UENF
Universidade Estadual do Norte Fluminense Darcy Ribeiro

CCT
LCMAT

Probabilidade e Estatística
Estatística Descritiva

Prof. Fermín Alfredo Tang Montané

Medidas de dispersão

- As medidas de tendência central, por si só, não são suficientes para caracterizar um conjunto de dados.
- Em particular, se comparamos dois conjuntos de dados as medidas de tendência central não fornecem qualquer informação referente a distribuição dos mesmos.
- **Exemplo:**
- Dois candidatos à emprego fizeram 5 provas e desejamos comparar seus rendimentos com base na media aritmética.
  - Candidato A: 70, 71, 69, 70, 70 Média = 70
  - Candidato B: 40, 80, 98, 62, 70 Média = 70
- Com base somente na média aritmética diríamos que os dois candidatos apresentaram o mesmorendimento. Porém, como podemos observar o candidato A apresentou notas mais uniformes.

Medidas de dispersão

- Permitem avaliar quantitativamente o grau de variabilidade ou dispersão dos valores de um conjunto de números em torno do valor médio.
- As principais estatísticas de dispersão são:
  - Amplitude total
  - Desvio médio
  - Variância
  - Desvio-padrão
  - Coeficiente de variação

Amplitude Total

- Amplitude total é a diferença entre o maior e o menor valor dos dados.
- **Exemplo 1:**
- A tabela abaixo apresenta o rendimento diário (em %) de três empregados:

Empregado	Dia 1	2	3	4	5	Média	Máx.	Mín.	Amplitude Total
A	92	80	85	70	87	82,8	92	70	22
B	90	87	78	72	87	82,8	90	72	18
C	85	81	87	82	79	82,8	87	79	8

- Os dados mostram que embora os três empregados possuem a mesma média, os seus desempenhos são diferentes.

Amplitude Total

- No entanto, a medida de Amplitude total pode não ser sempre a mais adequada para medir a dispersão dos dados.
- **Exemplo 2:**
- A tabela abaixo apresenta o rendimento diário (em %) de três empregados:

Empregado	Dia					Média	Máx.	Mín.	Amplitude Total
	1	2	3	4	5				
A	82	70	65	60	73	70	82	60	22
B	60	78	68	62	82	70	82	60	22
C	53	72	75	75	75	70	75	53	22

- Embora os desempenhos sejam diferentes, nem a média aritmética nem a amplitude total mostram qualquer variação.

Desvio Médio

- O desvio médio de um conjunto de n valores $x_1, x_2, x_3, \ldots, x_n$ é definido pela expressão:

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

- No caso de dados agrupados em intervalos de classe:

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

- onde, k é o número de classes.
- Esta medida de dispersão considera todos os valores do conjunto de dados, e não apenas os valores extremos.

Desvio Médio

- **Exemplo:**
- A tabela abaixo apresenta o rendimento diário (em %) de três empregados:

Empregado	Dia					Média	Desvio Médio (d)	Amplitude Total
	1	2	3	4	5			
A	92	80	85	70	87	82,8	6,24	22
B	90	87	78	72	87	82,8	6,24	18
C	85	81	87	82	79	82,8	2,56	8

- Com objetivo ilustrativo mostram-se os valores dos desvios.

9,2	2,8	2,2	12,8	4,2
7,2	4,2	4,8	10,8	4,2
2,2	1,8	4,2	0,8	3,8

Variância amostral (s^2)

- A variância de um conjunto de n valores $x_1, x_2, x_3, \ldots, x_n$ (que representam uma amostra) é definido pela expressão:

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

- No caso de dados agrupados em intervalos de classe:

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

- onde, k é o número de classes.

- Esta medida de dispersão intensifica o valor dos desvios ao considerar o termo quadrático. Magnifica desvios muito grandes.

Variância amostral (s^2)

- **Exemplo:**
- A tabela abaixo apresenta o rendimento diário (em %) de três empregados:

Empregado	Dia 1	2	3	4	5	Média	Desvio Médio (d)	Variância Amostral
A	92	80	85	70	87	82,8	6,24	69,7
B	90	87	78	72	87	82,8	6,24	56,7
C	85	81	87	82	79	82,8	2,56	10,2

- Os valores dos quadrados dos desvios são os seguintes.

84,64	7,84	4,84	163,84	17,64
51,84	17,64	23,04	116,64	17,64
4,84	3,24	17,64	0,64	14,44

Desvio Padrão Amostral (s)

- O desvio padrão amostral é a raiz quadrada da variância amostral.
- Para dados não agrupados é definido pela expressão:

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

- No caso de dados agrupados em intervalos de classe:

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i.\left(x_i - \bar{x}\right)^2}{n-1}}$$

- onde, k é o número de classes.

Coeficiente de variação (CV)

- O coeficiente de variação é a razão entre o desvio-padrão e a média aritmética, em porcentagem. Definido pela seguinte expressão:

$$CV = \frac{s}{\bar{x}}.100$$

Variância amostral (s^2), Desvio Padrão Amostral (s), Coeficiente de variação (CV)

- b) Para dados agrupados em classes discretas.

Idade (X_i)	Número de Alunos (f_i)	$x_i f_i$	$(x_i - \bar{x})^2$	$f_i.(x_i - \bar{x})^2$
20	1	20	13,80	13,80
21	3	63	7,37	22,10
22	4	88	2,94	11,76
23	7	161	0,51	3,57
24	9	216	0,08	0,73
25	6	150	1,65	9,92
26	4	104	5,22	20,90
27	0	0	10,80	0,00
28	1	28	18,37	18,37
Total	**35**	**830**		**101,14**

$$\bar{x} = \frac{830}{35} = 23,714$$

$$s^2 = \frac{101,14}{34} = 2,9747$$

$$s = \sqrt{2,9747} = 1,7247$$

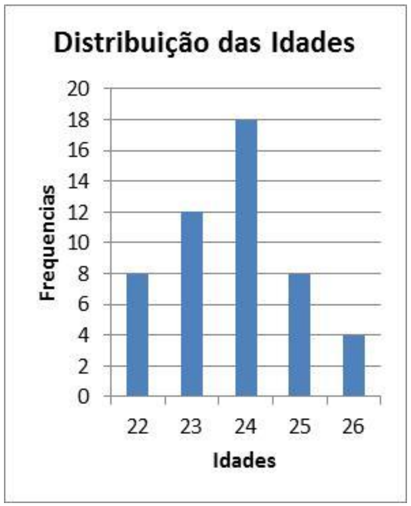
$$s = 1,72 \text{ anos}$$

$$CV = \frac{1,7247}{23,714}.100$$

$$= 7,2729\%$$

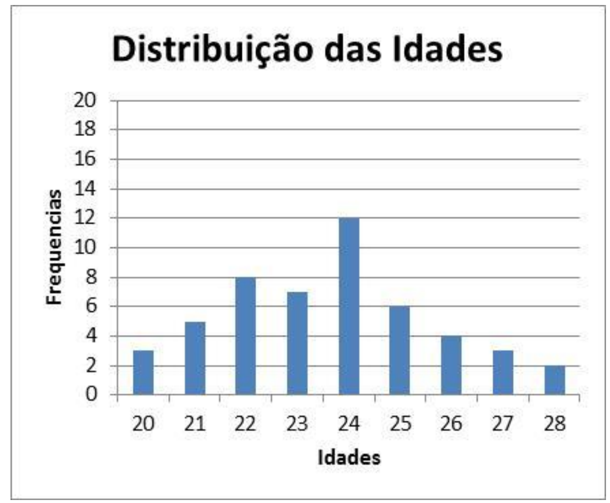
Variância amostral (s^2), Desvio Padrão Amostral (s), Coeficiente de variação (CV)

- O desvio padrão aumenta para dados com maior dispersão.



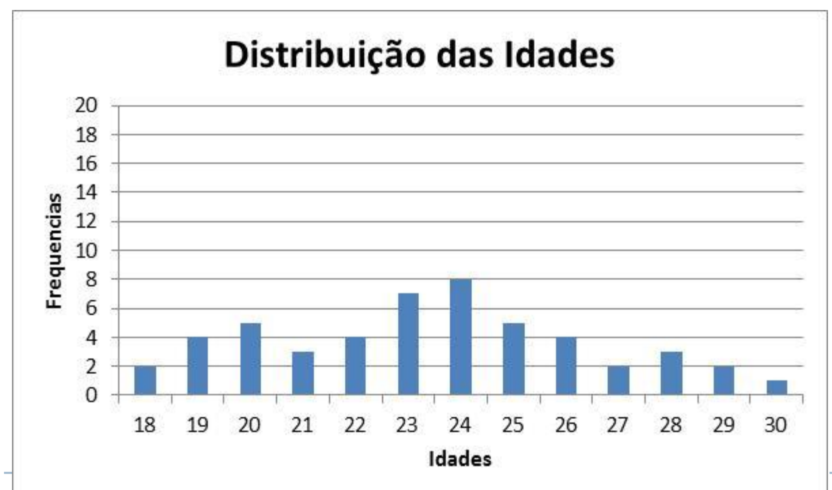
$\bar{x} = 23,76$

$s = 1,15$



$\bar{x} = 23,62$

$s = 2,05$



$\bar{x} = 23,42$

$s = 3,07$

Variância amostral (s^2), Desvio Padrão Amostral (s), Coeficiente de variação (CV)

- b) Para dados agrupados em intervalos de classes.

Intervalo de Classe Alturas(m)	Número de Alunos (f_i)	x_i	$f_i x_i$	$(x_i - \bar{x})^2$	$f_i.(x_i - \bar{x})^2$
[1,45 1,49 >	4	1,47	5,88	0,009216	0,03686
[1,49 1,53 >	8	1,51	12,08	0,003136	0,02509
[1,53 1,57 >	4	1,55	6,20	0,000256	0,00102
[1,57 1,61 >	5	1,59	7,95	0,000576	0,00288
[1,61 1,65 >	4	1,63	6,52	0,004096	0,01638
[1,65 1,69 >	5	1,67	8,35	0,010816	0,05408
Total	**30**		**46,98**		**0,13632**

$$\bar{x} = \frac{46,98}{30} = 1,566$$

$$s^2 = \frac{0,1362}{30} = 0,00454$$

$$s = \sqrt{0,00454} = 0,06737$$

$$s = 0,07 \text{ metros}$$

$$CV = \frac{0,06737}{1,566}.100$$

$$= 4,3020\%$$

Medidas de Posição ou Separatrizes

- São medidas que dividem um conjunto de valores em um certo número de partes iguais.
- A mediana, por exemplo, divide um conjunto de dados em duas partes iguais.

- Outras medidas de posição importantes são:
  - Quartis
  - Decis
  - Centis ou Percentis

Quartis

- O quartil divide um conjunto de valores **ordenado** em quatro partes iguais.
  - O primeiro quartil (Q_1) é o valor que antecede 25% da freqüência abaixo dele e sucede 75%,
  - O segundo quartil (Q_2) é igual ao valor da mediana e
  - O terceiro quartil (Q_3) é o valor que antecede 75% da freqüência abaixo dele e sucede 25%.

Quartis

- a) O cálculo do quartil i-êsimo é semelhante ao cálculo da mediana para dados não agrupados.
- **Exemplo:** Determine o 3° quartil das idades dos 35 alunos **(Ímpar)**.
- A definição do quartil Q_i, depende da existência de um número par ou impar de elementos.

20	21	21	21	22	22	22	22	23	23	23	23	23	23	23	24	24	24	24	24	24	24	24	24	25	25	25	25	25	25	26	26	26	26	28

$P_1 = 9$ $\quad$ $P_2 = 18$ $\quad$ $P_3 = 27$

$Q_1 = 23$ $\quad$ $Q_2 = 24$ $\quad$ $Q_3 = 25$

- No caso ímpar, a posição P_i do quartil Q_i existe:

$$\boxed{P_i = \left\lfloor \frac{i.n}{4} \right\rfloor + 1, \quad \forall \text{ i=1, 2, 3.}}$$

$$P_1 = \left\lfloor \frac{1(35)}{4} \right\rfloor + 1 = \lfloor 8,75 \rfloor + 1 = 9$$

$$P_2 = \left\lfloor \frac{2(35)}{4} \right\rfloor + 1 = \lfloor 17,5 \rfloor + 1 = 18$$

$$P_3 = \left\lfloor \frac{3(35)}{4} \right\rfloor + 1 = \lfloor 26,25 \rfloor + 1 = 27$$

Quartis

- a) O cálculo do quartil i-êsimo é semelhante ao cálculo da mediana para dados não agrupados.
- **Exemplo:** Determine o 3° quartil das idades dos 34 alunos **(Par)**.
- A definição do quartil Q_i, depende da existência de um número par ou impar de elementos.

20	21	21	21	22	22	22	22	23	23	23	23	23	23	23	24	24	24	24	24	24	24	24	24	25	25	25	25	25	25	26	26	26	26

$P_1 = 8$ $\quad$ $P_2 = 17$ $\quad$ $P_3 = 26$

$Q_1 = 22,5$ $\quad$ $Q_2 = 24$ $\quad$ $Q_3 = 25$

- No caso par, o valor do quartil Q_i é calculado como a média dos valores de duas posições. Não existe uma formula clara para as posições P_i.

Quartis

- **Exemplo:** Determine o 3° quartil das idades dos 35 alunos:

Idade (x_i)	Número de Alunos (f_i)	f_{ac}
20	1	1
21	3	4
22	4	8
23	7	15
24	9	24
25	6	30
26	4	34
27	0	34
28	1	35
Total	**35**	

A posição do Q_3:

$$P_i = \left\lfloor \frac{i.n}{4} \right\rfloor + 1, \qquad \text{para i=1, 2, 3.}$$

$$P_3 = \left\lfloor \frac{3(35)}{4} \right\rfloor + 1 = \lfloor 26,25 \rfloor + 1 = 27$$

O quartil Q_3 se encontra na posição 27.

$$Q_3 = 25 \text{ anos.}$$

Quartis

- b) Para dados agrupados em intervalos de classe a expressão para o cálculo do quartil i-êsimo é uma generalização da expressão para o cálculo da mediana.

$$Q_i = LI_i + \left(\frac{P_i - f'_{ac}}{f_{Q_i}} \right) .h, \qquad \text{para i=1, 2, 3.}$$

onde:

$$P_i = \frac{i.n}{4}, \qquad \text{para i=1, 2, 3.}$$

- LI_i: limite inferior da classe Q_i;
- P_i: posição do quartil i-êsimo;
- f'_{ac}: frequência acumulada da classe anterior a classe Q_i;
- f_{Qi}: frequência da classe Q_i;
- h: amplitude do intervalo de classe.

Decis

- O decil divide um conjunto de valores ordenados em dez partes iguais e são representados por $D_1, D_2, \ldots, D_9$.
- Sendo que o 5° decil é a mediana.

Decis

- b) Para dados agrupados em intervalos de classes a expressão para o cálculo do decil i-ésimo é o seguinte:

$$D_i = LI_i + \left(\frac{P_i - f'_{ac}}{f_{D_i}} \right).h, \qquad \text{para i=1, ..., 9.}$$

$$P_i = \frac{i.n}{10}, \qquad \text{para i=1, ..., 9.}$$

onde:

- LI_i: limite inferior da classe D_i;
- P_i: posição do decil i-êsimo;
- f'_{ac}: frequência acumulada da classe anterior a classe D_i;
- f_{Di}: frequência da classe D_i;
- h: amplitude do intervalo de classe.

Centis ou Percentis

- O centil ou percentil divide um conjunto de valores ordenados em 100 partes iguais;
- São representados por $C_1, C_2, \dots, C_{99}$.
  - o 50° centil é a mediana e
  - o 25° e 75° centis correspondem ao 1° e ao 3° quartis, respectivamente.

Centis ou Percentis

- b) Para dados agrupados em intervalos de classes a expressão para o cálculo do decil i-êsimo é o seguinte:

$$C_i = LI_i + \left(\frac{P_i - f'_{ac}}{f_{C_i}} \right).h, \qquad \text{para i=1, ..., 99.}$$

$$P_i = \frac{i.n}{100}, \qquad \text{para i=1, ..., 99.}$$

onde:

- LI_i: limite inferior da classe C_i;
- P_i: posição do centil i-êsimo;
- f'_{ac}: frequência acumulada da classe anterior a classe C_i;
- f_{Ci}: frequência da classe C_i;
- h: amplitude do intervalo de classe.

Exemplo - Quartil

- No exemplo das alturas dos 30 alunos determine o 3° quartil.

Intervalo de Classe Alturas (m)	Número de alunos f_i	f_{ac}
[1,45 1,49 >	4	4
[1,49 1,53 >	8	12
[1,53 1,57 >	4	16
[1,57 1,61 >	5	21
[1,61 1,65 >	4	25
[1,65 1,69 >	5	30
Total	30	

Posição do 3º Quartil:

$$P_3 = \frac{3(30)}{4} = 22,5$$

Calculo do 3º Quartil:

$$Q_i = LI_i + \left(\frac{P_i - f'_{ac}}{f_{Q_i}} \right).h, \text{ i=1,...,3.}$$

$$Q_3 = 1,61 + \left(\frac{22,5 - 21}{4} \right).0,04$$

$$= 1,61 + 0,015 = 1,625 \text{ metros}$$

Exemplo - Decil

- No exemplo das alturas dos 30 alunos determine o 6° decil.

Intervalo de Classe Alturas (m)	Número de alunos f_i	f_{ac}
[1,45 1,49 >	4	4
[1,49 1,53 >	8	12
[1,53 1,57 >	4	16
[1,57 1,61 >	5	21
[1,61 1,65 >	4	25
[1,65 1,69 >	5	30
Total	**30**	

Posição do 6º Decil:

$$P_6 = \frac{6(30)}{10} = 18$$

Calculo do 6º Decil:

$$D_i = LI_i + \left(\frac{P_i - f'_{ac}}{f_{D_i}} \right).h, \text{ i=1,...,9.}$$

$$D_6 = 1,57 + \left(\frac{18 - 16}{5} \right).0,04$$

$$= 1,57 + 0,016 = 1,586 \text{ metros}$$

Exemplo - Centil

- No exemplo das alturas dos 30 alunos determine o 20° centil.

Intervalo de Classe Alturas (m)				Número de alunos f_i	f_{ac}
[	1,45	1,49	>	4	4
[	1,49	1,53	>	8	12
[	1,53	1,57	>	4	16
[	1,57	1,61	>	5	21
[	1,61	1,65	>	4	25
[	1,65	1,69	>	5	30
Total				30	

Posição do 20º Centil:

$$P_{20} = \frac{20(30)}{100} = 6$$

Calculo do 20º Centil:

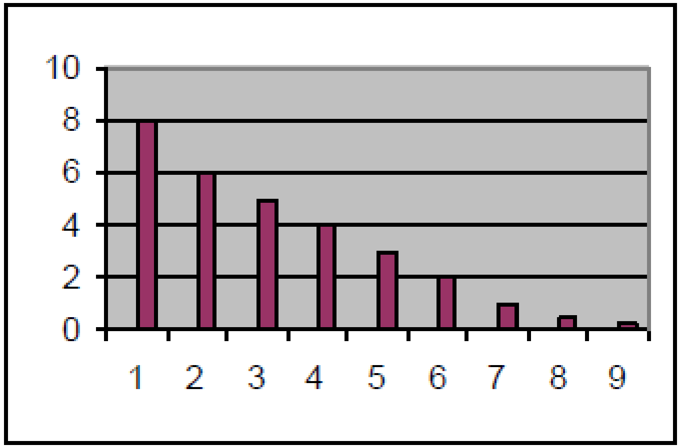
$$C_i = LI_i + \left(\frac{P_i - f'_{ac}}{f_{C_i}} \right).h, \text{ i=1, ..., 99}$$

$$C_{20} = 1,49 + \left(\frac{6-4}{8} \right).0,04$$

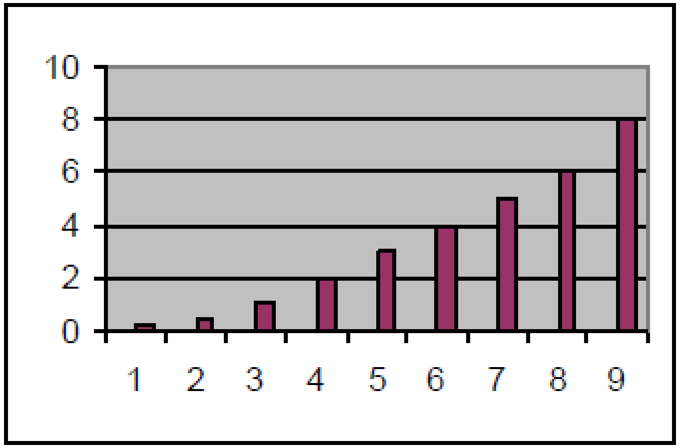
$$= 1,49 + 0,01 = 1,5 \text{ metros}$$

Medidas de Assimetria

- As medidas de assimetria procuram caracterizar o quanto o histograma de uma distribuição de freqüência se afasta da condição de simetria em relação à uma medida de tendência central.



Distribuição assimétrica positiva



Distribuição assimétrica negativa

Coeficiente de Assimetria de Pearson (A)

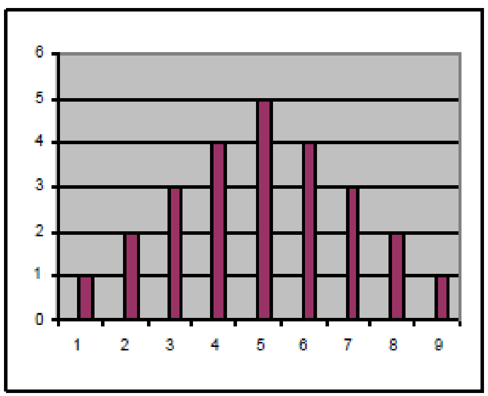
- O grau de assimetria de uma distribuição de frequência pode ser avaliada utilizando o coeficiente de Pearson:

$$A = \frac{\overline{x} - M_o}{s}$$

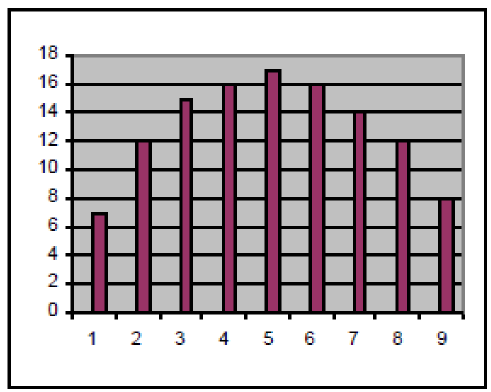
- onde:
  - |A| < 0,15 : distribuição praticamente simétrica;
  - 0,15 < |A| < 1 : distribuição assimétrica moderada;
  - |A| > 1 : distribuição fortemente assimétrica.

Medidas de Curtose

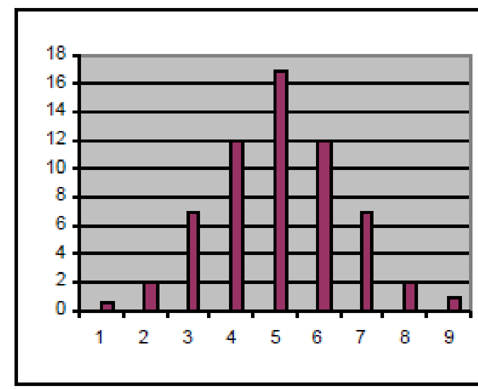
- As medidas de curtose caracterizam uma distribuição simétrica ou aproximadamente simétrica quanto ao seu achatamento, tomando como referência uma distribuição normal, que será objeto de estudo mais adiante.



Mesocúrtica
(normal)



Platicúrtica



Leptocúrtica

Coeficiente Percentílico de Curtose (C)

- O grau de achatamento com relação a distribuição normal de uma distribuição de frequência pode ser avaliado através do coeficiente percentílico:

$$C = \frac{C_{75} - C_{25}}{2(C_{90} - C_{10})}$$

- Onde, C_{10}, C_{25}, C_{75} e C_{90} são os 10°, 25°, 75° e 90° centis (ou percentis)

- Sendo que:
  - Se C = 0,263: distribuição é mesocúrtica (normal)
  - Se C < 0,263: distribuição leptocúrtica (alongada)
  - Se C > 0,263: distribuição platicúrtica (achatada)

Exemplo - Assimetria

- No exemplo das alturas dos 30 alunos classifique a distribuição quanto a assimetria.

Intervalo de Classe Alturas (m)	Número de alunos f_i	f_{ac}
[1,45 1,49 >	4	4
[1,49 1,53 >	8	12
[1,53 1,57 >	4	16
[1,57 1,61 >	5	21
[1,61 1,65 >	4	25
[1,65 1,69 >	5	30
Total	**30**	

Coeficiente de Assimetria de Pearson:

$$A = \frac{\overline{x} - M_o}{s}$$

$$A = \frac{1,566 - 1,51}{0,07} = 0,8$$

Distribuição com assimetria moderada.

Exemplo - Curtose

- No exemplo das alturas dos 30 alunos classifique a distribuição quanto a curtose.

Intervalo de Classe Alturas (m)	Número de alunos f_i	f_{ac}
[1,45 1,49 >	4	4
[1,49 1,53 >	8	12
[1,53 1,57 >	4	16
[1,57 1,61 >	5	21
[1,61 1,65 >	4	25
[1,65 1,69 >	5	30
Total	**30**	

$$P_{10} = \frac{10(30)}{100} = 3 \qquad P_{25} = \frac{25(30)}{100} = 7,5$$

$$P_{75} = \frac{75(30)}{100} = 22,5 \qquad P_{90} = \frac{90(30)}{100} = 27$$

Calcula-se, C_{10}, C_{25},C_{75} e C_{90} :

$$C_{10} = 1,45 + \left(\frac{3-0}{4}\right).0,04 = 1,48$$

$$C_{25} = 1,49 + \left(\frac{7,5-4}{8}\right).0,04 = 1,5075$$

$$C_{75} = 1,61 + \left(\frac{22,5-21}{4}\right).0,04 = 1,625$$

$$C_{90} = 1,65 + \left(\frac{27-25}{5}\right).0,04 = 1,666$$

Exemplo - Curtose

- No exemplo das alturas dos 30 alunos classifique a distribuição quanto a curtose.

Intervalo de Classe Alturas (m)	Número de alunos f_i	f_{ac}
[1,45 1,49 >	4	4
[1,49 1,53 >	8	12
[1,53 1,57 >	4	16
[1,57 1,61 >	5	21
[1,61 1,65 >	4	25
[1,65 1,69 >	5	30
Total	**30**	

Calcula-se o coeficiente de curtose:

$$C = \frac{C_{75} - C_{25}}{2(C_{90} - C_{10})}$$

$$= \frac{1,625 - 1,5075}{2(1,666 - 1,48)}$$

$$= 0,3159$$

Como C > 0,263,
a distribuição é platicúrtica.