

Práctica 1

Introducción a R Commander

Contenido de la práctica

1.1. Introducción	1
1.2. Iniciar R y R Commander	2
1.3. La ventana de R Commander	3
1.4. Conjunto de datos	3
1.4.1. Abrir un archivo de datos con R	3
1.4.2. Guardar un archivo de datos con R	4
1.4.3. Importar ficheros	5
1.5. Variables en R: tipos y su manipulación.	5
1.5.1. Tipos de variables	5
1.5.2. Cambiar variable numérica a factor	6
1.5.3. Calcular una nueva variable a partir de otras	7
1.5.4. Recodificar variables	7
1.6. Filtrar el fichero de datos	8
1.7. Salir de R Commander	10
1.8. Crear de un conjunto/fichero de datos	10

1.1. Introducción

El software estadístico que vamos a utilizar en la asignatura se llama R Commander, también conocido por su abreviatura Rcmdr. Se basa en un lenguaje de programación llamado R, que está enfocado al cálculo estadístico, aunque puede utilizarse para muchas otras tareas. El manejo de R se realiza mediante la escritura de comandos que el usuario debe conocer y con los que se indica a R la tarea que se quiere realizar. Para poder utilizarlo de una forma sencilla, sin necesidad de aprender los comandos, usaremos R Commander. Este es un programa basado en un sistema de menús y ventanas, que se encarga de mandar a R las órdenes de lo que el usuario quiere y de recoger los resultados obtenidos para mostrárselos. R Commander es software libre y puede instalarse en cualquier ordenador.

1.2. Iniciar R y R Commander

Para iniciar *R* desde el escritorio:

1. Se selecciona en la parte inferior izquierda de la pantalla:

Inicio > Todos los programas > R > R

2. Aparece la ventana *R Console* que se muestra en la figura 1.1 y es posible que a continuación se abra automáticamente la ventana *R Commander* que se muestra en la figura 1.2, que es en la que se trabajará y donde se realizarán los análisis estadísticos.

3. En caso de que *R Commander* no se haya abierto automáticamente, se selecciona, en la parte superior de la ventana *R Console*, la opción:

Paquetes > Cargar paquete

En la lista se elige el paquete *Rcmdr* y se pulsa el botón OK.

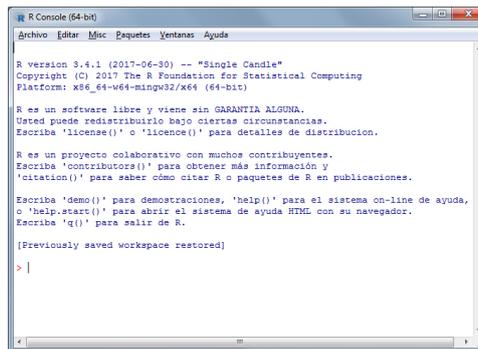


Figura 1.1: Ventana de R.

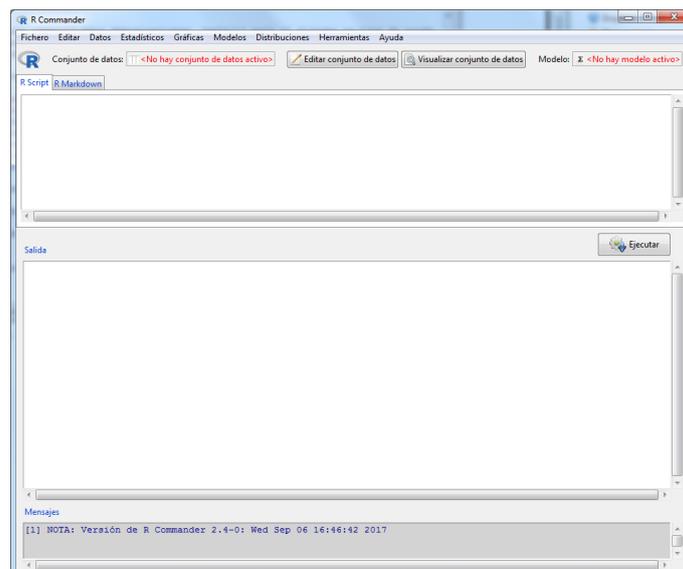


Figura 1.2: Ventana de R Commander.

1.3. La ventana de R Commander

La parte superior de la ventana de R Commander contiene el menú principal.

- En el menú *Datos* se encuentran las opciones para crear, cargar, guardar o modificar un conjunto de datos.
- En el menú *Estadísticos* las técnicas estadísticas más habituales.
- En el menú *Gráficas* diferentes tipos de representaciones gráficas.
- En el menú *Distribuciones* las herramientas para resolver problemas de cálculo de probabilidades.
- En el menú *Modelos* se encuentran algunas técnicas de selección de variables y validación de un modelo lineal.

Debajo del menú principal aparecerá el nombre del archivo de datos que vayamos a analizar, si ya se ha cargado, y los botones para poder editar o visualizar el conjunto de datos.

A continuación, en la ventana *R Script* se escribirán manual o automáticamente (si se usan las opciones del menú principal) las órdenes que debe ejecutar R. Debajo de esta, en la ventana *Salida*, se obtendrán los resultados de las instrucciones ejecutadas. Finalmente, en la ventana *Mensajes* aparecen mensajes informativos (en color azul), mensajes de error (en color rojo) o mensajes de precaución (en color verde). Observa que el primer mensaje que aparece se indica la versión de R que se está utilizando.

1.4. Conjunto de datos

1.4.1. Abrir un archivo de datos con R

El primer paso cuando se desea analizar un conjunto de datos con R Commander es crear o cargar el archivo que los contiene. Observa en la parte superior izquierda de la figura 1.2 el mensaje que aparece cuando todavía no se ha seleccionado un conjunto de datos: `<No hay conjunto de datos activo >` .

La extensión *.RData* corresponde a un fichero propio de R. Para abrirlo vamos al menú:

Datos > Cargar conjunto de datos

y, en la carpeta donde se encuentre, lo seleccionamos. Una vez ejecutado, pulsamos el botón: **Visualizar conjunto de datos**.

Ejercicio 1: Abre y visualiza el fichero *beneficios.RData*.

Al abrir el fichero y visualizarlo, obtenemos la figura 1.3. Este es el aspecto de un conjunto de datos (matricial, rectangular). En las filas aparecen los casos (empresas) y en las columnas las variables estudiadas. Este fichero contiene información sobre empresas (EMPRESA) agrupadas por sectores (SECTOR) y de las que se proporcionan datos sobre beneficios antes de impuestos (BAI) y ventas (VENTAS) en los años 94 y 95.

- *EMPRESA*: nombre de la empresa
- *SECTOR*: sector al que pertenece la empresa

- $BAI94$ y $BAI95$: beneficios antes de impuestos en los años 94 y 95.
- $VENTAS94$ y $VENTAS95$: ventas en los años 94 y 95.

El conjunto de datos se llama *beneficios*.

	SECTOR	EMPRESA	BAI94	VENTAS94	BAI95	VENTAS
1	ALIMENTARIAS	Bodegas y Bebidas	831	27004	1368	333
2	ALIMENTARIAS	Campofrio	2718	50739	2635	536
3	<NA>	Koipe	1969	49640	2559	1131
4	ALIMENTARIAS	OMSA Alimentación	1064	29423	509	310
5	ALIMENTARIAS	Pascual Herm.	-1050	9425	-1376	19
6	ALIMENTARIAS	San Miguel	3330	36494	1557	355
7	ALIMENTARIAS	Viscofán	2335	44039	10	465
8	AUTOMOVILES	Citroën Hispania	3093	359321	2013	3602
9	AUTOMOVILES	Fasa Renault	4645	622771	1694	6710
10	CEMENTO Y CONSTRUCCION	Asland C.	577	8497	3930	96
11	CEMENTO Y CONSTRUCCION	Cem. Molins	3178	32612	2606	347
12	CEMENTO Y CONSTRUCCION	Cem. Portland	7029	35669	9014	380
13	CEMENTO Y CONSTRUCCION	Cristaleria	4246	124807	16464	1448
14	CEMENTO Y CONSTRUCCION	Uniland	3065	18454	5938	215
15	CEMENTO Y CONSTRUCCION	Uralita	163	120702	12505	1409
16	CEMENTO Y CONSTRUCCION	Valenciana	13171	89147	16917	1071
17	CEMENTO Y CONSTRUCCION	Vidrala	1755	11259	2399	116
18	COMERCIO Y VARIOS	CLH (REP Campsa)	26759	62907	29038	593
19	COMERCIO Y VARIOS	Continente	13867	392971	14686	4070
20	COMERCIO Y VARIOS	C. Financ. Reunida	1596	2029	1674	
21	COMERCIO Y VARIOS	Mapfre	12574	199030	17868	2413
22	COMERCIO Y VARIOS	G. D. Anaya	4677	18691	4904	196
23	COMERCIO Y VARIOS	G. D. Felguera	-4206	32706	612	443
24	COMERCIO Y VARIOS	G. Hispano Suiza	45	54	32	12
25	COMERCIO Y VARIOS	Pryca	26444	520283	31602	5433
26	COMERCIO Y VARIOS	Prosegur	3371	43844	4402	557
27	COMERCIO Y VARIOS	Tabacalera	17467	783724	18748	7791
28	CONSTRUCTORAS	Agromán	-19554	125804	-5340	1025
29	CONSTRUCTORAS	Cubiertas	6095	193986	5321	2113
30	CONSTRUCTORAS	Construc. Lain	-930	39754	1115	566

Figura 1.3: Visualización del fichero beneficios

Nota 1: En una casilla de la variable *SECTOR*, aparece el símbolo *NA*. Para R Commander, éstos son los datos que no se conocen (datos perdidos, NA = Not Available en inglés).

Como se trata de un error, lo corregimos, Koipe es una empresa del sector *ALIMENTARIAS*. Si no se tuviera información para corregir el error el dato se dejaría con *<NA >* y en los cálculos que realicemos dicho caso no se consideraría.

1.4.2. Guardar un archivo de datos con R

Cuando se crea un fichero con R Commander, o se realizan modificaciones útiles sobre un fichero de datos existente, conviene guardar el fichero generado. Para ello, estando activo el fichero que se desea guardar, se utiliza la opción:

Datos > Conjunto de datos activo > Guardar el conjunto de datos activo

Se pueden guardar todos los ficheros generados en una sesión de trabajo en un único conjunto de ficheros que contenga a todos ellos (una opción muy útil si se han generado ficheros por filtrado, etc., por ejemplo). Para ello en el menú, se utiliza la opción:

Fichero > Guardar el entorno de trabajo de R como...

1.4.3. Importar ficheros

R utiliza un formato propio para almacenar los datos (extensión *.RData*), sin embargo es capaz de leer datos de otros tipos de ficheros, por ejemplo: *Excel*, *SPSS*, *SAS*, *Minitab*, *STATA* y también desde ficheros de texto con formato *.csv*. Para ello se utiliza el menú *Importar datos* que se encuentra en:

Datos > Importar datos

A continuación, importamos el fichero Excel *estanciasCCA.xlsx*, contiene información proporcionada por el INE de la encuesta de Ocupación en Alojamientos de Turismo Rural (comunidades y ciudades autónomas). Antes de importarlo le podemos echar un vistazo. Para importar seleccionamos:

Datos > Importar datos > desde un archivo Excel...

En la ventana que aparece (figura 1.4) le indicamos el nombre que queremos ponerle a nuestro conjunto de datos, por ejemplo *estanciasCAA*. También hay que indicarle que el nombre de las variables (columnas) aparece en la primera fila del documento Excel y que las variables carácter las convierta en *factores* (variables cualitativas de R). Tras pulsar *Aceptar*, aparece un diálogo de apertura de fichero, buscamos el fichero *estanciasCCA.xlsx* y lo seleccionamos. Si el fichero contiene más de una hoja, entonces nos aparecerá una ventana para que seleccionemos la hoja desde la que queremos importar los datos (en nuestro fichero la que se denomina *tabla*).

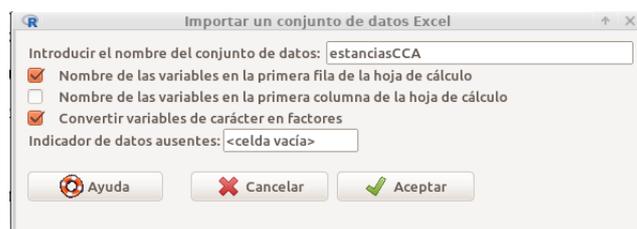


Figura 1.4: Importando archivo desde Excel

Una vez importado, si queremos guardarlo en el formato de R para posteriormente abrirlo directamente debemos guardar el fichero de datos:

Datos > Conjunto de datos activo > Guardar el conjunto de datos activo

Ejercicio 2: Importa la información contenida en el fichero Excel *datosLab.xls*. La primera hoja impórtala y almacena su información en un conjunto de datos denominado *Lab23* y la segunda hoja en uno denominado *Lab21*. Posteriormente, guarda los conjuntos de datos en sendos ficheros de nombre *Lab23.RData* y *Lab21.RData*.

1.5. Variables en R: tipos y su manipulación.

1.5.1. Tipos de variables

En R las variables se clasifican en *numéricas* (cuantitativas) y *factores* (cualitativas). En general, son *numéricas* aquellas cuyos valores figuran en el archivo con números y son *factores* aquellas cuyos valores figuran en el archivo con caracteres alfanuméricos (letras acompañadas o no de números).

Cuidado con las codificaciones: algunas veces las variables cualitativas se codifican con números por comodidad, pero deben ser tratadas adecuadamente como variables cualitativas.

Para saber de qué manera ha clasificado R las variables es útil realizar un estudio básico de cada una de ellas con la opción:

Estadísticos > Resúmenes > Conjunto de datos activo

Para las variables de tipo son *factor* R Commander realiza un recuento de los distintos valores y para las *numéricas* calcula algunas medidas.

Ejercicio 3: Realiza un resumen de las variables del fichero *beneficios.Rdata* y coméntalo.

Ejercicio 4: Carga el fichero *CCAA0910.RData* que contiene información variada sobre un conjunto de estudiantes. Haz un resumen numérico y clasifica las variables que contiene el fichero. ¿Crees que todo es correcto?

1.5.2. Cambiar variable numérica a factor

En el último ejemplo hemos observado que la variable *sexo* está mal codificada, ya que se trata de una variable cualitativa pero está representada con números. Para convertir una variable numérica en una variable factor se utiliza:

Datos > Modificar variables del conjunto de datos activo > Convertir variable numérica en factor...

Este menú permite utilizar el mismo nombre de la variable (o uno nuevo) y que los números sean tratados como caracteres (o bien asignarles otros valores).

Para cambiar la variable *sexo*, seleccionamos la opción, en el cuadro de diálogo que aparece seleccionamos la variable y dejamos marcada la opción *Asignar nombres a los niveles*. En *Nuevo nombre o prefijo para variables múltiples* no ponemos nada, de forma que la codificación correcta sustituirá a la antigua. Pulsando en *Aceptar* aparece un diálogo confirmando que queremos sustituir la antigua variable *sexo* por la nueva y después pide las equivalencias entre los números y los nuevos valores (ver figura 1.5).

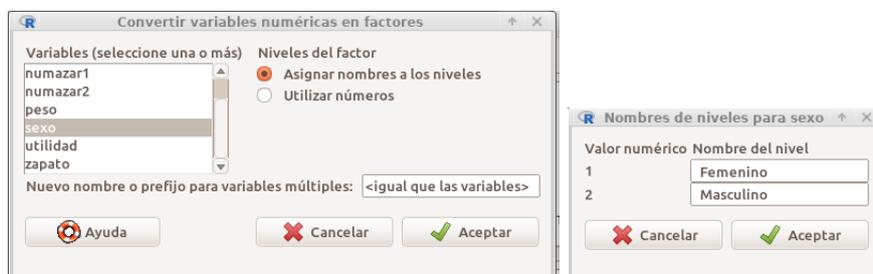


Figura 1.5: Conversión de variable numérica en factor.

Ejercicio 5: Corrige la variable *gafas* convirtiéndola a factor sin cambiarla de nombre. Los nuevos valores de las variable deben ser *Sí* y *No*, para los valores 1 y 0 respectivamente.

1.5.3. Calcular una nueva variable a partir de otras

En ocasiones resulta necesario analizar nuevas variables de interés, que se obtienen a partir de las ya existentes. El menú es:

Datos > Modificar variables del conjunto de datos > Calcular una nueva variable...

Las operaciones que se pueden realizar con variables son las habituales.

Ejercicio 6: Con los datos del fichero *CCAA0910.RData*, calcula el Índice de Masa Corporal (IMC) de los estudiantes. Este se define como:

$$\text{IMC} = \text{peso en kg} / (\text{altura en m})^2$$

El proceso se muestra en la figura 1.6.

Nota 2: En la casilla *Expresión a calcular* debes escribir las variables tal y como aparecen en el fichero (fíjate bien en las mayúsculas y minúsculas).

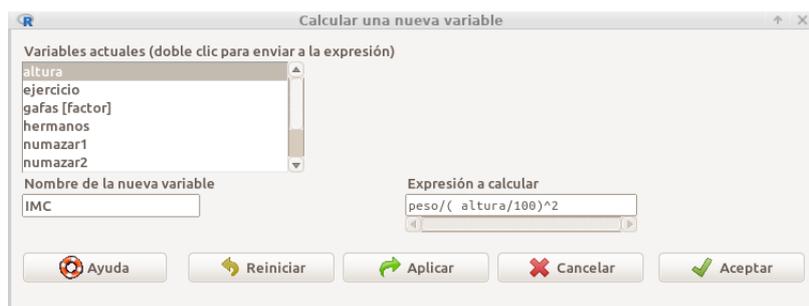


Figura 1.6: Creación de una nueva variable.

1.5.4. Recodificar variables

En ocasiones, se requiere agrupar los valores de una variable cuantitativa en categorías. Por ejemplo dada la información del gasto per capita en comisiones bancarias nos puede interesar agrupar dicho gasto en niveles, gasto bajo, gasto medio, gasto alto y gasto muy alto, dependiendo de los valores registrados. Para esto se utiliza la opción de recodificación de variables. Veamos un ejemplo con los datos recién calculados del *IMC* de los estudiantes.

El menú que permite realizar esta transformación es:

Datos > Modificar variables del conjunto de datos activo > Recodificar variables

Los pasos a seguir son:

1. Elegir la variable que se quiere recodificar (se pueden varias variables siempre que sean del mismo tipo y se les desee asignar la misma recodificación).
2. Escribir el nombre de la nueva variable.

3. Si la nueva variable tiene carácter numérico, se elimina la marca de la opción **Convertir cada nueva variable en factor**. En otro caso, deja marcado.
4. Escribir las reglas de recodificación.

A continuación se muestra las reglas de recodificación. Hay que tener en cuenta que R Commander recodifica en el orden en el que se incluyen las directrices de recodificación.

- **lo:a="Categoría 1"** indica que la categoría 1 de la nueva variable serán los valores $\leq a$, si el valor **a** no ha sido previamente recodificado. Si ha sido incluido en una recodificación previa, entonces incluye $<a$. El valor **Categoría 1** se escribe entre comillas si la variable nueva es de tipo factor.
- **a:b="Categoría 2"** indica que los valores de la variable original dentro del intervalo **[a,b]** pasarán a ser **Categoría 2** en la nueva variable. Si el valor *a* (o bien *b* o ambos) ha sido utilizado en alguna regla previa, entonces incluye $(a, b]$ (o bien $[a, b)$ o (a, b)).
- **b:hi="Categoría 3"** indica que la categoría 3 comprenderá los valores $\geq b$, siempre que **b** no se haya recodificado previamente.

Ejercicio 7: En la tabla siguiente se muestra la clasificación en cuatro niveles de las personas según el valor de su Índice de Masa Corporal (IMC). Calcula una nueva variable, denominada *Clasificación*, para los datos de los estudiantes contenidos en *CCAA0910*.

Valores IMC	Categoría en IMC
Menor que 18.5	BajoPeso
[18.5, 25)	PesoNormal
[25, 30)	Sobrepeso
30 o más	Obesidad

Las directrices que debes escribir son:

30:hi = "Obesidad"

25:30 = "Sobrepeso"

18.5:25 = "PesoNormal"

lo:18.5 = "BajoPeso"

1.6. Filtrar el fichero de datos

En algunas ocasiones es interesante estudiar un subconjunto de la población o de la muestra por separado. Aunque R Commander permite hacer estudios por grupos (definidos por una variable) en algunas ocasiones es mejor construir un fichero de datos independiente con la submuestra de interés. Además, esta opción de *Filtrado de ficheros*, permite también eliminar del fichero datos erróneos o atípicos.

La creación de subficheros se realiza utilizando la opción:

Datos > Conjunto de datos activo > Filtrar conjunto de datos activo

Para realizar el filtrado o extracción de los casos es necesario indicar en el cuadro *Expresión de selección* la condición que deben cumplir los datos que queremos extraer. Tras ello hay que indicar un nombre para el nuevo conjunto de datos.

Vamos a crear, a partir del fichero *CCAA0910.RData*, un conjunto de datos llamado *Mujeres* que contenga los datos de *las* estudiantes. La condición que impondremos es que la variable *sexo* tome valor *Femenino* y lo indicaremos como:

`sexo == "Femenino"`

El diálogo queda cómo el mostrado en la figura 1.7.



Figura 1.7: Filtrado/extracción de registros del conjunto *CCAA0910*

Tras darle a *Aceptar*, observa que, si no se ha producido ningún error, el nombre del conjunto de datos activo ha cambiado a *Mujeres* y, si lo visualizáis, solo veréis los registros que tenían el valor *Femenino* en la variable *sexo*. Es este momento el conjunto de datos está creado pero no guardado.

Ejercicio 8: Crea un fichero que contenga la información correspondiente a los hombres del fichero *CCAA0910.RData*. **Importante:** antes de proceder, selecciona el conjunto de datos *CCAA0910* como conjunto de datos activo (actualmente pone *Mujeres*).

Nota 3: En la expresión de selección puede ser necesario utilizar fórmulas más complejas. A continuación se detallan las expresiones y operaciones lógicas más habituales:

Expresiones lógicas				Operadores lógicos	
<	Menor	<=	Menor o igual	&	Conjunción (y)
>	Mayor	>=	Mayor o igual		Disyunción (o)
==	Igualdad lógica	!=	Distinto	!	Negación

Si, por ejemplo, quisiéramos crear un fichero con las mujeres cuya altura sea mayor o igual que 160 cm utilizaríamos la expresión:

`sexo=="Femenino" & altura>=160`

Observa que cuando la variable es de tipo cualitativo (factor) el valor debe ponerse entre comillas y cuando la variable es de tipo cuantitativo (numérico) el valor se pone sin comillas.

Ejercicio 9: Crea un conjunto de datos con los estudiantes que utilizan la talla de camiseta *L* y que tienen 2 hermanos y que usan gafas.

1.7. Salir de R Commander

Para salir (de un modo ortodoxo) de R y R Commander se utilizará:

Fichero > Salir > De Commander y R

Previamente el programa preguntará si se desean guardar las ventanas abiertas. Si se quiere guardar algún conjunto de datos creado ya se ha indicado cómo hacerlo.

Nota 4: Si en algún momento se cierra por error la ventana de R Commander, quedando abierta la de R, se puede volver a cargar R Commander ejecutando la instrucción **Commander()**.

1.8. Crear de un conjunto/fichero de datos

Veamos cómo introducir un conjunto de datos directamente en R Commander con el siguiente ejemplo¹: “Un sistema multimedia para aprender estadística incluye una prueba para valorar la destreza de los sujetos en la utilización del ratón (mouse). El programa informático hace que aparezca, al azar, un círculo en la pantalla. El sujeto tiene que situarse sobre el círculo y clicar tan rápido como pueda. Tan pronto como el usuario ha clicado sobre el círculo, aparece uno nuevo.” El tiempo de reacción se mide en milisegundos y la distancia desde el ratón al círculo en unidades de pantalla.

Prueba	Tiempo ms	Distancia (up)	Mano
1	115	190,7	derecha
2	96	138,52	derecha
3	110	165,08	derecha
4	100	126,19	derecha
5	111	163,19	derecha
6	101	305,66	derecha
7	111	176,15	derecha
8	106	162,78	derecha
9	96	147,87	derecha
10	96	271,46	derecha
11	240	190,7	izquierda
12	190	138,52	izquierda
13	170	165,08	izquierda
14	125	126,19	izquierda
15	315	163,19	izquierda
16	141	176,15	izquierda
18	210	162,78	izquierda
19	200	147,87	izquierda
20	401	271,46	izquierda

Seleccionamos:

Datos > Nuevo conjunto de datos

¹Extraído de Estadística aplicada básica. David S. More. Antoni Bosch Ed., pp. 201

En la ventana que aparece introducimos el nombre del conjunto de datos que vamos a crear, por ejemplo, *Multimedia*. Pulsamos en *Añadir columna* hasta tener 3 columnas y en *Añadir fila* hasta tener 20 filas. Observa que la primera columna se llama *rowname* y su contenido corresponde al nombre de la fila, que si no se indica lo contrario contiene el orden en la tabla, que corresponde con nuestra columna prueba.

Ahora si pulsamos sobre la cabecera de la primera columna, donde aparece *V1* podemos escribir el nombre, *Tiempo*, picando en la cabecera de la segunda, donde aparece *V2* podemos escribir el nombre de la segunda variable *Distancia* y análogamente en la última *Mano*. A partir de aquí se introduce en cada casilla el valor de la tabla (en R commander el separador decimal es el punto).

Al finalizar la edición de los datos tendremos la ventana que se muestra en la figura 1.8. A continuación, si pulsas *Aceptar*, observa cómo el conjunto de datos introducido es el nuevo *Conjunto de datos* activo (recuerda que este conjunto de datos no está guardado en el ordenador)

rowname	Tiempo	Distancia	Mano	
1	1	115	190.70	derecha
2	2	96	138.52	derecha
3	3	110	165.08	derecha
4	4	100	126.19	derecha
5	5	111	163.19	derecha
6	6	101	305.66	derecha
7	7	111	176.15	derecha
8	8	106	162.78	derecha
9	9	96	147.87	derecha
10	10	96	271.46	derecha
11	11	240	190.70	izquierda
12	12	190	138.52	izquierda
13	13	170	165.08	izquierda
14	14	125	126.19	izquierda
15	15	315	163.19	izquierda
16	16	240	305.66	izquierda
17	17	141	176.15	izquierda
18	18	210	162.78	izquierda
19	19	200	147.87	izquierda
20	20	401	271.46	izquierda

Figura 1.8: Ventana tras la introducción de datos.

Si por alguna razón habéis definido alguna columna de más o alguna fila de más no hay problema puesto que se pueden borrar, respectivamente, con las opciones:

Datos > Modificar variables del conjunto de datos > Eliminar variables del conjunto de datos...

Datos > Conjunto de datos activos > Borrar fila(s) del conjunto de datos activo

Ejercicio 10: Construye una variable que represente la velocidad de reacción definida como:

$$\text{Distancia} / \text{Tiempo}$$

Después, calcula la velocidad media de reacción con cada mano.

