

Using R for Regression Course —

ACSPRI 2012 Winter Program,

UTS, Sydney, July 2-6 2012

<http://www.acspri.org.au/winterprogram2012>

Tentative Schedule

John Maindonald

The discussion will in all cases be based around specific examples. The intention is to flesh out the tentative schedule in more detail closer to the time of the course.

Monday

- Check R installations
- Overview of course content, and some important preliminaries
- What is regression?
- Chapter 1 examples: broach statistical issues and use R graphical abilities
- Review of R basics, using both the R Commander GUI and the command line
- The Data Frame — commonly used with small to medium sized data sets
- Straight line regression – there's more to it than you might think

Tuesday

- Review of Monday
- Regression calculation preliminaries — exploratory data analysis
- Statistical issues – data generation and generalization from regression results
 - Distributions, and graphical display of distributions, with columns from datasets that will be used later in the course.
 - Populations and samples
 - The fixed and the random part of models
- Regression in R — model formulae, model matrix, and interpretation of R output.
- Leverage and why it matters
- When and why to use transformations
- Practice with multiple regression

Wednesday

- Simulation – models as recipes for creating data generation machines!
- Diagnostics – how credible is it that model A generated data B?
- Plots that summarize regression results
- Identification of non-linear relationships, and ways to handle them
- Using models to make predictions, as against interpretation of model parameters
- Multicollinearity and near-multicollinearity, and how to handle it
- The limitations of R^2 as a performance measure, and alternatives
- Training set, validation set, and test set
- Test set, cross-validation, and bootstrap accuracy estimates

Thursday

- Variable selection methods – cautions against unthinking use, traps, and how to get realistic standard errors.

- Several examples will be worked through in detail, illustrating, using and testing knowledge of the ideas that have been covered earlier.

Friday

- Regression with binary data, binomial data, and count data
- Review of course
- Where next – what the course did not cover!

ACSPRI Course – Modern Regression, Classification and Multivariate Exploration, September 26-30 2011.

Installation of R Using the R Installation Binary R-2.13.1patched-win.exe

Go to the directory **win-binaries**, click on **R-2.13.1patched-win.exe**, and follow instructions. If in doubt, accept the defaults. Alternatively go to a CRAN site (in Australia, use

<http://cran.ms.unimelb.edu.au/> or <http://cran.csiro.au/>) download **R-2.13.1patched-win.exe** which has any more recent patches, and install that.

A limited number of packages, those that are on the recommended list, will be installed as part of the initial installation. The DVD packages directory should have other packages that may be needed for the course.

To install packages from the DVD, start R, click on the Packages menu, then on Install package(s) from local zip files..., then navigate to the relevant Packages directory (packages-2.13) on the DVD. Select some or all of the packages on the DVD for installation. To update packages from the internet (more recent versions may in some cases be available), go to the Packages menu and select Update packages...

For updating or installing packages from the internet, ensure a live internet connection! Australian users should use an Australian CRAN mirror.

Several packages require Gtk2 in order to run. For 32-bit R-2.12.0 or later, install Gtk2 from the executable **gtk2-runtime-2.22.0-2010-10-21-ash.exe**. Change the install directory so that the path does not have spaces; use for example C:\gtkwin32.

The R package rggobi provides an interface to the dynamic graphics package Ggobi. Install Ggobi from **ggobi-2.1.8.exe**.

Adding an R Icon

A convenient way to add a new R icon is to start by copying an existing icon. Then right click on the icon, click on Properties, and set the target (click on New | Target) to the directory that is wanted as the working directory. Change the name of the shortcut as required.

Accessing packages from the DVD:

It is sometimes useful to allow another copy of R, perhaps the main R installation on the computer, access to packages from the **R-2.13.1patched** (or other) installation on the DVD. If the DVD is in drive D:, enter, from the R command line:

```
.libPaths("D:/R-2.13.1pat/library")
```

For execution whenever a session is started in a working directory, create a `.First()` function that includes this statement, for inclusion in the workspace image that is saved at the end of the session.

Note - Use of the Installation Tree on the DVD to Run R:

Running R from a DVD, or copying a directory tree from a DVD onto a hard or other drive and running R from there, can be a quick way to get started. It makes it possible to run R without doing a full installation onto the computer. The downside is that there is no configuration for the specific computer. Either run R directly from the DVD, or copy the complete R-2.13.1pat directory tree to another device such as the hard drive, or a thumb drive. Wherever the directory tree may be located, the choices are then: Either (for 64-bit R):

- click on **R-2.13.1pat\bin\x64\Rgui**, or
- give the full path (e.g., **D:\R-2.13.1pat\bin\i386\Rgui.exe**), as the target for an R icon.