

3. Planejamento estatístico

CONCEITOS BÁSICOS

3.1

O que é estatística?



Arte de torturar os dados até obter os resultados esperados!



A **estatística** é uma ciência que se preocupa com a coleta, organização, simplificação, análise e interpretação dos dados, assim como tirar conclusões sobre as características das fontes de onde estes foram retirados, para melhor compreender as situações.



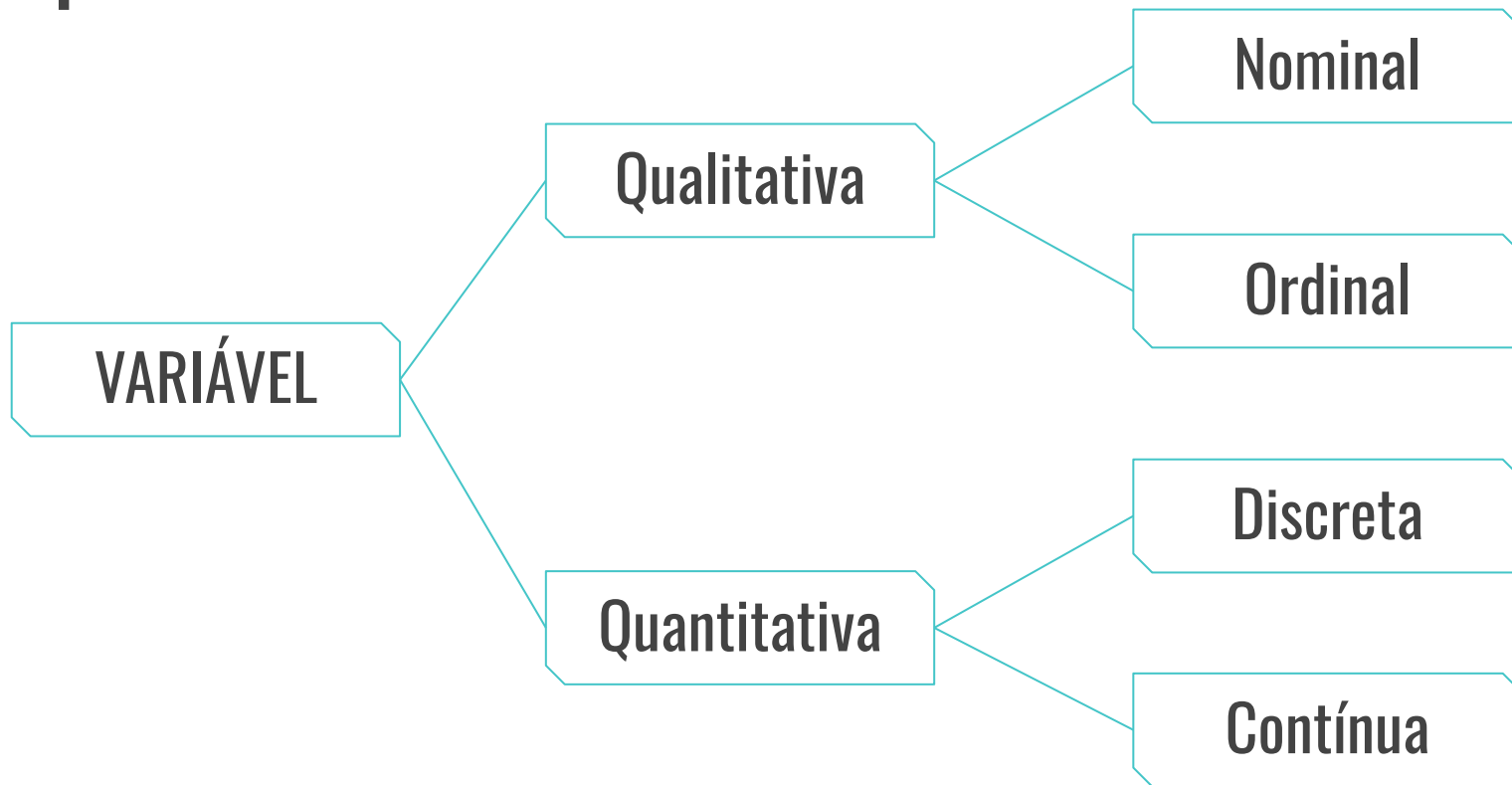
Conjunto de dados

Um conjunto de dados estatísticos consiste de uma ou mais medidas, escores ou valores observados (coletados) de certo número de indivíduos, animais, objetos, ensaios, experimentos, etc.

Variável:

Variável é a característica de interesse que é medida em cada elemento da amostra ou população, é aquilo que se deseja observar para tirar algumas conclusões. Como o nome diz, seus valores **variam** de elemento para elemento.

Tipos de variáveis



Variável qualitativa:

tem como característica categorias ou classes.

Nominal:

Apenas identifica um atributo à unidade observacional sem qualquer outra propriedade.

Ex: região de procedência, cor dos olhos.

Ordinal:

Identifica um atributo que estabelece uma estrutura de ordem nas unidades de observação.

Ex: nível de escolaridade, classe social.

Variável quantitativa:

pode ser medida em uma escala quantitativa, ou seja, apresenta valores numéricos que fazem sentido.

Discreta:

Pode assumir um conjunto finito ou enumerável de valores.

Ex: número de filhos, quantidade de televisores na residência.

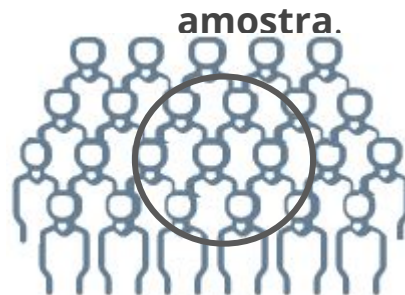
Contínua:

Pode assumir infinitos valores num intervalo de números reais.

Ex: peso, altura.

População e amostra

- ✔ **Problema típico de estatística aplicada:** conhecer certas características dos elementos de uma população, com base nos dados de uma amostra.
- ✔ Chamamos de **população** o conjunto de elementos que formam o universo de nosso estudo e que são passíveis de serem observados. Uma parte destes elementos é dita uma **amostra**.
- ✔ **N:** tamanho da população.
- ✔ **n:** tamanho da
- ✔ **Censo:** coleta de dados realizada com toda a população.
- ✔ **Amostragem:** coleta de dados realizada com uma parte da população.



Tipos de levantamento

- ✔ **Estudos Retrospectivos:** os dados são obtidos a partir de uma coleta de informações históricas relacionadas ao problema investigado.
- ✔ **Estudos Observacionais:** as características da população são observadas ou medidas sem que se faça manipulação
- ✔ **Experimentos Planejados:** as características da população são observadas ou medidas por pesquisadores que manipulam os dados para avaliar o efeito de diferentes tratamentos.

Amostragem

Nas pesquisas científicas, em que se deseja conhecer algumas características da população, essas características chamamos de **parâmetros**.

Podemos observar apenas uma amostra de seus elementos e, com base nos resultados da amostra, obter valores aproximados, ou **estimativas**, para os parâmetros de interesse. Esse tipo de pesquisa é usualmente chamada de **levantamento por amostragem**.

O uso apropriado dos dados de uma amostra para se ter conhecimento sobre parâmetros da população onde foi extraída a amostra é chamado de **inferência**.

Exemplos: proporção de pessoas contaminadas por uma doença, tempo médio de serviço de funcionários de uma empresa.

Plano de Amostragem

Para elaborar um **plano de amostragem**, devemos ter bem definidos:

- ✔ os objetivos da pesquisa,
- ✔ a população a ser amostrada,
- ✔ os parâmetros que precisamos estimar para atingir os objetivos da pesquisa,
- ✔ a unidade de amostragem,
- ✔ a forma de seleção dos elementos da população e
- ✔ o tamanho da amostra.

Unidade de amostragem pode ser os próprios elementos da população ou outras unidades que sejam mais fáceis de serem selecionadas, mas que tenham correspondência com os elementos da população.

Características importantes para o sucesso da pesquisa por amostragem

- ✔ **Representatividade:** a amostra precisa conter todas as subdivisões da população,
- ✔ **Suficiência:** quantidade tal que permita caracterizar a variabilidade, mesmo dentro das subdivisões da população e
- ✔ **Aleatoriedade:** necessária para a generalização estatística.

Tipos de amostragem

Amostragem Aleatória

A amostragem **probabilística** ou **aleatória** reúne todas as técnicas que usam mecanismos aleatórios de seleção dos elementos de uma amostra distribuindo a cada um deles uma probabilidade conhecida a priori de pertencer a amostra.

Aqui podemos fazer inferência a respeito da população.

Amostragem Não Aleatória

Na amostragem **não probabilística** ou **não aleatória** os elementos da população são selecionados de forma não aleatória, assim a probabilidade de seleção não é conhecida.

Aqui não é possível fazer inferência a respeito da população.

Tipos de amostragem aleatória:

**Amostragem aleatória simples
(AAS)**

Amostragem sistemática

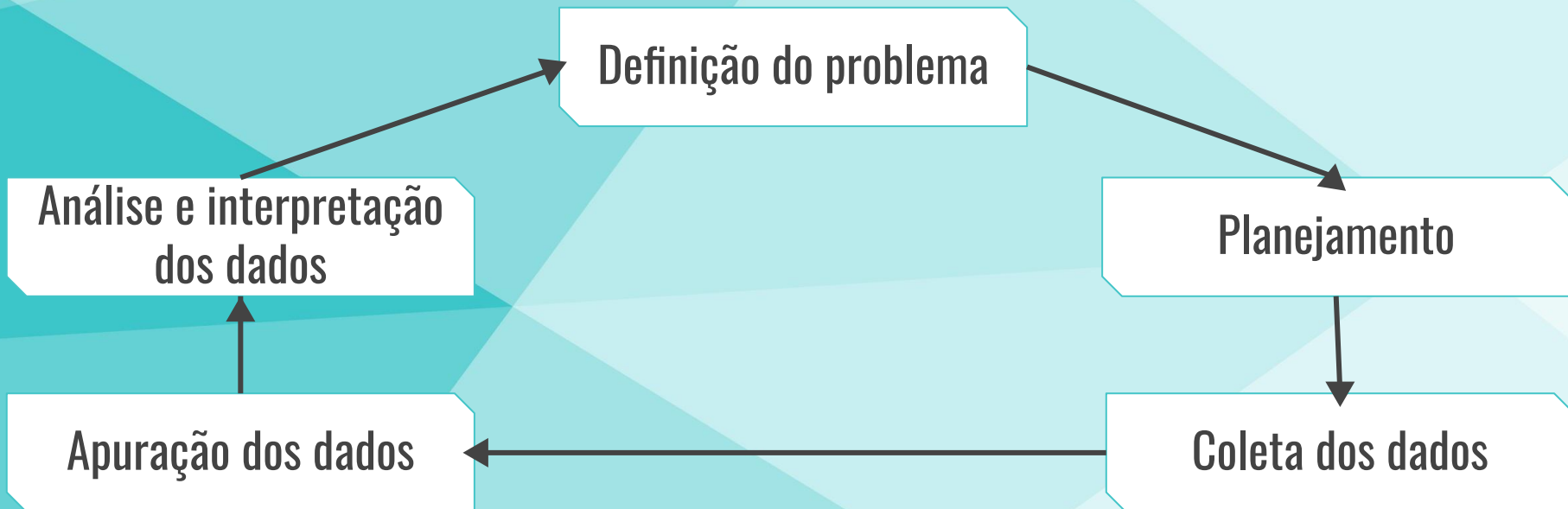
Amostragem estratificada

Amostragem por conglomerados

Metodologia Científica

Conjunto de métodos e técnicas para analisar dados e tomar decisões com base na incerteza.

Etapas da metodologia científica:



Organização de dados em planilhas

3.2

A importância da Visualização de Dados



flai



**DADOS
BRUTOS**



**DADOS
ORDENADOS**



**DADOS
ORGANIZADOS**



**DADOS
VISUALIZADOS**

Fonte: FLAI Inteligência Artificial <https://www.instagram.com/p/CErdQJJAAbj/>

Exemplo de um conjunto de dados

	B	C	D	E	F	G	H	I
2	Os dados a seguir são referentes à informações de uma amostra de equipamentos produzidos pela fábrica X.							
3								
4	Identificação do equipamento	Comprimento (em mm)	Tempo de funcionamento até a quebra (em dias)	Tipo	Cor	Potência	Empresa compradora	Número de parafusos utilizados
5	1	90	44,8	B	prata	50	C	11
6	2	90	45,8	B	preto	65	B	12
7	3	94	47,2	B	prata	65	B	12
8	4	94	44,9	B	prata	50	B	14
9	5	94	44,1	C	preto	65	A	15
10	6	92	57,8	C	prata	60	B	15
11	7	95	52,6	B	preto	65	B	10
12	8	92	44,5	C	prata	60	C	14
13	9	93	48,3	B	prata	55	B	11
14	10	90	46,2	C	preto	55	A	13
15	11	90	45,2	A	preto	65	A	15
16	12	93	50,3	B	preto	55	A	12
17	13	87	54,9	A	prata	60	C	16
18	14	93	52,5	B	prata	60	A	12
19	15	96	46,4	B	prata	55	B	12
20	16	92	52,0	A	prata	65	A	14
21	17	86	50,6	B	preto	65	B	12
22	18	94	42,5	C	preto	55	B	13

Como organizar dados em planilhas

- ✔ **Uma variável por coluna, uma unidade amostral por linha.**
- ✔ Cabeçalho com **nomes informativos**.
 - ▽ Evite utilizar letras maiúsculas, caracteres especiais (acentos, barras, vírgulas).
 - ▽ Evite espaços (use, por exemplo, underline ou hífen).
 - ▽ Caso opte por usar siglas, adicione uma aba na planilha explicando o que é cada coluna (metadados).
- ✔ A **aba dos metadados** pode conter também as unidades de medida adotadas, informações do delineamento experimental, datas de atividades e todas as informações relevantes daqueles dados.
- ✔ **ID** geralmente é uma coluna inserida na planilha e representa a identificação da unidade amostral. Essa identificação não é utilizada para analisar os dados, mas é útil para sabermos a qual das unidades amostrais estamos nos referindo.

Exemplo

Usando as dicas:



	A	B	C	D
1		Tratamento 1	Tratamento 2	
2	1	35	15	
3	2	40	18	
4	3	32	20	
5	4	30	12	
6	5	38	15	
7				

	A	B	D
1	grupo_experimental	comprimento_corpo	
2	com_predadores	35	
3	com_predadores	40	
4	com_predadores	32	
5	com_predadores	30	
6	com_predadores	38	
7	sem_predadores	15	
8	sem_predadores	18	
9	sem_predadores	20	
10	sem_predadores	12	
11	sem_predadores	15	
12			

Elementos de inferência

3.3

Problemas de inferência

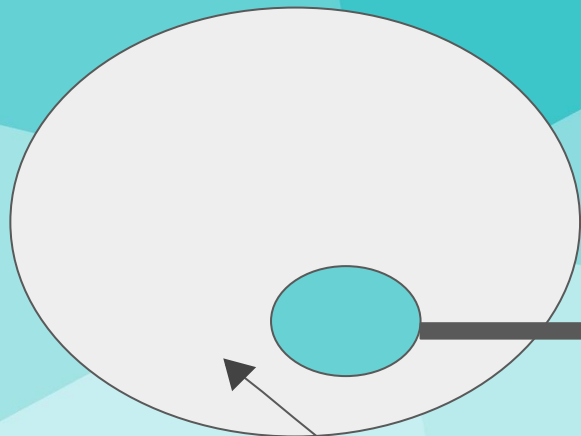
Inferir significa fazer afirmações sobre algo **desconhecido**.

A inferência estatística tem como objetivo fazer afirmações sobre uma característica de uma **população** a partir do conhecimento de dados de uma parte desta população (isto é, uma **amostra** de n observações).

A população é representada por uma distribuição de probabilidade com **parâmetro(s)** cujo(s) valor(es) é (são) **desconhecido(s)**.

Fazemos inferências sobre o(s) parâmetro(s).

População



Amostra

**Inferência
Estatística**

A Inferência estatística, portanto, tem como objetivo estudar generalizações sobre uma população através de evidências fornecidas por uma amostra retirada desta população. A amostra contém os elementos que podem ser observados e é onde as quantidades de interesse podem ser medidas.

Definições

Parâmetro: característica numérica (desconhecida) da distribuição dos elementos da população.

Espaço paramétrico: conjunto de todos os valores possíveis que o parâmetro pode assumir.

Estatística: qualquer função dos elementos de uma amostra aleatória, a qual não depende de parâmetros desconhecidos.

Estimador: Função da amostra, construída com a finalidade de representar, ou estimar um parâmetro de interesse na população.

Estimativa: Valor numérico que um estimador assume.