Curso R

Modelos Lineares Múltiplos

Alexandre Adalardo de Oliveira

Ecologia- IBUSP maio 2017

Modelos Lineares: múltiplas preditoras

Modelos Lineares: múltiplas preditoras

- preditoras: contínuas e categóricas
- interação entre preditoras
- matriz do modelo
- simplificação do modelo
- colinearidade
- diagnóstico do modelo

Testes Clássicos

Tipo de Variável		Estatística Clássica	
Resposta	Preditora	Teste	Hipótese
Categórica	Categórica	Qui-quadrado	independência
Contínua	Categórica (2 niveis)	Teste t	$\mu_1 = \mu_2$
Contínua	Categórica	Anova	$\mu_1 = \mu_2 = \mu_n$
Contínua	1 Contínua	Regressão	$\beta_1 = 0$
Contínua	>1 Contínua	Reg. múltipla	$\beta_1 = 0 ; \beta_n = 0$
Contínua	Cont + Categ	Ancova	$\beta_1 = \beta_2 ; \alpha_1 = \alpha 2$
Proporção	Contínua	Reg. Logística	$logit(\beta_1) = 1$

O modelo de regressão múltipla

$$y = \hat{\alpha} + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_n x_n + \epsilon$$
$$\epsilon = N(0, \sigma)$$


R

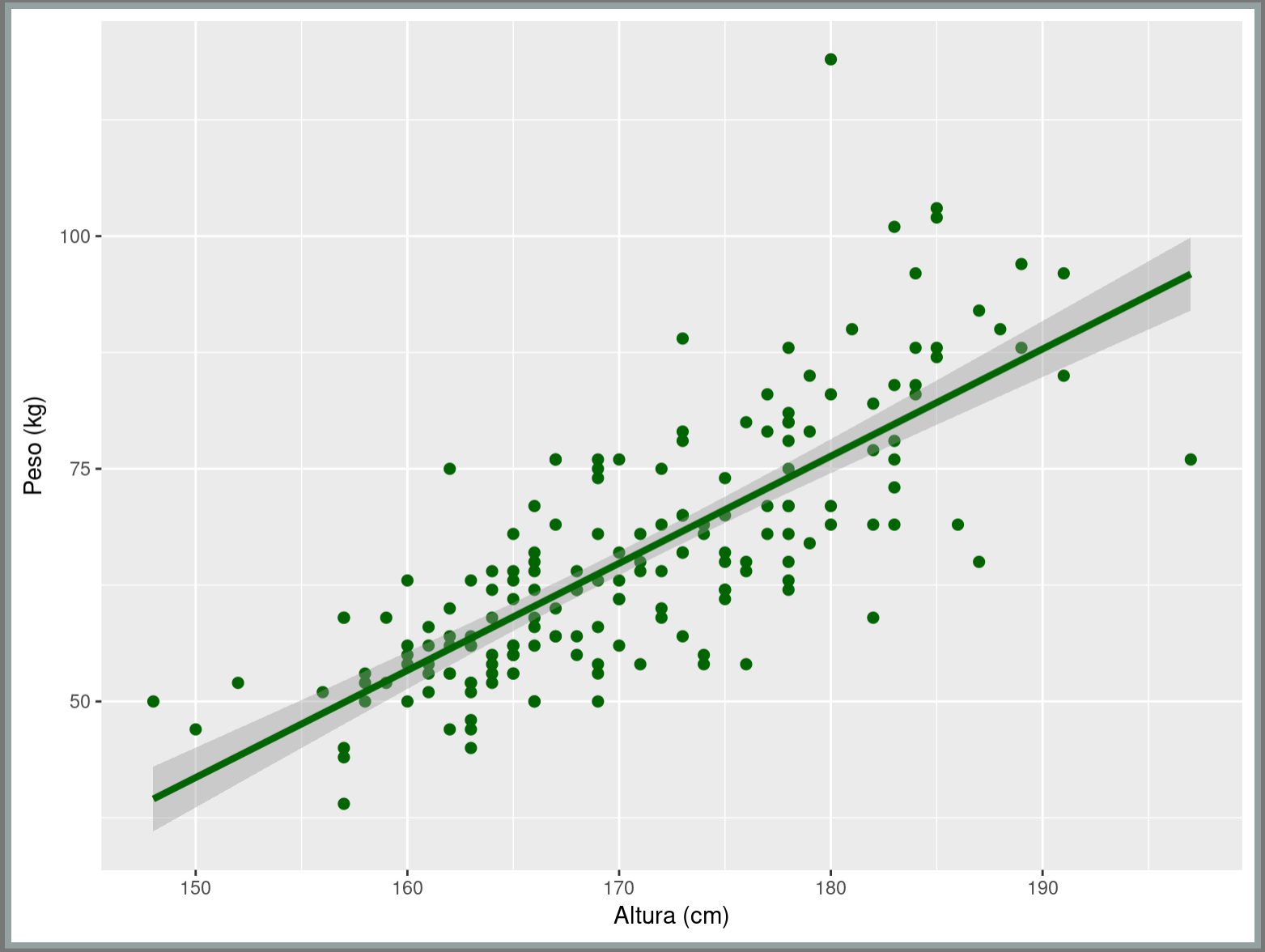
Retomando os modelos lineares

Peso ~ altura

```
library(car)
data(Davis)
str(Davis)
```

```
## 'data.frame':    200 obs. of  5 varia
##  $ sex   : Factor w/ 2 levels "F","M"
##  $ weight: int  77 58 53 68 59 76 76
##  $ height: int  182 161 161 177 157 1
##  $ repwt : int  77 51 54 70 59 76 77
##  $ repht : int  180 159 158 175 155 1
```

Gráfico: modelo linear



Resumo do lm

```
lmdavis <- lm(weight~height, data = Davi
summary(lmdavis)

##
## Call:
## lm(formula = weight ~ height, data =
##
## Residuals:
##     Min      1Q  Median      3Q     M
## -19.928  -5.406  -0.651   4.891  42.6
##
## Coefficients:
##             Estimate Std. Error t v
```

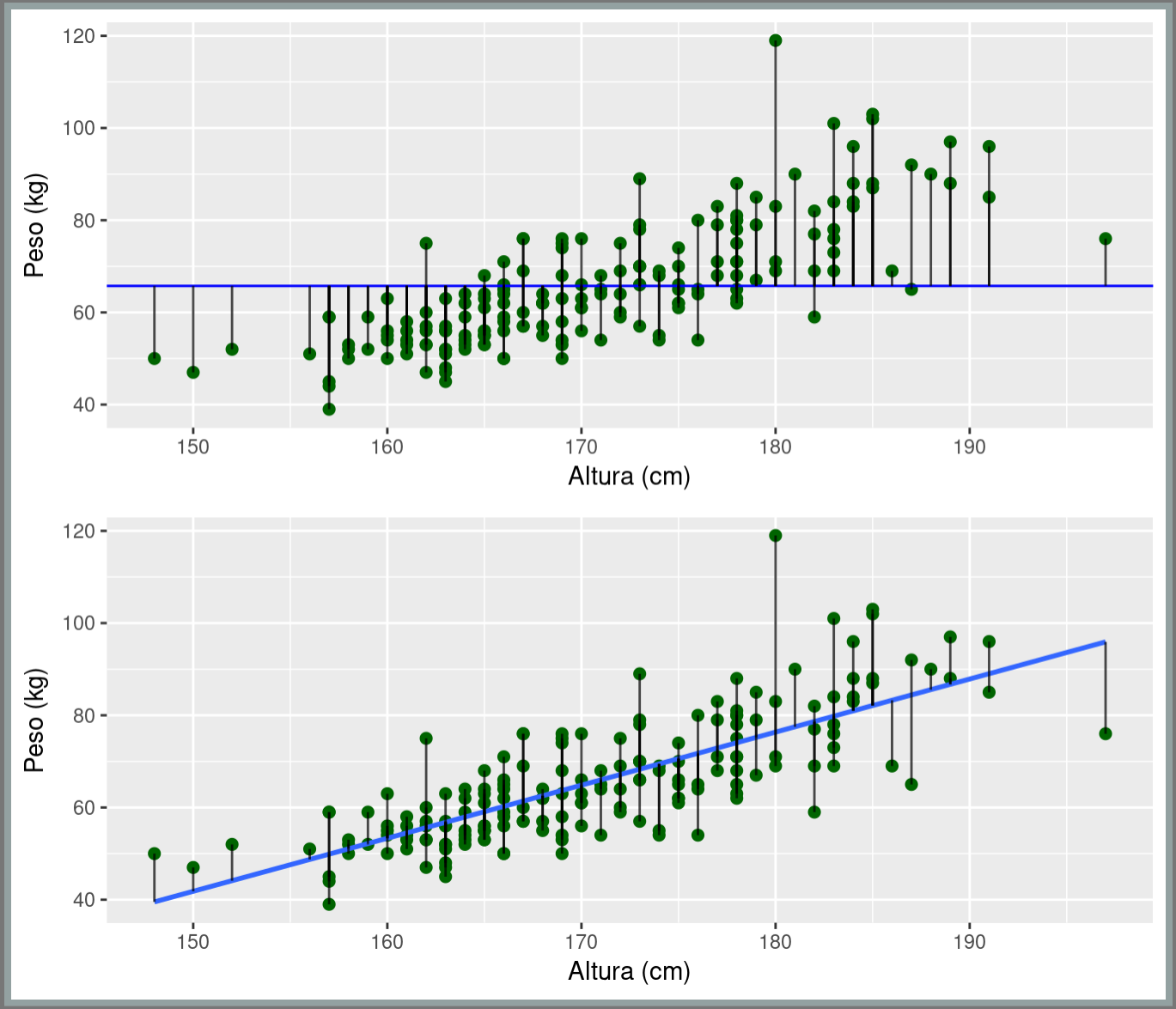
Anova do lm

```
lmdavis <- lm(weight~height, data = Davi

anova(lmdavis)

## Analysis of Variance Table
##
## Response: weight
##             Df Sum Sq Mean Sq F value
## height       1  19095 19095.0  256.08
## Residuals 178  13273    74.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Modelo Linear: peso ~ altura



ANOVA: comparando modelos

```
davisNull <- lm(weight ~ 1, data= Davis)
anova(davisNull,lmdavis)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ 1
## Model 2: weight ~ height
##   Res.Df   RSS Df Sum of Sq      F
## 1    179 32368
## 2    178 13273  1     19095 256.08 <
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

$$p_{valor} = 2.2e{-}16 = 2.2x10^{-16}$$

$$r^2 = 0.587$$

Anova: comparando modelos

Particionando Variância do Modelo

```
kable(anova(lmdavis),digits = 2, align =
```

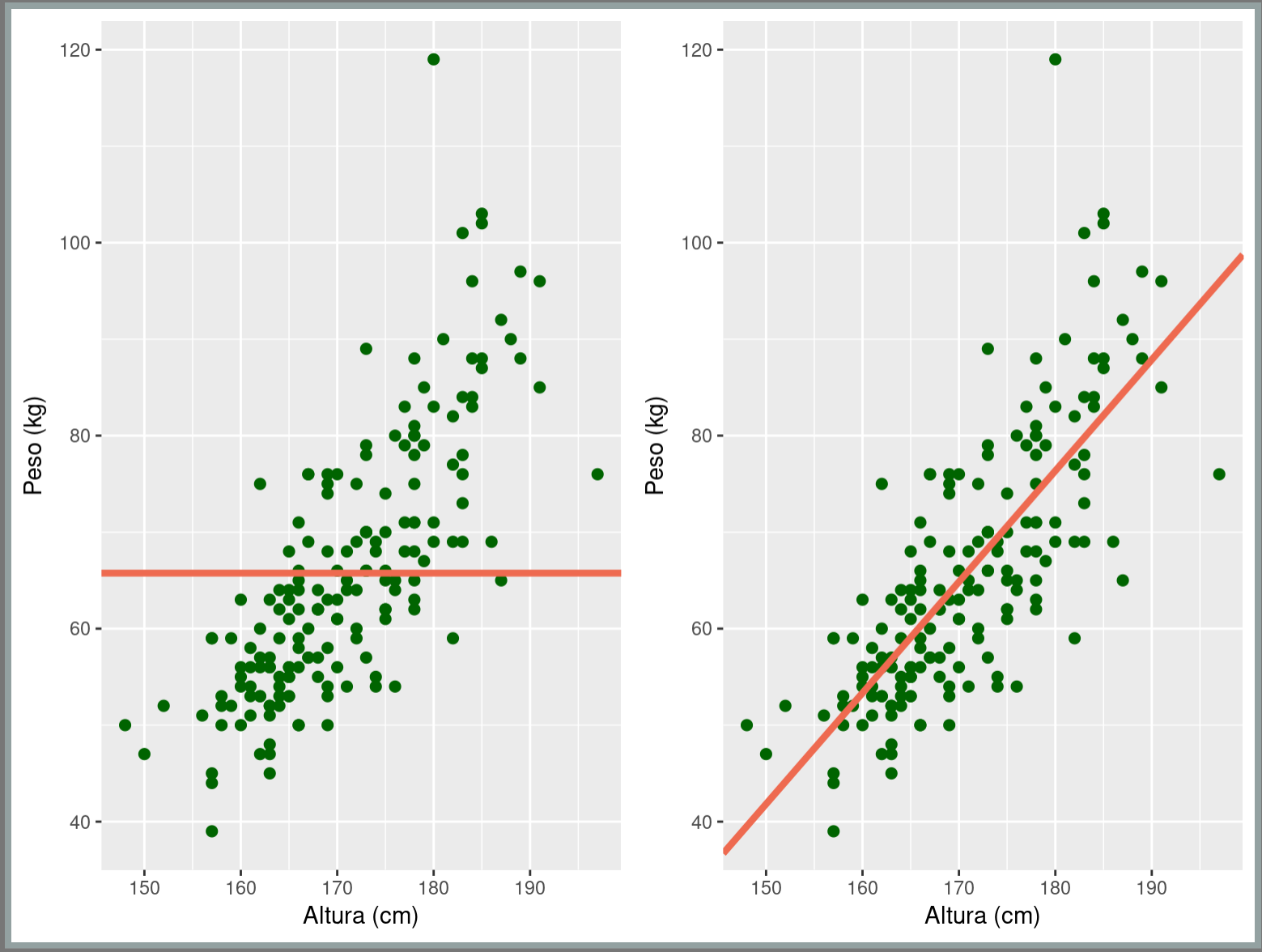
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
height	1	19095.04	19095.04	256.08	0
Residuals	178	13272.71	74.57	NA	NA

Confrontando com o modelo mínimo

```
kable(anova(davisNull, lmdavis), digits
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
179	32367.75	NA	NA	NA	NA
178	13272.71	1	19095.04	256.08	0

Comparando modelos

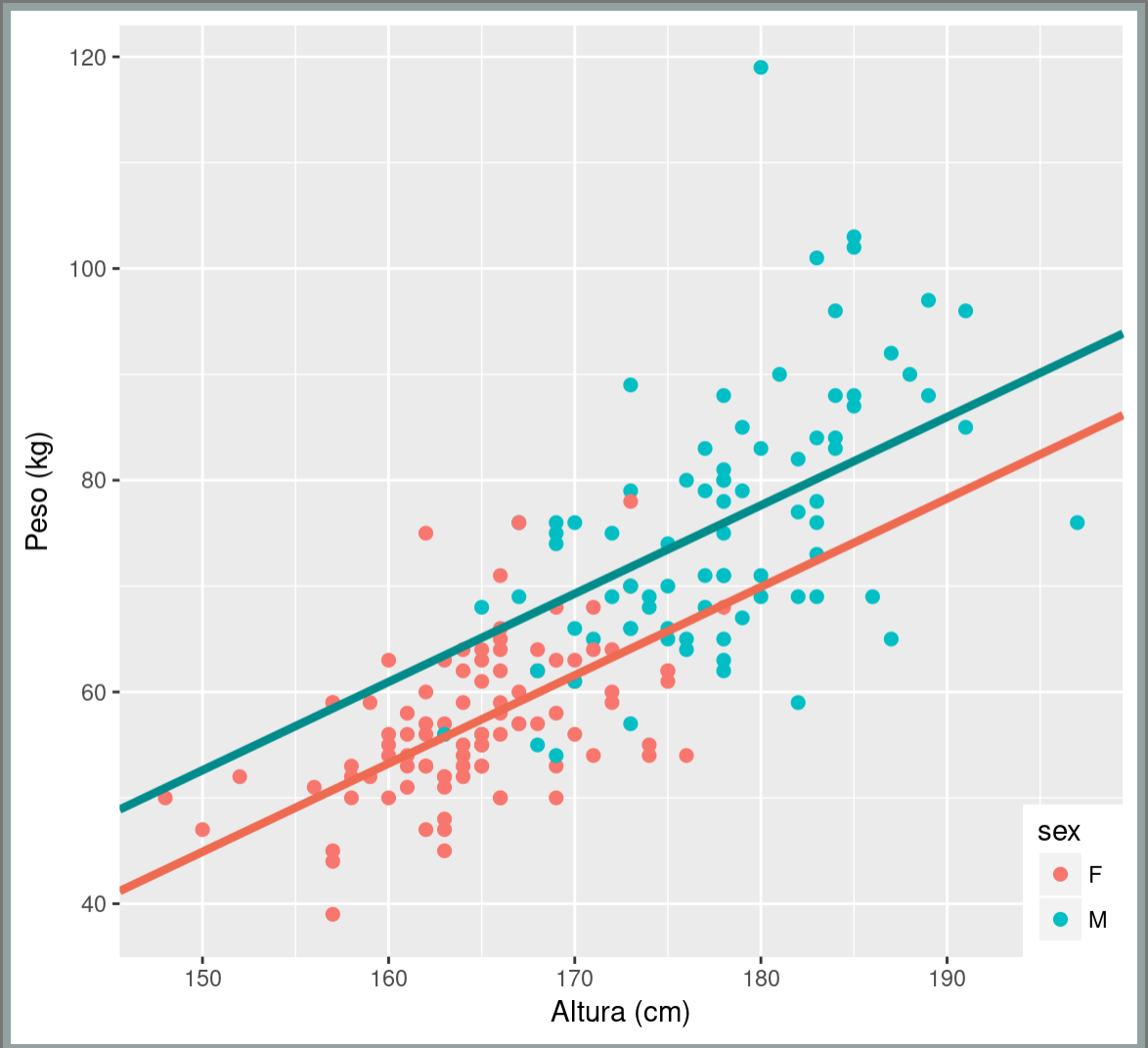



R

Multiplas Preditoras

Múltiplas Preditoras

```
lmdavis01 <- lm(weight~ height + sex, da
```



Preditora: contínua + fator

```
summary(lmdavis01)
```

```
##
## Call:
## lm(formula = weight ~ height + sex, d
##
## Residuals:
##     Min      1Q  Median      3Q     M
## -20.302  -4.808  -0.335   5.239  41.3
##
## Coefficients:
##             Estimate Std. Error t val
```

lm(weight ~ height + sex, data = Davis)

```
anova(lmdavis01)
```

```
## Analysis of Variance Table
##
## Response: weight
##             Df  Sum Sq Mean Sq F value
## height       1 19095.0 19095.0  280.04
## sex          1  1203.5  1203.5   17.65
## Residuals  177 12069.2    68.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

lm(weight ~ sex + height, data = Davis)

```
lmdavis01a <- lm(weight~sex + height, da
anova(lmdavis01a)

## Analysis of Variance Table
##
## Response: weight
##             Df  Sum Sq Mean Sq F value
## sex          1 15748.5 15748.5 230.958
## height       1  4550.1  4550.1  66.728
## Residuals 177 12069.2    68.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Anova do modelo

```
anova(lmdavis01a, lmdavis01)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ sex + height
## Model 2: weight ~ height + sex
##   Res.Df   RSS Df Sum of Sq F Pr(>F)
## 1    177 12069
## 2    177 12069  0         0
```

Anova: múltiplas preditoras

```
kable(anova(lmdavis01))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
height	1	19095.041	19095.04065	280.0366	0.0e+00
sex	1	1203.492	1203.49189	17.6497	4.2e-05
Residuals	177	12069.217	68.18767	NA	NA

```
kable(anova(lmdavis01a))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	15748.48	15748.48245	230.95792	0
height	1	4550.05	4550.05009	66.72834	0
Residuals	177	12069.22	68.18767	NA	NA

Anova do Modelo

```
anova(lmdavis01)
```

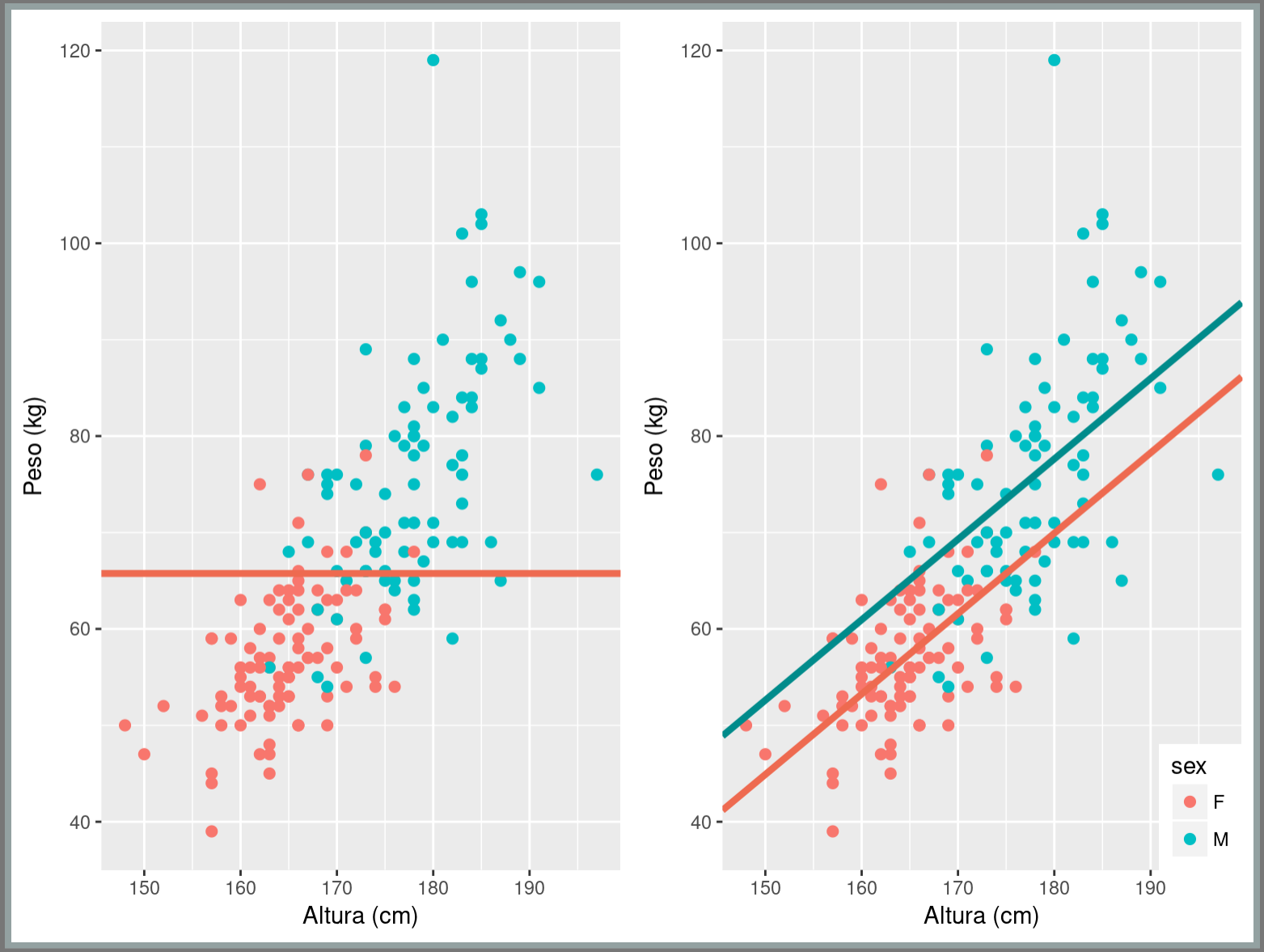
```
## Analysis of Variance Table
##
## Response: weight
##             Df  Sum Sq Mean Sq F value
## height       1 19095.0 19095.0  280.04
## sex          1  1203.5  1203.5   17.65
## Residuals  177 12069.2    68.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Confrontando com o modelo mínimo

```
anova(davisNull, lmdavis01)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ 1
## Model 2: weight ~ height + sex
##   Res.Df   RSS Df Sum of Sq      F
## 1    179 32368
## 2    177 12069  2     20298 148.84 <
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Comparação de Modelos



Estimativas do lm

```
coeflm01 <- coef(lmdavis01)
kable(coeflm01, digits = 2, align = "c")
```

(Intercept)	-80.21
height	0.83
sexM	7.71

Feminino ($sex = 0$)

$$w_f = \hat{\alpha} + \hat{\beta}_s sex + \hat{\beta}_h * height$$
$$w_f = \hat{\alpha} + \hat{\beta}_h * height$$

Estimativas do lm

(Intercept)	-80.21
height	0.83
sexM	7.71

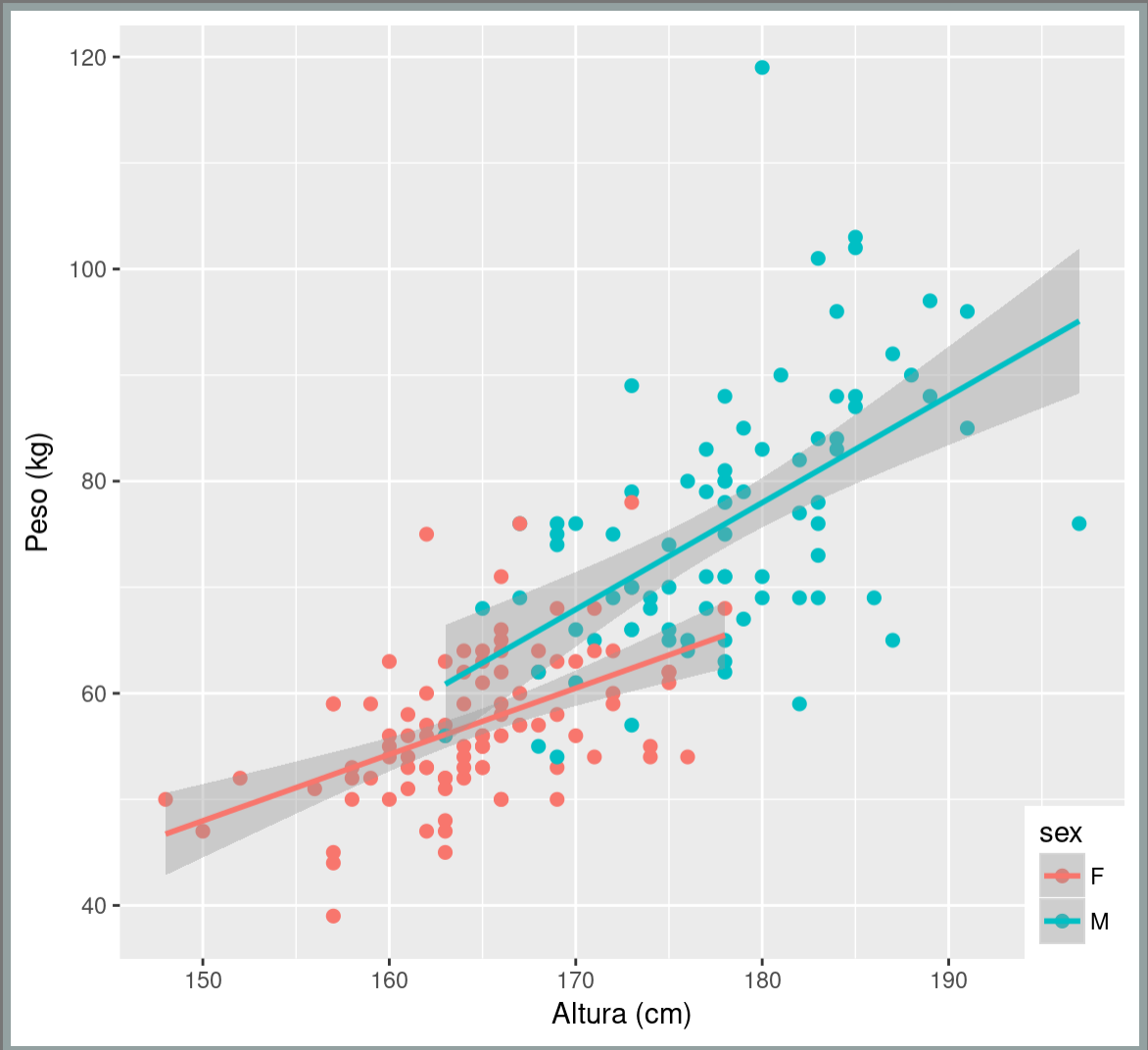
Masculino ($sex = 1$)

$$w_h = \hat{\alpha} + \hat{\beta}_s * sex + \hat{\beta} * height$$
$$w_h = \hat{\alpha} + \hat{\beta}_s + \hat{\beta}_h * height$$

Interação

Interação

```
lmdavisfull <- lm(weight ~ height + sex
```



Interação

```
summary(lmdavisfull)
```

```
##
## Call:
## lm(formula = weight ~ height + sex +
##
## Residuals:
##     Min      1Q  Median      3Q     M
## -20.990  -4.548  -0.926   4.821  41.0
##
## Coefficients:
##             Estimate Std. Error t val
```

Anova: comparando modelos

Multiplos testes

```
anova(lmdavisfull)
```

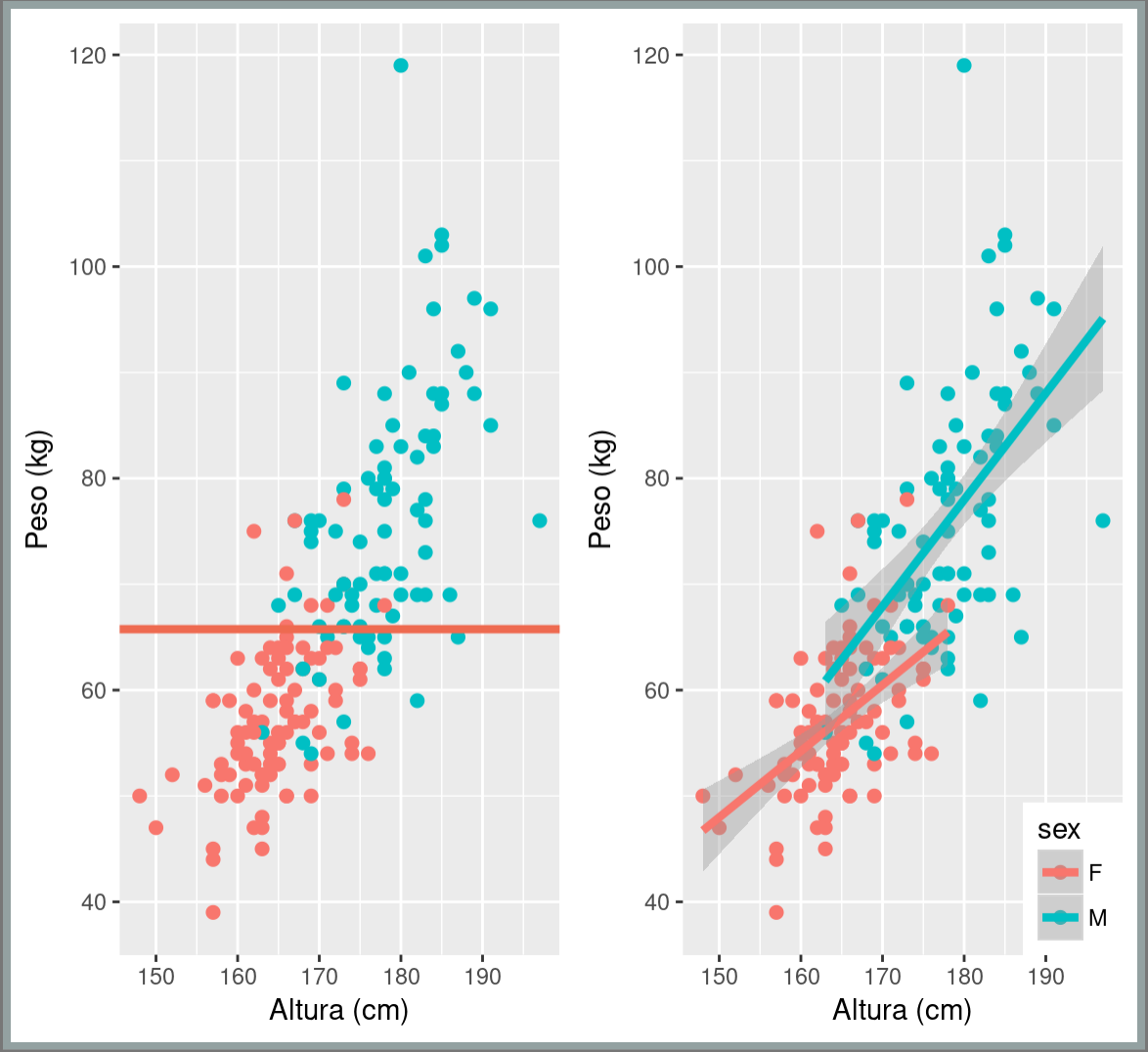
```
## Analysis of Variance Table
##
## Response: weight
##               Df  Sum Sq Mean Sq  F val
## height         1 19095.0 19095.0 284.00
## sex            1  1203.5  1203.5  17.89
## height:sex     1   235.8   235.8   3.50
## Residuals    176 11833.4    67.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0
```

Anova: comparando modelos

```
anova(lmdavisfull, davisNull)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ height + sex + sex:
## Model 2: weight ~ 1
##   Res.Df   RSS Df Sum of Sq     F
## 1    176 11833
## 2    179 32368 -3    -20534 101.8 < 2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Interação x nulo



lm(weight ~ height + sex*height, data=Davis)

```
## (Intercept)      height        sexM h
## -45.7988220   0.6252035 -57.4326307
```

Feminino ($sex = 0$)

$$w = \hat{\alpha} + \hat{\beta}_s sex + \hat{\beta}_h height + \hat{\beta}_{s:h} sex * height$$
$$w_m = \hat{\alpha} + \hat{\beta}_h height$$

Masculino ($sex = 1$)

$$w = \hat{\alpha} + \hat{\beta}_s sex + \hat{\beta}_h height + \hat{\beta}_{h:s} sex * height$$
$$w_h = \hat{\alpha} + \hat{\beta}_s + (\hat{\beta}_h + \hat{\beta}_{h:s}) * height$$

Predição do modelo

Uma mulher de 161cm de altura

$$w = \hat{\alpha} + \hat{\beta}_s sex + \hat{\beta}_h height + \hat{\beta}_{s:h} sex * height$$
$$sex = 0$$

```
(coefull <- coef(lmdavisfull))
```

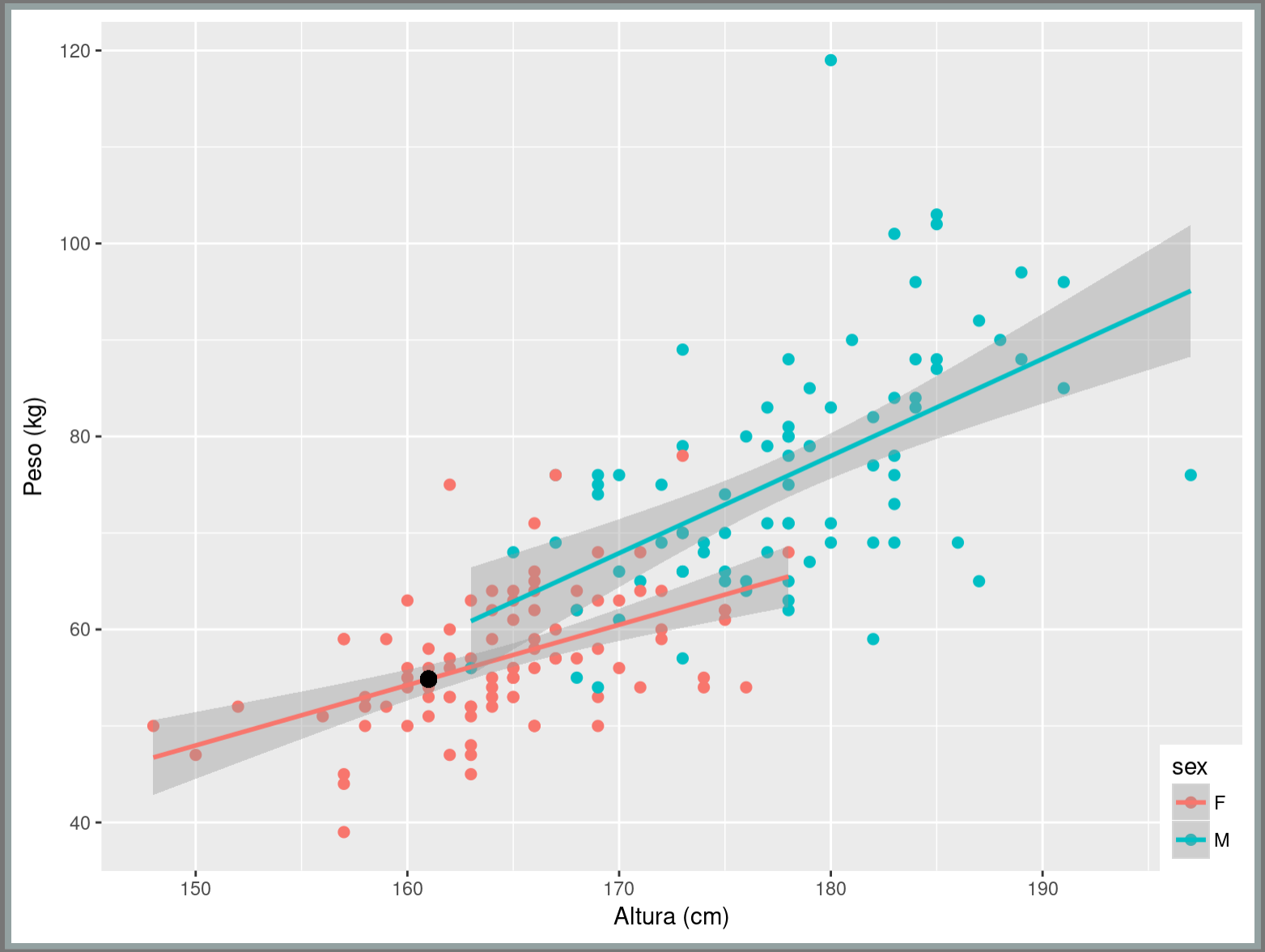
```
## (Intercept)      height        sexM h
## -45.7988220   0.6252035 -57.4326307
```

```
predMulher <- coefull[1] + coefull[2] *
(predMulher <- as.numeric(predMulher))
```

```
## [1] 54.85893
```

lm(weight ~ height + sex*height, data=Davis)

- Uma mulher com 161cm de altura tem peso 54.86 kg.



Predito do Modelo

Homem com 182cm

$$w = \hat{\alpha} + \hat{\beta}_s sex + \hat{\beta}_h height + \hat{\beta}_{s:h} sex * height$$
$$sex = 1$$

```
coefull
```

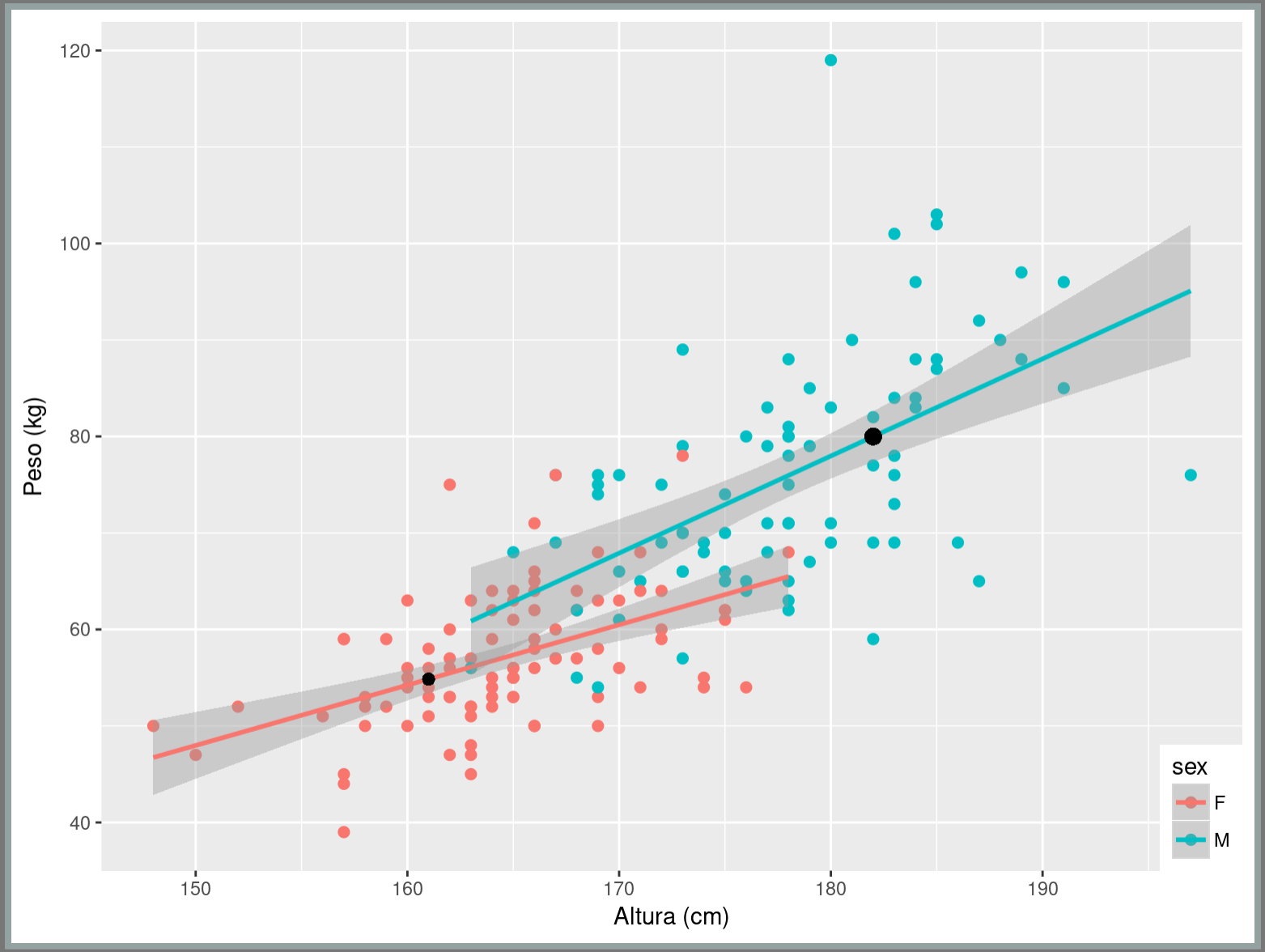
```
## (Intercept)      height        sexM h
## -45.7988220   0.6252035 -57.4326307
```

```
predHomem <- (coefull[1]+ coefull[3]) +
(predHomem <- as.numeric(predHomem))
```

```
## [1] 79.99018
```

lm(weight ~ height + sex*height, data=Davis)

- Um homem com 182cm de altura tem peso 79.99 kg.



Matriz do Modelo

```
Davis[1:2,1:3]
```

```
##   sex weight height
## 1   M     77    182
## 2   F     58    161
```

```
model.matrix(lmdavisfull)[1:2,]
```

```
##   (Intercept) height sexM height:sexM
## 1           1    182    1         182
## 2           1    161    0           0
```

```
coef(lmdavisfull)
```

```
## (Intercept)      height        sexM h
## -45.7988220   0.6252035 -57.4326307
```

Matriz do Modelo

```
kable(Davis[1:2,1:3])
```

sex	weight	height
M	77	182
F	58	161

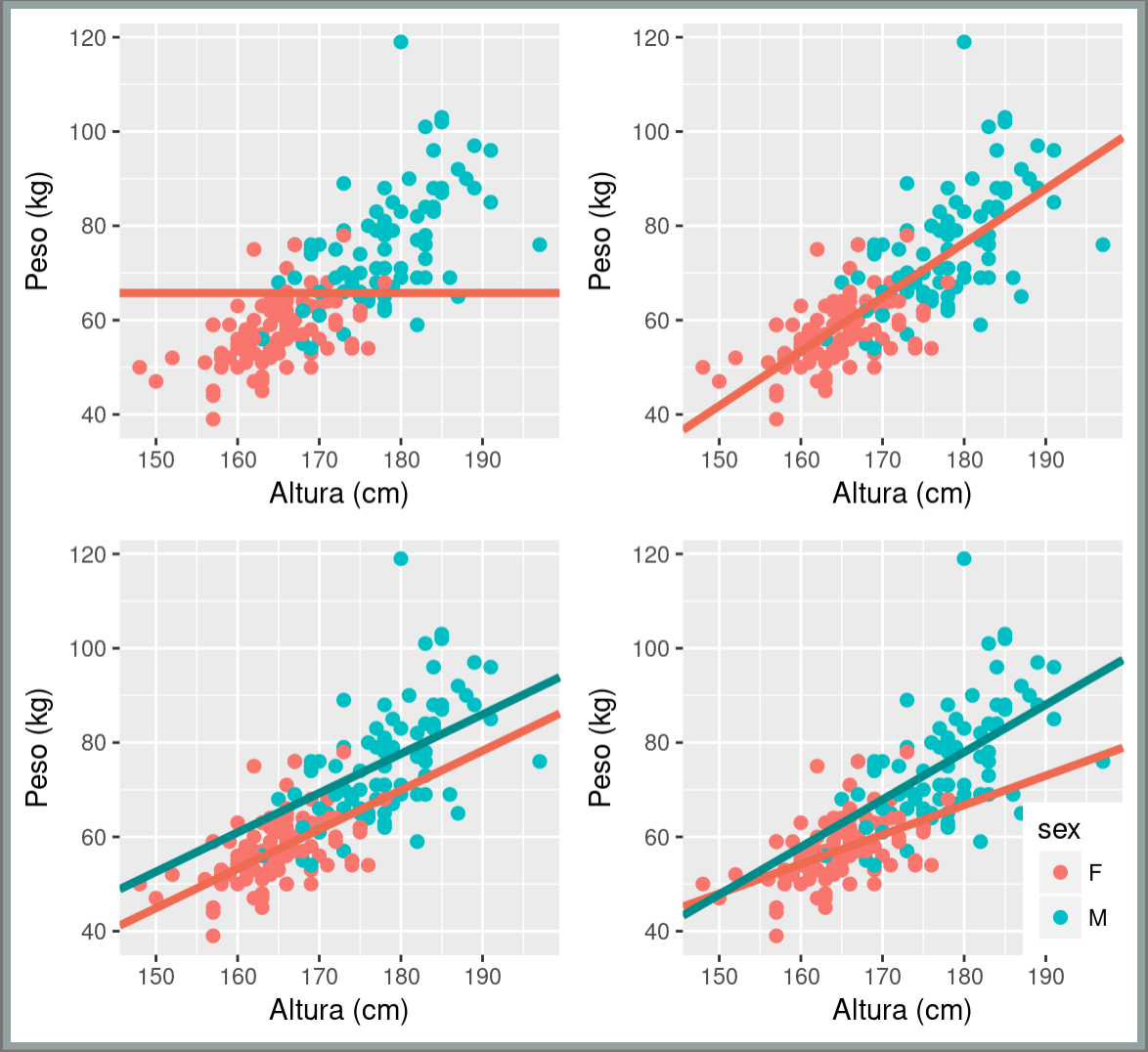
```
model.matrix(lmdavisfull)[1:2,] %*% coef
```

```
##        [,1]
## 1 79.99018
## 2 54.85893
```

```
predict(lmdavisfull)[1:2]
```

```
##        1        2
## 79.99018 54.85893
```

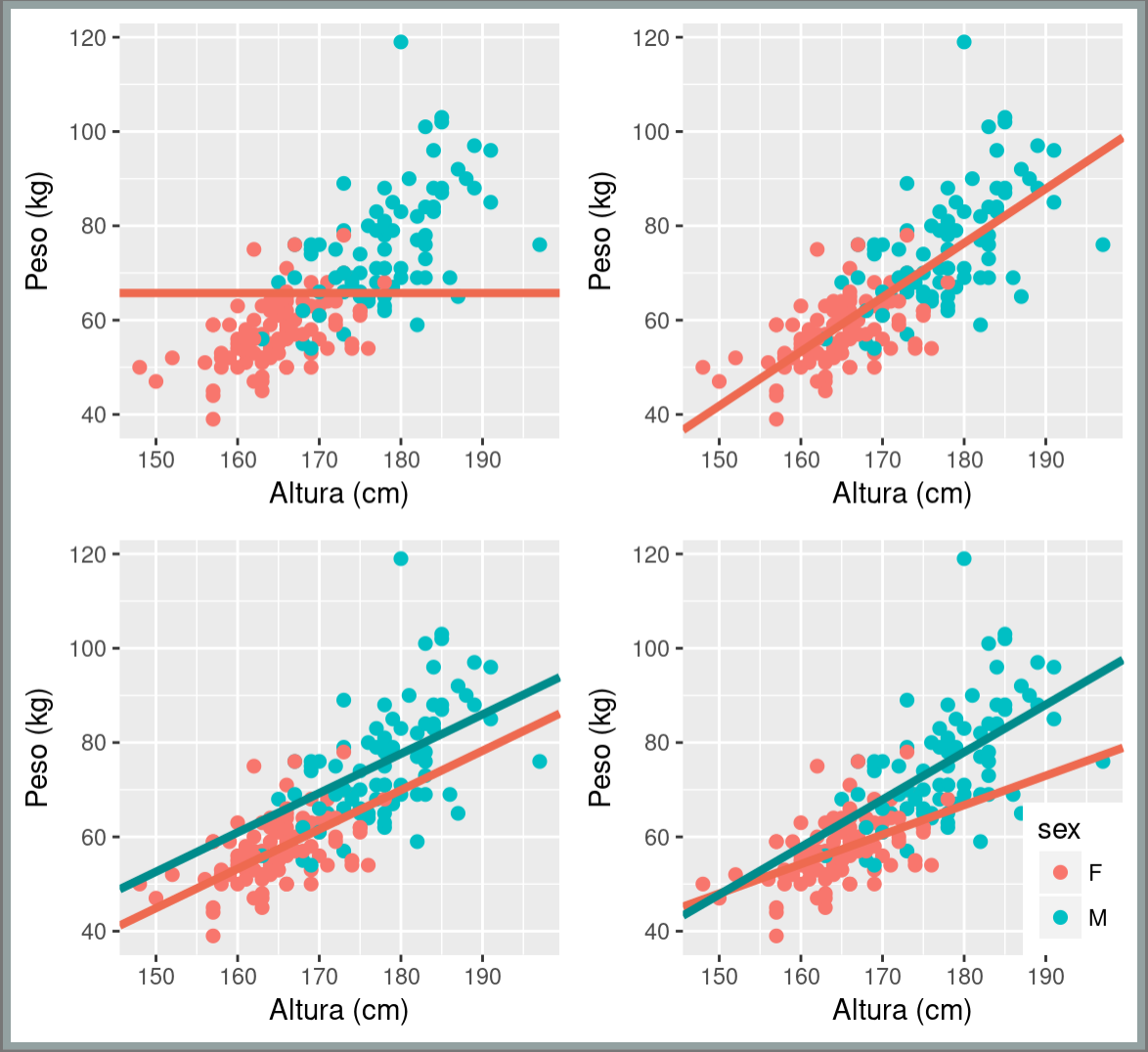
Modelos Concorrentes




R

Seleção de Modelo

Qual o modelo?



Seleção de modelo

Defina a priori os modelos concorrentes

- sentido biológico
- parâmetros com significado

Compare usando o Princípio da parcimônia

(Navalha de Occam)

- devem ter menos parâmetros possível
- linear é melhor que não-linear
- reter menos pressupostos
- simplificado ao mínimo adequado
- explicações mais simples são preferíveis

Simplificação do modelo

Método do modelo cheio ao mínimo adequado

1. ajuste o modelo máximo (cheio)
2. simplifique o modelo:
    - inspecione os coeficientes (summary)
    - remova termos não significativos
3. ordem de remoção de termos:
    - interação não significativos (maior ordem)
    - termos quadráticos ou não lineares
    - variáveis explicativas não significativas
    - agrupe níveis de fatores sem diferença
    - ANCOVA: intercepto não significativoa -> 0

Simplificação do modelo: continuação

Compare o modelo anterior com o simplificado

A diferença não é significativa:

* **retenha o modelo mais simples**
* **continue simplificando**

A difereça é significativa

* **retenha o modelo complexo**
* **este é o modelo MINÍMO ADEQUADO**

Simplificando Modelo: exemplo

```
anova(lmdavisfull, lmdavis01)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ height + sex + sex:
## Model 2: weight ~ height + sex
##   Res.Df   RSS Df Sum of Sq      F  P
## 1    176 11833
## 2    177 12069 -1   -235.82 3.5075 0.
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Simplificando Modelo: exemplo

```
anova(lmdavis01, lmdavis)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ height + sex
## Model 2: weight ~ height
##   Res.Df   RSS Df Sum of Sq     F
## 1    177 12069
## 2    178 13273 -1   -1203.5 17.65 4.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.
```

Modelo Mínimo Adequado

```
summary(lmdavis01)
```

```
##
## Call:
## lm(formula = weight ~ height + sex, d
##
## Residuals:
##     Min      1Q  Median      3Q     M
## -20.302  -4.808  -0.335   5.239  41.3
##
## Coefficients:
##             Estimate Std. Error t val
```

Anova

Anova do modelo: comparação entre modelos

```
kable(anova(lmdavisfull), align= "c", di
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
height	1	19095.0407	19095.0407	284.0037	0.0000
sex	1	1203.4919	1203.4919	17.8997	0.0000
height:sex	1	235.8241	235.8241	3.5075	0.0628
Residuals	176	11833.3933	67.2352	NA	NA
### Anova ent	re mode	los			

```
kable(anova(lmdavis01, lmdavisfull), ali
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
177	12069.22	NA	NA	NA	NA
176	11833.39	1	235.8241	3.5075	0.0628

Anova sequencial

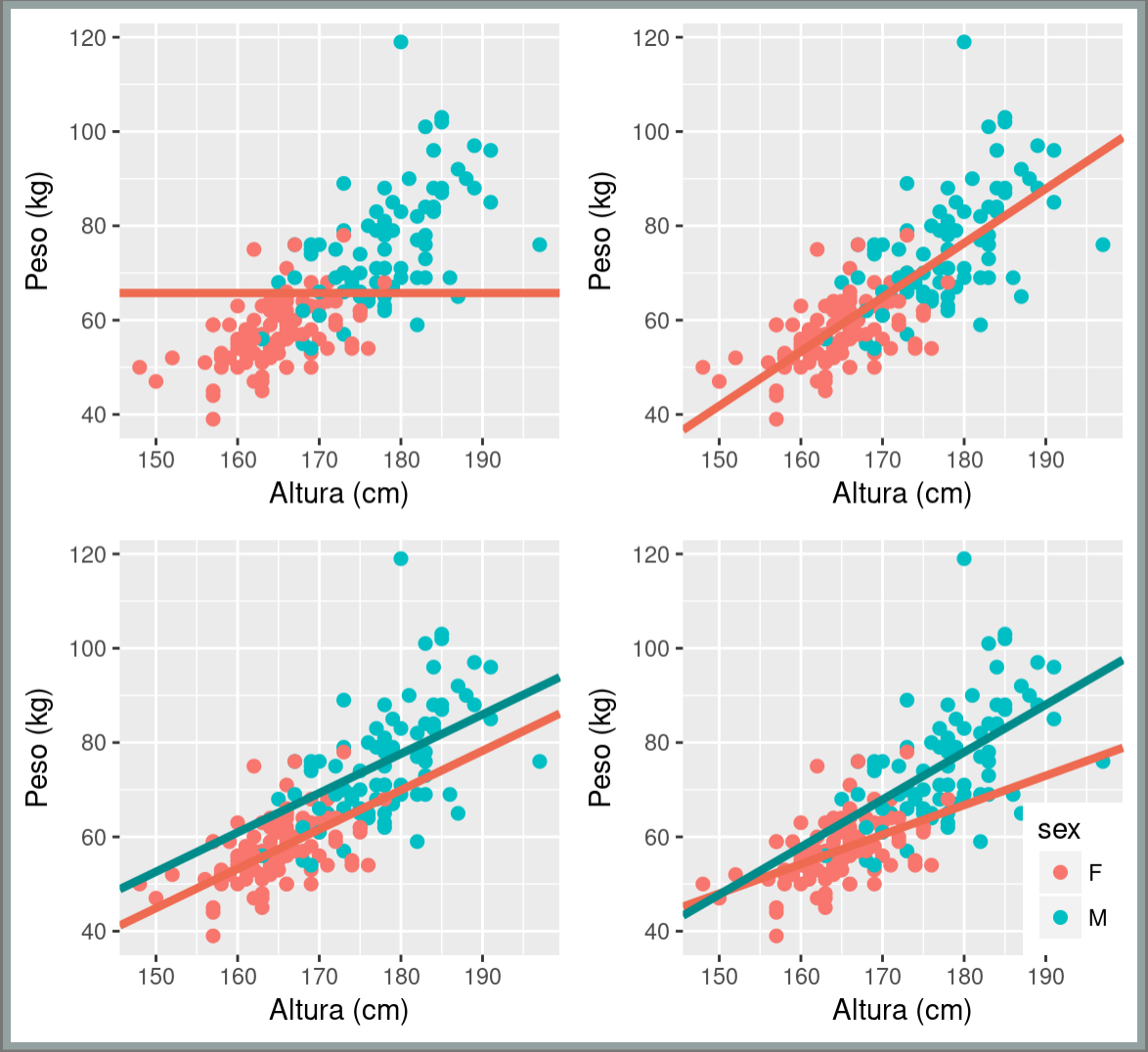
```
kable(anova(davisNull, lmdavis,lmdavis01
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
179	32367.75	NA	NA	NA	NA
178	13272.71	1	19095.0407	284.0037	0.0000
177	12069.22	1	1203.4919	17.8997	0.0000
176	11833.39	1	235.8241	3.5075	0.0628

Anova do cheio

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
height	1	19095.0407	19095.0407	284.0037	0.0000
sex	1	1203.4919	1203.4919	17.8997	0.0000
height:sex	1	235.8241	235.8241	3.5075	0.0628
Residuals	176	11833.3933	67.2352	NA	NA

Modelo sem interação!



Modelo Mínimo Adequado

```
coef(lmdavis01)
```

```
## (Intercept)      height        sexM
## -80.2107328   0.8340964   7.7070166
```

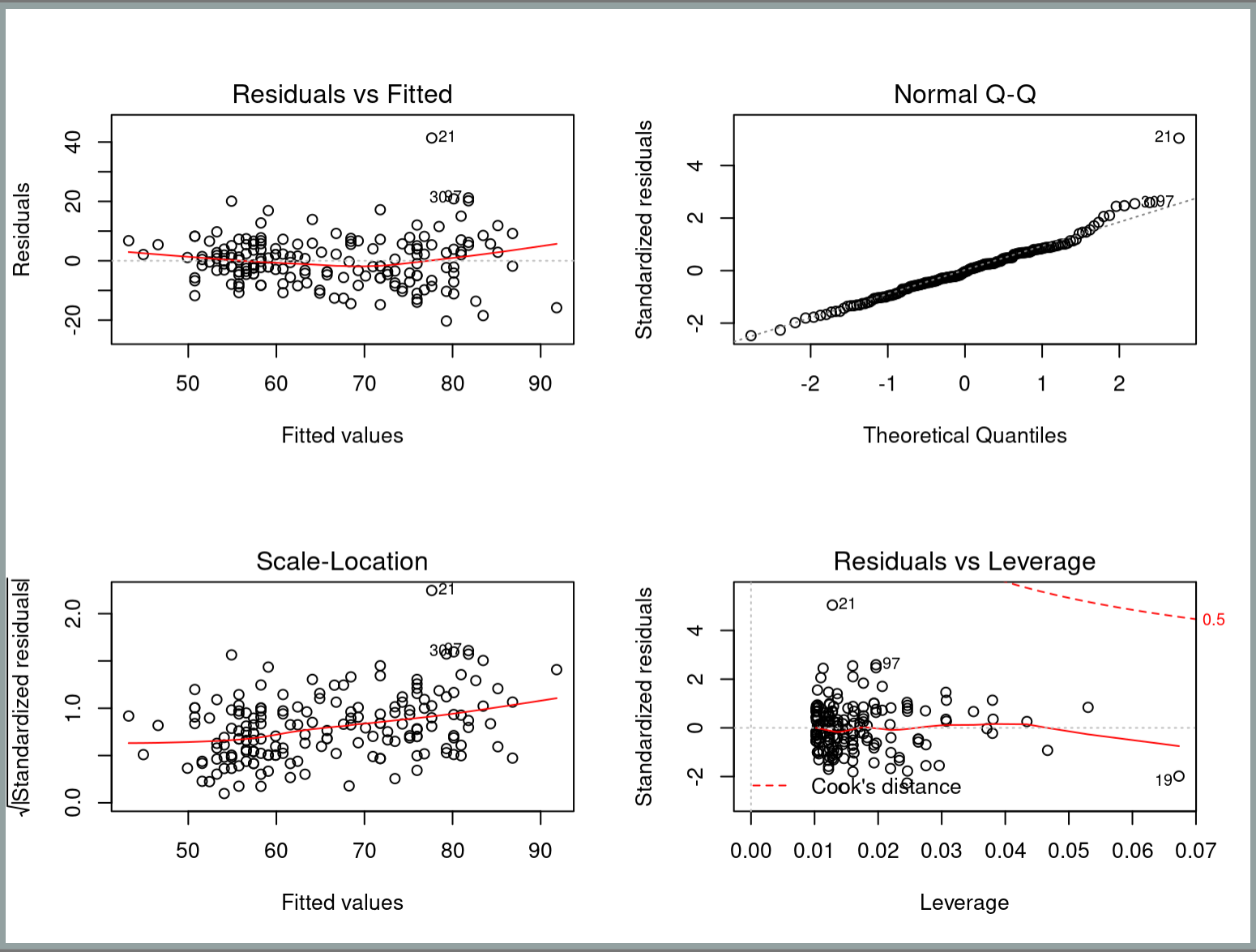
```
confint(lmdavis01)
```

```
##                   2.5 %     97.5 %
## (Intercept) -113.44661 -46.974852
## height         0.63259   1.035603
## sexM           4.08671  11.327323
```

Diagnóstico do Modelo: plot(modelo)

```
par(mfrow = c(2,2))
plot(lmdavis01)
```

Diagnóstico: plot(modelo)



Diagnóstico: plot(modelo)

```
##     sex weight height repwt repht     p
## 20    F     61    175    61   171 70.6
## 21    M    119    180   124   178 76.3
## 22    F     61    170    61   170 64.8
## 23    M     65    175    66   173 70.6
## 24    M     66    173    70   170 68.3
## 25    F     54    171    59   168 65.9
## 26    F     50    166    50   165 60.2
```


R

Atividades da Tarde

- APOSTILA
- Tutorial
- Exercícios