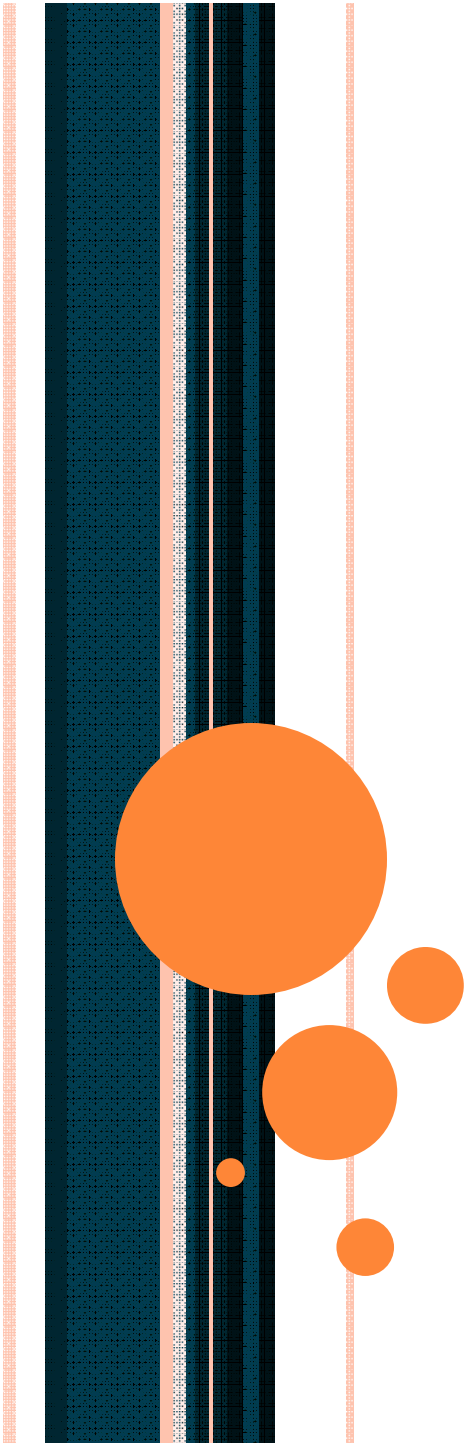


# INTRODUÇÃO À ANÁLISE EXPLORATÓRIA DE DADOS

**Anne Magály de Paula Canuto**

# INTRODUÇÃO

- O que é estatística?

*Estatística é a Ciência que permite obter conclusões a partir de dados*

Paul Velleman

*Conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos realizados em qualquer área do conhecimento*

Magalhães e de Lima

# INTRODUÇÃO

- A análise e a interpretação de dados permitem inferir algo sobre o objeto em estudo (problema, fenômeno, experimento, etc.) do qual os dados foram coletados.
  - Situações onde exista uma grande quantidade de informações
  - Exemplo: as transmissões esportivas: a interpretação dos dados podem ajudar a concluir qual foi o melhor time.

# INTRODUÇÃO

- A estatística pode ser dividida em três áreas
  - Estatística descritiva
  - Probabilidade
  - Inferência estatística (estatística inferencial)

# ESTATÍSTICA DESCRITIVA

- Responsável pela organização, descrição, apresentação e resumo dos dados.
  - Utilizam gráficos, tabelas e medidas descritivas como ferramentas.
- Utilizada na etapa inicial da análise
  - Obter informações (conclusões) a respeito de características dos dados

# Probabilidade e Estatística inferencial

- Probabilidade: a teoria matemática que estuda a incerteza oriunda de fenômenos de caráter aleatório.
- Estatística inferencial: o estudo de técnicas para extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas a partir de um subconjunto dos dados
- Para a inferirmos a partir dos dados, fazemos algumas suposições sobre as chances (ou probabilidades) de obtermos valores de dados diferentes.
  - O conjunto dessas suposições é denominado modelo de probabilidade.

# CONCEITOS PRELIMINARES

- Na terminologia estatística, o grande conjunto de dados que contém a característica que temos interesse recebe o nome de *população.*

- Esse termo refere-se não somente a uma coleção de indivíduos, mas também ao alvo sobre o qual reside nosso interesse.

  - Exemplos: todos os habitantes de Natal,  todas as lâmpadas produzidas por uma fábrica em certo período de tempo.

# Conceitos preliminares

- Algumas vezes podemos acessar toda a população para estudarmos características de interesse

- Em muitas situações, tal procedimento não pode ser realizado.

  - Em geral, razões econômicas são determinantes dessas situações a impossibilidade de se acessar toda a população de interesse é incontornável.

  - Como podemos fazer uma pesquisa qualquer com a população brasileira (ou de qualquer outro país)?

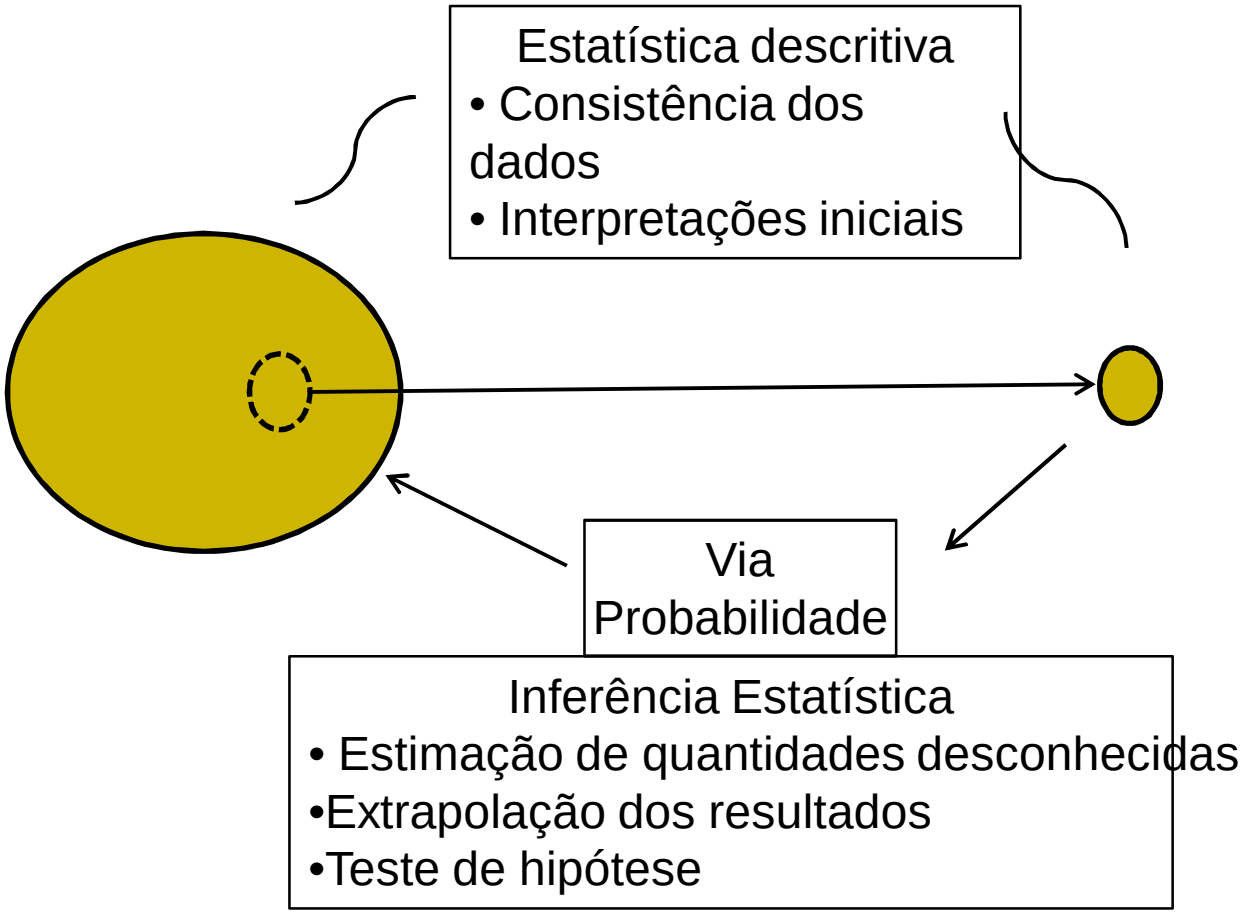
# Conceitos preliminares

- Tendo em vista as dificuldades de várias naturezas para se observar todos os elementos da população, tomaremos alguns deles para formar um grupo a ser estudado.

- Este subconjunto da população, em geral com dimensão menor, é denominado *amostra.*

- Porém, precisamos tomar muito cuidado com esta amostra: precisa ser representativa

# CONCEITOS PRELIMINARES



# Conceitos preliminares

- Como selecionar as amostras?
  - Objetivo: fornecer um subconjunto de valores o mais parecido possível com a população original
- Principais métodos
  - Seleção aleatória
  - Amostragem  estratificada
  - Amostragem sistemática
  - Outros métodos encontrados na *Teoria das amostragens*

# Organização dos dados

- Dado um conjunto de dados, como tratar os valores para se extrair informações importantes?
  - Técnicas da estatística, funções e gráficos
- Exemplo: suponha que um questionário seja aplicado a vocês solicitando as informações:
  - Id: Identificação do aluno
  - Turma a que o aluno foi alocado
  - Gênero: masculino ou feminino
  - Idade: idade em anos

# ORGANIZAÇÃO DOS DADOS

- Alt: altura em metros
- Peso: peso em quilogramas
- Filhos: número de filhos na família
- Fuma: sim ou não
- Toler: tolerância ao cigarro (I – indiferente, P – incomoda um pouco, M – incomoda muito)
- Exerc: horas de atividades fisica por semana
- Cine: Numero de vezes que vai ao cinema por semana
- OpCine: Opinião sobre as salas de cinema (B ou MB)
- TV: horas por semana na televisão
- OptTV: opinião sobre a programação da TV aberta (Ruim, Média, Boa, Não sabe)

| Id | Turma | Sexo | Idade | Alt | Peso | Filhos | Fuma | Toler | Exerc | Cine | OpCine | TV | OpTV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | F | 17 | 1,60 | 60,5 | 2 | NAO | P | 0 | 1 | B | 16 | R |
| 2 | A | F | 18 | 1,69 | 55,0 | 1 | NAO | M | 0 | 1 | B | 7 | R |
| 3 | A | M | 18 | 1,85 | 72,8 | 2 | NAO | P | 5 | 2 | M | 15 | R |
| 4 | A | M | 25 | 1,85 | 80,9 | 2 | NAO | P | 5 | 2 | B | 20 | R |
| 5 | A | F | 19 | 1,58 | 55,0 | 1 | NAO | M | 2 | 2 | B | 5 | R |
| 6 | A | M | 19 | 1,76 | 60,0 | 3 | NAO | M | 2 | 1 | B | 2 | R |
| 7 | A | F | 20 | 1,60 | 58,0 | 1 | NAO | P | 3 | 1 | B | 7 | R |
| 8 | A | F | 18 | 1,64 | 47,0 | 1 | SIM | I | 2 | 2 | M | 10 | R |
| 9 | A | F | 18 | 1,62 | 57,8 | 3 | NAO | M | 3 | 3 | M | 12 | R |
| 10 | A | F | 17 | 1,64 | 58,0 | 2 | NAO | M | 2 | 2 | M | 10 | R |
| 11 | A | F | 18 | 1,72 | 70,0 | 1 | SIM | I | 10 | 2 | B | 8 | N |
| 12 | A | F | 18 | 1,66 | 54,0 | 3 | NAO | M | 0 | 2 | B | 0 | R |
| 13 | A | F | 21 | 1,70 | 58,0 | 2 | NAO | M | 6 | 1 | M | 30 | R |
| 14 | A | M | 19 | 1,78 | 68,5 | 1 | SIM | I | 5 | 1 | M | 2 | N |
| 15 | A | F | 18 | 1,65 | 63,5 | 1 | NAO | I | 4 | 1 | B | 10 | R |
| 16 | A | F | 19 | 1,63 | 47,4 | 3 | NAO | P | 0 | 1 | B | 18 | R |
| 17 | A | F | 17 | 1,82 | 66,0 | 1 | NAO | P | 3 | 1 | B | 10 | N |
| 18 | A | M | 18 | 1,80 | 85,2 | 2 | NAO | P | 3 | 4 | B | 10 | R |
| 19 | A | F | 20 | 1,60 | 54,5 | 1 | NAO | P | 3 | 2 | B | 5 | R |
| 20 | A | F | 18 | 1,68 | 52,5 | 3 | NAO | M | 7 | 2 | B | 14 | M |
| 21 | A | F | 21 | 1,70 | 60,0 | 2 | NAO | P | 8 | 2 | B | 5 | R |
| 22 | A | F | 18 | 1,65 | 58,5 | 1 | NAO | M | 0 | 3 | B | 5 | R |
| 23 | A | F | 18 | 1,57 | 49,2 | 1 | SIM | I | 5 | 4 | B | 10 | R |
| 24 | A | F | 20 | 1,55 | 48,0 | 1 | SIM | I | 0 | 1 | M | 28 | R |
| 25 | A | F | 20 | 1,69 | 51,6 | 2 | NAO | P | 8 | 5 | M | 4 | N |
| 26 | A | F | 19 | 1,54 | 57,0 | 2 | NAO | I | 6 | 2 | B | 5 | R |
| 27 | B | F | 23 | 1,62 | 63,0 | 2 | NAO | M | 8 | 2 | M | 5 | R |
| 28 | B | F | 18 | 1,62 | 52,0 | 1 | NAO | P | 1 | 1 | M | 10 | R |
| 29 | B | F | 18 | 1,57 | 49,0 | 2 | NAO | P | 3 | 1 | B | 12 | R |
| 30 | B | F | 25 | 1,65 | 59,0 | 4 | NAO | M | 1 | 2 | M | 2 | R |
| 31 | B | F | 18 | 1,61 | 52,0 | 1 | NAO | P | 2 | 2 | M | 6 | N |
| 32 | B | M | 17 | 1,71 | 73,0 | 1 | NAO | P | 1 | 1 | B | 20 | R |
| 33 | B | F | 17 | 1,65 | 56,0 | 3 | NAO | M | 2 | 1 | B | 14 | R |
| 34 | B | F | 17 | 1,67 | 58,0 | 1 | NAO | M | 4 | 2 | B | 10 | R |
| 35 | B | M | 18 | 1,73 | 87,0 | 1 | NAO | M | 7 | 1 | B | 25 | B |
| 36 | B | F | 18 | 1,60 | 47,0 | 1 | NAO | P | 5 | 1 | M | 14 | R |
| 37 | B | M | 17 | 1,70 | 95,0 | 1 | NAO | P | 10 | 2 | M | 12 | N |
| 38 | B | M | 21 | 1,85 | 84,0 | 1 | SIM | I | 6 | 4 | B | 10 | R |
| 39 | B | F | 18 | 1,70 | 60,0 | 1 | NAO | P | 5 | 2 | B | 12 | R |
| 40 | B | M | 18 | 1,73 | 73,0 | 1 | NAO | M | 4 | 1 | B | 2 | R |
| 41 | B | F | 17 | 1,70 | 55,0 | 1 | NAO | I | 5 | 4 | B | 10 | B |
| 42 | B | F | 23 | 1,45 | 44,0 | 2 | NAO | M | 2 | 2 | B | 25 | R |
| 43 | B | M | 24 | 1,76 | 75,0 | 2 | NAO | I | 7 | 0 | M | 14 | N |
| 44 | B | F | 18 | 1,68 | 55,0 | 1 | NAO | P | 5 | 1 | B | 8 | R |
| 45 | B | F | 18 | 1,55 | 49,0 | 1 | NAO | M | 0 | 1 | M | 10 | R |
| 46 | B | F | 19 | 1,70 | 50,0 | 7 | NAO | M | 0 | 1 | B | 8 | R |
| 47 | B | F | 19 | 1,55 | 54,5 | 2 | NAO | M | 4 | 3 | B | 3 | R |
| 48 | B | F | 18 | 1,60 | 50,0 | 1 | NAO | P | 2 | 1 | B | 5 | R |
| 49 | B | M | 17 | 1,80 | 71,0 | 1 | NAO | P | 7 | 0 | M | 14 | R |
| 50 | B | M | 18 | 1,83 | 86,0 | 1 | NAO | P | 7 | 0 | M | 20 | B |

Informações dos questionários – Dados brutos

# ORGANIZAÇÃO DOS DADOS - VARIÁVEIS

- **VARIÁVEIS QUALITATIVAS**
  - A variável assume “valores” em categorias, classes ou rótulos
  - São dados não numéricos.
  - Oferece um vasto espectro de aplicação nas ciências sociais e do comportamento
    - É considerada de baixo nível de mensuração, do ponto de vista da aplicação de instrumental estatístico
  - Denotam características individuais das unidades sob análise
    - Exemplo:  sexo, estado civil, naturalidade, raça, grau de instrução, dentre outras
  - Permite estratificar as unidades para serem analisadas de acordo com outras variáveis

# Organização dos dados - variáveis

- Classificação
  - Ordinal: Existe uma ordenação natural  que indica intensidades crescentes
    - Exemplo: Tamanho, classe social, entre outros
  - Nominal: Não é possível estabelecer uma ordem natural entre seus valores
    - Exemplo: turma, gênero, Fuma, entre outros

# ORGANIZAÇÃO DOS DADOS - VARIÁVEIS

- **VARIÁVEIS QUANTITATIVAS**
  - Variáveis de natureza numéricas
- Classificação
  - Discretas: resultante de contagens (valores inteiros) e com conjuntos de valores finito e enumerável
    - Exemplo: Número de filhos, número de defeitos
  - Contínuas: assumem valores em intervalos de números (mensuração)
    - Exemplo: Altura, peso, etc.
- Classificação depende de particularidades das variáveis: Idade pode ser discreta ou contínua

# Organização dos dados

- Os dados vem sempre bem “organizados” como os mostrados na tabela mostrada?
- O que precisamos fazer para deixar os dados bem organizados como os da tabela?
- Preparação dos dados

# Por que preparar os dados?

- Os dados estão sujos
  - Incompletos
    - Ausência de atributos de interesse
    - Apenas dados agregados
    - Ausência de valores
  - Ruidosos
    - Erros aleatórios
    - Valores aberrantes (outliers)
  - Inconsistentes
    - Discrepâncias nas codificações ou nos nomes
- Sem dados de boa qualidade o raciocínio indutivo é pobre

# Etapas da preparação

- Limpeza dos dados
  - preencher dados ausentes, “alisar” ruído, identificar e/ou remover valores aberrantes, resolver inconsistências
- Integração e transformação de Dados
  - integração de múltiplas bases de dados, cubos e arquivos
  - Normalização e agregação
- Redução de Dados
  - redução no volume de dados com resultados similares

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Algumas técnicas que podemos utilizar para trabalhar com os dados brutos
  - Depende do tipo de variável

| variável qualitativa* | variável quantitativa |
|---|---|
| tabela de frequências<br>gráfico de barras<br>diagrama circular (pizza) | medidas de posição: média, mediana, moda<br>medidas de dispersão: variância, desvio-padrão, amplitude, coeficiente de variação<br>tabela de frequências<br>histograma<br>boxplot<br>gráfico de linha ou sequência<br>polígono de frequências |

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Um método simples e eficaz: a tabela de frequências
- P[illegible] qualitativas ou discretas

| Gênero | $n_i$ | $f_i$ |
|---|---|---|
| F | 37 | 0,74 |
| M | 13 | 0,26 |
| Total | n=50 | 1 |

- Para as variáveis ordinais: frequências acumuladas

| Idade | $n_i$ | $f_i$ | $f_{ac}$ |
|---|---|---|---|
| 17 | 9 | 0,18 | 0,18 |
| 18 | 22 | 0,44 | 0,62 |
| 19 | 7 | 0,14 | 0,76 |
| 20 | 4 | 0,08 | 0,84 |
| 21 | 3 | 0,06 | 0,9 |
| 22 | 0 | 0 | 0,9 |
| 23 | 2 | 0,04 | 0,94 |
| 24 | 1 | 0,02 | 0,96 |
| 23 | 2 | 0,04 | 1,00 |
| Total | n=50 | 1 | |

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Para variáveis quantitativas contínuas: intervalo de valores

| Peso | $n_i$ | $f_i$ | $f_{ac}$ |
|---|---|---|---|
| 40 ⊢ 50 | 8 | 0,16 | 0,16 |
| 50 ⊢ 60 | 22 | 0,44 | 0,60 |
| 60 ⊢ 70 | 8 | 0,16 | 0,76 |
| 70 ⊢ 80 | 6 | 0,12 | 0,88 |
| 80 ⊢ 90 | 5 | 0,10 | 0,98 |
| 90 ⊢ 100 | 1 | 0,02 | 1,00 |
| Total | n=50 | 1 | |

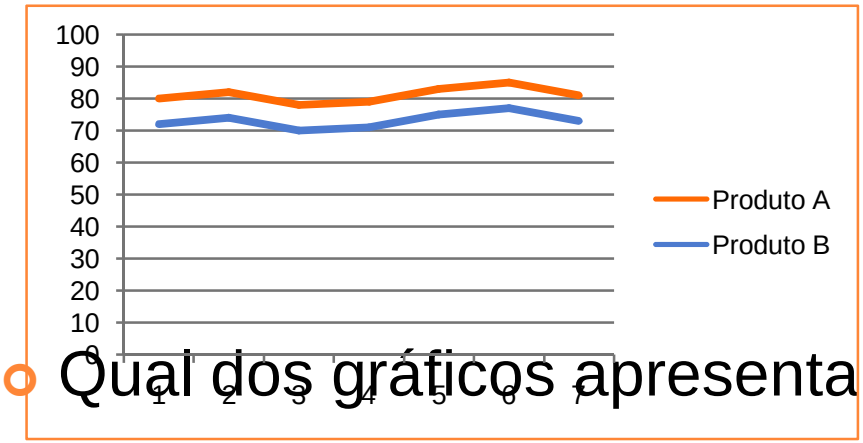
- Podemos usar com variáveis discretas com conjunto de valores grandes. Então como ficaria a tabela com a variável TV?

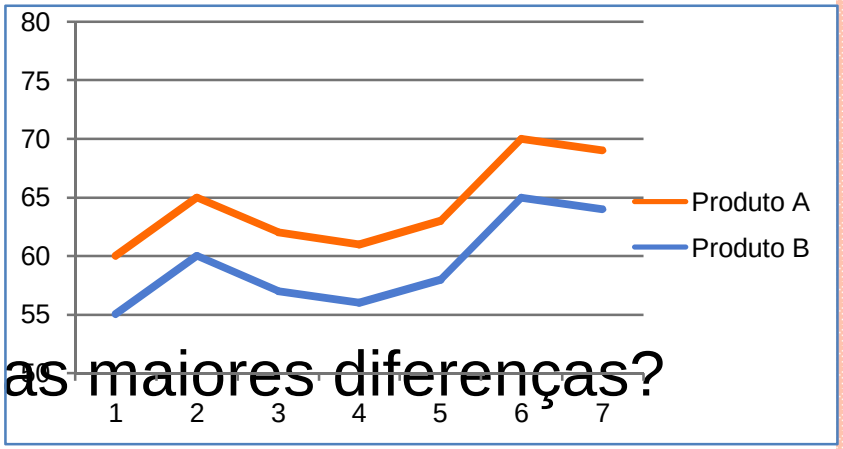
# COMO TRABALHAR COM OS DADOS BRUTOS?

- Podemos utilizar uma outra técnica eficiente: gráficos
- Os gráficos constituem uma das formas mais eficientes de apresentação de dados.
- Um gráfico é, essencialmente, uma figura constituída a partir de uma tabela, pois é quase sempre possível localizar um dado tabulado num gráfico.

# Como trabalhar com os dados brutos?

- Podemos utilizar uma infinidade de gráficos
  - Cuidado: gráficos desproporcionais podem nos dar uma falsa impressão das coisas




- Qual dos gráficos apresentam as maiores diferenças?

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Três tipos básicos de gráficos: disco ou pizza, barras e histogramas
- Os gráficos de pizza, diagrama circular, se adapta muito bem as variáveis qualitativas (porcentagem)

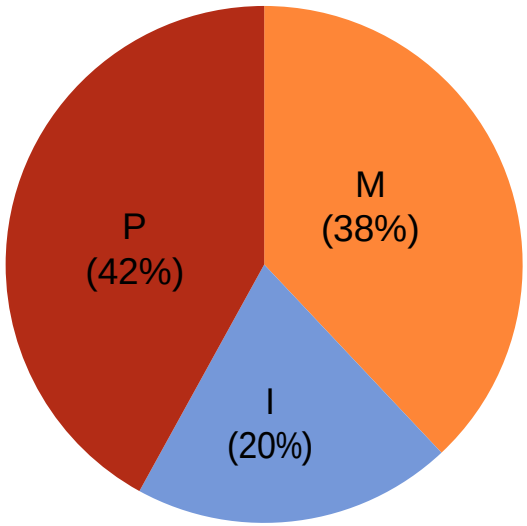
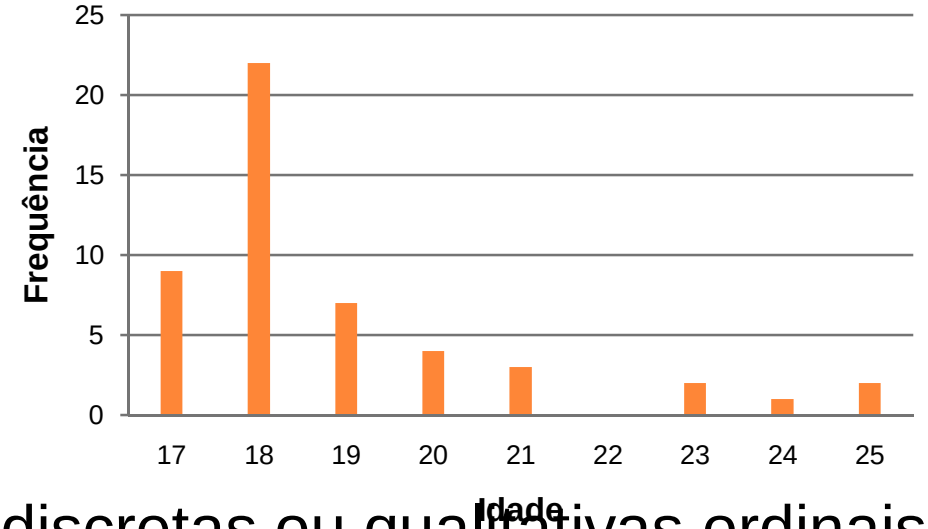



Gráfico para a variável Toler

# COMO TRABALHAR COM OS DADOS BRUTOS?

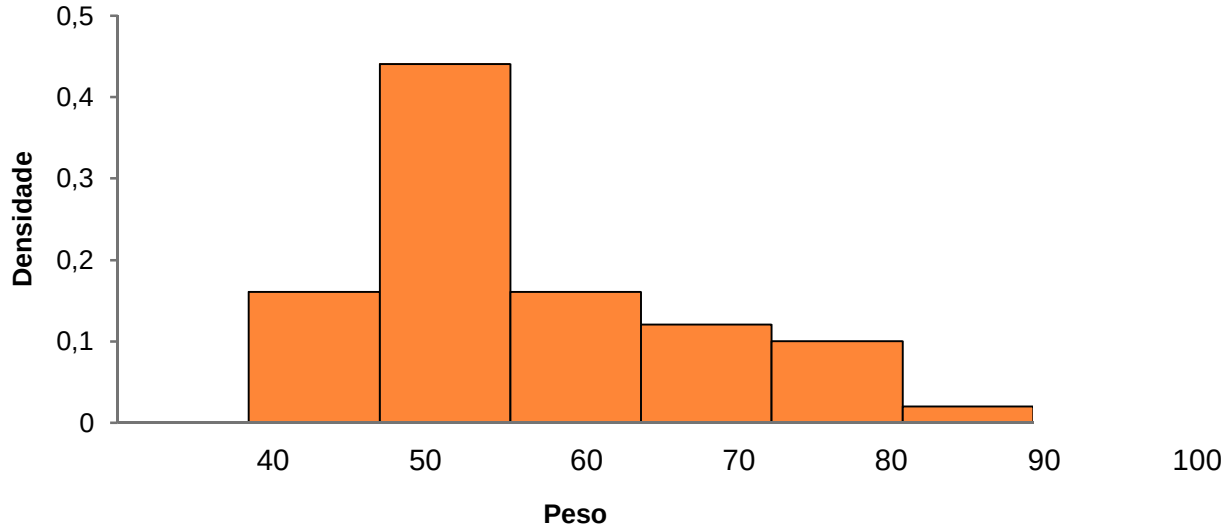
- Gráfico de barras: plano cartesiano com os valores das variáveis no eixo das abcissas e a frequência no eixo das ordenadas




- Variáveis discretas ou qualitativas ordinais

# COMO TRABALHAR COM OS DADOS BRUTOS?

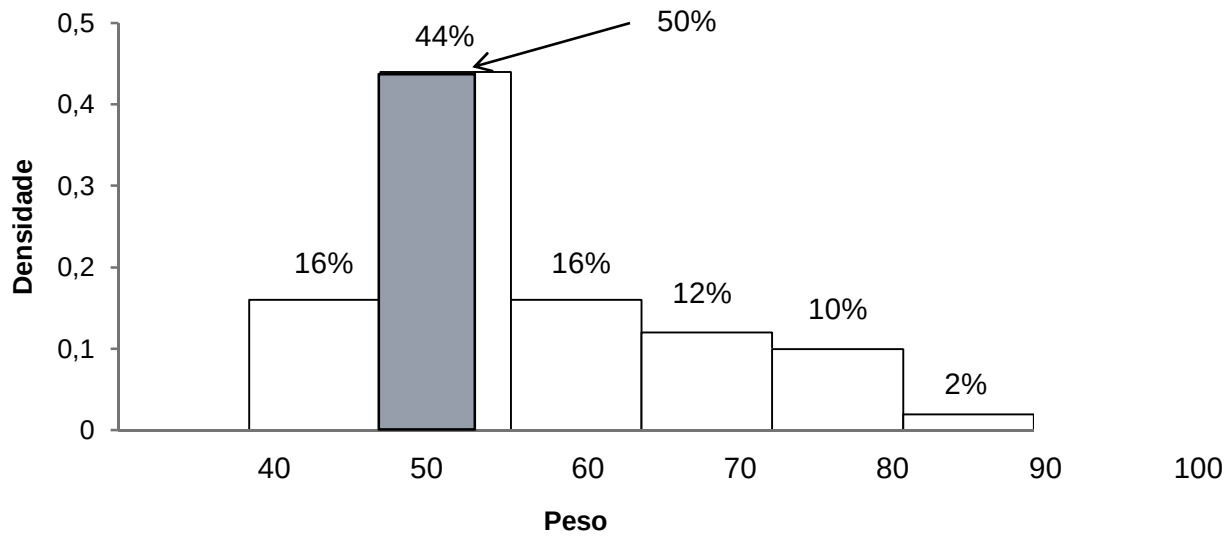
- Histograma: retângulos contíguos de base dada pelas faixas de valores e área igual a frequência relativa na respectiva faixa




- A frequência absoluta também pode ser utilizada

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Como calcular a mediana através de um histograma:




$$\frac{med - 50}{0,34} = \frac{60 - 50}{0,44} = 57,73\ Kg$$

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Como calcular os quartils?
  - Divisão do conjunto de dados em 4 partes (primeiro, segundo e terceiro quartils)
    - O segundo quartil é a mediana
- Exemplo:
  22, 29, 33, 35, 35, 37,38, 43, 43, 44, 48, 48, 52, 53, 55, 57, 61, 62, 67 e 69
  - Número de observações: 20
  - Mediana: Média da 10ª e 11ª observações (44+48)/2 = 46
  - Q1: Média da 5ª e 6ª observações (35+37)/2 = 36
  - Q3: Média da 15ª e 16ª observações (55+57)/2 = 46

# COMO TRABALHAR COM OS DADOS BRUTOS?

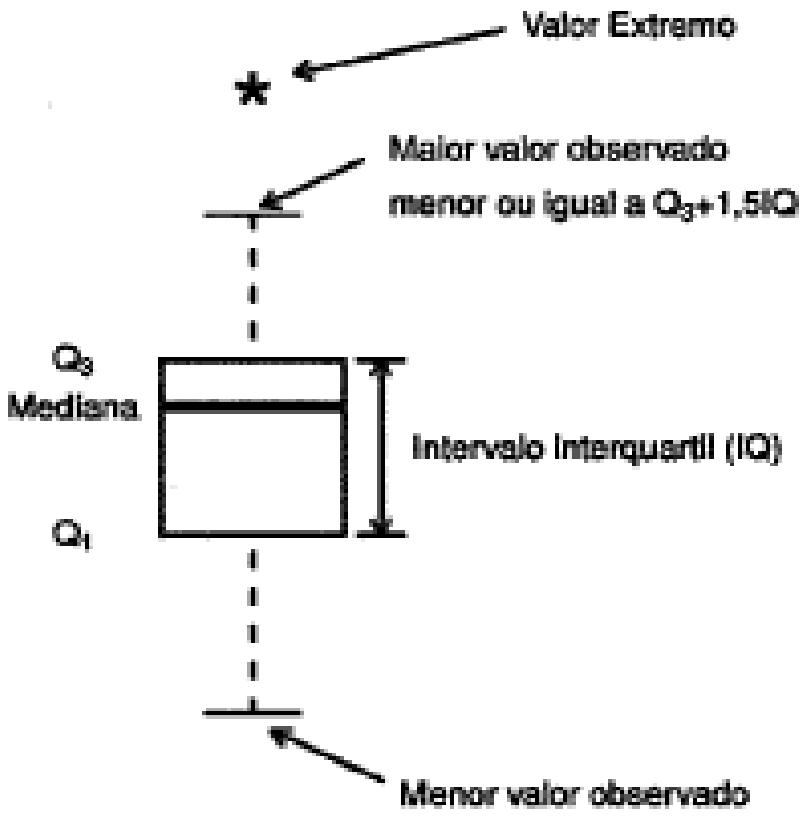
- Outro exemplo: 1,70; 1,71; 1,73; 1,73; 1,76; 1,76; 1,78; 1,80; 1,80; 1,83; 1,85; 1,85; 1,85
  - Qual seria a mediana?
  - O Q1 e o Q3?
- Para este exemplo:
  - Mediana = 7ª observação = 1,78
  - Q1 (conjunto inferior incluindo a mediana) = 4ª observação = 1,73
  - Q3 (conjunto superior incluindo a mediana) = 10ª observação = 1,83

# COMO TRABALHAR COM OS DADOS BRUTOS?

- Para que eu preciso dos valores dos quartils?
  - Analisar a distribuição dos dados
  - Gráfico de Box-plot (gráfico de caixa)
    - Retângulo em que a aresta inferior coincide com o primeiro quartil (Q1) e a superior, com o terceiro quartil
    - Permite visualizar diversos aspectos da distribuição dos dados, como posição, variabilidade, assimetria e a ocorrência de valores atípicos
    - Intervalo interquatil (IQ): contém 50% das observações e dá uma ideia da dispersão dos dados
    - Representa também os *outliers*: os dados que estiverem fora do intervalo

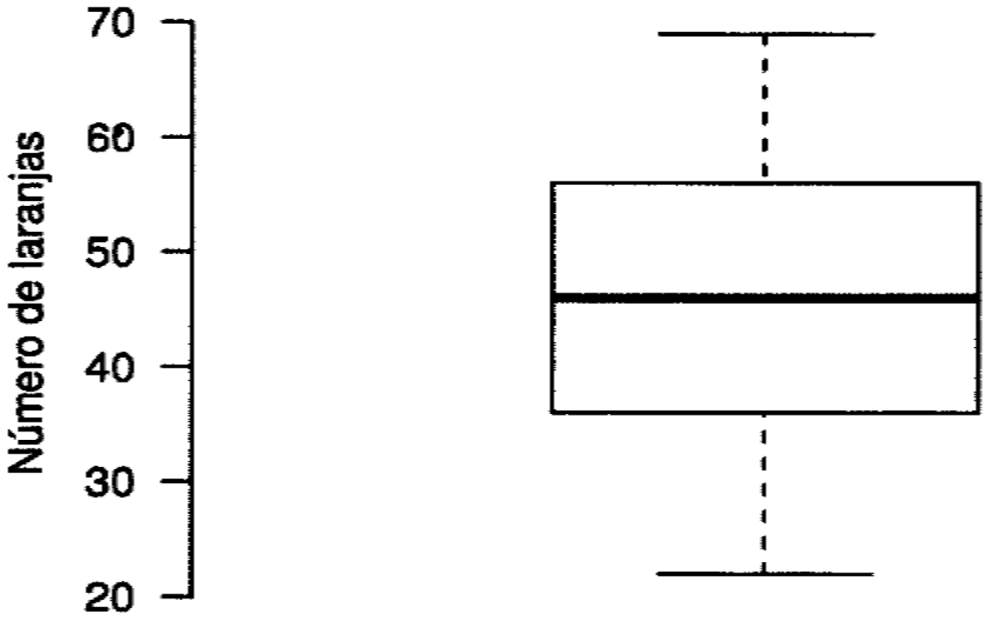
$$[Q_1 - 1.5 * IQ; Q_3 + 1.5 * IQ]$$

# Box-plot


Valor Extremo
Maior valor observado
menor ou igual a $Q_3$+1,5IQ
$Q_3$
Mediana
$Q_1$
Intervalo Interquartil (IQ)
Menor valor observado

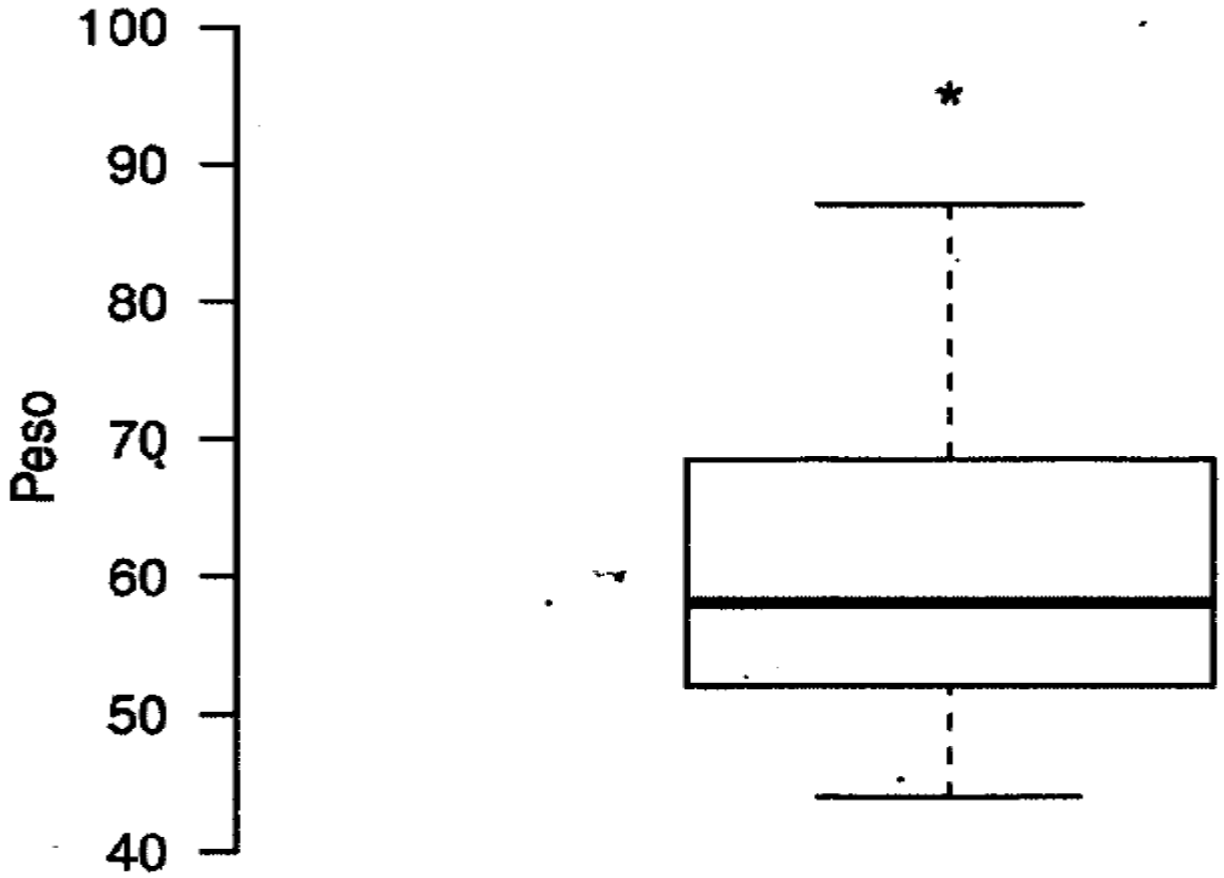
# BOX-PLOT

- Exemplo:
  22, 29, 33, 35, 35, 37,38, 43, 43, 44, 48, 48, 52, 53, 55, 5
  7, 61, 62, 67 e 69



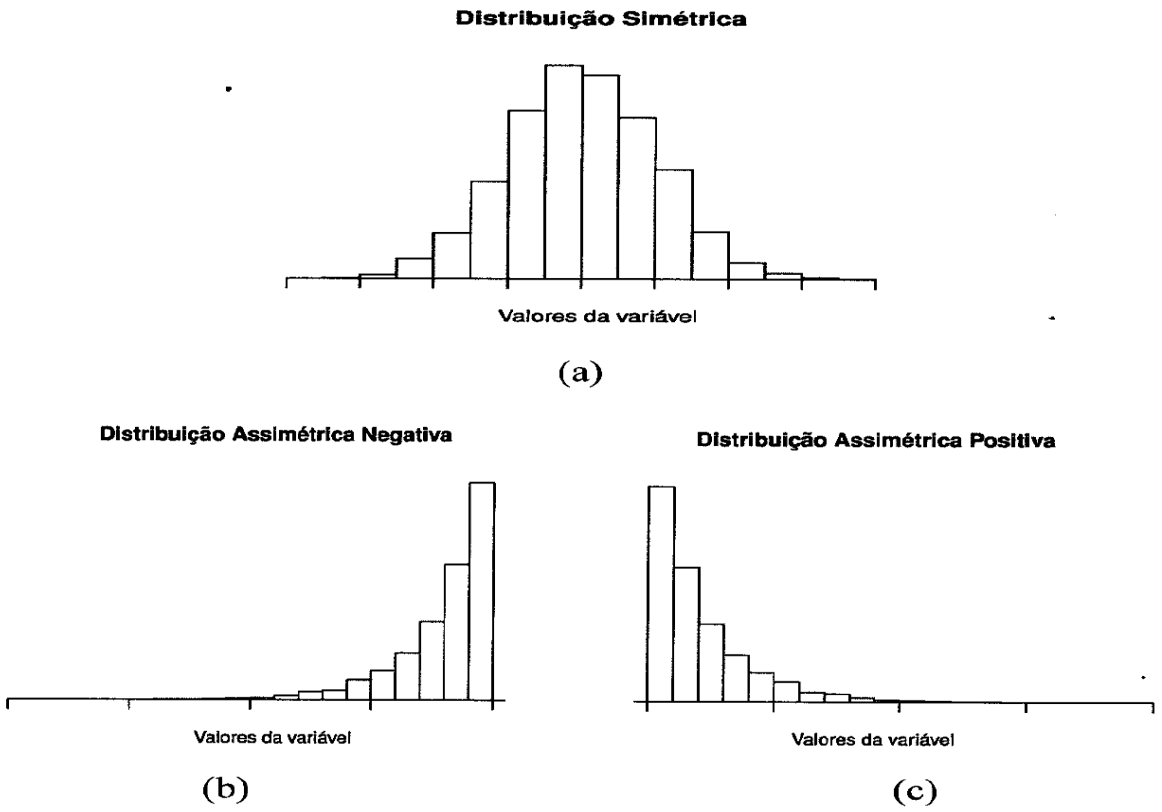
# BOX-PLOT

- Box-plot para a variável Peso



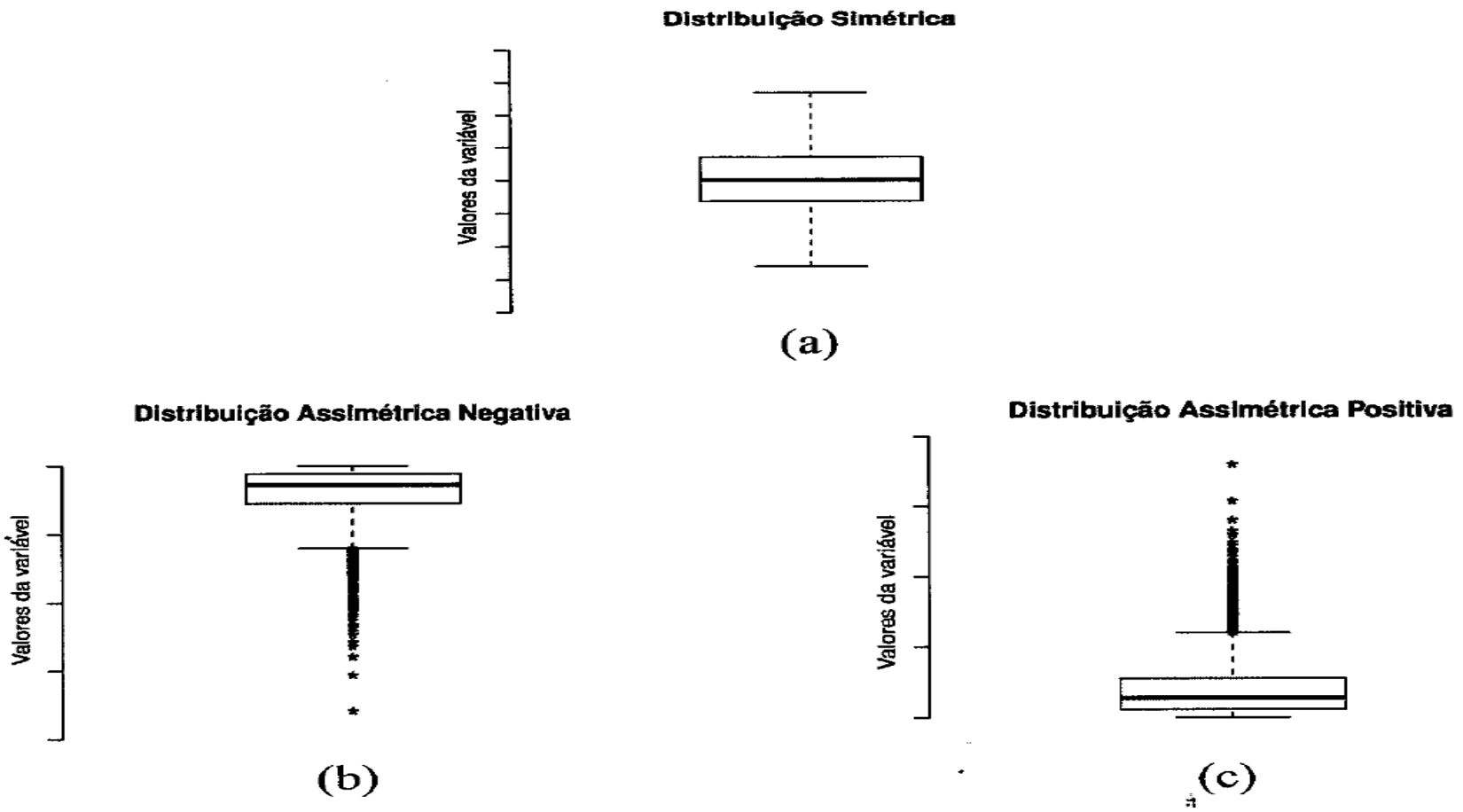
# BOX-PLOT

- Um aspecto importante a ser analisada: existência ou não de simetria na distribuição de seus valores



# BOX-PLOT

- Os gráficos de box-plot servem para detectar diferenças de simetria



# BOX-PLOT

- Podemos quantificar a assimetria (coeficiente de assimetria de Bowley)

$$gb = \frac{(Q_3 - med) - (med - Q_1)}{Q_3 - Q_1}$$

  - Com gb $\rightarrow$ -1, assimetria negativa
  - Com gb $\rightarrow$ 1, assimetria positiva
  - Com gb = 0, simetria

# COMPARANDO DUAS VARIÁVEIS

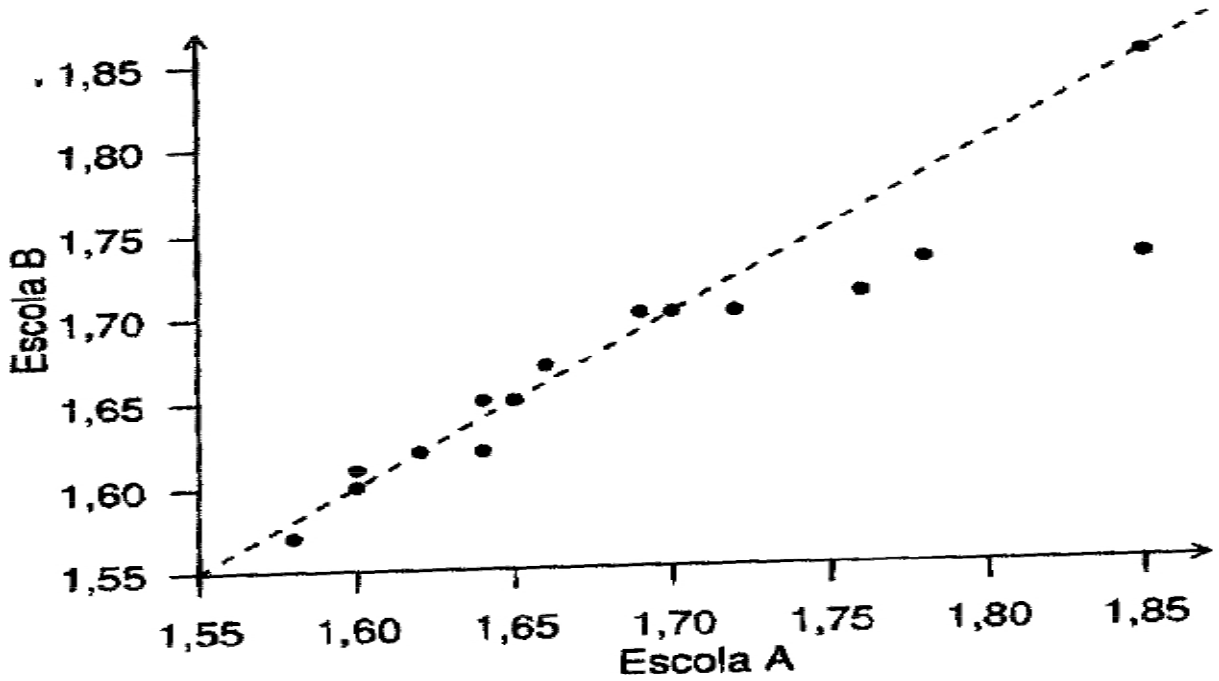
- E como comparar duas variáveis?
  - Uma nova definição,  quantis (qd): valores que limitam uma certa porcentagem  de observações da variável
    - O valor d define a porcentagem definida pelo quantil. Exemplo: $q_{12\%}$ é o quantil que limita 12% dos valores inferiores do conjunto de observações ordenadas
- Para a comparação: gráfico de quantis (Q-Q plot)
  - Representa, no plano cartesiano, quantis das respectivas variáveis como pares ordenados

# GRÁFICO Q-Q PLOT

- Se os pontos estiverem perto da reta de 45°
  - As duas distribuições se aproximam
- Exemplo: as alturas dos alunos de duas escolas
  - A:1,60;1,69;1,85;1,85;1,58;1,76;1,60;1,64;1,62;1,64;1,72;1,66; 1,70;1,78;1,65
  - B:1,62;1,62;1,57;1,65;1,61;1,71;1,65;1,67;1,73;1,60;1,70;1,85; 1,70;1,73;1,70
  - Ordenação:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 1,58 | 1,60 | 1,60 | 1,62 | 1,64 | 1,64 | 1,65 | 1,66 | 1,69 | 1,70 | 1,72 | 1,76 | 1,78 | 1,85 | 1,85 |
| B | 1,57 | 1,60 | 1,61 | 1,62 | 1,62 | 1,65 | 1,65 | 1,67 | 1,70 | 1,70 | 1,70 | 1,71 | 1,73 | 1,73 | 1,85 |

# GRÁFICO Q-Q PLOT




- As distribuições são similares para as alturas baixas
- Diferenças importante em alturas mais altas

# GRÁFICO Q-Q PLOT

- E para variáveis com tamanhos diferentes??
  - Usa a ideia de quantil para cada amostra
    - Usar os dados brutos, histograma ou tabela de frequencia
    - Dados brutos: definir o intervalo e fazer a interpolação
- Exemplo: variável Peso separado para os dois sexos
  - 13 observações para o feminino
  - 37 observações para o masculino

# GRÁFICO Q-Q PLOT

Masculino

| Ordem | Peso | Freq Ac. |
|---|---|---|
| 1 | 60,0 | 0,08 |
| 2 | 68,5 | 0,15 |
| 3 | 71,0 | 0,23 |
| 4 | 72,8 | 0,31 |
| 5 | 73,0 | 0,38 |
| 6 | 73,0 | 0,46 |
| 7 | 75,0 | 0,54 |
| 8 | 80,9 | 0,62 |
| 9 | 84,0 | 0,69 |
| 10 | 85,2 | 0,77 |
| 11 | 86,0 | 0,85 |
| 12 | 87,0 | 0,92 |
| 13 | 95,0 | 1,00 |

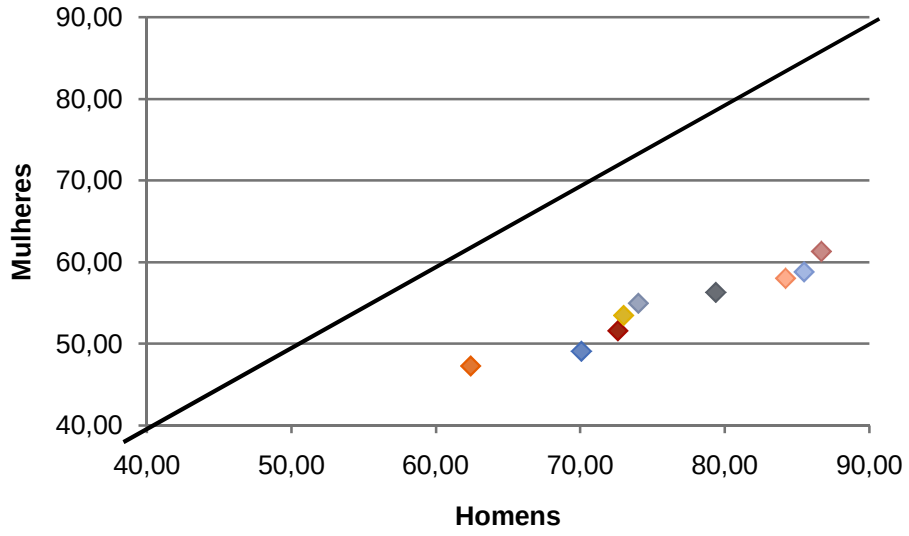
- Escolher intervalos de 10%. Então o primeiro quantil é 10%

$$\frac{68,5 - 60,0}{0,15 - 0,08} = \frac{d_{10\%} - 60,0}{0,10 - 0,08} \Rightarrow d_{10\%} = 62,4$$

- Calcula-se todos os demais quantils
- Fazer o mesmo para o sexo feminino (37 observações)

# GRÁFICO Q-Q PLOT

| Decis | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|
| Masc | 62,4 | 70,1 | 72,6 | 73,0 | 74,0 | 79,4 | 84,2 | 85,5 | 86,7 |
| Fem | 47,3 | 49,1 | 51,6 | 53,5 | 55,0 | 56,3 | 58,0 | 58,8 | 61,3 |




Percebe-se claramente que os pesos dos homens são maiores, em todos os decis

# Uso de computadores em Estatística

- O desenvolvimento da industria de computadores deu grande impulso ao uso da Estatistica
- Vários programas utilizam rotinas estatísticas
  - Planilhas eletronicas
- Programas com análise estatistica
  - Difererentes áreas de aplicação: humanas e biomédicas
- Qualquer programa, devem seguir as etapas
  1. Entrada de dados
  2. Execução da análise estatística
  3. Interpretação dos resultados obtidos

# Uso de computadores em Estatística

- Entrada de dados
  - Como os dados serão fornecidos e organizados
  - Uma matriz com unidades experimentais e características
- Execução da análise estatística
  - Trabalhar com as informações
  - Que tipo de informações (e como) podemos extrair dos dados
- Interpretação dos resultados
  - Verificar se resultados absurdos não estão acontecendo
  - O que eu posso extrair de “conhecimento” dos resultados obtidos

# EXERCÍCIO 1

- Quinze pacientes operados de uma clinica de cardíaca foram entrevistados quanto ao numero de meses previstos de tratamento, se haverá ou não sequelas após o tratamento e o grau de complexidade da cirurgia, de acordo com a tabela

| Paciente | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tratam | 7 | 8 | 5 | 6 | 4 | 5 | 7 | 7 | 6 | 8 | 6 | 5 | 5 | 4 | 5 |
| Seq | S | S | N | N | N | S | S | N | N | S | S | N | S | N | N |
| Cirurgia | A | M | A | M | M | B | A | M | B | M | B | B | M | M | A |

- Vamos analisar esses dados:
  - Classifique cada uma das variáveis
  - Para cada variável, construa a tabela de frequência e faça uma representação gráfica
  - Para o grupo de pacientes que não ficaram com sequelas, faça um gráfico de barras para a variável tratamento. Você acha que ela se comporta de modo diferente nesse grupo?

# Exercício 2

- Em um estudo clínico, dois medicamentos estão sendo avaliados. Cada um deles, $\alpha$ e $\beta$, foi aplicado a um grupo de 18 pessoas e todos tinham aproximadamente as mesmas características físicas, incluindo peso e idade. O tempo para o completo efeito do medicamento foi medido em segundos. Os dados, ordenados crescentemente, são apresentados na tabela

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 24 | 24 | 24 | 25 | 25 | 26 | 26 | 27 | 28 | 29 | 30 | 30 | 30 | 31 | 31 | 32 | 32 | 33 |
| $\beta$ | 19 | 19 | 19 | 20 | 22 | 25 | 26 | 26 | 27 | 29 | 29 | 31 | 34 | 34 | 37 | 40 | 41 | 42 |

  - Construa um box-plot para o tempo de efeito de cada medicamento e comente as diferenças encontradas
  - Determine o coeficiente de assimetria de Bowley para cada medicamento e comente os resultados
  - Compare os desempenhos usando um Q-plot