

# Estatística Descritiva

# Sumário

- ❑ Apresentação
- ❑ Conceitos fundamentais
- ❑ Prática

# Apresentação



# Stefano Mozart

Cientista de dados no Conselho Administrativo de Defesa Econômica, atuando no combate a cartéis e outras práticas anticompetitivas com o uso de modelos econométricos e estatísticos.

**Disclaimer:**

O conteúdo deste curso reflete apenas a visão pessoal do autor e, portanto, não representa a posição oficial do Cade ou de qualquer outra entidade do Governo Federal.

❑ **Método de ensino:**

- ❑ Aulas expositivas (8 aulas, no total), seguidas de exemplos e exercícios práticos;
- ❑ Os exercícios serão realizados em aula, com auxílio do professor e de monitores;

❑ **Avaliação:**

- ❑ Trabalho final consistindo de apresentação, composta por caderno de análise e vídeo expositivo, de um conjunto de dados à escolha do aluno;
- ❑ Os exercícios contribuirão para a realização do trabalho final, por isso recomenda-se a escolha do conjunto de dados o quanto antes.

# Estrutura do Curso

# Material

## ❑ Programa da disciplina:

[https://docs.google.com/document/d/17YDqld6qE9PeF57uk8CzDxIx\\_FzTHYITmBXGPg5wddRQ/edit?usp=sharing](https://docs.google.com/document/d/17YDqld6qE9PeF57uk8CzDxIx_FzTHYITmBXGPg5wddRQ/edit?usp=sharing)

## ❑ Ambiente de prática/exercícios:

- ❑ Serviço com Jupyter Notebook publicamente acessível (para permitir interação e correção dos exercícios, bem como avaliação do trabalho final);
- ❑ Grupo do WhatsApp, para facilitar a comunicação com o professor e os monitores durante a realização de exercícios;
- ❑ Cadernos Jupyter com exercícios e exemplos, no repositório git do curso: [https://github.com/stefanomozart/estatistica\\_descritiva](https://github.com/stefanomozart/estatistica_descritiva)

## ❑ Bibliografia básica:

- ❑ BUSSAB W. MORETTIN, P. **Estatística Básica**. São Paulo: ED Atual, 1986, 3ª ed.
- ❑ McKINNEY, W. **Python Para Análise de Dados: Tratamento de Dados com Pandas, NumPy & Jupyter**. São Paulo: Novatec Editora, 2023, 3ª ed.
- ❑ YWATA, A. X; CAJUEIRO, D. O; CAMARGO, R. S. **Introdução aos Métodos Estatísticos Para Economia e Finanças**. Brasília: Editora UnB, 2015, 1ª ed.

# Conceitos fundamentais





# Estatística

É a disciplina que se ocupa da coleta, organização, análise, interpretação e apresentação de dados a fim de dar suporte ao entendimento de fenômenos complexos, à previsão de eventos futuros e à tomada de decisão.

## ❑ **Estatística descritiva:**

- ❑ Ramo da estatística que foca na descrição e sumarização de conjuntos de dados.
- ❑ Envolve a coleta, tabulação e a apresentação de dados de modo que seja possível fornecer uma visão compreensível de informações através de tabelas, gráficos e medidas numéricas.

## ❑ **Estatística Inferencial:**

- ❑ Foca em técnicas que produzem generalizações acerca de uma população a partir de uma amostra de dados.
- ❑ Utiliza modelos probabilísticos com o fim de produzir de estimativas, testar hipóteses e inferir comportamentos futuros, ou que extrapolem os limites amostrais, para as variáveis analisadas.





# População

Totalidade de indivíduos ou elementos de interesse: que são objeto de estudo ou que possuem uma característica em comum.

- ❑ Finita (e.g. todos os habitantes de uma cidade);
- ❑ Infinita (todos os produtos com uma certa classificação CGCE/IBGE);
- ❑ Teórica (possíveis observações em 50 lançamentos simultâneos de três dados);
- ❑ Concreta (pacientes tratados com um determinado medicamento);
- ❑ Alvo (estudantes com renda mensal *per capita* inferior a R\$ 300,00);
- ❑ Acessível (estudantes com renda mensal *per capita* inferior a R\$ 300,00 que informaram sua renda e forneceram dados para contato);



# Amostra

- ❑ **Amostra:**

- ❑ Subconjunto representativo da população, selecionado para viabilizar a análise sem a necessidade de arrolar a totalidade dos elementos de interesse.
- ❑ As medidas estatísticas observadas na amostra viabilizam a realização de estimativas confiáveis para as medidas populacionais;

- ❑ **Amostragem:**

Técnica empregada na seleção/arrolamento dos elementos da amostra.

# Amostragem

## ❑ Aleatória Simples (AAS):



Cada membro da população tem igual probabilidade de ser escolhido. A seleção é completamente ou *quasi*-aleatória, o que pode ser feito com suporte de tabelas de números aleatórios ou software gerador de números aleatórios ou pseudo-aleatórios.



## Estratificada:

A população é dividida em estratos, ou grupos, que são homogêneos em relação a alguma característica. Uma amostra aleatória simples é então retirada de cada estrato. Este método assegura que cada categoria da população seja adequadamente representada na amostra final.



## Sistemática:

A partir de um ponto de partida aleatório, são selecionados elementos em intervalos regulares. Por exemplo, em uma lista de 1.000 pessoas, pode-se escolher uma pessoa a cada 10 para formar uma amostra de 100 pessoas.



# Amostragem

## ❑ **Por Conglomerados:**



A população é dividida em conglomerados que são representativos da população inteira. Alguns desses conglomerados são selecionados aleatoriamente e todos os elementos dentro dos conglomerados escolhidos são incluídos na amostra. Este método é muitas vezes usado quando a população é geograficamente dispersa.

## ❑ **Multi-estágio:**



Método que pode combinar diferentes técnicas de amostragem probabilística, especialmente estratificação e clusterização, em sequência:

- ❑ Geralmente associada censos e outros grandes estudos;
- ❑ Introduce bastante complexidade nos cálculos de erros amostrais e intervalos de confiança;



# Amostragem

## ❑ Não Probabilística:



Método ou heurística sem suporte de processos aleatórios:

- ❑ Amostragem por Conveniência: seleção de elementos mais fáceis de acessar.
- ❑ Amostragem por Julgamento ou Intencional: critérios de utilidade ou representatividade para a análise.
- ❑ Amostragem por Quotas: busca garantir que certas características sejam representadas na amostra em proporções semelhantes à população.
- ❑ Amostragem Bola de Neve: indivíduos indicam novos participantes;



# Variável

Uma característica, ou atributo, associada a cada elemento de uma população ou amostra. Também pode ser entendida como qualquer dimensão do conjunto de dados em análise que permite medir e expressar características de interesse de forma quantificável.

- ❑ Essas características podem se apresentar de diversas formas, manifestando uma natureza qualitativa ou quantitativa;
- ❑ Identificar e classificar corretamente as variáveis é crucial para o sucesso de qualquer análise estatística, pois isso influencia desde o design da coleta de dados até as técnicas de análise e a interpretação dos resultados;
  - ❑ Na coleta, a definição das variáveis garante a precisão e relevância dos dados, tendo em vista as perguntas de pesquisa;
  - ❑ Na análise, a correta interpretação e comunicação das variáveis ajuda tanto a moldar quanto a comunicar resultados;



# Variável qualitativa

Informação de natureza categórica, não quantificável. Tem forte relação com a análise de frequência e de distribuição de frequência.

- ❑ **Nominal:** Representam categorias sem qualquer ordem natural entre elas. A definição da variável é complicada pelo fato de que elementos podem ser associados a um ou mais valores. Exemplos incluem cor dos cabelos, gênero, ocupação, nacionalidade.
- ❑ **Ordinal:** Semelhantes às variáveis nominais, mas com uma ordem ou hierarquia definida. Elementos podem estar associados a apenas um valor. Exemplos incluem níveis de educação (fundamental, médio, superior); e classificações de percepção de qualidade (uma, duas, três estrelas) ou intensidade (discordo totalmente, discordo, etc);



# Variável quantitativa

Informação de natureza numérica que quantifica uma característica do elemento. São fundamentais numa análise estatística descritiva, uma vez que permitem a aferição direta de medidas estatísticas.

- ❑ **Discreta:** Assumem valores inteiros ou contáveis. Geralmente resultam de processos de contagem e têm um número finito ou enumerável de valores possíveis. Exemplos incluem o número de filhos, número de carros em um estacionamento.
- ❑ **Contínua:** Podem assumir qualquer valor dentro de um intervalo contínuo. Geralmente resultam de medições e podem incluir valores fracionários. Exemplos incluem altura, peso, temperatura;
- ❑ **Obs.:** Toda variável quantitativa pode ser expressa na forma de uma variável qualitativa por meio da especificação de intervalos;





# Frequência

Quantidade de observações de um valor específico para uma variável. Obtida a partir da simples contagem das observações na amostra, permite a identificação de padrões, bem como da existência de eventos anômalos, a partir da esporadicidade ou raridade de certas observações.

- ❑ **Absoluta:**

Número exato de vezes que um valor específico aparece em um conjunto de dados.

- ❑ **Relativa:**

Proporção ou porcentagem de ocorrências de um valor específico frente ao total de observações. É calculada dividindo a frequência absoluta pelo total de elementos.

- ❑ **Frequência Acumulada:**

Soma das frequências de todas as classes anteriores até a classe atual. Ajuda a entender quantas observações recaem em intervalos ou categorias.



# Distribuição de Frequência

Uma sumarização ou tabulação dos dados brutos que tem por objetivo demonstrar como as observações são distribuídas ao longo de diferentes categorias ou intervalos numéricos.

- ❑ Para variáveis quantitativas, os valores são agrupados em intervalos (também conhecidos como *bins*).
- ❑ Para variáveis qualitativas, as categorias são geralmente definidas pelas próprias características dos dados (os distintos valores admitidos ou efetivamente observados para cada variável).
- ❑ É crucial para determinar a forma da distribuição dos dados, como simetria, assimetria e a presença de modas.



# Medidas estatísticas

Valores numéricos calculados a partir de um conjunto de dados com o objetivo de resumir ou descrever características relevantes desses dados. Em última instância, essas medidas visam dar suporte à compreensão de características da população representada.

## ❑ Medidas de Tendência Central:

- ❑ Média;
- ❑ Mediana;
- ❑ Moda;

## ❑ Medidas de Dispersão:

- ❑ Variância;
- ❑ Desvio Padrão;
- ❑ Amplitude;
- ❑ Intervalo Interquartílico (IIQ);

## ❑ Medidas de Posição:

- ❑ Quartis;
- ❑ Decis
- ❑ Percentis;
- ❑ Quantis;

## ❑ Medidas de Forma:

- ❑ Curtose;
- ❑ Assimetria;

## ❑ Medidas de Relação:

- ❑ Covariância;
- ❑ Correlação.



# Viés

Distorção sistemática que ocorre durante a coleta, organização ou análise dos dados, levando a uma representação imprecisa ou não representativa da realidade ou da população estudada. Esse erro não é fruto do acaso, mas sim de falhas nos métodos de amostragem, registro ou apresentação.

- ❑ **Viés de Seleção:**  
Quando a amostra escolhida não representa adequadamente a população, fazendo com que os resultados fiquem sistematicamente distorcidos.
- ❑ **Viés de Medição:**  
Quando os instrumentos ou métodos de coleta de dados possuem falhas ou imprecisões que afetam consistentemente as medições.
- ❑ **Viés de Apresentação:**  
Quando a forma de organizar ou exibir os dados enfatiza ou minimiza certas características, podendo levar a interpretações equivocadas.

# Prática





## Tópicos para discussão:

- ❑ Qual é a utilidade da estatística descritiva na análise de políticas públicas?
- ❑ Qual a importância de uma amostra representativa?
- ❑ Quais os desafios mais relevantes no levantamento de uma amostra representativa?
- ❑ Em que situações, no contexto da análise de políticas públicas, cada técnica de amostragem pode ser mais apropriada?
- ❑ Quais seriam os maiores desafios para definição, coleta e análise de dados em políticas públicas?



## Fontes de dados:

- ❑ <https://dados.gov.br/dados/conjuntos-dados>  
(exige autenticação com Gov.br)
- ❑ <https://sidra.ibge.gov.br/>
- ❑ <https://basedosdados.org>
- ❑ [https://github.com/stefanomozart/estatistica\\_descritiva/tree/main/dados/originais](https://github.com/stefanomozart/estatistica_descritiva/tree/main/dados/originais)
- ❑ Sugestões?

“Statistics is the  
grammar of  
Science”

Karl Pearson



# Obrigado

**Stefano Mozart**

[linkedin.com/in/stefano-mozart/](https://linkedin.com/in/stefano-mozart/)

[github.com/stefanomozart](https://github.com/stefanomozart)

