Aula 1 - Somatório, medidas de tendência central e medidas de dispersão

Filipe Chagas

2 - Junho - 2022

Sumário

1 Somatório

1.1 Definição

Somatório é a soma de uma sequência de valores. Por exemplo: o somatório da sequência $[1, 2, 3, 4]$ é $1 + 2 + 3 + 4$.

A simbologia padrão do somatório é a seguinte:

$$\sum_{\textbf{índice}=\textbf{índice inicial}}^{\textbf{índice final}} \textbf{expressão} \tag{1}$$

Onde

- Σ é a letra grega 'Sigma' maiúscula;
- **índice** é uma variável usada para enumerar a sequência de valores do somatório. As letras i e j costumam ser usadas como variáveis de índice.
- **índice inicial** é o valor inicial da variável de índice.

- **índice final** é o valor final da variável de índice.

Em um somatório, o valor da variável de índice é incrementado de 1 em 1 até atingir o índice final.

Exemplo:

$$\sum_{i=1}^{4} i^2 = 1^2 + 2^2 + 3^2 + 4^2 \tag{2}$$

Neste exemplo, a variável de índice é i, o índice inicial é 1, o índice final é 4 e a expressão é i^2. Observe que os valores que são atribuídos a i ao longo do somatório são, respectivamente, 1, 2, 3 e 4.

Quando queremos representar o somatório de uma sequência, podemos usar um índice subscrito no símbolo da sequência para referenciar qual valor da sequência estamos somando. **Exemplo:** o somatório da sequência $x = [2, 0, 2, 2]$ pode ser escrito como:

$$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 2 + 0 + 2 + 2 = 6 \tag{3}$$

Outro exemplo: o somatório dos quadrados de uma sequência $u = [...]$ de n valores pode ser escrito como:

$$\sum_{i=1}^{n} {u_i}^2 = {u_1}^2 + {u_2}^2 + ... + {u_n}^2 \tag{4}$$

1.2 Propriedades

Algumas propriedades importantes do somatório são:

- $\sum_{i=m}^{n} ax_i = a \sum_{i=m}^{n} x_i$ (distributividade)
- $(\sum_{i=m}^{n} x_i) + (\sum_{i=m}^{n} y_i) = \sum_{i=m}^{n}(x_i + y_i)$ (associatividade)
- $(\sum_{i=u}^{n} x_i)(\sum_{j=v}^{m} y_j) = \sum_{i=u}^{n} \sum_{j=v}^{m} x_i y_j$ (distributividade/fatoração)

1.3 Questões

Responda os itens **a** e **b** considerando a seguinte equação:

$$y = \sum_{x=1}^{n} (-1)^x \tag{5}$$

(a) Sabendo apenas que n é um número positivo par, é possível determinar o valor de y? Se sim, qual é o valor de y?

(b) E se n for ímpar, qual é o valor de y?

2 Medidas de tendência central

2.1 Média

Média é a medida de tendência central mais usada. Existem dois tipos de média: média simples e média ponderada. A média simples de uma sequência $x = [x_1, x_2, ..., x_n]$ é definida como:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + ... + x_n}{n} \tag{6}$$

Já a média ponderada considera pesos. A média ponderada de uma sequência $x = [x_1, x_2, ..., x_n]$ com pesos $w = [w_1, w_2, ..., w_n]$ é definida como:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1x_1 + w_2x_2 + ... + w_nx_n}{w_1 + w_2 + ... + w_n} \tag{7}$$

2.2 Mediana

Mediana é uma medida de tendência central que corresponde ao elemento central de uma sequência **em ordem crescente**. Por exemplo: a mediana de uma sequência $[1, 3, 4, 6, 9]$ é 4, pois 4 é o elemento central da sequência.

A definição de mediana para uma sequência $x = [x_1, x_2, ...x_n]$ depende da divisibilidade de n por dois. Se n é impar, a sequência tem apenas um elemento central, que é $x_{(n+1)/2}$, portanto a mediana é definida como:

$$\textbf{mediana}(x) = x_{(n+1)/2} \tag{8}$$

Se n é par, então há dois elementos centrais na sequência, que são $x_{(n/2)}$ e $x_{(n/2)+1}$. Neste caso, a mediana é definida como a média simples dos dois elementos centrais da sequência, ou seja:

$$\textbf{mediana}(x) = \frac{x_{(n/2)} + x_{(n/2)+1}}{2} \tag{9}$$

2.3 Moda

Moda é o valor mais frequente de uma sequência. Exemplo: a moda da sequência $[5, 2, 4, 2, 3, 5, 4, 5, 5, 6]$ é 5, pois 5 é o valor que aparece mais vezes na sequência.

3 Medidas de dispersão

Medidas de dispersão servem para medir o quão discrepantes ou 'espalhados' os valores de uma sequência estão. Nesta aula, veremos duas medidas de dispersão: variância e desvio padrão.

3.1 Variância

A variância é uma medida de dispersão que indica o 'quão longe' os valores de uma sequência estão da média. A variância de uma sequência $x = [x_1, x_2, ..., x_n]$ é definida como:

$$\textbf{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{10}$$

Obs: lembre-se de que $\bar{x}$ é a média de x.

var(x) é um dos símbolos que representam a variância da sequência x. Outros símbolos comuns são **Var**$[x]$ e σ^2 ('σ' é a letra grega Sigma minúscula).

A ideia da variância é basicamente ser a média de todas as diferenças $x_i - \bar{x}$ da sequência elevadas ao quadrado. Elevar as diferenças ao quadrado faz com que não haja valores negativos no somatório. Quanto mais 'longe' os valores da sequência estiverem da média da sequência, maior será a variância.

Exemplo: considere a sequência $x = [5, 5, 5, 5]$. A média da sequência é:

$$\bar{x} = \frac{1}{4}\sum_{i=1}^{4} x_i = \frac{5+5+5+5}{4} = \frac{4 \times 5}{4} = 5 \tag{11}$$

A variância da sequência é:

$$\mathbf{var}(x) = \frac{1}{4}\sum_{i=1}^{4}(x_i - \bar{x})^2 = \frac{(5-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2}{4} = 0 \tag{12}$$

A variância de x é nula pois não há diferença entre os valores da sequência e a média.

Outro exemplo: considere a sequência $y = [1, 5, 1, 9, 9,]$. A média da sequência é:

$$\bar{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{1+5+1+9+9}{5} = 5 \tag{13}$$

A variância da sequência é:

$$\mathbf{var}(x) = \frac{1}{5}\sum_{i=1}^{5}(y_i - \bar{y})^2 = \frac{(1-5)^2 + (5-5)^2 + (1-5)^2 + (9-5)^2 + (9-5)^2}{5} = 12.8 \tag{14}$$

Como há valores em y 'longe' da média $\bar{y}$, a variância de y é maior que zero.

3.2 Desvio padrão

Desvio padrão é uma medida de tendência central que segue a mesma ideia da variância, mas é definida de forma um pouco diferente. A definição do desvio padrão para uma sequência $x = [...]$ de n valores é:

$$\sigma = \sqrt{\mathbf{var}(x)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{15}$$

σ é o principal símbolo usado para o desvio padrão. Outros símbolos menos comuns são **dp**(x) e **std**(x) (acrônimo de *standard deviation*).

O desvio padrão é a medida de dispersão mais usada na prática. Duas vantagens que o desvio padrão tem em relação à variância são:

- **Resultados menos 'desproporcionais' e mais fáceis de interpretar.**

  A variância pode dar resultados muito grandes por usar o quadrado de $x_i - \bar{x}$. O desvio padrão não dá resultados tão grandes, sendo mais próximo da média de $|x_i - \bar{x}|$.

- **Aplicações no cálculo de probabilidades.**

  O desvio padrão é muito utilizado em cálculos de probabilidade que envolvem funções de densidade de probabilidade. Este tema é complexo e não é cobrado no ENEM.

Os exemplos a seguir mostram comparações entre variância, desvio padrão e a média de $|x_i - \bar{x}|$ para algumas sequências. **Exemplos:**

Sequência	$x = [4, 4, 4, ..., 4]$
Média	$\bar{x} = 4$
Média de $\|x_i - \bar{x}\|$	0
Variância	$\sigma^2 = 0$
Desvio padrão	$\sigma = 0$

Sequência	$x = [1, 1, 5, 5]$
Média	$\bar{x} = 3$
Média de $\|x_i - \bar{x}\|$	2
Variância	$\sigma^2 = 4$
Desvio padrão	$\sigma = 2$

Sequência	$x = [3, 5, 6, 9, 4, 2, 1, 2, 1, 6]$
Média	$\bar{x} = 3.9$
Média de $\|x_i - \bar{x}\|$	2.1
Variância	$\sigma^2 \approx 6.09$
Desvio padrão	$\sigma \approx 2.47$

Sequência	$x = [8, 3, 4, 2, 3, 6, 8, 6, 4, 3]$
Média	$\bar{x} = 4.7$
Média de $\|x_i - \bar{x}\|$	≈ 1.84
Variância	$\sigma^2 \approx 4.21$
Desvio padrão	$\sigma \approx 2.05$