Medidas Estatísticas

Vilfredo Pareto (Paris, França, 15 de julho de 1848 — 19 de agosto de 1923)
Nascido em Paris como Marquês Vilfredo Frederico Damaso Pareto, mas educado na Itália onde se formou Engenheiro pela escola politécnica de Turim. O curso de sólida base matemática teve uma influência intelectual profunda que se refletiu em seu trabalho posterior. Em 1870 tornou-se doutor com a tese "Os princípios fundamentais do equilíbrio em corpos sólidos", que despertaram seu futuro interesse na análise de equilíbrio na Economia e Sociologia. Ele estudou a distribuição de riquezas em diferentes países, concluindo que uma minoria (20%) das pessoas controlam a grande maioria (80%) da riqueza. Esta mesma distribuição foi observada em outras áreas como na distribuição de terras na Itália e acabou sendo conhecido como efeito ou lei de Pareto. O efeito de Pareto pode ser observado também no Controlo de Qualidade onde normalmente 80% dos problemas se originam de apenas 20% das causas. O gráfico de Pareto é utilizado para mostrar a ação do princípio. Os dados são colocados de tal forma que os poucos fatores que causam a maioria dos problemas podem ser observados. A transposição do princípio para o Controle Estatístico de Qualidade e consequentemente para a Estatística é devida a J. M. Duran que cunhou a frase "poucos vitais muitos triviais" e que observou que é um princípio universal e o batizou como "princípio de Pareto".
Partindo do diagrama de Pareto uma destruição de probabilidade que mantém basicamente o formato 80/20 foi criada e nomeada em sua homenagem.
http://www.pucrs.br/famat/statweb/historia/daestatistica/biografias/Pareto.htm

- Somatório
- Médias
- Moda
- Separatrizes
- Diagrama de Caixa ou Boxplot
- Amplitude Total
- Desvio Padrão
- Variância
- Medidas de Dispersão relativa
- Medidas de Assimetria
- Medidas de Curtose

5 - MEDIDAS ESTATÍSTICAS

As medidas estatísticas são valores numéricos utilizados para descrever o conjunto de dados em estudo.

Como um valor numérico pode descrever um conjunto de dados?

Ao estudar um fenômeno é de grande ajuda saber como os valores se comportam, qual o valor mínimo e máximo, se os valores estão agrupados em torno de um determinado valor, se variam muito ou pouco, se estão concentrados numa extremidade do intervalo ou no centro deste. Com as medidas estatísticas também é possível saber qual será o formato do gráfico que representa estes dados antes de construí-lo.

As medidas estatísticas podem ser classificadas em Medidas de Posição (ou de Tendência central), Medidas de dispersão e Medidas de Assimetria e Curtose.

As Medidas de Posição carregam com si a noção de "ordem de grandeza", isto é, nos dá a dimensão dos valores. Com elas podemos nos localizar no eixo X e saber se os valores são da ordem de dezenas, centenas ou milhares. Já as Medidas de Dispersão indicam como os valores estão agrupados entorno de uma Medida de Posição. Os valores podem estar concentrados entorno da Medida de Posição tomada com referência ou podem estar mais espalhados (dispersos).

Com as Medidas de Assimetria e de Curtose temos uma imagem prévia da curva descrita pelo conjunto de valores. Esta curva pode ter deformações à esquerda ou à direita comparada com uma curva padrão. Ou pode estar mais aguçada ou mais achatada em relação a esta mesma curva padrão.

No cálculo de várias medidas estatísticas utilizam-se somas de várias parcelas. Para facilitar a representação destas somas são utilizados os Somatórios.

5.1 - SOMATÓRIO

O Somatório é um operador cuja principal finalidade é de simplificar a notação de certas equações.

Para representar uma soma de n valores do tipo $x_1 + x_2 + x_3 + \ldots + x_n$ pode-se codificá-la através da expressão:

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

onde $\sum$ $\equiv$ (sigma maiúsculo) é o símbolo utilizado para representar o somatório;

x_i $\equiv$ é a parcela genérica e "i" é um índice que assume os valores inteiros consecutivos entre os limites inferior e superior do somatório.

A expressão $\sum_{i=a}^{n} x_i$ deve ser lida "Somatório de x índice i, com i variando de a até n"

Exemplo 5.1 - Os exemplos a seguir ilustram o desenvolvimento de um somatório.

1) $\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$

2) $\sum_{1=3}^{5} x_i^2 = x_3^2 + x_4^2 + x_5^2$

3) $\sum_{i=2}^{4} (x_i + 5)^3 = (x_2 + 5)^3 + (x_3 + 5)^3 + (x_4 + 5)^3$

4) $\sum_{i=1}^{3} \frac{y_i}{x_i} = \frac{y_1}{x_1} + \frac{y_2}{x_2} + \frac{y_3}{x_3}$

Para se determinar o número de parcelas de um somatório do tipo $\sum_{i=a}^{n} ?$ utiliza-se a expressão: $P = n - a + 1$

5.2 - MÉDIAS

As médias são Medidas de Posição e indicam a ordem de grandeza dos valores. Além disso, são a melhor opção para representar um conjunto de dados. As médias podem ser simples ou ponderadas, dependendo se os valores estão na forma de conjuntos (Dados Brutos ou Rol) ou representados numa Distribuição de Frequências. Se os dados estão na forma de conjuntos deve-se calcular uma Média Simples (não tem frequência) se estão na forma de Distribuição deve-se calcular uma Média Ponderada (tem frequência).

5.2.1 - *MÉDIA ARITMÉTICA* $(\bar{x})$

É a medida estatística mais importante. Mais a diante será demonstrado que os valores de um conjunto de dados tendem a estar próximos do valor da Média aritmética.

A Média Aritmética é definida como sendo o quociente entre a soma de todos os valores e o número total de valores.

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

Média Aritmética Simples (dados na forma de conjuntos)

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

Média Aritmética Ponderada (dados na forma de Distribuição de Frequências)

Exemplo 5.2 - *Seja a série* $X = \{3,\ 5,\ 6,\ 8,\ 8,\ 9,\ 11,\ 13,\ 15\}$*, sua média aritmética é:*

<u>*Solução:*</u>

$$\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{3+5+6+8+8+9+11+13+15}{9} = \frac{78}{9} = 8,6667$$

Exemplo 5.3 - *Determine a média aritmética da distribuição de valores individuais a seguir:*

x_i	f_i
3	6
5	8
7	2
9	3
13	5
	24

Solução

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i f_i}{24} = \frac{3\times6+5\times8+7\times2+9\times3+13\times5}{24} = \frac{164}{24} = 6{,}8333...$$

Exemplo 5.4 - *Determine a média aritmética da distribuição a seguir.*

CLASSES	f_i
3 ├── 9	7
9 ├── 15	6
15 ├── 21	4
21 ├── 27	8
27 ├── 33	5
33 ├── 39	3
	33

Solução

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i f_i}{33} = \frac{6\times7+12\times6+18\times4+24\times8+30\times5+36\times3}{33}$$

$$\overline{x} = \frac{636}{33} = 19{,}272727...$$

5.2.2 - *Média Quadrática* $(\overline{x}_q)$

A média Quadrática é definida como sendo a raiz quadrada da Média Aritmética Simples dos Quadrados.

$$\overline{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$ Média Quadrática Simples (dados na forma de conjuntos)

$$\overline{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{n}}$$ Média Aritmética Ponderada (dados na forma de Distribuição de Frequências)

Exemplo 5.5 - *Determine a Média Quadrática da série* $X = \{5,\ 6,\ 8,\ 9,\ 10,\ 11,\ 14,\ 18,\ 21\}$ *:*

Solução

$$\overline{x}_q = \sqrt{\frac{\sum_{i=1}^{9} x_i^2}{9}} = \sqrt{\frac{5^2+6^2+8^2+9^2+10^2+11^2+14^2+18^2+21^2}{9}}$$

$$\overline{x}_q = \sqrt{\frac{1388}{9}} = \sqrt{154,2222} = 12,41862401$$

Exemplo 5.6 - Seja a distribuição a seguir. Determine sua média Quadrática.

CLASSES	f_i
7 ├── 12	4
12 ├── 17	3
17 ├── 22	6
22 ├── 27	5
27 ├── 32	2
	20

Solução

$$\overline{x}_q = \sqrt{\sum_{i=1}^{5}\frac{x_i^2 \times f_i}{20}} = \sqrt{\frac{9,5^2\times 4+14,5^2\times 3+19,5^2\times 6+24,5^2\times 5+29,5^2\times 2}{20}}$$

$$\overline{x}_q = \sqrt{\frac{8015}{20}} = \sqrt{400,75} = 20,01874122$$

5.2.3 - MÉDIA HARMÔNICA $(\overline{x}_h)$

A média Harmônica é definida como sendo o inverso da média aritmética dos inversos.

$$\overline{x}_h = \frac{1}{\frac{\sum_{i=1}^{n}\frac{1}{x_i}}{n}} \quad \text{ou} \quad \overline{x}_h = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

Média Harmônica Simples (dados na forma de conjuntos)

$$\overline{x}_h = \frac{1}{\frac{\sum_{i=1}^{n}\frac{1}{x_i}f_i}{n}} \quad \text{ou} \quad \overline{x}_h = \frac{n}{\sum_{i=1}^{n}\frac{f_i}{x_i}}$$

Média Harmônica Ponderada (dados na forma de Distribuição de Frequências)

Exemplo 5.7 - *Seja a distribuição a seguir. Determine sua média Harmônica.*

CLASSES	f_i
7 ├── 12	4
12 ├── 17	3
17 ├── 22	6
22 ├── 27	5
27 ├── 32	2
	20

Solução

$$\overline{x}_h = \frac{n}{\sum_{i=1}^{k}\frac{f_i}{x_i}} = \frac{20}{\sum_{i=1}^{5}\frac{f_i}{x_i}} = \frac{20}{\frac{4}{9,5}+\frac{3}{14,5}+\frac{6}{19,5}+\frac{5}{24,5}+\frac{2}{29,5}} = \frac{20}{1,20752} = 16,56288$$

5.2.4 - *MÉDIA GEOMÉTRICA* $\left(\overline{x}_g\right)$

A média Geométrica é definida como sendo a raiz n-ézima dos produtos dos valores.

$\overline{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i}$ Média Geométrica Simples (dados na forma de conjuntos)

$\overline{x}_g = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}}$ Média Geométrica Ponderada (dados na forma de Distribuição de Frequências)

Exemplo 5.8 - *Seja a distribuição a seguir. Determine sua média Harmônica.*

CLASSES	f_i
7 ├── 12	4
12 ├── 17	3
17 ├── 22	6
22 ├── 27	5
27 ├── 32	2
	20

Solução

$$\overline{x}_g = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}} = \sqrt[20]{9,5^4 \times 14,5^3 \times 19,5^6 \times 24,5^5 \times 29,5^2} = \sqrt[20]{1,04877 \times 10^{25}} = 17,82518424$$

Dependendo dos valores algumas calculadoras não são potentes o suficientes para realizar o cálculo de uma média Geométrica. Nestes casos podemos utilizar a fórmula a seguir.

$\overline{x}_g = 10^{\frac{\sum_{i=1}^{n} \log(x_i)}{n}}$ Média Geométrica Simples (dados na forma de conjuntos)

$\overline{x}_g = 10^{\frac{\sum_{i=1}^{k} f_i \times \log(x_i)}{n}}$ Média Geométrica Ponderada (dados na forma de Distribuição de Frequências)

5.2.5 - *PROPRIEDADES DAS MÉDIAS*

A principal propriedade das médias é:

$$\overline{x}_h \leq \overline{x}_g \leq \overline{x} \leq \overline{x}_q$$

5.3 - MODA (MO)

Num conjunto de observações ou numa Distribuição de Frequências a Moda é o valor de ocorrência mais frequente. A Moda também pode ser denominada por *Norma*, *Valor Dominante* ou *Valor Típico.*

5.3.1 - *MODA DE VALORES NÃO - TABULADOS*

Exemplo 5.9 - Determine a moda de uma das séries a seguir:

X = { 4, 5, 5, 6, 6, 6, 7, 7, 8, 9 }
Y = { 4, 4, 5, 5, 6, 6 }
Z = { 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6 }
W = { 1, 2, 3, 4, 5 }
Q = { 1 }
R = { 1, 3, 4, 4, 6, 7, 8, 8, 9, 10, 11, 11, 17 }

Solução:

Moda de X ⇨ Mo = 6
Moda de Y ⇨ Y é uma série Amodal
Moda de Z ⇨ Mo1 = 2 e Mo2 = 5 (Bimodal)
Moda de W ⇨ W é uma série Amodal
Moda de Q ⇨ Mo = 1
Moda de R ⇨ Mo1 = 4; Mo2 = 8 e Mo3 = 11 (Trimodal)

5.3.2 - *MODA DE DISTRIBUIÇÕES DE FREQUÊNCIAS DE VALORES DISTINTOS*

Neste caso a determinação da Moda é imediata bastando observar o valor de maior frequência da distribuição.

Exemplo 5.10 - Determine a Moda da distribuição a seguir

x_i	f_i
3	2
4	8
7	5
9	7
11	8
13	3

Solução:

A distribuição é Bimodal e apresenta Modas $Mo_1 = 4$ e $Mo_2 = 11$

5.3.3 - Moda de Distribuição de Frequências de Valores Agrupados em Classes

Existem três métodos para a determinação da moda neste caso.

5.3.3.1 - Moda Bruta

É definida como sendo o Ponto Médio da Classe Modal (Classe de maior frequência).

Exemplo 5.11 - Determine a moda da distribuição a seguir:

Classes	f_i
2 |--- 8	4
8 |--- 14	7
14 |--- 22	5
22 |--- 28	12
28 |--- 34	9
34 |--- 40	6

Solução:

A Classe modal é a 4ª, logo $Mo = \frac{22+28}{2} = 25$.

5.3.3.2 - Moda por King.

A teoria de King Admite que o valor modal desloca-se dentro do intervalo de classe de maneira que as distâncias desse ponto aos limites de classe sejam inversamente proporcionais às frequências das respectivas classes adjacentes.

$$Mo = l_i + \left(\frac{c \times f_{post}}{f_{ant} + f_{post}} \right)$$

Onde: $l_i \equiv$ Limite inferior da classe modal;

$c \equiv$ Intervalo de classe;

$f_{ant} \equiv$ Frequência Absoluta Simples da classe adjacente anterior à classe modal;

$f_{post} \equiv$ Frequência Absoluta Simples da classe adjacente posterior à classe modal.

Exemplo 5.12 - Determine a moda da distribuição a seguir:

Classes	fi
2 |--- 8	4
8 |--- 14	7
14 |--- 22	5
22 |--- 28	12
28 |--- 34	9
34 |--- 40	6

Solução:

Exemplo 5.13 -

Sendo a 4ª classe a classe modal temos:

$$Mo = l_i \frac{c \times f_{post}}{f_{ant} + f_{post}} = 22 + \frac{6 \times 9}{5 + 9} \Rightarrow Mo \approx 25{,}87514$$

5.3.3.3 - Moda por Czuber

O ponto que corresponde à Moda divide o intervalo de Classe Modal em duas partes, as quais são proporcionais às diferenças entre a Frequência Modal e as frequências das classes adjacentes.

$$Mo = l_i + \left(\frac{c \times \left(f_{mo} - f_{ant} \right)}{2 \times f_{mo} - \left(f_{ant} + f_{post} \right)} \right)$$

Exemplo 5.14 - Determine a Moda da distribuição a seguir:

Classes			f_i
2	$\vdash$	8	4
8	$\vdash$	14	7
14	$\vdash$	22	5
22	$\vdash$	28	12
28	$\vdash$	34	9
34	$\vdash$	40	6

Solução:

Sendo a classe modal a 4ª classe temos:

$$Mo = l_i + \left(\frac{c \times (f_{mo} - f_{ant})}{2 \times f_{mo} - (f_{ant} + f_{post})} \right)$$

$$Mo = 22 + \frac{6 \times (12 - 5)}{2 \times 12 - (5 + 9)}$$

$$Mo = 26,2$$

5.4 - SEPARATRIZES

Separatrizes são medidas que dividem um conjunto de dados “ordenado” em um determinado número de intervalos, sendo que estes intervalos conterão uma porcentagem de valores menores ou iguais ao valor separatriz. Note que as Separatrizes são influenciadas apenas pelo número de valores do conjunto (ou da Distribuição) e não pela ordem de grandeza desses valores.

São quatro tipos mais comuns: Mediana, Quartís, Decís e Percentís.

TABELA 19.- SEPARATRIZ, INTERVALOS, PORCENTAGENS E SIMBOLOGIAS

Medida/ Simbologia	Número de Intervalos	Porcentagem de Valores em cada intervalo	Pontos Separatrizes
Mediana (Md)	2	50%	Md
Quartil (Q_i)	4	25%	Q_1, Q_2 e Q_3
Decil (D_i)	10	10%	D_1, D_2,..., D_8 e D_9
Percentil ou Centil (C_i)	100	1%	C_1, C_2, ... C_{98} e C_{99}

A Figura 19 ilustra as porcentagens correspondentes a cada intervalo das Separatrizes. Neste caso foi considerado que os valores estão distribuídos de maneira uniforme, isto é, cada intervalo tem o mesmo comprimento. Na prática é raro um fato desses. A figura 20 ilustra uma situação mais real.

As Separatrizes podem ser utilizadas para determinar intervalos de maior e menor concentração de valores. Para os dados da figura 20 o intervalo que vai do valor mínimo até o 1º quartil é o intervalo de maior concentração e o intervalo que vai do 3º Quartil até o valor máximo é o intervalo onde os valores estão mais dispersos.

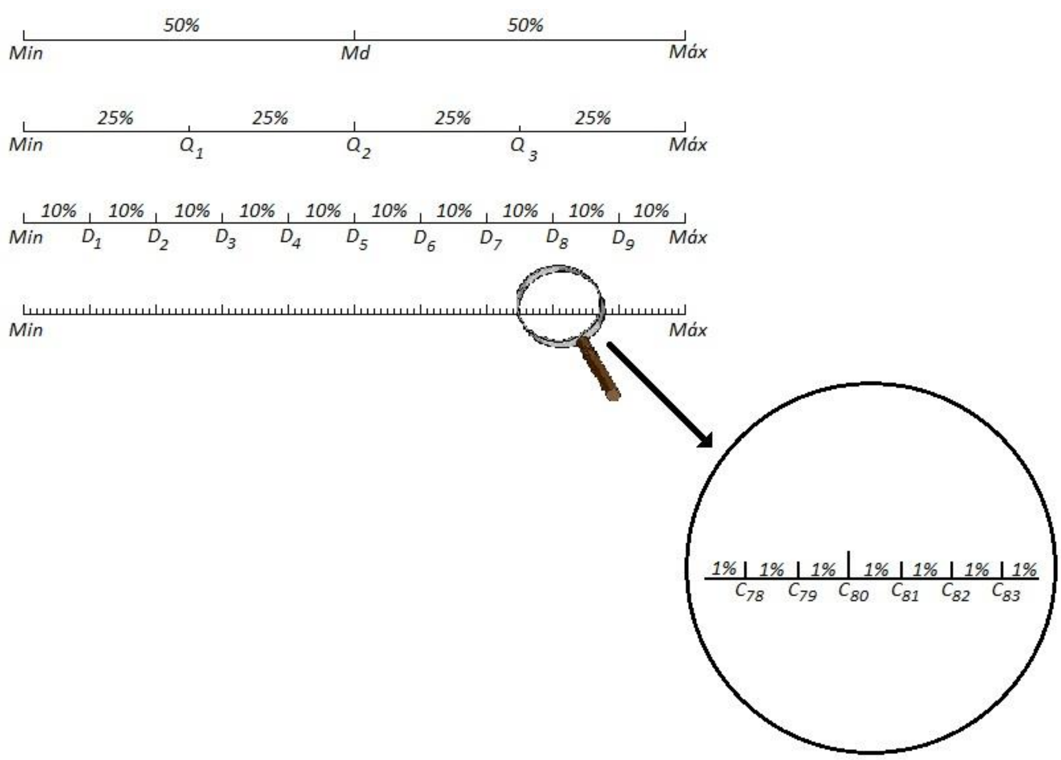


Figura 19.- Porcentagens correspondentes às separatrizes

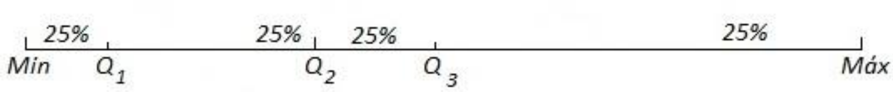


Figura 20.- Uma situação real em relação às separatrizes

Para se determinar qualquer valor separatriz os dados devem estar ordenados, ou em ordem crescente ou decrescente, porém suas repetições devem ser mantidas, isto é, deve-se obter o Rol do conjunto de dados.

5.4.1 - *Determinação da Mediana Para Conjuntos de Dados ou Distribuições de Valores Individuais*

A posição do valor mediano é determinada pelo "Elemento Mediano" (E_{MD}), onde:

- $E_{Md} = \dfrac{n}{2}$ se o número de valores observados da série ou distribuição for PAR.
- $E_{Md} = \dfrac{n+1}{2}$ se o número de valores observados da série ou distribuição for ÍMPAR

Exemplo 5.15 - *Determine a Mediana do conjunto de dados a seguir:*

$$X = \{ 2,\ 2,\ 3,\ 3,\ 4,\ 6,\ 6,\ 6,\ 8,\ 9,\ 10,\ 13,\ 13,\ 13,\ 15,\ 18,\ 19 \}$$

<u>*Solução*</u>

O número de valores observados é ÍMPAR, $n = 17$*, portanto*

$$E_{Md} = \frac{n+1}{2} = \frac{17+1}{2} = 9^{\circ}$$

Assim, o valor mediano é o 9º valor da série ordenada, ou seja,

$$Md = 8$$

Exemplo 5.16 - *Determine a Mediana do conjunto a seguir:*

Seja X = { 3, 4, 4, 6, 7, 7, 11, 15, 17, 17, 18, 19 }

<u>*Solução*</u>

O número de valores observados é PAR, n = 12, assim o Elemento Mediano é

$$E_{Md} = \frac{n}{2} = \frac{12}{2} = 6^{\circ}$$

Entretanto, à esquerda do 6º valor da série existem apenas 5 valores e a sua direita existem 6 valores, logo o 6º elemento da série não a divide em duas partes iguais. A Mediana será obtida, então fazendo - se a média entre o 6º e o 7º elementos da série, assim:

$$Md = \frac{7+11}{2} = 9$$

Exemplo 5.17 - *Considere a distribuição de valores individuais a seguir:*

x_i	f_i
2,3	5
7,2	8
8,5	4
9,6	3
9,7	6
10,3	4
12,7	5
	35

<u>*Solução*</u>

O número de valores observados é ÍMPAR, logo:

$$E_{Md} = \frac{n+1}{2} = \frac{35+1}{2} = 18^{\circ}$$

Portanto a Mediana é $Md = 9{,}6$.

5.4.2 - DETERMINAÇÃO DA MEDIANA PARA DISTRIBUIÇÕES DE VALORES AGRUPADOS EM CLASSES

O Elemento Mediano será dado por $E_{Md} = \frac{n}{2}$, independentemente se o número de valores observados da distribuição for par ou ímpar. A Mediana é obtida por:

$$Md = l_i + \left(\frac{c \times \left(E_{Md} - F_{ant} \right)}{f_{Md}} \right)$$

onde: $l_i \equiv$ Limite inferior da classe que contém a Mediana;
$c \equiv$ Intervalo de Classe;
$F_{ant} \equiv$ Frequência Absoluta Acumulada "Abaixo de" da classe anterior à classe que contém a Mediana;
$f_{Md} \equiv$ Frequência Absoluta Simples da classe que contém a Mediana;
$E_{Md} \equiv$ Elemento Mediano ($E_{Md} = \frac{n}{2}$).

Exemplo 5.18 - Determinar a Mediana da distribuição abaixo.

CLASSES	f_i	F_i
2 ⊢ 11	7	*7*
11 ⊢ 20	5	*12*
20 ⊢ 29	7	*19*
29 ⊢ 38	9	*28*
38 ⊢ 47	11	*39*
47 ⊢ 56	3	*42*
56 ⊢ 65	5	*47*
	47	

<u>*Solução:*</u>

$$E_{Md} = \frac{n}{2} = \frac{47}{2} = 23,5^{\circ}$$

Portanto a classe que contém a Mediana é a 4ª classe, assim

$$Md = l_i + \frac{c \times \left(E_{Md} - F_{ant} \right)}{f_{Md}}$$
$$Md = 29 + \frac{9 \times \left(23,5 - 19 \right)}{9}$$
$$Md = 33,5$$

5.4.3 - *Determinação dos Quartis, Decís e Centís de Séries e de Distribuições de Valores Não Agrupados em Classes*

No caso de conjuntos, o primeiro passo é a ordenação dos valores, isto é, passar de Dados Brutos par Rol.

No caso de distribuições de valores não agrupados em classes os valores já se encontram ordenados.

O passo seguinte é a determinação da posição do valor procurado dentro do conjunto e/ou distribuição. Para tal utilizaremos a tabela a seguir:

TABELA 20.- DETERMINAÇÃO DA POSIÇÃO DE UMA SEPARATRIZ

Medida	Posição do Elemento	Valores de i
Quartil	$E_{Q_i} = \frac{i \times n}{4}$	1, 2 ou 3
Decil	$E_{D_i} = \frac{i \times n}{10}$	1, 2, 3, 4, 5, 6, 7, 8 ou 9
Centil	$E_{C_i} = \frac{i \times n}{100}$	1, 2, 3, ... ,98 ou 99

Note que se o Elemento calculado for INTEIRO então a Separatriz procurada é um valor da série ou da distribuição. Se o Elemento não for um valor INTEIRO significa que a Separatriz procurada é um valor intermediário entre os valores que ocupam as posições aproximadas. Neste caso a Separatriz é definida através da interpolação entre o valor da posição calculada e o próximo valor da lista.

Exemplo 5.19 - Considere a lista de dados a seguir:

0,00	0,01	0,02	0,04	0,12	0,13	0,14	0,14	0,15	0,17	0,23	0,24	0,26	0,28	0,30	0,30	0,34	0,34	0,36	0,36
0,38	0,39	0,44	0,46	0,47	0,47	0,47	0,48	0,48	0,50	0,53	0,53	0,55	0,58	0,58	0,63	0,63	0,73	0,74	0,75
0,75	0,76	0,79	0,81	0,83	0,84	0,85	0,87	0,90	0,93	0,97	1,05	1,05	1,10	1,13	1,19	1,19	1,20	1,24	1,28
1,30	1,32	1,34	1,39	1,42	1,42	1,43	1,45	1,50	1,52	1,52	1,53	1,53	1,57	1,67	1,69	1,73	1,83	1,86	1,88
1,88	1,88	1,88	1,90	1,90	2,03	2,08	2,13	2,20	2,20	2,23	2,27	2,28	2,29	2,34	2,38	2,40	2,42	2,42	2,45
2,46	2,47	2,48	2,51	2,52	2,55	2,57	2,57	2,60	2,60	2,61	2,61	2,63	2,67	2,68	2,74	2,76	2,80	2,82	2,83
2,85	2,86	2,87	2,87	2,87	2,89	2,94	2,97	2,97	2,98	3,01	3,02	3,08	3,08	3,09	3,10	3,12	3,16	3,25	3,29
3,31	3,34	3,34	3,37	3,38	3,40	3,41	3,43	3,45	3,47	3,49	3,49	3,53	3,60	3,62	3,63	3,68	3,69	3,78	3,82
3,87	3,93	3,96	3,98	4,00	4,09	4,09	4,31	4,32	4,36	4,37	4,41	4,43	4,52	4,57	4,57	4,62	4,62	4,62	4,63
4,67	4,68	4,69	4,82	4,83	4,86	4,88	4,89	4,91	4,95	5,00	5,09	5,17	5,21	5,21	5,24	5,34	5,38	5,41	5,42
5,49	5,49	5,52	5,56	5,68	5,69	5,69	5,72	5,75	5,82	6,05	6,08	6,15	6,16	6,19	6,30	6,41	6,45	6,56	6,58
6,58	6,63	6,66	6,69	6,71	6,80	6,85	6,89	6,89	6,92	6,94	6,94	7,05	7,11	7,27	7,29	7,35	7,44	7,48	7,60
7,66	7,67	7,67	7,82	7,86	7,91	7,99	8,16	8,17	8,25	8,29	8,38	8,44	8,51	8,54	8,55	8,61	8,73	8,77	8,78
8,96	9,05	9,11	9,15	9,20	9,22	9,27	9,28	9,28	9,29	9,34	9,38	9,40	9,44	9,52	9,56	9,64	9,89	10,20	10,46
10,66	10,70	10,79	10,83	10,86	11,18	11,23	11,32	11,43	11,49	11,59	11,73	11,78	11,81	11,85	11,98	12,01	12,13	12,19	12,34
12,57	12,59	12,73	12,77	12,79	12,83	12,87	12,95	13,03	13,06	13,24	13,27	13,35	13,44	13,51	13,55	13,79	14,02	14,38	14,45
14,54	14,65	14,67	14,75	14,79	14,84	15,15	15,22	15,23	16,67	16,69	16,97	17,20	17,24	17,39	17,43	17,72	17,94	18,06	18,11
18,29	18,43	18,54	18,58	18,79	18,88	19,08	19,32	19,38	19,44	19,59	19,88	20,23	20,59	20,95	20,96	20,99	21,32	21,38	21,75
23,26	23,45	24,05	24,21	24,37	24,61	24,63	24,87	25,59	26,27	26,41	26,89	26,94	28,34	29,11	29,68	29,96	30,03	30,96	31,14
31,28	31,58	33,53	33,58	34,79	35,29	35,54	35,60	35,60	36,33	36,99	37,82	38,88	39,88	41,01	46,01	46,28			

Determine: Q_1 ; Q_3 ; D_7 e C_{83} .

<u>*Solução:*</u>

A lista contém $n = 397$ *valores e já está ordenada.*

a) A posição do 1º Quartil é dada por:

$$E_{Q_1} = \frac{i \times n}{4} = \frac{1 \times 397}{4} = 99,25^{\circ}$$

Logo o 1º Quartil é um valor que está ente o 99º valor (2,42) e o 100º valor (2,45) da lista, isto é,

$$Q_1 = 2,42 + 0,25 \times (2,45 - 2,42) = 2,4275 .$$

A interpretação deste resultado é: 25% dos valores desta lista são menores ou iguais a 2,4275 e 75% dos valores maiores ou iguais a 2,4275.

b) *A posição do 3º Quartil é dada por:*

$$E_{Q_3} = \frac{3 \times 397}{4} = 297,75^{\circ}$$

Logo o 3º Quartil é um valor que está entre o 297º (12,01) e o298º valor (12,13) da lista, isto é,

$$Q_3 = 12,01 + 0,75 \times (12,13 - 12,01) = 12,10 .$$

A interpretação deste resultado é: 75% dos valores desta lista são menores ou iguais a 12,10 e 25% dos valores maiores ou iguais a 12,10.

c) A posição do 7º Decil é dada por:

$$D_7 = \frac{7 \times 397}{10} = 277,9^{\circ}$$

Logo o 7º Decil é um valor que está entre o 277º (9,64) e o 278º valor (9,89) da lista, isto é,

$$D_7 = 9,64 + 0,9 \times (9,89 - 9,64) = 9,865 .$$

A interpretação deste resultado é: 70% dos valores desta lista são menores ou iguais a 9,865 e 30% dos valores maiores ou iguais a 9,865.

d) A posição do 83º Centil é dada por:

$$C_{83} = \frac{83 \times 397}{100} = 329,51^{\circ}$$

Logo o 83ºCentil é um valor que está entre o 329º (15,23) e o 330º valor (16,67) da lista, isto é,

$$C_{83} = 15,23 + 0,51 \times (16,67 - 15,23) = 15,9644 .$$

A interpretação deste resultado é: 83% dos valores desta lista são menores ou iguais a 15,9644 e 17% dos valores maiores ou iguais a 15,9644.

Exemplo 5.20 - *Calcule o 45º Centil da distribuição a seguir:*

x_i	f_i	F_i
2	4	4
3	3	7
5	6	13
6	8	21
8	5	26
11	3	29
	29	

Solução:

A posição do 45º Centil é dada por:

$$E_{C_{45}} = \frac{45 \times 29}{100} = 13,05$$

Logo o 45º Centil será o valor que está entre o 13º valor (6,0) e o 14º valor (8,0) da distribuição.

$$C_{45} = 6 + 0,05 \times (8-6) = 6,1)$$

5.4.4 - Determinação dos Quartís, Decís e Centís de Distribuições de Valores Agrupados em Classes

Assim como na Mediana o cálculo dos Quartís, Decís e Centís utilizará uma fórmula, uma vez que não é possível saber quais foram os valores observados. A Tabela a seguir traz todas as fórmulas necessárias

TABELA 21.- FORMULAS PARA O CÁLCULO DOS QUARTÍS, DECÍS E CENTÍS

Medida	Posição do Elemento	Fórmula
Quartil	$E_{Q_i} = \frac{i \times n}{4}$	$Q_i = l_i + \left(\frac{c \times \left(E_{Q_i} - F_{ant} \right)}{f_{Q_i}} \right)$
Decil	$E_{D_i} = \frac{i \times n}{10}$	$D_i = l_i + \left(\frac{c \times \left(E_{D_i} - F_{ant} \right)}{f_{D_i}} \right)$
Centil	$E_{C_i} = \frac{i \times n}{100}$	$C_i = l_i + \left(\frac{c \times \left(E_{C_i} - F_{ant} \right)}{f_{C_i}} \right)$

Onde: $l_i \equiv$ Limite inferior da classe que contém a separatriz desejada;

$c \equiv$ Intervalo de Classe;

$F_{ant} \equiv$ Frequência Absoluta Acumulada "Abaixo de" da classe anterior à classe que contém a Separatriz desejada;

$f_? \equiv$ Frequência Absoluta Simples da classe que contém a Separatriz desejada;

$E_{Q_i} \equiv$ Elemento Quartílico;

$E_{Q_i} \equiv$ Elemento Decílico;

$E_{C_i} \equiv$ Elemento Centílico.

Exemplo 5.21 - *Dada a distribuição a seguir determine:*

a) *Q1 e Q3*
b) *D2,*
c) *C37*

CLASSES			f_i	F_i
0,3	⊢	1,5	8	*8*
1,5	⊢	2,7	11	*19*
2,7	⊢	3,9	15	*34*
3,9	⊢	5,1	19	*53*
5,1	⊢	6,3	21	*74*
6,3	⊢	7,5	17	*91*
7,5	⊢	8,7	14	*105*
			105	

Solução:

a) Q1 = ?

O Elemento Quartílico de 1ª ordem é:

$$E_{Q_1} = \frac{1\times105}{4} = 26,25^{\circ}$$

Logo o 1º Quartil está na 3ª classe.

$$Q_1 = l_i + \frac{c\left(E_{Q_1} - F_{ant}\right)}{f_{Q_1}}$$

$$Q_1 = 2,7 + \frac{1,2\times\left(26,25-19\right)}{15}$$

$$Q_1 = 3,28$$

Q3 = ?

O Elemento Quartílico de 3ª ordem é:

$$E_{Q_3} = \frac{3\times105}{4} = 78,75^{\circ}$$

Logo o 3º Quartil está na 6ª classe.

$$Q_3 = l_i + \frac{c\times\left(E_{Q_\#} - F_{ant}\right)}{f_{Q_3}}$$

$$Q_3 = 6,3 + \frac{1,2\times\left(78,75-74\right)}{17}$$

$$Q_3 = 6,635294118$$

b) D2 = ?

$$E_{D_2} = \frac{2 \times 105}{10} = 21^{\circ}$$

$$D_2 = l_i + \frac{c \times \left(E_{D_2} - F_{ant}\right)}{f_{D_2}}$$

$$D_2 = 2,7 + \frac{1,2 \times (21 - 19)}{15}$$

$$D_2 = 2,86$$

c) C37 = ?

$$E_{C_{37}} = \frac{37 \times 105}{100} = 38,85$$

$$C_{37} = l_i + \frac{c\left(E_{C_{37}} - F_{ant}\right)}{f_{C_{37}}}$$

$$C_{37} = 3,9 + \frac{1,2 \times (38,85 - 34)}{19}$$

$$C_{37} = 4,20631579$$

As separatrizes são usadas para se achar um valor que determina certa porcentagem de valores. Mas também pode ser utilizada ao contrário, isto é, dado um valor, qual a porcentagem de valores é menor que o valor dado?

Exemplo 5.22 - *A distribuição a seguir representa as notas de 66 alunos em um determinado teste. Qual a porcentagem de alunos que obtiveram uma nota inferior a 35?*

TABELA 22.- NOTAS DE ALUNOS EM UM TESTE

CLASSES			f_i	F_i
7	⊢	19	7	*7*
19	⊢	31	11	*18*
31	⊢	43	8	*26*
43	⊢	55	6	*32*
55	⊢	67	15	*47*
67	⊢	79	10	*57*
79	⊢	91	5	*62*
91	⊢	103	4	*66*
			66	

<u>*Solução:*</u>

Usaremos a equação dos Centís para determinar as porcentagens.
O valor observado "35" se encontra na 3ª classe, assim

$$C_i = l_i + \frac{c \times \left(E_{C_i} - F_{ant}\right)}{f_{C_i}}$$

$$35 = 31 + \frac{12 \times \left(E_{C_i} - 18\right)}{8} \Rightarrow E_{C_i} = \frac{(35-31) \times 8}{12} + 18 = 20,666...$$

Isto significa que do início da distribuição até o valor "35" existem 20,666... valores observados. Usando a equação do Elemento Centílico temos:

$$E_{C_i} = \frac{i \times n}{100} \Rightarrow i = \frac{E_{C_i} \times 100}{n} = \frac{20,666.... \times 100}{66}$$

$$i = 31,313131...\%$$

5.5 - Diagrama de Caixa ou Boxplot

O *Boxplot* (ou Diagrama de Caixa) é um gráfico de um conjunto de dados que consiste em uma linha que se estende do valor mínimo ao valor máximo, em uma caixa com linhas traçadas no primeiro Quartil (Q_1), na mediana e no terceiro Quartil (Q_3).

O *Boxplot* oferece outra maneira de visualizar a concentração dos valores dentro dos intervalos.

Para construir um *Boxplot* calcule as medidas estatísticas ($\min, \quad Q_1, \quad Md, \quad Q_3, \quad Max$), construa uma escala que abranja o valores mínimos e máximo, trace um retângulo que estenda desde o Q_1 até o Q_3. Divida o retângulo onde a Mediana se encontra. Por último, estenda dois traços, um do valor mínimo até o Q_1 e outro do Q_3 até o valor máximo.

Os *Boxplots* podem ser representados tanto na posição horizontal como na vertical.

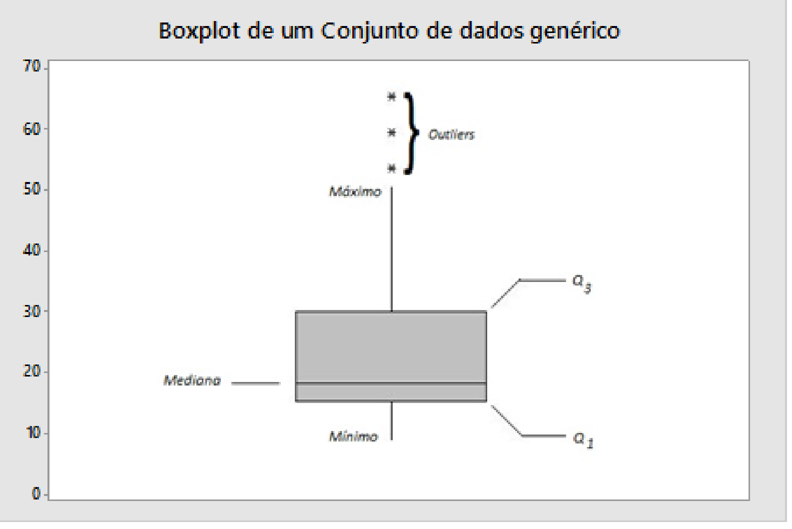


Figura 21.- Boxplot construído pelo Software Minitab

A diferença entre Q_3 e Q_1 é chamada *Amplitude Interquartílica*

$$AIQ = Q_3 - Q_1$$

Utilizando a Amplitude Interquartílica podemos determinar os Valores Atípicos (*Outliers*), isto é, valores que se presentam muito afastados dos demais valores de uma série. Usualmente os *Outliers* são valores que estão a uma distância de $1,5 \times AIQ$ abaixo do Q_1 ou acima do Q_3.

5.6 - AMPLITUDE TOTAL (A_t)

A Amplitude Total em conjunto com a Média Aritmética nos possibilita determinar um intervalo máximo onde os valores observados se encontram. Pode - se afirmar com um grau elevado de certeza que os valores observados pertencem ao intervalo:

$$[\bar{x} - A_t ; \bar{x} + A_t]$$

5.7 - DESVIO PADRÃO $(s \text{ ou } \sigma)$

O Desvio Padrão determina como os valores estão dispersos entorno da Média Aritmética de um conjunto de dados. Se o desvio Padrão for pequeno implica que os valores, em sua maioria, estão próximos da Média, caso contrário estes valores estão mais afastados entorno dessa Média.

É definido como sendo a média quadrática dos desvios tomados em relação à média aritmética.

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} \qquad \sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2 f_i}{n}}$$

O denominador n é utilizado para o cálculo do desvio padrão de populações. No caso de amostras utiliza - se o denominador $n-1$. Assim temos:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$ para conjuntos de valores

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}}$$ para valores tabulados

Existem duas outras fórmulas obtidas do desenvolvendo destas duas que determinam o valor do desvio - padrão. Estas fórmulas são denominadas *"Processo Abreviado"*.

$$s = \sqrt{\frac{1}{n-1}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right]}$$ para conjuntos de valores

$$s = \sqrt{\frac{1}{n-1}\left[\sum \left(x_i^2 f_i\right) - \frac{\left(\sum x_i f_i\right)^2}{n}\right]}$$ para valores tabulados

5.8 - VARIÂNCIA (s^2 ou σ^2)

A Variância é definida como sendo o quadrado do Desvio - Padrão, portanto suas fórmulas são idênticas às do desvio - padrão, exceto pela ausência da raiz quadrada.

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$$ para conjuntos de valores

$$s^2 = \frac{\sum(x_i - \overline{x})^2 f_i}{n-1}$$ para valores tabulados

Ou

$$s^2 = \frac{1}{n-1}\left[\sum\left(x_i^2\right) - \frac{\left(\sum x_i\right)^2}{n}\right]$$ para conjuntos de valores

$$s^2 = \frac{1}{n-1}\left[\sum\left(x_i^2 f_i\right) - \frac{\left(\sum x_i f_i\right)^2}{n}\right]$$ para valores tabulados

Exemplo 5.23 - Determine o Desvio - Padrão e a Variância da distribuição a seguir.

CLASSES			f_i
3	⊢—	12	7
12	⊢—	21	13
21	⊢—	30	17
30	⊢—	39	22
39	⊢—	48	28
48	⊢—	57	25
57	⊢—	66	14
66	⊢—	75	6
75	⊢—	84	8
			140

Solução:

O cálculo do desvio padrão direto na calculadora pode ser complicado e corre-se o risco de errar. Para minimizar este risco o cálculo será realizado através de uma tabela, isto é, desenvolveremos a fórmula ao construir a tabela. Como existem duas fórmulas para se calcular o Desvio Padrão construiremos duas tabelas,, uma para cada fórmula. Os resultados devem ser idênticos.

1ª maneira:

$$s = \sqrt{\frac{\sum(x_i - \overline{x})^2 f_i}{n-1}}$$

CLASSES	f_i	x_i	$x_i f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
3 ⊢— 12	7	7,5	52,5	-34,714286	1205,08163	8435,57143
12 ⊢— 21	13	17	214,5	-25,714286	661,22449	8595,91837
21 ⊢— 30	17	26	433,5	-16,714286	279,367347	4749,2449
30 ⊢— 39	22	35	759	-7,7142857	59,5102041	1309,22449
39 ⊢— 48	28	44	1218	1,28571429	1,65306122	46,2857143
48 ⊢— 57	25	53	1313	10,2857143	105,795918	2644,89796
57 ⊢— 66	14	62	861	19,2857143	371,938776	5207,14286
66 ⊢— 75	6	71	423	28,2857143	800,081633	4800,4898
75 ⊢— 84	8	80	636	37,2857143	1390,22449	11121,7959
	140		5910			46910,5714

A Média Aritmética da distribuição é:

$$\bar{x} = \frac{5910}{140} = 42,21428571$$

O Desvio - Padrão será:

$$s = \sqrt{\frac{46910,5714}{139}} = \sqrt{337,4861252} = 18,37079544$$

2ª maneira:

$$s = \sqrt{\frac{1}{n-1}\left[\sum\left(x_i^2 f_i\right) - \frac{\left(\sum x_i f_i\right)^2}{n}\right]}$$

CLASSES	f_i	x_i	$x_i f_i$	x_i^2	$x_i^2 f_i$
3 ⊢— 12	7	7,5	52,5	56,25	393,75
12 ⊢— 21	13	17	214,5	272,25	3539,25
21 ⊢— 30	17	26	433,5	650,25	11054,25
30 ⊢— 39	22	35	759	1190,25	26185,5
39 ⊢— 48	28	44	1218	1892,25	52983
48 ⊢— 57	25	53	1313	2756,25	68906,25
57 ⊢— 66	14	62	861	3782,25	52951,5
66 ⊢— 75	6	71	423	4970,25	29821,5
75 ⊢— 84	8	80	636	6320,25	50562
	140		5910		296397

O Desvio - Padrão da distribuição será:

$$s = \sqrt{\frac{1}{139}\left[296397 - \frac{5910^2}{140}\right]} = \sqrt{\frac{1}{139}\left(296397 - 249486,4286\right)} = \sqrt{\frac{46910,57143}{139}}$$

$$s = \sqrt{337,4861254} = 18,37079545$$

A Variância, como definido, é o quadrado do Desvio - Padrão, logo:

$$s^2 = 18,37079545^2 = 337,4861254$$

Uma opção de se obter, de maneira rápida, um valor aproximado do Desvio Padrão de um conjunto de dados é através da fórmula empírica:

$$\sigma = \frac{At}{4} = \frac{Max - Min}{4}$$

No caso do exemplo $s = \frac{84-3}{4} = 20,25$

5.9 - MEDIDAS DE DISPERSÃO RELATIVAS

Seja uma série X com média $\bar{x} = 10$ e desvio–padrão $\sigma_x = 2$ e uma série Y com média $\bar{y} = 100$ e desvio–padrão $\sigma_y = 5$. Do ponto de vista da dispersão absoluta, a série Y apresenta maior dispersão que a série X.

No entanto, levando-se em consideração as médias das séries, o desvio-padrão de Y que é 5 em relação a 100 é um valor menos significativo que o desvio-padrão de X que é 2 em relação a 10.

Isto leva-nos a definir as medidas de dispersão relativas: Coeficiente de Variação de Pearson e Variância Relativa.

O Coeficiente de Variação de Pearson de uma série é definido como:

$$CV_{(x)} = \frac{\sigma_x}{\bar{x}}$$

A Variância Relativa de uma série é indicada por:

$$V_{(x)} = \frac{\sigma^2(x)}{(\bar{x})^2}$$

Assim, se calcularmos o Coeficiente de Variação de Pearson das séries X e Y teremos:

$$CV_{(x)} = \frac{2}{10} = 0,2 = 20\%$$
$$CV_{(y)} = \frac{5}{100} = 0,05 = 5\%$$

Logo, apesar da série X ter uma menor dispersão absoluta ($\sigma_x = 2$) ela apresenta uma maior dispersão relativa (20%) do que a série Y (5%).

Como a medida de dispersão relativa leva em consideração a medida de dispersão absoluta e a média da série, é uma medida mais completa que a medida de dispersão absoluta.

Podemos afirmar que a série que tem a maior medida de dispersão relativa, tem em geral a maior dispersão.

5.10 - MEDIDAS DE ASSIMETRIA

Estas medidas referem - se à forma da curva de uma distribuição de frequências, mais especificamente do Polígono de Frequências ou do Histograma. Uma distribuição é dita simétrica quando sua Média é igual à Moda e a Mediana.

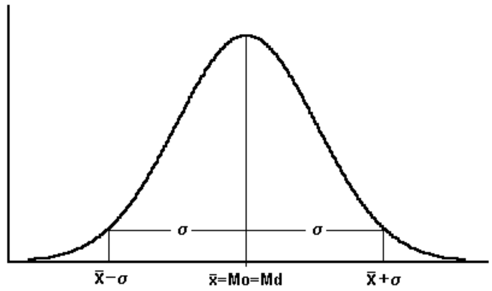


Figura 22.- Gráfico de uma curva Simétrica

Quando a Média, Mediana e a Moda recaem em pontos diferentes da distribuição, isto é, apresentam valores diferentes, a distribuição é dita Assimétrica.

Existem dois tipos de assimetria, dependendo dos valores da Média, da Mediana e da Moda.

Se $\bar{x} > Md > Mo$ a distribuição apresenta Assimetria Positiva.

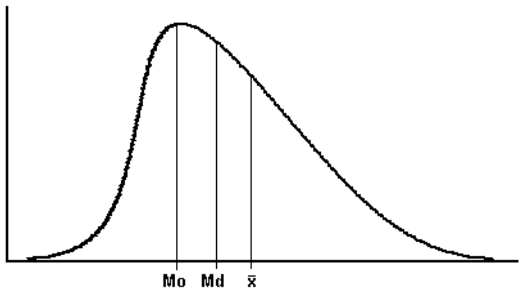


Figura 23.- Gráfico de uma curva com Assimetria Positiva

Se $\bar{x} < Md < Mo$ a distribuição apresenta Assimetria Negativa.

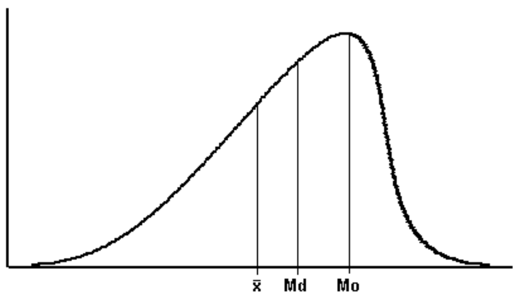


Figura 24.- Gráfico de uma curva com Assimetria Negativa

Porém a simples observação dos valores da Média, da Mediana e da Moda nos informa apenas o tipo de assimetria da distribuição, não nos sendo possível concluir sobre o grau de deformação do Polígono de Frequências.

Para determinarmos o grau de deformação da curva utilizaremos algumas medidas.

5.10.1 - *COEFICIENTES DE ASSIMETRIA DE PEARSON*

5.10.1.1 - - 1º Coeficiente de Assimetria de Pearson (e_1)

$$e_1 = \frac{\overline{x} - Mo}{\sigma}$$

Se $e_1 = 0 \Rightarrow$ a curva é simétrica.
Se $e_1 < 0 \Rightarrow$ a curva apresenta assimetria negativa.
Se $e_1 > 0 \Rightarrow$ a curva apresenta assimetria positiva.

Uma curva apresentará maior deformação quanto mais afastado de zero estiver o valor do Coeficiente de Assimetria.

5.10.1.2 - - 2º Coeficiente de Assimetria de Pearson (e_2)

$$e_2 = \frac{3(\overline{x} - Md)}{\sigma}$$

As mesmas interpretações feitas para o 1º Coeficiente de Assimetria de Pearson são válidas para o 2º Coeficiente.

5.11 - MEDIDAS DE CURTOSE

Além das deformações referentes aos diferentes valores da Média, da Mediana e da Moda, uma distribuição pode apresentar outro tipo de deformação que depende da concentração de valores em torno da sua Moda. De acordo com essa concentração uma curva pode apresentar - se mais achatada ou mais afilada. Existem três tipos de curvas:

- <u>Curva Mesocúrtica</u>: Apresenta um grau de achatamento médio, comparável a uma curva padronizada, dita Curva Normal ou Curva de Gauss.

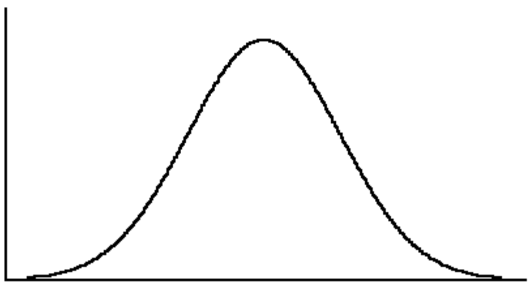


Figura 25.- Gráfico de uma curva Mesocúrtica

- <u>Curva Platicúrtica</u>: Apresenta um grau de achatamento MAIOR do que a curva padronizada.

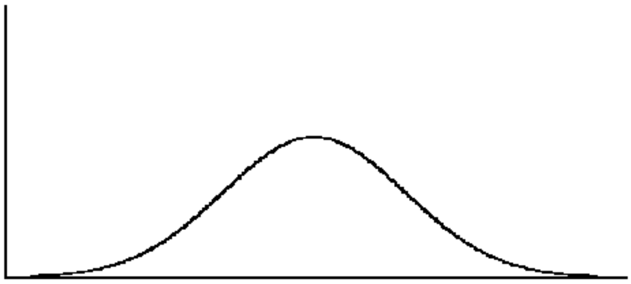


Figura 26.- Gráfico de uma curva Platicúrtica

- <u>Curva Leptocúrtica</u>: Apresenta um grau de achatamento MENOR do que a curva padronizada.

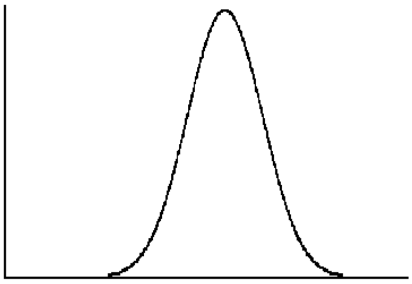


Figura 27.- Gráfico de uma curva Leptocúrtica

O grau de achatamento das curvas pode ser obtido pelo Coeficiente Percentílico de Curtose

5.11.1 - *Coeficiente Percentílico de Curtose* (K)

$$k = \frac{Q_3 - Q_1}{2(C_{90} - C_{10})}$$

Onde: Se K = 0,263 a curva será Mesocúrtica
Se K < 0,263 a curva será Leptocúrtica
Se K > 0,263 a curva será Platicúrtica.

Exemplo 5.24 - Dada a distribuição a seguir determines os Coeficientes de Assimetria e o Coeficiente de Curtose

CLASSES	f_i
3 ├── 9	7
9 ├── 15	12
15 ├── 21	17
21 ├── 27	28
27 ├── 33	21
33 ├── 39	15
39 ├── 45	8
	108

Solução:

a) *Cálculo do 1ºCoeficiente de Pearson*

CLASSES			f_i	F_i	x_i	$x_i f_i$	x_i^2	$x_i^2 f_i$
3	⊢	9	7	7	6	42	36	252
9	⊢	15	21	28	12	252	144	3024
15	⊢	21	28	56	18	504	324	9072
21	⊢	27	15	71	24	360	576	8640
27	⊢	33	12	83	30	360	900	10800
33	⊢	39	8	91	36	288	1296	10368
39	⊢	45	5	96	42	210	1764	8820
			96			2016		50976

$$\bar{x} = \frac{2016}{96} = 21$$

$$s = \sqrt{\left[\frac{1}{95}\left(50976 - \frac{2016^2}{96}\right)\right]} = 9,536632971$$

$$Mo = 15 + \frac{6\times(28-21)}{2\times 28-(21+15)} = 17,1$$

O 1º Coeficiente de Pearson é:

$$e_1 = \frac{\bar{x} - Mo}{\sigma} = \frac{21-17,1}{9,536632971} = 0,4089493652$$

Logo a distribuição é Assimétrica Positiva.

b) *Cálculo do2º Coeficiente de Pearson*

$$E_{Md} = \frac{96}{2} = 48^{\circ} \Rightarrow Md = 15 + \frac{6(48-28)}{28} = 19,28571429$$

O 2º Coeficiente de Pearson é:

$$e_2 = \frac{3(\bar{x} - Md)}{\sigma} = \frac{3(21-19,28571429)}{9,536632971} = 0,5392738869$$

c) *Cálculo do Coeficiente Percentílico de Curtose*

$$E_{Q_1} = \frac{96}{4} = 24^{\circ} \Rightarrow Q_1 = 9 + \frac{6(24-7)}{21} = 13,85714286$$

$$E_{Q_3} = \frac{3\times 96}{4} = 72^{\circ} \Rightarrow Q_3 = 27 + \frac{6(72-71)}{12} = 27,5$$

O Coeficiente Percentílico de Curtose será:

$$k = \frac{Q_3 - Q_1}{2(C_{90} - C_{10})} = \frac{27,5-13,85714286}{2\times(35,55-9,742857143)} = 0,2643232771$$

k> 0,263 logo a distribuição é Platicúrtica.

5.12 - Exercícios Propostos

5.12.1) Para cada variável indicada a seguir, determine sua Média Aritmética, sua Média Quadrática, sua Média Harmônica, sua Média Geométrica, sua Moda, sua Mediana e seus Quartís.

a) Altura dos Homens → Conjunto de dados 1 do Apêndice B
b) Altura das Mulheres → Conjunto de dados 1 do Apêndice B
c) Duração dos Filmes → Conjunto de dados 5 do Apêndice B
d) Consumo de Energia Elétrica → Conjunto de dados 9 do Apêndice B
e) Peso das Balas Azuis → Conjunto de dados 13 do Apêndice B

5.12.2) Para cada Distribuição a seguir determine:

a) Suas médias aritméticas;
b) Suas médias quadráticas;
c) Suas médias harmônicas;
d) Suas médias geométricas;
e) Suas Modas Bruta, por King e Por Czuber;
f) Suas Medianas;
g) Seus Quartís;
h) Os Decís de ordens 2, 6 e 8;
i) Os Centís de ordens 5; 10; 68; 90 e 95;
j) Seus Desvios-Padrão;
k) Suas Variâncias;
l) Os Coeficientes de Assimetria;
m) Os coeficientes de Curtose;
n) Os Coeficientes de Variação de Pearson.
o) Construa o Bloxplot correspondente.
p)

Distribuição A:

CLASSES	f_i					
4 ⊢ 13	9					
13 ⊢ 22	5					
22 ⊢ 31	7					
31 ⊢ 40	4					
40 ⊢ 49	19					
49 ⊢ 58	11					
58 ⊢ 67	8					
67 ⊢ 76	2					
76 ⊢ 85	6					

Distribuição B:

CLASSES	f_i					
2 ⊢ 13	6					
13 ⊢ 24	5					
24 ⊢ 35	12					
35 ⊢ 46	19					
46 ⊢ 57	11					
57 ⊢ 68	17					
68 ⊢ 79	3					
79 ⊢ 90	6					
90 ⊢ 101	8					

Distribuição C:

CLASSES	f_i					
6 ⊢ 19	2					
19 ⊢ 32	7					
32 ⊢ 45	4					
45 ⊢ 58	9					
58 ⊢ 71	3					
71 ⊢ 84	9					
84 ⊢ 97	5					
97 ⊢ 110	1					
110 ⊢ 123	8					

5.12.3) A tabela de Distribuição de Frequências a seguir se refere aos problemas detectados em uma amostra de 350 aparelhos de TV numa linha de produção de um fábrica durante um período de 60 meses.

APARELHOS DE TELEVISÃO DA MARCA KLX QUE APRESENTARAM PROBLEMAS NUM PERÍODO DE 60 MESES – 1998 A 2002

MÊS	QUANTIDADE
0 ⊢ 7,5	3
7,5 ⊢ 15	16
15 ⊢ 22,5	23
22,5 ⊢ 30	36
30 ⊢ 37,5	54
37,5 ⊢ 45	169
45 ⊢ 52,5	49
	350

FONTE: Laboratório de testes da fábrica

Determine:

a) A vida média de um aparelho de televisão;
b) Qual o tempo, em meses, necessário para que 50% dos aparelhos apresentem um problema;
c) A Moda por Czuber, isto é, qual a data onde se observou o maior número de aparelhos com problemas (em meses e em dias);
d) Quanto tempo é necessário para que 63% dos aparelhos apresentem problemas;
e) Se o fabricante repõe todos os aparelhos que apresentam problemas dentro do prazo de garantia, qual deve se o prazo estipulado no termo de garantia para que apenas 6% dos aparelhos sejam substituídos;
f) Qual a porcentagem de aparelhos apresenta defeitos com mais de 48 meses.

5.12.4) O departamento de controle de qualidade da empresa KLX fez um teste em 255 componentes fabricados por ela para detectar vários aspectos sobre seu produto. Depois do teste foi construída a seguinte tabela:

TESTE DE FUNCIONAMENTO DOS COMPONENTES FABRICADOS PELA

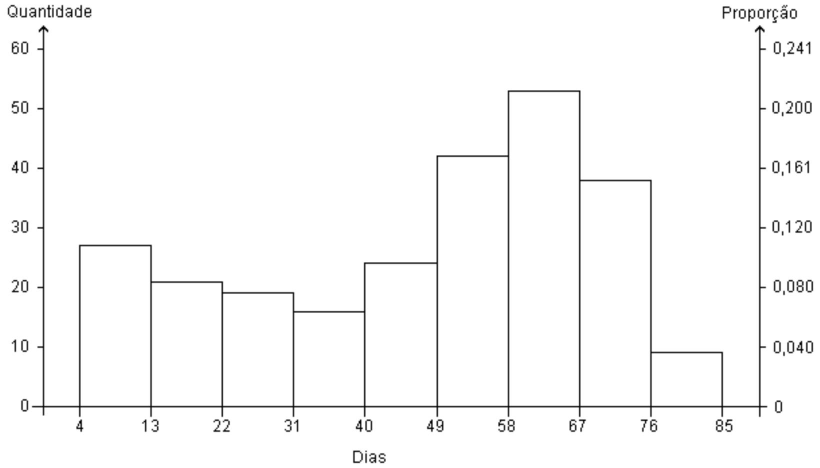
DIAS			COMPONENTES EM FUNCIO-NAMENTO
0	⊢	95	255
95	⊢	190	240
190	⊢	285	228
285	⊢	380	183
380	⊢	475	171
475	⊢	570	163
570	⊢	665	55
665	⊢	760	15
760	⊢	855	3

Determine:

a) A vida média de um componente;
b) Em que dia observou-se o maior número de componentes danificados;
c) O tempo necessário para que 30% dos componentes estejam danificados;
d) O custo total de substituição de um componente danificado é de R$ 97,00. De acordo com os cálculos do financeiro, num lote de 255 componentes a empresa não pode gastar mais do que R$ 9.506,00 para que o lucro não seja inferior à meta estipulada pela diretoria. Determine o prazo de garantia para que isto aconteça;
e) A porcentagem de componentes danificados no prazo de um ano;

5.12.5) Uma empresa deseja determinar o quão eficientes são suas embalagens. Para tal ela realizou um teste embalando legumes e observando quando eles apresentam traços de deterioração. O gráfico abaixo ilustra os resultados.

TESTE DE DETERIORAÇÃO EM LEGUMES EMBALADOS À VACUO E MANTIDOS À TEMPERATURA DE 10 GRAUS CELSUS NAS – JANEIRO DE 2003



FONTE: Laboratório da empresa

a) Construa a Tabela de Distribuição de Frequências que ele representa;
b) Calcule o tempo médio necessário para um legume apresentar algum traço de deterioração;
c) Determine o dia que apresentou o maior número de legumes com traços de deterioração;
d) O tempo médio quadrático para que os legumes apresentassem traços de deterioração;
e) O desvio padrão do tempo necessário para que os legumes apresentem traços de deterioração;
f) A porcentagem de legumes com traços de deterioração com no máximo 45 dias;
g) O tempo mediano de deterioração dos legumes.

5.12.6) A tabela a seguir representa a distribuição salarial dos empregados da empresa XYZ e o quadro a seguir ilustra duas propostas de aumentos salariais que foram apresentadas,

SALÁRIOS MÍNIMOS			EMPREGADOS
0	├──	4,5	363
4,5	├──	9,0	491
9,0	├──	13,5	312
13,5	├──	18,0	226
18,0	├──	22,5	130
22,5	├──	27,0	47
27,0	├──	31,5	9
	$\sum$		1578

1ª Proposta		2ª Proposta	
Faixa salarial	Aumento	Faixa salarial	Aumento
Até 5,2	20%	Até 3,7	20%
De 5,2 a 11	15%	De 3,7 a 6,3	17%
De 11 a 17,3	10%	De 6,3 a 10,5	12%
Mais de 17,3	0%	De 10,5 a 15	8%
		De 15 a 19	4%
		De 19 a 23	2%
		Mais de 23	0%

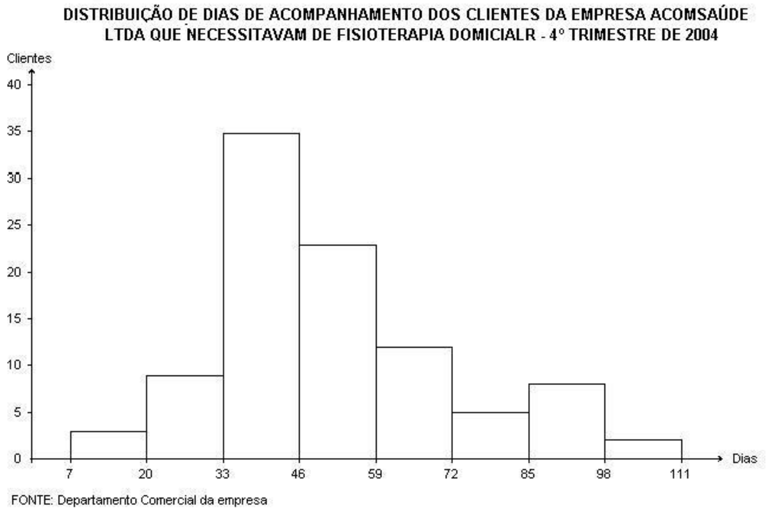
Determine:

a) O salário médio dos empregados;
b) O salário mediano dos empregados;
c) Os salários modais, por King e Czuber, dos empregados;
d) O desvio-padrão dos salários dos empregados;
e) A porcentagem de empregados que recebem um salário no intervalo $\bar{x} \pm s$;
f) A porcentagem de empregados agraciados com algum aumento na 1ª proposta?
g) A porcentagem de empregados agraciados com algum aumento na 2ª proposta?
h) A previsão da folha de pagamento após os aumentos da 1ª proposta?
i) A previsão da folha de pagamento após os aumentos da 2ª proposta?
j) A variação do salário médio, após o aumento da 1ª proposta, em relação ao salário médio antes do aumento?
k) A variação do salário médio, após o aumento da 2ª proposta, em relação ao salário médio antes do aumento?

5.12.7) A Empresa AcompSaude Ltda. presta serviço de acompanhamento médico domiciliar pós-cirúrgico a todos os clientes de necessitam de enfermeiras, fisioterapeutas, transporte às clinicas e/ou laboratórios e outros serviços do gênero. Devido às dificuldades em atender seus clientes no fim do ano passado, a empresa resolveu fazer um estudo a fim de estimar o número necessário de empregados para que não haja problemas. O gráfico a seguir apresenta a distribuição de dias de atendimento apenas dos clientes que necessitaram de fisioterapia domiciliar.

Determine:

a) A Tabela de Distribuição de Frequências que represente estes dados;
b) O tempo médio de atendimento aos clientes, em dias;
c) Qual o número modal de atendimento;
d) Qual a porcentagem de clientes que necessitou de acompanhamento por mais de 72 dias?
e) Qual a porcentagem de clientes que necessitou de acompanhamento por menos de 46 dias?



5.12.8) O gerente da loja XYK está desconfiado que sua loja esteja perdendo vendas devido ao número de clientes que entram na loja e não são atendidos por não haver vendedores suficientes, no entanto a contratação de vendedores está vinculada ao número de clientes não atendidos aos valores das vendas não efetuadas devido ao não atendimento. Para determinar se deve contratar mais vendedores ele realizou uma pesquisa durante um mês inteiro, anotando o número de clientes que ao foram atendidos e qual o possível valor que estes clientes gastariam na loja. Além disso determinou algumas diretivas para a contratação. Os dados coletados com relação ao número de clientes não atendidos diariamente e um gráfico representando o valor das vendas não realizadas estão a seguir.

Número médio de Clientes não atendidos diariamente	Valor das Vendas não Efetuadas Diariamente (R$)	Número de vendedores a serem contratados
Menor que 6	------	0
De 6 a 8	Menos de 110,00	1
De 8 a 10	De 110,00 a 190,00	2
De 11 a 12	De 190,00 a 270,00	3
De 12 a 14	De 270,00 a 360,00	4
De 14 a 17	De 360 a 450,00	5
------	Mais de 450,00	6

NÚMERO DIÁRIO DE CLIENTES NÃO ATENDIDOS NA LOJA XYK – AGOSTO DE 2006

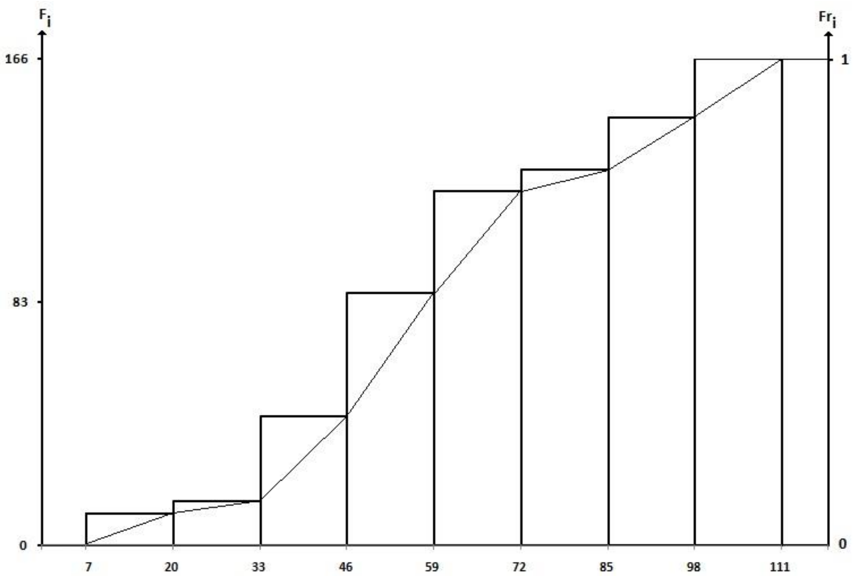
15	16	12	14	12	14
16	13	11	14	10	14
13	14	8	14	15	14
13	13	20	8	18	15
15	13	13	10	20	17

Obs.: Pessoas que estavam acompanhando os clientes não foram anotadas.

FONTE: Pesquisa realizad na própria loja

a) Construa uma Distribuição de Frequências e um Histograma representando o número de clientes não atendidos diariamente;
b) Construa uma Distribuição de Frequências representando o valor das vendas não efetuadas;
c) Determine se o gerente deve contratar novos vendedores e quantos devem ser contratados, se for o caso;
d) Determine, também, as Modas, os Quartís e os Centís 10 e 90 das duas distribuições.

5.12.9) Uma empresa realizou um teste em seus equipamentos em diferentes temperaturas, anotando o número total de aparelhos que apresentaram defeitos até a temperatura naquele instante. O gráfico a seguir ilustra os resultados.



FONTE: Departamento de controle de qualidade da empresa

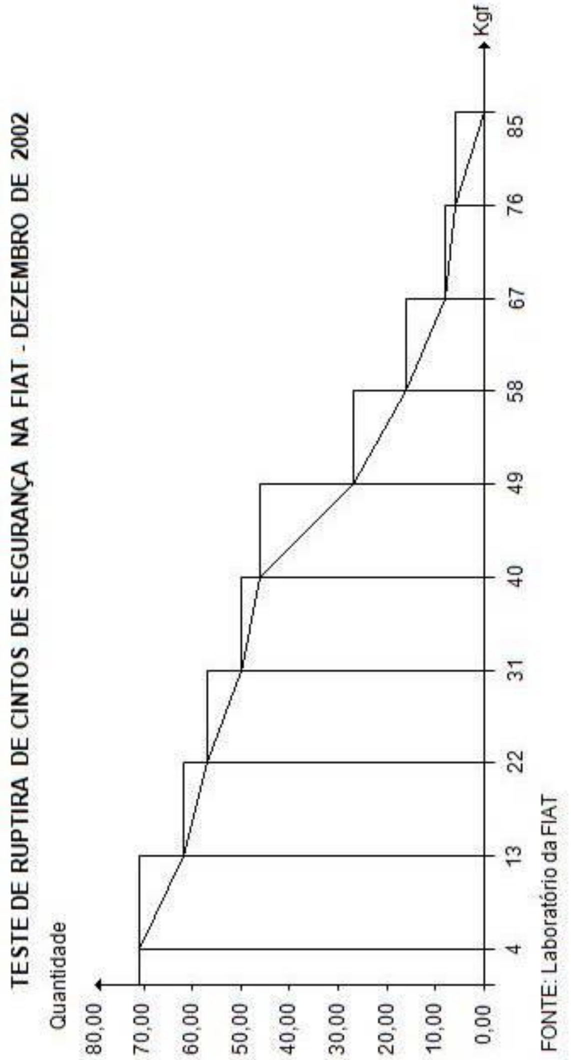
Determine:
a) A Distribuição de Frequências que representa este gráfico;
b) A temperatura média em que os aparelhos apresentaram defeitos;

c) Em que temperatura ocorreu a maior ocorrência de defeitos;
d) A porcentagem de aparelhos que apresentaram defeitos acima de 85 graus;
e) A porcentagem de aparelhos que apresentaram defeitos entre 46 e 72 graus;
f) A porcentagem de aparelhos que apresentaram defeitos com menos de 87,3 graus;
g) O desvio padrão da temperatura em que os aparelhos apresentaram defeitos.

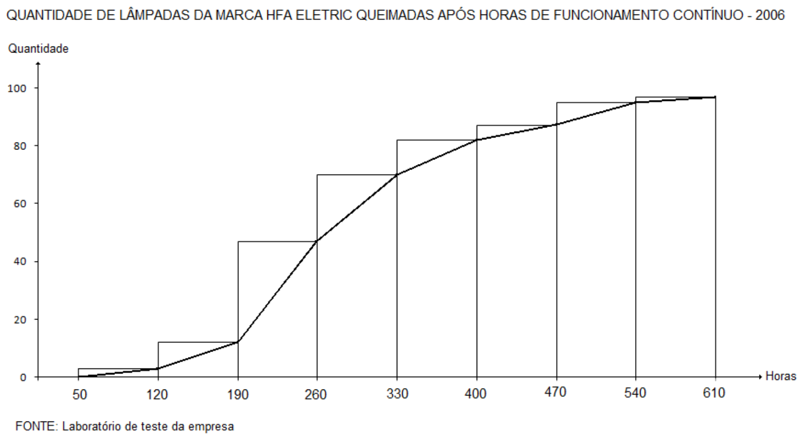
5.12.10) O gráfico a seguir representa os resultados obtidos em testes com cintos de segurança realizado no laboratório da FIAT Automóveis em dezembro de 2002.

Determine:

a) A tabela de Distribuição de Frequências que represente os dados do gráfico;
b) A tensão média que um Cinto de Segurança suporta em Kgf;
c) Qual o valor de tensão em que se observou o maior número de rupturas nos Cintos de Segurança;
d) Qual a força necessária para que 30% dos Cintos de Segurança fossem rompidos;
e) A variação média da tensão, em Kgf, necessária para a ruptura dos Cintos de Segurança;
f) A porcentagem de Cintos de Segurança que suportaram uma tensão superior $\overline{x}+2s$; e
g) A porcentagem de Cintos de Segurança que romperam com menos de 50 Kgf.



5.12.11) O departamento de compras de uma empresa deve decidir entre continuar usando as lâmpadas da marca Velha e passar a utilizar lâmpadas de uma nova marca. As lâmpadas antigas apresentam uma vida útil de 270 com uma variação média de 67,5 horas. A empresa da nova marca enviou o gráfico a seguir ilustrando a duração de suas lâmpadas.



Determine:

a) Utilizando o gráfico enviado pela empresa da nova marca, construa uma distribuição de Frequências para os tempos de duração das lâmpadas dessa empresa;
b) Determine em que horário ocorreu o maior número de lâmpadas da nova marca queimadas;
c) O número de horas necessárias para se ter 15% das lâmpadas queimadas;
d) A duração média das lâmpadas da nova marca;
e) A variação média do tempo de vida de uma lâmpada da nova marca;
f) Determine qual marca deve ser escolhida pelo departamento de compras, baseado na duração média de cada marca, na sua variação média e considerando, ainda que, o custo das duas marcas é o mesmo e que a confiabilidade é um item importante.