

# Guia Rapida Stata 5.0

## 1. Elementos basicos (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 menu la opcion **Help** y siga las instrucciones en pantalla. Para terminar la sesion, cierre el programa clickeando en el boton superior derecho (como cualquier otro programa Windows).

## 2. Lectura de datos

La estructura basica de datos en Stata es una tabla, en donde cada columna representa a una variable y cada fila a una observacion. Si bien existen numerosas formas de incorporar datos a Stata, en esta nota discutiremos la importacion de datos desde una planilla de calculo tipo Excel. Supongamos que los datos se encuentran en un archivo excel (de planilla de calculo) 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 2 metodos de importacion de datos desde Excel.

### a) Utilizando Copy & Paste

Abrir simultaneamente 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 **Edicion/Copiar**. Pasar a Stata y abrir el editor de Stata clickeando en el boton Editor. En esta ventana, clickear **Edit/Paste**, tras lo cual los datos apareceran en pantalla. Luego cerrar la ventana del editor. Los nombres de las variables apareceran en la ventana Variables. Este metodo puede tener problemas si no se dispone de suficiente memoria RAM, por lo que es util chequear si todos los datos han sido importado correctamente. El metodo siguiente no deberia 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 creara una nueva base de datos, y los nombres de las variables apareceran 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 ultimas puede obtenerse con

```
list prom toefl
```

Para interrumpir el listado, clicar el boton *break* en el menu

### 3. Estadisticas basicas y regresiones

El siguiente comando produce una serie de **medidas estadisticas basicas** para cada una de las variables

```
summarize
```

Para estimar los parametros de un modelo lineal por el metodo de **minimos cuadrados ordinarios**, el comando es:

```
regress y gre prom toefl
```

Este comando producira estimaciones minimo-cuadraticas de un model en donde *y* es la variable dependiente y las restantes son las variables explicativas. El modelo incluye automaticamente 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 util guardar resultados en un archivo de texto (*log file*) que puede ser editado con un procesador de textos comun. Para ello, clicar en el menu el boton *Log*, tras lo cual aparecera una ventana en la cual se debe ingresar un nombre para este archivo. Esto creara una nueva

pantalla en donde se guardaran los resultados. A partir de esto, los resultados apareceran en pantalla y seran 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 momentaneamente el envio 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 mas que archivos de texto que se guardan con extension “do”, y que contienen una lista de comandos de Stata. Esto permite ejecutar una sucesion 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 conversion desde Excel al comienzo de una nueva sesion. Para iniciar nuevamente la sesion con los datos en formato Stata, luego de abrir Stata, clickear

File/Open

y elegir la base de datos deseada.

## 6. Donde encontrar mas informacion

Stata es un paquete estadistico muy extenso y esto es solo una breve introduccion. 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 introduccion. El resto es util tenerlos solo como referencia y acudir solo si es necesario.
- Stata mantiene un muy buen *web-site* con abundante informacion 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.

## Guia para la elaboracion del trabajo practico No.1 en Stata 5.0

A continuación se presentan todos los detalles y comandos necesarios para obtener todos los resultados para la elaboración del Trabajo Practico No.1 (Retornos a la educación).

*Cuidado:* esto no es la solución del trabajo practico. La idea es que lean detenidamente cada comando, entendiendo que hace y como se puede volver a utilizar en otras situaciones. Les queda a Uds la parte mas dificil e interesante que es la interpretacion de los resultados. Un consejo: no empiecen a trabajar con la computadora si no leyeron detenidamente el trabajo practico. Entiendan exactamente lo que hay que hacer, diseñen una estrategia econometrica para la estimación y recién ahí empiecen a trabajar con la computadora.

### a) Lectura de datos y descripcion de los mismos

El primer paso consiste en leer los datos. Supondremos que ya estan convertidos al formato csv (ver Seccion 2) y que el archivo se llama EPHOct97.csv y que se encuentra en un subdirectorio llamado c:\data. Como se explicara anteriormente, se puede utilizar cualquiera de los dos metodos descriptos para leer los datos. Si todo sale bien, aparecera una lista de variables. Para obtener una descripcion basica de las variables utilizamos el comando

```
describe
```

y para obtener estadisticas basicas usamos

```
summarize
```

### b) Creación de variables nuevas

Luego de leer los datos, el primer paso consiste en elaborar variables nuevas para ser utilizadas en el analisis. Primero generamos las variables binarias correspondientes a los maximos niveles de educacion alcanzado por el individuo. El comando `generate` (abreviado `gen`) genera variables nuevas. El comando `replace` altera valores. La instrucción `if` funciona como operador logico, lo mismo que las instrucciones “==” y “|”. La mejor forma de entender como funcionan estos operadores es con un ejemplo. Supongamos que queremos crear una variable binaria que tome valor 1 si el individuo no tiene educacion o tiene como nivel educativo maximo el primario incompleto. Este nivel educativo corresponde a los codigos 10 (no educacion) y 12 (primario incompleto) de la variable `nivel` (en la EPH). La estrategia es la siguiente. Primero generamos una variable nueva (`edupi`) que toma todos valores cero. Luego hacemos que la variable tome valor igual a 1 si el individuo no tiene educacion o tiene educacion primaria incompleta. Los dos siguientes comandos hacen este trabajo

```
gen edupi=0
replace edupi=1 if nivel==10 | nivel==12
```

Primero, es importante notar la diferencia entre “=” y “==”. El operador “=” asigna valores a una variable (en el primer caso, por ejemplo, el valor 0 a todas las observaciones de la variable `edupi`).

El operador “==” compara que dos expresiones sean iguales (en la segunda línea, por ejemplo, que el nivel de educación de un individuo sea igual a 10 o a 12 años)

De esta manera es posible construir variables binarias para todos los niveles educativos:

```
gen edup=0
replace edup=1 if nivel==11
gen edusi=0
replace edusi=1 if nivel==22 | nivel==42
gen edus=0
replace edus=1 if nivel==21 | nivel==41
gen eduui=0
replace eduui=1 if nivel==32
gen eduu=0
replace eduu=1 if nivel==31
```

Luego creamos una variable (yearse) que contenga los años de educación para cada individuo computados según Pessino, 1995, pp. 23). A tal efecto, primero creamos yearse con valores iguales a 3 años (no educación o primaria incompleta) y luego modificamos de acuerdo a los otros niveles educacionales.

```
gen yearse=3
replace yearse=7 if edup==1
replace yearse=9.5 if edusi==1
replace yearse=12 if edus==1
replace yearse=14.5 if eduui
replace yearse=17 if eduu==1
```

Para el análisis es necesario utilizar algunas variables modificadas. En algunos casos solo cambiaremos el nombre para compatibilizarlo con los de Pessino. También utilizaremos operadores de multiplicación (\*)

```
gen wage = inghora
gen lwage = log(inghora)
gen age = edad
gen agesq = edad*edad
gen exper = age - yearse - 6
gen exper2 = exper*exper
```

La variable sexo toma valores 1 para hombres y 2 para mujeres. Si bien no es crucialmente importante, crearemos una variable nueva (sexo01) que toma valores 0 para hombres y 1 para mujeres

```
gen sexo01 = sexo-1
```

Por razones que serán evidentes más adelante, es conveniente crear una variable binaria que tome valores 1 si el individuo pertenece a la muestra a estudiar y 0 si no. Recordemos que el análisis se basa en hombres con ingreso positivo con edades entre 24 y 54 años. La estrategia para construir esta variable es similar a la anterior.

```

gen muest=0
replace muest=1 if edad>24
replace muest=0 if edad>54
replace muest=0 if sexo01==1
replace muest=0 if inghora==0
replace muest=0 if wage==.

```

La ultima linea requiere explicacion. Las observaciones faltantes se denotan en Stata con “.”. Entonces, la ultima linea indica que no deben entrar en la muestra individuos que no hayan declarado salario.

Esto completa la creación de variables nuevas. En este punto, es recomendable guardar los datos siguiendo las indicaciones de la sección 5.

### c) Datos para las preguntas del práctico

#### 1) Estadísticas básicas para la muestra relevante

Utilizaremos el comando `summarize`, pero restringido a las observaciones que entran en el análisis. Esto es:

```

summarize edupi edup edusi edus eduui eduu yearse age exper wage lwage
if muest==1

```

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

#### 2) Relación ingresos edad

El comando `table` crea tablas de datos. En este caso, el comando muestra el promedio del ingreso horario (`inghora`) para cada valor de la variable `edad`, restringiendo la muestra de acuerdo a la variable `muest`:

```

table edad if muest==1, c(mean inghora)

```

La relación edad salario se puede ver gráficamente con

```

graph lwage edad if muest==1

```

#### 3) Estimación de Ecuaciones de Mincer (corresponde a Pessino (1995, pp. 31 columnas 1 a 4))

Utilizaremos el comando `regress` descrito en la sección 3

```

regress lwage yearse exper exper2 if muest==1
regress lwage yearse age agesq if muest==1
regress lwage edup edusi edus eduui eduu exper exper2 if muest==1
regress lwage edup edusi edus eduui eduu age agesq if muest==1

```

#### 4) Algunos tests

El comando `test` permite evaluar hipotesis lineales sobre el ultimo modelo de regresion estimado. Es conveniente reestimar el modelo sobre el cual se desean implementar los tests, y luego implementar los tests.

```
regress lwage yearse exper exper2 if muest==1
test exper exper2
test exper2
```

La primer linea testea la hipotesis conjunta de que `exper` y `exper2` son no significativas. La segunda linea evalua la hipotesis nula de que `exper2` es no significativa.

### **Do file para el practico 1**

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

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

```
/* Econometria. Walter Sosa Escudero
```

```
Do file para el Trabajo Practico No.1: RETORNOS A LA EDUCACION
```

```
=====
```

```
Primera version: 2 de Febrero de 1998
```

```
DATOS: Encuesta Permanente de Hogares (EPH), Octubre de 1997, Gran Buenos Aires.
```

```
La fuente es el archivo EPHOct97.csv (Comma separated ASCII file)
```

```
Los datos se pueden leer con el comando:
```

```
insheet using c:\down\EPHOct97.csv,comma
```

```
o con Copy/Paste desde Excel.
```

```
Para una descripcion de las variables ver la "Presentacion de Base Usuarios" elaborada por el INDEC y los formularios personal y de familias.
```

```
El trabajo consiste en reestimar algunos resultados en Pessino (1995). Ver guia del trabajo practico para mas detalles.
```

```
*/
```

```
* Descripcion de variables (luego de haber leido los datos)
```

```
describe
```

```
summarize
```

```
* I) Generacion de variables nuevas
```

```
* -----
```

```
* a) Variables Dummy para educacion
```

```
gen edupi=0
```

```
replace edupi=1 if nivel==10 | nivel==12 /* No educacion o primario incompleto */
```

```
gen edup=0
```

```
replace edup=1 if nivel==11 /* Primario completo */
```

```
gen edusi=0
```

```
replace edusi=1 if nivel==22 | nivel==42 /* Secundario incompleto */
```

```
gen edus=0
```

```
replace edus=1 if nivel==21 | nivel==41 /* Secundario completo */
```

```
gen eduui=0
```

```
replace eduui=1 if nivel==32 /* Universitario o superior incompleto */
```

```
gen eduu=0
```

```
replace eduu=1 if nivel==31 /* Universitario completo */
```

```
* Yearse = años de educacion (computados segun Pessino, 1995, pp. 23)
```

```
gen yearse=3
```

```
replace yearse=7 if edup==1
```

```
replace yearse=9.5 if edusi==1
```

```
replace yearse=12 if edus==1
```

```
replace yearse=14.5 if eduui
```

```
replace yearse=17 if eduu==1
```

```
* Otras variables
gen wage = inghora
gen lwage = log(inghora)
gen age = edad
gen agesq = edad*edad
gen exper = age - yearse - 6
gen exper2 = exper*exper
gen sexo01 = sexo-1 /* sexo01=1 si mujer, 0 si varon */

/* Dummy para denotar la muestra relevante
muest=1 si es varon, edad entre 24 y 54, ingresos mayores que cero y
declarados */

gen muest=0
replace muest=1 if edad>24
replace muest=0 if edad>54
replace muest=0 if sexo01==1
replace muest=0 if inghora==0
replace muest=0 if wage==.

* II) Preguntas del Trabajo Practico No.1
-----

* 1) Estadisticas basicas para la muestra relevante
summarize edupi edup edusi edus eduui eduu yearse age exper wage lwage
if muest==1

* 2) Relacion ingresos edad
table edad if muest==1, c(mean inghora)
graph lwage edad if muest==1

* 3) Ecuaciones de Mincer (corresponde a Pessino (1995, pp. 31 columnas
* 1 a 4)
regress lwage yearse exper exper2 if muest==1
regress lwage yearse age agesq if muest==1
regress lwage edup edusi edus eduui eduu exper exper2 if muest==1
regress lwage edup edusi edus eduui eduu age agesq if muest==1

* 4) Algunos tests
regress lwage yearse exper exper2 if muest==1
test exper exper2
test exper2
```