Aula 05. Testes de Hipóteses para Duas Amostras

Estatística Inferencial

MBA CDIA

ENAP - Escola Nacional de Administração Pública

2025

Revisão: Identificando Resultados Significativos

Identifique se as afirmações são verdadeiras ou falsas:

a. Se p-valor $=$ 0,03, o resultado é estatisticamente significativo ao nível $\alpha = 0,05$.
b. Se p-valor $=$ 0,03, a hipótese nula é rejeitada ao nível $\alpha = 0,05$.
c. Se p-valor $=$ 0,03, o resultado é estatisticamente significativo ao nível $\alpha = 0,01$.
d. Se p-valor $=$ 0,03, a hipótese nula é rejeitada ao nível $\alpha = 0,01$.

Identifique se as afirmações são verdadeiras ou falsas:

a. Se p-valor $=$ 0,03, o resultado é estatisticamente significativo ao nível $\alpha = 0,05$.
b. Se p-valor $=$ 0,03, a hipótese nula é rejeitada ao nível $\alpha = 0,05$.
c. Se p-valor $=$ 0,03, o resultado é estatisticamente significativo ao nível $\alpha = 0,01$.
d. Se p-valor $=$ 0,03, a hipótese nula é rejeitada ao nível $\alpha = 0,01$.

Respostas:

- a. Verdadeiro (0,03 $<$ 0,05)
- b. Verdadeiro (p-valor $\leq \alpha \rightarrow$ rejeitar H_0)
- c. Falso (0,03 $>$ 0,01)
- d. Falso (não rejeitar H_0 quando p-valor $> \alpha$)

Introdução: Comparando Duas Populações

- Até agora, realizamos testes de hipóteses para uma única população (ex: média populacional μ).
- Agora, vamos estender os testes de hipóteses para testar afirmações sobre a **diferença entre duas médias populacionais**, $\mu_1 - \mu_2$.
- Aplicações práticas:
  - Comparar eficácia de dois tratamentos médicos
  - Avaliar diferença salarial entre grupos
  - Testar se dois métodos de ensino produzem resultados diferentes
  - Verificar se políticas públicas têm impactos diferentes em regiões distintas

Tipos de Amostras para Comparação

1. **Amostras Independentes:**
    - Duas amostras de populações distintas
    - Sem relação entre as observações
    - Exemplo: Salários de servidores públicos vs. privados
2. **Amostras Pareadas (Dependentes):**
    - Mesmos sujeitos medidos duas vezes
    - Observações relacionadas/pareadas
    - Exemplo: Desempenho antes e depois de um treinamento

Importante: A escolha do teste depende do tipo de amostra!

Teste de Hipóteses para $\mu_1 - \mu_2$: Amostras Independentes

Objetivo: Testar se as médias populacionais μ_1 e μ_2 são iguais.
Hipótese nula: $H_0 : \mu_1 = \mu_2$ (ou $H_0 : \mu_1 - \mu_2 = 0$)
Hipóteses alternativas possíveis:

- $H_1 : \mu_1 < \mu_2$ (unilateral à esquerda)
- $H_1 : \mu_1 > \mu_2$ (unilateral à direita)
- $H_1 : \mu_1 \neq \mu_2$ (bilateral)

Premissa requerida:

- Cada amostra tem tamanho grande ($n > 30$), ou
- Cada população é aproximadamente normal

A estatística de teste é:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(pois $\mu_1 = \mu_2$ sob H_0)

Onde:

- $\bar{x}_1, \bar{x}_2$ = médias amostrais
- s_1, s_2 = desvios padrão amostrais
- n_1, n_2 = tamanhos amostrais

Distribuição: t de Student com $gl = \min(n_1 - 1, n_2 - 1)$

Cálculo do p-valor: Amostras Independentes

Unilateral à esquerda:
$H_1 : \mu_1 < \mu_2$
p-valor $= P(T < t)$

Unilateral à direita:
$H_1 : \mu_1 > \mu_2$
p-valor $= P(T > t)$

Bilateral:
$H_1 : \mu_1 \neq \mu_2$
p-valor $= 2 \cdot P(T > |t|)$

Exemplo 1: Uso de Computadores na Educação

Um estudo comparou o desempenho de estudantes que usaram computadores nas aulas de matemática com aqueles que não usaram.

Com Computador:

- $n_1 = 60$
- $\bar{x}_1 = 309$
- $s_1 = 29$

Sem Computador:

- $n_2 = 40$
- $\bar{x}_2 = 303$
- $s_2 = 32$

Questão: Ao nível de 5% de significância, há diferença entre as médias dos dois grupos?

Solução: Exemplo 1

Passo 1: $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ (bilateral)
Passo 2: $\alpha = 0,05$
Passo 3: Calcular estatística de teste:

$$t = \frac{309 - 303}{\sqrt{\frac{29^2}{60} + \frac{32^2}{40}}} = \frac{6}{\sqrt{14,02 + 25,6}} \approx 0,953$$

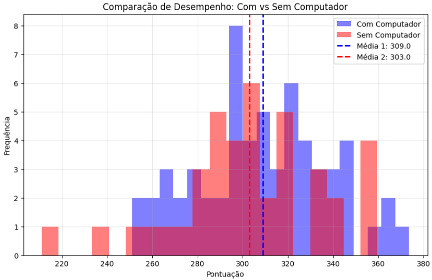
Passo 4: $gl = \min(59, 39) = 39$. Da tabela t: $0,20 <$ p-valor $< 0,50$
Passo 5: Como p-valor $> 0,05$, não rejeitamos H_0.
Conclusão: Não há evidência suficiente para afirmar que o uso de computadores afeta o desempenho.

Implementação em Python: Teste t para Amostras Independentes

```
import numpy as np
from scipy import stats

# Dados do problema
n1, x1_bar, s1 = 60, 309, 29
n2, x2_bar, s2 = 40, 303, 32

# Criar amostras artificiais com essas características
np.random.seed(42)
amostra1 = np.random.normal(x1_bar, s1, n1)
amostra2 = np.random.normal(x2_bar, s2, n2)

# Ajustar para ter exatamente as médias e desvios desejados
amostra1 = (amostra1 - amostra1.mean()) * (s1/amostra1.std()) + x1_bar
amostra2 = (amostra2 - amostra2.mean()) * (s2/amostra2.std()) + x2_bar

# Teste t para amostras independentes
t_stat, p_valor = stats.ttest_ind(amostra1, amostra2, equal_var=False)

print(f"Estatística t: {t_stat:.3f}")
print(f"p-valor: {p_valor:.3f}")
print(f"Conclusão: {'Rejeitar H0' if p_valor < 0.05 else 'Não rejeitar H0'}")
```

Implementação em Python: Teste t para Amostras Independentes (cont.)

```
import matplotlib.pyplot as plt
# Visualização das distribuições
plt.figure(figsize=(10, 6))
plt.hist(amostra1, bins=20, alpha=0.5, label='Com Computador', color='blue')
plt.hist(amostra2, bins=20, alpha=0.5, label='Sem Computador', color='red')
plt.axvline(amostra1.mean(), color='blue', linestyle='--', linewidth=2, label=f'Média 1: {amostra1.mean():.1f}')
plt.axvline(amostra2.mean(), color='red', linestyle='--', linewidth=2, label=f'Média 2: {amostra2.mean():.1f}')
plt.xlabel('Pontuação')
plt.ylabel('Frequência')
plt.title('Comparação de Desempenho: Com vs Sem Computador')
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```



Quando usar amostras pareadas?

- Mesmos sujeitos medidos em dois momentos diferentes
- Pares naturais (gêmeos, casais, antes/depois)
- Matching de sujeitos por características similares

Exemplo: Eficácia de um Programa de Treinamento

Funcionário	1	2	3	4	5	6	7	8
Produtividade Antes	75	82	68	71	79	85	73	77
Produtividade Depois	78	85	72	74	83	88	76	81
Diferença	3	3	4	3	4	3	3	4

Notação para Amostras Pareadas

- $d_i = x_{1i} - x_{2i}$ = diferença para o i-ésimo par
- μ_d = média populacional das diferenças (parâmetro desconhecido)
- $\bar{d}$ = média amostral das diferenças
- s_d = desvio padrão amostral das diferenças

Para o exemplo anterior:

$$\bar{d} = \frac{3+3+4+3+4+3+3+4}{8} = 3,375$$

$$s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}} \approx 0,518$$

Procedimento: Teste t para Amostras Pareadas

Passo 1: Hipóteses sobre μ_d:

- $H_0 : \mu_d = 0$ (sem diferença)
- $H_1 : \mu_d \neq 0$ ou $\mu_d > 0$ ou $\mu_d < 0$

Passo 2: Escolher nível de significância α

Passo 3: Estatística de teste:

$$t = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$$

com distribuição t de Student, $gl = n - 1$

Passo 4: Calcular p-valor

Passo 5: Decisão e conclusão

Exemplo 2: Economia de Combustível Após Revisão

Automóvel	1	2	3	4	5	6	7	8
Após Revisão	35,44	35,17	31,07	31,57	26,48	23,11	25,18	32,39
Antes da Revisão	33,76	34,30	29,55	30,90	24,92	21,78	24,30	31,25
Diferença	1,68	0,87	1,52	0,67	1,56	1,33	0,88	1,14

Teste ao nível $\alpha = 0,01$ se a economia de combustível melhorou após a revisão.
$\bar{d} = 1,206$ km/l, $s_d = 0,373$ km/l

Solução: Exemplo 2

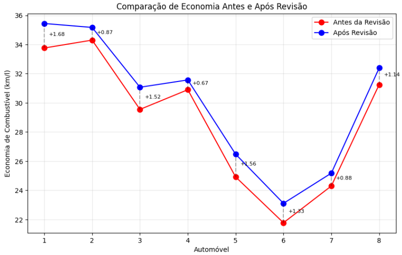
Passo 1: $H_0 : \mu_d = 0$ vs $H_1 : \mu_d > 0$ (melhoria após revisão)
Passo 2: $\alpha = 0,01$
Passo 3: Estatística de teste:

$$t = \frac{1,206 - 0}{0,373/\sqrt{8}} = \frac{1,206}{0,132} \approx 9,142$$

Passo 4: Com $gl = 7$, p-valor $< 0,0005$
Passo 5: Como p-valor $< 0,01$, rejeitamos H_0.
Conclusão: Há forte evidência de que a economia de combustível melhorou significativamente após a revisão.

Implementação em Python: Teste t Pareado

```
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
# Dados do problema
depois = np.array([35.44, 35.17, 31.07, 31.57, 26.48, 23.11, 25.18, 32.39])
antes = np.array([33.76, 34.30, 29.55, 30.90, 24.92, 21.78, 24.30, 31.25])
diferencas = depois - antes

# Teste t pareado
t_stat, p_valor = stats.ttest_rel(depois, antes, alternative='greater')
print(f"Estatística t: {t_stat:.3f}")
print(f"p-valor: {p_valor:.4f}")
print(f"Média das diferenças: {diferencas.mean():.3f} km/l")
print(f"Desvio padrão das diferenças: {diferencas.std(ddof=1):.3f} km/l")
# Visualização antes/depois
plt.figure(figsize=(10, 6))
automoveis = np.arange(1, 9)
plt.plot(automoveis, antes, 'ro-', label='Antes da Revisão', markersize=8)
plt.plot(automoveis, depois, 'bo-', label='Após Revisão', markersize=8)
for i in range(8):
    plt.plot([i+1, i+1], [antes[i], depois[i]], 'k--', alpha=0.3)
    plt.text(i+1.1, (antes[i]+depois[i])/2, f'+{diferencas[i]:.2f}', fontsize=8)
plt.xlabel('Automóvel')
plt.ylabel('Economia de Combustível (km/l)')
plt.title('Comparação de Economia Antes e Após Revisão')
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```



Escolhendo o Teste Correto

Amostras Independentes:

- Grupos diferentes de sujeitos
- Sem relação entre observações
- Exemplos:
  - Homens vs. Mulheres
  - Tratamento A vs. B (grupos diferentes)
  - Região Norte vs. Sul

Amostras Pareadas:

- Mesmos sujeitos, duas medições
- Pares naturalmente relacionados
- Exemplos:
  - Antes vs. Depois
  - Olho direito vs. esquerdo
  - Gêmeo 1 vs. Gêmeo 2

Dica: Amostras pareadas geralmente têm maior poder estatístico!

Exercício Prático: Avaliação de Política de Capacitação

O governo implementou um programa de capacitação profissional. Você tem dados de 150 participantes do programa e 120 não-participantes (grupo controle).

Dados disponíveis:

- `programa_capacitacao.csv`
- Variáveis: salário_mensal, grupo (programa/controle), experiência_anos

Tarefa:

1. Teste se há diferença significativa no salário médio entre os grupos

Solução: Avaliação de Política de Capacitação

```
import pandas as pd
import numpy as np
from scipy import stats
from google.colab import files

# Fazer upload do arquivo no Colab
print("Selecione o arquivo programa_capacitacao.csv")
uploaded = files.upload()

# Ler o arquivo
df = pd.read_csv('aula05-programa_capacitacao.csv')

# Verificar se foi carregado corretamente
print(f"Dataset carregado com {len(df)} registros")
print("\nPrimeiras linhas:")
print(df.head())
print("\nInformações do dataset:")
print(df.info())

# 1. Teste t para amostras independentes
grupo_programa = df[df['grupo'] == 'programa']['salario_mensal']
grupo_controle = df[df['grupo'] == 'controle']['salario_mensal']
t_stat, p_valor = stats.ttest_ind(grupo_programa, grupo_controle, equal_var=False)
print(f"Teste t para diferença de médias:")
print(f"Estatística t: {t_stat:.3f}")
print(f"p-valor: {p_valor:.4f}")
```

Pressupostos e Limitações dos Testes t

Pressupostos dos Testes t:

- **Normalidade:** Dados devem seguir distribuição normal
- Menos crítico para amostras grandes (n $>30) Verificar com histogramas ou testes de normalidade$
- **Independência:** Observações independentes
- **Variâncias iguais:** Para amostras independentes (pode ser relaxado)

Quando os pressupostos são violados:

- Usar testes não-paramétricos:
  - Mann-Whitney U (independentes)
  - Wilcoxon signed-rank (pareadas)
- Transformar os dados (log, raiz quadrada)
- Usar métodos de bootstrap

Resumo: Testes para Duas Amostras

Amostras Independentes:

- Estatística: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- GL: $\min(n_1 - 1, n_2 - 1)$
- Python: `stats.ttest_ind()`

Amostras Pareadas:

- Estatística: $t = \frac{\bar{d}}{s_d / \sqrt{n}}$
- GL: $n - 1$
- Python: `stats.ttest_rel()`

Pontos-chave:

- Identificar corretamente o tipo de amostra
- Verificar pressupostos antes de aplicar o teste
- Interpretar resultados no contexto do problema
- Considerar significância prática além da estatística

Próxima Aula

Aula 06: Testes χ^2 e ANOVA - Análise de Variância

Comparação de mais de duas médiasANOVA de um fatorTestes post-hoc

- **Obrigado!**

Dúvidas?