

Práctica 0: Introducción a R

Grado de Ingeniería Civil
Universidad de Cantabria

Capítulo 1

Introducción a R

El paquete R¹ es una colección de programas libres² diseñada para el análisis estadístico de datos, que permite desde los análisis descriptivos más sencillos (como tablas de frecuencias simples, cálculo de la media o correlaciones) a procedimientos inferenciales más complejos (como contraste de hipótesis, análisis de la varianza o el análisis de componentes principales). Solo unas pocas de las posibilidades de análisis que ofrece R serán abordadas en este curso.

R realiza tres funciones esenciales: (1) leer datos, (2) especificar el tipo de análisis que se quiere realizar con esos datos y (3) mostrar los resultados obtenidos tras los análisis. La selección de una metodología apropiada al caso de estudio, así como la interpretación de los resultados es tarea del investigador. Por este motivo resulta necesario que el investigador conozca el fundamento teórico de los distintos métodos estadísticos disponibles con el objeto de que la aplicación del análisis y la interpretación de los resultados se realice de forma satisfactoria.

Este primer tema introductorio de R ha sido adaptado de *El Manual de R de Estadística* de EUITIO.

¹Sitio de referencia: <http://www.r-project.org>.

²En el sentido GNU: <http://gnu.org/philosophy/free-sw.es.html>.

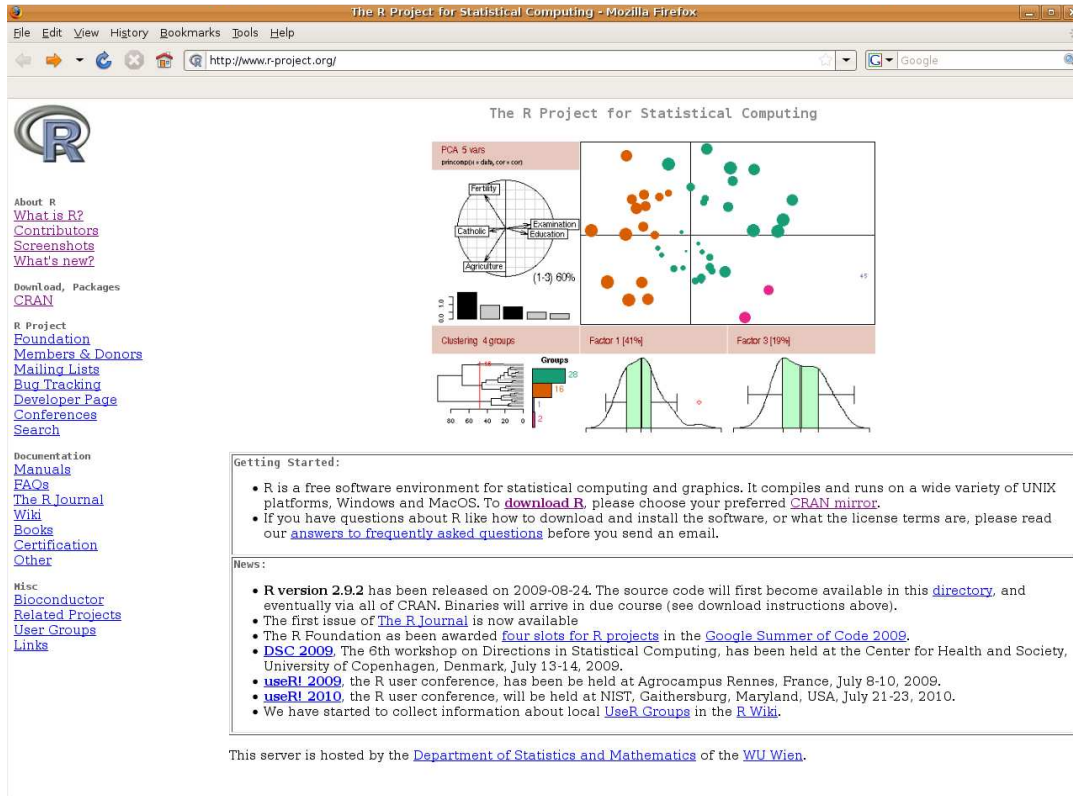


Figura 1.1: Página web del proyecto R

1.1. Objetivos del curso

Con este manual se pretende que el alumno se familiarice con el paquete R y sea capaz de utilizarlo como una herramienta de aplicaciones estadísticas a través de las prácticas que se plantean en clase. Para ello se propone al alumno que ejecute cada una de las órdenes que se van introduciendo en el manual, así como los ejercicios propuestos en las secciones **Practica tu mismo** que servirán para asimilar contenidos.

Todos los ficheros utilizados en este manual se pueden descargar de la web de la asignatura: <http://moodle.unican.es>.

1.2. Instalación de R

1.2.1. Microsoft Windows

Sigue los siguientes pasos para la instalación del software:

1. Descarga el fichero ejecutable desde la página del proyecto R

`http://cran.r-project.org/bin/windows/base/`

2. Ejecuta el fichero descargado, teniendo en cuenta:

- a) Cuando pregunta si deseamos establecer opciones de configuración, escoge *Sí*.
- b) Para el modo de presentación (MDI o SDI), escoge *SDI* (es conveniente por la implementación de la interfaz gráfica Rcommander que veremos después).

3. Ejecuta el programa R, ya instalado.
4. En el menú *Paquetes*, pincha en *Seleccionar espejo CRAN* para seleccionar un repositorio desde el cual se pueda descargar paquetes adicionales.
5. En el cuadro de diálogo, escoge *Spain (Madrid)*, *France (Toulouse)*, *Portugal* o algún otro cercano, y pulsa OK.
6. En el menú *Paquetes*, pincha en *Instalar paquete(s)*
7. Escoge *fBasics* y *Rcmdr* y acepta³. Estos son dos paquetes adicionales que necesitaremos tener instalados para algunas de las prácticas.

1.2.2. Linux

A través de la página inicial de R se accede a la descarga de R para diferentes versiones de Linux: Debian, Redhat, Suse y Ubuntu

`http://cran.r-project.org/bin/linux/`

Se recomienda seguir las instrucciones que se indican en la web en cada uno de los casos. Posteriormente instalar los paquetes `r-cran-rcmdr` y `r-cran-fbasics`.

³Si falta algún paquete aparecerá una ventana de aviso con los paquetes necesarios y la opción de instalarlos directamente. Los paquetes se instalan desde el espejo CRAN seleccionado anteriormente.

1.2.3. Mac Os X

En la siguiente dirección se accede a la instalación de R para Mac OS X:

<http://cran.r-project.org/bin/macosx/>

Seguid las instrucciones indicadas en la misma página web. Además en la siguiente web podeis encontrar los requisitos necesarios para Mac OS X antes de proceder a la instalación de R.

<http://r.research.att.com/tools/>

1.3. Estructura de R

El objetivo de esta primera práctica es que el alumno aprenda a manejar el programa R. Para ello, primero hablaremos de su estructura: sus ventanas y los elementos que las constituyen (barras de menús, de elementos activos, etcétera).

1.3.1. Comienzo de sesión

Tras arrancar el programa, aparece una ventana de órdenes titulada *Console R* que indica la versión de R y cómo obtener información acerca de la licencia de uso.

```
R version 2.9.2 (2009-08-24)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
```

```
R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.
```

```
R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener m'as informaci'on y
'citation()' para saber c'omo citar R o paquetes de R en publicaciones.
```

```
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.
```

```
>
```

Además al arrancar R vemos que también se especifican varias órdenes muy útiles para obtener ayuda acerca de las distintas funciones de R. Es el caso de:

```
>help() Muestra una ventana de ayuda general sobre R
>help.start() Arranca un manual de ayuda completo en formato html,
utilizando el navegador del sistema.
>help(mean) Muestra una ventana de ayuda sobre la función "media aritmética".
>?mean Lo mismo que el ejemplo anterior.
>help("[") Muestra una ayuda sobre el carácter [ , que es un carácter especial que forma parte del lenguaje R.
> apropos("mean") Muestra las funciones relacionadas con la función mean.
> help.search("mean") Busca ayuda sobre objetos o funciones que tengan nombre o título que contenga la cadena "mean".
```

1.3.2. Ventana de órdenes (consola)

Por debajo del título, esta ventana contiene una barra con los siguientes menús, cuyas opciones principales destacamos:

Archivo Operaciones básicas con los ficheros. Sólo usaremos:

Salir Para salir del programa.

Editar Típico menú con opciones de edición (copiar, pegar, ...).

Misc Opciones avanzadas.

Paquetes Permite gestionar los paquetes adicionales de R. Nos interesará la opción:

Seleccionar espejo CRAN Para activar una URL de donde descargar distintos paquetes de R.

Instalar paquete Permite seleccionar los paquetes que se quieren instalar.

Cargar paquete Para activar un paquete en concreto.

Ayuda Información abundante sobre R.

1.3.3. Interfaz gráfica (R-Commander)

R permite el uso de una interfaz gráfica⁴ que facilita su manejo permitiendo acceder fácilmente a muchas de las órdenes de gestión y análisis de

⁴Basada en Tcl/Tk.

datos del lenguaje R en lugar de escribir las instrucciones directamente en la consola. Para su uso se debe cargar el correspondiente paquete. Desde el menú *Paquetes*, pinche en *Cargar paquete* y escoja *Rcmdr*. Tanto en Linux como en Windows se puede arrancar este paquete escribiendo `library(Rcmdr)` en la línea de comandos. Este paquete se podrá cargar directamente siempre que antes haya sido instalado como se indicó en la sección 1.2. La pantalla presenta el aspecto de la figura 1.2 que consta de las siguientes partes:

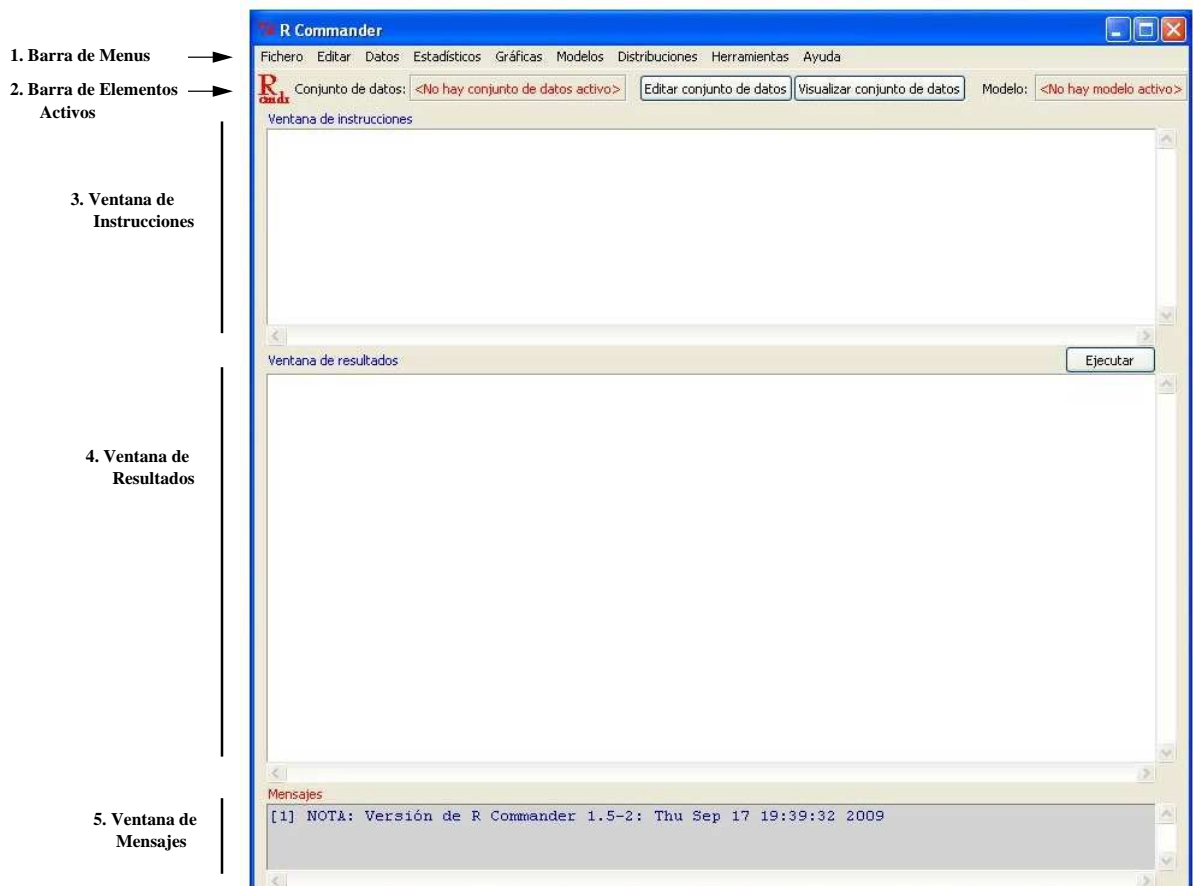


Figura 1.2: Aspecto inicial de la interfaz gráfica

- **Barra de Menús:** En la parte superior puede observarse una barra que consta de una serie de menús (*Archivo*, *Editar*, *Datos*, etc.). Si se selecciona con el ratón cada una de ellas aparece un menú desplegable donde se ofrecen otros submenús, cada uno de los cuales tiene a su vez un cuadro de diálogo que es el lugar donde se especifican los detalles de cada procedimiento.
- **Barra de Elementos Activos:** Inmediatamente debajo aparece otra barra que indica el conjunto de datos activo (*Datos:*), es decir el conjun-

to de datos con el que estamos trabajando. Hay además botones para ver los propios datos (*Visualizar datos*) y modificarlos (*Editar datos*).

- **Ventana de Instrucciones:** Muestra las órdenes de R correspondientes a las opciones de los menús escogidas por el usuario. Además, tales instrucciones se pueden modificar directamente en esta ventana, y ejecutar mediante el botón *Ejecutar*.
- **Ventana de Resultados:** Contiene aquellas salidas de las órdenes ejecutadas que se muestran en formato de texto.
- **Ventana de Mensajes:** Recoge la información adicional que R nos quiere hacer llegar (por ejemplo, errores, advertencias o mensajes administrativos).

Si por cualquier motivo salimos de la interfaz gráfica, podemos volver a acceder a ella escribiendo en la consola la orden `Commander()`.

A continuación, repasamos con más detalle cada una de las partes.

1.3.3.1. Barra de menús

En cada menú describimos solamente las opciones de interés para este curso:

Fichero Hay opciones para cargar o grabar instrucciones, resultados o el entorno de trabajo. También para salir de la interfaz gráfica, o también de R.

Editar Típico menú de edición. Permite seleccionar, cortar, copiar, pegar y buscar.

Datos Permite la gestión de los datos que se van a analizar. R mantiene distintos conjuntos de datos abiertos dentro del mismo entorno de trabajo. Sin embargo, uno (y sólo uno) de ellos se considera *activo*, siendo aquel al que se le aplicarán cada una de las funciones seleccionadas.

Un conjunto de datos es una matriz con variables como columnas y casos como filas. Sobre este tema haremos más hincapié en la sección 2.

Estadísticas Recoge los diferentes métodos de análisis que se pueden aplicar al conjunto de datos activo. Sobre este tema se profundizará en la siguiente práctica.

Gráficas Recoge los diferentes tipos de gráfico que se pueden obtener. Aprenderemos a manejar distintos tipos de gráficos en la siguiente práctica.

Modelos Un conjunto de datos puede tener asociados varios modelos estadísticos. Este menú sirve para la gestión de los mismos.

Distribuciones Para trabajar con funciones de distribución de probabilidad: cuantiles, probabilidades y gráficas asociadas a las distribuciones normal, t , χ^2 , F , binomial...

Herramientas Este menú de utilidades contiene:

Cargar paquete(s) Para cargar paquetes de R adicionales.

Opciones Para ajustar diferentes características de la interfaz, por ejemplo, el tamaño tipográfico.

Ayuda La ayuda de la interfaz gráfica es una extensión de la ofrecida por la consola.

Ayuda de R Commander Uso de la interfaz gráfica.

Introducción a R commander Artículo introductorio con imágenes.

Ayuda sobre los datos activos (si es posible) Misma opción que bajo el menú *Datos*.

Información sobre Rcmdr Versión y autores de la interfaz gráfica.

Además, la mayoría de los cuadros de diálogo dispone de un botón *Ayuda* que ofrece información sobre las órdenes de R asociadas a la acción correspondiente.

Con el fin de entender mejor el resto de menús de la ventana de R-Commander vamos a cargar un fichero de datos muy simple llamado *ejemplo.rda*⁵. Para ello escribimos en la ventana de instrucciones el siguiente comando:

```
DATA <- load('ejemplo.rda')
```

indicando delante del nombre del fichero la ruta de acceso completa donde se ha guardado el fichero de datos⁶. Ejecutamos esa línea dejando el cursor en la misma línea y pinchando el botón ejecutar. Con esta orden habremos cargado los datos contenidos en el fichero *ejemplo.rda* y le habremos asignado el nombre *DATA*. En la ventana de instrucciones se pueden escribir diferentes órdenes una en cada línea, sin embargo solo se ejecutarán aquellas que estén seleccionadas con el ratón al pinchar en el botón ejecutar.

⁵El fichero de datos se puede descargar de la página web de la asignatura, <http://moodle.unican.es>

⁶Si el fichero se ha guardado en el escritorio de Windows la ruta de acceso a los datos sería algo así: `C:/Documents and Settings/TuNombreDeUsuario/Escritorio/ejemplo.rda`

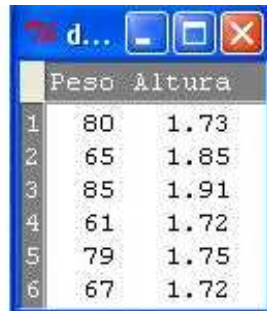
1.3.3.2. Barra de elementos activos

Como vemos en la figura 1.2 la barra de elementos activos consta de:

Conjunto de datos Nombre del conjunto de datos activo, es decir, el que se toma por omisión a la hora de ejecutar una orden. En nuestro caso aparece el fichero de datos que acabamos de cargar (*DATA*).

Editar conjunto de datos Hace aparecer una cuadrícula donde es posible modificar el contenido del conjunto actual de datos. Se trata de datos de las variables *Peso* y *Altura* de una muestra de individuos. Vemos como en la primera fila aparecen los nombres de las variables.

Visualizar conjunto de datos Muestra el contenido del conjunto actual de datos. En nuestro caso aparecen los valores del fichero que acabamos de cargar.



	Peso	Altura
1	80	1.73
2	65	1.85
3	85	1.91
4	61	1.72
5	79	1.75
6	67	1.72

En este caso también se pueden ver los datos en la ventana de resultados escribiendo en la ventana de instrucciones el nombre del fichero de datos (*DATA*) y pinchando en el botón ejecutar. También se puede seleccionar alguna de sus columnas del fichero activo mediante la orden `datos$Peso` por ejemplo. Si ejecutamos esa orden en la ventana de mensajes aparecen los datos de esa columna. Vemos así como el comando `$` permite seleccionar columnas de un fichero.

Modelo Para un mismo conjunto de datos se pueden crear diferentes modelos de análisis (de regresión lineal, de componentes principales...). Este menú permite escoger el modelo activo, es decir, aquél considerado por omisión cuando se ejecuta una orden.

1.3.3.3. Ventana de instrucciones

Como hemos visto se puede acceder a muchas órdenes desde los menús y los cuadros de diálogo. Además cada vez que se selecciona un opción de un menú la función correspondiente aparece escrita en la ventana de instrucciones. No obstante, algunas órdenes no están disponibles en los propios menús,

sino que sólo se pueden ejecutar mediante el uso del lenguaje R, es decir escribiendo la función correspondiente en la ventana de instrucciones. Además, es posible guardar las ordenes que aparecen en la ventana de instrucciones como un guión del análisis en un fichero de texto (habitualmente con extensión `.R`). De esta manera se podrá repetir los análisis en otro momento o ejecutarlos en un trabajo automatizado.

Un fichero `.R` es simplemente un fichero de texto que contiene órdenes. Es posible escribir órdenes directamente en la ventana de instrucciones (también en la consola), sin embargo, en muchos casos es más sencillo permitir que el programa le ayude a construir un guión aprovechando que la realización de una acción desde un cuadro de diálogo añade la orden a la ventana de instrucciones. En ésta la instrucción puede ser modificada para su posterior ejecución. Para ello, como ya se ha mencionado antes, ha de seleccionar con el ratón la orden u órdenes que se quieren ejecutar y después ha de pinchar en el botón *Submit* o *Ejecutar*.

Por ejemplo podemos calcular los estadísticos básicos del fichero de datos que tenemos activo, en nuestro caso el fichero `DATA`. Para ello seleccionamos el menú Estadísticos \Rightarrow Resúmenes \Rightarrow Conjunto de Datos activo. Vemos como en la ventana de instrucciones aparece el comando correspondiente a esta orden `summary(datos)`. Los resultados como veremos en la siguiente sección aparecen en la ventana de resultados (figura 1.3). La próxima vez que se quieran obtener estos estadísticos podemos escribir directamente la función en la ventana de instrucciones, sin necesidad de ir a los menús.

En el cuadro de diálogo de un procedimiento determinado, pulse en el botón *Help* o *Ayuda* para saber qué opciones del lenguaje R están disponibles (si hay alguna) para ese procedimiento. Si desea información completa sobre el lenguaje de órdenes, consulte el manual de referencia de R incluido en la web de la asignatura.

1.3.3.4. Ventana de resultados

Una vez que se solicita un análisis con los datos, los resultados obtenidos se muestran en la ventana inferior (*ventana de resultados*), ver figura 1.3. Como podemos comprobar, al ejecutar el comando anterior para calcular los estadísticos el resultado se ha mostrado en esta ventana.

El texto en rojo son las órdenes correspondientes que aparecen en la ventana de instrucciones. El texto en azul es el resultado de cada orden.

Los contenidos de la ventana de resultados son texto puro, que puede ser copiado a cualquier editor de texto para su procesamiento.

Podemos usar las ventanas de instrucciones y resultados a manera de calculadora. El lenguaje R permite las operaciones aritméticas básicas, la definición de variables o creación de funciones y dispone de una colección



```
> load("datos.rda")
> summary(datos)
      Peso      Altura
Min.   :61.00  Min.   :1.720
1st Qu.:65.50  1st Qu.:1.722
Median :73.00  Median :1.740
Mean   :72.83  Mean   :1.780
3rd Qu.:79.75  3rd Qu.:1.825
Max.   :85.00  Max.   :1.910
```

Figura 1.3: Ventana de resultados.

muy extensa de funciones ya definidas.

1.3.3.5. Ventana de mensajes

Recoge indicaciones, errores o advertencias que resultan de la ejecución de un comando de R.

1.3.4. Fin de sesión

En el menú *Archivo*, óptese por (*Salir*)

De Commander Se abandona la interfaz gráfica, pero no la consola de R.

Recuerde que para volver a la interfaz gráfica puede escribir `Commander()` en la consola.

De Commander y R Abandona el entorno R completamente.

En ambos casos se pide confirmación del abandono, y se pregunta si se quiere guardar el contenido de las ventanas de instrucciones y de resultados.

Capítulo 2

Manejo de datos

En este tema aprenderemos a manejar diferentes conjuntos de datos y a leer y almacenar en un fichero los datos necesarios para realizar un análisis. Estas tareas se realizan principalmente a través del menú *Datos* de la barra de menús, sin embargo también resulta interesante aprender a trabajar con vectores como veremos al final del capítulo.

2.1. Obtención de datos

Con R podemos introducir datos directamente, leerlos de un fichero ya existente, o bien utilizar datos que R trae de ejemplo.

Las opciones del menú *Datos* se muestran someramente a continuación:

Nuevo conjunto de datos Para introducir nuevos datos por el teclado. Requiere dar un nombre a los datos nuevos, que no puede contener espacios ni caracteres especiales. Ver sección 2.1.1 para más detalles.

Cargar conjunto de datos Para leer datos con formato propio de R.

Importar datos Para leer datos contenidos en un fichero. Soporta varios formatos: texto puro, SPSS, Minitab, Excel... Ver sección 2.1.2 para más detalles.

Conjunto de datos en paquetes R contiene una colección de datos de ejemplo, por si queremos ejercitarnos con el programa pero no disponemos de datos propios adecuados. Ver sección 2.1.3 para más detalles.

Conjunto de datos activos Aquí se gestiona el conjunto de datos activo.

Seleccionar conjunto de datos activos Elegir el conjunto de datos activo entre los que hay disponibles en ese momento en la sesión.

Esto mismo se puede hacer pinchando en la opción *Conjunto de datos* incluido en *Barra de Elementos Activos*.

Actualizar conjunto de datos activos Genera un mensaje con la estructura del conjunto de datos.

Ayuda sobre el conjunto de datos activos (si es posible) Algunos conjuntos de datos (como los de ejemplo) contienen una descripción.

VARIABLES DEL CONJUNTO DE DATOS ACTIVOS Lista los nombres de las variables del conjunto de datos.

Establecer nombres de casos A veces una variable no es tal, sino que contiene los nombres de los casos. Esta opción permite indicárselo a R.

Filtrar el conjunto de datos activos Si queremos que los análisis subsiguientes se realicen sobre un subconjunto de los casos, aquí podemos indicar una expresión de filtro. El filtro construye un nuevo conjunto de datos, cuyo nombre conviene indicar; en caso contrario, la selección se hace permanente (se eliminan los casos que no pasan el filtro).

Borrar fila(s) del conjunto de datos activos

Apilar variables del conjunto de datos activos Devuelve un conjunto de datos con las variables seleccionadas apiladas en una variable y acompañados de un factor que los distingue.

Eliminar los casos con valores omitidos En algunas variables, puede que se desconozca el valor para cierto caso: se trata de un dato ausente (*missing*). Esta opción elimina los casos con algún dato ausente.

Guardar el conjunto de datos activos Para guardar el fichero de datos activos con formato de R.

Exportar el conjunto de datos activos Para guardar una tabla con el conjunto de datos activo en un fichero de texto.

Modificar variables del conjunto de datos activos Para realizar transformaciones en los datos. Ver sección 2.2 para más detalles.

Recodificar variables Crea una nueva variable a partir de una ya existente. Sirve para agrupar datos cuantitativos en intervalos.

Calcular una nueva variable Crea una nueva variable a partir de una ya existente.

Añadir números de observaciones al conjunto de datos Etiqueta con números cada una de las filas del conjunto de datos.

Tipificar variables Para tipificar variables cuantitativas.

Convertir variable numérica en factor Indica al programa que los números no representan cantidades, sino categorías.

Segmentar variable numérica Simplifica la agrupación de datos cuantitativos en intervalos.

Renombrar variables Cambia el nombre de la variable.

Eliminar variables del conjunto de datos Elimina la variable.

A continuación veremos alguna de estas opciones con más detalle.

2.1.1. Creación de un conjunto de datos nuevo

En el menú *Datos* se encuentra la opción *Nuevo conjunto de datos*. Esta opción es conveniente cuando el conjunto de datos es pequeño. Para conjuntos de datos mayores, es más cómodo crear un fichero de datos por otros medios (por ejemplo, desde una hoja de cálculo o una base de datos) y luego importarlo ya que la interfaz de R no es muy amigable.

Lo primero que hay que tener en cuenta y no olvidar es que los conjuntos de datos (*data set*, *data frame*) están organizados de forma matricial, donde las filas se refieren a los casos (individuos, unidades u observaciones) de la muestra y las columnas a las variables.

Para introducir nuevos datos ha de escogerse la opción *Nuevo conjunto de datos* del menú *Datos*. Se nos pide entonces un nombre para el conjunto de datos (pues pueden manejarse varios simultáneamente).

Para introducir los datos simplemente se coloca el cursor en la celda correspondiente a cada individuo y variable. Para moverse de una celda a otra se puede utilizar el ratón, o la tecla de retorno para el desplazamiento vertical, o el tabulador para el desplazamiento horizontal o las flechas de desplazamiento del teclado.

Al introducir los datos, se observa que R da por omisión nombre a las variables (*var1*, *var2*, ...) y define sus características. En principio, una variable puede ser numérica (*numeric*) o de caracteres (*character*). Si se desea cambiar el nombre o definir el tipo de variable hay que pulsar en la cabecera de la columna correspondiente.

Llamaremos *factores* a las variables de caracteres. Nos servirán para representar variables cualitativas, es decir, aquéllas cuyo valores toman un número finito de modalidades.

Practica tu mismo:

2.1.1 Crea un nuevo conjunto de datos, según las indicaciones anteriores, con los siguientes valores. Define el tipo de dato que consideres oportuno en cada caso (numérica o de caracteres).

Peso	Altura	Edad	Sexo
80	1.73	20	V
85	1.91	19	V
61	1.72	19	M
79	1.75	25	M
67	1.72	21	V
65	1.70	19	M
55	1.59	25	V
75	1.75	19	M

2.1.2. Importar datos de un fichero externo

El fichero externo puede contener datos en formato de texto puro (ASCII) o en alguno de los formatos binarios soportados. En ambos casos ha de recurrirse al menú *Datos* ⇒ *Importar datos*.

Los ficheros de texto (columnas de números) representan la forma más universal para intercambio de datos. Para importar datos de texto se elegirá la opción *desde archivo de texto*, que abrirá el cuadro de diálogo *Leer archivo de texto* que se muestra en la figura 2.1.

Es necesario indicar:

Introducir el nombre del conjunto de datos: Escribir un nombre al conjunto construido a partir de los datos del fichero.

Nombres de las variables en el fichero: Marcar en caso de que el fichero contenga los nombres de las variables en la primera fila. Para ello se debe abrir el fichero antes con un editor de texto.

Indicador de datos ausentes: Cómo se indica si un campo no contiene valores, esto es, que se considera un valor ausente. Por omisión, el indicador es *NA* (*not available*, no disponible). Puede dejarse así a menudo, pues si un campo de una variable numérica está vacío, también se considera ausente.

Localización del archivo de datos: Indique la localización que corresponda.

Separador de campos: Indique el carácter que separa los campos, bien espacio en blanco, comas, tabuladores, o cualquier otro carácter que se puede especificar.

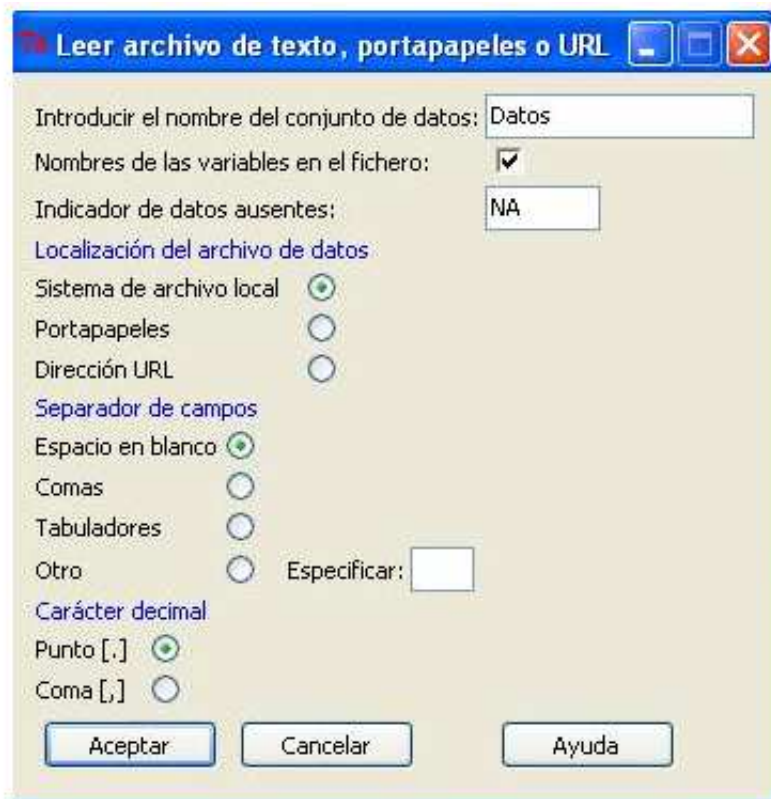


Figura 2.1: Leer archivo de texto.

Carácter decimal: Si se utiliza punto o coma para separar los decimales de la parte entera.

Se pueden abrir ficheros guardados desde otros programas estadísticos, como SPSS, Minitab o Excel. Por ejemplo, para abrir un fichero SPSS elegimos *desde datos SPSS* en el menú *Datos* \Rightarrow *Importar datos*.

Practica tu mismo:

2.1.2 Desde la página de moodle puedes descargarte ficheros con diferentes formatos para probar estas opciones. Prueba con los más comunes, un fichero de texto *datos.txt* y un fichero creado con Excel *datos.xls*. Visualiza los datos.

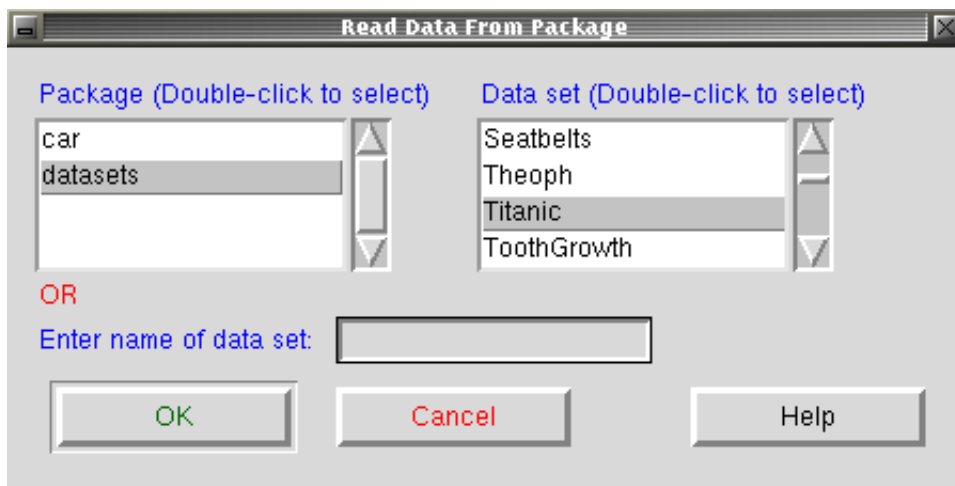
2.1.3. Utilizar datos incluidos en R

R incluye en su distribución una colección importante de datos de todo tipo. Para ver una descripción sucinta de los datos disponibles, elija la opción

Lista de conjuntos de datos en paquetes del menú Conjunto de datos en paquetes.

Data sets in package 'car':	
Adler	Experimenter Expectations
Angell	Moral Integration of American Cities
Anscombe	U. S. State Public-School Expenditures
Baumann	Methods of Teaching Reading Comprehension
Bfox	Canadian Women's Labour-Force Participation
Blackmoor	Exercise Histories of Eating-Disordered and Control Subjects
Burt	Fraudulent Data on IQs of Twins Raised Apart
Can.pop	Canadian Population Data
Chile	Voting Intentions in the 1988 Chilean Plebiscite
Chirot	The 1907 Romanian Peasant Rebellion
Cowles	Cowles and Davis's Data on Volunteering
Davis	Self-Reports of Height and Weight
DavisThin	Davis's Data on Drive for Thinness
Duncan	Duncan's Occupational Prestige Data
Ericksen	The 1980 U.S. Census Undercount
Florida	Florida County Voting
Freedman	Crowding and Crime in U. S. Metropolitan Areas
Friendly	Format Effects on Recall
Ginzberg	Data on Depression
Greene	Refugee Appeals
Guyer	Anonymity and Cooperation

Si alguno resulta de interés, escoja, en el mismo menú, la opción *Leer conjunto de datos desde paquete adjunto*. Indique el paquete y el conjunto de datos buscado, que se convertirá en el conjunto de datos activo.



Practica tu mismo:

2.1.3 Carga desde el paquete *datasets* de R el fichero *cars*. Para obtener información acerca del fichero consultar la ayuda donde se explican las características de las variables que contiene, unidades de las mismas ...
Visualiza el conjunto de datos.

2.2. Transformaciones de las variables

Vamos a utilizar las opciones básicas del menú *Datos / Modificar variables del conjunto de datos activos* que resultan más interesantes para este curso. Para ello vamos a trabajar con el fichero *cars* cargado anteriormente. Como habrás observado al cargar el fichero en la *Ventana de Mensajes* aparece el número de filas y columnas del fichero, es decir el número de casos y variables respectivamente. También se puede conocer la dimensión del fichero mediante la orden `dim(cars)`.

2.2.1. Calcular una nueva variable

Aquí podemos definir una nueva variable (o sobrescribir una antigua) mediante una expresión arbitraria.

Como hemos visto en la ayuda del fichero *cars*, las unidades de medida de la velocidad y la distancia de frenado de los coches son millas por hora y pies respectivamente. Supongamos que queremos pasar de las unidades de medida inglesa a unidades del Sistema Internacional más comunes, como son Km/h y metros. Teniendo en cuenta la equivalencia entre ambas¹, la opción correspondiente en este caso es *Calcular una nueva variable*. Por ejemplo para transformar la velocidad a Km/h deberíamos rellenar los campos de la ventana emergente como se muestra en la figura 2.2.

Si visualizamos de nuevo los datos vemos como aparece una nueva columna en el fichero.

Practica tu mismo:

2.2.4 Utilizando el comando que se ha generado en la ventana de instrucciones al hacer el cambio de unidad para la velocidad, escribe la orden adecuada para transformar la unidad de la variable *dist* de pies a metros.
Visualiza de nuevo el conjunto de datos.

¹1 mph = 1.61Km/h y 1 pie = 0,3048 metros

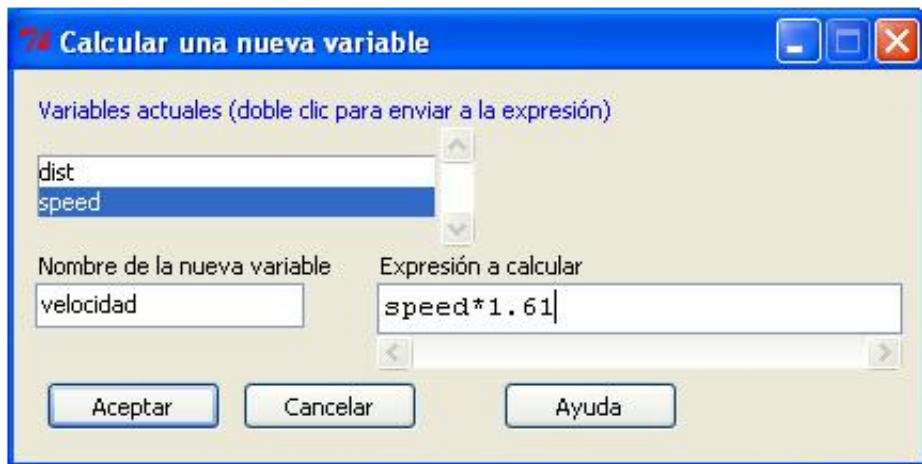


Figura 2.2: Calcular una nueva variable.

2.2.2. Renombrar

La opción *Renombrar variables* permite cambiar el nombre a una o varias de las variables del conjunto de datos activo.

Practica tu mismo:

2.2.5 Añade las unidades a los nombres de cada una de las variables del fichero *cars*.

Visualiza de nuevo el conjunto de datos.

2.2.3. Eliminar

La opción *Eliminar variables del conjunto de datos* permite eliminar una o varias variables del conjunto de datos activo.

2.3. Gestión del conjunto de datos activos

Sólo explicaremos en detalle la opción *Filtrar el conjunto de datos activos* del menú *Datos / Conjunto de datos activos*. Como ya se ha mencionado esta opción permite seleccionar un subconjunto de datos del fichero de datos activo de acuerdo a una expresión. Por ejemplo, de los datos del fichero *cars* nos puede interesar la distancia de frenado de aquellos vehículos que circulaban a una velocidad mayor de 30Km/h. Es decir, queremos obtener una subconjunto de los datos originales que cumplan esa condición. Si seleccionamos la opción *Filtrar el conjunto de datos activos* nos aparece una ventana como la de la figura 2.3 en la que debemos seleccionar únicamente la

variable *distancia* que es la que nos interesa. En el cuadro *expresión* debemos indicar la condición que queremos que cumplan estos datos, en nuestro caso `velocidad>30`. Por último asignar un nombre para guardar el subconjunto de datos resultante.



Figura 2.3: Filtrar el conjunto de datos activo.

En la ventana de instrucciones aparece escrita la orden que hemos ejecutado:

```
> vel30 <- subset(cars, subset=velocidad>30, select=c(distancia))
```

Además en la *Barra de Elementos Activos* aparece el nuevo conjunto de datos que acabamos de crear.

Practica tu mismo:

2.3.6 Intenta adivinar que ocurre si escribes las siguientes instrucciones:

- `cars$distancia[cars$velocidad>30]`
- `subset(cars, subset=velocidad>30)`
- `cars$distancia[cars$velocidad>30 & cars$velocidad<35]`
- `cars$velocidad[cars$distancia<=10]`

2.4. Estructuras de datos en R

R utiliza diferentes estructuras de datos. La estructura más simple es el vector, que es una colección ordenada de números. Para crear un vector, por ejemplo `x`, consistente en cinco números, por ejemplo 10.4, 5.6, 3.1, 6.4 y 21.7, use la orden

```
> x <- c(10.4, 5.6, 3.1, 6.4, 21.7)
```

Esta es una asignación en la que se utiliza la función `c()` que, en este contexto, puede tener un número arbitrario de vectores como argumento y cuyo valor es el vector obtenido mediante la concatenación de todos ellos. Un número, por sí mismo, se considera un vector de longitud uno.

Advierta que el operador de asignación, `(<-)`, no es el operador habitual en otros lenguajes, `=`, que se reserva para otro propósito, sino que consiste en dos caracteres, `<` (menor que) y `'-'` (guión), que obligatoriamente deben ir unidos y apuntan hacia el objeto que recibe el valor de la expresión. La asignación puede realizarse también mediante la función `assign()`. Una forma equivalente de realizar la asignación anterior es:

```
> assign('x', c(10.4, 5.6, 3.1, 6.4, 21.7))
```

El operador usual, `<-`, puede interpretarse como una abreviatura de la función `assign()`.

Si una expresión se utiliza como una orden por sí misma, su valor se imprime y se pierde. Así pues, la orden

```
> 1/x
```

simplemente imprime los inversos de los cinco valores anteriores en la pantalla

pero al no haberle asignado ninguna variable al resultado este no queda guardado, (por supuesto, el valor de `x` no se modifica).

Si a continuación hace la asignación

```
> y <- c(x, 0, x)
```

creará un vector, `y`, con 11 elementos, consistentes en dos copias de `x` con un cero entre ambas.

Se puede estar interesado en seleccionar un determinado valor del vector, por ejemplo el valor 5.6 del vector `x`. Para ello únicamente tendremos que saber la posición de dicho elemento en el vector y escribir la orden

```
> x[2]
```

Otra estructura común es la matriz de datos. Podemos crear una matriz de la forma:

```
> M <- matrix(c(1,2,3,4,5,6),nrow = 2, ncol = 3)
```

y al igual que con vectores podemos seleccionar elementos de la matriz como la fila 2:

```
> M[2,]
```

la columna 3:

```
> M[,3]
```

o el elemento de la fila 1 y la columna 2:

```
> M[1,2]
```

Los vectores y matrices pueden usarse en expresiones aritméticas, en cuyo caso las operaciones se realizan elemento a elemento. Los operadores aritméticos elementales son los habituales `+`, `-`, `*`, `/` y `^` para elevar a una potencia. Además están disponibles las funciones `log`, `exp`, `sin`, `cos`, `tan` y `sqrt` para calcular la raíz cuadrada, bien conocidas. Existen muchas más funciones, entre otras, las siguientes: `max` y `min` que seleccionan respectivamente el mayor y el menor elemento de un vector; `range` cuyo valor es el vector de longitud dos conteniendo el mínimo y el máximo, `c(min(x), max(x))`; `length(x)` que es el número de elementos o longitud de `x`; `sum(x)` que es la suma de todos los elementos de `x`; `prod(x)` que es el producto de todos ellos y `sort(x)` que ordena los elementos de `x` de menor a mayor. Como ya se comentó al principio se puede obtener más ayuda acerca de estas funciones con la orden `help()`

Practica tu mismo:

2.4.7 Crea un vector `x` con la columna `speed` del fichero `cars` que tienes cargado. Para ello selecciona ese conjunto de datos como fichero activo y ejecuta la siguiente línea de comandos:

```
>x<- cars$speed
```

Prueba todas las funciones mencionadas en esta sección y observa los resultados que obtienes.

2.4.8 Con el mismo fichero `cars` analiza los resultados de las siguientes instrucciones:

- `cars[2,]`

- `cars[3,4]`

- `cars[cars$velocidad>30,]`

- Los resultados de esta última orden ya los has obtenido antes con otro comando de R. ¿Recuerdas cuál?

2.4.9 Introduce el vector: `>x<-c(1,3,5,7,9)` e intenta adivinar antes de calcular los resultados de las siguientes instrucciones:

- `sum(x>5)`
- `sum(x[x>5])`
- `which(x>5)`
- `x>5`

- Extrae el elemento de `x` que ocupa la posición 2.

- Extrae utilizando un único comando, los elementos 3 elementos de `x` que ocupan las posiciones de la 2 a la 4.

2.4.10 Utiliza la función `seq` (consulta la ayuda para ver como se ejecuta esta función, recuerda: `>?seq`) para crear los siguientes vectores:

- 55, 56, 57, 58, 59
- 1, 3, 5, 7, 9, 11

2.4.11 Crea la siguiente matriz de datos: `>m<-matrix(1:15, nrow=3)` e intenta adivinar antes de calcular los resultados de las siguientes instrucciones:

- `dim(m)`
- `ncol(m)`
- `nrow(m)`
- `which(m>=5 & m<=8)`
- Extrae la primera fila.
- Extrae el segundo elemento de la tercera columna.