

# Introducción a STATA 9.0

## 1. Estructura

Cuando abrimos el programa STATA aparecen 4 ventanas:

- i) stata command*: ventana donde se escriben los comandos
- ii) stata results*: presenta los resultados de cada comando que usamos
- iii) review*: guarda los comandos anteriores que pueden ser reenviados a la ventana de *stata command* clickeando sobre ellos
- iv) stata variables*: lista de variables que pueden ser reenviados a la ventana de *stata command* clickeando sobre ellos

## 2. Entorno de trabajo

Hay 2 maneras de trabajar en stata:

- i) escribir directamente los comandos a utilizar en la ventana de *stata command*
- ii) “programando” en *stata do files* (archivos con extensión *.do*). Esto es muy útil pues podemos guardar los archivos y por lo tanto si queremos correr nuevamente las regresiones no precisamos tipear los comandos nuevamente sino que abrimos el archivo *do file* y lo corremos. Para usar esta opción hacer click en *do file* y se abra una pantalla donde podemos empezar a escribir el programa. Obviamente debemos guardar este archivo.

Si queremos conservar las salidas de todo lo que hacemos en *stata* debemos guardar los resultados en un archivo (*log file*) que puede ser editado en *notepad* o en *word*. El *log file* es un archivo de tipo *ASCII* donde aparece todo el output como texto. Para usar esta opción primero debemos crear un archivo *log file* al inicio y luego al final debemos cerrarlo:

```
log using prueba.log, replace  
log close
```

donde *prueba* es el nombre del archivo del archivo que guardara los resultados de *stata*.

\* *outreg* sirve para que las salidas de las regresiones queden con formato de tablas que se pueden copiar directamente al documento.

También es útil incrementar la memoria utilizada por *stata* especialmente cuando trabajamos como archivos de gran tamaño.

```
Set mem 300m
```

```
clear borra todos los datos en memoria.
```

Si empezamos las instrucciones en el do file con un \*, entonces stata no lo lee. Esto es útil para agregar comentarios al do file.

Ej: \* Esto no lo lee stata

### 3. 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. Suponiendo que tenemos los datos que queremos utilizar en Excel hay 2 posibilidades para que puedan ser leídos por stata:

i) Utilizar el programa Stat Transfer que es un programa utilitario muy práctico que sirve para transformar bases de datos grabadas en cualquier formato a cualquier otro formato. Los formatos posibles entre otros incluyen: stata, excel, dbase, access, matlab, gauss, sas, spss, access, limdep, s-plus, statsitica, minitab. Con dicho programa en un segundo y de manera automática se transforma un archivo excel (extensión xls) en un archivo stata (extensión dta).

ii) En stata no se puede importar datos directamente desde excel. Para importar un archivo excel en stata seguir los siguientes pasos:

- Abrir archivo en excel
- Herramientas-opciones-internacional-separador decimal .-separador de miles ,
- Guardar como
- Text (tab delimited)
- Ir a stata
- File-import
- ASCII data created by spreadsheet, marcar tab delimitado
- Listo

Es importante que se guarde el archivo como tab delimitado y no comma delimito pues se confunden las comas del delimitado con las comas de los números decimales en la base de datos de excel.

### 4. Signos útiles

Arithmetic	Logical	Relational (numeric and string)
+ addition	~ not	> greater than
- subtraction	! not	< less than
* multiplication	or	>= > or equal
/ division	& and	<= < or equal
^ power		= equal
		~= not equal

Dos signos de igual se usan para comparaciones lógicas y una solo para asignar un valor.

. representa un missing value (variable no disponible)

## 5. Comandos útiles

list = lista de las variables. También se puede listar un subconjunto de variable con list nombre de las variables a listar

sum = summarize. Presenta estadísticos descriptivos de todas las variables o de las variables requeridas.

sum variable , detail. Aparecen más estadísticos descriptivos como kurtosis, etc.

tab = tabulate, frecuencias

keep, drop, save

keep if educ>12 & dpto==1

drop in 10, borra solo la décima observación

gen = generate. Genera una nueva variable.

gen educ2 = educ\*educ

egen xx = mean(educ) ,by(sexo)

replace = modifica una variable existente.

Para generar una variable dummy que vale 1 si la persona trabaja:

gen dummy\_trabaja=0

replace dummy\_trabaja=1 if horas >0

o gen dummy\_trabaja=(horas >0)

sort horas = ordena la base de datos de manera creciente por la variables horas.

Para ordenar descendentemente usar gsort:

gsort -horas (reverse sort)

list in 1/10

merge permite combinar bases de datos que tiene una variable en común.

Requiere que las dos bases de datos estén ordenadas de manera ascendente por la variable común.

merge nhog using base\_de\_datos\_2

## 6. Gráficos

Para hacer gráficos existen diversos formatos, colores, escalas, etc. Les dejo a ustedes que se fijen en la ayuda los distintos formatos y opciones en cuanto a gráficos.

graph variable1 variable2

histogram horas = histograma de la variable horas

## 7. Regresiones

correlate x y = coeficiente de correlación entre las variables x y

Pruebas de diferencias de medias:

ttest horas = 20

ttest horas = wporhora

Estimación por MCO:

reg wporhora educ educ2 sexo edad , robust

la opción robust da errores estándares según la matriz de White (robustos a cualquier forma de heterocedasticidad).

Estimación por variables instrumentales:

ivreg wporhora men6 (educ=edad)

Para obtener predicciones luego de correr la regresión:

predict horas\_hat

probit dummy\_trabaja educ educ2 sexo edad

logit dummy\_trabaja educ educ2 sexo edad

tobit horas educ educ2 ing\_tot\_res , ll(0)

ll(0) indica censura a la izquierda en 0

tobit horas educ educ2 ing\_tot\_res , ul(100)

ul(100) indica censura a la derecha en 100

## 8. do files

Estructura:

clear

set mem 500m

cd "C:\varios" Esto fija el directorio de trabajo

log using tables.log ,replace output en un archivo

\* para escribir comentarios

comandos