



# LADE

LABORATÓRIO DE ANÁLISES DE DADOS  
EDUCACIONAIS E ESTATÍSTICA APLICADA  
—— IFCE - CAMPUS FORTALEZA ——

## Estatística Descritiva com Python

# Análise de dados : Teoria e Tecnologias

1. Importação de Dados
2. Limpeza e Preparação dos Dados
3. Análise Exploratória de Dados (EDA)
4. Modelagem de Dados
5. Visualização de Resultados

Fonte: <https://www.cursospm3.com.br/blog/python-para-analise-de-dados/>

1. Planejamento
2. Coleta de Dados
3. Crítica dos dados
4. Apuração dos dados
5. Análises de Dados
6. Emissão do Relatório Final
7. Comunicação dos Resultados

Fonte: Costa, Giovani. Estatística Aplicada A Educacao Com Abordagem Além da análise descritiva

1. Olhar para o quadro geral;
2. Obter os dados;
3. Descobrir e visualizar os dados para obter informações;
4. Preparar os dados para os algoritmos do Aprendizado de Máquina;
5. Selecionar e treinar um modelo;
6. Ajustar o seu modelo;
7. Apresentar sua solução;
8. Lançar, monitorar e manter seu sistema.

Fonte :Géron, Aurélien. Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow

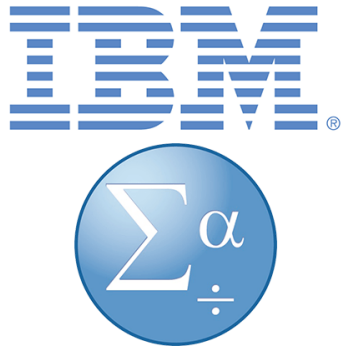


- 1. Defina seus objetivos:** Antes de começar a análise, é crucial entender o que você quer alcançar. Qual é a pergunta ou problema que você está tentando resolver?
- 2. Coleta de dados:** Reúna os dados relevantes para sua análise. Isso pode envolver a obtenção de dados de fontes diferentes, como bancos de dados, arquivos CSV, APIs da web, etc.
- 3. Limpeza de dados:** Os dados geralmente precisam ser limpos e preparados para análise. Isso pode incluir a remoção de dados duplicados, tratamento de valores ausentes, padronização de formatos de dados, entre outras tarefas.
- 4. Exploração de dados:** Nesta etapa, você explora seus dados para entender melhor sua estrutura e identificar padrões, tendências e relações. Isso pode envolver a criação de visualizações de dados, como gráficos e tabelas.
- 5. Análise estatística:** Use técnicas estatísticas para analisar seus dados de maneira mais aprofundada. Isso pode incluir medidas de tendência central, dispersão, correlação, regressão, entre outras.
- 6. Modelagem de dados (se aplicável):** Se você estiver realizando análises preditivas ou inferenciais, pode ser necessário criar modelos estatísticos ou de aprendizado de máquina para fazer previsões ou extrair insights dos dados.
- 7. Interpretação dos resultados:** Analise os resultados de sua análise e interprete o que eles significam em relação aos seus objetivos iniciais. Isso pode envolver a elaboração de conclusões e recomendações com base nos insights obtidos.
- 8. Comunicação dos resultados:** Comunique seus resultados de forma clara e compreensível para as partes interessadas relevantes. Isso pode incluir a criação de relatórios, apresentações ou visualizações de dados para transmitir suas descobertas.
- 9. Iteração e refinamento:** Às vezes, é necessário revisitar etapas anteriores, refinar suas análises ou coletar mais dados para obter uma compreensão mais completa do problema em questão.

(ChatGPT 3.5)



# Análise de dados : Teoria e Tecnologias



**Usaremos :**





Olá, este é o Colaboratory

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda



Índice



Vamos começar



Ciência de dados



Machine learning



Mais recursos

Exemplos em destaque

+ Seção

+ Código

+ Texto


Copiar para o Drive

## Conheça o Colab

### (Novidade) Teste a API Gemini

- [Generate a Gemini API key](#)
- [Talk to Gemini with the Speech-to-Text API](#)
- [Gemini API: Quickstart with Python](#)
- [Gemini API code sample](#)
- [Compare Gemini with ChatGPT](#)
- [More notebooks](#)

# Importação de Dados

 AulaCursoEstatisticaDecritiva.ipynb ☆  
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas /

Índice




+ Seção

{x}



Clica em arquivos

+ Código + Texto  
Comece a progr

 AulaCursoEstatisticaDecritiva.ipynb ☆  
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda [Todas as alterações foram salvas](#)

Arquivos



{x}



..

sample\_data

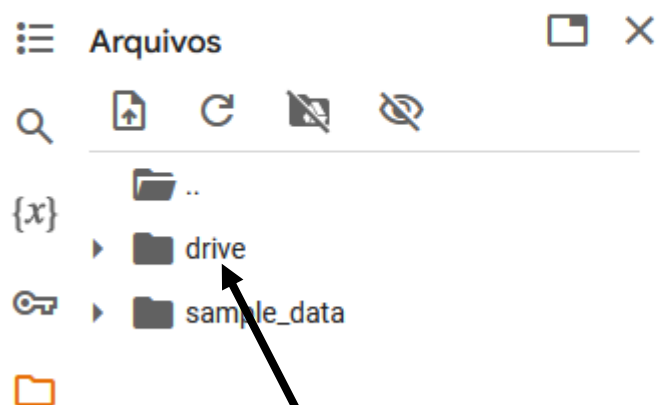


+ Código + Texto

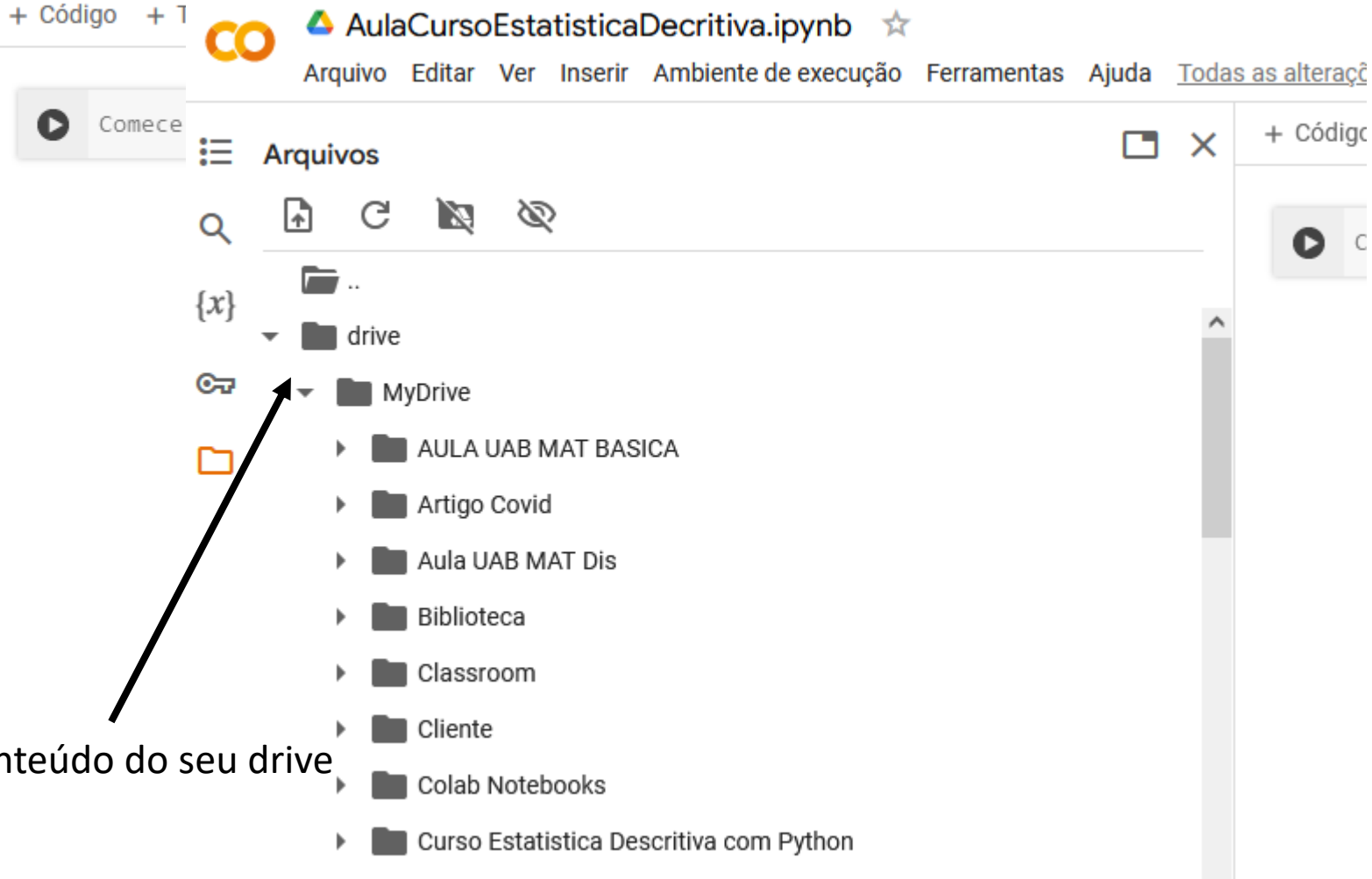


Comece a programar ou [gere código](#) com IA.

Montar seu Drive, para ter acesso aos seus arquivos.



Drive Montado



Acessando ao conteúdo do seu drive



AulaCursoEstatisticaDecritiva.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alteraç

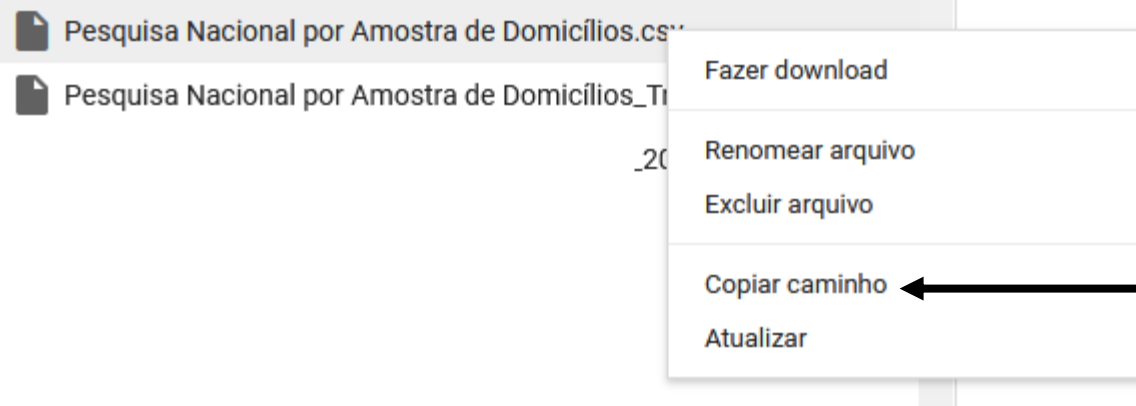
Arquivos



- Curso Estatística Descritiva com Python
- Curso de Introdução a Python
- Curso de programação
- Cursos
- Despesas
- Disciplina de Probabilidade e Estatística
  - PROBABILIDADE E ESTATISTICA
    - Aula02Exercicio\_Pro\_Est\_PesquisaAmostragem.ipynb
    - Chamada dos Alunos 2024\_1\_Probabilidade\_Computação.d...
    - Chamada dos Alunos 2024\_1\_Probabilidade\_Matemática.doc
    - Disciplina de Probabilidade e Estatística.docx
    - Horário das Aulas\_trabalho\_LADE\_estudos\_2024\_1.docx
    - Pesquisa Nacional por Amostra de Domicílios.csv

Procurando arquivo





Copiar o caminho do arquivo

/content/drive/MyDrive/Disciplina de Probabilidade e Estatística/Pesquisa Nacional por Amostra de Domicílios.csv

```
[2] import pandas as pd
import numpy as np
```

```
[ ] # Outra forma de montar seu drive

# from google.colab import drive
# drive.mount('/content/drive')
```

```
[3] PesquisaNacional = pd.read_csv('/content/drive/MyDrive/Disciplina de Probabilidade e Estatística/Pesquisa Nacional por Amostra de Domicílios.csv')
```

▶ PesquisaNacional



	UF	Sexo	Idade	Cor	Anos de Estudo	Renda	Altura
0	11	0	23	8	12	800	1.603808
1	11	1	23	2	12	1150	1.739790
2	11	1	35	8	15	880	1.760444
3	11	0	46	2	6	3500	1.783158
4	11	1	47	8	9	150	1.690631
...	...	...	...	...	...	...	...
76835	53	1	46	2	11	812	1.687030
76836	53	0	30	4	7	1500	1.792934
76837	53	0	32	8	12	1300	1.830587



**Temos a nossa base de dados, o quê fazer?**  
**Só um pouco de calma, vamos ao Pandas e Numpy.**

## Pandas - <https://pandas.pydata.org/>

Uma das bibliotecas mais conhecidas e mais usadas por profissionais de dados, ela permite que o usuário **manipule, transforme e analise dados** de maneira muito otimizada.

A **Pandas** possibilita a leitura em vários formatos, como SQL, CSV, Excel, etc., além de funcionar, principalmente, com dois tipos de estrutura de dados: *Series* e *DataFrames*.

**DataFrames** seguem uma estrutura semelhante a uma planilha de **Excel**.

**Series**, se refere a um **array unidimensional**, que pode ser entendido como uma lista simples de valores.

## NumPy - <https://numpy.org/>

A biblioteca **Numpy compila funções** relacionadas à álgebra linear e computação numérica, trabalhando com *arrays* multidimensionais, cálculos rápidos, entre outras funcionalidades.

Além disso, a biblioteca **NumPy está no núcleo de basicamente todos** os programas e bibliotecas que lidam com operações matemáticas e usam a linguagem de programação Python.

Como por exemplo, a própria biblioteca Pandas baseia sua estrutura de dados (*DataFrames* e *Series*) em *arrays* de NumPy.

Fonte: <https://www.cursospm3.com.br/blog/python-para-analise-de-dados/>



```
[2] import pandas as pd
import numpy as np
```

```
[ ] # Outra forma de montar seu drive

# from google.colab import drive
# drive.mount('/content/drive')
```

```
[3] PesquisaNacional = pd.read_csv('/content/drive/MyDrive/Disciplina de Probabilidade e Estatística/Pesquisa Nacional por Amostra de Domicílios.csv')
```

▶ PesquisaNacional



	UF	Sexo	Idade	Cor	Anos de Estudo	Renda	Altura
0	11	0	23	8	12	800	1.603808
1	11	1	23	2	12	1150	1.739790
2	11	1	35	8	15	880	1.760444
3	11	0	46	2	6	3500	1.783158
4	11	1	47	8	9	150	1.690631
...	...	...	...	...	...	...	...
76835	53	1	46	2	11	812	1.687030
76836	53	0	30	4	7	1500	1.792934
76837	53	0	32	8	12	1300	1.830587

**Vamos conhecer no base de dados, para isso usaremos**

`PesquisaNacional.info()`



PesquisaNacional.info()



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 76840 entries, 0 to 76839
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   UF                    76840 non-null  int64
1   Sexo                  76840 non-null  int64
2   Idade                 76840 non-null  int64
3   Cor                   76840 non-null  int64
4   Anos de Estudo        76840 non-null  int64
5   Renda                 76840 non-null  int64
6   Altura                76840 non-null  float64
dtypes: float64(1), int64(6)
memory usage: 4.1 MB
```

## Tipos de Variáveis

### Qualitativas

#### Nominal

- Profissão
- Sexo
- Religião

#### Ordinal

- Escolaridade
- Estágio da doença
- Classe social

### Quantitativas

#### Discreta

- Nº de filhos
- Nº de acessos à plataforma

#### Contínua

- Altura
- Peso
- Salário

# Iremos a documentação/dicionário da nossa base de dados

## Pesquisa Nacional por Amostra de Domicílios - PNAD

Dicionário – Códigos.

### Variáveis utilizadas

Renda

Documento na página do curso/Google sala de aula

Rendimento mensal do trabalho principal para pessoas de 10 anos ou mais de idade.

Idade

Idade do morador na data de referência em anos.

Altura (elaboração própria)

Altura do morador em metros.



# Trocaremos os códigos pelos significados

`PesquisaNacional['Sexo'].replace({0: 'Masculino', 1: 'Feminino'}, inplace=True)`

```
[ ] PesquisaNacional['Sexo'].replace({0: 'Masculino', 1: 'Feminino'}, inplace=True)
```



PesquisaNacional



	UF	Sexo	Idade	Cor	Anos de Estudo	Renda	Altura
0	11	Masculino	23	8	12	800	1.603808
1	11	Feminino	23	2	12	1150	1.739790
2	11	Feminino	35	8	15	880	1.760444
3	11	Masculino	46	2	6	3500	1.783158
4	11	Feminino	47	8	9	150	1.690631
...	...	...	...	...	...	...	...
76835	53	Feminino	46	2	11	812	1.687030
76836	53	Masculino	30	4	7	1500	1.792934
76837	53	Masculino	32	8	12	1300	1.830587
76838	53	Masculino	57	8	4	1500	1.726344
76839	53	Masculino	38	8	4	900	1.658305

76840 rows × 7 columns

## **Exercício :**

**Rotular as variáveis qualitativas , usando o dicionário, do arquivo :**

**Pesquisa Nacional por Amostra de Domicílios.csv**

## **Exercício para Casa :**

**Rotular as variáveis qualitativas , usando o dicionário, do arquivo :**

**Enem\_Fortaleza\_2021\_DisciplinaProbabilidadeEstatística.xlsx**

**Agora temos um arquivo xlsx, como abri-lo usando o pandas?**



# Aula 03

# Distribuição de Frequência

Quando se estuda uma variável, o maior interesse do pesquisador é conhecer o comportamento dessa variável, analisando a ocorrência de suas possíveis realizações. Nesta seção veremos uma maneira de se dispor um conjunto de realizações, para se ter uma ideia global sobre elas, ou seja, de sua distribuição.

Exemplo:

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (× sal. mín.)	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	—	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	—	5,73	20	10	outra
5	solteiro	ensino fundamental	—	6,26	40	07	outra
6	casado	ensino fundamental	0	6,66	28	00	interior
7	solteiro	ensino fundamental	—	6,86	41	00	interior
8	solteiro	ensino fundamental	—	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio	—	7,44	23	06	outra
11	casado	ensino médio	2	8,12	33	06	interior
12	solteiro	ensino fundamental	—	8,46	27	11	capital
13	solteiro	ensino médio	—	8,74	37	05	outra
14	casado	ensino fundamental	3	8,95	44	02	outra
15	casado	ensino médio	0	9,13	30	05	interior
16	solteiro	ensino médio	—	9,35	38	08	outra
17	casado	ensino médio	1	9,77	31	07	capital
18	casado	ensino fundamental	2	9,80	39	07	outra
19	solteiro	superior	—	10,53	25	08	interior
20	solteiro	ensino médio	—	10,76	37	04	interior
21	casado	ensino médio	1	11,06	30	09	outra
22	solteiro	ensino médio	—	11,59	34	02	capital
23	solteiro	ensino fundamental	—	12,00	41	00	outra
24	casado	superior	0	12,79	26	01	outra
25	casado	ensino médio	2	13,23	32	05	interior
26	casado	ensino médio	2	13,60	35	00	outra
27	solteiro	ensino fundamental	—	13,85	46	07	outra
28	casado	ensino médio	0	14,69	29	08	interior
29	casado	ensino médio	5	14,71	40	06	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior	—	16,22	31	05	outra
32	casado	ensino médio	1	16,61	36	04	interior
33	casado	superior	3	17,26	43	07	capital
34	solteiro	superior	—	18,75	33	07	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	02	interior

## Variáveis Categóricas / Qualitativas

```
▶ Frequencia= CompanhiaMB['Estado Civil'].value_counts()  
Frequencia
```

```
casado      20  
solteiro    16  
Name: Estado Civil, dtype: int64
```

```
▶ Frequencia= CompanhiaMB['Grau de Instrução'].value_counts()  
Frequencia
```

```
ensino médio      18  
ensino fundamental 12  
superior           6  
Name: Grau de Instrução, dtype: int64
```

# Distribuição de Frequência

Quando se estuda uma variável, o maior interesse do pesquisador é conhecer o comportamento dessa variável, analisando a ocorrência de suas possíveis realizações. Nesta seção veremos uma maneira de se dispor um conjunto de realizações, para se ter uma ideia global sobre elas, ou seja, de sua distribuição.

Exemplo:

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (× sal. mín.)	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	—	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	—	5,73	20	10	outra
5	solteiro	ensino fundamental	—	6,26	40	07	outra
6	casado	ensino fundamental	0	6,66	28	00	interior
7	solteiro	ensino fundamental	—	6,86	41	00	interior
8	solteiro	ensino fundamental	—	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio	—	7,44	23	06	outra
11	casado	ensino médio	2	8,12	33	06	interior
12	solteiro	ensino fundamental	—	8,46	27	11	capital
13	solteiro	ensino médio	—	8,74	37	05	outra
14	casado	ensino fundamental	3	8,95	44	02	outra
15	casado	ensino médio	0	9,13	30	05	interior
16	solteiro	ensino médio	—	9,35	38	08	outra
17	casado	ensino médio	1	9,77	31	07	capital
18	casado	ensino fundamental	2	9,80	39	07	outra
19	solteiro	superior	—	10,53	25	08	interior
20	solteiro	ensino médio	—	10,76	37	04	interior
21	casado	ensino médio	1	11,06	30	09	outra
22	solteiro	ensino médio	—	11,59	34	02	capital
23	solteiro	ensino fundamental	—	12,00	41	00	outra
24	casado	superior	0	12,79	26	01	outra
25	casado	ensino médio	2	13,23	32	05	interior
26	casado	ensino médio	2	13,60	35	00	outra
27	solteiro	ensino fundamental	—	13,85	46	07	outra
28	casado	ensino médio	0	14,69	29	08	interior
29	casado	ensino médio	5	14,71	40	06	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior	—	16,22	31	05	outra
32	casado	ensino médio	1	16,61	36	04	interior
33	casado	superior	3	17,26	43	07	capital
34	solteiro	superior	—	18,75	33	07	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	02	interior

## Variáveis Categóricas / Qualitativas

### #Frequencia Relativa do Estado Civil

```
FrequenciaCivil_Re=FrequenciaCivil / len(CompanhiaMB['Estado Civil'])  
FrequenciaCivil_R=FrequenciaCivil_Re*100  
FrequenciaCivil_R
```

```
casado      55.555556  
solteiro    44.444444  
Name: Estado Civil, dtype: float64
```



### #Frequencia Relativa do Grau de Instrução

```
FrequenciaIns_Re=FrequenciaIns / len(CompanhiaMB['Grau de Instrução'])  
FrequenciaIns_R=FrequenciaIns_Re*100  
FrequenciaIns_R
```



```
ensino médio      50.000000  
ensino fundamental 33.333333  
superior          16.666667  
Name: Grau de Instrução, dtype: float64
```

# Distribuição de Frequência

## Amplitude Total

A amplitude de um conjunto, em Estatística, é a diferença entre o maior elemento desse conjunto e o menor.

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (× sal. mín.)	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	—	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	—	5,73	20	10	outra
5	solteiro	ensino fundamental	—	6,26	40	07	outra
6	casado	ensino fundamental	0	6,66	28	00	interior
7	solteiro	ensino fundamental	—	6,86	41	00	interior
8	solteiro	ensino fundamental	—	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio	—	7,44	23	06	outra
11	casado	ensino médio	2	8,12	33	06	interior
12	solteiro	ensino fundamental	—	8,46	27	11	capital
13	solteiro	ensino médio	—	8,74	37	05	outra
14	casado	ensino fundamental	3	8,95	44	02	outra
15	casado	ensino médio	0	9,13	30	05	interior
16	solteiro	ensino médio	—	9,35	38	08	outra
17	casado	ensino médio	1	9,77	31	07	capital
18	casado	ensino fundamental	2	9,80	39	07	outra
19	solteiro	superior	—	10,53	25	08	interior
20	solteiro	ensino médio	—	10,76	37	04	interior
21	casado	ensino médio	1	11,06	30	09	outra
22	solteiro	ensino médio	—	11,59	34	02	capital
23	solteiro	ensino fundamental	—	12,00	41	00	outra
24	casado	superior	0	12,79	26	01	outra
25	casado	ensino médio	2	13,23	32	05	interior
26	casado	ensino médio	2	13,60	35	00	outra
27	solteiro	ensino fundamental	—	13,85	46	07	outra
28	casado	ensino médio	0	14,69	29	08	interior
29	casado	ensino médio	5	14,71	40	06	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior	—	16,22	31	05	outra
32	casado	ensino médio	1	16,61	36	04	interior
33	casado	superior	3	17,26	43	07	capital
34	solteiro	superior	—	18,75	33	07	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	02	interior

### Variáveis Quantitativas

Para a variável **salário** temos uma amplitude de:

Amplitude total = 23,30 – 4,00 = 19,30

Determine a amplitude total para variável Idade em anos.



# Classe

Classe de frequência, ou simplesmente, classe, é cada um dos grupos de valores em que se subdivide a amplitude total do conjunto de valores observados da variável.

Uma determinada classe pode ser identificada por seus extremos.

Classe de salários	Frequência $n_i$	Porcentagem $100 f_i$
4,00 ┤ 8,00	10	27,78
8,00 ┤ 12,00	12	33,33
12,00 ┤ 16,00	8	22,22
16,00 ┤ 20,00	5	13,89
20,00 ┤ 24,00	1	2,78
Total	36	100,00

(Fórmula de Sturges) Regra para determinar o número de classes :

K – Número de classes

n- número total de observações

$$K = 1 + 3.3 * \log_{10} n$$

Regra para determinar o número de classes :

K – Número de classes

n- número total de observações

$$K = 1 + 3.3 * \log_{10} n$$

Exemplos:

a) Se o número de observações for 500, determine o número de classes.

b) Se o número de observações for 50, determine o número de classes.

Truman L. Kelley, em *The Grouping Data for Graphic Portrayal*, sugere os seguintes números de classes, com base no número total de observações, para efeito de representação gráfica:

<i>N</i>	5	10	25	50	100	200	500	1 000
<i>k</i>	2	4	6	8	10	12	15	15

### Nota

Existem outras fórmulas empíricas que pretendem determinar o número de classes que deve ter a distribuição, como por exemplo:  $i \approx \sqrt{h}$  ( $h$  = amplitude do intervalo de classe). Mesmo conhecendo estes critérios de determinação do número de classes, o analista deverá ter em mente que a escolha dependerá antes da natureza dos dados e da unidade de medida em que eles forem expressos do que de regras muitas vezes arbitrárias e pouco flexíveis.

## Limites de Classes

Os limites de classe são seus valores extremos

Classe de salários	Frequência $n_i$	Porcentagem $100 f_i$
4,00 ┤ 8,00	10	27,78
8,00 ┤ 12,00	12	33,33
12,00 ┤ 16,00	8	22,22
16,00 ┤ 20,00	5	13,89
20,00 ┤ 24,00	1	2,78
Total	36	100,00

Limite Inferior

Limite Superior

4,00 ┤ 8,00



Fechado no 4,00 e aberto no 8,00



# Amplitude do Intervalo de Classe

A amplitude do intervalo de classe ou simplesmente intervalo de classe é o comprimento da classe, sendo geralmente definido como a diferença entre limites superior e inferior

Amplitude do intervalo de classe:

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (× sal. mín.)	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	—	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	—	5,73	20	10	outra
5	solteiro	ensino fundamental	—	6,26	40	07	outra
6	casado	ensino fundamental	0	6,66	28	00	interior
7	solteiro	ensino fundamental	—	6,86	41	00	interior
8	solteiro	ensino fundamental	—	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio	—	7,44	23	06	outra
11	casado	ensino médio	2	8,12	33	06	interior
12	solteiro	ensino fundamental	—	8,46	27	11	capital
13	solteiro	ensino médio	—	8,74	37	05	outra
14	casado	ensino fundamental	3	8,95	44	02	outra
15	casado	ensino médio	0	9,13	30	05	interior
16	solteiro	ensino médio	—	9,35	38	08	outra
17	casado	ensino médio	1	9,77	31	07	capital
18	casado	ensino fundamental	2	9,80	39	07	outra
19	solteiro	superior	—	10,53	25	08	interior
20	solteiro	ensino médio	—	10,76	37	04	interior
21	casado	ensino médio	1	11,06	30	09	outra
22	solteiro	ensino médio	—	11,59	34	02	capital
23	solteiro	ensino fundamental	—	12,00	41	00	outra
24	casado	superior	0	12,79	26	01	outra
25	casado	ensino médio	2	13,23	32	05	interior
26	casado	ensino médio	2	13,60	35	00	outra
27	solteiro	ensino fundamental	—	13,85	46	07	outra
28	casado	ensino médio	0	14,69	29	08	interior
29	casado	ensino médio	5	14,71	40	06	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior	—	16,22	31	05	outra
32	casado	ensino médio	1	16,61	36	04	interior
33	casado	superior	3	17,26	43	07	capital
34	solteiro	superior	—	18,75	33	07	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	02	interior

Fonte: Dados hipotéticos.

*Amplitude total*

*k*

## Exercício 01

Usando a Fórmula de Sturges determine:

- Amplitude total da variável idade em anos.
- Número de classes para as variável idade em anos.
- Amplitude de cada classe.
- Determine a frequência absoluta e relativa.
- Determine a frequência absoluta acumulada e relativa acumulada.
- Determine os Pontos médios das classes.



## Exercício em sala

### Exercício 02

Os dados seguintes representam 20 observações relativas ao índice pluviométrico em determinados municípios do Estado:

<i>Milímetros de chuva</i>			
144	152	159	160
160	151	157	146
154	145	141	150
142	146	142	141
141	150	143	158

Determinar o número de classes pela regra de Sturges.

Construir a tabela de frequências absolutas simples.

Determinar as frequências absolutas acumuladas (“abaixo de”).

Determinar as frequências simples relativas.

Determinar as frequências relativas acumuladas (“abaixo de”).

**Exercício 03 - Os dados seguintes representam as idades de um conjunto de pessoas.**

35 36 59 18 41 9 17 18 55 59 13 26 59 6 13 28 28 30 19 24 3 55 43 35 41  
20 19 27 40 30

Determine:

- a) Amplitude total da variável idade.
- b) Número de classes para a variável idade.
- c) Amplitude de cada classe.
- d) Determine a frequência absoluta e relativa.
- e) Determine a frequência absoluta acumulada e relativa acumulada.
- f) Determine os Pontos médios das classes.

**Exercício 04 - Os dados seguintes representam os pesos, em Kg, de um conjunto de pessoas.**

80 32 100 91 52 6 72 53 40 94 94 61 42 68 13 72 56 49 69 53 1 51 75 89 33

Determine:

- a) Amplitude total da variável peso.
- b) Número de classes para a variável peso.
- c) Amplitude de cada classe.
- d) Determine a frequência absoluta e relativa.
- e) Determine a frequência absoluta acumulada e relativa acumulada.
- f) Determine os Pontos médios das classes.

# Tratando os dados no Python

```
[3] CompanhiaMB=pd.read_excel("/content/drive/MyDrive/Curso Estatística Descritiva com Python/Companhia_MB.xlsx")
```



CompanhiaMB

	N	Estado Civil	Grau de Instrução	N de Filhos	Salario (x Sal Min)	Anos	Meses	Região de Procedência
0	1	solteiro	ensino fundamental	NaN	4.00	26	3	interior
1	2	casado	ensino fundamental	1.0	4.56	32	10	capital
2	3	casado	ensino fundamental	2.0	5.25	36	5	capital
3	4	solteiro	ensino médio	NaN	5.73	20	10	outra
4	5	solteiro	ensino fundamental	NaN	6.26	40	7	outra
5	6	casado	ensino fundamental	0.0	6.66	28	0	interior
6	7	solteiro	ensino fundamental	NaN	6.86	41	0	interior
-	-	-	-	-	-	-	-	-



## CompanhiaMB.info()

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 36 entries, 0 to 35  
Data columns (total 8 columns):  
#   Column                Non-Null Count  Dtype  
---  -  
0   N                      36 non-null    int64  
1   Estado Civil          36 non-null    object  
2   Grau de Instrução     36 non-null    object  
3   N de Filhos           20 non-null    float64  
4   Salario (x Sal Min)   36 non-null    float64  
5   Anos                  36 non-null    int64  
6   Meses                 36 non-null    int64  
7   Região de Procedência 36 non-null    object  
dtypes: float64(2), int64(3), object(3)  
memory usage: 2.4+ KB
```

```
[13] FrequenciaCivil= CompanhiaMB['Estado Civil'].value_counts()  
FrequenciaCivil
```

```
casado      20  
solteiro    16  
Name: Estado Civil, dtype: int64
```

```
▶ FrequenciaIns= CompanhiaMB['Grau de Instrução'].value_counts()  
FrequenciaIns
```

```
➡ ensino médio      18  
ensino fundamental  12  
superior            6  
Name: Grau de Instrução, dtype: int64
```

```
[16] #Frequencia Relativa do Estado Civil
FrequenciaCivil_Re=FrequenciaCivil / len(CompanhiaMB['Estado Civil'])
FrequenciaCivil_R=FrequenciaCivil_Re*100
FrequenciaCivil_R
```

```
casado      55.555556
solteiro    44.444444
Name: Estado Civil, dtype: float64
```

```
[18] #Frequencia Relativa do Grau de Instrução
FrequenciaIns_Re=FrequenciaIns / len(CompanhiaMB['Grau de Instrução'])
FrequenciaIns_R=FrequenciaIns_Re*100
FrequenciaIns_R
```

```
ensino médio      50.000000
ensino fundamental 33.333333
superior           16.666667
Name: Grau de Instrução, dtype: float64
```

```
[21] # Amplitude Total
      # Valor Máximo = CompanhiaMB['Salario (x Sal Min)'].max()
      # Valor Mínimo = CompanhiaMB['Salario (x Sal Min)'].min()
      Amplitude=CompanhiaMB['Salario (x Sal Min)'].max()-CompanhiaMB['Salario (x Sal Min)'].min()
      Amplitude
```

19.3

```
[23] # Número de Classes
      import math
      n=36
      k=1+3.3*math.log10(n)
      k
```

6.135798252532048

```
[25] # Amplitude de cada Classe
```

```
k=6
```

```
AmplitudeClasse=Amplitude/k
```

```
AmplitudeClasse
```

```
3.216666666666667
```

```
[31] # Colocaremos 3.22 e teremos 6 Classes
```

```
ValorMáximo = CompanhiaMB['Salario (x Sal Min)'].max()
```

```
ValorMínimo = CompanhiaMB['Salario (x Sal Min)'].min()
```

```
print(ValorMínimo)
```

```
print(ValorMáximo)
```

```
4.0
```

```
23.3
```





```
# Colocaremos 3.22 e teremos 6 Classes
```

```
ValorMáximo = CompanhiaMB['Salario (x Sal Min)'].max()
```

```
ValorMínimo = CompanhiaMB['Salario (x Sal Min)'].min()
```

```
print(ValorMínimo)
```

```
print(ValorMáximo)
```

4.0

23.3

```
[39] # Construindo as Classes
```

```
# 4 - 7,22
```

```
# 7,22 - 10,44
```

```
# 10,44 - 13,66
```

```
# 13,66 - 16,88
```

```
# 16,88 - 20,1
```

```
# 20,1 - 23,32
```

```
bins = [4,7.22,10.44,13.66,16.88,20.1,23.32]
```

# Aqui, ao passar right=False para a função pd.cut(), você está especificando que os intervalos são fechados à esquerda e abertos à direita. Isso garante que os intervalos sejam conforme você deseja, divididos em seis classes.

```
CompanhiaMB['Intervalo de Classe'] = pd.cut(CompanhiaMB['Salario (x Sal Min)'],bins,right=False)
```

```
CompanhiaMB
```

	N	Estado Civil	Grau de Instrução	N de Filhos	Salario (x Sal Min)	Anos	Meses	Região de Procedência	Intervalo de Classe
0	1	solteiro	ensino fundamental	NaN	4.00	26	3	interior	[4.0, 7.22)
1	2	casado	ensino fundamental	1.0	4.56	32	10	capital	[4.0, 7.22)
2	3	casado	ensino fundamental	2.0	5.25	36	5	capital	[4.0, 7.22)
3	4	solteiro	ensino médio	NaN	5.73	20	10	outra	[4.0, 7.22)
4	5	solteiro	ensino fundamental	NaN	6.26	40	7	outra	[4.0, 7.22)
5	6	casado	ensino fundamental	0.0	6.66	28	0	interior	[4.0, 7.22)

```
[41] CompanhiaMB['Intervalo de Classe'].value_counts()
```

```
[7.22, 10.44)    11
```

```
[10.44, 13.66)    8
```

```
[4.0, 7.22)       7
```

```
[13.66, 16.88)    6
```

```
[16.88, 20.1)     3
```

```
[20.1, 23.32)     1
```

```
Name: Intervalo de Classe, dtype: int64
```

## Exercício :

Na Base de dados Pesquisa Nacional por Amostra de Domicílios.csv faça:

- a) Tabela de frequência para variáveis qualitativas
- b) Categorizar as variáveis quantitativas e criar uma nova variáveis com essas categorias.
- c) Tabela de frequência para as novas variáveis categóricas

## Exercício para Casa :

Na Base de dados Enem\_Fortaleza\_2021\_DisciplinaProbabilidadeEstatística.xlsx faça:

(OBS.: Use sua Imaginação/Criatividade e o Bom Senso )

- a) Tabela de frequência para variáveis qualitativas
- b) Categorizar as variáveis quantitativas e criar uma nova variáveis com essas categorias.
- c) Tabela de frequência para as novas variáveis categóricas







**LADE**

LABORATÓRIO DE ANÁLISES DE DADOS  
EDUCACIONAIS E ESTATÍSTICA APLICADA  
— IFCE - CAMPUS FORTALEZA —

**OBRIGADO!!!**