

Data Frames



Data Frames



- É mais geral que uma matriz
 - ✓ Colunas diferentes de mesmo comprimento podem ter diferentes tipos de dados (numérico, caracter, lógico, fator, etc.)

```
d <- c(1,2,3,4)
e <- c("vermelho", "branco", "vermelho", NA)
f <- c(TRUE,TRUE,TRUE,FALSE)
meubanco <- data.frame(d,e,f)
names(meubanco) <- c("ID","Cor","verificou") # nome das variáveis
```

Laboratório de Estatística - 2020

84



Data Frames – Indexação



- Há várias maneiras para identificar os elementos de um data frame

```
meubanco[2:3]           # colunas 2 e3 do data frame
meubanco[c("ID", "verificou")] # columns ID e do data frame
meubanco$Cor             # variable verificou do data frame
```

Laboratório de Estatística - 2020

85



Exemplo – Consumo de Energia



- Consumo de energia pré-menstrual:
 - ✓ consumo.pre = c(5260, 5470, 5640, 6180, 6390)
- Consumo de energia pós-menstrual:
 - ✓ consumo.pos = c(3910, 4220, 3885, 5160, 5645)
- Conjunto de dados (Data frame):
 - ✓ dados.altman=data.frame(consumo.pre,consumo.pos)
- Tamanho do conjunto de dados:
 - ✓ dim(dados.altman)

Laboratório de Estatística - 2020

86



```
> consumo.pre = c(5260, 5470, 5640, 6180, 6390)
> consumo.pos = c(3910, 4220, 3885, 5160, 5645)
>
> dados.altman = data.frame(consumo.pre, consumo.pos)
>
> dados.altman
  consumo.pre consumo.pos
1      5260       3910
2      5470       4220
3      5640       3885
4      6180       5160
5      6390       5645
> |
```



- Dados são emparelhados
- Variáveis acessíveis usando notação \$
√ dados.altman\$consumo.pre

Indexação



- Consumo pré-menstrual mulher #5
√ consumo.pre[5]
- Consumo pré-menstrual mulheres #1, 3, 5
√ consumo.pre[c(1, 3, 5)]
- Consumo pré-menstrual 3 primeiras mulheres
√ consumo.pre[1:3]



- Consumo pré-menstrual todas exceto mulheres # 1, 3 e 5



√ consumo.pre[-c(1, 3, 5)]

```
> consumo.pre[5]
[1] 6390
>
> consumo.pre[c(3,5,7)]
[1] 5640 6390 NA
>
>
> consumo.pre[1:3]
[1] 5260 5470 5640
>
> consumo.pre[-c(1,3,5)]
[1] 5470 6180
>
> impar=c(1,3,5)
> consumo.pre[-impar]
[1] 5470 6180
>
> dados.altman$consumo.pre[5]
[1] 6390
> |
```



- Pode-se indexar pelo do Data frame
√ dados.altman\$consumo.pre[5]
- Todas as medidas da mulher #5:
√ dados.altman[5,]
- Consumos de energia pós-menstrual 3 primeiras mulheres
√ dados.altman[1:3, 2]





```

> dados.altman$consumo.pre[5]
[1] 6390
>
> dados.altman[5,]
  consumo.pre  consumo.pos
5       6390      5645
>
>
> dados.altman[1:3,2]
[1] 3910 4220 3885
>
> |

```

91

Laboratório de Estatística - 2020



Seleção Condicional



```

consumo.pos [consumo.pre>6000]

consumo.pos [consumo.pre>6000 & consumo.pre<=6500]

consumo.pre>6000 & consumo.pre<=6500

```

92

Laboratório de Estatística - 2020



```

> consumo.pos[consumo.pre>6000]
[1] 5160 5645
>
>
> consumo.pos[consumo.pre>6000 & consumo.pos <=6500]
[1] 5160 5645
>
>
> consumo.pre>6000 & consumo.pos <=6500
[1] FALSE FALSE FALSE TRUE TRUE
>
> condicao=consumo.pre>6000 & consumo.pos <=6500
>
> consumo.pos[condicao]
[1] 5160 5645
>
> |

```

93

Laboratório de Estatística - 2020



Ordenação



- Comando sort: ordena os elementos do vetor
 - ✓ Default é crescente
 - ✓ sort (consumo.pos)
- Ordenação descrescente
 - ✓ sort (consumo.pos, decreasing=TRUE)


```

> consumo.pos
[1] 3910 4220 3885 5160 5645
>
>
> sort(consumo.pos)
[1] 3885 3910 4220 5160 5645
>
>
> sort(consumo.pos,decreasing=TRUE)
[1] 5645 5160 4220 3910 3885
>
> |

```

94

Laboratório de Estatística - 2020

Comando order

- ✓ Fornece número de ordem de cada elemento do vetor
- ✓ `order(consumo.pos)`

```
>
> order(consumo.pos)
[1] 3 1 2 4 5
>
>
> ordem<-order(consumo.pos)
>
>
> consumo.pos[ordem]
[1] 3885 3910 4220 5160 5645
>
>
> consumo.pre[ordem]
[1] 5640 5260 5470 6180 6390
>
> |
```

Laboratório de Estatística - 2020

95



Funções Úteis

- Utilizem os comandos em objetos criados

```
length(objeto) # numero de elementos ou componentes
str(objeto) # estrutura de um objeto
class(objeto) # classe ou tipo de um objeto
names(objeto) # nomes

c(objeto,objeto,...) # combina objetos em um vetor
cbind(objeto, objeto, ...) # combina objetos como colunas
rbind(objeto, objeto, ...) # combina objetos como linhas

objeto # prints the object

ls() # lista objetos em uso
rm(objeto) # delete um objeto

novoobjeto <- edit(objeto) # edit copy and save as newobject
fix(objeto) # edit in place
```

Laboratório de Estatística - 2020

96



Criação de Data-frame



- Maneira diferente de criar data-frame
 - ✓ Comando `edit(data.frame())`
 - ✓ `dados.novo <- edit(data.frame())`
- É aberta janela com planilha e pode-se digitar informações do banco de dados.
- Para nomear a 1^a. variável, clica-se em “var1”
 - ✓ Janela para nomear a variável e seu tipo (numérica ou caractere).

Laboratório de Estatística - 2020

97

Exemplo

```
edit(data.frame())
```

R Editor de dados

Arquivo Editar Ajuda

var1	var2	var3	var4	var5
1				
2				
3				

Laboratório de Estatística - 2020

98



Carregando Pacotes

- Pacotes:
 - ✓ Conjuntos de funções específicas do R
 - ✓ No repositório do R está armazenada uma quantidade muito grande de pacotes que tem funções para um certo conjunto de tarefas
- ✓ Para usar um pacote:
 - baixar o pacote (download) do repositório
 - carregar o pacote na sua área de trabalho.

99

Laboratório de Estatística - 2020

Pacotes

- Quais pacotes estão disponíveis na sua instalação de R?
 - ✓ `library()`
- Interface hipertexto de ajuda:
 - ✓ `help.start()`
 - ✓ Escolher o link “Packages”
 - ✓ Clique no nome de um dos pacotes
 - Lista todos os objetos que este pacote contém.

100

Laboratório de Estatística - 2020

• Página do R Project

Package Index 

Packages in C:\Program Files\R\R-3.3.1\library

		A B C D E F G H I J K L M N O P Q R S T U V X Y Z
<code>abund</code>		Combine Multidimensional Arrays
<code>abundant</code>		Abundant regression and high-dimensional principal fitted components
<code>acepack</code>		ace() and avas() for selecting regression transformations
<code>ade4</code>		Analysis of Ecological Data : Exploratory and Euclidean Methods in Environmental Sciences
<code>abio</code>		A Collection of Statistical Tools for Biologists
<code>assertthat</code>		Easy pre and post assertions.
<code>base</code>		The R Base Package
<code>base64enc</code>		Tools for base64 encoding
<code>BB</code>		Solving and Optimizing Large-Scale Nonlinear Systems
<code>BH</code>		Boost C++ Header Files
<code>BHH2</code>		Useful Functions for Box, Hunter and Hunter II
<code>BiasedUrn</code>		Biased Urn Model Distributions
<code>BioGenerics</code>		S4 generic functions for Bioconductor
<code>BiocInstaller</code>		Install/Update Bioconductor, CRAN, and github Packages

101

Laboratório de Estatística - 2020

• Quais pacotes estão carregados na sua sessão?

- ✓ `search()`

• Instalação de pacote direto do R

- ✓ `install.packages("ISwR")`

102

Laboratório de Estatística - 2020

Exemplo

- Geração de amostra aleatória:

```
# 15 números aleatórios de uma distribuição normal, com média 1 e
# desvio-padrão=3

x1 <- rnorm(n = 15, mean = 1, sd = 3)
hist(x1)      # histograma de x1
truehist(x1)  # outro tipo de histograma
```
- Uso de função do pacote MASS:

```
search()
library(MASS)      # carrega pacote MASS
search()
truehist(x1)
help(package=MASS)  # ajuda sobre o pacote
```

103

Laboratório de Estatística - 2020

Demos

- Oferece demonstrações de como usar funções

```
# todos os demos (pacotes carregados)
demo()
# todos demos (pacotes instalados)
demo(package = .packages(all.available = TRUE))
demo(plotmath)          # demo sobre anotações em gráfico
demo(graphics)         # demos sobre tipos de gráficos
```

104

Laboratório de Estatística - 2020

Help

- Sempre perguntam para o Google, de preferência em inglês

105

Laboratório de Estatística - 2020

Conjuntos de Dados do R

- R traz vários conjuntos de dados internos, que são geralmente usados em demos ou exemplos
- Comando para ver a lista dos conjuntos de dados carregados:
√ data()

106

Laboratório de Estatística - 2020



Carregamento Built-in Data Set



- Carregamento do conjunto de dados:
✓ mtcars: Motor Trend Car Road Tests

```
# Conjunto de dados: mtcars: Motor Trend Car Road Tests
data(mtcars)           # carregamento
head(mtcars, 8)         # Print das primeiras 8 linhas
help(mtcars)            # informações sobre o banco
```

- Manipulação do conjunto de dados

```
nrow(mtcars)           # Número de linhas (observações)
ncol(mtcars)            # Número de colunas (linhas)
str(mtcars)              # estrutura do objeto
```

Laboratório de Estatística - 2020

107



• Colunas como vetores



```
mtcars$mpg                 # acesso à variável mpg
mtcars$disp                 # acesso à variável disp
pairs(mtcars, main = "Conjunto mtcars")
plot(mpg, disp)
# acesso às variáveis sem especificar banco
attach(mtcars)
plot(mpg, disp)             # plot das variáveis mpg e disp
objeto <- lm(disp ~ mpg)     # regressão linear
objeto
# resultados
summary(objeto)             # ajuste do modelo de regressão
abline(lm(disp ~ mpg))      # plot da reta de regressão
# estratificação do plot pela variável cyl
coplot(mpg ~ disp | as.factor(cyl), data = mtcars,
        panel = panel.smooth, rows = 1)
detach(mtcars)               # interromper acesso (final do trabalho)
```

Laboratório de Estatística - 2020

108