

# Modelos de Regressão e Previsão

## Introdução ao Modelo de Regressão Linear Simples

Prof. Carlos Trucíos  
carlos.trucios@facc.ufrj.br  
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis  
Universidade Federal do Rio de Janeiro

### Aula 1

# Introdução

# Introdução

# Introdução

**Data Is The New Oil --  
And That's A Good  
Thing**

The world's most valuable resource is  
no longer oil, but data

**Data Scientist: The Sexiest  
Job of the 21st Century**

# Introdução

Algumas razões para estudar análise de dados:

- A análise de dados é muito importante para as grandes empresas.
- Oportunidades de trabalho em alta
- Salario competitivo
- Oportunidades de trabalho em diversos setores
- As tomadas de decisão nas empresas são influenciadas pelos dados.

# Introdução

A modelagem entra em cena quando:

- Temos uma teoria econômica para testar
- Temos em mente uma relação que apresenta alguma importância na tomada de decisão
- Queremos explicar determinados fenômenos
- Queremos saber como o aumento/diminuição em uma variável influencia em outra.
- Queremos fazer previsão

# Modelagem

Estamos interessado em relações do tipo

$$y = f(x_1, x_2, \dots, x_n)$$

onde  $y$  é a variável de interesse e  $x_i$ 's são as variáveis que nos ajudarão a entender/explicar/prever  $y$

# Estrutura de dados

No processo de modelagem, lidamos com diversos tipos de dados, eles podem ser classificados em:

- Corte transversal
- Séries temporais
- Corte transversal agrupados
- Painel (ou longitudinais)

Para cada tipo de dado, teremos uma abordagem de modelagem diferente que nos permitirá explorar a informação contida nos dados.



# Estrutura de dados

## Corte transversal

Consiste em uma amostra de indivíduos\* (consumidores, empresas, cidades, países, etc) tomadas em determinado ponto no tempo. Podemos pensar nesse conjunto de dados como quando tiramos uma **foto** (*panorâmica*).

# Estrutura de dados

## Corte transversal

Consiste em uma amostra de indivíduos\* (consumidores, empresas, cidades, países, etc) tomadas em determinado ponto no tempo. Podemos pensar nesse conjunto de dados como quando tiramos uma **foto** (*panoramica*).

## Séries temporais

Consiste em observações sobre uma (ou várias) variáveis ao **longo do tempo**. A diferença dos dados e corte transversal, dados de séries temporais são ordenados de forma cronológica.

# Estrutura de dados

## Corte transversal agrupados

Agrupar várias amostras de corte transversal (cada uma tomada em diferentes períodos de tempo)

# Estrutura de dados

## Corte transversal agrupados

Agrupar várias amostras de corte transversal (cada uma tomada em diferentes períodos de tempo)

## Painel (ou longitudinais)

Consiste em uma série temporal para cada observação de corte transversal. A diferença dos dados de *corte transversal agrupados*, nos dados de painel as **mesmas unidades** são acompanhadas ao longo do tempo

*Nessa disciplina, estudaremos dados de corte transversal e de séries temporais*

# Modelo de Regressão Linear Simples

# Modelo de Regressão Linear Simples

Sejam  $x$  e  $y$  duas variáveis e suponha que queremos **explicar**  $y$  **em termos de**  $x$ , **estudar como varia**  $y$  **com variações de**  $x$ .

# Modelo de Regressão Linear Simples

Sejam  $x$  e  $y$  duas variáveis e suponha que queremos **explicar  $y$  em termos de  $x$ , estudar como varia  $y$  com variações de  $x$ .**

## MRL Simples

$$y = \beta_0 + \beta_1 x + u$$

onde

- $y$  variável dependente,
- $x$  variável independente,
- $\beta_0$  é parâmetro de intercepto,
- $\beta_1$  é o parâmetro de inclinação e
- $u$  é o termo de erro ou perturbação

$u$  representa outros fatores, além de  $x$ , que afetam  $y$ .

# Modelo de Regressão Linear Simples

y	x
v. dependente	v. independente
v. explicada	v. explicativa
v. resposta	v. de controle
v. prevista	v. previsora
regressando	regressor
v. <i>target</i>	covariável



# Modelo de Regressão Linear Simples

## Exemplos

- Produção de soja e os fertilizantes

$$\text{produção} = \beta_0 + \beta_1 \text{fertilizante} + u$$

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

# Modelo de Regressão Linear Simples

## Exemplos

- Produção de soja e os fertilizantes

$$\text{produção} = \beta_0 + \beta_1 \text{fertilizante} + u$$

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

Nos exemplos acima, podemos estar interessados em saber o efeito do fertilizante sobre a produção (mantendo fixos os outros fatores) ou o efeito dos anos em educação sobre o salário (mantendo fixos os outros fatores).

# Modelo de Regressão Linear Simples

## Ceteris Paribus

Outros fatores permanecendo iguais

Em muitas aplicações, o interesse é se uma variável (exemplo: fertilizante ou educação) tem efeito causal sobre outra (exemplo: produção ou salário). Nesse sentido estamos interessados em conclusões ceteris paribus de  $x$  sobre  $y$ .

# Modelo de Regressão Linear Simples

## Ceteris Paribus

Outros fatores permanecendo iguais

Em muitas aplicações, o interesse é se uma variável (exemplo: fertilizante ou educação) tem efeito causal sobre outra (exemplo: produção ou salário). Nesse sentido estamos interessados em conclusões ceteris paribus de  $x$  sobre  $y$ .

No MRLS

$$y = \beta_0 + \beta_1 x + u$$

se todos os outros fatores são mantidos fixos (de modo que  $\Delta u = 0$ ),

$$\Delta y = \beta_1 \Delta x$$

# Modelo de Regressão Linear Simples

*Mas como podemos manter fixos todos os outros fatores quando na verdade estamos ignorando eles?*

Veremos que na verdade, somente podemos obter estimadores confiáveis de  $\beta_0$  y  $\beta_1$  quando fazemos hipotesis sobre  $u$  e como se relaciona com  $x$

# Modelo de Regressão Linear Simples

*Mas como podemos manter fixos todos os outros fatores quando na verdade estamos ignorando eles?*

Veremos que na verdade, somente podemos obter estimadores confiáveis de  $\beta_0$  y  $\beta_1$  quando fazemos hipotesis sobre  $u$  e como se relaciona com  $x$

## Hipóteses

- $E(u) = 0$

# Modelo de Regressão Linear Simples

*Mas como podemos manter fixos todos os outros fatores quando na verdade estamos ignorando eles?*

Veremos que na verdade, somente podemos obter estimadores confiáveis de  $\beta_0$  y  $\beta_1$  quando fazemos hipotesis sobre  $u$  e como se relaciona com  $x$

## Hipóteses

- $E(u) = 0$
- $E(u|x) = E(u)$

# Modelo de Regressão Linear Simples

## O qué implicam essas hipoteses?

- 1 No modelo com intercepto, sem perda de generalidade, sempre podemos assumir que o valor médio de  $u$  é zero ( $E(u) = 0$ )



# Modelo de Regressão Linear Simples

## O qué implicam essas hipoteses?

- 1 No modelo com intercepto, sem perda de generalidade, sempre podemos assumir que o valor médio de  $u$  é zero ( $E(u) = 0$ )
- 2  $E(u|x) = E(u)$  diz que o valor médio dos fatores não observáveis ( $u$ ) é o mesmo para todo valor de  $x$  e que é igual à media de  $u$ .

# Modelo de Regressão Linear Simples

## O qué implicam essas hipoteses?

- 1 No modelo com intercepto, sem perda de generalidade, sempre podemos assumir que o valor médio de  $u$  é zero ( $E(u) = 0$ )
- 2  $E(u|x) = E(u)$  diz que o valor médio dos fatores não observáveis ( $u$ ) é o mesmo para todo valor de  $x$  e que é igual à media de  $u$ .
- 3 No MRLS, se aplicarmos  $E(\cdot|x)$ , temos que

$$E(y|x) = E(\beta_0 + \beta_1 x + u) = \beta_0 + \beta_1 \underbrace{E(x|x)}_x + \underbrace{E(u|x)}_0 = \beta_0 + \beta_1 x$$

(o valor médio de  $y$  muda com  $x$ )

# Modelo de Regressão Linear Simples

## Exemplos

- Suponha que  $u$  seja **aptidão**, então  $E(\text{aptidão}|\text{educação} = 5)$  representa a aptidão média para o grupo de pessoas com 5 anos de educação e  $E(\text{aptidão}|\text{educação} = 12)$  a aptidão média para o grupo de pessoas com 12 anos de educação.  $E(u|x) = E(u)$  implica que ambas as médias devem ser as mesmas.
- Mas se entendemos que a média da aptidão aumenta com os anos de educação formal, então  $E(u|x) \neq E(u)$

# Modelo de Regressão Linear Simples

Suponha que

$$y = 1.05 + 0.5x + u$$

com  $E(u|x) = E(u) = 0$ . Então

$$E(y|x) = \underbrace{1.05}_{\beta_0} + \underbrace{0.5}_{\beta_1} x$$

(a média de  $y$  aumenta em 0.5 por unidade em  $x$ ).

# Modelo de Regressão Linear Simples

Suponha que

$$y = 1.05 + 0.5x + u$$

com  $E(u|x) = E(u) = 0$ . Então

$$E(y|x) = \underbrace{1.05}_{\beta_0} + \underbrace{0.5}_{\beta_1} x$$

(a média de  $y$  aumenta em 0.5 por unidade em  $x$ ).

- Na prática, nunca conhecemos os valores de  $\beta_0$  e  $\beta_1$  e devemos estima-los.
- Estimar  $\beta_0$  e  $\beta_1$  é o tópico da nossa próxima aula

# Leituras recomendadas

## Leituras recomendadas

- Wooldridge, Jeffrey M. *Introdução à Econometria: Uma abordagem moderna*. (2016). Cengage Learning. – **Cap 1** e **Cap 2.1**