

Módulo II Descritiva

Importação de dados

Estatística descrtiva

pacotes gráficos

Análise gráfica Argumentos gráficos Outros

Introdução ao uso do software R

Fernando de Pol Mayer¹ Rodrigo Sant'Ana²

¹Universidade Federal de Santa Catarina (UFSC) Departamento de Ecologia e Zoologia (ECZ/CCB) fernando.mayer@qmail.com

> ²Instituto Albatroz oc.rodrigosantana@gmail.com

26 e 27 de Novembro, 2012



Sumário

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica

Argumentos gráficos Outros pacotes gráficos

Importação de dados

Estatística descrtiva

- Análise gráfica
 - Argumentos gráficos
 - Outros pacotes gráficos



Sumário

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos Importação de dados

2 Estatística descrtiva

- 3 Análise gráfica
 - Argumentos gráficos
 - Outros pacotes gráficos



Antes de importar para o R

Módulo II Descritiva

Importação de dados

Estatística descrtiva

- Se houverem valores perdidos, preencha com NAs
- A matriz de dados deve formar um bloco só. Se houverem colunas de diferentes comprimentos, preencha com NAs
- Salve o arquivo como "valores separados por vírgula" (.csv), mas atenção:
 - Se o separador de decimal for ",", o separador de campos será ";" automaticamente (o que é mais comum nos sistemas em português).



A função read.table()

Módulo II Descritiva

Importação de dados

Estatística descrtiva

gráficos

Análise gráfica Argumentos gráficos Outros O método mais comum de importação de dados para o R, é utilizando a função read.table(). Para importar um arquivo .csv faça:

Argumentos:

- "crabs.csv": nome do arquivo
- header = T: significa que a primeira linha do arquivo deve ser inrpretada como os nomes das colunas
- sep = ";": o separador de colunas (também pode ser "\t" para tabulação e "" para espaços)
- dec = ",": o separador de decimais



Estrutura dos dados importados

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes

gráficos

A função str() serve para demonstrar a estrutura de um objeto, como o nome das colunas e suas classes:

```
> str(dados)
'data.frame': 156 obs. of 7 variables:
 $ especie: Factor w/ 2 levels "azul","laranja": 1 1 1 1 1 1
          : Factor w/ 2 levels "F", "M": 2 2 2 2 2 2 2 2 2 2 ...
 $ sexo
  FL
                8.1 8.8 9.2 9.6 10.8 11.6 11.8 12.3 12.6 12.8
  RW
                6.7 7.7 7.8 7.9 9 9.1 10.5 11 10 10.9 ...
          : num
                 16.1 18.1 19 20.1 23 24.5 25.2 26.8 27.7 27.4
  CL
           num
  CW
                 19 20.8 22.4 23.1 26.5 28.4 29.3 31.5 31.7 31
           num
 $ BD
                7 7.4 7.7 8.2 9.8 10.4 10.3 11.4 11.4 11 ...
```



Estrutura dos dados importados

Módulo II Descritiva

Para visualizar as 10 primeiras linhas do data frame:

Importação de dados

Estatística descrtiva

```
> head(dados, 10) # ou: dados[1:10,]
                   FI
                                   CW
   especie sexo
                        RW
                             CL
                                        BD
                       6.7 16.1 19.0
      azul
                 8.1
                                       7.0
2
      azul
                8.8
                       7.7 18.1 20.8
3
              M 9.2
                       7.8 19.0 22.4
      azul
4
      azul
                  9.6
                       7.9 20.1 23.1
5
      azul
                       9.0 23.0 26.5
              M 10.8
6
      azul
                       9.1 24.5 28.4 10.4
                11.6
7
      azul
                11.8 10.5 25.2 29.3 10.3
8
      azul
                12.3 11.0 26.8 31.5 11.4
9
      azul
              M 12.6 10.0 27.7 31.7 11.4
10
      azul
              M 12.8 10.9 27.4 31.5 11.0
```



Sumário

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos 1 Importação de dados

2 Estatística descrtiva

- 3 Análise gráfica
 - Argumentos gráficos
 - Outros pacotes gráficos



Medidas de centro

Módulo II Descritiva Média (utilizando a coluna CL):

portação de

[1] 32.004

Média (utilizando a coluna BD):

> mean(dados\$BD)

[1] NA

Observe que a presença de NAs interfere no resultado. O argumento na.rm pode ser utilizado para que o cálculo seja feito sem os NAs:

> mean(dados\$CL) # ou sum(dados\$CL)/length(dados\$CL)

> mean(dados\$BD, na.rm=T)

[1] 14.019

O argumento na.rm é utilizado também em outras funções como median(), sum() e var()

Importação de dados

Estatística descrtiva



Medidas de variação

Módulo II Descritiva

Estatística descrtiva Análise gráfica

Argumentos gráficos Outros pacotes

gráficos

Desvio padrão (coluna CL):

Importação de dados \$CL)

[1] 7.2278

Variância:

> var(dados\$CL) # sd(dados\$CL)^2

[1] 52.242

Coeficiente de variação:

> sd(dados\$CL)/mean(dados\$CL)

[1] 0.22584



Medidas de posição relativa

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos

Quartis:

> quantile(dados\$CL)

```
0% 25% 50% 75% 100% 14.700 27.075 32.300 37.050 47.600
```

Decis:

```
> quantile(dados$CL, probs = seq(0, 1, 0.1))
```

```
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 14.70 22.10 25.40 28.15 30.10 32.30 34.20 36.10 38.10 41.40 100% 47.60
```



Sumários

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos A função summary() resume os dados de forma lógica:

> summary(dados\$CL)

```
Min. 1st Qu. Median Mean 3rd Qu. Max. 14.7 27.1 32.3 32.0 37.0 47.6
```

1st Qu. e 3rd Qu. se referem ao 1^o e 3^o quartis. Se houverem NAs, eles também serão contados:

```
> summary(dados$BD)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
6.1 11.5 13.8 14.0 16.6 21.6 2
```



Sumários

Módulo II Descritiva

Importação de

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos A função summary() pode ser utilizada para resumir todo o data frame:

```
> summary(dados)
    especie
                             FL
                                             RW
               sexo
 azul
        :77
               F:87
                      Min.
                             : 7.2
                                      Min.
                                              : 6.5
 laranja:79
               M:69
                      1st Qu.:12.9
                                      1st Qu.:11.0
                      Median: 15.6
                                      Median: 12.8
                      Mean
                              :15.6
                                      Mean
                                              :12.8
                      3rd Qu.:18.2
                                      3rd Qu.:14.5
                      Max.
                             :23.1
                                      Max.
                                              :20.2
                      NA's
                              :2
       CL
                       CW
                                        BD
         :14.7
                                 Min.
 Min.
                 Min.
                         :17.1
                                         : 6.1
 1st Qu.:27.1
                 1st Qu.:31.4
                                 1st Qu.:11.5
 Median:32.3
                 Median:36.9
                                 Median:13.8
 Mean
         :32.0
                 Mean
                         :36.3
                                 Mean
                                         :14.0
 3rd Qu.:37.0
                 3rd Qu.:41.9
                                 3rd Qu.:16.6
         :47.6
 Max.
                 Max.
                         :54.6
                                 Max.
                                         :21.6
                                                              13 / 49
                                 NA's
```



Tabelas de contingência

Módulo II Descritiva

A função table() é usada para formar tabelas de contingência:

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos > table(dados\$especie)

Pode também ser utilizada para a contagem de combinações de fatores:

> table(dados\$especie, dados\$sexo)

```
F M azul 43 34 laranja 44 35
```



As funções *apply()

Módulo II Descritiva A função lapply() retorna uma lista com os resultados de uma função aplicada a elemento(s) de um data frame:

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos

```
> lapply(dados[, 3:7], mean) # na.rm = T para lidar com NAs
$FL
```

\$RW

[1] NA

[1] 12.815

\$CL

[1] 32.004

\$CW

[1] 36.31

\$BD

[1] NA



As funções *apply()

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos sapply() simplifica os resultados em um vetor:

```
> sapply(dados[, 3:7], mean, na.rm = T)

FL    RW    CL    CW    BD
15.577 12.815 32.004 36.310 14.019
```

apply() aplica a função especificada por linha (1) ou coluna (2):

```
> apply(dados[, 3:7], 2, mean, na.rm=T)

FL    RW    CL    CW    BD
15.577 12.815 32.004 36.310 14.019
```



As funções *apply()

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos A função tapply() aplica a função (mean) a um vetor (dados\$CL) para cada combinação dos níveis dos fatores (list()):

> tapply(dados\$CL, list(dados\$especie, dados\$sexo), mean)

F M azul 28.044 32.176 laranja 34.570 33.474



A função aggregate() ("tabela dinâmica")

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos A função aggregate() agrega valores utilizando uma função, de acordo com uma lista de fatores especificados:

```
> aggregate(cbind(FL, RW, CL, CW, BD) ~ especie + sexo,
            data = dados, median, na.rm=T)
  especie sexo
                 FI
                      RW
                             CL
                                   CW
                                        BD
     azul
             F 13.3 12.2 28.15 32.65 11.8
 laranja
             F 18.0 14.6 34.70 39.40 15.7
3
     azul
             M 15.3 12.0 32.75 37.35 13.6
4 laranja
             M 16.3 12.0 32.30 35.30 14.7
```



Sumário

Módulo II Descritiva

Importação de dados Estatística

descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos 1 Importação de dados

2 Estatística descrtiva

- 3 Análise gráfica
 - Argumentos gráficos
 - Outros pacotes gráficos



Gráficos

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos Os comandos para gráficos se dividem em três grupos:

- Alto-nível: são as funções que criam novos gráficos, com eixos, títulos, etc. Exemplo: plot(), hist(), barplot()
- Baixo-nível: funções que adicionam mais informações (pontos, linhas, texto, etc.) a um gráfico já existente Exemplo: lines(), points(), legend()
- Interativo: funções que permitem adicionar ou extrair informações a um gráfico já existente, de modo interativo. Exemplo: locator()

Ainda existem uma série de parâmetros gráficos que podem ser manipulados e customizados.



Gráficos Terminologia

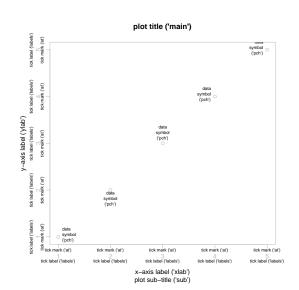
Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos

Outros pacotes gráficos





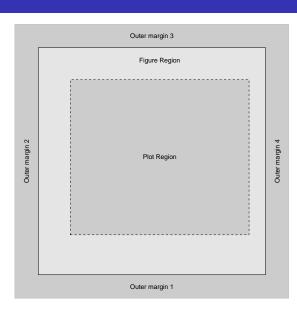
Gráficos Regiões do gráfico

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise grá<u>fica</u>





Gráficos Regiões de múltiplos gráficos

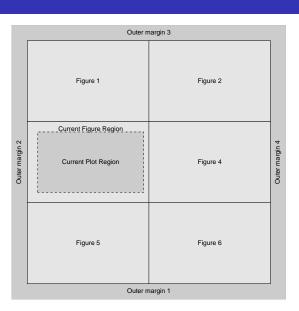
Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos

gráficos Outros pacotes gráficos





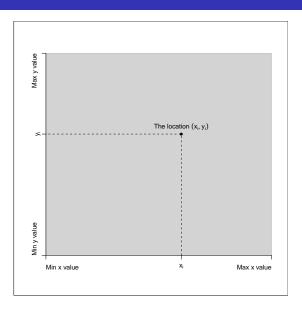
Gráficos Sistema de coordenadas

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica





Gráficos Tipos de linhas (lty)

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica

Integer	Sample line	String
Predefine	d	
0		"blank"
1		"solid"
2		"dashed"
3		"dotted"
4		"dotdash"
5		"longdash"
6		"twodash"
Custom		
		"13"
	$ \cdot$	"F8"
		"431313"
		"22848222"



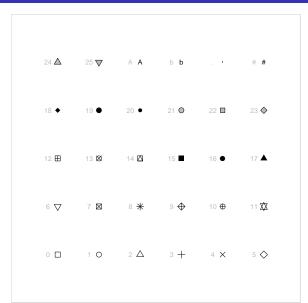
Gráficos Tipos de pontos (pch)

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise grá<u>fica</u>



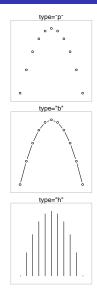


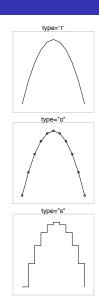
Gráficos Tipos de plot (type)

Módulo II Descritiva

Importação de dados

Estatística descrtiva







Gráficos Expressões matemáticas (expression())

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica

Argumentos gráficos Outros pacotes gráficos

Temperature (°C) in 2003

expression(paste("Temperature (", degree, "C) in 2003"))

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

expression(bar(x) == sum(frac(x[i], n), i==1, n))

$$\hat{\beta} = (X^t X)^{-1} X^t y$$

expression(hat(beta) == $(X^t * X)^{-1} * X^t * y$)

$$\mathbf{z} = \sqrt{\mathbf{x}^2 + \mathbf{v}^2}$$

expression(z[i] == sqrt(x[i]^2 + y[i]^2))



Módulo II Descritiva

Importação de dados Estatística

descrtiva

Análise gráfica

Argumentos gráficos Outros pacotes gráficos A função plot() é uma **função genérica**. Os gráficos serão representados de forma diferente dependendo da **classe** dos objetos. A forma geral desta função é:

Note que y não precisa ser necessariamente especificado. O mesmo resultado pode ser obtido em formato de **fórmula**:

Onde lê-se "y descrito utilizando-se x", ou "y em função de x"

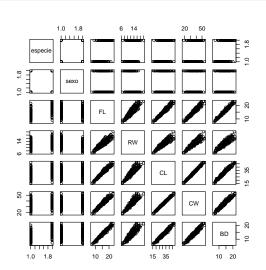


Módulo II Descritiva

> plot(dados)

Importação de dados

Estatística descrtiva





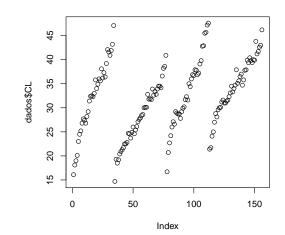
Módulo II Descritiva

> plot(dados\$CL)

uma variável

Importação de dados

Estatística descrtiva





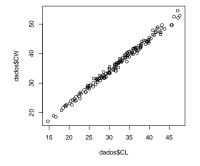
Módulo II Descritiva

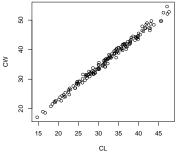
Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos > plot(dados\$CL, dados\$CW) # duas variáveis, ou:

> plot(CW ~ CL, data = dados) # mesmo resultado





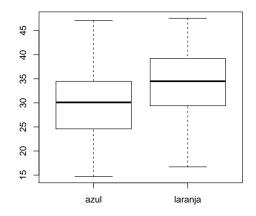


Módulo II Descritiva

> plot(dados\$especie, dados\$CL) # fator, numérico

Importação de dados

Estatística descrtiva





Módulo II Descritiva

Importação de dados

Estatística descrtiva Análise gráfica

Argumentos gráficos

Os **boxplots** são úteis para revelar o centro, a dispersão e a distribuição dos dados, além de outliers. São construídos da seguinte forma:

- A linha central mais escura representa a mediana. Os extremos da caixa são o 1° (q1) e o 3° (q3) quartis.
- As linhas que se extendem das caixas são definidas como:

$$q1 - 1, 5 \cdot IQR$$
 e $q3 + 1, 5 \cdot IQR$

$$q3+1, 5 \cdot IQR$$

onde IQR é o intervalo inter-quartil. As linhas vão até os valores máximo e mínimo que ainda se encontram dentro deste intervalo.

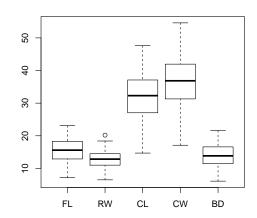


Módulo II Descritiva

> boxplot(dados[, 3:7])

Importação de dados

Estatística descrtiva



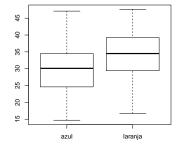


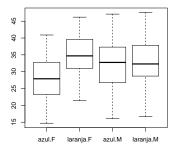
Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos > boxplot(CL ~ especie, data = dados) # um fator
> boxplot(CL ~ especie + sexo, data = dados) # dois fatores







Módulo II Descritiva

Importação de dados Estatística

descrtiva

Análise grá<u>fica</u>

Argumentos gráficos Outros pacotes gráficos Na função hist(), dois argumentos controlam a configuração das classes:

Argumentos	resultado
<pre>include.lowest = T, right = T</pre>	
include.lowest = F, right = T	(a,b],, (c,d]
include.lowest = F, right = F	[a,b),, [c,d)
include.lowest = T, right = F	[a,b),, [c,d]

O argumento breaks também pode ser utilizado para especificar as "quebras" das classes.



Módulo II Descritiva

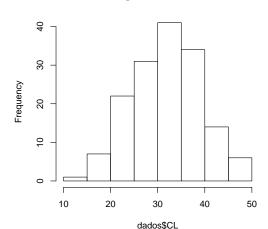
> hist(dados\$CL)

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos

Histogram of dados\$CL





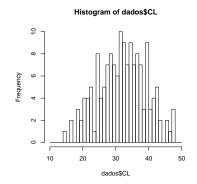
Módulo II Descritiva

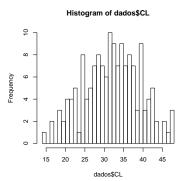
Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos > hist(dados\$CL, breaks = seq(10, 50, 1)) # ou

> hist(dados\$CL, nclass = 40) # aproximado







> h <- hist(dados\$CL)</pre>

Módulo II Descritiva

Importação de dados **F**statística

descrtiva Análise gráfica

Argumentos gráficos Outros pacotes gráficos

\$breaks [1] 10 15 20 25 30 35 40 45 50

> h

\$counts [1] 1 7 22 31 41 34 14 6

\$intensities

[1] 0.0012821 0.0089744 0.0282051 0.0397436 0.0525641 0.0435897

\$density

\$mids

[7] 0.0179487 0.0076923

0.0179487 0.0076923

uma lista com as uma série de informações:

[1] 0.0012821 0.0089744 0.0282051 0.0397436 0.0525641 0.0435897

40 / 49

A função hist(), além de produzir histogramas também retorna



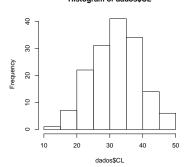
Módulo II Descritiva

- Importação de dados
- Estatística

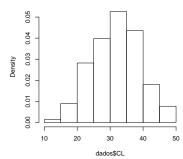
descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos > hist(dados\$CL) # contagem
> hist(dados\$CL, freg = F) # densidade

Histogram of dados\$CL



Histogram of dados\$CL





A função barplot()

Módulo II Descritiva

Importação de dados

Estatística

descrtiva Análise gráfica

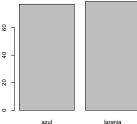
Argumentos gráficos Outros pacotes gráficos

A função barplot() gera gráficos de barra a partir de uma tabela

> table(dados\$especie)

azul laranja 77 79

> barplot(table(dados\$especie))





A função barplot()

Módulo II Descritiva

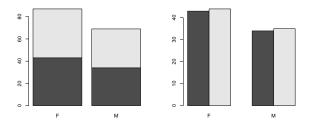
Importação de dados

Estatística descrtiva

Análise gráfica

Argumentos gráficos Outros pacotes gráficos > table(dados\$especie, dados\$sexo)

- > barplot(table(dados\$especie, dados\$sexo))
- > barplot(table(dados\$especie, dados\$sexo), beside = T)





Sumário

Módulo II Descritiva

Importação de dados Estatística descrtiva

Análise gráfica

Argumentos gráficos Outros pacotes gráficos Importação de dados

2 Estatística descrtiva

- 3 Análise gráfica
 - Argumentos gráficos
 - Outros pacotes gráficos



Argumentos gráficos

Módulo II Descritiva

Importação de dados Estatística

descrtiva

Análise gráfica

Argumentos gráficos

Outros pacotes gráficos Os principais argumentos que servem uma grande quantidade de funções gráficas são

- xlab, ylab: ("caracter") alteram o nome dos eixos (x e y label)
- xlim, ylim: (c(min, max)) alteram os limites dos eixos (x e y limits)
- main: ("caracter") altera o título do gráfico



Argumentos gráficos

Módulo II Descritiva

Importação de dados

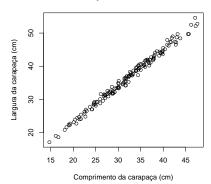
Estatística descrtiva

Análise gráfica

Argumentos gráficos Outros

Outros pacotes gráficos

Relação entre CL e CW





Argumentos gráficos

Módulo II Descritiva

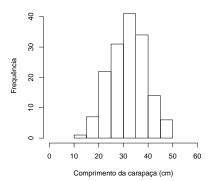
Importação de dados

Estatística descrtiva

Análise gráfica

Argumentos gráficos

Outros pacotes gráficos





Sumário

Módulo II Descritiva

Importação de dados Estatística

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes gráficos

- 1 Importação de dados
- Estatística descrtiva
- 3 Análise gráfica
 - Argumentos gráficos
 - Outros pacotes gráficos



Ouros pacotes gráficos

Módulo II Descritiva

Importação de dados

Estatística descrtiva

Análise gráfica Argumentos gráficos Outros pacotes

gráficos

Além do pacote de gráficos padrão do R, **graphics**, existem outros que incorporam opções mais avançadas

- lattice (mais antigo)
- ggplot2 (mais recente)