

# Introdução à estatística

**Prof. Anderson Rodrigo da Silva**

`anderson.silva@ifgoiano.edu.br`

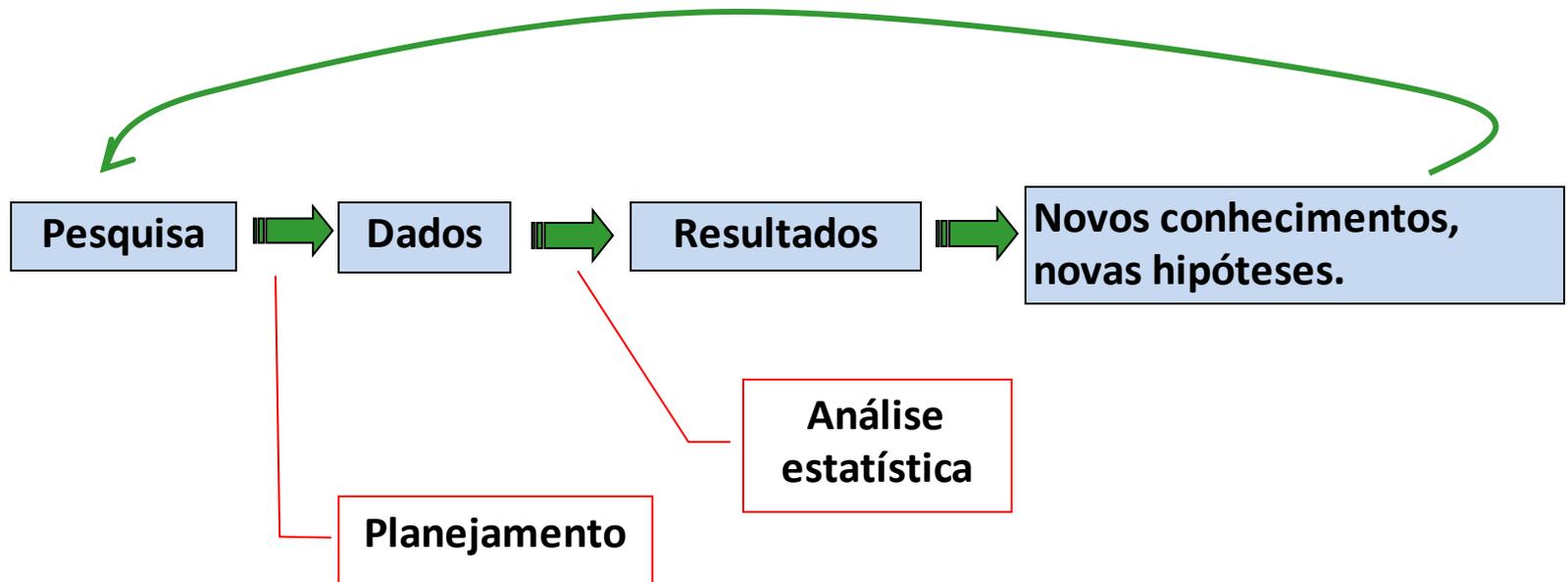
# Sobre Estatística

- Conjunto de conceitos e métodos.
- Ciência que recolhe fatos e fenômenos, analisa-os e interpreta-os quanto ao número de frequência, estabelecendo as leis que os regem.
- **Origem:** vem do latim (*status*).
  - Da necessidade que tinha o Estado Político de conhecer os seus domínios em termos de: número de pessoas (M, F); quantos pagavam impostos; demografia; número de animais; dimensão territorial.

# Importância nas atividades humanas

- **Científica:** pesquisas experimentais e/ou observacionais para adoção de novas práticas ou metodologias. Por exemplo: cultivares, linhagens, insumos agrícolas...
- **Comercial:** importação/exportação, seguros, previsão de safras, estimativas de produção etc.
- **Governamental:** indicadores e taxas. Exemplo: IDH, taxa de natalidade, de mortalidade, de crescimento econômico, atuária, de desemprego etc.

# Metodologia da Pesquisa

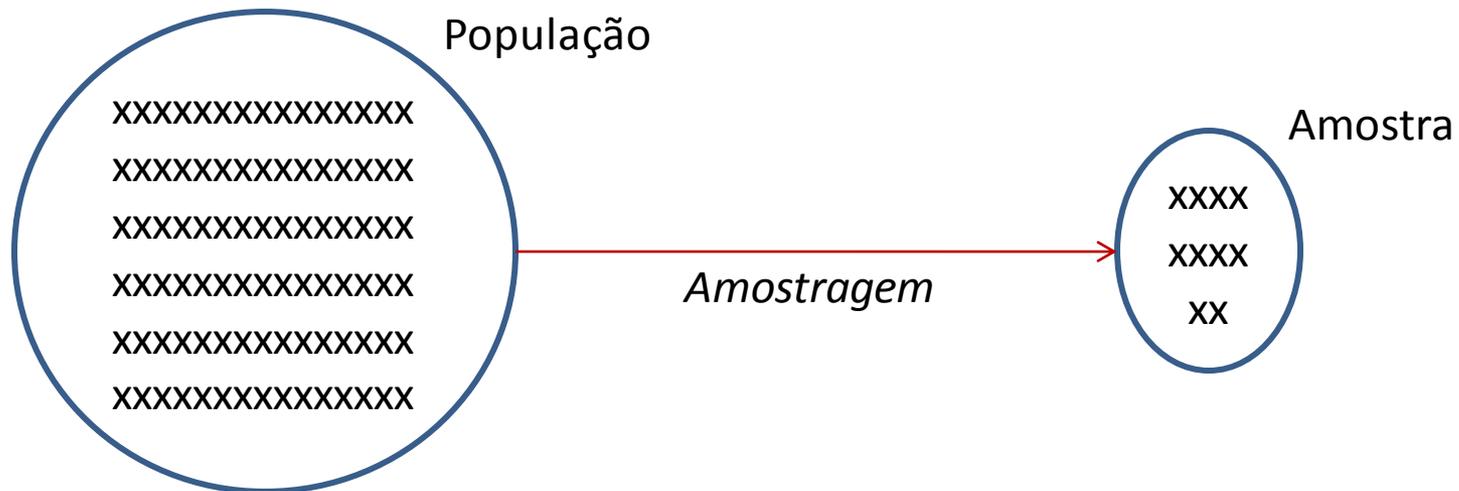


# Definições importantes

- **Variabilidade:** está relacionada ao material experimental. Exemplo: avaliar uma plantação e tomar as medidas: altura de plantas, número de frutos/planta etc. – variáveis!
- **Variável:** é uma característica da unidade observacional. Exemplo: altura da planta, diâmetro do colmo, massa corporal de leitões, cor da pelagem etc.
- **Unidade observacional:** na qual se tomam as medidas das variáveis. Exemplo: uma planta, um vaso, um canteiro, uma vaca, 5 vacas, um rebanho etc.

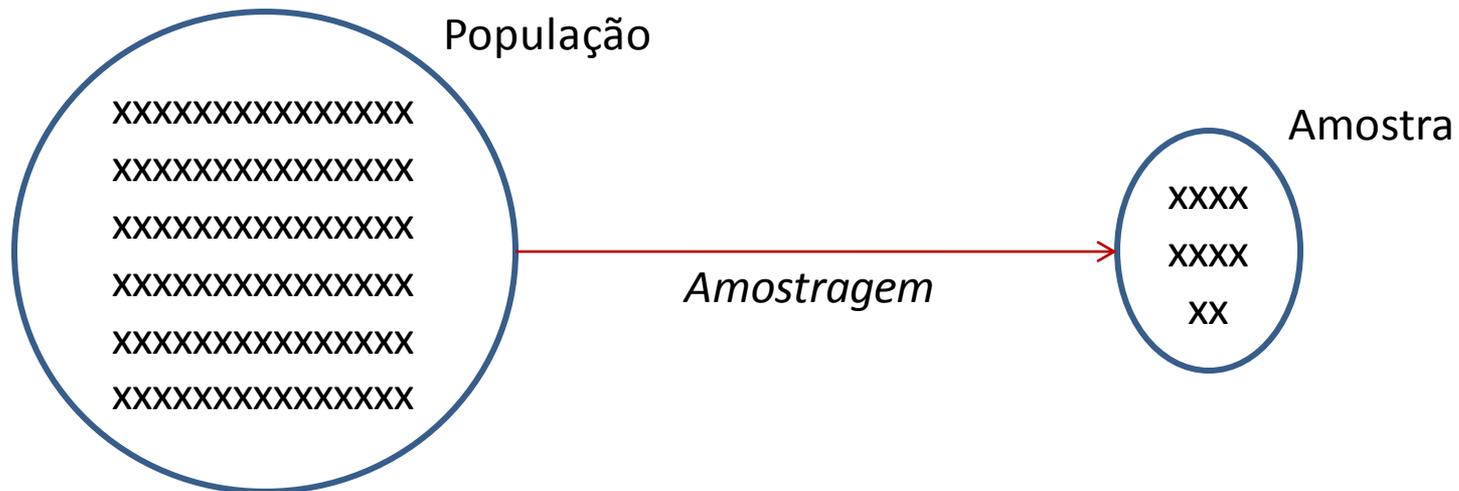
# Definições importantes

- **População:** é o universo estatístico – o conjunto de todos os possíveis valores de uma variável.
- **Amostra:** é uma parte ou subconjunto de uma população.



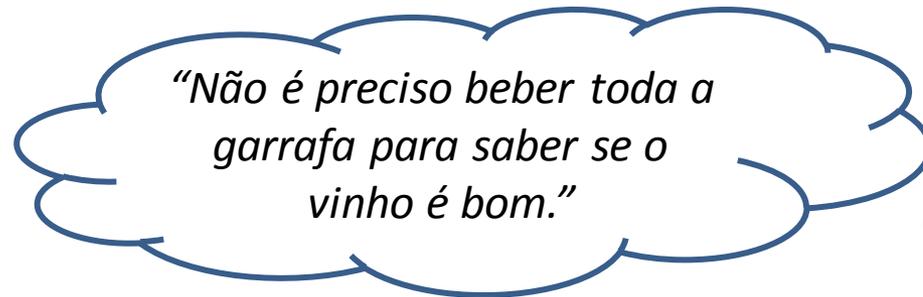
# Definições importantes

- **Amostragem:** é o levantamento que se faz de uma parte da população, a amostra. O ato de tomar uma amostra.
- **Censo:** é o levantamento que se faz sobre a população.

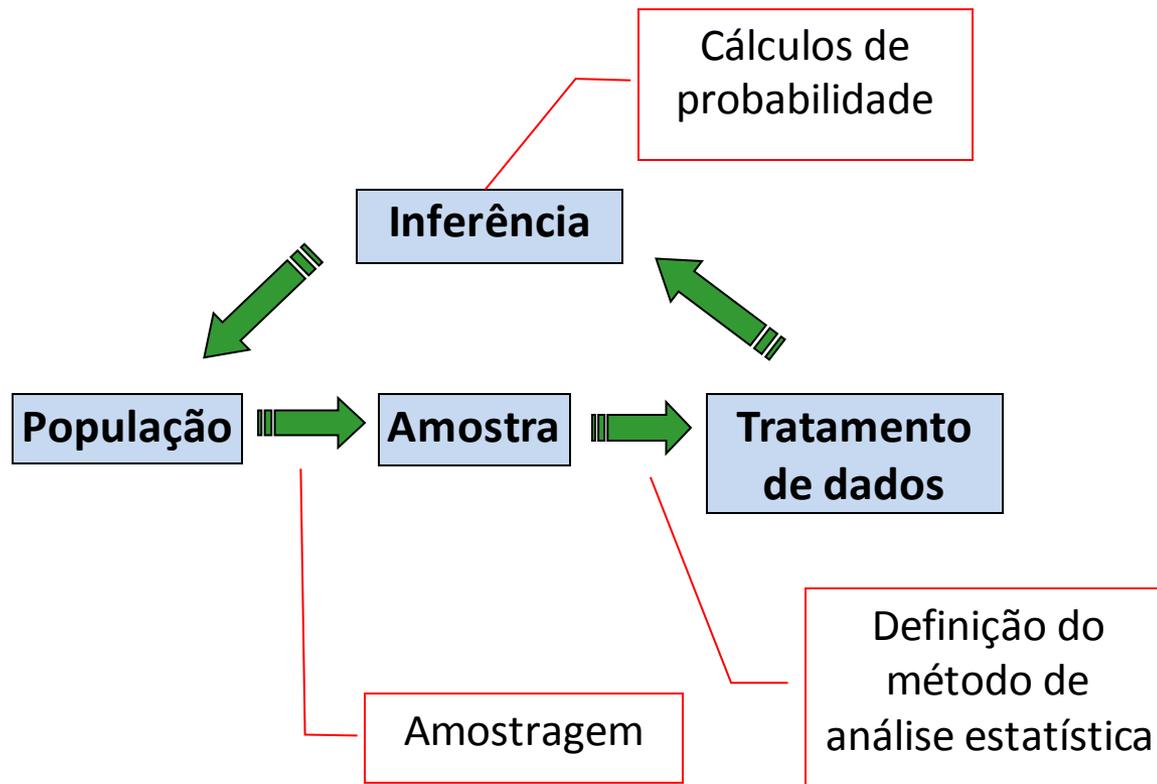


# Definições importantes

- **Inferência:** o ato de concluir algo a respeito da população com base nos dados amostrais.



# Metodologia estatística e o processo de tomada de decisão



# Métodos de análise estatística

- **Exploratória**
  - Gráficos, tabelas
  - Medidas estatísticas descritivas
- **Inferencial**
  - Cálculos de probabilidade
  - Testes de hipóteses
  - Intervalos de confiança

# Sobre a escolha do método

- Qual(is) o(s) objetivo(s) da pesquisa?
- Como foram obtidos os dados? Por simples observação? De um experimento?
- Quais os fatores que estão envolvidos na coleta dos dados? Locais, épocas, tratamentos...
- Quantos dados estão disponíveis para análise?
- Qual(is) o(s) tipo(s) da(s) variável(is)?

# Classificação das variáveis

- **Qualitativas ou categóricas:** Referem-se variáveis não numéricas.
- **Quantitativas:** quantidades medidas numa escala numérica.

# Variáveis qualitativas

Podem ser classificadas como nominais e ordinais.

- **Variáveis nominais:** Não tem ordenamento nem hierarquia. Ex: sexo, raça, doente ou não doente etc.
- **Variáveis ordinais:** São equivalentes as variáveis nominais, porém incluindo uma ordem.
  - Estádio fenológico: VC, V2, R1, R2, R3
  - Severidade de doença: baixa, média, alta

# Variáveis quantitativas

Podem ser do tipo: Discretas e Contínuas.

- **Variáveis Discretas:** variáveis numéricas que assumem somente números inteiros positivos. Em geral surgem de contagem. Ex: número de plantas doentes numa canteiro, quantidade de vacas leiteiras num rebanho, número de partos, número de grãos por vagem etc.

*OBS.: São também discretas as variáveis binárias (0 ou 1).*

- **Variáveis Contínuas:** Assumem qualquer valor dos números reais. Surgem de mensuração. Ex: peso do colmo, altura das plantas, litros de leite produzidos diariamente, consumo mensal de ração etc.

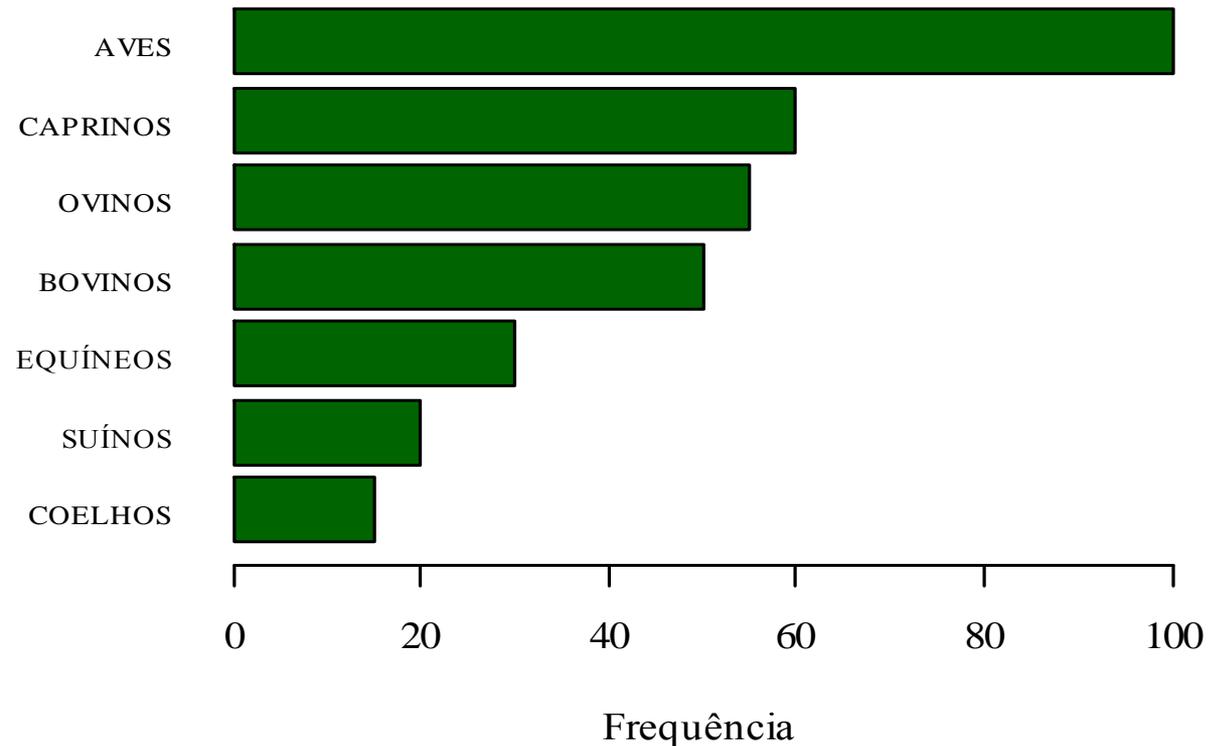
# Distribuição de frequências

**Uma das formas mais simples de se fazer a análise exploratória de uma série de dados é por meio da distribuição de frequências, que pode ser representada por gráficos ou tabelas contendo a frequência absoluta ou relativa de ocorrência de valores de uma determinada categoria ou intervalo.**

# Gráficos para variáveis qualitativas

## Gráfico de barras

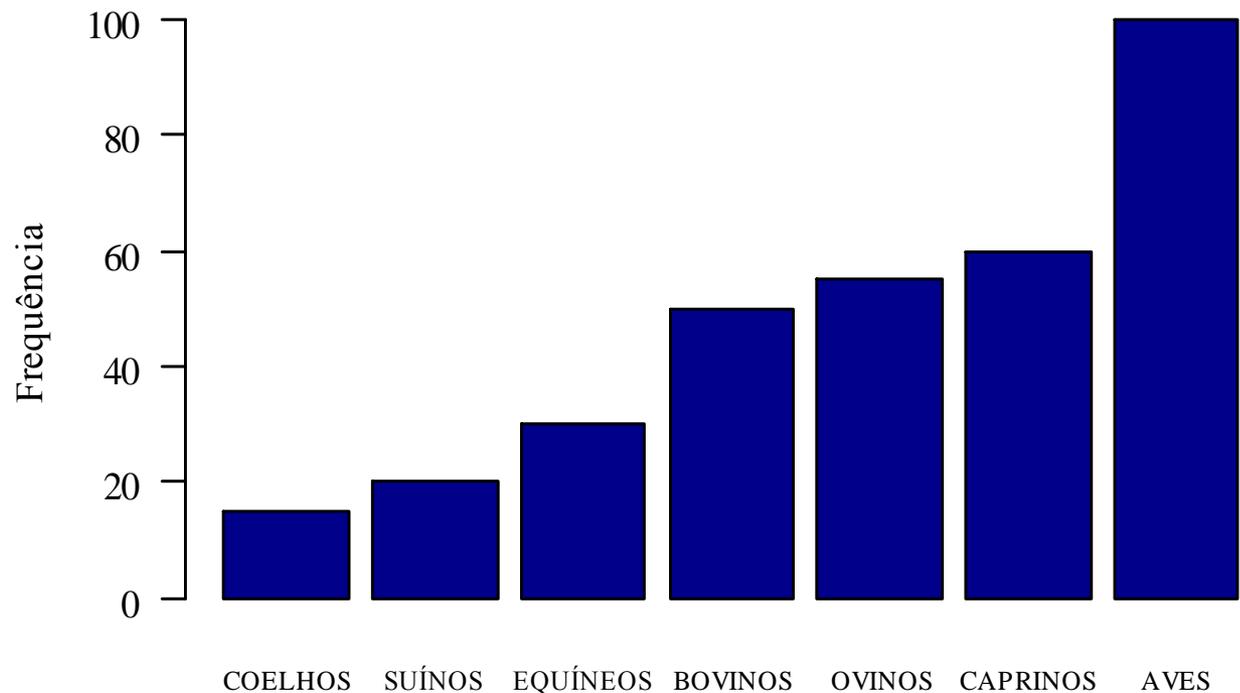
*Exemplo: Gráfico de barras referente a distribuição de frequências **relativas (%)** das espécies de animais exploradas em uma fazenda, em diversas propriedades agrícolas existentes no Estado do Paraná, cuja atividade principal é a pecuária.*



# Gráficos para variáveis qualitativas

## Gráfico de colunas

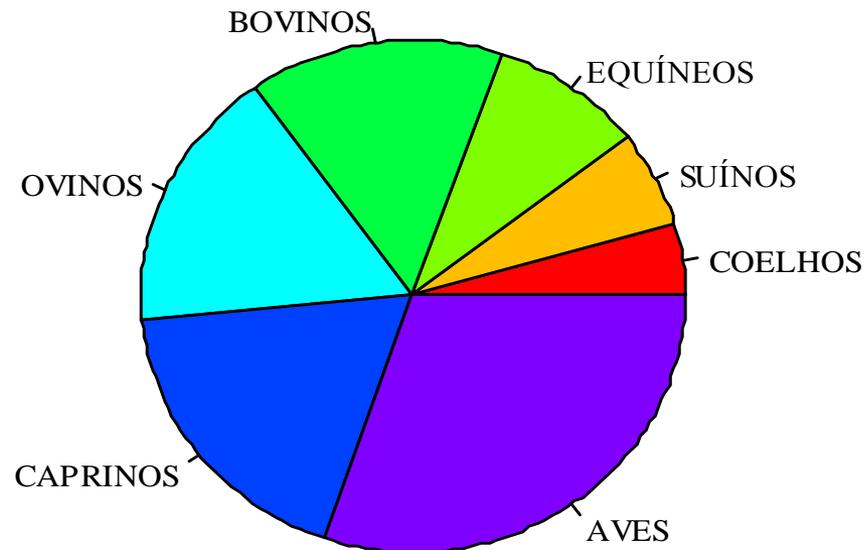
*Exemplo: Gráfico de colunas referente a distribuição de frequências relativas das espécies de animais exploradas em uma fazenda, em diversas propriedades agrícolas existentes no Estado do Paraná, cuja atividade principal é a pecuária.*



# Gráficos para variáveis qualitativas

## Gráfico de setores (“pizza”)

*Exemplo: Gráfico de setores referente a distribuição de frequências relativas das espécies de animais exploradas em uma fazenda, em diversas propriedades agrícolas existentes no Estado do Paraná, cuja atividade principal é a pecuária.*



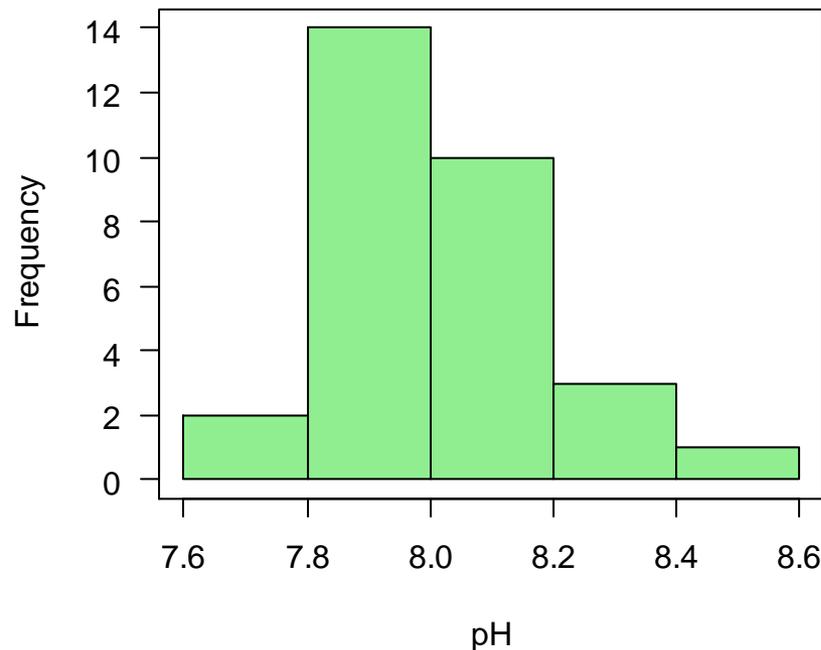
# Gráficos para variáveis quantitativas

EXEMPLO: dados de pH da água em 30 pontos do rio Doubs.

```
7.9 8.0 8.3 8.0 8.1 7.9 8.1 8.1 8.0 7.7 8.1 7.9 8.1  
8.3 8.6 8.0 8.0 8.0 8.1 8.0 7.9 8.1 8.1 8.0 7.9 7.9  
      8.1 8.3 7.8 8.2
```

# Gráficos para variáveis quantitativas

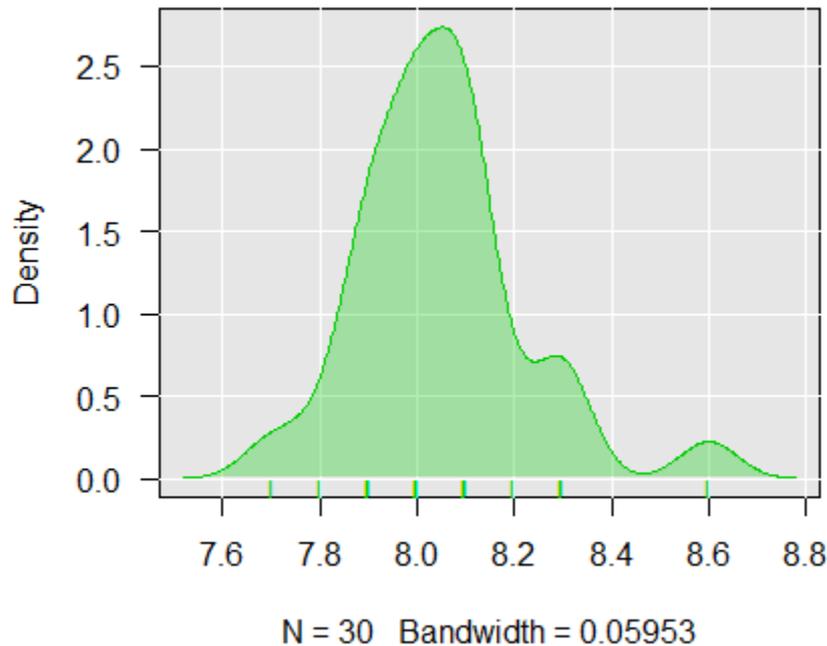
Histograma/polígono de frequências



*Exemplo: Distribuição de frequências do pH da água do rio Doubs*

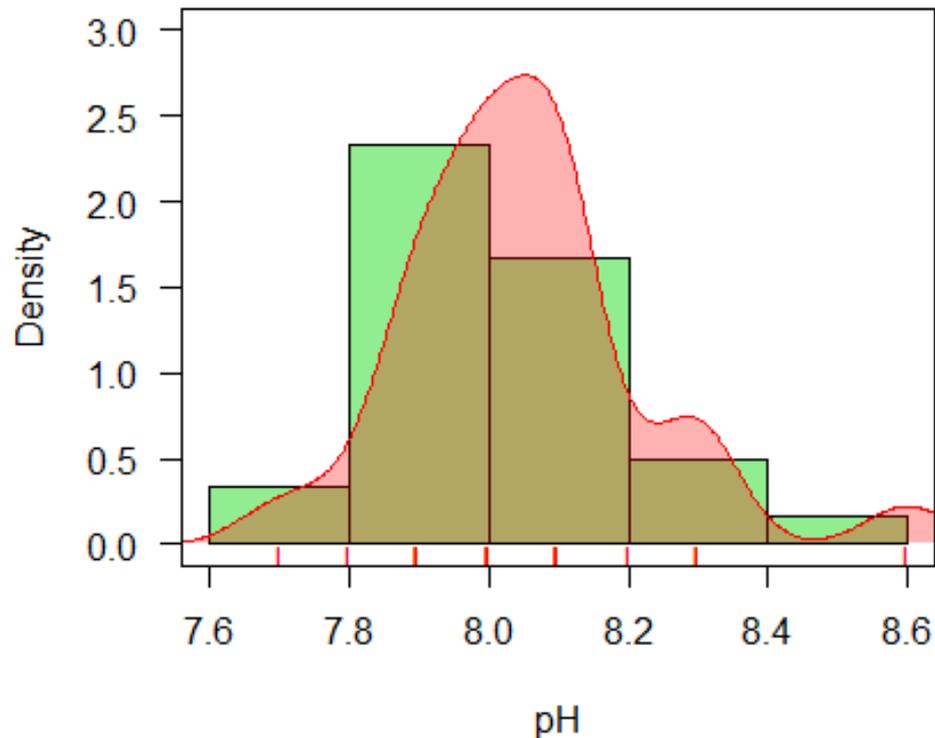
# Gráficos para variáveis quantitativas

## Densidade kernel



*Exemplo: Distribuição suavizada de frequências do pH da água do rio Doubs*

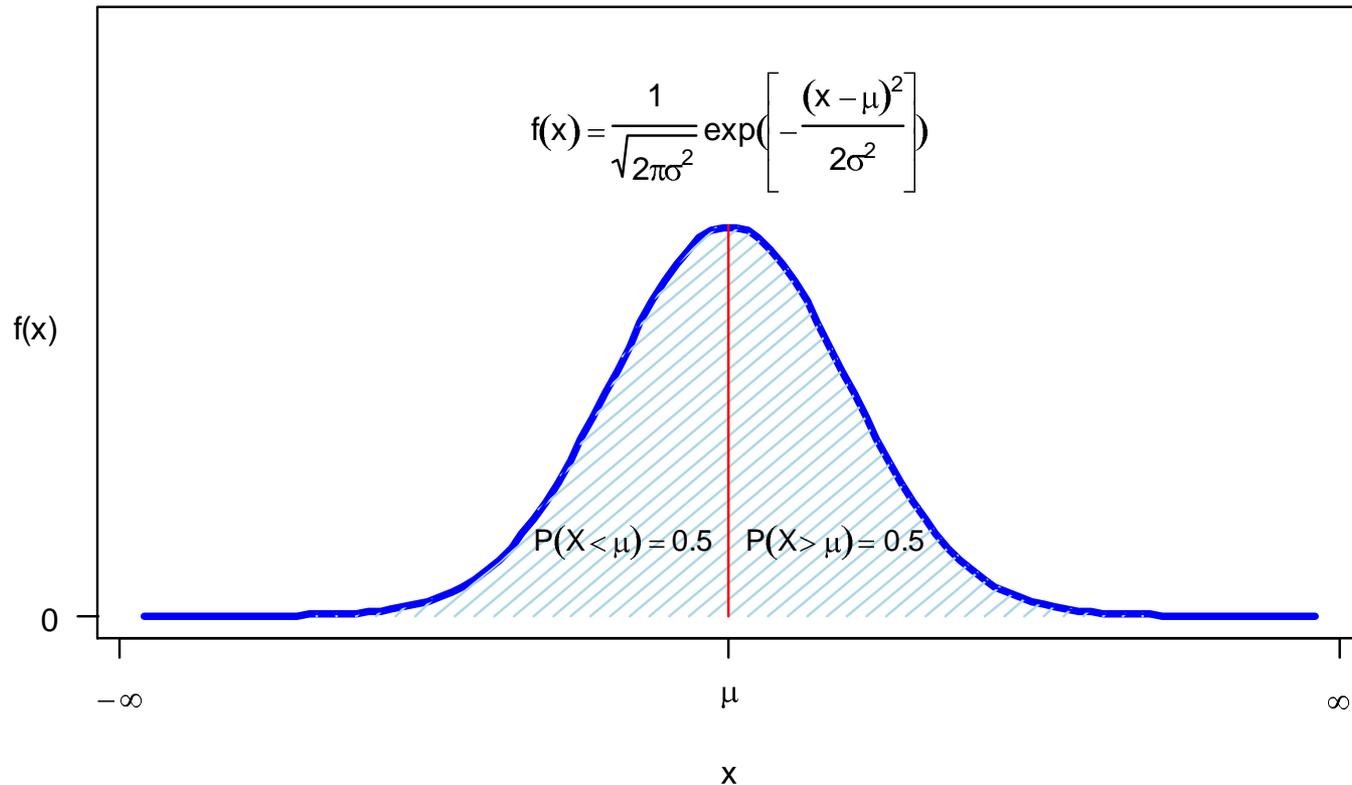
# Gráficos para variáveis quantitativas



# A distribuição Normal de probabilidades

- A “real” distribuição de frequências de muitas das variáveis aleatórias **contínuas** analisadas na pesquisa agrícola pode ser descrita por um modelo de probabilidades já conhecido, o modelo *normal*.
- É o mais importante modelo de distribuição de probabilidades.
- Outros modelos, como qui-quadrado, t-Student e F são derivados da normal.
- O modelo é função de dois parâmetros: média ( $\mu$ ) e variância ( $\sigma^2$ ).

# A distribuição Normal de probabilidades



- Distribuição simétrica em torno de  $\mu$
- Área sob a curva = 1
- Formato da curva determinado totalmente por  $\mu$  e  $\sigma^2$  (ver func. R `normalcurve()`)

# Recursos computacionais

Microsoft Excel® (2007 ou 2010), suplemento *Análise de dados*

- Arquivo >> Opções >> Suplementos >> Ir >> Ferramentas de Análise



## Software R (livre)

- <http://www.r-project.org/> >> CRAN >> Brazil >> USP

# Materiais de estudo



Materiais de estudo, listas de exercícios e *scripts* do R serão disponibilizados na página da disciplina, em: [www.arsilva.weebly.com](http://www.arsilva.weebly.com) >> *Materiais*

# Estatística Descritiva

É a parte da análise exploratória de dados que procura resumir a informação contida nos dados em medidas descritivas, de...

... POSIÇÃO - Ex.: média, mediana, moda

... DISPERSÃO - Ex.: amplitude total, variância, desvio padrão, CV%, AIQ

Suponha realizar um estudo de diversidade em que se deseja computar o nº de espécies vegetais que ocorrem num círculo amostral de raio cinco metros, em 8 locais de uma fitofisionomia. Obteve-se:

5 8 8 7 6 7 59 8

*Quais medidas descritivas você utilizaria para resumir a informação dessa amostra?*