

Escuela de Verano Métodos Mixtos



IV Versión de Escuela de Verano en Métodos Mixtos

Apuntes Básicos de Stata

Benjamín Muñoz

Santiago de Chile

05 de Enero de 2016

Índice

1. Introducción	2
2. Stata: Lo básico	2
2.1. Interfaz de Stata	3
2.2. Tipos de Archivos	3
2.3. La Hoja de Cálculo	4
2.4. Estructura de Comandos	4
2.5. Expresiones Lógicas	5
2.6. Stata como Calculadora	6
2.7. Ayuda	6
3. Manejo de Datos	7
4. Análisis Descriptivo	7
4.1. Gráficos	7
5. Inferencia Estadística Básica	8
6. Referencias para Trabajo Posterior	8
7. Listado de Comandos	8

1. Introducción

La presente guía fue diseñada como parte del Taller de Stata, actividad optativa de la Escuela de Verano en Métodos Mixtos. Dicha asignatura es de carácter aplicado y será dictada íntegramente en el laboratorio computacional del Instituto de Ciencia Política, localizado junto al Auditorio en que se desarrollan las actividades de la Escuela de Verano.

La clase tiene una estructura eminentemente práctica, dónde el profesor ejemplifica el uso de distintos comandos de Stata, mientras que los alumnos van replicando estos pasos. Para lograr que los estudiantes dominen los contenidos de la clase, se desarrollarán actividades prácticas de manera individual.

La presente guía reseña los aspectos introductorios del uso de Stata, cubriendo todas las etapas del trabajo empírico. En este material se resumen los principales comandos, incluyendo una explicación para su uso adecuado. **El objetivo fundamental es que los estudiantes sean capaces de implementar una rutina de trabajo estadística sencilla de manera autónoma.** Dentro de este marco, se tienen como objetivos específicos:

1. Introducirse a la lógica de operación de Stata.
2. Comprender la estructura general de los comandos de Stata.
3. Utilizar la documentación elaborada por Stata y los recursos disponibles en la web para resolver dudas y trabajar autónomamente.
4. Aprender los aspectos básicos del manejo de datos en Stata, incluyendo la importación y exportación de archivos.
5. Realizar operaciones estadísticas básicas con Stata.
6. Generar gráficos y tablas con Stata.
7. Comprender los aspectos básicos del trabajo reproducible.

La presente guía tiene la siguiente estructura.

2. Stata: Lo básico

Stata es un paquete de software estadístico que permite manejar y analizar datos de distinta naturaleza. Su nombre proviene de la combinación de las palabras statistics y data. El programa es elaborado por Stata Corp, y dicha empresa se encarga del lanzamiento de las actualizaciones oficiales. Periódicamente se lanzan nuevas versiones de Stata, la más reciente es la número 14. Los cambios entre versiones se centran en aspectos avanzados del

trabajo estadístico y en el manejo de bases de datos.

2.1. Interfaz de Stata

Al iniciar Stata se abre una pantalla, la que se subdivide en distintos paneles¹:

1. **Command** (abajo): Permite ingresar los comandos. Es la interfaz de línea de comandos.
2. **Results** (centro): Registra todos los comandos ejecutados y resultados obtenidos. Las notas aparecen en azul, los errores en rojo y todo el resto en negro. No es posible borrar nada del panel central.
3. **Review** (izquierda): Registra los comandos ingresados durante una sesión de Stata y permite ir construyendo un código automatizado. Los errores aparecen en rojo (con un código breve) y pueden ser borrados.
4. **Variables** (derecha, arriba): Describe las variables de la base de datos abierta.
5. **Properties** (derecha, abajo): Describe la base de datos abierta.

En la parte superior aparece la “Barra de Herramientas”. Esta se compone de dos aspectos. La barra de Herramientas propiamente tal y los íconos de atajo a funciones. Prácticamente todas las funciones de Stata pueden ser utilizadas por medio del Menú de la Barra de Herramientas. Las funciones estadísticas y de gestión de datos se encuentran en “Data”, “Graphics” y “Statistics”.

Las versiones recientes de Stata incorporan una Interfaz gráfica de usuario (GUI por sus siglas en inglés) donde se presentan los gráficos generados en Stata. En dicha ventana se accede al Editor de gráficos. También existe una ventana específica para examinar visualmente la base (fichero) de datos.

2.2. Tipos de Archivos

Log-files: es un tipo de archivos que permite almacenar los comandos y los resultados obtenidos en una sesión de Stata.

Do-file: es un tipo de archivo orientado al almacenamiento de comandos (código de sintaxis) para su uso posterior.

This facility allows you to print or save all commands you used for a session with STATA

¹Los cuales pueden ajustar y reubicar según sus preferencias.

Sin importar el paquete estadístico utilizado, una buena práctica de trabajo es la reproducibilidad y transparencia de los procedimientos. Es fundamental elaborar esquemas de trabajos que permitan mantener un control de los cambios/procesos realizados.

No resulta recomendable guardar todos los cambios introducidos en la base de datos. **Es una buena práctica almacenar una versión bruta de los datos, un archivo de procesamiento de datos (do file) y una versión " limpia" de la base de datos.**

2.3. La Hoja de Cálculo

Stata tiene un fichero (base) de datos activo a la vez, el cual puede ser observado por medio de la hoja de cálculo. Los datos son almacenados en filas y columnas, donde cada celda corresponde a un valor o dato específico. La estructura más frecuente de las bases de datos consisten en que cada fila corresponde a una unidad de observación y cada columna a una característica (variable) observada-.

A la hoja de cálculo se puede acceder por medio del Menú en la parte superior de la pantalla o por medio de los comandos:

- *browse* Abre la Hoja de cálculo (nueva ventana) para examinar los datos.
- *edit* Permite editar manualmente la información contenida en la hoja de cálculo.

No se pueden ejecutar comandos si la hoja de cálculo está abierta. En la hoja de cálculo se pueden realizar cambios en los valores almacenados, los nombres de las variables y las etiquetas utilizadas. **No se recomienda realizar cambios por este mecanismo, ya que es más fácil cometer errores.** Es preferible operar por medio de comandos en la edición de datos.

2.4. Estructura de Comandos

Stata es un software que puede ser utilizado por Menú como por medio del uso de la Interfaz de línea de comandos. Es preferible utilizar la interfaz de comandos para el trabajo empírico, por lo que dicho método de trabajo será usado en el Taller. **Un comando es una instrucción estructurada del usuario que le indica que procedimiento (o conjunto de procedimientos) desea implementar.** La estructura típica de los comandos está compuesta de cuatro elementos, cada uno de los cuales correspondientes a un tipo de pregunta particular:

1. **Comando:** ¿Qué quiero hacer?
2. **Nombres:** ¿Cuáles son los nombres de las variables o archivos que quiero ocupar?

3. **Condiciones:** ¿Qué observaciones deseas ocupar?

4. **Opciones:** ¿Tienes otros requisitos respecto a este programa?

La estructuración de las instrucciones en comandos siguen ciertas reglas de sintaxis. Se recomienda al usuario siempre comenzar por especificar con sus propias palabras que desea hacer y luego determinar el comando adecuado para Stata. Cada comando se debe ingresar en una línea separada, salvo que se esté trabajando en un archivo .do y se utilice /// al final de la línea para indicar que el comando sigue en la línea siguiente.

De manera genérica², la sintaxis de un comando tiene la siguiente estructura:

[by varlist:] command [variable/s] [if exp] [in rango] [,opciones]

Los paréntesis cuadrados (brackets) indican especificaciones opcionales. La especificación *[by varlist]* alude a la repetición del comando indicado sobre una lista de variables. Luego, *[variable/s]* corresponde al **nombre** de la o las variables sobre las cuales se aplica el comando.

Luego, se indican expresiones para aplicar el comando sobre un subconjunto de observaciones de la base de datos. Por un lado, *exp* se refiere a una expresión lógica, la cual se combina con el operador condicional *if*, mientras que *rango* corresponde a un rango de observaciones indicados numéricamente. En el primer caso se usan reglas lógicas para seleccionar las observaciones que cumplen una o más observaciones, en el segundo se utiliza la posición de las observaciones (número de fila) para la selección. Por último, *, opciones* indica argumentos adicionales al comando para activar o desactivar determinadas funciones. Deben especificarse después de una coma.

Todo lo que se ingrese en la Interfaz de líneas de comandos será interpretado por Stata como una instrucción salvo que se indique lo contrario. Para lograr esto es necesario señalizar que se ingresarán comentarios (textos que Stata ignorará).

- **comentario* el asterisco indica el inicio de un comentario, los cuales Stata no interpretará como instrucciones.
- **comentario** señala el comienzo y final de un comentario (solamente en archivo .do)

2.5. Expresiones Lógicas

Si se usan la opción *[if exp]* al especificar un comando, es necesario conocer la sintaxis de las expresiones lógicas en Stata:

²Hay comandos que tienen una estructura distinta.

Cuadro 1: Expresiones Lógicas en Stata

Símbolo	Descripción
==	Igual a
~=	No igual a
>=	Mayor o igual a
<=	Menor o igual a
>	Mayor a
<	Menor a
~	Negación
!	Negación
&	Y
	O

2.6. Stata como Calculadora

Stata puede ser utilizado para cálculos aritméticos sencillos. Sin embargo, es necesario utilizar el comando *display*, el cual permite imprimir los valores de variables de cadena y/o numéricas.

Cuadro 2: Operaciones Aritméticas y Funciones Básicas en Stata

Símbolo	Descripción
+	Adición
-	Sustracción
*	Producto
/	Cociente
	Potencia
exp()	Exponenciación
ln()	Logaritmo Natural
abs()	Valor Absoluto
sqr()	Raíz Cuadrada

Los operadores + y * pueden utilizarse con variables de cadena. Mientras + permite concatenar expresiones, * se usa para repetir una expresión de cadena un determinado número de veces (0 o más).

2.7. Ayuda

El comando *help* es el instrumento por defecto para utilizar la documentación de Stata. Sin embargo, no siempre resulta sencillo entender dichos documentos de ayuda. Se abrirá

una nueva ventana con dicha información. El esquema general del comando es:

- *help comando* Como argumento de nombre se debe indicar algún comando de Stata.

Ejemplo Solicite ayuda para utilizar el comando generate:

```
help generate
```

3. Manejo de Datos

Stata contiene un conjunto de bases de datos como ejemplos. Se puede acceder a ellos por medio del Menú en Archivo, Bases de Datos de Ejemplo. Otra alternativa es el comando: *sysuse dir*. Se listarán todas las bases de datos de ejemplo³. Para obtener más información sobre una de dichas bases se utiliza el comando *sysdescribe* que da las principales características del fichero de datos y las variables que lo componen. Por último, el comando *sysuse nombre* activará una de las bases de datos.

Ejemplo Ingrese los siguientes comandos:

```
sysuse dir  
sysdescribe census  
sysuse census
```

4. Análisis Descriptivo

4.1. Gráficos

Tal como se mencionó, Stata incorpora una Interfaz Gráficos de Usuario (GUI), el cual se materializa en una ventana separada para los gráficos.

Graph Editor: When you draw a graph in Stata, it opens in a separate window known as the Graph window. From there you can print the graph, copy it to the clipboard, and export it to formats such as PDF, EPS, PNG, and TIFF. You can click on a button which transforms the Graph window into the Graph Editor which gives you full point-and-click access to edit every aspect of your graphs. You can even ?record? the edits you make and apply them to another graph, or to hundreds of other graphs. Read more [here](#) and [here](#).

³Si se han utilizado otras bases de datos en el sistema también aparecerán listadas.

5. Listado de Comandos

Cuadro 3: Ficha de Comandos de Uso Frecuente

Comando	Descripción
browse	
clear	
collapse	
correlate	
count	
describe	
display	
drop	
edit	
keep	
generate	
label	
list	Lista el valor de las variables
merge	
regress	
rename	
sort	
summarize	
tabulate	

Fuente: Elaboración propia.