



**BIE5782**

Aula 7:

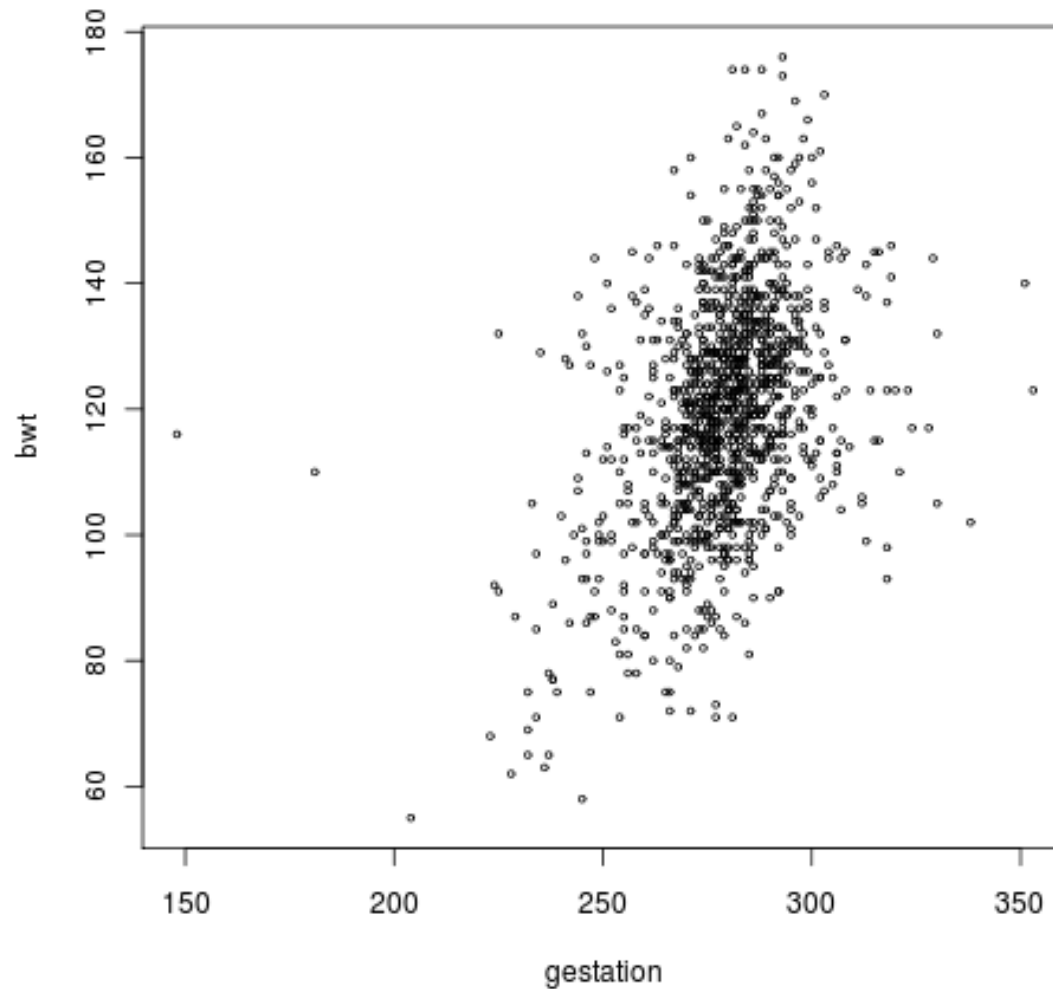
**REGRESSÃO LINEAR SIMPLES**

# OBJETIVOS

- Regressão linear simples:
  - Quando se aplica?
  - Como funciona?
  - Como interpretar seus resultados?
  - Como verificar suas premissas?
  - Como fazer tudo isto no R?

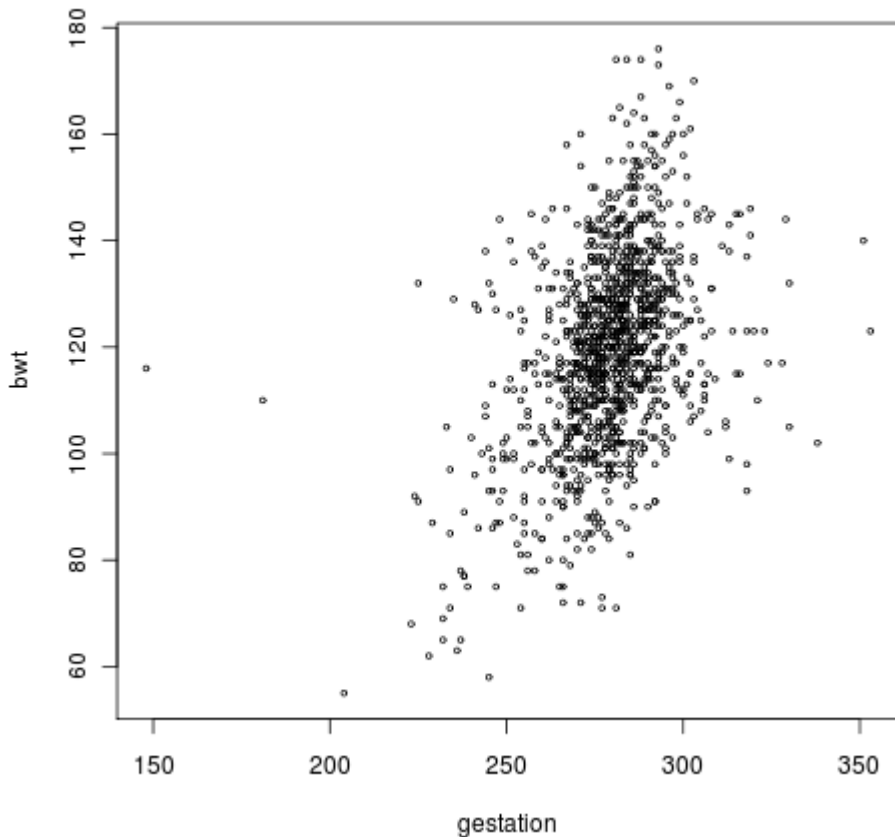
# `plot(y~x, data)`

## Scatterplot com fórmula



```
> plot(bwt~gestation, data=babies, cex=0.5)
```

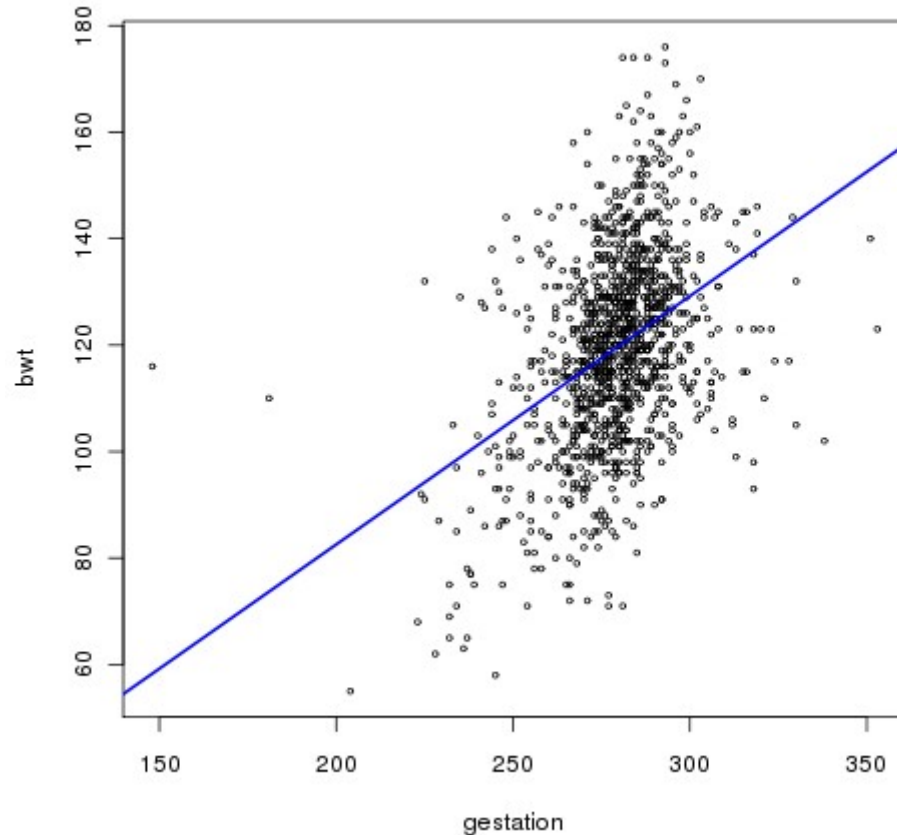
# Regressão linear: um modelo simples



- Há uma resposta à variável preditora?
- Esta resposta é linear?
- Qual a reta que melhor descreve esta resposta?
- O quão boa é esta descrição?

# 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
```

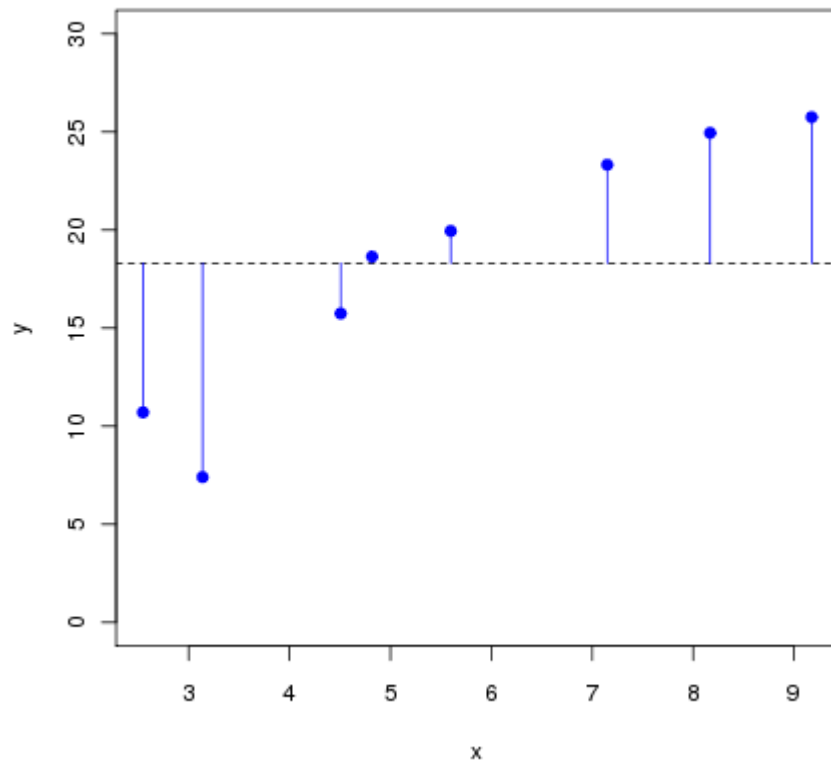
|           | Df   | Sum Sq | Mean Sq | F value | Pr(>F)    |     |
|-----------|------|--------|---------|---------|-----------|-----|
| gestation | 1    | 65450  | 65450   | 233.43  | < 2.2e-16 | *** |
| Residuals | 1172 | 328608 | 280     |         |           |     |

```
---
```

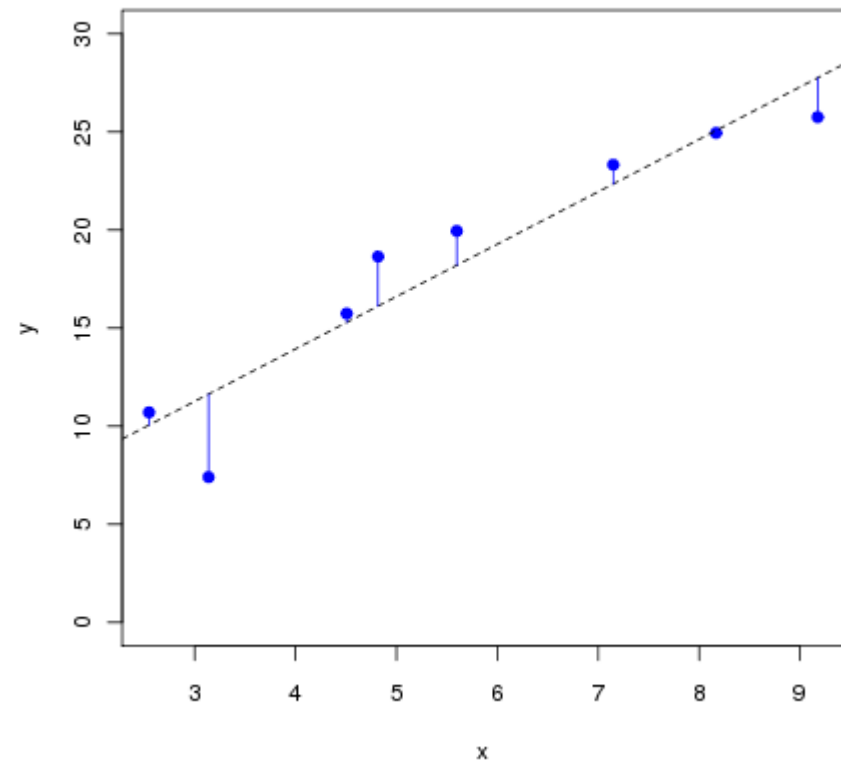
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'  
0.1 ' ' 1
```

# Para Compreender a Tabela da Regressão Somamos dos (Desvios) Quadrados

SS Total



SS Erro





Cálculos da ANOVA para regressão



# 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
> 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

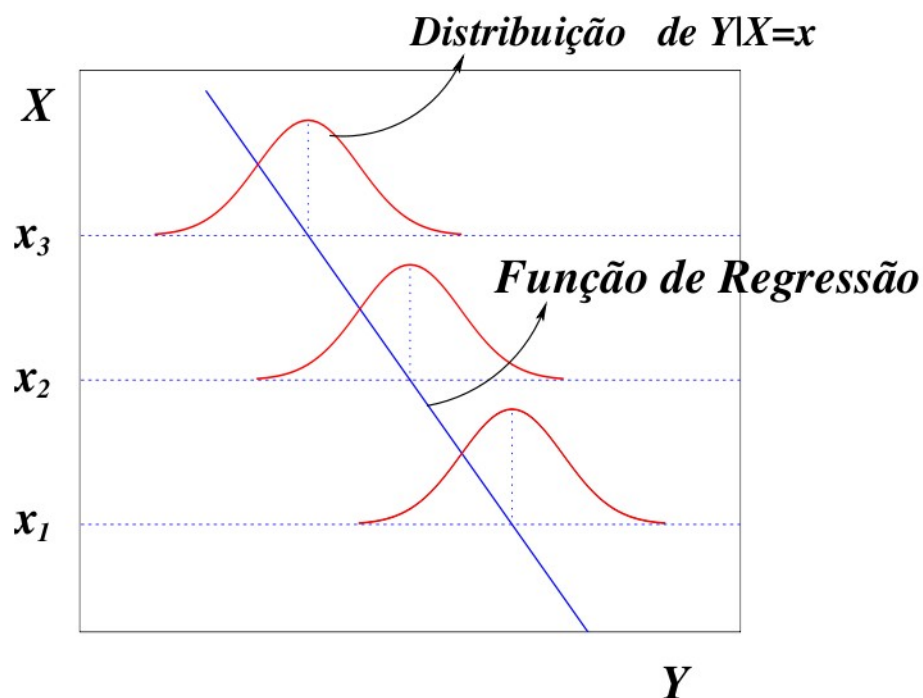
Multiple R-squared: 0.1661, Adjusted R-squared: 0.1654

F-statistic: 233.4 on 1 and 1172 DF, p-value: < 2.2e-16



# O MÉTODO DOS MÍNIMOS QUADRADOS

# Premissas do Modelo de Regressão Linear Gaussiana

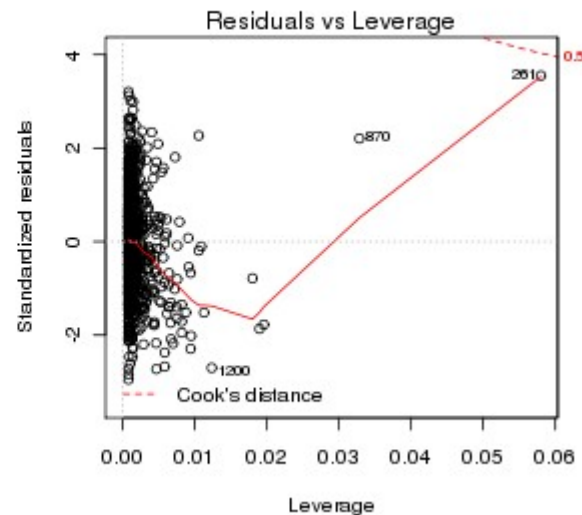
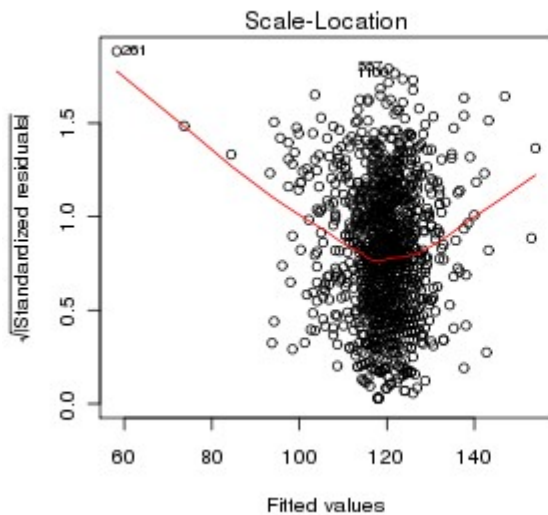
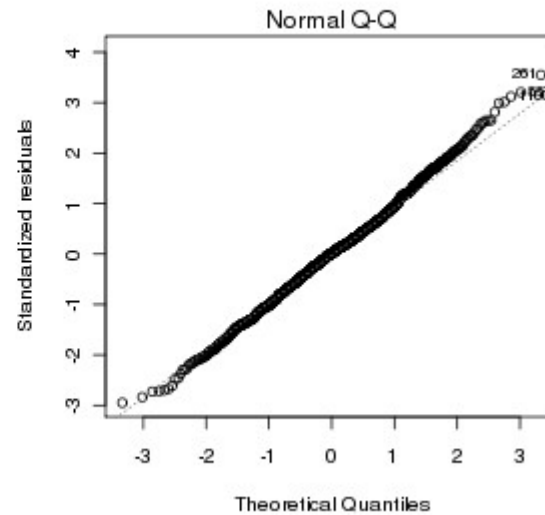
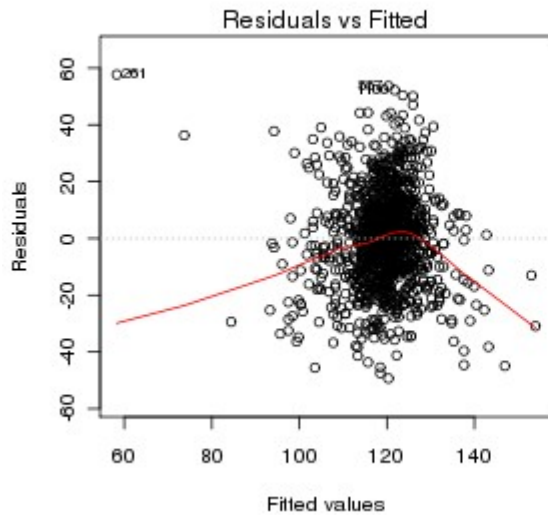


A variável resposta é uma variável normal (Gaussiana) sendo que:

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

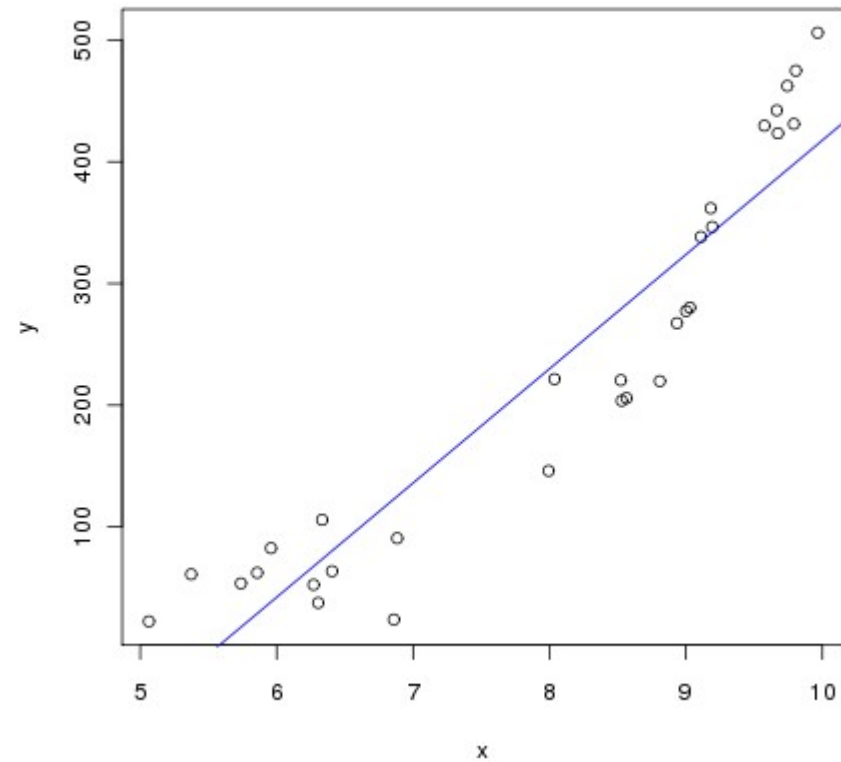
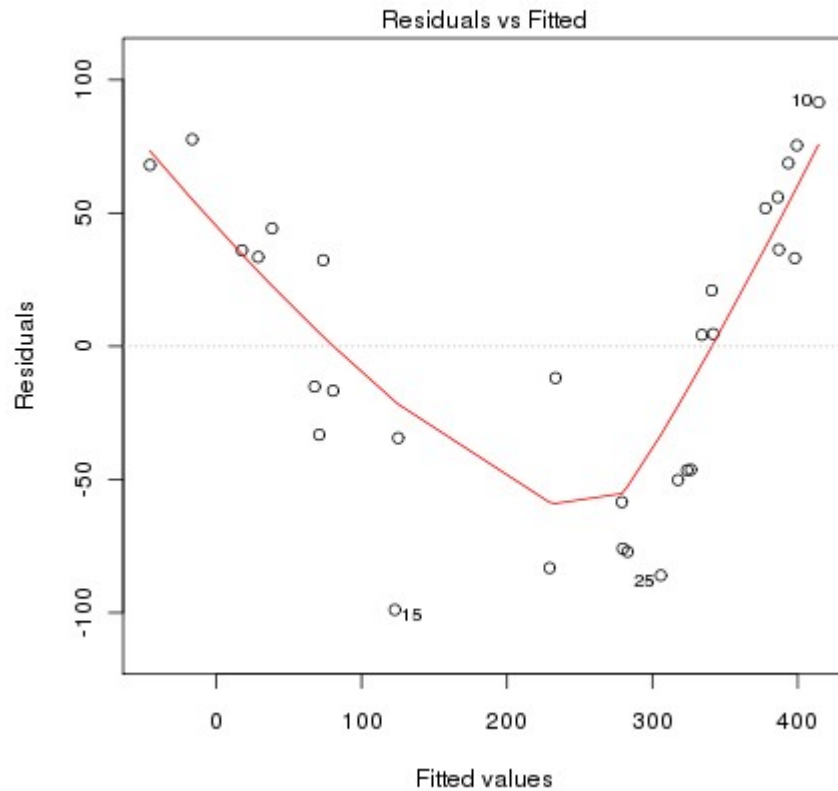
# plot.lm()

## Gráficos de Diagnóstico



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

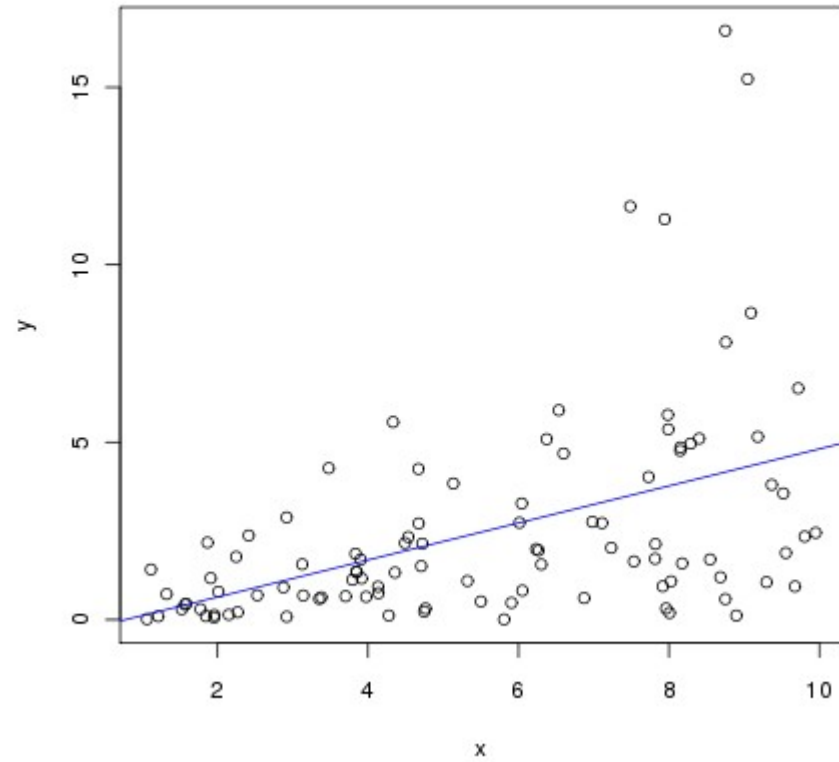
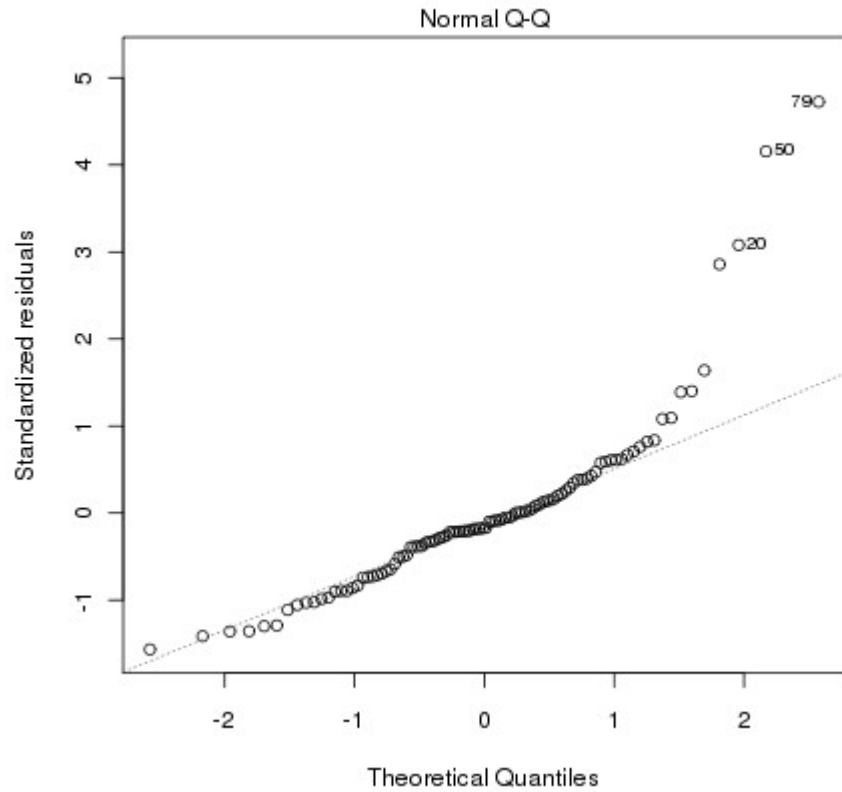
# Resíduos x Estimado



## Detecta:

- Tendências não-lineares
- Variâncias não homogêneas

# Gráfico de Quantis Resíduos x Normal



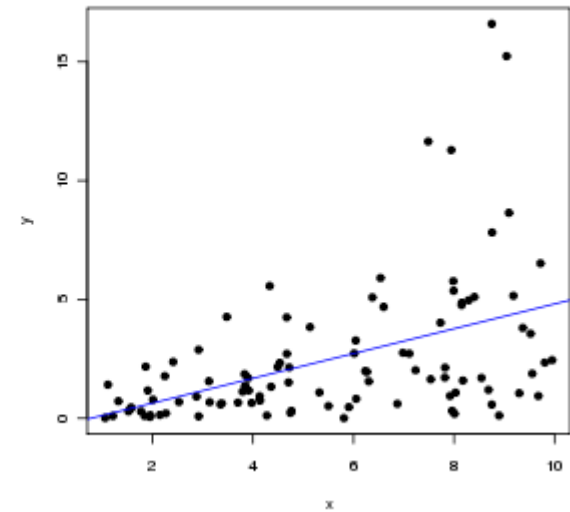
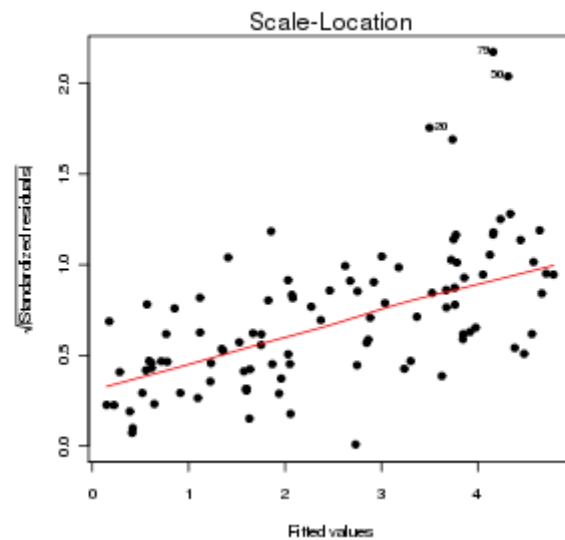
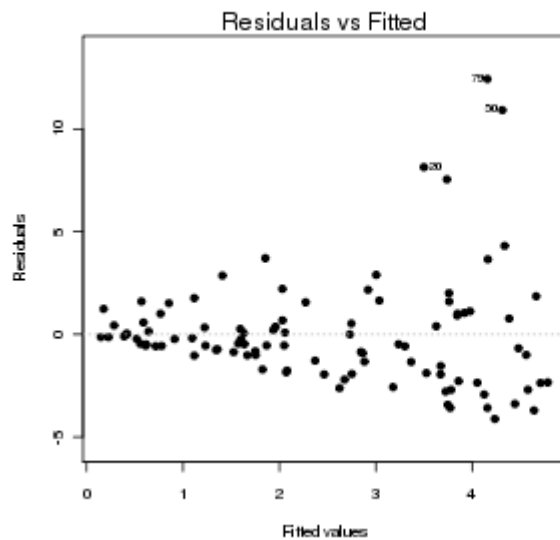
## Detecta:

- Desvios da normalidade nos resíduos



# Resíduos x Estimado

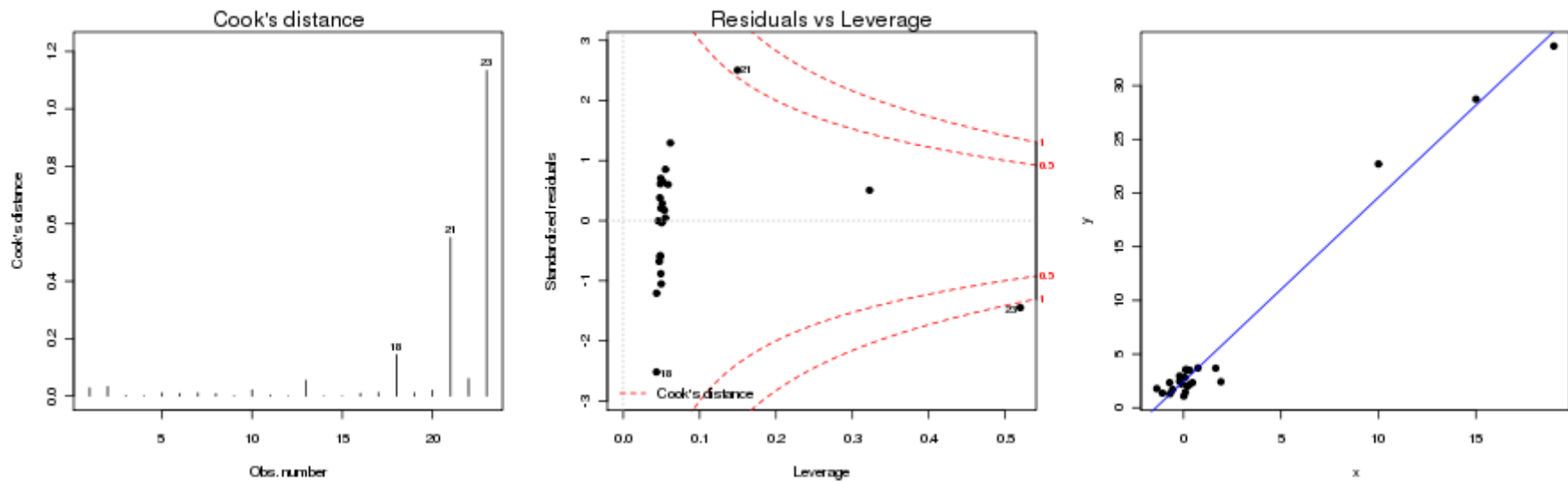
## Raiz dos Resíduos Padronizados x Estimado



### Detectam:

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

# Influência e Alavancagem



Detecta:

- Pontos influentes



Tutorial 1: ajuste e diagnóstico

Tutorial 2: simulação com dados não-normais

# FIM DA AULA 7

Para a tarde:

Aula 8 - Regressão Múltipla