

Comandos de utilidad en Stata

Héctor Tavárez

Nota: El propósito de este documento es informar sobre la existencia de algunos comandos de utilidad en Stata. Se proveen ejemplos de como se utilizan. Sin embargo, el objetivo no es explicar la prueba correcta para probar tu hipótesis, ni explicar detalladamente como funciona cada uno de los comandos. Se exhorta a la persona interesada a buscar en detalle como funciona cada comando en el manual de Stata debido a que cada comando tiene sus propias especificaciones y limitaciones. A continuación, se provee una lista de comandos divididos por temas que se discuten en ocasiones en los cursos.

Operadores de aritmética

- + Suma
- - Resta
- * Multiplicación
- / División
- ^ Potencia

Funciones matemáticas

- abs (x) calcula el valor absoluto
- sqrt (x) calcula la raíz cuadrada
- ln (x) calcula el logaritmo natural
- round (x) redondea al numero entero mas cercano

Comandos para establecer condiciones

- if quiere decir “si”
- in quiere decir “para las observaciones”

Se establecen condiciones con operadores lógicos y operadores de relación:

Operadores lógicos

- & quiere decir “y”
- | quiere decir “o”
- ! quiere decir “no”

Operadores de relación

- > quiere decir “mayor a”
- < quiere decir “menor a”
- >= quiere decir “mayor o igual a”
- <= quiere decir “menor o igual a”
- == quiere decir “igual a”
- != quiere decir “no es igual a”

* Los ejemplos de estos operadores lógicos y de relación se pueden ver abajo cuando se presenta cada comando

Manipulación de variables y manejo de variables

- generate

Útil para crear una nueva variable

Ejemplo: generate newvar1 = 0

Otros ejemplos serían el siguiente:

```
generate newvar1 = ingreso ^ 2
```

```
generate newvar1 = varname2 * varname3
```

```
generate newvar1 = varname2 + varname3 if varname4>1
```

- replace

Útil para cambiar valores dentro de una variable

Ejemplo: replace varname1 = 0 if varname == 2

*Note que en este comando luego del “if” se debe utilizar doble signo de igualdad, según la definición arriba en la sección de “operadores de relación”

Otros ejemplos serían los siguientes:

```
replace varname1 = 0 if varname1 == 2
```

```
replace varname1 = 1 if varname2 > 0
```

```
replace varname1 = 0 if varname2 != 1
```

```
replace varname1 = 1 if varname2 == 5 & varname3 == 10
```

- rename

Útil para cambiar el nombre a una variable

Ejemplo: rename ingrexo ingreso

*Note que en este comando no se utiliza el signo de igual

- drop

Útil para eliminar variables u observaciones

Ejemplos: drop ingreso

drop if género == "hombre"

- keep

Este comando trabaja de forma opuesta al comando drop. Permite dejar en Stata solo las variables u observaciones que se mencionan en este comando.

Ejemplos: keep ingreso edad educación genero dependientes

keep if genero == "hombre"

- sort

Útil para ordenar los datos en orden ascendente

Ejemplo: sort genero ingreso edad

Descripción de datos

- list

Útil para ver una variable en columna en la pantalla de resultados

Ejemplo: list varname1

El comando permite ver mas de una variable en columnas

list ingreso educación genero

- summarize

Útil para resumir los datos de una variable. Ofrece el número de observaciones, la media, desviación estándar, valor mínimo y máximo de la variable

Ejemplo: summarize varname1

Además, el comando permite evaluar múltiples variables a la misma vez:

summarize varname1 varname2 varname3

Se puede utilizar este comando cuando una variable toma unos valores en particular.
summarize varname1 if varname2<=1

Además, se puede utilizar este comando para ciertas observaciones en particular. Por ejemplo, se puede utilizar el comando “in” para resumir la variable solo durante las primeras 40 observaciones

summarize varname1 in 1/40

* Note que estas condiciones se pueden establecer con muchos comandos

El comando también se puede combinar con “, detail” para obtener mas información de la variable, como la mediana, varianza, etc.

summarize varname1, detail

- tabulate

Resumen de probabilidades de cada nivel dentro de una variable categórica

Ejemplo: tabulate varname1

El comando permite hacer tabulación cruzada de dos variables

tabulate varname1 varname2

- inspect

Útil para resumen rápido de una variable numérica. Presenta información diferente a los comandos summarize y tabulate. Presenta el número de valores negativos, positivos, y ceros. Además, provee información sobre el número de valores enteros y no enteros, y números de valores omitidos (missing values). Produce un pequeño histograma.

Ejemplo: inspect varname1

Otros ejemplos serían el siguiente:

inspect varname1 if varname2 == 0

inspect varname2 in 50/100

Pruebas estadísticas

El uso de estas pruebas va a depender de la hipótesis que se quiere probar, si se cumple con las suposiciones correspondientes a cada prueba y de la naturaleza de los datos. Se recomienda consultar con su libro de estadísticas para seleccionar la prueba correcta.

- swilk

Estima la prueba de Shapiro-Wilk para normalidad de los datos

Ejemplo: swilk varname1

También puedes verificar distribución de los datos de mas de una variable a la vez
swilk varname1 varname2 varname3

- correlate

Presenta la correlación de los datos

Ejemplo: correlate varname1 varname2

También puedes estimar la matriz correlación entre múltiples variables
correlate varname1 varname2 varname3 varname4 varname5 varname6

- ttest

Realiza la prueba t

Ejemplo: ttest varname1==50

Otros ejemplos serían los siguientes:

ttest varname1==varname2
ttest ingreso, by(género)

* Nota: La prueba a escoger va a depender de la naturaleza de los datos y la hipótesis a probar.

- tabulate

Calcula y muestra el chi-cuadrado de Pearson

Ejemplo: tabulate varname1 varname2, chi2

El comando también puede ser utilizado para calcular la Prueba Exacta de Fisher
tabulate varname1 varname2, exact

- anova

Realiza el one-way anova

Ejemplo: anova varname1 varname2

También realiza el two-way anova
anova varname1 varname2 varname3

- ranksum

Realiza la prueba Wilcoxon-Mann-Whitney. Prueba no paramétrica.

Ejemplo: ranksum ingreso, by(genero)

- **kwallis**

Realiza la prueba Kruskal-Wallis. Prueba no paramétrica.

Ejemplo: kwallis ingreso, by(educacion)

Regresiones

- **regress**

Estima una regresión múltiple (Mínimos Cuadrados Ordinarios) para variables dependientes que son continuas

Ejemplo: regress depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían los siguientes:

```
regress depvar indepvar1 indepvar2 if varname1>1
```

```
regress depvar indepvar1 indepvar3 if depvar<=300
```

* Nota: *El último ejemplo sería de utilidad si, por ejemplo, determinamos que los números por arriba de 300 en nuestra variable dependiente son “outliers”.*

- **qreg**

Estima una regresión por cuantiles. La regresión por cuantil tiene como objetivo estimar la mediana condicional u otros cuantiles de la variable de respuesta como la percentila 75th, por ejemplo.

Ejemplo: qreg depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían los siguientes:

```
qreg depvar indepvar1 indepvar2 if depvar<200, quantile(.75)
```

```
qreg depvar indepvar1 indepvar2 indepvar3 i.indepvar4
```

* Nota: *En el último ejemplo, la variable indepvar4 es categórica y se crea un variable dummy para cada categoría. La regresión por cuantiles, y otras regresiones, estima el modelo con estas dummies cuando se añade “i.” antes de la variable de interés.*

- **poisson**

Estima una regresión Poisson para variables dependientes de conteo de un evento. La variable dependiente tiene una distribución Poisson.

Ejemplo: poisson depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían los siguientes:

poisson depvar indepvar1 indepvar2 in 100/300

poisson depvar indepvar1 indepvar2 indepvar3, vce(robust)

* Nota: En el último ejemplo, vce indica el tipo de error estándar reportado. En este caso robustos.

- nbreg

Estima una regresión binomial negativa para variables dependientes de conteo de un evento. Se utiliza comúnmente para corregir problemas de sobre-dispersión.

Ejemplo: nbreg depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían los siguientes:

nbreg depvar indepvar1 indepvar2 in 100/300

nbreg depvar indepvar1 indepvar2 indepvar3, irr

* Nota: En el último ejemplo, se reportan los incident rate ratios

- logit

Estima un modelo logit para variables dependientes que son binarias

Ejemplo: logit depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían los siguientes:

logit depvar indepvar1 indepvar2 indepvar3 in 1/200

logit depvar indepvar1 indepvar2 indepvar3 indepvar4 in 100/300

* Nota: En el último ejemplo, la regresión logit se estima solo desde la observación 100 hasta la 300.

- probit

Estima un modelo probit para variables dependientes que son binarias

Ejemplo: probit depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían los siguientes:

probit depvar indepvar1 indepvar2 indepvar3 in 1/200

probit depvar indepvar1 indepvar2 indepvar3 indepvar4 in -300/1

* Nota: En el último ejemplo la regresión probit se estima para las últimas 300 observaciones

- **mlogit**

Estima un modelo logit multinomial para variables dependientes que son categóricas

Ejemplo: mlogit depvar indepvar1 indepvar2 indepvar3

Otros ejemplos serían:

mlogit depvar indepvar1 indepvar2 indepvar3 if varname1!=5

mlogit depvar indepvar1 indepvar2 indepvar3, baseoutcome(1)

* Nota: En el último ejemplo, el modelo logit multinomial deja depvar=1 como la base.

- **ologit**

Estima un modelo logit ordenado para variables dependientes que son ordinal

Ejemplo: ologit depvar indepvar1 indepvar2 indepvar3

Otro ejemplo serían los siguientes:

ologit depvar indepvar1 indepvar2 indepvar3 if varname1<4

ologit depvar indepvar1 indepvar2 indepvar3 indepvar4 indepvar5 a b a##b

* Nota: El último ejemplo incluye términos de interacción entre las variables a y b.

- **clogit**

Estima un modelo logit condicional para casos emparejados – datos de control.

Comúnmente utilizado en el método de experimentos de elección.

Ejemplos: clogit depvar indepvar1 indepvar2 indepvar3, group(groupvar)

Otros ejemplos serían los siguientes:

clogit depvar indepvar1 indepvar2 indepvar3 if varname2==1, group(groupvar)

clogit depvar indepvar1 *var, group(groupvar)

* Nota: En este comando hay que especificar una variable que identifica un grupo, es obligatorio. Para el último ejemplo, *var indica que todas las variables que terminan con "var" en el conjunto de datos entran al modelo como variables independientes.

* Ver el comando asclogit, el cual es un caso específico del clogit

- **mixlogit**

Estima un modelo logit mixto, conocido como un modelo logit con parámetros aleatorios. Comúnmente utilizado en el método de experimentos de elección.

Ejemplos: mixlogit depvar indepvar1 indepvar2, group(groupvar) rand(indepvar3
indepvar4 indepvar5)

Otros ejemplos serían los siguientes:

```
mixlogit depvar indepvar1 indepvar2 indepvar3 if varname1>=4,  
group(groupvar) rand(indepvar)
```

```
mixlogit depvar indepvar1 var*, group(groupvar) rand(indepvar)
```

* Nota: En este comando hay que especificar una variable que identifica un grupo y se tiene que especificar las variables que se van a estimar con parámetros aleatorios.
Para el último ejemplo, var* indica que todas las variables del conjunto de datos que comienzan con “var” van a entrar al modelo como variables independientes.

Para hacernos la vida más fácil...

A continuación, se presentan algunas abreviaciones para escribir en la ventana de comandos en Stata:

summarize → sum
tabulate → tab
generate → gen
rename → ren
correlate → corr
anova → an
regress → reg

Para entender como funciona cada comando de los presentados anteriormente, puede escribir “help command_name” en la ventana de comandos de Stata. Por ejemplo, luego de escribir “help logit” en la ventana de comandos de Stata, al oprimir “enter” esto le abre una venta donde le explica como funciona el comando logit. Similarmente, luego de escribir “help poisson” en la ventana de comandos de Stata, al oprimir “enter” esto le abre una venta donde le explica como funciona el comando poisson. ¡INTENTELO!