

| Estatística | População | Amostra | Censo ou Recenseamento | Sondagem |
|--|--------------------------------------|---|---|---|
| É uma ciência (ramo da matemática) que estuda uma ou várias características ou propriedades de uma população tendo por base a recolha, classificação, apresentação e interpretação dos dados sobre o fenómeno em estudo. | É o conjunto dos elementos em estudo | É uma parte da população em que incide o estudo Dimensão de uma amostra Número de elementos da amostra | É um estudo estatístico realizado sobre a totalidade da população | É um estudo estatístico realizado a partir de uma amostra |

Variável: característica comum que pode assumir valores de modalidades diferentes, de indivíduo para indivíduo

Variável qualitativa- Representa a informação que identifica alguma qualidade, categoria ou característica, não susceptível de medida, mas de classificação, assumindo várias modalidades. (não podem ser expressas numericamente)
Por ex; a cor da pele, cor dos olhos, marca de refrigerante, marca de automóvel, preferência musical entre outras)

Variável quantitativa- aquela a que é possível atribuir uma medida e que se apresenta com diferentes intensidades ou valores. Pode ser discreta ou contínua.

Variável quantitativa discreta- Variável cujos valores se podem ordenar de tal modo que entre dois valores consecutivos não possa existir um valor intermédio. (toma valores isolados)

Ex.: Número de irmãos de uma pessoa.

Variável quantitativa contínua- variável que pode tomar qualquer valor de um intervalo.

Ex.: Altura de uma pessoa.

A **frequência absoluta** de um acontecimento é o nº de vezes que esse acontecimento se repete

A **frequência relativa** de um acontecimento é o quociente da frequência absoluta desse acontecimento pelo nº. total de elementos em estudo

(Quando o total não dá 100 na frequência relativa: **retira-se um ao mais pequeno**, por exemplo:

0,30769 = 0,31

0,07692 = 0,08 <- retira-se aqui, faz-se batota)

Nos gráficos de barras, os rectângulos que as constituem variam apenas numa das dimensões de acordo com a frequência absoluta ou relativa. As barras devem estar separadas por espaços iguais.

Para descobrir o número de classes:

- $n = 24$ (efectivo da população/total)

- $2^k \geq n$ (sendo K o numero de classes): $2^1 = 2$; $2^2 = 4$; $2^3 = 8$; $2^4 = 16$; $2^5 = 32$

- $K = 5$ classes

- $x_{max} - x_{min} = 102 - 60 = 42$

- $\frac{x_{max} - x_{min}}{K} = \frac{42}{5} = 8,4$ <- amplitude da classe

8,5 <- amplitude da classe definitiva

Medidas de localização

Média (\bar{x}): número que se obtém dividindo a soma dos valores de todos os dados pelo número de dados

Moda (Mo): valor que ocorre com maior frequência. (se os dados forem apresentados em classes considera-se a **classe modal**)

Quando existem **duas modas:** chama-se **bimodal**.

Quando não há moda (nenhum dos dados apresenta maior frequência que os outros), chama-se **amodal**.

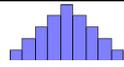
Mediana (\tilde{x}) é o valor central da amostra (dados organizados por ordem crescente ou decrescente)

Se o número de dados é ímpar, a mediana é o valor que ocupa a posição central. ($k = \frac{n+1}{2}$ e $\tilde{x} = x_k$)

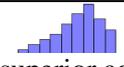
Se o número de dados é par, a mediana é a **média aritmética** dos dois valores centrais ($k = \frac{n}{2}$ e $\tilde{x} = \frac{x_k + x_{k+1}}{2}$)

Considerações gerais sobre a média, a moda e a mediana.

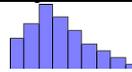
Se o valor da média for igual ao da mediana e este for igual ao da moda, a distribuição diz-se simétrica:



Se o valor da média for inferior ao da mediana e este for inferior ao da moda, a distribuição diz-se assimétrica negativa:



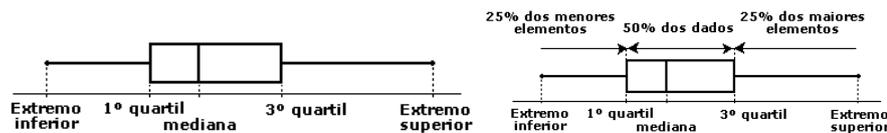
Se o valor da média for superior ao da mediana e este for superior ao da moda, a distribuição diz-se assimétrica positiva:



Quartis - Os **quartis** são os valores que dividem a distribuição em quatro partes iguais.

- **1º Quartil (Q₁):** É o valor que divide a sequência em duas partes, de tal modo a que pelo menos 25% das observações sejam inferiores ou iguais a esse valor e 75% das observações sejam iguais ou superiores a esse valor;
- **2º Quartil (Q₂ ou \tilde{x})-** Coincide com a mediana;
- **3º Quartil (Q₃)-** É o valor que divide a sequência em duas partes, de tal modo a que pelo menos 75% das observações sejam inferiores ou iguais a esse valor e 25% das observações sejam iguais ou superiores a esse valor.

Diagrama de extremos e quartis e Caixa dos bigodes: É um tipo de representação gráfica, em que se realçam algumas características da amostra. O conjunto dos valores da amostra compreendidos entre o 1º e o 3º QUARTIS, que vamos representar por Q₁ e Q₃ é representado por um rectângulo (caixa) com a MEDIANA indicada por uma barra. A largura do rectângulo não dá qualquer informação, pelo que pode ser qualquer. Consideram-se seguidamente duas linhas que unem os meios dos lados do rectângulo com os extremos da amostra. Para obter esta representação, começa por se recolher da amostra, informação sobre 5 números, que são: os 2 extremos (mínimo e máximo), a mediana e o 1º e 3º quartis. A representação do diagrama de extremos e quartis tem o seguinte aspecto:



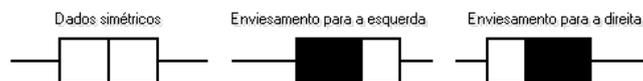
O **extremo inferior** é o **mínimo da amostra**, enquanto que o **extremo superior** é o **máximo da amostra**.

Como é que se pode reconhecer a simetria ou o enviesamento dos dados, a partir do Diagrama de extremos e quartis?

Existem fundamentalmente 3 características, que nos dão ideia da simetria ou enviesamento e da sua maior ou menor concentração

- distância entre a linha indicadora da mediana e os lados do retângulo
- comprimento das linhas que saem dos lados dos retângulos
- comprimento da caixa

Apresentamos seguidamente 3 exemplos de **diagramas de extremos e quartis**, correspondentes a tipos diferentes de distribuição de dados.



O que é um **outlier**? É um valor que não esteja compreendido no intervalo [mínimo, máximo]

Medidas de dispersão - permitem medir a variabilidade dos dados. A **amplitude** e o **desvio padrão**

Amplitude é a diferença entre o maior e o menor valor. Se os dados estiverem agrupados em classes, a amplitude é a diferença entre o limite superior da última classe e o inferior da primeira.

Amplitude inter-quartil-Diferença entre o 3º quartil e o 1º quartil.

Desvio padrão-medida que permite saber como é que se afasta cada um dos dados em relação à média. $\sigma = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}}$

Variância: medida de dispersão que mostra a distância que os valores estão da média. $\delta = \sqrt{\text{variância}} = \sqrt{\delta^2}$