

# **ANÁLISE MULTIVARIADA**

## **DE DADOS COM R**

Breno Cauã Rodrigues da Silva

# Índice

<b>Prefácio</b>	<b>3</b>
<b>I MULTIVARIADA I</b>	<b>4</b>
<b>II MULTIVARIADA II</b>	<b>5</b>
<b>1 Unidade I – Introdução</b>	<b>6</b>
1.1 Leitura de Dados Multivariados . . . . .	6
<b>References</b>	<b>7</b>

# Prefácio

Em um mundo cada vez mais impulsionado por dados, a capacidade de extrair *insights* significativos de conjuntos de dados complexos tornou-se uma habilidade indispensável. A análise multivariada, com sua gama de técnicas para explorar relações entre múltiplas variáveis, figura como uma ferramenta poderosa nesse cenário.

Este livro, “**Análise Multivariada de Dados com R**”, foi concebido com o propósito de desmistificar e tornar acessíveis os conceitos e aplicações da análise multivariada, utilizando a linguagem R como nossa principal ferramenta. Acreditamos que, ao combinar a teoria com exemplos práticos e implementações em R, podemos capacitar estudantes, pesquisadores e profissionais a aplicar essas técnicas em suas próprias áreas de interesse.

Nosso objetivo foi criar um recurso que não apenas introduza os fundamentos da análise multivariada, mas que também sirva como um guia prático para a resolução de problemas reais. Esperamos que este material seja um valioso companheiro em sua jornada de exploração de dados, abrindo novas perspectivas e *insights*.

---

Este é um *Quarto book*. Para saber mais sobre *Quarto books*, visite <https://quarto.org/docs/books/>.

**Part I**

**MULTIVARIADA I**

**Part II**

**MULTIVARIADA II**

# 1 Unidade I – Introdução

A *análise multivariada* de dados refere-se a um conjunto de técnicas estatísticas que possibilita a análise simultânea de múltiplas medidas para indivíduos, objetos ou fenômenos diversos observados. De maneira geral, a análise multivariada é usada para:

1. **Redução de dados ou simplificação estrutural:** explora a correlação entre as variáveis originais para construir índices ou outro conjunto de variáveis que sintetizam as variáveis originais, sem perder a informação.
  - **Ex:** Análise Fatorial.
2. **Classificação e discriminação:** agrupa indivíduos ou objetos ou variáveis similares de acordo com as suas características, pode ser utilizada para dados com variável resposta (dados supervisionados), ex: análise discriminante, ou para dados sem variável resposta (dados não supervisionados).
  - **Ex:** Análise de Agrupamentos.
3. **Analisar a relação entre as variáveis:** avalia a relação de dependência entre uma variável e um conjunto de outras variáveis, ou a dependência mútua entre grupos de variáveis.
  - **Ex:** Modelos de Regressão ou Equações Estruturais.

Em análise multivariada o interesse da análise pode ser na estrutura das variáveis ou dos indivíduos.

## 1.1 Leitura de Dados Multivariados

A matriz de dados

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix},$$

onde  $x_{jk}$  representa o  $j$ -ésimo “indivíduo” para a  $k$ -ésima variável.

Essa matriz pode ser simplificada por meio de um vetor aleatório, cujos elementos são as variáveis aleatórias:  $\mathbf{X}^\top = [\mathbf{X}_1 \ \mathbf{X}_2 \ \dots \ \mathbf{X}_p]^\top$ .

# References