

Aula 2 – Estatística

Objetivo da Aula

Compreender a importância da estatística no raciocínio científico e os conceitos básicos e fundamentais que tornam a estatística possível.

Apresentação

Estatística é uma área da matemática responsável por levantar dados de um determinado fenômeno, população ou evento. Para alcançar esses objetivos, essa área da matemática relaciona fatos com números. Uma série de métodos permite que dados sejam coletados e então analisados para que possa haver alguma interpretação em cima dos dados tratados.

Dividida em duas partes, a **estatística diferencial** organiza, analisa e apresenta os dados, enquanto a **estatística inferencial** estuda uma amostra para que sejam realizadas as análises e apresentados os dados coletados.

1. Importância da Estatística

Estatística é a ciência de aprender a partir de dados coletados por um conjunto de métodos. Esses dados coletados, tratados e por fim apresentados, fornecem auxílio para a tomada de decisões baseada em matemática. Podemos ver exemplo de estatística sendo aplicada em todos os campos de atividade na sociedade moderna:

- Renda per capita.
- Desemprego.
- Taxa de crescimento populacional.
- Hospitais.
- Escolaridade.

Acima estão alguns poucos exemplos que já tomam uma enorme fatia da atuação humana quanto à estatística. Todos esses fatores listados acima são essenciais para que se possa acessar as condições de um país, estado ou cidade de acordo com a população que ali está. Munido dos dados tratados pelos métodos de estatística, é possível que os poderes competentes tomem decisões que, potencialmente, irão impactar a vida de toda a população local.

Tendo isso em vista, fica evidente a importância que esta escola de matemática, a estatística, exerce em nossas vidas, mesmo quando não a notamos diretamente. Os dados coletados pelos métodos de estatística são muito específicos e difíceis de serem lidos de forma singular, muitas vezes aquele que possui o poder de tomar decisões necessita de uma visão mais direcionada sobre as áreas que lhe cabem, é exatamente para isso que serve a estatística.

A estatística é dividida em duas partes essenciais e distintas:

- **Estatística Descritiva**, ou diferencial.
 - Sintetiza os dados. Por exemplo, gráficos e tabelas.
 - Possui medidas descritivas que permitem a fácil compreensão de grandes quantidades de informação. Por exemplo, média, desvio padrão e mediana.
- **Estatística Indutiva**, ou inferencial.
 - Com uma amostra de parte da população se infere e estima comportamentos de toda a população. Exemplos:
 - Amostras de lotes de mercadorias.
 - Controle de qualidade.
 - Amostras demográficas humanas.
 - Informações sobre o que a população pensa sobre determinado assunto.
 - Amostras demográficas comerciais.
 - Pesquisas sobre interesses em cima de produtos.

Como podemos ver, ambas as partes da estatística são ferramentas não somente extremamente úteis, mas absolutamente necessárias para a tomada de decisões informada sobre qualquer assunto em larga escala. O conhecimento dos conceitos básicos pode auxiliar na melhor compreensão dos resultados de pesquisas estatísticas.

Mas qual a função da estatística diante da ciência? Bem, a natureza apresenta variabilidade entre indivíduos e até mesmo o próprio indivíduo. Os padrões de variação coletados com os dados são então utilizados em estudos experimentais, ou observacionais, de forma a responder uma questão científica.

Destaque

Podemos enxergar a estatística como a *tecnologia da ciência*. Ou seja, estatística é o método que valida a pesquisa científica.



Você Sabia?



Aqui no Brasil o órgão responsável por estatística é o IBGE (Instituto Brasileiro de Geografia e Estatística). O IBGE disponibiliza diversos vídeos, artigos e sites mostrando o resultado de suas pesquisas, sendo a maior delas o grande Censo Demográfico, que ocorre a cada 10 anos e busca contabilizar toda a população brasileira e suas variáveis. Vamos ver um vídeo de seu canal no Youtube <https://www.youtube.com/watch?v=s5yrDV_c2-4>

2. Conceitos Básicos

Para compreender corretamente os resultados, é necessário estar preparado e munido de conhecimentos fundamentais. Mesmo que não seja quem faça a pesquisa estatística, pode chegar um ponto em que será necessário interpretar resultados ou até mesmo exigir pesquisas, portanto é importante saber navegar nesse mundo. Vamos começar a compreender os conceitos básicos de estatística definindo um pequeno glossário de termos utilizados nessa ciência:

Dados: Observações de pelo menos uma variável; o que é efetivamente coletado para as pesquisas.

Variável: Aquilo que se observa no indivíduo.

EXEMPLO

Sexo, peso, tamanho, sabor, salário e outras.

Variáveis qualitativas	Características de uma população que não tem como serem medidas.	Ordinais (Classificação)	Pressão sanguínea, grau de gravidade de uma doença.
		Nominais (Classificação)	Sexo, raça, escolaridade
Variáveis Quantitativas	Características que podem ser quantificadas em uma população.	Discretas (Contagem)	Quantidade de filhos, quantidade de imóveis
		Contínuas (Mensuração)	Peso, altura, salário

População: Conjunto de indivíduos ou objetos que apresenta ao menos uma característica em comum. O objetivo da pesquisa é analisar ou inferir algum aspecto do comportamento da população.

EXEMPLO

Para pesquisar quantos adolescentes estão matriculados em escola pública ou particular de um determinado município.

População alvo – Todos os adolescentes matriculados em escolas do município.

Amostra: Subconjunto da população.

- Deve se levar em consideração a representatividade da amostra, seu tamanho diante da população original e sua aleatoriedade.
- A única forma de garantir a lisura do processo é que todos os elementos tenham chance de participar e que não exista um sorteio viciado para quem participa.

Parâmetro: Alguma medida numérica capaz de descrever uma característica de uma determinada população. Costuma ser demonstrado por letras gregas.

EXEMPLO

Para determinar a qualidade do ensino de matemática das escolas públicas e particulares, é necessário estabelecer a média de notas dos alunos matriculados no município.

Estatística: Medida numérica que descreve alguma característica de interesse de uma amostra.

EXEMPLO

A média de nota da amostra de alunos acima destacados.

Destaque

O levantamento da população pode ocorrer de duas formas: censo ou amostragem. O censo é a totalidade da população alvo, enquanto a amostragem é apenas parte que representa e infere o todo.



Os termos que encontramos nesse glossário são, em maior parte, intuitivos, porém é necessário saber diferenciá-los para que se garanta a eficácia da pesquisa estatística. Esse glossário nos possibilita ler um resultado estatístico, mas ainda não nos diz nada sobre os tipos de pesquisa que exigem o auxílio de estatística.

3. Pesquisas Científicas e a Influência da Estatística

Uma pesquisa científica pode ser de levantamento ou experimental. A **pesquisa de levantamento** é utilizada quando observamos, ou medimos, características de interesse de uma população sem manipulá-la. Dessa forma, podemos dizer que as características de interesse são **levantadas** da amostragem. Um exemplo disso é um estudo demográfico de consumidores de um produto em um país. Neste exemplo, temos como população todo o país. Seria inviável e muito caro analisar todos os indivíduos, então é feita uma amostragem que é analisada e processada para então inferirmos o resultado de toda a população original.

Já uma **pesquisa experimental** é quando um grupo de indivíduos, que podem ser humanos, animais ou até mesmo objetos, são manipulados para avaliar o efeito dessas manipulações e suas variações. Um exemplo disso é uma pesquisa de remédio, em que manipulamos indivíduos que recebem tratamentos para que possamos avaliar suas variações e, assim, testar corretamente a eficácia.

Ao iniciarmos uma pesquisa, é necessário muito preparo para garantir que os resultados não serão viciados e que sejam comprováveis. Devemos prestar atenção ao fato de que uma pesquisa não é sobre provar uma teoria ou hipótese do pesquisador, mas testar e colocar a prova isso tudo. Uma pesquisa em cima de uma teoria, ou tese, que venha a se provar negativa ainda mantém o seu valor, pois foi atestado que aquele não é o caminho “correto” e que a hipótese estava errada.

Para iniciar esse processo, primeiro é necessário **definir qual é o problema** em questão. Ao reconhecer um problema a ser testado, deve-se procurar formulá-lo da forma mais correta possível e definir qual a população a ser estudada. Dentro dessa população, precisamos definir quais variáveis vão ser observadas. Isso tudo é para que fique claro quais são as hipóteses testadas e exatamente o que vai ser investigado. Somente então os estudos e a pesquisa efetivamente começam.

Entra-se então em uma fase mais burocrática, a de **planejamento da pesquisa**. Nessa etapa, são definidos quais objetivos, bibliografias, materiais, equipamentos e local de trabalho para execução da pesquisa. Se faz necessário também definir se vai ser levantado um censo ou amostragem devido ao tamanho total da população alvo. Por fim, mas não menos importante, também são definidos quem vai estar envolvido no trabalho e qual o cronograma de execução. Ainda no planejamento, é firmado de vez as características a serem observadas, as variáveis, os elementos, a população e como isso vai ser feito, qual vai ser o questionário, ou entrevista, a ser estruturado como elemento de pesquisa.

Com tudo descrito acima definido, chegamos agora na execução dos estudos, para que isso ocorra é necessário que haja a **coleta de dados**, que podemos chamar de variáveis.

Agora é “mão na massa”. Com os instrumentos de pesquisa, que são os prontuários, protocolos e questionários, começam a se coletar as informações primárias e secundárias dos objetivos da pesquisa e alvo dos estudos. Um trabalho quase mecânico e um tanto repetitivo, principalmente dependendo do tamanho da população/amostragem, porém muito importante para que o processo ocorra.

Com os dados em mãos, agora começa a etapa de **apuração de dados**. Nesse momento, os dados são agrupados, resumidos e validados, normalmente com auxílio computacional (como as tabelas do Excel) para melhor visualização e compreensão dos dados. É formado um banco de dados que compreende toda a amostragem/população e seus dados, ou variáveis, coletados.

As próximas duas etapas são parcialmente misturadas, pois uma auxilia muito a compreensão da outra, estou me referindo a **análise e interpretação dos dados**, assim como **apresentação dos dados**. A análise dos dados deve ser feita utilizando programas estatísticos, tais como R, STATA, SAS e outros, seu objetivo é chegar a conclusões que auxiliem o pesquisador a concluir sua pesquisa e efetivamente testar os objetivos da pesquisa e do estudo, provar ou desaprovar algo. A análise estatística volta a ser dividida em suas duas partes de acordo com a necessidade, a estatística descritiva e a estatística indutiva.

Então a apresentação dos dados toma forma com as duas partes da estatística. Enquanto a estatística indutiva conclui os testes estatísticos, a estatística descritiva busca organizar e descrever os dados coletados e concluídos pela pesquisa. Repare como uma depende da outra, para conseguirmos chegar a conclusões da estatística indutiva, é necessário que sejamos capazes de olhar uma síntese compreensiva dos dados relevantes, tal qual a estatística descritiva nos proporciona.

Para garantir a plenitude do estudo, é necessário analisar todas as variáveis desejáveis. A partir disso, são montados alguns elementos básicos, como tabelas, gráficos e medidas numéricas.

4. Gráficos e Medidas

Sem entrar nas contas e informações específicas para alguns destes métodos, é possível compreendê-los, e até mesmo necessário em alguns casos, para que se entenda o que está sendo afirmado pelo estudo.

O objetivo na criação de um gráfico é que ele seja autoexplicativo, seja lá qual for o seu formato ou método de aplicação, ele deve ser autossuficiente para que só de olhar já fique clara a distribuição das variáveis nele. Acontece que, enquanto seres humanos, temos mais facilidade de compreender algumas informações, principalmente em se tratando de grandes populações, ao observá-las organizadas em gráficos que sintetizam bem os achados do estudo.

Uma **Tabela de Frequências** é quando é avaliado os valores das variáveis e então é definida a frequência em que determinado valor aparece. É possível fazer também a frequência em qual as variáveis não alcançam um valor específico determinado com as necessidades dos estudos.

Para garantir que o gráfico possua essa qualidade de autossuficiência, existem alguns elementos que devem ser incluídos obrigatoriamente. Estes são: **título**, que antecede e explica tudo referente à tabela; **cabeçalho**, que especifica o conteúdo das colunas; e o **corpo**, que corresponde ao conjunto de colunas, linhas, fatias, ou seja lá qual for o método de apresentação do gráfico.

Além desses elementos que tornam a leitura do gráfico possível para aquele não integrado com o projeto, existem elementos complementares a serem adicionados de acordo com a necessidade e relevância. Estes são; a **fonte**, que especifica de onde os dados surgiram; as **notas**, que esclarecem o conteúdo ou a metodologia adotada na coleta e preparo dos dados; e as **chamadas**, que explicam itens específicos na tabela conforme necessário.

É necessário também algumas padronagens, como a utilização de algarismos arábicos, a uniformidade das distâncias indicadoras de unidades e o fato de que o crescimento deve ser sempre coerente, ou seja, da esquerda para a direita e de baixo para cima. Dessa forma, garantimos que os gráficos não necessitem de nenhum esclarecimento adicional quanto ao que representam e têm a informar.

4.1. Medidas

- Medidas de Tendência Geral.
- Medidas Separatrizes.
- Medidas de Dispersão ou Variabilidade.

Quando falamos de medidas, estamos falando de métricas que nos oferecem informações quanto aos dados coletados para estudo. Estas podem ser divididas em três tipos que vamos ver sintetizadas aqui.

Começando pelas **Medidas de Tendência Geral**, seus objetivos são definir valores médios das variáveis estudadas. Dessa forma, é possível sintetizar todos os dados coletados relacionados àquela variável em uma única medida. Existem três tipos de medidas de tendência geral:

- Média Amostral
 - Média matemática simples, soma-se os valores da amostra e divide pela quantidade de indivíduos.

- Mediana
 - Divide a amostragem em duas partes iguais e diz qual valor é o mais central.
 - Coloca-se as variáveis em ordem crescente e então:
 - Caso de dimensão ímpar, o valor que se encontra absoluto no meio é a mediana.
 - Caso de dimensão par, é calculada a média entre os dois valores mais centrais e essa média é definida como mediana.
- Moda
 - Bem frequente em estudos estatísticos.
 - Refere-se à frequência obtida na tabela de frequência.
 - A mesma amostra pode ter mais de uma moda.
 - A moda é basicamente o valor, ou valores, mais frequente dentro de uma amostragem.

Agora vamos para as **Medidas Separatrizes**. Como o nome sugere, essas medidas separam e dividem as distribuições, sempre em partes iguais. Estas são mais simples conceitualmente:

- Quartis
 - São os valores que ficam na divisa ao separar a amostra em quatro partes iguais.
- Decis
 - São os valores que ficam na divisa ao separar a amostra em dez partes iguais.
- Percentis
 - São os valores que ficam na divisa ao separar a amostra em cem partes iguais.

Por fim, vamos às **Medidas de Variabilidade**. Esta busca trazer mais informações que ficam faltando nas medidas de tendência central. Acontece que é comum muita variabilidade dos valores dentro da mesma amostra, então essas medidas buscam demonstrar como é essa variação de fato ao medir a variação e não as médias. Essa medida pode ser comparada entre amostras para ver qual delas é mais heterogênea. Estas são:

- Amplitude total
 - Diferença calculada entre o maior e o menor valor das variáveis da amostra.
- Variância
 - Indica o quanto uma amostra se dispersa e varia da média.

- Desvio Padrão
 - É a raiz quadrada da variância.
 - É a medida mais usada na comparação de diferenças entre grupos.
 - Facilita a especificação de quanto acima ou abaixo da média está um determinado valor.
 - Quanto maior o desvio padrão, maior a variabilidade dos valores.

Considerações Finais da Aula

Ao compreendermos o básico de estatística se torna evidente o porquê essa ciência é tão importante para a tomada de decisões importantes de governos e empresas de diversos tamanhos. Os termos “métricas” e “valores” empregados em estatística são essenciais para que sejam tomadas decisões informadas de grande escala.

Estruturada para coletar, analisar, apurar e apresentar os dados, a estatística está entremeada a tal ponto com o processo de estudo e com a pesquisa científica que um não existiria sem o outro, são codependentes.

Considerando a natureza autossuficiente tão almejada nos gráficos, um conhecimento básico sobre os termos, métricas e valores já sustenta a compreensão da metodologia ali aplicada.

Materiais Complementares



Vídeo

IBGE Explica

2015, IBGE.

A série “IBGE Explica” é uma série de vídeos que explica vários fatores estatísticos (e geográficos, como é de natureza do IBGE) de forma extremamente didática, rápida e simples, vale checar.

Link para acesso: https://www.youtube.com/playlist?list=PLAvMMJyHZEaE_gzGaoD5R-kmCxO6rBexl6.

Referências

BECKER, João L. *Estatística Básica*. Porto Alegre: Grupo A, 2015. E-book.

ISBN 9788582603130. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788582603130/>. Acesso em: 09 mar. 2023.

COSTA, Giovani Glaucio de O. *Curso de Estatística Básica*, 2ª edição. São Paulo: Grupo GEN, 2015. E-book. ISBN 9788522498666. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788522498666/>. Acesso em: 09 mar. 2023.

SPIEGEL, Murray R.; STEPHENS, Larry J. *Estatística*. Porto Alegre: Grupo A, 2009. E-book. ISBN 9788577805204. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788577805204/>. Acesso em: 09 mar. 2023.