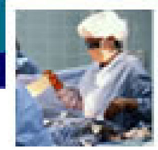
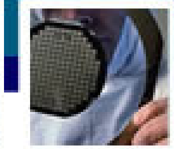
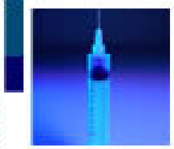
Bioestatística
Aula 2
Medidas de Tendência Central

Prof. Tiago A. E. Ferreira

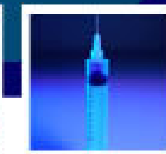
Introdução

- Como já visto na aula anterior, estamos interessados em estudar algum fenômeno sobre uma dada população.
  - Dada uma população, gera-se uma amostra para a caracterização das informações da população!
    - Assim, procura-se, a partir da amostra, características centrais da população.
    - **Medidas de tendência Central** são uma primeira caracterização dos conjuntos populacionais ou amostrais

Média Aritmética

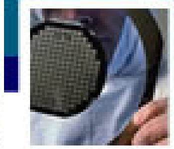
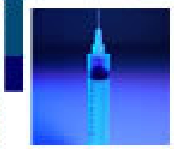




- Dado um conjunto de **n** valores da variável aleatória **X**, $X=\{x_1, x_2, ..., x_n\}$, a média aritmética de X, $\overline{X}$ é

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Média Aritmética

- Exemplo:
  - Dado um estudo foram medidas as concentrações percentuais GC de trechos de igual tamanho de seqüenciamentos genéticos, dados por:

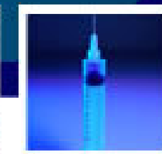
seqüência	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
GC (%)	56,3	45,5	46,7	49,8	54,2	42,1	56,6	44,7	61,2

  - Pergunta-se: Qual a concentração percentual média de GC do Estudo?

$$\overline{sequência} = \frac{1}{9}\sum_{i=1}^{9} sequência_1 = 50{,}79$$

Média Ponderada





- Dado que um certo experimento gere os seguintes dados:

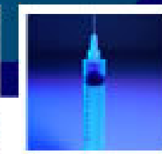
Valores	Frequência
2	11
4	24
5	43
Total:	78

  – Pode-se notar que foram gerados 78 dados, ou seja, n=78. Assim, a média aritmética seria:

$$\bar{x} = \frac{\overbrace{2+2+\cdots 2}^{\text{11 vezes}}+\overbrace{4+4+\cdots 4}^{\text{24 vezes}}+\overbrace{5+5+\cdots+5}^{\text{43 vezes}}}{78} = 4{,}269$$

Média Ponderada





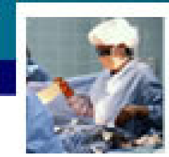
- No exemplo anterior, é equivalente:

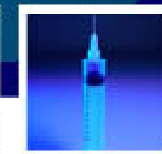
$$\bar{x} = \frac{11 \cdot 2 + 24 \cdot 4 + 43 \cdot 5}{11 + 24 + 43} = 4{,}269$$

  – Desta forma, a frequência de cada valor representa a importância que cada fator tem no calculo da média

$$\bar{x} = \frac{\sum_{i=1}^{qtd.valores} x_i \cdot f_i}{\sum_{j=1}^{qtd.valores} f_j} = \frac{\sum_{i=1}^{qtd.valores} x_i \cdot f_i}{n}$$

Média Ponderada





- Desta forma, é possível adotar pesos para os valores dos dados.
  - Quanto maior a importância de uma dada observação (ou valor) maior o seu peso.
  - Define-se a **Média Ponderada**

$$\overline{x}_p = \frac{\sum_{i=1}^{n} x_i \cdot p_i}{\sum_{j=1}^{n} p_j}$$

- onde $\mathbf{p_i}$ é o peso para o valor **i**

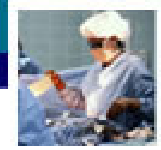
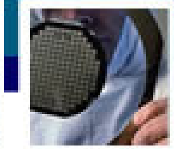
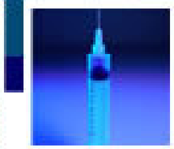
Propriedades da Média Aritmética

- A soma dos desvios de um conjunto X, com N números, em relação a sua média aritmética é sempre nula!

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = 0$$

- Prova Matemática?

Propriedades da Média Aritmética

- A soma dos quadrados dos desvios de um conjunto de números xi, em relação a qualquer número A, é um mínimo quando A = $\overline{x}$ e somente neste caso

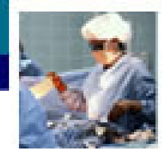
$$\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2 = mínimo$$

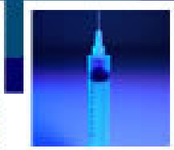
Propriedades da Média Aritmética

- Se f_1 números têm média m_1, f_2 números têm média m_2, ..., f_k números têm média m_k, então a média global pode ser calculada por:

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i \cdot f_i}{\sum_{j=1}^{k} f_j} = \frac{\sum_{i=1}^{k} m_i \cdot f_i}{n}$$

Propriedades da Média Aritmética



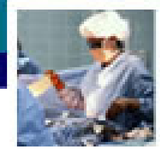
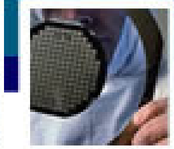
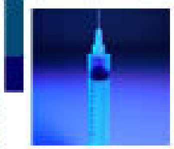


- A média aritmética de números é também igual à média de cada um destes números menos (ou mais) uma constante, somada depois a essa mesma constante.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} \left(x_i \mp A\right)}{n} \pm A$$

  – Esta propriedade também está afirmando que somar ou subtrair uma constante a um conjunto de valores numéricos, estamos somando ou subtraindo esta constante ao valor médio do conjunto!

Média Geométrica

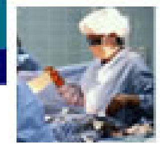
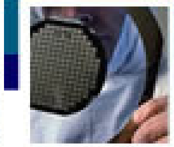
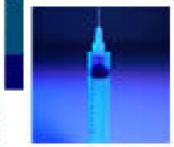
- A Média Geométrica é definida por,

$$G = \left[\prod_{i=1}^{n} x_i \right]^{\frac{1}{n}} \quad ou \quad G = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}$$

- Aplicando o logaritmo na base *b* a esta expressão, é possível mostrar que,

$$G = b^{\overline{\log_b x}} \quad onde \quad \overline{\log_b x} = \frac{\sum_{i=1}^{n} \log_b x_i}{n}$$

Exemplo de Média Geométrica

- A média geométrica é utilizada para casos onde a escala de referência é não linear, em particular, exponencial.
  - Imagine que uma experiência com 25 indivíduos consistiu em determinar a **Concentração Mínima Inibitória (CMI)** de um certo antibiótico.

CMI (μg/ml)	Número de Indivíduos
1,00000	1
0,50000	2
0,25000	6
0,12500	4
0,06250	3
0,03125	9
Total	25

Exemplo de Média Geométrica

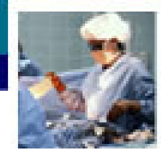
- Observe que escala não é linear, logo deve-se aplicar a média geométrica.

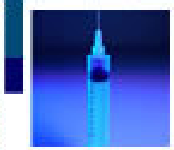
$$\overline{\log x} = \frac{1 \cdot \log 1 + 2 \cdot \log 0{,}5 + 6 \cdot \log 0{,}25 + \cdots + 9 \cdot \log 0{,}03125}{25} = -0{,}9994$$

- Logo,

$$G = 10^{\overline{\log x}} = 10^{-0{,}9994} \therefore G = 0{,}1001$$

Média Harmônica

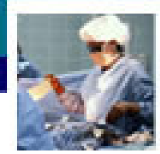


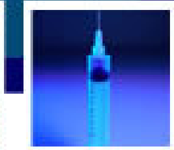


- A média harmônica (h) é dada por:

$$H = n \cdot \left[\sum_{i=1}^{n} (x_i)^{-1} \right]^{-1} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

Mediana





- A mediana é outra forma para selecionar um valor representativo e central de uma conjunto numérico.
  - Dado que um conjunto de observações esteja ordenado (crescente ou decrescente), o valor que ocupar a posição equidistante dos extremos é um valore representativo central deste conjunto.
    - **Este valor é a mediana:** $\hat{x}$

Mediana

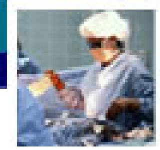
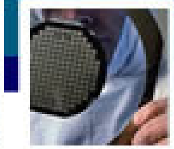
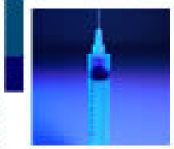
- Matematicamente:

$$X = \{x_1, x_2, \ldots, x_n\}$$

$$com,\ \{x_1 < x_2 < \cdots < x_n\}$$

$$tem-se,\quad \hat{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & se\, n\; for\; ímpar \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & se\, n\; for\; par \end{cases}$$

Mediana

- Exemplo:
  - Dado os conjuntos de pontos:
    - X={2, 8, 3, 5, 4}
    - Y={2, 8, 3, 5, 4, 0}

    Calcule as respectivas medianas
  - Resposta:

$$\hat{x} = 4$$

$$\hat{y} = 3{,}5$$

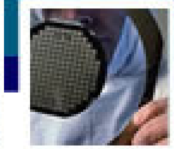
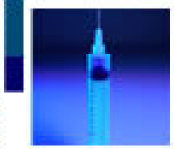
Mediana

- A medida da Média é influenciada diretamente pelos valores do conjunto.
  - Ou seja, se houver um valor atípico (muito grande ou muito pequeno) a medida da média poderá ser bastante modificada
- Já o cálculo da mediana não irá depender dos valores extremos
  - Desta forma, a mediana é bem mais robusta quando há valores atípicos nos dados.

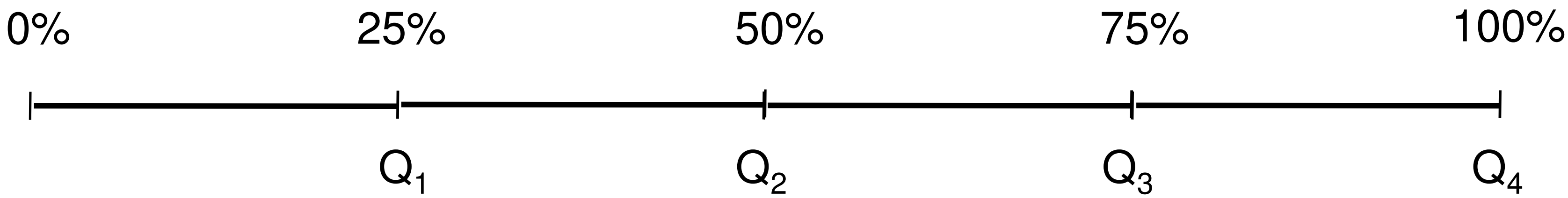
Quartis, Decis e Percentis

- A mediana é uma separatriz de um conjunto
  - 50% dos valores estão abaixo da mediana e 50% estão acima!
  - Ou seja, o conjunto foi dividido em 2 parte
- Podem ser definidas outras separtrizes
  - Tal que o conjunto seja dividido em:
    - **4 partes – Quartis**
    - **10 partes – Decis**
    - **100 partes - Percentis**

Quartis, Decis e Percentis

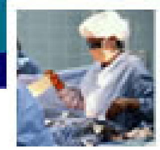
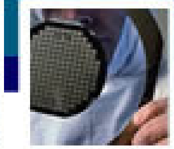
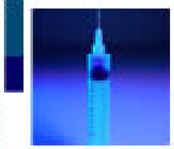




- Quartis: Q_1, Q_2, Q_3 e Q_4



- De forma análoga:
  - Decis: D_1, D_2, ..., D_9, D_{10}
  - Percentis: P_1, P_2, ..., P_{99}, P_{100}

Determinando as Separatrizes

- Forma geral para a determinação de separatrizes:

$$R_i = (N+1)\frac{i}{C}, \quad com\, i = 1,2,\ldots,C-1$$

  – Onde:
    - Ri é a ordem do número que separa a i-ésima separatriz
    - i é a separatriz desejada
    - C é o número de divisões do conjunto

Determinando as Sepatarizes

- Nos casos específicos:
  - Quartis:

$$R_i = (n+1)\frac{i}{4}$$

    - Com i=1, 2 e 3 (o quarto quartil é o último elemento!)
  - Decis:

$$R_i = (n+1)\frac{i}{10}$$

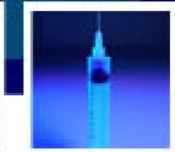
    - Com i=1, 2, ..., 9 (o décimo decil é o último elemento!)
  - Percentis:

$$R_i = (n+1)\frac{i}{100}$$

    - Com i=1, 2, ..., 99 (o centésimo percentil é o último elemento!)

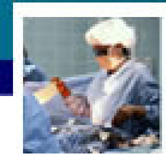
Exemplos:

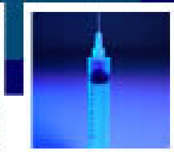




- Dado o conjunto de dados:
  - X={2,2,3,5,6,6,6,7,7,7,7,8,8,9,10,12,12,15,18, 25,27,28,29,30,33}
  - Determine os Quartis, o Decil 9 e o Percentil 95.

Moda





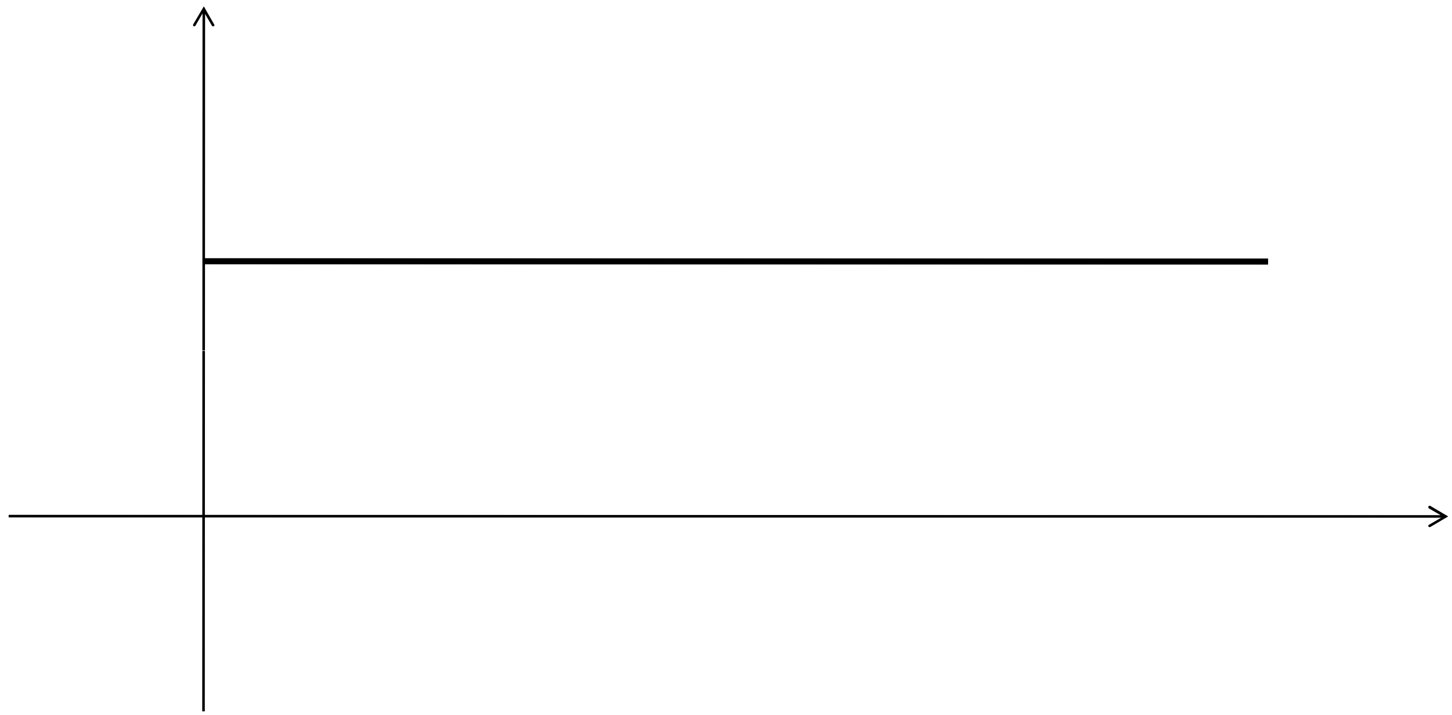
- A **Moda** é outra forma para selecionar um valor representativo e central de uma conjunto numérico.
  - A moda de um conjunto de dados é o seu valor mais frequênte
    - Ou seja, é o elemento que aparece o maior número de vezes.
    - Assim, dado um conjunto de valores, $x_1, x_2, \ldots, x_k$, com as respectivas frequências $f_1, f_2, \ldots, f_k$
    - Então defini-se a moda por:

$$\tilde{x} = x_i \,\middle|\, f_i > f_1, f_2, \ldots, f_k$$

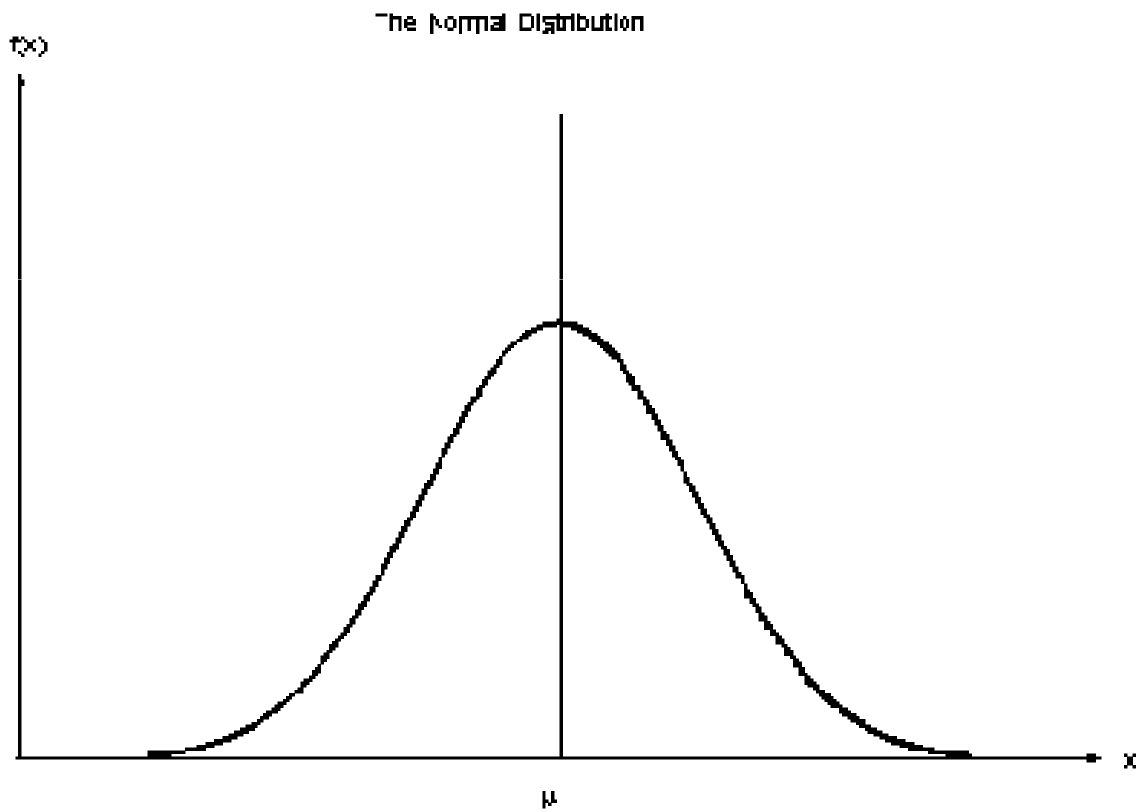
Moda

- Dado um conjunto de dados, este pode gerar distribuições:
  - Amodais
    - Quando todas as frequência são iguais
  - Unimodal
    - Quando há apenas uma frequência máxima
  - Bimodais
    - Quando existem duas frequências máximas
  - Multimodais
    - Quando existem mais de duas frequências máximas

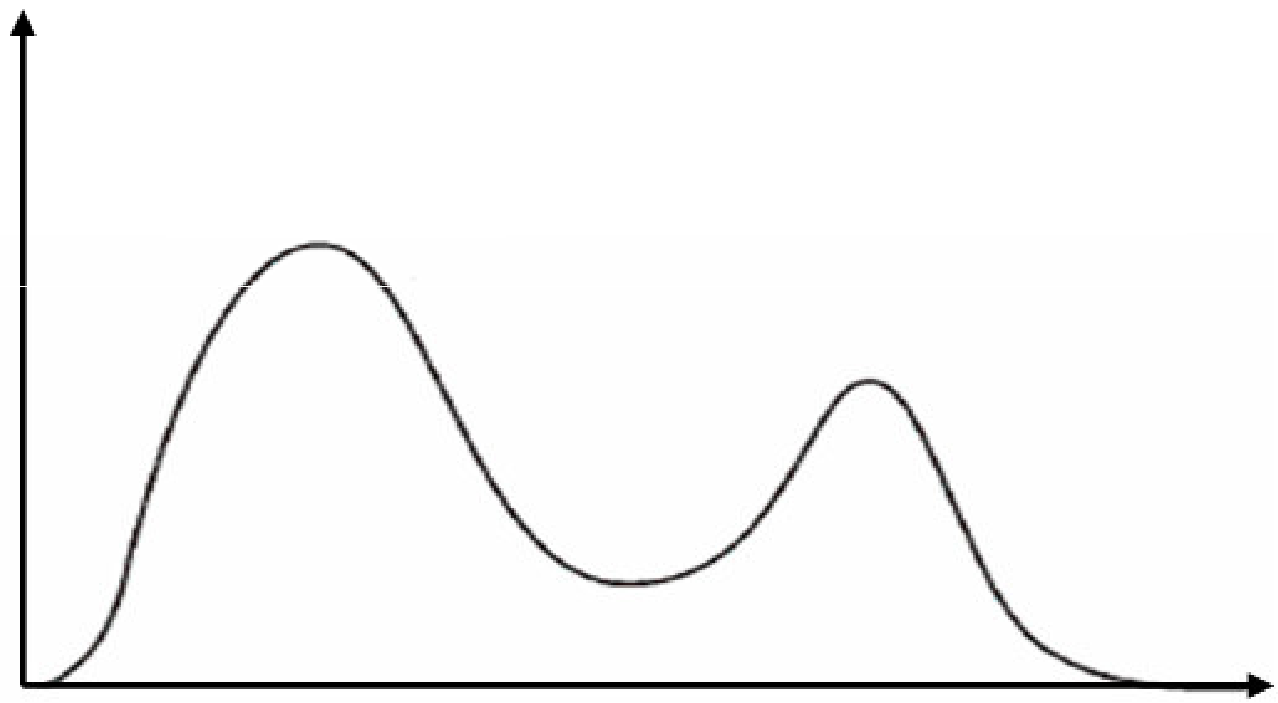
Distribuição Amodal



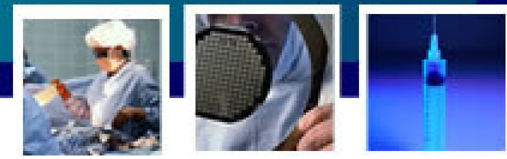
Distribuição Unimodal

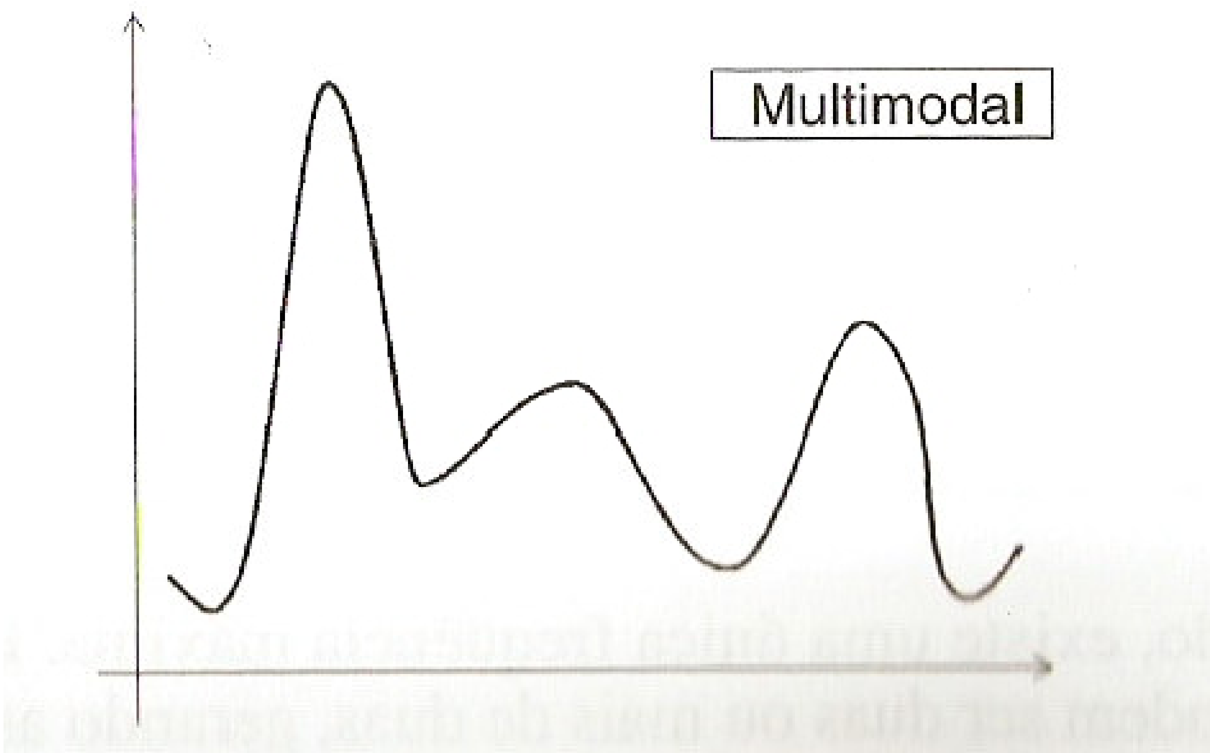


Distribuição Bimodal

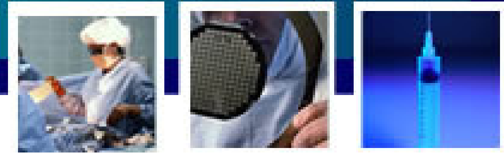


Distribuição Multimodal





Moda e as Demais Medidas



- Imagine um doença que verifica-se na infância e na velhice.
  - Tipicamente a sua distribuição será:

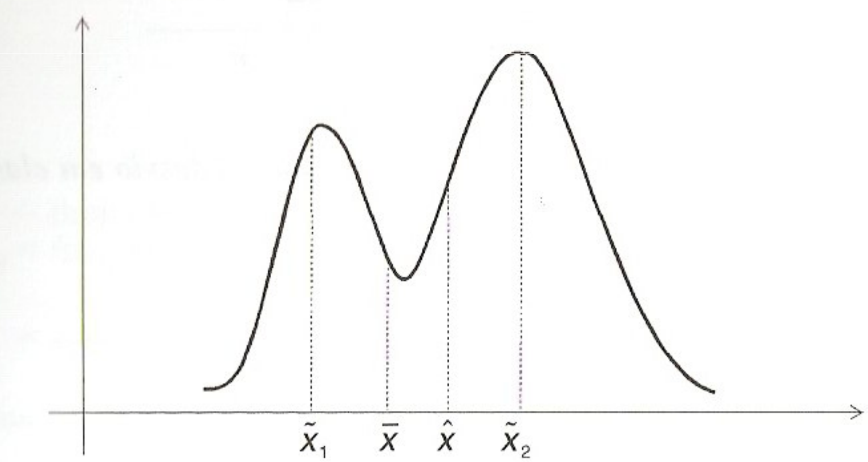
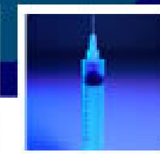


Fig. 4.3 Medidas de tendência central em distribuições bimodais.

  - Como calcular as idades normais da doença?

Tratamento de Dados Agrupados





- Se os dados são obtidos de forma agrupada, como calcular sua médias?
  - Suponha, por exemplo, que tenha-se uma tabela

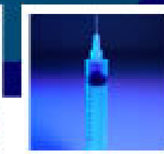
Classes	PM	fa
$Li_1 – Ls_1$	pm_1	fa_1
$Li_2 – Ls_2$	Pm_2	fa_2
.	.	.
.	.	.
.	.	.
$Li_n – Ls_n$	pm_n	fa_n
Total	—	n

Onde PM é o ponto médio da Classe, dado por:

$$pm_i = \frac{Ls_i + Li_i}{2}$$

Tratamento de Dados Agrupados



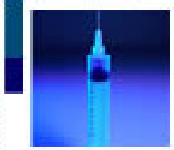


- A média aritmética para dados agrupados pode ser calculada adotando-se o ponto médio da classe como seu valor representativo

$$\bar{x} = \frac{\sum_{i=1}^{m} pm_i \cdot fa_i}{\sum_{j=1}^{m} fa_j} = \frac{\sum_{i=1}^{m} pm_i \cdot fa_i}{n}$$

Tratamento de Dados Agrupados



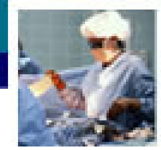
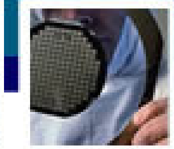
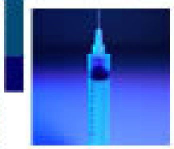


- Outra forma de calcular a média de dados agrupados é a partir do **calculo abreviado ou desvio em classes:**

$$\bar{x} = K + \left(\frac{\sum_{i=1}^{m} fa_i \cdot v_i}{n} \right) \cdot c$$

  – Onde:
    - K é o ponto médio de uma classe
    - fai é a frequência absoluta da classe i
    - νi é um valor de incremento ou decremento calculado a partir da classe escolhida para K
    - c é o intervalo de classe

Procedimento para o Cálculo Abreviado

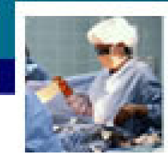
- Escolhe-se uma classes, preferencialmente a de maior frequência absoluta. O ponto média desta classe será K
- Atribui-se o valor zero ao ν a classe escolhida
- Os demais valores de ν são obtidos decrementando o seu valor para as classes anteriores à classe de K, e incrementando para as classes superiores

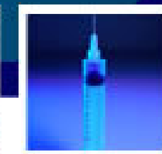
Exemplo

Tabela 4.5

Classe	fa	v	$fa.v$
0,0-0,2	3	−3	−9
0,2-0,4	6	−2	−12
0,4-0,6	18	−1	−18
0,6-0,8	35	0	0
0,8-1,0	16	+1	+16
1,0-1,2	2	+2	+4
Total	**80**	—	**−19**

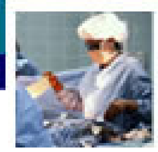
Mediana para Dados Agrupados

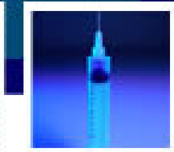




- Calcula-se a mediana pela expressão:
- $$\hat{x} = \hat{L}ir + \left(\frac{\frac{n}{2} - \sum_{i=1}^{med-1} fa_i}{fa_{med}} \right) \cdot c$$
  - Onde
    - $\hat{L}ir$ = limite inferior real da classe mediana
    - fa_{med} = frequência absoluta da classe mediana
    - Sendo a classe mediana aquela que contém o valor mediano
      - O valor mediano é o valor de ordem n/2

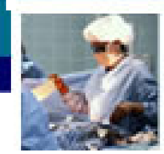
Exemplo - Mediana

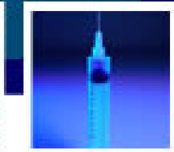




- Para a última tabela apresentada, qual o valor da mediana?

Moda para Dados Agrupados





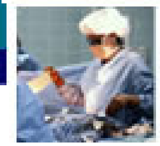
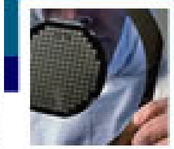
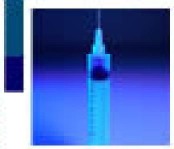
- Calcula-se a moda para dados agrupados a partir da expressão:

$$\tilde{x} = \tilde{L}ir + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2} \right) \cdot c$$

– Onde:

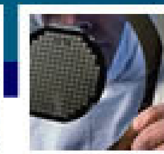
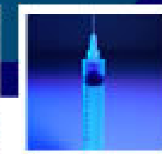
- Lir =limite inferior da classe modal
- $\Delta 1$ = Excesso de frequência da classe modal sobre a classe imediatamente inferior
- $\Delta 2$ = Excesso de frequência da classe modal sobre a classe imediatamente Superior.

Moda para Dados Agrupados

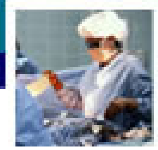
- Se existirem mais de uma frequência relevante na Tabela de dados, de tal forma que esta frequência não seja adjacente à frequência modal, então a distribuição dos dados terá mais de uma moda.
  - E todos os valores das modas devem ser calculados!

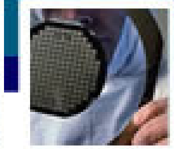
Exemplo

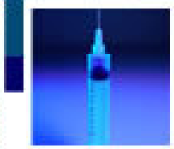




- Para a última tabela, exitem quantas modas?
  - E qual o valor destas?

Referências da Aula







- Livro Texto: Bioestatística. Teoria e Computacional (Héctor G. Arango). Guanaba
  - Seção 1 – Medidas de Tendência Central
- Leituras Complementares:
  - Livro: Bioestatística. Princípios e Aplicações (Sidia M. Callegari-Jaqcques)
    - Capítulo 2.
  - Livro: Estatística Básica (Wilton de O. Bussab e Pedro A. Morettin). Saraiva
    - Capítulo: 3