

Universidade Federal de Lavras

Departamento de Ciências Exatas

Prof. Daniel Furtado Ferreira

1^a Aula Prática

Técnicas de somatório

Notação e propriedades:

- 1) Variáveis e índices: o símbolo x_j (leia x índice j) representa qualquer um dos n valores x_1, x_2, \dots, x_n resultantes da mensuração de uma variável aleatória X na amostra (conjunto de dados). A letra j , usada como índice, indica um dos possíveis valores, de 1 a n , da variável aleatória. Assim, por exemplo, se for considerada uma amostra de tamanho $n = 3$ de coelhos ao abate aos noventa dias e se X representa uma variável relativa ao peso em kg, então uma possibilidade de resultados é: 2,56, 2,43 e 2,60. Logo, $x_1 = 2,56$, $x_2 = 2,43$ e $x_3 = 2,60$. Os valores da variável aleatória são representados por letras minúscula e as variáveis aleatórias em si, por letras maiúscula.
- 2) Notação de somatório: para representarmos a soma de n variáveis aleatórias podemos utilizar o símbolo \sum , letra grega maiúscula *sigma*. Assim, a soma $x_1 + x_2 + \dots + x_n$ pode ser sintetizada por $\sum_{j=1}^n x_j$, ou seja,

$$\sum_{j=1}^n x_j = x_1 + x_2 + \dots + x_n.$$

A variação do índice j pode não ir de 1 a n , mas estar em qualquer subintervalo desses limites.

- 3) Algumas propriedades:

$$\text{a) } \sum_{j=1}^n ax_j = ax_1 + ax_2 + \dots + ax_n = a \sum_{j=1}^n x_j;$$

$$\text{b) } \sum_{j=1}^n x_j y_j = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \neq \left(\sum_{j=1}^n x_j \right) \left(\sum_{j=1}^n y_j \right);$$

$$\text{c) } \sum_{j=1}^n (ax_j + by_j) = a \sum_{j=1}^n x_j + b \sum_{j=1}^n y_j;$$

$$\text{d) } \sum_{j=1}^n k = nk,$$

em que a , b e k são constantes.

Exercícios propostos:

- 1) Considere as amostras de tamanho $n = 5$ dadas por:

$$X = [2, 6, 4, 7, 10]$$

$$Y = [42, 82, 79, 31, 55]$$

obter:

$$\text{a) } \sum_{j=1}^4 x_j$$

$$\text{b) } \sum_{j=1}^5 y_j$$

$$\text{c) } \sum_{j=1}^5 2x_j^2$$

$$\text{d) } \sum_{j=1}^5 x_j y_j$$

$$\text{e) } \sum_{j=1}^5 (3x_j + 2y_j)$$

$$\text{f) } \sum_{j=2}^4 x_j y_j + \sum_{j=1}^5 y_j^2$$

2) Considere

$$\bar{X} = \frac{\sum_{j=1}^n X_j}{n} \quad \text{e} \quad S^2 = \frac{1}{n-1} \left[\sum_{j=1}^n X_j^2 - \frac{\left(\sum_{j=1}^n X_j \right)^2}{n} \right],$$

os estimadores da média e da variância populacionais, respectivamente, obtidos em uma amostra aleatória de tamanho n . Considerando os dados $X = \{2, 4, 5, 6, 1, 8\}$, calcular a média e a variância.

3) Mostrar numericamente, a partir do conjunto $X = [1, 3, 2]$ e de forma algébrica, para qualquer amostra de tamanho n , que $\sum_{j=1}^n (X_j - \bar{X}) = 0$.

4) Demonstrar que o valor de

$$Q = \frac{\sum_{j=1}^n (X_j - A)^2}{n-1}$$

representa um ponto de mínimo se o valor de A for igual a \bar{X} . Representar em um gráfico o esboço da função Q .

5) Criar um conjunto de valores de tamanho $n = 3$ para que a seguinte igualdade se verifique:

$$Q = \frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n-1} = 0.$$

6) Desenvolver a expressão

$$Q = \frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n-1}$$

considerando as propriedades de somatório e mostrar que $Q = S^2$. A partir deste resultado e daquele obtido no exercício 5, qual é o significado e interpretação você atribui à variância S^2 ?

Resolução

1) Sejam as amostras de tamanho $n = 5$ dadas por:

$$X = [2, 6, 4, 7, 10]$$

$$Y = [42, 82, 79, 31, 55]$$

então:

$$\text{a) } \sum_{j=1}^4 x_j = x_1 + x_2 + x_3 + x_4 = 19 ;$$

$$\text{b) } \sum_{j=1}^5 y_j = y_1 + \dots + y_5 = 289 ;$$

$$\text{c) } \sum_{j=1}^5 2x_j^2 = 2 \sum_{j=1}^5 x_j^2 = 410 ;$$

$$\text{d) } \sum_{j=1}^5 x_j y_j = 1659 ;$$

$$\text{e) } \sum_{j=1}^5 (3x_j + 2y_j) = 3 \sum_{j=1}^5 x_j + 2 \sum_{j=1}^5 y_j = 665 ; \quad \text{e}$$

$$\text{f) } \sum_{j=2}^4 x_j y_j + \sum_{j=1}^5 y_j^2 = 19740 .$$

2) Considerando o conjunto de dados $X = \{2, 4, 5, 6, 1, 8\}$ a média e a variância são:

$$\bar{X} = \frac{\sum_{j=1}^n X_j}{n} = 4,3333 \text{ e}$$

$$S^2 = \frac{1}{n-1} \left[\sum_{j=1}^n X_j^2 - \frac{\left(\sum_{j=1}^n X_j \right)^2}{n} \right] = 6,6667$$

3) Mostrar numérica e algebricamente que as somas de desvios em relação a média aritmética é nula, qualquer que seja a amostra.

a) Numericamente

A média é igual a $\bar{X} = 2$, assim

$$\sum_{j=1}^n (X_j - \bar{X}) = (1 - 2) + (3 - 2) + (2 - 2) = -1 + 0 + 1 = 0;$$

b) algebricamente

$$\begin{aligned} \sum_{j=1}^n (X_j - \bar{X}) &= \sum_{j=1}^n X_j - \sum_{j=1}^n \bar{X} = \sum_{j=1}^n X_j - n\bar{X} \\ &= \sum_{j=1}^n X_j - \frac{\sum_{j=1}^n X_j}{1} = \sum_{j=1}^n X_j - \sum_{j=1}^n X_j = 0 \quad \text{C.Q.M.} \end{aligned}$$

4) Expandindo o somatório e derivando Q em relação a A tem-se:

$$\begin{aligned} Q &= \frac{1}{n-1} \sum_{j=1}^n (X_j - A)^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j^2 - 2AX_j + A^2) \\ &= \frac{1}{n-1} \left(\sum_{j=1}^n X_j^2 - \sum_{j=1}^n 2AX_j + \sum_{j=1}^n A^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{j=1}^n X_j^2 - 2A \sum_{j=1}^n X_j + nA^2 \right) \\ \\ \frac{dQ}{dA} &= \frac{1}{n-1} \left(-2 \sum_{j=1}^n X_j + 2nA \right) \end{aligned}$$

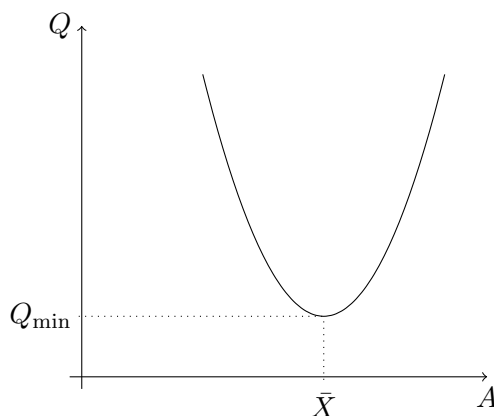
Igualando a derivada a zero, e resolvendo em relação a A , tem-se:

$$\begin{aligned} \frac{dQ}{dA} &= \frac{1}{n-1} \left(-2 \sum_{j=1}^n X_j + 2nA \right) = 0 \\ 2nA &= 2 \sum_{j=1}^n X_j \\ A &= \frac{\sum_{j=1}^n X_j}{n} = \bar{X} \end{aligned}$$

O ponto ótimo, obtido igualando a derivada primeira a zero, pode ser de máximo, de mínimo ou de inflexão. Para certificar-se de que o valor de Q é um valor mínimo, quando A é igual à média amostral, basta mostrar que a segunda derivada é positiva. A segunda derivada de Q em relação a A é dada por:

$$\frac{d^2Q}{dAdA} = \frac{2n}{n-1} > 0$$

ou seja, a segunda derivada para qualquer tamanho de amostra será positiva, ficando concluída assim a demonstração. Veja o gráfico da função a seguir, em que $Q_{\min} = S^2$.



5) Para que o somatório em questão seja nulo é necessário que cada parcela seja igual a zero. Para isso acontecer é preciso que cada valor x_j seja igual a média da amostra, ou seja, $x_j = \bar{X}$. Assim, concluímos

que os n valores da amostra têm de ser iguais. Logo, podemos construir quaisquer amostra de tamanho $n = 5$ com valores iguais, como, por exemplo, $X = \{1,1,1,1,1\}$.

6) Desenvolvendo

$$\begin{aligned}
 Q &= \frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n-1} = \frac{1}{n-1} \sum_{j=1}^n (X_j^2 - 2\bar{X}X_j + \bar{X}^2) \\
 &= \frac{1}{n-1} \left(\sum_{j=1}^n X_j^2 - \sum_{j=1}^n 2\bar{X}X_j + \sum_{j=1}^n \bar{X}^2 \right) \\
 &= \frac{1}{n-1} \left(\sum_{j=1}^n X_j^2 - 2\bar{X} \sum_{j=1}^n X_j + n\bar{X}^2 \right) \\
 &= \frac{1}{n-1} \left[\sum_{j=1}^n X_j^2 - 2 \frac{\sum_{j=1}^n X_j}{n} \sum_{j=1}^n X_j + n \left(\frac{\sum_{j=1}^n X_j}{n} \right)^2 \right] \\
 &= \frac{1}{n-1} \left[\left(\sum_{j=1}^n X_j \right)^2 - \frac{2 \left(\sum_{j=1}^n X_j \right)^2}{n} + \frac{\mathcal{N} \left(\sum_{j=1}^n X_j \right)^2}{n^2} \right] \\
 &= \frac{1}{n-1} \left[\sum_{j=1}^n X_j^2 - \frac{\left(\sum_{j=1}^n X_j \right)^2}{n} \right] = S^2
 \end{aligned}$$

Assim, a variância é função da soma de quadrados de desvios em relação a média e é um valor mínimo, se for considerada outra constante no lugar da média. Se tomarmos Q como uma função de A , que representa uma parábola, a variância representa o ponto de mínimo desta parábola, quando $A = \bar{X}$. A variância é tanto menor, próximo de zero, quanto maior for a semelhança dos dados amostrais, ou seja, será próxima de zero em amostras com pouca variação. Crescerá para infinito, quando as diferenças entre os elementos da amostra aumentar.