

# Introdução aos Modelos Lineares em Ecologia

Prof. Adriano Sanches Melo - Dep. Ecologia – UFG  
asm.adrimelo no gmail.com

Página do curso: [www.ecologia.ufrgs.br/~adrimelo/lm/](http://www.ecologia.ufrgs.br/~adrimelo/lm/)

Livro-texto: Crawley, M.J. 2005. Statistics: An Introduction using R.  
John Wiley & Sons.

Página do livro na internet:

<http://www3.imperial.ac.uk/naturalsciences/research/statisticsusingr>

# AULA 07

## 1. Introdução

### Regressão Múltipla:

Variável resposta contínua

Várias variáveis preditoras contínuas

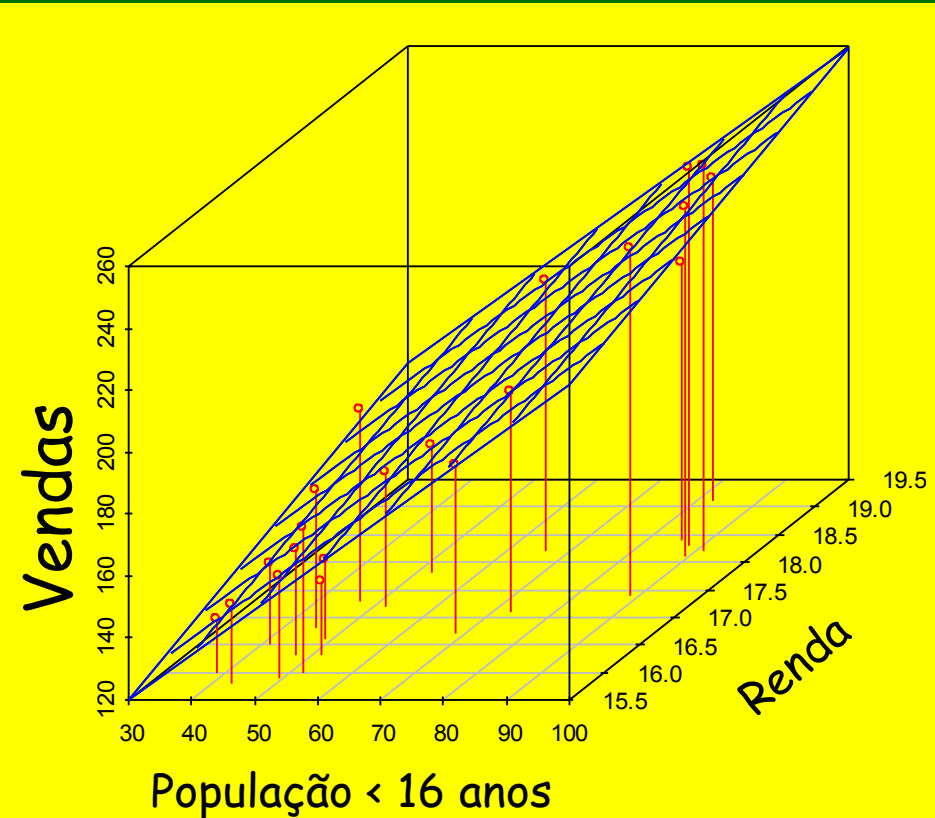
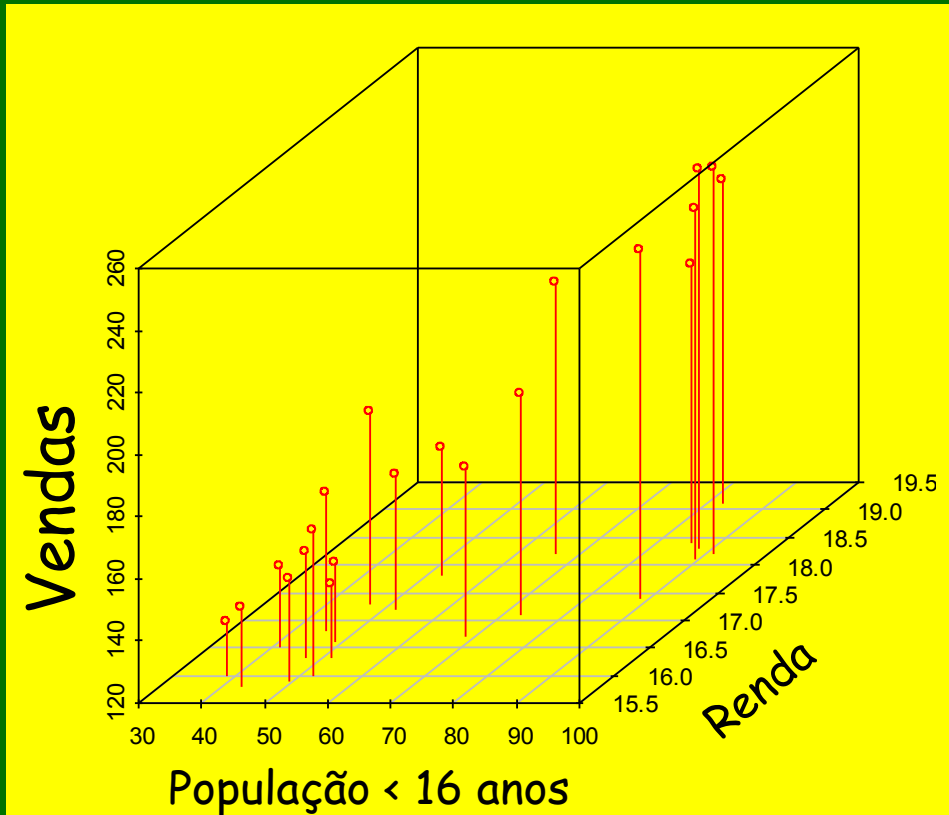
Exemplo:

$$Y = a + X_1 + X_2 + X_1 \cdot X_2$$

$Y =$  Vendas

$X_1 =$  População < 16 anos

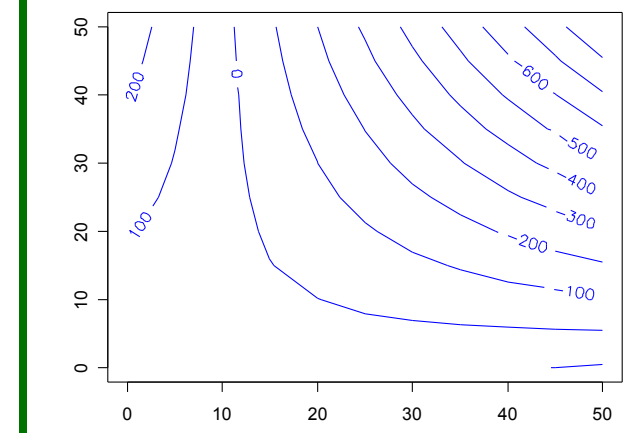
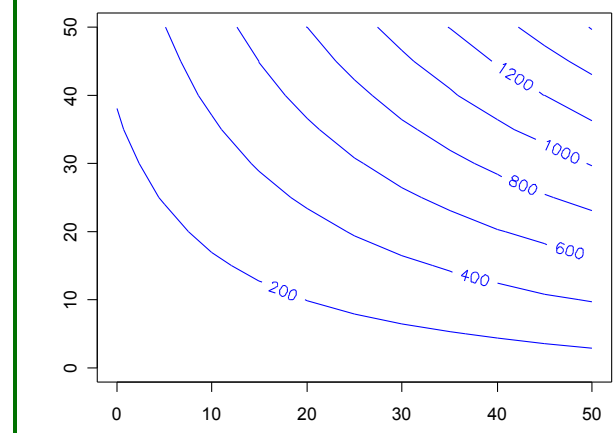
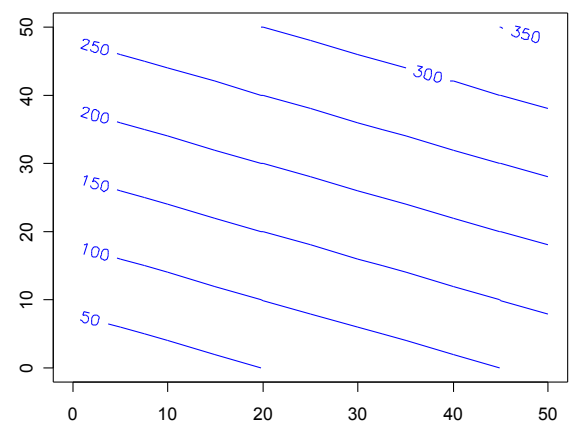
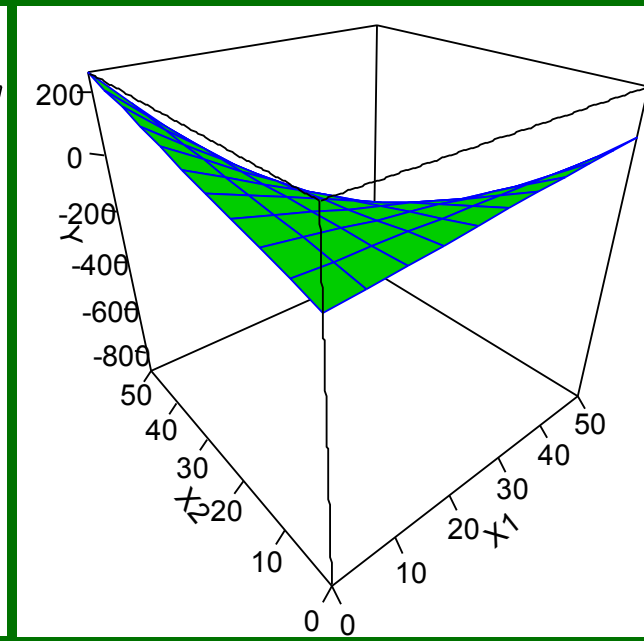
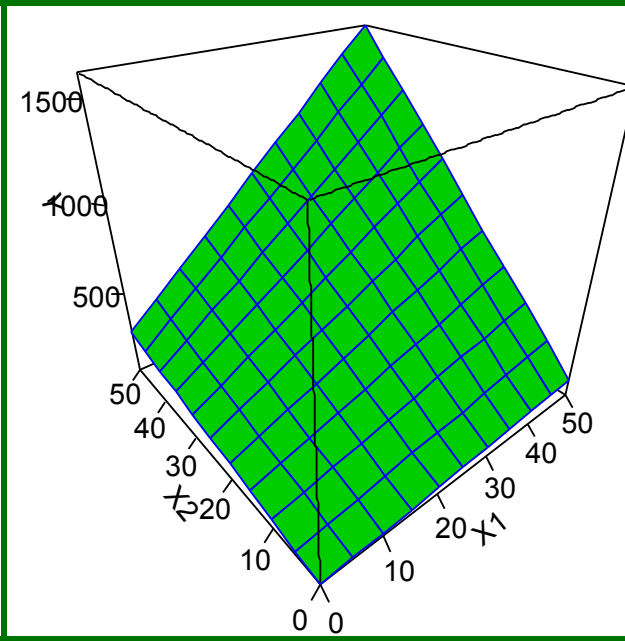
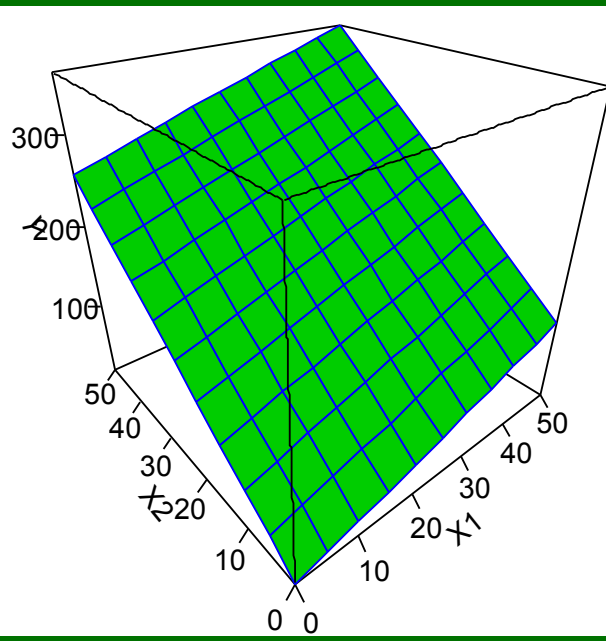
$X_2 =$  Renda



Modelo sem interação  
 $Y=10 + 2X_1 + 5X_2$

Com interação-sinérgica  
 $Y=10 + 2X_1 + 5X_2 + 0.5X_1X_2$

Com interação-interferência  
 $Y=10 + 2X_1 + 5X_2 - 0.5X_1X_2$



Para as figuras do slide anterior:

```
x1<-seq(0,50,5)
```

```
x2<-x1
```

```
x12<-expand.grid(x1, x2)
```

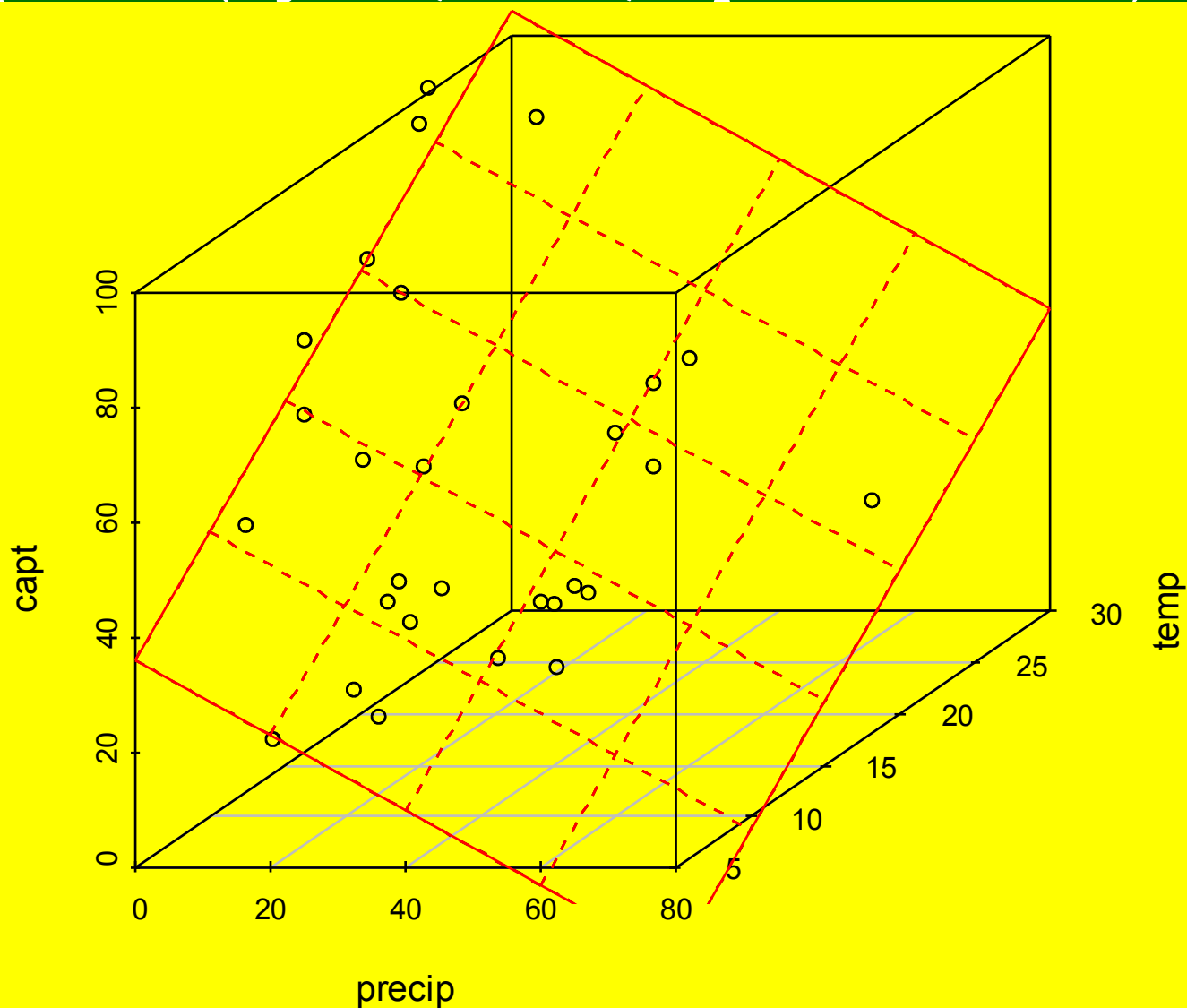
```
aditivo<-10 + 2*x12[,1] + 5*x12[,2]
```

```
aditivo.m<-matrix(aditivo,11,11)
```

```
persp(seq(0,50,5),seq(0,50,5),z=aditivo.m,theta=-40, phi=30,  
col=3,xlab="X1", ylab="X2", zlab="Y", ticktype="detailed",  
border=4) #Figura 3D
```

```
contour(seq(0,50,5), seq(0,50,5), z=aditivo.m, col=4, labcex=1)  
#Figura Relevo
```

```
library(scatterplot3d)
ajuste<-lm(capt~precip+temp)
resu<-scatterplot3d(x=precip,y=temp,z=capt)
resu$plane3d(ajuste,col=2,lty.box="solid")
```



## 2. Comentários gerais

- Amplo uso em Ecologia: Exemplos
- Talvez uma das análises mais subjetivas... (ver Cap. 11 Crawley)
- Pergunta que se faz: Vale a pena incluir a variável?
- Os parâmetros ( $\beta_1$  e  $\beta_2$ ) são chamados coeficientes parciais de regressão, visto que refletem o efeito de uma variável preditora quando as outras já estão no modelo e são mantidas no mesmo nível (mesmo valor de X).
- Um dos maiores problemas no uso:  
Correlação entre preditores (Multicolinearidade)

### 3. Multicolinearidade e seus efeitos

*Situação em que não exista correlação entre variáveis preditoras*

Quando  $r_{12} = 0$ , os coeficientes  $b_1$  e  $b_2$  são os mesmos não importa se ajustados de forma separada (regressão com 1 X) ou conjunta (regressão com 2 Xs).

Ainda, o extra SS obtido com  $X_2$  quando  $X_1$  já está no modelo é igual ao obtido com um modelo de regressão simples contendo apenas  $X_2$

### 3. Multicolinearidade e seus efeitos

#### *Efeitos da Multicolinearidade*

- a) Quando existe multi-colinearidade, pode-se encontrar a situação em que nenhum coeficiente é significativo no modelo, embora exista de fato uma relação. Isto pois os coeficientes são estimados com baixa precisão. Usando apenas um grupo reduzido de variáveis independentes pode revelar coeficientes de fato significativos.
- b) A interpretação de que o aumento de uma unidade em  $X_1$  quando  $X_2$  é mantido constante não é válida; ou melhor ainda, nem se pode fazer. Um exemplo: não se pode prever quanto uma cultura produzirá a mais aumentando uma hora de chuva SEM diminuir uma hora de sol.



### 3. Multicolinearidade e seus efeitos

#### *Efeitos no Extra SS*

Quando duas variáveis são correlacionadas, a entrada da segunda será responsável por apenas um pequeno valor de SS. O mesmo para  $R^2$ .

#### *Efeitos sobre $s\{b_k\}$ :*

Aumento da variabilidade do coeficiente.

### 3. Multicolinearidade e seus efeitos

*Comentário (extraído de Kutner et al. 2004)*

Another illustration of how intercorrelated variables that are omitted from model can influence the regression coefficients in the model is provided by an analyst who was perplexed about the sign of a regression coefficient in the fitted regression model. The analyst had found in a regression of territory company sales on territory population size, per capita income, and some other predictor variable that the regression coefficient for population size was negative, and this conclusion was supported by a confidence interval for the regression coefficient. A consultant noted that the analyst did not include the major competitor's market penetration as a predictor variable in the model. The competitor was most active and effective in territories with large populations, thereby keeping company sales down in these territories. The result of the omission of this predictor variable from the model was a negative coefficient for the population size variable.

## 4. Seleção de Modelo

-- Manual

-- Automatico

**forward**

inicia-se com 1 variável e vai incorporando a cada passo outras variáveis

**backward elimination**

inicia-se o processo com o modelo completo (saturado) e então tenta-se simplificá-lo a cada passo.

-- Procedimento automático no R:

```
step(modelo)
```

## 5. Critérios para remoção/inclusão de variável

### a) F parcial

Calcula-se o modelo inteiro, com todas variáveis (Full model)

Calcula-se o modelo sem uma das variáveis (Reduced model)

Testa-se com F:

$$F = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

O teste acima é dito ser parcial. Testa-se apenas uma variável. O nome “parcial” distingue de “geral ou total” quando todas as variáveis são testadas em conjunto.

## 5. Critérios para remoção/inclusão de variável

### b) $R^2_p$ ou $SSE_p$

O subscrito “ $p$ ” no símbolo do Coeficiente de Determinação indica o número de parâmetros no modelo.

Visto que  $SSTO$  é constante para todos modelos,  $R^2$  varia inversamente com  $SSE_p$ . Visto que  $R^2$  sempre aumentará ou ficará igual conforme incluimos mais parâmetros, a idéia de usar  $R^2$  é identificar o ponto em que a adição de um novo parâmetro não aumente sensivelmente o valor de  $R^2$ .

Bons modelos possuem altos valores de  $R^2_p$ .

Veja exemplo de uso em Hawkins et al. (2003. Ecology 84:1608-23), disponível na página da disciplina.

## 5. Critérios para remoção/inclusão de variável

### c) $R^2_{ap}$ ou $MSE_p$

Visto que  $R^2_p$  não leva em consideração o número de parâmetros, uma alternativa seria o uso de  $R^2_{ap}$  ou seja do Coeficiente de Determinação Ajustado. Bons modelos possuem altos valores de  $R^2_{an}$

$$R^2_{a,p} = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left( \frac{n-1}{n-p} \right) \frac{SSE}{SSTO}$$

$n$  = número observações  
 $p$  = número de termos  
(variáveis  $X + 1$ ,  
pois inclui constante)  
no modelo

O coeficiente ajustado de determinação múltipla pode ficar menor quando acrescentamos uma variável  $X$  ao modelo já existente, visto que cada decréscimo em  $SSE$  pode ser contra-balanceado pela perda de 1 grau de liberdade no denominador “ $n-p$ ”.

## 5. Critérios para remoção/inclusão de variável

### d) AIC (Akaike Information Criterion)

$$\text{AIC} = -2 \cdot \log\text{-likelihood} + 2p$$

- É uma Deviance ( $-2 \cdot \log\text{-likelihood}$ ) penalizada ( $+2p$ )
- Bons modelos possuem baixos valores de AIC
- É o critério usado na função `step( )` do R.

## 6. Modelos polinomiais de regressão

Usos de modelos polinomiais

- 1) Quando a função de resposta é verdadeiramente uma função polinomial.
- 2) Quando a função resposta é desconhecida ou complexa, mas uma função polinomial é uma boa aproximação.

Um perigo no uso de regressão polinomial é que sua extrapolação pode ser imprevisível. Modelos polinomiais podem oferecer bons ajustes para os dados em mãos, mas podem assumir direções não-esperadas quando extrapolados além do intervalo observado.

$$Y_i = a + b_1 x_i + b_2 x_i^2$$

$$\text{onde, } x_i = X_i - \bar{X}$$

A razão de usar a variável preditora centrada é que  $X$  and  $X^2$  frequentemente serão altamente correlacionados. Centrando a variável preditora frequentemente reduz a multicolinearidade.



## 6.1. Comentários gerais

### Implementação de modelos de regressão polinomiais

#### *Ajuste de Modelos Polinomiais*

`lm(y ~ x + I(x^2))` # A função `I()` faz com que o sinal expoente “^” de fato tenha efeito dentro de uma formula.

*Abordagem Hierárquica de Ajuste:* Com uma abordagem hierárquica, se um termo polinomial é retido no modelo, então todos os termos de ordens inferiores também serão retidos. O mesmo para interações: se mantém a interação, deve-se manter os termos de primeira potência.

## Exercícios e estudo individual:

- Lista em sala de aula
- Crawley: Cap. 11
- Gotelli e Ellison: pp. 275-279  
282-287