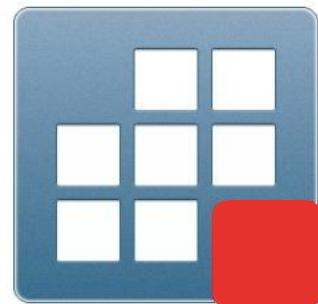


Programación en STATA®

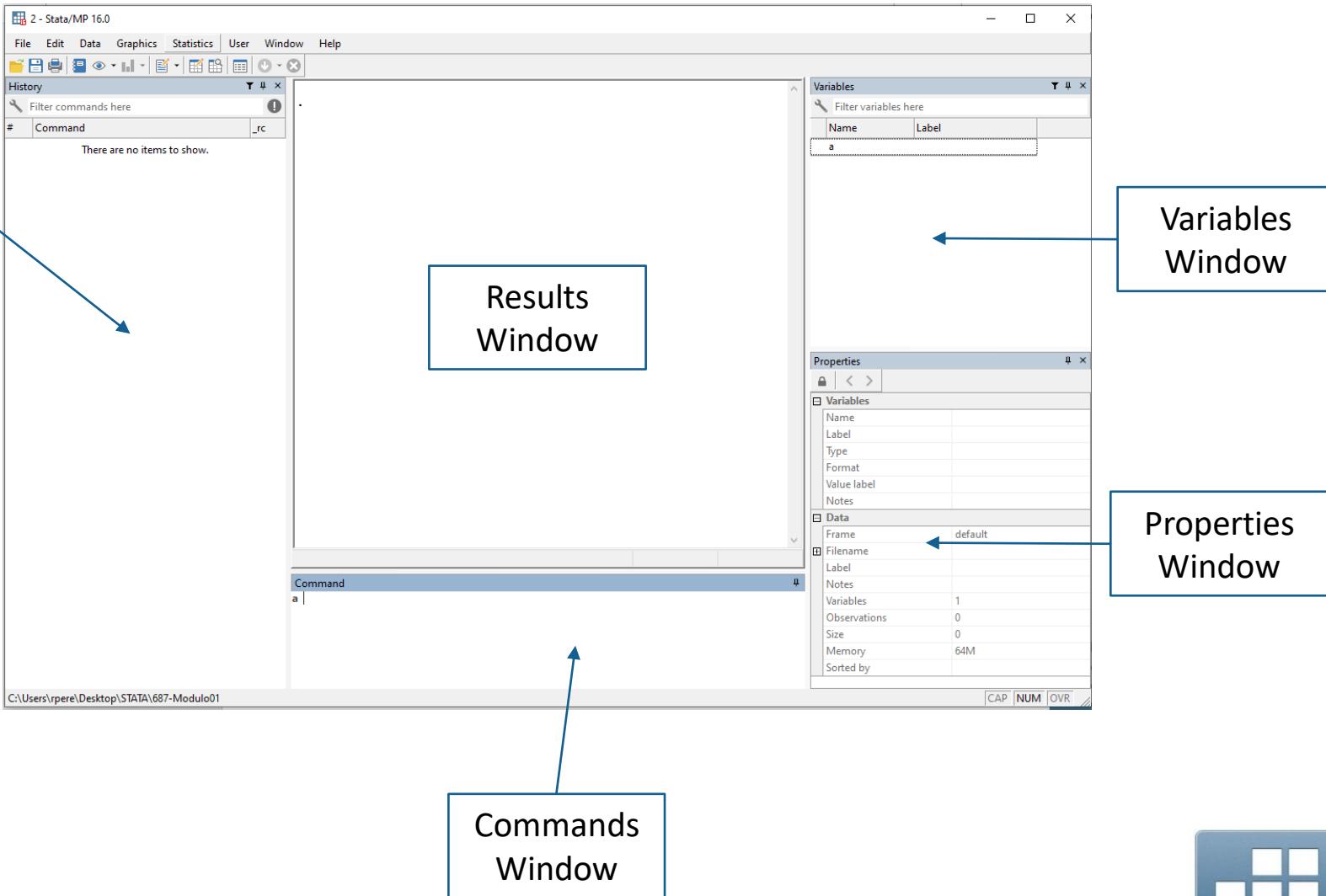
Richard P. Pérez Palma Ponce



Parte 1: Introducción



Interfaz Gráfica de Usuarios (GUI) de Stata



Métodos Abreviados (Shortcuts)

F2

Describe data

Ctrl + 8

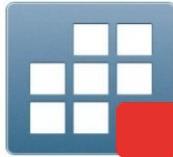
Abrir el data editor

Ctr + 9

Abrir nuevo do-file

Ctr + D

Ejecuta texto seleccionado



Establecer los parámetros iniciales

cd

Cambiar el directorio actual

dir (*.dta)

Mostrar nombre de archivos en la ruta de trabajo

capture log close

Cierra el log existente de cualquier do-file

log using “.txt”, replace

Crea un nuevo log file para guardar resultados y códigos

search (findit)

Encontrar el paquete a instalar

ssc install

Instalar paquete



Parte 2: Elementos de Programación



Creación de objetos: Scalar y Matrix

Scalar

Scalar A = 3

Matrix

Llenar Matrices

(nx1) : matrix A = ($a_1 \backslash a_2, \backslash \cdots \backslash a_n$)
(1xn) : matrix A = (a_1, a_2, \dots, a_n)

Operaciones con matrices:

Matriz Inversa: invsym

Determinante: det

Operator	Symbol
Unary operators	
negation	-
transposition	,
Binary operators	
(lowest precedence)	
row join	\
column join	,
addition	+
subtraction	-
multiplication	*
division by scalar	/
Kronecker product	#
(highest precedence)	



Bucles y Condicionales

Loops

- Textos y listas

```
foreach x of varlist mpg weight foreign {  
    summarize `x'  
}  
  
foreach x in "Richard" "Manuel" "Gabriela" {  
    display length("`x'")  
}
```

- Numérico

```
forvalues i = 10(10)50 {  
    generate d`i' = 0  
}
```

Condicionales

```
local m = 1  
forvalues i = 10(10)40 {  
    if `i' == 30 {  
        display "Se ha elegido el número `m'"  
    }  
    local m = `++m'  
}
```



Parte 3: Importación de datos



Importación de datos: .dta y bases externas.

use

```
. use sumaria-2019.dta, replace  
  
. describe  
  
Contains data from sumaria-2019.dta  
obs:      34,565  
vars:      158          10 Sep 2020 02:23  
  
           storage  display   value  
variable name    type    format   label     variable label  
  
aÑo          str4    %4s      año de la encuesta  
mes          str2    %2s      mes de ejecución de la encuesta  
nconglome   str6    %6s      número de conglomerado (proveniente del marco)  
conglome    str6    %6s      número de conglomerado  
vivienda    str3    %3s      número de selección de vivienda  
hogar        str2    %2s      número secuencial del hogar  
ubigeo       str6    %6s      ubicación geográfica  
dominio     byte    %8.0g    dominio  
estrato     byte    %46.0g   estrato  
percepho    byte    %8.0g    total de perceptores de ingresos
```

import

```
. import delimited "Ubigeo.csv", stringcols(1) clear  
(7 vars, 1,874 obs)  
  
. save Ubigeo.dta, replace  
file Ubigeo.dta saved
```



Parte 4: Manejo de Base de Datos



Análisis de Base de Datos

describe

describe

Contains data from https://stats.idre.ucla.edu/stat/data/patient_pt1_stata_dm.dta

obs: 120

vars: 25

6 May 2017 19:05

size: 11,400

variable name	storage type	display format	value label	variable label
hospital	str14	%14s		
hospid	byte	%8.0g		hospid
docid	str5	%9s		

list

. list dominio estrato mieperho percepho in 100/120

	dominio	estrato	mieperho	percepho
100.	selva Área de empadronamiento rural (aer) simple	2	1	
101.	selva Área de empadronamiento rural (aer) simple	5	2	
102.	selva Área de empadronamiento rural (aer) simple	3	2	
103.	selva Área de empadronamiento rural (aer) simple	5	3	
104.	selva Área de empadronamiento rural (aer) simple	3	2	



Análisis de Base de Datos

codebook

dominio	dominio geográfico
<pre>type: numeric (byte) label: dominio range: [1,8] units: 1 unique values: 8 missing .: 0/34,565 tabulation: Freq. Numeric Label 4,782 1 costa norte 3,258 2 costa centro 2,196 3 costa sur 2,362 4 sierra norte 6,049 5 sierra centro 4,898 6 sierra sur 6,990 7 selva 4,030 8 lima metropolitana</pre>	

tabulate

. tabulate dominio	Freq.	Percent	Cum.
costa norte	4,782	13.83	13.83
costa centro	3,258	9.43	23.26
costa sur	2,196	6.35	29.61
sierra norte	2,362	6.83	36.45
sierra centro	6,049	17.50	53.95
sierra sur	4,898	14.17	68.12
selva	6,990	20.22	88.34
lima metropolitana	4,030	11.66	100.00
Total	34,565	100.00	



Creación de Variables

generate

```
. gen d_gf = 1 if mieperho >= 5  
(25,464 missing values generated)
```

replace

```
. gen d_gf = 1 if mieperho >= 5  
(25,464 missing values generated)
```

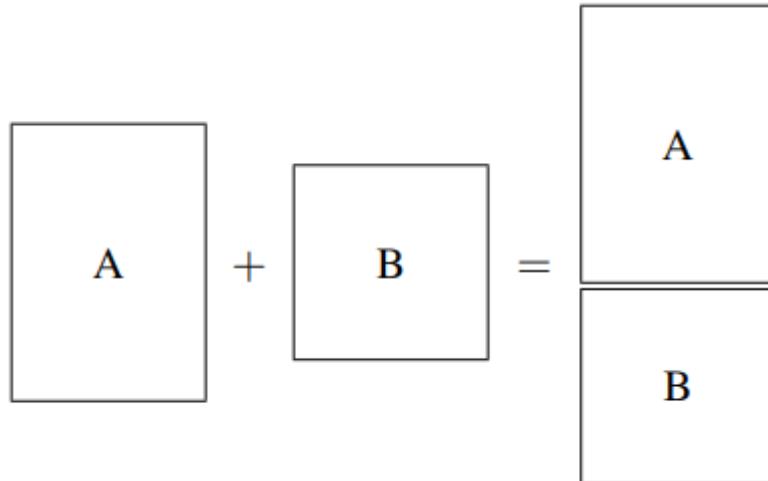
values

```
label values d_gf etiquetas  
  
label variable d_gf "Hogares con 5 o más miembros"  
  
label define etiquetas 0 "Hogares con menos de 5 miembros" 1 "Hogares con 5 o más miembros"
```

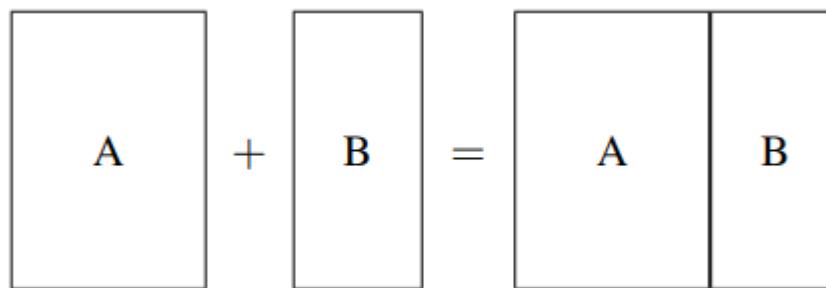


Combinar Base de Datos

Append



Merge



Descargar datos de ENAHO: Encuesta Nacional de Hogares

<https://www.inei.gob.pe/bases-de-datos/>

The screenshot shows the INEI Microdatos Base de Datos homepage. At the top left is the INEI logo. The header features "MICRODATOS" and "BASE DE DATOS". Below the header is a world map. A red box highlights the "Consulta por Encuestas" button in the top right menu bar. Below the menu are links for "PRESENTACIÓN" and "GUÍA DE USUARIO". The main content area is titled "PRESENTACIÓN" and contains two paragraphs about the ENAHO survey, followed by a large red bracket on the right side.

MICRODATOS
BASE DE DATOS

Consulta por Encuestas Documentación

PRESENTACIÓN GUÍA DE USUARIO

PRESENTACIÓN

El Instituto nacional de Estadística e Informática (INEI), en el marco de sus actividades para la promoción y difusión de las investigaciones estadísticas que realiza, pone a disposición del público en general el sistema de Microdatos.

Este sistema, proporciona las bases de datos y la documentación derivada de las investigaciones y encuestas ejecutadas por el INEI durante los últimos años, resguardando el secreto estadístico de la información.

The screenshot shows the "CONSULTA POR ENCUESTA" section. It includes a message to select the survey, year, and period, and a note about viewing modules. Below is a search interface with dropdowns for "ENCUESTA" (set to "ENAHO Metodología ACTUALIZADA"), "AÑO" (set to "2019"), and "Período" (set to "Anual - (Ene-Dic)"). A large red bracket on the right side covers the search area and the results table.

CONSULTA POR ENCUESTA

Sírvase seleccionar Encuesta, Año y Período y a continuación se mostrarán todas los Módulos de la Encuesta Seleccionada. Luego proceda a descargar el módulo de su interés.

ENCUESTA: ENAHO Metodología ACTUALIZADA

AÑO: 2019 Período: Anual - (Ene-Dic)

Nro	Año	Periodo	Código Encuesta	Encuesta	Código Módulo	Módulo	Ficha	Descarga
1	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	1	Características de la Vivienda y del Hogar		
2	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	2	Características de los Miembros del Hogar		
3	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	3	Educación		
4	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	4	Salud		
5	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	5	Empleo e Ingresos		
6	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	7	Gastos en Alimentos y Bebidas (Módulo 601)		
7	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	8	Instituciones Beneficas		
8	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	9	Mantenimiento de la Vivienda		
9	2019	55	687	Condiciones de Vida y Pobreza - ENAHO	10	Transportes y Comunicaciones		



Parte 5: Personalización de Gráficos



Cheat Sheet: Visualización de información

Data Visualization with Stata

For more info, see Stata's reference manual (stata.com)

ONE VARIABLE

CONTINUOUS

```
histogram mpg, width(5) freq kdensity kdensity(bwidth(5))
histogram
    bin(#) • width(#) • density • fraction • frequency • percent • addlabels
    addplot(plot) • options • normals(options) • kdensity
    kdensity(bwidth(n)) • smooth(n)
```

DISCRETE

```
graph bar (count), over(foreign, gap(.5)) intensity(.5)
bar plot
    count • (percent) • (count) • over(variables) • options: gap(#)
    label • descending • reverse • cw • missing • nofill • all/categories •
    percentages • stack • bargap(#) • intensity(#) • vallernate • alternate
graph bar (percent), over(rep78) over(foreign)
grouped bar plot
    (axis) • (percent) • (count) • over(variables) • options: gap(#)
    label • descending • reverse • cw • missing • nofill • all/categories •
    percentages • stack • bargap(#) • intensity(#) • vallernate • alternate
```

DISCRETE X, CONTINUOUS Y

```
graph bar (median) price, over(foreign)
bar plot
    (axis) • (percent) • (count) • (stat: mean median sum min max ... )
    over(variables) • options: gap(#) • relabel • descending • reverse
    sort(variable(n)) • cw • missing • nofill • all/categories • percentages
    stack • bargap(#) • intensity(#) • vallernate • alternate
graph dot (mean) length headroom, over(foreign) m1(, ml(5))
dot plot
    (axis) • (percent) • (count) • (stat: mean median sum min max ... )
    over(variables) • options: gap(#) • relabel • descending • reverse
    sort(variable(n)) • cw • missing • nofill • all/categories • percentages
    stack • bargap(#) • intensity(#) • vallernate • alternate
graph hbox mpg, over(rep78, descending) by(foreign) missing
box plot
    over(variables) • options: total • gap(#) • relabel • descending • reverse
    sort(variable(n)) • missing • nofill • all/categories • percentages
    stack • bargap(#) • intensity(#)
    violin plot
    over(variables) • options: total • missing • nofill •
    vertical • horizontal • qdots • kernel(options) • bandwidth(#)
    width(n) • plotwidth(n) • plotgap(n) • intensity(options)
    bar(options) • median(options) • showout(options)
```

Plot Placement

JUXTAPOSE (FACET)

```
twoway scatter mpg price, by(foreign, norescale)
range plot (y, - y) with area shading
total • missing • colfirst • rows(n) • cols(n) • holes(nmlist)
compact • nolegend • byscale • noyrescale • noyscale
noxaxes • noyaxes • noytitle • noylabel
noxlabel • noxticks • noxticks • noymargin(options)
```

SUPERIMPOSE

```
graph combine plot1.gph plot2.gph...
combine two or more saved graphs into a single plot
scatter y3 y2 y1 x, msymbol(i o) mlabel(var3 var2 var1)
    plot several y values for a single x value
graph twoway scatter mpg price in 27/74 || scatter mpg price /* *
    * if mpg < 15 & price > 12000 in 27/74, mlabel(make) ml()
    combine twoway plots using ||
```

BASIC PLOT SYNTAX:

```
graph <plot type> variables: y first [in] [if], <plot-specific options> - facet -
    title("title") subtitle("subtitle") xtitle("x-axis title") ytitle("y axis title") xscale(range(low high) log reverse off noline) yscale(<options>)
    custom appearance
    axes
    plot size
    save
    annotations
```

<marker, line, text, axis, legend, background options> scheme(s1mono) play(customTheme) xsize(5) ysize(4) saving("myPlot.gph", replace)

Two+ CONTINUOUS VARIABLES

graph matrix mpg price weight, half

```
scatterplot of each combination of variables
half • jitter(#) • jitterseed(#)
diagonal • leastsq(variables)
```

twoway scatter mpg weight, jitter(?)

```
scatterplot
jitter(#) • jitterseed(#) • sort • crossings(gap)
connect(options) • (weight<variable>)
```

twoway scatter mpg weight, mlabel(mpg)

```
scatterplot with labelled values
jitter(#) • jitterseed(#) • sort • crossings(gap)
connect(options) • (weight<variable>)
```

twoway connected mpg price, sort(price)

```
scatterplot with connected lines and symbols
jitter(#) • jitterseed(#) • sort • less also line
connect(options) • crossings(gap) •
```

twoway area mpg price, sort(price)

```
line plot with area shading
sort • crossings(gap) • vertical • horizontal
base(#)
```

twoway bar price rep78

```
bar plot
vertical • horizontal • base(#) • barwidth(#)
```

twoway dot mpg rep78

```
dot plot
vertical • horizontal • base(#) • rdots(#)
color(<color>) • dcolor(<color>) • scolor(<color>)
dsize(markers(n)) • symbol(marker type)
dwidth(strokewidth(n)) • dotdiameter(n)
```

twoway dropline mpg price in 1/5

```
dropped line plot
vertical • horizontal • base(#)
```

twoway rcpssym length headroom price

```
range plot (y, - y) with capped lines
vertical • horizontal
see also rcap
```

twoway rarea length headroom price, sort

```
range plot (y, - y) with area shading
vertical • horizontal • sort
crossings(gap) •
```

twoway rbar length headroom price

```
range plot (y, - y) with bars
vertical • horizontal • barwidth(#) • midwidth
marker size)
```

twoway pccspike wage68 ttl_exp68 wage88 ttl_exp88

```
Parallel coordinates plot
vertical • horizontal
```

twoway pccapsym wage68 ttl_exp68 wage88 ttl_exp88

```
Slope/bump plot
vertical • horizontal • headlabel
```

THREE VARIABLES

twoway contour mpg price weight, level(20) crule(intensity)

```
3D contour plot
contour(#) • levels(#) • minmax • crule(hue | chue | linear)
color(<color>) • color(<color>) • colors(colorlist) • heatmap
interpolatepline | Shepard | none
```

regress price mpg trunk weight length turn, nocons

```
matrix regmat = e(V)
ssc install plotmatrix
plotmatrix, mat(regmat) color(green)
heatmap
mu(variable) • plot(options) • color(<color>) • freq
```

SUMMARY PLOTS

twoway mband mpg weight || scatter mpg weight

```
plot median of the y values
bands()
```

biscatter weight mpg, line(none)

```
ssc install biscatter
plot a single value (mean or median) for each x value
medians • quantities • discrete • rootdist(<variables>)
log(pettifit) | fit | connect | none) • smagfit(<variable>)
```

FITTING RESULTS

twoway ifitci mpg weight || scatter mpg weight

```
calculate and plot linear fit to data with confidence intervals
level(#) • stdp • stdf • nofit • fitplot(plottype) • plot(plottype)
gauge(#) • n(#) • atobs • plot(options) • smopt(options)
```

twoway lowess mpg weight || scatter mpg weight

```
calculate and plot lowess smoothing
width(#) • open • smooth • logit • adjust
```

twoway qfclt mpg weight, alwidth(none) || scatter mpg weight

```
calculate and plot quadratic fit to data with confidence intervals
level(#) • stdp • stdf • nofit • fitplot(plottype) • plot(plottype)
gauge(#) • n(#) • atobs • plot(options) • smopt(options)
```

REGRESSION RESULTS

regress price mpg headroom trunk length turn

```
coefplot, drop(_cons) xline(0)
Plot regression coefficients
Plot(jevels) • bl(options) • all(options) • nos • levels(#)
keep(<variables>) • drop(<variables>) • rename(<list>)
horizontal • vertical • generated(<variable>)
```

regress mpg weight length turn

margins, eyex(weight) at(weight = (1800/200)/4800)

marginsplot, nocl

Plot marginal effects of regression

horizontal • nos

Cheat Sheet: Personalización de Gráficos

Plotting in Stata

Customizing Appearance

For more info, see Stata's reference manual ([stata.com](#))



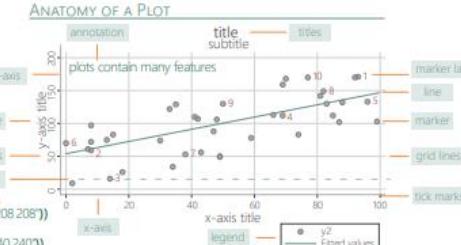
`scatter price mpg, graphregion(color(192 192 192)) ifcolor((208 208 208))`
specify the fill of the background in RGB or with a Stata color

`scatter price mpg, plotregion(color(224 224 224)) ifcolor((240 240 240))`
specify the fill of the plot background in RGB or with a Stata color

SYNTAX	MARKERS	LINES / BORDERS	TEXT														
<code>marker <marker options></code>	arguments for the plot objects (in green) go in the options portion of these commands (in orange)	<code>line <line options> <marker options> axes xscale(.) yscale(.)</code> <code>marker mcolor(...)</code> <code>xline(.)</code> <code>yline(.)</code> <code>legend legend(region...)</code>	<code>marker label <marker options></code> <code>titles title(.) subtitle(.)</code> <code>text annotation text(.)</code>														
<code>mcolor("145 168 208")</code>	specify the fill and stroke of the marker in RGB or with a Stata color	<code>color("145 168 208")</code> specify the stroke color of the line or border	<code>color("145 168 208")</code> specify the color of the text														
<code>mfcolor("145 168 208")</code>	specify the fill of the marker	<code>marker mcolor("145 168 208")</code> <code>tick marks tcolor("145 168 208")</code> <code>grid lines gcolor("145 168 208")</code>	<code>marker label mlabel("145 168 208")</code> <code>axis labels xlabel(...) ylabel(...)</code> <code>text annotation text(...)</code>														
<code>msize(medium)</code>	specify the marker size:	<code>width(medthick)</code> specify the thickness (stroke) of a line: <table border="1" style="margin-left: 20px; border-collapse: collapse;"> <tr><td>ehuge</td><td>medlarge</td></tr> <tr><td>vhuge</td><td>medium</td></tr> <tr><td>huge</td><td>medsmall</td></tr> <tr><td>vlarge</td><td>small</td></tr> <tr><td>large</td><td>vsmall</td></tr> <tr><td></td><td>tiny</td></tr> <tr><td></td><td>vtiny</td></tr> </table>	ehuge	medlarge	vhuge	medium	huge	medsmall	vlarge	small	large	vsmall		tiny		vtiny	<code>size(medsmall)</code> specify the size of the text: <code>marker label mlabelsize(medsmall)</code> <code>axis labels labsizes(medsmall)</code>
ehuge	medlarge																
vhuge	medium																
huge	medsmall																
vlarge	small																
large	vsmall																
	tiny																
	vtiny																
<code>msymbol(Dh)</code>	specify the marker symbol:	<code>ipattern(dash)</code> specify the line pattern: <table border="1" style="margin-left: 20px; border-collapse: collapse;"> <tr><td>solid</td><td>longdash</td><td>longdash_dot</td></tr> <tr><td>dash</td><td>shortdash</td><td>shortdash_dot</td></tr> <tr><td>dot</td><td>dash_dot</td><td>blank</td></tr> </table>	solid	longdash	longdash_dot	dash	shortdash	shortdash_dot	dot	dash_dot	blank	<code>text Text</code> <code>Text huge</code> <code>Text vlarge</code> <code>Text large</code> <code>Text medlarge</code> <code>Text medium</code>					
solid	longdash	longdash_dot															
dash	shortdash	shortdash_dot															
dot	dash_dot	blank															
<code>Position</code>	<code>jitter(#)</code> randomly displace the markers	<code>label(#10, tposition(crossing))</code> number of tick marks, position (outside crossing inside)	<code>marker label mlabposition(5)</code> marker location relative to marker (clock position: 0 – 12)														

Laura Hughes (ljhughes@usaid.gov) • Tim Essam (tessam@usaid.gov) • inspired by RStudio's awesome Cheat Sheets (rstudio.com/resources/cheatsheets)
 follow us @flaneuseks and @StataRGIS

Anatomy of a Plot



Apply Themes

Schemes are sets of graphical parameters, so you don't have to specify the look of the graphs every time.

USING A SAVED THEME

```
twoway scatter mpg price, scheme(customTheme)
```

help scheme entries Create custom themes by saving options in a .scheme file
see all options for setting scheme properties

adopath ++ " /<location>/StataThemes" set path of the folder (StataThemes) where custom .scheme files are saved
set as default scheme

set scheme customTheme, permanently change the theme

net inst breswscheme, from("https://wbuchanan.github.io/breswscheme/") replace install William Buchanan's package to generate custom schemes and color palettes (including ColorBrewer)

USING THE GRAPH EDITOR

`twoway scatter mpg price, play(graphEditorTheme)`



Select the Graph Editor
Click Record
Double-click on symbols and areas on plot, or regions on sidebar to customize
Unclick Record
Save theme as a .grec file

Save Plots

graph twoway scatter y x, saving("myPlot.gph") replace save the graph when drawing

graph save "myPlot.gph", replace save current graph to disk

graph combine plot1.gph plot2.gph... combine two or more saved graphs into a single plot

graph export "myPlot.pdf", as(pdf) see options to set export the current graph as an image file

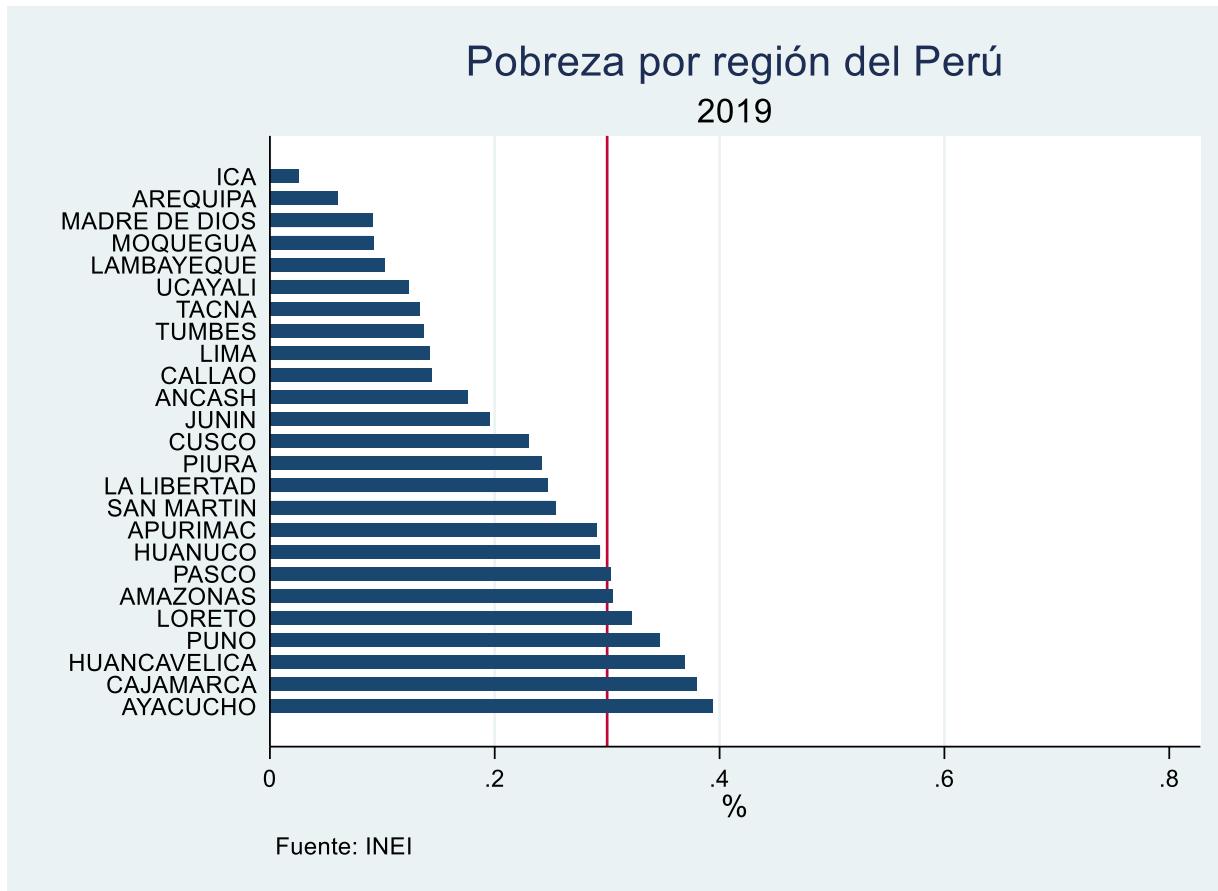
geocenter.github.io/StataTraining
Disclaimer: we are not affiliated with Stata. But we like it.

updated July 2019
CC BY 4.0

<https://www.stata.com/bookstore/statacheatsheets.pdf>

Ejemplo 1: Gráfico de Barra + Parámetros

```
graph hbar pobreza_n [w = fp], over(departamento, sort((mean) pobreza_n)
label(labsize(small))) ytitle("%") title("Pobreza por región del Perú") note("Fuente:
INEI") subtitle("2019") ylabel(0(0.2)0.8, labsize(small)) yline(0.3)
```



Ejemplo 2: Dos Gráficos en 1 sólo + Parámetros

```
graph twoway (scatter mieperho inghog2d[w = factor07],  
mcolor("0 155 255") msize(tiny)) (lfit mieperho inghog2d [w = factor07],  
lwidth(vthick)), title("Relación: Miembros del hogar e Ingreso") xtitle("Ingresos")  
ytitle("Miembros del hogar")
```

