1. MEDIDAS DE TENDÊNCIA CENTRAL (OU DE POSIÇÃO)

Conceito: São aquelas que mostram o valor em torno do qual se agrupam as observações.

1.1 MÉDIA ARITMÉTICA ($\overline{X}$)

- Seja (x_1, x_2, ... , x_n), uma amostra de n dados:

$$\boxed{\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

- Se os dados estão apresentados em uma distribuição de freqüências com k classes cada uma com freqüência genérica f_j, j=1,2,...,k, então a média aritmética será calculada como:

$$\boxed{\overline{X} = \frac{\sum_{j=1}^{k} PM_j f_j}{\sum_{j=1}^{k} f_j} = \frac{\sum_{j=1}^{k} PM_j f_j}{n}}$$

Ex: Cálculo da média dos pesos dos terneiros da fazenda Canoas-SC, à partir dos dados originais:

$$\overline{x} = \frac{22 + 22{,}5 + ... + 29{,}5}{20} = \frac{501{,}5}{20} = 25{,}075\text{kg}$$

Cálculo da média dos pesos dos terneiros da fazenda Canoas-SC, à partir da distribuição de freqüências:

$$\overline{x} = \frac{22{,}75(4) + 24{,}25(5) + 25{,}75(8) + 27{,}25(1) + 28{,}75(2)}{20} = \frac{503}{20} = 25{,}15\text{kg}$$

1.2 MÉDIA PONDERADA ($\overline{X}$)

- Se aos dados (x_1, x_2, ... ,x_k), associam-se certos fatores de ponderação ou pesos (w_1, w_2, ... , w_k) que dependem do significado ou importância atribuída aos números, então temos:

$$\boxed{\overline{X} = \frac{\sum_{j=1}^{k} x_j w_j}{\sum_{j=1}^{k} w_j}}$$

Ex:

Notas de um aluno

PROVA	*Peso (w_j)*	*Nota*
1ª.	3	8,5
2ª.	1	7,5
3ª.	1	9,0

A média nesse caso deve ser a ponderada:

$$\bar{x} = \frac{3(8,5) + 1(7,5) + 1(9,0)}{3 + 1 + 1} = \frac{42}{5} = 8,4$$

enquanto que a média aritmética é:

$$\bar{x} = \frac{8,5 + 7,5 + 9,0}{3} = 8,3$$

PROPRIEDADES DA MÉDIA ARITMÉTICA

(a) A soma algébrica dos desvios de um conjunto de dados (x_1, x_2, ... ,x_n), em relação à média aritmética, é zero.

Ex: 4 5 6 $\rightarrow \bar{x} = 5 \quad \therefore (4-5)+(5-5)+(6-5) = -1+0+1 = 0$

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Verificando:

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - n\frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0 \text{ cqv}$$

(b) A soma dos quadrados dos desvios de um conjunto de dados X_i, i=1, 2, ... ,n em relação a qualquer número '**a**', é um mínimo quando **a**= $\bar{X}$ e somente nesse caso.

Verificando:

$f(a) = \sum_{i=1}^{n}(X_i - a)^2$, que derivando e igualando a zero, temos:

$$f'(a) = 2\sum_{i=1}^{n}(X_i - a)(-1) = 0$$

$$\sum_{i=1}^{n} X_i - na = 0$$

$$a = \frac{\sum_{i=1}^{n} X_i}{n}\text{, ou seja, } a = \overline{X} \quad \text{cqv}$$

Ex: 4 5 6, cuja média é 5:

$(4-5)^2 + (5-5)^2 + (6-5)^2=1+0+1=2$

Em torno de outro valor: $(4-4)^2 + (5-4)^2 + (6-4)^2=0+1+4=5$

(c) Se f_1 dados têm média m_1, f_2 dados têm média m_2, ... , f_k dados têm média m_k, a média de todos os dados é:

$$\boxed{\overline{X} = \frac{\sum_{j=1}^{k} f_j m_j}{\sum_{j=1}^{k} f_j}}$$

(média ponderada das médias)

Ex: Dados de precipitação média de três cidades, em pluviômetros espalhados em pontos de cada uma cidade.

Cidade	Média de precipitação (mm)	Número de pontos
Piracicaba	200	5
Limeira	300	10
Campinas	150	6

Qual foi a média de chuva na região?

Ex:

$$X:\{\underbrace{4\ \ 5\ \ 6}_{5}\ \ \underbrace{4\ \ 2}_{3}\ \ \underbrace{7\ \ 2}_{4,5}\}$$

Médias

DESVANTAGEM DA MÉDIA – é sensível a valores extremos

Ex: Considere os dados: 5, 6, 8, 4, 7, 100

$$\overline{x} = 21,6666$$

1.3 MEDIANA (Md)

- É o valor tal que metade do conjunto de dados é igual ou inferiores a esse valor e a outra metade é igual ou superior a esse valor. É o valor que divide o conjunto de dados ao meio.

-Se n é IMPAR

Md é o valor central

Ex: Peso de terneiros (kg)

22 23 21 20 25 Rol $\longrightarrow$ 20 21 22 23 25

md=22

Posição da mediana em uma amostra de tamanho n impar: $\frac{n+1}{2}$

-Se n é PAR

Md é a média aritmética dos dois valores centrais

Ex: 5 3 7 6 2 10 Rol $\longrightarrow$ 2 3 5 6 7 10

md=$\frac{5+6}{2} = 5{,}5$

Posições dos valores para o cálculo da mediana em uma amostra de tamanho n par:

$\frac{n}{2}$ e $\frac{n+2}{2}$

Ex: Peso dos terneiros (kg), n=20, par. A mediana é a média dos pesos dos animais de posições no rol: 20/2=10^{o}. e (20+2)/2=11^{o}. Portanto, como o rol dos dados ordenados de forma crescente é:

22,0	22,5	23,3	23,4	23,5
24,0	24,5	24,5	24,7	25,0
25,0	25,5	25,6	25,7	25,8
25,9	26,0	27,1	28,0	29,5

logo a md=$\frac{25+25}{2} = 25$kg

CÁLCULO DA MEDIANA ATRAVÉS DA DISTRIBUIÇÃO DE FREQUÊNCIA

PASSOS: (1^{o}) Localizar a classe mediana (h-ésima)

$$(2^o)\ Md = LI_h + \left(\frac{\frac{n}{2} - \sum_{j=1}^{h-1} f_j}{f_h} \right) i_h$$

onde:

LI_h: é o limite inferior da classe mediana;

n: é o número de dados observados;

$\sum_{j=1}^{h-1} f_j$: é a soma de todas as freqüências das classes inferiores à classe mediana;

f_h: é a freqüência da classe mediana;

i_h: é a amplitude do intervalo de classe da classe mediana.

Ex: Distribuição de freqüências do peso de terneiros (kg) da raça crioula preta, Fazenda Canoas-SC, 1985.

Classes Peso (kg)	**Freqüência Absoluta**
22,0 |-- 23,5	4
23,5 |-- 25,0	5
25,0 |-- 26,5	8
26,5 |-- 28,0	1
28,0 |--| 29,5	2
Total	20

$$md = 25 + \left(\frac{\frac{20}{2} - (4+5)}{8} \right) 1{,}5 \ = \ 25{,}1875 kg$$

VANTAGEM DA MEDIANA EM COMPARAÇÃO COM A MÉDIA – não é sensível a valores extremos. Não é afetada em presença de valores extremos.

Ex: 5, 6, 8, 4, 7, 100 Rol $\longrightarrow$ 4, 5, 6, 7, 8, 100

$$md = \frac{6+7}{2} = 6{,}5$$

1.4 MODA (Mo)

- É o valor que ocorre com maior freqüência no conjunto de dados. É o valor mais comum, se existir.

  -A moda pode NÃO EXISTIR, Ex: {2, 3, 5, 6, 7, 10}

  -Uma ÚNICA MODA (unimodal), Ex: {1, 2, 2, 3, 4} mo=2 **com freqüência modal 2.**

  -DUAS MODAS (bimodal)

  Ex_1: {1, 1, 2, 3, 4, 4} mo=1 e mo=4 com freqüência modal 2.

  Ex_2: {1, 1, 2, 4, 3, 1, 4, 4} mo=1 e mo=4 com freqüência modal 3.

  -MAIS DE DUAS MODAS (multimodal)

  Ex: {1, 1, 1, 2, 2, 2, 3, 4, 4, 4} mo=1, mo=2 e mo=4 com freqüência modal 3.

Ex: Peso vivo dos terneiros (kg), n=20, par. Portanto, como o rol dos dados ordenados de forma crescente é:

22,0	22,5	23,3	23,4	23,5
24,0	24,5	24,5	24,7	25,0
25,0	25,5	25,6	25,7	25,8
25,9	26,0	27,1	28,0	29,5

Temos mo=24,5 e mo=25,0 com freqüência modal 2.

CÁLCULO DA MODA ATRAVÉS DA DISTRIBUIÇÃO DE FREQUÊNCIAS

PASSOS: (1^o) Localizar a classe modal (h-ésima).

Obs. Se as amplitudes de classes forem diferentes ($i \neq$'s), calcular todos os $\left(\frac{f_j}{i}\right)$ e tomar como classe modal aquela que fornecer o maior valor desse quociente.

(2^o) $$Mo = LI_h + \left[\frac{f_h - f_{h-1}}{2f_h - f_{h-1} - f_{h+1}}\right] i_h$$

onde:

LI_h: é o limite inferior da classe modal;

f_h: é a freqüência da classe modal;

f_{h-1}: é a freqüência da classe imediatamente inferior à classe modal;

f_{h+1}: é a freqüência da classe imediatamente superior à classe modal;

i_h: é a amplitude do intervalo de classe da classe modal.

Ex: Distribuição de freqüências do peso de terneiros (kg) da raça crioula preta, Fazenda Canoas-SC, 1985.

Classes Peso (kg)	**Freqüência Absoluta**
22,0 |-- 23,5	4
23,5 |-- 25,0	5
25,0 |-- 26,5	8
26,5 |-- 28,0	1
28,0 |--| 29,5	2
Total	20

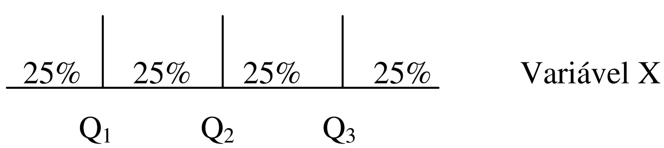
Ex: Intervalos de classes diferentes

Classes Peso (kg)	**Freqüência Absoluta**	**f_j/i**
22,0 |-- 24,0	5	2,5
24,0 |-- 27,0	8	2,6666
27,0 |-- 29,0	10	5
29,0 |-- 33,0	15	3,75
33,0 |--| 35,0	8	4
Total	46	

$$mo = 25 + \frac{8-5}{2\times 8-5-1}1{,}5 = 25{,}45kg$$

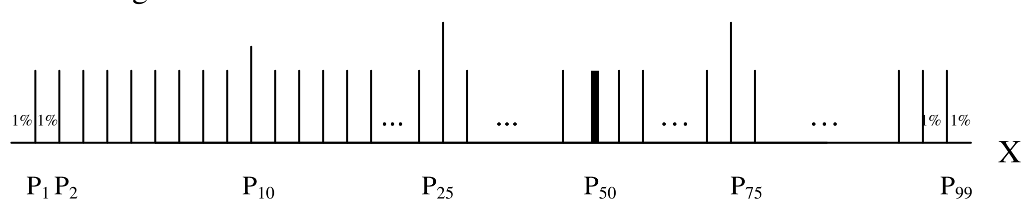
A classe modal é a terceira.

1.5 QUARTIS – são separatrizes que dividem o conjunto de dados em quatro partes iguais.



$$Q_1 \Rightarrow \frac{n}{4} \qquad Q_2 \Rightarrow \frac{n}{2} \qquad Q_3 \Rightarrow \frac{3n}{4}$$

1.6 PERCENTIS – são separatrizes que dividem o conjunto de dados em cem partes iguais



$$Pr \Rightarrow \frac{r\times n}{100}$$

Se np for inteiro $P_{100p} = \dfrac{x_{[np]} + x_{[np+1]}}{2}$

Se np for não inteiro $P_{100p} = x_{[\text{int}[np]+1]}$, onde int é a função inteiro. Ex.int[5,23]=5.

Ex: Um agrônomo precisa fazer um desbaste em mudas de mamão que estão em uma estufa, retirando 15% das mudas menores. Qual a altura abaixo da qual todas as mudas serão removidas?

Altura de mudas de mamão em cm

10,3	11,0	11,9	12,6	13,1
13,5	14,9	15,3	16,0	16,9
17,7	18,1	18,7	19,9	20,2
20,8	23,5	23,8	24,2	24,5
24,6	26,0	26,3	26,8	28,3
28,5	29,3	30,2	31,7	32,2

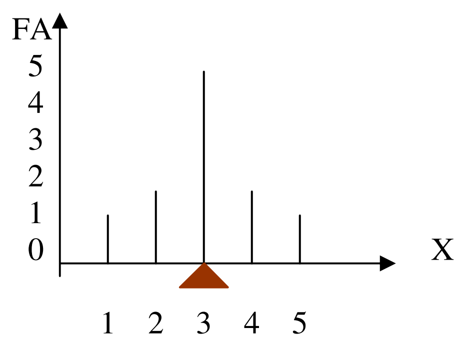
P_{15}=?

n = 30 p=0,15 np=4,5 $\therefore P_{15}=x_{[\text{int}[4,5]+1]}=x_{[5]}=13{,}1$cm

RELAÇÃO ENTRE A MÉDIA, MODA E MEDIANA e a ASSIMETRIA DA DISTRIBUIÇÃO

a) DISTRIBUIÇÃO DE FREQUÊNCIA SIMÉTRICA

X	f
1	1
2	2
3	5
4	2
5	1
	11



$$\bar{x} = \frac{1(1)+2(2)+3(5)+4(2)+5(1)}{11} = \frac{33}{11} = 3$$

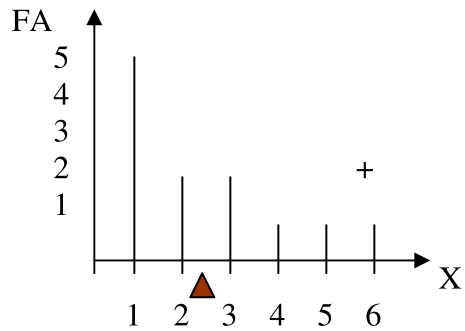
mo=3

md=3

$\therefore$ $\boxed{\bar{X} = Mo = Md}$

b) DISTRIBUIÇÃO DE FREQUÊNCIA ASSIMÉTRICA À DIREITA (ASSIMETRIA POSITIVA) CASO UNIMODAL

X	f
1	5
2	2
3	2
4	1
5	1
6	1
	12



$$\bar{x} = \frac{1(5)+2(2)+3(2)+4(1)+5(1)+6(1)}{12} = 2,5$$

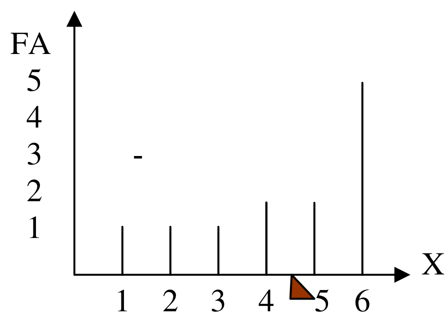
md=2

mo=1

$\therefore$ $\boxed{Mo < Md < \bar{X}}$

c) DISTRIBUIÇÃO DE FREQUÊNCIA ASSIMÉTRICA À ESQUERDA (ASSIMETRIA NEGATIVA) CASO UNIMODAL

X	f
1	1
2	1
3	1
4	2
5	2
6	5
	12



$$\bar{x} = \frac{1(1)+2(1)+3(1)+4(2)+5(2)+6(5)}{12} = 4,5$$

md=5,0

mo=6,0

$\therefore$ $\boxed{\bar{X} < Md < Mo}$

Ex: Peso vivo dos terneiros ao nascer

$$\begin{cases} \bar{x} = 25,15kg \\ md = 25,18kg \\ mo = 25,45kg \end{cases}$$

$\therefore$ O Peso dos terneiros apresenta uma assimetria levemente negativa.

Obs_1: Na distribuição simétrica, se o histograma for apoiado na mediana, haverá equilíbrio. Mas o ponto de apoio é a média que nesse caso é igual a mediana.

Obs_2: Na distribuição positivamente assimétrica, se o histograma for apoiado na mediana, ele penderá para direita. O peso – área do histograma- está igualmente dividido entre os dois lados, mas o momento no sentido horário é maior, pois os "braços de alavanca" dados por $[X_j\text{-}Md]$ são maiores, ou seja o "Torque" é maior. O ponto de apoio está mais à direita, na média.

Obs_3: Na distribuição negativamente assimétrica, se o histograma for apoiado na mediana, ele penderá para esquerda. O peso – área do histograma- está igualmente dividido entre os dois lados, mas o momento no sentido anti-horário é maior, pois os "braços de alavanca" dados por $[X_j\text{-}Md]$ são maiores, ou seja o "Torque" é maior. O ponto de apoio está mais à esquerda, na média.

Obs_4: A distribuição de variáveis econômico-sociais é geralmente assimétrica:

Ex: - Distribuição dos SALÁRIOS
- Distribuição da RENDA
- Distribuição dos ESTABELECIMENTOS AGROPECUÁRIOS CONFORME SUA ÁREA NO BRASIL.

Obs_5: Qual das medidas de tendência central deve ser usada?

R. Depende do enfoque de análise.

Ex: Dissídio salarial em uma empresa $\left(Mo<Md<\overline{X} \right)$

- EMPRESÁRIO
- REPRESENTANTE DOS TRABALHADORES

OUTROS TIPOS DE MÉDIAS

1. Média Geométrica

a média geométrica de n valores não-negativos $\left(X_1, X_2, \ldots, X_n \right)$ é:

$$G = \sqrt[n]{X_1.X_2.\cdots.X_n} = \sqrt[n]{\prod_{i=1}^{n} X_i}$$

Obs: $\boxed{G \le \overline{X}}$ e $G = \overline{X}$ se todos os X_i forem iguais.

Ex_1: { 5 8 3 7 2}

$$g = \sqrt[5]{5\times8\times3\times7\times2} = \sqrt[5]{1680} \approx 4{,}4163$$

Ex_2: A quantidade de laranjas em um pomar, aumentou de 1.000 para 4.000 em três dias. Qual foi a porcentagem média de acréscimo por dia?

Sol.: Seja r o acréscimo:

- Quantidade de laranjas depois do 1º dia = 1.000 + 1.000r = 1.000(1+r)
- Quantidade de laranjas depois do 2º dia = 1.000(1+r) + 1.000(1+r)r=1.000$(1+r)^2$
- Quantidade de laranjas depois do 3º dia = 1.000$(1+r)^2$+1.000$(1+r)^2$r=1.000$(1+r)^3$

$$\therefore \quad 1.000(1+r)^3=4.000$$

$$(1+r)^3=4$$

$$1+r=\sqrt[3]{4}$$

$$r=\sqrt[3]{4}-1$$

$$r=0{,}587 \quad \text{ou} \quad 58{,}7\%$$

Assim, se a partir de uma quantidade P, que cresce a uma taxa constante r por unidade de tempo, tem-se após n unidades de tempo, um total

$$A = P(1+r)^n \quad \text{ou} \quad \boxed{r = \sqrt[n]{\frac{A}{P}} - 1} \quad \text{média geométrica}$$

2. Média Harmônica

a média harmônica de n valores diferentes de zero $(X_1, X_2, \cdots, X_n)$ é o inverso da média aritmética dos inversos dos valores.

$$H = \left(\frac{\sum_{i=1}^{n} X_i^{-1}}{n}\right)^{-1} \quad \text{ou} \quad H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

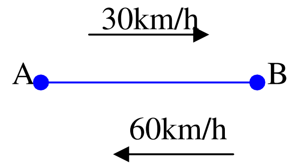
$$H \le G$$

$$H \le G \le \overline{X}$$

Ex_1: { 5 8 3 7 2}

$$h = \left(\frac{5^{-1}+8^{-1}+3^{-1}+7^{-1}+2^{-1}}{5}\right)^{-1} \approx 3{,}8426$$

Ex_2: Uma máquina agrícola realiza uma operação entre dois pontos: A----B.



Qual é a velocidade média da máquina para o percurso?

Sol. Supor que a distância entre os dois pontos seja de 60km.

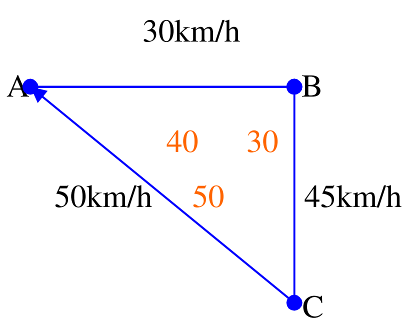
$$\text{Tempo } \overline{AB} = \frac{60km}{30km/h} = 2h$$

$$\text{Tempo } \overline{BA} = \frac{60km}{60km/h} = 1h$$

$$\therefore \text{ velocidade} = \frac{120km}{3} = 40km/h$$

$$h = \frac{2}{\frac{1}{30} + \frac{1}{60}} = 40km/h$$

Obs: Se as distâncias (d_i) não são iguais, deve-se ponderar as velocidades por essas distâncias:



$$H = \frac{k}{\sum_{i=1}^{n} \frac{d_i}{X_i}}$$ onde: $k = \sum_{i=1}^{n} d_i$ e n é o tamanho da amostra.

$$h = \frac{120}{\frac{40}{30} + \frac{30}{45} + \frac{50}{50}} = 40km/h$$

3. Média Quadrática ou Raiz da Média Quadrática

a média quadrática de n valores $(X_1, X_2, \cdots, X_n)$ é a raiz quadrada da média dos quadrados dos valores. É muito usada em estatística (desvio padrão) e física aplicada.

$$MQ = RMQ = \sqrt{\frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}} \quad \text{ou} \quad MQP = \sqrt{\frac{X_1^2(f_1) + X_2^2(f_2) + \cdots + X_k^2(f_k)}{\sum_{i=1}^{k} f_k}}$$

Ex: {3, 5, 6, 6, 7, 10, 12 }

$$mq = \sqrt{\frac{3^2 + 5^2 + 6^2 + 6^2 + 7^2 + 10^2 + 12^2}{7}} = \sqrt{57} \approx 7{,}5493 \qquad H \le G \le \overline{X} \le MQ$$

4. Média Móvel

Seja v_t (v_1, v_2, ... , v_n) os valores de uma série temporal (Preços mensais de um produto ou a série de produções de um bem econômico). Os valores da média aritmética móvel de k termos são obtidos somando k termos consecutivos da série e dividindo por k.

Valores da média móvel:

$$\frac{v_1 + v_2 + \cdots v_k}{k}$$

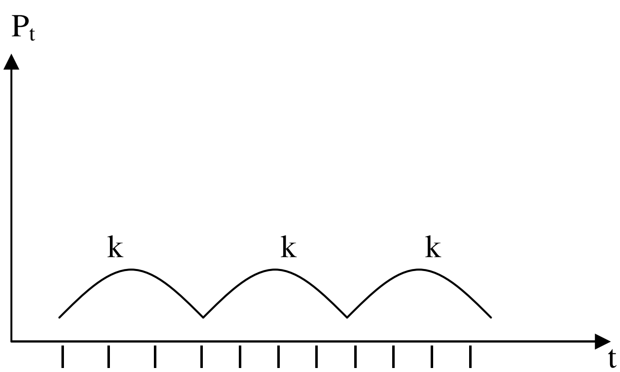
$$\frac{v_2 + v_3 + \cdots v_{k+1}}{k}$$

$$\frac{v_3 + v_4 + \cdots v_{k+2}}{k}$$

$$\vdots$$

$$\frac{v_{n-k+1} + \cdots + v_{n-1} + v_n}{k}$$

Obs: k geralmente é o número de termos que corresponde a um período de flutuação



Bimestral

Trimestral

Ex: Preços anuais: {2, 6, 7, 2, 6, 7, 2, 6, 7}

k=3anos

Mm={5, 5, 5, 5, 5, 5, 5}

O número de termos na série de médias móveis é $\boxed{n-k+1}$

Obs: Interessa comparar os valores da série temporal dada com os valores da média móvel. Para isso é necessário definir o valor da média móvel correspondente a um valor v_t qualquer da série dada.

1º CASO) k IMPAR:

Qualquer número impar pode ser escrito como $2\lambda+1$, $\lambda \in I^+$. Assim, a média móvel correspondente a v_t é:

$$M_t = \left(\frac{v_{t-\lambda} + v_{t-\lambda+1} + \cdots + v_t + \cdots + v_{t+\lambda-1} + v_{t+\lambda}}{k} \right)$$

Ex: k=3anos $\Leftrightarrow \lambda = 1$

$$M_t = \left(\frac{v_{t-1} + v_t + v_{t+1}}{3} \right), \quad t \geq 2$$

2º CASO) k PAR:

Qualquer número par pode ser escrito como 2λ, $\lambda \in I^+$. Um valor qualquer da média móvel não corresponde exatamente a nenhum dos termos da série dada.

Ex: {4 5 6 7| 5 7 8 9| ...}   k=4

A média móvel corresponderá a um intervalo e não a um v_t da série de dados.

Solução: Média móvel centralizada de k= 2λ termos correspondente ao valor v_t é:

$$M_t = \left(\frac{0{,}5v_{t-\lambda} + v_{t-\lambda+1} + \cdots + v_t + \cdots + v_{t+\lambda-1} + 0{,}5v_{t+\lambda}}{k} \right)$$

Ex: {4 5 6 7}

Se k=2, então $\lambda = 1$

$$M_t = \left(\frac{0{,}5v_{t-1} + v_t + 0{,}5v_{t+1}}{2} \right)$$

$$M_2 = \left(\frac{0{,}5 \times 4 + 5 + 0{,}5 \times 6}{2} \right) = 5$$

$$M_3 = \left(\frac{0{,}5 \times 5 + 6 + 0{,}5 \times 7}{2} \right) = 6$$

Obs: O número de médias móveis é: $\boxed{\frac{n-k}{2} + 1}$