Lista 1C

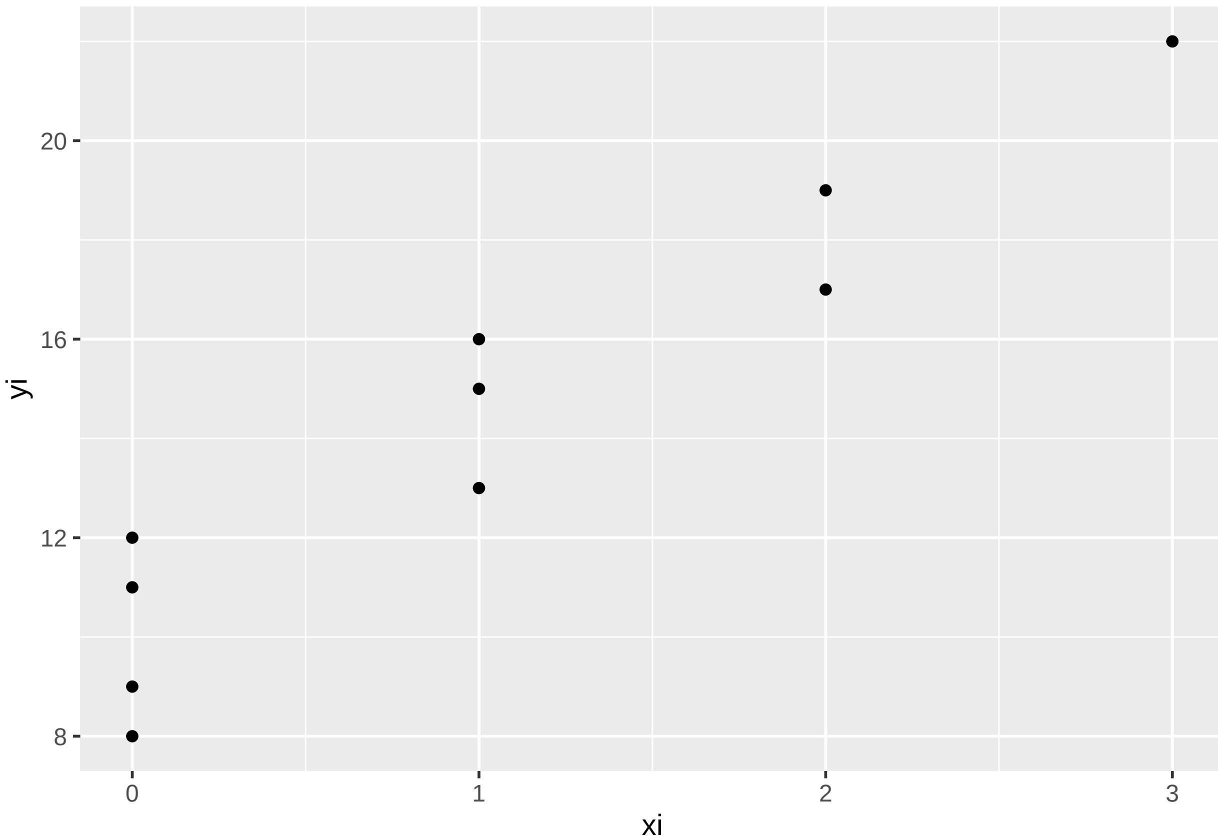
Davi Wentrick Feijó

2023-04-22

Uma substância usada em pesquisas biológicas e médicas é enviada por avião aos usuários em caixas de 1000 ampolas. Os dados abaixo, envolvendo 10 remessas, contém as informações coletadas sobre o número de vezes que a caixa foi transferida de uma aeronave durante o percurso (X) e o número de ampolas quebradas até a chegada (Y). Assumir que um modelo de regressão de primeira ordem é apropriado.

xi	yi
1	16
0	9
2	17
0	12
3	22
1	13
0	8
1	15
2	19
0	11

a) Represente graficamente os dados e analise-o.



b) Calcule o coeficiente de correlação linear simples entre as variáveis e analise-o. Os dados confirmam, a um nível de significância de 0,05, que existe associação linear entre o número de ampolas quebradas até a chegada da remessa e o número de vezes que a caixa foi transferida de uma aeronave durante o percurso?

```
## O coeficiente de correlacao é: 0.949158
```

```
##
##  Pearson's product-moment correlation
##
## data:  xi and yi
## t = 8.528, df = 8, p-value = 2.749e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7940941 0.9882133
## sample estimates:
##      cor
## 0.949158
```

c) Modele o estudo acima, especificando a unidade de informação, o tipo de estudo, a variável resposta e a variável explicativa e a hipótese de trabalho. A unidade de informacao sao as caixas com ampolas, o tipo de estudo é observacional, a variavel resposta é a quantidade de ampolas quebradas e a variavel explicativa é o número de transferencias.

d) Obter as estimativas dos parâmetros do MRLS. Faça um gráfico com a função de regressão estimada e os dados observados. A função de regressão estimada apresenta um bom ajuste neste caso?

$$\bar{Y} = 14.2 \quad \bar{X} = 1$$

$$\sum_{n=1}^{n} X_i = 10 \quad \sum_{n=1}^{n} Y_i = 142$$

$$\sum_{n=1}^{n} X_i^2 = 20 \quad \sum_{n=1}^{n} Y_i^2 = 2194 \quad \sum_{n=1}^{n} X_i Y_i = 182$$

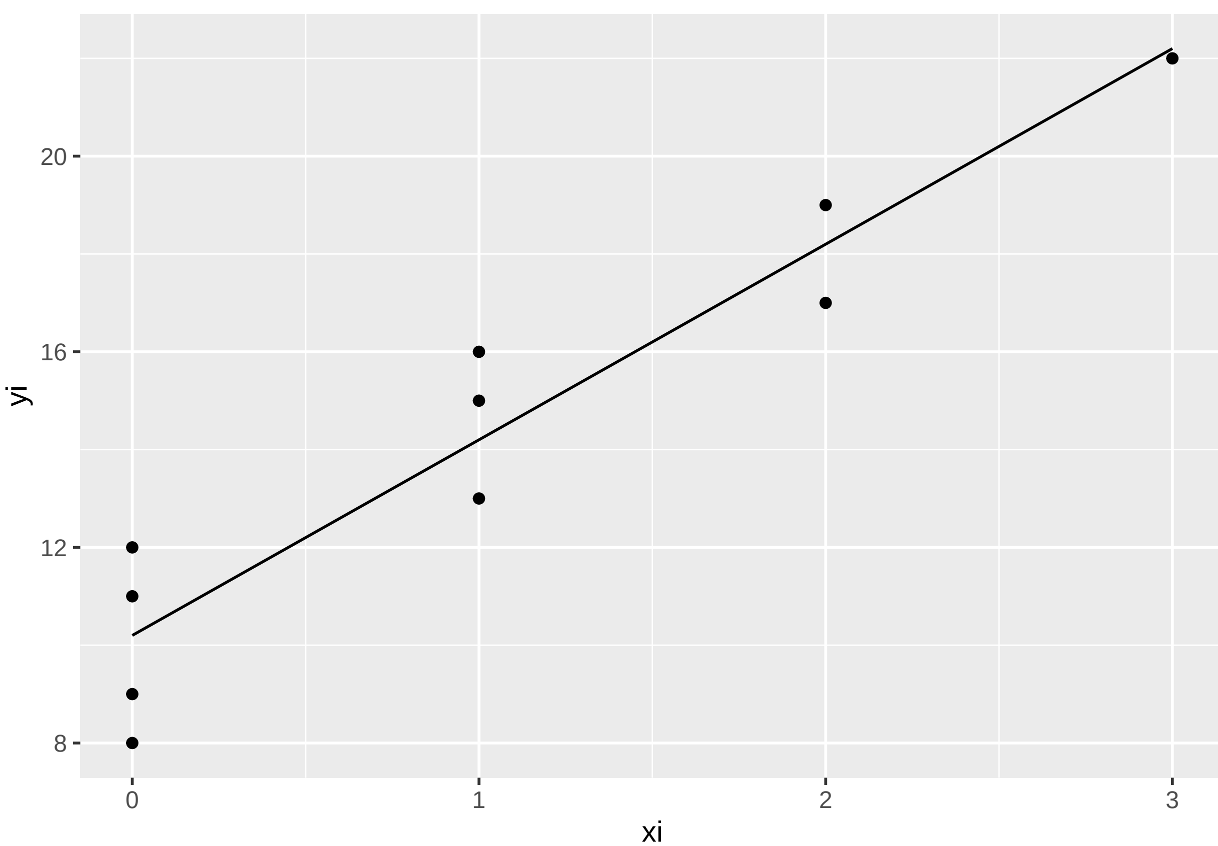
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

```
## O valor de Beta 0 é: 10.2
```

```
## O valor de Beta 0 é: 4
```

Com os novos coeficientes obtidos de β_1 e β_0 a nossa equacao ficou assim:

$$\hat{Y} = 10,2 + 4x$$



e) Obtenha a estimativa pontual do número esperado de ampolas quebradas quando 1 transferência for feita.

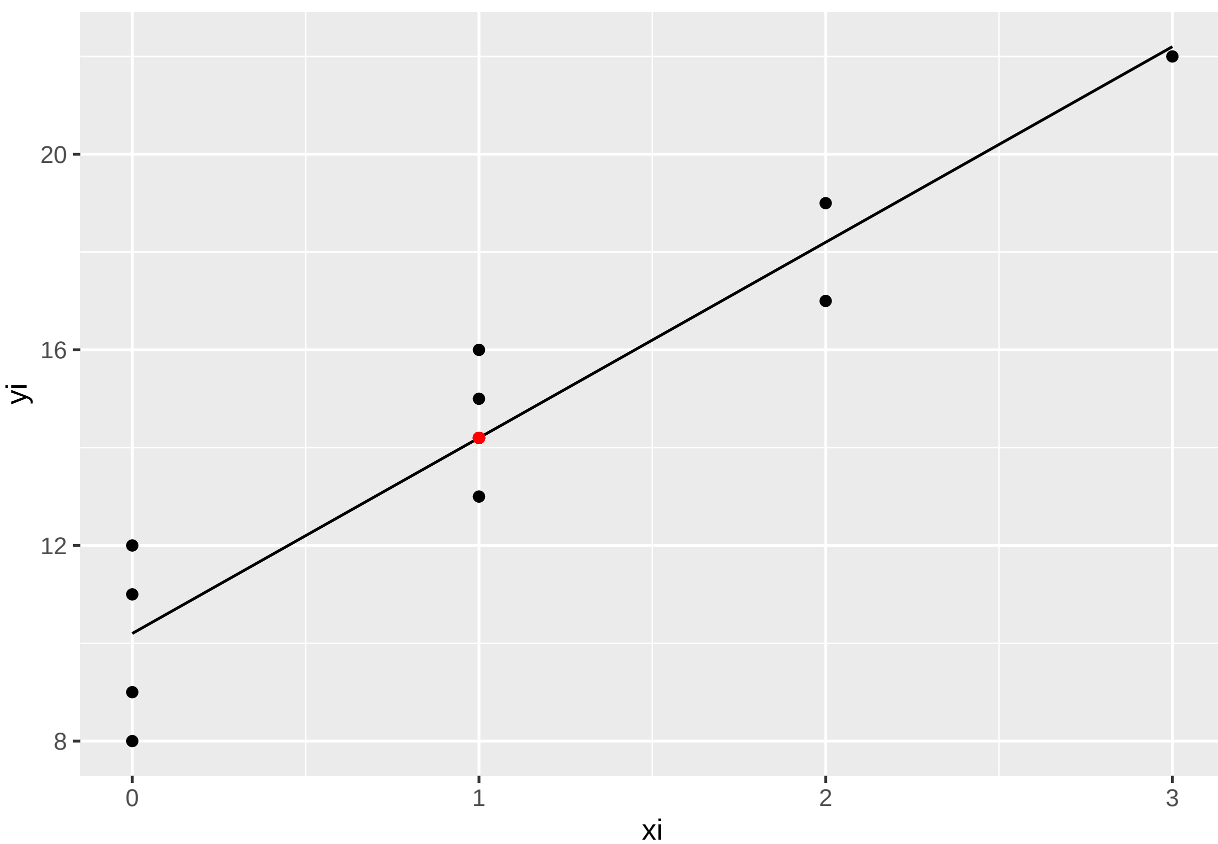
```
## O valor estimado pontualmente de Y quando X = 1 é: 14.2
```

f) Estime o aumento no número esperado de ampolas quebradas quando forem feitas 2 transferências comparada com 1 transferência.

```
## O valor estimado pontualmente de Y quando X = 1 é: 18.2
```

```
## O aumento estimado entre X = 1 e X = 2 é: 4
```

g) Verifique se a sua função de regressão ajustada passa pelo ponto $(\bar{X}, \bar{Y})$.



h) Obtenha o resíduo para o primeiro caso. Qual é a sua relação com ϵ_1?

```
## Os residuos quando X = 1 sao: 1.8 -1.2 0.8
```

i) Calcule os resíduos estimados e a soma deles

```
## A soma dos residuos é: 7.105427e-15
```

Calcule MSE. O que é estimado pelo MSE?

$$MSE = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

k) Estimar $V(\beta_0)$ e $V(\beta_1)$.

```
## A variancia de Beta 0 é: 0.44
##  O desvio padrao é: 0.663325
##  e o coeficiente de variacao é: 0.06503186
```

```
## A variancia de Beta 0 é: 0.22
##  O desvio padrao é: 0.4690416
##  e o coeficiente de variacao é: 0.1172604
```

l) Calcular o coeficiente de correlação linear entre $\hat{\beta}_1$ e $\hat{\beta}_0$.

m) Estime b1 com 95% de confiança. Faça a interpretação do intervalo encontrado.

```
## Intervalo de confiança para Beta 1 com Alfa de 0.05 é: [ 2.918388 ,  5.081612 ]
```

n) Faça o teste t para decidir se existe ou não regressão linear. Use um nível de significância de 5%. Formule as hipóteses, regra de decisão e a conclusão. Qual é o p-valor do teste? Interprete.

```
## O P-Valor observado foi: 1.374335e-05 O valor observado pertence a regiao crítica de 0.05 ?: TRUE
```

```
#estimacao pontual quando xi=0

x_escolhido = 0

#intervalo de confianca
sigma_erro = sqrt(sigma_quadrado_erro)
t_value <- qt(0.975, n - 2,lower.tail = TRUE) # Valor crítico da distribuição t-Student para o nível de
y_new_hat <- beta_0 + beta_1*x_escolhido # Estimativa pontual da média de Y para o valor de X escolhido
ic_lower <- y_new_hat - t_value*sigma_erro*sqrt((1/n)+((x_escolhido- x_barra)^2)/(x_quadrado - n*x_barra
ic_upper <- y_new_hat + t_value*sigma_erro*sqrt((1/n)+((x_escolhido- x_barra)^2)/(x_quadrado - n*x_barra
#arrumar os parenteses da equacao (acho que ja arrumei)
cat("O valor estimado Y quando X=", x_escolhido, "é:",y_new_hat)
```

p) Um especialista na área de medicina sugeriu, baseado em experiências, que o número médio de ampolas quebradas não deveria exceder 9 quando não há transferência. Faça um teste, usando a=0,025%. Formule as hipóteses, regra de decisão e conclusão. Qual é o p-valor do teste?

```
## O valor estimado Y quando X= 0 é: 10.2
```

```r
cat("Intervalo de confiança para a média de Y quando X =", x_escolhido, ": [", ic_lower, ",", ic_upper,
```

```
## Intervalo de confiança para a média de Y quando X = 0 : [ 8.67037 , 11.72963 ]
```

```r
#teste de hipoteses
#Ho: ampolas qubradas quando (x=0) > 9
#H0: ampolas qubradas quando (x=0) <= 9
t_obs = (9 - y_new_hat)/(n-2)

pt(t_obs,n-2,lower.tail = FALSE)
```

```
## [1] 0.5577613
```

q) Devido a mudanças na rota do avião, as remessas podem ter que ser transferidas mais frequentemente do que atualmente. Estime a média de ampolas quebradas para os seguintes números de transferências: X=2 e 4. Faça os intervalos com 99% de confiança. Interprete os resultados. X = 2

```r
#estimacao pontual quando xi=0

x_escolhido = 2

#intervalo de confianca
sigma_erro = sqrt(sigma_quadrado_erro)
t_value <- qt(0.99, n - 2,lower.tail = TRUE) # Valor crítico da distribuição t-Student para o nível de
y_new_hat <- beta_0 + beta_1*x_escolhido # Estimativa pontual da média de Y para o valor de X escolhido
ic_lower <- y_new_hat - t_value*sigma_erro*sqrt((1/n)+((x_escolhido- x_barra)^2)/(x_quadrado - n*x_barra
ic_upper <- y_new_hat + t_value*sigma_erro*sqrt((1/n)+((x_escolhido- x_barra)^2)/(x_quadrado - n*x_barra
#arrumar os parenteses da equacao (acho que ja arrumei)
cat("O valor estimado Y quando X=", x_escolhido, "é:",y_new_hat)
```

```
## O valor estimado Y quando X= 2 é: 18.2
```

```r
cat("Intervalo de confiança para a média de Y quando X =", x_escolhido, ": [", ic_lower, ",", ic_upper,
```

```
## Intervalo de confiança para a média de Y quando X = 2 : [ 16.27871 , 20.12129 ]
```

X = 4

```r
#estimacao pontual quando xi=0

x_escolhido = 4

#intervalo de confianca
sigma_erro = sqrt(sigma_quadrado_erro)
t_value <- qt(0.99, n - 2,lower.tail = TRUE) # Valor crítico da distribuição t-Student para o nível de
y_new_hat <- beta_0 + beta_1*x_escolhido # Estimativa pontual da média de Y para o valor de X escolhido
ic_lower <- y_new_hat - t_value*sigma_erro*sqrt((1/n)+((x_escolhido- x_barra)^2)/(x_quadrado - n*x_barra
ic_upper <- y_new_hat + t_value*sigma_erro*sqrt((1/n)+((x_escolhido- x_barra)^2)/(x_quadrado - n*x_barra
#arrumar os parenteses da equacao (acho que ja arrumei)
cat("O valor estimado Y quando X=", x_escolhido, "é:",y_new_hat)
```

```
## O valor estimado Y quando X= 4 é: 26.2
```

```
cat("Intervalo de confiança para a média de Y quando X =", x_escolhido, ": [", ic_lower, ",", ic_upper,
```

```
## Intervalo de confiança para a média de Y quando X = 4 : [ 21.90386 , 30.49614 ]
```

2) As estimativas obtidas para os parâmetros do MRLS ajustado aos dados do exercício 1 são, respectivamente, 10.20 e 4.00 e $\sum_i e_i = 17.60$ consequentemente. Determine o valor do criterio de mínimo quadrado $Q = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$ para as seguintes estimativas:

a) $\hat{\beta}_0 = 9$ e $\hat{\beta}_1 = 3$
b) $\hat{\beta}_0 = 11$ e $\hat{\beta}_1 = 5$

O valor do critério Q é maior do que para as estimativas de Mínimos Quadrados? Justifique sua resposta.

```
## O valor Q com Beta 0 = 9 e Beta 1 = 3 é: 22
```

```
## O valor Q com Beta 0 = 11 e Beta 1 = 5 é: -18
```

O valor Q é maior que a estimativa de minimos quadrados pois a estimativa é um minimo!

Dando continuidade ao estudo realizado no Exercício 1:

```
#regiao de confiança para superficie de regressao
sigma = sqrt(sigma_quadrado_erro)

x_h = 4

w = 2*(qf(1-alfa,2,n-2))

y_h = beta_0 + beta_1*x_h

s_pred = sigma*sqrt(1 + (1/n) + ((x_h - x_barra)^2)/(x_quadrado - n*x_barra^2))

ic_superficie_inf = y_h - w * s_pred

ic_superficie_sup = y_h + w * s_pred


cat("O valor estimado Yh quando Xh=", x_h, "e:",y_h)
```

a) Determine os valores limites da faixa de confiança de 99% para a linha de regressão quando Xh=2 e quando Xh=4. A sua faixa de confiança é mais ampla nestes dois pontos do que os correspondentes intervalos de confiança calculados no item e.? Ela deveria ser?

```
## O valor estimado Yh quando Xh= 4 e: 26.2
```

```r
cat("Intervalo de confiança de Bonferroni para a média de Y quando X =", x_h, ": [", ic_superficie_inf,
```

```
## Intervalo de confiança de Bonferroni para a média de Y quando X = 4 : [ 7.493571 , 44.90643 ]
```

```r
reg = lm(yi ~ xi, data = dados)
```

b) Faça a ANOVA.

c) Teste ausência de regressão usando o teste F a um nível se significância de 0,05.

i) Construa o quadro de ANOVA. Quais elementos são aditivos?

ii) Formule as hipóteses, regra de decisão e conclusão.

d) Verifique o valor da estatística t obtida para o teste realizado no item n do exercício 1 e mostre numericamente sua equivalência com a estatística F* obtida no item c deste exercício.

e) Calcule r2 e r. Qual a proporção da variação em Y é atribuída pela introdução da variável X no modelo de regressão?