

Aula Prática: Análise Descritiva

Bioestatística F

André Gabriel F. C. da Costa
Dr. Enrico A. Colossimo



Sumário da apresentação

- 1 **Introdução ao R**
 - O Software R
 - Conhecendo os objetos do R
 - Importando Dados
- 2 **Análise Descritiva - Uma Visão Univariada**
 - Variáveis Qualitativas
 - Variáveis Quantitativas
- 3 **Análise Descritiva - Uma Visão Bivariada**
 - Quantitativa x Qualitativa
 - Qualitativa x Qualitativa
 - Quantitativa x Quantitativa



O Software R

O que é o R

- R é um software livre para análise de dados totalmente gratuito. O R apareceu inicialmente em 1996, com os professores de estatística Ross Ihaka e Robert Gentleman, da Universidade de Auckland que desenvolveram a nova linguagem computacional, similar a linguagem S desenvolvida por John Chambers.



O Software R

O que é o R

- R é um software livre para análise de dados totalmente gratuito. O R apareceu inicialmente em 1996, com os professores de estatística Ross Ihaka e Robert Gentleman, da Universidade de Auckland que desenvolveram a nova linguagem computacional, similar a linguagem S desenvolvida por John Chambers.

Por que utilizar o R

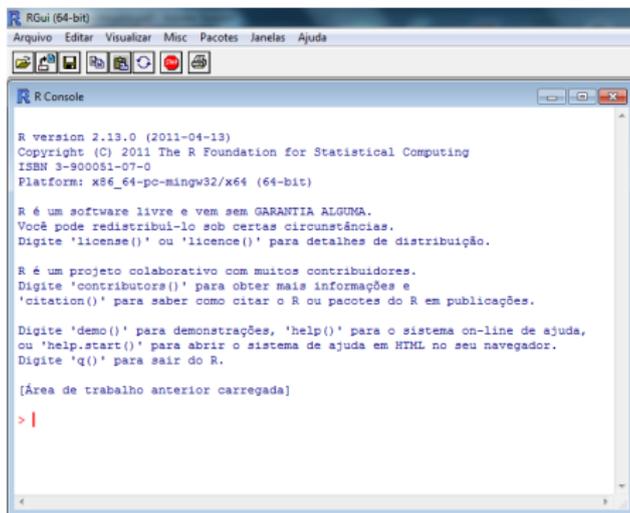
- Software gratuito com código aberto com uma linguagem acessível;
- Expansão exponencial entre pesquisadores, engenheiros e estatísticos;
- Novas aplicações - mais de 3543 pacotes;
- Cobertura inigualável - tecnologia de ponta;
- Capacidade Gráfica;
- Está disponíveis para diferentes plataformas: Windows, Linux e Mac.



O Software R

Instalação:

O site chamado de CRAN The Comprehensive R Archive Network <http://cran.r-project.org/> fornece o instalador do programa. Basta seleccionar o arquivo de acordo com sua plataforma (Windows, MAC ou Linux), baixar o pacote base e executar o programa de instalação.



```
RGui (64-bit)
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda

R Console

R version 2.13.0 (2011-04-13)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

[Área de trabalho anterior carregada]

> |
```



Operadores do R

Operadores	Significado
<-	Atribui um nome a um objeto
*;/;+;-	Multiplicação, divisão, soma e subtração respectivamente
x^b ou x**b	x elevado à b
>, <, =, >=, <=, !=	Maior, menor, igual, maior ou igual, menor ou igual e diferente respectivamente
&	Utilizado em programação lógica, significa e (adiciona uma condição)
	Utilizado em programação lógica, significa ou (adiciona outra possibilidade)
y~ x	Indica y em função de x
,	Separa um argumento de outro dentro da mesma função
"	Delimita caractere
()	Delimita os argumentos dentro de uma função
{}	Indica o início e fim de uma função
[]	Seleciona parte de um objeto
#	Adiciona algum comentário
?	Obtém ajuda



Funções Básicas do R

Funções básicas	Significado
<code>require(nome do pacote)</code>	Ativa o pacote entre parênteses
<code>install.packages("pacotes")</code>	Instala o pacote selecionado, quando está conectada a internet
<code>Help(nome da função)</code>	Fornecer informações da função entre parênteses
<code>read.table(nome do arquivo)</code>	Importação de arquivos no formato txt, para o R
Funções estatísticas	Significado
<code>max(x)</code>	Valor máximo
<code>min(x)</code>	Valor mínimo
<code>sum(x)</code>	Soma dos valores
<code>mean(x)</code>	Média
<code>median(x)</code>	Mediana
<code>sd(x)</code>	Desvio padrão
<code>var(x)</code>	Variância
Funções matemáticas	Significado
<code>length(x)</code>	Tamanho do vetor
<code>log(x)</code>	Logaritmo na base e
<code>exp(x)</code>	Exponencial
<code>sqrt(x)</code>	Raiz quadrada



Conhecendo os objetos do R

Os objetos do R:

Vetores - Fatores - Matrizes - Listas - DataFrames



Conhecendo os objetos do R

Os objetos do R:

Vetores - Fatores - Matrizes - Listas - DataFrames

Vetores: Uma seqüência de valores numéricos

```
x <- c(2, 3, 5, 7, 11)
```

Fatores: Armazenar dados categóricos

```
c5 <- c("M", "F", "F", "F", "M", "M")
```

```
F5 <- as.factor(c5)
```

Matrizes: Combinações de vetores.

```
matrix(1:12, ncol=3).
```

	[,1]	[,2]	[,3]
[,2]	1	5	9
[,3]	2	6	10
[,4]	3	7	11
[,5]	4	8	12



Conhecendo os objetos do R

Listas: Utilizados para combinar diferentes objetos em um objeto único.

```
peessoa <- list(idade=21, nome='Fred', score=c(65,78,55))
```

- `peessoa$idade`

21

- `peessoa$nome`

"Fred"

- `peessoa$score`

65 78 55



Conhecendo os objetos do R

DataFrame: É a melhor forma de armazenar dados, pois diferentemente da matriz pode-se armazenar dados numéricos e não numéricos.

"Talvez o mais importante para analisar seus próprios dados corretamente está em começar montando seu dataframe de maneira correta"



Conhecendo os objetos do R

DataFrame: É a melhor forma de armazenar dados, pois diferentemente da matriz pode-se armazenar dados numéricos e não numéricos.

"Talvez o mais importante para analisar seus próprios dados corretamente está em começar montando seu dataframe de maneira correta"

Exemplo: Suponha um experimento com três tratamentos (controle, pré-aquecido e pré-refrigerados) e quatro medições por tratamento, a forma a abaixo está correta?

Controle	Pré-aquecido	Pré-refrigerado
6,1	6,3	7,1
5,9	6,2	8,2
5,8	5,8	7,3
5,4	6,3	6,9



Conhecendo os objetos do R

A melhor maneira seria:

Tratamento	Resposta
Controle	6,1
Controle	5,9
Controle	5,8
Controle	5,4
Pré-aquecido	6,3
Pré-aquecido	6,2
Pré-aquecido	5,8
Pré-aquecido	6,3
Pré-refrigerado	7,1
Pré-refrigerado	8,2
Pré-refrigerado	7,3
Pré-refrigerado	6,9



Importando Dados

Importando Dados

- Após tabular os dados no Excel, salvar como "texto (separado por tabulações)"
- Mudar na barra de ferramentas diretório do R, para a pasta que salvou o arquivo do Excel
- Utilizar o comando:

```
read.table(file, header = FALSE, dec = ".")
```

Para utilizar a função `read.table()` não pode haver nomes compostos nem células vazias.



Importando Dados

Importando Dados

- Após tabular os dados no Excel, salvar como "texto (separado por tabulações)"
- Mudar na barra de ferramentas diretório do R, para a pasta que salvou o arquivo do Excel
- Utilizar o comando:

```
read.table(file, header = FALSE, dec = ".")
```

Para utilizar a função `read.table()` não pode haver nomes compostos nem células vazias.

Caso o banco de dados apresente células vazias ou nomes composto, salvar o arquivo do excel como "CSV (separado por vírgulas)" e utilizar o comando:

```
read.csv2(file, header = TRUE, dec=",")
```



Importando Dados

Com o pacote `foreign` pode-se importar dados de outros softwares, como:

- `read.dta()` Importa dados do Stata
- `read.epiinfo()` Importa dados do Epi Info
- `read.mtp()` Importa dados do Minitab
- `read.spss()` Importa dados do SPSS
- `read.ssd()` Importa dados do SAS

O pacote `RODBC` fornece acesso a bases de dados como SQL Server e Microsoft Access.



Importando Dados

```
data<- read.csv2("dados.csv")
```

```
head(data)
```

```

Amostra Doença      Sexo idade peso recusa.de.alimento
1      Não Masculino  29 0.83      Não Recusa
2      Não Feminino  24 1.92      Não Recusa
3      Sim  Feminino  26 0.86      Não Recusa
4      Sim  Feminino  26 0.93      Não Recusa
5      Não Masculino  29 1.03      Não Recusa
6      Não Masculino  31 1.45      Não Recusa

```

Doença = $\begin{cases} \text{Não} \\ \text{sim} \end{cases}$; Sexo = $\begin{cases} \text{Masculino} \\ \text{Feminino} \end{cases}$; Idade (Semanas); Peso (Kg)

Recusa de alimentos = $\begin{cases} \text{Não Recusa} \\ \text{Às vezes} \\ \text{Sempre} \end{cases}$



Análise Descritiva Univariada: Variáveis Qualitativas

```
attach(data)
doença.tb <- table(Doença)
doença.tb
```

```
Doença
Não Sim
83 48
```

```
doença.tbprop <- prop.table(doença.tb)
doença.tbprop
```

```
Doença
      Não      Sim
0.6335878 0.3664122
```

```
sexo.tb <- table(Sexo)
sexo.tb
```

```
Sexo
Feminino Masculino
60          71
```



Análise Descritiva Univariada: Variáveis Qualitativas

```
sexo.tb <- prop.table(sexo.tb)  
sexo.tb
```

```
Sexo  
Feminino Masculino  
0.4580153 0.5419847
```

```
recusa.tb <- table(recusa.de.alimento)  
recusa.tb
```

```
recusa.de.alimento  
As vezes Não Recusa Recusa  
13 109 9
```

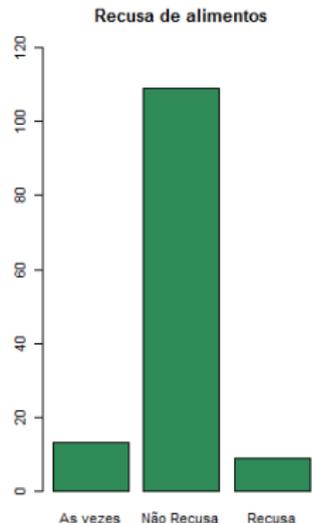
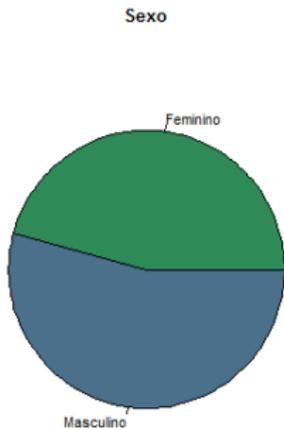
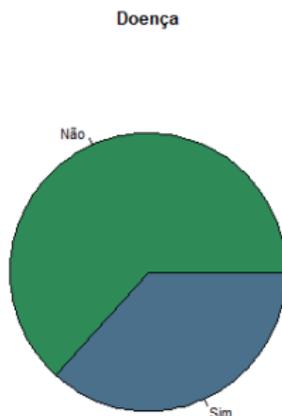
```
recusa.tb <- prop.table(recusa.tb)  
recusa.tb
```

```
recusa.de.alimento  
As vezes Não Recusa Recusa  
0.09923664 0.83206107 0.06870229
```



Análise Descritiva Univariada: Variáveis Qualitativas

```
par(mfrow=c(1,3))  
pie(doença.tb,col=c("seagreen4","skyblue4"), main="Doença")  
pie(sexo.tb,col=c("seagreen4","skyblue4"), main="Sexo")  
barplot(recusa.tb, col="seagreen4", ylim=c(0,120), main="Recusa  
de alimentos")
```



Análise Descritiva Univariada: Variáveis Quantitativas

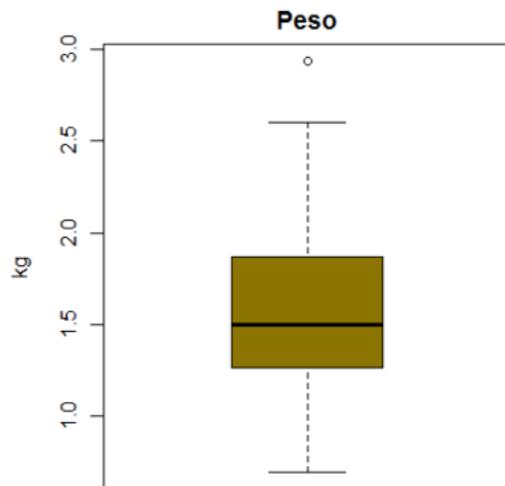
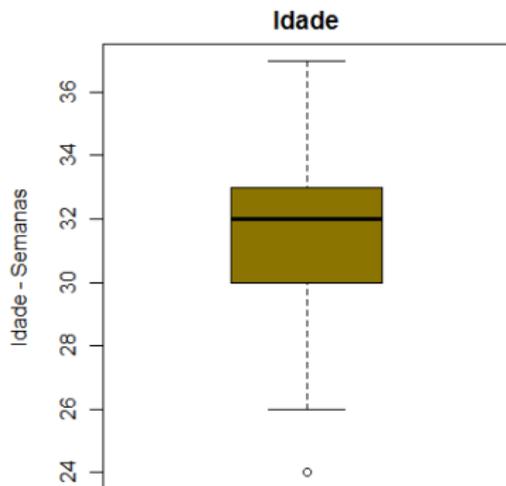
```
install.packages("fields")  
require(fields)  
tab<- cbind(stats(idade), stats(peso))  
colnames(tab)<- c("Idade","Peso")  
round(tab,2)
```

	Idade	Peso
N	131.00	131.00
mean	31.63	1.56
Std.Dev.	2.49	0.43
min	24.00	0.69
Q1	30.00	1.26
median	32.00	1.50
Q3	33.00	1.86
max	37.00	2.94
missing values	0.00	0.00



Análise Descritiva Univariada: Variáveis Quantitativas

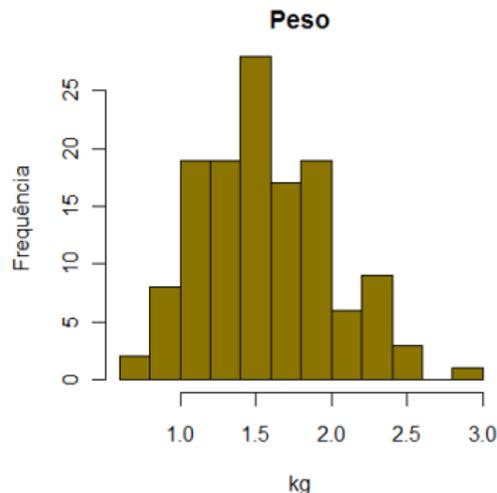
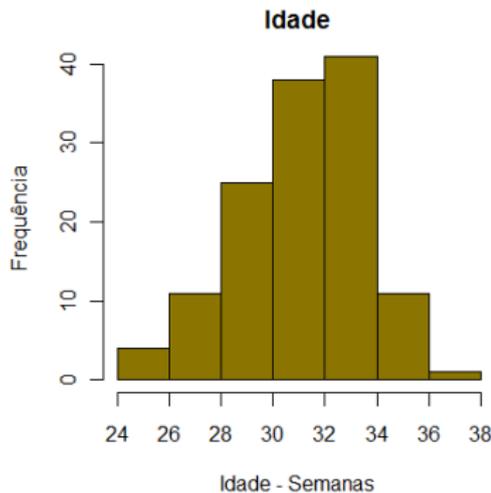
```
par(mfrow=c(1,2))  
boxplot(idade, main="Idade", ylab="Idade - Semanas",  
col="gold4")  
boxplot(peso, main="Peso", ylab="kg", col="gold4")
```



Análise Descritiva Univariada: Variáveis Quantitativas

```
hist(idade, main="Idade", xlab="Idade - Semanas",  
     ylab="Frequência", col="gold4")
```

```
hist(peso, main="Peso", xlab="kg", ylab="Frequência",  
     col="gold4")
```



Análise Descritiva Bivariada: Quantitativa x Qualitativa

```
tab1<-tapply(peso, Doença, stats)  
round(data.frame(tab1[1], tab1[2]),3)
```

	Não	Sim
N	83.000	48.000
mean	1.635	1.438
Std.Dev.	0.448	0.383
min	0.690	0.760
Q1	1.320	1.135
median	1.600	1.410
Q3	1.930	1.657
max	2.940	2.530
missing values	0.000	0.000



Análise Descritiva Bivariada: Quantitativa x Qualitativa

```
tab1<-tapply(peso, Doença, stats)  
round(data.frame(tab1[1], tab1[2]),3)
```

	Não	Sim
N	83.000	48.000
mean	1.635	1.438
Std.Dev.	0.448	0.383
min	0.690	0.760
Q1	1.320	1.135
median	1.600	1.410
Q3	1.930	1.657
max	2.940	2.530
missing values	0.000	0.000

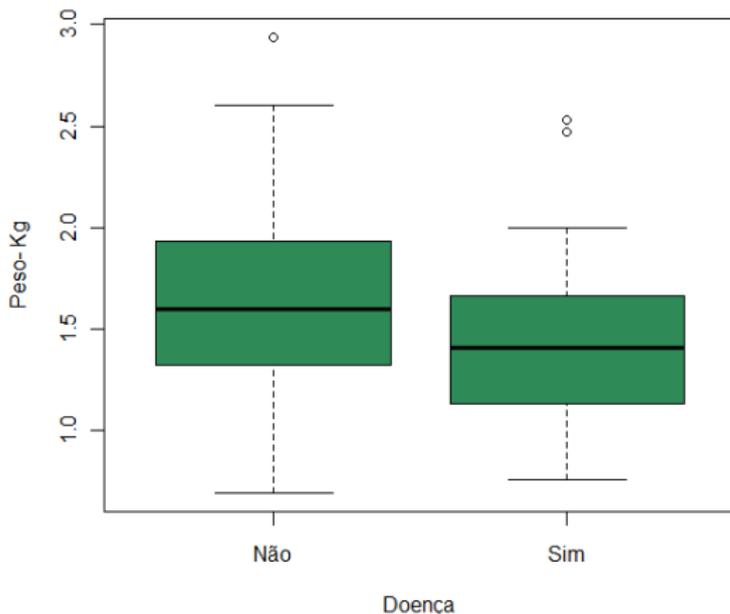
```
t(round(data.frame(tab1[1], tab1[2]),3))[, -9]
```

	N	mean	Std.Dev.	min	Q1	median	Q3	max
Não	83	1.635	0.448	0.69	1.320	1.60	1.930	2.94
Sim	48	1.438	0.383	0.76	1.135	1.41	1.657	2.53



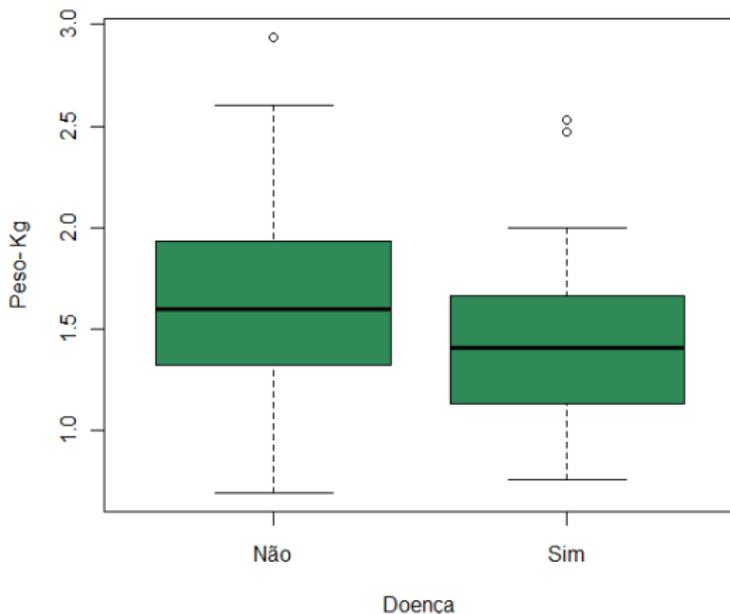
Análise Descritiva Bivariada: Quantitativa x Qualitativa

```
par(mfrow=c(1,1))  
boxplot(peso ~ Doença, col="seagreen4", xlab="Doença",  
ylab="Peso(Kg)")
```



Análise Descritiva Bivariada: Quantitativa x Qualitativa

```
par(mfrow=c(1,1))  
boxplot(peso ~ Doença, col="seagreen4", xlab="Doença",  
ylab="Peso(Kg)")
```



Análise Descritiva Bivariada: Qualitativa x Qualitativa

```
t1<- table(Doença, recusa.de.alimento)  
t1
```

```
      recusa.de.alimento  
Doença As vezes Não Recusa Recusa  
Não      9      69      5  
Sim      4      40      4
```



Análise Descritiva Bivariada: Qualitativa x Qualitativa

```
t1<- table(Doença, recusa.de.alimento)
t1
```

```
      recusa.de.alimento
Doença As vezes Não Recusa Recusa
Não      9      69      5
Sim      4      40      4
```

```
prop.table(t1)
```

```
      recusa.de.alimento
Doença As vezes Não Recusa Recusa
Não  0.0687  0.5267 0.0382
Sim  0.0305  0.3053 0.0305
```



Análise Descritiva Bivariada: Qualitativa x Qualitativa

```
t1<- table(Doença, recusa.de.alimento)
t1
```

```
      recusa.de.alimento
Doença As vezes Não Recusa Recusa
Não      9      69      5
Sim      4      40      4
```

```
prop.table(t1)
```

```
      recusa.de.alimento
Doença As vezes Não Recusa Recusa
Não  0.0687  0.5267 0.0382
Sim  0.0305  0.3053 0.0305
```

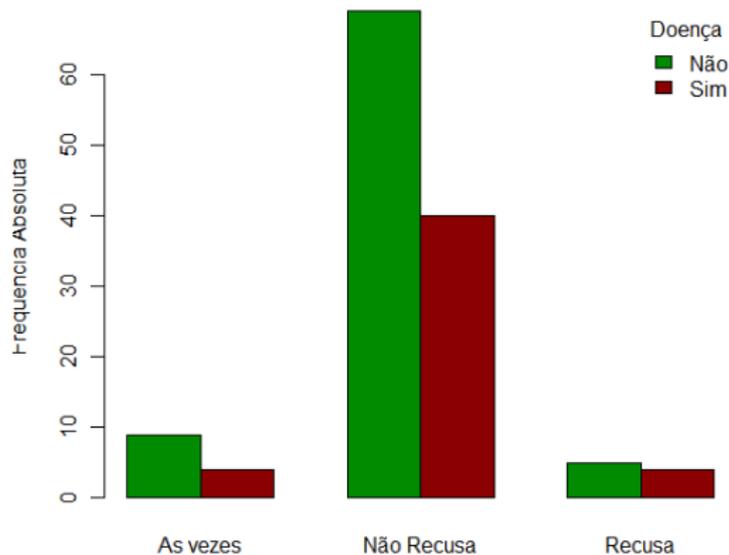
```
prop.table(t1, mar=1)
```

```
      recusa.de.alimento
Doença As vezes Não Recusa Recusa
Não  0.1084  0.8313 0.0602
Sim  0.0833  0.8333 0.0833
```



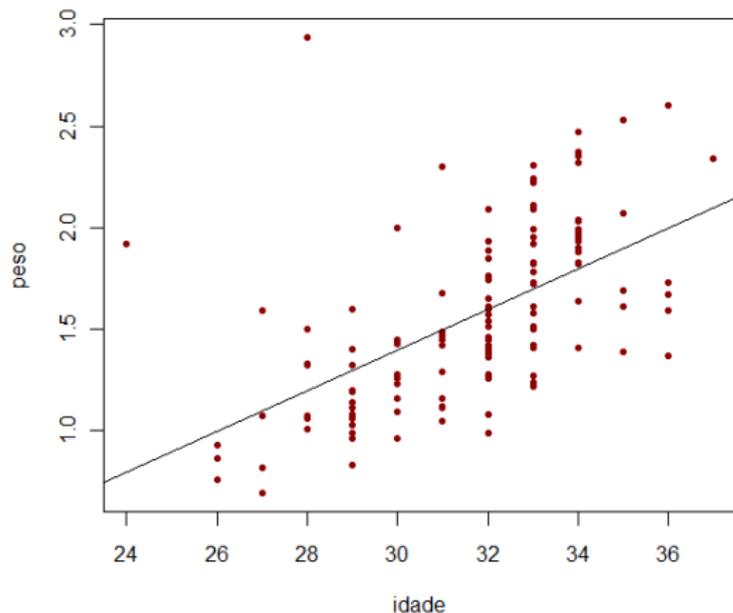
Análise Descritiva Bivariada: Qualitativa x Qualitativa

```
barplot(t1, beside=T, col=c("green4","red4"),  
ylab="Frequência Absoluta")  
legend("topright", fill=c("green4","red4"),  
c("Não","Sim"), title="Doença", bty="n")
```



Análise Descritiva Bivariada: Quantitativa x Quantitativa

```
plot(idade,peso, pch=20, col="red4")  
abline(lm(peso ~ idade))
```



Prática no R

