

Data science aplicada aos negócios

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

Ciências de dados é ciência

O que é e como fazer ciência?

A ciência está baseada na formulação de hipóteses que podem ser testadas para assim fornecerem evidências que corroboram ou desprovam essa hipótese, levando a uma nova hipótese.

Entendendo o problema

Nosso trabalho é responder perguntas mas para fazer as perguntas certas precisamos entender o nosso problema.

- Não adianta resolver o problema errado

Análise Exploratória

Análise exploratória

Etapas da estatística descritiva.



Análise descritiva dos dados

Uma descrição óbvia de um conjunto de dados está nos próprios dados.

- Número de observações
- Valores mínimos e máximos
- Ordenação dos valores

Análise descritiva dos dados

Quando temos muitos dados para analisar cada observação, precisamos de novas ferramentas.

- Tendências centrais
- Dispersão
- Correlação

Análise descritiva dos dados

A análise descritiva utiliza a estatística descritiva para construir conhecimento.

Tendências centrais

- Média
- Mediana

Análise descritiva dos dados

Dispersão

- Amplitude
- Variância
- Desvio-padrão

Análise descritiva dos dados

Correlação

- Covariância
- Correlação

A correlação varia entre -1 (anticorrelação perfeita) e 1 (correlação perfeita)

Pandas e Numpy



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA