


Tópico de Aprendizagem – **Organização e tratamento de dados**

Tarefa nº6 – **Simetria e enviesamento**

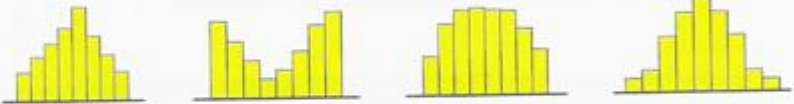
**NOTA**  
 Dado um histograma, é fácil obter a posição da mediana, pois esta está numa posição tal, que passando uma linha vertical por esse ponto, o histograma fica dividido em duas partes com áreas iguais, como se representa na figura seguinte:




Nas representações gráficas das distribuições podemos, por vezes, encontrar semelhanças ou diferenças relativamente à sua forma. Essa forma também nos pode fazer optar por uma medida de localização ou dispersão mais representativa.

**Histogramas**


Nas **distribuições simétricas** a média e a mediana assumem valores muito idênticos ou iguais, podendo optar-se por qualquer uma delas para representar os dados.



Nas **distribuições enviesadas**, a mediana é a que representa mais correctamente os dados, pois como vimos anteriormente a média é pouco resistente aos outliers.

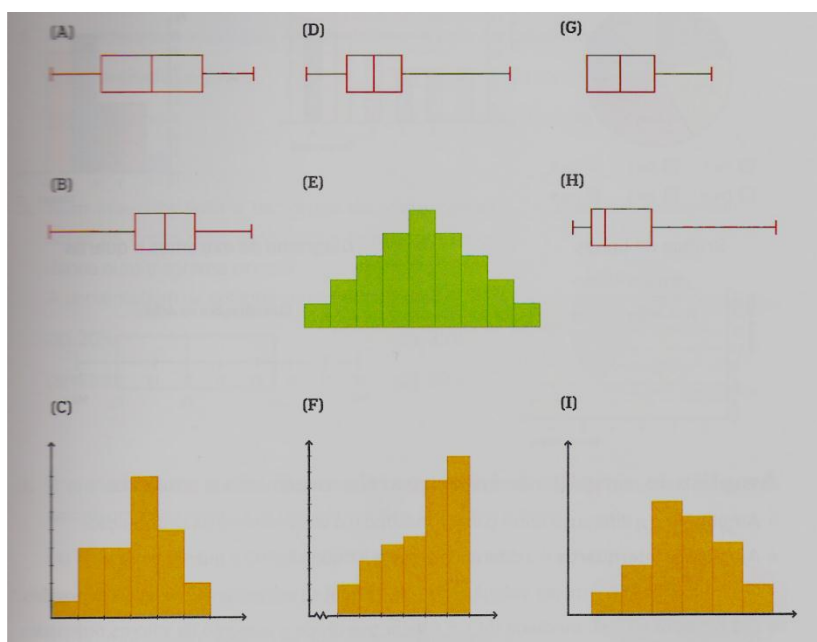


**Diagramas de extremos e quartis**



**Exercício:** Para representar algumas distribuições estatísticas, elaboraram-se histogramas e diagramas de extremos e quartis. Classifica-os relativamente à sua simetria, escolhendo um dos seguintes rótulos:

«ligeiramente enviesada à esquerda», «simétrica», «enviesada à esquerda», «ligeiramente enviesada à direita» e «enviesada à direita».



## Medidas para resumir a informação contida numa amostra

Para descrever o comportamento dos dados de uma amostra utilizam-se as medidas de localização e as medidas de dispersão.

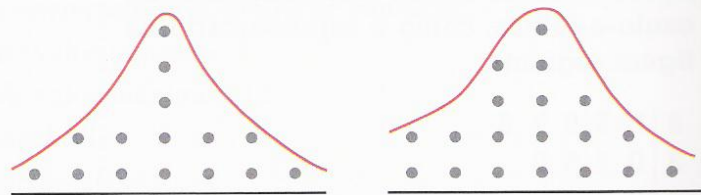
Uma questão que deve ser colocada consiste em decidir quais as medidas mais adequadas para cada caso concreto.

Para responder à questão sugere-se a seguinte metodologia:

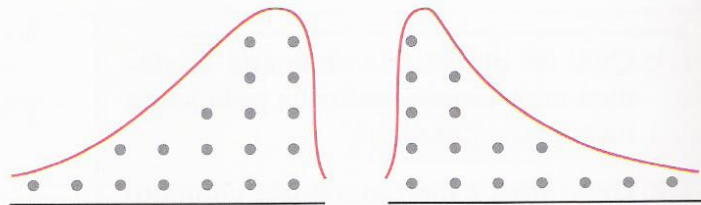
1.º Fazer uma representação gráfica dos dados.

2.º Observando o gráfico:

a) Se a distribuição dos dados se apresentar aproximadamente simétrica, então a média e a mediana coincidem.



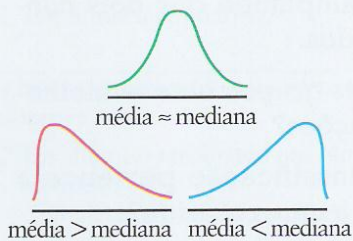
b) Se a distribuição apresentar um enviesamento, deve utilizar-se a mediana e a amplitude interquartis, já que a mediana é mais resistente que a média, pois não é tão sensível aos dados.



c) Se se verificar a existência de alguns (ou algum) valores (ou valor) extremos muito maiores ou muito menores que os restantes (*outliers*), calculamos a média desprezando esses *outliers* e acrescentamos ao estudo uma observação sobre o assunto.

### OBSERVAÇÃO

A utilização da média e da mediana depende do contexto em que estão a ser utilizadas. Como a média é influenciada quer por valores muito grandes, quer por valores muito pequenos, se a distribuição dos dados for enviesada para a direita (alguns valores grandes como *outliers*), a média tende a ser maior que a mediana; se for aproximadamente simétrica, a média aproxima-se da mediana e se for enviesada para a esquerda (alguns valores pequenos como *outliers*), a média tende a ser inferior à mediana. De um modo geral, temos que:



### Exercícios:

1. O João estava a usar uma folha de cálculo para um estudo sobre o tempo, em minutos, de atraso do autocarro que todos os dias utiliza para ir para a escola.

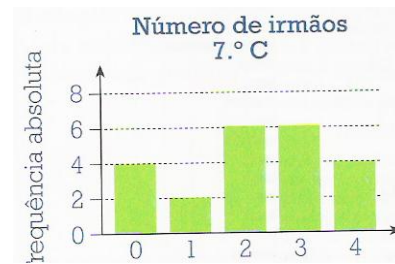
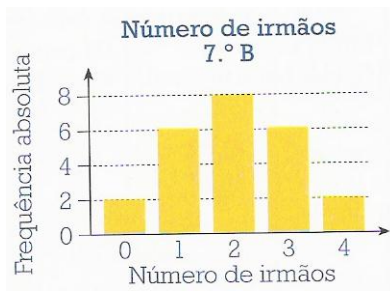
Os valores representados variavam entre 0 e 15 minutos.

Ao digitar um dado enganou-se e escreveu 77 em lugar de 7.

Qual o efeito deste engano sobre as medidas de localização (média, moda e mediana) e dispersão (amplitude e amplitude interquartis) estudadas?

**2.** A cada um dos alunos das turmas A, B e C do 7º ano de uma escola perguntou-se quantos irmãos tinha.

Com os dados recolhidos construiu-se, para cada turma, um gráfico de barras como se mostra nas figuras seguintes:



**2.1.** A partir das representações gráficas diz em qual das turmas se espera maior média? E maior mediana?

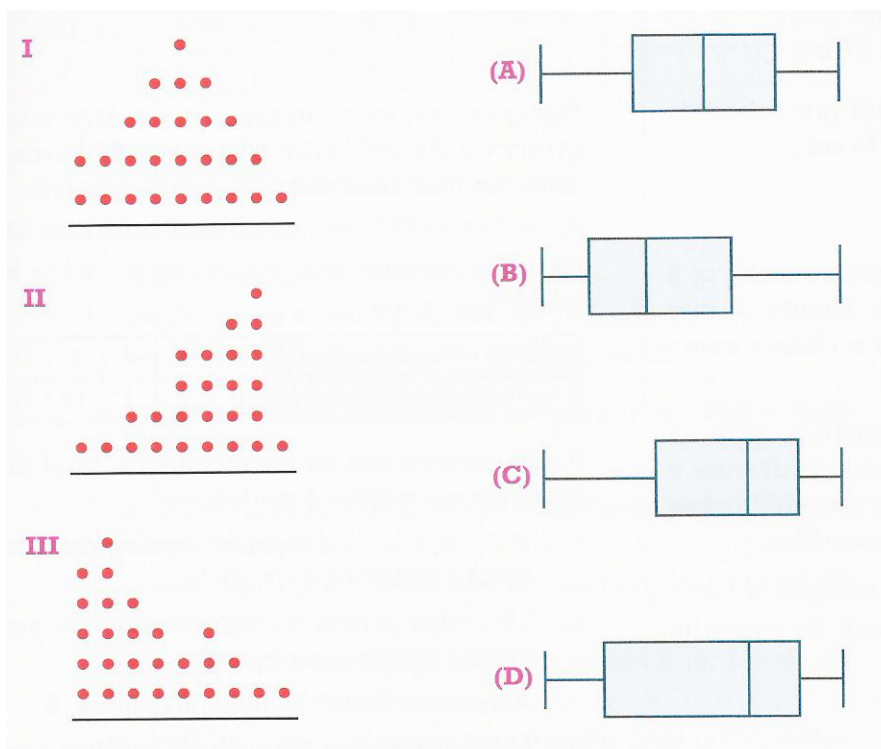
E em qual das turmas se espera que a mediana esteja mais afastada da média? Explica como obtiveste as tuas respostas.

**2.2.** Verifica a tua resposta calculando a média (com uma casa decimal) e a mediana para cada uma das turmas.

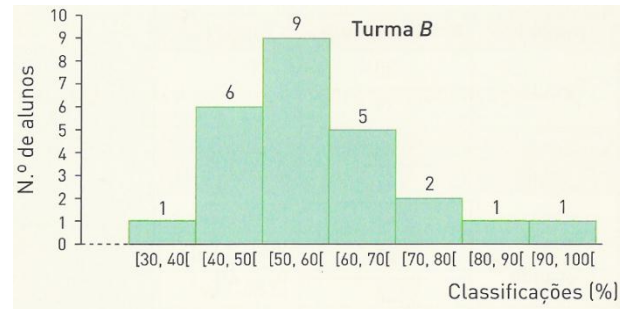
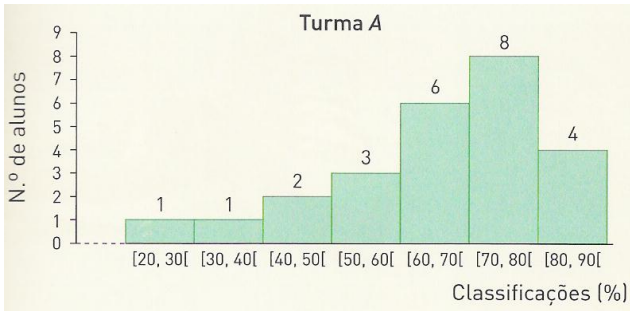
**2.3.** Calcula para cada uma das turmas a amplitude interquartis.

**3.** Considera os três digramas de pontos e os quatro diagramas de extremos e quartis apresentados a seguir. Faz corresponder a cada diagrama de pontos um diagrama de extremos e quartis.

Explica, para cada resposta, a tua opção.



4. A seguir são apresentados gráficos relativos às classificações, em percentagem, obtidas pelos das turmas A e B de uma escola.



4.1. Em qual das turmas se verifica que mais de metade dos alunos obteve classificações inferiores a 60%?

4.2. Qual é a percentagem de negativas (classificações inferiores a 50%) em cada turma?

4.3. Atendendo à assimetria (enviesamento) que é possível observar em cada gráfico, para cada uma das turmas compara a média e a mediana, indicando qual delas se espera que seja menor?

4.4. Relativamente à turma B, sabe-se que os resultados são os seguintes:

34	40	41	42	45
45	47	50	52	55
54	54	55	56	58
58	59	62	65	67
68	74	78	86	94

Confirma o resultado esperado para esta turma na alínea anterior, indicando os valores da média e da mediana.

5. Na figura estão representados três diagramas de extremos e quartis que representam três distribuições.

Atendendo à simetria, diz qual é o diagrama que pode traduzir

cada uma das seguintes situações:

- a)  $\bar{x} = 12$  e Med = 10
- b)  $\bar{x} = 26$  e Med = 29
- c)  $\bar{x} = 4$  e Med = 4

