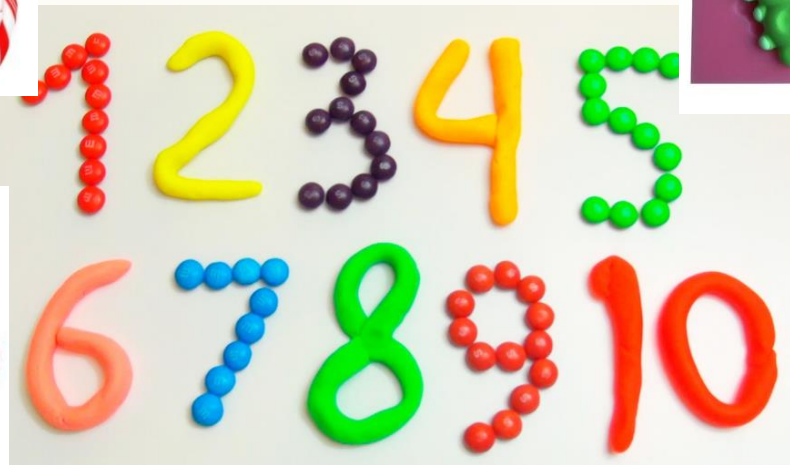


Aula 15 Goodies*



* Goodies related to animals, plants and numbers...



Realwokieleaks

@realwokieleaks

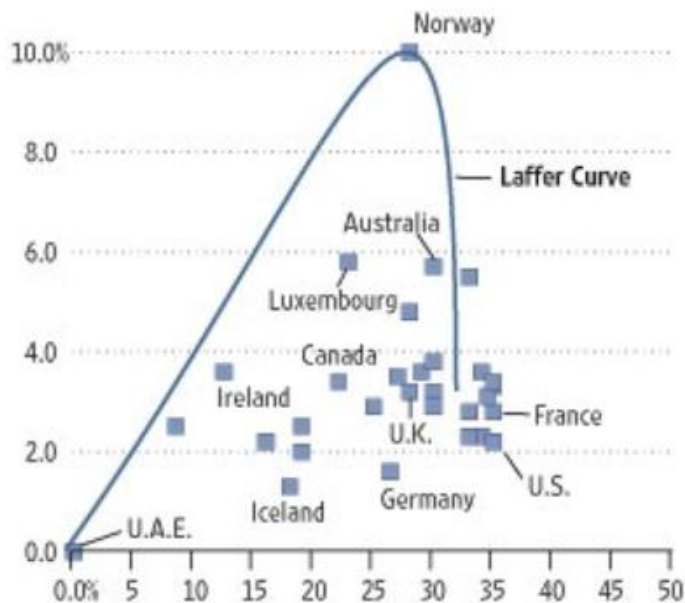
Follow



This is the greatest economics graph I have ever seen. i've been laughing about this for like 3 days

Corporate Taxes and Revenue, 2004

Left scale represents tax revenues as a percentage of GDP. Bottom scale represents central government corporate tax rates.



Sources: OECD Revenue Statistics, Kevin Hassett, American Enterprise Institute

https://swirlstats.com/

Home

Learn

Teach

Contribute

Blog

FAQ

Help

{swirl}

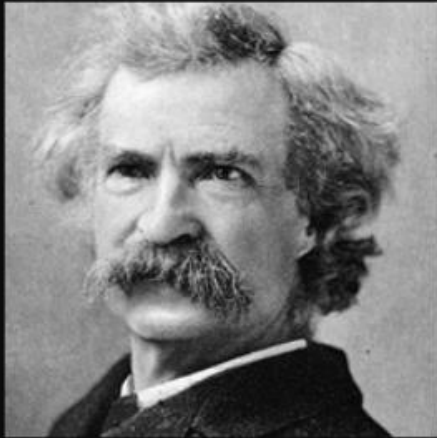
Learn R, in R.

swirl teaches you R programming and data science
interactively, at your own pace, and right in the R console!

 Follow @swirlstats

Got questions? Join our [discussion group!](#)

```
install.packages("swirl")  
library("swirl")  
swirl()
```



Data is like garbage. You'd better know what you are going to do with it before you collect it.

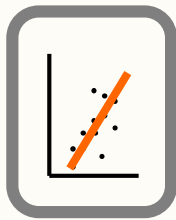
~ Mark Twain

AZ QUOTES

<https://www.azquotes.com/quote/660697>

A decorative graphic of a feather, rendered in a light beige or tan color, positioned on the left side of the slide. The feather has a central rachis with numerous barbs extending outwards, creating a fan-like shape. It is oriented vertically, pointing downwards.

Regressão Multipla



Análise da regressão múltipla

Âmbito:

Avaliar a relação entre a variável dependente e múltiplas variáveis independentes

Objectivos:

- Modelação dum evento ecológico
- Testes de hipóteses
- Predição (modelo preditivo)



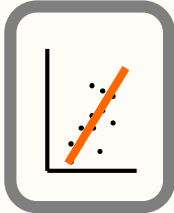
Análise da regressão múltipla

Quando existe mais do que uma variável independente

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

cada observação é dada por:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

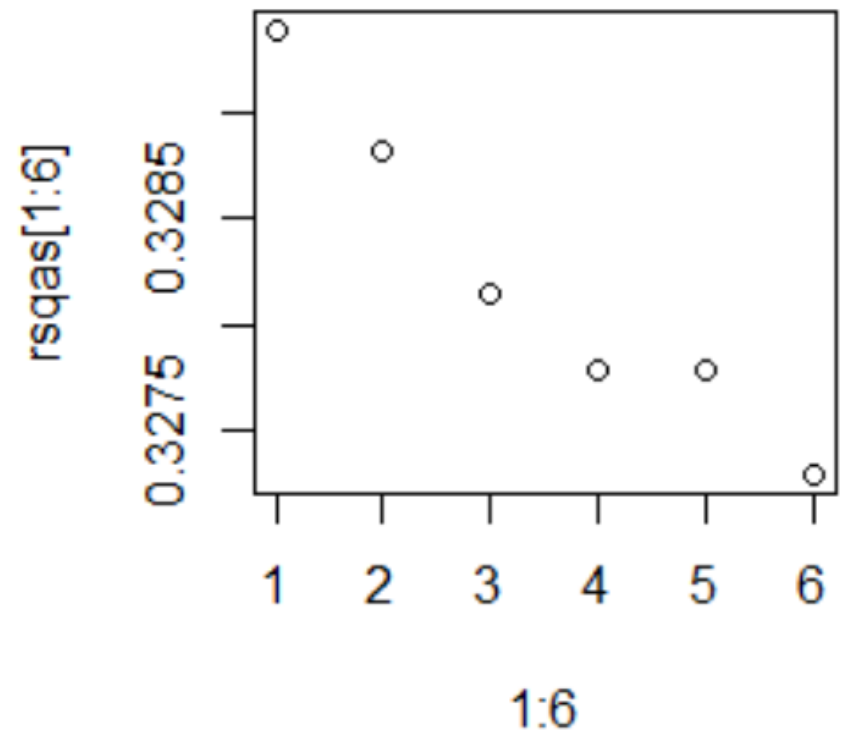
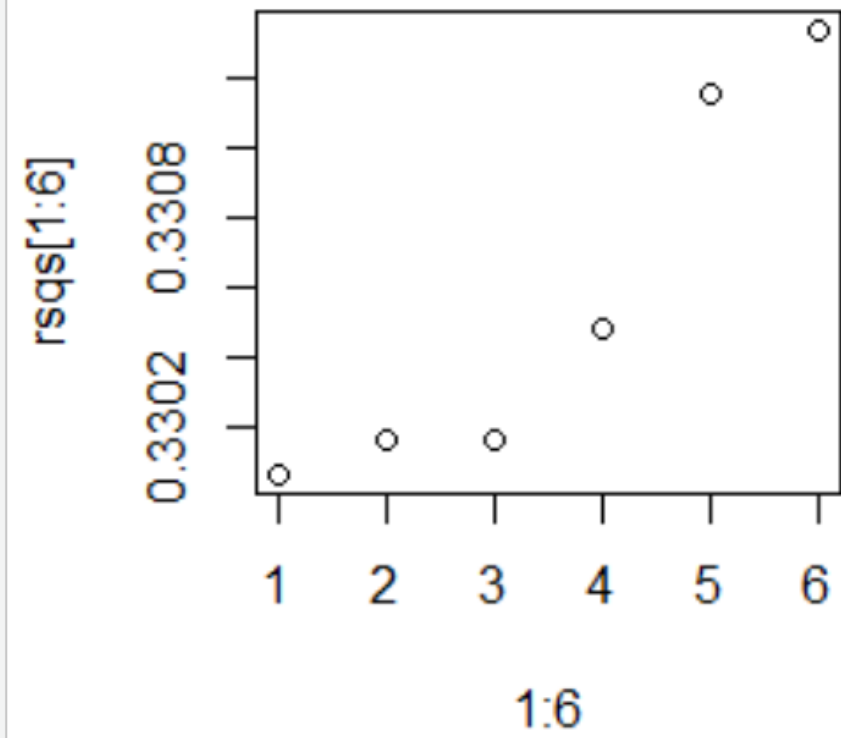


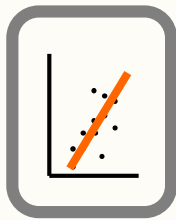
Análise da regressão múltipla

Variabilidade explicada pelo modelo

O coeficiente de determinação R^2 depende de p e n e, por isso, uma estimativa mais adequada é o R^2 ajustado, dado por:

$$R_a^2 = 1 - \frac{MS_{RESIDUAL}}{MS_{TOTAL}} = 1 - \frac{n-1}{n-p-1} (1 - R^2)$$





Testes de hipóteses na análise da regressão

Testes F

Hipóteses:

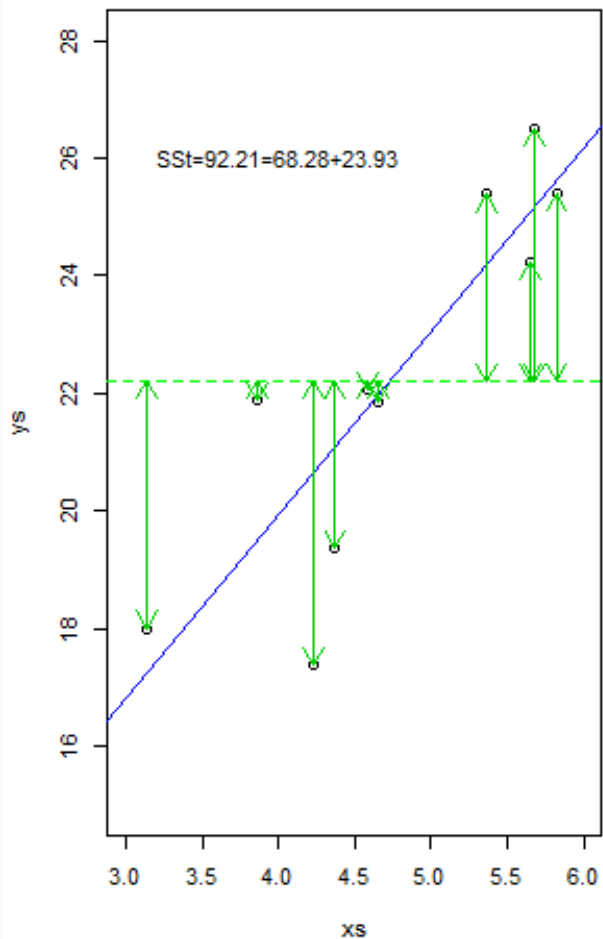
$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H_1 : Pelo menos um dos β_i é diferente de 0

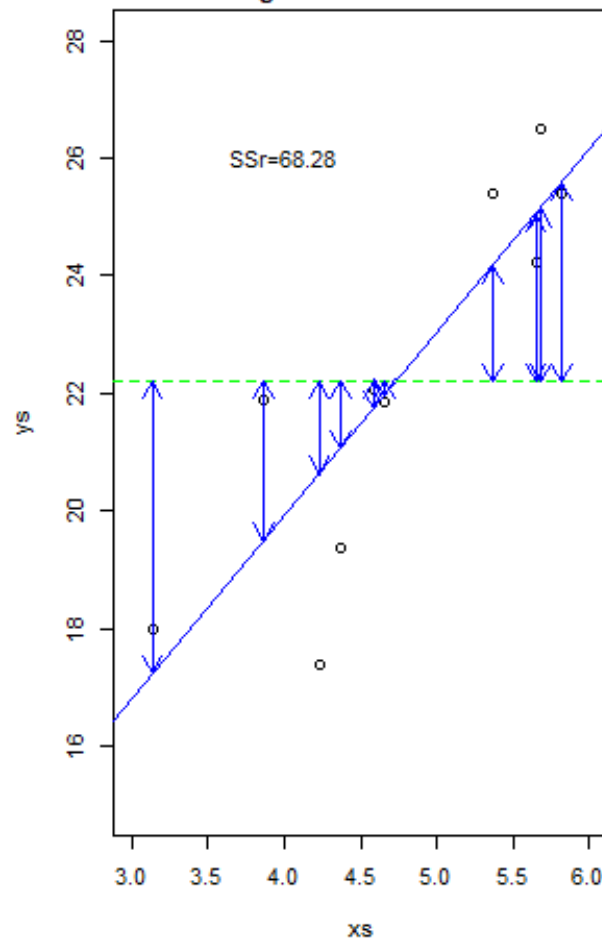
$$SQ_{TOTAL} = SQ_{REGRESSÃO} + SQ_{residual (erro)}$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

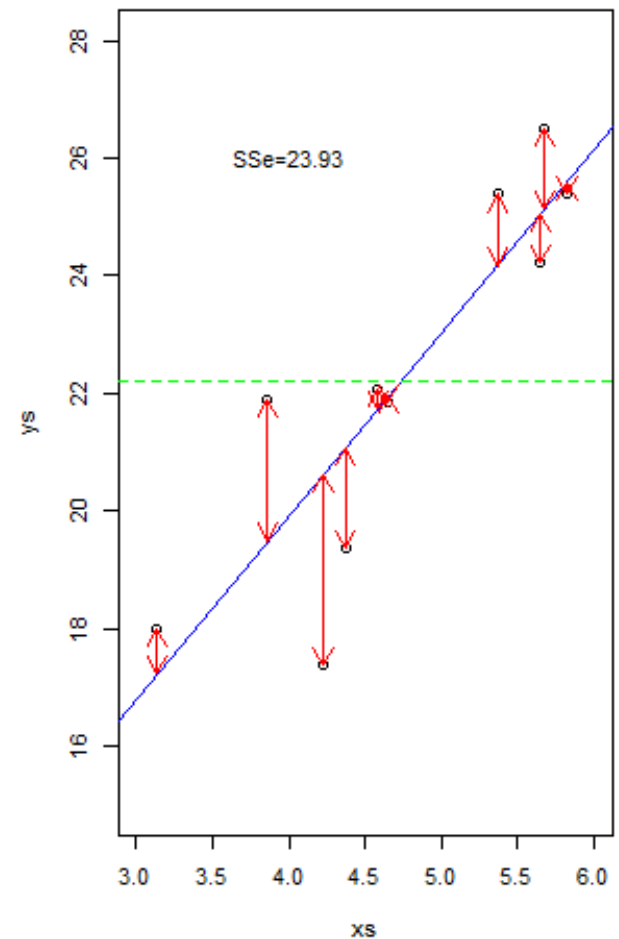
Total Variation



Regression Variation



Error Variation





Testes de hipóteses na análise da regressão

Estatística de teste:

$$F = \frac{\frac{SQ_{REGRESSÃO}}{v_{REGRESSÃO}}}{\frac{SQ_{ERRO}}{v_{ERRO}}} = \frac{QM_{REGRESSÃO}}{QM_{ERRO}}$$

sendo $gl_{regressão} = p - 1$ ($p = n^{\circ}$ de coeficientes) e
 $gl_{erro} = n - 2$ ($n = n^{\circ}$ observações)

Valor crítico:

$$F_{\alpha, v_{reg}, v_{erro}}$$

Critério de decisão:

Rejeitar H_0 se:

$$F > F_{\alpha, v_{reg}, v_{erro}}$$

Não rejeitar H_0 caso contrário

$$gl=p-1=2-1=1$$

$$gl=n-2=10-2=8$$

```
> summary(aov(ys~xs))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
xs	1	68.28	68.28	22.82	0.0014 **
Residuals	8	23.93	2.99		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

H0: xs e ys são independentes, ou xs não influencia ys

```
> summary(lm(ys~xs))
```

Call:

```
lm(formula = ys ~ xs)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.2431	-0.6694	0.1048	1.1045	2.3948

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.454	3.135	2.377	0.0447 *
xs	3.115	0.652	4.778	0.0014 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.73 on 8 degrees of freedom

Multiple R-squared: 0.7405, Adjusted R-squared: 0.708

F-statistic: 22.82 on 1 and 8 DF, p-value: 0.001395

```
> 1-pf(22.82,1,8)
[1] 0.001395977
> 2*(1-pt(4.778,8))
[1] 0.00139424
```

H0: $\beta_0=0$
H0: $\beta_1=0$



Testes de hipóteses na análise da regressão

Hipótese:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Estatística de teste:

$$t = \frac{\hat{\beta} - 0}{s_{\hat{\beta}}}$$

onde

$$s_{\hat{\beta}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}}{\sum x^2}$$

Valor crítico:

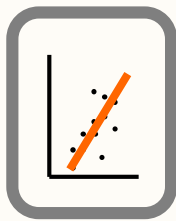
$$t_{\alpha(2), n-2}$$

Critério de decisão:

Rejeitar H_0 se:

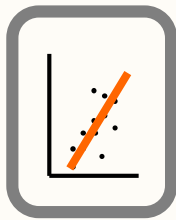
$$t > t_{\alpha(2), n-2}$$

Não rejeitar H_0 caso contrário



Pressupostos da análise da regressão

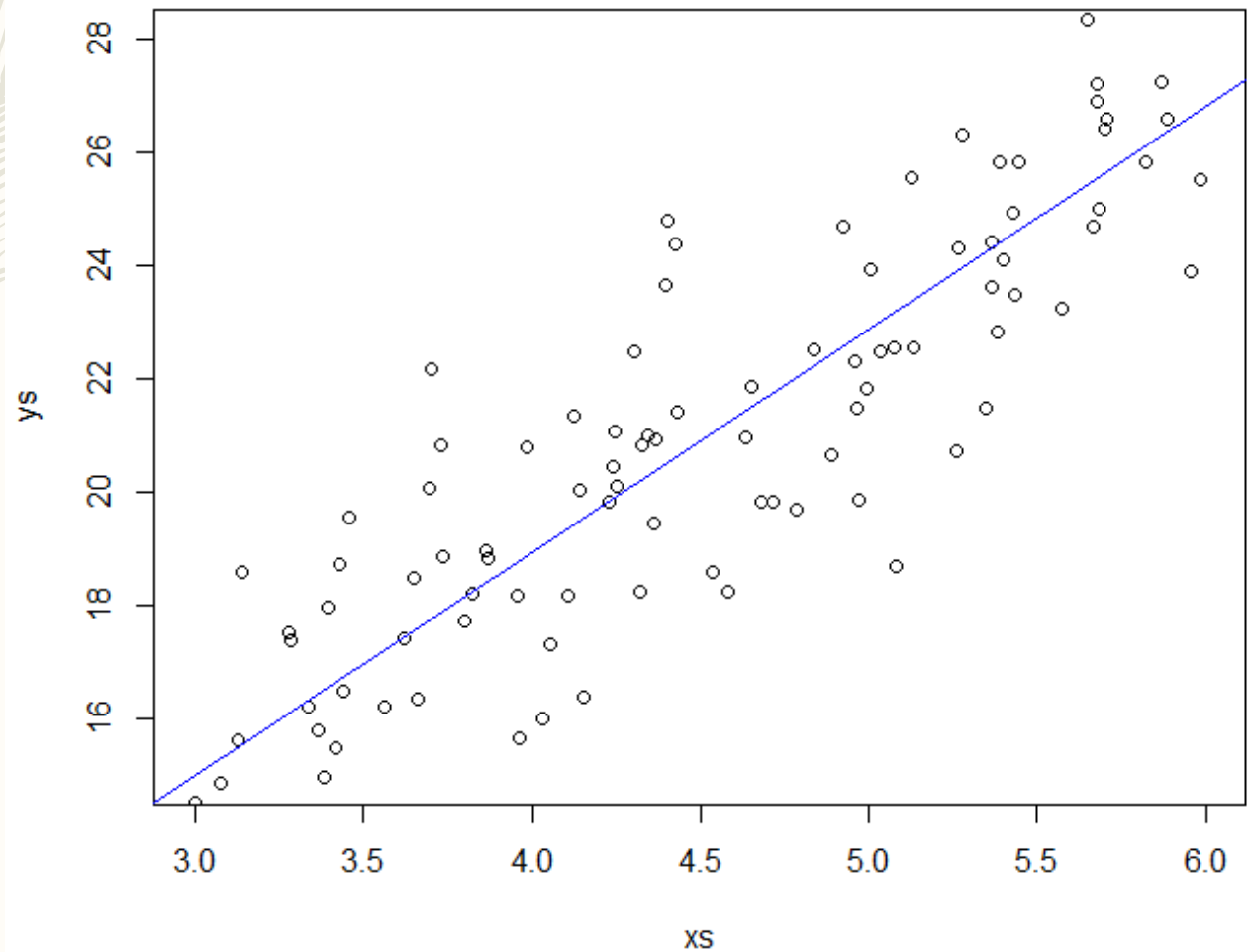
- A distribuição dos erro é normal de média 0 e variância constante σ^2 ;
- Os erros são independentes, i.e. não correlacionados;
- Não há erros em X (ou na prática as medições de X são obtidas com erro negligível comparativamente às de Y)



Análise de resíduos

- Normalidade
- Homogeneidade de variâncias
- Erros independentes (inexistência de autocorrelação dos resíduos)

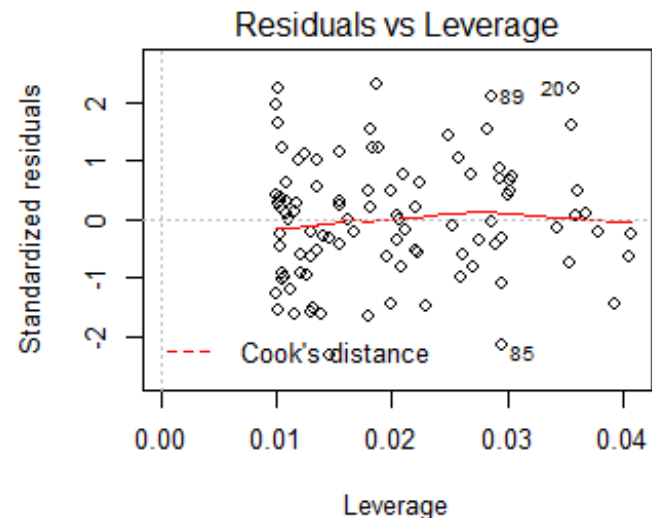
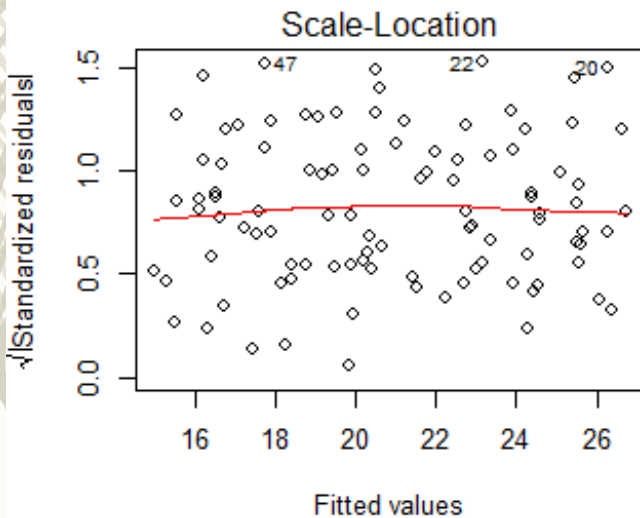
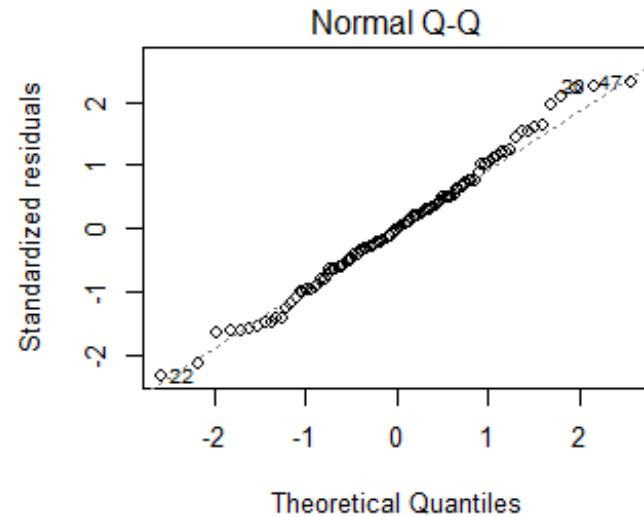
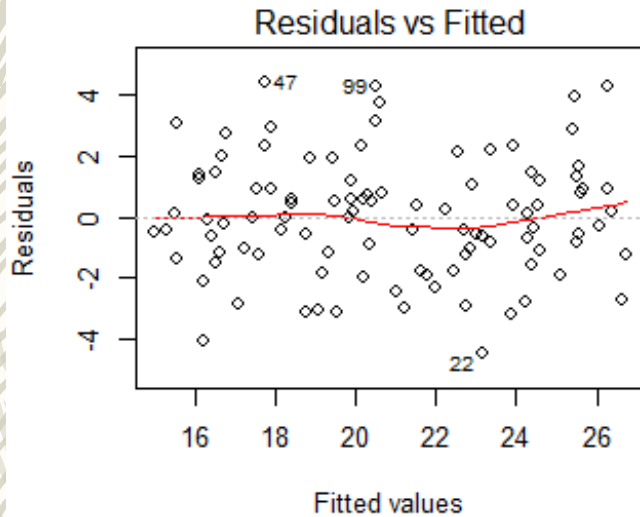
```
par(mfrow=c(1,1),mar=c(4,4,2,1))
set.seed(123)
n=100;slope=4;intercept=3
xs=runif(n,3,6)
ys=intercept+slope*xs+rnorm(n,mean=0,sd=2)
plot(xs,ys,ylim=c(15,28),xlim=c(3,6),main="")
mylm1=lm(ys~xs)
abline(mylm1,col=4)
```



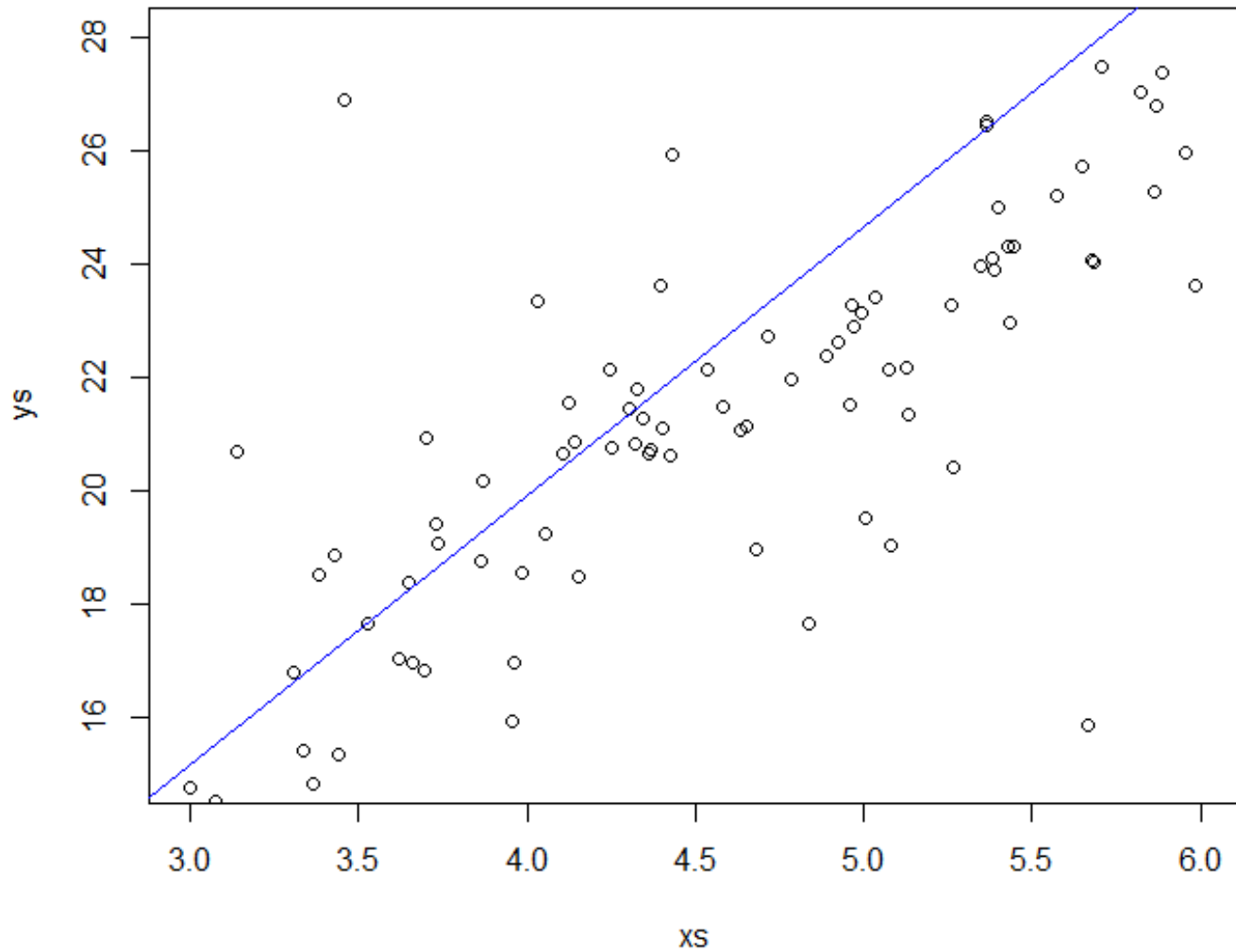
```
par(mfrow=c(2,2),mar=c(4,4,2,1))
```

```
plot(mylm1)
```

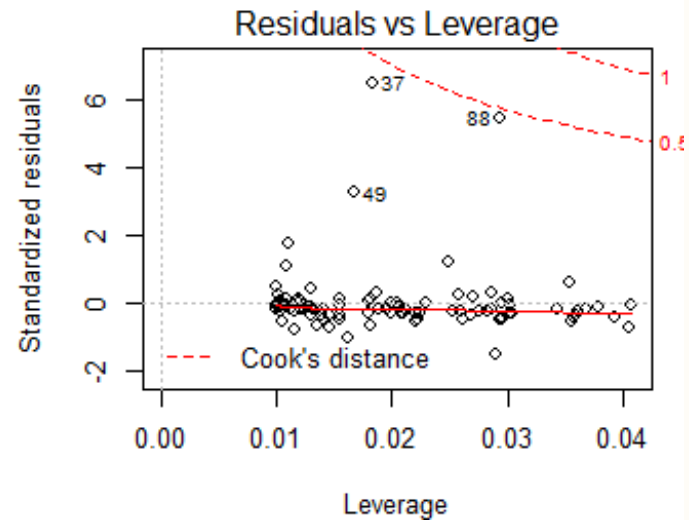
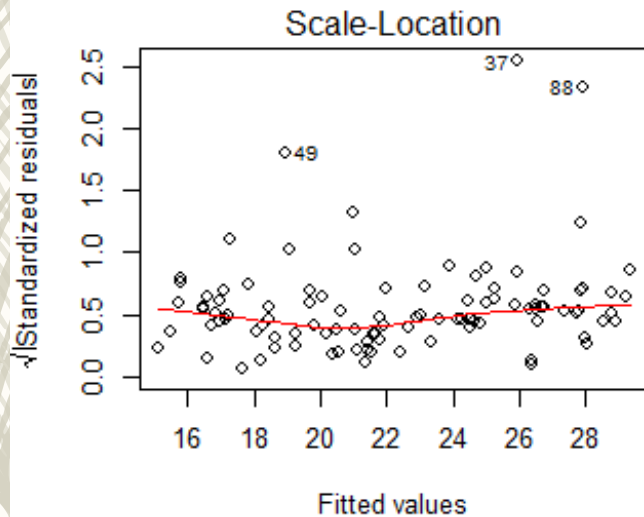
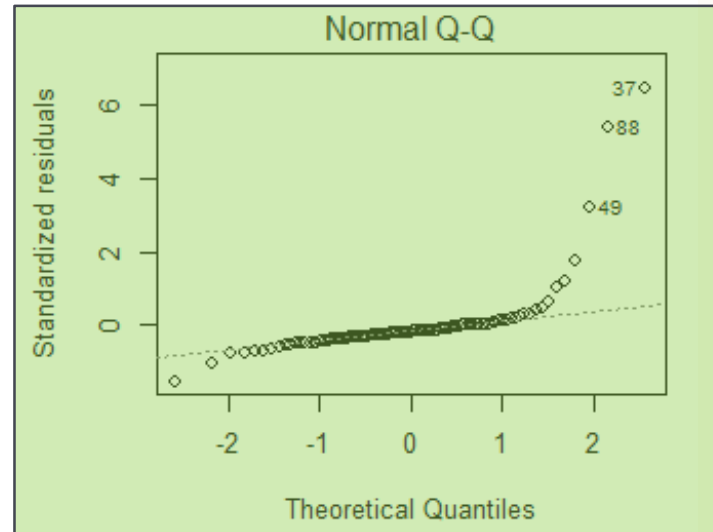
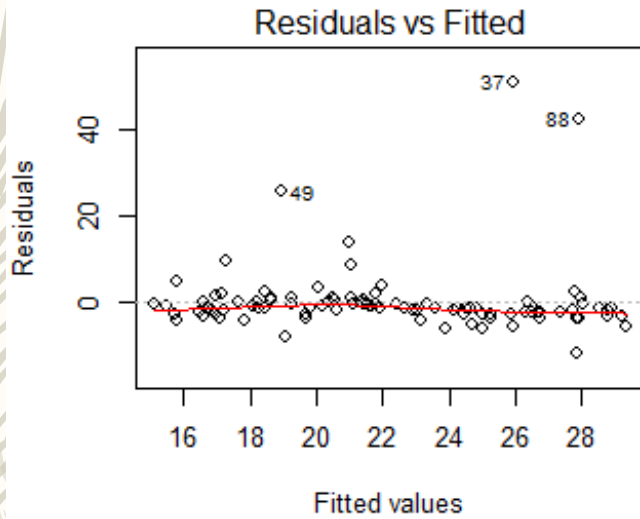
Análise de resíduos: normalidade, homocedasticidade e independência



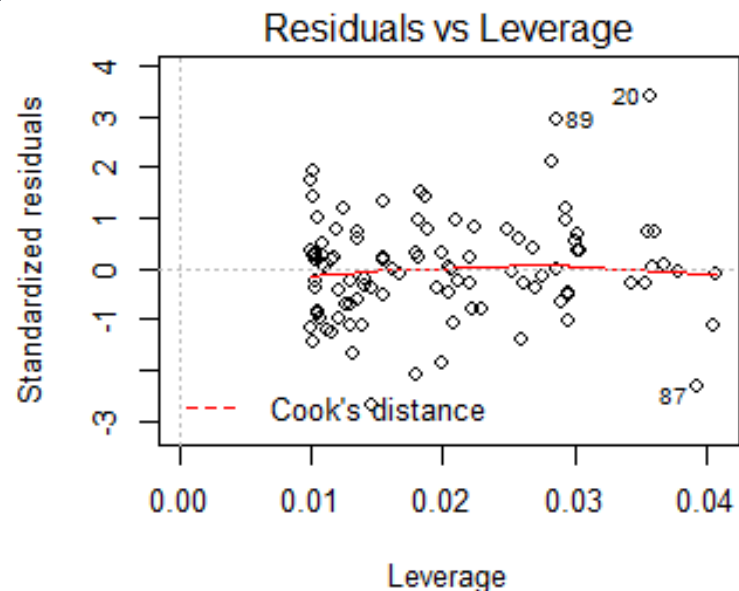
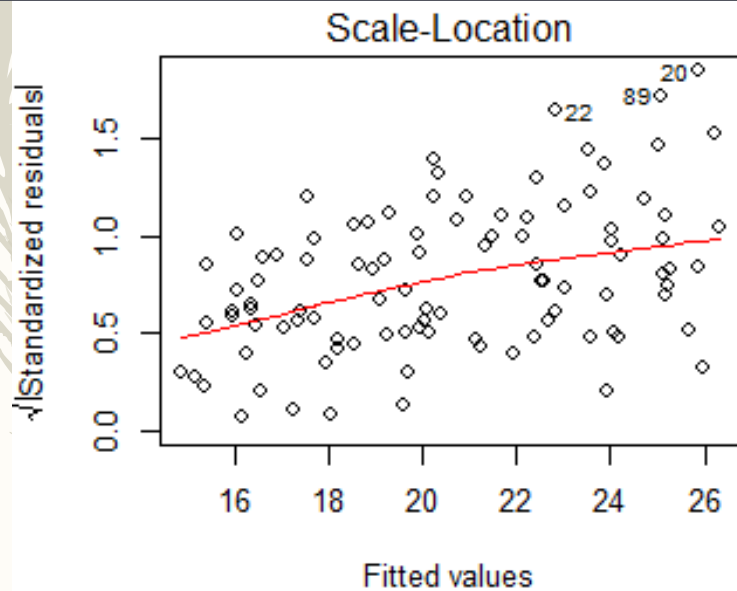
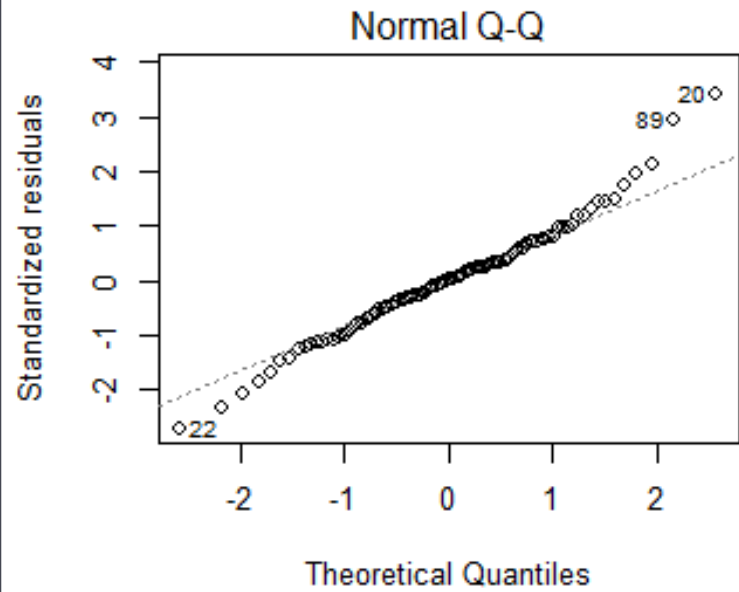
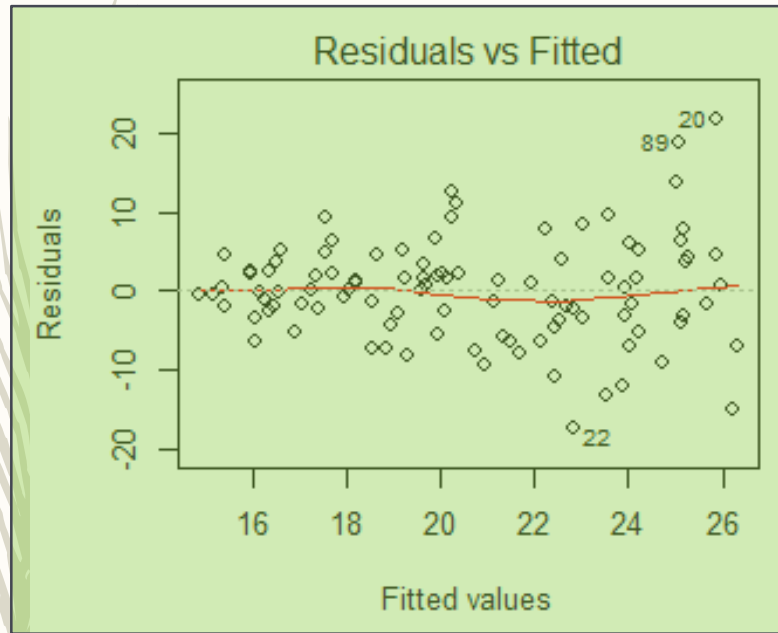
Non-Gaussian Errors



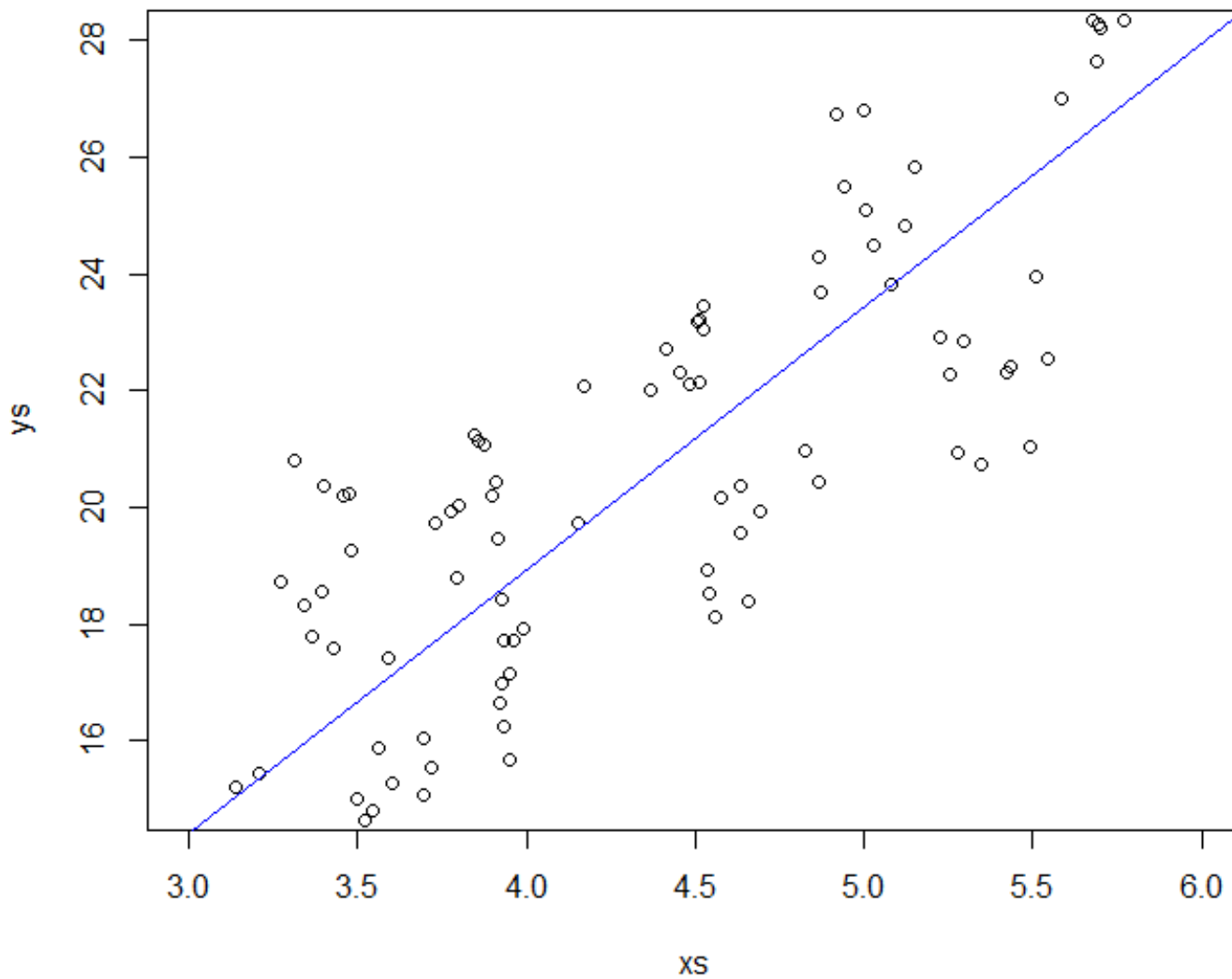
Non-Gaussian Errors



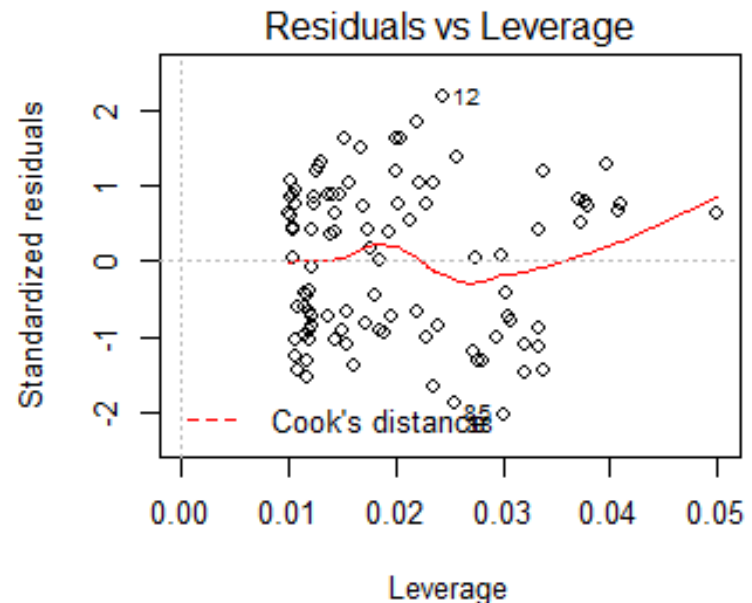
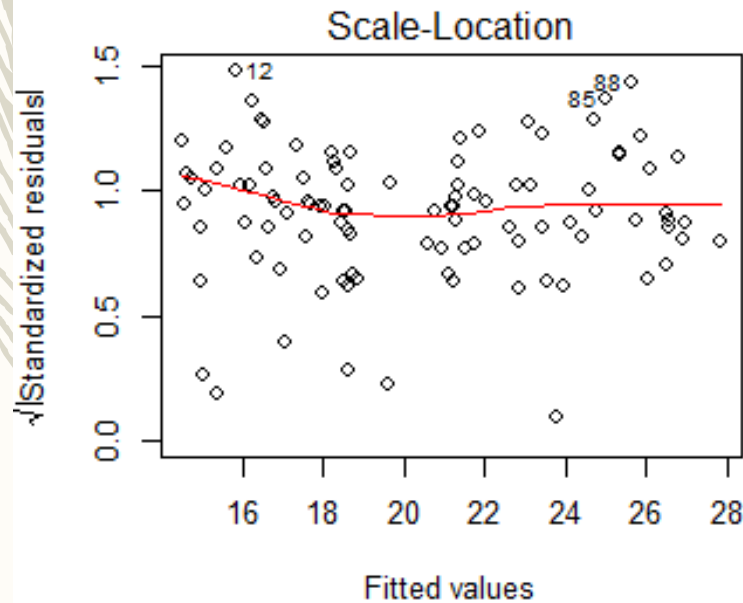
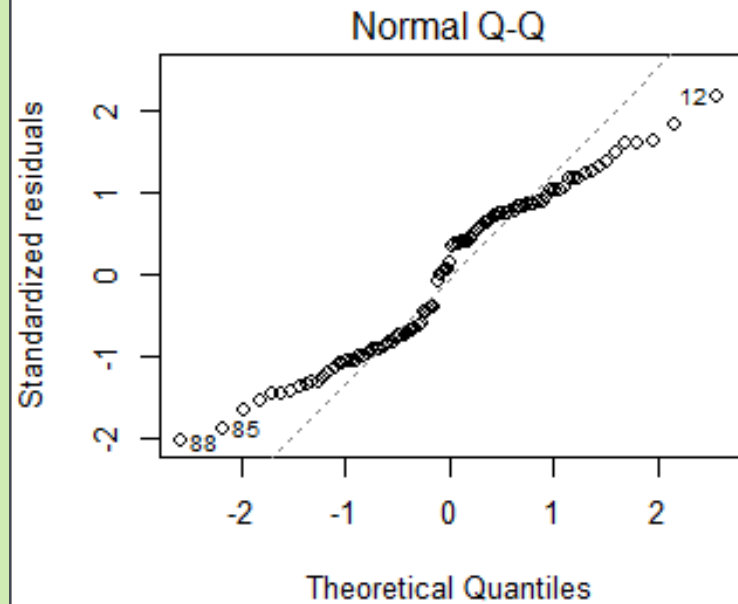
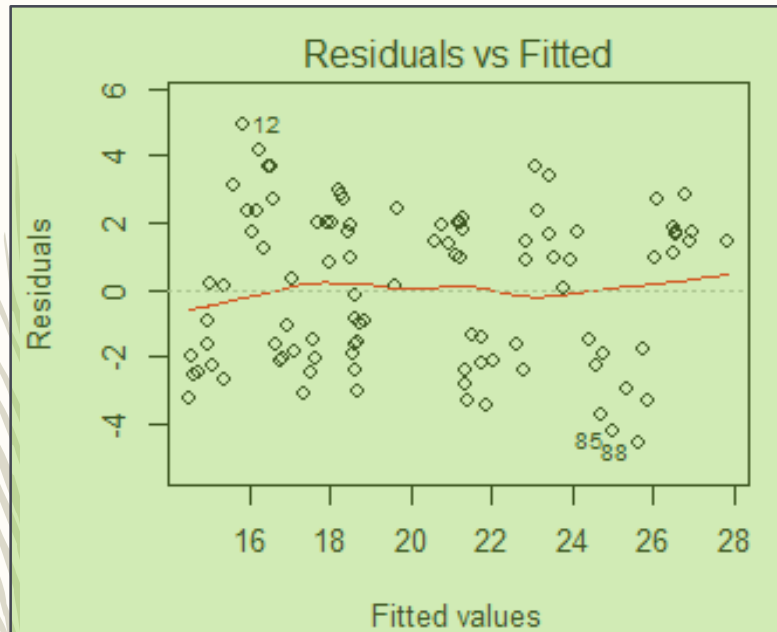
Heterocedastic errors



Non-independent errors



Non-independent errors





Análise de resíduos

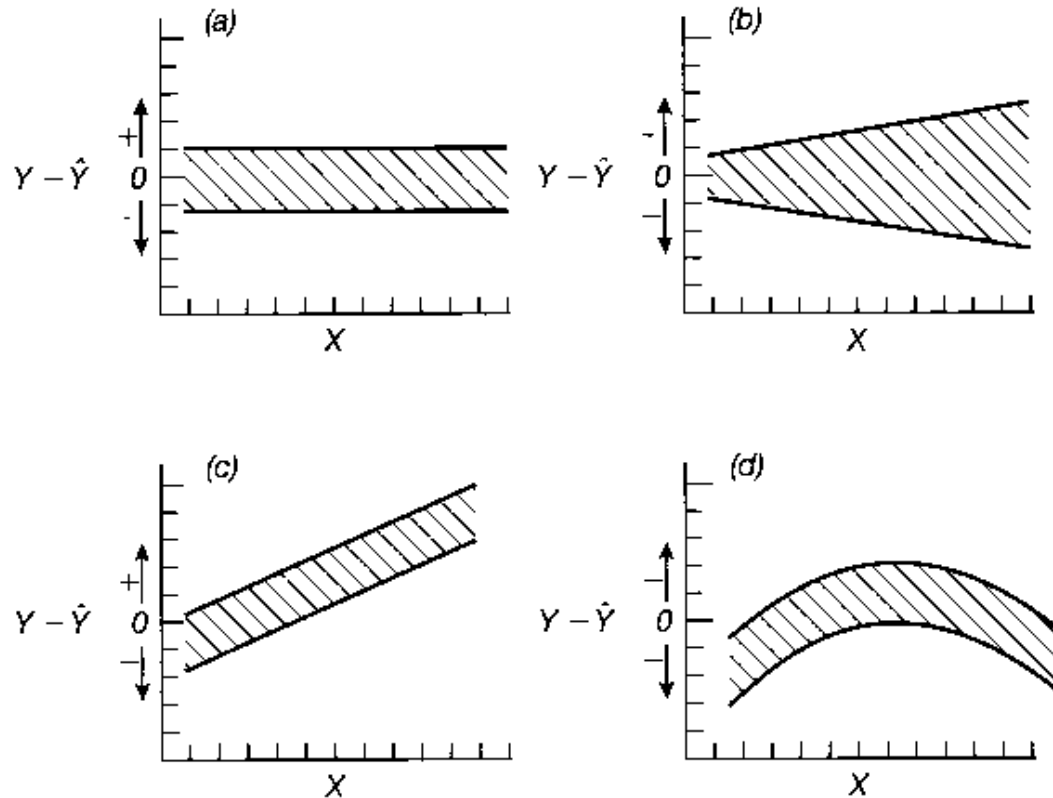
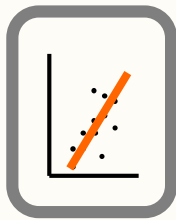


Figure 16.12 The plotting of residuals. (a) Data exhibiting homoscedasticity. (b) Data with heteroscedasticity of the sort in Example 16.9. (c) Data for which there was likely an error in the regression calculations, or an additional variable is needed in the regression model. (d) Data for which a linear regression does not accurately describe the relationship between Y and X , and a curvilinear relationship should be considered.

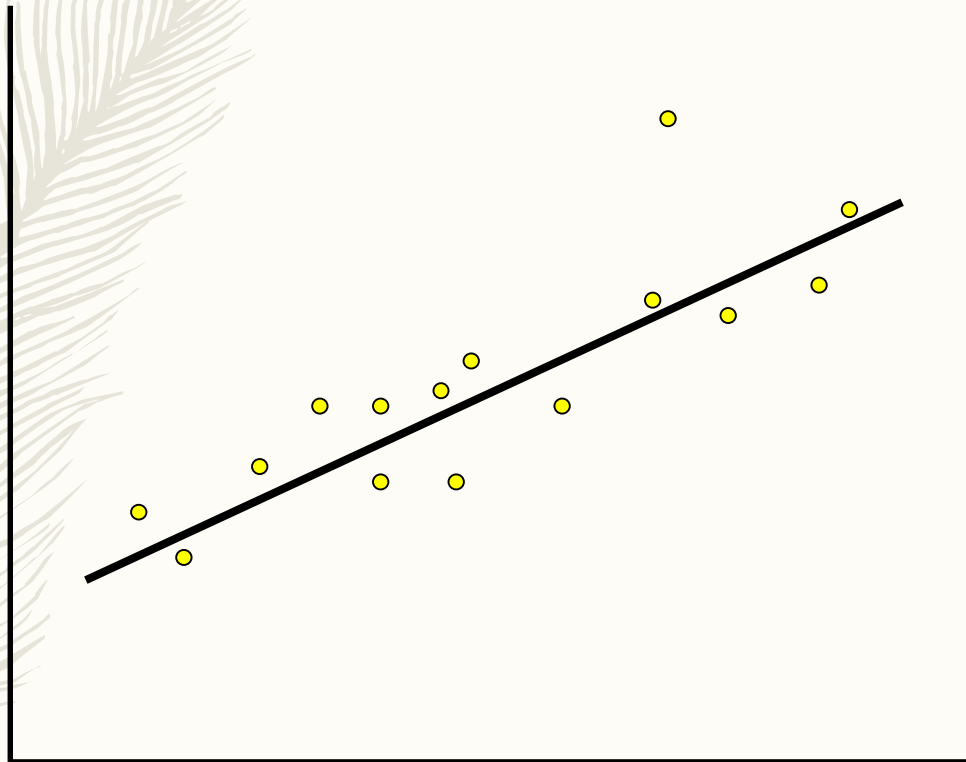


Análise de resíduos: outliers e observações influentes

- Outliers (análise exploratória e análise de resíduos)
- Observações influentes (por exemplo através da distância de Cook)

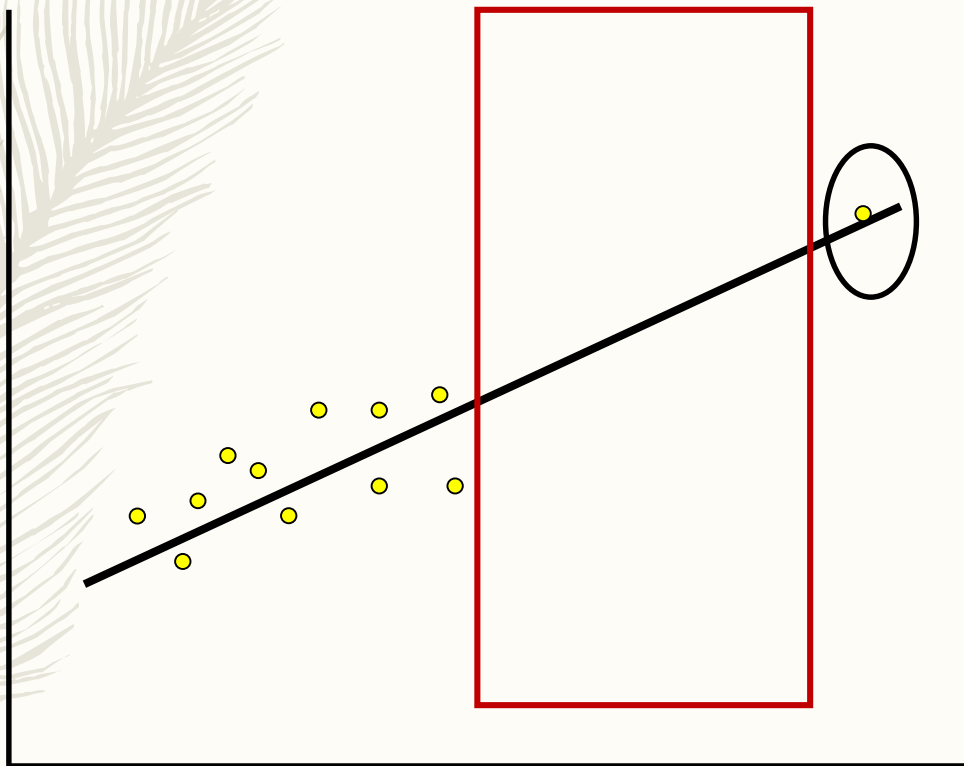


Análise de resíduos: outliers e observações influentes

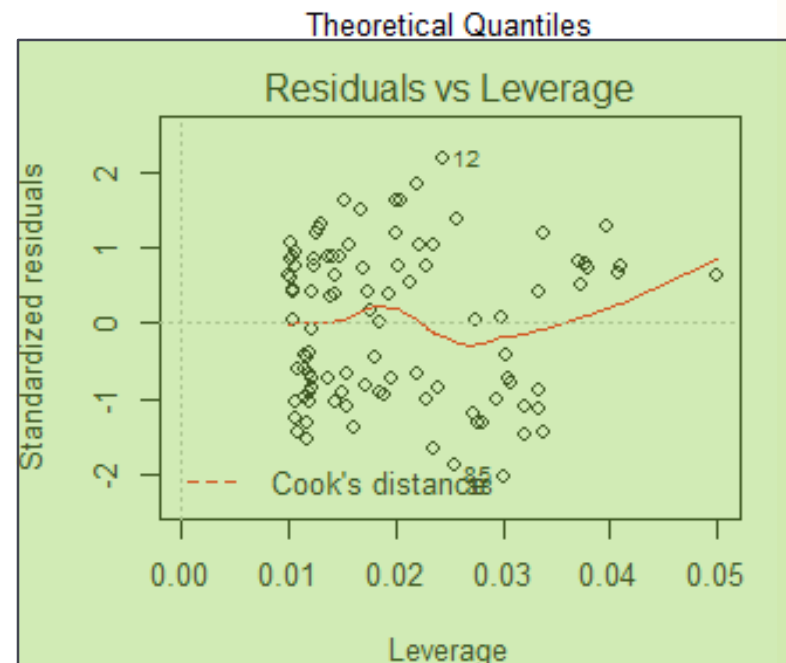
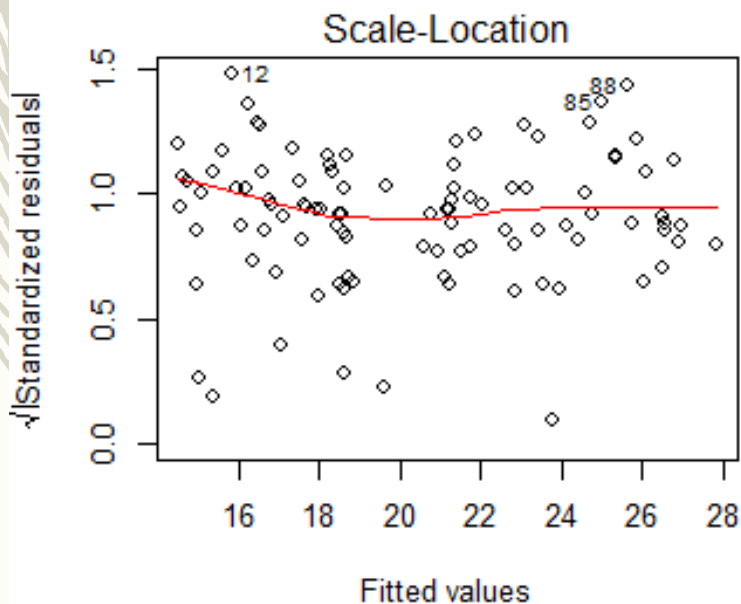
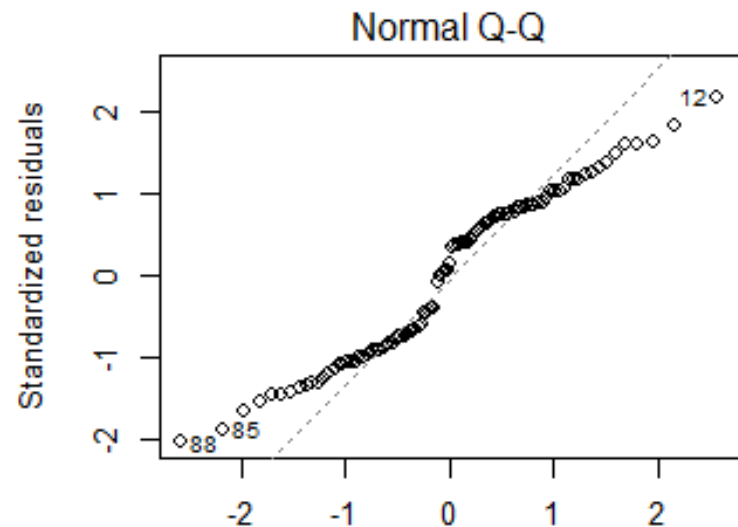
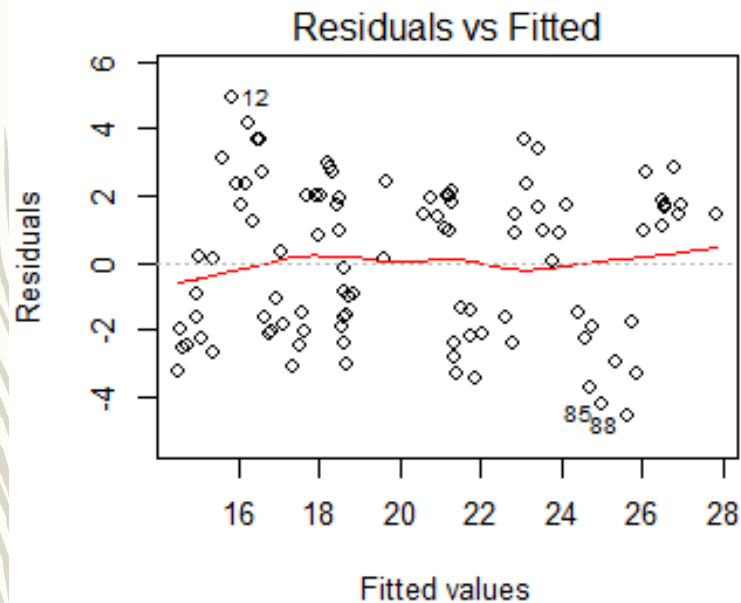




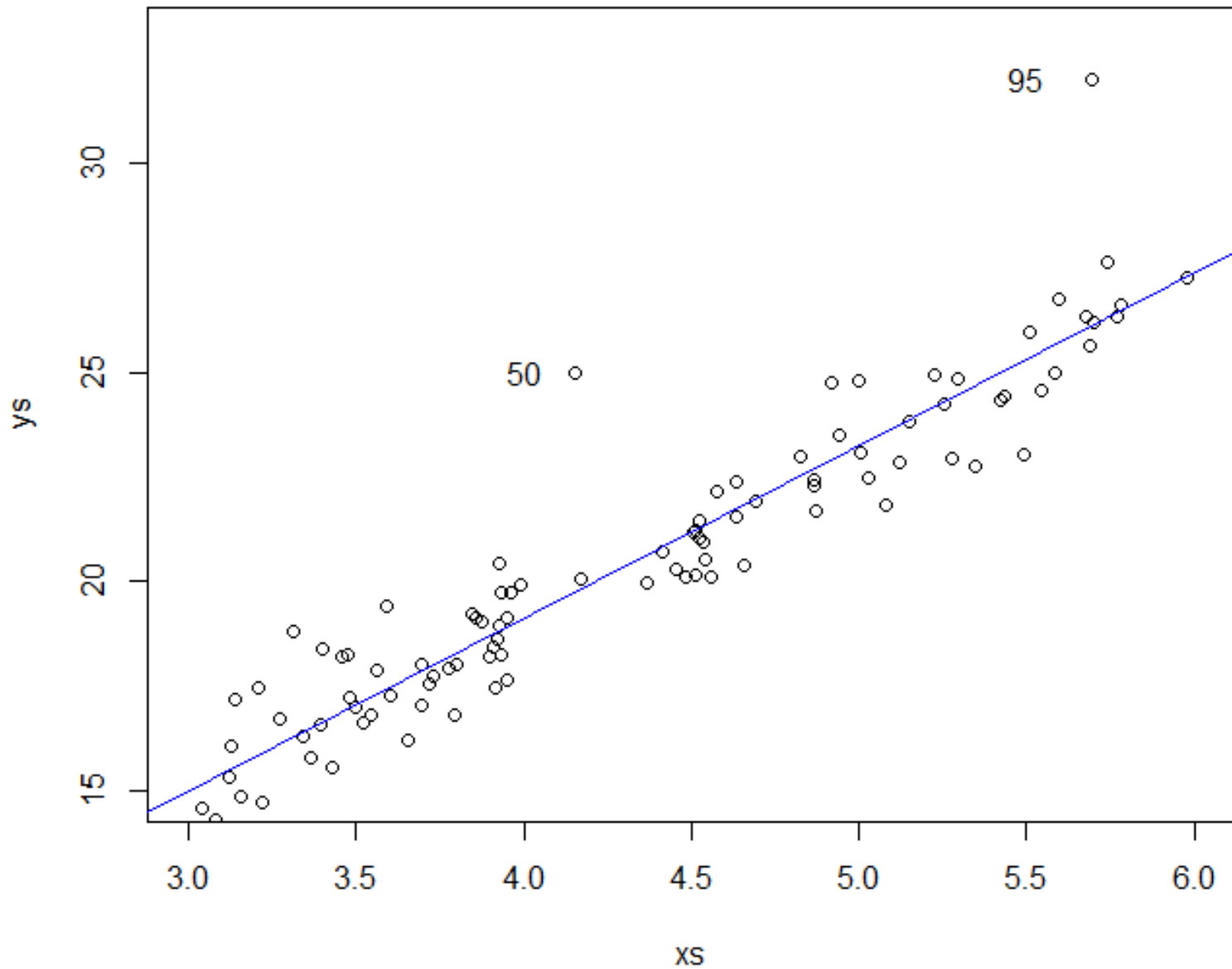
Análise de resíduos: outliers e observações influentes



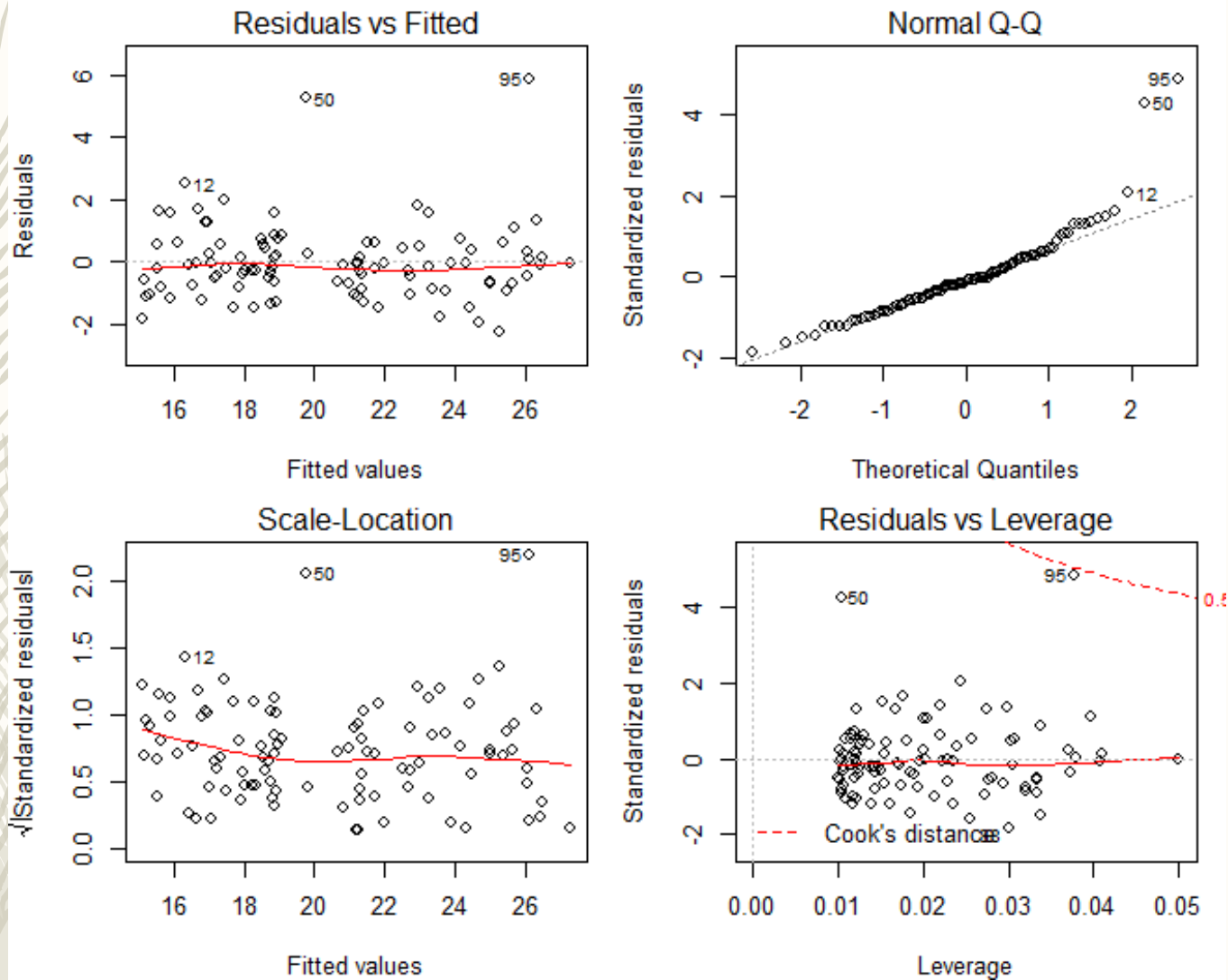
Influential observations

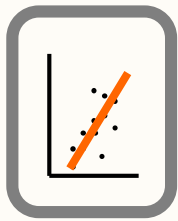


Influential observations



Influential observations





Seleccção de modelos de regressão

Quando pretendemos seleccionar um sub-conjunto de variáveis (as mais importantes) e obter um modelo simplificado.

Os coeficientes de regressão do modelo simplificado são diferentes dos do modelo original (modelo saturado).

- Forward selection
- Backward elimination
- Stepwise selection
- Test all combinations