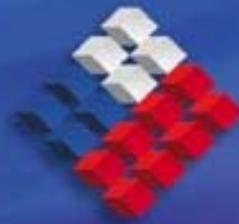


ENCUESTA DE CARACTERIZACION
SOCIOECONOMICA NACIONAL

Casen 2003



GOBIERNO DE CHILE
MIDEPLAN

Manual STATA



Manual STATA

El objetivo de este pequeño manual es introducir al usuario en el manejo básico del programa estadístico STATA, en particular para su uso con la base de datos CASEN. Se utilizará como plataforma la versión 7.0 del programa.

1. Introducción a Stata

Al trabajar en stata ocuparemos principalmente los siguientes archivos:

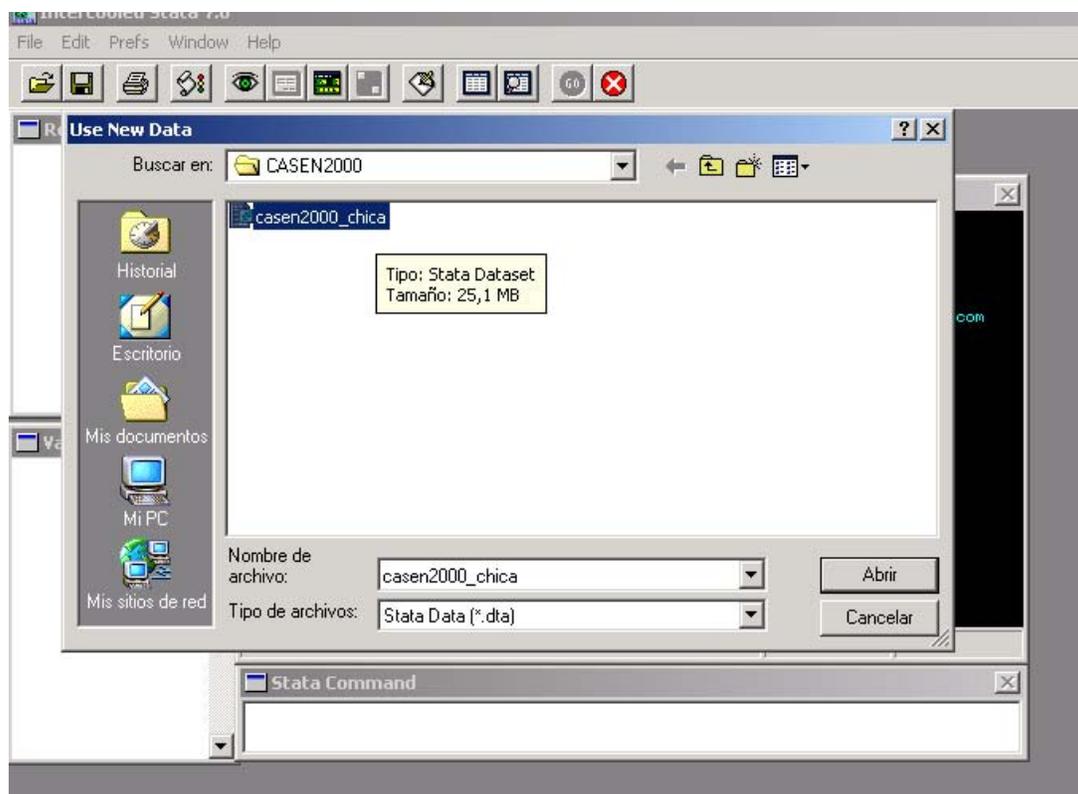
name. dta	Base de datos en stata
name. log	Archivo en el que podemos guardar comandos y resultados, excepto gráficos.
name. do	Archivo con comandos, generalmente hecho en el editor de stata, block de notas o word (pero guardado con formato sólo texto). Nos permite hacer un programa para luego hacerlo correr en stata, lo cual es más rápido que trabajar en forma interactivo.
name. ado	Archivo que contiene un mini programa ya insertado a stata, el cual nos permite hacer cálculos sin tener que programar en el momento las instrucciones para hacer por ej. un índice, sino que simplemente ocupamos un comando que nos hace funcionar el programa. Éste está inserto en la librería de stata.

1.1 Abrir la base de datos

El insumo básico para el trabajo estadístico es la base de datos. En este programa no se puede abrir la base directamente (desde el explorador, por ejemplo). Antes de abrir una base de datos (y de comenzar a trabajar) es necesario colocar memoria al programa. Para ello hacemos:

```
. set mem 300m
```

Es importante notar que se ha colocado una cantidad de memoria referencial, que permite abrir la base CASEN completa. Dependiendo de la velocidad y capacidad de procesamiento del computador en que se trabaje se podrá trabajar en la base o no. Si su computador no tiene mucha potencia, se recomienda reducir el tamaño de la base utilizando el programa "STATA TRASFER". (Ver anexo1)



Para abrir una base .dta debemos hacer (como se ve en figura anterior):
 →File→Open→abrir

o bien tipear directamente:

. use " [disco en que lo guardaremos] \ [ruta de acceso] \ [nombre de archivo.dta]", clear
 por ejemplo:

. use "C:\Mideplan\Casen\Casen2003.dta", clear

Luego, para guardar una base de datos, reescribiendo sobre la base antigua, hacemos:

. save " [disco en que lo guardaremos] \ [ruta de acceso] \ [nombre de archivo.dta]", replace
 Si queremos ver los datos de la base, podemos hacer click en el *Data browse*, y si queremos modificar los datos debemos abrir el *Data editor*.

1.2 Archivo log:

Para guardar todo lo que hagamos podemos hacer un "log", el cual puede ser abierto después en un procesador de texto. Para ello hacemos:

→open log→open

o bien tipeamos directamente:

log using [disco en que lo guardaremos] \ [ruta de acceso] \ [nombre de archivo.log]

por ejemplo:

log using C:\mideplan\clase1.log

Aquí agregamos `,replace` para que cada vez que hagamos correr un programa `.do` los resultados se reescriban sobre lo hecho antes.

log using C:\mideplan\clase1.log, replace

Para terminar de usar el log debemos tipear:

`. log close`

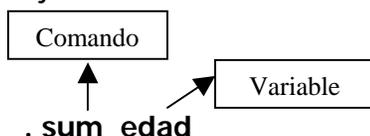
2. Principales comandos:

help (h) Este es un comando con el cual podemos obtener ayuda de stata. Si hacemos `help comando`, obtendremos ayuda específica del comando. Ej:
`. help graph`

describe (d) Este comando nos despliega un resumen de los contenidos de la base de datos que tenemos abierta.

summarize (sum) Computa estadísticas descriptivas

Ej:



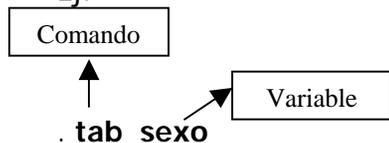
Variable	Obs	Mean	Std. Dev.	Min	Max
edad	257077	31.94582	21.39316	0	107

`. sum edad yopraj`

Variable	Obs	Mean	Std. Dev.	Min	Max
edad	257077	31.94582	21.39316	0	107
yopraj	86483	229203.7	517893.7	1500	5.40e+07

tabulate (tab) Computa frecuencias

Ej:



sexo	Freq.	Percent	Cum.
-----+-----			
hombre	127,505	49.60	49.60
mujer	129,572	50.40	100.00
-----+-----			
Total	257,077	100.00	

. tab pco1 sexo

Al incluir dos variables el comando hace un cruce de ambas.

Parentesco con el Jefe de hogar	sexo		
	hombre	mujer	Total
-----+-----+-----			
jefe(a) de hogar	51,699	16,454	68,153
cónyuge o pareja	2,709	45,155	47,864
hijo(a), hijastro(a)	54,775	49,600	104,375
padre o madre	350	1,426	1,776
suegro(a)	313	853	1,166
yerno o nuera	2,020	1,729	3,749
nieto(a)	9,932	9,112	19,044
hermano(a)	1,491	1,337	2,828
cuñado(a)	567	375	942
otro familiar	2,585	2,290	4,875
no familiar	1,028	647	1,675
s. doméstico p.adentr	36	594	630
-----+-----+-----			
Total	127,505	129,572	257,077

2.1 Creación de Variables

generate(gen) Este es un comando que nos permite crear variables, las cuales pueden ser una combinación de otras o una variable totalmente nueva. Por ejemplo, para crear una variable totalmente nueva puedo hacer:

```
. gen a=0
```

y en el editor de stata coloco el valor de la nueva variable.

Si quiero crear una variable a partir de otra puedo hacer:

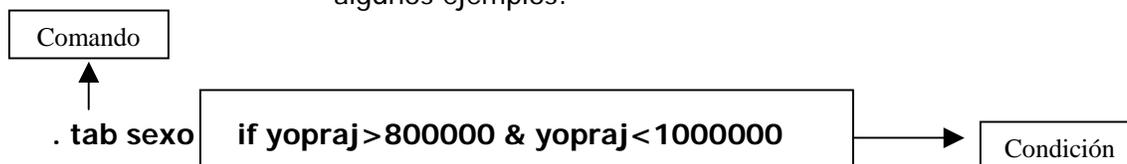
```
. gen b=3 + esc
```

Expresiones lógicas y matemáticas más usadas

+	suma	>=	Mayor o igual	&	y
-	resta	>	Mayor estricto que		o
*	multiplicación	<=	Menor o igual que	exp()	exponencial
/	división	<	Menor estricto que	log()	logaritmo
= =	igual	!=	Distinto que		

2.2 Expresiones condicionales

if Utilizar *if* nos permite hacer una instrucción condicional, ya sea con un comando descriptivo o de creación. Veremos más claramente esto mediante algunos ejemplos:



sexo	Freq.	Percent	Cum.
hombre	490	74.70	74.70
mujer	166	25.30	100.00
Total	656	100.00	

```
. gen g=0 if esc <= 8
```

(168115 missing values generated)

```
. replace g=1 if esc >= 9
```

(168115 real changes made)

by() Este es un conectivo que también nos sirve tanto para comandos de estadística descriptiva como para comandos de creación.

Ej:

```
. sort sexo
. by sexo: tab rama. egen h= sum(yopraj)
. by(comu). sum hVariable
```

2.3 Algunos Comandos Adicionales

drop (drop) Borra la variable especificada

Ej:

```
. drop sexo
```

preserve, restore Con estos comando puedo borrar una variable temporalmente, y luego recuperarlas.

Ej:

```
preserve
drop edad
restore
```

Borro temporalmente la variable edad. Al poner *restore*, la vuelvo a recuperar en la base

replace (replace) Con este comando puedo reemplazar una variable por otra, o el valor de la variable en un caso determinado por otro.

Ej:

```
. replace a=b
```

rename (rename) Con "rename" puedo cambiarle el nombre a una variable. Para ello puedo hacer:

```
. rename nombre antiguo nombre nuevo
```

egen (egen) Este comando también nos permite crear variables, pero a partir de la aplicación de una función conocida por stata sobre alguna variable en particular.

Ej:

```
. egen e = max(edad)
```

```
. sum e
```

Variable	Obs	Mean	Std. Dev.	Min	Max
e	257077	107	0	107	107

sort (sort) Nos permite ordenar la variable especificada en orden ascendente.

graph (graph) Nos permite hacer gráficos. Para llevarlos a word sólo debemos copiar y pegarlos, esto es:

→edit→copy

Para hacer un gráfico de barras:

```
. sort pco1
. graph yopraj, bar by(pco1)
```

Un “truco” muy utilizado para que las tablas aparezcan ordenadas en word, es colocar todo en letra tamaño 9.

2. Base de Datos

Trabajaremos con una parte de la encuesta **Casen 2003**. Es fundamental para ello saber relacionar los resultados que buscamos con las respuestas del cuestionario.

Comenzaremos por hacer un breve análisis del cuestionario para familiarizarnos con éste.

Identificadores:

segmento: Segmento al cual pertenece la encuesta (subdivisión de las comunas)
f : Folio de la encuesta
r : Región
z: Zona (1:urbano; 2:rural)
o: Orden de la persona dentro de la encuesta

Hogar:

La encuesta Casen es dirigida al jefe de hogar, por lo que generalmente contesta él o su cónyuge por todos los miembros de la familia. Por esto hay que tener cuidado con las conclusiones que se saquen a partir de ciertas preguntas, porque como no es el afectado necesariamente quien las responde, puede haber sesgo.

En cada encuesta hay sólo un hogar. Éste lo constituyen las personas que viven bajo el mismo techo y comparten el presupuesto para comida. En un mismo hogar pueden haber más de un núcleo, por lo que es necesario distinguir entre el jefe de hogar y el jefe de núcleo. Usualmente se trabaja más con la variable “relación con el jefe de hogar” (pco1).

Educación

Para conocer el nivel de educación de cada miembro del hogar se le pregunta cuál es el curso y tipo de estudio actual (para los que están estudiando) o último curso aprobado (para los que no están estudiando). A partir de estas dos variables (e7t y e7c) se construye la variables "años de escolaridad" (en la base de clases: esc). Así, si por ejemplo:

Si una persona tiene como último curso probado 2° medio técnico profesional: e7t=8 y e7c=2. En este caso $esc = 8 + 2 = 10$ (ya que como está en enseñanza media tiene 8 años aprobados + los dos de enseñanza media)

Ingresos

Los ingresos de la persona se dividen en una serie de categorías, ya sea tomando en cuenta el ingreso de la ocupación principal (yopraj) o agregándole otra serie de ingresos (ing. de otros trabajos, subsidios, regalías, etc.) que junto al yopraj constituyen el ingreso monetario total por persona (ymoneaj).

Para conseguir el ingreso total familiar debemos tomar otro ingreso más en cuenta: alquiler imputado por la casa en que vive (si es que no paga arriendo por ella): yaimhaj

Luego, el ingreso total familiar será:

$ytothaj = \sum ymoneaj + yaimhaj$

Empleo

Para saber si una persona está con empleo se le pregunta si trabajó la semana pasada (o1), y luego se corrobora su situación preguntándole (a los que contestaron que no antes) si es que tenían empleo, pero estaban ausentes temporalmente (o2). A los que contestaron no, se les pregunta si buscó trabajo remunerado en los últimos dos meses (o3). Con todo esto podemos hacer un índice de tasa de participación en el mercado laboral.

Características del empleo.

En esta encuesta hay tres variables que nos indican características del empleo:

Oficio (viene de la pregunta o7): oficio

Actividad de la empresa (viene de la pregunta o8), con la cual podemos crear la variable sector económico (agrupando por actividad): rama

Categoría ocupacional (viene de la pregunta o9), nos permite separar entre asalariados y no asalariados, y distinguir categorías particulares: o9

Factor de expansión:

En esta encuesta se utiliza un "factor de expansión" para hacer la muestra representativa. Así, cada encuesta será ponderada por el factor que le corresponda [$w=expr$]. Por esto, cada vez que

queramos hacer estadística descriptiva, debemos ocupar este factor de expansión, de lo contrario nuestros resultados no serán representativos.

Ya habíamos visto la forma general de calcular estadísticas, pero como nuestra base utiliza factor de expansión, entonces la forma de hacer estadísticos es, por ejemplo:

```
. tab sexo [w=expr]
. tab pco1 [w=expr] if yopraj !=.
```

Formando Hogares.

Muchos análisis de pobreza y distribución se hacen a nivel de hogares, por lo que es fundamental que veamos la forma de agrupar cada hogar.

En primer lugar es necesario agrupar a todos los miembros de un hogar a través de una clave (o número) que los identifique. Así creamos una "macro" que da a cada miembro de un mismo hogar (misma encuesta) un mismo número, utilizando los identificadores antes vistos. Para ello utilizamos la instrucción:

```
. macro define id "identificadores"
```

lo que en nuestro caso sería:

```
. macro define id "segmento folio"
```

Luego, si quisiéramos contar la cantidad de personas que constituyen cada hogar podemos hacer:

```
. egen nt= count(f) , by ($id)
```

Ej:

Construir ingreso per cápita:

```
. macro define id "segmento folio"
```

```
. egen nt= count(f) , by ($id)
```

```
. gen ytothaj= ymonehaj + yaimhaj
```

```
. gen ypc=ytothaj/nt
```

3. Archivos .do

En esta sección haremos un programa .do, con él podremos hacer un programa (para quienes están familiarizados con SPSS corresponde a una sintaxis) y hacerlo correr en stata, lo cual es un método mucho más rápido que el interactivo. El do se puede hacer en el editor de stata, en word (guardándolo como sólo texto) o en el block de notas. En estos ejemplos trabajaremos en el block. Trate de entender cada una de las instrucciones que se presentan, cuándo hay que modificarlas, y la pertinencia de cada una de ellas, ya que este no es un programa tan estándar, y cada vez que usted quiera programar, deberá analizar bien qué es lo que buscar, y una vez que tenga claro cómo se "haría a mano", puede comenzar a programar.

Haremos un ejemplo en que buscamos analizar hogares:

Primero indicamos la base que usaremos y luego el log:

```
use c:\mideplan\casen\casen2003.dta
log using c:\MIDEPLAN\ejercicio1.log, replace
```

Para trabajar con estadísticas por hogares usualmente se excluye al servicio doméstico puertas adentro, por lo que hacemos:

```
. preserve
. drop if nucleo==0
```

Luego anidamos hogares:

```
. macro define id "segmento folio"
. egen nt= count(f) , by ($id)
```

Si queremos hacer indicaciones para guiarnos en el programa podemos escribir cosas entre asteriscos. Por ejemplo:

```
* excluye servicio dom adentro *
```

Generamos el ingreso per cápita por hogar:

```
**ingreso per cápita simple**
. gen ytothaj= ymonehaj + yaimhaj
. gen ypc=ytothaj/nt
. restore
```

Forma paralela de calcular el ingreso per cápita:

```
*** Ingreso per cápita número 2, considera servicio domestico puertas adentro como hogar independiente, corrige por servicio domes externo, que se había excluido de otros procedimientos*****
```

```
. egen nt2= count(f) if nucleo!=0, by ($id)
. sum nt2 [w=expr]
. replace ytothaj = yopraj if nucleo!=0
. gen ypc2=ytothaj/nt2
. replace ypc2 = yopraj if nucleo!=0
```

Al guardarlo, debemos guardarlo con un nombre *.do, y si trabajan en word, asegurarse que escogen la alternativa "tipo de archivo: sólo texto"

Para hacer correr el programa debemos hacer:
do disco\ubicación del archivo

es decir:

```
do c:\mideplan\programa1.do
```

ANEXO 1

El programa "Stata transfer" permite traspasar una base de datos desde formato SPSS o excel a una base de datos de stata. Dentro de las opciones se pueden escoger el número de variables que se pueden incluir en la nueva base que se generará. Por ejemplo si tengo una base en stata que es muy "pesada", puedo pasarla a través de stata transfer y seleccionar las variables con las que necesito para trabajar.

