MEDIDAS DE POSIÇÃO

Expressam a característica dos dados observados ***tenderem*** a se ***agrupar em torno dos valores centrais***, indicado a ***posição*** da série em relação ao eixo dos valores assumidos pela variável ou característica em estudo. Em síntese, podemos dizer que as **MEDIDAS DE POSIÇÃO** tentam traduzir a ***semelhança*** que os ***dados estatísticos*** referentes à observação de um fenômeno apresentam ***entre si***, conforme se pode notar pela observação das séries abaixo.

Série	Valores										Média	Mediana	Moda
1	1	3	7	10	10	11	15	18	20	35	**13**	**10,5**	**10**
2	12	12	13	13	13	13	13	13	14	14	**13**	**13**	**13**
3	13	13	13	13	13	13	13	13	13	13	**13**	**13**	**13**

A julgar apenas pela **MÉDIA**, teríamos que concluir pela ***igualdade*** entre as três séries **1**, **2** e **3**. Se estendermos nossa análise, incluindo as medidas **MÉDIA**, **MEDIANA** e **MODA**, teríamos que concluir pela ***igualdade*** entre as séries **2** e **3**. Mas, como os conjuntos são pequenos, conseguimos observar que eles não são iguais.

SEPARATRIZES

Outras medidas de posição são as **separatrizes**, que englobam:
- a própria mediana;
- os quartis;
- os decis;
- os percentis.

Mediana: divide a série em duas partes iguais.

Quartis: denominamos quartis os valores de uma série que a dividem em quatro partes iguais.

Há, portanto, três quartis:

- O **primeiro quartil (Q_1) :** é o valor situado de tal modo na série que uma quarta parte (25%) dos dados é menor que ele e as três quartas partes restantes (75%) são maiores.
- O **segundo quartil (Q_2) :** é exatamente o valor da mediana, ou seja, o valor situado de tal modo na série que deixa metade (50%) dos dados a esquerda dele e a outra metade à direita **(Q_2 =Md).**
- O **terceiro quartil (Q_3) :** é o valor situado de tal modo na série que as três quartas partes (75%) dos dados são menores que ele e uma quarta parte restante (25%) é maior.

FÓRMULA DO QUARTIL PARA DADOS BRUTOS

$$Q_i = x_i.\frac{n}{4}$$

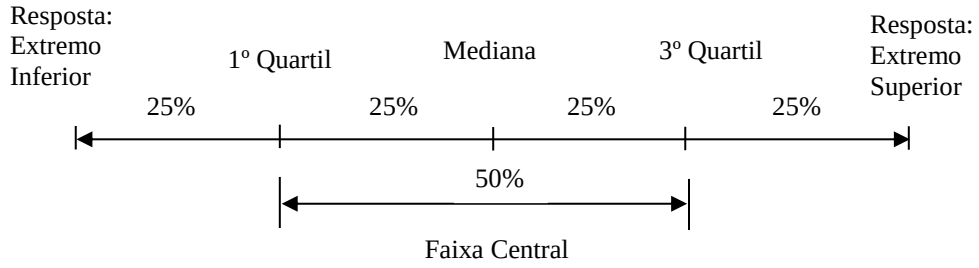
FÓRMULA DO QUARTIL PARA TABELA COM INTERVALO DE CLASSE

$$Q_i = l_i + \frac{\left(i.\frac{\Sigma f_i}{4} - Fant\right)}{fi_{\text{classe considerada}}}.h$$

$\frac{\Sigma f_i}{4}$ = somatório das frequências dividido por quatro;
Li = limite inferior da classe do quartil considerado;
Fant = frequência acumulada da classe anterior à classe do quartil considerado;
h = amplitude do intervalo de classe do quartil considerado;
fi = frequência simples da classe do quartil considerado.

Os quartis são valores de um conjunto de dados ordenados, que os dividem em quatro partes iguais. É necessário, portanto, três quartis (Q_1, Q_2 e Q_3) para dividir um conjunto de dados ordenados em quatro partes iguais.
Q_1 : deixa 25% dos elementos abaixo dele.
Q_2 : deixa 50% dos elementos abaixo dele e coincide com a mediana.
Q_3 : deixa 75% dos elementos abaixo dele.

A figura abaixo mostra bem o quartis:



Decis: os decis por sua vez, são os dez valores que dividem a série em 10 partes iguais, onde, cada uma delas contém 10% dos dados.

FÓRMULA DO DECIL PARA DADOS BRUTOS

$$D_i = x_i . \frac{n}{10}$$

FÓRMULA DO DECIL PARA TABELA COM INTERVALO DE CLASSE

$$D_i = l_i + \frac{\left(i . \frac{\Sigma f_i}{10} - Fant\right)}{fi_{\text{classe considerada}}} . h$$

$\frac{\Sigma f_i}{10}$ = somatório das frequências dividido por dez;
Li = limite inferior da classe do decil considerado;
Fant = frequência acumulada da classe anterior à classe do decil considerado;
h = amplitude do intervalo de classe do decil considerado;
fi = frequência simples da classe do decil considerado.

Percentis: denominamos percentis os noventa e nove valores que separam uma série em 100 partes iguais, ou seja:

$P_1, P_2, P_3, \ldots, P_{99}$, onde $P_{50} = Md = Q_2$, $P_{25} = Q_1$ e $P_{75} = Q_3$

FÓRMULA DO PERCENTIL PARA DADOS BRUTOS

$$P_i = x_i . \frac{n}{100}$$

FÓRMULA DO PERCENTIL PARA TABELA COM INTERVALO DE CLASSE

$$P_i = l_i + \frac{\left(i . \frac{\Sigma f_i}{100} - Fant\right)}{fi_{\text{classe considerada}}} . h$$

$\frac{\Sigma f_i}{100}$ = somatório das frequências dividido por cem;
Li = limite inferior da classe do percentil considerado;
Fant = frequência acumulada da classe anterior à classe do percentil considerado;
h = amplitude do intervalo de classe do percentil considerado;
fi = frequência simples da classe do percentil considerado.

Sintetizando o modo de encontrar as medidas de posição de acordo com a forma de apresentação dos dados, vemos que as medidas descritas abaixo devem ser obtidas:

Quando os dados se apresentarem em:	Média	Moda	Mediana	Quartis, Decis e Percentis
Rol	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	Pela observação dos dados	Pela observação dos dados	Pela observação dos dados
Agrupamento Simples	$\overline{x} = \frac{\sum_{i=1}^{n} x_i . f_i}{\sum f_i}$	Pela observação dos dados	Pela observação dos dados	Pela observação dos dados
Ramo e Folhas	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	Pela observação dos dados	Pela observação dos dados	Pela observação dos dados
Agrupamento Em Classes	$\overline{x} = \frac{\sum_{i=1}^{n} x_i . f_i}{\sum f_i}$	Fórmula $\text{Mo} = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} . h$	Fórmula $\text{Md} = l_i + \frac{\left(\frac{\Sigma f_i}{2} - Fant\right)}{fmd} . h$	Fórmula $Q_i = l_i + \frac{\left(i.\frac{\Sigma f_i}{4} - Fant\right)}{fi_{\text{classe considerada}}} . h$ $D_i = l_i + \frac{\left(i.\frac{\Sigma f_i}{10} - Fant\right)}{fi_{\text{classe considerada}}} . h$ $P_i = l_i + \frac{\left(i.\frac{\Sigma f_i}{100} - Fant\right)}{fi_{\text{classe considerada}}} . h$

Frequentemente, as **MEDIDAS DE TENDÊNCIA CENTRAL** ***não são suficientes*** para caracterizar completamente uma série numérica, conforme pode ser observado nas séries de dados acima.

O que se constata, é que os ***fenômenos*** passíveis de análise pelo método estatístico, bem como ***os dados estatísticos*** a eles referentes, ***caracterizam-se*** tanto pela sua ***semelhança*** quanto pela sua ***variabilidade***.

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

Vimos que a média a moda e a mediana podiam ser usadas para resumir, num único número, aquilo que é "médio" ou "típico" de um conjunto de dados. Mas a informação contida fornecida pelas medidas de posição necessita em geral ser complementada pelas medidas de dispersão. Estas servem para indicar o quanto os dados se apresentam dispersos em torno da região central. Caracterizam, portanto, o grau de variação existente no conjunto de valores. As medidas de dispersão que nos interessam são:

- a amplitude total;
- o desvio médio;
- a variância;
- o desvio-padrão;
- e o coeficiente de variação;
- Box Plot.

A dispersão mede quão próximos os valores estão uns dos outros no grupo.

(A) Pequena Dispersão

(B) Grande Dispersão

A variabilidade de B é maior que a de A.
Para termos uma boa representação dos dados, temos que ter:
Uma medida de posição (quase sempre a Média) mais uma medida de dispersão (quase sempre o Desvio Padrão).

AMPLITUDE TOTAL

Dados não Agrupados

A amplitude total é a diferença entre o maior e o menor valor observado:

$$\mathbf{A_T = x(máx) - x(mín)}$$

Exemplo: Para os valores: 40, 45, 48, 52, 54, 62 e 70
Temos: AT = 70 – 40 = 30

Quando dizemos que a amplitude total dos valores é 30, estamos afirmando alguma coisa do grau de sua concentração. É evidente que, quanto maior a amplitude total, maior a dispersão ou variabilidade dos valores da variável.

Dados Agrupados

- ✓ **Sem intervalos de classe:**

Neste caso, ainda temos: $A_T = x(máx) - x(mín)$
Exemplo: Considerando a tabela abaixo:

x_i	0	1	2	3	4
f_i	2	6	12	7	3

Temos: $A_T = 4 - 0 = 4$

- ✓ **Com intervalos de classe:**

Neste caso, a amplitude total é a diferença entre o **limite superior da última** classe e o **limite inferior da primeira** classe: $A_T = L_{sup}(máx) - l_{inf}(mín)$
Exemplo: Considerando a distribuição abaixo:

i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\sum = 40$

Temos: $A_T = 174 - 150 = 24$

VARIÂNCIA E DESVIO PADRÃO

Duas medidas de variação que usam todas as entradas de dados são a variância e o desvio padrão. Contudo, antes de aprender essas medidas de variação, você precisa saber qual é o significado do desvio de uma entrada em um conjunto de dados.

Desvio de um entrada *x* em um conjunto de dados de uma população é diferença entre a entrada e a média μ do conjunto de dados, ou seja, parâmetro que indica o grau de variação de um conjunto de elementos.

Exemplo: Dada a temperatura máxima durante 3 dias em uma cidade A, obteve-se os seguintes valores: 28º, 29º e 30º, a média calculada é de: 29º.

Em outra cidade B, foram coletadas as temperaturas máximas de 22º, 29º e 35º, obtendo de média 29º.

Logo as médias das duas cidades tem o mesmo valor. Para podermos diferenciar uma média da outra, foi criada a noção de desvio padrão, que serve para dizer o quanto os valores dos quais se extraiu a média são próximos ou distantes da própria média.

Quanto menor o desvio padrão, mais homogênea é a minha amostra.

DESVIO MÉDIO (D_M): É a média aritmética dos desvios.

Formulação matemática:

DESVIO MÉDIO (DADOS BRUTOS)	**DESVIO MÉDIO (DADOS TABELADOS)**
$D_m = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n}$	$D_m = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert . f_i}{\sum_{i=1}^{n} f_i}$

VARIÂNCIA (s^2) OU (σ^2)

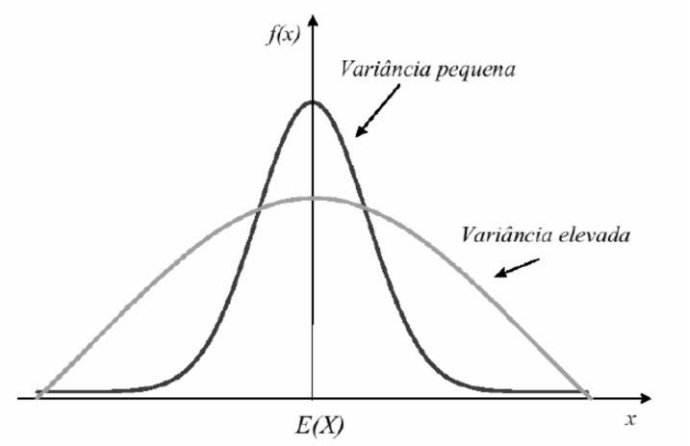
A **VARIÂNCIA** é uma medida que tem pouca utilidade como estatística descritiva, porém é extremamente importante na inferência estatística. A variância leva em consideração os valores extremos e os valores intermediários, isto é, expressa melhor os resultados obtidos.

Quando a série de dados representa uma **AMOSTRA**, a **VARIÂNCIA** é denotada por $\mathbf{s^2}$, e quando provém de uma **POPULAÇÃO**, a **VARIÂNCIA** é denotada por $\boldsymbol{\sigma^2}$ (σ = sigma minúsculo, caractere do alfabeto grego, equivalente ao s minúsculo no alfabeto arábico). Observe que há uma ***diferença*** no método de ***cálculo das duas VARIÂNCIAS***: quando se trata de uma **POPULAÇÃO**, o denominador da equação de $\boldsymbol{\sigma^2}$ representa a quantidade total de elementos na população (**N**), enquanto no caso de uma **AMOSTRA**, o denominador da equação de $\mathbf{s^2}$ é o total de elementos na amostra menos 1 (**n-1**).

Formulação matemática:

VARIÂNCIA AMOSTRAL s^2	**VARIÂNCIA POPULACIONAL σ^2**
$s^2 = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert^2}{(n-1)}$	$\sigma^2 = \frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert^2}{N}$

Em várias situações, torna-se necessário visualizar como os dados estão dispersos. Tomando como exemplo várias empresas que apresentem salários médios iguais, podemos concluir, então, que a contribuição social (% do salário) será a mesma? Somente com base no salário médio, sim, mas estaríamos chegando a uma conclusão errada. A variação em termos de faixas salariais pode ser diferente, apesar de apresentarem a mesma média.



DESVIO PADRÃO (S) OU (σ)

O desvio-padrão é a medida mais usada na comparação de diferenças entre grupos, por ser mais precisa e estar na mesma medida do conjunto de dados. Ele determina a dispersão dos valores em relação a média. Sua formulação é dada pela raiz quadrada da média aritmética dos quadrados dos desvios, ou seja:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}}$$, logo temos:

DESVIO PADRÃO AMOSTRAL (*S*)	DESVIO PADRÃO POPULACIONAL (σ)
$\mathbf{s} = \sqrt[2]{s^2}$	$\boldsymbol{\sigma} = \sqrt[2]{\sigma^2}$

Importante!

Condição para se usar o desvio-padrão ou variância para comparar a variabilidade entre grupos:

- ✓ mesmo número de observações;
- ✓ mesma unidade;
- ✓ mesma média.

COEFICIENTE DE VARIAÇÃO (CV)

Podemos considerar uma situação na qual se avalia o custo indireto de fabricação (CIF) de um produto em reais e o tempo gasto em uma máquina para fabricação deste produto em segundos.

	$\bar{x}$	s
CIF	R$ 175,00	R$ 5,00
Tempo	68 segundos	2 segundos

A princípio, você poderia concluir que o CIF apresenta maior variabilidade. Entretanto, as condições citadas anteriormente deveriam ser satisfeitas para que pudesse utilizar o desvio padrão para comparar a variabilidade. Como as condições não são satisfeitas, devemos tentar expressar a dispersão dos dados em torno da média, em termos porcentuais. Então, utilizaremos uma medida estatística chamada **de coeficiente de variação**.

O coeficiente de variação (cv) é definido como o quociente entre o desvio-padrão e a média. É expresso em porcentagem.

A grande utilidade do COEFICIENTE DE VARIAÇÃO é permitir a *comparação de variabilidade* de *diferentes conjuntos de dados*.

$$cv = \frac{S}{\bar{x}} \text{ x } 100, \text{ logo temos:}$$

$$CV_{amostra} = \frac{s}{\bar{x}} \times 100 \quad ou \quad CV_{população} = \frac{\sigma}{\bar{x}} \times 100$$

Para a situação do CIF e Tempo, teremos:

$$CV_{CIF} = \frac{s}{\bar{x}} \times 100 \quad = \frac{5}{175} \times 100 = 2{,}85\%$$

$$CV_{t} = \frac{s}{\bar{x}} \times 100 \quad = \frac{2}{68} \times 100 = 2{,}94\%$$

Portanto, neste caso, o tempo de horas da máquina apresenta maior dispersão do que o custo indireto de fabricação (CIF), mudando a conclusão anterior.

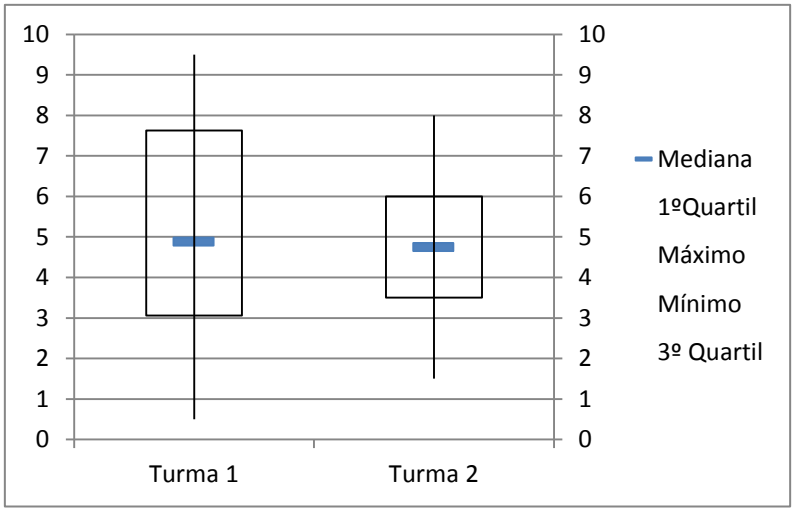
O **ESQUEMA DOS CINCO NÚMEROS** de um conjunto de dados consiste na menor observação, no primeiro quartil, na mediana, no terceiro quartil e na maior observação, escritos do menor para o maior. Sendo representado como:

Min. Q_1 Md Q_3 Max.

Embora as três medidas **Q_1**, **MEDIANA** e **Q_3** mostrem a forma da distribuição de ***50% dos valores ao redor da mediana***, a adição dos valores **MÍNIMO** e **MÁXIMO** a estas três medidas permite obter um conjunto mais completo de informações sobre a forma da distribuição. O **BOX PLOT** é a **forma gráfica** de representar estas cinco medidas estatísticas num único conjunto de resultados conforme ilustrado abaixo.

	Turma 1	Turma 2
Mediana	4,88	4,75
1ºQuartil	3,06	3,5
Máximo	9,5	8
Mínimo	0,5	1,5
3ºQuartil	7,63	6

Notas de duas turmas de matemática.



Na verdade, o gráfico **BOX PLOT** nos fornece informações sobre a posição central, dispersão e assimetria da respectiva distribuição de frequências dos dados.

Se estivermos diante de uma situação na qual essas três medidas apresentam o mesmo valor, tal fato nos informa que a distribuição dos dados é **simétrica**; quando resultam em valores diferentes, porém **muito próximos**, indica que a forma dessa distribuição é **aproximadamente simétrica**. Nesses casos, optaremos por qualquer uma das três: média, moda ou mediana.

OBS: Quartis em dados não agrupados

➔ O método mais prático é utilizar **o princípio do cálculo da mediana** para os **3 quartis**. Na realidade **serão calculadas " 3 medianas " em uma mesma série**.

*Ex 1: Calcule os **quartis** da série: { 5, 2, 6, 9, 10, 13, 15 }*

- O primeiro passo a ser dado é o da ordenação (crescente ou decrescente) dos valores:

{ 2, 5, 6, 9, 10, 13, 15 }

*- O valor que divide a série acima em duas partes iguais é igual a **9**, logo a **Md = 9 que será = Q_2 = 9***

*- Temos agora **{2, 5, 6 }** e **{10, 13, 15 }** como sendo os dois grupos de valores iguais proporcionados pela mediana (**2º quartil**). Para o cálculo do **1º e 3º quartis** basta **calcular as medianas das partes iguais provenientes da verdadeira Mediana da série (2º quartil).***

*Logo em **{ 2, 5, 6 }** a mediana é = **5** . Ou seja: será o **1º quartil = Q_1 = 5** em **{10, 13, 15 }** a mediana é =**13** . Ou seja: será o**3º quartil = Q_3= 13***

*Ex 2: Calcule os **quartis** da série: { 1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13 }*

*A série já está ordenada, então calcularemos o **2º Quartil = Md = (5+6)/2 = 5,5***

*- O **1º quartil** será a mediana da série à esquerda de Md **: { 1, 1, 2, 3, 5, 5 }***

Q_1 = (2+3)/2 = 2,5

*- O **3º quartil** será a mediana da série à direita de Md : **{6, 7, 9, 9, 10, 13 }***

Q_3 = (9+9)/2 = 9

Exercícios

1-Para as distribuições:

a)-Calcule D_6, P_{65}, e Q_1

Classes	4|-- 6	6|-- 8	8 |-- 10	10 |--12
f_i	4	11	15	5

(resp. D_6 =8,8; P_{65} = 9,03 e Q_1 = 6,86)

b) Calcule D_2, P_{43}, e Q_3

Classes	20|-- 30	30|-- 40	40 |-- 50	50 |--60	60 |--70
f_i	3	8	18	22	24

(resp. $D_{2=}$ 33,6; P_{43} = 42,32 e $Q_{3=}$ 50)

2-Dada a distribuição, determinar os quartis, D_2, D_6, D_8, P_{37} , P_5 , P_{86}, P_{47} e P_{93}. (resp. Q_1 = 653,33 ; Q_2 = 723; Q_3 = 796,76 ; $D_{2=}$ 638,67; D_6 =750,18; $D_{8=}$ 814,93; P_{37}= 688,53 ; P_5 = 558; P_{86} = 837,56; P_{47} = 715, 08 e P_{93} = 868,93).

n	Valores	f_i	f_a
1	525 |-- 580	8	
2	580 |-- 635	10	
3	635 |-- 690	18	
4	690 |-- 745	20	
5	745 |-- 800	17	
6	800 |-- 855	14	
7	855 |-- 910	9	
Σ		96	

3- Dada a tabela abaixo calcule: moda, Q_2, D_7, e P_{82}. (Mo = 75 kg, Q_2 = 74 kg, D_7 = 78,909 kg, e P_{82} = 81,8545 kg).

Ganho de peso de suínos.		
kg	f_i	f_a
59 |-- 63	3	
63 |-- 71	14	
71 |-- 83	22	
83 |-- 90	6	
Σ	45	

4-Os salários de 160 professores estão distribuídos conforme a tabela a seguir; determine o Q_1, D_4, e P_{85}. (Q_1 = 4, D_4 = 5,13, e P_{85} = 8,07)

Salários mínimos.		
Salário	f_i	f_a
1 |-- 3	20	
3 |-- 5	40	
5 |-- 7	60	
7 |-- 9	30	
9 |-- 11	10	
Σ	160	

5- Tomemos os resultados das estaturas e dos pesos de um mesmo grupo de indivíduos:

	$\bar{x}$	s
Estatura	175 cm	5,0 cm
Peso	68 kg	2,0 kg

Qual das medidas possui maior homogeneidade?

6-) Suponha que uma das famílias de Akhiok vá para outra cidade. As idades dos membros da família são 33, 34, 6, 58, 11, 6 e 56. Supondo que eles representam:

a) Uma amostra b) uma população

Calcule as medidas de:

- Posição: média, moda, mediana, 1º quartil, 3º quartil, 10º percentil e 90º percentil ;
- Dispersão: Amplitude total, variância, desvio padrão e coeficiente de variação (CV)

7-) Dada a distribuição amostral. Calcular a média, moda, mediana, desvio-padrão, e os quartis. (resp. $\bar{x}$ = 66,875; Mo = 41,429; S^2 = 1.021,62 ; S = 31,96; Q_1 = 40 ; md = Q_2 = 50; Q_3 = 90)

Salários (R$1.000,00)	30|-- 50	50|--100	100 |-- 150
Empregados	80	50	30

8-) Os tempos de reação de um indivíduo a determinados estímulos, medidos por um psicólogo, foram: 0,53 – 0,46 – 0,50 – 0,49 – 0,52 – 0,44 – 0,55 segundos. Determine o tempo médio e o desvio padrão de reação do indivíduo a esses estímulos. ($\bar{x} = 0{,}50seg.\ \sigma = 0{,}036s$)

9-) Dados os conjuntos de números: A={220, 230, 240, 250, 260} e B = {20, 30, 40, 50, 60}.

a) Calcule o desvio padrão do conjunto A; ($\boldsymbol{\sigma A = 14,14}$)
b) Calcule o desvio padrão do conjunto B; ($\boldsymbol{\sigma B = 14,14}$)
c) Que relação existe entre os desvios padrões dos dois conjuntos de números?

10-) Tomando-se os pedidos de combustível dos postos de certa região (20 postos) obtiveram-se os seguintes valores (em 1.000 litros): 20, 20, 21, 21, 21, 21, 22, 22, 22, 22,22, 22, 23, 23, 23, 23, 23, 24, 24, 26. Monte a distribuição de frequência e calcule a média, a moda, a mediana, desvio padrão.($\boldsymbol{\bar{x} = 22,25\ mil\ litros,\ Mo = 22\ mil\ litros, Md = 22\ mil\ litros}$)

11-) A renda média mensal na localidade A é de R$ 750,00 e na localidade B é de R$ 500,00.Os desvios padrões são R$ 100,00 e R$ 80,00. Faça uma análise comparativa quanto ao grau de homogeneidade da renda nestas duas localidades.

12-) O risco de uma ação de uma empresa pode ser devidamente avaliado através da variabilidade dos retornos esperados. Portanto, a comparação das distribuições probabilísticas dos retornos, relativas a cada ação individual, possibilita a quem toma decisões perceber os diferentes graus de risco. Analise os dados estatísticos relativos aos retornos de 5 ações descritas na tabela a seguir e diga qual é a menos arriscada:

Tabela – Avaliação do Risco das Ações – Empresa TECOL – 2010.

Discriminação	Ação A	Ação B	Ação C	Ação D	Ação E
Valor esperado	15%	12%	5%	10%	4%
Desvio padrão	6%	6,6%	2,5%	3%	2,6%
Coeficiente de Variação	0,40	0,55	0,50	0,30	0,65

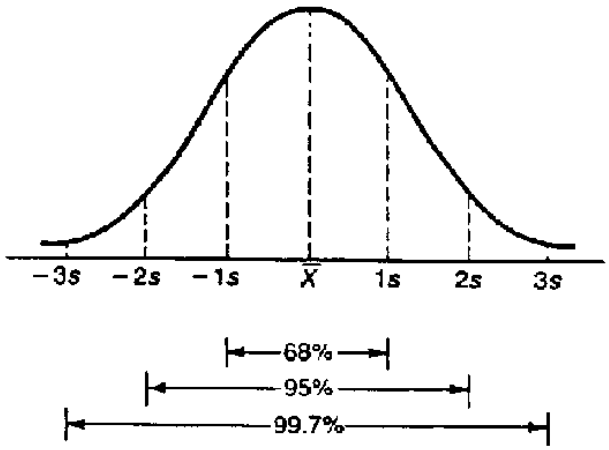
13-) A tabela a seguir indica as notas de uma turma, na disciplina de Matemática. Calcule o desvio padrão amostral para a média das notas destes alunos. Calcule a moda, mediana e as medidas de separatrizes (Q_1; Q_3 , D_5, P_{10} e P_{90}). ($\bar{x} = \frac{312}{50} = 6{,}24\ \ Samostral = 1{,}4957$)

Avaliação dos alunos da Escola Básica Lauro Bitencurcchi

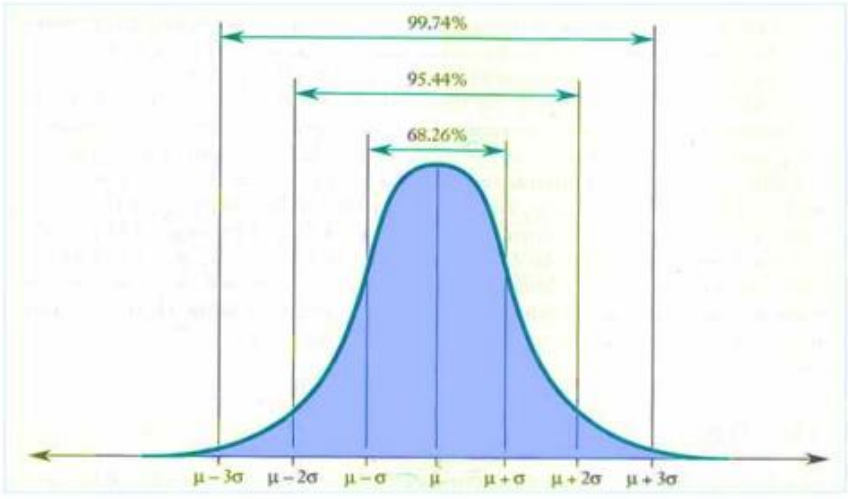
Notas	3|--4	4|--5	5|--6	6|--7	7|--8	8|--9	9|--10
f_i	4	7	9	15	10	3	2

REGRA EMPÍRICA

Para dados com distribuição (simétrica) na forma de sino, o desvio padrão tem as seguintes características. Aproximadamente 68% das medidas (dados) cairão dentro de um desvio padrão da média, 95% cairão dentro de dois desvios padrões, e 99,7% (ou quase 100%) ficam dentro de três desvios padrões. Ver a figura seguinte:



Por exemplo: Seja o peso médio de uma pessoa de 70 kg com um desvio padrão de 3,4 kg.
Então, 68% dos pesos ficam entre 66,6 e 73,4 kg, um desvio padrão, ou seja, (média + 1 desvio padrão) = (70 + 3,4) = 73,4, e (média - 1 desvio padrão) = 66,6. Noventa e cinco por cento (95%) dos pesos ficam entre 63,2 e 76,8 kg, dois desvios padrões. Noventa e nove e sete décimos de porcentagem (99.7%) ficam entre 59,8 e 80,2 kg, três desvios padrões. Veja a figura seguinte:



Uma curva em forma de sino simétrica, mostrando as Relações entre o Desvio Padrão e a Média.

ESCORE PADRÃO ou ESCORE $\mathcal{Z}$ ou $\mathcal{Z}$ SCORE

Podemos pegar qualquer ponto no eixo X da figura acima e descobrir quantos desvios padrões acima ou abaixo da média aquele ponto se encontra. Em outras palavras, um **Z** score representa o número de desvios padrões que uma observação (X) está acima ou abaixo da média. Quanto maior o valor de **Z**, mais distante o valor estará da média. Note que valores além de três desvios padrões são muito improváveis. Se um **Z** score for negativo, a observação (X) está abaixo da média. O **Z** score é encontrado usando a seguinte relação:

$$\mathcal{Z} = \frac{(x-\mu)}{\sigma}, \text{ onde}$$

$\mathcal{X}$ = valor dado;
μ = média;
σ = desvio padrão.