**UNIVERSIDADE DE SÃO PAULO**

**ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES**

**ACH-2036 - Métodos Quantitativos para Análise Multivariada**

Prof. Dr. Régis Rossi Alves Faria

**Entrega 1 - Preparo e Análise de Banco de Dados**

Alunos:

Diego Aguiar Garcia

Gabriel Fernandes Marchesano

Mateus de Sales Francisco

Nicolas Garcia

SÃO PAULO  
 2025

**Sumário**

[**1. Motivação de Pesquisa 2**](#_76l2csyb90h5)

[**2. Revisão Teórica 2**](#_spqz3x37q0lu)

[**3. Metodologia 2**](#_55upe36bsuyd)

[**4. Análise Descritiva 2**](#_l3hjngtduj4f)

[**5. Resultados e Observações 2**](#_2791uaarcad4)

[**Referências 2**](#_yelpwzxio0pp)

# 1. Motivação de Pesquisa

# 2. Revisão Teórica

A análise descritiva de dados constitui a base fundamental da estatística aplicada, fornecendo ferramentas essenciais para a compreensão inicial de conjuntos de dados e identificação de padrões relevantes. A estatística descritiva tem como objetivo organizar, resumir e apresentar os dados de forma clara e objetiva, permitindo ao pesquisador extrair informações significativas sobre as características principais das variáveis em estudo¹.

## 2.1 Classificação das Variáveis

O processo de análise descritiva inicia-se com a identificação e classificação das variáveis, que podem ser categorizadas como qualitativas (nominais ou ordinais) ou quantitativas (discretas ou contínuas). Esta distinção é fundamental pois determina os métodos estatísticos apropriados para cada tipo de variável¹.

## 2.2 Medidas de Tendência Central

As medidas de tendência central constituem indicadores que representam o valor típico ou central de um conjunto de dados. As principais medidas são:

**Média Populacional (μ):** É definida como a soma de todos os valores populacionais dividida pelo número total de observações da população:

onde Xᵢ representa cada valor individual da população e N o número total de observações populacionais.

**Média Amostral (x̄):** É definida como a soma de todos os valores amostrais dividida pelo número total de observações da amostra:

onde xᵢ representa cada valor individual da amostra e n o número total de observações amostrais.

**Mediana (Md):** Valor que divide o conjunto de dados ordenados em duas partes iguais, representando o percentil 50.

**Moda (Mo):** Valor ou valores que aparecem com maior frequência no conjunto de dados.

## 2.3 Medidas de Variabilidade

As medidas de variabilidade quantificam a dispersão dos dados em torno das medidas de tendência central:

**Variância Populacional (σ²):** Mede a dispersão dos dados populacionais em relação à média populacional:

onde N representa o número total de observações na população.

**Variância Amostral (s²):** Mede a dispersão dos dados amostrais em relação à média amostral:

onde o denominador (n-1) representa os graus de liberdade da amostra, fornecendo um estimador não viesado da variância populacional.

**Desvio-padrão Populacional (σ):** Raiz quadrada da variância populacional:

**Desvio-padrão Amostral (s):** Raiz quadrada da variância amostral, expressa na mesma unidade dos dados originais:

Esta medida fornece uma interpretação mais intuitiva da variabilidade por estar na mesma escala dos dados.

**Coeficiente de Variação (CV):** Medida de dispersão relativa que permite comparar variabilidades de diferentes conjuntos de dados:

## 2.4 Medidas Separatrizes

As medidas separatrizes dividem o conjunto de dados ordenados em partes iguais:

**Quartis:** Dividem os dados em quatro partes iguais. O primeiro quartil (Q₁) representa o percentil 25, o segundo quartil (Q₂) corresponde à mediana (percentil 50), e o terceiro quartil (Q₃) representa o percentil 75.

**Amplitude Interquartílica (AIQ):** Diferença entre o terceiro e primeiro quartis:

## 2.5 Representação Gráfica

### **2.5.1 Histogramas**

O histograma é uma representação gráfica da distribuição de frequências de uma variável quantitativa contínua. Consiste em retângulos adjacentes onde a área de cada retângulo é proporcional à frequência da classe correspondente. O histograma permite identificar a forma da distribuição, presença de assimetrias, multimodalidade e valores atípicos².

### **2.5.2 Boxplots**

O boxplot, também conhecido como diagrama de caixa, é uma representação gráfica que exibe simultaneamente várias medidas descritivas: mediana, quartis, amplitude interquartílica e valores extremos. A caixa representa 50% dos dados centrais (entre Q₁ e Q₃), com uma linha interna indicando a mediana. As "bigodes" estendem-se até os valores máximo e mínimo dentro de 1,5 × AIQ dos quartis, e pontos além desse limite são considerados outliers².

## 2.6 Avaliação da Normalidade

A avaliação da normalidade dos dados representa aspecto fundamental na análise descritiva. Uma distribuição normal é caracterizada por ser simétrica, unimodal e com forma de sino, seguindo a função de densidade de probabilidade:

onde μ é a média populacional e σ o desvio-padrão populacional.

A verificação da normalidade pode ser realizada através de métodos gráficos (histogramas, gráficos Q-Q) e testes estatísticos formais (Shapiro-Wilk, Kolmogorov-Smirnov)³. A presença de distribuições não normais pode indicar a necessidade de transformações dos dados ou utilização de métodos não paramétricos.

# 3. Metodologia

# 4. Análise Descritiva

# 5. Resultados e Observações

# Referências

¹ MORETTIN, P. A.; BUSSAB, W. O. Estatística Básica.

² CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. Análise Multivariada. 1ª ed. Editora Atlas, 2009.

³ HAIR JR., J. F. et al. Análise Multivariada de Dados. 6ª ed. Porto Alegre: Bookman, 2009.