

SOCIETY OF ACTUARIES

Article from:

CompAct

October 2008 – Issue No. 29



by Steven Craighead

¹ has become the lingua franca of the statistical world. Personally, I believe that it also has a lot to offer us actuaries. Most of our models can be constructed within Excel, but there are still areas where other modeling languages can be useful. R is a good candidate because:

- 1. It is open source.
- 2. It runs on multiple platforms.
- 3. It is free.
- 4. It has over 1000 various packages available.
- 5. It can easily be integrated into multiple packages including Excel.
- It is constantly improving with new statistical tools constantly being developed for R by researchers.

I am going to write a series of columns on R for the readers of CompAct and Expanding Horizons explaining how to use R as well as describing various packages available.

By the end of this article you should know how to locate, download and install R. I will also tell you how to start and end R and how to access the help features. The next article I will dedicate to how to get data into and out of R and how to access the data from a dataframe, which is the most common data structure within R. Other articles will follow outlining how to create modeling formulae and using R to model not just linear regression but other predictive models such as GLM and GAM.

Figure 1. The cran.r-project.org web page.						
R	The Comprehensive R Archive Network Frequently used pages					
CRAN Mirrors What's new? Task Views Search About R R Homepage Software R Sources R Binaries Packages Other Document at ion Manuals FAQs Contributed Newsletter	Download and Install R Precompiled binary distributions of the base system and contributed packages, Windows and Macusers most likely want one of these versions of R: • Linux • MacOS X • Windows Source Code for all Platforms Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it! • The latest release (2008-06-23): <u>R-2.7 1.tar.gz</u> (read <u>what's new</u> in the latest version). • Sources of <u>R alpha and beta releases</u> (daily snapshots, created only in time periods before a planned release). • Daily snapshots of current patched and development versions are <u>available here</u> . Please read about <u>new features and bug fixes</u> before filing corresponding feature requests or bug reports. • Source code of older versions of R is <u>available here</u> .					

R is located on "The Comprehensive R Archive Network," whose Web site is *http://cran.r-project.org/*. When you go to the site, the homepage displays the frequently used pages. See Figure 1. Note that the first set of choices is on how to download and install R from existing pre-compiled binaries for Linux, Mac and Windows. Choose your favorite operating system platform and you will go to a Web page that allows you to choose between base and contributed packages. Due to my limited access to testing other operating systems, in the remainder of this article I will only describe how to install the Windows binary. Notice in Figure 2, you will have the two choices of base and contrib. Initially, choose base and download the most current release. At the time of the writing of this article, version 2.7.1 is available.

Figure 2. The R for Windows Page					
R	R for Windows This directory contains binaries for a base distribution and packages to run on i336/x64 Windows.				
CRAN Mirrors	Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.				
What's new? Task Views Search	Subdirectories: base Binaries for base distribution (managed by Duncan Murdoch)				
About R R Homepage	contrib Binaries of contributed packages (managed by Uwe Ligges)				
Software	Please do not submit binaries to CKAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.				
R. Binaries Packages	You may also want to read the <u>K FAQ</u> and <u>K for Windows FAQ</u> .				
Documentation	Last moduled April 4, 2004, by Friedrich Leisch				
<u>FAQs</u> <u>Contributed</u> <u>Newsletter</u>					

Observe from Figure 3, the availability of the file R-2.7.1-win-32.exe (or whatever the latest version is). Download this file to some location on your hard disk. Next run the executable and begin the installation process.

(continued on page 26)

	Fig	ure 3. R-2.7.1 for Windows				
		R-2.7.1 for Windows				
PE	This directory contains a Windows binary distribution of R-2.7.1 (r45970).					
	Patches to this releas	e are incorporated in the r-patched snapshot build.				
CRAN <u>Mirrors</u> What's new?	A build of the development version (which will eventually become the next major release of R) is available in the <u>r-devel snapshot</u> build.					
Task Views Search	In this directory.					
About R	README.R-2.7.1	Installation and other instructions.				
R Homepage	CHANGES	New features of this Windows version.				
Software <u>R Sources</u> R Binaries	NEWS New features of all versions. R-2.7.1-win32 exe Setup program (about 29 megabytes). Please download this from a mirror near you. old Previous releases.					
Packages Other	md5sum.txt	md5sum output for the setup program. A Windows GUI version of md5sum is available at <u>http://www.md5summer.org/;</u> a Windows command line version is available at <u>http://www.etree.org/md5com.html</u> .				
Documentation						
<u>Manuals</u> <u>FAQs</u> Contributed	Please see the <u>R FAQ</u> for general information about R and the <u>R Windows FAQ</u> for Windows-specific information, including upgrade advice.					
Newsletter	Note to webmasters: A stable link which will redirect to the current Windows binary release is < <u>CRAN MIRCOR>/bin/windows/base/release.htm.</u>					

If you choose the default installation directory, R will be installed in C:\Program Files\ R\R-2.7.1. Note how the release is installed under a general R folder. I suggest that you make sure to keep your releases separate because of potential future conflicts with new releases.

The installer will then ask what support files you want to have installed with the base installation. I suggest that you should include the technical manuals as well as the reference manual. These will become handy as you expand your modeling skills. See Figure 4. I also suggest that you use the default startup options as well.

To run R from your Start button, choose the Programs tab and select the R selection and then choose the current release under that selection.

Figure 5 displays an empty R console that will then appear. R will suggest that you should run demo() for some examples, or help() for on-line help or help-start() for an html based help brower for R. It also states that to exit, you type q() and press return.





The demos() will give a list of various demonstrations available. You run these demonstrations by type "demo ("demoprogram") in R. For instance to look at the "plotmath" demo, type:

demo("plotmath")

You will see multiple displays of various mathematical formulas and symbols available, displayed in a separate graphics window. See Figure 6 for an example. The window may pause and you need to press enter to continue to the next example within the demo. Note the various commands that appear in the R environment.

ext					
• @ -					
R Console	Click or hit ENTER for next	page			
<pre>draw.plotmath.cell(expression(x t==t y), i, nr); i <</pre>	1 states and a				
draw.nlotmath.cell(expression/x %pron% v), i, pr); i	Anthmetic C	Anthmetic Operators		Radicals	
	X + Y	x+y	sqrf(x)	√X.	
draw.title.cell("Typeface", i, nr); i <- i + 1	X - Y	x-y	sqrt(x, y)	¥x	
<pre>draw.plotmath.cell(expression(plain(x)), i, nr); i <</pre>	X*Y	XY 	Relati	ons	
draw.plotmath.cell(expression(italic(x)), i, nr); i	XVY	x/y	х у	x= y	
	X %+-% Y	x±y	x 1= y	X≠ y	
<pre>draw.plotmath.cell(expression(bold(x)), i, nr); i <-</pre>	X%0/%0Y	x÷ y	x < y	x <y< td=""></y<>	
<pre>draw.plotmath.cell(expression(bolditalic(x)), i, nr)</pre>)(%°*% Y	x×y	χ <= γ	x≤y	
draw wlotwath call (avaragaion (underline (v)) i mr);	X % % Y	х.у	x>y	x>	
draw.prochach.ceri(expression(underine(x)), i, ht);	-X	-x	$X \ge = \Lambda$	x≥y	
# Need fever, wider columns for ellipsis	+X	+X	ж%~~%у	х≈у	
hr <-20	Sub/Super	Sub/Superscripts		X≅Y	
nc <- 2	54[1]	X,	x %==% y	X=Y	
make.table(nr, nc)	x*2	x2	x %prop% y	xœy	
aiting to confirm page change	Juxtapos	Juxtaposition		ace	
	X* Y	ху	plain(x)	х	
	paste(x, y, z)	xyz	italic(x)	x	
	Lists		bold(x)	x	
	list(x, y, z)	x, y, z	bolditalic(x)	x	
	L. CONTRACTOR	200000000	underlineba	x	

To use help, either type help(command) or ?command in the R environment. For example, to find out help on the linear regression method, you can type:

Figure 7 displays an example of the Web browser that pops up with a description of Fitting Linear Models. If you page down through various help articles, various R commands will be displayed, which you can copy and paste into R.

(continued on page 28)



Steven Craighead, ASA, MAAA, CERA, is a consultant at Towers Perrin. He can be contacted at steven.craighead@ towersperrin.com.

[?]lm or help(lm)



Also, in the help browser you are able to index and search for various other objects or methods available within the base system.

Another helpful help command is "help.search("phrase"), where in the R environment you can look for various packages and methods and objects that contain the phrase that you enter.

For example, if you enter: help.search("phrase")

Look at Figure 8, for the R information screen that results from the help.search command.

Figure 8. help.search results					
R Information					
Help files with alias matching:	or concept or title matching 'phrase' using fuzzy				
plot.spec(stats) RSiteSearch(utils)	Plotting Spectral Densities Search for Key Words or Phrases in the R-help Mailing List Archives or Documentation				
Type 'help(FOO, packag	ge = PKG)' to inspect entry 'FOO(PKG) TITLE'.				

Figure 8 is the R information screen that results from the help.search command.

Observe from Figure 8 that if you load the stats package and invoke help on the plot. spec function, you will find some use of the word "phrase."

By the way, you load packages by choosing the Load Package option off of the main menu Package choice. See Figure 9.

If you want to specify the subdirectory in which you want to work in Windows, choose the Change Directory option off of the File option of the main menu. From the main menu, you can observe how to save and load workspaces and history files. The workspace will contain your data and various functions and objects that you build, but the history files will store the command you enter in the R environment. As you develop your own models you will need to exploit these features.

One final thing, if you want to stop R from running some long process, press the Stop Sign Icon at the top of the environment.

Remember to type q() to exit.

I'm going to stop for now. Next column, we will look at various ways to get data into and out of R. \blacksquare

References:

 R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

