

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import pearsonr, linregress
```

✓ Exercício 6.5 da Apostila - Página 105

Os dados adiante fornecem, para 11 países, o consumo de cigarros per capita em 1930 e as mortes por 1.000.000 de habitantes em 1950 causadas por câncer no pulmão.

País	Consumo de cigarros	Mortes (por 1.000.000 de hab.)
Islândia	240	63
Noruega	255	100
Suécia	340	140
Dinamarca	375	175
Canadá	510	160
Austrália	490	180
Holanda	490	250
Suíça	180	180
Finlândia	1.125	360
Grã-Bretanha	1.150	470
EUA	1.275	200

```
# Dados fornecidos
```

```
países = ["Islândia", "Noruega", "Suécia", "Dinamarca", "Canadá", "Austrália", "Holanda", "Suíça", "Finlândia", "Grã-Bretanha", "EUA"]
consumo_cigarros_1930 = [240, 255, 340, 375, 510, 490, 490, 180, 1125, 1150, 1275] # Consumo per capita em 1930
mortes_cancer_1950 = [63, 100, 140, 175, 160, 180, 250, 180, 360, 470, 200] # Mortes por câncer no pulmão por 1.000.000 habitantes em 1950
```

(a) Escolha adequadamente X e Y

```
X = consumo_cigarros_1930
Y = mortes_cancer_1950
```

```
print(f"Variável Independente (X): Consumo de cigarros per capita em 1930 = {X}")
print(f"Variável Dependente (Y): Mortes por câncer no pulmão em 1950 = {Y}")
```

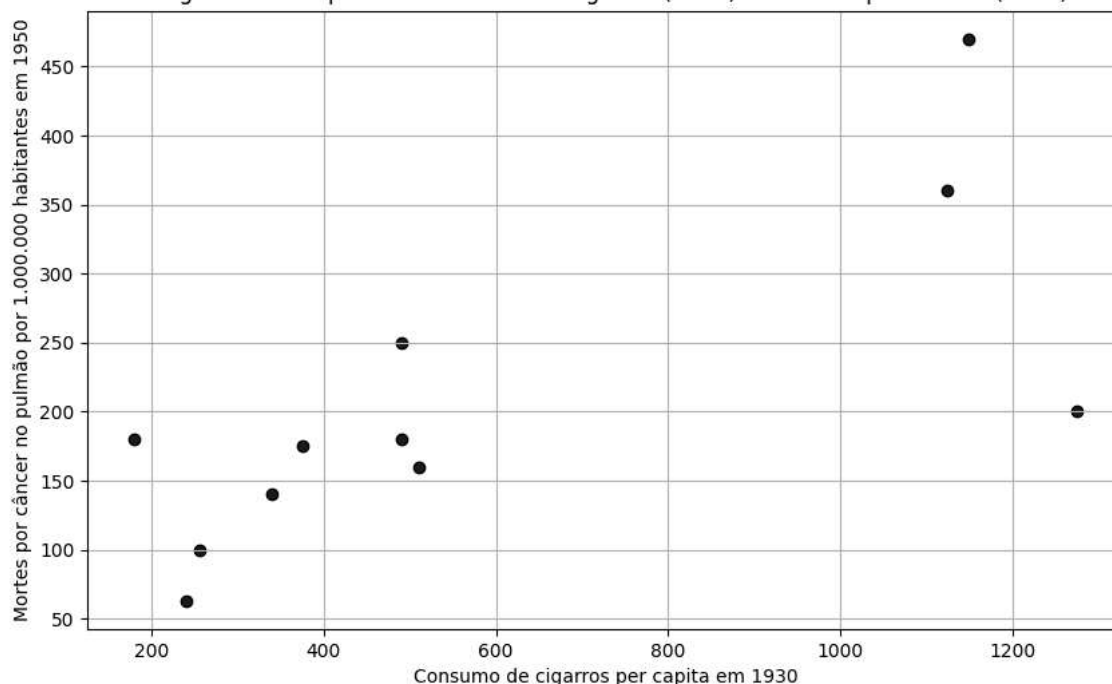
↔ Variável Independente (X): Consumo de cigarros per capita em 1930 = [240, 255, 340, 375, 510, 490, 490, 180, 1125, 1150, 1275]
Variável Dependente (Y): Mortes por câncer no pulmão em 1950 = [63, 100, 140, 175, 160, 180, 250, 180, 360, 470, 200]

(b) Construa o diagrama de dispersão entre X e Y.

```
plt.figure(figsize=(10, 6))
plt.scatter(X, Y, color='blue', edgecolor='black')
plt.title('Diagrama de Dispersão: Consumo de Cigarros (1930) vs. Mortes por Câncer (1950)')
plt.xlabel('Consumo de cigarros per capita em 1930')
plt.ylabel('Mortes por câncer no pulmão por 1.000.000 habitantes em 1950')
plt.grid(True)
plt.show()
```



Diagrama de Dispersão: Consumo de Cigarros (1930) vs. Mortes por Câncer (1950)



(c) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.

```
r, p_value = pearsonr(X, Y)
print(f"Coeficiente de Pearson: {r:.4f}")
print(f"Valor-p: {p_value:.4g}")
if r > 0.7:
    interpretacao = "Correlação linear positiva forte."
elif r > 0.4:
    interpretacao = "Correlação linear positiva moderada."
elif r > 0:
    interpretacao = "Correlação linear positiva fraca."
elif r < -0.7:
    interpretacao = "Correlação linear negativa forte."
elif r < -0.4:
    interpretacao = "Correlação linear negativa moderada."
elif r < 0:
    interpretacao = "Correlação linear negativa fraca."
else:
    interpretacao = "Correlação linear muito fraca ou inexistente."
print(f"Interpretação: {interpretacao}")
```



Coeficiente de Pearson: 0.7437
Valor-p: 0.008692
Interpretação: Correlação linear positiva forte.

(d) Obtenha a equação de regressão que melhor descreve a associação entre X e Y. Interprete os valores de b_0 e b_1 .

```
slope, intercept, r_value, p_value, std_err = linregress(X, Y)
print(f"Coeficiente angular (b1): {slope:.4f}")
print(f"Intercepto (b0): {intercept:.4f}")
print(f"Equação estimada: Y = {intercept:.2f} + {slope:.2f}*X")
print(f"Interpretação de b1: A cada unidade a mais de consumo, espera-se {slope:.2f} mortes adicionais.")
print(f"Interpretação de b0: Se o consumo fosse zero, estima-se {intercept:.2f} mortes.")
```



Coeficiente angular (b1): 0.2163
Intercepto (b0): 80.6802
Equação estimada: $Y = 80.68 + 0.22 \cdot X$
Interpretação de b1: A cada unidade a mais de consumo, espera-se 0.22 mortes adicionais.
Interpretação de b0: Se o consumo fosse zero, estima-se 80.68 mortes.

(e) Se, no ano de 1930, o consumo de cigarros per capita no Brasil foi 630 estime o número de mortes causadas por câncer de pulmão no ano de 1950.

```
X_brasil = 630
Y_brasil_estimado = intercept + slope * X_brasil
print(f"Estimativa de mortes por câncer: {Y_brasil_estimado:.2f} por 1.000.000 de habitantes")
```



Estimativa de mortes por câncer: 216.92 por 1.000.000 de habitantes

(f) Calcule e interprete o coeficiente de determinação R^2 para a equação de regressão estimada no item (d).

```
R2 = r ** 2  
print(f"Coeficiente de Determinação ( $R^2$ ): {R2:.4f} (ou {R2*100:.2f}%)")
```

↔ Coeficiente de Determinação (R^2): 0.5531 (ou 55.31%)