

Modelos de Regressão Não Linear

Dr. Walmes Marques Zeviani

walmes@ufpr.br

Departamento de Estatística - UFPR

Laboratório de Estatística e Geoinformação (LEG)

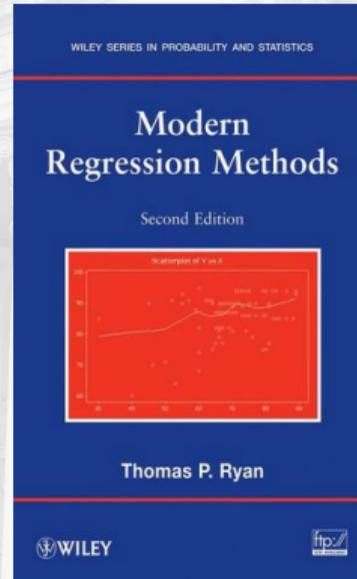


Agradecimentos



- ▶ Comissão organizadora do I Workshop de Regressão Não Linear.
- ▶ Aos participantes do evento.
- ▶ Colegas do DEST pelo incentivo e colaboração.

O que é regressão?



História

- ▶ Francis Galton (1822 – 1911) era primo de Charles Darwin e tinha competências em medicina e matemática.
- ▶ Galton era fascinado pela biometria humana e herdabilidade.
- ▶ Inventou a identificação pela impressão digital.
- ▶ Estudou dados de altura dos pais e filhos adultos (1886).
- ▶ “Law of universal regression”.
- ▶ Altura do filhos *regrediam* para a média.
- ▶ Com ajuda de Karl Pearson, ajustou a reta.
- ▶ A técnica recebeu o nome **regressão**.

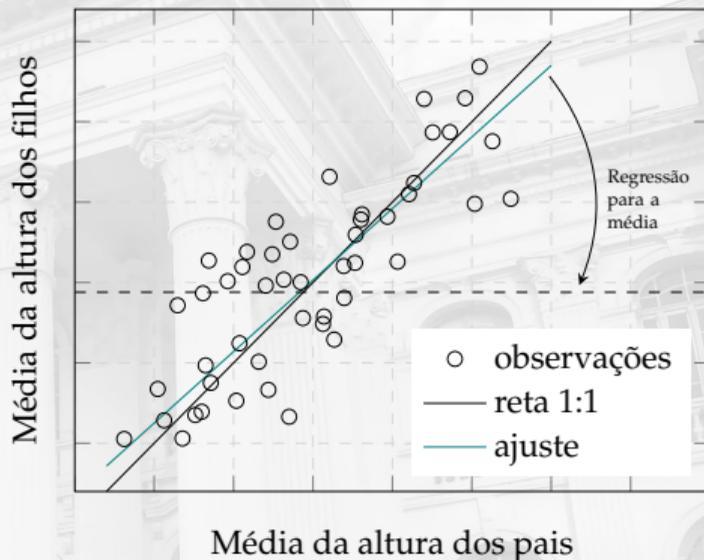


Figura 1: A regressão para a média de Galton.

Motivação

- ▶ Estudar a associação entre uma variável y e um conjunto de variáveis $x_i, i \geq 1$.
- ▶ Fazer a previsão de y .





Figura 2: Representação das variáveis relacionadas com o valor de venda de um imóvel.



Figura 3: Representação das variáveis relacionadas com o rendimento escolar de uma criança.

Mais exemplos

- ▶ Tempo de uma substância no sangue:
 - ▶ Concentração aplicada;
 - ▶ Paciente (idade, sexo, pressão, hábitos).
- ▶ Produtividade de um pomar:
 - ▶ Fertilizante (dose, tipo, frequência);
 - ▶ Preparo do solo;
 - ▶ Manejo de pragas e doenças;
 - ▶ Poda e tratos culturais;
 - ▶ Irrigação;
 - ▶ Clima.
- ▶ Desempenho de time de futebol:
 - ▶ Quantidade, intensidade e qualidade de treino;
 - ▶ Nutrição e preparo físico dos jogadores;
 - ▶ Entozamento entre jogadores;
 - ▶ Estratégia de jogo;
 - ▶ Experiência dos jogadores e equipe;
 - ▶ Condição do gramado;
 - ▶ Apoio da torcida.

Forma genérica



Figura 4: Representação esquemática genérica de um modelo de regressão.

Organização do modelo de regressão

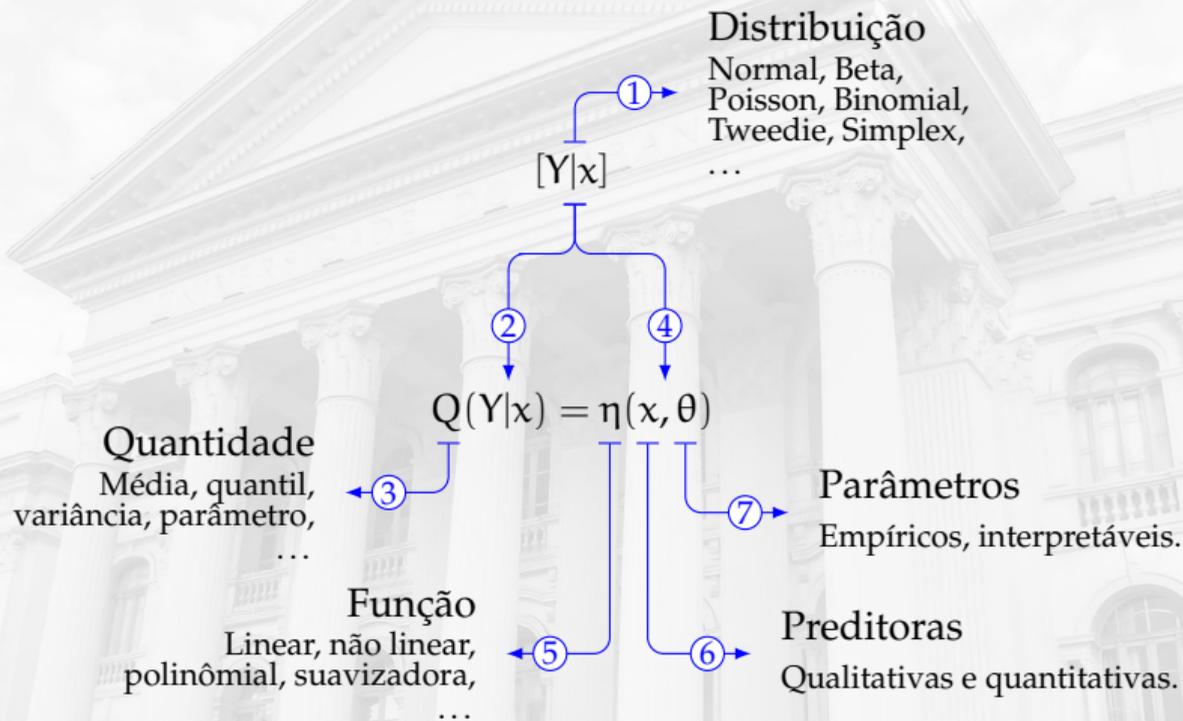
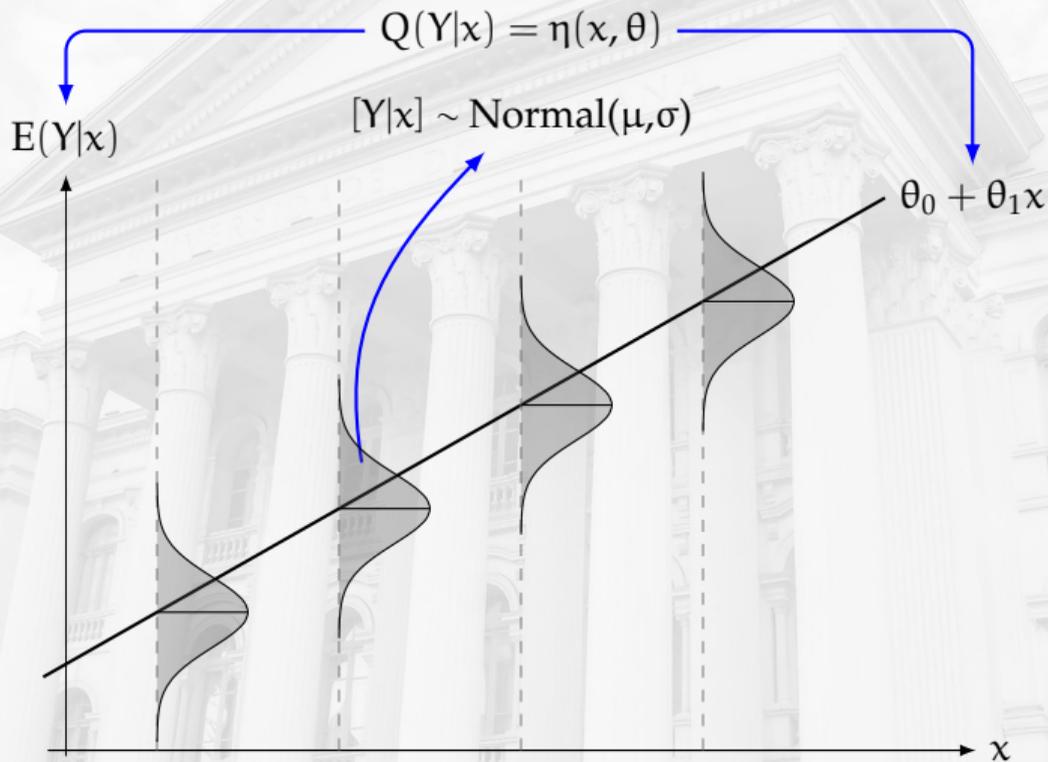
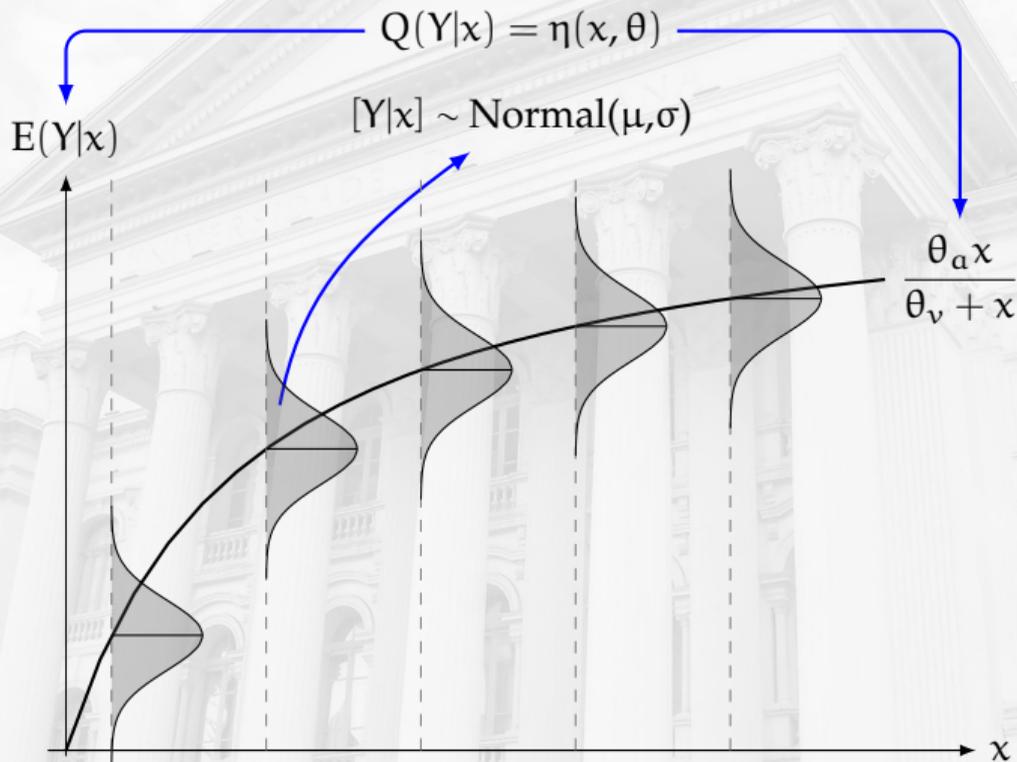


Figura 5: Representação esquemática da construção de um modelo de regressão.

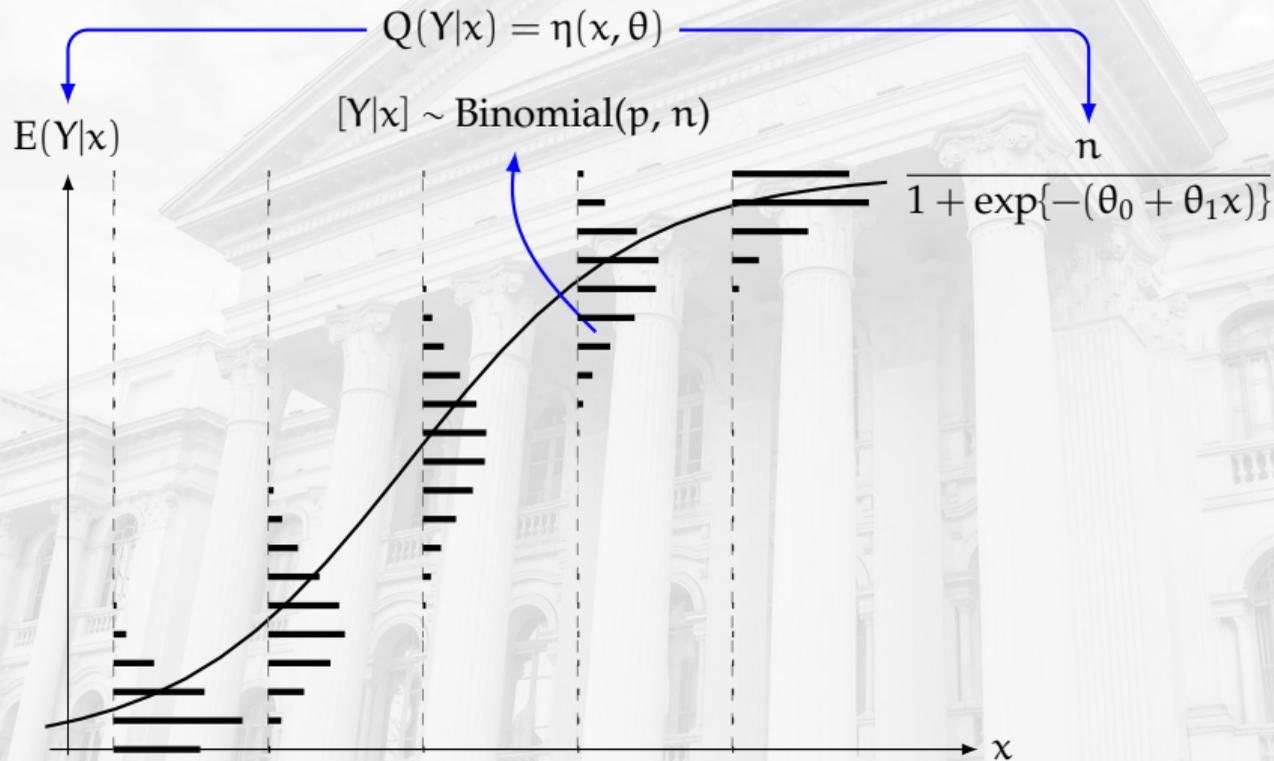
Regressão Linear Simples



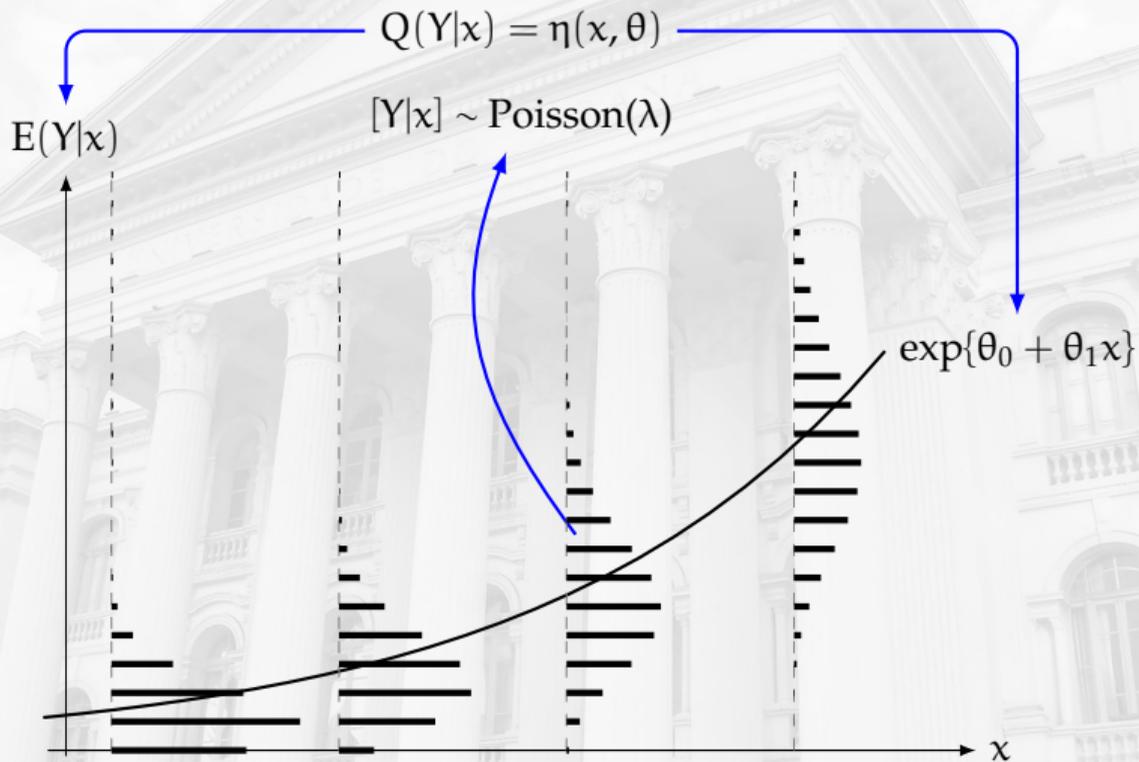
Regressão Não Linear



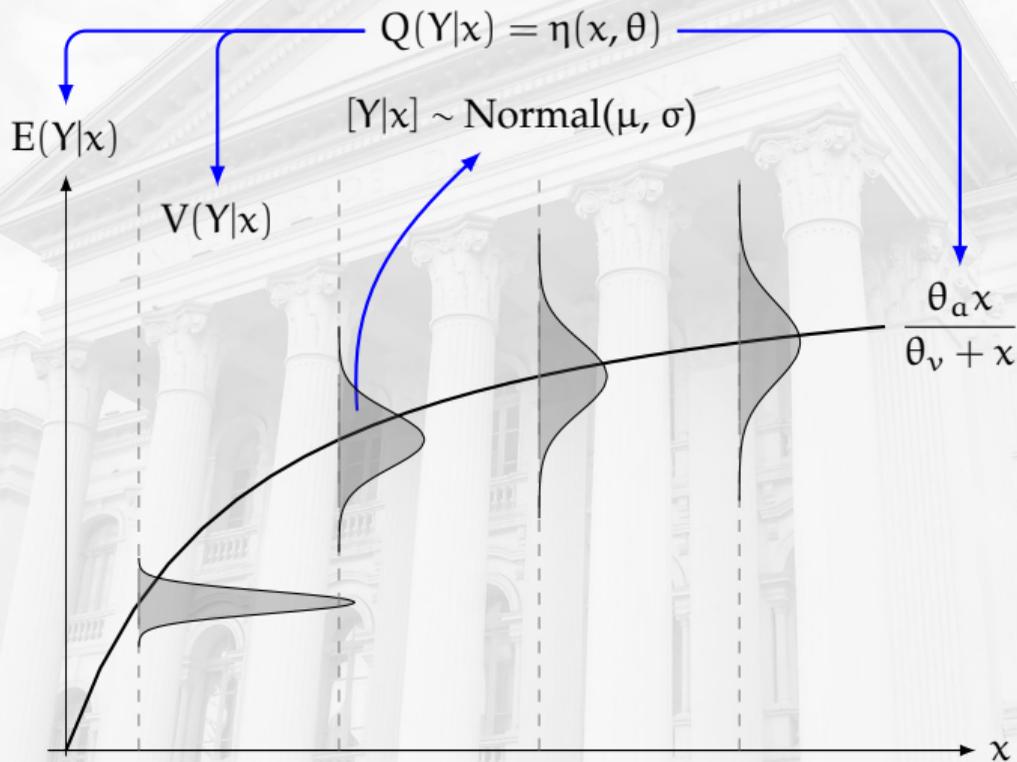
Regressão Binomial



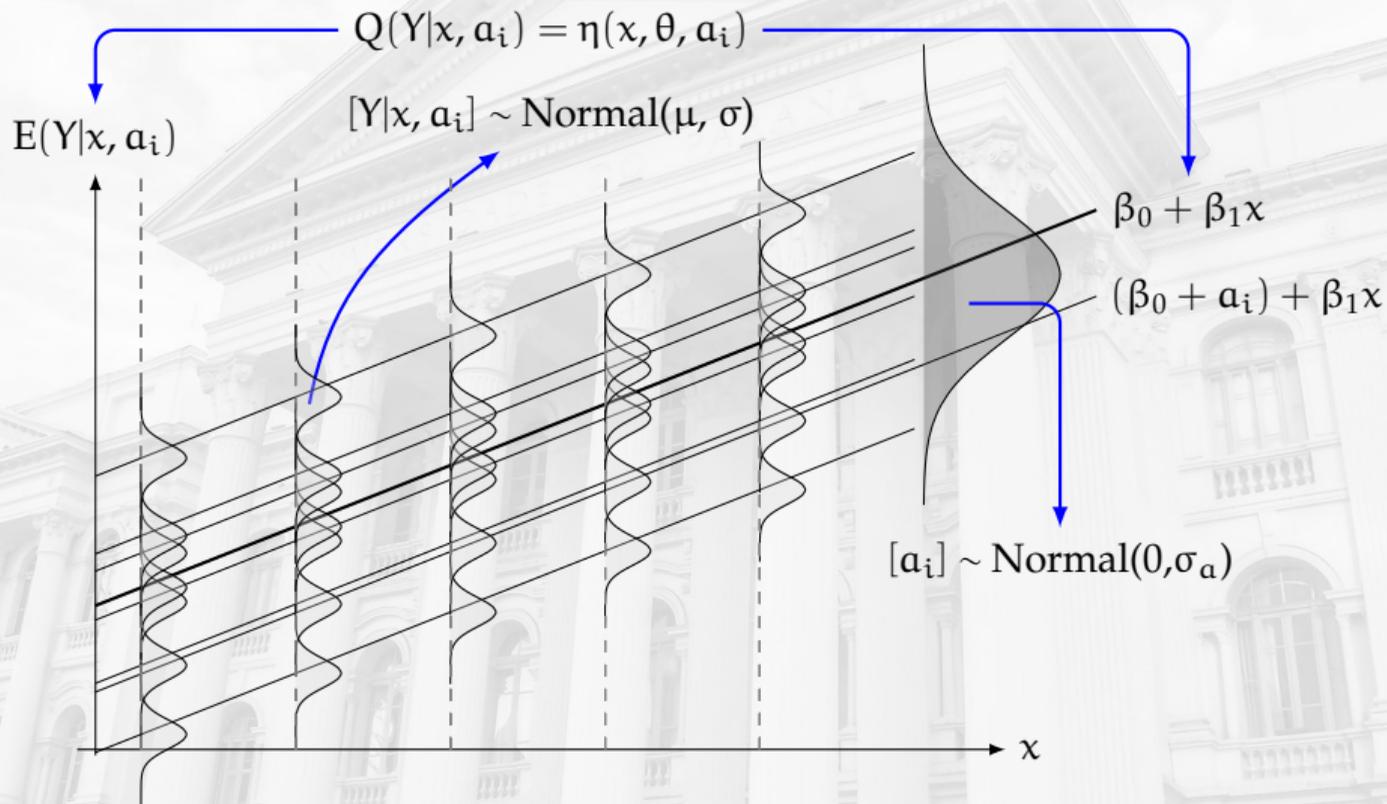
Regressão Poisson



Regressão Não Linear com Variância Não Constante



Regressão Linear com Efeito Aleatório



Especificação

▶ Regressão linear simples

Resposta ou variável
dependente [y]

Preditora ou variável
independente [x]

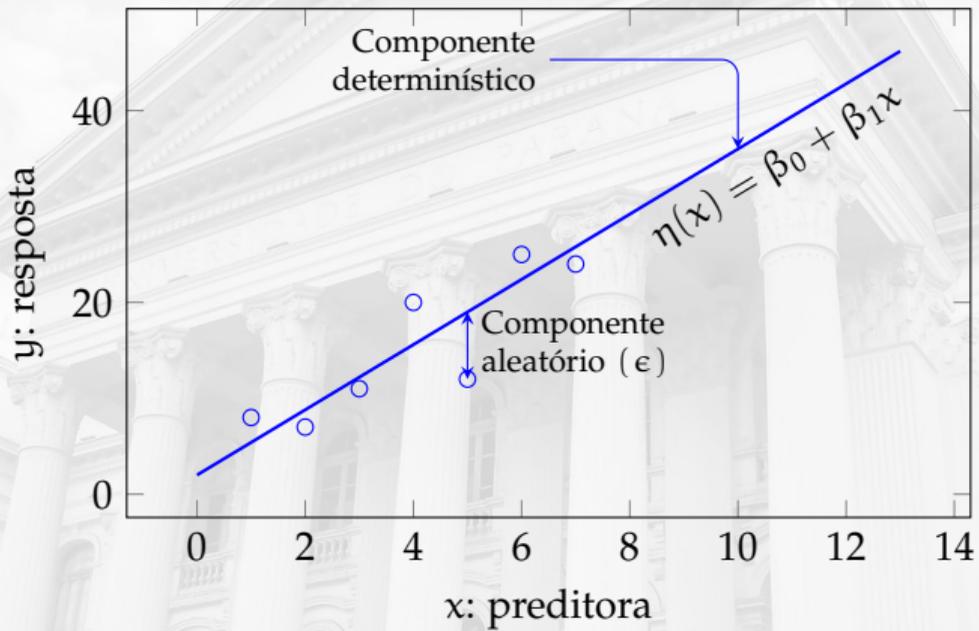
$$Y = \beta_0 + \beta_1 x + \epsilon$$

Intercepto [y]

Erro aleatório [y]
 $E(\epsilon) = 0,$
 $V(\epsilon) = \sigma^2$

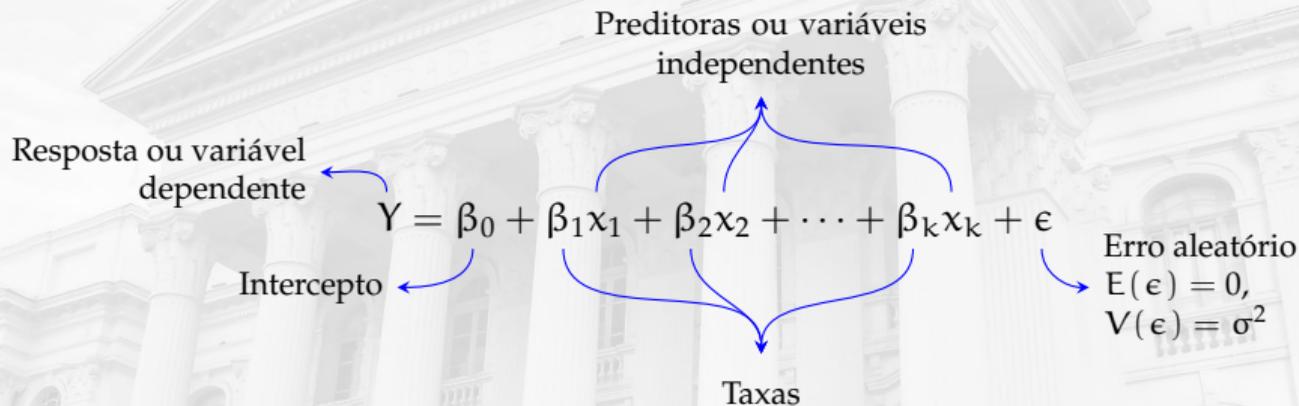
Taxa [$y x^{-1}$]

- ▶ $E(Y|x) = \beta_0 + \beta_1 x.$
- ▶ $V(Y|x) = V(\epsilon) = \sigma^2.$



Especificação

► Regressão linear múltipla



- $E(Y|x_i, i = 1, \dots, k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k.$
- $V(Y|x_i, i = 1, \dots, k) = V(\epsilon) = \sigma^2.$

Representação matricial

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

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

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}_{\substack{n \times p \\ (p = k + 1)}} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{p \times 1} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n \times 1}$$

Estimação

- ▶ Critério de mínimos quadrados (ordinários)

$$\begin{aligned} \text{SSE}(\beta) &= \sum_{i=1}^n (y_i - x_i^\top \beta)^2 \\ &= (y - X\beta)^\top (y - X\beta) = \|y - X\beta\|^2 \end{aligned}$$

- ▶ Estimador

$$\begin{aligned} \hat{\beta} &= \arg \min_{\beta \in \mathbb{R}^p} \text{SSE}(\beta) \\ &= (X^\top X)^{-1} X^\top y \end{aligned}$$

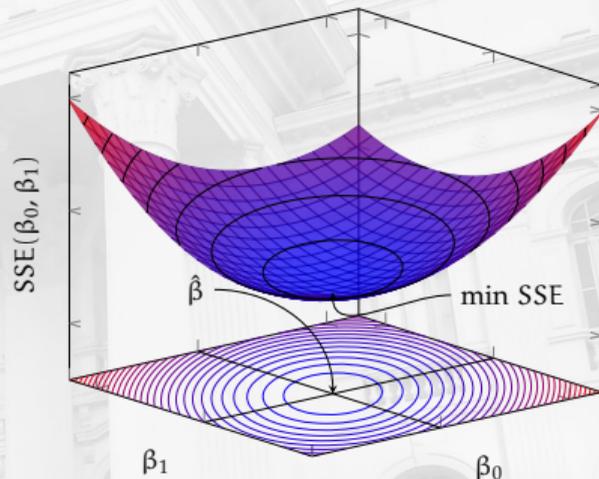
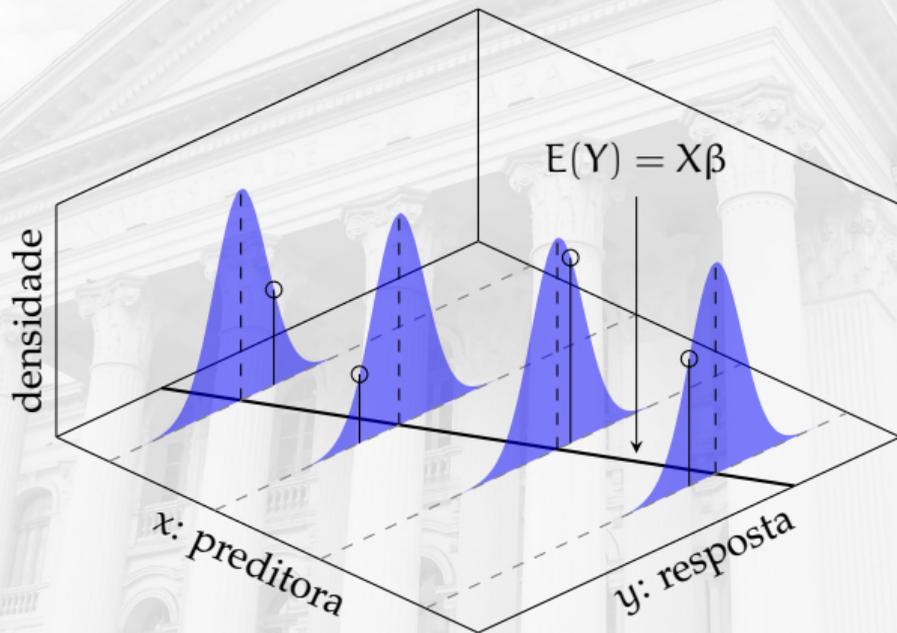


Figura 6: A superfície e mínimos quadrados.

Estimação baseada na verossimilhança



- ▶ Se $Y \sim \text{Normal}(\mu = X\beta, \sigma^2 = \sigma^2)$, então a log-verossimilhança é

$$\ell(\beta, \sigma) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{(y - X\beta)^\top (y - X\beta)}{2\sigma^2}. \quad (1)$$

- ▶ Os estimadores correspondem ao máximo da $\ell(\theta)$, $\theta = (\beta, \sigma^2)^\top$,

$$\frac{\partial \ell(\theta)}{\partial \theta^\top} = 0, \quad \hat{\beta} = (X^\top X)^{-1} X^\top y, \quad \hat{\sigma}^2 = \frac{\text{SSE}}{n}. \quad (2)$$

Medidas de ajuste

- ▶ $R^2 = 1 - \frac{SSE(\beta)}{SSE(\beta_0)} = 1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|}$.
- ▶ $R^2_{\text{adj}} = 1 - \frac{n-1}{n-p} (1 - R^2)$.
- ▶ $\text{PRESS} = \sum_{i=1}^n (y - \hat{y}_{i(-i)})^2$, menor é melhor;
- ▶ Log-verossimilhança (maior é melhor)

$$\begin{aligned} \ell &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}^2) - \frac{\|y - X\hat{\beta}\|}{2\hat{\sigma}^2} \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(SSE/n) - \frac{n}{2}, \quad \hat{\sigma}^2 = SSE/n = \|y - X\hat{\beta}\|/n. \end{aligned}$$

- ▶ $\text{AIC} = 2(p+1) - 2\ell$, menor é melhor;
- ▶ $\text{BIC} = \log(n)(p+1) - 2\ell$, menor é melhor;

Medidas de diagnóstico e influência

- ▶ Matriz de projeção

$$\hat{y} = Hy, \quad H = X(X^T X)^{-1}X^T,$$

H é simétrica e indepotente. O posto de H é $\text{tr}(H) = p$.

- ▶ Alavancagem (*leverage*)

$$h_i = H_{ii}$$

$$h = \text{diag}(H).$$

- ▶ Resíduos ordinários, $V(\hat{e}) = \sigma^2(I - H)$,

$$\hat{e}_i = y_i - \hat{y}_i$$

$$\hat{e} = y - \hat{y}$$

$$\hat{e} = y - X\hat{\beta}.$$

- ▶ Resíduos padronizados (ou internamente studentizados),

$$r_i = \frac{\hat{e}_i}{s(\hat{e}_i)} = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_i}}$$

- ▶ Resíduos studentizados (ou externamente studentizados),

$$t_i = \frac{\hat{e}_i}{s(\hat{e}_{i(-i)})} = \frac{\hat{e}_i}{\hat{\sigma}_{-i}\sqrt{1-h_i}}$$
$$\hat{\sigma}_{-i}^2 = \frac{(n-p)\hat{\sigma}^2 - \frac{\hat{e}_i^2}{1-h_i}}{(n-1)-p}$$

► Distância de Cook

$$D_i = \frac{(\hat{y} - \hat{y}_{i(-i)})^\top (\hat{y} - \hat{y}_{i(-i)})}{p \hat{\sigma}^2} = \frac{1}{p} \cdot \frac{h_i}{(1 - h_i)} \cdot \frac{\hat{e}_i^2}{\hat{\sigma}^2 (1 - h_i)}.$$

► DFfits

$$dffits_i = \frac{\hat{y}_i - \hat{y}_{i(-i)}}{\hat{\sigma}_{-i} \sqrt{h_i}} = t_i \left(\frac{h_i}{1 - h_i} \right)^{1/2}.$$

► DFbetas

$$dbetas_i = \frac{\hat{\beta} - \hat{\beta}_{-i}}{\hat{\sigma}_{-i} \sqrt{\text{diag}((X^\top X)^{-1})}}$$
$$\hat{\beta}_{-i} = \hat{\beta} - \frac{\hat{e}_i}{1 - h_i} \cdot (X^\top X)^{-1} x_i.$$

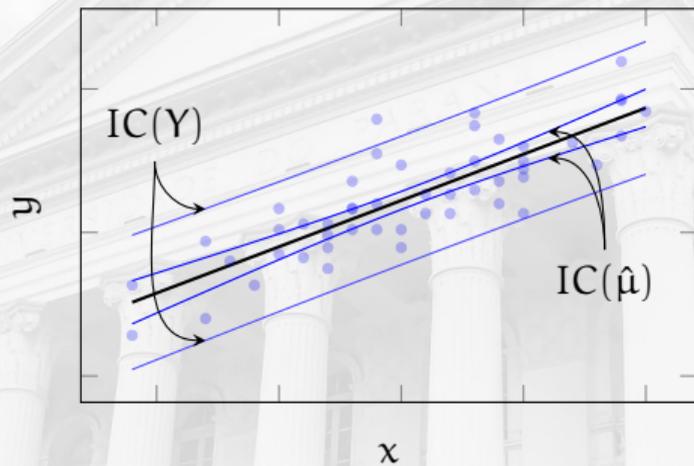


Figura 7: Bandas de confiança para $\hat{\mu}$ e de predição para Y .

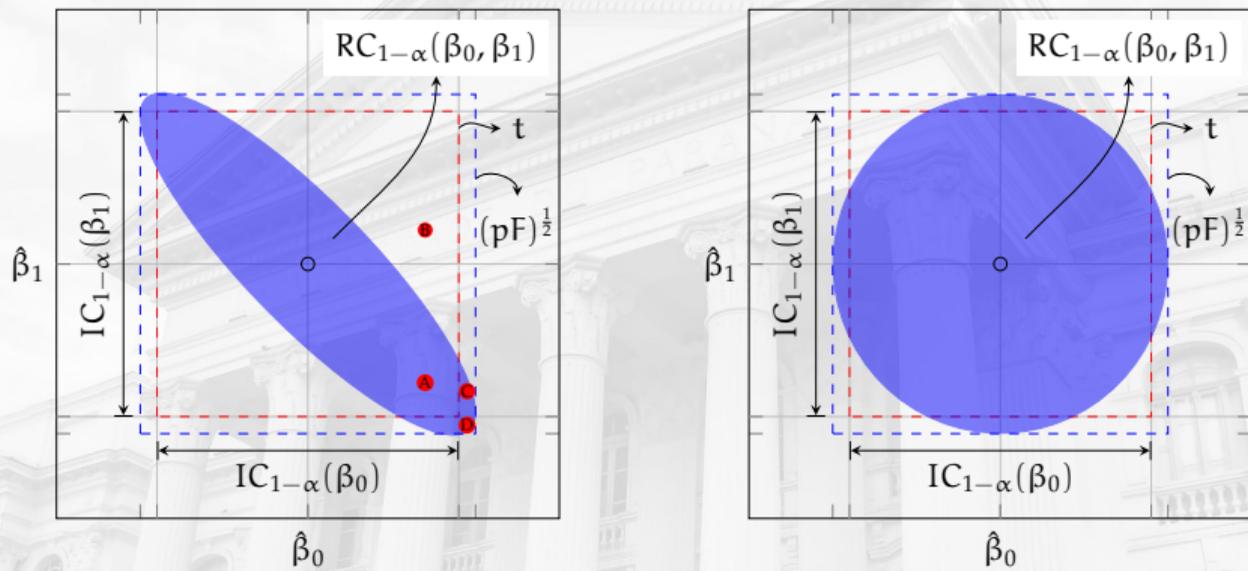


Figura 8: Região e intervalos de confiança para β .

Modelos de regressão não lineares

Benefícios

- ▶ Baseados em teoria ou princípios que dão uma relação funcional mais específica entre y e x ;
- ▶ Parâmetros são interpretáveis;
- ▶ São parsimoniosos;
- ▶ Podem ser feitas previsões fora do domínio observado de x ;

Custos

- ▶ Requerem procedimentos iterativos de estimação;
- ▶ Métodos de inferência são aproximados;

Definição

Linear nos parâmetros

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

$$\frac{\partial \eta}{\partial \theta_0} = 1,$$

$$\frac{\partial \eta}{\partial \theta_1} = x,$$

$$\frac{\partial \eta}{\partial \theta_2} = x^2.$$

Não linear nos parâmetros

$$\eta(x, \theta) = \theta_a (1 - \exp\{-\theta_e (x - \theta_c)\}).$$

$$\frac{\partial \eta}{\partial \theta_a} = 1 - \exp\{-\theta_e (x - \theta_c)\}$$

$$\frac{\partial \eta}{\partial \theta_e} = -\theta_a (\theta_c - x) \exp\{-\theta_e (x - \theta_c)\}$$

$$\frac{\partial \eta}{\partial \theta_c} = -\theta_a \theta_e \exp\{-\theta_e (x - \theta_c)\}.$$

Pontos característicos e formas

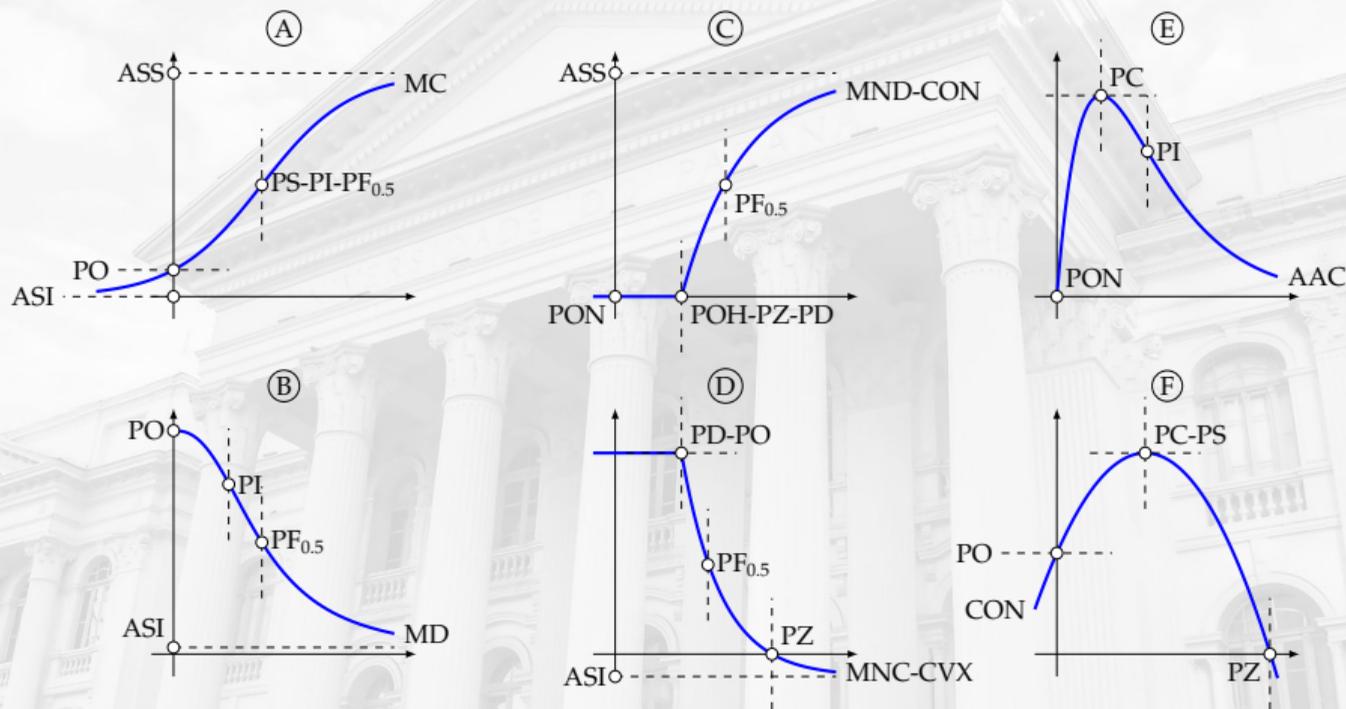


Figura 9: Funções não lineares com destaque para os pontos característicos e formas.

Determinação das unidades de medida (dimensionalidade)

► Modelo Michaelis-Menten

$$y = \frac{\theta_a x}{\theta_v + x}$$

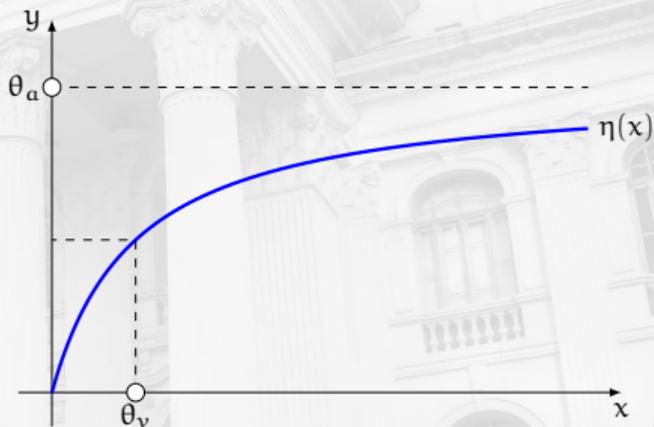
Resposta [y] ←

Assíntota [y] ←

Meia vida [x] ←

Preditora [x] ←

- $\lim_{x \rightarrow \infty} \eta(x) = \theta_a.$
- $\eta(x = \theta_v) = \theta_a/2.$



Aplicações

- ▶ Detalhes do procedimento de estimação.
- ▶ Medidas de diagnóstico e qualidade de ajuste.
- ▶ Métodos de inferência.