

CURSO DE FORMACIÓN
“MINERÍA DE DATOS CON 'R' ”
UNIVERSIDAD DE VIGO

**UNIDAD PRELIMINAR. BREVE INTRODUCCIÓN A “R”.
INSTALACIÓN Y FUNCIONAMIENTO BÁSICO**

En esta unidad preliminar realizaremos algunas operaciones estadísticas sencillas con el programa R, con el objetivo de que el alumno se familiarice con el entorno y el manejo elemental de los datos.

R es un lenguaje de programación para estadística y matemáticas desarrollado inicialmente por Ihaka y Gentleman en 1997. Toma como base las características del software comercial S-PLUS, es de código abierto y distribución gratuita, y dispone de facilidades para manejo de datos, aplicación de técnicas estadísticas, gráficos, etc.

Aunque al principio puede parecer difícil de manejar, o poco amistoso, esta sensación suele atenuarse después de un par de meses de uso, cuando se empieza a valorar más la flexibilidad, el control de las operaciones y la potencia estadística que la sencillez de manejo.

Su popularidad es cada vez mayor dentro de la comunidad científica. Su éxito se debe principalmente a las siguientes razones: estabilidad, bajo consumo de recursos, dinamismo y gratuidad. R es dinámico porque permite a cualquier usuario programar nuevas aplicaciones (paquetes) y ponerlas a disposición de la comunidad de usuarios; R se desarrolla y evoluciona de forma continua apoyado en una gran comunidad de usuarios que programan y estudian el funcionamiento de sus aplicaciones. Actualmente, su desarrollo corre a cargo del *R Development Core Team*.

Podemos encontrar la página web de R en la dirección: <http://www.r-project.org/>; allí veremos un enlace ([download R](#)) para la instalación de R. El enlace pide una localización (CRAN Mirror): elegimos generalmente la más próxima, por ejemplo Spain <https://ftp.cixug.es/CRAN/> o <https://cran.rediris.es/>.

Nos pide el sistema operativo (Windows, Linux, o MacOS): pinchamos [Download R for Windows](#) → [install R for the first time](#) → [Download R 3.5.3 for Windows](#) (79 mbytes, 32/64 bit)

El archivo de instalación se descargará -guardar archivo- en pocos segundos. El gran número de usuarios que actualmente tiene R en todo el mundo, con miles de accesos simultáneos, causa en ocasiones dificultades en las descargas, y dependiendo de la hora y las características de la red puede ocurrir que el proceso sea lento, y en algunos casos se interrumpa o se produzcan descargas incompletas, por lo que podría ser necesario repetirlo, tanto en la instalación inicial como en la instalación posterior de cualquier paquete adicional que se requiera. Es preciso tener paciencia en esta fase y esperar todo el tiempo necesario.

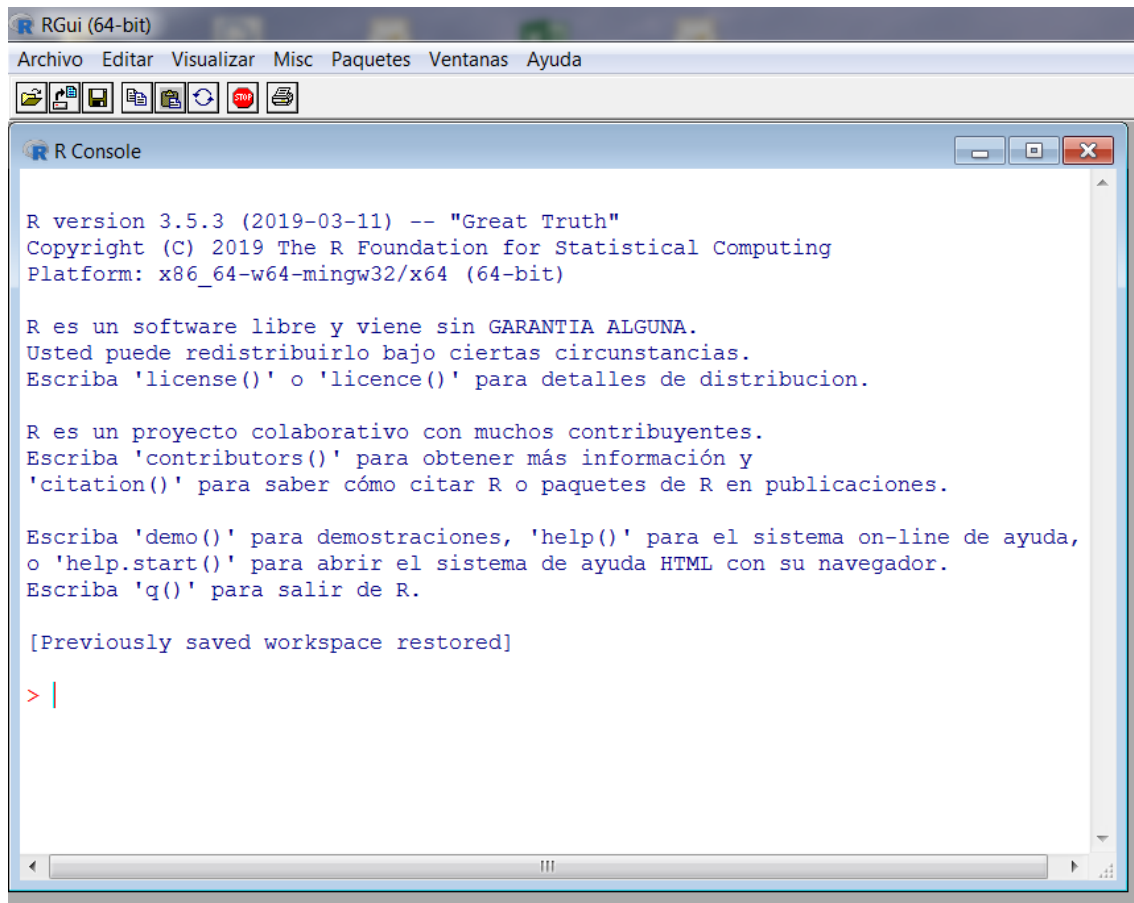
Ejecutamos finalmente el archivo descargado, y se recomienda aceptar todas las opciones por defecto durante la instalación, marcando “crear un icono en el escritorio” sin utilizar otras opciones de configuración; en general es relativamente rápida y no suele plantear problemas. Se instalan dos versiones, con sus correspondientes iconos en el escritorio, una para procesador de 32 bits y otra para 64 bits; en algunos equipos antiguos quizás solo funcione correctamente la primera.

Para la instalación completa de R se requiere la descarga de numerosos archivos, que se realiza de forma automática con las utilidades de instalación. Sin embargo es posible que los sistemas de protección de nuestro ordenador (antivirus, cortafuegos), consideren una amenaza el acceso continuado durante varios minutos y la descarga de ficheros, e impidan o limiten la correcta instalación: puede ser conveniente en ese caso –lo advertiremos porque aparecen sucesivos mensajes de error de instalación- reducir temporalmente el nivel de protección (todos los archivos de R están libres de virus y son completamente fiables) hasta que el programa R quede instalado por completo.

Una vez instalado, R se puede utilizar desde una consola, con una interfaz de línea de instrucciones. Es el método de mayor fiabilidad y estabilidad, pero incómodo y poco atractivo para muchos usuarios. Como alternativa existen diferentes entornos gráficos que facilitan la interacción con el usuario (llamados GUI, *Graphical Users Interface*). Entre los más populares se encuentran *RGui*, *Tinn-R*, *R Commander*, *Emacs*, *RKward*, o *RStudio*. *RGui* se instala de forma automática con el paquete *base* de Windows. El resto de GUIs es necesario instalarlas específicamente, y se encuentran gratuitamente en Internet. En este curso utilizaremos *R Commander* desde un entorno Windows.

Instalado R, se dispone ya de las librerías o paquetes básicos, que permiten muchas operaciones de gestión de datos y técnicas estadísticas; algunas aplicaciones requieren sin embargo paquetes adicionales, que se instalan fácilmente desde la consola. Activaremos R, iniciando la sesión, mediante el icono correspondiente en el escritorio:

Con ello entramos en la ventana principal de R:



A continuación del símbolo “>” podemos escribir las instrucciones de R para realizar cualquier operación. Para instalar un paquete nuevo, como R Commander, utilizaremos la opción de menú Paquetes → Instalar paquetes. Aparece una ventana en la que debemos elegir el servidor o repositorio desde el que queremos hacer la descarga, por ejemplo Spain (A Coruña) o Spain (Madrid); se suele elegir el más próximo, y en general todos ellos funcionan correctamente. Una nueva ventana (puede tardar algunos segundos en aparecer) indica –en orden alfabético– los paquetes disponibles (cerca de diez mil): elegiremos Rcmdr (R commander). Se instalará normalmente en uno o dos minutos, aunque el creciente número de usuarios de R puede hacer que la red funcione con lentitud (debemos ser pacientes si tarda más). También instala automáticamente otros paquetes relacionados (unos 100 para R Commander), por lo que tenemos que esperar hasta que termine. Sabremos que ha terminado cuando aparece de nuevo el símbolo “>” en la consola, indicando que está disponible. Los mensajes de error nos advertirán de cualquier problema (quizás algún paquete no instalado correctamente)

Cualquier otro paquete adicional que se necesite en adelante lo instalaremos del mismo modo. Dependiendo del tamaño del paquete y del estado de la red, algunos pueden requerir un tiempo de varios minutos (incluso horas con una red muy deficiente en horario conflictivo).

La instalación de cada paquete solo se hace una vez, y queda instalado de forma permanente en ese ordenador. Sin embargo es necesario en cada sesión de R “cargar” en la memoria los paquetes que se necesiten (los cuales deberán estar “instalados” previamente). Para ello utilizamos la orden del menú Paquetes → cargar paquete, eligiendo en la lista alfabética de paquetes instalados en nuestro ordenador el que

queremos cargar, Rcmdr. Aparecerá la ventana de R Commander, que utilizaremos en adelante como entorno de usuario en lugar de la ventana principal de R o consola de R.

Alternativamente, en lugar del menú, para cargar el paquete podríamos haber escrito y ejecutado la orden `library(nombre del paquete)`.

(Atención: en R es necesario siempre respetar las mayúsculas y minúsculas)

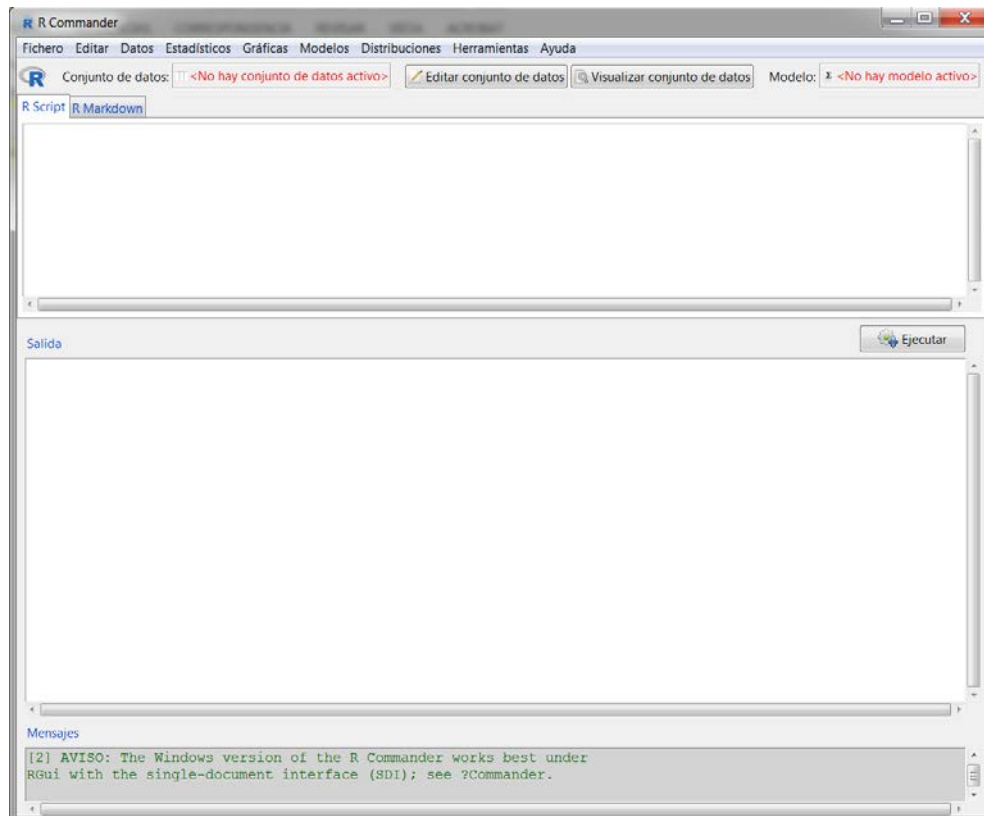
```
> library(Rcmdr)
```

Una vez escrita la orden a continuación del símbolo `>`, se ejecuta pulsando `<ENTER>`.

Cuando se carga un paquete por primera vez puede detectar que no están instalados otros paquetes complementarios que son utilizados por algunas de sus funciones, en cuyo caso preguntará si deseamos instalarlos en ese momento (debemos aceptar, y esperar a que acabe de instalarlos todos). Si no se instalan todos los paquetes relacionados, por el deficiente estado de la red, seguirá intentándolo cada vez que se inicie la sesión con ese paquete.

R Commander, paquete de R que genera instrucciones mediante menú para las operaciones estadísticas más sencillas y frecuentes del mismo modo que lo hacen otros programas estadísticos, facilita la utilización de R a los usuarios de la estadística que no quieren dedicar mucho tiempo a aprender el lenguaje de programación, pero es criticado por muchos expertos en R, que consideran que el aprendizaje de este lenguaje es el mejor modo de sacar partido al programa, y que R Commander genera malos hábitos impidiendo un verdadero progreso en el conocimiento de R. En este curso utilizaremos R Commander, y también muchas instrucciones que no pueden construirse mediante menú, recomendando al mismo tiempo al alumno que se inicie cuanto antes en los elementos de programación.

Por otra parte R Commander, que como todos los paquetes de R es una contribución voluntaria de carácter no comercial, no está optimizado para cualquier configuración del equipo de usuario, y en ocasiones presenta fallos (por ejemplo, se cuelga, no responde, alguna opción de menú no aparece, o no permite navegar por todas las carpetas). En general esos problemas desaparecen al cargar de nuevo R Commander (o R) y volver a ejecutar.

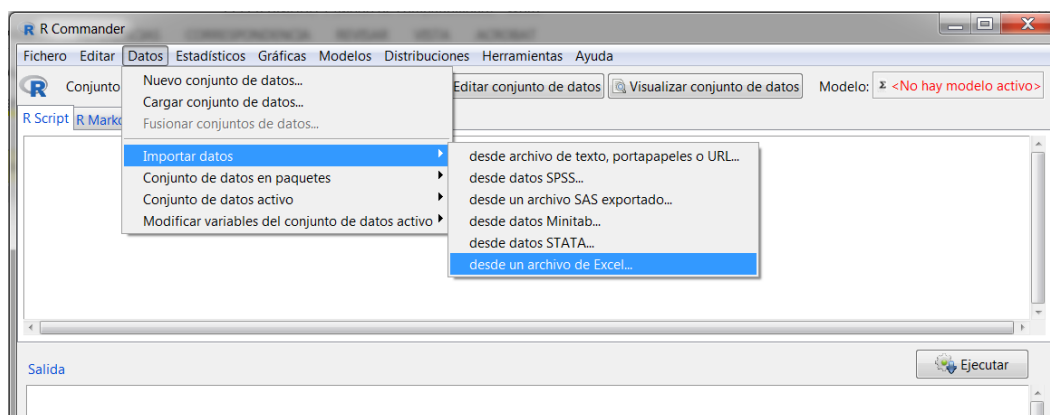


Todos nuestros datos y resultados estarán en un directorio o carpeta determinada, que denominaremos C:\CURSO DM. Lo primero que haremos será crear esa carpeta (con el explorador de Windows), cargando en ella los ficheros de datos que utilizaremos durante el curso, y configurar desde R Commander la carpeta de trabajo para facilitar el intercambio de archivos, mediante la opción Fichero → Cambiar directorio de trabajo, con la cual elegimos aquella carpeta. Alternativamente podríamos haber utilizado, en lugar del menú de R Commander, la orden (lo que hace R Commander es generar y ejecutar esa orden):

```
setwd("C:/CURSO DM")
```

Ahora vamos a leer un conjunto de datos desde un archivo Excel. Para ello utilizamos la opción de menú de R Commander:

Datos → Importar datos → Desde un archivo de Excel



Aparece un pequeño cuadro de diálogo en el que debemos asignar un nombre al conjunto de datos. Aceptamos el que pone por defecto R Commander, "Dataset", y elegimos a continuación el archivo "deudas.xls", hoja de calculo Excel, que habremos situado previamente en la citada carpeta de trabajo. Si queremos tener en R varios conjuntos de datos simultáneamente, debemos asignar nombres diferentes (como DatosA, DatosB, o cualquier otro nombre más descriptivo que no contenga espacios o símbolos aritméticos).

Hemos leído los datos desde un archivo de Excel (xls). En algunos equipos, dependiendo del sistema operativo o de la configuración, puede haber dificultades para leer un archivo directamente con ese formato (por ejemplo, porque no están instalados los drivers necesarios la primera vez que se utiliza). Existen otras alternativas, de las cuales quizás la más práctica consiste en transformar el archivo en formato de texto separado por punto y coma, o formato csv (desde Excel se crea al guardar con la opción o tipo de archivo csv-delimitado por comas). El archivo -del cual se ha añadido una copia en la unidad con el mismo nombre "deudas" - tiene la extensión csv y se lee desde R con la orden siguiente, que puede ser escrita o copiada en la consola de R, pulsando a continuación <ENTER>:

```
> Dataset <- read.csv2("deudas.csv",header=TRUE,encoding="latin1")
```

O bien copiada en la ventana de instrucciones de R Commander (R Script), y se ejecuta marcándola con el ratón (o poniendo el cursor en cualquier punto de la orden) y pulsando <EJECUTAR>

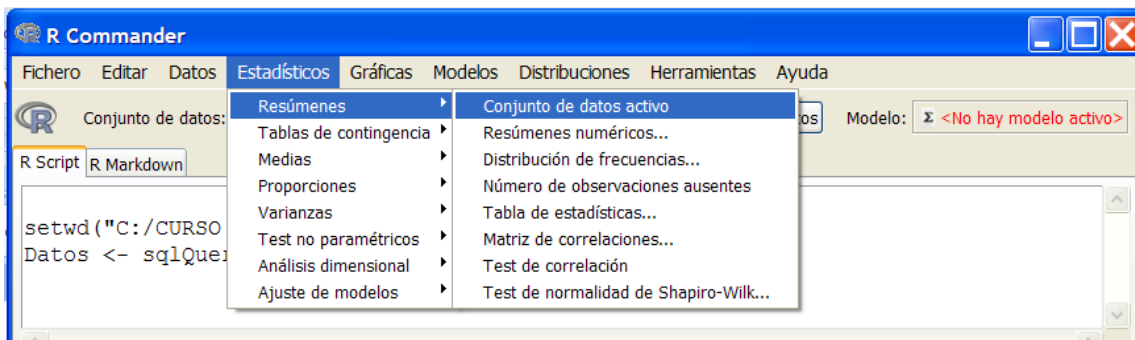
Con la opción de menú "Visualizar conjunto de datos" podremos inspeccionar nuestro archivo. Si en "conjunto de datos", en la parte superior de la ventana de R Commander, indica que no hay conjunto de datos activo, podemos elegir allí el que queremos activar, ya que el menú solo actuará sobre ese conjunto:

	Edad	Formacion	Empleo	Residencia	Ingreso	Deud_ing	Deud_tarj	Deud_otr	Impago
1	mas de 40	estudios universitarios	más de 10 años	más de 10 años	176	9.3	11.36	5.01	Si
2	mas de 40	elemental	más de 10 años	más de 10 años	55	5.5	0.86	2.17	No
3	mas de 40	elemental	más de 10 años	más de 10 años	120	2.9	2.66	0.82	No
4	menos de 30	bachillerato	menos de 5 años	menos de 5 años	28	17.3	1.79	3.06	Si
5	entre 30 y 40	elemental	más de 10 años	entre 5 y 10 años	67	30.6	3.83	16.67	No
6	mas de 40	elemental	más de 10 años	más de 10 años	38	3.6	0.13	1.24	No
7	menos de 30	elemental	menos de 5 años	menos de 5 años	19	24.4	1.36	3.28	Si
8	menos de 30	elemental	menos de 5 años	menos de 5 años	16	1.7	0.18	0.09	No
9	mas de 40	elemental	más de 10 años	más de 10 años	100	9.1	3.70	5.40	No
10	entre 30 y 40	bachillerato	más de 10 años	entre 5 y 10 años	41	16.4	2.92	3.81	Si
11	mas de 40	elemental	más de 10 años	más de 10 años	72	7.6	1.18	4.29	No
12	entre 30 y 40	elemental	entre 5 y 10 años	entre 5 y 10 años	61	5.7	0.56	2.91	No
13	mas de 40	estudios universitarios	menos de 5 años	más de 10 años	26	1.7	0.10	0.34	No
14	entre 30 y 40	elemental	más de 10 años	menos de 5 años	52	3.2	1.15	0.51	No
15	mas de 40	elemental	más de 10 años	más de 10 años	43	5.6	0.59	1.82	No
16	menos de 30	bachillerato	menos de 5 años	menos de 5 años	16	18.0	0.24	2.64	Si
17	mas de 40	bachillerato	entre 5 y 10 años	más de 10 años	69	6.7	0.71	3.92	No
18	mas de 40	elemental	más de 10 años	más de 10 años	64	16.7	0.95	9.74	No
19	mas de 40	elemental	menos de 5 años	más de 10 años	20	2.1	0.11	0.32	No
20	menos de 30	elemental	entre 5 y 10 años	menos de 5 años	35	2.9	0.08	0.94	No
21	menos de 30	bachillerato	menos de 5 años	entre 5 y 10 años	24	17.1	1.34	2.77	Si
22	mas de 40	elemental	más de 10 años	menos de 5 años	50	4.2	0.56	1.54	No
23	menos de 30	elemental	entre 5 y 10 años	menos de 5 años	31	6.6	0.34	1.71	No
24	mas de 40	bachillerato	entre 5 y 10 años	más de 10 años	41	4.6	0.94	0.94	No
25	entre 30 y 40	elemental	más de 10 años	más de 10 años	35	4.5	0.43	1.14	No
26	menos de 30	bachillerato	menos de 5 años	menos de 5 años	21	11.4	0.78	1.62	Si
27	entre 30 y 40	elemental	más de 10 años	menos de 5 años	42	7.4	0.21	2.90	No
28	entre 30 y 40	elemental	menos de 5 años	más de 10 años	25	12.6	0.57	2.58	Si
29	mas de 40	elemental	más de 10 años	más de 10 años	52	12.9	3.03	3.68	Si
30	entre 30 y 40	estudios universitarios	menos de 5 años	más de 10 años	25	3.7	0.31	0.61	No

R ha transformado el archivo Excel en un conjunto de datos (data frame), que consta de casos (filas) y variables (columnas). En este ejemplo se trata de una muestra de 100 clientes de un banco con datos económicos (ingresos, deudas, impagos) y personales (edad, formación, ...)

Realizaremos algunas operaciones estadísticas sencillas con estos datos. En primer lugar un resumen rápido:

Estadísticos → Resúmenes → conjunto de datos activo



```
> summary(Dataset)
```

Edad		Formacion		Empleo		Residencia	
entre 30 y 40:	36	bachillerato	:28	entre 5 y 10 años:	24	entre 5 y 10 años:	21
mas de 40	:37	elemental	:56	más de 10 años	:41	más de 10 años	:39
menos de 30	:27	estudios universit.:	16	menos de 5 años	:35	menos de 5 años	:40

Ingreso		Deud_ing		Deud_tarj		Deud_otr		Impago	
Min.	: 15.00	Min.	: 0.600	Min.	: 0.0300	Min.	: 0.090	No:	68
1st Qu.:	27.00	1st Qu.:	3.550	1st Qu.:	0.3675	1st Qu.:	0.940	Si:	32
Median	: 38.50	Median	: 6.400	Median	: 0.8700	Median	: 1.780		
Mean	: 46.27	Mean	: 8.915	Mean	: 1.2622	Mean	: 2.448		
3rd Qu.:	57.00	3rd Qu.:	12.725	3rd Qu.:	1.5225	3rd Qu.:	3.027		
Max.	:176.00	Max.	:35.300	Max.	:11.3600	Max.	:16.670		

Los resultados, que aparecen en la ventana de “salida” de R Commander, consisten en un recuento o tabla de frecuencias para las variables cualitativas o con valores agrupados (Edad, Formacion, Empleo, Residencia, Impago) y algunas medidas de resumen (mínimo, máximo, media, mediana, cuartiles 1 y 3) para las variables cuantitativas (Ingreso, Deud_ing, Deud_tarj, Deud_otr). En la ventana de instrucciones (“R Script”) de R Commander aparecen las órdenes que el programa genera a partir del menú, que pueden ser ejecutadas de nuevo marcándolas con el cursor y pulsando el botón <ejecutar>. Existe una tercera ventana de “Mensajes”, en la que R Commander advierte de cualquier error. En cualquiera de las tres ventanas de R Commander es posible borrar, modificar o copiar (se trata de ventanas de texto).

La ventana de instrucciones tiene como complemento una utilidad que no emplearemos en este curso denominada *R Markdown* que sirve para transformar el texto de estas ventanas en formato HTML para la construcción de páginas web.

Ahora calcularemos los coeficientes de correlación entre las variables cuantitativas:

Estadísticos → Resúmenes → Matriz de correlaciones

En el cuadro de diálogo que aparece elegimos todas las variables (solo aparecen las cuantitativas, ya que las variables cualitativas o no numéricas no son adecuadas para el cálculo de correlaciones), elegimos el tipo de coeficiente (Pearson) y marcamos “p-valores pareados” para contrastar la significación estadística. Finalmente pulsamos “aceptar”. Ese cuadro tiene también la opción “aplicar”, que ejecuta la orden con el mismo efecto que “aceptar”, pero mantiene abierto el cuadro de diálogo para repetir la orden con opciones diferentes:

```
rcorr.adjust(Dataset[,c("Deud_ing", "Deud_otr", "Deud_tarj", "Ingreso")], type="pearson", use="complete")
```

```
Pearson correlations:
      Deud_ing Deud_otr Deud_tarj Ingreso
Deud_ing  1.0000  0.6548  0.4367 -0.2044
Deud_otr  0.6548  1.0000  0.5456  0.3298
Deud_tarj 0.4367  0.5456  1.0000  0.5370
Ingreso   -0.2044  0.3298  0.5370  1.0000
```

```
Number of observations: 100
```

```
Pairwise two-sided p-values:
      Deud_ing Deud_otr Deud_tarj Ingreso
Deud_ing      <.0001  <.0001  0.0414
Deud_otr      <.0001      <.0001  0.0008
Deud_tarj      <.0001      <.0001      <.0001
Ingreso      0.0414  0.0008      <.0001
```

```
Adjusted p-values (Holm's method)
      Deud_ing Deud_otr Deud_tarj Ingreso
Deud_ing      <.0001  <.0001  0.0414
Deud_otr      <.0001      <.0001  0.0016
Deud_tarj      <.0001      <.0001      <.0001
Ingreso      0.0414  0.0016      <.0001
```

La primera matriz muestra los coeficientes de correlación entre cada par de variables, la segunda y la tercera los valores P y valores P ajustados de los contrastes de nulidad. Un valor P menor que el nivel de significación elegido (habitualmente 0,05 ó 0,01) indica que el coeficiente de correlación correspondiente no es nulo y por lo tanto la correlación es estadísticamente significativa. Deben utilizarse los valores P ajustados para comparaciones múltiples cuando buscamos en toda la matriz las correlaciones significativas, y los no ajustados cuando solo queremos contrastar de antemano un único coeficiente concreto. Observando esta última matriz vemos que todas las variables están correlacionadas entre sí al nivel de significación 0,05.

Veamos si los impagos están relacionados con la edad. Para ello construimos una tabla de contingencia y aplicamos el test de independencia Ji cuadrado:

Estadísticos → Tablas de contingencia → Tabla de doble entrada

En el cuadro de diálogo elegimos la variable de fila “Impago”, y la variable de columna “Edad”. Marcamos en “estadísticos” “porcentajes por columnas” y “test de indepen-

dencia Chi-cuadrado”. En general, cuando tenemos una variable que puede ser explicada por la otra es conveniente utilizar siempre el mismo formato para facilitar la interpretación de los resultados, poniendo la variable dependiente en filas y la explicativa en columnas, e interpretando los porcentajes de columna como en el ejemplo:

```
.Table <- xtabs(~Impago+Edad, data=Dataset); .Table; colPercents(.Table)
.Test <- chisq.test(.Table, correct=FALSE) ; .Test
```

	Edad		
Impago	entre 30 y 40	mas de 40	menos de 30
No	24	32	12
Si	12	5	15

Impago	entre 30 y 40	mas de 40	menos de 30
No	66.7	86.5	44.4
Si	33.3	13.5	55.6
Total	100.0	100.0	100.0
Count	36.0	37.0	27.0

Pearson's Chi-squared test

```
data: .Table
X-squared = 12.725, df = 2, p-value = 0.001725
```

Los resultados muestran que el 55,6% de los clientes con menos de 30 años ha tenido impagos, y estos porcentajes son menores entre los clientes con edad entre 30 y 40 (33,3%) o con más de 40 años (13,5%). El valor P del contraste 0.001725 (menor que el nivel de significación 0,05 o 0,01) indica que las dos variables no son independientes, sino que el impago está relacionado significativamente con la edad.

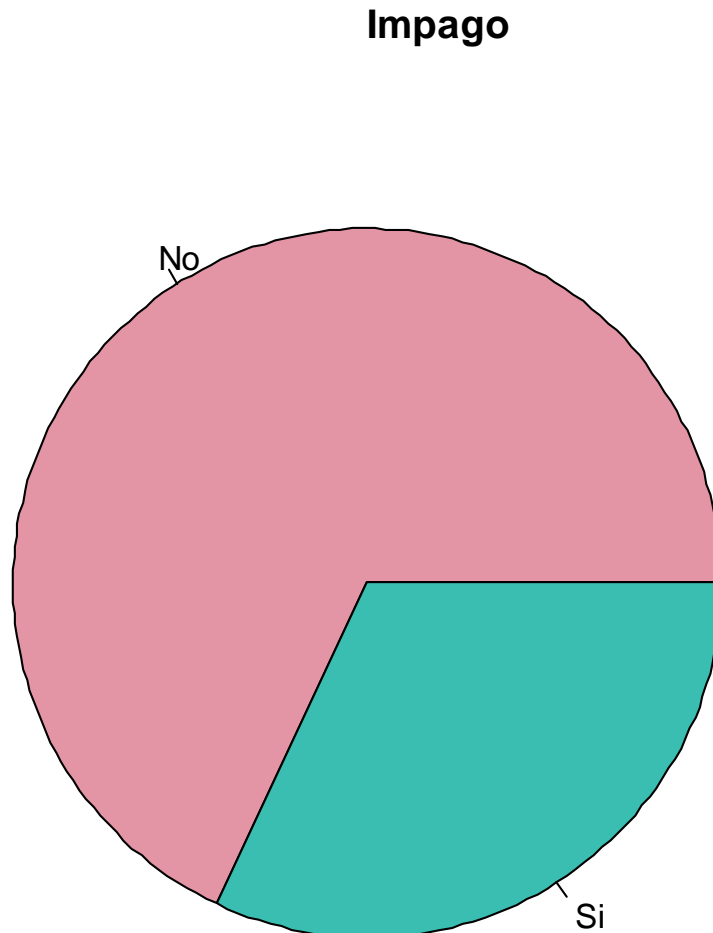
Esta prueba o contraste de independencia, que puede utilizarse con carácter general para valorar la relación entre dos caracteres cualitativos a través de la tabla de frecuencias bidimensionales o tabla de contingencia, requiere por otra parte que el valor esperado de las frecuencias en cada casilla no sea menor que cinco, ya que la distribución del estadístico de contraste es aproximada; cuando no se cumple aparecerá un aviso en color verde en la ventana inferior (lo que no ocurre en este ejemplo).

Ahora haremos un par de gráficos. En primer lugar representamos la variable cualitativa “Impago”, para la cual es adecuado un gráfico de sectores:

Gráficas → Gráfica de sectores

En el cuadro de diálogo elegimos la variable a representar “Impago”

```
with(Dataset, pie(table(Impago), labels=levels(Impago), xlab="", ylab="", main="Impago",  
col=rainbow_hcl(2)))
```



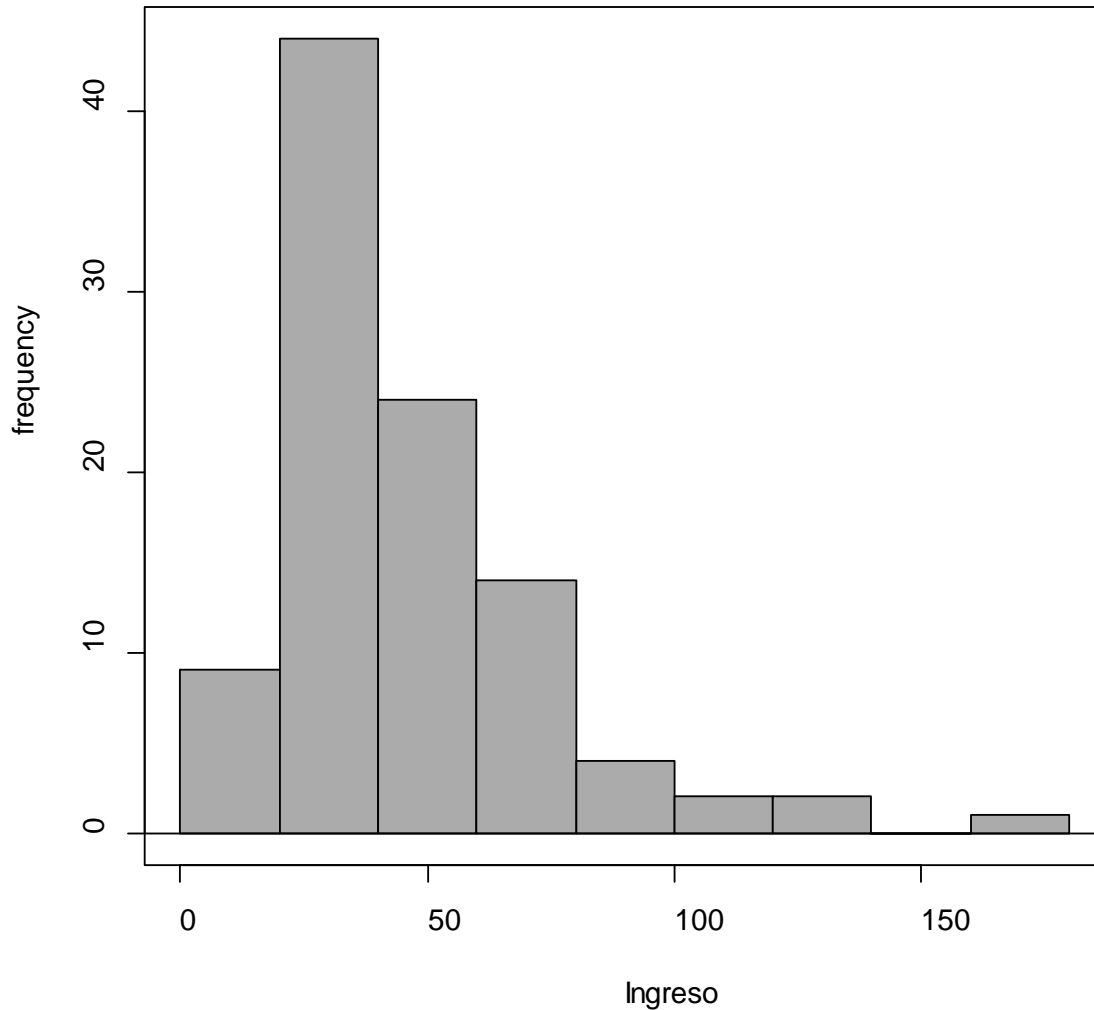
El gráfico aparece en una ventana de gráficos, que quizás estará oculta detrás de la propia ventana de R Commander. Se puede copiar fácilmente mediante Ctrl+W sobre la ventana del gráfico, y Ctrl+V (o edición → pegar; botón derecho del ratón → pegar) en cualquier aplicación de Windows (como Word o Excel) en la que se deseen insertar los gráficos contruidos. Deben ser copiados antes de elaborar otro gráfico, ya que generalmente la ventana de gráficos nueva sustituye a la antigua, que desaparece.

Para una variable cuantitativa es más adecuado el histograma:

Gráficas → Histograma ...

En el cuadro de diálogo elegimos la variable “Ingreso”. Con la pestaña de “opciones” podríamos cambiar los títulos.

```
with(Dataset, Hist(Ingreso, scale="frequency", breaks="Sturges", col="darkgray"))
```

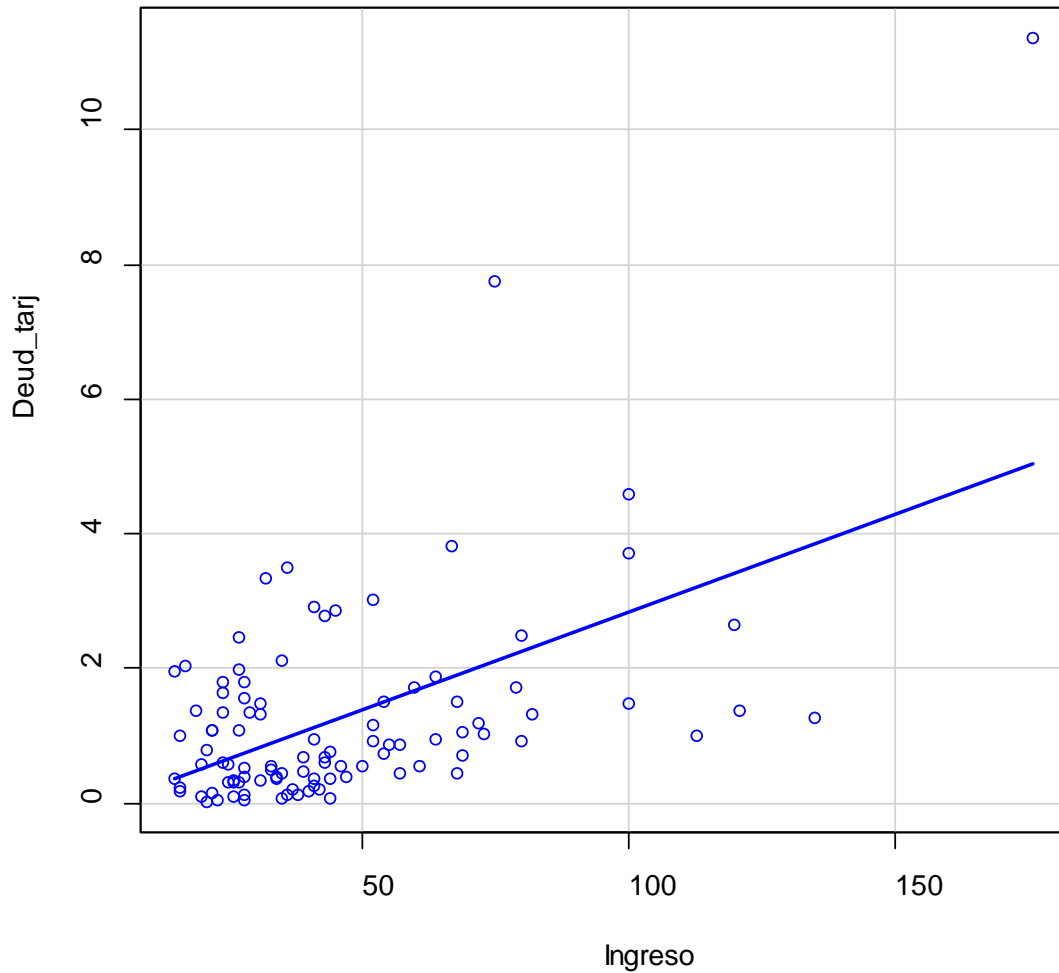


Veremos ahora gráficamente la relación, observada anteriormente mediante el coeficiente de correlación, entre las deudas por tarjeta y los ingresos; para ello el gráfico más adecuado es el diagrama de dispersión, y es habitual que la variable X (eje horizontal) sea la explicativa e Y (eje vertical) la variable dependiente:

Gráficas → Diagrama de dispersión

Elegimos como variable X “Ingreso”, y como variable Y “Deud_tarj”. Marcamos en opciones únicamente “línea de mínimos cuadrados”.

```
scatterplot(Deud_tarj~Ingreso, regLine=TRUE, smooth=FALSE, boxplots=FALSE, data=Dataset)
```



La línea recta con pendiente positiva expresa la relación entre ambas variables.

Por último, construiremos un modelo de regresión lineal para intentar explicar mejor la relación entre el cociente deudas/ingresos y las restantes variables cuantitativas.

Estadísticos → Ajuste de modelos → Regresión lineal

En el cuadro de diálogo que aparece elegimos la variable explicada “Deud_ing”, y las variables explicativas (las tres restantes):



```
> RegModel.1 <- lm(Deud_ing~Deud_otr+Deud_tarj+Ingreso, data=Dataset)
```

```
> summary(RegModel.1)
```

```
Call: lm(formula = Deud_ing ~ Deud_otr + Deud_tarj + Ingreso, data = Dataset)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.7619	-2.0973	-0.7929	1.4115	12.3010

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.2730	0.7464	12.423	< 2e-16 ***
Deud_otr	1.8619	0.1857	10.025	< 2e-16 ***
Deud_tarj	2.0405	0.3225	6.328	0.00000000789 ***
Ingreso	-0.1619	0.0155	-10.444	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.727 on 96 degrees of freedom

Multiple R-squared: 0.7368, Adjusted R-squared: 0.7286

F-statistic: 89.58 on 3 and 96 DF, p-value: < 2.2e-16

En los resultados observamos en primer lugar que todos los coeficientes son significativos, ya que los valores P en la última columna son prácticamente cero así como el valor P del modelo en su conjunto, y en segundo lugar comprobamos que el coeficiente de determinación (adjusted R-squared) o porcentaje de variabilidad del diámetro -variable dependiente- explicado por el ajuste es del 72,86%, moderadamente alto.

R tiene miles de funciones, por lo que no es posible aprender o memorizar el lenguaje completo. Es conveniente conocer las órdenes básicas, que se repiten de forma similar en cualquier aplicación, como las que se refieren al manejo de los datos, cálculo de estadísticos y tablas, o representación gráfica elemental mostradas en esta breve introducción. Para las aplicaciones más complejas es recomendable utilizar las funciones, órdenes, y procedimientos de algún ejemplo conocido (que es conveniente conservar) y cambiar en ellas el nombre del archivo o conjunto de datos, los nombres de las variables, y quizás algún elemento de configuración.

Para los usuarios nuevos de R suele ser frustrante encontrar que los resultados no aparecen, o se multiplican los avisos y mensajes de error. La utilización de órdenes que deben ser programadas en lugar de un menú completo de usuario hace necesario habitualmente corregir pequeños defectos de sintaxis y ejecutar de nuevo la instrucción. Debe prestarse atención al mensaje de error que aparece cuanto una orden no es ejecutada correctamente (cuando no hay mensaje debemos pensar que la ejecución es correcta), por lo que conviene borrar con frecuencia los mensajes antiguos, para que podamos reconocer el mensaje nuevo. Algunos de los errores más habituales son:

- a) Los nombres de las funciones y objetos (variables, conjuntos de datos, resultados) deben ser exactos. Un cambio inadvertido en una mayúscula, un espacio, o un punto, harán que R no reconozca ese nombre.
- b) Las variables pertenecen a un determinado conjunto de datos. La función `attach` sirve para establecer un conjunto por defecto, al que se refieren los nombres de variables. Si no los reconoce quizás hemos olvidado esa orden `attach` previa.
- c) Las funciones pertenecen a un determinado paquete de R. Si no reconoce una función quizás hemos olvidado cargar (`library`) el paquete correspondiente.
- d) Algunos procedimientos y funciones exigen que los datos no contengan valores perdidos (NA).
- e) Las órdenes de una aplicación suelen estar encadenadas: cada una utiliza los resultados producidos por las anteriores, y si se produce un error repercutirá en todo lo que sigue. Es necesario ejecutar las instrucciones una a una comprobando que no hay mensaje de error.
- f) Algunas órdenes no producen resultados visibles, sino que los almacenan en un objeto de resultado. Podemos ver los resultados en general ejecutando el nombre del objeto creado.
- g) Una orden puede tener más de una línea. Debemos marcar la orden completa para ejecutarla.
- h) Los argumentos de una función de R están separados por comas. Debemos comprobar que todos ellos son correctos. También debemos comprobar que no falta o sobra ningún paréntesis, comillas, punto y coma, u otro símbolo necesario.
- i) Los textos (etiquetas, títulos, etc.) deben ponerse entre comillas. El símbolo de comillas (" ") puede ser distinto del que se obtiene con el teclado del editor de textos (" "), en cuyo caso deberá ser reescrito desde R, ya que el segundo no es válido.

Finalizamos esta sesión guardando las instrucciones y resultados; debe indicarse un nombre de archivo (por ejemplo: `instrucciones1`, `resultados1`). Los resultados se guardan en un archivo de texto (extensión `".txt"`), y las instrucciones en un archivo de órdenes de R (extensión `".R"`), y pueden ser abiertos y utilizados posteriormente.

Fichero → Guardar las instrucciones como...

Fichero → Guardar los resultados como...

Por último cerramos R commander.

Fichero → Salir