

Curso de Estatística Básica

Aula 08 (?) Ajuste de modelos

Pavel Dodonov
pdodonov@gmail.com
anotherrecoblog.wordpress.com

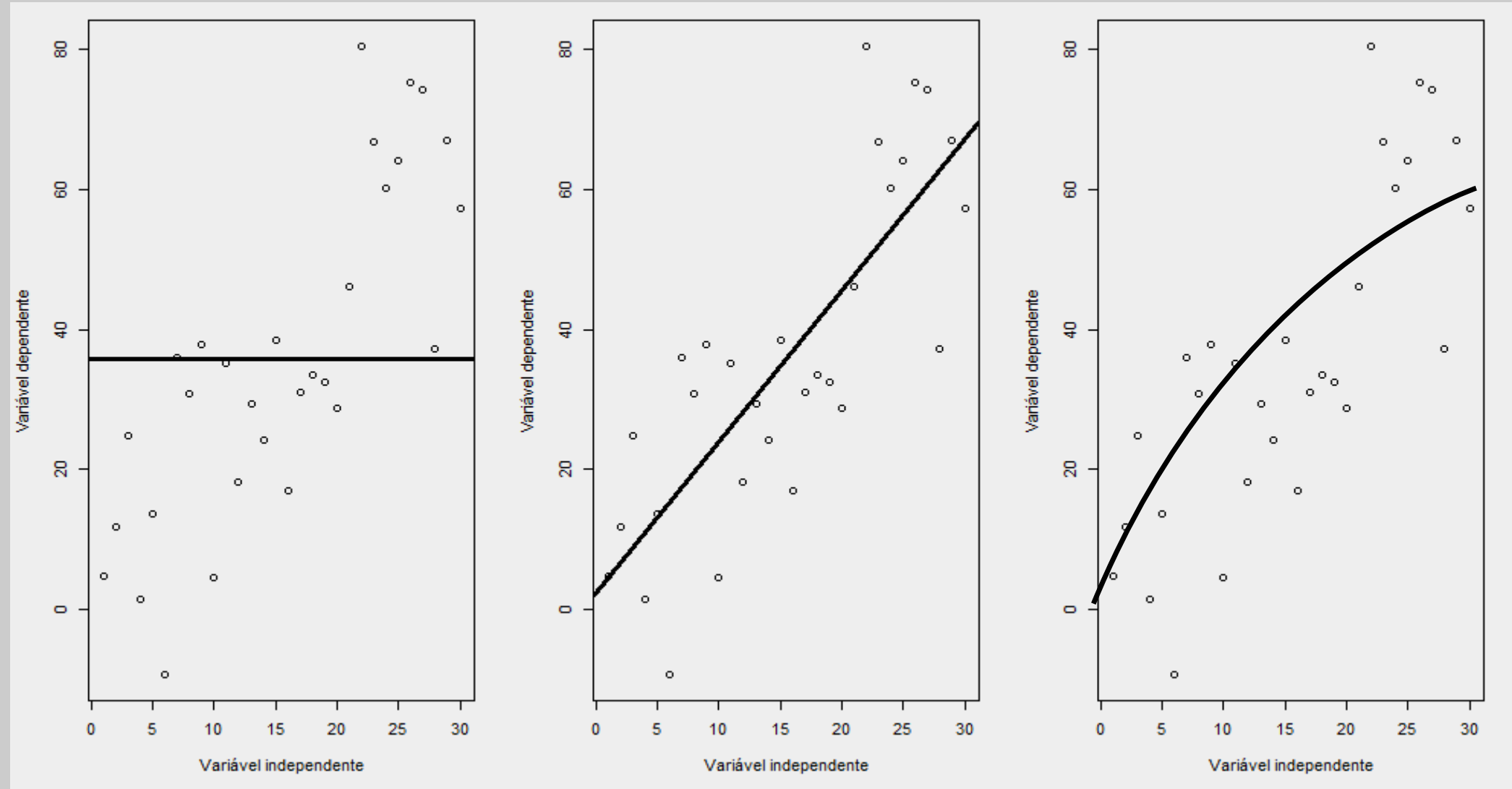
Ajuste de modelos

- 1) Múltiplas variáveis explanatórias
- 2) Relações não-lineares
- 3) Distribuições não-normais

Seleção de modelos

Baseada em
verossimilhança
(*likelihood*) e na
teoria de
informação

Seleção de modelos

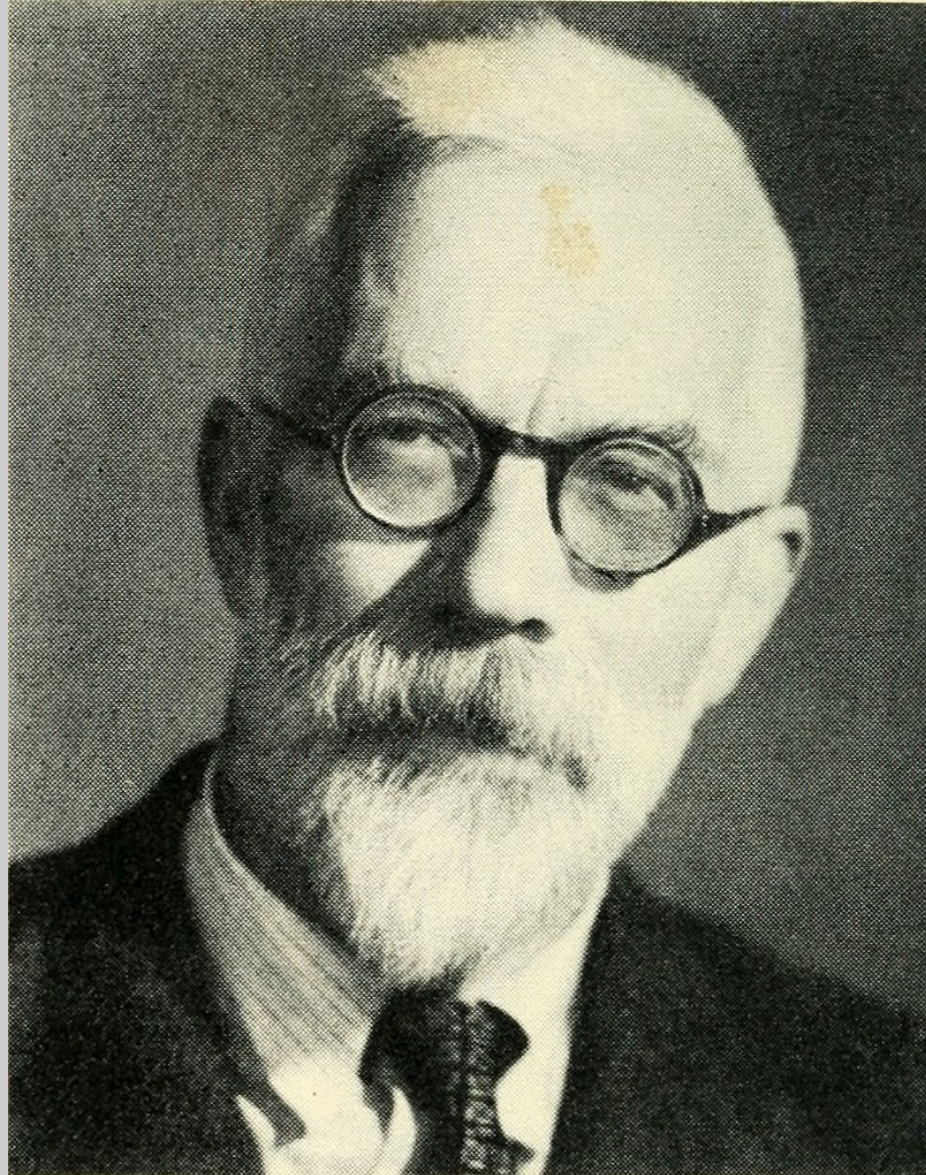


Explicação mais provável?

Verossimilhança

**Probabilidade de
que um modelo
tenha gerado os
dados observados**

Verossimilhança



Ronald Fisher

Comparando modelos

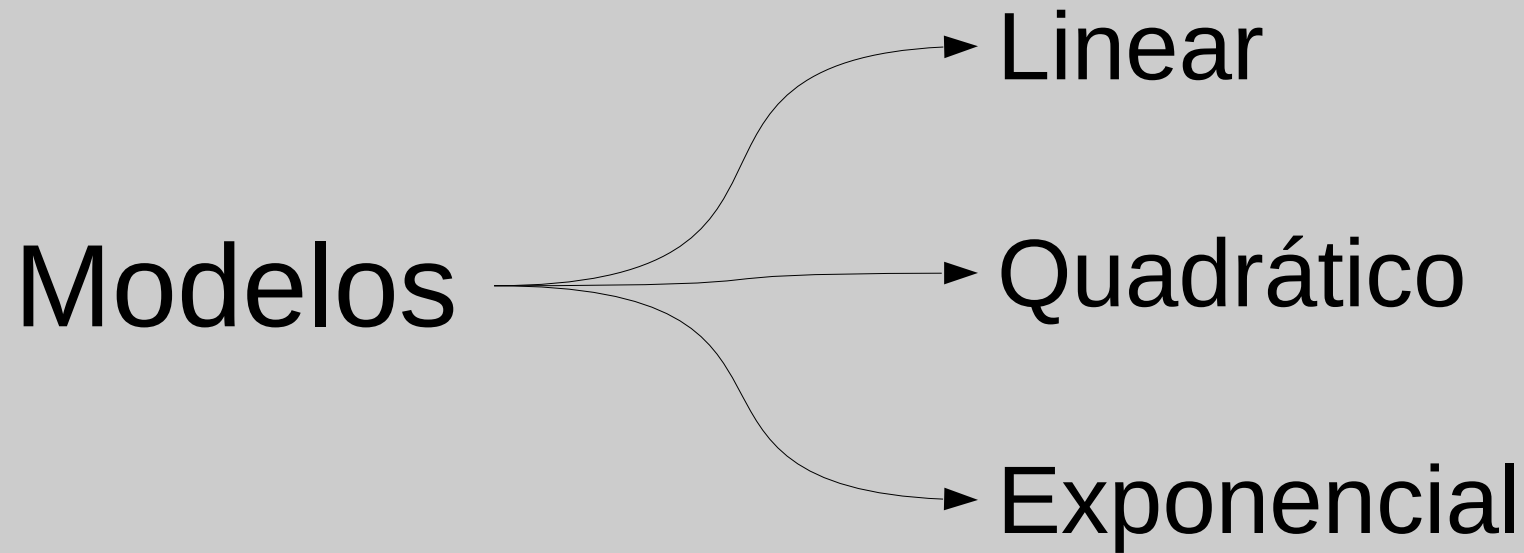
**Melhor modelo: o
modelo mais próximo
da realidade biológica,
considerando os dados
coletados**

Ajuste de modelos

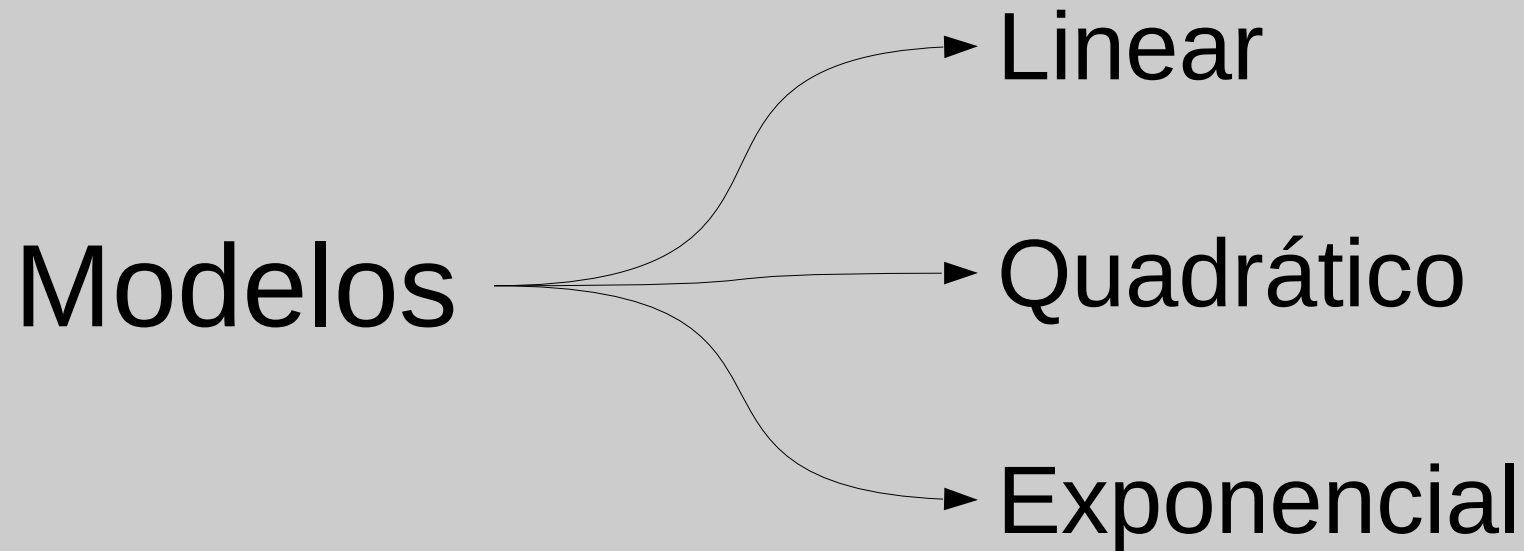
Modelos

Parâmetros

Ajuste de modelos

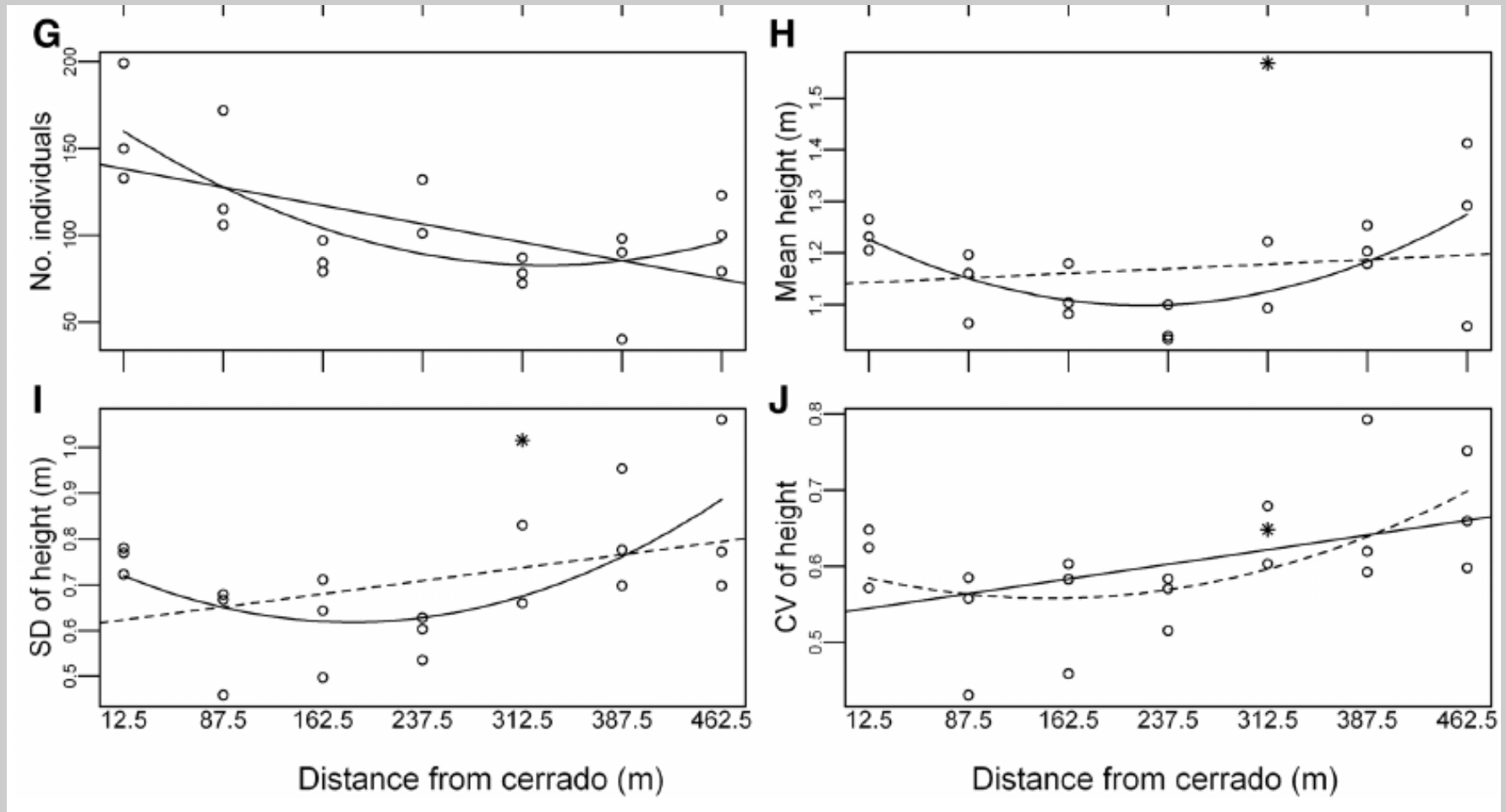


Ajuste de modelos




Implicam diferentes **relações**
entre as variáveis

Ajuste de modelos



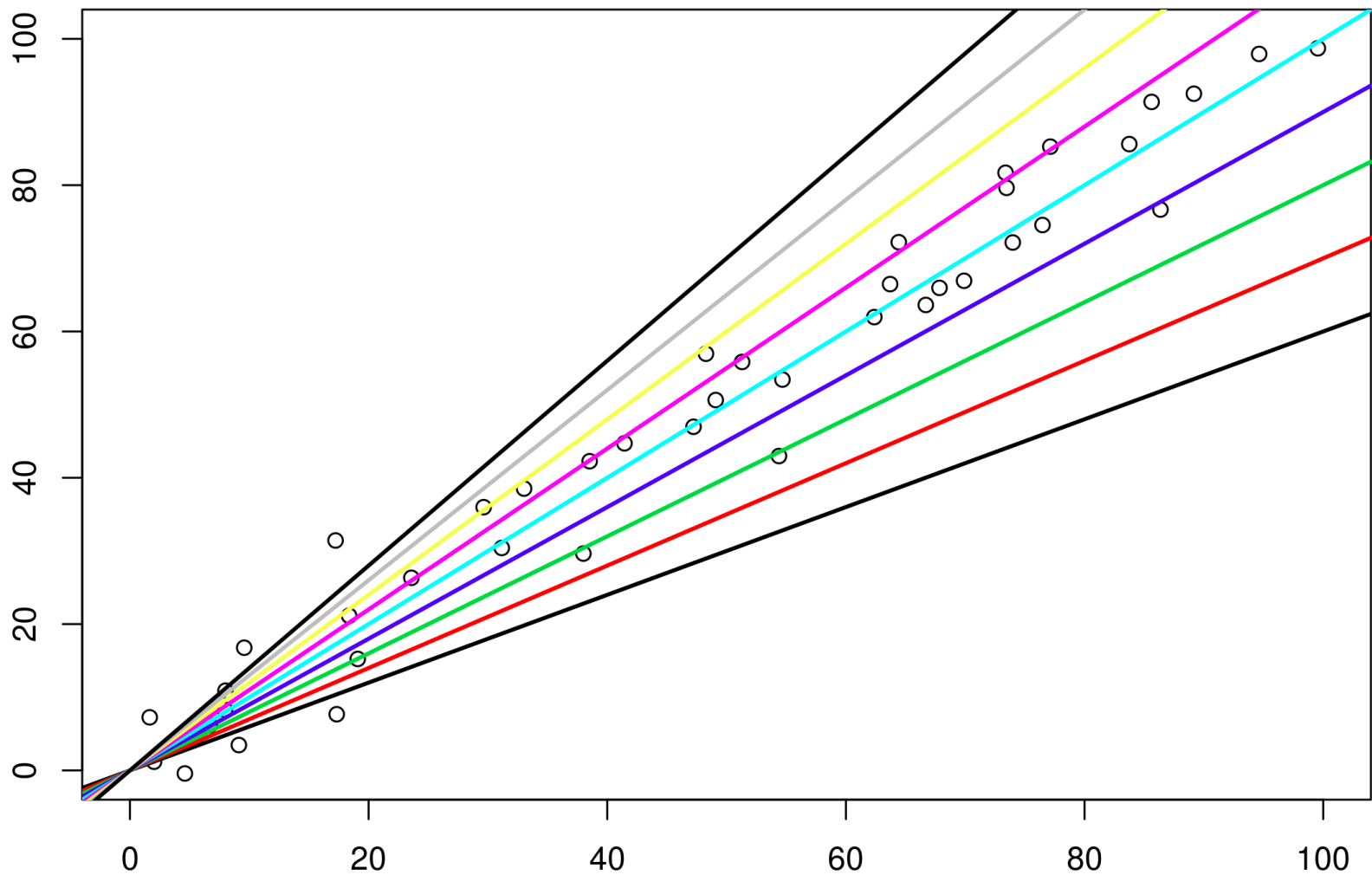
Exemplo: Comparando modelos lineares e quadráticos
(Dodonov et al. 2014)

Ajuste de modelos

Modelos  Tipo de relação

Parâmetros  ?

Ajuste de modelos

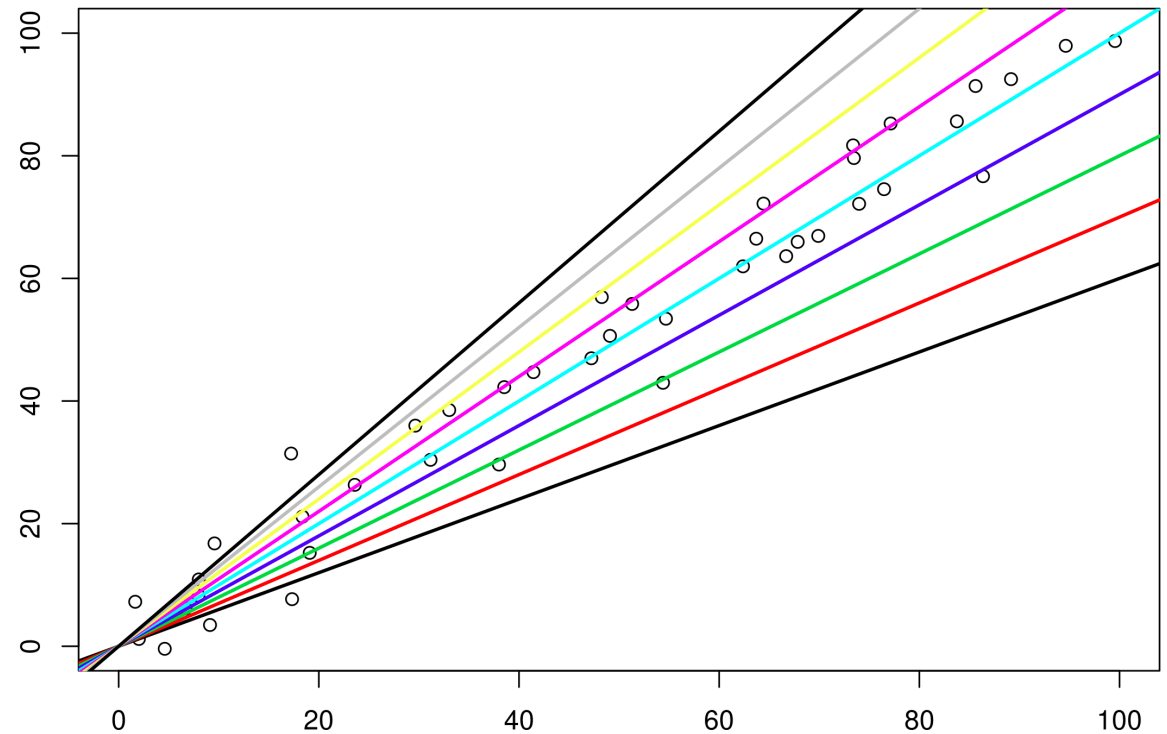


Ajuste de modelos

Modelo linear

Intercepto = 0

Inclinações variando de 0.6 a 1.4

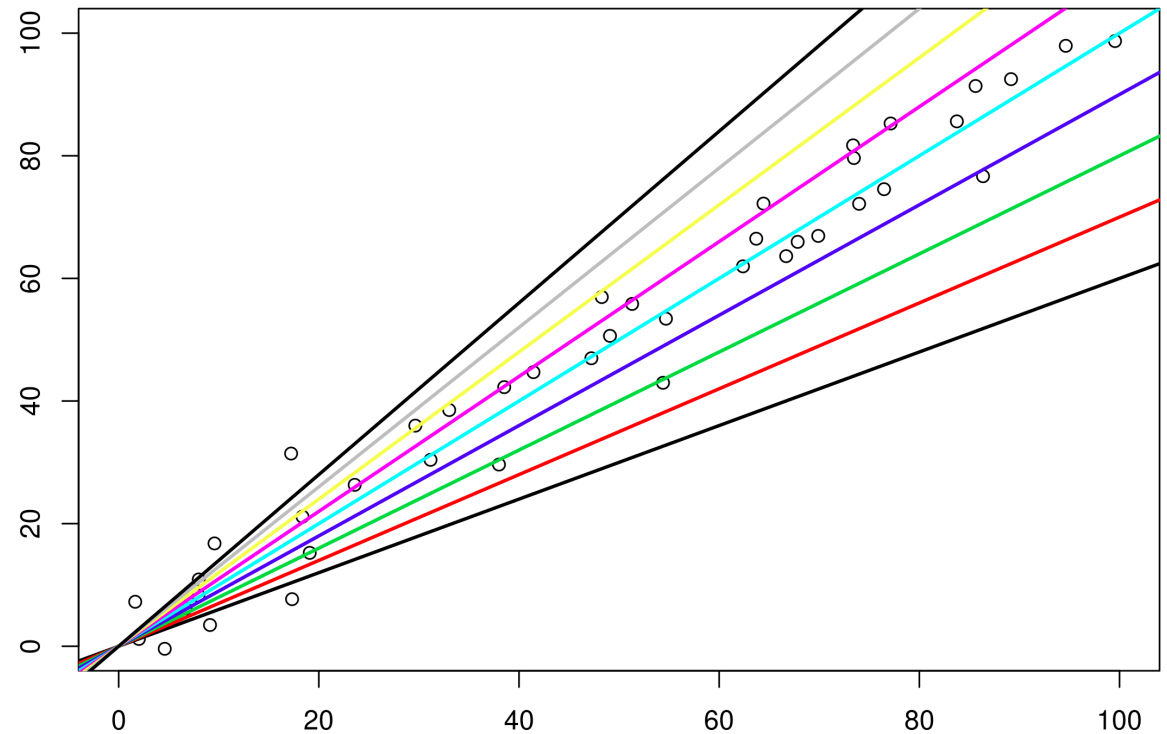


Ajuste de modelos

Modelo linear

Intercepto = 0


Inclinações variando de
0.6 a 1.4



Os **valores** do intercepto e da
inclinação são **parâmetros**

Ajuste de modelos

Modelos  Tipo de relação

Parâmetros  A forma exata da relação

Ajuste de modelos

Modelos

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

$$Y_i = \beta_0 + \beta_1 * X_i + \beta_2 * X_i^2 + \epsilon_i$$

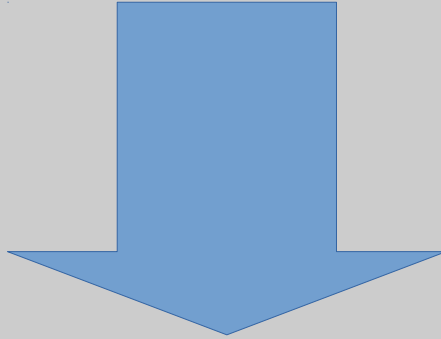
$$Y_i = \beta_0 * e^{(\beta_1 * X_i)} + \epsilon_i$$

Parâmetros

Os valores
exatos de
 β_0 , β_1 e β_2 .

Mais de uma variável

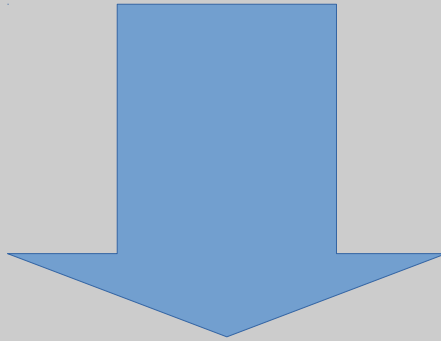
Tamanho do manguezal



Biodiversidade de formigas

Mais de uma variável

Tamanho do manguezal

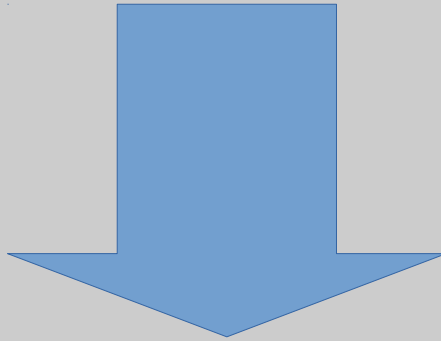


Espera-se maior
diversidade em
manguezais
maiores

Biodiversidade de formigas

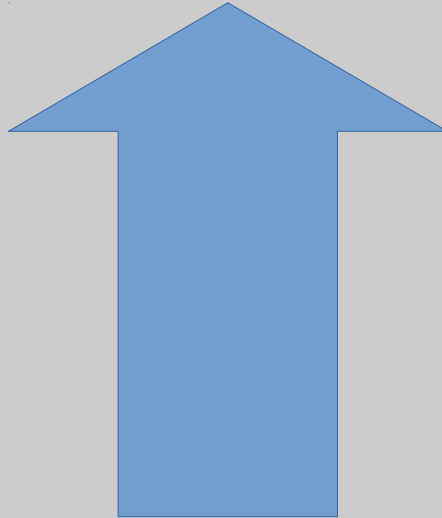
Mais de uma variável

Tamanho do manguezal



Espera-se maior
diversidade em
manguezais
maiores

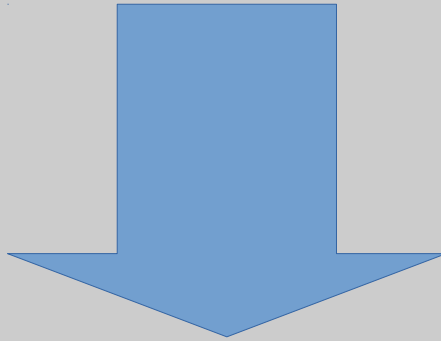
Biodiversidade de formigas



Estado de conservação do manguezal

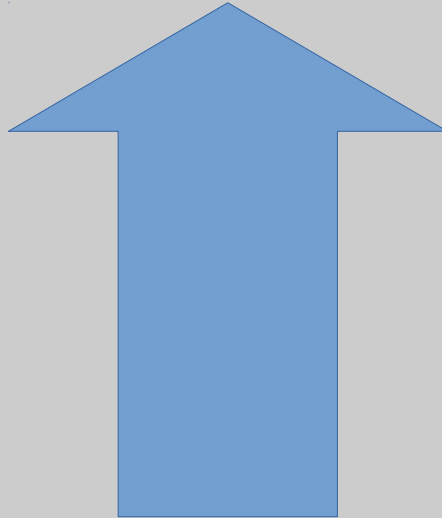
Mais de uma variável

Tamanho do manguezal



Espera-se maior
diversidade em
manguezais
maiores

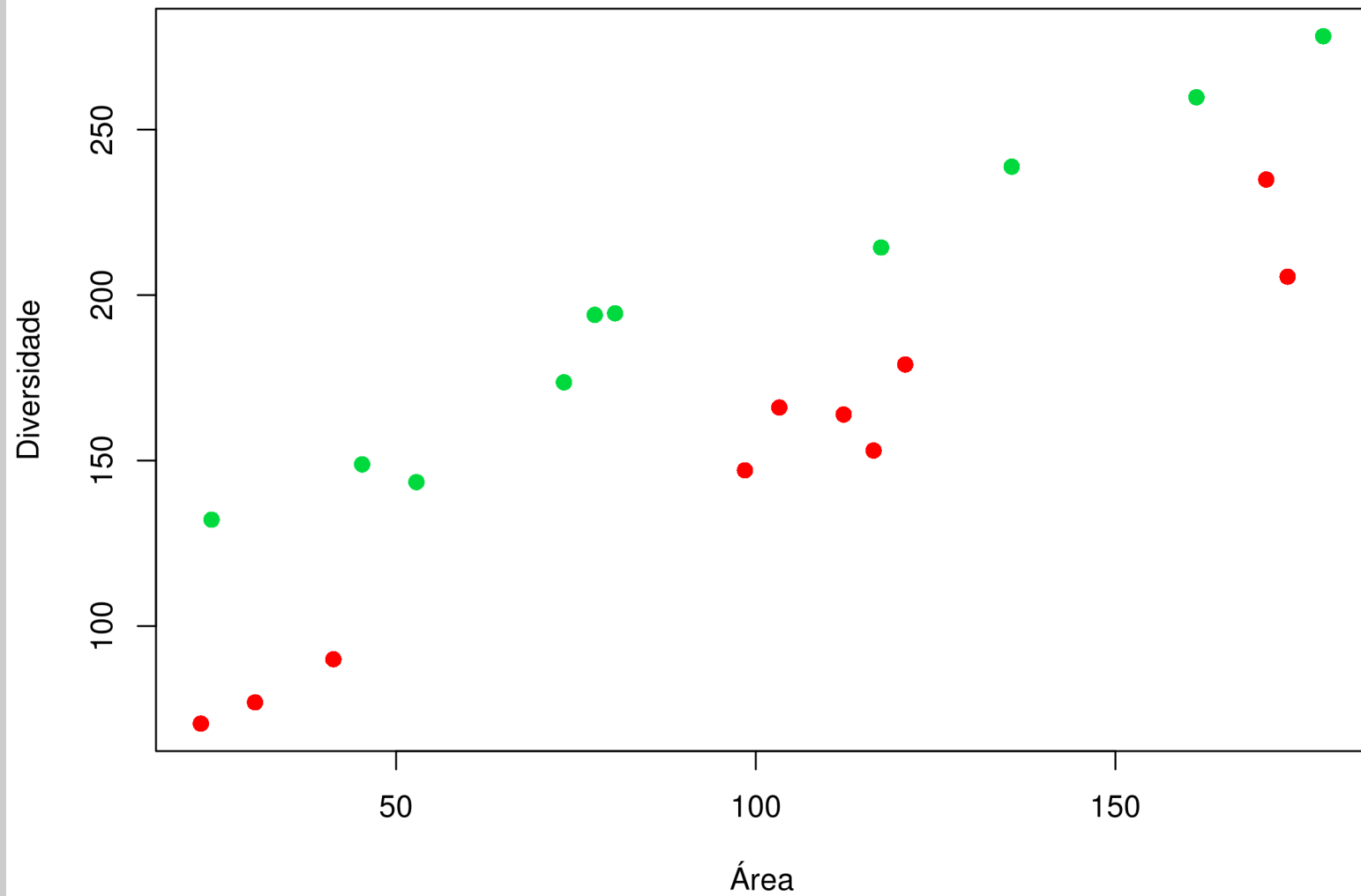
Biodiversidade de formigas



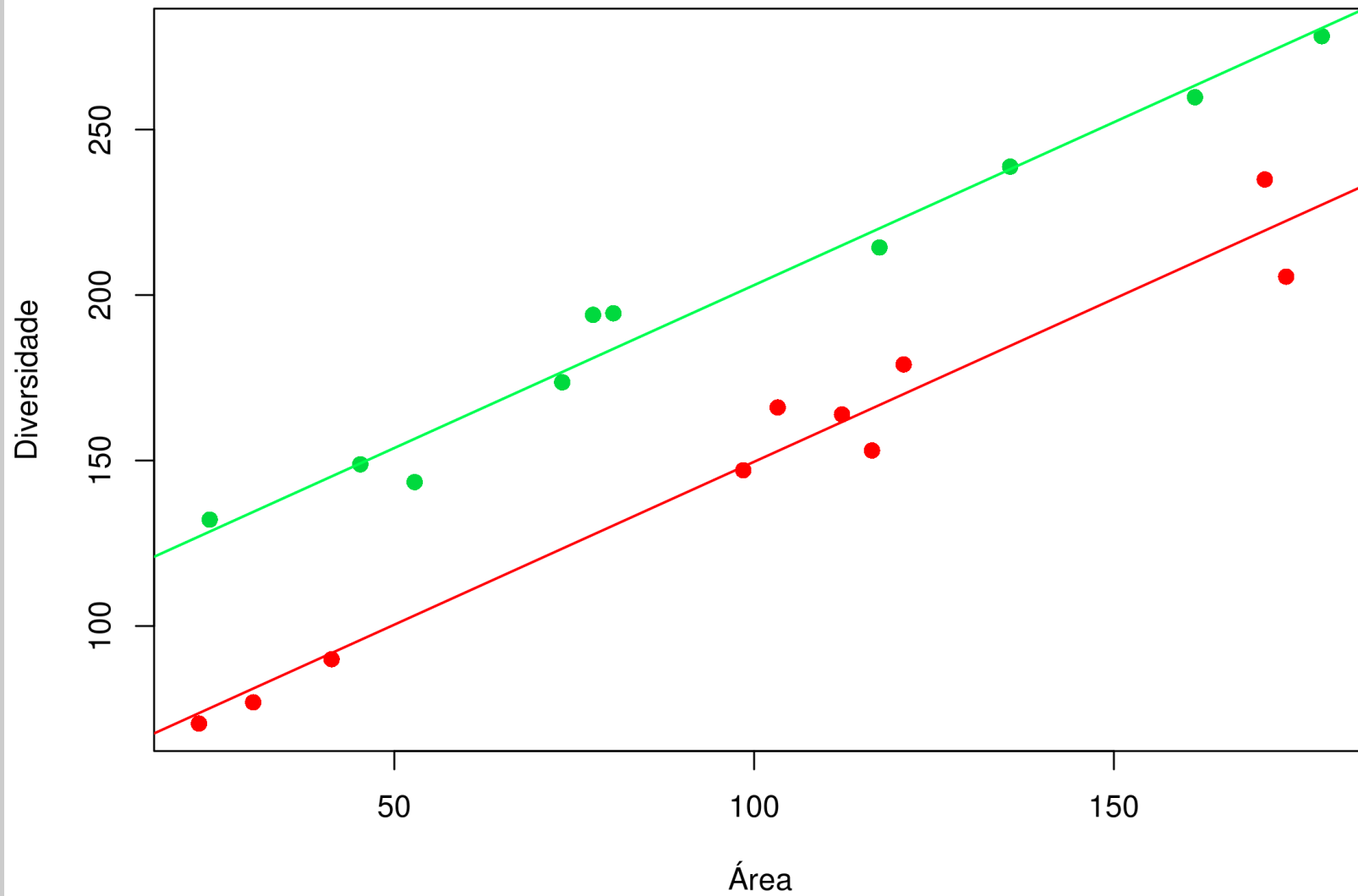
Espera-se menor
biodiversidade em
manguezais
antropizados

Estado de conservação do manguezal

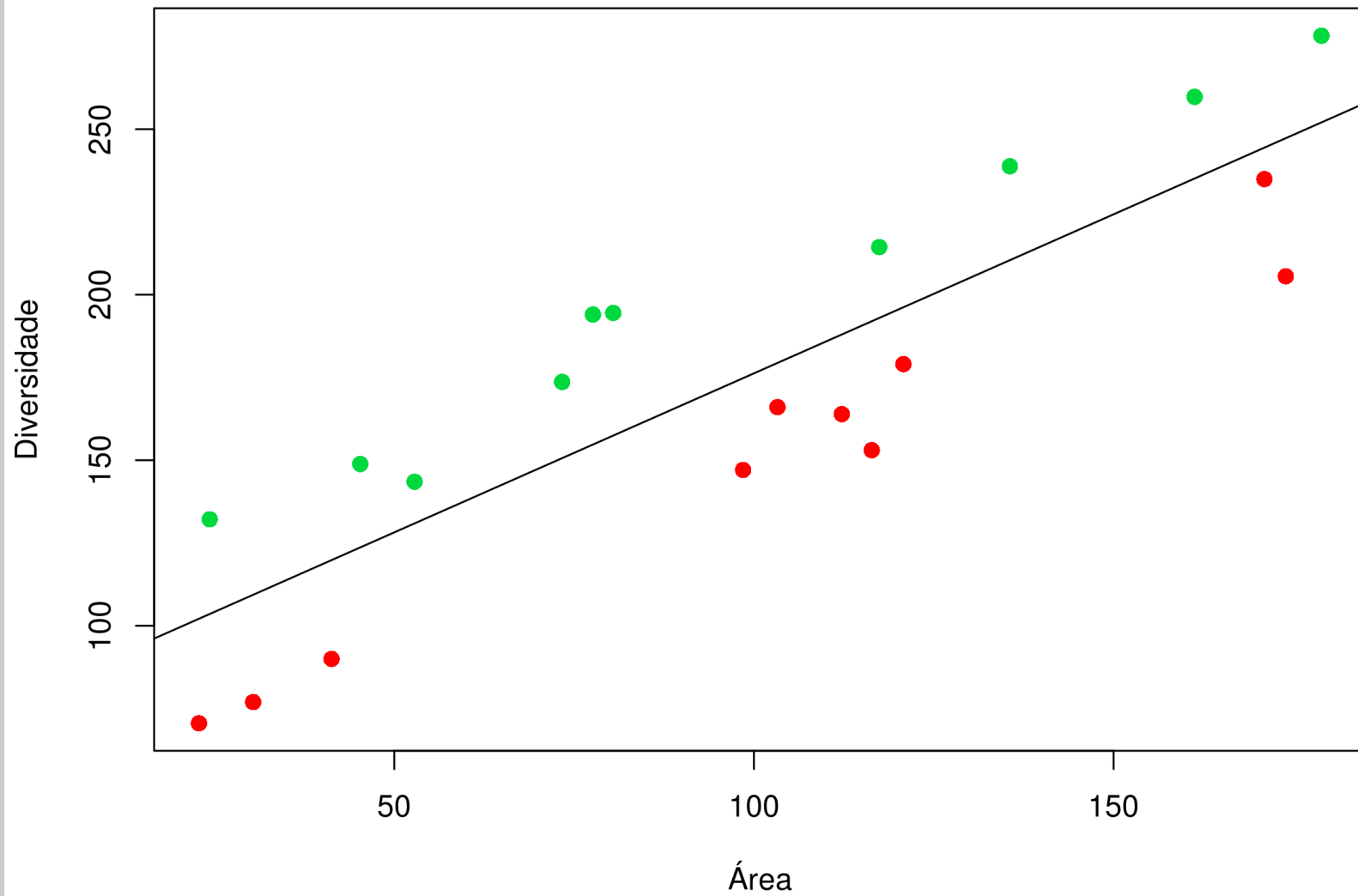
Mais de uma variável



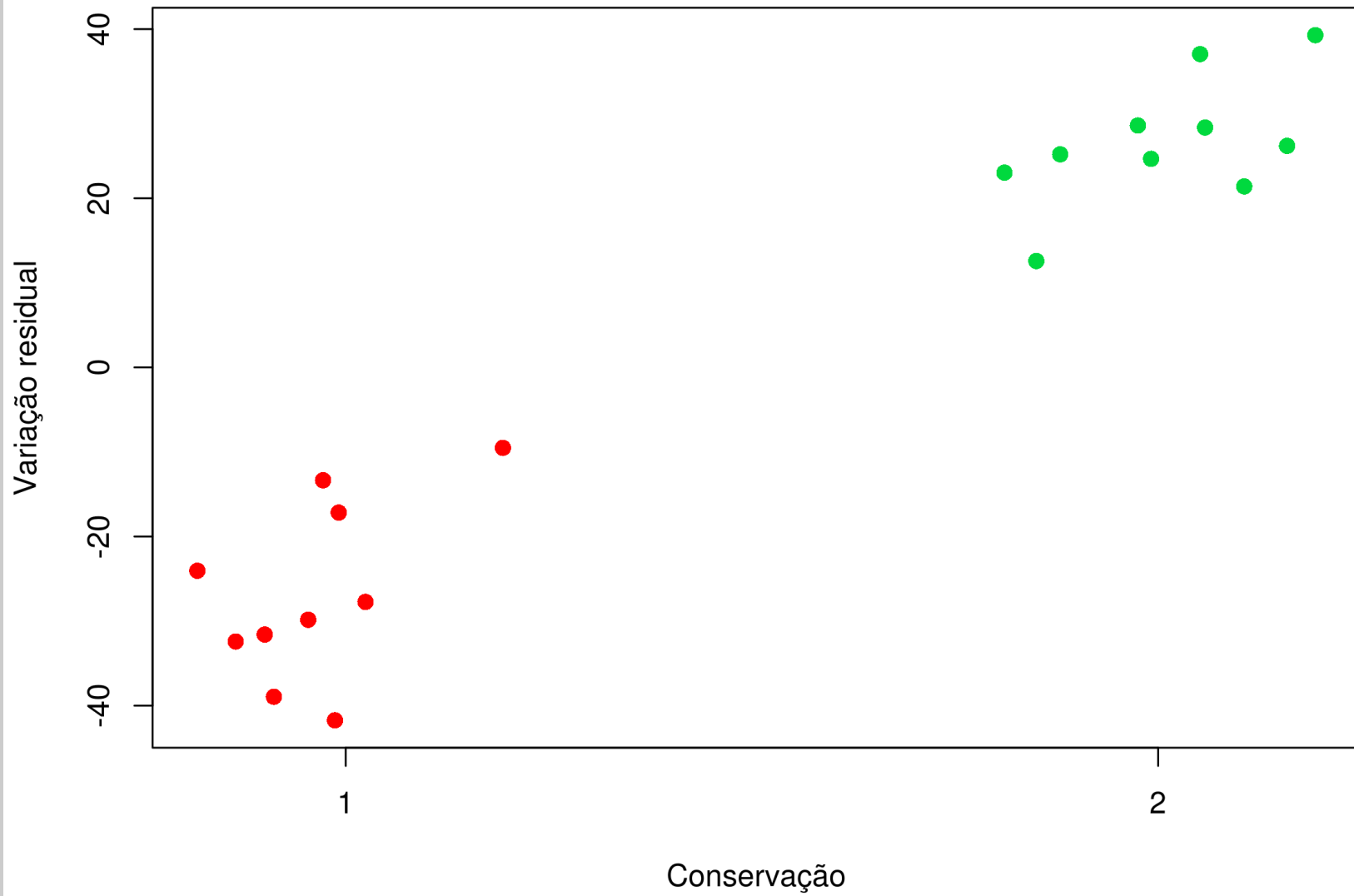
Mais de uma variável



Modelo mais simples...

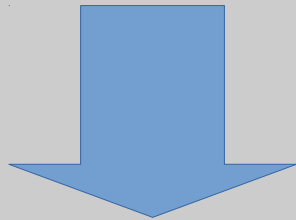


Variação residual

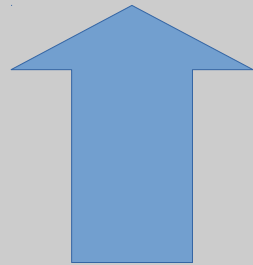


Mais de uma variável

Tamanho do manguezal



Biodiversidade de formigas



Estado de conservação do manguezal

Explicam a **variação residual**
– a variação que a
outra variável não
explica.

Mais de uma variável

Y: Biodiversidade de formigas (**contínua**)

X_1 : Tamanho do manguezal (**contínua**)

X_2 : Grau de conservação (**categórica** – preservado ou antropizado)

Mais de uma variável

Y: Biodiversidade de formigas (**contínua**)

X_1 : Tamanho do manguezal (**contínua**)

X_2 : Grau de conservação (**categórica** – preservado ou antropizado)

$$Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \epsilon$$

Mais de uma variável

Y: Biodiversidade de formigas (**contínua**)

X_1 : Tamanho do manguezal (**contínua**)

X_2 : Grau de conservação (**categórica** – preservado ou antropizado)

$$Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \epsilon$$

$$Y = \beta * X + \epsilon$$


Notação matricial

Diferentes nomes, mesmo princípio...

Regressão múltipla: diversas
variáveis explanatórias
contínuas

ANOVA multifatorial:
diversas variáveis
explanatórias categóricas

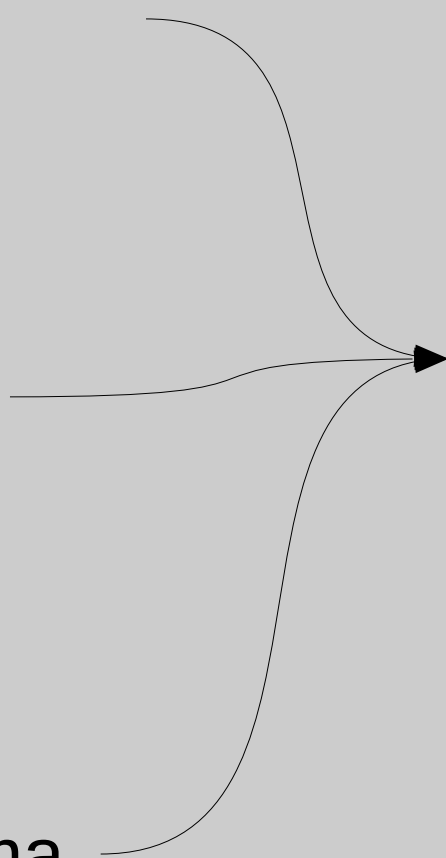
ANCOVA: Uma variável
explanatória contínua e uma
categórica

Diferentes nomes, mesmo princípio...

Regressão múltipla: diversas variáveis explanatórias contínuas

ANOVA multifatorial: diversas variáveis explanatórias categóricas

ANCOVA: Uma variável explanatória contínua e uma categórica



Modelo linear: engloba essas situações (e outras mais complexas)

Linear?

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

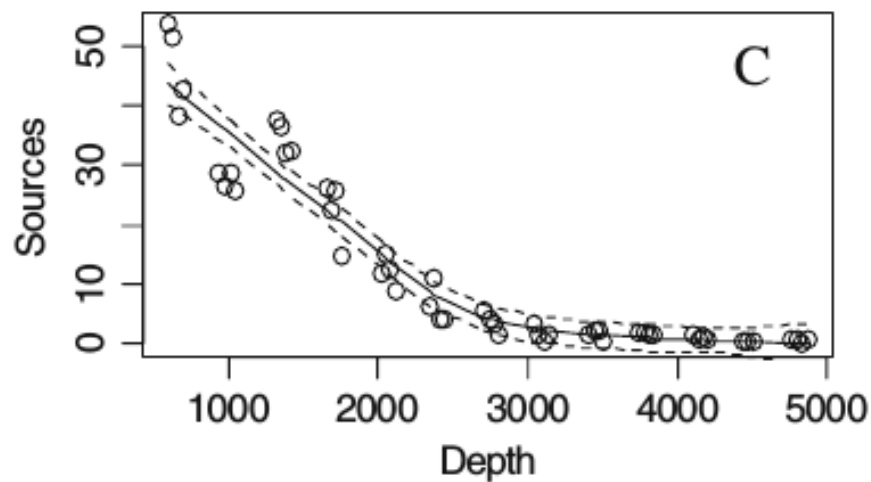
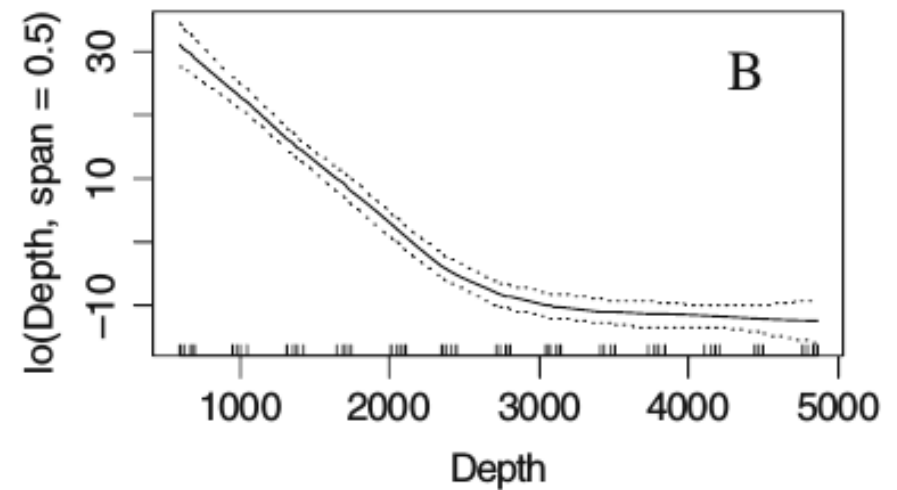
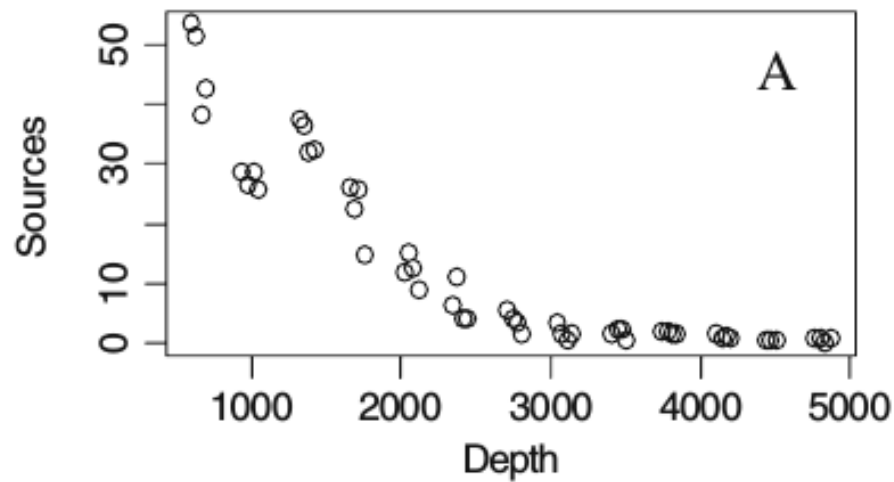
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \epsilon_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

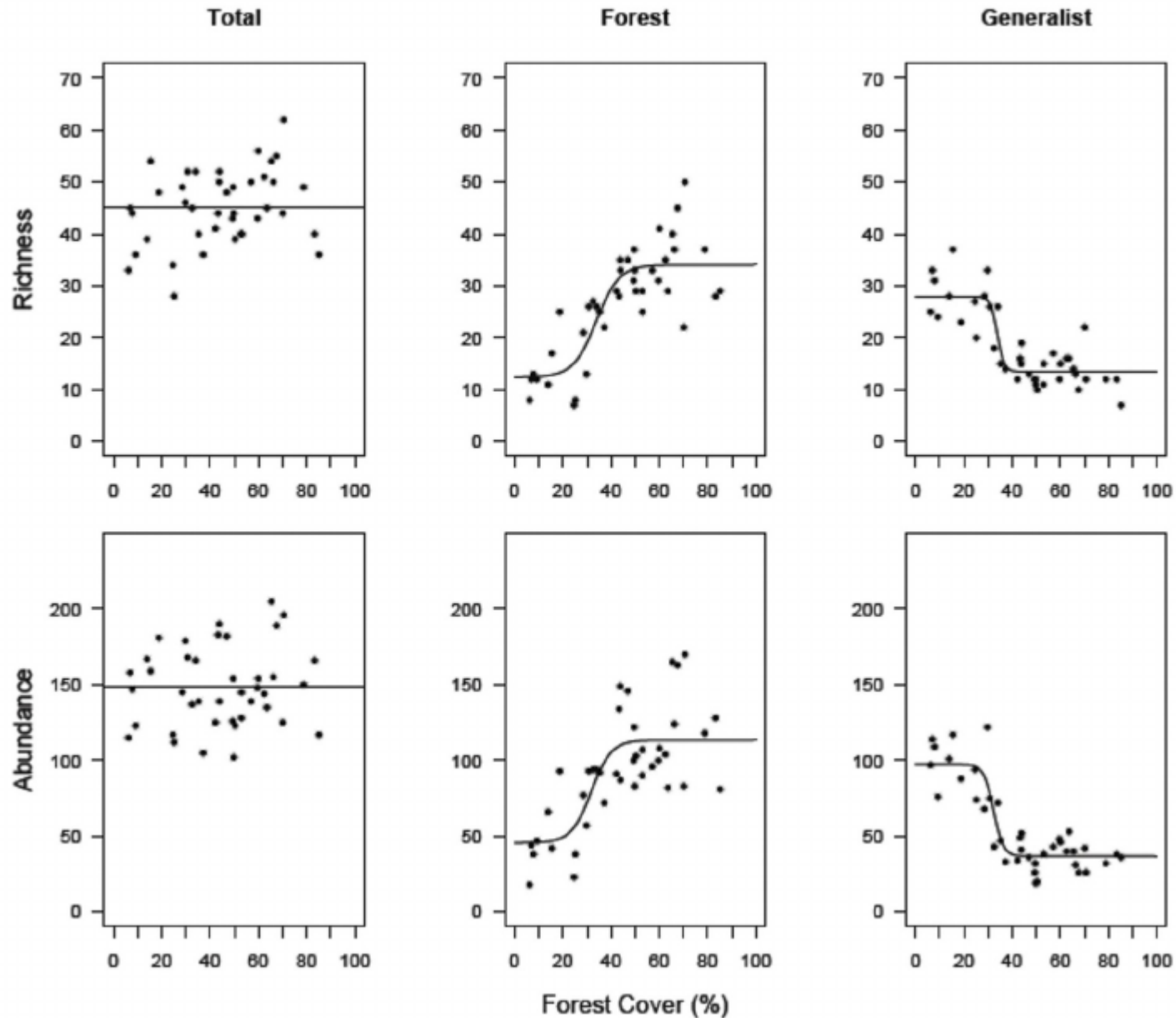
$$Y_i = \beta_0 e^{(\beta_1 X_i)} + \epsilon_i$$

“As coisas nem sempre são lineares!”



Zuur et al. 2009

Modelos não-lineares



Modelos não-lineares

Model	Equation
Null	$Y=a+\varepsilon$
Linear	$Y=a+bX+\varepsilon$
Power	$Y=aX^b+\varepsilon$
Logistic	$Y = \frac{a}{1+e^{(X-b)c}} + \varepsilon$
Piecewise	$\begin{cases} Y = a + bX + \varepsilon & \text{if } X \leq z \\ Y = c + dX + \varepsilon & \text{if } X > z \end{cases}$

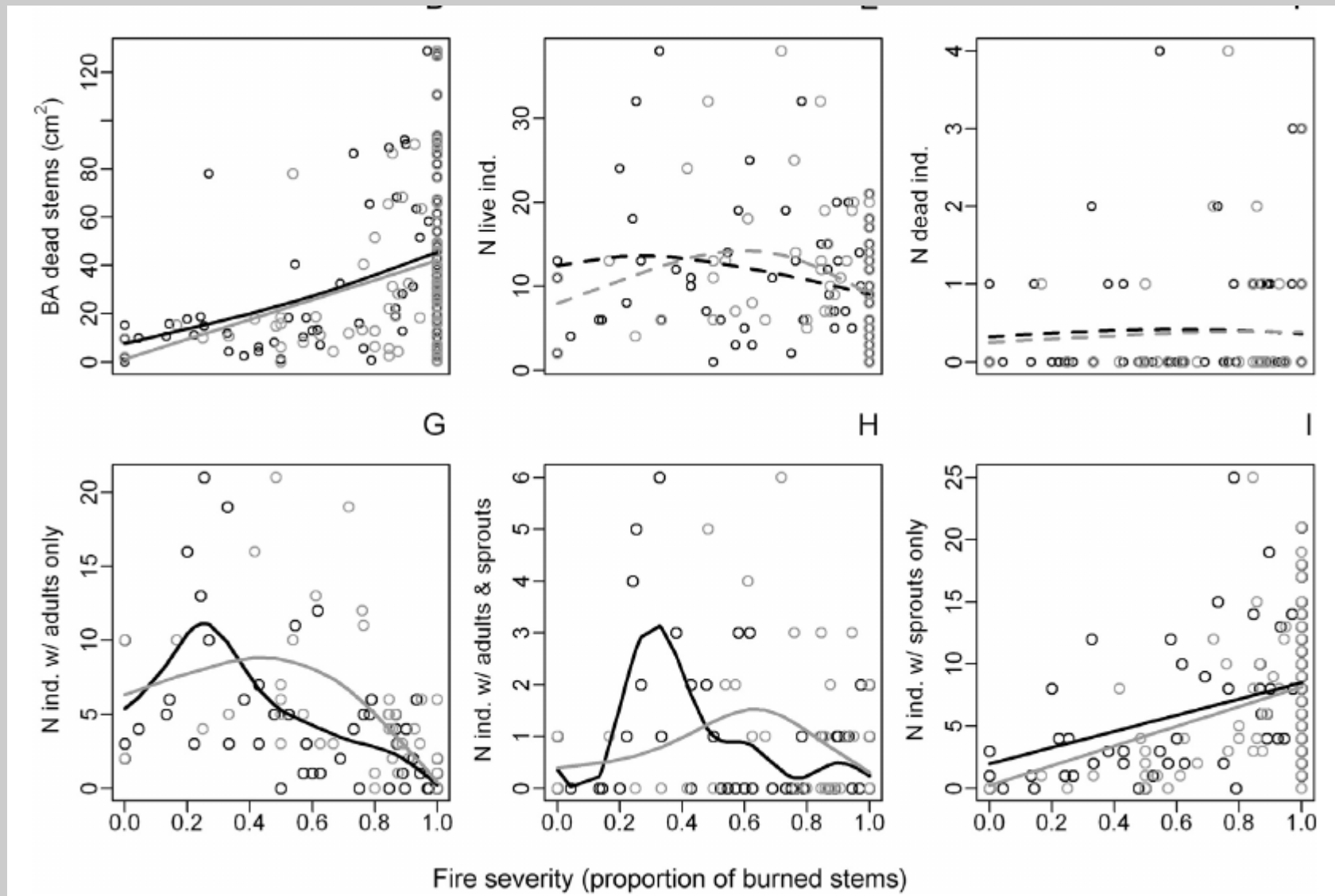
Dodonov et al. 2016

Têm uma equação definida

Modelos aditivos

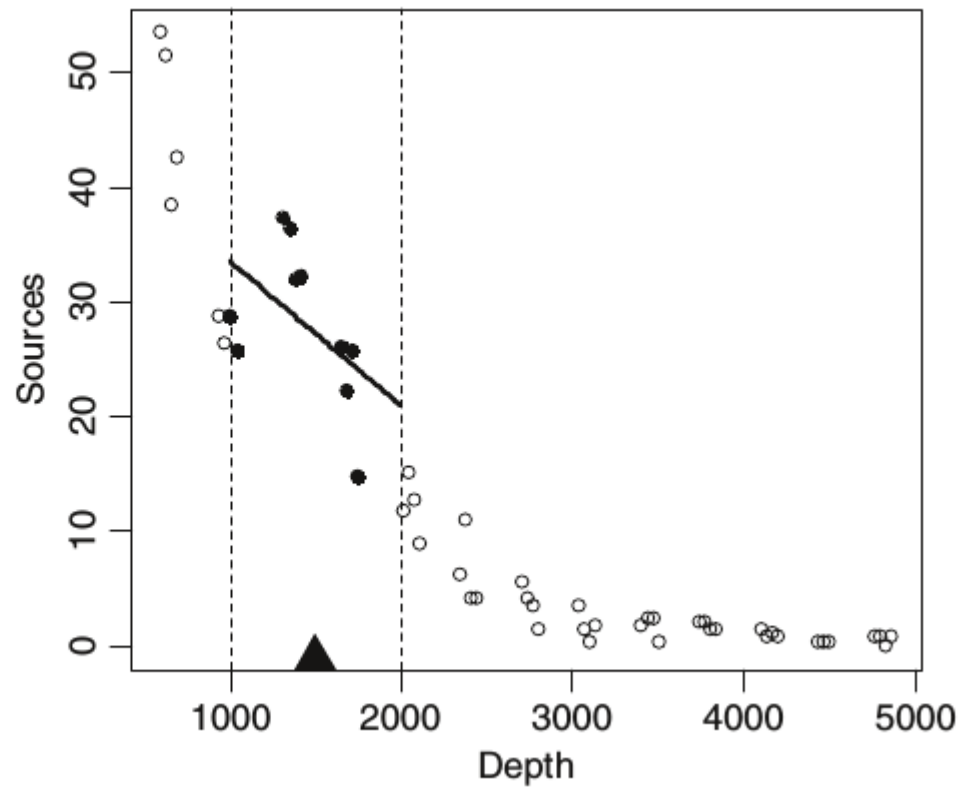
Não têm uma equação
definida

Modelos aditivos



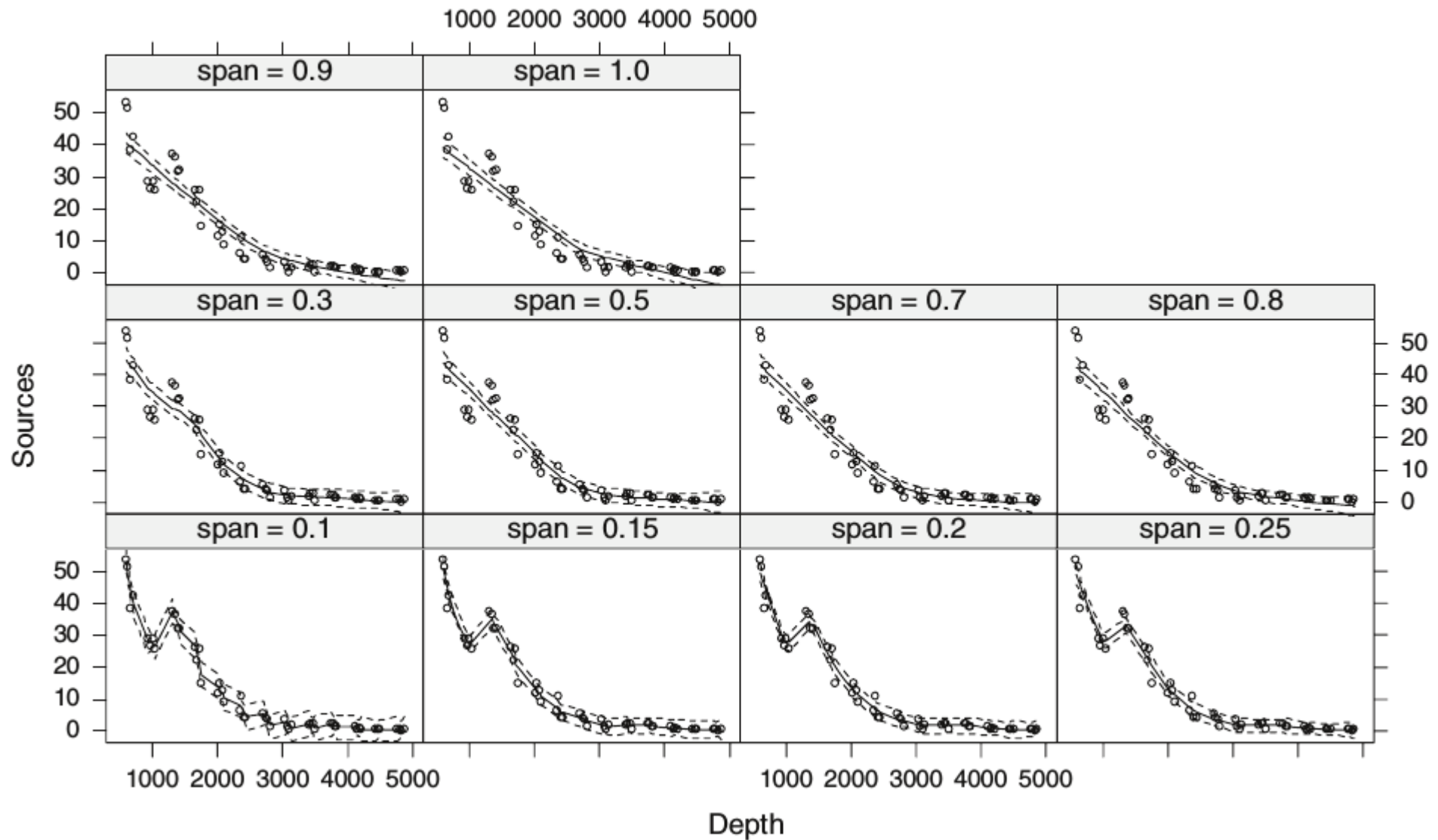
Modelos aditivos

Fig. 3.2 Illustration of LOESS smoothing. We want to predict a value of sources at the target value of 1500 m depth (denoted by the *triangle*). A window around this target value is chosen (denoted by the *dotted lines*) and all points in this window (the *black dots*) are used in the local linear regression analysis



Zuur et al. 2009

Modelos aditivos



Modelos aditivos

$$Y_i = \beta_0 + f(X_i) + \epsilon_i$$

Modelos aditivos

$$Y_i = \beta_0 + f(X_i) + \epsilon_i$$



“Função suavizada”, ou *smoother*

Modelos aditivos

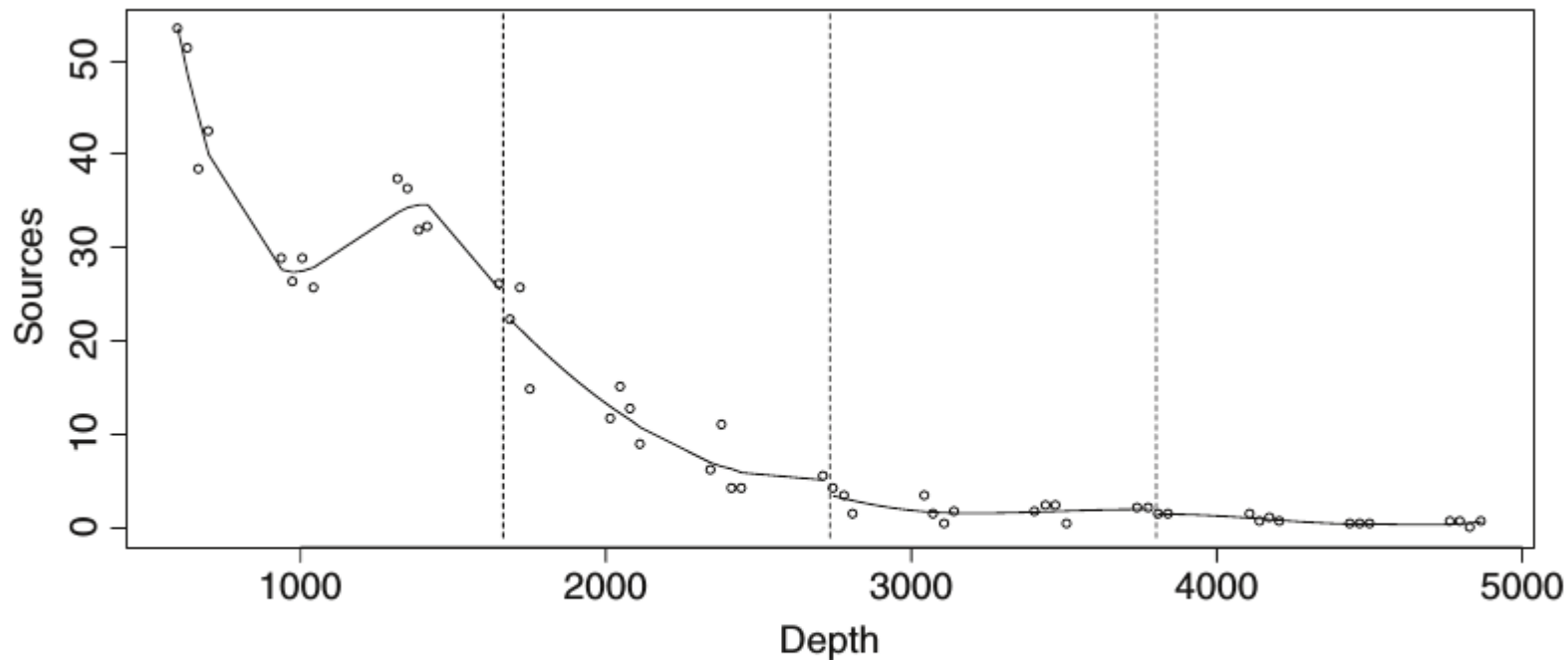
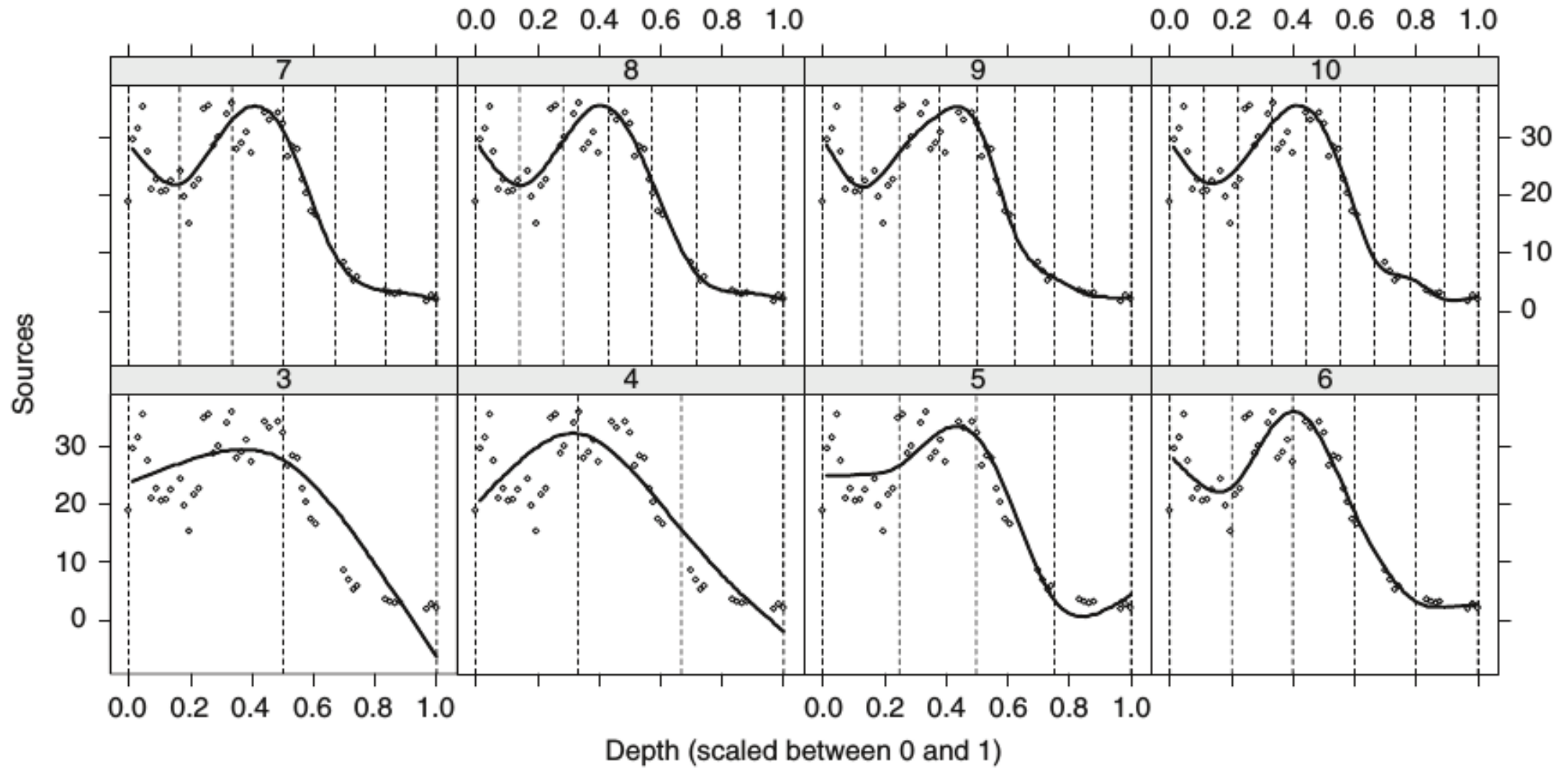


Fig. 3.8 Illustration of fitting a cubic polynomial on four segments of data using the ISIT data from station 19. We arbitrarily choose four segments along the depth gradient. The dotted lines mark these segments, and the line in each segment is the fit from the cubic polynomial model. R code to create this graph can be found on the book website

Modelos aditivos

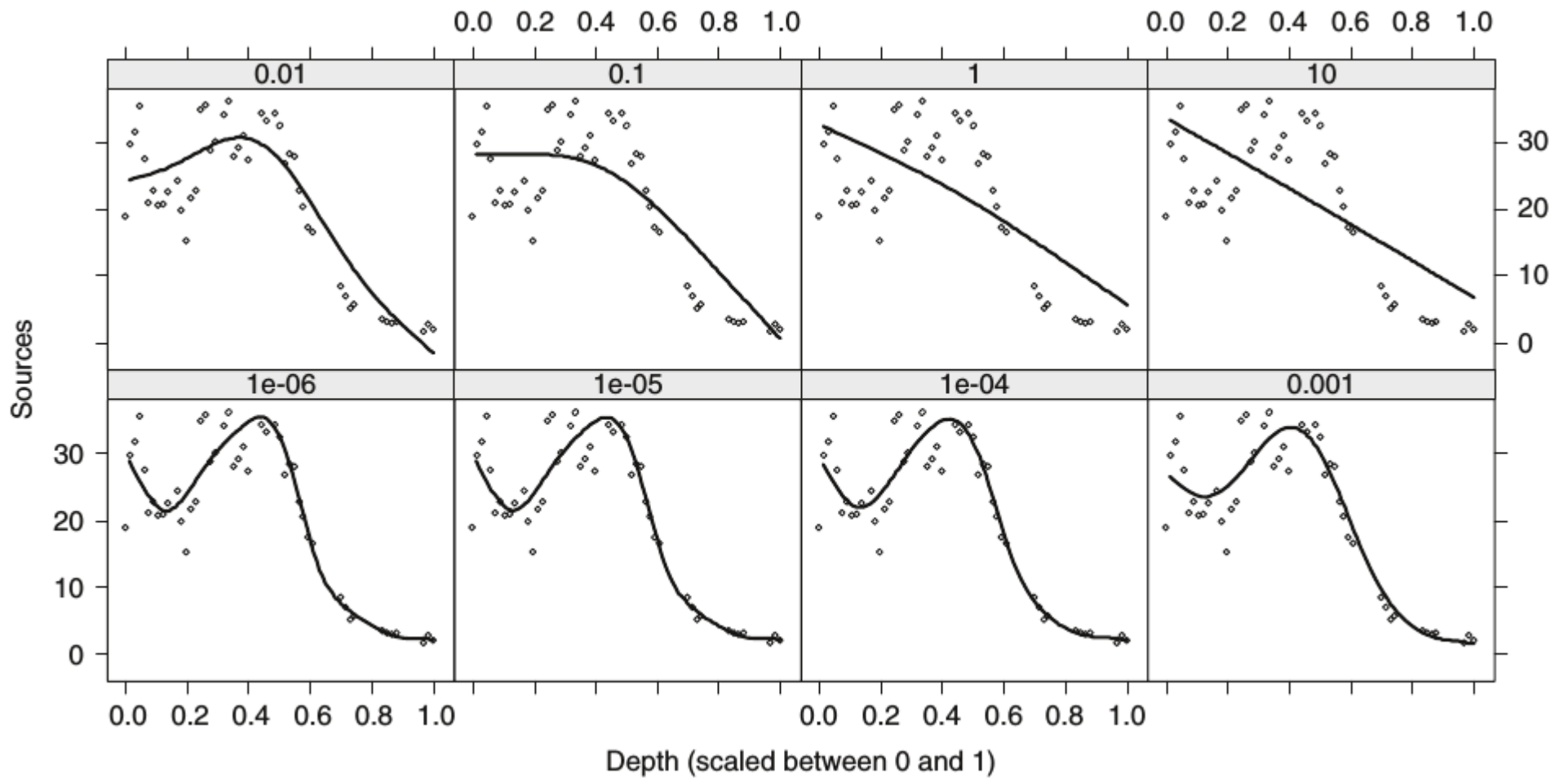


Modelos aditivos

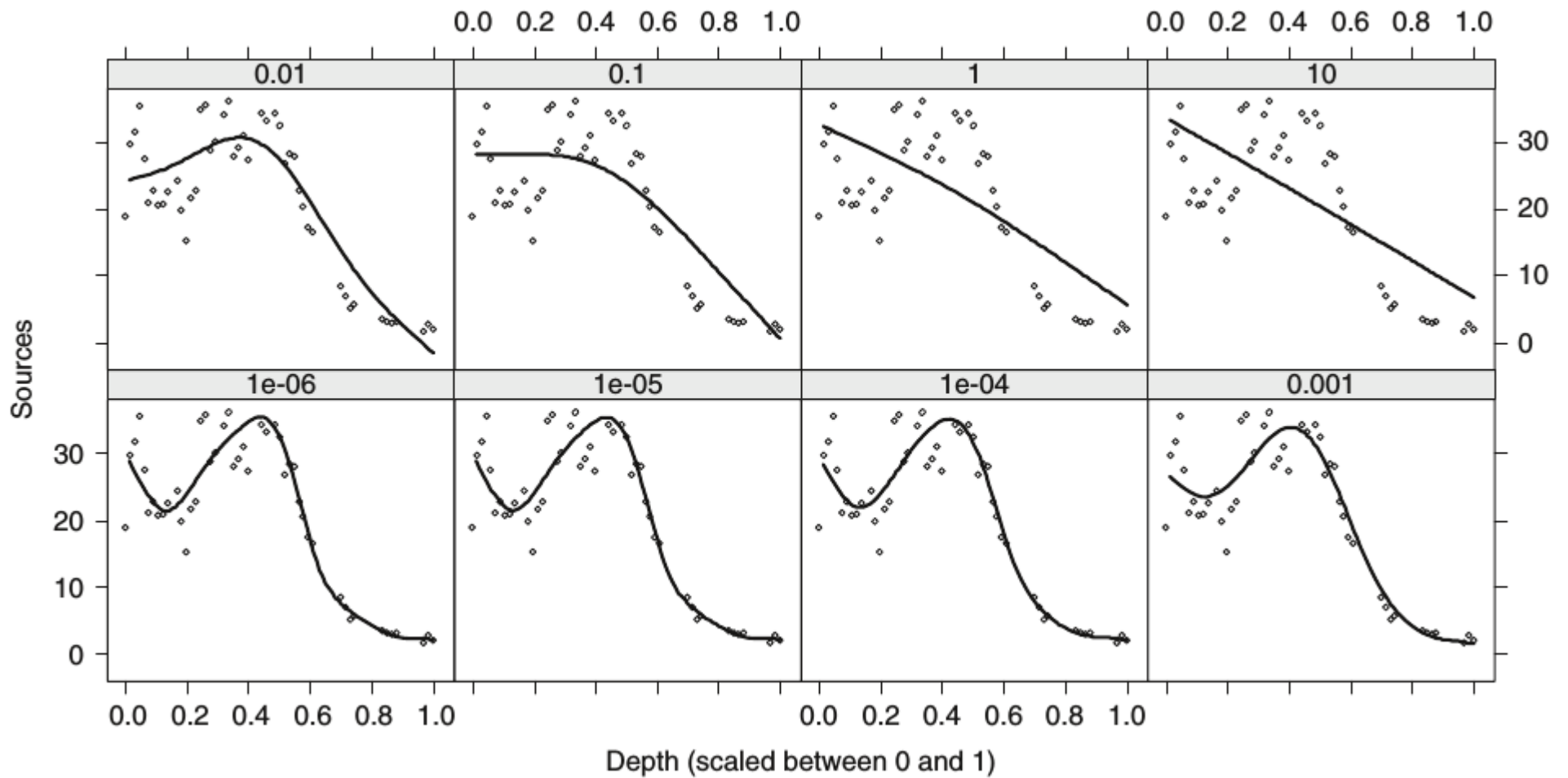
Smoothing splines: É provavelmente o modelo aditivo mais usado. Modelos mais complexos (menos lineares) são penalizados.

λ : o quanto modelos mais complexos são penalizados.

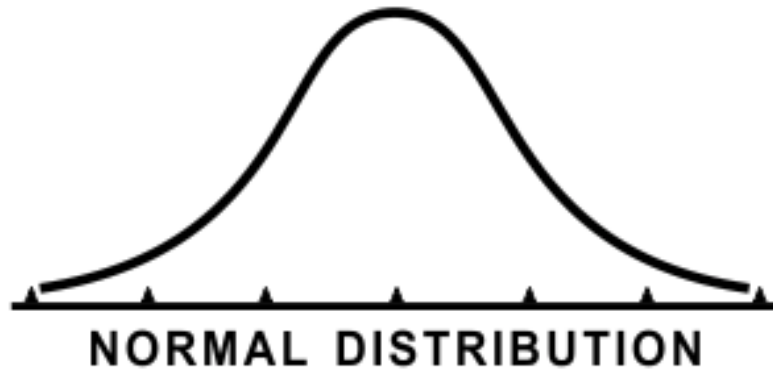
Modelos aditivos



Modelos aditivos



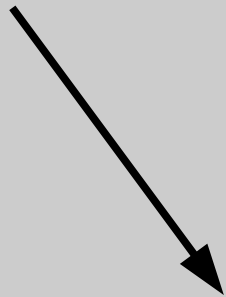
Desvios de normalidade



Desvios de normalidade

$$Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

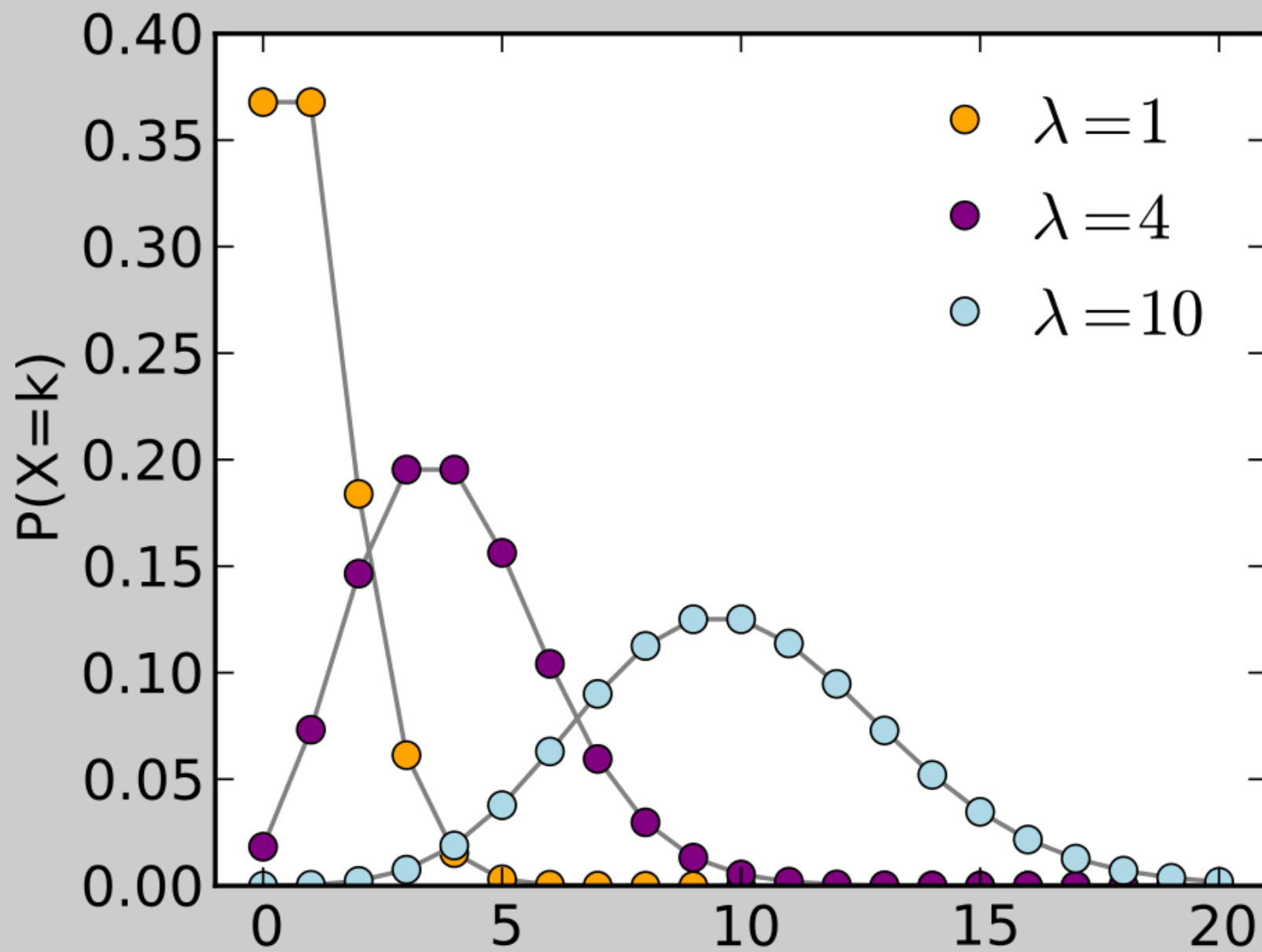


Resíduos seguem distribuição normal

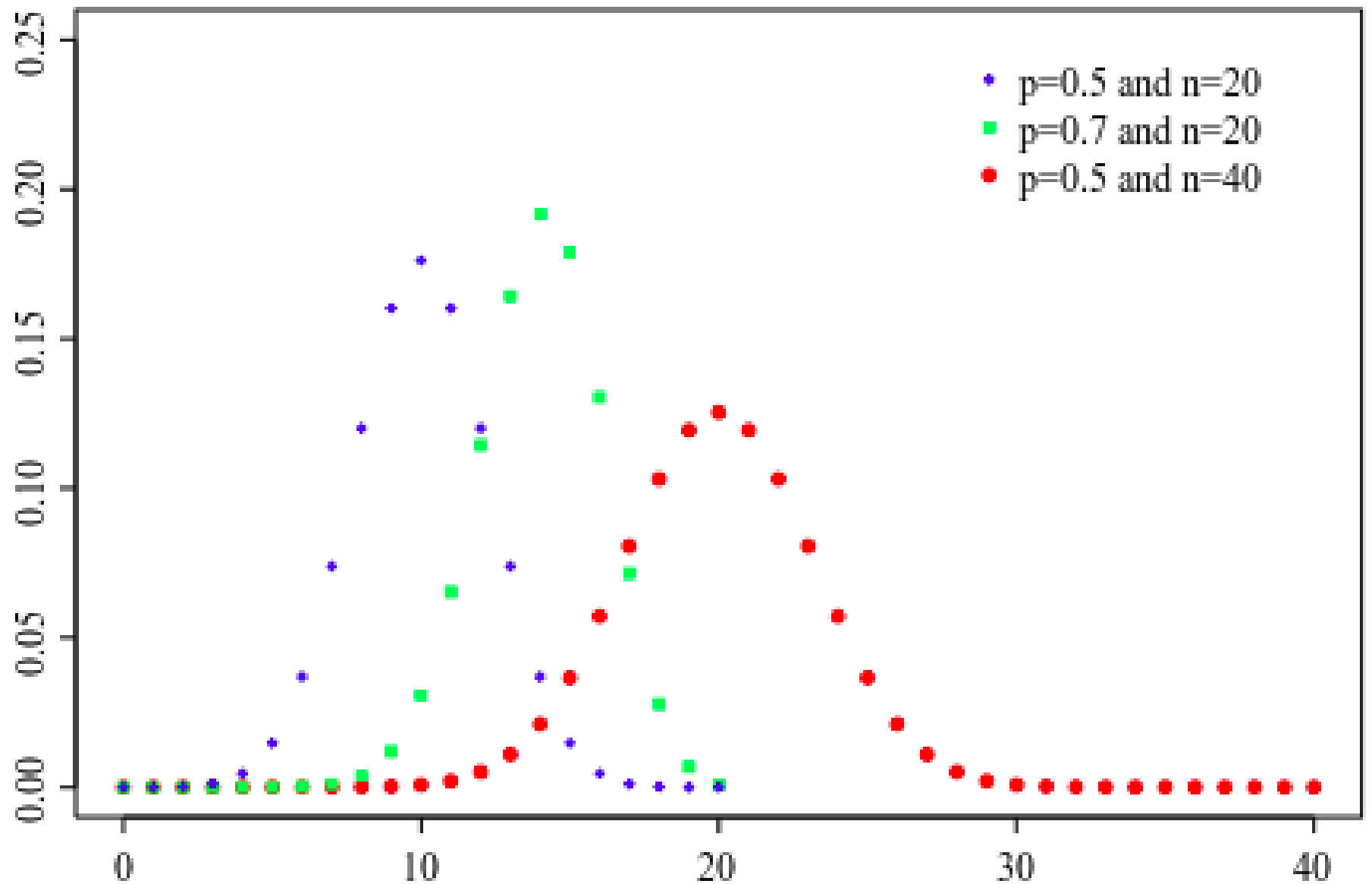
Desvios de normalidade

GLMs (Generalized Linear Models):
Generalizam os modelos lineares para
distribuições não-normais de resíduos

Poisson



Binomial



Gamma

