ME414 - Estatística para Experimentalistas

Teste de Hipóteses II

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 14

Teste de Hipóteses para a proporção

Diferença de médias para polulações não relacionadas

Diferença de médias para amostras relacionadas.

Teste de Hipóteses para a proporção

Teste de Hipóteses para a proporção

- Até agora temos visto testes de hipóteses para a média populacional μ.

Teste de Hipóteses para a proporção

- Até agora temos visto testes de hipóteses para a média populacional μ.
- Outro parâmetro populacional de interesse é a proporção p

Teste de Hipóteses para a proporção

- Até agora temos visto testes de hipóteses para a média populacional μ.
- Outro parâmetro populacional de interesse é a proporção p
- Seja p_0 o valor hipotético da proporção populacional, estamos interessados em testes da forma:

$$H_0 : p = p_0 \quad vs \quad H_1 : p \neq p_0,$$

$$H_0 : p \leq p_0 \quad vs \quad H_1 : p > p_0,$$

$$H_0 : p \geq p_0 \quad vs \quad H_1 : p < p_0.$$

Teste de Hipóteses para a proporção

- Até agora temos visto testes de hipóteses para a média populacional μ.
- Outro parâmetro populacional de interesse é a proporção p
- Seja p_0 o valor hipotético da proporção populacional, estamos interessados em testes da forma:

$$H_0 : p = p_0 \quad vs \quad H_1 : p \neq p_0,$$

$$H_0 : p \leq p_0 \quad vs \quad H_1 : p > p_0,$$

$$H_0 : p \geq p_0 \quad vs \quad H_1 : p < p_0.$$

- Assim como no caso do teste para a média, para o caso da proporção também precisamos de uma **estatística de teste**.

Teste de Hipóteses para a proporção

Sabemos que se $X_1, \ldots, X_n \sim N(\mu, \sigma)$, no teste para μ com σ conhecido a estatística de teste é da forma

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Teste de Hipóteses para a proporção

Sabemos que se $X_1, \ldots, X_n \sim N(\mu, \sigma)$, no teste para μ com σ conhecido a estatística de teste é da forma

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Sejam $X_1, \ldots, X_n \sim Bernoulli(p)$, embora os dados não tenham uma distribuição normal, pelo TCL temos que

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim_{approx} N(0,1)$$

Teste de Hipóteses para a proporção

No caso de $X_1, \ldots, X_n \sim Bernoulli(p)$, temos que $\bar{x} = \bar{p}$ e $\mu_0 = p_0$.

Teste de Hipóteses para a proporção

No caso de $X_1, \ldots, X_n \sim Bernoulli(p)$, temos que $\bar{x} = \bar{p}$ e $\mu_0 = p_0$.

Sob $H_0 : p = p_0$, temos que $\sigma = p_0(1 - p_0)$. Então, a estatística de teste é da forma

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \sim_{approx} N(0, 1)$$

Teste de Hipóteses para a proporção

No caso de $X_1, \ldots, X_n \sim Bernoulli(p)$, temos que $\bar{x} = \bar{p}$ e $\mu_0 = p_0$.

Sob $H_0 : p = p_0$, temos que $\sigma = p_0(1 - p_0)$. Então, a estatística de teste é da forma

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim_{approx} N(0, 1)$$

Com isso, podemos tester as hipóteses:

- $H_0 : p = p_0 \quad vs \quad H_1 : p \neq p_0,$
- $H_0 : p \leq p_0 \quad vs \quad H_1 : p > p_0,$
- $H_0 : p \geq p_0 \quad vs \quad H_1 : p < p_0.$

como usual.

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

- Por padrão assumimos $\alpha = 0.05$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

- Por padrão assumimos $\alpha = 0.05$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.175 - 0.2}{\sqrt{\dfrac{0.2(1-0.2)}{400}}} = -1.25$$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

- Por padrão assumimos $\alpha = 0.05$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.175 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{400}}} = -1.25$$

- Como $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

- Por padrão assumimos $\alpha = 0.05$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.175 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{400}}} = -1.25$$

- Como $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

- Por padrão assumimos $\alpha = 0.05$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.175 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{400}}} = -1.25$$

- Como $H_0 : p = 0.2 \quad vs. \quad H_1 : p \neq 0.2$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$

```
alpha = 0.05
qnorm(1-alpha/2)
```

```
## [1] 1.959964
```

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0: p = 0.2 \quad vs. \quad H_1: p \neq 0.2$. Uma amostra de tamanho 400 produziu $\bar{p} = 0.175$. **Rejeitamos H_0 ou não?**

- Por padrão assumimos $\alpha = 0.05$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.175 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{400}}} = -1.25$$

- Como $H_0: p = 0.2 \quad vs. \quad H_1: p \neq 0.2$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$

```
alpha = 0.05
qnorm(1-alpha/2)
```

```
## [1] 1.959964
```

$|$-1.25$| = 1.25 > 1.959964$ **?** Não, então não rejeitamos H_0

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

- $\alpha = 0.01$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

- $\alpha = 0.01$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.72 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -1.2$$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

- $\alpha = 0.01$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.72 - 0.75}{\sqrt{\dfrac{0.75(1-0.75)}{300}}} = -1.2$$

- Como $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$, rejeitamos H_0 se $z < z_\alpha$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

- $\alpha = 0.01$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{0.72 - 0.75}{\sqrt{\dfrac{0.75(1 - 0.75)}{300}}} = -1.2$$

- Como $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$, rejeitamos H_0 se $z < z_\alpha$

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

- $\alpha = 0.01$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.72 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -1.2$$

- Como $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$, rejeitamos H_0 se $z < z_\alpha$

```
alpha = 0.01
qnorm(alpha)
```

```
## [1] -2.326348
```

Teste de Hipóteses para a proporção

Considere o seguinte teste $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$. Uma amostra de tamanho 300 produziu $\bar{p} = 0.72$. Considerando um nível de significância de 1%, **rejeitamos H_0 ou não?**

- $\alpha = 0.01$
- Definos a estatística de teste

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.72 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -1.2$$

- Como $H_0 : p \geq 0.75 \quad vs. \quad H_1 : p < 0.75$, rejeitamos H_0 se $z < z_\alpha$

```
alpha = 0.01
qnorm(alpha)
```

```
## [1] -2.326348
```

-1.2 < -2.326348 **?** Não, então não rejeitamos H_0

Diferença de médias para polulações não relacionadas

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.
- As duas amostras são tomadas separada e independentemente de duas populações diferentes.

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.
- As duas amostras são tomadas separada e independentemente de duas populações diferentes.
- Calcularemos Intervalos de Confiança e fazeremos testes de hipóteses para $\mu_x - \mu_y$.

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.
- As duas amostras são tomadas separada e independentemente de duas populações diferentes.
- Calcularemos Intervalos de Confiança e fazeremos testes de hipóteses para $\mu_x - \mu_y$.

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.
- As duas amostras são tomadas separada e independentemente de duas populações diferentes.
- Calcularemos Intervalos de Confiança e fazeremos testes de hipóteses para $\mu_x - \mu_y$.

Como faremos isto?

- Selecionamos uma amostra de tamanho n_1 da população 1 e calculamos $\bar{x}$

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.
- As duas amostras são tomadas separada e independentemente de duas populações diferentes.
- Calcularemos Intervalos de Confiança e fazeremos testes de hipóteses para $\mu_x - \mu_y$.

Como faremos isto?

- Selecionamos uma amostra de tamanho n_1 da população 1 e calculamos $\bar{x}$
- Selecionamos uma amostra de tamanho n_2 da população 2 e calculamos $\bar{y}$

Duas populações: σ_x e σ_y conhecidos.

- Seja μ_x a média da população 1 e μ_y a média da população 2
- Estamos interessados em fazer inferência para a diferença $\mu_x - \mu_y$.
- As duas amostras são tomadas separada e independentemente de duas populações diferentes.
- Calcularemos Intervalos de Confiança e fazeremos testes de hipóteses para $\mu_x - \mu_y$.

Como faremos isto?

- Selecionamos uma amostra de tamanho n_1 da população 1 e calculamos $\bar{x}$
- Selecionamos uma amostra de tamanho n_2 da população 2 e calculamos $\bar{y}$
- Com isso, temos $\bar{x} - \bar{y}$ um estimador por ponto de $\mu_x - \mu_y$

Duas populações: σ_x e σ_y conhecidos.

- Sabemos que $\bar{x} \sim N(\mu_x, \sigma_{\bar{x}})$ e $\bar{y} \sim N(\mu_y, \sigma_{\bar{y}})$, então

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}\right)$$

Duas populações: σ_x e σ_y conhecidos.

- Sabemos que $\bar{x} \sim N(\mu_x, \sigma_{\bar{x}})$ e $\bar{y} \sim N(\mu_y, \sigma_{\bar{y}})$, então

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}\right)$$

- Padronizando,

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} \sim N(0, 1)$$

Duas populações: σ_x e σ_y conhecidos.

- Sabemos que $\bar{x} \sim N(\mu_x, \sigma_{\bar{x}})$ e $\bar{y} \sim N(\mu_y, \sigma_{\bar{y}})$, então

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}\right)$$

- Padronizando,

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} \sim N(0, 1)$$

Duas populações: σ_x e σ_y conhecidos.

- Sabemos que $\bar{x} \sim N(\mu_x, \sigma_{\bar{x}})$ e $\bar{y} \sim N(\mu_y, \sigma_{\bar{y}})$, então

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}\right)$$

- Padronizando,

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} \sim N(0, 1)$$

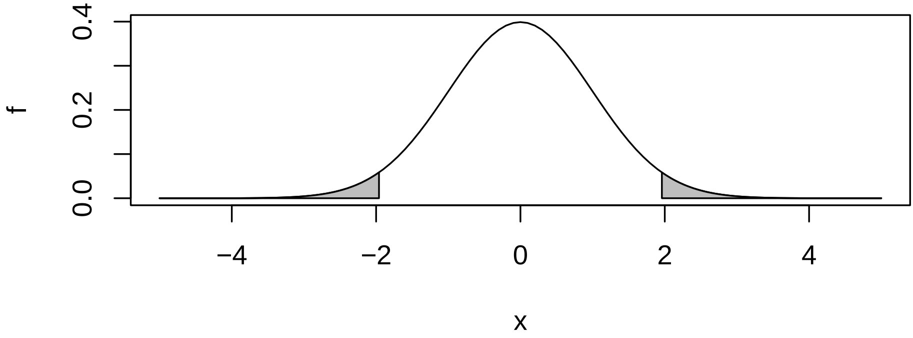
z nos ajudará tanto a construir intervalos de confiança quanto testes de hipóteses.

Duas populações: σ_x e σ_y conhecidos.

Se quisermos um intervalo de confiança $\delta = 1 - \alpha$ para $\mu_x - \mu_y$ faremos $P(|Z| < k) = 1 - \alpha$

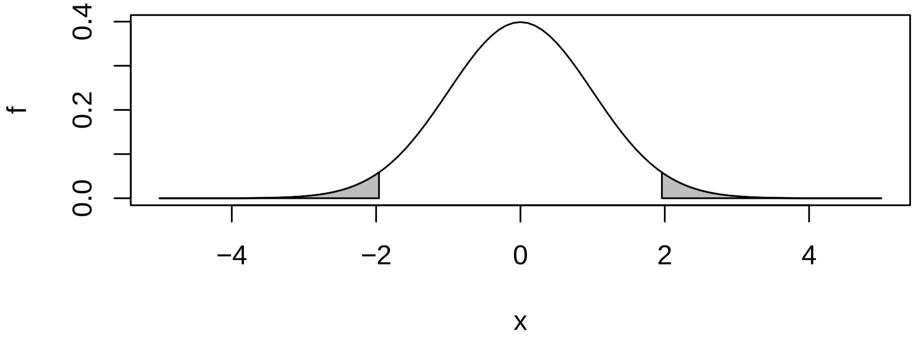
Duas populações: σ_x e σ_y conhecidos.

Se quisermos um intervalo de confiança $\delta = 1 - \alpha$ para $\mu_x - \mu_y$ faremos $P(|Z| < k) = 1 - \alpha$



Duas populações: σ_x e σ_y conhecidos.

Se quisermos um intervalo de confiança $\delta = 1 - \alpha$ para $\mu_x - \mu_y$ faremos $P(|Z| < k) = 1 - \alpha$



$$-z_{1-\alpha/2} \leq \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} \leq z_{1-\alpha/2}$$

Duas populações: σ_x e σ_y conhecidos.

$$-z_{1-\alpha/2} \leq \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\sigma_x^2/n_1+\sigma_y^2/n_2}} \leq z_{1-\alpha/2}$$

Duas populações: σ_x e σ_y conhecidos.

$$-z_{1-\alpha/2} \leq \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} \leq z_{1-\alpha/2}$$

$$(\bar{x}-\bar{y})-z_{1-\alpha/2}\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2} \leq \mu_x-\mu_y \leq (\bar{x}-\bar{y})+z_{1-\alpha/2}\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}$$

Duas populações: σ_x e σ_y conhecidos.

$$-z_{1-\alpha/2} \leq \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\sigma_x^2/n_1+\sigma_y^2/n_2}} \leq z_{1-\alpha/2}$$

$$(\bar{x}-\bar{y})-z_{1-\alpha/2}\sqrt{\sigma_x^2/n_1+\sigma_y^2/n_2} \leq \mu_x-\mu_y \leq (\bar{x}-\bar{y})+z_{1-\alpha/2}\sqrt{\sigma_x^2/n_1+\sigma_y^2/n_2}$$

Então, o intervalo de confiança $\delta = 1-\alpha$ para $\mu_x-\mu_y$ é da forma

$$\left\langle (\bar{x}-\bar{y}) \pm \underbrace{z_{1-\alpha/2}\sqrt{\sigma_x^2/n_1+\sigma_y^2/n_2}}_{\text{Margem de erro}} \right\rangle$$

Duas populações: σ_x e σ_y conhecidos.

E se quisermos um teste para a diferença de médias?

Duas populações: σ_x e σ_y conhecidos.

E se quisermos um teste para a diferença de médias? Sejam as hipóteses:

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|z| > k = z_{1-\alpha/2}$ (equivalentemente se $z > k$ ou $z < -k$);

Duas populações: σ_x e σ_y conhecidos.

E se quisermos um teste para a diferença de médias? Sejam as hipóteses:

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|z| > k = z_{1-\alpha/2}$ (equivalentemente se $z > k$ ou $z < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $z > k_1 = z_{1-\alpha}$;

Duas populações: σ_x e σ_y conhecidos.

E se quisermos um teste para a diferença de médias? Sejam as hipóteses:

- $H_0: \mu_x - \mu_y = D_0 \quad vs. \quad H_1: \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|z| > k = z_{1-\alpha/2}$ (equivalentemente se $z > k$ ou $z < -k$);
- Se $H_0: \mu_x - \mu_y \leq D_0 \quad vs \quad H_1: \mu_x - \mu_y > D_0$, rejeitamos H_0 se $z > k_1 = z_{1-\alpha}$;
- Se $H_0: \mu_x - \mu_y \geq D_0 \quad vs \quad H_1: \mu_x - \mu_y < D_0$, rejeitamos H_0 se $z < k_2 = z_{\alpha}$.

Duas populações: σ_x e σ_y conhecidos.

E se quisermos um teste para a diferença de médias? Sejam as hipóteses:

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|z| > k = z_{1-\alpha/2}$ (equivalentemente se $z > k$ ou $z < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $z > k_1 = z_{1-\alpha}$;
- Se $H_0 : \mu_x - \mu_y \geq D_0 \quad vs \quad H_1 : \mu_x - \mu_y < D_0$, rejeitamos H_0 se $z < k_2 = z_{\alpha}$.

Duas populações: σ_x e σ_y conhecidos.

E se quisermos um teste para a diferença de médias? Sejam as hipóteses:

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|z| > k = z_{1-\alpha/2}$ (equivalentemente se $z > k$ ou $z < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $z > k_1 = z_{1-\alpha}$;
- Se $H_0 : \mu_x - \mu_y \geq D_0 \quad vs \quad H_1 : \mu_x - \mu_y < D_0$, rejeitamos H_0 se $z < k_2 = z_{\alpha}$.

Em que

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}}$$

Duas populações: σ_x e σ_y conhecidos.

Considere os seguintes resultados:

- **Amostra 1**: $n_1 = 50$, $\bar{x} = 13.6$, $\sigma_1 = 2.2$
- **Amostra 2**: $n_2 = 35$, $\bar{y} = 11.6$, $\sigma_1 = 3.0$

a. Qual é a estimação por ponto de $\mu_x - \mu_y$?
b. Calcule um IC 90% para $\mu_x - \mu_y$
c. Teste $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$ (considera $\alpha = 0.10$)

Duas populações: σ_x e σ_y conhecidos.

Considere os seguintes resultados:

- **Amostra 1**: $n_1 = 50$, $\bar{x} = 13.6$, $\sigma_1 = 2.2$
- **Amostra 2**: $n_2 = 35$, $\bar{y} = 11.6$, $\sigma_1 = 3.0$

a. Qual é a estimação por ponto de $\mu_x - \mu_y$?
b. Calcule um IC 90% para $\mu_x - \mu_y$
c. Teste $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$ (considera $\alpha = 0.10$)

Solução:

Duas populações: σ_x e σ_y conhecidos.

Considere os seguintes resultados:

- **Amostra 1**: $n_1 = 50$, $\bar{x} = 13.6$, $\sigma_1 = 2.2$
- **Amostra 2**: $n_2 = 35$, $\bar{y} = 11.6$, $\sigma_1 = 3.0$

a. Qual é a estimação por ponto de $\mu_x - \mu_y$?
b. Calcule um IC 90% para $\mu_x - \mu_y$
c. Teste $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$ (considera $\alpha = 0.10$)

Solução:

a. A estimação por ponto de $\mu_x - \mu_y$ é $\bar{x} - \bar{y} = 13.6 - 11.6 = 2$

Duas populações: σ_x e σ_y conhecidos.

Considere os seguintes resultados:

- **Amostra 1**: $n_1 = 50$, $\bar{x} = 13.6$, $\sigma_1 = 2.2$
- **Amostra 2**: $n_2 = 35$, $\bar{y} = 11.6$, $\sigma_1 = 3.0$

a. Qual é a estimação por ponto de $\mu_x - \mu_y$?
b. Calcule um IC 90% para $\mu_x - \mu_y$
c. Teste $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$ (considera $\alpha = 0.10$)

Solução:

a. A estimação por ponto de $\mu_x - \mu_y$ é $\bar{x} - \bar{y} = 13.6 - 11.6 = 2$

b. IC 90%, istp implica que $\alpha = 0.10$,

$$\left\langle \underbrace{(\bar{x} - \bar{y})}_{2} \pm z_{1-\alpha/2} \underbrace{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}}_{\sqrt{2.2^2/50 + 3^2/35} = 0.594931} \right\rangle$$

Duas populações: σ_x e σ_y conhecidos.

b. Calcularemos $z_{1-\alpha/2}$, e como $\alpha = 0.10$

```
alpha = 0.10
qnorm(1-alpha/2)
```

```
## [1] 1.644854
```

Duas populações: σ_x e σ_y conhecidos.

b. Calcularemos $z_{1-\alpha/2}$, e como $\alpha = 0.10$

```
alpha = 0.10
qnorm(1-alpha/2)
```

```
## [1] 1.644854
```

$$\langle \underbrace{(\bar{x} - \bar{y})}_{2} \pm \underbrace{z_{1-\alpha/2}}_{1.64} \underbrace{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}}_{0.594931} \rangle = \langle 1.024313; 2.975687 \rangle$$

Duas populações: σ_x e σ_y conhecidos.

b. Calcularemos $z_{1-\alpha/2}$, e como $\alpha = 0.10$

```
alpha = 0.10
qnorm(1-alpha/2)
```

```
## [1] 1.644854
```

$$\left\langle \underbrace{(\bar{x}-\bar{y})}_{2} \pm \underbrace{z_{1-\alpha/2}}_{1.64} \underbrace{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}}_{0.594931} \right\rangle = \langle 1.024313; 2.975687 \rangle$$

c. Queremos testar $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$, então

$$z = \frac{(\bar{x}-\bar{y}) - D_0}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} = \frac{2-0}{0.594931} = 3.361734$$

Duas populações: σ_x e σ_y conhecidos.

- $z = 3.361734$

Duas populações: σ_x e σ_y conhecidos.

- $z = 3.361734$
- Como $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$

Duas populações: σ_x e σ_y conhecidos.

- $z = 3.361734$
- Como $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$
- $z_{1-\alpha/2} = 1.6448536$ (já calculamos isto antes para o IC)

Duas populações: σ_x e σ_y conhecidos.

- $z = 3.361734$
- Como $H_0 : \mu_x = \mu_y \quad vs \quad H_1 : \mu_x \neq \mu_y$, rejeitamos H_0 se $|z| > z_{1-\alpha/2}$
- $z_{1-\alpha/2} = 1.6448536$ (já calculamos isto antes para o IC)
- $3.361734 > 1.6448536$ **?** Sim, então rejeitamos H_0 e concluimos que $\mu_x \neq \mu_y$

Duas populações: σ_x e σ_y conhecidos.

Considere as seguintes hipóteses $H_0 : \mu_x - \mu_y \leq 0 \quad vs. \quad H_1 : \mu_x - \mu_y > 0$ e considere os seguintes resultados:

- **Amostra 1**: $n_1 = 40$, $\bar{x} = 25.2$, $\sigma_1 = 5.2$
- **Amostra 2**: $n_2 = 50$, $\bar{y} = 22.8$, $\sigma_1 = 6.0$

Rejeitamos H_0? (considere $\alpha = 0.01$)

Duas populações: σ_x e σ_y conhecidos.

Considere as seguintes hipóteses $H_0 : \mu_x - \mu_y \leq 0 \quad vs. \quad H_1 : \mu_x - \mu_y > 0$ e considere os seguintes resultados:

- **Amostra 1**: $n_1 = 40$, $\bar{x} = 25.2$, $\sigma_1 = 5.2$
- **Amostra 2**: $n_2 = 50$, $\bar{y} = 22.8$, $\sigma_1 = 6.0$

Rejeitamos H_0? (considere $\alpha = 0.01$)

Solução

- Estatística de teste:

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}} = \frac{(25.2 - 22.8) - 0}{\sqrt{5.2^2/40 + 6^2/50}} = \frac{2.4}{1.181524} = 2.031275$$

Duas populações: σ_x e σ_y conhecidos.

- $z = 2.031275$

Duas populações: σ_x e σ_y conhecidos.

- $z = 2.031275$
- Como $H_0 : \mu_x - \mu_y \leq 0 \quad vs. \quad H_1 : \mu_x - \mu_y > 0$ rejeitamos H_0 se $z > z_{1-\alpha}$

Duas populações: σ_x e σ_y conhecidos.

- $z = 2.031275$
- Como $H_0 : \mu_x - \mu_y \leq 0 \quad vs. \quad H_1 : \mu_x - \mu_y > 0$ rejeitamos H_0 se $z > z_{1-\alpha}$
- $\alpha = 0.01$

Duas populações: σ_x e σ_y conhecidos.

- $z = 2.031275$
- Como $H_0 : \mu_x - \mu_y \leq 0 \quad vs. \quad H_1 : \mu_x - \mu_y > 0$ rejeitamos H_0 se $z > z_{1-\alpha}$
- $\alpha = 0.01$

Duas populações: σ_x e σ_y conhecidos.

- $z = 2.031275$
- Como $H_0 : \mu_x - \mu_y \leq 0 \quad vs. \quad H_1 : \mu_x - \mu_y > 0$ rejeitamos H_0 se $z > z_{1-\alpha}$
- $\alpha = 0.01$

```
alpha = 0.01
qnorm(1-alpha)
```

```
## [1] 2.326348
```

- $z = 2.031275 > 2.3263479$ **?** Não, então não rejeitamos H_0 (nível de significância $\alpha = 0.01$).

Duas populações: σ_x e σ_y desconhecidos e diferentes.

- O que acontece quando não conhecemos σ_x nem σ_y?

Duas populações: σ_x e σ_y desconhecidos e diferentes.

- O que acontece quando não conhecemos σ_x nem σ_y?
- Devemos estimar esses valores pela variância amostral, assim teremos $\hat{\sigma}_x$ e $\hat{\sigma}_y$

Duas populações: σ_x e σ_y desconhecidos e diferentes.

- O que acontece quando não conhecemos σ_x nem σ_y?
- Devemos estimar esses valores pela variância amostral, assim teremos $\hat{\sigma}_x$ e $\hat{\sigma}_y$
- Substituir σ_x e σ_y por $\hat{\sigma}_x$ e $\hat{\sigma}_y$ terá um custo.

Duas populações: σ_x e σ_y desconhecidos e diferentes.

- O que acontece quando não conhecemos σ_x nem σ_y?
- Devemos estimar esses valores pela variância amostral, assim teremos $\hat{\sigma}_x$ e $\hat{\sigma}_y$
- Substituir σ_x e σ_y por $\hat{\sigma}_x$ e $\hat{\sigma}_y$ terá um custo.
- O custo é não podermos mais utilizar a distribuição normal, no caso utilizaremos uma distribuição t

Duas populações: σ_x e σ_y desconhecidos e diferentes.

- O que acontece quando não conhecemos σ_x nem σ_y?
- Devemos estimar esses valores pela variância amostral, assim teremos $\hat{\sigma}_x$ e $\hat{\sigma}_y$
- Substituir σ_x e σ_y por $\hat{\sigma}_x$ e $\hat{\sigma}_y$ terá um custo.
- O custo é não podermos mais utilizar a distribuição normal, no caso utilizaremos uma distribuição t

Duas populações: σ_x e σ_y desconhecidos e diferentes.

- O que acontece quando não conhecemos σ_x nem σ_y?
- Devemos estimar esses valores pela variância amostral, assim teremos $\hat{\sigma}_x$ e $\hat{\sigma}_y$
- Substituir σ_x e σ_y por $\hat{\sigma}_x$ e $\hat{\sigma}_y$ terá um custo.
- O custo é não podermos mais utilizar a distribuição normal, no caso utilizaremos uma distribuição t

Intervalo de confiança

O intervalo de confiança $\delta = 1 - \alpha$ para $\mu_x - \mu_y$ é da forma

$$\left\langle (\bar{x} - \bar{y}) \pm \underbrace{t_{1-\alpha/2,gl}\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}}_{\text{Margem de erro}} \right\rangle$$

Duas populações: σ_x e σ_y desconhecidos e diferentes.

Teste de Hipóteses:

Duas populações: σ_x e σ_y desconhecidos e diferentes.

Teste de Hipóteses:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}}$$

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$ (equivalentemente se $t > k$ ou $t < -k$);

Duas populações: σ_x e σ_y desconhecidos e diferentes.

Teste de Hipóteses:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}}$$

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$ (equivalentemente se $t > k$ ou $t < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $t > k_1 = t_{1-\alpha,gl}$;

Duas populações: σ_x e σ_y desconhecidos e diferentes.

Teste de Hipóteses:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}}$$

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$ (equivalentemente se $t > k$ ou $t < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $t > k_1 = t_{1-\alpha,gl}$;
- Se $H_0 : \mu_x - \mu_y \geq D_0 \quad vs \quad H_1 : \mu_x - \mu_y < D_0$, rejeitamos H_0 se $t < k_2 = t_{\alpha,gl}$.

Duas populações: σ_x e σ_y desconhecidos e diferentes.

Teste de Hipóteses:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}}$$

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$ (equivalentemente se $t > k$ ou $t < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $t > k_1 = t_{1-\alpha,gl}$;
- Se $H_0 : \mu_x - \mu_y \geq D_0 \quad vs \quad H_1 : \mu_x - \mu_y < D_0$, rejeitamos H_0 se $t < k_2 = t_{\alpha,gl}$.

Duas populações: σ_x e σ_y desconhecidos e diferentes.

Teste de Hipóteses:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}}$$

- $H_0 : \mu_x - \mu_y = D_0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq D_0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$ (equivalentemente se $t > k$ ou $t < -k$);
- Se $H_0 : \mu_x - \mu_y \leq D_0 \quad vs \quad H_1 : \mu_x - \mu_y > D_0$, rejeitamos H_0 se $t > k_1 = t_{1-\alpha,gl}$;
- Se $H_0 : \mu_x - \mu_y \geq D_0 \quad vs \quad H_1 : \mu_x - \mu_y < D_0$, rejeitamos H_0 se $t < k_2 = t_{\alpha,gl}$.

Quem é gl?

Duas populações: σ_x e σ_y desconhecidos e diferentes.

$$gl = \frac{\left(\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{\hat{\sigma}_x^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{\hat{\sigma}_y^2}{n_2}\right)^2}$$

Duas populações: σ_x e σ_y desconhecidos e iguais.

- O procedimento descrito anteriormente é valido para o caso das variâncias desconhecidas serem diferentes.

Duas populações: σ_x e σ_y desconhecidos e iguais.

- O procedimento descrito anteriormente é valido para o caso das variâncias desconhecidas serem diferentes.
- Quando as variâncias desconhecidas são iguais, utilizamos outra estatística de teste dada por

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}, \quad \text{em que } s_p^2 = \frac{(n_1-1)\hat{\sigma}_x^2 + (n_2-1)\hat{\sigma}_y^2}{n_1+n_2-2}$$

Duas populações: σ_x e σ_y desconhecidos e iguais.

- O procedimento descrito anteriormente é valido para o caso das variâncias desconhecidas serem diferentes.
- Quando as variâncias desconhecidas são iguais, utilizamos outra estatística de teste dada por

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t_{n_1+n_2-2}, \quad \text{em que } s_p^2 = \frac{(n_1 - 1)\hat{\sigma}_x^2 + (n_2 - 1)\hat{\sigma}_y^2}{n_1 + n_2 - 2}$$

- Na prática precisamos fazer um teste de hipóteses para verificar se as variâncias são iguais ou diferentes.

Duas populações: σ_x e σ_y desconhecidos e iguais.

- O procedimento descrito anteriormente é valido para o caso das variâncias desconhecidas serem diferentes.
- Quando as variâncias desconhecidas são iguais, utilizamos outra estatística de teste dada por

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t_{n_1+n_2-2}, \quad \text{em que } s_p^2 = \frac{(n_1 - 1)\hat{\sigma}_x^2 + (n_2 - 1)\hat{\sigma}_y^2}{n_1 + n_2 - 2}$$

- Na prática precisamos fazer um teste de hipóteses para verificar se as variâncias são iguais ou diferentes.
- Por enquanto essa informação será dada e não precisamos nos preocupar com isso.

Duas populações: σ_x e σ_y desconhecidos

Considere o seguinte teste $H_0 : \mu_x - \mu_y = 0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq 0.$
Considere as seguintes informações:

- **Amostra 1**: $n_1 = 35$, $\bar{x} = 13.6$ e $\hat{\sigma}_x = 5.2$
- **Amostra 2**: $n_2 = 40$, $\bar{y} = 10.1$ e $\hat{\sigma}_y = 8.5$

Rejeitamos ou não H_0? (considere $\alpha = 0.05$ e que $\sigma_x \neq \sigma_y$)

Duas populações: σ_x e σ_y desconhecidos

Considere o seguinte teste $H_0 : \mu_x - \mu_y = 0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq 0$.
Considere as seguintes informações:

- **Amostra 1**: $n_1 = 35$, $\bar{x} = 13.6$ e $\hat{\sigma}_x = 5.2$
- **Amostra 2**: $n_2 = 40$, $\bar{y} = 10.1$ e $\hat{\sigma}_y = 8.5$

Rejeitamos ou não H_0? (considere $\alpha = 0.05$ e que $\sigma_x \neq \sigma_y$)

Solução

- Estatística de teste:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\hat{\sigma}_x^2/n_1 + \hat{\sigma}_y^2/n_2}} = \frac{(13.6 - 10.1) - 0}{\sqrt{5.2^2/35 + 8.5^2/40}} = \frac{3.5}{1.605871} = 2.179503$$

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$
- Como estamos testando $H_0 : \mu_x - \mu_y = 0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq 0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$
- Como estamos testando $H_0 : \mu_x - \mu_y = 0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq 0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2,gl}$
- $$gl = \frac{\left(\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{\hat{\sigma}_x^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{\hat{\sigma}_y^2}{n_2}\right)^2} = \underbrace{\frac{\left(\frac{5.2^2}{35} + \frac{8.5^2}{40}\right)^2}{\frac{1}{35-1}\left(\frac{5.2^2}{35}\right)^2 + \frac{1}{40-1}\left(\frac{8.5^2}{40}\right)^2}}_{65.70829}$$

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$, $gl = 65.70829$

```
alpha = 0.05
qt(1-alpha/2,65.70829)
```

```
## [1] 1.99673
```

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$, $gl = 65.70829$
- Para $\alpha = 0.05$

```
alpha = 0.05
qt(1-alpha/2,65.70829)
```

```
## [1] 1.99673
```

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$, $gl = 65.70829$
- Para $\alpha = 0.05$

```
alpha = 0.05
qt(1-alpha/2,65.70829)
```

```
## [1] 1.99673
```

- $2.179503 > 1.9967299$? Sim, então rejeitamos H_0

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$, $gl = 65.70829$
- Para $\alpha = 0.05$

```
alpha = 0.05
qt(1-alpha/2,65.70829)
```

```
## [1] 1.99673
```

- $2.179503 > 1.9967299$ **?** Sim, então rejeitamos H_0

Duas populações: σ_x e σ_y desconhecidos

- $t = 2.179503$, $gl = 65.70829$
- Para $\alpha = 0.05$

```
alpha = 0.05
qt(1-alpha/2,65.70829)
```

```
## [1] 1.99673
```

- $2.179503 > 1.9967299$ **?** Sim, então rejeitamos H_0

Antigamente, as pessoas arredondavan `gl` para baixo e assim poder olhar nas tabelas da distribuição T (que só tinha os valores para graus de liberade inteiros). Hoje em dia não precisamos mais disso.

Duas populações: σ_x e σ_y desconhecidos

Resolveremos o mesmo exercícios mas **assumindo** que $\sigma_x = \sigma_y$:

Duas populações: σ_x e σ_y desconhecidos

Resolveremos o mesmo exercícios mas **assumindo** que $\sigma_x = \sigma_y$:

- **Amostra 1**: $n_1 = 35$, $\bar{x} = 13.6$ e $\hat{\sigma}_x = 5.2$
- **Amostra 2**: $n_2 = 40$, $\bar{y} = 10.1$ e $\hat{\sigma}_y = 8.5$

Duas populações: σ_x e σ_y desconhecidos

Resolveremos o mesmo exercícios mas **assumindo** que $\sigma_x = \sigma_y$:

- **Amostra 1**: $n_1 = 35$, $\bar{x} = 13.6$ e $\hat{\sigma}_x = 5.2$
- **Amostra 2**: $n_2 = 40$, $\bar{y} = 10.1$ e $\hat{\sigma}_y = 8.5$

$$s_p^2 = \frac{(n_1 - 1)\hat{\sigma}_x^2 + (n_2 - 1)\hat{\sigma}_y^2}{n_1 + n_2 - 2} = \frac{(35 - 1) \times 5.2^2 + (40 - 1) \times 8.5^2}{35 + 40 - 2} = 51.19329$$

Então

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{13.6 - 10.1}{\sqrt{51.19329}\sqrt{\frac{1}{35} + \frac{1}{40}}} = 2.113464$$

Duas populações: σ_x e σ_y desconhecidos

Resolveremos o mesmo exercícios mas **assumindo** que $\sigma_x = \sigma_y$:

- **Amostra 1**: $n_1 = 35$, $\bar{x} = 13.6$ e $\hat{\sigma}_x = 5.2$
- **Amostra 2**: $n_2 = 40$, $\bar{y} = 10.1$ e $\hat{\sigma}_y = 8.5$

$$s_p^2 = \frac{(n_1 - 1)\hat{\sigma}_x^2 + (n_2 - 1)\hat{\sigma}_y^2}{n_1 + n_2 - 2} = \frac{(35 - 1) \times 5.2^2 + (40 - 1) \times 8.5^2}{35 + 40 - 2} = 51.19329$$

Então

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{13.6 - 10.1}{\sqrt{51.19329}\sqrt{\frac{1}{35} + \frac{1}{40}}} = 2.113464$$

Como $H_0 : \mu_x - \mu_y = 0 \quad vs. \quad H_1 : \mu_x - \mu_y \neq 0$, rejeitamos H_0 se $|t| > k = t_{1-\alpha/2, n_1+n_2-2} = 1.9929971$

Diferença de médias para amostras relacionadas.

Diferença de médias para amostras relacionadas.

- Até agora temos trabalhado com inferência para a diferença de médias quando as duas populações são distintas (ou independentes);

Diferença de médias para amostras relacionadas.

- Até agora temos trabalhado com inferência para a diferença de médias quando as duas populações são distintas (ou independentes);
- Em ocasiões, precisamos fazer inferência para a diferená de médias quando as amostras são relacionadas.

Diferença de médias para amostras relacionadas.

- Até agora temos trabalhado com inferência para a diferença de médias quando as duas populações são distintas (ou independentes);
- Em ocasiões, precisamos fazer inferência para a diferená de médias quando as amostras são relacionadas.
- **Exemplo:** um mesmo grupo de funcionários antes e depois de um treinamento, um mesmo grupo de pacientes antes e depois de um medicamento, opinião de um memso numero de pessoas antes de depois um anuncio publicitário, etc.

Diferença de médias para amostras relacionadas.

- Até agora temos trabalhado com inferência para a diferença de médias quando as duas populações são distintas (ou independentes);
- Em ocasiões, precisamos fazer inferência para a diferená de médias quando as amostras são relacionadas.
- **Exemplo:** um mesmo grupo de funcionários antes e depois de um treinamento, um mesmo grupo de pacientes antes e depois de um medicamento, opinião de um memso numero de pessoas antes de depois um anuncio publicitário, etc.
- Nestes casos, a estatística de teste é dada por

$$t = \frac{\bar{d} - \mu_0}{\hat{\sigma}_d / \sqrt{n}} \sim t_{n-1},$$

com $d_i = x_i - y_i$, $\bar{d}$ e $\hat{\sigma}_d$ são a média e desvio padrão amostral de $d_1, \ldots, d_n$.

Diferença de médias para amostras relacionadas.

$$t = \frac{\bar{d} - \mu_0}{\hat{\sigma}_d / \sqrt{n}} \sim t_{n-1},$$

- $H_0 : \mu_d = \mu_0 \quad vs. \quad H_1 : \mu_d \neq \mu_0$, rejeitamos H_0 se $|\mathrm{t}| > t_{1-\alpha/2, n-1}$

Diferença de médias para amostras relacionadas.

$$t = \frac{\bar{d} - \mu_0}{\hat{\sigma}_d/\sqrt{n}} \sim t_{n-1},$$

- $H_0 : \mu_d = \mu_0 \quad vs. \quad H_1 : \mu_d \neq \mu_0$, rejeitamos H_0 se $|t| > t_{1-\alpha/2,n-1}$
- $H_0 : \mu_d \leq \mu_0 \quad vs. \quad H_1 : \mu_d > \mu_0$, rejeitamos H_0 se t $> t_{1-\alpha,n-1}$

Diferença de médias para amostras relacionadas.

$$t = \frac{\bar{d} - \mu_0}{\hat{\sigma}_d/\sqrt{n}} \sim t_{n-1},$$

- $H_0 : \mu_d = \mu_0 \quad vs. \quad H_1 : \mu_d \neq \mu_0$, rejeitamos H_0 se $|t| > t_{1-\alpha/2,n-1}$
- $H_0 : \mu_d \leq \mu_0 \quad vs. \quad H_1 : \mu_d > \mu_0$, rejeitamos H_0 se $t > t_{1-\alpha,n-1}$
- $H_0 : \mu_d \geq \mu_0 \quad vs. \quad H_1 : \mu_d < \mu_0$, rejeitamos H_0 se $t < t_{\alpha,n-1}$

Diferença de médias para amostras relacionadas.

Considere o seguinte teste de hipóteses: $H_0 : \mu_d \leq 0 \quad vs. \quad H_1 : \mu_d > 0.$
Os dados a seguir são amostras relacionadas.

Elemento	antes	depois
1	21	20
2	28	26
3	18	18
4	20	20
5	26	24

Rejitamos H_0 ou não? (considere $\alpha = 0.05$)

Diferença de médias para amostras relacionadas.

Elemento	antes	depois	d_i
1	21	20	1
2	28	26	2
3	18	18	0
4	20	20	0
5	26	24	2

Diferença de médias para amostras relacionadas.

Elemento	antes	depois	d_i
1	21	20	1
2	28	26	2
3	18	18	0
4	20	20	0
5	26	24	2

- Então $n = 5$ $\bar{d} = 1$ e $\hat{\sigma}_d = 1$

Diferença de médias para amostras relacionadas.

Elemento	antes	depois	d_i
1	21	20	1
2	28	26	2
3	18	18	0
4	20	20	0
5	26	24	2

- Então $n = 5$ $\bar{d} = 1$ e $\hat{\sigma}_d = 1$
- Estatística de teste

$$t = \frac{\bar{d} - \mu_0}{\hat{\sigma}_d/\sqrt{n}} = \frac{1}{1/\sqrt{5}} = 2.236068$$

Diferença de médias para amostras relacionadas.

- $t = 2.236068$

```r
alpha = 0.05; n = 5
qt(1-alpha, n-1)
```

```
## [1] 2.131847
```

Dica

Diferença de médias para amostras relacionadas.

- $t = 2.236068$
- Com $H_0 : \mu_d \leq 0 \quad vs. \quad H_1 : \mu_d > 0$, rejeitamos H_0 se t $> t_{1-\alpha,n-1}$

```
alpha = 0.05; n = 5
qt(1-alpha, n-1)
```

```
## [1] 2.131847
```

Dica

Diferença de médias para amostras relacionadas.

- $t = 2.236068$
- Com $H_0 : \mu_d \leq 0 \quad vs. \quad H_1 : \mu_d > 0$, rejeitamos H_0 se t $> t_{1-\alpha,n-1}$

```
alpha = 0.05; n = 5
qt(1-alpha, n-1)
```

```
## [1] 2.131847
```

- $2.236068 > 2.1318468$ **?** Sim, então rejeitamos H_0 e concluimos que $\mu_d > 0$

Dica

Diferença de médias para amostras relacionadas.

- $t = 2.236068$
- Com $H_0 : \mu_d \leq 0 \quad vs. \quad H_1 : \mu_d > 0$, rejeitamos H_0 se t $> t_{1-\alpha,n-1}$

```
alpha = 0.05; n = 5
qt(1-alpha, n-1)
```

```
## [1] 2.131847
```

- $2.236068 > 2.1318468$ **?** Sim, então rejeitamos H_0 e concluimos que $\mu_d > 0$

Dica

- Em alguns casos temos visto que para fazer inferência precisamos da distribuição t.

Diferença de médias para amostras relacionadas.

- $t = 2.236068$
- Com $H_0 : \mu_d \leq 0$ $\quad$ *vs.* $\quad$ $H_1 : \mu_d > 0$, rejeitamos H_0 se t $> t_{1-\alpha,n-1}$

```r
alpha = 0.05; n = 5
qt(1-alpha, n-1)
```

```
## [1] 2.131847
```

- $2.236068 > 2.1318468$ **?** Sim, então rejeitamos H_0 e concluimos que $\mu_d > 0$

Dica

- Em alguns casos temos visto que para fazer inferência precisamos da distribuição t.
- Quando o tamanho da amostra for grande, sempre podemos aproximar a distribuição t pela distribuição Normal.

Leituras recomendadas

- Anderson, D. R; Sweeney, D. J.; e Williams, T. A. (2008). *Estatística Aplicada à Administração e Economia*. 2ed. Cengage Learning. **Cap 10**
- Morettin, P.A; e Bussab, W. de O. (2004). *Estatística Básica*. 5ed, Saraiva. **Cap 13**