

1

Introdução ao R

O R é uma linguagem e ambiente utilizada na computação estatística e na elaboração de gráficos. É disponibilizado sob os termos da GNU – *General Public License* – da Fundação do Software Livre (www.gnu.org/fsf/fsf.pt.html). O R é um *software* livre, ou seja, os usuários têm liberdade para executá-lo, copiá-lo, distribuí-lo, estudá-lo, modificá-lo e aperfeiçoá-lo. Todo o código fonte do R é aberto.

É exatamente por esse motivo que o R vem se tornando, nos últimos anos, o ambiente mais utilizado pelos estatísticos para o desenvolvimento de novos métodos. Por um lado, isso significa que todas as funções e bibliotecas do R são amplamente utilizadas pelos especialistas da área, o que contribui para aumentar a qualidade do produto. Por outro, garante que os métodos mais recentes apareçam, em primeiro lugar, no R.

O R está disponível na Internet no *website* do CRAN – *Comprehensive R Archive Network*, ou Rede Completa de Arquivos do R – em www.r-project.org e diversos *mirrors*, inclusive no Brasil. Pode-se baixar o código fonte e compilar localmente, ou instalar o *software* pré-compilado, atualmente disponível para a família Unix (incluindo Linux), a maior parte dos MacOS e ainda Windows 9x/NT/2000.

Para instalar, entre no link CRAN, da página principal do R. Escolha o local mais próximo de você. Em seguida, escolha o sistema operacional. No *link base* está o R (neste momento é o `rw2011.exe`, com 25Mb), e no *link contrib* estão os vários pacotes e bibliotecas adicionais. Toda a instalação está muito bem documentada no texto *README.rw2011*. O procedimento básico é baixar o programa em

algum diretório temporário na sua máquina, dar um duplo clique no arquivo e seguir o procedimento usual de instalação de qualquer programa.

A biblioteca *survival* é uma das bibliotecas que estão pré-instaladas com o pacote básico. Outras bibliotecas podem ser instaladas, sempre que necessário. Por exemplo, a biblioteca *chron*, usada na preparação de bancos de dados com variável tempo-dependente (Apêndice ??), deve ser instalada. Na versão para Windows basta acessar o *menu Packages* na barra de ferramentas do R para fazer a instalação direto do CRAN (para isso você deve estar conectado a Internet).

Gostaríamos de chamar a atenção para alguns pontos em que usuário iniciante costuma errar:

- No R letras maiúsculas e minúsculas **são** diferentes;
- Para ler um arquivo, por exemplo, um banco de dados com o comando `read.table()`, o diretório *default* é onde o *workspace* do R está localizado. Para indicar outro diretório escreva o caminho completo com o separador de diretório / ou dupla barra invertida \\;
- Tudo o que você fez no R só será salvo após o comando `save.image()`, ou ao responder *yes* para salvar ao sair do R, ficando arquivado em um arquivo chamado *.Rdata*, no seu diretório de trabalho;
- No Windows, ao abrir o R clicando no ícone do *desktop*, o objeto carregado é o que está no diretório *default*, e não o que vinha sendo trabalhado;
- Deve-se mudar o diretório de trabalho usando a barra de ferramentas (*File > Change dir...*) e escolhendo seu diretório de trabalho e lendo o objeto *.Rdata* com o comando `Load Workspace`;
- Não use o mesmo diretório e o mesmo arquivo *.Rdata* para todos seus trabalhos, pois como todos os objetos ficam em memória, a *performance* do programa pode diminuir com objetos muito grandes;

- Para cada projeto seu, use um diretório diferente, onde ficam arquivados os arquivos do R (*.Rdata* e *.Rhistory*), seus *scripts*, seus bancos de dados, textos e figuras salvas;
- Uma boa forma de direcionar o R sempre para seu diretório de trabalho é acessar o programa clicando sobre o arquivo *.Rdata*, que está associado ao R.

Nossa sugestão mais enfática é que você se acostume a salvar todos os comandos usado em um pequeno arquivo texto. Contribui para lembrar o que foi feito e o que deu certo, e permite refazer rapidamente o trabalho em casos de *crash*, que são raríssimos em ambiente Linux, mas relativamente frequentes em ambiente Windows. O arquivo *.Rhistory* guarda uma longa lista de comandos recentes, mas arquiva também todos os comandos digitados erradamente. Este arquivo também é salvo conjuntamente com o *.Rdata*.

Uma biblioteca adicional que facilita a entrada de dados para o R é a *foreign*, que importa dados de algumas versões do SAS, SPSS, Stata, Systat, Epi-Info. Maiores informações no *help*.

Existem dezenas de manuais na *webpage* do R (clique nos *links* em *documentation*) e apostilas, algumas em português. Além disso, alguns livros-texto de estatística importantes utilizam o R. Uma referência das mais úteis é o livro *Modern Applied Statistics with S*, de Venables & Ripley, em sua quarta edição. Outra referência muito boa é o livro *An R Companion to Applied Regression*, de John Fox & Sanford Weisberg, segunda edição. A editora *Springer* publica uma série chamada *Use R!*, com dezenas de títulos, sobre variados tópicos estatísticos, sempre usando o R.

Esse livro foi elaborado na versão 2.13 (13/4/2011). As atualizações do R e das bibliotecas são muito frequentes. Caso algum comando deixe de funcionar, a causa pode ser a mudança de versão. Sempre que algum erro no livro ou nos exercícios for detectado, entre em contato com as autoras. Tentaremos manter uma versão dos exercícios, bancos de dados e cadernos de respostas sempre atualizada na página do livro: sobrevida.fiocruz.br.