

Notas de clase – Taller de Stata¹
Clase 10 – Gráficos I - Clasificación

Contenido

1. Introducción
 - 1.1. Estructura básica de un gráfico en *Stata*
 - 1.2. Menús y editor de gráficos
2. Gráficos básicos
 - 2.1. Barra y caja de distribución
 - 2.2. Pastel
3. Familia de gráficos de doble entrada (*twoway*)
 - 3.1. Sintaxis del comando
 - 3.2. Barra y Línea
 - 3.3. Dispersión
 - 3.4. Otros

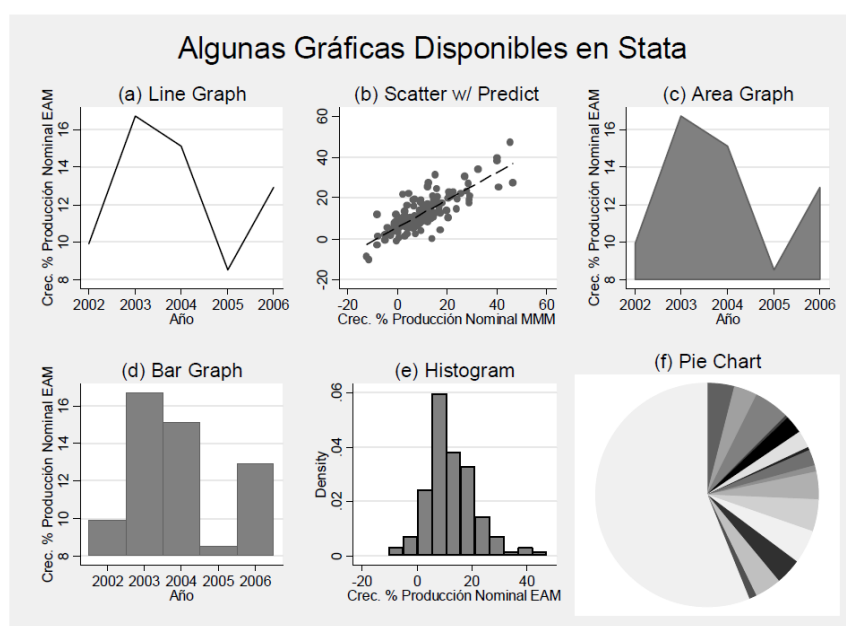
[Estas notas se diseñaron con base en la versión 16 de *Stata*, conforme el software cambie las notas deben ser actualizadas].

¹ Estas notas están basadas en la guía desarrollada previamente para este curso por Rodrigo Azuero Melo, Nicolás de Roux, Luis Roberto Martínez, Román Andrés Zárate y Santiago Gómez.

1. Introducción

Stata cuenta con herramientas para construir una gran cantidad de gráficos. En esta clase vamos a explorar las capacidades más importantes de este *software*, iniciando con gráficos sencillos y el uso sucesivo de opciones para agregar nuevas características o modificar las características disponibles por defecto. El objetivo es sintetizar los conceptos más importantes para entender cómo se realizan gráficos en *Stata*. De ninguna manera se busca reemplazar la documentación existente, la cual se sugiere consultar de forma sistemática para encontrar las soluciones deseadas.

Figura 1. Algunos Tipos de Gráficos disponibles en *Stata*.



1.1. Tipos de gráficos

Podemos pensar en dos grandes tipos de gráficos en *Stata*, los gráficos *básicos* y los gráficos *twoway*. En primer lugar, explicamos cómo realizar gráficos básicos junto con las opciones para editar, guardar, exportar, usar, mostrar, etc. Luego mostramos cómo realizar gráficos *twoway* sencillos. Finalmente mostramos cómo realizar otros gráficos más complejos por el tratamiento numérico que utilizan.

Los gráficos tienen dos ejes, el Y y el X, de tal manera que se puede analizar el comportamiento de una variable respecto a la otra. Los gráficos básicos son utilizados principalmente para mostrar el comportamiento numérico de *una* variable (eje Y) respecto a otra (eje X). La variable del eje

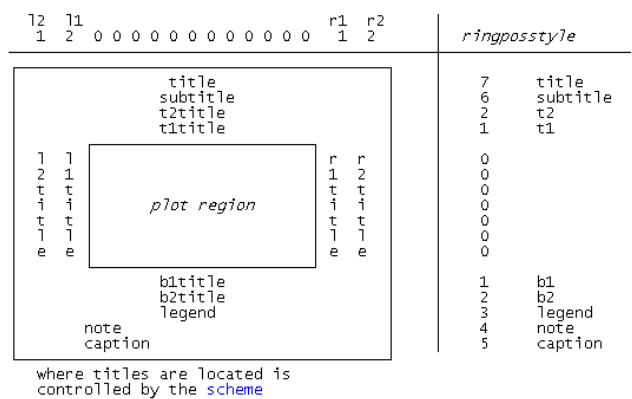
X puede entenderse como una variable independiente y puede ser categórica (ordinal o nominal). Por ejemplo, estos gráficos permiten ver el comportamiento de una variable por categorías en barras, puntos, cajas de distribución, gráficos de pastel (también llamados gráficos de tarta, de torta, de 360° o *graph pie-chart*), entre otros.

Los gráficos *twoway* (o de doble entrada) son utilizados principalmente para mostrar la relación entre datos numéricos, esto es, gráficos que tienen escalas numéricas en sus ejes X y Y. En la *Figura 1* se muestran algunos tipos de gráficos. Dada esta clasificación, solo el gráfico de pastel (f) es un gráfico exclusivamente básico, el gráfico de barra puede ser básico o *twoway* y los demás son gráficos del tipo *twoway*.

1.2. Estructura básica de un gráfico en *Stata*

En la *Figura 2* se muestra la estructura básica de un gráfico en *Stata*. Un gráfico tiene dos áreas básicas, el área de dibujo que muestra la información a ilustrar y el área descriptiva que permite leer el gráfico de forma directa a través del título, los nombres de los ejes, las leyendas, etc. Cada característica del gráfico es un objeto que puede ser controlado a través de alguna orden en la sintaxis del comando. *Stata* construye los gráficos con un formato por defecto que puede ser cambiado a través de la opción *scheme*.

Figura 2: Estructura básica de un gráfico en *Stata*

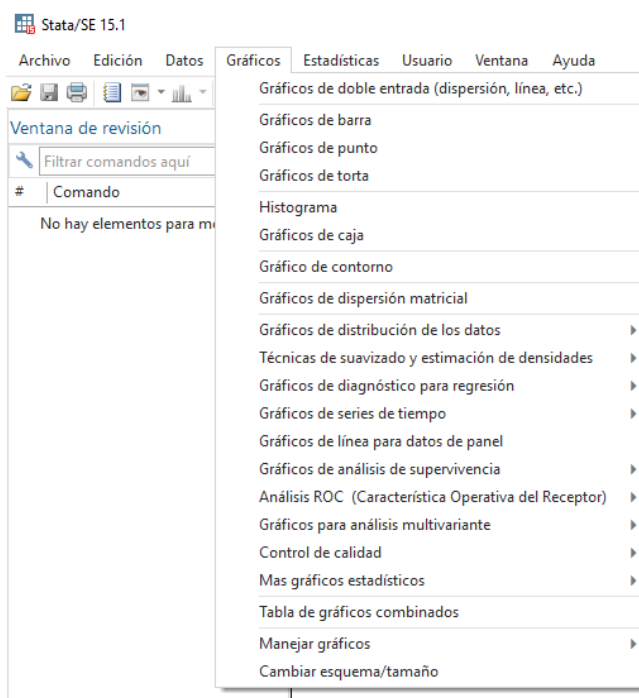


1.3. Menús y editor de gráficos

En clase nos hemos acostumbrado a manejar todas las órdenes dadas a *Stata* a través de líneas de código ya que es la forma más eficaz de utilizar *Stata* para trabajar de forma trazable y replicable. Sin embargo, *Stata* cuenta con algunas opciones en sus menús desplegables para realizar gráficos con la asistencia de ventanas y un editor manual. Estas opciones pueden ser

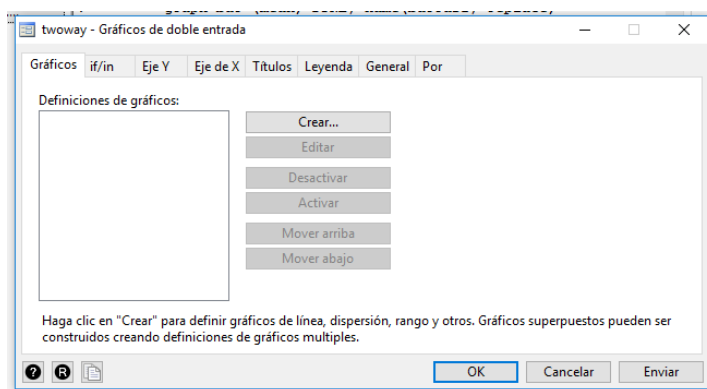
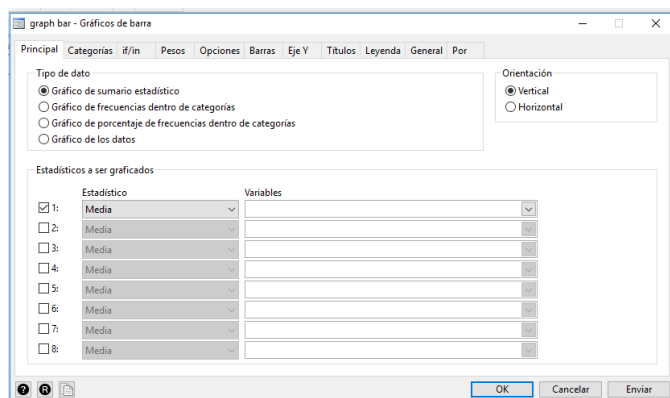
útiles para los usuarios que no están familiarizados con los comandos de *Stata* y para realizar gráficos sencillas y pequeños cambios sin comprometer algún código ya escrito. Aquí simplemente mostramos cómo acceder a estas herramientas y queda para el usuario su exploración. En la *Figura 3* se muestran las opciones del menú desplegable para gráficos básicos. En la *Figura 4* se muestra la ventana donde se ingresan las opciones para la construcción de un gráfico de barra y también para la construcción de un gráfico de doble entrada (*twoway*).

Figura 3: Menú desplegable para gráficos en Stata



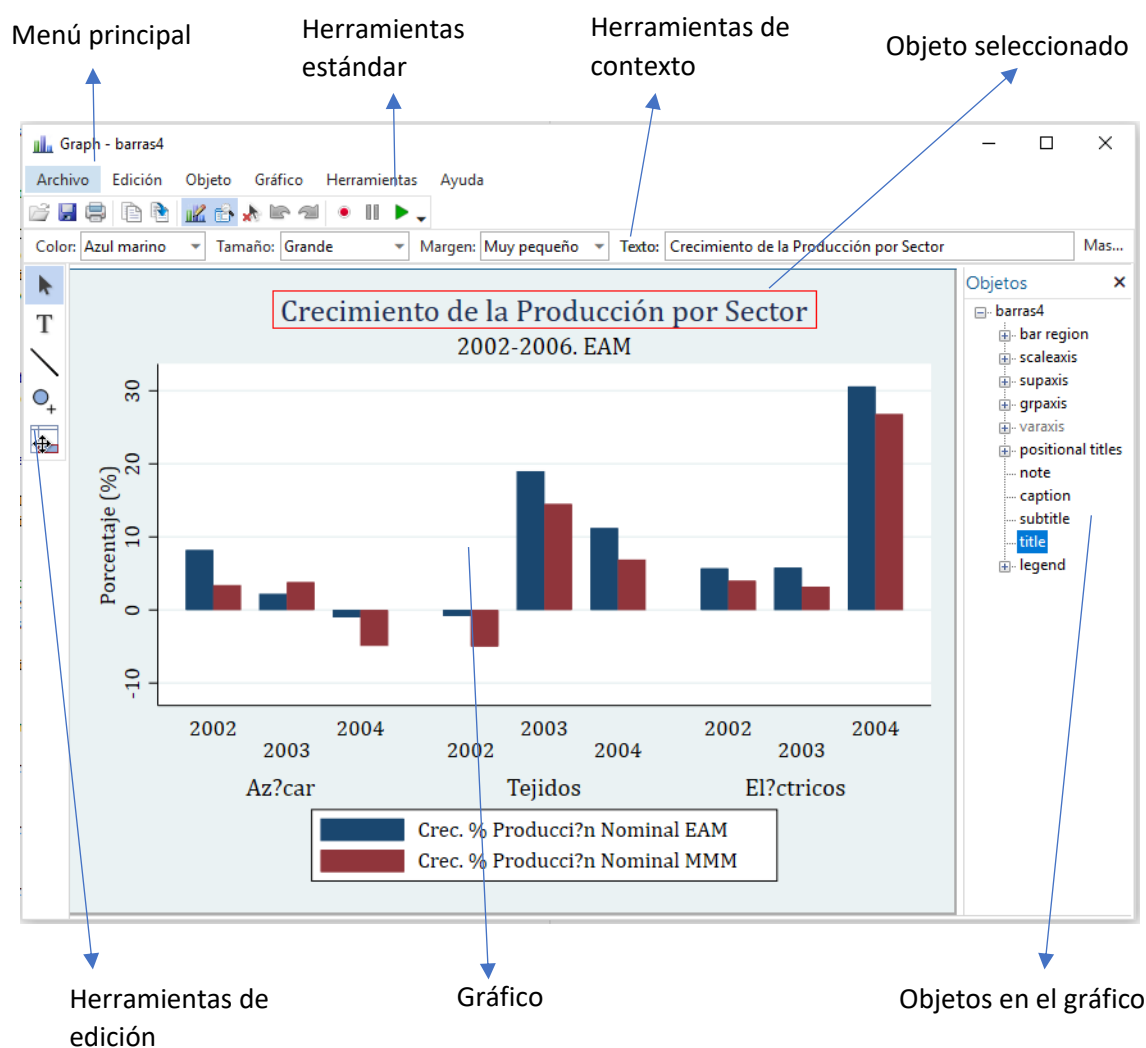
4

Figura 4: Ventanas para construir gráficos de barras y de doble entrada (*twoway*)



Cuando se crea un gráfico, por defecto, *Stata* abre una ventana para ver el resultado. En la barra de herramientas estándar de la ventana hay un botón llamado “Iniciar editor de gráficos”. En la *Figura 5* se muestra el editor de gráfico iniciado y se señalan las partes de la ventana. Se puede seleccionar cualquier objeto del gráfico y en las herramientas de contexto se pueden hacer algunas modificaciones sencillas, al hacer doble clic a algún objeto se abre otra ventana con más opciones de contexto para modificar el objeto.

Figura 5: Ventana del editor de gráficos



2. Gráficos básicos

Vimos que estos gráficos son utilizados principalmente para mostrar el comportamiento numérico de una variable respecto a otra que puede ser categórica. Los principales tipos de gráficos básicos son las barras, los pasteles y las cajas de distribución. La sintaxis de un gráfico básico es la siguiente:

graph *plottype* [(*stat*)] [*varlist*], *over*[*var*, *options*] *over*[*var*, *options*] (*general options*)

La sintaxis contiene el tipo de gráfico (*plottype*), y [*varlist*] se refiere a la o las variables de interés. Si el gráfico es de barras, antes de cada variable de interés se enuncia el estadístico a utilizar, el estadístico por defecto es el promedio. Además, antes de la coma se puede especificar algún condicional. La variable independiente (o categórica, que en un gráfico de barras o en la caja de distribución se ubica en el eje X) se declara con la opción *over*, en su interior se puede especificar otras opciones, como la ubicación de la leyenda del eje X. Las opciones generales corresponden a las características generales, como títulos, leyendas y otros elementos.

2.1 Barra y caja de distribución

El gráfico básico más utilizado es la gráfica de barras. A continuación, se muestra un ejemplo de una gráfica de barra con algunas de sus opciones más importantes.

Ej. 1 →

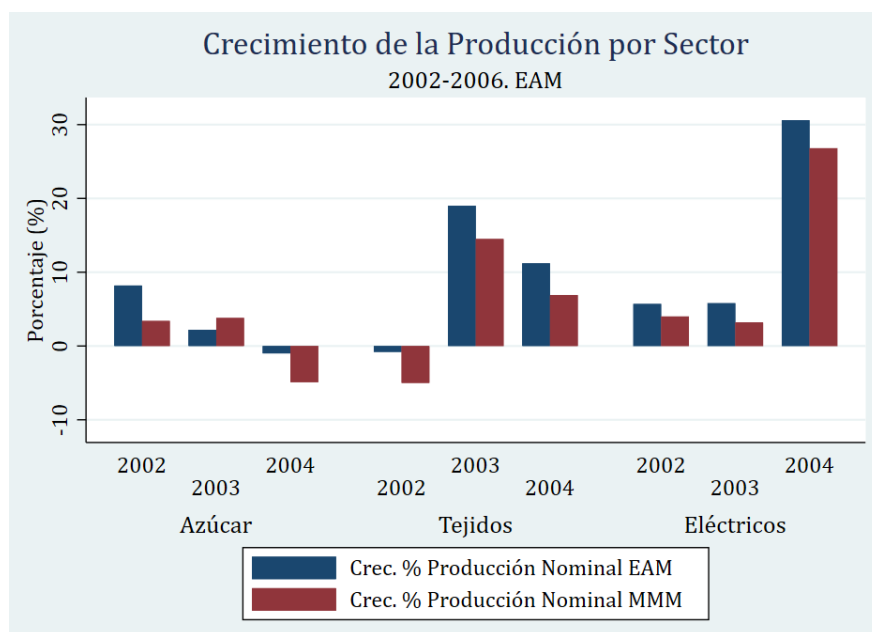
```
graph bar (asis) DPNE DPNM /// tipo de gráfico, estadístico y variables de interés
if (SECTOR==15|SECTOR==20|SECTOR==33) & YEAR<2005, /// Condicionales
over(YEAR, label(alt)) over(SECTOR) /// Doble over
legend(r(2)) /// Colocar en columna la leyenda
ytittle(Porcentaje (%)) /// Título del eje Y
title(Crecimiento de la Producción por Sector) subtitle(2002-2006. EAM) /// Opciones
name(barras, replace) /// guarda en la memoria de la sesión
saving(barras, replace) // guarda en el directorio de trabajo
```

En esta gráfica se muestra el porcentaje de crecimiento de la producción nominal para tres sectores de la industria colombiana, Azúcar, Tejidos y Eléctrico. Las barras azules son datos de la Encuesta Anual Manufacturera y las barras rojas son datos de la Encuesta Muestra Mensual Manufacturera. En la primera línea del gráfico se declaran las variables de interés y por defecto se asigna un color diferente para cada una. En el condicional se limitan los datos de las categorías

a considerar (variables independientes, eje X). Note que para este gráfico se utilizan dos categorías, sector y año. El orden de categorización está dado por el orden en que se declaran las opciones *over*. Para incluir la leyenda en la gráfica se debe utilizar la opción *legend(r(2))*, donde 1 denota que la leyenda se coloca de forma horizontal y 2 denota que la leyenda se coloca de forma vertical.

En la primera línea del comando, donde se declara el estadístico se colocó la instrucción *axis*. Esta permite que se calcule el promedio de las variables de interés según las categorías y que en la leyenda aparezcan las etiquetas de las variables.

Figura 6: Gráfico de barras en *Stata*



Cuando creamos un gráfico en *Stata* debemos guardarla, ya que *Stata* la borrará al momento de generar una nueva gráfica. Los gráficos se pueden guardar en la memoria temporal o en el disco duro. Si se guardan en la memoria, los gráficos se borrarán una vez se cierre la sesión de *Stata*, si se guardan en el disco duro, éstas estarán disponibles para futuras sesiones. Para guardar en memoria se usa la opción `[, name(nombre deseado)]`. Para guardar en el disco duro se emplea la opción `[, saving(archivo deseado)]`. En ambos casos se puede añadir la subopción `[, replace]`. En el caso de guardar el gráfico en el disco duro, si no se especifica la ruta completa para el archivo, *Stata* asumirá que se desea guardar la gráfica en el *current directory*.

El uso de las cajas de distribución es homólogo a los gráficos de barras, con la diferencia de que no se requiere algún estadístico.

2.2 Pastel

El gráfico de pastel se utiliza para mostrar la contribución de diferentes categorías en una variable. Solo se puede utilizar una variable de interés y una variable categórica. En nuestro caso, en el siguiente ejemplo podemos ver cuáles sectores contribuyen más al crecimiento de la producción en el año 2003.

Ej. 2 →

```
graph pie DPNE if YEAR==2003 & SECTOR!=0, over(SECTOR) /// Puede usar
condicionales
scheme(economist) /// Estilos de colores, más en la documentación (p. 689)
sort descending /// Organiza las categorías
title(Crecimiento de la producción por sector en 2003, size(12)) ///
saving(pie2, replace) /// Guarda la gráfica en el directorio
name(pie2, replace)
```

Figura 7: Gráfico de pastel en Stata



En este gráfico se muestra la distribución del crecimiento de la producción nominal por sector del año 2003. Se observa que la sintaxis es muy similar a la mostrada para el gráfico de barras, las diferencias radican en las limitaciones de la gráfica misma. Se utiliza la opción *scheme* para

cambiar el formato por defecto, en este ejemplo se utiliza el formato de la revista *The Economist*. Además, se muestra cómo cambiar el tamaño del título del gráfico.

3. Familia de gráficos de doble entrada (*twoway*)

Esta familia de gráficos es la más utilizada en investigación porque ofrece una gran cantidad de opciones para mostrar relaciones entre variables, especialmente numéricas.

3.1 Sintaxis de comando

Respecto a los gráficos básicos, la sintaxis del comando cambia significativamente. A continuación, se explica la sintaxis general.

graph twoway (plottype varlist,options) (plottype varlist,options), options

Todas las líneas de código para realizar gráficos empiezan con el comando *graph*, para los gráficos de doble entrada se puede simplificar escribiendo solo “*twoway*” o “*tw*”, en algunos tipos de gráfico solo es necesario escribir el nombre del tipo. En la sintaxis se muestran paréntesis de órdenes, esto indica que en la misma línea de código se puede construir varios gráficos (dibujos) que van a ser mostrados bajo los mismos ejes. *plottype* hace referencia al tipo de gráfico a realizar, por ejemplo, de línea, de caja, de dispersión (*scatterplot*), de área bajo la curva, etc. *varlist* hace referencia a las variables que se desean graficar. Antes de la coma en cada paréntesis se puede agregar condicionales y después de la coma se escriben las opciones para cada (gráfico) dibujo. Se infiere entonces que existen opciones específicas a cada dibujo, como colores y tamaños de los objetos, y para otras opciones que aplican el gráfico compuesto, como los títulos y escala de ejes y títulos, subtítulos, notas, etc., que son objetos fuera del área de dibujo.

9

3.2 Barra, línea

A continuación, se muestra la sintaxis de dos gráficos sencillos, de barras y de línea respectivamente:

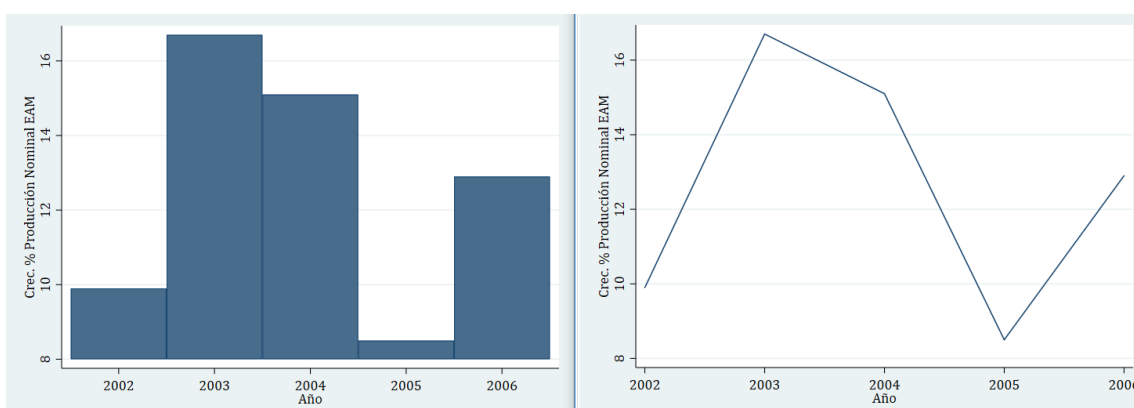
Ej. 3 → tw bar DPNE YEAR if SECTOR==0, name(barras, replace)

 line DPNE YEAR if SECTOR==0, name(linea, replace)

Estos comandos se muestran de manera abreviada. En la lista de variables, la última de ellas se coloca en el eje X, las demás se muestran sobre el eje Y. Para ambos gráficos observamos datos

Respecto al gráfico de barras *twoway*, se quiere ver la evolución del crecimiento de la producción de la industria colombiana a través del tiempo. El sector 0 corresponde al valor del crecimiento de la producción para toda la industria y es una observación existente para cada año. La gráfica de línea de interpreta de manera homóloga.

Figura 8: Gráficos de barras y línea *twoway* en *Stata*



3.3 Dispersión

El gráfico de dispersión o *scatterplot* ilustra a través de puntos el comportamiento de una variable numérica respecto a otra. Del comportamiento de los puntos se podría leer una tendencia.

A continuación, se utilizan ejemplos de construcción de gráficos de dispersión para mostrar sus principales opciones.

Ej. 4 →

`twoway (scatter DPNE YEAR) (lfit DPNE YEAR) if SECTOR==00, name(scatter3, replace)`

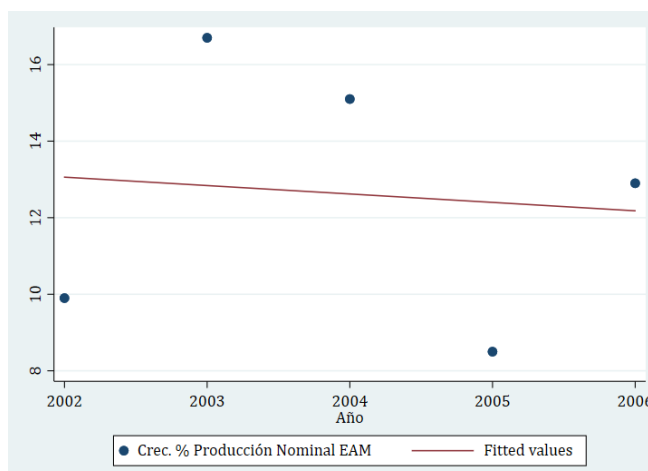
En la *Figura 9* muestra la evolución del crecimiento de la producción nominal con los datos de la Encuesta Anual Manufacturera entre los años 2002 y 2006. Observe que se utilizan dos paréntesis de órdenes, cada uno realiza un dibujo diferente en la gráfica. El primer paréntesis realiza un gráfico de dispersión, se especifican las variables igual que en ejemplos anteriores. En

el segundo paréntesis se realiza una línea de tendencia con las mismas variables. Para todo el gráfico se condiciona que se tomen los datos para todos los sectores (SECTOR==0).

Ej. 5 →

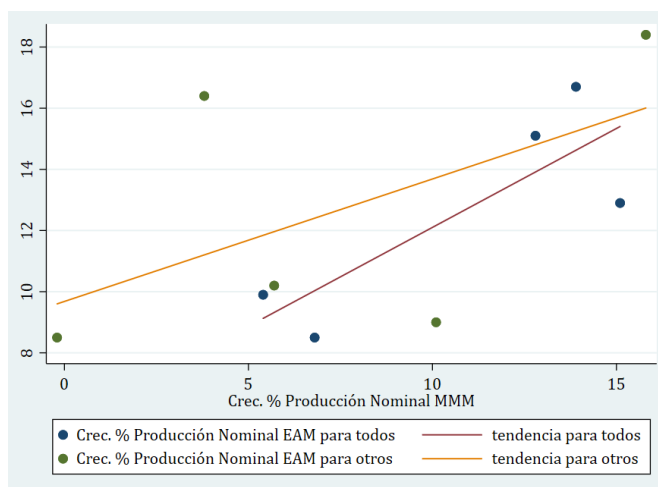
```
local t: variable label DPNE
twoway      (scatter DPNE DPNM if SECTOR==00) ///
            (lfit DPNE DPNM if SECTOR==00) ///
            (scatter DPNE DPNM if SECTOR==36) ///
            (lfit DPNE DPNM if SECTOR==36) ///
, name(scat_3, replace) ///
legend(order(1 "`t' para todos" 2 "tendencia para todos" 3 "`t' para otros" 4 "tendencia para otros"))
```

Figura 9: Gráfico de dispersión y la línea de tendencia



11

Figura 10: Varios gráficos de dispersión y tendencia



Para todos los gráficos de la familia *twoway* se puede observar que la leyenda se agrega de manera automática y que, a diferencia de los gráficos básicos, por defecto utiliza las etiquetas de las variables utilizadas. En el *ejemplo 5*, cuyo resultado se muestra en la *Figura 10*, se puede ver que la agregación de más dibujos dentro de un gráfico *twoway* se hace de manera homóloga. Para los dos gráficos de dispersión se especifica directamente el contenido de la leyenda a través de un objeto local que contiene la etiqueta de la variable y para las líneas de tendencia se escribe directamente el contenido.

3.4 Otros

En clase y en la documentación² se muestra que *Stata* puede hacer una gran variedad de gráficos. De los más frecuentemente utilizados son el histograma, el gráfico de puntos conectados y el gráfico de área. La sintaxis y las opciones de estos tipos de gráficos son muy similares a las ya estudiadas, las diferencias de sintaxis radican en las limitaciones de cada gráfico y lo que se quiere ilustrar con cada uno.

² Disponible en: <https://www.stata.com/bookstore/graphics-reference-manual/>
Taller de Stata
Clase 10 – Gráficos 1 - Clasificación
Miguel Garzón Ramírez, Cristhian Acosta Pardo
Facultad de Economía, Universidad de los Andes