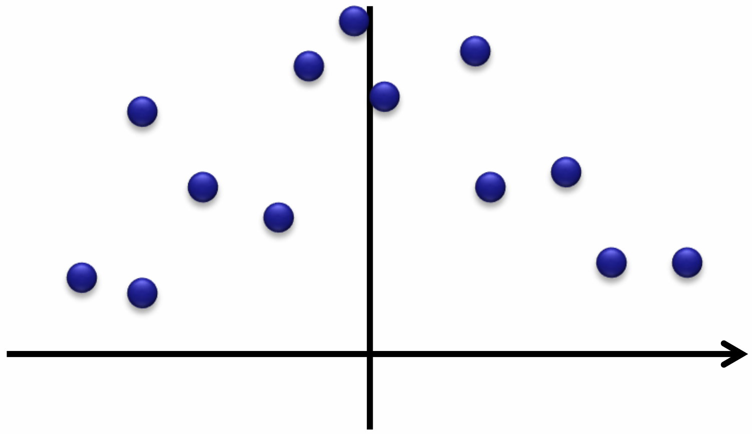
Estatística Descritiva

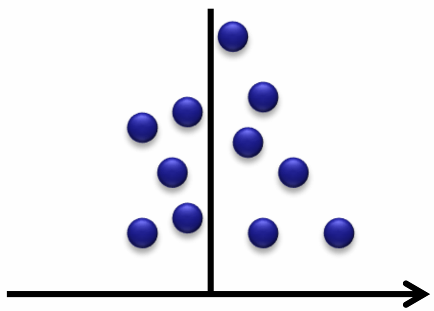
Tendência Central

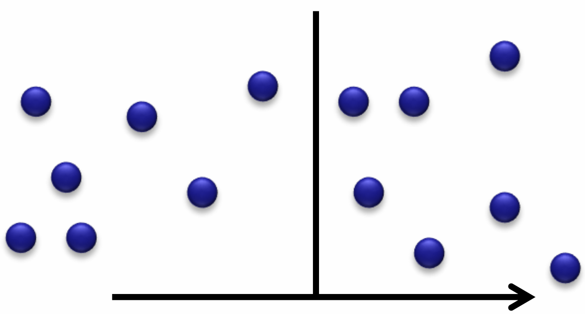
- Medidas de Tendência Central: Mostram o valor representativo de um conjunto de dados, em torno do qual os demais dados se distribuem.



Variação

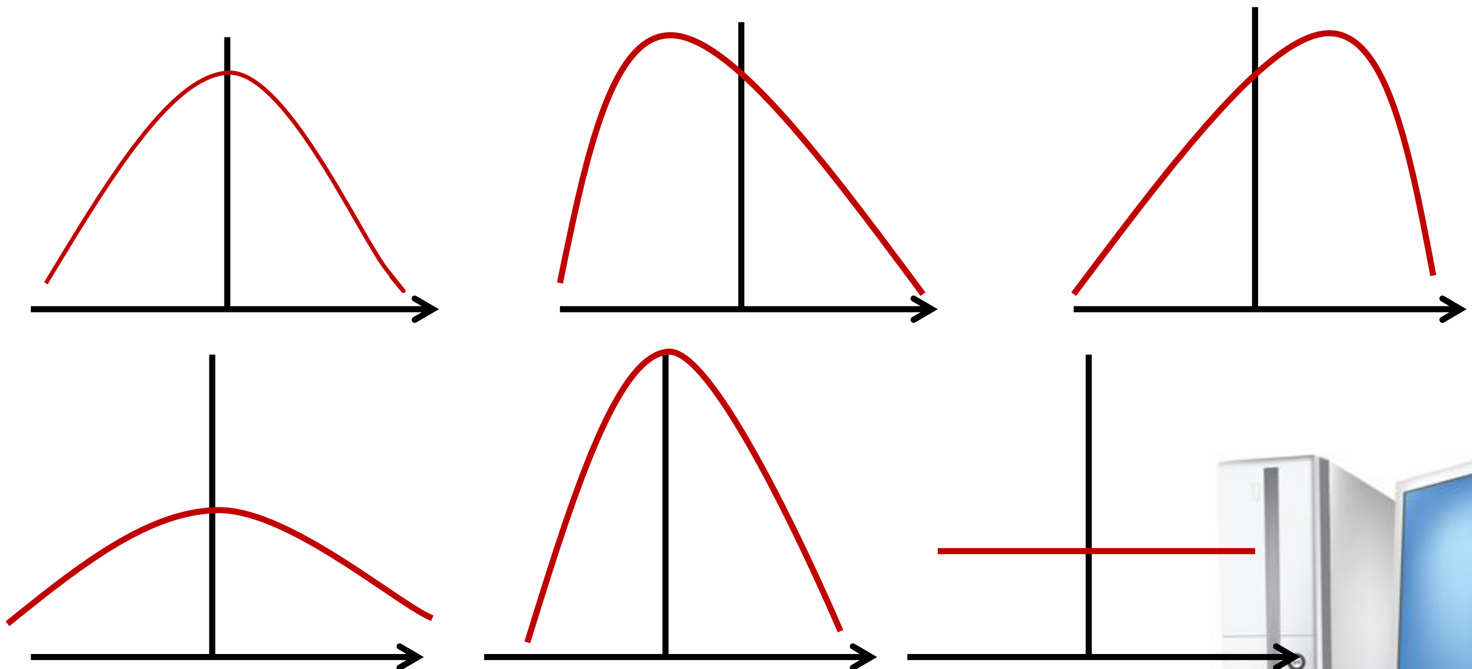
- Medidas de Variação: Mostram como os dados se distribuem em torno de um valor central; mostra se os dados estão concentrados ou espalhados em torno de um valor representativo.





Formato

- Medidas de Formato: Mostram o formato da distribuição dos dados:



Medidas de Tendência Central

- A *média aritmética* $\bar{x}$ ou simplesmente ***Média*** de uma **amostra** de *n* itens pode ser obtida pela seguinte expressão:

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

Média Aritmética

- A *média aritmética* μ ou simplesmente **Média** de uma **população** de *N* itens pode ser obtida pela seguinte expressão

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Exemplo – Calcule a Média

- A tabela a seguir mostra o tempo que você leva para se arrumar para sair de casa no período da manhã, desde o momento que acorda até a saída de casa, durante 10 dias:

Dia	Tempo (minutos)
1	39
2	29
3	43
4	52
5	39
6	44
7	40
8	31
9	44
10	35

Média: Observação

- A média é uma medida muito sensível ao valores extremos de uma distribuição de dados:

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

$$\bar{x} = \frac{1}{4}(1 + 2 + 3 + 40) = 11,5$$

Exemplo

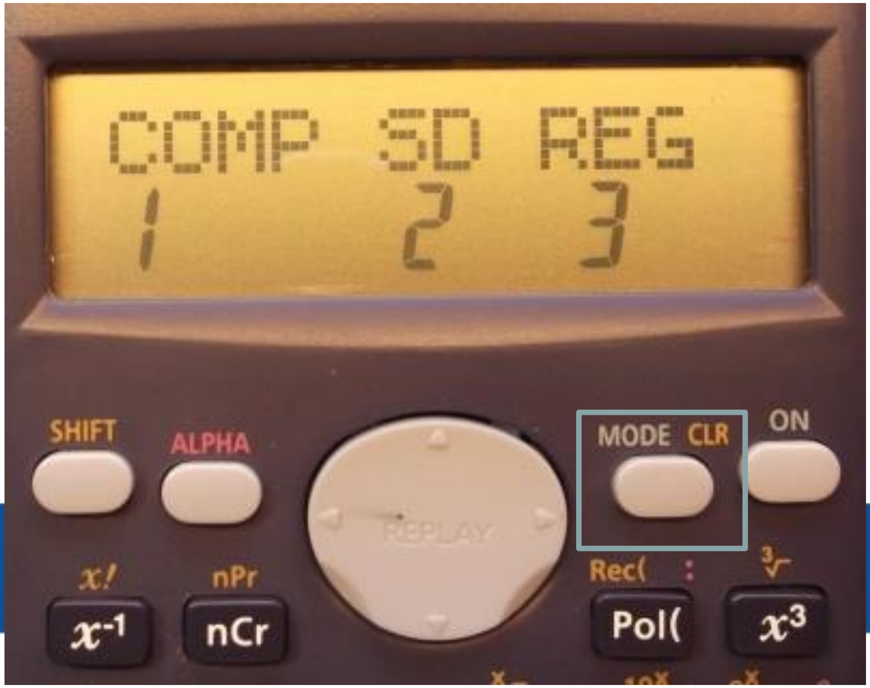
- A tabela a seguir mostra os salários pagos em uma empresa:

Função	Salário (R$)
Gerente	30.000
Funcionário 1	600
Funcionário 2	600
Funcionário 3	600
Funcionário 4	600

- A empresa alega em um anúncio que o salário médio é de aproximadamente R$ 6000,00. Isto representa a realidade?

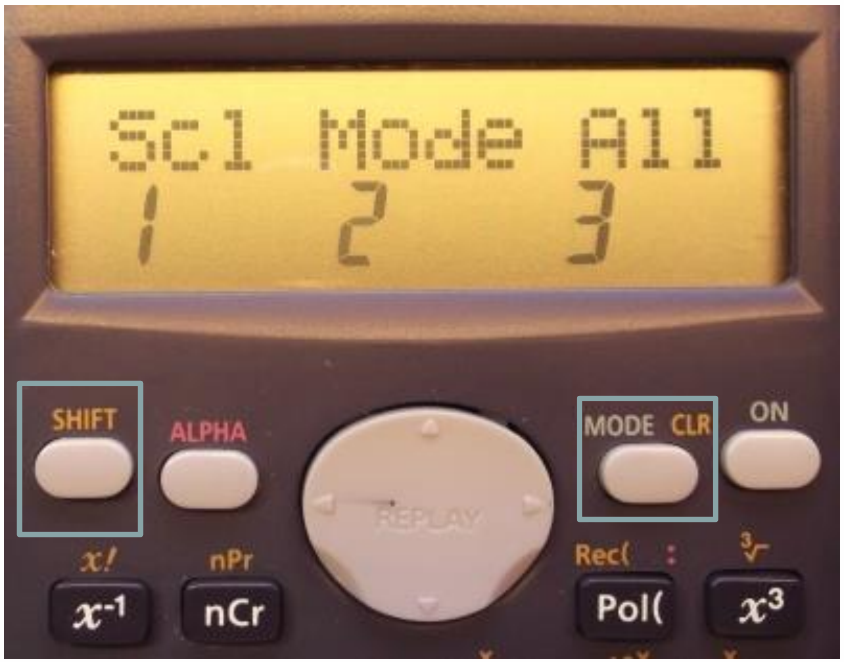
Cálculo da Média com Calculadora

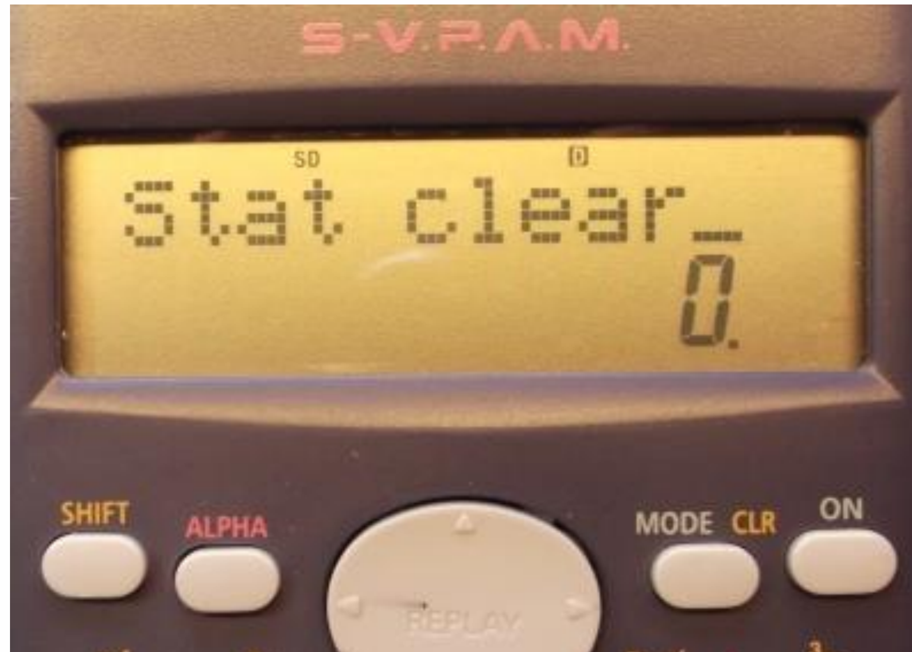
- Selecionar o modo Estatístico: MODE→SD (tecla **2**)



Usando a calculadora

- Limpar a memória estatística: SHIFT→MODE→Scl (Tecla **1**)

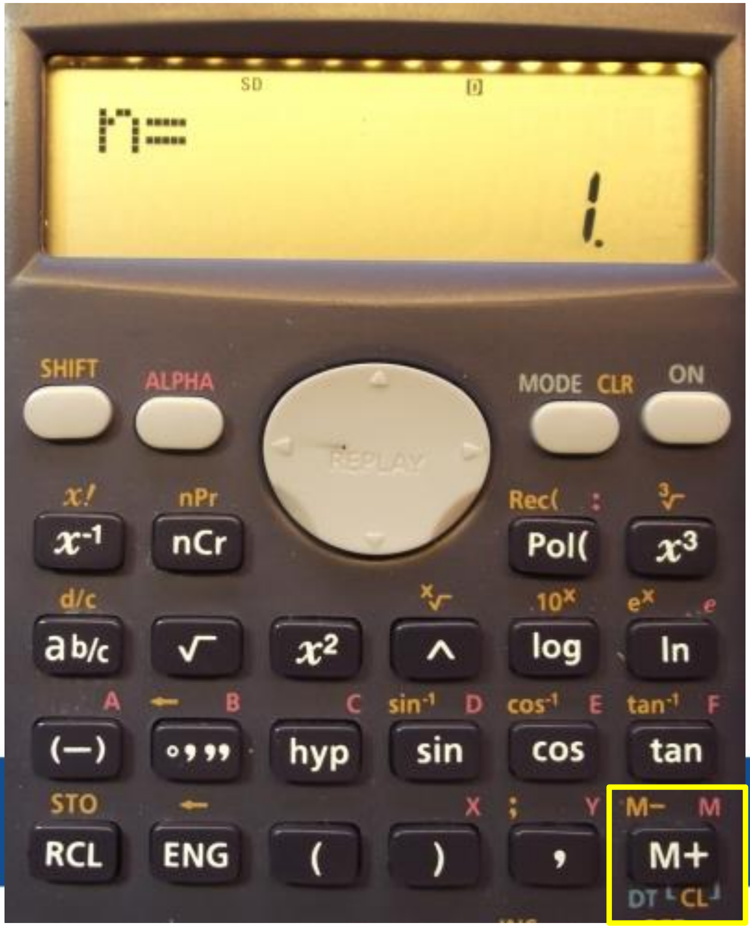




- Pressionar a tecla **=** para limpar a memória.

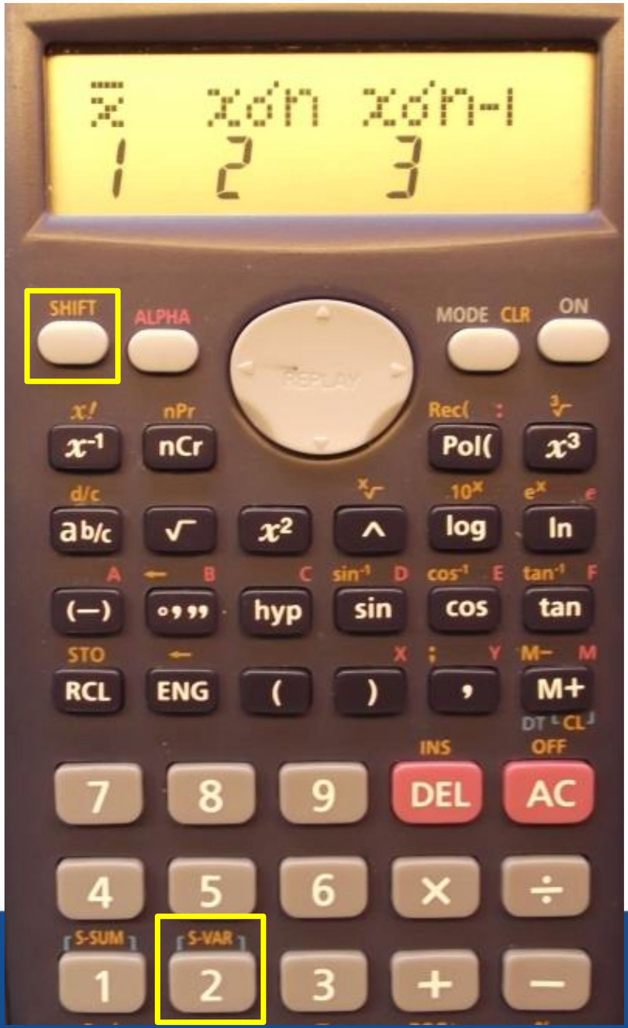
Usando a calculadora

- Inserir os dados pressionando a tecla **DT (M+)** após a digitação de cada valor. O display informa quantos valores foram digitados:



Usando a calculadora

- Para obter o menu estatístico, pressionar SHIFT→S-VAR (Tecla **2**)



Usando a calculadora

- Pressione a tecla **1** ($\bar{x}$) e em seguida a tecla **=** e será exibido no display o valor da média aritmética dos valores digitados.
- Não esquecer de limpar a memória antes do início da digitação de uma nova série.

Mediana

- É o **valor central** de uma distribuição *ordenada* de dados (rol).
- 50% dos valores da distribuição estão abaixo da mediana e os outros 50% dos valores estão acima da mediana.
- Em uma amostra de n itens, a POSIÇÃO da mediana pode ser obtida por:

$$Posição\ Mediana = \frac{n+1}{2}$$

Exemplo

$$x_1 = 10, \quad x_2 = 20, \quad x_3 = 30 \quad x_4 = 40 \quad x_5 = 50$$

$$n = 5 \rightarrow \mathit{Posição\ Mediana} = \frac{5+1}{2} = 3$$

$$\mathit{Mediana} : \boxed{Md = x_3 = 30}$$

Mas e se a quantidade de itens (n) for par?

Determinação da Mediana

- *n é ímpar:* A posição da mediana é a dada pela expressão (n+1)/2:

$$\left.\begin{array}{l} x_1 = 2 \\ x_2 = 20 \\ x_3 = 42 \\ x_4 = 60 \\ x_5 = 71 \\ x_6 = 80 \\ x_7 = 90 \end{array}\right\} \begin{array}{l} n = 7\ (\text{ímpar}) \rightarrow \text{Posição Mediana} = \dfrac{7+1}{2} = 4 \\ Md = x_4 = 60 \end{array}$$

Determinação da Mediana

- *n é par:* A Mediana é dada pela média aritmética dos dois elementos centrais da distribuição ordenada:

$$\left.\begin{aligned} x_1 &= 2 \\ x_2 &= 20 \\ x_3 &= 42 \\ x_4 &= 60 \\ x_5 &= 71 \\ x_6 &= 80 \end{aligned}\right\} \begin{aligned} & n = 6\,(\text{par}) \rightarrow \text{Elementos centrais}: x_3 = 42 \text{ e } x_4 = 60 \\ & Md = \frac{42+60}{2} = 51 \rightarrow \boxed{Md = 51} \end{aligned}$$

Exemplos:

1. Calcule a Mediana dos tempos que você leva para se arrumar no período da manhã.
2. Calcule a Mediana dos salários da empresa vista anteriormente.

O que você pode dizer agora sobre a média e a mediana dos salários?

Moda

- É o valor que mais se repete em uma amostra ordenada (rol).
- Uma amostra pode conter apenas uma moda, ou pode conter mais de uma moda ou mesmo não conter moda (amodal).
- Exemplo:
  - ❖ **3, 3, 3**, 4, 5, 6, 6, 7, 8, 9, 9 → Moda Mo=3
  - ❖ **3, 3**, 4, 5, **6, 6**, 7, 8, **9, 9** → Modas Mo_1=3, Mo_2=6, Mo_3=9
  - ❖ 3, 4, 5, 6, 7, 8, 9 → Sem Moda

Exemplos

1. Calcule a moda dos tempos que você leva para se arrumar no período da manhã.
2. Calcule a moda dos salários da empresa vista.

O que você pode dizer sobre a média, mediana e moda dos salários?

Média Ponderada

- Na média ponderada são atribuídos *pesos* para cada dado. Esses pesos representam quantas vezes esses valores aparecem repetidos na sequência de dados de tamanho ***n***.
- Admitindo que cada valor x_i possui peso n_i, (cada valor x_i aparece repetido n_i vezes), a média ponderada pode ser calculada da seguinte forma:

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

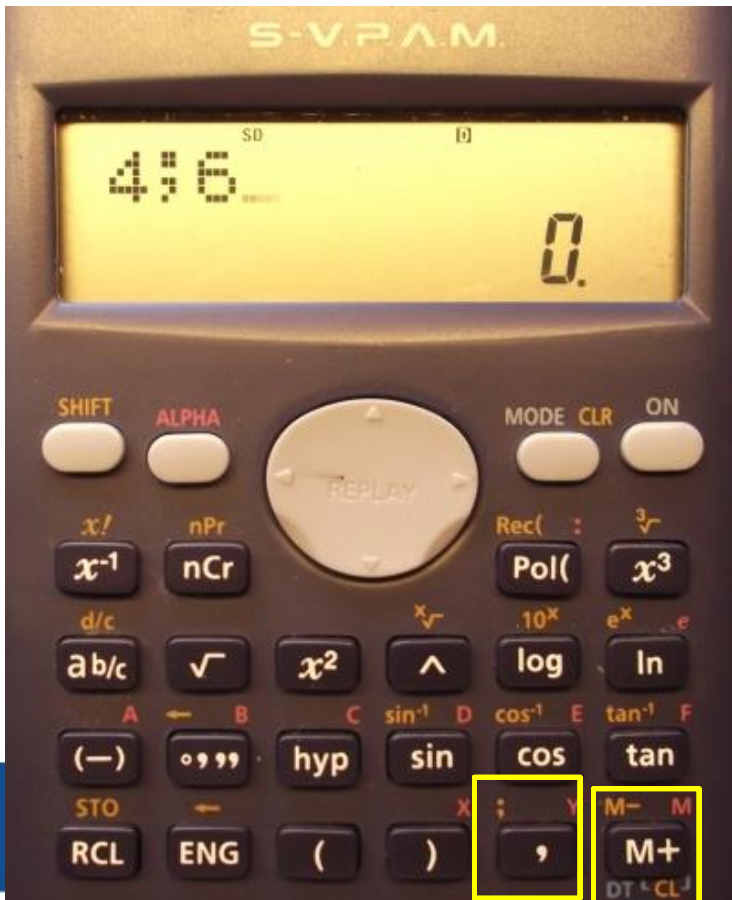
Exemplo

- Dada a tabela abaixo, calcule sua média ponderada:

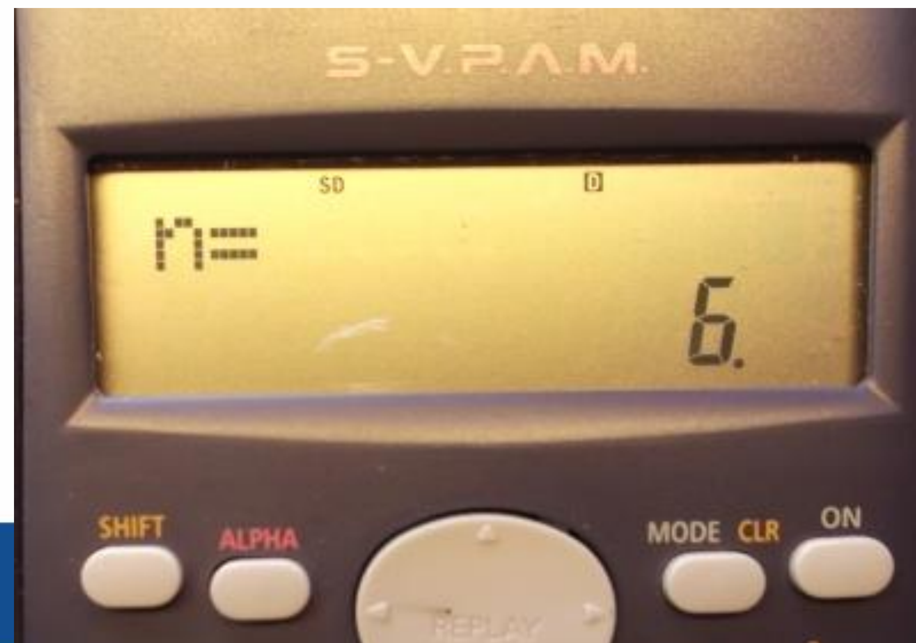
Valor	Frequência
12	2
20	1
28	4
10	3
8	5

Usando a calculadora

- Para o calculo da média ponderada usando a calculadora, devemos mudar a forma que introduzimos os dados.
- Devemos digitar o valor e a sua frequência separados por um **;** (**SHIFT→,**) e em seguida pressionar DT, como já visto.



O display exibira a quantidade (frequência ou peso do valor digitado. Daí em diante o procedimento de cálculo é idêntico ao da média aritmética comum.



Medidas de Dispersão

- Variância s^2 de uma amostra de n itens pode ser obtido pela seguinte expressão:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Seu desvio-padrão s pode ser obtido por:

$$s = \sqrt{s^2}$$

Variância e Desvio-Padrão

- Variância σ^2 de uma população de *N* itens pode ser obtida pela seguinte expressão:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

- Seu desvio-padrão σ pode ser obtido por:

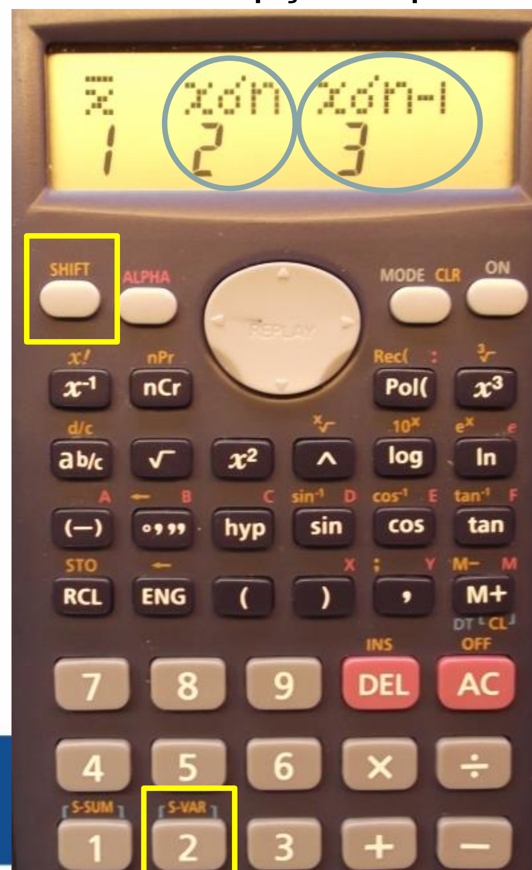
$$\sigma = \sqrt{\sigma^2}$$

Exemplo

1. Calcule a variância dos tempos que você leva para se aprontar no período da manhã.
2. Calcule o desvio-padrão dos tempos que você leva para se aprontar no período da manhã.

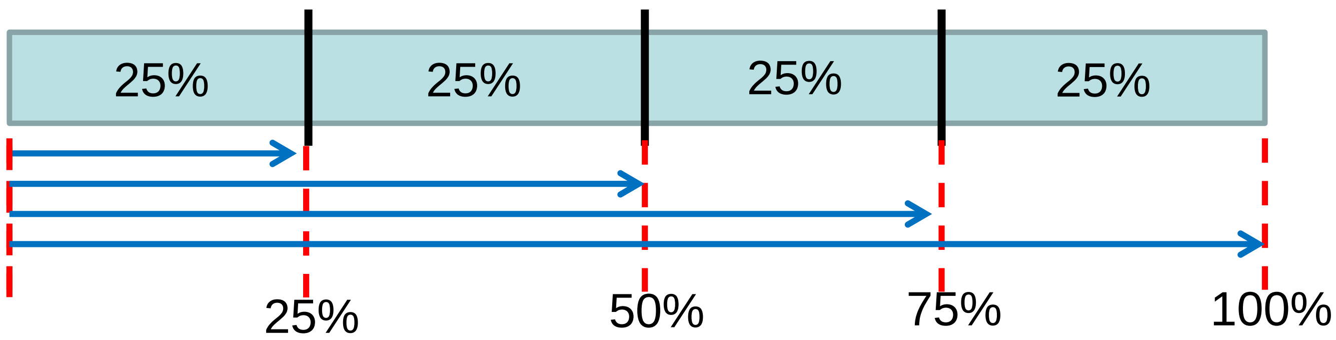
Usando a calculadora

- Para o calculo da variância e do desvio-padrão usando calculadora, deve-se proceder da mesma fora que no cálculo da média aritmética. No menu estatístico deveremos escolher a opção 2 para a população ou 3 para a amostra:



Quartis

- Divide um grupo ordenado de dados (ROL) em quatro partes iguais, cada uma contendo 25% dos dados.



- $\boldsymbol{Q_1}$ engloba 25% dos dados.
- $\boldsymbol{Q_2}$ engloba 50% dos dados (mediana).
- $\boldsymbol{Q_3}$ engloba 75% dos dados.

Determinação dos Quartis

- Posição de $Q_1=(n+1)/4$
- Posição de $Q_2=(n+1)/2$
- Posição de $Q_3=3(n+1)/4$
- Onde n é o tamanho da amostra.
- Arredondamento:
  - Se o resultado for um número inteiro, então o quartil é igual ao valor na ordem de classificação.
  - Se o resultado for metade fracionada **.5**, então o quartil é obtido pela média entre os valores correspondentes no ROL (regra igual à regra da Mediana).
  - Se o resultado for fracionado **.25** ou **.75** arredondar a posição para o inteiro mais próximo.

Exemplos

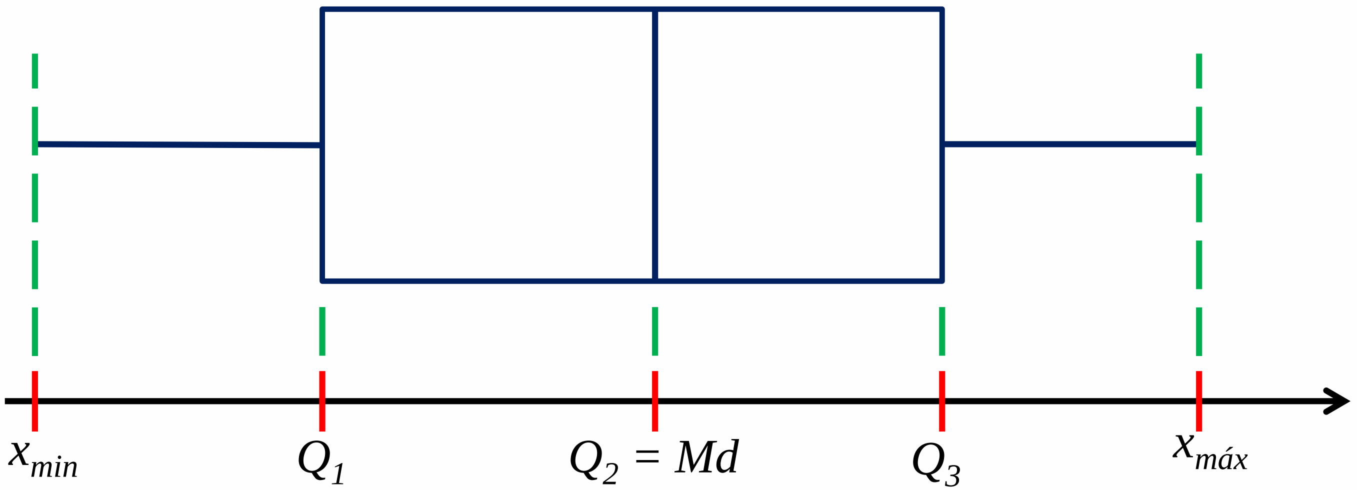
- Dados: (2,6,10, 23,45,67,87,93,100,112,145) com $\boldsymbol{n=11}$:
  - Posição de $Q_1=(11+1)/4=3$ e o quartil $Q_1=x_3=10$
  - Posição de $Q_2=(11+1)/2=6$ e o quartil $Q_2=x_6=67$
  - Posição de $Q_3=3(11+1)/4=9$ e o quartil $Q_3=x_9=100$
- Dados: (2,6,10, 23,45,67,87,93,100,112,145, 157, 189) com $\boldsymbol{n=13}$:
  - Posição de $Q_1=(13+1)/4=3{,}5$ e o quartil $Q_1=(x_3+x_4)/2=(10+23)/2$ e $Q_1=16{,}5$
  - Posição de $Q_2=(13+1)/2=7$ e o quartil $Q_2=x_7=87$
  - Posição de $Q_3=3(13+1)/4=10{,}5$ e o quartil $Q_3=(x_{10}+x_{11})/2$ e $Q_3=(100+112)/2$ e $Q_3=106$
- Dados: (2,6,10,23,45,67,87,93,100,112) com $\boldsymbol{n=10}$:
  - Posição de $Q_1=(10+1)/4=2{,}75$ e $Q_1=x_3=10$
  - Posição de $Q_2=(10+1)/2=5{,}5$ e $Q_2=(x_5+x_6)/2=(45+67)/2=56$
  - Posição de $Q_3=3(10+1)/4=8{,}25$ e $Q_3=x_8=93$

Resumo de 5 valores

- x_{min}
- Q_1
- Q_2 ou mediana
- Q_3
- $x_{máx}$

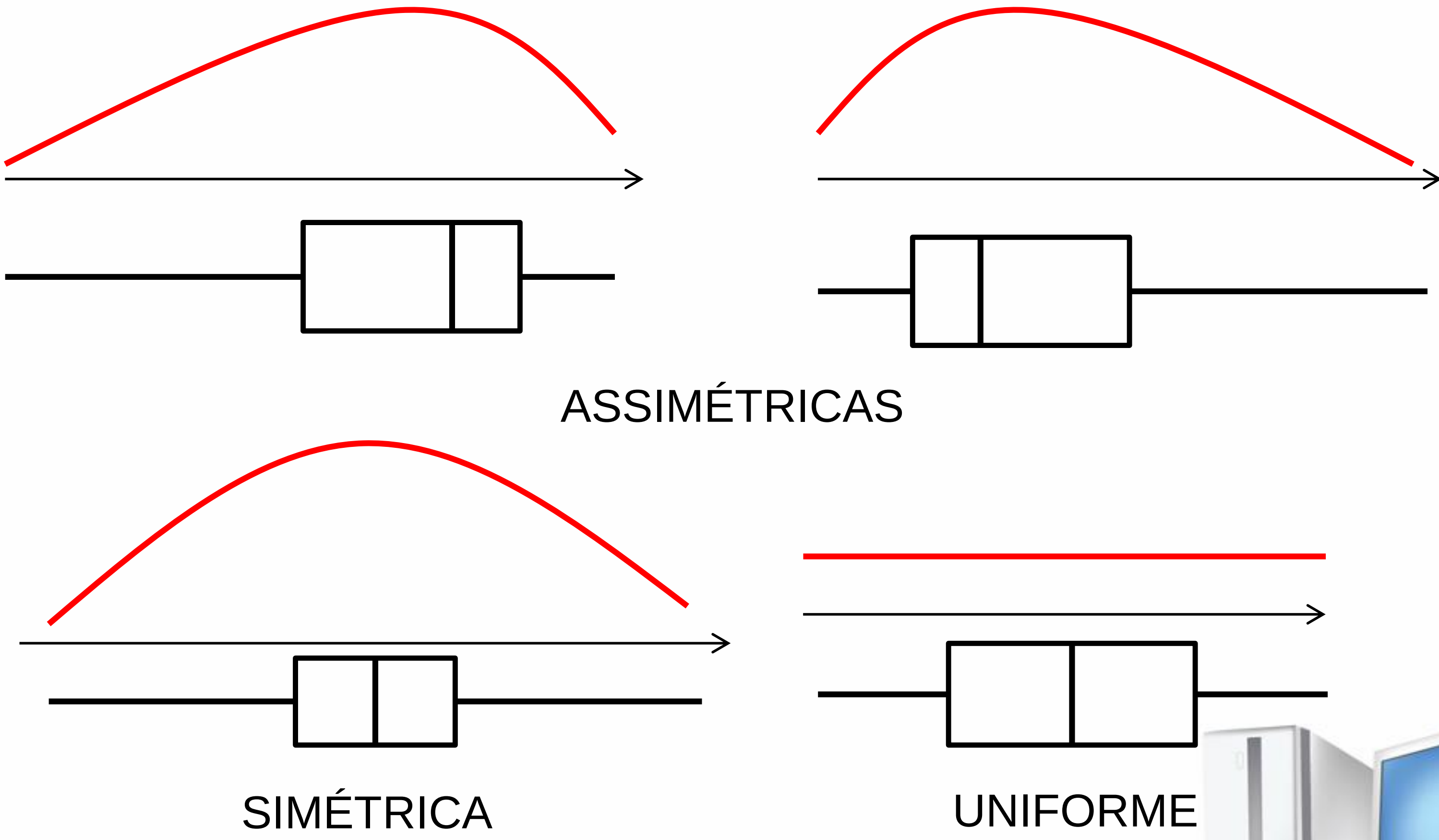
Box Plot

- Representação gráfica em escala do resumo dos 5 pontos:



- O Box Plot permite observar a assimetria da distribuição.

Box Plot e Assimetria



Percentil

- Em uma amostra de **n** itens, o *k-ésimo* percentil P_k é dado pelo valor x_k que corresponde à frequência cumulativa de $nk/100$ de dados em um rol:
    - O 1º percentil corresponde a 1% dos dados.
    - O 20º percentil corresponde a 20% dos dados.
    - O 98º percentil corresponde a 98% dos dados.

Percentil e Decil

- O 25º percentil corresponde ao primeiro quartil $\mathbf{Q_1}$.
- O 50º percentil corresponde à **mediana**.
- O 75º percentil corresponde ao terceiro quartil $\mathbf{Q_3}$.
- ***DECIL:*** é qualquer um dos nove valores que dividem os dados ordenados em dez partes iguais, de modo que cada parte representa ***1/10*** da amostra ou população.
- O **10º percentil** corresponde ao **1º decil**.
- O **20º percentil** corresponde ao **2º decil**.
- ...
- O **80º percentil** corresponde ao **8º decil**.
- O **90º percentil** corresponde ao **9º decil**.

Percentil

- A definição de ***Mendenhall e Sincich*** para o *k-ésimo* percentil de **n** valores ordenados é o valor que se encontra na posição:

$$posição = \frac{k(n+1)}{100}$$

- Este valor deve ser arredondado para o inteiro mais próximo.
- Alguns softwares podem usar definições diferentes para o cálculo do percentil.

Exemplo:

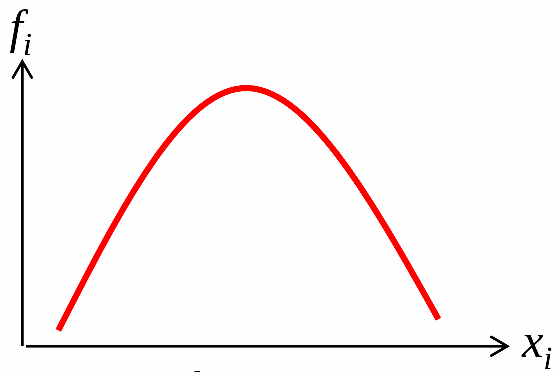
- Considere o seguinte rol de dados:

91,93,94,101,101,101,101,102,103,103,104,105,106,107,107,107,
109, 110,110,111,111,113,114,115,115,116,118,,118,118,119,119,119,
119,120,121,121,121,122,123,124,126,127,128,129,131,132,133,135,
141,141

- Determine a Média, Mediana, Moda, Q_1 e Q_3.
- Determine P_1, P_{10}, P_{30}, P_{50}, P_{80} e P_{75}.

Assimetria e Formato

- As medidas de assimetria possibilitam analisar uma distribuição de acordo com as relações entre suas medidas de moda, média, mediana observadas graficamente.



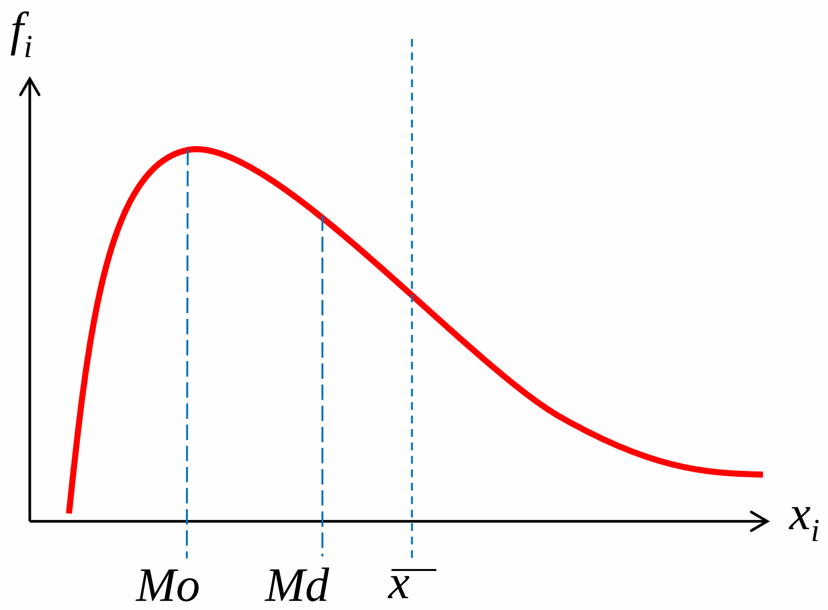
- Uma distribuição é dita simétrica quando a moda, média e mediana apresentam o mesmo valor:

$$\bar{x} = Mo = Md$$

- Neste caso dizemos que a distribuição tem **assimetria nula**.

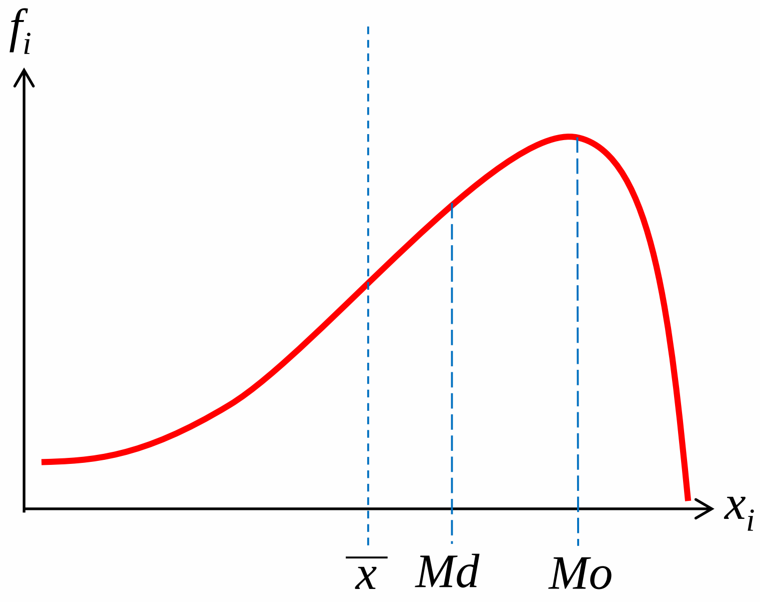
Assimetria

- Se $Mo < Md < \bar{x}$:
- $\bar{x} - Mo > 0$: ***Assimetria Positiva*** ou ***à direita***:



Assimetria

- Se $Mo > Md > \bar{x}$:
- $\bar{x} - Mo < 0$: **Assimétrica Negativa** ou **à esquerda:**



Coeficiente de Assimetria

- A medida de assimetria pode ser realizada pelo 1º Coeficiente de Assimetria de Pearson:

$$As = \frac{(\mu - Mo)}{\sigma} \quad (população)$$

$$As = \frac{(\bar{x} - Mo)}{s} \quad (amostra)$$

- Se ***As=0*** a distribuição é simétrica.
- Se ***As>0*** a distribuição é assimétrica positiva.
- Se ***As<0*** a distribuição é assimétrica negativa.

Coeficiente de Assimetria

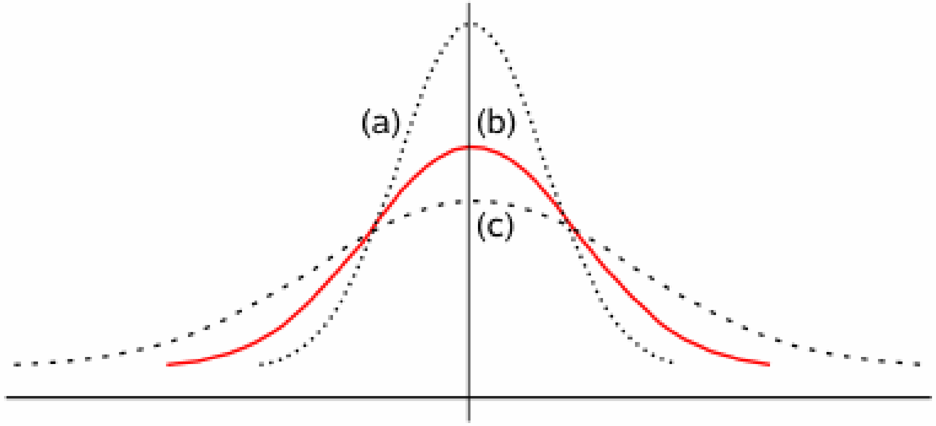
- O coeficiente de Pearson também pode ser representado por:

$$As = \frac{3(\bar{x} - Md)}{s}$$

- Se $|As| < 0{,}15$ a distribuição é simétrica.
- Se $0{,}15 \leq |As| < 1{,}0$ a distribuição é a assimétrica moderada.
- Se $|As| \geq 1{,}0$ a distribuição é assimétrica forte.

Coeficiente de Curtose

- Curtose é o grau de achatamento da distribuição, em relação a curva normal (gaussiana) de referência:



- (a) Distribuição Leptocúrtica
- (b) Distribuição Mesocúrtica (Normal)
- (c) Distribuição Platicúrtica

Coeficiente de Curtose

- A curtose pode ser medida pela seguinte expressão:

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

- C=0,263: curva mesocúrtica
- C<0,263: curva leptocúrtica
- C>0,263: curva platicúrtica

Exercícios

- 1. Calcule os coeficientes de Pearson para os exemplos dos tempos que você leva para se arrumar no período da manhã.
- 2. Calcule o coeficiente de curtose do mesmo problema.
- 3. Calcule os coeficiente de Pearson e de curtose para o rol de dados usado no exemplo do percentil.

Bibliografia

- *Levine, Stephen, Krehbiel, Berenson*. Estatística Teoria e Apliações. Quinta Edição. Ed. LTC.
- Apostila de Estatística *Prof. Cristina Vidal* Accioly. Aula 10. Obtida na Internet.
- Apostila de Estatística e Probabilidade *Prof. Paulo Alessio*. Obtida na Internet.
- Imagens da Internet.