BIE5782

Unidade 7:

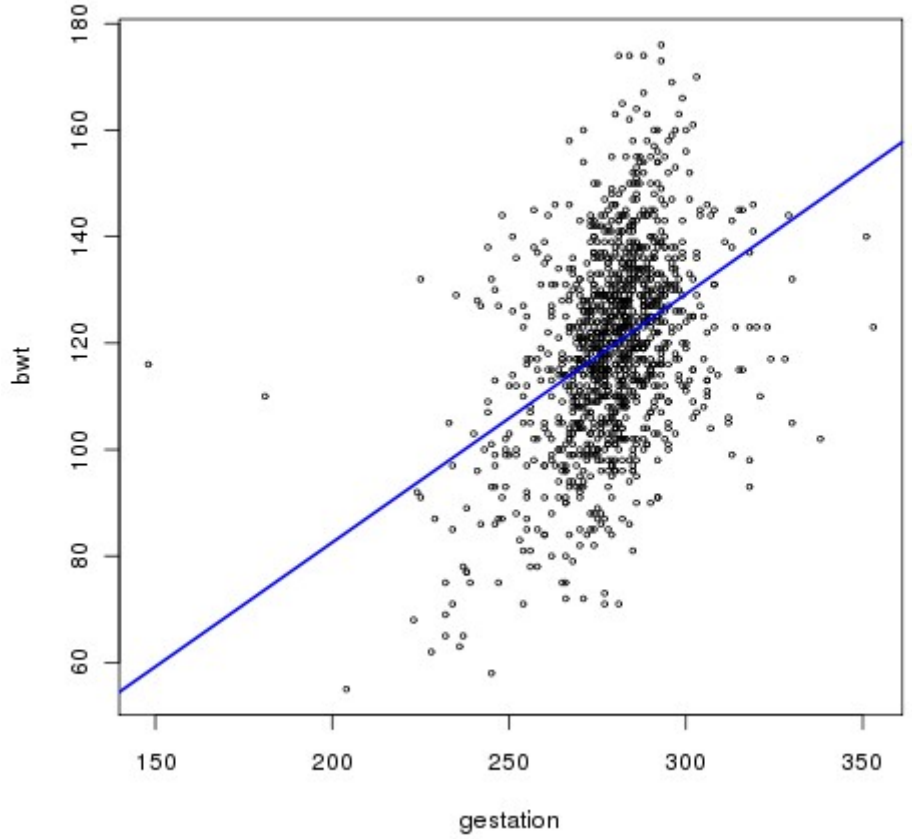
INTRODUÇÃO AOS MODELOS LINEARES

ROTEIRO

1. Motivação
2. Método dos mínimos quadrados
3. Ajuste no R: função lm
4. Resultado no R: objeto lm
5. Premissas, interpretação e diagnóstico
6. Duas variáveis: efeitos aditivos e interação

lm()

Ajusta Modelo Linear Gaussiano



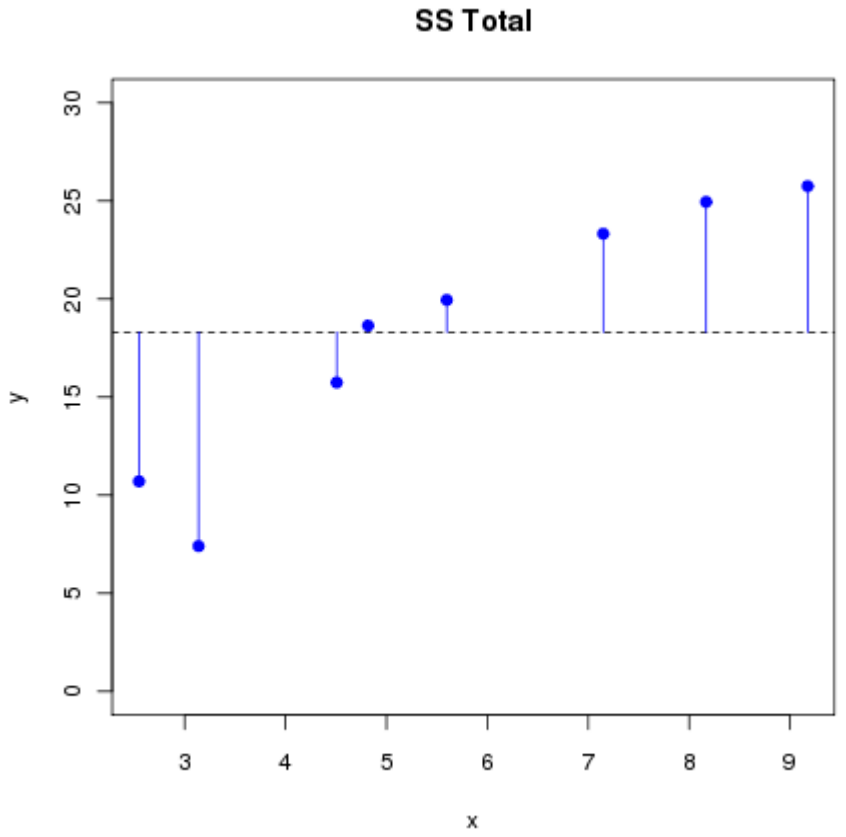

```
> plot(bwt~gestation, data=babies, cex=0.5)
> babies.m1 <- lm(bwt~gestation, data=babies)
> abline(babies.m1, col="blue", lwd=2)
```

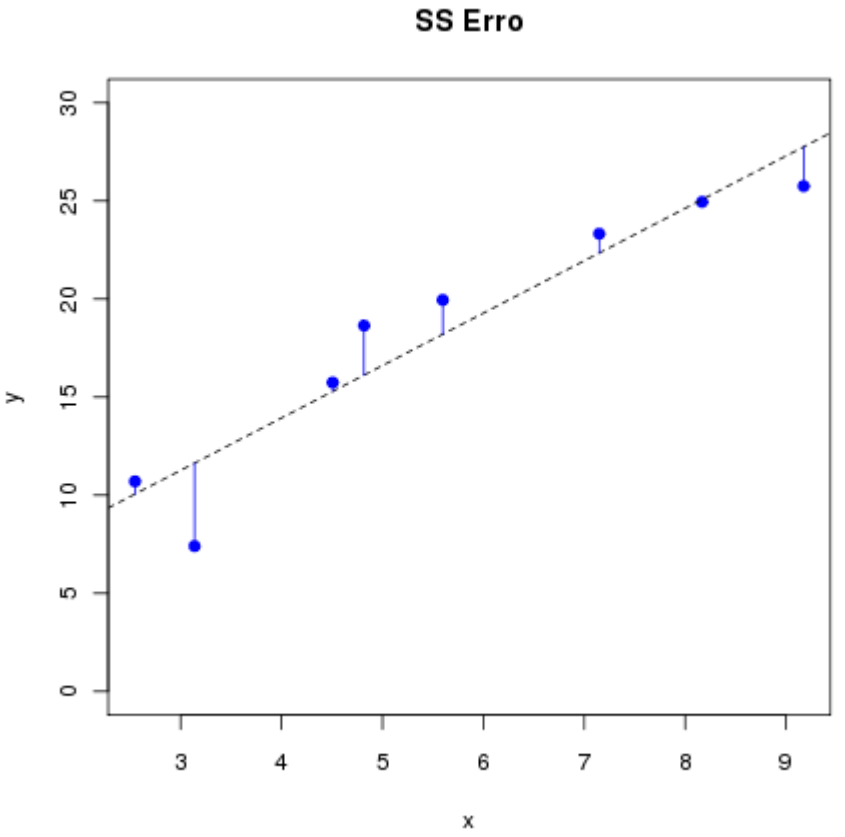
anova.lm()

Avalia o Modelo

```
> anova(babies.m1)
Analysis of Variance Table

Response: bwt
             Df Sum Sq Mean Sq F value    Pr(>F)
gestation     1  65450   65450  233.43 < 2.2e-16 ***
Residuals  1172 328608     280
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

Somas dos (Desvios) Quadrados





Classe `lm`

```
> names(babies.m1)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"

> babies.m1$coefficients
(Intercept)   gestation
-10.7541389   0.4665569

> babies.m1$residuals[1:4]
         1          2          3          5
 -1.748014  -7.814900   8.584770 -12.814900

> babies.m1$fitted.values[1:4]
       1        2        3        5
121.7480 120.8149 119.4152 120.8149

> babies.m1$call
lm(formula = bwt ~ gestation, data = babies)
```

Objetos da classe `lm` são listas com todos os objetos resultantes do ajuste de um modelo linear Gaussiano.

`coef()`, `confint()`, `residuals()`, `fitted()`, `logLik()`, `AIC()` …

Funções de Extração

```
> coef(babies.m1)
(Intercept)   gestation
-10.7541389   0.4665569

> confint(babies.m1)
                    2.5 %    97.5 %
(Intercept) -27.5035066 5.9952288
gestation     0.4066435 0.5264702
```

`coef()`, `confint()`, `residuals()`, `fitted()`, `logLik()`, `AIC()` …

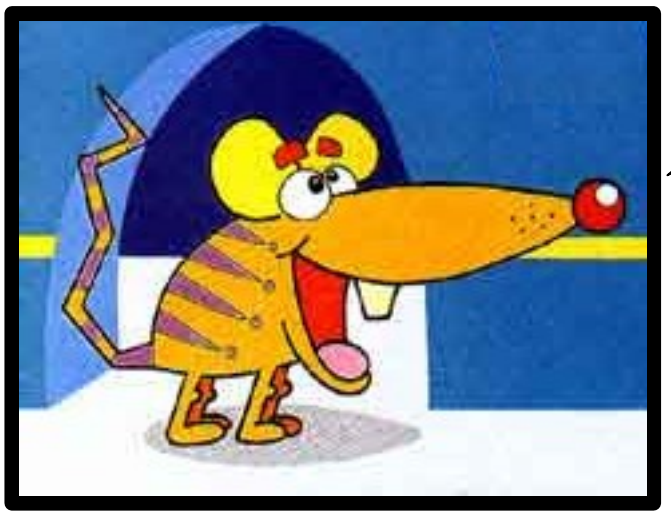
Funções de Extração

```
> residuals(babies.m1)[1:4]
         1          2          3          5
 -1.748014  -7.814900   8.584770 -12.814900

> fitted(babies.m1)[1:4]
       1        2        3        5
121.7480 120.8149 119.4152 120.8149

> logLik(babies.m1) ## pacote MASS
'log Lik.' -4973.256 (df=3)

> AIC(babies.m1)
[1] 9952.512
```

summary.lm()

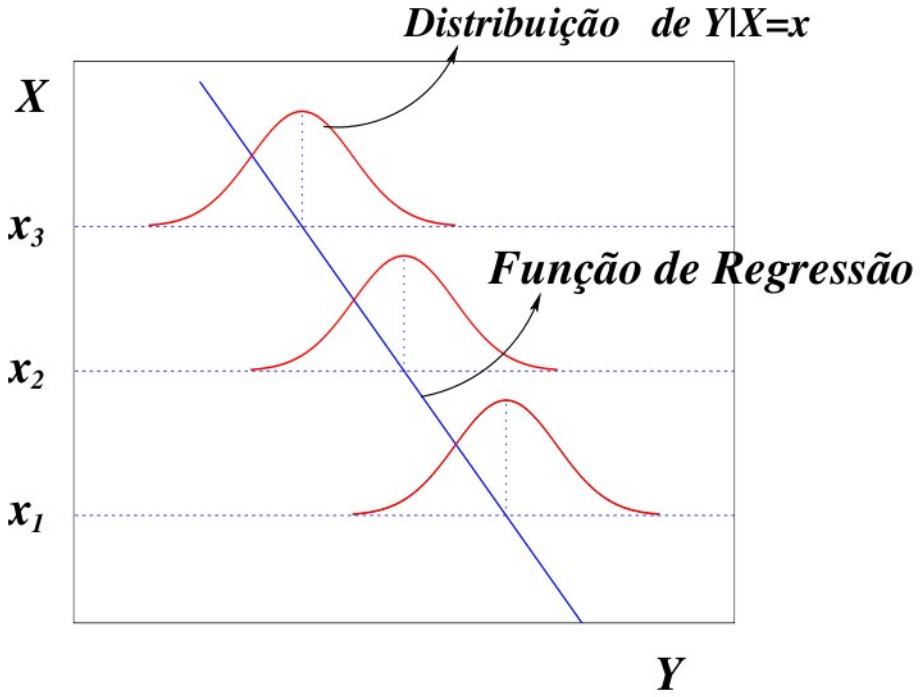
Resumo do Modelo

```
> summary(babies.m1)

Call:
lm(formula = bwt ~ gestation, data = babies)

Residuals:
     Min       1Q   Median       3Q      Max
-49.3483 -11.0653   0.2177  10.1015  57.7037

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.75414    8.53693   -1.26    0.208
gestation     0.46656    0.03054   15.28   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 16.74 on 1172 degrees of freedom
Multiple R-squared: 0.1661,  Adjusted R-squared: 0.1654
F-statistic: 233.4 on 1 and 1172 DF,  p-value: < 2.2e-16
```



Premissas do Modelo de Regressão Linear Gaussiana

João L.F. Batista

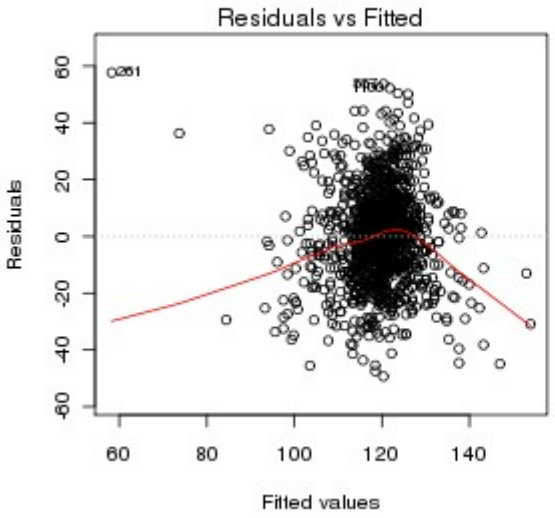


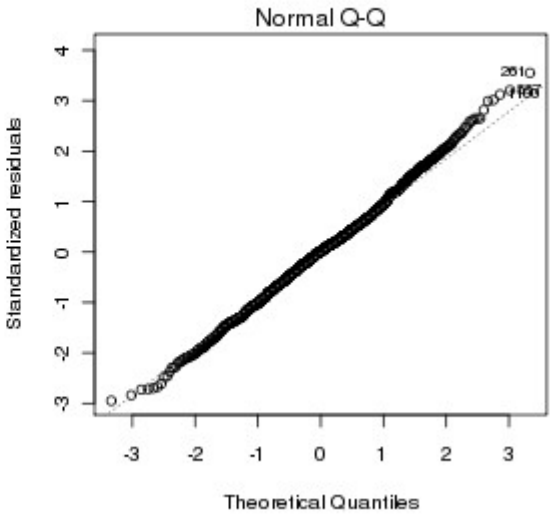
A variável resposta é uma variável normal (Gaussiana) sendo que:

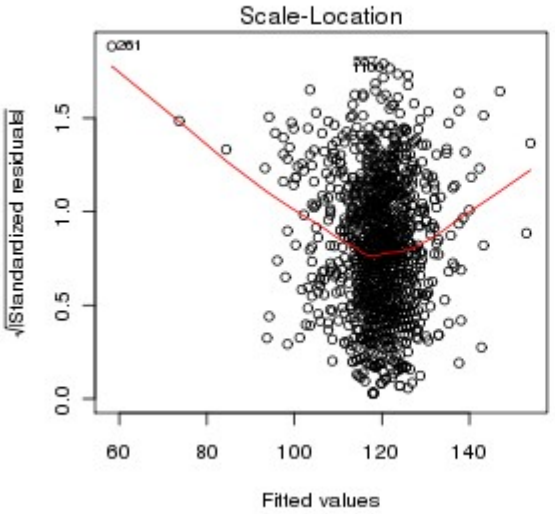
- Sua média é uma função linear das variáveis preditoras;
- Seu desvio-padrão é constante;
- LOGO: resíduos com média zero e variância constante

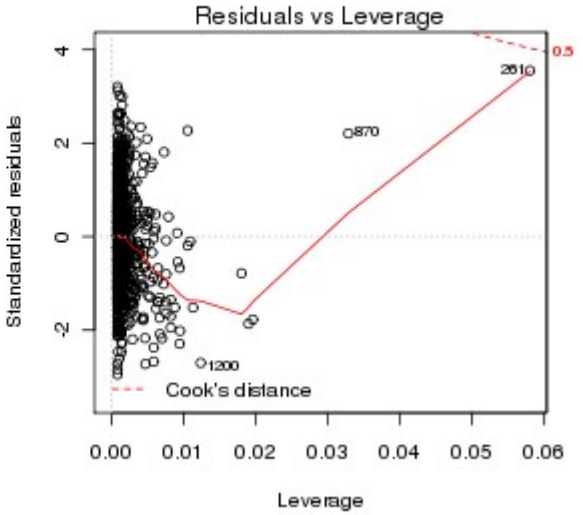
plot.lm()

Gráficos de Diagnóstico



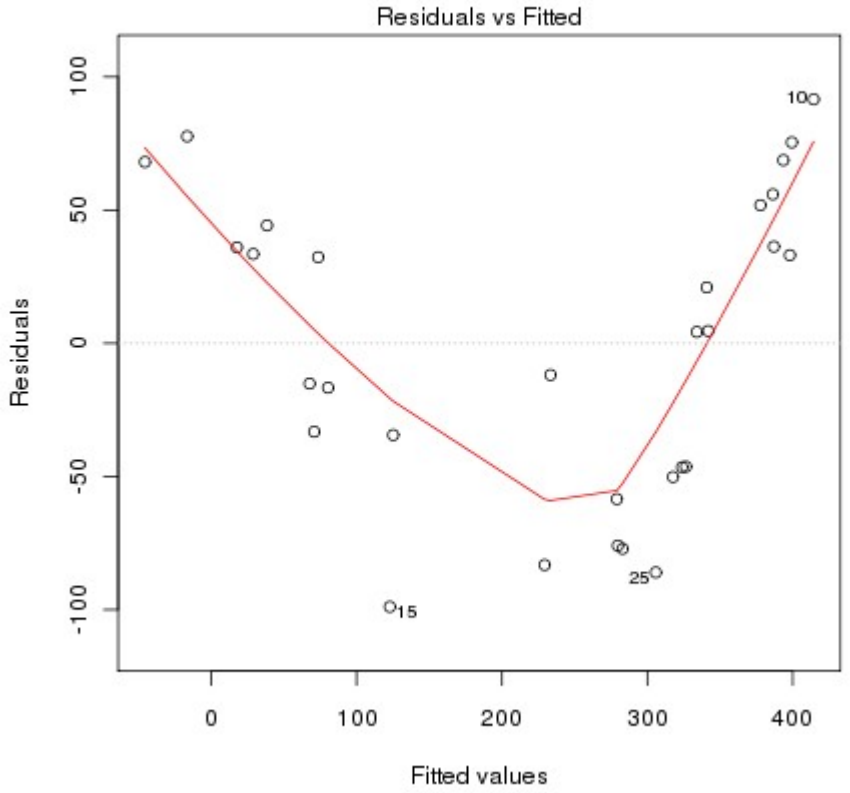


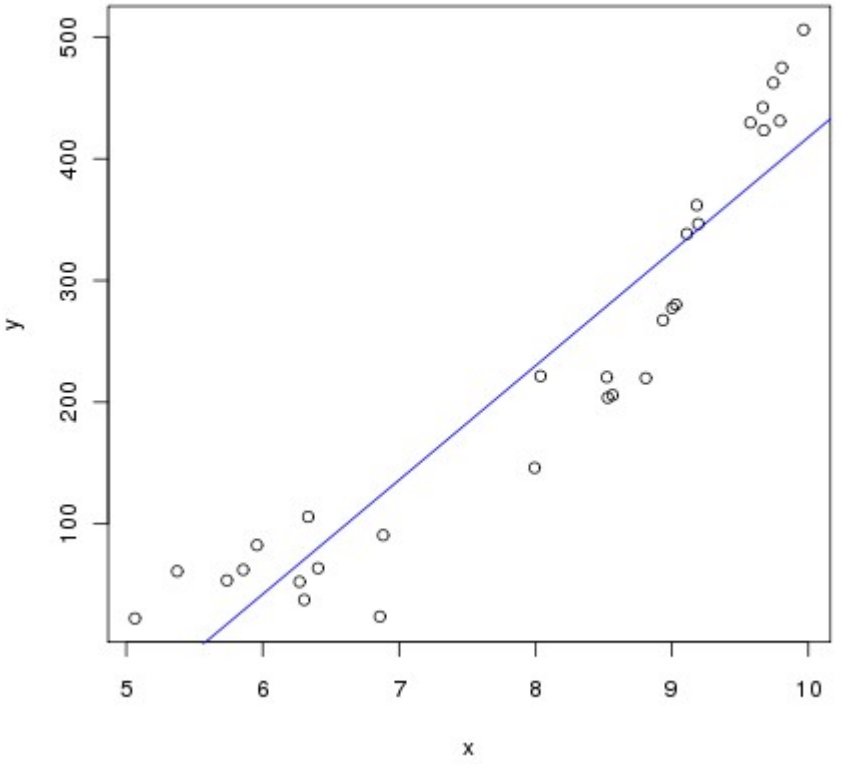





```
> par(mfrow=c(2,2))
> plot(babies.m1)
> par(mfrow=c(1,1))
```

Resíduos x Estimado

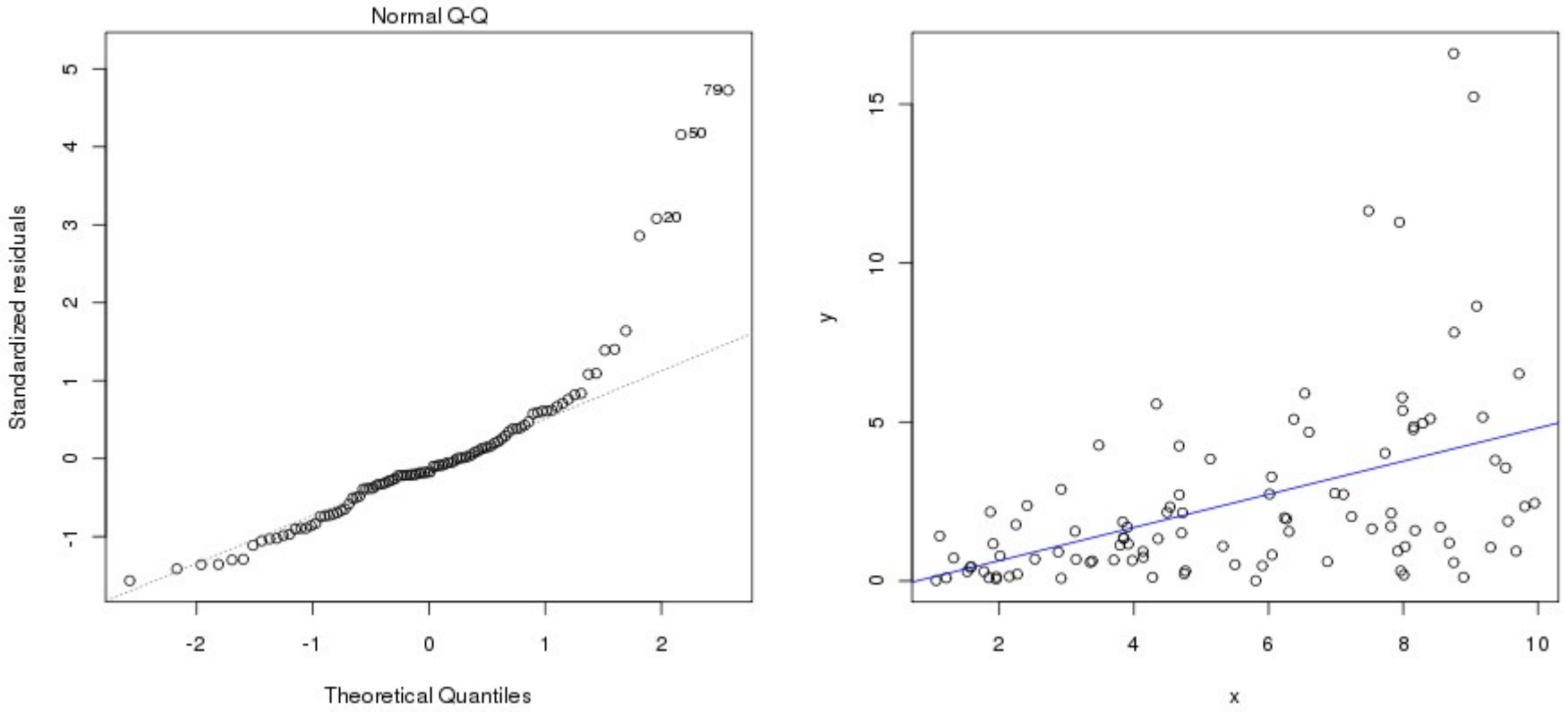




<u>Detecta:</u>

- Tendências não-lineares
- Variâncias não homogêneas

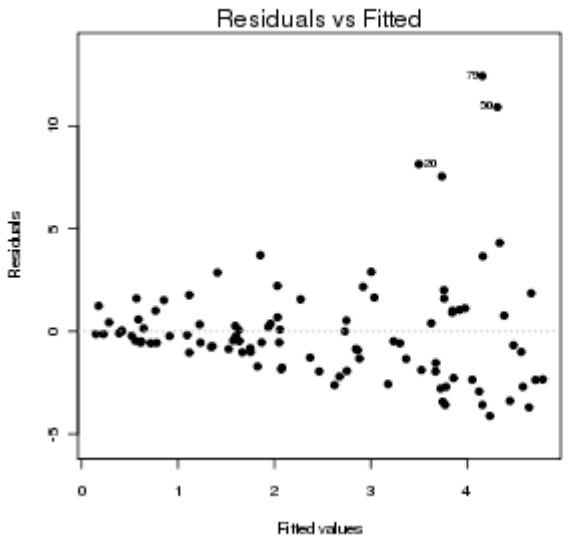
Gráfico de Quantis Resíduos x Normal

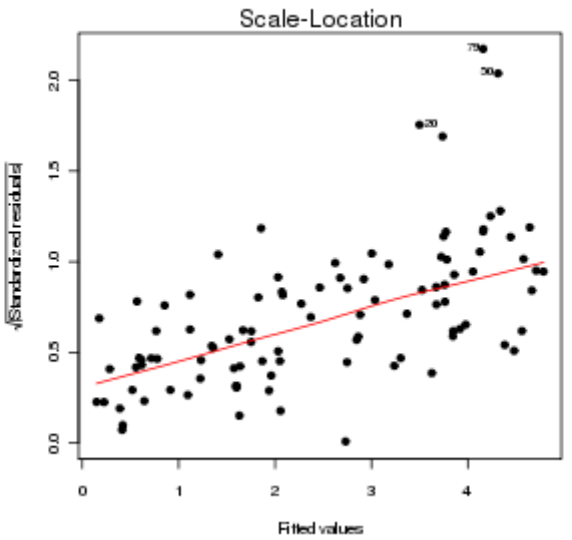


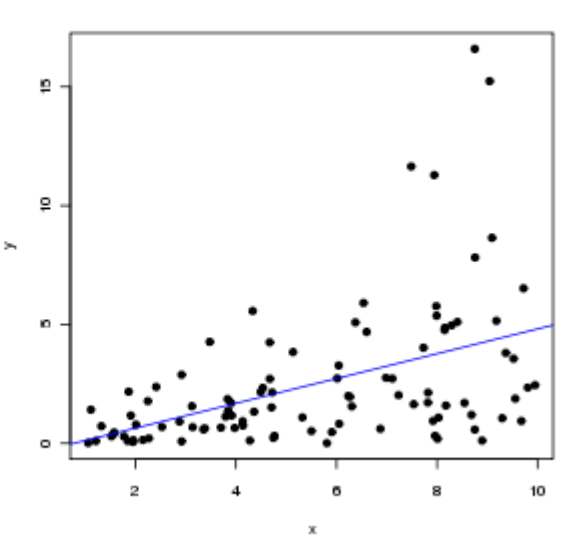
<u>Detecta:</u>

- Desvios da normalidade nos resíduos

Resíduos x Estimado
Raiz dos Resíduos Padronizados x Estimado



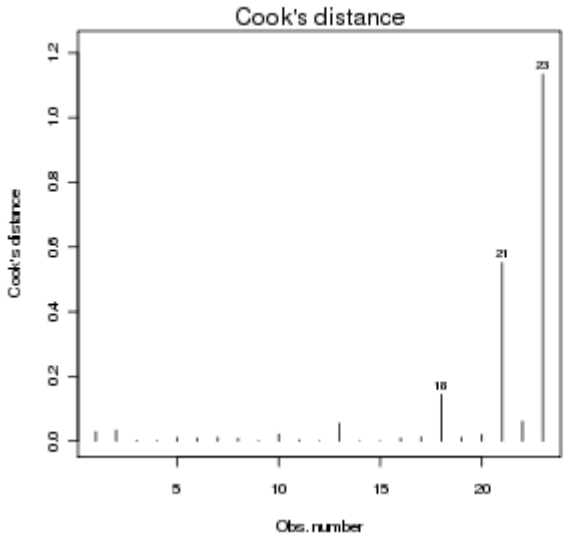


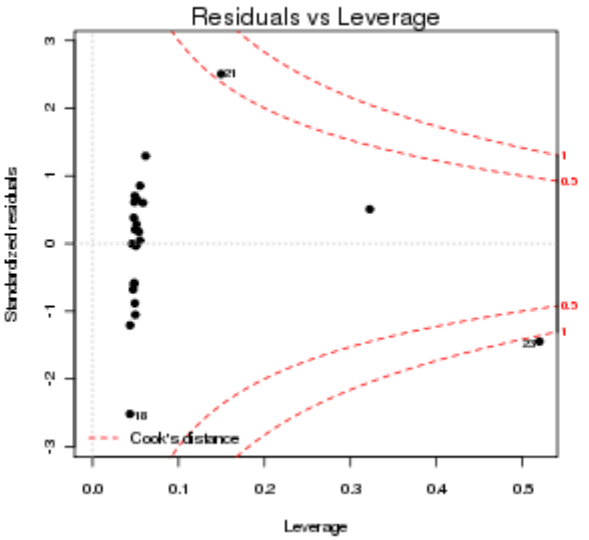


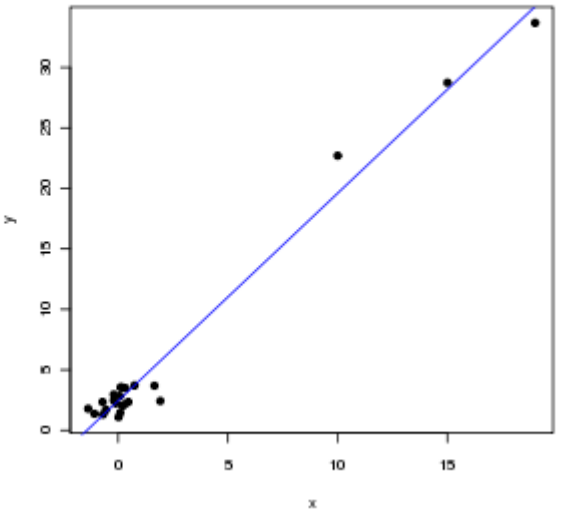
<u>Detectam:</u>

- Mudanças na variância (heteroscedasticidade);
- Valores extremos não esperados (*outliers*).

Influência e Alavancagem

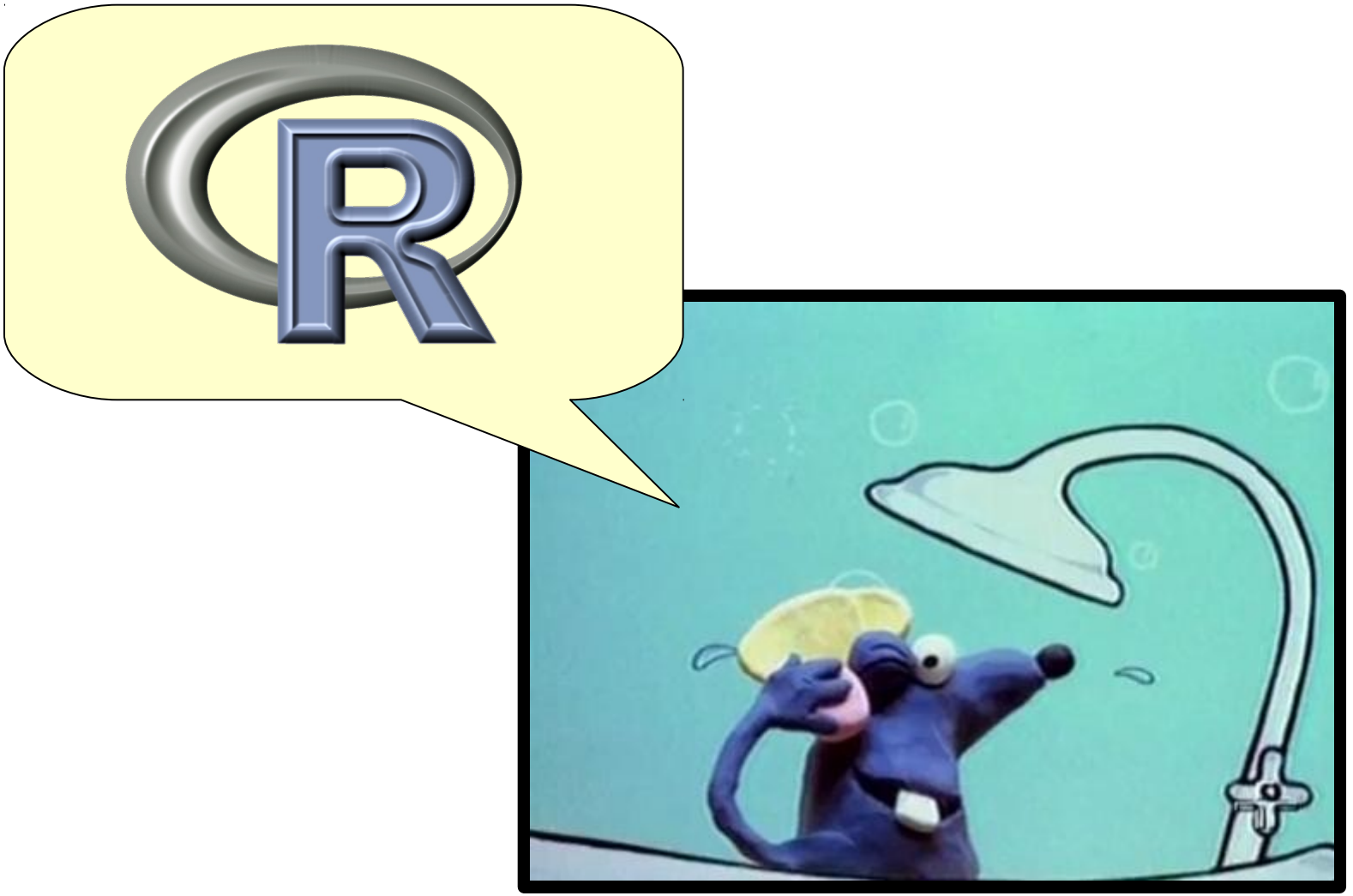




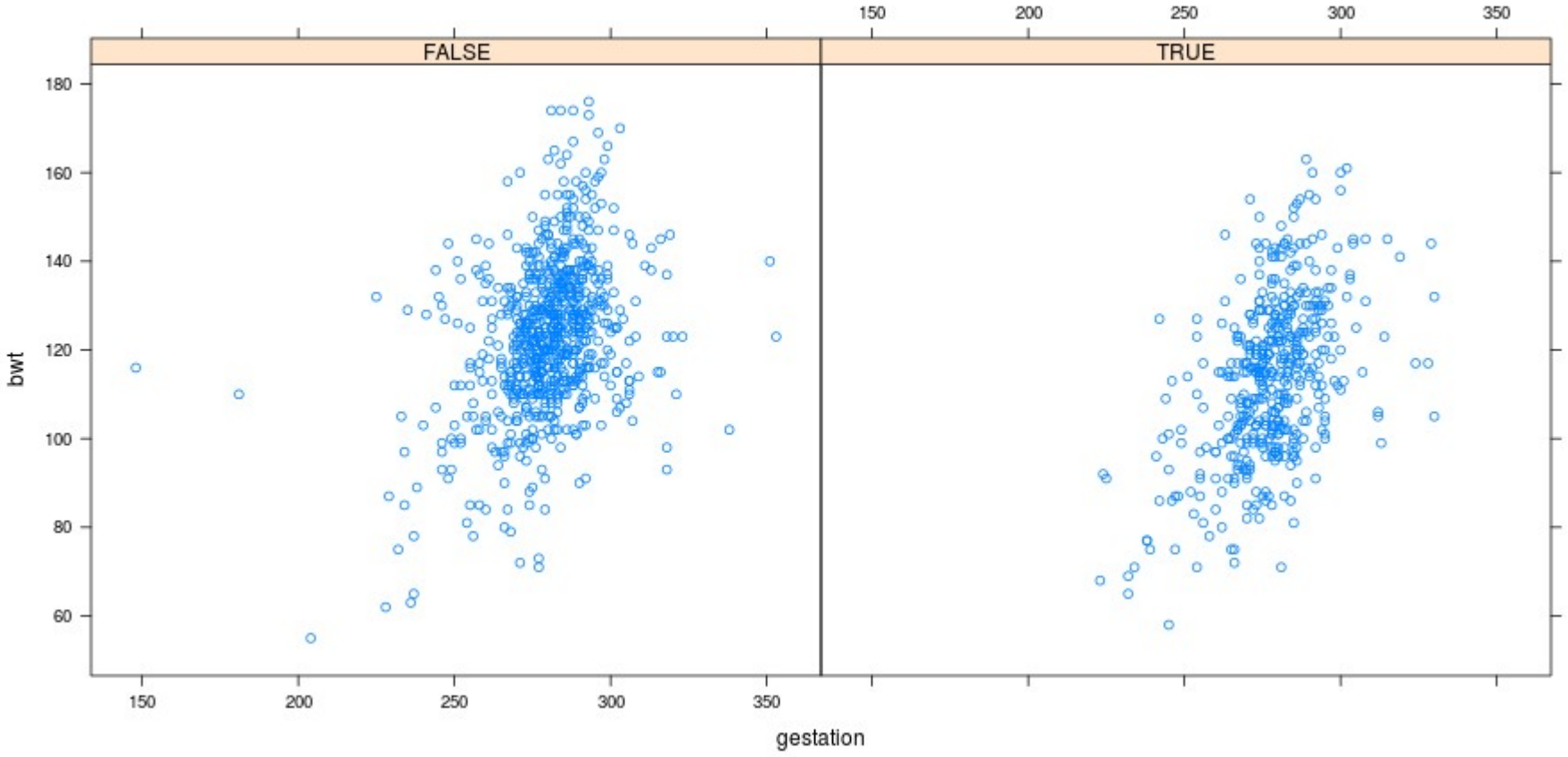


<u>Detecta:</u>

- Pontos influentes



Adicionando mais uma variável




```
xyplot( bwt~gestation|smoke, data=babies )
```

lm ☞ update ☞ anova

O Ciclo de Ajuste e Avaliação

```
> babies.m1 <- lm(bwt~gestation,data=babies)
> babies.m2 <- update(babies.m1, .~.+smoke,data=babies)

> anova(babies.m2, babies.m1)
Analysis of Variance Table

Model 1: bwt ~ gestation + smoke
Model 2: bwt ~ gestation
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1171 309075
2   1172 328608 -1    -19533 74.007 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
```

```
> summary(babies.m2)

Call:
lm(formula = bwt ~ gestation + smoke, data = babies)

Residuals:
    Min      1Q  Median      3Q     Max
-50.789 -11.035  -0.211  10.053  52.412

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.18492    8.32945  -0.382    0.702
gestation    0.45117    0.02968  15.200   <2e-16 ***
smokeTRUE   -8.37440    0.97346  -8.603   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Residual standard error: 16.25 on 1171 degrees of
freedom
Multiple R-squared: 0.2157,  Adjusted R-squared: 0.2143
F-statistic:   161 on 2 and 1171 DF,  p-value: < 2.2e-16
```

model.matrix()

A Matriz do Modelo

```
> names(babies.m2)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "contrasts"     "xlevels"       "call"          "terms"
[13] "model"
```

```
> formula(babies.m2)
bwt ~ gestation + smoke

> model.matrix(babies.m2)
     (Intercept) gestation smokeTRUE
1              1       284         0
2              1       282         0
3              1       279         1
5              1       282         1
...
```

A MATRIZ

(do modelo)

(Intercept)	gestation	smokeTRUE
1	284	0
1	282	0
1	279	1
1	282	1
1	286	0
1	244	0
1	245	0
1	289	0
1	299	1
1	351	0

Cálculo Matricial dos Esperados

y = a.X

$$\begin{bmatrix} 1 & 284 & 0 \\ 1 & 282 & 0 \\ 1 & 279 & 1 \end{bmatrix} \text{ x } \begin{bmatrix} 3{,}19 \\ 0{,}45 \\ -8{,}37 \end{bmatrix} =$$

$$= \begin{bmatrix} 1\times 3{,}19 + 0{,}45\times 284 - 8{,}37\times 0 \\ 1\times 3{,}19 + 0{,}45\times 282 - 8{,}37\times 0 \\ 1\times 3{,}19 + 0{,}45\times 279 - 8{,}37\times 1 \end{bmatrix}$$

Sugestão de leitura

John Fox (2002). An R and S-Plus Companion to Applied Regression. Sage Publications, Thousand Oaks, CA, USA.