# Notas de Estudo - Econometria (Versão Preliminar)

João P. M. Silva*

30 de outubro de 2025

**Resumo**

Esse documento compila as minhas notas de estudo para a disciplina de Econometria do curso de graduação em economia da UFPE. As aulas foram ministradas pelo prof. Breno Sampaio no semestre 2023.1.
Os materiais destas notas NÃO são originais. Trata-se de uma coletânea dedd pedaços de livros texto e materiais, por vezes copiados *ipsis literis* do original. Eventuais erros nas notas são de minha responsabilidade.

Os livros/materiais que uso como referência são:

- Notas de Aula
- Introdução à Econometria, 6 ed., Jeffery M. Wooldridge, 2019.

*Email: jpmsilva02@gmail.com

# Conteúdo

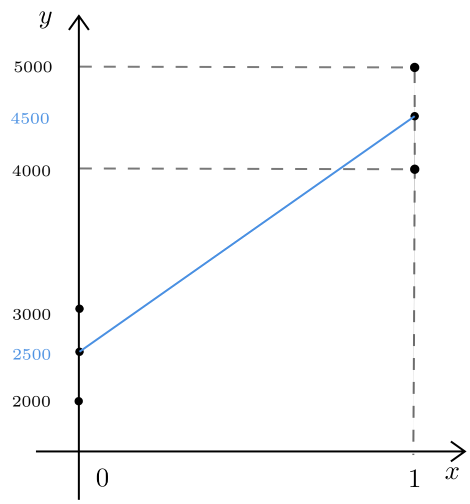
# 1 Regressão Linear Simples

Grande parte da análise econométrica começa com a seguinte premissa: $y$ e $x$ são duas variáveis, que representam alguma população, e estamos interessados em "explicar $y$ em termos de $x$", ou em "estudar como $y$ varia com a variação em $x$".

Vamos supor o seguinte exemplo:

| Renda | Educação |
|---|---|
| 2000 | 0 |
| 3000 | 0 |
| 4000 | 1 |
| 5000 | 1 |

Queremos saber qual o efeito da educação na renda. Note que educação é uma variável binária podendo assumir valores 0 para os indivíduos que não fizeram graduação e 1 para os que fizeram.




$$\text{Renda} = a + b * educ$$

$$\text{Renda} = 2500 + 2000 * educ$$

Notamos que "a"é o intercepto e "b"o coeficiente angular da reta, ou seja representa o quanto "y"varia em termos de "x". Observamos também que existem dois pontos que representam a média dos dois grupos. 4500 para os que fizeram graduação e 2500 para os que não fizeram.

Em termos simples a reta azul representa uma regressão linear simples, ou seja, a reta que melhor se ajusta ao conjunto de pontos.

Vamos começar a formalizar as coisas.

$$y_i = \beta_0 + \beta_1 x_i + d_i \tag{1}$$

A equação acima representa um modelo de regressão linear simples, onde:

$y_i$ : Variável dependente.

$\beta_0$ : Parâmetro de intercepto (constante).

$\beta_1$ : Parâmetro de inclinação.

$x_i$ : Variável independente.

$d_i$ : Desvios em relação à média (termo de erro).

Vamos calcular os desvios em relação à média das rendas:

| Renda | di |
|---|---|
| 2000 | -500 |
| 3000 | 500 |
| 4000 | -500 |
| 5000 | 500 |
| $\sum d_i = 0$ | |

Em outras palavras:

$$\sum (x_i - \bar{x}) = 0 \tag{2}$$

Vamos provar que a expressão acima e verdadeira:

$$\begin{aligned}
\sum (x_i - \bar{x}) &= \sum x_i - \sum \bar{x} \\
&= \frac{n}{n} \sum x_i - \sum \bar{x} \\
&= n \underbrace{\frac{\sum x_i}{n}}_{\bar{x}} - \sum \bar{x} \\
&= n\bar{x} - \underbrace{\sum \bar{x}}_{n\bar{x}} \\
&= n\bar{x} - n\bar{x} = 0
\end{aligned}$$

Acabamos de provar que o somatório dos desvios em relação a média são sempre zero.

Agora vamos substituir nosso $d_i$ por um novo termo $\varepsilon_i$, que chamaremos de termo de erro. Portanto, nossa nova fórmula será:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{3}$$

## 1.1 Exogeneidade Estrita

Nossa hipótese $H_1$ será:

$$H1 : E[\epsilon|x] = 0 \tag{4}$$

Ou seja, A média do erro para todo $X$ é igual a zero visto que $\sum \epsilon_i = 0$.
Voltando ao nosso exemplo inicial de duas variáveis binárias assumindo valores de 0 ou 1. Como sabemos que a média dos desvios para todo $x$ é igual a zero, então temos que

$$E(\epsilon|G = 0) = 0$$
$$E(\epsilon|G = 1) = 0$$
$$E(\epsilon|G) = 0$$

**Chamamos isso de exogeneidade estrita.**

$$E(\epsilon|x = 0) = 0$$
$$E(\epsilon|x = 1) = 0$$
$$\vdots$$
$$E(\epsilon|x = n) = 0$$

Vamos voltar ao nosso exemplo (1). Note que estaremos frequentemente usando esse exemplo durante estas notas de aula.

Sabemos que a média para o grupo de pessoas que não possui graduação é de 2500 e, para os que possuem graduação, a renda média será de 4500. E a média do todo é de 3500.

$$\bar{y} = 3500$$
$$\bar{y}|x = 0 = 2500$$
$$\bar{y}|x = 1 = 4500$$

Disso podemos extrair a seguinte propriedade matemática:

$$3500 = \frac{1}{2}2500 + \frac{1}{2}4500$$

**A média do todo é a média das médias de subgrupos (condicionais).**

Isso nos leva para a lei das expectativas iteradas, onde:

**Lei das Expectativas Iteradas**

$$E[y] = E[E[y|x]]$$

A média de y é igual a média das médias condicionais de y em relação a x.

Agora temos dois pressupostos,

$$P1 : E[\epsilon] = E[E[\epsilon|x]] = E[0] = 0 \tag{5}$$
$$P2 : E[x\epsilon] = E[E[x\epsilon|x]] = E[xE[\epsilon|x]] = E[0] = 0 \tag{6}$$

## 1.2 Método dos Momentos

### 1.2.1 Somatório

Dada a reta estimada por:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{7}$$

E os pressupostos 1 e 2 da lei das expectativas iteradas, temos:

$$E[\epsilon_i] = \frac{1}{n}\sum \epsilon_i \implies \frac{1}{n}\sum (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{8}$$

$$E[x\epsilon_i] = \frac{1}{n}\sum x_i\epsilon_i \implies \frac{1}{n}\sum x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \tag{9}$$

De (8), temos:

$$\frac{1}{n}\sum (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\bar{y} - \beta_0 - \beta_1 \bar{x} = 0$$

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}} \tag{10}$$

Agora (10) em (9):

$$\frac{1}{n}\sum x_i[y_i - (\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}) - \beta_1 x_i] = 0$$

$$\sum x_i[(y_i - \bar{y}) - \beta_1(x_i - \bar{x})] = 0$$

$$\beta_1 = \frac{\sum x_i(y_i - \bar{y})}{\sum x_i(x_i - \bar{x})}$$

$$\beta_1 = \frac{\sum x_i(y_i - \bar{y}) - \sum \bar{x}(y_i - \bar{y})}{\sum x_i(x_i - \bar{x}) - \sum \bar{x}(x_i - \bar{x})}$$

$$\boxed{\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}} \tag{11}$$

Podemos concluir que,

$$\hat{\beta}_1 = \frac{Cov(x,y)}{Var(x)} \tag{12}$$

### 1.2.2 Matriz

$$y = x\beta + \varepsilon$$

$$x'y = x'x\beta + x'\varepsilon$$

$$E(x'y) = E(x'x\beta) + \underbrace{E(x'\varepsilon)}_{=0}$$

$$x'y = x'x\beta$$

$$(x'x)^{-1}x'y = (x'x)^{-1}x'x\beta$$

$$\boxed{\beta = (x'x)^{-1}x'y} \tag{13}$$

Matricialmente...

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \mathbb{1}' & \mathbb{1} & x_i \\ x_i' & & \end{pmatrix}^{-1} \begin{bmatrix} \mathbb{1}' \\ \\ x_i' \end{bmatrix} \begin{bmatrix} y_i \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

## 1.3 Ordinary Least Squares

### 1.3.1 Somatório

A reta de uma regressão é melhor ajustada quando:

$$\beta_0, \beta_1 = \min \sum \varepsilon_i \tag{14}$$

Sendo $\sum \varepsilon_i = 0$, temos:

$$\beta_0, \beta_1 = \min \sum \varepsilon_i^2$$

$$\beta_0, \beta_1 = \min \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

Agora derivando parcialmente,

$$\frac{\partial}{\partial \beta_0} = (2) \sum (y_i - \beta_0 - \beta_1 x_i)(-1) = 0 \tag{15}$$

$$\frac{\partial}{\partial \beta_1} = (2) \sum (y_i - \beta_0 - \beta_1 x_i)(-1)(x_i) = 0 \tag{16}$$

Agora resolvendo (15) para $\beta_0$:

$$(2) \sum (y_i - \beta_0 - \beta_1 x_i)(-1) = 0$$

Após cortarmos o 2 e o -1, multiplicamos por 1/n.

$$\frac{1}{n} \sum (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\bar{y} - \beta_0 - \beta_1 \bar{x} = 0$$

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}} \tag{17}$$

Agora jogando (17) em (16) e cortando 2 e -1, temos:

$$\sum [y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i](x_i) = 0$$

$$\sum x_i [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x}) = 0$$

$$\hat{\beta}_1 = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})}$$

$$\hat{\beta}_1 = \frac{\sum x_i (y_i - \bar{y}) - \sum \bar{x} (y_i - \bar{y})}{\sum x_i (x_i - \bar{x}) - \sum \bar{x} (x_i - \bar{x})}$$

$$\boxed{\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}} \tag{18}$$

### 1.3.2 Matriz

O ajuste ótimo é dado por:

$$\beta = \arg\min_{\beta} \varepsilon'\varepsilon \tag{19}$$

$$\beta = \arg\min_{\beta} (y - x\beta)'(y - x\beta)$$

$$\beta = \arg\min_{\beta} [y'y - (x\beta)'y - y'(x\beta) + (x\beta)'(x\beta)]$$

$$\beta = \arg\min_{\beta} [y'y - 2\beta'x'y + \beta x'x\beta]$$

$$\frac{\partial}{\partial\beta} = (-2)x'y + (2)x'x\beta = 0$$

$$x'x\beta = x'y$$

$$(x'x)^{-1}x'x\beta = (x'x)^{-1}x'y$$

$$\boxed{\beta = (x'x)^{-1}x'y} \tag{20}$$

Matricialmente...

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \mathbb{1}' & \mathbb{1} \quad x_i & \\ x_i' & & \end{pmatrix}^{-1} \begin{bmatrix} \mathbb{1}' \\ \\ x_i' \end{bmatrix} \begin{bmatrix} y_i \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

## 1.4 R Squared

### 1.4.1 Demonstração

Sendo,

$$y_i = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\hat{y}_i} + \hat{\varepsilon}_i \tag{21}$$

$$y_i = \hat{y}_i + \hat{\varepsilon}_i \tag{22}$$

Multiplicamos ambos os lados da equação (22) por $-\bar{y}$.

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + \hat{\varepsilon}_i \tag{23}$$

Agora elevamos a equação (23) ao quadrado, assim obtendo:

$$(y_i - \bar{y})^2 = (\hat{y}_i - \bar{y} + \hat{\varepsilon}_i)^2 \tag{24}$$

Abrindo o produto notável da eq. (24):

$$(y_i - \bar{y})^2 = (\hat{y}_i - \bar{y})^2 + 2(\hat{y}_i - \bar{y})\hat{\varepsilon}_i + \hat{\varepsilon_i}^2 \tag{25}$$

Aplicando somatório em ambos os lados da eq. (25):

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \underbrace{2\sum (\hat{y}_i - \bar{y})\hat{\varepsilon}_i}_{\sum \varepsilon_i = 0 \text{ via H1}} + \sum \hat{\varepsilon_i}^2 \tag{26}$$

Assim obtemos,

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{\varepsilon_i}^2 \tag{27}$$

Portanto,

$$\boxed{R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}} \tag{28}$$

Ou, manipulando...

$$\frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{(\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} + \frac{\sum \hat{\varepsilon}^2}{\sum (y_i - \bar{y})^2} \tag{29}$$

Assim, obtemos:

$$\boxed{R^2 = 1 - \frac{\sum \hat{\varepsilon}^2}{\sum (y_i - \bar{y})^2}} \tag{30}$$

## 1.5 Viés

### 1.5.1 Hipóteses de Gauss-Markov

Vamos iniciar estabelecendo a inexistência de viés do método OLS sob um conjunto de hipóteses. É importante notar que apenas as hipóteses de 1 até 4 são necessárias para mostrar que $\hat{\beta_0}$ e $\hat{\beta_1}$ são não viesadas. A hipótese 5 de homoscedasticidade foi adicionada para obtermos as fórmulas habituais de variância de OLS.

**Hipótese 1: Linear em Parâmetros.**
No modelo populacional, a variável dependente, y, está relacionada com a variável independente, x, e com o erro, u, como:

$$y = \beta_0 + \beta_1 x + u \tag{31}$$

em que $\beta_0$ e $\beta_1$ são os parâmetros do intercepto e da inclinação populacionais, respectivamente.

**Hipótese 2: Amostragem Aleatória.**
Temos uma amostra aleatória de tamanho $n$, $\{(x_i, y_i) : i = 1, 2, ..., n\}$, seguindo o modelo populacional na Hipótese 1.

**Hipótese 3: Variação Amostral na Variável Explicativa.**
Os resultados amostrais em $x$ não são todos do mesmo valor.

**Hipótese 4: Média Condicional Zero.**
O erro $u$ tem zero como valor esperado, quaisquer que sejam os valores das variáveis. Em outras palavras:

$$E(u|x) = 0 \tag{32}$$

**Hipótese 5: Homoscedasticidade.**
O erro $u$ tem a mesma vari6ancia quaisquer qu sejam os valores das variáveis explicativas. Em outras palavras:

$$Var(u|x) = \sigma^2 \tag{33}$$

### 1.5.2 Demonstração de Ausência de Viés

A ausência de viés é garantida por:

$$E[\beta] = \beta$$

Sabendo que

$$y = x\beta + \varepsilon \tag{34}$$

$$\hat{\beta} = (x'x)^{-1} x\prime y \tag{35}$$

Substituindo (34) em (35), temos:

$$\hat{\beta} = (x'x)^{-1} x'(x\beta + \varepsilon)$$

$$\hat{\beta} = \beta + (x'x)^{-1}x'\varepsilon$$

$$E[\hat{\beta}] = E[\beta] + E[(x'x)^{-1}x'\varepsilon]$$

$$E[\hat{\beta}|x] = E[\beta|x] + E[(x'x)^{-1}x'\varepsilon|x]$$

$$E[\hat{\beta}|x] = \beta + \underbrace{(x'x)^{-1}x'E[\varepsilon|x]}_{=0}$$

$$E[\hat{\beta}|x] = \beta \tag{36}$$

da Lei das Expectativas Iteradas, temos que:

$$E[\hat{\beta}] = E[E(\hat{\beta}|x)] = E[\beta] = \beta \tag{37}$$

## 1.6 Modelos de RLS com Transformações Logarítmicas

Considerando[1] um modelo bi-variado simples $Y_i = \alpha + \beta X_i + \varepsilon_i$, existem quatro possibilidades de combinação de transformações envolvendo logaritmos: O caso linear sem transformações, o modelo linear-log, log-linear e o modelo log-log.

| | $X$ | |
|---|---|---|
| $Y$ | $X$ | log$X$ |
| $Y$ | *linear*<br>$\hat{Y}_i = \alpha + \beta X_i$ | *linear-log*<br>$\hat{Y}_i = \alpha + \beta \log X_i$ |
| log$Y$ | *log-linear*<br>$\log \hat{Y}_i = \alpha + \beta X_i$ | *log-log*<br>$\log \hat{Y}_i = \alpha + \beta \log X_i$ |

Table 1: Four varieties of logarithmic transformations

Lembrando que estamos usando o Log Natural, onde a base é $e \approx 2.71828$. Logaritmos podem possuir outras bases.

A transformação logarítmica de variáveis em um modelo de regressão é uma maneira muito comum de lidar com situações onde existe uma relação não-linear entre as variáveis independentes e dependentes. Usar o logaritmo de uma ou mais variáveis em vez de sua forma sem Log, torna o relacionamento entre as variáveis não-linear, enquanto ainda preserva o modelo linear.

A seguir vamos entender como interpretar cada um dos modelos.

[1] Essa seção é baseada no material "Linear Regression Models with Logarithmic Transformations"de Kenneth Benoit (LSE), 2011.

### 1.6.1 Modelo Linear

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{38}$$

Este é o modelo de regressão no qual estamos acostumados. Relembre que em um modelo de regressão linear simples, o coeficiente $\beta$ nos dá diretamente o quanto Y varia por uma mudança em uma unidade em X.

Vale notar que essa interpretação se manterá nas variáveis que foram transformadas logaritmicamente, porém faz mais sentido interpretar as mudanças não em unidades de log, e sim, em variações percentuais.

### 1.6.2 Modelo Linear-Log

$$Y_i = \alpha + \beta log X_i + \varepsilon_i \tag{39}$$

No modelo linear-log, a interpretação do coeficiente estimado $\hat{\beta}$ é de que a mudança em uma unidade em $logX$ produzirá um aumento esperado em Y de $\hat{\beta}$ unidades.

### 1.6.3 Modelo Log-Linear

$$log Y_i = \alpha + \beta X_i + \varepsilon_i \tag{40}$$

No modelo log-linear, a interpretação do $\hat{\beta}$ estimado é de que um aumento em uma unidade de X produzirá um aumento esperado em logY de $\hat{\beta}$ unidades. Em termos de Y, isso significa que o valor esperado de Y é multiplicado por $e^{\hat{\beta}}$. Em outras palavras, cada aumento em uma unidade em X multiplica o valor esperado de Y em $e^{\hat{\beta}}$.

### 1.6.4 Modelo Log-Log

$$log Y_i = \alpha + \beta log X_i + \varepsilon_i \tag{41}$$

Em casos em que tanto a variável dependente quanto a(s) variável(is) independente(s) estão logaritmicamente transformadas, a interpretação é uma combinação dos casos linear-log e log-linear acima mencionados. Em outras palavras, a interpretação é dada como uma mudança percentual esperada em Y quando X aumenta em certo percentual. Tais relações, em que tanto Y quanto X estão logaritmicamente transformados, são comumente referidas como elásticas na econometria, e o coeficiente de log X é chamado de elasticidade.

# 2 Regressão Múltipla

A análise de regressão múltipla nos permite controlar *explicitamente* muitos outros fatores que, de maneira simultânea, afetam a variável dependente. Em razão dos modelos de regressão múltipla acomodarem mais variáveis explicativas que estar correlacionadas, esperamos inferir causalidade nos casos em que a análise de regressão simples seria enganosa.

Naturalmente, ao adicionar ao nosso modelo mais fatores que são úteis para explicar y, então mais da variação de y poderá ser explicada.

## 2.1 Viés de Variável Omitida

Dado o modelo estimado:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon \tag{42}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = (x'x)^{-1}x'y \tag{43}$$

E o modelo verdadeiro:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{44}$$

Então,

$$\hat{\beta} = (x'x)^{-1}x'(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon)$$

$$\hat{\beta} = (x'x)^{-1}x'[x\beta + \beta_2 x_2 + \varepsilon]$$

$$\hat{\beta} = \beta + (x'x)^{-1}x'\beta_2 x_2 + (x'x)^{-1}x'\varepsilon$$

$$E[\hat{\beta}|x] = E[\beta|x] + E[(x'x)^{-1}x'\beta_2 x_2|x] + E[(x'x)^{-1}x'\varepsilon|x]$$

$$E[\hat{\beta}|x] = \beta_1 + \beta_2 \underbrace{(x'x)^{-1}x'x_2}_{\text{A}}$$

De A temos,

$$\beta(x'x)^{-1}x'\beta_2 x_2$$

$$x_2 = \gamma_0 + \gamma_1 x_1 + \varepsilon$$

$$E[\hat{\beta}|x] = \beta + \beta_2\hat{\gamma} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \beta_2 \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{pmatrix}$$

- $\beta_1$ é superestimado quando:

$$\beta_2\hat{\gamma} < 0 \tag{45}$$

- $\beta_1$ é subestimado quando:

$$\beta_2\hat{\gamma} > 0 \tag{46}$$

# A Matriz inversa