The only drawback of the book follows directly from the target audience of possible S novices: More experienced S users will probably be disappointed by the ratio of S introduction to material on regression analysis. Only about 130 out of 300 pages deal directly with linear models, the major part is an introduction to S. The online appendix at the book's homepage (mirrored on CRAN) contains several extension chapters on more advanced topics like bootstrapping, time series, nonlinear or robust regression.

In summary, I highly recommend the book to anyone who wants to learn or teach applied regression analysis with S.

Friedrich Leisch
Universität Wien, Austria
Friedrich.Leisch@R-project.org

Changes in R 1.7.0

by the R Core Team

User-visible changes

- `solve()`, `chol()`, `eigen()` and `svd()` now use LAPACK routines unless a new back-compatibility option is turned on. The signs and normalization of eigen/singular vectors may change from earlier versions.
- The 'methods', 'modreg', 'mva', 'nls' and 'ts' packages are now attached by default at startup (in addition to 'ctest'). The option `"defaultPackages"` has been added which contains the initial list of packages. See `?Startup` and `?options` for details. Note that `.First()` is no longer used by R itself.

  `class()` now always (not just when 'methods' is attached) gives a non-null class, and `UseMethod()` always dispatches on the class that `class()` returns. This means that methods like foo.matrix and foo.integer will be used. Functions `oldClass()` and `oldClass<-()` get and set the "class" attribute as R without 'methods' used to.
- The default random number generators have been changed to 'Mersenne-Twister' and 'Inversion'. A new `RNGversion()` function allows you to restore the generators of an earlier R version if reproducibility is required.
- Namespaces can now be defined for packages other than 'base': see 'Writing R Extensions'. This hides some internal objects and changes the search path from objects in a namespace. All the base packages (except methods and tcltk) have namespaces, as well as the recommended packages 'KernSmooth', 'MASS', 'boot', 'class', 'nnet', 'rpart' and 'spatial'.
- Formulae are not longer automatically simplified when `terms()` is called, so the formulae in results may still be in the original form rather than the equivalent simplified form (which may have reordered the terms): the results are now much closer to those of S.
- The tables for plotmath, Hershey and Japanese have been moved from the help pages (`example(plotmath)` etc) to `demo(plotmath)` etc.
- Errors and warnings are sent to stderr not stdout on command-line versions of R (Unix and Windows).
- The R_X11 module is no longer loaded until it is needed, so do test that `x11()` works in a new Unix-alike R installation.

New features

- `if()` and `while()` give a warning if called with a vector condition.
- Installed packages under Unix without compiled code are no longer stamped with the platform and can be copied to other Unix-alike platforms (but not to other OSes because of potential problems with line endings and OS-specific help files).
- The internal random number generators will now never return values of 0 or 1 for `runif()`. This might affect simulation output in extremely rare cases. Note that this is not guaranteed for user-supplied random-number generators, nor when the standalone Rmath library is used.
- When assigning names to a vector, a value that is too short is padded by character NAs. (Wishlist part of PR#2358)
- It is now recommended to use the 'SystemRequirements:' field in the DESCRIPTION file for specifying dependencies external to the R system.
- Output text connections no longer have a line-length limit.

- On platforms where vsnprintf does not return the needed buffer size the output line-length limit for `fifo()`, `gzfile()` and `bzfile()` has been raised from 10k to 100k chars.
- The Math group generic does not check the number of arguments supplied before dispatch: it used to if the default method had one argument but not if it had two. This allows `trunc.POSIXt()` to be called via the group generic `trunc()`.
- Logical matrix replacement indexing of data frames is now implemented (interpreted as if the lhs was a matrix).
- Recursive indexing of lists is allowed, so `x[[c(4,2)]]` is shorthand for `x[[4]][[2]]` etc. (Wishlist PR#1588)
- Most of the time series functions now check explicitly for a numeric time series, rather than fail at a later stage.
- The postscript output makes use of relative moves, and so is somewhat more compact.
- `%*%` and `crossprod()` for complex arguments make use of BLAS routines and so may be much faster on some platforms.
- `arima()` has `coef()`, `logLik()` (and hence AIC) and `vcov()` methods.
- New function `as.difftime()` for time-interval data.
- `basename()` and `dirname()` are now vectorized.
- `biplot.default()` mva allows 'xlab' and 'ylab' parameters to be set (without partially matching to 'xlabs' and 'ylabs'). (Thanks to Uwe Ligges.)
- New function `capture.output()` to send printed output from an expression to a connection or a text string.
- `ccf()` (pckage ts) now coerces its `x` and `y` arguments to class `"ts"`.
- `chol()` and `chol2inv()` now use LAPACK routines by default.
- `as.dist(.)` is now idempotent, i.e., works for `"dist"` objects.
- Generic function `confint()` and 'lm' method (formerly in package MASS, which has 'glm' and 'nls' methods).
- New function `constrOptim()` for optimisation under linear inequality constraints.
- Add 'difftime' subscript method and methods for the group generics. (Thereby fixing PR#2345)
- `download.file()` can now use HTTP proxies which require 'basic' username/password authentication.
- `dump()` has a new argument 'envir'. The search for named objects now starts by default in the environment from which `dump()` is called.
- The `edit.matrix()` and `edit.data.frame()` editors can now handle logical data.
- New argument 'local' for `example()` (suggested by Andy Liaw).
- New function `file.symlink()` to create symbolic file links where supported by the OS.
- New generic function `flush()` with a method to flush connections.
- New function `force()` to force evaluation of a formal argument.
- New functions `getFromNamespace()`, `fixInNamespace()` and `getS3method()` to facilitate developing code in packages with namespaces.
- `glm()` now accepts 'etastart' and 'mustart' as alternative ways to express starting values.
- New function `gzcon()` which wraps a connection and provides (de)compression compatible with gzip.

  `load()` now uses `gzcon()`, so can read compressed saves from suitable connections.
- `help.search()` can now reliably match individual aliases and keywords, provided that all packages searched were installed using R 1.7.0 or newer.
- `hist.default()` now returns the nominal break points, not those adjusted for numerical tolerances.

  To guard against unthinking use, 'include.lowest' in `hist.default()` is now ignored, with a warning, unless 'breaks' is a vector. (It either generated an error or had no effect, depending how prettification of the range operated.)
- New generic functions `influence()`, `hatvalues()` and `dfbeta()` with lm and glm methods; the previously normal functions `rstudent()`, `rstandard()`, `cooks.distance()` and `dfbetas()` became generic. These have changed behavior for glm objects – all originating from John Fox' car package.

- `interaction.plot()` has several new arguments, and the legend is not clipped anymore by default. It internally uses `axis(1,*)` instead of `mtext()`. This also addresses "bugs" PR#820, PR#1305, PR#1899.
- New `isoreg()` function and class for isotonic regression ('modreg' package).
- `La.chol()` and `La.chol2inv()` now give interpretable error messages rather than LAPACK error codes.
- `legend()` has a new 'plot' argument. Setting it 'FALSE' gives size information without plotting (suggested by U.Ligges).
- `library()` was changed so that when the methods package is attached it no longer complains about formal generic functions not specific to the library.
- `list.files()/dir()` have a new argument 'recursive'.
- `lm.influence()` has a new 'do.coef' argument allowing *not* to compute casewise changed coefficients. This makes `plot.lm()` much quicker for large data sets.
- `load()` now returns invisibly a character vector of the names of the objects which were restored.
- New convenience function `loadURL()` to allow loading data files from URLs (requested by Frank Harrell).
- New function `mapply()`, a multivariate `lapply()`.
- New function `md5sum()` in package tools to calculate MD5 checksums on files (e.g. on parts of the R installation).
- `medpolish()` package eda now has an 'na.rm' argument (PR#2298).
- `methods()` now looks for registered methods in namespaces, and knows about many objects that look like methods but are not.
- `mosaicplot()` has a new default for 'main', and supports the 'las' argument (contributed by Uwe Ligges and Wolfram Fischer).
- An attempt to `open()` an already open connection will be detected and ignored with a warning. This avoids improperly closing some types of connections if they are opened repeatedly.
- `optim(method = "SANN")` can now cover combinatorial optimization by supplying a move function as the 'gr' argument (contributed by Adrian Trapletti).
- PDF files produced by `pdf()` have more extensive information fields, including the version of R that produced them.
- On Unix(-alike) systems the default PDF viewer is now determined during configuration, and available as the 'pdfviewer' option.
- `pie(...)` has always accepted graphical pars but only passed them on to `title()`. Now `pie(, cex=1.5)` works.
- plot.dendrogram ('mva' package) now draws leaf labels if present by default.
- New `plot.design()` function as in S.
- The `postscript()` and `PDF()` drivers now allow the title to be set.
- New function `power.anova.test()`, contributed by Claus Ekstrøm.
- `power.t.test()` now behaves correctly for negative delta in the two-tailed case.
- `power.t.test()` and `power.prop.test()` now have a 'strict' argument that includes rejections in the "wrong tail" in the power calculation. (Based in part on code suggested by Ulrich Halekoh.)
- `prcomp()` is now fast for nm inputs with $m >> n$.
- `princomp()` no longer allows the use of more variables than units: use `prcomp()` instead.
- `princomp.formula()` now has principal argument 'formula', so `update()` can be used.
- Printing an object with attributes now dispatches on the `class(es)` of the attributes. See `?print.default` for the fine print. (PR#2506)
- `print.matrix()` and `prmatrix()` are now separate functions. `prmatrix()` is the old S-compatible function, and `print.matrix()` is a proper print method, currently identical to `print.default()`. `prmatrix()` and the old `print.matrix()` did not print attributes of a matrix, but the new `print.matrix()` does.
- `print.summary.lm()` and `print.summary.glm()` now default to `symbolic.cor = FALSE`, but `symbolic.cor` can be passed to the print methods from the summary methods. `print.summary.lm()` and `print.summary.glm()` print correlations to 2 decimal places, and the symbolic printout avoids abbreviating labels.

- If a `prompt()` method is called with 'filename' as 'NA', a list-style representation of the documentation shell generated is returned. New function `promptData()` for documenting objects as data sets.
- `qqnorm()` and `qqline()` have an optional logical argument 'datax' to transpose the plot (S-PLUS compatibility).
- `qr()` now has the option to use LAPACK routines, and the results can be used by the helper routines `qr.coef()`, `qr.qy()` and `qr.qty()`. The LAPACK-using versions may be much faster for large matrices (using an optimized BLAS) but are less flexible.
- QR objects now have class "qr", and `solve.qr()` is now just the method for `solve()` for the class.
- New function `r2dtable()` for generating random samples of two-way tables with given marginals using Patefield's algorithm.
- `rchisq()` now has a non-centrality parameter 'ncp', and there's a C API for `rnchisq()`.
- New generic function `reorder()` with a dendrogram method; new `order.dendrogram()` and `heatmap()`.
- `require()` has a new argument named `character.only` to make it align with library.
- New functions `rmultinom()` and `dmultinom()`, the first one with a C API.
- New function `runmed()` for fast runnning medians ('modreg' package).
- New function `slice.index()` for identifying indexes with respect to slices of an array.
- `solve.default(a)` now gives the dimnames one would expect.
- `stepfun()` has a new 'right' argument for right-continuous step function construction.
- `str()` now shows ordered factors different from unordered ones. It also differentiates "NA" and `as.character(NA)`, also for factor levels.
- `symnum()` has a new logical argument 'abbr.colnames'.
- `summary(<logical>)` now mentions NA's as suggested by Göran Broström.
- `summaryRprof()` now prints times with a precision appropriate to the sampling interval, rather than always to 2dp.
- New function `Sys.getpid()` to get the process ID of the R session.
- `table()` now allows exclude= with factor arguments (requested by Michael Friendly).
- The `tempfile()` function now takes an optional second argument giving the directory name.
- The ordering of terms for

```
terms.formula(keep.order=FALSE)
```

  is now defined on the help page and used consistently, so that repeated calls will not alter the ordering (which is why `delete.response()` was failing: see the bug fixes). The formula is not simplified unless the new argument 'simplify' is true.
- added "[" method for terms objects.
- New argument 'silent' to `try()`.
- `ts()` now allows arbitrary values for `y` in `start/end = c(x, y)`: it always allowed `y < 1` but objected to `y > frequency`.
- `unique.default()` now works for POSIXct objects, and hence so does `factor()`.
- Package tcltk now allows return values from the R side to the Tcl side in callbacks and the R_eval command. If the return value from the R function or expression is of class "tclObj" then it will be returned to Tcl.
- A new HIGHLY EXPERIMENTAL graphical user interface using the tcltk package is provided. Currently, little more than a proof of concept. It can be started by calling "R -g Tk" (this may change in later versions) or by evaluating `tkStartGUI()`. Only Unix-like systems for now. It is not too stable at this point; in particular, signal handling is not working properly.
- Changes to support name spaces:
  - Placing base in a name space can no longer be disabled by defining the environment variable R_NO_BASE_NAMESPACE.
  - New function `topenv()` to determine the nearest top level environment (usually .GlobalEnv or a name space environment).
  - Added name space support for packages that do not use methods.

- Formal classes and methods can be 'sealed', by using the corresponding argument to setClass or setMethod. New functions `isSealedClass()` and `isSealedMethod()` test sealing.
- packages can now be loaded with version numbers. This allows for multiple versions of files to be installed (and potentially loaded). Some serious testing will be going on, but it should have no effect unless specifically asked for.

Installation changes

- TITLE files in packages are no longer used, the Title field in the DESCRIPTION file being preferred. TITLE files will be ignored in both installed packages and source packages.
- When searching for a Fortran 77 compiler, configure by default now also looks for Fujitsu's frt and Compaq's fort, but no longer for cf77 and cft77.
- Configure checks that mixed C/Fortran code can be run before checking compatibility on ints and doubles: the latter test was sometimes failing because the Fortran libraries were not found.
- PCRE and bzip2 are built from versions in the R sources if the appropriate library is not found.
- New configure option '`--with-lapack`' to allow high-performance LAPACK libraries to be used: a generic LAPACK library will be used if found. This option is not the default.
- New configure options '`--with-libpng`', '`--with-jpeglib`', '`--with-zlib`', '`--with-bzlib`' and '`--with-pcre`', principally to allow these libraries to be avoided if they are unsuitable.
- If the precious variable R_BROWSER is set at configure time it overrides the automatic selection of the default browser. It should be set to the full path unless the browser appears at different locations on different client machines.
- Perl requirements are down again to 5.004 or newer.
- Autoconf 2.57 or later is required to build the configure script.
- Configure provides a more comprehensive summary of its results.
- Index generation now happens when installing source packages using R code in package tools. An existing 'INDEX' file is used as is; otherwise, it is automatically generated from the `\name` and `\title` entries in the Rd files. Data, demo and vignette indices are computed from all available files of the respective kind, and the corresponding index information (in the Rd files, the 'demo/00Index' file, and the `\VignetteIndexEntry{}` entries, respectively). These index files, as well as the package Rd contents data base, are serialized as R objects in the 'Meta' subdirectory of the top-level package directory, allowing for faster and more reliable index-based computations (e.g., in `help.search()`).
- The Rd contents data base is now computed when installing source packages using R code in package tools. The information is represented as a data frame without collapsing the aliases and keywords, and serialized as an R object. (The 'CONTENTS' file in Debian Control Format is still written, as it is used by the HTML search engine.)
- A NAMESPACE file in root directory of a source package is copied to the root of the package installation directory. Attempting to install a package with a NAMESPACE file using '`--save`' signals an error; this is a temporary measure.

Deprecated & defunct

- The assignment operator '_' will be removed in the next release and users are now warned on every usage: you may even see multiple warnings for each usage.

  If environment variable R_NO_UNDERLINE is set to anything of positive length then use of '_' becomes a syntax error.
- `machine()`, `Machine()` and `Platform()` are defunct.
- `restart()` is defunct. Use `try()`, as has long been recommended.
- The deprecated arguments 'pkg' and 'lib' of `system.file()` have been removed.
- `printNoClass()` methods is deprecated (and moved to base, since it was a copy of a base function).
- Primitives `dataClass()` and `objWithClass()` have been replaced by `class()` and `class<-()`; they were internal support functions for use by package methods.
- The use of SIGUSR2 to quit a running R process under Unix is deprecated, the signal may need to be reclaimed for other purposes.

Utilities

- R CMD check more compactly displays the tests of DESCRIPTION meta-information. It now reports demos and vignettes without available index information. Unless installation tests are skipped, checking is aborted if the package dependencies cannot be resolved at run time. Rd files are now also explicitly checked for empty `\name` and `\title` entries. The examples are always run with T and F redefined to give an error if used instead of TRUE and FALSE.

- The Perl code to build help now removes an existing example file if there are no examples in the current help file.

- R CMD Rdindex is now deprecated in favor of function `Rdindex()` in package tools.

- `Sweave()` now encloses the Sinput and Soutput environments of each chunk in an Schunk environment. This allows to fix some vertical spacing problems when using the latex class slides.

C-level facilities

- A full double-precision LAPACK shared library is made available as -lRlapack. To use this include `$(LAPACK_LIBS) $(BLAS_LIBS)` in PKG_LIBS.

- Header file R_ext/Lapack.h added. C declarations of BLAS routines moved to R_ext/BLAS.h and included in R_ext/Applic.h and R_ext/Linpack.h for backward compatibility.

- R will automatically call initialization and unload routines, if present, in shared libraries/DLLs during `dyn.load()` and `dyn.unload()` calls. The routines are named R_init_<dll name> and R_unload_<dll name>, respectively. See the Writing R Extensions Manual for more information.

- Routines exported directly from the R executable for use with `.C()`, `.Call()`, `.Fortran()` and `.External()` are now accessed via the registration mechanism (optionally) used by packages. The ROUTINES file (in src/appl/) and associated scripts to generate FFTab.h and FFDecl.h are no longer used.

- Entry point Rf_append is no longer in the installed headers (but is still available). It is apparently unused.

- Many conflicts between other headers and R's can be avoided by defining STRICT_R_HEADERS and/or R_NO_REMAP – see 'Writing R Extensions' for details.

- New entry point R_GetX11Image and formerly undocumented ptr_R_GetX11Image are in new header R_ext/GetX11Image. These are used by package tkrplot.

Changes on CRAN

by Kurt Hornik and Friedrich Leisch

New contributed packages

Davies useful functions for the Davies quantile function and the Generalized Lambda distribution. By Robin Hankin.

GRASS Interface between GRASS 5.0 geographical information system and R, based on starting R from within the GRASS environment using values of environment variables set in the GISRC file. Interface examples should be run outside GRASS, others may be run within. Wrapper and helper functions are provided for a range of R functions to match the interface metadata structures. Interface functions by Roger Bivand, wrapper and helper functions modified from various originals by interface author.

MCMCpack This package contains functions for posterior simulation for a number of statistical models. All simulation is done in compiled C++ written in the Scythe Statistical Library Version 0.3. All models return coda mcmc objects that can then be summarized using coda functions or the coda menu interface. The package also contains some useful utility functions, including some additional PDFs and pseudo-random number generators for statistical distributions. By Andrew D. Martin, and Kevin M. Quinn.

RSvgDevice A graphics device for R that uses the new w3.org xml standard for Scalable Vector Graphics. By T Jake Luciani.

SenSrivastava Collection of datasets from Sen & Srivastava: Regression Analysis, Theory, Methods and Applications, Springer. Sources for individual data files are more fully documented in