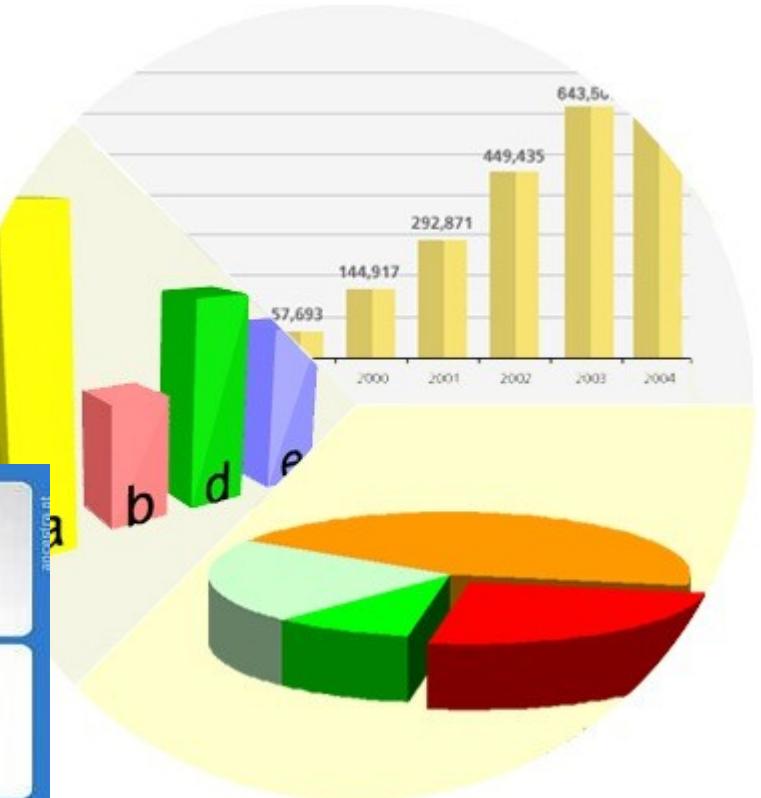
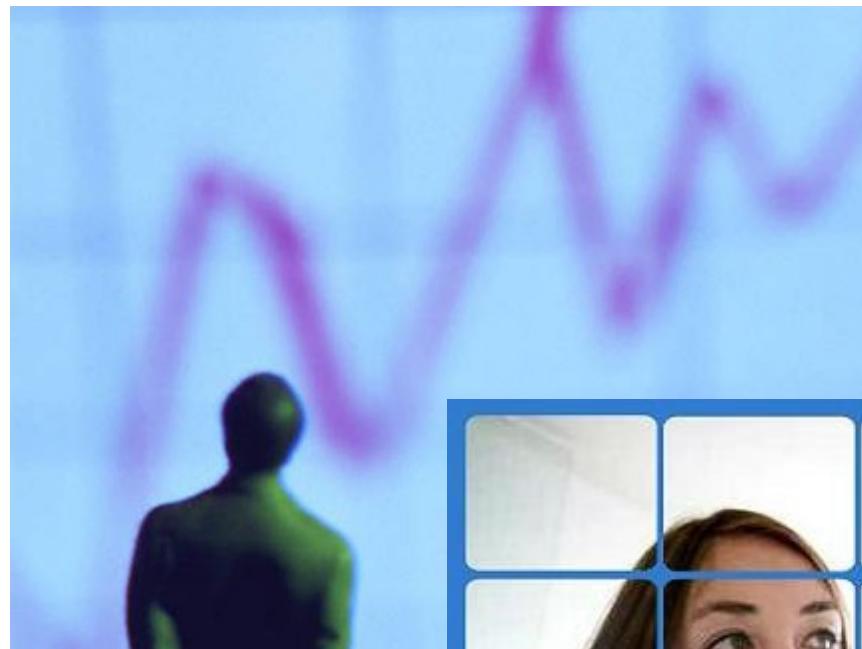


# Estatística Aplicada

## 19/01/2013



# Aula anterior - 05/01/2013

- Definição e Conceito Histórico.
- População e Amostra: conceito e exemplos.
- Censo e Estimação: qual usar?
- Dados estatísticos: valores numéricos, métodos, informações do FC, conclusões, decisões.
- Estatística Descritiva e Indutiva;
- Dados Qualitativos e Quantitativos;
- Técnicas de Amostragem: aleatória, estrat. e sistemática.

# Aula anterior - 05/01/2013

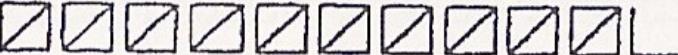
- Organização dos dados: Tabelas e Gráficos
  - Características de uma tabela;
  - Tabela de frequência e contingência;
  - Gráficos:
    - Linhas
    - Barras
    - Pontos
    - Histograma

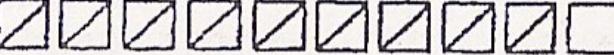
# Variáveis

- **Qualitativa:** dados distribuídos em categorias mutuamente exclusivas. Ex: sexo (M ou F), cor, causa de morte, grupo sanguíneo.
- **Ordinal:** dados distribuídos em categorias mutuamente exclusivas *que tem ordenação natural*. Ex: grau de instrução (primário, secundário e superior), status social, estágio da doença.
- **Quantitativa:** dados expressos em números. Ex: idade, estatura, peso corporal, etc.

# Apuração dos Dados

- Variável qualitativa ou ordinal: contagem.
- Variável quantitativa: anotar cada valor.

Masculino -  = 52

Feminino -  = 48

Nº do prontuário	Peso ao nascer
10 525	3,25
10 526	2,00
.	.
.	.
10 624	3,20

# População e Amostra

Entende-se por *população* o conjunto de elementos que têm em comum, determinada *característica*. Todo subconjunto não vazio e com menor número de elementos do que a população constitui uma *amostra*.

**Censo:** é o conjunto de dados obtidos através do recenseamento.

**Recenseamento:** quando se faz a coleta de informações sobre toda a população.

**Amostragem:** quando são coletadas informações de apenas parte da população.

# Técnicas de Amostragem

Definida a população, é preciso estabelecer a técnica de amostragem.

- Amostra casual simples: elementos retirados ao acaso da população.
- Amostra sistemática: não por acaso, mas por um *sistema*.
- Amostra estratificada: elementos provenientes de todos os *estratos* da população.
- Amostra de conveniência: elementos que o pesquisador reuniu simplesmente porque dispunha deles.

O pesquisador que trabalha com amostra sempre pretende fazer **inferência**, isto é, estender os resultados da amostra para toda a população.

# Aplicações

1.5.1 - Os prontuários dos pacientes de um hospital estão organizados em um arquivo, por ordem alfabética. Qual é a maneira mais rápida de amostrar 1/3 do total de prontuários?

1.5.2 - Um pesquisador tem dez gaiolas que contêm, cada uma, seis ratos. Como o pesquisador pode selecionar dez ratos para uma amostra?

# Aplicações

1.5.3 - *Para levantar dados sobre o número de filhos por casal, em uma comunidade, um pesquisador organizou um questionário que enviou, pelo correio, a todas as residências. A resposta ao questionário era facultativa, pois o pesquisador não tinha condições de exigir a resposta. Nesse questionário perguntava-se o número de filhos por casal morador na residência. Você acha que os dados assim obtidos têm algum tipo de tendenciosidade?*

# Aplicações

1.5.5 - Muitas pessoas acreditam que as famílias se tornaram menores. Suponha que, para estudar essa questão, foi selecionada uma amostra de 2.000 casais e perguntou-se quantos filhos eles tinham, quantos filhos tinham seus pais e quantos filhos tinham seus avós. O procedimento introduz tendenciosidade nos dados. Por quê?

# Apresentação de Dados em Tabelas

# Apresentação de Dados em Tabelas

Cabeçalho

Coluna indicadora

Corpo

Causa	Freqüência
Acidente	29.601
Abuso	2.604
Suicídio	7.965
Profissional	3.735
Outras	1.959
Ignorada	1.103

Fonte: MS/FIOCRUZ/SINITOX

Casos registrados de intoxicação humana, segundo a causa determinante. Brasil, 1993

Causa	Freqüência	Freqüência relativa
Acidente	29.601	63,03
Abuso	2.604	5,54
Suicídio	7.965	16,96
Profissional	3.735	7,95
Outros	1.959	4,17
Ignorada	1.103	2,35
Total	46.967	100,00

Nascidos vivos registrados segundo o ano do registro

Ano do registro	Freqüência
1984	2 559 038
1985	2 619 604
1986	2 779 253

Fonte: IBGE (1988)

Nota: Nascimentos ocorridos no ano de registro

# Tabela de Contingência

Nascidos vivos registrados segundo o ano de registro e o sexo

Ano de registro	Sexo		Total
	Masculino	Feminino	
1984	1 307 758	1 251 280	2 559 038
1985	1 339 059	1 280 545	2 619 604
1986	1 418 050	1 361 203	2 779 253

Fonte: IBGE (1988)

Nota: Nascimentos ocorridos no ano de registro

Dois fatores

Estimativa  
de Risco

Recém-nascidos segundo a época do ataque de rubéola na gestante e a condição de normal ou defeituoso

Época do ataque	Condição		Total	Freqüência relativa de defeituosos
	Normal	Defeituoso		
Até o 3º mês.....	36	14	50	28,0%
Depois do 3º mês	51	3	54	5,6%

Fonte: HILL et alii (1958)

# Tabela de Distribuição de Frequência

As tabelas com grande número de dados são cansativas e não dão ao leitor visão rápida e global do fenômeno. Para isso, é preciso que os dados estejam organizados em uma *tabela de distribuição de freqüências*. Nesta seção se explica, passo a passo, a construção desse tipo de tabela usando, como exemplo, os dados da Tabela 2.6.

Imagine que, para dar uma idéia geral sobre peso ao nascer de nascidos vivos, o pesquisador irá apresentar não os pesos observados, mas o número de nascidos vivos por faixas de peso. Deve, então, construir uma tabela de distribuição de freqüências.

Tabela 2.0

Peso ao nascer de nascidos vivos, em quilogramas

2,522	3,200	1,900	4,100	4,600	3,400
2,720	3,720	3,600	2,400	1,720	3,400
3,125	2,800	3,200	2,700	2,750	1,570
2,250	2,900	3,300	2,450	4,200	3,800
3,220	2,950	2,900	3,400	2,100	2,700
3,000	2,480	2,500	2,400	4,450	2,900
3,725	3,800	3,600	3,120	2,900	3,700
2,890	2,500	2,500	3,400	2,920	2,120
3,110	3,550	2,300	3,200	2,720	3,150
3,520	3,000	2,950	2,700	2,900	2,400
3,100	4,100	3,000	3,150	2,000	3,450
3,200	3,200	3,750	2,800	2,720	3,120
2,780	3,450	3,150	2,700	2,480	2,120
3,155	3,100	3,200	3,300	3,900	2,450
2,150	3,150	2,500	3,200	2,500	2,700
3,300	2,800	2,900	3,200	2,480	-
3,250	2,900	3,200	2,800	2,450	-

Primeiro, é preciso definir as faixas de peso que recebem, tecnicamente, o nome de classes. Observe os dados apresentados na Tabela 2.6. O menor valor é 1,570kg e o maior valor é 4,600kg. Podem então ser definidas classes de 1,5 a 2,0kg, de 2,0 a 2,5kg, e assim por diante, como mostra o esquema dado a seguir:

1,5	—	2,0
2,0	—	2,5
2,5	—	3,0
3,0	—	3,5
3,5	—	4,0
4,0	—	4,5
4,5	—	5,0

Na classe de 1,5 a menos de 2,0kg são colocados desde nascidos com 1,5kg até os que nasceram com 1,999kg; na classe de 2,0 a menos de 2,5kg são colocados desde nascidos com 2,0kg até os que nasceram com 2,499kg, e assim por diante. Logo, cada classe cobre um intervalo de 0,5kg, ou seja, cada intervalo de classe é de 0,5kg. É mais fácil trabalhar com intervalos de classe iguais. A distribuição das freqüências, obtida a partir da Tabela 2.6, é dada a seguir.

Classe	Freqüência
1,5 — 2,0	█ = 3
2,0 — 2,5	█ █ █ = 16
2,5 — 3,0	█ █ █ █ █ █ █ = 31
3,0 — 3,5	█ █ █ █ █ █ █ █ = 34
3,5 — 4,0	█ █ = 11
4,0 — 4,5	█ = 4
4,5 — 5,0	█ = 1

Nascidos vivos segundo o peso ao nascer, em quilogramas

Classe	Ponto médio	Freqüência
1,5—2,0	1,75	3
2,0—2,5	2,25	16
2,5—3,0	2,75	31
3,0—3,5	3,25	34
3,5—4,0	3,75	11
4,0—4,5	4,25	4
4,5—5,0	4,75	1

$$\frac{1,5 + 2,0}{2} = 1,75$$

$$k = 1 + 3,222 \cdot \log n,$$

$$k = 1 + 3,222 \cdot \log 100 = 7,444,$$

2.4.1 - De acordo com o IBGE (1988), a distribuição dos suicídios ocorridos no Brasil em 1986, segundo a causa atribuída, foi a seguinte: 263 por alcoolismo, 198 por dificuldade financeira, 700 por doença mental, 189 por outro tipo de doença, 416 por desilusão amorosa e 217 por outras causas. Apresente essa distribuição em uma tabela.

Tabela 2.9

Suicídios ocorridos no Brasil em 1986, segundo a causa atribuída

Causa atribuída	Freqüência	Percentagem
Alcoolismo .....		
Dificuldade financeira .....		
Doença mental .....		
Outro tipo de doença .....		
Desilusão amorosa .....		
Outras .....		

Fonte: IBGE (1988)

2.4.2 - Construa uma tabela de distribuição de freqüências para apresentar os dados da Tabela 2.10.

**Tabela 2.10**  
Pressão arterial, em milímetros de mercúrio, de cães adultos  
anestesiados e após laparotomia

130,0	105,0	120,0	111,5	99,0	116,0	82,5
107,5	125,0	100,0	107,5	120,0	143,0	115,0
135,0	130,0	135,0	127,5	90,5	104,5	136,5
100,0	145,0	125,0	104,5	101,5	102,5	101,5
134,5	158,5	110,0	102,5	90,5	107,5	124,0
121,5	135,0	102,0	119,5	115,5	125,5	117,5
107,5	140,0	121,5	107,5	113,0	93,0	103,5

Fonte: ARAÚJO e HOSSNE (1977)

Para determinar o número de classes pode ser usada a fórmula:

$$k = 1 + 3,222 \cdot \log n,$$

onde  $n$  é 49. Então,

$$k = 1 + 3,222 \cdot 1,6902 = 6,4458.$$

De acordo com a fórmula, podem ser constituídas 6 ou 7 classes. Como o menor valor observado é 82,5 e o maior valor é 158,5, é razoável construir classes com intervalos iguais a 10, a partir de 80. O número de classes será, então, 8, um pouco maior do que o estabelecido pela fórmula.

**Tabela 2.11**

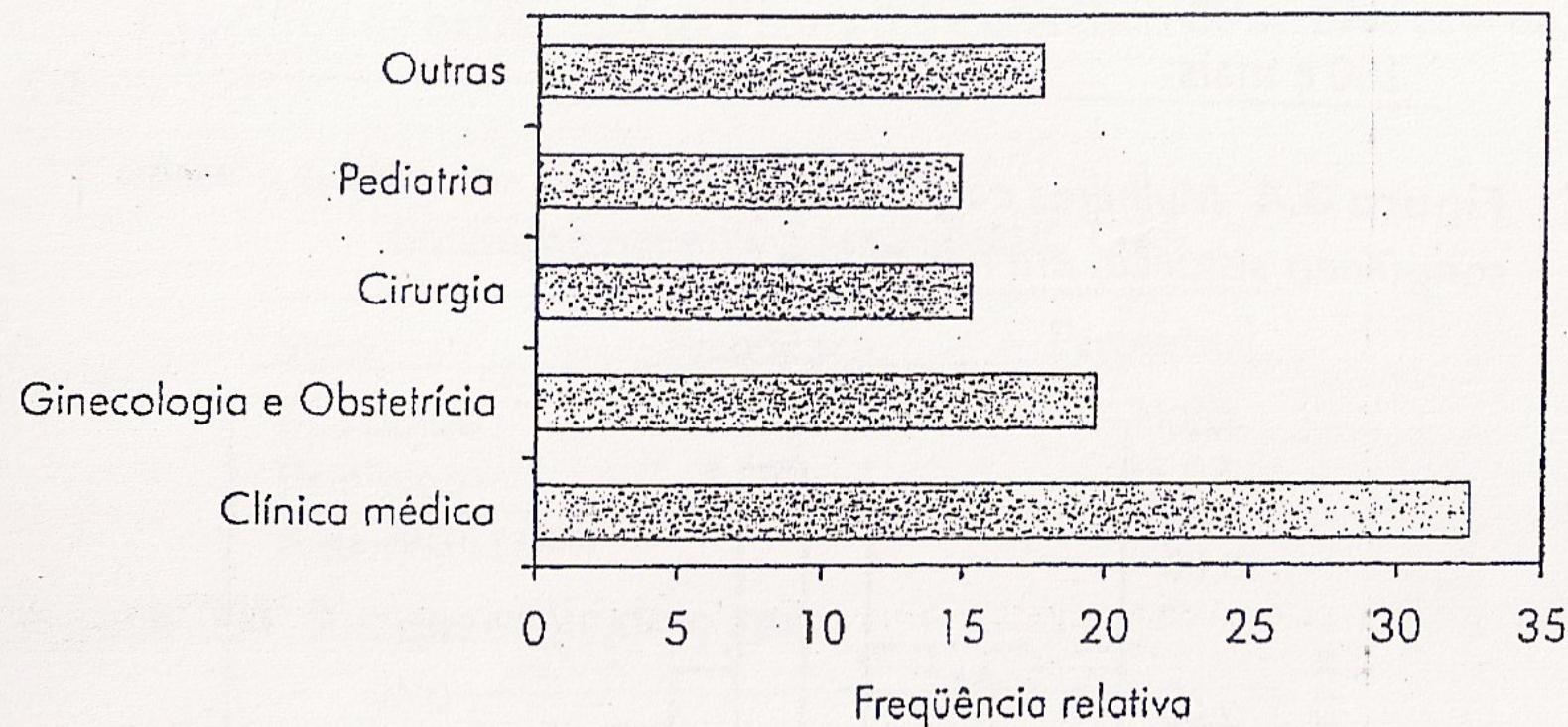
Cães adultos anestesiados e após laparotomia segundo a pressão arterial, em milímetros de mercúrio

Classe	Ponto médio	Freqüência
80— 90	85	1
90— 100	95	4
100— 110	105	16
110— 120	115	8
120— 130	125	9
130— 140	135	7
140— 150	145	3
150— 160	155	1

# Apresentação de Dados em Gráficos

# Gráfico de Barras

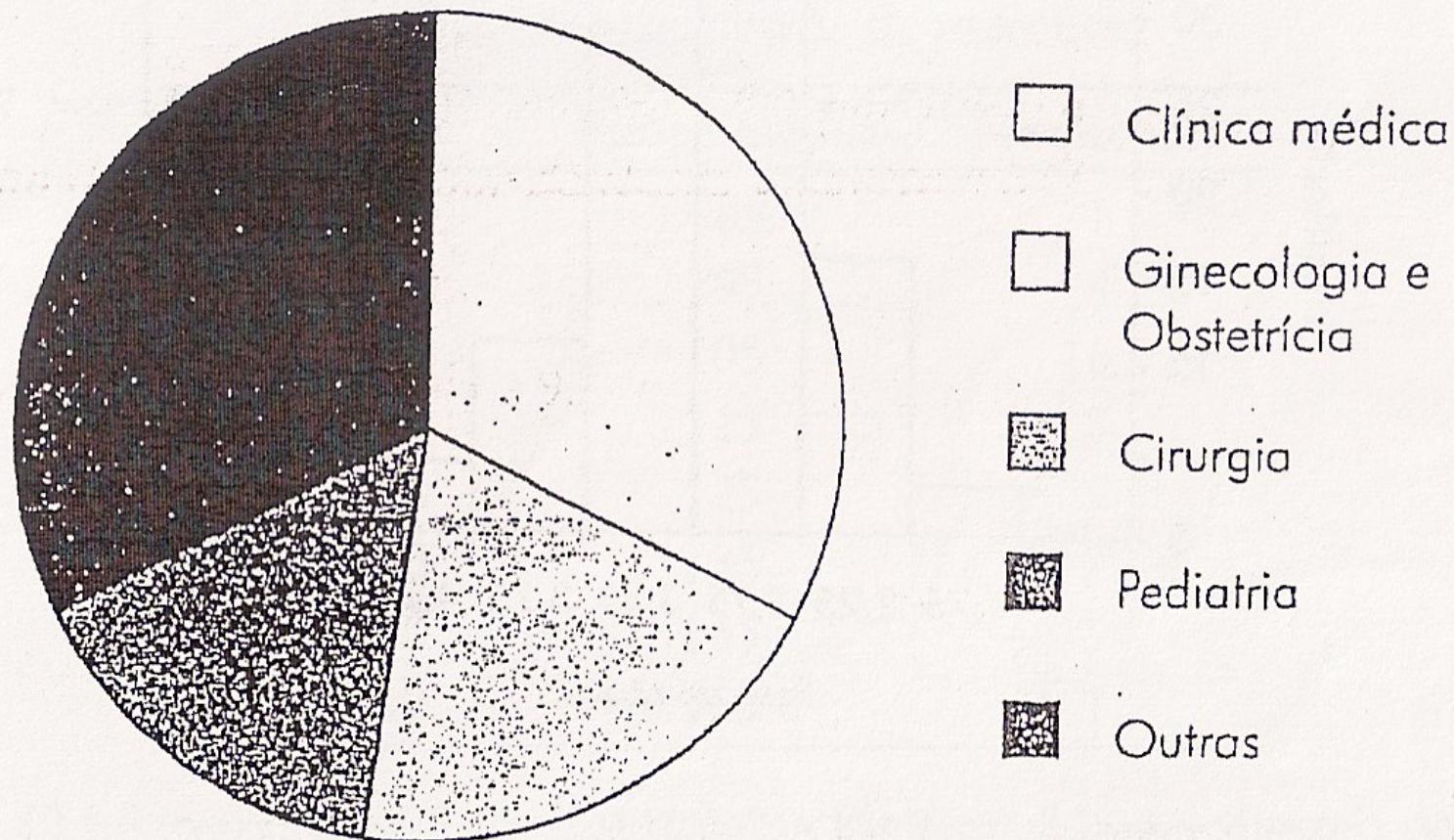
**Figura 3.1** Internações em estabelecimentos de saúde, por espécie de clínica. IBGE 1992



Variáveis qualitativas ou ordinais.

# Gráficos de Setores

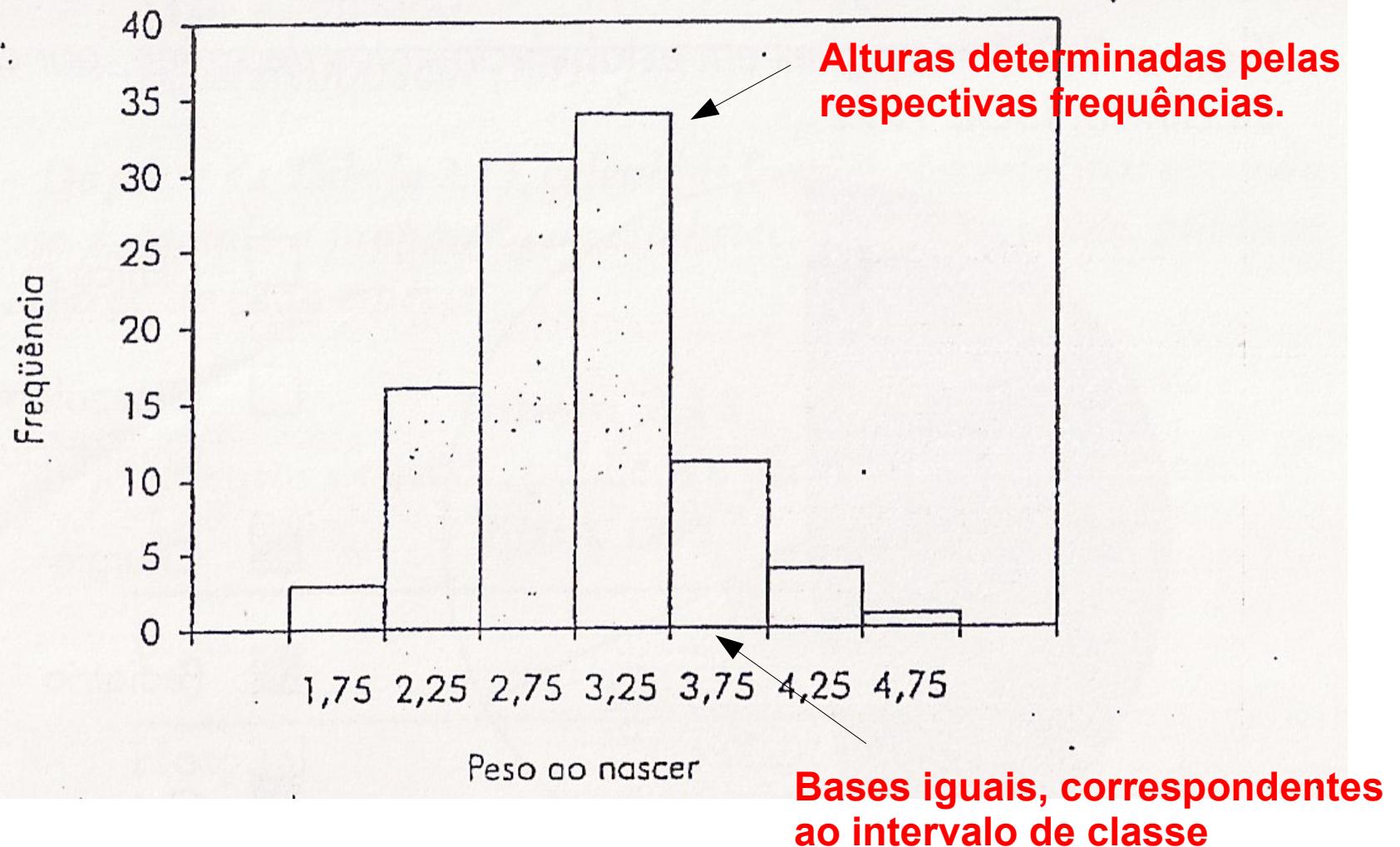
**Figura 3.2** Internações em estabelecimentos de saúde, por espécie de clínica. IBGE 1992



Variáveis qualitativas ou ordinais.

# Histograma

Figura 3.3 Nascidos vivos segundo o peso ao nascer, em quilogramas



# Aplicações

3.5.1 - Faça um gráfico de barras e um gráfico de setores para apresentar os dados da Tabela 3.3.

**Tabela 3.3**  
Suicidas segundo o sexo. Brasil, 1986

Sexo	Freqüência	Percentual
Masculino	3 562	74,93
Feminino	1 192	25,07

*Fonte: IBGE (1988)*

# Medidas de Tendência Central

Os dados quantitativos, apresentados em tabelas e gráficos, constituem a informação básica do problema em estudo. Mas é conveniente apresentar, além dos dados, medidas que mostrem a informação de maneira resumida. As medidas de tendência central, definidas neste Capítulo, dão o valor do ponto em torno do qual os dados se distribuem. São medidas de tendência central: a média aritmética (ou simplesmente média), a mediana e a moda.

## Média Aritmética

## Mediana

## Moda

# Média Aritmética

Tabela 4.1

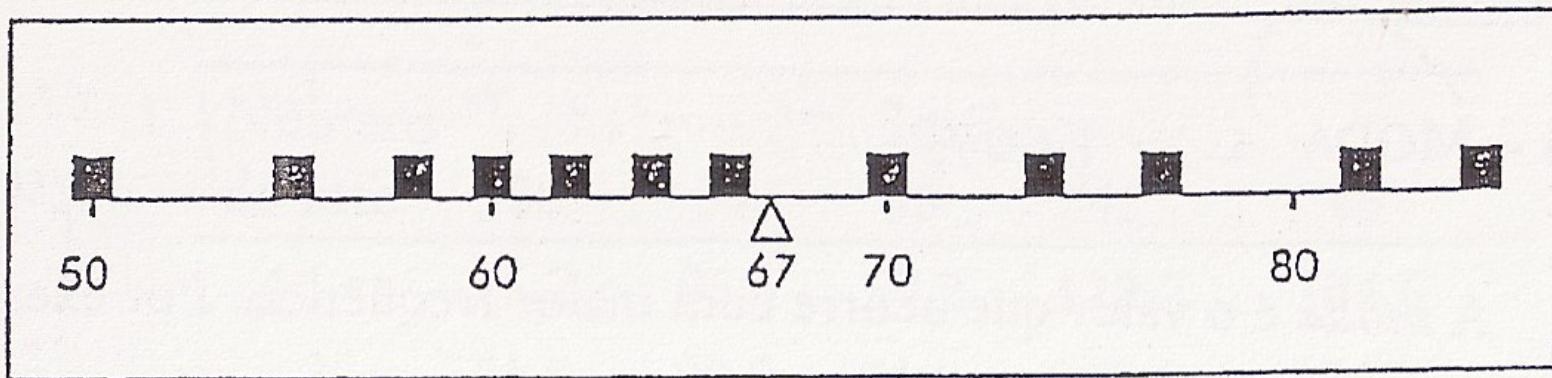
Peso, em gramas, de ratos machos da raça Wistar com 30 dias de idade

50	62	70
86	60	64
66	77	58
55	82	74

A média aritmética dos dados apresentados na Tabela 4.1 é:

$$\frac{50 + 86 + \dots + 74}{12} = \frac{804}{12} = 67$$

**Figura 4.1** Distribuição de dados sobre o eixo e a respectiva média



# Média de dados em tabelas de distribuição de frequências

Tabela 4.2

Nascidos vivos segundo o peso ao nascer, em quilogramas

Classe	Ponto médio	Freqüência
1,5 — 2,0	1,75	3
2,0 — 2,5	2,25	16
2,5 — 3,0	2,75	31
3,0 — 3,5	3,25	34
3,5 — 4,0	3,75	11
4,0 — 4,5	4,25	4
4,5 — 5,0	4,75	1

Tabela 4.3  
Distribuição de freqüências

Ponto médio	Freqüência
$x_1$	$f_1$
$x_2$	$f_2$
.	.
.	.
$x_k$	$f_k$

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}, \quad \text{onde } n = \sum_{i=1}^k f_i$$

# Mediana

Se a amostra é constituída por um número ímpar de dados, a *mediana* é o valor que fica no centro dos dados ordenados. Por exemplo, a mediana dos valores

1, 2, 3, 5 e 9

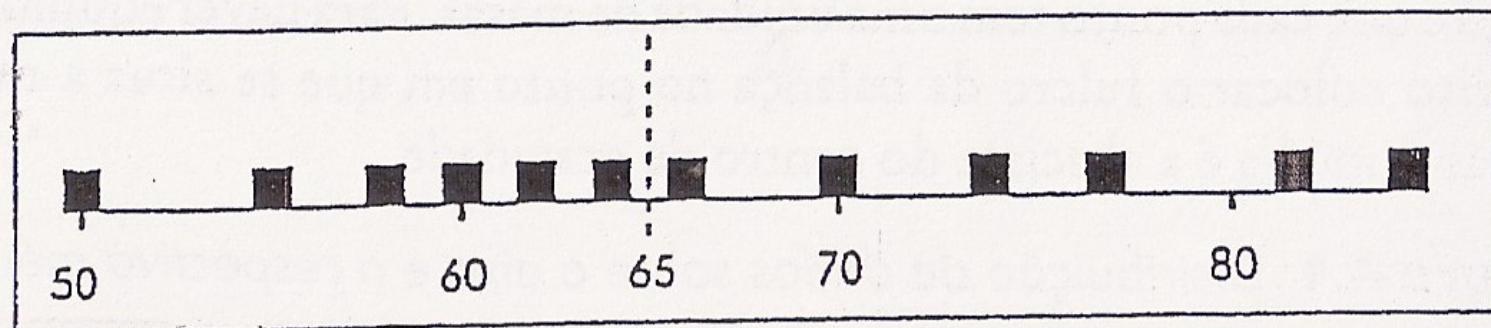
é 3.

Se a amostra é constituída por um número par de dados, a *mediana* é a média aritmética dos dois valores que ficam na posição central dos dados ordenados. Por exemplo, a mediana dos valores

1, 2, 3, 4, 7 e 9

é a média aritmética dos números 3 e 4, ou seja, a mediana é 3,5.

**Figura 4.2** Distribuição de dados sobre o eixo e a respectiva mediana



# Moda

A *moda* é o valor que ocorre com maior freqüência. Por exemplo, dados os números

3, 4, 5, 7, 7, 7, 9 e 9,

a moda é 7, porque 7 é o número que ocorre maior número de vezes.

Existem conjuntos de dados que não apresentam moda, porque nenhum valor se repete maior número de vezes, e existem conjuntos de dados com duas ou mais modas. Assim, o conjunto de números

1, 2, 3, 4 e 5

não tem moda, e o conjunto de números

1, 2, 2, 3, 4, 4 e 5

tem duas modas, 2 e 4.

A moda, diferentemente das outras medidas de tendência central, pode ser obtida mesmo que a variável seja qualitativa. Veja os dados apresentados na Tabela 4.8. O sangue tipo O ocorreu com maior freqüência. Então a moda dessa amostra é sangue tipo O.

**Tabela 4.4**  
Indivíduos segundo o tipo de sangue

Tipo de sangue	Freqüência
O	547
A	441
B	123
AB	25

*Fonte: GARCIA (1977)*

# Medidas de Dispersão

# Medidas de Dispersão

Tabela 5.1

Notas de quatro alunos em cinco provas

Aluno	Notas					Média
Antônio	5	5	5	5	5	5
João	6	4	5	4	6	5
José	10	5	5	5	0	5
Pedro	10	10	5	0	0	5

Todos os alunos obtiveram média igual a 5, mas a dispersão das notas em torno da média não é a mesma para todos os alunos. A Tabela 5.1 mostra claramente que:

- a) As notas de Antônio não variaram (a dispersão é nula).
- b) As notas de João variaram menos do que as notas de José (a dispersão das notas de João é menor do que a dispersão das notas de José).
- c) As notas de Pedro variaram mais do que as notas de todos os outros (a dispersão das notas de Pedro é a maior).

# Medidas de dispersão

- Amplitude: é a diferença entre o maior e o menor dado observado;
- Variância ( $s^2$ ): soma dos quadrados dos desvio em relação à média, dividido por (n-1);
- Desvio Padrão (s) - (raiz da variância);
- Coeficiente de Variação (CV) –  $s/x * 100$ ;

**Tabela 5.5**

Média e variância das notas de 4 alunos em 5 provas

Aluno	Notas					Média	Variância
Antônio	5	5	5	5	5	5	0
João	6	4	5	4	6	5	1
José	10	5	5	5	0	5	12,5
Pedro	10	10	5	0	0	5	25

FIM