



SHORT COURSE INFORMATION

Training course in Advanced Statistics using R

Mercure Grosvenor Hotel, Adelaide

5-7 May 2009

Introduction

R is a language and environment for statistical computing and graphics. It provides a wide variety of utilities for summarisation, graphing and analysis of data (including a wide range of statistical methods). Even more importantly, R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License and runs on a wide variety of systems.

The down-side of R is that it takes some time and effort to become familiar with how it works and with the various R commands and syntax. However, once users are familiar with the R environment and data structures, the system is relatively easy and quick to use and is extremely powerful.

This course is aimed at helping R users become confident and proficient in using R for some of the more sophisticated statistical analyses, particularly multiple regression analyses, linear models and mixed effect models.

About the Course

The course is targeted at scientists and students who need tools for analysis of complex datasets. The course will be of particular value to students undertaking postgraduate degrees, as well as institutional researchers. The course assumes some familiarity with R, equivalent to the introductory course being run as part of this series. It also assumes an understanding of the principles and application of multivariable statistical analyses and statistical modelling. This is not a statistics course – rather it is a course about how to use R to undertake a range of statistical analyses.

Computers

Participants will be required to bring their own laptop computer to use during the course. Anyone wishing to install R before the course can download version 2.7.1 (the version we will be using) from <http://cran.r-project.org/> (follow the links to windows, base and old to find the windows version 2.7.1). Participants can also

download the Tinn-R text editor which will be used during the workshop from <http://sourceforge.net/projects/tinn-r>.

Participants will also be provided with a memory stick containing all the required software.

Course Outline

Topics covered in the course include:

- Introduction
- Review of R functionality and approach
- Linear Regression
- Logistic/Poisson regression
- Survival analysis
- Multi-level/mixed effects models

Training Approach

The course will be based heavily around the practical use of R through a wide range of hands-on exercises. Software and a series of training data sets will be provided. Each new topic will be briefly introduced by the lead trainers, and then participants will undertake a number of exercises. Participants are also encouraged to bring their own datasets and to apply techniques and functions learned to these.

Price

The cost of the course will be free to AB-CRC students, \$726 (inc. GST) for other students, \$825 (inc. GST) for AB-CRC members and \$990 (inc. GST) for non-members, which will cover tuition, course materials, lunch and morning and afternoon tea each day. Participants are responsible for the cost of their travel, accommodation and other meals.

Course Coordinator

Dr Evan Sergeant
AusVet Animal Health Services
59 Turner Crescent, Orange, NSW, 2800, Australia
Phone: (+61) 2 6362 1598
Email: evan@ausvet.com.au

Contact Details

For registration and administrative information, please contact:

Debra Gendle
Australian Biosecurity CRC for Emerging Infectious Disease
Ph: 08 9266 1634
Fax: 08 9266 1650
Email: debra.gendle@abcrc.org.au

About R

[Based on the description of R at <http://www.r-project.org>]

R is a language and environment for statistical computing and graphics. It can be considered as a different implementation of the S language and environment. It provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, and others) and graphical techniques, and is highly extensible. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

The R environment

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The term 'environment' is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

R, like S, is designed around a true computer language, and it allows users to add additional functionality by defining new functions. Much of the system is itself written in the R dialect of S, which makes it easy for users to follow the algorithmic choices made. For computationally-intensive tasks, C, C++ and Fortran code can be linked and called at run time. Advanced users can write C code to manipulate R objects directly.

Many users think of R as a statistics system. We prefer to think of it as an environment within which statistical techniques are implemented. R can be extended (easily) via *packages*. There are a number of packages supplied with the R distribution and anyone can produce add-on packages for particular purposes. As a result many packages are available through the CRAN family of Internet sites covering a very wide range of modern statistics.