

Peculiaridade na determinação do desvio médio em função da mediana

E' fáto de todos conhecido que a determinação do desvio médio em função da mediana, pelo método rápido classico, exige que a origem escolhida coincida com a classe em que se encontra o valor daquela medida de posição. Em alguns casos, como no exemplo abaixo, pode succeder entretanto, que o valor da mediana coincida com o do limite de classe, colocando desta maneira o estatisticista na duvida quanto a escolha da origem, se na classe que tem por limite superior o valor da mediana, se na classe imediatamente seguinte.

Atentando-se porém para o fáto de que tal valor da mediana pertence tanto a uma como a outra destas duas classes, concluiremos que tal duvida não tem razão de ser, a origem podendo ser colocada indiferentemente, em qualquer delas: esta afirmativa é demonstrada in-

teiramente não só pelos quadros abaixo, como pela exposição que a eles se seguem.

QUADRO I

Determinação da mediana das populações das sédes dos 258 municípios do Estado de São Paulo:

VALORES DA VARIÁVEL	FREQUÊNCIAS
(X)	(F)
2.000 — 8.000	54
8.000 — 14.000	75
	129
14.000 — 20.000	36
20.000 — 26.000	29
26.000 — 32.000	24
32.000 — 38.000	8
38.000 — 44.000	8
44.000 — 50.000	2
50.000 — 56.000	6
56.000 — 62.000	4
62.000 — 68.000	5
68.000 — 74.000	2
74.000 — 80.000	1
80.000 — 86.000	1
86.000 — 92.000	1
92.000 — 98.000	0
98.000 — 104.000	0

VALORES DA VARIÁVEL	FREQUÊNCIAS
(x)	(f)
104.000 — 110.000	0
110.000 — 116.000	0
116.000 — 122.000	0
122.000 — 128.000	0
128.000 — 134.000	1
134.000 — 140.000	0
140.000 — 146.000	1
Total:	258

Temos então:

Ordem da observação, cujo valor dará a mediana $\frac{258 + 1}{2} = 129,5$. Esta observação não é definida; este fato, entretanto não importa, visto como, estando as observações distribuídas em classes, vamos procurar, por interpolação, o valor da observação de ordem $\frac{N}{2}$ ou seja 129 que se acha na 2.^a classe. Servindo-nos da fórmula clássica:

$M_i = V + \frac{(N - N_1)}{N_2 - N_1} \cdot h$ temos para o nosso caso:

$$M_i = 8.000 + \frac{(129 - 54)}{75} \cdot 6.000 = 14.000.$$

Deve-se notar a coincidência do valor da mediana com o limite superior da 2.^a classe.

QUADRO II

Determinação direta do desvio médio em função da mediana com os dados acima:

VALORES DA VARIÁVEL (x)	Frequências (f)	Desvios a partir da mediana, ob- servação feita dos sinais (x)	PRODUTOS (f. x)
2.000 — 8.000	54	9.000	486.000
8.000 — 14.000	75	3.000	225.000
14.000 — 20.000	36	3.000	108.000
20.000 — 26.000	29	9.000	261.000
26.000 — 32.000	24	15.000	360.000
32.000 — 38.000	8	21.000	168.000
38.000 — 44.000	8	27.000	216.000
44.000 — 50.000	2	33.000	66.000
50.000 — 56.000	6	39.000	234.000
56.000 — 62.000	4	45.000	180.000
62.000 — 68.000	5	51.000	255.000
68.000 — 74.000	2	57.000	114.000
74.000 — 80.000	1	63.000	63.000
80.000 — 86.000	1	69.000	69.000
86.000 — 92.000	1	75.000	75.000
92.000 — 98.000	0	81.000	0
98.000 — 104.000	0	87.000	0
104.000 — 110.000	0	93.000	0
110.000 — 116.000	0	99.000	0
116.000 — 122.000	0	105.000	0
122.000 — 128.000	0	111.000	0

VALORES DA VARIÁVEL (x)	Fre- quências (f)	Desvios a partir da mediana, obs- tração feita dos sinais (x)	PRODUTOS (f. x)
128.000 — 134.000	1	117.000	117.000
134.000 — 140.000	0	123.000	0
140.000 — 146.000	1	129.000	129.000
Totais:	258		3.126.000

Temos então: $\Sigma (f \cdot x) = 3.126.000$, por-
tanto D. M. $= \frac{3.126.000}{258} = 12.116$

QUADRO III

*Determinação do desvio médio, com os dados
acima, pelo método rápido classico.*

A origem tomada é a classe cujo limite su-
perior corresponde á mediana.

VALORES DA VARIÁVEL (x)	Fre- quências (f)	Desvios a partir de A + (d)	Produtos + (f. d)
2.000 — 8.000	54	1	54
8.000 — 14.000	75	0	0
14.000 — 20.000	36	1	36
20.000 — 26.000	29	2	58
26.000 — 32.000	24	3	72
32.000 — 38.000	8	4	32

VALORES DA VARIÁVEL (x)	Fre- quências (h)	Desvios a partir de A + (h)	Produtos + (f x)
38.000 — 44.000	8	5	40
44.000 — 50.000	2	6	12
50.000 — 56.000	6	7	42
56.000 — 62.000	4	8	32
62.000 — 68.000	5	9	45
68.000 — 74.000	2	10	20
74.000 — 80.000	1	11	11
80.000 — 86.000	1	12	12
86.000 — 92.000	1	13	13
92.000 — 98.000	0	14	0
98.000 — 104.000	0	15	0
104.000 — 110.000	0	16	0
110.000 — 116.000	0	17	0
116.000 — 122.000	0	18	0
122.000 — 128.000	0	19	0
128.000 — 134.000	1	20	20
134.000 — 140.000	0	21	0
140.000 — 146.000	1	22	22
Totais:	258		521

Temos pois: $\Sigma (f \cdot x) = 521$

$$d = \frac{14.000 - 11.000}{6.000} = 0,5$$

$$N_1 = 129$$

$$N_2 = 129$$

$$D. M. = \frac{521 + 0,5 (129 - 129)}{258} = \frac{521}{258}$$

$$= 2,019379 \text{ intervalos} = 12.116.$$

Deve-se notar a absoluta coincidência deste resultado com o do Quadro II.

QUADRO IV

Determinação do desvio médio com os dados acima pelo método rápido classico.

A origem tomada é a classe cujo limite inferior, corresponde á mediana.

VALORES DA VARIÁVEL (x)	Fre- quências (f)	Desvios a partir de A + (z)	Produtos + (f. z)
2.000 — 8.000	54	2	108
8.000 — 14.000	75	1	75
14.000 — 20.000	36	0	0
20.000 — 26.000	29	1	29
26.000 — 32.000	24	2	48
32.000 — 38.000	8	3	24
38.000 — 44.000	8	4	32
44.000 — 50.000	2	5	10
50.000 — 56.000	6	6	36
56.000 — 62.000	4	7	28
62.000 — 68.000	5	8	40
68.000 — 74.000	2	9	18
74.000 — 80.000	1	10	10
80.000 — 86.000	1	11	11
86.000 — 92.000	1	0	12
92.000 — 98.000	0	12	0
98.000 — 104.000	0	13	0
104.000 — 110.000	0	14	0
110.000 — 116.000	0	15	0

VALORES DA VARIÁVEL (x)	Fre- quências (f)	Desvios a partir de A + (f. s)	Produtos + (f. s)
116.000 — 122.000	0	17	0
122.000 — 128.000	0	18	0
128.000 — 134.000	1	19	19
134.000 — 140.000	0	20	
140.000 — 146.000	1	21	21
Totais:	258		521

Temos, pois: $\Sigma + (f. s) = 521$

$$d = \frac{14.000 - 17.000}{6.000} = -0,5$$

$$N_1 = 129$$

$$N_2 = 129$$

$$D. M. = \frac{521 - 0,5 (129 - 129)}{258} = 12.116$$

Deve-se notar a absoluta coincidência deste resultado com o do Quadro II.

Para que possamos compreender a razão do fato que ora apresentamos, mister se faz que analisemos as formulas da mediana e do desvio médio em função da mediana, calculado pelo método rápido classico: $Mi = V + \frac{(\frac{N}{2} - N_1)_n}{N_2}$
e $D. M. = \frac{\Sigma + (f. s) + d (N_1 - N_2)_i}{N}$

Inicialmente, precisamos resaltar a diferença dos significados de N_1 e N_2 , numa e noutra destas formulas: de fato, enquanto na 1.^a,

N_1 significa o total das freqüências das classes com valores inferiores á classe em que se encontra a mediana, na 2.^a ele significa o total das freqüências com valores menores ao da mediana; quanto a N_2 , ao passo que na 1.^a formula representa o numero de observações da classe em que se acha a mediana, na 2.^a, ele significa o total das freqüências com valores superiores a esta medida de posição.

Com esta consideração preliminar, podemos proseguir afirmando que, evidentemente, a condição necessaria e suficiente para que o valor da mediana coincida com um limite superior de classe é dada pela igualdade abaixo, tirada da formula da mediana:

$$\frac{\left(\frac{N}{2} - N_1\right)}{N_2} = n, \text{ donde se deduz:}$$

$$\frac{\frac{N}{2} - N_1}{N_2} = 1 \text{ e portanto } N_1 = N_2 - \frac{N}{2}$$

Vejamos como interpretar este resultado; desde que N_1 representa o total das freqüências das classes com valores inferiores á classe em que se encontra a mediana, e N_2 a freqüência da classe em que se acha a mediana, e desde que todas estas N_2 observações valendo o ponto médio da classe são inferiores á mediana, que vale o limite superior de tal classe, podemos concluir: a) a soma de N_1 e N_2 nada mais é do que

o total das observações com valor inferior á mediana, isto é, nada mais é do que o simbolo N_1 da formula do desvio médio; b) o valor de N_1 da formula do desvio médio precisa ser $\frac{N}{2}$ para que seja possível a coincidência da mediana com um limite superior de classe; c) evidentemente sendo N_2 da formula do desvio médio igual a $N - N_1$, o seu valor será para que a hipótese permaneça: $N_2 = N - N_1 = N - \frac{N}{2} = \frac{N}{2}$; d) desde que N_1 e N_2 na formula do desvio médio são iguais, a sua diferença é 0 e o unico factor suscetivel de vir influenciar o resultado é $\Sigma + (f \cdot \xi)$ pois N é constante qualquer que seja a origem adotada.

Passemos portanto á demonstração da invariabilidade de $\Sigma + (f \cdot \xi)$; adotando-se uma ou outra das origens; para tal, imaginemos a distribuição abaixo, na qual, $C_0, C_1, \dots, C_{(n-1)}$ representam n classes com valores inferiores á classe C_0 , (suposta ser aquela cujo limite superior coincide com a mediana), e $C_1, C_2, \dots, C_{(n-1)}, C_m$ outras classes de valores superior a C_0 .

Se a classe C_0 é tomada como origem dos desvios, temos evidentemente os valores de $+\xi$ dados pela 3.^a coluna e os de $+f \cdot \xi$ pela 4.^a.

CLASSES	Frequências (f)	Desvios a partir de C_0 + (ξ)	Produtos + ($f \cdot \xi$)
C_n	f_n	n	$f_n \cdot n$
$C_{(n-1)}$	$f_{(n-1)}$	(n-1)	$f_{(n-1)} \cdot (n-1)$
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
C_1	f_1	1	$f_1 \cdot 1$
C_0	f_0	0	$f_0 \cdot 0$
C'_1	f'_1	1	$f'_1 \cdot 1$
C'_2	f'_2	2	$f'_2 \cdot 2$
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
$C'_{(m-1)}$	$f'_{(m-1)}$	(m-1)	$f'_{(m-1)} \cdot (m-1)$
C'_m	f'_m	m	$f'_m \cdot m$

Pois bem, se a origem passa de C_0 para a classe C'_1 , cujo limite inferior coincide com a mediana, as $n + 1$ primeiras classes têm os seus $+ \xi$ aumentados, cada qual, de uma unidade, e os seus $+ f \cdot \xi$ transformados nos valores da 4.^a coluna do quadro abaixo:

CLASSES	Frequências (f)	Desvios a partir de C'_1 + (ξ)	Produtos + ($f \cdot \xi$)
C_n	f_n	$(n + 1)$	$f_n \cdot (n + 1)$
$C_{(n-1)}$	$f_{(n-1)}$	n	$f_{(n-1)} \cdot n$
.	.	.	.

CLASSES	Frequências (f)	Desvios a partir de C'_1 + (ξ)	Produtos + (f, ξ)
.	.	.	.
.	.	.	.
C_1	f_1	2	$f_1 \cdot 2$
C_0	f_0	1	$f_0 \cdot 1$

Ha, desta forma na soma dos produtos + $f \cdot \xi$ das (n+1) primeiras classes, pela mudança da origem um aumento igual a:

$$\begin{aligned}
 & [f_n \cdot (n+1) + f_{(n-1)} \cdot n + \dots + f_1 \cdot 2 + f_1 \cdot 1 + f_0 \cdot 1] = \\
 & [f_n \cdot n + f_{(n-1)} \cdot (n-1) + \dots + f_1 \cdot 1 + f_0 \cdot 0] = \\
 & f_n \cdot n - f_n \cdot n + f_n + f_{(n-1)} \cdot n - f_{(n-1)} \cdot (n-1) + f_{(n-1)} \cdot n - f_{(n-1)} \cdot (n-1) \\
 & + \dots + 2 f_1 - f_1 + f_0 = f_n + f_{(n-1)} + \dots + f_1 + f_0
 \end{aligned}$$

isto é, um aumento igual á soma das frequências com valores inferiores ao da mediana, portanto um aumento igual a N_1 , que como já sabemos vale $\frac{N}{2}$.

Em contraposição a este aumento as m classes restantes com a antiga origem, têm, pela mudança desta origem, os seus + ξ diminuídos de uma unidade e os seus + $f \cdot \xi$ transformados nos resultados da 4.^a coluna do quadro seguinte:

CLASSES	Frequências (f)	Desvios a partir de C' ₁ + (f. ξ)	Produtos + (f. ξ)
C' ₁	f' ₁	0	f' ₁ · 0
C' ₂	f' ₂	1	f' ₂ · 1
.	.	.	.
.	.	.	.
.	.	.	.
C' (m - 1)	f' (m - 1)	m-2	f' (m-1) · (m-2)
C' m	f' m	m-1	f' m · (m-1)

A diminuição do valor de $\sum + f. \xi$ nas m classes que completam a distribuição com a adoção da nova origem, será pois de:

$$\begin{aligned}
 & [f'_1 \cdot 1 + f'_2 \cdot 2 + \dots + f'_{(m-1)} \cdot (m-1) + f'_m \cdot m] - \\
 & [f'_1 \cdot 0 + f'_2 \cdot 1 + \dots + f'_{(m-1)} \cdot (m-2) + f'_m \cdot (m-1)] = \\
 & f'_1 + 2 f'_2 - f'_2 + \dots + f'_{(m-1)} \cdot m - f'_{(m-1)} \cdot m - f'_{(m-1)} \\
 & + 2 f'_{(m-1)} + f'_m \cdot m - f'_m \cdot m + f'_m = \\
 & f'_1 + f'_2 + \dots + f'_{(m-1)} + f'_m
 \end{aligned}$$

isto é, uma diminuição igual a soma das observações com valores superiores aos da mediana, ou seja uma diminuição igual a N_2 , que como já sabemos vale $\frac{N}{2}$.

Como vemos, a mudança da origem provocou na soma dos produtos $+ f. \xi$ um aumento e

uma diminuição iguais a $\frac{N}{2}$, ou em outras palavras ela não alterou o resultado. Demonstrada desta forma a invariabilidade do valor de $\Sigma + (f. \xi)$ podemos portanto concluir que quando a mediana coincide com o limite de classe, a origem a se adotar para o calculo rápido classico do desvio médio em função da mediana pode ser dupla.