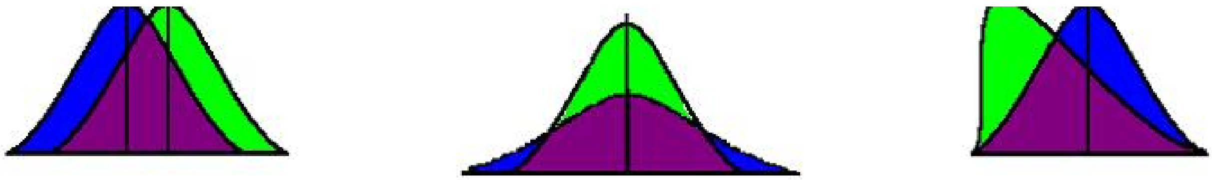


Medidas Descritivas

Apontamentos sobre as medidas descritivas: média, moda, mediana, quartis, variância, desvio padrão, coeficiente de variação e assimetria

Page

- As medidas descritivas podem ser 3:
    - Medidas de posição ou localização
    - Medidas de variabilidade
    - Medidas de assimetria

Propriedades:

- **Objetividade -** de modo a que 2 observadores distintos cheguem à mesma conclusão
- **Depender de todas as observações -** de modo a refletir a informação contida no conjunto de todas as observações
- **Ter um significado concreto -** de modo a facilitar a interpretação
- **Eficiência computacional e algébrica -** otimização e velocidade dos cálculos e simplicidade dos mesmos
- **Insensibilidade às variações das amostras -** de modo a que as medidas sejam robustas e estáveis

Medidas de Posição ou Localização Central

- As medidas de posição ou localização central são 3:
    - Média
    - Moda
    - Mediana

MÉDIA ARITMÉTICA (ou Média)

- **Média -** é obtida dividindo a soma de todos os valores numéricos observados pelo número total de observações.

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

- Nem sempre é possível considerar todas as observações para o cálculo da média, uma vez que as observações já podem ser dadas sob a forma de tabela de frequência, assim:

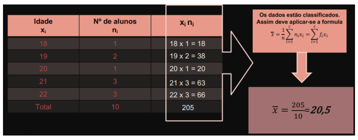
$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{c} n_i x_i}$$

Exemplo 1:

Exemplo 2

Considere a tabela seguinte que representa a idade de 10 alunos de um grupo de teatro

Calcula a média dos alunos



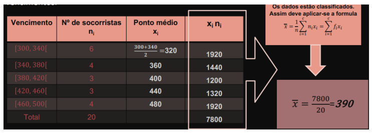
Idade x_i	Nº de alunos n_i		$x_i n_i$
18	1		18 x 1 = 18
19	2		19 x 2 = 38
20	1		20 x 1 = 20
21	3		21 x 3 = 63
22	3		22 x 3 = 66
Total	10		205

Os dados estão classificados. Assim deve aplicar-se a formula $\bar{x} = \frac{1}{n}\sum_{i=1}^{c} n_i x_i = \sum_{i=1}^{c} f_i x_i$

$\bar{x} = \frac{205}{10} = 20,5$

Os socorristas de uma praia estão classificados em 5 categorias de vencimento (em euros) consoante o nº de horas. Conforme mostra a tabela seguinte.

Calcula a média de vencimentos



Vencimento	Nº de socorristas n_i	Ponto médio x_i	$x_i n_i$
[300, 340[	6	$\frac{300+340}{2}=320$	1920
[340, 380[	4	360	1440
[380, 420[	3	400	1200
[420, 460[	3	440	1320
[460, 500[	4	480	1920
Total	20		7800

Os dados estão classificados. Assim deve aplicar-se a formula $\bar{x} = \frac{1}{n}\sum_{i=1}^{c} n_i x_i = \sum_{i=1}^{c} f_i x_i$

$\bar{x} = \frac{7800}{20} = 390$

MEDIANA

- **Mediana -** é o valor que particiona os dados previamente dispostos por ordem crescrente ou decrescente, em 2 grupos em que metade dos valores (50%) são iguais ou inferiores à mediana e a outra metade são iguais ou superiores à mediana

$$\tilde{x} = \begin{cases} x_{(\frac{n}{2})}+x_{(\frac{x}{2}+1)}, & se\ n\ é\ par \\ x_{(\frac{n+1}{2})}, & se\ n\ é\ impar \end{cases}$$

- Quando a distribuição é apresentada em dados agrupados em classes, indica-se a classe mediana

$$\tilde{x} = l + \frac{\frac{n}{2} - N_a}{n_i} \text{ x } amp\ onde\ l = limite\ inferior\ da\ classe\ mediana$$

amp = amplitude da classe mediana; n = nº de elementos ;ni = frequência absoluta da classe mediana; Na = frequência absoluta acumulada da classe anterior à classe mediana

Exemplo 3:

Considere-se a tabela com a altura de 5 pessoas definida em cm e ordenada por ordem crescente:

x1
x2
x3
x4
x5
125
134
145
153
167

Neste caso:

A mediana é o valor central, o nº de observações é ímpar pelo que x3, é o valor central, logo:

$$\tilde{x} = x_3 = 145$$

Exemplo 4:

Considere-se agora a tabela com a altura de 6 pessoas definidas em cm e ordenada por ordem crescente

x1	x2
x3	x4
x5	x6
125	134
145	153
167	171

Neste caso:

Como o nº de observações é par, não existe um valor central, assim o cálculo da mediana é dado pela média aritmética dos valores centrais, ou seja:

$$\tilde{x} = \frac{x_3 + x_4}{2} = 149$$

Exemplo 5:

Considere a seguinte tabela:

O número de observações é par. Assim:

28/2 = 14, logo pertence à classe [10, 15[- Classe Mediana

Classes	ni	Ni
[0, 5[	6	6
[5, 10[	5	11
[10, 15[	7	18
[15, 20[	4	22
[20, 25[	2	24
[25, 30[	4	28
Total (**Σ**)	28	

Aplicar a fórmula:

$$\tilde{x} = l + \frac{\frac{n}{2} - N_a}{n_i} \times amp$$

$$\tilde{x} = 10 + \frac{\frac{28}{2} - 11}{7} \times 5 \ \approx 12,1$$

MODA

- É o elemento que num conjunto de dados ocorre com maior frequência

- **Moda (dados não classificados e dados classificados discretos) -** é o valor que ocorre com maior frequência num conjunto de observações
- **Classe Modal (dados classificados quantitativos contínuos) -** numa distribuição de frequência, com intervalos de classe de igual amplitude, a classe é a classe com maior frequência

- Quando existe um único valor ou classe modal, a distribuição diz-se unimodal
- Se existem várias modas a distribuição diz-se plurimodal
- Caso n

- Para calcular a moda quando se tem uma classe modal recorre-se à formula de King:

$$\boxed{moda = l + \frac{n_p}{n_a + n_p} \times amp}$$

l = limite inferior da classe modal; np = frequência absoluta da classe posterior; na = frequência absoluta da classe anterior; amp = amplitude da classe

Exemplo 6:

Considere a seguinte tabela:

Classes	ni	Ni
[0, 5[	6	6
[5, 10[	5	11
[10, 15[	7	18
[15, 20[	4	22
[20, 25[	2	24
[25, 30[	4	28
Total (**Σ**)	28	

A classe modal é com maior frequência [10, 15[- Classe Modal

Aplicar a fórmula:

$$moda = l + \frac{n_p}{n_a + n_p} \times amp$$

$$moda = 10 + \frac{4}{5+4} \times 5 \approx 12,2$$

Exercício 4:

Considere a seguinte tabela:

4. Identificar os seguintes elementos da tabela:

4.1 Frequência simples absoluta, da quinta classe

Resposta: 24

4.2 Frequência absoluta acumulada, da penúltima classe

Resposta: 87

4.3 Limite inferior da sexta classe

Resposta: 3,00

4.4 Limite superior da quarta classe

Resposta: 2,95

4.8 Média

Resposta:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{c} n_i x_i$$

$$= \frac{270,1}{90} = 3,001$$

4.9 Mediana

Resposta:

$$\tilde{x} = l + \frac{\frac{n}{2} - N_a}{n_i} \times amp$$

$$= 2,9895$$

4. 10 Moda

Classes	ni
[2,75; 2,80[	2
[2,80; 2,85[	3
[2,85; 2,90[	10
[2,90; 2,95[	11
[2,95; 3,00[	24
[3,00; 3,05[	14
[3,05; 3,10[	9
[3,10; 3,15[	8
[3,15; 3,20[	[illegible]
[3,20; 3,25[	3
Total	90

4.5 Amplitude das classes

Resposta: 0,05

4.6 Amplitude total

Resposta: 0,5

4.7 Ponto médio da terceira classe

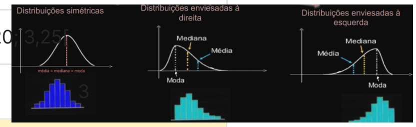
Resposta: 2,875

Resposta:

$$moda = l + \frac{n_p}{n_a + n_p} \text{ x } amp = 2,978$$

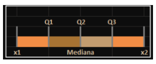
Comparação entre Média, Mediana e Moda

- As distribuições de frequências podem ser simétricas ou não, em relação ao eixo
  - A posição relativa da média, mediana e moda dá informação sobre a forma da curva da distribuição
    - Distribuições simétricas (média = mediana = moda)
    - Distribuições enviesadas à direita (moda < mediana < média)
    - Distribuições enviesadas à esquerda (moda > mediana > média)



Medidas de posição não central

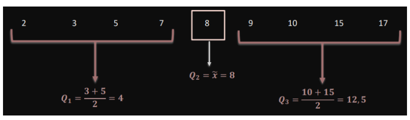
- São pontos que dividem a distribuição em frações de quantidades de dados idênticas. São divididas em:
  - **Quartis -** assim como a mediana divide a distribuição em 2 partes idênticas estes dividem a distribuição em 4 partes idênticas:



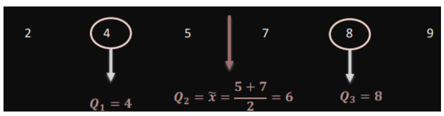
  - **Decis -** dividem a distribuição em 10 partes idênticas
  - **Percentis (ou centis) -** dividem a distribuição em 100 partes idênticas

- Numa distribuição há outras medidas designadas por quartis que dividem o conjunto de dados em quatros partes (contendo em cada uma delas, 25% dos dados)
  - O 1º Quartil representa-se por Q1
  - O 2º Quartil representa-se por Q2 ou Mediana
  - O 3º Quartil representa-se por Q3

Exemplo 1:



Exemplo 2:



- **Assim para dados discretos agrupados ou não:**
  - Para poder determinar os fractis, é necessário calcular primeiro as suas posições absolutas na distribuição, ou as respetivas classes

    $$\boxed{P_{Q_i} = \frac{i(n+1)}{K} = u,d \qquad i = 1,2,3}$$

    K → 4 para quartis, 10 para decis, 100 para percentis

  - Se a posição do quartil i é inteira, então:

    $$\boxed{K_i = x(P_{k_i})}$$

  - Se a posição do quartil i é decimal, então é decomposto em
    - Parte inteira (u)
    - Parte decimal (d)

    $$\boxed{K_i = x(u) + 0,d \times [x(u+1) - x(u)]}$$

Exemplo 3:

Considere os seguintes dados e calcule os quartis

Variável	Frequência (ni)	Frequência acumulada (Ni)
0	2	2
1	6	8
2	9	17
3	13	30
4	5	35
Total	35	

1º Quartil:

$$PQ_1 = \frac{n+1}{4} = \frac{35+1}{4} = 9;\ Ni = 9 \Rightarrow Q_1 = 2$$

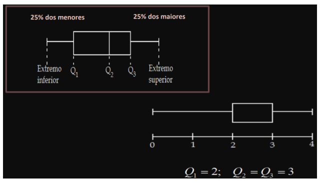
2º Quartil:

$$PQ_2 = 2 \times \frac{n+1}{4} = 4 \times \frac{35+1}{4} = 18;\ Ni = 18 \Rightarrow Q_2 = 3$$

3º Quartil:

$$PQ_3 = 3 \times \frac{n+1}{4} = 3 \times \frac{35+1}{4} = 27;\ Q_3 = 3$$

Representação gráfica do resultado do Exemplo 3:



- **Assim para dados discretos agrupados em classes:**
  - Para determinar os fractis, é necessário calcular primeiro as suas posições absolutas na distribuição, ou as respetivas classes

$$\boxed{P_{Q_i} = \frac{i \times n}{K}}$$

K → 4 para quartis, 10 para decis, 100 para percentis

$$\boxed{Q_j = l_i + \frac{j \times \frac{n}{k} - N_a}{n_i} \times amp}$$

Exemplo 4:

Considere os seguintes dados:

Variável
Frequência (ni)
Frequência acumulada (Ni)
[54; 58[
9
9
[58; 62[
11
20
[62; 66[
8
28
[66; 70[
5
33
Total
33

Calcula os quartis

$$P_{Q_1} = \frac{n}{4} = \frac{33}{4} = 8,25; \quad Q_1 \Rightarrow 1^{a} Classe$$

$$Q_1 = 54 + \frac{1 \times \frac{33}{4} - 0}{9} \times 4 = 57.67$$

$$P_{Q_2} = 2 \times \frac{n}{4} = 2 \times \frac{33}{4} = 16,5; \quad Q_2 \Rightarrow 2^{a} classe$$

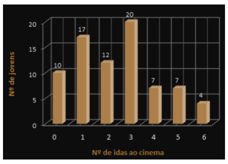
$$Q_2 = 58 + \frac{2 \times \frac{33}{4} - 9}{11} \times 4 = 60,73$$

$$P_{Q_3} = 3 \times \frac{n}{4} = 3 \times \frac{33}{4} = 24,75; \quad Q_3 \Rightarrow 3^{a} classe$$

$$Q_3 = 62 + \frac{3 \times \frac{33}{4} - 20}{8} \times 4 = 64,38$$

Exercício 1:

Fez-se um inquérito a um grupo de jovens sobre as idas ao cinema no último mês e os resultados estão sintetizados no seguinte gráfico:



1.1 Calcula os quartis

$$P_{Q_1} = \frac{n}{4} = \frac{77}{4} = 19,25; \quad Q_1 = 1$$

$$P_{Q_2} = 2 \times \frac{n}{4} = 2 \times \frac{77}{4} = 38,5; \quad Q_2 = 2$$

$$P_{Q_3} = 3 \times \frac{n}{4} = 3 \times \frac{77}{4} = 57,75; \quad Q_3 = 3$$

Exercício 3:

De um pomar de macieiras jovens selecionaram-se macieiras e contou-se o nº de maçãs de cada uma

Os dados obtidos foram as seguintes:

7, 6, 13, 5, 6, 5, 9, 9 ,15, 9, 8, 8, 12, 13, 5

Determine:

3.1 P30

5, 5, 5, 6, 6, 7, 8, 8, 9, 9, 9, 12, 13, 13, 15

$$P_{30} = \frac{30(15+1)}{100} = 4,8; \quad x(4) + 0,8[x(5) - x(4)] = 6 + 0,8(6-6) = 6, \quad logo\ P_{30} = 6$$

3.2 P50

$$P_{50} = \frac{50(15+1)}{100} = 8; \quad x(8) = 8, \quad logo P_{50} = 8$$

3.3 P80

$$P_{80} = \frac{80(15+1)}{100} = 12,8; \quad x(12) + 0,8[x(13) - x(12)] = 12 + 0,8(1) = 12,8, \quad logo P_{80} = 12,8$$

Exercício 4:

Considere os resultados de um teste de Estatística, assim agrupados:

4.1 Calcula:

a) os quartis

$$P_{Q_1} = \frac{1 \times 50}{4} = 12,5; \quad Q_1 = 65 + \frac{1 \times \frac{50}{4} - 11}{12} \times 5 = 65,625$$

$$P_{Q_2} = \frac{2 \times 50}{4} = 25; \quad Q_2 = 70 + \frac{2 \times \frac{50}{4} - 23}{15} \times 5 = 70,66(6)$$

Classes	ni
[55; 60[	3
[60; 65 [	8
[65; 70[	12
[70; 75[	15
[75; 80[	6
[80; 85[	4
[85; 90[	2
Total	50

$$P_{Q_3} = \frac{3 \text{ x } 50}{4} = 37,5; \quad Q_3 = 75 + \frac{3 \text{ x } \frac{50}{4} - 23}{15} \text{ x } 5 = 74,83(3)$$

b) 8º decil

$$P_{D_8} = \frac{8 \text{ x } 50}{100} = 40, \quad logo\ D_8 \in [75,80[$$

c) 23ºcentil

$$P_{C_{23}} = \frac{23 \text{ x } 50}{100} = 11,5, \quad logo\ C_{23} \in [65,70[$$

Medidas de variabilidade

- Estas medidas servem para medir a variabilidade (dispersão) num conjunto de dados. E são:

Amplitude:

- **Amplitude total:** é a diferença entre o valor máximo e o valor mínimo do conjunto de observações
  - Depende apenas das observações extremas
  - Depende do nº de observações: quanto maior for o nº de observações maior tende a ser a amplitude total
- **Amplitude Inter-Quartis:** é a diferença entre o 3º quartil e o 1º quartil (Q3-Q1)

$$\boxed{Amplitude_{inter-quartil} = Q_3 - Q_1}$$

50% dos elementos do meio da amostra, estão contidos num intervalo com esta amplitude.

Esta medida é não negativa e será tanto maior for a variabilidade nos dados.

Exemplo 3:

Variável	Frequência (ni)
0	2
1	6
2	9
3	13
4	5
Total	35

Amplitude total = 4 - 0 = 4

1º Quartil: Q1 = 2

3º Quartil: Q3 = 3

Amplitude inter-quartil = Q3 - Q4 = 3 - 2 = 1

Desvio Absoluto Médio

- É a média aritmética dos valores absolutos dos desvios de cada um dos dados em relação à média

- Em dados simples tem-se:

$$\boxed{\delta_{\overline{x}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|}$$

- Em dados organizados em tabelas de frequência tem-se:

$$\boxed{\delta_{\overline{x}} = \frac{1}{n}\sum_{i=1}^{k}|x_i - \overline{x}| \times n_i}$$

Variância

- Obtém-se somando os quadrados dos desvios das observações da amostra, relativamente à sua média, e dividindo pelo nº de observações da amostra

- Em dados simples tem-se:

$$\boxed{S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

- Em dados organizados em tabelas de frequência tem-se:

$$\boxed{S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \times n_i}$$

Desvio Padrão:

- É a raiz quadrada da variância. Dá informação sobre a variabilidade dos dados em relação à média

- Em dados simples tem-se:

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

- Em dados organizados em tabelas de frequência tem-se:

$$\boxed{s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \times n_i}}$$

Exercício 7:

O número de golos marcados por uma equipa nas primeiras 20 jornadas do campeonato foi registado na tabela seguinte

Nº de golos	0	1	2	3	4	5
Nº de jogos	2	5	4	5	3	1

7. Calcula o desvio-padrão

xi	ni
xini	(xi - ni)^2 x ni
0	2
0	10.125
1	5
5	7.8125
2	4
8	0.25
3	5
15	2.8425
4	3
12	9.1825
5	1
5	7.5625
Total	20
45	37.75

$$\overline{x} = \sum_{i=1}^{n}(x_i n_i) = 2,25$$

$$S^2 = \frac{\sum^n (x_i - n_i) \text{ x } n_i}{n} = \frac{37,75}{20} = 1.8875$$

$$s = 1.3739$$

Coeficiente de Variação de Pearson

- O coeficiente de variação de Pearson, é dado pela relação, em termos percentuais, entre o desvio padrão e a média de distribuição

$$\boxed{C_{VP} = \frac{S}{\overline{x}} \text{ x } 100}$$

Exemplo 1:

Observe-se a tabela seguinte. Qual das 2 variáveis: altura ou peso apresenta menor variabilidade, sabendo valores da média e do desvio padrão?

	Média	Desvio-padrão
Altura	175 cm	5 cm
Peso	68 Kg	2 Kg

Calcule-se o coeficiente de variação de Pearson, para cada uma das variáveis:

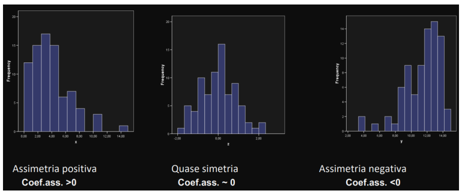
$$C_{VP} altura = \frac{s}{\overline{x}} \text{ x } 100 = 2,86\%$$

$$C_{VP} peso = \frac{s}{\overline{x}} \text{ x } 100 = 2,94\%$$

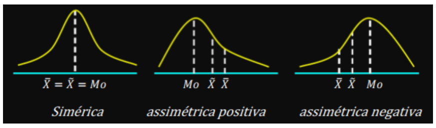
Conclusão: Neste caso, a altura apresenta menor dispersão que o peso

Medidas de assimetria

- Uma distribuição possui assimetria positiva quando existe uma concentração de valores na zona de valores mais reduzidos da amostra



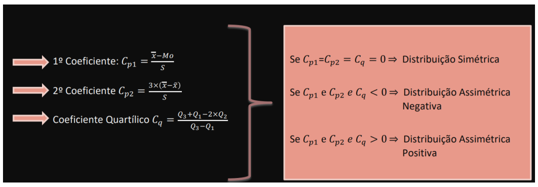
- Também se pode medir o grau de assimetria de uma distribuição comparam-se as 3 medidas de tendência central: média, moda e mediana



Assim:

- Sendo a distribuição simétrica, a média e a moda coincidem
- Sendo a distribuição assimétrica:
  - à esquerda ou negativa, a média é menor que a moda
  - à direita ou positiva, a média é maior que a moda

- Pode medir-se o grau de assimetria de uma distribuição recorrendo aos coeficientes de assimetria de Pearson:

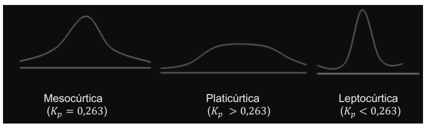


Medidas de Achatamento (Kurtosis)

- Esta é uma medida que descreve o grau de achatamento ou afunilamento da curva da distribuição. O seu valor diz-nos se a curva tende a ser muito afunilada/pontiaguda (e.g., com um pico), com uma elevada proporção dos dados aglomerados junto do centro, ou achatada, com os dados espalhando-se ao longo de uma grande amplitude da variável.

- Coeficiente percentil de Kurtosis (Kp):

$$\boxed{K_p = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}$$



Distribuições Bidimensionais