

*Conceitos Iniciais de Estatística –Módulo 4 :*  
**GENERALIDADES SOBRE ESTATÍSTICA DESCRITIVA**  
*Prof. Rogério Rodrigues*



## **CONCEITOS INICIAIS DE ESTATÍSTICA:**

► **GENERALIDADES SOBRE ESTATÍSTICA DESCRITIVA**

**CURSO : ADMINISTRAÇÃO**

**PERÍODO : 4º**

### **1) ESTATÍSTICA DESCRITIVA & ESTATÍSTICA INFERENCIAL :**

Como sabemos, a Estatística é a ciência que lida com dados; então ela envolve a coleta desses dados, a classificação, a sumarização, a organização, a análise e a interpretação desses dados. Nesse mister, a Estatística tem duas frentes básicas de abordagem: Descrição e Inferência.

**1.1) A Estatística Descritiva** é a frente de abordagem que cuida da organização, da sumarização e da descrição de um conjunto de dados. Fazem parte desse conjunto de procedimentos o cálculo de medidas de posição e de dispersão, a construção de tabelas e a construção de gráficos.

**Exemplo Ilustrativo 1:** Considere uma coleta de dados realizada numa classe de 50 alunos, cujo objetivo foi o levantamento do número de pessoas do grupo familiar de cada um desses indivíduos. O entrevistador, depois das entrevistas, saiu com a seguinte planilha de dados:

|   |   |   |   |   |
|---|---|---|---|---|
| 2 | 5 | 2 | 2 | 6 |
| 3 | 4 | 6 | 3 | 1 |
| 4 | 6 | 3 | 3 | 6 |
| 4 | 4 | 1 | 5 | 3 |
| 6 | 6 | 3 | 5 | 4 |
| 1 | 1 | 4 | 4 | 3 |
| 5 | 5 | 2 | 3 | 6 |
| 3 | 1 | 4 | 4 | 5 |
| 5 | 4 | 3 | 6 | 4 |
| 2 | 2 | 5 | 3 | 6 |

Como esses dados, desagrupados como se apresentam, têm pouca expressão objetiva, o pesquisador resolveu tabulá-los com as medidas legendadas a seguir:

$f_i$  → frequência absoluta       $x_i$  → cada uma das variáveis ( número de pessoas de cada grupo familiar)

$f_r$  → frequência relativa       $F_b$  → frequência acumulada para baixo

$F_c$  → frequência acumulada para cima       $F_{br}$  → frequência relativa acumulada para baixo

$F_{cr}$  → frequência relativa acumulada para cima       $\Delta x$  → desvio em relação à média =  $|x_i - \text{média}|$

$d_M$  → Desvio médio       $S^2$  → Variância       $S$  → Desvio Padrão       $C_v$  → Coeficiente de variação

| $x_i$        | $f_i$     | $f_i \cdot x_i$ | $f_r$ (%)  | $F_b$ | $F_{br}$ (%) | $F_c$ | $F_{cr}$ (%) | $\Delta x$ | $f_i \cdot \Delta x_i$ | $f_i \cdot \Delta x_i^2$ |
|--------------|-----------|-----------------|------------|-------|--------------|-------|--------------|------------|------------------------|--------------------------|
| 1            | 5         | 5               | 10         | 5     | 10           | 50    | 100          | 2,76       | 13,8                   | 38,09                    |
| 2            | 6         | 12              | 12         | 11    | 22           | 45    | 90           | 1,76       | 10,56                  | 18,59                    |
| 3            | 11        | 33              | 22         | 22    | 44           | 39    | 78           | 0,76       | 8,36                   | 6,35                     |
| 4            | 11        | 44              | 22         | 33    | 66           | 28    | 56           | 2,24       | 24,62                  | 55,19                    |
| 5            | 8         | 40              | 16         | 41    | 82           | 17    | 34           | 1,24       | 9,92                   | 12,30                    |
| 6            | 9         | 54              | 18         | 50    | 100          | 9     | 18           | 2,24       | 20,16                  | 45,16                    |
| <b>TOTAL</b> | <b>50</b> | <b>188</b>      | <b>100</b> |       |              |       |              |            | <b>87,44</b>           | <b>175,68</b>            |

► **Medidas de Posição (tendência central):**

a) **Média** =  $\frac{\sum f_i \cdot x_i}{\sum f_i} = \frac{188}{50} = 3,76$ .

b) **Moda** = 3 ou 4 pessoas (bimodal).

c) **Mediana** = 4 pessoas.

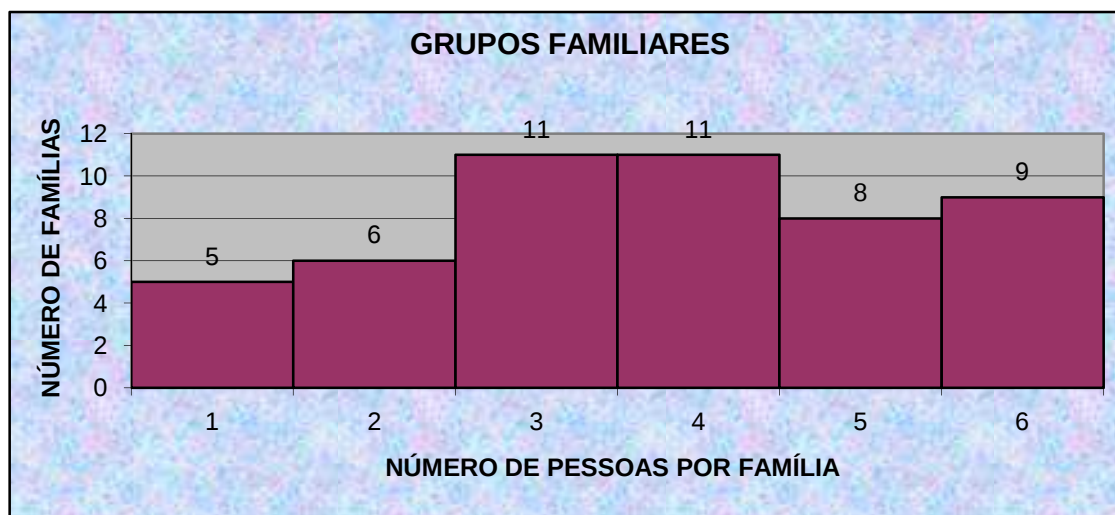
► **Medidas de dispersão**

a) **d<sub>M</sub>** =  $\frac{\sum f_i \cdot \Delta x_i}{\sum f_i} = \frac{87,44}{50} = 1,75$ .

b) **S<sup>2</sup>** =  $\frac{\sum f_i \cdot \Delta x_i^2}{\sum f_i - 1} = \frac{175,68}{50 - 1} = 3,59$ .

c) **S** =  $\sqrt{S^2} = \sqrt{3,59} = 1,9$ .

► **Representação Gráfica (Histograma):**



Todos os procedimentos gerados a partir da planilha primitiva, frequências, medidas e gráfico ilustram a abordagem da Estatística Descritiva.

**1.2) A Estatística Inferencial** cuida dos métodos que tornam possível estudar as características de uma população com base em uma amostra dessa população. Seu objetivo de ponta é estabelecer condições de confiabilidade de inferências sobre atributos de uma população através de um processo de amostragem. Para isso, a Teoria das Probabilidades provê e regula a possibilidade de acerto, de modo que os resultados obtidos com uma amostra reflitam, de fato, os resultados característicos da população.

**Exemplo Ilustrativo 2:** Numa pesquisa sobre o processo eleitoral no Brasil, os institutos encarregados dessa tarefa nunca entrevistam todos os eleitores da região onde ocorre a pesquisa, ou seja, não é toda a população que é entrevistada, mas uma amostra criteriosamente escolhida dessa população. Com base nas respostas dessa amostra, os resultados são entendidos a toda população, depois de análise criteriosa e metódica, em que as probabilidades de cada candidato são verificadas, inclusive com margens de erros determinadas.

## 2) COMPLEMENTOS PARA A ESTATÍSTICA DESCRITIVA :

### 2.1) Sobre as Medidas de Posição - Separatrizes :

A determinação da Mediana de uma série de variáveis implica no estabelecimento de uma variável que ocupa a posição central, quando essas variáveis estão ordenadas, e, conseqüentemente, divide o grupo de variáveis em dois subgrupos: o subgrupo das variáveis menores ou iguais à mediana e o subgrupo das variáveis maiores ou iguais à Mediana.

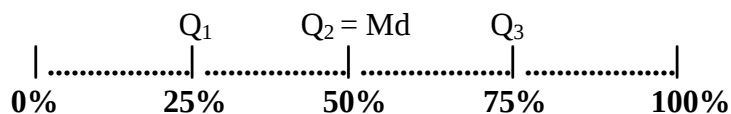
**Exemplo Ilustrativo 3:** As idades das 11 moças que trabalham em uma repartição pública são dadas pela tabela:

|    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|
| 18 | 25 | 19 | 22 | 27 | 20 | 19 | 22 | 20 | 30 | 29 |
|----|----|----|----|----|----|----|----|----|----|----|

Ordenando crescentemente essas idades, tem-se: 18 – 19 – 19 – 20 – 20 – **22** – 22 – 25 – 27 – 29 – 30 , em que 22 ocupa a posição central e separa os subgrupos 18 – 19 – 19 – 20 – 20 (menores ou iguais a 22) e 22 – 25 – 27 – 29 – 30 (maiores ou iguais a 22). A mediana é 22.

De modo análogo, uma série de variáveis pode ser dividida em 4 subgrupos (**Quartis**), 10 subgrupos (**Decis**) ou 100 subgrupos (**Percentis**).

**2.1.1) Quartis** : A série de variáveis fica assim dividida :



→ Fórmula para determinar o 1º quartil: 
$$Q_1 = \lambda_1 + \frac{\left(\frac{n}{4} - \sum F_{b.ant}\right).h}{f_1}$$

→ Fórmula para determinar o 3º quartil: 
$$Q_3 = \lambda_3 + \frac{\left(\frac{3n}{4} - \sum F_{b.ant}\right).h}{f_3}$$

**Exemplo Ilustrativo 4:** A distribuição abaixo registra as idades de 56 pessoas de uma grande família. Determine todos os quartis dessa distribuição.

| IDADES (anos)        | $f_i$ | $F_{bi}$ |
|----------------------|-------|----------|
| [7 , 17[             | 6     | 6        |
| [17 , 27[            | 15    | 21       |
| [27 , 37[            | 20    | 41       |
| [37 , 47[            | 10    | 51       |
| [47 , 57[            | 5     | 56       |
| $\Sigma \rightarrow$ | 56    |          |

1º) Como  $n = 56$  variáveis, tem-se que  $\frac{n}{4} = \frac{56}{4} = 14^{\text{o}}$  termo  $\Rightarrow$  Na coluna de  $F_{bi}$  identificamos a classe  $Q_1$ , ou seja, a classe em que  $F_{bi} \geq 14$ . Então, tem-se a 2ª classe:  $\lambda_1 = 17$ ;

2º) Como  $n = 56$  variáveis, tem-se que  $\frac{n}{2} = \frac{56}{2} = 28^{\text{o}}$  termo  $\Rightarrow$  Na coluna de  $F_{bi}$  identificamos a classe  $Q_2$ , ou seja, a classe em que  $F_{bi} \geq 28$ . Então, tem-se a 3ª classe:  $\lambda_2 = 27$ ;

3º) Como  $n = 56$  variáveis, tem-se que  $\frac{3n}{4} = \frac{3 \cdot 56}{4} = 42^{\text{o}}$  termo  $\Rightarrow$  Na coluna de  $F_{bi}$  identificamos a classe  $Q_3$ , ou seja, a classe em que  $F_{bi} \geq 42$ . Então, tem-se a 4ª classe:  $\lambda_3 = 37$ ;

Aplicando-se as fórmulas, tem-se:

$$\rightarrow Q_1 = 17 + \left( \frac{14 - 6}{15} \right) \cdot 10 = 22,33$$

$$\rightarrow Q_2 = 27 + \left( \frac{28 - 21}{20} \right) \cdot 10 = 30,5$$

$$\rightarrow Q_3 = 37 + \left( \frac{42 - 41}{10} \right) \cdot 10 = 38$$

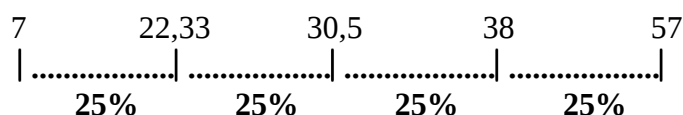
Então, podemos afirmar que:

25% das variáveis observadas estão entre 7 anos e 22,33 anos;

25% das variáveis observadas estão entre 22,33 anos e 30,5 anos;

25% das variáveis observadas estão entre 30,5 anos e 38 anos;

25% das variáveis observadas estão entre 38 anos e 57 anos;



**2.1.2) Decis** : A série de variáveis fica dividida em 10 subgrupos de 10% em 10%. Cada decil é calculado de modo análogo aos quartis, pela fórmula:

$$D_i = \lambda_i + \frac{\left( \frac{i \cdot n}{10} - \sum F_{b.ant} \right) \cdot h}{f_i}$$

**2.1.3) Percentis** : A série de variáveis fica dividida em 100 subgrupos de 1% em 1%. Cada Percentil é calculado de modo análogo aos decis, pela fórmula:

$$P_i = \lambda_i + \frac{\left(\frac{i}{100} - \sum F_{b.ant}\right).h}{f_i}$$

**Exemplo Ilustrativo 5:** Determinar o 4º decil e o 72º percentil da distribuição abaixo. A distribuição registra o número de goals de 40 artilheiros internacionais.

| Nº DE GOOLS | f <sub>i</sub> | F <sub>bi</sub> |
|-------------|----------------|-----------------|
| [4 , 9[     | 8              | 8               |
| [9 , 14[    | 12             | 20              |
| [14 , 19[   | 17             | 37              |
| [19 , 24[   | 3              | 40              |
| Σ→          | 40             |                 |

$$\text{Sabe-se que } n = 40 \Rightarrow \begin{cases} \frac{i}{10} = \frac{4.40}{10} = 16 \Rightarrow \lambda_4 = 9, F_{b.ant} = 8 \text{ e } h = 5 \\ \frac{i}{100} = \frac{72.40}{100} = 28,8 \Rightarrow \lambda_{100} = 14, F_{b.ant} = 20 \text{ e } h = 5 \end{cases}$$

$$\text{Então, } D_4 = 9 + \frac{(16-8).5}{12} = 12,33 \quad \text{e} \quad P_{72} = 14 + \frac{(28,8-20).5}{17} = 16,59$$

**OBS :** A Mediana sempre está entre os limites  $\lambda$  e L da classe mediana, enquanto  $\frac{\sum f_i}{2}$  está entre  $F_{b.ant}$  e

$F_{b.post}$ . Numa regra de três, pode-se escrever que  $\frac{Md - \lambda}{L - \lambda} = \frac{\frac{\sum f_i}{2} - F_{b.ant}}{F_b - F_{b.ant}}$ , em que  $L - \lambda = h$  e

$F_b - F_{b.ant} = f_i$ . Daí, tem-se  $(Md - \lambda).f_i = \left(\frac{\sum f_i}{2} - F_{b.ant}\right).h \Rightarrow (Md - \lambda) = \frac{\left(\frac{\sum f_i}{2} - F_{b.ant}\right).h}{f_i}$  que será a

fórmula

$$Md = \lambda + \frac{\left(\frac{\sum f_i}{2} - F_{b.ant}\right).h}{f_i}$$

### Exercício Proposto:

Em cada série de variáveis apresentadas a seguir, determine o que se pede.

a) A mediana, os quartis e o sexto decil das notas de Estatística de 35 alunos, dadas no quadro a seguir:

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 6  | 6  | 7  | 8  | 8  | 8  |
| 9  | 10 | 10 | 11 | 12 | 13 |
| 14 | 14 | 15 | 16 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 29 | 30 | 31 | 32 |    |

b) A mediana, os quartis, o terceiro decil e o vigésimo percentil das massas, em kg, de 180 pacientes de uma clínica:

| MASSAS         | Nº DE PACIENTES | F <sub>b</sub> |
|----------------|-----------------|----------------|
| 30             | 40              | 40             |
| 40             | 32              | 72             |
| 50             | 35              | 107            |
| 60             | 24              | 131            |
| 70             | 20              | 151            |
| 80             | 29              | 180            |
| <b>TOTAL →</b> | <b>180</b>      |                |

c) A mediana, os quartis, o sétimo decil e o octogésimo percentil dos salários de 22 funcionários de uma empresa:

| i              | SALÁRIOS  | f <sub>i</sub>  | f <sub>b</sub> |
|----------------|-----------|-----------------|----------------|
| 1              | 300 — 320 | 4               |                |
| 2              | 320 — 340 | 3               |                |
| 3              | 340 — 360 | 5               |                |
| 4              | 360 — 380 | 8               |                |
| 5              | 380 — 400 | 2               |                |
| <b>TOTAL →</b> |           | $\sum f_i = 22$ |                |

d) A mediana, os quartis, o quarto decil e o trigésimo percentil do número de jogos realizados por 45 jogadores profissionais:

| i              | JOGOS     | f <sub>i</sub>  | f <sub>b</sub> |
|----------------|-----------|-----------------|----------------|
| 1              | 125 — 175 | 8               |                |
| 2              | 175 — 225 | 6               |                |
| 3              | 225 — 275 | 9               |                |
| 4              | 275 — 325 | 7               |                |
| 5              | 325 — 375 | 9               |                |
| 6              | 375 — 425 | 6               |                |
| <b>TOTAL →</b> |           | $\sum f_i = 45$ |                |

e) A mediana, os quartis, o nono decil e o sexagésimo quinto percentil das contribuições de 121 empregados ao INSS, assim distribuídos por classe: [54, 74[ → 12, [74, 94[ → 11, [94, 114[ → 11, [114, 134[ → 13, [134, 154[ → 16, [154, 174[ → 19, [174, 194[ → 21 e [194, 214[ → 18.

## 2.2) Sobre as Medidas de Dispersão :

**2.2.1) Sobre a Variância amostral :** A variância  $S^2$  é, genericamente, a soma dos quadrados dos desvios das variáveis, em relação à média dessas variáveis, dividida pelo número de variáveis, para variáveis não agrupadas, ou seja,

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Nessa fórmula, pode-se desenvolver o somatório  $\sum (x_i - \bar{x})^2$  como  $\sum (x_i^2 + \bar{x}^2 - 2x_i\bar{x}) = \sum x_i^2 + n\bar{x}^2 -$

$$- 2\bar{x}\sum x_i = \sum x_i^2 + n\left(\frac{\sum x_i}{n}\right)^2 - \frac{2(\sum x_i)(\sum x_i)}{n} = \sum x_i^2 + \frac{n\sum x_i^2}{n^2} - \frac{2\sum x_i^2}{n} = \sum x_i^2 - \frac{\sum x_i^2}{n}.$$

. Então, tem-se  $S^2 = \frac{1}{n} \left[ \sum x_i^2 - \frac{\sum x_i^2}{n} \right]$  para variáveis não agrupadas. Para variáveis agrupadas, basta acrescentar o fator relativo à frequência absoluta, ou seja,

$$S^2 = \frac{\sum f_i \cdot (x_i - \bar{x})^2}{n} = \frac{1}{n} \left[ \sum f_i \cdot x_i^2 - \frac{\sum f_i \cdot x_i^2}{n} \right]$$

Lembre-se que  $n = \sum f_i$

**2.2.2) Sobre o Desvio Padrão :** Como o desvio padrão é a raiz quadrada da variância, temos então:

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

► **Observação Importante :** No desenvolvimento anterior da fórmula da variância, a última etapa, que foi  $S^2 = \sum x_i^2 - \frac{\sum x_i^2}{n}$  pode ser escrita como  $S^2 = \frac{(n-1) \sum x_i^2}{n}$  para variáveis não agrupadas. Essa

expressão implica que o desvio padrão é  $S = \sqrt{\frac{(n-1)}{n} \sum x_i^2}$ ; isso parece razoável quando as variáveis são representativas da população e não da amostra. Por isso, aplica-se a correção de Bessel, que considera o denominador  $n - 1$  no lugar de  $n$ . Então, tem-se:

|           | VARIÂNCIA                                                                                    | DESVIO PADRÃO                                                                                     |
|-----------|----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| POPULAÇÃO | $S^2 = \frac{1}{n} \left[ \sum x_i^2 \cdot f_i - \frac{(\sum x_i \cdot f_i)^2}{n} \right]$   | $S = \sqrt{\frac{1}{n} \left[ \sum x_i^2 \cdot f_i - \frac{(\sum x_i \cdot f_i)^2}{n} \right]}$   |
| AMOSTRA   | $S^2 = \frac{1}{n-1} \left[ \sum x_i^2 \cdot f_i - \frac{(\sum x_i \cdot f_i)^2}{n} \right]$ | $S = \sqrt{\frac{1}{n-1} \left[ \sum x_i^2 \cdot f_i - \frac{(\sum x_i \cdot f_i)^2}{n} \right]}$ |

### 2.2.3) Interpretação do Desvio Padrão :

a) **Regra Empírica:** Veja a tabela abaixo, considerando uma distribuição amostral com média  $\bar{x}$  e desvio padrão S.

| INTERVALO       | CONTEUDO DO INTERVALO                           | DISTRIBUIÇÕES APROX. SIMÉTRICAS     | DISTRIBUIÇÕES FORT. SIMÉTRICAS      |
|-----------------|-------------------------------------------------|-------------------------------------|-------------------------------------|
| $\bar{x} \pm S$ | Entre 60% e 80% de todas as variáveis amostrais | 70% de todas as variáveis amostrais | 90% de todas as variáveis amostrais |

| INTERVALO        | CONTEUDO DO INTERVALO PARA DISTRIBUIÇÃO SIMÉT. | CONTEÚDO P/ DIST. COM ASSIMETRIA EL. |
|------------------|------------------------------------------------|--------------------------------------|
| $\bar{x} \pm 2S$ | 95% de todas as variáveis amostrais            | 100% de todas as variáveis amostrais |

| INTERVALO        | CONTEUDO DO INTERVALO PARA DISTRIBUIÇÃO SIMÉT. |
|------------------|------------------------------------------------|
| $\bar{x} \pm 3S$ | 100% de todas as variáveis amostrais           |

### b) Teorema de Tchebycheff :

Para qualquer distribuição amostral com média  $\bar{x}$  e desvio padrão S:

- O intervalo  $\bar{x} \pm 2S$  contém, no mínimo, 75% de todas as variáveis amostrais.
- O intervalo  $\bar{x} \pm 3S$  contém, no mínimo, 89% de todas as variáveis amostrais.

**Exemplo Ilustrativo 6:** Calcular a variância e o desvio padrão do número de goals de 40 jogadores de futebol profissional ao longo de um mesmo período. Interpretar depois o desvio padrão da distribuição pelas duas regras anteriores.

| Nº DE GOOLS          | $f_i$ | $x_i$ | $f_i \cdot x_M$ | $x_M^2 \cdot f_i$ |
|----------------------|-------|-------|-----------------|-------------------|
| [4 , 9[              | 8     | 6,5   | 52              | 338               |
| [9 , 14[             | 12    | 11,5  | 138             | 1.587             |
| [14 , 19[            | 17    | 16,5  | 280,5           | 4.628,25          |
| [19 , 24[            | 3     | 21,5  | 64,5            | 1.386,75          |
| $\Sigma \rightarrow$ | 40    |       | 535             | 7.940,00          |

1º) A média amostral será  $\frac{535}{40} = 13,38$  gols.

2º) A variância será  $S^2 = \frac{1}{39} \left[ 7.940 - \frac{(535)^2}{40} \right] = 20,11 \Rightarrow$  desvio padrão  $\Rightarrow S = 4,48$  gols.

3º)  $\bar{x} \pm S = 13,38 \pm 4,48 = (8,9 ; 17,86) \Rightarrow$  entre 9 e 18 gols há 29 jogadores  $\Rightarrow 29/40 = 72,5\%$

$\bar{x} \pm 2S = 13,38 \pm 8,96 = (4,42 ; 22,34) \Rightarrow$  entre 4 e 22 gols há 37 jogadores  $\Rightarrow 37/40 = 92,5\%$

$\bar{x} \pm 3S = 13,38 \pm 13,44 = (-0,06 ; 26,82) \Rightarrow$  entre 0 e 26 gols há 40 jogadores  $\Rightarrow 100\%$

Os resultados confirmam o critérios empírico e de Tchebycheff.

**Exercício Proposto:**

Calcular a variância e o desvio padrão de cada série de dados a seguir. Interpretar depois o desvio padrão da distribuição pelas duas regras anteriores.

a) Notas de Estatística de 35 alunos:

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 6  | 6  | 7  | 8  | 8  | 8  |
| 9  | 10 | 10 | 11 | 12 | 13 |
| 14 | 14 | 15 | 16 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 29 | 30 | 31 | 32 |    |

b) Massas, em kg, de 180 pacientes de uma clínica:

| MASSAS         | Nº DE PACIENTES | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ |
|----------------|-----------------|-----------------|-------------------|
| 30             | 40              |                 |                   |
| 40             | 32              |                 |                   |
| 50             | 35              |                 |                   |
| 60             | 24              |                 |                   |
| 70             | 20              |                 |                   |
| 80             | 29              |                 |                   |
| <b>TOTAL →</b> | <b>180</b>      |                 |                   |

c) Salários de 22 funcionários de uma empresa:

| i              | SALÁRIOS  | $f_i$           | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ |
|----------------|-----------|-----------------|-----------------|-------------------|
| 1              | 300 — 320 | 4               |                 |                   |
| 2              | 320 — 340 | 3               |                 |                   |
| 3              | 340 — 360 | 5               |                 |                   |
| 4              | 360 — 380 | 8               |                 |                   |
| 5              | 380 — 400 | 2               |                 |                   |
| <b>TOTAL →</b> |           | $\sum f_i = 22$ |                 |                   |

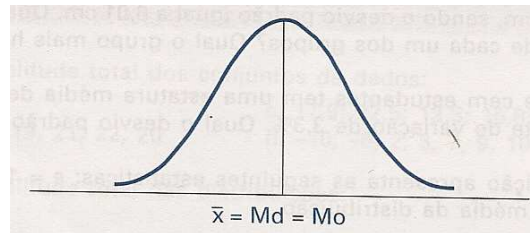
d) Número de jogos realizados por 45 jogadores profissionais:

| i              | JOGOS     | $f_i$           | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ |
|----------------|-----------|-----------------|-----------------|-------------------|
| 1              | 125 — 175 | 8               |                 |                   |
| 2              | 175 — 225 | 6               |                 |                   |
| 3              | 225 — 275 | 9               |                 |                   |
| 4              | 275 — 325 | 7               |                 |                   |
| 5              | 325 — 375 | 9               |                 |                   |
| 6              | 375 — 425 | 6               |                 |                   |
| <b>TOTAL →</b> |           | $\sum f_i = 45$ |                 |                   |

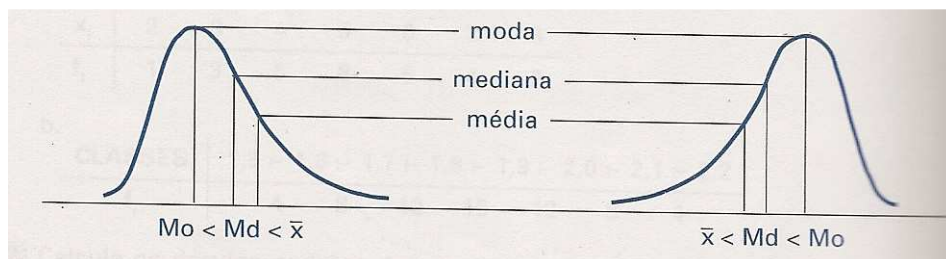
e) Contribuições de 121 empregados ao INSS, assim distribuídos por classe:  $[54, 74[ \rightarrow 12$ ,  $[74, 94[ \rightarrow 11$ ,  $[94, 114[ \rightarrow 11$ ,  $[114, 134[ \rightarrow 13$ ,  $[134, 154[ \rightarrow 16$ ,  $[154, 174[ \rightarrow 19$ ,  $[174, 194[ \rightarrow 21$  e  $[194, 214[ \rightarrow 18$ .

### 2.3) Sobre as Medidas de Assimetria e Curtose :

**2.3.1) Assimetria :** Uma distribuição é simétrica quando coincidem a **média, a moda** e a **mediana**. Então, a assimetria acontece quando há divergência entre essas medidas. Considerando a média  $\bar{x}$ , a moda  $Mo$  e a mediana  $Md$ , tem-se:



**Gráfico de distribuição simétrica**



**Gráfico de distribuição Assimétrica positiva (à direita)**

**Gráfico de distribuição Assimétrica negativa (à esquerda)**

Com base nessas relações entre a média e a moda, pode-se, com a diferença  $\Delta M = \bar{x} - Mo$ , determinar o tipo de assimetria:

- ▶ Se  $\Delta M = 0 \Rightarrow$  D. Simétrica
- ▶ Se  $\Delta M < 0 \Rightarrow$  Assimetria negativa
- ▶ Se  $\Delta M > 0 \Rightarrow$  Assimetria positiva

**Exemplo Ilustrativo 7:** Sejam as distribuições  $D_1$ ,  $D_2$  e  $D_3$ , dadas a seguir:

**DISTRIBUIÇÃO  $D_1$**

| IDADES (anos)        | $f_i$ |
|----------------------|-------|
| [2, 6[               | 6     |
| [6, 10[              | 12    |
| [10, 14[             | 24    |
| [14, 18[             | 12    |
| [18, 22[             | 6     |
| $\Sigma \rightarrow$ | 60    |

$\rightarrow \bar{x} = 12$  anos  
 $\rightarrow MD = 12$  anos  
 $\rightarrow \text{Moda} = 12$  anos  
 $\rightarrow \Delta M = 12 - 12 = 0 \Rightarrow$  SIMÉTRICA

**DISTRIBUIÇÃO D<sub>2</sub>**

| IDADES (anos)        | f <sub>i</sub> |
|----------------------|----------------|
| [2 , 6[              | 6              |
| [6 , 10[             | 12             |
| [10 , 14[            | 24             |
| [14 , 18[            | 30             |
| [18 , 22[            | 6              |
| $\Sigma \rightarrow$ | 78             |

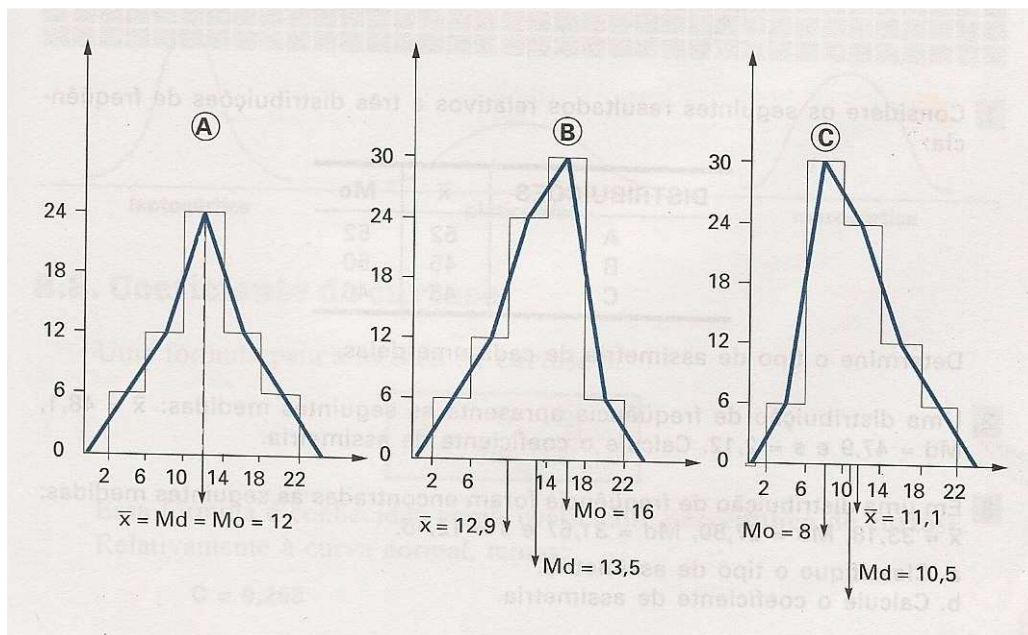
$\rightarrow \bar{x} = 12,9$  anos  
 $\rightarrow MD = 13,5$  anos  
 $\rightarrow \text{Moda} = 16$  anos  
 $\rightarrow \Delta M = 12,9 - 16 = -3,1$  anos  $\Rightarrow$  ASSIMÉTRICA NEGATIVA

**DISTRIBUIÇÃO D<sub>3</sub>**

| IDADES (anos)        | f <sub>i</sub> |
|----------------------|----------------|
| [2 , 6[              | 6              |
| [6 , 10[             | 30             |
| [10 , 14[            | 24             |
| [14 , 18[            | 12             |
| [18 , 22[            | 6              |
| $\Sigma \rightarrow$ | 78             |

$\rightarrow \bar{x} = 11,1$  anos  
 $\rightarrow MD = 10,5$  anos  
 $\rightarrow \text{Moda} = 8$  anos  
 $\rightarrow \Delta M = 11,1 - 8 = 3,1$  anos  $\Rightarrow$  ASSIMÉTRICA POSITIVA

Vejamos os gráficos dessas distribuições :

**2.3.1.1) coeficiente de Assimetria de Pearson:**

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

Se  $0,15 < |As| < 1 \Rightarrow$  ASSIMETRIA MODERADA.

Se  $|As| > 1 \Rightarrow$  ASSIMETRIA FORTE.

**Exemplo Ilustrativo 8:** Para as distribuições anteriores, tem-se:  $S(D_1) = 4,42$ ,  $S(D_2) = 4,20$  e  $S(D_1) = 4,20$ . Então, os coeficientes de assimetria serão:

$$\blacktriangleright D_1 \blacktriangleright A_s = \frac{3(12 - 12)}{4,42} = 0 \Rightarrow \text{SIMÉTRIA}$$

$$\blacktriangleright D_2 \blacktriangleright A_s = \frac{3(12,9 - 13,5)}{4,20} = -0,420 \Rightarrow \text{ASSIMÉTRIA NEGATIVA}$$

$$\blacktriangleright D_3 \blacktriangleright A_s = \frac{3(11,1 - 10,5)}{4,20} = 0,429 \Rightarrow \text{ASSIMÉTRIA POSITIVA}$$

**Exercício proposto:** Avalie a assimetria de cada distribuição a seguir, usando os dois métodos abordados.

a) Notas de Estatística de 35 alunos:

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 6  | 6  | 7  | 8  | 8  | 8  |
| 9  | 10 | 10 | 11 | 12 | 13 |
| 14 | 14 | 15 | 16 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 29 | 30 | 31 | 32 |    |

b) Massas, em kg, de 180 pacientes de uma clínica:

| MASSAS         | Nº DE PACIENTES | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ |
|----------------|-----------------|-----------------|-------------------|
| 30             | 40              |                 |                   |
| 40             | 32              |                 |                   |
| 50             | 35              |                 |                   |
| 60             | 24              |                 |                   |
| 70             | 20              |                 |                   |
| 80             | 29              |                 |                   |
| <b>TOTAL →</b> | <b>180</b>      |                 |                   |

c) Salários de 22 funcionários de uma empresa:

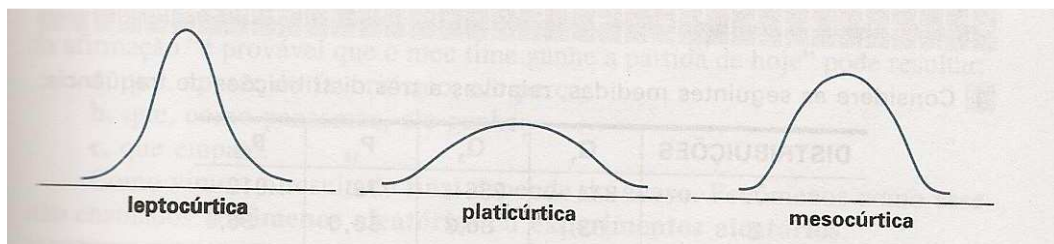
| i              | SALÁRIOS  | $f_i$           | $x_i$ | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ | $f_{bi}$ |
|----------------|-----------|-----------------|-------|-----------------|-------------------|----------|
| 1              | 300 — 320 | 4               |       |                 |                   |          |
| 2              | 320 — 340 | 3               |       |                 |                   |          |
| 3              | 340 — 360 | 5               |       |                 |                   |          |
| 4              | 360 — 380 | 8               |       |                 |                   |          |
| 5              | 380 — 400 | 2               |       |                 |                   |          |
| <b>TOTAL →</b> |           | $\sum f_i = 22$ |       |                 |                   |          |

d) Número de jogos realizados por 45 jogadores profissionais:

| i       | JOGOS     | $f_i$           | $x_i$ | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ | $f_{bi}$ |
|---------|-----------|-----------------|-------|-----------------|-------------------|----------|
| 1       | 125 — 175 | 8               |       |                 |                   |          |
| 2       | 175 — 225 | 6               |       |                 |                   |          |
| 3       | 225 — 275 | 9               |       |                 |                   |          |
| 4       | 275 — 325 | 7               |       |                 |                   |          |
| 5       | 325 — 375 | 9               |       |                 |                   |          |
| 6       | 375 — 425 | 6               |       |                 |                   |          |
| TOTAL → |           | $\sum f_i = 45$ |       |                 |                   |          |

e) Contribuições de 121 empregados ao INSS, assim distribuídos por classe: [54 , 74[ → 12, [74 , 94[ → 11, [94 , 114[ → 11, [114 , 134[ → 13 , [134 , 154[ → 16 , [154 , 174[ → 19 , [174 , 194[ → 21 e [194 , 214[ → 18.

**2.3.2) Curtose :** É o grau de achatamento da curva de uma distribuição em relação a uma distribuição padrão, chamada de **curva normal** ou **mesocúrtica**. A curva mais estreita que a normal recebe o nome de **leptocúrtica** e a mais aberta que a normal é a **platicúrtica**.



**2.3.2.1) Coeficiente de Curtose :**

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

Em que C é o **coeficiente percentílico de curtose** ( para a curva normal,  $C = 0,263$ );  $Q_1$  e  $Q_3$  são quartis,  $P_{10}$  e  $P_{90}$  são percentis. Ainda temos que:

→ Para a curva leptocúrtica,  $C < 0,263$  e

→ Para a curva platicúrtica,  $C > 0,263$ .

**Exemplo Ilustrativo 9:** Para a distribuição abaixo, calcule o grau de curtose e classifique a curva em relação à normal. Determine também o tipo de assimetria dessa curva.

| IDADES (anos)        | $f_i$ | $x_i$ | $F_{bi}$ | $f_i \cdot x_i$ | $x_i^2 \cdot f_i$ |
|----------------------|-------|-------|----------|-----------------|-------------------|
| [7 , 17[             | 6     | 12    | 6        | 72              | 864               |
| [17 , 27[            | 15    | 22    | 21       | 330             | 7.260             |
| [27 , 37[            | 20    | 32    | 41       | 640             | 20.480            |
| [37 , 47[            | 10    | 42    | 51       | 420             | 17.640            |
| [47 , 57[            | 5     | 52    | 56       | 260             | 13.520            |
| $\Sigma \rightarrow$ | 56    |       |          | 1.722           | 59.764            |

$$\rightarrow \bar{x} = \frac{1.722}{56} = 30,75 \text{ anos.}$$

$$\rightarrow Mo = 27 + \frac{5 \cdot 10}{5+10} = 30,33 \text{ anos.}$$

$$\rightarrow Md = 27 + \frac{(28-21) \cdot 10}{20} = 30,5 \text{ anos.}$$

$$\rightarrow S^2 = \frac{1}{55} \left[ 59.764 - \frac{(1.722)^2}{56} \right] = 123,86 \Rightarrow S = \sqrt{123,86} = 11,13 \text{ anos.}$$

$$\rightarrow \text{Quartil 1 : } \left\{ \frac{n}{4} = 14, \lambda_1 = 17, F_{i.\text{ant}} = 6 \text{ e } f_1 = 15 \Rightarrow Q_1 = 17 + \frac{(14-6) \cdot 10}{15} = 22,33 \text{ anos} \right.$$

$$\rightarrow \text{Quartil 3 : } \left\{ \frac{3n}{4} = 42, \lambda_3 = 37, F_{i.\text{ant}} = 41 \text{ e } f_3 = 10 \Rightarrow Q_3 = 37 + \frac{(42-41) \cdot 10}{10} = 38 \text{ anos.} \right.$$

$$\rightarrow \text{Percentil 10 : } \left\{ \frac{in}{100} = 5,6, \lambda_{10} = 7, F_{i.\text{ant}} = 0 \text{ e } f_{10} = 6 \Rightarrow P_{10} = 7 + \frac{(5,6-0) \cdot 10}{6} = 16,33 \text{ anos} \right.$$

$$\rightarrow \text{Percentil 90 : } \left\{ \frac{in}{100} = 50,4, \lambda_{10} = 37, F_{i.\text{ant}} = 41 \text{ e } f_{10} = 10 \Rightarrow P_{90} = 37 + \frac{(50,4-41) \cdot 10}{10} = 46,4 \text{ anos} \right.$$

$$\rightarrow As = \frac{3 \cdot (30,75 - 30,5)}{11,13} = 0,0673 \quad \text{e} \quad \Delta M = 30,75 - 30,33 = 0,42 \Rightarrow \text{Leve assimetria à direita, quase simétrica.}$$

$$\rightarrow C = \frac{(38 - 22,33)}{2(46,4 - 16,33)} = \frac{15,67}{60,14} = 0,261 \Rightarrow \text{Curva leptocúrtica, quase normal.}$$

#### **Exemplo de Reciclagem do módulo 4:**

O número  $\frac{1+\sqrt{5}}{2}$ , aproximadamente igual a 1,618, é conhecido como *Número de ouro*, pois, segundo a concepção do classicismo grego, tem sua origem na divisão de um todo em duas partes, tais que a menor caiba na maior o mesmo número de vezes que esta caiba no todo. Constatou-se a presença deste número na morfologia dos seres vivos. Um exemplo disso é a razão entre o comprimento total das duas primeiras falanges de um dedo humano e o comprimento da terceira falange do mesmo dedo. Um professor resolveu usar a sua turma de 42 alunos como amostra. Para fazer a coleta de dados, ele mediu os dedos de todos da turma e calculou a razão entre as falanges. Os resultados das proporções constatadas estão ordenadas na tabela abaixo:

|      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|
| 1,25 | 1,27 | 1,28 | 1,28 | 1,29 | 1,30 | 1,32 |
| 1,33 | 1,33 | 1,35 | 1,36 | 1,36 | 1,37 | 1,38 |
| 1,38 | 1,38 | 1,39 | 1,40 | 1,42 | 1,44 | 1,44 |
| 1,47 | 1,52 | 1,54 | 1,57 | 1,62 | 1,63 | 1,63 |
| 1,63 | 1,64 | 1,64 | 1,64 | 1,65 | 1,65 | 1,66 |
| 1,66 | 1,67 | 1,68 | 1,68 | 1,69 | 1,71 | 1,72 |

Construir uma distribuição com classes que possibilite determinar a média dessas variáveis, sua moda, sua mediana, sua variância, seu desvio padrão, seu coeficiente de variação, sua medida e tipo de assimetria, sua medida de curtose e o tipo de curva gerada pela distribuição.

Resolução:

A tabulação desses dados, em forma de distribuição com classes, ficou assim:

| i               | RAZÃO         | f <sub>i</sub> | x <sub>M</sub> | f <sub>i</sub> ·x <sub>M</sub> | (x <sub>M</sub> - $\bar{x}$ ) <sup>2</sup> | f <sub>i</sub> · (x <sub>M</sub> - $\bar{x}$ ) <sup>2</sup> | F <sub>b</sub> |
|-----------------|---------------|----------------|----------------|--------------------------------|--------------------------------------------|-------------------------------------------------------------|----------------|
| 1               | [1,25 ; 1,33[ | 7              | 1,29           | 9,03                           | 0,04                                       | 0,28                                                        | 7              |
| 2               | [1,33 ; 1,41[ | 11             | 1,37           | 15,07                          | 0,0144                                     | 0,1584                                                      | 18             |
| 3               | [1,41 ; 1,49[ | 4              | 1,45           | 5,80                           | 0,0016                                     | 0,0064                                                      | 22             |
| 4               | [1,49 ; 1,57[ | 2              | 1,53           | 3,06                           | 0,0016                                     | 0,0032                                                      | 24             |
| 5               | [1,57 ; 1,65[ | 8              | 1,61           | 12,88                          | 0,0144                                     | 0,1152                                                      | 32             |
| 6               | [1,65 ; 1,73[ | 10             | 1,69           | 16,90                          | 0,04                                       | 0,40                                                        | 42             |
| <b>TOTAIS →</b> |               | <b>42</b>      |                | <b>62,74</b>                   |                                            | <b>0,9632</b>                                               |                |

Então, teremos:

$$1^{\circ}) \bar{x} = \frac{\sum f_i \cdot x_M}{\sum f_i} = \frac{62,74}{42} = 1,49.$$

$$2^{\circ}) s^2 = \frac{0,9632}{41} = 0,023 \Rightarrow s = 0,15.$$

3<sup>o</sup>) O intervalo  $\bar{x} \pm s$  é  $1,34 \leq x_i \leq 1,64$  e contém 23 dos 42 dados, ou seja, 54,76% da amostra;  
O intervalo  $\bar{x} \pm 2s$  é  $1,19 \leq x_i \leq 1,79$  e contém os 42 dados, ou seja, 100% da amostra.

$$4^{\circ}) \text{A moda da distribuição é } Mo = 1,33 + \frac{4 \cdot (0,08)}{4+7} = 1,36.$$

$$5^{\circ}) \text{A mediana da distribuição será } Md = 1,41 + \frac{(21-18) \cdot 0,08}{4} = 1,47.$$

$$5^{\circ}) \text{O coeficiente de variação será } C_V = \frac{0,15}{1,49} \times 100\% = 10,1\%.$$

7<sup>o</sup>) Quanto à assimetria do histograma, temos

→ Pela diferença  $\Delta M = \bar{x} - Mo = 1,49 - 1,36 = 0,13 \Rightarrow$  Assimetria à direita.

→ Pelo coeficiente de assimetria de Pearson:  $As = \frac{3(\bar{x} - Md)}{S} = \frac{3(1,49 - 1,47)}{0,15} = 0,4 \Rightarrow$  Assimetria à direita (leve assimetria).

8<sup>o</sup>) Quanto à medida de curtose, teremos:

$$a) Q_1 = 1,33 + \frac{(10,5-7) \cdot 0,08}{11} = 1,35 \quad e \quad Q_3 = 1,57 + \frac{(31,5-24) \cdot 0,08}{8} = 1,65$$

$$b) P_{10} = 1,25 + \frac{(4,2-0) \cdot 0,08}{7} = 1,34 \quad e \quad P_{90} = 1,65 + \frac{(37,8-32) \cdot 0,08}{10} = 1,70$$

$$c) \text{Pelo coeficiente de curtose } C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{1,65 - 1,35}{2(1,70 - 1,34)} = 0,42 \Rightarrow \text{curva platicúrtica.}$$

**Exercício proposto:** Para cada distribuição do Exercício proposto da página 9, verifique a assimetria pelos dois processos estudados e classifique a curva quanto à curtose.