Modelos Lineares Múltiplos

unificação metodológica

Alexandre Adalardo de Oliveira
PlanECO 2017

Conceitos

- preditoras contínuas e categóricas
- interação entre preditoras
- matriz do modelo (algebra linear)
- simplificação do modelo
- colinearidade

Modelos Lineares Múltiplos

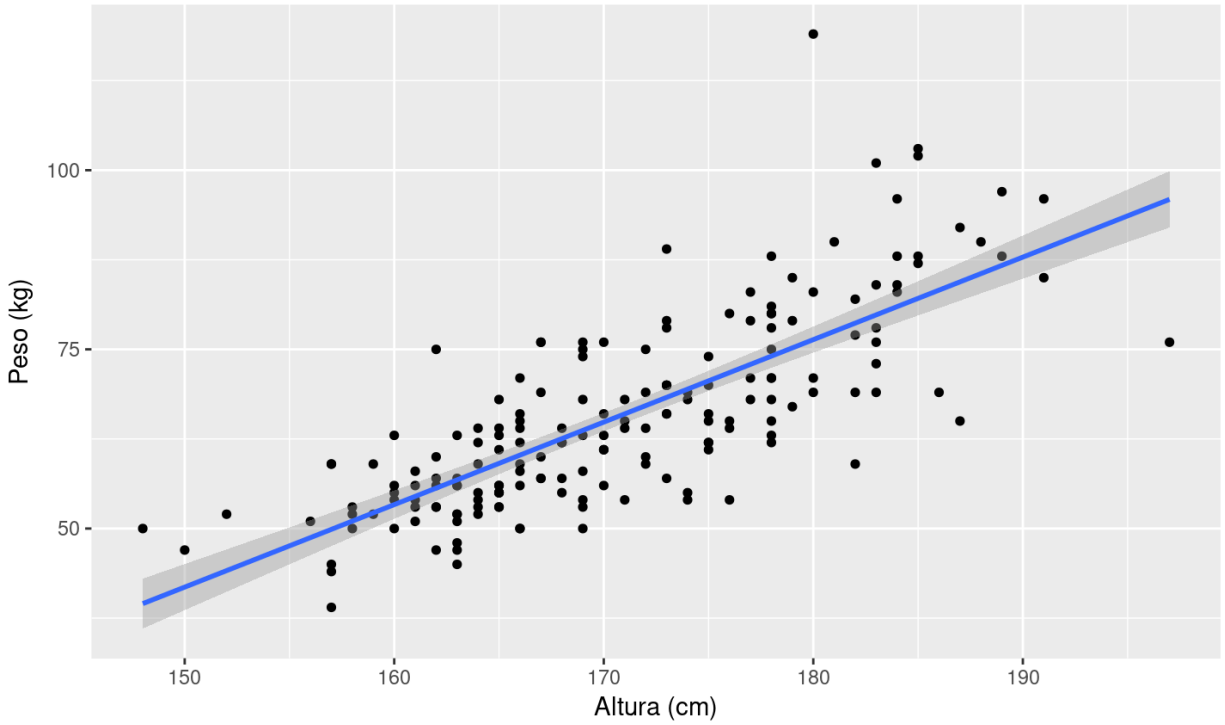
Retomando a regressão

Davis (peso ~ altura)

```
## 'data.frame':    180 obs. of  5 variables:
##  $ sex   : Factor w/ 2 levels "F","M": 2 1 1 2 1 2 2 2 2 2 ...
##  $ weight: int  77 58 53 68 59 76 76 69 71 65 ...
##  $ height: int  182 161 161 177 157 170 167 186 178 171 ...
##  $ repwt : int  77 51 54 70 59 76 77 73 71 64 ...
##  $ repht : int  180 159 158 175 155 165 165 180 175 170 ...
```

Há relação entre peso e altura

Gráfico da Regressão: peso ~ altura



lmdavis <- lm(weight~height, data = Davis)

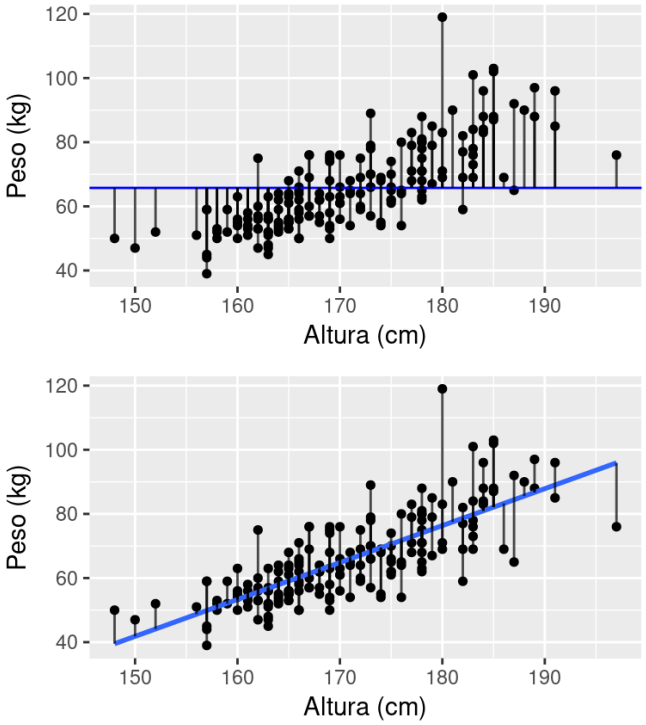
Modelo da Regressão

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

```
##
## Call:
## lm(formula = weight ~ height, data = Davis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.928  -5.406  -0.651   4.891  42.641
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -130.84185   12.30184  -10.64   <2e-16 ***
## height         1.15112    0.07193   16.00   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

5/70

Modelo da Regressão: peso ~ altura



anova(davisNull,lmdavis)

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
179	32367.75				
178	13272.71	1	19095.04	256.0832	0

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

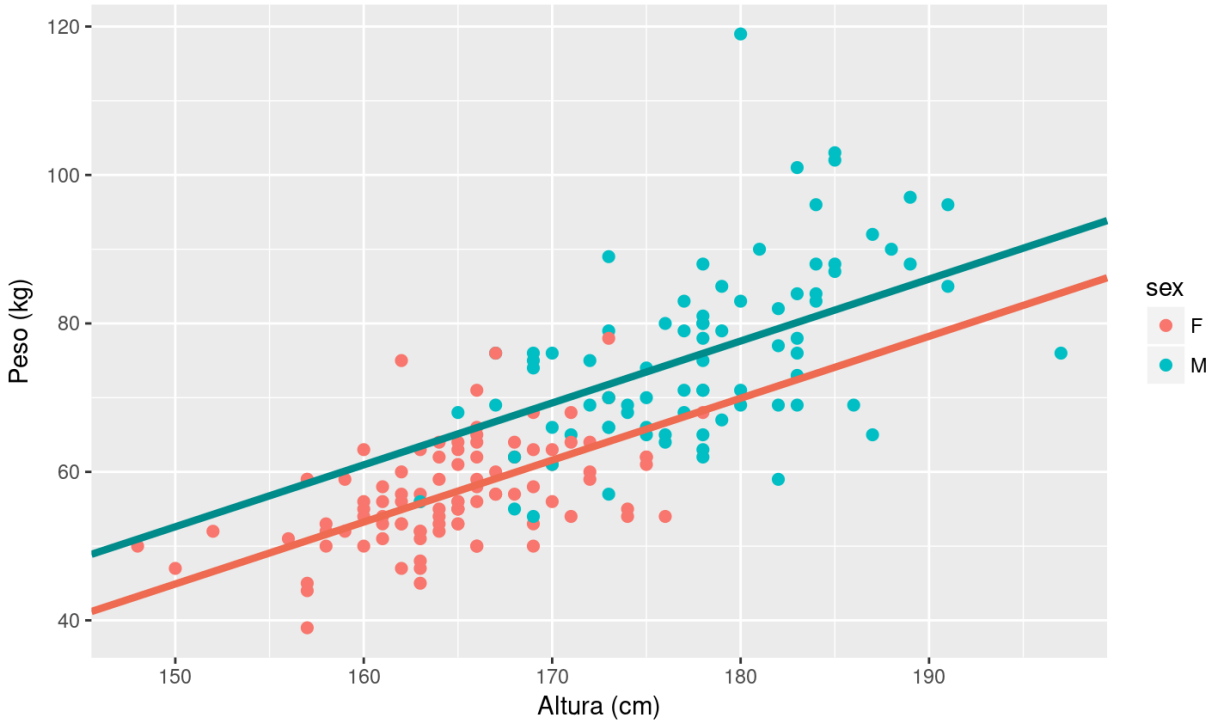
$$p_{valor} = 2.2 * 10^{-16}$$

$$r^2 = 0.587$$

Modelo de Regressão:

A relação entre peso e altura é diferente para os sexos

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



7/70

sexo: variável dummy com dois níveis (mulher = 0, homem = 1)

```
lmdavis01 <- lm(weight~ height + sex, data = Davis)
summary(lmdavis01)
```

```
##
## Call:
## lm(formula = weight ~ height + sex, data = Davis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.302  -4.808  -0.335   5.239  41.366
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -80.2107    16.8415  -4.763 3.96e-06 ***
## height        0.8341     0.1021   8.169 5.71e-14 ***
## sexM          7.7070     1.8345   4.201 4.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.258 on 177 degrees of freedom
## Multiple R-squared:  0.6271, Adjusted R-squared:  0.6229
```

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

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

Mulher ($sex = 0$)

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

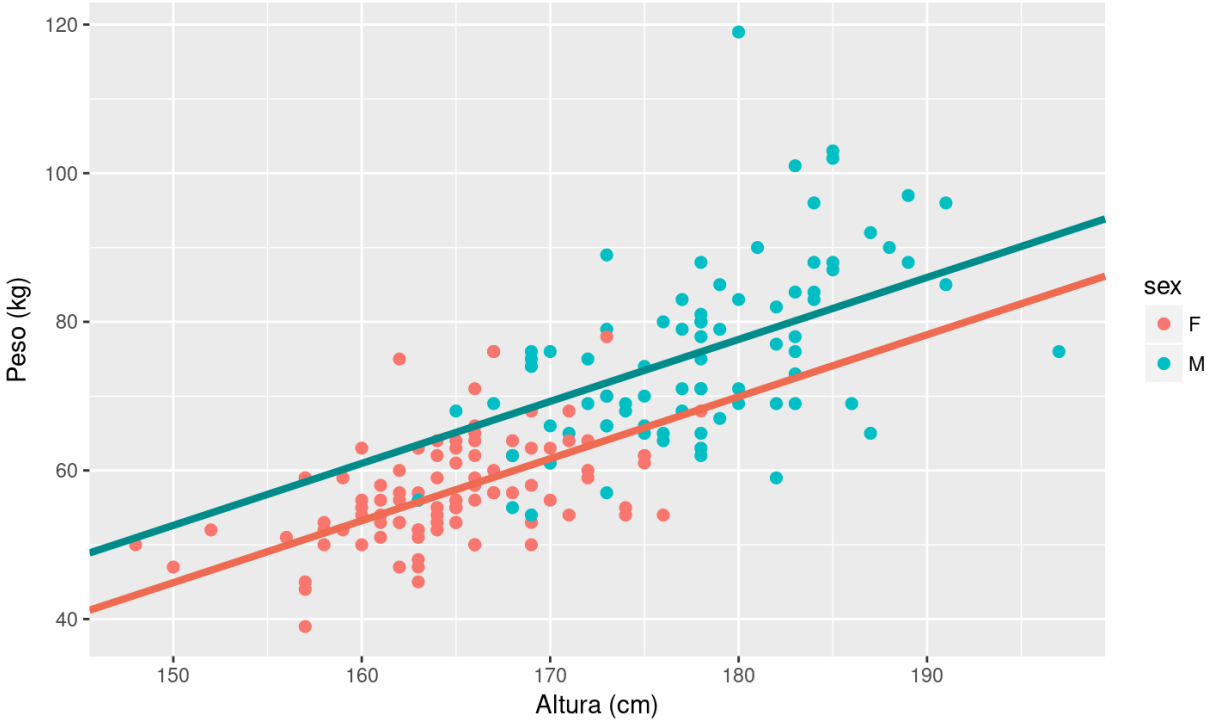
$$w_f = \hat{\alpha} + \hat{\beta}_h * height$$

Homem ($sex = 1$)

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

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

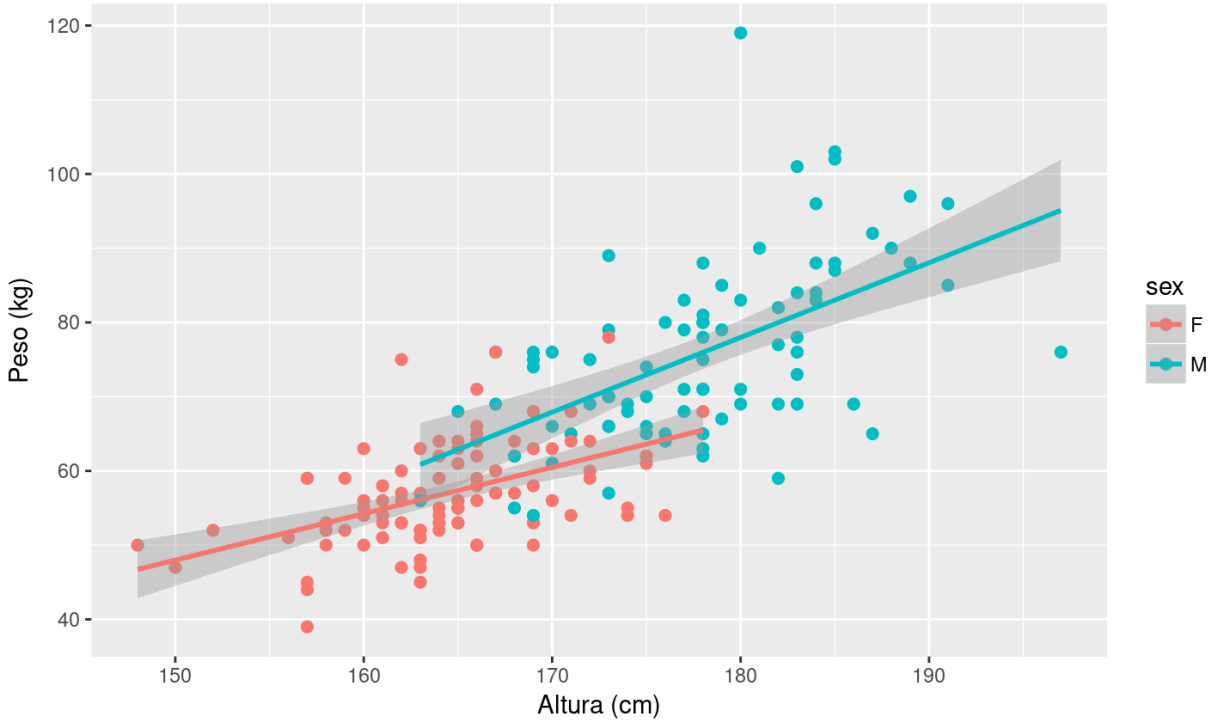
lm(weight ~ height + sex, data = Davis)



10/70

Interação

```
lmdavisfull <- lm(weight ~ height + sex + sex:height, data = Davis)
```



11/70

```
lmdavisfull <- lm(weight ~ height + sex*height, data=Davis)
summary(lmdavisfull)
```

```
##
## Call:
## lm(formula = weight ~ height + sex * height, data = Davis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.990  -4.548  -0.926   4.821  41.023
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -45.7988    24.8453  -1.843   0.0670 .
## height        0.6252     0.1507   4.148 5.22e-05 ***
## sexM        -57.4326    34.8293  -1.649   0.1009
## height:sexM   0.3815     0.2037   1.873   0.0628 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.2 on 176 degrees of freedom
## Multiple R-squared:  0.6344, Adjusted R-squared:  0.6282
## F-statistic: 101.8 on 3 and 176 DF,  p-value: < 2.2e-16
```

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

```
## (Intercept)      height        sexM height:sexM
## -45.7988220   0.6252035 -57.4326307   0.3815088
```

Mulher ($sex = 0$)

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

$$w_m = \hat{\alpha} + \hat{\beta}_h height$$

Homem ($sex = 1$)

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

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

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))
```

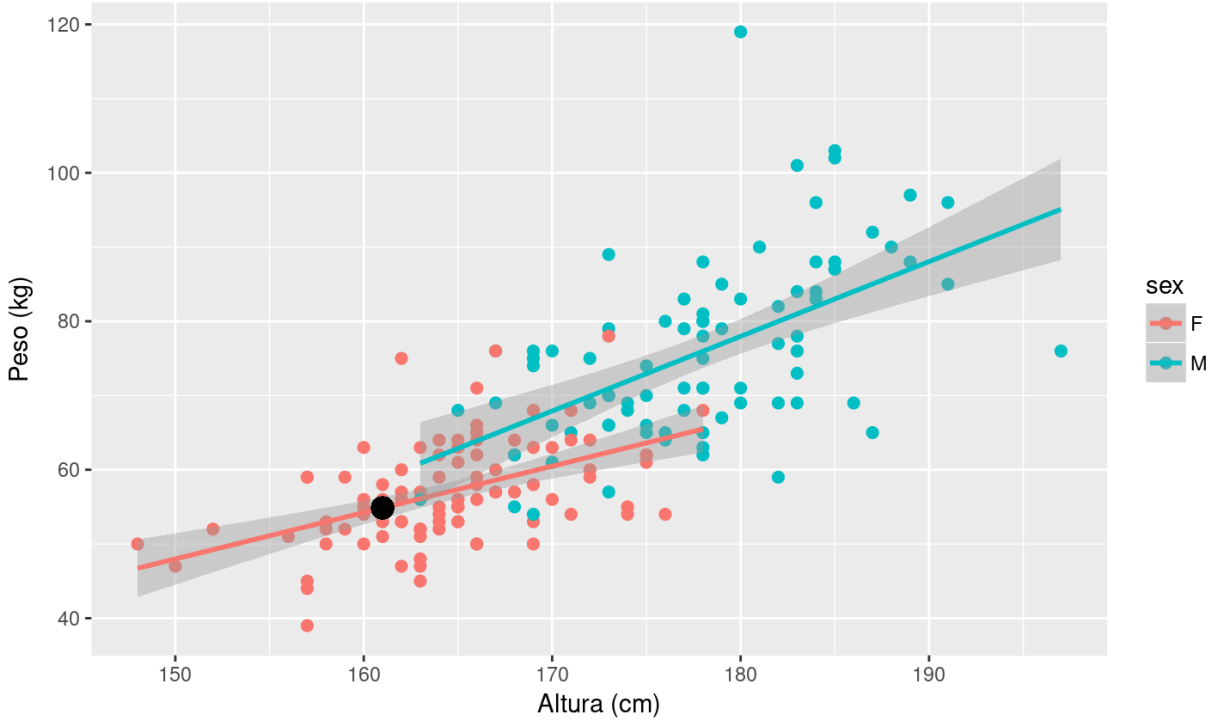
```
## (Intercept)      height        sexM height:sexM
## -45.7988220   0.6252035 -57.4326307   0.3815088
```

```
predMulher <- coefull[1] + coefull[2] * 161
(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
```

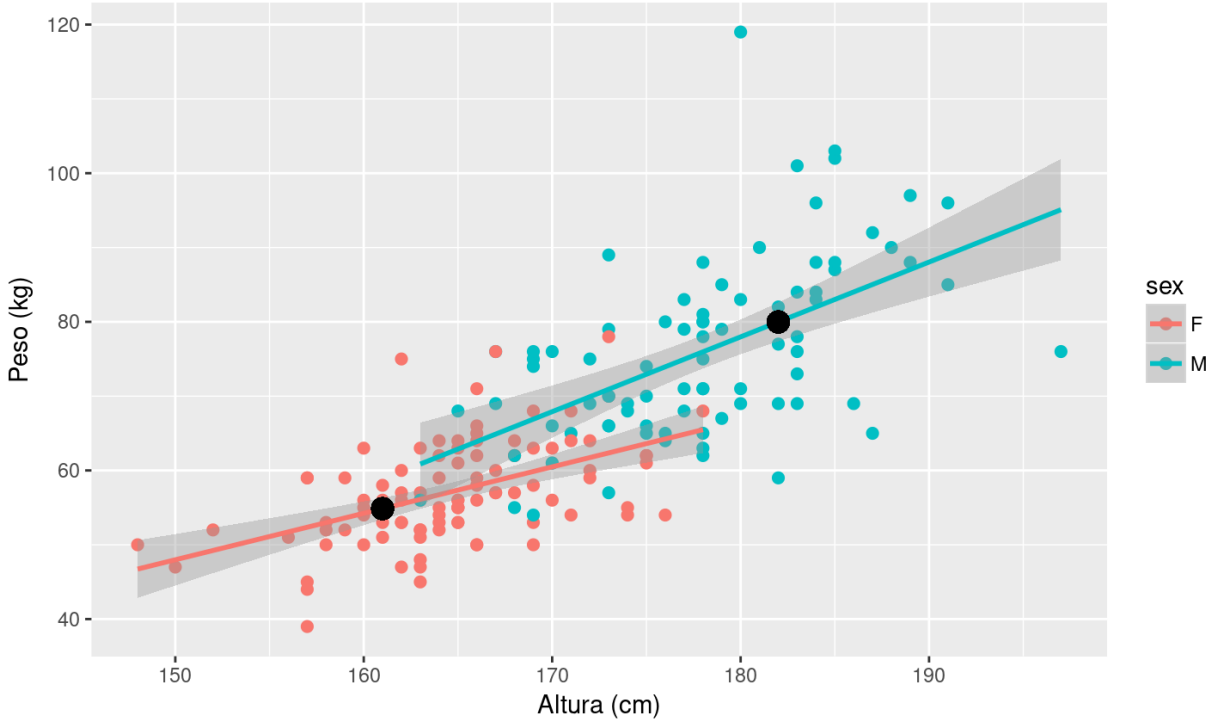
```
## (Intercept)      height        sexM height:sexM
## -45.7988220   0.6252035 -57.4326307   0.3815088
```

```
predHomem <- (coefull[1]+ coefull[3]) + (coefull[2]+ coefull[4]) * 182
(predHomem <- as.numeric(predHomem))
```

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

Predito pelo modelo

lm(weight ~ height + sex*height, data=Davis)* Um homem com 182cm de altura tem peso 79.99 kg.



17/70

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 height:sexM
## -45.7988220   0.6252035 -57.4326307   0.3815088
```

Matriz do Modelo

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

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

```
predHomem
```

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

```
predMulher
```

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

Qual o melhor modelo?

Princípio da parcimônia (Navalha de Occam)

- minimizar número de parâmetros
- 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 significativas (maior ordem)
   - termos quadráticos ou não lineares
   - variáveis explicativas não significativas
   - agrupe níveis de fatores sem diferença
   - ANCOVA: intercepto não significativo -> 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**

Modelo cheio

```
summary(lmdavisfull)
```

```
##
## Call:
## lm(formula = weight ~ height + sex * height, data = Davis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.990  -4.548  -0.926   4.821  41.023
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -45.7988    24.8453  -1.843   0.0670 .
## height        0.6252     0.1507   4.148 5.22e-05 ***
## sexM        -57.4326    34.8293  -1.649   0.1009
## height:sexM   0.3815     0.2037   1.873   0.0628 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.2 on 176 degrees of freedom
## Multiple R-squared:  0.6344,  Adjusted R-squared:  0.6282
```

Simplificando Modelo: exemplo

weight ~ height + sex + sex:height

weight ~ height + sex

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

```
## Analysis of Variance Table
##
## Model 1: weight ~ height + sex * height
## Model 2: weight ~ height + sex
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    176 11833
## 2    177 12069 -1   -235.82 3.5075 0.06275 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Simplificando Modelo: exemplo

```
summary(lmdavis01)
```

```
##
## Call:
## lm(formula = weight ~ height + sex, data = Davis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.302  -4.808  -0.335   5.239  41.366
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -80.2107    16.8415  -4.763 3.96e-06 ***
## height        0.8341     0.1021   8.169 5.71e-14 ***
## sexM          7.7070     1.8345   4.201 4.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.258 on 177 degrees of freedom
## Multiple R-squared:  0.6271, Adjusted R-squared:  0.6229
## F-statistic: 148.8 on 2 and 177 DF,  p-value: < 2.2e-16
```

Simplificando Modelo: exemplo

weight ~ height + sex

weight ~ height

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

```
## Analysis of Variance Table
##
## Model 1: weight ~ height + sex
## Model 2: weight ~ height
##   Res.Df   RSS Df Sum of Sq     F    Pr(>F)
## 1    177 12069
## 2    178 13273 -1   -1203.5 17.65 4.204e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelo Mínimo Adequado

```
summary(lmdavis01)
```

```
##
## Call:
## lm(formula = weight ~ height + sex, data = Davis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.302  -4.808  -0.335   5.239  41.366
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -80.2107    16.8415  -4.763 3.96e-06 ***
## height        0.8341     0.1021   8.169 5.71e-14 ***
## sexM          7.7070     1.8345   4.201 4.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.258 on 177 degrees of freedom
## Multiple R-squared:  0.6271, Adjusted R-squared:  0.6229
## F-statistic: 148.8 on 2 and 177 DF,  p-value: < 2.2e-16
```

Modelo Mínimo Adequado

```
anova(lmdavis01)
```

```
## Analysis of Variance Table
##
## Response: weight
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## height      1 19095.0 19095.0  280.04 < 2.2e-16 ***
## sex         1  1203.5  1203.5   17.65 4.204e-05 ***
## Residuals 177 12069.2    68.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelo Mínimo Adequado

```
coef(lmdavis01)
```

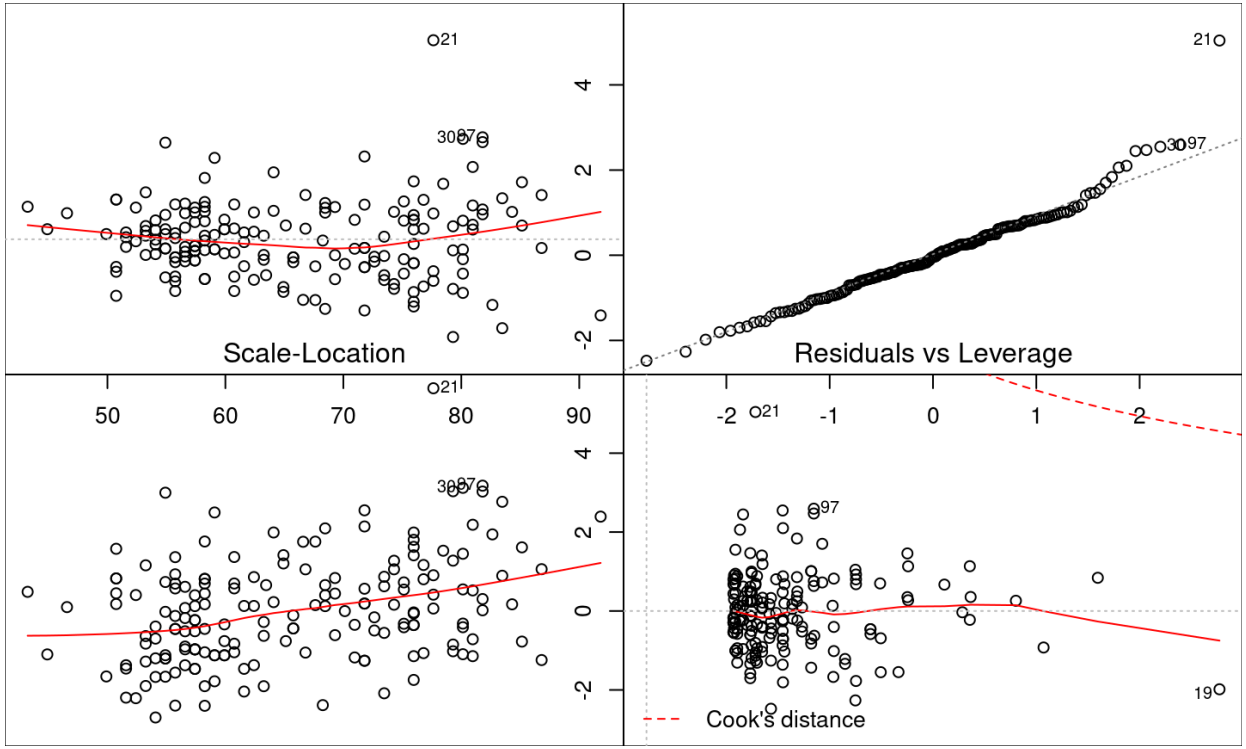
```
## (Intercept)      height        sexM
## -80.2107328   0.8340964   7.7070166
```

```
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)

```
oldpar <- par(mfrow=c(2,2), mar=c(0,0,0,0))
plot(lmdavis01)
```



30/70

```
par(oldpar)
```

Atividade

Modelo Linear Múltiplo:

- quais variáveis incluir
- curvatura em resposta a variável preditora
- interações entre variáveis
- correlação entre variáveis preditoras (colinearidade)
- saturação do modelo (complexidade)

Poluição: ozônio

Quais variáveis climáticas estão relacionadas à concentração de ozônio

```
ozo <- read.table("ozone.data.txt", header=TRUE)
str(ozo)
```

```
## 'data.frame':    111 obs. of  4 variables:
##  $ rad  : int  190 118 149 313 299 99 19 256 290 274 ...
##  $ temp : int  67 72 74 62 65 59 61 69 66 68 ...
##  $ wind : num  7.4 8 12.6 11.5 8.6 13.8 20.1 9.7 9.2 10.9 ...
##  $ ozone: int  41 36 12 18 23 19 8 16 11 14 ...
```

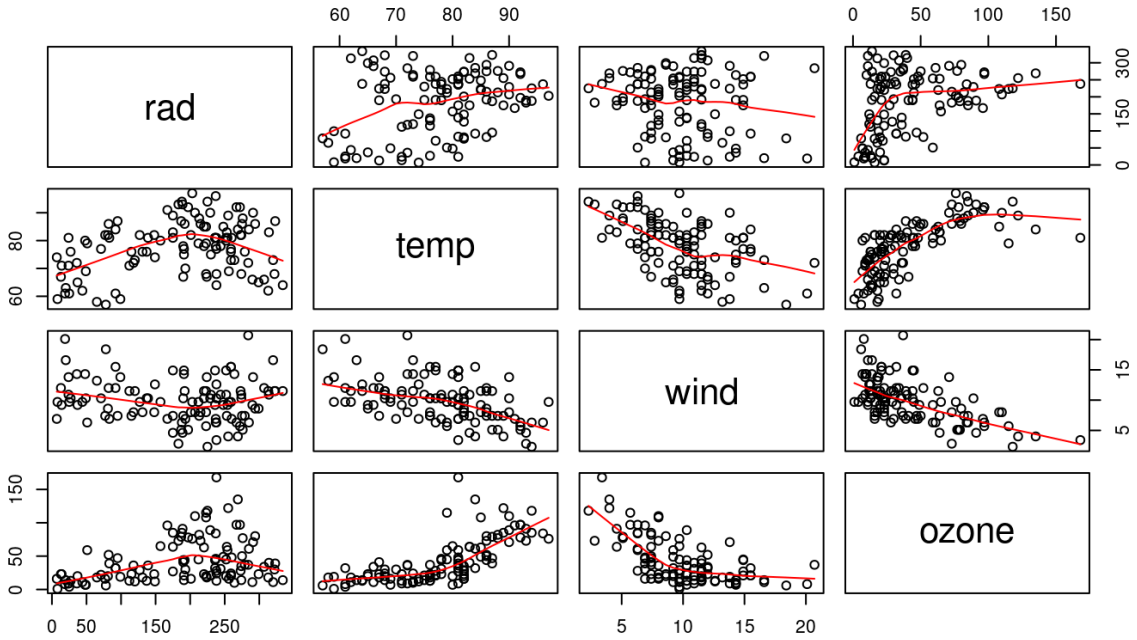
Ozônio data

```
tabozo <- data.frame(var = names(ozo), natureza = c("pred", "pred","pred","resposta"), tipo
kable(tabozo)
```

var	natureza	tipo	descrição
rad	pred	contínua	radiação
temp	pred	contínua	temperatura
wind	pred	contínua	vento
ozone	resposta	contínua	ozônio

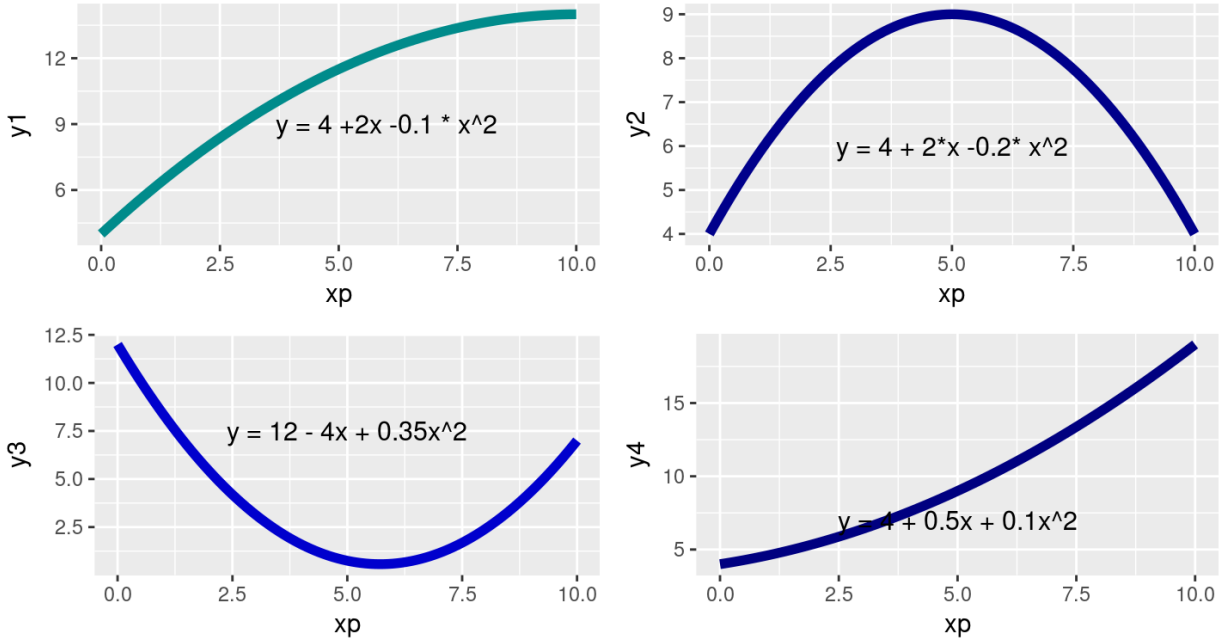
Linearidade

```
pairs(ozo, panel= panel.smooth)
```



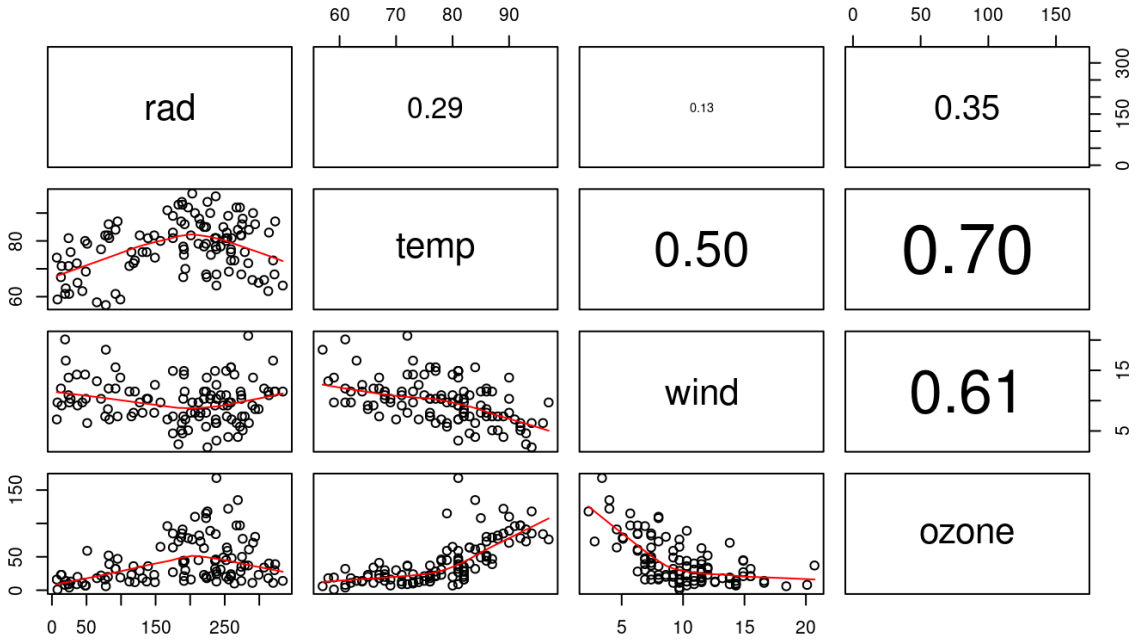
Curvatura da relação: polinômios

```
xp <- seq(0,10, 0.1)
y1 <- 4 + 2*xp -0.1* xp^2
y2 <- 4 + 2*xp -0.2* xp^2
y3 <- 12 - 4*xp + 0.35* xp^2
y4 <- 4 + 0.5 * xp + 0.1 * xp^2
```



36/70

Correlação entre preditoras



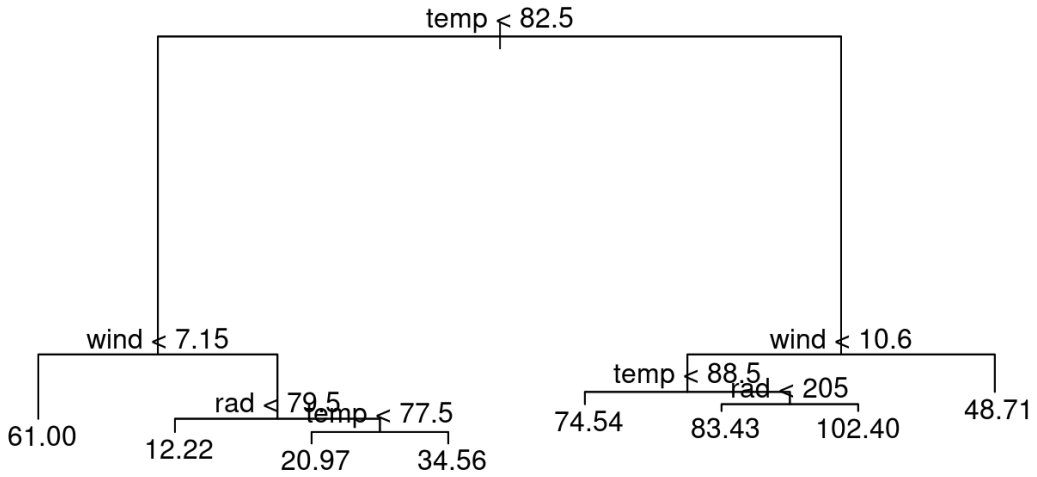
37/70

Correlação entre preditoras

- indíce de colinearidade (confirmar)
- reter apenas uma das variáveis colineares
- reduzir as dimensões das variáveis colineares (PCA)

Interação

```
mt <- tree(ozone ~., data = ozo)
par(mfrow= c(1,1))
plot(mt); text(mt)
```



Definir os termos do modelo cheio

Modelo para concentração de ozônio:

Modelo Cheio:

- temp
- wind
- rad
- temp^2
- wind^2
- rad^2
- temp:wind
- temp:rad
- wind:rad
- temp:wind:rad

Modelo Cheio: Ozônio

```
lmozfull <- lm(ozone ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) + I(rad^2) +
                 temp:wind + temp:rad + wind:rad +
                 temp:wind:rad, data = ozo)
lmozfull <- lm(ozone ~  temp * wind * rad +
                  I(temp^2) + I(wind^2) + I(rad^2), data = ozo)
```

```
##
## Call:
## lm(formula = ozone ~ temp * wind * rad + I(temp^2) + I(wind^2) +
##     I(rad^2), data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.894 -11.205  -2.736   8.809  70.551
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.683e+02  2.073e+02   2.741  0.00725 **
## temp          -1.076e+01  4.303e+00  -2.501  0.01401 *
## wind          -3.237e+01  1.173e+01  -2.760  0.00687 **
## rad           -3.117e-01  5.585e-01  -0.558  0.57799
## I(temp^2)      5.833e-02  2.396e-02   2.435  0.01668 *
## I(wind^2)      6.106e-01  1.469e-01   4.157 6.81e-05 ***
## I(rad^2)      -3.619e-04  2.573e-04  -1.407  0.16265
## temp:wind      2.377e-01  1.367e-01   1.739  0.08519 .
## temp:rad       8.403e-03  7.512e-03   1.119  0.26602
## wind:rad       2.054e-02  4.892e-02   0.420  0.67552
## temp:wind:rad -4.324e-04  6.595e-04  -0.656  0.51358
## ---
```

Simplificando o modelo: Ozônio

```
lmoz01 <- lm(ozone ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) + I(rad^2) +
                 temp:wind + temp:rad + wind:rad, data = ozo)
anova(lmozfull, lmoz01)
```

```
## Analysis of Variance Table
##
## Model 1: ozone ~ temp * wind * rad + I(temp^2) + I(wind^2) + I(rad^2)
## Model 2: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + temp:rad + wind:rad
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    100 31742
## 2    101 31879 -1   -136.44 0.4298 0.5136
```

```
##
## Call:
## lm(formula = ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) +
##     I(rad^2) + temp:wind + temp:rad + wind:rad, data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.611 -11.455  -2.901   8.548  70.325
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.245e+02  1.957e+02   2.680   0.0086 **
## temp        -1.021e+01  4.209e+00  -2.427   0.0170 *
## wind        -2.802e+01  9.645e+00  -2.906   0.0045 **
## rad          2.628e-02  2.142e-01   0.123   0.9026
## I(temp^2)    5.953e-02  2.382e-02   2.499   0.0141 *
## I(wind^2)    6.173e-01  1.461e-01   4.225 5.25e-05 ***
## I(rad^2)    -3.388e-04  2.541e-04  -1.333   0.1855
## temp:wind    1.734e-01  9.497e-02   1.825   0.0709 .
## temp:rad     3.750e-03  2.459e-03   1.525   0.1303
## wind:rad    -1.127e-02  6.277e-03  -1.795   0.0756 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
""
```

Simplificando o modelo: Ozônio

```
lmoz02 <- lm(ozone ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) + I(rad^2) +
                 temp:wind +  wind:rad, data = ozo)
anova(lmoz01, lmoz02)
```

```
## Analysis of Variance Table
##
## Model 1: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + temp:rad + wind:rad
## Model 2: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + wind:rad
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    101 31879
## 2    102 32613 -1   -734.23 2.3262 0.1303
```

```
##
## Call:
## lm(formula = ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) +
##     I(rad^2) + temp:wind + wind:rad, data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.040 -11.962  -2.863   9.661  70.475
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.488e+02  1.963e+02   2.796  0.00619 **
## temp        -1.144e+01  4.158e+00  -2.752  0.00702 **
## wind        -2.876e+01  9.695e+00  -2.967  0.00375 **
## rad          3.061e-01  1.113e-01   2.751  0.00704 **
## I(temp^2)    7.145e-02  2.265e-02   3.154  0.00211 **
## I(wind^2)    6.363e-01  1.465e-01   4.343 3.33e-05 ***
## I(rad^2)    -2.690e-04  2.516e-04  -1.069  0.28755
## temp:wind    1.840e-01  9.533e-02   1.930  0.05644 .
## wind:rad    -1.381e-02  6.090e-03  -2.268  0.02541 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Simplificando o modelo: Ozônio

```
lmoz03 <- lm(ozone ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) +
                 temp:wind +  wind:rad, data = ozo)
anova( lmoz02, lmoz03)
```

```
## Analysis of Variance Table
##
## Model 1: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + wind:rad
## Model 2: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + temp:wind +
##     wind:rad
##   Res.Df   RSS Df Sum of Sq     F Pr(>F)
## 1    102 32613
## 2    103 32978 -1   -365.45 1.143 0.2875
```

```
##
## Call:
## lm(formula = ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) +
##     temp:wind + wind:rad, data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.379 -11.375  -2.217   8.921  71.247
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 514.401470 193.783580   2.655  0.00920 **
## temp        -10.654041   4.094889  -2.602  0.01064 *
## wind        -27.391965   9.616998  -2.848  0.00531 **
## rad           0.212945   0.069283   3.074  0.00271 **
## I(temp^2)     0.067805   0.022408   3.026  0.00313 **
## I(wind^2)     0.619396   0.145773   4.249 4.72e-05 ***
## temp:wind     0.169674   0.094458   1.796  0.07538 .
## wind:rad     -0.013561   0.006089  -2.227  0.02813 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.89 on 103 degrees of freedom
```

Simplificando o modelo: Ozônio

```
lmoz04 <- lm(ozone ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) +
                  wind:rad, data = ozo)
anova( lmoz03, lmoz04)
```

```
## Analysis of Variance Table
##
## Model 1: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + temp:wind +
##     wind:rad
## Model 2: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + wind:rad
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    103 32978
## 2    104 34011 -1   -1033.1 3.2267 0.07538 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) +
##     wind:rad, data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.478 -10.735  -2.437   9.685  77.543
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 223.573855 107.618223   2.077 0.040221 *
## temp         -5.197139   2.775039  -1.873 0.063902 .
## wind        -10.816032   2.736757  -3.952 0.000141 ***
## rad           0.173431   0.066398   2.612 0.010333 *
## I(temp^2)     0.043640   0.018112   2.410 0.017731 *
## I(wind^2)     0.430059   0.101767   4.226 5.12e-05 ***
## wind:rad     -0.009819   0.005783  -1.698 0.092507 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.08 on 104 degrees of freedom
## Multiple R-squared:  0.7208, Adjusted R-squared:  0.7047
```

Simplificando o modelo: Ozônio

```
lmoz05 <- lm(ozone ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2),
                  data = ozo)
anova( lmoz04, lmoz05)
```

```
## Analysis of Variance Table
##
## Model 1: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2) + wind:rad
## Model 2: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2)
##   Res.Df   RSS Df Sum of Sq     F  Pr(>F)
## 1    104 34011
## 2    105 34954 -1   -942.85 2.883 0.09251 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = ozone ~ temp + wind + rad + I(temp^2) + I(wind^2),
##     data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.044 -10.796  -4.138   8.131  80.098
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 291.16758  100.87723   2.886  0.00473 **
## temp         -6.33955    2.71627  -2.334  0.02150 *
## wind        -13.39674    2.29623  -5.834 6.05e-08 ***
## rad           0.06586    0.02005   3.285  0.00139 **
## I(temp^2)     0.05102    0.01774   2.876  0.00488 **
## I(wind^2)     0.46464    0.10060   4.619 1.10e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.25 on 105 degrees of freedom
## Multiple R-squared:  0.713,  Adjusted R-squared:  0.6994
## F-statistic: 52.18 on 5 and 105 DF,  p-value: < 2.2e-16
```

52/70

Simplificando o modelo: Ozônio

```
lmoz06 <- lm(ozone ~ temp + wind + rad +
                  I(wind^2)
                  , data = ozo)
anova(lmoz05, lmoz06)
```

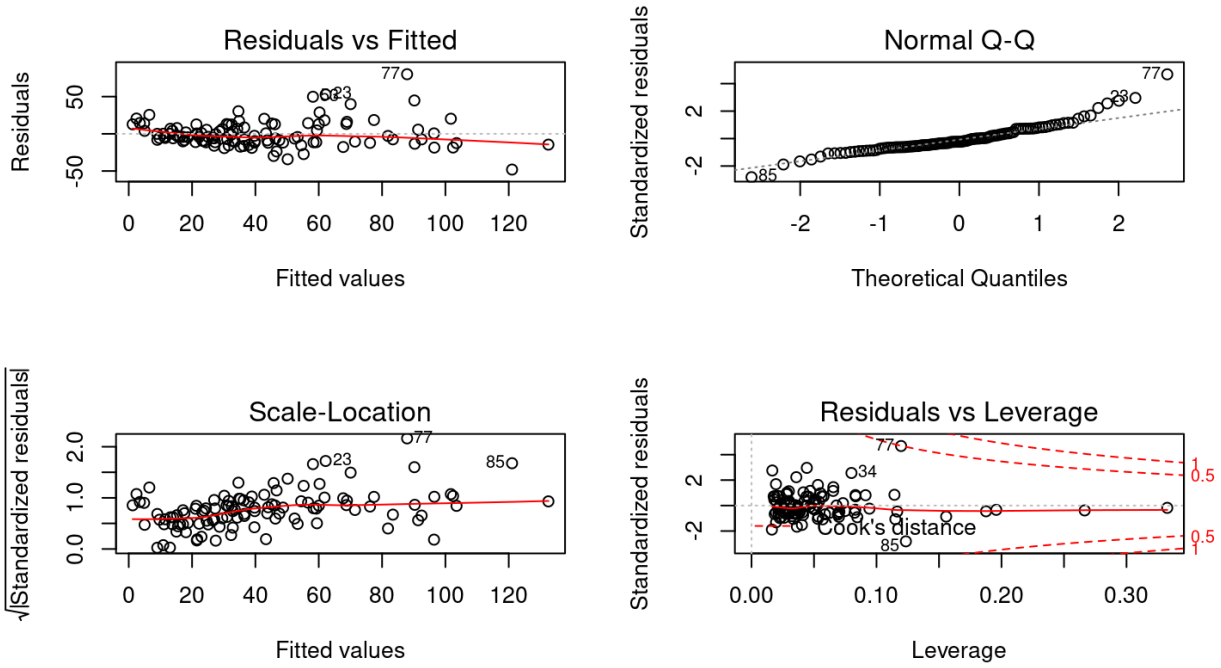
```
## Analysis of Variance Table
##
## Model 1: ozone ~ temp + wind + rad + I(temp^2) + I(wind^2)
## Model 2: ozone ~ temp + wind + rad + I(wind^2)
##   Res.Df   RSS Df Sum of Sq      F   Pr(>F)
## 1    105 34954
## 2    106 37708 -1   -2753.7 8.2718 0.004877 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelo Mínimo Adequado

```
##
## Call:
## lm(formula = ozone ~ temp + wind + rad + I(temp^2) + I(wind^2),
##     data = ozo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.044 -10.796  -4.138   8.131  80.098
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 291.16758  100.87723   2.886  0.00473 **
## temp         -6.33955    2.71627  -2.334  0.02150 *
## wind        -13.39674    2.29623  -5.834 6.05e-08 ***
## rad           0.06586    0.02005   3.285  0.00139 **
## I(temp^2)     0.05102    0.01774   2.876  0.00488 **
## I(wind^2)     0.46464    0.10060   4.619 1.10e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.25 on 105 degrees of freedom
```

DIAGNÓSTICO DO MODELO

```
par(mfrow=c(2,2))
plot(lmoz05)
```



Transformando variável

```
 lmoz07 <- lm(log(ozone) ~ temp + wind + rad +
    I(temp^2)+ I(wind^2) + I(rad^2) + temp:wind +
      temp:rad + wind:rad + temp:wind:rad, data = ozo)
summary(lmoz07)
```

```
##
## Call:
## lm(formula = log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) +
##     I(rad^2) + temp:wind + temp:rad + wind:rad + temp:wind:rad,
##     data = ozo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.91943 -0.24169 -0.01742  0.28213  1.11802
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.803e+00  5.676e+00   0.494   0.6225
## temp          -3.018e-02  1.178e-01  -0.256   0.7983
## wind          -9.812e-02  3.211e-01  -0.306   0.7605
## rad            2.771e-02  1.529e-02   1.812   0.0729
```

Simplificando o modelo

```
 lmoz08 <- lm(log(ozone) ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) + I(rad^2) +
                 temp:wind + temp:rad + wind:rad, data = ozo)
anova(lmoz07, lmoz08)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + temp:rad + wind:rad + temp:wind:rad
## Model 2: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + temp:rad + wind:rad
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    100 23.787
## 2    101 24.256 -1  -0.46883 1.9709 0.1634
```

Simplificando o modelo

```
 lmoz09 <- lm(log(ozone) ~ temp + wind + rad +
                  I(temp^2)+ I(wind^2) + I(rad^2) +
                 temp:wind +  wind:rad, data = ozo)
anova(lmoz08, lmoz09)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + temp:rad + wind:rad
## Model 2: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + wind:rad
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    101 24.256
## 2    102 24.281 -1  -0.02515 0.1047 0.7469
```

Simplificando o modelo

```
 lmoz10 <- lm(log(ozone) ~ temp + wind + rad +
                 I(temp^2)+ I(wind^2) + I(rad^2) +
                 wind:rad, data = ozo)
anova(lmoz09, lmoz10)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     temp:wind + wind:rad
## Model 2: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     wind:rad
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    102 24.281
## 2    103 24.401 -1  -0.11987 0.5035 0.4796
```

Simplificando o modelo

```
 lmoz11 <- lm(log(ozone) ~ temp + wind + rad +
          I(temp^2)+ I(wind^2) + I(rad^2), data = ozo)
anova(lmoz10, lmoz11)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2) +
##     wind:rad
## Model 2: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2)
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    103 24.401
## 2    104 24.522 -1  -0.12081 0.51 0.4768
```

Simplificando o modelo

```
 lmoz12 <- lm(log(ozone) ~ temp + wind + rad +
          I(wind^2) + I(rad^2), data = ozo)
anova(lmoz11, lmoz12)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(temp^2) + I(wind^2) + I(rad^2)
## Model 2: log(ozone) ~ temp + wind + rad + I(wind^2) + I(rad^2)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    104 24.522
## 2    105 24.707 -1  -0.18512 0.7851 0.3776
```

Simplificando o modelo

```
summary( lmoz12)
```

```
##
## Call:
## lm(formula = log(ozone) ~ temp + wind + rad + I(wind^2) + I(rad^2),
##     data = ozo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85551 -0.25578  0.00248  0.31349  1.16251
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.724e-01  6.350e-01   1.216 0.226543
## temp         4.193e-02  6.237e-03   6.723 9.52e-10 ***
## wind        -2.211e-01  5.874e-02  -3.765 0.000275 ***
## rad          7.466e-03  2.323e-03   3.215 0.001736 **
## I(wind^2)    7.390e-03  2.585e-03   2.859 0.005126 **
## I(rad^2)    -1.470e-05  6.734e-06  -2.183 0.031246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 0.1 ' ' 1
```

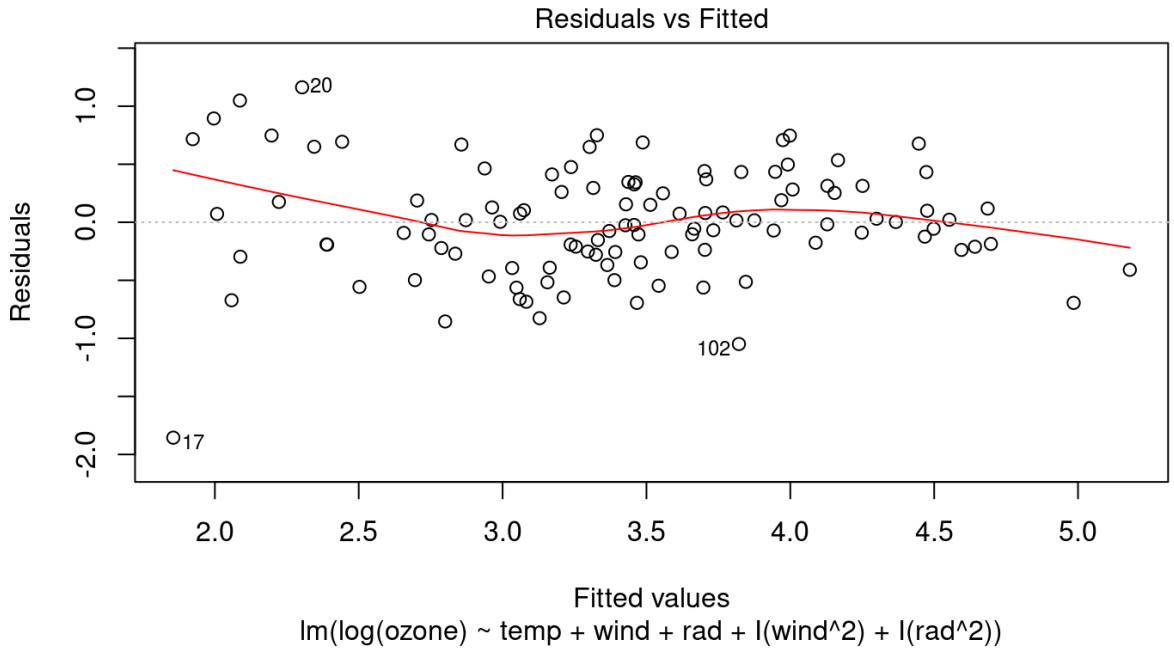
Simplificando o modelo

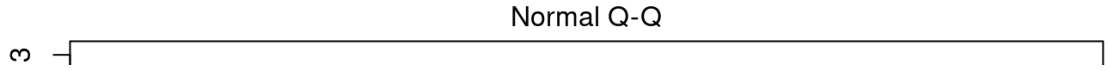
```
 lmoz13 <- lm(log(ozone) ~ temp + wind + rad +
          I(wind^2), data = ozo)
anova(lmoz12, lmoz13)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(wind^2) + I(rad^2)
## Model 2: log(ozone) ~ temp + wind + rad + I(wind^2)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    105 24.707
## 2    106 25.828 -1   -1.1216 4.7665 0.03125 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnóstico do Modelo

```
plot(lmoz12)
```





64/70

Modelo Mínimo Adequado

```
summary(lmoz12)
```

```
##
## Call:
## lm(formula = log(ozone) ~ temp + wind + rad + I(wind^2) + I(rad^2),
##     data = ozo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85551 -0.25578  0.00248  0.31349  1.16251
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.724e-01  6.350e-01   1.216 0.226543
## temp         4.193e-02  6.237e-03   6.723 9.52e-10 ***
## wind        -2.211e-01  5.874e-02  -3.765 0.000275 ***
## rad          7.466e-03  2.323e-03   3.215 0.001736 **
## I(wind^2)    7.390e-03  2.585e-03   2.859 0.005126 **
## I(rad^2)    -1.470e-05  6.734e-06  -2.183 0.031246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 0.1 ' ' 1
```

IMPORTÂNCIA DAS VARIÁVEIS

Escalas diferentes

```
lmoz12a <- lm(log(ozone) ~ I(temp/100) + wind + rad +
          I((wind/100)^2) + I(rad^2), data = ozo)
anova(lmoz12, lmoz12a)
```

```
## Analysis of Variance Table
##
## Model 1: log(ozone) ~ temp + wind + rad + I(wind^2) + I(rad^2)
## Model 2: log(ozone) ~ I(temp/100) + wind + rad + I((wind/100)^2) + I(rad^2)
##   Res.Df    RSS Df  Sum of Sq F Pr(>F)
## 1    105 24.707
## 2    105 24.707  0 1.0658e-14
```

IMPORTÂNCIA DAS VARIÁVEIS

Escalas diferentes: problema

```
coef(lmoz12)
```

```
##   (Intercept)          temp          wind           rad     I(wind^2)
##  7.723892e-01  4.193355e-02 -2.211428e-01  7.465764e-03  7.390204e-03
##      I(rad^2)
## -1.470231e-05
```

```
coef(lmoz12a)
```

```
##     (Intercept)     I(temp/100)            wind             rad
##    7.723892e-01    4.193355e+00   -2.211428e-01    7.465764e-03
## I((wind/100)^2)        I(rad^2)
##    7.390204e+01   -1.470231e-05
```

Rescalonando os coeficientes:

```
tempR <- scale(ozo$temp)
windR <- scale(ozo$wind)
radR <- scale (ozo$rad)
## modelo rescalonado
lmoz12R <- lm(log(ozone) ~ tempR + windR + radR + I(radR^2)+ I(windR^2), data = ozo)
```

Modelo rescalonado

```
##
## Call:
## lm(formula = log(ozone) ~ tempR + windR + radR + I(radR^2) +
##     I(windR^2), data = ozo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85551 -0.25578  0.00248  0.31349  1.16251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.44421    0.07836  43.951  < 2e-16 ***
## tempR        0.39963    0.05944   6.723 9.52e-10 ***
## windR       -0.26425    0.05688  -4.646 9.86e-06 ***
## radR         0.18520    0.05268   3.516 0.000649 ***
## I(radR^2)   -0.12216    0.05595  -2.183 0.031246 *
## I(windR^2)   0.09362    0.03274   2.859 0.005126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4851 on 105 degrees of freedom
```

69/70

Atividade