1.11 Medidas de Dispersão

Devemos ter em mente que as medidas de tendência central não são suficientes para caracterizar totalmente uma sequência numérica. Uma breve reflexão nos permite concluir isso.

Vamos observar as sequências a seguir:

X: 10,1,18,20,35,3,7,15,11,10 → m=13, 10 elementos.

Y: 12,13,13,14,12,14,12,14,13,13 → m=13, 10 elementos.

Z: 13,13,13,13,13,13,13,13,13,13 → m=13, 10 elementos.

Elas possuem a mesma média, mas são completamente distintas do ponto de vista da variabilidade de dados.

A média 13 representa melhor Z, do que Y e X; e Y é melhor representada do que X.

As medidas de dispersão avaliam a representatividade da média, quando há concentração e dispersão em torno desse valor.

Medidas de Dispersão Absoluta

Amplitude total, Desvio Médio Simples e Desvio Padrão

Amplitude Total

É a diferença entre o maior e o menor valor da sequência. Apesar da facilidade de obtenção da amplitude total, esta medida tem pouca sensibilidade estatística.

Desvio Médio Simples

A dispersão dos dados em relação a média de uma sequência pode ser avaliada através dos desvios de cada elemento da sequência em relação a média. Indicaremos por DMS. É definido como sendo uma média aritmética dos desvios de cada elemento da série para a média da série.

1.12 Cálculo do Desvio Médio Simples

1.12.1 Dados Brutos ou Rol

Inicialmente calculamos a média da sequência, em seguida identificamos a distância de cada elemento da sequência para sua média. Finalmente, calculamos a média destas distâncias.

Ex: X: $x_1, x_2, x_3, \cdots, x_n$

$$\text{DMS} = \frac{\sum |x_i - \overline{x}|}{n}$$

- Em média, cada elemento da sequência está afastado do valor médio $\overline{x}$ por "DMS" unidades.

1.12.2 Variável Discreta

Devemos ter em mete que no caso da variável discreta, aparecerá uma frequência, ou seja, repetições. Assim, devemos trabalhar com uma média aritmética ponderada.

$$\text{DMS} = \frac{\sum |x_i - \overline{x}| f_i}{\sum f_i}$$

Ex:

x_i	f_i
1	2
3	5
4	2
5	1

$$\overline{x} = \frac{(1\cdot2)+(3\cdot5)+(4\cdot2)+(5\cdot1)}{2+5+2+1} = 3$$

$\lvert x_i - \overline{x}\rvert f_i$
$\lvert 1-3\rvert \cdot 2 = 4$
$\lvert 3-3\rvert \cdot 5 = 0$
$\lvert 4-3\rvert \cdot 2 = 2$
$\lvert 5-3\rvert \cdot 1 = 2$

$$\sum |x_i - \overline{x}| f_i = 8$$

$$DMS = \frac{8}{10} = 0.8$$

- Em média, cada elemento da série está afastado de 3 por $0,8$ unidades.

1.12.3 Variável Contínua

Pelo fato de desconhecermos os valores individuais dos elementos, substituimos estes valores pelos pontos médios das classes.

$$\text{DMS} = \frac{\sum |x_i - \overline{x}| f_i}{\sum f_i}$$

O procedimento do cálculo é análogo ao da variável discreta. Podemos entender $|x_i - \overline{x}|$ como distância!

1.13 Variância e Desvio Padrão

A variância é uma média aritmética calculada a partir dos quadrados dos desvios obtidos entre os elementos da série e sua média.

O desvio padrão é a raiz quadrada positiva da variância.

Para esta medida devemos levar em consideração o fato de uma sequência de dados representar toda uma população ou apenas uma amostra de uma população.

$$\begin{aligned} \text{População} &\rightarrow \sigma^2(x) : \text{variância}, \\ &\rightarrow \sigma(x) : \text{desvio padrão} \\ \text{Amostra} &\rightarrow S^2(x) : \text{variância}, \\ &\rightarrow S(x) : \text{desvio padrão} \end{aligned}$$

Cálculo da Variância e Desvio Padrão

1.13.1 Dados Brutos ou Rol

a) Se a sequência representa uma **População**, a variância é calculada pela fórmula:

$$\sigma^2(x) = \frac{\sum(x_i - \overline{x})^2}{n}$$

e o desvio padrão por:

$$\sigma(x) = \sqrt{\sigma^2(x)}$$

b) Se a sequência representa uma **amostra**, a variância é dada por:

$$S^2(x) = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

e o desvio padrão por:

$$S(x) = \sqrt{S^2(x)}$$

1.13.2 Variável Discreta

a) **População:** $\sigma^2(x) = \frac{\sum(x_i - \overline{x})^2 f_i}{\sum f_i}$; $\sigma(x) = \sqrt{\sigma^2(x)}$.

a) **Amostra:** $S^2(x) = \frac{\sum(x_i - \overline{x})^2 f_i}{\sum f_i - 1}$; $S(x) = \sqrt{S^2(x)}$.

1.13.3 Variável Contínua

Análogo ao cálculo para a variável discreta, entretanto *"x_i representa o valor médio da classe."*

Algumas Considerções

No cálculo da variância, quando elevamos ao quadrado a diferença $(x_i - \overline{x})$, a unidade da série fica também elevada ao quadrado. Portanto, a variância é dada sempre no quadrado da unidade da série.

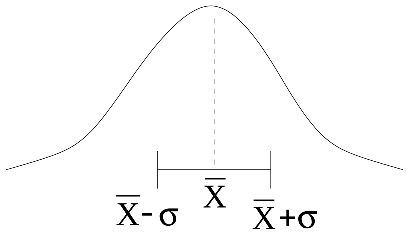
Existem casos onde a unidade de medida não tem sentido. Ex.: qudando os dados estão expressos em litros. A variância será em "litros quadrados".

- A variância não pode ser comparada diretamente com os dados da série, ou seja: variância não tem interpretação.
- Com o propósito de suprimir esta deficiência é que se define o desvio padrão.

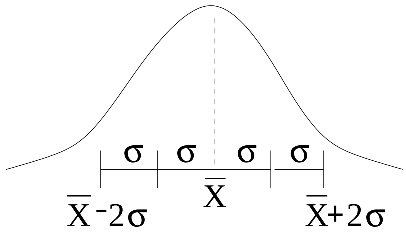
1.13.4 Interpretação do Desvio Padrão

Quando a curva de frequência representativa da série é prefeitamente simétrica, podemos fazer alguns estudos.

Ex:



Podemos afirmar que nesse intervalo $[\overline{x}-\sigma, \overline{x}+\sigma]$ contém 68% dos valores da série.



Podemos afirmar que nesse intervalo $[\overline{x} - 2\sigma, \overline{x} + 2\sigma]$ contém 95% dos valores da série.

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

Ao afirmarmos que uma série apresenta uma média $\overline{x} = 100$ e desvio padrão $\sigma(x) = 5$, podemos interpretar esses valores da seguinte forma:

1. Os valores da série estão concentrados em torno de 100.
2. 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.
3. As medidas de dispersão vistas até agora são medidas absolutas e portanto avaliam a dispersão absoluta da série.
4. Para justificar que o denominador da variância amostral deve ser $n - 1$ e não n, argumentamos:

O modelo matemático que calcula a variância de uma amostra **não pode ser**:

$$S^2(x) = \frac{\sum(x_i - \overline{x})^2}{n}$$

pois caso isto seja verdadeiro, este modelo deveria determinar a variância para qualquer tamanho de amostra.

Suponha uma amostra construída por um único elemento x_1. O valor médio da amostra também é x_1.

O cálculo pela expressão acima, seria então:

$$S^2(x) = \frac{\sum(x_i - x_1)^2}{1} = 0$$

Seríamos induzidos a afirmar que a dispersão da população de onde provém a amostra é zero, isto é, a população é cosntruída em sua totalidade por elementos idênticos. O que é, em geral, uma afirmação falsa. Para corrigir, basta colocar $n-1$. O que dá:

$$S^2(x) = \frac{\sum(x_i - x_1)^2}{n-1}$$

Caso a amostra tenha um elemento $S^2(x)$ nos dá $\frac{0}{0}$ que é uma indeterminação, o que significa que a variância existe mas não está determinada. Significa também que amostras de apenas um elemento não nos fornece informações sobre a variância da série.

1.13.5 Medidas de Dispersão Relativa

Se uma série X apresenta $\overline{x} = 10$ e $\sigma(x) = 2$ e uma série apresenta $\overline{x} = 100$ e $\sigma(x) = 5$, do ponto de vista da dispersão absoluta, a série apresenta maior dis persão que a série.

Levando em consideração as médias das séries, oesvio p adrão de que é 5 em relação a 100 é um

$\overline{}$
$\overline{x}$

Variância Relativa: $V(x)$ definida por $V(x) = \dfrac{\sigma^2(x)}{(\overline{x})^2}$

Devido a disvisão de elementos com mesma unidade, obtemos como resultado um número puro. O que nos permite escrevê-lo como percentual.

Calculando para os exemplos citados no início, temos que:

$$\text{CV}_{(x)} = \frac{2}{10} = 0,2 \quad \text{ou} \quad 20\%$$
$$\text{CV}_{(y)} = \frac{5}{100} = 0,05 \quad \text{ou} \quad 5\%$$

Comparando ambos, concluímos que a sérieX admite maior di spersão relativa. Sendo uma informação mais completa que a medida de dispersão absoluta.

Podemos consccluir:

A série apresenta maior dispersão absoluta.

A série apresenta maior dispersão relativa.

Portanto, a série apresenta maior dispersão.

Exercícios Propostos

1. Calcule a amplitude total da série: 2,8,10,15,20,22,3 0.

2. Calcule a amplitude total da série Y: 12,9,15,40,22,34,8.

3. Calcule a amplitude total da série:

x_i	f_i
3	4
8	7
12	9
15	10
20	3

4. Calcule a amplitude total da série:

Classe	Salário US$	N° de vendedores
1	70 |— 120	8
2	120 |— 170	28
3	170 |— 220	54
4	220 |— 270	32
5	270 |— 320	12
6	320 |— 370	6

5. Considerando as séries X e Y da questão 1 e 2, qual delas apresenta maior dispersão absoluta?

6. Calcule o DMS da série X: 3,8,12,3,9,7.

7. Interprete o valor obtido questão anterior.

8. Calcule o DMS da série:

x_i	f_i
2	3
4	8
5	10
6	6
8	2
10	1

9. Interprete o valor obtido na questão anterior.

10. Calcule o DMS da série da questão 4 e interprete o valor obtido.

11. Calcule a variância e o desvio padrão da População X: 2,3,7,9,11,13.

12. c Y: 5,12,4,20,13,17.

13. Calcule a variância e o desvio padrão da amostra Z: 15,16,17,20,21.

14. Calcule a variância e o desvio padrão da população abaixo, e interprete o seu valor:

Idade	N° de alunos
17	3
18	18
19	17
20	8
21	4

15. Calcule a variância e o desvio padrão para a distribuição de valores de 54 notas fiscais emitidas na mesma data, selecionadas em uma loja de departamentos. (Amostra)

Classe	Consumo por nota US$	N° de notas
1	0 |— 50	10
2	50 |— 100	28
3	100 |— 150	12
4	150 |— 200	2
5	200 |— 250	1
6	250 |— 300	1

16. Interprete o valor obtido na questão anterior.

17. Responda, justificando em cada passo, as questões abaixo:
    a) Qual das séries apresenta maior dispersão absoluta?
    b) Qual das séries apresenta maior dispersão relativa?
    c) Qual das séries apresenta maior dispersão?

(a) A: $\begin{cases} \overline{X}_A = 20 \\ \sigma(A) = 2 \end{cases}$ B: $\begin{cases} \overline{X}_B = 20 \\ \sigma(B) = 5 \end{cases}$

(b) A: $\begin{cases} \overline{X}_A = 50 \\ \sigma(A) = 2 \end{cases}$ B: $\begin{cases} \overline{X}_B = 100 \\ \sigma(B) = 3 \end{cases}$

(c) A: $\begin{cases} \overline{X}_A = 20 \\ \sigma^2(A) = 9 \end{cases}$ B: $\begin{cases} \overline{X}_B = 30 \\ \sigma^2(B) = 16 \end{cases}$

(d) A: $\begin{cases} \overline{X}_A = 30 \\ \sigma(A) = 5 \end{cases}$ B: $\begin{cases} \overline{X}_B = 50 \\ \sigma^2(B) = 9 \end{cases}$

(e) A: $\begin{cases} \overline{X}_A = 20 \\ \sigma^2(A) = 9 \end{cases}$ B: $\begin{cases} \overline{X}_B = 40 \\ \sigma(B) = 3 \end{cases}$

(f) A: $\begin{cases} \overline{X}_A = 20 \\ \sigma(A) = 3 \end{cases}$ B: $\begin{cases} \overline{X}_B = 60 \\ \sigma(B) = 9 \end{cases}$