

# ME951 - Estatística e Probabilidade I

## Parte 20

Notas de aula de ME414 produzidas pelos professores **Samara Kiihl**, **Tatiana Benaglia** e **Benilton Carvalho** modificadas e alteradas pela Profa. **Larissa Avila Matos**

## Inferência para duas populações: Teste de hipótese para duas médias

# Teste de hipótese para duas médias

**População 1:** Coletamos uma amostra aleatória  $X_1, X_2, \dots, X_n$  de uma população com média  $\mu_1$  e a variância  $\sigma_1^2$  e usamos  $\bar{X}$  para estimar  $\mu_1$ .

**População 2:** Coletamos uma amostra aleatória  $Y_1, Y_2, \dots, Y_m$  de uma população com média  $\mu_2$  e a variância  $\sigma_2^2$  e usamos  $\bar{Y}$  para estimar  $\mu_2$ .

A população 1 é independente da população 2.

# Teste de hipótese para duas médias

## Condições:

- 1 As populações 1 e 2 são aproximadamente normais ou
- 2 Os tamanhos amostrais  $n$  e  $m$  são suficientemente grandes.

Se pelo menos uma das condições acima é satisfeita, temos:

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \quad \text{e} \quad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right)$$

# Teste de hipótese para duas médias ( $\sigma_1^2 \neq \sigma_2^2$ )

## Caso 1: Variâncias diferentes e conhecidas

Assumindo que as duas amostras  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_m$  são independentes com  $\sigma_1^2 \neq \sigma_2^2$  conhecidas, temos:

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

# Teste de hipótese para duas médias ( $\sigma_1^2 \neq \sigma_2^2$ )

## Caso 1: Variâncias diferentes e conhecidas

Temos interesse em:  $H_0: \mu_1 - \mu_2 = \Delta_0$  vs  $H_a: \mu_1 - \mu_2 \neq \Delta_0$  (ou  $H_a: \mu_1 - \mu_2 < \Delta_0$  ou  $H_a: \mu_1 - \mu_2 > \Delta_0$ ).

E daí, sob  $H_0$ , temos que:

$$Z = \frac{(\bar{X} - \bar{Y}) - \overbrace{(\mu_1 - \mu_2)}^{\Delta_0}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

## Teste de hipótese para duas médias ( $\sigma_1^2 \neq \sigma_2^2$ )

Se temos interesse em:  $H_0 : \mu_1 - \mu_2 = \Delta_0$  vs  $H_a : \mu_1 - \mu_2 > \Delta_0$ .

Uma amostra aleatória de tamanho  $m$  é coletada da população  $X$  e calcula-se a média amostral,  $\bar{x}$ . Similarmente, para a população  $Y$ , temos  $\bar{y}$  obtida a partir de uma amostra aleatória de tamanho  $n$ .

$$\begin{aligned}\text{p-valor} &= P\left(\frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \geq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}\right) \\ &= P\left(Z \geq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}\right) = 1 - \Phi\left(\frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}\right),\end{aligned}$$

em que  $\Phi(z) = P(Z \leq z)$  para  $Z \sim N(0, 1)$ .

## Teste de hipótese para duas médias ( $\sigma_1^2 \neq \sigma_2^2$ )

Se temos interesse em:  $H_0 : \mu_1 - \mu_2 = \Delta_0$  vs  $H_a : \mu_1 - \mu_2 < \Delta_0$ .

$$\begin{aligned}\text{p-valor} &= P\left(\frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \leq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right) \\ &= P\left(Z \leq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right) = \Phi\left(\frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right)\end{aligned}$$

em que  $\Phi(z) = P(Z \leq z)$  para  $Z \sim N(0, 1)$ .



## Teste de hipótese para duas médias ( $\sigma_1^2 \neq \sigma_2^2$ )

Se temos interesse em:  $H_0 : \mu_1 - \mu_2 = \Delta_0$  vs  $H_a : \mu_1 - \mu_2 \neq \Delta_0$ .

$$\begin{aligned}\text{p-valor} &= P\left(\frac{|\bar{X} - \bar{Y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \geq \frac{|\bar{x} - \bar{y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right) \\ &= P\left(|Z| \geq \frac{|\bar{x} - \bar{y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right) = 2 \times \left[1 - \Phi\left(\frac{|\bar{x} - \bar{y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right)\right]\end{aligned}$$

em que  $\Phi(z) = P(Z \leq z)$  para  $Z \sim N(0, 1)$ .

# Teste de hipótese para duas médias ( $\sigma_1^2 = \sigma_2^2$ )

## Caso 2: Variâncias iguais e conhecidas

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$$

Temos interesse em:  $H_0: \mu_1 - \mu_2 = \Delta_0$  vs  $H_a: \mu_1 - \mu_2 \neq \Delta_0$  (ou  $H_a: \mu_1 - \mu_2 < \Delta_0$  ou  $H_a: \mu_1 - \mu_2 > \Delta_0$ ).

E daí, sob  $H_0$ , temos que:

$$Z = \frac{(\bar{X} - \bar{Y}) - \overbrace{(\mu_1 - \mu_2)}^{\Delta_0}}{\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}} \sim N(0, 1)$$

# Teste de hipótese para duas médias ( $\sigma_1^2 = \sigma_2^2$ desconhecidas)

## Caso 3: Variâncias iguais e desconhecidas

Assim como no caso de uma média com variância desconhecida, usamos uma estimativa de  $\sigma^2$  e a distribuição normal é substituída pela distribuição  $t$ .

No caso de duas populações, o estimador da variância  $\sigma^2$  é a combinação das variâncias amostrais de cada população, ou seja,

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2},$$

sendo  $S_i^2$  é a variância amostral da população  $i$ .

## Teste de hipótese para duas médias ( $\sigma_1^2 = \sigma_2^2$ desconhecidas)

Quando  $\sigma^2$  é conhecida:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n + 1/m)}} \sim N(0, 1)$$

Quando  $\sigma^2$  é desconhecida:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n + 1/m)}} \sim t_{n+m-2}$$

## Teste de hipótese para duas médias ( $\sigma_1^2 = \sigma_2^2$ desconhecidas)

Temos interesse em:  $H_0: \mu_1 - \mu_2 = \Delta_0$  vs  $H_a: \mu_1 - \mu_2 \neq \Delta_0$  (ou  $H_a: \mu_1 - \mu_2 < \Delta_0$  ou  $H_a: \mu_1 - \mu_2 > \Delta_0$ ).

E daí, sob  $H_0$ , temos que:

$$T = \frac{(\bar{X} - \bar{Y}) - \overbrace{(\mu_1 - \mu_2)}^{\Delta_0}}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}} \sim t_{n+m-2}$$

**Observação:** Se  $n$  e  $m$  são pequenos, as duas amostras devem vir de populações aproximadamente normais. Se  $n$  e  $m$  são grandes, então a distribuição  $t$  com  $n + m - 2$  graus de liberdade aproxima-se de uma normal.

# Resumo: Teste de hipótese para duas médias

Para  $H_0: \mu_1 - \mu_2 = \Delta_0$  vs  $H_1: \mu_1 - \mu_2 \neq \Delta_0$

Variâncias	Estatística do teste	Valor crítico para $\alpha$	Valor de p
Diferentes e conhecidas ( $\sigma_1^2 \neq \sigma_2^2$ )	$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$	rejeitar se $ z_{obs}  \geq z_{\alpha/2}$	$2P(Z \geq  z_{obs} )$
Iguais e conhecidas ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )	$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}} \sim N(0, 1)$	rejeitar se $ z_{obs}  \geq z_{\alpha/2}$	$2P(Z \geq  z_{obs} )$
Iguais e desconhecidas ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )	$T = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{S_p^2(1/m + 1/n)}} \sim t_{n+m-2}$	rejeitar se $ t_{obs}  \geq t_{n+m-2, \alpha/2}$	$2P(T \geq  t_{obs} )$

# Resumo: Teste de hipótese para duas médias

Para  $H_0: \mu_1 - \mu_2 = \Delta_0$  vs  $H_1: \mu_1 - \mu_2 \leq \Delta_0$

Variâncias	Estatística do teste	Valor crítico para $\alpha$	Valor de p
Diferentes e conhecidas ( $\sigma_1^2 \neq \sigma_2^2$ )	$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$	rejeitar se $z_{obs} \leq -z_\alpha$	$P(Z \leq z_{obs})$
Iguais e conhecidas ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )	$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}} \sim N(0, 1)$	rejeitar se $z_{obs} \leq -z_\alpha$	$P(Z \leq z_{obs})$
Iguais e desconhecidas ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )	$T \sim \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{S_p^2(1/m + 1/n)}} \sim t_{n+m-2}$	rejeitar se $t_{obs} \leq -t_{n+m-2, \alpha}$	$P(T \leq t_{obs})$

# Resumo: Teste de hipótese para duas médias

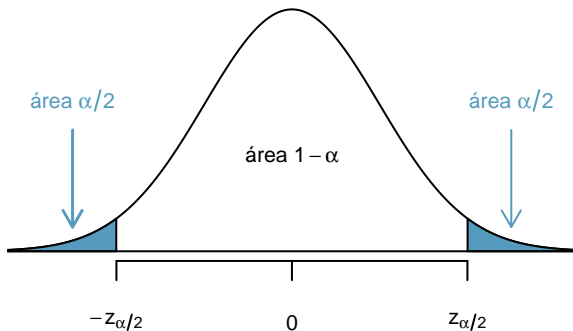
Para  $H_0: \mu_1 - \mu_2 = \Delta_0$  vs  $H_1: \mu_1 - \mu_2 \geq \Delta_0$

Variâncias	Estatística do teste	Valor crítico para $\alpha$	Valor de p
Diferentes e conhecidas ( $\sigma_1^2 \neq \sigma_2^2$ )	$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$	rejeitar se $z_{obs} \geq z_\alpha$	$P(Z \geq z_{obs})$
Iguais e conhecidas ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )	$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}} \sim N(0, 1)$	rejeitar se $z_{obs} \geq z_\alpha$	$P(Z \geq z_{obs})$
Iguais e desconhecidas ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )	$T \sim \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{S_p^2(1/m + 1/n)}} \sim t_{n+m-2}$	rejeitar se $t_{obs} \geq t_{n+m-2, \alpha}$	$P(T \geq t_{obs})$



## Relembrando: Como encontrar $z_{\alpha/2}$

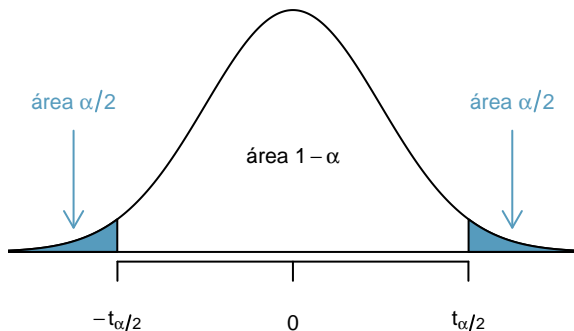
$$P(|Z| \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



Procure na tabela o valor de  $z$  tal que a probabilidade acumulada até o valor de  $z$ , isto é  $P(Z \leq z) = \Phi(z)$ , seja  $1 - \alpha/2$ .

## Relembrando: Como encontrar $t_{\nu, \alpha/2}$

$$P(-t_{\nu, \alpha/2} < T < t_{\nu, \alpha/2}) = 1 - \alpha$$



Nesse caso,  $\nu = n + m - 2$  e os valores da distribuição  $t$  encontram-se tabelados.

## Exemplo: tempo de incubação de dois vírus

O tempo de incubação do vírus 1 segue uma distribuição normal com média  $\mu_1$  e desvio padrão  $\sigma_1 = \sqrt{2}$ .

Por outro lado, o tempo de incubação do vírus 2 segue uma distribuição normal com média  $\mu_2$  e desvio padrão  $\sigma_2 = 1$ .

Os tempos de incubação de ambos os vírus são considerados independentes.

Afirma-se que em média, o tempo de incubação do vírus 1 é 3 meses depois do tempo médio de incubação do vírus 2.

# Exemplo: tempo de incubação de dois vírus

Realizaram um estudo de controle e os tempos de incubação registrados foram (tempo em meses):

- X: tempo de incubação do vírus 1 (20 observações)

```
[1] 4.56 3.72 3.45 2.86 4.03 4.08 6.56 4.31 0.42 5.56 5.92 2.65 4.54 4.04  
[15] 4.23 6.24 6.16 5.46 3.22 2.28
```

- Y: tempo de incubação do vírus 2 (22 observações)

```
[1] 2.44 1.49 2.68 2.60 1.51 1.60 1.47 3.70 2.22 1.78 2.36 1.56 2.98 3.33  
[15] 2.22 0.58 2.26 2.26 1.92 0.50 1.17 1.70
```

## Exemplo: tempo de incubação de dois vírus

Recentemente, pacientes contaminados com os vírus foram avaliados e suspeita-se que talvez o tempo de incubação do vírus 1 não seja 3 meses depois do tempo médio de incubação do vírus 2.

Definindo as hipóteses as serem testadas:

$$H_0 : \mu_1 - \mu_2 = 3 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq 3$$

Os dados coletados serão usados para avaliar se temos ou não evidências contra  $H_0$ .

Vamos calcular a média amostral das duas populações:  $\bar{x} = 4.21$  e  $\bar{y} = 2.02$ .

Pelo enunciado, as duas populações são normais e as variâncias são conhecidas:  $\sigma_1^2 = 2$  e  $\sigma_2^2 = 1$ . Veja que as populações são normais, variâncias diferentes mas conhecidas. Além disso,  $n = 20$  e  $m = 22$ .

## Exemplo: tempo de incubação de dois vírus

$$\begin{aligned}\text{p-valor} &= P\left(\frac{|\bar{X} - \bar{Y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \geq \frac{|4.21 - 2.02 - 3|}{\sqrt{\frac{2}{22} + \frac{1}{20}}}\right) \\ &= P(|Z| \geq 2.12) = 2 \times P(Z \geq 2.12) \\ &= 2 \times [1 - \Phi(2.12)] = 2 \times [1 - 0.983] = 0.034\end{aligned}$$

Se  $\alpha = 0.01$ , como  $\text{p-valor} = 0.034 > \alpha = 0.01$ , não temos evidência para rejeitar  $H_0 : \mu_1 = 3 + \mu_2$  com nível de significância 0.01.

Valor crítico para  $\alpha = 0.01$ : 2.58, ou seja, se  $|z_{obs}| \geq 2.58$  temos evidências para rejeitar  $H_0$  com nível de significância  $\alpha = 0.01$ .

## Exemplo: Tecidos

Dois tipos diferentes de tecido devem ser comparados. Uma máquina de testes *Martindale* pode comparar duas amostras ao mesmo tempo. O peso (em miligramas) para sete experimentos foram:

Tecido	1	2	3	4	5	6	7
A	36	26	31	38	28	20	37
B	39	27	35	42	31	39	22

Construa um teste de hipótese com nível de significância 5% para testar a hipótese nula de igualdade entre os pesos médios dos tecidos. Admita que a variância é a mesma, e igual a 49.

Quais outras suposições são necessárias para que o teste seja válido?

*Adaptado de: Profa. Nancy Garcia, Notas de aula.*

## Exemplo: Tecidos

Os tecidos do tipo A tem uma média amostral igual a  $\bar{x}_A = 30.86$ . Já os tecidos do tipo B têm média amostral de  $\bar{x}_B = 33.57$ .

A variância populacional é igual a 49, enquanto as variâncias amostrais são 44.14 e 52.62, respectivamente.

**Suposições:** Como os tamanhos amostrais  $n = m = 7$  são pequenos, devemos assumir os pesos dos tecidos dos dois tipos são normalmente distribuídos ou seja,  $X_A \sim N(\mu_A, \sigma^2)$  e  $X_B \sim N(\mu_B, \sigma^2)$ . Além disso são independentes e com variâncias iguais.



## Exemplo: Tecidos

Assumimos que as variâncias são iguais e **conhecidas** ( $\sigma_1^2 = \sigma_2^2 = 49$ ).

Além disso,  $n = 7$  e  $m = 7$ .

Definindo as hipóteses as serem testadas:

$$H_0 : \mu_A - \mu_B = 0 \quad vs \quad H_1 : \mu_A - \mu_B \neq 0.$$

Como a variância é conhecida, a estatística do teste é dada por

$$Z = \frac{\bar{X}_A - \bar{X}_B - \Delta_0}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Se a hipótese nula é verdadeira, temos que  $\Delta_0 = \mu_A - \mu_B = 0$  e  $Z \sim N(0, 1)$ . Note que a hipótese alternativa é do tipo  $\neq$ , então o teste é bilateral.

## Exemplo: Tecidos

$$\begin{aligned}\text{p-valor} &= P \left( \frac{|\bar{X}_A - \bar{X}_B - \Delta_0|}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \geq \frac{|30.86 - 33.57 - 0|}{7 \sqrt{\frac{1}{7} + \frac{1}{7}}} \right) \\ &= P(|Z| \geq 0.72) = 2 \times P(Z \geq 0.72) \\ &= 2 \times [1 - \Phi(0.72)] = 2 \times [1 - 0.7642] = 0.4716\end{aligned}$$

Se  $\alpha = 0.05$ , como  $\text{p-valor} = 0.4716 > \alpha = 0.05$ , não temos evidência para rejeitar  $H_0 : \mu_A = \mu_B$  com nível de significância 0.05.

Valor crítico para  $\alpha = 0.05$ : 1.96, ou seja, se  $|z_{obs}| \geq 1.96$  temos evidências para rejeitar  $H_0$  com nível de significância  $\alpha = 0.05$ .

## Exemplo: Tecidos

Vamos assumir agora que a variância populacional não fosse conhecida.

Assumindo ainda que as variâncias são iguais mas **desconhecidas**, vamos então estimar a variância amostral combinada.

Sabendo que  $s_1^2 = 44.14$ ,  $s_2^2 = 52.62$  e  $n = m = 7$  temos:

$$\begin{aligned}s_p^2 &= \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2} \\ &= \frac{(7-1)44.14 + (7-1)52.62}{7+7-2} \\ &= 48.38\end{aligned}$$

## Exemplo: Tecidos

Nesse caso, a estatística do teste, sob  $H_0$ , é dada por:

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_p^2 \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}} \sim t_{n+m-2}$$
$$t_{obs} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{s_p^2(1/n_A + 1/n_B)}}$$
$$= \frac{30.86 - 33.57}{\sqrt{48.38(1/7 + 1/7)}} = -0.73$$

## Exemplo: Tecidos

Considerando nível de significância 0.05, rejeitamos  $H_0$  se  $|t_{obs}| \geq t_{n+m-2, 0.025}$ .

Valor crítico para  $\alpha = 0.05$ : 2.18, ou seja, se  $|t_{obs}| \geq 2.18$  temos evidências para rejeitar  $H_0$  com nível de significância  $\alpha = 0.05$ .

No caso,  $|t_{obs}| = 0.73 < 2.18$ , portanto não encontramos evidências para rejeitar a hipótese de que as médias são iguais.

## Exemplo: tempo de adaptação

Num estudo comparativo do tempo médio de adaptação (em anos), uma amostra aleatória, de 50 homens e 50 mulheres de um grande complexo industrial, produziu os seguintes resultados:

Estatística	Homens	Mulheres
Média	3.2	3.7
Desvio Padrão	0.8	0.9

Construa um teste de hipótese com nível de significância de 5% para a diferença entre o tempo médio de adaptação para homens e mulheres.

*Fonte:* Adaptado de Morettin & Bussab, Estatística Básica 5ª edição, pág 365.

## Exemplo: tempo de adaptação

Veja que não sabemos a variância populacional, mas temos os desvios-padrão amostrais e estes são bem próximos. Então iremos assumir que as variâncias são iguais porém desconhecidas.

Nesse caso, vamos então estimar a variância amostral combinada.

Sabendo que  $s_1 = 0.8$ ,  $s_2 = 0.9$  e  $n = m = 50$  temos:

$$\begin{aligned}s_p^2 &= \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2} \\&= \frac{(50-1)(0.8)^2 + (50-1)(0.9)^2}{50+50-2} \\&= 0.73\end{aligned}$$

## Exemplo: tempo de adaptação

Nesse caso, a estatística do teste, sob  $H_0$ , é dada por:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}} \sim t_{n_A+m_B-2}$$

$$\begin{aligned} t_{obs} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2(\frac{1}{n} + \frac{1}{m})}} \\ &= \frac{3.2 - 3.7}{\sqrt{0.73(\frac{1}{50} + \frac{1}{50})}} = -2.93 \end{aligned}$$



## Exemplo: tempo de adaptação

Considerando nível de significância 0.05 e  $H_a: \mu_1 \neq \mu_2$ , rejeitamos  $H_0$  se  $|t_{obs}| \geq t_{n+m-2, 0.025} = 1.98$ .

Valor crítico para  $\alpha = 0.05$ : 1.98, ou seja, se  $|t_{obs}| \geq 1.98$  temos evidências para rejeitar  $H_0$  com nível de significância  $\alpha = 0.05$ .

No caso,  $|t_{obs}| = 2.93 > 1.98$ , portanto encontramos evidências para rejeitar a hipótese de que as médias são iguais.

## Inferência para duas populações: Teste de hipótese para duas proporções

# Teste de hipótese para duas proporções

Considere  $X_1, \dots, X_{n_1}$  e  $Y_1, \dots, Y_{n_2}$  duas amostras independentes de ensaios de Bernoulli tal que  $X \sim b(p_1)$  e  $Y \sim b(p_2)$ , com probabilidade  $p_1$  e  $p_2$  de apresentarem uma certa característica.

Queremos testar:  $H_0: p_1 - p_2 = 0$  vs  $H_a: p_1 - p_2 \neq 0$  (ou  $H_a: p_1 - p_2 < 0$  ou  $H_a: p_1 - p_2 > 0$ ).

Em aulas anteriores vimos que:

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right) \quad \text{e} \quad \hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

Como as variâncias de  $\hat{p}_1$  e  $\hat{p}_2$  dependem de  $p_1$  e  $p_2$  e, portanto, não são conhecidas, iremos usar uma estimativa dessas variâncias.

# Teste de hipótese para duas proporções

Sob  $H_0$ ,  $p_1 = p_2 = p$ , portanto:

$$\hat{p}_1 \sim N\left(p_1, \frac{p(1-p)}{n_1}\right) \quad \text{e} \quad \hat{p}_2 \sim N\left(p_2, \frac{p(1-p)}{n_2}\right)$$

No entanto,  $p$  é desconhecido. Iremos utilizar como estimativa para  $p$ :  $\hat{p}$ , definido como o número de sucessos entre todos os elementos amostrados. Ou seja, o estimador é a proporção de sucessos na amostra toda, sem levar em consideração as populações, pois, sob  $H_0$ ,  $p_1 = p_2$  (não há diferença entre as proporções das duas populações).

# Teste de hipótese para duas proporções

Então, para  $H_0: p_1 = p_2$  usamos a estatística do teste a seguir:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1)$$

em que  $\hat{p}$  é a proporção de sucessos entre os  $n_1 + n_2$  elementos amostrados.

**Condições:** Todas as quantidades

$n_1\hat{p}_1$ ,  $n_1(1 - \hat{p}_1)$ ,  $n_2\hat{p}_2$  e  $n_2(1 - \hat{p}_2)$  devem ser pelo menos igual a 10 para que a aproximação pela normal seja válida.

# Teste de hipótese para duas proporções

Resumindo:

Para  $H_0: p_1 - p_2 = 0$

$H_a$	Valor crítico para $\alpha$	Valor de p
$p_1 - p_2 \neq 0$	rejeitar se $ z_{obs}  \geq z_{\alpha/2}$	$2P(Z \geq  z_{obs} )$
$p_1 - p_2 < 0$	rejeitar se $z_{obs} \leq -z_{\alpha}$	$P(Z \leq z_{obs})$
$p_1 - p_2 > 0$	rejeitar se $z_{obs} \geq z_{\alpha}$	$P(Z \geq z_{obs})$

## Exemplo: decisão sobre gastos

O dinheiro que não é gasto hoje pode ser gasto depois.

Será que ao lembrar o aluno deste fato faz com que tome a decisão sobre uma compra de maneira diferente?

## Exemplo: decisão sobre gastos

O dinheiro que não é gasto hoje pode ser gasto depois.

Será que ao lembrar o aluno deste fato faz com que tome a decisão sobre uma compra de maneira diferente?

O cético pode pensar que lembrar não irá influenciar na decisão.



## Exemplo: decisão sobre gastos

O dinheiro que não é gasto hoje pode ser gasto depois.

Será que ao lembrar o aluno deste fato faz com que tome a decisão sobre uma compra de maneira diferente?

O cético pode pensar que lembrar não irá influenciar na decisão.

Podemos utilizar um teste de hipótese:

## Exemplo: decisão sobre gastos

O dinheiro que não é gasto hoje pode ser gasto depois.

Será que ao relembrar o aluno deste fato faz com que tome a decisão sobre uma compra de maneira diferente?

O cético pode pensar que relembrar não irá influenciar na decisão.

Podemos utilizar um teste de hipótese:

- $H_0$ : Relembrar o aluno de que ele pode poupar para comprar algo especial depois não irá influenciar na decisão de gasto do aluno.
- $H_a$ : Relembrar o aluno de que ele pode poupar para comprar algo especial depois irá aumentar a chance dele não gastar em algo no presente.

## Exemplo: decisão sobre gastos

Alunos de ME414 do segundo semestres de 2015 foram recrutados para um estudo e cada um recebeu a seguinte informação através do Google Forms:

*Imagine que você estivesse poupando para comprar algo especial. Em uma visita ao shopping você encontra um DVD da sua série/filme favorita que estava na sua “lista de desejos” há tempos. O DVD está em promoção, custando R\$ 20,00. O que você faria?*

56 alunos (grupo 1) selecionados ao acaso receberam a seguinte opção de resposta:

- Compraria o DVD.
- Não compraria o DVD.

54 alunos (grupo 2) selecionados ao acaso receberam a seguinte opção de resposta:

- Compraria o DVD.
- Não compraria o DVD. Pouparia os R\$ 20,00 para algo especial.

Obs: estudo adaptado do artigo *Frederick S, Novemsky N, Wang J, Dhar R, Nowlis S. 2009. Opportunity Cost Neglect. Journal of Consumer Research 36: 553-561.*

## Exemplo: decisão sobre gastos

	Compraria	Não compraria
grupo1	31	25
grupo2	29	25

Entre os alunos do grupo 1, a proporção que decide não comprar foi 0.45.

Entre os alunos do grupo 2, a proporção que decide não comprar foi 0.46.

## Exemplo: decisão sobre gastos

	Compraria	Não compraria
grupo1	31	25
grupo2	29	25

Entre os alunos do grupo 1, a proporção que decide não comprar foi 0.45.

Entre os alunos do grupo 2, a proporção que decide não comprar foi 0.46.

Temos evidências contra a hipótese nula, ou seja, relembrar o aluno não influencia na decisão?

## Exemplo: decisão sobre gastos

Para realizar o teste de hipótese, devemos fazer algumas suposições.

Considere duas populações:  $X$  e  $Y$  tal que:

- $X_i \sim b(p_1)$  indica se o  $i$ -ésimo aluno do **grupo 1** decide não comprar o DVD e  $p_1$  é a probabilidade de decidir por não comprar.
- $Y_i \sim b(p_2)$  indica se o  $i$ -ésimo aluno do **grupo 2** decide não comprar o DVD e  $p_2$  é a probabilidade de decidir por não comprar.

Queremos testar:

$$H_0: p_1 = p_2$$

$$H_a: p_1 < p_2$$

## Exemplo: decisão sobre gastos

Seja  $\hat{p}_1$  a proporção que decide não comprar entre os alunos  $n_1$  amostrados do grupo 1.

Seja  $\hat{p}_2$  a proporção que decide não comprar entre os  $n_2$  alunos amostrados do grupo 2.



## Exemplo: decisão sobre gastos

Seja  $\hat{p}_1$  a proporção que decide não comprar entre os alunos  $n_1$  amostrados do grupo 1.

Seja  $\hat{p}_2$  a proporção que decide não comprar entre os  $n_2$  alunos amostrados do grupo 2.

Relembrando o TCL:

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right) \quad \text{e} \quad \hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

**Condições:** Todas as quantidades

$n_1\hat{p}_1$ ,  $n_1(1-\hat{p}_1)$ ,  $n_2\hat{p}_2$  e  $n_2(1-\hat{p}_2)$  devem ser pelo menos igual a 10 para que a aproximação pela normal seja válida.

Então, para  $H_0: p_1 = p_2$  usamos a estatística do teste a seguir:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1)$$

em que  $\hat{p}$  é a proporção que decide não comprar entre os  $n_1 + n_2$  alunos amostrados.

## Exemplo: decisão sobre gastos

$H_0: p_1 = p_2$  vs  $H_a: p_1 < p_2$ , que é equivalente a testar:  $H_0: p_1 - p_2 = 0$  vs  $H_a: p_1 - p_2 < 0$ .

$$\begin{aligned} \text{p-valor} &= P \left( \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq \frac{25/56 - 25/54}{\sqrt{5/11(1-5/11)(\frac{1}{56} + \frac{1}{54})}} \right) \\ &= P(Z \leq -0.17) = 0.4325 \end{aligned}$$

Se  $\alpha = 0.05$ , como  $\text{p-valor} = 0.4325 > \alpha = 0.05$ , não temos evidência para rejeitar  $H_0$  com nível de significância 0.05.

Valor crítico para  $\alpha = 0.05$ : -1.64, ou seja, se  $z_{obs} \leq -1.64$  temos evidências para rejeitar  $H_0$  com nível de significância  $\alpha = 0.05$ .

- [Ross](#): seções 10.1, 10.2, 10.3, 10.4 e 10.6.
- [OpenIntro](#): seções 3.2 e 4.3.
- Magalhães: capítulo 9.