Características Amostrais

Estatística → Conjunto de métodos que permitem recolher, organizar, analisar e interpretar dados.

População → Conjunto de elementos (designados de unidades estatísticas) sobre os quais podem ser recolhidos dados relativos a uma característica comum.

Variável Estatística → Característica que admite valores distintos (número ou modalidade).

1. Quantitativa (dados contáveis)
    - Contínua
    - Discreta
2. Qualitativa (dados não contáveis)

Amostra → Subconjunto de uma população formado pelos elementos relativamente aos quais são recolhidos dados, designados por unidades estatísticas.

Dimensão da Amostra → Número de unidades estatísticas pertencentes à amostra.

Média de uma Amostra

Dados $n \in \mathbb{N}$ e uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ de uma variável estatística, a média da amostra $\underset{\sim}{x}$ representa-se por $\overline{x}$ e é dada por :
$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Fórmula da Média para Dados Agrupados

Dados $n \in \mathbb{N}$ e uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n); \tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_m$ com $(1 \leq m \leq n)$, são os m valores distintos da amostra $\underset{\sim}{x}$ e representa-se por $\tilde{x}$.

$$\overline{x} = \frac{\sum_{j=1}^{m} \tilde{x}_j n_j}{n}$$

onde n_j com $(1 \leq j \leq m)$ é a frequência absoluta do valor $\tilde{x}_j$.

Propriedades da Média de uma Amostra

Dados $n \in \mathbb{N}$ e uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ temos que:

1. Sendo a amostra $\underset{\sim}{y} = (ax_1 + h, ax_2 + h, ..., ax_n + h)$ então $\overline{y} = a\overline{x} + h$ onde h e a são dois números reais ;
2. A média é uma caracerística amostral pouco resistente pois para uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ basta alterar um dado x_i $i \in 1, ..., n$ para o valor da média também se alterar;
3. O valor da média está entre o máximo e o mínimo da amostra e não pode ser igual ao mínimo sem ser também igual ao máximo (o que acontece quando a amostra é constante).

Desvio em Relação à Média

Dados $n \in \mathbb{N}$, uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ e $i \in 1, ..., n$, designa-se por desvio de em relação à média a quantidade $x_i - \overline{x}$ e representa-se por d_i.

A soma dos desvios em relação à média dos valores observados é nula
$\hookrightarrow \sum\limits_{i=1}^{n} d_i = 0$

Soma dos Quadrados dos Desvios

Dados $n \in \mathbb{N}$ e uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$, a soma dos quadrados dos desvios dos x_i em relação à média representa-se por SS_x e é dada por :

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

Soma dos Quadrados dos Desvios para Dados Agrupados

Dados $n \in \mathbb{N}$ e uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$, a soma dos quadrados dos desvios para dados agrupados é dada por:

$$SS_x = \sum_{j=1}^{m}(\tilde{x}_j - \overline{x})^2 n_j$$

onde $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_m$ representam os m valores distintos da amostra $\underset{\sim}{x}$ e n_j a frequência absoluta de $\tilde{x}_j$.

Propriedades da Variância da Soma dos Quadrados dos Desvios

Dados $n \in \mathbb{N}$, uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ e os números reia h e a temos que:

1. $SS_x = 0$ se e só se $x_1 = x_2 = ... = x_n$;
2. Se $\underset{\sim}{y} = \underset{\sim}{x} + h$ então $SS_y = SS_x$;
3. Se $\underset{\sim}{y} = a\underset{\sim}{x}$ então $SS_y = a^2 SS_x$.

Variância e Desvio-Padrão de uma Amostra

Dados $n \in \mathbb{N}$ e uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ temos que:

- ${S_x}^2 = \frac{SS_x}{n-1}$ diz-se a variância da amostra $\underset{\sim}{x}$;
- $S_x = \sqrt{\frac{SS_x}{n-1}}$ diz-se o desvio-padrão da amostra $\underset{\sim}{x}$.

Propriedades da Variância e do Desvio-Padrão de uma Amostra

Dados $n \in \mathbb{N}$, uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ e os números reais h e a temos que:

1. $S_x = 0$ se e só se $x_1 = x_2 = ... = x_n$;
2. Se $\underset{\sim}{y} = \underset{\sim}{x} + h$ então $S_y = S_x$;
3. Se $\underset{\sim}{y} = a\underset{\sim}{x}$ então $S_y = |a| S_x$.

Quartis e Percentis

A mediana é o valor que separa a quantidade de dados em duas partes iguais: 50% dos dados abaixo dela e 50% acima. Assim como a mediana, existem outros valores que separam os dados em partes iguais.

Quartis $\longrightarrow$ dividem os dados em quartas partes (cada parte tem 25% dos dados).

Percentis $\longrightarrow$ dividem amostra ordenada em 100 partes onde cada parte tem 1% dos dados. (Os percentis são um caso particular dos quantis.). São indicados por $P_1, P_2, ..., P_{99}, P_{100}$.
Percentil de Ordem k

Dados $n \in \mathbb{N}$, uma amostra $\underset{\sim}{x} = (x_1, x_2, ..., x_n)$ e um número natural k∈]0,100] designa-se por **percentil de ordem k** e representa-se por P_k:

- O valor máximo da amostra é igual ao percentil 100 ou seja P_{100};
- O 1º e o 3º quartis também são conhecidos como P_{25} e P_{75}, respetivamente;
- A mediana de $\underset{\sim}{x}$ é igual ao percentil 50 (P_{50}).

Processo para Determinar o Percentil de Ordem k

1. Calcular o valor $\frac{kn}{100}$;
2. Para $k \neq 100$ se o valor $\frac{kn}{100}$ **for inteiro** então P_k é a média dos elementos de ordem $\frac{kn}{100}$ e $\frac{kn}{100}+1$, na amostra ordenada, ou seja $P_k = \frac{x_{(\frac{kn}{100})}+x_{(\frac{kn}{100}+1)}}{2}$;
3. Para $k \neq 100$ se $\frac{kn}{100}$ **não for inteiro** então $P_k = x_{([\frac{kn}{100}]+1)}$.

Exemplo:

Considera a amostra $\underset{\sim}{x} = (0, 10, 3, 4, 2, 0, 14, 3)$

Ordenando os dados da amostra ficamos com $\tilde{x} = (0, 0, 2, 3, 3, 4, 10, 14)$.

A mediana de $\underset{\sim}{x}$ é igual P_{50}.

Tem-se que $\frac{kn}{100} = \frac{50\times 8}{100} = 4$. Como 4 é um número inteiro então $P_{50} = \frac{x_4+x_5}{2} = \frac{3+3}{2} = 3$