ESTATÍSTICA

- 2.3.2 - Medidas de Dispersão

O objetivo das medidas de dispersão é medir quão próximos uns dos outros estão os valores de um grupo (e algumas mensuram a dispersão dos dados em torno de uma medida de posição).

Intervalo

É a medida mais simples de dispersão. Consiste em identificar os valores extremos do conjunto (mínimo e máximo), podendo ser expresso:

- pela diferença entre o valor máximo e o mínimo;

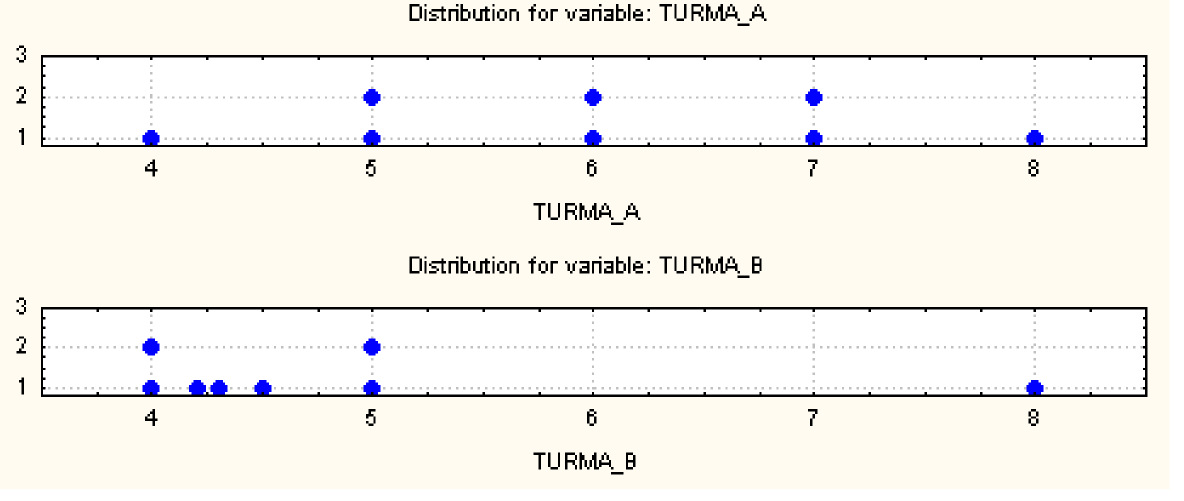
- pela simples identificação dos valores.

EX1.:Observe o conjunto abaixo, referente às notas de duas turmas:

Turma	Valores	Intervalo
A	4 5 5 6 6 7 7 8	4 [4,8]
B	4 4 4,2 4,3 4,5 5 5 8	4 [4,8]

ESTATÍSTICA

Desvantagem do uso do intervalo



Apesar de sua simplicidade o intervalo não dá idéia de como os dados estão agrupados entre os extremos.

No caso acima ambos os grupos têm o mesmo intervalo (4, [4,8]), mas no primeiro grupo os dados estão bem dispersos, enquanto no primeiro estão próximos do valor mínimo.

ESTATÍSTICA

Variância (s^2)

A variância é uma das medidas de dispersão mais importantes. É a média aritmética dos quadrados dos desvios de cada valor em relação à **média**: proporciona uma mensuração da dispersão dos dados em torno da média.

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \textbf{(amostra)}$$

- se os dados referem-se a uma POPULAÇÃO usa-se **n** no denominador da expressão.

- a unidade da variância é o quadrado da unidade dos dados (e portanto o quadrado da unidade da média) causando dificuldades para avaliar a dispersão:

Por exemplo:

Média = 75 kg

Variância = 12 kg^2

ESTATÍSTICA

Desvio Padrão (s)

É a raiz quadrada positiva da variância, apresentando a mesma unidade dos dados e da média, permitindo avaliar melhor a dispersão.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad \textbf{(amostra)}$$

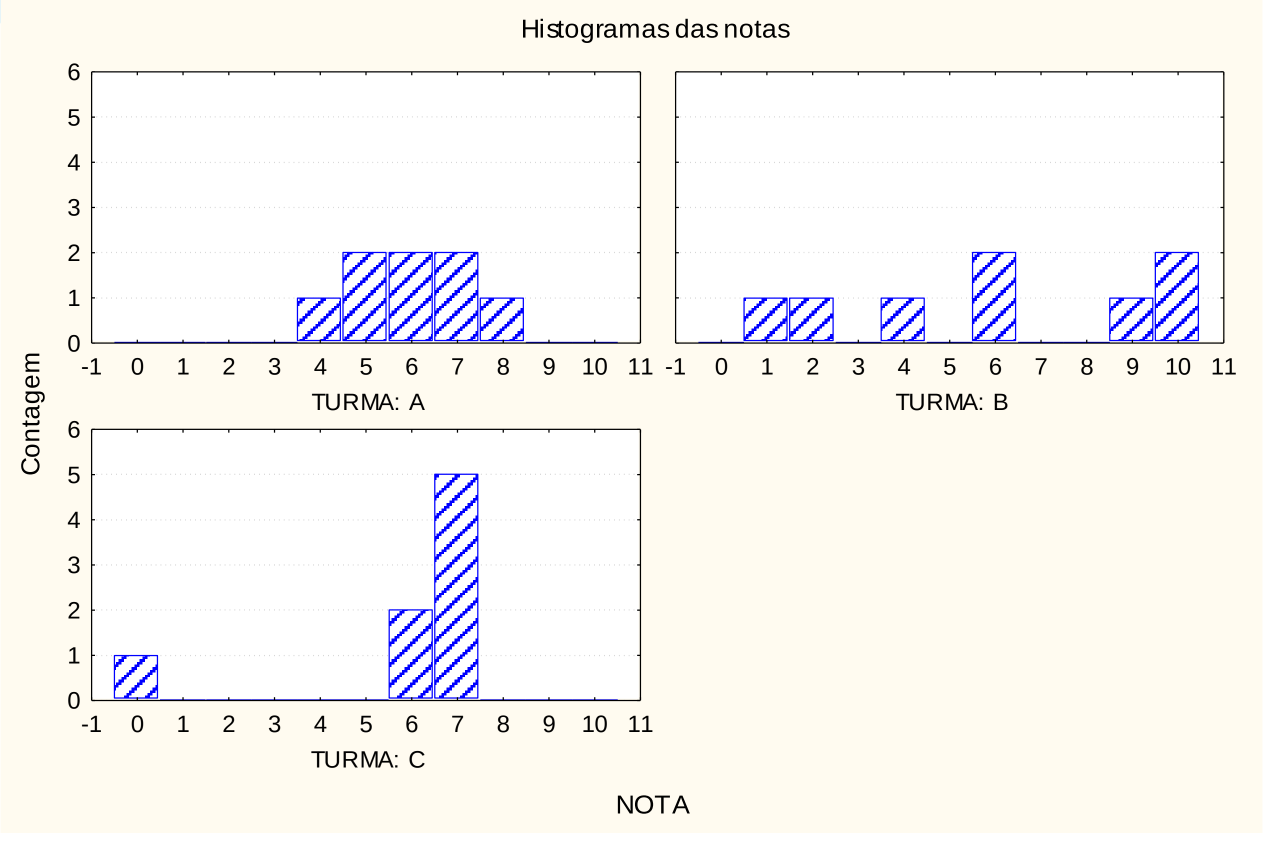
- se os dados referem-se a uma POPULAÇÃO usa-se **n** no denominador da expressão.

- quanto maior o desvio padrão, maior a dispersão dos dados em torno da média.

EX.2 A tabela abaixo refere-se às notas finais de 3 turmas de estudantes.

Turma	Valores	Média	Desvio Padrão
A	4 5 5 6 6 7 7 8	6,0	1,31
B	1 2 4 6 6 9 10 10	6,0	3,51
C	0 6 6 7 7 7 7,5 7,5	6,0	2,49

ESTATÍSTICA



Quanto maior a dispersão dos dados maior o valor do desvio padrão:

- na figura acima, a turma A é a mais homogênea, tendo o menor desvio padrão, 1,31.

- a turma B tem o maior desvio padrão por apresentar maior dispersão, 3,51.

- o valor discrepante (0) na turma C aumenta a dispersão dos dados.

ESTATÍSTICA

Fórmula alternativa para cálculo do desvio padrão

Se a média apresentar um valor fracionário os desvios de cada valor em relação à média acumularão erros de arredondamento.

Fórmula equivalente, que reduz o erro:

$$s = \sqrt{\frac{\sum(x^2) - [\frac{(\sum x)^2}{n}]}{n-1}} \quad \textbf{(amostra)}$$

EX.3 Calcular o desvio padrão para o conjunto de dados abaixo:

									Soma
X	4	5	5	6	6	7	7	8	48
X^2	16	25	25	36	36	49	49	64	300

$$s = \sqrt{\frac{300 - \left[(48)^2 / 8\right]}{8-1}} = \sqrt{\frac{300 - 288}{7}} = 1{,}31$$

ESTATÍSTICA

Cálculo do desvio padrão a partir de uma tabela de freqüências

Tal como no caso da média os valores da variável (ou os pontos médios das classes), e os quadrados desses valores, serão multiplicados por suas respectivas freqüências:

$$s = \sqrt{\frac{\sum(x^2 \times f) - \{\frac{[\sum(x \times f)]^2}{n}\}}{n-1}} \quad \textbf{(amostra)}$$

Onde **x** é o valor da variável (discreta) ou do ponto médio da classe, e **f** a sua freqüência.

Novamente se o conjunto de dados referir-se a uma população deve-se usar **n** no denominador.

ESTATÍSTICA

EX.4 - Calcular o desvio padrão do número de pessoas por residência, com base na tabela abaixo:

Pessoas X	Residências f	X x f	X^2	X^2 x f
1	1	1	1	1
2	3	6	4	12
3	6	18	9	54
4	13	52	16	208
5	11	55	25	275
6	4	24	36	144
7	0	0	49	0
8	2	16	64	128
Total	40	172	-	822

$$\sum(x^2 \times f) = 822 \qquad \sum(x \times f) = 172 \qquad n = 40$$

$$s = \sqrt{\frac{\sum(x^2 \times f) - \{\frac{[\sum(x \times f)]^2}{n}\}}{n-1}}$$

$$s = \sqrt{\frac{822 - [(172)^2 / 40]}{40-1}} = \sqrt{\frac{822 - 739{,}6}{39}} \cong 1{,}45$$

ESTATÍSTICA

EX.5.Calcular o desvio padrão das taxas de mortalidade infantil em municípios do Oeste de SC em 1982

Classes	Freq f	Ponto médio X	X x f	X^2	X^2 x f
9,9 |-- 18,62	10	14,26	142,6	203,3476	2033,476
18,62 |-- 27,34	13	22,98	298,74	528,0804	6865,0452
27,34 |-- 36,06	6	31,7	190,2	1004,89	6029,34
36,06 |-- 44,78	4	40,42	161,68	1633,7764	6535,1056
44,78 |-- 53,5	0	49,14	0	2414,7396	0
53,5 |-- 62,2	1	57,86	57,86	3347,7796	3347,7796
Total	34	-	851,08	-	24810,7464

$$\sum(x^2 \times f) = 24810{,}7464 \quad \sum(x \times f) = 851{,}08 \quad n = 34$$

$$s = \sqrt{\frac{\sum(x^2 \times f) - \{\frac{[\sum(x \times f)]^2}{n}\}}{n-1}}$$

$$s = \sqrt{\frac{24810{,}7464 - [(851{,}08)^2 / 34]}{34-1}} \cong 10{,}31$$

ESTATÍSTICA

Teorema de Chebyshev

"A proporção (ou fração) de qualquer conjunto de dados a menos de K desvios padrões a contar da média é sempre *ao menos* 1 - 1/K^2, onde K é um número positivo maior do que 1."

$$\bar{x} - K \times s \text{ a } \bar{x} + K \times s \Rightarrow \geq 1 - \frac{1}{K^2}$$

- Para K = 2: ao menos 3/4 (75%) de todos os valores estão no intervalo que vai de 2 desvios padrão abaixo da média até 2 desvios padrões acima da média.

$$\bar{x} - 2 \times s \text{ a } \bar{x} + 2 \times s \Rightarrow \geq 1 - \frac{1}{2^2} = \frac{3}{4}$$

- Para K = 3: ao menos 8/9 (89%) de todos os valores estão no intervalo que vai de 3 desvios padrões abaixo da média até 3 desvios padrões acima da média.

$$\bar{x} - 3 \times s \text{ a } \bar{x} + 3 \times s \Rightarrow \geq 1 - \frac{1}{3^2} = \frac{8}{9}$$

ESTATÍSTICA

Coeficiente de Variação Percentual (c.v.%)

O coeficiente de variação percentual é uma medida de dispersão relativa, pois permite comparar a dispersão de diferentes distribuições (com diferentes médias e desvios padrões).

$$c.v.\% = \frac{s}{\bar{x}} \times 100$$

Onde $\bar{x}$ é a média e **s** é o desvio padrão do conjunto de dados.

Quanto menor o valor do c.v.% mais os dados estão concentrados em torno da média (conjunto mais homogêneo.

EX.6 Sejam 2 turmas. As notas da turma A apresentam média 6 com desvio padrão 2,5, e as da turma B média 9 e desvio padrão 3. Qual é a mais homogênea em termos de notas.

$$c.v.\%_A = \frac{2{,}5}{6} \times 100 \cong 41{,}67\%$$

$$c.v.\%_B = \frac{3}{9} \times 100 \cong 33{,}33\%$$