

Linguagem R

R-Gui e R-Commander

1 – INSTALAÇÃO DO PROGRAMA BÁSICO E PACOTES

O sistema é formado por um programa básico (R-Gui) e muitos pacotes com procedimentos adicionais. Tudo é gratuito e com código livre, voltado para comunidade acadêmica.

A página central do R é:

<http://www.r-project.org/>

Para fazer o download, você deve usar um “Cran Mirror”, clicando em CRAN na margem esquerda. No Brasil, temos:

<http://cran-r.c3sl.ufpr.br/> Universidade Federal do Parana

<http://cran.fiocruz.br/> Oswaldo Cruz Foundation, Rio de Janeiro

<http://www.vps.fmvz.usp.br/CRAN/> University of Sao Paulo, Sao Paulo

<http://brieger.esalq.usp.br/CRAN/> University of Sao Paulo, Piracicaba

No Cran escolhido, você escolhe o sistema operacional de sua máquina (p. ex., Windows) e faça o download do programa básico: “Base”. Observe que na página do Download tem algumas informações úteis.

Para instalar o programa básico, basta executar o arquivo *R-(versão) - (sistema operacional).exe* que você baixou. Para executar, pode usar o ícone no desktop. A Figura 1, lado esquerdo, mostra a tela inicial do R-Gui.

Os comandos da linguagem R podem ser feitos diretamente no R-Gui, mas para pessoas que não estão muito acostumadas com o R (como eu!), pode fazer uso do pacote R-Commander ou Rcmdr, o qual é uma máscara dos comandos, permitindo usar os procedimentos básicos através de Menu. O lado direito da Figura 1 mostra como fazer a instalação do Rcmdr.

Qualquer pacote do R que você instala, o próprio sistema irá instalar outros pacotes que são necessários para o pacote que você está instalando. Esta instalação pode ocorrer no momento da instalação ou quando você vai usar pela primeira vez.

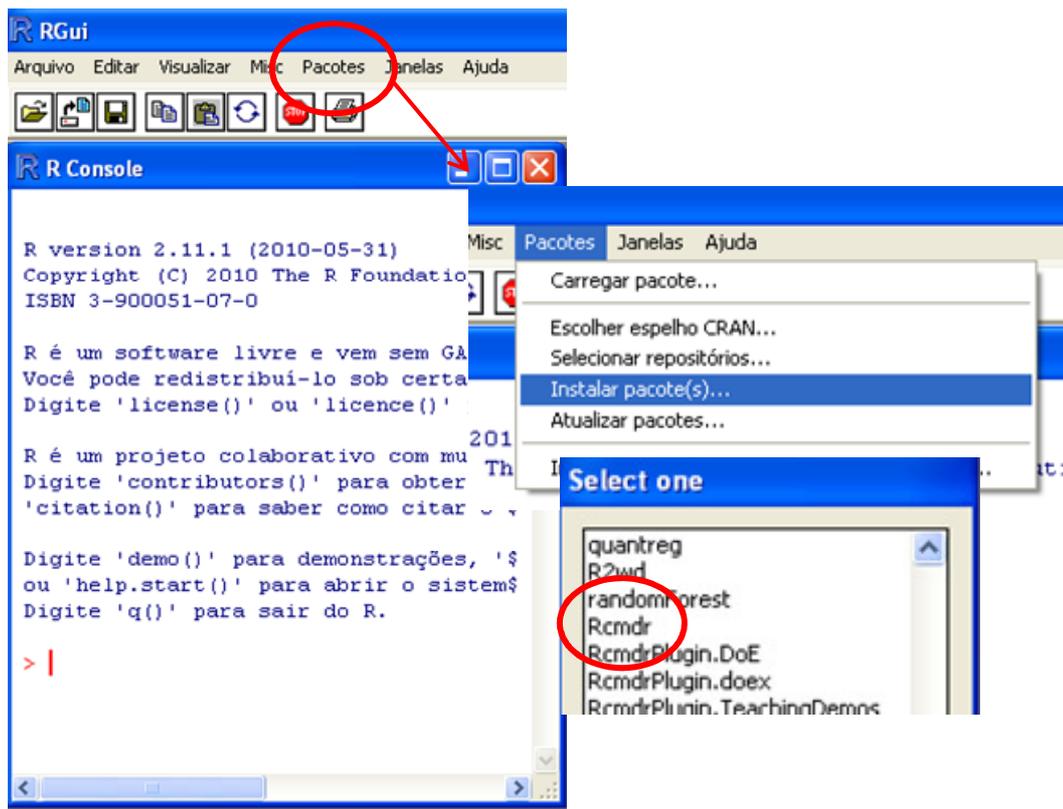


Figura 1 – Tela do R-Gui e ilustração de como instalar o pacote Rcmdr via menu.

Instalando o Rcmdr através do R-Gui:

- Pacotes
- Instalar pacotes
- Escolher o “cran mirror”
- Escolher o pacote Rcmdr

Para usar o R-Commander, voltar ao Menu do R-Gui e clicar em:

- Pacotes
- Carregar pacotes
- Rcmdr

Na primeira vez, o sistema deve carregar alguns pacotes para o funcionamento do R-Commander. O sistema pode demorar um pouco para carregar os pacotes.

2 – INICIANDO O R-COMMANDER

Abrir o R-Gui (no ícone R)

Carregar o R-Commander (no Menu do R-Gui: *Pacotes >> Carregar Pacotes >> Rcmdr*)

O R-Commander abre uma tela, como mostrada na Figura 2. Você pode solicitar uma tarefa ao sistema pelo Menu, sendo que o Script dessa tarefa é registrado automaticamente na *janela Script*, e os resultados na *janela de Resultados* (abaixo).

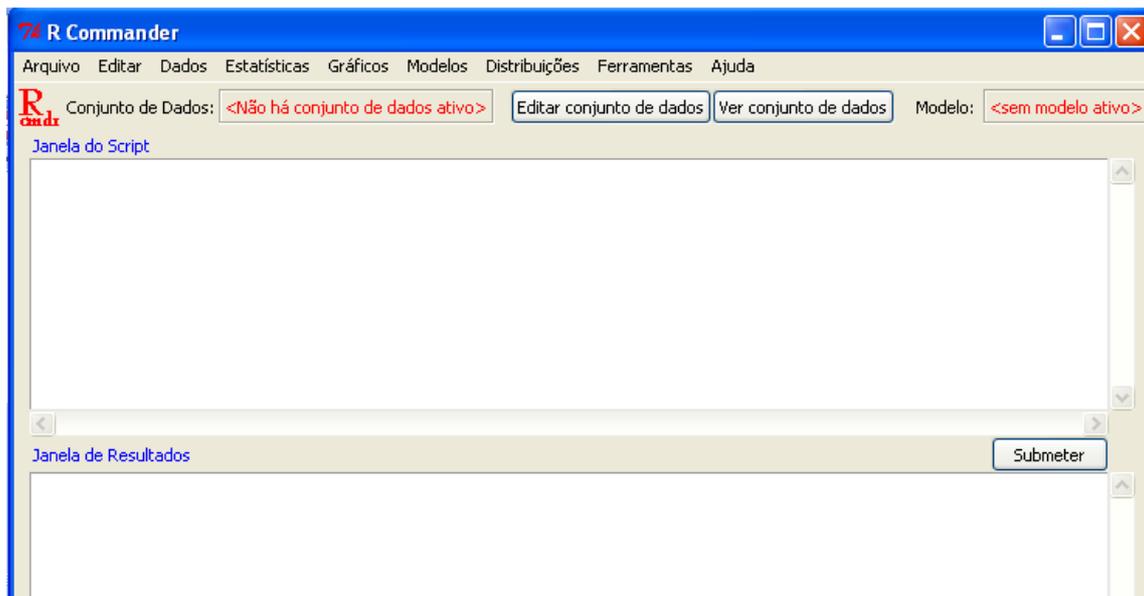


Figura 1 – Tela inicial do Rcmdr.

2.1 – Importando um arquivo de dados

Normalmente, as análises estatísticas são feitas sobre um conjunto de dados. Em geral, os dados são digitados usando programas próprios para digitação ou planilhas eletrônicas. Em se usando planilhas eletrônicas, na primeira linha devem ser colocados os nomes das variáveis. Considere que seus dados estão armazenados na *Planilha 1* do arquivo *Dinheiro_no_Bolso.xls*, do Microsoft Excel.

A Figura 3 ilustra o processo, que se inicia no Menu Dados. Você deve atribuir um nome para o arquivo de dados sem espaço e sem caracteres especiais (atribuímos o nome *dados*). O sistema vai te orientar para você localizar seu arquivo de dados. Se o arquivo que você escolher tiver mais de uma planilha, o sistema abre nova janela para você marcar a planilha desejada.

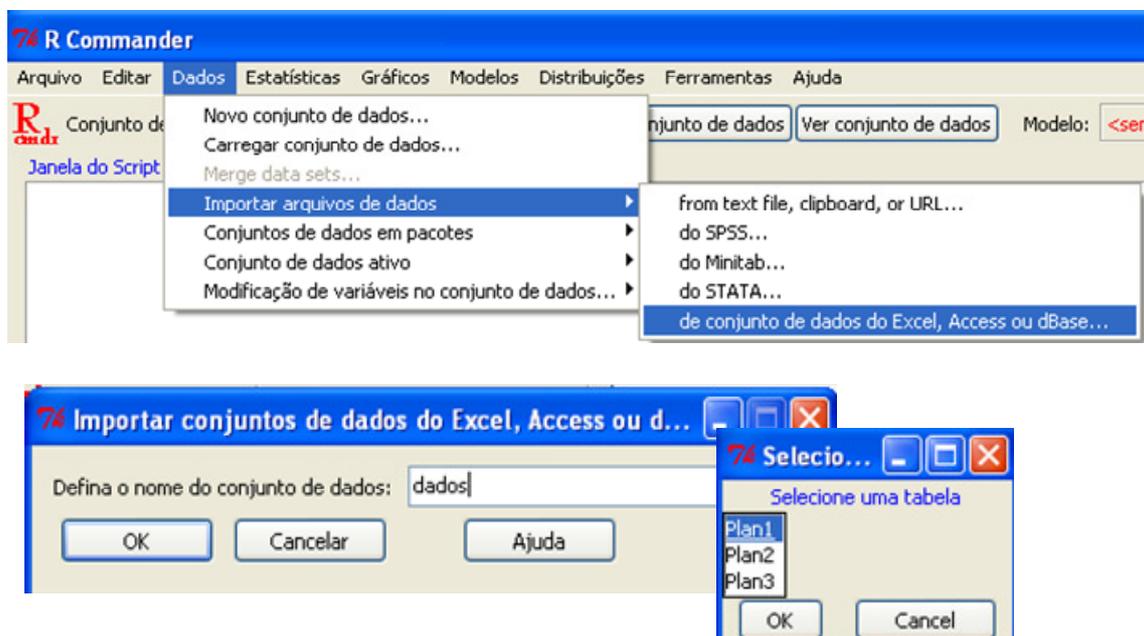


Figura 3 – Importando um arquivo do Excel.

Observe, na Figura 3, que você pode importar dados de várias fontes. Na importação são usados arquivos de fontes de dados ODBC do próprio sistema operacional, de tal forma que há compatibilidade de formatos. Por exemplo, se seu sistema está configurado com decimais no formato português (decimal com vírgula), ao passar para o R, é automaticamente colocado decimal em formato de ponto. Por outro lado, pode haver problema se misturar ambientes de 32 e 64 bits.

Para ver ou fazer edição simples no arquivo de dados, pode usar os botões “Ver conjunto de dados” e “Editar conjunto de dados” do Menu do R-Commander. A Figura 4 mostra o arquivo no Excel e no R. Observe que na janela do Script aparecem os comandos para importação e para mostrar o conjunto de dados. Se você quiser conhecer melhor o comando, basta digitar:

? (nome do comando)

Marcar esse texto e clicar “submeter”. O R mostra em HTML toda a sintaxe do comando.

The image shows a screenshot of the R Commander interface. On the left, the 'Janela do Script' (Script Window) contains the following R code:

```
dados <- sqlQuery(channel =  
library(relimp, pos=4)  
showData(dados, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,  
maxheight=30)
```

On the right, there is a preview of an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	Ident	Curso_sem	sexo	altura	idade	dinheirobolso
2	1	CCO_2009_2	M	182	19	3,00
3	2	CCO_2009_2	F	177	17	20,00
4	3	CCO_2009_2	M	172	21	13,00
5	4	CCO_2009_2	M	183	25	15,00
6	5	CCO_2009_2	M	182	18	22,00

Below the spreadsheet, a window titled 'dados' displays the same data in a table format:

	Ident	Curso_sem	sexo	altura	idade	dinheirobolso
1	1	CCO_2009_2	M	182	19	3.00
2	2	CCO_2009_2	F	177	17	20.00
3	3	CCO_2009_2	M	172	21	13.00
4	4	CCO_2009_2	M	183	25	15.00
5	5	CCO_2009_2	M	182	18	22.00

Figura 4 – As cinco primeiras observações do arquivo de dados no Excel e no R.

2.2 – Manipulando e salvando um arquivo de dados

Várias manipulações de dados podem ser feitas através do Menu do R-Commander:

- Dados >> Conjunto de Dados Ativo, e
- Dados >> Modificação de variáveis no conjunto de dados

Como exemplo, vamos considerar que queremos criar a variável *faixa_etaria*, através da recodificação de *idade*, com duas categorias: *até 18* e *mais de 18*. Caminhamos pelo Menu:

- Dados >> Modificação de variáveis no conjunto de dados >> Recodificar variáveis.

Na janela que se abre, marcar a variável *idade*, preencher o nome da nova variável (*faixa_etaria*) e as definições p/ recodificação (Figura 5). É claro que eu não lembrava como deveriam ser feitas as “definições p/ recodificação”, então cliquei em “Ajuda”. Após executar a recodificação, se clicarmos em “Ver conjunto de dados”, veremos uma janela como está do lado direito da Figura 5.

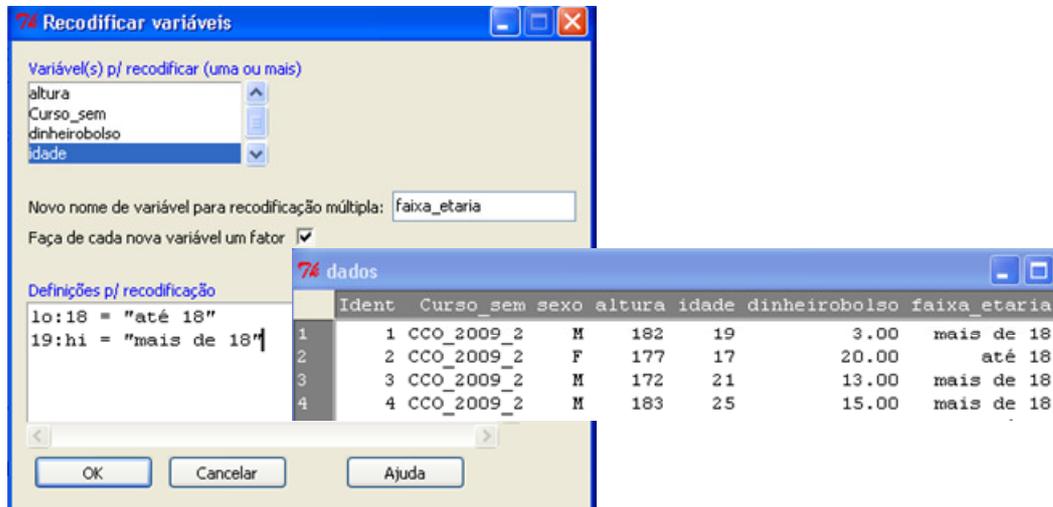


Figura 5 – Processo de recodificação de variável e o novo arquivo de dados.

Muitas vezes representamos as categorias por códigos: 1, 2, ... Para o sistema “entender” que são códigos e não números propriamente dito, você deve identificar a variável como **fator**. Caso estiver como numérica, você pode transformá-la em fator usando:

- Dados >> Modificação de variáveis no conjunto de dados >> Converter variável numérica em fator.

Para **salvar** o novo conjunto de dados, fazemos:

- Dados >> Conjunto de Dados Ativo >> Salvar conjunto de dados ativo

O arquivo será salvo em formato próprio do R (.rda), mas também pode escolher

- Dados >> Conjunto de Dados Ativo >> Exportar conjunto de dados ativo

para gravar em outro formato.

O R pode carregar mais de um conjunto de dados. Se estiver trabalhando com vários arquivos de dados, certifique-se qual deles está ativo: no Menu, “Conjunto de Dados:” Para alterar o conjunto de dados ativo,

- Dados >> Conjunto de dados ativo >> Selecionar conjunto de dados ativo

3 – DISTRIBUIÇÃO DE FREQUÊNCIAS

Para o leitor não ter uma decepção com alguns procedimentos introdutórios, observamos que o R não é voltado para simples descrição de dados, mas para

procedimentos estatísticos propriamente ditos. Distribuições de freqüências em forma tabular, ou gráficos de distribuição de freqüências de variáveis qualitativas (gráfico em barras, colunas e setores) são melhores construídos nas planilhas eletrônicas (p. ex., Excel), com a função Tabela Dinâmica.

3.1 – Tabelas de distribuição de freqüências simples

Para obter uma tabela de distribuição de freqüências, com classificação simples, de **variável qualitativa**, no Menu do R-Commander, fazer:

- Estatística >> Resumos >> Distribuições de frequências

Marcar a variável qualitativa ou fator de interesse e OK. Os resultados aparecem em formato texto.

Para **variável quantitativa discreta**, você pode transformá-la em fator (Dados >> Modificação de variáveis no conjunto de dados >> Converter variável numérica em fator) e proceder de forma similar ao processo com variável qualitativa.

Para **variável quantitativa contínua** (ou mesmo uma discreta com muitas categorias), você precisa recodificá-la (Dados >> Modificação de variáveis no conjunto de dados >> Recodificar variáveis, como ilustrado na seção 2). Com a variável recodificada, proceder de forma similar ao processo com variável qualitativa.

3.2 – Tabelas de distribuição de freqüências com dupla classificação

Quando se quer a distribuição de frequências conjunta de duas variáveis qualitativas, fazer:

- Estatística >> Tabelas de contingência >> Tabela de dupla entrada

Na janela que se abre, você deve marcar a variável qualitativa (ou fator) que vai na linha; idem para a que vai na coluna. Você também pode escolher se quer as freqüências absolutas (marcar “sem percentual”) ou escolher o tipo de percentual que quer.

3.3 – Gráfico de distribuição de freqüências

Para **variável qualitativa**, pode-se fazer um gráfico de colunas (ou barras verticais). No Menu do R-Commander, fazer:

- Gráficos >> Gráfico de barras

Marcar a variável qualitativa ou fator de interesse e OK.

Para **variável quantitativa contínua** (ou mesmo uma discreta com muitas categorias), o gráfico padrão de distribuição de freqüências é o histograma. Fazer:

- Gráficos >> Histograma

Na janela que se abre, marcar a variável de interesse. Você pode escolher o número de classes e escolher a medida de frequência no eixo vertical: frequência absoluta (contagem), frequência relativa (Porcentagem) ou, ainda, a densidade (padronização para a área total ser igual à unidade). Ver Figura 6.

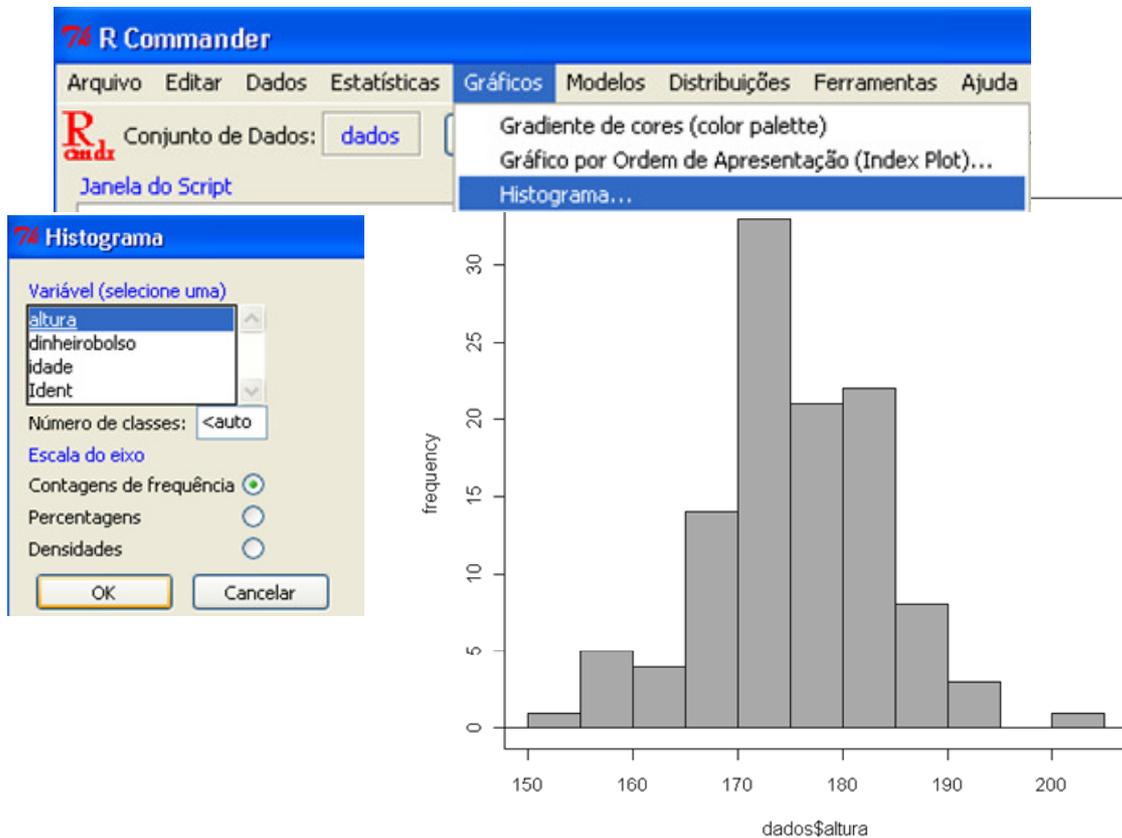


Figura 6 – Histograma da variável *altura* feito através do Menu do R-Commander.

No R não se tem flexibilidade de editar o gráfico, mas pode-se usar o **comando** que aparece na janela de Script para fazer o gráfico com aparência diferente. Observando essa janela, verifica-se que o comando é “Hist”. Então, digitar:

? Hist

Marcar e clicar em submeter. Na ajuda sobre o comando, verifica-se, por exemplo, que com os argumentos “xlab=” e “ylab=” você pode escolher os títulos dos eixos. Mudamos também a cor (col=”blue”). Então, repetimos o comando com esses argumentos e submetemos (Figura 7).

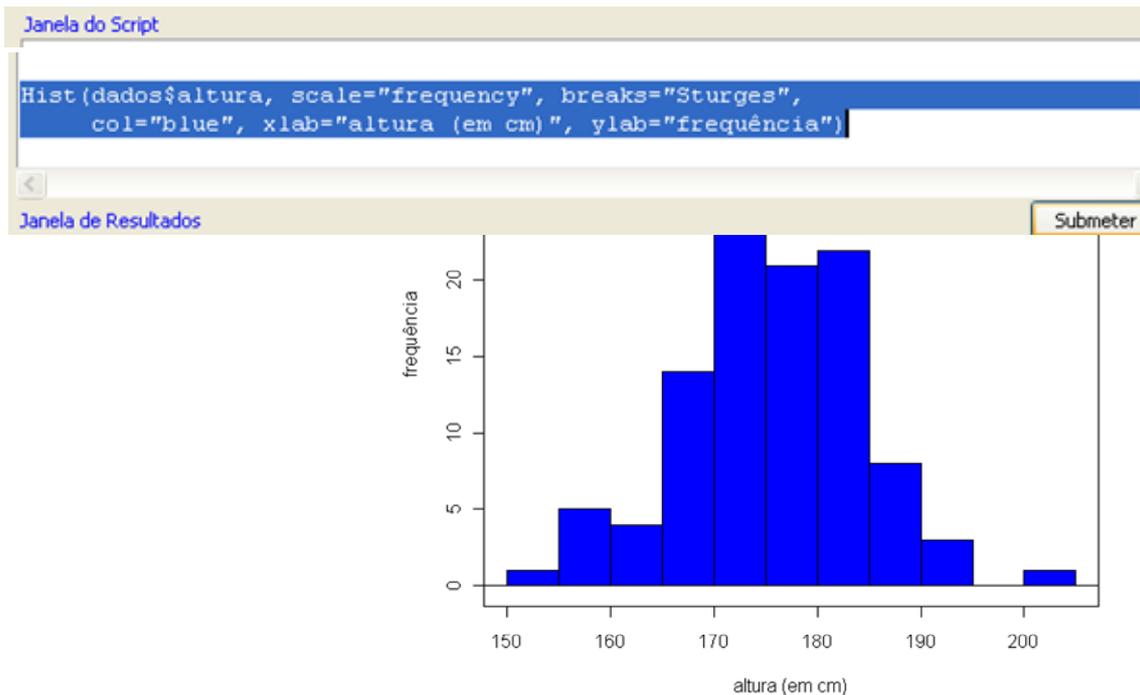


Figura 7 – Histograma da variável *altura* feito através da janela do Script.

O histograma é um gráfico bastante típico da Estatística, porém você teria alguma dificuldade de fazer um histograma adequado numa planilha eletrônica.

4 – MEDIDAS DESCRITIVAS

Medidas descritivas de uma ou mais variáveis quantitativas podem ser facilmente feitas pelo Menu do R-Commander:

- Estatística >> Resumos >> Resumos numéricos

Na janela que se abre você pode desmarcar medidas que não queira ou acrescentar outros percentis (quantis). Se quiser, você pode obter separadamente as medidas por categoria de uma variável qualitativa ou fator. O resultado sai em formato texto na janela de Resultados. Como exemplo, escolhemos a variável quantitativa *altura*, com resultados separados por *sexo* (Figura 8).



	mean	sd	0%	25%	50%	75%	100%	n
F	164.0833	6.529073	153	160	162.5	168.25	177	12
M	177.7300	6.955639	160	173	177.0	182.00	202	100

Figura 8 – Medidas descritivas da *altura*, por *sexo*.

4 – DIAGRAMA EM CAIXAS

Você pode ter a representação de sua distribuição através de diagramas em caixas (Boxplots). Seguir o caminho:

- Gráficos >> Boxplot

Como exemplo, marcamos a variável *altura* e escolhemos a “variável de grupo” *sexo*. Ver Figura 9.

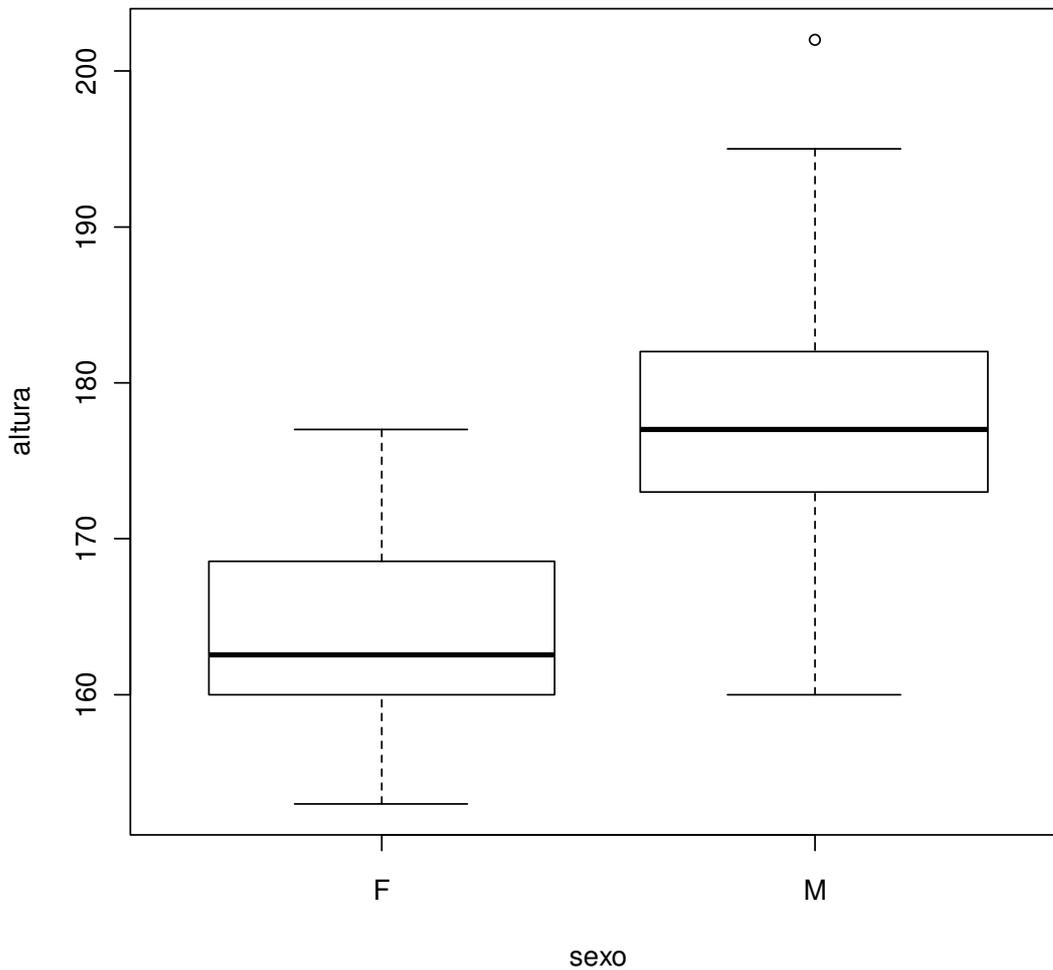


Figura 9 –Diagramas em caixas (Boxplots) de *altura* das mulheres e dos homens.

5 – OUTROS PROCEDIMENTOS

Os tópicos anteriores mostraram como baixar, instalar e realizar alguns procedimentos elementares pelo R-Commander. A medida que você aprende novos métodos estatísticos ou tenha novas necessidades de manipulação do arquivo de dados, você pode explorar o Menu do R-Commander para realizar as tarefas.

Voltamos a lembrar que nos próprios repositórios ou Crans Mirror do R têm documentações para aprendizagem dessa linguagem. Materiais em português também podem ser localizados pela Internet.