Ministério da Saúde
Instituto Nacional de Câncer

MEDIDAS DE POSIÇÃO E VARIAÇÃO SEPARATRIZES

MANUAL DIDÁTICO NÚMERO 2

INCA
INSTITUTO NACIONAL DE CÂNCER
MINISTÉRIO DA SAÚDE

Manual didático nº 2

MEDIDAS DE POSIÇÃO E VARIAÇÃO - SEPARATRIZES

Texto: Pedro Carvalho Rodrigues
Divisão de Ensino e Divulgação Científica
Setor de Pós Graduação

ÍNDICE

1. INTRODUÇÃO

Identificamos com certa freqüência algumas características, as quais designamos por variáveis, tais como, a pressão arterial sistólica, a altura de indivíduos, o peso de crianças, o diâmetro de uma árvore. É também observado que quantitativamente estas variáveis apresentam valores que se caracterizam pela variação.

Nestes casos, conceituam-se estes tipos de variáveis, como quantitativas contínuas, as quais ensejam ao cálculo de medidas de posição e de variação e que serão descritas e apresentadas a seguir.

2. MEDIDAS DE POSIÇÃO

Quando obtemos um conjunto de dados numéricos, provenientes de uma amostra ou população, quanto a uma variável contínua, utilizamos medidas que de certa forma possam resumir todo o conjunto. Quando estas medidas pertencem a uma população, são identificadas como PARÂMETROS e quando pertencem a uma amostra são consideradas como ESTIMATIVAS.

As medidas de posição ou tendência central mais usadas para representar um conjunto de dados quantitativos contínuos, são a média aritmética, a mediana e a moda.

2.1 - MÉDIA ARITMÉTICA SIMPLES

Para obtenção da média aritmética procedemos ao somatório dos valores de todos os dados e dividimos este somatório pelo número de observações do conjunto.

Exemplo: Peso em gramas de camundongos médios da raça Reedly, com 20 dias de idade.

50, 62, 65, 74, 55, 66, 58, 80, 72, 74, 60, 64

A média aritmética dos dados é:

$$\frac{50+62+65+74+55+66+58+80+72+74+60+64}{12} = \frac{780}{12} = 65$$

Assim, o peso médio dos camundongos é 65 gramas.

O símbolo usado para a média aritmética é μ para a população e $\bar{X}$ para a amostra, sendo calculada na amostra pela expressão:

$$\boxed{\bar{X} = \frac{\sum x}{n}}$$

em que:

$\bar{X}$ = *média aritmética*

$\sum x$ = *somatório dos valores*

n = *número de observações*

2.2 - Média Aritmética de Dados Agrupados

Quando obtem-se uma série de valores e estes formam uma distribuição de freqüências, a determinação da média aritmética dos dados, levará em consideração a freqüência com que os valores se apresentam.

Dada a distribuição:	Variável (X)	Freqüência (f)
	X_1	f_1
	X_2	f_2
	X_3	f_3
	:	:
	X_n	f_n

A expressão a usar será:

$$\bar{X} = \frac{\Sigma xf}{n} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots f_n}$$

Assim podemos calcular a média de idade de indivíduos atendidos no ambulatório do Posto de Saúde no Município de Rio Bonito - RJ.

Idade (X)	Freqüência (f)
20	5
21	6
22	8
23	10
24	7
25	6
26	6

Nestes casos de distribuição de freqüências simples, podemos também calcular freqüências relativa e acumulada.

Idade (X)	Freqüência simples	Xf	Freqüência relativa	Freqüência acumulada
20	5	100	5/48	5
21	6	126	6/48	11
22	8	176	8/48	19
23	10	230	10/48	29
24	7	168	7/48	36
25	6	150	6/48	42
26	6	156	6/48	48
Total	48	1106	1	-

A média aritmética da idade será:

$$\bar{X} = \frac{\Sigma xf}{n} = \frac{1.106}{48} = 23,0$$

Portanto a média de idade do conjunto corresponde a 23 anos.

2.3 - Média Aritmética de Dados Agrupados

Em muitas das vêzes, os dados obtidos numa análise não são apresentados com seu valor individual. Nestas situações, procede-se como no caso anterior (2.2) em que temos um valor representativo e este então será conseguido, utilizando-se como representante do intervalo o seu ponto médio.

Assim, se tivermos a distribuição de freqüências abaixo:

Níveis de Anticorpos	Freqüência (f)
0 ⊢5	4
5 ⊢10	3
10 ⊢15	5
15 ⊢20	6

Para determinarmos a média aritmética necessitamos utilizar a expressão: $\bar{X} = \frac{\Sigma xf}{n}$, em que:

$\bar{X}$ = média aritmética

f = freqüência

x = valor representativo da cada intervalo (ponto médio)

n = número de observações.

Então:

Níveis de Anticorpos	Valor representativo (x)	Freqüência	x.f
0 |-5	2,5	4	10,0
5 |-10	7,5	3	22,5
10 |-15	12,5	5	62,5
15 |-20	17,5	6	105,0
Total	-	18	200,0

$$\bar{X} = \frac{\Sigma xf}{n} = \frac{200}{18} = 11,1$$

Logo, o nível de anticorpos apresenta valor médio de 11,1.

2.4 - MÉDIA GEOMÉTRICA

Quando dispomos de valores que apresentam acentuada variação podemos utilizar a média geométrica, que expressa um valor representativo do conjunto.

A média geométrica é determinada através da expressão:

$$M.G. = \sqrt[n]{X_1 . X_2 . X_3 .... X_n}$$

em que n é o tamanho da amostra e X_1, X_2, X_3, ...X_n são os valores individuais da amostra.

Por exemplo, a média geométrica dos valores 10, 18, 26, 40, será:

$$M.G. = \sqrt[4]{10 . 18 . 26 . 40} = \sqrt[4]{187.200}$$

Na prática a média geométrica é calculada utilizando-se logarítmos, e assim teríamos, aplicando-se a propriedade de uma raiz quadrada:

$$\log(M.G.) = \frac{\log 187.200}{4}$$

$$\log(M.G.) = \frac{5,2723}{4}$$

$\log(M.G.) = 1,318075$, o seu antilogaritmo será: $M.G. = 20,8$

2.5 - Mediana

Trata-se de uma medida de posição, usada para indicação do centro da distribuição dos dados. A mediana de um conjunto de observações é o valor que fica exatamente no meio da série, quando os valores já estiverem ordenados.

Esta medida é útil quando os dados do conjunto apresentam uma distribuição assimétrica.

Para seu cálculo:

a) Ordenamos os valores;

b) Verificamos o total de valores (n);

c) Se o número de observações é ímpar, considera-se o valor mediano igual a $\frac{n+1}{2}$

Exemplo: 4 - 12 - 15 - 20 - 70

sendo $n = 5$, faz-se $\frac{5+1}{2} = \frac{6}{2} = 3^{o}$

Portanto, o 3º elemento é o valor mediano, neste caso igual a 15.

d) Se o número de observações for par, consideramos o valor mediano igual a média aritmética dos valores centrais:

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

Exemplo: 7, 15, 18, 22, 30, 90

Sendo n = 6, procede-se fazendo $\frac{6}{2} = 3$ que corresponde ao 3º e $\frac{6}{2} + 1 = 3 + 1 = 4$, ou seja o 4º.

A média aritmética dos 3º e 4º valores, ou seja 18 e 22 será igual a 20. Portanto, o valor mediano corresponde a 20. A mediana é também chamada de percentil 50 .

Pressupõe-se que o cálculo da mediana indica que os dados apresentem uma distribuição assimétrica e, então o seu valor deve se acompanhar dos valores mínimo e máximo. Logo, nos exemplos mostrados, apresentaríamos:

$$M_e = 15\,(4 - 70) \quad \text{e} \quad M_e = 20\,(7 - 90)$$

Para os casos de dados agrupados o cálculo da mediana é realizado através da expressão :

$$M_e = L_i + \frac{\frac{n}{2} - F_a}{F_{Me}} \cdot h_{Me}$$

Em que: L_i *: limite inferior da classe que contém a mediana;*

n *: número de observações do conjunto;*

F_a *: soma das freqüências das classes anteriores à classe da mediana;*

F_{Me} *: freqüência simples da classe da mediana;*

h_{Me} *: amplitude da classe que contém a mediana;*

Exemplo: distribuição de freqüências:

Idade (anos)	Freqüência Simples	Freqüência Acumulada
10 ⊢ 20	10	10
20 ⊢ 30	15	25
30 ⊢ 40	20	45
40 ⊢ 50	15	60
50 ⊢ 60	8	68
60 ⊢ 70	2	70
Total	70	-

Considerando-se que temos 70 observações:

Classe da mediana:	30 ⊢ 40
Limite inferior:	30
Nº de observações:	70
Freqüência acumulada anterior à classe mediana:	25
Freqüência simples da classe mediana:	20
Amplitude de classe:	10

$$M_e = L_i + \frac{\frac{n}{2} - F_a}{F_{Me}} \cdot h_{Me}$$

$$M_e = 30 + \frac{\frac{70}{2} - 25}{20} \cdot 10$$

$\frac{70}{2} - 25$

$$M_e = 30 + \frac{35 - 25}{20} \cdot 10 = 30 + \frac{10}{20} \cdot 10$$

$$M_e = 30 + \frac{100}{20} \therefore M_e = 30 + 5 = 35$$

A mediana corresponde a 35 anos.

2.6 - MODA

Coerente com o termo moda. Corresponde ao valor mais freqüente do conjunto. Para uma série de valores torna-se simples a verificação de seu valor.

Exemplo: 5, 8, 7, 7, 15, 12, 7, 8, 9, 14, 7, 3, 7, 2, 7, 10, 7, 7

Percebe-se claramente a grande presença do valor 7, com 8 freqüências. Portanto, neste caso o valor 7 corresponde ao valor modal.

Para uma série de dados agrupados, temos que utilizar a expressão:

$$M_o = L_i + \frac{d_1}{d_1 + d_2} \cdot h$$

L_i = limite inferior da classe modal;
d_1 = diferença entre a freqüência da classe da moda e da classe anterior;
d_2 = diferença entre a freqüência da classe da moda posterior:
h = amplitude das classes.

Exemplo: distribuição de freqüências

Idade (anos)	Freqüência	
	Simples	Acumulada
10 ⊢ 20	10	10
20 ⊢ 30	15	25
30 ⊢ 40	20	45
40 ⊢ 50	15	60
50 ⊢ 60	8	68
60 ⊢ 70	2	70
Total	70	–

Classe da moda : 30 ⊢ 40

Limite inferior da classe modal : 30

$$d_1 = 20 - 15 = 5$$

$$d_2 = 20 - 15 = 5$$

$$h = 10$$

$$\boxed{M_0 = L_i + \frac{d_1}{d_1 + d_2} \cdot h}$$

$$M_0 = 30 + \frac{5}{5+5}\ 10 = 30 + \frac{5}{10}\ 10$$

$$M_0 = 30 + 5 \therefore M_0 = 35$$

A idade modal corresponde a 35 anos.

3 - MEDIDAS DE VARIAÇÃO

Considerando que os valores quantitativos contínuos apresentam como característica principal a variação, utilizamos algumas medidas que servem para expressar esta variabilidade.

3.1- AMPLITUDE TOTAL

Define-se amplitude total como a diferença entre o maior e menor valor do conjunto. Sendo um conjunto de dados igual a :

10, 18, 13, 20, 6, 40, 60, 25, 70

Verifica-se que o maior valor é 70 e o menor é 6. Portanto, a amplitude total é igual a 70 - 6 = 64.

3.2 - DESVIO MÉDIO

O desvio médio indica a distância média das diferenças individuais de cada valor do conjunto em relação a média aritmética.

A partir da média verifica-se que a soma das diferenças de cada valor do conjunto em relação a média aritmética é igual a 0 (zero).

Então, no caso do desvio médio, leva-se em consideração, apenas os valores absolutos das referidas diferenças. A partir destes valores absolutos, calculamos a média aritmética dos desvios.

Assim, para o conjunto: 7, 8, 5, 4, 10, 2, a média aritmética corresponde a 6. O desvio médio seria:

$$D.M. = \frac{|7-6| + |8-6| + |5-6| + |4-6| + |10-6| + |2-6|}{6}$$

$$D.M. = \frac{1+2+1+2+4+4}{6} = \frac{14}{6} \qquad D.M. = 2,3$$

3.3 - DESVIO PADRÃO

Na população, simbolizado por (σ) e na amostra (s), indica a variação ocorrida num conjunto de dados contínuos.

Para obtenção do desvio padrão, tomamos os seguintes passos:

a) verificam-se os desvios dos valores do conjunto em relação a média aritmética;

b) eleva-se estes desvios ao quadrado;

c) procede-se ao somatório dos desvios ao quadrado;

d) calcula-se a média da soma dos desvios ao quadrado;

e) extrai-se a raiz quadrada.

Nos casos de dados agrupados onde têm-se uma distribuição de freqüências, o cálculo do desvio padrão é realizado através da expressão:

$$s = \sqrt{\frac{\Sigma fx^2}{n} - \left(\frac{\Sigma fx}{n}\right)^2}$$

Para calcular o desvio padrão quando tivermos a distribuição de freqüências abaixo, faremos:

X	f
3	10
4	15
5	12
6	8
7	6
8	4

Há um total de 55 valores, agrupados em 6 categorias. Faz-se então:

X	f	fx	fx^2
3	10	30	90
4	15	60	240
5	12	60	300
6	8	48	288
7	6	42	294
8	4	32	256
Total	55	272	1.468

$$s = \sqrt{\frac{\Sigma fx^2}{n} - \left(\frac{\Sigma fx}{n}\right)^2} \qquad s = \sqrt{\frac{1468}{55} - \left(\frac{272}{55}\right)^2}$$

$$s = \sqrt{26,68 - 24,01} = \sqrt{2,67} \qquad s = 1,6$$

3.4 - VARIÂNCIA

Dado um conjunto de dados quantitativos contínuos, observamos um desvio de cada valor em relação a média aritmética. Estes desvios (+) e (-) são elevados ao quadrado. Posteriormente, procederemos ao somatório e finalmente calculamos a média aritmética destes valores encontrados. O resultado obtido corresponderá a variância.

Tendo-se os valores: 7, 3, 4, 6, 1, 6, 5, 8

$$\bar{X} = \frac{\Sigma X}{n} = \frac{40}{8} = 5$$

$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
7 - 5 = +2	4
3 - 5 = -2	4
4 - 5 = -1	1
6 - 5 = +1	1
1 - 5 = -4	16
6 - 5 = +1	1
5 - 5 = 0	0
8 - 5 = +3	9
Total	36

$$s^2 = \frac{\Sigma (X_i - \bar{X})^2}{n - 1} = \frac{36}{8 - 1} = \frac{36}{7} \qquad s^2 = 5,14$$

O denominador corresponde a n-1, pois trata-se de uma pequena amostra

($n \leq 30$). A variância é igual a 5,14.

3.5 - Coeficiente de Variação

Trata-se de uma medida que é também conhecida como variação relativa. Relaciona o desvio padrão e a média aritmética. Calcula-se através da expressão:

$$C.V. = \frac{s}{\bar{X}} \cdot 100$$

Não basta identificar num conjunto de dados quantitativos contínuos, o valor do desvio padrão para verificar a variabilidade do conjunto. É necessário, que além do desvio padrão, verifique-se também a média aritmética, para que se observe em torno de que valor estão situados os elementos do conjunto.

Assim, se identificamos um desvio padrão igual a 5, não há como concluir se é elevado ou não o seu valor. Isto dependerá dos valores originais do conjunto observado.

Exemplo:

Conjunto A	Conjunto B
$\bar{X} = 40$	$\bar{X} = 400$
$s = 5$	$s = 5$

Nos dois conjuntos A e B, os desvios padrões são iguais a 5. Entretanto, é facilmente entendido que o conjunto B, com média aritmética igual a 400, indica que os valores deste conjunto se situam em torno daquele valor. Já o conjunto A, com o mesmo valor do desvio padrão, igual a 5, tem como valor médio 40.

Portanto, concluímos que o conjunto B apresenta maior homogeneidade em seu conjunto, pois os coeficientes de variação correspondem a $C.V._A$ = 12,4% e $C.V._B$ = 1,25%, respectivamente.

Nas séries simples, a estimativa do desvio padrão é obtida através da expressão:

$$s = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n}} \quad \text{ou} \quad s = \sqrt{\frac{\Sigma (X_i - \bar{X})^2}{n}}$$

Onde: X = variáveis do conjunto
Σ = somatório
n = número de observações

Exemplificando: Calcular o desvio padrão dos valores 1, 2, 6, 12, 16, 18, 22

$$\bar{X} = \frac{\Sigma X}{n}, \quad \bar{X} = \frac{77}{7}, \quad \bar{X} = 11$$

X	$(X_i - \bar{X})$	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	1 - 11	- 1	100
2	2 - 11	- 9	81
6	6 - 11	- 5	25
12	12 - 11	+ 1	1
16	16 - 11	+ 5	25
18	18 - 11	+ 7	49
22	22 - 11	+ 11	121

$$\Sigma (X_i - \bar{X})^2 = 81 + 25 + 1 + 25 + 49 + 121 + 100$$
$$\Sigma (X_i - \bar{X})^2 = 402 \qquad n = 7$$

$$s = \sqrt{\frac{402}{7-1}} = \sqrt{67} \qquad s = 8,1$$

3.6 - ERRO PADRÃO DA MÉDIA

De um modo geral, o trabalho de pesquisa envolvendo amostragem, utiliza as variáveis em estudo usando amostra e nos casos em que utiliza dados quantitativos contínuos, faz uso da média aritmética para representar o conjunto de valores. Portanto a média aritmética é a medida de posição tradicionalmente usada.

Devemos considerar que se outras amostras fossem estudadas, estas apresentariam valores médios, diferentes do encontrado da primeira amostra. Há portanto, uma variação entre as médias provenientes das várias amostras.

Então, determina-se o erro padrão da média, que corresponde a esta variação mencionada.

Este erro padrão da média é obtido, com informações da primeira amostra considerada em estudo.

A expressão para cálculo do erro padrão da média corresponde a:

$$s(\bar{X}) = \frac{s}{\sqrt{n}} \quad \text{em que :}$$

$s(\bar{X})$ = erro padrão da média

s = desvio padrão da amostra

n = número de observações

Exemplo: Uma amostra de 100 indivíduos apresenta média de teor de hemoglobina igual a 13,8 mg e desvio padrão de 2 mg.

Verificar o erro padrão da média.

$$s(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \frac{2}{10} = 0{,}2\ mg$$

3.7 - ERRO PADRÃO DA PERCENTAGEM

Similar ao erro padrão da média, em trabalhos envolvendo amostragem, algumas variáveis da amostra fornecem dados quantitativos discretos, os quais são transformados em percentagem. Mas sabe-se que os percentuais encontrados em outras amostras, apresentam valores diferentes.

Portanto, também os valores percentuais sofrem variações considerando as várias amostras possíveis.

A variação entre estes valores percentuais é então chamada de erro padrão da percentagem.

A expressão para cálculo do erro da percentagem corresponde a:

$$s(p) = \sqrt{\frac{p \cdot q}{n}}$$

$s(p)$ = erro padrão da percentagem

p = percentual obtido na amostra

q = 100% - p

n = número de observações

Exemplo: Uma amostra com 100 indivíduos apresentou quanto a taxa de uréia uma percentagem de 20% de valores, acima dos valores normais.

Verificar o erro padrão da percentagem.

$$s(p) = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{20 \times 80}{100}} = \sqrt{\frac{1600}{100}} = \sqrt{16} = 4\%$$

O erro padrão da porcentagem corresponde a 4%.

4. SEPARATRIZES

Abordando as medidas de posição, fizemos referência com relação a mediana, que caracteriza-se por dividir o conjunto de observações em 2 conjuntos, com igual número de valores.

Apresentamos em seguida outras medidas, chamadas __separatrizes__, as quais realizam divisão dos valores do conjunto de acordo com o número de elementos, considerando-os já ordenados:

- quartis: divide em quatro partes;

- decis: divide em 10 partes;

- centis ou percentis: divide em 100 partes.

4.1 QUARTIS

O conjunto de dados é dividido em 4 partes iguais, quanto ao número de elementos.

$$\begin{array}{ccccccc} & \frac{n}{4} & & \frac{n}{2} & & \frac{3n}{4} & \\ | \text{———} & | & \text{———} & | & \text{———} & | & \text{———} | \\ & Q_1 & & Q_2 & & Q_3 & \\ & 1^{o}\ \text{quartil} & & 2^{o}\ \text{quartil} & & 3^{o}\ \text{quartil} & \end{array}$$

Exemplo: Um conjunto é composto dos seguintes valores, já ordenados:

7, 13, 18, 22, 40, 50, 74, 82, ou seja n = 8.

1^{o} quartil: $\frac{n}{4} = \frac{8}{4} = 2^{o}$, $\quad 2^{o}$ quartil: $\frac{n}{2} = \frac{8}{2} = 4^{o}$,

3^{o} quartil: $\frac{3n}{4} = \frac{24}{4} = 6^{o}$

O 2^{o} elemento do conjunto, igual a 13, corresponde ao 1^{o} quartil; o 4^{o} elemento 22 é o 2^{o} quartil ou mediana e o 6^{o} elemento, igual a 50, representa o 3^{o} quartil.

4.2 - Decis

Dividindo-se o conjunto em 10 partes, isto proporcionará a obtenção de 9 decis.

$\frac{n}{10}$	$\frac{2n}{10}$	$\frac{3n}{10}$	$\frac{4n}{10}$	$\frac{5n}{10}$	$\frac{6n}{10}$	$\frac{7n}{10}$	$\frac{8n}{10}$	$\frac{9n}{10}$
D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9

Exemplo: Um conjunto com 20 elementos, ordenados de forma crescente:

15, 18, 24, 30, 36, 39, 43, 44, 50, 53, 58, 60, 64, 67, 71, 73, 78, 80, 82, 86.

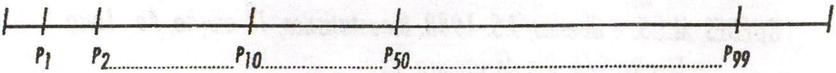
Identificar o 4^{o} decil.

$$4^{o} \text{ decil} = \frac{4n}{10} = \frac{4 \times 20}{10} = \frac{80}{10} = 8^{o}$$

Portanto, o 8^{o} elemento da série corresponde ao 4^{o} decil, ou seja, o valor 44.

4.3 - Centis ou Percentis

Resta então dividir o conjunto em 100 partes iguais, quanto ao número de valores. Surge então a última medida separatriz, que é o percentil. Encontramos então 99 percentis.



Exemplo: Um conjunto contendo 200 valores identificará o P_{50}, como o valor mediano, ou 2º quartil, como seu correspondente. Neste caso, o P_{50} será obtido fazendo-se $P_{50} = \frac{n}{2}$

Sendo n = 200, então o $P_{50} = \frac{200}{2} = 100^{o}$

Conclui-se que ordenados os 200 valores do conjunto o 100º valor será considerado o P_{50} (Mediana).

5. BIBLIOGRAFIA

COSTA NETO, P.L.O. 1977. Estatística, 2ª edição. Editora Edgard Blucher Ltda., São Paulo, SP.

GUEDES, M.L.S. e Guedes, J.S. 1988. Bioestatística, 1ª edição, Ao Livro Técnico S/A, Rio de Janeiro, RJ.

VIEIRA, S. 1980. Introdução à Bioestatística, 2ª edição, Editora Campus Ltda, Rio de Janeiro, RJ.

80025 75540

INCA

Ministério da Saúde
Instituto Nacional de Câncer

SÉRIE MANUAIS DIDÁTICOS
Divisão de Ensino e Divulgação Científica
Setor de Pós Graduação
Comitê de Padronização de Impressos

Texto - Pedro Carvalho Rodrigues