Introdução à Bioestatística

Marcelo Goulart Correia

Instituto Nacional de Cardiologia

December 1, 2015

- As medidas de posição mais utilizadas são:
  - Amplitude
  - Desvio padrão
  - Variância
  - Coeficiente de variação
  - Erro padrão da média
  - Intervalo interquartílico

- Amplitude
  - É a subtração do maior valor observado

$$\boxed{A_{mp} = Maior_{valor} - Menor_{valor}} \tag{1}$$

- Exemplo:

Categorias	x_i	n_i	f_i	N_i	F_i
130cm |- 140cm	135cm	2	0,10	2	0,10
140cm |- 150cm	145cm	3	0,15	5	0,25
150cm |- 160cm	155cm	7	0,35	12	0,60
160cm |- 170cm	165cm	4	0,20	16	0,80
170cm |- 180cm	175cm	4	0,20	20	1,00
Total		20			

$$A_{mp} = 180cm - 130cm$$

$$A_{mp} = 50cm$$

- Pontos importantes sobre o uso da amplitude:
  - Fácil de calcular
  - Não utiliza todas as observações
  - Sofre influência de valores extremos

- Variância
  - É a média do quadrado das diferenças entre os valores observados em relação à média aritimética.

$$\boxed{S^2 = \frac{1}{n} * \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2}$$

- x_i = I-ésimo valor observado
- $\bar{x}$ = Média aritimética dos valores observados

- Exemplo:

Categorias	x_i	n_i	f_i	N_i	F_i	$(x_i - \bar{x})^2$
130cm |- 140cm	135cm	2	0,10	2	0,10	506,25
140cm |- 150cm	145cm	3	0,15	5	0,25	156,25
150cm |- 160cm	155cm	7	0,35	12	0,60	6,25
160cm |- 170cm	165cm	4	0,20	16	0,80	56,25
170cm |- 180cm	175cm	4	0,20	20	1,00	306,25
Total		20				

$$S^2 = \frac{1}{20} * (506,25 * 2 + 156,25 * 3 + 6,25 * 7 + 56,25 * 4 + 306,25 * 4)$$

$$S^2 = 148,75$$

- Desvio Padrão
  - Torna a medida da variância com a mesma magnitude das medidas observadas

$$\boxed{S = \sqrt{S^2}} \tag{3}$$

- S^2 = Variância calculada

- Exemplo:

Categorias	x_i	n_i	f_i	N_i	F_i	$(x_i - \bar{x})^2$
130cm |- 140cm	135cm	2	0,10	2	0,10	506,25
140cm |- 150cm	145cm	3	0,15	5	0,25	156,25
150cm |- 160cm	155cm	7	0,35	12	0,60	6,25
160cm |- 170cm	165cm	4	0,20	16	0,80	56,25
170cm |- 180cm	175cm	4	0,20	20	1,00	306,25
Total		20				

$$S^2 = \frac{1}{20} * (506,25 * 2 + 156,25 * 3 + 6,25 * 7 + 56,25 * 4 + 306,25 * 4)$$

$$S^2 = 148,75$$
$$S = \sqrt{148,75} = 12,20cm$$

- Pontos importantes sobre o uso da variância e do desvio padrão:
  - Ambas as medidas são correlacionadas
  - A expressão $\bar{x} \pm 2S$ indica que pelo menos 75% das observações estão incluídas nesse intervalo. Caso a distribuição seja Normal, esse número pode chegar aproximadamente à 95%.
  - Não utilizar essas medidas quando a mediana for utilizada como medida de tendência central

- Exercício

- Coeficiente de variação
  - É uma medida que permite verificar a dispersão dos dados em torno da média aritimética

$$\boxed{CV = \frac{S_X}{\bar{x}} * 100} \tag{4}$$

- S_X = Desvio padrão da variável X
- $\bar{x}$ = Média da variável X

- Exemplo:

Categorias	x_i	n_i	f_i	N_i	F_i	$(x_i - \bar{x})^2$
130cm |- 140cm	135cm	2	0,10	2	0,10	506,25
140cm |- 150cm	145cm	3	0,15	5	0,25	156,25
150cm |- 160cm	155cm	7	0,35	12	0,60	6,25
160cm |- 170cm	165cm	4	0,20	16	0,80	56,25
170cm |- 180cm	175cm	4	0,20	20	1,00	306,25
Total		20				

$$S^2 = \frac{1}{20} * (506,25*2 + 156,25*3 + 6,25*7 + 56,25*4 + 306,25*4)$$
$$S^2 = 148,75$$
$$S = \sqrt{148,75} = 12,20cm$$

$$CV = \frac{12,20}{157,5} = 7,75\%$$

- Pontos importantes sobre o uso do coeficiente de variação:
  - Só pode ser utilizada em variáveis com valores positivos
  - Os resultados são sensíveis ao acréscimo de novos dados
  - Os resultados são invariantes à mudança de escala

- Erro padrão da média
  - Calcular a dispersão das médias de diferentes amostras (da mesma população) com mesmo tamanho

$$\boxed{EPM = \frac{S}{\sqrt{n}}} \tag{5}$$

- S = Desvio padrão calculado
- n = Tamanho da amostra

- Exemplo:

Categorias	x_i	n_i	f_i	N_i	F_i	$(x_i - \bar{x})^2$
130cm |- 140cm	135cm	2	0,10	2	0,10	506,25
140cm |- 150cm	145cm	3	0,15	5	0,25	156,25
150cm |- 160cm	155cm	7	0,35	12	0,60	6,25
160cm |- 170cm	165cm	4	0,20	16	0,80	56,25
170cm |- 180cm	175cm	4	0,20	20	1,00	306,25
Total		20				

$$S^2 = \frac{1}{20} * (506,25 * 2 + 156,25 * 3 + 6,25 * 7 + 56,25 * 4 + 306,25 * 4)$$
$$S^2 = 148,75$$
$$S = \sqrt{148,75} = 12,20cm$$
$$EPM = 12,20/\sqrt{20} = 2,73$$

- Pontos importantes sobre o uso do erro padrão da média:
  - Medida essencial para o cálculo do intervalo de confiança
  - Seu uso tem melhor aplicação em variáveis que seguem distribuição normal
  - Desvio padrão $\neq$ Erro padrão da média

- Por que?
  - Desvio padrão –> Descrever variabilidade dos dados amostrais
  - Erro padrão –> Descreve a imprecisão de uma medida
  - Cuidado com quem publica o erro padrão ao invés do desvio padrão

- Intervalo interquartílico
  - É uma medida de variabilidade baseada na divisão dos dados em quartis

$$\boxed{IQR = Q_3 - Q_1} \tag{6}$$

- Q_1 = Quartil 1 ou Percentil 25
- Q_3 = Quartil 3 ou Percentil 75

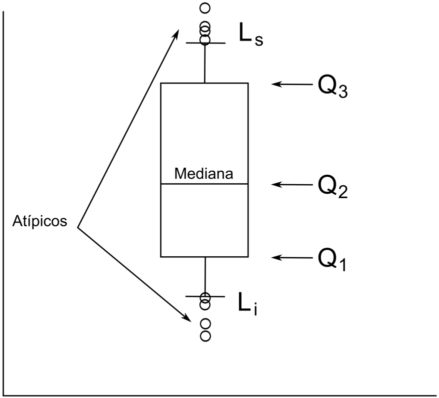
- Exemplo:

Categorias	x_i	n_i	f_i	N_i	F_i	$(x_i - \bar{x})^2$
130cm |- 140cm	135cm	2	0,10	2	0,10	506,25
140cm |- 150cm	145cm	3	0,15	5	0,25	156,25
150cm |- 160cm	155cm	7	0,35	12	0,60	6,25
160cm |- 170cm	165cm	4	0,20	16	0,80	56,25
170cm |- 180cm	175cm	4	0,20	20	1,00	306,25
Total		20				

$$P_{25} = 140cm + \frac{20*25/100-2}{3} * 10$$
$$P_{25} = 150,00cm$$

$$P_{75} = 160cm + \frac{20*75/100-12}{4} * 10$$
$$P_{75} = 167,50cm$$
$$IQR = 167,50 - 150,00 = 17,5cm$$

- Pontos importantes sobre o uso do intervalo interquartílico:
    - Melhor utilizada quando a distribuição dos dados não seja normal
    - Mediana é a medida de tendência central desse intervalo
    - Auxilia no detecção de *outliers*



- L_s ou $L_i = \pm 1,5 * IQR$

- Exercício

- Assimetria

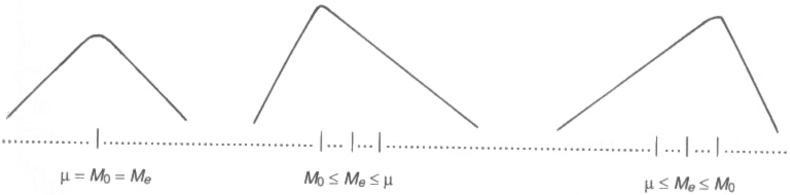
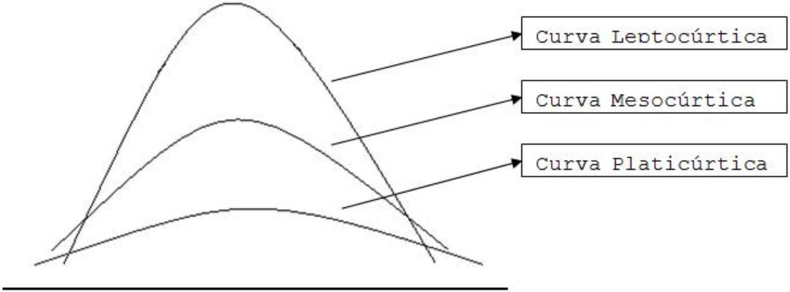


Figura 1: Distribuição simétrica

Figura 2: Assimetria positiva

Figura 3: Assimetria negativa

- Curtose


Curva Leptocúrtica
Curva Mesocúrtica
Curva Platicúrtica