



*Microsoft Power BI Para Business Intelligence e Data Science*

# Microsoft Power BI Para Business Intelligence e Data Science

## Métodos Estatísticos Para Análise de Dados

Existem diversos métodos estatísticos para análise de dados que podem ser utilizados de acordo com o objetivo do projeto de Ciência de Dados. Aqui estão alguns dos métodos estatísticos mais comuns:

### **Análise Descritiva**

A análise descritiva envolve a descrição e resumo dos dados por meio de medidas de tendência central (média, mediana, moda), medidas de dispersão (variância, desvio padrão, coeficiente de variação), e medidas de posição relativa (percentis, quartis, z-scores). Essa análise fornece uma visão geral dos dados e ajuda a entender sua distribuição e características básicas.

### **Análise Exploratória de Dados (EDA – Exploratory Data Analysis)**

A EDA é uma abordagem para analisar conjuntos de dados, geralmente com o objetivo de identificar padrões, tendências, outliers e relações entre variáveis. Essa análise envolve a criação de gráficos, como histogramas, gráficos de dispersão, gráficos de caixa e gráficos de barras, para visualizar os dados e gerar insights.

### **Testes de Hipóteses**

Os testes de hipóteses são métodos estatísticos que envolvem a formulação de hipóteses nulas e alternativas sobre os parâmetros de uma população e o uso de dados amostrais para testar a validade dessas hipóteses. Exemplos de testes de hipóteses incluem o teste t de Student, o teste qui-quadrado e o teste de Mann-Whitney.

### **Regressão**

A análise de regressão é um método estatístico utilizado para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. A regressão linear é a forma mais simples de regressão e descreve a relação linear entre as variáveis.

### **Análise de Variância (ANOVA)**

A ANOVA é um método estatístico utilizado para comparar as médias de três ou mais grupos, determinando se existem diferenças significativas entre eles. Existem vários tipos de ANOVA, como ANOVA de um fator, ANOVA de dois fatores e ANOVA de medidas repetidas.

### **Análise de Séries Temporais**

A análise de séries temporais envolve a análise de dados coletados ao longo do tempo para identificar padrões, tendências e ciclos. Essa análise pode incluir a decomposição da série temporal em componentes sazonais e de tendência, a aplicação de modelos autorregressivos e de médias móveis (ARIMA) e a previsão de valores futuros.

## **Análise de Agrupamento (Clusterização)**

A análise de agrupamento é um método de aprendizado não supervisionado que agrupa observações com base em suas características e semelhanças. Existem vários algoritmos de agrupamento, como k-means, agrupamento hierárquico e DBSCAN.

## **Análise de Componentes Principais (PCA)**

A PCA é uma técnica de redução de dimensionalidade que transforma um conjunto de dados com muitas variáveis correlacionadas em um conjunto de dados com variáveis não correlacionadas chamadas de componentes principais.

Todos esses métodos são usados em Data Science e são estudados na prática através de projetos aqui na DSA.