
Probabilidade e Estatística para Ciência da Computação

1. Estatística Descritiva

Conceitos Fundamentais

- **Definição:** Ramo da estatística que visa sumarizar e descrever conjuntos de dados.
- **Objetivo:** Organizar, visualizar e interpretar dados coletados em pesquisas ou observações.
- **Aplicações:** Análise inicial de dados, criação de gráficos e tabelas, cálculo de medidas resumo.

Tipos de Variáveis

- **Quantitativas (Numéricas)**
 - **Discretas:** Valores resultantes de contagem (ex.: número de filhos, peças defeituosas).
 - **Contínuas:** Valores resultantes de medição (ex.: altura, peso, renda).
- **Qualitativas (Categóricas)**
 - **Nominais:** Sem ordenação (ex.: sexo, cor dos olhos).
 - **Ordinais:** Com ordenação (ex.: escolaridade, estágio de doença).

Representação de Dados

- **Tabelas**
 - **Séries Cronológicas:** Dados ordenados por tempo (ex.: vendas mensais).
 - **Séries Geográficas:** Dados agrupados por localização (ex.: vendas por estado).

- Séries Específicas: Dados categorizados por fenômeno (ex.: vendas por linha de produto).

- **Gráficos**

- Barras: Comparação entre categorias.
- Pizza: Proporções de categorias.
- Linhas: Tendências temporais.
- Dispersão: Relação entre variáveis quantitativas.
- Histograma: Distribuição de frequências.
- Boxplot: Resumo de medidas (quartis, outliers).

Medidas Resumo

- **Tendência Central**

- Média: Soma dos valores ÷ número de observações.
- Mediana: Valor central quando os dados estão ordenados.
- Moda: Valor mais frequente.

- **Dispersão**

- Amplitude: Diferença entre máximo e mínimo.
- Variância e Desvio Padrão: Variabilidade em torno da média.
- Coeficiente de Variação: Desvio padrão relativo à média.

Exemplo Prático em R

```
# Cálculo de medidas descritivas
dados <- c(23, 24, 25, 24, 30, 28, 31)
media <- mean(dados)
mediana <- median(dados)
desvio padrão <- sd(dados)
```

2. Probabilidade

Conceitos Fundamentais

- **Definição:** Medida da incerteza associada a eventos aleatórios.
- **Aplicações:** Modelagem de fenômenos incertos, previsões, tomada de decisão.

Problemas Clássicos

- **Paradoxo dos Aniversários:**
Em um grupo de 23 pessoas, a probabilidade de duas compartilharem aniversário é $\approx 50\%$.
- **Caixa de Bertrand:**
Probabilidade condicional: a segunda moeda ser de ouro dado que a primeira é de ouro = $2/3$.
- **Paradoxo de Monty Hall:**
Trocar a porta aumenta a chance de ganhar de $1/3$ para $2/3$.

Exemplo Prático em R

```
# Simulação do Paradoxo de Monty Hall
monty_hall <- function(trocas) {
  premio <- sample(1:3, 1)
  escolha <- sample(1:3, 1)
  if (trocas) {
    portas <- setdiff(1:3, c(escolha, premio))
    abertura <- sample(portas, 1)
    escolha <- setdiff(1:3, c(escolha, abertura))
  }
  escolha == premio
}
```

3. Estatística Inferencial

Conceitos Fundamentais

- **Definição:** Uso de amostras para inferir propriedades de populações.
- **Aplicações:** Testes de hipóteses, intervalos de confiança, modelos preditivos.

População vs. Amostra

- **População:** Conjunto completo de elementos.
- **Amostra:** Subconjunto representativo da população.
- **Parâmetros vs. Estatísticas:**
 - Parâmetro: Medida da população (ex.: média populacional μ).
 - Estatística: Medida da amostra (ex.: média amostral \bar{x}).

Métodos

- **Estimação:** Pontual ou intervalar.
- **Testes de Hipóteses**
 - Hipótese nula (H_0)
 - Hipótese alternativa (H_1)
 - Valor-p: Probabilidade de observar os dados sob H_0

Exemplo Prático em R

```
# Teste t para média
amostra <- c(23, 24, 25, 24, 30, 28, 31)
teste <- t.test(amostra, mu = 25)
```

4. Visualização de Dados em R

Gráficos Comuns

```
# Gráfico de Barras
barplot(table(dados), main = "Distribuição de Frequências")
```

```
# Histograma
hist(dados, breaks = "Sturges", col = "lightblue")
```

```
# Boxplot
boxplot(dados, col = "orange", notch = TRUE)
```

```
# Dispersão
plot(x, y, xlab = "Variável X", ylab = "Variável Y")
```

Pacotes Úteis

- **ggplot2**: Gráficos avançados
 - **dplyr**: Manipulação de dados
 - **corrplot**: Visualização de correlações
-

5. Aplicações Práticas

Ciência da Computação

- Aprendizado de Máquina: Modelos preditivos
- Análise de Dados: Big Data
- Simulações: Modelagem probabilística

Exemplo Final

```
# Regressão Linear Simples  
modelo <- lm(y ~ x, data = dados)  
summary(modelo)  
plot(modelo)
```
