BIOESTATÍSTICA

Testes de hipóteses para comparação das médias de duas populações

GLEICE M S CONCEIÇÃO

FSP USP

BIOESTATÍSTICA

Teste de hipóteses para comparação das médias de duas populações com observações independentes

GLEICE M S CONCEIÇÃO

FSP USP

Comparando dois grupos

- Comparar técnicas usuais com métodos alternativos
- Comparação de drogas, de métodos cirúrgicos, de dietas, de procedimentos de laboratório, de tratamentos, etc.
- Considera-se o melhor tratamento aquele que produz bons resultados para a grande maioria da população em estudo

Comparando dois grupos

BIOESTATÍSTICA

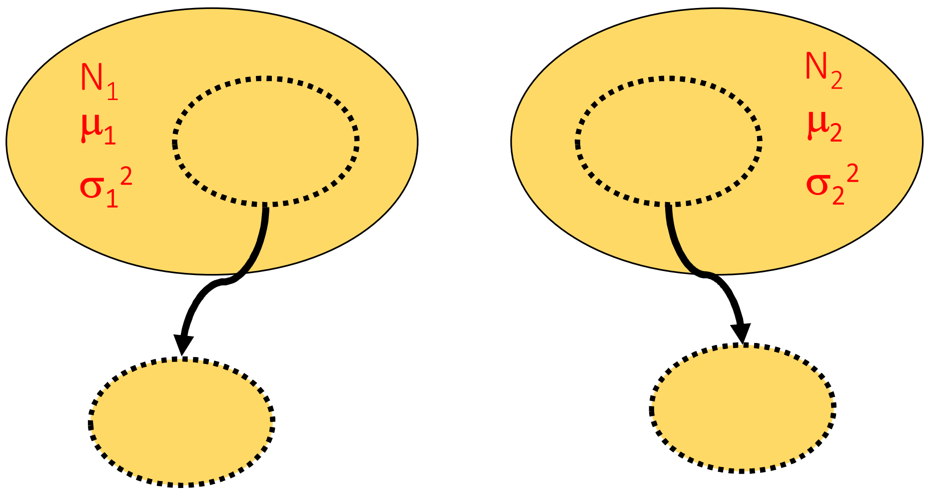
X_1 : v.a. de interesse na população 1, com $E(X_1) = \mu_1$ e $Var(X_1) = \sigma_1^2$

X_2 : v.a. de interesse na população 2, com $E(X_2) = \mu_2$ e $Var(X_2) = \sigma_2^2$

Para saber se $\mu_1 = \mu_2$

População 1

População 2



Observações independentes

Amostra 1

$n_1 \rightarrow \quad \bar{X}_1 \; e \; S_1^2$

Amostra 2

$n_2 \rightarrow \quad \bar{X}_2 \; e \; S_2^2$

Testes t-Student para a comparação das médias de duas populações

(observações independentes)

BIOESTATÍSTICA

1. Definir a(s) variável(eis) de interesse e o(s) parâmetro(s) que a(s) caracteriza(m).

X_1 : v.a. de interesse na população 1, com $E(X_1) = \mu_1$ e $Var(X_1) = \sigma_1^2$

X_2 : v.a. de interesse na população 2, com $E(X_2) = \mu_2$ e $Var(X_2) = \sigma_2^2$

2. Estabelecer as hipóteses nula e alternativa e interpretá-las.

$$\begin{array}{l} H_0\colon \mu_1 = \mu_2 \\ H_a\colon \mu_1 \begin{array}{c} > \\ < \\ \neq \end{array} \mu_2 \end{array} \Rightarrow \begin{array}{l} H_0\colon \mu_1 - \mu_2 = 0 \\ H_a\colon \mu_1 - \mu_2 \begin{array}{c} > \\ < \\ \neq \end{array} 0 \end{array} \Rightarrow \begin{array}{l} H_0\colon \mu_D = 0 \\ H_a\colon \mu_D \begin{array}{c} > \\ < \\ \neq \end{array} 0 \end{array} \quad \text{onde } \mu_D = \mu_1 - \mu_2$$

GLEICE M S CONCEIÇÃO
FSP - USP

Testes t-Student para a comparação das médias de duas populações

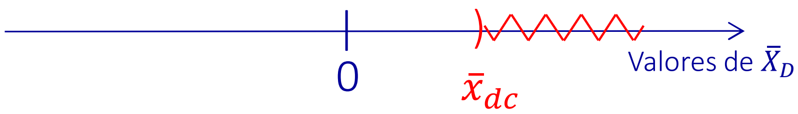
(observações independentes)

3. Definir a forma da região crítica, com base na hipótese alternativa

Parâmetro: $\mu_D = \mu_1 - \mu_2$

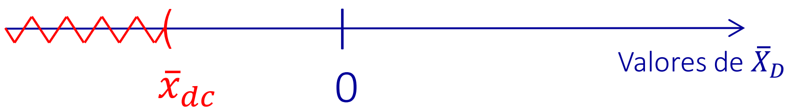
Estimador: $\bar{X}_D = \bar{X}_1 - \bar{X}_2$

✓ $H_a: \mu_1 > \mu_2$



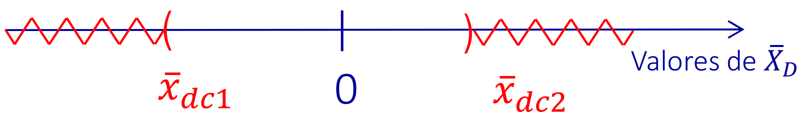
$RC = \{\bar{X}_D \in R \mid \bar{X}_D \geq \bar{x}_{dc}\}$

✓ $H_a: \mu_1 < \mu_2$



$RC = \{\bar{X}_D \in R \mid \bar{X}_D \leq \bar{x}_{dc}\}$

✓ $H_a: \mu_1 \neq \mu_2$



$RC = \{\bar{X}_D \in R \mid \bar{X}_D \leq \bar{x}_{dc1} \text{ ou } \bar{X}_D \geq \bar{x}_{dc2}\}$

Testes t-Student para a comparação das médias de duas populações

observações independentes

BIOESTATÍSTICA

Antes de definir a estatística do testes, alguns resultados ...

Sejam X_1 e X_2 variáveis aleatórias. Então:

- ✓ $E(X_1 + X_2) = E(X_1) + E(X_2)$
- ✓ $E(X_1 - X_2) = E(X_1) - E(X_2)$

Além disso, se X_1 e X_2 forem **independentes**:

- ✓ $VAR(X_1 + X_2) = VAR(X_1) + VAR(X_2)$
- ✓ $VAR(X_1 - X_2) = VAR(X_1) + VAR(X_2)$

GLEICE M S CONCEIÇÃO
FSP - USP

Testes t-Student para a comparação das médias de duas populações

(observações independentes)

BIOESTATÍSTICA

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador;
   definir a estatística do teste e sua distribuição; especificar as suposições assumidas.

Parâmetro: $\mu_D = \mu_1 - \mu_2$

Estimador e distribuição do estimador: $\bar{X}_D = \bar{X}_1 - \bar{X}_2 \sim N(\quad ? \quad , \quad ? \quad)$

Estatística do teste e sua distribuição:

GLEICE M S CONCEIÇÃO
FSP - USP

Testes t-Student para a comparação das médias de duas populações

(observações independentes)

BIOESTATÍSTICA

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador;
   definir a estatística do teste e sua distribuição; especificar as suposições assumidas.

Parâmetro: $\mu_D = \mu_1 - \mu_2$

Estimador e distribuição do estimador: $\bar{X}_D = \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\,, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$

Estatística do teste e sua distribuição:

✓ Se as variâncias populacionais (σ_1^2 e σ_2^2) fossem conhecidas (pouco provável!!):

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{\bar{X}_D - \mu_D}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0{,}1)$$

GLEICE M S CONCEIÇÃO
FSP - USP

Testes t-Student para a comparação das médias de duas populações

observações independentes

BIOESTATÍSTICA

4...

Estatística do teste e sua distribuição:

✓ Se as variâncias populacionais (σ_1^2 e σ_2^2) forem desconhecidas, precisamos substituí-las por seus estimadores (S_1^2 e S_2^2) na expressão abaixo, chegando a uma distribuição t-Student.

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Mas temos que considerar duas situações possíveis para as variâncias populacionais:

a) as variâncias populacionais são diferentes, isto é, $\sigma_1^2 \neq \sigma_2^2$

b) as variâncias populacionais são iguais, isto é, $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Testes t-Student para a comparação das médias de duas populações

observações independentes

BIOESTATÍSTICA

4...

Estatística do teste e sua distribuição:

a) as variâncias populacionais são diferentes, isto é, $\sigma_1^2 \neq \sigma_2^2$

Fácil! Substituímos cada variância pelo seu respectivo estimador:

$$Z = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad \Rightarrow T = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\vartheta$$

$$\text{onde } \vartheta = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Testes t-Student para a comparação das médias de duas populações

(observações independentes)

4...

Estatística do teste e sua distribuição:

b) as variâncias populacionais são iguais, isto é, $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Para estimar a variância única σ^2, utilizamos uma média ponderada de S_1^2 e S_2^2 :

$$S_{comb}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

E substituímos cada variância por S_{comb}^2:

$$Z = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad \Rightarrow T = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{S_{comb}^2}{n_1} + \frac{S_{comb}^2}{n_2}}} \sim t_{n_1+n_2-2}$$

IV. Teste t-Student para comparação de médias, quando as variâncias são diferentes ($\sigma_1^2 \neq \sigma_2^2$)

V. Teste t-Student para comparação de médias, quando as variâncias são iguais ($\sigma_1^2 = \sigma_2^2$)

observações independentes

BIOESTATÍSTICA

Estatística do teste IV

$$T = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\vartheta$$

onde $\vartheta = \dfrac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$

Estatística do teste V

$$T = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{S_{comb}^2}{n_1} + \frac{S_{comb}^2}{n_2}}} \sim t_{n_1+n_2-2}$$

onde $S_{comb}^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

Testes t-Student para a comparação das variâncias de duas populações

observações independentes

Para decidir se as variâncias populacionais são iguais ou diferentes, é necessário um teste de hipóteses do tipo

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_a : \sigma_1^2 \neq \sigma_2^2$$

Por ora, em lugar de testar se as variâncias são iguais ou não (Teste VI), usaremos o seguinte recurso:

Obtemos o quociente entre a maior e a menor variância amostral. Se este quociente for menor do que 3, assumimos que as variâncias populacionais são iguais. Caso contrário, assumimos que as variâncias populacionais são diferentes.

Testes t-Student para a comparação das médias de duas populações

observações independentes

BIOESTATÍSTICA

4...

Especifique as suposições assumidas.

✓ X_1 e X_2 têm distribuição *Normal.*

Isto garante que $\bar{X}_1$ e $\bar{X}_2$ terão distribuição *Normal* e, consequentemente, $\bar{X}_D$ terá distribuição *Normal*.

✓ Se o tamanho da amostra for grande, podemos usar o TLC.

Para avaliar se X_1 e X_2 têm distribuição *Normal:*

- *Histogramas*
- *Testes de Normalidade*

Testes t-Student para a comparação das médias de duas populações

observações independentes

BIOESTATÍSTICA

5. Fixar α

6. Obter a região crítica com base no valor de $\bar{x}_c$

7. Tomar a decisão, comparando o valor de $\bar{X}_{obs}$ com a região crítica

ou

6. Obter a região crítica com base no valor de t_c

7. Tomar a decisão, comparando o valor de t_{obs} com a região crítica

ou

6. Obter o nível descritivo (p-valor)

7. Tomar a decisão, comparando o p-valor com o valor de α

GLEICE M S CONCEIÇÃO
FSP - USP

Exemplo

Para avaliar a ação antidepressiva da tianeptina, foi realizado um ensaio clínico aleatorizado e duplo cego. Durante 42 dias, um grupo de pacientes recebeu a droga enquanto outro grupo recebeu placebo. A depressão foi medida através da escala de Montgomery-Asberg (MADRS). Os resultados estão apresentados abaixo.

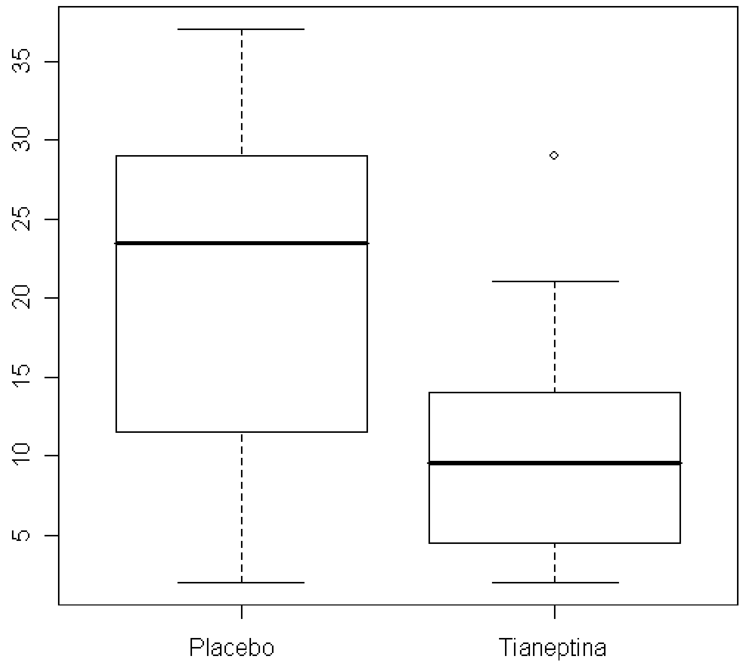
Teste as hipóteses correspondentes.

Placebo	Tianeptina
$n_1 = 21$	$n_2 = 31$
$\bar{X}_1 = 21{,}06$	$\bar{X}_2 = 10{,}63$
$S_1^2 = 90{,}72$	$S_2^2 = 52{,}65$

Exemplo

Análise descritiva

Box plots para os escores de depressão nos dois grupos de pacientes.



GLEICE M S CONCEIÇÃO
FSP - USP

Exemplo

Análise descritiva

BIOESTATÍSTICA

Intervalos de confiança para os escores médios de depressão nos dois grupos de pacientes.

$$IC(\mu_1; 0{,}95) = (16{,}72 : 25{,}40)$$

$$IC(\mu_2; 0{,}95) = (\ 7{,}97 : 13{,}29)$$

Conclusões iniciais:

É possível que a droga funcione pois, como os intervalos de confiança não se interceptam, as médias devem ser diferentes. Isto sugere que o escore médio populacional no grupo tianeptina é menor do que no grupo placebo .

Exemplo

Teste de hipóteses

1. Definir a(s) variável(eis) de interesse e o(s) parâmetro(s) que a(s) caracteriza(m).

X_1 – escore de depressão no grupo placebo, com $E(X_1) = \mu_1$ e $Var(X_1) = \sigma_1^2$

X_2 – escore de depressão no grupo tianeptina, com $E(X_2) = \mu_2$ e $Var(X_2) = \sigma_2^2$

2. Estabelecer as hipóteses nula e alternativa e interpretá-las.

$$\begin{matrix} H_0\colon \mu_1 = \mu_2 \\ H_a\colon \mu_1 > \mu_2 \end{matrix} \Rightarrow \begin{matrix} H_0\colon \mu_1 - \mu_2 = 0 \\ H_a\colon \mu_1 - \mu_2 > 0 \end{matrix} \Rightarrow \begin{matrix} H_0\colon \mu_D = 0 \\ H_a\colon \mu_D > 0 \end{matrix}$$

onde $\mu_D = \mu_1 - \mu_2$

$H_0\colon \mu_D = 0$ A droga não funciona

$H_a\colon \mu_D > 0$ A droga funciona

Exemplo

Teste de hipóteses

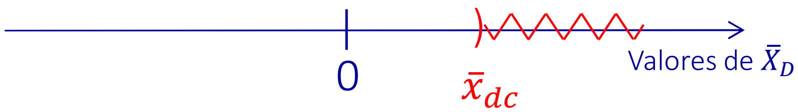
BIOESTATÍSTICA

3. Definir a forma da região crítica, com base na hipótese alternativa

Parâmetro: $\mu_D = \mu_1 - \mu_2$

Estimador: $\bar{X}_D = \bar{X}_1 - \bar{X}_2$

$H_a: \mu_1 > \mu_2$



$RC = \{\bar{X}_D \in R |\ \bar{X}_D \geq \bar{x}_{dc}\}$

Exemplo

Teste de hipóteses

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador; definir a estatística do teste e sua distribuição especificar as suposições assumidas.

Parâmetro: $\mu_D = \mu_1 - \mu_2$

Estimador e distribuição do estimador: $\bar{X}_D = \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\,, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Estatística do teste e sua distribuição:

Primeiro, precisamos decidir se as variâncias populacionais são iguais ou diferentes.

$$Q = \frac{90{,}72}{52{,}65} = 1{,}723$$

Como, Q < 3, decidimos que as variâncias populacionais são iguais, isto, é $\sigma_1^2 = \sigma_2^2 = \sigma^2$

GLEICE M S CONCEIÇÃO
FSP - USP

Exemplo

Teste de hipóteses

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador; definir a estatística do teste e sua distribuição especificar as suposições assumidas.

Estatística do teste e sua distribuição:

variâncias iguais $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$

$$\Rightarrow T = \frac{\bar{X}_D - \mu_D}{\sqrt{\frac{S_{comb}^2}{n_1} + \frac{S_{comb}^2}{n_2}}} \sim t_{n_1+n_2-2}$$

$$S_{comb}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(21-1)*90{,}72 + (31-1)*52{,}65}{21+31-2} = 67{,}878$$

$$S_{comb} = 8{,}239$$

Exemplo

Teste de hipóteses

4. Identificar o parâmetro, o seu estimador e a distribuição do estimador; definir a estatística do teste e sua distribuição especificar as suposições assumidas.

Suposições assumidas:

- ✓ X_1 e X_2 têm distribuição *Normal.*

  Isto garante que $\bar{X}_1$ e $\bar{X}_2$ terão distribuição *Normal* e,

  consequentemente, $\bar{X}_D$ terá distribuição *Normal*.

5. Fixar α

α=0,05 e α=0,01

Exemplo

Teste de hipóteses

6. ii) Obter a região crítica com base no valor de t_c

Da tabela da t_{50}:

Para α=0,05 $\Rightarrow$ RC={t ∈ $\mathcal{R}$ $|t \geq$1,6775}

Para α=0,01 $\Rightarrow$ RC={t ∈ $\mathcal{R}$ $|t \geq$2,4065}

Da amostra:

$$\bar{X}_{Dobs} = \bar{X}_{1\ obs} - \bar{X}_{2\ obs} = 21{,}06 - 10{,}63 = 10{,}43$$

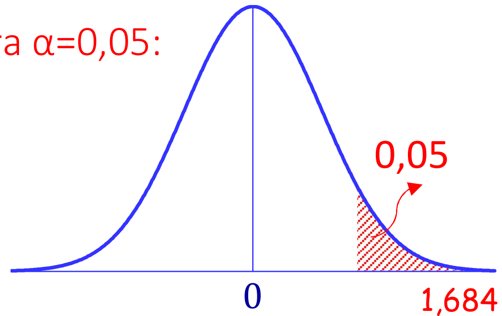
Sob H_0, o valor de t observado é:

$$t_{obs} = \frac{\bar{X}_{Dobs} - \mu_D}{\sqrt{\frac{S^2_{comb}}{n_1} + \frac{S^2_{comb}}{n_2}}} = \frac{10{,}43 - 0}{\sqrt{\frac{67{,}878}{21} + \frac{67{,}878}{31}}} = 4{,}4793$$

Sob H_0

t_{50}

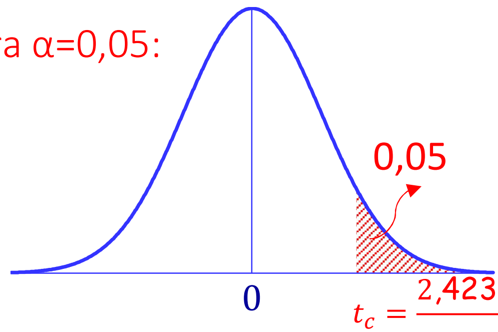
Para α=0,05:



$$t_c = \frac{1{,}684 + 1{,}671}{2} = 1{,}6775$$

t_{50}

Para α=0,05:



$$t_c = \frac{2{,}423 + 2{,}390}{2} = 2{,}4065$$

Exemplo

Teste de hipóteses

7. ii) Tomar a decisão, comparando o valor de t_{obs} com a região crítica

Para α=0,05 $\Rightarrow$ RC={t $\in \mathcal{R}$ $|t \geq$1,6775}

Como $t_{obs} = 4{,}4793$, $t_{obs} \in RC$, então rejeito H_0 e decido que **a droga funciona,** isto é, $\mu_1 > \mu_2$.

Para α=0,01 $\Rightarrow$ RC={t $\in \mathcal{R}$ $|t \geq$2,4065}

Como $t_{obs} = 4{,}4793$, $t_{obs} \in RC$, então rejeito H_0 e decido que **a droga funciona,** isto é, $\mu_1 > \mu_2$.

Exemplo

Teste de hipóteses

6. iii) Obter o nível descritivo (p-valor)

Da amostra: $\bar{X}_{Dobs} = \bar{X}_{1\ obs} - \bar{X}_{2\ obs} = 21{,}06 - 10{,}63 = 10{,}43$

$$p - valor = P(\bar{X}_D \geq \bar{X}_{Dobs} | H_0 \text{ é } verdadeira) =$$

$$= P(\bar{X}_D \geq 10{,}43\ |\mu_D = 0)$$

$$= P\left(\frac{\bar{X}_D - 0}{\sqrt{\frac{67{,}878}{21} + \frac{67{,}878}{31}}} \geq \frac{10{,}43 - 0}{\sqrt{\frac{67{,}878}{21} + \frac{67{,}878}{31}}}\right) = P(t_{50} \geq 4{,}4793) \cong 0{,}00$$

t_{obs}

Exemplo

Teste de hipóteses

7. iii) Tomar a decisão, comparando o p-valor com o valor de α

$$p - valor \cong 0{,}00$$

Como $p - valor < \alpha$, rejeito H_0 e decido que a droga funciona.

BIOESTATÍSTICA

Exemplo

Teste de hipóteses

O que significa o p-valor, neste caso?

$$p - valor \cong 0,00$$

Se, de fato, droga não funciona, a probabilidade de que a diferença entre os escores médios de depressão nos dois grupos seja tão grande ou maior que a observada é aproximadamente zero.

BIOESTATÍSTICA

Comparando as médias de duas populações

Em resumo...

- Amostras Dependentes ?
  - sim → Teste t-pareado **(Teste III)**
  - não → Variâncias desconhecidas → Teste para variâncias
    - Variâncias diferentes → Teste t-Student com variâncias diferentes **(Teste IV)**
    - Variâncias iguais → Teste t-Student com variâncias Iguais **(Teste V)**