

Análise de Regressão

Biologia Quantitativa
Departamento de Zoologia – UnB

02 de março de 2021

Referências

- Andrade e Ogliari cap. 9
- Sokal & Rohlf cap 14 (sec 14.1 a 14.7)

Conceitos Centrais

- Equação Linear: equação em que cada termo é uma constante ou o produto de uma constante e uma única variável
- Modelo Linear Geral (GLM)

$$Y = XB + U$$

- Y é matriz de observações
- X é matriz de variáveis
- B é matriz de parâmetros
- U é matriz de erros

Conceitos Centrais

- Mínimos Quadrados
- Ajuste por Mínimos Quadrados
- Resíduos e interpretação
- Hipóteses em modelos lineares
- Tamanho amostral e universo amostral

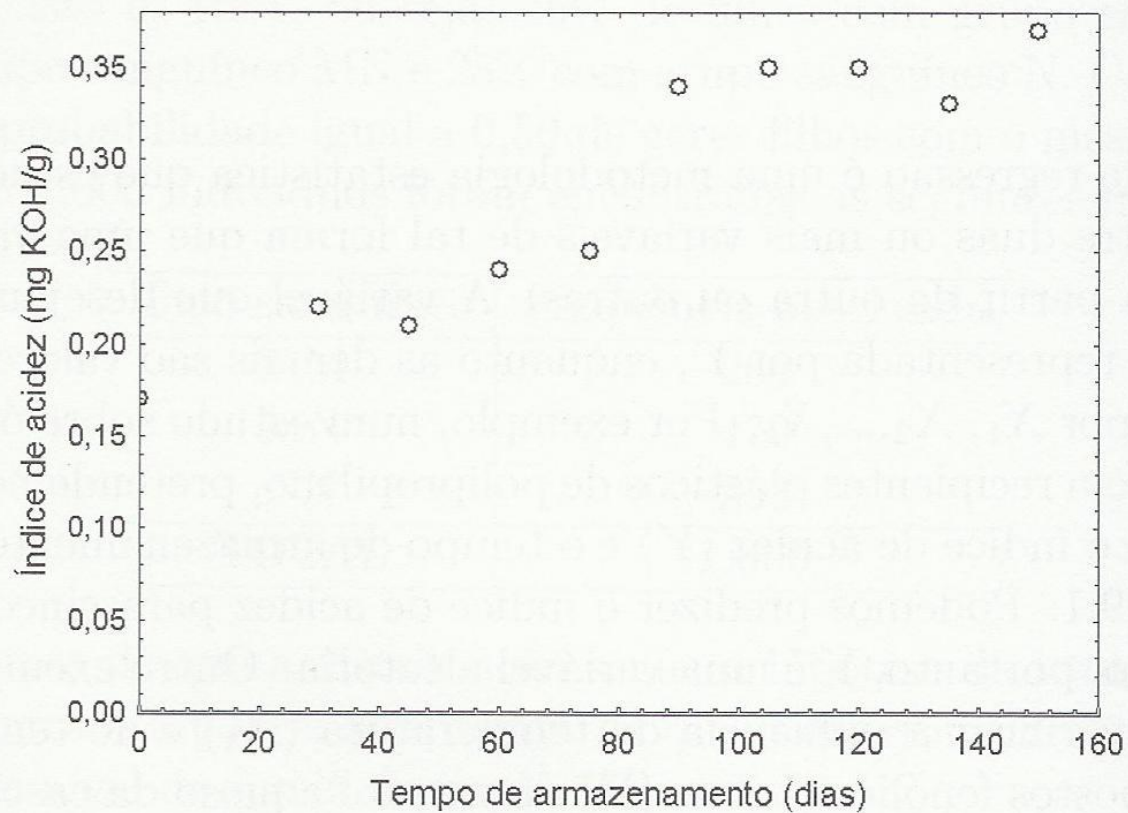
Conceito

- Ajuste de uma equação linear a um conjunto de dados usando o método dos mínimos quadrados.
- Uma variável dependente e uma ou mais independentes
- Estimativa dos parâmetros e teste de hipóteses relativas aos mesmos

Equação Regressão

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i.$$

Distribuição de Variáveis



Cálculo do modelo

$$b_1 = \frac{n \sum (X_i Y_i) - [(\sum X_i)(\sum Y_i)]}{n \sum X_i^2 - (\sum X_i)^2} \quad (9.3)$$

e

$$b_0 = \frac{\sum Y_i - b_1 \sum X_i}{n} \quad (9.4)$$

O modelo de regressão ajustado com os dados da amostra é dado por:

$$\hat{Y}_i = b_0 + b_1 X_i. \quad (9.5)$$

Ajuste

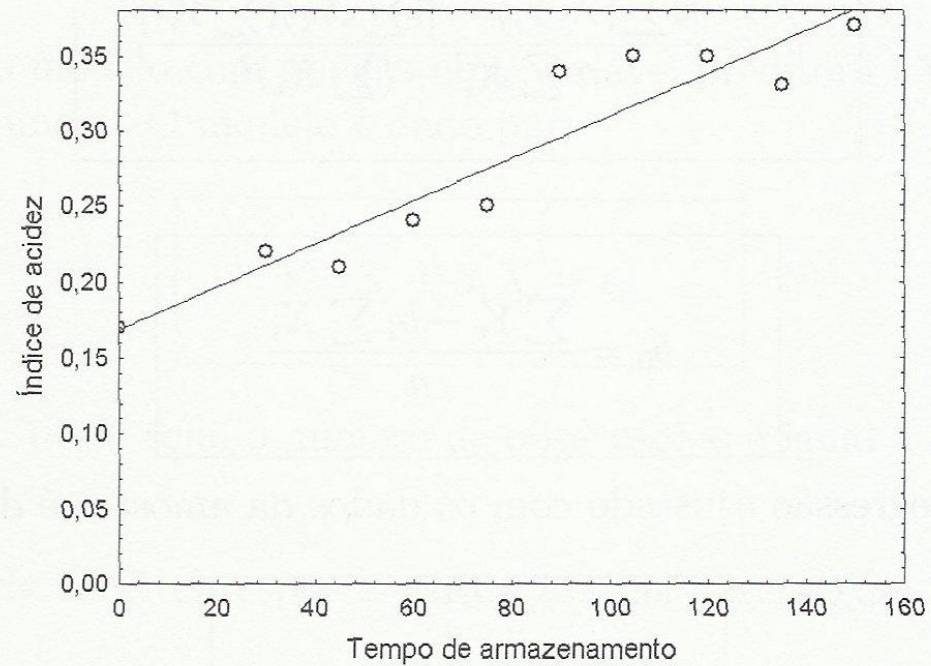


Figura 9.3 – Reta de regressão ajustada das variáveis índice de acidez e tempo de armazenamento

Premissas

- Variável independente medida com exatidão
- Variável dependente com distribuição de erros normal e homoscedástica

Premissas

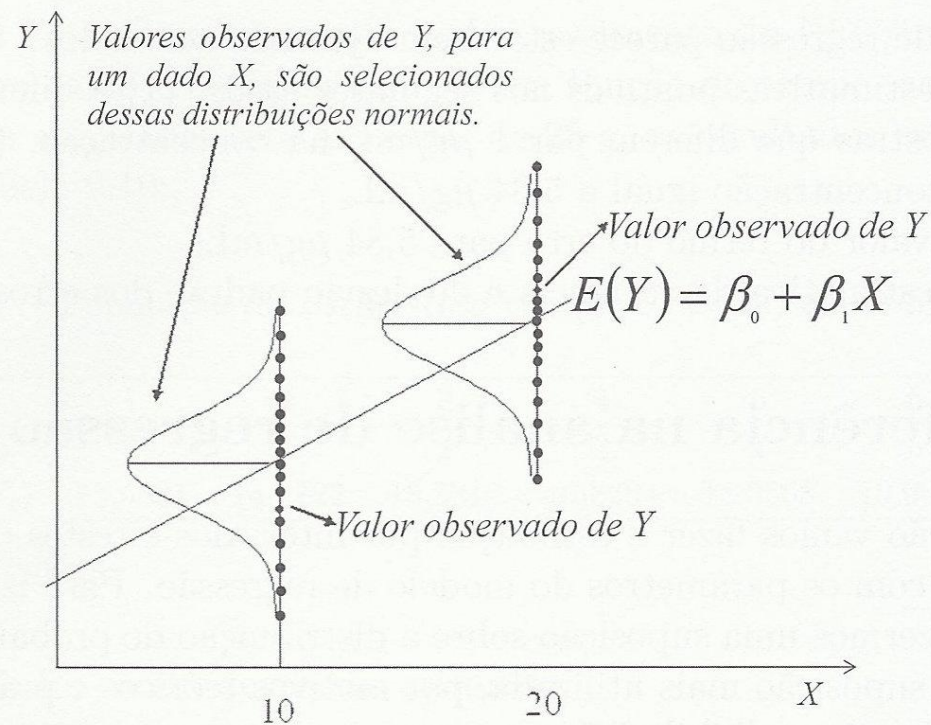
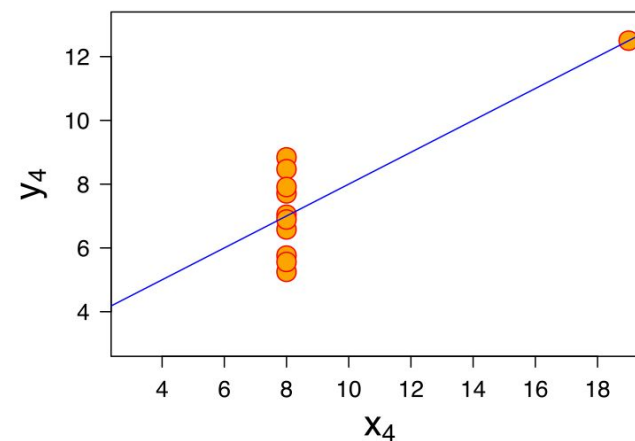
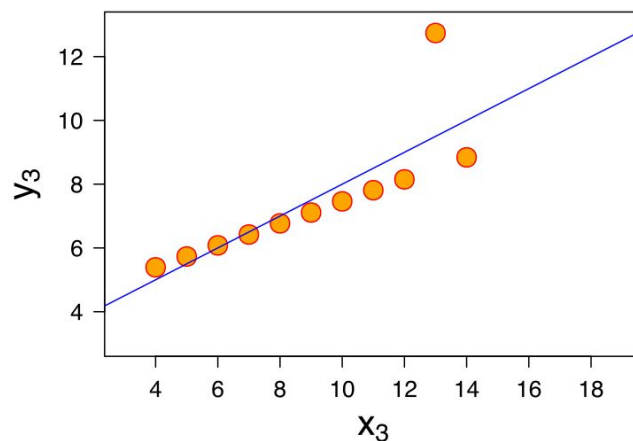
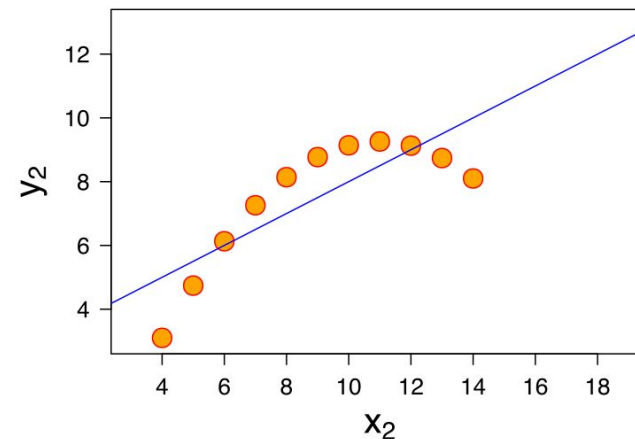
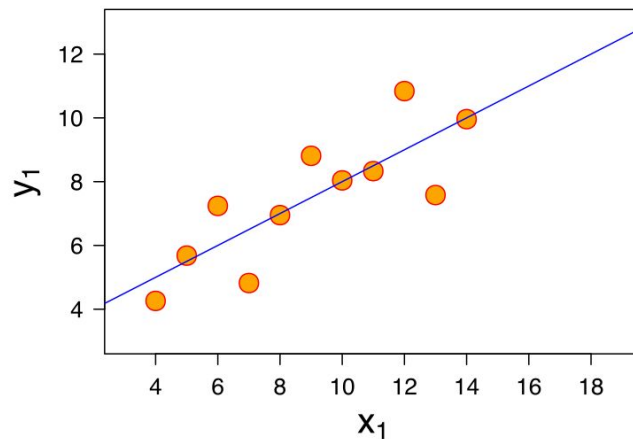


Figura 9.5 – Distribuição de probabilidade da variável Y para diferentes valores de X

Quarteto de Anscombe

4 distribuições mesma med var

Qual é linear?



Hipóteses

- Hipóteses a respeito da inclinação
- Hipóteses a respeito do intercepto
- Hipóteses sobre a linearidade (análise dos resíduos)

Usos

- Examinar relação entre variáveis
- Ajustar/testar modelo aos dados observados
- Estimar valor e variância de variável dependente para determinados valores da variável independente
- Projetar valores da variável dependente em situações não amostradas

Exemplo – Predação em Lobeira

- Quais fatores influenciam a intensidade do ataque de roedores a frutos de lobeira (*Solanum lycocarpum*) no cerrado?
- Distância da planta-mãe? Quantidade de frutos produzidos?
- Armadilhas para roedores na árvore, e a 0m, 5m, 10m
- Regressão linear para relação entre tamanho da safra e proporção de frutos consumidos

Exemplo – Predação em Lobeira

- Ajuste da equação $y = ax^b$
- y = número frutos intactos
- a = parâmetro
- x = número total de frutos
- b = parâmetro: $b=1$ não há efeito do tamanho da safra sobre proporção de ataque; $b>1$ efeito positivo ataque maior nas safras maiores; $b<1$ efeito negativo
- Equação foi linearizada para poder computar:

Exemplo – Predação em lobeira

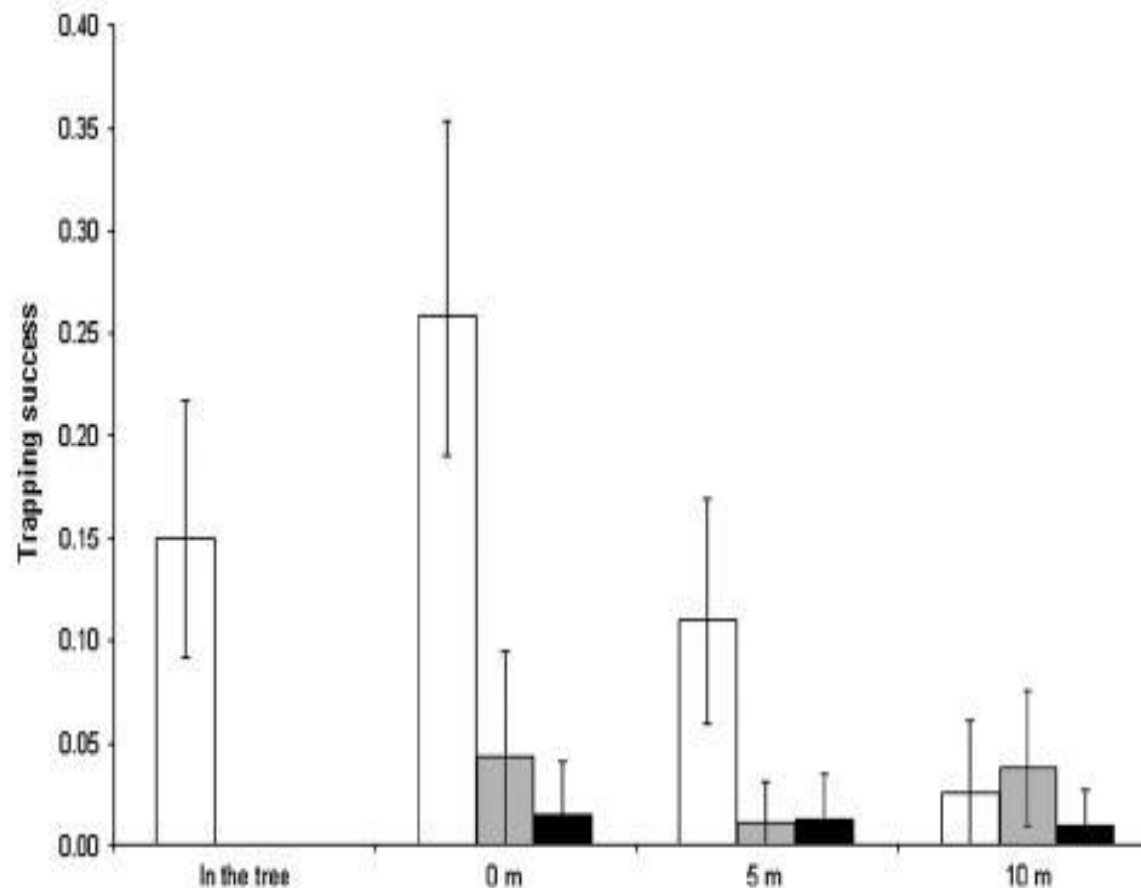


Fig. 1 – The success rate in trapping rodents as a function of distance to the nearest *Solanum lycocarpum* plant. The rodent species are *Oryzomys scotti* (white columns), *Necomys lasiurus* (grey columns) and *Calomys callosus* (black columns). The bars indicate the 95% confidence intervals estimated using bootstrapping procedures.

Exemplo – Predação em lobeira

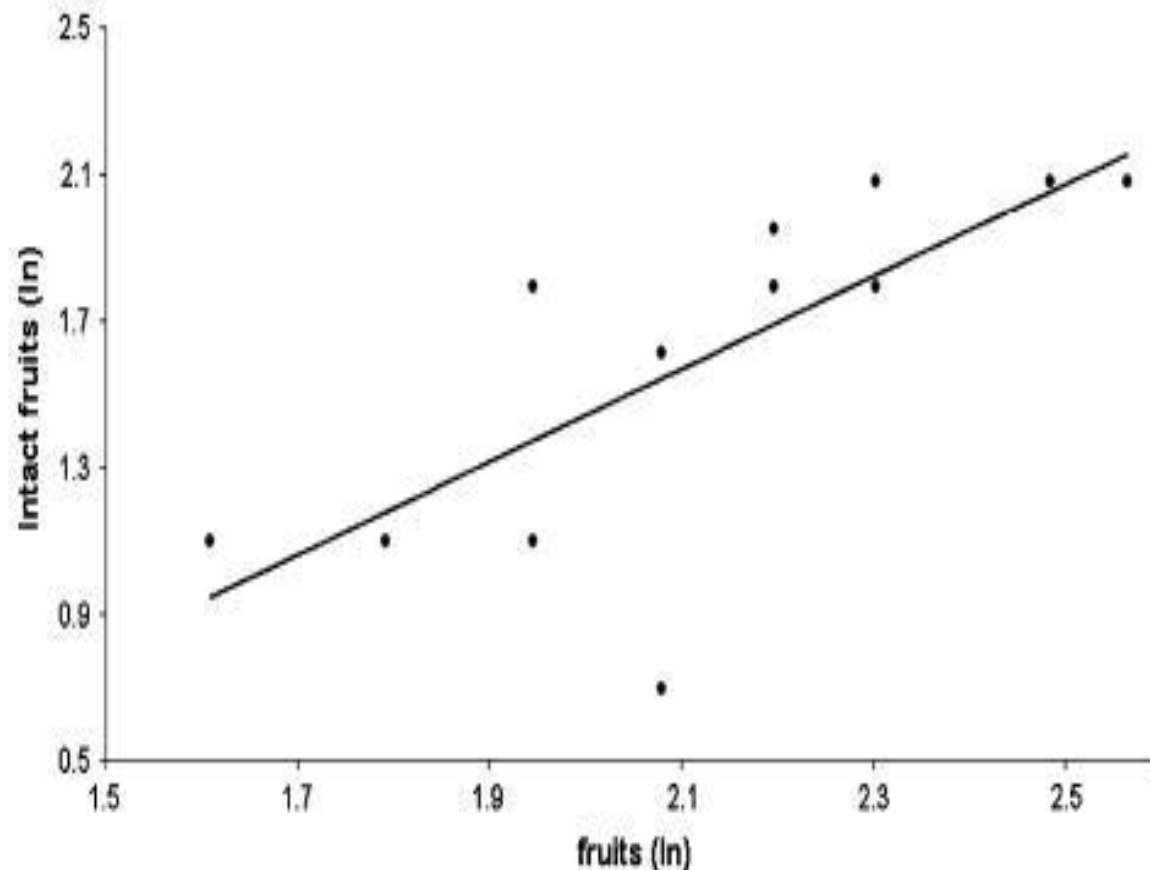


Fig. 2 – The relationship between crop size (number of fruits) and intact crop size of shrubs of *Solanum lycocarpum*. Solid line indicates the power-law that best fits the data. The slope of 1.26 is significantly greater than 1.0, suggesting that larger crop sizes are attacked proportionally less than smaller crops.

Aplicações

- Análises com múltiplas variáveis dependentes
- Modelos lineares em combinação com outros métodos como Anova (análise de covariância)
- Quadro lógico para partição de variâncias e estimativas de erros