ESTATÍSTICA

para Psicologia – Parte 2

Luiz A. Bertolo

Cap 02 - Medidas Estatísticas

A distribuição de frequências permite-nos descrever, de modo geral, os grupos de valores (classes) assumidos por uma variável. Com ela, por exemplo, podemos localizar se a maior concentração de valores de uma dada distribuição se encontra no início, no meio, ou no final dos valores.

Quando confrontamos distribuições e queremos destacar as tendências de cada uma, isoladamente, necessitamos de conceitos que expressem através de números estas tendências. Esses conceitos são denominados **elementos típicos** da distribuição (ou estatísticas) e são:

- Medidas de Posição (locação ou tendência central)
- Medidas de Dispersão (variabilidade)
- Medidas de Assimetria
- Medidas de Curtose

2.1 – Medidas de Posição (ou tendência central)

Mostram o valor representativo em torno do qual os dados tendem a agrupar-se com maior ou menor freqüência.

A medida de tendência central é um número que está representando todo o conjunto de dados; nas pesquisas tal número pode ser encontrado a partir das medidas:

a) **média aritmética,**

b) **moda,**

c) **mediana.**

O uso de cada uma delas é mais conveniente de acordo com o nível de mensuração, o aspecto ou forma da distribuição de dados e o objetivo da pesquisa.

Outras medidas de posição são as **separatrizes**, que englobam:

- a própria mediana;
- os quartis;
- os percentis.

2.1.1 – Média Aritmética Simples ($\bar{x}$)

É a medida de centralidade mais comum, porém deve ser usada em dados representando variáveis quantitativas, pois não haveria sentido utilizá-la em uma distribuição em que a variável fosse, por exemplo, time de futebol ou sexo. A média representa, ainda, o ponto de distribuição no qual se equilibram as discrepâncias (diferenças) positivas e negativas de cada dado, ou seja, as discrepâncias positivas somadas se **anulam** com as negativas somadas.

2.1.1.1 dados **NÃO agrupados**

Definida da seguinte forma:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

é a soma de todos os números, dividida pelo número de parcelas. É uma das medidas de tendência central de maior emprego.

EX: 4 15 20 20 24 27 30 $\bar{x} = \frac{4 + 15 + 20 + 20 + 24 + 27 + 30}{7} = 20$

Observe que: (20-4) + (20-15) + (20 – 24) + (20 – 27) + (20 – 30) = 0

Média Aritmética Simples Dados Agrupados

Vamos dividi-los em duas categorias: sem intervalo de classe e com intervalo de classe.

2.1.1.2.1 – Sem intervalos de classe

Seja a distribuição de frequências associada a uma amostra de 34 famílias de quatro filhos, tomando para a variável o *número de filhos do sexo masculino*:

Nº de filhos	f_i
0	2
1	6
2	10
3	12
4	4
	Σ = 34

Neste caso, como as frequências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva à **Média Aritmética Ponderada.**

Média Aritmética Ponderada

É um tipo de média aritmética de vários valores com pesos diferentes, dada por:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_nx_n}{f_1 + f_n + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

f_i = frequência do valor x_i na amostra.

Nº de filhos	f_i	x_if_i
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
	Σ = 34	Σ = 78

Um modo rápido de obtermos a média ponderada é abrir, na tabela, uma coluna correspondente aos produtos x_if_i:

$$\bar{x} = = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} \Rightarrow \bar{x} = \frac{78}{34} = 2{,}29$$

Média Aritmética Simples Dados Agrupados

2.1.1.2.1 – Com intervalos de classe

Aqui, convencionamos que *todos os valores incluídos em um determinado intervalo de classe* ***coincidem*** *com o seu ponto médio*, e determinamos a média aritmética ponderada por meio da fórmula:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Onde x_i é o ponto médio da classe.
Consideremos a distribuição:

i	X	f_i	x_i	$x_i f_i$
1	150 |— 154	4	152	608
2	154 |— 158	9	156	1.404
3	158 |— 162	11	160	1.760
4	162 |— 166	8	164	1.312
5	166 |— 170	5	168	840
6	170 |—174	3	172	516
		Σ = 40		Σ = 6.440

$$\bar{x} = = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \Rightarrow \bar{x} = \frac{6.440}{40} = 161$$

A média aritmética simples pode ser vista como a média ponderada com todos os pesos iguais. Para efeito de nomenclatura sempre trataremos a média aritmética simples ou ponderada simplesmente por **média** representada por ($\bar{x}$).

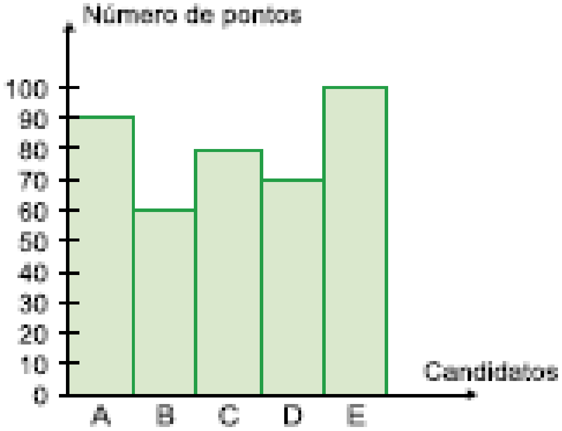
Exercícios de Aplicação 1

01. Temos um gráfico que nos mostra o desempenho dos 5 melhores classificados em um determinado concurso, no qual a pontuação varia de zero a cem pontos.

a) Qual é a soma dos pontos dos candidatos A, B, C, D e E?

b) Determine a média **aritmética** dos pontos dos candidatos discriminados no gráfico.

c) Mostre qual o candidato que fez mais e o que fez menos pontos.



Resposta:

a. 90 +60 + 80 + 70 + 100 = 400

b. $\bar{x} = \dfrac{90 + 60 + 80 + 70 + 100}{5} = 80 \text{ pontos}$

c. O candidato que fez mais pontos foi o candidato E (100 pontos), e o candidato que fez menos pontos foi o candidato B (60 pontos)

Exercício de Aplicação 2

Um professor de uma determinada disciplina resolveu que suas provas bimestrais terão pesos diferentes em cada bimestre e que seus alunos, só no final do 4º bimestre, receberão a média final. Escolhendo aleatoriamente um aluno desse professor, vamos, de acordo com suas notas e respectivos pesos, verificar sua média final.

O aluno no primeiro bimestre tirou 6 e a prova tinha peso 2, no 2º bimestre tirou 5 e o peso era 4, no 3º bimestre o aluno tirou 3 e o peso era 2 e, finalmente, no 4º bimestre tirou 10 e o peso era 4. Calcule sua média final.

Para resolver esse exercício, considere a média ponderada:

$$\bar{x}_p = \frac{6\,.2+5\,.4+3\,.2+10\,.4}{2+4+2+4} = \frac{78}{12} = 6{,}5$$

A média final do aluno foi 6,5.

Exercício de Aplicação 3

A tabela a seguir apresenta a distribuição de freqüências dos salários de um grupo de 50 empregados de uma empresa, num certo mês.

Número da classe	Salário do mês em reais	Número de empregados
1	1.000 ⊢ 2.000	20
2	2.000 ⊢ 3.000	18
3	3.000 ⊢ 4.000	9
4	4.000 ⊢ 5.000	3

O salário médio desses empregados, nesse mês, foi de:

a. R$ 2.637,00
b. R$ 2.500,00
c. R$ 2.420,00
d. R$ 2.400,00

Os valores centrais das classes 1, 2, 3 e 4 são, respectivamente, 1.500, 2.500, 3.500 e 4.500 reais, obtidos da seguinte maneira:

$$\left.\begin{aligned}
&\text{classe1} \Rightarrow \frac{1.000 + 2.000}{2} = \frac{3.000}{2} = 1.500\\
&\text{classe2} \Rightarrow \frac{2.000 + 3.000}{2} = \frac{5.000}{2} = 2.500\\
&\text{classe3} \Rightarrow \frac{3.000 + 4.000}{2} = \frac{7.000}{2} = 3.500\\
&\text{classe4} \Rightarrow \frac{4.000 + 5.000}{2} = \frac{9.000}{2} = 4.500
\end{aligned}\right\}$$

Encontramos a média aritmética simples dos limites das classes, para cada classe

Para determinar o **salário médio**, precisamos encontrar a <u>média aritmética ponderada</u> (os pesos serão as freqüências).

$$\bar{x}_p \frac{1.500 \,.\, 20 + 2.500 \,.\, 18 + 3.500 \,.\, 9 + 4.500 \,.\, 3}{20 + 18 + 9 + 3} \Rightarrow \bar{x}_p \frac{30.000 + 45.000 + 31.500 + 13.500}{50} = \frac{120.000}{50} = 2.400$$

Portanto, o salário médio é de R$ 2.400

Outras Médias

Média Geométrica (X_G)

É definida como a raiz de ordem n do produto desses números.

$$\bar{x}_G = \sqrt[n]{x_1 . x_2 . x_3 \ldots . x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Média Harmônica

É definida assim:

$$\bar{x}_H = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}} = \frac{1}{\sum \frac{1}{x}}$$

Exemplo de Aplicação 4

Calcule a média geométrica da série (2, 4, 8)

$$\bar{x}_G = \sqrt[n]{x_1 . x_2 . x_3 \ldots . x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Calcule a média harmônica da série (2, 4, 8)

$$\bar{x}_H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{1}{\sum \frac{1}{x}}$$

Médiana ($\tilde{x}$)

É o valor "do meio" de um conjunto de dados, quando os dados estão dispostos em ordem crescente ou decrescente, cortando, assim, a distribuição em duas partes com o mesmo número de elementos.

É também uma **medida separatriz** definida e exata, de fácil compreensão. Ela serve para análise comparativa e é representada por $\tilde{x}$.

Para dados **não agrupados** em classes:

Se n é **ímpar** → é o termo $\tilde{x} = \left(\dfrac{n+1}{2}\right)^{\circ} termo$

Se *n* é **par** → é o termo

$$\tilde{x} = \frac{\frac{n}{2} termo + (\frac{n}{2}+1) termo}{2}$$

EX1: Em um colégio, estão matriculados numa determinada classe 21 alunos. Durante o 1º bimestre foi feito um levantamento da freqüência destes alunos e foram observadas as seguintes faltas: 0, 0, 3, 5, 7, 9, 0, 1, 2, 3, 11, 2, 3, 5, 6, 4, 10, 12, 0, 1, 2. Qual a mediana das faltas? Dica: Primeiro construa o ROL.

Resposta: 3

EX2: As idades dos atletas amadores de uma determinada modalidade esportiva são 14, 12, 16, 13, 17, 16 anos. Encontre a mediana da série. Dica: Primeiro construa o ROL

Resposta: 15 anos

Médiana ($\tilde{x}$) cont...

2.1.3.2 – Dados agrupados

Se os dados se agrupam em uma *distribuição de frequência*, o cálculo da mediana se processa de modo muito semelhante àquele dos dados não-agrupados, implicando, porém, a determinação prévia das **frequências acumuladas**. Ainda aqui, temos que determinar um valor tal que divida a distribuição em dois grupos que contenham o mesmo *número de elementos*.

Para o caso de uma distribuição, porém, **a ordem**, a partir de qualquer um dos extremos, é dada por:

$$\boxed{\frac{\sum f_i}{2}}$$

2.1.3.2.1 – Sem intervalos de classe

Neste caso, é o bastante identificar a frequência acumulada **imediatamente superior** à metade da soma das frequências (ordem). A mediana será aquele valor da variável que corresponde a tal frequência acumulada.

Por exemplo,

Nº de filhos	f_i	FA_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\Sigma = 34$	

Sendo

$$\frac{\sum f_i}{2} = \frac{34}{2} = 17$$

A menor freqüência acumulada que supera esse valor é 18, que corresponde ao valor 2 da variável nº de filhos, sendo este o valor mediano. Logo, Md = 2 filhos.

Médiana ($\tilde{x}$) cont...

2.1.3.2.2 – Com intervalos de classe

Neste caso, o problema consiste em determinar o **ponto do intervalo** em que está compreendida a mediana. Para tanto, temos inicialmente que determinar a classe na qual se acha a mediana – **classe mediana**. Tal classe será, evidentemente, aquela correspondente à frequência acumulada imediatamente superior a $\frac{\sum f_i}{2}$.

Seja a distribuição:

i	X	f_i	F_i
1	150 |— 154	4	4
2	154 |— 158	9	13
3	158 |— 162	11	24 $\rightarrow$ Classe mediana
4	162 |— 166	8	32
5	166 |— 170	5	37
6	170 |—174	3	40

Temos:

$$\frac{\sum f_i}{2} = \frac{40}{2} = 20$$

Encontramos a Classe Mediana. E o valor da **MEDIANA** ?

Médiana ($\tilde{x}$) cont...

Existem três (03) maneiras de encontrarmos o valor da mediana quando os dados estão agrupados com intervalo de classe:

#01. Como há 24 valores incluídos nas três primeiras classes da distribuição e como pretendemos determinar o valor que ocupa o 20º lugar, a partir do início da série, vemos que este deve estar localizado na terceira classe (i = 3), supondo que as frequências dessas classes estejam uniformemente distribuídas.

Como há 11 elementos nessa classe e o intervalo de classe é igual a 4, devemos tomar, a partir do limite inferior, a distância:

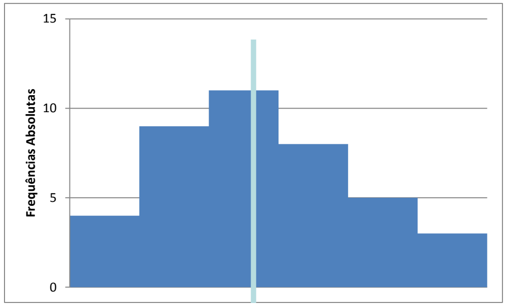
11 ....4 $\Rightarrow$ x = $\frac{7}{11}.4 = 2{,}54$

7 ... x

E a mediana será dada por:

$$\tilde{x} = Md = 158 + 2{,}54 = \mathbf{160,54}$$

#02. Poderíamos num histograma determinar graficamente a mediana como sendo aquele ponto do eixo das abcissas por onde passa a vertical que divide o histograma em duas áreas iguais:



4 . 4 + 4 . 9 + x . 11 = (4-x) . 11 + 4 . 8 + 4 . 5 + 4 . 3 $\Rightarrow$ 52 + 11x = 44 – 11x + 64 ou 22x = 56 $\Rightarrow$ x = 2,5454

Md = 158 + 2,5454 = **160,54.**

Médiana ($\tilde{x}$) cont...

#03. Existe, também, uma fórmula para calcularmos a mediana diretamente da tabela de distribuição de frequências:

$$Md = l_i^* + \frac{\left[\frac{\sum f_i}{2} - FA_{anterior}\right] h}{f^*}$$

Onde: l_i^* é o **limite inferior** da classe mediana;
$FA_{anterior}$ é a frequência acumulada da classe anterior à classe mediana;
f^* é a *frequência absoluta* da classe mediana;
h é a amplitude do intervalo da classe mediana.

No exemplo anterior:

$$Md = 158 + \frac{[20 - 13]4}{11} = \mathbf{160,54}$$

Exercícios de Aplicação

01. Encontre a mediana para as seguintes séries de dados:

. {35, 36, 37, 38, 40, 40, 41, 43, 46} $\Rightarrow \tilde{x} =$ 40

{12, 14, 14, 15, 16, 16,17, 20} $\Rightarrow \tilde{x} = \frac{15+16}{2}$ = 15,5

02. Em um colégio, estão matriculados numa determinada classe 21 alunos. Durante o 1º bimestre foi feito um levantamento da frequência destes alunos e foram observadas as seguintes faltas: 0, 0, 3, 5, 7, 9, 0, 1, 2, 3, 11, 2, 3, 5, 6, 4, 10, 12, 0, 1, 2. Qual a mediana $\tilde{x}$ das faltas? Resposta: 3

03. As idades dos atletas amadores de uma determinada modalidade esportiva são 14, 12, 16, 13, 17, 16 anos. Encontre a mediana da série. Resposta: 15 anos

04. Calcule a mediana da seguinte distribuição de frequências:

Custos (R$)	450 |-- 550	550 |-- 650	650 |-- 750	750 |-- 850	750 |-- 850	850 |-- 950	950 |-- 1.050	1.050 |-- 1.150
f_i	8	10	11	16	13	13	5	1

Média versus Mediana

A média é muito sensível a valores extremos de um conjunto de observações, enquanto a mediana não sofre muito com a presença de alguns valores muito altos ou muito baixos. A mediana é mais “robusta” do que a média.

Devemos preferir a **mediana** como medida sintetizadora quando o histograma do conjunto de valores é **assimétrico**

Ex.: { 200, 250, 250, 300, 450, 460, 510 }
$\bar{x}$=345,7 $\tilde{x}$=300
Tanto $\bar{x}$ como $\tilde{x}$, são boas medidas de posição.

Ex.: { 200, 250, 250, 300, 450, 460, 2300 }
$\bar{x}$ = 601 $\tilde{x}$ = 300
Devido ao valor 2300, $\tilde{x}$ é preferível a $\bar{x}$.

Moda e Classe Modal

É o valor que ocorre com maior freqüência em um conjunto de observações individuais. Para dados agrupados temos a **classe modal**. Em alguns casos pode haver mais de uma moda. Assim temos uma distribuição bimodal, trimodal, etc...

A moda é o valor em torno do qual os dados estatísticos tendem a estar mais pesadamente concentrados e é representada por Mo, também conhecida pelo nome de norma ou modo.

O termo moda foi introduzido por Pearson.

Exemplos para daados NÃO agrupados

01 - Em um grupo de pessoas cujas idades são: 3, 2, 5, 2, 6, 2, 4, 4, 2, 7, 2 anos, a moda é **2** anos (Mo = 2). Portanto, denomina-se **unimodal**.

02 - Algumas pessoas freqüentaram a escola por estes números de anos: 5, 3, 7, 5, 5, 8, 5, 3, 1, 1, 3, 3, 10, 3, 5. Nesta série de números, podem-se ter duas modas: Portanto **bimodal**.

Moda de Dados Agrupados

2.1.5.2.1 – Sem intervalos de classe

Ex. 3. Temos um grupo de pessoas cujas idades são: 3, 2, 5, 2, 6, 2, 4, 4, 2, 7, 2 anos:

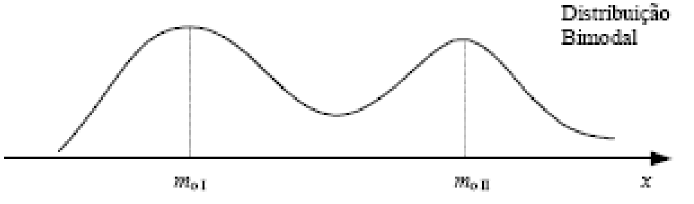
Idade	2	3	4	5	6	7
Freqüência	5	1	2	1	1	1

Fica claro que a moda é **2 anos**.

4. Tempo, em anos, que um grupo de pessoas freqüentou a escola.

Tempo de Escolaridade	
Tempo em anos de permanência na escola	**Freqüência**
1	2
3	5
5	5
7	1
8	1
10	1

Nesse exemplo, afirmamos que há duas modas, 3 e 5, portanto o conjunto de dados é **bimodal**.



Nota importante

Quando não houver repetição de números, não haverá moda (o conjunto de dados é **amodal**).

Moda de Dados Agrupados

2.1.5.2.2 – Com intervalos de classe

Quando os dados estão agrupados em classes,

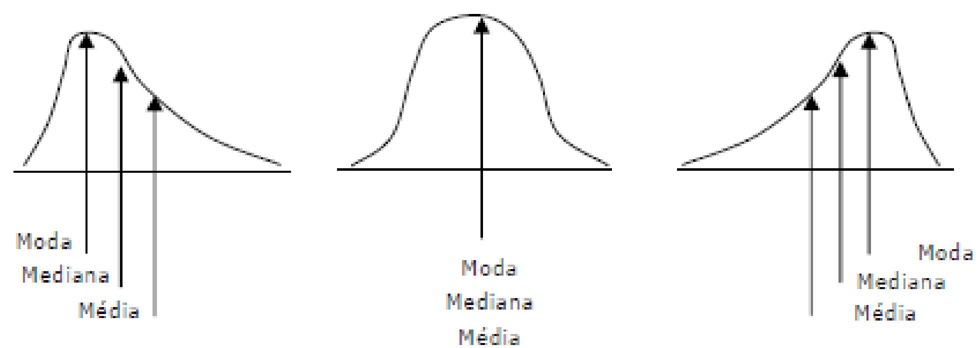
X	x_i	n_i	
10 |— 20	15	2	
20 |— 30	25	4	
30 |— 40	35	10	⇒ Classe Modal
40 |— 50	45	6	
50 |— 60	55	2	

Se precisarmos de um número representativo, tomamos o ponto médio do intervalo de classe.

Entretanto, temos a fórmula de **Czuber:**

$$M_o = l_i^* + \frac{D_1}{D_1 + D_2} h$$

$D_1 = f^* - f_{anterior}$ e $D_2 = f^* - f_{posterior}$



Exercícios de Aplicação de Moda

1. Considere os números 621, 310, 621, 201 e calcule:

a) a média aritmética ($\bar{x}$));

c) a moda (Mo).

Resposta

Primeiramente, monta-se a tabela de frequências:

Números	621	310	201
Freqüência	2	1	1

Números	frequência f_i	$x_i \cdot f_i$
621	2	1.242
310	1	310
201	1	201
Σ	4	1.753

a. $\bar{x} = (621 + 310 + 621 + 201)/4 = 1.753/4 = 438,25$

ou

$$\bar{x}_p = (621.2 + 310.1 + +201.1)/(2 + 1 + 1) = 1.753/4 = 438,25$$

c. Observando a tabela com os dados do exercício, verificamos que o número 621 aparece 2 vezes. Essa é a maior freqüência de acordo com a tabela, portanto Mo = 621.

Exercícios de Aplicação de Moda

2. Considere a tabela de frequência com os dados agrupados em intervalos de classe como mostrado na tabela abaixo e calcule a moda:

i	X	f_i	F_i
1	150 |— 154	4	4
2	154 |— 158	9	13
3	158 |— 162	11	24
4	162 |— 166	8	32
5	166 |— 170	5	37
6	170 |—174	3	40

classe Modal

aplicando a fórmula de **Czuber:**

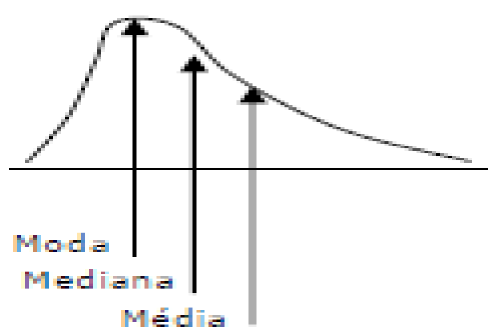
$$Mo = l_i^* + \frac{D_1}{D_1 + D_2} h$$

$D_1 = f^* - f_{anterior}$ e $D_2 = f^* - f_{posterior}$

temos:

$$M_0 = 158 + \frac{10 - 4}{(10 - 4) + (10 - 6)} = 158 + \frac{6}{10}$$
$$= 158{,}60$$

Para esta Tabela de Frequências com dados agrupados com intervalos de classe a média = 161 e mediana = 160,54 (encontrados anteriormente) e agora a moda = 158,60, mostra, claramente que os dados estão distribuídos assimetricamente com distorção (assimetria ou *skewness*) à esquerda.



Medidas Separatrizes

Como vimos, a mediana caracteriza uma série de valores devido à sua posição central. No entanto, ela apresenta outra característica, tão importante quanto a primeira: ela separa a série em dois grupos que apresenta o mesmo número de valores.

Assim há outras medidas que não são de tendência central, mas que estão ligadas à mediana. Essas medidas, juntamente com a mediana são chamadas *separatrizes*. São elas: os **quartis**, os **percentis** e os **decis**.

Percentis

- *"O percentil de ordem p, 0 ≤ p≤ 100, de um conjunto de valores dispostos em ordem crescente é um valor tal que p% das observações estão nele ou abaixo dele e (1 - p)% estão nele ou acima dele."*
- **Ex**: Para valores de 51 a 100, ordenados crescentemente:
- $P_{25} = 25$ deixa 25% dos dados (12,5 ⇒ 13 valores) nele ou abaixo dele e 75% dos dados (37,5 ⇒ 38 valores) nele ou acima dele. Assim: $P_{25} = 63$.
- Similarmente, P_{80} deixa 80% dos dados (40 valores) nele ou abaixo dele e 20% dos dados (10 valores) nele ou acima dele. Assim:

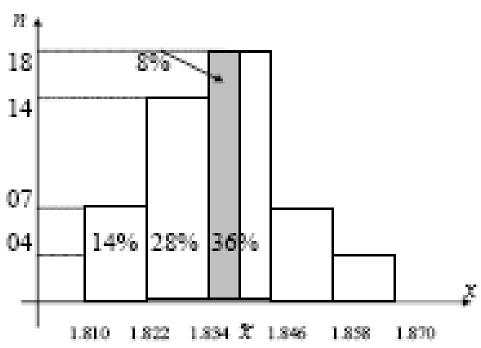
$$P_{80} = \frac{(90 + 91)}{2} = 90,5$$

Percentis de dados agrupados

Para dados agrupados em classes, os percentís podem ser obtidos por interpolação linear (regra de três simples).
Ex.: Dada a distribuição de freqüência de uma variável *X* qualquer:

X			x_i	N_i	N_i
1,810	|—	1,822	1,816	7	7
1,822	|—	1,834	1,828	14	21
1,834	|—	1,846	1,840	18	39
1,846	|—	1,858	1,852	7	46
1,858	|—	1,870	1,864	4	50

Temos que, para P_{50} (50% de 50) será o 25º elemento, está na terceira classe. Isto porque a segunda classe contém 21 elementos e a terceira, 39 elementos. Logo, o 25º elemento estará na 3ª



$$\bar{x} = (621 + 310 + 621 + 201)/4 = 1.753/4 = 438,25$$

Um outro processo gráfico pode ser usado para o cálculo desses percentis. (Veja Ogiva de Galton). Tal processo exige rigor no traçado e deve-se preferir papel milimetrado.
Obs.: As calculadoras geralmente não fornecem mediana e percentis.

Colocar os Exercícios Propostos da p. 10 da Apostila

2.2 – Medidas de Dispersão ou Variabilidade

- Vimos que a **moda**, a **mediana** e a **média aritmética** possuem a função de representar, a partir de <u>um único número</u>, a seqüência a ser analisada. Porém, tal método ainda é muito incompleto para que nós possamos tirar alguma conclusão sobre o trabalho. É necessário que possamos enxergar algo mais nessa seqüência que estamos analisando, como, por exemplo, certa "personalidade" da seqüência.

2.2 – Medidas de Dispersão ou Variabilidade – Cont...

- Observe a seguinte situação: quatro turmas, uma de cada um dos cursos Ciência da Computação, Matemática, Ciências Contábeis e Fisioterapia, fizeram uma prova de estatística e quando o professor verificou a média das notas de cada turma, constatou que, em cada uma das quatro turmas, a média dos alunos foi igual a 6,0. E aí? Será que podemos concluir que o desempenho das quatro turmas foi o mesmo? Será que todos os alunos, de todas as turmas, tiraram nota 6,0 na prova? É óbvio que, nesse momento, o bom senso fala mais alto e podemos, no mínimo, desconfiar de que não. Pois é exatamente aí que reside a tal "personalidade" que podemos atribuir a cada turma em relação ao comportamento das notas.

2.2 – Medidas de Dispersão ou Variabilidade – Cont...

- O que quero dizer é que, com as **medidas de dispersão**, seremos capazes de verificar que, por mais que a média das turmas na prova de estatística tenha sido 6,0, poderemos com tais medidas determinar as turmas que tiveram um comportamento homogêneo, em que os alunos tiraram notas próximas de 6,0, como também determinar as turmas que tiveram um comportamento heterogêneo em relação à nota 6,0, ou seja, por mais que a média tenha sido 6,0, as notas não foram próximas de 6,0. Em outras palavras, torna-se necessário estabelecer medidas que indiquem o grau de dispersão em relação ao valor central. Algumas medidas de dispersão que sintetizam essa variabilidade são:

2.2.1 – Amplitude (H)

- É uma medida de dispersão muito rápida e, ao mesmo tempo, **muito imprecisa**, pois consiste simplesmente em verificar a diferença entre o maior valor e o menor valor obtido na coleta de dados. Essa é nossa velha conhecida. Mesmo assim um exemplo

Pessoas	Peso (kg)
Agulha	30
Aderbal	15
Corá	55
Renato	52
Guilherme	60
Bruno	53
Bertolo	75
Alexandre	20
Fábio Thomáz	40

Na tabela ao lado, temos o peso das pessoas de um determinado grupo analisado e podemos verificar que a amplitude total foi de:AT = 75 – 15 = 60

2.2.2 – Desvio Médio

- Como a palavra desvio está associada à diferença, temos que, no contexto da nossa matéria, o desvio deve ser empregado com a diferença do elemento analisado em relação à média, ou seja, o quanto o elemento se afasta da média da seqüência. Daí é importante perceber que essa diferença deve ser necessariamente trabalhada em módulo, pois não tem sentido a distância negativa. E o desvio médio, então, passa a ser encontrado a partir da média aritmética de todos os desvios.

$$\text{Desvio Médio} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \cdots + |x_N - \bar{x}|}{N} = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N}$$

Desvio Médio - Exemplo

$$\text{Desvio Médio} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \cdots + |x_N - \bar{x}|}{N} = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N}$$

Com os dados do exercício anterior, temos:

$$\bar{x} = \frac{30 + 15 + 55 + 52 + 60 + 53 + 75 + 20 + 40}{9} = 44{,}4$$

$$\text{Desvio Médio} = \frac{|30 - 44{,}4| + |15 - 44{,}4| + |55 - 44{,}4| + |52 - 44{,}4| + |60 - 44{,}4| + |53 - 44{,}4| + |75 - 44{,}4| + |20 - 44{,}4| + |40 - 44{,}4|}{9}$$

$$= 16{,}17$$

E também porque é fácil ver que a soma dos desvios, é identicamente nula e que, portanto, não serve como medida de dispersão:

2.2.2 – Variância

- A variância é uma medida de dispersão muito parecida com o desvio médio, a única diferença em relação a este é que, na variância, ao invés de trabalharmos em <u>módulo</u> as diferenças entre cada elemento e a média, tomamos os <u>quadrados</u> das diferenças. Isso se dá pelo fato de que, elevando cada diferença ao quadrado, continuamos trabalhando com números não negativos, como também pelo fato de que, em procedimentos estatísticos mais avançados, tal método facilita futuras manipulações algébricas.

$$\text{Variância } \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2}{N} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$$

Variância - Exemplo

$$\text{Variância} = \frac{(30-44{,}4)^2+(15-44{,}4)^2+(55-44{,}4)^2+(52-44{,}4)^2+(60-44{,}4)^2+(53-44{,}4)^2+(75-44{,}4)^2+(20-44{,}4)^2+(40-44{,}4)^2}{9}$$

$$= 345{,}57$$

Desvio Padrão

- Para entendermos o procedimento para o cálculo do desvio-padrão, é interessante percebermos que, no cálculo da variância, tal como vimos no tópico anterior, cometemos um “erro técnico” que será corrigido pelo desvio-padrão, ou seja, no momento em que elevamos ao quadrado as dispersões (diferenças) de cada elemento em relação à média, automaticamente alteramos a **unidade** de trabalho. Por exemplo: se estivermos trabalhando com a coleta das alturas, em metro, das pessoas de uma determinada comunidade, a unidade da variância encontrada será o m^2 (metro quadrado), que representa áreas. E é aí que entra o desvio-padrão, ou seja, extraindo a raiz quadrada da variância.

$$\text{Desvio} - \text{padrão}\ \sigma = \sqrt{\text{Variância}}$$

Desvio Padrão - Exemplo

- Então, se no exemplo do item anterior a variância encontrada foi 345,57, temos que o desvio-padrão foi de

$$\sqrt{345,57} = 18,58$$

Observação: O uso do Desvio Médio pode causar dificuldades quando comparamos conjuntos de dados com números diferentes de observações: Exemplo: Em A = {3,4,5,6,7} temos o Desvio Médio (DM) como 6/5 = 1,2 e σ^2 = 10/5 = 2

Em D = {3,5,5,7} temos o Desvio Médio (DM) = 1,0 e σ^2 = 2

Assim, podemos dizer que, segundo o Desvio Médio, o grupo *D* é mais homogêneo (tem menor dispersão) do que *A*, enquanto que ambos têm a mesma homogeneidade segundo a variância. O desvio médio possui pequena utilização em estatística e em geral vale 0,8 vezes o desvio padrão .

2.2.4 – Momentos de uma distribuição de freqüências

Definimos o momento de ordem *t* de um conjunto de dados como:

$$M_t = \frac{\sum_{i=1}^{N} (x_i)^t}{N}$$

Definimos o momento de ordem *t* centrado em relação a uma constante *a* como

$$M_t = \frac{\sum_{i=1}^{N} (x_i - a)^t}{N}$$

Especial interesse tem o caso do momento centrado em relação a , dado por:

$$m_t = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^t}{N}$$

Mais Momentos

- Conforme já vimos nos casos da média e da variância, as expressões precedentes podem ser reescritas levando-se em consideração as freqüências dos diferentes valores existentes. Temos então respectivamente,

$$M_t = \frac{\sum_{i=1}^{N} (x_i)^t . f_i}{N}$$

$$M_t = \frac{\sum_{i=1}^{N} (x_i - a)^t . f_i}{N}$$

$$m_t = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^t . f_i}{N}$$

É fácil ver que $M_1 = \bar{x}$; $m_1 = 0$; $m_2 = \sigma^2$.

2.2.5 – Coeficiente de variação (CV)

O coeficiente de variação exprime a variabilidade em termos relativos. É uma medida adimensional e sua grande utilidade é permitir a comparação das variabilidades em diferentes conjuntos de dados.

$$CV = \frac{\sigma}{\bar{x}}$$

Exemplo: Testes de resistência à tração, aplicados a dois tipos diferentes de aço:

	$\bar{x}$ (kg/mm^2)	σ (kg/mm^2)
Tipo I	27,45	2,0
Tipo II	147,00	17,25

$CV_I = 2/27{,}45 = 7{,}29\%$
$CV_{II} = 17{,}25/145 = 11{,}73\%$
Assim, apesar do Tipo I ser menos resistente, é ele mais estável, mais consistente.