



1

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

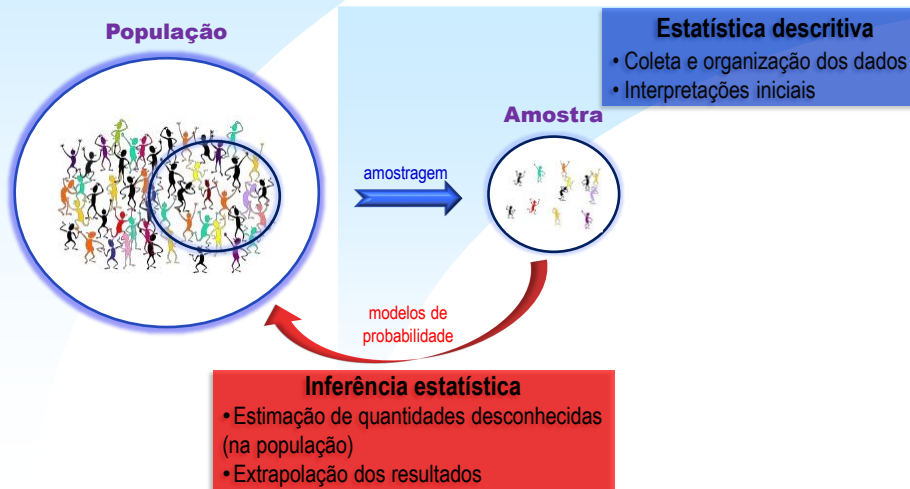
Parâmetro: medida que descreve alguma característica da população

Estimador ou estatística: medida que descreve alguma característica da amostra

2

População x Amostra

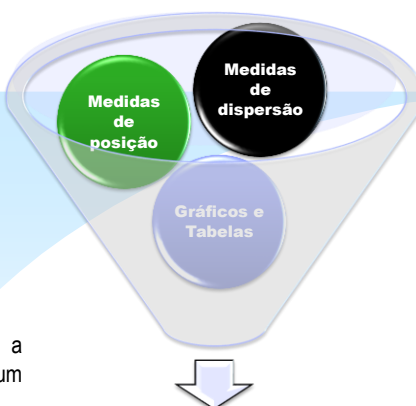
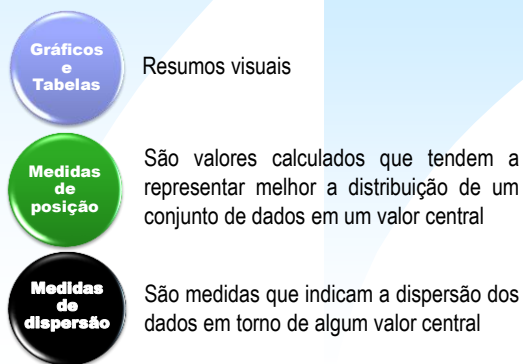
Para inferir (deduzir) certas características de uma população (pessoas, peças, etc...) deve-se trabalhar com uma amostra que seja representativa dessa população.



3

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.



Estatística descritiva

4

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), ...

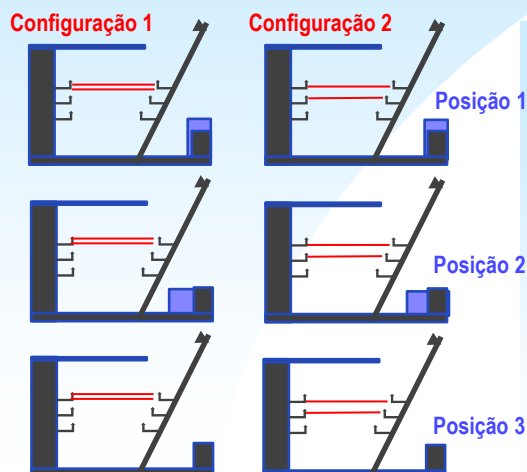
Qualitativa: as respostas desse tipo de variável representam diferentes categorias que se distinguem por alguma característica não numérica.

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

5

Exercício: classificação das variáveis

A partir de um experimento realizado em anos anteriores, bolas de tênis de mesa foram lançadas por uma catapulta controlando algumas condições: a **configuração do elástico** (1: dois em um único gancho ou 2: um elástico em dois ganchos) e a **posição de um bloco limitador** (1: vertical, 2: horizontal e 3: sem).



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...).

6

Exercício: classifique das variáveis do experimento.

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

7

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, ...

8

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)

9

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:



US\$ 3,321 bilhões

P I X A R

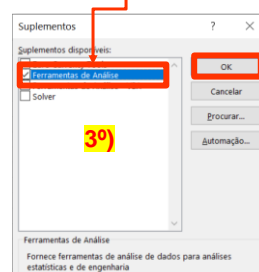
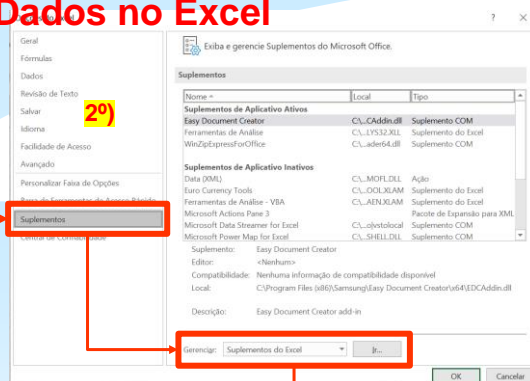
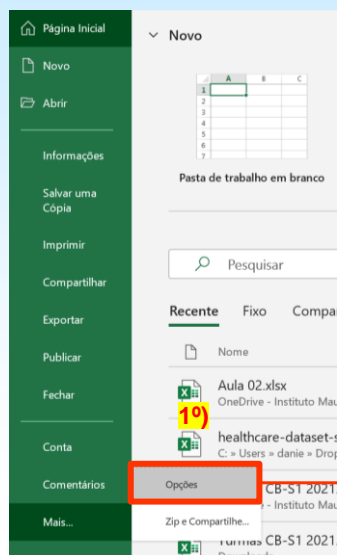
US\$ 3,231 bilhões

O desempenho financeiro é parecido entre os dois estúdios?

10

Instalar a Análise de Dados no Excel

→ clicar em “Arquivo” e...



11

Exercício: por quê calcular média e mediana?

Na aba bônus do arquivo “Aula02.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.



=média(Axx:Axx)

=med(Axx:Axx)

Média aritmética dos dados

Mediana dos dados

12

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

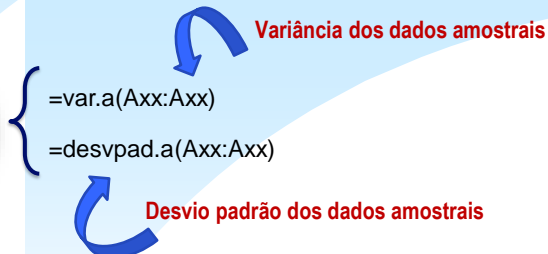
Não existe um consenso, mas na prática, uma variável com $CV_{\%}$ superior a 40% é considerada como tendo alta variabilidade

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

13

Exercício: por quê medir a variabilidade?

Na aba bônus do arquivo “Aula02.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.

14

Exercício (Excel - ferramentas de análise / análise de dados)

De acordo com a Organização Mundial da Saúde, o AVC é a 2ª causa de morte no mundo, responsável por aproximadamente 11% do total de mortes.

Este conjunto de dados é usado para prever se um paciente tem probabilidade de desenvolver AVC com base em parâmetros de entrada como sexo, idade, várias doenças e tabagismo. Cada linha nos dados fornece informações relevantes sobre o paciente.

Na aba AVC do arquivo "Aula02.xlsx" são apresentadas as informações de 5110 pessoas.

- Classifique as variáveis e calcule os resumos numéricos univariados usando a ferramenta Análise de Dados do Excel.
- Quais são os prós e contras dessa ferramenta?

15

Distribuição de frequências

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

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

(não tem fórmula pronta no Excel)

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?

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

16

Distribuição de frequências

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

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

(não tem fórmula pronta no Excel)

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.

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

17

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.

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

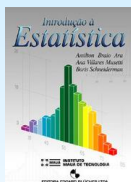
$$s^2 \cong 0,013 \%^2 \rightarrow s \cong 0,115 \%$$

18

Estudo recomendado

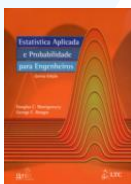


Fazer o(s) exercício(s) não finalizados na aula



Cap. 1

Seção 1.1 a 1.5 e seus respectivos exercícios



Cap. 1 e Cap. 6

Seção 6.1 e seus respectivos
exercícios

19



20