

Introducción a Stata

Econometría I

Pavel Solís

Universidad Anáhuac

2025

Índice

Interfaz del Usuario

Sintaxis

Archivos do

Limpieza de Datos

Actividad

1. Crea folder **eco3404** en Documentos
2. Descarga el archivo **auto.xlsx** de la página del curso
3. En el archivo en Excel, calcula lo siguiente:
 - Número de variables
 - Número de observaciones
 - Promedios
 - Desviaciones estándar
 - Valores mínimos
 - Valores máximos

Interfaz del Usuario

Programas para el análisis estadístico

Comercial (licencia):

- Excel
- Stata
- Matlab
- EViews
- SAS

Código libre:

- R
- Python

Stata

Permite manipular, visualizar y analizar datos, así como generar reportes

- La primera versión se lanzó en 1985

Ventajas:

- Fácil de aprender
- Ampliamente usado (economía, biomedicina, epidemiología, sociología)
- Actualizado (nueva versión cada 2 años)
- Permite usar funciones desarrolladas por usuarios

Experimentar conforme van aprendiendo es la mejor forma de aprender

Pantalla principal

La pantalla principal de Stata muestra 5 ventanas:

- Variables
- Propiedades
- Comandos
- Resultados
- Historial

Cambiar ventanas: Ctrl+1-5 o Ctrl+Tab

Ajustar ventanas:

- Tamaño (orillas), ocultar/mostrar (tachuela), reubicar (arrastrar)

Ruta de trabajo y barra de estado

La ruta de trabajo es donde Stata buscará y guardará archivos

- El comando `pwd` despliega la ruta de trabajo (útil si conexión vía Citrix)

La barra de estado está en la parte inferior de la pantalla principal

- La esquina inferior izquierda muestra la ruta de trabajo

El comando `cd` permite cambiar la ruta de trabajo

```
cd ~\Documentos\eco3404
```

- También se puede copiar la ruta del Explorador de archivos
- Si hay espacios en el nombre de la ruta, se usan comillas
- Se puede verificar el cambio de ruta en la barra de estado

Menús y barra de herramientas

Menús

- Acciones: Archivo, Edición, Ventana, Ayuda
- Instrucciones: Datos, Gráficos, Estadísticas
- Personalización: Usuario

Barra de herramientas: Acceso rápido a funciones más utilizadas

- Acciones: Abrir, Guardar, Imprimir, Bitácora (log), Continuar, Detener
- Ventanas independientes:
 - Visor de ayuda (archivos smcl), Gráfica, Editor de archivos do, Editor de datos (editar, ver), Editor de variables

Dar instrucciones

Hay 2 opciones para dar instrucciones a Stata

- Menús y ventanas: Selección vía clics
 - Es natural, reporta comandos (útil para aprender)
 - Ej. Importar archivo **auto.xlsx** y usar el Editor de datos para visualizar observaciones y variables
- **Comandos**: Instrucciones escritas para que Stata realice una acción
 - A menudo toman argumentos (variables, opciones, expresiones) para especificar los detalles de la acción
 - Una vez que se conocen, es más rápido que vía clics
 - **Funciones** (ej. `sqrt`) calculan valores y deben usarse con un comando

Interfaz del usuario: Variables

Muestra la lista de variables en la base de datos

Opciones:

- Configuración de columnas (clic derecho en nombre de columna)
- Selección de variables (clic, Shift+clic vs Ctrl+clic)
- Enviar a v. de comandos (doble clic vs flecha de pegado vs clic derecho)
- Retener y descartar variables (clic derecho, genera comandos)
- Búsqueda de variables (filtro y llave inglesa refina)
- Ordenar (oscila entre ascendente, descendente, original).

Interfaz del usuario: Propiedades

Muestra las propiedades de las variables y de la base de datos

El candado permite editar (genera comandos)

- Opciones para variables:
 - Cambiar nombre, etiqueta, tipo de valor (ej. float, double, int), formato
 - Agregar notas
- Opciones para bases de datos:
 - Cambiar nombre, etiqueta
 - Agregar notas
 - Ordenar por lista de variables

Interfaz del usuario: Comandos

Características:

- Permite editar texto (Ctrl+c, v, x, z),
- Permite borrar todo (Esc)
- Permite autocompletar variables (Tab)
 - No funciona para comandos
- Permite acceder a la historia de comandos (Re Pág, Av Pág)

Stata puede usarse como calculadora

```
display 5+25+3.2
```

```
display sqrt(4)+exp(0)+5.23
```

```
display invttail(25,0.1)
```

Interfaz del usuario: Resultados

Muestra los resultados tras ejecutar los comandos

Características:

- Permite buscar comandos y resultados (Ctrl+F)
- Menú (clic derecho)
 - Seleccionar todo
 - Limpiar pantalla (igual que comando `cls`)
 - Preferencias (color de fondo)
- Tras hacer una selección, menú (clic derecho)
 - Copiar, copiar cuadro, copiar tabla, copiar como imagen

Interfaz del usuario: Historial

Permite revisar la historia de comandos

Características:

- Código de colores: Resultados en negro, errores en rojo
- Búsqueda de comandos (filtro y llave inglesa refina)
- Permite ocultar errores (signo de exclamación)
- Permite reutilizar comandos (1 vs 2 clics, Shift+clic vs Ctrl+clic)
- Menú (clic derecho)
 - Cortar, copiar, eliminar, seleccionar todo, borrar todo, ejecutar lo seleccionado, enviar lo seleccionado a archivo do

Sintaxis

Sintaxis

En general, la sintaxis en Stata es (en minúsculas, en inglés):

```
[by varlist:] command [varlist] [if exp] [in range] [, options]
```

Ej. `help list`

- Subrayado: Abreviación mínima (para comandos y opciones)
- Corchetes: Indican elementos opcionales
 - No corchetes significa requerido
- Lista de variables (opcional o explícita): Permite notación corta
- Calificativo `if`: Usa expresiones lógicas
- Calificativo `in`: Define rango de observaciones
- Las opciones del comando se incluyen después de la coma

Comandos más usados

Para obtener ayuda y navegar directorios

- help, search, which, pwd, cd

Para usar y guardar datos

- clear, use, append, merge, import, save

Para obtener información básica de los datos

- describe, codebook, list, browse, count, tabulate, by, summarize, correlate

Para manipular de datos

- generate, replace, egen, rename, drop, keep, sort, encode, order, reshape

Otros

- log, display, histogram, scatter

Archivos do

Archivos auxiliares de Stata

Tipos de archivos:

- .do (comandos)
- .dta (datos)
- .ado (adiciones, definen comandos/funciones)
- Otros: .smcl (bitácora, ayuda), .dct (diccionario), .stbcal (calendario), .sthlp
 - smcl por Stata Markup Control Language

Comandos y paquetes escritos por usuarios se pueden descargar gratis (ssc):

- estout, outreg2, coefplot, distinct

Archivos do

Archivos de texto con comandos de Stata

- Líneas en blanco se ignoran
- Código de colores:
 - Azul (palabras clave), rojo (comillas), verde (comentarios), negro (resto)
- Comentarios: * (una línea), /* */ (varias líneas), // (misma línea)
 - Comentar bloque de texto (Ctrl + }), para correr parte no comentada
 - Seccionar: ***, *===, *---, *%%%
 - Documentar: Título, autor, fecha, propósito, versión
 - Continuación de líneas largas (vinculador): /// (precedido por espacio)
- Guardar archivo frecuentemente (Ctrl + S): Nombre del archivo sin espacios

Archivos do

- Ejecutar código (Ctrl + D, botón Ejecutar, comando do):
 - Línea (selección con triple clic)
 - Bloque de líneas (seleccionar partes de primera y última líneas)
 - Todo (sin selección)
- Bitácora (log): Al terminar de editar el archivo
 - Guarda resultados (no gráficas) como texto
 - Evita volver a ejecutar el archivo do (útil si conexión vía Citrix)
 - log using C1-Intro.log, replace (al inicio)
 - log off, log on (intermedio, opcional)
 - log close (al final para guardarlo)
 - translate C1-Intro.log C1-Intro.pdf, replace (opcional)

Limpieza de Datos

Limpieza de Datos

Buen uso de archivos do

- Seguir buenas prácticas para ellos (ej. seccionar)

Buenas prácticas:

- Estudia la base de datos antes de realizar cualquier análisis
- Mantén la base de datos original, guarda la base limpia en otro archivo
- Documenta tu trabajo para que sea reproducible y corregible (archivos do)
- Comenta tu código para otras personas y para ti mismo en el futuro
- Implementa verificaciones para validar que el código hace lo que esperas

Convención para nombrar variables

Caracteres permitidos:

- Los nombres de variables pueden contener letras, números y guiones bajos
- El primer caracter no puede ser un número
- Los guiones bajos pueden servir para separar palabras

Stata distingue entre mayúsculas y minúsculas

- Ej. age, Age y AGE se consideran tres variables distintas
- Es común que los nombres de las variables estén en minúsculas

Se recomienda usar nombres descriptivos cortos

- Los nombres de las variables no pueden superar los 32 caracteres