

Medidas de Dispersão¹

Introdução

Uma breve reflexão sobre as medidas de tendência central permite-nos concluir que elas não são suficientes para caracterizar totalmente uma sequência numérica.

Se observarmos as seqüências:

X: 10, 1, 18, 20, 35, 3, 7, 15, 11, 10.

Y: 12, 13, 13, 14, 12, 14, 12, 14, 13, 13.

Z: 13, 13, 13, 13, 13, 13, 13, 13, 13, 13.

concluiremos que todas possuem a mesma média 13.

No entanto, são seqüências completamente distintas do ponto de vista da variabilidade de dados.

Na seqüência Z não há variabilidade de dados.

A média 13 representa bem qualquer valor da série.

Na seqüência Y, a média 13 representa bem a série, mas existem elementos da série levemente diferenciados da média 13.

Na seqüência X existem muitos elementos bastante diferenciados da média 13.

Concluimos que a média 13 representa otimamente a seqüência Z, representa bem a seqüência Y, mas não representa bem a seqüência X.

Precisaremos, então, de medidas que avaliem a representatividade da média. Para isto usaremos as **medidas de dispersão**.

Observe que na seqüência Z os dados estão totalmente concentrados sobre a média 13. Não há dispersão de dados.

Na seqüência Y há forte concentração dos dados sobre a média 13, mas há fraca dispersão de dados. Já na série X há fraca concentração de dados em torno da média 13 e forte dispersão de dados em relação à média.

As principais medidas de dispersão absolutas são: **desvio médio, variância e desvio padrão**.

Variância (Va ou σ^2) e Desvio padrão (Dp ou σ)

A dispersão dos dados em relação à média de uma seqüência pode ser avaliada através dos desvios de cada elemento da seqüência em relação à média da seqüência

Uma forma de se conseguir que as diferenças $x_i - \bar{x}$ se tornem sempre positivas ou nulas é considerar o quadrado destas diferenças, isto é, $(x_i - \bar{x})^2$.

A variância é definida como sendo uma média aritmética dos quadrados dos desvios de cada elemento da série para a média da série

O desvio padrão é a raiz quadrada da variância.

¹ Baseada no material do Prof. Conrad E. Pinheiro.

Exemplo 1: Considere as notas 2 – 8 – 5 – 6 obtidas por 4 alunos, numa avaliação de Biologia, distribuídas na tabela abaixo. Calcule o desvio padrão.

Exemplo 2: Calcule o desvio padrão da série abaixo.

x_i	f_i
2	3
3	5
4	8
5	4
Total	

Exemplo 3: Calcule o desvio padrão da série abaixo.

Class e	Int. classe	f_i	
1	0 — 4	1	
2	4 — 8	3	
3	8 — 12	5	
4	12 — 16	1	
	Total		

Comentários:

No cálculo da variância, quando elevamos ao quadrado a diferença $(x_i - \bar{x})$, a unidade de medida da série fica também elevada ao quadrado. Portanto, a variância é dada sempre no quadrado da unidade de medida da série. Se os dados são expressos em metros, a variância é expressa em metros quadrados. Em algumas situações, a unidade de medida da

variância nem faz sentido. É o caso, por exemplo, em que os dados são expressos em litros. A variância será expressa em litros quadrados.

Portanto, o valor da variância não pode ser comparado diretamente com os dados da série, ou seja: **variância não tem interpretação.**

Exatamente para suprir esta deficiência da variância é que se define o desvio padrão. Como o desvio padrão é a raiz quadrada da variância, o desvio padrão terá sempre a mesma unidade de medida da série e, portanto admite interpretação.

Coeficiente de variação (CV)

Se uma série X apresenta $\bar{x} = 10$ e $\sigma_x = 2$ e uma série Y apresenta $\bar{y} = 100$ e $\sigma_y = 5$, do ponto de vista da dispersão absoluta, a série Y apresenta maior dispersão que a série X.

No entanto, se levarmos em consideração as médias das séries, o desvio padrão de Y que é 5 em relação a 100 é um valor menos significativo que o desvio padrão de X que é em relação a 10.

Isto nos leva a definir a medida de dispersão relativa, o **coeficiente de variação**. Esta medida é útil quando queremos saber se a dispersão é muito grande em relação à média.

O coeficiente de variação é indicado por

$$CV = \frac{\sigma}{\mu}$$

Calculando, então, o coeficiente de variação das séries citadas tem:

$$CV_x = \frac{2}{10} = 0,2 \text{ ou } 20\%$$

$$CV_y = \frac{5}{100} = 0,05 \text{ ou } 5\%$$

Comparando os valores destes dois coeficientes concluímos que a série X admite maior dispersão relativa.

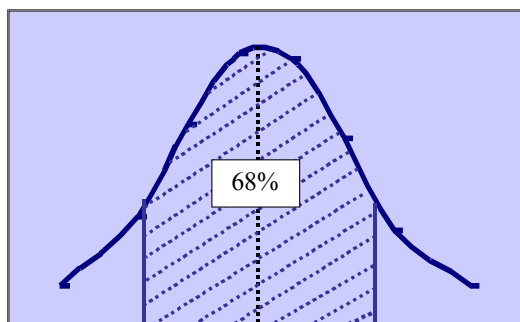
Como a medida de dispersão relativa leva em consideração a medida de dispersão absoluta e a média da série, é uma medida mais completa que a medida de dispersão absoluta.

O uso do desvio padrão

O desvio padrão é a mais importante das medidas de dispersão.

Na curva normal o valor da média aritmética é igual ao valor da moda e igual ao valor da mediana. $\bar{x} = Md = Mo$.

Quando uma curva de frequência representativa da série é perfeitamente simétrica como a curva abaixo, podemos afirmar que o intervalo $[\bar{x} - \sigma, \bar{x} + \sigma]$ contém aproximadamente **68%** dos valores da série.



O

intervalo

[$- 2 \sigma$, $+ 2 \sigma$] contém aproximadamente **95%** dos valores da série.

O intervalo [$- 3 \sigma$, $+ 3 \sigma$] contém aproximadamente **99%** dos valores da série.

Esses percentuais 68%, 95% e 99% citados na interpretação serão comprovadas, com maior precisão, no estudo da **distribuição normal de probabilidades**.

Quando a distribuição não é perfeitamente simétrica estes percentuais apresentam pequenas variações para mais ou para menos, segundo o caso.

De modo que, quando se afirma que uma série média $\bar{x} = 100$ e desvio padrão $\sigma = 5$, podemos interpretar estes valores da seguinte forma:

- Os valores da série estão concentrados em torno de 100.
- O intervalo [95, 105] contém aproximadamente, 68% dos valores da série.
- O intervalo [90, 110] contém aproximadamente, 95% dos valores da série.
- O intervalo [85, 115] contém aproximadamente, 99% dos valores da série.

É importante perceber que, ao aumentar o tamanho do intervalo, aumenta-se o percentual de elementos contido no intervalo.

Exemplo: Foi observado que as contas de luz para uma área municipal, no mês de junho, são normalmente distribuídas. Se a média das contas for \$ 42,00 e o desvio padrão \$ 12,00, entre que intervalo de valores estão 68% das contas? E 95% das contas?

$$- \sigma = 42,00 - 12,00 = 30,00$$

$$+ \sigma = 42,00 + 12,00 = 54,00$$

68% das contas estão entre os valores de \$ 30,00 e \$ 54,00

$$- 2 \sigma = 42,00 - 2 \cdot 12,00 = 42,00 - 24,00 = 18,00$$

$$+ 2 \sigma = 42,00 + 2 \cdot 12,00 = 42,00 + 24,00 = 66,00$$

95% das contas estão entre os valores de \$ 18,00 e \$ 66,00

Exercícios

1) Calcule o desvio padrão da distribuição populacional:

Classes	2	6	10	14	18	22
f_i	5	12	21	15	7	

2) Em um exame final de Matemática, o grau médio de um grupo de 150 alunos foi 7,8 e o desvio padrão, 0,80. Em Estatística, entretanto, o grau médio final foi 7,3 e o desvio padrão, 0,76. Em que disciplina foi maior a dispersão?

3) Medidas as estaturas de 1017 indivíduos, obtivemos $\bar{x} = 162,2$ cm e $s = 8,01$ cm. O peso médio desses mesmos indivíduos é 52 kg, com um desvio padrão de 2,3 kg. Esses indivíduos apresentam maior variabilidade em estatura ou em peso?

4) Um grupo de 85 moças tem estatura média de 160,6 cm, com um desvio padrão igual a 5,97 cm. Outro grupo de 125 moças tem uma estatura média de 161,9 cm, sendo o desvio padrão igual a 6,01 cm. Qual é o coeficiente de variação de cada um dos grupos? Qual o grupo mais homogêneo?

5) Um grupo de cem estudantes tem uma estatura média de 163,8 cm, com um coeficiente de variação de 3,3%. Qual o desvio padrão desse grupo?

6) Uma distribuição apresenta as seguintes estatísticas: $\sigma = 1,5$ e $CV = 2,9\%$. Determine a média da distribuição.

7) Numa fábrica de rolamentos, retirou-se da produção de um determinado dia uma amostra de 10 rolamentos, dos quais se mediu o diâmetro externo, em mm, obtendo-se:

20,2 21,4 20,8 19,6 22,1
21,7 20,4 22,0 20,5 19,3

Calcular a média e o desvio padrão desta amostra.

8) Calcular a média e o desvio padrão da seguinte distribuição amostral de uma variável X.

faixas de observações	freqüência
0 — 10	25
10 — 20	48
20 — 30	66
30 — 40	44
40 — 50	17
Total	200

9) Em 120 experimentos, onde cada um consiste em lançar 3 moedas e contar o número de caras, obtivemos os seguintes resultados:

Nº de caras	0	1	2	3
Nº de experimentos	18	40	49	13

Calcular a média, a variância e o desvio padrão do número de caras observado em cada experimento.

10) Uma amostra de 900 lâmpadas foi testada para se determinar a durabilidade. Os dados foram:

Durabilidade em horas	freqüência
1000 — 1100	15
1100 — 1200	30
1200 — 1300	82
1300 — 1400	94
1400 — 1500	107
1500 — 1600	136
1600 — 1700	158
1700 — 1800	144
1800 — 1900	95
1900 — 2000	39
Total	900

O fabricante garante que no máximo 3% das lâmpadas têm durabilidade menor que 1200 horas e que a durabilidade média é de 1575 horas. Na amostra testada

- qual é a porcentagem de lâmpadas que duraram menos de 1200 horas?
- qual é a durabilidade média?
- qual é o desvio padrão?

11) A tabela representa as estaturas de 35 crianças nascidas numa mesma maternidade numa certa semana.

estatura (cm)	n° de crianças
45 — 46	1
46 — 47	4
47 — 48	6
48 — 49	12
49 — 50	8
50 — 51	3
51 — 52	0
52 — 53	1

Determinar a média e o desvio-padrão das estaturas destas crianças:

- ao nascerem.
- ao completarem o primeiro mês de vida, supondo que no primeiro mês cada criança cresce 4 cm.

12) Um restaurante cobra o almoço de cada cliente através do peso (por quilo) da quantidade de alimento consumida. Foi observado, durante um mês, que as quantidades de alimento consumidas são normalmente distribuídas. Se a média consumida for 550 g e o desvio padrão 200 g, calcular:

- a amplitude do intervalo da zona de normalidade;
- a amplitude dos 95% centrais.

13) Os pratos produzidos por uma indústria têm diâmetro médio de 19 cm e desvio padrão de 0,2 cm. Dois pratos A e B cujos diâmetros medem respectivamente 19,8 cm e 18,3cm

serão testados pelo Controle Estatístico de Qualidade, que admite uma tolerância de três desvios acima e três abaixo da média. Assinale a alternativa correta:

- a) O prato A será aprovado
- b) Ambos os pratos serão reprovados
- c) o prato A será reprovado e o prato B aprovado
- d) o prato B será reprovado.

14) O desvio padrão de um conjunto de dados é 16. A variância será:

- a) 16
- b) 64
- c) 256
- d) 4

15) A variância de um conjunto de dados é 16. O desvio padrão será:

- a) 4
- b) 256
- c) 36
- d) 2

16) Calcule o desvio padrão das seguintes populações:

- a) X: 2, 3, 7, 9, 11, 13.
- b) Y: 5, 12, 4, 20, 13, 17.

17) Calcule o desvio padrão das seguintes amostras:

- a) Z: 15, 16, 17, 20, 21.
- b) T: 6, 5, 10, 12, 19.

18) Uma fábrica corta bambus para a confecção de cercas. Cada corte deve ter um comprimento médio de 180cm e apresenta um desvio-padrão de 1,5cm. Após cortados, os bambus passam por um controle de qualidade que rejeita cortes que estejam com 2 desvios-padrão acima ou abaixo da média especificada. Seis bambus, A, B, C D, E e F foram medidos pelo controle de qualidade e os valores obtidos são apresentados na tabela a seguir. Quais deles o controle deve aprovar e quais deve rejeitar?

bambu	comprimento
A	178,5cm
B	183,4cm
C	176,2cm
D	175,8cm
E	182,7cm
F	180,0 cm