

AMOSTRAGEM



Objetivo da Estatística: fornecer informação (conhecimento), utilizando quantidades numéricas.

1. Obtenção dos dados ⇒ Amostragem
2. Descrição, classificação e apresentação dos dados ⇒ Estatística descritiva
3. Conclusão a tirar dos dados ⇒ Inferência Estatística



Amostragem Aleatória: processo que garante que todos os elementos da população têm as mesmas hipóteses de ser integrados na amostra.

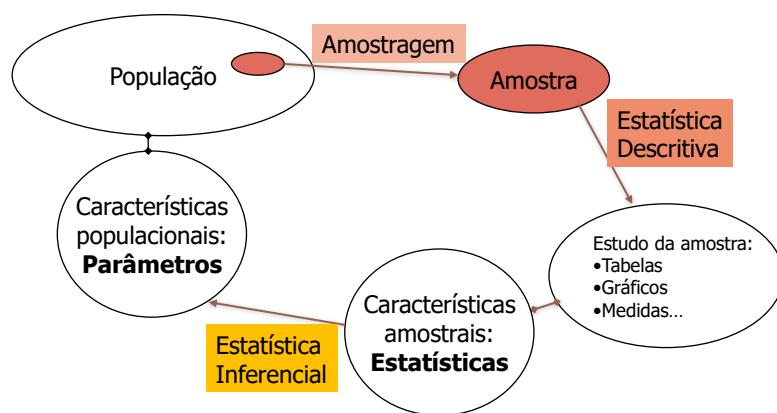
Estatística Descritiva: síntese e representação de uma forma compreensível da informação contida num conjunto de dados – construção de tabelas, gráficos ou cálculo de medidas centrais e de dispersão.

Inferência Estatística: a partir de um conjunto limitado de dados (amostra), pretende-se caracterizar o todo a partir do qual os dados foram obtidos (população).

3



Procedimento estatístico



4



População

Coleção de unidades individuais, com uma ou mais características comuns, que se pretendem analisar.

- Exemplos:

- Alunos da Universidade do Minho
- Potenciais eleitores para as eleições presidenciais

5



Parâmetros

Características numéricas que descrevem a população. Estas características são, em geral desconhecidas.

- Exemplos:

- Alunos da Universidade do Minho
 - Altura média dos alunos
 - Classificações médias obtidas a Estatística Aplicada...
- Potenciais eleitores para as eleições presidenciais
 - Proporção de eleitores decididos a votar
 - Idade média dos eleitores...

6



Amostra

Subconjunto da população, que se observa com o objetivo de tirar conclusões para a população de onde foi retirada.

- **Exemplos:**

- Alunos da Universidade do Minho
 - Amostra de 100 alunos inscritos na Universidade do Minho
- Potenciais eleitores para as eleições legislativas
 - Amostra de 1000 potenciais eleitores recenseados.

7



Estatística

Característica numérica que descreve a amostra. Calcula-se o valor da estatística a partir dos valores observados na amostra. Utiliza-se a estatística para estimar um parâmetro desconhecido.

- **Exemplos:**

- Alunos da Universidade do Minho
 - Altura média de 100 alunos da amostra
- Potenciais eleitores para as eleições legislativas
 - Proporção de eleitores que estão decididos a votar dos 1000 eleitores da amostra

8



Recenseamento ou censo

Estudo científico de um universo de pessoas, instituições ou objetos físicos com o propósito de adquirir conhecimentos, observando todos os seus elementos, e fazer juízos quantitativos acerca das características importantes desse universo.

- Exemplo:

- XV Recenseamento Geral da População Portuguesa (2011) (Instituto Nacional de Estatística – INE)
<http://www.ine.pt>

9



Sondagem

Estudo científico de uma parte de uma população com o objetivo de estudar atitudes, hábitos e preferências da população relativamente a acontecimentos, circunstâncias e assuntos de interesse comum.

- Exemplos:

- Sondagens para obter informação acerca da atitude dos eleitores
- Sondagens para testar as preferências dos consumidores

10



Fases de uma sondagem

- Escolha da amostra (amostragem)
- Obtenção de informação
- Análise dos dados
- Relatório final

11

Representatividade da amostra



A amostra deve ser tão representativa quanto possível da população que se está a estudar – deve-se evitar o enviesamento.

- Exemplos de amostras enviesadas ou tendenciosas:
 - Amostragem por conveniência
 - Utilizar uma amostra de sócios do SCP para prever o vencedor de um “derby” Benfica-Sporting
 - Utilizar uma amostra de alunos de um curso para tirar conclusões acerca do aproveitamento dos alunos universitários
 - Amostragem por resposta voluntária
 - Certas “sondagens” realizadas pelas estações de televisão utilizando respostas voluntárias, por exemplo, por SMS

12



Amostra aleatória simples

Dada uma população, uma amostra aleatória simples de dimensão n é um conjunto de n unidades da população, tal que qualquer outro conjunto de n unidades teria igual probabilidade de ser selecionado.

- Minimiza o enviesamento
- Tem em conta o princípio da aleatoriedade
- Recolha sem reposição

13

ESCALAS DE MEDIDA





ESCALAS

- Nominal
- Ordinal
- Intervalar
- Proporcional

15



NOMINAL

- Dados em categorias não ordenadas
- Variáveis classificadas por uma qualidade que possuem, um atributo
- Podem ser representadas por números sem significado
- Exemplos:
 - Preferência musical; cor dos olhos; sexo; classes sociais...

16



ORDINAL

- Ordem das categorias é importante
- Diferenças relativas e não quantitativas
- Podem ser representadas por números sem significado a não ser pela ordem

- Exemplos:
 - Classificação de ferimentos: 1-fatal, 2-grave, 3-moderado, 4-ligeiro; queimaduras, graus 1,2 e 3; alturas ou pesos ordenados em classes; classificação nos exames: 1-mau, 2-insuficiente, 3-suficiente, 4-bom e 5-muito bom.

17



INTERVALAR

- Escalas que possuem um intervalo constante mas não têm um zero absoluto
- Não é possível calcular razões porque o zero é arbitrário

- Exemplos:
 - Temperaturas em graus Celsius ou Fahrenheit – 20°C (68°F), 25°C (77°F), 5°C (41°F), 10°C (50°F); dados circulares, tempo ou orientação

18



PROPORTIONAL

- Existe um intervalo de tamanho constante entre unidades adjacentes
- Existe um zero com significado físico
- Exemplos:
 - Comprimentos – 30 cm (11,8 in), 60 cm (23,6 in); pesos; contagens, volumes, capacidades, velocidades, tempos de duração

19



DADOS

- Contínuos – existe um valor possível entre dois valores possíveis
 - um comprimento pode tomar uma qualquer valor entre dois limites
- Discretos – a variável só pode tomar certos valores
 - número de folhas de uma planta, o número de glóbulos brancos

20

ESTATÍSTICA DESCritIVA



DISTRIBUIÇÕES DE FREQUÊNCIA



- Tabelas de Frequência
 - listagem de todos os valores observados e determinação do número de vezes que um valor é observado



EXEMPLO

- Número de peixes tabulados de acordo com a pigmentação preta

Classe	Pigmentação	f_i
0	Sem	13
1	Ligeira	68
2	Moderada	44
3	Forte	21
4	Cheia	8

23



FREQUÊNCIAS

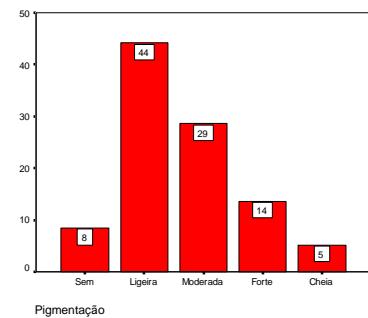
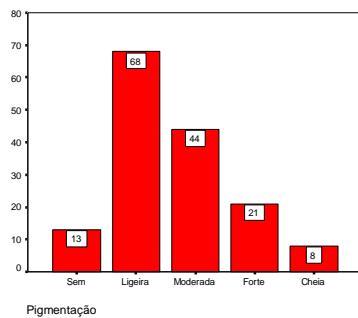
Pigmentação

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Sem	13	8.4	8.4	8.4
	Ligeira	68	44.2	44.2	52.6
	Moderada	44	28.6	28.6	81.2
	Forte	21	13.6	13.6	94.8
	Cheia	8	5.2	5.2	100.0
	Total	154	100.0	100.0	

24



GRÁFICOS DE BARRAS



25

EXEMPLO

- A tabela apresenta os tempos de espera (X) numa fila de supermercado de sujeitos selecionados aleatoriamente

4	18	8	25	5,5	7
7	26	8	16	2	1
12	3	2	9	16	4
21	7	13	27	8	8
27	4	34,5	19	7	5
18	9	12	16	2	6
12	10	7	21	3	1
0,5	11	10	13	4	5
20	1,5	5	7	12	2
8,5	12	5	10	18	0,5

26

CONSTRUÇÃO



- Número de observações, n

- Amplitude, R

- Número de classes, k

Regra de Sturges

$$k = 1 + 3.3 \log(n)$$

- Intervalo de classe, R/k

- Extremos de classe

n	k
25	5-6
50	6-7
100	7-8
500	9-10
1000	10-11

27

CONSTRUÇÃO



- Número de observações, n=60
- Amplitude, R=34,5-0,5=34,0
- Número de classes, k=7
- Intervalo de classe, R/k=34/7≈4,8≈5,0
- Extremos de classe, min=0,5

0- 5 ($0 \leq X < 5$)	15
5- 10	19
10- 15	11
15- 20	7
20- 25	3
25- 30	4
30- 35	1

28



ESTATÍSTICAS SUMÁRIAS

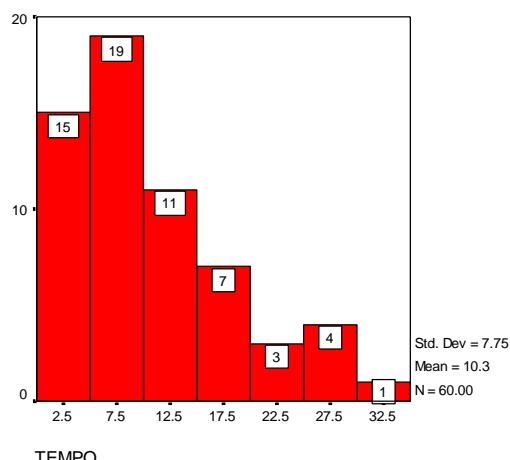
Statistics

TEMPO		
N	Valid	60
	Missing	0
Mean		10.267
Median		8.000
Std. Deviation		7.7462
Variance		60.0040
Range		34.0
Minimum		.5
Maximum		34.5
Percentiles	10	2.000
	20	4.000
	25	4.250
	30	5.000
	40	7.000
	50	8.000
	60	10.000
	70	12.000
	75	15.250
	80	17.600
	90	21.000

29



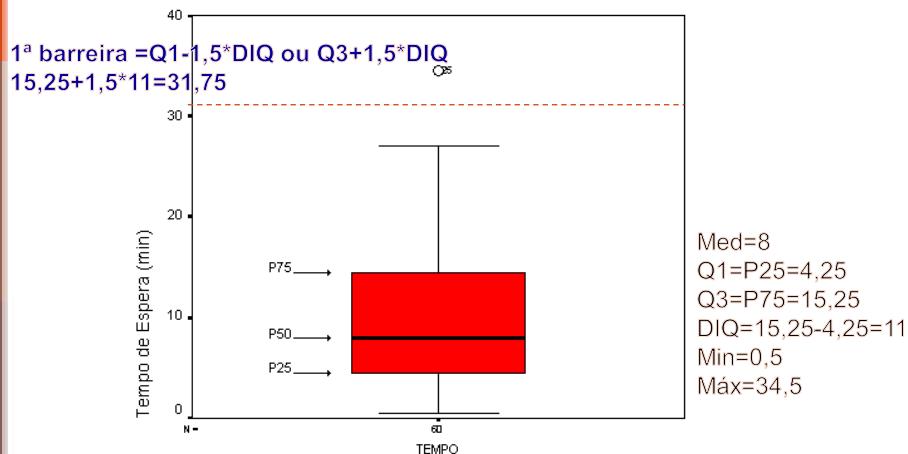
HISTOGRAMA



30



CAIXA DE BIGODES



31

MEDIDAS DE LOCALIZAÇÃO

■ **Média aritmética** $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ $\bar{X} \approx \sum_k f_{r_k} M_k = \frac{1}{n} \sum_k f_k M_k$

■ **Mediana** $Med = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+1}{2}\right)}}{2}$ n par

$$Med = x_{\left(\frac{n+1}{2}\right)} \quad n \text{ ímpar}$$

$$Med = LI + \frac{0.5 - F_{r_A}^-}{F_{r_{Med}}^- - F_{r_A}^-} \Delta \quad \text{representação histograma}$$

■ **Moda**

$$Mod = LI + \frac{d_1}{d_1 + d_2} \Delta \quad \text{com} \quad d_1 = f_{Mod} - f_A^-, \quad d_2 = f_{Mod} - f_D^+$$

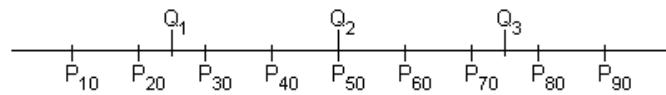
32

MEDIDAS DE LOCALIZAÇÃO



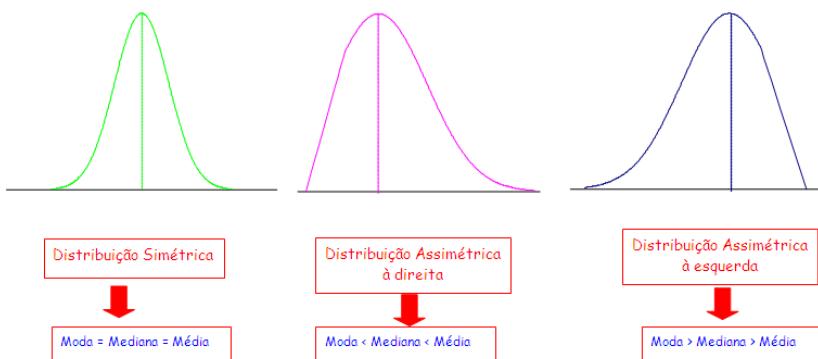
- Quartis $Q_1 = X_{(n+1)/4}$ $Q_2 = \text{Med}$ $Q_3 = X_{3(n+1)/4}$

- Percentis



33

Comparação entre as medidas de localização



34

MEDIDAS DE DISPERSÃO



■ Amplitude

$$R = x_{(n)} - x_{(1)}$$

■ Distância interquartílica

$$DIQ = Q_3 - Q_1$$

■ Variância

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

■ Desvio padrão

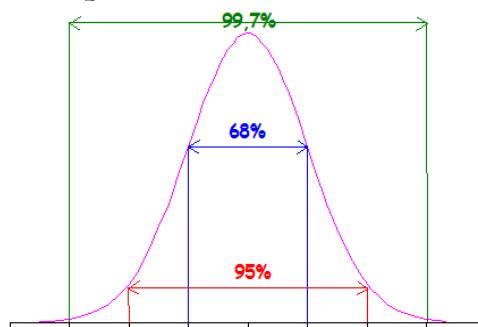
$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

■ Coeficiente de variação

$$V = \frac{s}{\bar{X}}$$

35

DISTRIBUIÇÃO NORMAL



❖ Distribuição simétrica em forma de sino, centrada em μ .

❖ 68% das observações pertencem a $\mu - \sigma ; \mu + \sigma [$;

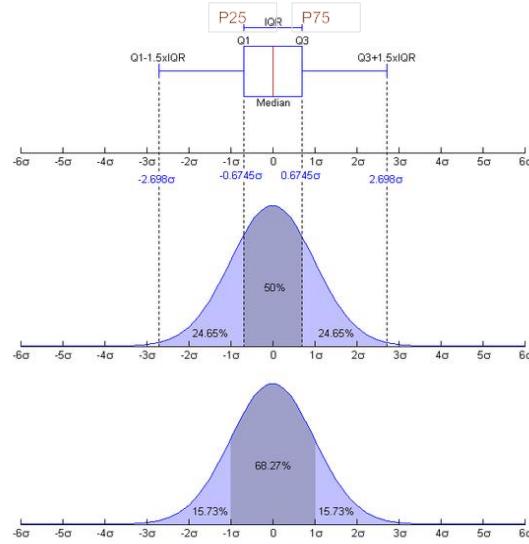
❖ 95% das observações pertencem a $\mu - 2\sigma ; \mu + 2\sigma [$

❖ 99,7% das observações pertencem a $\mu - 3\sigma ; \mu + 3\sigma [$

36



Densidade e box-plot



37



Alguns sites interessantes...

- ALEA - Ação Local de Estatística Aplicada
 - <http://alea-estp.ine.pt/>
- INE – Instituto Nacional de Estatística
 - <http://www.ine.pt/>

38