

El Uso de la CASEN para el Análisis Socioeconómico y una Introducción a Stata

Damian Clarke
Análisis Económico y Social usando Stata



UNIVERSIDAD DE CHILE

18–20 Agosto, 2020

Contenidos

Encuestas de Hogar en la Economía

Encuestas de Hogar

La CASEN

Idiomas de programación estadística

Modos de uso de Stata y syntaxis general

Cómo Conseguir Ayuda?

Trabajando en Stata

1. Encuestas de Hogar en la Economía

Medición de Bienestar en Economía

Para entender el comportamiento humano y cambios en bienestar social, es necesario *medir*. Existen diversas fuentes de datos frecuentemente utilizadas.

- ▶ Indicadores a nivel agregado
- ▶ Datos administrativos
- ▶ Datos generados a partir de plataformas web/textos/fuentes externas
- ▶ Censos
- ▶ ⋮
- ▶ **Encuestas de hogar**

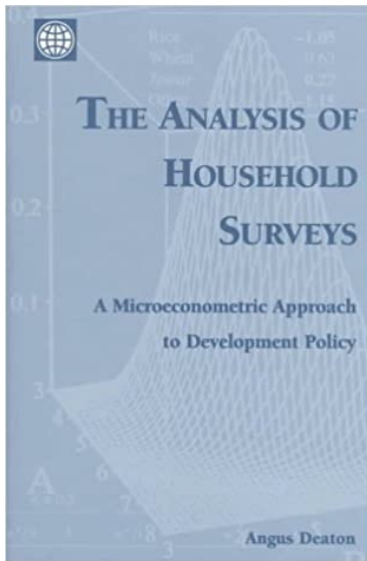
Encuestas de Hogar

Las encuestas de hogar son una de las herramientas principales para medir y entender el comportamiento humano en investigaciones en ciencias sociales.

- ▶ Permiten tener una medida mucho más extendido acerca de las personas
- ▶ Recopilan información *auto-reportado* que sería difícil observar sin preguntar directamente
- ▶ Son diseñadas para permitir inferencia acerca de la población del país o algún sub-*area*
 - ▶ Esto requiere un marco muestral diseñado específicamente para hacer esto
 - ▶ Hay varias consideraciones prácticos importantes: clusterización, estratificación, factores de expansión
- ▶ Encuestas proporcionan medidas *con errores muestrales*

Por Ejemplo...

Los errores muestrales efectivos a nivel nacional son 0,4 puntos porcentuales en el caso del error absoluto y 4,3%, el error relativo, para la estimación de la tasa de pobreza por ingreso (CASEN 2017, Anexo 1, Manual del investigador – Guía práctica para el uso y análisis de información).



** Leemos introducción y las secciones 1.1-1.3 (pdf del libro gratis/libre en línea!)
<https://openknowledge.worldbank.org/handle/10986/30394>

El Libro de Deaton

Este libro está enfocado en encuestas de hogar, específicamente relacionados con el tema de desarrollo económico.

"I do not see how one can look at figures like these without seeing them representing possibilities. Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what exactly? If not, what is it about the "nature of India" that makes it so? The consequences for human welfare involved in questions like these are simply staggering: once one starts to think about them, it is hard to think about anything else."

Robert E. Lucas, "On the Mechanics of Economic Development" *Journal of Monetary Economics* 22(1), 1988, p. 5.

Encuestas de Hogar

También son herramientas fundamentales utilizadas por entidades públicas para medir cambios a través del tiempo

- ▶ Por ejemplo, la tasa de población viviendo abajo de la línea de pobreza es un indicador público/político importante
- ▶ Pero ojo con la “Ley de Goodhart” y la “Ley de Cambell”: *“The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”*
- ▶ A menudo, las encuestas de hogar son financiados por gobiernos (eg DHS, CPS, CASEN, EPS, ...)

Objetivo e información CASEN

- ▶ Conocer periódicamente la situación socioeconómica de los hogares y de la población que reside en viviendas particulares, recolectando información de aspectos como:
 - ▶ composición de los hogares y familias
 - ▶ salud,
 - ▶ vivienda
 - ▶ trabajo e ingresos
- ▶ Evaluar el impacto de la política social mediante la estimación la cobertura, la focalización y la distribución del gasto fiscal de los principales programas sociales en el ingreso de los hogares y su distribución
- ▶ http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Manual_del_Investigador_Casen_2017.pdf

Periodicidad y cobertura CASEN

- ▶ La Encuesta CASEN es realizada desde el año 1990 con una periodicidad bianual o trianual
- ▶ Hasta ahora, las encuestas aplicadas corresponden a los años: 1990, 1992, 1994, 1996, 1998, 2000, 2003, 2006, 2009, 2011, 2013, 2015 y 2017
- ▶ Su cobertura abarca el territorio nacional, exceptuando algunas zonas muy alejadas o de difícil acceso
- ▶ Por ejemplo, en el año 2017 el proceso de levantamiento de la encuesta se realizó en todas las regiones del país, en 324 comunas

Elementos estadísticos CASEN

- ▶ La población objetivo de la encuesta es la población que reside en viviendas particulares a lo largo del territorio nacional
 - ▶ Al interior de cada vivienda seleccionada, se entrevista a todos los hogares y se consulta por todos los residentes habituales
- ▶ Selección de viviendas con la finalidad de representar de manera adecuada a la población de interés a nivel nacional y regional
 - ▶ De acuerdo con el diseño de la muestra, corresponde aplicar este factor de expansión a cada hogar y persona seleccionada, lo que depende del número de viviendas que tiene la sección geográfica y el número de secciones que tiene el estrato
 - ▶ El factor de expansión corresponde a las variable *EXPR* si el análisis de los datos se realiza a a nivel nacional, regional o zona urbana-rural

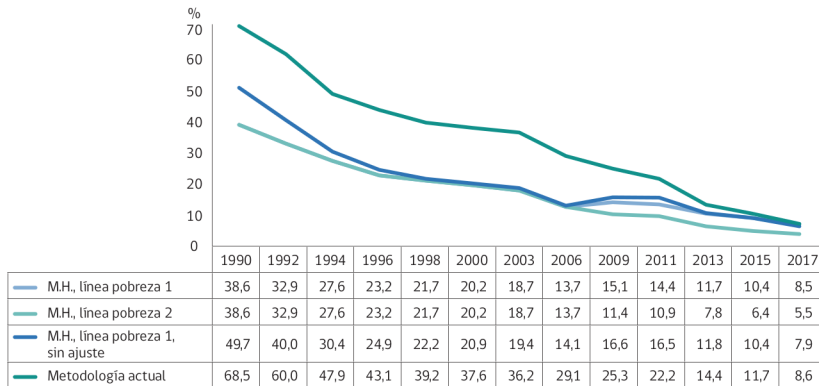
Elementos estadísticos CASEN

Figure: Distribución de la muestra Casen 2017 por región

nº	Región	Viviendas
XV	Región de Arica y Parinacota	2.408
I	Región de Tarapacá	2.974
II	Región de Antofagasta	2.511
III	Región de Atacama	2.331
IV	Región de Coquimbo	3.028
V	Región de Valparaíso	6.717
VI	Región del Libertador Gral. Bdo. O'Higgins	5.099
VII	Región del Maule	5.007
VIII	Región del Biobío	6.901
XVI	Región de Ñuble	2.834
IX	Región de La Araucanía	5.136
XIV	Región de Los Ríos	3.624
X	Región de Los Lagos	4.129
XI	Región Aysén del Gral. Carlos Ibáñez del Campo	1.862
XII	Región de Magallanes y la Antártica Chilena	2.301
XIII	Región Metropolitana de Santiago	12.954
Total	País	69.816

La CASEN – Evolución de Incidencia de Pobreza por Ingresos

(Porcentaje, personas)

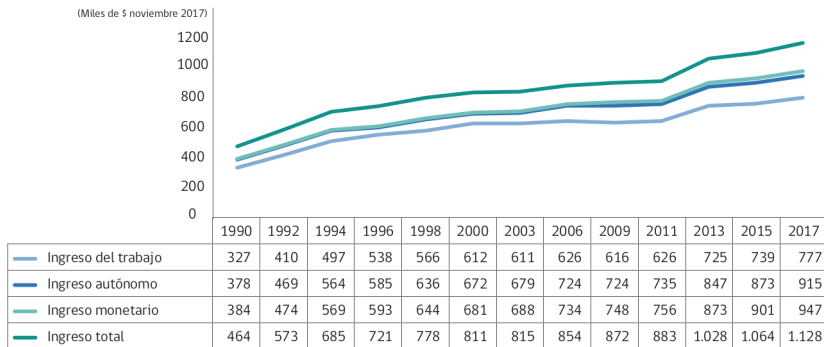


Nota: M.H: Metodología histórica.

Fuente: Ministerio de Desarrollo Social y Familia, Observatorio Social, Encuesta Casen 1990-2017 y CEPAL. Información del Sistema de Cuentas Nacionales provista por el Banco Central. Para el año 2017, se emplea información preliminar, disponible a la fecha

La CASEN – Evolución de Ingreso Promedio Real de los Hogares

(Miles de \$ de noviembre 2017)



Fuente: Ministerio de Desarrollo Social y Familia, Observatorio Social, Encuesta Casen 1990-2017

Otras encuestas a hogares en Chile

- ▶ Encuesta de Presupuestos Familiares (EPF), Instituto Nacional de Estadísticas (INE)
- ▶ Encuesta Nacional de Empleo (ENE), INE
- ▶ Encuesta de población activa (EPA), INE
- ▶ Encuesta de Protección Social (EPS), Subsecretaría Previsión Social
- ▶ Encuesta Nacional de Victimización por Violencia Intrafamiliar y Delitos Sexuales (ENVIF), Centro de Estudios y Análisis del Delito (CEAD)
- ▶ Encuesta de Ocupación y Desocupación en el Gran Santiago (EOD), Centro de Microdatos, Universidad de Chile
- ▶ Encuesta Financiera de Hogares (EFH), Banco Central de Chile

2. Idiomas de programación estadística

Idiomas de programación estadística

Existen muchas distintas opciones al elegir un programa de computación estadístico

Varias Consideraciones/Diferencias

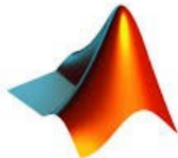
- ▶ Costo
- ▶ Nivel de apoyo/calidad de documentación
- ▶ Código fuente abierta/cerrada
- ▶ Fortalezas en distintas áreas
- ▶ Nivel de actividad de comunidad de usuarios (librerías contribuidas)
- ▶ Aspectos de redes (¿qué usan mis colaboradores?)

STATA



GNU Octave

julia



MATLAB



3. Modos de uso de Stata y syntaxis general

Stata/MP 15.1



Review

Filter commands here

#	Command	rc
---	---------	----

Results

Filter results here

(R)

15.1

Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user 4-core Stata perpetual license:
Serial number: 501506301632
Licensed to: Damian Clarke
USACH

Notes:
1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set_maxvar](#).

running /usr/local/stata15/profile.do ...

.

Variables

Filter variables here

#	Name	Label
---	------	-------

Properties

Name

Label

Type

Format

Value label

Notes

▶ Data

/home/damianINS

_____ (R)
/ / / / /
/ / / / /
Statistics/Data Analysis 15.1
MP - Parallel Edition
Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user 4-core Stata perpetual license:

Serial number: 501506301632
Licensed to: Damian Clarke
Universidad de Chile

Notes:

1. Unicode is supported; see `help unicode_advice`.
2. More than 2 billion observations are allowed; see `help obs_advice`.
3. Maximum number of variables is set to 5000; see `help set_maxvar`.

. █

Modos de uso de Stata y syntaxis general

`nombre_comando` *argumentos* [`ponderadores`] [`if`] [`in`] [, *opciones*]

- ▶ **nombre_comando**: Comandos definidos por Stata (eg use, summarize, regress)
- ▶ **argumentos**: Los “insumos” directos a los comandos
- ▶ **ponderadores**: “Pesos” para distintas observaciones en los comandos (eg [pweight=factor])
- ▶ **if**: Condiciones de inclusión en el comando (eg if anio==2017)
- ▶ **in**: Rango de observaciones utilizados en el comando (eg in 1/10)
- ▶ **opciones**: Otras opciones para el comando (eg robust, clear)

** [] indica que son elementos opcionales

Algunos Ejemplos Ilustrativos

```
use casen2017.dta
summarize ytrabajocor [pw=expr] if anio==2017
regress ytrabajocor edad [pw=expr] if anio==2017, robust
save salarios_promedios.dta, replace
```


4. Cómo Conseguir Ayuda?

Cómo Conseguir Ayuda?

La documentación de Stata realmente es increíble. Pero también existen varios otros canales para resolver problemas.

- ▶ Documentación de Stata (`help`, también en línea)
- ▶ Foros en línea
 - ▶ Statalist
 - ▶ StackOverflow
- ▶ Los mismos mensajes de error de Stata
- ▶ Algunas opciones más técnicas/para consultas más avanzadas
 - ▶ Stata Journal/Stata Technical Bulletin
 - ▶ `capture`

Viewer - help use

Navigation icons: back, forward, search, print, etc. Search bar: help use

help use ✕

Dialog ▾ Also see ▾ Jump to ▾

[D] use — Load Stata dataset
([View complete PDF manual entry](#))

Syntax

Load Stata-format dataset

```
use filename [, clear nlabel]
```

Load subset of Stata-format dataset

```
use [varlist] [if] [in] using filename [, clear nlabel]
```

Menu

File > Open...

Description

use loads into memory a Stata-format dataset previously saved by **save**. If *filename* is specified without an extension, **.dta** is assumed. If your *filename* contains embedded spaces, remember to enclose it in double quotes.

In the second syntax for **use**, a subset of the data may be read.

Cómo Conseguir Ayuda?

Muchas veces el principal desafío es saber qué preguntar para encontrar ayuda

- ▶ Aquí el comando `search` puede ser útil
- ▶ Generalmente los nombres de los comandos 'hacen sentido', pero a veces una pregunta es más elaborada
- ▶ Con una pregunta clara, a menudo una búsqueda en línea da buenos recursos
 - ▶ Por ejemplo "cómo abro un archivo csv en Stata?"
- ▶ Hay una comunidad activa de usuari@s en Stata que quieren ayudar, pero fíjense en normas de foros en línea

5. Trabajando en Stata

Trabajando en Stata

Metas para ahora en Stata...

1. Uso básico de Stata y sus lugares de trabajo
2. Entender la estructura de directorios, y cómo mover entre directorios
3. Entender cómo importar y exportar datos de diversos tipos
4. Algunas detalles iniciales de datos en Stata

Uso básico de Stata y sus lugares de trabajo (1)

Stata tiene varias ventanas de trabajo: la línea de comando, la ventana de salida, un editor de texto, etc.

- ▶ Hasta versión 15 de Stata, solo se permite trabajar con una base de datos (o planilla) a la vez
- ▶ Adentro de la base de datos, a menudo tenemos columnas que son “variables” y filas que son observaciones
 - ▶ Esto implica que todas tus “variables” tienen que estar incorporado en una planilla
 - ▶ O se puede ir cambiando entre planillas para sacar información necesaria
 - ▶ Implica que a menudo queremos hacer bastante de trabajo de unir/resumir/juntar varias bases en una
 - ▶ Volveremos a ver todo esto en la unidad 4 del curso
 - ▶ Esto es diferente a otras idiomas (como R, Python, ...) donde se puede tener varias planillas a la vez
 - ▶ Esto tiene ventajas y desventajas

Uso básico de Stata y sus lugares de trabajo (2)

Hay muchas maneras de “visualizar” los datos que están en Stata en un momento determinado

- ▶ Se puede examinar todos los datos que están en memoria (`browse`, `list`)
- ▶ Todos los datos abiertos en Stata están guardados en memoria virtual (RAM) no en el disco duro al momento de trabajar
- ▶ Stata trabaja de forma muy simple con datos que son numéricas y datos que son de caracteres de texto

Data Editor (Browse) - Casen 2017.dta

folio[1] 110110020201

	folio	o	id_vivienda	hogar	region	provincia	c
1	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
2	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
3	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
4	1.101e+11	2	1.101e+09	1	Región ...	Iquique	
5	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
6	1.101e+11	2	1.101e+09	1	Región ...	Iquique	
7	1.101e+11	3	1.101e+09	1	Región ...	Iquique	
8	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
9	1.101e+11	2	1.101e+09	1	Región ...	Iquique	
10	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
11	1.101e+11	2	1.101e+09	1	Región ...	Iquique	
12	1.101e+11	3	1.101e+09	1	Región ...	Iquique	
13	1.101e+11	4	1.101e+09	1	Región ...	Iquique	
14	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
15	1.101e+11	2	1.101e+09	1	Región ...	Iquique	
16	1.101e+11	3	1.101e+09	1	Región ...	Iquique	
17	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
18	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
19	1.101e+11	2	1.101e+09	1	Región ...	Iquique	
20	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
21	1.101e+11	1	1.101e+09	1	Región ...	Iquique	
22	1.101e+11	2	1.101e+09	1	Región ...	Iquique	

Variables

Filter variables here

	Name	Label
☒	folio	Identificación hogar (c
☒	o	Orden
☒	id_vivienda	Identificador de la vivier
☒	hogar	Identificación del hoga
☒	region	Región
☒	provincia	Provincia
☒	comuna	Comuna
☒	zona	Zona
☒	expr	Factor de Expansión Re
☒	expc	Factor de Expansión Co
☒	expr_div	Factor de expansión or
☒	varstrat	Estratos de Varianza

Properties

Name: folio

Label: Identificación hogar (c

Type: double

Format: %10.0g

Value label:

Notes: No notes

► Data

Ready Vars: 804 Order: Dat... Obs: 216,439 Filter: Off Mode: Browse

Manejando por el sistema operativo

Hay 3 comandos que son bastante fundamentales al estar interactuando con el sistema operativo desde Stata:

- ▶ `pwd`: escribir directorio actual (*print working directory*)
- ▶ `cd`: cambiar directorio
- ▶ `ls`: escribir contenidos del directorio actual

¿Donde “busca” Stata en mi sistema?

```
. adopath
[1]  (BASE)      "/usr/local/stata15/ado/base/"
[2]  (SITE)      "/usr/local/ado/"
[3]              "."
[4]  (PERSONAL)  "~/ado/personal/"
[5]  (PLUS)      "~/ado/plus/"
[6]  (OLDPLACE)  "~/ado/"
```

Trabajando con Datos en Stata

Existen varias maneras de importar datos a Stata dependiendo del formato de los datos. Puede ser ilustrativo ver el manual aquí:

<https://www.stata.com/manuals13/u21.pdf>

- ▶ Si los datos ya están en formato Stata (.dta), es simple: use
 - ▶ Pero aún así, hay varias posibles extensiones...
 - ▶ webuse, sysuse
- ▶ Si los datos están en otro formato, la importación depende del formato
 - ▶ Excel: `import excel`
 - ▶ csv, txt: `import delimited`
 - ▶ Otros formatos menos comunes: `infile`, `odbc`
 - ▶ En el caso de otros formatos de programas externas (eg Rdat), es más complejo

```
. input y x
```

y

x

```
1. 2 5
```

```
2. 3 4
```

```
3. 1 7
```

```
4. 8 4
```

```
6. end
```

```
. list
```

```
+-----+
```

```
| y   x |
```

```
|-----|
```

```
1. | 2   5 |
```

```
2. | 3   4 |
```

```
3. | 1   7 |
```

```
4. | 8   4 |
```

```
+-----+
```

Un Ejemplo Trabajado

```
webuse set http://www.damianclarke.net/teaching/stata/
```

```
webuse casen1990
```

```
⋮
```