

Inferência Estatística II

Testes de hipóteses específicos



Prof. Paulo Cerqueira Jr - cerqueirajr@ufpa.br

Faculdade de Estatística - FAEST

Instituto de Ciências Exatas e Naturais - ICEN

<https://github.com/paulocerqueirajr> 

Introdução

Introdução

- Existem uma série de testes de hipóteses, cada um com sua finalidade.
- Para uma amostra, temos alguns testes:
 - Teste de normalidade
 - Teste-t comparação de grupos (independentes e pareados);
 - Testes não-paramétricos Qui-quadrado;

Testes para duas amostra

Testes para duas amostra

- Para duas amostra, temos alguns testes:
 - Teste para a comparação de médias com variância conhecida (Teste Z);
 - Teste para a comparação de médias com variância desconhecida (Teste t de **Student** comparação de grupos);
 - Teste para a comparação de variâncias;
 - Teste para comparação de amostras dependentes (Teste t de **Student** comparação de grupos pareado)

Teste para Comparação das médias

Teste para Comparação das médias

- Queremos comparar:

$$\begin{array}{ll} & \times \quad (i) H_1 : \mu_1 \neq \mu_2. \text{ (bilateral)} \\ H_0 : \mu_1 = \mu_2 & \times \quad (ii) H_1 : \mu_1 > \mu_2. \text{ (Unilateral à direita)} \\ & \times \quad (iii) H_1 : \mu_1 < \mu_2. \text{ (Unilateral à esquerda)} \end{array}$$

CASO 1: Mesma variância, conhecida. ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).

- Estatística para o teste:

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

Exemplo

Exemplo: Vamos supor que os níveis de QI entre meninos e meninas da décima série sejam normalmente distribuídos, cada um com desvio padrão populacional de 25. Uma professora quer saber se a média do QI entre meninos e meninas da turma é diferente, então ela seleciona duas amostras aleatórias — uma de meninos e outra de meninas — cada uma com tamanho 40, e registra os níveis de QI. Vamos realizar o teste Z para duas amostras para determinar se a média dos níveis de QI é diferente entre meninos e meninas, com nível de significância de 5%.

Os dados:

```
1 # Definindo os dados:
2 qi_meninos = c( 79,118,99,117,98,102,112,111,102,73,114,97,114,122,
3                90,115,84,105,84,126,83,96,111,151,147,103,104,118,
4                132,108,95,118,121,88,94,92,94,109,105,123)
5 qi_meninas = c( 99,128,89,107,99,104,119,112,105,93,84,91,113,129,
6                100,105,94,115,114,106,113,116,116,131,117,123,134,
7                128,112,101,105,89,101,118,124,72,104,119,145,133)
8 dados <- data.frame(grupo = rep(c("Meninas", "Meninos"), each=40),
9                    qi = c(qi_meninos, qi_meninas))
10 head(dados)
```

	grupo	qi
1	Meninas	79
2	Meninas	118
3	Meninas	99
4	Meninas	117
5	Meninas	98
6	Meninas	102

Exemplo

Passo 1: $H_0 : \mu_A = \mu_B$ contra $H_1 : \mu_A \neq \mu_B$

Passo 2:

$$Z = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$$

Passo 3: Sob H_0 , temos $Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{50} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$,

$$RC = \{Z \in \mathbb{R} : |Z| < z_c\}$$

Logo,

$$0,025 = P(Z < z_1 | Z \sim N(0, 1)) \quad \text{e} \quad 0,025 = P(Z > z_2 | Z \sim N(0, 1)).$$

Assim, temos $z_1 = -1,96$ e $z_2 = 1,96$. E a regra de decisão é dada por

Rejeitar H_0 se $|Z| < 1,96$.

Exemplo

Passo 4: O valor observado da estatística é

$$z_0 = \frac{106.350 - 110.175}{25 \sqrt{\frac{1}{40} + \frac{1}{40}}} = -0.6842368$$

Passo 5: Como o valor observado de Z pertence a RC, rejeitamos H_0 , e concluímos que há evidências de que a média dos níveis de QI é a mesma entre meninos e meninas.

Importante

- Poderíamos construir um I.C. para a diferença $\theta = \mu_A - \mu_B$!!!
- Em que:

$$\text{IC}(\theta, \gamma) = \left(\bar{X}_A - \bar{X}_B \pm z_{\alpha/2} \sigma \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

Exemplo

```
1 library(BSDA)
2 z.test(x=qi_meninos, y=qi_meninas, mu=0, sigma.x=25, sigma.y=25, alternative = "two.sided")
```

Two-sample z-Test

```
data:  qi_meninos and qi_meninas
z = -0.68424, p-value = 0.4938
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.781532    7.131532
sample estimates:
mean of x mean of y
 106.350   110.175
```

Teste para Comparação das médias

CASO 2: Mesma variância, desconhecida.

- Neste caso, a estatística para o teste:

$$T = \frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2},$$

onde

$$S_p^2 = \frac{(n-1)S_A^2 + (m-1)S_B^2}{n+m-2}.$$

- Este teste é conhecido como **teste t para duas amostras!**

Teste para Comparação das médias

CASO 3: Variâncias desiguais e desconhecidas.

- Pode-se provar que a estatística

$$T = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{m}}},$$

sob H_0 , tem uma distribuição aproximadamente t-Student com graus de liberdade, dados aproximadamente por

$$v = \frac{(x + y)^2}{\frac{x^2}{n-1} + \frac{y^2}{m-1}},$$

com $x = \frac{S_A^2}{n}$ e $y = \frac{S_B^2}{m}$.

- Este teste é conhecido como **Problema de Behrens-Fisher!**.

Exemplo

- Os dados:

```
1 # Peso de homens e mulheres
2 peso.m <- c(38.9, 61.2, 73.3, 21.8, 63.4,
3            64.6, 48.4, 48.8, 48.5)
4 peso.h <- c(67.8, 60, 63.4, 76, 89.4,
5            73.3, 67.3, 61.3, 62.4)
6 # Criando um data frame
7 dados <- data.frame(
8     grupo = rep(c("Mulheres",
9                  "Homens"), ea
10    peso = c(peso.m, peso.h)
11    )
12 head(dados)
```

```
  grupo peso
1 Mulheres 38.9
2 Mulheres 61.2
3 Mulheres 73.3
4 Mulheres 21.8
5 Mulheres 63.4
6 Mulheres 64.6
```

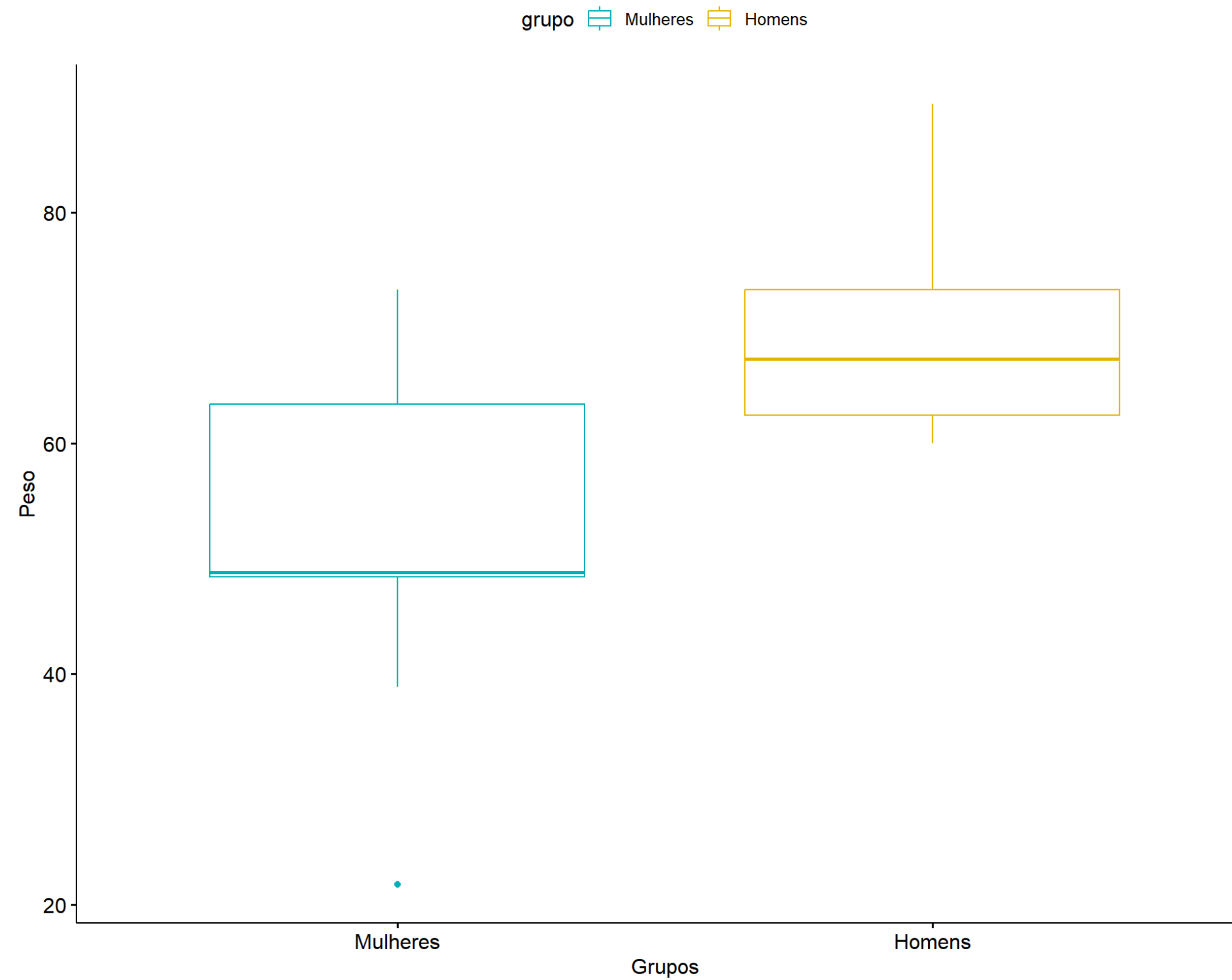
- Medidas de resumo:

```
1 library(dplyr)
2 group_by(dados, grupo) %>%
3   summarise(
4     freq = n(),
5     media = mean(peso, na.rm = TRUE),
6     vari = var(peso, na.rm = TRUE)
7   )
```

```
# A tibble: 2 × 4
  grupo      freq media  vari
  <chr>    <int> <dbl> <dbl>
1 Homens         9  69.0  87.9
2 Mulheres        9  52.1 243.
```


Exemplo

```
1 library("ggpubr")
2 ggboxplot(dados, x = "grupo", y = "peso",
3           color = "grupo", palette = c("#00AFBB", "#E7B800"),
4           ylab = "Peso", xlab = "Grupos")
```



Exemplo

A construção formal do teste.

Passo 1: $H_0 : \mu_A = \mu_B$ contra $H_1 : \mu_A \neq \mu_B$

Passo 2:

$$T = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{m}}}$$

Exemplo

Passo 3: Sob H_0 , temos $T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{n} + \frac{s_B^2}{m}}} \sim t_v$

$$RC = \{T \in \mathbb{R} : T < t_1 \text{ ou } T > t_2\},$$

com

$$0,025 = P(T < t_1 | T \sim t_v) \quad \text{e} \quad 0,025 = P(T > t_2 | T \sim t_v).$$

Agora,

$$v = \frac{((87.9/9) + (243/9))^2}{(87.9/9)^2/8 + (243/9)^2/8} = 13.1179798 \simeq 14$$

Portanto,

$$t_1 = -2.1447867 \quad \text{e} \quad t_2 = 2.1447867.$$

E a regra de decisão é dada por: Rejeitar H_0 se $T < -2.1447867$ ou $T > 2.1447867$.

Exemplo

Passo 4: O valor observado da estatística é

$$t_0 = \frac{69.0 - 52.1}{\sqrt{\frac{87.9}{9} + \frac{243}{9}}} = 2.7871451.$$

Passo 5: Como o valor observado de T pertence a RC, rejeitamos H_0 , e concluimos que há evidências de que homens e mulheres tem pesos diferentes.

Exemplo

- Caso 3:

```
(res <- t.test(peso ~ grupo, data = dados, var.equal = FALSE))
```

Welch Two Sample t-test

data: peso by grupo

t = 2.7842, df = 13.114, p-value = 0.01538

alternative hypothesis: true difference in means between group Homens and group Mulheres is not equal to 0

95 percent confidence interval:

3.795858 29.981920

sample estimates:

mean in group Homens	mean in group Mulheres
68.98889	52.10000

Exemplo

- Caso 2:

```
(res <- t.test(peso ~ grupo, data = dados, var.equal = TRUE))
```

Two Sample t-test

```
data:  peso by grupo
t = 2.7842, df = 16, p-value = 0.01327
alternative hypothesis: true difference in means between group Homens and group Mulheres is not equal to 0
95 percent confidence interval:
 4.029759 29.748019
sample estimates:
mean in group Homens mean in group Mulheres
      68.98889      52.10000
```

Obs: Para realizar o teste t com variâncias iguais temos que usar o argumento `var.equal=TRUE`.

Teste para Comparação das Variâncias

Teste para Comparação das Variâncias

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

A estatística do teste será

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{n-1, m-1}$$

Exemplo: Queremos verificar se duas máquinas produzem peças com a mesma homogeneidade quanto à resistência à tensão. Para isso, sorteamos duas amostras de seis peças de cada máquina, e obtivemos as seguintes resistências:

- Máquina A: 145, 127, 136, 142, 141, 137
- Máquina B: 143, 128, 132, 138, 142, 132

Exemplo

Passo 1: $H_0 : \sigma_A^2 = \sigma_B^2$ contra $H_1 : \sigma_A^2 \neq \sigma_B^2$

Passo 2:

$$F = \frac{S_A^2 / \sigma_A^2}{S_B^2 / \sigma_B^2} \sim F_{5,5}$$

Passo 3: Sob H_0 , temos que $F = \frac{S_A^2}{S_B^2} \sim F_{5,5}$

Fixando $\alpha = 0,05$, a RC é dada por

$$RC = \{F < F_1 \text{ ou } F > F_2\},$$

com F_1 e F_2 tais que

$$0,025 = P(F < F_1 | F \sim F_{5,5}) \quad \text{e} \quad 0,025 = P(F > F_2 | F \sim F_{5,5})$$

Exemplo

Temos então, $F_2 = 7,15$ e $F_1 = 1/7,15 = 0,14$. Assim, a regra de decisão é:

Rejeitar H_0 se $F < 0,14$ ou $F > 7,15$.

Passo 4: Com os dados apresentados, temos $S_A^2 = 40$ e $S_B^2 = 37$. Portanto, o valor observado da estatística é $F_o = 40/37 = 1,08$.

Passo 5: Como o valor observado da estatística não pertence a RC, aceitamos H_0 e concluimos que as máquinas produzem com a mesma variabilidade.

Exemplo

```
A <- c(145, 127, 136, 142, 141, 137)
B <- c(143, 128, 132, 138, 142, 132)
dados <- data.frame(grupo = rep(c("A", "B"), each=6), res = c(A, B))
head(dados)
```

	grupo	res
1	A	145
2	A	127
3	A	136
4	A	142
5	A	141
6	A	137

```
(res.ftest <- var.test(res ~ grupo, data = dados))
```

F test to compare two variances

```
data: res by grupo
F = 1.0821, num df = 5, denom df = 5, p-value = 0.9331
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1514131 7.7327847
sample estimates:
ratio of variances
      1.082056
```

Duas Populações Normais dependentes

Duas Populações Normais dependentes

Aqui temos duas amostras X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_n , só que agora as observações são pareadas, isto é, temos uma amostra de pares

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

Se definirmos a v.a. $D = X - Y$, teremos uma amostra D_1, D_2, \dots, D_n , resultante da diferença dos valores entre cada par. Reduzimos o problema de duas populações a um problema de uma única população, já visto anteriormente. Assim,

$$\overline{D} = \frac{1}{n} \sum_{i=1}^n D_i = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n Y_i = \overline{X} - \overline{Y}$$

terá distribuição $N(\mu_D, \sigma_D^2/n)$.

Duas Populações Normais dependentes :

Considerando

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2,$$

temos que

$$T = \frac{\sqrt{n}(\bar{D} - \mu_D)}{S_D} \sim t_{n-1}$$

Como $\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_1 - \mu_2$, testar $H_0 : \mu_D = 0$ é equivalente a testar $H_0 : \mu_1 = \mu_2$.

Exemplo

Exemplo: Cinco operadores de certo tipo de máquina são treinados em máquinas de duas marcas diferentes, A e B. Mediu-se o tempo em que cada um deles gasta na realização de uma mesma tarefa, e os resultados estão na tabela a seguir.

Operador	Marca A	Marca B
A	80	75
B	72	70
C	65	60
D	78	72
E	85	78

Ao nível de significância de 10%, poderíamos afirmar que a tarefa realizada na Máquina A demora mais que na Máquina B?

Exemplo

Passo 1:

$$H_0 : \mu_A = \mu_B \times H_1 : \mu_A > \mu_B$$

Essas hipóteses são equivalentes a

$$H_0 : \mu_D = 0 \times H_1 : \mu_D > 0$$

Passo 2:

$$T = \frac{\sqrt{n}(\bar{D} - \mu_D)}{S_D} \sim t_4$$

Passo 3: Como é o mesmo operador que realiza a tarefa nas duas máquinas, dizemos que as variáveis são emparelhadas. Sob H_0 , temos $T = \frac{\sqrt{n}\bar{D}}{S_D} \sim t_4$. Assim, com $\alpha = 0,10$, temos

$$P(T > t_c | T \sim t_4) = 0,10.$$

Portanto, $t_c = 1,533$, logo, a regra de decisão é: Rejeitar H_0 se $T > 1,533$.

Exemplo

Passo 4: Da Tabela de dados acima, obtemos os valores de D :

$$d_i : \quad 5, 2, 5, 6, 7$$

e, portanto,

$$\bar{d} = 5, \quad \text{e} \quad s_D^2 = 3,5$$

Logo, o valor observado da estatística \hat{e}

$$t_o = (\sqrt{5} \times 5) / \sqrt{3,5} = 5,98$$

Passo 5: Como o valor observado pertence a RC, rejeitamos H_0 , ou seja, demora-se mais para realizar a tarefa na máquina A.

Podemos construir um I.C. para μ_D , adotando $\gamma = 0,90$:

$$IC(\mu_D; 90\%) = \bar{D} \pm t_{\alpha/2} \times \sqrt{s_D^2} / \sqrt{n}$$

$$IC(\mu_D; 90\%) = 5 \pm 1,78 = [3,22 ; 6,78]$$

Exemplo

```
ma <- c(80, 72, 65, 78, 85)
mb <- c(75, 70, 60, 72, 78)
```

```
dados <- data.frame(grupo = rep(c("antes", "depois"), each=5), tempo = c(ma, mb) )
head(dados)
```

	grupo	tempo
1	antes	80
2	antes	72
3	antes	65
4	antes	78
5	antes	85
6	depois	75

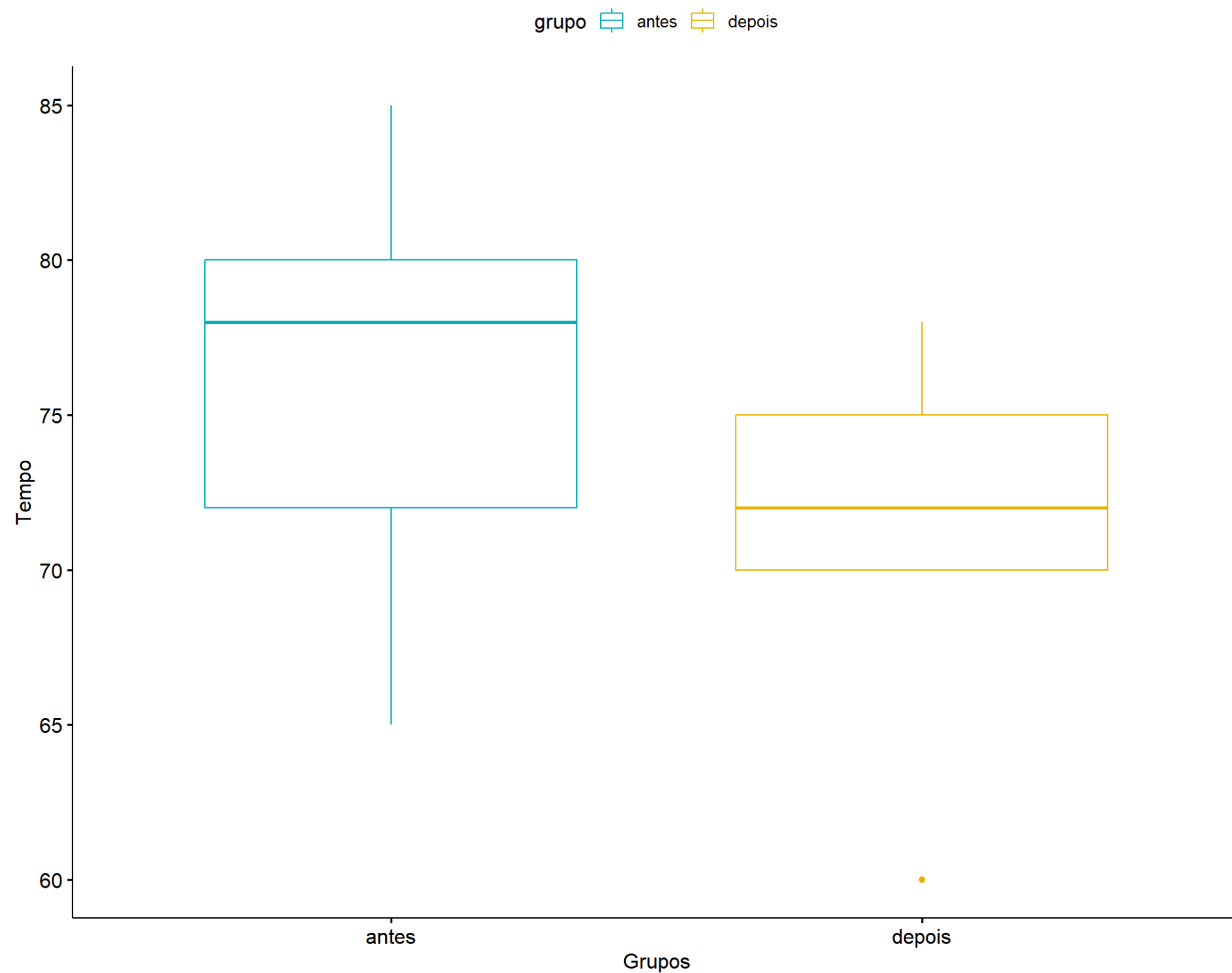
```
group_by(dados, grupo) %>%
  summarise(
    freq = n(),
    media = mean(tempo, na.rm = TRUE),
    desvio = sd(tempo, na.rm = TRUE)
  )
```

A tibble: 2 × 4

	grupo	freq	media	desvio
	<chr>	<int>	<dbl>	<dbl>
1	antes	5	76	7.71
2	depois	5	71	6.86

Exemplo

```
ggboxplot(dados, x = "grupo", y = "tempo",  
          color = "grupo",  
          palette = c("#00AFBB", "#E7B800"),  
          order = c("antes", "depois"),  
          ylab = "Tempo", xlab = "Grupos")
```



```
t.test(x=ma, y=mb, alternative = "greater",  
       paired = TRUE)
```

Paired t-test

data: ma and mb

t = 5.9761, df = 4, p-value = 0.00197

alternative hypothesis: true mean difference is
greater than 0

95 percent confidence interval:

3.216369 Inf

sample estimates:

mean difference

5

Testes Não-paramétricos

Teste de Aderência de Kolmogorov-Smirnov:

Objetivo: Testar se a amostra observada provém de um modelo de probabilidade especificado, com função de distribuição acumulada (f.d.a.) $F_0(x)$.

- X_1, \dots, X_n : a.a. da v.a. X , cujo modelo de probabilidade é desconhecido.
- Seja $f(x)$ a f.d.p. (ou f.p.) da v.a. X e $F(x) = P(X \leq x)$ a sua f.d.a. Queremos testar a seguinte hipótese:

$$H_0 : F(x) = F_0(x), \quad \text{para todo } x$$

- Um possível estimador para $F(x)$ é a função de distribuição empírica (f.d.e.), definida por

$$F_e(x) = \frac{N(x)}{n},$$

onde $N(x)$ é o número de observações menores ou iguais a x , e n é o número total de observações.

- Se $F_e(x)$ for um bom estimador de $F(x)$, então as duas curvas devem estar próximas.

Teste de Aderência de Kolmogorov-Smirnov:

ⁱ Passos de construção do teste:

1. Ordenar os valores da amostra;
2. Obter os valores de $F(x_i)$, a f.d.a. especificada na hipótese H_0 , para cada observação x_i da amostra;
3. Obter os valores de $F_e(x_i) = \frac{i}{n}$, a f.d.a. empírica para cada observação x_i da amostra;
4. Obter as diferenças $|F(x_i) - F_e(x_i)|$;
5. Obter o valor da diferença máxima, $D_{obs} = \max |F(x_i) - F_e(x_i)|$;
6. Obter na Tabela da distribuição de Kolmogorov-Smirnov o valor tabelado da variável D , para um nível fixado ($\alpha = 0,05$ ou $\alpha = 0,01$), D_{tab} ;
7. Rejeitar H_0 se $D_{obs} > D_{tab}$, caso contrário, aceitar H_0 .

Determinação do Valor Tabelado para o Teste de Kolmogorov-Smirnov

- Para Amostras Pequenas ($n \leq 50$) tem-se a seguinte tabela de Valores Críticos:

n	$D_{0.05}$	$D_{0.01}$
5	0.565	0.669
10	0.409	0.490
20	0.294	0.352
30	0.242	0.290
50	0.188	0.226

- Fórmula Aproximada

$$D_{\alpha} \approx \frac{c(\alpha)}{\sqrt{n}}$$

Determinação do Valor Tabelado para o Teste de Kolmogorov-Smirnov

- Para Amostras Grandes ($n > 50$), temos a aproximação Assintótica:

$$D_{\alpha} = \frac{c(\alpha)}{\sqrt{n}}$$

- Valores de $c(\alpha)$:

- $\alpha = 0.05 \Rightarrow \frac{1.36}{\sqrt{n}}$

- $\alpha = 0.01 \Rightarrow \frac{1.63}{\sqrt{n}}$

- A fórmula geral para qualquer tamanho de amostra:

$$D_{\alpha} = \sqrt{-\frac{\ln(\alpha/2)}{2n}}$$

Exemplo

Objetivo: Verificar se uma amostra de dados segue uma distribuição normal padrão $\mathcal{N}(0, 1)$ utilizando o teste de aderência de Kolmogorov-Smirnov.

- Amostra de tamanho $n = 5$:

$$X = \{-0.5, 0.1, 0.3, 0.8, 1.2\}$$

- Passo a Passo do Teste:
 - Passo 1: Ordenar os dados

$$X_{\text{ordenado}} = \{-0.5, 0.1, 0.3, 0.8, 1.2\}$$

Exemplo

- Passo 2: Calcular $F(x_i)$ (f.d.a. teórica sob H_0) Assumindo $H_0 : F(x) = \Phi(x)$ (distribuição normal padrão):

$$F(-0.5) = \Phi(-0.5) \approx 0.3085$$

$$F(0.1) = \Phi(0.1) \approx 0.5398$$

$$F(0.3) = \Phi(0.3) \approx 0.6179$$

$$F(0.8) = \Phi(0.8) \approx 0.7881$$

$$F(1.2) = \Phi(1.2) \approx 0.8849$$

- Passo 3: Calcular $F_e(x_i)$ (f.d.a. empírica)

$$F_e(x_i) = \frac{i}{n} = \begin{cases} F_e(-0.5) = \frac{1}{5} = 0.2 \\ F_e(0.1) = \frac{2}{5} = 0.4 \\ F_e(0.3) = \frac{3}{5} = 0.6 \\ F_e(0.8) = \frac{4}{5} = 0.8 \\ F_e(1.2) = \frac{5}{5} = 1.0 \end{cases}$$

Exemplo

- Passo 4: Calcular as diferenças $|F(x_i) - F_e(x_i)|$

$$|F(-0.5) - F_e(-0.5)| = 0.1085$$

$$|F(0.1) - F_e(0.1)| = 0.1398$$

$$|F(0.3) - F_e(0.3)| = 0.0179$$

$$|F(0.8) - F_e(0.8)| = 0.0119$$

$$|F(1.2) - F_e(1.2)| = 0.1151$$

- Passo 5: Encontrar a diferença máxima D_{obs}

$$D_{\text{obs}} = \max\{0.1085, 0.1398, 0.0179, 0.0119, 0.1151\} = 0.1398$$

- Passo 6: Valor crítico D_{tab} ($\alpha = 0.05$) Para $n = 5$ e $\alpha = 0.05$:

$$D_{\text{tab}} \approx 0.565$$

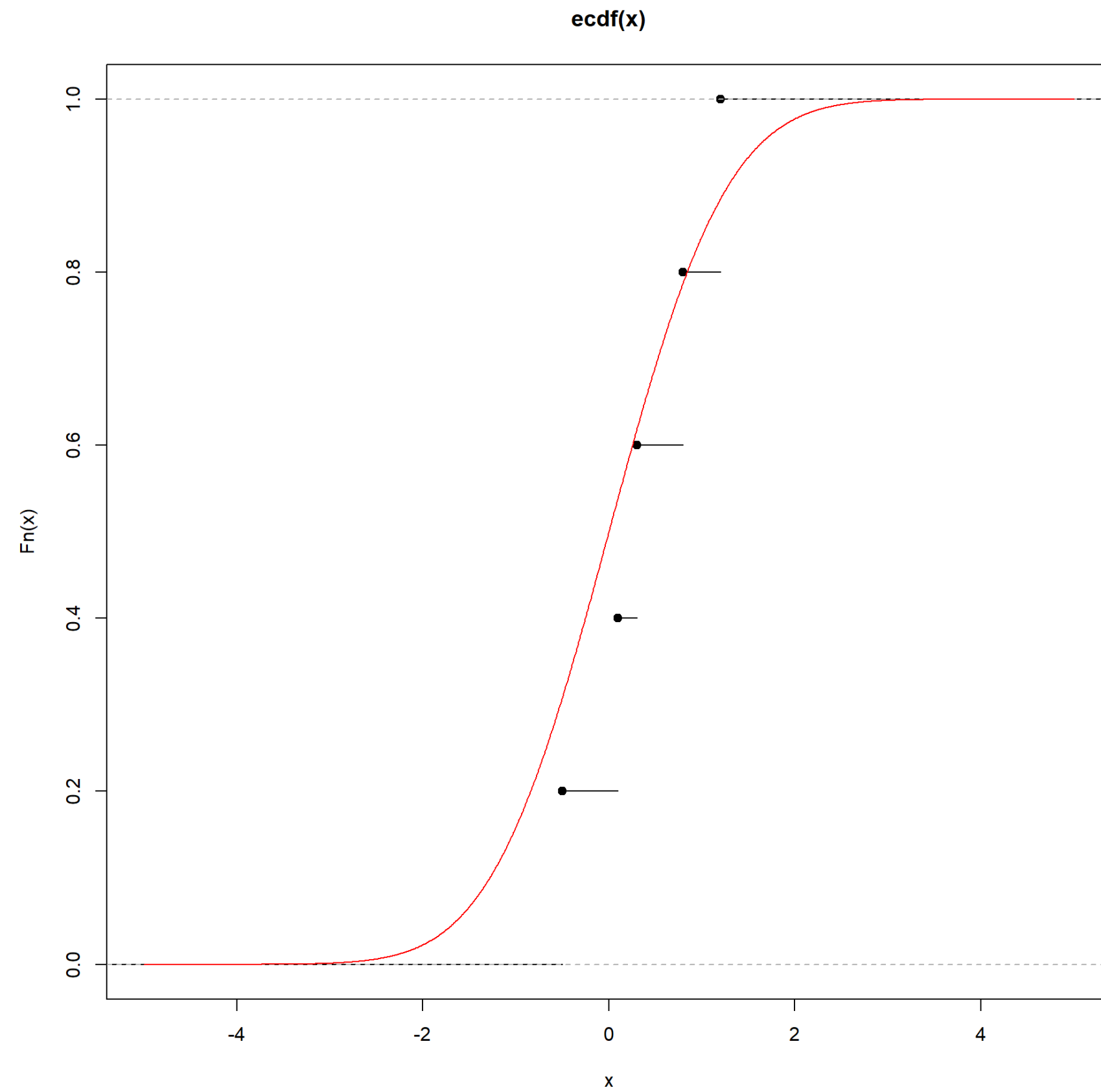
Exemplo

- Passo 7: Decisão Como $D_{\text{obs}} = 0.1398 < D_{\text{tab}} = 0.565$, não rejeitamos H_0 .
- **Conclusão:** Não há evidências estatísticas para afirmar que a amostra não segue a distribuição normal padrão $N(0, 1)$.

Observações Adicionais:

1. **Amostras maiores:** Para $n > 50$, utilize aproximações assintóticas (ex: $D_{\text{tab}} \approx \frac{1.36}{\sqrt{n}}$ para $\alpha = 0.05$).
2. **Outras distribuições:** O teste pode ser adaptado para qualquer distribuição contínua (exponencial, uniforme, etc.).
3. O teste é mais conservativo para amostras pequenas.
4. $F_0(x)$ deve ser completamente especificada.

Exemplo



Exact one-sample Kolmogorov-Smirnov test

```
data: x
D = 0.33983, p-value = 0.5079
alternative hypothesis: two-sided
```

Testes de Normalidade Mais Usuais

Teste de Shapiro-Wilk

- **Indicado para:** Amostras pequenas a moderadas ($n \leq 50$, mas pode ser usado até $n \approx 2000$).
- **Hipótese nula (H_0):** Os dados seguem uma distribuição normal.
- **Características:**
 - Muito sensível a desvios da normalidade.
 - **Recomendado quando** se precisa de maior poder estatístico em amostras pequenas.
- **Vantagem:** Um dos mais poderosos para n pequeno.
- **Desvantagem:** Perde eficiência com amostras grandes.

Kolmogorov-Smirnov (K-S)

- **Indicado para:** Amostras de qualquer tamanho, mas menos poderoso para detectar pequenas violações.
- **Hipótese nula (H_0):** A distribuição empírica dos dados é igual à normal.
- **Características:**
 - Baseia-se na maior distância entre a função de distribuição empírica e a teórica.
 - A versão com correção de **Lilliefors** é usada quando os parâmetros da normal são estimados da amostra.
- **Vantagem:** Aplicável a várias distribuições.
- **Desvantagem:** Menos sensível que Shapiro-Wilk.

Anderson-Darling

- **Indicado para:** Amostras pequenas a grandes.
- **Hipótese nula (H_0):** Os dados seguem uma distribuição normal.
- **Características:**
 - **Dá mais peso às caudas** da distribuição (útil se há suspeitas de problemas nas caudas).
- **Vantagem:** Mais sensível a desvios nas extremidades (caudas).
- **Desvantagem:** Nem sempre retorna um valor-p diretamente (depende do software).

Lilliefors

- **Descrição:** É uma adaptação do teste K-S, usada quando a média e desvio padrão são estimados da amostra.
- **Usado frequentemente em:** Softwares como SPSS.
- **Similaridades:** Mesmo princípio do K-S, mas com ajuste.

Jarque-Bera

- **Indicado para:** Amostras grandes.
- **Hipótese nula (H_0):** Os dados têm **assimetria (skewness)** e **curtose (kurtosis)** compatíveis com a normal.
- **Características:**
 - Usa estatísticas de **assimetria e curtose** para verificar desvio da normalidade.
- **Vantagem:** Simples e computacionalmente leve.
- **Desvantagem:** Fraco com amostras pequenas; ignora outros aspectos da forma da distribuição.

Exercícios:

Baixe o banco de dados [dadosInf2_ex.xlsx](#) no repositório do github e responda as seguintes perguntas:

1. O peso médio antes do tratamento difere entre os participantes do grupo **Controle** e do grupo **Tratamento**?
 2. A média da pressão arterial **antes do tratamento** difere entre os **homens** e as **mulheres**?
 3. Existe diferença na altura média entre os participantes do grupo **Controle** e os do grupo **Tratamento**?
 4. Há diferença significativa na idade média entre **homens** e **mulheres**?
 5. O peso médio dos participantes mudou após o tratamento?
 6. O tratamento foi eficaz em reduzir a pressão arterial dos participantes?
- Coloque as respostas no formato relatório usando o [quarto](#), por exemplo.