Funções e cálculos estatísticos

Estatística Descritiva



**Sub-grupo do conteúdo estatístico**

Ana Clara Lacerda da Silva, Emanuel Gonçalves Menezes, João Pedro Inacio Porto Vidigal, Luiz Gustavo Alves Alencar e Uigor Teodoro Martins.

# Distribuição de frequências

* **Pontual, sem perda de informação**

A construção de uma distribuição de freqüência pontual é equivalente à construção de uma tabela simples, onde se listam os diferentes valores observados da variável, com suas freqüências absolutas, denotadas por Fi, onde o índice i corresponde ao número de linhas da tabela,

* **frequência relativa**

onde n é o tamanho da amostra, devendo ser substituída por N se os dados forem populacionais. A soma das freqüências relativas de todas as categorias é igual a 1; 

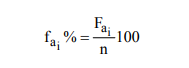
* **relativa em percentual**



representando o percentual de observações que pertencem àquela categoria. A soma das freqüências deve, agora, ser igual a 100%;

* **absoluta acumulada**

A frequência absoluta acumulada é a soma das frequências absolutas ao decorrer das linhas da tabela. Essa frequência é bastante útil para obter alguns dados de determinada tabela.

* **acumulada relativa**

A frequência relativa acumulada é o acúmulo da frequência relativa. Para encontrar a frequência relativa acumulada, acrescentamos uma nova coluna à tabela. Copiamos a primeira frequência relativa na primeira linha, a segunda linha será a soma da frequência relativa da linha com a frequência acumulada da linha anterior, e assim sucessivamente.

* **Em classes, com perda de informações.**

O menor valor da classe é denominado limite inferior (li) e o maior valor da classe é denominado limite superior (Li). O intervalo ou classe pode ser representado das seguintes maneiras:

a) li | \_\_\_\_ Li, onde o limite inferior da classe é incluído na contagem da freqüência absoluta mas o superior não;

b) li \_\_\_\_| Li, onde o limite superior da classe é incluído na contagem mas o inferior não;

c) li | \_\_\_\_| Li, onde tanto o limite inferior quanto o superior são incluídos na contagem;

d) li \_\_\_\_ Li, onde os limites não fazem parte da contagem.

Pode-se escolher qualquer uma destas opções sendo o importante tornar claro no texto ou na tabela qual está sendo usada.

Milone (2004, p.36) apresenta os seguintes critérios para a determinação do número de intervalos, denotado por k:

1. Raiz quadrada: k n =

2. Log (Sturges): k =1+ 3,3log n

3. ln (Milone): k 1 2 ln n =− + ⋅

4. k 1 10 AT d = + ,

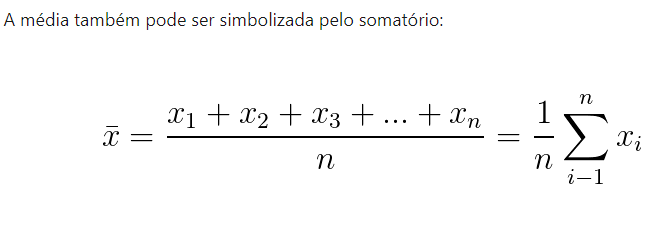
onde n é o número de elementos da amostra, AT é a amplitude total dos dados e d é o número de decimais de seus elementos.

# Medidas Descritivas

Medidas de tendência central.

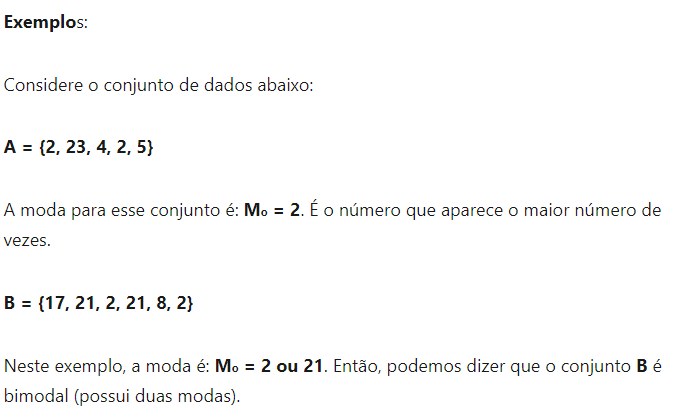
* **Média aritmética (amostral e populacional).**

A Média Aritmética de um conjunto de dados é obtida somando todos os valores e dividindo o valor encontrado pelo número de dados desse conjunto.



* **Moda.**

Para calcular a moda de um conjunto de dados só é preciso observar os dados que aparecem com maior frequência no conjunto.



* **Mediana.**

A Mediana (**Md**) é o valor de centro de um conjunto de dados. Para calcular, primeiro devemos ordenar o conjunto de dados.

Para calcular a mediana:

* Devemos ordenar o conjunto de dados em ordem crescente;
* Se o número de elementos for par, então a mediana é a média dos dois valores centrais. Soma os dois valores centrais e divide o resultado por 2: **(a + b)/2**.
* Se o número de elementos for ímpar, então a mediana é o valor central.

# Medidas Separatrizes

* **Quartil.**

Os quartis dividem o conjunto de dados em quatro partes iguais.



* **Decil.**

Os decis dividem o conjunto de dados em dez partes iguais.



* **Percentil. (cálculo complexo)**

Os percentis dividem o conjunto de dados em cem partes iguais.



# Medidas de Dispersão

* **Amplitude total.**

A amplitude total de um conjunto de dados é a diferença entre o maior e o menor valor observado. A medida de dispersão não levar em consideração os valores intermediários perdendo a informação de como os dados estão distribuídos e/ou concentrados.



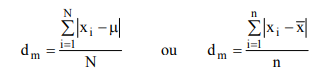
* **Amplitude interquartílica.**

A amplitude interquartílica é a diferença entre o terceiro e o primeiro quartil. Esta medida é mais estável que a amplitude total por não considerar os valores mais extremos. Esta medida abrange 50% dos dados e é útil para detectar valores discrepantes. Por outras palavras, é a distância entre o terceiro quartil e o primeiro quartil.

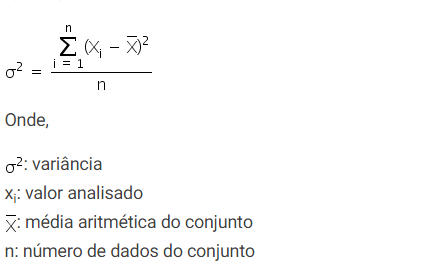


* **Desvio médio.**

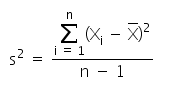
Ao somar todos os desvios, ou seja, ao somar todas as diferenças de cada valor observado em relação a média, o resultado é igual a zero (propriedade 5 da média). Isto significa que esta medida não mede a variabilidade dos dados. Para resolver este problema, pode-se desconsiderar o sinal da diferença, considerando-as em módulo e a média destas diferenças em módulo é denominada desvio médio.



* **Variância populacional.**



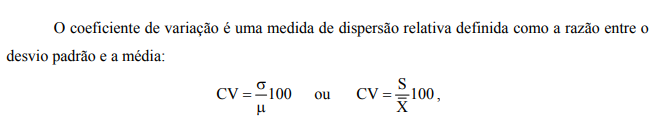
* **Variância amostral.**



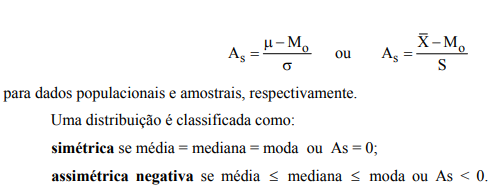
* **Desvio Padrão.**



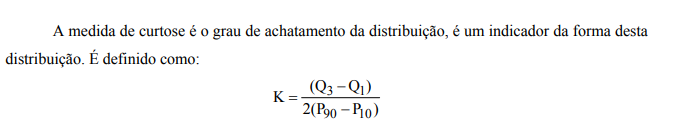
* Coeficiente de variação.



* **Medidas de assimetria.**



* **Medidas de curtose.**



# 

# 