

# ESCOLA DE ENGENHARIA MAUÁ

## EFB803 Estatística

2020  
Aula 02



(Análise descritiva: medidas resumo)

### Definições

**População:** conjunto de todos os elementos de interesse em estudo

**Amostra:** subconjunto representativo da população que será estudado para tirar conclusões para a população toda

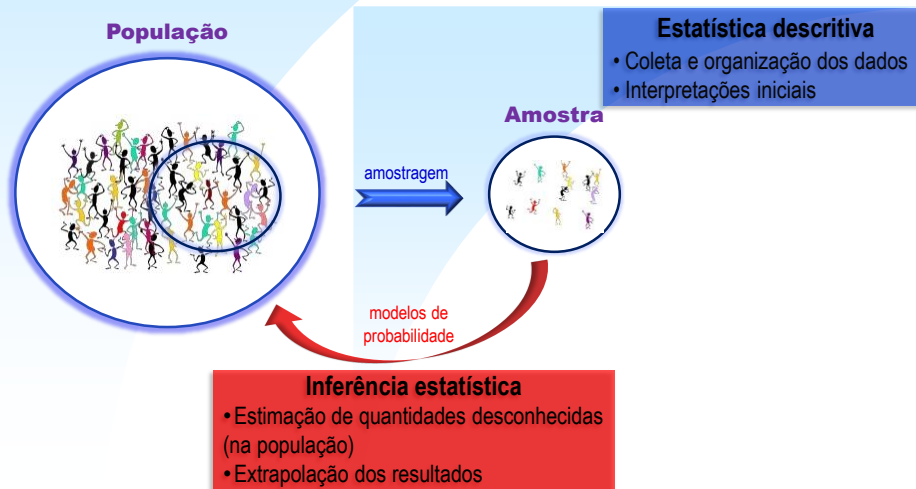
**Variável:** toda característica que, observada em uma unidade experimental, pode variar de uma unidade para outra

**Parâmetro:** medida que descreve alguma característica da população

**Estimador ou estatística:** medida que descreve alguma característica da amostra

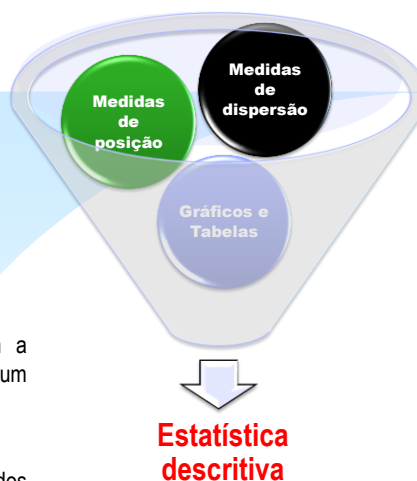
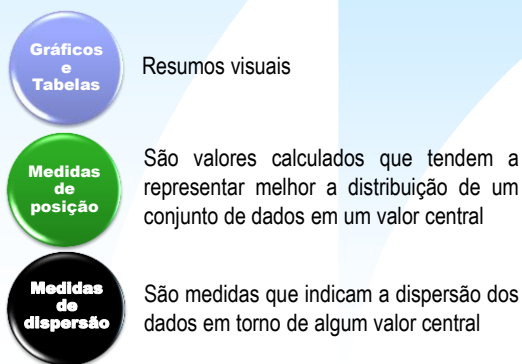
## População x Amostra

Para inferir (deduzir) certas características de uma população (pessoas, peças, etc...) deve-se trabalhar com uma amostra que seja representativa dessa população.



## Estatística descritiva

- ✓ Trata-se da organização, apresentação e descrição de um conjunto de dados (para uma ou mais variáveis);
- ✓ Os resumos descritivos podem ser organizados em tabelas, apresentados graficamente ou a partir de estimadores ou estatísticas de parâmetros da população.



## Classificação das variáveis

Saber classificar cada tipo de variável auxilia na busca de técnicas estatísticas mais adequadas para o resumo dos dados.

### Qualitativa

Exs.: Sexo, Classe social, Qualidade da peça (boa ou defeituosa), Escolaridade, ...

### Quantitativa

Exs.: Salário (R\$), Nº defeitos por peça, Volume (mL), Estatura (m), Nº filhos, ...

### Discreta

Ex.: Nº defeitos por peça, Nº de filhos...

### Contínua

Ex.: Salário (R\$), Peso (kg), Estatura (m), ...

**Qualitativa:** as respostas desse tipo de variável representam diferentes categorias que se distinguem por alguma característica não numérica.

**Quantitativa:** as respostas desse tipo de variável consistem em números que representam, em geral, contagem (discretas) ou medidas provenientes de alguma mensuração (contínuas).

## Exercício. Classifique as variáveis (2 minutos)

A partir do experimento realizado na aula anterior, as distâncias alcançadas pelas bolas de tênis de mesa foram mensuradas controlando algumas condições: a **configuração do elástico** (1: dois no 1º gancho ou 2: um em cada gancho) e a **posição do bloco limitador** (1: vertical, 2: horizontal e 3: sem).

Lançamento	Configuração do elástico	Posicao	Turma	Periodo	Dia	Distancia
1	1	1	EL	D	SEG	82,0
...	...	...	...	...	...	...
6	1	2	EL	D	SEG	79,0
...	...	...	...	...	...	...
10	1	3	CA	N	SEG	109,0
...	...	...	...	...	...	...
90	2	1	CV	D	SEG	51,0
...	...	...	...	...	...	...
200	2	2	QM	T	SEG	60,0
...	...	...	...	...	...	...
300	2	3	CM	T	SEG	79,0
...	...	...	...	...	...	...

Não é variável

Qualitativas

Quantitativa contínua

Além disso, também foram observadas as **turmas** (CV, CA, QM, MC, etc...), **período** (D, T e N) e **dia** em que as medidas foram tomadas (SEG, TER, etc...). Os dados foram tabulados e um resumo é apresentado ao lado.

Classifique as variáveis do estudo.

## Classificação das variáveis

Para cada tipo de variável existem técnicas estatísticas mais adequadas para o resumo dos dados.

### Qualitativa

Exs.: Sexo, Classe social, Qualidade da peça (boa ou defeituosa), Escolaridade, ...

#### Resumos estatísticos que podem ser feitos

- ✓ Contar a frequência absoluta de cada categoria
- ✓ Contar a frequência relativa (%) de cada categoria
- ✓ Construir gráficos de pizza, coluna, barras, ...

### Quantitativa

Exs.: Salário (R\$), Nº defeitos por peça, Volume (mL), Estatura (m), Nº filhos, ...

#### Resumos estatísticos que podem ser feitos

- ✓ Cálculo de medidas de posição (ou de localização)
- ✓ Cálculo de medidas de dispersão (ou variabilidade)
- ✓ Construir gráficos: boxplot, histograma, linha, dispersão, ...

## Medidas de posição

Estatísticas que tendem a representar melhor a distribuição dos dados de uma variável X em um único valor central. Fornecem uma idéia do "centro de gravidade" dos dados.

### ✓ Média da amostra ( $\bar{x}$ )

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$x_i$  : valor da i-ésima observação da variável X  
n : tamanho da amostra

### ✓ Mediana (Md)

É o valor que ocupa a posição central quando os dados estão ordenados

$$Md = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ for ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n+2}{2})}}{2}, & \text{se } n \text{ for par} \end{cases}$$

$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)} \leq x_{(5)} \leq \dots \leq x_{(n)}$   
menor valor da variável X      maior valor da variável X

### ✓ Moda (Mo)

É o valor (ou valores) de maior frequência na amostra (OBS.: pouco usada na prática)

**Por quê calcular a média? É “bobinho”...**

A Disney comprou os Estúdios Pixar em 2006 em um negócio de mais de 7 bilhões. Em uma análise dos filmes produzidos pela Disney e pela Pixar nos 10 anos anteriores ao negócio, os faturamentos resultam em:

**Disney** **P I X A R**

**US\$ 3,321 bilhões** **US\$ 3,231 bilhões**

O desempenho financeiro é parecido entre os dois estúdios?

### Exercício (4 minutos)

Na aba Bonus do arquivo “Aula 02.xlsx” são apresentados quatro cenários (A, B, C e D) de valores de bônus de Natal pagos a uma amostra de estagiários. Calcule a média e mediana da variável “bônus de final de ano de estagiário” e compare os resultados dessas duas medidas de posição para cada cenário.



## Medidas de dispersão (ou de variabilidade)

São medidas que indicam a dispersão dos dados em torno de algum valor central

### ✓ Variância amostral ( $s^2$ )

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \text{ ou } \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right]$$

### ✓ Desvio padrão (s)

O desvio padrão é definido pela raiz quadrada positiva da variância:  $s = \sqrt{s^2}$

### ✓ Coeficiente de variação ( $CV_{\%}$ )

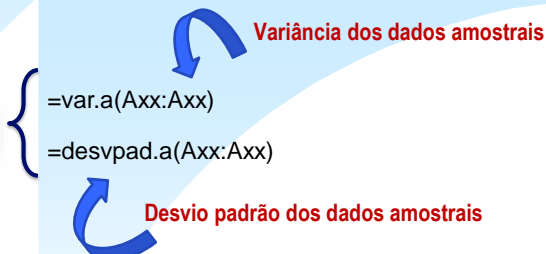
Indica a dispersão de um conjunto de dados em relação à sua média

Não existe um consenso, mas na prática, uma variável com  $CV_{\%}$  superior a 40% é considerada como tendo alta variabilidade

$$CV_{\%} = \frac{s}{\bar{x}} \cdot 100$$

## Exercício (2 minutos)

Na aba Bonus do arquivo “Aula 02.xlsx” são apresentados quatro cenários (A, B, C e D) de valores de bônus de Natal pagos a uma amostra de estagiários. Calcule o desvio padrão e o coeficiente de variação e avalie em quais situações a variabilidade dos dados é maior.



### OBSERVAÇÕES

- Com as fórmulas acima, obtemos a variância e o desvio padrão amostrais de uma variável de interesse;
- O Excel oferece calcular a variância e o desvio padrão populacionais (`var.p` e `desvpad.p`, respectivamente). Na prática, não são muito utilizadas;
- Não existe uma fórmula pronta no Excel para calcular o CV.

### Exercício (Por que calculamos a raiz quadrada da variância?)

Foram selecionadas aleatoriamente as notas da P<sub>1</sub> de Estatística de um grupo de alunos que prestava atenção na aula (X, com pontuação de 0 a 10) e de um grupo de alunos que só ficava na conversa/celular durante as aulas (Y, com pontuação de 0 a 10). Os dados são apresentados abaixo:

Presta Atenção	6,0	6,5	8,0	9,0	5,0	5,5	8,0	9,0	9,5	9,5	10,0	10,0
Conversa e celular	2,5	3,0	4,5	5,5	1,5	2,0	4,5	7,5	6,0	6,0	6,5	8,5

Dados para auxiliar os cálculos:  $\sum X_i = 96,00$  pontos  $\sum X_i^2 = 804,00$  pontos<sup>2</sup>  
 $\sum Y_i = 58,00$  pontos  $\sum Y_i^2 = 335,00$  pontos<sup>2</sup>

- Calcule a média da P1 para cada grupo.
- A variância das notas em cada grupo parece alta? É melhor apresentar a variância ou o desvio padrão?

### Exercício. Tabela dinâmica (15 minutos)

Neste exercício iremos trabalhar com uma amostra de 570 lançamentos. O conjunto de dados está disponível em "Aula 02.xlsx". Com eles, trace um perfil resumido dos resultados obtidos apenas para o que se pede abaixo:

- Verifique se há efeito da configuração dos elásticos do experimento. Ou seja, compare a distância média somente entre as duas configurações utilizadas (desconsiderando os demais fatores presentes no estudo). Calcule também o desvio padrão das distâncias obtidas nas duas configurações. Quais são as suas interpretações iniciais? Parece existir ou não tal efeito?
- Compare as distâncias médias somente entre as três posições utilizadas do bloco limitador e também calcule os desvios padrão de cada um dos três grupos. Parece existir o efeito do limitador no alcance das bolinhas?
- Repita o que foi pedido nos itens acima, agora comparando as distâncias médias (e calcule os respectivos desvios padrão) entre as habilitações que participaram dos lançamentos na segunda. As distâncias parecem similares entre os cursos?

## Distribuição de frequências

Quando os dados estão dispostos em uma tabela de frequências com k classes, calcula-se a **média amostral** pela expressão

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

7 classes

Número de defeitos	Frequência
0	13
1	11
2	9
3	8
4	6
5	1
6	2

**Exercício.** Uma amostra de 50 peças foi selecionada pelo controle de qualidade de uma empresa. A variável X de interesse é o número de defeitos por peça. Em média, quantos defeitos há nessa amostra?

$$\bar{x} \cong 1,9 \text{ defeitos}$$

**Pense em como calcular a mediana nesse caso. Quanto ela vale?**

$$Md = 2 \text{ defeitos}$$

## Distribuição de frequências

Numa tabela de frequências, composta de k classes, a **variância amostral** pode ser calculada por:

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}{n-1} \text{ ou } \frac{1}{n-1} \left[ \sum_{i=1}^k x_i^2 f_i - \frac{\left( \sum_{i=1}^k x_i f_i \right)^2}{n} \right]$$

7 classes

Número de defeitos	Frequência
0	13
1	11
2	9
3	8
4	6
5	1
6	2

**Exercício.** Uma amostra de 50 peças foi selecionada pelo controle de qualidade de uma empresa. A variável X de interesse é o número de defeitos por peça. Qual é o valor do desvio padrão de X? Calcule.

$$s \cong 1,7 \text{ defeitos}$$



## Distribuições de frequências: dados agrupados

Quando os dados estão dispostos em uma tabela de frequências com  $k$  classes, porém com valores agrupados, utiliza-se as mesmas expressões o cálculo da média e da variância apresentados nas duas telas anteriores.

Nesse caso, cada  $x_i$  representa o ponto médio da classe  $i$

Rendimento mensal (%)	Frequência
0,60  — 0,70	4
0,70  — 0,80	2
0,80  — 0,90	4
0,90  — 1,00	2

**Como calcular a média e a variância da amostra agora?**

**Exercício** Uma amostra do rendimentos mensais de certa aplicação financeira foi selecionada e os dados foram apresentados consolidados numa tabela de frequência agrupada. Como calcular a média e o desvio padrão do retorno financeiro agora? Calcule.

$$\bar{x} \cong 0,783 \%$$

$$s^2 \cong 0,013 \%^2 \rightarrow s \cong 0,115 \%$$

## Estudo recomendado

➡ Fazer o(s) exercício(s) não finalizados na aula



### Cap. 1

Seção 1.1 a 1.5 e seus respectivos exercícios



### Cap. 1 e Cap. 6

Seção 6.1 e seus respectivos exercícios