

# Primeiro Exercício

## Estatísticas de um arquivo de dados

2020

### Objetivo

Você vai desenvolver um programa que, dado um arquivo de entrada com um conjunto de valores, calcula a média, o desvio padrão e um histograma desses valores.

### Dados de entrada

Os dados serão fornecidos em um arquivo cujo nome dever ser lido da linha de comando. Esse arquivo irá consistir em uma sequência de tamanho arbitrário de valores de ponto flutuante separados por espaços em branco (espaço, mudança de linha ou tabulação).

### O que calcular

Dando o nome  $x_i$  para o  $i$ -ésimo valor no arquivo de entrada, o programa deve calcular:

**Número de valores** O número de valores fornecidos,  $N$ .

**Média** A média dos valores fornecidos:

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

**Desvio padrão amostral** O desvio padrão amostral dos valores fornecidos, definido como:

$$\sigma = \sqrt{\frac{\sum_{x=0}^{N-1} (x_i - \bar{x})^2}{N - 1}}.$$

**Histograma** Um histograma com  $B$  caixas dos valores fornecidos. Esse histograma deve ser computado como segue:

1. O número de caixas  $B$  deve ser lido da linha de comando, como argumento seguinte ao nome do arquivo de dados.
2. Encontram-se os valores mínimo  $x_{\min}$  e máximo  $x_{\max}$  entre os fornecidos.
3. Computa-se o tamanho das caixas do histograma:

$$\Delta = \frac{x_{\max} - x_{\min}}{B}.$$

4. Inicia-se um vetor de contadores inteiros de tamanho igual ao número de caixas com seus elementos em zero.
5. Para cada um dos valores lidos,  $x_i$ , computa-se a caixa  $k_i$  a que ele pertence:

$$k_i = \begin{cases} \lfloor \frac{x_i - x_{\min}}{\Delta} \rfloor, & x_i \neq x_{\max} \\ B - 1, & x_i = x_{\max} \end{cases},$$

onde  $\lfloor y \rfloor$  indica o inteiro igual ou imediatamente abaixo de  $y$ .  
Incrementa-se então o contador de índice  $k_i$ .

## Escrita dos resultados

Os resultados devem ser escritos na saída padrão `std::cout` da seguinte forma:

1. O número de elementos lidos,  $N$ , em uma linha.
2. A média dos valores,  $\bar{x}$ , na linha seguinte.
3. O desvio padrão,  $\sigma$ , na linha seguinte.
4. Os dados do histograma nas  $B$  linhas seguintes, onde cada linha tem as informações de uma caixa da seguinte forma:
  - Início da caixa.
  - Final da caixa.
  - Contador de elementos na caixa.

Esse valores devem ser separados na linha por um espaço em branco.

## Organização do código

- Coloque todo o código em um único arquivo `.cpp`.
- Divida o código em funções da seguinte forma:
  - Uma função que recebe o nome do arquivo, lê os dados desse arquivo e os retorna em um vetor de `double`.
  - Uma função que, dado o vetor de dados, computa a média e o desvio padrão, e retorna esses dois valores.
  - Uma função que recebe o vetor de dados e o número  $B$  de caixas desejado e computa o histograma. Essa função deve retornar o vetor de informações das caixas (veja a seguir) e o vetor de contadores. O vetor de contadores tem uma entrada para cada uma das  $B$  caixas, que indica quantos valores caem dentro dessa caixa. O vetor de informação de caixa consiste de um vetor de `double` com  $B + 1$  elementos, indicando que a caixa  $k$  inicia no valor da entrada  $k$  e termina no valor da entrada  $k + 1$ .
  - A função `main` deve ler o nome do arquivo e o número de caixas da linha de comando, chamar as funções auxiliares e escrever os resultados da forma pedida.
- Comece com o protótipo das funções usadas na `main`, seguido do código da `main` e depois dos códigos das outras funções.
- Use nomes de variáveis e funções adequados (ajudam a entender para que a variável ou função é usada).
- Use comentários nos locais apropriados do código para ajudar a sua compreensão.