

Introdução à Ciência de Dados

Projeto 2

- 1 - Gere dados a partir de uma distribuição de Poisson. Varie a taxa λ no intervalo $[1,10]$ e mostre o gráfico da média em função da variância.
- 2 - Considere os dados da Iris. Calcule a média, variância e IQR para cada atributo.
- 3 - Obtenha o boxplot de todas as variáveis da flor Iris, para cada espécie.
- 4 - Para a função `log()`, investigue como as correlações de Pearson e Spearman variam de acordo com o intervalo dos dados.
- 5 - Considere o código que mostra como a correlação de Pearson muda com a inclusão de ruídos (ver Python: Análise Exploratória de Dados). Modifique a função para $Y=0.5*X+Y=0.5*X+$ ruído. Varie o ruído e calcule os coeficientes de Pearson e Spearman, mostrando os respectivos scatterplots com os valores dos coeficientes (como feito no exemplo).
- 6- Considere os dados abaixo, chamado quarteto de Ascomb. Calcule a média, variância, correlação de Pearson e Spearman entre as variáveis x e y . O que você pode dizer sobre esses dados?

```
import matplotlib.pyplot as plt
import numpy as np

x = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y1 = [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68]
y2 = [9.14, 8.14, 8.74, 8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74]
y3 = [7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73]
x4 = [8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8]
y4 = [6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.50, 5.56, 7.91, 6.89]

datasets = {
    'I': (x, y1),
    'II': (x, y2),
    'III': (x, y3),
    'IV': (x4, y4)
}
```

```

fig, axs = plt.subplots(2, 2, sharex=True, sharey=True, figsize=(10
, 5),
                        gridspec_kw={'wspace': 0.08, 'hspace': 0.08
})

for ax, (label, (x, y)) in zip(axs.flat, datasets.items()):
    ax.text(0.1, 0.9, label, fontsize=20, transform=ax.transAxes, v
a='top')
    ax.tick_params(direction='in', top=True, right=True)
    ax.plot(x, y, 'o')
plt.show()

```

