**Introduction to R Programming (for SEC)**

Introduction

**R** is a computer language not entirely unlike the S language developed at AT&T Bell Laboratories by Rick Becker, John Chambers and Allan Wilks. The two languages are implemented quite differently, but bear enough superficial resemblance that users should be able to switch between the two with relative ease. Currently the software is undergoing active development. Discussion of the development process is carried out on the "r-devel" mailing list. See the [resources page](http://127.0.0.1:18627/help/doc/html/resources.html) for details on how to subscribe to this list.

We have implemented **R** in what we hope is a very portable fashion and in way which requires relatively little in the way of machine resources. Implementations exist for many for many members of the Unix family of operating systems, including **AIX**, **FreeBSD**, **GNU/Linux**, **HPUX**, **Irix**,**macOS**, **Solaris**, and **Tru64**. In addition there is a version for **Microsoft Windows** (9x, ME, NT4, 2000, XP).

Present Status

The present version implements most of the functionality in the 1988 S book (the "Blue Book") and many of the applications. In addition, we have implemented much of the functionality from the 1992 S book (the "White Book"). In particular we have versions of "lm", "glm", "aov" and "loess", and versions of "gam" and "tree" are available in contributed packages. There are several manuals in the distribution, plus a comprehensive set of help pages in "output independent" form which can be used to create versions for HTML, LaTeX, text, PDF etc.

**Introduction and preliminaries**

### 1.1 The R environment

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. Among other things it has

* 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 directly at the computer or on hardcopy, and
* a well developed, simple and effective programming language (called ‘S’) which includes conditionals, loops, user defined recursive functions and input and output facilities. (Indeed most of the system supplied functions are themselves written in the S language.)

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 is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R are essentially ephemeral, written for a single piece of data analysis.

### 1.2 Related software and documentation

R can be regarded as an implementation of the S language which was developed at Bell Laboratories by Rick Becker, John Chambers and Allan Wilks, and also forms the basis of the S-PLUS systems.

The evolution of the S language is characterized by four books by John Chambers and coauthors. For R, the basic reference is The New S Language: A Programming Environment for Data Analysis and Graphics by Richard A. Becker, John M. Chambers and Allan R. Wilks. The new features of the 1991 release of S are covered in Statistical Models in S edited by John M. Chambers and Trevor J. Hastie. The formal methods and classes of the **methods** package are based on those described in Programming with Data by John M. Chambers. See [References](http://127.0.0.1:18627/help/doc/manual/R-intro.html#References), for precise references.

There are now a number of books which describe how to use R for data analysis and statistics, and documentation for S/S-PLUScan typically be used with R, keeping the differences between the S implementations in mind. See [What documentation exists for R?](http://127.0.0.1:18627/help/doc/manual/R-FAQ.html#What-documentation-exists-for-R_003f) in The R statistical system FAQ.

&&&