

Departamento de Estadística e I.O.

Máster en Estadística Aplicada



**UNIVERSIDAD
DE GRANADA**

**MODELOS DE RESPUESTA DISCRETA
APLICACIONES BIOSANITARIAS**

**Tema 1 de prácticas
Introducción al programa R**

Profesores

Ana María Aguilera del Pino

Manuel Escabias Machuca

Título original: Modelos de Respuesta Discreta. Aplicaciones Biosanitarias.
Tema 1 de prácticas: Introducción al programa R

© Los profesores

Todos los derechos reservados. Esta publicación es de uso personal del alumno y no puede ser reproducida, ni registrada, ni transmitida en ninguna forma ni por ningún medio, sin el permiso de los autores

Índice general

1. Introducción al programa R	1
1.1. Introducción	1
1.2. Descarga e instalación de R	1
1.3. Tecleando las primeras sentencias	4
1.4. Directorio de trabajo y ayuda	5
1.5. Conceptos básicos de R	5
1.5.1. Vectores	6
1.5.2. Matrices	6
1.5.3. Data Frame	8
1.5.4. Listas	8
1.5.5. Factores	8
1.5.6. Funciones	9
1.6. Importar y exportar datos	10
1.6.1. El formato CSV	10
1.7. Cálculo de probabilidades y cuantiles de distribuciones con R .	13
1.8. Relación de ejercicios	14

Capítulo 1

Introducción al programa R

1.1. Introducción

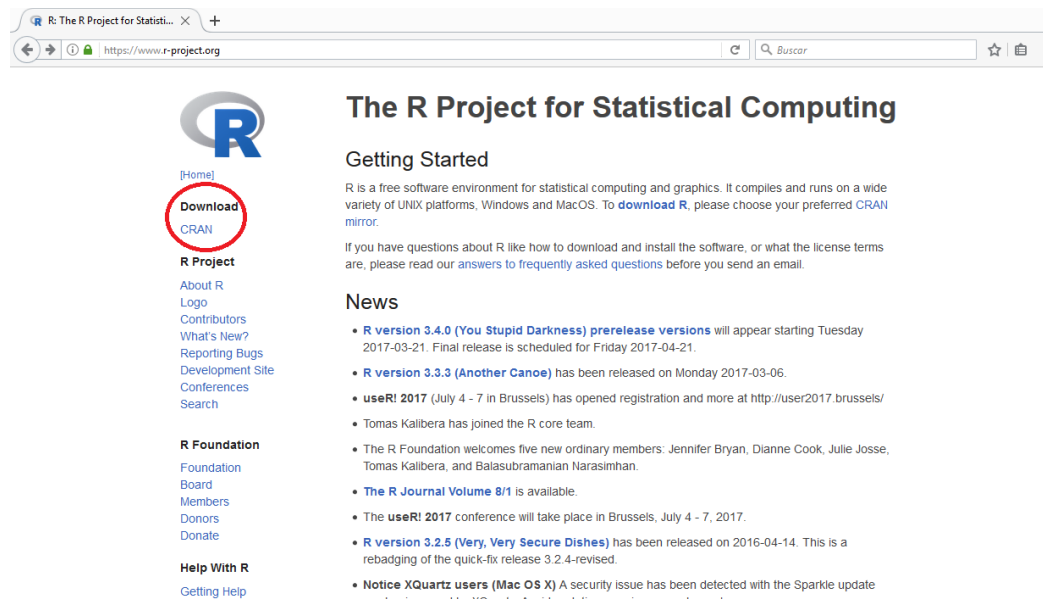
R es un lenguaje de programación y un entorno para el análisis estadístico y gráfico, libre, gratuito y multiplataforma (está disponible para Windows, Macintosh y Linux). En los últimos años se ha convertido en un software muy popular no sólo en la universidad y la docencia, sino también en el mundo de la empresa en donde empresas como Google, Facebook, Twitter, Microsoft, IBM... están utilizando R en sus desarrollos. Una de las características que hacen muy interesante este software es que R es ampliamente extensible mediante funciones y paquetes que crea la comunidad de usuarios.

El ser tan popular ha hecho que surjan multitud de páginas y tutoriales que enseñan a utilizar este software. Una de ellas es *Curso de Introducción a R* impartido por Vicente Coll Serrano, profesor Titular de Universidad del Área de Métodos Cuantitativos para la Economía y la Empresa (Departamento de Economía Aplicada) de la Universidad de Valencia, junto con Pedro J. Pérez Vázquez, profesor Titular de Universidad del Departamento de Análisis Económico de la Universidad de Valencia. El curso completo se puede seguir con explicaciones detalladas en la dirección: https://www.uv.es/vcoll/curso_r.html

En este documento se muestran una pequeña introducción con unos primeros pasos para comenzar a trabajar con R. Algunas de las explicaciones han sido extraídas del curso anteriormente indicado.

1.2. Descarga e instalación de R

La descarga e instalación de R se puede llevar a cabo desde la página web de R project: <http://www.r-project.org>.



The R Project for Statistical Computing

Getting Started

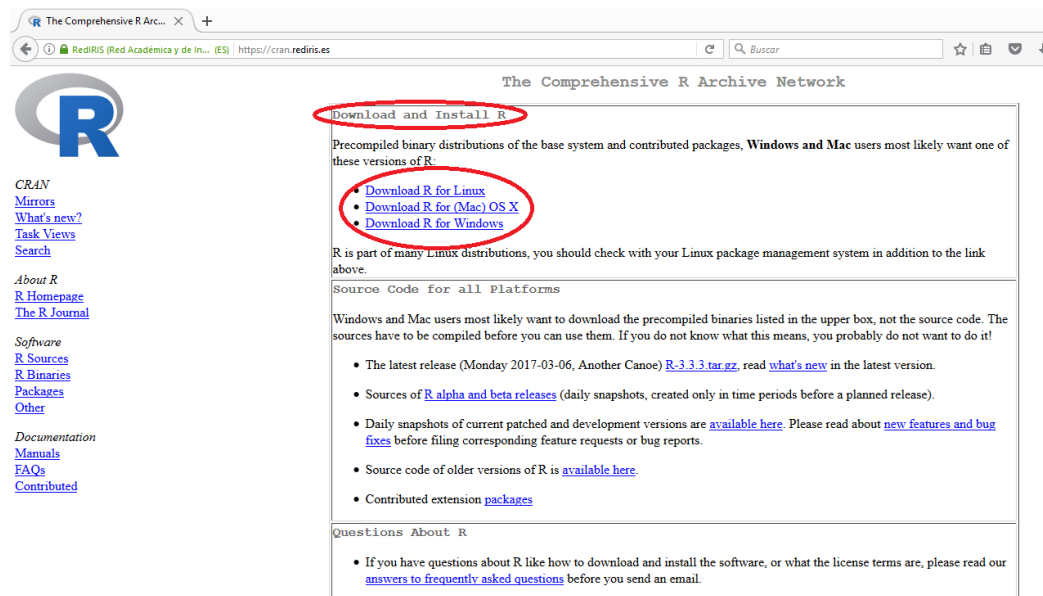
R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 3.4.0 (You Stupid Darkness) prerelease versions** will appear starting Tuesday 2017-03-21. Final release is scheduled for Friday 2017-04-21.
- **R version 3.3.3 (Another Canoe)** has been released on Monday 2017-03-06.
- **useR! 2017** (July 4 - 7 in Brussels) has opened registration and more at <http://user2017.brussels/>
- Tomas Kalibera has joined the R core team.
- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- **The R Journal Volume 8/1** is available.
- The **useR! 2017** conference will take place in Brussels, July 4 - 7, 2017.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.

Para descargar la aplicación debemos hacer clic en *Cran* y pinchar sobre el enlace del repositorio más próximo, por ejemplo Spain. Spanish National Research Network, Madrid (<http://cran.rediris.es/>).



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

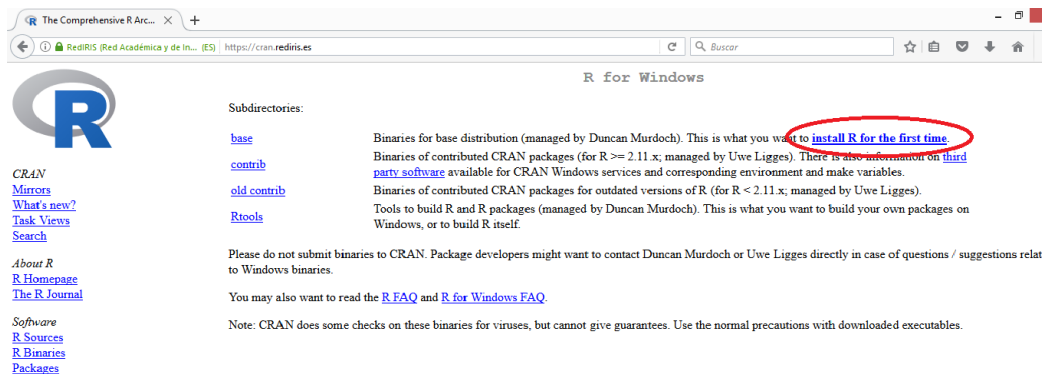
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Monday 2017-03-06, Another Canoe) [R-3.3.3.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

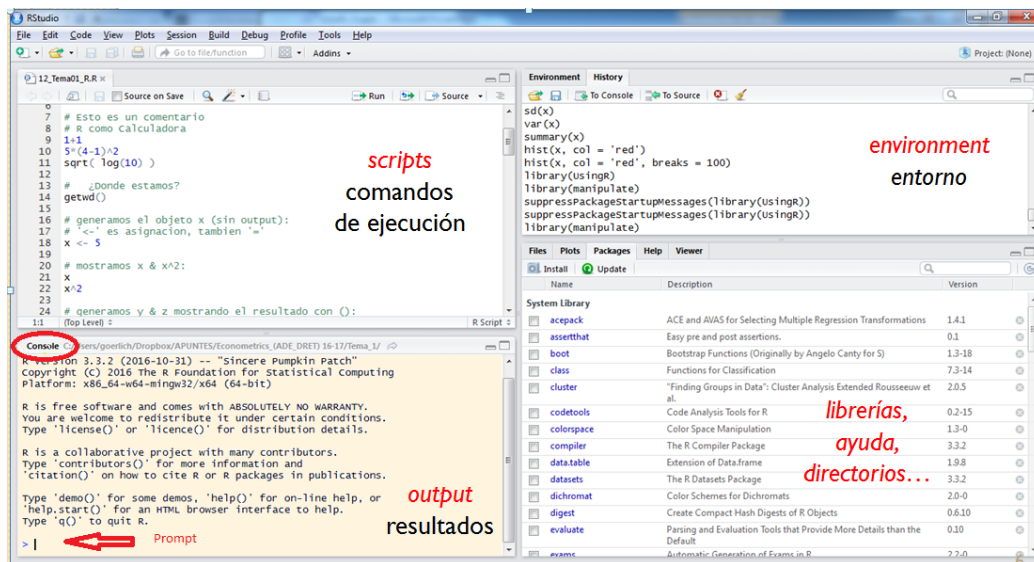
En función del sistema operativo, hemos de seleccionar la correspondiente opción. Para Windows, al hacer clic sobre Download R for Windows se tiene la siguiente imagen



En la siguiente ventana, hemos de hacer clic sobre Download R 4.0.3 for Windows (a fecha 14 de Octubre de 2020), guardar el archivo de instalación y ejecutarlo para proceder a la instalación de R.

La interfaz de usuario (GUI, Graphical User Interface) de R no es muy amigable ni versátil. RStudio es un programa que permite interactuar con R de forma más amigable, facilitando muchas de las tareas de programación y análisis de datos en R. Para Instalar RStudio, se descarga la aplicación desde la página web de RStudio según el sistema operativo, se ejecuta y se siguen las indicaciones. Antes de la instalación de RStudio, hemos de tener instalado R.

RStudio está (normalmente) dividido en 4 paneles.



La consola se encuentra en el panel inferior-izquierdo, y es el espacio para ejecutar sentencias y donde se muestran los resultados de estas ejecuciones.

Existen muchas formas de trabajar con R y RStudio. La forma más elemental y fácil para iniciarse en el manejo del programa es a través de la consola aunque pronto se puede pasar al uso de *Scripts* de R que se explicarán más adelante.

1.3. Tecleando las primeras sentencias

Al abrir RStudio se pueden comenzar a ejecutar sentencias en la consola a modo de calculadora. Al pulsar la tecla intro tras la escritura de las sentencias, se muestra el resultado.

Ejemplo:

- $2+5$
- $3*7$
- $8/2$
- $3^4 : 3^4$
- $\sqrt{9} : \text{sqrt}(9)$. `sqrt()` es una función que calcula la raíz cuadrada a la cantidad que hay dentro del paréntesis
- $\log 5 : \log(5)$. `log()` es una función que calcula el logaritmo neperiano a lo que hay dentro del paréntesis. Si en lugar de un logaritmo neperiano se quiere calcular un logaritmo decimal la función es `log10()`

Trabajar en la consola es muy limitado ya que las instrucciones se han de introducir una a una. Lo habitual es trabajar con scripts o ficheros de instrucciones. Estos ficheros tienen extensión `.R` y se pueden guardar.

El panel del script se sitúa en la parte superior-izquierda de RStudio, las instrucciones se escriben línea por línea y se pueden ejecutar una a una o en bloque. Para ejecutar una línea es suficiente con seleccionarla con el ratón (o colocar el cursor en la línea a ejecutar) y pulsar la tecla *Run* (o Ctrl+ Intro)

Ejemplo. Escribimos las sentencias siguientes en un nuevo *Script* vacío y ejecutamos una a una:

```
2+5
3*7
8/2
3^4
sqrt(9)
log(5)
```

1.4. Directorio de trabajo y ayuda

Como se ha indicado previamente, existen muchas formas de trabajar con R y con RStudio. Una de las más habituales, fácil de comprender y útil para comenzar a familiarizarse con el programa, consiste en establecer el entorno de trabajo como primer paso.

El entorno de trabajo (o *Workspace*) es una carpeta o directorio de nuestro ordenador. Para cada proyecto que comencemos fijar en el programa RStudio cuál es el espacio de trabajo. Para ello desde el mismo menú de RStudio en las opciones **Session->Set Working directory->Choose directory** (o bien con las teclas Shift+Ctrl+h) y seleccionando en el menú la carpeta que elijamos.

La fijación del entorno de trabajo es importante entre otras cosas para guardar por defecto los scripts que creamos o bien para que nuestros scripts encuentren objetos (funciones, conjuntos de datos, etc...) que utilicen en un determinado procedimiento.

Una alternativa es trabajar con proyectos de R, pero esto se dejará para usuarios avanzados.

Ayuda en R. Para obtener ayuda sobre cómo funciona una determinada función, cuáles son sus argumentos, etc, hay varias sentencias que se pueden ejecutar tanto en un script como en la consola: **help()** poniendo dentro del paréntesis el nombre de la función, escribiendo **?** y a continuación el nombre de la función o bien escribiendo el nombre de la función y pulsando la tecla F1

1.5. Conceptos básicos de R

Se pretende aquí mostrar la filosofía y la práctica del trabajo con R.

- R trabaja tecleando sentencias y mostrando el resultado en pantalla
- Se puede trabajar con números, operadores básicos (+, -, *, /...) y funciones (log(), sqrt(), sin(), cos(), exp(), round(),...) e incluso algunas constantes (pi).
- Es importante conocer cómo funcionan las funciones, qué parámetros necesitan (las cosas que se ponen dentro del paréntesis) y qué resultado muestran. A veces la función puede utilizar más de un parámetro en cuyo caso suelen escribirse separados por comas.

- **Asignación.** El símbolo `<-` es el operador para asignar. En la filosofía de R, debemos asignar un nombre a un elemento que vayamos a utilizar con frecuencia.
- En la lógica de R, todo elemento asignado se llama *objeto*. Existen muchos tipos de objetos: vectores, matrices, listas, `data_frame`, o funciones

A continuación se describen algunos de estos tipos de objetos

1.5.1. Vectores

Para crear un vector se utiliza la función `c()`. Por ejemplo: `Conjunto1 <- c(1.5,2.7,3.1,4.6)` crea un objeto denominado `Conjunto1` que es de tipo vector y que tiene cuatro elementos. Para ver su contenido es suficiente con escribir su nombre bien en la consola, bien en un script.

Los vectores se pueden...

- Crear y asignar
- Operar (sumar, restar, multiplicar, dividir, ordenar,...)
- Averiguar su longitud con la función `length()`
- Averiguar el elemento que ocupa una cierta posición
- ...

Se puede acceder al contenido de una posición concreta. La sentencia es `Nombre_del_objeto[posicion]`. Por ejemplo `Conjunto1[2]`

Los datos para hacer estadística de una variable suelen ser vectores. Algunos procedimientos de R devuelven objetos en formato de vector.

1.5.2. Matrices

Para crear una matriz se utiliza la función `matrix()` indicando: los elementos de la matriz (mediante un vector), el número de filas que tiene y el número de columnas. Por ejemplo:

- `Matriz1<-matrix(c(1,2,3,4,5,6),nrow=2,ncol=3)` crea la matriz

$$\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

- `Matriz2<-matrix(c(1,2,3,4,5,6),nrow=2,ncol=3,byrow=T)` crea la matriz

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

- `Matriz3<-matrix(c(1,2,3,4,5,6),nrow=3,ncol=2)` crea la matriz

$$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

- `Matriz4<-matrix(c(1,2,3,4,5,6),nrow=3,ncol=2,byrow=T)` crea la matriz

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Las matrices se pueden...

- Crear y asignar
- Operar (sumar, restar, multiplicar, dividir, ordenar,...)
- Averiguar su dimensión (número de filas y de columnas) con la función `dim()`
- Averiguar el elemento que ocupa una cierta posición
- ...

Se puede acceder al contenido de una posición concreta. La sentencia es `Nombre_del_objeto[filas,columna]`. Por ejemplo `Matriz4[2,1]`

Se puede acceder al contenido de una fila. La sentencia es `Nombre_del_objeto[filas,]`. Por ejemplo `Matriz3[2,]`. El contenido de una fila, R lo toma como un vector.

Se puede acceder al contenido de una columna. La sentencia es `Nombre_del_objeto[,columna]`. Por ejemplo `Matriz2[,2]`. El contenido de una columna, R lo toma como un vector.

Se puede asignar nombres a las filas y a las columnas. Por ejemplo `Matriz2<-matrix(c(1,2,3,4,5),dimnames=list(c("Fila1","Fila2"),c(Columna1", Columna2",Columna3")))`

Algunos métodos estadísticos se hacen con matrices, y algunos otros devuelven objetos de tipo matriz.

La ayuda para la función, se obtiene con `?matrix`

1.5.3. Data Frame

El problema que presentan las matrices para un trabajo estadístico eficaz con R, es que el contenido de las matrices debe ser del mismo tipo en todas las posiciones: siempre números, siempre letras o siempre palabras.

Habitualmente un conjunto de datos para la estadística es una tabla configurada en filas y columnas donde cada columna representa una variable (y cada una puede ser de diferente tipo: numérica, carácter...) y cada fila un caso.

Habitualmente cuando se carga un conjunto de datos en R, el objeto que resulta es un Data Frame.

Las cuestiones de acceso a posiciones concretas, filas, columnas, etc. funcionan de manera similar a las matrices.

Sin embargo el `data.frame` permite más funcionalidades.

Algunas funcionalidades:

- Para acceder a los elementos de una columna (variable) se usa el símbolo `$` separando el nombre del data frame y de la columna.
- Se puede añadir una columna (variable) `Nombre_dataframe$Nombre_variable<-vector` con el contenido
- Ver el contenido de un data frame: `head()` (primeras filas) y `tail()` (últimas filas)
- Ver/trabajar con los nombres de un data.frame: `names()`. El resultado es un vector de nombres.
- Selección de datos (Subsetting): `subset()`.

1.5.4. Listas

Las listas son vectores que en cada posición albergan cualquier objeto (vector, matriz, data frame...)

Se utiliza el doble corchete `[[]]` para acceder al contenido concreto de una posición de la lista.

Cada elemento de la lista puede tener un nombre y utilizar la sentencia `names()` para trabajar con los nombres.

1.5.5. Factores

Tipo de dato que se utiliza para representar variables de naturaleza categórica. Pueden ser ordenados o no ordenados. Si no se indica nada se ordenan alfabéticamente aunque se puede cambiar el orden. Habitualmente

cuando se carga un conjunto de datos en el que aparece una columna con al menos un carácter, el programa R considera toda la columna como una variable de tipo factor.

Para realizar un análisis de regresión con variables categóricas es conveniente guardar las variables categóricas como factores (R codificará internamente los distintos niveles del factor como enteros).

1.5.6. Funciones

El programa R permite crear funciones propias que faciliten realizar determinados cálculos que se repiten con cierta frecuencia. Para ilustrar cómo escribir tales funciones, guardarlas y usarlas vamos a utilizar un ejemplo.

Supongamos que queremos crear una función que calcule el cociente de ventajas a partir de una matriz 2×2 que representa una tabla de frecuencias:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

Está claro que el cociente de ventajas se calcula de la forma

$$\frac{A \times C}{B \times D}$$

Vamos a crear una función llamada `Cociente.Ventajas` para su cálculo:

```
Cociente.Ventajas<-function(Matriz)
{
  Resultado<-(Matriz[1,1]*Matriz[2,2])/(Matriz[1,2]*Matriz[2,1])
  return(Resultado)
}
```

Este código se escribe en un script de R, se guarda en el espacio de trabajo en el que se vaya a utilizar (con extensión `.R`), y para que pueda ser utilizada, se pulsa la tecla *source* de la ventana. Este último paso sólo es necesario realizarlo una vez en cada sesión de R.

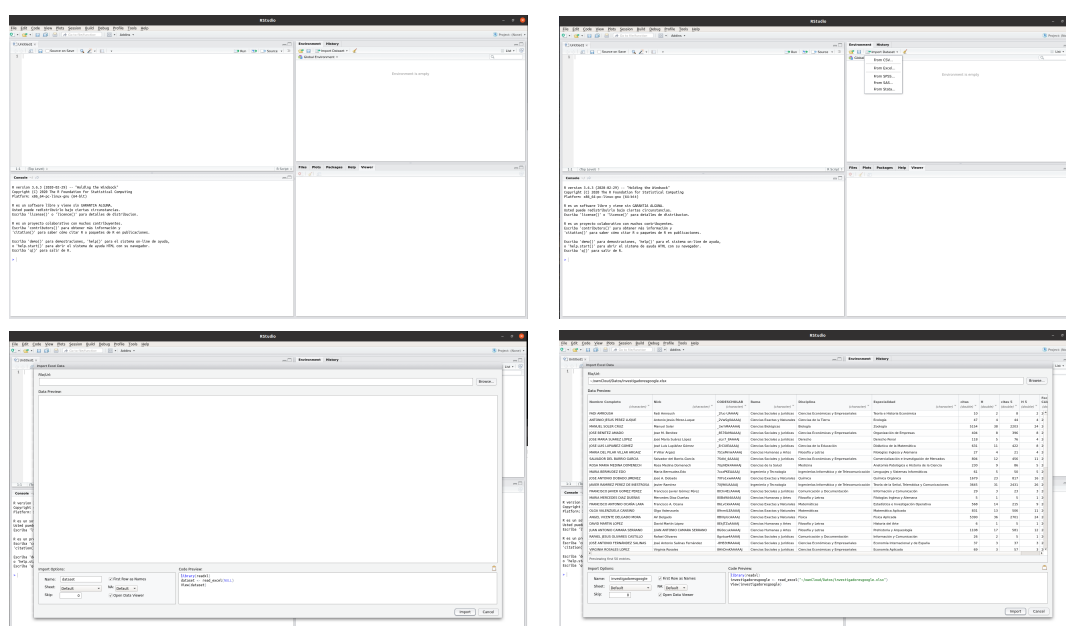
De esta manera la función `Cociente.Ventajas()` es una función más de R que se puede utilizar.

Para utilizar la función, supongamos que tenemos definida una matriz 2×2 llamada `Matriz1`, entonces bastaría ejecutar la sentencia `Cociente.Ventajas(Matriz1)` y nos devolverá como resultado el valor de la operación que hemos definido.

Hay que tener en cuenta que si `Matriz1` es una matriz que tiene menos de dos filas o menos de dos columnas, la ejecución de la función nos proporcionará un error. Si por contra tiene más de dos filas y más de dos columnas, nos devolverá un valor numérico, aunque éste no será un cociente de ventajas pues la matriz original no es 2×2 .

1.6. Importar y exportar datos

Para el trabajo de análisis estadístico de datos hemos de estar familiarizados con los distintos formatos de datos que existen. Es habitual encontrar un conjunto de datos en el formato de Hoja de Excell, pero también en formato de otros software estadísticos como SPSS, SAS o Stata. RStudio permite la importación de un conjunto de datos en los formatos indicados anteriormente a base de clicks de ratón.



El resultado es un data.frame con los datos importados.

Al usar los clicks de ratón, para importar datos en realidad se está llamando a unas funciones que son las que importan realmente los datos. Esas funciones suelen pertenecer a paquetes no básicos de R, y puede que la primera vez que se utilizan, el programa RStudio solicite instalar tales paquetes. Bastaría con aceptar los pasos que se proponen y ya estaríamos listos. Estos paquetes suelen ser `readr` y `haven`.

1.6.1. El formato CSV

Uno de los formatos de datos más popular en el trabajo Estadístico es el formato `.csv`. El formato `.csv` significa *Comma Separated Values*. Es un fichero de tipo texto plano organizado en filas y columnas donde una columna se separa de la siguiente utilizando un carácter que puede ser la coma, el

espacio en blanco, el tabulador, el punto y coma o bien otro carácter como & o \$ o cualquier otro.

Los ficheros de tipo `.csv` suelen utilizarse en Estadística para conjuntos de datos porque ocupan poca memoria para la cantidad de información que suelen contener. Por defecto, cuando se hace doble click sobre un fichero csv para abrirlo, la mayoría de los sistemas operativos los abren con Excell, a pesar de no ser ficheros de Excell. Así mismo, un fichero creado con el programa Excell puede guardarse en formato csv.

Las principales características que hay que observar en un fichero csv antes de proceder a importarlo a R son:

- Si aparecen los nombres de las variables en la primera fila
- Carácter separador de campos
- Carácter decimal

A continuación se muestran algunos ejemplos de conjuntos de datos de tipo csv.

```

1 Edad,Sexo,TiempoConexion,Interes,Dias,Institucion
2 43,Mujer,185.82,Moderado,76,Mixta
3 50,Mujer,154.17,Bajo,66,Publica
4 21,Mujer,200.14,Bajo,69,Publica
5 21,Mujer,224.85,Moderado,71,Mixta
6 58,Hombre,147.78,Bajo,72,Publica
7 35,Mujer,296.67,Moderado,73,Publica
8 36,Mujer,206.31,Alto,100,Privada
9 37,Hombre,300.02,Bajo,86,Privada
10 46,Mujer,185.97,Bajo,58,Publica
11 53,Hombre,194.87,Bajo,114,Privada
12 22,Hombre,256.49,Bajo,55,Publica
13 58,Mujer,166.83,Bajo,110,Publica
14 49,Hombre,184.19,Moderado,53,Privada
15 56,Mujer,201.34,Moderado,101,Publica
16 58,Mujer,241.07,Bajo,44,Publica
17 45,Hombre,174.67,Alto,98,Mixta
18 31,Mujer,225.25,Moderado,102,Publica
19 56,Hombre,222.43,Moderado,68,Privada
20 54,Mujer,211.7,Alto,48,Mixta
21 50,Hombre,177.72,Moderado,66,Privada
22 38,Hombre,141.29,Bajo,62,Publica
23 33,Hombre,174.66,Moderado,72,Mixta

```

a

```

1 Edad,Sexo,TiempoConexion,Interes,Dias,Institucion
2 47,Mujer,202.98,Moderado,115,Mixta
3 37,Mujer,104.22,Bajo,55,Publica
4 50,Hombre,168.9,Moderado,73,Publica
5 39,Hombre,115.07,Alto,86,Mixta
6 50,Hombre,220.02,Moderado,94,Publica
7 49,Mujer,186.86,Bajo,71,Publica
8 39,Mujer,174.7,Moderado,50,Privada
9 22,Mujer,186.33,Moderado,88,Privada
10 31,Hombre,96.87,Moderado,109,Publica
11 37,Hombre,198.79,Moderado,66,Mixta
12 28,Mujer,153.89,Bajo,126,Publica
13 31,Hombre,111.73,Bajo,88,Publica
14 37,Mujer,138.62,Alto,108,Mixta
15 29,Mujer,154.8,Bajo,36,Publica
16 38,Mujer,120.36,Alto,54,Publica
17 45,Mujer,257.8,Alto,113,Mixta
18 21,Hombre,141.93,Moderado,106,Privada
19 50,Hombre,128.02,Bajo,49,Privada
20 40,Mujer,321.44,Moderado,78,Mixta
21 25,Hombre,128.67,Bajo,78,Privada

```

b

```

1 Edad,Sexo,TiempoConexion,Interes,Dias,Institucion
2 55,Hombre,198.09,Moderado,90,Privada
3 43,Mujer,119.71,Bajo,87,Publica
4 48,Mujer,167.21,Bajo,75,Publica
5 46,Mujer,203.21,Alto,87,Privada
6 53,Mujer,227.31,Bajo,102,Privada
7 36,Hombre,207.14,Bajo,52,Publica
8 21,Hombre,229.7,Alto,71,Mixta
9 42,Hombre,153.54,Moderado,67,Publica
10 50,Hombre,157.29,Moderado,107,Publica
11 51,Hombre,264.22,Moderado,78,Mixta
12 35,Mujer,240.32,Moderado,75,Publica
13 45,Mujer,168.68,Bajo,114,Publica
14 33,Hombre,232.67,Moderado,114,Publica
15 58,Hombre,155.33,Moderado,76,Privada
16 36,Hombre,154.73,Bajo,63,Privada
17 33,Mujer,174.64,Moderado,94,Privada
18 33,Hombre,253.29,Moderado,63,Publica
19 35,Mujer,183.75,Bajo,84,Publica
20 37,Hombre,168.97,Alto,60,Mixta
21 59,Hombre,242.7,Moderado,76,Publica

```

c

```

1 36,Mujer,188.94,Alto,70,Privada
2 31,Mujer,175.26,Moderado,93,Privada
3 23,Mujer,211.81,Bajo,43,Publica
4 58,Mujer,247.56,Moderado,76,Mixta
5 25,Hombre,270.63,Bajo,72,Privada
6 53,Hombre,203.03,Bajo,120,Publica
7 51,Mujer,182.66,Moderado,108,Mixta
8 24,Hombre,221.03,Alto,68,Publica
9 35,Mujer,140.13,Bajo,99,Publica
10 57,Mujer,200.16,Moderado,57,Mixta
11 27,Hombre,121.95,Bajo,83,Publica
12 21,Mujer,223.64,Bajo,73,Publica
13 23,Mujer,197.78,Moderado,85,Mixta
14 29,Hombre,205.88,Bajo,107,Privada
15 20,Hombre,215.07,Bajo,132,Privada
16 28,Mujer,193.89,Alto,85,Mixta
17 56,Mujer,141.57,Bajo,70,Publica
18 27,Hombre,177.31,Bajo,81,Publica
19 52,Hombre,187.62,Moderado,68,Mixta
20 38,Mujer,271.79,Bajo,90,Publica
21 22,Mujer,261.8,Bajo,59,Publica

```

d

Las principales características de los ficheros anteriores son:

- a
 - Nombres de las variables: Sí
 - Separador de campos: coma
 - Carácter decimal: punto
- b
 - Nombres de las variables: Sí
 - Separador de campos: punto y coma
 - Carácter decimal: punto
- c

- Nombres de las variables: Sí
- Separador de campos: punto y coma
- Carácter decimal: coma
- d
 - Nombres de las variables: No
 - Separador de campos: coma
 - Carácter decimal: punto

1.7. Cálculo de probabilidades y cuantiles de distribuciones con R

R tiene implementadas funciones para el cálculo de probabilidades y cuantiles de las principales distribuciones de probabilidad. A continuación se muestran algunas de ellas.

Distribución Normal Si X es una distribución normal de media μ y desviación estándar σ , ($X \rightsquigarrow N[\mu, \sigma]$)

- La probabilidad $P\{X \leq a\}$ se calcula con la sentencia `pnorm(a, mean= μ , sd= σ , lower.tail=T)`. Se pueden omitir las palabras `mean=`, `sd=` y `lower.tail=`, en cuyo caso la función considera los valores que se introducen según la posición que ocupan.
- La probabilidad $P\{X > a\}$ se calcula con la sentencia `pnorm(a, mean= μ , sd= σ , lower.tail=F)`.
- El cuantil a tal que la probabilidad $P\{X \leq a\} = p$ se calcula con la sentencia `qnorm(p, mean= μ , sd= σ , lower.tail=T)`.
- El cuantil a tal que la probabilidad $P\{X > a\} = p$ se calcula con la sentencia `qnorm(p, mean= μ , sd= σ , lower.tail=F)`.

Distribución Chi-Cuadrado Si X es una distribución Chi-Cuadrado con n grados de libertad ($X \rightsquigarrow \chi_n^2$)

- La probabilidad $P\{X \leq a\}$ se calcula con la sentencia `pchisq(a, df= n , lower.tail=T)`. Se pueden omitir las palabras `df=` y `lower.tail=`, en cuyo caso la función considera los valores que se introducen según la posición que ocupan.

- La probabilidad $P\{X > a\}$ se calcula con la sentencia `pchisq(a,df=n,lower.tail=F)`.
- El cuantil a tal que la probabilidad $P\{X \leq a\} = p$ se calcula con la sentencia `qchisq(p,df=n,lower.tail=T)`.
- El cuantil a tal que la probabilidad $P\{X > a\} = p$ se calcula con la sentencia `qchisq(a,df=n,lower.tail=T)`.

El interesado puede investigar a partir de las ayudas del programa las funciones para el cálculo de probabilidades y cuantiles de otras distribuciones como Binomial, Poisson, Geométrica, T-Student, F-Snedecor, etc. Así mismo, se asume que el lector está familiarizado con las propiedades del cálculo de probabilidades de distribuciones para otro tipo de probabilidades como $P\{X < a\}$, $P\{X \geq a\}$, $P\{a \leq X \leq b\}$ y otro tipo de cuantiles.

1.8. Relación de ejercicios

A continuación se proponen una serie de ejercicios para que el lector pueda practicar e investigar su posible resolución. Ninguno de éstos ejercicios es será evaluable, pero sí es conveniente que un usuario de R sepa dar una respuesta, pues algunas de las operaciones que aquí se muestran pueden aparecer en métodos estadísticos objeto de este curso.

1. Crear un vector llamado "Pesos" con los los siguientes datos: 65, 71, 63, 84, 69, 75, 77, 85.
 - aplica la sentencia `sort()` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `order()` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `2.4+Pesos` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `1.7*Pesos` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - Cree una matriz 2×4 llamada "MatrizPesos"

$$\begin{pmatrix} 65 & 63 & 69 & 77 \\ 71 & 84 & 75 & 85 \end{pmatrix}$$

- Investiga cómo se crearía la matriz

$$\begin{pmatrix} 65 & 71 & 63 & 84 \\ 69 & 75 & 77 & 85 \end{pmatrix}$$

- Sume las dos matrices. Asígneles un nombre
 - Reste las dos matrices. Asígneles un nombre
 - aplica la sentencia `t()` a "MatrizPesos". ¿Qué se observa? Asígnele un nombre
 - Realiza el producto de matrices que puedas
2. Crear un vector llamado `.Alturas` con los los siguientes datos: 155, 165, 171, 173, 180, 191, 181, 175.
- Calcula el el imc sabiendo que $imc = Peso/Altura^2$ teniendo en cuenta que la altura debe ser en metros (dividida por 100)
 - aplica la sentencia `sort()` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `order()` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `2.4*Alturas` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `1.7*Alturas` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - Cree una matriz 2×4 llamada "MatrizAlturas"

$$\begin{pmatrix} 155 & 171 & 180 & 181 \\ 165 & 173 & 191 & 175 \end{pmatrix}$$

- Investiga cómo se crearía la matriz

$$\begin{pmatrix} 155 & 165 & 171 & 173 \\ 180 & 191 & 181 & 175 \end{pmatrix}$$

- Sume las dos matrices. Asígneles un nombre
- Reste las dos matrices. Asígneles un nombre
- aplica la sentencia `t()` a "MatrizAlturas". ¿Qué se observa? Asígnele un nombre
- Realiza el producto de matrices que puedas

3. multiplica vectorialmente "Pesos" .Alturas"
4. Crear un vector llamado "Provincias" con los los siguientes datos: GR, JA, GR, AL, JA, GR, AL, GR.
 - aplica la sentencia `sort()` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `order()` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `2.4+Provincias` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `1.7*Provincias` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `2.4+sort(Provincias)` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - aplica la sentencia `1.7*order(Provincias)` a dicho vector. ¿Qué se observa? Asígnele un nombre
 - Cree una matriz 2×4 llamada "MatrizProvincias". Cree la misma anterior pero por filas.
 - Cree una matriz 4×2 llamada "MatrizProvincias2". Cree la misma anterior pero por filas.
 - Sume las dos matrices. Asígneles un nombre
 - Reste las dos matrices. Asígneles un nombre
 - aplica la sentencia `t()` a "MatrizProvincias". ¿Qué se observa? Asígnele un nombre
5. En las distribuciones binomial ($X \rightsquigarrow B(n, p)$), Poisson ($X \rightsquigarrow \mathcal{P}(\lambda)$), geométrica ($X \rightsquigarrow G(p)$), e Hipergeométrica ($X \rightsquigarrow H(m, n, k)$), siguientes, calcule la probabilidades y cuantiles que se indican:

$X_1 \rightsquigarrow B(7, 0, 35)$	$X_2 \rightsquigarrow \mathcal{P}(2, 52)$	$X_3 \rightsquigarrow G(0, 6)$	$X_4 \rightsquigarrow H(2, 6, 3)$
$P\{X_1 = 5\}$	$P\{X_2 = 1\}$	$P\{X_3 = 6\}$	$P\{X_4 = 3\}$
$P\{X_1 \leq 4\}$	$P\{X_2 \leq 1\}$	$P\{X_3 \leq 6\}$	$P\{X_4 \leq 3\}$
$P\{X_1 > 2\}$	$P\{X_2 > 3\}$	$P\{X_3 > 5\}$	$P\{X_4 > 4\}$
$P\{X_1 < 3\}$	$P\{X_2 < 2\}$	$P\{X_3 < 7\}$	$P\{X_4 < 4\}$
$P\{X_1 \geq 6\}$	$P\{X_2 \geq 2\}$	$P\{X_3 \geq 8\}$	$P\{X_4 \geq 6\}$
$P\{2 \leq X_1 < 5\}$	$P\{1 \leq X_2 < 3\}$	$P\{4 \leq X_3 < 8\}$	$P\{2 \leq X_4 < 6\}$
$P\{4 < X_1 \leq 6\}$	$P\{2 < X_2 \leq 3\}$	$P\{5 < X_3 \leq 7\}$	$P\{2 < X_4 \leq 5\}$
$P\{1 \leq X_1 \leq 3\}$	$P\{1 \leq X_2 \leq 3\}$	$P\{2 \leq X_3 \leq 5\}$	$P\{4 \leq X_4 \leq 7\}$
$P\{5 < X_1 < 7\}$	$P\{2 < X_2 < 4\}$	$P\{5 < X_3 < 9\}$	$P\{5 < X_4 < 7\}$
a: $P\{X_1 \leq a\} = 0,35$	a: $P\{X_2 \leq a\} = 0,23$	a: $P\{X_3 \leq a\} = 0,19$	a: $P\{X_4 \leq a\} = 0,63$
a: $P\{X_1 < a\} = 0,63$	a: $P\{X_2 < a\} = 0,44$	a: $P\{X_3 < a\} = 0,33$	a: $P\{X_4 < a\} = 0,59$
a: $P\{X_1 > a\} = 0,44$	a: $P\{X_2 > a\} = 0,33$	a: $P\{X_3 > a\} = 0,65$	a: $P\{X_4 > a\} = 0,21$
a: $P\{X_1 \geq a\} = 0,81$	a: $P\{X_2 \geq a\} = 0,64$	a: $P\{X_3 \geq a\} = 0,49$	a: $P\{X_4 \geq a\} = 0,54$

6. En las distribuciones de Normales ($X \rightsquigarrow N[\mu, \sigma]$) y Chi cuadrado ($X \rightsquigarrow \chi_n^2$), calcule la probabilidades y cuantiles que se indican:

$X_1 \rightsquigarrow N[2,5, 1,3]$	$X_2 \rightsquigarrow N[12,4, 4,9]$	$X_3 \rightsquigarrow \chi_5^2$	$X_4 \rightsquigarrow \chi_{8,5}^2$
$P\{X_1 \leq 4,1\}$	$P\{X_2 \leq 10,3\}$	$P\{X_3 \leq 6,9\}$	$P\{X_4 \leq 7,3\}$
$P\{X_1 > 2,5\}$	$P\{X_2 > 13,1\}$	$P\{X_3 > 5,4\}$	$P\{X_4 > 5,4\}$
$P\{X_1 < 3,9\}$	$P\{X_2 < 12,54\}$	$P\{X_3 < 7,2\}$	$P\{X_4 < 6,9\}$
$P\{X_1 \geq 6,3\}$	$P\{X_2 \geq 9,2\}$	$P\{X_3 \geq 6,1\}$	$P\{X_4 \geq 6,1\}$
$P\{2,8 \leq X_1 < 5,4\}$	$P\{8,1 \leq X_2 < 13,9\}$	$P\{4,5 \leq X_3 < 7,8\}$	$P\{8,2 \leq X_4 < 9,6\}$
$P\{4,3 < X_1 \leq 6,5\}$	$P\{10,2 < X_2 \leq 13,5\}$	$P\{3,5 < X_3 \leq 5,7\}$	$P\{7,4 < X_4 \leq 8,5\}$
$P\{1,4 \leq X_1 \leq 3,2\}$	$P\{11,6 \leq X_2 \leq 14,3\}$	$P\{4,2 \leq X_3 \leq 6,5\}$	$P\{8,4 \leq X_4 \leq 10,7\}$
$P\{5,5 < X_1 < 7,9\}$	$P\{12,0 < X_2 < 14,0\}$	$P\{5,6 < X_3 < 7,9\}$	$P\{6,5 < X_4 < 7,4\}$
a: $P\{X_1 \leq a\} = 0,35$	a: $P\{X_2 \leq a\} = 0,23$	a: $P\{X_3 \leq a\} = 0,19$	a: $P\{X_4 \leq a\} = 0,63$
a: $P\{X_1 < a\} = 0,63$	a: $P\{X_2 < a\} = 0,44$	a: $P\{X_3 < a\} = 0,33$	a: $P\{X_4 < a\} = 0,59$
a: $P\{X_1 > a\} = 0,44$	a: $P\{X_2 > a\} = 0,33$	a: $P\{X_3 > a\} = 0,65$	a: $P\{X_4 > a\} = 0,21$
a: $P\{X_1 \geq a\} = 0,81$	a: $P\{X_2 \geq a\} = 0,64$	a: $P\{X_3 \geq a\} = 0,49$	a: $P\{X_4 \geq a\} = 0,54$

7. En una distribución t-student ($Z \rightsquigarrow t_{10}$):

a: $P\{Z \leq a\} = 0,19$	a: $P\{-a \leq Z \leq a\} = 0,33$	a: $P\{-a \leq Z \leq a\} = 0,65$	a: $P\{Z \geq a\} = 0,49$
---------------------------	-----------------------------------	-----------------------------------	---------------------------

8. Crear un fichero .csv con 10 filas y 3 columnas que contenga la edad, el peso y la altura de 10 personas. Utilice el separador de campos y el carácter decimal que decida y el nombre de las variables. Posteriormente impórtelo con RStudio. Puede utilizar una ventana de tipo **text file** de RStudio y guardarla en su espacio de trabajo con extensión .csv