**Matemática Aplicada às Ciências Sociais (M.A.C.S.)**

**2 – Estatística**

2.1 – Dados, tabelas e gráficos

2.1.1 – Noções básicas

2.1.1.1 - Variável

A **variável** ou **caráter estatístico** (**xi**) é uma característica ou propriedade da população em estudo, à qual se pode atribuir um nº ou uma categoria. As variáveis estatísticas podem ser:

* **Qualitativas** – são atributos que, por se relacionarem com qualidades, não se podem traduzir numericamente;
* **Quantitativas** – são atributos que se podem traduzir numericamente, através de uma contagem ou de uma medição. Estas podem ainda ser:
	+ **Discretas** – a variável só toma valores correspondentes a pontos isolados;
	+ **Contínuas** – a variável só pode tomar qualquer valor de um intervalo.

2.1.1.2 – População

A **população** ou **universo estatístico** é o conjunto de elementos com pelo menos uma característica comum que se pretende analisar.
Chama-se **unidade estatística** a cada um dos elementos da população, e **efetivo ou dimensão da população** ao nº de elementos da população.
Faz-se um **censo** quando se quer realizar um estudo estatístico que incida sobre todos os elementos da população.

2.1.1.3 – Amostra

A **amostra** é o subconjunto finito da população. **Amostragem** é todo o processo de recolha de uma parte, geralmente pequena, dos elementos que constituem um dado conjunto. Da análise dessa parte pretende obter-se informações para todo o conjunto. Nesse processo, é importante distinguir entre população e **população inquirida** – aquela sobre a qual é efetivamente feita a amostragem.
Faz-se uma **sondagem** quando se quer realizar um estudo estatístico em que se utilize apenas uma amostra da população.

2.1.2 – Construção de tabelas de frequência

2.1.2.1 – Frequência absoluta

A **frequência absoluta** (**fi**) é o nº de vezes que o valor da variável se repete ou é observado. A **frequência absoluta acumulada** (**Fi**) é igual à soma das frequências absolutas.

2.1.2.2 – Frequência relativa

A **frequência relativa** (**fri**) é igual ao quociente entre a frequência absoluta e o nº total de observações. A **frequência relativa acumulada** (**Fri**) é igual à soma das frequências relativas.

2.1.2.3 – Regra de Sturges

Para organizar uma amostra de dados contínuos, de dimensão n, pode considerar-se para o nº de classes o valor k, onde k é o menor nº inteiro, tal que 2k ≥ n.

2.1.3 – Representações gráficas

2.1.3.1 – Diagramas de caule-e-folhas

1 – Traça-se uma linha vertical e regista-se do lado esquerdo o(s) dígito(s) da ordem de maior grandeza;

2 – Regista-se do lado direito da linha o dígito do valor da variável em estudo correspondente às unidades (podem ou não aparecer por ordem crescente).

2.1.3.2 – Gráficos circulares

São gráficos construídos a partir da divisão de um círculo em setores circulares cuja área é proporcional à frequência de cada um dos valores da variável. A amplitude do ângulo de cada setor é igual a $\frac{fi}{n}$ x 360° ou a fri x 360°.

2.1.3.3 – Pictogramas

São gráficos onde se utilizam figuras o símbolos alusivos à variável em estudo.

2.1.3.4 – Gráficos de barras

São gráficos formados por um conjunto de barras cuja altura é proporcional às frequências de cada um dos valores da variável. Num dos eixos, marcam-se as frequências, e no outro, os valores da variável.
As variáveis devem ser qualitativas ou quantitativas discretas.

2.1.3.5 – Gráficos de linhas

São usados para representar informação que varia ao longo do tempo.

2.1.3.6 – Histogramas

São gráficos formados por um conjunto de retângulos adjacentes, tendo por base o intervalo da classe e uma área proporcional às respetivas frequências (absolutas e relativas) de cada classe. Estes gráficos utilizam-se sempre que os dados estão agrupados em classes na forma de intervalos (variáveis contínuas).
Para se calcularem o nº de classes de uma variável utiliza-se a Regra de Sturges, e para se calcular o intervalo das classes, faz-se a amplitude do conjunto dos dados.

20

18

16

14

12

10

8

6

4

2

2.1.3.7 – Polígonos de frequências

O polígono de frequências resulta da união sucessiva dos pontos médios dos lados superiores dos diferentes retângulos de um histograma através de segmentos de reta. Para que o polígono fique concluído, deverá ainda unir-se o extremo esquerdo do polígono com o ponto médio da classe anterior à primeira cuja frequência é nula, procedendo-se analogamente para o extremo direito do polígono.

20

18

16

14

12

10

8

6

4

2

2.2 – Cálculo de estatísticas

2.2.1 – Medidas de localização

$$\sum\_{i=1}^{n}x\_{i}$$

2.2.1.1 – Média

Casio: Stat – Calc – 1 Var (dados simples) / 2 Var (dados agrupados) - $\overline{x}$

$\overline{x}$ =\_ \_ n

* Para dados simples ou não classificados:
* Para dados agrupados ou classificados:

$$\sum\_{i=1}^{n}f\_{i}×x\_{i}$$

$\overline{x}$ =\_\_ \_ n

2.2.1.2 – Moda

A moda (**Mo**) é o valor da variável ao qual corresponde uma maior frequência. A amostra é:

* **Bimodal** se houver dois valores com a maior frequência;
* **Plurimodal** se existirem vários valores com a frequência mais alta;
* **Amodal** se os valores da variável têm todos a mesma frequência.

Quando os dados estão agrupados em classes, sob a forma de intervalos, a **classe modal** é a classe à qual corresponde a maior frequência.

2.2.1.3 – Mediana

Casio: Stat – Calc – 1 Var/ 2 Var - Med

A mediana(**Md ou** $\tilde{x}$)é preferível à média quando se está interessado em conhecer o ponto médio da distribuição dos dados, ou seja, o valor que divide os dados em duas partes iguais (**Q2**).

2.2.1.4 – Quartis

* O **1º quartil** (**Q1**) é o valor que divide a amostra (ordenada por ordem crescente) em duas partes, de tal modo que 25% das observações sejam inferiores ou iguais a esse valor. Quando n é:

Casio: Stat – Calc – 1 Var/ 2 Var – Q1

* Par – Q1 = xk, k = $\frac{n+ 2}{4}$;
* Ímpar – Q1 = xk, k = $\frac{n+ 1}{4}$.
* O **2º quartil** (**Q2**) é o valor que divide a amostra (ordenada) em duas partes, de tal modo que 50% das observações sejam inferiores ou iguais a esse valor. O 2º quartil corresponde assim à mediana;

Casio: Stat – Calc – 1 Var/ 2 Var - Med

* Par – Q2 = $\tilde{x}$ = $\frac{x\_{k}+x\_{k}+1}{2}$, k = $\frac{n}{2}$;
* Ímpar – Q2 = $\tilde{x}$ = xk, k = $\frac{n+1}{2}$.
* O **3º quartil** (**Q3**) é o valor que divide a amostra (ordenada por ordem crescente) em duas partes, de tal modo que 75% das observações sejam inferiores ou iguais a esse valor.

Casio: Stat – Calc – 1 Var/ 2 Var – Q3

* Par – Q3 = xk, k = $\frac{3n+ 2}{4}$;
* Ímpar – Q3 = xk, k = 3 x $\frac{n+ 1}{4}$.

Diagrama de extremos e quartis

1 – Desenha-se uma linha vertical ou horizontal onde se marcam alguns dos valores da variável, nomeadamente os extremos e quartis;

2 – Constrói-se um retângulo correspondente ao intervalo entre o 1º e o 3º quartil;

3 – Constrói-se um segmento de reta entre o extremo inferior e o 1º quartil e outro segmento de reta entre o 3º quartil e o extremo superior.

A partir do diagrama de extremos e quartis, podem-se tirar conclusões quanto à dispersão e concentração dos conjuntos de dados:



* **Dados simétricos** – os dados estão distribuídos de forma simétrica;
* **Enviesamento para a esquerda** – os dados estão mais dispersos à esquerda de Q2 e mais concentrados à direita de Q2;
* **Enviesamento para a direita** – os dados estão mais dispersos à direita de Q2 e mais concentrados à esquerda de Q2.

2.2.2 – Medidas de dispersão

2.2.2.1 – Amplitude

Representa-se por **R** (**range**) de um conjunto de dados à diferença entre o valor máximo e o valor mínimo desse conjunto.

2.2.2.2 – Amplitude interquartil

A **amplitude interquartil** (**Aq**) é a diferença entre o 3º quartil (Q3) e o 1º quartil (Q1), isto é, Q3 – Q1.

2.2.2.3 – Desvio médio

* Para dados simples:
* Para dados agrupados:

Na qual k representa o nº de valores diferentes que surgem na amostra.
Se os dados estão agrupados em classes, xi e fi são, respetivamente, o ponto médio e a frequência absoluta da classe i; k é o número de classes.

2.2.2.4 – Desvio padrão

Casio: Stat – Calc – 1 Var/ 2 Var – xσn (σ)/ xσn-1(s)

O **desvio padrão** representa-se por **s** quando os dados da variável representam uma amostra, ou por **σ** quando os dados da variável representam a população, e é igual à raiz quadrada da **variância** (**s2** ou **σ2**), ou seja, σ = √σ2. Quando os dados da variável representam:

* Uma amostra, a variância é igual a
* A população, a variância é igual a

2.3 – Dados bivariados

2.3.1 – Análise gráfica de dados bivariados

**Dados bidimensionais** ou **bivariados** são dados obtidos de pares de variáveis.
**Diagrama de dispersão** ou **gráfico de correlação** é um gráfico em que a cada ponto correspondem duas coordenadas, que são os valores das duas variáveis em estudo (xi e yi). Este gráfico permite analisar de que forma se relacionam duas variáveis.

2.3.1.1 – Correlação linear
Diz-se que existe **correlação** entre duas variáveis quando a variação de uma delas implica uma alteração na outra.
A correlação diz-se **linear** se a nuvem de pontos se ajustar em torno de uma linha reta, que é a **reta de regressão**.
A correlação linear pode ser:

* **Linear positiva** – quando a variável x aumenta, a variável y também aumenta; diz-se que há uma associação positiva entre as variáveis se aos maiores valores de uma correspondem, de uma maneira geral, os maiores valores da outra;
* **Nula** – não há associação entre as variáveis quando a nuvem de pontos se encontra bastante dispersa;
* **Linear negativa** – quando a variável x aumenta, a variável y diminui; diz-se que há uma associação negativa entre as variáveis se aos maiores valores de uma correspondem, de um modo geral, os menores valores de outra, e vice-versa.

2.3.1.2 – Coeficiente de correlação

O **coeficiente de correlação linear**, ou **coeficiente de Pearson**, mede o grau de associação linear entre duas variáveis. Representa-se por **r** e varia entre -1 e 1.

* Se r=**1**, a correlação diz-se **total** (ou **perfeita**) **positiva**;

Casio: Stat – Grph – Gph1 – X - r

* Se r=**0**, a correlação diz-se **nula**, isto é, não há correlação linear;
* Se r=**–1**, a correlação diz-se **total** (ou **perfeita**) **negativa**.

2.3.2 – Tabelas de contingência

**Tabela de contingência** é um processo de organizar informação correspondente a dados bivariados qualitativos, quantitativos ou ambos.