

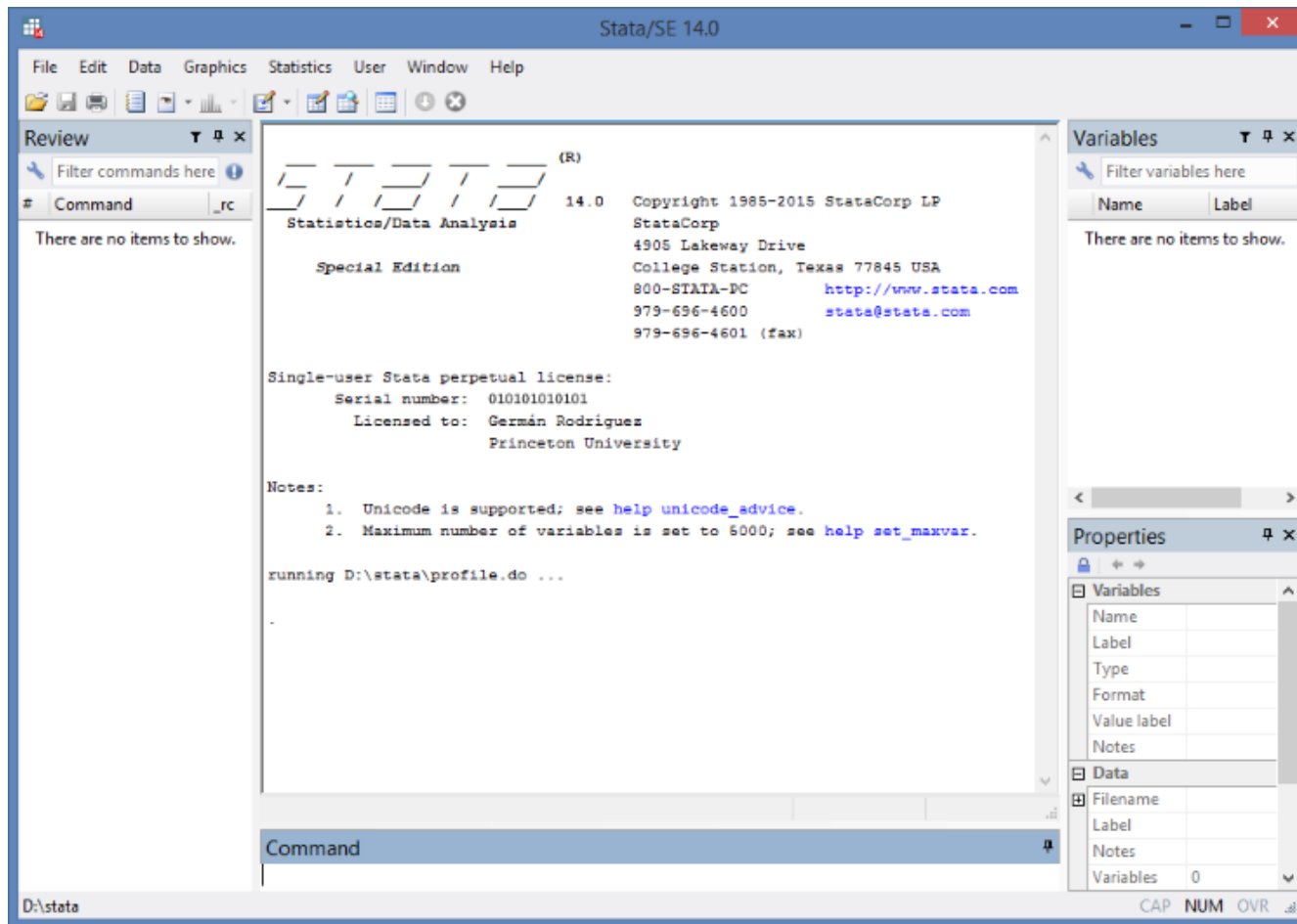
INTRODUCCIÓN A STATA



Panorama general

- Stata es un paquete estadístico de uso general ampliamente usado en economía, sociología, biomedicina y epidemiología
- Principales características:
 - ▣ Multiplataforma (Linux, Mac, Windows)
 - ▣ Extensible
 - ▣ Múltiples ediciones (MP, IC, SE, Small)
 - ▣ Comunidad activa (Statalist, Stack Overflow)

Interfaz Gráfica



Interfaz Gráfica

- Comandos
- Variables
- Output
- Historial
- Menús
- Propiedades

Comandos

- Sirve como calculadora

```
. display 2 + 2
```

```
4
```

- Tiene funciones estadísticas y matemáticas incorporadas

```
. display ttail(100, 0)
```

```
.5
```

Comandos (2)

- Son case-sensitive
- Se pueden abreviar (pero las funciones no)
- Page up y Page Down para ver la historia
- También está la ventana historial

Ayuda

- Es muy importante saber orientarse con la ayuda
 - . help ttail
 - . help regression (q para salir)

Cargar un archivo de muestra

- Stata tiene algunos archivos incorporados como muestra

. sysuse lifeexp

- Para ver la estructura del archivo

. desc

- ▣ Variable name
- ▣ Variable label
- ▣ Storage Type
- ▣ etc.

Estadísticas descriptivas

- . summarize

- gnppc tiene datos faltantes

- . list country if missing(gnppc)

- Muestra la variable country para las observaciones que cumplen la condición missing(gnppc)

- En mayor detalle

- . summ gnppc, detail

Factores

- . tab region
- Es categórica pero aparece como numérica
- . sort region
- . by region: summ gnppc

Pregunta de análisis

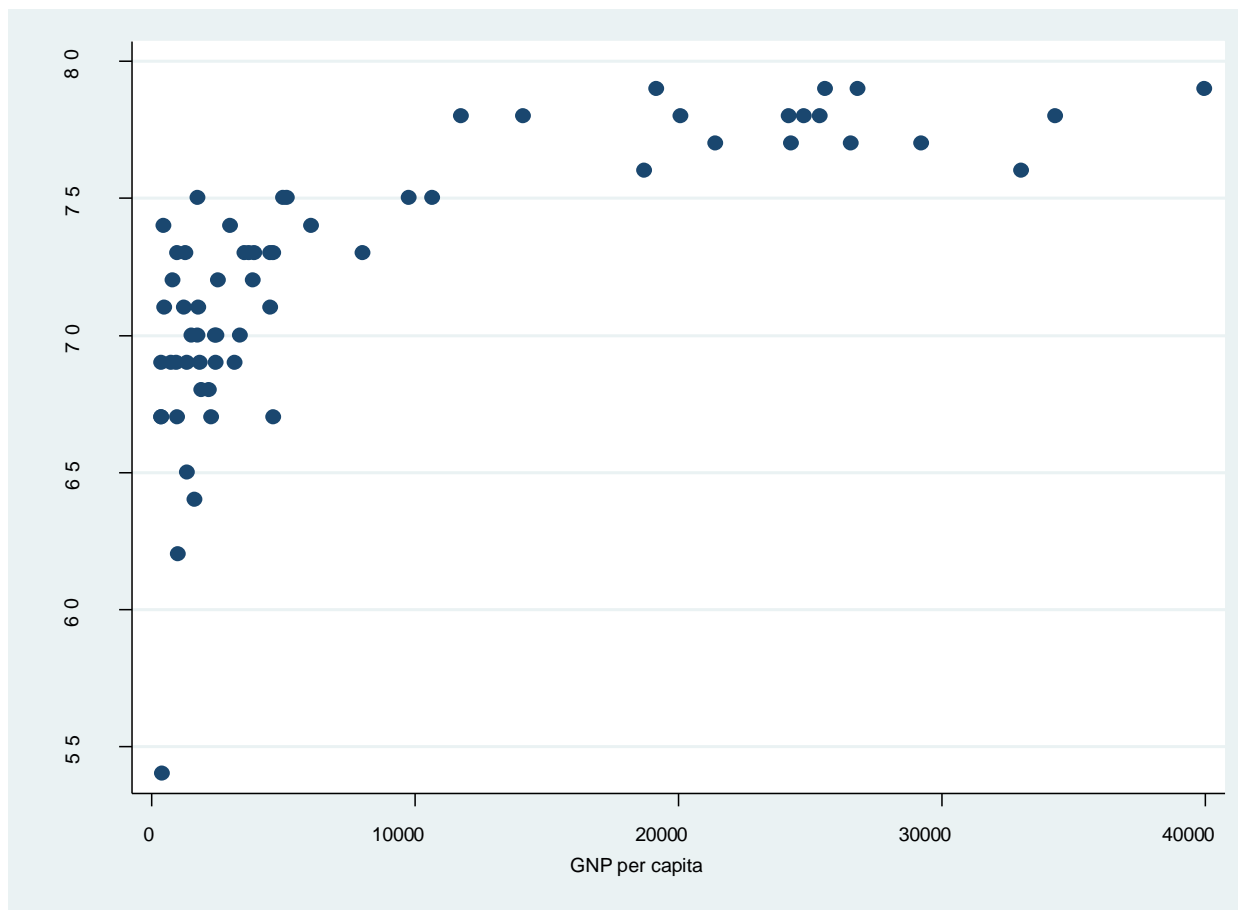
- Los países ricos tienen mayor expectativa de vida?

Gráfico

- Para entender la relación entre el PIB per cápita y la esperanza de vida podemos hacer un gráfico
- ¿Qué esperamos ver?
- . graph twoway scatter lexp gnppc

Gráfico (2)

- La correlación es positiva pero no lineal



Transformación

- Para linealizar la relación calculamos el log del PIB per cápita
- generate log_gnppc = log(gnppc)

Un Modelo para la relación

. reg lexp log_gnppc

- Stata omite las observaciones con datos faltantes de alguna de las variables del modelo
- Inferencia
- Interpretación

Comandos post-estimación

- Luego de hacer una estimación, tenemos acceso a los resultados es *macros*.

. *return list*

. *ereturn list*

- Estos resultados pueden ser útiles para otros cálculos

Comandos post-estimación (1)

- También podemos crear nuevas variables en base al modelo estimado
 - . predict lexp_hat
 - . predict resid, residuals
- Y hacer pruebas de hipótesis
 - . test log_gnppc
- Este test es equivalente a la salida anterior, pero en modelos más complejos podemos querer hacer otras cosas que no aparecen en la salida normal

Otro Gráfico

- Podemos graficar las variables y la recta de regresión
- `graph twoway (scatter lexp log_gnppc) (lfit lexp log_gnppc)`

Outlier

- Hay una observación que tiene una esperanza de vida baja respecto al PIB per cápita
- ```
. list country lexp lexp_hat if lexp < 55
```

# Uruguay

---

. list if country == "Uruguay"

# No trabajar con los menús

- Por qué no trabajar con el mouse/con comandos interactivos
- La ayuda online esta pensada para trabajar con comandos
- No es reproducible
  - ▣ Trabajo en equipo
  - ▣ Cambios en los datos
  - ▣ Apagar la PC

# Log files

- Guardan los comandos y el output

- . log using esperanza\_de\_vida.log

- ...

- . log close



# Comentarios

---

- Difícil saber que comentar y que no, pero importante tratar

# Sintaxis

- [by varlist:] command [varlist] [=exp] [if exp] [in range] [weight] [using filename] [,options]
- -[] indica un elemento opcional
- command nombre del comando (obligatorio)

# Sintaxis (2)

- varlist lista de variables
- . summarize gnppc lexp
- =exp si generamos una nueva variable es la fórmula
- . generate log\_gnppc = log(gnppc)

# Sintaxis (3)

- if exp para correr el comando solo usando un subconjunto de observaciones
- . list country gnppc if missin(gnppc)
- , options son opciones que modifican el comportamiento del comando, van dps de la coma
- . summarize gnppc vs summarize gnppc, detailed
- by varlist

# ¿Qué hicimos?

- Familiarizamos con la interfaz de Stata
- Analizamos un data set para resolver una pregunta
- Estadística descriptiva
- Gráficos
- Especificación y estimación de un modelo para contestar la pregunta de análisis