

# Uso básico del software Stata

14° Taller de Investigación  
Universidad Austral – Sede Rosario

Dictado por: Lic. Jorge Camusso

2022

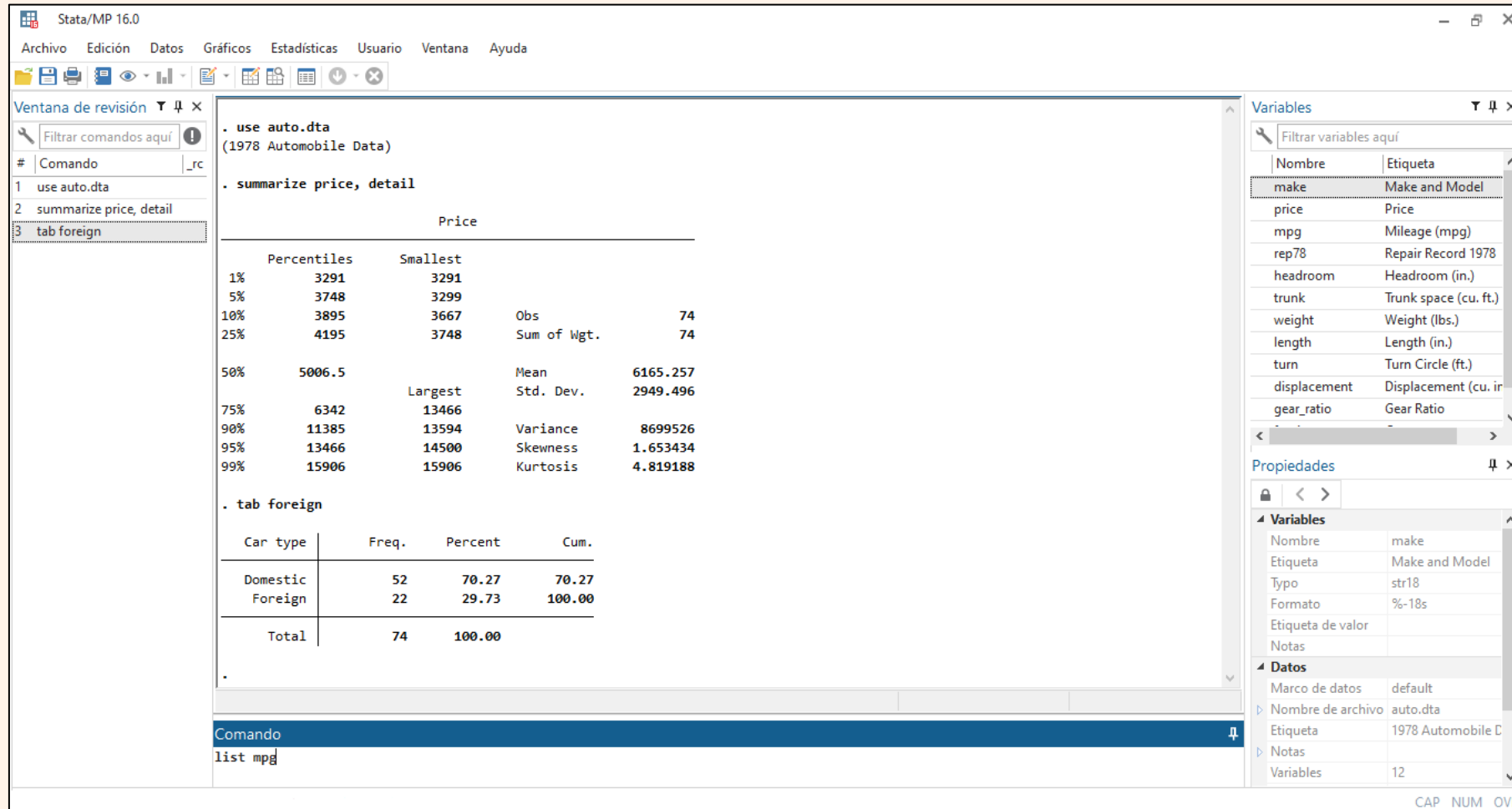


# ¿Qué es Stata?

Stata es un paquete de software estadístico orientado al análisis, manejo, descripción, visualización y reporte de datos.

- Web: <https://www.stata.com/>
- Última versión: 17
- Ediciones: Stata/BE, Stata/SE y Stata/MP
- Manual completo:  
<https://www.stata.com/manuals/r.pdf>
- Foros: <https://www.statalist.org/forums/>

# Elementos principales de la interfaz de Stata



Stata/MP 16.0

Archivo Edición Datos Gráficos Estadísticas Usuario Ventana Ayuda

Ventana de revisión

Filtrar comandos aquí

#	Comando	_rc
1	use auto.dta	
2	summarize price, detail	
3	tab foreign	

```
. use auto.dta
(1978 Automobile Data)

. summarize price, detail
```

Price			
Percentiles	Smallest		
1%	3291	3291	
5%	3748	3299	
10%	3895	3667	Obs 74
25%	4195	3748	Sum of Wgt. 74
50%	5006.5		Mean 6165.257
75%	6342	Largest 13466	Std. Dev. 2949.496
90%	11385	13594	Variance 8699526
95%	13466	14500	Skewness 1.653434
99%	15906	15906	Kurtosis 4.819188

```
. tab foreign
```

Car type	Freq.	Percent	Cum.
Domestic	52	70.27	70.27
Foreign	22	29.73	100.00
Total	74	100.00	

```
.
Comando
list mpg
```

Variables

Filtrar variables aquí

Nombre	Etiqueta
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu. in.)
gear_ratio	Gear Ratio

Propiedades

Variables

Nombre	make
Etiqueta	Make and Model
Typo	str18
Formato	%-18s
Etiqueta de valor	
Notas	

Datos

Marco de datos	default
Nombre de archivo	auto.dta
Etiqueta	1978 Automobile D
Notas	
Variables	12

CAP NUM OVR



# Bases de datos, *do-files* y *log-files*

- **Bases de datos:** Stata trabaja con bases de datos en formato .dta.
- ***Do-files*:** son archivos de texto sin formato (ASCII) que contienen una secuencia de comandos de Stata. El software interpreta estos comandos tal y como si hubieran sido ingresados en la ventana de comandos, siguiendo el orden de la secuencia. También es posible escribir comentarios explicativos acerca de qué se está haciendo con la secuencia de comandos del *do-file*, lo cual puede ser útil para el propio usuario o para un tercero.
- ***Log-files*:** son archivos que permiten mantener un registro de todo lo que aparece en la ventana de resultados. Registran tanto el historial de comandos como el historial de salidas.

# Ejemplo de do-file

```
1 // A Quick Tour of Stata
2 // Germán Rodríguez - Fall 2019
3 version 16
4 clear
5 capture log close
6 log using QuickTour, text replace
7 display 2+2
8 display 2 * ttail(20,2.1)
9 // load sample data and inspect
10 sysuse lifeexp
11 desc
12 summarize lexp gnppc
13 list country gnppc if missing(gnppc)
14 graph twoway scatter lexp gnppc, ///
15 title(Life Expectancy and GNP ) xtitle(GNP per capita)
16 // save the graph in PNG format
17 graph export scatter.png, width(400) replace
18 gen loggnppc = log(gnppc)
19 regress lexp loggnppc
20 predict plexp
21 graph twoway (scatter lexp loggnppc) (lfit lexp loggnppc) ///
22 , title(Life Expectancy and GNP) xtitle(log GNP per capita)
23 graph export fit.png, width(400) replace
24 list country lexp plexp if lexp < 55, clean
25 list gnppc loggnppc lexp plexp if country == "United States", clean
26 log close
27
```

# Ejemplo de *log-file*

```
name: <unnamed>
log: C:\Users\mydir\Documents\base.smcl
log type: smcl
opened on: 4 Feb 2019, 23:36:17

. sysuse auto
(1978 Automobile Data)

. by foreign, sort: summarize price mpg
```

---

```
-> foreign = Domestic
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	52	6072.423	3097.104	3291	15906
mpg	52	19.82692	4.743297	12	34

---

```
-> foreign = Foreign
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	22	6384.682	2621.915	3748	12990
mpg	22	24.77273	6.611187	14	41

```
. * be sure to include the above stats in report!
. * now for something completely different
. corr price mpg
(obs=74)
```

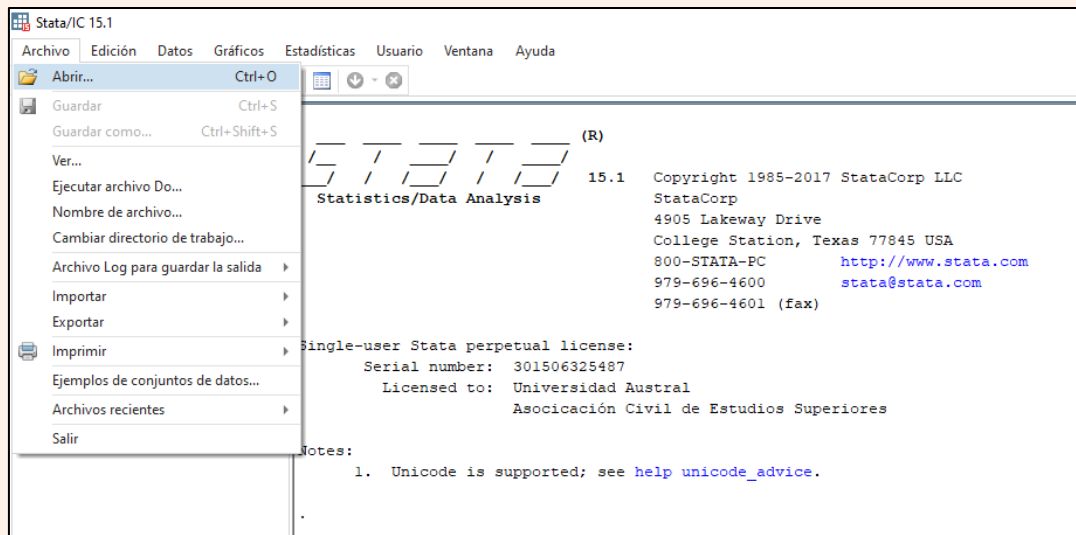
	price	mpg
price	1.0000	
mpg	-0.4686	1.0000

```
. log close
name: <unnamed>
log: C:\Users\mydir\Documents\base.smcl
log type: smcl
closed on: 4 Feb 2019, 23:37:41
```

# ¿Cómo cargar bases de datos .dta?

Si la base de datos con la que queremos trabajar está en formato .dta:

Stata también tiene precargadas bases que sirven como ejemplos para ejecutar diferentes comandos:

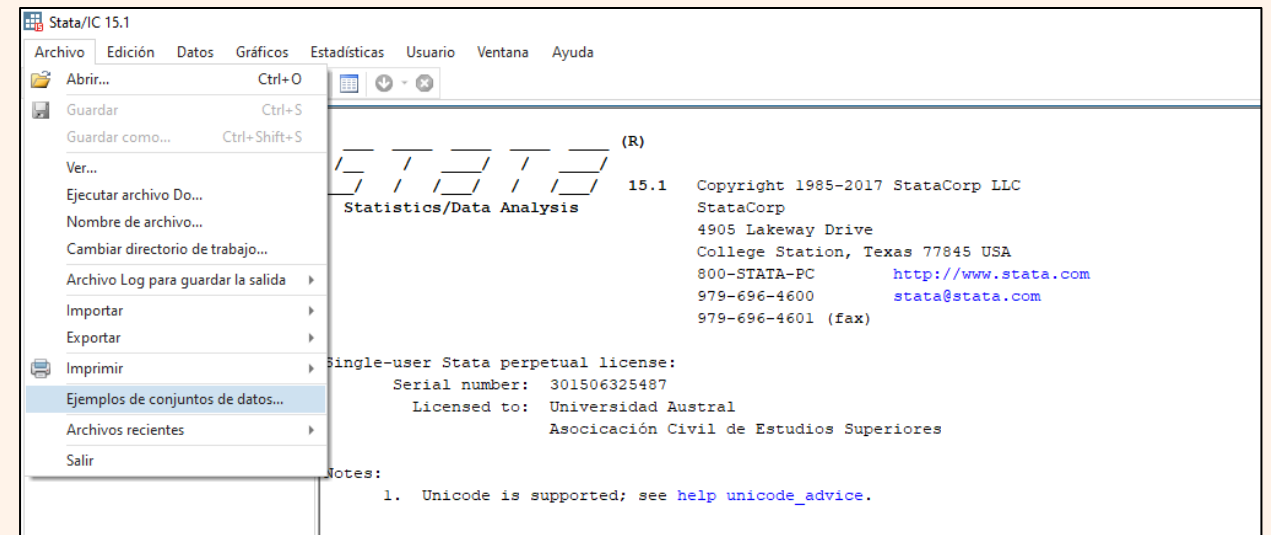


```
Stata/IC 15.1
Archivo | Edición | Datos | Gráficos | Estadísticas | Usuario | Ventana | Ayuda
Abrir... Ctrl+O
Guardar Ctrl+S
Guardar como... Ctrl+Shift+S
Ver...
Ejecutar archivo Do...
Nombre de archivo...
Cambiar directorio de trabajo...
Archivo Log para guardar la salida
Importar
Exportar
Imprimir
Ejemplos de conjuntos de datos...
Archivos recientes
Salir

----- (R)
-----
Statistics/Data Analysis 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 Stata perpetual license:
Serial number: 301506325487
Licensed to: Universidad Austral
Asociación Civil de Estudios Superiores

Notes:
1. Unicode is supported; see help unicode_advice.
```



```
Stata/IC 15.1
Archivo | Edición | Datos | Gráficos | Estadísticas | Usuario | Ventana | Ayuda
Abrir... Ctrl+O
Guardar Ctrl+S
Guardar como... Ctrl+Shift+S
Ver...
Ejecutar archivo Do...
Nombre de archivo...
Cambiar directorio de trabajo...
Archivo Log para guardar la salida
Importar
Exportar
Imprimir
Ejemplos de conjuntos de datos...
Archivos recientes
Salir

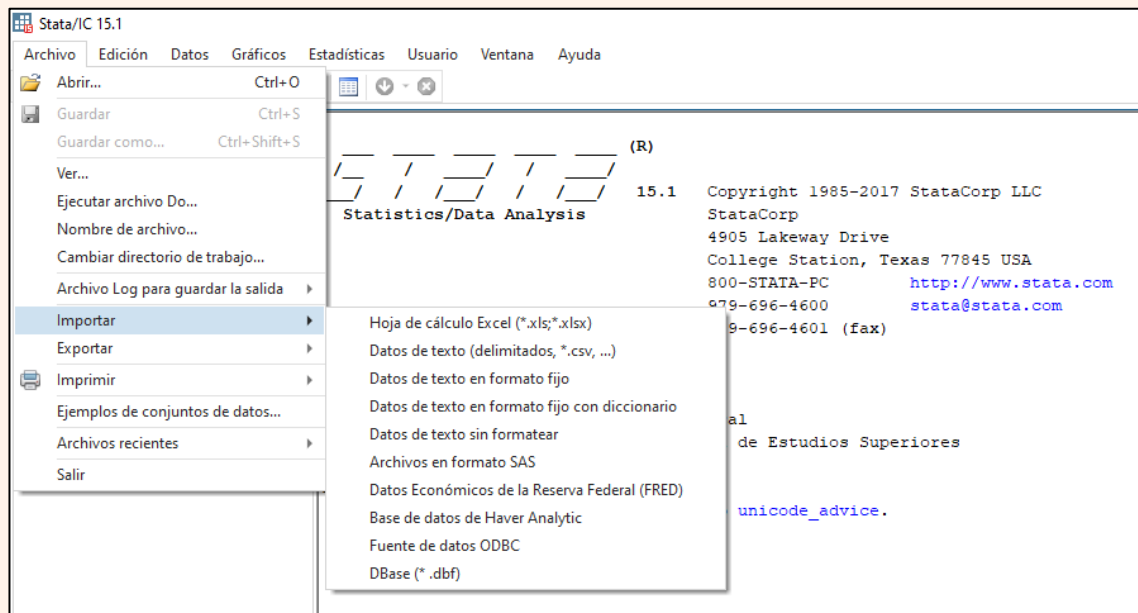
----- (R)
-----
Statistics/Data Analysis 15.1 Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user Stata perpetual license:
Serial number: 301506325487
Licensed to: Universidad Austral
Asociación Civil de Estudios Superiores

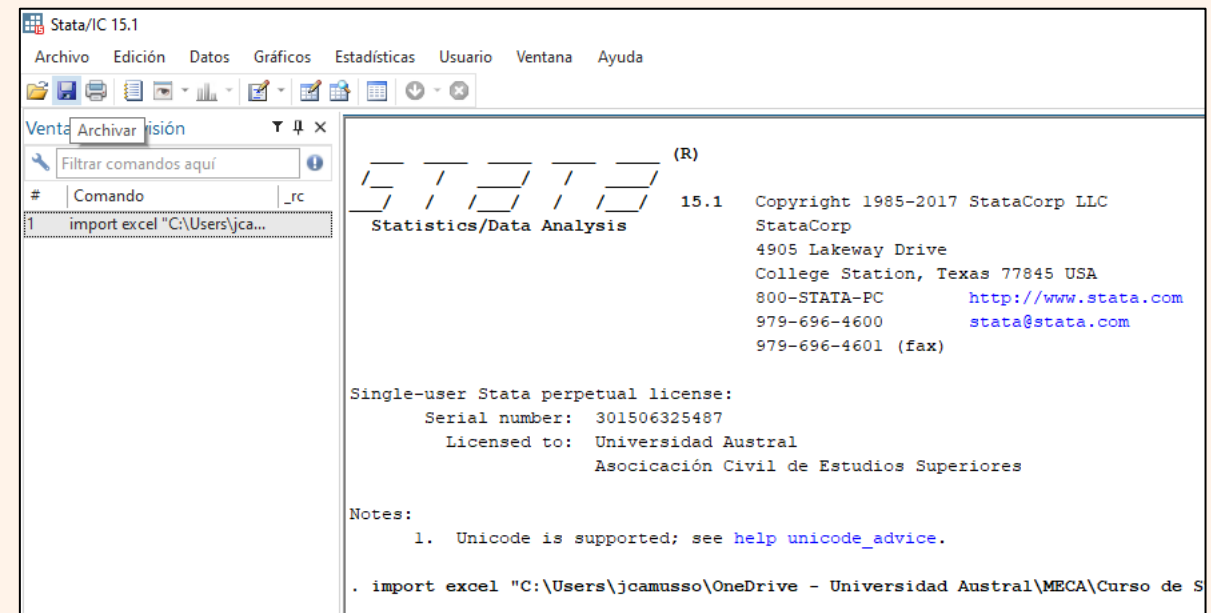
Notes:
1. Unicode is supported; see help unicode_advice.
```

# ¿Cómo importar bases de datos en otro formato?

En ocasiones la base de datos se encuentra en algún formato diferente (hoja de cálculo, texto con formato, texto sin formato, etc.). Si es así, deberemos importar este tipo de archivos. También podremos convertirlos en .dta.



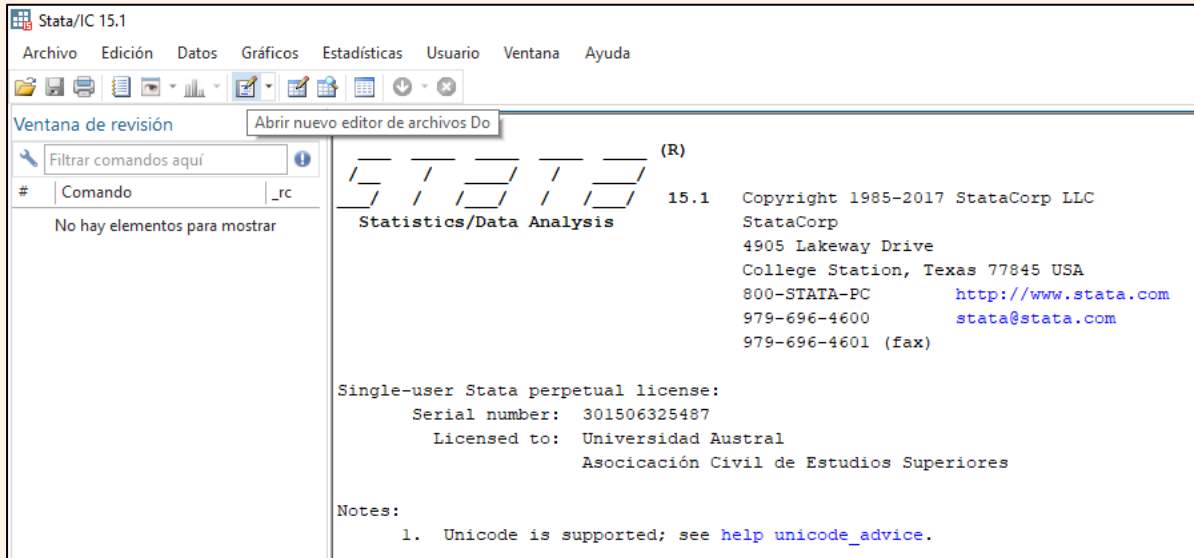
Importar



Guardar base importada en formato .dta



# ¿Cómo abrir un nuevo do-file?



Stata/IC 15.1

Archivo Edición Datos Gráficos Estadísticas Usuario Ventana Ayuda

Ventana de revisión

Abrir nuevo editor de archivos Do

Filtrar comandos aquí

#	Comando	_rc
No hay elementos para mostrar		

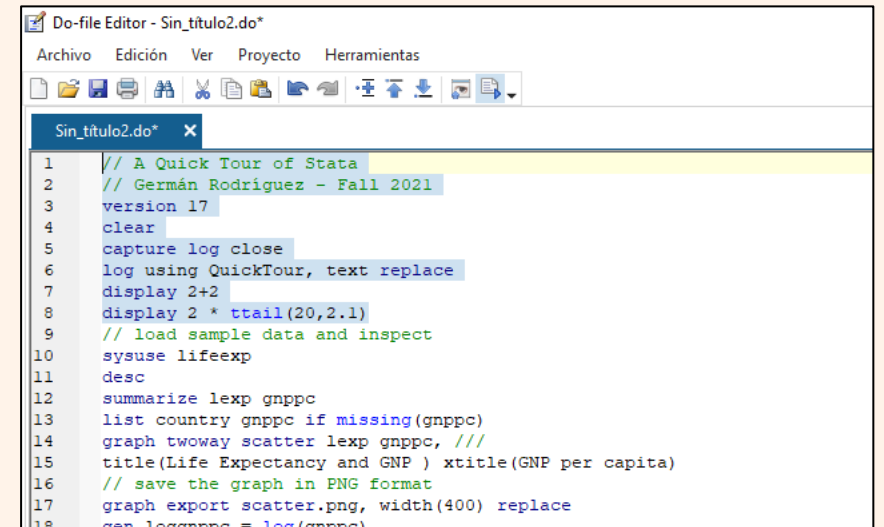
Statistics/Data Analysis 15.1 (R)

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](mailto:stata@stata.com)  
979-696-4601 (fax)

Single-user Stata perpetual license:  
Serial number: 301506325487  
Licensed to: Universidad Austral  
Asociación Civil de Estudios Superiores

Notes:  
1. Unicode is supported; see [help unicode\\_advice](#).

Abrir nuevo do-file



Do-file Editor - Sin\_titulo2.do\*

Archivo Edición Ver Proyecto Herramientas

Sin\_titulo2.do\* x

```
1 // A Quick Tour of Stata
2 // Germán Rodríguez - Fall 2021
3 version 17
4 clear
5 capture log close
6 log using QuickTour, text replace
7 display 2+2
8 display 2 * ttail(20,2.1)
9 // load sample data and inspect
10 sysuse lifeexp
11 desc
12 summarize lexp gnppc
13 list country gnppc if missing(gnppc)
14 graph twoway scatter lexp gnppc, ///
15 title(Life Expectancy and GNP ) xtitle(GNP per capita)
16 // save the graph in PNG format
17 graph export scatter.png, width(400) replace
18 gen loggnppc = log(gnppc)
```

Ejecutar parte de un do-file

# Sintaxis general de las líneas de comandos

[**by** *varlist*:] *command* [*varlist*] [=*exp*] [**if** *exp*] [**in** *range*] [*weight*] [**using** *filename*] [,*options*]

- **command**: es el único elemento de la sintaxis que está presente en todas las líneas de comandos. En general refiere a un verbo de acción que antecede el nombre de una o más variables.
- **varlist**: lista de una o más variables.
- **=exp**: expresión lógica que refiere a una fórmula matemática.
- **if exp & in range**: son elementos que permiten restringir la acción de un comando a un subconjunto de observaciones de la base de datos.
- **weight**: elemento de la sintaxis que indica que el comando debe ejecutarse usando ponderadores.
- **using filename**: esta parte de la línea de comandos introduce el nombre de un archivo, el cual puede estar en nuestra PC o en internet.
- **options**: funciones opcionales para los comandos. Las opciones suelen ser específicas para cada comando.
- **by varlist**: elemento que indica a Stata que debe repetir el comando para cada grupo de observaciones definido por los distintos valores de las variables de la lista.

El elemento *command* puede abreviarse en la escritura, siempre y cuando la expresión resultante no sea ambigua.

Puede usarse el asterisco (\*) como comodín para hacer referencia a una lista de variables que comparten un prefijo.



# Algunos comandos generales útiles

- **pwd:** muestra el directorio de trabajo (la carpeta en la cual, por defecto, Stata lee y guarda diferentes tipos de datos):

```
. pwd  
C:\Users\jcamusso\OneDrive - Universidad Austral\Documents
```

- **cd:** cambia el directorio de trabajo a la ruta que especifiquemos:

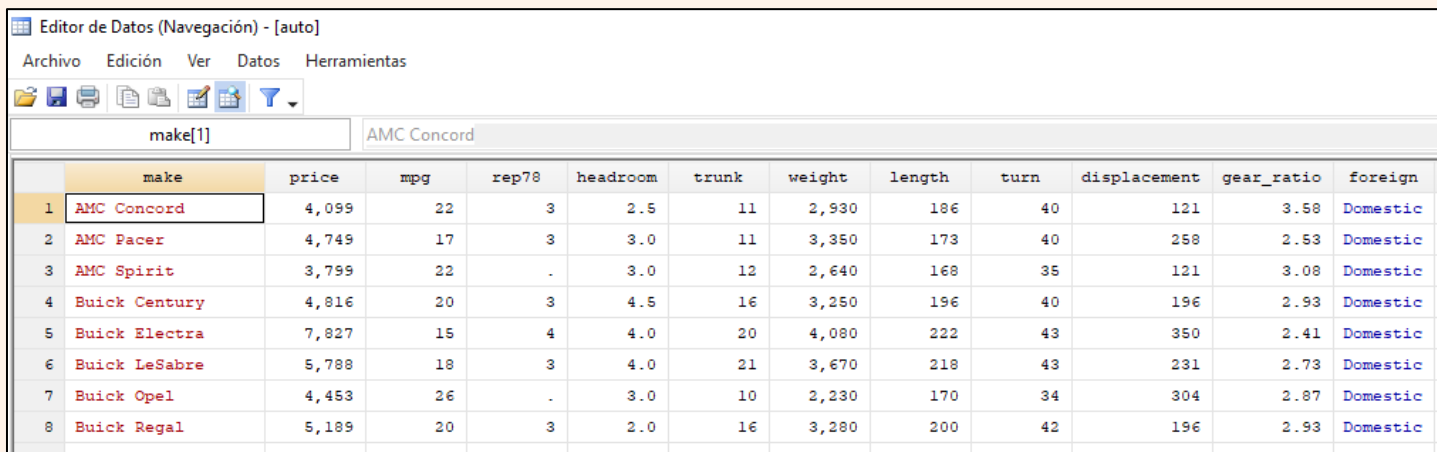
```
. cd "C:\Users\jcamusso\OneDrive - Universidad Austral\Talleres\Uso básico del software Stata"  
C:\Users\jcamusso\OneDrive - Universidad Austral\Talleres\Uso básico del software Stata  
  
. pwd  
C:\Users\jcamusso\OneDrive - Universidad Austral\Talleres\Uso básico del software Stata
```

# Algunos comandos generales útiles

- **browse:** ver la base de datos cargada, en modo lectura (no permite editar celdas).
- **edit:** ver la base de datos cargada, en modo edición (permite editar celdas a mano).

```
. sysuse auto.dta
(1978 Automobile Data)

. browse
```



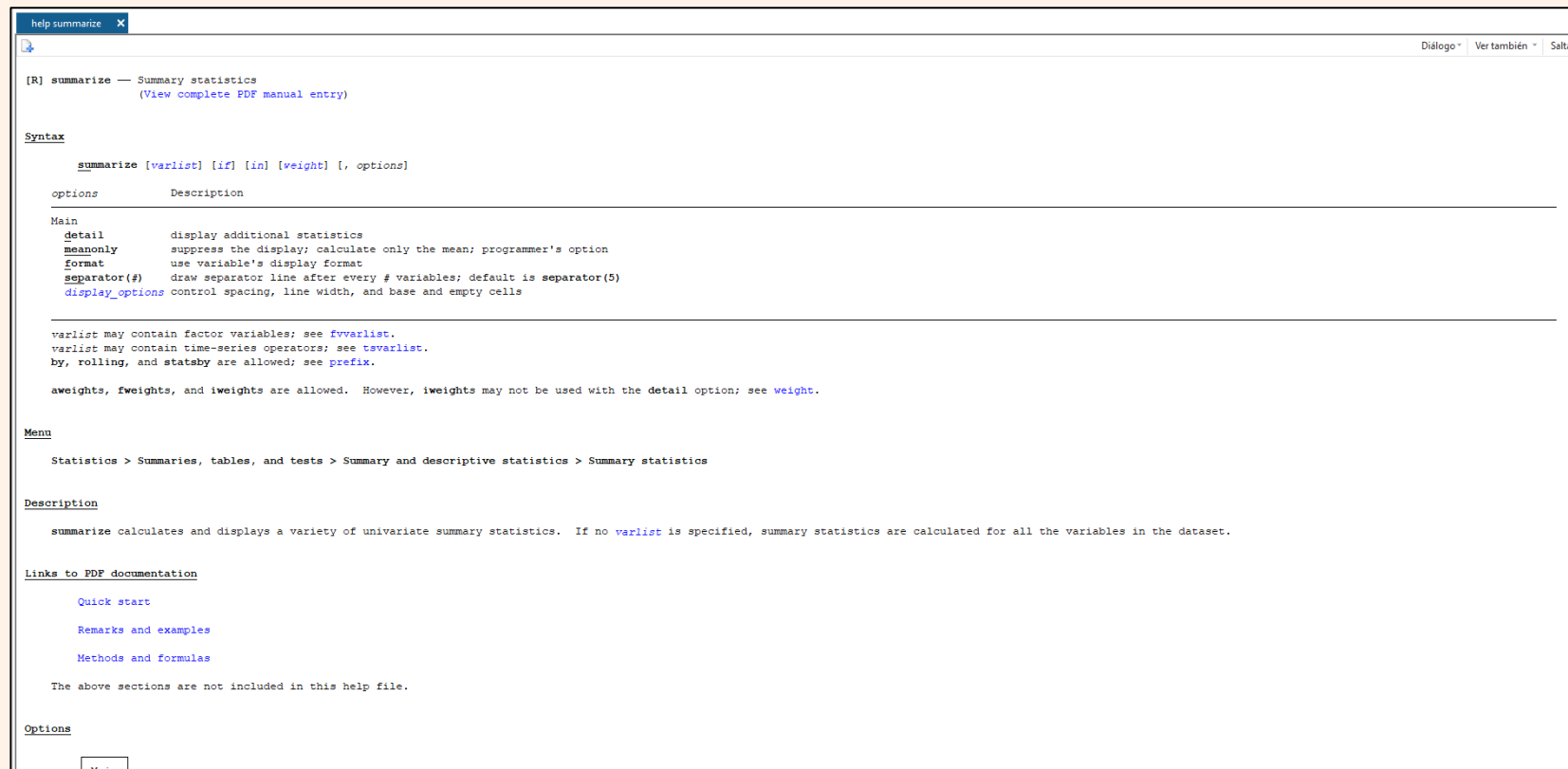
	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic

Las filas representan observaciones y las columnas representan variables

# Algunos comandos generales útiles

- **help:** obtener información detallada de un comando.

```
. help summarize
```



```
help summarize x
[
  Diálogo Ver también Saltar a
]
[R] summarize — Summary statistics
  (View complete PDF manual entry)

Syntax

  summarize [varlist] [if] [in] [weight] [, options]

options      Description
-----
Main
  detail      display additional statistics
  meonly      suppress the display; calculate only the mean; programmer's option
  format      use variable's display format
  separator(#) draw separator line after every # variables; default is separator(5)
  display_options control spacing, line width, and base and empty cells

varlist may contain factor variables; see fvarlist.
varlist may contain time-series operators; see tsvarlist.
by, rolling, and statsby are allowed; see prefix.

aweights, fweights, and iweights are allowed. However, iweights may not be used with the detail option; see weight.

Menu

  Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics

Description

  summarize calculates and displays a variety of univariate summary statistics. If no varlist is specified, summary statistics are calculated for all the variables in the dataset.

Links to PDF documentation

  Quick start
  Remarks and examples
  Methods and formulas

The above sections are not included in this help file.

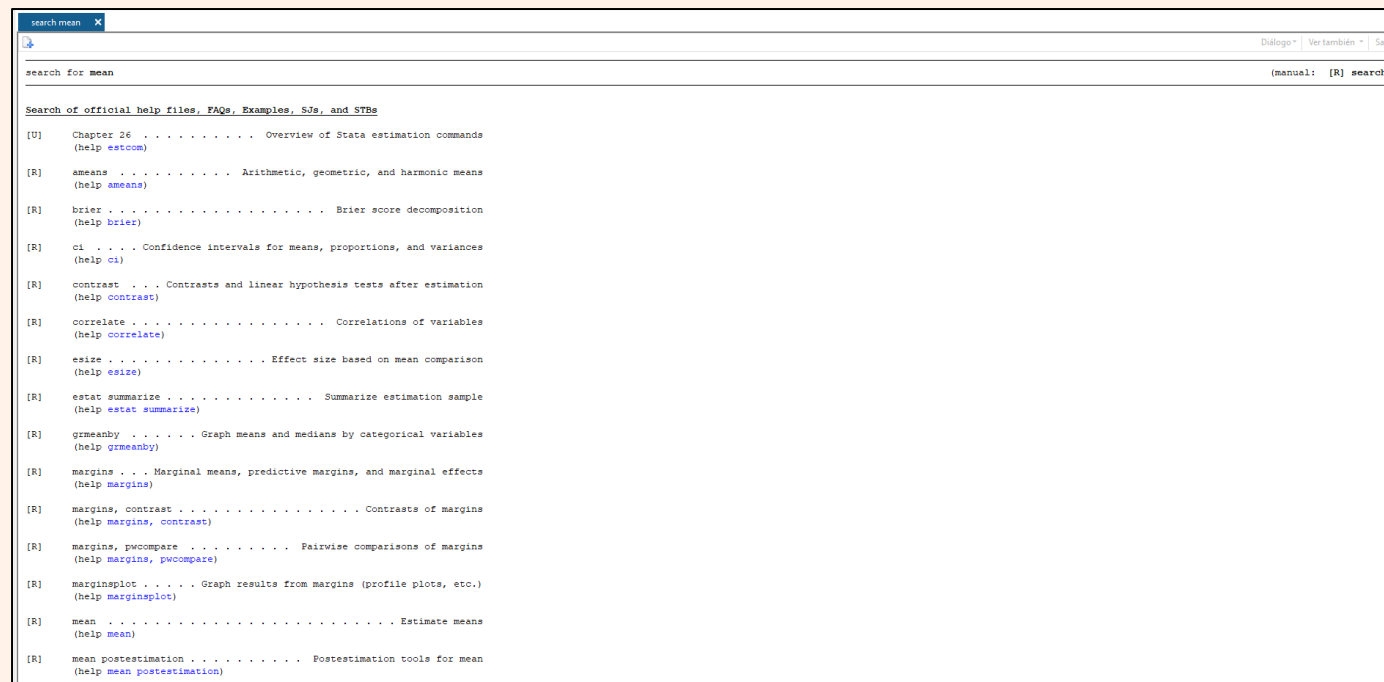
Options

  Main
```

# Algunos comandos generales útiles

- **search**: comando útil para buscar qué comando debemos ejecutar si queremos realizar una determinada acción. Por ejemplo, **search mean** nos dará información sobre qué comando ejecutar para calcular la media de una variable.

```
. search mean
```



```
search mean
search for mean
(manual: [R] search)

Search of official help files, FAQs, Examples, SJs, and STBs

[U] Chapter 26 . . . . . Overview of Stata estimation commands
(help estcom)

[R] ameans . . . . . Arithmetic, geometric, and harmonic means
(help ameans)

[R] brier . . . . . Brier score decomposition
(help brier)

[R] ci . . . . . Confidence intervals for means, proportions, and variances
(help ci)

[R] contrast . . . . . Contrasts and linear hypothesis tests after estimation
(help contrast)

[R] correlate . . . . . Correlations of variables
(help correlate)

[R] esize . . . . . Effect size based on mean comparison
(help esize)

[R] estat summarize . . . . . Summarize estimation sample
(help estat summarize)

[R] grmeanby . . . . . Graph means and medians by categorical variables
(help grmeanby)

[R] margins . . . . . Marginal means, predictive margins, and marginal effects
(help margins)

[R] margins, contrast . . . . . Contrasts of margins
(help margins, contrast)

[R] margins, pwcompare . . . . . Pairwise comparisons of margins
(help margins, pwcompare)

[R] marginsplot . . . . . Graph results from margins (profile plots, etc.)
(help marginsplot)

[R] mean . . . . . Estimate means
(help mean)

[R] mean postestimation . . . . . Postestimation tools for mean
(help mean postestimation)
```

# Algunos comandos generales útiles

Stata no permite deshacer acciones sobre la base de datos mediante CTRL+Z, pero pueden utilizarse los siguientes comandos:

- **preserve:** guarda temporalmente la base de datos.
- **restore:** reestablece el estado de la base de datos guardado con **preserve**.

```
. sysuse auto.dta
(1978 Automobile Data)

. preserve

. keep make price

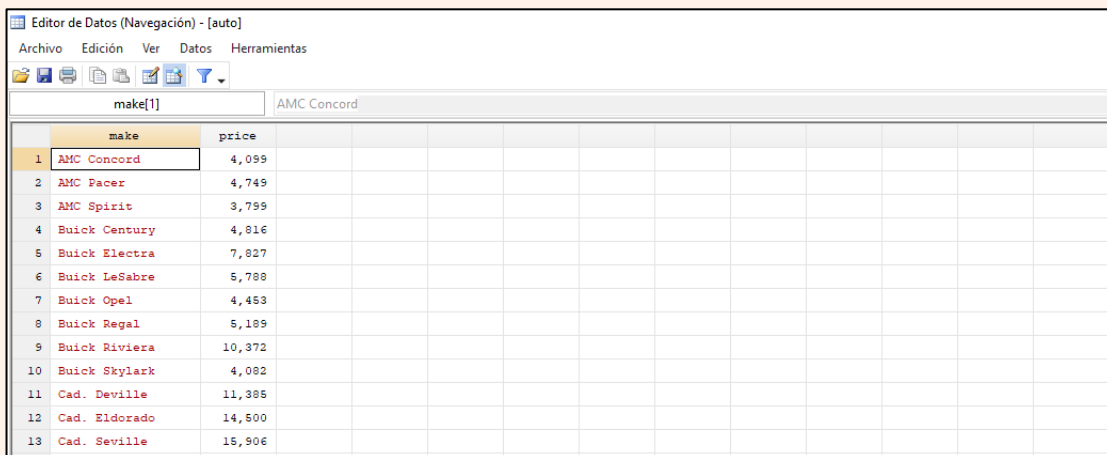
. browse
```

Borramos todas las variables, excepto *make* y *price*, pero previamente preservamos la base

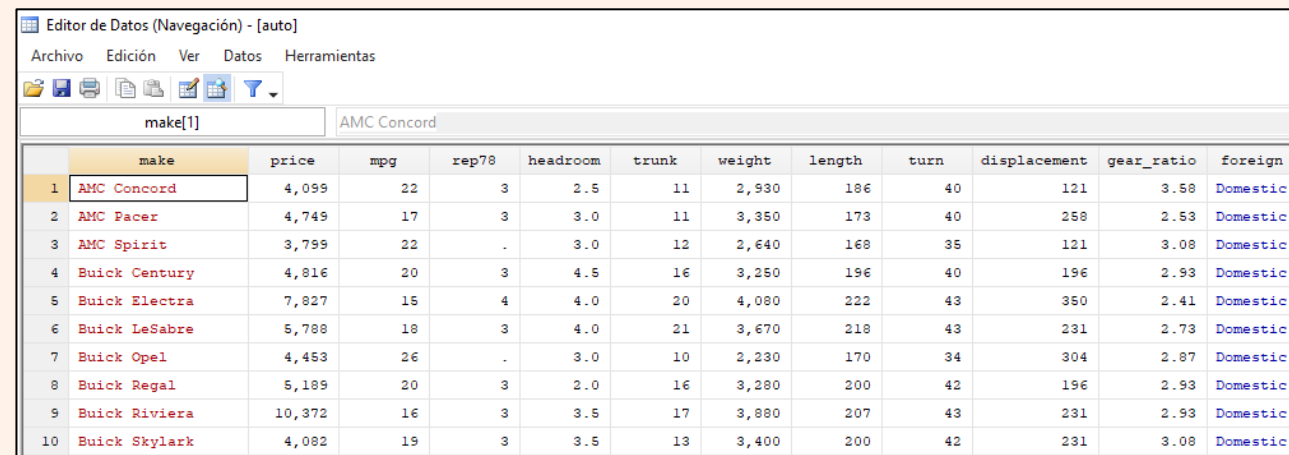
```
. restore

. browse
```

Restauramos la base



	make	price
1	AMC Concord	4,099
2	AMC Pacer	4,749
3	AMC Spirit	3,799
4	Buick Century	4,816
5	Buick Electra	7,827
6	Buick LeSabre	5,788
7	Buick Opel	4,453
8	Buick Regal	5,189
9	Buick Riviera	10,372
10	Buick Skylark	4,082
11	Cad. Deville	11,395
12	Cad. Eldorado	14,500
13	Cad. Seville	15,906



	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9	Buick Riviera	10,372	16	3	3.5	17	3,880	207	43	231	2.93	Domestic
10	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic



# Algunos comandos generales útiles

- **set more off:** indica a Stata que no debe exhibir el mensaje "more" cuando está mostrando los resultados de un comando en pantalla
- **clear all:** borra datos, resultados, etc., que estén guardados en la memoria.





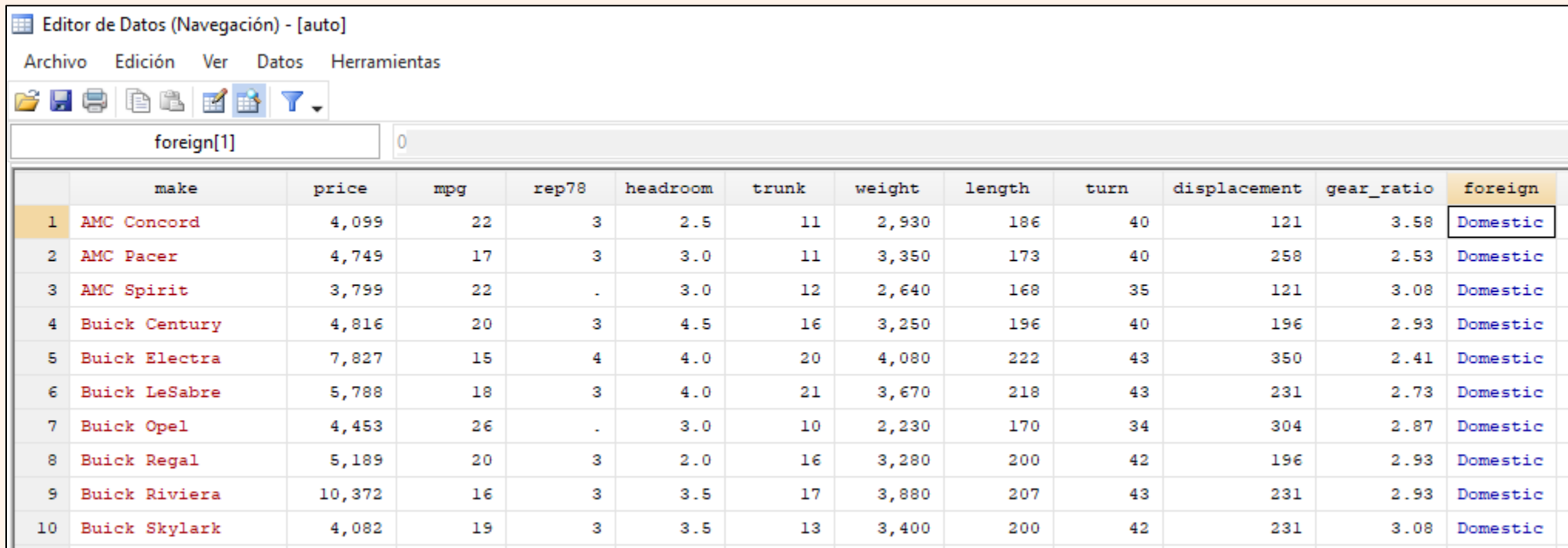
# Tipos de variables

En términos generales, Stata trabaja con dos tipos de variables:

- **numeric:** variables cuyos valores son números. Por defecto, el software considera el punto como separador decimal y la coma como separador de miles.
- **string:** son variables cuyos valores no contienen únicamente números, sino también letras y otros tipos de caracteres especiales.

# Tipos de variables

En el visor de la base de datos, Stata muestra los valores de las variables numéricas en **color negro** (si no están etiquetados). Si los valores están etiquetados, se muestran las etiquetas en **azul**. Los valores de las variables string se muestran en **rojo**:



Editor de Datos (Navegación) - [auto]

Archivo Edición Ver Datos Herramientas

foreign[1] 0

	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9	Buick Riviera	10,372	16	3	3.5	17	3,880	207	43	231	2.93	Domestic
10	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic



# Operadores aritméticos, lógicos y relacionales

Arithmetic	Logical	Relational
+ add	! not (also ~)	== equal
- subtract	or	!= not equal (also ~=)
* multiply	& and	< less than
/ divide		<= less than or equal
^ raise to power		> greater than
+ string concatenation		>= greater than or equal

Fuente: extraído de Rodríguez (2019, p. 21)

# Algunas funciones matemáticas

Function	Description
$\text{abs}(x)$	the absolute value of $x$
$\text{exp}(x)$	the exponential function of $x$
$\text{int}(x)$	the integer obtained by truncating $x$ towards zero
$\ln(x)$ or $\log(x)$	the natural logarithm of $x$ if $x > 0$
$\log_{10}(x)$	the log base 10 of $x$ (for $x > 0$ )
$\text{logit}(x)$	the log of the odds for probability $x$ : $\text{logit}(x) = \ln(x/(1-x))$
$\text{max}(x_1, x_2, \dots, x_n)$	the maximum of $x_1, x_2, \dots, x_n$ , ignoring missing values
$\text{min}(x_1, x_2, \dots, x_n)$	the minimum of $x_1, x_2, \dots, x_n$ , ignoring missing values
$\text{round}(x)$	$x$ rounded to the nearest whole number
$\text{sqrt}(x)$	the square root of $x$ if $x \geq 0$

Fuente: extraído de Rodríguez (2019, p. 22)

# Información básica de las variables

- **describe** y **codebook**: proporcionan cierta información sobre la base de datos y las variables (tipo, etiquetas, valores perdidos, etc.). Se pueden aplicar sobre un subconjunto específico de variables.

```
. sysuse auto.dta
(1978 Automobile Data)

. describe

Contains data from C:\Program Files (x86)\Stata15\ado\base/a/auto.dta
  obs:          74          1978 Automobile Data
  vars:          12          13 Apr 2016 17:45
  size:         3,182        (_dta has notes)
```

variable name	storage type	display format	value label	variable label
make	str18	%-18s		Make and Model
price	int	%8.0gc		Price
mpg	int	%8.0g		Mileage (mpg)
rep78	int	%8.0g		Repair Record 1978
headroom	float	%6.1f		Headroom (in.)
trunk	int	%8.0g		Trunk space (cu. ft.)
weight	int	%8.0gc		Weight (lbs.)
length	int	%8.0g		Length (in.)
turn	int	%8.0g		Turn Circle (ft.)
displacement	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type

```
Sorted by: foreign
```

```
. codebook make price
```

```
make
-----
      type:  string (str18), but longest is str17
unique values:  74                missing "":  0/74

examples:  "Cad. Deville"
           "Dodge Magnum"
           "Merc. XR-7"
           "Pont. Catalina"

warning:  variable has embedded blanks
```

```
price
-----
      type:  numeric (int)
      range: [3291,15906]          units:  1
unique values:  74                missing .:  0/74

      mean:  6165.26
      std. dev:  2949.5

percentiles:  10%    25%    50%    75%    90%
              3895   4195   5006.5  6342   11385
```

El valor perdido para una variable string es la ausencia de caracteres ("").  
El valor perdido para una variable numérica es el punto (.)

# Información básica de las variables

- **list:** permite listar, en formato tabla, los valores de algunas o todas las observaciones de un conjunto de variables.

```
. sysuse lifeexp.dta  
(Life expectancy, 1998)
```

```
. list in 1/5
```

	region	country	popgro~h	lexp	gnppc	safewa~r
1.	Eur & C.Asia	Albania	1.2	72	810	76
2.	Eur & C.Asia	Armenia	1.1	74	460	.
3.	Eur & C.Asia	Austria	.4	79	26830	.
4.	Eur & C.Asia	Azerbaijan	1.4	71	480	.
5.	Eur & C.Asia	Belarus	.3	68	2180	.

```
. list country popgrowth gnppc in 1/5
```

	country	popgro~h	gnppc
1.	Albania	1.2	810
2.	Armenia	1.1	460
3.	Austria	.4	26830
4.	Azerbaijan	1.4	480
5.	Belarus	.3	2180

# Ordenamiento

**sort:** ordena las observaciones de acuerdo a una o más variables (de forma numérica o alfabética, según corresponda). Algunos comandos requieren que las observaciones estén ordenadas de determinada forma para poder funcionar.

```
. sysuse census.dta
(1980 Census data by state)

. sort region

. list in 1/15
```

	state	state2	region	pop	poplt5	pop5_17	pop18p	pop65p	popurban	medage	death	marriage	divorce
1.	Pennsylvania	PA	NE	11,863,895	747,458	2,375,838	8,740,599	1,530,933	8,220,851	32.10	123,261	93,673	34,922
2.	Rhode Island	RI	NE	947,154	56,692	186,159	704,303	126,922	824,004	31.80	9,300	7,490	3,606
3.	Connecticut	CT	NE	3,107,576	185,188	637,731	2,284,657	364,864	2,449,774	32.00	26,005	26,048	13,488
4.	Massachusetts	MA	NE	5,737,037	337,215	1,153,174	4,246,648	726,531	4,808,339	31.20	54,919	46,273	17,873
5.	Maine	ME	NE	1,124,660	78,514	242,873	803,273	140,918	534,072	30.40	10,768	12,040	6,205
6.	New York	NY	NE	17,558,072	1,135,925	3,551,938	12,870,209	2,160,767	14,858,068	31.90	171,769	144,518	61,972
7.	New Hampshire	NH	NE	920,610	62,512	195,570	662,528	102,967	480,325	30.10	7,594	9,251	5,254
8.	Vermont	VT	NE	511,456	35,998	109,320	366,138	58,166	172,735	29.40	4,587	5,226	2,623
9.	New Jersey	NJ	NE	7,364,823	463,289	1,527,572	5,373,962	859,771	6,557,377	32.20	68,762	55,794	27,796
10.	Illinois	IL	N Cntrl	11,426,518	842,241	2,400,796	8,183,481	1,261,885	9,518,039	29.90	102,230	109,823	50,997
11.	Nebraska	NE	N Cntrl	1,569,825	122,946	324,224	1,122,655	205,684	987,859	29.70	14,465	14,239	6,442
12.	Kansas	KS	N Cntrl	2,363,679	180,877	468,158	1,714,644	306,263	1,575,899	30.10	21,910	24,847	13,410
13.	Missouri	MO	N Cntrl	4,916,686	354,144	1,008,339	3,554,203	648,126	3,349,588	30.90	49,329	54,625	27,595
14.	S. Dakota	SD	N Cntrl	690,768	58,446	147,160	485,162	91,019	320,777	28.90	6,523	8,800	2,811
15.	Indiana	IN	N Cntrl	5,490,224	418,764	1,199,554	3,871,906	585,384	3,525,298	29.20	47,300	57,853	40,006

Datos ordenados según *region*

```
. sort region pop

. list in 1/15
```

	state	state2	region	pop	poplt5	pop5_17	pop18p	pop65p	popurban	medage	death	marriage	divorce
1.	Vermont	VT	NE	511,456	35,998	109,320	366,138	58,166	172,735	29.40	4,587	5,226	2,623
2.	New Hampshire	NH	NE	920,610	62,512	195,570	662,528	102,967	480,325	30.10	7,594	9,251	5,254
3.	Rhode Island	RI	NE	947,154	56,692	186,159	704,303	126,922	824,004	31.80	9,300	7,490	3,606
4.	Maine	ME	NE	1,124,660	78,514	242,873	803,273	140,918	534,072	30.40	10,768	12,040	6,205
5.	Connecticut	CT	NE	3,107,576	185,188	637,731	2,284,657	364,864	2,449,774	32.00	26,005	26,048	13,488
6.	Massachusetts	MA	NE	5,737,037	337,215	1,153,174	4,246,648	726,531	4,808,339	31.20	54,919	46,273	17,873
7.	New Jersey	NJ	NE	7,364,823	463,289	1,527,572	5,373,962	859,771	6,557,377	32.20	68,762	55,794	27,796
8.	Pennsylvania	PA	NE	11,863,895	747,458	2,375,838	8,740,599	1,530,933	8,220,851	32.10	123,261	93,673	34,922
9.	New York	NY	NE	17,558,072	1,135,925	3,551,938	12,870,209	2,160,767	14,858,068	31.90	171,769	144,518	61,972
10.	N. Dakota	ND	N Cntrl	652,717	54,752	136,239	461,726	80,445	318,310	28.30	5,596	6,094	2,142
11.	S. Dakota	SD	N Cntrl	690,768	58,446	147,160	485,162	91,019	320,777	28.90	6,523	8,800	2,811
12.	Nebraska	NE	N Cntrl	1,569,825	122,946	324,224	1,122,655	205,684	987,859	29.70	14,465	14,239	6,442
13.	Kansas	KS	N Cntrl	2,363,679	180,877	468,158	1,714,644	306,263	1,575,899	30.10	21,910	24,847	13,410
14.	Iowa	IA	N Cntrl	2,913,808	221,628	604,245	2,087,935	387,584	1,708,232	30.00	26,348	27,474	11,854
15.	Minnesota	MN	N Cntrl	4,075,970	307,249	864,559	2,904,162	479,564	2,725,202	29.20	33,412	37,641	15,371

Datos ordenados según *region* y *pop*



# Borrar/mantener observaciones o variables

- **drop:** comando que permite borrar variables u observaciones.
- **keep:** comando que hace lo opuesto al anterior, es decir, mantiene dentro la base aquellas observaciones o variables que le indiquemos.





# Creación y modificación de variables

- **rename:** renombra una variable.
- **generate:** genera nuevas variables utilizando operadores, funciones matemáticas y condiciones.
- **egen:** extensión del comando anterior, que permite trabajar con otro tipo de funciones matemáticas.
- **replace:** reemplaza valores de una variable de acuerdo a determinadas condiciones.
- **label:** asigna etiquetas a los nombres de las variables para describirlas con mayor precisión.
- **label define:** define un conjunto de etiquetas para valores numéricos.
- **label values:** asigna etiquetas a los valores de una variable numérica.
- **recode:** crea categorías de valores de una variable numérica, según ciertas condiciones.
- **destring y tostring:** el primer comando permite convertir, siempre y cuando sea posible, una variable string en una variable numérica, mientras que el segundo comando realiza el proceso inverso.
- **xpose:** transpone los datos, convirtiendo las variables en observaciones, y las observaciones en variables.

# Creación y modificación de variables

Etiquetar nombres de variables:

```
. use "C:\Users\jcamusso\OneDrive - Universidad Austral\MECA\Curso de STATA\Curso de Stata - MECA 2020\Datasets\EPH_base_individual_4T_2019.dta"

. describe ipcf p21
```

variable name	storage type	display format	value label	variable label
ipcf	str9	%9s		IPCF
p21	long	%12.0g		P21

```
. label variable ipcf "Ingreso per cápita familiar"

. label variable p21 "Ingreso de la ocupación principal"

. describe ipcf p21
```

variable name	storage type	display format	value label	variable label
ipcf	str9	%9s		Ingreso per cápita familiar
p21	long	%12.0g		Ingreso de la ocupación principal



# Creación y modificación de variables

Etiquetar valores de una variable numérica:

```
. use "C:\Users\jcamusso\OneDrive - Universidad Austral\MECA\Curso de STATA\Curso de Stata - MECA 2020\Datasets\EPH_base_individual_4T_2019.dta"  
. browse region
```

Variable numérica sin etiquetar →

region[1]		1
	region	
1	1	
2	1	
3	43	
4	43	
5	43	
6	43	
7	43	
8	43	
9	42	
10	43	

```
. label define regiones 1 "GBA" 40 "NOA" 41 "NEA" 42 "Cuyo" 43 "Pampeana" 44 "Patagonia"  
. label values region regiones  
. browse region
```

Variable numérica etiquetada →

region[1]		1
	region	
1	GBA	
2	GBA	
3	Pampeana	
4	Pampeana	
5	Pampeana	
6	Pampeana	
7	Pampeana	
8	Pampeana	
9	Cuyo	
10	Pampeana	

# Creación y modificación de variables

## Recode de variable numérica:

```
. use "C:\Users\jcamusso\OneDrive - Universidad Austral\MECA\Curso de STATA\Curso de Stata - MECA 2020\Datasets\EPH_base_individual_4T_2019.dta"  
  
. recode ch06 (min/25=1) (26/30=2) (31/35=3) (36/40=4) (41/44=5) (45/50=6) (51/max=7) if ch06>0, generate(edad_grupos)  
(57166 differences between ch06 and edad_grupos)  
  
. list ch06 edad_grupos in 1/15
```

	ch06	edad_g~s
1.	74	7
2.	78	7
3.	16	1
4.	16	1
5.	15	1
6.	41	5
7.	2	1
8.	44	5
9.	56	7
10.	15	1
11.	44	5
12.	12	1
13.	25	1
14.	71	7
15.	74	7

Aclaración: la variable ch06 es la edad de la persona en la base de datos utilizada.



# Creación y modificación de variables

Transponer una base de datos:

```
. webuse xposexmpl  
  
. list  


|    | county | year1 | year2 | year3 |
|----|--------|-------|-------|-------|
| 1. | 1      | 57.2  | 11.3  | 19.5  |
| 2. | 2      | 12.5  | 8.2   | 28.9  |
| 3. | 3      | 18    | 14.2  | 33.2  |


```
. xpose, clear  
  
. list  


	v1	v2	v3
1.	1	2	3
2.	57.2	12.5	18
3.	11.3	8.2	14.2
4.	19.5	28.9	33.2


```


```



# Estadística descriptiva

Algunos ejemplos:

- Conteo de observaciones (comando **count**)
- Cálculo de media, mediana, desvío estándar, etc. (comando **summarize**)
- Cálculo de cuantiles y creación de grupos cuantílicos (comandos **xtile** y **pctile**)
- Construcción de tablas de una y doble entrada (comando **tabulate**)
- Uso de ponderadores (elemento **weight** de la sintaxis de un comando)

# Estadística descriptiva

Tablas con media, mediana, desvío estándar, etc., de una variable o conjunto de variables:

```
. sysuse auto.dta
(1978 Automobile Data)

. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

```
. summarize price, detail
```

		Price	
		Percentiles	Smallest
1%	3291	3291	
5%	3748	3299	
10%	3895	3667	Obs 74
25%	4195	3748	Sum of Wgt. 74
50%	5006.5		Mean 6165.257
		Largest	Std. Dev. 2949.496
75%	6342	13466	
90%	11385	13594	Variance 8699526
95%	13466	14500	Skewness 1.653434
99%	15906	15906	Kurtosis 4.819188

# Estadística descriptiva

Tablas de frecuencias de una y doble entrada:

```
. sysuse auto.dta
(1978 Automobile Data)

. xtile grupo_quintilico_price = price, nquantiles(5)

. tab grupo_quintilico_price
```

5 quantiles of price	Freq.	Percent	Cum.
1	15	20.27	20.27
2	15	20.27	40.54
3	15	20.27	60.81
4	15	20.27	81.08
5	14	18.92	100.00
Total	74	100.00	

```
. tab foreign
```

Car type	Freq.	Percent	Cum.
Domestic	52	70.27	70.27
Foreign	22	29.73	100.00
Total	74	100.00	

```
. tab grupo_quintilico_price foreign
```

5 quantiles of price	Car type		Total
	Domestic	Foreign	
1	11	4	15
2	12	3	15
3	12	3	15
4	8	7	15
5	9	5	14
Total	52	22	74

Aclaración: la variable grupo\_quintilico\_price fue creada con el comando **xtile**.



# Estadística descriptiva

Otro tipo de tablas de estadística descriptiva, usando comando **tabulate**:

```
. tab grupo_quintilico_price, summarize (price)
```

5 quantiles of price	Summary of Price		
	Mean	Std. Dev.	Freq.
1	3,834.067	253.47146	15
2	4,397.267	155.77019	15
3	5,052.6	297.79662	15
4	6,301.733	587.6315	15
5	11,603.14	2,261.872	14
Total	6,165.257	2,949.496	74

```
. tab grupo_quintilico_price foreign, summarize (price)
```

Means, Standard Deviations and Frequencies of Price

5 quantiles of price	Car type		Total
	Domestic	Foreign	
1	3,825	3,859	3,834.067
	293.29916	109.29166	253.47146
	11	4	15
2	4,381.25	4,461.333	4,397.267
	159.33277	150.08775	155.77019
	12	3	15
3	5,051.333	5,057.667	5,052.6
	300.87458	350.48728	297.79662
	12	3	15
4	6,324.375	6,275.857	6,301.733
	661.93589	541.36384	587.6315
	8	7	15
5	12,211.67	10,507.8	11,603.14
	2,287.997	1,955.619	2,261.872
	9	5	14
Total	6,072.423	6,384.682	6,165.257
	3,097.104	2,621.915	2,949.496
	52	22	74

Aclaración: la variable grupo\_quintilico\_price fue creada con el comando **xtile**.



# Gráficos

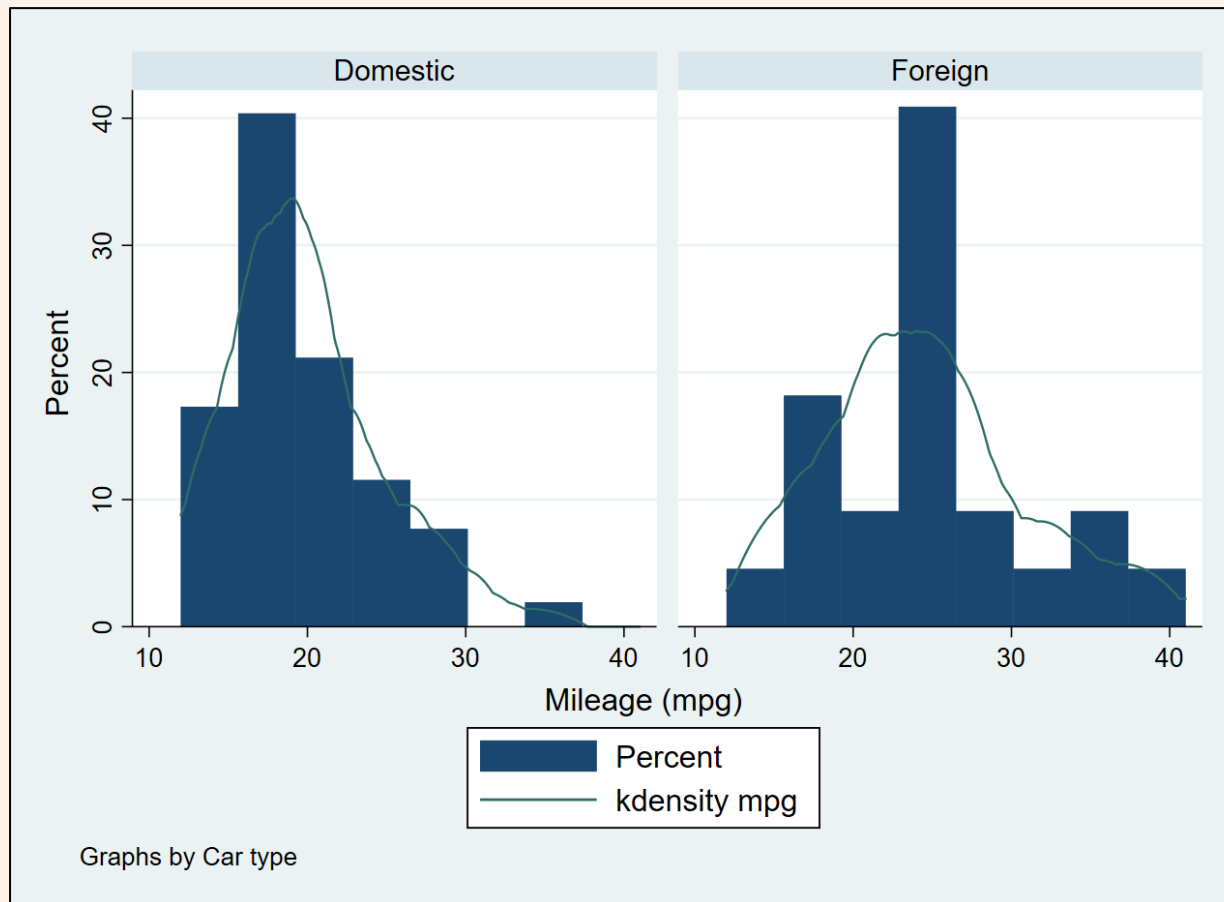
Stata permite construir diversos tipos de gráficos para nuestros datos, los cuales pueden exportarse en diferentes formatos (.png, .pdf, .emf, etc.). Además, el software permite combinar dos o más gráficos en uno solo. Algunos ejemplos:

- Histograma
- Box-Plot
- Gráfico de barras
- Gráfico de líneas
- Diagrama de dispersión

# Gráficos

Ejemplo de histograma:

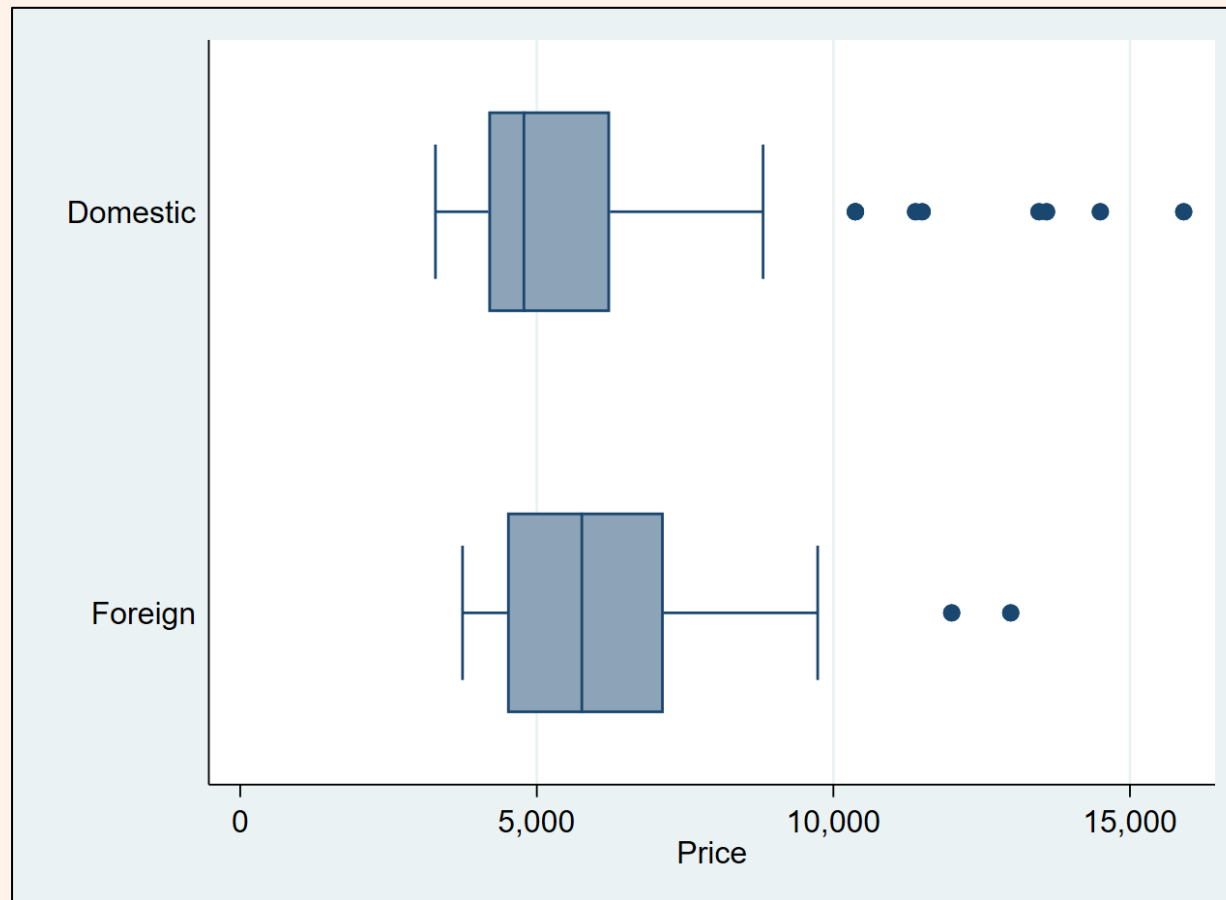
```
. sysuse auto.dta  
(1978 Automobile Data)  
  
. histogram mpg, percent color(navy) by(foreign) kden
```



# Gráficos

Ejemplo de box-plot:

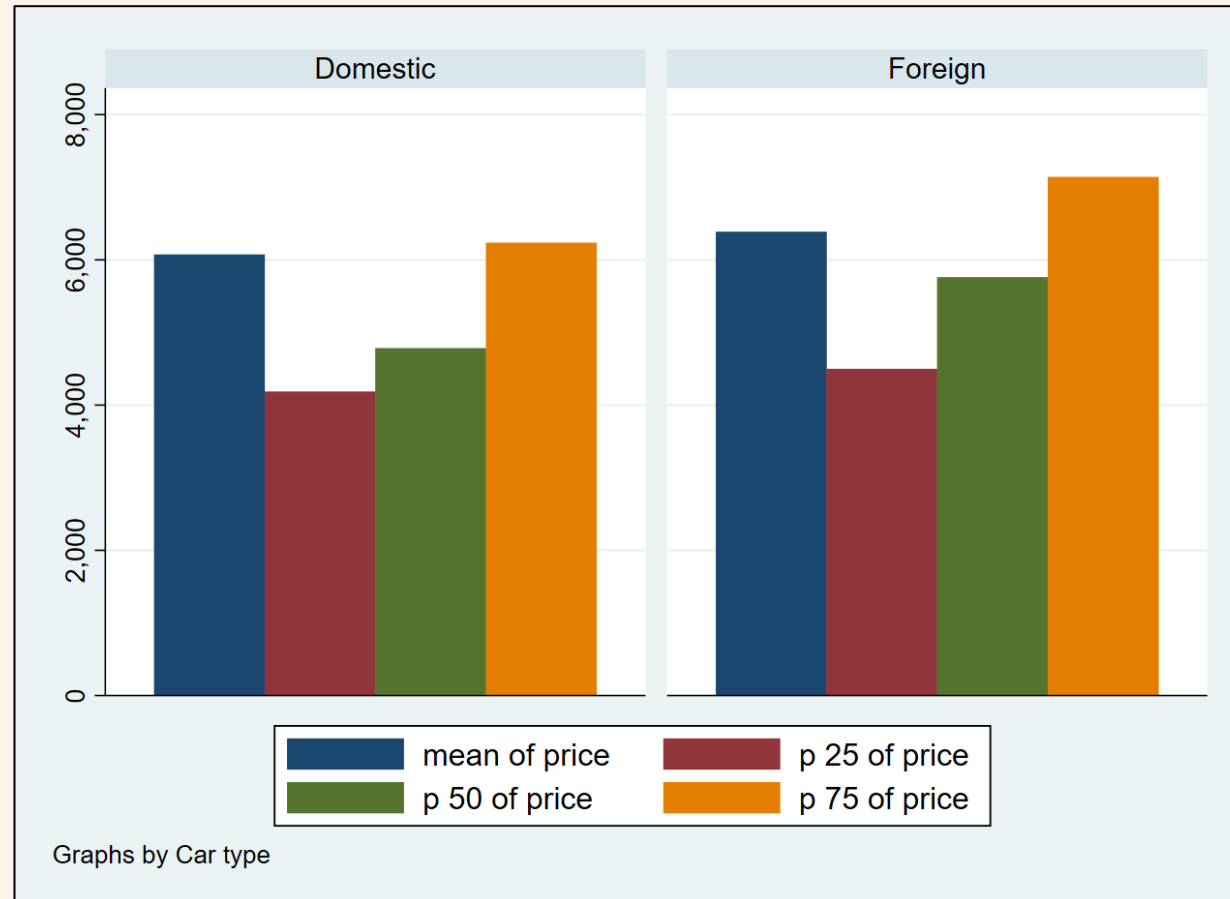
```
. sysuse auto.dta  
(1978 Automobile Data)  
  
. graph hbox price, over(foreign)
```



# Gráficos

Ejemplo de gráfico de barras:

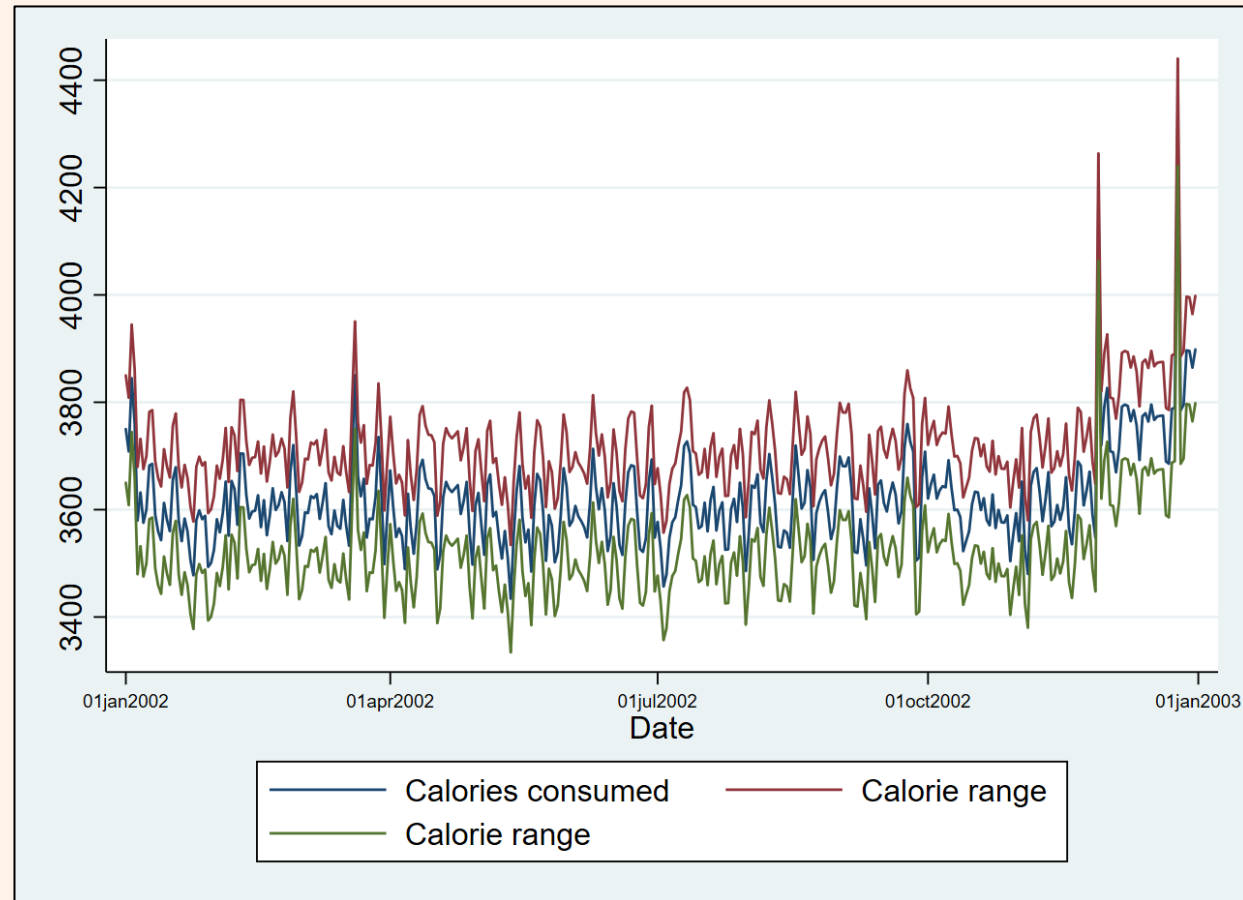
```
. sysuse auto.dta  
(1978 Automobile Data)  
  
. graph bar (mean) price (p25) price (median) price (p75) price, by(foreign)
```



# Gráficos

Ejemplo de gráfico de líneas:

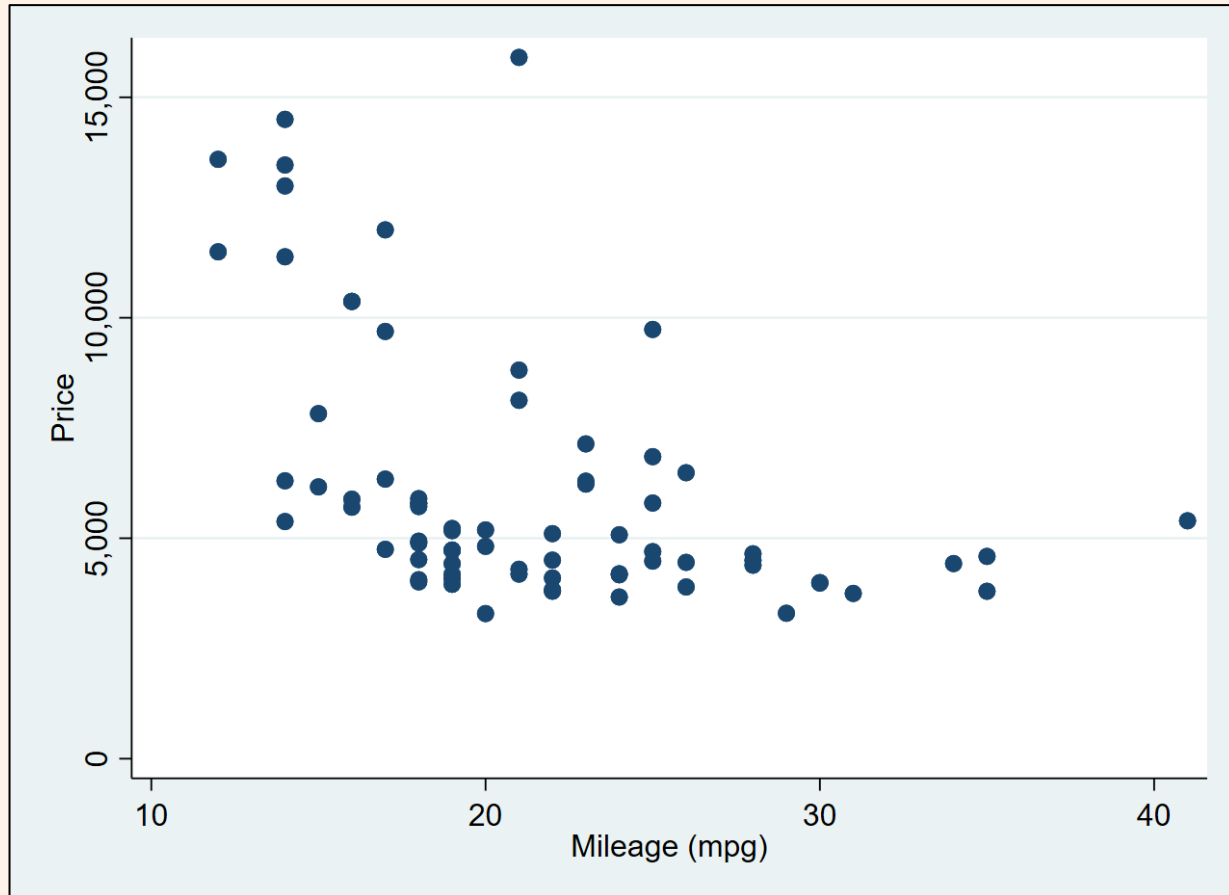
```
. sysuse tsline2.dta  
  
. tsset day  
   time variable: day, 01jan2002 to 31dec2002  
       delta: 1 day  
  
. tsline calories ucalories lcalories
```



# Gráficos

Ejemplo de diagrama de dispersión:

```
. sysuse auto.dta  
(1978 Automobile Data)  
  
. scatter price mpg
```

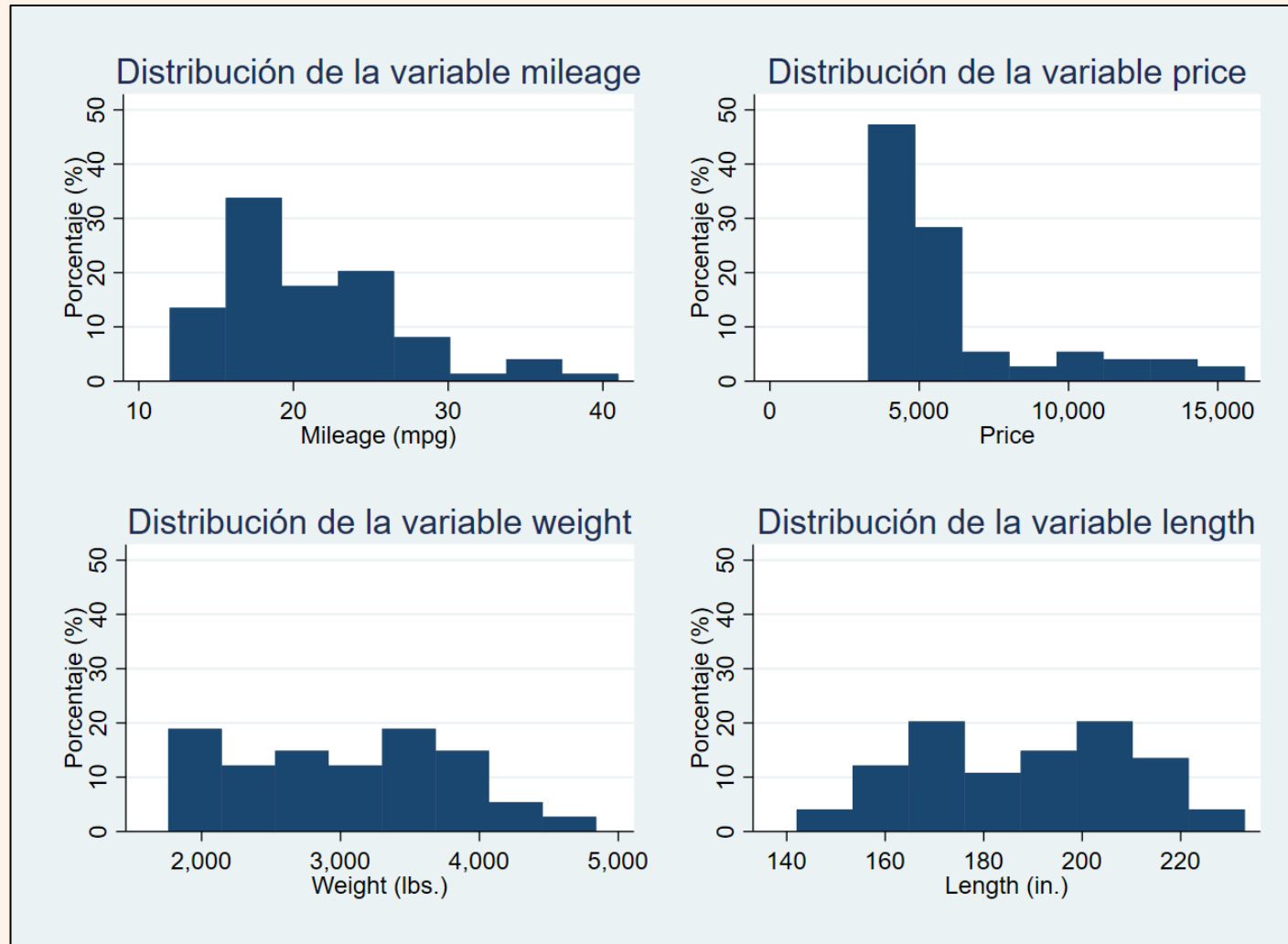


```
. graph twoway (scatter price mpg) (lfit price mpg)
```



# Gráficos

Ejemplo de gráfico combinado:







# Instalación de paquetes creados por usuarios

Stata permite la instalación de paquetes creados por usuarios, los cuales usualmente incorporan funciones que el software no trae incorporadas oficialmente.

Cuando se conoce el nombre del paquete (previa búsqueda en Google, por ejemplo), los pasos para instalarlo son:

```
. search tabout
```

*tabout* es el nombre del paquete que quiero instalar

Clickeamos →

```
search for tabout

Search of official help files, FAQs, Examples, SJs, and STBs

Web resources from Stata and other users

(contacting http://www.stata.com)

1 package found (Stata Journal listed first)
-----
tabout from http://fmwww.bc.edu/RePEc/bocode/t
  'TABOUT': module to export publication quality cross-tabulations / tabout
  is a table building program for oneway and twoway / tables of frequencies
  and percentages, and for summary tables. It / produces publication quality
  tables for export to a text file. / These tables can then be used with

(click here to return to the previous screen)

(end of search)
```

# Instalación de paquetes creados por usuarios

```
package about from http://fmwww.bc.edu/RePEc/bocode/t

TITLE
    'ABOUT': module to export publication quality cross-tabulations

DESCRIPTION/AUTHOR(S)

    about is a table building program for oneway and twoway
    tables of frequencies and percentages, and for summary tables. It
    produces publication quality tables for export to a text file.
    These tables can then be used with spreadsheets, word processors,
    web browsers or compilers like LaTeX. The tables produced by
    about allow multiple panels so that a number of variables can be
    included in the one table. about also provides standard errors
    and confidence intervals, as well as a range of table statistics
    (chi2 etc). The output from about matches Stata's tabulate, most
    of tabstat and some of table.

KW: tables
KW: latex
KW: html
KW: delimited text

Requires: Stata version 9.2

Distribution-Date: 20190316

Author: Ian Watson , Macquarie University, Australia
Support: email mail@ianwatson.com.au

INSTALLATION FILES                                     (click here to install)
    about.ado
    tabstatout.ado
    about.hlp
    ../f/figout.ado
    ../f/figout.hlp

ANCILLARY FILES                                       (click here to get)
    about_tutorial.pdf
    ../e/examples_tab.do
    ../e/examples_tex.do
    ../b/bot.tex
    top.tex
```

← *cliqueamos*

# Exportar resultados a Excel

Algunos comandos útiles:

- **putexcel:** exporta expresiones, matrices, imágenes y resultados guardados por Stata a una planilla Excel.
- **tabout:** exporta tablas de una y doble entrada a una planilla Excel (paquete creado por usuarios, por lo que requiere instalación previa para su uso).
- **outreg2:** exporta resultados de un modelo de regresión a una planilla Excel (paquete creado por usuarios, por lo que requiere instalación previa para su uso).



# Resumen y fusión de bases de datos

- **collapse:** “resumir” (colapsar) una base de datos en un *dataset* más pequeño que contenga diferentes estadísticas (media, mediana, percentiles, etc.) de las variables.
- **append:** permite fusionar verticalmente (es decir, apilar) dos o más bases de datos, agregando nuevas observaciones a las ya existentes en el *master dataset* (la base de datos que está cargada en la memoria antes de realizar el apilado).
- **merge:** permite fusionar horizontalmente dos o más bases de datos, agregando nuevas variables a las ya existentes en el *master dataset*. Para ello se necesita una variable identificatoria que “linkee” las observaciones de las diferentes bases de datos. Además, existen diferentes tipos de fusiones horizontales según la forma en que esta variable identifique las observaciones.



# by | bysort

La opción **by** indica que un determinado comando debe ejecutarse para varios grupos de observaciones, delimitados de acuerdo al valor que tienen en una o más variables (por ejemplo, si queremos calcular la media del salario por región geográfica, la opción **by** permite que el comando **summarize** pueda ejecutarse para cada valor posible de la variable “región”). Dependiendo del comando, la opción **by** debe especificarse al comienzo de la línea o al final de la misma

En ocasiones, algunos comandos requieren que las observaciones estén previamente ordenadas (de acuerdo a los valores de una o más variables) para poder ejecutarse en conjunto con la opción **by**. Si bien, como vimos previamente, esto puede realizarse mediante el comando **sort**, existe una alternativa más directa dada por el comando **bysort**.

# Bucles

En ocasiones, uno o más pasos de nuestro análisis de datos con Stata tienen que hacerse repetidamente, con ligeras modificaciones. Un bucle (*loop*) permite que un mismo comando pueda ejecutarse para varios elementos o *ítems* de un conjunto (variables, valores numéricos, categorías, etc.), sin necesidad de escribir un comando separado para cada elemento. Existen dos comandos que nos serán muy útiles para generar bucles simples y anidados:

- **foreach:** asigna la macro local -abreviación para una cadena de caracteres o números- *lname* (nombre de la macro definida por el usuario) a cada elemento de una lista de variables, categorías, etc., y ejecuta el comando indicado entre llaves.
- **forvalues:** asigna la macro local -abreviación para una cadena de caracteres o números- *lname* (nombre de la macro definida por el usuario) a cada valor de un rango de números y ejecuta el comando indicado entre llaves.



# Otros tópicos

- Programación en Stata
- Macros locales y globales
- Escalares
- Matrices y vectores
- Integración oficial con Python (nueva característica de Stata 17)
- Escribir documentos LaTeX: paquete *texdoc* (creado por usuarios)
- Tablas personalizadas (nueva característica de Stata 17)



# Referencias

- Rodríguez, G. (2022). *Stata Tutorial: Updated for Version 17*. Princeton University.
- Rodríguez, G. (2019). *Stata Tutorial: Updated for Version 16*. Princeton University.
- StataCorp. (2021). *Getting Started with Stata for Windows: Release 17*. Texas: Stata Press.
- StataCorp. (2019). *Getting Started with Stata for Windows: Release 16*. Texas: Stata Press.



# SOMOSAUSTRAL

# Muchas gracias.

[www.austral.edu.ar](http://www.austral.edu.ar)

