

# **ESTATÍSTICA**

## **Introdução à Estatística**

**Prof. Dr. Ricardo da Silva Manca**

# Estatística: O que é?

- Estatística pode ser pensada como a ciência de aprendizagem a partir de dados.
- Em linhas gerais, a Estatística fornece métodos que auxiliam o processo de tomada de decisão.
- A Estatística está presente em todas as áreas da ciência que envolvam a coleta e análise de dados.

# Introdução à Estatística

A Estatística está compreendida em duas partes:

- **Estatística Descritiva:** Reúne um conjunto de técnicas para sumarizar os dados (**tabelas**, **gráficos**) e medidas descritivas que permitem tirar muitas informações contidas nos dados.
- **Estatística Indutiva:** Produzir afirmações sobre uma dada característica da população, na qual estamos interessados, a partir de informações colhidas de uma parte dessa população.

# Conceitos Básicos de Estatística

- A finalidade da pesquisa é coletar **dados** para obter informações.
  - ▷ **Dados** – observações de uma ou mais **variáveis**.
  - ▷ **Variável** é aquilo que se deseja observar para se tirar algum tipo de conclusão, por ex., idade, sexo, peso e outras.
  - ▷ Dados usualmente provem de uma **amostra**, a qual representa uma **população** de interesse.

# Conceitos Básicos de Estatística

- **População:** É o conjunto de indivíduos (ou objetos) que apresentam pelo menos uma característica em comum, cujo comportamento deseja-se analisar ou inferir.

**Exemplo:** Estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de São Luís.

- **População alvo** – todas as crianças nesta faixa etária deste município.
- **População de estudo** – crianças matriculadas em escolas.

- **Amostra:** É um subconjunto da população.

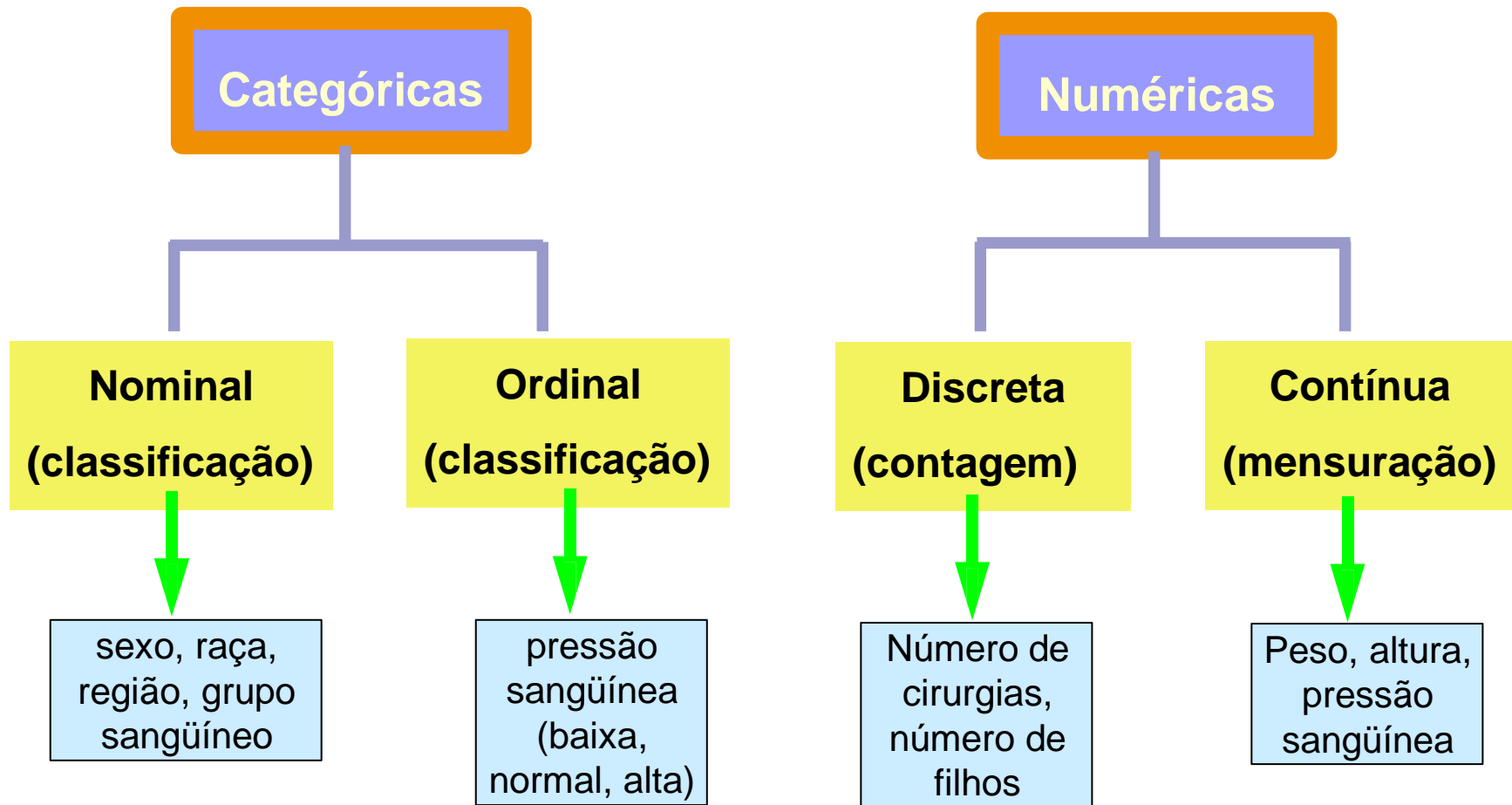
# Conceitos Básicos de Estatística

- **Parâmetro:** uma medida numérica que descreve alguma característica de uma população.
  - D Frequentemente desconhecido e denotado por **letras gregas**
  - D **Exemplo:** **Peso médio** ao nascer de crianças que nascem no município de São Luís
- **Estatística:** uma medida numérica que descreve alguma característica de uma amostra.
  - D É habitualmente representada por letras latinas
  - D **Exemplo:** **Peso médio** ao nascer, calculado em uma amostra de 120.000 crianças nascidas no Município de São Luís

# Tipos de Variáveis

- As variáveis podem ser categóricas (**qualitativas**) ou numéricas (**quantitativas**)
- **Variáveis qualitativas**: São características de uma população que não pode ser medidas.
  - Ordinais** – **Ex**: Grau de gravidade de uma doença
  - Nominais** – **Ex**: Presença de um sintoma
- **Variáveis quantitativas**: São características de uma população que pode ser quantificadas.
  - Discretas** – **Ex**: Número de cirurgias
  - Contínuas** – **Ex**: Idade, Pressão Arterial

# Esquemáticamente





# Tipos de variáveis

Classifique as variáveis apresentadas na tabela:

<b>Idade</b>	<b>Sexo</b>	<b>Hemoglobina</b>	<b>Tipo de urticária</b>	<b>Duração</b>
34	masc	14,2	física	curta
58	masc	14,4	física	longa
31	fem	15,1	idiopática	média
49	masc	10,9	idiopática	média
39	fem	14,4	física	longa
33	masc	14,1	física	curta
35	fem	14,0	idiopática	longa

O tipo da variável irá indicar a melhor forma para apresentação em tabelas e gráficos, em medidas de resumo e a análise estatística mais adequada.

# Qual o papel da Estatística na ciência?

- O propósito da investigação é responder uma *questão científica*.
- Na ciência, são realizados estudos *experimentais* ou *observacionais*, levando à coleção de *dados* numéricos.
- O padrão de *variação* nos dados faz com que a resposta não seja óbvia.

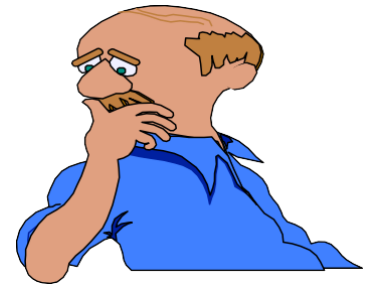
# Por que usar Estatística?

- Por que a natureza apresenta **VARIABILIDADE**:
  - Variações de indivíduo para indivíduo;
  - Variações no mesmo indivíduo;
- Segundo Pereira (1997), a estatística é a tecnologia da ciência e, portanto, a estatística deve estar presente desde o início da pesquisa.
- Sem Métodos Estatísticos, sem validade científica!

# Tipos de pesquisas científicas

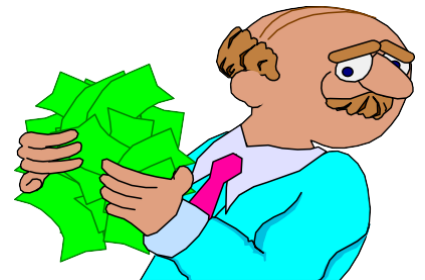
- **DE LEVANTAMENTO**

Características de interesse de uma população são levantadas (observadas ou medidas), mas sem manipulação.



- **EXPERIMENTAL**

Grupos de indivíduos (ou animais, ou objetos) são manipulados para se avaliar o efeito de diferentes tratamentos.



# Pesquisas de levantamento

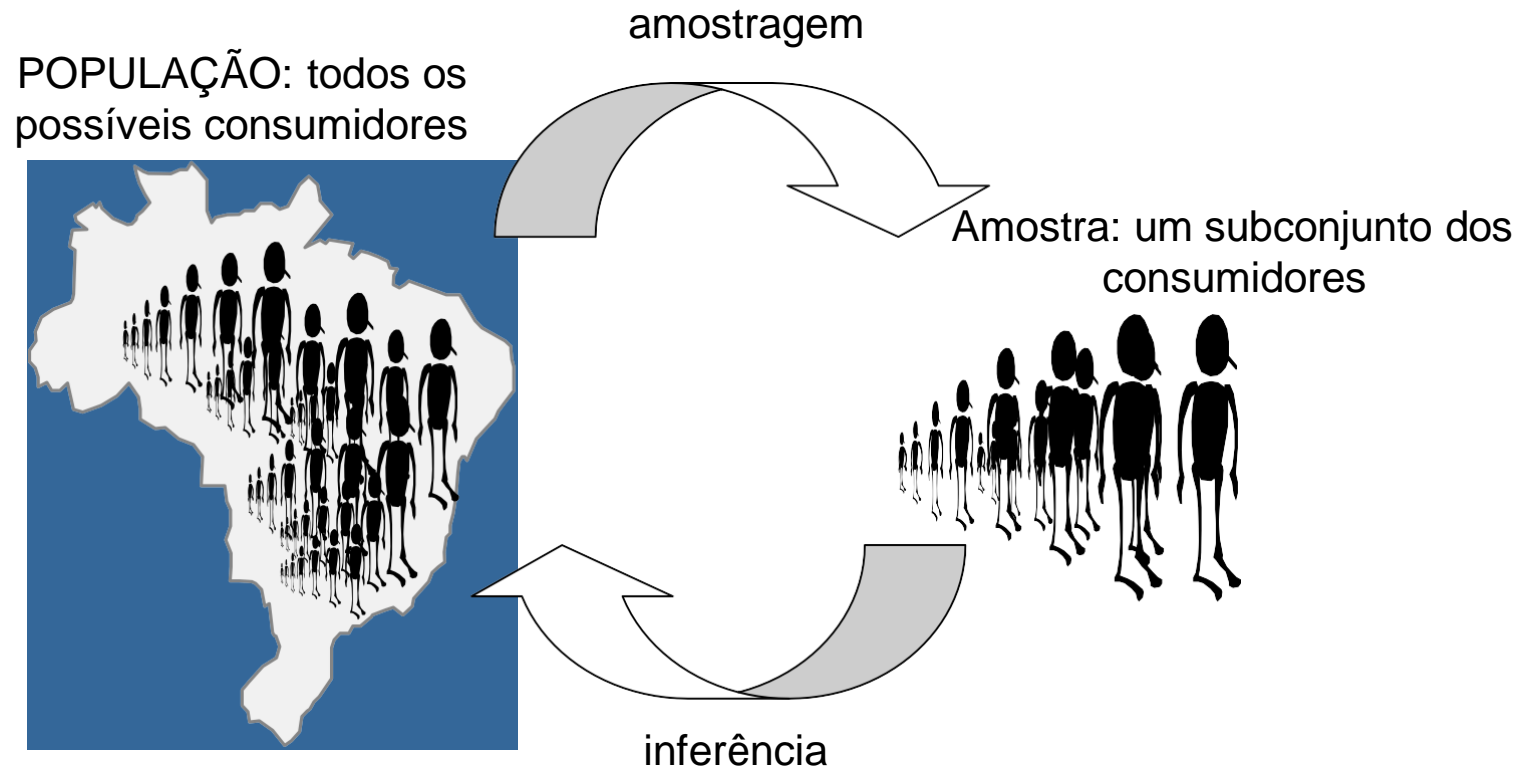
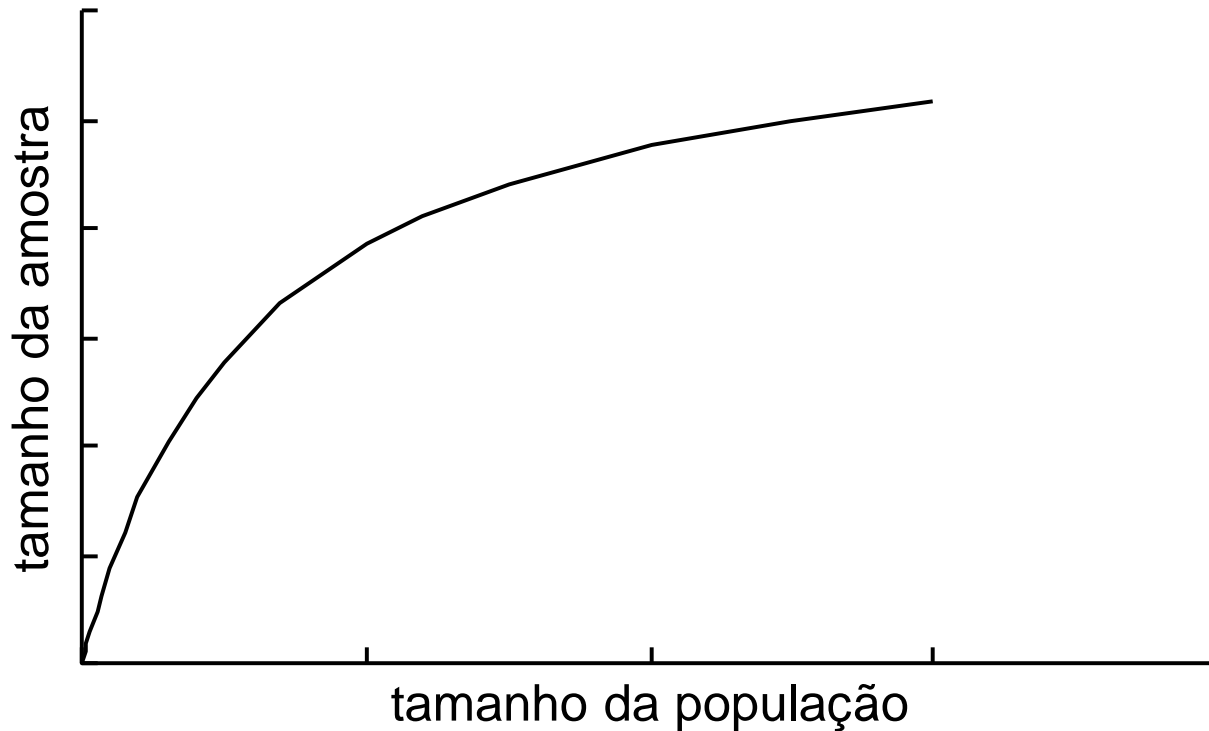


Ilustração de um levantamento por amostragem

# Amostragem

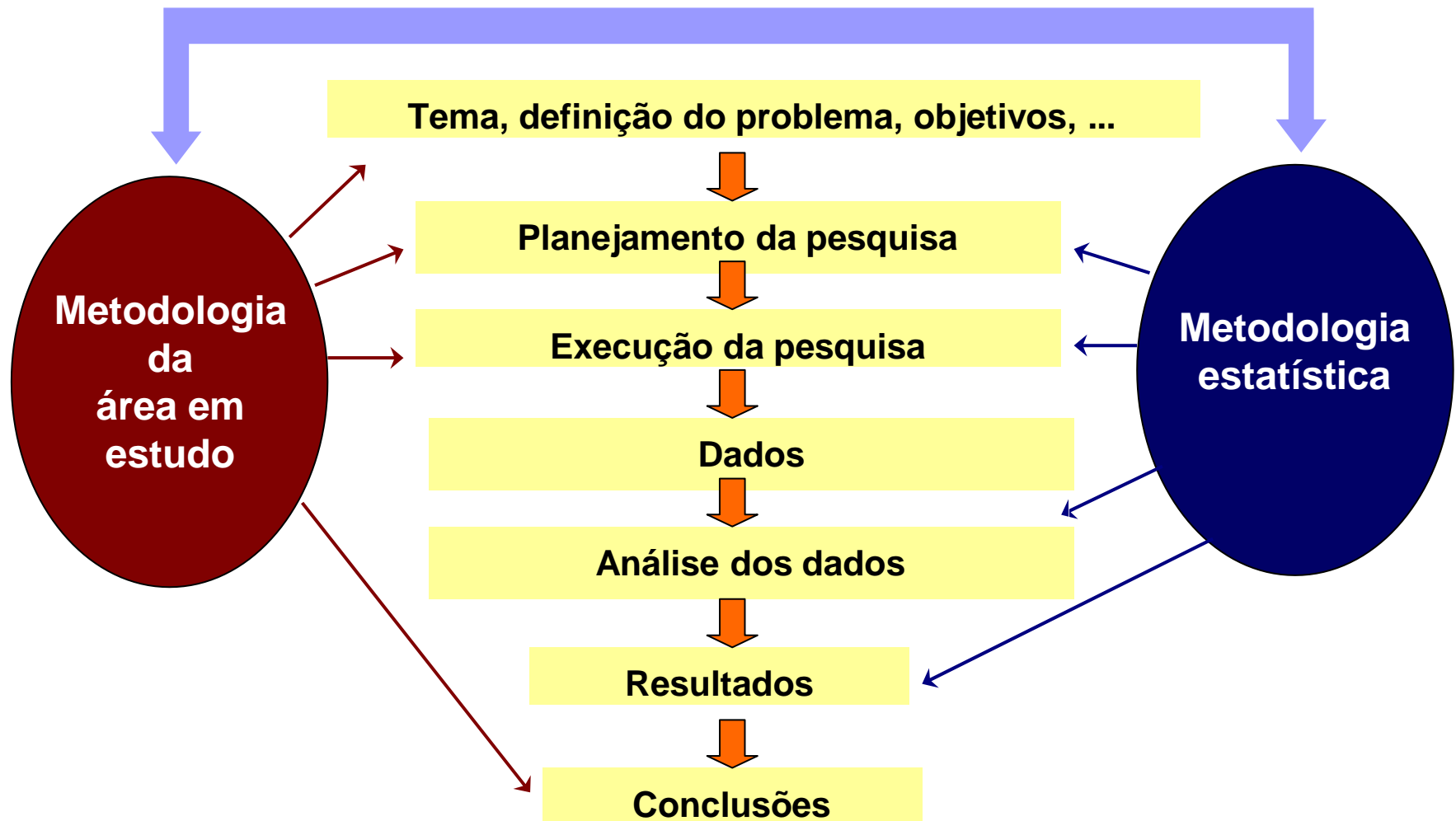
- Representatividade da amostra
- Tamanho da amostra
- Aleatoriedade da amostra
  - Garantir que TODOS os elementos da população tenham chance de pertencer à amostra.
  - Sorteio NÃO VICIADO.
  - Única forma de poder generalizar estatisticamente os resultados para a população.

# Censo ou amostragem



Relação entre tamanho da população e tamanho da amostra para garantir determinada margem de erro

# Etapas usuais de uma pesquisa científica





# Fases de uma Pesquisa



- Definição do problema
- Planejamento
- Coleta dos dados
- Apuração dos dados
- Apresentação dos dados
- Análise e interpretação dos dados

# Definição do Problema

- Formular corretamente o problema.
- Definir a população a ser estudada.
- Quais variáveis serão observadas?
- Quais hipóteses serão avaliadas?
- Determinar o que se pretende investigar.
- Estudos realizados ([revisão da literatura](#)).

# Planejamento da Pesquisa

- Nesta fase, são definidos:
  - Objetivos a serem alcançados
  - Bibliografia, materiais, impressos, equipamentos a serem utilizados
  - Tipo de levantamento (censo ou amostragem)
  - Pessoal que vai ser envolvido no trabalho
  - Locais de trabalho
  - Cronograma da execução

Perguntas que precisam ser respondidas no planejamento de uma pesquisa

## ■ O quê?

D características a serem observadas → VARIÁVEIS

## ■ Quem?

D os elementos a serem pesquisados →  
POPULAÇÃO

## ■ Como?

D o instrumento de coleta de dados →  
QUESTIONÁRIO / ENTREVISTA ESTRUTURADA

# Coleta dos dados



- Definir o instrumento de pesquisa:
  - Prontuários
  - Protocolos
  - Questionários
- Tipos de informações:
  - Primárias
  - Secundárias

# Apuração dos Dados

- Consiste em resumir os dados, através de contagem ou agrupamento.
- Frequentemente, exige um programa computacional.
- Por exemplo, [Access](#), [Excel](#) ou [Epi Info](#).
- As variáveis categóricas devem ser codificadas.
- Codificar dados ausentes.

# Exemplo de Banco de Dados

Paciente	Sexo	Peso	Tipo de Tratamento	Nº de Convulsões	Classificação da Doença
1	M	89,79	A	1	Leve
2	F	64,20	A	3	Severa
3	M	91,00	B	2	Moderada
4	F	51,68	A	0	Moderada
5	F	48,52	B	1	Leve
...					
58	M	71,00	B	0	Severa
59	M	78,80	A	2	Leve
60	F	71,00	B	3	Moderada

# Análise e Interpretação dos Dados

- Tirar conclusões que auxiliem o pesquisador.
- Necessidade de um programa estatístico (STATA, EPIINFO, BIOESTAT, SAS, SPSS, MINITAB)
- Análise estatística:
  - Estatística Descritiva: tabelas ou gráficos, média, mediana, desvio padrão.
  - Estatística Indutiva: testes estatísticos.





# Apresentação dos dados

# Análise Descritiva dos Dados

- A análise descritiva consiste basicamente na **organização** e **descrição** dos dados.
- Elementos básicos: **tabelas**, **gráficos** e **medidas numéricas**.
- Começaremos a análise com apenas uma variável em estudo.

# Tabela de Frequências

- Forma de representação da frequência de cada valor distinto da variável em estudo.
- Juntamente com as frequências simples, a tabela poderá ainda incluir:
  - Frequências relativas
  - Frequências acumuladas
  - Frequências relativas acumuladas.

# Tabela de Freqüências

- **Freqüência relativa:** percentagem relativa à freqüência.
- **Freqüência acumulada:** número de vezes que uma variável assume um valor inferior ou igual a esse valor.
- **Freqüência relativa acumulada:** percentagem relativa à freqüência acumulada.

# Tabela de Frequências

## ■ Exemplo: Consideremos a seguinte tabela

Nome	Sexo	Nome	Sexo
Paula	F	Gonçalo	M
Manuel	M	Pedro	M
Carla	F	Cristina	F
Maria	F	Sofia	F
João	M	Susana	F

Temos,

### **Sexo Masculino:**

Frequência absoluta : **4**

Frequência relativa: 4 em 10 = **40%**

### **Sexo Feminino:** Frequência absoluta : **6**

Frequência relativa: 6 em 10 = **60%**

# Tabela de Frequências

- Assim a **tabela de frequências** da variável Sexo, para o exemplo anterior, será:

Sexo	Freq. Simples (n)	Freq. Relativa (%)
Feminino	6	60
Masculino	4	40
Total	10	100

# Elementos essenciais de uma

## tabela

- **Título:** uma indicação que antecede a tabela e explique tudo referente a tabela.
- **Cabeçalho:** colocado na parte superior da tabela, especificando o conteúdo das colunas.
- **Corpo:** corresponde ao conjunto de colunas e de linhas que contêm informações sobre o fenômeno estudado.

# Elementos complementares da tabela

- **Fonte:** é a indicação do órgão ou entidade responsável pelo fornecimento dos dados ou pela sua elaboração. É colocada no rodapé da tabela.
- **Notas:** são informações destinadas a esclarecer o conteúdo das tabelas, ou indicar a metodologia adotada na coleta ou preparo dos dados.
- **Chamadas:** são informações de natureza específica referindo-se a um item específico da tabela, colocado no rodapé da página.



# Elementos essenciais de um gráfico

- Todo gráfico deve ter **título**, **escala** e **fonte** de dados, de forma a dispensar qualquer esclarecimento adicional.
- A numeração dos gráficos é feita utilizando-se algarismos arábicos.
- As escalas devem crescer da esquerda para a direita e de baixo para cima.
- As distâncias que indicam as unidades devem ser rigorosamente uniformes.

# Variável Qualitativa

- Podemos sumarizar a variável em:
  - Tabelas – usando contagens ou porcentagens
  - Gráfico de Barras ou Gráfico de Setores

# Tabelas: Variável Qualitativa

**Tabela 1. Tipo de parto em nascidos vivos de parto único.  
São Luís- MA, 1997/98**

Tipo de parto	Freqüências	%
Vaginal	1619	66,27
Cesáreo	824	33,73
Total	2443	100,00

# Tabelas: Variável qualitativa

**TABELA 2 – Número e porcentagem de causas de morte de residentes em São Luís, no período de 10 de agosto a 31 de dezembro de 2005.**

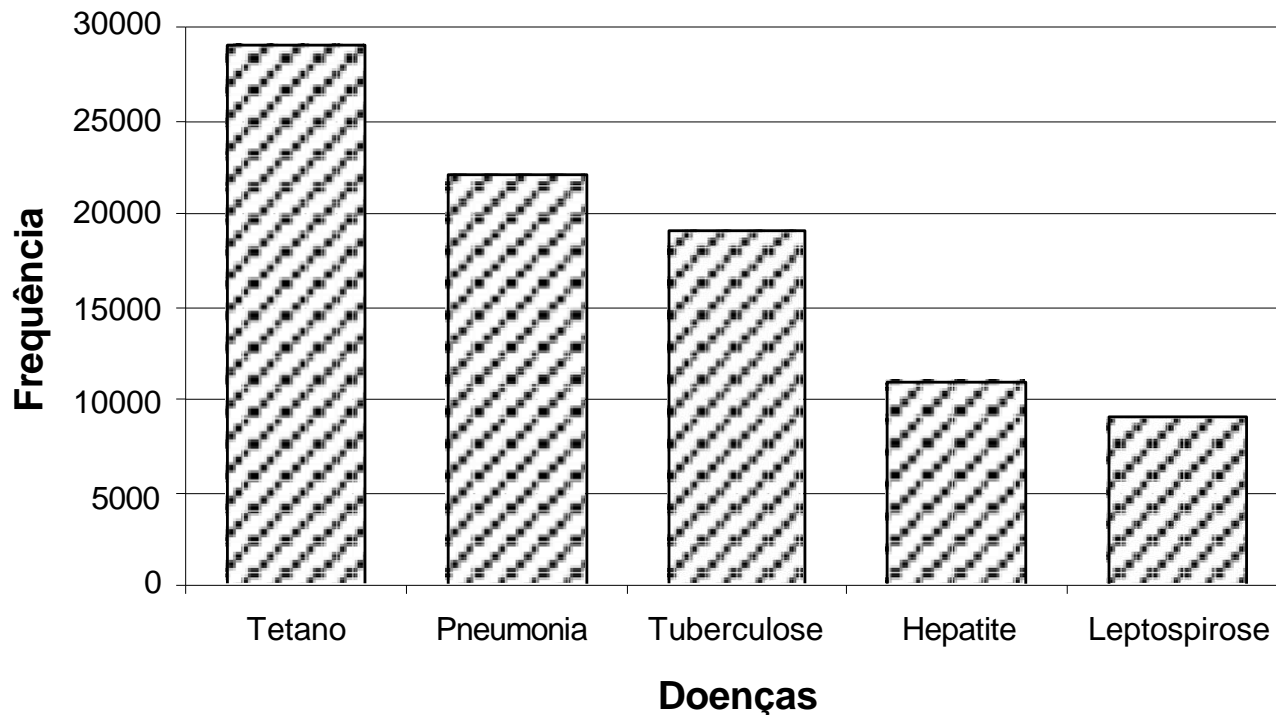
<b>CAUSAS DA MORTE</b>	<b>Frequência</b>	<b>%</b>
Doenças do ap. circulatório	281	33,5
Neoplasias	115	13,7
Causas externas	92	11,0
Doenças do ap. respiratório	87	10,4
Doenças das glând. endócrinas	56	6,7
Doenças do ap. digestivo	54	6,4
Doenças e infec. e parasitárias	46	5,5
Afecções do per. Perinatal	26	3,1
Demais grupos	82	9,8
<b>TOTAL</b>	<b>839</b>	<b>100,0</b>

**Fonte: Desconhecida**

# Gráficos: Variável Qualitativa

## Gráfico de Barras

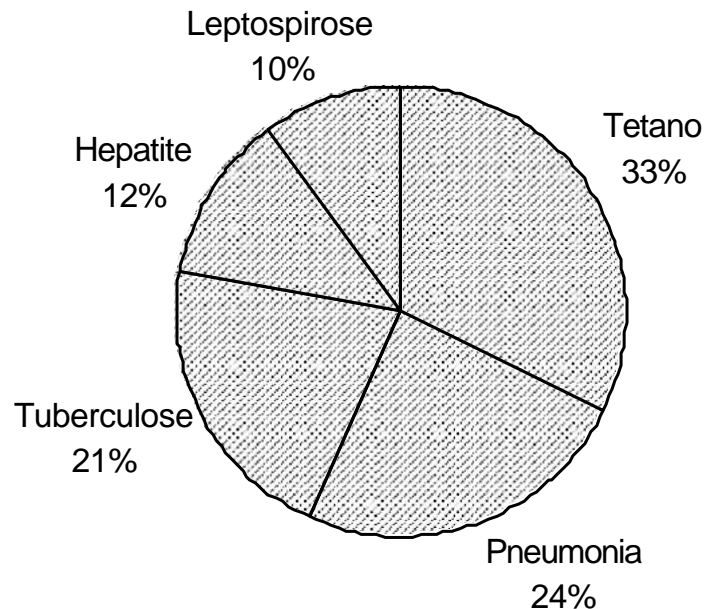
**Figura 1:** Dados sobre as doenças mais comuns ocorridas no Estado de São Paulo



# Gráficos: Variável Qualitativa

## Gráfico de Setores

**FIGURA 2:** Dados sobre as doenças mais comuns ocorridas no Estado de São Paulo



# Variável Quantitativa

- Podemos sumarizar a variável em:
  - ▶ Tabelas de Freqüências
  - ▶ Histograma ou Polígono de Freqüências
  - ▶ Gráficos de linhas
  - ▶ Box plot

# Tabela de Frequências

**TABELA 2: Tempo de Internação (em dias) de 160 pacientes  
no Hospital X**

Tempo de Internação ( dias)	Nº de pacientes ( $f_i$ )
10  --- 20	38
20  --- 30	45
30  --- 40	30
40  --- 50	22
50  --- 60	10
60  --- 70	15
Total	160

**Fonte: Divisão de Estatística (Março-1990)**



# Determinação das classes de uma tabela de frequências

- Critério para determinar a quantidade de classes:

$$k = 1 + 3,3 \log(n)$$

- Amplitude das classes

$$a = \frac{\text{maior valor} - \text{menor valor}}{\text{número de classes}}$$

# Exemplo

- Considere os seguintes dados, referentes ao peso de 30 crianças com sete anos, em kg:

13,00	13,63	14,10	14,10	14,70	15,35	15,54	16,00	16,00	16,30
17,40	17,40	17,70	17,70	17,90	17,90	18,20	18,35	19,10	19,30
19,50	19,70	20,00	20,32	20,50	21,45	21,50	22,00	22,25	24,00

Construa uma tabela de freqüências para os dados acima.

# Histograma

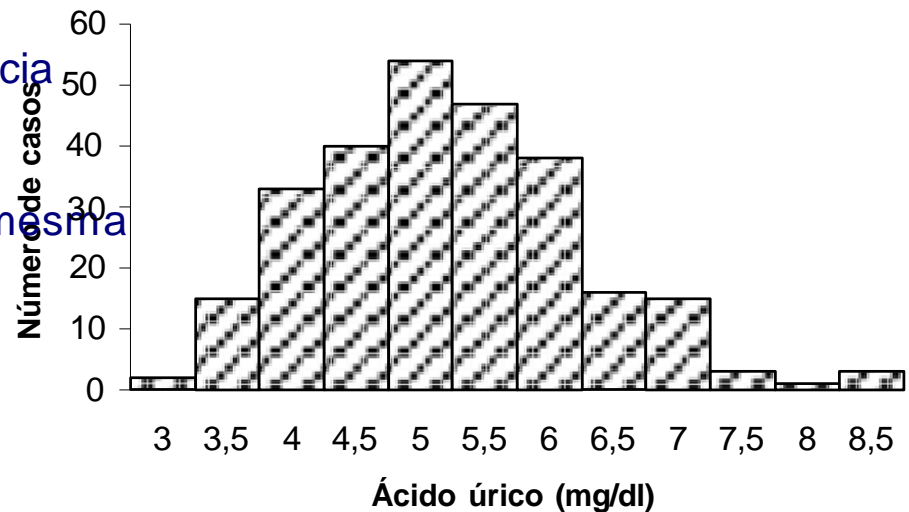
- D Representação gráfica da distribuição das frequências absolutas ou relativas
- D Normalmente utilizado para variáveis contínuas.
- D Características:

- as barras devem estar todas juntas;

- cada barra representa a frequência de um intervalo de valores;

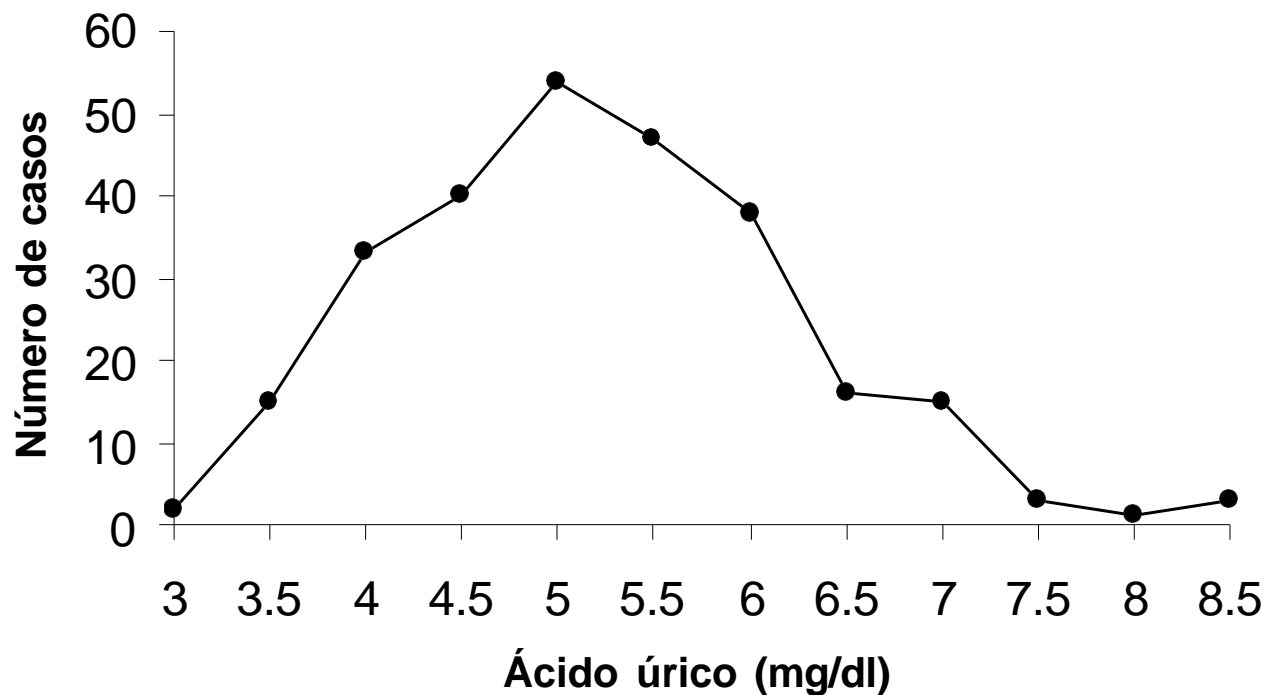
- os intervalos devem ter todos a mesma amplitude.

**FIGURA 3: Ácido úrico dos pacientes internados no Hospital X**



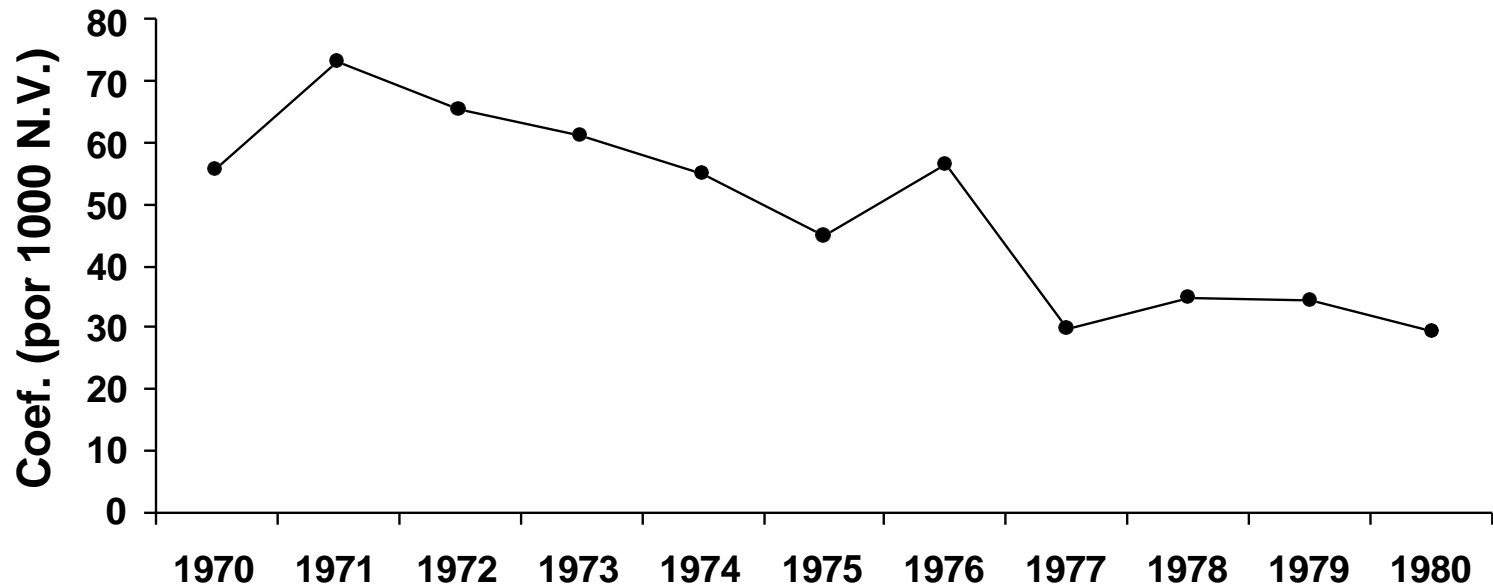
# Polígono de Freqüências

**FIGURA 4: Ácido úrico dos pacientes internados no Hospital X**



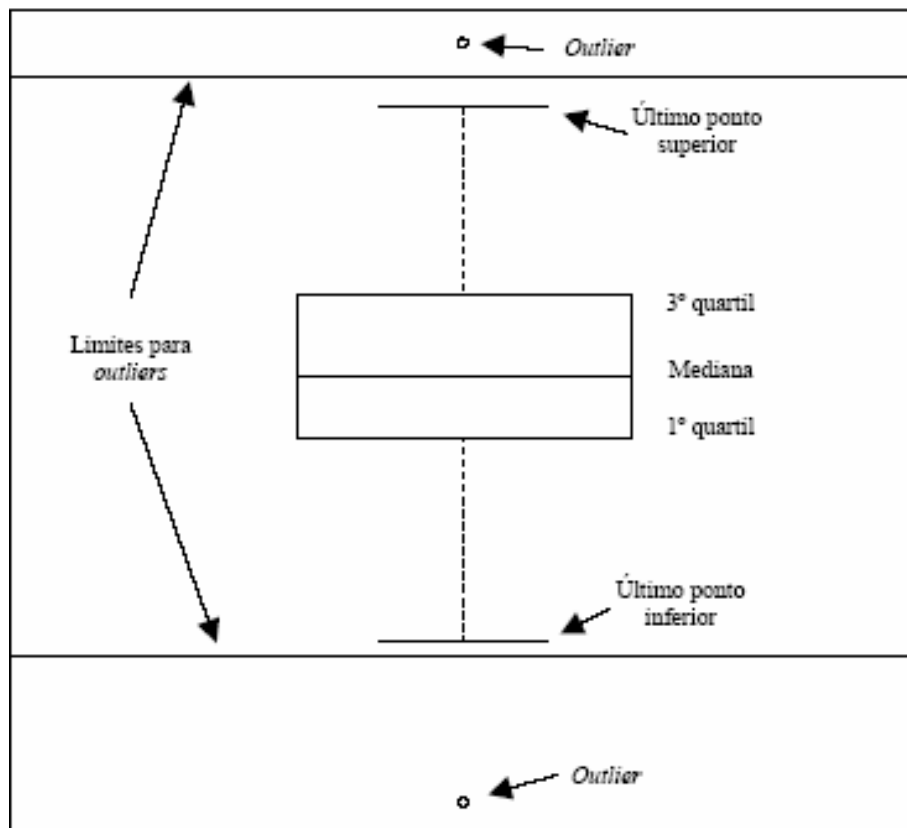
# Gráfico de linhas

**Mortalidade Infantil, São Caetano do Sul (SP), 1970-80**



# Box-Plot

Representação gráfica de cinco medidas: **mínimo**, **quartil inferior**, **mediana**, **quartil superior**, **máximo**



$$\text{limite1} = Q1 - 1,5 \cdot (Q3 - Q1)$$

$$\text{limite2} = Q3 + 1,5 \cdot (Q3 - Q1)$$

# Exemplo Box-Plot

Figura 1: Boxplot do nível de Hemoglobina glicosilada, segundo grupo de gestantes.

