



3ª LISTA DE EXERCÍCIOS (Use sempre 2 casas decimais)

Turma: Engenharia de Software
RGA: 2021.1906.069-7
Aluno: Maycon Felipe da Silva Mota

Lista de Exercicios

```
In [94]: import pandas as pd
import numpy as np
import statistics as stats
import matplotlib.pyplot as plt # Importar a biblioteca Matplotlib
import collections
plt.style.use('seaborn-whitegrid') # Definir um estilo para a Biblioteca Matplotlib
```

1ª Questão – Há duas turmas de Estatística numa certa Universidade. A turma A possui 8 alunos, que na primeira prova tiraram: 6, 8, 5, 7, 4, 6, 6 e 7. A turma B que possui 7 alunos, houve notas de 4, 5, 3, 6, 5, 4 e 5. Calcule as medidas de posição que você conhece, e explique por meio destas, qual turma possui maiores notas de Estatística?

```
In [95]: dados = {'turma_a': [6, 8, 5, 7, 4, 6, 6, 7], 'turma_b': [4, 5, 3, 6, 5, 4, 5, np.NaN]} # Definir os dados conforme o exercicio pede
dados = pd.DataFrame(data=dados) # Criar um DataFrame com basen os dados

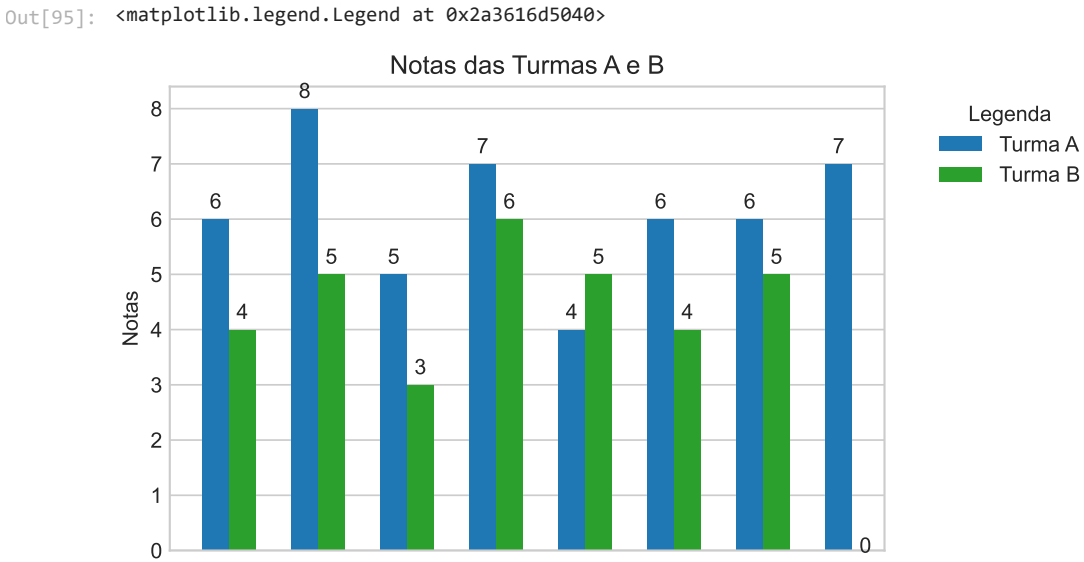
rotulos = ['Turma A', 'Turma B'] # Definir rótulos
x = np.arange(len(rotulos)) # Utilizar o NumPy para criar uma lista de números até a quantidade dos rótulos para colocarmos a Localização das barras // [0, 1]
width = 0.3 # Definir o tamanho das barras

x = np.arange(len(dados['turma_a'])) # Criar uma Lista de números até a quantidade que temos na Lista de Rotulos
width = 0.30 # Definir a Largura das barras

fig, axes = plt.subplots() # Criar o gráfico
plt.title('Notas das Turmas A e B') # Definir um título
plt.ylabel('Notas') # Definir um rótulo para o eixo X

rects1 = axes.bar(x, dados['turma_a'], width=1.*width, align='edge', label="Turma A") # Criando as barras para turma A
rects2 = axes.bar(x, dados['turma_b'], width=width, align='edge', color=list(plt.rcParams['axes.prop_cycle'])[2]['color'], label="Turma B") # Criando as barras para turma B e adicionado cores diferentes para turma B

axes.bar_label(rects1, padding=3) # Criando rótulo dos valores da turma A
axes.bar_label(rects2, padding=3) # Criando rótulo dos valores da turma B
axes.get_xaxis().set_visible(False) # Ocultando o eixo X
plt.legend(handles=[rects1, rects2], title='Legenda', bbox_to_anchor=(1.05, 1), loc='upper left') # Criando a Legenda, definindo o titulo e deixando ela fora do box do gráfico
```



```
In [96]: # Gerar a estatística descritiva
dados.describe()
```

Out[96]:

	turma_a	turma_b
count	8.000000	7.000000
mean	6.125000	4.571429
std	1.246423	0.975900
min	4.000000	3.000000
25%	5.750000	4.000000
50%	6.000000	5.000000
75%	7.000000	5.000000
max	8.000000	6.000000

Com base no gráfico e na estatística descritiva gerado acima, podemos definir que a **Turma A** possui notas melhores em relação a Turma B. Partindo da Média, identificamos que a soma de todas as notas dividindo pela quantidade de alunos resulta em 6.12/4.57, importante atentarmos que a Turma B possui um aluno a menos. Analisando o desvio padrão, percebemos que a dispersão das notas dos alunos da turma A é de 1.24, ou seja, podemos definir que estatisticamente, os alunos não estão alinhados conforme o padrão da sala, diferente da Turma B onde há um valor de 0.97. Percebemos pela análise dos quartis que em ambas turmas, os alunos low-level, isso é, os de baixo desempenho estão similares, enquanto os de alto desempenho são diferentes, na turma A, ter bom desempenho (p75>) é ter nota 7 enquanto na turma B, é 5. Em valores mínimos, a Turma A está parecida com a Turma B, entretanto em valores máximo, a discrepância é por dois pontos.

2ª Questão – Qual a diferença do 1º e 3º quartil?

O primeiro quartil tende a apresentar os valores de 25% da amostra ordenada, isso é, percentil 25. Já o terceiro quartil, conhecido como quartil superior, tente a mostrar os 25% dos valores mais elevados da amostra ordenada, isso é, percentil 75.

3ª Questão – Qual a diferença entre mediana e o 2º quartil? Por quê?

```
In [97]: dados.median()
```

Out[97]:

turma_a	6.0
turma_b	5.0
dtype:	float64

Não há diferenças entre a mediana e o 2º quartil, a literatura entende que as duas são a mesma coisa.

```
In [98]: ### 4ª Questão - Considere os seguintes dados, referentes ao peso de 12 crianças com sete anos, em kg, calcule a média, mediana, quartis e moda, e interprete-as de maneira correta.

medidas_peso = [13, 13, 14, 14, 14, 15, 15, 16, 16, 16, 17, 17]
medidas_peso = pd.DataFrame(medidas_peso, columns=['Peso'])
medidas_peso['Peso'].describe()
```

Out[98]:

count	12.000000
mean	15.000000
std	1.414214
min	13.000000
25%	14.000000
50%	15.000000
75%	16.000000
max	17.000000
Name:	Peso, dtype: float64

```
In [99]: # mediana
print("Mediana: {}".format(stats.median(medidas_peso['Peso'])))
print("Moda: {}".format(stats.mode(medidas_peso['Peso'])))
print("Media: {}".format(stats.mean(medidas_peso['Peso'])))
```

Mediana: 15.0
Moda: 14
Media: 15

Inicialmente, temos 12 crianças, a média de peso dessas crianças são 15kgs. O desvio padrão, isso é a dispersão entre os valores de peso das crianças são 1.41, isso indica que as crianças possuem pesos muito diferentes entre si. Para os quartis, vamos dividir os pesos das crianças em 25%, dando 3 frações. A primeira parte nos trás que o peso é 14kg, a segunda parte, 15kg e a terceira parte, 16kg. A moda, o peso mais comum entre as crianças é 14kg. a mediana, o meio do conjunto de medidas do peso dessas crianças, é 15kg. A média, a soma dividido pela quantidade de crianças pesadas, é 15kg.

5ª Questão – Da Questão 4, classifique o conjunto de dados em relação a moda.

Não entendi muito bem, mas imagino que seja o que segue

```
In [100]: def frequency_sort(a): # cria uma função para ordenar por frequência/moda
f = collections.Counter(a) # conta quantos elementos se repetem
a.sort(key = lambda x: (~f[x], x)) # ordena pela qtd de elementos que se repetiram
return a # devolta f
```

```
print(frequency_sort([13, 13, 14, 14, 14, 15, 15, 16, 16, 16, 17, 17])) # imprime o resultado da função frequency_sort
```

[14, 14, 14, 16, 16, 16, 13, 13, 15, 15, 17, 17]

6ª Questão – Pra que serve uma medida de posição?

A medida de posição ou medida de tendência central são conjuntos de valores para analisarmos a distribuição de valores de um determinado conjunto. A MTC tem o objetivo de caracterizar um conjunto de dados.