Inferência Estatística

O teste t de *Student*

Fabio Cop (fabiocopf@gmail.com)
Instituto do Mar - UNIFESP
Última atualização em 11 de junho de 2021

Conteúdo da aula

1. Distribuição t de *Student*

2. Estatística t *versus* Estatística z: aumento dos graus de liberdade

3. Teste t para uma média amostral

4. Teste t para duas médias independentes

    - Teste t de Welch - variâncias heterogêneas

    - Comparando variâncias

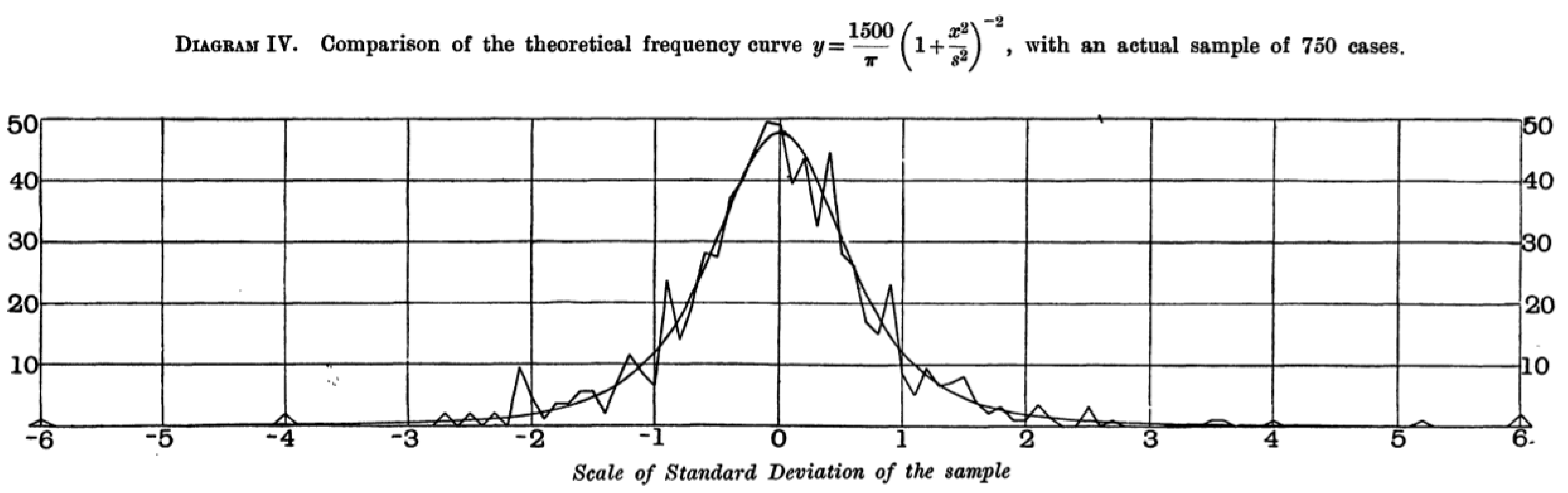
    - Teste t variâncias homogêneas

5. Teste t pareado para duas médias dependentes

Distribuição t de Student: Biometrika. 1908, vol. 6, 1-15

Distribuição *t* de Student: Biometrika. 1908, vol. 6, 1-15

> O Teorema Central do Limite garante que a distribuição de $\overline{X}$ tende à normalidade à medida que n aumenta.
>
> No entanto, para amostras pequenas a distribuição t de Student fornece uma aproximação melhor para a distribuição das médias amostrais.



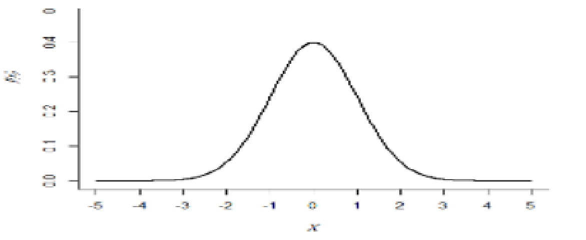
Distribuição *t* de Student: Biometrika. 1908, vol. 6, 1-15

Estatística Z

$$z = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

→

Dist. Normal Padronizada



→

$$\mathrm{P}(-z \leq Z \geq +z)$$

Estatística T

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

→

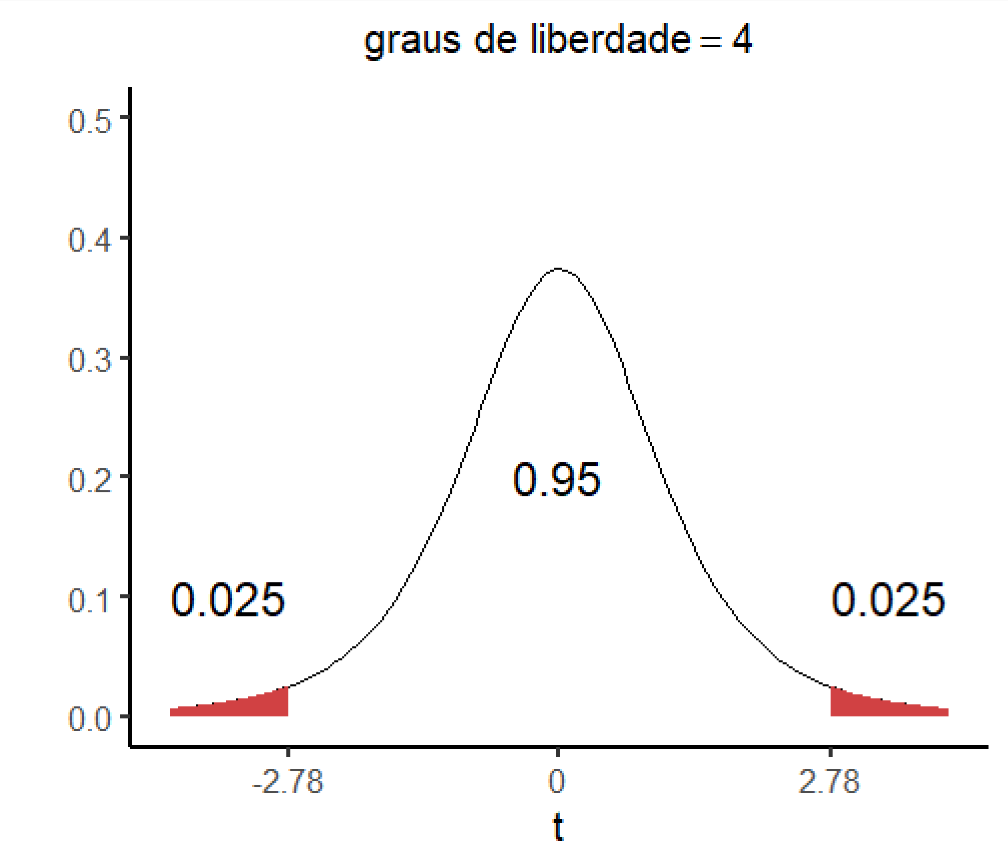
Distribuição t de Student

→

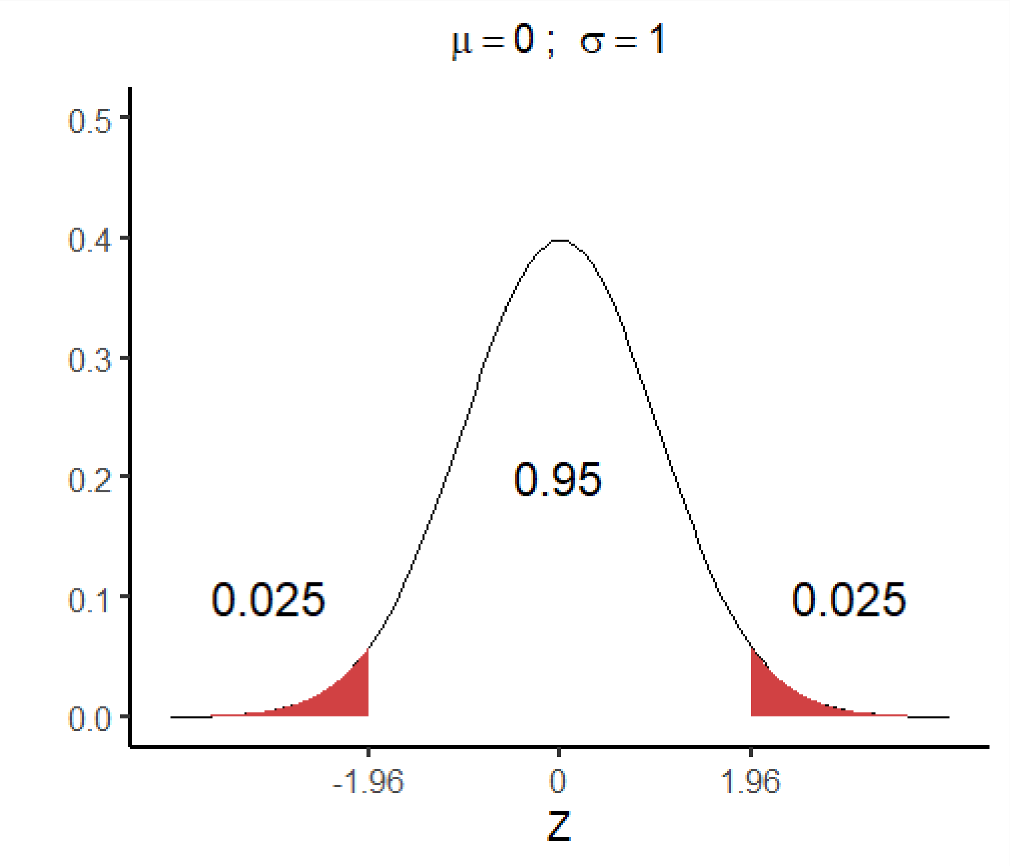
$$\mathrm{P}(-t \leq T \geq +t)$$

Distribuição *t* de Student: Biometrika. 1908, vol. 6, 1-15

Distribuição t

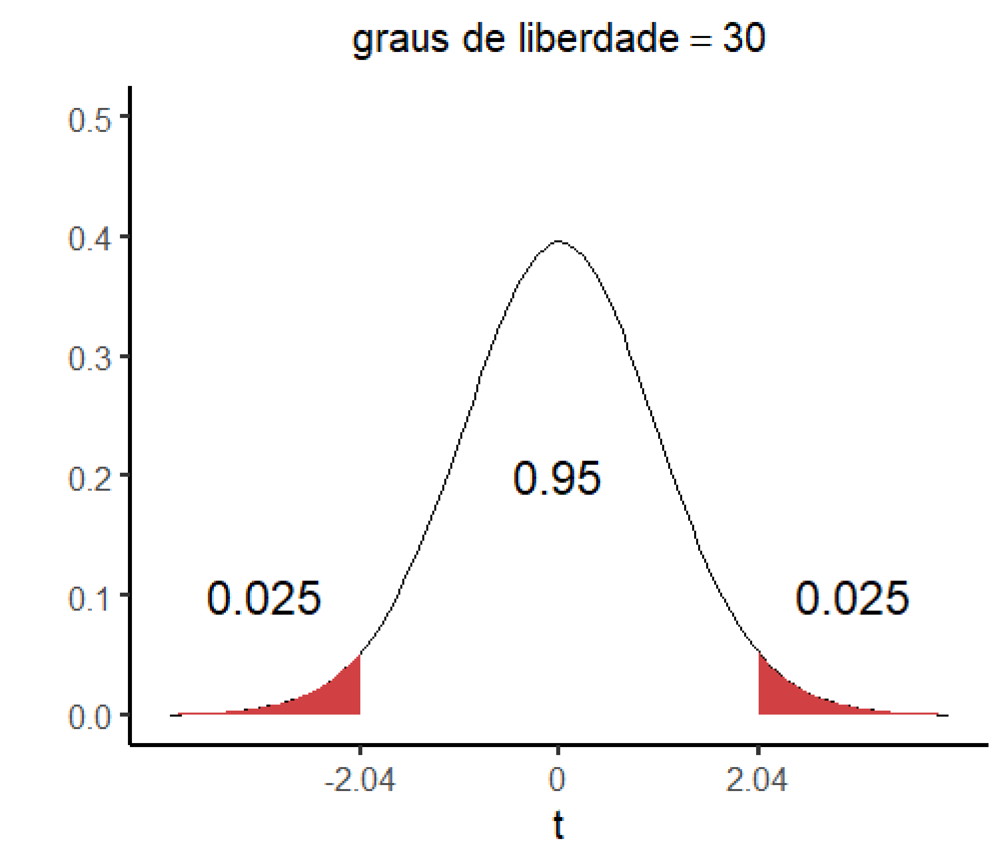


Distribuição Normal Padronizada

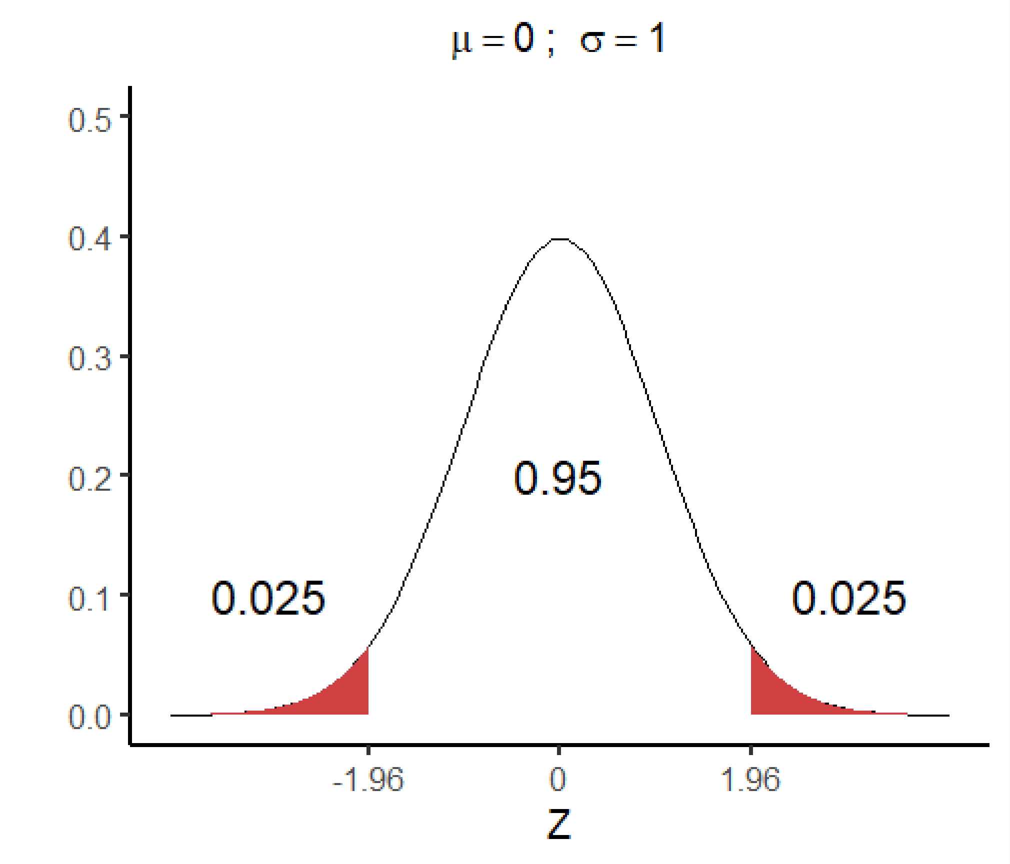


Distribuição *t* de Student: Biometrika. 1908, vol. 6, 1-15

Distribuição t

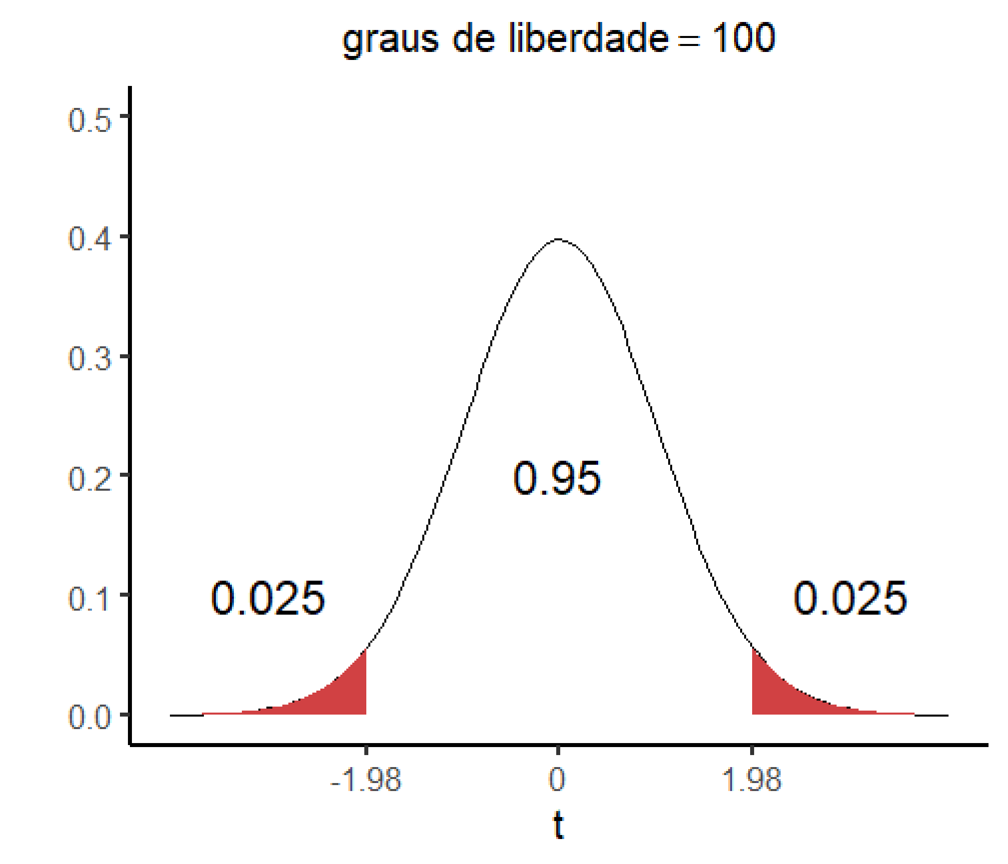


Distribuição Normal Padronizada

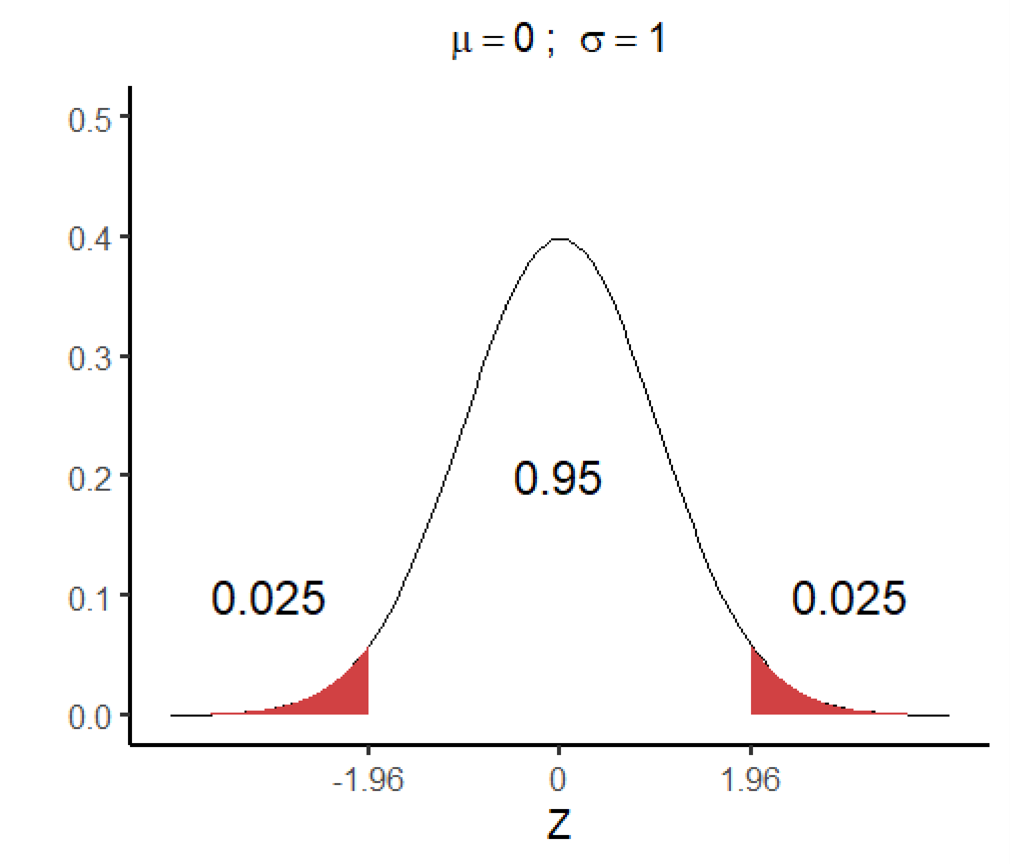


Distribuição *t* de Student: Biometrika. 1908, vol. 6, 1-15

Distribuição t



Distribuição Normal Padronizada



Teste t para uma média amostral

A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você dispõe de amostras para 10 riachos que obtidas na década de 70. Com estas amostras, deseja testar se houve mudança na diversidade média ao longo dos últimos 50 anos.

Hipóteses estatísticas

$H_0 : \mu = 2.65$ (Hipótese nula)

$H_a : \mu \neq 2.65$ (Hipótese alternativa)

$\alpha = 0.05$ (nível de significância)

Diversidade nos riachos				
2.92	2.99	2.83	2.53	2.67
2.69	3.60	2.64	2.99	2.55

A média amostral nos 10 riachos:

$$\overline{X} = \frac{\sum X_i}{n} = 2.84$$

com erro padrão de:

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.32}{3.16} = 0.1$$

e um valor de $t_{calculado}$:

$$t_c = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{2.84 - 2.65}{0.1} = 1.91$$

Teste t para uma média amostral

A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você dispõe de amostras para 10 riachos que obtidas na década de 70. Com estas amostras, deseja testar se houve mudança na diversidade média ao longo dos últimos 50 anos.

Hipóteses estatísticas

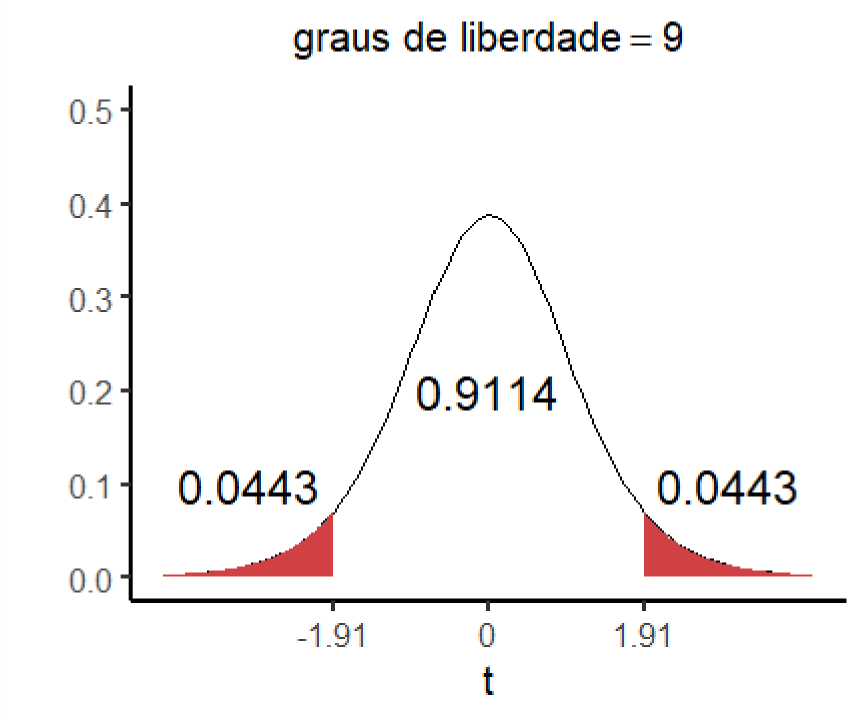
$H_0 : \mu = 2.65$ (Hipótese nula)

$H_a : \mu \neq 2.65$ (Hipótese alternativa)

$\alpha = 0.05$ (nível de significância)

Utilizando a Tabela t, encontramos a probabilidade de obtermos valores tão ou mais extremos que -1.91 e $+1.91$.

Segundo H_0



Teste t para uma média amostral

A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você dispõe de amostras para 10 riachos que obtidas na década de 70. Com estas amostras, deseja testar se houve mudança na diversidade média ao longo dos últimos 50 anos.

Hipóteses estatísticas

$H_0 : \mu = 2.65$ (Hipótese nula)

$H_a : \mu \neq 2.65$ (Hipótese alternativa)

$\alpha = 0.05$ (nível de significância)

Média da amostra

$$\overline{X} = 2.84$$

Resultado do teste

$$p = 0.0443 + 0.0443 = 0.0886$$

Como $0.0886 > 0.05$

Aceito H_0 e concluo que:

> Não há evidências na amostra que me permita dizer que a diversidade há 50 anos fosse diferente da diversidade atual.

Teste t para uma média amostral

A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você dispõe de amostras para 10 riachos que obtidas na década de 70. Com estas amostras, deseja testar se houve mudança na diversidade média ao longo dos últimos 50 anos.

---

amostra: 2.92, 2.69, 2.99, 3.6, 2.83, 2.64, 2.53, 2.99, 2.67, 2.55

```r
t.test(amostra, mu = 2.65, alternative = "two.sided")
```

```
##
##      One Sample t-test
##
## data:  amostra
## t = 1.9093, df = 9, p-value = 0.08857
## alternative hypothesis: true mean is not equal to 2.65
## 95 percent confidence interval:
##  2.614698 3.067302
## sample estimates:
## mean of x
##     2.841
```

Teste t para uma média amostral

(TESTE **UNICAUDAL**) A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você supõe que na década de 70 a diversidade fosse **maior** que atualmente devido a fatores como desmatamento e aumento na ocupação urbana ao longo dos últimos 50 anos.

Hipóteses estatísticas

$H_0 : \mu = 2.65$ (Hipótese nula)

$H_a : \mu > 2.65$ (Hipótese alternativa)

$\alpha = 0.05$ (nível de significância)

Diversidade nos riachos

2.92	2.99	2.83	2.53	2.67
2.69	3.60	2.64	2.99	2.55

A média amostral nos 10 riachos:

$$\overline{X} = \frac{\sum X_i}{n} = 2.84$$

com erro padrão de:

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.32}{3.16} = 0.1$$

e um valor de $t_{calculado}$:

$$t_c = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{2.84 - 2.65}{0.1} = 1.91$$

Teste t para uma média amostral

(TESTE **UNICAUDAL**) A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você supõe que na década de 70 a diversidade fosse **maior** que atualmente devido a fatores como desmatamento e aumento na ocupação urbana ao longo dos últimos 50 anos.

Hipóteses estatísticas

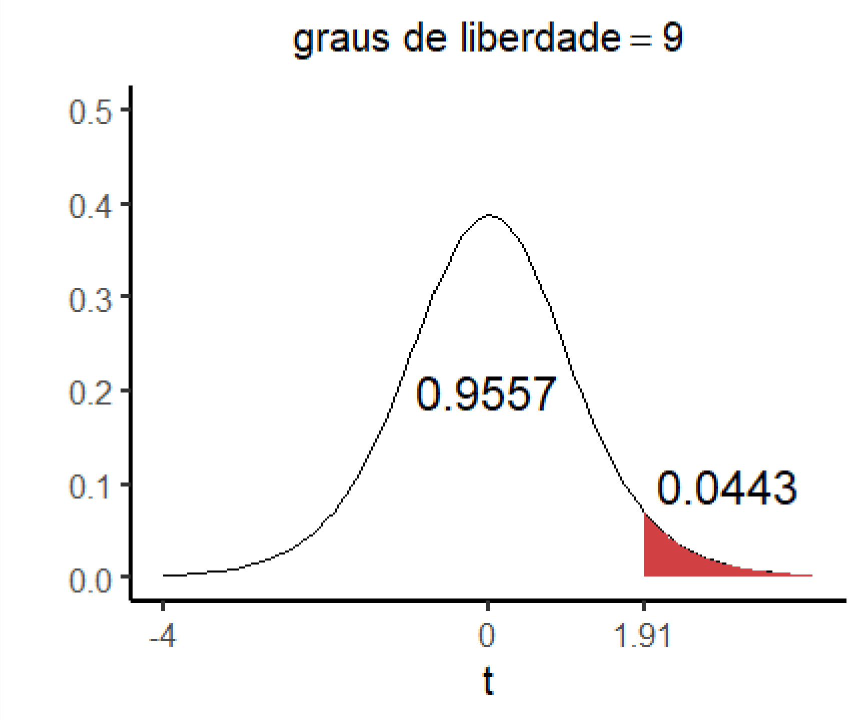
$H_0 : \mu = 2.65$ (Hipótese nula)

$H_a : \mu > 2.65$ (Hipótese alternativa)

$\alpha = 0.05$ (nível de significância)

Utilizando a Tabela t, encontramos a probabilidade de obtermos valores tão ou mais extremos que $+1.91$.

Segundo H_0



Teste t para uma média amostral

(TESTE **UNICAUDAL**) A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você supõe que na década de 70 a diversidade fosse **maior** que atualmente devido a fatores como desmatamento e aumento na ocupação urbana ao longo dos últimos 50 anos.

Hipóteses estatísticas

$H_0 : \mu = 2.65$ (Hipótese nula)

$H_a : \mu > 2.65$ (Hipótese alternativa)

$\alpha = 0.05$ (nível de significância)

Média da amostra

$\overline{X} = 2.84$

Resultado do teste

$p = 0.0443$

Como $0.0443 \leq 0.05$

Rejeito H_0 e concluo que:

> Há evidências na amostra para dizer que a diversidade há 50 anos era **maior** que diversidade atual.

Teste t para uma média amostral

(TESTE **UNICAUDAL**) A diversidade média de peixes em riachos costeiros de Mata Atlântica tem distribuição normal com $\mu = 2.65$. Você supõe que na década de 70 a diversidade fosse **maior** que atualmente devido a fatores como desmatamento e aumento na ocupação urbana ao longo dos últimos 50 anos.

---

amostra: 2.92, 2.69, 2.99, 3.6, 2.83, 2.64, 2.53, 2.99, 2.67, 2.55

```
t.test(amostra, mu = 2.65, alternative = "greater")
```

```
##
##      One Sample t-test
##
## data:  amostra
## t = 1.9093, df = 9, p-value = 0.04428
## alternative hypothesis: true mean is greater than 2.65
## 95 percent confidence interval:
##  2.657618      Inf
## sample estimates:
## mean of x
##     2.841
```

Teste t para duas amostras independentes

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$H_0 : \mu_{macho} = \mu_{fêmea}$

$H_a : \mu_{macho} \neq \mu_{fêmea}$

$\alpha = 0.05$

ID	1	2	3	4	5	6	7	8	9	10
Femea	110	111	107	108	110	105	107	106	111	111
Macho	120	107	110	116	114	111	113	117	114	112

$\overline{X}_{fêmea} = 108.6$; $s_{fêmea} = 2.27$; $n_{fêmea} = 10$

$\overline{X}_{macho} = 113.4$; $s_{macho} = 3.72$; $n_{macho} = 10$

Teste t de Welch - variâncias heterogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$H_0 : \mu_{macho} = \mu_{fêmea}$

$H_a : \mu_{macho} \neq \mu_{fêmea}$

$\alpha = 0.05$

1. Erro padrão da diferença de médias

$$s_{\overline{X}_f - \overline{X}_m} = \sqrt{\frac{s_f^2}{n_f} + \frac{s_m^2}{n_m}}$$

2. Graus de liberdade

$$gl = \frac{\left(\frac{s_f^2}{n_f} + \frac{s_m^2}{n_m}\right)^2}{\frac{\left(\frac{s_f^2}{n_f}\right)^2}{n_f - 1} + \frac{\left(\frac{s_m^2}{n_m}\right)^2}{n_m - 1}}; \text{ ou } gl = min(n_f - 1, n_m - 1)$$

3. Estatística t

$$t_c = \frac{\overline{X}_f - \overline{X}_m}{s_{\overline{X}_f - \overline{X}_m}}$$

Teste t de Welch - variâncias heterogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$H_0 : \mu_{macho} = \mu_{fêmea}$

$H_a : \mu_{macho} \neq \mu_{fêmea}$

$\alpha = 0.05$

ID	1	2	3	4	5	6	7	8	9	10
Femea	110	111	107	108	110	105	107	106	111	111
Macho	120	107	110	116	114	111	113	117	114	112

1. Erro padrão da diferença de médias

$$s_{\overline{X}_f - \overline{X}_m} = \sqrt{\frac{2.27}{10} + \frac{3.72}{10}} = 1.38$$

2. Graus de liberdade

$gl = 14.89$, ou $gl = 9$ (método simplificado)

3. Estatística t

$$t_c = \frac{108.6 - 108.6}{1.38} = -3.48$$

Teste t de Welch - variâncias heterogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

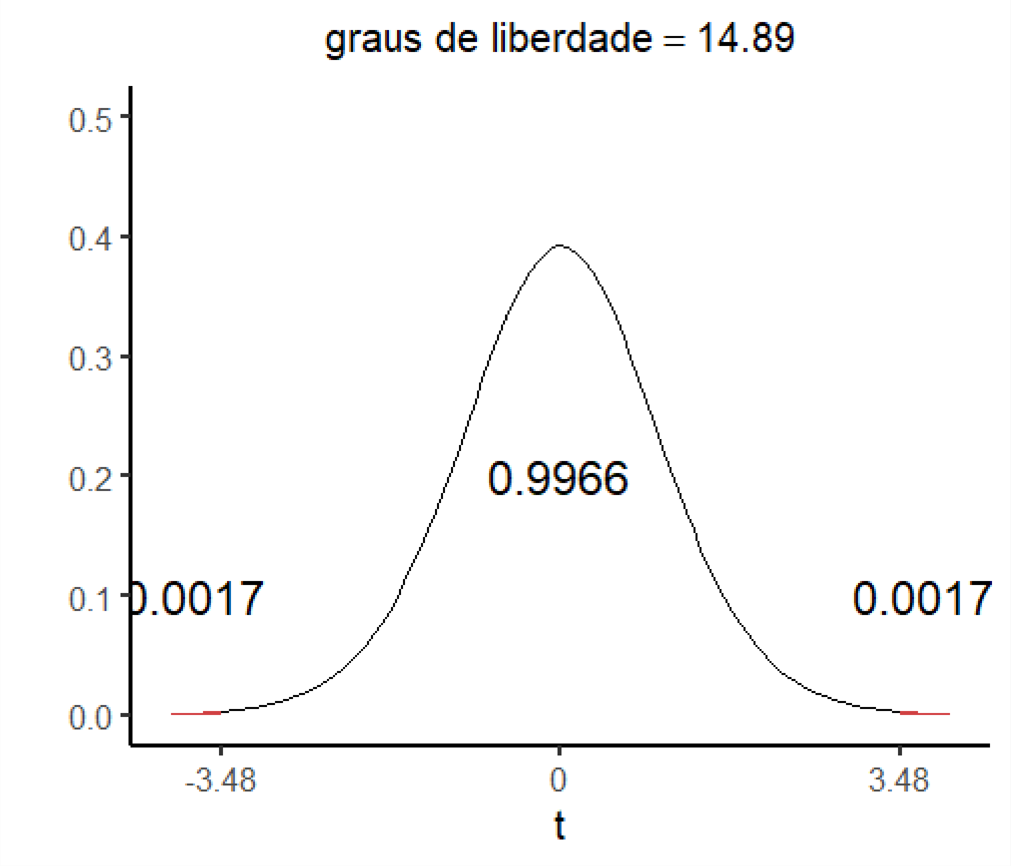
$\overline{X}_{fêmea} = 108.6$; $\overline{X}_{macho} = 113.4$

Resultado do teste

$p = 0.0017 + 0.0017 = 0.0034$

Rejeito H_0 e concluo que:

> **Há evidências** na amostra para assumir que existem diferenças no comprimento médio entre machos e fêmeas. O comprimento da mandíbula das fêmeas é, em média, 4.8 mm menor.



Teste t de Welch - variâncias heterogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

```r
t.test(Comprimento ~ Sexo, data = jackal,
                       alternative = 'two.sided',
                       var.equal = FALSE)
```

```
##
##      Welch Two Sample t-test
##
## data:  Comprimento by Sexo
## t = -3.4843, df = 14.894, p-value = 0.00336
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.738105 -1.861895
## sample estimates:
## mean in group Femea mean in group Macho
##               108.6               113.4
```

Comprimento	Sexo
120	Macho
107	Macho
110	Macho
116	Macho
114	Macho
111	Macho
113	Macho
117	Macho
114	Macho
112	Macho
110	Femea
111	Femea
107	Femea
108	Femea
110	Femea
105	Femea

Teste de hipóteses para duas variâncias - Teste F

Existe evidência de que a variância no tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

ID	1	2	3	4	5	6	7	8	9	10
Femea	110	111	107	108	110	105	107	106	111	111
Macho	120	107	110	116	114	111	113	117	114	112

Hipóteses estatísticas

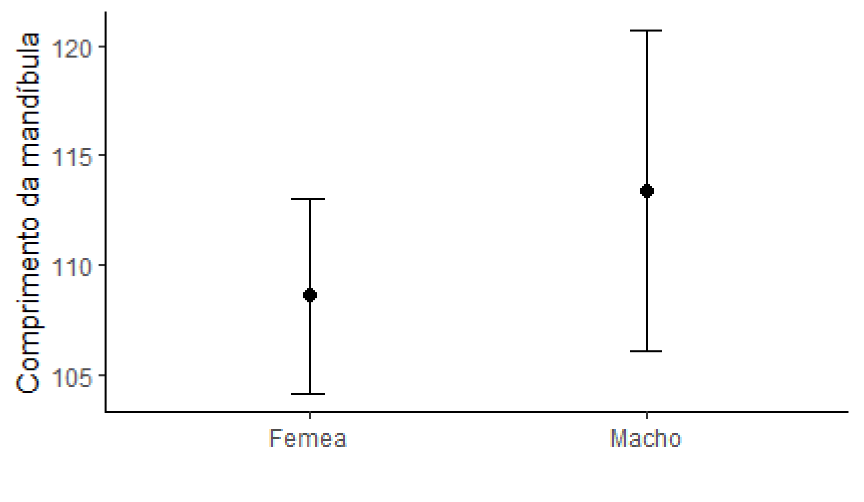
$H_0 : \sigma^2_{macho} = \sigma^2_{fêmea}$

$H_a : \sigma^2_{macho} \neq \sigma^2_{fêmea}$

$\alpha = 0.05$

$s^2_{fêmea} = 5.153$; $n_{fêmea} = 10$

$s^2_{macho} = 13.838$; $n_{macho} = 10$



Teste de hipóteses para duas variâncias - Teste F

Existe evidência de que a variância no tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$H_0 : \sigma^2_{macho} = \sigma^2_{fêmea}$

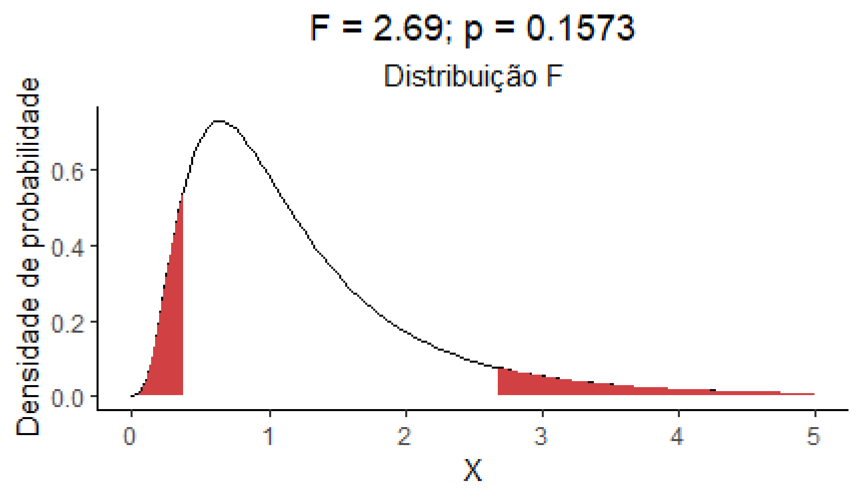
$H_a : \sigma^2_{macho} \neq \sigma^2_{fêmea}$

$\alpha = 0.05$

$s^2_{fêmea} = 5.153$; $n_{fêmea} = 10$

$s^2_{macho} = 13.838$; $n_{macho} = 10$

$$F = \frac{s^2_{maior}}{s^2_{menor}} = \frac{13.838}{5.153} = 2.69$$



Teste de hipóteses para duas variâncias - Teste F

Existe evidência de que a variância no tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

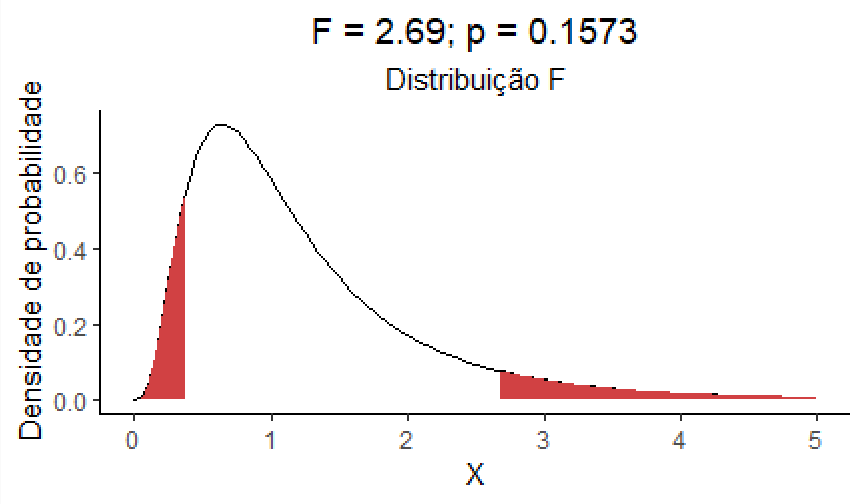
```
vmaior = jackal$Comprimento[jackal$Sexo == "Macho"]
vmenor = jackal$Comprimento[jackal$Sexo == "Femea"]
var.test(x = vmaior, y = vmenor, data = jackal)
```

```
##
##      F test to compare two variances
##
## data:  vmaior and vmenor
## F = 2.681, num df = 9, denom df = 9, p-value = 0.1579
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   0.665931 10.793829
## sample estimates:
## ratio of variances
##           2.681034
```

$$s^2_{fêmea} = 5.153; n_{fêmea} = 10$$

$$s^2_{macho} = 13.838; n_{macho} = 10$$

$$F = \frac{s^2_{maior}}{s^2_{menor}} = \frac{13.838}{5.153} = 2.69$$



Teste de hipóteses para duas variâncias - Teste F

Existe evidência de que a variância no tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$H_0 : \sigma^2_{macho} = \sigma^2_{fêmea}$

$H_a : \sigma^2_{macho} \neq \sigma^2_{fêmea}$

$\alpha = 0.05$

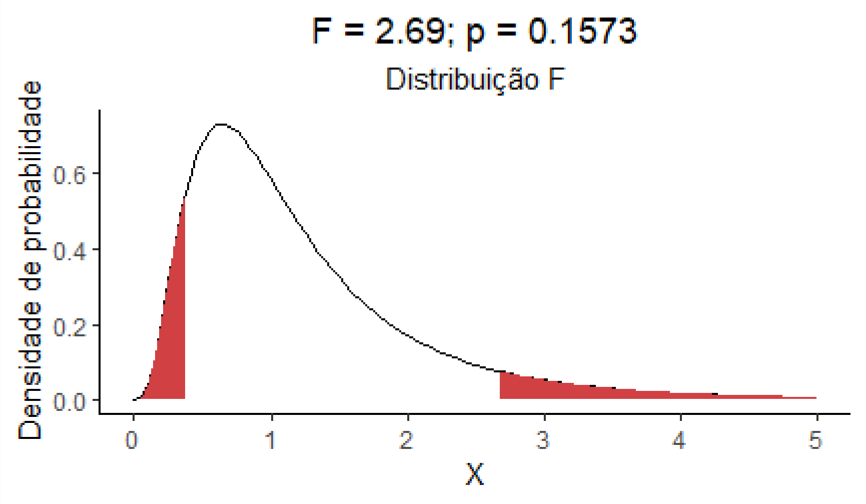
Aceito H_0 e concluo que:

> Ainda que a variância amostral em machos seja 2.69 vezes maior que em fêmeas, a diferença é **não significativa**. Portanto posso assumir variâncias **homogêneas**, isto é, $\sigma^2_{fêmea} = \sigma^2_{macho}$.

$s^2_{fêmea} = 5.153$; $n_{fêmea} = 10$

$s^2_{macho} = 13.838$; $n_{macho} = 10$

$$F = \frac{s^2_{maior}}{s^2_{menor}} = \frac{13.838}{5.153} = 2.69$$



Teste t - Variâncias homogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$H_0 : \mu_{macho} = \mu_{fêmea}$

$H_a : \mu_{macho} \neq \mu_{fêmea}$

$\alpha = 0.05$

ID	1	2	3	4	5	6	7	8	9	10
Femea	110	111	107	108	110	105	107	106	111	111
Macho	120	107	110	116	114	111	113	117	114	112

$\overline{X}_{fêmea} = 108.6$; $\overline{X}_{macho} = 113.4$

$s_{fêmea} = 2.27$; $s_{macho} = 3.72$

$n_{fêmea} = n_{macho} = 10$

Teste t - Variâncias homogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

Hipóteses estatísticas

$$H_0 : \mu_{macho} = \mu_{fêmea}$$

$$H_a : \mu_{macho} \neq \mu_{fêmea}$$

$$\alpha = 0.05$$

1. Variância conjunta

$$s_p^2 = \frac{\sum(X_{i,f} - \overline{X}_f)^2 + \sum(X_{i,m} - \overline{X}_m)^2}{(n_f - 1) + (n_m - 1)}$$

2. Erro padrão da diferença de médias

$$s_{\overline{X}_f - \overline{X}_m} = \sqrt{\frac{s_p^2}{n_f} + \frac{s_p^2}{n_m}}$$

3. Estatística t

$$t_c = \frac{\overline{X}_f - \overline{X}_m}{s_{\overline{X}_f - \overline{X}_m}}$$

Teste t - Variâncias homogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

ID	1	2	3	4	5	6	7	8	9	10
Femea	110	111	107	108	110	105	107	106	111	111
Macho	120	107	110	116	114	111	113	117	114	112

Hipóteses estatísticas

$H_0 : \mu_{macho} = \mu_{fêmea}$

$H_a : \mu_{macho} \neq \mu_{fêmea}$

$\alpha = 0.05$

1. Variância conjunta

$$s_p^2 = \frac{46.4+124.4}{9+9} = 9.49$$

2. Erro padrão da diferença de médias

$$s_{\overline{X}_f - \overline{X}_m} = \sqrt{\frac{9.49}{10} + \frac{9.49}{10}} = 1.38$$

3. Estatística t

$$t_c = \frac{108.6-108.6}{1.38} = -3.48$$

Teste t - Variâncias homogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

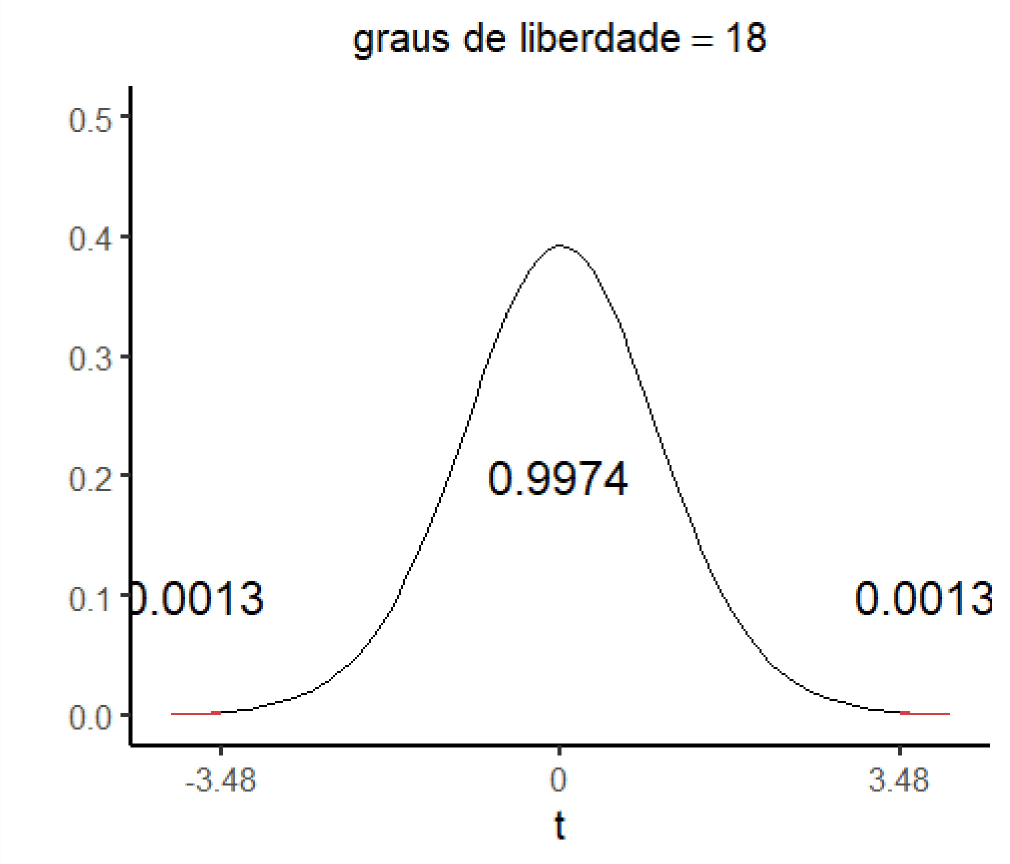
$\overline{X}_{f\hat{e}mea} = 108.6$; $\overline{X}_{macho} = 113.4$

Resultado do teste

$p = 0.0013 + 0.0013 = 0.0026$

Rejeito H_0 e concluo que:

> **Há evidências** na amostra para assumir que existem diferenças no comprimento médio entre machos e fêmeas. O comprimento da mandíbula das fêmeas é, em média, 4.8 mm menor.



Teste t - Variâncias homogêneas

Existe evidência de que o tamanho das mandíbulas de Chacais dourados seja diferente em machos e fêmeas?

```r
t.test(Comprimento ~ Sexo, data = jackal,
                       alternative = 'two.sided',
                       var.equal = TRUE)
```

```
##
##      Two Sample t-test
##
## data:  Comprimento by Sexo
## t = -3.4843, df = 18, p-value = 0.002647
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.694227 -1.905773
## sample estimates:
## mean in group Femea mean in group Macho
##               108.6               113.4
```

Comprimento	Sexo
120	Macho
107	Macho
110	Macho
116	Macho
114	Macho
111	Macho
113	Macho
117	Macho
114	Macho
112	Macho
110	Femea
111	Femea
107	Femea
108	Femea
110	Femea
105	Femea

Teste t para duas amostras independentes

Pressupostos

1. Amostras independentes;

2. A população de origem tem distribuição Normal;

3. As variâncias populacionais de fêmeas e machos são **homogêneas**, $\sigma^2_{fêmea} = \sigma^2_{macho}$.