

## **Aula 03 – (Aula Síncrona - Teórica)** (23/08)

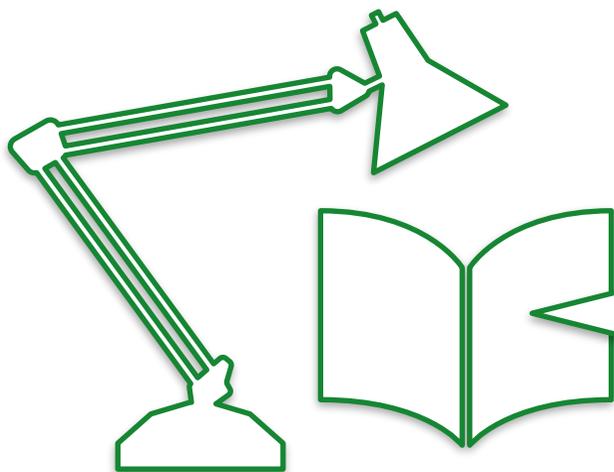
Davi R. de Moura Costa



USP



fea-RP



## Inferência Estatística

- Coleta
- Tratamento
- Análise
- Modelagem

## População e Amostra

- Teste de Hipótese

## Tomada de decisão

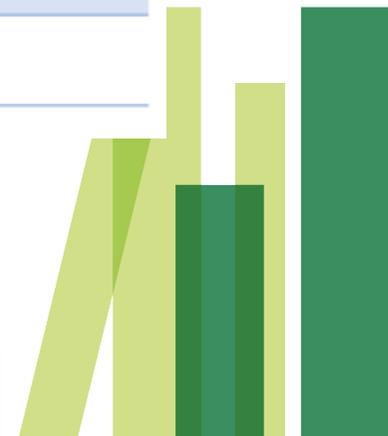


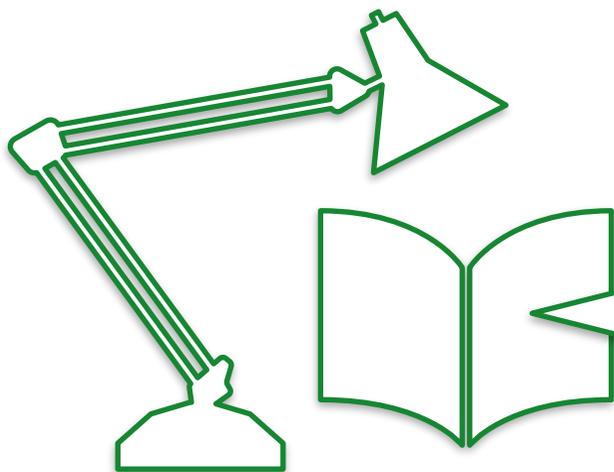
Aula de hoje - **23/08** está  
embasado em:  
BUSSAB, Wilton O.; MORETTIN,  
Pedro A. "**Estatística Básica**"  
6ed. São Paulo Sartaiva, 2010  
- Cap. 1 e 2.



Área	Analista de BI	Cientista de Dados
<b>Foco</b>	Relatórios, KPI's, Tendências	Padrões, Correlações, Modelos Preditivos
<b>Processo</b>	Estático, Comparativo	Exploratório, Experimental, Visual
<b>Fontes de Dados</b>	Data Warehouses, Bancos Transacionais	Big Data, Dados Não-Estruturados, Bancos Transacionais e NoSQL, Dados Gerados em Tempo Real
<b>Qualidade dos Dados na Fonte</b>	Alta	Baixa ou Média (requer processo de limpeza e transformação)
<b>Modelo de Dados</b>	Esquema de dados bem definido na fonte	Esquema de dados definido no momento da consulta
<b>Transformações nos Dados</b>	Pouca ou nenhuma (dados já organizados na fonte)	Transformação sob demanda, necessidade de complementar os dados
<b>Análise</b>	Descritiva, Retrospectiva	Preditiva, Prescritiva
<b>Responde à pergunta:</b>	O que aconteceu?	O que pode acontecer?

Fonte: <http://datascienceacademy.com.br/blog/qual-a-diferenca-entre-o-analista-de-bi-e-o-cientista-de-dados/>



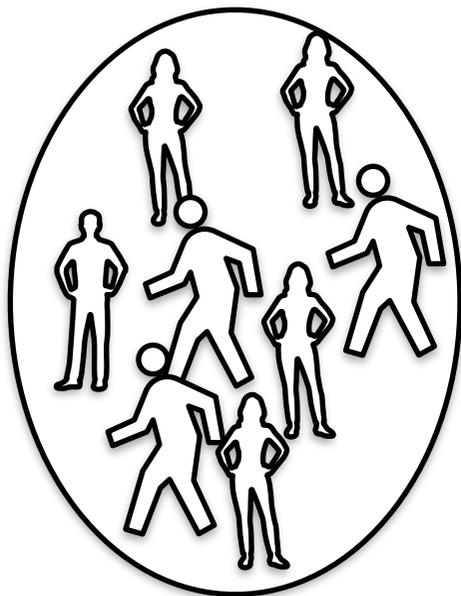


Inferência Estatística  
- Coleta

Questionário



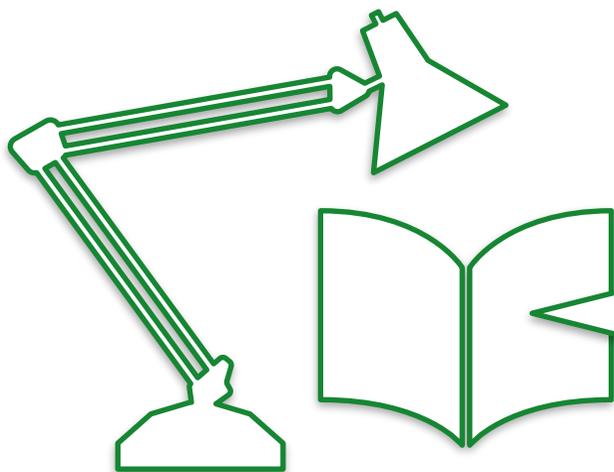
## Amostra - Estatísticas



Amostra Aleatória Simples - c/ e s/ reposição

Evita viés

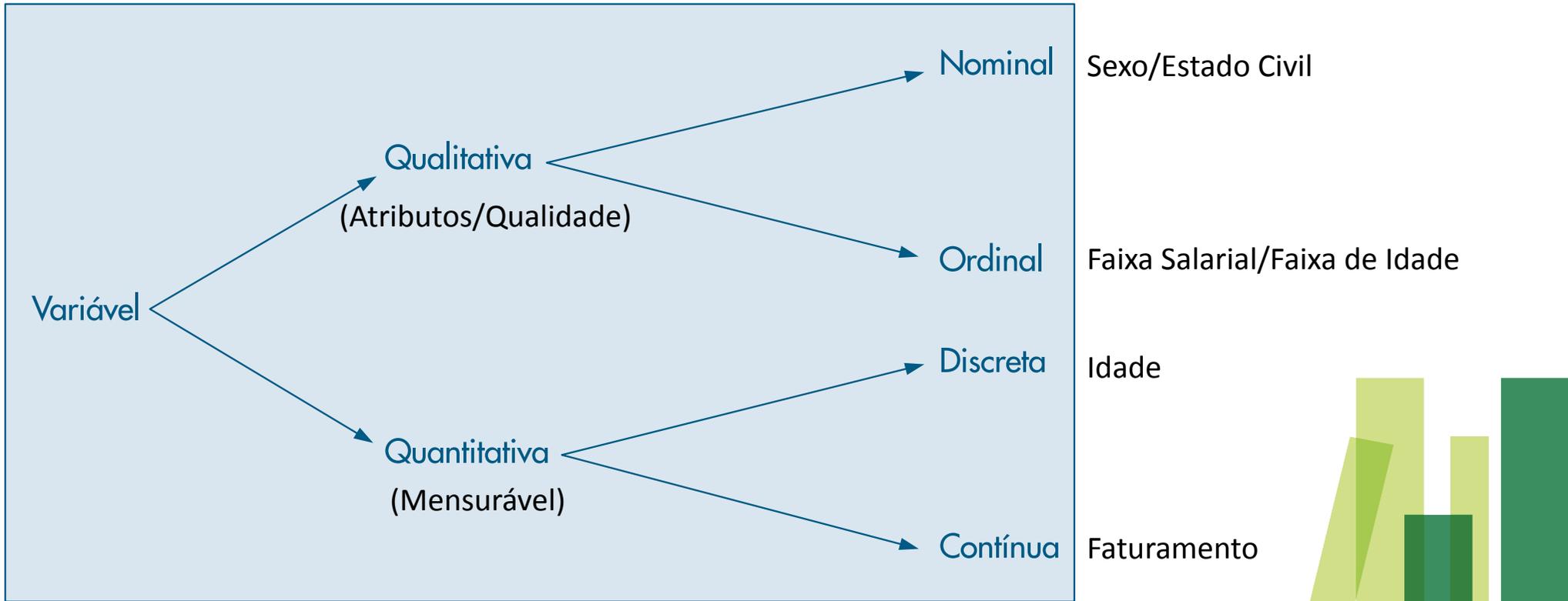
**Definição.** Uma amostra aleatória simples de tamanho  $n$  de uma variável aleatória  $X$ , com dada distribuição, é o conjunto de  $n$  variáveis aleatórias independentes  $X_1, X_2, \dots, X_n$ , cada uma com a mesma distribuição de  $X$ .

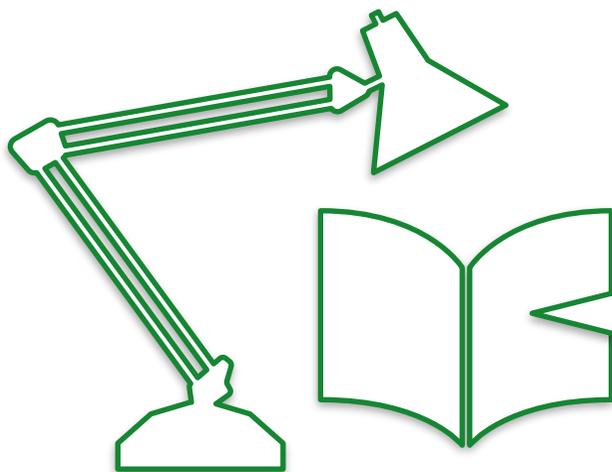


Inferência Estatística  
- Tratamento

As Variáveis e seus significados







## Inferência Estatística

- Análise (Qualitativa e Quantitativa)

Quais medidas são relevantes? Medidas de Posição e Medidas de Dispersão

Há padrões nas Respostas?



## Distribuição de Frequência

### Dados discretos

Grau de instrução	Frequência $n_i$ (Absoluta)	Proporção $f_i$ (Relativa)	Porcentagem $100 f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

Como chegou no 12, 18 e 6?

$$\sum_{i=1}^n (n_i) \quad f_i = \frac{\sum_{i=1}^n (n_i)}{n} \quad f_i = 100 * \left( \frac{n_i}{\sum_{i=1}^n (n_i)} \right)$$

O que Média?

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

O que é Variância ou Desvio-padrão?

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$



Fábrica 1

Grau de instrução	Freqüência $n_i$	Proporção $f_i$	Porcentagem $100 f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

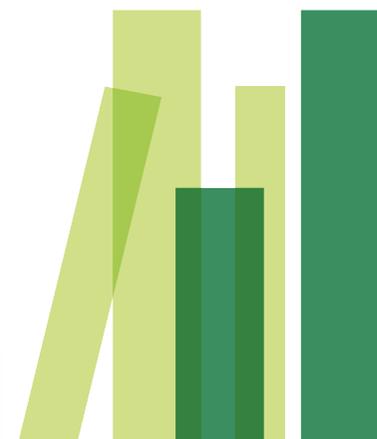
Fábrica 2

Grau de instrução	Freqüência $n_i$	Porcentagem $100 f_i$
Fundamental	650	32,50
Médio	1.020	51,00
Superior	330	16,50
Total	2.000	100,00

Como comparar os dados?

O que os dados nos dizem?

## Importância



## Dados Contínuos

Classe de salários	Frequência $n_i$	Porcentagem $100 f_i$
4,00 ─ 8,00	10	27,78
8,00 ─ 12,00	12	33,33
12,00 ─ 16,00	8	22,22
16,00 ─ 20,00	5	13,89
20,00 ─ 24,00	1	2,78
Total	36	100,00

Como chegou no 10, 12, 8, 5 e 1?

Qual o ganho ou perda  
em relação a tabela  
anterior?

## Distribuição de Frequência

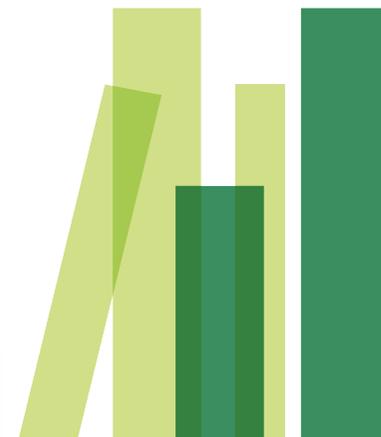


Classe de salários	Frequência $n_i$	Porcentagem $100 f_i$
4,00 ─ 8,00	10	27,78
8,00 ─ 12,00	12	33,33
12,00 ─ 16,00	8	22,22
16,00 ─ 20,00	5	13,89
20,00 ─ 24,00	1	2,78
Total	36	100,00

Classes de salários	Ponto médio $s_i$	Frequência $n_i$	Porcentagem $100 f_i$
4,00 ─ 8,00	6,00	10	27,78
8,00 ─ 12,00	10,00	12	33,33
12,00 ─ 16,00	14,00	8	22,22
16,00 ─ 20,00	18,00	5	13,89
20,00 ─ 24,00	22,00	1	2,78
Total	—	36	100,00

A importância do ponto médio

## Distribuição de Frequência



## Tipos de Gráficos

Variáveis Qualitativas - agrupamentos

Gráfico - Barras

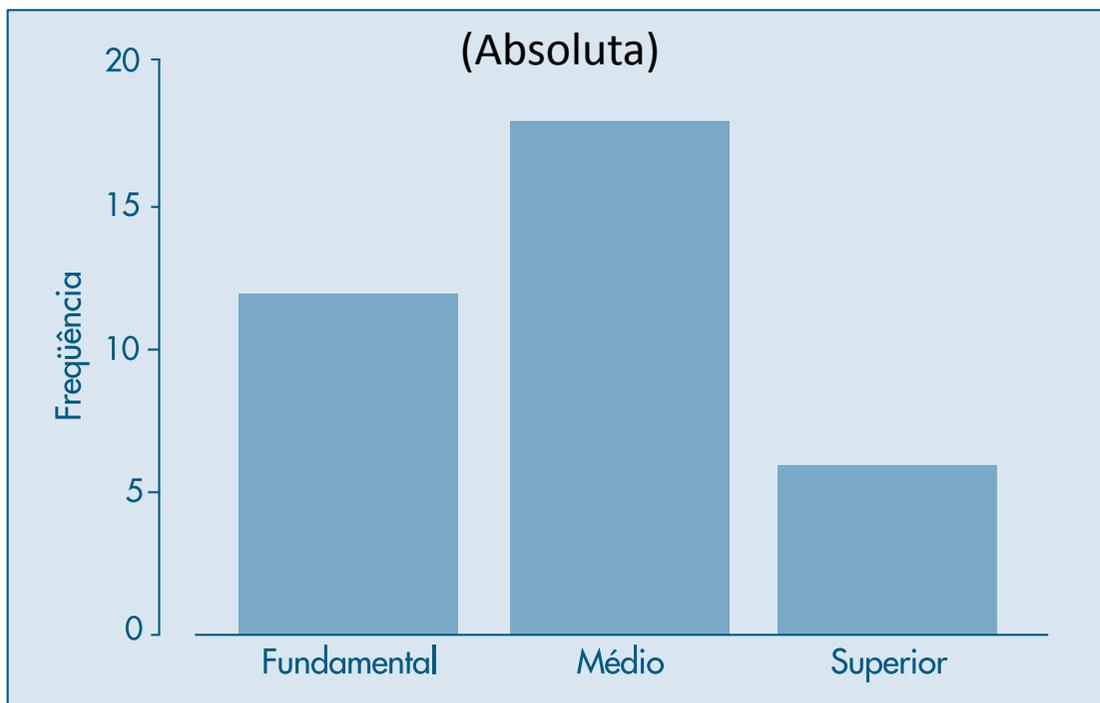
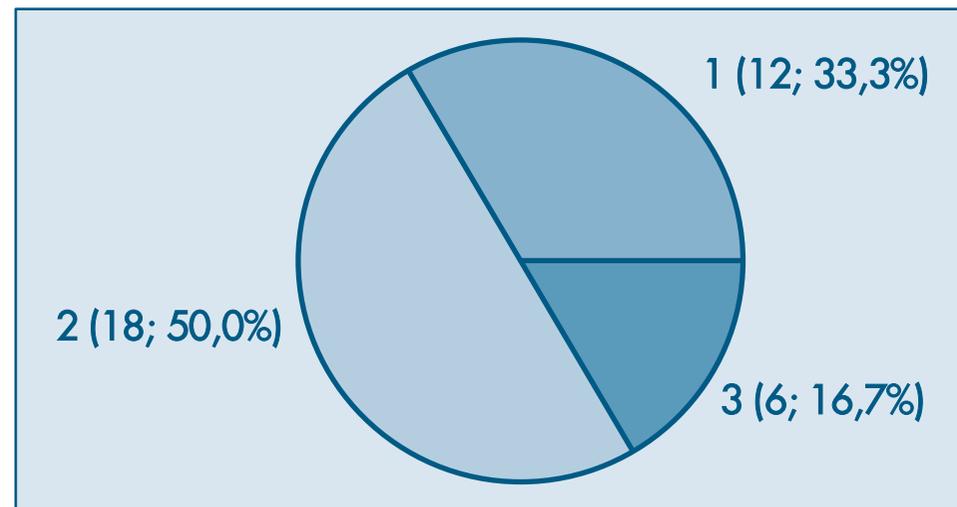


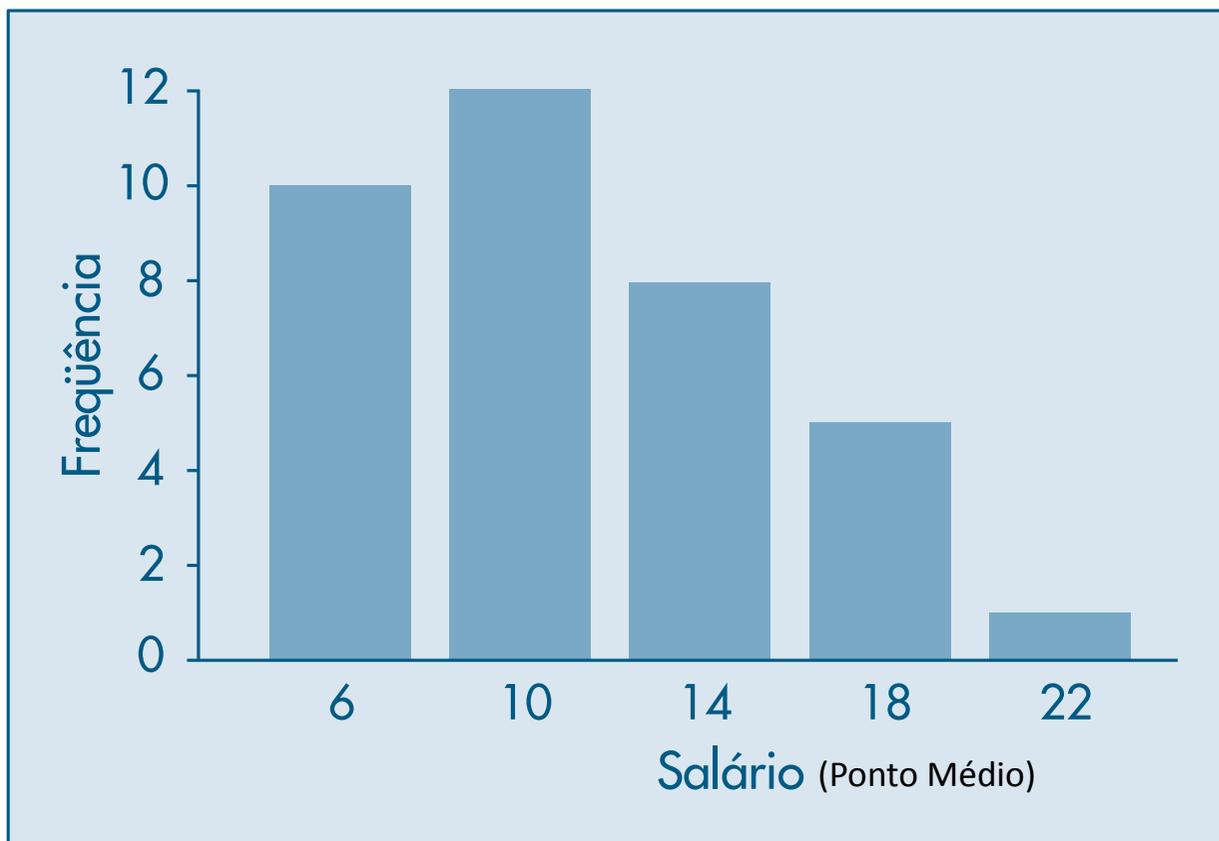
Gráfico - 2



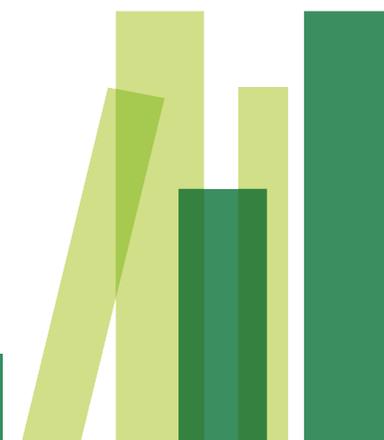
1 = Fundamental, 2 = Médio e 3 = Superior

## Tipos de Gráficos

Gráfico - Barras

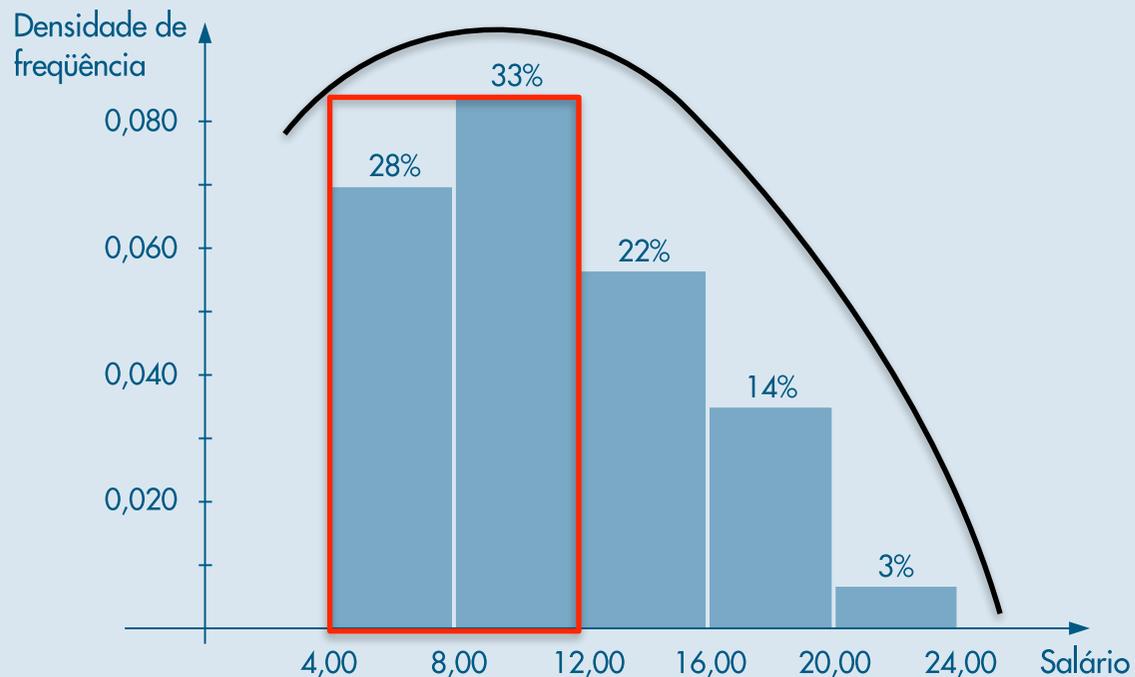


Classes de salários	Ponto médio $s_i$	Frequência $n_i$	Porcentagem $100 f_i$
4,00 ┤ 8,00	6,00	10	27,78
8,00 ┤ 12,00	10,00	12	33,33
12,00 ┤ 16,00	14,00	8	22,22
16,00 ┤ 20,00	18,00	5	13,89
20,00 ┤ 24,00	22,00	1	2,78
Total	—	36	100,00



## Tipos de Gráficos

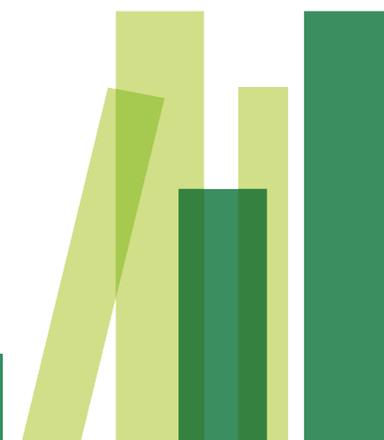
Gráfico - Histograma

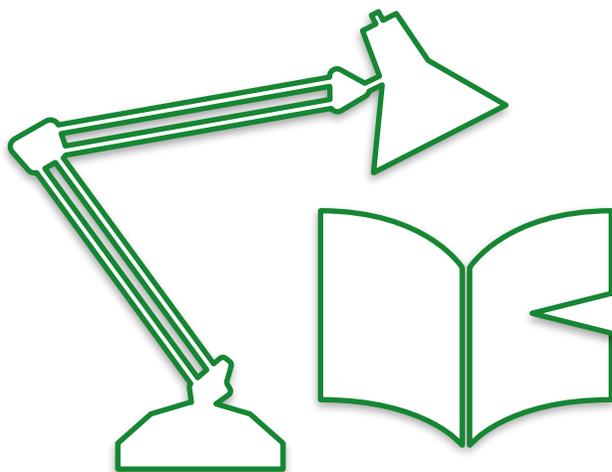


61% dos funcionários recebem menos de 12 sm

Área embaixo da curva = 100%

Classes de salários	Ponto médio $s_i$	Freqüência $n_i$	Porcentagem $100 f_i$
4,00 – 8,00	6,00	10	27,78
8,00 – 12,00	10,00	12	33,33
12,00 – 16,00	14,00	8	22,22
16,00 – 20,00	18,00	5	13,89
20,00 – 24,00	22,00	1	2,78
Total	—	36	100,00

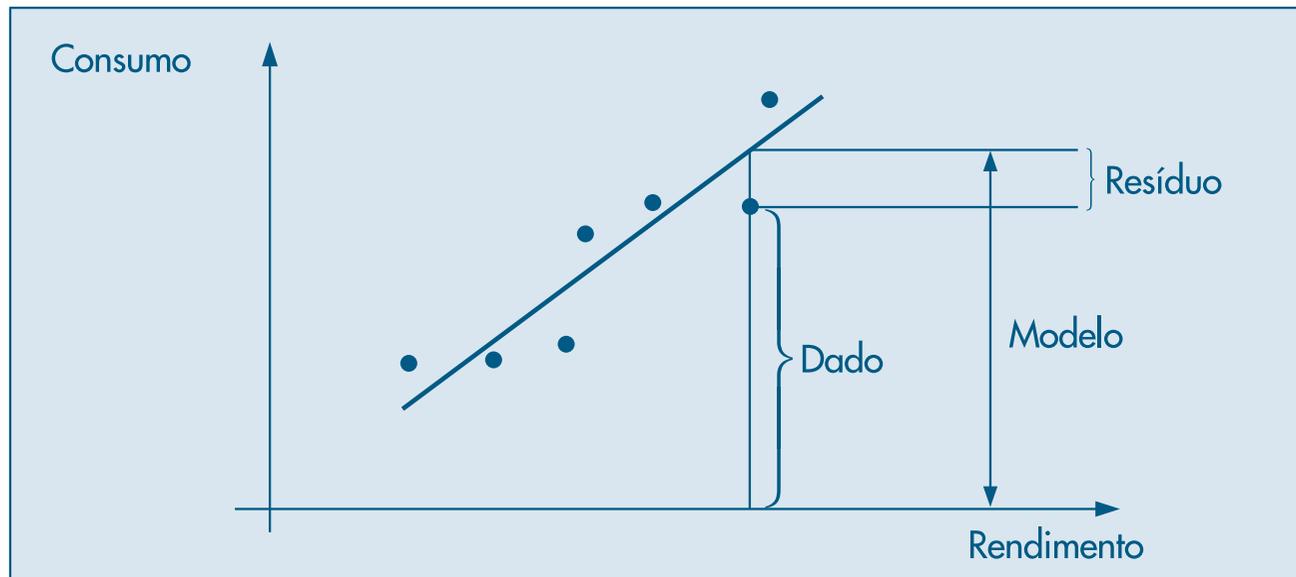




Inferência Estatística  
- Modelagem

É possível pensar em algo preditivo?





Podemos, então, escrever de modo esquemático:

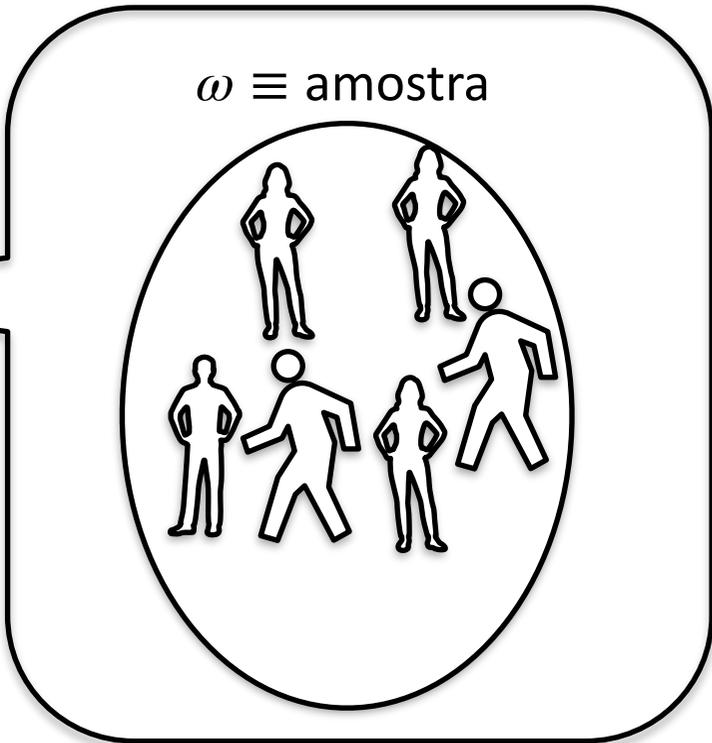
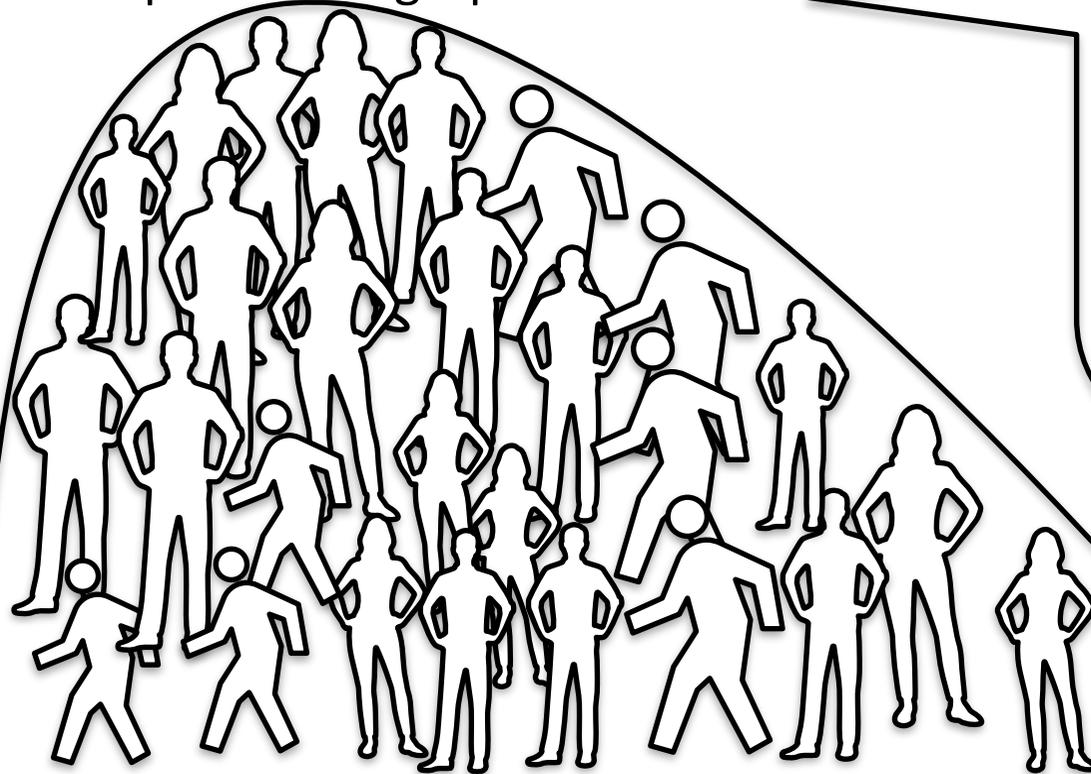
ou, ainda,

$$C = f(R) = a + bR$$

(1.1)

➔ População e Amostra

$\Omega \equiv$  Remuneração dos - CEO's  
Cooperativas Agropecuárias



com  $\omega \in \Omega$



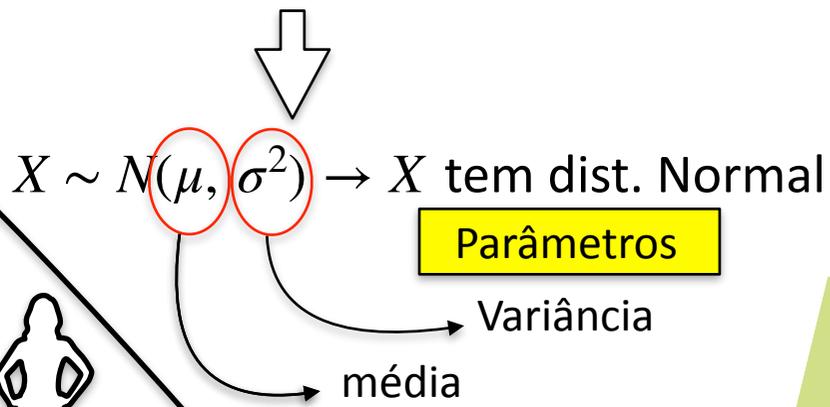
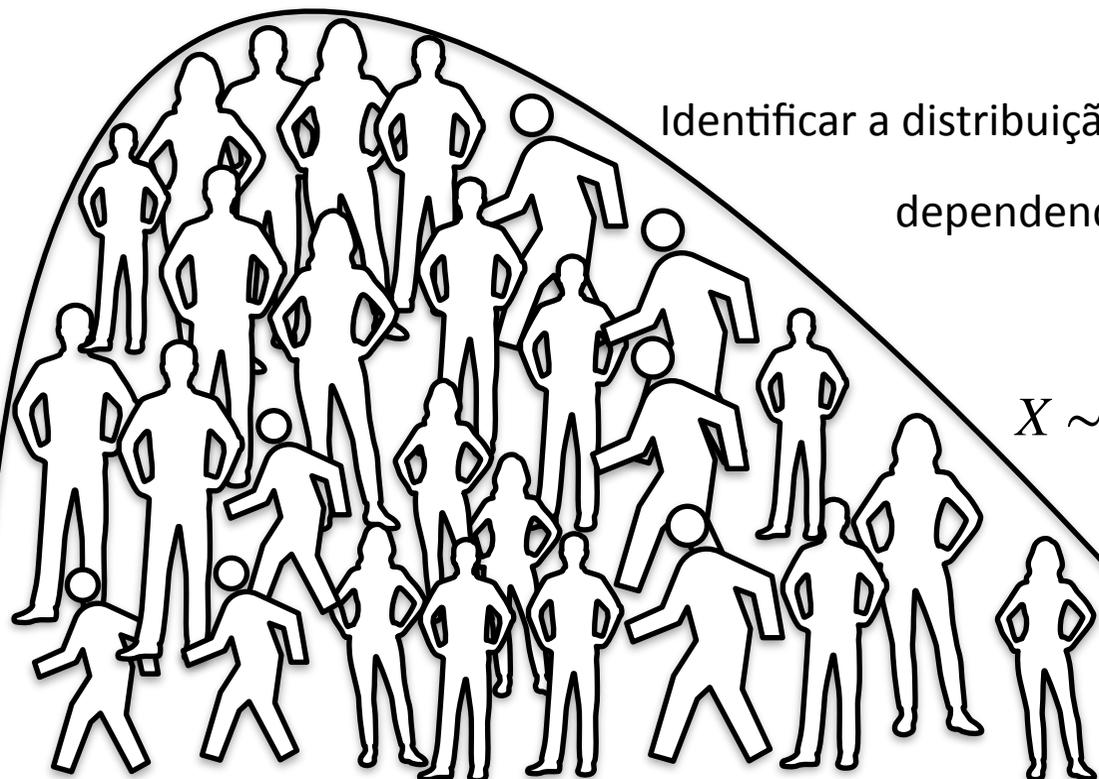
→ População e Amostra

→ Precisamos de mais informações

Inferência: Fazer afirmações sobre a população como um todo, a partir de dados coletados de uma parte da população.

$\Omega \equiv f(x)$  – função densidade de probabilidade

Podemos plotar os dados e verificar o formato da curva

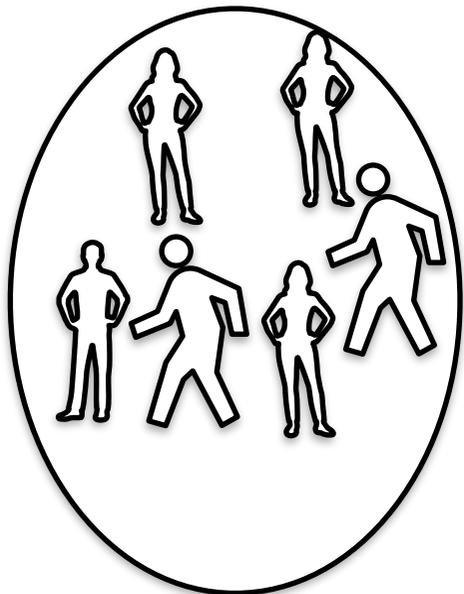


Em geral - desconhecidos



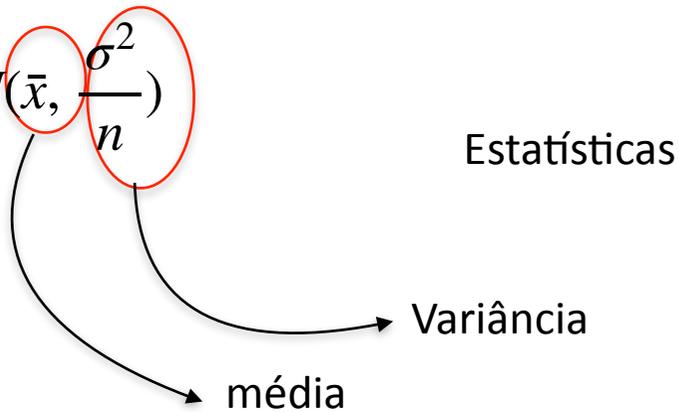
# População x Amostra

Amostra



$$\text{Se } X \sim N(\mu, \sigma^2)$$

$$\Rightarrow x \sim N(\bar{x}, \frac{\sigma^2}{n})$$



Mensuráveis

Deveríamos poder fazer afirmações sobre a população a partir da sua análise



## Background - Distribuição Amostral da Média

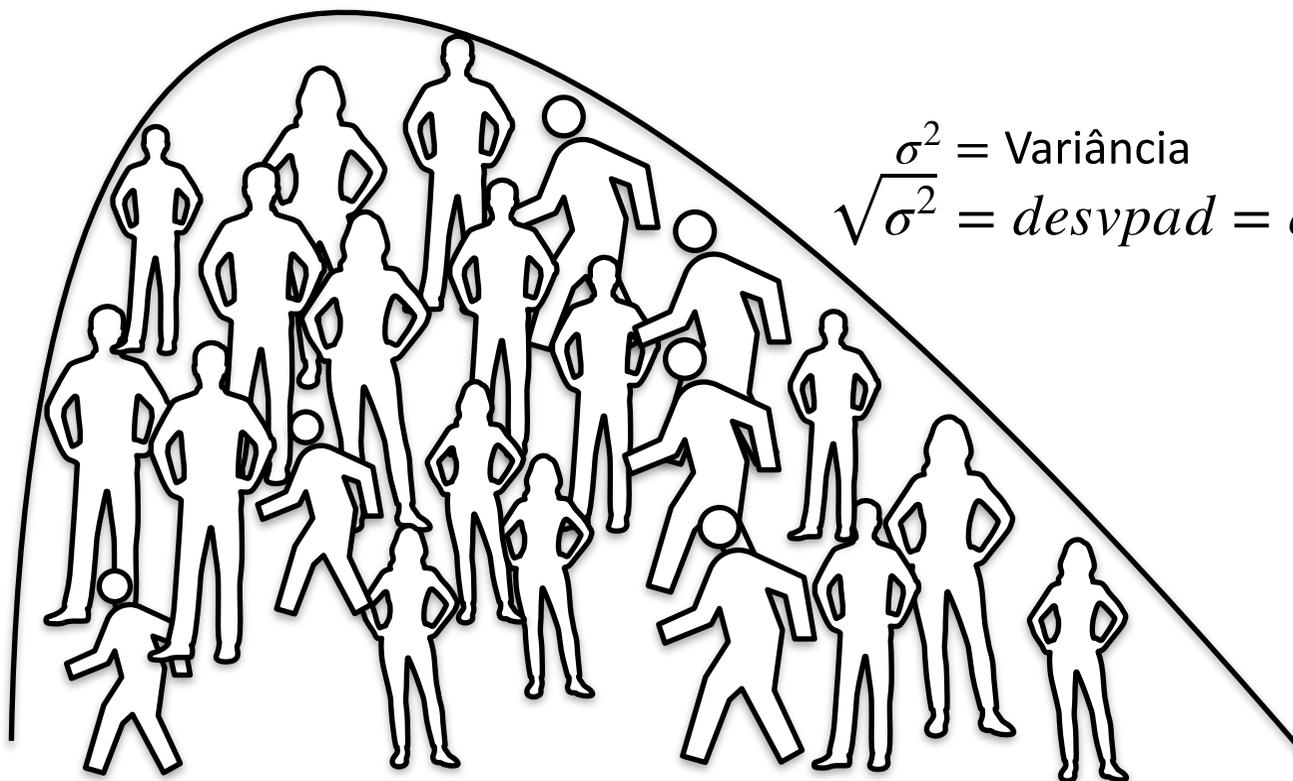
→ População - parâmetros

$\mu = \text{Média}$

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

$\sigma^2 = \text{Variância}$   
 $\sqrt{\sigma^2} = \text{desvpad} = \sigma$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$



## Material Complementar

**Recomendação:** Não se atenham aos videos abaixo. Fiquem a vontade a procurar outros.

### Tema - Coleta de Dados (videos - Inglês e Português)

1º [https://www.youtube.com/watch?v=nmL6m6iknBI&list=TLPQMTkwODIwMjAEODufmyI\\_EA&index=1](https://www.youtube.com/watch?v=nmL6m6iknBI&list=TLPQMTkwODIwMjAEODufmyI_EA&index=1)

2º [https://www.youtube.com/watch?v=kuyt0u\\_tFi4&list=TLPQMTkwODIwMjC2psnoZa\\_ilw&index=3](https://www.youtube.com/watch?v=kuyt0u_tFi4&list=TLPQMTkwODIwMjC2psnoZa_ilw&index=3)

3º <https://www.youtube.com/watch?v=Tr8VgJmcylQ>

### Link para textos a respeito

1º <https://mydatascope.com/blog/pt/conheca-os-diferentes-metodos-de-coleta-de-dados/>

2º <https://www.ibccoaching.com.br/portal/mercado-trabalho/metodos-objetivos-coleta-de-dados-empresas-por-que-como-sao-feitas/>

## Material Complementar

**Recomendação:** Não se atenham aos videos abaixo. Fiquem a vontade a procurar outros.

### Tema - Banco de Dados e Tratamento

1º <https://www.youtube.com/watch?v=YhlEVEPsZqI>

2º - <https://www.youtube.com/watch?v=9PuO-oijkCI>

3º - [https://www.youtube.com/watch?v=SLY\\_ZcNYjIY&list=PL7iAT8C5wumpQWB8AFW7CwK2nlzh8ZdP9&index=3](https://www.youtube.com/watch?v=SLY_ZcNYjIY&list=PL7iAT8C5wumpQWB8AFW7CwK2nlzh8ZdP9&index=3)

4º - <https://www.youtube.com/watch?v=kKykrTV17oU>

Excel - planilha de dados  
Access  
SQL (Serve)  
Postgree SQL  
MySQL

Municípios do Brasil

Acentos;  
Espaços;  
Erros;  
Relações



## Material Complementar

**Recomendação:** Siga a ordem dos vídeos

### Vídeos

1º - [https://www.youtube.com/watch?v=mm2A5tKVIpq&list=TLPQMTkwODIwMjC2psnoZa\\_ilw&index=2](https://www.youtube.com/watch?v=mm2A5tKVIpq&list=TLPQMTkwODIwMjC2psnoZa_ilw&index=2)

2º - <https://www.youtube.com/watch?v=q3wmL9quPwU>

### Textos

1º - <https://www.voitto.com.br/blog/artigo/business-intelligence>

2º - <http://datascienceacademy.com.br/blog/qual-a-diferenca-entre-o-analista-de-bi-e-o-cientista-de-dados/>



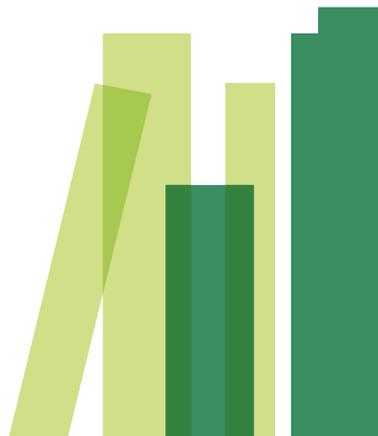
## Visualização de Dados, Relatórios e BI



## Data Science / Machine Learning / Deep Learning



Fonte: <http://datascienceacademy.com.br/blog/qual-a-diferenca-entre-o-analista-de-bi-e-o-cientista-de-dados/>



Atividade 2 - estará disponível no Moodle na quinta-feira

Data de entrega - 25/08/2022