ME414 - Estatística para Experimentalistas

Parte 20

Inferência para duas populações:
Teste de hipótese para duas médias

Teste de hipótese para duas médias

População 1: Coletamos uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com média μ_1 e a variância σ_1^2 e usamos $\bar{X}$ para estimar μ_1.

População 2: Coletamos uma amostra aleatória $Y_1, Y_2, \ldots, Y_m$ de uma população com média μ_2 e a variância σ_2^2 e usamos $\bar{Y}$ para estimar μ_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 $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ 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 testar as hipóteses:

$$H_0 : \mu_1 - \mu_2 = \Delta_0 \qquad \text{vs} \qquad \begin{array}{l} H_A : \mu_1 - \mu_2 \neq \Delta_0 \text{ ou} \\ H_A : \mu_1 - \mu_2 < \Delta_0 \text{ ou} \\ H_A : \mu_1 - \mu_2 > \Delta_0. \end{array}$$

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 ($\sigma_1^2 \neq \sigma_2^2$)

Definidas as hipóteses, coletamos as informações das duas populações.

Para a população X: uma amostra aleatória de tamanho n é coletada e calcula-se a média amostral $\bar{x}$.

Para a população Y: similarmente, uma amostra aleatória de tamanho m é coletada e calcula-se a média amostral $\bar{y}$.

Calcula-se a estatística do teste:

$$Z = \frac{(\bar{X} - \bar{Y}) - \overbrace{(\mu_1 - \mu_2)}^{\Delta_0}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \qquad \Longrightarrow \qquad z_{obs} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

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

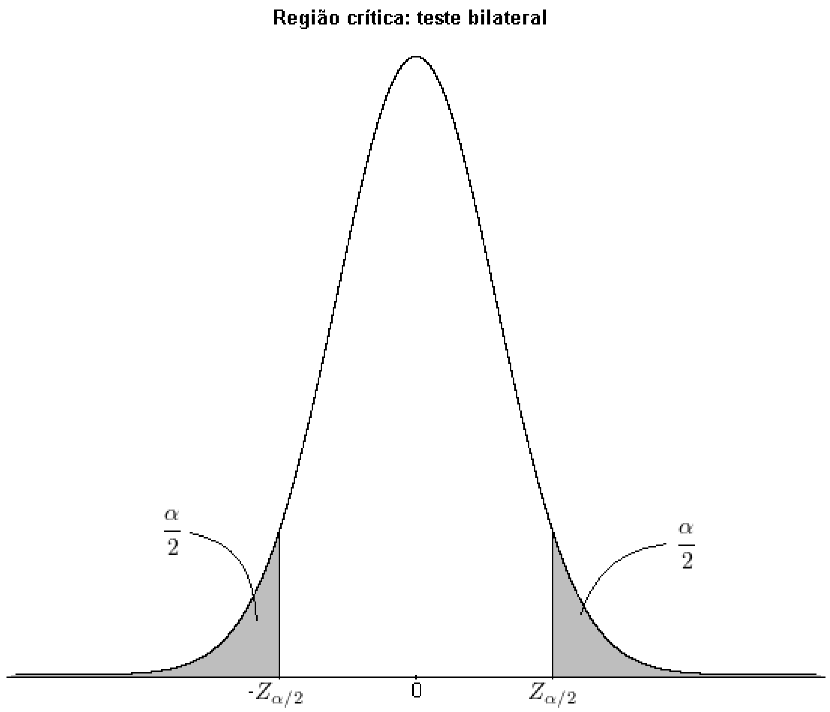
O p-valor é calculado de acordo com a hipótese alternativa.

Se o teste é **bilateral**, ou seja,

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$
$$H_A : \mu_1 - \mu_2 \neq \Delta_0$$

Usando o valor observado da estatística do teste:

$$\begin{aligned}\text{p-valor} &= P(|Z| \geq |z_{obs}|) \\ &= 2P(Z \leq -|z_{obs}|)\end{aligned}$$



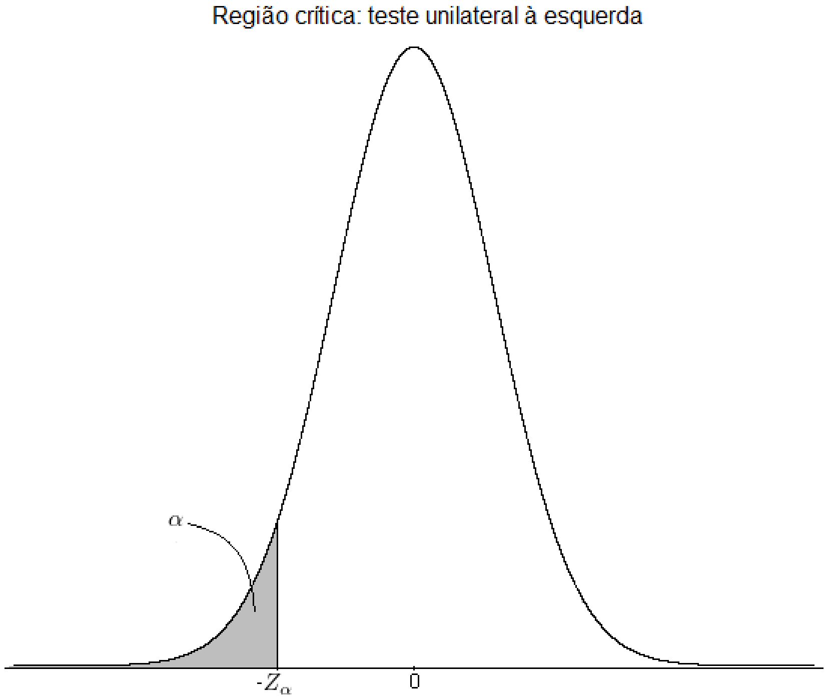
Conclusão: Rejeita-se H_0 se p-valor $\leq \alpha$ ou, de forma equivalente, se z_{obs} cai na região crítica (área cinza do gráfico).

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

O p-valor é calculado de acordo com a hipótese alternativa.

Se o teste é **unilateral à esquerda**, ou seja,

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$
$$H_A : \mu_1 - \mu_2 < \Delta_0$$



Usando o valor observado da estatística do teste:

$$\text{p-valor} = P(Z \leq z_{obs})$$

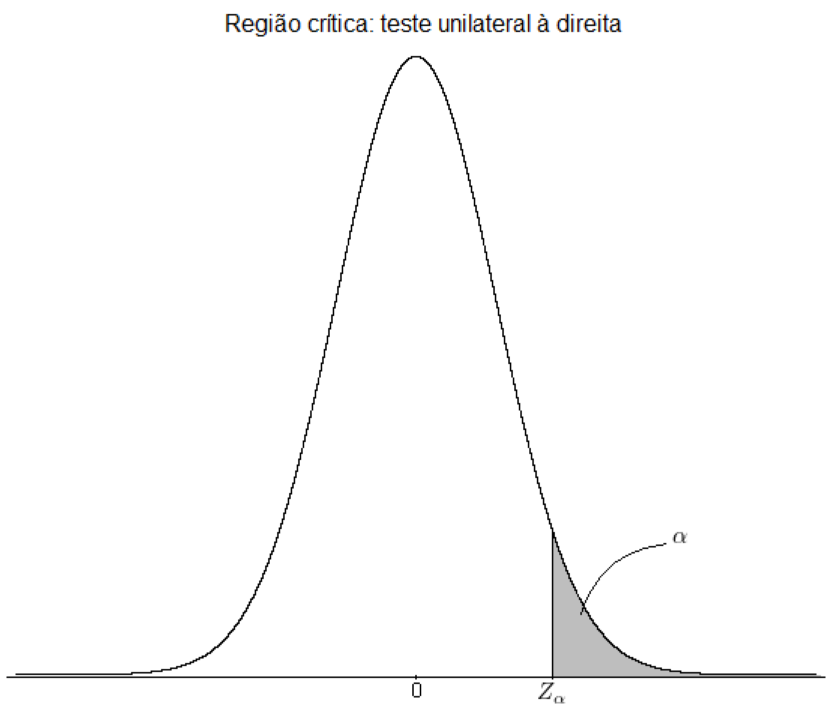
Conclusão: Rejeita-se H_0 se p-valor $\leq \alpha$
ou, de forma equivalente, se z_{obs} cai na região crítica (área cinza à esquerda do gráfico).

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

O p-valor é calculado de acordo com a hipótese alternativa.

Se o teste é **unilateral à direita**, ou seja,

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$
$$H_A : \mu_1 - \mu_2 > \Delta_0$$



Usando o valor observado da estatística do teste:

$$\text{p-valor} = P(Z \geq z_{obs})$$

Conclusão: Rejeita-se H_0 se p-valor $\leq \alpha$ ou, de forma equivalente, se z_{obs} cai na região crítica (área cinza à direita do gráfico).

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

Caso 2: Variâncias iguais e conhecidas

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

As hipóteses são as mesmas que as testadas no caso 1.

E daí, sob H_0, temos que:

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

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

Caso 3: Variâncias iguais e desconhecidas

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

No caso de duas populações, o estimador da variância σ^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

As variâncias são iguais $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Quando σ^2 é conhecida:

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

Quando σ^2 é desconhecida:

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

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

Temos interesse em testar:
$H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_A : \mu_1 - \mu_2 \neq \Delta_0 \quad \text{(ou hipóteses unilaterais)}$

E daí, sob H_0, temos que:

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

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.

Esse teste é conhecido como **teste t** para amostras independentes.

Resumo: Teste de hipótese para duas médias

Para $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_A : \mu_1 - \mu_2 \neq \Delta_0$

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

Resumo: Teste de hipótese para duas médias

Para $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_A : \mu_1 - \mu_2 < \Delta_0$

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

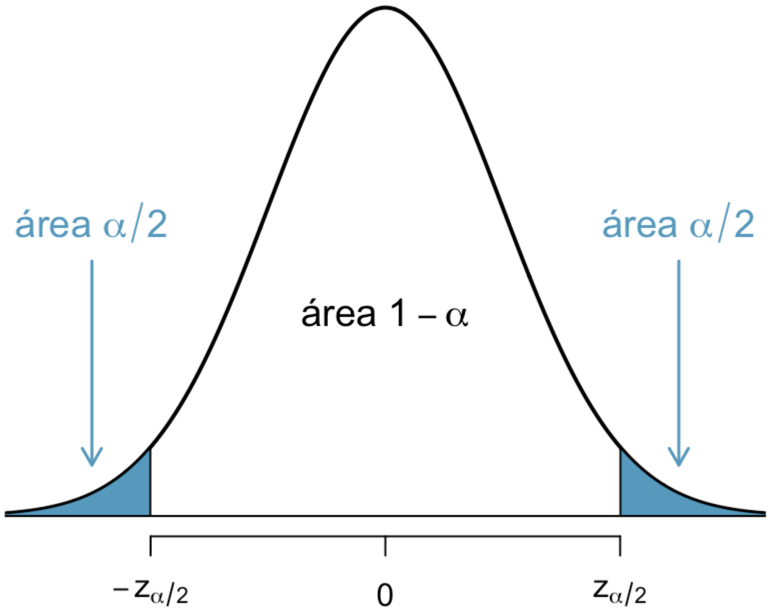
Resumo: Teste de hipótese para duas médias

Para $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad \text{vs} \quad H_A : \mu_1 - \mu_2 > \Delta_0$

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

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

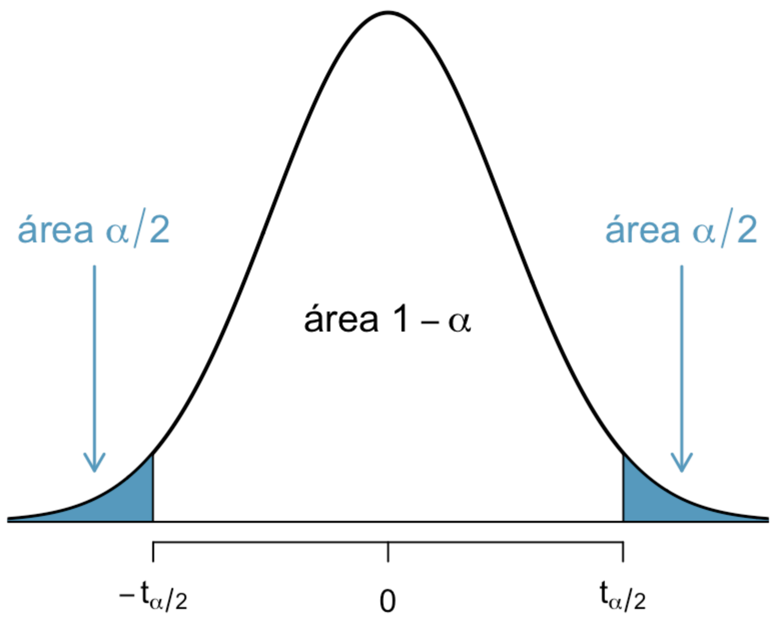
$$P(|Z| \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



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

O tempo de incubação do vírus 1 segue uma distribuição normal com média μ_1 e desvio padrão $\sigma_1 = \sqrt{2}$.

Por outro lado, o tempo de incubação do vírus 2 segue uma distribuição normal com média μ_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

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

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

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	4.23	6.24	6.16	5.46	3.22	2.28

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

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	2.22	0.58	2.26	2.26	1.92	0.50	1.17	1.70

Exemplo: Tempo de Incubação

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:

$$H_0 : \mu_1 - \mu_2 = 3 \qquad \text{vs} \qquad H_A : \mu_1 - \mu_2 \neq 3$$

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

Cálculo da estatística do teste:

$$z_{obs} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{(4.21 - 2.02) - 3}{\sqrt{\frac{2}{20} + \frac{1}{22}}} = -2.12$$

Como é um teste bilateral, o p-valor é dado por:

$$\text{p-valor} = P(|Z| \geq |z_{obs}|) = P(|Z| \geq 2.12) = 2P(Z \leq -2.12) = 0.034$$

Para um nível de significância $\alpha = 0.01$: Como p-valor = 0.034 $> \alpha = 0.01$, não temos evidência para rejeitar $H_0 : \mu_1 = 3 + \mu_2$.

Valor crítico para $\alpha = 0.01$: $z_{0.005} = 2.58$, ou seja, rejeita-se H_0 se $z_{obs} \geq 2.58$ ou $z_{obs} \leq -2.58$.

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?
Fonte: Adaptado das Notas de aula, Profa. Nancy Garcia.

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 a serem testadas:
$H_0 : \mu_A - \mu_B = 0 \qquad \text{vs} \qquad H_A : \mu_A - \mu_B \neq 0$

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

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \Delta_0}{\sqrt{\sigma^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

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

Cálculo da estatística do teste:

$$z_{obs} = \frac{(\bar{x}_A - \bar{x}_B) - \Delta_0}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} = \frac{(30.86 - 33.57) - 0}{\sqrt{49 \left(\frac{1}{7} + \frac{1}{7}\right)}} = -0.72$$

Como é um teste bilateral, o p-valor é dado por:

$$\text{p-valor} = P(|Z| \geq |z_{obs}|) = P(|Z| \geq 0.72) = 2P(Z \leq -0.72) = 0.4716$$

Para um nível de significância $\alpha = 0.05$: Como p-valor = $0.4716 > \alpha = 0.05$, não temos evidência para rejeitar $H_0 : \mu_A = \mu_B$.

Valor crítico para $\alpha = 0.05$: $z_{0.025} = 1.96$, ou seja, rejeita-se H_0 se $|z_{obs}| \geq 1.96$.

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:

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \sim t_{n+m-2}$$

Então,

$$t_{obs} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{S_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{\mathbf{30.86 - 33.57}}{\sqrt{\mathbf{48.38}\left(\frac{1}{7} + \frac{1}{7}\right)}} = -\mathbf{0.73}$$

Para um nível de significância $\alpha = 0.05$, rejeitamos H_0 se $|t_{obs}| \geq t_{n+m-2,0.025}$. No caso, $|t_{obs}| = 0.73 < 2.18 = t_{12,0.025}$. Portanto, não temos evidências para rejeitar a hipótese de que as médias dos dois tecidos são iguais.

Inferência para duas populações: Teste de hipótese para duas proporções

Teste de hipótese para duas proporções

Considere $X_1, \ldots, X_{n_1}$ e $Y_1, \ldots, Y_{n_2}$ 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.

Temos interesse em testar as hipóteses:

$$
H_0: p_1 - p_2 = 0 \qquad \text{vs} \qquad \begin{array}{l} H_A: p_1 - p_2 \neq 0 \text{ ou} \\ H_A: p_1 - p_2 < 0 \text{ ou} \\ H_A: p_1 - p_2 > 0 \end{array}
$$

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, a proporção de sucessos na amostra toda ($\widehat{p}$), sem levar em consideração as populações, ou seja,

$$\widehat{p} = \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2}.$$

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

Teste de hipótese para duas proporções

Para testar $H_0 : p_1 - p_2 = 0$, calcula-se a estatística do teste

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

O p-valor e a conclusão se dá de acordo com a hipótese alternativa na tabela:

Hipótese Alternativa	Valor crítico para α	Valor de p
$H_A : p_1 - p_2 \neq 0$	rejeitar H_0 se $\mid z_{obs} \mid \geq z_{\alpha/2}$	$2P(Z \geq \mid z_{obs} \mid)$
$H_A : p_1 - p_2 < 0$	rejeitar H_0 se $z_{obs} \leq -z_\alpha$	$P(Z \leq z_{obs})$
$H_A : p_1 - p_2 > 0$	rejeitar H_0 se $z_{obs} \geq z_\alpha$	$P(Z \geq z_{obs})$

IMECC

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 uma turma de ME414 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	Total
Grupo1	31	25	56
Grupo2	29	25	54

Entre os alunos do Grupo 1, a proporção que decide não comprar foi

$$\widehat{p}_1 = 25/56 = 0.45$$

Entre os alunos do Grupo 2, a proporção que decide não comprar foi

$$\widehat{p}_2 = 25/54 = 0.46$$

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:

$$H_0 : p_1 = p_2 \qquad \text{vs} \qquad H_A : p_1 < p_2$$

Exemplo: decisão sobre gastos

Seja $\hat{p}_1$ a proporção que decide não comprar entre os n_1 alunos **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 $n_1\hat{p}_1,\ n_1(1-\hat{p}_1),\ n_2\hat{p}_2\ \text{ e }\ n_2(1-\hat{p}_2)$ devem ser pelo menos igual a 10 para que a aproximação pela normal seja válida.

Exemplo: decisão sobre gastos

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:

$$\widehat{p} = \frac{25+25}{56+54} = \frac{50}{110} = 0.45$$

Exemplo: decisão sobre gastos

$$\begin{cases} H_0 : p_1 = p_2 \\ H_A : p_1 < p_2 \end{cases} \iff \begin{cases} H_0 : p_1 - p_2 = 0 \\ H_A : p_1 - p_2 < 0 \end{cases}$$

Calculando a estatística do teste:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.45 - 0.46}{\sqrt{(0.45)(0.55)\left(\frac{1}{56} + \frac{1}{54}\right)}} = -0.11 = z_{obs}$$

Para um nível de significância $\alpha = 0.05$, rejeitamos H_0 se $z_{obs} \leq -z_{0.05}$. No caso, $z_{obs} = -0.11 > -1.64 = -z_{0.05}$. Portanto, não temos evidências para rejeitar a hipótese de que as duas proporções são iguais.

p-valor = $P(Z \leq z_{obs}) = P(Z \leq -0.11) = 0.4562 > \alpha$. Portanto, não rejeitamos H_0.

Leituras

- Ross: seções 10.1, 10.2, 10.3, 10.4 e 10.6.
- OpenIntro: seções 3.2 e 4.3.
- Magalhães: capítulo 9.

Slides produzidos pelos professores:

- Samara Kiihl
- Tatiana Benaglia
- Larissa Matos
- Benilton Carvalho