# Econometria Aplicada

## Regressão linear múltipla e formas funcionais

João Ricardo Costa Filho

*“The most important questions of life are, for the most part, really only problems in probability."*

Laplace (1812)

*“In God we trust. All others must bring data."*

William Edwards Deming

# A regressão linear

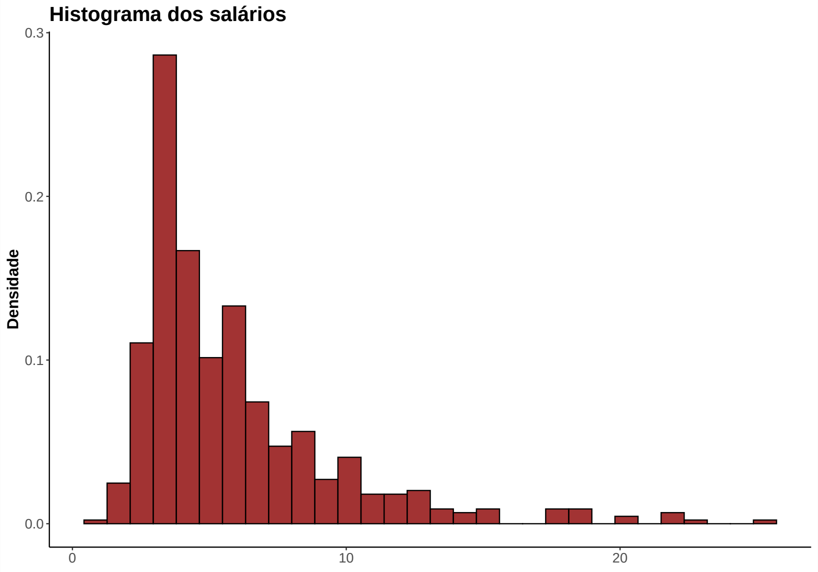
## Motivação (tudo começa com uma pergunta)

Qual é o efeito da educação nos salários dos trabalhadores?

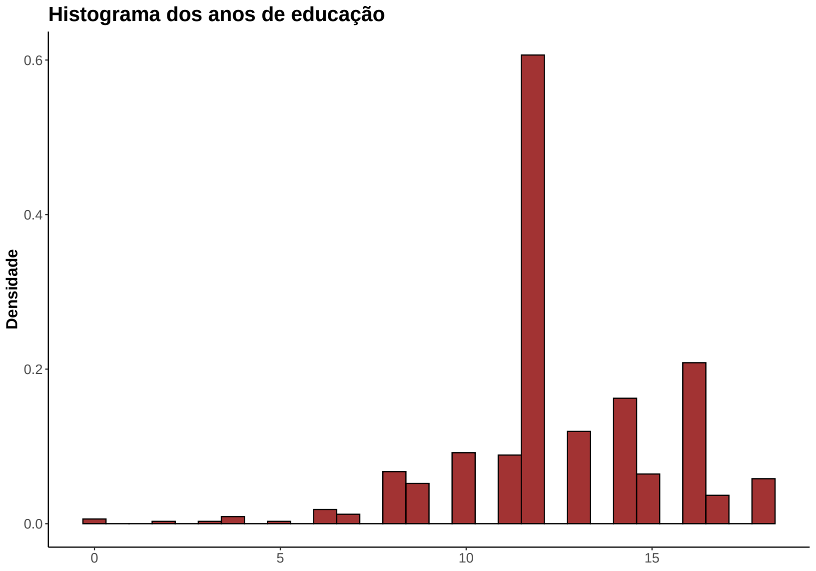
# Dados

```r
library(wooldridge)

data(wage1) # base de dados

attach( wage1 )
```

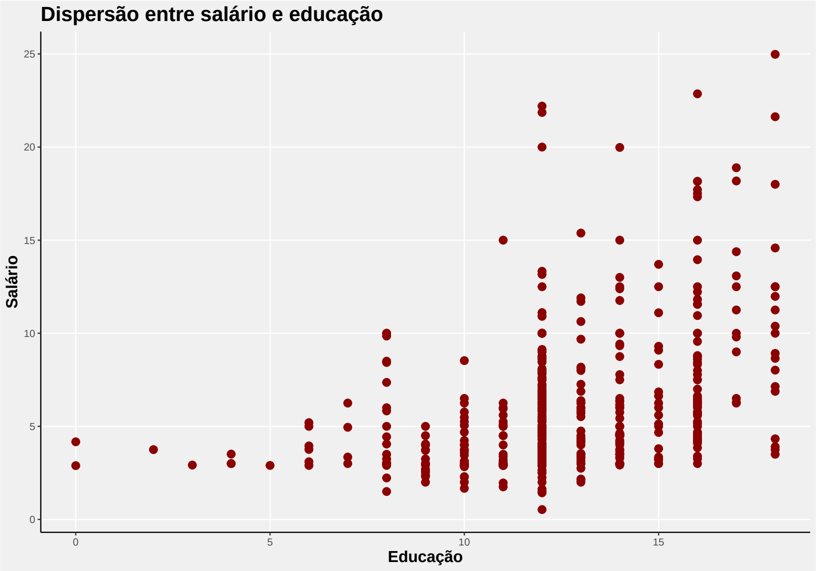
# Visualização dos dados (super importante!)



# Visualização dos dados (super importante!)



# Visualização dos dados (super importante!)



# Estatísticas descritivas (super importante!)

```
##                            Salário Educação
## Média                         5.90    12.56
## Variância                    13.64     7.67
## Desvio-padrão                 3.69     2.77
## Coeficiente de Variação       0.63     0.22

## [1] "Covariância entre salários e educação"

## [1] 4.15

## [1] "Correlação entre salários e educação"

## [1] 0.41
```

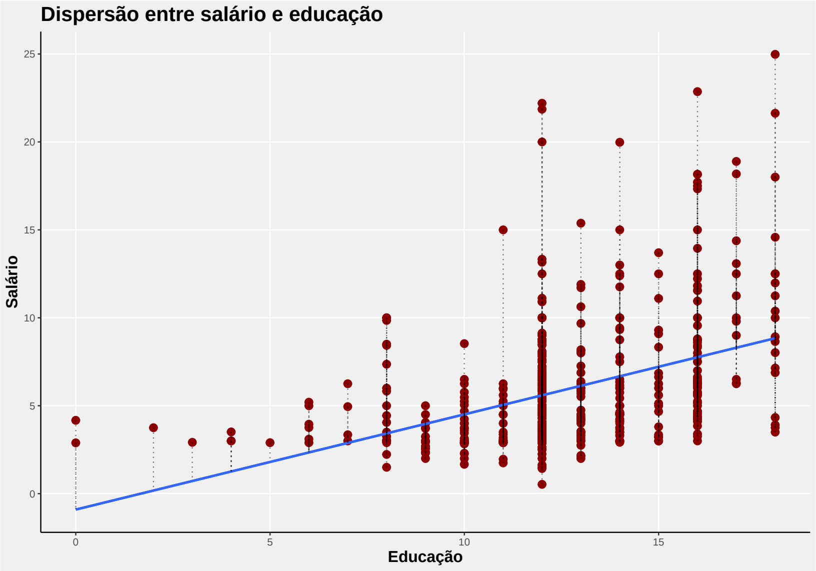
# Regressão linear simples

$$salario_i = \beta_0 + \beta_1 Educ_i + \varepsilon_i.$$

## Regressão linear simples no R

```r
reg = lm( wage ~ educ, data = wage1)
```

# Regressão linear simples no R



# Inferência

A relação entre salários e educação é estatísticamente significativa?

Para $\hat{\beta}_1$:

$$\mathcal{H}_0 : \beta_1 = \mu$$
$$\mathcal{H}_a : \beta_1 \neq \mu$$

E a significância conjunta?

$$\mathcal{H}_0 : \beta_0 = \beta_1 = 0$$
$$\mathcal{H}_a : \beta_j \neq 0, \text{ para } j = [0, 1]$$

Vamos olhar para os resíduos.

# Só a variável 'educação' na regressão é suficiente?

E se quisermos estimar a seguinte equação?

$$salario_i = \beta_0 + \beta_1 Educ_i + \beta_2 Exper_i + \varepsilon_i$$

# Regressão múltipla no R

```r
reg = lm( wage ~ educ + exper, data = wage1)
```

# Formas funcionais e não-linearidades

Se a realação entre as variáveis for não-linear, ainda podemos utilizar uma regressão linear?

## Termo quadrático

Será que a experiência possui retornos marginais constantes, crescentes ou decrescentes? Como estimar?

# Termo quadrático

Será que a experiência possui retornos marginais constantes, crescentes ou decrescentes? Como estimar?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 Exper_i + \beta_3 Exper_i^2 + \varepsilon_i$$

# Termo quadrático

Será que a experiência possui retornos marginais constantes, crescentes ou decrescentes? Como estimar?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 Exper_i + \beta_3 Exper_i^2 + \varepsilon_i$$

No R:

```
reg = lm( wage ~ educ + exper + expersq, data = wage1)
```

# (Semi-)Elasticidade

Imagine que estamos interessados na semi-elasticidade salário-educação e na elasticidade salário-experiência. Como estimar essas relações?

# (Semi-)Elasticidade

Imagine que estamos interessados na semi-elasticidade salário-educação e na elasticidade salário-experiência. Como estimar essas relações?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 \ln[Exper_i] + \varepsilon_i$$

# (Semi-)Elasticidade

Imagine que estamos interessados na semi-elasticidade salário-educação e na elasticidade salário-experiência. Como estimar essas relações?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 \ln[Exper_i] + \varepsilon_i$$

No R:

```r
reg = lm( wage ~ educ + exper, data = wage1)
```

# Interpretação dos coeficientes

Quais são os significados de $\beta_1$ e $\beta_2$ agora?

# Interpretação dos coeficientes

Quais são os significados de $\beta_1$ e $\beta_2$ agora?

- $\beta_1$ (educação): Em média, se aumentarmos a **educação** de um(a) trabalhador(a) em 1 ano, o salário aumentará em $\beta_1 * 100\%$ dólares por hora.

# Interpretação dos coeficientes

Quais são os significados de $\beta_1$ e $\beta_2$ agora?

- $\beta_1$ (educação): Em média, se aumentarmos a **educação** de um(a) trabalhador(a) em 1 ano, o salário aumentará em $\beta_1 * 100\%$ dólares por hora.
- $\beta_2$ (experiência): Em média, se aumentarmos a **experiência** de um(a) trabalhador(a) em 1%, o salário aumentará em $\beta_1\%$.

# Dummy

Será que há alguma diferença entre trabalhadores casos e solteiros?

# Dummy

Será que há alguma diferença entre trabalhadores casos e solteiros?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 \ln[Exper_i] + \delta D_i + \varepsilon_i$$

# Dummy

Será que há alguma diferença entre trabalhadores casos e solteiros?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 \ln[Exper_i] + \delta D_i + \varepsilon_i$$

No R:

```
reg = lm( wage ~ educ + exper + married, data = wage1)
```

# Interpretação dos coeficientes

# Interpretação dos coeficientes

- Linear (nível): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
    - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1$ unidades em Y.

## Interpretação dos coeficientes

- Linear (nível): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1$ unidades em Y.
- Log-linear: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.

# Interpretação dos coeficientes

- Linear (nível): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1$ unidades em Y.
- Log-linear: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.
- Log-Log: $\ln[Y_i] = \beta_0 + \beta_1 \ln[X_i] + \varepsilon_i$
  - Em média, o aumento de 1% em X está associado à um aumento de $\beta_1\%$ em Y.

## Interpretação dos coeficientes

- Linear (nível): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1$ unidades em Y.
- Log-linear: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.
- Log-Log: $\ln[Y_i] = \beta_0 + \beta_1 \ln[X_i] + \varepsilon_i$
  - Em média, o aumento de 1% em X está associado à um aumento de $\beta_1\%$ em Y.
- Linear com dummy: $Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \varepsilon_i$
  - Em média, a dummy sair de 0 para 1 está associado à Y aumentar em $\delta$ unidades.

## Interpretação dos coeficientes

- Linear (nível): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1$ unidades em Y.
- Log-linear: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.
- Log-Log: $\ln[Y_i] = \beta_0 + \beta_1 \ln[X_i] + \varepsilon_i$
  - Em média, o aumento de 1% em X está associado à um aumento de $\beta_1\%$ em Y.
- Linear com dummy: $Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \varepsilon_i$
  - Em média, a dummy sair de 0 para 1 está associado à Y aumentar em $\delta$ unidades.
- Log-linear com dummy: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \delta D_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\exp[\delta - 1]\%$ em Y.

Vamos para a atividade em grupo!

# Referências i