

# PE22NB\_T1\_DIOVANI\_COELLI

April 6, 2022

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ CAMPUS PATO BRANCO.  
PROF<sup>a</sup>. DR<sup>a</sup>. CLEONIS VIATER FIGUEIRA. ENGENHARIA DA COMPUTAÇÃO - 2  
PERÍODO. PE22NB-2CP. ALUNO: DIOVANI EDUARDO COELLI.

Para a realização desse trabalho eu utilizei a linguagem de programação [Python](#) em conjunto com [Jupyter Notebooks](#) para gerar as visualizações e esse arquivo pdf, as bibliotecas usadas para a leitura e transformação das informações foram [pandas](#) e [numpy](#). A interpretação das informações segue a ordem, primeiro vem o código com comentários explicando o funcionamento e em seguida o resultado gerado por aquele código.

```
[ ]: # Abre arquivo data.txt
f = open('data/data.txt', 'r')

f_string = f.read()

f.close()

# Limpa e transforma o arquivo em csv (Comma-separated values) para facilitar
↳ o uso
f_string = f_string.replace('\r', '\n')
f_string = f_string.replace('\t', '')
f_string = f_string.replace(' ', ',')

# Escreve o novo arquivo como data.csv
novo_arquivo = open('data/data.csv', 'x')

novo_arquivo.write(f_string)

novo_arquivo.close()
```

```
[ ]: import pandas as pd

pd.options.display.float_format = '{:,.2f}'.format

# Lê a tabela limpa
data = pd.read_csv('data/data.csv')

# estado civil: solteiro 1    casado 2
```

```

data['civil'] = data['civil'].replace(1, 'solteiro').replace(2, 'casado')

# nivel de instrução: ensino fundamental 1, ensino médio 2, ensino superior 3
data['instrucao'] = data['instrucao'].replace(1, 'ensino fundamental').
↳replace(2, 'ensino médio').replace(3, 'ensino superior')

# origem de procedencia: interior capital outra
data['regiao'] = data['regiao'].replace(1, 'interior').replace(2, 'capital').
↳replace(3, 'outra')

# Imprime os dados
data

```

```

[ ]:      funcionario      civil      instrucao      filhos      salario      ano      mes      \
0          1      solteiro      ensino fundamental      NaN      4.00      26      3
1          2      casado      ensino fundamental      1.00      4.56      32      10
2          3      casado      ensino fundamental      2.00      5.25      36      5
3          4      solteiro      ensino médio      NaN      5.73      20      10
4          5      solteiro      ensino fundamental      NaN      6.26      40      7
5          6      casado      ensino fundamental      0.00      6.66      28      0
6          7      solteiro      ensino fundamental      NaN      6.86      41      0
7          8      solteiro      ensino fundamental      NaN      7.39      43      4
8          9      casado      ensino médio      1.00      7.59      34      10
9         10      solteiro      ensino médio      NaN      7.44      23      6
10         11      casado      ensino médio      2.00      8.12      33      6
11         12      solteiro      ensino fundamental      NaN      8.46      27      11
12         13      solteiro      ensino médio      NaN      8.74      37      5
13         14      casado      ensino fundamental      3.00      8.95      44      2
14         15      casado      ensino médio      0.00      9.13      30      5
15         16      solteiro      ensino médio      NaN      9.35      38      8
16         17      casado      ensino médio      1.00      9.77      31      7
17         18      casado      ensino fundamental      2.00      9.80      39      7
18         19      solteiro      ensino superior      NaN      10.53      25      8
19         20      solteiro      ensino médio      NaN      10.76      37      4
20         21      casado      ensino médio      1.00      11.06      30      9
21         22      solteiro      ensino médio      NaN      11.59      34      2
22         23      solteiro      ensino fundamental      NaN      12.00      41      0
23         24      casado      ensino superior      0.00      12.79      26      1
24         25      casado      ensino médio      2.00      13.23      32      5
25         26      casado      ensino médio      2.00      13.60      35      0
26         27      solteiro      ensino fundamental      NaN      13.85      46      7
27         28      casado      ensino médio      0.00      14.69      29      8
28         29      casado      ensino médio      5.00      14.71      40      6
29         30      casado      ensino médio      2.00      15.99      35      10
30         31      solteiro      ensino superior      NaN      16.22      31      5
31         32      casado      ensino médio      1.00      16.61      36      4
32         33      casado      ensino superior      3.00      17.26      43      7

```

33	34	solteiro	ensino superior	NaN	18.75	33	7
34	35	casado	ensino médio	2.00	19.40	48	11
35	36	casado	ensino superior	3.00	23.30	42	2

	regiao
0	interior
1	capital
2	capital
3	outra
4	outra
5	interior
6	interior
7	capital
8	capital
9	outra
10	interior
11	capital
12	outra
13	outra
14	interior
15	outra
16	capital
17	outra
18	interior
19	interior
20	outra
21	capital
22	outra
23	outra
24	interior
25	outra
26	outra
27	interior
28	interior
29	capital
30	outra
31	interior
32	capital
33	capital
34	capital
35	interior

**0.0.1 PARTE 1.** faça uma descrição, usando as medidas descritivas (media, mediana, moda, amplitude, variância e desvio-padrão, quando forem adequadas) com respeito a idade e nível de instrução (apenas estas duas)

```
[ ]: import numpy as np

# Cria tabela descritiva da idade
idade_frame = pd.DataFrame(data['ano'].describe()).rename(columns={'ano': 'idade'})

idade_frame = idade_frame.transpose()

# Gera medidas descritivas adicionais
idade_frame['amplitude'] = np.subtract(idade_frame['max'], idade_frame['min'])
idade_frame['variancia'] = data['ano'].var()

# Traduz colunas para o portugues
idade_frame = idade_frame.rename(columns={'mean': 'média', 'std': 'desvio-padrão', '50%': 'mediana', 'min': 'menor', 'max': 'maior' })

# Remove colunas desnecessarias
idade_frame = idade_frame.drop('25%', axis=1)
idade_frame = idade_frame.drop('75%', axis=1)
idade_frame = idade_frame.drop('count', axis=1)

print('Descrição idade')
idade_frame.transpose()
```

Descrição idade

```
[ ]:      idade
média      34.58
desvio-padrão  6.74
menor       20.00
mediana     34.50
maior       48.00
amplitude    28.00
variancia    45.39
```

```
[ ]: # Cria tabela descritiva de instrucao
instrucao_frame = pd.DataFrame(data['instrucao'].describe()).transpose()

# Traduz cabecalhos
instrucao_frame = instrucao_frame.rename(columns={'count': 'frequencia absoluta total', 'top': 'mais frequente', 'freq': 'frequencia absoluta mais frequente'})

# Remove colunas desnecessarias
```

```
instrucao_frame = instrucao_frame.drop('unique', axis=1)

print('Descrição grau de instrução')
instrucao_frame.transpose()
```

Descrição grau de instrução

```
[ ]:          instrucao
frequencia absoluta total          36
mais frequente          ensino médio
frequencia absoluta mais frequente    18
```

## 0.0.2 PARTE 2. organize uma tabela de distribuição de frequências para a idade e para o nível de instrução, com as frequências absolutas.

```
[ ]: # Gera a frequencia absoluta
instrucao_absoluta = data['instrucao'].value_counts()

# Gera os cabecalhos
instrucao_absoluta_frame = pd.DataFrame({'grau de instrucao':_,
    ↳instrucao_absoluta.index, 'frequencia absoluta': instrucao_absoluta.values})

# Gera o total
instrucao_absoluta_frame.loc['total'] = instrucao_absoluta_frame.
    ↳sum(numeric_only=True)

# Imprime a tabela
print('Tabela de frequencia absoluta grau de instrução')
instrucao_absoluta_frame
```

Tabela de frequencia absoluta instrução

```
[ ]:      grau de instrucao  frequencia absoluta
0          ensino médio          18.00
1      ensino fundamental          12.00
2          ensino superior           6.00
total                NaN          36.00
```

```
[ ]: # Gera a frequencia absoluta
idade_absoluta = data['ano'].value_counts()

# Gera os cabecalhos
idade_absoluta_frame = pd.DataFrame({'idade': idade_absoluta.index, 'frequencia_
    ↳absoluta': idade_absoluta.values})

# Gera o total
idade_absoluta_frame.loc['total', 'frequencia absoluta'] =_
    ↳idade_absoluta_frame['frequencia absoluta'].sum()
```

```
# Imprime a tabela
print('Tabela de frequencia absoluta idade')
idade_absoluta_frame
```

Tabela de frequencia absoluta idade

```
[ ]:      idade  frequencia absoluta
0      26.00             2.00
1      34.00             2.00
2      35.00             2.00
3      31.00             2.00
4      30.00             2.00
5      32.00             2.00
6      33.00             2.00
7      37.00             2.00
8      43.00             2.00
9      41.00             2.00
10     40.00             2.00
11     36.00             2.00
12     23.00             1.00
13     27.00             1.00
14     28.00             1.00
15     44.00             1.00
16     38.00             1.00
17     20.00             1.00
18     39.00             1.00
19     25.00             1.00
20     46.00             1.00
21     29.00             1.00
22     48.00             1.00
23     42.00             1.00
total   NaN             36.00
```

**0.0.3 PARTE 3.** organize uma tabela de cruzamento ou tabela de contingencia que apresente informações sobre o nível de instrução e idade, com as frequencias relativas.

```
[ ]: # Cria uma tabela de cruzamento
crosstable = pd.crosstab(data['ano'], data['instrucao'], margins=True,
    ↪ margins_name='total', normalize=True)

# Cria a tabela total
crosstable['idade'] = crosstable.rename(index={'total': 0}, columns={'ano':
    ↪ 'idade'}).index

# Agrupa as idade em intervalos de 5
```

```

crosstable_agrupada = crosstable.groupby(pd.cut(crosstable['idade'], np.
    ↳arange(20, 55, 5))).sum()

# Limpa informações extras da tabela
crosstable_agrupada = crosstable_agrupada.drop('idade', axis=1)

total_frame = pd.DataFrame(crosstable.loc['total']).transpose().drop('idade',
    ↳axis=1)

# Concatena o total a tabela
crosstable_agrupada = pd.concat([crosstable_agrupada, total_frame])

crosstable_agrupada.columns.name = 'idade'

# Imprime a tabela
print('Tabela de contingencia de idade e grau de instrução')
crosstable_agrupada

```

Tabela de contingencia de idade e grau de instrução

```

[ ]: idade      ensino fundamental  ensino médio  ensino superior  total
(20, 25]          0.00           0.03           0.03      0.06
(25, 30]          0.08           0.08           0.03      0.19
(30, 35]          0.03           0.19           0.06      0.28
(35, 40]          0.08           0.14           0.00      0.22
(40, 45]          0.11           0.00           0.06      0.17
(45, 50]          0.03           0.03           0.00      0.06
total             0.33           0.50           0.17      1.00

```

#### 0.0.4 PARTE 4. Em qual situação (cite um exemplo) em que é melhor em optar pela apresentação da tabela com base na frequência relativa e em que é melhor em optar pela apresentação da tabela com base na frequência absoluta?

A frequência relativa é utilizada quando comparando dados de duas tabelas distintas, não podendo assim comparar os dados diretamente mas sim por porcentagem relativo ao total da tabela. Exemplo: comparando idade dos funcionarios de uma empresa com 200 funcionarios com outra empresa com 1000 funcionarios, fica impossivel comparar com a frequência absoluta.

A frequência absoluta é utilizada quando se quer ter uma visão geral dos dados, por exemplo quando se quer fazer uma descrição inicial dos dados.

#### 0.0.5 PARTE 5. os dados apresentados na Parte 3, servem de base para as frequências marginais das variáveis descritas. Escolha uma delas e descreva em termos de probabilidades.

```

[ ]: porcentagem_30a34anos = 0.28
    ensino_medio_30a34anos = 0.19 + 0.06

```

```

ensino_medio_geral = 0.5 + 0.17

# 100 / 28 = 3.57
# 3.57 * 23 = 89.29
chance_ensino_medio = ensino_medio_30a34anos / porcentagem_30a34anos

print(f'A porcentagem de funcionarios entre 30 e 34 anos é de_
↳ {porcentagem_30a34anos * 100:.2f}% e a porcentagem de funcionarios entre 30_
↳ e 34 anos que possuem ensino médio ou superior é de {ensino_medio_30a34anos_
↳ * 100:.2f}%, então dado um funcionario aleatório entre 30 e 34 anos a_
↳ probabilidade dele possuir ao menos o ensino médio é de {chance_ensino_medio_
↳ * 100:.2f}%, a chance de um funcionario qualquer da empresa possuir ensino_
↳ médio ou superior é de {ensino_medio_geral * 100:.2f}%. ' )

```

A porcentagem de funcionarios entre 30 e 34 anos é de 28.00% e a porcentagem de funcionarios entre 30 e 34 anos que possuem ensino médio ou superior é de 25.00%, então dado um funcionario aleatório entre 30 e 34 anos a probabilidade dele possuir ao menos o ensino médio é de 89.29%, a chance de um funcionario qualquer da empresa possuir ensino médio ou superior é de 67.00%.