

PCAM - Cálculos Estatísticos sobre Dados de Experimentos

Um pesquisador possui um conjunto de amostras agrupadas em colunas de uma matriz de dados (cada coluna representa as amostras de um experimento realizado). Há n amostras em cada experimento e há um total de E experimentos. Os dados coletados nos experimentos são inteiros que variam entre 0 e 99. No intuito de inferir sobre os dados obtidos nos experimentos, o pesquisador precisa calcular, para cada experimento (coluna) as seguintes métricas: **média aritmética**, **média harmônica**, **mediana**, **moda**, **variância**, **desvio padrão** e **coeficiente de variação**. Cada métrica pode ser definida da seguinte forma:

Média aritmética: Somatório de todos os elementos da amostra, divididos pelo tamanho da amostra (somas das linhas da coluna da matriz divididos pela quantidade de linhas);

$$\frac{1}{n} \sum_{i=1}^n x_i$$

Média harmônica: Razão entre o tamanho da amostra e o somatório do inverso das amostras:

$$\frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Mediana: Elemento médio da amostra (elemento médio da **coluna ordenada**). Para um número par de elementos, a mediana de um vetor **vet** seria a média entre os elementos do meio **((vet[n/2]+vet[n/2+1])/2)**. Em outras palavras seriam as posições do vetor:

$$\frac{n+1}{2} \quad \text{ou} \quad \frac{n}{2} \text{ e } \frac{n}{2} + 1$$

Moda: Elemento mais frequente da amostra (elemento que mais aparece na coluna, se houver mais de um, considera-se somente o primeiro. Se não houver, retorna -1).

Variância: Soma dos quadrados das diferenças entre o elemento da amostra e a **média aritmética** calculada, dividido pelo tamanho da amostra menos 1 unidade.

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Desvio padrão: Raiz quadrada da **variância**.

$$S = \sqrt{s^2}$$

Coeficiente de variação: Razão entre o **desvio padrão** e a **média aritmética**.

$$CV = \frac{s}{\bar{x}}$$

Considere este exemplo com 04 experimentos ($E = 4$) e 06 amostras por experimento ($n = 6$). Dada a matriz $A[6,4]$:

9	8	4	5
4	12	20	40
8	8	4	4
8	12	4	21
33	44	20	1
10	18	17	10

A saída é formada pelas métricas calculadas em cada coluna (ou experimento):

Média aritmética: 12.0, 17.0, 11.5, 13.5

Média harmônica: 8.1, 12.1, 6.6, 3.7

Mediana: 8.5, 12.0, 10.5, 7.5

Moda: 8.0, 8.0, 4.0, -1.0

Variância: 110.0, 188.4, 68.7, 217.9

Desvio Padrão: 10.5, 13.7, 8.3, 14.8

Coeficiente de variação: 0.9, 0.8, 0.7, 1.1

O objetivo deste exercício é desenvolver um projeto baseado na metodologia PCAM.

O algoritmo será desenvolvido com OpenMP e executará em um processador multicore. Os números de experimentos, amostras e núcleos de processamento podem ser grandes, porém, assume-se que todas as amostras (sem replicação delas) podem ser armazenadas na memória para execução.

Submeta no e-disciplinas uma solução por grupo, contendo detalhamento do particionamento, comunicação, aglomeração e mapeamento. Indique no início do PDF submetido os nomes dos integrantes do grupo que fizeram a atividade.
