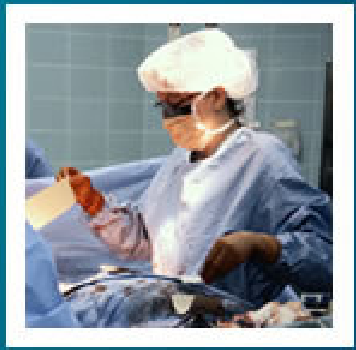
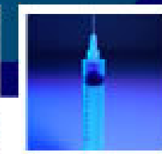
Bioestatística
Aula 3
Medidas de Dispersão ou Variabilidade



Prof. Tiago A. E. Ferreira

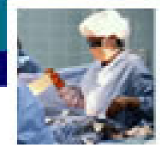
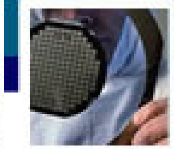
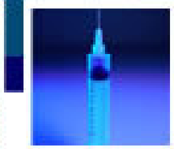
Aspectos Gerais





- **Dispersão** ou **Variabilidade** nada mais é do que a diferença observada entre os valores de um conjunto de dados
  - A diversidade de valores contidos em dados observados irá depender do fenômeno gerador dos próprios dados.
    - Esta diversidade será então uma característica da população em estudo, sendo responsável pela dispersão ou variabilidade dos dados.
  - Questão: Como medir esta diversidade?

Amplitude total

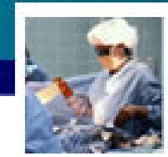




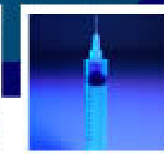
- Uma das formas mais diretas e simples de medir a dispersão é através da **Amplitude Total (AT)** do conjunto de dados.
  - Como já visto, sendo $X=\{x_1, x_2, ..., x_n\}$ com $\{x_1 < x_2 < ... < x_n\}$, então:
    - Para dados simples: $AT = x_n - x_1$
      - Para dados agrupados em tabelas: $AT = Lsr_m - Lir_1$
      - onde:
        - » Lsr_m = limite superior real da classe m (última Classe)
        - » Lir_b = limite inferior real da classe 1

Amplitude Total

- Desta forma, quanto maior a dispersão ou variabilidade dos valores no conjunto de dados, maior será a Amplitude Total, e vice-versa.
  - Apesar de sua simplicidade, a AT tem suas limitações:
    - A AT considera apenas dois pontos de todo o conjunto de dados, não sendo sensível aos “n-2” demais valores do conjunto de dados.
    - No caso de dados agrupados, se os limites forem abertos, como calcular AT?

Problemas da Amplitude Total

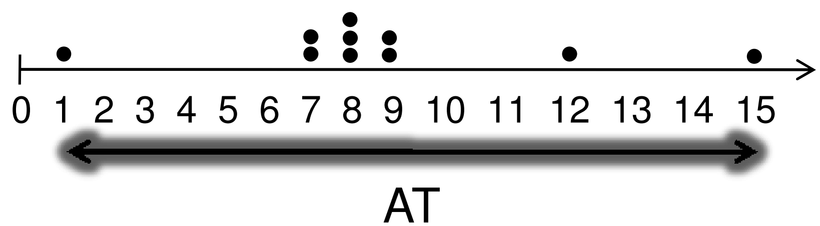




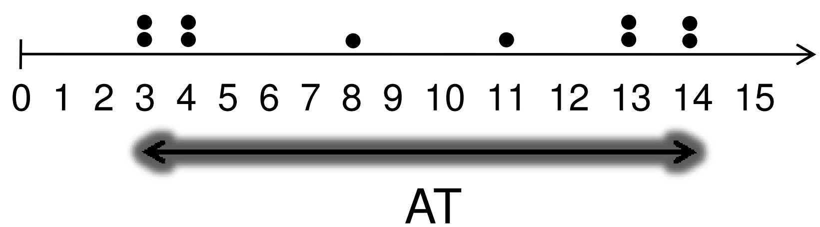
- Dado os dois conjuntos exemplos:
  - A={1,7,7,8,8,8,9,9,12,15}
  - B={3,3,4,4,8,11,13,13,14,14}
- Segundo a medida AT, a dispersão é:
  - Dispersão[A]=AT_A=15-1=14
  - Dispersão[B]=AT_B=14-3=11
    - Será que estes resultados refletem a realidade?

Problemas da Amplitude Total

- Conjunto A: $AT_A = 14$



- Conjunto B: $AT_B = 11$



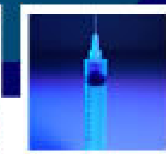
Soma dos Desvios Simples

- Uma alternativa à AT é calcular a diferença entre os dados do conjunto, dois a dois.
  - Este processo necessita de todas as combinações n 2 a 2.
  - Porém estas diferenças podem ser calculadas a partir de um valor central como referência, por exemplo a média.
  - Assim, a Soma dos Desvios Simples (SDS) é

$$SDS = \sum_{i=1}^{n} \left(x_i - \bar{x} \right)$$

Soma dos Desvios Simples





- Contudo, já foi visto que, para qualquer conjunto de dados, SDS será sempre nulo!
  - A ideia do SDS está coerente, porém o SDS sempre se anula devido ao fato de que para toda variação negativa (desvios à esquerda da média) existirá uma variação positiva (desvios à direita da média)

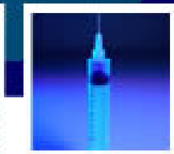
Soma dos Desvios Absolutos

- Desta forma, se todos os desvios forem considerados de forma absoluta, é possível ter a Soma dos Desvios Absolutos (SDA):

$$SDA = \sum_{i=1}^{n} \left| x_i - \bar{x} \right|$$

Desvio Médio



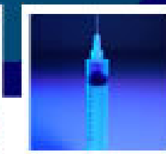


- A partir da Soma dos Desvios Absolutos ainda é possível se calcular a variabilidade em média!
  - Ou seja, em média quanto cada ponto do conjunto de dados está se afastando do valor médio.
- Define-se o Desvio Médio (DM)

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

Soma dos Quadrados dos Desvios

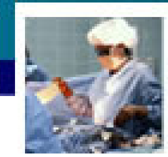


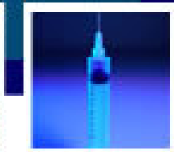


- Uma outra forma de resolver o problema da SDS é através da Soma dos Quadrados dos Desvios (SQD).
  - Uma vez que todo número quadrado é não negativos (≥ 0).
- Define-se SQD como,

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

SDA e SQD, Qual utilizar?

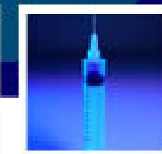




- Tanto a Soma dos Desvios Absolutos com a Soma Quadrada dos Desvios satisfazem os requisitos para serem medidas de dispersão ou variabilidade.
  – Porém a SQD tem maior capacidade de destrinçar os desvios em relação a SDA.
  – Desta forma, a SQD é preferida na grande maioria dos casos!

Variância



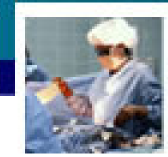


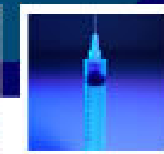
- Também é possível se calcular o Desvio Quadrático Médio, conhecido comumente com **Variância** (Var[x]):

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

  – Esta expressão é a Variância populacional

Desvio Padrão





- A raiz quadrada da variância é o desvio padrão:

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

- O desvio padrão e variância são os dois indicadores de variabilidade mais utilizados

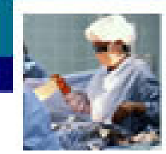
Variância e Desvio Padrão Amostrais

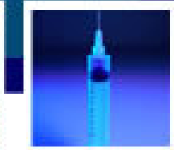
- Quando se trata de amostras extraídas de uma população, a variância e o desvio padrão sofrem uma correção amostral, sendo representados por S^2 e S

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

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

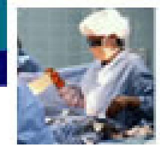
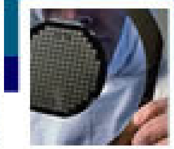
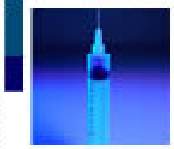
Variância e Desvio Padrão Amostrais





- Na prática, as expressões populacionais e amostrais só terão uma diferença significativa para amostras muito pequenas
  - Tipicamente amostras com menos de 30 elementos (ou valores)

Cálculos Abreviados

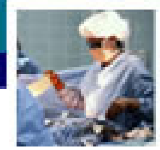
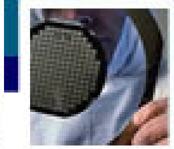
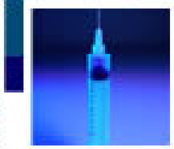
- Variância populacional

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

- Variância Amostral

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

Cálculos Abreviados

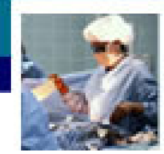
- Desvio Padrão Populacional

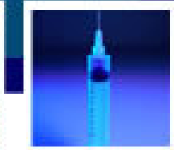
$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}}$$

- Desvio Padrão Amostral

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

Dados Agrupados



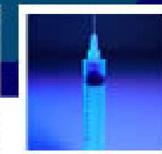


- Variância populacional para dados agrupados em classes homogêneas

$$\sigma^2 = \left[\frac{\sum_{i=1}^{n} (fa_i \cdot v_i)^2}{n} - \left(\frac{\sum_{i=1}^{n} fa_i \cdot v_i}{n} \right)^2 \right] \cdot c^2$$

Dados Agrupados



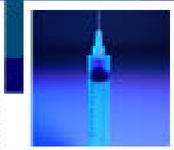


- Variância populacional para dados agrupados em classes não homogêneas:

$$\sigma^2 = \frac{\sum_{i=1}^{n} fa_i \cdot (pm_i - \bar{x})^2}{\sum_{i=1}^{n} fa_i} = \frac{\sum_{i=1}^{n} fa_i \cdot (pm_i - \bar{x})^2}{n}$$

Medida de Variabilidade Normalizada





- Como comparar duas grandezas diferentes?
  - Matematicamente (e até fisicamente) não há sentido em se comparar duas grandezas distintas
    - Como comparar, por exemplo, Kg com ml, ou seja, massa com volume?
    - Para se poder afirmar que uma dada grandeza é maior ou menor, ou ainda igual, a outra grandeza ambas devem ter a mesma unidade ou serem adimensionais.

Medida de Variabilidade Normalizada

- Exemplo: Dada a tabela de comprimento e peso

Recém-Nascido	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Comprimento (cm)	52	48	45	49	51	54	47	50	46	51
Peso (g)	3300	3200	2950	3150	3350	3450	2900	3300	3150	3250

- É possível calcular:
  - $\overline{\text{Comprimento}}$ = 49,3 cm
  - $\overline{\text{Peso}}$ = 3200g
  - $s_H^2 = 7{,}24\ cm^2$ $\qquad$ $s_P^2 = 26500{,}58\ g^2$
  - $s_H = 2{,}69\ cm$ $\qquad$ $s_P = 162{,}79\ g$
- É possível, observando os resultados, afirmar qual grandeza tem a maior variabilidade?

Medida de Variabilidade Normalizada

- Observe que as grandezas Peso (gramas) e Comprimentos (centímetros) têm unidades distintas!
  - Além deste fato, todos os cálculos são baseados em um fator de referência, no caso a média!
    - Assim, toda medida de variação vista tem como referência básica o valor da média.
    - Logo toda comparação deve ser feita em relação a média da grandeza.

Coeficiente de Variação

- É possível definir o coeficiente de variação (grandeza admensional):
  - Caso populacional:

$$cv = \frac{\sigma}{\bar{x}}$$

  - Caso amostral:

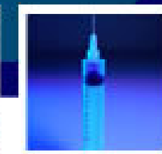
$$cv = \frac{s}{\bar{x}}$$

Coeficiente de Variação

- Desta forma, para o exemplo anterior:
  - $cv_H = 0{,}545$
  - $cv_P = 0{,}0509$
    - Com a utilização do coeficiente de variação é possível comparar qual grandeza teve efetivamente uma variabilidade maior!
    - Este coeficiente de variação também nos dá alguma noção de estabilidade, ou seja, qual é a ordem de grandeza da variabilidade em relação a média
      - Quanto mais cv se aproximar de zero maior a estabilidade do sistema!

Medidas de Assimetria





- Como as medidas estudadas até então é possível determinarmos um valor típico (ou central) dos dados e o comportamento dos demais pontos em torno deste valor típico.

Medidas de Assimetria

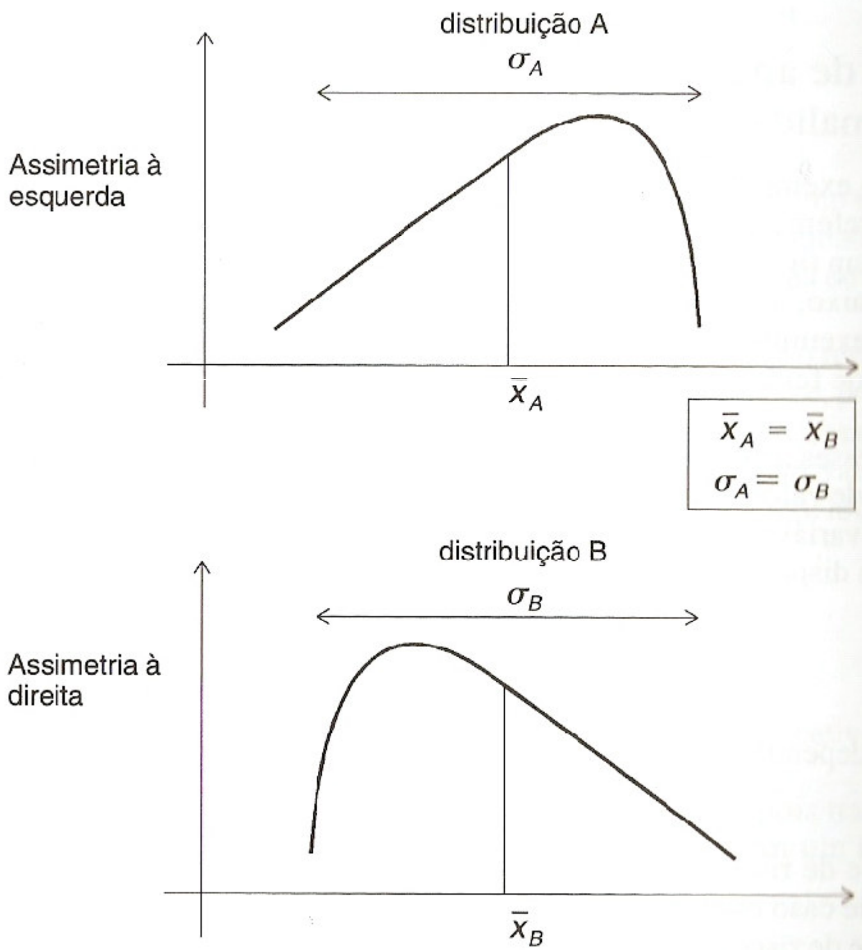
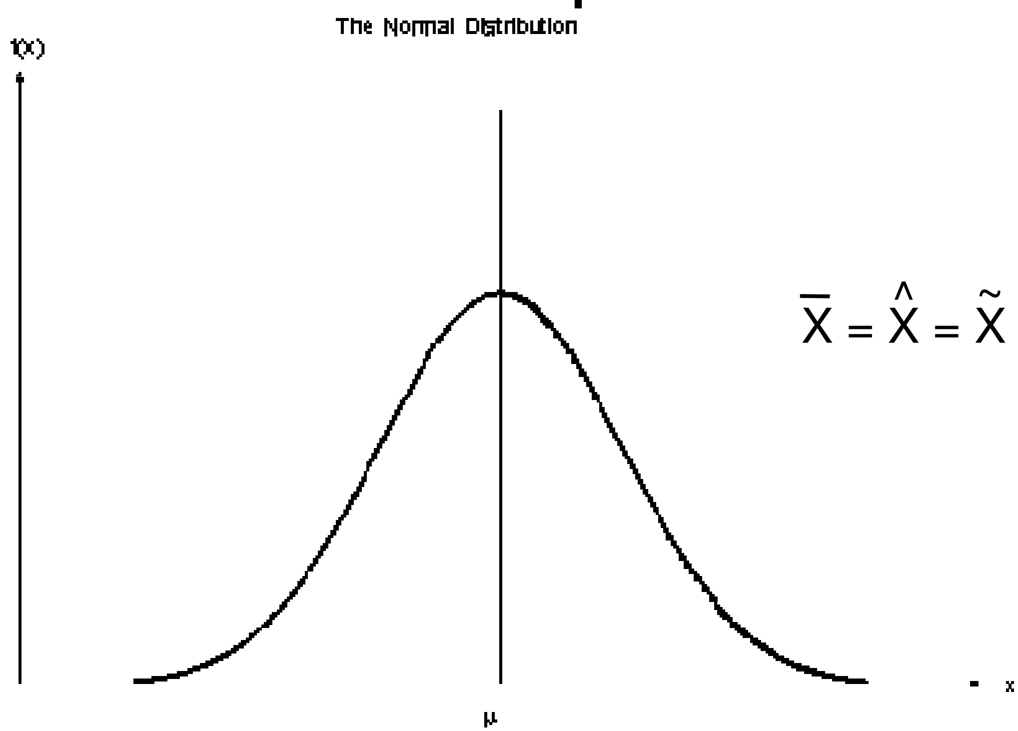


Fig. 4.7 Distribuições à esquerda e à direita.

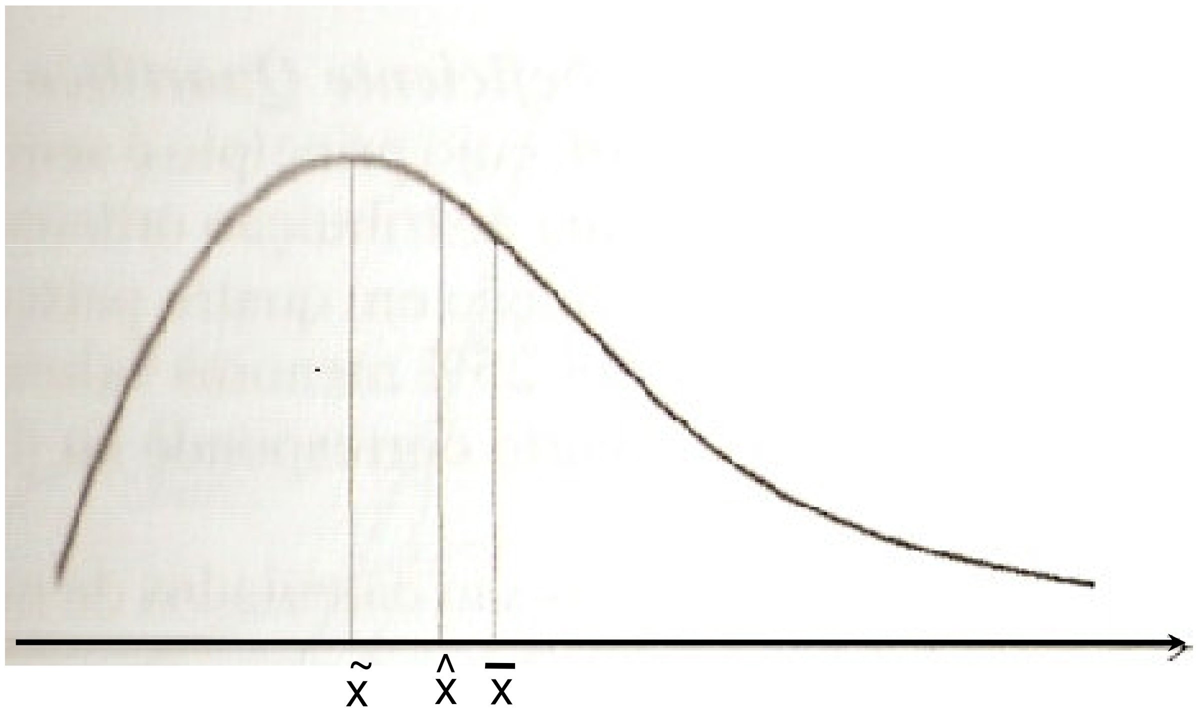
Medida de Assimetria

- Defini-se simetria a identidade de comportamento de curva a ambos os lados de um eixo ou plano de simetria

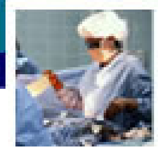
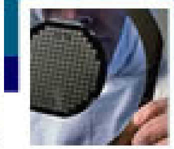
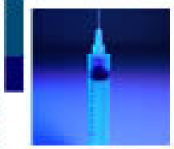


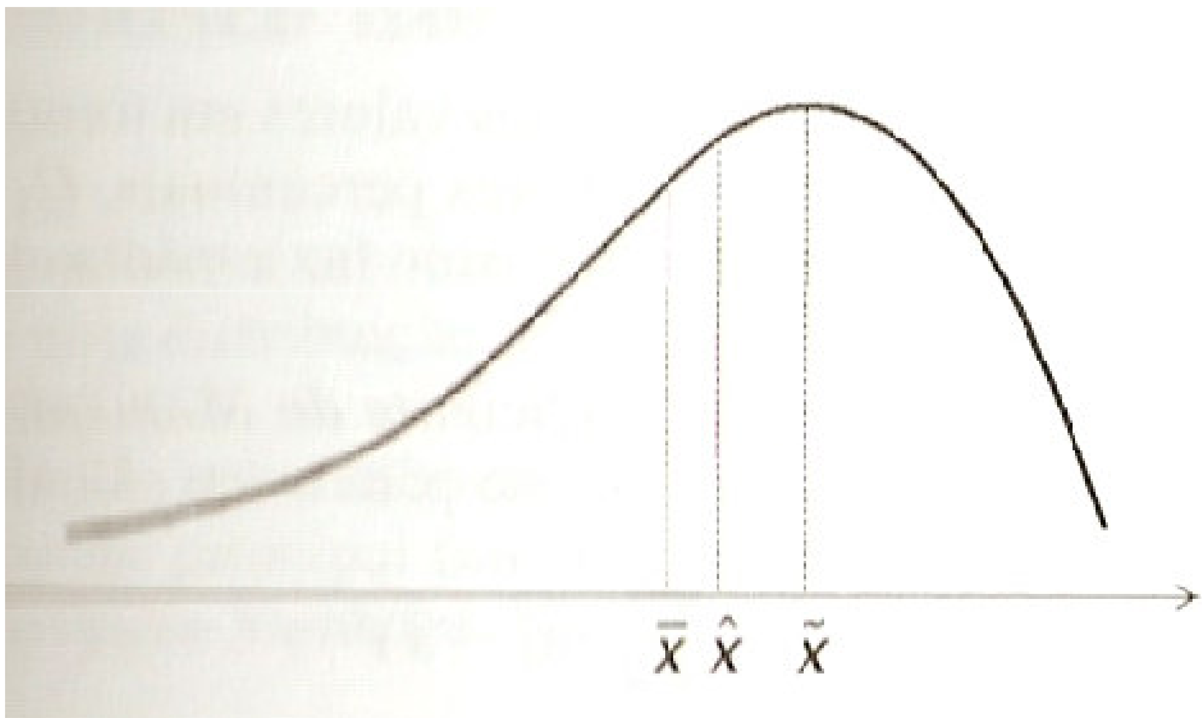
$$\bar{X} = \hat{X} = \tilde{X}$$

Assimetria à Direita

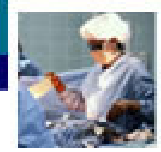
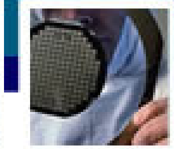
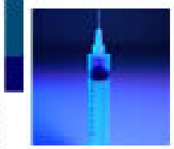


Assimetria à Esquerda



Coeficiente de Assimetria

- Como medir a assimetria de um conjunto de dado?
  - **Coeficiente de Assimetria de Pearson**
    - Quanto maior a assimetria maior a diferença entre a média e a moda
      - Assimetria = Média – Moda
  - Este coeficiente pode ser melhorado se dividido pelo desvio padrão, tornando-se adimensional

$$P = \frac{\bar{x} - \tilde{x}}{\sigma}$$

Coeficiente de Assimetria

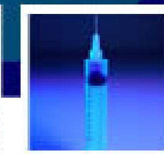
- Dado:

$$P = \frac{\bar{x} - \tilde{x}}{\sigma}$$

  – É interessante notar:
    - Distribuição simétrica, P = 0 (média = moda)
    - Distribuição é à esquerda, P < 0 (média < moda)
    - Distribuição é à direita, P > 0 (média > moda)
    - Quanto maior a assimetria, maior |P|

Coeficiente Quartílico de Assimetria (Q)





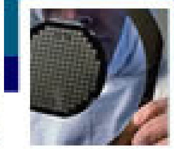
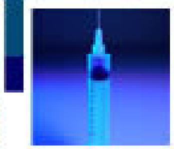
- O coef. Quartílico de assimetria (Q) é calculado a partir dos valores dos quartis:

$$Q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

$$ou$$

$$Q = \frac{Q_3 - 2 \cdot Q_2 - Q_1}{Q_3 - Q_1}$$

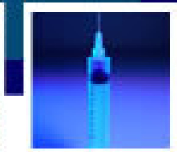
Coeficiente do Momento de Assimetria

- O coeficiente do momento de assimetria :

$$a_3 = \frac{\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{n}}{\sigma^3}$$

Curtose

- O conceito de curtose é relativo ao “achatamento” ou “alongamento”de uma dada distribuição.
  - Distribuições achatadas: **platicúrtica**
  - Distribuições normais: **mesocúrticas**
  - Distribuições Alongadas: **leptocúrtica**

Curtose

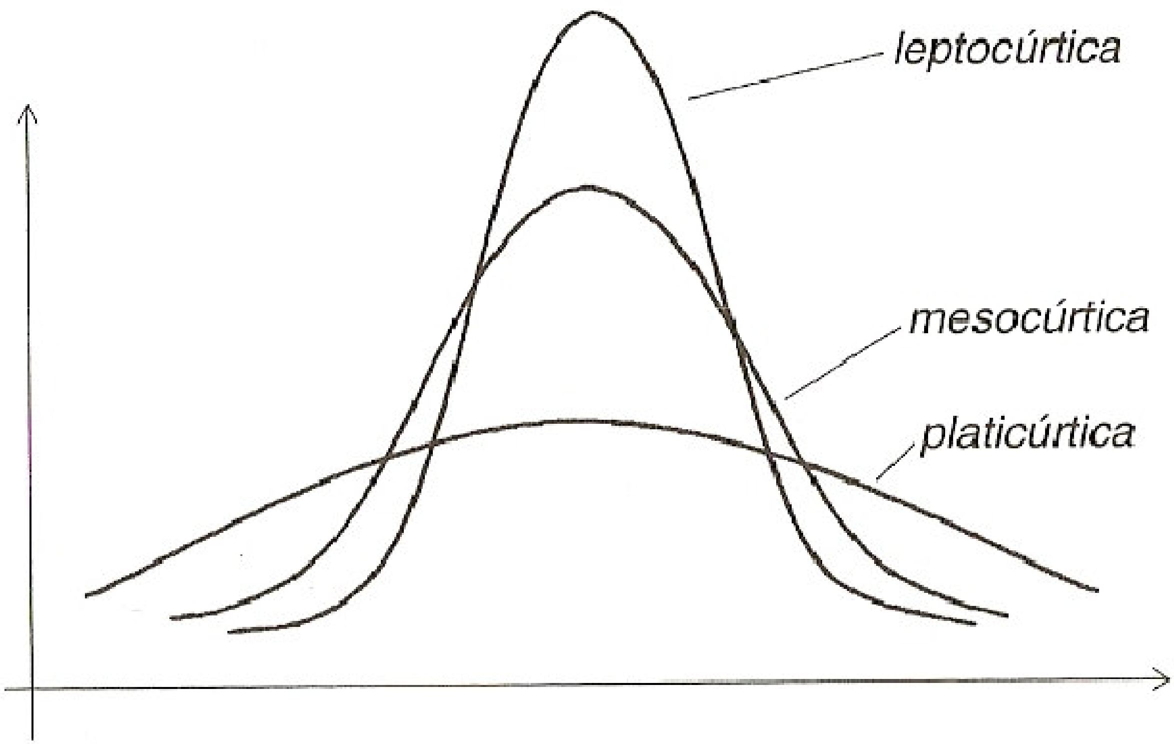
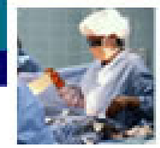
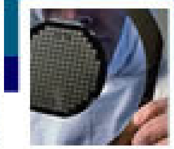
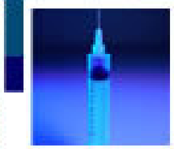


Fig. 4.9 Tipos de distribuições quanto à curtose.

Referências

- Livro Texto: Bioestatística. Teoria e Computacional (Héctor G. Arango). Guanaba
  - Seção 2, 3 e 4 – Capítulo 4
- Leituras Complementares:
  - Livro: Bioestatística. Princípios e Aplicações (Sidia M. Callegari-Jaqcques)
    - Capítulo 3.