma (xi - \mu)^2 / (n - 1)$. Em que:
 - Σ = somatório.
 - Xi = cada valor da série de dados.
 - μ = média aritmética da série.
 - n = número de observações da série.
- Substituindo os valores da tabela acima, nessa fórmula, obtemos
 - Variância = $[(64 - 61,80)^2 + (90 - 61,80)^2 + (35 - 61,80)^2 + (50 - 61,80)^2 + (70 - 61,80)^2] / (5 - 1)$
 - Variância = $669,60 / 4$
 - Variância = 167,40

• **Desvio-padrão.**

- O desvio-padrão é a raiz quadrada da variância, e isso é feito para facilitar a leitura do dado.
 - No caso do exemplo acima, chegaríamos ao valor de 12,93.

4. Afinal, o que É o Método Estatístico?

É muito aceita a divisão de estatística em dois principais grupos, a **estatística descritiva** e a **estatística indutiva (ou inferencial)**. Ambas são importantes para a realização de

pesquisas e análises estatísticas adequadas para a tomada de decisões embasadas em evidências, porém existem diferenças entre elas.

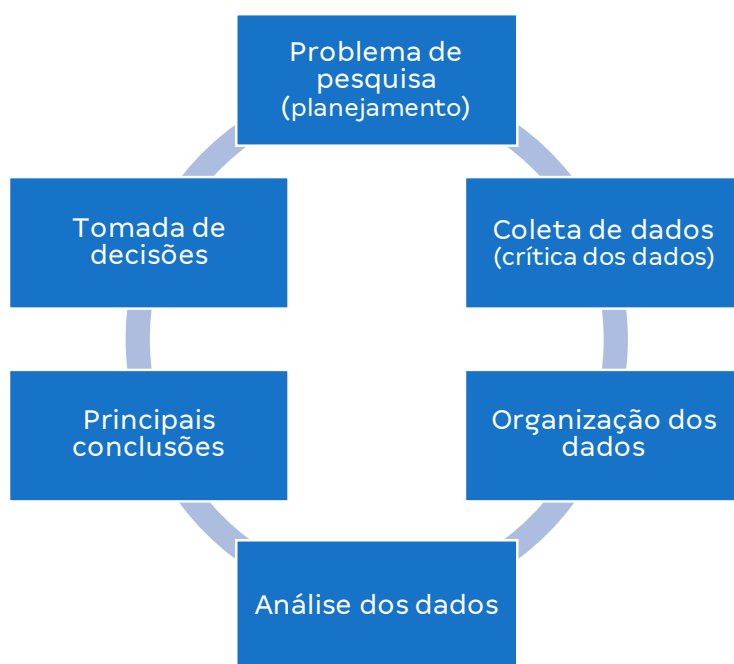
A estatística descritiva envolve a organização, sumarização e descrição de dados coletados em uma pesquisa ou estudo. Utilizando métodos como tabelas e gráficos, podemos analisar e medir dispersão e tendências dos dados. Por isso, a estatística descritiva é utilizada para fornecer informações úteis sobre os dados disponíveis e os tornar mais compreensíveis e acessíveis aos pesquisadores e tomadores de decisão.

Já a estatística indutiva envolve a análise e interpretação de dados de uma amostra selecionada a partir de uma população maior. Utilizando o método estatístico descritivo, nessa amostra, é possível fazer inferências sobre a população, a partir dos dados coletados na amostra. Esse ramo da estatística é mais usado para testar hipóteses, estimar parâmetros desconhecidos e fazer previsões. Tudo isso permite que os tomadores de decisão tenham uma boa noção das possibilidades antes de efetivamente tomarem a decisão.

O método estatístico, então, é um passo a passo, baseado em técnicas e medidas, para que possamos chegar a conclusões e inferir decisões a serem tomadas, de acordo com observações empíricas. Seja por meio da estatística descritiva ou da indutiva, é necessário seguir o passo a passo para garantir os melhores resultados.

4.1. Passo a Passo da Metodologia

Figura 3: Passo a passo da metodologia



Fonte: Elaboração própria.

O primeiro passo do método estatístico é detectar e definir o problema a ser estudado; aí, então, começa a etapa de **planejamento**. Nesse momento, é necessário estabelecer como os dados serão coletados, bem como o cronograma das atividades, os custos relacionados à pesquisa e a forma como será feita a análise dos dados e das informações já disponíveis antes de a pesquisa começar.

Com o planejamento feito, passamos para a **coleta de dados**. Essa é a busca e/ou compilação dos dados relevantes das variáveis e dos componentes do fenômeno em questão. Podemos separar o tipo de coleta de dados de duas formas distintas:

- Quanto à fonte.
 - Direta.
 - Os dados são obtidos diretamente da fonte.
 - Pode ser por meio de questionários, formulários, entrevistas e experimentos.
 - Indireta.
 - Os dados são obtidos de fontes secundárias ou registros existentes.
 - Eles podem ser conjuntos de dados governamentais ou registros médicos, por exemplo.
- Quanto ao tempo.
 - Contínuo.
 - É realizado de forma permanente, sem parar.
 - Como a coleta de informações do clima.
 - Periódico.
 - É realizado dentro de determinado intervalo de tempo.
 - Como o censo demográfico feito pelo IBGE de 10 em 10 anos.
 - Ocasional.
 - É realizado sem época preestabelecida.
 - Como uma pesquisa de popularidade sobre programas de televisão.

Faz-se necessário, então, que ocorra a crítica dos dados. Essa é uma subetapa da coleta que tem como objetivo eliminar erros que podem causar equívocos de apresentação e análise. Em resumo, trata-se de uma rigorosa revisão crítica dos dados.

Seguimos para a **organização dos dados**. Essa etapa é a de apresentação dos dados, para que se facilitem a leitura e a compreensão deles para os pesquisadores ou tomadores de decisão. Ao organizar os dados em tabelas ou gráficos, de acordo com as necessidades, de forma prática, sintética e racional, todos os envolvidos conseguem melhor entender tudo feito até o momento.

O que nos permite passar para a **análise de dados**. Nesse momento, aplicamos técnicas estatísticas para extrair informações relevantes a partir dos dados coletados e criticados. O objetivo, agora, é identificar padrões, tendências, relações e outras informações úteis que possam ser usadas para fazer inferências sobre a população em estudo. Todas as etapas são importantes para que a metodologia seja corretamente aplicada, mas é correto afirmar que a análise de dados é crucial, pois, a partir dessa etapa, são obtidas as conclusões que nos levam às possíveis decisões a serem tomadas.

Com a análise dos dados feita, é possível chegar à etapa das **principais conclusões**. Para que as conclusões sejam sólidas, é importante levar em consideração a correta interpretação dos dados e de suas análises, assim como o contexto de onde os dados foram coletados.

Por fim, depois de todo o processo estatístico, podemos chegar, finalmente, à **tomada de decisão** principal. A essa altura, devemos ter dados coletados de fatos concretos e corretamente criticados, bem como análises utilizando técnicas estatísticas que nos levam a conclusões baseadas em contexto e interpretação das análises. Sendo assim, os tomadores de decisão podem, enfim, tomar uma decisão em cima do problema que originou a necessidade da metodologia em primeiro lugar. Essa decisão será informada e baseada em todo um processo conclusivo e que pode ser conferido em qualquer uma de suas etapas para testar a validade de suas conclusões.

Considerações Finais da Aula

Nesta aula, vimos a área de conhecimento da matemática denominada de “estatística”. Apresentamos ferramentas e técnicas que são importantes em diversas áreas do conhecimento.

Conversamos sobre gráficos e tabelas, que são elementos essenciais para uma correta visualização dos dados. Entender qual é o melhor gráfico e em que situação cada um deve ser utilizado simplifica a compreensão da informação que queremos apresentar. Como vimos, gráficos em colunas são úteis quando se quer mostrar alterações de dados em um período ou simplesmente fazer comparações entre itens distintos.

Materiais Complementares

Artigo

Tipos de gráfico disponíveis no Office

2023, Microsoft.

Veja o site indicado que apresenta um material sobre cada tipo de gráfico disponível no Excel, bem como quando e onde cada um deve ser utilizado.

Link para acesso: <https://support.microsoft.com/pt-br/office/tipos-de-gr%C3%A1fico-dispon%C3%ADveis-no-office-a6187218-807e-4103-9e0a-27cdb19afb90> (acesso em 2 jun. 2023.)

Vídeo

Como Fazer GRÁFICOS no EXCEL em 2020 | Office 365

2020, Curso de Excel Online.

Apresenta o processo de criação de gráficos no Office 365.

Link para acesso: <https://youtu.be/Lk7raY4VuaA> (acesso em 2 jun. 2023.)

Referências

MOORE, David S.; NOTZ, William I.; FLINGER, Michael A. A estatística básica e sua prática. 7. ed. Rio de Janeiro: LTC, 2017. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788521634294>. Acesso em: 9 abr. 2023.