UNIVERSIDADE FEDERAL DE ALAGOAS
Instituto de Computação

MEDIDAS DESCRITIVAS

Parte II - UNIDADE I

PROFESSOR: PETRUCIO ANTONIO MEDEIROS BARROS
TURMA: CIÊNCIA DA COMPUTAÇÃO / ENG. DA COMPUTAÇÃO

MEDIDAS DESCRITIVAS

- Em qualquer área do conhecimento, as informações costumam não ser constantes.
- Assim, é necessário que variável avaliada esteja associada a uma medida que faça referência à variabilidade que reflita a flutuação dos dados.
  - Principais características analisadas:
    - A tendência central dos dados;
    - A dispersão ou variação em relação a esse centro;
    - Os dados que ocupam certas posições;
    - Simetria dos dados;
    - A forma na qual os dados se agrupam.

MEDIDAS DESCRITIVAS

- Mostram um valor representativo em torno do qual os dados tendem a agrupar-se com maior ou menor frequência.
- São utilizadas para sintetizar em um único número o conjunto de dados observados.
- As principais medidas descritivas ou de tendência central, são:
  - a) **Média;**
  - b) **Mediana;**
  - c) **Moda.**
- As principais medidas de dispersão ou variabilidade, são:
  - a) **Variância;**
  - b) **Desvio padrão;**
  - c) **Coeficiente de variação.**

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
- Média Aritmética Simples

Para uma sequência numérica X: $x_1, x_2, \ldots, x_n$ a média aritmética simples, que designaremos por $\overline{x}$ onde:

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n}.$$

- Propriedades da média:
  - A soma algébrica dos desvios tomados em relação a media (a diferença entre cada elemento de um conjunto de valores e a média aritmética) é nula;
  - Somando-se ou subtraindo-se uma constante (c) de todos os valores uma variável, a média do conjunto fica aumenta ou diminuída dessa constante;
  - Multiplicando-se ou dividindo-se todos os valores uma variável, por uma constante (c), a média do conjunto fica multiplicada ou dividida por essa constante.

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
- Média Aritmética Ponderada

Para uma sequência numérica $X : x_1, x_2, \ldots, x_n$, em que cada valor possui um peso $p_1, p_2, \ldots, p_n$ respectivamente, a média aritmética ponderada, será calculada por:

$$\overline{x} = \frac{x_1 p_1 + \ldots + x_n p_n}{p_1 + \ldots + p_n}.$$

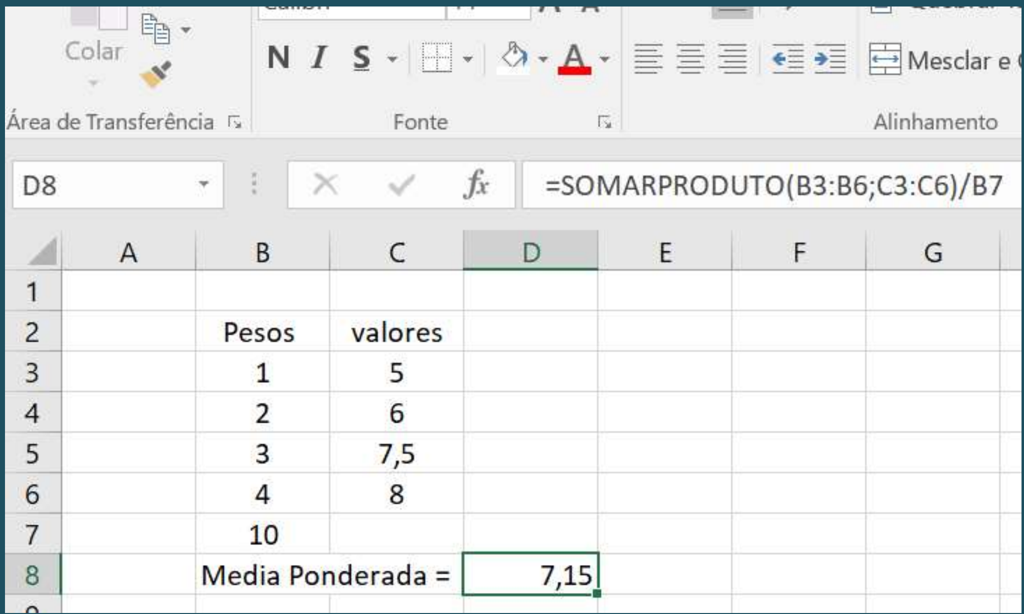
- **Exemplo:** O exame de seleção pode ser composto de 3 provas onde as duas primeiras tem peso 1 e a terceira tem peso 2. Um candidato com notas 70, 75 e 90 terá média final:

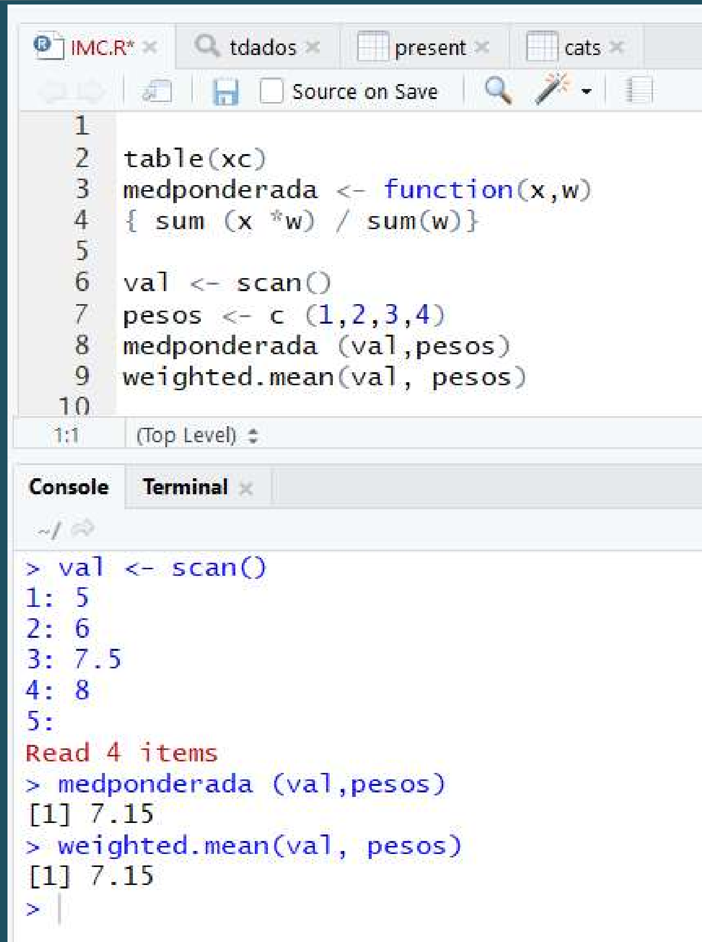
$$\overline{X} = \frac{1(70) + 1(75) + 2(90)}{4} = 81{,}25$$

MEDIDAS DESCRITIVAS

➢ **Medidas Descritivas ou de Tendência Central:**

- Média Aritmética Ponderada





MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**

- Média aritmética para dados tabelados

Ao calcular a média aritmética para dados tabelados, deve-se levar em conta que o número de elementos pertencentes a classe pode diferir, o que pode ser compreendido como um peso. Nesse caso, da expressão que fornece a média aritmética ponderada de um conjunto de dados brutos, tem-se

$$\overline{x} = \frac{x_1 p_1 + \ldots + x_n p_n}{p_1 + \ldots + p_n},$$

e adaptando-a para o caso de dados tabelados, surge a expressão a seguir,

$$\overline{x} = \frac{x_1 f_1 + \ldots + x_n f_n}{f_1 + \ldots + f_n},$$

sendo x_i a média dos extremos do I_c da i-ésima linha e f_i a frequência absoluta da i-ésima linha.

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**

  - Média aritmética para dados agrupados ou tabelados

Classes			Frequências (fa)	Freq. Relativa (fr)	Xi	Xi (fa)
69,2	|--	94,8	3	7,50%	82,0	246,0
94,8	|--	120,4	8	20,00%	107,6	860,8
120,4	|--	146	16	40,00%	133,2	2.131,2
146	|--	171,6	7	17,50%	158,8	1.111,6
171,6	|--	197,2	4	10,00%	184,4	737,6
197,2	|--	222,8	2	5,00%	210,0	420,0
Total			40	100,00%		5.507,2
			Média Ponderada = 5.507,2 / 40 = 137,68			

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
- **Média Geométrica**
  - ✓ **É utilizada principalmente para calcular médias de razões, de taxas de variação e de índices econômicos.**

Dados *n* valores x¹, x², ..., xn, a média geométrica desses valores será:

$$\bar{x}_g = \sqrt[n]{x1 * x2 * x \ldots xn}$$

$$\bar{x}_g = \sqrt[3]{x1 * x2 * x3} = \sqrt[3]{10 * 60 * 360} = \sqrt[3]{216000} = 60$$

- ✓ **Média Geométrica Ponderada**

$$\bar{x}_g = \sqrt[\sum_{j=1}^{k} fj]{x1^{f1} \times x2^{f2} \times \ldots \times xk^{fk}}$$

Exemplo: Calcule a média geométrica dos valores constantes da seguinte tabela:

xf	fj
1	2
3	4
9	2
27	1
	$\sum_{j=1}^{k} fj = 9$

$$\bar{x}_g = \sqrt[9]{1^2 \times 3^4 \times 9^2 \times 27^1}$$

$$\bar{x}_g = \sqrt[9]{1 \times 81 \times 81 \times 27} = \sqrt[9]{177147}$$

$$\bar{x}_g = 3{,}829554$$

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
- **Média Harmônica**

Sejam x_1, x_2, x_3,……x_n, valores de x, associados às freqüências absolutas n_1, n_2, n_3,……n_n, respectivamente.

A média harmônica de X é definida por:

$$Mh = \frac{n}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \ldots\ldots + \frac{n_n}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{n_i}{x_i}}$$

- Para dados não agrupados n = 1.
- Para dados agrupados sem intervalo de classe x_i é o valor da variável.
- Para dados agrupados com intervalo de classe x_i é o ponto médio da classe.

- Normalmente, ele é adequado para situações em que a **média das taxas** é desejada

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
- **Média Harmônica**

  - **Exemplo:**

Um investidor compra $ 18.000 de ações de uma empresa a $ 45 a ação, e em seguida compra $18.000 a $ 36 a ação. Qual o preço médio por ação, pago pelo investidor?

$$\overline{x_h} = \frac{2}{\frac{1}{45} + \frac{1}{36}} = 2 \,.\, \frac{180}{9} = 40$$

MEDIDAS DESCRITIVAS

- **Relação entre as Médias:**

- A média geométrica e a média harmônica são menores, ou no máximo igual, à aritmética.
- A igualdade só ocorre no caso em que todos os valores da amostra são idênticos.
- Quanto maior a variabilidade, maior será a diferença entre as médias harmônica e geométrica e a média aritmética.

$$H \leq G \leq \bar{X}$$

- Exemplo: Para a amostra 12 , 14 e 16 temos:

$$\boldsymbol{H = 13,81 < G = 13,90} < \bar{X} = 14{,}00$$

MEDIDAS DESCRITIVAS

➢ **Relação entre as Médias: Exemplo:**

Excel

Medias - Exemplo 1		
7,5		
8,0	Media Aritmética	7,000000
7,0	Media Geométrica	6,696892
3,0	Media Harmônica	6,311044
5,0		
10,0		
8,0		
7,5		
6,0		
8,0		

Source on Save

```r
media3<-function(x){
  n<-length(x)
  ma<-mean (x)
  mg<-(prod(x))^(1/n)
  mh<-(1/n*sum(1/x))^(-1)

  cat("média aritm=",ma," média geom=",mg," média harm=",mh)
}
x <- scan ()

media3 (x)
```

31:1 (Top Level)

Console Terminal

```
> x <- scan ()
1: 7.5
2: 8
3: 7
4: 3
5: 5
6: 10
7: 8
8: 7.5
9: 6
10: 8
11:
Read 10 items
> media3 (x)
média aritm= 7  média geom= 6.696892  média harm= 6.311044
```

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
  - Mediana "Md"
  - A mediana de um conjunto de valores, dispostos segundo uma ordem (crescente ou decrescente) é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.
    - **Se a série dada tiver número ímpar de termos -** O valor mediano será o termo de ordem dado pela fórmula:

$$Md = \frac{n+1}{2}$$

Na amostra: 32 36 42 42 58 Md = 42

    - **Se a série dada tiver número par de termos -** O valor mediano será o termo de ordem dado pela fórmula:

$$Md = \frac{\left[\left(\frac{n}{2}\right) + \left(\frac{n}{2} + 1\right)\right]}{2}$$

Na amostra: 30 40 42 45 50 80 Md= $\frac{42+45}{2}$ = 43,5

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
  - Mediana "Md" para dados agrupados ou tabelados
    - Calculada através da Fórmula:

$$M_d = l_i + \left[\frac{\left(\frac{n}{2}\right) - f_{abs\ ant\ ac}}{f_{abs}}\right] . c$$

Sendo:

- l_i : limite inferior da classe Mediana;
- n : número total de elementos da série;
- $f_{abs\ ant\ ac}$: frequência absoluta anterior acumulada a classe Mediana;
- f_{abs} : frequência absoluta da classe Mediana;
- c : amplitude da classe Mediana.

MEDIDAS DESCRITIVAS

Medidas Descritivas ou de Tendência Central:

- Mediana "Md" para dados agrupados ou tabelados

Classes			Frequências (fa)	Freq. Relativa (fr)	Freq. Acumulada (Fa)	Freq. Acumulada Relativa (Fr)	
69,2	|--	94,8	3	7,50%	3	7,50%	
94,8	|--	120,4	8	20,00%	11	27,50%	
120,4	|--	146	16	40,00%	27	67,50%	Classe Md
146	|--	171,6	7	17,50%	34	85,00%	
171,6	|--	197,2	4	10,00%	38	95,00%	
197,2	|--	222,8	2	5,00%	40	100,00%	
Total			40	100,00%			
			Md=	**134,8**	**120,4 + (25,6/16) * 9**		

$$M_d = l_i + \left[\frac{\left(\frac{n}{2}\right) - f_{abs\ ant\ ac}}{f_{abs}}\right] . c$$

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
  - Moda "Mo"
    - É o valor que ocorre com mais frequência, em uma série de valores.
    - Numa amostra, a Moda pode não existir ou ser múltipla.
      - Exemplos:
        - Na amostra 21 24 27 27 28 28 31 31 31 Mo = 31 → Unimodal
        - Na amostra 45 46 49 52 52 60 60 76 79 Mo = 52 e 60 → Bimodal
        - Na amostra 3 5 8 10 12 13 → Amodal

MEDIDAS DESCRITIVAS

Medidas Descritivas ou de Tendência Central:

- Moda "Mo" para dados agrupados ou tabelados
  - A moda é o ponto médio da classe modal (classe que apresenta maior frequência).

Classes			Frequências (fa)	Freq. Relativa (fr)	Freq. Acumulada (Fa)	Freq. Acumulada Relativa (Fr)
69,2	|--	94,8	3	7,50%	3	7,50%
94,8	|--	120,4	8	20,00%	11	27,50%
120,4	|--	146,0	16	40,00%	27	67,50%
146,0	|--	171,6	7	17,50%	34	85,00%
171,6	|--	197,2	4	10,00%	38	95,00%
197,2	|--	222,8	2	5,00%	40	100,00%
Total			40	100,00%		
			Mo =	**133,2**	**(120,4 + 146) /2**	

Método de Czuber

$$Mo = l + \left(\frac{d1}{d1 + d2}\right) . h$$

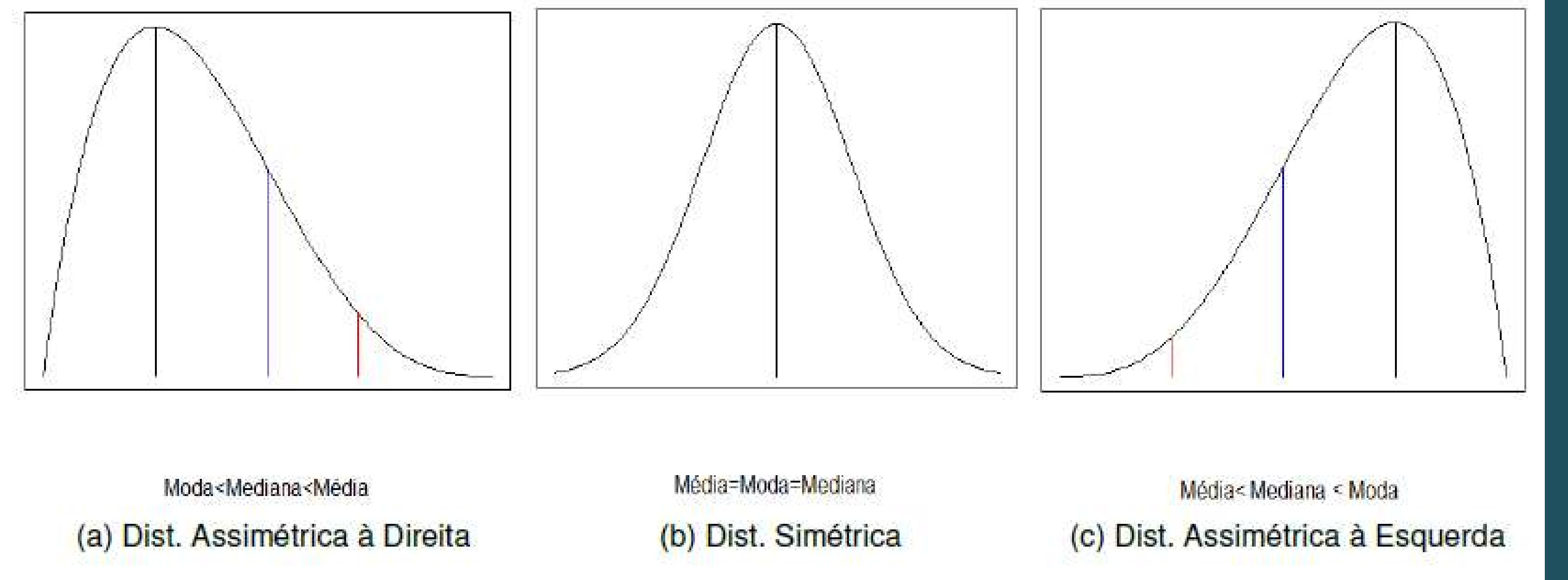
Onde:

- l → limite inferior da classe de maior frequência.
- d1 → subtração da maior frequência pela frequência da classe anterior.
- d2 → subtração da maior frequência pela frequência da classe posterior.
- h → amplitude da classe de maior frequência.

Moda Czuber = 120,4 + $\frac{8}{(8+9)}$ 25,6 = 132,44

MEDIDAS DESCRITIVAS

- **Medidas Descritivas ou de Tendência Central:**
  - Relação entre média, moda e mediana



Relação entre média, moda e mediana

MEDIDAS DESCRITIVAS

➢ Medidas de Dispersão ou Variabilidade

- As medidas de dispersão medem a variabilidade dos dados em estudo. Permitem verificar se o conjunto de dados é homogêneo ou heterogêneo. São medidas estatísticas que medem a dispersão dos dados, em torno de um valor central.
- Considere o valor (em reais) ganho de três grupos de empregados:
  - A: 70, 70, 70, 70, 70
  - B: 50, 60, 70, 80, 90
  - C: 5, 15, 50, 120, 160
- Podemos verificar que, apesar de apresentarem a mesma média (70), os três grupos apresentam comportamento diferenciado, pois o grupo A é o mais homogêneo, e o grupo C é o que apresenta maior variação de ganho.
- Portanto, se faz necessário uma medida de posição que avalie esta distribuição, ou seja, a variabilidade de um conjunto de dados.
- Quanto maior a variabilidade, maior será a dispersão das observações.

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - As medidas utilizadas para representar dispersão são:
    - **Amplitude;**
    - **Variância;**
    - **Desvio Padrão;**
    - **Coeficiente de Variação.**

  - **Amplitude**

    - Amplitude total ou máxima é a diferença entre o maior e o menor valor de um conjunto de dados.

A = (valor maior) - (valor menor)

- Depende apenas dos valores maior e menor, não é tão útil quanto as outras medidas de variação que usam todos os valores.
- Para dados agrupados ou tabelados: A = (limite inferior da classe inferior – limite superior da classe superior)

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade

  - **Variância**

    - É uma medida de dispersão que mede a variabilidade de um conjunto de dados. Quando desejamos obter a variância de uma população, o que geralmente não é possível pelo fato de desconhecermos toda a população, representamos variância por δ^2. Quando desejamos obter a variância amostral, representamos por s^2, sendo s^2 uma estimativa de δ^2.
    - A variância é definida como a média das diferenças quadráticas de n valores em relação à sua média aritmética.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Fórmula alternativa:

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

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - **Variância**
    - Propriedades da Variância
      - Para qualquer distribuição a variância é sempre uma quantidade positiva .
      - Se os valores das observações são todos iguais então a variância é zero.
      - Variância de uma constante é zero.
      - Se somarmos ou subtrairmos uma mesma constante de cada elemento de um conjunto de dados sua variância não se altera.
      - Se multiplicarmos ou dividirmos a cada elemento de um conjunto de dados por uma mesma constante sua variância fica multiplicada ou dividida pelo quadrado da constante.

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - Variância
    - Variância para dados tabelados
    - Para dados tabelados, considerar que os dados estão divididos em intervalos. Sendo assim, cada um contêm um número diferente de elementos. Portanto devemos considerar a frequência absoluta de cada um. A variância é obtida através da expressão:

$$s^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{\sum_{i=1}^{k} f_i - 1}$$

Fórmula alternativa:

- $$S^2 = \frac{\sum x_i^2 . f_i - \frac{(\sum x_i . f_i)^2}{n}}{n-1}$$

    - Sendo k o número de intervalos de classe, x_i a média de cada intervalo de classe e $\overline{x}$ a média do conjunto.

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - Desvio padrão
    - A variância não tem a mesma magnitude que as observações, pois está elevada ao quadrado. Se quisermos que a medida de dispersão seja da mesma dimensão que as das observações, basta extrair a raiz quadrada.

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

- Exemplo: para a amostra 10 12 14 16 18
- A média é 14 e o desvio-padrão é calculado:

$$S = \sqrt{\frac{(10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2}{n-1}} = 3{,}16$$

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - Coeficiente de Variação
    - O Coeficiente de Variação caracteriza a dispersão ou variabilidade dos dados em termos relativos ao valor médio.
      - Dado pela fórmula: $\boldsymbol{CV = \frac{desvio\ padrão}{média} . 100}$
    - Quanto maior o coeficiente de variação, maior é a variação entre os dados do grupo avaliado. A variabilidade de um conjunto de dados depende da área de pesquisa. Contudo, alguns autores apresentam que:
      - CV ≤ 20% a amostra é homogênea;
      - CV > 20% a amostra é heterogênea.

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - Coeficiente de Variação
    - Exemplo: Considere as valores de ocupação de dois hotéis A e B.
    - Comente as diferenças entre os hotéis?

Meses	Unidades habitacionais (média mensal)	
	Hotel A	Hotel B
Janeiro	760	420
Fevereiro	690	450
Março	380	510
Abril	280	460
Maio	320	470
Junho	300	440
Julho	710	480
Agosto	270	430
Setembro	360	410
Média	**452,2222**	**452,2222**
Desvio	**204,6202**	**31,5348**
Coeficiente de Variacão	**45,25%**	**6,97%**

MEDIDAS DESCRITIVAS

- Medidas de Dispersão ou Variabilidade
  - Coeficiente de Variação
    - Calculo das variâncias e dos desvios para os hotéis A e B.

Meses	Unidades habitacionais (média mensal)	
	Hotel A	Hotel B
Janeiro	760	420
Fevereiro	690	450
Março	380	510
Abril	280	460
Maio	320	470
Junho	300	440
Julho	710	480
Agosto	270	430
Setembro	360	410
Média	452,2222	452,2222
Desvio	204,6202	31,5348
Coeficiente de Variacão	45,25%	6,97%

Meses	Hotel A	(Hotel A) ^2	Hotel B	(Hotel B) ^2			
Janeiro	760	577600	420	176400			
Fevereiro	690	476100	450	202500			
Março	380	144400	510	260100		Variância	Desvio padrão
Abril	280	78400	460	211600			
Maio	320	102400	470	220900	Hotel A	41869,44	204,62
Junho	300	90000	440	193600			
Julho	710	504100	480	230400	Hotel B	994,4444	31,5348
Agosto	270	72900	430	184900			
Setembro	360	129600	410	168100			
	4070	2175500	4070	1848500			

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

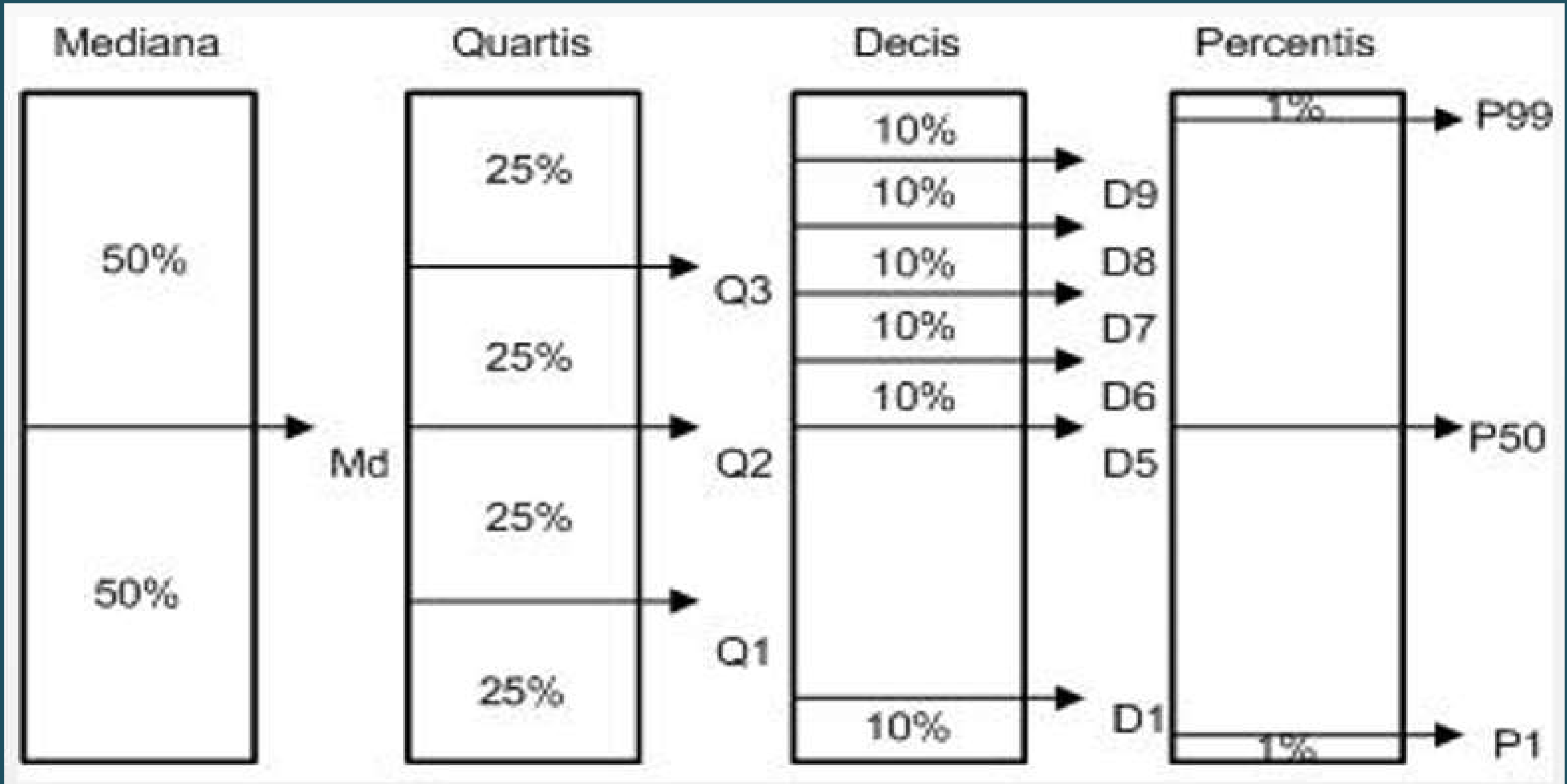
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

MEDIDAS DESCRITIVAS

- Medidas Separatrizes
  - A principal característica das medidas separatrizes consiste na separação da série ordenada em ordem crescente em partes iguais que apresentam o mesmo número de valores.
  - De um modo geral, os percentis, são números reais que dividem a sequência ordenada de dados em partes que contêm a mesma quantidade de elementos da série. É possível ter os seguintes múltiplos.
    - ✓ Mediana - Divide um conjunto em dois grupos com 50% dos dados cada;
    - ✓ Quartis - Divide o conjunto em quatro partes Iguais; (Q)
    - ✓ Quintis - Cada parte fica com 20%; (K)
    - ✓ Decis - Dividimos em 10 partes iguais o conjunto, onde cada parte fica com 10% dos dados.(D)
    - ✓ Percentis - Dividimos um conjunto de dados em cem partes, onde cada uma ficará com 1% dos elementos.
    - ✓ Os quartis, quintis e decis são múltiplos dos percentis.

MEDIDAS DESCRITIVAS

- Medidas Separatrizes



MEDIDAS DESCRITIVAS

- Medidas Separatrizes

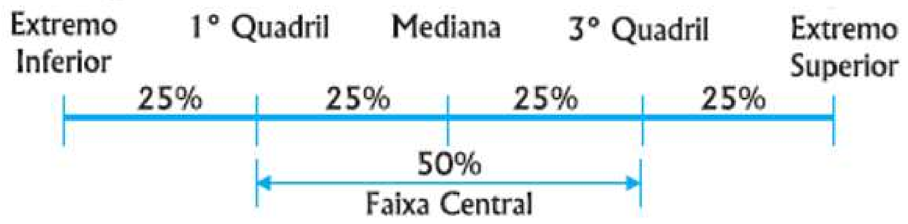
Quartis (Q_1, Q_2, e Q_3)

são valores de um conjunto de dados ordenados, que os dividem em quatro partes iguais.

Q1 : deixa 25% dos elementos abaixo dele.

Q2 : deixa 50% dos elementos abaixo dele e coincide com a mediana.

Q3 : deixa 75% dos elementos



$$Q_1 = l_i + \left[\frac{\left(\frac{\sum f_i}{4}\right) - f_{abs\ ant\ ac}}{f_{abs}}\right] . C$$

$$Q_2 = l_i + \left[\frac{\left(\frac{2\sum f_i}{4}\right) - f_{abs\ ant\ ac}}{f_{abs}}\right] . C$$

$$Q_3 = l_i + \left[\frac{\left(\frac{3\sum f_i}{4}\right) - f_{abs\ ant\ ac}}{f_{abs}}\right] . C$$

Sendo:

- l_i : limite inferior da classe que contém Q_i;
- n : número total de elementos da série;
- $f_{abs\ ant\ ac}$: frequência absoluta anterior acumulada da a classe que contém Q_i;
- f_{abs} : frequência absoluta da classe que contém Q_i;
- c : amplitude da classe que contém Q_i.

$$Q_k = l_i + \left[\frac{\left(\left(\frac{k . \sum f_i}{4}\right)\right) - f_{abs\ ant\ ac}}{f_{abs}}\right] . C$$

MEDIDAS DESCRITIVAS

- Medidas Separatrizes

Dercis (D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, e D_9)

- São valores que separam uma série em 10 partes iguais.
- Para cálculo usa-se a mesma técnica do calculo de quartis, substituído $\frac{\sum f_i}{2}$ por $\frac{k \sum f_i}{10}$, sendo k o número do dercil.

$$\boxed{D_k = l_i + \left[\frac{\left(\frac{k \sum f_i}{10}\right) - f_{abs\,ant\,ac}}{f_{abs}}\right] . \, C}$$

Sendo:

- l_i : limite inferior da classe que contém D_k;
- $\sum f_i$: número total de elementos da série;
- i : número do dercil (inteiro de 1 a 9);
- $f_{abs\,ant\,ac}$: frequência absoluta anterior acumulada a classe que contém D_k;
- f_{abs} : frequência absoluta da classe que contém D_k;
- c : amplitude da classe que contém D_k.

MEDIDAS DESCRITIVAS

- Medidas Separatrizes

Percentis $(P_1, P_2, P_3, \ldots P_{99})$

- São valores que separam uma série em 100 partes iguais.
- Para cálculo usa-se a mesma técnica do calculo de quartis, substituído $\frac{\sum f_i}{4}$ por $\frac{k \sum f_i}{100}$, sendo k o número do percentil.

$$P_k = l_i + \left[\frac{\left(\frac{k \sum f_i}{100} \right) - f_{abs\ ant\ ac}}{f_{abs}} \right] . C$$

Sendo:

- l_i : limite inferior da classe que contém P_k;
- $\sum f_i$: número total de elementos da série;
- i : número do percentil (inteiro de 1 a 99);
- $f_{abs\ ant\ ac}$: frequência absoluta anterior acumulada a classe que contém P_k;
- f_{abs} : frequência absoluta da classe que contém P_k;
- c : amplitude da classe que contém P_k.

MEDIDAS DESCRITIVAS

- Medidas Separatrizes

  - No Excel:

    = Quartil (valores ; n), n= 1,2,3 e 4

    = Percentil (valores ; k) k = 0,1; 0,2; 0,3; ..., 0,99)

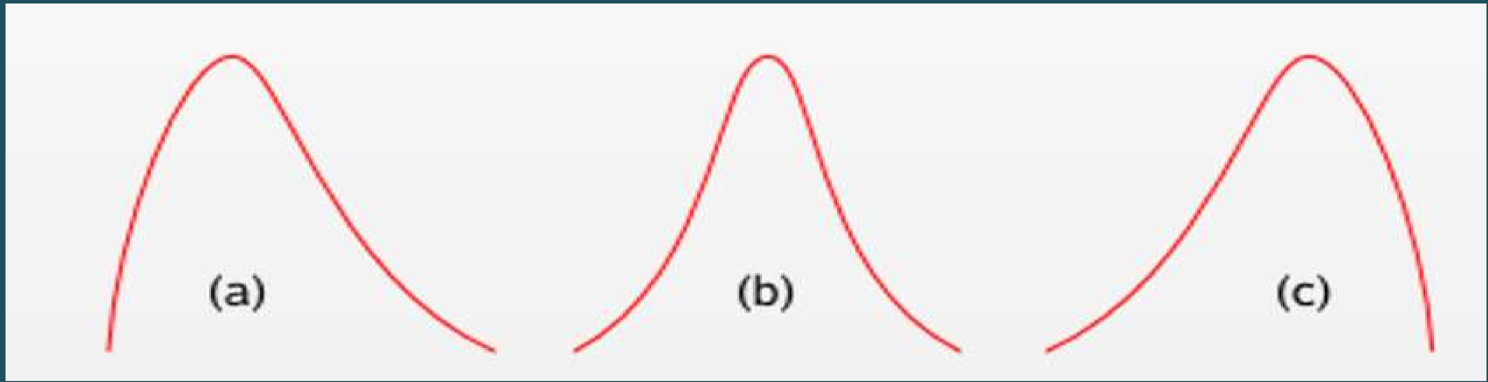
  - No R:

    > quantile (valores, probs = 0,25)

    > quantile (valores, probs = c (0.10, 0.35, 0.50, 0.75, 0.90))

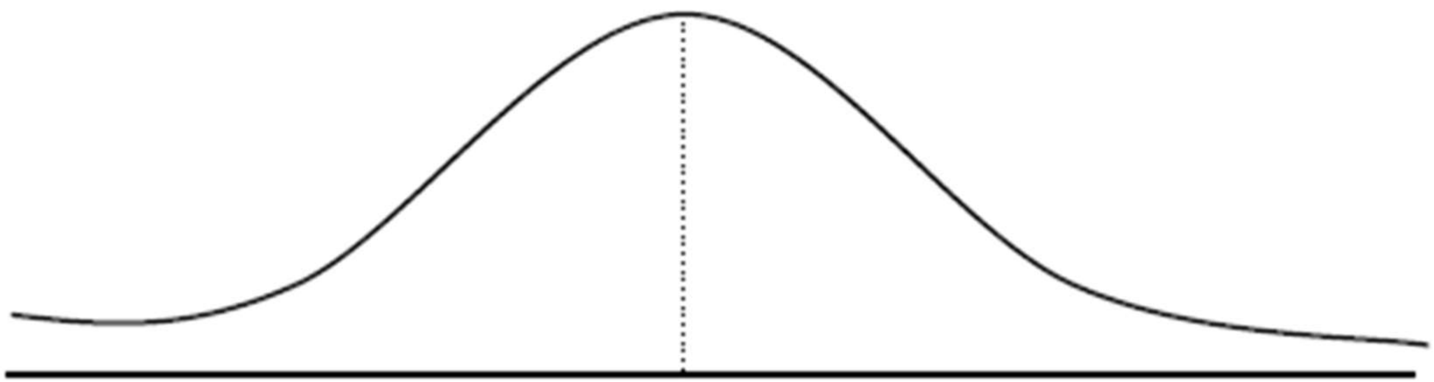
MEDIDAS DE ASSIMETRIA

- A medida de assimetria é baseada nas relações entre a média, mediana e moda.
- A distância entre a média e a moda pode ser usada para medir a assimetria, ou seja, quanto maior é a distância, seja negativa ou positiva, maior é a assimetria da distribuição.
- Estas medidas referem-se à forma da curva de uma distribuição de frequência, mas especificamente do polígono de frequência ou do histograma.
- Denomina-se assimetria o grau de afastamento de uma distribuição da unidade de assimetria.



MEDIDAS DE ASSIMETRIA

Simetria



$$\overline{X} = Mo = Md$$

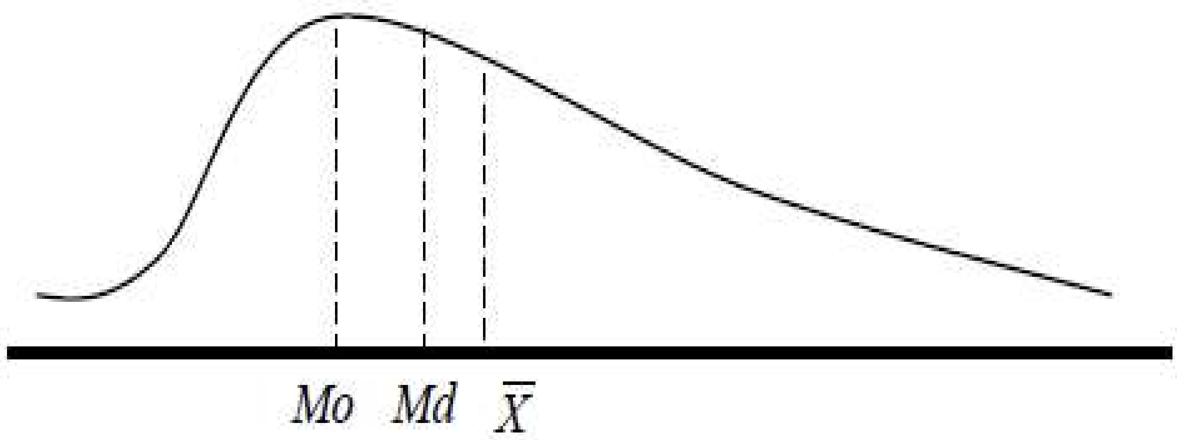
- Em uma distribuição simétrica, tem-se igualdade dos valores da média, mediana e moda.

MEDIDAS DE ASSIMETRIA

- Em uma distribuição assimétrica positiva, ou assimetria à direita, tem-se:

Assimetria à direita (ou positiva)

$$Mo < Md < \overline{X}$$

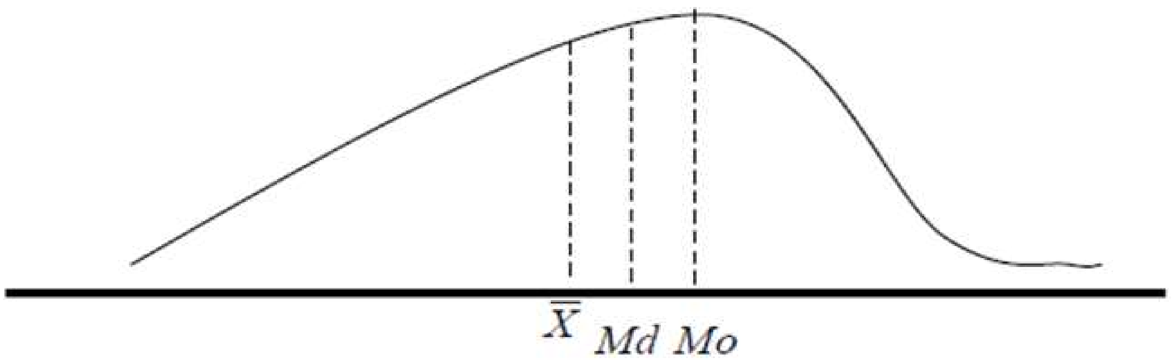


MEDIDAS DE ASSIMETRIA

- Em uma distribuição assimétrica negativa, ou assimetria à esquerda predominam valores inferiores à Moda.

Assimetria à esquerda (ou negativa)

$\overline{X} < Md < Mo$

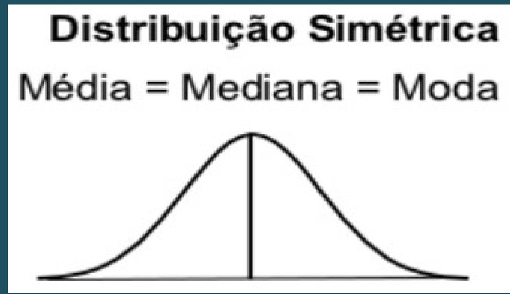


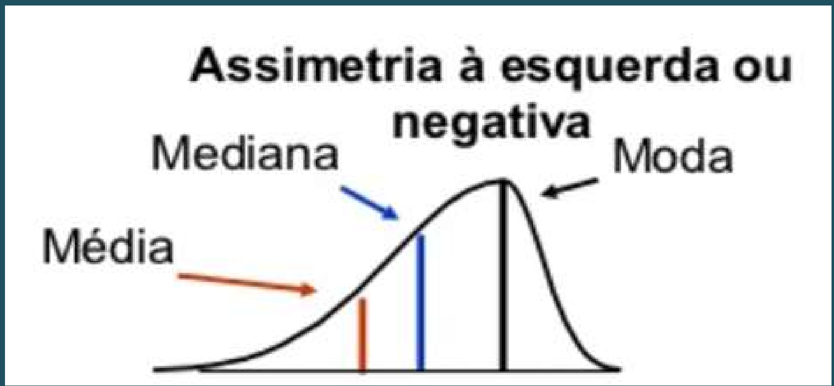
MEDIDAS DE ASSIMETRIA

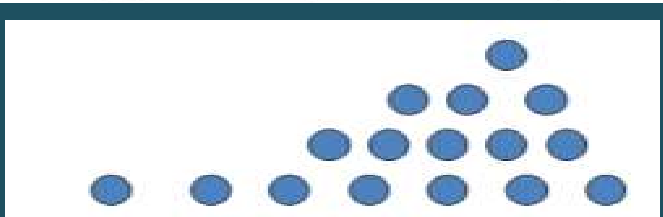
Tipos de Assimetria:

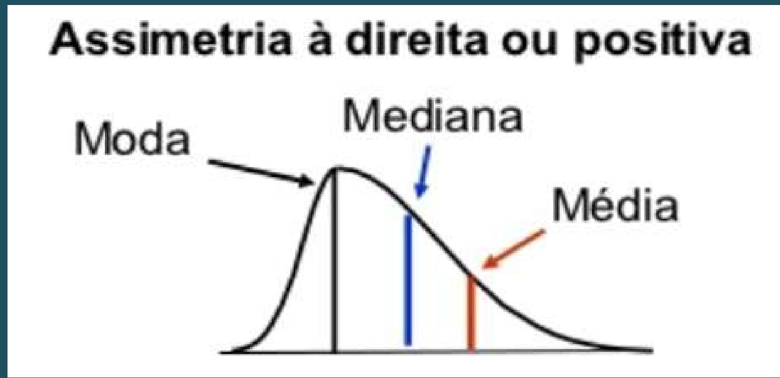
- $\bar{x} - Mo = 0 \rightarrow$ assimetria nula ou distribuição simétrica
- $\bar{x} - Mo < 0 \rightarrow$ assimetria negativa ou esquerda
- $\bar{x} - Mo > 0 \rightarrow$ assimetria positiva ou à direita

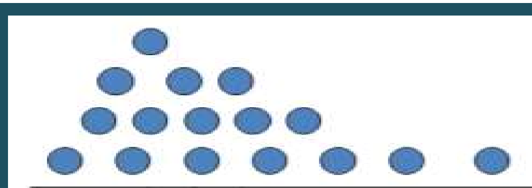
onde: Mo – Moda



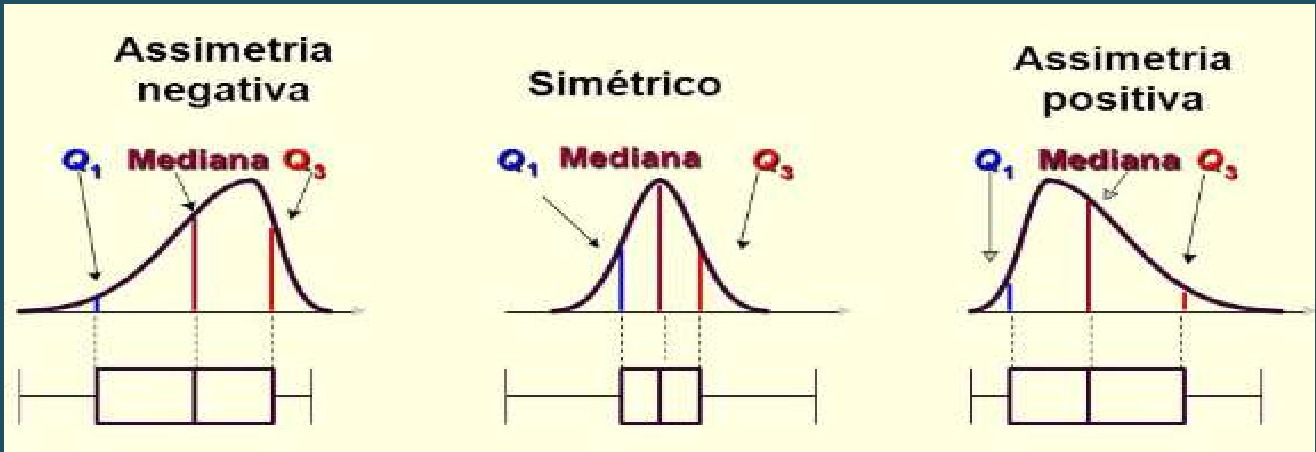


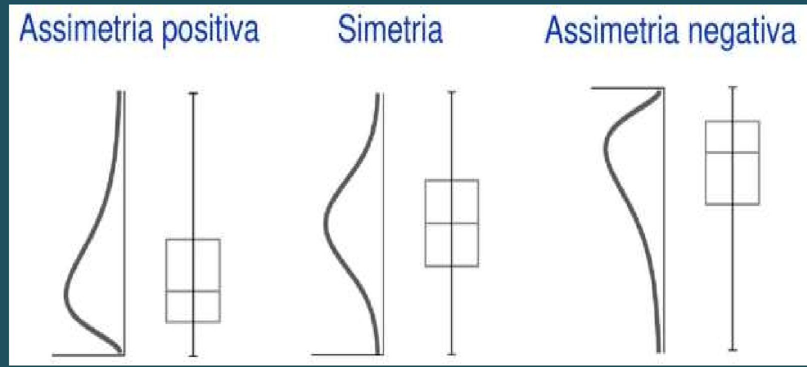


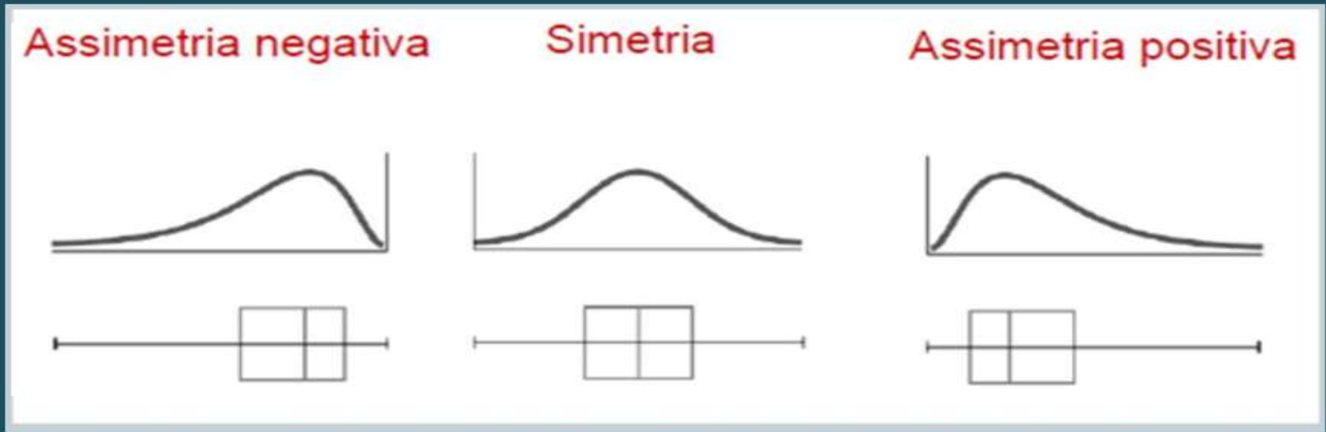




MEDIDAS DE ASSIMETRIA e BOXPLOT







MEDIDAS DE ASSIMETRIA

Coeficiente de assimetria baseados nas medidas de tendência central

O Coeficiente de Assimetria de Pearson, A_p, baseia-se na posição relativa das medidas de tendência central de acordo com o tipo de assimetria dos dados . Ele é definido como

$$A_P = \frac{\bar{x} - m_o}{S}.$$

Temos

a) Distribuições simétricas unimodais: $\bar{x} = m_d = m_o$; nesse caso, $A_P = 0$
b) Distribuições assimétricas positivas: $\bar{x} > m_d > m_o$; então $A_P > 0$
c) Distribuições assimétricas negativas: $\bar{x} < m_d < m_o$, fazendo com que $A_P < 0$.

A determinação da moda para dados contínuos não é trivial. Uma alternativa é utilizar o coeficiente

$$A_{P2} = \frac{\bar{x} - m_d}{S}.$$

MEDIDAS DE ASSIMETRIA

Coeficiente de assimetria baseado em quartis

Para distribuições simétricas, temos que $(Q_3 - m_d) = (m_d - Q_1)$. Por outro lado,

a) Para distribuições assimétricas positivas $(Q_3 - m_d) > (m_d - Q_1)$.
b) Para distribuições assimétricas negativas $(Q_3 - m_d) < (m_d - Q_1)$.

Observando esses fatos, foi proposto o seguinte coeficiente:

$$A_Q = \frac{(Q_3 - m_d) - (m_d - Q_1).}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2m_d}{Q_3 - Q_1}.$$

A função do denominador, assim como em A_P é fazer com que este coeficiente seja adimensional, permitindo a comparação entre conjuntos de dados medidos em diferentes escalas.

A interpretação é feita da seguinte maneira

a) Se a distribuição foi simétrica, então $A_Q = 0$.
b) Se a distribuição foi assimétrica positiva, então $A_Q > 0$.
c) Se a distribuição foi assimétrica negativa, então $A_Q < 0$.

MEDIDAS DE ASSIMETRIA

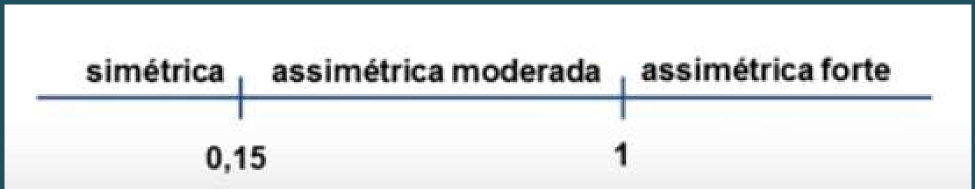
- **Coeficiente de assimetria (b_1)**

Seja $X_1, X_2, \cdots, X_n$ um conjunto de dados e $Z_i = \frac{X_i - \bar{X}}{S}$, $i = 1, 2, \cdots, n$.

$$b_1 = \sum_{i=1}^{n} \frac{Z_i^3}{n} = \frac{m_3}{S^3}$$

- Para dados agrupados:

$$m_3 = \sum_{j=1}^{k} f_i\,(c_i - \bar{x})^3,$$ sendo c_i o ponto médio da faixa i e k, o número de faixas



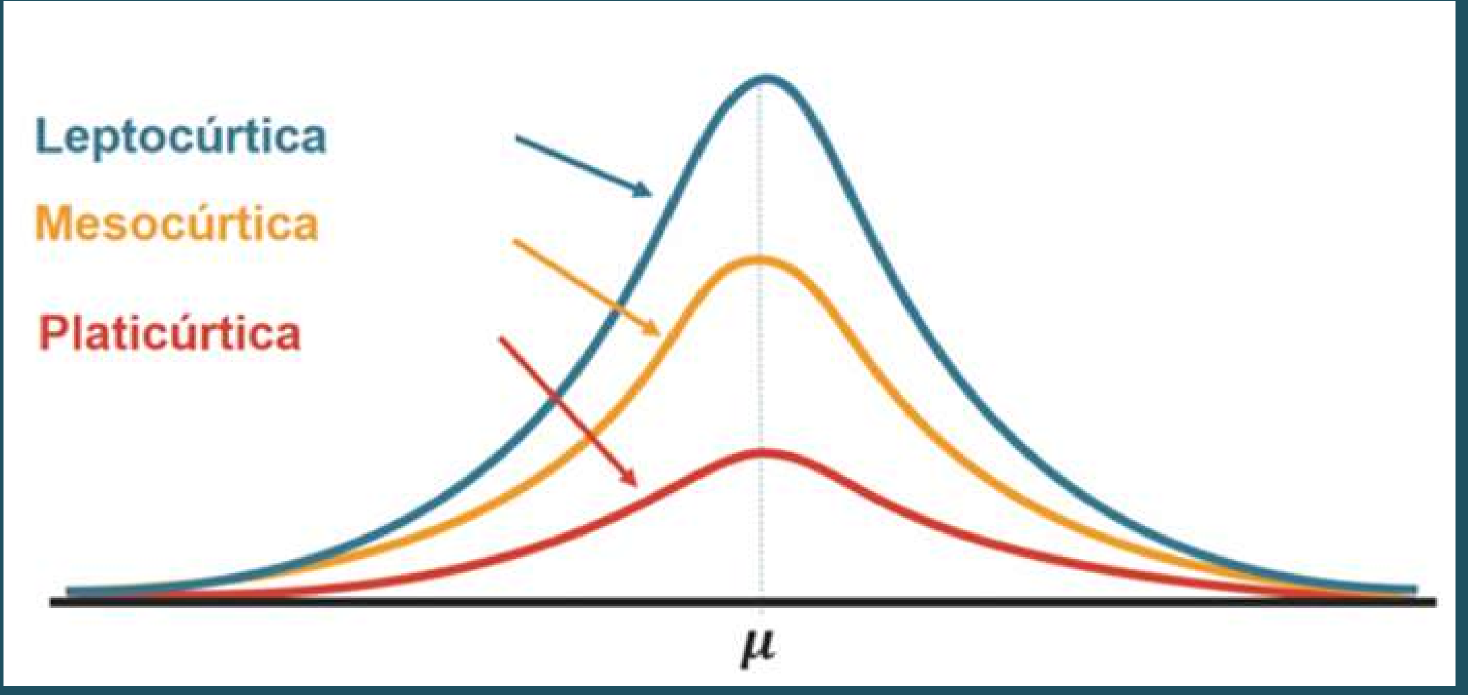

```
package 'moments' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
 C:\Users\daiane.mattos\AppData\Local\Temp\RtmpCUZ57S\downloaded_packages
> require(moments)
Carregando pacotes exigidos: moments
Warning message:
package 'moments' was built under R version 3.1.2
> skewness(renda[,31])
[1] 6.105393
> hist(renda[,31])
>
```

MEDIDAS DE CURTOSE

- A curtose é o grau de achatamento (ou afilamento) de uma distribuição em comparação com uma distribuição padrão (curva normal).

  - De acordo com o grau de curtose, podemos ter:



MEDIDAS DE CURTOSE

Para medir o grau de curtose utiliza-se o coeficiente:

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

Q_1 : valor do 1º Quartil
Q_3 : valor do 3º Quartil
P_{10} : valor do percentil 10
P_{90} : valor do percentil 90

- Se K= 0,263, diz-se que a curva correspondente à distribuição de freqüência é mesocúrtica.
- Se K > 0,263, diz-se que a curva correspondente à distribuição de freqüência é platicúrtica.
- Se K < 0,263, diz-se que a curva correspondente à distribuição de freqüência é leptocúrtica.

Exercício Resolvido

$$Q_k = l_i + \left[\frac{\left(\left(\frac{k \,.\, \sum f_i}{4}\right)\right) - f_{abs\ ant\ ac}}{f_{abs}}\right].\mathrm{c}$$

IMC*
15,6807
17,7096
18,3655
18,5108
21,3382
21,5139
22,5466
23,1481
23,2254
23,5304
23,5898
23,6742
24,8016
24,8016
24,9135
25,1429
25,7630
26,0375
26,1286
26,4463
27,3588
28,0816
28,6501
29,0487
29,2421
29,3629
29,3848
30,6394
30,8444
33,4622

ORDEM	CLASSES (IMC)	fa(i)	fr(i)	fac(i)	frac(i)	X (i)	x (i) * fa(i)	fa(i) * X(i)^2
1	15,68|---18,68	4	13,33%	4	13,33%	17,18	68,72	1.180,61
2	18,68|---21,68	2	6,67%	6	20,00%	20,18	40,36	814,46
3	21,68|---24,68	6	20,00%	12	40,00%	23,18	139,08	3.223,87
4	24,68|---27,68	9	30,00%	21	70,00%	26,18	235,62	6.168,53
5	27,68|---30,68	7	23,33%	28	93,33%	29,18	204,26	5.960,31
6	30,68|---33,68	2	6,67%	30	100,00%	32,18	64,36	2.071,10
		30	100,00%				752,40	19.418,89

Menor valor	15,68
Maior valor	33,46
Amplitude	17,78
N° de classes	5,48
Amp. Classe	3,0

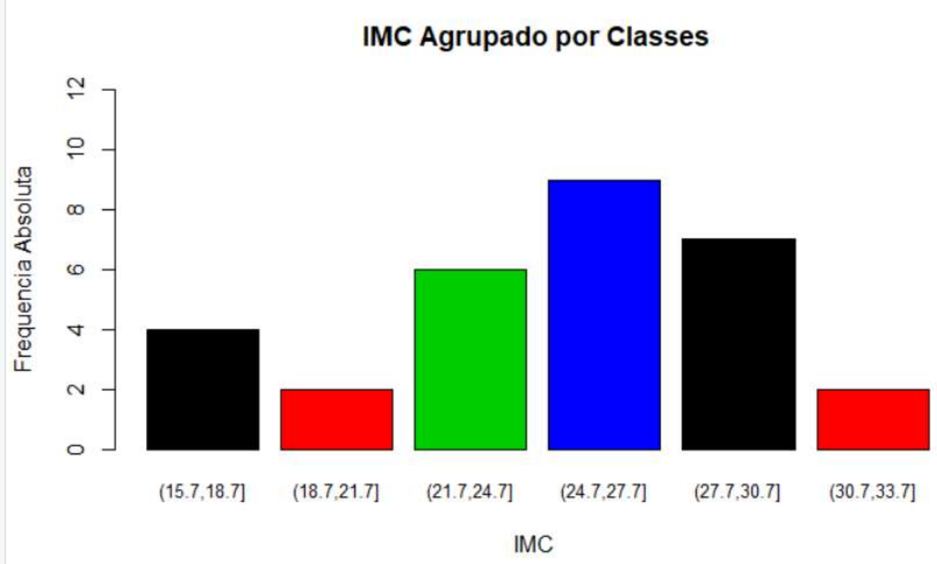
MÉDIA	25,08
MEDIANA	25,68
MODA	26,18
MODA de CRUZBER	26,48
DESVIO PADRÃO	4,35
COEFICIENTE de VARIAÇÃO	17,34
Q3	28,32
D5	25,68

Exercício Resolvido

Usando o software R

```
>
>
> limcl <- c( 15.68, 18.68, 21.68, 24.68, 27.67, 30.68, 33.68)
> freqcl <- table(cut(x, breaks = limcl))
> freqcl

(15.7,18.7] (18.7,21.7] (21.7,24.7] (24.7,27.7] (27.7,30.7] (30.7,33.7]
          4           2           6           9           7           2
>
> barplot (freqcl,
+          col=1:4 ,
+          ylim=c(0,12) ,
+          xlab="IMC ",ylab=" Frequencia Absoluta ",
+          main=" IMC Agrupado por Classes ",
+          cex.names = 0.8)
> 
```

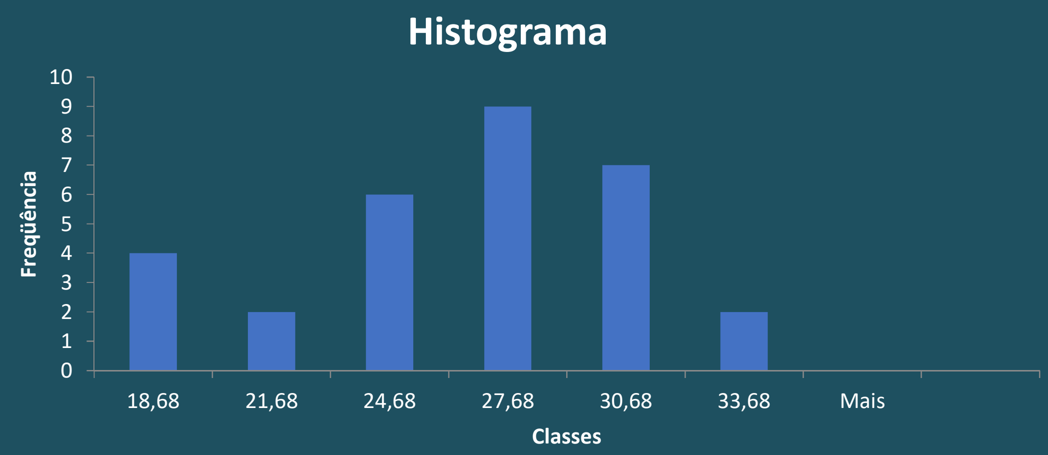



```
> x
 [1] 15.6807 17.7096 18.3655 18.5108 21.3382 21.5139 22.5466 23.1481 23.2254 23.5304 23.5898
[12] 23.6742 24.8016 24.8016 24.9135 25.1429 25.7630 26.0375 26.1286 26.4463 27.3588 28.0816
[23] 28.6501 29.0487 29.2421 29.3629 29.3848 30.6394 30.8444 33.4622
> 
```

Exercício Resolvido

	IMC*	Classes		
	15,6807	18,68	Classes	Freqüência
	17,7096	21,68	18,68	4
	18,3655	24,68	21,68	2
	18,5108	27,68	24,68	6
	21,3382	30,68	27,68	9
	21,5139	33,68	30,68	7
	22,5466		33,68	2
	23,1481		Mais	0
	23,2254			
	23,5304			
	23,5898			
	23,6742			
	24,8016			
	24,8016			
	24,9135			
	25,1429			
	25,7630			
	26,0375			
	26,1286			
	26,4463			
	27,3588			
	28,0816			
	28,6501			
	29,0487			
	29,2421			
	29,3629			
	29,3848			
	30,6394			
	30,8444			
	33,4622			

Usando o software Excel



Exercício Resolvido

Classificação do IMC:

Menor que 18,5 - Abaixo do peso

Entre 18,5 e 24,9 - Peso normal

Entre 25 e 29,9 - Sobrepeso (acima do peso desejado)

Igual ou acima de 30 - Obesidade

Cálculo do IMC:

IMC=peso (kg) / altura (m) x altura (m)

Exemplo: João tem 83 kg e sua altura é 1,75 m

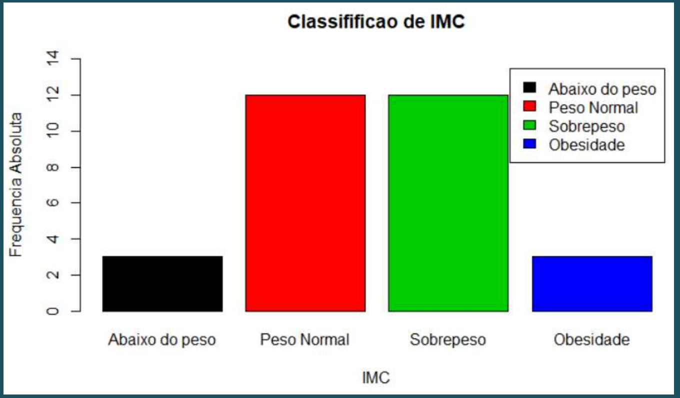
Altura x altura = 1,75 x 1,75 = 3.0625

IMC = 83 divididos por 3,0625 = 27,10

O resultado de 27,10 de IMC indica que João está acima do peso desejado (sobrepeso).

```
>
> rotulox <- c ( "Abaixo do peso",
+               "Peso Normal",
+               "Sobrepeso",
+               "Obesidade")
> limval <- c( 0, 18.5, 24.99, 29.99, 50)
> freqc <- table(cut(x, breaks = limval))
> freqc

 (0,18.5] (18.5,25]   (25,30]   (30,50]
        3        12        12         3
> barplot(freqc, names.arg = rotulox,
+         col = 1:4,
+         ylim = c(0,14),
+         legend.text = rotulox,
+         xlab = "IMC",
+         ylab = "Frequencia Absoluta",
+         main = "Classifificao de IMC")
>
```



...

- **DÚVIDAS ???**

E-mail: Petrucio.barros@ic.ufal.br

Materiais das Aulas na plataforma Moodle

Créditos ao professor Estevam Vilar por parte do material das aulas.