

Introducir e importar datos con R-Commander

En esta práctica, aprenderemos, por una parte a introducir datos y crear nuestro propio conjunto de datos usando R-Commander, y por otra parte a importar datos desde un fichero de texto.

0. Primeros pasos.

Empezaremos por:

1. Ejecutar R
2. Cambiar el directorio de trabajo al directorio que hayamos escogido (por ejemplo Mis Documentos)
3. En el caso en que no lo tengamos hecho, descargar el fichero datos_para_importar.zip.Rdata desde el Aula Virtual (en Datos necesarios para las prácticas), lo copiamos en el directorio de trabajo y lo descomprimos.
4. Limpiar el entorno de trabajo de cualquier objeto heredado de una sesión anterior:

`rm(list=ls())`

5. Comprobamos los objetos que tenemos definidos con la instrucción `ls()`
6. Lanzamos R-Commander con la instrucción :

`library(Rcmdr)`

2. Introducir datos manualmente.

2.1 Desde R, usando las instrucciones `c` y `data.frame`

Desde la línea de comandos de R, podemos construir nuestros propios conjuntos de datos, introduciendo los valores manualmente, tal como lo vimos en el documento: [Transparencias, introducción a R. \(Aula virtual\)](#)

Por ejemplo:

```
peso = c(54, 70, 65, 78, 68, 85)
```

```
altura = c(160, 170, 172, 185, 160, 175)
```

```
constitución = data.frame(altura, peso)
```

2.2 Desde R-Commander, usando el editor de datos.

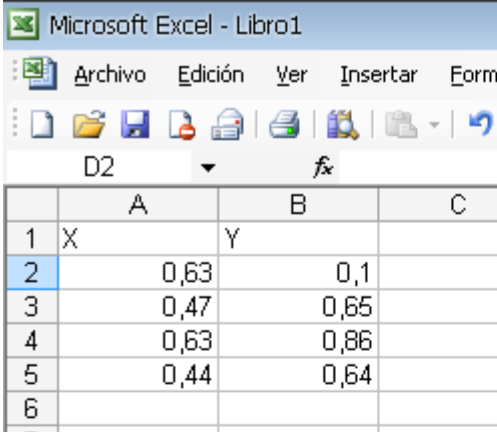
- En la primera fila de botones, usamos el comando Datos > Nuevo conjunto de datos
- Escogemos el nombre del conjunto de datos que queremos construir.
- Aparece una ventana de edición de datos.
 - Pinchando en la cabecera (`var1`, `var2`, ...) podemos especificar el nombre de las variables y su tipo (numérico o caracteres). Para cada variable, debemos pulsar `intro`, una vez que hemos especificado estas dos características
 - Basta entonces con introducir los datos en cada celda para cada variable.

Cuando hemos acabado de introducir los datos, usamos la instrucción Cerrar en el menú Archivo del editor de datos.

2.3 Pasando primero por Excel.

El editor de datos de R-Commander es muy rudimentario, puede ser más cómodo preparar primero un fichero Excel que contenga los nombres de las variables y los datos para importarlo después desde R-Commander.

Abrimos Excel y en una hoja de cálculo, introducimos los nombres de las variables en la primera fila y los datos a continuación:



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - Libro1". The menu bar includes "Archivo", "Edición", "Ver", "Insertar", and "Form". The toolbar contains icons for file operations and editing. The active cell is D2. The spreadsheet has columns A, B, and C, and rows 1 through 6. Column A is labeled "X" and column B is labeled "Y".

	A	B	C
1	X	Y	
2		0,63	0,1
3		0,47	0,65
4		0,63	0,86
5		0,44	0,64
6			

Guardamos el fichero con el nombre: ejemploparaR.xls.

Desde R-Commander, podemos ahora importar los datos:

Usamos

Datos > Importar datos > desde conjunto de datos Excel, Access o dBase.

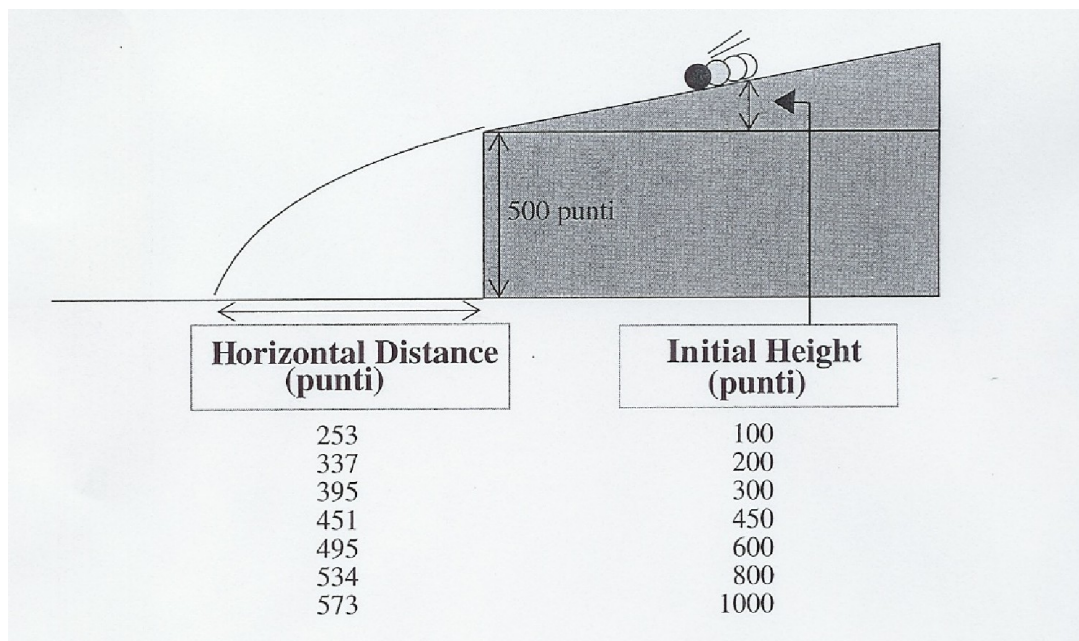
Especificamos el nombre del conjunto (por ejemplo datosexcel), recorremos las carpetas hasta encontrar el fichero ejemploparaR.xls y finalmente sólo nos queda especificar qué hoja queremos importar del fichero Excel (en nuestro caso, la hoja 1).

2.4 Ejercicios

- 1) Introducir los datos siguientes siguiendo cada uno de los tres métodos y creando tres conjuntos de datos llamados datosmetodo1, datosmetodo2 y datosmetodo3.

x	8,9	5,5	3,1	0,7	1,5
y	8,7	7,9	3	0,2	8,2
z	5,1	9,7	6,7	8,2	0,7

- 2) Introducir los datos siguientes, escogiendo cualquiera de los tres métodos
 - a) En 1609 Galileo demostró matemáticamente que la trayectoria de un cuerpo que cae con un componente de velocidad horizontal es una parábola. Su descubrimiento tuvo su origen en observaciones empíricas que realizó casi un año antes. Para estas observaciones, ideó un experimento en el que una bola empapada de tinta rodaba en un plano inclinado para luego caer desde una altura de 500 punti (1 punti= 169/189mn). Galileo estudió la distancia horizontal que alcanza la bola en función de la altura desde la que sale. Un diagrama ilustrativo, extraído de Ramsey, Schafer (2002), "The statistical Sleuth" p 268, se enseña a continuación:



b) La hidrólisis de un cierto éster tiene lugar en medio ácido según un proceso cinético de primer orden. Partiendo de una concentración inicial de 30 mM del éster, se han medido las concentraciones del mismo a diferentes tiempos obteniéndose los resultados siguientes:

T	3	4	10	15	20	30	40	50	60	75	90
C	25,5	23,4	18,2	14,2	11	6,7	4,1	2,5	1,5	0,7	0,4

3. Importar datos desde ficheros de texto

) 3.1 Importar los datos

En la gran mayoría de las veces, es probable que tengamos los datos ya introducidos en un fichero. En esta asignatura, los datos que usaremos se presentarán en un archivo de texto, con campos separados por algún carácter determinado (por ejemplo “;”).

Para importar datos desde un fichero de texto, usamos el comando:

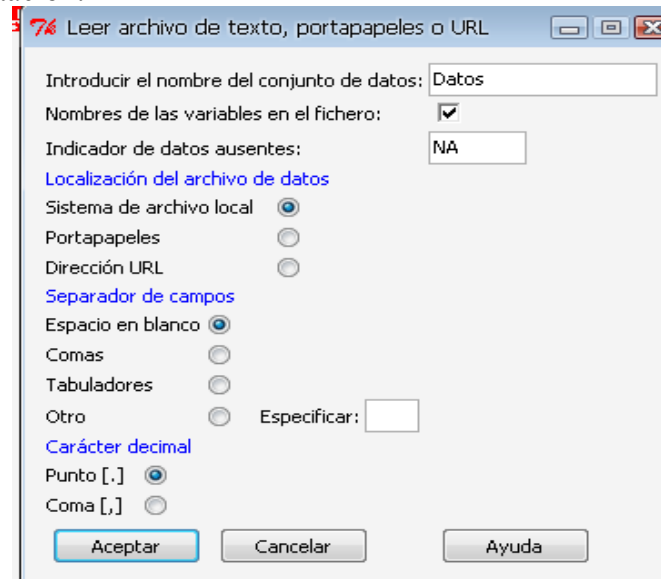
Datos > Importar datos > desde archivo de texto, portapapeles o URL

Necesitamos ser conscientes del formato que tienen los datos en el fichero antes de poder importarlo.

En particular debemos fijarnos en :

1. ¿Contiene la primera fila los nombres de las variables?
2. ¿Hay algún símbolo para indicar datos faltantes?
3. ¿Cuál es el separador de campos? (es decir ¿cómo están separadas las diferentes columnas de datos en el fichero?)
4. ¿Cuál es el separador decimal? (es decir ¿se usa el punto decimal o la coma para indicar decimales?)

Una vez que sabemos contestar a todas estas preguntas para nuestro fichero en particular, podemos rellenar los datos que nos pide R-Commander para iniciar el proceso de importación:



Cambiamos los campos para que correspondan a nuestro fichero de datos, especificando en particular el nombre del conjunto de datos que queremos definir y le damos a Aceptar...

) **3.2 Establecer una columna para los nombres de los individuos (opcional)**

En algunos casos, el fichero contiene una columna que recoge los “nombres” para cada uno de los individuos (es decir los nombres de las filas).

Considere por ejemplo el fichero `proteinas.txt`, que describe la composición de la dieta de varios países.

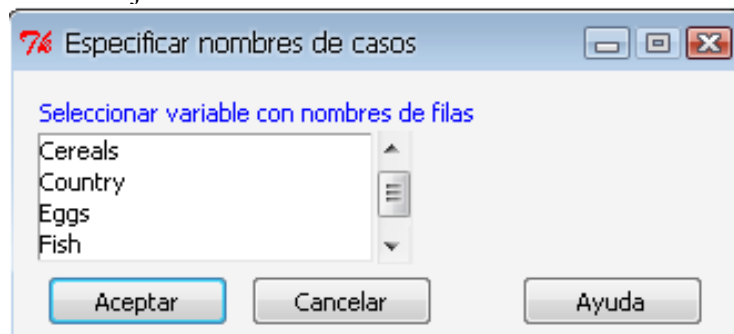
Importar los datos en un conjunto que se llame proteínas

Los individuos son los países, y hay una columna el fichero que contiene sus nombres.

Una vez que hemos importado los datos, vamos a especificar que la columna `Country`, es una columna que contiene los nombres de los individuos.

Siendo `proteinas` el conjunto de datos activo, escogemos

Datos > Conjunto de datos activo > Establecer nombres de caso



Basta con seleccionemos la variable `Country` y le damos a aceptar...

Es interesante hacerlo cuando es posible, porque se identificarán los datos atípicos en las gráficas etc... usando sus nombres y no sus índices, lo que hace mucho más sencilla la interpretación.

. Ejercicios

Para todos los archivos de texto contenidos en el fichero comprimido `datos_para_importar.zip` (descargable desde el Aula virtual), importar los datos (se podrá por ejemplo, usar para el conjunto de datos un nombre parecido al fichero que se importa).