



10.º ANO | MATEMÁTICA A

RESUMOS ESTATÍSTICA

ANTÓNIO LEITE | 2025

ESTATÍSTICA

População e amostra

- População é um conjunto de elementos, designados por unidades estatísticas, que podem ser pessoas, animais, instituições, objetos físicos, resultados experimentais, com uma ou mais características em comum, que se pretendem analisar.
- Amostra é um subconjunto da população formado por unidades estatísticas que se observam com o objetivo de obter informações para estudar a(s) característica(s) pretendida(s).
- A dimensão da amostra é o número de unidades estatísticas a ela pertencentes.
- Censo: são recolhidos dados sobre todos os elementos da população.
- Sondagem: são recolhidos dados sobre uma parte da população, a amostra, e generalizam-se conclusões para a restante parte da população.

Fases de um procedimento estatístico

- Produção ou aquisição de dados;
 - Organização e representação de dados;
 - Interpretação tendo por base as representações obtidas.
-
- Os métodos para a seleção de amostras, devem ser no sentido de que estas sejam representativas das populações subjacentes, de modo a evitar amostras enviesadas cujo estudo deve inferir conclusões erradas para as populações.
Exemplos de amostras enviesadas: amostras por conveniência e amostras de resposta voluntária.
 - Diferentes tipos de amostragem: aleatória simples, aleatória sistemática e estratificada ou proporcional.

Dados univariados. Dados quantitativos discretos ou contínuos

- Variável discreta: assume um número finito ou infinito numerável de valores distintos e a observação assume a forma de uma contagem.
- Variável contínua: assume qualquer valor num intervalo de número reais e a observação assume a forma de uma medição.

Organização de dados

- A frequência absoluta é o número de vezes que um valor (número ou modalidade) é observado.
- A frequência relativa é igual ao quociente entre a respectiva frequência absoluta e o número de dados.
- A frequência absoluta acumulada correspondente ao máximo do conjunto de dados é igual ao número de dados.
- A frequência relativa acumulada correspondente ao máximo do conjunto de dados é igual a 1.

Organizar e representar a informação contida em dados quantitativos discretos e contínuos pode ser feita recorrendo a tabelas de frequências absolutas, absolutas acumuladas, relativas e relativas acumuladas.

Pode, ainda, recorrer-se a representações gráficas adequadas, como, por exemplo, gráficos de barras, diagramas de caule-e-folhas e diagramas de extremos e quartis.

Histograma

É um diagrama de áreas, em que para a sua construção é necessária uma organização prévia dos dados em classes na forma de intervalo.

São constituídos por retângulos, em que um dos lados coincide com a marcação das classes no eixo horizontal e a área de cada retângulo corresponde à frequência absoluta ou relativa da respectiva classe.

Medidas de Localização

- A moda de um conjunto de dados quantitativos ou qualitativos é a categoria (ou a classe) que tem maior frequência absoluta.

Classificação da moda:

- amodal: todos os valores têm igual frequência absoluta;
 - unimodal: existe um e um só valor (ou modalidade) com maior frequência;
 - bimodal: existem exatamente dois valores (ou modalidades) que têm maior frequência absoluta;
 - plurimodal: existem mais de dois valores (ou modalidades) com frequência absoluta maior que outro ou outros.
- Média aritmética (\bar{x}) é dada por

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N},$$

onde $x_1 + x_2 + \dots + x_n$ é a soma de todos dados e N é o número total de dados. Ou então:

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N},$$

em que a variável toma k valores distintos, x_1, x_2, \dots, x_k são os valores da variável, n_1, n_2, \dots, n_k são os respetivos valores da frequência absoluta e N é o número total de dados.

Quando os dados estão agrupados em classes na forma de intervalo de número reais determina-se um valor aproximado da média aritmética recorrendo a um representante de cada classe, a marca da classe, e em seguida procede-se como se os dados não estivessem agrupados.

- Marca da classe:

Sendo $[a, b[$ uma classe, a sua marca é igual a $\frac{a+b}{2}$.

- Mediana (\tilde{x})

- Se o número de dados é ímpar, a mediana é o valor do dado que ocupa a posição central, quando estes estão ordenados por ordem crescente ou decrescente.
- Se o número de dados é par, a mediana é a média aritmética dos dois dados que ocupam as posições centrais, quando estes estão ordenados por ordem crescente ou decrescente.

- Percentis

- Percentil de ordem k (P_k)

Seja x_1, x_2, \dots, x_n uma amostra ordenada de dimensão n e k um número inteiro tal que $1 \leq k \leq 100$. Assim, P_k é um número tal que:

- Pelo menos $k\%$ dos elementos da amostra são menores ou iguais a P_k ;
- No mínimo $(100 - k)\%$ dos elementos da amostra são superiores a P_k .

- Como calcular percentis?

Tomamos, como exemplo, uma amostra de dimensão 20.

- Calculamos P_{30} .

Ora 30% de 20 são 6, como 6 é um número inteiro, então $P_{30} = \frac{x_6 + x_7}{2}$.

- Calculamos agora P_{72} .

Ora 72% de 20 são 14.4, como 14.4 não é um número inteiro, o percentil é o valor do dado de ordem igual ao número inteiro seguinte, isto é 15, portanto $P_{72} = x_{15}$.

- O 1º quartil é o P_{25} .
- O 2º quartil é o P_{50} ou \tilde{x} .
- O 3º quartil é o P_{75} .

Medidas de dispersão

- A amplitude de um conjunto de dados é igual à diferença entre o maior valor e o menor valor desse conjunto de dados.
- A amplitude interquartil é igual à diferença entre o 3º quartil e o 1º quartil.
- Desvio padrão (amostral) representa-se por s e é tal que:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

A variância é igual a s^2 .

Propriedades das medidas

- Dado um conjunto de dados cuja média é \bar{x} , ao adicionarmos k a cada um dos dados, obtém-se um novo conjunto de dados em que a média é igual a $\bar{x} + k$.
- Dado um conjunto de dados cuja média é \bar{x} , ao multiplicarmos cada dado por k , obtém-se um novo conjunto de dados em que a média é igual a $k \times \bar{x}$.
- Numa amostra x_1, x_2, \dots, x_n de dimensão n e média \bar{x} , a soma dos desvios $x_i - \bar{x}$, $i = 1, 2, \dots, n$ é igual a zero.

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$$

- Quanto maior for o desvio-padrão, maior é a dispersão de dados.
- O desvio padrão é igual a zero quando todos os dados são iguais, isto é, quando não há variabilidade.
- A amplitude interquartil igual a zero não implica a não existência de variabilidade.
- Sejam x_1, x_2, \dots, x_n , os dados de uma amostra de dimensão n e desvio padrão s

Se a é um número real, não nulo, tem-se:

- a amostra $x_1 + a, x_2 + a, \dots, x_n + a$ tem desvio padrão s , ou seja, o desvio padrão mantém-se inalterado.
- a amostra ax_1, ax_2, \dots, ax_n , tem desvio padrão igual a $|a| \times s$, ou seja, o desvio padrão é alterado.

Dados bivariados

- Dados quantitativos: para estudar a associação entre duas variáveis quantitativas de uma população, observam-se essas variáveis sobre cada unidade estatística, obtendo-se uma amostra de pares ordenados.
- Diagrama de dispersão (nuvem de dispersão): é uma representação gráfica em que cada par (x_i, y_i) é representado por um ponto de coordenadas (x, y) num sistema de eixos coordenados.

No eixo das abcissas representa-se a variável independente e no eixo das ordenadas a variável dependente, como habitualmente.

- Coeficiente de correlação linear (r) e é tal que:

$$r = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2} \times \sqrt{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}}$$

Note-se que $r \in [-1, 1]$.

- Se $r < 0$, diz-se que há correlação negativa. Portanto, à medida que uma variável aumenta, a outra diminui. Quanto mais próximo de -1 for o valor de r , maior é a correlação.
- Se $r = 0$, a correlação é nula.
- Se $r > 0$, diz-se que há correlação positiva. Portanto, à medida que uma variável aumenta, a outra também aumenta. Quanto mais próximo de 1 for o valor de r , maior é a correlação.
- Se $r = -1$ ou $r = 1$, os pontos do diagrama de dispersão estão sobre uma reta. Essa reta tem declive negativo se $r = -1$ e declive positivo se $r = 1$.

- Reta de regressão

No caso do diagrama de dispersão mostrar uma forte associação linear entre as variáveis, essa associação pode ser descrita pela reta de regressão ou reta dos mínimos quadrados. Normalmente, a equação dessa reta de regressão linear pode ser obtida usando a tecnologia, por exemplo, uma calculadora gráfica.

- Gráfico de linhas

É um caso particular de um diagrama de dispersão em que se pretende estudar a evolução de uma das variáveis relativamente a outra variável, de um modo geral o tempo, e em que se unem, por linhas, os pontos representados.