Lista 4 - MAE0217

Todos os exercícios são do capítulo 6

Exercício 1

a)

Considerando o modelo dado

$$y_i = \beta x_i + e_i, \qquad i = 1,\ldots, n$$

onde $\mathrm{E(e_i)} = 0$ e $\mathrm{Var(e_i)} = \sigma^2$. Então, assim como apresentado no texto base iremos considerar a soma dos quadrados dos erros e_i,

$$\mathrm{Q}(\beta) = \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \beta x_i)^2$$

Para determinar o estimador de mínimos quadrados de β, faremos o seguinte precedimento:

1. Derivar $\mathrm{Q}(\beta)$ em relação a β
2. Igualar o resultado dessa derivação à zero, obtendo assim a equação de estimação

E a solução dessa equação será $\hat{\beta}$, o estimador desejado.

(1)

$$\frac{\partial Q}{\partial \beta} = \sum_{i=1}^{n} \frac{\partial (y_i - \beta x_i)^2}{\partial \beta} = \sum_{i=1}^{n} -2x_i(y_i - \beta x_i) = -2\sum_{i=1}^{n}(x_i y_i - \beta x_i^2) = -2(\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2)$$

Assim,

$$\frac{\partial Q}{\partial \beta} = -2(\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2)$$

(2) Agora, igualando a derivada à zero obtemos:

$$\frac{\partial Q}{\partial \beta} = -2(\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

Por fim, o estimador para $\hat{\beta}$ é:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Seguindo o mesmo raciocínio adotado no livro para a proposição de um estimador não enviesado para σ^2, podemos afimar que um estimador não viesado é:

$$\mathrm{S}^2 = \frac{1}{n-1} Q(\hat{\beta}) = \frac{1}{n-1}\sum_{i=1}^{n} \hat{e}_i^2 = \frac{1}{n-1}\sum_{i=1} (y_i - \hat{\beta} x_i)^2$$

pois perdemos um grau de liberdade na estimação de um parâmetro ($\hat{\beta}$)

b)

Para determinar a distribuição aproximada de $\hat{\beta}$ devemos considerar alguns pontos, dentre os quais alguns já estão satisfeitos pelas suposições do próprio exercício,

1. $\mathrm{E(e_i)} = 0$
2. $\mathrm{Var(e_i)} = \sigma^2$ constante (Homoscedasticidade)
3. Os erros (e_i) são não correlacionados
4. Suposição de normalidade $(e_i \sim N(0, \sigma^2))$

Dado que a condição 4 é a única não satisfeita de modo direto,existem duas maneiras de obter uma distribuição para $\hat{\beta}$

1. Utilizando uma aproximação assintótica pelo Teorema Central do Limite (TLC). Isto é, para uma grande quantidade de observações, pode-se chegar à uma distribuição normal aproximada.

2. Forçando que 4. seja válida.

Podemos obter a seguinte distribuição

$$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2})$$

sendo esta aproximada pela maneira 1. e exata pela 2. Considere ainda a seguinte expressão

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta} x_i$$

Desta forma, como o estimador de β é uma transformação linear da distribuição dos resíduos, então podemos dizer que a distribuição aproximada, em ambos os casos, para $\hat{\beta}$ é,

$$\hat{\beta} \sim N(\beta, \frac{S^2}{\sum_{i=i}^{n} x_i^2})$$

E por fim, uma outra abordagem para a distribuição de $\hat{\beta}$ seria padronizá-la, e apartir dessa padronização obter uma distribuição t-Student com n-1 graus de liberdade.

Deste modo, fazendo o que foi descrito acima obtemos,

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{S^2}{\sum_{i=1}^{n} x_i^2}}} \sim t_{(n-1)}$$

c)

Considerando a distribuição aproximada de $\hat{\beta}$ obtida no item **b)**, iremos construir um intervalo de confiança para o parâmetro β com coeficiente de confiança γ,$0 < \gamma < 1$

$$\mathrm{IC}(\beta, 100\gamma\%) = [\hat{\beta} - z_\gamma (\frac{S^2}{\sum_{i=1}^{n} x_i^2})^{\frac{1}{2}}, \hat{\beta} + z_\gamma (\frac{S^2}{\sum_{i=1}^{n} x_i^2})^{\frac{1}{2}}]$$

Exercício 2

Considerando o modelo especificado no **Exercício 1**, definamos a reta dos valores esperados de y_i, $i = 1, \ldots, n$ como sendo,

$$\mathrm{E(y_i)} = \beta x_i$$

Então, fixando dois valores x_i e x_{i+1} faremos a seguinte manipulação algébrica,

$$\mathrm{E}(y_i \mid x_{i+1}) - \mathrm{E}(y_i \mid x_i) = \beta x_{i+1} - \beta x_i$$

$$\mathrm{E}(y_i \mid x_{i+1}) - \mathrm{E}(y_i \mid x_i) = \beta (x_{i+1} - x_i)$$

$$\beta = \frac{\mathrm{E}(y_i \mid x_{i+1}) - \mathrm{E}(y_i \mid x_i)}{(x_{i+1} - x_i)}$$

Com o resultado acima, fica mais claro que β corresponde a variação esperada para a variável Y por unidade de variação da variável X, uma vez que o denominador corresponde a variação em uma unidade da variável X e o numerador corresponde à diferença (variação) dos valores esperados de Y, para a correspondente variação da variável X.

Exercício 3

Dado,

$$y_i = \alpha + \beta x_i + e_i, \qquad i = 1, \ldots, 18$$

onde os e_i são erros aleatórios e não correlacionados com $E(e_i) = 0$ e $Var(e_i) = \sigma^2$.

i)

Considerando o modelo dado, podemos interpretar α e β da seguinte forma:

1. $\alpha :=$ "Média das notas independentemente do tipo de escola onde o aluno estudou"
2. $\beta :=$ "Variação média da nota com relação à que tipo de escola se estuda"

ii)

Considere o seguinte script no R:

```
##
## Call:
## lm(formula = notas ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6778 -0.7333 -0.1833  1.0556  1.4667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.8556     0.2552  26.864 9.71e-15 ***
## x             0.3222     0.2552   1.263    0.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.083 on 16 degrees of freedom
## Multiple R-squared:  0.09062,    Adjusted R-squared:  0.03378
## F-statistic: 1.594 on 1 and 16 DF,  p-value: 0.2248
```

Logo, temos,

$$\hat{\alpha} \approx 6.86$$

$$\hat{\beta} \approx 0.32$$

E uma estimativa para σ^2 é,

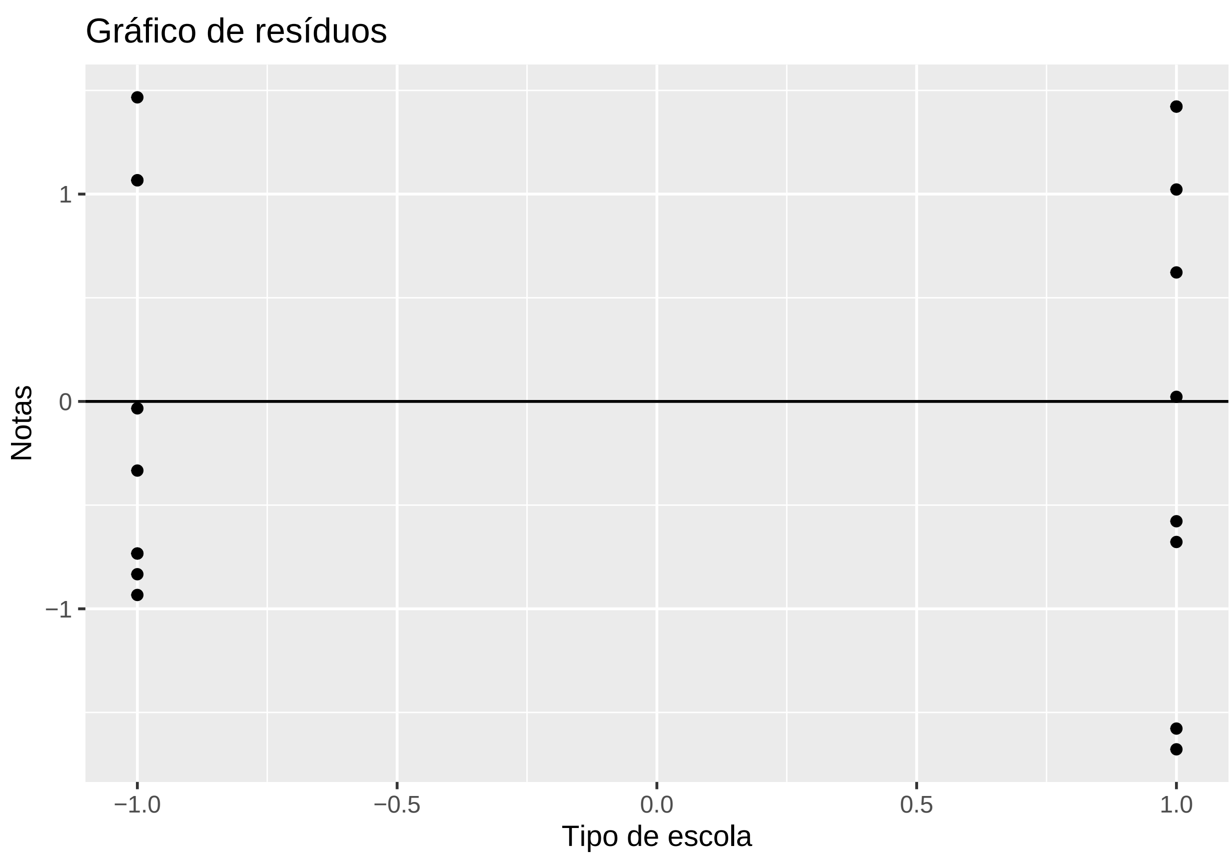
$$\hat{\sigma}^2 \approx 1.17$$

iii)

Em um primeiro instante, como consequência quase direta da observação do script apresentado o item *ii)*, temos,

$$R^2 = 0.090 \quad e \quad R^2_{ajustado} = 0.034$$

Dado que o coeficiente de determinação é bem pequeno, podemos concluir que pouquíssima variabilidade dos dados é explicada pelo modelo. Então iremos em um segundo instante avaliar algumas suposições do modelo que não são detectadas por essa ferramenta.



Com o gráfico de resíduos e o coeficiente de determinação apresentados acima, é razoável considerar a qualidade do ajuste como sendo bem ruim, pois os resíduos não parecem satisfazer a premissa de homocedasticidade. Além disso o coeficiente de determinação ($R^2_{ajustado}$) é extremamente baixo, como já havia sido constatado.

iv)

$$\text{IC}(\alpha, 95\%) \approx [6.31, 7.40]$$
$$\text{IC}(\beta, 95\%) \approx [-0.22, 0.86]$$

v)

Os intervalos de confiança obtidos estão apresentados abaixo

$$\text{IC}(y_1, 95\%) = [6.41, 7.94]$$

$$\text{IC}(y_2, 95\%) = [5.77, 7.30]$$

Onde y_1 representa a nota dos alunos de escola particular e y_2 representa a nota dos alunos de escola pública

vi)

De acordo com o modelo dado, temos apenas dois valores de y_i ajustados, uma vez que a variável explicativa x_i assume apenas 2 valores, sendo assim

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \implies \hat{y}_1 = \hat{y}_2 \iff \hat{\beta} = 0$$

Logo, basta testar se $\hat{\beta} = 0$ e, utilizando o script apresentado no item **ii)** verificamos que o teste $H_0 : \beta = 0$ contra $H_0 : \beta \neq 0$ fornece um $p \approx 0.225$ e deste modo é possível afimar que não há evidência amostral de que $\beta \neq 0$ logo, aceita-se H_0.Podendo então afimar que não há evidência amostral de que os valores esperados das notas sejam diferentes.

vii)

1)

A interpretação de α e β é a mesma apresentado no item **i)**.

2)

Considere o seguinte script no R:

```
##
## Call:
## lm(formula = notas ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6778 -0.7333 -0.1833  1.0556  1.4667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.5333     0.3609  18.103 4.42e-12 ***
## x             0.6444     0.5104   1.263    0.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.083 on 16 degrees of freedom
## Multiple R-squared:  0.09062,    Adjusted R-squared:  0.03378
## F-statistic: 1.594 on 1 and 16 DF,  p-value: 0.2248
```

Logo, temos,

$$\hat{\alpha} \approx 6.53$$

$$\hat{\beta} \approx 0.64$$

E uma estimativa para σ^2 é,

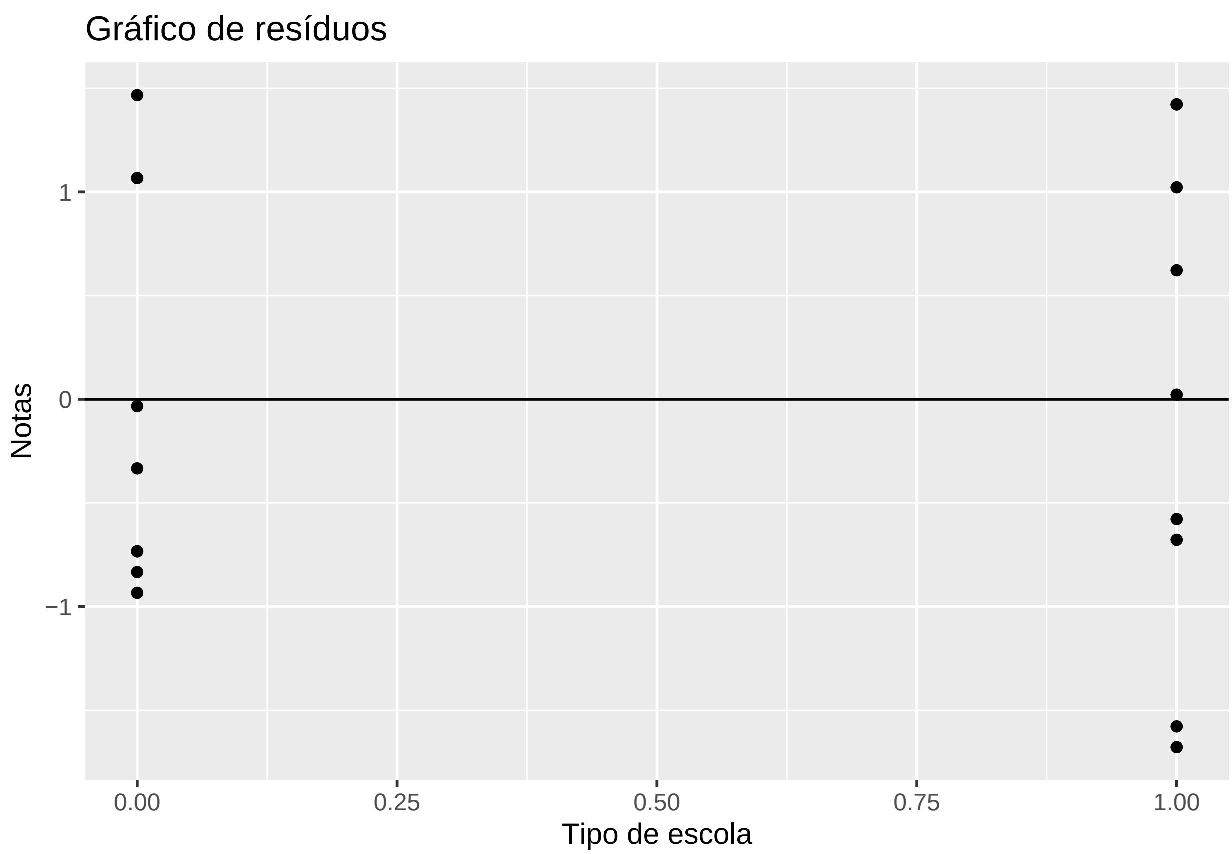
$$\hat{\sigma}^2 \approx 1.08$$

3)

Em um primeiro instante, como consequência quase direta da observação do script apresentado o item *ii)*, temos,

$$R^2 = 0.090 \quad e \quad R^2_{ajustado} = 0.034$$

Dado que o coeficiente de determinação é bem pequeno, podemos concluir que pouquíssima variabilidade dos dados é explicada pelo modelo. Então iremos em um segundo instante avaliar algumas suposições do modelo que não são detectadas por essa ferramenta.



A interpretação do gráfico de resíduos é semelhante à feita em **iii)**.

4)

$$\text{IC}(\alpha, 95\%) \approx [5.77, 7.30]$$
$$\text{IC}(\beta, 95\%) \approx [-0.44, 1.73]$$

5)

Os intervalos de confiança obtidos estão apresentados abaixo

$$\text{IC}(y_1, 95\%) = [6.41, 7.94]$$

$$\text{IC}(y_2, 95\%) = [5.77, 7.30]$$

Onde y_1 representa a nota dos alunos de escola particular e y_2 representa a nota dos alunos de escola pública

6)

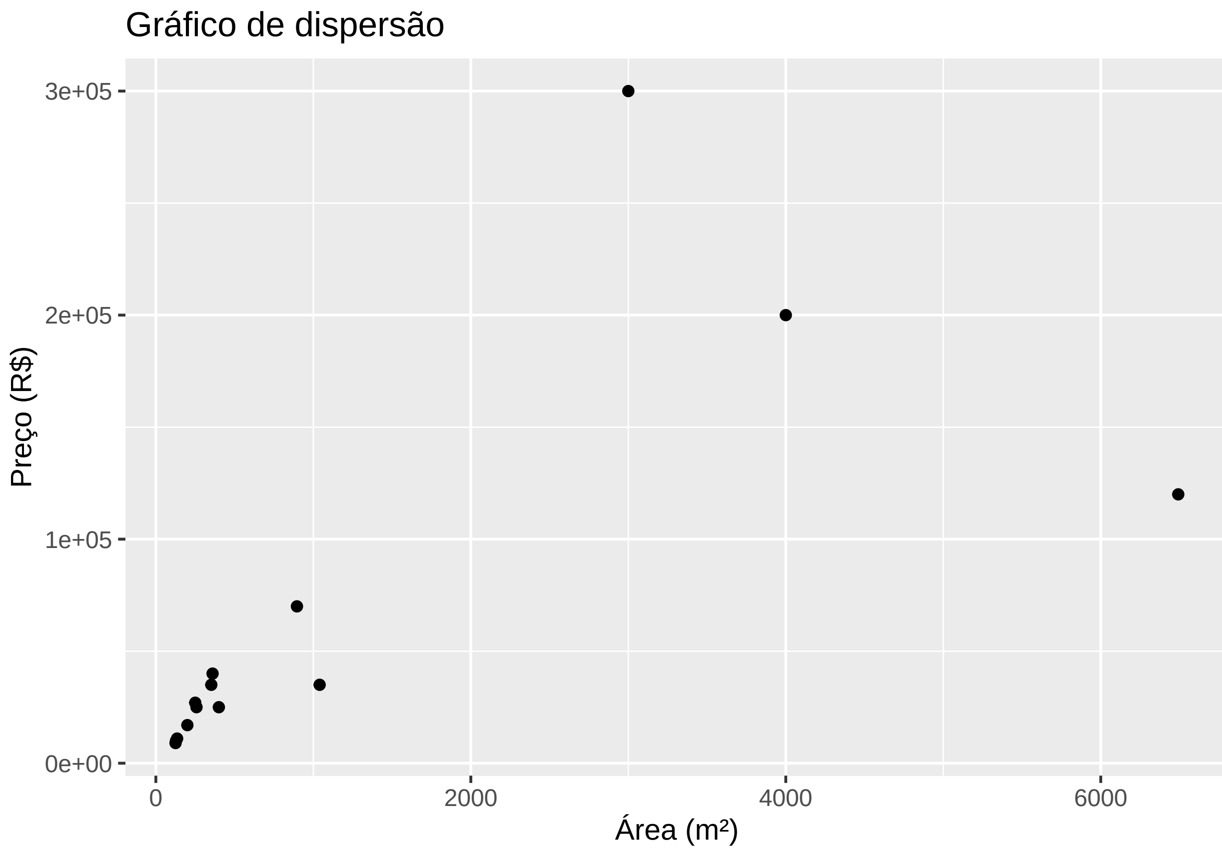
De acordo com o modelo dado, temos apenas dois valores de y_i ajustados, uma vez que a variável explicativa x_i assume apenas 2 valores, sendo assim

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \implies \hat{y}_1 = \hat{y}_2 \iff \hat{\beta} = 0$$

Logo, basta testar se $\hat{\beta} = 0$ e, utilizando o script apresentado no item **ii)** verificamos que o teste $H_0 : \beta = 0$ contra $H_0 : \beta \neq 0$ fornece um $p \approx 0.225$ e deste modo é possível afimar que não há evidência amostral de que $\beta \neq 0$ logo, aceita-se H_0.Podendo então afimar que não há evidência amostral de que os valores esperados das notas sejam diferentes.

Exercício 5

i)



ii)

Considere o script do R,

```
##
## Call:
## lm(formula = preco ~ area)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -108260  -20708  -12137     713  180031
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26934.568  20758.379   1.298  0.21884
## area           31.011      9.306   3.332  0.00597 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64100 on 12 degrees of freedom
```

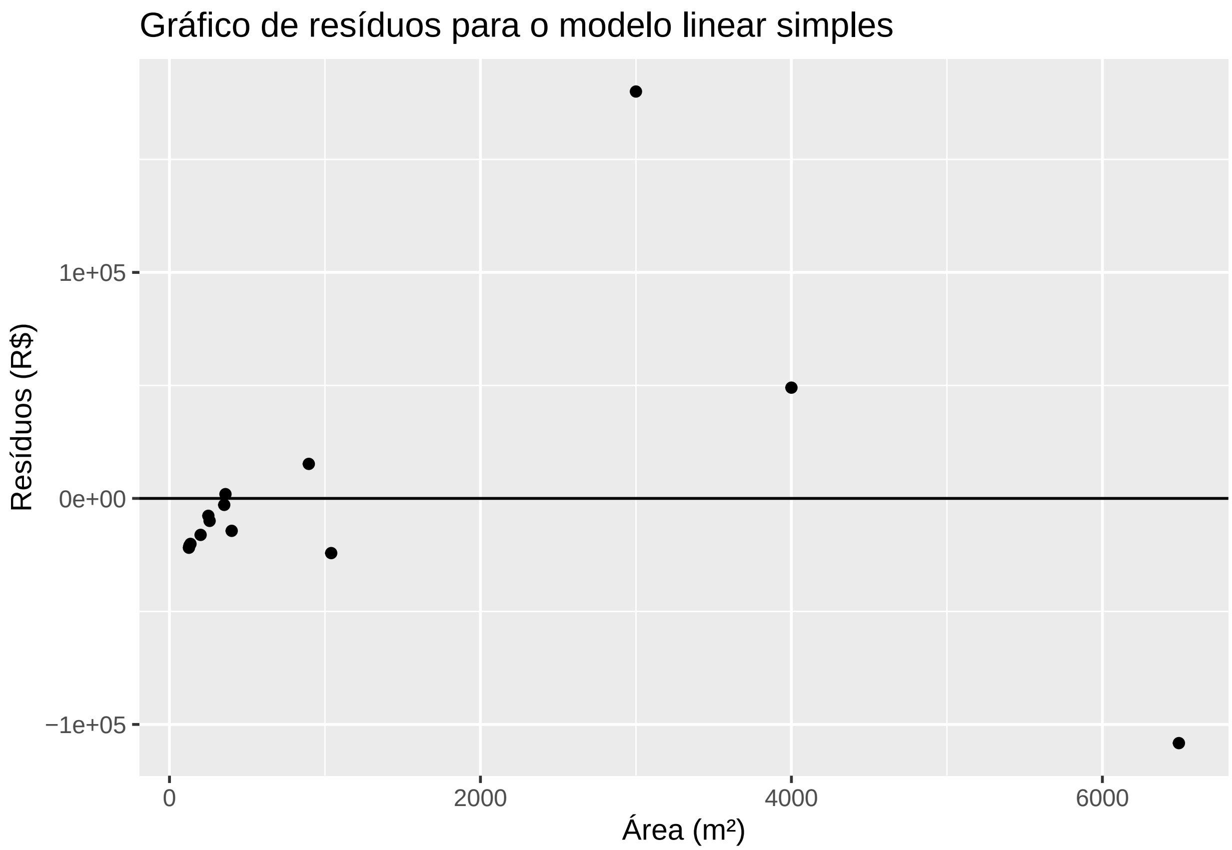
```
## Multiple R-squared:  0.4806, Adjusted R-squared:  0.4374
## F-statistic: 11.11 on 1 and 12 DF,  p-value: 0.005971
```

Assim podemos obter as seguintes estimativas,

$$\hat{\alpha} \approx 26934.6 \qquad \mathrm{EP}_{\hat{\alpha}} = 20758.38$$

$$\hat{\beta} \approx 31.0 \qquad \mathrm{EP}_{\hat{\beta}} = 9.31$$

$$R^2 = 0.48 \quad e \quad R^2_{ajustado} = 0.44$$



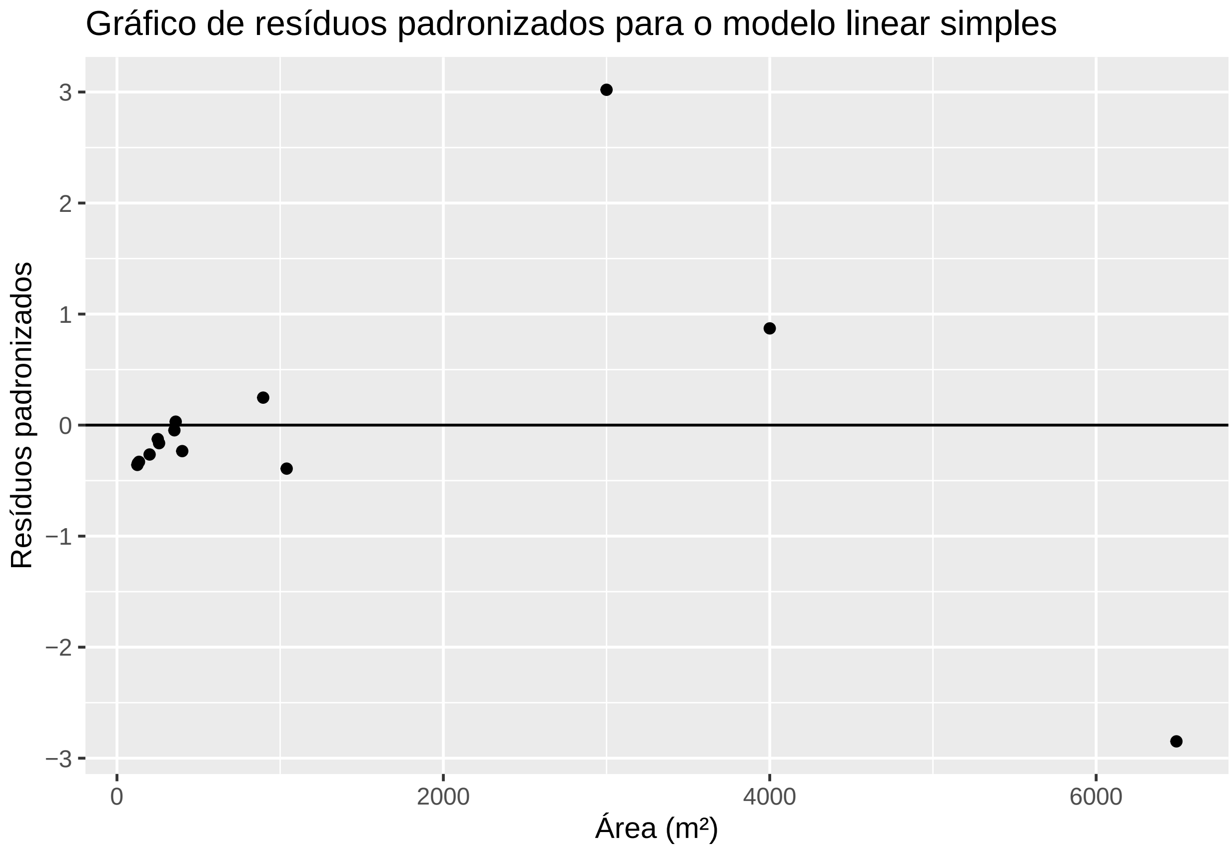
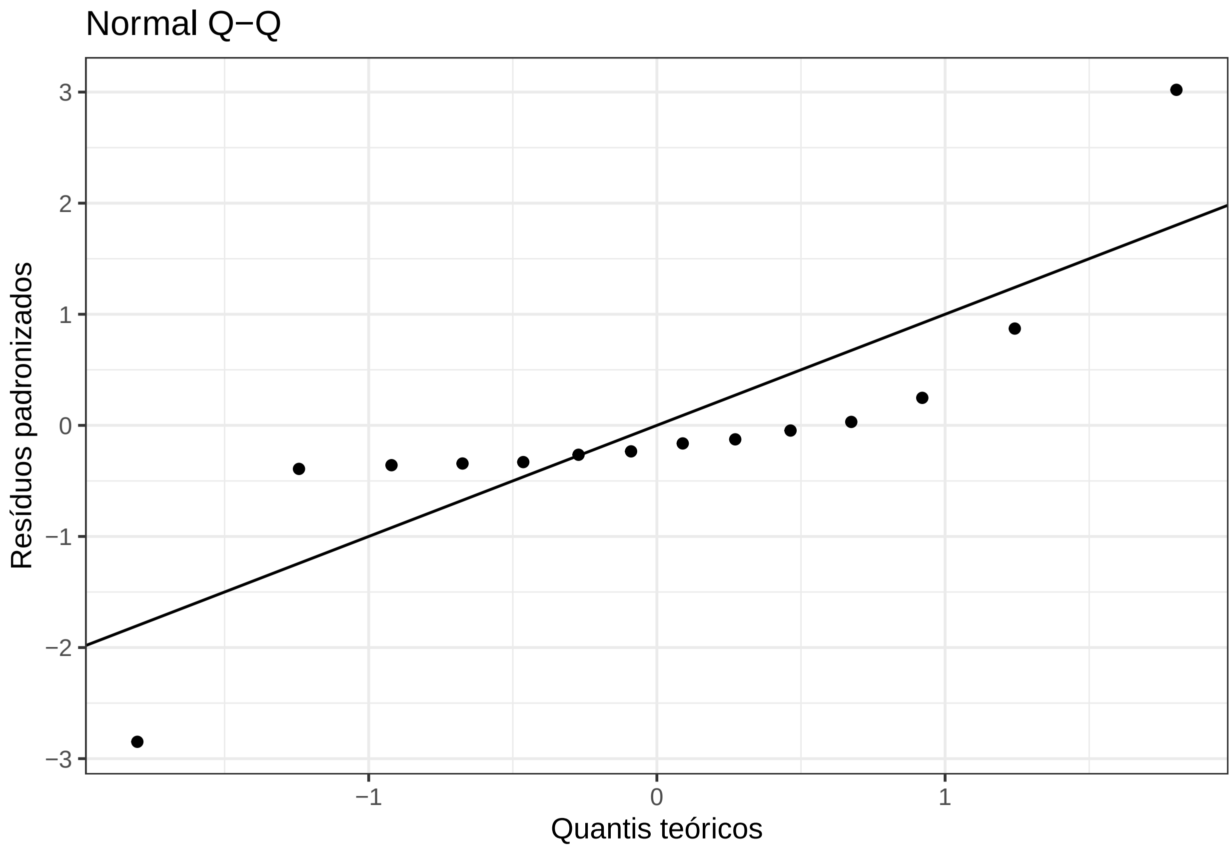
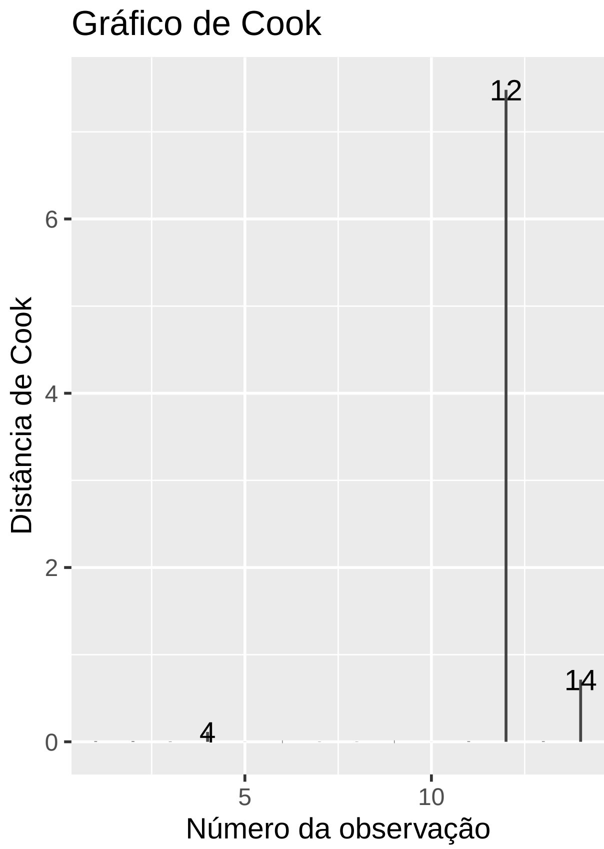


Gráfico de resíduos padronizados para o modelo linear simples
Resíduos padronizados
3
2
1
0
−1
−2
−3
0
2000
4000
6000
Área (m²)


Normal Q−Q
Resíduos padronizados
Quantis teóricos
3
2
1
0
−1
−2
−3
−1
0
1



O script do R a seguir está relacionado ao mesmo modelo de regressão linear simples porém, com a remoção dos pontos de alavanca indicados pelo **gráfico de Cook** acima (as observações 4, 12 e 14 foram removidas).

```
##
## Call:
## lm(formula = preco ~ area, data = newdat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21639  -6414  -2909   6675  19658
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11163.50    5858.57   1.905  0.08910 .
## area           43.73      12.25   3.570  0.00602 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11970 on 9 degrees of freedom
## Multiple R-squared:  0.5862, Adjusted R-squared:  0.5402
## F-statistic: 12.75 on 1 and 9 DF,  p-value: 0.006021
```

Deste modo podemos então observar um novo $R^2_{ajustado}$,removendo os pontos de alavanca,

$$R^2_{ajustado} = 0.54$$

Como o $R^2_{ajustado}$, com a remoção das observações específicadas, é maior entende-se que este é um modelo com mais qualidade do que o seu anterior (com todas as observações).

iii)

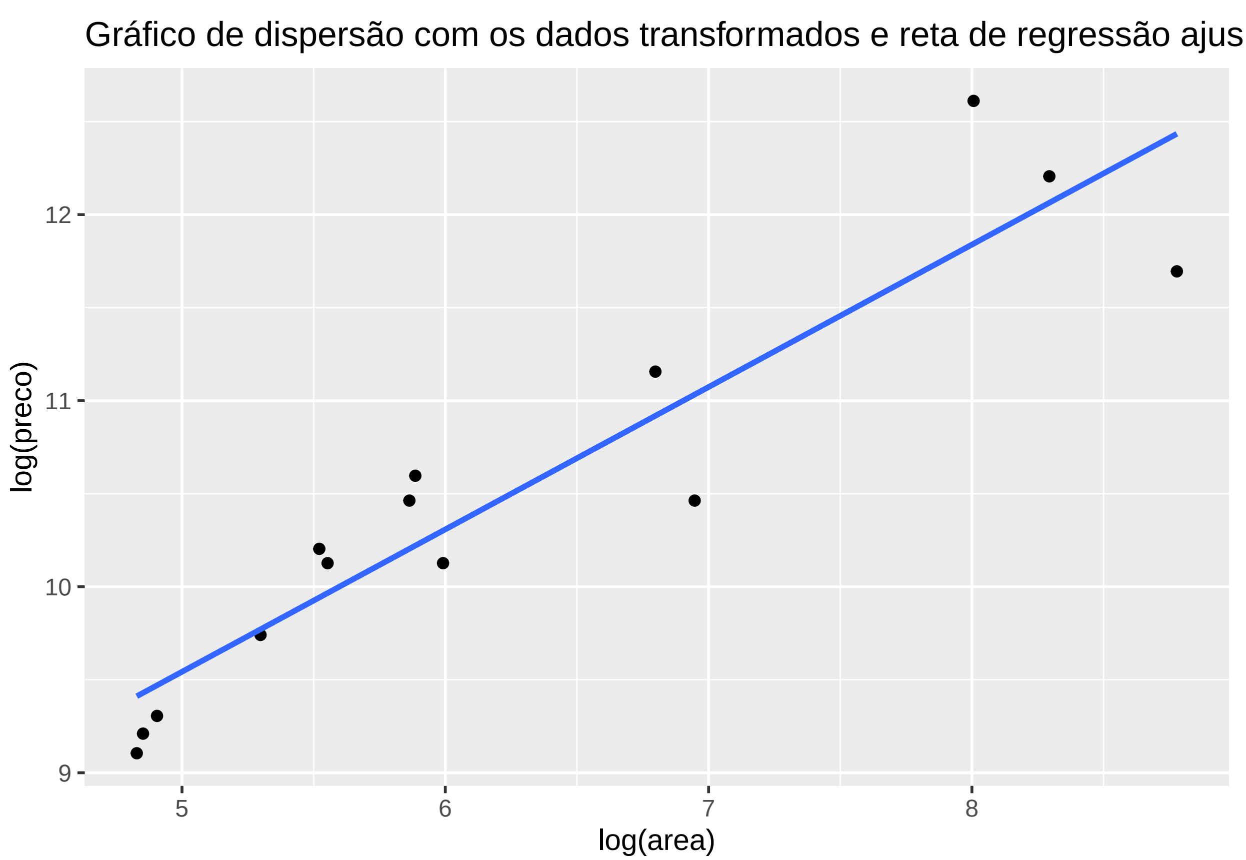
Considere o modelo,

$$y_i = \beta x_i^{\gamma} e_i$$

Linearizando-o por meio de uma transformação logarítimica obtemos,

$$log(y_i) = log(\beta) + \gamma log(x_i) + log(e_i)$$

```
##
## Call:
## lm(formula = log(preco) ~ log(area), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73940 -0.20812  0.05598  0.25384  0.76774
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.71583    0.55460  10.306 2.58e-07 ***
## log(area)    0.76539    0.08694   8.803 1.39e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4124 on 12 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8547
## F-statistic:  77.5 on 1 and 12 DF,  p-value: 1.393e-06
```



Temos então, que para a linearização neste modelo,

$$R^2_{ajustado} = 0.85$$

Comparando o $R^2_{ajustado}$ do modelo linear simples do item **ii)** com o do modelo linearizado do item **iii)** podemos dizer o que o modelo linearizado possui maior qualidade no ajuste.É válido evidenciar o fato de que, mesmo com a remoção dos pontos de alavanca, o modelo de regressão linear simples, não explicar melhor a variabilidade do que a linearização do modelo dado em **iii)**.

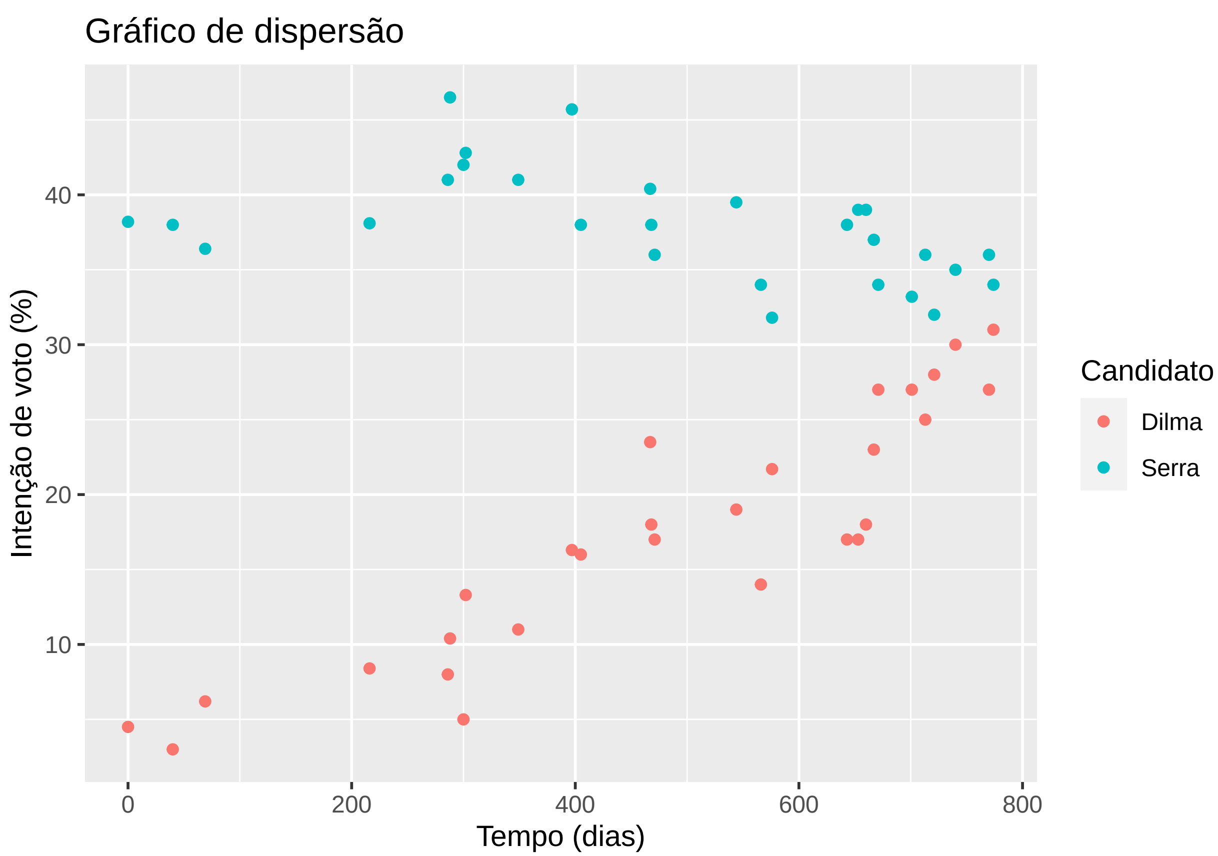
iv)

Logo, usando o modelo do item **iii)**,

$$\text{IC}(y_{200}, 95\%) = [441.9, 2678.5]$$
$$\text{IC}(y_{500}, 95\%) = [514.0, 2939.6]$$
$$\text{IC}(y_{1000}, 95\%) = [568.1, 3126.6]$$

Exercício 7

a)



b)

Considere os seguintes modelos

$$y_{it} = \alpha_i + \beta_i x_t + \gamma_i x_t^2 + e_{it} \qquad i = 1, 2 \quad e \quad t = 1, ..., 28$$

onde y_{it} representa a porcentagem de intenção de voto no canditado i no instante t, α_i denota o valor esperado da intenção de voto no candidato i no começo da disputa eleitoral, β_i e γ_i representam os coeficientes dos termos linear e quadrático da curva que rege a variação temporal de intenção de voto no canditado i no intervalo de tempo estudado e e_{it} denota um erro aleatório com média 0 e variância σ^2 (isto é, $E(e_{it}) = 0$ e $Var(e_{it}) = \sigma^2$). Utilizaremos t como índice para salientar que as observações são colhidas sequencialmente ao longo do tempo.Os candidatos estão codificados por 1 e 2 onde 1 representa a Dilma e 2 representa o Serra.

Assim temos dois modelos,

$$y_{1t} = \alpha_1 + \beta_1 x_t + \gamma_1 x_t^2 + e_{1t} \qquad t = 1, ..., 28$$

$$y_{2t} = \alpha_2 + \beta_2 x_t + \gamma_2 x_t^2 + e_{2t} \qquad t = 1, ..., 56$$

c)

Para a candidata Dilma temos,

```
##
## Call:
## lm(formula = porcent ~ tempo + I(tempo^2), data = data, subset = (cand ==
##     "Dilma"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9941 -1.2144  0.4919  2.1134  7.5737
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.819e+00  2.176e+00   1.755   0.0915 .
## tempo       1.729e-02  1.123e-02   1.540   0.1362
## I(tempo^2)  1.849e-05  1.302e-05   1.420   0.1680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.413 on 25 degrees of freedom
## Multiple R-squared:  0.8404, Adjusted R-squared:  0.8277
## F-statistic: 65.84 on 2 and 25 DF,  p-value: 1.088e-10
```

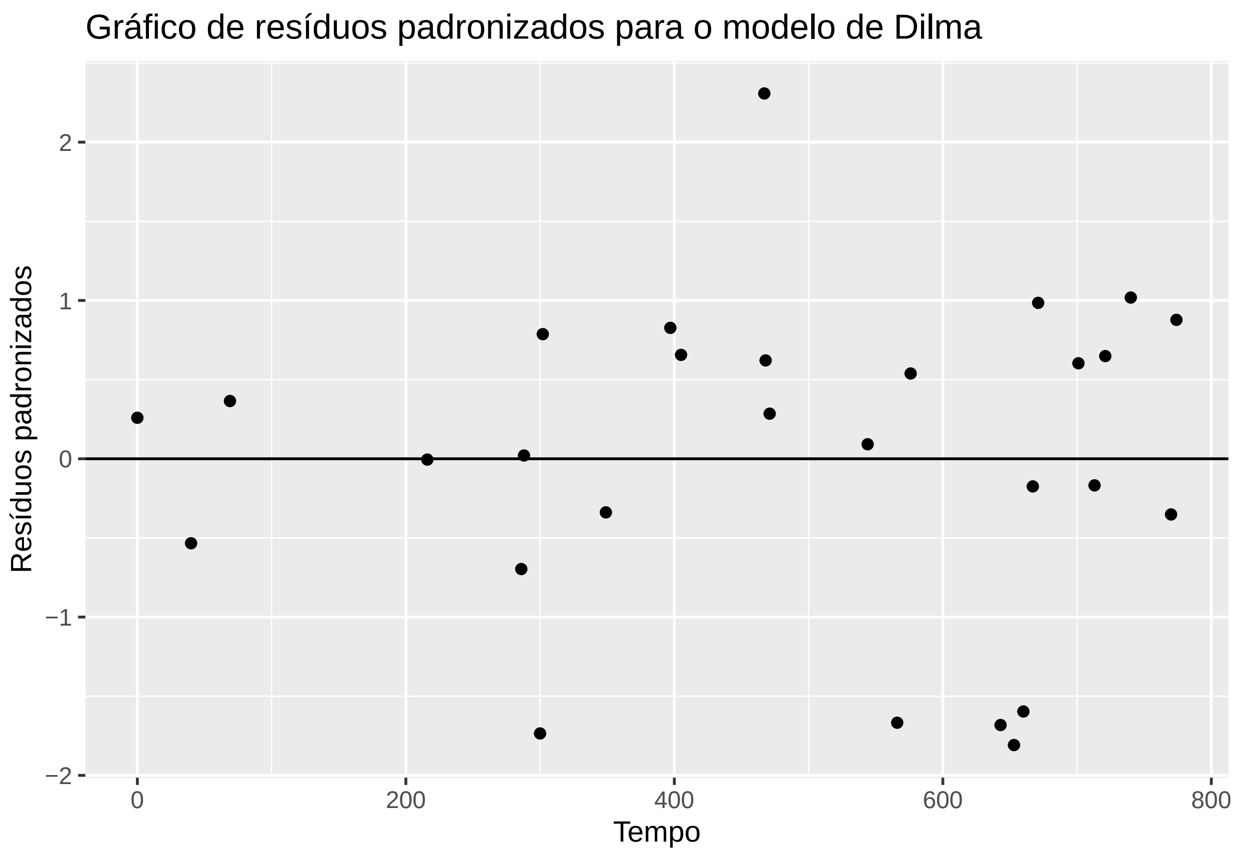
Para o candidato Serra temos,

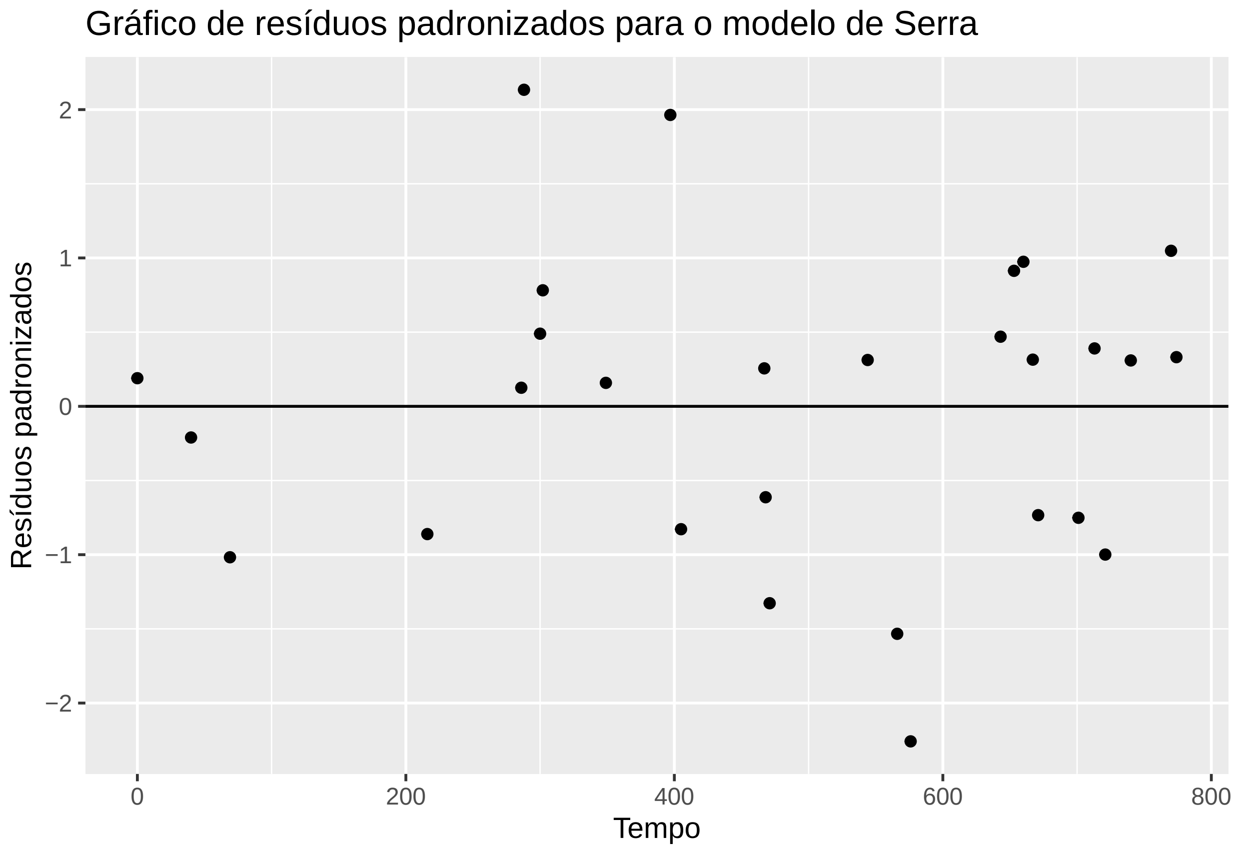
```
##
## Call:
## lm(formula = porcent ~ tempo + I(tempo^2), data = data, subset = (cand ==
##     "Serra"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -6.2778 -2.1120  0.5685  1.3137  5.8420
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.778e+01  1.823e+00  20.722  < 2e-16 ***
## tempo        1.946e-02  9.409e-03   2.068  0.04913 *
## I(tempo^2)  -3.289e-05  1.091e-05  -3.015  0.00582 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.859 on 25 degrees of freedom
## Multiple R-squared:  0.4451, Adjusted R-squared:  0.4007
## F-statistic: 10.03 on 2 and 25 DF,  p-value: 0.0006351
```

d)

Primeiramente analisando os resíduos dos dois modelos,





Com base na analise dos gráficos de resíduos, é razoável considerar satisfeita a hipóteses de homocedasticidade dos erros. Além disso, olhando os scripts do item **c)**, é coerente verificar como ficaria o modelo com a remoção do termo quadratico para o modelo que explica a variação da intenção de votos em porcentagem na candidata Dilma.E também é notório que os parâmetros do modelo ajustado para o candidato Serra, por causa dos valores -p bem pequenos dos testes de hipótese que testam se os parâmetros são nulos, é coerente refutar a ideia de que o modelo pode ser simplificado. Deste modo o novo modelo para Dilma,

```
##
## Call:
## lm(formula = porcent ~ tempo, data = data, subset = (cand ==
##     "Dilma"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4249 -0.9812  0.1408  2.3627  6.6129
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.612618   1.553208   1.038    0.309
## tempo       0.032708   0.002928  11.171 2.01e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.479 on 26 degrees of freedom
## Multiple R-squared:  0.8276, Adjusted R-squared:  0.8209
## F-statistic: 124.8 on 1 and 26 DF,  p-value: 2.015e-11
```

E ainda, seguindo este novo resultado é razoável supor que trata-se de uma regressão simples sem intercepto (regressão linear simples que passa pela origem),pois o valor-p para um teste de $H_0 : \alpha = 0$ contra $H_1 : \alpha \neq 0$ é $p \approx 0.305$ indicando que é uma suposição comedida. Um novo modelo mais simples para intenção de voto da canditada Dilma é

$$y_{1t} = \beta_1 x_{1t}$$

E não há evidências de que seja possível ajustar um modelo mais simples para o candidato Serra.

e)

Em 3 de outubro de 2010, a quantidade de dias a partir de 16 de fevereiro de 2008 é 960 dias

$$\text{IC}(y_{1,960^*}, 95\%) = [29.82, 36.20]$$
$$\text{IC}(y_{2,960^*}, 95\%) = [20.14, 32.16]$$

Logo,

$$\text{IC}(y_{2,960^*} - y_{1,960^*}, 95\%) = [20.14 - 36.20, 32.16 - 29.82] = [-16.06, 2.34]$$
$$\text{IC}(y_{2,960^*} - y_{1,960^*}, 95\%) = [-16.06, 2.34]$$

Onde, $y_{i,960^*}$ representa a porcentagem da intenção de voto no candidato i, no dia 960, contado a partir da primeira pesquisa eleitoral.

f)

Criticar?

Exercício 15

$\vdash SQTot = SQRes + SQReg$

Para isso, consideraremos as seguintes expressões

$$SQTot = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad SQRes = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad SQReg = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Note que as expressões acima são definições para o caso de *Regressão linear simples* então podemos especificar o modelo associado,

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \qquad (1)$$

Além disso iremos enfatizar as equações de estimação(obtidas através das derivadas parciais em α e β ,igualadas à zero,da soma dos quadrados dos resíduos),pois serão usadas em um ponto essencial dessa demonstração.

$$n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (2)$$

$$\hat{\alpha} \sum_{i=1}^{n} x_i + \hat{\beta} \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (3)$$

Assim, desenvolvendo o primeiro membro da equação dada temos,

$$SQTot = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2$$

$$SQTot = \sum_{i=1}^{n}[(y_i - \hat{y}_i)^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2]$$

$$SQTot = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$SQTot = SQRes + SQReg + \sum_{i=1}^{n}2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

Então basta, $\vdash \sum_{i=1}^{n} 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$

$$\sum_{i=1}^{n}2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 2\sum_{i=1}^{n}[\hat{y}_i(y_i - \hat{y}_i) - \bar{y}(y_i - \hat{y}_i)]$$

Denominando $e_i = y_i - \hat{y}_i$ então,

$$2\sum_{i=1}^{n}[\hat{y}_i(y_i - \hat{y}_i) - \bar{y}(y_i - \hat{y}_i)] = 2\sum_{i=1}^{n}[\hat{y}_i e_i - \bar{y}e_i] = 2\sum_{i=1}^{n}\hat{y}_i e_i - 2\bar{y}\sum_{i=1}^{n}e_i$$

Agora, basta que,

$$\sum_{i=1}^{n}\hat{y}_i e_i = 0 \quad (*) \quad e \quad \bar{y}\sum_{i=1}^{n}e_i = 0 \quad (**)$$

Primeiro mostremos que a soma dos resíduos é nula,

$$\sum_{i=1}^{n}e_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)$$

$$\sum_{i=1}^{n}e_i = \sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i) = \sum_{i=1}^{n}y_i - n\hat{\alpha} - \sum_{i=1}^{n}\hat{\beta}x_i$$

Usando (2),

$$\sum_{i=1}^{n}e_i = 0$$

Como consequência direta desse resultado a equação $(**)$ vale.

Para $(*)$,

$$\sum_{i=1}^{n}\hat{y}_i e_i = \sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i)e_i = \hat{\alpha}\sum_{i=1}^{n}e_i + \hat{\beta}\sum_{i=1}^{n}x_i e_i = 0 + \hat{\beta}\sum_{i=1}^{n}x_i e_i$$

Por fim,

$\vdash \sum_{i=1}^{n} x_i e_i = 0$

$$\sum_{i=1}^{n}x_i e_i = \sum_{i=1}^{n}x_i(y_i - \hat{y}_i) = \sum_{i=1}^{n}x_i(y_i - \hat{\alpha} - \hat{\beta}x_i)$$

$$\sum_{i=1}^{n}x_i e_i = \sum_{i=1}^{n}x_i(y_i - \hat{\alpha} - \hat{\beta}x_i) = \sum_{i=1}^{n}x_i y_i - \hat{\alpha}\sum_{i=1}^{n}x_i - \hat{\beta}\sum_{i=1}^{n}x_i^2$$

Usando (3) temos que

$$\sum_{i=1}^{n}x_i e_i = 0$$

Sendo assim como as equações $(*)$ $(**)$ valem então isso implica que,

$$\sum_{i=1}^{n} 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$$

Logo, fica provado que,

$$SQTot = SQRes + SQReg$$

Observação: Nota-se que as equações (2) e (3) valem pois, essas equações são utilizadas na **construção** dos **estimadores de mínimos quadrados** ($\hat{\alpha}$ e $\hat{\beta}$) e portanto são *"forçadas"* a serem válidas.