MEDIDAS DE TENDÊNCIA CENTRAL

Prof. ALFREDO DE OLIVEIRA PEREIRA

(*Súmula da 4.ª aula do Curso de Extensão Universitária para Técnicos de Educação e Orientadores Educacionais, organizado pelos Cursos "Magister", sob o patrocínio da Secretaria de Educação e Cultura de São Paulo*)

UNIDADE IV — MEDIDAS DE TENDÊNCIA CENTRAL: 1 — Noção geral sôbre os promédios. 2 — Média aritmética. 3 — Mediana. 4 — Moda. 5 — Relação empírica de Pearson. 6 — Emprêgo dos promédios.

1 — NOÇÃO GERAL SÔBRE OS PROMÉDIOS

Quem examinar uma distribuição de freqüência, aproximadamente simétrica, notará que os valores de freqüência tendem a se agrupar no centro da distribuição, dispersando-se gradativamente no sentido das classes extremas que, em geral, acusam baixas freqüencias. Verifica-se, portanto, que os valores centrais da distribuição reúnem maior número de casos, isto é, apresentam maior probabilidade de ocorrência que os seus valores limites; logo, o conhecimento exato dêsses valores típicos da distribuição fornece valiosas indicações do fenômeno que representa.

A *média aritmética*, a *mediana* e a *moda* também conhecidas pelo têrmo genérico de *promédios* são as principais medidas de tendência central, de locação ou de posição. Em geral, destinam-se essas estatísticas a representar todos os valores de uma série. Assim, quando se calcula a média aritmética de uma série, procura-se substituir todos os seus têrmos por um único valor que bem a represente. Naturalmente, essa substituição implica a presunção de que todos os valores da série considerada são equivalentes a essa estatística.

Além da média aritmética, a Estatística Geral se utiliza da *média geométrica* e da *média harmônica* de uso restrito no campo da Psicologia e da Educação. Por êsse motivo não constará do presente estudo qualquer referência, além desta, sôbre essas últimas es-

tatísticas; conseqüentemente, dirá respeito à média aritmética tôda e qualquer indicação que fizermos, empregando o têrmo genérico de *média.*

2 — MÉDIA

Média (m) de uma série de valores é o quociente da divisão de sua soma pelo número de valores:

$$m = \frac{\Sigma(X)}{n} \qquad (3)$$

em que Σ é o sinal de somatório; (X), todos os valores que integram a série, e n, o número dêles. A fórmula enunciada parece suficientemente clara para demonstrar como se calcula a média de uma série de valores.

Exemplo I: Sejam os números
31 37 39 42 56

cuja soma é 205. Dividindo-se 205 pelo número de valores (5), obtém-se a média 41, que é o número mais representativo da série, isto é, a constante pela qual poderão ser substituidos todos os seus valores sem alteração do somatório. Realmente, se somarmos cinco parcelas iguais a 41 (ou multiplicarmos a média pelo número de valores da série), obteremos a soma inicial (205).

Em se tratando de valores grupados, é evidente que cada um dêles deverá ser, prèviamente, multiplicado pela respectiva freqüência, a fim de a média resultante expressar todos os valores da série. Nesse caso, a fórmula de cálculo passará a ser:

$$m = \frac{\Sigma(Xf)}{\Sigma f} \qquad (4)$$

Exemplo II:

Notas	*Freqüência*
20	3
30	5
40	6
50	8
60	11
70	4
80	2
90	1
	40

Chamando as notas obtidas de X, as freqüências de f e o produto dêsses valores de Xf:

X	f	Xf
20	3	60
30	5	150
40	6	240
50	8	400
60	11	660
70	4	280
80	2	160
90	1	90
	40	2040

Da aplicação da fórmula 4 resulta:

$$m = \frac{2040}{40} = 51$$

As freqüências, neste caso, representam *pesos de ponderação,* daí resultando a denominação de *média ponderada* para a média de uma série de valores grupados. A rigor, tôda média é ponderada, uma vez que resulta do quociente

da divisão de *todos* os valores da série pelo número dêles, fato que se verifica, igualmente, no cálculo de uma série simples de valores, não se justificando, em conseqüência, a denominação corrente de *média aritmética simples.*

Não difere o critério de cálculo para a determinação da média de uma série de valores tabulados em classes. Seja a distribuição-modêlo que vimos estudando. A marcha do cálculo é idêntica à do exemplo precedente, sendo as classes representadas, respectivamente, por seus pontos médios, conforme se poderá verificar no seguinte quadro:

X	Pm	f	fPm
0 — 9	4,5	3	13,5
10 — 19	14,5	6	87,0
20 — 29	24,5	10	245,0
30 — 39	34,5	15	517,5
40 — 49	44,5	16	712,0
50 — 59	54,5	29	1580,5
60 — 69	64,5	12	774,0
70 — 79	74,5	4	298,0
80 — 89	84,5	3	253,5
90 — 99	94,5	2	189,0
Σ	—	100	4670,0

$$m = \frac{4670}{100} = 46{,}7$$

Até aqui temos aplicado o método comum ou aritmético de calcular a média. Além dêsse método, um outro de grande utilidade nas elaborações estatísticas é empregado. Baseia-se êsse método de cálculo na principal propriedade da média: a soma algébrica dos afastamentos dos valores da série em relação à verdadeira média é nula.

Consideremos, por exemplo, a mesma série de valores do exemplo I. As diferenças, afastamentos, desvios ou discrepâncias (D) dos valores da série em relação à média são:

$$\begin{aligned} X - m &= D \\ 31 - 41 &= -10 \\ 37 - 41 &= -4 \\ 39 - 41 &= -2 \\ 42 - 41 &= +1 \\ 56 - 41 &= +15 \\ &\quad\overline{} \\ &\quad 0 \end{aligned}$$

A soma algébrica dos desvios é nula. E' evidente que não obteríamos o mesmo resultado se, ao invés da média verdadeira 41, tomássemos outra qualquer como ponto de origem dos desvios. Admitamos, só para exemplificar, que essa média, tomada arbitràriamente, fôsse 39.

Os desvios em tôrno dela seriam os seguintes:

$$\begin{aligned} X - m_o &= D \\ 31 - 39 &= -8 \\ 37 - 39 &= -2 \\ 39 - 39 &= 0 \\ 42 - 39 &= +3 \\ 56 - 39 &= +17 \\ &\quad\overline{} \\ &\quad +10 \end{aligned}$$

Tendo em vista a propriedade fundamental da média, verifica-se que 39 não é a média verdadeira da série considerada, porque, necessàriamente, o somatório dos desvios deveria ser nulo e não + 10. O excesso dos valores positivos obtidos na soma dos desvios indica que a média adotada como origem (39) é inferior à média verdadeira, devendo ser acrescida,

conseqüentemente, de determinada quantidade que a corrija.

Essa quantidade (k), denominada *coeficiente de correção*, é o resultado da divisão do somatório algébrico dos desvios pelo número dêles, isto é, a sua média.

$$k = \frac{\pm \Sigma(D)}{n} \qquad (5)$$

$$k = \frac{+\ 10}{5} = +\ 2$$

Adicionando o coeficiente de correção (+ 2) à média de origem arbitrária (39), obtém-se a média verdadeira da série (39 + 2 = 41).

Do que foi exposto se pode concluir que a média de uma série de valores poderá ser também obtida, partindo-se de uma *média auxiliar, fictícia* ou de origem arbitrária (m_o) à qual se adicionará algèbricamente o coeficiente de correção:

$$m = m_o \pm k$$

ou

$$m = m_o \pm \frac{\Sigma(D)}{n} \qquad (6)$$

Nas elaborações estatísticas êste é o critério comum de calcular a média, uma vez que oferece grande simplificação aos cálculos, especialmente quando se trata de uma série de valores fracionários ou de freqüências elevadas. Além disso, facilita a determinação de outras estatísticas como veremos adiante.

O cálculo para uma série de valores grupados não oferece dificuldade. Seja a série já citada no exemplo II em que encontramos a média 51. O quadro seguinte sintetiza a marcha do cálculo em função da média auxiliar 60.

X	f	D	Df
20	3	−40	−120
30	5	−30	−150
40	6	−20	−120
50	8	−10	− 80
60	11	0	0
70	4	+10	+ 40
80	2	+20	+ 40
90	1	+30	+ 30
Σ	40	—	−360

$$m = 60 - \frac{360}{40} = 51$$

O produto Df é indispensável por se tratar de uma série de valores grupados, indicando a multiplicação dos desvios pelas respectivas freqüências. A fórmula do cálculo, pràticamente, é a mesma:

$$m = m_o \pm \frac{\Sigma(Df)}{\Sigma f} \qquad (7)$$

Observe-se que todos os desvios são múltiplos de intervalo de classe (10) existente entre as notas. Isto sempre ocorre quando a série de valores apresenta intervalo constante e se adota como média auxiliar um dos seus valores.

Ora, dividindo-se os desvios pelo intervalo, o cálculo fica consideràvelmente simplificado, uma vez que os quocientes formam a série natural dos números inteiros, positivos e negativos, a partir da média de origem arbitrária a que

corresponde, naturalmente, o desvio zero. Em conseqüência, o coeficiente de correção fica dividido pelo intervalo de classe, de vez que o produto Df sofre a mesma alteração, sendo necessário, portanto, corrigi-lo, no resultado, com a operação inversa, transformando-se a fórmula inicial em

$$m = m_o \pm \frac{\Sigma(df)}{\Sigma f} \cdot h \qquad (8)$$

A média do exemplo II calculada por êsse método não se altera:

X	f	D	$d = \frac{D}{h}$	df
20	3	−40	− 4	−12
30	5	−30	− 3	−15
40	6	−20	− 2	−12
50	8	−10	− 1	− 8
60	11	0	0	0
70	4	+10	+ 1	+ 4
80	2	+20	+ 2	+ 4
90	1	+30	+ 3	+ 3
Σ	40	—	—	−36

$$m = 60 - \frac{36}{40} \times 10 = 51$$

A simplificação dos cálculos mais se evidencia na determinação da média de uma distribuição de freqüência em classes. A marcha do cálculo poderá ser acompanhada no quadro abaixo (Exemplo III), que reproduz a distribuição-modêlo que vimos examinando, com as seguintes observações sôbre as colunas:

Coluna X — Classes de distribuição

Coluna Pm — Pontos médios das classes

Coluna f — Freqüência das classes

Coluna D — Diferenças Pm — m, isto é, afastamentos dos pontos médios em relação à média de origem arbitrária adotada (54,5).

Coluna d — Desvios resultantes da divisão dos desvios da coluna D pelo intervalo da classe =

$$d = \frac{D}{h}$$

Coluna df — Produto das freqüências (coluna f) pelos desvios (coluna d)

$$m_o = 54{,}5$$

X	*Pm*	*f*	*D*	*d*	*df*
0 — 9	4,5	3	−50	− 5	−15
10 — 19	14,5	6	−40	− 4	−24
20 — 19	24,5	10	−30	− 3	−30
30 — 39	34,5	15	−20	− 2	−30
40 — 49	44,5	16	−10	− 1	−16
50 — 59	54,5	29	0	0	0
60 — 69	64,5	12	+10	+ 1	+12
70 — 79	74,5	4	+20	+ 2	+ 8
80 — 89	84,5	3	+30	+ 3	+ 9
90 — 99	94,5	2	+40	+ 4	+ 8
Σ	—	100	100	—	−78

$$m = 54{,}5 + \frac{-78}{100} \times 10 = 46{,}7$$

Adotado êsse método de simplificação, pode-se prescindir da coluna D, por desnecessária ao cálculo.

3 — MEDIANA

Mediana (md) de uma série de valores ordenados é o seu valor central, aquêle valor que divide a série em duas partes iguais. Se o número de valores é par, a mediana é, por convenção, a média dos dois valores centrais. Exemplo IV — Na série ordenada

31 37 39 42 56

a mediana é o terceiro valor (39), uma vez que ocupa a posição central da série. Generalizando o conceito de mediana, podemos dizer que ela, numa série de valores ordenados, crescente ou decrescentemente é o valor de ordem

$$P = \frac{n+1}{2} \qquad (9)$$

Cumpre esclarecer, desde logo, que a mediana não pode ser confundida com a média e, só excepcionalmente, os dois valores se igualam.

Exemplo V — Na série:
2 23 39 85 101

a mediana é a mesma da série anterior, isto é, 39. Enquanto na série do exemplo IV a média é igual a 41, nesta ela tem o valor de 50. Conclui-se do exposto que a média é mais sensível que a mediana, uma vez que do seu cálculo participam todos os valores numéricos da série, o mesmo não se verificando com a mediana, cujo valor independe de grandeza dos demais valores da série sòmente sendo influenciada pelo número dêles.

A mediana de um conjunto de valores ordenados pode ser obtida por simples inspeção do valor que, no conjunto, ocupa a posição

$$\frac{n + 1.}{2}$$

Se a série contiver um número ímpar de valores, 7, por exemplo, a mediana será exatamente o valor que ocupar a 4.ª ordem:

$$P = \frac{7+1}{2} = 4$$

No caso de uma série par de valores, a ordem

$$\frac{n + 1}{2}$$

será um ponto intermediário entre dois valores centrais: logo, a mediana corresponderá à média dêsses dois valores centrais. Exemplo VI:

37 35 31 27 25 20 19 17

O valor de ordem

$$\frac{n + 1}{2}$$

da série é o valor que corresponde à posição

$$P = \frac{8+1}{2} = 4{,}5$$

Essa posição não está representada por qualquer número da sé-

rie, ficando num ponto entre os valores 27 e 25, ou seja, a média dêsses valores, 26.

O critério anterior também se aplica a uma série de valores grupados. Para facilidade da localização do valor a que corresponde a freqüência de ordem

$$\frac{n + 1}{2}$$

convém, prèviamente, calcular as freqüências acumuladas.

Exemplo VII:

X	f	fa
10	3	3
30	5	8
45	6	14
50	6	20
60	13	33
75	4	37
80	2	39
100	1	40

O valor de posição, neste caso, é:

$$P = \frac{40 + 1}{2} = 20{,}5$$

e a mediana será o valor da variável que ocupar a posição 20,5. Observando-se as freqüências acumuladas, verifica-se que as freqüências de ordem 15, 16, 17, 18, 19 e 20 correspondem ao valor 50 da variável, e que as freqüências de ordem 21, 22, 23, 24, 25, 26, 27, 28, 29 e 30, correspondem ao valor 60. Como se deseja a posição 20,5, intermediária entre os valores de ordem 20 e 21, segue-se que a mediana será, igualmente, o valor compreendido entre 50 e 60, melhor, a sua média, 55. Calculemos, agora, a mediana da série de valores grupados do exemplo II:

X	f	fa
20	3	3
30	5	8
40	6	14
50	8	22
60	11	33
70	4	37
80	2	39
90	1	40

O valor de posição, neste caso, também é 20,5. Essa freqüência de ordem

$$\frac{n + 1}{2}$$

corresponde ao valor 50 da variável, da mesma forma que as freqüências de ordem 15, 16, 17 18, 19, 20, 21 e 22; conseqüentemente, a mediana da série é 50.

Em se tratando de valores grupados em classes, a mediana poderá ser calculada pela fórmula de King:

$$md = l + \frac{h(P - f_a)}{f} \qquad (10)$$

em que os símbolos têm a seguinte significação:

P — valor de posição, ou seja, freqüência de ordem

$$\frac{n}{2}$$

l — limite real inferior da classe mediana, isto é, da classe a que corresponde a freqüência de ordem P

h — intervalo de classe da distribuição

fa — freqüência acumulada da classe que precede imediatamente a classe mediana

f — freqüência simples da classe mediana

O quadro seguinte, relativo à distribuição-modêlo, enuncia os elementos integrantes da fórmula 10:

X	f	fa
0 — 9	3	3
10 — 19	6	9
20 — 29	10	19
30 — 39	15	34
40 — 49	16	50
50 — 59	29	79
60 — 69	12	91
70 — 79	4	95
80 — 89	3	98
90 — 99	2	100
Σ	100	—

Comecemos por calcular a freqüência (P) de ordem

$$\frac{n}{2}$$

$$P = \frac{100}{2} = 50$$

A freqüência de ordem 50 integra a freqüência acumulada da classe 40-49 (Classe mediana). O limite real inferior dessa classe é 39,5 e o intervalo de classe da distribuição é 10. A freqüência acumulada da classe imediatamente anterior à mediana é 34, sendo a freqüência simples da classe mediana 16; logo:

$l = 39{,}5$ h $= 10$ fa $= 34$ f $= 16$

Substituindo os símbolos da fórmula 10 pelos respectivos valores:

$$md = 39{,}5 + \frac{10(50 - 34)}{16} = 49{,}5$$

4 — MODA

Moda, norma ou *valor prevalente* (mo) de uma distribuição de freqüência de tendência central é o valor da variável que ocorre com maior freqüência; nas representações gráficas, corresponde à ordenada máxima. As distribuições que não apresentam tendência central, a rigor, não têm moda, conquanto possam apresentar uma freqüência máxima.

Numa série de valores grupados em classes ou não, a moda pode ser determinada por simples inspeção das freqüências. Exemplo VIII: Distribuição do número de famílias segundo o número de filhos.

N.º de filhos	*N.º de famílias*
0	180
1	235
2	347
3	564
4	308
5	211
6	195
7	98
8	10

A moda dessa distribuição é *3 filhos* porque é o valor da variável a que corresponde a freqüência máxima (564). Na amostra examinada, a moda é o valor de maior probabilidade de ocorrência.

Nas distribuições em classes, a moda poderá ser obtida pela fórmula de Czuber:

$$mo = l + \frac{h(f_m - f_{m-1})}{2f_m - (f_{m-1} + f_{m+1})} \qquad (11)$$

em que os símbolos significam, respectivamente:

mo — moda

l — limite real inferior da classe modal

h — intervalo de classe

fm — freqüência máxima

fm—1 — freqüência imediatamente anterior à freqüência máxima

fm+1 — freqüência imediatamente posterior à freqüência máxima.

A moda da distribuição-modêlo é igual a 53,8.

5 — RELAÇÃO EMPÍRICA DE PEARSON

Nas distribuições rigrosamente simétricas, em que se verifica uma exata repartição de valores de freqüência em tôrno da freqüência máxima, a média, a mediana e a moda são iguais. Conclui-se dessa observação experimental que, quanto mais próxima do tipo normal fôr uma dada distribuição, tanto mais próximos estarão seus valores típicos da ordenada máxima. Partindo dêsse princípio, Pearson propôs a expressão:

$$mo = 3md - 2m \qquad (12)$$

para definir a moda em função dos outros dois valores de tendência central. Essa fórmula, conhecida como *relação empírica de Pearson*, fornece, aproximadamente, o valor da moda, e das outras medidas de tendência central, desde que a distribuição seja mais ou menos normal.

6 — EMPRÊGO DOS PROMÉDIOS

O emprêgo das medidas de tendência central que acabamos de estudar depende, essencialmente, da natureza da variável examinada e da forma de sua distribuição. Se uma série apresenta certa regularidade quanto à distribuição de seus valores, a média deve ser a preferida, porque ela reflete a magnitude de todos os valores da série; se, todavia, não se verificar entre os valores da série pequena dispersão, discrepando consideràvelmente os valores extremos, é conveniente preferir-se a mediana ou a moda, que não sofrem a influência dêsses valores extremos. Numa distribuição de classes extremas indefinidas, a média nada expressará, devendo-se dar preferência, neste caso, à mediana ou à moda. A moda é geralmente indicada nos casos em que se deseja conhecer o valor típico ou prevalente da distribuição, qualquer que seja a sua forma.

A rigor, portanto, não existem indicações quanto à utilização preferencial desta ou daquela medida

de tendência central, devendo-se evitar o emprêgo de uma ou de outra nos casos acima apontados. Dada a facilidade de tratamento algébrico a que está sujeito e a generalidade de seu emprêgo na determinação das demais constantes estatísticas, a média é por natureza o valor típico de emprêgo mais recomendado.