

Estadística
Básica
con
R y R-Commander
(Versión Febrero 2008)

Autores:

A. J. Arriaza Gómez
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
S. Pérez Plaza
A. Sánchez Navas



Universidad
de Cádiz

Servicio de Publicaciones

Copyright ©2008 Universidad de Cádiz. Se concede permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre de GNU, Versión 1.2 o cualquier otra versión posterior publicada por la Free Software Foundation. Una traducción de la licencia está incluida en la sección titulada "Licencia de Documentación Libre de GNU".

Copyright ©2008 Universidad de Cádiz. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation. A copy of the license is included in the section entitled "GNU Free Documentation License".

Edita: Servicio de Publicaciones de la Universidad de Cádiz
C/ Dr. Marañón, 3
11002 Cádiz

<http://www.uca.es/publicaciones>

ISBN:

Depósito legal:

Estadística Básica con R y R-commander
(Versión Febrero 2008)
*Autores: A. J. Arriaza Gómez, F. Fernández Palacín,
M. A. López Sánchez, M. Muñoz Márquez, S. Pérez Plaza,
A. Sánchez Navas*
©2008 Servicio de Publicaciones de la Universidad de Cádiz
<http://knuth.uca.es/ebrcmdr>

Capítulo 1

Comenzando con R

1. Introducción

El que un libro que pretende incidir sobre los aspectos prácticos de la Estadística, comience con un capítulo dedicado al software, no debería sorprender, aun cuando en el Prólogo se haya dejado claro que no es un objetivo fundamental enseñar a manejar un programa informático. De hecho, este manual seguiría teniendo utilidad aun cuando se usara otra interfaz gráfica distinta a la que se propone o, incluso, otro software; bastaría en ese caso con acomodar los menús y/o la sintaxis. No obstante, el que existan varias soluciones informáticas, no quiere decir que optar por una de ellas no tenga un interés determinante y, por tanto, deben emplearse para su elección criterios objetivos de eficiencia, no solo de carácter estadístico, sino que atiendan también a su facilidad de uso.

Para la elección de **R** se han evaluado pues distintos aspectos, siendo especialmente destacables sus bondades en lo que se refiere a calidad, a la cantidad de técnicas y funciones implementadas, a que es libre y a la gran comunidad científica que lo usa como estándar para el análisis de datos. Dicha comunidad ha desarrollado y desarrolla herramientas integradas en paquetes—en la actualidad más de 800—, que dan solución a una gran variedad de problemas estadísticos.

2 Capítulo 1. Comenzando con R

R es un lenguaje de programación y un entorno para análisis estadístico y la realización de gráficos. Debido a su naturaleza es fácilmente adaptable a una gran variedad de tareas. Fue inicialmente escrito por Robert Gentleman y Ross Ihaka del *Departamento de Estadística* de la *Universidad de Auckland* en Nueva Zelanda. **R** actualmente es el resultado de un esfuerzo de colaboración de personas del todo el mundo. Desde mediados de 1997 se formó lo que se conoce como núcleo de desarrollo de **R**, que actualmente es el que tiene la posibilidad de modificación directa del código fuente. Por otra parte, **R** es un proyecto GNU similar a S, desarrollado éste por los *Laboratorios Bell*. Las diferencias entre **R** y S son importantes, pero la mayoría del código escrito para S corre bajo **R** sin modificaciones.

R abarca una amplia gama de técnicas estadísticas que van desde los modelos lineales a las más modernas técnicas de clasificación pasando por los test clásicos y el análisis de series temporales. Proporciona una amplia gama de gráficos que además son fácilmente adaptables y extensibles. La calidad de los gráficos producidos y la posibilidad de incluir en ellos símbolos y fórmulas matemáticas, posibilitan su inclusión en publicaciones que suelen requerir gráficos de alta calidad.

El código de **R** está disponible como software libre bajo las condiciones de la licencia GNU-GPL. Además está disponible precompilado para una multitud de plataformas. La página principal del proyecto es <http://www.r-project.org>.

Una diferencia importante entre **R**, y también S, con el resto del software estadístico es el uso del **objeto** como entidad básica. Cualquier expresión evaluada por **R** tiene como resultado un objeto. Cada objeto pertenece a una clase, de forma que las funciones pueden tener comportamientos diferentes en función de la clase a la que pertenece su objeto argumento. Por ejemplo, el resultado de la función `print` evaluada sobre un vector da como resultado la impresión de todos los elementos del vector mientras que la misma función evaluada sobre una función muestra información sobre ella. De la misma manera, la función `plot` no se comporta igual cuando su argumento es un vector que cuando es un fichero de datos o una función.

A continuación se dan unas breves instrucciones que permitirán comenzar a usar **R** y su interfaz gráfica **R-Commander**, que se denotará abreviadamente como **Rcmdr**. Instrucciones más detalladas y actualizadas pueden encontrarse en <http://knuth.uca.es/R> en la sección R Wiki. Por último, existen multitud de documentos que ilustran sobre el manejo de **R**, algunos de ellos pueden descargarse desde <http://knuth.uca.es/R> en la sección Documentación. Los autores de este manual han redactado un somero documento técnico sobre el uso de **R**, a cuyo repositorio puede accederse en la dirección <http://knuth.uca.es/R-basico>.

2. Instalación de R y R-Commander

2.1. Instalación en GNU/Linux

Para la instalación, distribuciones derivadas de debian (Ubuntu, Guadalinex,...), en una consola se introduce en una sola línea:

```
sudo apt-get install r-base-html r-cran-rcmdr r-cran-rodbc  
r-doc-html r-recommended
```

Otra opción es utilizar el gestor de paquetes de la propia distribución e instalar los paquetes `r-base-html`, `r-cran-rcmdr`, `r-cran-rodbc`, `r-doc-html` y `r-recommended`.

2.2. Instalación en Windows

La descarga de **R** en el equipo se efectúa desde:

<http://cran.es.r-project.org/bin/windows/base/release.htm>

Luego se procede con la ejecución, siguiendo las instrucciones. Para la instalación de **Rcmdr**, se arranca **R** desde Inicio→Todos los programas→**R**. A continuación, Paquetes→Instalar Paquete(s) y elegido el mirror desde el cual se quiere instalar el paquete, por ejemplo *Spain (Madrid)*, se selecciona **Rcmdr**.

R-Nota 1.1

Harán falta más paquetes para la instalación completa de **Rcmdr**, pero se instalarán automáticamente la primera vez que se ejecute.

3. Ejecución de Rcmdr

En ambos sistemas operativos, la carga de la librería se efectuará mediante la instrucción `library("Rcmdr")`.

R-Nota 1.2

Si se cierra **Rcmdr** (sin cerrar **R**), para volver a cargarlo se debe ejecutar la instrucción `Commander()`.