EXERCICIOS_Pos_3_10_20.R

rstudio-user

2020-10-12

```r
# Modelo binomial

# 1) Qual é a probabilidade de se obter 3 caras em 5 lançamentos de uma moeda honesta?
x <- 3
n <- 5
p <- 0.5
dbinom(x, n, p)
```

```
## [1] 0.3125
```

```r
# 2) Qual é a probabilidade de se obter menos que 3 caras em 5 lançamentos da mesma moeda?
n <- 5
p <- 0.5
dbinom(0, n, p) + dbinom(1, n, p) + dbinom(2, n, p)
```

```
## [1] 0.5
```

```r
# 3) Se a probabilidade de atingir um alvo num único disparo é 0,3, qual é a probabilidade de que em
# 4 disparos o alvo seja atingido no mínimo 3 vezes?
x <- 3
n <- 4
p <- 0.3
dbinom(x, n, p)
```

```
## [1] 0.0756
```

```r
# Modelo Poisson

# 1) Um departamento de polícia recebe em média 5 solicitações por hora. Qual a probabilidade de
# receber 2 solicitações numa hora selecionada aleatoriamente?
dpois(2,5)
```

```
## [1] 0.08422434
```

```r
# 2) A experiência passada indica que um número médio de 6 clientes por hora para
# para colocar gasolina numa bomba.

# a) Qual é a probabilidade de 3 clientes pararem qualquer hora?
dpois(3,6)
```

```
## [1] 0.08923508
```

```r
# b) Qual é a probabilidade de 3 clientes ou menos pararem em qualquer hora?
dpois(0,6) + dpois(1,6) + dpois(2,6) + dpois(3,6)
```

```
## [1] 0.1512039
```

```r
# Distribuição normal

# 3) A duração de um certo componente eletrônico tem média de 859 dias e desvio padrão de 40
# dias. Sabendo que a duração é normalmente distribuída, calcule a probabilidade desse componente
# durar:

# a) Entre 700 e 1000 dias
E <- 1 - pnorm(700, 859, 40)
E
```

```
## [1] 0.9999648
```

```r
F <- 1 - pnorm(1000, 859, 40)
F
```

```
## [1] 0.0002117414
```

```r
P <- E - F
P
```

```
## [1] 0.9997531
```

```r
# b) Mais de 800 dias
G <- 1 - pnorm(800, 859, 40)
G
```

```
## [1] 0.9298937
```

```r
# 1) Um artigo do periódico Materials Engineering descreve os resultados de teste de tensão quanto
# à adesão em 22 corpos de prova de liga U-700. A carga no ponto de falha do corpo de prova é
# dada pelo arquivo "carga no ponto de falha.txt". Verifique se os dados sugerem que a carga
# média na falha excede 10 MPa. Considere o nível de significância de 5%. Interprete o resultado.
library(readr)
Carga_no_ponto_de_falha <- read_table2("Carga_no_ponto_de_falha.txt")
```

```
## Parsed with column specification:
## cols(
##   Carga = col_double()
## )
```

```r
t.test(
  Carga_no_ponto_de_falha$Carga,
  mu = 10,
  alternative = "greater"
)
```

```
##
##  One Sample t-test
##
## data:  Carga_no_ponto_de_falha$Carga
## t = 4.9017, df = 21, p-value = 3.781e-05
## alternative hypothesis: true mean is greater than 10
## 95 percent confidence interval:
##  12.40996      Inf
## sample estimates:
## mean of x
##  13.71364
```

```r
# Resposta: Utilizando o teste t é possível verificar que o valor de p ficou abaixo do nível de
# significância utilizado (0.05). Nossa hipótese alternativa indica que a média excede 10 MPa,
# como o valor de p ficou abaixo do nível de significância, podemos considerar a hipótese
# alternativa como verdadeira, ou seja, a carga média na falha excede 10 MPa.

# 2) Dois catalisadores estão sendo analisados para determinar como eles afetam o rendimento
# médio de um processo químico. Especificamente, o catalisador 1 está corretamente em uso,
# mas o catalisador 2 é aceitável. Uma vez que o catalisador 2 é mais barato, ele deve ser
# adotado, desde que ele não mude o rendimento do processo. Um teste é feito em uma planta
# piloto, resultando nos dados do arquivo "catalisadores". Há alguma diferença entre os
# rendimentos médios? Use a = 0,05 e considere variâncias iguais. Formule antes as hipóteses.
library(readxl)
catalisadores <- read_excel(
  "catalisadores.xlsx",
  col_types = c("numeric", "numeric")
)
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A2 / R2C1: '91.5'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B2 / R2C2: '89.19'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A3 / R3C1: '94.18'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B3 / R3C2: '90.95'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A4 / R4C1: '92.18'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B4 / R4C2: '90.46'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A5 / R5C1: '95.39'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B5 / R5C2: '93.21'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A6 / R6C1: '91.79'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B6 / R6C2: '97.19'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A7 / R7C1: '89.07'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B7 / R7C2: '97.04'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A8 / R8C1: '94.72'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B8 / R8C2: '91.07'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in A9 / R9C1: '89.21'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in B9 / R9C2: '92.75'
```

```r
t.test(
  catalisadores$catalisador_1,
  catalisadores$catalisador_2,
  conf.level = 0.95
)
```

```
##
##  Welch Two Sample t-test
##
## data:  catalisadores$catalisador_1 and catalisadores$catalisador_2
## t = -0.35359, df = 13.353, p-value = 0.7292
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.387118  2.432118
## sample estimates:
## mean of x mean of y
##   92.2550   92.7325
```

```r
# Resposta: Temos duas hipóteses, H0 indica que não há diferença nos rendimentos médios e nossa
# hipótese alternativa H1 indica que há diferença. Foi utilizado o teste t nas duas amostras dos
# catalisadores 1 e 2, com nível de significância de 0.05. O resultado é o valor de p maior que o
# nível de significância 0.05, sendo assim podemos afirmar que nossa hipótese alternativa H1 não
# é verdadeira e que não há diferença nos rendimentos médios.


# 3) Uma companhia fabrica propulsores para uso em motores de turbinas de avião. Uma das
# operações envolve esmerilhar o acabamento de uma superfície particular para um componente
# de liga de titânio. Dois processos diferentes para esmerilhar podem ser usados, podendo
# produzir peças com iguais rugosidades médias na superfície. Uma amostra aleatória de n1 =
# 11 peças, proveniente do primeiro processo, resulta em um desvio-padrão de s1 = 5,1
# micropolegadas. Uma amostra aleatória de n2 = 16 peças, proveniente do segundo processo,
# resulta em um desvio-padrão de s2 = 4,7 micropolegadas. Verifique se a razão entre as duas
# variâncias é diferente de 1 com um nível de confiança de 90%. Considere que os dois processos
# sejam diferentes e a rugosidade na superfície seja normalmente distribuída.
amostraN1 <- rnorm(11, sd = 5.1)
amostraN2 <- rnorm(16, sd = 4.7)

var.test(
  amostraN1,
  amostraN2,
  ratio = 1,
  alternative = "t",
  conf.level = 0.9
)
```

```
##
##  F test to compare two variances
##
## data:  amostraN1 and amostraN2
## F = 0.86892, num df = 10, denom df = 15, p-value = 0.8428
## alternative hypothesis: true ratio of variances is not equal to 1
## 90 percent confidence interval:
##  0.3415962 2.4721045
```

```
## sample estimates:
## ratio of variances
##          0.8689245
```

```r
# Resposta: As hipóteses para a razão entre as duas variâncias são H0 igual a 1 e H1 diferente de 1.
# Como o valor de p é maior que 0.10 sendo que o nível de confiança é 90%, então não podemos rejeitar
# a hipótese H0, sendo que podemos considerar que a razão entre as duas variâncias é 1.


# 4) Verifique se os dados do arquivo "carga no ponto de falha" segue distribuição normal.
shapiro.test(Carga_no_ponto_de_falha$Carga)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Carga_no_ponto_de_falha$Carga
## W = 0.96981, p-value = 0.7067
```

```r
# Resposta: Segue distribuição normal (pois o valor de p é maior que 0.05, o que indica que
# hipótese nula é verdadeira)

# 5) Verifique se os dados do arquivo "catalisadores" seguem distribuição normal.
catalisadores1e2 <- c(catalisadores$catalisador_1, catalisadores$catalisador_2)
shapiro.test(catalisadores$catalisador_1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  catalisadores$catalisador_1
## W = 0.92171, p-value = 0.4439
```

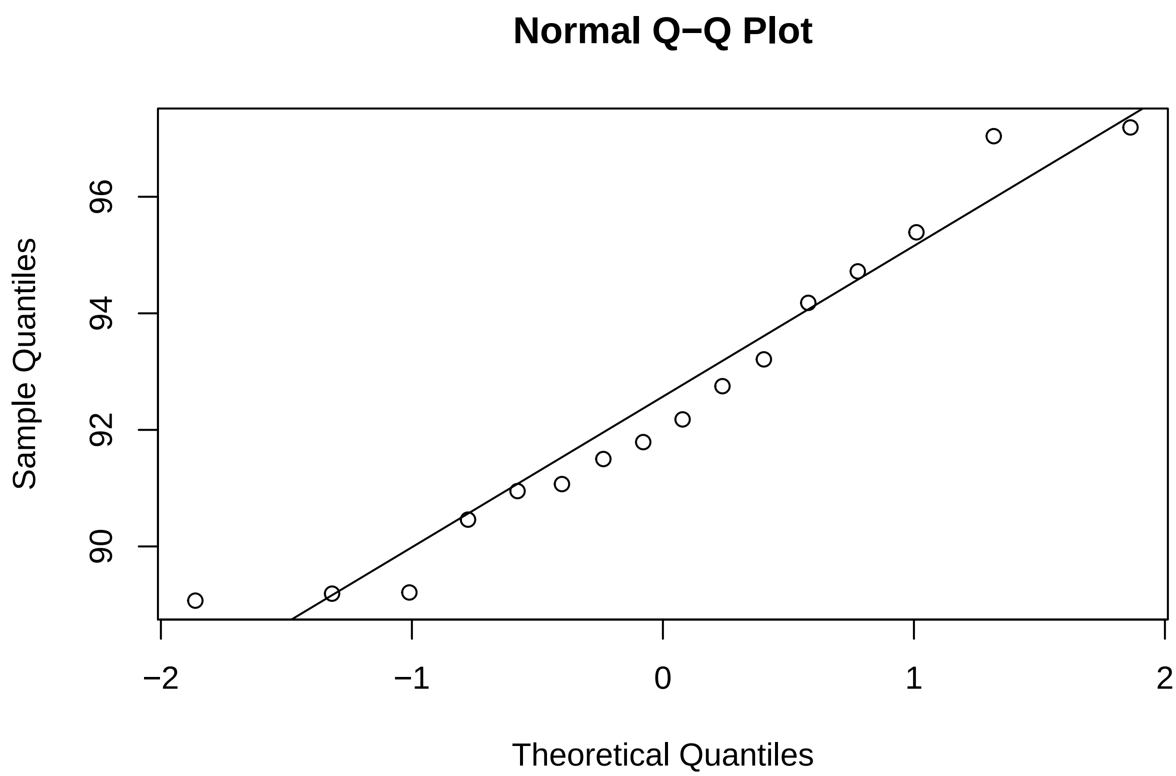
```r
shapiro.test(catalisadores$catalisador_2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  catalisadores$catalisador_2
## W = 0.88182, p-value = 0.196
```

```r
shapiro.test(catalisadores1e2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  catalisadores1e2
## W = 0.93907, p-value = 0.3378
```

```r
# Os dados do catalisador 1 e 2 separadamente ou juntos seguem distribuição normal.

qqnorm(catalisadores1e2)
qqline(catalisadores1e2)
```




```r
# ANOVA parte 1

# 1) Uma companhia deseja testar quatro tipos diferentes de válvulas: A, B, C e D. As vidas médias,
# em horas, constam na tabela abaixo. Cada tipo foi testado, aleatoriamente, em seis aparelhos
# idênticos.
# A  B  C  D
# 53 52 51 49
# 58 60 57 54
# 56 52 55 52
# 60 58 53 50
# 51 50 54 53
# 55 54 50 51

# a) Crie o objeto "vida_media" e "válvula" (este como fator).
vida_media <- c(
  53, 52, 51, 49,
  58, 60, 57, 54,
  56, 52, 55, 52,
  60, 58, 53, 50,
  51, 50, 54, 53,
  55, 54, 50, 51
)
vida_media
```

```
##  [1] 53 52 51 49 58 60 57 54 56 52 55 52 60 58 53 50 51 50 54 53 55 54 50 51
```

```r
valvula <- factor(rep(paste("valvula", LETTERS[1:4], sep = ""), 6))
valvula
```

```
##  [1] valvulaA valvulaB valvulaC valvulaD valvulaA valvulaB valvulaC valvulaD
##  [9] valvulaA valvulaB valvulaC valvulaD valvulaA valvulaB valvulaC valvulaD
```

```
## [17] valvulaA valvulaB valvulaC valvulaD valvulaA valvulaB valvulaC valvulaD
## Levels: valvulaA valvulaB valvulaC valvulaD
```

```r
# b) Teste se há diferença significativa entre as válvulas, ao nível de 5%. Interprete o resultado.
resultado <- aov(vida_media~valvula)
anova(resultado)
```

```
## Analysis of Variance Table
##
## Response: vida_media
##           Df  Sum Sq Mean Sq F value Pr(>F)
## valvula    3  51.667 17.2222  1.9171 0.1593
## Residuals 20 179.667  8.9833
```

```r
# Resposta: Como o valor de p foi maior que 0.05, então não existe diferença significativa
# entre as válvulas

# c) Crie um data.frame com "válvula" e "vida_media".
df <- data.frame(vida_media = vida_media, valvula = valvula)
df
```

```
##    vida_media  valvula
## 1          53 valvulaA
## 2          52 valvulaB
## 3          51 valvulaC
## 4          49 valvulaD
## 5          58 valvulaA
## 6          60 valvulaB
## 7          57 valvulaC
## 8          54 valvulaD
## 9          56 valvulaA
## 10         52 valvulaB
## 11         55 valvulaC
## 12         52 valvulaD
## 13         60 valvulaA
## 14         58 valvulaB
## 15         53 valvulaC
## 16         50 valvulaD
## 17         51 valvulaA
## 18         50 valvulaB
## 19         54 valvulaC
## 20         53 valvulaD
## 21         55 valvulaA
## 22         54 valvulaB
## 23         50 valvulaC
## 24         51 valvulaD
```

```r
# d) Ordene o data.frame de acordo com o nome da válvula.
df[order(df$valvula),]
```

```
##    vida_media  valvula
## 1          53 valvulaA
## 5          58 valvulaA
## 9          56 valvulaA
## 13         60 valvulaA
## 17         51 valvulaA
## 21         55 valvulaA
```

```
## 2          52 valvulaB
## 6          60 valvulaB
## 10         52 valvulaB
## 14         58 valvulaB
## 18         50 valvulaB
## 22         54 valvulaB
## 3          51 valvulaC
## 7          57 valvulaC
## 11         55 valvulaC
## 15         53 valvulaC
## 19         54 valvulaC
## 23         50 valvulaC
## 4          49 valvulaD
## 8          54 valvulaD
## 12         52 valvulaD
## 16         50 valvulaD
## 20         53 valvulaD
## 24         51 valvulaD
```

```r
# 2) São feitas cinco misturas da mesma liga metálica e para cada mistura foram efetuadas seis
# medidas de densidade.
# A   B   C   D   E
# 3,6 3,3 3,5 3,5 3,7
# 3,5 3,5 3,3 3,4 3,4
# 3,7 3,4 3,4 3,0 3,6
# 3,1 3,2 3,4 3,3 3,5
# 3,1 3,4 3,3 3,3 3,6
# 3,2 3,4 3,2 3,8 3,4

# Há evidência de que certas misturas tenham densidade média maior do que de outras? Adote a = 5%.
medidasDensidade <- c(
  3.6, 3.3, 3.5, 3.5, 3.7,
  3.5, 3.5, 3.3, 3.4, 3.4,
  3.7, 3.4, 3.4, 3.0, 3.6,
  3.1, 3.2, 3.4, 3.3, 3.5,
  3.1, 3.4, 3.3, 3.3, 3.6,
  3.2, 3.4, 3.2, 3.8, 3.4
)
medidasDensidade
```

```
##  [1] 3.6 3.3 3.5 3.5 3.7 3.5 3.5 3.3 3.4 3.4 3.7 3.4 3.4 3.0 3.6 3.1 3.2 3.4 3.3
## [20] 3.5 3.1 3.4 3.3 3.3 3.6 3.2 3.4 3.2 3.8 3.4
```

```r
misturas <- factor(rep(paste("mistura", LETTERS[1:5], sep = ""), 6))
misturas
```

```
##  [1] misturaA misturaB misturaC misturaD misturaE misturaA misturaB misturaC
##  [9] misturaD misturaE misturaA misturaB misturaC misturaD misturaE misturaA
## [17] misturaB misturaC misturaD misturaE misturaA misturaB misturaC misturaD
## [25] misturaE misturaA misturaB misturaC misturaD misturaE
## Levels: misturaA misturaB misturaC misturaD misturaE
```

```r
resultado <- aov(medidasDensidade~misturas)
anova(resultado)
```

```
## Analysis of Variance Table
##
```

```
## Response: medidasDensidade
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## misturas   4 0.13667 0.034167   0.967 0.4429
## Residuals 25 0.88333 0.035333
```

```r
# Resposta: Como o valor de p foi maior que 0.05, então não há evidência de que certas misturas
# tenham densidade média maior do que de outras.


# 3) Os dados a seguir, representam, em segundos, o tempo gasto por cinco operários para realizar
# certa tarefa, usando três máquinas diferentes. Considerando a = 5%, verifique se há diferenças
# entre as máquinas e entre os operários.

# Máquinas
# Operários A  B  C
# 1         40 59 42
# 2         39 55 51
# 3         47 55 45
# 4         45 50 40
# 5         52 52 41

dados <- c(
  40, 59, 42,
  39, 55, 51,
  47, 55, 45,
  45, 50, 40,
  52, 52, 41
)

operarios <- gl(5, 3, label = c(paste("operario", 1:5)))
maquinas <- rep(paste("maquina", LETTERS[1:3]), 5)
tabela <- data.frame(
  operarios = operarios,
  maquinas = factor(maquinas),
  dados = dados
)
tabela
```

```
##     operarios  maquinas dados
## 1  operario 1 maquina A    40
## 2  operario 1 maquina B    59
## 3  operario 1 maquina C    42
## 4  operario 2 maquina A    39
## 5  operario 2 maquina B    55
## 6  operario 2 maquina C    51
## 7  operario 3 maquina A    47
## 8  operario 3 maquina B    55
## 9  operario 3 maquina C    45
## 10 operario 4 maquina A    45
## 11 operario 4 maquina B    50
## 12 operario 4 maquina C    40
## 13 operario 5 maquina A    52
## 14 operario 5 maquina B    52
## 15 operario 5 maquina C    41
```

```r
resultado <- aov(
  dados~operarios+maquinas,
  tabela
)
resultado
```

```
## Call:
##    aov(formula = dados ~ operarios + maquinas, data = tabela)
##
## Terms:
##                 operarios maquinas Residuals
## Sum of Squares    30.4000 334.9333  208.4000
## Deg. of Freedom         4        2         8
##
## Residual standard error: 5.10392
## Estimated effects may be unbalanced
```

```r
anova(resultado)
```

```
## Analysis of Variance Table
##
## Response: dados
##           Df Sum Sq Mean Sq F value  Pr(>F)
## operarios  4  30.40    7.60  0.2917 0.87539
## maquinas   2 334.93  167.47  6.4287 0.02164 *
## Residuals  8 208.40   26.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Resposta: Entre os operários não há diferença, pois o valor de p ficou acima de 0.05. Entretanto,
# há diferença entre as máquinas, na qual o valor de p ficou menor que 0.05


# 4) Plantam-se quatro tipos diferentes de sementes de café em cinco tipos de solo. Cada solo é
# dividido em quatro lotes, pelos quais se distribuem, aleatoriamente, os quatro tipos de
# sementes. Ao nível de 5%, teste se a produção varia devido ao solo e/ou devido à variedade
# do café.

# Tipos de café
# SOLO  I  II III IV
# A     15 12 10  14
# B     19 15 12  11
# C     18 14 15  12
# D     16 11 12  16
# E     17 16 11  14

dados <- c(
  15, 12, 10,  14,
  19, 15, 12,  11,
  18, 14, 15,  12,
  16, 11, 12,  16,
  17, 16, 11,  14
)

solos <- gl(5, 4, label = c(paste("solo", LETTERS[1:5])))
```

```r
tiposCafe <- rep(paste("tipo_cafe", c(1:4)))

tabela <- data.frame(
  solos = solos,
  tiposCafe = factor(tiposCafe),
  dados = dados
)
tabela
```

```
##     solos   tiposCafe dados
## 1  solo A tipo_cafe 1    15
## 2  solo A tipo_cafe 2    12
## 3  solo A tipo_cafe 3    10
## 4  solo A tipo_cafe 4    14
## 5  solo B tipo_cafe 1    19
## 6  solo B tipo_cafe 2    15
## 7  solo B tipo_cafe 3    12
## 8  solo B tipo_cafe 4    11
## 9  solo C tipo_cafe 1    18
## 10 solo C tipo_cafe 2    14
## 11 solo C tipo_cafe 3    15
## 12 solo C tipo_cafe 4    12
## 13 solo D tipo_cafe 1    16
## 14 solo D tipo_cafe 2    11
## 15 solo D tipo_cafe 3    12
## 16 solo D tipo_cafe 4    16
## 17 solo E tipo_cafe 1    17
## 18 solo E tipo_cafe 2    16
## 19 solo E tipo_cafe 3    11
## 20 solo E tipo_cafe 4    14
```

```r
resultado <- aov(
  dados~solos+tiposCafe,
  tabela
)
resultado
```

```
## Call:
##    aov(formula = dados ~ solos + tiposCafe, data = tabela)
##
## Terms:
##                  solos tiposCafe Residuals
## Sum of Squares   10.0      67.6      46.4
## Deg. of Freedom     4         3        12
##
## Residual standard error: 1.966384
## Estimated effects may be unbalanced
```

```r
anova(resultado)
```

```
## Analysis of Variance Table
##
## Response: dados
##           Df Sum Sq Mean Sq F value  Pr(>F)
## solos      4   10.0  2.5000  0.6466 0.63989
## tiposCafe  3   67.6 22.5333  5.8276 0.01075 *
```

```
## Residuals 12   46.4  3.8667
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Resposta: Entre os diferentes solos não há diferença na variação da produção, pois o valor de p
# ficou acima de 0.05. Entretanto, há variação entre os tipos de café, na qual o valor de p ficou
# menor que 0.05.


# ANOVA parte 2

# 1) Considere os dados de índice de mudança de cor para um experimento para comparar duas
# marcas de caneta e quatro diferentes tratamentos de lavagem em relação à capacidade de
# remover tinta de um determinado tipo de tecido. Há diferenças ao nível de 5% devido aos
# tratamentos de lavagem? Plote o gráfico do teste Tukey e interprete o resultado.

#    A    B
# T1 0,97 0,78
# T2 0,68 0,76
# T3 0,10 0,14
# T4 0,15 0,05

dados <- c(
  0.97, 0.78,
  0.68, 0.76,
  0.10, 0.14,
  0.15, 0.05
)

tratamentos <- gl(4, 2, label = c(paste("tratamento", c(1:4))))
marcas <- rep(paste("marca", LETTERS[1:2]))

tabela <- data.frame(
  tratamentos = tratamentos,
  marcas = factor(marcas),
  dados = dados
)
tabela
```

```
##    tratamentos  marcas dados
## 1 tratamento 1 marca A  0.97
## 2 tratamento 1 marca B  0.78
## 3 tratamento 2 marca A  0.68
## 4 tratamento 2 marca B  0.76
## 5 tratamento 3 marca A  0.10
## 6 tratamento 3 marca B  0.14
## 7 tratamento 4 marca A  0.15
## 8 tratamento 4 marca B  0.05
```

```r
resultadoAov <- aov(
  dados~tratamentos,
  tabela
)
resultadoAov
```

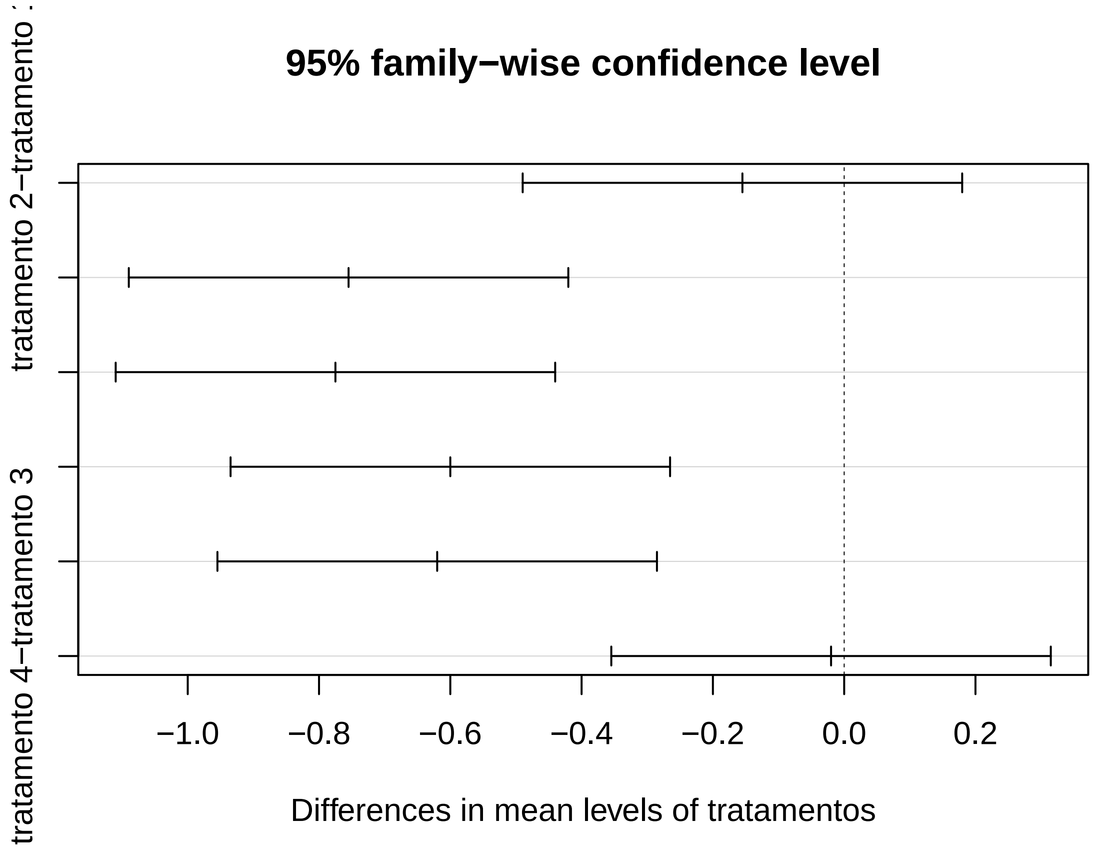
```
## Call:
```

```
##    aov(formula = dados ~ tratamentos, data = tabela)
##
## Terms:
##                 tratamentos Residuals
## Sum of Squares    0.9697375 0.0270500
## Deg. of Freedom           3         4
##
## Residual standard error: 0.08223442
## Estimated effects may be unbalanced
```

```
resultado <- TukeyHSD(resultadoAov, "tratamentos")
resultado
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = dados ~ tratamentos, data = tabela)
##
## $tratamentos
##                              diff        lwr        upr     p adj
## tratamento 2-tratamento 1 -0.155 -0.4897644  0.1797644 0.3594796
## tratamento 3-tratamento 1 -0.755 -1.0897644 -0.4202356 0.0027235
## tratamento 4-tratamento 1 -0.775 -1.1097644 -0.4402356 0.0024642
## tratamento 3-tratamento 2 -0.600 -0.9347644 -0.2652356 0.0064810
## tratamento 4-tratamento 2 -0.620 -0.9547644 -0.2852356 0.0057354
## tratamento 4-tratamento 3 -0.020 -0.3547644  0.3147644 0.9941576
```

```
plot(resultado, cex = 0.6)
```



```r
# Resposta: Na comparação entre os tratamentos de lavagem, há diferenças
# significativas entre todos os tratamentos exceto os:
# tratamento 2-tratamento 1
# tratamento 4-tratamento 3
# No gráfico é possível observar os tratamentos que contém o 0 (e não possuem
# diferenças significativas) e os tratamentos que não contém o 0 (logo possuem
# diferenças significativas)


# 2) Uma companhia deseja testar 4 diferentes tipos de pneus, A, B, C e D. As vidas médias dos pneus
# (em milhares de milhas) constam na tabela abaixo, onde cada tipo foi testado aleatoriamente em 6
# automóveis idênticos. Determine:

# a) Qual a conclusão do teste Tukey, ao nível de significância de 5%? Explique.
# A  B  C  D
# 23 32 31 28
# 18 40 37 34
# 16 42 35 32
# 10 38 33 30
# 11 30 34 33
# 15 34 30 31

dados <- c(
  23, 32, 31, 28,
  18, 40, 37, 34,
  16, 42, 35, 32,
  10, 38, 33, 30,
  11, 30, 34, 33,
  15, 34, 30, 31
)

pneus <- rep(paste("pneu", LETTERS[1:4]))

tabela <- data.frame(
  pneus = pneus,
  dados = dados
)
tabela
```

```
##     pneus dados
## 1  pneu A    23
## 2  pneu B    32
## 3  pneu C    31
## 4  pneu D    28
## 5  pneu A    18
## 6  pneu B    40
## 7  pneu C    37
## 8  pneu D    34
## 9  pneu A    16
## 10 pneu B    42
## 11 pneu C    35
## 12 pneu D    32
## 13 pneu A    10
## 14 pneu B    38
```

```
## 15 pneu C    33
## 16 pneu D    30
## 17 pneu A    11
## 18 pneu B    30
## 19 pneu C    34
## 20 pneu D    33
## 21 pneu A    15
## 22 pneu B    34
## 23 pneu C    30
## 24 pneu D    31
```

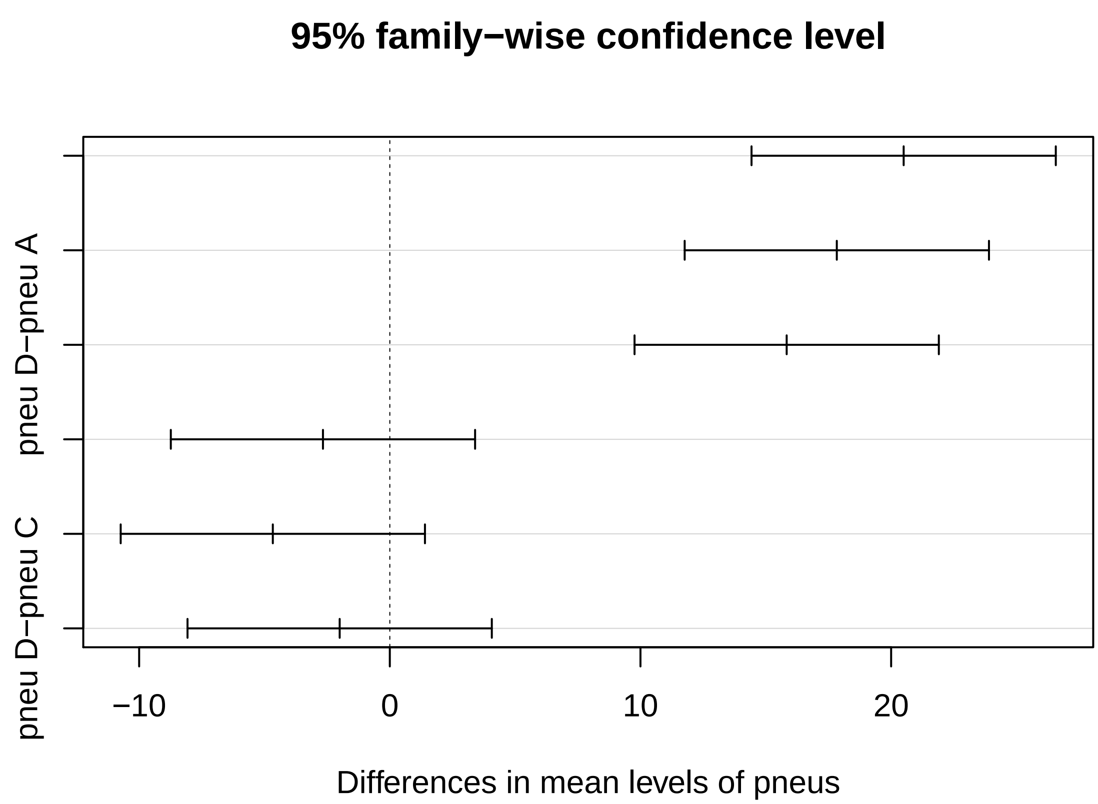
```r
resultadoAov <- aov(
  dados~pneus,
  tabela
)
resultadoAov
```

```
## Call:
##    aov(formula = dados ~ pneus, data = tabela)
##
## Terms:
##                     pneus Residuals
## Sum of Squares  1532.7917  282.1667
## Deg. of Freedom         3        20
##
## Residual standard error: 3.756106
## Estimated effects may be unbalanced
```

```r
resultado <- TukeyHSD(resultadoAov, "pneus")
resultado
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = dados ~ pneus, data = tabela)
##
## $pneus
##                      diff        lwr       upr     p adj
## pneu B-pneu A 20.500000  14.430258 26.569742 0.0000000
## pneu C-pneu A 17.833333  11.763591 23.903075 0.0000004
## pneu D-pneu A 15.833333   9.763591 21.903075 0.0000026
## pneu C-pneu B -2.666667  -8.736409  3.403075 0.6159777
## pneu D-pneu B -4.666667 -10.736409  1.403075 0.1711926
## pneu D-pneu C -2.000000  -8.069742  4.069742 0.7933521
```

```r
# Resposta: Há diferenças significativas entre os:
# B-pneu A, C-pneu A e D-pneu A, nas quais o valor de p ficou abaixo de 0.05
# (não contém 0), já os demais não possuem diferenças significativas.

# b) Faça o gráfico do teste.

plot(resultado, cex = 0.6)
```




```r
# Regressão

# 1) Os dados apresentados na tabela abaixo relacionam o teor de vitamina C (mg de ácido
# ascórbico/100ml de suco de maçã) em função do período de armazenamento em dias.

# Período de armazenamento (dias)
# Teor de vitamina C
# 1   4,09
# 45  3,27
# 90  2,45
# 135 3,27
# 180 1,64

# a) Crie um data.frame e plote o gráfico.
periodo <- c(1, 45, 90, 135, 180)
teor <- c(4.09, 3.27, 2.45, 3.27, 1.64)
dados <- data.frame(periodo, teor)
plot(dados)

# b) Ache a equação da reta que relaciona dos dados.
reglin <- lm(teor~periodo, dados)
reglin
```

```
##
## Call:
## lm(formula = teor ~ periodo, data = dados)
##
## Coefficients:
## (Intercept)      periodo
```

```
##     3.92980     -0.01093
```

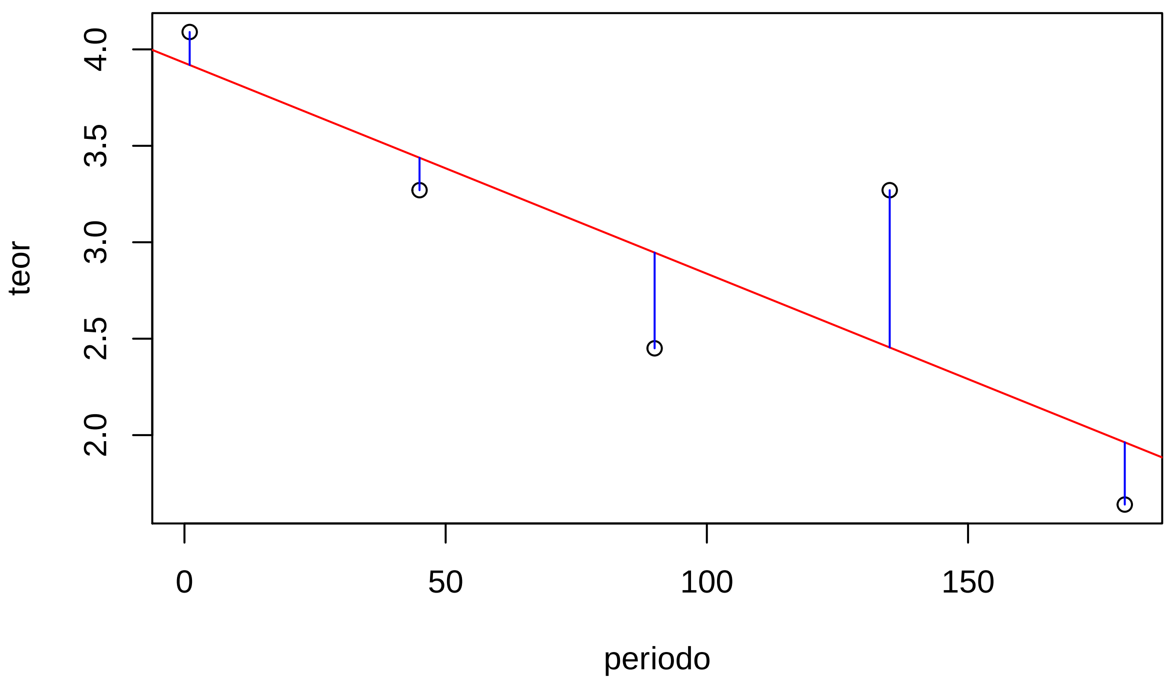
```r
# c) Qual seria o teor de vitamina C se o suco ficar armazenado durante 20 dias?

# Utilizando a função predict
predict(reglin, newdata = data.frame(teor = c(NA), periodo = c(20)))
```

```
##        1
## 3.711217
```

```r
# Calculando manualmente
teorPeriodo20 = 3.92980 + -0.01093 * 20
teorPeriodo20
```

```
## [1] 3.7112
```

```r
# d) Agora plote novamente os dados e acrescente ao gráfico, além da reta de regressão
# ajustada, segmentos de reta representando os resíduos, ou seja, segmentos de reta que
# vão dos valores observados (pontos) aos calculados (reta).
plot(dados)
abline(reglin, col = 2)
segments(
  dados$periodo,
  dados$teor,
  dados$periodo,
  predict(reglin),
  col = 4
)
```




```r
# e) Qual a conclusão dessa regressão?
# Resposta: podemos concluir que o período armazenamento influencia no teor de vitamina C, quanto
# maior o período, menor o teor de vitamina.


# 2) Uma cadeia de padarias queria saber se a quantidade de dinheiro gasto em propaganda faz
```

```r
# aumentar as vendas. Durante seis semanas, fez, em ordem aleatória, gastos com propaganda
# de valores variados conforme a tabela abaixo e anotou os valores recebidos nas vendas.
# Encontre a equação da reta e faça o gráfico de dispersão com a reta ajustada e resíduos.
# Ache o valor do coeficiente de determinação e interprete o valor. O que pode ser concluído
# sobre o comportamento dos valores?

# Gastos Valores recebidos nas vendas
# 100,00 1020,00
# 150,00 1610,00
# 200,00 2030,00
# 250,00 2560,00
# 300,00 2800,00

gastos <- c(100, 150, 200, 250, 300)
vendas <- c(1020, 1610, 2030, 2560, 2800)
dados <- data.frame(gastos, vendas)
dados
```

```
##   gastos vendas
## 1    100   1020
## 2    150   1610
## 3    200   2030
## 4    250   2560
## 5    300   2800
```

```r
reglin <- lm(vendas~gastos, dados)
reglin
```

```
##
## Call:
## lm(formula = vendas ~ gastos, data = dados)
##
## Coefficients:
## (Intercept)       gastos
##      200.00         9.02
```

```r
predict(reglin)
```

```
##    1    2    3    4    5
## 1102 1553 2004 2455 2906
```

```r
resid(reglin)
```

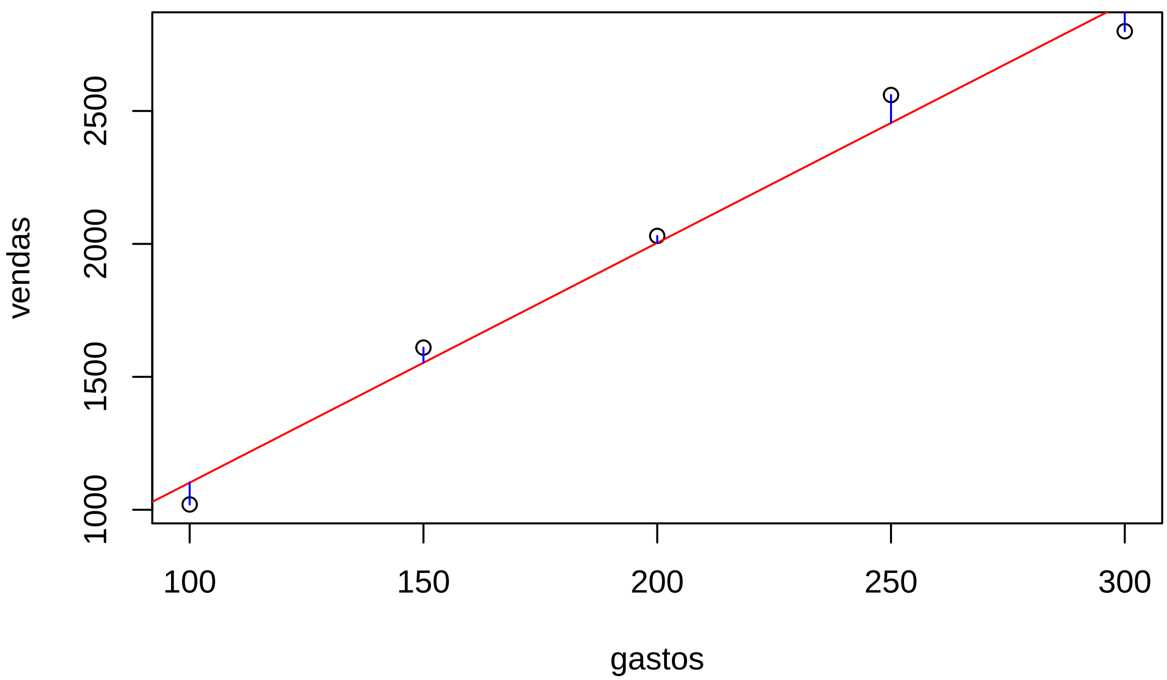
```
##    1    2    3    4    5
##  -82   57   26  105 -106
```

```r
resultado <- data.frame(
  gastos,
  vendas,
  calculado = predict(reglin),
  residuos= resid(reglin)
)
resultado
```

```
##   gastos vendas calculado residuos
## 1    100   1020      1102      -82
## 2    150   1610      1553       57
```

```
## 3    200   2030       2004       26
## 4    250   2560       2455      105
## 5    300   2800       2906     -106
```

```
plot(gastos, vendas)
abline(reglin, col = 2)
segments(
  resultado$gastos,
  resultado$vendas,
  resultado$gastos,
  resultado$calculado,
  col = 4
)
```




```
summary(reglin)
```

```
##
## Call:
## lm(formula = vendas ~ gastos, data = dados)
##
## Residuals:
##    1    2    3    4    5
##  -82   57   26  105 -106
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 200.0000   140.5205   1.423 0.249828
## gastos        9.0200     0.6624  13.617 0.000857 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 104.7 on 3 degrees of freedom
## Multiple R-squared:  0.9841, Adjusted R-squared:  0.9788
```

```
## F-statistic: 185.4 on 1 and 3 DF,  p-value: 0.0008568
```

```r
# Resposta: podemos concluir que quanto maior é o gasto com propagandas, maior é o valor de vendas.
# Através do coeficiente de determinação, podemos entender que o modelo linear se ajusta bem
# aos dados, com 0.9841 (98% indica o quanto pode ser explicado o aumento do valor de vendas em
# relação ao valor gasto com propagandas)


# 3) Para uma amostra de 8 operadores de máquina, foram coletados o número de horas de
# treinamento (x) e o tempo necessário para completar o trabalho (y). Os dados encontram-se na
# tabela abaixo:

# X   Y
# 5,2 13
# 5,1 15
# 4,9 18
# 4,6 20
# 4,7 19
# 4,8 17
# 4,6 21
# 4,9 16

horasTreinamento <- c(5.2, 5.1, 4.9, 4.6, 4.7, 4.8, 4.6, 4.9)
tempoCompletarTrabalho <- c(13, 15, 18, 20, 19, 17, 21, 16)
dados <- data.frame(horasTreinamento, tempoCompletarTrabalho)
dados
```

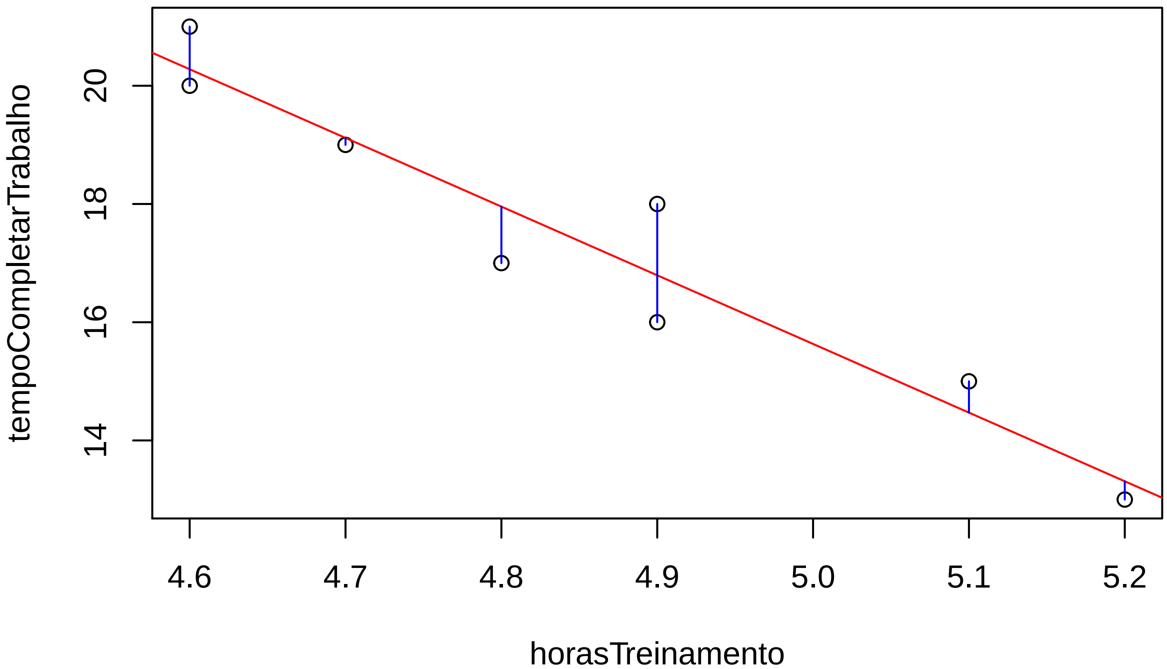
```
##   horasTreinamento tempoCompletarTrabalho
## 1              5.2                     13
## 2              5.1                     15
## 3              4.9                     18
## 4              4.6                     20
## 5              4.7                     19
## 6              4.8                     17
## 7              4.6                     21
## 8              4.9                     16
```

```r
# a) Faça o gráfico de dispersão para esses dados.
plot(dados)

# b) Determine a equação da reta.
reglin <- lm(tempoCompletarTrabalho~horasTreinamento, dados)
reglin
```

```
##
## Call:
## lm(formula = tempoCompletarTrabalho ~ horasTreinamento, data = dados)
##
## Coefficients:
##      (Intercept)  horasTreinamento
##            73.72            -11.62
```

```r
# c) Trace no gráfico anterior, a reta de regressão.
plot(dados)
abline(reglin, col = 2)
segments(
```

```r
  dados$horasTreinamento,
  dados$tempoCompletarTrabalho,
  dados$horasTreinamento,
  predict(reglin),
  col = 4
)
```




```r
# d) Calcule e interprete o coeficiente de determinação.
summary(reglin)
```

```
##
## Call:
## lm(formula = tempoCompletarTrabalho ~ horasTreinamento, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9559 -0.4301 -0.1985  0.5772  1.2059
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        73.721      6.785  10.865  3.6e-05 ***
## horasTreinamento  -11.618      1.398  -8.312 0.000164 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.815 on 6 degrees of freedom
## Multiple R-squared:  0.9201, Adjusted R-squared:  0.9068
## F-statistic: 69.09 on 1 and 6 DF,  p-value: 0.0001645
```

```r
# Resposta: quanto maior a quantidade de horas de treinamento, menor o tempo para completar
# o trabalho. De acordo com o coeficiente de determinação, 0.9201 (92% indica o quanto pode ser
# explicado a diminuição do tempo de trabalho em relação ao maior tempo de treinamento).
# O modelo linear se ajusta bem aos dados.
```