

Relatório Técnico: Correlação entre Grupos A e B

Autora: Morgana Souza Francisco de Assis

Data: 11/04/2025

1. Hipóteses

Este estudo teve como objetivo principal investigar se existe uma relação estatisticamente significativa entre dois grupos de dados quantitativos.

- Hipótese Nula (H_0): Não há correlação significativa entre os valores dos grupos A e B.
- Hipótese Alternativa (H_1): Existe uma correlação significativa entre os valores dos grupos A e B.

2. Materiais e Métodos

2.1. Dados

Foram utilizados dois grupos numéricos hipotéticos compostos por 5 elementos cada:

- Grupo A: [3, 5, 7, 9, 11]
- Grupo B: [2, 4, 6, 8, 10]

2.2. Testes Estatísticos

1. Teste de Shapiro-Wilk (Shapiro, Wilk, 1965): Avaliou a normalidade dos dados em cada grupo, condição essencial para definir o teste de correlação mais adequado.
2. Correlação de Pearson (Pearson, 1895): Aplicado para medir a intensidade da relação linear entre os dois grupos, já que ambos apresentaram distribuição normal.

2.3. Ferramentas

As análises estatísticas foram conduzidas na linguagem Python 3.9 (Python Software Foundation, 2025), utilizando as seguintes bibliotecas:

- pandas (McKinney, 2011) para estruturação dos dados
- scipy.stats (Virtanen, 2020) para os testes estatísticos
- matplotlib e seaborn (Hunter, 2007; Waskom, 2021) para geração dos gráficos

3. Resultados

3.1. Estatísticas Descritivas:

- Grupo A: Média = 7.0 | Mediana = 7.0 | Desvio Padrão = 3.16
- Grupo B: Média = 6.0 | Mediana = 6.0 | Desvio Padrão = 3.16

3.2. Normalidade dos Dados (Shapiro-Wilk):

- Grupo A: $p = 0.967$
- Grupo B: $p = 0.967$

Como $p > 0.05$, ambos os grupos podem ser considerados normalmente distribuídos.

3.3. Correlação entre os Grupos (Pearson):

- Coeficiente de correlação (r) = 1.0
- Valor- p = 0.0

O resultado aponta uma correlação perfeita positiva, estatisticamente significativa, entre os valores dos Grupos A e B.

4. Referências

HUNTER, J. D. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, v. 9, n. 3, p. 90-95, 2007.

McKINNEY, W. pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing, v. 14, p. 1-9, 2011.

PEARSON, K. Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, v. 58, p. 240-242, 1895. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rspl.1895.0041>.

PYTHON SOFTWARE FOUNDATION. Python Language Reference. Disponível em: <https://www.python.org/>. Acesso em: 11 abril 2025.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). Biometrika, v. 52, n. 3-4, p. 591-611, 1965.

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, v. 17, p. 261-272, 2020.

WASKOM, M. L. seaborn: statistical data visualization. Journal of Open Source Software, v. 6, n. 60, p. 3021, 2021.

5. Anexo

Notebook Python: <https://github.com/MorganaSouza/DasafioDataScience2>

Colab:

https://colab.research.google.com/drive/10PysThlthQWhy8r3_kDDCHk8su9ZLzQj?usp=sharing