REMat
ISSN 2177-5095
nº3 - 2013
REVISTA ELETRÔNICA DE MATEMÁTICA
www2.jatai.ufg.br/ojs/index.php/matematica
contato: remat.ufg@gmail.com

Um susto no cálculo da Mediana

Rogério César dos Santos
Professor da UNB-FUP.
professorrogeriocesar@gmail.com

Resumo

Costuma-se dizer que a mediana é o valor que "divide" o conjunto em dois subconjuntos de mesmo tamanho, ou de mesma ordem. Essa interpretação nem sempre é esclarecida nos livros. Vejamos como deve ser entendida essa divisão em dois subconjuntos para cada caso, n par e n ímpar. Vamos explorar também o conceito de mediana a partir de dados agrupados em classes, isto é, a partir de um histograma.

Palavras-chaves: Mediana, Estatística, Histograma, Medidas de Tendência Central.

A sudden scare in median's computation

Abstract

It is said that the median is the dividing value of a given set in two subsets having the same cardinality. This interpretation isn't always clear in the books. We'll see how to understand this division in two subsets for the cases of n even and n odd. We'll also see how to understand the concept of median for clustered data, that is, a histogram.

Keywords: Median, Estatistics, Histogram, Measures of Central Tendency.

1 Definição de Mediana a partir de um conjunto numérico finito

De acordo com [LEONARDECZ – 2004], a mediana de um conjunto numérico finito e ordenado $\{y_1 \leq y_2 \leq ... \leq y_n\}$ é, por definição:

- O elemento de ordem $\frac{n+1}{2}$, se n for ímpar, ou seja, $y_{\frac{(n+1)}{2}}$;
- A média aritmética dos elementos de ordem $\frac{n}{2}$ e $\frac{n}{2}+1$, se n for par, ou seja, $\frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$.

REMat
ISSN 2177-5095
nº3 - 2013
REVISTA ELETRÔNICA DE MATEMÁTICA
www2.jatai.ufg.br/ojs/index.php/matematica
contato: remat.ufg@gmail.com

1.1 Mediana – o valor que divide um conjunto ordenado em dois subconjuntos de mesma ordem

Para estudarmos como a mediana "divide" um certo conjunto em dois, considere os três exemplos: o conjunto $A = \{2, 2, 3\}$, cuja ordem $n = 3$ é ímpar, que possui mediana $y_2 = 2$; o conjunto $B = \{1, 2, 2, 5\}$, com $n = 4$ par, que possui mediana $\frac{(2+2)}{2} = 2$, e o conjunto $C = \{1, 3, 6, 7\}$, $n = 4$ também par, que possui mediana $\frac{3+6}{2} = 4,5$.

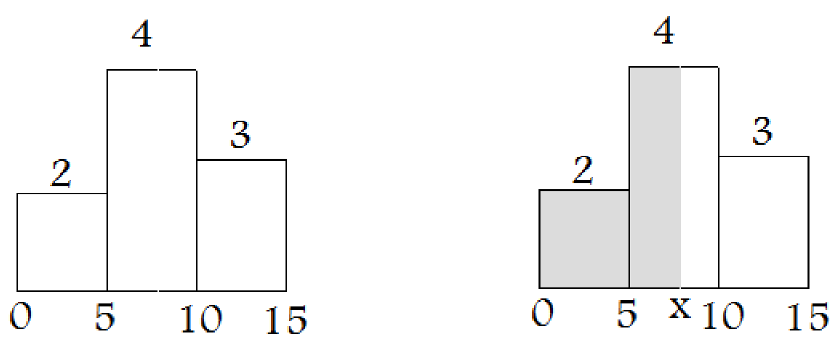
Em A, a mediana é 2, e temos dois elementos menores ou iguais do que ela, e dois elementos maiores ou iguais do que ela, isto é, a mediana 2 "divide" A em dois subconjuntos, mas a intersecção de ambos não é vazia, e isso acontece toda vez que n é ímpar.

No caso em que n é par, a mediana pode (como em B) ou não (como em C) coincidir com algum elemento do conjunto. Observe que $\frac{y_{\frac{n}{2}} - y_{\frac{n}{2}+1}}{2} = y_{\frac{n}{2}}$ ou $y_{\frac{n}{2}+1}$, se e somente se $y_{\frac{n}{2}} = y_{\frac{n}{2}+1}$.

Em B, há três valores menores ou iguais à mediana 2 e três valores maiores ou iguais do que 2, os dois subconjuntos têm um elemento em comum: a própria mediana 2. Em C, há dois valores menores do que a mediana $4,5$ e dois valores maiores do que $4,5$, os dois subconjuntos são disjuntos. Em resumo, se n for ímpar, os subconjuntos não são disjuntos, mas quando n é par, os subconjuntos podem ser disjuntos ou não. Em todo caso, os dois subconjuntos têm mesma ordem ou tamanho.

1.2 Dados agrupados em classes, através de um Histograma

Estava eu ministrando uma aula de estatística, e pedi para meus alunos calcularem a mediana a partir do histograma ilustrado abaixo:



Como os dados estão organizados em classes através de um histograma, a mediana é o valor x que deixa a área à esquerda do histograma igual à área à direita. Então, os alunos deveriam fazer a seguinte conta: a soma das áreas dos retângulos hachurados, na figura da direita, é igual à soma das áreas dos demais retângulos. Observe que o valor x deve pertencer à *classe mediana*, cujo conceito falarei adiante. A classe mediana nesse exemplo é a classe de intervalo $]5, 10]$. No entanto, um dos alunos encontrou um valor x que não pertencia à classe mediana, ou seja, estava entre 0 e 5, e não entre 5 e 10 como deveria. Daí eu pensei: o valor x que divide a área do histograma ao meio sempre pertencerá à classe mediana?

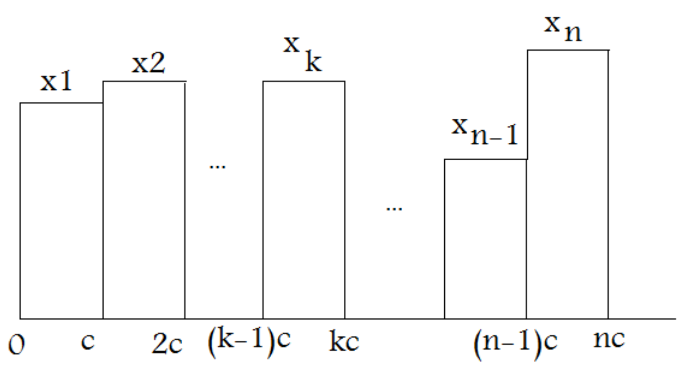


Figura 1: Em cada intervalo semi-aberto $](i-1)c,\ ic]$ há x_i valores, para $i = 1, ..., n$.

Considere o histograma acima, onde o vértice inferior esquerdo do primeiro retângulo da esquerda foi colocado na origem do semi-eixo x positivo. Para tanto, considere que:

- $x_1, x_2, \cdots, x_k, \cdots, x_n$ são as frequências de cada classe, que coincidem com as alturas dos retângulos;
- c é o comprimento de cada classe, que coincide com o tamanho da base dos retângulos;
- Os valores $0, c, 2c, \cdots, kc, \cdots, nc$ são as absissas correspondentes aos vértices dos retângulos, localizados no semi-eixo x positivo;
- Os dados coincidentes com o extremo esquerdo de cada intervalo ou classe pertencem à classe da esquerda, e os dados iguais ao extremo direito de cada intervalo pertence à própria classe correspondente, isto é, se há um dado $y = i \cdot c$, então y pertence à classe da esquerda, isto é, a classe correspondente ao intevalo semi-aberto $](i-1) \cdot c, i \cdot c]$ para $i = 1, \ldots, n$.

2 Definição de classe mediana

De acordo com a definição, a classe mediana é a primeira das classes, da esquerda para a direita, ou da direita para a esquerda, cuja frequência acumulada é maior ou igual do que a metade da soma S das frequências. Isto é, se $S = \sum_{i=1}^{n} x_i$ então a classe mediana, que na figura 1 está denotada pelo intervalo semi-aberto $](k-1)\cdot c, k\cdot c]$, é tal que $\sum_{i=1}^{k-1} x_i < \frac{S}{2}$ e $\sum_{i=1}^{k} x_i \geq \frac{S}{2}$.

2.1 Definição de mediana a partir de um histograma

Dado o histograma, a mediana é o valor x tal que a soma das áreas dos retângulos hachurados na figura abaixo (os da esquerda) é igual à soma das dos retângulos à direita. A nossa pergunta é: existe um tal valor x? E se existe, ele se localiza obrigatoriamente na classe mediana, ou seja, $(k-1)\cdot c < x \leq k\cdot c$?

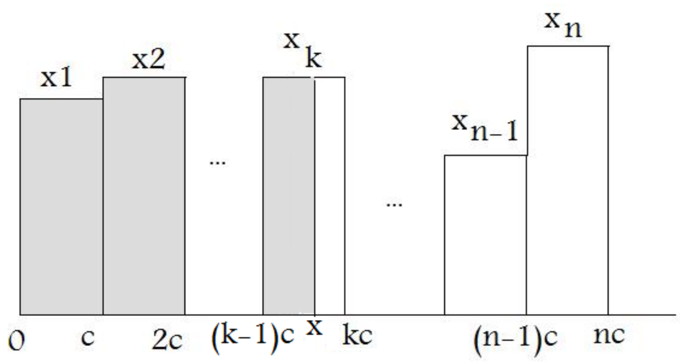


Figura 2: O valor x que divide a área do histograma ao meio pertence à classe mediana?

Vamos mostrar que sim. O meu aluno, que citei acima, havia feito o cálculo errado. A mediana sempre pertence à classe mediana, como demonstrarei abaixo:

Proposição. *Suponha que a classe* $](k-1)\cdot c, k\cdot c]$ *seja a classe mediana, isto é, suponha (I)* $\sum_{i=1}^{k-1} x_i < \frac{S}{2}$ *e (II)* $\sum_{i=1}^{k} x_i \geq \frac{S}{2}$. *Então, o valor x que deixa metade da área do histograma à sua esquerda e metade da área à sua direita pertence à classe mediana. Em outras palavras a equação (III)* $c\cdot x_1 + \cdots + c\cdot x_{k-1} + (x-(k-1)\cdot c)\cdot x_k = (k\cdot c - x)\cdot x_k + c\cdot x_{k+1} + \cdots + c\cdot x_n$ *tem solução x pertencente ao intervalo semi-aberto* $](k-1)\cdot c, k\cdot c]$.

Demonstração. Primeiro, observe que a hipótese (I) é equivalente a $2\cdot\sum_{i=1}^{k-1} x_i < \sum_{i=1}^{n} x_i$, ou seja, $\sum_{i=1}^{k-1} x_i < \sum_{i=1}^{n} x_i$. Também a hipótese (II) é equivalente a $2 \cdot \sum_{i=1}^{k} x_i \geq \sum_{i=1}^{n} x_i$, ou seja, $\sum_{i=1}^{k} x_i \geq \sum_{i=k+1}^{n} x_i$. Agora, resolvendo a equação (III) para x, temos:

$$c \cdot x_1 + \cdots + c \cdot x_{k-1} + x \cdot x_k - k \cdot c \cdot x_k + c \cdot x_k = \\ k \cdot x_k \cdot c - x \cdot x_k + c \cdot x_{k+1} + \cdots + c \cdot x_n$$

ou, $x = \dfrac{2 \cdot k \cdot c \cdot x_k + c \cdot \sum_{i=k+1}^{n} x_i - c \cdot \sum_{i=1}^{k-1} x_i - c \cdot x_k}{2 \cdot x_k}$. Este valor de x satisfaz $(k-1) \cdot c < x < k \cdot c$, como veremos a seguir:

A desigualdade da esquerda, $(k-1) \cdot c < \dfrac{2 \cdot k \cdot c \cdot x_k + c \cdot \sum_{i=k+1}^{n} x_i - c \cdot \sum_{i=1}^{k-1} x_i - c \cdot x_k}{2 \cdot x_k}$, é equivalente a

$$\begin{aligned} 2 \cdot k \cdot c \cdot x_k + c \cdot \sum_{i=k+1}^{n} x_i - c \cdot \sum_{i=1}^{k-1} x_i - c \cdot x_k > 2 \cdot x_k \cdot k \cdot c - 2 \cdot x_k \cdot c &\iff \\ \sum_{i=k+1}^{n} x_i - \sum_{i=1}^{k-1} x_i - x_k + 2 \cdot x_k > 0 &\iff \\ \sum_{i=k+1}^{n} x_i - \sum_{i=1}^{k-1} x_i + x_k > 0 &\iff \\ \sum_{i=k}^{n} x_i > \sum_{i=1}^{k-1} x_i & \end{aligned}$$

o que é verdade pela hipótese (I). Já a desigualdade da direita, $\dfrac{2 \cdot k \cdot c \cdot x_k + c \cdot \sum_{i=k+1}^{n} x_i - c \cdot \sum_{i=1}^{k-1} x_i - c \cdot x_k}{2 \cdot x_k} \leq k \cdot c$ é equivalente a

$$\begin{aligned} 2 \cdot k \cdot c \cdot x_k + c \cdot \sum_{i=k+1}^{n} x_i - c \cdot \sum_{i=1}^{k-1} x_i - c \cdot x_k \leq 2 \cdot x_k \cdot k \cdot c &\iff \\ \sum_{i=k+1}^{n} x_i - \sum_{i=1}^{k-1} x_i - x_k \leq 0 &\iff \\ \sum_{i=k+1}^{n} x_i \leq \sum_{i=1}^{k} x_i & \end{aligned}$$

que é verdade pela hipótese (II). Assim, está provado que $x \in](k-1) \cdot c, k \cdot c]$. $\square$

3 Conclusão

A demonstração aqui apresentada não deixa dúvidas. A mediana sempre pertencerá à classe mediana. Não tenham receio, professores, se o valor encontrado por algum aluno estiver fora desta classe, com certeza houve algum erro de cálculo.

Referências

[LEONARDECZ-2004] LEONARDECZ, E. **Estatística para ciências da vida.** Brasília: Universa, 2004