

Introdução

Fernando B. Sabino da Silva

March 6, 2018

1 Objetivos

2 Introdução

- Econometria e Teoria Econômica
- Modelo Econômico
- Modelo Econométrico
- Ciência de Dados
 - Definição
 - Dados e Tomada de Decisão
 - O Processo de Ciência de Dados
- Big Data
 - Dados Gerados por Minuto
 - Velocidade de Crescimento dos Dados

Ao final desta aula os alunos deverão ser capazes de:

- Conhecer alguns campos de atuação da Econometria e sua relação com a teoria econômica;
- Diferir sobre Modelos Econômicos e Modelos Econométricos;
- Entender o papel dos dados na definição do método econométrico a ser utilizado;
- Entender conceitos de Ciência de Dados e sua relação com a Econometria;
- Conhecer a plataforma DataCamp que será usada para aprendizagem da linguagem R.

- Por meio da Econometria é possível avaliar empiricamente a teoria econômica:
 - Explicar fatos passados;
 - Testar teorias e hipóteses;
 - Prever resultados de políticas ou eventos futuros;
 - Estimar relações entre variáveis econômicas.
- Isso é viável porque, em geral, **há relações de equilíbrio de longo prazo** entre variáveis econômicas.
- O economista pode fazer uso de diversos campos da econometria de acordo com o **fundamento econômico**:
 - Econometria básica (regressão linear múltipla, classificação...)
 - Econometria usando Séries Temporais (ARIMA, GARCH, MIDAS, VAR, VEC, ...)
 - Econometria Não Paramétrica (Estimação de Densidades e Regressões não paramétricas, ...)
 - Microeconometria (Dados em painel, ...)
 - Macroeconometria (DSGE, DSGE-VAR, ...)

Quais os efeitos do **treinamento** sobre a **produtividade** do trabalhador?

$$\textit{saláριο}h = f(\textit{educ}, \textit{exper}, \textit{treina}) \quad (1)$$

onde:

- *saláριο*h é o salário-hora;
- *educ* representa os anos de educação formal;
- *exper* refere-se aos de experiência no mercado de trabalho;
- *treina* corresponde as semanas ocupadas em treinamento.

Hipótese: Os trabalhadores são pagos de acordo com sua produtividade.

Um modelo econométrico para o exemplo anterior

$$\text{salário}_i = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{treina} + u \quad (2)$$

onde termo u contém todos os outros fatores não incluídos na equação (sejam observados diretamente ou não), mas que podem influenciar a produtividade, tais como:

- aptidão inata;
- qualidade da educação;
- formação da família.

Objetivo: Testar hipótese sobre o parâmetro β_3 . Exemplo: O efeito marginal é diferente de zero?

- Com a migração para a internet, produzimos um fluxo constante e exaustivo de informação digital;
- Estima-se que **90% dos dados armazenados no mundo** foram produzidos apenas nos **últimos dois anos**;
- Informação tornou-se a moeda mais poderosa e exige que o mercado saiba interpretá-la a seu favor;
- Logo, a profissão do Cientista de Dados se torna crucial;
- Áreas de atuação: varejo, saúde, finanças, telecomunicações, segurança, transporte, economia, recursos humanos, ...

Definição 1: Ciência de Dados é o termo usado para o processo de extrair *insights (valor)* de dados que são coletados de várias fontes (estruturados e não estruturados)

Definição 2: Consiste de um conjunto de especialidades, tais como estatística, matemática, programação, computação e *business*

- Leia mais em

<http://datascience.columbia.edu/data-for-good-preface>

Exemplo: Recrutamento Interno de Colaboradores

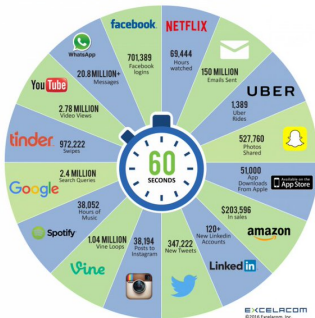
- **O que?:** Construir algoritmo que recomende colaboradores internos e/ou em processo de desligamento com maior probabilidade de performar bem nas vagas disponíveis considerando as restrições de negócio e exigências das vagas.
- **Por quê?:** Auxiliar instituição financeira em gestão de pessoas por meio da redução do turnover (em média 7.000 pessoas/ano) e, conseqüentemente, minimização dos seguintes custos médios por colaborador:
 - Desligamento: R\$ 48.443,91
 - Contratação: R\$ 4.458,00
 - Trabalhista: R\$ 93.810,54

Custo Total no Ano: R\$ 1.026.987.150,00

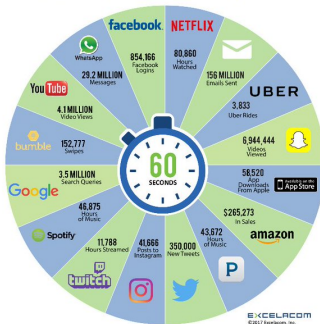
- 1 Identificar o problema da área de negócio;
- 2 Compreender o problema (entidades e atributos);
 - Exemplos de entidades: Cliente, Produto, Contrato, Vendas, etc.
 - Exemplos de atributos: Código do Produto (Entidade Produto), Nome do Cliente (Entidade Cliente).
- 3 Coletar conjuntos de dados que representem as entidades;
- 4 Limpar e transformar os dados;
- 5 Compreender os relacionamentos entre os dados;
- 6 Criar modelos estatísticos e matemáticos que representem os relacionamentos;
- 7 Utilizar os modelos para fazer previsões;
- 8 Entregar valor e resultado.

Dados Gerados por Minuto

2016 What happens in an INTERNET MINUTE?



2017 What happens in an INTERNET MINUTE?



Velocidade de Crescimento dos Dados

