

Guía Rápida Stata

1. Elementos básicos (obtener ayuda, salir)

Stata presenta cuatro ventanas llamadas *Review*, *Variables*, *Stata Results* y *Stata command*. Stata se maneja con comandos, los cuales se tipean en la ventana *Stata command*, y los resultados aparecen en la ventana *Stata Results*. La ventana *Review* acumula comandos anteriores, los cuales pueden ser reenviados a la ventana de *Stata command* clickeando sobre ellos.

Para obtener **ayuda**, clickee en el menú la opción **Help** y siga las instrucciones en pantalla. Para terminar la sesión, cierre el programa clickeando en el botón superior derecho (como cualquier otro programa Windows).

2. Lectura de datos

La estructura básica de datos en Stata es una tabla, en donde cada columna representa a una variable y cada fila a una observación. Si bien existen numerosas formas de incorporar datos a Stata, en esta nota discutiremos la importación de datos desde una planilla de cálculo tipo Excel. Supongamos que los datos se encuentran en un archivo Excel (de planilla de cálculo) llamado *mydata.xls*, en el cual en la primer fila se encuentran los nombres de las variables y en las restantes las observaciones, de modo que cada columna corresponde a una variable. Discutiremos dos métodos de importación de datos desde Excel.

a) Utilizando Copy & Paste

Abrir simultáneamente Excel y Stata. Una vez en Excel, pintar todo el rango de datos que se desea llevar a Stata (incluyendo los nombres de las variables). Clickear **Edición/Copiar**. Pasar a Stata y abrir el editor de Stata clickeando en el botón Editor. En esta ventana, clickear **Edit/Paste**, tras lo cual los datos aparecerán en pantalla. Luego cerrar la ventana del editor. Los nombres de las variables aparecerán en la ventana Variables. Este método puede tener problemas si no se dispone de suficiente memoria RAM, por lo que es útil chequear si todos los datos han sido importado correctamente. El método siguiente no debería presentar problemas de memoria

b) Utilizando insheet

Primeramente, hay que abrir el archivo conteniendo los datos en Excel y proceder a grabarlo nuevamente con la secuencia

```
Archivo, Guardar como, CSV (delimitado por comas), Guardar
```

Tras lo cual se guardara una copia en formato ASCII, que contiene las variables separando las columnas con comas. Supongamos que este archivo se llama *mydata.csv* y que se encuentra en el subdirectorio `c:\misdatos\mydata.csv`

Para leer los datos desde Stata, ingresar el siguiente comando:

```
insheet using c:\misdatos\mydata.csv, comma
```

Tras lo cual, si todo sale bien, Stata creará una nueva base de datos, y los nombres de las variables aparecerán en la ventana *Variables*.

Si se desea listar los datos, el comando es

```
list
```

Alternativamente, se pueden listar un subgrupo de variables. Por ejemplo, si la base de datos contiene las variables *y*, *gre*, *prom*, y *toefl*, un listado de estas dos últimas puede obtenerse con

```
list prom toefl
```

Para interrumpir el listado, clickear el botón *break* en el menú

3. Estadísticas básicas y regresiones

El siguiente comando produce una serie de **medidas estadísticas básicas** para cada una de las variables

```
summarize
```

Para estimar los parámetros de un modelo lineal por el método de **mínimos cuadrados ordinarios**, el comando es:

```
regress y gre prom toefl
```

Este comando producirá estimaciones mínimo-cuadráticas de un modelo en donde *y* es la variable dependiente y las restantes son las variables explicativas. El modelo incluye automáticamente una constante.

Los modelos **probit** y **logit** pueden ser estimados con

```
logit y gre prom toefl
```

o

```
probit y gre prom toefl
```

para lo cual es necesario que la variable dependiente sea un indicador binario

4. Archivos de salida y lectura de comandos desde un archivo

A veces es útil guardar resultados en un archivo de texto (*log file*) que puede ser editado con un procesador de textos común. Para ello, clickear en el menú el botón *Log*, tras lo cual aparecerá una ventana en la cual se debe ingresar un nombre para este archivo. Esto creará una nueva

pantalla en donde se guardaran los resultados. A partir de esto, los resultados apareceran en pantalla y serán guardados en este archivo. Para volver a la pantalla de resultados, clickear en el marco de la ventana de Stata results. Para ver el contenido de log, clickear *Log/Bring log window to top*. Para interrumpir momentáneamente el envío de resultados al log, clickear *Log/Suspend log file*, y para reanudarlo, clickear *Log/Resume suspended log file*. Para cerrar el log file, clickear *Log/Close log file*. El log file es un archivo ASCII que puede ser editado con cualquier procesador o editor (tipo Word, Wordpad o Notepad).

Otra práctica altamente recomendable consiste en armar *do files*. Estos no son más que archivos de texto que se guardan con extensión “do”, y que contienen una lista de comandos de Stata. Esto permite ejecutar una sucesión de comandos Stata de una sola vez. La elaboración de archivos “do” es recomendable por dos razones. Primero, permite construir programas y rutinas elementales, que pueden ser reutilizadas en diversas situaciones o con otras bases de datos. Segundo, proveen una forma eficiente de progresar en el análisis de datos, ya que es fácil corregir errores o reproducir resultados obtenidos con anterioridad

5. Guardar datos

Si se desea, se pueden guardar los datos en formato Stata. Para esto clickear

File/Save

tras lo cual (luego de elegir un nombre para el archivo) los datos son guardados en formato Stata, y no es necesario volver a realizar la conversión desde Excel al comienzo de una nueva sesión. Para iniciar nuevamente la sesión con los datos en formato Stata, luego de abrir Stata, clickear

File/Open

y elegir la base de datos deseada.

6. Donde encontrar mas información

Stata es un paquete estadístico muy extenso y esto es solo una breve introducción. Estas son algunas recomendaciones para aquellos interesados en explorar el software a fondo:

- Los *manuales* originales, aunque muy voluminosos, son una excelente fuente de referencia, con numerosos ejemplos. Uno de ellos (*Getting Started*) es recomendable para aquellos que desean una breve introducción. El resto es útil tenerlos solo como referencia y acudir solo si es necesario.
- Stata mantiene un muy buen *web-site* con abundante información sobre el producto, para usuarios iniciados y avanzados. El mismo esta en:

<http://www.stata.com>

- Stata dispone de una serie de *tutoriales on-line*. En la ventana de comandos, tipear `tutorial` para ver una lista de tutoriales disponibles. Se recomienda altamente trabajar con estos tutoriales.

c) Datos para las preguntas del práctico

1) Estadísticas básicas para la muestra relevante

Utilizaremos el comando `sum` para obtener estadísticas descriptivas de las variables. Esto es:

```
sum ln_tma94 n_civ_soltero n_sexo n_edad n_maxeduc_mp n_trabv1 n_trabv2  
n_comercial n_sec_pri n_reside n_cambio_reside n_h_estd
```

Nota: este comando ocupa dos líneas, pero en Stata se debe ingresar en una línea sola.

2) Relación horas de estudio performance

Utilizaremos el comando `reg` para llevar a cabo la regresión pedida en el punto 2.

```
reg ln_tma94 n_civ_soltero n_sexo n_edad n_maxeduc_mp n_trabv1 n_trabv2  
n_comercial n_sec_pri n_reside n_cambio_reside n_h_estd n_h_estd2
```

La relación entre el logaritmo natural de la performance y las horas de estudio se puede ver gráficamente con:

```
graph ln_tma94 n_h_estd
```

3) Algunos tests

El comando `test` permite evaluar hipótesis lineales sobre el último modelo de regresión estimado. En algunas ocasiones, es conveniente reestimar el modelo sobre el cual se desean implementar los tests, y luego implementar los tests.

```
test n_h_estd2  
test n_trabv1 n_trabv2
```

La primera línea evalúa la hipótesis nula de que `n_h_estd2` es no significativa. Mientras que la segunda línea testea la hipótesis conjunta de que `n_trabv1` y `n_trabv2` son no significativas.

Do file para el práctico 1

Como dijéramos anteriormente, es muy importante crear archivos “do” que permitan documentar todo lo que se hace en Stata. A continuación se presenta un ejemplo de archivo “do” que permite reproducir todos los resultados del práctico 1.

Un consejo útil para la elaboración de estos archivos es agregar abundantes comentarios. Esto se hace encerrando el texto con los símbolos `/*` y `*/` (Stata ignora el contenido entre dichos símbolos) o comenzando la línea con `*`.

/* Econometria. Walter Sosa Escudero

Do file para el Trabajo Practico No.1: PERFORMANCE DE ESTUDIANTES

=====

Primera versión: 28 de Febrero de 2007

DATOS: Sosa Escudero, Porto y Giovagnoli (2007).

La fuente es el archivo rendimientouba.dta

El trabajo consiste en reestimar algunos resultados en Sosa Escudero, Porto y Giovagnoli (2007). Ver guía del trabajo practico para mas detalles.

*/

* I) Preguntas del Trabajo Practico No.1

* 1) estadísticas basicas para la muestra relevante
sum n_tma94 n_civ_soltero n_sexo n_edad n_maxeduc_mp n_trabv1 n_trabv2
n_comercial n_sec_pri n_reside n_cambio_reside n_h_estd

* 2) Relación perfomance - horas de estudio

*Genero variables nuevas
gen ln_tma94 = log(n_tma94)
gen n_h_estd2 =n_h_estd^2

*Modelo de regresion
reg ln_tma94 n_civ_soltero n_sexo n_edad n_maxeduc_mp n_trabv1 n_trabv2
n_comercial n_sec_pri n_reside n_cambio_reside n_h_estd n_h_estd2

* 4) Algunos tests

*a) Se puede utilizar un t-test
test n_h_estd2

*b) Debe ser un joint test
test n_trabv1 = n_trabv2 =0