

```

watershed.c > ...
1  #include <stdio.h>
2  #include <string.h>
3  #include <stdlib.h>
4  #include "../utils/image
5
6  /*-----
7   * Fila de prioridade
8   *-----
9  typedef struct no *ptno;
10 typedef struct no
11 {
12     int i, j;
13     ptno next;
14 } no;
15
16 /*-----
17 * init *Q new in
18 * [a:] [b:] + [c:] -> [a:]
19 * ^
20 *-----
21 void InsQ(ptno *Q, int i, int j)
22 {
23     ptno new = malloc(sizeof(no));
24     new->i = i;
25     new->j = j;
26     if (!(*Q))
27         new->next = new;
28 }

```

# Estrutura de dados: Aplicações para Vetores

Luiz Eduardo da Silva

Algoritmos e Estrutura de Dados I

Ciência da Computação

UNIFAL-MG

# Agenda

## 1 Aplicação para vetores

- Ciência de Dados
- Estimativa de Localização
- Estimativa de Dispersão

## 2 Atividade

- Exercício
- Exemplo

# Agenda

## 1 Aplicação para vetores

- Ciência de Dados
- Estimativa de Localização
- Estimativa de Dispersão

## 2 Atividade

## Ciência de Dados

- A **ciência de dados** é uma nova área de aplicação que funde diversas disciplinas, incluindo ciência da computação, estatística e outras áreas específicas.
- Uma das primeiras tarefas de um cientista de dados é **explorar os dados**, ou seja, analisar os dados buscando encontrar alguns padrões e características que possam ajudar na sua interpretação.
- Os dados a serem analisados podem estar em várias formas, como num conjunto de valores numéricos, num **vetor**.
- A partir desse conjunto de valores podem-se extrair algumas estimativas, como estimativa de **localização** e de **dispersão**.

## Média

- Um passo fundamental na análise e exploração dos dados é a **estimativa de localização**, isto é, um "valor típico" de onde a maioria dos dados estão localizados.
- Dentre as diversas estimativas de localização, temos a média que é dado pela seguinte fórmula:

$$\bar{x} = \frac{\sum_{i=0}^{n-1} x_i}{n}$$

- onde:
  - $\bar{x}$  é o valor da média
  - $n$  é a quantidade de valores
  - $x_i$  é cada valor do vetor de dados

## Média aparada

- Uma **outra estimativa de localização** é a média aparada que é menos afetada por valores extremos, pois exclui uma faixa fixada de valores "piores" e "melhores" de um vetor de dados. Sua fórmula é:

$$\bar{x} = \frac{\sum_{i=p}^{n-p-1} x_i}{n - 2p}$$

- onde:
  - $\bar{x}$  é o valor da média
  - $n$  é a quantidade de valores
  - $x_i$  é cada valor do vetor
  - $p$  é a quantidade de valores descartados no início e no fim do vetor de dados

## Desvio padrão

- A **estimativa de dispersão** ou variabilidade é usada pelo cientista de dados para verificar se os dados estão compactados e espalhados.
- Dentre as estimativas de dispersão, temos o desvio padrão, que é dado pela fórmula:

$$s = \sqrt{\frac{\sum_{i=0}^{n-1} (x_i - \bar{x})^2}{n - 1}}$$

- onde:
  - $n$  é a quantidade de valores
  - $\bar{x}$  é o valor da média
  - $x_i$  é cada valor do vetor de dados

# Agenda

## 1 Aplicação para vetores

## 2 Atividade

- Exercício
- Exemplo



## Exercício

### Exercício

- Completar o algoritmo seguinte que define um vetor  $X$  com até 100 valores numéricos. A parte do algoritmo (que já está pronto) lê o valor de  $N$  e  $P$ , que são, respectivamente, a quantidade de valores que serão efetivamente utilizados no vetor e a quantidade de valores a serem desprezados no cálculo da média aparada. E em seguida, os  $N$  valores para preencher o vetor (**para simplificar, considere que os valores lidos serão digitados de maneira ordenada**).
  - 1 Fazer uma repetição para o cálculo do média
  - 2 Fazer uma repetição para o cálculo da média aparada e
  - 3 Fazer uma repetição para o cálculo do desvio padrão.

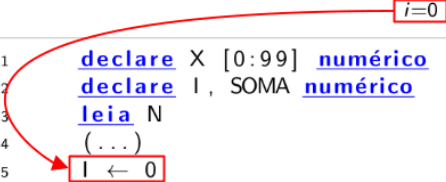
```
1 algoritmo
2   declare X [0:99] numérico
3   declare I, N, P, MEDIA, MEDAPARA, DESVIO numérico
4   declare ...
5   leia N
6   leia P
7   I  $\leftarrow$  0
8   repita
9     se I  $\geq$  N
10      então interrompa
11      fim-se
12      leia X[I]
13      I  $\leftarrow$  I + 1
14  fim-repita
15  (...)
16 fim-algoritmo
```

$$soma = \sum_{i=0}^{n-1} x_i$$

```
1  declare X [0:99] numérico
2  declare I, SOMA numérico
3  leia N
4  (...)
5  I ← 0
6  SOMA ← 0
7  repita
8      se I >= N
9          então interrompa
10     fim-se
11     SOMA ← SOMA + X[I]
12     I ← I + 1
13 fim-repita
14 (...)
```

$$soma = \sum_{i=0}^{n-1} x_i$$

```
1  declare X [0:99] numérico
2  declare I, SOMA numérico
3  leia N
4  (...)
5  I ← 0
6  SOMA ← 0
7  repita
8      se I >= N
9          então interrompa
10         fim-se
11         SOMA ← SOMA + X[I]
12         I ← I + 1
13 fim-repita
14 (...)
```



$$soma = \sum_{i=0}^{n-1} x_i$$

```
1  declare X [0:99] numérico
2  declare I, SOMA numérico
3  leia N
4  (...)
5  I ← 0
6  SOMA ← 0
7  repita
8      se I >= N
9          então interrompa
10     fim-se
11     SOMA ← SOMA + X[I]
12     I ← I + 1
13 fim-repita
14 (...)
```

