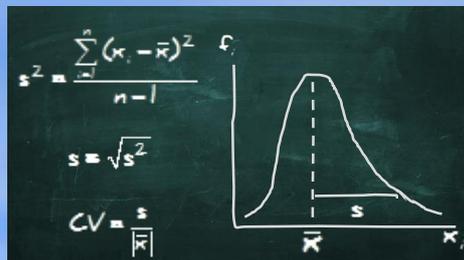


# METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA

---

Pedro López-Roldán  
Sandra Fachelli





# METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA

---

Pedro López-Roldán  
Sandra Fachelli

Bellaterra (Cerdanyola del Vallès) | Barcelona  
Dipòsit Digital de Documents  
Universitat Autònoma de Barcelona

**UAB**





Este libro digital se publica bajo licencia *Creative Commons*, cualquier persona es libre de copiar, distribuir o comunicar públicamente la obra, de acuerdo con las siguientes condiciones:

-  *Reconocimiento.* Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
-  *No Comercial.* No puede utilizar el material para una finalidad comercial.
-  *Sin obra derivada.* Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

No hay restricciones adicionales. No puede aplicar términos legales o medidas tecnológicas que legalmente restrinjan realizar aquello que la licencia permite.

Pedro López-Roldán

Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (<http://quit.uab.cat>)

Institut d'Estudis del Treball (<http://iet.uab.cat/>)

Departament de Sociologia. Universitat Autònoma de Barcelona

[pedro.lopez.rolan@uab.cat](mailto:pedro.lopez.rolan@uab.cat)

Sandra Fachelli

Departament de Sociologia i Anàlisi de les Organitzacions

Universitat de Barcelona

Grup de Recerca en Educació i Treball (<http://grupsderecerca.uab.cat/gret>)

Departament de Sociologia. Universitat Autònoma de Barcelona

[sandra.fachelli@ub.edu](mailto:sandra.fachelli@ub.edu)

Edició digital: <http://ddd.uab.cat/record/129382>

1ª edición, febrero de 2015

Edifici B · Campus de la UAB · 08193 Bellaterra  
(Cerdanyola del Vallés) · Barcelona · España  
Tel. +34 93 581 1676

# Índice general

## **PRESENTACIÓN**

### **PARTE I. METODOLOGÍA**

- I.1. FUNDAMENTOS METODOLÓGICOS
- I.2. EL PROCESO DE INVESTIGACIÓN
- I.3. PERSPECTIVAS METODOLÓGICAS Y DISEÑOS MIXTOS
- I.4. CLASIFICACIÓN DE LAS TÉCNICAS DE INVESTIGACIÓN

### **PARTE II. PRODUCCIÓN**

- II.1. LA MEDICIÓN DE LOS FENÓMENOS SOCIALES
- II.2. FUENTES DE DATOS
- II.3. EL MÉTODO DE LA ENCUESTA SOCIAL
- II.4. EL DISEÑO DE LA MUESTRA
- II.5. LA INVESTIGACIÓN EXPERIMENTAL

### **PARTE III. ANÁLISIS**

- III.1. SOFTWARE PARA EL ANÁLISIS DE DATOS: SPSS, R Y SPAD
- III.2. PREPARACIÓN DE LOS DATOS PARA EL ANÁLISIS
- III.3. ANÁLISIS DESCRIPTIVO DE DATOS CON UNA VARIABLE
- III.4. FUNDAMENTOS DE ESTADÍSTICA INFERENCIAL
- III.5. CLASIFICACIÓN DE LAS TÉCNICAS DE ANÁLISIS DE DATOS
- III.6. ANÁLISIS DE TABLAS DE CONTINGENCIA
- III.7. ANÁLISIS LOG-LINEAL
- III.8. ANÁLISIS DE VARIANZA
- III.9. ANÁLISIS DE REGRESIÓN
- III.10. ANÁLISIS DE REGRESIÓN LOGÍSTICA
- III.11. ANÁLISIS FACTORIAL
- III.12. ANÁLISIS DE CLASIFICACIÓN



# Metodología de la Investigación Social Cuantitativa

---

Pedro López-Roldán  
Sandra Fachelli

## PARTE III. ANÁLISIS

### Capítulo III.1 Software para el análisis de datos: SPSS, R y SPAD

Bellaterra (Cerdanyola del Vallès) | Barcelona  
Dipòsit Digital de Documents  
Universitat Autònoma de Barcelona

**UAB**



Cómo citar este capítulo:

López-Roldán, P.; Fachelli, S. (2015). Software para el análisis de datos: SPSS, R y SPAD. En P. López-Roldán y S. Fachelli, *Metodología de la Investigación Social Cuantitativa*. Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona. Capítulo III.1. 1ª edición. Edición digital: <http://ddd.uab.cat/record/129380>

Capítulo redactado en febrero de 2015

# Índice de contenidos

<b>1. El paquete estadístico SPSS</b>	<b>7</b>
1.1. Ayuda y referencias de SPSS	9
1.2. El entorno de trabajo de SPSS	9
1.3. Configuración de las opciones de SPSS	16
1.4. Análisis de datos con SPSS: el editor de resultados	17
1.5. Copiar y exportar los resultados de SPSS	20
1.6. La programación con SPSS: el editor de sintaxis	21
1.6.1. Principales características	21
1.6.2. El lenguaje de comandos de SPSS	23
1.7. Errores de ejecución	27
<b>2. El paquete estadístico R</b>	<b>28</b>
2.1. Características generales de R	31
2.1.1. Instalación de R	31
2.1.2. Iniciando R	32
2.1.3. La ayuda y documentación de R	33
2.1.4. Instalación de paquetes	38
2.1.5. Funcionamiento de R: objetos y workspace	41
2.2. Deducer	43
2.2.1. Instalación de Deducer	43
2.2.2. La consola de Deducer	46
2.2.3. El visor de datos	50
2.2.4. Análisis de datos con la consola de Deducer	57
2.3. RStudio	62
2.3.1. Ejecución de programas de instrucciones con R-Studio	63
2.3.2. Comandos y programación en R	64
<b>3. El paquete estadístico SPAD</b>	<b>66</b>
<b>4. Bibliografía</b>	<b>78</b>



## Software para el análisis de datos: SPSS, R y SPAD

**E**l análisis de datos cuantitativos requiere la utilización de software específico para la realización de diferentes tareas que genéricamente englobamos en la fase del análisis de los datos. Entre ellas se encuentra el registro de los datos, su depuración, el tratamiento de ficheros, la transformación de las variables y el análisis propiamente dicho a través de las distintas técnicas implementadas, ya sean de procesamiento de un cálculo numérico o estadístico o de representación gráfica.

Existen numerosos programas que a lo largo del tiempo, desde los años 60, han ido apareciendo. Unos se han consolidado como referentes a nivel mundial con millones de usuarios/as, otros también se mantienen en ámbitos más reducidos de público. Muchos responden a la lógica del mercado comercial y otros son de libre acceso.

En general se utilizan en un entorno de trabajo que emplea una interfaz de ventanas y menús, lo que facilita sobre todo el aprendizaje y su uso en un entorno “amigable”, especialmente en la etapa de iniciación. Pero el uso más avanzado, profesional, exhaustivo y eficiente del software implica habitualmente que se utilice el lenguaje de programación y la ejecución de comandos, ya sea uno a uno o a través de diversos de ellos que se reúnen en programas de instrucciones (sintaxis). También es de interés la utilización combinada, de menús y lenguaje de comandos, para optimizar el esfuerzo de procesamientos de los datos.

En la tabla adjunta se presentan los programas informáticos más utilizados en el ámbito de las ciencias sociales para el análisis de datos estadísticos y que se suelen denominar como paquetes estadísticos.

En este texto manejaremos en distintos momentos los tres primeros: SPSS, R y SPAD, presentados en un entorno de *MS-Windows*. El primero se denomina actualmente como **IBM SPSS Statistics**, y sus siglas se corresponden con *Statistical Product and Service Solutions*. Creado en 1968, SPSS, junto a SAS, es el software estadístico comercial más completo y con mayor presencia a nivel mundial. Una licencia de uso es cara y su precio se expresa en miles de euros. Por su parte **R** significa la alternativa de software libre que se ha ido consolidando desde su creación en 1993 para convertirse en la actualidad en uno de los más utilizados en todos los ámbitos y disciplinas. Por su parte **SPAD**

(*Système Portable pour l'Analyse des Données*) es un software comercial de origen francés con una mucha menor presencia en el contexto internacional pero que ofrece algunos procedimientos de análisis de interés como alternativa o complemento de los ofrecidos por los dos anteriores.

Tabla III.1.1 Software para el análisis de datos: paquetes estadísticos

Nombre	Desarrollo	Web
	IBM	C <a href="http://www.spss.com">www.spss.com</a>
	R Foundation	L <a href="http://www.r-project.org">www.r-project.org</a>
	Coheris	C <a href="http://www.coheris.fr/fr/page/produits/Spad.html">www.coheris.fr/fr/page/produits/Spad.html</a>
	Projecto GNU	L <a href="http://www.gnu.org/software/pspp">www.gnu.org/software/pspp</a>
	SAS Institute	C <a href="http://www.sas.com">www.sas.com</a>
	StataCorp	C <a href="http://www.stata.com">www.stata.com</a>
	Statsoft	C <a href="http://www.statsoft.com">www.statsoft.com</a>
	Systat Software	C <a href="http://www.systat.com">www.systat.com</a>
	Minitab Inc.	C <a href="http://www.minitab.com">www.minitab.com</a>
	Statpoint Technologies	C <a href="http://www.statgraphics.net">www.statgraphics.net</a>
	Microsoft	C <a href="http://www.microsoft.com">www.microsoft.com</a>

Nota: C: distribución comercial | L: software libre

Fuente: Elaboración propia

La información sobre el resto del software se puede consultar a través de las direcciones web que se adjuntan o realizando una búsqueda en internet<sup>1</sup>.

En los siguientes apartados, de forma introductoria y de presentación, destacamos las principales características de los tres paquetes estadísticos seleccionados. Nos referiremos al entorno de trabajo de cada uno y a las diferentes ventanas con las que se trabaja: veremos los editores de los datos, el sistema de menú con el que realizar las distintas tareas del paquete estadístico, la visualización de los resultados de tablas y gráficos, así como el editor de sintaxis que posibilita la ejecución de las distintas tareas con el lenguaje de comandos.

Después de este primer recorrido, en el capítulo siguiente profundizaremos en distintas tareas destinadas a la preparación de los datos para el análisis: la identificación de los datos, el tratamiento de ficheros, la depuración de los datos y su transformación. Todas ellas previas o posteriores, pero necesarias y complementarias, al análisis propiamente dicho de los datos. Después se introduce, capítulo a capítulo, desde el análisis de datos más sencillo que es el univariable, pasando por el análisis bivivariable de datos estadísticos y llegando a tratar diversas técnicas más complejas de análisis multivariable, ya sea en modelos de dependencia o interdependencia. Dedicaremos un capítulo específico a presentar de forma conjunta y general una clasificación y descripción de las distintas técnicas de análisis de datos.

<sup>1</sup> En [http://en.wikipedia.org/wiki/Comparison\\_of\\_statistical\\_packages](http://en.wikipedia.org/wiki/Comparison_of_statistical_packages) se puede encontrar una comparativa de éstos y otros softwares existentes. Se puede consultar también la página [http://en.wikipedia.org/wiki/List\\_of\\_statistical\\_packages](http://en.wikipedia.org/wiki/List_of_statistical_packages). Salas (2008) ofrece una comparativa entre SPSS, SAS y R.

## 1. El paquete estadístico SPSS

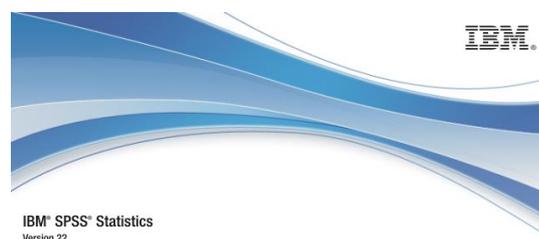


se creó en 1968 con este acrónimo para identificar lo que originalmente se denominó como *Statistical Package for the Social Sciences* y se ha convertido en el programa estadístico de referencia obligada a nivel mundial. La diversificación de productos de la compañía y una marcada orientación hacia la comercialización en múltiples ámbitos explica el cambio de denominación hacia la nueva identificación de *Statistical Product and Service Solutions*.

La primera versión del programa fue desarrollada por Norman H. Nie, Dale H. Bent y C. Hadlai Hull<sup>2</sup> en 1968 en la *Stanford University*. Entre 1969 y 1975 la *University of Chicago* por medio de su *National Opinion Research Center* estuvo a cargo del desarrollo, distribución y venta. A partir de 1975 se crea la compañía *SPSS Inc.* Las primeras versiones funcionaban en grandes ordenadores y no es hasta el año 1984 que aparece la primera versión para ordenadores personales, llamado SPSS/PC+ para la plataforma de MS-DOS. Con la versión 10, en 1992, empezó a correr en el entorno gráfico de *Windows*. Desde entonces la empresa fue creciendo y consolidándose, e introdujo distintas aplicaciones analíticas que ampliaron considerablemente la oferta de productos que se expresan bajo el concepto de análisis predictivo. La etapa actual está marcada por la adquisición de SPSS Inc. por IBM en 2009.

Una de las constantes de esta evolución es la progresiva introducción de mejoras destinadas a facilitar la realización de las diferentes tareas de análisis así como la continua incorporación de nuevos procedimientos. También ha significado la diversificación de los usuarios, inicialmente circunscrito al ámbito de la investigación más académica para generalizarse a todo tipo de usuario e institución.

Muchas de las características básicas de SPSS no se han modificado en el tiempo. La mayoría de los procedimientos que se ejecutaban hace 20 o 30 años siguen funcionando, muchos con mejoras introducidas y otros, si bien continúan existiendo, han entrado en desuso. No obstante, en una perspectiva de 20 o 30 años los cambios son relevantes, en el entorno de trabajo y la resolución gráfica, en la cantidad y calidad de los procedimientos disponibles o en la capacidad de procesamiento<sup>3</sup>.



En estas páginas daremos cuenta de la versión de **IBM SPSS Statistics 22**, introducida en el mercado en el agosto del 2013.

<sup>2</sup> Hay quien ha señalado (Wellman, 1998) que el manual de estos tres autores (Nie, Bent y Hull, 1970) ha sido uno de los libros más influyentes entre los sociólogos en aquel entonces, tras su amplia difusión en el ámbito universitario estadounidense.

<sup>3</sup> El relato de la historia de SPSS se puede consultar en: <http://www.spss.com.hk/corpinfo/history.htm>.

SPSS es un paquete estadístico que integra un conjunto de programas y subprogramas organizados de forma que cada uno de ellos está interconectado con todos los demás. Así, la utilidad del conjunto integrado es mayor que la suma de las partes. De esta forma el paquete estadístico permite que se puedan aplicar a un mismo fichero de datos un conjunto muy amplio de procedimientos estadísticos de manera sincronizada, sin salir nunca del programa.

En SPSS existe un módulo base y diversos módulos adicionales que amplían las posibilidades de análisis y configuran distintas ofertas de procedimientos adecuados para distintas necesidades y ámbitos profesionales. Si entramos en la página de IBM SPSS podemos ver los distintos módulos y productos (<http://www-01.ibm.com/software/analytics/spss/>). Junto al módulo base se dispone de los módulos siguientes:



*Statistics, Advanced Statistics, Regression, Categories, Exact Test, Custom Tables, Decision Trees, Complex Samples, Neural Networks, Forecasting, Conjoint, Bootstrapping, Missing Values, Data Preparation.*

En el esquema adjunto que aparece en el Gráfico III.1.1 se puede observar la lógica general de tareas que satisface un paquete estadístico como SPSS dentro de un mismo proceso analítico, desde la planificación y producción de los datos que son introducidos en el soporte informático, pasando por la gestión y análisis de los mismos, hasta su presentación y distribución de los resultados.

Gráfico III.1.1 Esquema del proceso analítico de datos con SPSS



Fuente: IBM SPSS

## 1.1. Ayuda y referencias de SPSS

En este capítulo damos cuenta de una breve introducción al software. Para ampliar la información sobre el mismo se pueden consultar diversos recursos a través de los enlaces que disponibles en: <http://pagines.uab.cat/plopez/content/spss#links>, donde en particular se dispone del acceso a los distintos manuales en pdf de cada versión. Para una primera lectura se puede consultar la guía breve (IBM Corporation, 2015c).

El propio software estadístico dispone de una ayuda muy completa que se puede obtener a través del menú de **Ayuda** y que enlaza con diversas direcciones de SPSS en internet:

- La ayuda por temas e índice.
- El tutorial de las funciones básicas del SPSS.
- Estudio de casos con ejemplos prácticos de análisis estadístico y su interpretación. Es accionable desde el menú contextual de una tabla pivote en la ventana de resultados.
- El asesor estadístico que orienta la realización de algunas tareas fundamentales con SPSS.
- *Command Syntax Reference Guide* es el manual del SPSS de la sintaxis de comandos, en inglés y en formato PDF, fundamental para un conocimiento exhaustivo y más avanzado del software con más de 280 comandos que son explicados a lo largo de más de 2.000 páginas.
- La explicación de los algoritmos aplicados en los distintos procedimientos de análisis.
- El acceso a la comunidad de SPSS:  
<http://www.ibm.com/developerworks/spssdevcentral>
- La ayuda para trabajar con R.
- La página principal de IBM SPSS.
- Manuales de programación con otras aplicaciones.

Gráfico III.1.2 Menú ayuda de SPSS



Además se dispone en distintos entornos de trabajo:

- Del botón de ayuda en los cuadros de diálogo.
- De la ayuda en el menú contextual.
- De los esquemas de sintaxis de los comandos de SPSS.
- Del asesor de resultados.

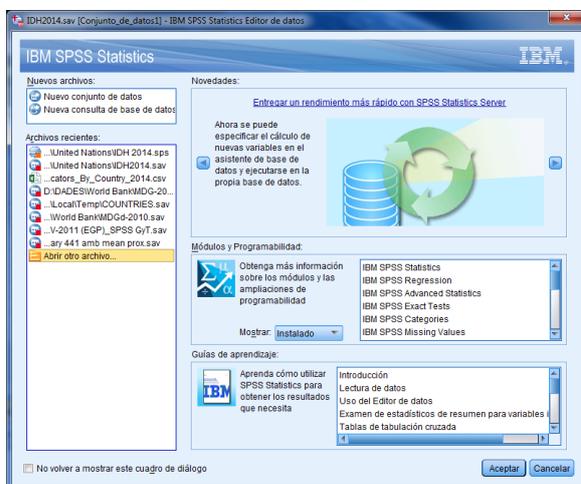
## 1.2. El entorno de trabajo de SPSS

El entorno de trabajo del SPSS es una interface gráfica de ventanas, menús y cuadros de diálogo propios del sistema operativo de *Microsoft Windows*.

Cuando iniciamos una sesión con el paquete estadístico nos aparece una ventana de bienvenida que nos permite ver los productos instalados y no instalados, acceder a

guías de aprendizaje, ver las novedades de la versión e iniciar nuestra sesión de trabajo a partir de archivos trabajados con anterioridad o con otros nuevos.

### Gráfico III.1.3 Ventana de bienvenida de SPSS



En SPSS existen **tres ventanas principales** que se corresponden con los tres archivos principales del software estadístico:

- El editor de datos: **Nombre\_archivo.SAV**.
- El editor o visor de resultados: **Nombre\_archivo.SPV** que incluye a su vez tres editores, de:  
 Tablas , Gráficos  y Texto .
- El editor de sintaxis: **Nombre\_archivo.SPS**



En el Gráfico III.1.4 se pueden apreciar las imágenes de tres capturas de pantalla de las ventanas principales de SPSS. En el Gráfico III.1.5 se amplían las barras de menús e iconos de cada una de ellas. Los iconos que aparecen se irán comentando a lo largo del texto y están destinados a ejecutar algunas de las tareas más habituales o de mayor interés de los menús.

Se puede abrir más de una ventana de cada tipo de editor. La ventana que se considera activa (designada) incorpora el signo **+** en el icono que la representa.

### Gráfico III.1.4 Ventanas principales del SPSS: datos, resultados y sintaxis

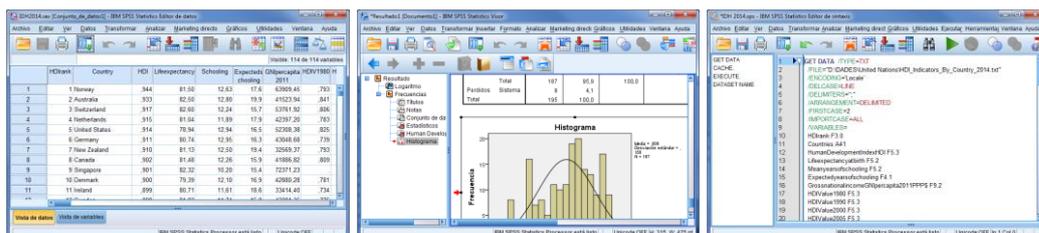


Gráfico III.1.5 Menús y barras de iconos de las ventanas de datos, resultados y sintaxis del SPSS

#### Editor de datos



#### Visor de resultados



#### Editor de sintaxis



A continuación se describen las características generales de cada menú. En particular cabe destacar como principales los correspondientes a: **Datos**, **Transformar**, **Analizar** y **Gráficos** que aparecen desplegados como imágenes en el Gráfico III.1.6.

**ARCHIVO:** Menú destinado a crear, abrir y guardar los archivos, pedir información de los mismos o imprimirlos.

**EDITAR:** Menú de edición habitual del entorno *Windows* que incluye también la configuración de las opciones generales del SPSS.

**VER:** Menú que permite activar diferentes opciones de visualización de la ventana.

**DATOS:** Menú que recoge un conjunto de comandos de manipulación del archivo de datos (definición de los datos, selección y ponderación de casos, segmentación del archivo, agregación de datos, unión de ficheros, reestructuración, etc).

**TRANSFORMAR:** Menú de los procedimientos de transformación de los datos de un archivo y que implican en general la generación de nuevas variables como resultado de su recodificación o la realización de cálculos.

**ANALIZAR:** Menú de todos los procedimientos de análisis estadístico con SPSS.

**GRÁFICOS:** Menú para obtener representaciones gráficas de los datos.

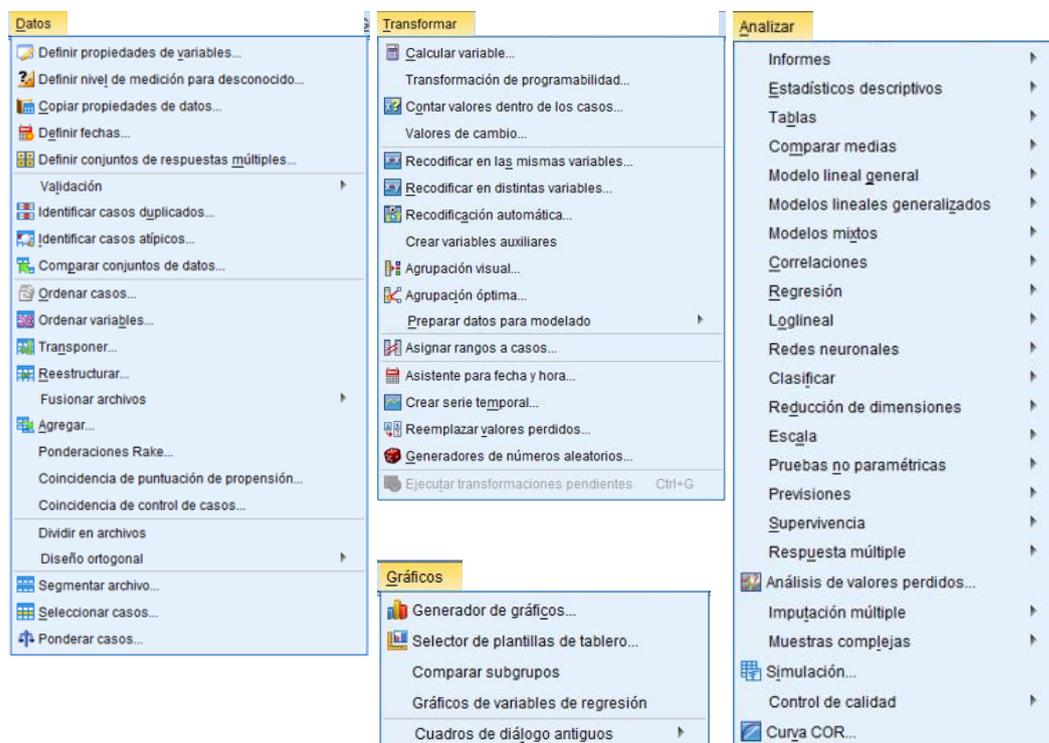
**FORMATO:** Menú de utilidades de formato en la ventana de resultados.

**UTILIDADES:** Menú de utilidades para pedir información de las variables, definir conjuntos de variables, ejecutar procesos, configurar los menús, etc.

**VENTANA:** Menú de organización y de desplazamiento entre las diferentes ventanas y de división del visor de datos.

**AYUDA:** Menús de ayuda del SPSS comentado.

Gráfico III.1.6 Menús Datos, Transformar, Analizar y Gráficos de SPSS



Los diferentes procedimientos o comandos del SPSS están destinados a la realización de 4 tipos principales de tareas:

1. La **generación y transformación de archivos de datos**: identificación de los datos, importación y exportación de archivos, su combinación, así como comandos de información de los mismos.
2. La **transformación de los datos**: para la modificación y creación de nuevas variables que preparan los datos para el análisis.
3. El **análisis estadístico de los datos**: procedimientos de análisis y de representación gráfica.
4. Un conjunto de **comandos adicionales** de operación para la configuración, el control y la obtención de información del entorno de trabajo del SPSS, o para ser utilizados en la programación con el lenguaje de comandos.

Para trabajar con el SPSS tenemos que empezar necesariamente por crear o abrir un archivo de datos:

- Introduciéndolos en la ventana de datos.
- Importándolos de hojas de cálculos, bases de datos, archivos **txt**,...
- Abriendo un archivo previamente creado en formato SPSS.
- O incluso a partir de instrucciones en un programa de sintaxis.

Para **abrir una matriz de datos** en SPSS creada previamente:

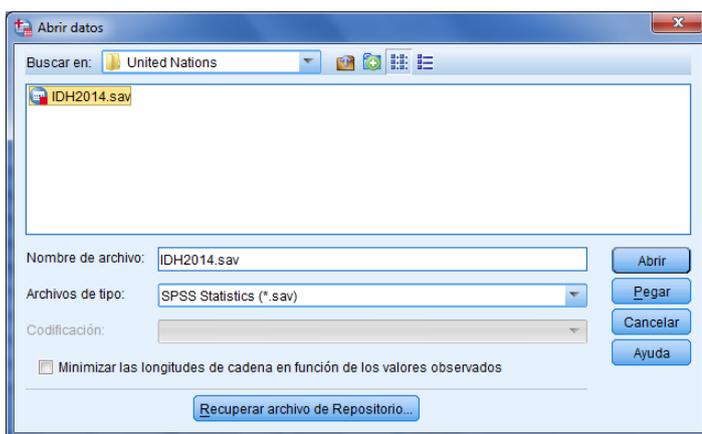
- A través del menú: **Archivo / Abrir / Datos**
- Clicando sobre el botón  desde la ventana del editor de datos.

### ► Ejercicio 1. Abrir una matriz de datos en SPSS

Desde el editor de datos abriremos dos de las matrices de datos que utilizamos en este manual: **IDH2014.sav** y **CIS3041.sav**.

Si abrimos la primera de las matrices accedemos a un cuadro como el siguiente donde, en particular, hemos seleccionado la matriz **IDH2014.sav**:

Gráfico III.1.7 Cuadro de diálogo “Abrir datos” del SPSS



Este archivo en la ventana del editor de datos se considera el **conjunto de datos activo** y se visualiza en el gráfico Gráfico III.1.8.

Gráfico III.1.8 Ventana del editor de datos del SPSS. Archivo IDH2014.sav.

	HDrank	Country	HDI	Lifeexpectancy	Schooling	Expected schooling	GNIpercapita 2011	HDIV1980	HDIV1990	HDIV2000	HDIV2005	HD
1	1	Norway	,944	81,50	12,63	17,6	63909,45	,793	,841	,910	,935	
2	2	Australia	,933	82,50	12,80	19,9	41523,94	,841	,866	,898	,912	
3	3	Switzerland	,917	82,60	12,24	15,7	53761,92	,806	,829	,886	,901	
4	4	Netherlands	,915	81,04	11,89	17,9	42397,20	,783	,826	,874	,888	
5	5	United States	,914	78,94	12,94	16,5	52308,38	,825	,858	,883	,897	
6	6	Germany	,911	80,74	12,95	16,3	43048,68	,739	,782	,854	,887	
7	7	New Zealand	,910	81,13	12,50	19,4	32569,37	,793	,821	,873	,894	
8	8	Canada	,902	81,48	12,26	15,9	41886,82	,809	,848	,867	,892	
9	9	Singapore	,901	82,32	10,20	15,4	72371,23		,744	,800	840	
10	10	Denmark	,900	79,39	12,10	16,9	42880,28	,781	,806	,859	,891	
11	11	Ireland	,899	80,71	11,61	18,6	33414,40	,734	,775	,862	,890	
12	12	Sweden	,898	81,82	11,74	15,8	43201,35	,776	,807	,889	,887	
13	13	Iceland	,895	82,09	10,41	18,7	35116,46	,754	,800	,858	,888	
14	14	United Kingdom	,892	80,55	12,31	16,2	35001,63	,735	,768	,863	,888	

Recordemos que una matriz de datos es un conjunto de filas y columnas a partir de la cual se organiza la información estadística. El formato más habitual de matriz de datos es aquél donde cada línea, llamada registro (*record*), constituye un caso (individuo o unidad) sobre el que se expresan los valores codificados correspondientes a las variables consideradas a las columnas. En este caso cada fila representa un país y cada columna un atributo o propiedad que caracteriza a cada uno de ellos. Los datos provienen de Naciones Unidas y corresponden a los indicadores que se utilizan para construir el **Índice de Desarrollo Humano** (IDH) correspondiente al año 2014<sup>4</sup>. Podemos comprobar que disponemos de información para 195 países en relación a 114 variables. La primera variable corresponde a la posición de cada país en el ranking que establece el índice y la segunda es el identificador del país. Las 112 variables restantes son los indicadores que considera el *United Nations Development Programme* para construir el indicador sintético IDH, todas ellas son variables medidas en una escala numérica o cuantitativa (de escala en terminología del SPSS).

La segunda matriz de datos, **CIS3041.sav**, corresponde a los datos del **Barómetro del Centro de Investigaciones Sociológicas** de octubre de 2014 (estudio número 3041)<sup>5</sup>. Se trata de una muestra estadística de 2480 individuos que contestaron la encuesta del barómetro en relación a diversas cuestiones de la actualidad política, entre otras, y que dan lugar a un total de 210 variables. En esta matriz (ver Gráfico III.1.9) cada fila es un individuo anónimo que se identifica con un número correspondiente al código numérico asignado a su cuestionario. En este caso la mayor parte de las variables son cualitativas, medidas en una escala nominal u ordinal.

Gráfico III.1.9 Ventana del editor de datos del SPSS. Archivo CIS3041.sav.

	ESTU	CUES	CCAA	PROV	MUN	TAMUNI	AREA	DISTR
1	3041	1	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
2	3041	2	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
3	3041	3	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
4	3041	4	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
5	3041	5	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
6	3041	6	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
7	3041	7	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
8	3041	8	País Vasco	Araba-Ál...	Burgos / Vitoria-Gasteiz	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
9	3041	9	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
10	3041	10	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
11	3041	11	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
12	3041	12	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado

A lo largo de este manual abriremos diversas matrices de datos que se proporcionan y también crearemos pequeñas matrices para realizar algunas tareas concretas. En el próximo capítulo trabajaremos con más profundidad el tema de la creación de una

<sup>4</sup> Los datos han sido extraídos de la dirección: <http://hdr.undp.org/es/data>.

<sup>5</sup> La información del barómetro (datos, cuestionario, resultados,...) se puede consultar en la dirección: [http://www.cis.es/cis/openncm/ES/1\\_encuestas/estudios/ver.jsp?estudio=14119](http://www.cis.es/cis/openncm/ES/1_encuestas/estudios/ver.jsp?estudio=14119).

matriz a partir de los datos de una encuesta. No obstante veamos a continuación a través de un ejercicio de iniciación cómo introducir e identificar una matriz de datos.

### ► Ejercicio 2. Crear una matriz de datos en SPSS

En el editor de datos introduciremos **5 casos** correspondientes a personas imaginarias con la información de su **nombre**, **edad** y **sexo**.

Si tenemos abierta una matriz de datos previamente en el editor y queremos crear una nueva, procederemos en primer lugar a abrir una nueva ventana del **editor de datos en blanco**. A través del menú, haremos: **Archivo / Nuevo / Datos**.

Los datos se introducen en la pestaña **Vista de datos** y se identifican en términos de lo que llamaremos **diccionario de los datos** en la pestaña **Vista de variables**.

Desde la pestaña de datos introduciremos los 5 casos de esta forma:

	VAR00001	VAR00002	VAR00003
1	Betina	30,00	2,00
2	Mauri	16,00	1,00
3	Santi	9,00	1,00
4	Lidia	18,00	2,00
5	Nico	15,00	1,00

Cuando se introducen los datos el sistema asigna automáticamente un nombre a las variables (VAR00001, VAR00002 y VAR00003) y asigna dos decimales a los valores numéricos. Los datos pueden ser textuales (como el nombre) o numéricos (como la edad). En el caso del sexo hemos introducido un código numérico para diferenciar ambos sexos (el 1 para los varones y el 2 para las mujeres). Introducida así la información la pestaña de variables presenta esta imagen:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	VAR00001	Cadena	6	0		Ninguna	Ninguna	7	Izqui...	Nomi...	Entra...
2	VAR00002	Numérico	8	2		Ninguna	Ninguna	8	Dere...	Descono...	Entra...
3	VAR00003	Numérico	8	2		Ninguna	Ninguna	8	Dere...	Descono...	Entra...

Este diccionario que identifica las variables iniciales lo mejoraremos con los siguientes cambios: damos nombre a las variables, ajustamos la anchura y el número de decimales, ponemos una etiqueta identificativa para cada variable y también una etiqueta para identificar cada código de la variable sexo, finalmente precisaremos el nivel de medida:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	nombre	Cadena	6	0	Nombre de la persona	Ninguna	Ninguna	7	Izqui...	Nomi...	Entra...
2	edad	Numérico	2	0	Edad de la persona	Ninguna	Ninguna	8	Dere...	Escala	Entra...
3	sexo	Numérico	1	0	Sexo de la persona	{1, Varón}...	Ninguna	8	Dere...	Nomi...	Entra...

En los gráficos siguientes (Gráfico III.1.10 y Gráfico III.1.11) se indican algunos de los botones y espacios principales de la ventana de datos, tanto de la vista de datos como de la vista de variables. A lo largo del texto iremos profundizando en su uso.

Una vez introducidos los datos es necesario guardarlos y convertirlos en un fichero del sistema SPSS, por ejemplo con el nombre **Datos1.sav**. Para **guardar un archivo de datos**:

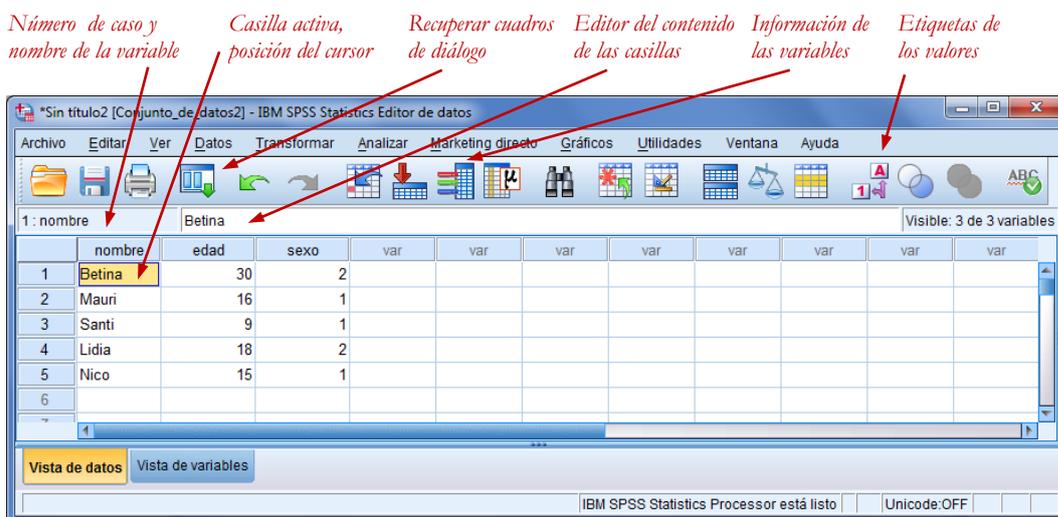
- A través del menú: **Archivo / Guardar**, o bien **Archivo / Guardar como**

- Con el teclado: <Ctrl>+<S>

- Clicando sobre el botón **Guardar este documento** .

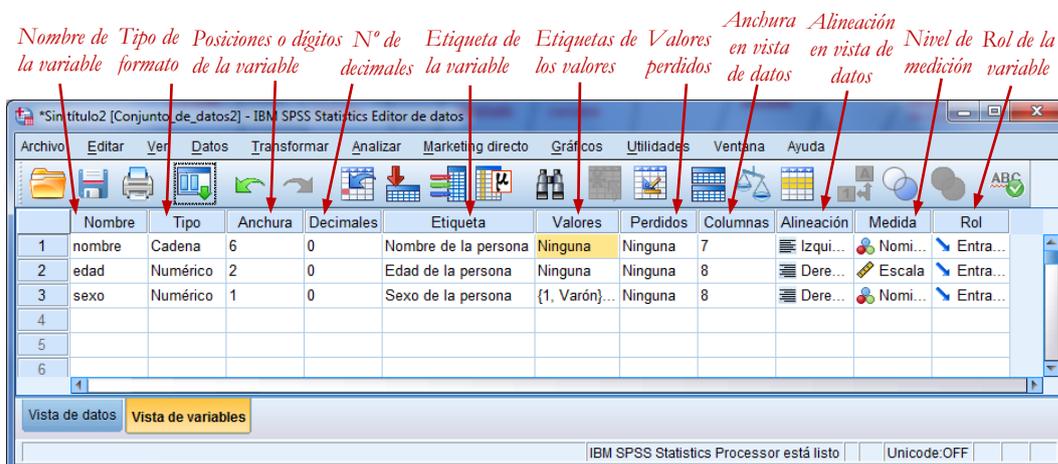
En el próximo capítulo veremos otra tarea importante y habitual del tratamiento de datos: la importación y exportación de matrices de datos, así como diversas tareas destinadas a la manipulación de una o varias matrices.

Gráfico III.1.10 Ventana del editor de datos del SPSS: vista de datos



*Barra de estado con información del procesador y del archivo de datos* 

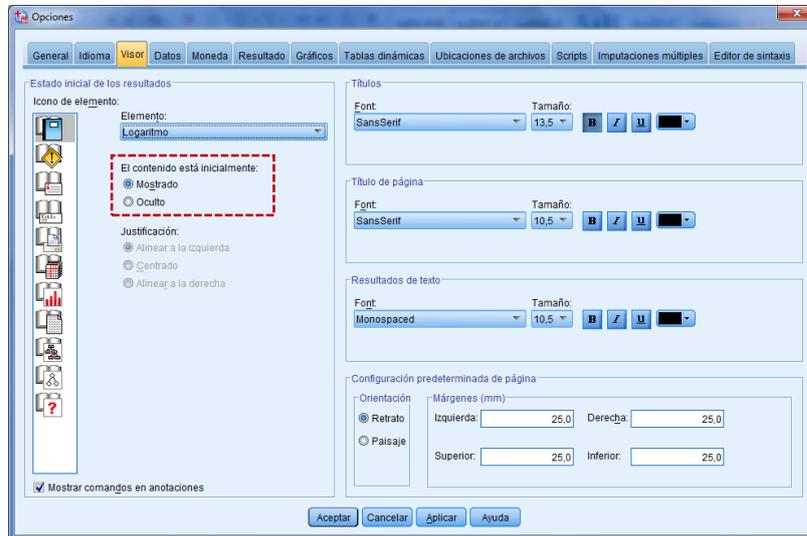
Gráfico III.1.11 Ventana del editor de datos del SPSS: vista de variables



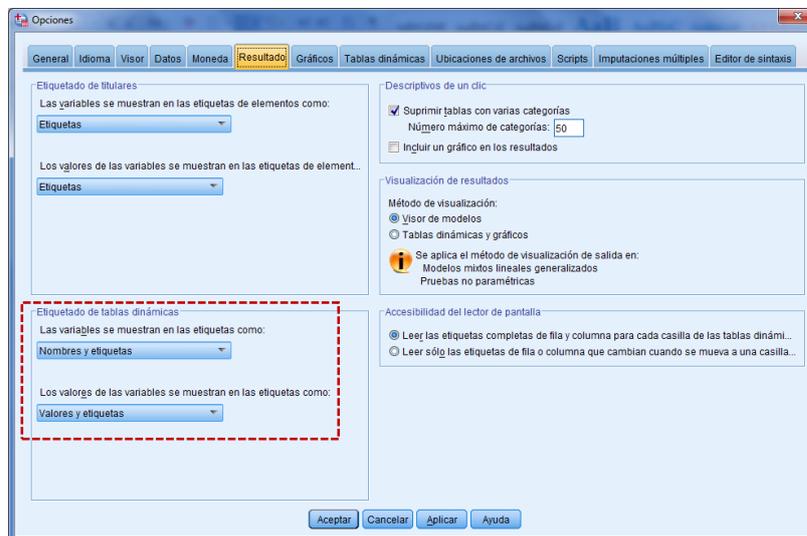
### 1.3. Configuración de las opciones de SPSS

En el menú **Editar** se encuentra la configuración de las **opciones** que controlan diversas alternativas de trabajo con el software. Con el objetivo de favorecer el aprendizaje del funcionamiento del paquete estadístico mantendremos activadas dos opciones de interés:

- a) En la pestaña de **Visor** mantendremos marcada la opción **Mostrado** en relación al elemento **Logaritmo** que hace referencia a las instrucciones que se ejecutan en el lenguaje de comandos del SPSS. Activado el mismo nos permitirá familiarizarnos con la sintaxis (el lenguaje de comandos) ya que cada vez que ejecutemos algún procedimiento podremos ver a qué sintaxis se corresponde.



- b) En la pestaña de Resultados, en el apartado de **Etiquetado de tablas dinámicas** (también llamadas **tablas pivote**) activaremos las opciones **Nombre y etiquetas** para las variables y **Valores y etiquetas** para los valores. De esta forma podremos identificar mejor las variables y sus valores en el momento de trabajar con los datos.



#### 1.4. Análisis de datos con SPSS: el editor de resultados

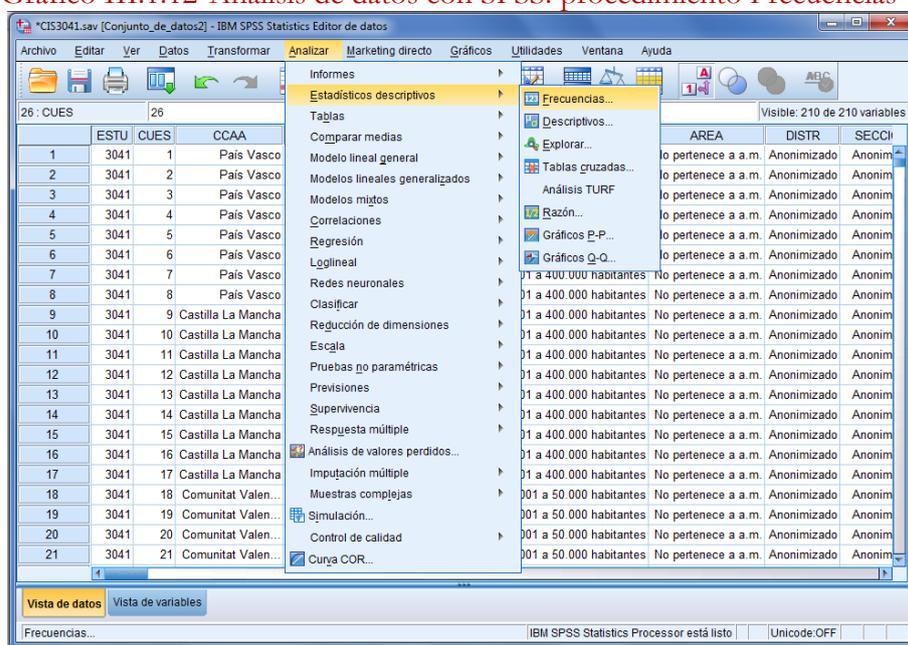
Con una matriz de datos que se abre o que acabamos de crear hay dos tipos principales de tareas que se pueden realizar con SPSS: manipularlas (menús **Datos** y **Transformar**) y analizarlas (menús **Analizar** y **Gráficos**). Todos los resultados obtenidos de estas tareas generan **tablas**, **gráficos** y **anotaciones de texto**, que aparecerán

automáticamente en la ventana de resultados o **Visor de resultados**<sup>6</sup>. Los tres tipos de materiales son editables y podemos adaptar su información para mejorar la lectura y la presentación formal. Realizaremos a continuación diversos ejercicios sencillos que nos permitirán obtener resultados a través de los menús de **Analizar** y **Gráficos**, los otros dos se verán en el próximo capítulo.

► **Ejercicio 3. Obtención de una tabla de frecuencias y un gráfico de sectores**  
Abrimos en primer lugar la matriz de datos **CIS3041.sav**. A través del menú **Analizar** se trata de obtener una tabla de frecuencias de la variable **sexo**: la variable **P31**, especificando además que extraiga un gráfico de sectores (llamados también circulares).

Para obtener las frecuencias desplegamos el menú **Analizar**, clicamos sobre **Estadísticos descriptivos** y sobre **Frecuencias** después:

Gráfico III.1.12 Análisis de datos con SPSS: procedimiento Frecuencias

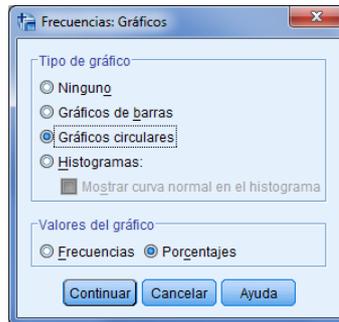


Aparece el cuadro de diálogo que sigue. A la izquierda disponemos del listado de variables de la matriz de datos, seleccionamos la variable **P31** y la colocamos en la derecha en el recuadro de **Variables** clicando sobre el botón .



<sup>6</sup> No obstante, es necesario tener presente que algunos procedimientos de los menús “Datos” i “Transformar” no suelen generar resultados visibles en el visor (más allá de presentar la sintaxis), pero sí que tienen un efecto sobre los datos que se pueden observar en la ventana de datos.

A continuación clicamos sobre el botón de **Gráficos** y especificamos en el nuevo cuadro de diálogo **Gráficos circulares** y **Porcentajes**.



Le damos a **Continuar**, volvemos al cuadro de diálogo principal anterior, y finalmente ejecutamos el procedimiento de obtención de frecuencias clicando sobre **Aceptar**. Automáticamente aparecerá en la ventana del visor de resultados el contenido del análisis que acabamos de ejecutar, de la forma que aparece en el Gráfico III.1.13 .

Gráfico III.1.13 Ventana del visor de resultados del SPSS

Botones para contraer o expandir - / +      Bloque de resultados (cabecera)

*Objetos de resultados (títulos)*

*Marca del objeto seleccionado. En este caso se suprimió la tabla de "Notas" por cuestión de espacio.*

*Anotación de texto (sintaxis)*

*Título del comando de SPSS*

*Tabla de frecuencias: tabla dinámica (o pivote)*

*Menú contextual*

*Gráfico circular*

*Esquema del visor*      *Contenido del visor*

Los resultados de texto , las tablas  y los gráficos  que aparecen en la ventana de resultados se pueden editar (o abrir en una ventana propia) de forma independiente si hacemos doble-clic sobre cada objeto (o a través del menú contextual).

Si queremos **guardar un archivo de resultados**, desde la ventana del visor de resultados optamos por:

- El menú: **Archivo / Guardar**, o bien **Archivo / Guardar como**.
- El teclado: <Ctrl>+<S>.
- El botón **Guardar este documento** .

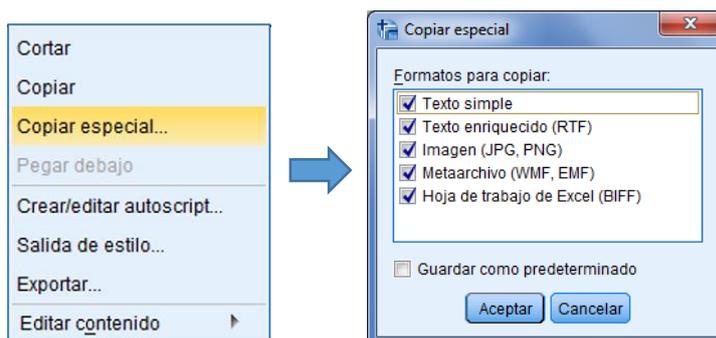
#### ► Ejercicio 4. Propuesto

Con la matriz de datos **CIS3041.sav** obtener las tablas de frecuencias de las variables: P1, P1101, P15, P18, P23, P28, P32, P33A, **P45** y OCUMAR11. ¿Qué nivel de medición tienen las diferentes variables?

### 1.5. Copiar y exportar los resultados de SPSS

Todos los objetos de resultados se pueden exportar o copiar a otras aplicaciones como procesadores de texto (*MS-Word*, *OpenOffice Writer*) o como hojas de cálculo (*MS\_Excel*, *OpenOffice Calc*). También se pueden exportar en formato **HTML** para incrustarlas en páginas web, o **PDF** para el intercambio de documentos.

Para copiar una tabla, un texto o un gráfico simplemente accionamos la opción copiar y pegamos en la aplicación de destino. Sobre el mismo elemento podemos abrir el menú contextual con el botón derecho del ratón y elegir **Copiar** o bien precisar la opción de copiado con **Copiar especial**. En este último caso podemos decidir y configurar si se copia como texto simple, enriquecido, imagen, metaarchivo o formato de Excel:

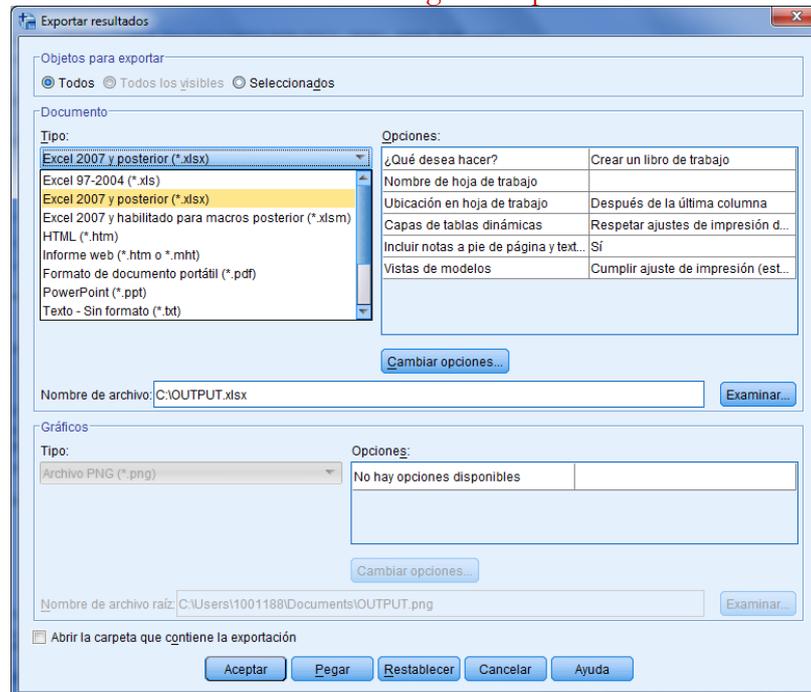


Al clicar sobre **Guardar como predeterminado** se mantendrá por defecto la configuración de tipos de copiado.

La opción de copiar como **Imagen (JPG, PNG)** es aconsejable cuando queremos mantener el aspecto con el que aparece en SPSS y facilita ajustar su tamaño en la aplicación de destino. Si la copiamos como **Texto enriquecido (RTF)** se puede adjuntar como una tabla del editor de textos. Si la queremos editar desde Excel optaremos por **Hoja de trabajo de Excel (BIFF)**.

También disponemos de la opción “Exportar” (a través del menú contextual, del menú Archivo / Exportar o con el botón ) que nos permite guardar los textos, las tablas y los gráficos conjuntamente en un archivo con diversidad de formatos: Word, Excel, HTML, Power-Point, TXT,...

Gráfico III.1.14 Cuadro de diálogo de exportación de resultados de SPSS



### ► Ejercicio 5. Propuesto

Copiar una tabla y un gráfico del visor de resultados mediante **Copiar especial** del menú contextual en todos los formatos posibles y pegarlos en un documento de *MS Word* como tabla de *Word* y como imagen. En segundo lugar, exportarlos con el procedimiento **Exportar**, eligiendo la opción **Todos**, en un documento de *Excel*, a continuación abrir el documento creado y comprobar la información obtenida.

## 1.6. La programación con SPSS: el editor de sintaxis

### 1.6.1. Principales características

Con el software SPSS existen dos formas posibles de trabajo:

- 1) **Interactivamente.** La ejecución de las instrucciones se realiza una a una, viendo los resultados de cada acción. Así podemos hacerlo a través de:
  - El **sistema de menús** donde la instrucción se elabora a través de los cuadros de diálogos y cuya ejecución se produce cuando clicamos **Aceptar**, como se hizo anteriormente. Esta es la manera de trabajar habitualmente en una etapa inicial de formación y cuando no se domina el lenguaje de comandos.
  - La ventana del **editor de sintaxis**, donde se escriben las instrucciones en el lenguaje de comandos del SPSS, a continuación se selecciona la instrucción y se

ejecuta a través del **botón de ejecución** . Es necesario por tanto conocer dicho lenguaje, pero sin conocerlo las instrucciones se pueden obtener automáticamente si clicamos sobre el botón **Pegar** desde los cuadros de diálogo de cada procedimiento de los menús.

- 2) **Procesando por lotes** o en modo *batch*. La ejecución de las instrucciones se realiza con ayuda del editor de sintaxis y donde se elaboran programas de instrucciones que se ejecutan conjuntamente. Una modalidad alternativa de esta forma de ejecución es la denominada unidad de producción (*Production mode*) donde las instrucciones se almacenan en un archivo de extensión **.SPP** y se ejecutan en bloque sin necesidad de abrir la aplicación del SPSS.

En un inicio, cuando se empieza a utilizar y aprender el uso del software, o en el caso de una utilización no experta, se puede sugerir que la opción más conveniente es emplear la modalidad de ejecución de instrucciones a través de los menús, ya que el aprendizaje inicial es mucho más fácil y rápido. A medida que se conoce y se domina el software se constata que de hecho la mejor opción es la utilización diversificada de las diferentes modalidades de trabajo. En ciertos momentos la utilización de los menús sigue siendo efectiva y la opción óptima, pero hay muchas situaciones y necesidades que no se pueden lograr de esta forma; la experiencia sugiere estrategias de trabajo mixtas que se adapten a las preferencias de cada usuario/a y los requerimientos de cada tarea a realizar, pero se constata a la larga que el uso avanzado del software implica trabajar con el lenguaje de comandos.

¿Por qué este interés por la programación? Un usuario avanzado o profesional del software precisará del uso de la sintaxis por varias razones, entre ellas cabe destacar:

- Algunos comandos sólo están disponibles a través del uso del lenguaje de comandos.
- Ciertas opciones de algunos comandos, los cuales se pueden ejecutar por menús, no se pueden ejecutar si no es a través de la sintaxis.
- En algunas ocasiones la programación es más fácil y rápida que la utilización de los menús, como es el caso de la ejecución de comandos relacionados con la transformación de los datos.
- El lenguaje de comandos permite guardar las instrucciones en un archivo de sintaxis que puede ser reutilizado o revisado con el objetivo de recordar, comprobar o repetir las tareas realizadas. Favorece por tanto la replicación y el control del trabajo.

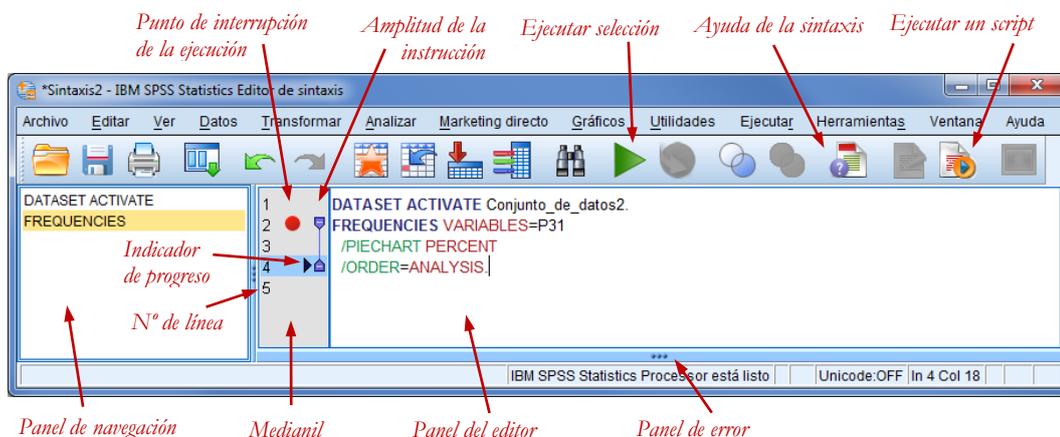
Veamos a continuación un ejercicio práctico de ejecución con el lenguaje de comandos.

### ► Ejercicio 6. Ejecución de una sintaxis de SPSS

Volver al menú de **Analizar** para ejecutar el procedimiento de frecuencias y pegar la sintaxis para ejecutarla después desde el editor de sintaxis.

Para volver al cuadro de diálogo de **Frecuencias** se puede ir por el camino de antes: **Analizar / Estadísticos descriptivos / Frecuencias**, o bien a través de una “atajo” si clicamos sobre el botón de **Recuperar los cuadros de diálogo recientes**  y eligiendo **Frecuencias**. Con la misma variable anterior o con otra de nuestra elección procedemos a ejecutar el procedimiento, pero esta vez en lugar de clicar sobre **Aceptar** lo haremos sobre **Pegar**. Se abrirá inmediatamente una ventana de sintaxis con las instrucciones del procedimiento de frecuencias en lenguaje de comandos.

Gráfico III.1.15 Ventana de sintaxis del SPSS



Se han incluido en el editor de sintaxis dos comandos: **DATASET ACTIVATE**, que hace que el conjunto de datos nombrado se convierta en el fichero activo para trabajar (el nombre asignado es automático: *Conjunto\_de\_datos2*), y **FRECUENCIAS**, que extrae la tabla de frecuencias y el gráfico.

Para **ejecutar instrucciones** se seleccionan y se clicla sobre el **botón de ejecución** . No es necesario seleccionar la instrucción completa, con enlazar el final de una con el inicio de la otra es suficiente. En el caso de la ejecución de un solo comando incluso no es necesaria la selección, basta con tener el cursor en cualquier posición entre el primer carácter y el último de la instrucción, el **punto**.

### 1.6.2. El lenguaje de comandos de SPSS<sup>7</sup>

El SPSS es un sistema que maneja, analiza y lista datos a través de instrucciones o mandatos que constituyen el **lenguaje de comandos del SPSS**. Constituyen de hecho palabras diseñadas especialmente para mantener un lenguaje sencillo y directo, y a las que se le asocian unas determinadas funciones. La programación con el SPSS se estructura a partir de la utilización de estas instrucciones mediante la elaboración de un fichero de comandos donde se disponen en sucesivas líneas cada una de ellas, desde unas cuantas hasta centenares.

El editor de sintaxis de SPSS es un sencillo editor de formato texto donde se escriben las instrucciones. Se rediseñó en la versión 17, con funciones como la de autocompletar, codificación de color, marcadores y puntos de separación, o el número de línea.

La estructura de los programas de SPSS es bastante flexible en relación al orden en que pueden colocarse las instrucciones del programa, se limita a las necesidades del sistema para respetar ciertas disposiciones lógicas. Así, por ejemplo, las variables tienen que ser definidas con anterioridad a su utilización en cualquier comando de transformación o procedimiento estadístico.

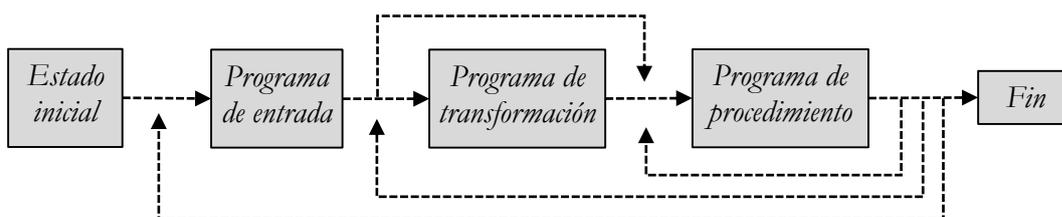
<sup>7</sup> Este apartado más avanzado puede saltarse para una lectura posterior si no se tiene un conocimiento suficiente del software.

En general, la mejor disposición de los comandos es aquella que se efectúa en el orden de las operaciones que el usuario/a desea realizar y considera lógica. Si las instrucciones están correctamente construidas, a medida que éstas aparecen, el sistema SPSS las irá ejecutando secuencialmente. No obstante, dependiendo del comando y de la configuración de las opciones del SPSS, cuando se ejecutan determinadas instrucciones no siempre el sistema las ejecuta de manera inmediata. El sistema SPSS trabaja de acuerdo con diferentes **estados del programa**, antes o después de encontrarse un comando, que estructuran el funcionamiento del proceso de los datos del sistema y que sirve como referencia básica de las tareas y de la disposición de las instrucciones en un fichero de comandos.

Estas tareas se pueden resumir en tres operaciones básicas: comandos que definen el **fichero activo**, comandos que transforman los datos y comandos que los analizan o listan. Específicamente, el sistema SPSS verifica el orden de estos comandos en función de los cuatro posibles estados que puede tener el programa: una sesión de trabajo con SPSS empieza en el llamado estado **inicial**, a continuación con el estado programa de **entrada** permite leer los datos, el estado de **transformación** posibilita la transformación de los datos, mientras que el estado de **procedimiento** permite iniciar la ejecución de un procedimiento (de análisis estadístico o de otro tipo).

Como se muestra en el esquema del Gráfico III.1.16, partiendo del estado inicial necesariamente tenemos que llamar a los datos que emplearemos; estos datos son leídos y constituyen el archivo activo. Con estos datos ejecutaremos comandos de transformación y de procedimiento, o bien sólo de procedimiento. Este proceso lógico puede repetirse tantas veces como sean necesarias antes de llegar a la finalización del programa de comandos, a la finalización de una sesión de trabajo. Entre todos estos estados del programa se podrán intercalar comandos llamados de configuración del entorno y de utilidades.

Gráfico III.1.16 Estados del programa en SPSS



Es importante destacar que los comandos que sitúen el programa en **estado de transformación** determinan que el sistema SPSS se limite a verificar la correcta sintaxis de las instrucciones y no es hasta que encuentra un comando de procedimiento que esta ejecución tiene lugar. Los comandos de procedimiento son los que provocan la lectura de los datos y es en ese momento en el que tienen efecto las anteriores instrucciones que pudiera haber de definición o de manipulación y transformación, las cuales se ejecutarán con anterioridad a la ejecución del procedimiento de que se trate. También es posible forzar su ejecución con el comando **EXECUTE**.

Las **instrucciones de SPSS** se componen de un comando al que se suele acompañar de una especificación adicional según una sintaxis establecida así como una serie de subcomandos que realizan funciones concretas dentro de cada comando, muchos de ellos con especificaciones también adicionales. De forma general un comando se puede representar de esta forma:

**COMANDO**  
/SUBCOMANDO = PALABRA\_CLAVE

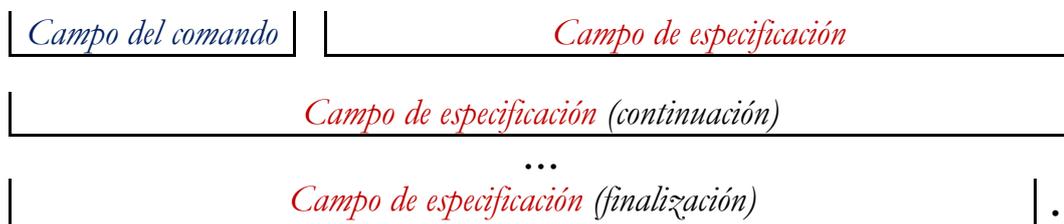
En el caso anterior la sintaxis ejecutada para pedir las frecuencias fue:

**FRECUENCIAS VARIABLES=P31**  
/PIECHART PERCENT  
/ORDER=ANALYSIS.

En la instrucción el comando es el procedimiento **FRECUENCIAS**, y en ella se especifican las variables con la palabra clave **VARIABLES**, y también dos subcomandos, el que nos proporciona el gráfico de sectores con los porcentajes, **/PIECHART PERCENT**, y el que determina cómo aparecen las tablas de los estadísticos, **/ORDER=ANALYSIS**.

Las instrucciones de SPSS se disponen en una o más líneas que deben terminar siempre con un **punto**<sup>8</sup>. En el formato general de estas instrucciones de SPSS se pueden distinguir dos zonas según la representación del Gráfico III.1.17.

Gráfico III.1.17 Esquema de una instrucción en SPSS



El **campo del comando** (o campo de control) constituye el espacio destinado a los comandos o palabras clave (*keywords*) predefinidas en el lenguaje de comandos de SPSS destinados a la realización de cada acción. El **campo de especificación** es el espacio reservado para completar cada comando y donde se incluye la información dada por el usuario/a según los objetivos del proceso. Incluye:

- Palabras clave: **ALL AND BY EQ GE GT LE LT NE NOT OR TO WITH**.
- Nombres de las variables: de 64 caracteres como máximo, empiezan con una letra del alfabeto (**A-Z**) o con los signos **@, #, \$**, el resto puede ser además un número, un **.** o un **\_**. Pero no pueden acabar en punto, ni valen los espacios o caracteres especiales como **!, ?, ' o \***. Es indiferente mayúsculas o minúsculas, y se conserva la forma.
- Etiquetas de valores y/o variables.
- Valores de las variables.
- Operadores aritméticos (**+ - \* / \*\***) y lógicos.

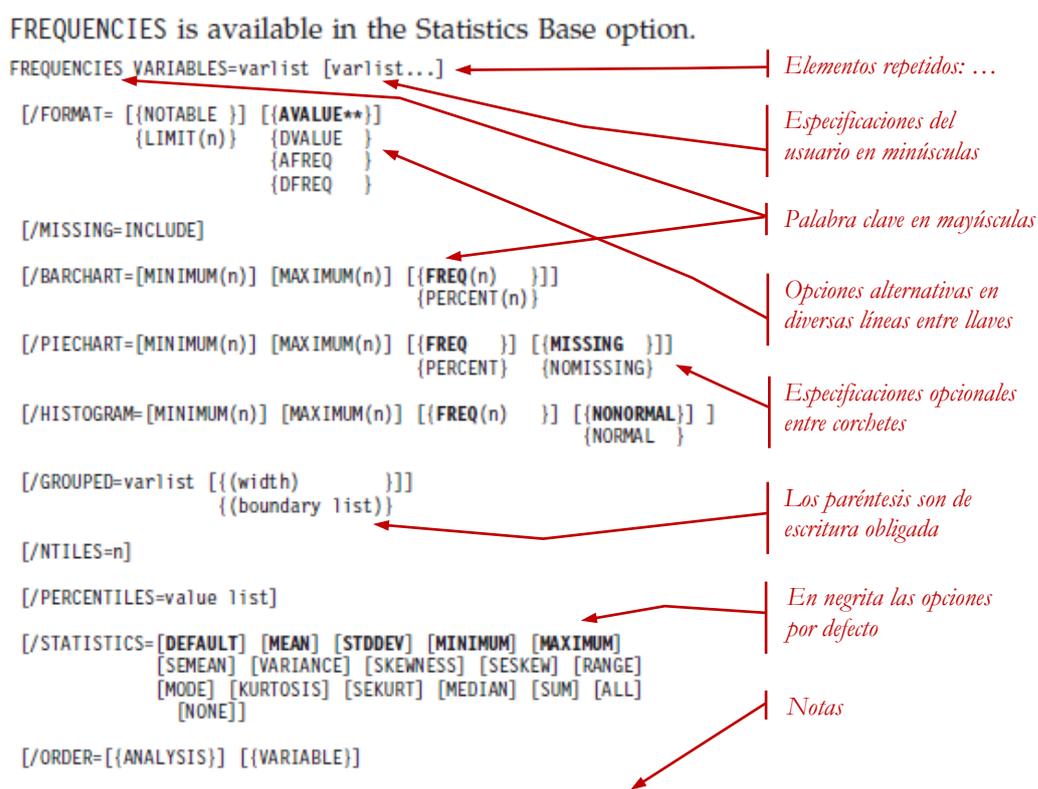
<sup>8</sup> Si la línea siguiente a la instrucción está en blanco no es necesario el punto, no obstante se recomienda su uso.

- Funciones.
- Subcomandos (se separan con un /).
- Delimitadores especiales: espacio en blanco, / ( ) = , \.

Estos elementos, en terminología del SPSS, se denominan **universales** (*universals*).

Cada comando tiene sus propias especificaciones que se resumen en un esquema denominado **diagrama de sintaxis** (*syntax diagram*) con el que se representa el contenido de las instrucciones de SPSS que se puede elaborar con cada uno de los comandos a partir de los universales proporcionando una ayuda específica de gran utilidad. Tomemos como ejemplo el comando destinado a obtener tablas de frecuencias, **FREQUENCIES**. En el Gráfico III.1.18 adjunto se presenta dicho esquema donde se han señalado diversos aspectos para interpretarlo. A este esquema se puede acceder a través del botón de ayuda de la sintaxis  del editor. La información completa sobre el lenguaje de comandos se puede consultar en el manual de referencia de sintaxis (*Command Syntax Reference Guide*).

Gráfico III.1.18 Diagrama de sintaxis de un comando de SPSS: **FREQUENCIES**



\*\* Default if subcommand is omitted or specified without keyword.

Fuente: IBM SPSS Corporation

Es habitual que la mayor parte de las especificaciones sean optativas por lo que una instrucción puede ser muy escueta. Por ejemplo para pedir sencillamente la tabla de frecuencias de la variable P31 del caso anterior es suficiente con:

```
FREQUENCIES P31.
```

Incluso es posible abreviar cada comando hasta los tres primeros caracteres:

**FRE P31.**

### 1.7. Errores de ejecución

Una de las ventajas que proporciona trabajar desde los menús es que no es posible cometer errores a la hora de solicitar la ejecución de un procedimiento. Pero cuando escribimos las instrucciones con el lenguaje de comando es habitual que nos equivoquemos en su elaboración por lo que podemos cometer diversos tipos de errores.

Algunos de estos errores sintácticos se evitan en el editor de sintaxis pues el sistema SPSS mantiene de color rojo el comando hasta que la instrucción no se escribe correctamente. Cuando el sistema SPSS encuentra algún tipo de error en la ejecución de una instrucción actúa en función de ésta y envía un mensaje de error justo en la ubicación donde éste ha tenido lugar e indicando la acción que ha tomado. Los mensajes y errores en SPSS son de dos tipos:

- 1) *Warning*: error no fatal que anula la ejecución de una parte de la instrucción, indica o avisa de la existencia de algún error no excesivamente importante, no afecta a la ejecución del programa pero sí a los resultados. Por ejemplo, la etiqueta de una variable no reconocida o la definición incorrecta del formato de una variable.
- 2) *Error*: error fatal que anula la ejecución de la instrucción donde se encuentra el error. Los errores cometidos pueden encontrarse en una de las tres situaciones siguientes:
  - El error puede obligar a saltarse la ejecución de un comando pero que no impiden que el proceso continúe. Por ejemplo, una variable no definida en el sistema dentro de la especificación de un comando de procesamiento estadístico.
  - Pueden ser errores que producen la interrupción del proceso pero continúan rastreando la existencia de errores sintácticos del fichero de instrucciones. Por ejemplo, un paréntesis en una expresión de operaciones aritméticas.
  - O pueden ser errores que producen la interrupción inmediata del proceso de ejecución del fichero de comandos. Por ejemplo, porque existen muchos errores acumulados o no se dispone de memoria suficiente.

Los errores aparecen notificados en el editor de resultados. También en el editor de sintaxis se dispone de un espacio denominado **panel de error** donde muestran los errores de ejecución indicando el número de línea (ver Gráfico III.1.15).

Por último, cabe indicar la presencia de los errores que no identifica el sistema SPSS pues no son errores de programación sino del usuario. Por ejemplo, si recodificamos el valor 4 con el valor 3 cuando en realidad debería ser un 2, el sistema lo considerará como 3, no dará ningún mensaje de error pero lo estaremos cometiendo. Por eso es importante la constante verificación de los resultados obtenidos. En otros casos los errores pueden derivarse de un uso indebido de las técnicas de análisis de datos, pongamos por caso calcular la media de la variable “estado civil”, el paquete estadístico da el resultado sin más pero cuya interpretación carece de sentido.

## 2. El paquete estadístico R

 es un software estadístico que se ha ido convirtiendo desde su creación en 1993 en uno de los programas informáticos para el análisis de datos más utilizados en todo el mundo. En todas las disciplinas científicas y ámbitos profesionales de aplicación, R empieza a ser una referencia obligada. ¿A qué se debe este éxito? Sugerimos la respuesta diciendo que cumple con esa máxima de ser **bueno, bonito y barato**. Principalmente por su inigualable relación calidad/precio. Es un producto de primera calidad, potente y eficiente, flexible, que se ha extendido mundialmente con una red de apoyo entre usuarios valiosísima, al mejor precio posible: software libre que se distribuye gratuitamente bajo licencia GNU<sup>9</sup>. La razón estética se podría discutir, pero la capacidad de elaboración de gráficos de alta calidad bien merece un reconocimiento destacado<sup>10</sup>.

R es un lenguaje y un entorno de trabajo destinado al tratamiento y análisis de datos estadísticos y a la elaboración de gráficos. Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka<sup>11</sup> del Departamento de Estadística de la *University of Auckland* (Nueva Zelanda) en 1993 (Ihaka y Gentleman, 1996) y actualmente es desarrollado por el *R Development Core Team*. Originalmente R se programó en el lenguaje S y ha heredado su carácter de lenguaje de programación orientado a objetos. Su página web es: <http://www.r-project.org/> (ver Gráfico III.1.19). La primera versión, la 1.0.0, apareció en internet en el año 2000 (Fox, 2009).

R es un paquete de programas que se va ampliando y mejorando de forma constante gracias a la incesante actividad, generosa y no lucrativa, de la comunidad R. En terminología de R cada programa concreto destinado a la realización de una tarea específica de análisis se denomina *package*. Cuando se instala el programa se dispone de un conjunto **base** de paquetes que se pueden ir ampliando desde el propio programa accediendo al CRAN (*Comprehensive R Archive Network*) en la dirección <http://cran.r-project.org>. En el momento de escribir este texto se pueden encontrar más de 6.000 paquetes en el CRAN (<http://cran.r-project.org/web/packages/>).<sup>12</sup>

Según Maurandi et al. (2013) destacan entre sus ventajas, además de ser un software libre:

- Es un programa multiplataforma y, por tanto, existen versiones para Mac, iPhone, Linux o Windows.
- A través del software R es posible analizar cualquier tipo de datos. Destaca por su compatibilidad con todos los formatos de datos (txt, csv, xls, sav, sas,...).

<sup>9</sup> Sistema operativo que es software completamente libre, para su uso, modificación y distribución, evitando también su comercialización (ver <http://www.gnu.org/>). GNU, en inglés “ñu”, es un acrónimo recursivo de “GNU No es Unix”.

<sup>10</sup> Seguramente R se va a imponer en el concierto mundial, Elosua y Etxebarria (2012: 9) señalan que “*La variedad y cantidad de modelos y análisis que ofrece R no es comparable con los ofrecidos por el software privativo. Ninguno de los paquetes comerciales podría competir con R. Como garantía de su validez y actualidad podríamos citar que los paquetes comerciales más importantes están incorporando módulos de conexión para R*”. SPSS entre ellos.

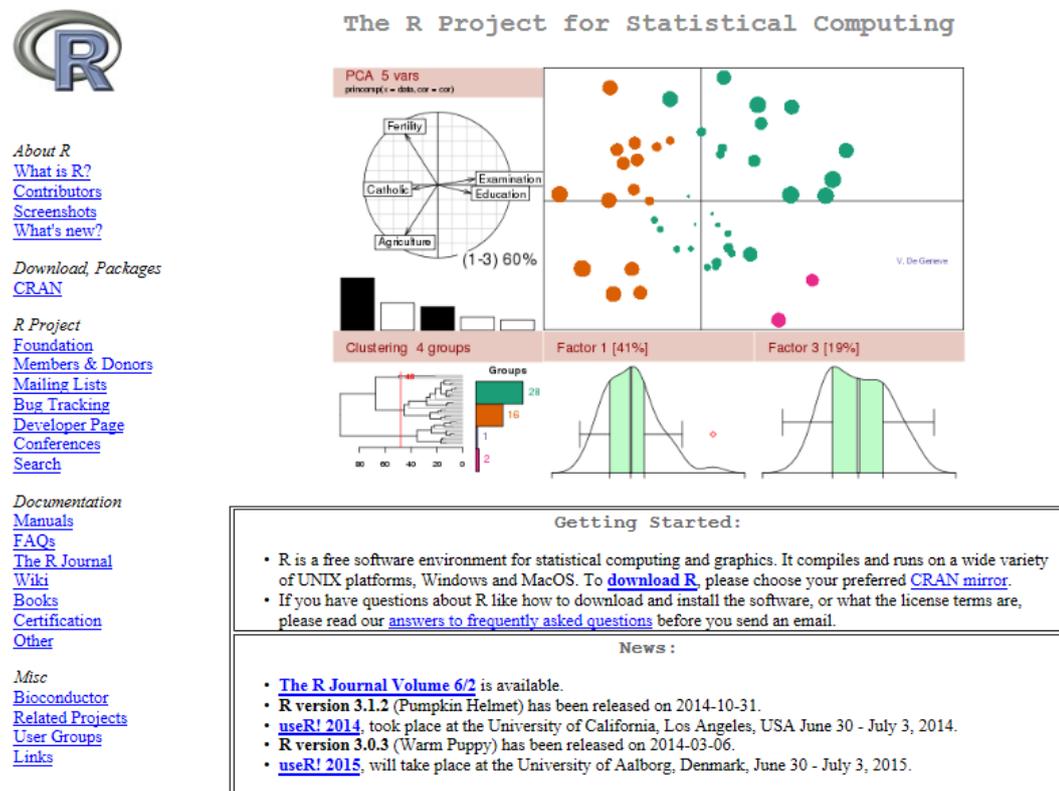
<sup>11</sup> El nombre del programa informático, R, responde en parte a las iniciales de los dos creadores, y a un juego de palabras con el lenguaje del cual deriva, S (Hornik, 2013).

<sup>12</sup> R-Bloggers (<http://www.r-bloggers.com/>) es una forma de seguir y estar al día sobre R. Es un *hub* de los contenidos recogidos de *bloggers* que escriben sobre R, un sitio que proporciona información y ayuda para los usuarios de R para conectar y seguir la “Blogosfera R”.

- Tiene una capacidad gráfica muy buena en relación a los otros paquetes estadísticos.
- Existen miles de técnicas estadísticas que se han implementado en R, y siguen aumentando día tras día. Así, la instalación base del programa es ampliable mediante extensiones para realizar análisis específicos para infinidad de disciplinas.

Peró et al. (2012: 2) señalan que su carácter de libre acceso además de permitir poder conocer con detalle, modificar y distribuir los distintos procedimientos y de facilitar una mayor difusión la reproductibilidad de los hallazgos científicos *tiene una clara connotación social ya que implica la distribución de recursos tecnológicos a colectivos que no pueden disponer de recursos económicos suficientes para adquirir un software comercial*. Es habitual encontrar a propósito de R una frase recurrente que lo identifica hoy en día como la *lingua franca* del procesamiento y análisis de datos.

Gráfico III.1.19 Página de inicio de la web de R (31 de enero 2015)



This server is hosted by the [Institute for Statistics and Mathematics](#) of [WU \(Wirtschaftsuniversität Wien\)](#).

R tampoco deja de tener inconvenientes: el principal, para un iniciado en ciencias sociales y humanas, el lenguaje de programación; también la interfaz gráfica básica, la no actualización de algunos paquetes, las dificultades de instalación y funcionamiento en un entorno, las dificultades o la imposibilidad de utilizar los acentos o las dificultades para trabajar con datos de encuesta ya no se conciben valores y etiquetas o valores perdidos diferenciados. Pero son obstáculos en parte soslayables, que esperamos se resuelvan con el tiempo, en relación con las ventajas que pueden obtenerse.

La curva de aprendizaje de R, como también sucede con las técnicas de análisis de datos estadísticos, es lenta y costosa al inicio. Trabajar con R implica conocer el **lenguaje de programación de R**, significa tener que elaborar programas de instrucciones que se irán ejecutando a medida que se vayan enviando para ser procesados, en vez de trabajar con menús y cuadros de diálogo donde a base de clics de ratón se modelizan o parametrizan los distintos procedimientos. Este aspecto tiene la gran ventaja de que permite al usuario un mayor control de los procedimientos que ejecuta<sup>13</sup>.

De esa forma, con una sintaxis particular, construiremos series de instrucciones más o menos detalladas y/o numerosas que nos permitirán ejecutar un tipo de análisis o una tarea determinada. Esta forma de trabajo es más larga y laboriosa, y para una persona formada en ciencias sociales o humanidades que no tiene un bagaje formativo formal en este tipo de conceptos y lenguaje puede resultar al principio algo complicado. Pero progresivamente se irá evidenciando que es una opción más flexible, rica y adaptable a nuevas necesidades que un entorno gráfico de menús y cuadros de diálogo predefinidos no nos facilita. No obstante, en una etapa inicial del aprendizaje y mediante la utilización de **interfaces gráficas** (denominados GUI, *Graphical User Interface*) veremos cómo se facilita el uso y el aprendizaje de R para realizar algunas tareas de análisis preestablecidas en un entorno gráfico más amigable.

Entre las diversas alternativas de interfaces gráficas trabajaremos con **Deducer**<sup>14</sup>. Se trata de una interfaz que nos facilita la utilización de unos pocos paquetes de R. Seguramente con el tiempo se irán ampliando las opciones de análisis con este tipo de entorno, pero hasta el momento es limitado. La práctica de la investigación se extiende a necesidades del analista que están limitadas por la configuración actual de estos editores, ya sea porque alguna opción del paquete en cuestión no está disponible ya sea porque sencillamente no se implementó en este entorno. El conocimiento progresivo de estas herramientas ilustrará la necesidad de realizar tareas fuera del mismo, circunstancia que obliga a combinar y conocer ambas forma de trabajo con el software.

Para trabajar con el lenguaje de programación de R el usuario/a maneja un intérprete de línea de comandos a través de una consola. La instalación inicial de R proporciona esta consola básica. Alternativamente se han creado editores de texto de R (denominados IDE, *Integrated Development Environment*) que trabajan con versiones mejoradas de la consola que nos facilitan el trabajo de programación y la visualización y tratamiento de los resultados. Entre las distintas posibilidades existentes optamos por **R-Studio**<sup>15</sup>.

En los apartados que siguen introduciremos las principales características del paquete estadístico y el procesamiento de los datos a través de las dos alternativas de ejecución de instrucciones: por menús con Deducer y con R-Studio para emplear el lenguaje de programación. Posteriormente, y en el contexto de cada procedimiento de análisis

---

<sup>13</sup> Este control en R se extiende a la posibilidad de disponer del código abierto con el que se programa todo el software lo que permite que cada paquete pueda ser ampliado o reprogramado, o también a crear nuevos paquetes, a partir de nuevas contribuciones con la ayuda de especialistas en programación.

<sup>14</sup> Una alternativa posible de editor gráfico es **R Commander**, desarrollada por el sociólogo John Fox (2005). Otras posibilidades son: *Reward*, *R.NET*, *Poor Man's GUI* o *RExcel*.

<sup>15</sup> Otros interfaces para la programación con R son: *Tinn-R*, *WinEdit* o *Emacs*.

veremos más específicamente y con mayor detenimiento las distintas técnicas de análisis que se tratan.

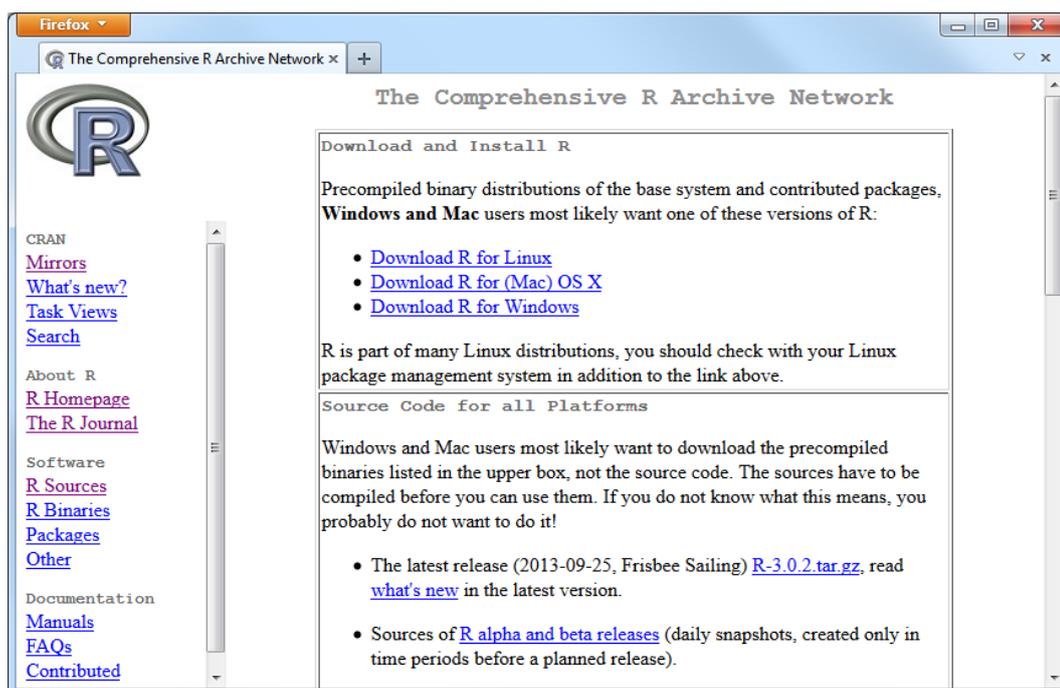
## 2.1. Características generales de R

Presentaremos algunos aspectos fundamentales para empezar a trabajar con R: cómo instalar el programa, practicaremos una primera toma de contacto con el programa base, dónde encontrar recursos de ayuda, y describiremos la lógica de funcionamiento general del software. Trabajaremos inicialmente con la **versión 3.1.2.**, la última versión existente al escribir el texto. Posteriormente veremos que al instalar el paquete Deducir deberemos trabajar con una versión anterior que es la compatible con algunas funciones de este entorno gráfico.

### 2.1.1. Instalación de R

R se descarga desde la página oficial: <http://cran.r-project.org>, eligiendo la plataforma de cada usuario/a (*Windows*, *Linux* o *MacOSX*).

Gráfico III.1.20 Página web de R: The Comprehensive R Archive Network (CRAN)



En el momento de escribir estas páginas manejamos la **versión 3.1.2**, que instalamos en una plataforma *Windows*, ya sea en un entorno de 32 o 64 bits.

#### R-3.1.2 for Windows (32/64 bit)

[Download R 3.1.2 for Windows](#) (54 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

Se descargará el archivo **R-3.1.2-win.exe** que se tendrá que ejecutar a continuación. Como administradores el proceso de instalación no conlleva ninguna dificultad:

1. Se siguen los sucesivos cuadros de diálogo.
2. Se elige la carpeta de instalación.
3. Se eligen los componentes (habitualmente *Core* y *32-bit Files*).
4. Se elige si se utilizan las opciones de configuración de la instalación. No es necesario (opción por defecto “No”) pero se pueden configurar o dejar igual:
  - El modo de “display”: una ventana (opción MDI) vs. ventanas separadas (opción SDI).
  - El estilo de ayuda: simple o HTML.
  - El acceso a Internet: Estándar o Internet2.
  - La carpeta del menú inicio.
  - Tareas adicionales: icono de escritorio o de inicio rápido y el registro de entradas.

Cuando acabemos nos aparecerá el mensaje de finalización de la instalación y dispondremos de un acceso a la aplicación.



Posteriormente veremos que al instalar la interfaz gráfica **Deducer** tendremos diversas opciones que nos harán elegir entre mantener esta instalación y añadir Deducer, instalar conjuntamente R y Deducer con un instalador único en la versión **2.15.0** o bien trabajar con Deducer y la versión la **2.15.3** para poder utilizar algunas ampliaciones de Deducer.

### 2.1.2. Iniciando R

Al abrir la aplicación para trabajar con R nos aparecerá una ventana como la siguiente que recibe el nombre de **consola de R**:

Gráfico III.1.21 Consola de R

 La imagen muestra una ventana de software con el título 'RGui (32-bit) - [R Console]'. La barra de menús incluye 'Archivo', 'Editar', 'Visualizar', 'Misc', 'Paquetes', 'Ventanas' y 'Ayuda'. La barra de herramientas contiene iconos para abrir, guardar, imprimir, refrescar y otros. El área principal muestra el texto de bienvenida de R:
 

```
R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribución.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```

En color rojo, al final, aparece el símbolo del sistema (el *prompt*): `>` y el cursor `|` posicionado esperando la introducción de una instrucción en el lenguaje de programación de R. La barra de menús nos ofrece algunas posibilidades de tareas a realizar pero no es el lugar donde se pueden encontrar las instrucciones de los procedimientos de R para el análisis estadístico y la obtención de representaciones gráficas.

Para trabajar en este entorno podemos hacerlo de forma interactiva ejecutando una instrucción cada vez y viendo el resultado a continuación, o podemos elaborar programas con diversas instrucciones (denominados *scripts*) que se pueden ejecutar en bloque.

Probemos ejecutar una instrucción tan sencilla como sumar 2 más 2. Escribimos `2+2` y apretamos sobre la tecla `<Enter>`, obtendremos el resultado esperado:

```
> 2+2
[1] 4
```

Ahora repetamos la misma operación pero asignando el resultado de la suma a un **objeto**. Si a este objeto le llamamos `suma` la asignación se realiza así:

```
> suma = 2+2
```

o bien:

```
> suma <- 2+2
```

El operador de asignación `=` es equivalente al operador `<-`<sup>16</sup>. A continuación para visualizar el resultado, basta con escribir `suma` y darle a `<Enter>`:

```
> suma
[1] 4
```

Evidentemente necesitamos conocer el lenguaje de comandos de R para seguir avanzando. Volveremos más adelante con más detalle a tratar la cuestión de cómo elaborar instrucciones con el lenguaje de programación pero con un entorno de trabajo distinto, el de R-Studio, que sigue en esencia la misma lógica que acabamos de ver. Mientras tanto veremos poco a poco algunas instrucciones con comandos de R.

### 2.1.3. La ayuda y documentación de R

Para trabajar con R y profundizar en su conocimiento podemos consultar diversas fuentes de información, referencias bibliográficas y recurrir a la ayuda del propio programa.

El primer recurso es la propia página web de R donde se dispone de una ayuda muy completa sobre todas las funciones, procedimientos y elementos que configuran el

<sup>16</sup> Este último se puede escribir en la consola de R mediante las teclas `<ALT><->`.

lenguaje, así como manuales y preguntas frecuentes. Desde la página principal se puede ir al apartado de documentación y encontraremos diversos recursos:

**Manuals:** para diversas plataformas y editados por el *R Development Core Team*, en <http://cran.r-project.org/manuals.html>. Por ejemplo se puede consultar el manual introductorio en línea: <http://cran.r-project.org/doc/manuals/R-intro.html>.

**FAQs:** preguntas más frecuentes, en <http://cran.r-project.org/faqs.html>.

**Books:** una selección extensa de libros relacionados con R y S, en <http://www.r-project.org/doc/bib/R-books.html>

**Other:** documentación diversa, revistas, manuales y tutoriales de los usuarios de R, en <http://www.r-project.org/other-docs.html>. En particular, en el CRAN: <http://cran.r-project.org/other-docs.html>

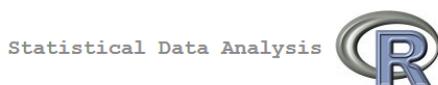
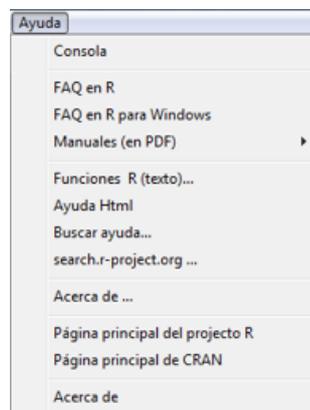
**The R Journal:** la revista *open access* del *R Project* para usuarios y desarrolladores, en <http://journal.r-project.org/>

En el CRAN se puede consultar también información de los diferentes **paquetes**, a través de su información general, en <http://cran.r-project.org/web/packages/>, o bien consultando las *Task Views* que son colecciones de paquetes por áreas específicas que permiten también su instalación conjunta: <http://cran.r-project.org/web/views/>.

Además de la bibliografía al final de este capítulo, otros recursos, manuales y enlaces en relación a R los recogemos y actualizamos en la página: <http://pagines.uab.cat/plopez/content/r>.

En la consola de R disponemos del menú de ayuda que nos permite acceder a algunos de los lugares que hemos comentado y a otros específicos.

En particular, podemos elegir el buscador **R Site Search** (<http://search.r-project.org/>), que tiene un comando específico en el lenguaje de R: **RSiteSearch**, o si clicamos sobre **Ayuda Html** accederemos a la página web de R con diversos enlaces a recursos de ayuda del software:



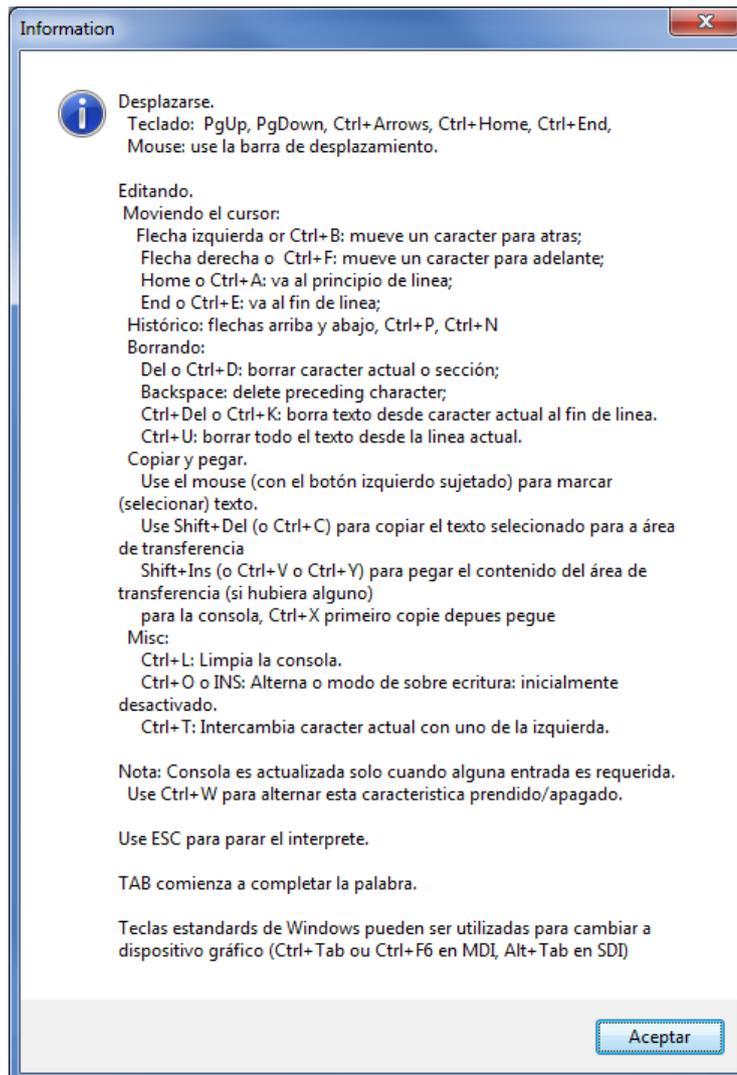
Manuals	
<a href="#">An Introduction to R</a> <a href="#">Writing R Extensions</a> <a href="#">R Data Import/Export</a>	<a href="#">The R Language Definition</a> <a href="#">R Installation and Administration</a> <a href="#">R Internals</a>
Reference	
<a href="#">Packages</a>	<a href="#">Search Engine &amp; Keywords</a>
Miscellaneous Material	
<a href="#">About R</a> <a href="#">License</a> <a href="#">NEWS</a>	<a href="#">Authors</a> <a href="#">Frequently Asked Questions</a> <a href="#">User Manuals</a>
<a href="#">Resources</a> <a href="#">Thanks</a> <a href="#">Technical papers</a>	
Material specific to the Windows port	
<a href="#">CHANGES up to R 2.15.0</a>	<a href="#">Windows FAQ</a>

Resultado que obtenemos también si ejecutamos el comando:

```
> help.start()
```

También es interesante en un inicio consultar la ayuda sobre el funcionamiento de la consola de R y la línea de comandos, menú **Ayuda / Consola**, donde nos aparece el resumen de las funciones del Gráfico III.1.22.

Gráfico III.1.22 Ayuda del funcionamiento de la consola de R



Desde la consola podemos igualmente utilizar el comando de ayuda: **help** o **?**. Cada vez que lo ejecutamos se accede a la ayuda que se encuentra en internet, lo que exige trabajar siempre con una conexión. El comando funciona escribiendo primero la palabra clave **help** y a continuación especificando entre paréntesis la palabra clave sobre la que deseamos ayuda, la cual debe ir entre dobles comillas. Por ejemplo, si pedimos ayuda sobre el propio comando de ayuda:

```
> help("help")
```

O de forma equivalente:

```
> ?help
```

Accedemos a la página de documentación de R sobre el comando. En el Gráfico III.1.23 se presenta un fragmento comentando la estructura general de la documentación de ayuda de los comandos de R.

Gráfico III.1.23 Extracto de la ayuda del comando **help** de R

The screenshot shows the R help page for the 'help' command. Red arrows point from descriptive text to specific parts of the page:

- Documentation**: Points to the title 'help {utils}' and the 'Documentation' header. Annotation: *Nombre la función o comando y entre {} el paquete al que pertenece*
- Description**: Points to the 'Description' header and the text 'help is the primary interface to the help systems.'. Annotation: *Breve descripción*
- Usage**: Points to the 'Usage' header and the code block for the 'help' function. Annotation: *Especifica cómo utilizar el comando a través de sus argumentos*
- Arguments**: Points to the 'Arguments' header and the list of arguments (topic, package, lib.loc, verbose, try.all.packages, help\_type). Annotation: *Detalla los argumentos y las opciones*
- Details**: Points to the 'Details' header and the list of help types. Annotation: *Comentarios adicionales de funcionamiento*
- Additional sections**: Points to the bottom of the page. Annotation: *La ayuda se completa con otros apartados más: "Offline help", "Note", "References", "See Also" y "Examples"*

Este comando exige conocer el comando u operador sobre el que se desea información. Si lo que queremos hacer es una búsqueda en la ayuda se puede emplear el comando **help.search** o **??**, por ejemplo para hacer una búsqueda de la palabra clave **survey**:

```
> help.search("survey")
```

O de forma equivalente:

```
> ??survey
```

El resultado es el siguiente:

Search Results 

---



The search string was "survey"

---

Help pages:

<a href="#">MASS::housing</a>	Frequency Table from a Copenhagen Housing Conditions Survey
<a href="#">MASS::survey</a>	Student Survey Data
<a href="#">survival::lines.survfit</a>	Add Lines or Points to a Survival Plot
<a href="#">survival::print.summary.survexp</a>	Print Survexp Summary
<a href="#">survival::summary.survexp</a>	Summary function for a survexp object
<a href="#">survival::survexp</a>	Compute Expected Survival
<a href="#">survival::survexp.fit</a>	Compute Expected Survival
<a href="#">survival::survexp.us</a>	Census Data Sets for the Expected Survival and Person Years Functions

Una forma particular de ayuda es ejecutar el comando **example** que nos proporciona información de cómo utilizar un comando. Por ejemplo, si nos interesamos sobre cómo utilizar el procedimiento que calcula medias, el comando o función **mean**, haremos:

```
> example("mean")
```

El resultado es:

```
mean> x <- c(0:10, 50)
mean> xm <- mean(x)
mean> c(xm, mean(x, trim = 0.10))
[1] 8.75 5.50
```

Se presenta un ejemplo donde se crea un objeto de R, un vector de nombre **x**, que contiene la concatenación (operador **c**) de los valores desde 0 hasta 10 y el 50 (es decir, 0 1 2 3 4 5 6 7 8 9 10 50), se calcula la media de este vector o variable y se pone en el objeto **xm**, operación que da 8,75, y finalmente se presenta concatenada la información de la media y la media recortada que elimina el 10% de casos extremos (en este caso el 50, que es un valor muy alto en relación a los demás) dando el valor 5,5. Podemos ejecutar cada uno de estos comandos:

```
> x=c(0:10,50)
> x
[1] 0 1 2 3 4 5 6 7 8 9 10 50
> xm=mean(x)
> xm
[1] 8.75
> c(xm,mean(x,trim=0.10))
[1] 8.75 5.50
```

Si queremos información sobre un paquete de R particular<sup>17</sup>, por ejemplo `foreign` que permite importar datos de otros softwares, podemos escribir en la consola:

```
> help(package="foreign")
```

Otra forma de ayuda es la `vignette`, una descripción breve que añade el creador de un paquete. Para visualizar las viñetas de los paquetes que tengamos instalados:

```
> vignette(all=TRUE)
```

Una opción interesante es visualizar las demostraciones de los paquetes que nos ayudan a aprender su uso a través de ejemplos con el comando o función `demo`. Si escribimos `demo()` se listan todas las demostraciones disponibles en los paquetes que tengamos instalados. Podemos ejecutar en particular las siguientes:

```
> demo("graphics")
> demo("persp")
> demo("plotmath")
> demo("colors")
```

La ayuda se puede obtener también a través de la comunidad R, a través del *mailing list* (<http://www.r-project.org/mail.html>) o mediante búsquedas en internet, imprescindibles y fructíferas.

#### 2.1.4. Instalación de paquetes

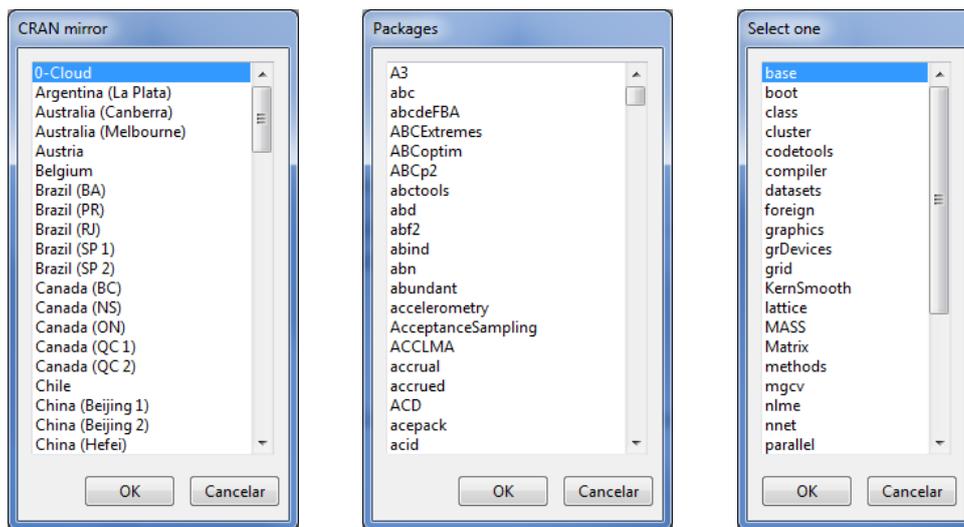
Los paquetes son colecciones de funciones de R, datos y código compilado en un formato definido. La primera instalación de R incorpora los paquetes base que se pueden ir ampliando. Los paquetes los podemos encontrar en el *Comprehensive R Archive Network* o CRAN (<http://cran.r-project.org/>)<sup>18</sup> que es el repositorio oficial de paquetes disponibles para R. Actualmente se encuentran más de 6.000 paquetes, que los diversos usuarios generan de forma abierta y colaborativa, y que engloban una gran diversidad de temáticas.

El CRAN se encuentra en diversos servidores espejo en todo el mundo y cuando queremos descargar uno para ampliar nuestra instalación de paquetes de R tendremos que seleccionar de una lista de servidores el que queramos: el de la nube (0-Cloud), el que esté más cercano físicamente (Spain, Madrid) o cualquier otro. Desde la consola se accede al *CRAN Mirror* a través del menú: **Paquetes / Seleccionar espejo CRAN** (ver Gráfico III.1.24a). El paquete que deseemos instalar del CRAN los encontraremos listado a través de: **Paquetes / Instalar paquete** (ver Gráfico III.1.24b). Los paquetes que se descargan están disponibles entre las librerías de nuestra instalación pero no están cargados en el sistema de R hasta que no se indica. Para ello es necesario seleccionar el paquete y cargarlo a través de **Paquetes / Cargar paquete** (ver Gráfico III.1.24c).

<sup>17</sup> El comando funciona para los paquetes que tenemos bajados en nuestra librería o carpeta donde está instalado R y se almacenan los paquetes. Las librerías se encuentran en nuestro ordenador en la subcarpeta `library` que cuelga de la carpeta principal donde esté instalado R.

<sup>18</sup> Ver Gráfico III.1.20

Gráfico III.1.24 Listado de servidores del CRAN y paquetes en la consola de R



a. Espejo del CRAN

b. Paquetes para descargar

c. Paquetes disponibles

Así pues para proceder a instalar un paquete adicional que necesitemos en R seguiremos los siguientes pasos:

- 1) En el menú **Paquetes / Instalar paquete** elegimos el nuevo paquete.
- 2) Nos aparecerá la imagen del **CRAN Mirror** y elegiremos el servidor de descarga.
- 3) Para usarlo deberemos cargarlo a través de **Paquetes / Cargar paquete**.

Haremos seguidamente un ejercicio de instalación y carga desde la consola de R. Previamente ejecutaremos el comando **search** que nos informa de los paquetes cargados en este momento en la sesión de trabajo:

```
> search()
[1] ".GlobalEnv"      "package:grid"    "package:stats"   "package:graphics"
[5] "package:grDevices" "package:utils"   "package:datasets" "package:methods"
[9] "Autoloads"       "package:base"
```

Y **library** que nos informa de los paquetes o **librerías** que tenemos descargados:

```
> library()
Packages in library 'C:/Program Files/R/R-3.1.2/library':

base           The R Base Package
boot           Bootstrap Functions (originally by Angelo Canty for S)
class          Functions for Classification
cluster        Cluster Analysis Extended Rousseeuw et al.
codetools      Code Analysis Tools for R
compiler       The R Compiler Package
datasets       The R Datasets Package
foreign        Read Data Stored by Minitab, S, SAS, SPSS, Stata,
               Systat, Weka, dBase, ...
graphics       The R Graphics Package
grDevices      The R Graphics Devices and Support for Colours and Fonts
grid           The Grid Graphics Package
KernSmooth     Functions for kernel smoothing for Wand & Jones (1995)
lattice        Lattice Graphics
MASS           Support Functions and Datasets for Venables and Ripley's
               MASS
Matrix         Sparse and Dense Matrix Classes and Methods
methods        Formal Methods and Classes
```

mgcv	Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation
nlme	Linear and Nonlinear Mixed Effects Models
nnet	Feed-forward Neural Networks and Multinomial Log-Linear Models
parallel	Support for Parallel computation in R
rpart	Recursive Partitioning and Regression Trees
spatial	Functions for Kriging and Point Pattern Analysis
splines	Regression Spline Functions and Classes
stats	The R Stats Package
stats4	Statistical Functions using S4 Classes
survival	Survival Analysis
tcltk	Tcl/Tk Interface
tools	Tools for Package Development
translations	The R Translations Package
utils	The R Utils Package

Si queremos obtener información de una librería específica, por ejemplo del paquete **base**, podemos ejecutar el comando anterior especificando:

```
> library(help="base")
```

Para instalar un paquete vamos al menú **Paquetes / Instalar paquete**, como indicamos anteriormente, o de forma equivalente podemos ejecutar el comando **install.packages**. Por ejemplo instalaremos el paquete **sna**, que incluye diversas herramientas para el análisis de redes sociales. Previamente podemos buscar documentación sobre el paquete mediante:

```
> RSiteSearch("sna")
```

Y a continuación realizamos la instalación:

```
> install.packages("sna")
--- Please select a CRAN mirror for use in this session ---
probando la URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/sna 2.3-2.zip'
Content type 'application/zip' length 848031 bytes (828 Kb)
URL abierta
downloaded 828 Kb

package 'sna' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Usuario\AppData\Local\Temp\RtmporwTBF\downloaded_packages
```

Una vez instalado correctamente cargamos la librería y ya podemos utilizarla:

```
> library("sna")
sna: Tools for Social Network Analysis
Version 2.3-2 created on 2014-01-13.
copyright (c) 2005, Carter T. Butts, University of California-Irvine
For citation information, type citation("sna").
Type help(package="sna") to get started.
```

Debido a la diversidad creciente de temáticas y paquetes el CRAN se organiza también por temas (*views*) agrupando así los paquetes según su similitud. Las *Task View* reúnen estas colecciones de paquetes de R que comparten un mismo tipo de temática o ámbito de trabajo. En <http://cran.r-project.org/web/views/> podemos ver la relación de temáticas. Cuando se instala una *Task View* automáticamente lo hacen cada uno de los paquetes que lo integran sin tener que hacerlo uno a uno.

Entre las diversas colecciones de paquetes podemos destacar:

- *SocialSciences*: <http://cran.r-project.org/web/views/SocialSciences.html>
- *OfficialStatistics*: <http://cran.r-project.org/web/views/OfficialStatistics.html>
- *Multivariate*: <http://cran.r-project.org/web/views/Multivariate.html>
- *Cluster*: <http://cran.r-project.org/web/views/Cluster.html>
- *Graphics*: <http://cran.r-project.org/web/views/Graphics.html>
- *Distributions*: <http://cran.r-project.org/web/views/Distributions.html>

A medida que se profundiza el conocimiento y se adquiere experiencia, y también se van ampliando las necesidades, tanto en R como en el tratamiento y análisis de datos cuantitativos, se irán incorporando distintos procedimientos de esta inmensidad de alternativas que nos ofrecen hoy en día los paquetes estadísticos y R en particular.

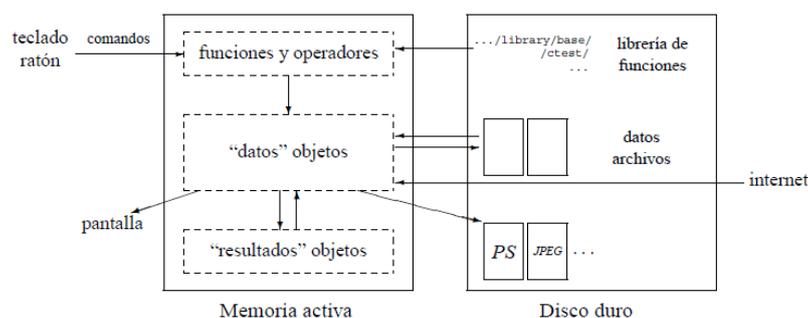
Finalmente comentaremos que un paquete también se puede actualizar (a través del menú **Paquetes / Actualizar paquetes** o con el comando **update.packages**) o desinstalar (el comando **uninstall.packages**). La desinstalación del software es la habitual de cualquier sistema operativo, en el caso de *MS-Windows* a través de **Panel de control / Programas y características**.

### 2.1.5. Funcionamiento de R: objetos y workspace

En el Gráfico III.1.25 se esquematiza la lógica general de funcionamiento de R. El software se caracteriza en primer lugar por ser un programa **orientado a objetos**, es decir, que se identifican como objetos cada uno de los elementos con los que se trabaja: ya sean datos, variables, matrices, cálculos, funciones, resultados de tablas de un análisis, gráficos, funciones, etc. Éstos se almacenan en la memoria activa del ordenador con un nombre específico que identifica al objeto. Los objetos se pueden utilizar total o parcialmente en cada momento y se manipulan a través de operadores (aritméticos, lógicos) y funciones (realizan una acción según unos argumentos, y se consideran también objetos).

Los objetos que son almacenados en la memoria activa del ordenador no usan archivos temporales. Los datos y los resultados pueden guardarse como objetos en la memoria activa, pueden visualizarse en pantalla o ser guardados en el disco duro. Los datos pueden cargarse desde el disco duro o pueden ser leídos desde un servidor remoto a través de red de internet. En el caso particular de los resultados, como son también objetos, pueden considerarse un caso específico de datos (Paradis, 2003: 5).

Gráfico III.1.25 Esquema del funcionamiento de R



Fuente: Paradis (2003: 4)

Cuando se vayan asimilando los principales conceptos del lenguaje de programación de R se irán precisando estas primeras ideas generales.

Un elemento importante y útil en R es el *Workspace*, el **área o espacio de trabajo de R**, es el lugar donde se incluyen los objetos con los que trabaja el programa (variables, matrices de datos,...). Cuando se cierra la aplicación siempre se pregunta si se quiere guardar y así poder volver a cargarlo en la siguiente sesión, lo que significa una gran ventaja y eficiencia en el trabajo. Por defecto el espacio de trabajo se localiza en el directorio de trabajo (*working directory*) que se haya definido en la instalación.

Para ver todos los objetos de nuestro espacio o área de trabajo podemos ejecutar por el menú **Misc / Listar objetos** o bien ejecutar el comando:

```
> ls()
```

Para eliminar objetos utilizaríamos el comando **rm**.

Para ver el espacio de trabajo donde estamos trabajando podemos ir a **Archivo / Guardar área de trabajo** o ejecutar el comando:

```
> getwd()
```

Para cambiar el espacio: **Archivo / Cambiar dir**, o bien ejecutamos el cambio a la carpeta, por ejemplo, **Espacio** de la unidad **c**<sup>19</sup>:

```
> setwd("c:/Espacio")
```

Para guardar el área de trabajo por el menú: **Archivo / Guardar área de trabajo** o con el comando, al que le damos el nombre **Espacio.RData**<sup>20</sup>:

```
> save.image("C:\\Espacio\\Espacio.RData")
```

Para cargar un espacio de trabajo guardado: **Archivo / Cargar área de trabajo** o bien:

```
> load("C:\\Espacio\\Espacio.RData")
```

Todo lo que hemos visto hasta ahora a través de los menús o de los comandos de R se podrá realizar de la misma forma en entornos de trabajo más amigables y con más prestaciones que la consola original de R. Veremos en primer lugar **Deducer** que nos permitirá trabajar con R algunos procedimientos fundamentales y otros básicos a través de un sistema de menús que sigue una forma de trabajo similar a la de SPSS. En segundo lugar veremos **R-Studio** que proporciona un entorno mejorado para programar con el lenguaje de comando de R.

---

<sup>19</sup> Nótese que la dirección de la carpeta en R tiene la barra **no invertida**.

<sup>20</sup> En este caso la dirección de la carpeta tiene **doble** barra invertida.

## 2.2. Deducer



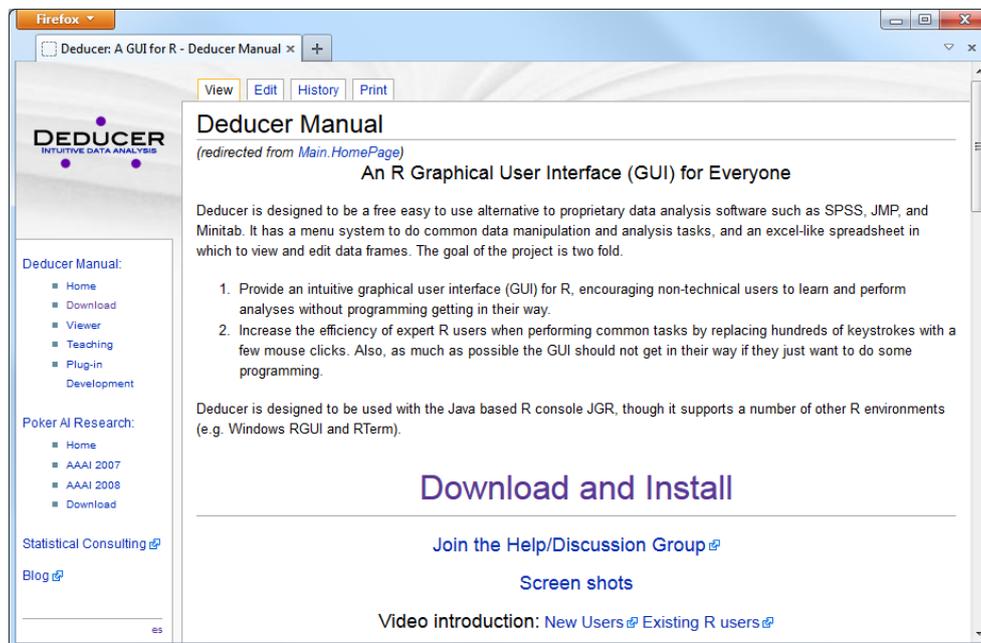
Para empezar a trabajar con R hemos optado por acercar al usuario que se encuentra en la etapa inicial de su aprendizaje a un entorno de trabajo más amigable como es **Deducer**. Se trata de un **GUI** (*Graphical User Interface*) que nos permite trabajar en un entorno gráfico de forma similar a otros paquetes estadísticos, como SPSS, cuando se ejecutan las instrucciones a través del menú y la ayuda de los cuadros de diálogo.

Deducer fue creado por Ian Fellows (Fellows, 2012) con el objetivo de poder manipular y analizar datos con R sin tener conocimientos previos del lenguaje de programación. Está diseñado para ser utilizado con la consola de R basada en Java (**JGR: Java Gui for R**, pronunciado “Jaguar”) si bien soporta otros entornos como *Windows*.

### 2.2.1. Instalación de Deducer

Para instalar el software tendremos tres opciones<sup>21</sup>. Se puede descargar desde la página <http://www.deducer.org>, eligiendo la plataforma de cada usuario/a (*Windows*, *Linux* o *MacOSX*). En nuestro caso trabajaremos con el entorno *Windows* y una alternativa de instalación es optar por ejecutar el *Download installer*. En ese caso la versión actual de Deducer se descarga con la versión 2.15.0 de R. Se descargará por tanto el archivo **Deducer-R-2.15.0-win.exe** que se tendrá que ejecutar a continuación. Si hemos instalado una versión más reciente de R no necesariamente es problemática la nueva instalación pues pueden coexistir dos versiones o más simultáneamente<sup>22</sup>.

### Gráfico III.1.26 Página web principal de Deducer



<sup>21</sup> En <http://www.rforge.net/JGR/index.html> se puede consultar el proceso de instalación.

<sup>22</sup> En todo caso si diera problemas se puede tener una sola versión de R que contenga Deducer y se pueden desinstalar las otras versiones de R.

También podemos optar directamente por instalar Deducer desde la consola de R en la última versión (la 3.2.1 en nuestro caso) instalando el paquete **Deducer**. Para ello ejecutaremos las instrucciones siguientes desde la consola:

```
> install.packages(c("JGR", "Deducer"))
> library(JGR)
> JGR()
```

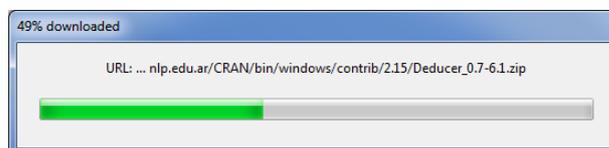
Veremos que además, en las dos opciones citadas, será posible instalar paquetes adicionales de Deducer que son ampliaciones de procedimientos de este entorno, lo haremos desde el menú de Deducer. Pero algunas de estas ampliaciones solamente funcionan en versiones anteriores a la 3.0.0.

Una tercera opción por tanto será instalar la versión anterior a la 3.0.0, la 2.15.3 (de marzo de 2013), es la que aquí utilizaremos<sup>23</sup>, de la siguiente manera:

- 1) Instalar la versión 2.15.3 desde el CRAN de R (<http://cran.r-project.org/bin/windows/base/old/2.15.3/>) como se indica en el apartado 2.1.1.
- 2) Ejecutar la instalación de Deducer, cargar la librería y ejecutar el paquete desde la consola de R con las tres líneas de instrucciones que acabamos de ver en la segunda opción:

```
> install.packages(c("JGR", "Deducer"))
```

Nos aparecen en la consola estos resultados:



```
There is a binary version available (and will be installed) but the
source version is later:
  binary source
Deducer 0.7-6.1 0.7-7
```

```
also installing the dependencies 'lattice', 'zoo', 'png', 'digest', 'gtable',
'reshape2', 'proto', 'RColorBrewer', 'stringr', 'dichromat', 'munsell',
'labeling', 'mvtnorm', 'TH.data', 'sandwich', 'colorspace', 'rJava', 'JavaGD',
'iplots', 'ggplot2', 'scales', 'car', 'multcomp', 'effects', 'plyr', 'e1071'
```

...

```
package 'JGR' successfully unpacked and MD5 sums checked
package 'Deducer' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\Usuario\AppData\Local\Temp\RtmpY19H9t\downloaded_packages
```

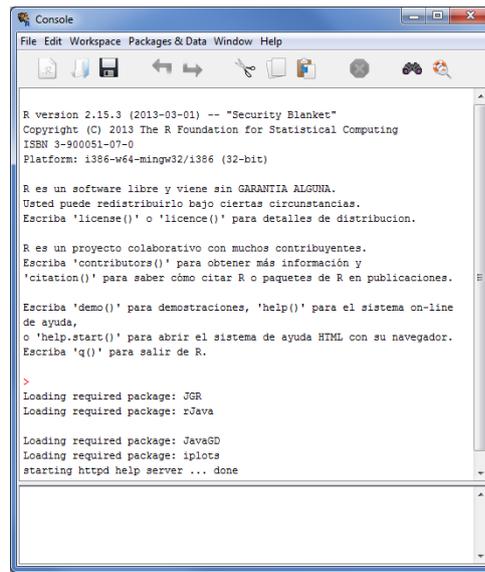
A continuación cargamos y ejecutamos Deducer:

<sup>23</sup> Todo lo que trabajaremos con la ayuda de Deducer en este texto se podrá ejecutar en cualquiera de las tres opciones.

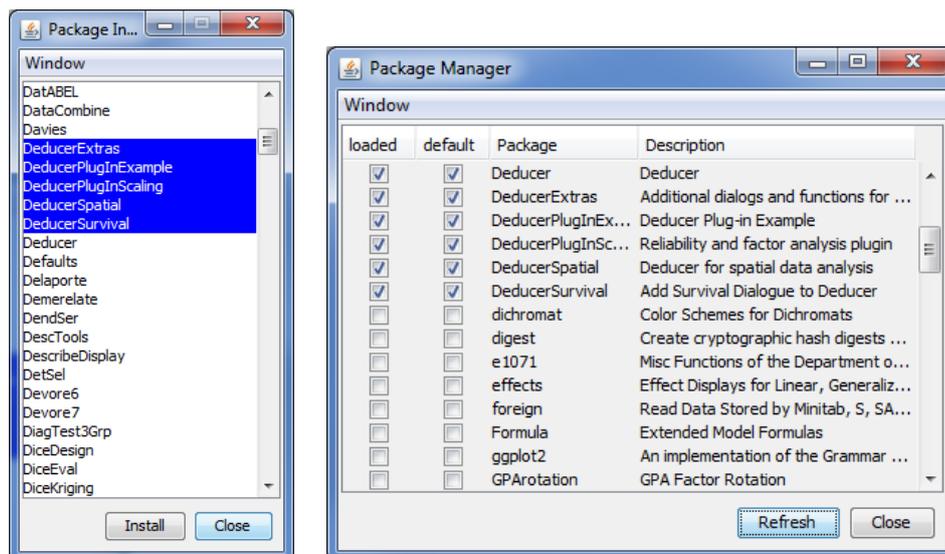
```
> library(JGR)
Loading required package: rJava
Loading required package: JavaGD
Loading required package: iplots

Please type JGR() to launch console. Platform specific launchers (.exe and .app)
can also be obtained at http://www.rforge.net/JGR/files/.
> JGR()
```

Se abrirá la ventana inicial de Deducer:



- 3) Instalar las ampliaciones desde el propio Deducer a través del menú. Primero instalaremos los paquetes de ampliación a través de **Package & Data / Package Installer**: elegimos el *CRAN mirror* y todos los paquetes de Deducer que aparecen en la imagen adjunta.



Cuando acabe la descarga los configuraremos a través de **Package & Data / Package Manager** donde marcamos todos los paquetes de Deducer, tanto en *default* (a partir

de ahora al iniciar Deducer se cargarán siempre) como en *loaded* (se cargan en este momento también). Después de cargarlos veremos cómo se amplían los desplegables del menú de la consola, como en el Gráfico III.1.27.

- 4) Instalaremos también el paquete **DeducerRichOutput**, un complemento interesante que mejora el entorno y la presentación de las salidas de resultados al convertirlos a formato **HTML**. No está disponible en el CRAN sino en la dirección: [https://r-forge.r-project.org/R/?group\\_id=1005](https://r-forge.r-project.org/R/?group_id=1005)<sup>24</sup>.

Para instalarlo bajaremos el archivo:

[http://download.r-forge.r-project.org/bin/windows/contrib/3.0/DeducerRichOutput\\_0.3.7.zip](http://download.r-forge.r-project.org/bin/windows/contrib/3.0/DeducerRichOutput_0.3.7.zip)

Se descomprime y la carpeta de nombre **DeducerRichOutput** se copia dentro de la carpeta **library** en donde esté instalado R (`...\R\R-2.15.3\library`). A continuación desde **Package & Data / Package Manager** activamos el paquete marcando en *default* y en *loaded*. Inmediatamente nos cambiará la imagen de la consola según el nuevo entorno (ver Gráfico III.1.27).

- 5) Si queremos trabajar con Deducer habitualmente necesitaremos abrir R y ejecutar cada vez los dos comandos de cargar la librería y abrir Deducer. Para lanzar o abrir Deducer directamente se puede bajar un pequeño programa: **JGR launcher**, para *MS-Windows* con el enlace [http://www.rforge.net/JGR/web-files/jgr-1\\_62.exe](http://www.rforge.net/JGR/web-files/jgr-1_62.exe). Al ejecutarse se abrirá Deducer.



Cuando se abre Deducer nos aparece la consola o ventana principal de R y se cargan los paquetes que se necesitan para poder utilizar los menús y las opciones de ampliación que estuvieran instaladas (Gráfico III.1.27), junto con la ventana del visor de datos (Gráfico III.1.31).



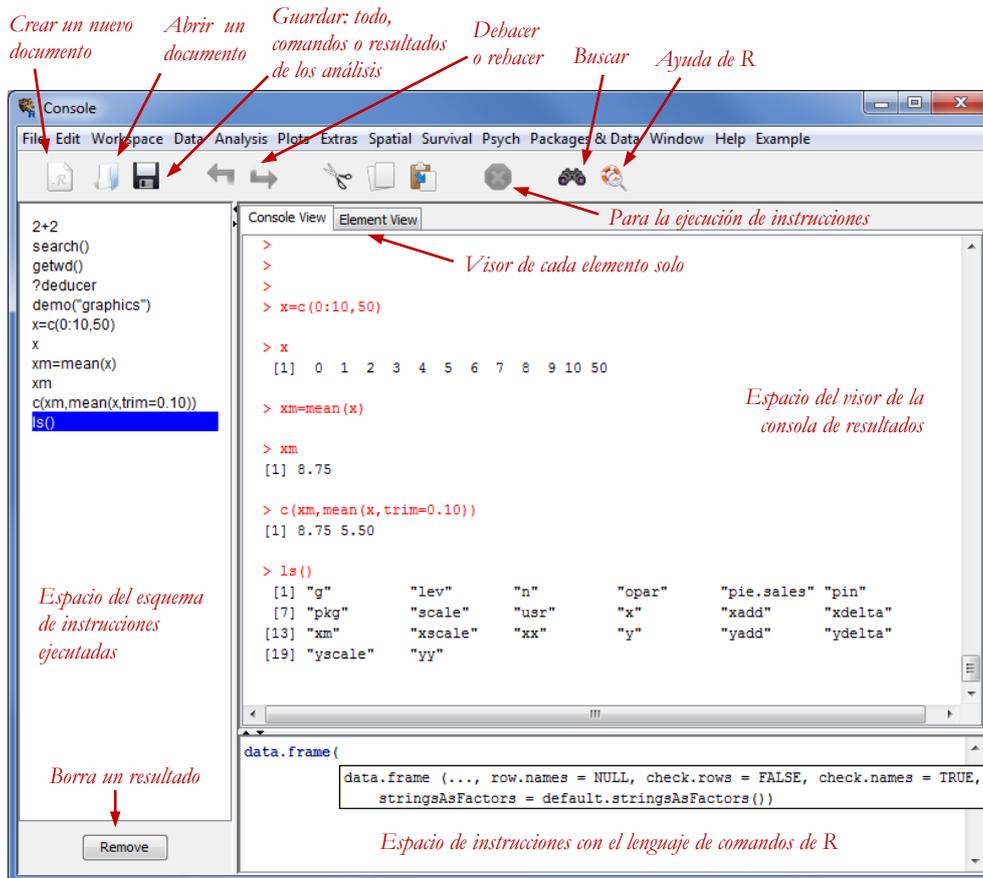
### 2.2.2. La consola de Deducer

La consola contiene una barra de menús a través de la cual elegiremos el procedimiento o tarea a realizar. Cada tarea se traduce en instrucciones que se expresan en el lenguaje de comandos y que van apareciendo en el espacio central de la consola, por lo que a través de Deducer cada procedimiento que utilizamos lo podemos seguir con su correspondiente versión en lenguaje de programación.

<sup>24</sup> Según se indica en la página se podría ejecutar desde la consola la instrucción:

> `install.packages("DeducerRichOutput", repos="http://R-Forge.R-project.org")` pero no nos funciona.

Gráfico III.1.27 La consola de Deducer



En este espacio aparecerán los resultados de nuestros análisis, si bien los gráficos se abren en ventanas independientes y otros procedimientos como la ayuda también se visualizan en una ventana propia. En la parte inferior, además, tenemos la opción de introducir directamente instrucciones con comandos de R y ejecutarlas. A la izquierda disponemos de un espacio del esquema de instrucciones ejecutadas que permite navegar a lo largo de los resultados, pudiendo visualizarlos uno a uno de forma independiente en la pestaña *Element View*. Por tanto disponemos de un entorno que nos permite simultáneamente ejecutar instrucciones a través de los menús y del lenguaje de programación. Progresivamente iremos comentando con más detalle los contenidos del menú y de la barra de iconos.

A continuación se describen las características generales de cada menú. En particular cabe destacar como principales los correspondientes a: **Data**, **Analysis**, **Plots** y **Extras** que aparecen desplegados como imágenes en el Gráfico III.1.28. **Error! Reference source not found.**

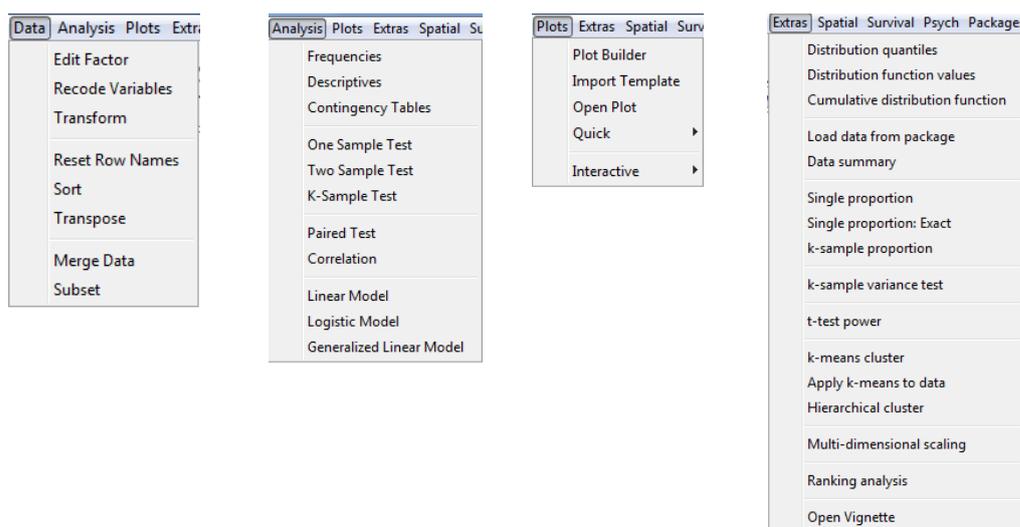
**FILE:** Menú destinado a crear, abrir y guardar los archivos, y a cambiar de directorio de trabajo.

**EDIT:** Menú de edición habitual del entorno *Windows* que incluye también la configuración de las opciones generales de R y Deducer.

**WORKSPACE:** Menú que abre, guarda o limpia el espacio de trabajo.

- DATA*: Menú que recoge un conjunto de comandos de transformación de los datos de un archivo y de manipulación del archivo de datos (selección, ordenación o unión de ficheros).
- ANALYSIS*: Menú de los procedimientos básicos de análisis estadístico con R.
- PLOTS*: Menú para obtener representaciones gráficas de los datos.
- EXTRAS*: Menú que amplía los procedimientos básicos de análisis estadístico.
- SPATIAL*: Menú adicional para el análisis de datos espaciales.
- SURVIVAL*: Menú adicional para el análisis de datos de supervivencia.
- PSYCH*: Menú adicional para el análisis de datos (fatorial, fiabilidad,...).
- PACKAGE & DATA*: Menú de utilidades para visualizar los objetos, abrir el editor de datos, instalar y gestionar los paquetes de R.
- WINDOW*: Menú para manejar las ventanas.
- HELP*: Menú de ayuda de R y Deducer.
- EXAMPLE*: Menú de ejemplo.

Gráfico III.1.28 Menús Data, Analysis, Plots y Extras de Deducer



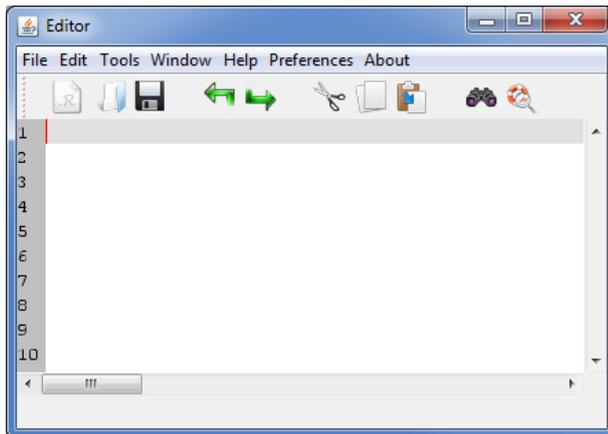
Los diferentes procedimientos o comandos del R y Deducer están destinados a la realización de 4 tipos principales de tareas:

1. La **generación y transformación de archivos de datos**: identificación de los datos, importación y exportación de archivos, su combinación, así como comandos de información de los mismos.
2. La **transformación de los datos**: para la modificación y creación de nuevas variables que preparan los datos para el análisis.
3. El **análisis estadístico de los datos**: procedimientos de análisis y de representación gráfica.
4. Un conjunto de **comandos adicionales** de operación para la configuración, el control y la obtención de información del entorno de trabajo de R, o para ser utilizados en la programación con el lenguaje de comandos.

Veámoslos brevemente algunos aspectos de los diferentes menús.

- Menú *Files*. Nos facilita en primer lugar crear, abrir y guardar archivos de datos desde la consola. Por otro lado si queremos trabajar con documentos de programas de instrucciones (*scripts*) podemos crear un nuevo documento  o abrir uno existente . En cualquier caso accederemos al editor:

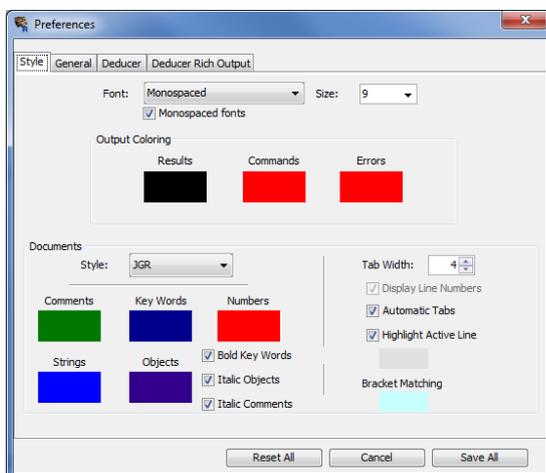
Gráfico III.1.29 Editor de documentos



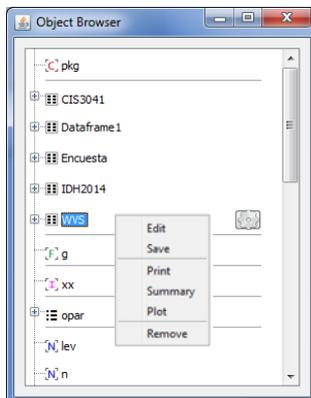
Con *Save* o con el icono  podemos guardar en un archivo externo toda la información de la consola (opción por defecto) o bien sólo los comandos o los resultados del análisis. *Set Working Directory* cambia la carpeta de trabajo del disco por defecto. Finalmente *Quit* cierra la aplicación.

- Menú *Edit*. Nos permite deshacer o rehacer  una acción, junto con las habituales tareas de cortar, copiar y pegar . Podemos copiar toda la salidas de la consola con *Copy Special* o bien sólo los comandos o sólo los resultados. El menú se completa con las opciones de borrar, seleccionarlo todo, buscar  y limpiar toda la consola de las salidas. Finalmente podemos aumentar o reducir el tamaño de la fuente y cambiar las preferencias que vienen por defecto :

Gráfico III.1.30 Preferencias de Deducer



- Menú *Workspace*. Permite guardar un área o espacio de trabajo, abrir uno existente o limpiar completamente el espacio actual.
- Menú *Package & Data*. Con el *Object Browser* podemos visualizar y editar los distintos objetos :



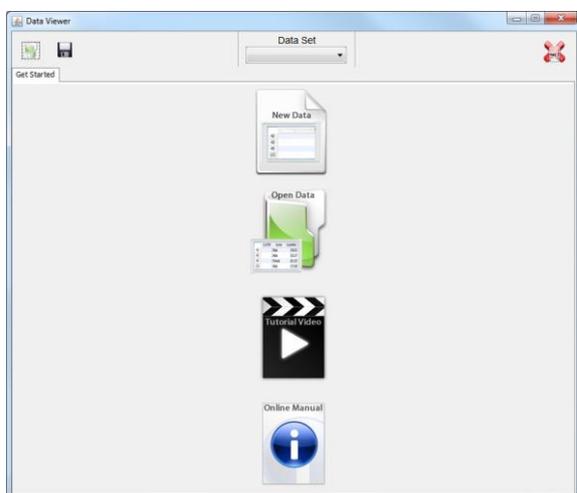
En este menú con *Data Viewer* simplemente accedemos al visor de datos que veremos a continuación mientras que *GUI Add-on*, *Package Manager* y *Package Installer* nos permiten gestionar los paquetes que se instalen en R y en Deducer como hemos visto anteriormente.

- Menú *Help*: la opción R Help  nos permite acceder a la ayuda general de R como vimos también en la consola original de R en el apartado 2.1.3 con el añadido de *Deducer Help* que es un enlace a la página web de inicio de Deducer donde aparece la ayuda.

### 2.2.3. El visor de datos

El visor de datos contiene cuatro vínculos que nos permiten acceder al editor de datos (a través de *New Data*), al disco para abrir un archivo de datos (*Open Data*) así como a la ayuda: del manual la página web de Deducer (*Online Manual*) así como tutoriales en video (*Tutorial Video*).

Gráfico III.1.31 Visor de datos de Deducer



Para empezar a realizar un análisis de datos necesitamos ante todo disponer de los mismos. Podemos obtener los datos de diferentes formas:

- Introduciendo la información directamente en el editor de datos, en este caso clicaremos sobre *New Data*.
- O bien abriendo una base o matriz de datos existente en formato de R (extensión **.rda** o **.rdata**), en este caso clicaremos sobre *Open Data*.
- O bien importándolos de hojas de cálculo (extensión de Excel **.xlsx** o con formato de separación por comas **.csv**), de matrices de datos en otros formatos de paquetes estadísticos (**.sav** de SPSS, **.xpt** de SAS, **.dta** de Stata...), datos en formato plano o texto (**.txt**), entre otros. La importación no requiere en Deducer un procesamiento especial, directamente clicando sobre *Open Data* podemos elegir el tipo de formato de archivo.
- Finalmente, también podemos incluso generar los datos a partir de comandos de R en un programa de instrucciones o *script*.

Los objetos más relevantes en el sistema de análisis R son los datos, que obedecen a estructuras diferentes. Los tipos de datos son:

- **Vectors**: los vectores son una serie de valores de una dimensión de tres tipos: numéricos, alfanuméricos (caracteres) y lógicos. Se definen con varias funciones o comandos de R, como **c** (concatenar), **seq** (secuencia) o con el operador **:**.
- **Factors**: los factores son vectores que contienen variables cualitativas o categóricas y los valores son caracteres (textuales).
- **Matrix**: las matrices son objetos de dos dimensiones, con filas y columnas, que contienen valores de un mismo tipo. Se definen con el comando **matrix** de R.
- **Data frame**: o marco de datos, es una generalización de las matrices donde cada columna puede contener tipos de datos diferentes. Se definen con el comando **data.frame** de R.
- **Arrays**: o arreglos, son objetos multidimensionales. Se definen con el comando **array** de R.
- **List**: la lista consiste en un conjunto de objetos, que pueden ser iguales o diferentes, y que pueden tener diferentes dimensiones y tipos de datos. Se definen con el comando **list** de R.

En general trabajaremos con *data frames* que denominaremos **matriz de datos** o **base de datos**.

Para **abrir una matriz de datos** en R creada previamente:

- A través del menú: **File / Open Data / Datos...**
- Clicando sobre el botón  desde la ventana del editor de datos.
- Con la teclas **<CTRL>+<L>**.

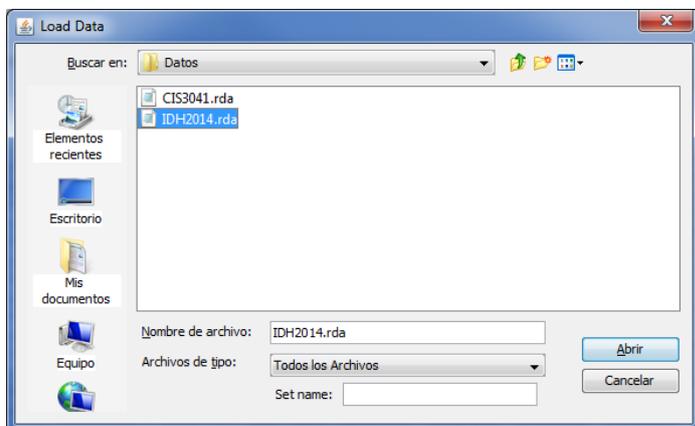
### ► Ejercicio 7. Abrir una matriz de datos en R

Desde el editor de datos abriremos dos de las matrices de datos que utilizamos en este manual: **IDH2014.rda** y **CIS3041.rda**<sup>25</sup>.

<sup>25</sup> Las matrices se pueden bajar desde la página web de este capítulo. Conviene recordar que una vez bajado el archivo si se intenta abrir desde una carpeta del disco cuyo nombre contenga caracteres acentuados la dirección no será identificada por el programa y no se podrá abrir. Tampoco el nombre del archivo deberá tenerlos.

Si abrimos la primera de las matrices accedemos a un cuadro como el siguiente donde, en particular, hemos seleccionado la matriz **IDH2014.rda**:

Gráfico III.1.32 Cuadro de diálogo *Load Data* de Deducer



Este archivo se alojará en la ventana del editor de datos (ver Gráfico III.1.33). Cuando existen varios archivos de datos abiertos a la vez se puede seleccionar desde el desplegable *Data Set*. El menú *Edit* por su parte nos permite copiar, cortar y pegar los datos, y sacar la matriz de datos del espacio de trabajo (*workspace*) , con lo que se cerraría la ventana de datos y desaparecería la línea correspondiente del *Data Set*.

Recordemos que una matriz de datos es un conjunto de filas y columnas a partir de la cual se organiza la información estadística. El formato más habitual de matriz de datos es aquél donde cada línea, llamada registro (*record*), constituye un caso (individuo o unidad) sobre el que se expresan los valores codificados correspondientes a las variables consideradas a las columnas (matriz  $X$  de  $n$  casos y  $p$  variables).

En este caso cada fila representa un país y cada columna un atributo o propiedad que caracteriza a cada uno de ellos. Los datos proviene de Naciones Unidas y corresponden a los indicadores que se utilizan para construir el **Índice de Desarrollo Humano** (IDH) correspondiente al año 2014<sup>26</sup>. Podemos comprobar que disponemos de información para 195 países en relación a 114 variables. La primera variable corresponde a la posición de cada país en el ranking que establece el índice y la segunda es el identificador del país. Las 112 variables restantes son los indicadores que considera el *United Nations Development Programme* para construir el indicador sintético IDH, todas ellas son variables medidas en una escala numérica o cuantitativa, identificadas como tipo *double* en R.

<sup>26</sup> Los datos han sido extraídos de la dirección: <http://hdr.undp.org/es/data>.

Gráfico III.1.33 Ventana del editor de datos de Deducir. Archivo IDH2014.rda

## a. Data View

	HDIrank	Country	HDI	Lifeexpectancy	Schooling	Expectedschooling	GNIpercapita2011	HDIV1980	HDIV1990	HDIV2000	HDIV2005	HC
1	1	Norway	0.944	81.50	12.63	17.6	63909.45	0.793	0.841	0.910	0.935	
2	2	Australia	0.933	82.50	12.80	19.9	41523.94	0.841	0.866	0.898	0.912	
3	3	Switzerland	0.917	82.60	12.24	15.7	53761.92	0.806	0.829	0.886	0.901	
4	4	Netherlands	0.915	81.04	11.89	17.9	42397.20	0.783	0.826	0.874	0.888	
5	5	United States	0.914	78.94	12.94	16.5	52308.38	0.825	0.858	0.883	0.897	
6	6	Germany	0.911	80.74	12.95	16.3	43048.68	0.739	0.782	0.854	0.887	
7	7	New Zealand	0.910	81.13	12.50	19.4	32569.37	0.793	0.821	0.873	0.894	
8	8	Canada	0.902	81.48	12.26	15.9	41886.82	0.809	0.848	0.867	0.892	
9	9	Singapore	0.901	82.32	10.20	15.4	72371.23	na	0.744	0.800	0.840	
10	10	Denmark	0.900	79.39	12.10	16.9	42880.28	0.781	0.806	0.859	0.891	
11	11	Ireland	0.899	80.71	11.61	18.6	33414.40	0.734	0.775	0.862	0.890	
12	12	Sweden	0.898	81.82	11.74	15.8	43201.35	0.776	0.807	0.889	0.887	
13	13	Iceland	0.895	82.09	10.41	18.7	35116.46	0.754	0.800	0.858	0.888	
14	14	United Kingdom	0.892	80.55	12.31	16.2	35001.63	0.735	0.768	0.863	0.888	

## b. Variable View

Variable	Type	Factor Levels
10	HDIV2000	Double
11	HDIV2005	Double
12	HDIV2008	Double
13	HDIV2010	Double
14	HDIV2011	Double
15	HDIV2012	Double
16	HDIV2013	Double
17	InequalityadjustedHDI	Double
18	CoefficientofhumanInequality	Double
19	InequalityLifeExpectancy	Double
20	InequalityAdjustedLifeExpectancyIndex	Double
21	InequalityEducation	Double
22	InequalityAdjustedEducationIndex	Double
23	InequalityIncome	Double

Una base o matriz de datos de estas características se identifica en R como un **data frame**, (mediante el símbolo **(df)** en *Data Set* se identifica el tipo de datos) traducido como **marco de datos**. Pero R también trabaja con el concepto de **matriz** en otro sentido y es preciso no confundirlo. Se denomina **matrix** a un tipo de objeto que también es un conjunto de filas y columnas, pero a diferencia del *data frame* en una *matrix* todos los elementos deben ser del mismo tipo. Por tanto, un *data frame* es una generalización de una *matrix* y puede contener, por ejemplo, columnas con variables de diferente tipo: numéricas, factor, carácter, etc. En *matrix* todas las variables deben ser del mismo tipo. Como hemos comentado en general trabajaremos con *data frames* que denominaremos **matriz de datos** o **base de datos**, si hacemos referencia en algún momento a una matriz de elementos iguales en el contexto de R emplearemos la expresión *matrix*.

La segunda matriz de datos, **CIS3041.rda**, corresponde a los datos del **Barómetro del Centro de Investigaciones Sociológicas** de octubre de 2014 (estudio número 3041)<sup>27</sup>.

<sup>27</sup> La información del barómetro (datos, cuestionario, resultados, la ficha técnica y los principales resultados que se obtienen de las tablas de frecuencias y de los cruces) se puede consultar en la dirección: [http://www.cis.es/cis/openem/ES/1\\_encuestas/estudios/ver.jsp?estudio=14119](http://www.cis.es/cis/openem/ES/1_encuestas/estudios/ver.jsp?estudio=14119). En la dirección

Se trata de una muestra de estadística de 2480 individuos que contestaron la encuesta del barómetro en relación a diversas cuestiones de la actualidad política, entre otras, y que dan lugar a un total de 210 variables. En esta matriz (ver Gráfico III.1.34) cada fila es un individuo anónimo que se identifica con un número correspondiente al número asignado a su cuestionario. En este caso la mayor parte de las variables son cualitativas, medidas en una escala nominal u ordinal.

Gráfico III.1.34 Ventana del editor de datos del SPSS. Archivo CIS3041.sav.

*a. Data View*

	ESTU	CUES	CCAA	PROV	MUN	TAMUNI	AREA	DISTR
1	17489	1	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
2	3041	2	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
3	3041	3	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
4	3041	4	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
5	3041	5	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
6	3041	6	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
7	3041	7	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
8	3041	8	Pais Vasco	Araba-alava	Burgos / Vitoria-Gasteiz ...	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
9	3041	9	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
10	3041	10	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
11	3041	11	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
12	3041	12	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
13	3041	13	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado
14	3041	14	Castilla La Mancha	Albacete	Albacete	100.001 a 400.000 habitantes	No pertenece a a.m.	Anonimizado

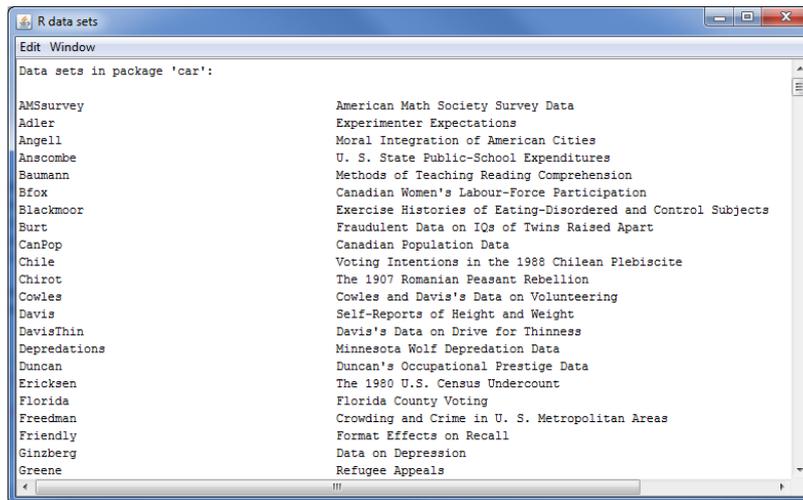
*b. Variable View*

	Variable	Type	Factor Levels
1	ESTU	Double	
2	CUES	Double	
3	CCAA	Factor	(1) Andalucía; (2) Aragón; (3) Asturias (Principado de); (4) Balears (Illes); (5) Canarias; (6) Ca...
4	PROV	Factor	(1) Araba-alava; (2) Albacete; (3) Alacant/Alicante; (4) Almeria; (5) avila; (6) Badajoz; (7) Illes...
5	MUN	Factor	(1) Mun. <100.000 hab.; (2) Ceuta / Melilla; (3) Albacete; (4) Algeciras; (5) Alcalá de Henares; ...
6	TAMUNI	Factor	(1) Menos o igual a 2.000 habitantes ; (2) 2.001 a 10.000 habitantes ; (3) 10.001 a 50.000 ha...
7	AREA	Factor	(1) No pertenece a a.m. ; (2) Pertenece a a.m. ;
8	DISTR	Factor	(1) Anonimizado;
9	SECCION	Factor	(1) Anonimizado;
10	ENTREV	Factor	(1) Anonimizado;
11	P0	Factor	(1) Española ; (2) Española y otra ; (3) Otra nacionalidad ;
12	POA	Factor	(1) Albanesa; (2) Austriaca; (3) Belga; (4) Bulgara; (5) Chipriota; (6) Danesa; (7) Española; (8)...
13	P1	Factor	(1) Muy buena; (2) Buena; (3) Regular; (4) Mala; (5) Muy mala;
14	P2	Factor	(1) Mejor ; (2) Igual ; (3) Peor ;

Los diferentes paquetes instalados de R disponen de archivos de datos que podemos utilizar en cualquier momento. Ejecutando el comando `data()` desde la consola obtenemos un extenso listado de los conjuntos de datos disponibles:

<http://pagines.uab.cat/plopez/content/bases-de-dades#cis> se puede consultar una pequeña nota explicativa de cómo trabajar con los archivos de datos que se ofrecen y en el capítulo siguiente de este manual se dedica un apartado específico para su tratamiento con SPSS.

```
> data()
```



Se puede abrir cualquier conjunto de datos ejecutando el mismo comando con el nombre del archivo de datos entre paréntesis. Por ejemplo en el paquete `effects` existe una matriz (*data frame*) con datos de la *World Values Survey*, si ejecutamos:

```
> data(WVS)
```

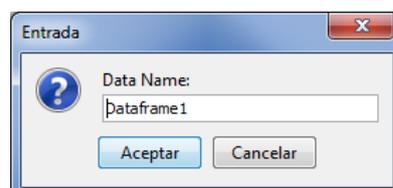
se cargará en el editor de datos, desplegamos el *Data Set* y veremos la matriz *WVS*.

A lo largo de este manual abriremos diversas matrices de datos que se proporcionan y también crearemos pequeñas matrices para realizar algunas tareas concretas. En el próximo capítulo trabajaremos con más profundidad el tema de la creación de una matriz a partir de los datos de una encuesta. No obstante veamos a continuación a través de un ejercicio de iniciación cómo introducir e identificar una matriz de datos.

### ► Ejercicio 8. Crear una matriz de datos en Deducer

En el editor de datos introduciremos **5 casos** correspondientes a personas imaginarias con la información de su **nombre**, **edad** y **sexo**.

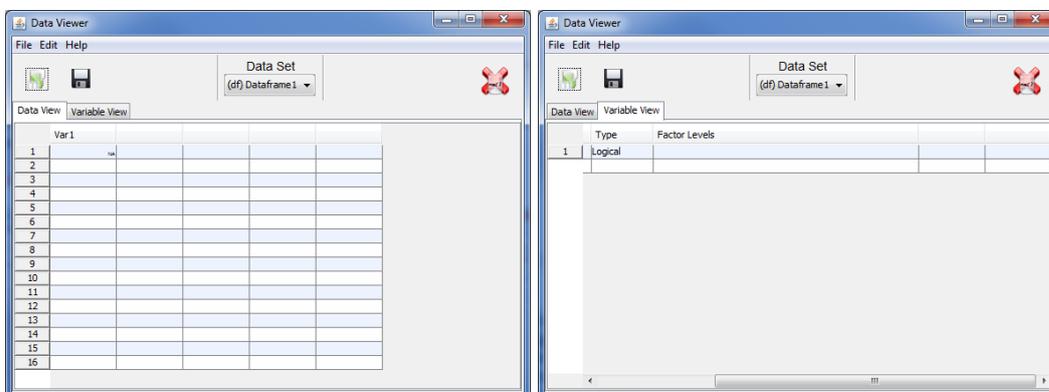
Si tenemos abierta una matriz de datos previamente en el editor y queremos crear una nueva procederemos en primer lugar a abrir una nueva ventana del **editor de datos en blanco**. A través del menú, haremos: **File / New Data**. Nos pedirá inmediatamente el nombre que le queremos dar a los nuevos datos, por ejemplo **Dataframe1**.



Los datos se introducen en la pestaña **Data View** y se identifican en términos de lo que llamaremos como el **diccionario de los datos** en la pestaña **Variable View**. Inicialmente

ambas ventanas tienen la apariencia del Gráfico III.1.35. A continuación introduciremos estos datos e identificaremos las variables.

Gráfico III.1.35 Estado inicial del visor de datos y de variables



Desde el visor de datos introduciremos los 5 casos de esta forma:

	Var1	V2	V3
1	Betina	30	Mujer
2	Mauri	16	Varón
3	Santi	9	Varón
4	Lidia	18	Mujer
5	Nico	15	Varón

Si una casilla de la matriz de datos no contiene información aparecen las letras **NA** (*Not Available*) indicando que se corresponden con los llamados **valores perdidos** o ausentes (*missing values*)<sup>28</sup>.

Cuando se introducen los datos el sistema asigna automáticamente un nombre a las variables (Var1, V2 y V3) y asigna un tipo de formato de variable. En la otra pestaña, en **Variable View**, disponemos de lo que podemos denominar como el diccionario de los datos, es decir, un conjunto de características que identifican a las variables de la matriz de datos: el nombre de la variable, el tipo de variable y, en el caso de una variable cualitativa o factor, los valores-etiquetas de cada categoría. Los datos introducidos, textuales (como el nombre y el sexo) o numéricos (como la edad) generan esta imagen en la pestaña de variables:

	Variable	Type	Factor Levels
1	Var1	Character	
2	V2	Double	
3	V3	Character	

Mejoraremos este diccionario de las variables iniciales que las identifica con los siguientes cambios: damos nombre a la variable **sexo** y cambiamos el tipo de variable

<sup>28</sup> Muchos procedimientos en el lenguaje de comandos de R necesitan que se indique su presencia para ejecutarse de forma correcta. Así por ejemplo la función o comando **na.omit** permitirá declarar la omisión de estos valores en los cálculos de un procedimiento determinado, o **is.na** permitirá localizarlos.

para considerarla como variable factor. Veremos cómo aparecen los valores o etiquetas de la variable:

	Variable	Type	Factor Levels
1	Nombre	Character	
2	Edad	Double	
3	Sexo	Factor	(1) Mujer; (2) Varón;

Una vez introducidos los datos es necesario guardarlos y convertirlos en un fichero del sistema R, por ejemplo con el nombre **Dataframe1.rda**. Para **guardar un archivo de datos** lo haremos a través del menú: **File / Save Data**, con las teclas <CTRL>+<S>

o clicando sobre el botón .

En el próximo capítulo veremos otra tarea importante y habitual del tratamiento de datos: la importación y exportación de matrices de datos, así como diversas tareas destinadas a la manipulación de una o varias matrices.

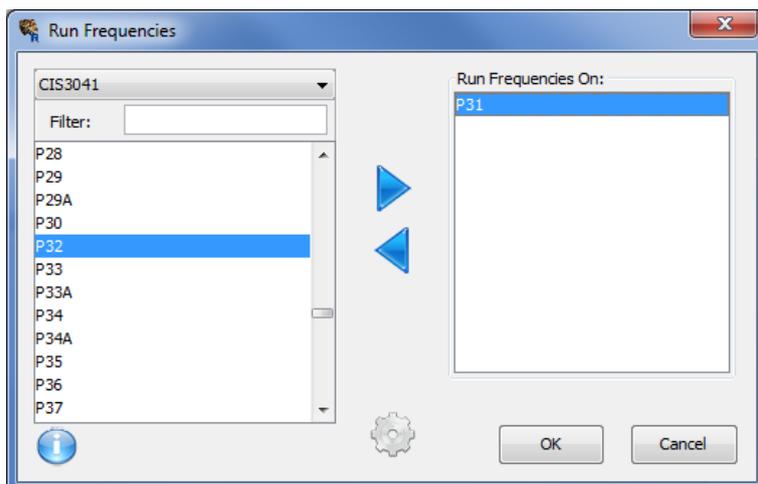
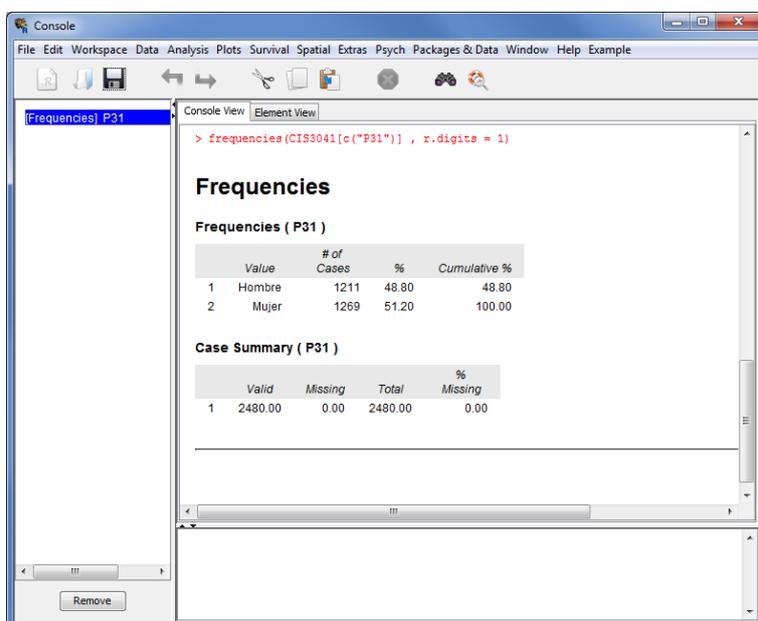
#### 2.2.4. Análisis de datos con la consola de Deducer

Con un software como R son muchos los procedimientos que podemos utilizar, tantos que intentar utilizarlos todos nos puede requerir más de una vida. Hemos comentado que podemos resumirlos en cuatro grandes tareas fundamentales: la creación e identificación, su manipulación y transformación, el análisis y otras tareas adicionales del entorno. A lo largo de este texto tendremos ocasión de detenernos en cada una de ellas. En este capítulo de presentación de R ya hemos visto sobre todo el primer tipo de tarea desde el punto de vista de Deducer. Daremos cuenta seguidamente, en una visión de presentación, de un ejercicio de análisis sencillo a través de los menús que nos ofrece la consola de Deducer.

Con los datos de la matriz **CIS3041.rda** solicitaremos una tabla de frecuencias que presenta la distribución del recuento del número de casos que tiene cada valor de la variable junto con los tantos por ciento. A través del menú **Analysis / Frequencies** accederemos al cuadro de diálogo del procedimiento estadístico descriptivo de una variable (Gráfico III.1.36). De la lista de variables de la izquierda buscaremos la variable **P31** (el sexo del entrevistado) y la colocaremos en el recuadro **Run Frequencies On:** clicando sobre  y a continuación sobre **OK**. Automáticamente aparecerá en la ventana de la consola el contenido del análisis (Gráfico III.1.37).

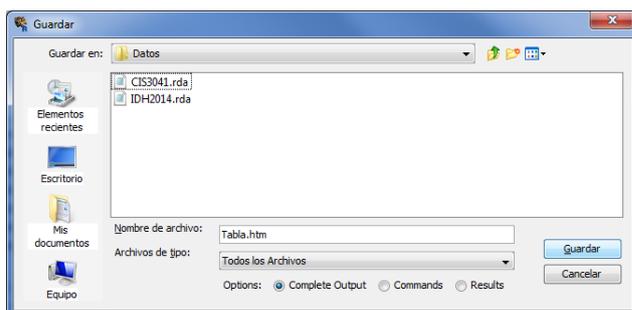
La salida presenta un formato enriquecido al convertirse en HTML pues utilizamos el paquete *DeducerRichOutput* que nos mejora notablemente la presentación de la tabla en relación al formato original. Para copiar la tabla e insertarla en un procesador de textos podemos capturar la pantalla para pegarla como imagen, si la copiamos simplemente y luego la pegamos no conservará esta apariencia pero se insertará como una tabla que luego podemos acabar de editar en un determinado formato. Pero la mejor opción es guardar los resultados como **html**<sup>29</sup> a través de menú **File / Save** (o <CTRL>+<S>) y elegir entre guardar toda la salida (comandos y resultados) o solo una parte.

<sup>29</sup> Ello es posible al trabajar con el *DeducerRichOutput*, sino aparecería en formato de texto plano.

Gráfico III.1.36 Cuadro de diálogo del procedimiento *Frequencies*Gráfico III.1.37 Resultados del procedimiento *Frequencies*

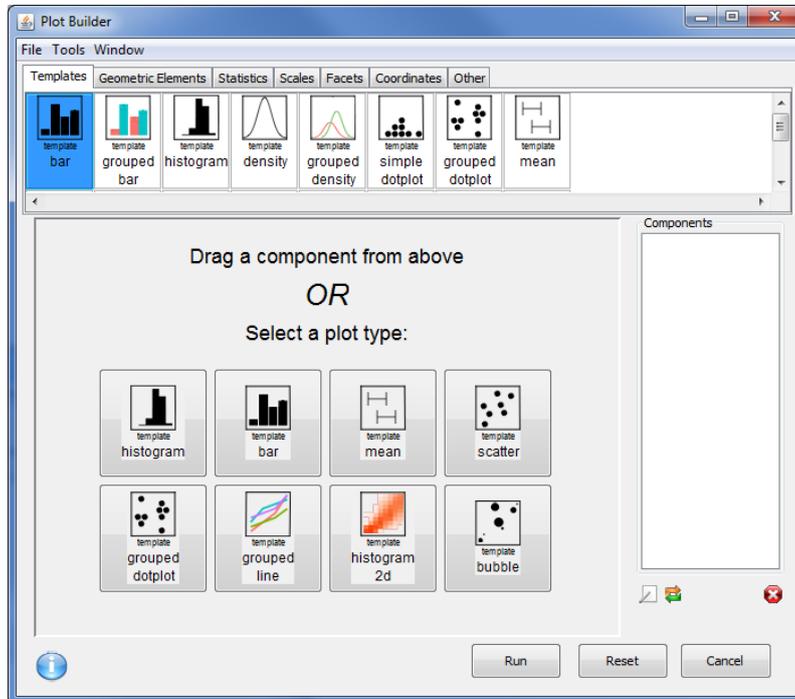
Una vez guardado con el documento con un nombre y la extensión **htm** (o **html**) se puede visualizar en nuestro navegador o también se puede abrir desde el procesador de textos que utilizemos.

Gráfico III.1.38 Cuadro de diálogo para guardar la salida de la consola



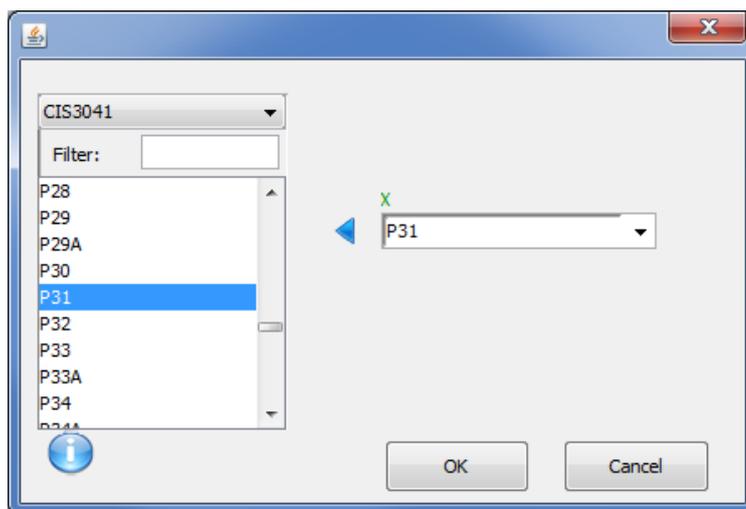
Para completar esta primera incursión en el análisis de datos con R representaremos gráficamente los datos de la variable **P31** mediante un diagrama de barras. Procederemos de la siguiente forma. Accederemos al menú a través de **Plot / Plot Builder** (Gráfico III.1.39) y elegiremos el icono  haciendo doble-clic:

Gráfico III.1.39 Ventana para elaborar gráficos *Plot Builder*



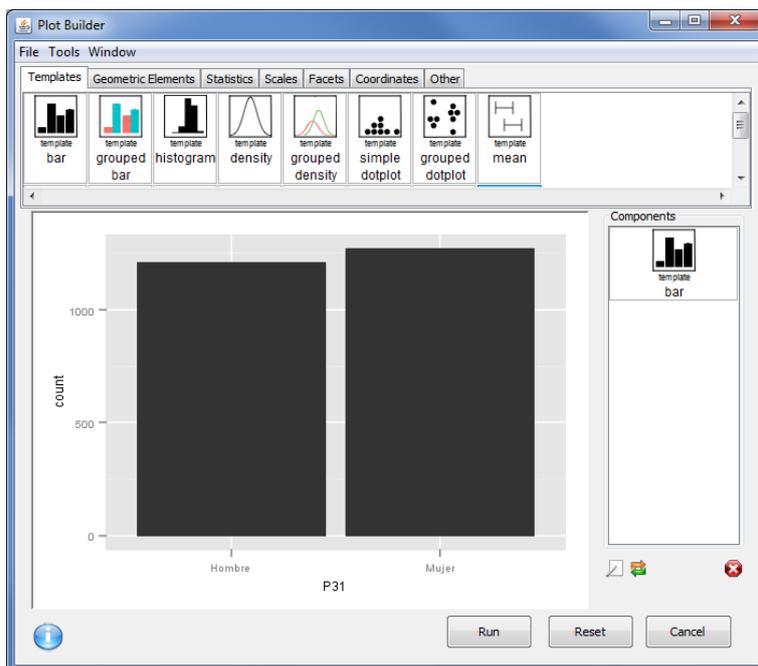
Accederemos al cuadro de diálogo y pasaremos la variable **P31** sobre el recuadro identificado con X y clicaremos sobre **OK**.

Gráfico III.1.40 Cuadro de diálogo del gráfico de barras (*bar*)



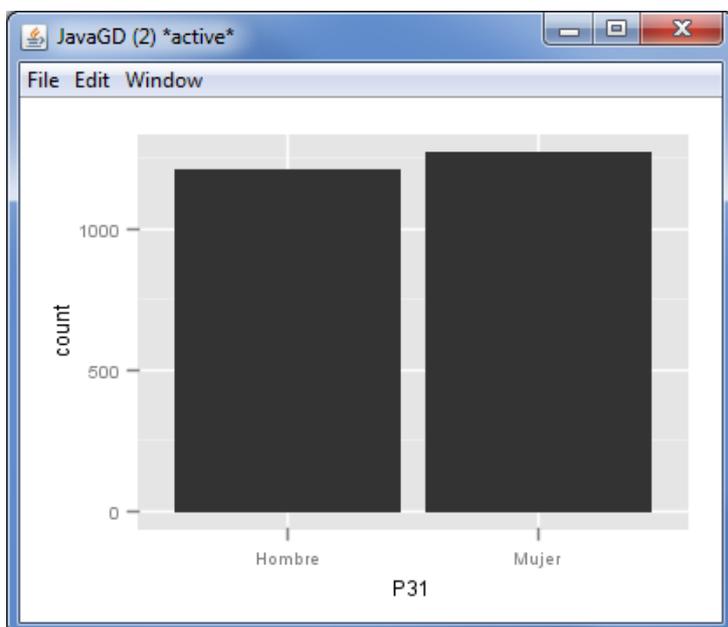
El gráfico que se obtiene aparecerá en la ventana del *Plot Builder*.

Gráfico III.1.41 Diagrama de barras con Deducer



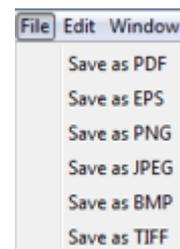
El gráfico se puede guardar a través del menú del editor gráfico: **File / Save**. Los gráficos se guardan con la extensión **.ggp** pues corresponden al paquete de gráficos de R **ggplot2**<sup>30</sup>. Si clicamos sobre el icono **Run** se ejecuta efectivamente el procedimiento y aparecerá el gráfico en una ventana propia:

Gráfico III.1.42 Ventana de resultados gráficos con Deducer



<sup>30</sup> Existen diversos paquetes y comandos que realizan gráficos, entre ellos el citado **ggplot2**, **plot**, **lattice** o **rggobi**.

Desde esta ventana también podemos copiarlo como imagen, a través del menú **Edit**, para pegarlo en un procesador de texto por ejemplo, o guardarlo en diversos formatos de imagen (JPG, PNG, BMP,...) en el menú **File**.

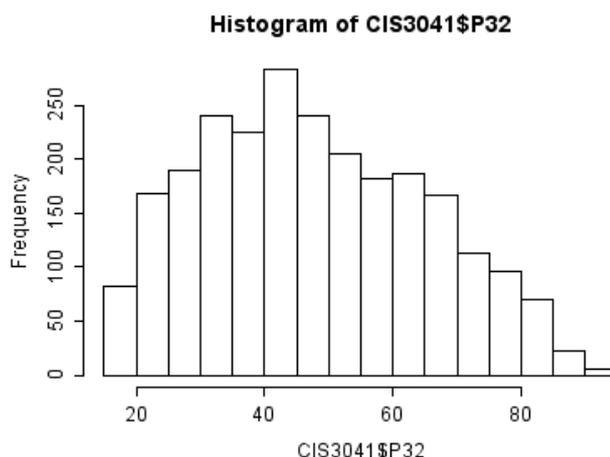


Una alternativa más rápida para llegar al mismo resultado de obtención del gráfico es ejecutar a través de **Plot / Quick / bar**.

Para el caso de una variable numérica este tipo de representación no es la adecuada sino el histograma. Podemos obtener esta representación de la variable **P32** que corresponde a la edad. En Deducer podemos obtenerlo a través de **Plot / Quick / histogram**. Alternativamente podemos obtenerlo de una forma rápida a través de línea de comandos de la consola con **hist**<sup>31</sup>:

```
> hist(CIS3041$P32)
```

Con especificaciones adicionales se puede configurar un resultado estéticamente distinto, en este caso es el siguiente:



A las tablas y los gráficos que generamos les podemos asignar un nombre de objeto pues constituyen también objetos de R.

Para finalizar, además de guardar la información de la consola (menú **File / Save**) podemos guardar el área de trabajo, el *workspace*, yendo al menú **Workspace / Save**, y en la próxima sesión recuperar el entorno de información con el que hemos estado trabajando.

### ► Ejercicio 9. Propuesto

Trabajando con la matriz de datos **CIS3041.rda** obtener las tablas de frecuencias de las variables: **P1**, **P1101**, **P15**, **P18**, **P23**, **P28**, **P32**, **P33A**, **P45** y **OCUMAR11**. ¿Qué nivel de medición tienen las diferentes variables? Obtener también una representación gráfica adecuada de las variables. Finalmente guardar los resultados en formato **html**, para a continuación editarlos y guardarlos en un documento de un procesador de textos incorporando los gráficos obtenidos.

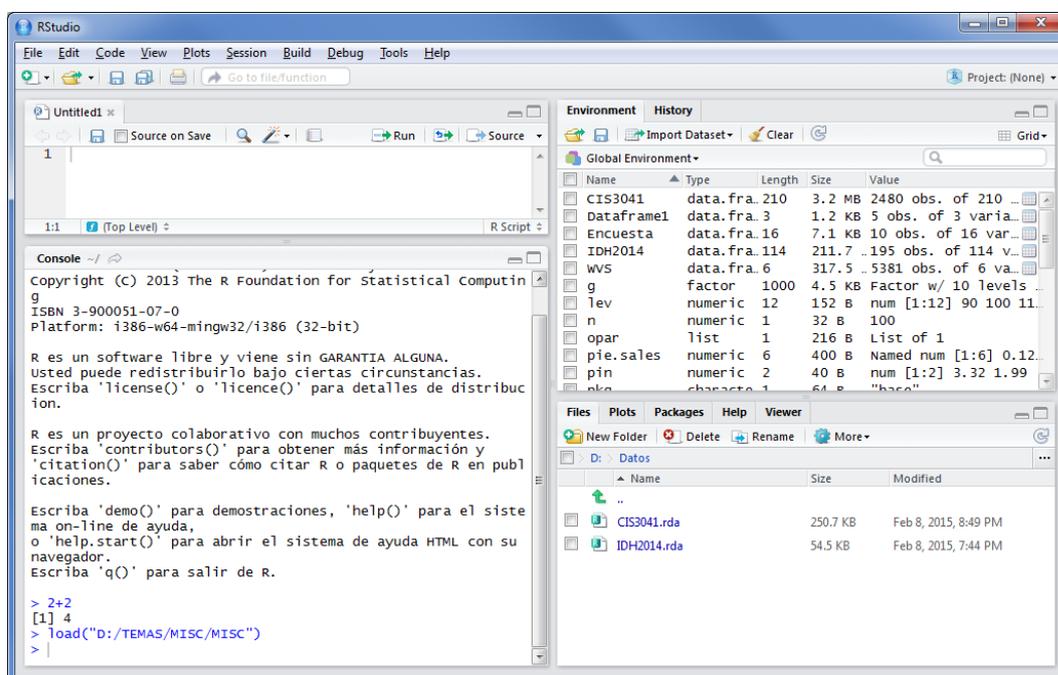
<sup>31</sup> En la instrucción, para indicar que se trata de la variable **P32** de la matriz de datos **CIS3041**, se utiliza esta expresión **CIS3041\$P32** con el operador **\$** que las vincula.

### 2.3. RStudio

que  **RStudio** es una consola para trabajar con el lenguaje de comandos de R que mejora notablemente las prestaciones de la consola original del programa. Se trata de un editor de textos IDE, *Integrated Development Environment*, que facilita la realización de las distintas tareas de programación en el lenguaje de R en un entorno gráfico integrado.

El software se descarga libremente desde su página principal: <http://www.rstudio.com/>. Elegiremos *Desktop* y después la plataforma sobre la que trabajamos, bajaremos la versión 0.98.1102 de febrero de 2015. Para *Windows* se corresponde con el fichero **RStudio-0.98.1102.exe**<sup>32</sup>. Una vez instalado y abierto el software nos aparece la ventana de la aplicación como en el Gráfico III.1.43.

Gráfico III.1.43 Entorno de trabajo de RStudio



El entorno de trabajo es una única ventana donde se disponen ordenadamente los distintos elementos que nos facilitan el trabajo de programación y de análisis con R. Este entorno incluye a su vez otras cuatro ventanas:

- 1) En la ventana superior izquierda se dispone el **editor de instrucciones** (*Source Editor*) aquí se abren y se editan archivos de programas de instrucciones con comandos de R (se denominan *scripts*). El editor permite también abrir otros tipos de archivos de texto plano o incluso bases de datos.
- 2) En la ventana inferior izquierda se localiza la **consola de R** (*Console*) donde se ejecutan los comandos de R.
- 3) En la ventana superior derecha disponemos de dos pestañas:
  - **Environment**, donde aparece la lista de los objetos disponibles en la memoria activa, en el espacio de trabajo (*workspace*).

<sup>32</sup> Desde la página principal podemos ir al apartado de soporte (<https://support.rstudio.com/hc/en-us>) donde consultar la documentación de ayuda, FAQ, etc.

- *History*, donde se registran todos los comandos ejecutados en una sesión con R.
- 4) La ventana inferior derecha dispone de cinco pestañas:
- *Files*: donde podemos acceder a las carpetas y archivos del disco duro.
  - *Plots*: donde aparecen los gráficos creados con R.
  - *Packages*: donde podemos administrar los paquetes instalados de R.
  - *Help*: donde accedemos a la ayuda de R.
  - *Viewer*: donde se visualiza el contenido web local.

Las ventanas se pueden cambiar de tamaño y ocultar según las necesidades del usuario/a.

### 2.3.1. Ejecución de programas de instrucciones con R-Studio

Para tener una primera aproximación a este entorno de trabajo se puede ejecutar el comando **demo** que vimos con anterioridad para la demostración de algunos procedimientos de los diversos paquetes instalados. Para ello utilizaremos la línea de comandos de la consola de R (ventana inferior izquierda).

En la línea de comandos de la consola podemos ir escribiendo instrucciones e ir ejecutándolas apretando *<Enter>*. Por ejemplo, ejecutemos:

```
> demo()
```

Como resultado vemos el listado de las distintas demos de los paquetes instalados. Podemos ejecutar en particular la demo **graphics** que muestra algunas de las capacidades gráficas del paquete *graphics* de R. Para introducir la nueva instrucción recuperaremos la que hemos escrito anteriormente tecleando la flecha hacia arriba. Cualquier instrucción introducida con anterioridad se puede recuperar simplemente moviéndonos con flecha hacia arriba *<↑>* y la flecha hacia abajo *<↓>*. Escribiremos a continuación el nombre de la demo dentro del paréntesis:

```
> demo(graphics)
```

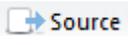
En la pestaña de *Plots* se puede navegar con la flechas   para visualizar los distintos gráficos obtenidos y en la pestaña *Environment* podemos ver los distintos objetos que se han generado.

Cuando se escriben instrucciones de R existe una función de autocompletado que nos ayuda en la elaboración de las instrucciones y que se completa efectivamente apretando la tecla del tabulador. Esta función está disponible igualmente en el editor de programas de instrucciones.

A continuación crearemos un programa de instrucciones de R, un *script*. En el editor de instrucciones podemos crear un archivo de instrucciones de R (a través de *File/New File/R Script* o clicando sobre  y eligiendo *R Script*), también podemos abrir uno existente (por el menú: *File/Open File* o clicando sobre el icono ) y lo podemos guardar (*File/Save* o clicando sobre   para guardar uno o todos los documentos abiertos).

Vamos a crear un programa de instrucciones para reproducir el análisis estadístico básico que vimos anteriormente con la matriz de datos `CIS3041.rda`. Primero abrimos la matriz, pedimos la tabla de frecuencias de la variable `P31` y el gráfico de barras. Las instrucciones que ejecutaremos son las siguientes:

```
#Ejemplo de programa de instrucciones de R
load("D:/Datos/CIS3041.rda")
sexo <- table(CIS3041[c("P31")]) #Objeto sexo
table(sexo)
prop.table(sexo)
barplot(sexo)
```

Introducimos todas las instrucciones en el editor primero<sup>33</sup> y procedemos a ejecutarlas. Clicando sobre  **Source** todas las instrucciones se envían a la consola para ejecutarse. Con  **Run** se ejecuta la línea sobre la que se sitúa el cursor y se posiciona inmediatamente en la línea siguiente del programa de instrucciones para permitirnos seguir ejecutando instrucciones<sup>34</sup>. Procederemos de esta segunda forma. Nos situamos en la primera línea y sucesivamente iremos clicando sobre *Run* para ejecutar cada instrucción y ver poco a poco los resultados que se generan.

La primera línea es un comentario simplemente, un texto libre que podemos incorporar en cualquier momento empezando la línea con el símbolo `#`. También se pueden incorporar al final de línea de una instrucción, como aparece en la tercera línea del programa. El comando `load` abre la matriz de datos `CIS3041.rda`. La siguiente instrucción asigna al objeto de nombre `sexo` la variable `P31` extraída de la matriz de datos. El comando `table` extrae el recuento de casos de la variable, `prop.table` las frecuencias relativas. Esta información aparece en la consola, ventana inferior izquierda. Finalmente `prop.table` obtiene el gráfico de barras que podemos ver en la ventana inferior derecha.

Cada documento nuevo de instrucciones que elaboremos desde el editor de instrucciones se abre en una pestaña. En un programa de instrucciones o *script* podemos ocultar secciones y bloques de comandos: **Edit / Folding**.

La ejecución de instrucciones se puede detener a través del icono  que aparece en un procesamiento largo de datos o presionando sobre la tecla `<Esc>`.

### 2.3.2. Comandos y programación en R

En este apartado recogeremos una selección de comandos de R que nos irán introduciendo en el conocimiento más profundo del lenguaje de programación y en el propio funcionamiento de R.

R es un lenguaje de programación llamado funcional, es decir, que el resultado que proporciona se corresponde con el objetivo deseado y que la implementación sea formalmente correcta (Peró et al., 2012: 20). Ya comentamos que como heredero del

<sup>33</sup> Se pueden copiar de este documento y pegar en el editor.

<sup>34</sup> Si se seleccionan diversas líneas y clicamos sobre *Run* se ejecutarían conjuntamente y el cursor se posicionaría en la línea siguiente de la última instrucción.

lenguaje S, el lenguaje R está orientado a objetos: todo la información que manejamos en un entorno de trabajo con el software es susceptible de identificarse con un objeto que le confiere una unidad que es generado o eliminado, consultado o transformado, unido a otro objeto, etc. Cada objeto tiene unas propiedades particulares que permite operar de forma distinta con él.

Consideremos el caso de un **vector**, que se define como un conjunto ordenado de valores numéricos, de tipo carácter o lógico. Por ejemplo un vector numérico (o tipo *double* como variable) y otro carácter (o tipo *factor* como variable) tienen naturalezas distintas, en el primer caso se pueden pedir cálculos estadísticos como una media y una desviación y en el segundo estos cálculos no tienen sentido y tan solo podemos solicitar del objeto el número de elementos y su tipo.

### ► Ejercicio 10. Creación de objetos vector de naturaleza distinta

Crearemos dos vectores de nombre **edad** y **sexo** y solicitaremos la ejecución del comando o función **summary** destinado a resumir la información de un objeto.

Podemos elaborar el siguiente programa de instrucciones y ejecutarlo:

```
# Creamos dos vectores con las edades de 5 personas y su sexo:
Edad=c(30,16,9,18,15)
Sexo=c("Mujer","Varón","Varón","Mujer","Varón")
# Pedimos el resumen de su contenido
summary(Edad)
summary(Sexo)
```

Su ejecución genera estos resultados:

```
> # Creamos dos vectores con las edades de 5 personas y su sexo:
> Edad=c(30,16,9,18,15)
> Sexo=c("Mujer","Varón","Varón","Mujer","Varón")
> # Pedimos el resumen de su contenido
> summary(Edad)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   9.0   15.0   16.0   17.6   18.0   30.0
> summary(Sexo)
  Length      Class      Mode
     5 character character
```

Un objeto que hemos ido trabajado en este capítulo es la matriz de datos o *dataframe* **CIS3041.rda** que tiene como nombre de objeto **CIS3041**. Si ejecutamos el comando **summary(CIS3041)** obtenemos un largo listado con el resumen de cada una de las variables de la matriz.

Un objeto entero puede ser transformado. Por ejemplo podríamos restar del vector **Edad** el valor de la media de las edades para determinar cuántos años más o menos tiene cada individuo respecto de la media. Del resultado anterior sabemos que la media es 17,6, el cálculo de la diferencia implica restar de cada dato del vector el valor de la media. La instrucción siguiente calcula la diferencia que comentamos:

```
> Edad-17.6
[1] 12.4 -1.6 -8.6  0.4 -2.6
```

Existen numerosos operadores y funciones en el lenguaje de comandos de R. En la Tabla III.1.1 adjunta se recoge un listado de algunos de ellos a efectos ilustrativos.

Tabla III.1.1. Operadores y funciones de R

Descripción	Comandos de R
<i>Operadores</i>	
<b>Aritméticos:</b> suma, resta, multiplicación, división, potencia	+ - * / ^
<b>Relacionales:</b> mayor que, mayor o igual que, menor que, menor o igual que, igual a, distinto de	> >= < <= == !=
<b>Lógicos:</b> y, o, no	&   !
<b>De asignación</b>	= <- ->
<b>De fórmulas:</b> relación entre variables	~
<b>De indexación:</b> subíndices y componentes de una base de datos	[ ] \$
<b>De secuencia</b>	:
<b>Aritméticos:</b> suma, resta, multiplicación, división, potencia	+ - * / ^
<i>Funciones</i>	
Concatenación de elementos	c
Creación de secuencias de números	seq ( <i>inicio, final, incremento</i> )
Repetir el valor $x$ según la especificación	rep ( $x, \dots$ )
Raíz cuadrada de un número $x$ real o complejo	sqrt ( $x$ )
Valor absoluto de un número $x$ real o complejo	abs ( $x$ )
Logaritmo en una base de un número $x$	log ( $x, base$ )
Exponencial en base $e$ de un número $x$	exp ( $x$ )
Media de $x$	mean ( $x$ )

Las funciones se consideran objetos también de R. La relación de funciones es enorme y depende de los paquetes que tengamos instalados. Con el tiempo y a medida que se tiene la necesidad de resolver un problema de análisis se va enriqueciendo la *expertise* del usuario/a.

### 3. El paquete estadístico SPAD

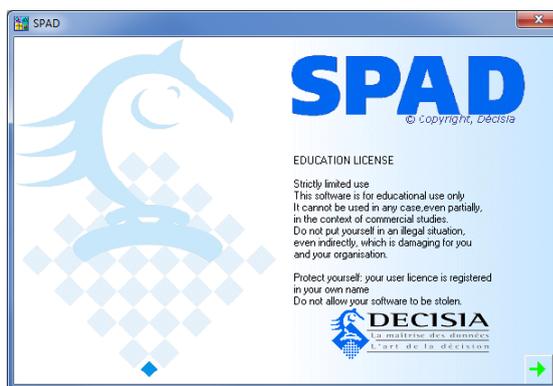
**SPAD** es un paquete estadístico creado originalmente a mediados de los años 70 en Francia. Una breve reseña de su historia se puede consultar en Morineau (2012). Sus creadores, Ludovic Lebart, Alain Morineau y Nicole Tabard, publicaron en 1970 el libro *Techniques de la Description Statistique* con un anexo que incluye el código **Fortran** con el que hacer “correr” las técnicas de análisis factorial presentadas en el texto (componentes principales y correspondencias). Se creó entonces una asociación sin ánimo de lucro (CESIA) para compartir y ampliar los programas que se habían elaborado. En 1979 Ludovic Lebart, Alain Morineau y Jean-Pierre Fénelon publican *Traitement des Données Statistiques* donde se amplían los procedimientos de análisis estadístico (modelos lineales, clasificación) y se elaboran nuevas versiones de los programas Fortran. A principios de los años 80 todos esos programas se integran en el software que Lebart y Morineau denominaron SPAD

(*Système Portable pour l'Analyse des Données*). La versión 1 de SPAD se publica en 1985. Millet (1987) reseña la referencia a la distinción entre SPAD-N (numérico) y SPAD-T (textual) de aquel momento y señalaba que las dos características más destacables eran la *potencia y especificidad de métodos en el tratamiento de información textual junto con la concordancia de los métodos multidimensionales reproducidos con trabajos de importante base teórica*.

Con el tiempo se profesionalizó y fueron mejorando sus prestaciones. Una sociedad comercial, CISIA (*Centre International de Statistique et d'Informatique Appliquées*), se constituye para asegurar su desarrollo y comercialización en 1987. Una referencia fundamental del momento del software es la publicación en 1995 del texto *Statistique Exploratoire Multidimensionnelle* por Ludovic Lebart, Alain Morineau y Marie Piron que se ha ido renovando desde entonces en sucesivas ediciones. En ese momento CISIA se fusiona con CERESTA (*Centre d'Enseignement et de Recherche en Statistiques*). En 1996 aparece la versión 3 para *Windows*. Ya en el nuevo siglo la empresa deviene DECISIA, en 2002, y en 2007 empezó a comercializarse con el grupo COHERIS, un año después se lanza la versión 7 y hasta la versión 8.2 actual (octubre de 2014), periodo donde se acentúa su carácter de software para el *data mining analytics*. En <http://www.coheris.com/produits/analytics/logiciel-data-mining/> se pueden consultar sus características.

En este manual no haremos un uso extenso del software como en el caso de SPSS y R. El interés principal se dirige a aplicar algunos procedimientos específicos de análisis multivariable de larga tradición en la llamada **escuela francesa del análisis de datos** donde en particular se utilizan de forma combinada el análisis de correspondencias junto al análisis de clasificación a partir de los trabajos seminales de Jean-Paul Benzécri. Esta articulación de procedimientos entronca con la perspectiva más actual del *data mining* (minería de datos) y el *data learning* (aprendizaje a partir de los datos). En el caso de su implementación en SPAD tiene un valor especial por la calidad de edición de sus gráficos factoriales y el procedimiento descriptivo de caracterización de variables que valoramos desde siempre.

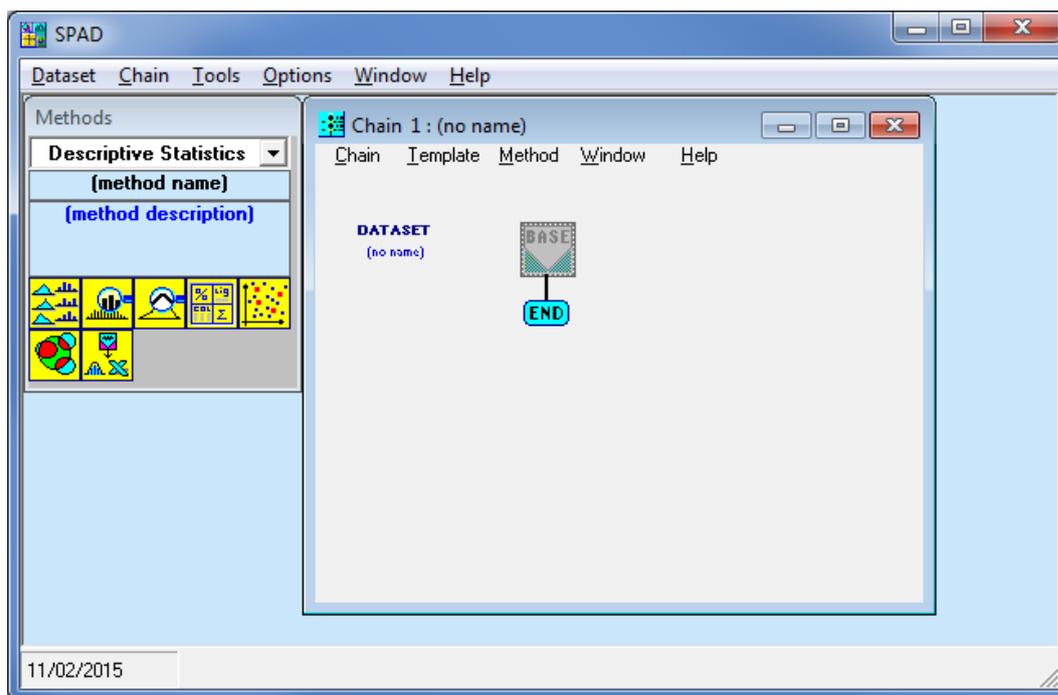
Además nos referiremos a la versión de **SPAD 5.6**, del año 2003, que en lo fundamental conserva las mismas características que los procedimientos de la versión más actual.



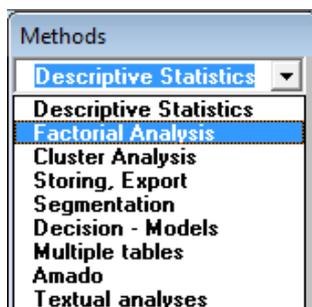
Seguidamente presentamos las características generales del software y del entorno de trabajo. Para complementar esta guía introductoria se pueden consultar las referencias de CISIA-CERESTA (2001a, 2001b) y Becue y Valls (2003). La referencia teórica de las técnicas estadísticas se encuentra en Lebart, Morineau y Piron (2006).

El entorno de trabajo de SPAD se presenta en el Gráfico III.1.44, es la imagen inicial que aparece cuando se abre la aplicación. La ventana principal incluye dos en su interior: la que corresponde a los diferentes procedimientos de tratamiento y análisis de los datos (**Methods**) y la de construcción de los programas de sintaxis (**Chain**), que en francés recibe el nombre de **filère** dando nombre a la extensión de los archivos de sintaxis que se guardan: **FIL**.

Gráfico III.1.44 Entorno de trabajo de SPAD 5.6



Los procedimientos o métodos en terminología de SPAD aparecen agrupados por temáticas más generales y cada método concreto tiene una simbología gráfica. Estos mismos métodos veremos que se pueden localizar en el menú **Method** o **Template** de la ventana de sintaxis para construir programas de sintaxis, en el segundo caso con combinaciones de métodos predefinidos.

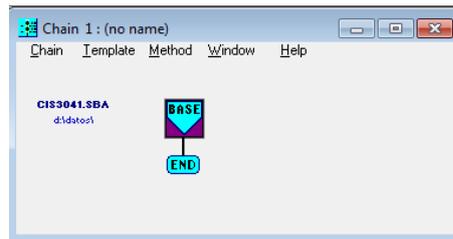


La ventana de sintaxis o **chain** es una representación gráfica de los procedimientos de análisis que se quieren encadenar. El primer paso siempre será especificar la base de datos (ícono **Base** ) y a continuación se insertarían los métodos de tratamiento y análisis de la base. Para seleccionar la base de datos hacemos doble clic sobre el icono,

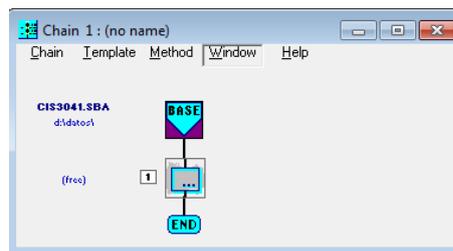
buscamos la matriz de datos en el disco y les damos a OK. Las bases de datos en SPAD tienen la extensión **SBA**.

Como ejercicio práctico para acompañar esta explicación abriremos la base de datos **CIS3041.sba** que se encuentra en la página web de este capítulo y que corresponde a los datos del **Barómetro del Centro de Investigaciones Sociológicas** de octubre de 2014 (estudio número 3041)<sup>35</sup>.

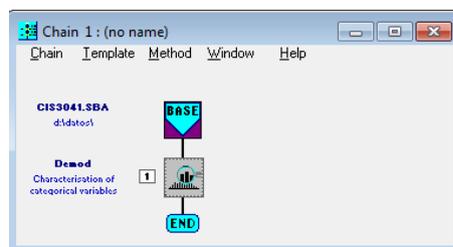
Cuando se abre la matriz de datos el icono **Base** queda parametrizado y pasa de gris a color.



Realizaremos un ejercicio de análisis para ilustrar el funcionamiento de la sintaxis. Necesitamos ante todo crear un espacio para insertar el método de análisis. Para ello clicaremos sobre el botón derecho del ratón en el icono **Base** y elegiremos **Insert method**. La nueva imagen es la siguiente.

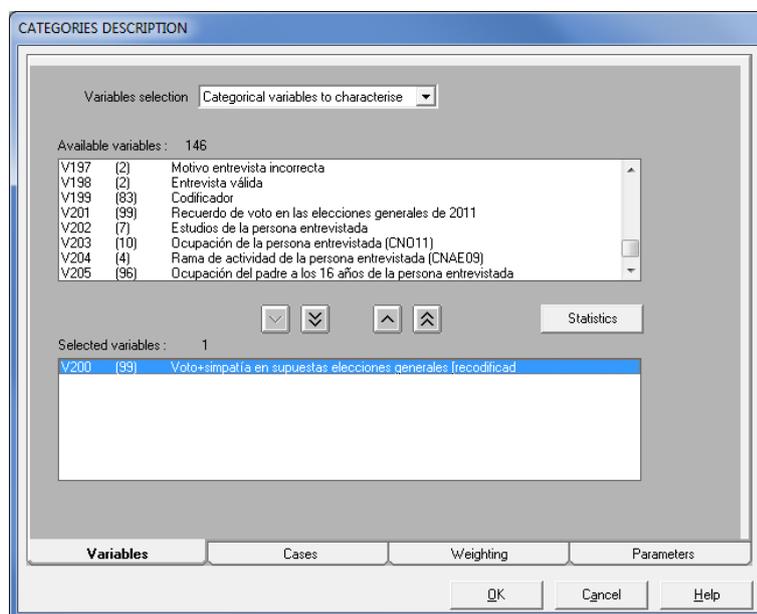


La propuesta es incluir un procedimiento de interés que posee SPAD: la caracterización de una variable, cuantitativa (método **DECOD**) o cualitativa (método **DEMODO**), a partir de un conjunto de otras variables que pueden ser también cuantitativas y cualitativas. En la ventana **Methods** elegiremos en el grupo de **Descriptive Statistics** y el método **DEMODO** : pinchamos sobre el icono, lo arrastramos hasta el icono gris que insertamos  y lo dejamos caer sobre él. La nueva apariencia del gráfico es la siguiente:



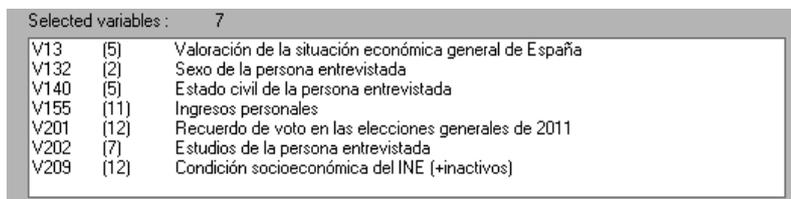
<sup>35</sup> La información del barómetro (datos, cuestionario, resultados,...) se puede consultar en la dirección: [http://www.cis.es/cis/openem/ES/1\\_encuestas/estudios/ver.jsp?estudio=14119](http://www.cis.es/cis/openem/ES/1_encuestas/estudios/ver.jsp?estudio=14119).

El siguiente paso es parametrizarlo, es decir, especificar el análisis que queremos realizar eligiendo las variables y configurando las opciones. Haciendo doble clic sobre el nuevo icono aparecerá el cuadro de diálogo del método:



En la pestaña de **Variables** elegimos en primer lugar la variable **V200** sobre la intención de voto (etiquetada como **Voto+simpatía en supuestas elecciones generales**) como variable para caracterizar (*Categorical variables to characterise*). Seguidamente en el desplegable **Variables Selection** elegiremos las variables que la caracterizarán, las cuales podrán ser de dos tipos y las ubicaremos en recuadros diferentes:

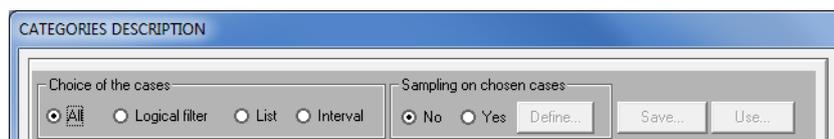
- las cualitativas (*characterising categorical variables*)



- y las cuantitativas (*characterising continuous variables*)<sup>36</sup>.

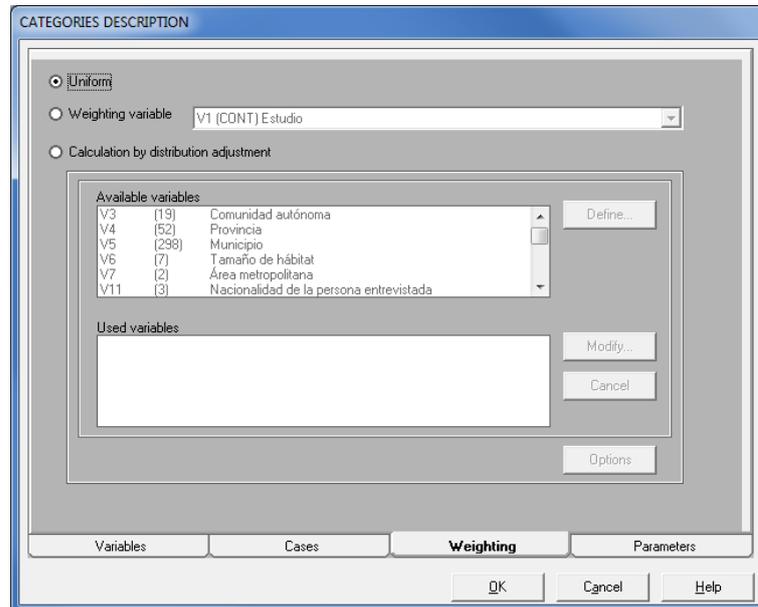


En la pestaña de **Cases** podríamos realizar una selección de casos según una condición de filtro, a partir del listado de casos o definiendo un intervalo de ellos, en este caso dejamos la opción por defecto:

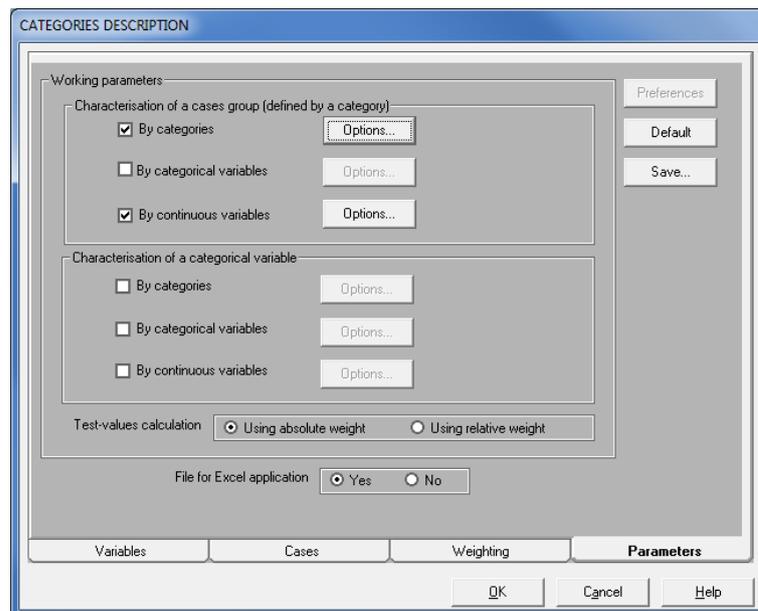


<sup>36</sup> El botón **Statistics** en la pestaña de **Variables** es muy útil pues nos permite acceder a una ventana emergente con la tabla de frecuencias de cada variable sin tener que consultar esta información fuera del cuadro de diálogo.

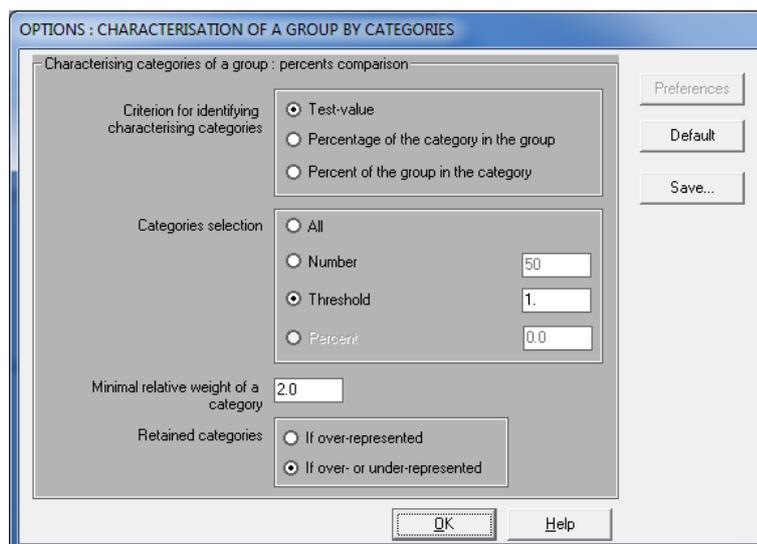
Trabajaremos con todos los casos y también sin la ponderación de los datos que nos ofrece la pestaña **Weighting**:



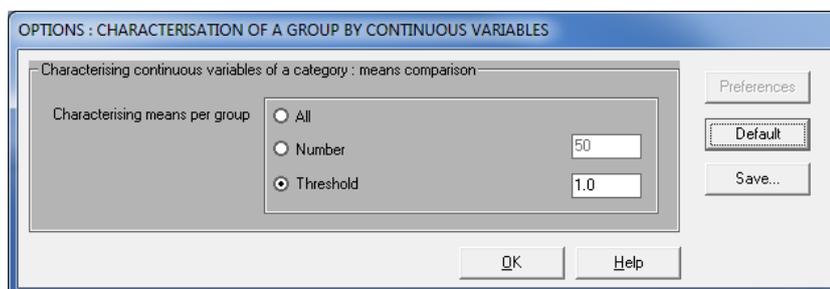
Finalmente completamos las especificaciones con la pestaña de **Parameters**:



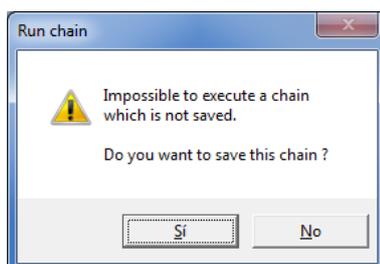
En este caso hemos optado por describir los individuos que pertenecen a cada categoría de la variable a caracterizar, en nuestro caso los votantes de cada partido. También se podría optar por la caracterización global de la variable. Dentro de la primera opción detallamos además el tratamiento que se hace de las variables cualitativas:



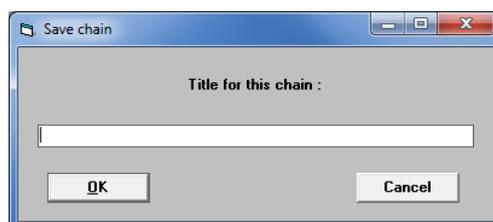
y cuantitativas:



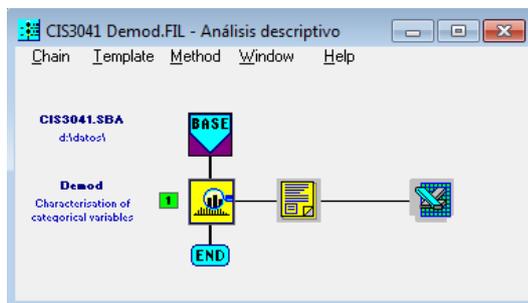
Dejando las opciones que aparecen por defecto y clicando en **OK** sucesivamente se acepta la ejecución del procedimiento, es decir, volvemos a la ventana **Chain** y vemos como el método parametrizado adquiere su color. El paso siguiente es ordenar la ejecución, de dos formas: con el botón derecho y clicando sobre **Run method** o bien con las teclas **<F5>**. Nos aparecerá el mensaje que indica la necesidad de guardar primero la **cadena** o archivo de sintaxis:



Después de clicar sobre **Sí**, y guardar el archivo con un nombre, por ejemplo **CIS3041-DEMODO.fl**, nos pedirá un título:



Podemos escribir por ejemplo Análisis descriptivo. Al clicar sobre OK se ejecuta y la ventana de sintaxis tiene esta imagen:



Donde vemos un documento de resultados (el LOG) y un documento en Excel<sup>37</sup>. En el primer caso accedemos a un editor de los resultados con un formato de texto plano que informa de los comandos que ejecutamos y de los resultados obtenidos en forma de tablas:

I.VALUE	PROB.	PERCENTAGES	CHARACTERISTIC	OF VARIABLES	IDEN WEIGHT
45.21	0.000	100.00	100.00	14.60 PP	Voto+simpatía en supuestas elecciones generales [recodificadV001 362
27.85	0.000	55.84	83.15	21.73 PP	Recuerdo de voto en las elecciones generales de 2011 RE01 539
9.15	0.000	30.83	33.98	16.09 Regular	Valoración de la situación económica general de España P103 399

Los resultados de tablas se exportan también en un formato de Excel aplicando una plantilla. Si clicamos doblemente sobre el icono se abrirá MS-Excel con la información siguiente:

<sup>37</sup> Cada procedimiento aparece numerado con un recuadro, cuando está blanco está pendiente de parametrización, en color verde ■ significa que el método es ejecutable y si se hace doble clic se pondrá en rojo que indica que el método está desactivado, es ejecutable pero el usuario bloquea la ejecución.

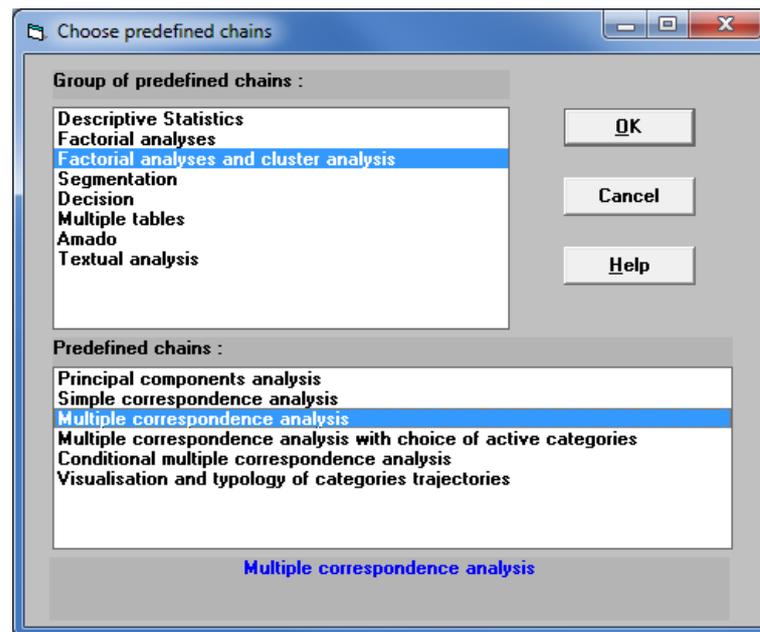
Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
Voto+simpatía en supuestas elecciones generales [recodificad	PP	100,00	14,60	100,00	45,21	0,000	362
Recuerdo de voto en las elecciones generales de 2011	PP	83,15	21,73	55,84	27,85	0,000	539
Valoración de la situación económica general de España	Regular	33,98	16,09	30,83	9,15	0,000	399
Condición socioeconómica del INE (+inactivos)	Jubilados/as y pensi	36,19	24,72	21,37	5,23	0,000	613
Condición socioeconómica del INE (+inactivos)	Trabajo doméstico no	12,98	7,30	25,97	4,09	0,000	181
Estado civil de la persona entrevistada	Casado/a	61,33	53,67	16,68	3,12	0,001	1331
Condición socioeconómica del INE (+inactivos)	Directores/as y prof	6,91	3,83	26,32	2,94	0,002	95
Estudios de la persona entrevistada	Primaria	25,69	20,28	18,49	2,64	0,004	503
Condición socioeconómica del INE (+inactivos)	Obreros/as no cualif	5,80	9,15	9,25	-2,39	0,008	227
Condición socioeconómica del INE (+inactivos)	Técnicos/as y cuadro	6,08	10,52	8,43	-3,05	0,001	261
Recuerdo de voto en las elecciones generales de 2011	No recuerda	1,38	4,44	4,55	-3,26	0,001	110
Recuerdo de voto en las elecciones generales de 2011	N.C.	2,76	6,81	5,92	-3,50	0,000	169
Voto+simpatía en supuestas elecciones generales [recodificad	ERC	0,00	2,22	0,00	-3,62	0,000	55
Recuerdo de voto en las elecciones generales de 2011	UPyD	0,28	2,94	1,37	-3,69	0,000	73
Voto+simpatía en supuestas elecciones generales [recodificad	CiU	0,00	2,34	0,00	-3,75	0,000	58
Estado civil de la persona entrevistada	Soltero/a	22,10	30,56	10,55	-3,81	0,000	758
Recuerdo de voto en las elecciones generales de 2011	GIU	0,00	2,46	0,00	-3,87	0,000	61
Voto+simpatía en supuestas elecciones generales [recodificad	UPyD	0,00	2,54	0,00	-3,94	0,000	63

Disponemos de tres pestañas: **Demod-5** con las tablas de caracterización de cada partido según las categorías de las variables cualitativas más características de cada uno y que define por tanto el perfil del votante de cada partido. Esta información se presenta en forma gráfica en la pestaña **Histogram**. Finalmente en **Demod-13** tenemos el perfil del votante a partir de las variables cuantitativas elegidas. No entraremos en la lectura de esta información ni en las características de la información presentada o de los criterios estadísticos que se aplican en el procedimiento. En el capítulo de análisis de clasificación daremos cuenta de ello. Pero el lector puede sacar rápidas e interesantes conclusiones teniendo presente que las líneas que aparecen en la columna **Variable label** corresponden, por orden de importancia, al perfil característico de cada **Group** o categoría de la variable de intención de voto, es decir, de cada partido. Después de una línea en blanco aparecen también los perfiles contrarios que más se oponen en relación a la categoría considerada. Esa ordenación se realiza de acuerdo con los valores test que aparecen en la columna titulada **Test-value**. El valor es tanto mayor cuanto más se aleja del comportamiento promedio de toda la muestra, por tanto, si una categoría no aparece es porque no difiere significativamente del promedio. Este criterio se aplica tanto a las variables categóricas como a las continuas:

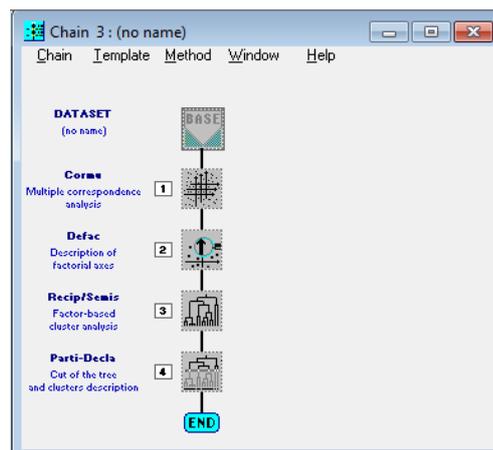
Characteristic variables	Category and mean	Overall mean	Category and Std. deviation	Overall Std. deviation	Test-value	Probability
Mariano Rajoy	6,358	2,305	2,192	2,788	29,57	0,000
Edad de la persona entrevistada	54,011	48,316	18,499	17,486	6,70	0,000
Escala de felicidad personal (0-10)	7,688	7,181	1,648	1,934	5,37	0,000
La política	4,331	3,906	3,043	3,289	2,62	0,004
Edad de la persona entrevistada	53,002	48,316	17,256	17,486	6,28	0,000
Mariano Rajoy	1,565	2,305	2,169	2,788	-6,15	0,000

Esta es la forma habitual de trabajo de cualquier procedimiento del software SPAD. Aparecerán resultados distintos según el método utilizado y veremos también que los resultados gráficos se podrán generar y visualizar con editores específicos.

En este caso hemos optado por construir la cadena de sintaxis nosotros mismos. Pero podemos optar por programas que combinan ciertos métodos que están predefinidos en el menú **Template**. Por ejemplo si queremos realizar un análisis de clasificación para construir una tipología con un análisis factorial de correspondencias múltiples previo tendríamos esta combinación de métodos que elegiremos desde el menú **Templates / Predefined chains**:

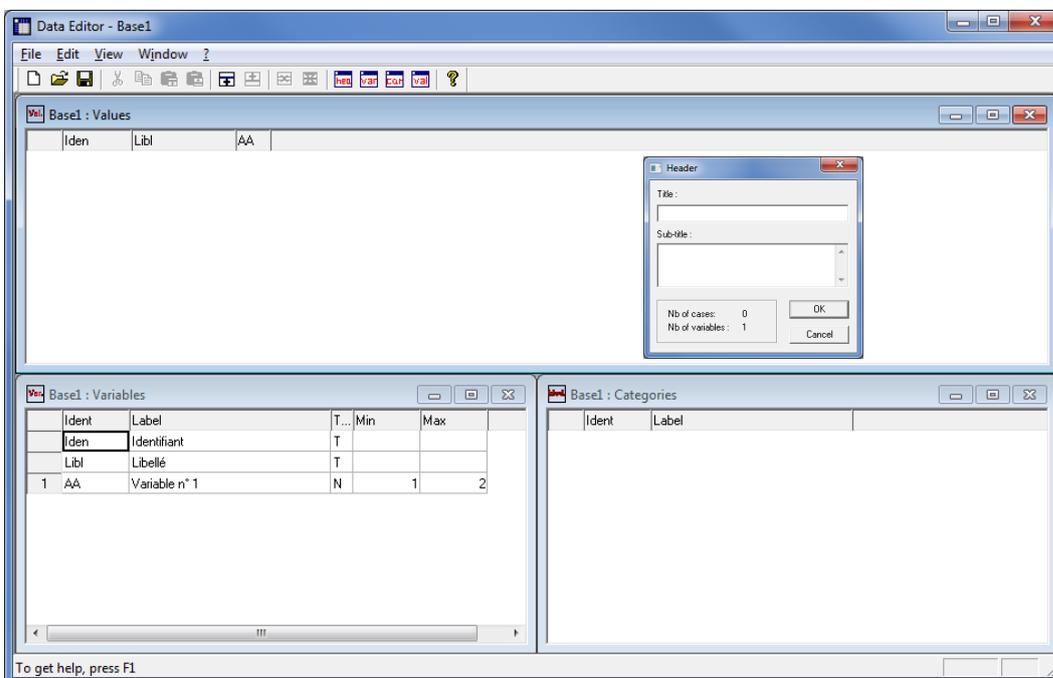


Clicando sobre **OK** se genera este gráfico de sintaxis que habrá de ser parametrizado:

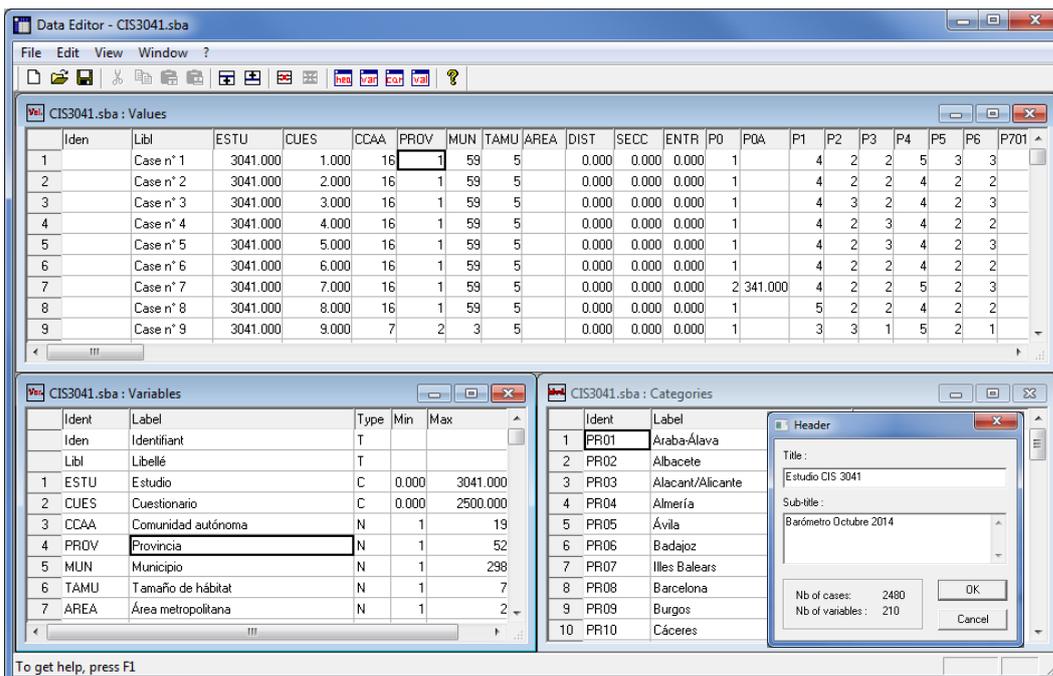


El otro aspecto relevante de todo software estadístico es el editor de datos, así como la importación y exportación de los mismos. En la ventana principal de SPAD no aparece el editor. Se abre a través del menú **Dataset** el cual nos permite también la importación y exportación. Nos centraremos en el funcionamiento del editor de datos y en la importación de ficheros de datos de SPSS.

Para crear una base de datos clicaríamos en el menú **Dataset / New dataset** mientras que para editar una existente los haríamos sobre **Dataset / Edit dataset**. En el primer caso accedemos directamente al editor que se abre en blanco para introducir los datos en la ventana de **Values**, para definir de las variables (nombre, etiqueta, tipo de variable, máximo y mínimo) en **Variables** y la definición de las etiquetas de los valores y de las variables cualitativas en **Categories**:



Existe una cuarta ventana, **Header**, que informa sobre el número de casos y de variables y permite titular la base. Si abrimos la base de datos **CIS3041.sba** a través de **File / Open** (o <CTRL>+<O> o clicando sobre ) veremos como se llena el editor:



En SPAD existen tres tipos de variables (**Type**): **T** para el formato textual, **N** para las variables cualitativas (nominales u ordinales) y **C** para las cuantitativas o continuas. De cada variable de tipo N y C es necesario definir su valor mínimo y máximo, teniendo en cuenta que las variables cualitativas no admiten el 0 como categoría ya que se reserva a los **valores perdidos**. Además es conveniente codificar las variables cualitativas con valores enteros consecutivos desde el 1, en caso contrario nos aparecerán también todos los valores intermedios en los resultados de algunos procedimientos.

Las variables se etiquetan con un máximo de 60 caracteres y también con un identificador interno de 4 caracteres que se utiliza en ciertos gráficos y métodos. Los individuos de la matriz también pueden ser etiquetados e identificados de la misma forma.

Desde SPAD se pueden importar fácilmente ficheros de SPSS y SAS, así como ficheros de texto o formato plano. En el caso de importar datos desde SPSS el procedimiento es tan sencillo como ir a abrir un archivo de datos y elegir la opción **SPSS file (\*.sav)** en **Tipo**. Una vez abierta se guarda como fichero de SPAD con la extensión **SBA**. No obstante deberemos tener en cuenta y respetar ciertos criterios como los que hemos citado. En particular:

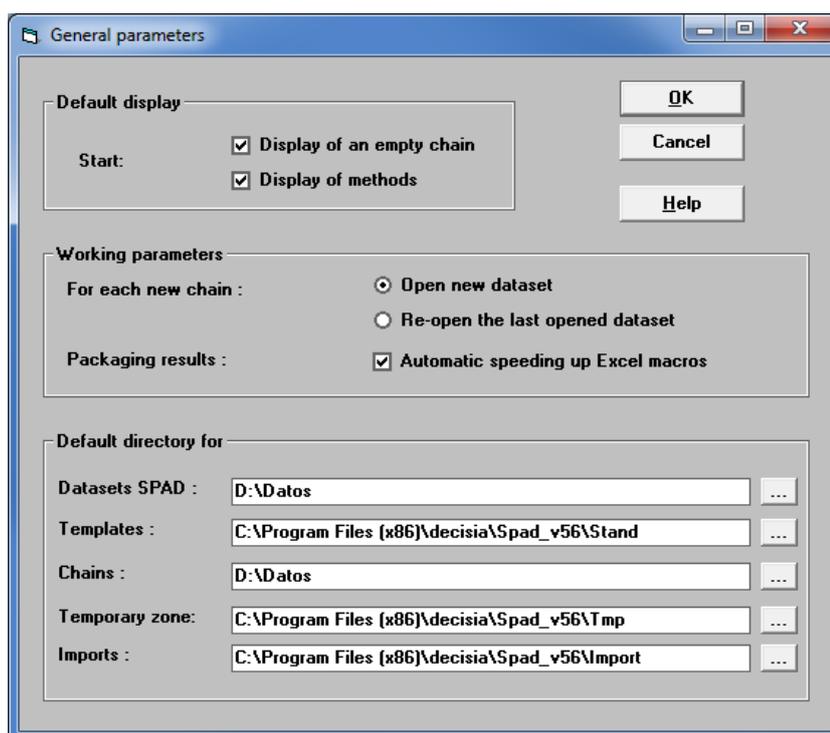
- Las variables cualitativas no pueden codificarse con 0 o valores negativos.
- Es conveniente codificar las variables cualitativas con valores enteros consecutivos.
- No se distingue entre perdidos del usuario y perdidos del sistema, todos se consideran perdidos del sistema (espacios en blanco).
- No se importan las variables cadena.
- Conviene verificar que el tipo de variable que automáticamente se asigna es el correcto.

Las etiquetas de variables y valores se conservan pero los nombres de las variables cambian todos por un nombre que empieza por V y añaden números enteros consecutivos desde el 1 hasta el número de variables de la matriz.

La matriz de datos **CIS3041.sba** es un caso de importación de datos desde SPSS donde se han tenido que realizar algunos ajustes como los reseñados para poder trabajar sin problemas en SPAD.

Para exportar desde SPAD hacia SPSS sencillamente se guarda la base de datos con la extensión **SAV**.

Para finalizar comentaremos un aspecto relevante de los parámetros generales que se configuran en el menú **Options** de la ventana principal. Entre ellos queremos insistir en la especificación de las carpetas de referencia que tiene por defecto el sistema. Conviene como regla general especificar un directorio en que trabajemos habitualmente, tanto para abrir las bases de datos como los archivos de sintaxis. El inconveniente con el que se encuentra el usuario es la obligación de navegar por el ordenador siempre hasta la carpeta donde está instalado el software. De hecho se podrían modificar las direcciones en función de la carpeta de trabajo que estemos utilizando en cada momento.



Relacionado con este comentario se puede sugerir otro aspecto característico del funcionamiento del software. A diferencia de otras aplicaciones, donde los archivos internos del sistema no son visibles, en el caso de SPAD aparecen numerosos ficheros en la carpeta de trabajo. En este sentido, sin que sea un problema necesariamente, se recomienda crear una carpeta para cada tipo de tarea y no mezclar ficheros de análisis y bases de datos diferentes.

En los capítulos dedicados al análisis factorial de correspondencias simples y múltiples y al análisis de clasificación veremos otros aspectos con mayor detenimiento, en particular la edición de gráficos factoriales que es una de las herramientas más interesantes de SPAD.

#### 4. Bibliografía

- Badiella, Ll. et al. (2015). *Manual de Introducció a Deducer: una interfaz gràfica para usuarios de R*. Bellaterra (Cerdanyola del Vallès). Servei d'Estadística Aplicada de la Universitat Autònoma de Barcelona. 5ª edición.  
<http://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/Manual%20curs%20Deducer.pdf>
- Bécue, M.; Valls, J. (2003). *Manual de introducció a los métodos factoriales y clasificación con SPAD*. Bellaterra (Cerdanyola del Vallès): Servei d'Estadística de la Universitat Autònoma de Barcelona.  
<http://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/manualSPAD.pdf>
- Bouso, J. (2013). *El paquete estadístico R*. Madrid: Centro de Investigaciones Sociológicas.

- Braña, T.; Rial, A.; Varela, J. (1998). *Diseños de investigación y proceso de datos con SPSS*. Madrid: Pirámide.
- Chapman, G. (2012). *Deducer Quick Start Guide*. Exploring Computer Science. National Science Foundation.  
<http://www.exploringcs.org/wp-content/uploads/2010/08/Deducer-Quick-Start-Guide.pdf>
- CISIA-CERESTA (2001a). *Introduction à SPAD Version 5.0. Manuel de Prise en Main*. Montreuil: CISIA-CERESTA.
- CISIA-CERESTA (2001b). *Système SPAD pour Windows Version 5.0 Base. Aide à l'interprétation*. Montreuil: CISIA-CERESTA.
- Díaz de Rada, V. (2002). *Técnicas de análisis de datos para investigadores sociales. Aplicaciones prácticas con SPSS para Windows*. Madrid: RA-MA.
- Elosúa, P.; Etxebarria, J. (2012). *R Commander. Gestión y análisis de datos*. (2012). Madrid: La Muralla.
- Etxeberria, J.; Garcia, E.; Gil, J.; Rodriguez, G. (1995). *Análisis de datos y textos*. Madrid: Ra-ma.
- Everitt, B. S.; Hothorn, T. (2014). *A Handbook of Statistical Analyses Using R*. Boca Raton, FL:Chapman & Hall/CRC. 3a. edición.
- Fellows, I. (2012). Deducer: A Data Analysis GUI for R. *Journal of Statistical Software*, 49 (8): 1-15.  
<http://www.jstatsoft.org/v49/i08/paper>
- Fellows, I. (2013). *Deducer. Reference Manual*.  
<http://cran.r-project.org/web/packages/Deducer/Deducer.pdf>  
<http://cran.r-project.org/web/packages/Deducer/>
- Filgueira, E. (2001). *Análisis de datos con SPSSWIN*. Madrid: Alianza.
- Fox, J. (2005). The R Commander: A Basic Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14 (9): 1-42.  
<http://CRAN.R-project.org/package=Rcmdr>
- Fox, J. (2009). Aspects of the social organization and trajectory of the R project. *The R Journal*, 1(2): 5-13.  
[http://journal.r-project.org/archive/2009-2/RJournal\\_2009-2\\_Fox.pdf](http://journal.r-project.org/archive/2009-2/RJournal_2009-2_Fox.pdf)
- Guisande González, C.; Vaamonde Liste, A.; Barreiro Felpeto, A. (2011). *Tratamiento de datos con R, STATISTICA y SPSS*. Madrid: Díaz Santos.
- Hornik, K. (2015). *R FAQ. Frequently Asked Questions on R. Version 3.1, 4-1-2015*.  
<http://cran.r-project.org/doc/FAQ/R-FAQ.html>
- IBM Corporation (2013). *IBM SPSS Statistics 22 Command Syntax Reference*.  
[ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/en/client/Manuals/IBM\\_SPSS\\_Statistics\\_Command\\_Syntax\\_Reference.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/en/client/Manuals/IBM_SPSS_Statistics_Command_Syntax_Reference.pdf)
- IBM Corporation (2015a). *IBM SPSS Statistics 22 Core System. Guía del usuario*.  
[ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM\\_SPSS\\_Statistics\\_Core\\_System\\_User\\_Guide.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM_SPSS_Statistics_Core_System_User_Guide.pdf)
- IBM Corporation (2015b). *IBM SPSS Statistics Base 22*.  
[ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM\\_SPSS\\_Statistics\\_Base.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM_SPSS_Statistics_Base.pdf)
- IBM Corporation (2015c). *Guía breve de IBM SPSS Statistics 22*.  
[ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM\\_SPSS\\_Statistics\\_Brief\\_Guide.pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/es/client/Manuals/IBM_SPSS_Statistics_Brief_Guide.pdf)

- Ihaka, R.; Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5: 299-314.
- Lebart, L.; Morineau, A.; Tabard, N. (1970). *Techniques de la Description Statistique*. Paris: Dunod.
- Lebart, L.; Morineau, A.; Fénélon, J.-P. (1995). *Traitement des Données Statistiques*. Paris: Dunod.
- Lebart, L.; Morineau, A.; Piron, M. (2006). *Statistique exploratoire multidimensionnelle. Visualisation et inférence en fouilles de données*. Paris: Dunod.
- Lizasoain, L.; Joaristi, L. (2003). *Gestión y análisis de datos con SPSS: versión 11*. Madrid: Paraninfo.
- López-Roldán, P. (2014). *Análisis de datos con SPSS*. En P. López-Roldán, *Recursos per a la investigació social*. Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona.  
<http://ddd.uab.cat/record/89349>
- Murillo Torrecilla, F. J.; Martínez-Garrido, C. (2012). *Análisis de datos cuantitativos con SPSS en investigación socioeducativa*. Madrid: Servicio de Publicaciones de la Universidad Autónoma de Madrid.
- Maurandi, A.; Río, L.; Balsalobre, C. (2013). *Fundamentos estadísticos para investigación. Introducción a R*. Bubok Publishing.  
[http://www.bubok.es/downloads/download\\_gratis?book=MjIzMjA3LTIwMTQwMjIzLQ==&tipo\\_portada=6](http://www.bubok.es/downloads/download_gratis?book=MjIzMjA3LTIwMTQwMjIzLQ==&tipo_portada=6)
- Millet, X. (1987). Crítica de programas de ordenador: El paquete estadístico SPAD. *Revista Estadística Española*, 116 (3): 133-139.
- Morineau, A. (2012). *Histoire du logiciel SPAD - bref historique*.  
<http://www.deenov.com/blog-deenov/histoire-du-logiciel-spad.aspx>
- Muenchen, R. A. (2011). *R for SAS and SPSS Users*. New York: Springer. 2ª edición.
- Paradis, E. (2003). *R para principiantes*. Traducción de J. A. Ahumada  
[http://cran.r-project.org/doc/contrib/rdebut\\_es.pdf](http://cran.r-project.org/doc/contrib/rdebut_es.pdf)
- Pardo, A. (2002). *SPSS 11 Guía para el análisis de datos*. Madrid: McGraw-Hill.
- Pardo, A.; Ruiz, M. A. (2005). *Análisis de datos con SPSS 13*. Madrid: McGraw-Hill.
- Pardo, A.; Ruiz, M. A. (2009). *Gestión de datos con SPSS Statistics*. Madrid: Síntesis.
- Pérez López, C. (2005). *Métodos estadísticos avanzados con SPSS*. Madrid: Thomson.
- Peró, M. et al. (2012). *Estadística aplicada a las ciencias sociales mediante R y R-Commander*. Madrid: Ibergarceta Publicaciones.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. The R Foundation for Statistical Computing Vienna, Austria. ISBN: 3-900051-07-0. <http://www.r-project.org/>.
- Rial, A.; Varela, J.; Rojas, A. J. (2001). *Depuración y análisis preliminares de datos en SPSS*. Madrid: RA-MA.
- Salas, C. (2008) ¿Por qué comprar un programa estadístico si existe R? *Ecología Austral*, 18, 223-231. <http://www.scielo.org.ar/pdf/ecoaus/v18n2/v18n2a07>
- SPSS Inc. (2001). *Guía para el análisis de datos*. Madrid: SPSS Hispanoportuguesa. (CD-Rom)
- Verzani, J. (2011). *Getting Started with RStudio. An Integrated Development Environment for R*. O'Reilly Media  
<http://it-ebooks.info/go.php?id=2253-1393279417-92f949c1f9c4deb4913f40010679fd3f>

- Visauta Vinacua, B. (2002). *Análisis estadístico con SPSS 11.0 para Windows. Volumen I. Estadística básica*. 2a. Edició. Madrid: McGraw-Hill.
- Wellman, B. (1998). Doing It Ourselves: The SPSS Manual as Sociology's Most Influential Recent Book. En D. Clawson. *Required Reading: Sociology's Most Influential Books*. Amherst: University of Massachusetts Press, 71–78.