

# Conceitos Básicos

Prof. Dr. Felipe Dias

Agosto  
2024

# Definição de Estatística

É a ciência que oferece uma coleção de métodos para produzir e obter dados, organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conhecimento.

(Retirado de Santos, A.B. (s.d.), disponível no [link](#))

# Definição de Método Estatístico

Os métodos estatísticos são usados para facilitar o entendimento da variabilidade existente nas variáveis investigadas, a obtenção de evidências fundamentadas em dados; e a explicação do fenômeno.

(Retirado de Santos, A.B. (s.d.), disponível no [link](#))

# Pensemos em algumas situações-problema...

- 1) Você foi contratado pela faculdade para auxiliar em uma pesquisa sobre as notas de disciplinas relacionadas à Matemática. Vocês seriam capazes de pensar em estratégias para apresentar resultados à esta faculdade?
- 2) Agora, mais experiente, você foi contratado pelo IBGE para auxiliar na elaboração de pesquisa censitária sobre a obesidade da população. Como você faria o levantamento destas informações?

# Ainda sobre as situações-problema...

- Percebam que os problemas, aparentemente simples apresentam enorme diferença:
  1. Nos dois casos é possível mapear todas as informações ou todas as pessoas envolvidas?
  2. Afinal, qual é o tema/foco do seu estudos?
  3. Existe algum valor ou medida que represente bem sua variável?

# Análise Estatística

Em geral para efetuar uma análise estatística é necessário um embasamento sobre **chances de ocorrência de uma situação** para:

1. Entendimento do comportamento das variáveis de interesse;
2. Coletar dados, para obter os subsídios empíricos para o estudo;
3. Realizar a análise descritiva, a fim de resumir e descrever como os dados coletados se comportam.

É parte da **inferência estatística**, produzir estimativas e testar hipóteses na busca por respostas para uma questão (Adaptado de Santos, A.B. (s.d.), disponível no [link](#)).

# População e Amostra

**População:** é uma coleção completa de todos os elementos (valores, pessoas, medidas, etc) a serem estudados. Pode ser finita ou infinita.

**Amostra:** é uma sub-coleção de elementos extraídos de uma população.

**Parâmetro:** é uma medida numérica que descreve uma característica de uma população.

**Estatística:** é uma medida numérica que descreve uma característica de uma amostra.

# Estatística Descritiva e Inferencial

**Estatística Descritiva (ou dedutiva):** é a parte da estatística que busca descrever e analisar um certo grupo, sem tirar conclusões sobre um grupo maior (população).

**Estatística Inferencial (ou indutiva):** ramo da estatística que busca tirar conclusões a partir de amostras de uma população



# Tipos de variáveis

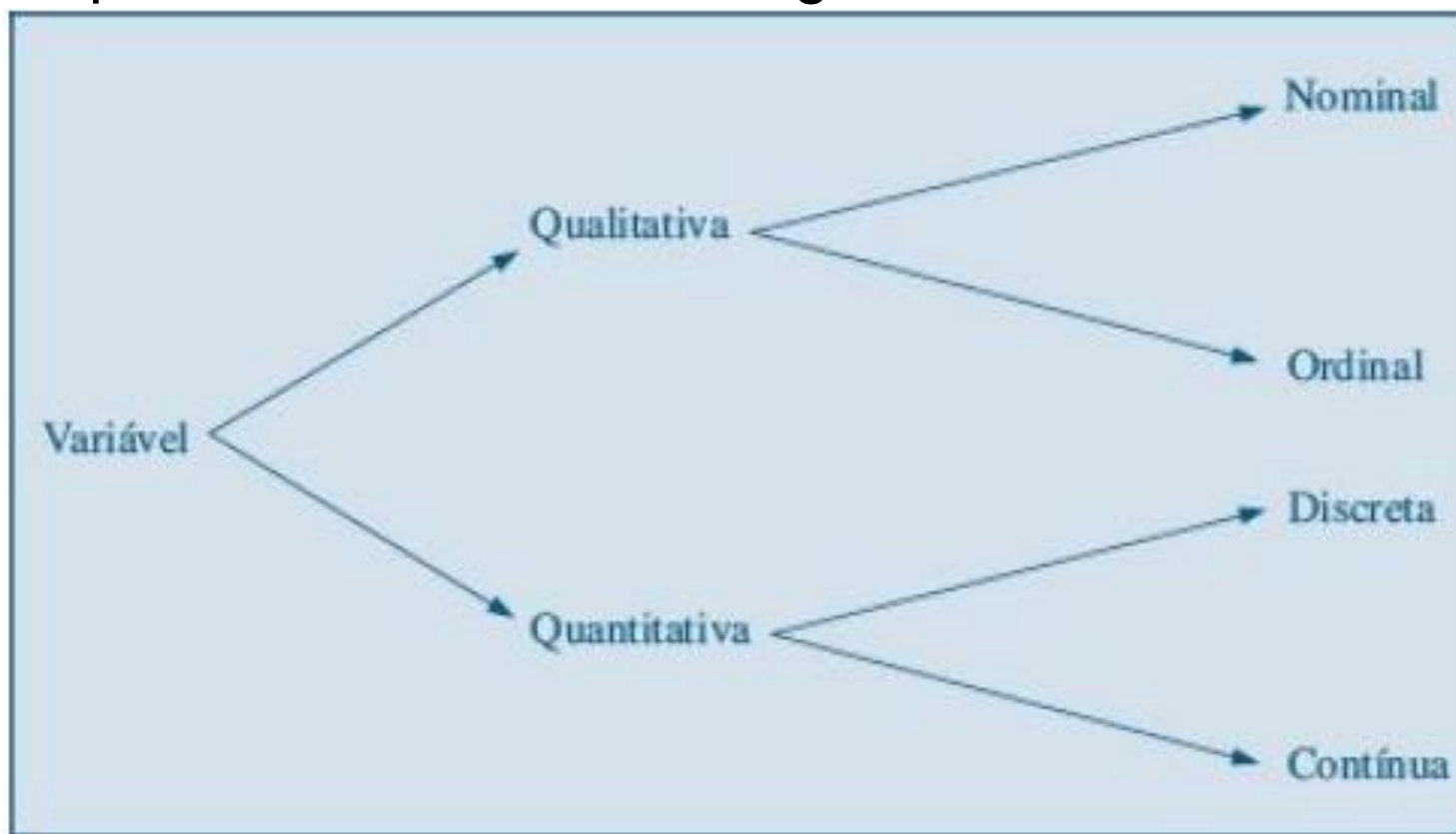
- As variáveis são os objetos do estudo da Estatística;
- São representadas por símbolos ou letras e, podem assumir um conjunto de valores;
- Uma variável que pode assumir qualquer valor entre dois dados é chamada **variável contínua**. Um exemplo o peso de um grupo de estudantes;
- Uma variável que admite apenas valores bem definidos, é chamada de **variável discreta**. Um exemplo o número de filhos de um conjunto de famílias.

# Tipos de variáveis

- É possível ainda classificar as variáveis como **qualitativas** e **quantitativas**:
- **Qualitativas:** apresentam como possíveis realizações uma qualidade (ou atributo) da variável;
- **Quantitativa:** apresentam como possíveis realizações números resultantes de uma contagem ou mensuração.

# Tipos de variáveis

Em resumo, pode-se classificar da seguinte forma:



Fonte: adaptado de Bussab & Morettin (2017)

# Distribuição de frequências

Vejamos o exemplo ao lado.

- É uma tabela com dados dos funcionários de uma empresa:

1. Estado civil;
2. Grau de instrução;
3. Nº de filhos
4. Salários;
5. Idade
6. Região de procedência

N	Estado Civil	Grau de Instrução	N de Filhos	Salario (x Sal Min)	Idade (em anos)	Meses	Região de Procedência
1	solteiro	ensino fundamental		4,00	26	3	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	5	capital
4	solteiro	ensino médio		5,73	20	10	outra
5	solteiro	ensino fundamental		6,26	40	7	outra
6	casado	ensino fundamental	0	6,66	28	0	interior
7	solteiro	ensino fundamental		6,86	41	0	interior
8	solteiro	ensino fundamental		7,39	43	4	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio		7,44	23	6	outra
11	casado	ensino médio	2	8,12	33	6	interior
12	solteiro	ensino fundamental		8,46	27	11	capital
13	solteiro	ensino médio		8,74	37	5	outra
14	casado	ensino fundamental	3	8,95	44	2	outra
15	casado	ensino médio	0	9,13	30	5	interior
16	solteiro	ensino médio		9,35	38	8	outra
17	casado	ensino médio	1	9,77	31	7	capital
18	casado	ensino fundamental	2	9,80	39	7	outra
19	solteiro	superior		10,53	25	8	interior
20	solteiro	ensino médio		10,76	37	4	interior
21	casado	ensino médio	1	11,06	30	9	outra
22	solteiro	ensino médio		11,59	34	2	capital
23	solteiro	ensino fundamental		12,00	41	0	outra
24	casado	superior	0	12,79	26	1	outra
25	casado	ensino médio	2	13,23	32	5	interior
26	casado	ensino médio	2	13,60	35	0	outra
27	solteiro	ensino fundamental		13,85	46	7	outra
28	casado	ensino médio	0	14,69	29	8	interior
29	casado	ensino médio	5	14,71	40	6	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior		16,22	31	5	outra
32	casado	ensino médio	1	16,61	36	4	interior
33	casado	superior	3	17,26	43	7	capital
34	solteiro	superior		18,75	33	7	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	2	interior

# Distribuição de frequências

- Dificuldade em estudar grande quantidade de dados;
- Tabelas de frequências possibilitam o resumo dos dados, de forma a facilitar a visualização;
- Para tanto, as variáveis são agrupadas de acordo com suas ocorrências;
- Embora útil, o processo de construção de distribuição de frequências pode causar perda de informações.

# Distribuição de frequências

## Regra Prática

1. Identificar o maior e o menor valor dos dados brutos;
2. Calcular a amplitude total dos dados;
3. Dividir a amplitude total dos dados para que todas as classes tenham a mesma amplitude de classe;
4. Agrupar os dados dentro das respectivas classes formadas.

# Distribuição de frequências

## Regra Prática

### OBSERVAÇÕES IMPORTANTES

1. O número de intervalos de classe é, geralmente tomado entre 5 a 20, dependendo dos dados;
2. Os intervalos de classe são, preferencialmente escolhidos, de forma que o ponto médio do intervalo coincidam com dados observados;

# Distribuição de frequência

## Voltando ao exemplo...

Retornando a tabela apresentada anteriormente, vamos agrupar os dados de acordo com o grau de instrução:

Grau de instrução	Frequência $n_i$	Proporção $f_i$	Porcentagem $100 f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

**Frequência:** número de observações;

**Proporção ou Frequência Relativa:** número de observações dividido pelo total;



# Distribuição de frequências

## Um problema...

Seria possível comparar essas informações da tabela (grau de instrução) com a tabela abaixo?

Grau de instrução	Frequência $n_i$	Porcentagem $100 f_i$
Fundamental	1.650	32,50
Médio	1.020	51,00
Superior	1.330	16,50
Total	2.000	100,00

# Distribuição de frequências

## Comparação de diferentes amostras

- É possível comparar a mesma variável para diferentes amostras;
- Para tanto não se pode utilizar a “Frequência” sim a **porcentagem!**
- Isso porque a porcentagem reduz as observações à uma mesma base, no caso, base 100...

# Gráficos de Frequências

Para os próximos tópicos, vocês deverão fazer a distribuição de frequências para os dados de salários, com 5 classes;

Pode utilizar Excel ou qualquer outro programa que tenham facilidade. Se preferirem pode fazer esse trabalho à mão...

A distribuição de vocês deve conter a frequência relativa, a frequência acumulada, a porcentagem relativa e a porcentagem acumulada de cada classe.

Qual a média salarial? E qual o valor salarial que mais se repete nesta empresa? O que você pode inferir da distribuição de salários dessa empresa?

# Gráficos de Frequências

O resultado será uma tabela parecida com essa:

Classe de salários	Frequência $n_i$	Frequência acumulada $N_i$	Porcentagem $100 f_i$	Porcentagem acumulada $100 F_i$
4,00 — 8,00	10	10	27,78	27,78
8,00 — 12,00	12	22	33,33	61,11
12,00 — 16,00	8	30	22,22	83,33
16,00 — 20,00	5	35	13,89	97,22
20,00 — 24,00	1	36	2,78	100,00
Total	36	—	100,00	—

# Gráficos de Frequências

O gráfico de porcentagem acumulada fica como o gráfico ao lado

Mas o que ele representa?

