# Cuadernos Metodológicos

45

# Análisis de datos con Stata

2.ª edición

Modesto Escobar Mercado Enrique Fernández Macías Fabrizio Bernardi



Consejo Editorial de la colección Cuadernos Metodológicos

Félix Requena Santos, Presidente del CIS

Franciso Alvira Martín, Universidad Computense de Madrid.

Consejeros

DIRECTOR

M.ª Ángeles Cea D'Ancona, Universidad Complutense de Madrid.

Jesús M. de Miguel Rodríguez, Universidad de Barcelona

Modesto Escobar Mercado, Universidad de Salamanca.

J. Sebastián Fernández Prados. Universidad de Almería

Juan Ignacio Martínez Pastor, Universidad Nacional de Educación a Distancia

SECRETARIA

M. a Paz Cristina Rodríguez Vela, Directora del Departamento de Publicaciones y Fomento de la Investigación, CIS

Todos los derechos reservados. Prohibida la reproducción total o parcial de esta obra por

Las normas editoriales y las instrucciones para los autores pueden consultarse en:

http://www.cis.es/publicaciones/CM/

cualquier procedimiento (ya sea gráfico, electrónico, óptico, químico, mecánico, fotocopia, etc.) y el almacenamiento o transmisión de sus contenidos en soportes magnéticos, sonoros,

visuales o de cualquier otro tipo sin permiso expreso del editor.

COLECCIÓN «CUADERNOS METODOLÓGICOS», NÚM. 45

Catálogo de Publicaciones de la Administración General del Estado http://publicacionesoficiales.boe.es

Segunda edición, mayo de 2012

Primera edición, diciembre de 2009

© CENTRO DE INVESTIGACIONES SOCIOLÓGICAS

Montalbán, 8, 28014 Madrid

© Modesto Escobar Mercado.

© Enrique Fernández Macías. © Fabrizio Bernardi.

DERECHOS RESERVADOS CONFORME A LA LEY

Impreso y hecho en España

Printed and made in Spain

NIPO: 004-12-004-0

ISBN: 978-84-7476-588-5 Depósito legal: M.

Fotocomposición e impresión:

Efca, S. A.

Parque Industrial "Las Monjas" Verano, 28

28850 Torrejón de Ardoz

### Índice

		Págs.
1.	INTRODUCCIÓN	9
2.	PRIMEROS PASOS CON STATA	17
	2.1. La información en los archivos de Stata	17
	2.2. La interfaz de Stata	20
	2.3. Las ventanas de Stata	25
	2.4. Modos de trabajo en Stata	33
	2.5. El fichero de resultados	39
	2.6. Las variables de la matriz de datos	43
	2.7. Ejercicios	56
3.	INTRODUCCIÓN DE DATOS	59
	3.1. Introducción manual de datos	59
	3.2. Lectura de datos con Stata	64
	3.3. Fusión de ficheros	78
	3.4. Ejercicios	82
4.	ESTADÍSTICAS DE UNA SOLA VARIABLE	85
	4.1. Clasificación de variables	85
	4.2. La tabla de distribución de frecuencias	87
	4.3. Estadísticos resúmenes de distribuciones	84
	4.4. Obtención de las medidas características de una	
	distribución	90
	4.5. La ponderación de los datos	93
	4.6. El error típico	99
	4.7. Ejercicios	108
5.	MANIPULACIÓN Y MODIFICACIÓN DE DATOS	111
	5.1. Manipulación de datos	111

			Págs.
	5.2.	Generación y modificación de variables	122
	5.3.	Características e instrucciones especiales	135
	5.4.	Ejercicios	141
6.	GRÁ	FICOS CON STATA	143
	6.1.	Características de los gráficos de Stata	144
	6.2.	Gráficos unidimensionales	147
	6.3.	Gráficos bidimensionales	163
	6.4.	Componentes de los gráficos	178
	6.5.	Esquemas	180
	6.6.	El editor de gráficos	184
	6.7.	Ejercicios	188
7.	LA P	PRUEBA ESTADÍSTICA Y LAS COMPARACIONES	189
	7.1.	Pruebas de una sola variable	191
	7.2.	Comparación de dos variables	198
	7.3.	Comparaciones de dos muestras (independientes)	207
	7.4.	Comparaciones de k muestras independientes	213
	7.5.	Comparaciones de k muestras dependientes	223
	7.6.	Ejercicios	230
8.	CON	FECCIÓN Y ANÁLISIS DE TABLAS CON STATA	231
	8.1.	Tablas de contingencia de dos variables	232
	8.2.	Más de dos variables	249
	8.3.	Otras tablas especiales	252
	8.4.	Las tablas de respuesta múltiple	259
	8.5.	Ejercicios	268
9.	LA R	REGRESIÓN	271
	9.1.	Nube de puntos, varianza y correlación entre dos variables	272
	9.2.	La regresión simple	277
	9.3.	Bondad del ajuste de la regresión	283
	9.4.	Inferencias en la regresión simple	287
	9.5.	Regresión múltiple	291
	9.6.	Regresión con variables ficticias	298
	9.7.	Regresiones con interacción	305
	9.8.	Otras relaciones funcionales de la regresión	312
	9.9.	Eiercicios	322

			Págs.
10.	DIAG	NÓSTICO DE LA REGRESIÓN	325
	10.1.	Supuestos de la regresión lineal	325
	10.2.	Análisis de los casos en la regresión	338
	10.3.	Regresiones especiales	347
	10.4.	Regresión robusta	355
	10.5.	Regresión de cuantiles	362
	10.6.	Regresión por bandas	366
	10.7.	Ejercicios	367
11.	LA RI	EGRESIÓN LOGÍSTICA	369
	11.1.	El modelo estadístico	369
	11.2.	Estimación del modelo	376
	11.3.	Diagnóstico del modelo	382
	11.4.	Comparación de modelos	394
	11.5.	Interpretación del modelo	400
	11.6.	Ejercicios	414
12.	REGI	RESIÓN LOGÍSTICA PARA VARIABLE ORDINAL	
	Y MU	LTINOMIAL	415
	12.1.	El modelo estadístico del logit ordinal	415
	12.2.	Estimación e interpretación del modelo	419
	12.3.	El supuesto de regresiones paralelas o razones	
		proporcionales	424
	12.4.	Regresión logística para variable dependiente	
		nominal	427
	12.5.	Estimación e interpretación del modelo	429
	12.6.	El supuesto de independencia de alternativas	
		irrelevantes	434
	12.7.	Ejercicios	436
13.	EL Al	NÁLISIS DE LA HISTORIA DE ACONTECIMIENTOS	
	CON	STATA	439
	13.1.	Qué es y cómo funciona el AHA	439
	13.2.	El AHA con Stata: instrucciones para definir los datos	445
	13.3.	La función de supervivencia	450
	13.4.	Modelos de la tasa de transición con tiempo continuo	452
	13.5.	Eiercicios	461

		Págs.
14.	ANÁLISIS DE DATOS DE ENCUESTA CON STATA	463
	14.1. Ajustes en el análisis de muestras complejas	464
	14.2. Ponderaciones, estratos y conglomerados	465
	14.3. Un ejemplo práctico con Stata. Las órdenes svy	470
	14.4. Ejercicios	480
15.	BIBLIOGRAFÍA COMENTADA	481
16.	ÍNDICE DE INSTRUCCIONES	489
17.	ÍNDICE DE CUADROS	493
18.	ÍNDICE DE ILUSTRACIONES	495
10	ÍNDICE DE CRÁFICOS	505

#### Primeros pasos con Stata

#### 2.1. La información en los archivos de Stata

Cualquier programa estadístico trabaja con información en muy diversos formatos almacenada en distintos tipos de fichero. Por ello, en este apartado se van a describir los principales archivos con los que trabaja Stata. Los seis tipos que se verán a continuación pueden dividirse en tres grandes grupos: los que guardan información sin procesar, los que conservan la información procesada y los que permiten o ayudan a transformar la información. En el primer grupo se incluyen los ficheros que contienen los datos individuales tal y como son introducidos en el ordenador después del trabajo de campo; en el segundo se consideran los archivos donde se guardan los estadísticos o gráficos que se generan con el análisis del programa y, finalmente, se consideran del tercer grupo los ficheros donde convenientemente se almacenan las instrucciones necesarias para realizar las tareas.

La base de trabajo es la matriz de datos, que consiste en una disposición ordenada de información, poco o nada procesada. Generalmente, el modo como esta se organiza para su tratamiento es de tal forma que los casos se encuentren expuestos en fila y las variables en columnas. Un ejemplo simple puede bastar para la comprensión de la estructura. Suponiendo que hubiera que analizar a dos personas, una mujer y un hombre de edades respectivas de 21 y 20 años, se pueden distinguir tres conceptos primordiales:

En primer lugar, el concepto de *caso*, esto es, cualquier unidad de la que se recoge información. En los datos anteriores existen dos, las dos personas de las que se saben sus características sociodemográficas. En segundo lugar, el concepto de *variable*, es decir, las características susceptibles de adquirir distintas modalidades. En el ejemplo presente, las dos variables disponibles son sexo y edad. Una de ellas es de naturaleza cualitativa, mientras la otra se presenta como cuantitativa. Cada una de las modalidades, cualitativas o cuantitativas, de estas variables recibe el nombre de *valor*. En este caso, son valores 20 y 21 años. También son valores "mujer" y "hombre", aunque por no ser de naturaleza numérica, también pueden denominarse *atributos*, *categorías* o, en conjunto, un *factor*.

#### ILUSTRACIÓN 2.1. Matriz literal de datos

Hombre 20 Mujer 21

Esta matriz de datos podría condensarse aún más si se representan los atributos con una serie de códigos. Así puede reducirse *Hombre*, poniendo a todos los casos con esta característica un símbolo que lo represente, que puede ser Ô, H o preferiblemente un dígito, para que la introducción de la información se pueda hacer del modo más rápido y, por costumbre, el 1 para las personas de género masculino y el 2 para las mujeres. De esta forma, la matriz de datos original presenta una estructura como la siguiente:

#### ILUSTRACIÓN 2.2. Matriz codificada de datos

1 20 2 21

Esta matriz o conjunto de datos, para que pueda ser tratada informáticamente más de una vez, ha de ser guardada en un *fichero de datos*. En principio, cualquier archivo que contenga información ordenada puede ser leído directa o indirectamente por Stata. Pero sólo pueden ser utilizados desde el interior del programa media docena de formatos: ASCII o Unicode, XML, ODBC, SAS XPORT y Excel. Para el resto de casos, existen otros programas que transforman los ficheros generados por aplicaciones como hojas de cálculo, bases de datos o incluso otros programas estadísticos en ficheros de trabajo aptos para Stata. Entre ellos, uno de los más conocidos es *Stat-Transfer*<sup>1</sup>, cuyo uso y utilidad se verá en la Sección 3.2.2.

Sin embargo, para el trabajo estadístico no basta con tener la matriz de datos bruta. Hay que añadirle al menos los nombres de las variables para que cuando se solicite una determinada tarea el programa sepa qué información se desea tratar. No es lo mismo solicitar una media del sexo, que de la edad. Se podría indicar que se desea sólo una media de la segunda variable; pero es mucho más cómodo solicitarla llamándola edad. Por ello, una de las operaciones imprescindibles en todo programa estadístico es la de convertir el fichero de datos brutos en otro con la matriz de datos ampliada con las definiciones y transformaciones de la información original que el usuario considere conveniente.

<sup>&</sup>lt;sup>1</sup> Stat-Transfer no es un producto de la casa *StataCorp.*, sino de *Circle System*, aunque fuera de Estados Unidos lo suelen comercializar las mismas empresas que venden Stata.

Estos específicos ficheros con información bruta, definiciones y transformaciones se denominan *ficheros de trabajo*. Sólo pueden construirse con el programa Stata o con otros pocos programas estadísticos que incluyen la posibilidad de guardar los datos en este formato. Generalmente se les reconoce por tener la extensión  $.dta^2$ . En ellos están almacenados los datos de las variables originales y de las creadas posteriormente por el usuario, junto con sus correspondientes nombres, etiquetas y formatos.

Para que puedan comprobarse los ejemplos de su manual, Stata permite acceder a todos los ficheros empleados. Se puede obtener una relación de los ficheros de datos incorporados en la instalación del programa mediante la instrucción *sysuse dir:* 

#### ILUSTRACIÓN 2.3. Directorio de los ficheros de datos en el sistema

	artificial.dta		network1.dta	sp500.dta	voter.dta
- 1	auto.dta	citytemp.dta	networkla.dta	surface.dta	xtline1.dta
-	autornd.dta	citytemp4.dta	nlsw88.dta	tsline1.dta	
1	bplong.dta	educ99gdp.dta	nlswidel.dta	tsline2.dta	
1	bpwide.dta	gnp96.dta	pop2000.dta	uslifeexp.dta	
- 1	cancer.dta	lifeexp.dta	sandstone.dta	uslifeexp2.dta	

Otros archivos de interés en el trabajo con Stata son los *ficheros de resultados* (con extensión *smcl* o *log*): siempre que así se le indique, los resultados de las órdenes dadas al programa son archivados en un fichero para que puedan quedar disponibles permanentemente, sin tener que volver a procesar de nuevo los datos mediante las instrucciones pertinentes. Stata dispone de un formato específico de grabación de los resultados en un fichero (*formatted log*) al que incorpora la extensión *smcl*, que consta de todos los elementos adicionales necesarios para una presentación idónea de las tablas estadísticas. Pero en ocasiones<sup>3</sup> es útil que los resultados se generen en un formato tratable universalmente, como es el caso de los ficheros en código ASCII. Por ello, también existe la posibilidad de grabar los resultados sin formato en archivos generados en Stata con la extensión *log*.

En cualquier caso, en ninguno de estos dos tipos de ficheros se incorporan los gráficos, pues cada uno de estos se guarda en un fichero independiente y específico para este tipo de representación de datos. Como se verá en el capítulo 6, cuando Stata genera un gráfico, el programa abre una ventana especial donde lo ubica y caso de que quiera conservarse, ha de grabarse como un *fichero gráfico*. Stata dispone de un formato propio (*gph*); pero, para que otros usuarios que no usen Stata lo puedan contemplar, también

<sup>&</sup>lt;sup>2</sup> Determinados ficheros de datos creados para las funciones de impulso/respuesta asociadas a modelos VAR y VEC. se guardan con la extensión *irf*, en lugar de *dta*.

<sup>&</sup>lt;sup>3</sup> Por ejemplo, cuando se desea trasladar los resultados a otro programa, como puede ser un procesador de textos, o cuando se quiere que sean leídos en algún ordenador que no disponga del programa Stata.

permite grabarlo en otros formatos tales como metaarchivo de Windows (wmf); metaarchivo mejorado (emf); portable de red (png); postcript (ps); postcript encapsulado (eps); formato de documento portátil (pdf) y el formato de fichero de imagen etiquetada (tif).

Finalmente, es importante cerrar la lista de ficheros de Stata con los denominados ficheros de programa (do y ado), que contienen conjuntos de instrucciones de Stata que pueden ejecutarse automáticamente sin necesidad de tenerlas que volver a introducir interactivamente. Los hay básicamente de dos tipos: los primeros permiten repetir los mismos análisis o transformaciones de datos cuantas veces se desee a los mismos datos; mientras que los segundos se utilizan para aplicar un tratamiento común a datos diferentes, como si fuera una instrucción más del programa; pues se incorporan automáticamente al arrancarlo. Hay cientos de ellos que pueden obtenerse en las páginas oficiales de Stata; pero un usuario avanzado puede construirlos para su propio uso y puede ponerlos a disposición de la comunidad científica. Además, tienen su complemento en los ficheros de ayuda, distinguibles tanto en Stata como en otros programas por su extensión sthlp (o hlp en las primeras versiones), donde se incluyen explicaciones concretas de cómo pueden usarse las instrucciones programadas.

Como resumen, puede confeccionarse el siguiente esquema de los seis tipos de ficheros acabados de describir:

Tipo	Subtipo	Extensiones
Datos	Brutos	.dat, .txt y otras
	Ampliados (Trabajo)	.dta e .irf
Resultados	Textuales	.smcl y .log
	Gráficos	.wmf, .emf, .png, .pdf, .ps y .eps
Programas	Programa	.do y .ado
	Ayuda	.sthlp y .hlp

CUADRO 2.1. Tipos de ficheros en Stata

#### 2.2. La interfaz de Stata

Al ejecutarse Stata, se muestra una pantalla compuesta por una serie de elementos cuyos usos y funciones se dan a continuación. Lo primero que hay que tener en cuenta son las cinco franjas horizontales que presenta la interfaz del programa. Todas estas divisiones, salvo la cuarta, que es la mayor y está compuesta por un conjunto de ventanas, presentan una sola línea de extensión vertical:

# The or fine fine figure before the protection of the bodies topy | Comment | Fine | F

ILUSTRACIÓN 2.4. Primera pantalla de Stata 12

La primera de las zonas presenta el color que por defecto le adjudique el sistema operativo a los programas que con él se ejecutan. En ella están indicados el nombre y la versión del programa que se ha puesto en marcha, el nombre del fichero de trabajo que estuviera abierto, así como la ruta o directorio del ordenador en el que se encuentra.

A continuación, en la segunda zona horizontal, aparece la franja del menú, compuesto por nueve apartados, que son los siguientes:

- 1) *File*: Este ítem del menú permite realizar la apertura, grabación e impresión de los distintos ficheros de trabajo analizados en el apartado anterior.
- 2) El segundo apartado del menú es *Edit*. Sirve para copiar y pegar fragmentos de texto. El uso más común que se da a esta instrucción es la de trasladar los resultados del análisis a otra aplicación como pueda ser un procesador de texto la mayor parte de veces, una hoja de cálculo o un programa de gráficos. También puede utilizarse para cortar y pegar determinados fragmentos de instrucciones de un lugar a otro. Las dos opciones principales de este menú son *copy* (copiar) y *paste* (pegar). Como en la mayor parte de los programas que se ejecutan con *Windows*, ambas pueden ser sustituidas respectivamente por la combinación de teclas Ctrl+c y Ctrl+v. También en este apartado se encuentra la opción de las preferencias (*Preferences*). Permite el cambio de determinados aspectos de las ventanas del programa. Dos son los principales apartados que pueden cambiarse. El relativo a los textos o ventanas y el relacionado con los gráficos. En relación con el primero, en las antiguas versiones de este

programa estaba asociada la imagen de los resultados con una pantalla de fondo negra en la que las instrucciones aparecían en blanco, los resultados estadísticos en amarillo, el texto complementario en verde y los errores en rojo. Todos los elementos de esta combinación pueden cambiarse tanto en la pantalla activa de resultados (Results Colors) como en el visor de otros ficheros (Viewer Colors) de modo independiente. En relación con los gráficos, puede cambiarse el esquema (véase la sección 6.5), la fuente de sus textos y algunos aspectos de la impresión o de su exportación directa<sup>4</sup> a otros programas. Finalmente, la disposición de las ventanas de Stata puede cambiarse si el usuario cambia manualmente el tamaño o la posición de estas y guarda su opción mediante Save Preference Set. A partir de ese momento, el programa se presentará de esa forma incluso después de salir al arrancar de nuevo. Por su lado, hay en la versión 12.0 incorporados seis modelos de disposición: para obtener la configuración mostrada en la ilustración 2.4 hay que optar por el Widscreen Layout, mientras que la disposición clásica de otras versiones se obtiene mediante la opción Combined Layout.

- En los tres apartados siguientes del menú (Data, Graphics y Statistics) se despliegan las múltiples operaciones estadísticas de la que es capaz Stata a través de cuadros de diálogos. En el primero (Data) se incluyen aquellas instrucciones que sirven para describir los datos, transformarlos o hacer manipulaciones al fichero donde están contenidos. Una parte sustancial de estas órdenes están explicadas en la presente obra a lo largo de los capítulos 3 y 5. El segundo de los tres apartados en cuestión (Graphics) está reservado a las instrucciones gráficas. Las más importantes están contempladas en el Capítulo 6. Y, bajo el rótulo de *Statistics*, se dispone la casi totalidad de operaciones estadísticas de la que es capaz este programa. Hay que tener en cuenta que esta posibilidad de obtener resultados estadísticos mediante menús y cuadros de diálogo sólo se ha incorporado a Stata a partir de su versión 8. A un usuario novel de Stata le resultará mucho más cómodo el empleo de estas guías. Sin embargo, un usuario experimentado preferirá escribir directamente las instrucciones una a una o recopilarlas en un fichero para ejecutarlas en serie.
- 4) La denominación *User*, situada en la sexta posición de la franja del menú, sirve para que un programador inserte allí sus propias utilidades. Por tanto, nada será dicho sobre este apartado en este libro introductorio.
- 5) El apartado *Windows* permite acceder a diez de los once tipos de ventanas que componen la estructura interna del programa Stata en su duodécima versión: instrucciones, resultados, historia *(review)*,

<sup>&</sup>lt;sup>4</sup> Se entiende por exportación directa, cuando se utiliza el portapapeles de *Windows* para pasar un objeto de una aplicación a otra. Otro modo de traspasar un gráfico a otra aplicación es grabándolo en un fichero que sea capaz de ser leído por el susodicho programa.

- variables, propiedades, gráficos, visor de ficheros, editor de datos, editor de programas y gestor de variables. Por su especial importancia, se dedicará el próximo apartado a su descripción.
- Finalmente, no falta en el menú el ítem correspondiente a la avuda (Help). En él se distinguen cinco partes diferenciadas: en la primera, sólo existe una línea que remite al manual completo de Stata, dividido en cada uno de sus volúmenes<sup>5</sup>; en la segunda, se ofrece toda la ayuda interna disponible del programa que se muestra en las ventanas de ayuda con un formato especial dotado de hipertextualidad, remitiendo tanto a otros contenidos de la misma ventana como a secciones concretas del manual en pdf; en la tercera, se ofrecen noticias del programa y detalles sobre el contenido de las actualizaciones desde la instalación inicial del software hasta la última versión instalada: en la cuarta aparecen posibles actualizaciones y extensiones del programa así como la ayuda ofrecida en la red, que será mostrada en el explorador de páginas web por defecto que se disponga, y la quinta ofrece el logotipo y dirección de la empresa, la cantidad de memoria física y la disponible en el ordenador, la versión de Stata y la información sobre la licencia en uso. Estas cinco partes se componen de once líneas distintas tal como se muestra en la ilustración 2.5:

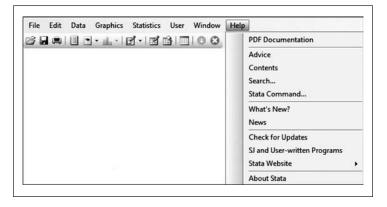


ILUSTRACIÓN 2.5. Menú de ayuda

<sup>&</sup>lt;sup>5</sup> En la versión 12 se cuentan dieciocho unidades en el índice conjunto del fichero contenedor en formato *pdf*: Contenidos, Guía de instalación [IG], Primeros pasos [GS] (una breve introducción con los aspectos básicos de Stata), Guía del usuario [U] (desarrollo de los elementos comunes más importantes de Stata: sintaxis, formatos, variables, funciones, macros y elementos básicos de programación), Gestión de datos [D], Gráficos [G], Imputación múltiple [MI], Estadísticas multivariantes [MV], Referencia básica [R] (listado alfabético de las instrucciones de Stata no contempladas en el resto de manuales), Modelos de ecuaciones estructurales [SEM], Análisis de supervivencia [ST], Datos de encuesta [SVY], Series temporales [TS], Datos longitudinales [XT], Programación [P], Mata [M] (lenguaje específico de programación para trabajar con matrices), Índice [I] y Tabla de contenidos.

Tras la documentación en *pdf*, hay cinco líneas correspondientes a la ayuda interna, que ofrece una serie de consejos para la obtención de distintos tipos de ayuda (*Advice*), un índice temático de los contenidos (*Contents*), un buscador de términos (*Search*), una referencia de todas las instrucciones del programa (*Stata Command*), un repertorio de novedades de la última versión instalada (*What's New*) y un noticiario relacionado con el programa (*News*). En cualquiera de los seis casos, aparece el visor de ayuda donde se expone lo solicitado a modo de hipertexto, de modo tal que se puede navegar por la ingente cantidad de información disponible tanto en el formato propio de Stata como en el formato pdf, que contiene la documentación completa del programa<sup>6</sup>. Como botón de muestra, la ilustración 2.6 contiene la pantalla obtenida al solicitar el índice temático de contenidos:



ILUSTRACIÓN 2.6. Índice temático de contenidos

<sup>&</sup>lt;sup>6</sup> La ayuda que se muestra en la ventana del visor de Stata incluye prácticamente toda la documentación de los manuales con excepción de los ejemplos aislados del texto, las notas técnicas y las referencias bibliográficas.

Las tres siguientes líneas de la ayuda son las últimas actualizaciones disponibles (*Check for Updates*); programas divulgados en los boletines y revistas del Programa (*SJ and User-written Programs*), que pueden incorporarse gratuitamente a los incorporados en el programa, y la página web oficial de la corporación Stata (*Stata Web Site*), subdividida por su parte en página principal, soporte al usuario, FAQ (preguntas respondidas frecuentemente), blog, revista y editorial.

La tercera franja horizontal de la interfaz de Stata es la barra de herramientas, que está constituida en las versiones 11 y 12 por doce iconos con operaciones útiles y frecuentes del programa. Estas son de izquierda a derecha las siguientes: apertura de un fichero de datos, grabación del fichero de datos activo, impresión de los resultados, apertura (visión o cierre) del fichero de resultados, apertura del visor de ayuda, activación de pantalla de gráficos, edición de programas, editor de datos, visor de datos, gestor de variables, botón de continuación en pantalla de resultados y botón de interrupción de resultados.

ILUSTRACIÓN 2.7. Barra de herramientas de Stata 12



En la cuarta franja de la pantalla se ubican entre dos (la de órdenes y la de resultados son inevitables) y cinco ventanas de Stata (las otras tres, optativas y flotantes, son la de variables, la del historial y la de propiedades) que serán descritas con más detalle en el próximo apartado.

Finalmente, en la franja inferior, con el mismo color de fondo que las líneas de menús e iconos, se encuentra la línea de estado, en la que se expone el nombre del directorio de trabajo donde se guardarán y leerán los distintos ficheros, a menos que se especifique un directorio distinto, además de los pilotos que indican en el extremo derecho si se encuentran pulsadas las teclas de fijación de mayúsculas (*CAP*), teclado numérico (*NUM*) e inserción de caracteres (*INS*).

#### 2.3. Las ventanas de Stata

Ya se ha dicho en el apartado anterior que Stata trabaja con once ventanas distintas, cinco internas y seis externas. No todas son igual de importantes, ni todas están presentes al mismo tiempo. De hecho, al empezar una sesión con Stata 12 sólo aparecen las cinco internas. Entre ellas, las más centrales para el trabajo son las de resultados y las de órdenes.

La ventana de órdenes (Stata Command), ubicada por defecto (Widescreen layout) en la parte inferior de la cuarta franja del programa, es un re-

cuadro en blanco donde deben escribirse las instrucciones u órdenes de Stata. Una instrucción básica para empezar es *dir*, para saber los ficheros contenidos en el directorio de trabajo. Si se escribe en la ventana en cuestión esta palabra seguida por la tecla de retorno como final de la orden, inmediatamente aparecerá un texto en la *ventana de resultados*.

dir

En este caso, aparecerá un texto similar al de la ilustración 2.8:

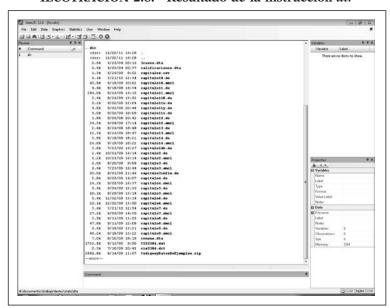


ILUSTRACIÓN 2.8. Resultado de la instrucción dir

La ventana de órdenes, donde se escribió la instrucción, se queda en blanco después de ejecutarla. Sin embargo, la orden queda guardada en otra pantalla, mostrada aquí en la parte superior derecha de la ilustración, en la llamada *ventana de historia (Review)*. Además, en la ventana de resultados (*Stata Results*) aparece el producto de la primera instrucción, esto es, un listado con todos los ficheros ubicados en el directorio por defecto. Si hay más ficheros que líneas permite el tamaño de la pantalla, aparece el texto —*more*— en color diferente al del resto. Ante este mensaje, hay tres posibilidades: la primera es apretar la tecla *l* o Intro (*Enter*), en cuyo caso, en la pantalla de resultados aparecerá una línea más. La segunda opción es apretar otra tecla distinta de las dos anteriores o el penúltimo icono (*Clearmore--condition*). De este modo, el texto, en lugar de avanzar una línea,

avanzará toda una pantalla. Finalmente, si se desea interrumpir la salida de resultados, en el caso de que no se haya obtenido lo deseado, se puede pulsar q, la combinación de teclas Ctrl+k, o el último icono de la barra de herramientas. Así se detendrá la orden, no aparecerán más líneas en pantalla y se estará en condiciones de escribir una nueva instrucción. Una línea con un solo punto en la pantalla de resultados indica que el sistema está listo para recibir otra orden.

Para ver algún contenido en la *ventana de variables (Variables)*, es preciso crear o recuperar un conjunto de datos. Como la segunda de estas dos tareas es más fácil, se pondrá como primer ejemplo la recuperación de una base de datos incorporada en el programa y se deja la creación de un fichero de datos para un capítulo posterior. Con el fin de recuperarla, en la ventana de órdenes hay que escribir *sysuse auto*, siendo esta última palabra el nombre con el que es conocida esta base de datos, que consiste en un listado de automóviles comercializados en los años setenta en Estados Unidos acompañado con una serie de variables cuyo contenido son las características de los modelos.

sysuse auto

Desde el momento en que se introduce esta instrucción, aparece en la ventana de variables la lista de ellas que están incluidas en el fichero *auto*. Es fácil advertir que cada línea corresponde a una variable y está dividida en cuatro columnas: la primera refleja su propio nombre, la segunda contiene su etiqueta, es decir un texto que la acompaña y que proporciona una descripción más extensa de su contenido, la tercera informa del tipo, mientras que la cuarta refleja su formato<sup>7</sup>. Estas propiedades de las variables serán descritas con más detalle en las secciones 2.6.1, 2.6.2 y 2.6.3.

Tanto la ventana de historia como la de variables permiten trasladar su contenido a la pantalla de órdenes. De este modo, si se lleva el cursor a la línea *dir* de la primera de las ventanas, aparecerá el texto en la ventana de órdenes y, si se pulsa *Intro* en esta, o si se pulsa un par de veces el texto en la ventana de historia, la instrucción mencionada será ejecutada de nuevo. En cambio, si se desea eliminar de la ventana de órdenes por haberla incluido por error, puede hacerse mediante la tecla *Esc*. En el caso de las variables, el funcionamiento es similar, salvo en que el doble *clic*, en lugar de ejecutar la instrucción, traslada el nombre de la variable a la ventana de órdenes.

<sup>&</sup>lt;sup>7</sup> La presentación de la ventana de variables con cuatro columnas (ilustración 2.9) no se obtiene por defecto. Para que aparezcan las dos últimas (*type y format*) se ha de colocar el cursor en la barra de contenido (VariablelLabel), hacer clic posteriormente en el botón derecho del ratón y, finalmente, marcar la(s) columna(s) deseada(s).

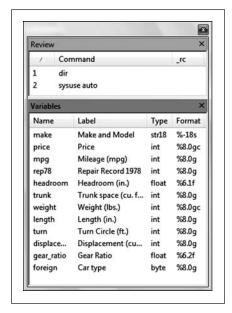


ILUSTRACIÓN 2.9. Ventanas de historia y variables

Otra ventana de frecuente uso en Stata, ya mostrada en la ilustración 2.6, es el visor de ayuda. Con la instrucción help orden (dir, por ejemplo) se obtiene la información correspondiente en una ventana independiente. También, en lugar de escribir la instrucción, se puede solicitar ayuda mediante el ítem Help/Stata Command de la barra de menús. De este modo, aparece un cuadro de diálogo que pide al usuario una orden de Stata y muestra de ella prácticamente toda la información contenida en el manual en una ventana independiente. Ésta posee además varios botones e iconos que realizan operaciones como búsqueda de otras órdenes, exploración de contenidos (search) y búsqueda de cadenas en el interior del visor.

La ventana visor no sólo se emplea para visualizar la ayuda, también puede utilizarse para ver un fichero en ASCII o en formato *smcl*, propio de los resultados grabados de Stata, como se explica con más detalle en la sección 2.5.

Hasta el momento, los textos generados por las instrucciones introducidas en la ventana de órdenes han aparecido en la ventana mayor de Stata, esto es, en la de resultados, o en el visor. Además, hay otro tipo de instrucciones, las gráficas, que muestran su resultado en una ventana diferente. De este modo, si se escribe la siguiente instrucción:

histogram price, freq

Aparece una nueva ventana por encima de la de resultados. Es la *ventana gráfica* de Stata, que se superpone a la anterior. Los resultados pueden ser vueltos a poner en primer plano, pulsando su primera franja. Alternativamente, el gráfico puede aparecer de nuevo pulsando el sexto icono de la barra de herramientas de la pantalla principal del programa.

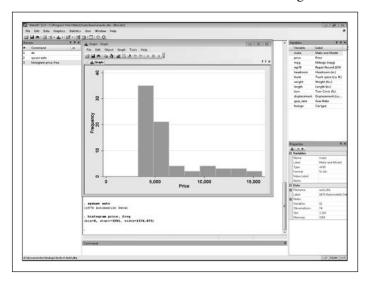


ILUSTRACIÓN 2.10. Ventana de resultados gráficos

También pueden ser considerados otro tipo de ventanas de Stata los *cuadros de diálogos* que aparecen al solicitar cualquier tarea mediante el menú. Las ilustraciones 2.13 y 2.15 muestran un par de ejemplos de este tipo de ventanas, que tienen la propiedad exclusiva de generar instrucciones de Stata que se acumulan en la ventana de historia y, una vez ejecutadas, muestran su producto en la ventana de resultados. Casos especiales de este tipo de ventanas, presentes sólo a partir de la versiones 11 y 12, son el gestor de variables y el módulo de propiedades, cuyo uso se verá al final de este capítulo.

La utilidad que en Stata crea o modifica la información analizable, conocida como *editor de datos (Data Editor)*, o la que los inspecciona (*Browser*), generan el octavo tipo de ventana. Su función es mostrar y permitir hacer modificaciones (esta última función sólo en la primera opción) de los datos cargados en la memoria. Como puede apreciarse en la ilustración 2.11, se trata de una ventana, similar a la de una hoja de cálculo, en la que los casos se representan en las líneas y las variables en las columnas. Así, los tres primeros casos corresponden a los modelos *Concord*, *Pacer* y *Spirit* de la casa de automóviles AMC, que tenían en 1979 precios

respectivos de 4.099, 4.749 y 3.799 dólares. En el caso de que se quiera realizar algún cambio, basta llevar el cursor a la casilla correspondiente y reemplazar el valor antiguo con uno nuevo. Tras realizar los cambios deseados, se puede cerrar la ventana con el botón situado en su extremo superior izquierda que tiene una figura de aspa o, si se prefiere, mantenerla abierta.

edit

ILUSTRACIÓN 2.11. Ventana del editor

Varios son los caminos para acceder a la ventana del editor. El más rápido es pulsar el octavo botón de la barra de herramientas (*Data Editor - Edit*). También puede hacerse escribiendo la orden *edit* en la ventana de instrucciones, mediante menú, seleccionando *Data/Data Editor/Edit*, o pulsando la combinación de teclas Ctrl+8. En los cuatro anteriores casos, se permite al usuario realizar modificaciones. Ahora bien, si el propósito es sólo contemplar los casos, sin realizar ningún cambio, es preferible entrar al editor mediante la orden *browse*, pulsando el noveno botón (*Data Editor - Browser*) o seleccionando del menú la entrada *Data/Data editor/Browser*.

Para finalizar la descripción de *ventanas* de Stata, es preciso explicar de antemano qué es un *programa*, ya que la última que se contempla en este apartado es la de su editor. Un programa de Stata se compone de un conjunto de instrucciones reunidas en un fichero. La forma más cómoda de gene-

rarlo es convirtiendo la historia de instrucciones en un fichero que las contenga. Siguiendo con este primer ejemplo, es decir, las cuatro órdenes para explicar las distintas ventanas de este paquete estadístico, es fácil convertirlas a un fichero de programas haciendo clic con el botón derecho del ratón en la ventana de la historia y señalando la opción *Save All* o *Save Selected*. Tras ello, sale el menú de grabación de ficheros propio del sistema operativo con el que se trabaje y se puede poner el nombre que se desee. Automáticamente Stata le pondrá la extensión *do*, que es con la que se reconoce a este tipo de archivos. Después de grabado, un fichero de programa puede revisarse o ejecutarse cuantas veces se desee. Otra opción más práctica es mandar este contenido de la ventana de la historia al editor de ficheros de programas de Stata mediante la línea del menú contextual *Send to Do-file editor*.

El editor de estos ficheros también puede ponerse en marcha abriendo su ventana, bien escribiendo en la ventana de órdenes la palabra *doedit*, bien a través de menú (*Window/Do-file Editor*), bien haciendo clic en el séptimo icono de la barra de herramientas (*Do-file Editor*) o pulsando la combinación de teclas Ctrl+9.

doedit

ILUSTRACIÓN 2.12. Ventana del editor con el contenido del "Primer programa.do"



Una vez ejecutado el programa y abierta su correspondiente ventana, hay que abrir el fichero ya guardado (o empezar a escribir uno nuevo)

mediante el menú (*File/Open*), el segundo icono de las herramientas (*Open*) o bien la combinación de teclas Ctrl+o. Después se selecciona el fichero con extensión *do* deseado en el directorio donde se encuentre y todas las instrucciones de las que se componen aparecerán en la nueva pantalla, de tal modo que aparecen remarcados con diferentes colores, entre otros elementos, las instrucciones, las cadenas, los operadores, las funciones y los comentarios<sup>8</sup>.

Dentro de esta ventana, pueden hacerse cuantas modificaciones se consideren oportunas escribiendo, borrando, copiando, cortando y pegando como en cualquier editor, y ejecutarlas cuantas veces se desee.

Para esto último hay dos modos: el primero es do, en cuyo caso aparecen las órdenes en la pantalla de resultados, y el segundo es run, se ejecutan las órdenes pero su contenido y resultados se ocultan. Ambas se encuentran bajo el rótulo del menú Tools, y son respectivamente el último (Execute (do)) y el penúltimo (Execute quietly (run)) icono de la barra de herramientas propias del editor, cuyo aspecto el lector atento habrá notado diferente del que aparece en la pantalla general de Stata. También es posible realizar la misma operación con las respectivas combinaciones de teclas Ctrl+d o Ctrl+r. Ambos modos funcionan con el conjunto del fichero o con una selección parcial de las órdenes que se consideren más apropiadas para una determinada tarea. Además, con el fin de que todas las modificaciones queden guardadas para uso posterior, también se permite en esta ventana la grabación de su contenido sea mediante menú (File/Save), icono (el tercero, Save), o teclas (Ctrl+s) o (Ctrl+May+s) en el caso de que se quiera dar un nombre distinto al fichero que se graba<sup>9</sup>.

Resumiendo el contenido de este apartado, son once los tipos de ventanas de Stata. Cinco de ellas son internas y aparecen directamente al iniciar el programa: la de *órdenes*, donde el usuario puede ir escribiendo una a una cuantas instrucciones considere relevantes; la de *resultados*, donde aparecerá la ejecución de la instrucción; la de *historia*, en la que se acumularán todas las instrucciones ejecutadas desde el comienzo de la puesta en marcha del programa; la de *variables*, donde se muestra la lista de ellas del fichero de datos que en cada momento se encuentre cargado en memoria y la de *propiedades*, cuyo uso se verá en la sección 2.6.4. Los otros seis tipos de ventanas, las externas, aparecen cuando se realiza una operación que las necesita. En esta categoría se encuentran: el *visor*, que es capaz de mostrar

<sup>&</sup>lt;sup>8</sup> En la pestaña *Syntax Color* del cuadro de diálogo que aparece con el menú *Edit/Preferences* del editor de programas aparecen todos los elementos remarcables y se permite cambiar su color y otras propiedades de su fuente.

<sup>&</sup>lt;sup>9</sup> Otra posibilidad importante para la elaboración, comparación y ejecución de programas en la capacidad de disponer de varios ficheros a la vez en distintas pestañas de la misma ventana. Para lograr abrir más ficheros, se puede pulsar Ctrl+o, el segundo icono u obtenerlo mediante menú (*File/Open*).

ayuda del programa y resultados grabados; la pantalla de *gráficos*, donde se muestran resultados que no son representables mediante caracteres de texto; el *editor de datos*, para ver o modificar los datos cargados en el programa; los *cuadros de diálogo*, para escribir instrucciones con más facilidad; el *editor de programas*, para la confección, grabación y ejecución de una secuencia de instrucciones que permita resolver peticiones complejas y el *gestor de variables*, que permite asignar o modificar las propiedades de las variables.

#### 2.4. Modos de trabajo en Stata

Hay tres formas distintas de proporcionar las instrucciones a la aplicación Stata para obtener los resultados deseados. En este apartado se contempla cómo se emplea cada una de ellas. Aunque prácticamente todo pueda realizarse con estos tres modos de trabajo, la elección de cuál usar dependerá de la tarea que se haga y de las preferencias del usuario. Los tres modos son el de *instrucción*, cuando se introducen literalmente una a una cada orden; el de *menú*, cuando se utiliza un cuadro de diálogo para efectuar una petición, y el de *programación*, en el caso de querer ejecutar automáticamente un conjunto de instrucciones.

#### 2.4.1. Modo instrucción

Este modo de trabajo se basa en la inserción manual de instrucciones en la ventana de órdenes. Se caracteriza por ser interactivo ya que cada línea introducida por el usuario genera un resultado y, hasta que este no se complete, no se puede introducir la siguiente orden.

Toda instrucción de Stata está compuesta al menos por una palabra, que es la *orden* propiamente dicha, a veces precedida por una *preinstrucción*, de la que se separa por dos puntos; seguida generalmente por unas *especificaciones*; matizada, si procede, por unos *calificadores*, y ampliada, si cabe, con una serie de *opciones* propias de cada orden, que deben separarse del resto de la instrucción mediante una coma.

La estructura, por tanto, de toda instrucción presenta el siguiente esquema:

[preinstrucción:] orden [especificaciones] [calificadores] [,opciones]

Puesto que lo expuesto en corchetes es optativo, por la sintaxis empleada cabe deducir que lo único obligatorio en cada instrucción es la orden. Ahora bien según sea ésta, las especificaciones serán obligato-

rias u optativas. Por ejemplo, puede darse la orden *help* sin ninguna especificación, pero no puede emplearse *histogram* seguida de ningún nombre, ya que al menos requiere que se le incluya el de una y sólo una variable.

Fijándose en las cuatro últimas instrucciones que quedan presentes en la ventana de historia de órdenes, éstas eran:

help sysuse auto histogram price, freq edit

Como fácilmente puede apreciarse, la primera y la última sólo constan de órdenes, mientras que las dos centrales tienen especificaciones. En la segunda la especificación es el nombre del fichero; en la tercera el nombre de una variable. Y la tercera instrucción contiene también una opción *frequency*, que ha sido abreviada<sup>10</sup> con sus cuatro primeras letras. Es fundamental retener que todas las opciones han de figurar detrás de la coma, separadas entre sí al menos por un espacio en blanco.

Además de la orden, sus especificaciones y opciones, la mayor parte de ellas pueden incorporar preinstrucciones, que modifican el funcionamiento de la instrucción, como, por ejemplo, aplicándola a distintas submuestras, y calificadores que restringen el uso de la instrucción a casos con una determinada característica. Todas estas posibilidades son tan importantes que serán tratadas con algo más de detenimiento en el capítulo 5.

#### 2.4.2. Modo menú

El segundo modo de proporcionar instrucciones al programa es mediante los menús. Cuando se habló de la interfaz de Stata y se mencionó la primera zona horizontal de su ventana, se dijo que desde el tercer al quinto de sus apartados (*Data, Graphics y Statistics*) se podían encontrar prácticamente la totalidad de las instrucciones propias del paquete, mientras que aquellas

Las abreviaturas en Stata pueden emplearse en la inmensa mayor parte de las ocasiones, siempre que no produzcan ambigüedad. De este modo, pueden abreviarse nombres de variables y opciones. Las órdenes sólo pueden recortarse de la forma que se indica en el manual o en la ayuda del programa. Por ejemplo, el programa entiende *hist*; pero interpreta como error *histo* o *histogra*. Sin embargo, en el fichero *auto*, la variable *price* puede ser escrita como *p, pr, pric* y *price*, ya que ninguna otra variable comienza por *p*. De la misma forma, la opción *frequency* puede ser abreviada, al menos, con *freq*. También entendería bien el programa *frecuenc*.

adicionales que el usuario considerara convenientes podían ser añadidas en el siguiente elemento (*User*).

Este modo de trabajo, que facilita la producción de instrucciones a quienes no conocen la sintaxis del programa, fue incorporado a partir de la versión 8, por lo que muchos usuarios anteriores de esta aplicación prefieren seguir utilizando el anterior modo de trabajo, porque es más rápido de usar si se conoce bien. También tratan de evitar este modo de trabajo quienes programan, ya que necesitan recordar continuamente las distintas palabras claves que hay que utilizar en la confección de los programas.

Con el sistema de los menús se han trasladado todas las posibilidades de una orden a un cuadro de diálogo, tanto más complejo cuanto más lo es la instrucción, que se obtiene bien presionando la línea correspondiente del menú, bien escribiendo una orden que lo ponga en funcionamiento.

Por ejemplo, si se desea un histograma de la variable *price*, habrá que optar por especificar *Graphics/Histogram* trasladando el cursor mediante el ratón a los correspondientes elementos que lo componen (primero a *Graphics*, a continuación a *Histogram*) y haciendo clic, una vez obtenido este último. El resultado es un cuadro de diálogo ubicado en una nueva ventana:

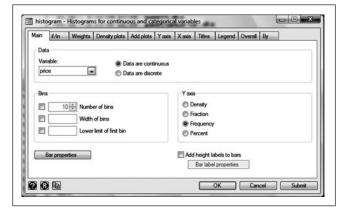


ILUSTRACIÓN 2.13. Cuadro de diálogo de la orden histogram

En la ilustración 2.13 aparece el cuadro de diálogo en el que se ha insertado el nombre de la variable en el correspondiente recuadro (Variable) y se ha cambiado la opción Y-axis, marcando Frequency, en lugar de Density, que es con la que opera por omisión. La orden histogram posee más posibilidades que las que se muestran en el cuadro de diálogo principal (main). Por ello en la línea superior hay otras pestañas referentes de otros conjuntos de opciones. Sólo a modo de ejemplo, se muestra a continuación el cuadro de diálogo de la pestaña Density Plots:

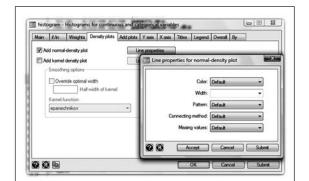


ILUSTRACIÓN 2.14. Cuadro de diálogo de la pestaña Density Plots

Puede apreciarse asimismo que en la línea inferior de cualquier cuadro de diálogo existen seis botones. Tres pequeños en la parte izquierda, el primero (?), para obtener ayuda; el segundo (R), para limpiar el contenido de todos los campos del cuadro y dejarlos en sus opciones por defecto. El tercero es para copiar en el portapapeles el texto de la instrucción que se está construyendo mediante el menú. En la parte derecha, son tres los botones rectangulares con fondo claro. El primero y el último mandan la instrucción: uno, *OK*, cerrando el cuadro de diálogo, el otro, *Submit*, manteniéndolo abierto. El del medio, *Cancel*, sirve para cerrar el cuadro de diálogo sin ningún efecto. Así, pues, tanto el primer como el tercer botón de la parte derecha sirven para obtener un histograma como el mostrado en el gráfico 2.1.

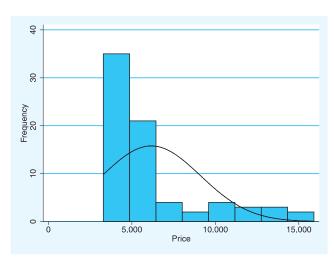


GRÁFICO 2.1. Histograma del precio de los coches

Lo más común es utilizar los cuadros de diálogos como se acaba de explicar, es decir, buscándolos a través del menú. Sin embargo, usuarios experimentados de Stata podrían hacerlo de otro modo, empezándolos desde la línea de instrucciones, sobre todo, en el caso en el que las órdenes sean muy complejas y no se recuerden todas sus modalidades y opciones. Para ello hay que escribir db seguido de la orden, cuyo cuadro de diálogo se desea obtener. De esta forma, si se quiere que aparezca el cuadro de diálogo de la instrucción sysuse, se escribirá una línea como sigue:

db sysuse

Inmediatamente aparecerá el cuadro de la ilustración 2.15, mucho más simple que el anterior, en la medida de que se trata de una orden con menos parámetros y opciones que la de *histogram*. En este cuadro de diálogo se ofrecen dos posibilidades: abrir un fichero (*Select File*), cuyo nombre hay que escribir en el cuadrado *File to select*, o mostrar todos los ficheros de datos incorporados en el programa disponibles para el usuario (*Show datasets available in the adoptath*). Por último, puede marcarse la opción de descartar datos previamente cargados en memoria (*Automatically discard data in memory*), ya que si existieran unos datos previos modificados con el programa, no podría abrirse el nuevo fichero, a menos que se grabaran las modificaciones o se especificara la opción en cuestión<sup>11</sup>.

Sysuse - Use shipped dataset

Options
Select file
Show datasets available in the adopath

File to select
autol
Automatically discard data in memory

OK
Cancel
Submit

ILUSTRACIÓN 2.15. Cuadro de diálogo de la orden sysuse

<sup>&</sup>lt;sup>11</sup> Stata no permite trabajar con dos ficheros al mismo tiempo. En su lugar, se puede ejecutar el programa varias veces con archivos de datos distintos. Obviamente, no habría interferencias, ni intercambios en el trabajo entre ellos.

#### Modo programación 2.4.3.

La tercera posibilidad de trabajo con Stata es el modo de programación, que consiste en escribir una serie de instrucciones necesarias para llevar a cabo una tarea, grabarlas en un fichero y desde este ejecutarlas cuantas veces se desee con o sin cambios en las órdenes que lo necesiten. En el apartado anterior, cuando se describió la ventana del editor de programas, se explicó que hay dos formas de ejecución de estos ficheros: uno, mediante la instrucción run, en cuyo caso no se muestran las líneas de instrucción; el otro, mediante la instrucción do, para obtener el mismo resultado, pero con las órdenes incluidas. Existe un tercer modo de ejecutar un programa escrito con el lenguaje de Stata. Se trata de hacerlo desde el sistema operativo. Cualquier fichero con extensión do, con sólo aplicarle un doble clic, es capaz de cargarse con Stata y ejecutarse.

Si además se incluye en el fichero la instrucción set more off, el usuario puede desentenderse del proceso y el programa ejecutará sin interrupción todas las instrucciones incluidas. También pueden añadirse comentarios: a) poniendo un asterisco al inicio de cualquier línea, b) colocando los comentarios entre /\* y \*/, c) después de espacio seguido de dos barras hasta el final de la línea y d) del mismo modo que en c) pero con tres barras, en cuyo caso la línea siguiente será considerada de la misma orden. Esta última opción es un modo óptimo en los programas para disponer en varias líneas las instrucciones muy largas.

Por ejemplo, si con el editor preferido<sup>12</sup> se escribe un fichero con las siguientes instrucciones llamado listauto.do:

#### ILUSTRACIÓN 2.16. Contenido del fichero listauto.do

\*\*\*\*\*\*\*\*\*

- \* M. Escobar, E. Fernández, F. Bernardi
- \* Análisis de datos con Stata
- \* Madrid, CIS, 2009

set more off //Esta instrucción sirve para que no se pare la pantalla. sysuse auto, clear /\*Lee uno de los ficheros ejemplos de Stata\*/

list make /// Con tres barras entiende que la orden continúa en la siguiente línea.

set more on //Vuelve a parar la pantalla de resultados cuando se llene.

<sup>12</sup> Incluso puede emplearse un procesador de texto, siempre y cuando a la hora de grabar se tenga la precaución de grabar el fichero en formato ASCII y se le ponga la extensión .do.

Se obtendrá un listado de las dos variables expresadas (*make* y *price*) para todos los casos del fichero *auto* del que a continuación se ofrece un extracto, escribiendo la instrucción:

do listauto

ILUSTRACIÓN 2.17. Listado de casos

-	+	
	make	price
1.	AMC Concord	4,099
	AMC Pacer	4,749
	AMC Spirit	
	Buick Century	4,816
5.	Buick Electra 	7,827
6.	Buick LeSabre	
7.	Buick Opel	4,453
	Buick Regal	
	Buick Riviera	
10.	Buick Skylark 	
11.	Cad. Deville	
	l <b></b>	
66.	   Subaru	3,798
67.	Toyota Celica	
	Toyota Corolla	
	Toyota Corona	
70.	VW Dasher	7,140
71.	   VW Diesel	5,397
	VW Rabbit	4,697
73.	VW Scirocco	6,850
74.	Volvo 260	11,995
-	+	

Este modo de programación puede hacerse tan flexible como se quiera, tanto por la posibilidad de intercambiar parámetros (enviar al programa información, que luego es devuelta para su uso en la ventana principal) como por la de incluir instrucciones de control de flujo, que permite poner en manos del usuario la posibilidad de escribir con un lenguaje sencillo sus propias rutinas, superando de este modo la rigidez que imponen otras aplicaciones estadísticas que no permiten obtener estadísticos distintos de los que ya vienen preprogramados en el paquete<sup>13</sup>.

Quienes estén interesados en este tipo de procesos pueden consultar los capítulos 16-18 de la guía del usuario (Stata Corporation, 2011c) y el volumen del manual de Stata dedicado a la programación (Stata Corporation, 2011i).

#### 2.5. El fichero de resultados

Quien haya trabajado una larga sesión con Stata y haya querido volver a mirar los resultados de los primeros análisis habrá advertido que la pantalla de resultados tiene una capacidad limitada, pues no es capaz de almacenar más allá de una determinada cantidad de caracteres (200.000 por defecto en la versión 12 de Stata aunque puede ser ampliada en *Edit/Preferences/General Preferences/Windowing* hasta 2 millones. En trabajos largos esto es un enorme inconveniente. Además, cualquiera que sea la longitud de los resultados, Stata no es capaz de grabarla mediante una instrucción. Si se desea guardar total o parcialmente su contenido, es preciso marcar el bloque deseado, y optar mediante menú de cabecera (*Edit/Copy Text*) o menú contextual (*Copy Text*) su traslado a otro programa, un procesador de texto, por ejemplo<sup>14</sup>.

Sin embargo, el modo en que Stata ha previsto que no se tenga que realizar esta tarea de cortar y pegar, cada vez que se genera un resultado que se quiera guardar, es mediante la grabación de la pantalla de resultados en un fichero. Esta operación no es automática y ha de ser el usuario quien inicie el proceso, lo detenga, lo continúe o lo cierre.

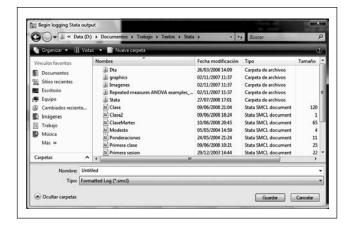


ILUSTRACIÓN 2.18. Pantalla de inicio de ficheros de resultados

Como otras operaciones frecuentes de Stata, la creación de un fichero de resultados se puede realizar de cuatro modos: mediante instrucción inte-

Caso de que se copie un resultado de Stata a un procesador de texto, es imprescindible darle una fuente con tipo de letra de espacios fijos (Courier o Lucida), ya que las fuentes proporcionales (Times, Arial, Century, entre otras muchas) producen textos de tamaño variable y, por tanto, no generan textos o números alineados verticalmente.

ractiva o programada (log using), con icono (el cuarto de la barra de herramientas, Log begin), teclado (Ctrl+L) o mediante menú (File/Log/Begin). Con las tres últimas modalidades, aparece una ventana (ilustración 2.18) con un listado de ficheros con extensión smcl, que son aquellos en los que Stata guarda sus resultados con un formato propio, en todo momento convertibles a ficheros con formato plano en ASCII. En la mencionada ventana, debe escribirse en la casilla Nombre el título que se quiera dar al fichero donde a partir de ese momento se grabarán todos los resultados. También puede elegirse el formato de este fichero. Aunque, si nada se indica, Stata utiliza su formato propio (smcl); se puede cambiar desde el principio de la grabación, optando en la casilla Tipo por la extensión log.

Esta operación también puede hacerse mediante instrucción, sea en la pantalla de órdenes, sea en un programa. Por ejemplo, si se desea generar un fichero de resultados llamado primero(.smcl)<sup>15</sup>, habrá que escribir la siguiente línea<sup>16</sup>:

```
log using primero
```

Si no existe ya ese fichero en el directorio actual de trabajo y si no se ha abierto con anterioridad algún otro fichero de resultados, aparecerá en la ventana de resultados un texto que advierte la operación realizada:

# ILUSTRACIÓN 2.19. Cabecera de la apertura de un fichero de resultados

```
log: C:\Documents and Settings\...\Mis documentos\stata\primero.smcl
log type: smcl
opened on: Jan 2004, 12:37:07
```

A partir de este momento, todo lo que aparece en la ventana de resultados, salvo la ayuda, será grabado en el fichero, directorio y disco del ordenador especificado. La grabación puede ser revisada, suspendida o finalizada. Si se intenta hacer cualquiera de estas operaciones mediante menú (File/Log), mediante icono de la barra de herramientas (Close/Suspend), o con el teclado (Ctrl+L), aparece un cuadro de diálogo para que el usuario opte por la fórmula deseada.

<sup>&</sup>lt;sup>15</sup> Caso de que el fichero contenga espacios en blancos, es obligatorio que su nombre sea escrito entre comillas.

Desde la versión 12 se puede añadir como opción un nombre interno al fichero de resultados [, name(nombre\_interno)]. De este modo se puede tener abierto más de uno al mismo tiempo. El control del uso sería con la instrucción log {offlon} nombre\_interno. También son útiles las opciones replace y append, que sirven respectivamente para regenerar un fichero ya existente o para añadirle los nuevos resultados.

## ILUSTRACIÓN 2.20. Cuadro de diálogo para un fichero de resultados ya abierto



Todas esas operaciones también pueden realizarse mediante instrucciones en la ventana de órdenes. Estas son:

view nombrefichero.smcl log close log off

Al igual que se puede grabar un fichero de resultados, también puede hacerse algo similar con todas las instrucciones de una sesión de Stata. Ya se ha visto cómo puede hacerse *a posteriori*, haciendo aparecer el menú de contexto en la ventana de historia. Pero también puede realizarse *a priori*, mediante la instrucción *cmdlog using* nombrefichero<sup>17</sup>. Si se quiere crear un fichero llamado primeras instrucciones, con extensión *do* habrá que escribir la siguiente orden.

cmdlog using "primeras instrucciones.do"

Y para suspender, reanudar o terminar la grabación, ha de usarse la instrucción *cmdlog* acompañada de *off, on* o *close* respectivamente.

Tanto en ficheros de resultados como en ficheros de instrucciones, otro aspecto que ha de tenerse en cuenta es que, en el caso de que se quiera dar un nombre de fichero ya existente, si se intenta con una línea de instrucción, dará un error, a menos que se añada la opción *append*, si se quiere

<sup>&</sup>lt;sup>17</sup> Si no se especifica ninguna extensión al fichero, el programa le pondrá .txt. Si se desea repetir una sesión, conviene cambiarle la extensión por la de .do. De este modo podrán ejecutarse de nuevo.

añadir lo nuevo a lo existente, o replace, si se prefiere suplantar el antiguo contenido por el nuevo.

En cualquier momento, por otro lado, se puede visualizar cualquier fichero de resultados, comandos e incluso de ayuda con la instrucción *view* nombrefichero. Esta posibilidad se encuentra también en el menú *File/View*. Si, a continuación, se aprieta en el botón *Browse*, aparece el explorador de ficheros, desde el que puede seleccionarse cualquier archivo escrito en ASCII o en SMCL para su visualización<sup>18</sup>. En cualquier caso, también puede llamarse directamente al fichero si en la instrucción *view* se especifica su nombre y extensión.

view primero.smcl

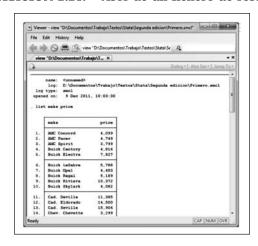


ILUSTRACIÓN 2.21. Visor de un fichero de resultados

#### 2.6. Las variables de la matriz de datos

Como la mayoría de programas informáticos, Stata es básicamente una herramienta para manipular datos: datos estadísticos en este caso. El funcionamiento de Stata consiste en manipular, modificar y realizar operaciones matemáticas sobre una matriz de datos que se almacena en la memoria del ordenador. Para que Stata pueda trabajar con estos datos, tienen que estar en el formato adecuado, en el formato de Stata. En el próximo capítulo, se verá cómo construir o traducir los ficheros propios de este programa, bien

<sup>&</sup>lt;sup>18</sup> Ha de tenerse en cuenta que se permite la visión de varios ficheros al mismo tiempo, tanto en nuevas ventanas como en nuevas pestañas (*Tab*).

sea introduciendo los datos manualmente o traduciéndolos desde ficheros construidos por otros programas (como SPSS o Excel) al formato de Stata. Pero, primero, es conveniente explicar operaciones más fáciles como su uso, el formato y la disposición de datos que en ellos se contiene.

Al igual que la mayor parte de programas estadísticos, Stata trabaja con un fichero de datos estructurado por casos y variables. Este fichero ha de residir en un determinado directorio de una unidad del ordenador. Para hacer uso de él y aplicarle cuantas operaciones estadísticas se desee, ha de cargarse en la memoria central de la máquina. Ello se consigue abriendo el fichero mediante el menú *File/Open*, el primer icono de la barra de herramientas (*Open*), la combinación de teclas Ctrl+0, o la instrucción *use* nombrefichero. Anteriormente se ha utilizado la instrucción *sysuse*, pero esta sólo sirve para ficheros suministrados por el programa para mostrar ejemplos incorporados en el manual. Si se quiere, en consecuencia, abrir un fichero previamente creado por el usuario, como puede ser el fichero con información de países, al que se tituló con el nombre "mundo", la instrucción debería incluir la opción *clear*, si se desea descartar¹9 posibles modificaciones realizadas en un conjunto de datos cargados en memoria.

use mundo, clear

A partir de ese momento, si el programa ha localizado correctamente el fichero, la matriz de datos se carga en la memoria del ordenador y, como prueba de ello, aparecen las variables en su ventana correspondiente, siempre que se encuentre abierta. Como se ha visto anteriormente, en la pantalla del editor de datos, esta matriz se representa habitualmente como una tabla de datos en la que las filas son casos y las columnas variables (así se representa por ejemplo en el editor de datos de Stata, como se verá con atención en el siguiente apartado). Los casos son los individuos de los que se dispone información, y las variables son las categorías en las que se organiza esa información. En este ejemplo, al disponer de un conjunto de indicadores de países, los casos serán cada uno de los países incluidos, y las variables cada uno de los indicadores disponibles como la superficie, la población, el PIB, etc.

Para conocer las características de las variables que están contenidas en un fichero, Stata posee una instrucción que construye una lista de todas ellas, salvo que se especifique tras la instrucción un subconjunto

<sup>&</sup>lt;sup>19</sup> Si se usa una versión de Stata anterior a la 12, podría ocurrir que el fichero fuera demasiado grande para que cupiera en la memoria que Stata reserva por defecto (10 Megabytes) al iniciarse el programa. Con la orden *set mem* #M, el usuario puede cambiar la capacidad reservada. A partir de dicha versión, el usuario no tiene que preocuparse de la ampliación de la gestión de la memoria, puesto que el programa la gestiona automáticamente.

de variables. Esta orden es *describe*. Usada sin argumento ni opción, proporciona de cada variable información sobre el tipo de almacenamiento, el formato de presentación y las etiquetas. Aplicada sobre el fichero *mundo* del presente ejemplo, muestra el resultado presentado en la ilustración 2.22.

Se advierte en primer lugar que el fichero consta de 213 observaciones, cada una de ellas corresponde a un país. También se indica que está compuesto por 17 variables y que el tamaño que ocupa en disco (y actualmente en memoria) es de 20.661 bytes.

En la ilustración 2.22 se lista el conjunto de variables disponibles en el fichero, ahora cargadas en la memoria del programa. Cada una de ellas aparece, junto con su nombre, con su tipo de almacenamiento, su formato de presentación y sus etiquetas, conceptos todos ellos que se explican en los tres próximos apartados. De estas propiedades de las variables, la más sencilla, útil, necesaria y empleada es el etiquetaje. Por ello, se aborda en primer lugar. Las otras dos, el tipo y el formato, son más complejas y no tan necesarias, por lo que si no se entienden en un primer momento, el lector puede continuar con los siguientes capítulos de este libro sin temor a perderse algo imprescindible.

#### ILUSTRACIÓN 2.22. Descripción de variables

Contains data obs:	from muno 213	do.dta		Indicadores de los países. Mundo (2002) Fuente: The World Bank 13 Aug 2009 20:02
		99.9% of	memory fre	
variable name			value	variable label
capital continente ocde fiocde ue fiue superficie poblacion densidad evn tmi	byte byte float byte float double float float float int	%19s %7.0f %2.0f %d %5.0f %d %12.0fc %9.3fc %9.3fc %8.0fc %3.0f %8.0g		Pertenece a la OCDE Fecha de ingreso en la OCDE
tascrec pib rnbpc rnbppa	long long	%6.2f %12.0fc		Tasa de analitabetismo Tasa de crecimiento Producto interior bruto (mil \$) Renta per cápita (\$) Renta per cápita (poder de compra) * indicated variables have notes

#### 2.6.1. Etiquetas de variables y de valores

En Stata pueden asignarse etiquetas a la base de datos, a las variables y a los valores. Estas etiquetas harán más fácil la comprensión de los análisis estadísticos, por lo que es conveniente ponerlas.

La instrucción general para etiquetas es *label*, tras la cual se especifica qué es lo que se quiere etiquetar y la etiqueta. Para poner una etiqueta a la base de datos, ha de escribirse *label data* y la etiqueta que se desee:

label data "Indicadores de los países. Mundo (2000)"

Las etiquetas de las variables se ponen con la instrucción compuesta *label variable*:

label variable pib "Producto interior bruto"

Como puede deducirse, la etiqueta debe ir entrecomillada, obligatoriamente si tiene espacios en blanco.

Un poco más complejo es poner etiquetas a los valores. Las etiquetas de valores se definen por listas, y luego se asignan a las variables deseadas. Esto permite que se asigne una misma lista de etiquetas de valores a varias variables con iguales respuestas. Por ejemplo, es posible asignar al mismo tiempo etiquetas a los valores de un conjunto de preguntas que tengan las mismas posibilidades de respuesta, como muy de acuerdo, de acuerdo, en desacuerdo y muy en desacuerdo.

Lo primero que debe hacerse es definir una lista de etiquetas de valores, al que se denominará *conti* (por continente). Esto ha de hacerse del siguiente modo:

label define conti 1 "Europa" 2 "Asia" 3 "África" 4 "América" 5 "Oceanía"

Como puede apreciarse, tras la orden *label* va la especificación *define*, el nombre de la lista de etiquetas de valores y luego los valores seguidos por sus respectivas etiquetas. Una vez que se introduzca esta instrucción, la lista de etiquetas de valores quedará en memoria junto con los datos. Si se guardan estos, las listas de etiquetas quedarán también grabadas, de tal modo que se recuperan en sesiones subsiguientes de Stata.

Una vez definida una lista de etiquetas, se puede asignar a tantas variables como se desee, en cualquier momento. En este caso concreto, la instrucción es:

label values continente conti

Este procedimiento tiene importantes ventajas cuando hay varias variables con el mismo tipo de etiquetas, como sucede con las variables ocde y ue, ambas relacionadas con la pertenencia o no a estos organismos. Para etiquetar sus valores, primero se definen las etiquetas y después se asignan a cada una de las variables de este modo:

```
label define perte 0 "No" 1 "Sí"
label values ocde perte
label values ue perte
```

En cualquier momento, el usuario puede ver las listas de etiquetas que están definidas en un determinado conjunto disponible de datos. Para ello hay que utilizar la orden *labelbook*, o para un resultado más escueto *label list*, que sólo muestra códigos y etiquetas para cada lista.

```
label list
```

#### ILUSTRACIÓN 2.23. Lista de etiquetas

```
conti:

1 Europa
2 Asia
3 África
4 América
5 Oceanía
perte:

0 No
1 Sí
```

Otras instrucciones útiles para trabajar con etiquetas son *label drop* (que elimina las listas de etiquetas que se declaren expresamente) y *label save* (que guarda la definición de las etiquetas en el archivo .do que se indique). Tecleando *label values* seguido sólo por un nombre de variable, se quitarán las asignaciones que tuvieran sus valores a una lista de etiquetas.

```
label values ocde
```

Mediante la anterior instrucción, la etiqueta *perte* dejará de estar asignada a la variable ocde. La lista de etiquetas *perte* no se borra de la base de datos, sólo deja de estar asignada a ocde (seguirá asignada a la otra variable asignada, a ue). A menos que se escriba la orden específica para hacerlo (con *label drop*), si una etiqueta no está asignada a ninguna variable no desaparece del archivo, con lo que puede ser usada siempre que se requiera.

#### 2.6.2. Tipos de almacenamiento de las variables

En Stata, cada variable tiene un formato según el tipo de datos que contenga. Los valores de las variables pueden componerse de: a) una cadena de caracteres (string), b) números o c) fechas. Estos son los tres tipos principales de variables en Stata.

Las variables numéricas y de cadena no sólo contienen la información de su tipo, sino también el tamaño máximo de dígitos de la variable. En las variables de cadena la norma y el procedimiento son sencillos: automáticamente Stata asignará a cada variable el tipo *str* y el número de caracteres que contenga el conjunto de caracteres más largo. Por ejemplo, en una variable que incluyera las provincias españolas, la cadena de caracteres más larga sería Santa Cruz de Tenerife, que se escribe con 22 caracteres, por lo que Stata asignaría a la variable provincia la extensión de 22: el tipo sería *str*22.

Para las variables numéricas es algo más complicado, puesto que el tipo no depende directamente del número de dígitos sino del valor máximo: desde *byte*, que puede almacenar desde el valor –127 hasta el 126; hasta *double*, que puede almacenar desde el número –9.0<sup>238</sup> hasta el 9.0<sup>307</sup>. En el cuadro 2.2 pueden verse los distintos nombres y características de los tipos de variables numéricas que utiliza Stata.

CUADRO 2.2. Tipos de almacenamiento de variables numéricas<sup>20</sup>

Tipo variable	Valor mínimo	Valor máximo	Valor más cercano a 0 (sin ser 0)	Bytes
byte	-127	100	+/- 1	1
int	-32.767	32.740	+/- 1	2
long	-2.147.483.647	2.147.483.620	+/- 1	4
float	$-1,7 \times 10^{38}$	$1,7x10^{38}$	$+/-10^{-38}$	4
double	$-9,0\times10^{307}$	$9,0\times10^{307}$	+/- 10-323	8

Fuente: Stata Corporation (2011c: 110).

En principio, no es necesario preocuparse por el tipo de las variables, porque Stata asigna automáticamente el tipo adecuado, e incluso lo cambia si es necesario (si se introduce un valor mayor que el máximo). Por ejemplo, si se añade un caso con el valor 101 en una variable *byte* (que puede alma-

<sup>&</sup>lt;sup>20</sup> Los valores mínimos y máximos de las variables flotantes y dobles han sido reducidos a un decimal para simplificar su exposición. El valor exacto puede consultarse en la sección 12.2.3 de la guía del usuario.

cenar hasta el valor 100, véase el cuadro 2.2), automáticamente Stata cambia el formato de la variable y la convierte en int, que sí puede almacenar un valor mayor que la centena. Pero, aunque no sea necesario asignar directamente el formato a las variables, porque Stata ya lo hace automáticamente, puede ocurrir que los formatos sean demasiado grandes para los datos, por lo que estos ocupen demasiada memoria. Por ejemplo, en el supuesto de que a una variable de edad, para la que en principio el formato byte (que puede almacenar de -127 a 100) es más que suficiente, por un error en la introducción de datos, se introdujera el valor 195; automáticamente, Stata cambiaría el tipo de la variable a int. Si luego se advierte el error y se corrige (cambiando el valor a 19), el tipo de la variable seguirá siendo int, aunque los datos que almacena la variable edad no superen en ningún caso la centena (y por tanto el tipo byte es suficiente). Esto es así porque Stata modifica el formato al alza pero no a la baja: cuando se introduce un valor superior al máximo permitido por el tipo de variable, cambia el formato para que se pueda almacenar el valor correctamente, pero a menos que se especifique nunca cambia el formato, si se reduce el valor máximo. Esto es importante porque el formato de la variable determina cuánta memoria ocupan los datos. Si los formatos de los datos son mayores de lo necesario, puede que la matriz ocupe tanto que no quepa en la memoria de trabajo y, en consecuencia, su proceso se ralentiza por la necesidad de usar el disco duro como memoria virtual.

La instrucción *compress* está específicamente diseñada para este problema. Cuando se introduce esta orden, Stata comprueba uno a uno los formatos de todas las variables de la matriz y asigna a cada una de ellas el formato más pequeño posible. Es una instrucción que nunca modifica los contenidos de la matriz, sólo el tipo de las variables. Caso de que este ya sea tan pequeño como posible, no modificará nada. Pero, en muchos casos, este comando puede hacer mejorar ostensiblemente el funcionamiento de Stata, al reducir el tamaño que ocupan los datos en memoria.

Finalmente hay que conocer el peculiar modo con que Stata trata las fechas. Estas pueden ser una variable de texto: "21 Mar 1952", un conjunto de tres variables numéricas, 21 para el día, 3 para el mes y 1952 para el año, o una sola variable numérica, en cuyo caso se necesita una referencia, una fecha de partida que represente el valor 0, que en Stata es el 1 de enero de 1960. En cualquier caso, para que este programa las trate como variable de fecha, especialmente en los análisis de series temporales, sólo es válida la última forma de almacenamiento.

Resulta evidente que el usuario no va a introducir la variable de fecha según los días que haya transcurrido desde comienzos del año 1960. Para la conversión se dispone de muchas funciones que permiten tanto pasar del formato usual al modo de trabajo como a la inversa.

Una de las operaciones más empleadas en este sentido es la de proporcionar un determinado formato de presentación a una variable. De este modo, una variable con el valor numérico temporal de 2 puede aparecer literalmente como "3 january 1960", si se le indica dicho formato tal como se indica en el próximo apartado.

#### 2.6.3. Formatos de presentación de las variables

Como en la mayor parte de las aplicaciones informáticas, hay que distinguir entre el modo en el que son almacenados los valores de las variables que presentan los casos y el formato en el que son presentados en la pantalla. El primero está determinado por el tipo de almacenamiento, mientras que el segundo es el que es denominado formato de presentación.

Una instrucción para la que es importante la utilización de los formatos de presentación es *list*. Como ya se ha visto su función es la de mostrar los valores que tienen los casos en unas determinadas variables. El modo de cambiar la presentación de las variables es mediante otra orden anterior a la mencionada. Se trata de *format*, que ha de presentar la siguiente estructura:

format listavariables %formato

Donde aparece listavariables, ha de figurar una o varias variables mediante las convenciones propias del programa y donde aparece %formato se especifica mediante claves el aspecto con el que se desean mostrar las variables. Desde el punto de vista del formato también es útil distinguir los tres tipos de variables: las numéricas, las textuales y las de fecha.

1) Las *variables numéricas* pueden presentarse a su vez de acuerdo a una de las siguientes modalidades:

%p.dg, para mostrar todo tipo de formatos.

%p.df, para mostrar formatos de un número determinado de decimales.

%p.de, para mostrar los números en notación científica.

donde p significa el número de posiciones que se desean obtener de un determinado número y d expresa el número de decimales que se quieren mostrar. A todas ellas se le puede añadir una c, si se desea que se añada una coma cada tres dígitos para mejorar la legibilidad de las cifras largas. Por defecto, Stata emplea los puntos para expresar los decimales y las comas para los millares. Esto puede cambiarse con la instrucción set dp comma.

En realidad, sólo existen dos tipos de formatos para los números, el decimal (f) y el científico (e). El formato (g) hace que sea el mismo programa quien se encargue de seleccionar la prestación más adecuada según las características del número mostrado.

Para que quede mejor aclarado es imprescindible un buen ejemplo con distintos tipos de formato. De la base de datos de los países del mundo se han seleccionado las variables *superficie*, *pib*, *tmi* y *tascrec*. El formato respectivo de cada una de ellas es el siguiente: tanto en *superficie* como en *pib*, se han puesto doce caracteres sin ningún decimal. En ambas también se ha añadido el carácter "c" a fin de mejorar la legibilidad. En la tasa de mortalidad se ha fijado un formato de 3 posiciones sin ningún decimal (está medida en tantos por mil). Y, finalmente, la tasa de crecimiento se expresa en formato de seis posiciones y dos decimales. Toda esta información está guardada en el fichero, por lo que no es necesaria su introducción, a menos que se desee cambiar.

Un listado de los diez primeros países de las variables con los formatos antedichos puede solicitarse con la siguiente instrucción:

list pais superficie pib tmi tascrec in 1/10

Nótese que a la orden *list* se le ha añadido la lista de variables y la partícula *in* seguida de 1/10, lo que significa desde el caso primero hasta el décimo.

Ī	pais	superfi~e	pib	tmi	tascrec
1					
1.	Afganistán	652,090	4,100	163	2.60
2.	Albania	27,400	4,114	20	0.40
3.	Alemania	349,300	1,873,854	4	0.30
4.	Andorra	500	950		
5.	Angola	1,246,700	9,471	128	3.10
- 1					
6.	Antigua y Barbuda	400	640	16	1.40
7.	Antillas Holandesas	800	2,360	13	0.86
8.	Arabia Saudí	2,149,690	173,287	18	2.80
9.	Argelia	2,381,700	53,009	33	1.90
10.	Argentina	2,736,700	268,773	17	1.30

ILUSTRACIÓN 2.24. Listado parcial de casos

En cambio, si antes de realizar ese mismo listado se le hubiera cambiado el formato de un modo similar a éste:

format %8.0g superficie pib tmi tascrec

el resultado hubiera sido diferente en las variables superficie, pib y tascrec.

-	+				+
	pais	superf~e	pib	tmi	tascrec
1.	Afganistán	652090	4100	163	2.6
2.	Albania	27400	4114	20	.4
3.	Alemania	349300	1.9e+06	4	.3
4.	Andorra	500	950		.
5.	Angola	1.2e+06	9471	128	3.1
_					
6.		400	640	16	1.4
7.	Antillas Holandesas	800	2360	13	.86
8.	Arabia Saudí	2.1e+06	173287	18	2.8
9.	Argelia	2.4e+06	53009	33	1.9
10.	Argentina	2.7e+06	268773	17	1.3
	+				

ILUSTRACIÓN 2.25. Listado formateado de casos (I)

En estos resultados, en los que se ha aplicado a todas las variables numéricas el formato %8.0g²¹, se puede ver que hay casos de las variables superficie y pib —Angola en la primera y Alemania en la segunda, entre otros— que son mostrados en notación científica, pues de otro modo no cabrían en los 8 espacios. Por otro lado, en la variable de la tasa de crecimiento se detecta otra de las peculiaridades del tratamiento del formato que hace Stata. Es de notar cómo, a pesar de que estén puestas en formato de cero decimales, Stata los muestra todos siempre y cuando estén así almacenados, sin efectuar operación de redondeo como lo hubiera hecho en el caso de haber utilizado el formato %p.df de presentación. Por ello, en el caso de las Antillas Holandesas aparecen dos decimales. En el caso de que hubiera algún país sin decimales, obviamente no los mostraría.

2) Los formatos de las *variables textuales* también pueden ser cambiados. Pero, en lugar de utilizar los caracteres *f*, *g* o *e*, hay que emplear *s*, abreviatura de *string*. Obviamente, en este caso, no han de aparecer cifras decimales; en cambio, es muy útil utilizar la opción del alineamiento a la izquierda, que se logra mediante el signo menos delante de la cifra que indica el número de posiciones necesarias para la presentación del texto. De este modo, con las dos siguientes instrucciones, se mostraría los cinco primeros países acompañados de su correspondiente tasa de mortalidad infantil:

format %-24s pais list pais tmi in 1/5, clean

<sup>&</sup>lt;sup>21</sup> Por defecto Stata asigna el formato %8.0g a todas las variables almacenadas como *byte* o *integer*, con %9.0g a las variables *float*, %10.0g a las *double* y %12.0g a las *long*.

Es preciso notar cómo en los resultados de la ilustración 2.26, el tamaño de la columna de la variable *país* no contiene 24 columnas. A menos que se le especifique la opción *fast*, el programa examina la longitud de los casos que va a mostrar y automáticamente ajusta el tamaño de la columna al máximo de caracteres. Otra opción interesante de la orden *list* es *clean*, que hace que en el listado los casos no queden separados por líneas horizontales.

#### ILUSTRACIÓN 2.26. Listado formateado de casos (II)

	pais	tmi
1.	Afganistán	163
2.	Albania	20
3.	Alemania	4
4.	Andorra	
5.	Angola	128

3) Finalmente, hay que referirse a los complejos formatos de *variables de fecha (date)*. Como se dijo anteriormente, Stata almacena los datos relacionados con fechas como un número de tal modo que el 0 representa el 1 de enero de 1960. Un valor negativo es una fecha anterior a la mencionada y todo positivo, en correspondencia, posterior. Así, como dicho año fue bisiesto, el número 366 equivale al 1 de enero de 1961 y el número –365 representa el primer día del año 1959.

Todo ello es fácilmente apreciable utilizando la orden *display*, que muestra el contenido de una variable o constante con la posibilidad de aplicarle un formato temporal. De este modo, si se escribe la siguiente línea en la ventana de órdenes:

```
display %d -365, %d 0, %d 366
```

se mostrará en la pantalla de resultados tres fechas consecutivas correspondientes a los primeros días de los años 1959, 1960 y 1961.

#### ILUSTRACIÓN 2.27. Exposición de fechas (I)

```
01jan1959 01jan1960 01jan1961
```

Como puede apreciarse, las fechas se muestran en inglés con dos dígitos para el día, seguidos por tres caracteres para el mes y cuatro dígitos para el año. Esto también puede ser cambiado al especificar el formato %d seguido con una lista formada por una combinación de las siguientes convenciones.

c / C	Muestra el siglo sin/con ceros a la izquierda	
y / Y	Muestra los dos dígitos del año sin/con ceros a la izquierda	
m/M	Muestra el mes en mayúscula abreviado con 3 letras/sin abreviar	
1 / L	Muestra el mes en minúscula abreviado con 3 letras/sin abreviar	
n/N	Muestra el mes numéricamente sin/con ceros a la izquierda	
d / D	Muestra el día del mes sin/con ceros a la izquierda	
j / J	Muestra el día del año (1/366) sin/con ceros a la izquierda	
w/W	Muestra la semana (1/52) del año sin/con ceros a la izquierda	
,:-/'	Caracteres permitidos directamente en el formato de fecha	
!	Prefijo para introducir cualquier otro carácter	

CUADRO 2.3. Formatos de fecha

La misma instrucción anterior puede ser empleada con otros formatos para que la presentación de cada una de las fechas sea totalmente diferente. El próximo ejemplo muestra tres formatos distintos en los que puede presentarse los segundos días de los años 1959, 1960 y 1961<sup>22</sup>.

display %dd\_M\_cY -364, %dM/d/Y 1, %dd-n-CY 367

#### ILUSTRACIÓN 2.28. Exposición de fechas (II)

2 January 1959	January/2/60	2-1-1961

#### 2.6.4. El gestor de variables

Se podrían sintetizar las tres últimas secciones, explicando una de las novedades introducidas en la versión 11 de Stata. Se trata del gestor de variables, que puede ser llamado mediante menú (*Data/Variables manager*), icono (el décimo, Variables manager) e incluso desde la ventana de órdenes, mediante la instrucción *varmanage*.

Esta instrucción genera un cuadro de diálogo que contiene tantas líneas como variables se encuentran en la matriz de datos y seis columnas correspondientes al nombre de la variable, su etiqueta y las de sus valores, su tipo, su formato y las notas que el usuario desee incorporar a cada una de ellas.

<sup>&</sup>lt;sup>22</sup> Es preciso añadir que Stata también puede considerar las fechas semanal, mensual, trimestral, semestralmente e incluso fechas con hora incluida. En cualquier caso, siempre el punto de referencia es el 1 de enero de 1960 y la variable queda guardada como numérica; pero en estos casos el número, en lugar de días, significa semanas, meses, trimestres, semestres o milisegundos. Sin embargo, como este libro no trata de series temporales, se considera que no es útil explicar su uso. Se sugiere pedir ayuda en Stata mediante la instrucción *help dates*.

#### Variables Manager Enter filter text here Drag a column header here to group by that column %7 O ertenece a la OCDE %2.06 echa de ingreso en la OCD %dd\_ldlell\_n\_ld... nece a la UE 95.06 %dd\_ldlell\_n\_ld... %12.06 %9.3fc %8.0fc %3.01 %8.0g Tasa de analfabetismo %8.0g < > %6.25 oducto interior bruto (mil. %12.06 nta per cápita (\$) Vars: 17 | CAP | NUM

ILUSTRACIÓN 2.29. Ventana del gestor de variables

Las variables de esta lista pueden ser ordenadas por su posición en la matriz (#), por su nombre, el de la etiqueta o el de la lista de etiquetas de los valores. Del mismo modo, esto es, apretando en su cabecera, pueden quedar ordenadas por su tipo, formato o la posesión o no de notas. También las variables pueden ser clasificadas por cualquiera de sus características menos el nombre y su orden, siempre y cuando se arrastre su encabezamiento a la franja situada por encima de las cabeceras. Incluso, en el caso de contar con muchas variables, puede realizarse una selección de ellas, escribiendo los caracteres deseados en la casilla situada arriba a la izquierda inmediatamente debajo del marco de la ventana. Todos estos cambios mencionados en este párrafo afectan sólo al gestor, ya que la cantidad de variables y el orden en la matriz quedarán inalterados<sup>23</sup>.

Otra característica sobresaliente del gestor de variables es la posibilidad de trasladar el nombre o la lista de un subconjunto de variables a la ventana de órdenes, al editor de programas o donde se requieran. Apretando la tecla mayúscula o control al tiempo que se pulsa el botón izquierdo del ratón en las respectivas líneas de las variables, quedará activada más de una variable. Una vez que se han seleccionado las variables pertinentes, apretando el botón derecho del ratón aparecerá un menú contextual, cuya última línea hace que todos los nombres de las variables seleccionadas aparezcan en la ventana de órdenes. Alternativamente, si se selecciona la línea *Copy varlist* o se pulsan las teclas Ctrl+c, se almacenarán en el portapapeles y, de este modo, podrán pegarse allá donde se requiera con la combinación Ctrl+v.

<sup>&</sup>lt;sup>23</sup> Si se desea borrar variables de la matriz podrá hacerse mediante las opciones *keep* y *drop* del menú contextual que se obtiene pulsando el botón derecho del ratón sobre una selección de variables. Ambas opciones son también instrucciones que pueden ejecutarse en la ventana de órdenes para mantener o borrar una lista de variables.

El gestor de variables permite también introducir y editar las características de las variables. En la parte derecha, se encuentra una ventana interna que contiene nombre, etiqueta, tipo, formato, etiqueta de valores y notas de la variable seleccionada. Todas estas características pueden ser modificadas por el usuario, del mismo modo que también pueden efectuarse cambios desde la ventana interna de las propiedades de las variables, que aparece en la parte inferior izquierda de la pantalla por defecto de la versión 12 de Stata (véase la ilustración 2.4), sin necesidad de entrar en el gestor de variables, siempre y cuando aparezca abierto el candado situado debajo del título (*Properties*) de la ventana.

**Properties** 2 + + **■** Variables Name Label Pertenece a la OCDE Type byte Format %2.0f Value Label perte □ Notes <Add note> □ Data Filename mundo.dta Label Indicadores de los paises. Mu **⊕** Notes Variables Observations 213 Size 19.34K Memory 32M

ILUSTRACIÓN 2.30. Ventana interna de las propiedades de las variables

#### 2.7. Ejercicios

- 1) Familiarízate con los ficheros de ejemplo con los que cuenta Stata mediante la instrucción *sysuse*. Mira qué variables contiene, el modo en que están grabadas, su formato de presentación y las etiquetas que contiene. Finalmente haz un listado de los diez primeros casos. (Ficheros propuestos: census, citytemp, educ99gdp, gnp96, lifeexp, pop2000, uslifeexp, voter).
- 2) Copia los ficheros que se proporcionan con este libro a un directorio. Arranca Stata desde ese directorio (o al menos, una vez dentro de Stata, escribe la instrucción cd "directorio"). Lista todos los ficheros que sean del tipo \*.dta. Ábrelos e inspecciona también las variables, su formato y etiquetas para terminar haciendo examen de ellos con la instrucción *browse*.
- 3) En el fichero que prefieras de los dos ejercicios anteriores, cambia las etiquetas del fichero, de las variables y de los valores. Por ejem-

plo, todos los ficheros del primer ejercicio tienen etiquetas en inglés, coge uno de ellos y reemplázalas por otro idioma. Si quieres conservar los cambios, no olvides terminar con la instrucción *save*, *replace*, que se explica en el capítulo 3.