EACS-Excel, R y R Commander

Índice

1. **Aplicaciones informáticas para el análisis de datos**
2. **Breves nociones de Excel**
3. **R y R Commander**
4. **Instalación de R y R Commander**
5. **Ventanas de R y R Commander**
6. **Manejo de ficheros de datos con R Commander**
7. **Importación de conjuntos de datos con R Commander**
8. **Los cuadros de sintaxis y resultados. La ventana de gráficos**
9. **Algunas transformaciones de variables**
10. **Resumen del conjunto de datos**
11. **Salir de R Commander**
12. **Algunos consejos prácticos**
13. **Ejercicios**

1. Aplicaciones informáticas para el análisis de datos

Existen muchos paquetes informáticos para la gestión de bases de datos y el análisis estadístico. En estas prácticas utilizaremos el paquete estadístico R con el intérprete R Commander, que permite una utilización sencilla de R con una estructura de ventanas y comandos. Sin embargo, la introducción y depuración de datos con R Commander es poco flexible, y utilizaremos Excel para este cometido.

2. Breves nociones de EXCEL

Excel es una aplicación desarrollada por Microsoft (*Microsoft Office*) para manejar hojas de cálculo. Seguramente, Excel es el software más extendido en entorno Windows para el manejo de dichas hojas. Una hoja de cálculo no es más que un programa que permite manipular datos numéricos y alfanuméricos dispuestos en forma de tablas, donde es posible realizar cálculos complejos con fórmulas y funciones, además de dibujar diferentes tipos de gráficos.

Aunque no es un software desarrollado específicamente para realizar análisis estadísticos detallados, *Excel* dispone de una serie de herramientas que os permitirán, en caso de necesidad, realizar sencillos análisis estadísticos. Estas breves notas, sólo pretenden daros una idea general de cómo se maneja el programa y dónde podéis encontrar estas herramientas estadísticas.

NOTA 2.1: En función del sistema operativo que estés empleando y la versión de *Office* que tengas instalada, las salidas y comentarios que se señalan a continuación pueden variar un poco, aunque la filosofía de funcionamiento es la misma. En nuestro caso, las salidas corresponden a la *Excel* del *Office 2016*, bajo sistema operativo *Windows 10*.

NOTA 2.2: Cualquier estudiante de la Universidad de Zaragoza puede utilizar parte del software licenciado por la misma. En particular, podéis instalar y usar Microsoft Office (MS Word, Excel, PowerPoint, …) en vuestros ordenadores. Para más información, podéis consultar los siguientes enlaces:

https://sicuz.unizar.es/soporte-equipamiento-puesto-de-trabajo/gestion-de-software-corporativo/microsoft

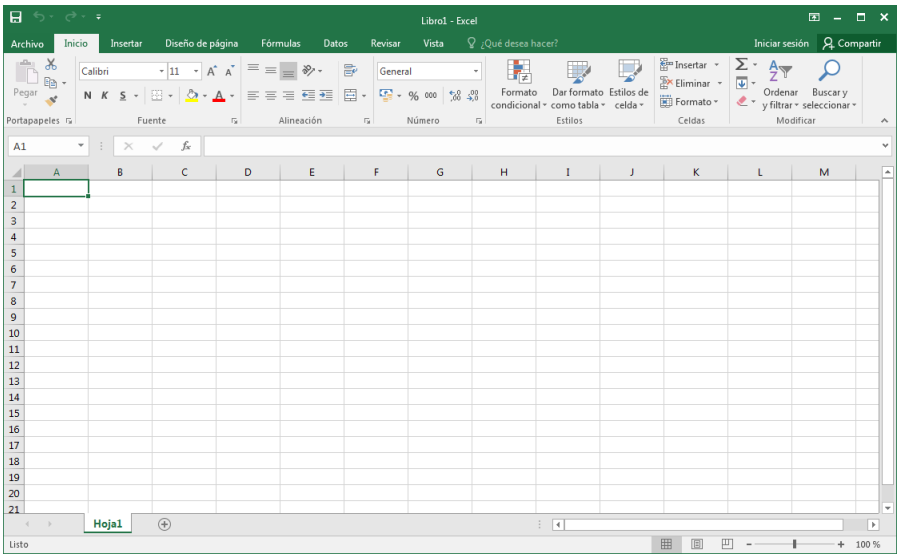
https://sicuz.unizar.es/soporte-equipamiento-puesto-de-trabajo/gestion-de-software-corporativo/software-para-estudiantes

Resumiendo, tenéis que crear una cuenta Microsoft educativa en el enlace

https://products.office.com/es-es/student/office-in-education,

cuyo identificador coincida con la cuenta de correo electrónico que os ha proporcionado la Universidad. Una vez dados de alta, ya podréis usar Office de forma on-line. También podréis instalaros el Office en vuestro ordenador a través de la opción *Instalar aplicaciones*, opción que suele estar en el menú de configuración.

Para acceder a *Excel*, selecciona *Excel* desde el menú *Inicio*. Se abrirá la siguiente ventana, que te permitirá acceder a todas las funcionalidades del software.



Observa que se ha abierto un documento (*Libro1*) conteniendo una hoja de cálculo (*Hoja1*). Para crear nuevas hojas de cálculo, no tienes más que pulsar sobre el botón "más" que se encuentra a la derecha del nombre *Hoja1*. Para pasar de una hoja de cálculo a otra, hay que pulsar sobre el nombre de la hoja de cálculo que se quiere activar.

Para moverte por una hoja de cálculo y pasar de una celdilla a otra, no tienes más que usar el ratón o las teclas de las flechas y la tecla <*return*>. Observa que las diferentes columnas vienen etiquetadas como A, B… y las diferentes filas como 1, 2…

Como ejemplo, vamos a introducir en la *Hoja1* los siguientes datos que contienen la altura, en metros, de un grupo de chicos y chicas.

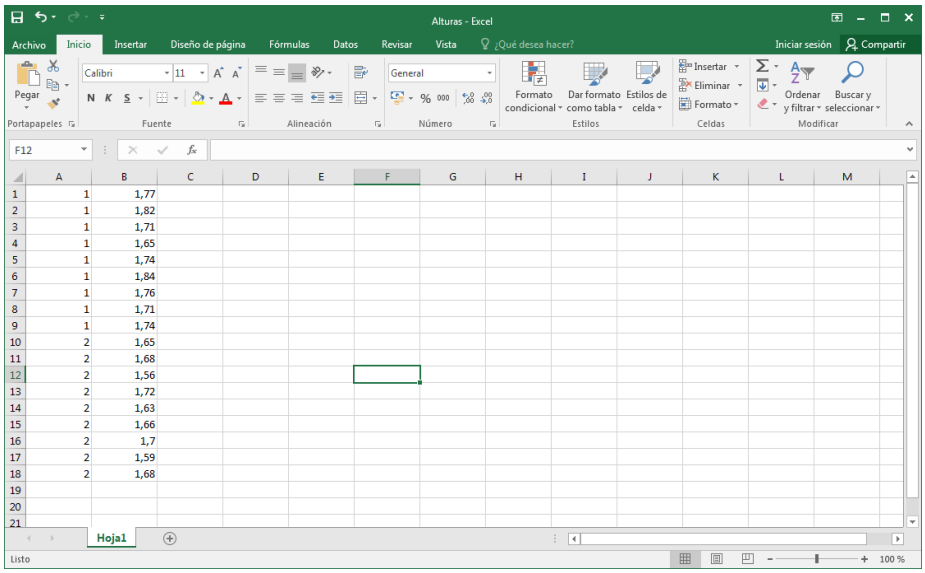
Chicos	Chicas
1.77	1.65
1.82	1.68
1.71	1.56
1.65	1.72
1.74	1.63
1.84	1.66
1.76	1.70
1.71	1.59
1.74	1.68

Nos planteamos primero qué variables nos están proporcionando. Vemos que hay dos: Sexo y Altura. Después, cuántos individuos hay en este grupo. Son 18, 9 chicos y 9 chicas. Por comodidad, codificaremos la variable Sexo con 1 para chicos y 2 para chicas. Crearemos dos columnas, una conteniendo unos y doses en función de si la altura es de un chico (1) o una chica (2) y una segunda conteniendo todas las alturas.

Para ***introducir los datos***, basta con que coloques el ratón en la primera casilla en blanco de la primera columna, teclea 1 y pulsa <*return*>. Se habrá introducido un 1 en la primera casilla y el ratón se habrá colocado en la segunda casilla. Repite la operación hasta que hayas introducido nueve unos. A continuación, repite el proceso hasta introducir nueve doses. Verás que tendrás 18 casos, los nueve primeros con unos y los nueve segundos con doses.

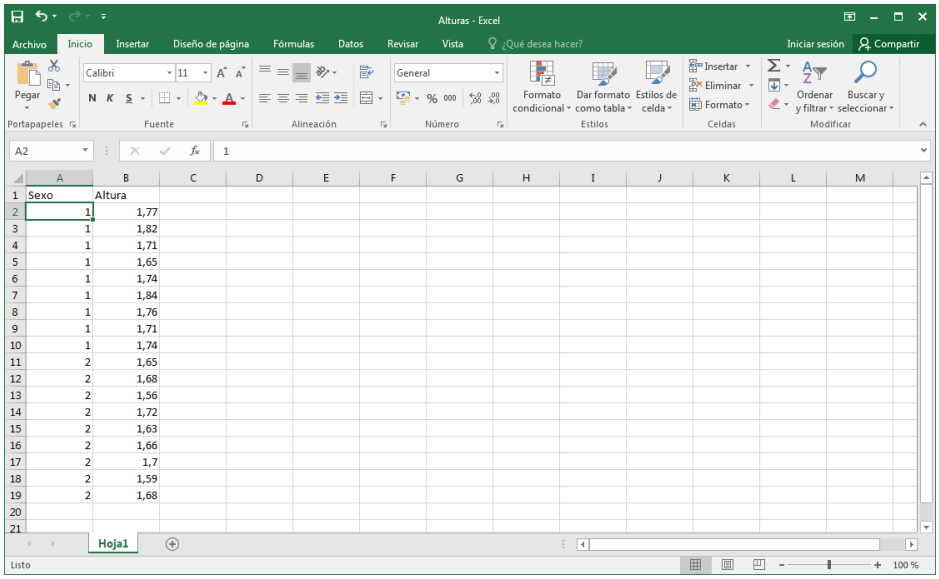
Luego, para introducir las alturas de las diferentes personas, colócate en la primera casilla de la segunda columna y ve introduciendo las alturas, tecleándolas y pulsando <return>. La parte decimal se codifica usando la *coma* (aunque depende de cómo sea la configuración de tu ordenador podría ser el *punto*); en cualquier caso, si introduces los datos con el teclado numérico, la tecla "." de este teclado siempre la identificará como el punto decimal. Si introduces la parte decimal de forma errónea, Excel entenderá que esa casilla es de tipo Texto y no podrás realizar operaciones matemáticas con ella.

En nuestro caso, la hoja Excel mostrará este aspecto.

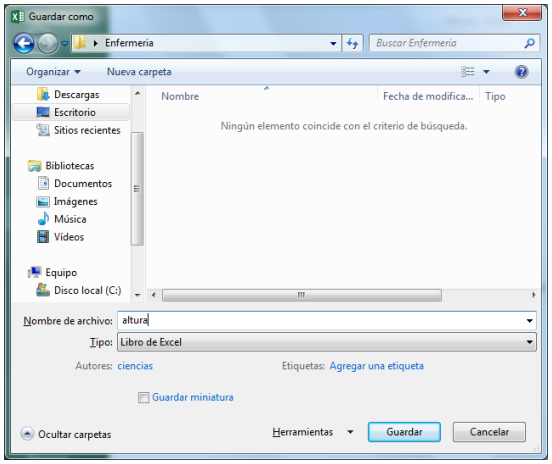


Para ***poner nombre a las columnas*** (por ejemplo, *sexo* y *altura*), tienes que utilizar siempre la primera fila. En nuestro caso, como la primera fila ya está con datos, tendremos que insertar una nueva fila para poner los nombres. Para realizar esta operación, sigue los siguientes pasos:

1) Coloca el ratón sobre la etiqueta 1 que identifica a la primera fila. Con el botón derecho del ratón, elige la opción *Insertar*. Verás que se crea una fila vacía en la primera posición.
2) Escribe la palabra *Sexo* en la primera casilla de la primera columna y la palabra *Altura* en la primera casilla de la segunda columna. Observa que los datos numéricos se alinean a la derecha, mientras que los datos alfanuméricos (o caracteres) se alinean a la izquierda. La hoja de cálculo te habrá quedado entonces de la siguiente forma.

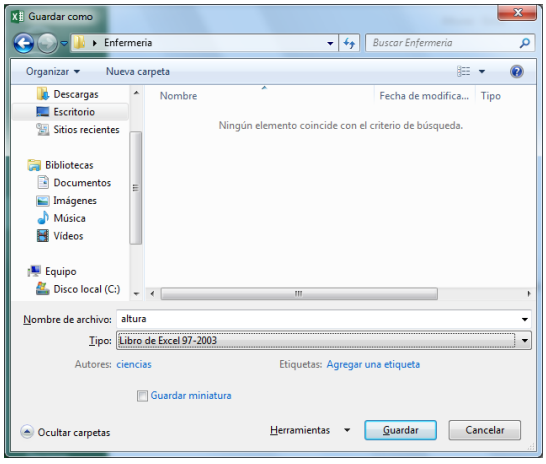


Para ***guardar estos datos*** en tu ordenador, (en estos momentos sólo los tienes en la memoria), selecciona la opción *Guardar como ...* en el menú Archivo (los menús aparecen como una pestaña en la parte superior de la ventana) o presiona directamente sobre el icono . En la nueva ventana que aparece, selecciona *Examinar* para elegir la carpeta donde quieres guardar los datos. Se te abrirá un cuadro de diálogo como el que figura a continuación.



En este cuadro de diálogo podrás elegir el nombre para tu archivo (en el ejemplo *altura*), el directorio donde lo vas a guardar (en el ejemplo en la carpeta *Enfermería*) y el formato (o tipo) con el que lo quieres guardar (en el ejemplo, *Libro de Excel (*.xlsx)*. Pulsando el botón *Guardar*, se almacenará el archivo en tu ordenador.

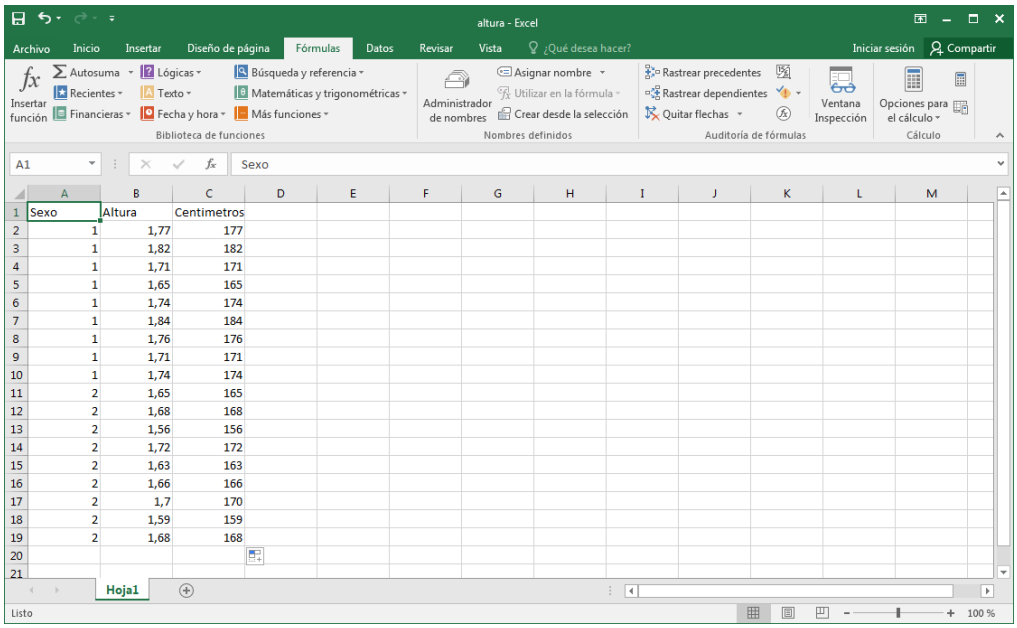
NOTA 2.3: Si no se especifica otra cosa, el formato con el que se guardan los datos es el propio de la versión de *Excel* con la que estéis trabajando. Si estáis trabajando con las versiones de *Excel 2007/2010/2013/2016*, se os guardará con formato *Excel 2007/2010/2013/2016 (*extensión .xlsx*)*, y si estáis trabajando con la versión de *Excel 2003*, se os guardará con formato *Excel 2003* (extensión .xls*)*. Mientras que las versiones de *Excel 2007/2010/2013/2016* son capaces de manejar los archivos generados con versiones anteriores de *Excel*, el recíproco no es cierto: un archivo con formato *Excel 2007/2010/2013/2016* no puede ser manejado por la versión de *Excel 2003*. Por ejemplo, si estamos manejando la versión *Excel 2016*, y queremos guardar los datos para trabajarlos con la versión *Excel 2003*, elegiríamos *Libro de Excel 97 – 2003 (*.xls)* en el campo *Tipo*.



La ***realización de transformaciones*** en Excel se hace a través del menú *Fórmulas* en la parte superior de la ventana y, una vez allí, pulsando en el símbolo f_x *(Insertar función)*. Si la transformación es sencilla (por ejemplo, sumar o multiplicar), se puede hacer directamente sin necesidad de pasar por el menú Fórmulas. Así, si queréis pasar las alturas a centímetros, creando una nueva columna con dichas alturas (la operación a realizar es altura*100), los pasos que seguiríais son los siguientes:

1) Coloca el ratón en la casilla *C2* (tercera columna, segunda fila).
2) Teclea =.
3) Con el ratón, pulsa sobre la casilla *B2* (contiene la altura en metros para el primero de los chicos).
4) Teclea **100*. Observa que la definición para la casilla *C2* te habrá quedado *=B2*100*.
5) Pulsa la tecla *<Return>*. Observa que en la casilla *C2* aparece el dato numérico *177*.
6) Coloca el ratón en la esquina inferior derecha de la casilla *C2* hasta que veas que cambia de forma (┼).

7) Con el ratón cambiado de forma, arrástralo hasta la casilla *C19*, con su botón izquierdo pulsado.
8) En la casilla *C1*, escribe la palabra Centímetros. La hoja de cálculo *Hoja1* te habrá quedado de la siguiente forma:

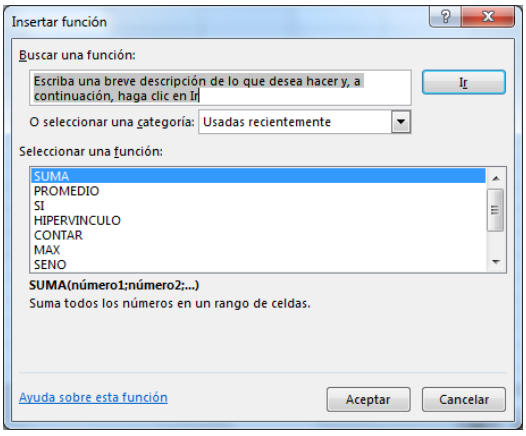


	A	B	C
1	Sexo	Altura	Centimetros
2	1	1,77	177
3	1	1,82	182
4	1	1,71	171
5	1	1,65	165
6	1	1,74	174
7	1	1,84	184
8	1	1,76	176
9	1	1,71	171
10	1	1,74	174
11	2	1,65	165
12	2	1,68	168
13	2	1,56	156
14	2	1,72	172
15	2	1,63	163
16	2	1,66	166
17	2	1,7	170
18	2	1,59	159
19	2	1,68	168

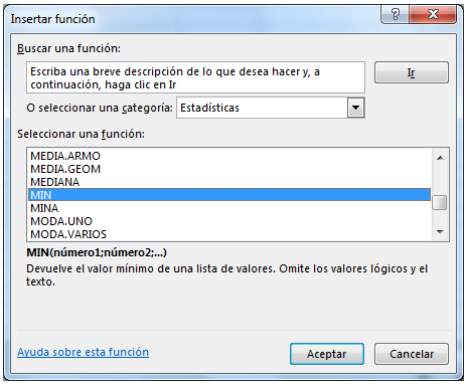
Observa que, para la creación de la nueva columna, hemos definido la transformación únicamente para una de las casillas (la primera que iba a contener datos numéricos) y hemos arrastrado hacia abajo esa definición. Esta forma de trabajar, arrastrando el resultado de una celda a una fila o columna, es muy habitual en Excel y es necesario familiarizarse con ella para poder sacar partido a las potencialidades del programa.

Excel tiene definidas muchas ***funciones*** para realizar cálculos complicados o para realizar transformaciones más complicadas. Para acceder a este conjunto de funciones, basta con pulsar sobre el símbolo f_x dentro del menú Fórmulas. Por ejemplo, vamos a utilizar una de estas funciones para calcular la más pequeña de todas las alturas expresadas en centímetros y la vamos a guardar en la casilla *C20*. Los pasos que seguirías serían los siguientes:

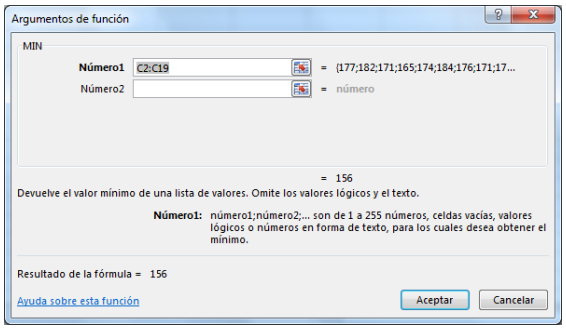
1) Coloca el ratón en la casilla *C20* y pulsa sobre el símbolo f_x. Te aparecerá un cuadro de diálogo similar a éste:



2) Dentro del campo *O seleccionar una categoría*, elige la opción *Estadísticas*. En el cuadro *Seleccionar una función*, te aparecerán todas las funciones estadísticas que lleva predefinidas *Excel*.
3) De entre todas las funciones que te aparecen, elige la función *MIN*.



4) Al pulsar *Aceptar*, te aparecerá el siguiente cuadro de diálogo:

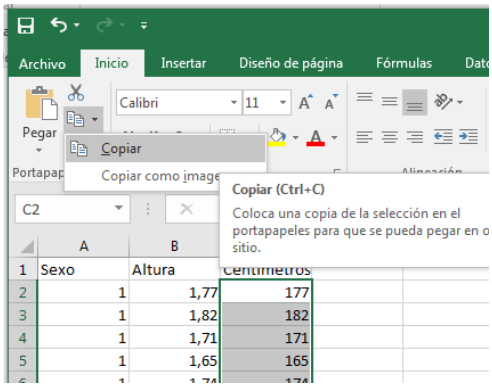


5) Cuando te aparece el cuadro de diálogo en *Número 1* aparece C2:C19 de manera automática. Esto es porque Excel cree que lo que queremos hacer es calcular el mínimo de esas casillas, ya que estamos situados justamente debajo de ellas. Si es así, simplemente tendremos que pulsar en *Aceptar*. Si lo que queremos es elegir otro conjunto de casillas para que nos calcule el mínimo, entonces las seleccionaremos con el ratón de la hoja de cálculo (la selección ha de hacerse bien de casillas adyacentes horizontales, de casillas adyacentes verticales, de una matriz de casillas, o de varias casillas dispersas, en cuyo caso se van pulsando las casillas manteniendo la tecla <ctrl> pulsada). Como en nuestro caso lo que aparece de forma automática es exactamente lo que queremos, pulsamos *Aceptar*.

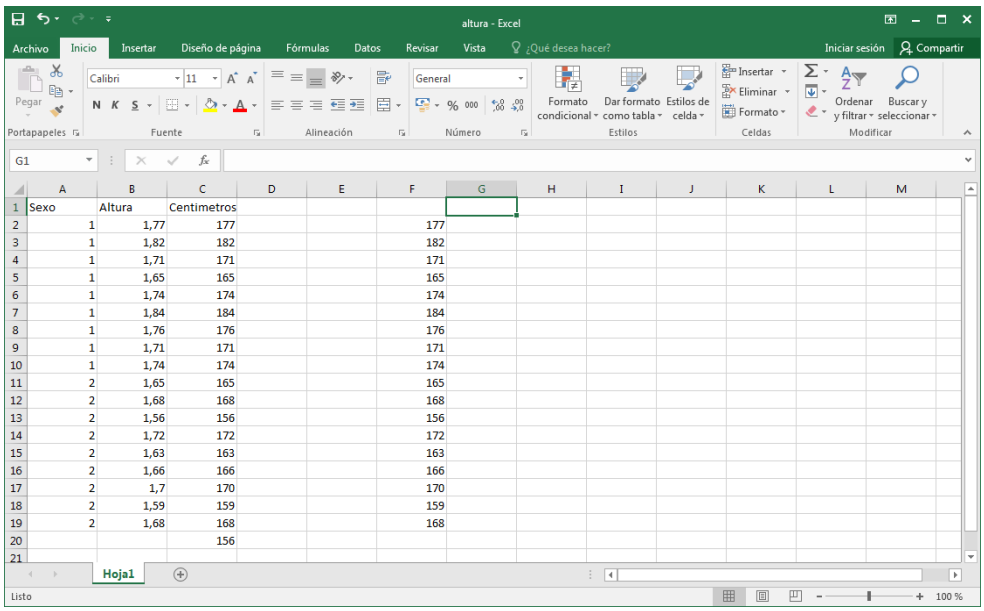
Como resultado verás que la casilla *C20* contiene el valor *156*, que es precisamente la más pequeña de todas las alturas.

Una forma alternativa de encontrar este mínimo sin utilizar esta función sería copiar los datos de la altura en centímetros en una nueva columna y luego ordenar los datos de esa columna. Para copiar los datos a una columna nueva, basta con que sigas los siguientes pasos:

1) Selecciona con el ratón los datos de la altura en centímetros (casillas *C2* a *C19*).
2) Presiona a la vez las teclas <*ctrl.*> + *c*, o elige la opción *Copiar* en el menú de *Inicio* para copiar los datos al portapapeles.



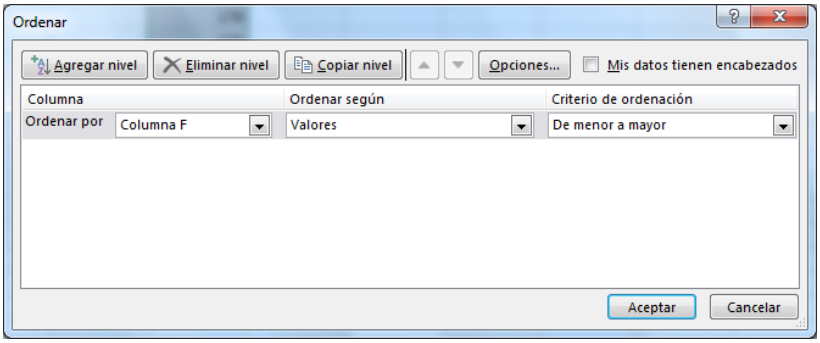
3) Sobre la casilla *F2*, la primera que va a contener la copia de los datos, pulsa el botón derecho del ratón y elige *Pegar valores* (aparece con una pizarra con el número 123) dentro de *Opciones de pegado*.
4) Verás que en la columna *F* te aparecen los datos copiados.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sexo	Altura	Centimetros										
2	1	1,77	177			177							
3	1	1,82	182			182							
4	1	1,71	171			171							
5	1	1,65	165			165							
6	1	1,74	174			174							
7	1	1,84	184			184							
8	1	1,76	176			176							
9	1	1,71	171			171							
10	1	1,74	174			174							
11	2	1,65	165			165							
12	2	1,68	168			168							
13	2	1,56	156			156							
14	2	1,72	172			172							
15	2	1,63	163			163							
16	2	1,66	166			166							
17	2	1,7	170			170							
18	2	1,59	159			159							
19	2	1,68	168			168							
20			156										
21													

Para ordenar ahora estos datos copiados, sigue los siguientes pasos:

1) Selecciona con el ratón los datos copiados anteriormente (casillas *F2* a *F19*). Te habrán quedado las casillas seleccionadas en azul.
2) Selecciona la opción *Ordenar* dentro del menú *Datos*. Te aparecerá el siguiente cuadro de diálogo.

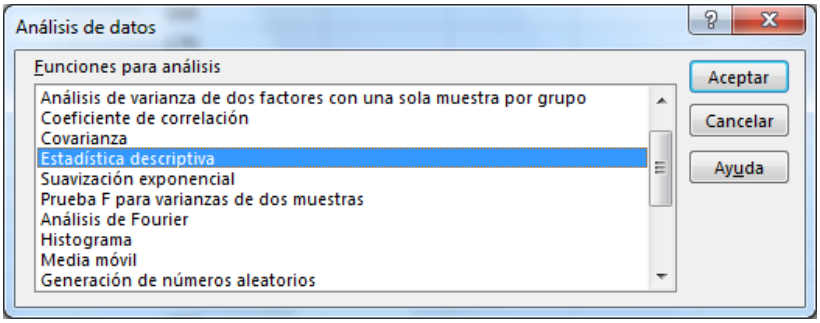


3) Pulsa *Aceptar* (se estarán ordenando de forma ascendente los datos de la columna *F*, los que hemos seleccionado). Comprobarás que en la casilla *F2* aparece el dato 156, el más pequeño de todos.

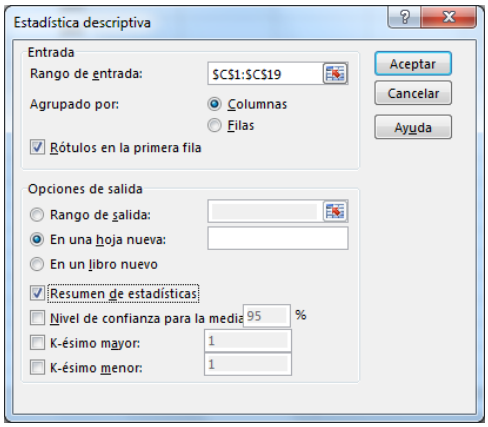
Además de este tipo de funciones, Excel lleva incorporadas una serie de ***macros*** que permiten realizar diferentes operaciones de tipo estadístico. Estas macros se pueden usar eligiendo la opción *Análisis de datos*, dentro del menú *Datos*.

Por ejemplo, vamos a obtener un conjunto de medidas descriptivas (media, desviación típica, mínimo, máximo…) de las alturas expresadas en centímetros. Para esto, tienes que seguir los siguientes pasos:

1) En el menú *Datos*, elige la opción *Análisis de datos*. En el cuadro de diálogo que te aparece, elige *Estadística descriptiva* y pulsa *Aceptar*.

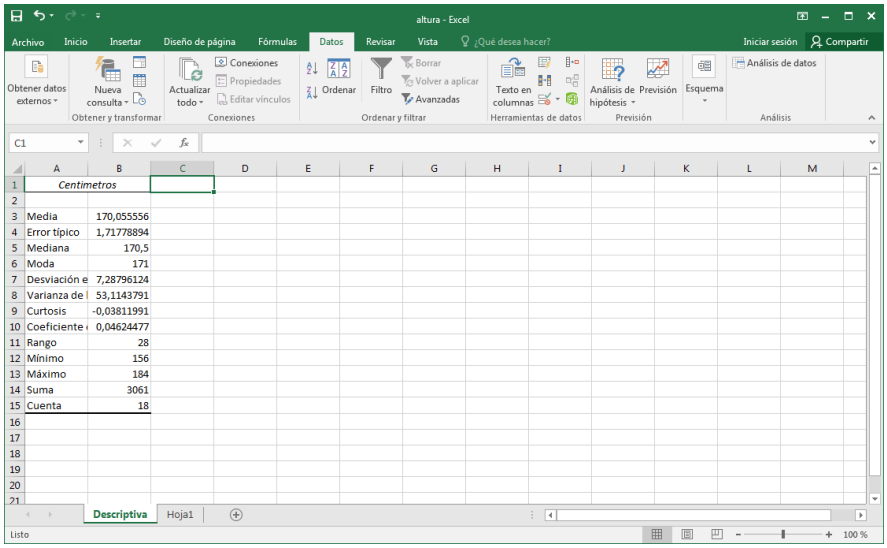


2) En el campo *Rango de entrada* del nuevo cuadro de diálogo que te aparece, hay que especificar las casillas para las que quieres obtener las descriptivas. Para ello, selecciona con el ratón todas las casillas correspondientes a la altura en centímetros, incluido el nombre. Verás que en el campo *Rango de entrada* aparece *$C$1:$C$19*.
3) Selecciona la casilla *Rótulos en la primera fila* (en la primera fila de las casillas seleccionadas se encuentra el nombre de la variable) y selecciona la casilla *Resumen de estadísticas* (para obtener en salida las medidas descriptivas). Observa que, por defecto, la salida va a parar siempre a una hoja de cálculo nueva, a no ser que se especifique otra cosa.



4) Al pulsa *Aceptar*, obtendrás las diferentes mediadas descriptivas. Observa cómo *Excel* ha llevado los resultados a la hoja de cálculo *Hoja2*.

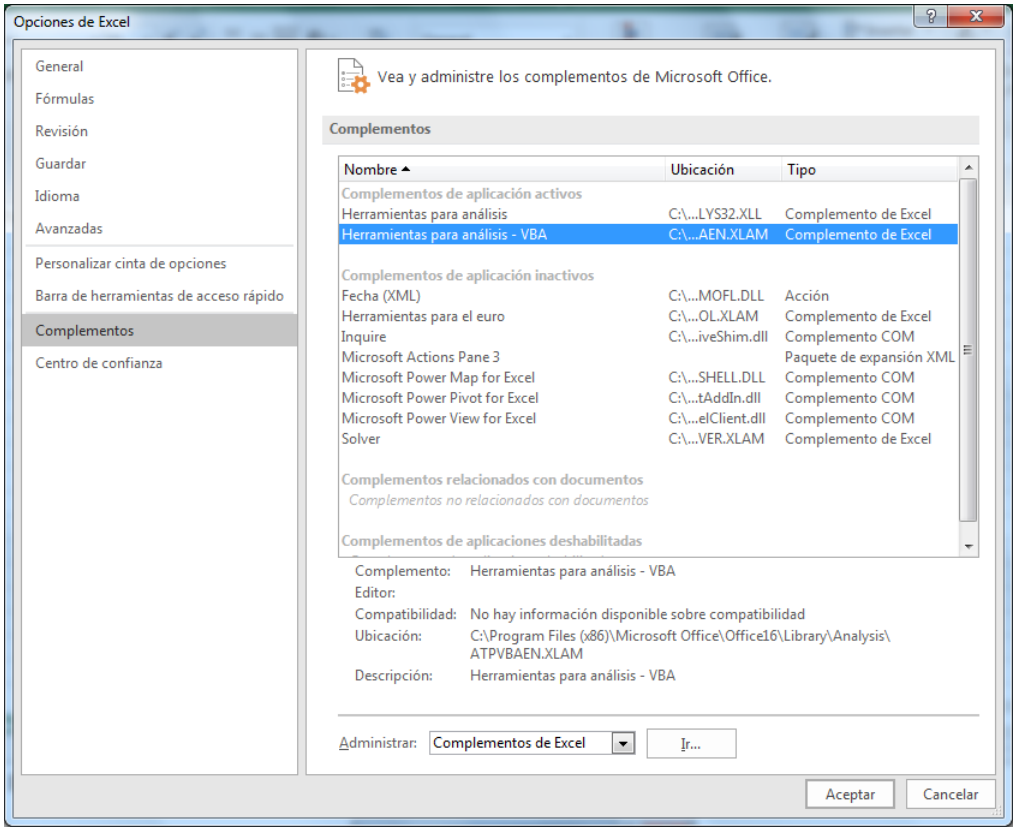
Si se quisiese cambiar el nombre de la hoja de cálculo *Hoja2* por el nombre *Descriptiva* (que identifica mejor el contenido de la hoja de cálculo), basta con que pulses dos veces con el ratón sobre el nombre *Hoja2* que aparece en la pestaña correspondiente. Verás que *Hoja2* pasa a negrita y entonces puedes escribir encima la palabra *Descriptiva*. A partir de ese momento, la hoja de cálculo *Hoja2* deja de tener ese nombre y pasa a denominarse *Descriptiva*.



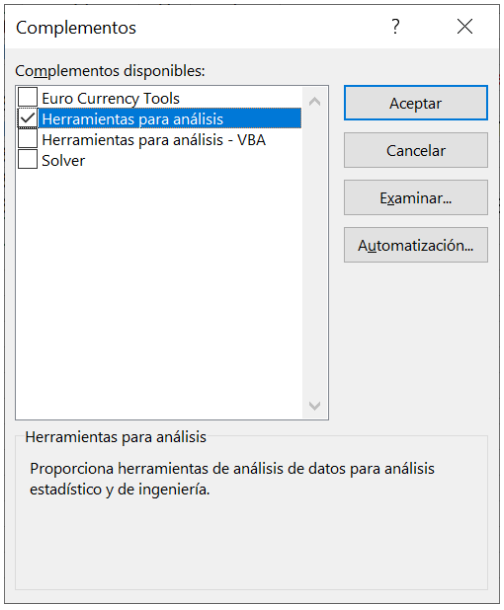
Para guardar las modificaciones que hemos realizado, así como los resultados obtenidos, puedes hacerlo pulsando el botón de *Guardar* en la parte superior izquierda de la ventana o a través del menú *Archivo*.

Para salir de *Excel*, basta con que presiones la **x** que aparece en la esquina superior derecha de la ventana de Excel.

NOTA 2.4: Si no tienes activa esta opción de *Análisis de datos* dentro del menú *Datos*, es porque no está activada o no está instalada esta herramienta. Para activarla o instalarla, tienes que elegir *Opciones* en el menú Archivo, y en el cuadro de diálogo que aparece seleccionar la opción *Complementos*. En el nuevo cuadro de diálogo que aparece, selecciona *Herramientas para el análisis – VBA* y presiona el botón *Ir…*



En el nuevo cuadro de diálogo, selecciona el complemento *Herramientas para análisis.*



3. R y R Commander

R es un lenguaje de programación especialmente indicado para realizar análisis estadísticos. A diferencia de la mayoría de los programas que solemos utilizar en nuestros ordenadores, que tienen interfaces tipo ventana, R es manejado a través de una consola en la que se introducen instrucciones y comandos propios de R para obtener los resultados deseados. R fue inicialmente escrito por Robert Gentleman y Ross Ihaka, miembros del Departamento de Estadística de la Universidad de Auckland en Nueva Zelanda. Actualmente R abarca una amplia gama de gráficos y técnicas estadísticas, que son el resultado del esfuerzo de miles de personas de todo el mundo que han colaborado en su desarrollo. 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, desde la que se puede acceder tanto a los archivos necesarios para su instalación como al resto de recursos del proyecto.

Para facilitar el uso de R sin necesidad de tener que conocer con detalle su propio lenguaje de comandos, John Fox, profesor de sociología de la Universidad de McMaster (Canadá), desarrolló en 2005 el paquete **Rcmdr**. Dicho paquete, también conocido como **R Commander**, es una adaptación de R mediante una interfaz tipo ventana como las que estamos acostumbrados a manejar en cualquier ordenador. R Commander cubre la mayor parte de los análisis estadísticos básicos más habituales, a través de los clásicos menús desplegables asociados a las interfaces tipo ventana, y es una forma amigable de manejar R sin necesidad de aprender su lenguaje de comandos o casi nada de él. R Commander puede descargarse desde la página principal del proyecto R: http://www.r-project.org, y para que funcione correctamente, debe estar instalado previamente en nuestro ordenador el software R.

4. Instalación de R y R Commander

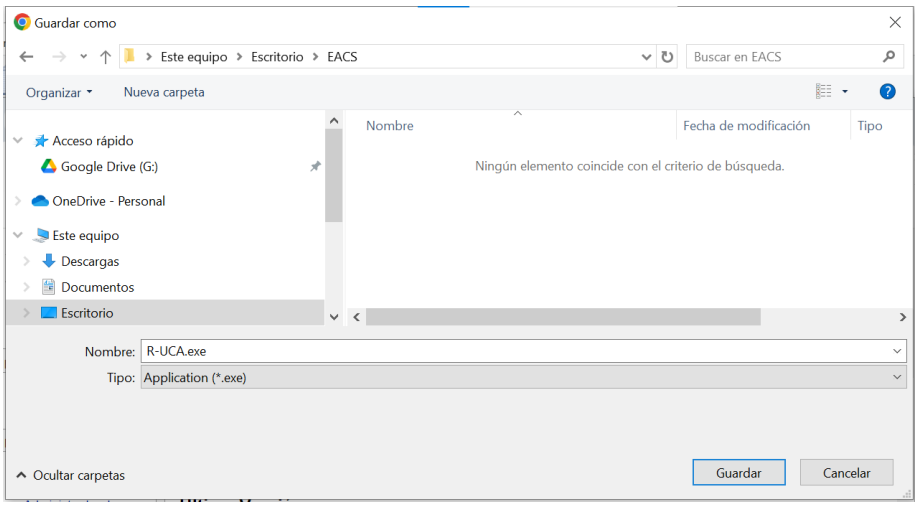
Tanto la instalación de R como de R Commander para cualquier plataforma (Windows, Linux, Mac, …) se puede realizar desde la página principal del proyecto http://www.r-project.org.

Sistemas operativos Windows.

Para sistemas operativos Windows, es mucho más sencillo sin embargo realizar la instalación de los dos paquetes usando el paquete R-UCA. El paquete *R-UCA para Windows* ha sido preparado por la Universidad de Cádiz y no es más que una recopilación de R y R Commander, junto con algunos paquetes adicionales de uso frecuente. Dicho paquete se puede obtener en la dirección: http://knuth.uca.es/R/doku.php?id=instalacion_de_r_y_rcmdr:r-uca.

Para instalar R-UCA en un sistema operativo Windows, los pasos que seguiríais son los siguientes.

3) Accede a la página web: http://knuth.uca.es/R/doku.php?id=instalacion_de_r_y_rcmdr:r-uca.
4) De entre las diferentes versiones de R-UCA, elige para descargar la misma versión que está instalada en la sala de ordenadores de las prácticas o, como alternativa, la **Última Versión**.
5) Al presionar con el ratón sobre el enlace **R-UCA**, aparece una ventana que permite elegir la carpeta o el directorio dónde guardar el archivo R-UCA.exe. Dicho archivo es el instalador del paquete R-UCA.



6) Al hacer doble click sobre el archivo recién descargado, se procede a la instalación del paquete R-UCA. Durante el proceso de instalación os aparecerán diferentes ventanas. Además de aceptar los términos de la licencia GNU en una de esas ventanas, conviene dejar por defecto las opciones que aparezcan en el resto de las ventanas. Si la instalación se ha realizado correctamente, aparecerá una carpeta llamada R en *Inicio>Programas*.

NOTA 4.1: Si se desea instalar cualquier otra versión de R-UCA (4.3.3, 4.3.1, 4.1.1, 4.0.5, 3.6.1, 3.5.0, …), los pasos a seguir son los mismos. Las diferencias de funcionamiento entre las diferentes versiones de R-UCA son mínimas, por lo que podéis trabajar con cualquiera de ellas.

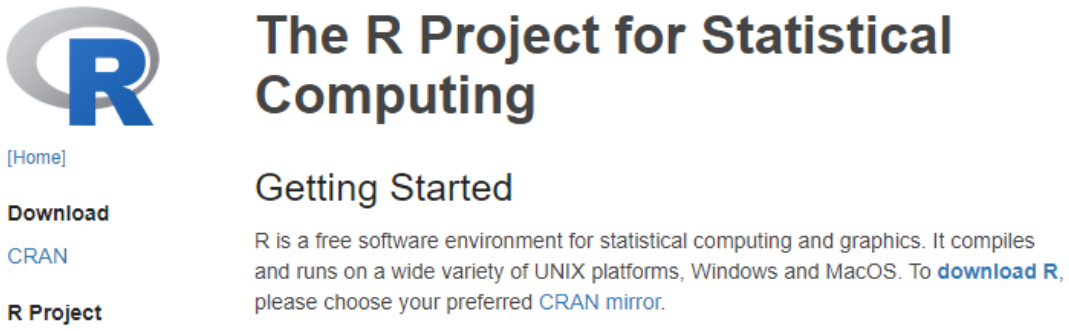
Sistemas operativos MacOS.

La instalación de R y R Commander en este tipo de sistemas es un poco más compleja, pues hay que instalar primero el paquete R, y luego, a través de la consola de R, instalar el paquete R Commander. Además, suele ser necesaria la instalación de algún paquete adicional, como XQuartz, que proporciona la información necesaria para generar la interfaz gráfica asociada a R Commander. En la web podéis encontrar varios vídeos explicativos, como, por ejemplo:

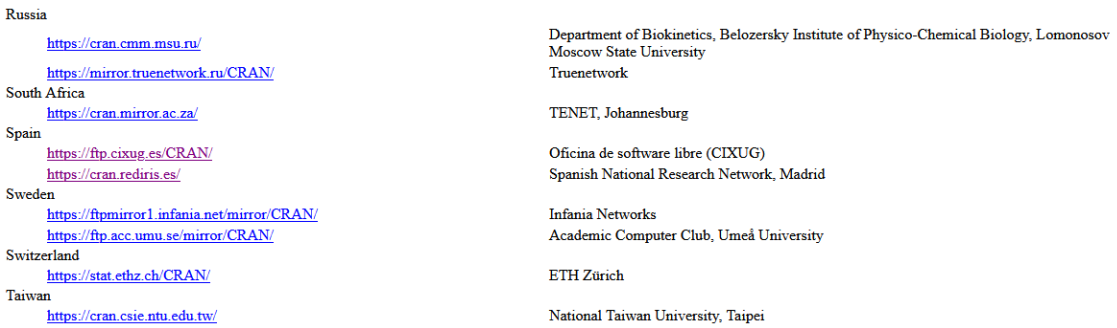
- https://www.youtube.com/watch?v=vrhJzBGXxjs (vídeo en el que se explica cómo instalar R en MacOS, dependiendo del tipo de procesador que lleve el MAC.)
- https://www.youtube.com/watch?v=e1xtjIAwj3s (video que explica cómo instalar R y R Commander, tanto en Windows como en MacOs.)

A modo de resumen, los pasos que seguiríais para instalar R y R Commander en un Mac son los siguientes.

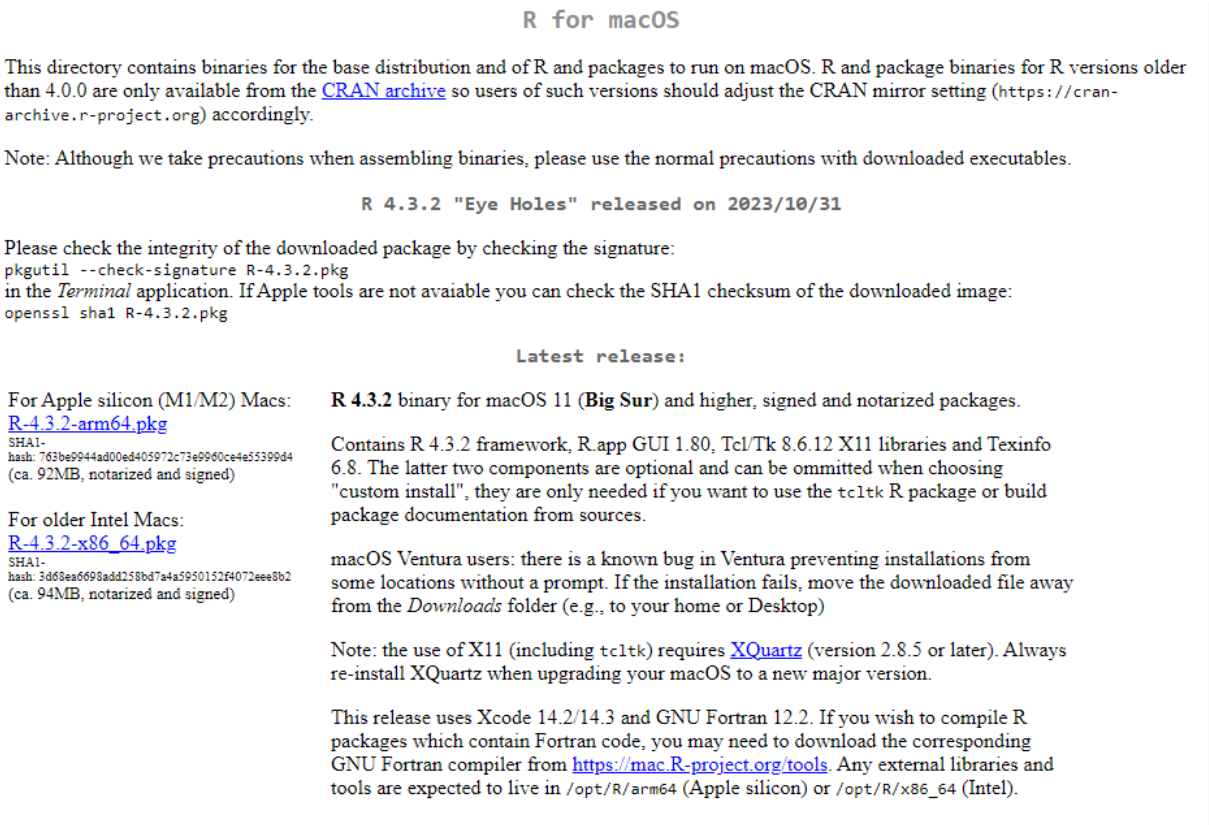
1) Accede a la página principal del proyecto R: http://www.r-project.org.
2) Pulsa sobre el enlace *Download R*, que aparece en las primeras líneas de esa página web, a la derecha.



3) Elige el "mirror" desde donde descargar el software. Lo normal es elegir uno de los "espejos" españoles.



4) En la página web que aparece, elige la opción *Download R for macOS*. Te aparecerá la ventana que aparece a continuación.



R for macOS

This directory contains binaries for the base distribution and of R and packages to run on macOS. R and package binaries for R versions older than 4.0.0 are only available from the CRAN archive so users of such versions should adjust the CRAN mirror setting (https://cran-archive.r-project.org) accordingly.

Note: Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

R 4.3.2 "Eye Holes" released on 2023/10/31

Please check the integrity of the downloaded package by checking the signature:
pkgutil --check-signature R-4.3.2.pkg
in the *Terminal* application. If Apple tools are not avaiable you can check the SHA1 checksum of the downloaded image:
openssl sha1 R-4.3.2.pkg

Latest release:

For Apple silicon (M1/M2) Macs:
R-4.3.2-arm64.pkg
SHA1-hash: 763be9944ad00ed405972c73e9960ce4e55399d4
(ca. 92MB, notarized and signed)

For older Intel Macs:
R-4.3.2-x86_64.pkg
SHA1-hash: 3d68ea6698add258bd7a4a5950152f4072eee8b2
(ca. 94MB, notarized and signed)

R 4.3.2 binary for macOS 11 (**Big Sur**) and higher, signed and notarized packages.

Contains R 4.3.2 framework, R.app GUI 1.80, Tcl/Tk 8.6.12 X11 libraries and Texinfo 6.8. The latter two components are optional and can be ommitted when choosing "custom install", they are only needed if you want to use the tcltk R package or build package documentation from sources.

macOS Ventura users: there is a known bug in Ventura preventing installations from some locations without a prompt. If the installation fails, move the downloaded file away from the *Downloads* folder (e.g., to your home or Desktop)

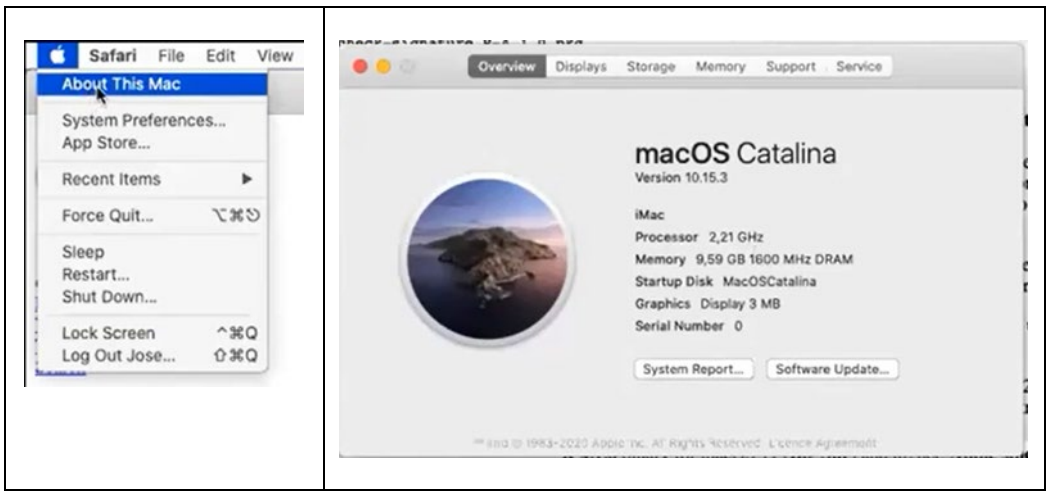
Note: the use of X11 (including tcltk) requires XQuartz (version 2.8.5 or later). Always re-install XQuartz when upgrading your macOS to a new major version.

This release uses Xcode 14.2/14.3 and GNU Fortran 12.2. If you wish to compile R packages which contain Fortran code, you may need to download the corresponding GNU Fortran compiler from https://mac.R-project.org/tools. Any external libraries and tools are expected to live in /opt/R/arm64 (Apple silicon) or /opt/R/x86_64 (Intel).

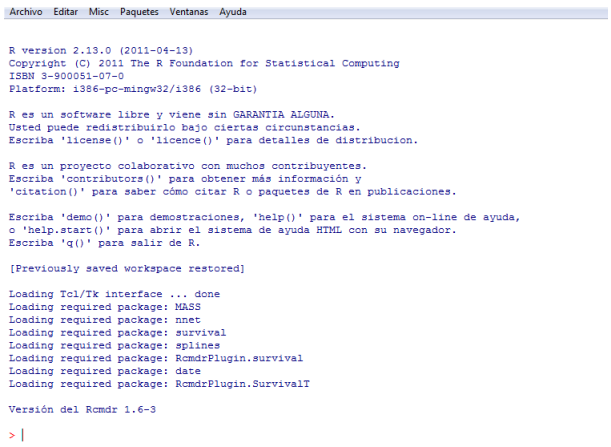
Observa que te aparecen dos posibilidades de descarga. El instalador que tendrás que descargar en tu Mac dependerá del tipo de procesador que lleve instalado tu Mac. Si tu Mac lleva instalado un procesador propio de Apple, el procesador Apple M1/M2, tendrás que descargarte el primer instalador, R-4.3.2-arm64.pkg. Si tu Mac lleva instalado un procesador Intel, tendrás que descargarte el segundo instalador, R-4.3.2-x86_64.pkg.

5) Descarga e instala la versión adecuada para tu Mac.
6) Descarga e instala en tu Mac el software XQuartz. Puedes descargar este software a través del enlace que aparece en la página web anterior, o través del enlace: https://www.xquartz.org/. Actualmente, la versión de XQuartz disponible es la versión 2.8.5, que puedes instalar en tu Mac a partir del instalador XQuartz-2.8.5.pkg (lo tendrás que descargar y ejecutar en tu Mac).
7) Tras la instalación de los dos paquetes, R y XQuartz, reinicia tu Mac.

NOTA 4.2: Una forma rápida de averiguar el tipo de procesador que lleva instalado tu Mac es elegir en la manzanita la opción "Acerca de …". En el ejemplo que aparece a continuación, el procesador instalado en el Mac es Intel, pues no aparecen a continuación de "Procesador" las palabras apple m1, o m2 o arm.



8) Ejecuta el software R en tu ordenador. Normalmente, el software lo encontrarás en la carpeta *Aplicaciones del Finder* y lo podrás identificar con el icono . Te aparecerá la siguiente ventana, que es la *consola* de R, o una similar.



9) Para instalar el software R Commander en tu Mac, escribe a continuación del símbolo > el comando: install.packages("Rcmdr", dependencies=TRUE) y pulsa la tecla "Intro".

Con el procedimiento anterior, ya tendrás instalados R y R Commander en tu Mac.

Una vez realizada la instalación, cada vez que quieras usar R Commander en tu Mac tendrás que hacer lo siguiente

1) Ejecuta el software R.
2) Desde la consola de R, ejecuta el comando: **Sys.setenv(LANG="es_ES.UTF-8").** Este comando lo que hace es pasar el entorno de trabajo de R del inglés al castellano.
3) Desde la consola de R, ejecuta el comando: **library(Rcmdr)**. Este comando lo que hace es abrir R Commander.

NOTA 4.3: Según cómo tengáis configurado vuestro sistema operativo, podríais observar una ralentización en la ejecución de R. Esto puede ser debido a que se tiene activada una opción del sistema operativo llamada "power nap", que hace funcionar el ordenador en modo "ahorro de energía". Para desactivar esa opción, podéis usar los ajustes del sistema operativo relacionados con la gestión del modo "ahorro de energía", o bien realizar lo siguiente. Abre R y en su consola ejecuta el comando *library(Rcmdr)* para abrir R Commander. En el menú *Tools* (*Herramientas*) de R Commander, elige la opción *Manage macOS app nap for R.app*. En el cuadro de diálogo que aparece, selecciona el botón que desconecta (off) la opción "app nat". Esta elección será permanente en las sesiones de R Commander hasta que la cambiéis.

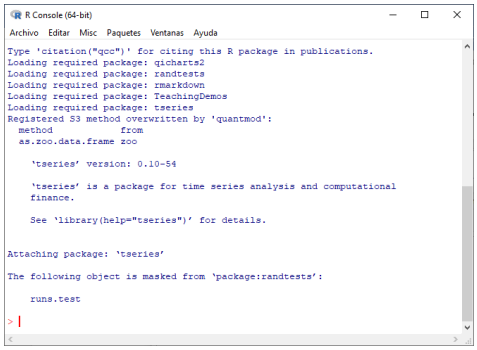
5. Ventanas de R y R Commander

En lo sucesivo, todas las explicaciones acerca del funcionamiento de R y R Commander se realizarán asumiendo que estamos trabajando en un entorno Windows. En entornos MacOS, las diferencias de funcionamiento son mínimas.

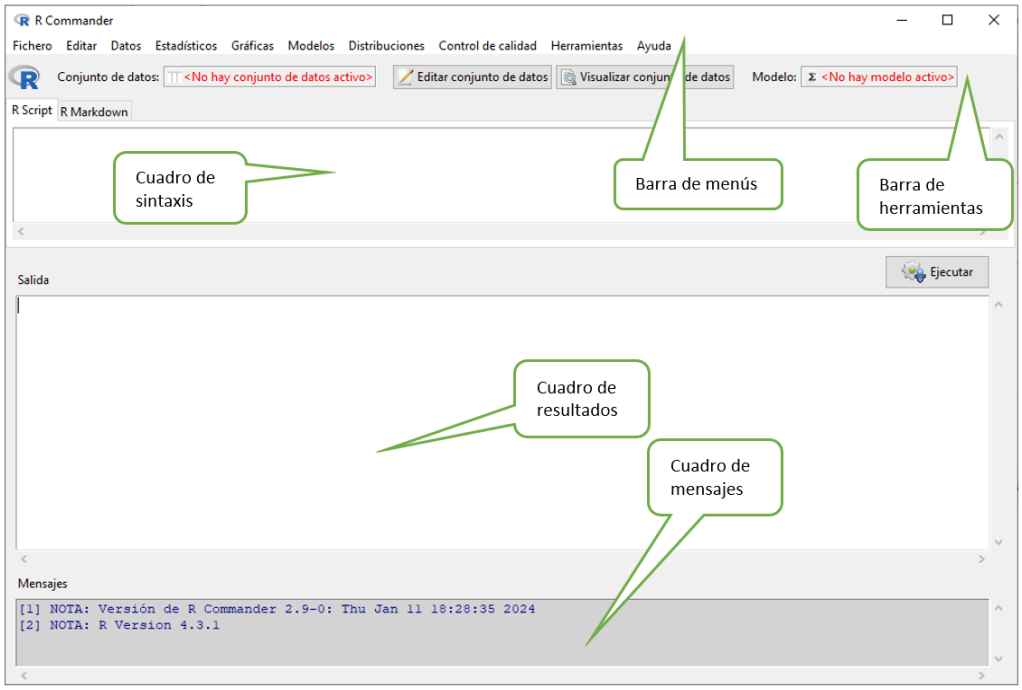
Para acceder a R y R Commander en sistemas operativos Windows, basta con seleccionar R desde el menú de inicio. Usualmente, el software aparece en la capeta R, dentro de *Inicio > Todos los Programas.*

Al ejecutar R, aparecen dos ventanas: una es la consola de R y la otra la ventana asociada a R Commander.

La consola de R tiene el aspecto que se muestra a continuación y es el lugar desde el que podríamos ejecutar cualquier comando de R, escribiéndolo directamente tras el símbolo >. Pero, de forma habitual, nosotros **no** la utilizaremos.



La segunda ventana es la asociada a R Commander, dispone de un sistema de menús y es la que utilizaremos de forma habitual. En esta pantalla podemos distinguir la *barra de menús*, la *barra de herramientas* y tres cuadros: el de *sintaxis*, el de *resultados* y el de *mensajes*.



Puede haber pequeñas diferencias en las imágenes si utilizamos otra versión de R Commander.

En la **barra de menús** encontrarás diferentes pestañas, con sus correspondientes desplegables con opciones, a través de los cuales podrás realizar todas las operaciones habituales. A modo de resumen, las principales opciones de los diferentes menús son las siguientes.

- **Fichero**: Permite cambiar el directorio de trabajo, y guardar las instrucciones y los resultados de una sesión de trabajo. Además, permite salir de la aplicación.
- **Editar**: Contiene las opciones habituales relacionadas con la edición: "Cortar", "Copiar", "Pegar", "Borrar", "Buscar...", "Seleccionar todo", "Deshacer", "Rehacer", "Limpiar ventana".
- **Datos**: Mediante las opciones de este menú se pueden cargar, editar y guardar datos. Además, se puede acceder a los datos de ejemplo que vienen con R. Otras opciones de este mismo menú permiten operaciones con los datos, como, por ejemplo: recodificación, tipificación, construcción de nuevas variables, ...
- **Estadísticos**: A través de este menú podemos acceder a las diferentes técnicas estadísticas disponibles: obtención de resúmenes descriptivos, obtención de tablas, ajuste de modelos, contrastes de hipótesis e intervalos de confianza, …
- **Gráficas**: Permite obtener diferentes tipos de gráficos.
- **Modelos**: En este menú, se puede acceder a una serie de utilidades para el análisis detallado de los modelos que se estén construyendo sobre un conjunto de datos.

- **Distribuciones**: Con las opciones de este menú, se pueden manejar diferentes tipos de variables aleatorias, así como obtener su función de distribución y sus cuantiles.
- **Control de calidad**: Procedimientos estadísticos que no utilizaremos en este curso.
- **Herramientas**: Mediante este menú se pueden cargar paquetes y plugins, y se puede configurar la interfaz.
- **Ayuda**: Muestra ayuda sobre R Commander.

En la **barra de herramientas** se muestra información sobre los datos y el modelo en uso. Los cuatro botones que aparecen son los siguientes.

- **Conjunto de datos**: Este botón muestra el nombre del conjunto de datos activo o la leyenda "No hay conjunto de datos activo" cuando todavía no se ha cargado o creado ningún conjunto de datos. Pulsando sobre este botón, se despliega un menú que permite activar cualquier conjunto de datos que esté disponible en ese momento.
- **Editar conjunto de datos**: Permite la edición del conjunto de datos activo en un entorno similar al de una hoja de cálculo. Durante la edición de los datos no es posible realizar ninguna otra operación con R. Por ello, es absolutamente imprescindible cerrar la ventana de edición de datos antes de intentar cualquier otra operación.
- **Visualizar conjunto de datos**: Muestra una ventana con los datos del conjunto de datos activo en formato similar al anterior. Esta ventana no permite la modificación de los datos, pero puede mantenerse abierta mientras se continúa haciendo operaciones.
- **Modelo**: La leyenda muestra el nombre del modelo activo o la leyenda "No hay modelo activo" cuando no se ha construido ningún modelo previamente. La pulsación sobre dicho botón permite la selección del modelo en uso de entre los disponibles.

Además de la barra de menús y la barra de herramientas, observamos tres cuadros. El primero es el **cuadro de sintaxis** o de instrucciones ***(R Script)***. Allí van apareciendo todos los comandos o instrucciones de R que se van ejecutando cuando realizamos las diferentes selecciones en los menús. Esto es algo que realiza R Commander automáticamente. Seleccionando cualquier instrucción de ese cuadro, se puede volver a ejecutar dicha instrucción pulsando el botón "*Ejecutar*".

En el **cuadro de resultados (*Salida)***, van apareciendo los resultados de las diferentes selecciones que vayamos realizando en los menús.

En el **cuadro de m*ensajes***, aparecen diferentes notas, avisos o errores. Los mensajes de este cuadro son importantes, pues nos permiten monitorizar los errores, para localizar su procedencia y subsanarlos.

Al lado de la pestaña *R Script*, hay otra pestaña llamada *R Markdown*. A través de esa pestaña, se pueden crear documentos, presentaciones e informes de R, en diferentes formatos (html, pdf, …), de una manera sencilla. Nosotros no vamos a utilizar esta herramienta. Los resultados que nos interesen los llevaremos directamente al Word para incluirlos en nuestros informes.

6. Manejo de ficheros de datos con R Commander

Con la opción *Fichero > Cambiar directorio de trabajo*, podemos elegir la carpeta donde, por defecto, R Commander irá a buscar archivos o los guardará. Si no cambiamos el directorio por defecto, cuando queramos abrir o guardar un fichero irá a la última carpeta utilizada y tendremos que buscar la carpeta que nos interese, si no es esa, en el árbol de carpetas.

R Commander permite crear conjuntos de datos y guardarlos en archivos con formato propio de R. La creación de un nuevo conjunto de datos se realiza a través de la opción *Datos > Nuevo conjunto de datos.* Podemos guardar este conjunto de datos como un archivo en nuestro pendrive o disco duro con la opción *Datos > Conjunto de datos activo > Guardar el conjunto de datos activo.*

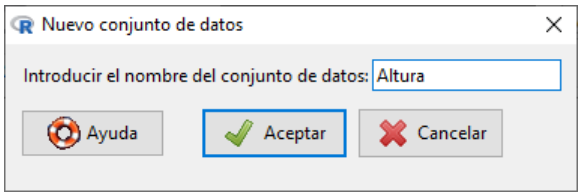
Aunque el editor de R Commander es bastante pobre y no suele utilizarse para crear conjuntos de datos, vamos a ver brevemente cómo crearlos. Supongamos que queréis crear un nuevo conjunto de datos conteniendo la información de la altura, en metros, de un grupo de chicos y chicas.

Chicos	Chicas
1.77	1.65
1.82	1.68
1.71	1.56
1.65	1.72
1.74	1.63
1.84	1.66
1.76	1.70
1.71	1.59
1.74	1.68

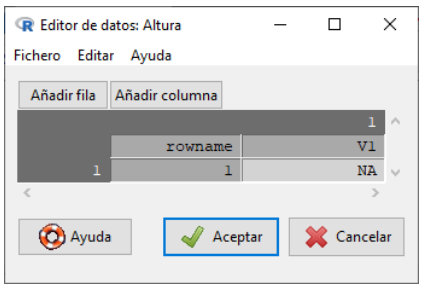
Antes de crear el fichero, nos planteamos cuántos individuos tiene la muestra y de qué variables disponemos. La estructura de cualquier conjunto de datos estadísticos es siempre la misma, cada columna representa los valores de una variable y cada fila los valores correspondientes a un individuo. El conjunto de datos que vais a crear tendrá que contener 2 variables (sexo y altura) y 18 casos (el total de chicos y chicas). Notad que la variable sexo es una variable nominal cualitativa, el sexo de la persona, y la variable altura es una variable numérica con decimales, la altura de la persona.

Una posible forma de crear el fichero de datos con R Commander sería la siguiente.

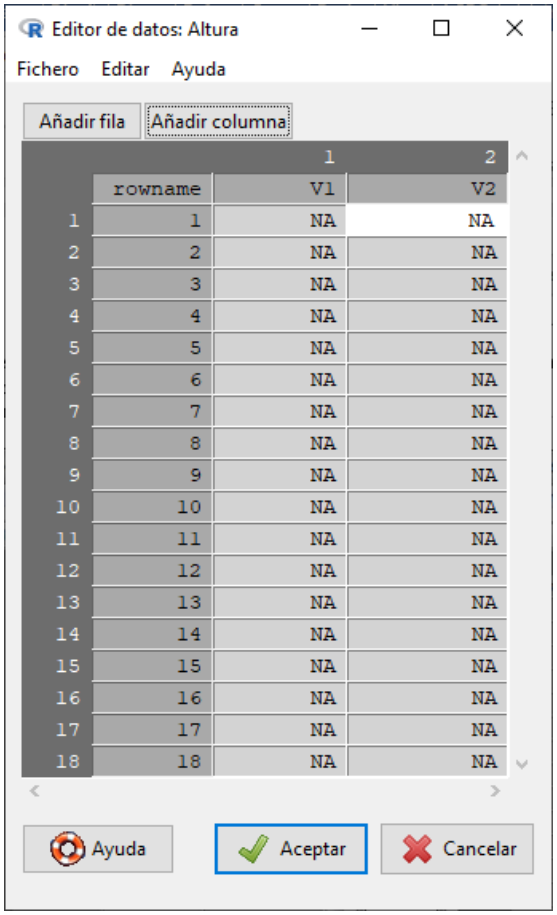
1) Desde el menú *Datos*, elige la opción *Nuevo conjunto de datos*. Te aparecerá la siguiente ventana, donde puedes dar nombre al conjunto de datos que vas a crear. En este ejemplo, estarías llamando *Altura* al nuevo conjunto de datos:

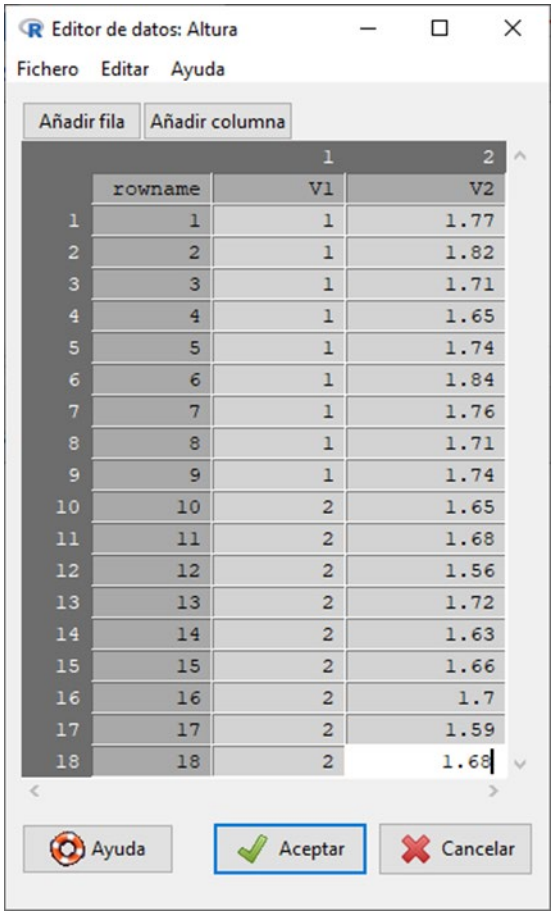


2) Al pulsar *Aceptar*, te aparece la siguiente ventana, que es el editor de R Commander. Por defecto, sólo aparece una fila y una columna.

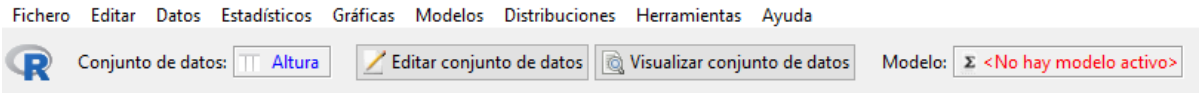


3) Como tenemos 18 casos y dos variables, pulsa 17 veces sobre el botón *Añadir fila* y una vez sobre el botón *Añadir columna*, para que el editor tenga el aspecto que se muestra a la izquierda de la siguiente figura. También se pueden añadir filas pulsando la tecla "*enter*" y columnas con la tecla de tabulación. Observa, a la izquierda en la imagen siguiente, que todas las celdas llevan el valor NA (del inglés, Not Available). Para R Commander, NA son "datos perdidos" o "datos desconocidos". En este caso indican que todavía no hemos introducido ningún dato.
4) Con el ratón y/o las flechas, muévete por el editor para modificar los valores de las variables (valores NA). En la primera columna (V1) habrá que introducir 1 cuando se trate de un chico y 2 cuando se trate de una chica. En la segunda columna (V2) habrá que introducir las alturas. Después de introducir los 18 casos, el editor tendrá el aspecto que se muestra a la derecha de la imagen siguiente. Observa que el punto es el indicador de la parte decimal en R.
5) Si te has confundido y has añadido alguna fila o columna de más, la pestaña *Editar* te permitirá borrar la fila o la columna seleccionada.





6) Por defecto, los nombres que se dan a las variables son V1, V2, … Para cambiarles el nombre no hay más que colocarse en la casilla correspondiente y escribir el nombre que se desee. Así, si te colocas sobre la celdilla que pone V1 y escribes la palabra *Sexo*, estarás llamando Sexo a la primera de las variables. Si tecleas *Altura* en la celdilla correspondiente a V2 estarás llamando Altura a la segunda variable.
7) Dentro del menú *Fichero*, elige la opción *Salir y guardar*, o pulsa directamente el botón *Aceptar* en la ventana del *Editor*. De esta forma, habrás creado el conjunto de datos *Altura* que contiene dos variables: *Sexo* y *Altura*. Observa cómo en la barra de herramientas, en el campo *Conjunto de datos*, aparece *Altura*, que es el conjunto de datos que acabas de crear y que es el que se encuentra activo en R Commander para realizar cualquier operación a partir de este momento.

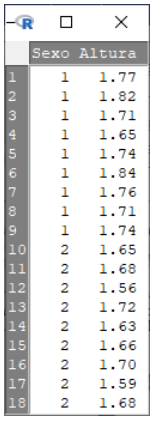


Cuando se manejan datos y variables en el editor, conviene tener en cuenta las siguientes observaciones.

- En el menú *Editar* del editor de R Commander encontrarás diferentes opciones que te permitirán añadir o borrar filas y columnas, así como copiar y pegar el contenido de las celdas.
- R Commander maneja dos tipos de variables: *numéricas* y *factores*. Las *numéricas* se corresponderían con las variables estadísticas cuantitativas, y los *factores* con las variables estadísticas cualitativas. Mientras que las variables de tipo numérico sólo pueden contener números, los factores pueden contener cualquier texto, incluidos blancos (aunque no es recomendable).
- La naturaleza de las variables estadísticas no determina la forma en la que las podemos guardar en el ordenador como variables de R. En muchos casos, conviene utilizar variables numéricas (de R) para codificar los niveles de una variable estadística cualitativa. Es lo que hemos hecho por ejemplo al crear la variable *Sexo* en R. La hemos creado numérica con valores 1 y 2, cuando la variable estadística Sexo es en realidad cualitativa. Veremos más adelante cómo convertir variables numéricas en variables de tipo factor y viceversa.

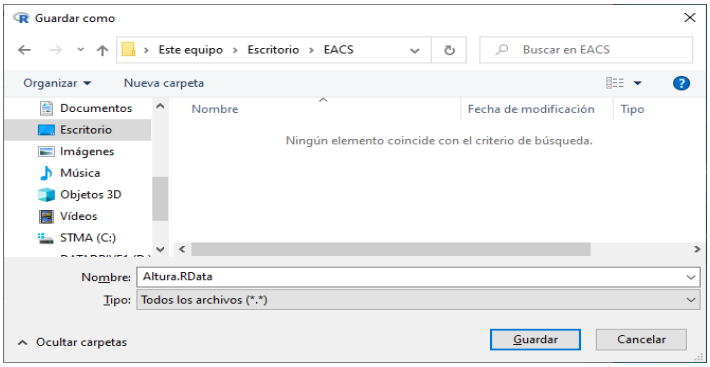
- La columna *rowname* hace referencia a las "etiquetas" de los casos. Así, si los nombres de nuestros individuos hubiesen sido Luis, Juan, …, podíamos haber utilizado esa columna rowname para etiquetar los casos y ponerles etiquetas Luis, Juan, … en lugar de 1, 2, …
- R Commander es sensible a mayúsculas y minúsculas. No es lo mismo usar altura como nombre de una variable que usar Altura.
- Para evitar posibles problemas en el correcto funcionamiento de R Commander, se recomienda dejar cerrada la ventana asociada al editor, una vez finalizada la edición de datos.

Para volver al editor y modificar los datos, basta con pulsar el botón *Editar conjunto de datos* que se encuentra en la barra de herramientas. En esa misma barra se encuentra el botón *Visualizar conjunto de datos*, que nos permite ver el contenido de nuestro conjunto de datos sin posibilidad de modificarlos. Si presionamos dicho botón, R Commander nos mostraría la siguiente ventana.



	Sexo	Altura
1	1	1.77
2	1	1.82
3	1	1.71
4	1	1.65
5	1	1.74
6	1	1.84
7	1	1.76
8	1	1.71
9	1	1.74
10	2	1.65
11	2	1.68
12	2	1.56
13	2	1.72
14	2	1.63
15	2	1.66
16	2	1.70
17	2	1.59
18	2	1.68

El conjunto de datos que acabáis de crear se encuentra únicamente en la memoria de R Commander. Para guardar los datos en nuestro ordenador, habrá que elegir la opción *Datos > Conjunto de datos activo > Guardar el conjunto de datos activo*. En la ventana que aparece, podemos elegir la carpeta donde guardar los datos y el nombre para nuestro archivo. En el ejemplo, cuando presionemos el botón *Guardar*, estaremos guardando los datos en un fichero que se va a llamar *Altura.RData*, dentro de la carpeta *EACS* del Escritorio.

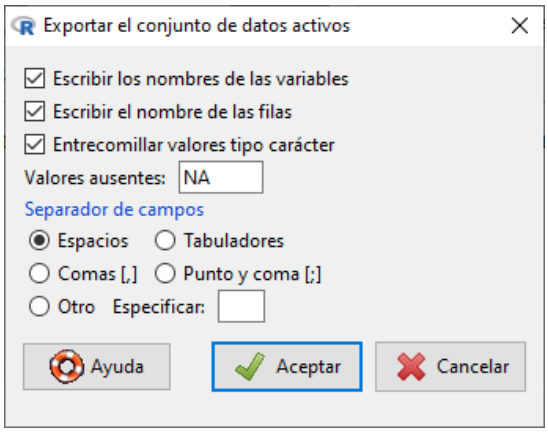


Por defecto, la extensión que R Commander añade al fichero que se va a guardar es .RData. R Commander identifica además cualquier fichero con alguna de estas extensiones: .rda, .Rda o .RDA, como un archivo que contiene conjuntos de datos creados con R Commander.

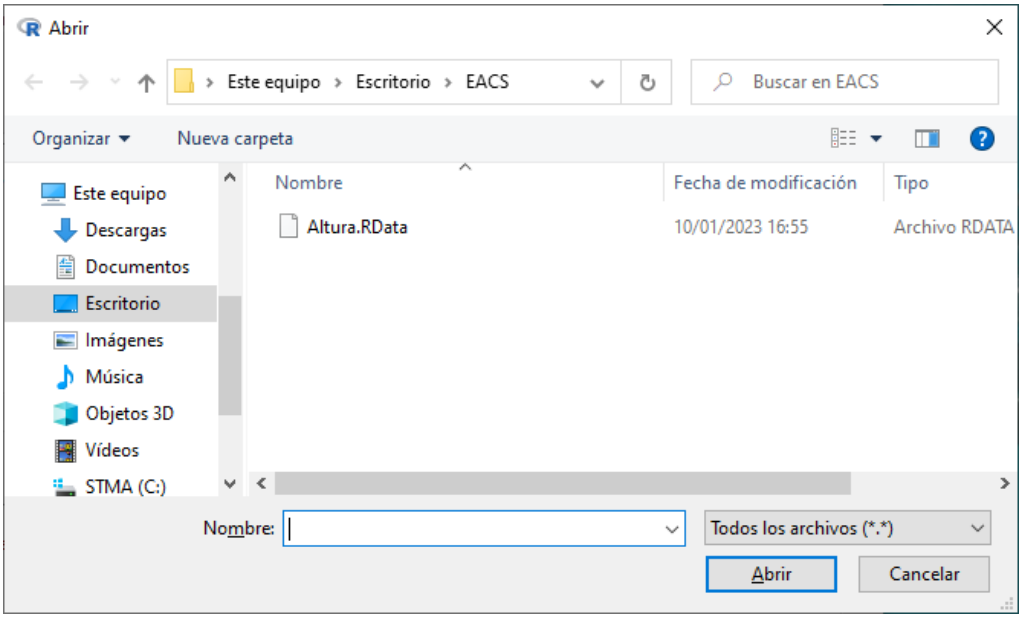
No hay que confundir el nombre con el que se guarda fichero (en el ejemplo *Altura.RData*) con el nombre interno del conjunto de datos (en el ejemplo es el mismo: *Altura*). El nombre "interno" es el que usa R Commander en sus opciones y menús y puede no coincidir con el nombre del fichero externo. Os recomendamos que uséis el mismo nombre para evitar confusiones.

Además de guardar los datos con formato propio de R Commander, podíamos haber guardado los datos en formato texto para importarlos en otras aplicaciones y trabajar los datos con ellas. Para generar estos ficheros de texto, bastaría con

elegir la opción *Datos* > *Conjunto de datos activo* > *Exportar el conjunto de datos activo* y rellenar convenientemente el correspondiente cuadro de diálogo.



Reseñar por último que la forma de abrir ficheros o archivos generados por R Commander y que contengan conjuntos de datos es a través de la opción *Datos* > *Cargar conjunto de datos*. Si seleccionamos esa opción y elegimos la carpeta *EACS*, podríamos cargar en R Commander el conjunto de datos que acabamos de crear y que tenemos guardado en el archivo *Altura.RData*.

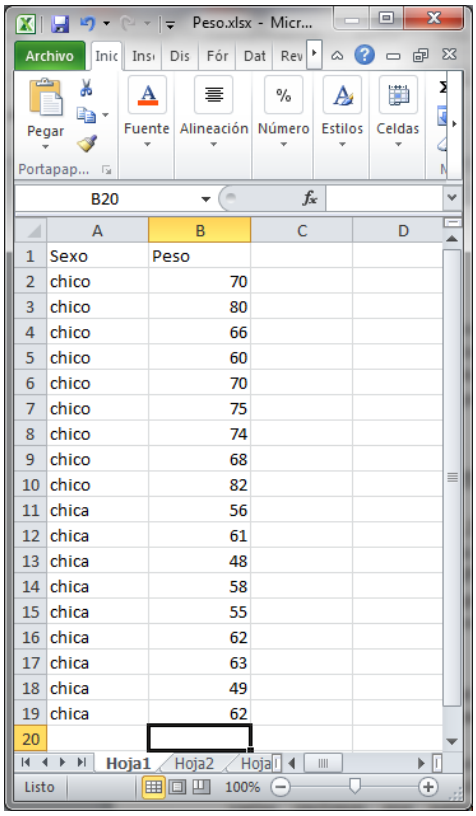


NOTA 6.1: La mayoría de los archivos de Windows se abren al hacer doble click sobre ellos con el ratón. Así, al hacer doble click sobre un fichero de tipo Word, automáticamente se abre dicho fichero con la aplicación MS-Word. Eso **podría no suceder** con los archivos de datos generados con R Commander. La forma de abrir los ficheros de datos generados con R Commander es la señalada anteriormente, a través de la opción *Datos* > *Cargar conjunto de datos.*

7. Importación de conjuntos de datos con R Commander

Como ya se ha comentado anteriormente, a través de la opción *Datos* > *Cargar conjunto de datos*, podemos cargar en R Commander cualquier archivo de datos propio de R que tengamos en nuestro ordenador. Además de esa opción, R Commander tiene la posibilidad de cargar datos generados con otras aplicaciones, como Excel, SPSS, Stata o SAS, para poder trabajar dichos datos posteriormente desde R Commander. Como ejemplo, vamos a importar los datos que se encuentran en el fichero *Peso.xlsx*, que es un fichero creado con Excel y que contiene los pesos de nuestro conjunto de

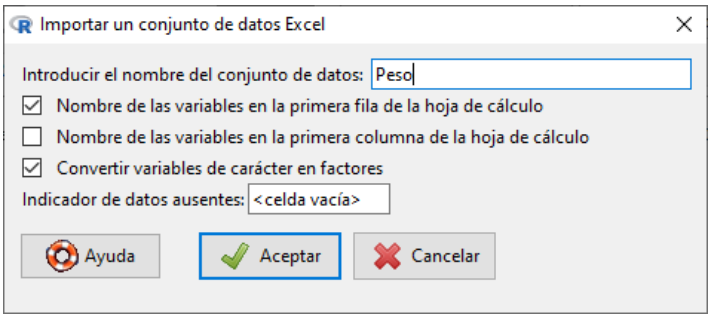
chicos y chicas. Observa que los nombres de las variables aparecen en la primera fila y que los datos se encuentran en la primera hoja Excel (Hoja 1).



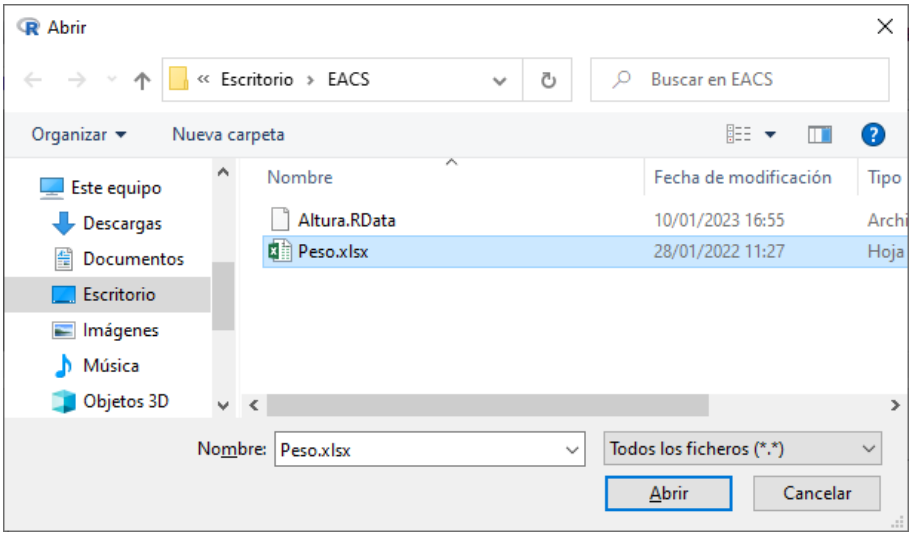
	A	B	C	D
1	Sexo	Peso		
2	chico	70		
3	chico	80		
4	chico	66		
5	chico	60		
6	chico	70		
7	chico	75		
8	chico	74		
9	chico	68		
10	chico	82		
11	chica	56		
12	chica	61		
13	chica	48		
14	chica	58		
15	chica	55		
16	chica	62		
17	chica	63		
18	chica	49		
19	chica	62		
20				

Para importar estos datos, los pasos que seguirías son los siguientes.

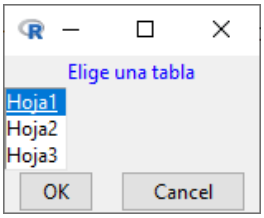
1) Dentro del menú *Datos*, elige la opción *Datos > Importar datos > desde un archivo de Excel*.



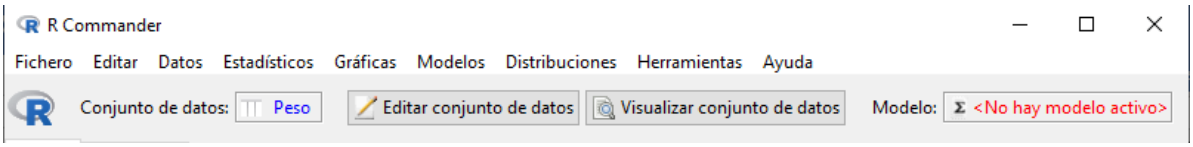
2) En el cuadro de diálogo que aparece, llama *Peso* al conjunto de datos que se va a crear en R Commander. Observa que va a tomar de la primera línea de la hoja Excel donde se encuentran los datos los nombres de las variables y que las variables de Excel que contengan caracteres las va a transformar en variables de tipo factor.
3) Al pulsar *Aceptar*, te aparecerá el siguiente cuadro de diálogo donde elegirás el archivo Excel que vas a importar. En el ejemplo, estarás importando el archivo Excel *Peso*, que se encuentra en la carpeta *EACS*.



4) Al pulsar *Abrir*, te aparece un cuadro de diálogo que te permite elegir la hoja de cálculo dentro del libro Peso que contiene los datos a importar. En nuestro caso, la primera hoja (Hoja1).

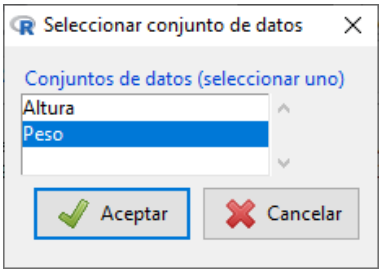


5) Tras pulsar *OK*, R Commander importa los datos contenidos en el archivo Excel *Peso* y crea el conjunto de datos *Peso*, que es el que queda activo en R Commander para trabajar.



Para guardar los datos de los pesos en un archivo propio de R Commander, basta elegir la opción *Datos* > *Conjunto de datos activo* > *Guardar el conjunto de datos activo*, como hemos hecho antes para guardar las alturas.

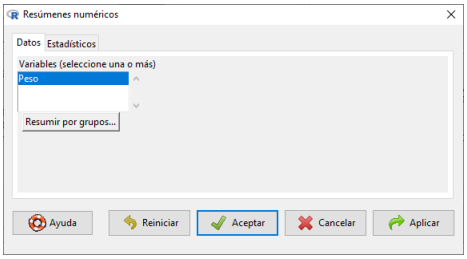
En R Commander podemos tener varios conjuntos de datos abiertos a la vez para poder trabajar con ellos, pero sólo podemos tener uno de ellos activo. De hecho, en este momento, R Commander tiene dos conjuntos de datos abiertos: el que hemos creado (*Altura*) y el que hemos importado (*Peso*). Para cambiar de un conjunto de datos a otro, basta con pulsar sobre el botón asociado a *Conjunto de datos* y elegir el conjunto de datos que queramos activar. Todo lo que se haga con R Commander (una transformación, una tabla de frecuencias, un gráfico, …) sólo afecta al conjunto de datos activo.



8. Los cuadros de sintaxis y resultados. La ventana de gráficos

Al realizar cualquier operación con R Commander, automáticamente se genera el código R asociado a esa operación. Dicho código siempre aparece en el *cuadro de sintaxis*. Así, cuando hemos guardado los datos de las alturas, se ha generado el comando "*save*", o cuando hemos importado los datos de los pesos desde el archivo Excel se ha generado el comando "*readXL*". Vamos a realizar un par de operaciones más para ver qué código R se genera y qué resultados se obtienen.

Con el conjunto de datos *Peso* activo, desde el menú *Estadísticos*, elige *Estadísticos* > *Resúmenes* > *Resúmenes numéricos*... En la ventana que aparece, elige la variable *Peso* (si no está seleccionada por defecto) y pulsa *Aceptar*.



Como resultado, verás en el cuadro de sintaxis el siguiente código R:

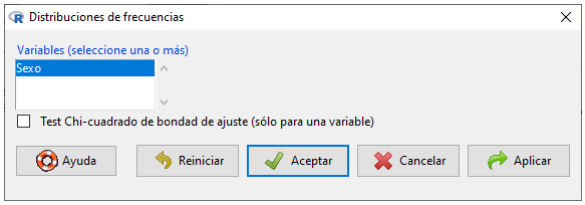
```
> numSummary(Peso[,"Peso", drop=FALSE], statistics=c("mean", "sd", "quantiles", "CV"), quantiles=c(0,.25,.5,.75,1))
```

y en la ventana de resultados verás el mismo código R y el resultado de la ejecución de dicho código.

```
> numSummary(Peso[,"Peso", drop=FALSE], statistics=c("mean", "sd", "quantiles", "CV"), quantiles=c(0,.25,.5,.75,1))
     mean       sd  IQR 0%  25%  50% 75% 100%  n
 64.38889 9.659734 11.5 48 58.5 62.5  70   82 18
```

En la siguiente práctica veremos las opciones que esta ventana de sintaxis proporciona para calcular estadísticos.

Para calcular la tabla de frecuencias de una variable factor podemos usar *Estadísticos* > *Resúmenes* > *Distribución de frecuencias*. En la ventana que aparece, elige (si no está seleccionada por defecto) la variable *Sexo* y pulsa *Aceptar*.



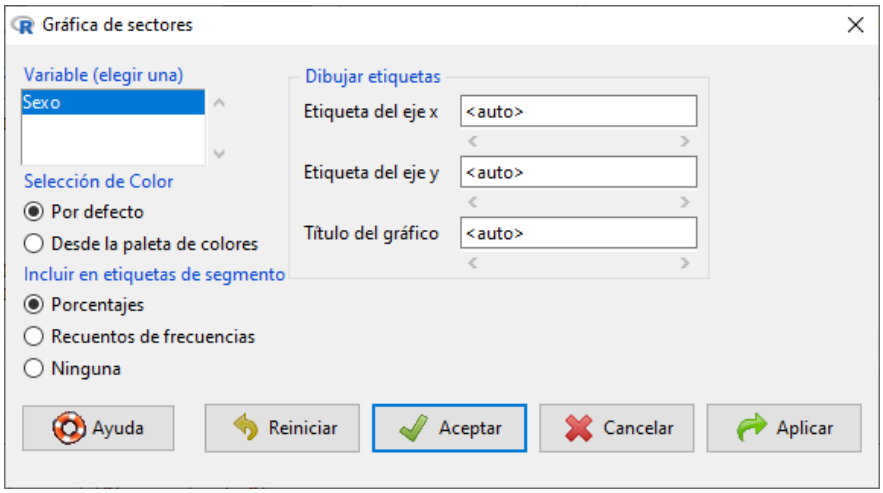
Como resultado, verás en el cuadro de sintaxis el código R generado, que se corresponde con la obtención de tablas de frecuencias. En el cuadro de resultados, además de escribir, en rojo, el código R que genera las tablas de frecuencias, verás, a continuación, las tablas de frecuencias absoluta y relativa (en porcentaje) asociadas a la variable *Sexo*: hay un 50% de chicos en nuestro conjunto de datos *Peso* (9 chicos) y otro 50% de chicas (9 chicas).

```
> local({
+   .Table <- with(Peso, table(Sexo))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table), 2))
+ })

counts:
Sexo
chica chico
    9     9

percentages:
Sexo
chica chico
   50    50
```

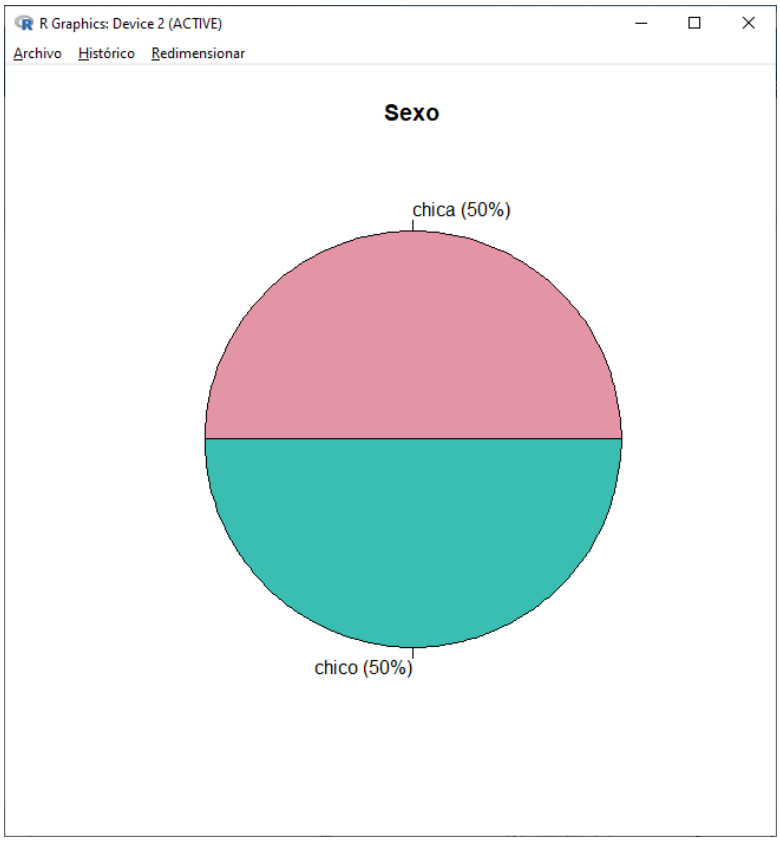
Para obtener un gráfico que represente esta distribución de frecuencias podemos usar, dentro del menú *Gráficas*, la opción *Gráficas* > *Gráfica de sectores*. En la ventana que aparece, elige la variable *Sexo* y pulsa *Aceptar*.



Con esta opción, lo que hemos conseguido es un gráfico de sectores de la variable *Sexo*. En el cuadro de sintaxis y en el de resultados aparece el código R asociado a la obtención del gráfico, basado en el comando "*piechart*":

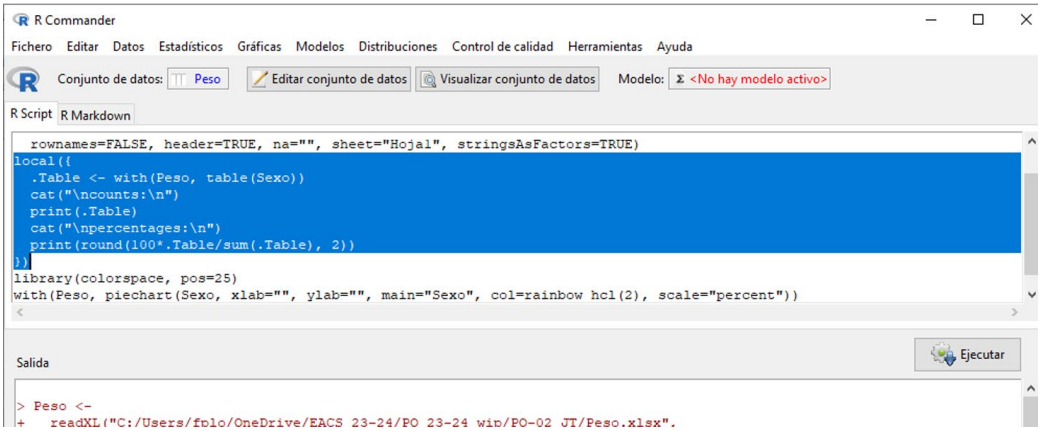
```
library(colorspace, pos=25)
with(Peso, piechart(Sexo, xlab="", ylab="", main="Sexo", col=rainbow_hcl(2), scale="percent"))
```

y el gráfico va a parar a una ventana nueva, llamada **ventana de gráficos**. Más adelante veremos cómo manejar los gráficos.



Cualquiera de los comandos generados (o cualquier código R) se puede volver a ejecutar directamente desde el cuadro de sintaxis, sin necesidad de tener que elegir opciones en los menús y rellenar los cuadros de diálogo correspondientes. Por ejemplo, si quisiésemos volver a obtener las tablas de frecuencias de la variable *Sexo* sin usar los menús, bastaría con seleccionar con el ratón el código asociado a la obtención de las tablas de frecuencias que se encuentra en el cuadro de

sintaxis, y pulsar el botón *Ejecutar*. Al realizar esta operación, verás que en el cuadro de resultados vuelven a aparecer las tablas de frecuencias de la variable *Sexo*.



El conjunto de instrucciones del cuadro de sintaxis se puede guardar en un fichero, para poder utilizar dicho conjunto de instrucciones en una sesión posterior de R Commander. Guardar o abrir ficheros de instrucciones se puede realizar a través del menú *Fichero*.

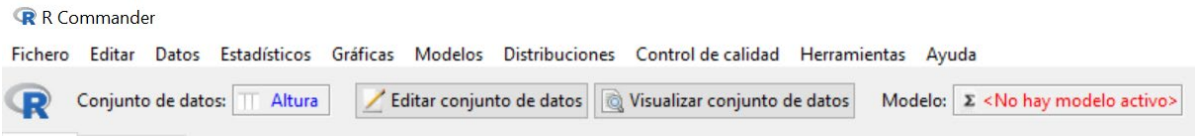
9. Algunas transformaciones de variables

Por diferentes motivos, en muchas ocasiones resulta necesario transformar nuestros datos para obtener nuevas variables que nos faciliten el análisis de nuestros datos. La opción *Modificar variables del conjunto de datos activo*, que se encuentra dentro del menú *Datos,* nos permitirá realizar diferentes tipos de sencillas transformaciones. Vamos a realizar algunas de ellas.

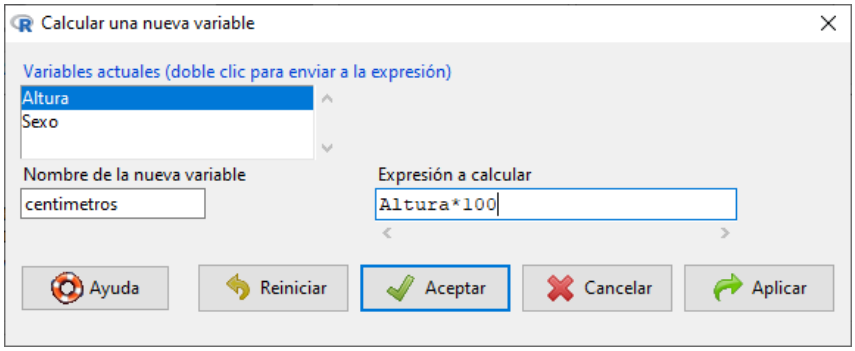
Calcular una nueva variable

Con los datos contenidos en el conjunto de datos *Altura*, crearemos una nueva variable llamada *centimetros* que será la altura en metros multiplicada por 100. Los pasos para llevar a cabo esta transformación son:

1) Asegúrate de que el conjunto de datos que se encuentra activo en R Commander es el conjunto de datos *Altura*. Recuerda que para activarlo hay que pulsar sobre el botón asociado a *Conjunto de datos* y seleccionarlo.



2) Selecciona *Datos* > *Modificar variables del conjunto de datos activo* > *Calcular una nueva variable*, desde el menú *Datos*.
3) En la casilla *Nombre de la nueva variable* teclea *centimetros*.
4) Pulsa dos veces con el ratón sobre la variable *Altura* para que dicha variable pase a la casilla *Expresión a calcular*.
5) Añade *100 para que la *Expresión a calcular* sea Altura*100. El cuadro de diálogo te habrá quedado de la siguiente manera.



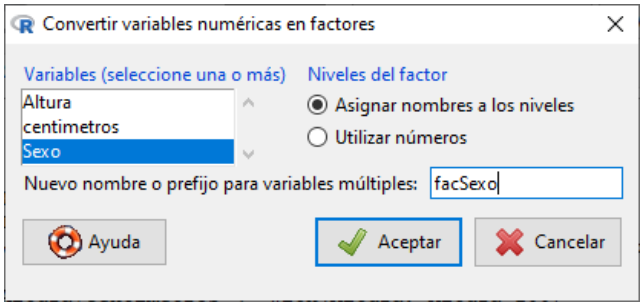
6) Pulsa *Aceptar* para realizar la transformación.

Observa, usando *Visualizar conjunto de datos*, cómo aparece la nueva variable *centimetros*, que ha sido calculada para todos los casos. Es importante destacar que, una vez hecha una transformación, **los datos de la nueva variable creada permanecen inalterados, aunque se modifiquen los datos originales usados en la transformación**.

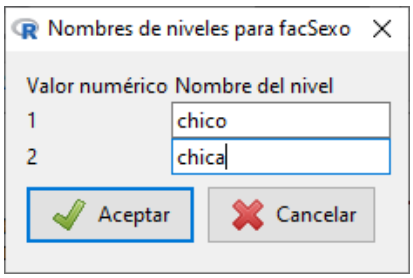
Convertir variable numérica en factor

Si quisiéramos obtener una tabla de frecuencias (o un gráfico de sectores) de la variable *Sexo* del conjunto de datos *Altura*, veríamos que no podemos. Eso es porque, por defecto, R Commander sólo permite realizar tablas de frecuencias de variables de tipo factor, y nosotros hemos creado la variable *Sexo* de tipo numérico. Para obtener la tabla o el gráfico de esa variable, hay que reconvertirla a una variable de tipo factor. Los pasos que seguirías para realizar esta reconversión serían los siguientes.

1) Selecciona en el menú principal *Datos* > *Modificar variables del conjunto de datos activo* > *Convertir variable numérica en factor.*
2) Rellena el cuadro de diálogo como se muestra a continuación. Vas a crear una nueva variable de tipo factor, llamada *facSexo*, a partir de la variable *Sexo*. Al elegir *Asignar nombre a los niveles*, podrás elegir las etiquetas (por ejemplo "chico" y "chica") para los niveles de la variable (1 y 2).



3) Al pulsar *Aceptar*, te aparece el siguiente cuadro de diálogo, donde puedes especificar las etiquetas para cada una de las modalidades de la variable *Sexo*. En este caso, estás asignado "chico" a todos los casos cuyo valor sea 1 en la variable *Sexo* y "chica" a todos los casos cuyo valor sea 2 en la variable *Sexo*.



Después de *Aceptar*, habrás creado una nueva variable de tipo factor llamada *facSexo*, que contendrá las etiquetas "chico" y "chica" como sus niveles o categorías. Ahora podrás obtener tablas de frecuencias y gráficos de barras o sectores para la variable estadística Sexo usando la variable de tipo factor *facSexo*.

10. Resumen del conjunto de datos

R Commander lleva incorporada una opción que permite obtener un conjunto de estadísticos básicos para todas las variables de un conjunto de datos. Dicha opción es *Estadísticos* > *Resúmenes* > *Conjunto de datos activo*. Aplicada esta opción sobre el conjunto de datos *Altura*, los resultados que se obtienen son los siguientes.

```
> summary(Altura)
      Sexo          Altura        centimetros     facSexo
 Min.   :1.0   Min.   :1.560   Min.   :156.0   chico:9
 1st Qu.:1.0   1st Qu.:1.653   1st Qu.:165.2   chica:9
 Median :1.5   Median :1.705   Median :170.5
 Mean   :1.5   Mean   :1.701   Mean   :170.1
 3rd Qu.:2.0   3rd Qu.:1.740   3rd Qu.:174.0
 Max.   :2.0   Max.   :1.840   Max.   :184.0
```

Para las variables numéricas, R Commander calcula el mínimo, el máximo, la media y los cuartiles. Para las variables de tipo factor, muestra la distribución de frecuencias absolutas. Observa que, como hemos definido la variable *Sexo* de tipo numérico, R Commander calcula su máximo, su mínimo, etc., medidas que no tienen sentido para una variable estadística cualitativa. Por el contrario, para la variable *facSexo*, R Commander obtiene la tabla de frecuencias, pues se ha definido como variable de tipo factor.

Cuando el número de variables del conjunto de datos es grande, la salida anterior es excesivamente voluminosa y resulta difícil ver en conjunto cómo se comportan las variables. Por ejemplo, esta sería la salida asociada al conjunto de datos *pulso,* que manejaremos a lo largo de las clases prácticas.

```
> summary(pulso)
     Altura           Peso             Edad            Sexo       Fumador Alcohol
 Min.   : 68.0   Min.   : 27.00   Min.   :18.00   Min.   :1.000   no:99   no:42
 1st Qu.:165.2   1st Qu.: 56.25   1st Qu.:19.00   1st Qu.:1.000   sí:11   sí:68
 Median :172.5   Median : 63.00   Median :20.00   Median :1.000
 Mean   :171.6   Mean   : 66.33   Mean   :20.56   Mean   :1.464
 3rd Qu.:180.0   3rd Qu.: 75.00   3rd Qu.:21.00   3rd Qu.:2.000
 Max.   :195.0   Max.   :110.00   Max.   :45.00   Max.   :2.000

    Ejercicio  Correr      Pulso1           Pulso2           Año
 alta    :14   no:64   Min.   : 47.00   Min.   : 56.0   Min.   :93.00
 baja    :37   sí:46   1st Qu.: 68.00   1st Qu.: 72.0   1st Qu.:95.00
 moderada:59           Median : 76.00   Median : 84.0   Median :96.00
                       Mean   : 75.69   Mean   : 96.8   Mean   :95.63
                       3rd Qu.: 82.00   3rd Qu.:125.0   3rd Qu.:97.00
                       Max.   :145.00   Max.   :176.0   Max.   :98.00
                       NA's   :1        NA's   :1
```

En esos casos, se recomienda crear tablas personalizadas en Word, conteniendo parte de las variables y/o parte de la información. Por ejemplo, podríamos separar las variables de tipo numérico de las de tipo factor, e incluir sólo algunas de las variables y algunos de los estadísticos en las tablas. A continuación, se muestran un par de ejemplos de dichas tablas.

En la primera sólo se ha incluido información del máximo, mínimo, media y mediana para las variables *Altura, Peso, Edad, Pulso1* y *Pulso2*.

	Mínimo	Media	Mediana	Máximo
Altura	68	171.6	172.5	195
Peso	27	66.33	63	110
Edad	18	20.56	20	45
Pulso1	47	75.69	76	145
Pulso2	56	96.8	84	176

En la segunda, se incluyen las tablas de frecuencias de las variables *Sexo, Fumador, Alcohol* y *Ejercicio*.

		Frecuencia
Sexo	Mujer	51
	Varón	59
Fumador	Sí	11
	No	99
Alcohol	Sí	68
	No	42
Ejercicio	Baja	37
	Moderada	59
	Alta	14

NOTA 10.1: Una vez insertada una tabla en Word, con el botón derecho del ratón se puede acceder a prácticamente todas las opciones de edición de tablas: insertar y borrar filas o columnas, cambio de colores, tamaños y tipos de las letras, modificación de los bordes de la tabla, alineación del texto dentro de las celdas, combinación de celdas, ...

11. Salir de R Commander

Para salir de R y R Commander desde la ventana de R Commander se utilizará la opción *Fichero> Salir> De Commander y R.* El programa preguntará si se desean guardar las ventanas abiertas. En general, no es necesario guardar ni el contenido del cuadro de sintaxis, ni el contenido de la pestaña Markdown, ni los resultados del cuadro de resultados. También se puede salir de R y R Commander desde la consola de R, a través de la opción *Archivo > Salir*.

Si se quiere guardar algún conjunto de datos creado en la sesión hay que comprobar primero que, efectivamente, es el conjunto de datos activo y aparece en el recuadro a la derecha de "*Conjunto de datos*", o seleccionarlo, si el conjunto de datos activo es otro. Lo guardaremos con la opción *Datos> Conjunto de datos activo> Guardar el conjunto de datos activo*

La opción *Fichero> Guardar el entorno de trabajo de R como,* permite guardar **todos** los conjuntos de datos abiertos durante la sesión de R Commander en un único fichero. Para cargar ese fichero con todos los conjuntos de datos en una nueva sesión de trabajo hay que utilizar la opción *Datos > Cargar conjunto de datos*, y seleccionar ese fichero. Aunque R Commander da un mensaje de error, realiza bien la operación de la carga de todos los conjuntos de datos del fichero. También se puede cargar este fichero desde la consola de R con la opción *Archivo > Cargar área de trabajo*.

12. Algunos consejos prácticos

1. En la mayoría de los cuadros de diálogo encontrarás los botones *Reiniciar, Aplicar* y *Ayuda.* El botón *Reiniciar* te permite elegir los valores preestablecidos para el cuadro de diálogo. El botón *Aplicar* te permite obtener los resultados asociados al cuadro de diálogo sin salir de él. Por último, el botón *Ayuda* te permite acceder a las páginas web (en inglés) donde se explica el uso de los comandos de R asociados al cuadro de diálogo. También puedes consultar las ayudas de R y R Commander a través del menú *Ayuda.*

2. No es lo mismo un dato faltante o *missing* (NA) que un dato con valor cero.

3. R y R Commander utilizan "." para representar la parte decimal, mientras que Excel utiliza ",".

4. Procura dar nombres cortos, pero significativos, a las variables y a los ficheros, sin emplear caracteres extraños, especialmente si el ordenador con el que estás trabajando es un Mac.

5. Para copiar cualquier texto, de los cuadros de sintaxis o resultados, a un documento Word basta con seguir los siguientes pasos.

- Selecciona con el ratón el texto que deseas copiar.
- Con el botón derecho del ratón, selecciona *Copiar* o pulsa simultáneamente las teclas <*control*>+<*c*>.
- Con un documento Word abierto, inserta los resultados poniendo el cursor en la posición deseada y seleccionado *Opciones de pegado* en la ventana que se abre al pulsar el botón derecho del ratón o pulsando simultáneamente las teclas <*control*>+<*v*>. También los puedes insertar seleccionando *Pegar* en el Menú.

6. Para copiar una ventana que se encuentre activa a un documento Word basta con seguir los siguientes pasos.

- Presiona simultáneamente las teclas <*Alt*>+<*Print Screen*> con esa ventana activa.
- Inserta la ventana en el documento Word poniendo el cursor en la posición deseada y seleccionando *Pegar* en el Menú, seleccionado *Opciones de pegado* en la ventana que se abre al pulsar el botón derecho del ratón o pulsando simultáneamente las teclas <*control*>+<*v*>.

Una vez insertada la ventana, puedes modificar su tamaño seleccionando el objeto y eligiendo la opción *tamaño y posición* con el botón derecho del ratón. También puedes modificar su tamaño seleccionando el objeto y usando el ratón. Al mover el ratón hacia los extremos de la caja que rodea al objeto, el ratón cambia de forma transformándose en una flecha. En ese momento, con el botón derecho del ratón podemos modificar el tamaño del objeto.

Otra alternativa para copiar una ventana (o una parte de ella) a un documento Word es utilizar la herramienta "*Recorte de pantalla"* de Windows, que se suele encontrar en *Notificaciones* o en *Inicio > Recorte y anotación*, o la herramienta "*Recortes"*, que suele encontrarse en *Inicio > Accesorios de Windows*.

7. Se puede copiar directamente desde R un conjunto de datos a un archivo Word (o a una hoja de cálculo Excel). Para ello:

- Desde el editor de R, selecciona con el ratón todos los datos que quieres copiar, desde el primero hasta el último.
- Pulsa simultáneamente <*control*>+<*c*> para llevar los datos al portapapeles.
- Seleccionando con el ratón la posición en el documento Word abierto (o pulsando sobre la casilla de la hoja de cálculo Excel a partir de la cuál quieres pegar los datos), inserta los datos pulsando <*control*>+<*v*>. También puedes usar la opción *Pegar* en el Menú, o las *Opciones de pegado* del botón derecho del ratón.

8. Para copiar una(s) columna(s) de datos de un archivo Word o una hoja de cálculo de Excel al editor de R Commander, basta con seguir los siguientes pasos.

- Selecciona con el ratón todos los datos de la(s) columna(s) que quieres copiar, desde el primero hasta el último.
- Selecciona *Copiar* en el menú o pulsa simultáneamente las teclas <*control*>+<*c*>.
- Define en el editor de R tantas filas y columnas como vayas a pegar.
- Coloca el ratón en el primer caso de la columna del editor de R Commander donde quieres insertar los datos, e insértalos pulsando simultáneamente las teclas <*control*>+<*v*>.

Si tienes varias columnas de datos en un documento Word, y sólo deseas copiar una de ellas para llevarla a un archivo de R, usa la combinación de la tecla <*Alt*> y el ratón.

13. Ejercicios

1. El archivo articuloperiodico.pdf contiene una información publicada el 11 de noviembre de 2009 en el Periódico de Aragón. Observa el uso que se hace de la estadística para describir a la población.

2. El archivo informe113.pdf contiene información sobre la evolución del covid-19 en España, a 12 de enero de 2022. Dicha información ha sido elaborada por el Instituto de Salud Carlos III de Madrid. Observa y comenta con tus compañeros el uso que se hace de la estadística para describir esa evolución.

3. Usando *Excel*, crea una nueva hoja de cálculo que contenga la altura (en metros), el peso (en kilos) y el sexo de varios de tus compañeros de clase. Genera unas medidas descriptivas de la altura y del peso. Guarda los datos y los resultados obtenidos en un archivo llamado *alturas.xlsx.*

4. Los siguientes datos, son las edades a las que 44 jugadores de tenis sufrieron la enfermedad conocida como *codo de tenista*. Introduce estas edades en una hoja de cálculo en *Excel*, obtén medidas descriptivas para estas edades y guarda los datos, junto con los resultados obtenidos, en un libro de *Excel*.

53, 21, 23, 37, 28, 37, 35, 31, 35, 41, 43, 48, 54, 57, 55, 54, 55, 43, 41, 31, 35, 28, 18, 24, 24, 37, 31, 36, 35, 41, 43, 48, 54, 46, 49, 44, 42, 43, 55, 43, 38, 46, 31, 28.

5. En un grupo de enfermos que sufren esquizofrenia paranoica se ha estudiado la edad en la que se produjo el primer ataque. Introduce los datos en una hoja de cálculo *Excel* y genera medidas descriptivas para la variable edad. Guarda los datos y los resultados obtenidos en un libro *Excel*.

Hombres: 24, 33, 23, 20, 26, 32, 35, 21 y 25.
Mujeres: 22, 34, 26, 31, 26, 35, 25, 38, 36, 22, 23 y 37.

6. Una muestra con 28 adultos alcohólicos que sufren hipertensión ha sido utilizada para estudiar el efecto de un antihipertensivo. Los individuos han sido asignados aleatoriamente a un grupo tratado con un placebo o a un grupo tratado con el medicamento. Introduce los datos en una hoja de cálculo *Excel* y genera medidas descriptivas para la variable tensión en el grupo Placebo y por separado en el grupo Medicamento. Guarda los datos y los resultados obtenidos en un libro *Excel*.

Placebo: 105, 107, 110, 117, 124, 153, 137, 174, 109, 119, 143, 162, 91, 146 y 109.
Medicamento: 92, 96, 104, 119, 106, 100, 93, 90, 98, 109, 106, 88 y 94.

7. En un grupo de enfermos que sufren esquizofrenia paranoica se ha estudiado la edad en la que se produjo el primer ataque. Introduce los datos en el editor de R Commander, pon nombres *Edad* y *Sexo* a las variables y para la variable *Sexo* utiliza los valores 1 y 2. Genera la variable *facSexo* a partir de la variable *Sexo*, para que contenga "hombre" y "mujer" en función de si la variable *Sexo* es 1 o 2. Guarda los datos en un archivo de R con nombre *esquizofrenia*.

Hombres: 24, 33, 23, 20, 26, 32, 35, 21 y 25.
Mujeres: 22, 34, 26, 31, 26, 35, 25, 38, 36, 22, 23 y 37.

8. Una muestra de 28 adultos alcohólicos que sufren hipertensión ha sido utilizada para estudiar el efecto de un antihipertensivo. Los individuos han sido asignados aleatoriamente a un grupo tratado con un placebo o a un grupo tratado con el medicamento. Introduce los datos en el editor de R, pon nombres *Tension* y *Grupo* a las variables y para la variable *Grupo* utiliza los valores 1 (para placebo) y 2 (para medicamento). Genera la variable *facGrupo* a partir de la variable *Grupo*, que contenga "placebo" y "medicamento" en función de si la variable *Grupo* es 1 o 2. Guarda los datos en un archivo de R con nombre *tension*.

Placebo: 105, 107, 110, 117, 124, 153, 137, 174, 109, 119, 143, 162, 91, 146 y 109.
Medicamento: 92, 96, 104, 119, 106, 100, 93, 90, 98, 109, 106, 88 y 94.

9. Los estudiantes de una clase de enfermería participaron en el siguiente experimento. Primero se tomaron su pulso cardíaco (*Pulso1*). Luego, lanzaron una moneda al aire. Si la moneda salía cara, entonces debían correr durante un minuto. En otro caso, se quedaban sentados durante ese minuto. Al cabo de ese minuto, los estudiantes se volvieron a tomar el

pulso (*Pulso2*). Los resultados del experimento, junto con información personal de los estudiantes, se encuentran en el fichero *Pulso.RData*.

Carga en R Commander el archivo Pulso.Rdata. Crea un fichero Word y guárdalo con nombre *Tarea*. En ese fichero Word, contesta a las siguientes preguntas.

a) Haz una lista con los nombres de las variables y su tipo (numérico o categórico). Si en alguna de ellas el tipo de variable en R (numérico o factor) no coincide con el tipo estadístico, indícalo también.
b) Observa las variables del caso 76. ¿Qué valores se registran para ese individuo en *Pulso1* y *Pulso2*?
c) Genera una nueva variable de tipo factor que se llame *facSexo*, que contenga chico o chica en función de si la variable *Sexo* es 1 ó 2. Obtén una tabla de frecuencias de esa variable y cópiala en el fichero Word.
d) Genera la variable *difer*, conteniendo la diferencia entre el pulso después del minuto y el pulso antes del minuto. Obtén un resumen numérico de la variable *difer* y cópialo en el fichero Word.

Una vez contestadas las preguntas, guarda los datos modificados en formato propio de R con el nombre *pulsonuevo*. Guarda también dichos datos modificados en un archivo de *texto* con nombre *pulsotexto*, donde, además de los valores de las variables separados por espacios, aparezcan los nombres de las variables, los valores de tipo carácter entrecomillados y no aparezcan los nombres de las filas. Por último, guarda el documento Word *Tarea.*

10. El archivo *calcio.RData* contiene información sobre un experimento realizado para comprobar el efecto del calcio sobre la presión sanguínea de las personas. Un grupo de 10 personas recibió un suplemento de calcio durante 12 semanas, mientras que un grupo de 11 personas recibió placebo durante el mismo periodo de tiempo. A todas las personas se les midió la presión sanguínea antes y después del tratamiento.

a) Abre este conjunto de datos con R Commander. Observa qué variables contiene y cuál es su tipo.
b) Genera la variable *difer*, conteniendo la diferencia entre la presión sanguínea después del tratamiento y la presión sanguínea antes del tratamiento.
c) Guarda como ficheros externos los datos modificados en formato propio de R, con el nombre *calcionuevo,* y en formato de hoja de tipo texto, con el nombre *calciotexto*.

11. El archivo *camisetas.xlsx* es un archivo con formato Excel que contiene los datos de una encuesta realizada a un grupo de jóvenes para estudiar el tipo de camiseta más usado. Las variables que contiene son: un número de identificación del encuestado, los años del encuestado, los meses transcurridos desde su último cumpleaños, su altura en centímetros y la talla de camiseta que utiliza (s, m, l, xl).

a) Importa estos datos con R Commander.
b) A partir de las variables años y meses trascurridos desde el último cumpleaños, crea una nueva variable llamada *edad* que sea la edad del encuestado aplicando la fórmula edad=años+meses/12.
c) Obtén la tabla de frecuencias asociada a la variable *camiseta* y cópiala en un fichero Word. A la vista de esa tabla de frecuencias, comenta qué podríamos decir sobre la distribución de las tallas de las camisetas en esa muestra.
d) Guarda este conjunto de datos, que incluye la variable *edad*, en el archivo de R Commander *Rcamisetas*.

Para pensar:

- *Si las estadísticas mostrasen que la mortalidad por tuberculosis es mayor en Segovia que en las demás provincias, ¿significaría esto que el clima segoviano favorece el contagio tuberculoso?*
- *Un estudio hizo ver que en cierta población europea se produjo simultáneamente un fuerte crecimiento en el nacimiento de niños y un notable incremento del número de nidos de cigüeñas. ¿Es el incremento en el número de cigüeñas la causa del incremento en el nacimiento de niños o es el incremento en el nacimiento de niños el responsable del aumento del número de cigüeñas?*