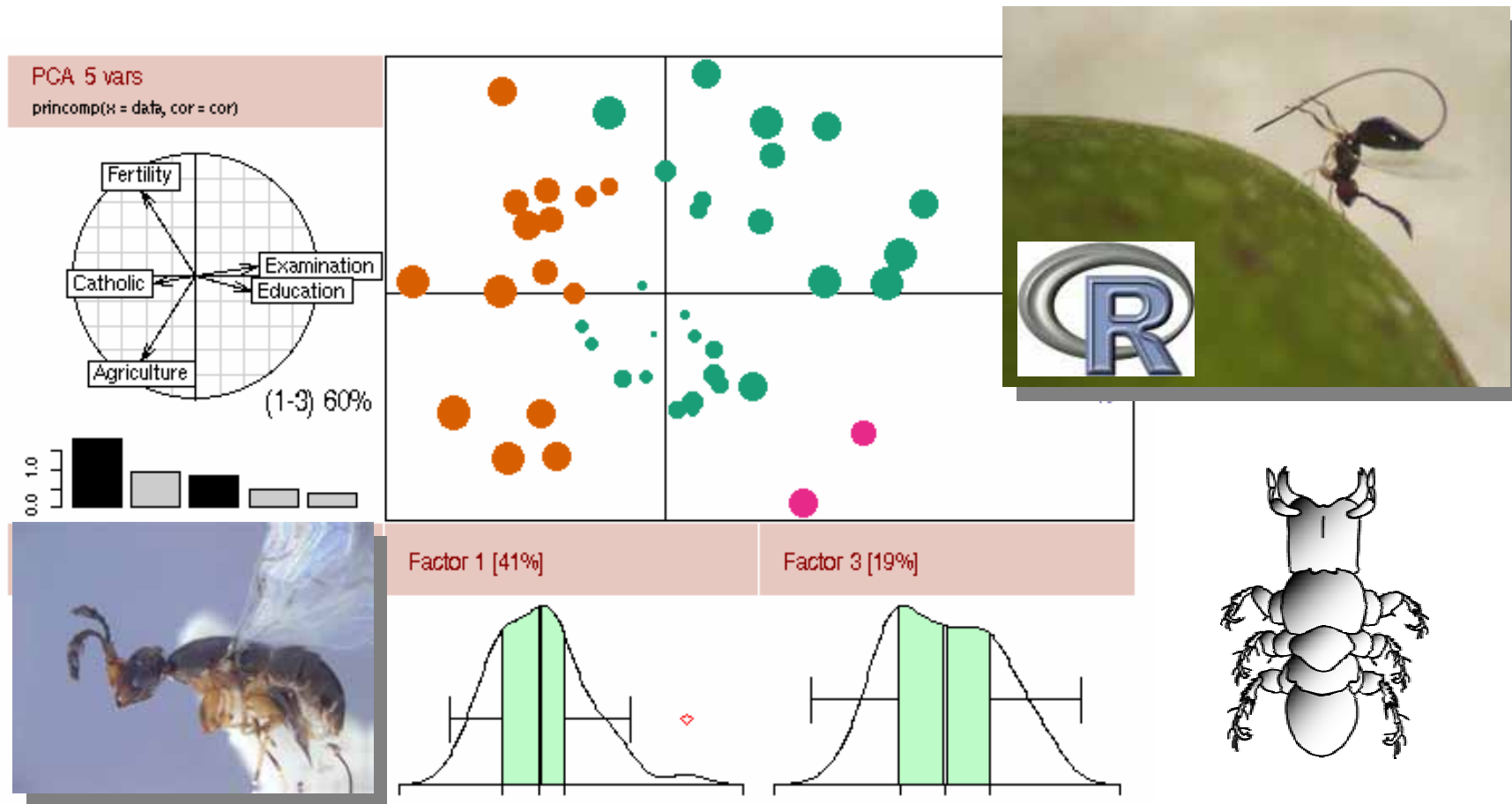


Uso da linguagem R para análise de dados em ecologia



Objetivo da aula

Apresentar os princípios básicos de modelagem numérica em Biologia.

Modelagem numérica em Biologia

✓ O que é um **modelo**?

➤ Uma **representação**, normalmente **simplificada**, da natureza (ou de um fenômeno natural);

✓ Por quê usar modelos?

➤ Para representar (**explicar**) de **forma mais simples** o fenômeno estudado;

✓ Vocês já usam modelagem matemática em seus trabalhos?

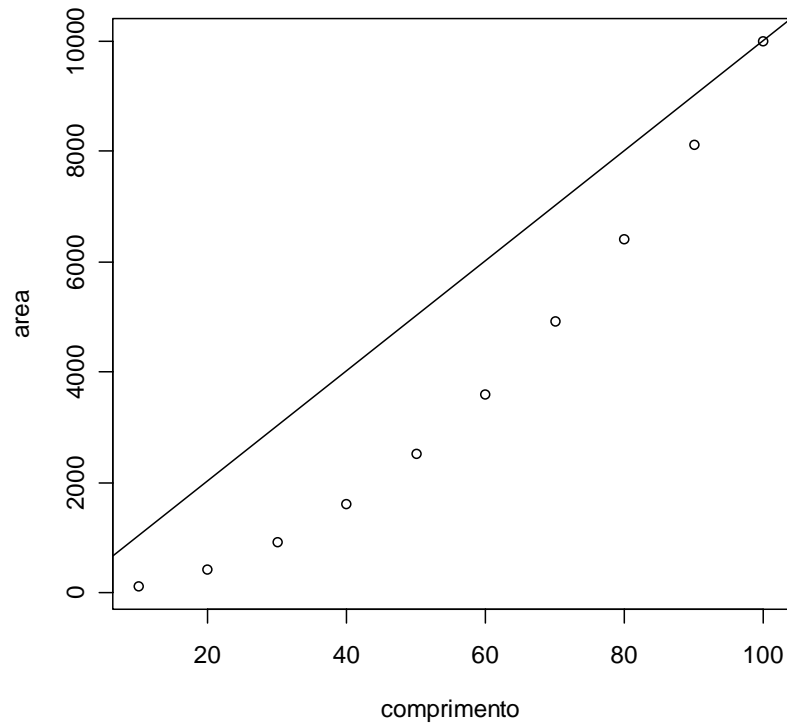
✓ Quando alguém usa ANOVA para comparar três grupos (comparar a mortalidade de um dado inseto após a aplicação de três inseticidas diferentes), ela está testando algum modelo? **Qual**?

✓ Quando eu digo que o peso dos alunos dessa turma é correlacionado a suas alturas, estou me referindo a algum modelo? **Qual**?

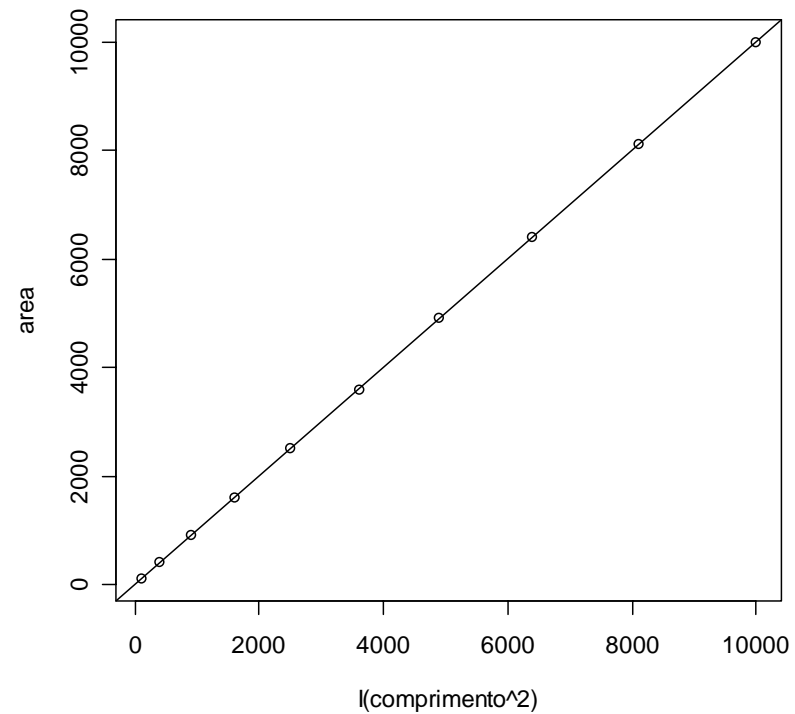
✓ Esse modelo é **linear**? Porque?

Modelagem numérica em Biologia

✓ E se eu disser que a **área de um quadrado** é correlacionado ao **comprimento de seu lado**, essa relação é linear?



Comprimento² vs. Área?



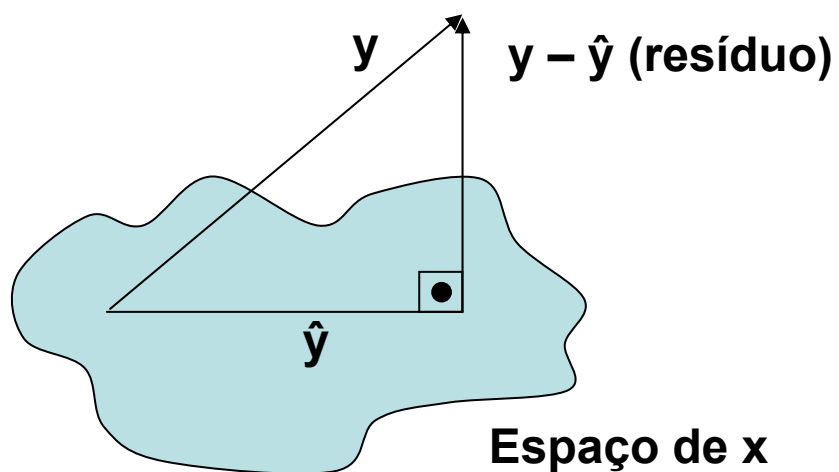
Modelos lineares

$$y \sim \alpha + \beta x + \varepsilon \quad \text{ou,}$$

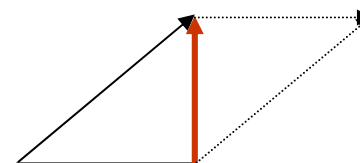
$$y \sim a + bx + \text{erro (estimativas de } \alpha \text{ e } \beta)$$

a = intercepto

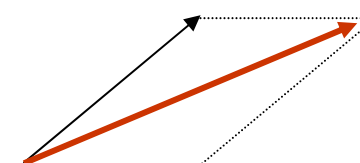
b = coeficiente angular (inclinação)



Subtração de vetores

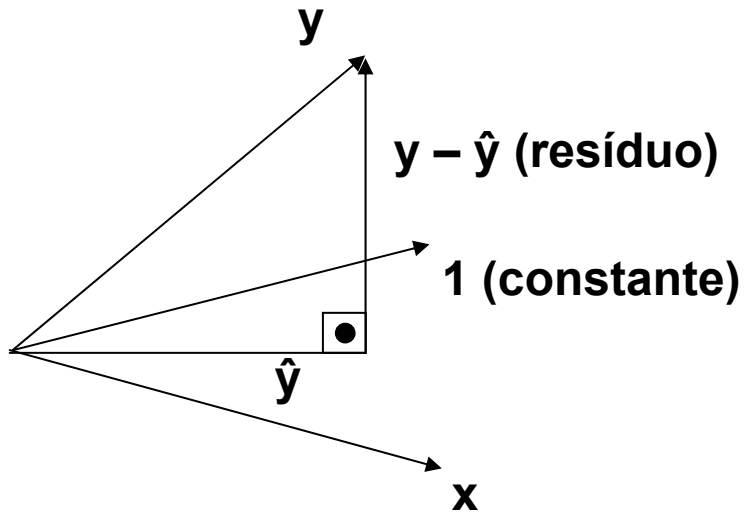


Soma de vetores



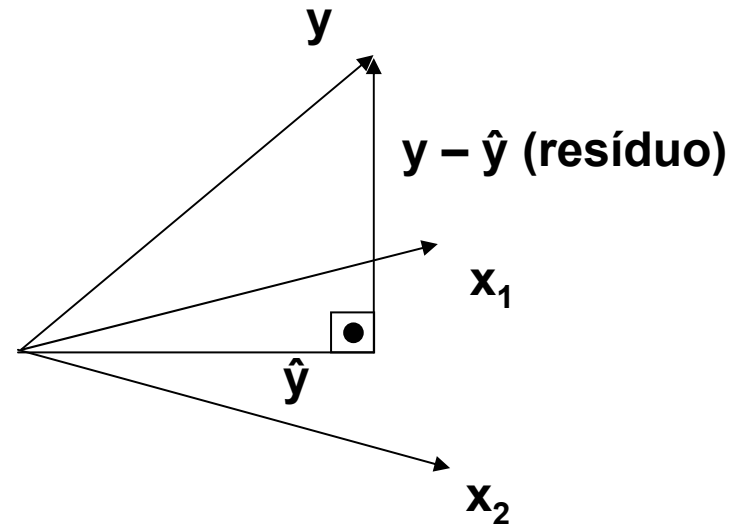
Modelos lineares

$$y \sim \alpha + \beta x + \varepsilon$$



\hat{y} = projeção de y no subespaço delimitado pelos vetores x e 1

$$y \sim \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$



\hat{y} = projeção de y no subespaço delimitado pelos vetores x_1 e x_2

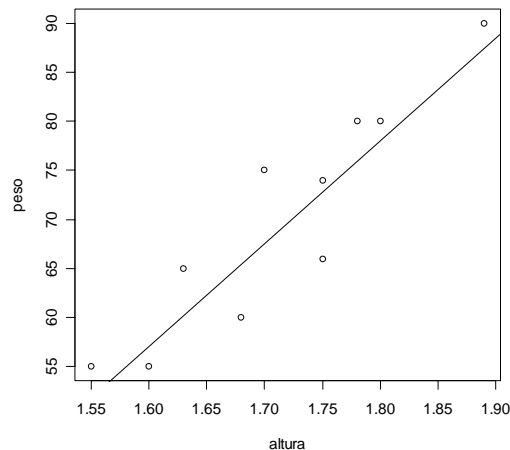
Modelos lineares

Representação matricial

$$\mathbf{y} \sim \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

- ✓ Como estimar α e β ?
- ✓ Minimizar a soma dos quadrados dos resíduos



Modelos lineares

Representação matricial

$$Y \sim X b$$

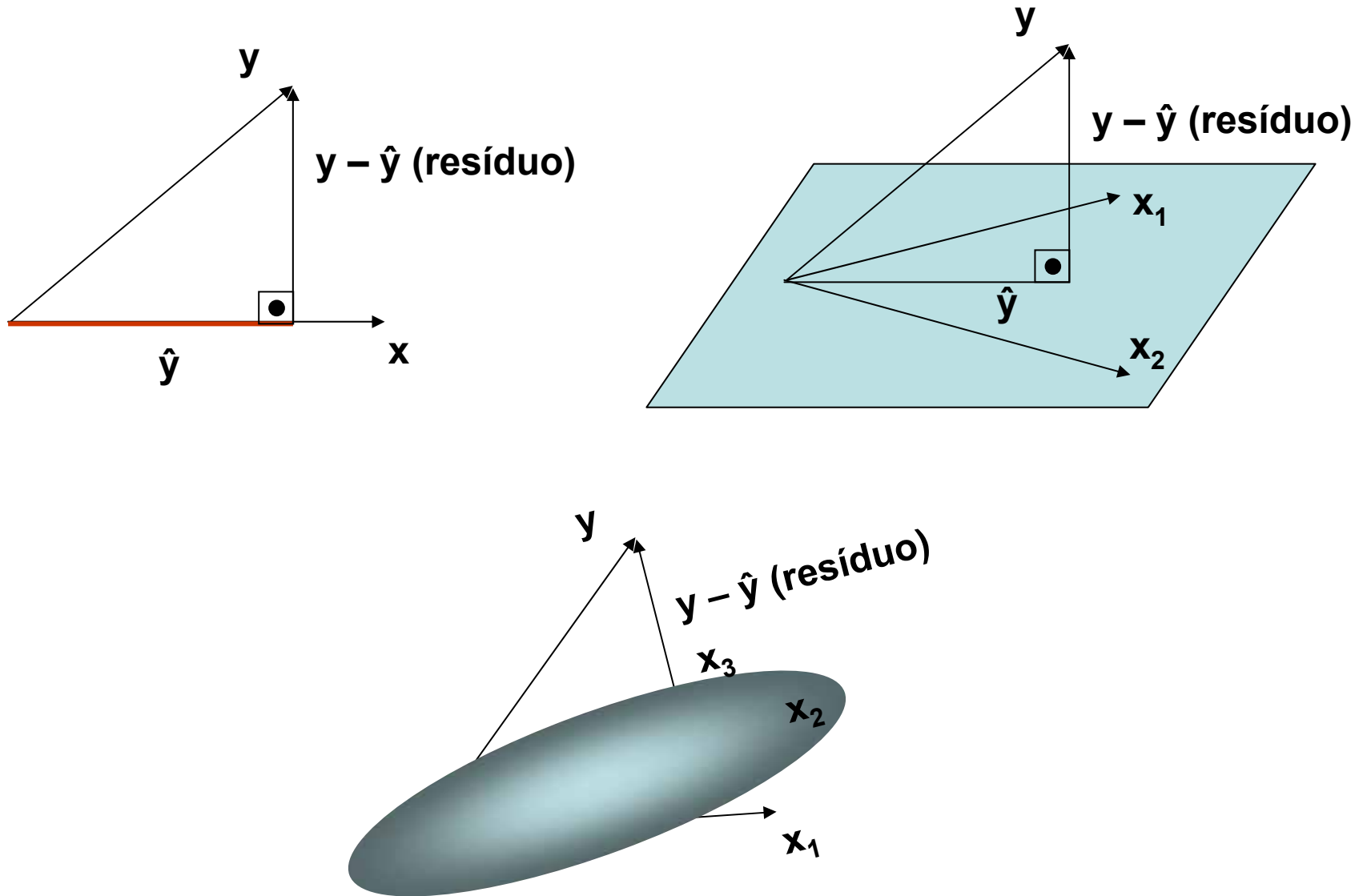
$$[X'Y] \sim [X'X] b$$

$$b \sim [X'X]^{-1} [X'Y]$$

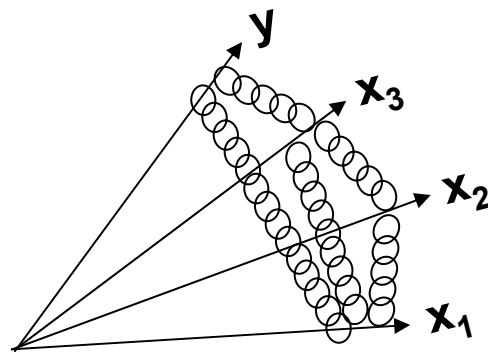
- ✓ Como estimar α e β ?
- ✓ Minimizar a soma dos quadrados dos resíduos

Exemplos no R

Modelagem numérica em Biologia



Modelagem numérica em Biologia



Modelos lineares

Quando x e y são contínuos...

✓ Regressão

Quando y é contínuo e x é categórico...

✓ ANOVA

Quando y é contínuo e X tem dados contínuos e categóricos...

✓ ANCOVA

Como fazer no R?

```
lm( formula, data, weights, subset, na.action )
```

'formula' - é uma fórmula estatística que indica o modelo a ser ajustado. Possui a mesma forma básica que foi vista na funções gráficas.

```
y ~ 1 + x
```

'data' - o conjunto de dados (data.frame).

'weights' - são os pesos para regressão ponderada.

'subset' - um vetor com as condições que definem um sub-conjunto dos dados.

'na.action' - função que especifica o que fazer no caso de observações perdidas (NA). O valor default é 'na.omit' que elimina as linhas (observações) que possuem observações perdidas nas variáveis definidas na fórmula.

Estrutura da 'fórmula' no R

$y \sim x$: modelo com constante implícita

$y \sim 1 + x$: explicitamente indicando a constante

$y \sim -1 + x$: modelo sem constante

$y \sim x_1 + x_2 + \dots + x_n$: modelo múltiplo

$y \sim I(x_1^2 * x_2^3)$: y como função de $(x_1^2 * x_2^3)$

Estrutura da 'fórmula' no R

Símbolos utilizados nas Fórmulas Estatísticas

Expressão

Significado

$Y \sim X$

Modele Y como função estatística de X

$A + B$

inclui ambos os fatores A e B

$A * B$

$A + B + A:B$

A / B

$A + B \%in\% (A)$ modelos hierárquicos

$A:B$

efeito da interação entre os fatores A e B

$B \%in\% A$

efeitos de B dentro dos níveis de A

A^m

todos os termos de A cruzados até à ordem m

Exemplos no R

Análise diagnóstica de Modelos lineares

Pressuposto da distribuição normal de erros

- ✓ Modelo linear pressupõe que o erro, ou resíduo, tenha distribuição Normal (média 0, DP 1)
- ✓ Deve-se testar se os **resíduos** têm distribuição **Normal**, e não as variáveis.

Modelagem numérica em Biologia

✓ **Análise diagnóstica de modelos lineares**

➤ **Avaliar a qualidade da análise (se atende aos pressupostos teóricos do método, se a relação é linear, presença de valores discrepantes)**

✓ **Princípios básicos**

➤ **Analisar o resíduo e transformar a variável resposta, se necessário**

✓ **O que é resíduo**

➤ **A diferença entre os valores observados e os estimados pelo modelo;**

Exemplos no R

Leitura

Fowler 1990 - The 10 most common statistical errors

Koricheva 2003 - Non-significant results in ecology

Zar 1996 – Cap 13, 3ª Ed.

Kleinbaum et al 1988 – Regression Diagnostics

Atkinson 1973 - Testing transformations to normality

Chatterjee & Hadi 1986 - Influential observations

Brown et al 1995 - Teaching concepts of accuracy