

# STATA 7 AL ALCANCE DE TODOS: GUIA DE USO.

Por: Paloma Rodríguez Campo ( [palo1980@hotmail.com](mailto:palo1980@hotmail.com) )  
y Gorka Otaño Marote ( [songorka@mundofree.com](mailto:songorka@mundofree.com) )

## 1. INTRODUCCIÓN.

El siguiente documento muestra los pasos básicos para poder realizar funciones sencillas de estimación con el programa STATA. Para ello, se utilizará de ejemplo los datos de panel *WAGEPAN* que se encuentra en la base de datos del libro Wooldridge, J. (2002) Econometric Analysis of Cross Section and Panel Data, The MIT Press Cambridge, Massachusetts.


➤ Para acceder al *programa*, seguir el siguiente camino:

- Inicio → Programas → Stata → [versión] Stata 7

“**versión**” dependerá de la versión instalada del programa (easy, intercooled...).

- También es posible que en el escritorio aparezca el acceso directo a STATA 7. Hacer doble clic en él para acceder al programa.



➤ Para acceder a una de las bases de datos que esté en formato STATA, se hará a través del botón situado en la parte superior izquierda . Por defecto se abre el directorio C: STATA, se recomienda situar en este directorio los ficheros \*.DTA (compatibles con STATA) del CD de bases de datos o de la pagina web.

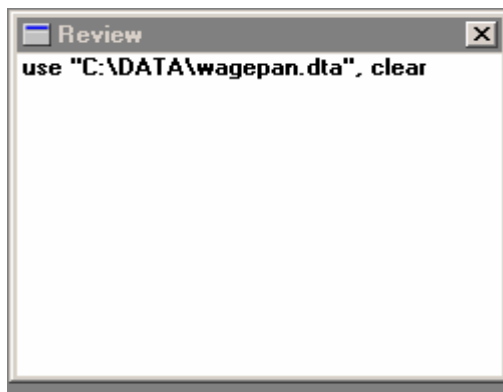
El programa tiene 4 ventanas, si resultan incómodas por su pequeño tamaño basta con situar el cursor en uno de los lados de cada ventana hasta que éste se convierta en una flecha ( — ) y poner el tamaño que más convenga. Para moverlas, situar el cursor en la parte de arriba de la ventana, presionar el botón izquierdo y según se mueve el ratón se moverá la ventana.

## 2. DESCRIPCIÓN DE LAS VENTANAS

- **Stata command:** se encuentra en la parte inferior derecha. En esta ventana vamos a escribir todas las instrucciones que ejecutemos.

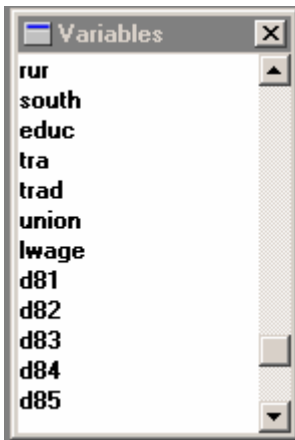


- **Review:** ventana situada en la parte superior izquierda, en ella quedarán registradas todas las órdenes que demos al programa:

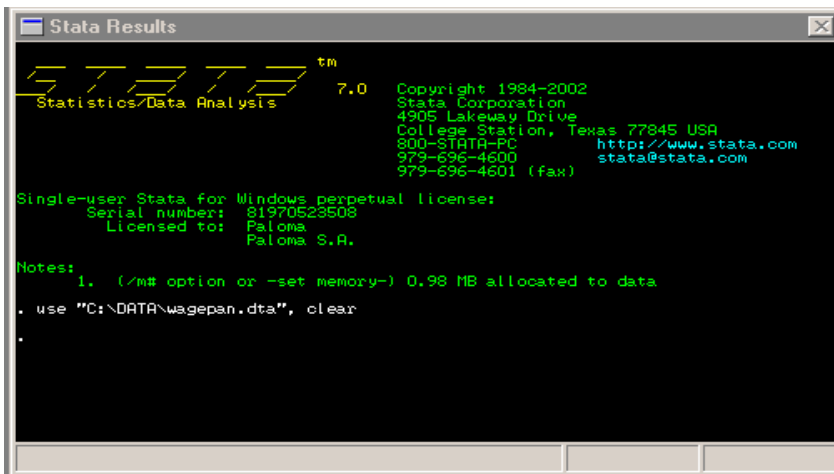


En este caso está registrada la orden anterior de abrir la base de datos “*wagepan*”

- **Variables:** situada en la parte inferior izquierda, en esta ventana aparecen todas las variables que hay en la base de datos elegida (en este caso las variables de “*wagepan*”). Para usarlas basta con hacer doble clic con el ratón sobre ellas.

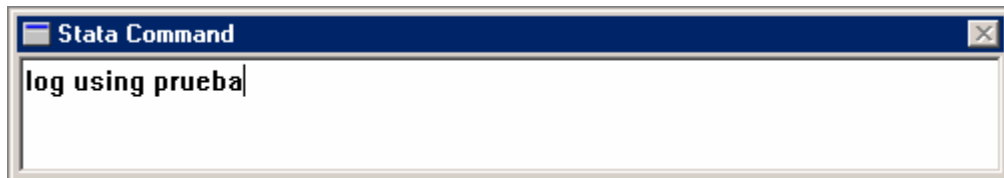


- **Stata Results:** situado en la parte superior derecha, en esta ventana aparecerán los resultados de las ejecuciones de las órdenes como las estimaciones de los modelos.



### 3. PASOS BÁSICOS DEL PROGRAMA

- 1) Lo primero que se debe hacer es usar la orden **“log using”** que se encarga de *guardar* toda la sesión en un fichero llamado “prueba” (en este caso) y con extensión \*.smcl



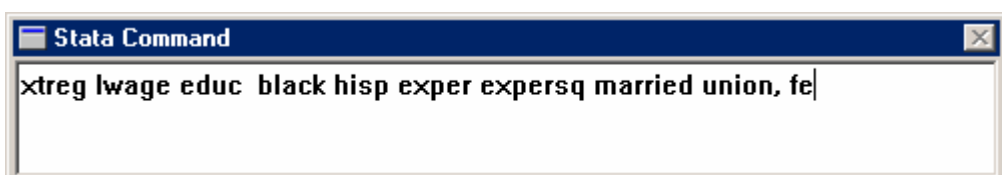
```
log using prueba
```

Se puede sustituir la palabra “prueba” por el nombre deseado. Es muy importante no usar el mismo nombre dos veces, ya que STATA no lo permite. En el caso de que el fichero ya exista, aparece el mensaje en la ventana **Stata Results** “file C:\DATA\prueba.smcl already exists”

- 2) Estimación de un modelo mediante **efectos fijos**, se toma como ejemplo el modelo:

$$\log(\text{wage}_{it}) = \mathbf{q}_t + \mathbf{b}_1 \text{educ} + \mathbf{b}_2 \text{black}_i + \mathbf{b}_3 \text{hispan}_i + \mathbf{b}_4 \text{exp } er_{it} + \mathbf{b}_5 \text{exp } er^2_{it} + \mathbf{b}_6 \text{married}_{it} + \mathbf{b}_7 \text{union}_{it} + c_i + u_{it}$$

Para ello hay que utilizar la orden **xtreg** seguida por la *variable dependiente* lwage y después las *variables independientes*. Después de la última variable debe ir una coma “,” y un espacio, y escribir **fe** para que se estime por efectos fijos en la ventana de **Stata Command**.



```
xtreg lwage educ black hisp exper expersq married union, fe
```

Para añadir las variables basta con hacer clic en el nombre de cada variable en la ventana de “variables”

Salida en Stata results:

```

. xtreg lwage educ black hisp exper expersq married union, fe
Fixed-effects (within) regression      Number of obs   =   4360
Group variable (i) : nr                Number of groups =   545
R-sq:  within = 0.1780                  Obs per group:  min =    8
      between = 0.0005                    avg =   8.0
      overall  = 0.0638                    max =    8
corr(u_i, Xb) = -0.1139                  F(4, 3811)      =   206.38
                                          Prob > F        =    0.0000

```

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	(dropped)					
black	(dropped)					
hisp	(dropped)					
exper	.1168467	.0084197	13.88	0.000	.1003392	.1333542
expersq	-.0043009	.0006053	-7.11	0.000	-.0054876	-.0031142
married	.0453033	.0183097	2.47	0.013	.0094056	.081201
union	.0820871	.0192907	4.26	0.000	.044266	.1199083
_cons	1.06488	.0266607	39.94	0.000	1.012609	1.11715
sigma_u	.4000539					
sigma_e	.35125535					
rho	.5646785	(fraction of variance due to u_i)				

Análisis por partes:

VARIABLES	COEFICIENTES	DESVIACIÓN TÍPICA	ESTADÍSTICO t	P-valor	INTERVALO DE CONFIANZA AL 95%
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	(dropped)				
black	(dropped)				
hisp	(dropped)				
exper	.1168467	.0084197	13.88	0.000	.1003392 .1333542
expersq	-.0043009	.0006053	-7.11	0.000	-.0054876 -.0031142
married	.0453033	.0183097	2.47	0.013	.0094056 .081201
union	.0820871	.0192907	4.26	0.000	.044266 .1199083
_cons	1.06488	.0266607	39.94	0.000	1.012609 1.11715

- En la primera columna se muestra los nombres de las variables
- **Coef.** → muestra los coeficientes estimados de las variables.

En *educ*, *black* e *hisp* no sale coeficiente, en su lugar aparece *dropped*, eso quiere decir que el programa a la hora de estimar, las ha eliminado ya que no varían en el tiempo, un individuo siempre es negro o hispano.

- **Std. Err.** → muestra la **desviación típica**
- **t** → es el valor del **estadístico t**

- $P>[t] \rightarrow$  es el **p-valor**, la probabilidad que deja a la derecha el valor estimado
- [95% Conf. Interval]  $\rightarrow$  es el **intervalo de confianza al 95%**. El centro del intervalo siempre coincide con el coeficiente.

```

Fixed-effects (within) regression              Number of obs   =   4360
Group variable (i) : nr                       Number of groups =    545
R-sq:  within = 0.1780                        Obs per group:  min =     8
        between = 0.0005                       avg =    8.0
        overall = 0.0638                       max =     8
corr(u_i, Xb) = -0.1139                       F(4,3811)       =   206.38
                                                Prob > F        =    0.0000

```

- **R-sq: within**  $\rightarrow$  Mide el  $R^2$ , el **coeficiente de determinación**. Utilizar el *within*.  
**Number of obs**  $\rightarrow$  Número de observaciones.
- **Number of groups**  $\rightarrow$  Número de grupos, en este caso de individuos
- **Obs per group**  $\rightarrow$  Observaciones temporales existentes para cada grupo, en este caso son los años.
- **F-estadístico**  $\rightarrow$  **Estadístico F**, es la significación conjunta de todas las variables explicativas del modelo, en esta caso es  $F(4,3811)$ , donde el 4 son los grados de libertad (número de variables sin tener en cuenta la constante), y 3811 es el número de observaciones menos el número de individuos menos el número de variables sin contar las constantes ( $4360-545-4=3811$ ). Debajo está **Prob > F** que indica la probabilidad que queda a la derecha del valor estimado.
- **Corr(u\_i, Xb)**  $\rightarrow$  Correlación entre errores y regresores, estimado con efectos aleatorios.

```

sigma_u  .4000539
sigma_e  .35125535
rho      .5646785 (fraction of variance due to u_i)

```

- *sigma\_u*: Es la desviación típica de los errores comunes.
- *sigma\_e*: Es la desviación típica de los errores individuales.

3) Estimación de un modelo mediante **efectos aleatorios**. Vamos a estimar el mismo modelo que antes pero esta vez por efectos aleatorios

```
Stata Command
xtreg lwage educ black hisp exper expersq married union
```

es igual que el anterior pero **sin** la parte final de **fe**, ya que el programa estima mediante efectos aleatorios por defecto.

### Salida en Stata results

```
Stata Results
. xtreg lwage educ black hisp exper expersq married union
Random-effects GLS regression
Group variable (i) : nr
Number of obs = 4360
Number of groups = 545
R-sq: within = 0.1774
      between = 0.1837
      overall = 0.1808
Obs per group: min = 8
               avg = 8.0
               max = 8
Random effects u_i ~ Gaussian
corr(u_i, X) = 0 (assumed)
Wald chi2(7) = 943.95
Prob > chi2 = 0.0000
```

	lwage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
educ		.1012246	.0089133	11.36	0.000	.0837549 .1186943
black		-.1441307	.0476148	-3.03	0.002	-.237454 -.0508073
hisp		.0201511	.0426011	0.47	0.636	-.0633456 .1036477
exper		.1121195	.0082609	13.57	0.000	.0959285 .1283105
expersq		-.0040689	.0005918	-6.88	0.000	-.0052288 -.0029089
married		.0627951	.0167729	3.74	0.000	.0299209 .0956693
union		.1073789	.01783	6.02	0.000	.0724327 .142325
_cons		-.1074643	.1107057	-0.97	0.332	-.3244435 .1095149
sigma_u		.32456727				
sigma_e		.35125535				
rho		.46057172				(fraction of variance due to u_i)

more

log on (smcl)

Comparado con el modelo anterior, las diferencias en la salida son:

- **z-estadístico** → Está en lugar del t-estadístico. Es muy similar y se utiliza de la misma manera.
- **Random effects u\_i** En este caso asumimos que son normales: Gaussianos. Y que las correlaciones entre los errores y las x son cero.
- **Chi<sup>2</sup>** → está en lugar del estadístico F, es similar y se utiliza de manera parecida.

4) **Contraste de Hausman.** Este contraste nos muestra si no existe (H<sub>0</sub>) o si existe (H<sub>1</sub>) correlación entre errores y regresores. En el caso de que exista, hay que utilizar el estimador de efectos fijos, si no, se puede utilizar el de efectos

aleatorios. Para que Stata realice el contraste, simplemente hay que escribir “xthausman” en la ventana Stata comand.

### Salida de **Stata Results**

```
Hausman specification test
```

lwage	Coefficients		Difference
	Fixed Effects	Random Effects	
exper	.1168467	.1121195	.0047272
exper <sup>2</sup>	-.0043009	-.0040689	-.000232
married	.0453033	.0627951	-.0174918
union	.0820871	.1073789	-.0252917

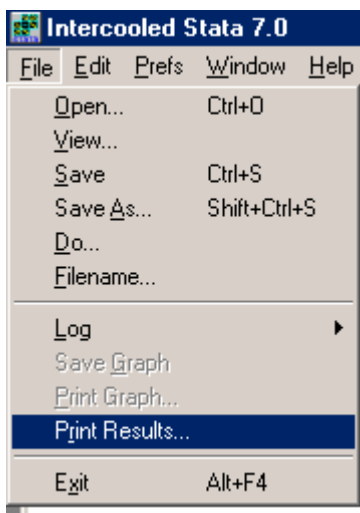
```

Test: Ho: difference in coefficients not systematic
      chi2( 4) = (b-B)'[S^(-1)](b-B), S = (S_fe - S_re)
      = 31.45
      Prob>chi2 = 0.0000

```

- **Prob>chi<sup>2</sup>** → es la probabilidad que queda a la derecha del valor estimado.

5) **Imprimir salidas.** A veces resulta más cómodo para analizar unos resultados verlos escritos sobre un papel. Para ello, ir a **File** y a **Print Results...** y se impriman todo lo que ha ido apareciendo en la ventana de **Stata Results**.

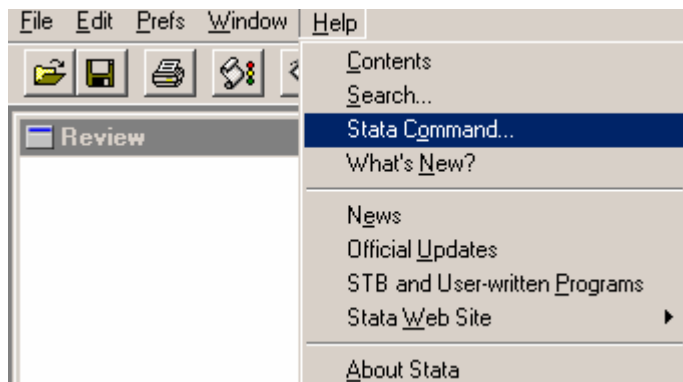


6) **Ver un archivo guardado.** En el paso 1, usamos la orden *log*, al hacer esto el programa ha ido guardando las salidas de STATA en formato ASCII en un fichero con extensión *smcl*. Por defecto, se guarda en C:/DATA. Para abrirlo, situar el ratón encima del fichero, presionar el botón derecho y elegir la opción

“abrir con...” y elegir un programa como el “el bloc de notas” o “Microsoft Word”.

Al abrir esos archivos, cada línea comienza con una abreviatura. Esta abreviatura dice que tipo de línea es. Así por ejemplo {com} se refiere a una orden escrita en la ventana **Stata Command**, {txt} son las líneas de texto de los resultados que aparecen en la ventana **Stata Results**, {err} se refiere a algún error que hayamos cometido...

- 7) **Uso de HELP:** Presionando con el ratón sobre la opción **help** o pulsando “alt” + “h”, se despliega el menú de ayuda de STATA. Dentro de ese menú se encuentra la opción “**stata command**”.



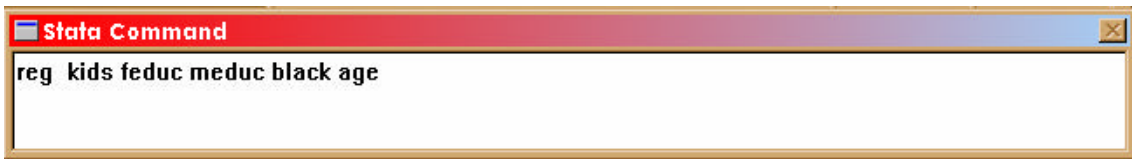
Una vez presionada la opción, se escribe el nombre de la instrucción que queremos utilizar para que nos salga una extensa descripción de cómo funciona y las posibilidades que tiene.

## CORTE TRANSVERSAL

Los pasos para trabajar con datos de corte transversal son similares a los empleados para datos de panel. La diferencia radica en las órdenes o comandos que se utilizan en el programa.

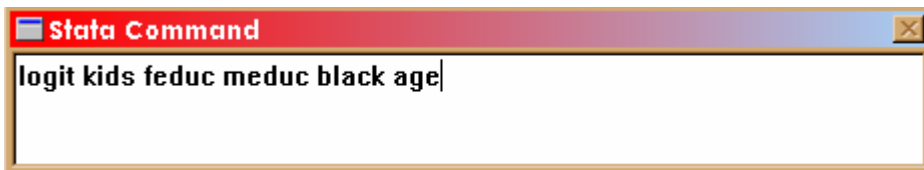
Para los ejemplos vamos utilizar la base de datos FERTIL1, usando *kids* como variable dependiente, y *feduc*, *meduc*, *black* y *age* como variables explicativas.

- Estimar un modelo de regresión normal, se utiliza la orden **reg** seguida de la variable dependiente y de las independientes. Se escribe en la ventana de “stata command”.



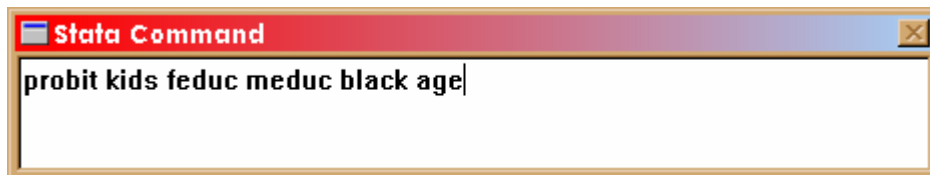
```
Stata Command
reg kids feduc meduc black age
```

- Estimar un modelo logit por máxima verosimilitud, se utiliza la orden **logit**, seguida de la variable dependiente y de las independientes. Se escribe en la ventana de “stata command”.



```
Stata Command
logit kids feduc meduc black age|
```

- Estimar un modelo probit por máxima verosimilitud, se utiliza la orden **probit**, seguida de la variable dependiente y de las independientes. Se escribe en la ventana de “stata command”.



```
Stata Command
probit kids feduc meduc black age|
```