

Introdução à Probabilidade e Estatística



Universidade de Évora

Departamento de Matemática

Ano lectivo 2015/16

Ana Isabel Santos

Estatística Descritiva

Aula 1

Materiais necessários para aula prática

Ficha de exercícios

Calculadora

Formulário

Tabelas (a partir do capítulo 4, inclusive)

Caderno

Computador com Software SPSS (optativo)

Motivação para trabalhar

Estatística Descritiva

A **Estatística** (EACHS) é uma área do conhecimento que inclui os instrumentos necessários para estudar determinados fenómenos da sociedade. Por exemplo, estudos sobre a natalidade, o emprego, a criminalidade, a opinião sobre ... , etc.

- ▶ **Estatística Descritiva:** conjunto de técnicas que permitem recolher, organizar, reduzir e apresentar dados estatísticos.
- ▶ **Inferência Estatística:** conjunto de técnicas que permitem caracterizar uma certa população com base na informação amostral.

População vs Amostra



Classificação das variáveis

Variáveis Qualitativas: dados com características não numéricas, identificam uma qualidade ou característica.

- ◆ **Escala nominal:** dados divididos por categorias sem ordem.
Exemplo: sexo, distrito de residência, cor dos olhos, nacionalidade.
- ◆ **Escala ordinal:** dados divididos por categorias com sequência.
Exemplo: opinião sobre algo, estado civil, grau de escolaridade.

Variáveis Quantitativas: dados com características numéricas.

- ◆ **Discretas:** tomam um n° finito ou infinito numerável de valores.
Exemplo: n° de filhos, n° de pessoas que vão ao hospital por hora, nota final.
- ◆ **Contínuas:** tomam um n° infinito não numerável de valores.
Exemplo: idade, altura, peso, temperatura do ar, salário.

Dados

Variável em estudo - População X



Amostra em bruto - (x_1, x_2, \dots, x_n)



Amostra ordenada - $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$



Dados agrupados - *Tabela de Frequências*

Tabela de frequências para variáveis qualitativas e quantitativas discretas

Dados Agrupados

X'_i	n_i	f_i	N_i	F_i
x'_1	n_1	f_1	$N_1 = n_1$	$F_1 = f_1$
x'_2	n_2	f_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
\vdots	\vdots	\vdots	\vdots	\vdots
x'_k	n_k	f_k	$N_K = n$	$F_k = 1$
	n	1		

➤ X'_i – categoria; k – n.º de categorias; n – dimensão da amostra;

➤ n_i – frequência absoluta simples;

➤ $N_i = \sum_{j=1}^i n_j$ – frequência absoluta acumuladas;

➤ $f_i = \frac{n_i}{n}$ – frequência relativas simples;

➤ $F_i = \frac{N_i}{n} = \sum_{j=1}^i f_j$ – frequência relativas acumuladas.

Exemplo 1: Regiões onde se localizam hotéis com SPA em Portugal

X – regiões onde existem hotéis com SPA ;

$k = 4$

X'_i	n_i	f_i	N_i	F_i
Região norte	258	0.2346	258	0.2346
Região centro	395	0.3591	653	0.5937
Região Sul	224	0.2036	877	0.7973
Ilhas	223	0.2027	1100	1
	1100	1		

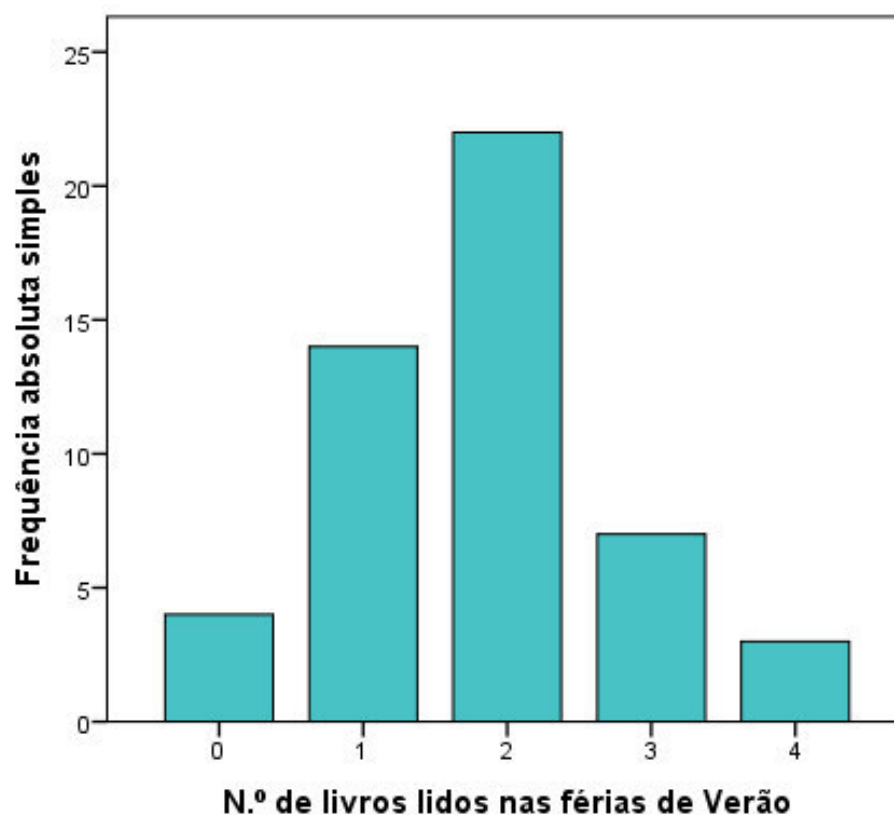
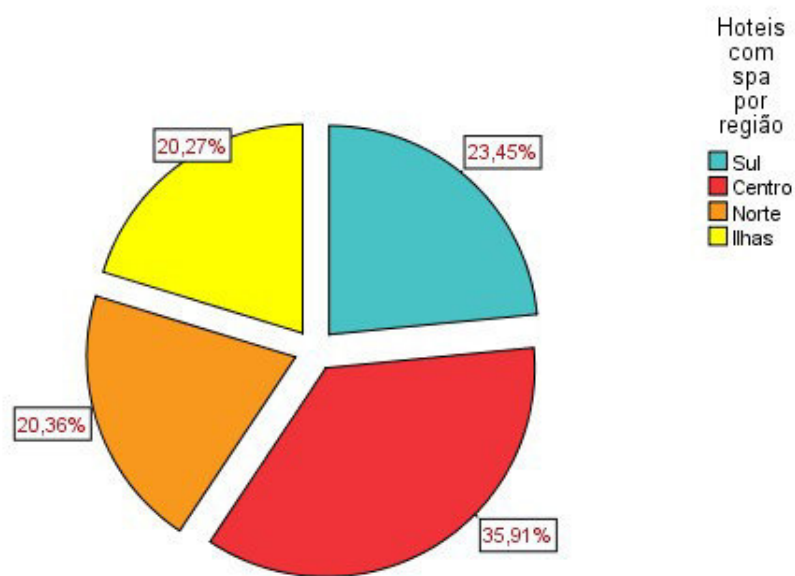
Exemplo 2: Número de livros lidos nas férias de Verão.

X – N.º de livros lidos nas férias de Verão, por 50 alunos de IPE.

X'_i	n_i	f_i	N_i	F_i
0	4	0.08	4	0.08
1	14	0.28	18	0.36
2	22	0.44	40	0.80
3	7	0.14	47	0.94
4	3	0.06	50	1
	50	1		

Representação gráfica de dados qualitativos e quantitativos discretos

Gráfico Circular e Gráfico de Barras



Para dados de natureza discreta, com um n.º pequeno de valores.

Exercício 1.1 :

Considere os resultados finais de Estatística de 20 estudantes de uma Universidade:

9 14 12 8 14 12 16 16 8 14 11 12 12 11 11 18 14 18 15 15

- a)** Os dados em estudo são de tipo qualitativo ou quantitativo?
- b)** Construa a tabela de frequências associada a estes dados.
- c)** Represente graficamente a informação disponibilizada.
- d)** Determine e interprete a média, a moda e a mediana.
- e)** Calcule a variância e o desvio padrão.
- f)** Calcule e interprete o valor do percentil 15 e do 8º decil.
- g)** Indique a amplitude da amostra e a amplitude interquartil.
- h)** Determine e interprete o coeficiente de variação.
- i)** Estude a distribuição dos dados quanto à assimetria e ao achatamento.
- j)** Apresente os dados numa caixa de bigodes.

Exercício 1.1:

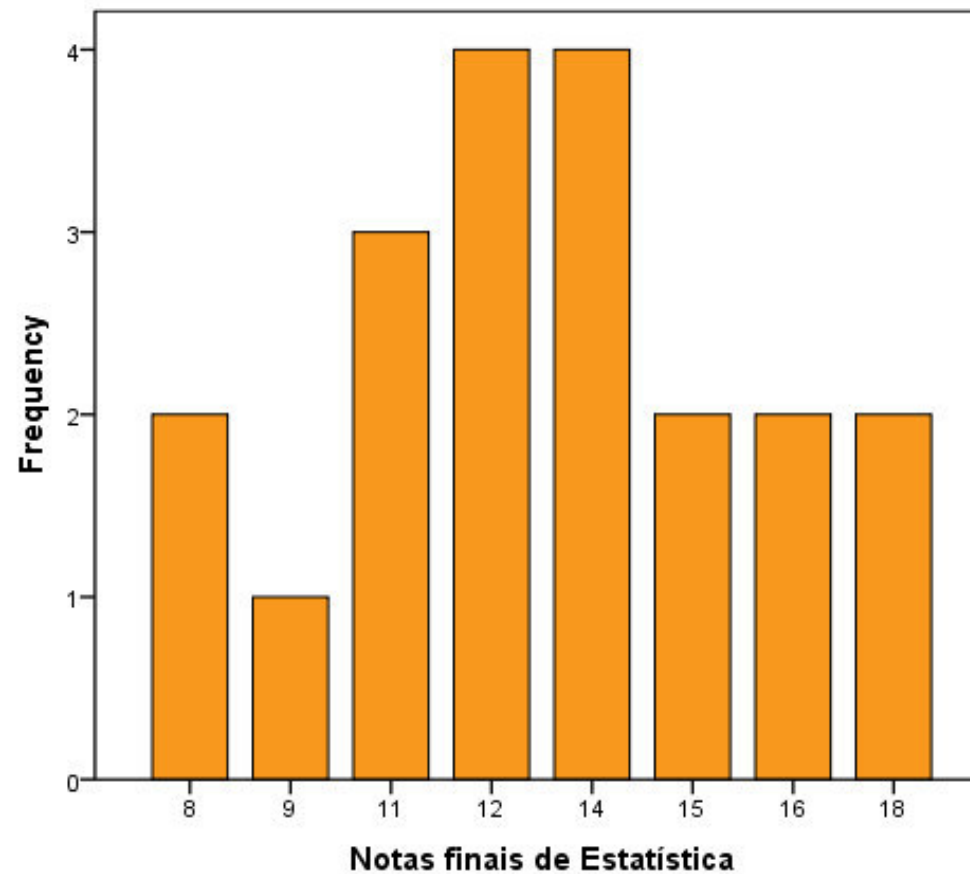
X – Notas finais de Estatística (variável quantitativa discreta)

Tabela de Frequências

X'_i	n_i	f_i	N_i	F_i
8	2	0.10	2	0.10
9	1	0.05	3	0.15
11	3	0.15	6	0.30
12	4	0.20	10	0.50
14	4	0.20	14	0.70
15	2	0.10	16	0.80
16	2	0.10	18	0.90
18	2	0.10	20	1
	20	1		

Exercício 1.1 : Notas finais de Estatística

Gráfico de Barras



Medida de Localização (Tendência Central)

Moda (mode) - M_0

Dados não agrupados:

$\hat{X} = M_0$ - Valor mais frequente

Dados agrupados qualitativos ou discretos:

$\hat{X} = M_0$ - Valor ou categoria com maior frequência simples.

Medida de Localização (Tendência Central)

Mediana (median) - M_e

Valor na amostra ordenada que tem 50% de valores inferiores ou iguais a ele e os restantes 50% superiores ou iguais.

Dados não agrupados ou discretos:

$$M_e = \tilde{X} = \begin{cases} \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2} + 1\right)}}{2} & \text{se } n \text{ par,} \\ X_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ ímpar.} \end{cases}$$

Medida de Localização (Tendência Central)

Média (mean) - \bar{X}

Dados não agrupados:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Dados agrupados:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i X'_i = \sum_{i=1}^k f_i X'_i$$

Medida de Localização (Tendência Não Central)

Quantis - Q_p - divide a amostra em duas partes.

Dados não agrupados ou discretos:

$$Q_p = \begin{cases} \frac{X_{(np)} + X_{(np+1)}}{2} & \text{se } np \text{ inteiro,} \\ X_{([np]+1)} & \text{se } np \text{ não inteiro,} \end{cases} \quad \text{para } p \in]0, 1[.$$

$$\text{Quantis} \begin{cases} \text{Quartis} = Q_i, & p = \frac{i}{4}, \quad i = 1, 2, 3; \\ \text{Decis} = D_i, & p = \frac{i}{10}, \quad i = 1, 2, \dots, 9; \\ \text{Percentis} = P_i, & p = \frac{i}{100} \quad i = 1, 2, \dots, 99; \end{cases}$$

Observações:

$$Q_1 = P_{25}, \quad Q_2 = \tilde{X} = D_5 = P_{50}, \quad Q_3 = P_{75}, \quad D_i = P_{i \times 10}.$$

Medidas de Dispersão

Amplitude total (range): $a = \text{máx} - \text{mín}$

Amplitude interquartil: $IQ = Q_3 - Q_1$

Intervalo de variação: $Q' = P_{90} - P_{10}$

Variância (variance):

Dados não agrupados

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

Dados agrupados

$$S^2 = \frac{1}{n-1} \sum_{i=1}^k n_i (X'_i - \bar{X})^2$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^k n_i X_i'^2 - n\bar{X}^2 \right)$$

Desvio-padrão (Std. Deviation): $S = \sqrt{S^2}$