



Pós Graduação *Lato Sensu*

# Planejamento de um Estudo Estatístico

**Prof.:** Wagner Pinheiro  
wagner2235@gmail.com

# Sumário

- 1 Noções e métodos estatísticos
- 2 Planejamento de um estudo estatístico
- 3 Questionário

# Noções e métodos estatísticos

Para entendimento dos conceitos que norteiam a Estatística é necessários que sejam estabelecidos as seguintes definições:

- **População**
- **Amostra**
- **Amostragem**
- **Parâmetro**
- **Estimador**
- **Estimativa**

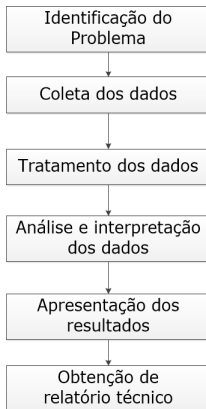
# Noções e métodos estatísticos

**Estatística**: é um conjunto de técnicas utilizadas para coletar, organizar, descrever, analisar e interpretar dados, ou provenientes de experimentos, ou vindos de estudos observacionais.

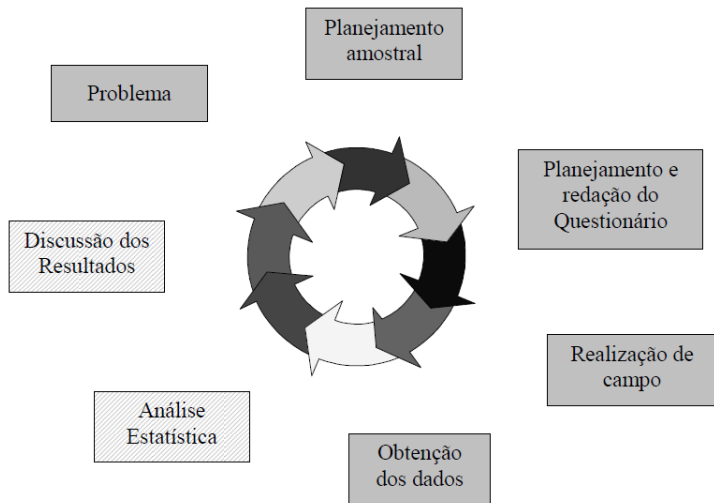
Pode-se dividir a ciência Estatística em dois grupos de estudo



# Planejamento de um estudo estatístico



# Planejamento de um estudo estatístico











## IMPORTANTE

Após a identificação do pesquisador, do pesquisado e dos filtros, são colocadas as questões propriamente ditas do questionário.

## OBSERVAÇÃO

Para elaborar as perguntas de um questionário é indispensável levar em conta que o informante não poderá contar com explicações adicionais do pesquisador. Por este motivo, as perguntas devem ser muito claras e objetivas, para evitar interpretações errôneas, e não devem ser invasivas.



# Estatística Descritiva

Resumo e apresentação de dados

**Prof.:** Wagner Pinheiro  
wagner2235@gmail.com

# Sumário

- 1 Resumo e apresentação de dados
- 2 Distribuição de frequência
  - Elementos de uma distribuição de frequência
  - Representação gráfica de distribuições de frequência

# Medidas Resumo

## Medidas de Tendência Central (Medida de posição ou nível)

i) Média:

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

ii) Mediana:

$$Md = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se par} \end{cases}$$

ii) **Moda:** A moda de um conjunto de observações é o valor de maior ocorrência dentro do conjunto.

# Medidas Resumo

## Medidas de Dispersão (Medida de variabilidade ou de escala

i) Variância ( $s^2$ ):

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

ii) Desvio padrão ( $s$ ):

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

iii) Coeficiente de variação ( $CV$ )

$$CV_{(\%)} = \frac{s}{\bar{X}} \cdot 100\%$$

# Distribuição de frequência

A apresentação da distribuição de frequência de forma tabular é uma apresentação numérica dos dados.

		<b>Título:</b> O que? Quando? Onde?	} <b>cabeçalho</b> O cruzamento de linha com a coluna chama-se casa ou célula.
<b>Corpo</b>	{		
<b>da</b>	{	<b>coluna</b>	
<b>Tabela</b>	{	<b>indicadora</b>	
		<b>Numérica</b>	
<b>Rodapé:</b> fonte, notas, observações.			



Assim uma distribuição de frequência apresenta o conjunto de todos os valores ou particularidades de uma variável, organizados segundo a frequência ou número de ocorrência correspondentes.

$X_i$	Frequência absoluta ( $n_i$ )	Frequência relativa ( $f_i$ )
$x_1$	$n_1$	$f_1 = \frac{n_1}{n}$
$x_2$	$n_2$	$f_2 = \frac{n_2}{n}$
$\vdots$	$\vdots$	
$x_p$	$n_p$	$f_p = \frac{n_p}{n}$
Total	$n$	1

Onde a *frequência absoluta* ( $n_i$ ),  
*frequência relativa* é encontrada pelo quociente entre a frequência absoluta ( $n_i$ ) e o total de observações ( $n$ ).

# Elementos de uma distribuição de frequência

- ① *Rol*: Organizar os dados em ordem crescente ou decresce.
- ② *Amplitude Total da distribuição de Frequência*:

$$AT = \text{máx}\{n\} - \text{mín}\{n\}$$

- ③ *Número de classes de uma distribuição de frequência* ( $K$ ):

$$K = \sqrt{n}$$

- ④ *Intervalo de classe ou amplitude do intervalo de classe* ( $A_c$ ):

$$A_c = \frac{AT}{K}$$

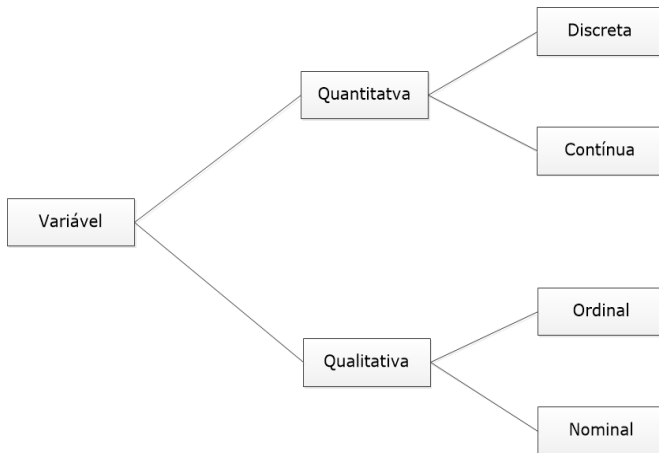
# Representação gráfica - Normas

Cada tipo de gráfico tem indicação específica em função da natureza da variável em estudo.

De modo geral, de acordo com as **normas brasileiras** (ABNT), os gráficos devem possuir os seguintes elementos:

- Todo gráfico deve apresentar título e escala;
- O título deve ser colocado abaixo da ilustração;
- As escalas devem crescer da esquerda para a direita;
- As legendas explicativas devem ser colocadas de preferência à direita da figura;
- Os gráficos devem ser numerados, na ordem em que são citados no texto.

# Representação gráfica - Normas



# Dados QUALITATIVOS

- Gráfico de BARRAS: O gráfico de barras é um excelente recurso para apresentar dados qualitativos coletados em um determinado momento. Por exemplo, respostas de um questionário
- Gráfico de SETORES: Os gráficos de setores são especialmente úteis para mostrar como se dividem o todo. São popularmente conhecidos como **gráfico de pizza**, em razão do seu aspecto, dividido em fatias.
- Gráfico de PARETO: O gráfico de Pareto é um gráfico de colunas que ordena as frequências das ocorrências, da maior para a menor, permitindo a priorização dos problemas, procurando levar a cabo o princípio de Pareto (80% das consequências advêm de 20% das causas), isto é, há muitos problemas sem importância diante de outros mais graves.

# Dados QUALITATIVOS

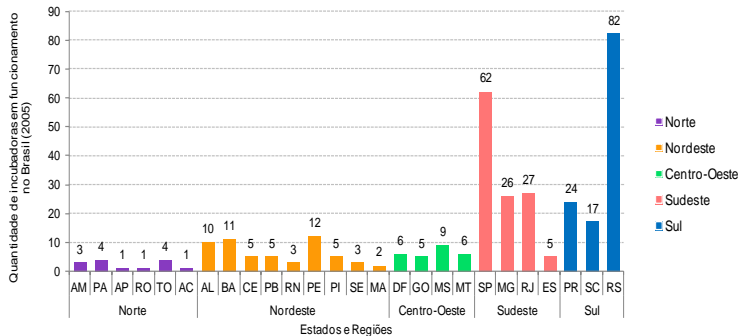
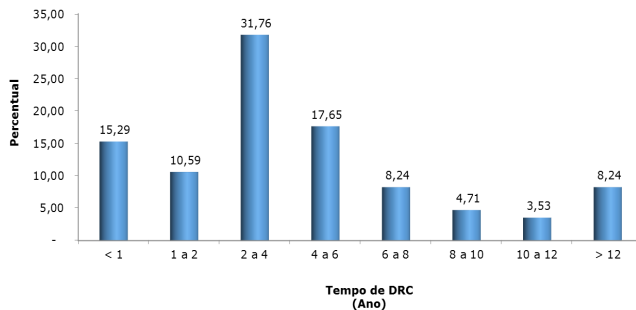
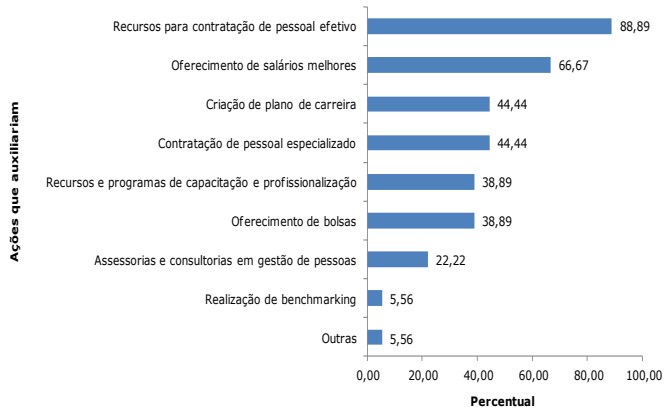


Figure: Quantidade de incubadoras brasileiras por estado e região, no ano de 2005.

# Dados QUALITATIVOS



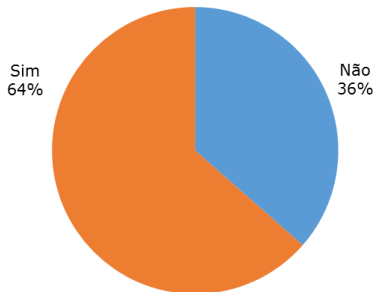
# Dados QUALITATIVOS





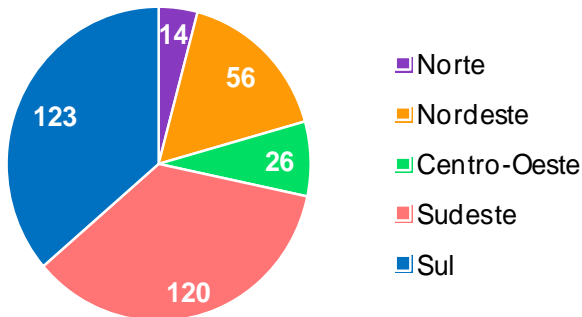
# Dados QUALITATIVOS

**IMPORTANTE:** É mais fácil comparar comprimentos de barras do que ângulos do gráfico de setores. Por isso, desenho o gráfico de setores somente quando o número de categorias for pequeno.

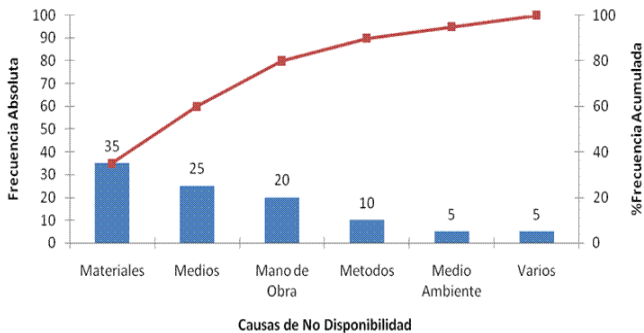


**Figure:** Percentual dos pacientes renais crônico com presença de prurido, no período de fevereiro a março de 2016.

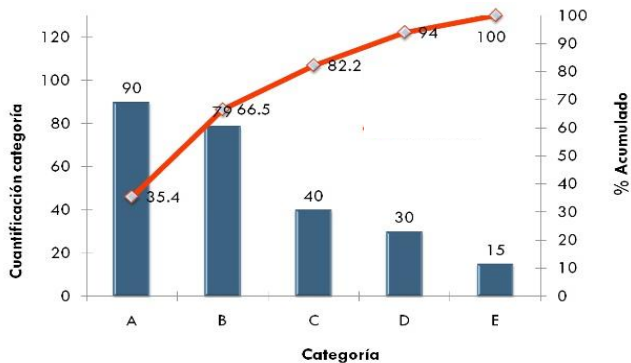
# Dados QUALITATIVOS



# Dados QUALITATIVOS



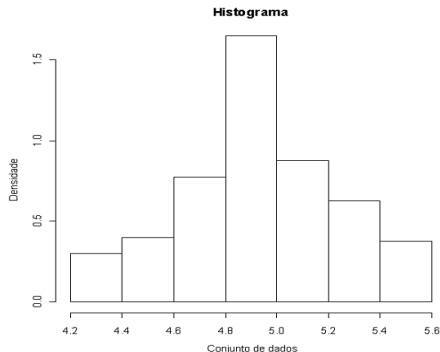
# Dados QUALITATIVOS



# Dados QUANTITATIVOS

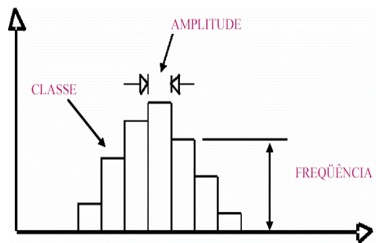
## Histograma:

O histograma é uma forma de mostrar a distribuição dos dados, apresentando-os sob barras justapostas sobre um eixo. Cada barra representa uma classe, ou um grupo de unidade.



# Dados QUANTITATIVOS

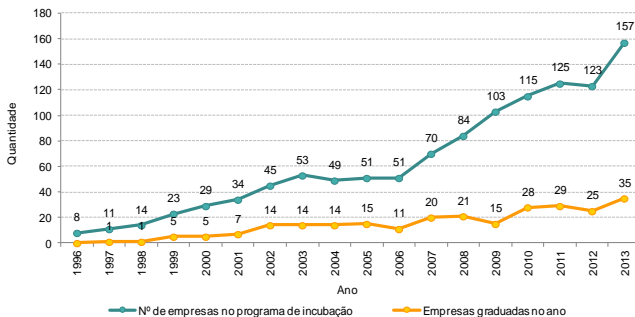
Classes	Frequência
5 ┤ 7	7
7 ┤ 9	5
9 ┤ 11	6
11 ┤ 13	4
13 ┤ 15	8
15 ┤ 17	7
17 ┤ 19	3
Total	40



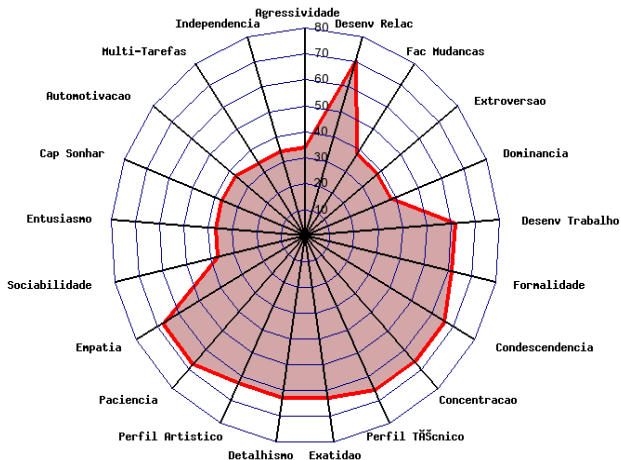
# Dados QUANTITATIVOS

## Série temporal:

Um gráfico de séries temporais é um gráfico que você pode usar para avaliar padrões e comportamento nos dados ao longo do tempo.

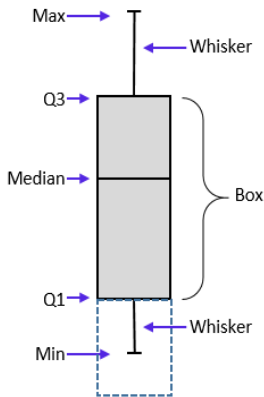


# OUTROS EXEMPLOS





# OUTROS EXEMPLOS



## Exercício

Os dados a seguir representam a temperatura de efluentes em dias consecutivos na descarga de uma estação de tratamento de esgoto.

43	44	44	45	45
46	46	46	47	48
48	49	49	49	49
49	50	50	50	50
51	51	51	52	52

Pede-se:

- Construa a distribuição de frequência.

# Estatística Descritiva

## Medidas de tendencia não central

**Prof.:** Wagner Pinheiro  
wagner2235@gmail.com

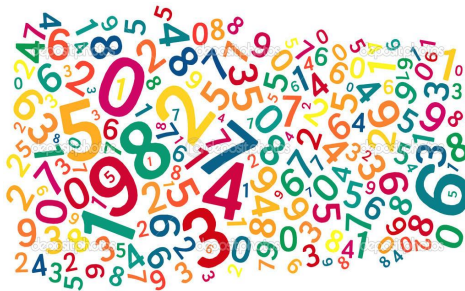
# Sumário

- 1 Medidas de Tendência Não Central
  - Quantis
  - Valores discrepantes (Outliers)
  - Exercícios

# Medidas de Tendência Não Central

Tanto a média quanto o desvio padrão, podem não ser medidas adequadas para representar um conjunto de dados, pois:

- a) são afetados, de forma exagerada, por valores extremos;
- b) apenas com esses dois valores não temos ideia da simetria ou assimetria da distribuição dos dados.



# Quantis

A designação de *quantil* encontra-se associada à ideia de que os quantis dividem a distribuição de frequência em quantidades iguais, isto é, com igual número de observações.

Chamaremos genericamente ao quantil de ordem  $p$ ,  $Q_p$ , ao valor tal que  $100p\%$  dos elementos da amostra são  $\leq Q_p$  e os restantes  $100(1 - p)\%$  são  $\geq Q_p$ .

Os quantis se dividem em:

- Quartil: divide a distribuição de frequência em quatro partes iguais;
- Decil: divide a distribuição de frequência em dez partes iguais;
- Centil: divide a distribuição de frequência em cem partes iguais.

# Quantis

De maneira genérica podemos então entender a formulação para o cálculo das medidas de tendência não central (quartil, decil e centil), onde chama-se de quantil empírico de ordem  $p$ , para  $p$  entre  $(0 < p < 1)$ , ao valor dados por

$$Q_p = \begin{cases} X_{([np]+1)}, & \text{para } np \text{ não inteiro} \\ \frac{X_{(np)} + X_{(np+1)}}{2}, & \text{para } np \text{ inteiro} \end{cases} \quad (1)$$

onde  $[np]$  representa a parte inteira de  $np$ ,  $n$  é o tamanho da amostra e  $p$  é o quantil de interesse a ser localizado na distribuição de frequência (quartil, decil ou centil).

# Quartil

O primeiro quartil,  $Q_1$ , é o número que deixa 25% das observações abaixo e 75% acima, enquanto que o terceiro quartil,  $Q_3$ , deixa 75% das observações abaixo e 25% acima. Já  $Q_2$  é a mediana, deixa 50% das observações abaixo e 50% das observações acima, como apresentado na Figura.

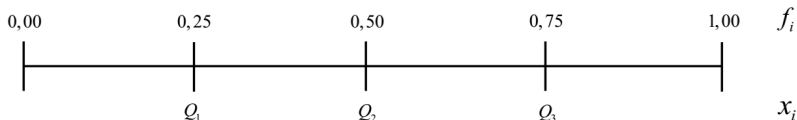


Figure: Representação gráfica da divisão de dados em quartis.



# Quartil

$Q_1$ : Primeiro quartil

$$Q_p = Q_{\frac{1}{4}} = \begin{cases} X_{([\frac{n}{4}]+1)}, & \text{para } \frac{n}{4} \text{ não inteiro} \\ \frac{X_{(\frac{n}{4})} + X_{(\frac{n}{4}+1)}}{2}, & \text{para } \frac{n}{4} \text{ inteiro} \end{cases}$$

$Q_2$ : Segundo quartil

$$Q_p = Q_{\frac{1}{2}} = \begin{cases} X_{([\frac{n}{2}]+1)}, & \text{para } \frac{n}{2} \text{ não inteiro} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & \text{para } \frac{n}{2} \text{ inteiro} \end{cases}$$

$Q_3$ : Terceiro quartil

$$Q_p = Q_{\frac{3}{4}} = \begin{cases} X_{([\frac{3n}{4}]+1)}, & \text{para } \frac{3n}{4} \text{ não inteiro} \\ \frac{X_{(\frac{3n}{4})} + X_{(\frac{3n}{4}+1)}}{2}, & \text{para } \frac{3n}{4} \text{ inteiro} \end{cases}$$

## Valores discrepantes (Outliers)

Para identificação de outliers devemos construir intervalos de valores a partir da amostra, para isso utilizaremos os quartis. Assim, teremos que calcular o intervalo interquartil que é a diferença entre o primeiro e o terceiro quartil, dada por

$$A = Q_3 - Q_1$$

Existem duas formas de outliers, os denominados outlier moderado e o outlier extremo o seu cálculo é dado por

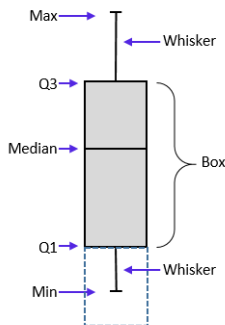
a) **Outlier moderado**

$$\begin{cases} OI = Q_1 - 1,5 \times A \\ OS = Q_3 + 1,5 \times A \end{cases}$$

b) **Outlier extremo**

$$\begin{cases} OI = Q_1 - 3 \times A \\ OS = Q_3 + 3 \times A \end{cases}$$

# Representação gráfica



**Figure:** Representação gráfica dos quartis em uma distribuição de frequência.

# Representação gráfica

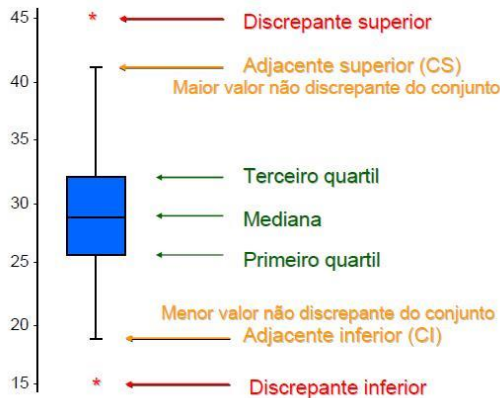


Figure: Representação gráfica dos quartis em uma distribuição de frequência.

# Exercício 1

Os dados a seguir representam a temperatura de efluentes em dias consecutivos na descarga de uma estação de tratamento de esgoto.

43	44	44	45	45
46	46	46	47	48
48	49	49	49	49
49	50	50	50	50
51	51	51	52	52

Calcule:

- a) O primeiro quartil e interprete.
- b) O valor mediano da temperatura de efluentes.
- c) O terceiro quartil e interprete.

## Exercício 2

Suponha que tenhamos os seguintes valores de uma determinada variável  $X$ , são eles: 15; 5; 3; 8; 10; 2; 7; 11; 12. Descreva o procedimento utilizado no cálculo dos quartis e encontre os valores do  $Q_1$ ,  $Q_2$  e  $Q_3$ , utilizando os dados da variável  $X$ .