



# Universidade de São Paulo

## Estimativa de Parâmetros I

## Método dos Quadrados Mínimos

**Prof. Adriano Azevedo Filho, Ph.D.**  
**azevedofilho@usp.br**  
**USP/ESALQ/DEAS**

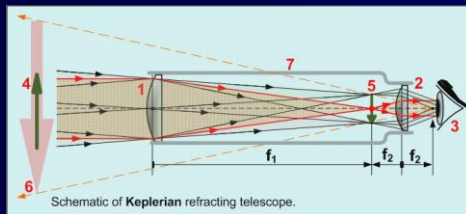
# Método dos Quadrados Mínimos (Ordinário)

- Em inglês: Ordinary Least Squares (OLS)
- Problemas que motivaram seu desenvolvimento:

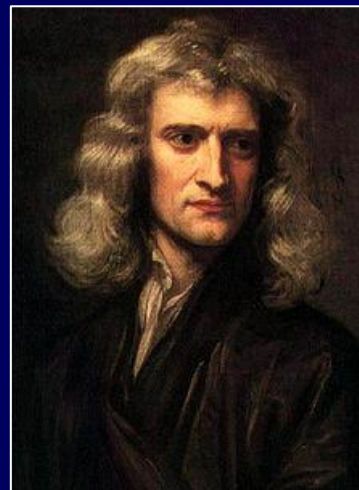
Estimativa de parâmetros de modelos na Física e Astronomia (1600 – início dos 1800)



Kepler (1571-1630)



telescópio de Kepler (1611)



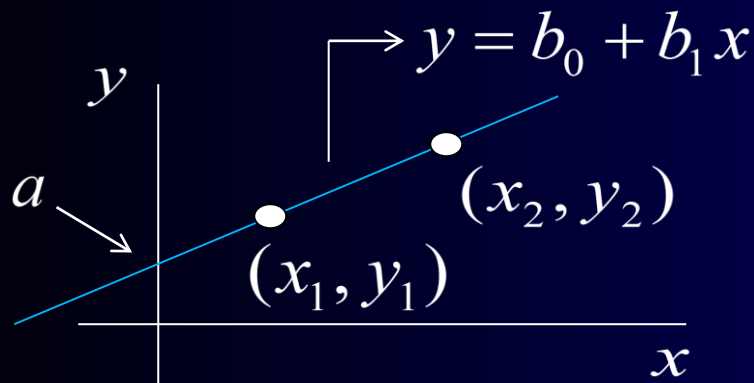
Newton (1642-1727)

Suponha que uma teoria estabelece que

Ex.  $y = b_0 + b_1 x$  como “estimar”  $b_0$  e  $b_1$  ?

Se o modelo/teoria está certa e sem erros de medida:

basta 2 observações de  $x$  e  $y$  :  $(x_1, y_1)$   $(x_2, y_2)$



$$\begin{cases} b_0 + b_1 x_1 = y_1 \\ b_0 + b_1 x_2 = y_2 \end{cases}$$

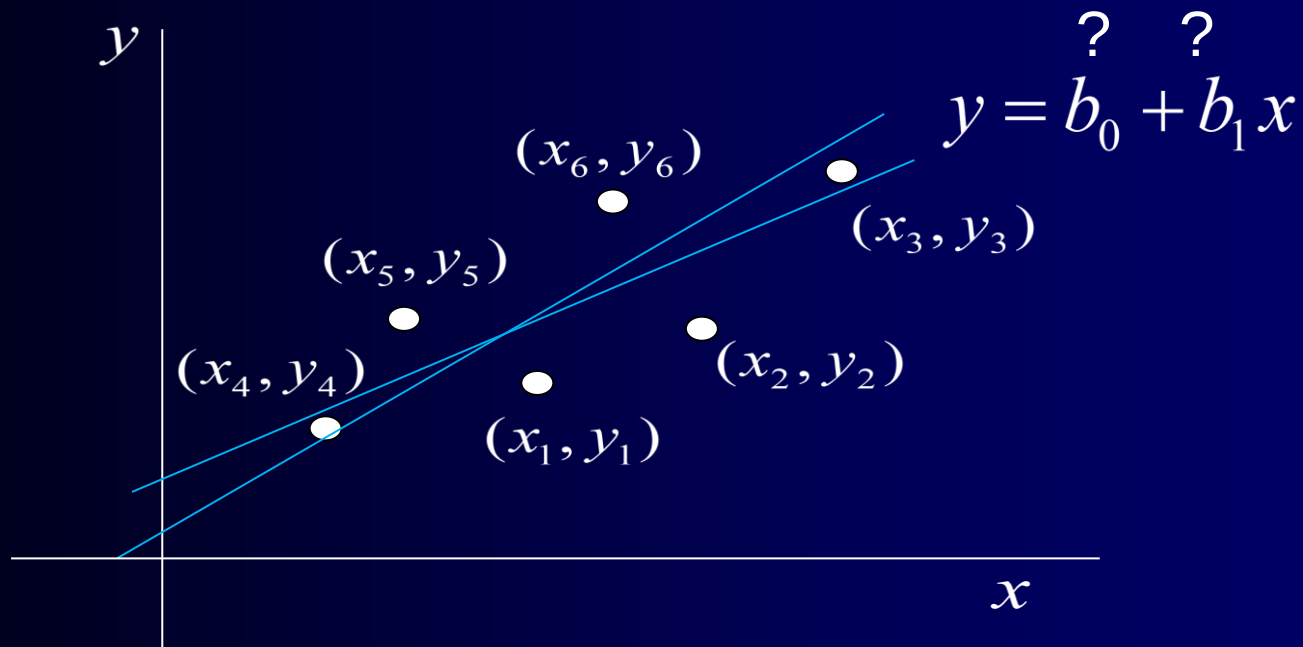
2 equações e  
2 incógnitas

$$b_1 = \frac{y_1 - y_2}{x_1 - x_2}$$

$$b_0 = y_1 - b_1 x_1$$

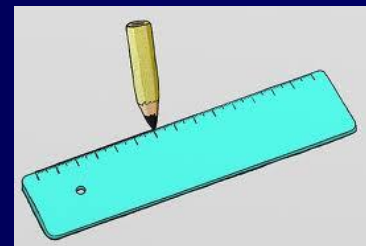
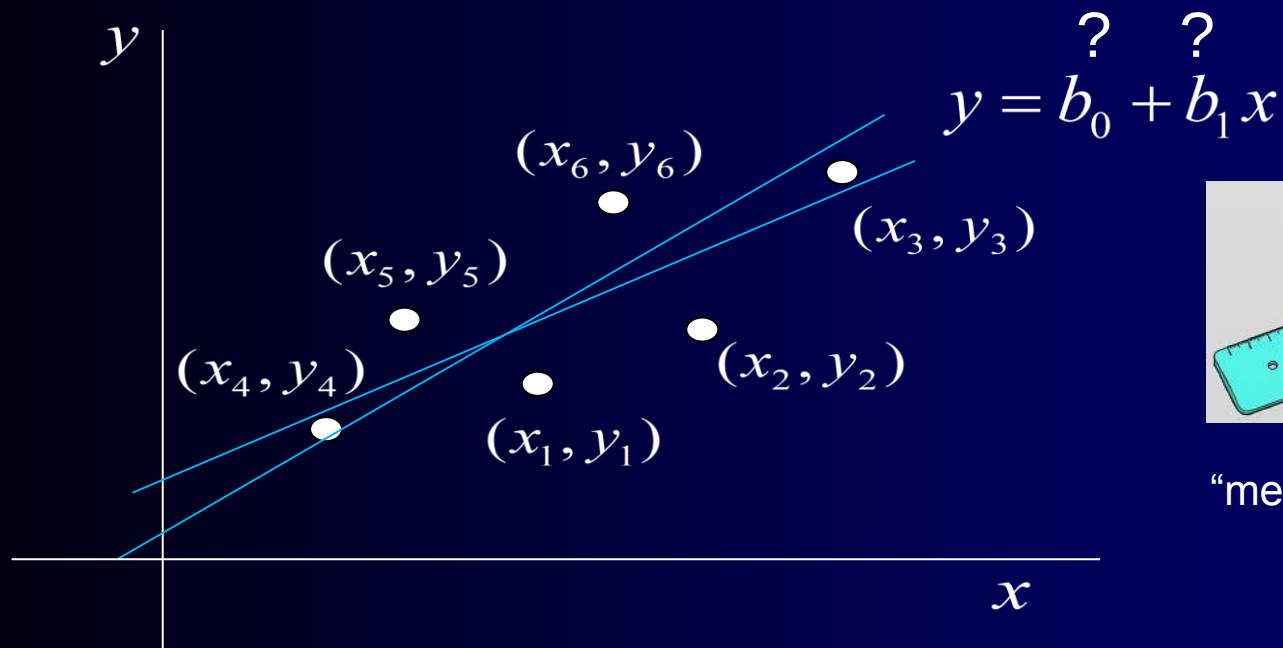
Solução exata !

E se tivermos erros de medida e mais de 2 observações ?



# E se tivermos erros de medida e mais de 2 observações ?

Alternativa 1: “no olhometro com régua” (problemas com repetibilidade e mais de 2 variáveis)



“meio dos pontos?”

# E se tivermos erros de medida e mais de 2 observações ?

Alternativa 2: sistema de equações (problema de identificação)

“solução” de Mayer (1750)



Mayer (1723-1762)

$$\left\{ \begin{array}{l} b_0 + b_1 x_1 = y_1 \\ b_0 + b_1 x_2 = y_2 \\ b_0 + b_1 x_3 = y_3 \end{array} \right\} + \rightarrow 3b_0 + b_1 \sum_{i=1}^3 x_i = \sum_{i=1}^3 y_i$$

$$\left\{ \begin{array}{l} b_0 + b_1 x_4 = y_4 \\ b_0 + b_1 x_5 = y_5 \\ b_0 + b_1 x_6 = y_6 \end{array} \right\} + \rightarrow 3b_0 + b_1 \sum_{i=3}^6 x_i = \sum_{i=3}^6 y_i$$

2 equações e  
2 incógnitas:  
há “solução” agora!



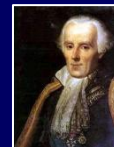
crítica

Laplace: arbitrariedade  
nas equações usadas  
nas 2 somas

6 equações e 2 incógnitas,  
como “solucionar”? sem solução!

# E se tivermos erros de medida e mais de 2 observações ?

Alternativa 2: sistema de equações (problema de identificação)



Laplace (1749-1827)

“solução” de Laplace (1788)

$$\left\{ \begin{array}{l} b_0 + b_1 x_1 = y_1 \\ b_0 + b_1 x_2 = y_2 \\ b_0 + b_1 x_3 = y_3 \\ b_0 + b_1 x_4 = y_4 \\ b_0 + b_1 x_5 = y_5 \\ b_0 + b_1 x_6 = y_6 \end{array} \right. \xrightarrow{\text{orange arrow}} \left\{ \begin{array}{l} \sum_{i=1}^6 b_0 + b_1 \sum_{i=1}^6 x_i = \sum_{i=1}^6 y_i \xrightarrow{\text{blue arrow}} \text{nova equação 1} \\ \text{soma das 6 equações} \\ \sum_{i=1}^6 (-1)^{i+1} b_0 + b_1 \sum_{i=1}^6 (-1)^{i+1} x_i = \sum_{i=1}^6 (-1)^{i+1} y_i \xrightarrow{\text{blue arrow}} \text{nova equação 2} \\ \text{eq. 1 - eq. 2 + eq. 3 - eq. 4 + eq. 5 - eq. 6} \end{array} \right.$$

2 equações formadas tem toda a informação das 6 equações (complicado para mais de 2 parâmetros)

As soluções de Mayer e Laplace  
não foram satisfatórias !

Solução definitiva: Legendre (1804)



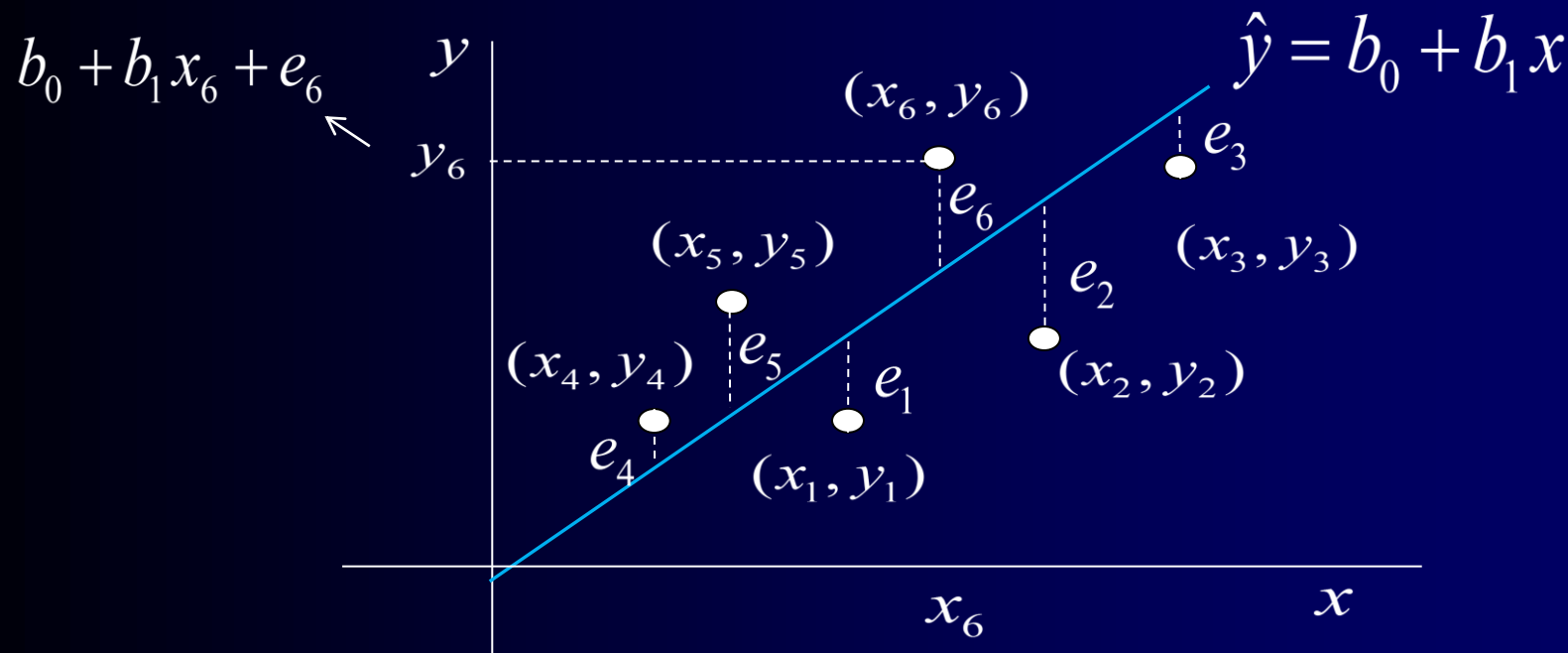
Legendre (1745-1833)

$$y_i = b_0 + b_1 x_i + e_i$$



introduziu a noção de “erro” associado à  
observação  $i$

$$y_i = b_0 + b_1 x_i + e_i \quad i = 1, \dots, 6$$



idéia fundamental de Legendre: achar os valores de  $b_0$  e  $b_1$  que minimizem uma medida agregada dos “erros” (qual ?)

Para o caso de n observações:

$$\begin{array}{llll} y_1 = b_0 + b_1 x_1 + e_1 & e_1 = y_1 - b_0 - b_1 x_1 & e_1^2 = (y_1 - b_0 - b_1 x_1)^2 & \\ y_2 = b_0 + b_1 x_2 + e_2 & e_2 = y_2 - b_0 - b_1 x_2 & e_2^2 = (y_2 - b_0 - b_1 x_2)^2 & \\ \vdots & \vdots & \vdots & \vdots \\ y_n = b_0 + b_1 x_n + e_n & e_n = y_n - b_0 - b_1 x_n & e_n^2 = (y_n - b_0 - b_1 x_n)^2 & \end{array}$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$



Medida agregada dos erros

Soma dos erros ao quadrado



Problema

achar  $b_0$  e  $b_1$  que  
minimizem essa medida

Matematicamente fácil de resolver

### Problema:

Achar os valores de  
 $b_0, b_1$   
que minimizam

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

a soma dos erros  
ao quadrado

Podemos usar as condições de primeira ordem

$$\begin{cases} \frac{\partial \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2}{\partial b_0} = 0 \\ \frac{\partial \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2}{\partial b_1} = 0 \end{cases}$$



$$\begin{cases} \sum_{i=1}^n 2 (y_i - b_0 - b_1 x_i)(-1) = 0 \\ \sum_{i=1}^n 2 (y_i - b_0 - b_1 x_i)(-x_i) = 0 \end{cases}$$

### Problema:

Achar os valores de  
 $b_0, b_1$   
que minimizam

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

a soma dos erros  
ao quadrado

Dividindo por -2 e passando o somatório

$$\left\{ \begin{array}{l} \sum_{i=1}^n y_i - n b_0 - b_1 \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n x_i y_i - b_0 \sum_{i=1}^n x_i - b_1 \sum_{i=1}^n x_i^2 = 0 \end{array} \right.$$



Rearranjando expressões

$$\left\{ \begin{array}{l} b_0 n + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{array} \right.$$

2 equações e 2 incógnitas

Expressando na forma matricial

Problema:

Achar os valores de  
 $b_0, b_1$   
que minimizam

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

a soma dos erros  
ao quadrado

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$



$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

## Problema:

Achar os valores de  
 $b_0, b_1$   
que minimizam

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

a soma dos erros  
ao quadrado

## Resultado final

médias

$$\begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{b}_1 \bar{x} \\ \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \end{pmatrix}$$

$\hat{b}_0, \hat{b}_1$   
são os valores de  
 $b_0, b_1$   
que minimizam !



$\hat{b}_0, \hat{b}_1 \rightarrow$

são os estimadores de quadrados  
mínimos ordinários (OLS) de

$\rightarrow b_0, b_1$

Método inicialmente não tinha  
uma interpretação estatística bem definida,  
era mais um método prático (muito útil)

### Exercício

$n$  medidas de uma quantidade  $b_0$  foram tomadas. Ache um estimador de quadrados mínimos para  $b_0$

$$y_i = b_0 + e_i \quad i = 1, \dots, n$$

Extensão ao caso de  $m$  variáveis explicativas

Extensão ao caso de m variáveis explicativas (c/ notação comum)

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon$$

$$\begin{array}{l} \text{n} \\ \text{obs.} \end{array} \left\{ \begin{array}{l} y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_m x_{1m} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \cdots + \beta_m x_{2m} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_m x_{nm} + \varepsilon_n \end{array} \right.$$

$\underbrace{\hspace{10em}}_{m \text{ variáveis "explicativas" } (m < n)}$

Extensão ao caso de m variáveis explicativas

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$\begin{matrix} \mathbf{y} \\ \left( \begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right) \end{matrix} = \begin{matrix} \mathbf{X} \\ \left( \begin{array}{cccc} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{array} \right) \end{matrix} \begin{matrix} \boldsymbol{\beta} \\ \left( \begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{array} \right) \end{matrix} + \begin{matrix} \mathbf{e} \\ \left( \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{array} \right) \end{matrix}$$

Extensão ao caso de  $m$  variáveis explicativas

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \longrightarrow \mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

$$\sum_{i=1}^n \varepsilon_i^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$



soma dos erros ao quadrado  
expresso na forma vetorial/matricial

Problema: encontrar  $\boldsymbol{\beta}$  que minimiza  $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Extensão ao caso de m variáveis explicativas

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon$$

Problema: encontrar  $\boldsymbol{\beta}$  que minimiza  $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

$$\frac{\partial (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$$

Condição de primeira ordem  
(caso multivariado)

Solução:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

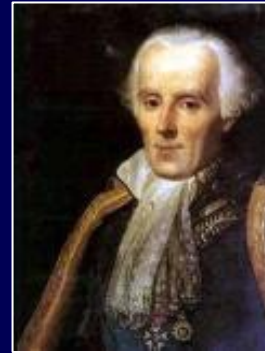
Estimador de quadrados  
mínimos ordinários (OLS) de  $\boldsymbol{\beta}$

# Premissas Estatísticas adicionadas ao Modelo

## Síntese Gauss-Laplace



Gauss (1777-1855)



Laplace (1749-1827)

Propriedades importantes do estimador  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i \quad i = 1, \dots, n$$

$y_i$     $\varepsilon_i$    variáveis aleatórias    $x_{ij}$  s   “valores não-aleatórios / fixos”

Premissas:  $\left\{ \begin{array}{l} \varepsilon_i \sim \mathbf{N}(0, \sigma) \\ \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j \end{array} \right.$    erros tem distribuição Normal com esperança zero e desvio padrão  $\sigma$   
ou, de forma mais restritiva, temos erros independentes

  $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$  e  $V[\hat{\boldsymbol{\beta}}]$  é mínima  
não-tendencioso

Desvios nas premissas invalidam uma ou as 2 condições acima

# Inferências relativas aos parâmetros

Inferências sobre os parâmetros de regressão baseiam-se na distribuição t-Student

$$\left\{ \begin{array}{l} y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \\ \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \end{array} \right. \quad \varepsilon_i \sim N(0, \sigma), \text{ i.i.d}$$
$$s = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} \quad \text{é um estimador de } \sigma$$

Pode-se demonstrar que  
(é uma quantidade pivotal)

$$\frac{\hat{\beta}_1 - \beta_1}{\frac{s}{\sqrt{\sum_{i=1}^n x_i^2}}} \sim \text{t-Student}(n-2)$$

Exercício: como dai chegar a um Intervalo de confiança para b?