

CÁLCULO DE ERRORES MUESTRALES CON



1. GENERALIDADES DE STATA

STATA es un software para el análisis estadístico, muy adecuado para el manejo de grandes volúmenes de datos. En general funciona mediante la escritura directa de comandos, aunque la versión 8 permite ejecutar una gran cantidad de procedimientos de manera interactiva, a través de menús.

Esta aplicación se puede utilizar en tres presentaciones: *Small-STATA* (dirigida a estudiantes), *STATA/Intercooled* (estándar) y *STATA-SE*. Esta última es la versión más completa, con mayor capacidad para trabajar con archivos de datos sumamente grandes y con análisis más complejos.

Una de las grandes ventajas que tiene el software es que permite calcular con un solo comando los ajustes necesarios que, de acuerdo con el diseño, establece la teoría del muestreo para las estimaciones que se hacen con datos de encuestas. Esto lo hace mediante el método de linealización (series de Taylor).

En este trabajo se presentan en forma resumida los comandos principales que se utilizan para el cálculo de los errores de muestreo y efectos del diseño asociados con dichas estimaciones. La intención es que el documento sirva como un pequeño "manual de bolsillo" que posteriormente pueda ser utilizado al analizar datos de encuestas donde se utilizó un diseño complejo. Debido a esto, se incluyen también algunos comandos útiles para el manejo de datos y el análisis estadístico básico (descriptivo).

2. COMANDOS PARA EL TRABAJO CON ARCHIVOS DE STATA

| Comando | Descripción |
|-----------------------------|---|
| <code>set memory 64m</code> | Amplía la memoria de trabajo. Se recomienda hacerlo siempre. |
| <code>use</code> | Abre el archivo de datos de STATA. Requiere la especificación del directorio donde se encuentra. |
| <code>C:\folder1</code> | Fija folder1 como directorio de trabajo. Si se hace esto primero, no es necesario especificar la ruta cuando se usa el comando <i>use</i> . |
| <code>clear</code> | Cierra el archivo. |
| <code>exit</code> | Cierra el programa. |

| Comando | Descripción |
|----------------------------------|---|
| <code>infile/ insheet</code> | Permite importar un archivo en formato ASCII. |
| <code>describe</code> | Describe el contenido del archivo: Número de observaciones, fecha de creación, espacio en memoria RAM, tipo de variables, etiquetas de variables, etiquetas de valores etc. |

3. MANEJO DE DATOS

| Comando | Descripción |
|---------------------------------------|--|
| <code>rename var var2</code> | Cambia el nombre de la variable <i>var</i> por el de <i>var2</i> |
| <code>gen var3=var1+2*var2</code> | Crea una nueva variable. En el ejemplo, crea <i>var3</i> como combinación lineal de <i>var1</i> y <i>var2</i> . |
| <code>recode</code> | Recodifica la variable. Sustituye los valores originales por los nuevos en el mismo campo, es decir, <u>se pierde la variable original</u> . |
| <code>list</code> | Lista un grupo de variables |
| <code>keep varlist</code> | Deja las variables que se indican (<i>varlist</i>) y elimina las demás. |
| <code>drop varlist</code> | Elimina las variables que se indican (<i>varlist</i>) y deja las demás. |
| <code>sort var1</code> | Ordena el archivo según la variable <i>var1</i> |

4. ESTADÍSTICAS DESCRIPTIVAS

| Comando | Descripción |
|-----------------------------|--|
| <code>sum var1</code> | Devuelve las principales estadísticas resumen (promedio, sd, etc). |
| <code>tab1 var1 var2</code> | Construye una distribución de frecuencias para cada |

| Comando | Descripción |
|--------------------------------------|---|
| | variable. |
| <code>tab var1 var2</code> | Construye un cuadro cruzado de las dos variables. |
| <code>correlate varlist</code> | Construye la matriz de correlaciones |
| <code>ttest var1, by var2</code> | Hace la prueba t para la igualdad de promedios de <i>var1</i> , para los grupos definidos por <i>var2</i> . |

5. ANÁLISIS DE DATOS DE ENCUESTAS: La familia de comandos *svy*.

STATA tiene una serie de comandos para el análisis de datos provenientes de encuestas complejas (*Survey data*) los cuales se conocen como *svy commands*.

Los *svy commands* permiten realizar análisis que incluyen los ajustes necesarios en los errores de muestreo, de acuerdo con el diseño utilizado en la encuesta, de esta manera, con dos sencillos comandos se realizan los cálculos apropiados de manera eficiente. Esto aplica para diferentes tipos de análisis inferencial:

- ✓ Estimación (medias, proporciones, totales)
- ✓ Pruebas de hipótesis (e.g. prueba de igualdad de medias)
- ✓ Análisis de regresión (lineal y no lineal).

Algunos de esos comandos son:

| Comando | Descripción | Comando | Descripción |
|-----------------|------------------------------|-------------------|-----------------------------------|
| svymean | Estimación de la media. | svyregress | Regresión lineal |
| svyprop | Estimación de una proporción | svylogit | Regresión logística (dicotómica) |
| svytotal | Estima totales | svymlogit | Regresión logística (multinomial) |

| | | | |
|-----------------|---------------------------|---------------|--|
| svyratio | Estima una media de razón | lincom | Estima combinaciones lineales (e.g. diferencias de medias) |
|-----------------|---------------------------|---------------|--|

5.1 Especificación del diseño de muestreo utilizado:

El diseño de muestreo complejo más común es el muestreo estratificado de conglomerados. STATA permite definir el diseño en estos términos, incluir correcciones para poblaciones finitas y especificar ponderadores. El comando utilizado para especificar el diseño es *svyset*.

```
svyset [pweight=weight] [, strata(varname) psu(varname)
fpc(varname) ]
      [clear(clear_options) | clear]
```

Aquí:

pweight: define el ponderador

strata: define los estratos

psu: define la unidad primaria de muestreo

fpc: define el factor de corrección para poblaciones finitas, aplicado a las UPM's (supone que no hay submuestreo). Se indica la variable que contiene el número de UPM's en la población/estrato

clear: permite eliminar cualquiera de las anteriores

Para ilustrar se usan los datos de la Encuesta Nacional de Salud Reproductiva realizada en 1999. En este archivo, el estrato viene definido por la variable *estrato*, la unidad primaria de muestreo por *upmc*. Además se utilizará como ponderador la variable *fwgt=N/n*

La especificación del diseño es:

```
. svyset [pweight=fwgt], strata(estrato) psu(upmc)
pweight is fwgt
strata is estrato
psu is upmc
```

6. CALCULO DE ERRORES MUESTRALES

Para fines de este trabajo interesan fundamentalmente los comandos que tienen que ver con el cálculo de errores en estimaciones de medias, proporciones, razones y totales.

6.1 Estimación de una media

```
svymean varlist [if exp][in range] by(varlist)
      subpop(varname) deff defl obs size
```

| | |
|-----------------------|--|
| <code>by()</code> | Permite hacer la estimación para subclases definidas por los valores de la variable que se indica (<i>varlist</i>) |
| <code>subpop()</code> | Permite hacer subclases definidas por la variable que se indica, la cual usualmente es <i>dummy</i> |
| <code>deff:</code> | Calcula el efecto del diseño |
| <code>defl:</code> | Calcula la raíz cuadrada del efecto del diseño |
| <code>obs:</code> | Devuelve el número de observaciones |
| <code>size:</code> | Estima el tamaño de la (sub)población |

Para estimar la edad promedio al momento de la primera relación sexual (*p717*):

```
. svy: mean p717
Survey mean estimation
pweight:  fwgt                Number of obs   =      886
Strata:   estrato              Number of strata =        5
PSU:      upmc                 Number of PSUs  =       51
                                   Population size = 701933.5

-----+-----
      Mean |      Estimate   Std. Err.   [95% Conf. Interval]      Deff
-----+-----
      p717 |      26.30248   .5966469   25.1015   27.50347   1.291908
-----+-----
```

6.2 Estimación de un total

```
svytotal varlist [if exp][in range] by(varlist)
      subpop(varname) deff deff obs size
```

Para estimar el total de mujeres que se han practicado la esterilización (*u_ester*, variable "dummy")

```
. svytotal u_ester
Survey total estimation
pweight:  fwgt                Number of obs   =      871
Strata:   estrato             Number of strata =       5
PSU:      upmc                Number of PSUs  =      52
                               Population size = 690049.75

-----
      Total | Estimate   Std. Err.   [95% Conf. Interval]      Deff
-----+-----
u_ester |    136267   11524.45   113082.8   159451.2   1.531191
-----
```

6.3 Estimación de una media de razón

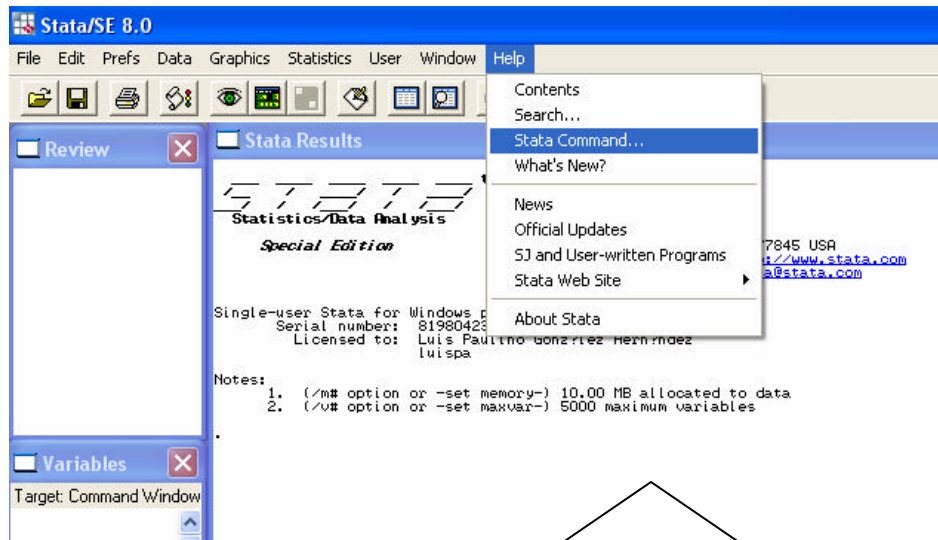
```
svyratio varname[/]varname [varname[/]varname...][if exp]
      [in range] [by(varlist) subpop(varname) deff deff obs size]
```

6.4 Estimación de una proporción

```
svyprop varlist [if exp][in range] [by(varlist)]
```

```
subpop(varname) format(%fmt)
```


PARA NO OLVIDAR...



Si no recuerda la sintaxis de un comando en particular, y no tiene una manual al alcance, STATA le ofrece la ventana STATA COMMAND en el menú Help. Esta herramienta le orientará en la escritura de los comandos que necesite