

Modelo linear

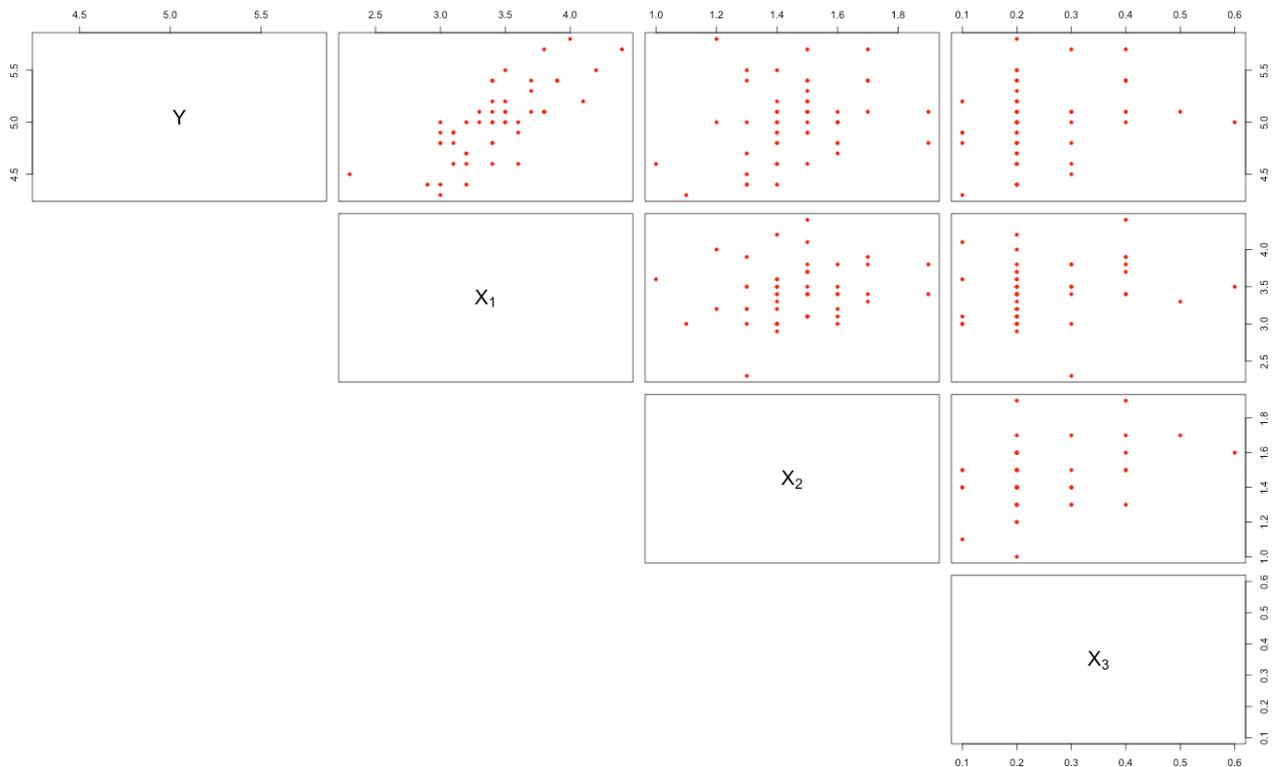
É apresentado um exemplo de ajuste em R com algumas ferramentas de diagnóstico aplicadas a um modelo linear normal. Os dados de $n = 50$ observações estão apresentados na p. 8, sendo que Y representa a variável resposta e X_1 , X_2 , e X_3 são as variáveis explicativas.

```
dados <- . . . # completar  
n <- nrow(dados)
```

```
summary(dados)
```

	Y	X1	X2	X3
Min.	:4.300	Min. :2.300	Min. :1.000	Min. :0.100
1st Qu.:	4.800	1st Qu.:3.200	1st Qu.:1.400	1st Qu.:0.200
Median	:5.000	Median :3.400	Median :1.500	Median :0.200
Mean	:5.006	Mean :3.428	Mean :1.462	Mean :0.246
3rd Qu.:	5.200	3rd Qu.:3.675	3rd Qu.:1.575	3rd Qu.:0.300
Max.	:5.800	Max. :4.400	Max. :1.900	Max. :0.600

```
pairs(dados, labels = c("Y", expression(X[1]), expression(X[2]),  
expression(X[3])), pch = 20, lower.panel = NULL, col = "red")
```



Nota 1. Comente os gráficos acima.

Um modelo linear é ajustado com a função `lm`.

```
m1 <- lm(Y ~ X1 + X2 + X3, data = dados)
names(m1)

[1] "coefficients" "residuals"      "effects"      "rank"
[5] "fitted.values" "assign"          "qr"           "df.residual"
[9] "xlevels"      "call"           "terms"        "model"
```

Nota 2. Explique cada um dos componentes do objeto `m1` acima.

Nota 3. Refaça o ajuste utilizando a função `glm`.

```
summary(m1)

Call:
lm(formula = Y ~ X1 + X2 + X3, data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-0.40662 -0.17721  0.01222  0.13388  0.49693

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.35189    0.39287   5.986 3.03e-07 ***
X1           0.65483    0.09245   7.083 6.83e-09 ***
X2           0.23756    0.20802   1.142  0.259
X3           0.25213    0.34686   0.727  0.471
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2371 on 46 degrees of freedom
Multiple R-squared: 0.5751, Adjusted R-squared: 0.5474
F-statistic: 20.76 on 3 and 46 DF, p-value: 1.192e-08
```

Nota 4. Comente os resultados gerados pela função `summary`.

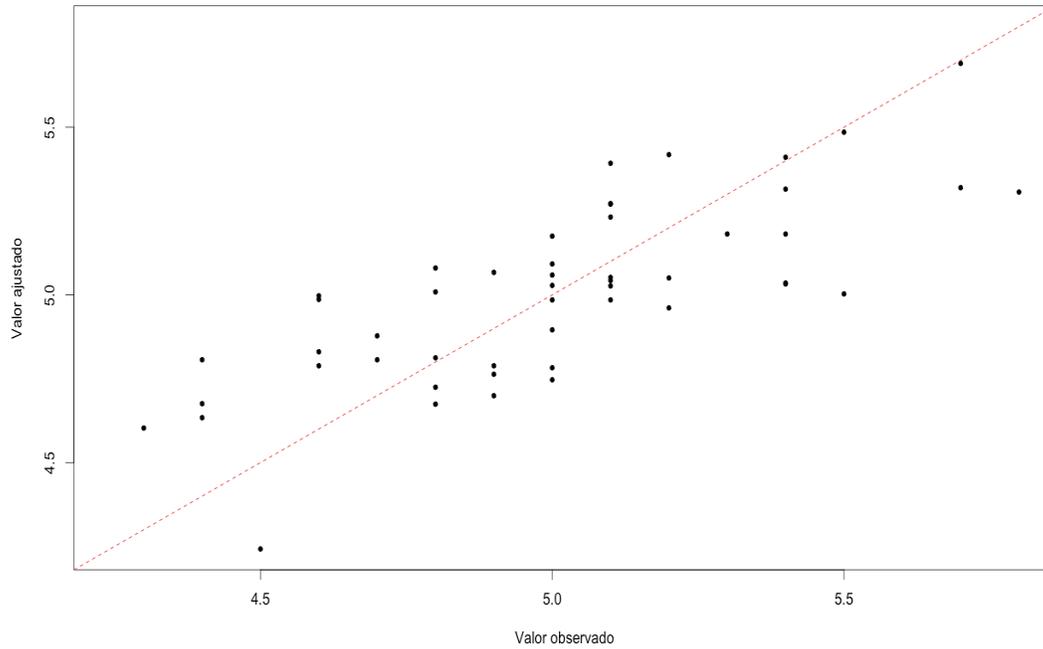
Nota 5. No modelo nulo a média da variável resposta é constante. Pode ser ajustado com o comando `m0 <- lm(Y ~ 1)`. Com os resultados do comando `anova(m0, m1)`, construa a tabela ANOVA do ajuste do modelo `m1`.

Nota 6. Apresente intervalos de confiança de 95% para os coeficientes de `X1`, `X2`, e `X3`.

Os valores observados e ajustados pelo modelo são apresentados abaixo.

```
rm1 <- range(dados$Y, m1$fitted.values)
plot(dados$Y, m1$fitted.values, pch = 20, xlab = "Valor observado",
      ylab = "Valor ajustado", xlim = rm1, ylim = rm1)
abline(0, 1, lty = 2, col = "red")
```

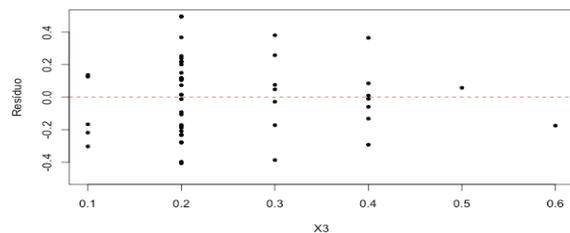
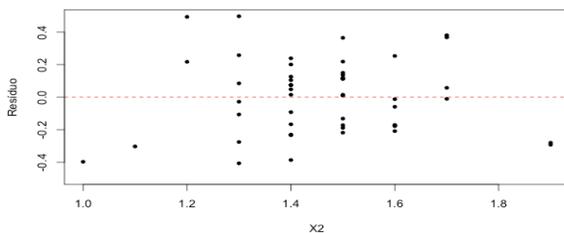
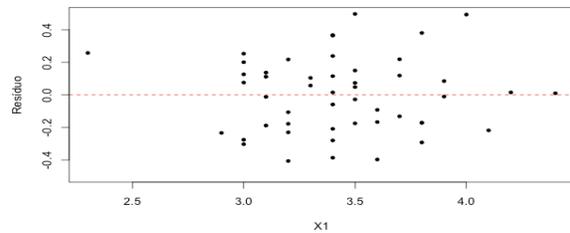
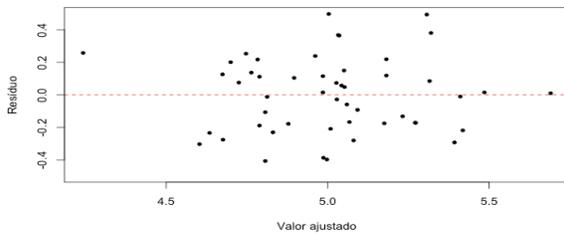
Nota 7. Justifique o uso da função `range` no trecho acima.



Nota 8. Comente o resultado do ajuste com base no gráfico acima.

Os gráficos dos resíduos ordinários *versus* valores preditos e também variáveis explicativas são mostrados em seguida.

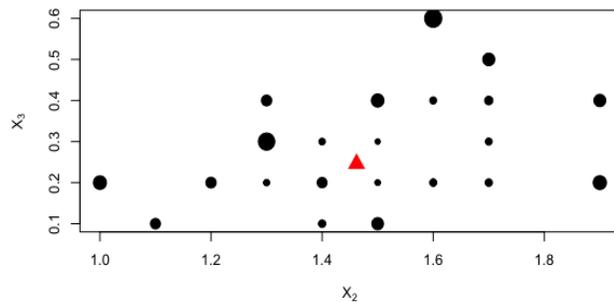
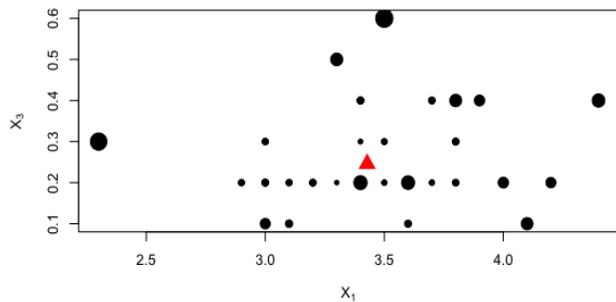
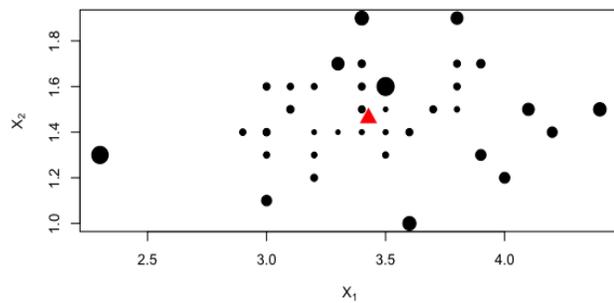
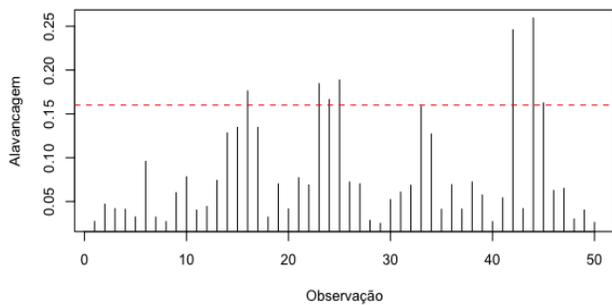
```
par(mfrow = c(2, 2))
maxe <- max(abs(m1$resid))
plot(m1$fitted.values, m1$residuals, pch = 20, ylab = "Resíduo",
     xlab = "Valor ajustado", ylim = c(-maxe, maxe))
abline(h = 0, lty = 2, col = "red")
for (j in 2:4) {
  plot(dados[, j], m1$residuals, pch = 20, ylab = "Resíduo",
       xlab = names(dados)[j], ylim = c(-maxe, maxe))
  abline(h = 0, lty = 2, col = "red")
}
```



Nota 9. Comente os gráficos acima relacionando-os com alguma suposição do modelo.

Alguns gráficos representando a medida de alavancagem são mostrados abaixo.

```
p <- length(m1$coefficients)
par(mfrow = c(2, 2))
minh <- min(lm.influence(m1)$h)
cexh <- 3 * (lm.influence(m1)$h - minh) / (max(lm.influence(m1)$h) - minh)
+ 1
plot(lm.influence(m1)$h, type = "h", xlab = "Observação",
     ylab = "Alavancagem")
abline(h = 2 * p / n, lty = 2, col = "red")
plot(X[, 2], X[, 3], pch = 20, cex = cexh, xlab = expression(X[1]), ylab =
expression(X[2]))
points(mean(X[, 2]), mean(X[, 3]), pch = 17, col = "red", cex = 2)
plot(X[, 2], X[, 4], pch = 20, cex = cexh, xlab = expression(X[1]), ylab =
expression(X[3]))
points(mean(X[, 2]), mean(X[, 4]), pch = 17, col = "red", cex = 2)
plot(X[, 3], X[, 4], pch = 20, cex = cexh, xlab = expression(X[2]), ylab =
expression(X[3]))
points(mean(X[, 3]), mean(X[, 4]), pch = 17, col = "red", cex = 2)
```



Nota 10. Explique os resultados gerados pela função `lm.influence`.

Nota 11. Nos gráficos acima identifique as observações com alavancagem mais alta.

Nota 12. Calcule a matriz chapéu (H). A função `model.matrix` fornece a matriz modelo. Para comparação, utilize as funções `hat` e `hatvalues`.

Dois gráficos com os resíduos studentizados deletados são apresentados a seguir.

```
# Resíduo studentizado deletado
ti <- m1$residuals / (lm.influence(m1)$sigma *
  sqrt(1 - lm.influence(m1)$h))
```

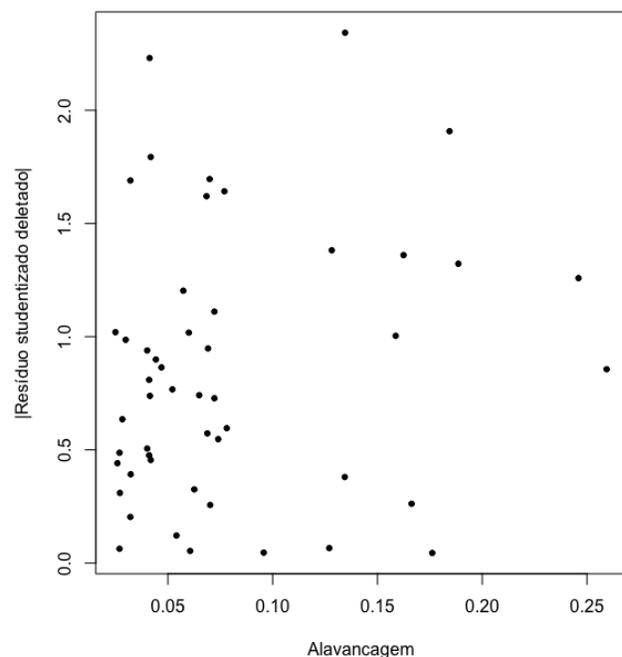
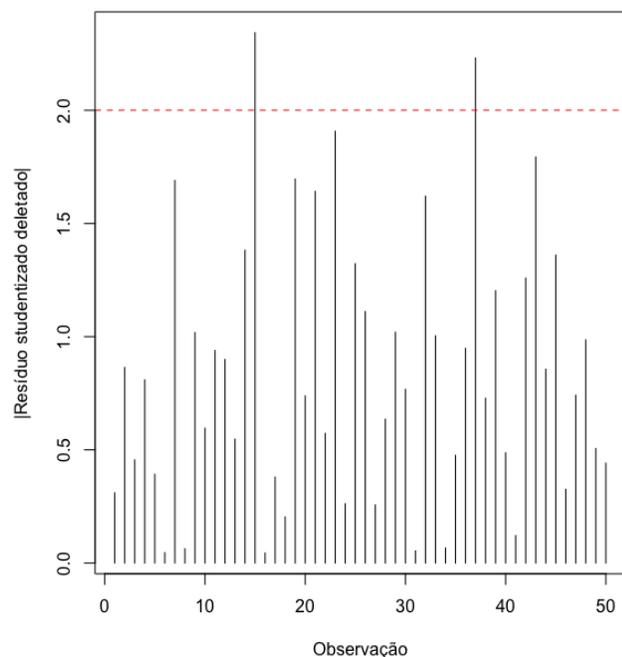
Nota 13. Verifique os resultados das funções `rstandard` e `rstudent`.

```
summary(ti)
```

```
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-1.908000 -0.798800  0.053680 -0.001755  0.583500  2.343000
```

```
par(mfrow = c(1, 2))
plot(abs(ti), type = "h", xlab = "Observação",
      ylab = "|Resíduo studentizado deletado|")
abline(h = qt(0.975, df = n - p - 1), lty = 2, col = "red")
```

```
plot(lm.influence(m1)$h, abs(ti), pch = 20, xlab = "Alavancagem",
      ylab = "|Resíduo studentizado deletado|")
```



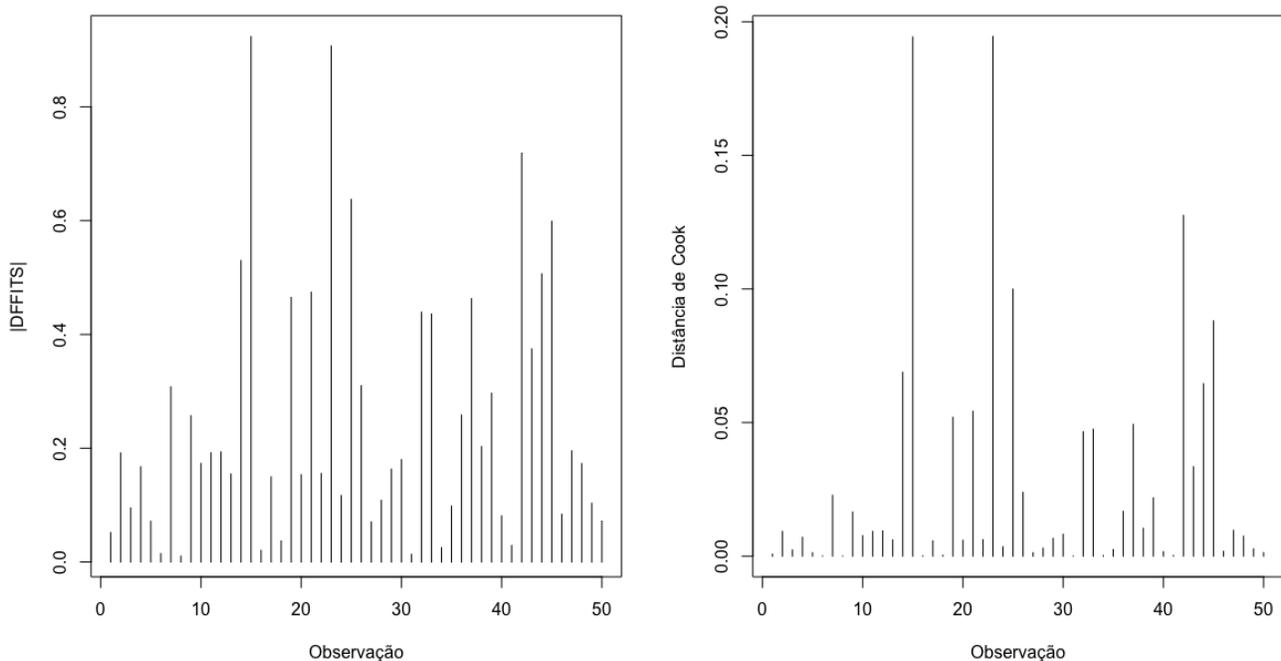
Nota 14. Comente os gráficos acima relacionando-os com suposições do modelo.

Para ajudar a identificar observações influentes, são apresentados gráficos de índices das medidas DFFITS e distância de Cook.

```
# Observações influentes
dffits <- ti * sqrt(lm.influence(m1)$h / (1 - lm.influence(m1)$h))
dcook <- m1$resid^2 * lm.influence(m1)$h /
  (p * summary(m1)$s^2 * (1 - lm.influence(m1)$h)^2)
```

Nota 15. Compare os resultados acima com os obtidos com as funções `dffits` e `cooks.distance`.

```
par(mfrow = c(1, 2))
plot(abs(dffits1), type = "h", xlab = "Observação", ylab = "|DFFITS|")
plot(dcook, type = "h", xlab = "Observação", ylab = "Distância de Cook")
```

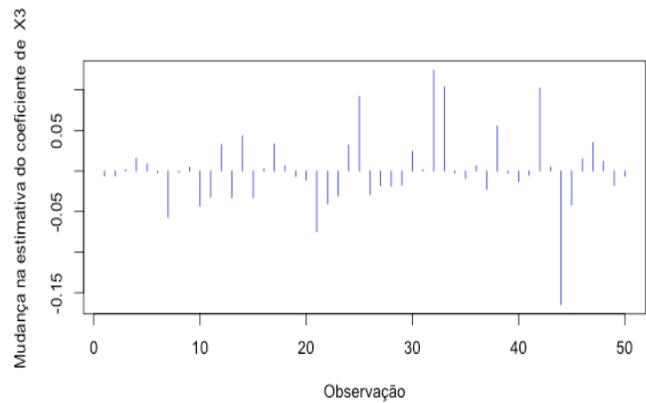
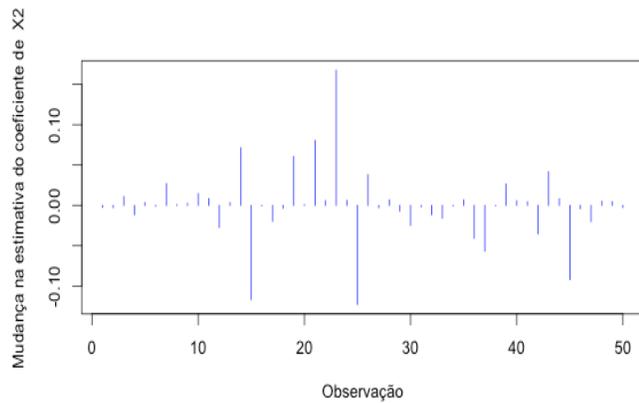
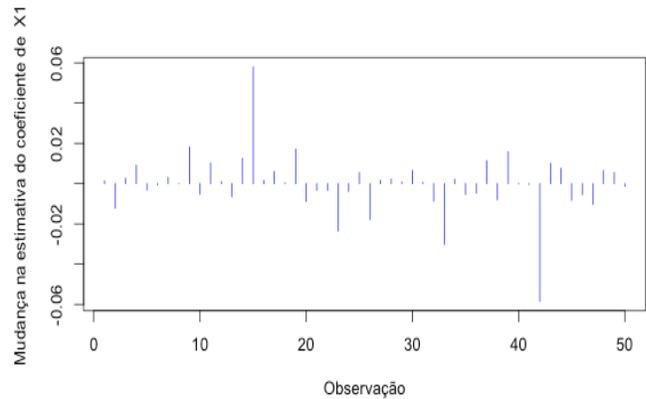
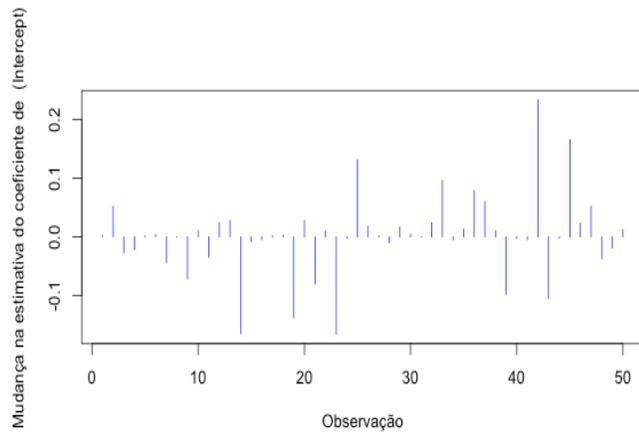


Nota 16. Identifique as observações mais influentes nos gráficos acima. Tente apontar alguma diferença marcante delas em relação às demais.

Em seguida, para cada coeficiente do modelo, é apresentada a mudança na estimativa decorrente da exclusão de uma observação de cada vez.

```
par(mfrow = c(2, 2))
for (j in 1:4) {
  plot(lm.influence(m1)$coef[, j], type = "h", xlab = "Observação",
       ylab = paste("Mudança na estimativa do coeficiente de ",
                    names(coef(m1))[j]), col = "blue")
}
```

Nota 17. Identifique as observações mais influentes nos gráficos abaixo. Tente apontar alguma diferença marcante delas em relação às demais.



Nota 18. Calcule DFBETAS e represente graficamente. Implemente a expressão e compare com os resultados da função `dfbetas`.

Nota 19. Efetue o teste da hipótese $H: \beta_{X2} = \beta_{X3} = 0$. Caso seja possível simplificar o modelo, refaça o exemplo com o modelo mais simples.

Nota 20. Procure reproduzir os resultados utilizando outro pacote estatístico (SAS, SPSS, Minitab, Statistika, ...).

Obs.	Y	X1	X2	X3
1	5,1	3,5	1,4	0,2
2	4,9	3,0	1,4	0,2
3	4,7	3,2	1,3	0,2
4	4,6	3,1	1,5	0,2
5	5,0	3,6	1,4	0,2
6	5,4	3,9	1,7	0,4
7	4,6	3,4	1,4	0,3
8	5,0	3,4	1,5	0,2
9	4,4	2,9	1,4	0,2
10	4,9	3,1	1,5	0,1
11	5,4	3,7	1,5	0,2
12	4,8	3,4	1,6	0,2
13	4,8	3,0	1,4	0,1
14	4,3	3,0	1,1	0,1
15	5,8	4,0	1,2	0,2
16	5,7	4,4	1,5	0,4
17	5,4	3,9	1,3	0,4
18	5,1	3,5	1,4	0,3
19	5,7	3,8	1,7	0,3
20	5,1	3,8	1,5	0,3
21	5,4	3,4	1,7	0,2
22	5,1	3,7	1,5	0,4
23	4,6	3,6	1,0	0,2
24	5,1	3,3	1,7	0,5
25	4,8	3,4	1,9	0,2
26	5,0	3,0	1,6	0,2
27	5,0	3,4	1,6	0,4
28	5,2	3,5	1,5	0,2
29	5,2	3,4	1,4	0,2
30	4,7	3,2	1,6	0,2
31	4,8	3,1	1,6	0,2
32	5,4	3,4	1,5	0,4
33	5,2	4,1	1,5	0,1
34	5,5	4,2	1,4	0,2
35	4,9	3,1	1,5	0,2
36	5,0	3,2	1,2	0,2
37	5,5	3,5	1,3	0,2
38	4,9	3,6	1,4	0,1
39	4,4	3,0	1,3	0,2
40	5,1	3,4	1,5	0,2
41	5,0	3,5	1,3	0,3
42	4,5	2,3	1,3	0,3
43	4,4	3,2	1,3	0,2
44	5,0	3,5	1,6	0,6
45	5,1	3,8	1,9	0,4
46	4,8	3,0	1,4	0,3
47	5,1	3,8	1,6	0,2
48	4,6	3,2	1,4	0,2
49	5,3	3,7	1,5	0,2
50	5,0	3,3	1,4	0,2