

Introdução à Bioestatística

Estatística Descritiva/Exploratória

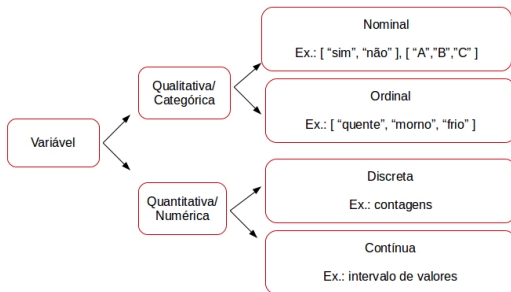
Enrico A. Colosimo/UFMG

<http://www.est.ufmg.br/~enricoc/>

Depto. Estatística - ICEx - UFMG

Descrição de Dados

- Variável: característica de interesse



Variáveis: resposta e covariáveis

Definições

- Variáveis quantitativas discretas: podem ser vistas como resultantes de contagens, assumindo assim, em geral, valores inteiros. Ex: número de ovos de um parasita por volume de fezes.

- Variáveis quantitativas numéricas: assumem valores em um intervalo de números e, geralmente, são provenientes de uma escala numérica. Ex.: peso, altura, glicemia, etc

Examinando descritivamente variáveis

1. Univariada: uma única variável é examinada por indivíduo.
2. Bivariada: duas variáveis são examinadas por indivíduo.
3. Multivariada: mais de duas variáveis é examinada por indivíduo.

Construção do Banco de Dados

1. Planilha ou Banco de Dados.
2. Tantas linhas quanto indivíduos (tamanho da amostra).
3. Tantas colunas quanto variáveis medidas.

Exemplo: Base de Dados

id	Nível.Educ.	Peso(Kg)	Altura(cm)	Idade	Fumante	Gênero
1	1	52	152	40	0	0
2	3	46	160	47	0	0
3	2	89	168	53	0	1
4	1	86	173	38	1	1
5	4	61	169	52	0	0
6	4	87	170	50	0	1
7	1	67	161	51	0	1
8	3	78	166	61	0	1
9	2	99	178	38	1	1
10	2	62	162	63	0	0
.
.
99	1	110	198	59	0	1
100	2	77	168	48	0	1

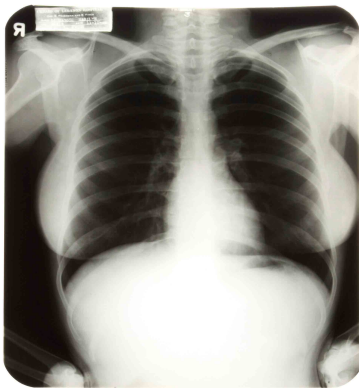
1:Ens.Fund.Inc.
2:Ens.Fund.
3:Ens.Medio
4:Ens.Superior

0: Não 0: Feminino
1: Sim 1: Masculino

1. Técnicas gráficas.

2. Síntese numérica.

Interpretando Gráficos



1 Variáveis Qualitativas/Categóricas

- Disco/ Torta/ Pizza
- Barras (variável vs frequência)

2 Variáveis Quantitativas/Contínuas

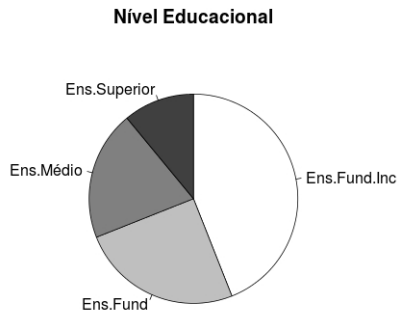
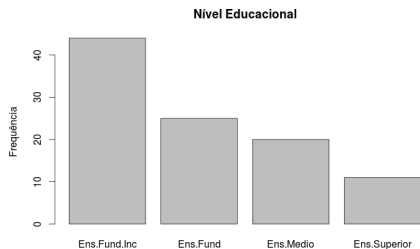
- Histograma (polígono de frequência)
- Boxplot

Variáveis Categóricas - Tabela de Frequência

Variável	n_i	f_i	fac
valor 1			
valor 2			
⋮			
Total	n	1	

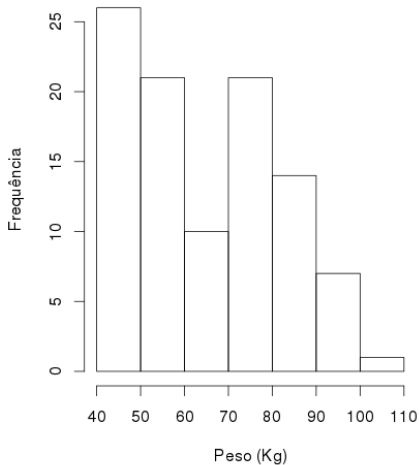
Nível Educacional	n_i	f_i	fac
Ens.Fund.Inc	44	0,44	0,44
Ens.Fund	25	0,25	0,69
Ens.Médio	20	0,20	0,89
Ens.Superior	11	0,11	1
Total	100	1	

Categórica Ordinal: Nível Educacional



Variável Numérica: Peso (kg)

Histograma



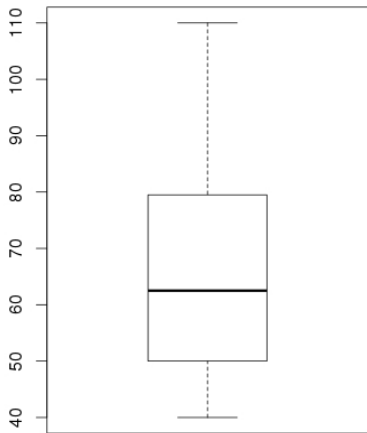
Dividir em classes:
Quantas? $\sqrt{(n)}$

Tamanho das classes:
 $\frac{x_{max} - x_{min}}{k}$

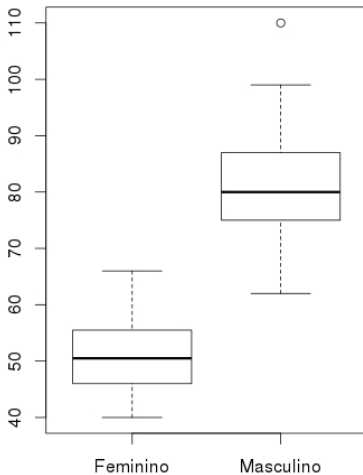
k : numero de classes.

Variável Numérica: Peso (kg)

Peso(Kg)



Peso por Gênero



Gênero

Variável Numérica: Peso (kg)

Estatística	Geral	Feminino	Masculino
Máximo	110,00	66,00	110,00
Q_3	79,25	55,25	87,00
Média	65,35	50,83	81,08
Mediana	62,50	50,50	80,00
Q_1	50,00	46,00	75,00
Mínimo	40,00	40,00	62,00
Desvio-Padrão	17,2371	6,7292	9,5155

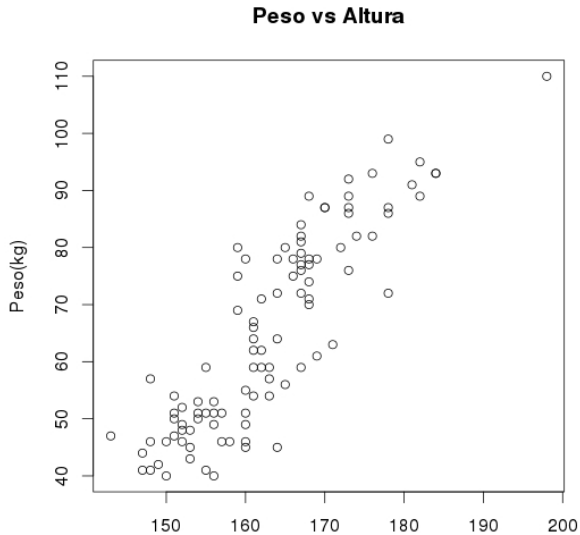
Gráficos Bivariados

1. Tabela de Contingência

		Gênero		
		Homem	Mulher	Total
Fumante	Não	48	36	84
	Sim	4	12	16
Total		52	48	100

Gráficos Bivariados

1. Gráfico de Dispersão



1. Medidas de Posição (Tendência Central)

1.1 Média amostral (\bar{x}):

Sejam x_1, x_2, \dots, x_n observações da variável X:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

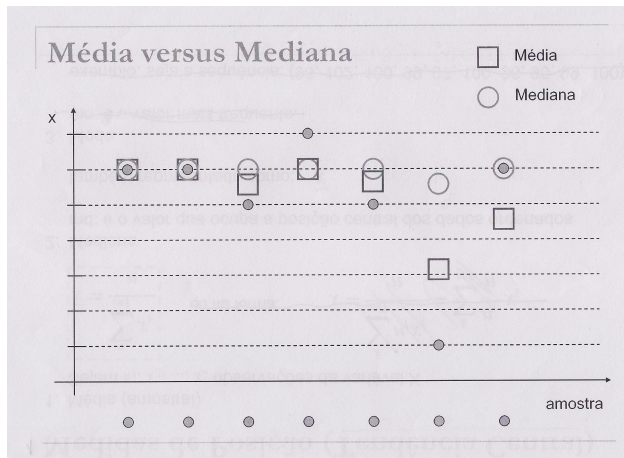
1.2 Mediana (md):

Valor que ocupa a posição central dos dados ordenados.

1.3 Moda (mo):

Valor de maior frequência dos dados.

Média vs Mediana



Medidas de Dispersão

1. Amplitude: Diferença entre o maior e menor valor do conjunto de dados:

Amp = maior - menor

2. Desvio-Padrão

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

- 1 Exemplo:

Dados 1: {5, 7, 9, 11, 13}

Dados 2: {1, 5, 9, 13, 19}

$\bar{x}_1 = \bar{x}_2 = 13$ (mesma média)

$s_1 = 3,2$ $Amp_1 = 8$

$S_2 = 7,0$ $Amp_2 = 18$

Percentis e Quartis

1 Percentil:

O percentil de ordem $100p\%$ de um conjunto de valores em ordem crescente é um valor tal que $(100p)\%$ das observações são menores ou iguais a ele e $100(1-p)\%$ são maiores ou igual a ele.

2 Quartis:

Os percentis de ordem 25, 50 e 75 são chamados quartis. São representados por Q_1 , Q_2 (mediana) e Q_3 .

OBS.: Q_1 deixa pelo menos 25% dos dados abaixo dele e pelo menos 75% dos dados acima dele.

Exemplos: Percentis e Quartis

Dados: 1,2,3,4,5,6,7,8,9,10

Encontrar: P_{10} , P_{25}

Exercício: P_{50} , $P_{75} = Q_3$

P_{10} :

Verdadeiro: qualquer valor entre 1 e 2.

$np=10 \times (0,1)=1$

Software: $1+1 \times (0,1)=1,1$

P_{25} :

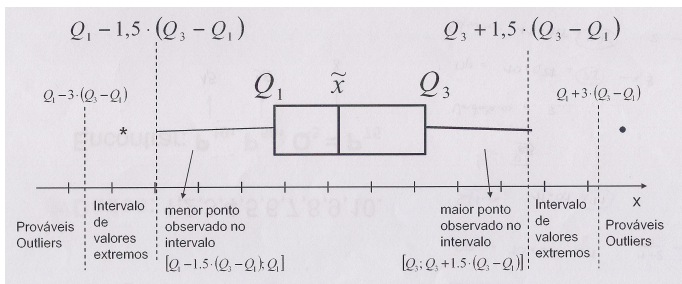
Verdadeiro: 3

$np=10 \times (0,25)=2,5 \rightarrow 3$

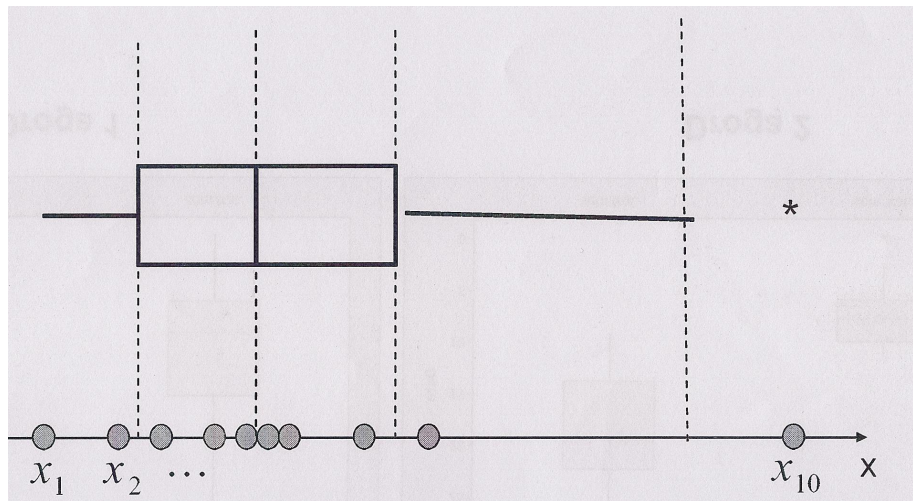
Software: $1+1 \times (0,25)=2,75$

Construção do Box-Plot

É um gráfico que apresenta simultaneamente várias características de dados: localização, dispersão, simetria e presença de observações discrepantes ("outliers")

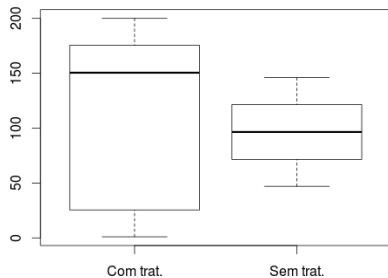


Box-Plot

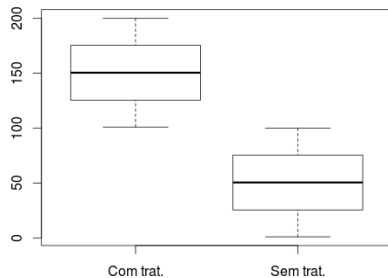


Comparação de Grupos

Droga 01



Droga 02

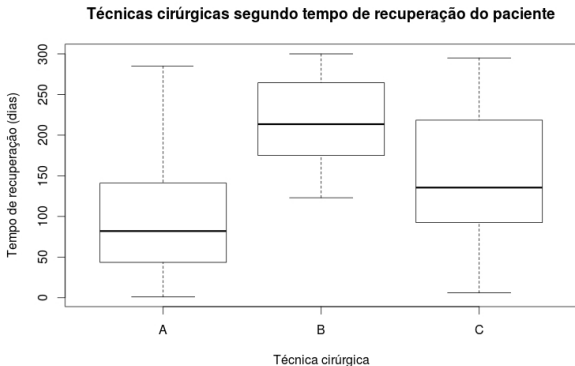


Causas do aparecimento de outliers

- Leitura, anotação ou transcrição incorreta dos dados.
- Erro na execução do experimento ou na tomada da medida.
- Mudanças não controláveis nas condições experimentais ou dos pacientes.
- Características inerentes à variável estudada (Ex.: Grande instabilidade do que está sendo medido).

Exemplo

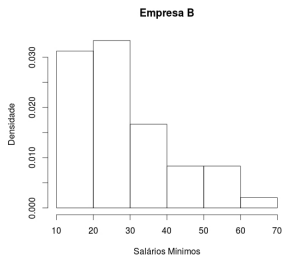
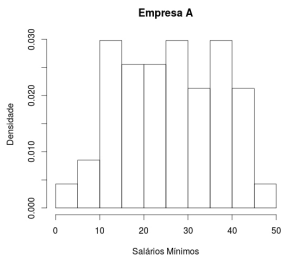
Deseja-se comparar três técnicas cirúrgicas para a extração de dente siso. Cada uma das técnicas foi aplicada em 20 pacientes e os resultados são apresentados a seguir:



Qual é a impressão inicial ao se comparar estas técnicas cirúrgicas?

Exemplo

Suponha que duas empresas desejam empregá-lo e após considerar as vantagens de cada uma você escolha aquela que lhe paga melhor. Após certa pesquisa, você consegue a distribuição de salário das empresas dadas segundo os gráficos abaixo. Com base nas informações de cada gráfico, qual seria a sua decisão?



Comando do R

```
# Lendo os dados
bd = read.table(file="/BancoDados.txt", dec=".", header=TRUE, sep=";")

# Gráfico de barras
barplot(table(bd$nivel_educ), main="Nível Educacional", ylab="Frequência")

# Gráfico de pizza
pie(table(bd$nivel_educ), main="Nível Educacional", col= gray( c(1, 0.75, 0.5, 0.25) ) )

# Histograma
hist(bd$peso, xlab="Peso (Kg)", ylab="Frequência", main="Histograma")

# Boxplot
boxplot(bd$peso, main="Peso (Kg)")

# Medidas descritivas
summary(bd$peso)
quantile(bd$peso)
sd(bd$peso)

# Tabela de Contingência
addmargins(table(bd$fumante, bd$genero))

# Gráfico de Dispersão
plot(bd$altura, bd$peso, xlab="Altura", ylab="Peso", main="Peso vs Altura")

# Boxplot que compara duas classes/fatores (tratamentos)
droga1 = read.table(file="/droga1.txt", dec=".", header=TRUE, sep=";")
names(droga1) = c("trat", "exp")
droga1$trat = as.factor(droga1$trat)
droga1$exp = as.numeric(droga1$exp)
boxplot(exp trat, data=droga1, main="Droga 01")
```