

---

# STATA 9.0



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

**Autores:**

**CATHERINE CRUZ PINZON**

**Director Unidad Informática:** Henry Martínez Sarmiento

**Tutor Investigación:** Álvaro Enrique Palacios

**Coordinadores:** María Alejandra Enríquez  
Leydi Diana Rincón

**Coordinador Servicios Web:** Miguel Ibañez

**Analista de Infraestructura  
y Comunicaciones:** Adelaida Amaya

**Analista de Sistemas de  
Información:** Álvaro Enrique Palacios Villamil

**Líder de Gestión de  
Recurso Humano:** Islena del Pilar Gonzalez

**UNIVERSIDAD NACIONAL COLOMBIA  
FACULTAD DE CIENCIAS ECONÓMICAS  
UNIDAD DE INFORMÁTICA Y COMUNICACIONES  
BOGOTÁ D.C.  
MARZO 2006**

# STATA 9.0

**Director Unidad Informática:** Henry Martínez Sarmiento

**Tutor Investigación:** María Alejandra Enríquez

## **Auxiliares de Investigación:**

ANDRES MAURICIO SALAMANCA	GUILLERMO ALBERTO ARIZA
BRAYAN RICARDO ROJAS	HENRY ALEXANDER RINCON
CAMILO ERNESTO LOPEZ	HOOVER QUITIAN
CAMILO IBAÑEZ	JORGE ELIECER ROJAS
CARLOS HERNAN PORRAS	JUAN FELIPE RINCON
CAROLINA RUBIANO OCHOA	LEIDY CAROLINA RINCÓN
CATHERINE CRUZ	LEIDY VIVIANA AVILÉS
CLAUDIA PATRICIA TOVAR	LUIS ALFONSO NIETO
CRISTIAN GERARDO GIL	LUZ KARINA RAMOS
CRISTIAN JAVIER PEÑALOZA	MILLER GIOVANNY FRANCO
DANIEL ENRIQUE QUINTERO	SANDRA LILIANA BARRIOS
DANIEL ERNESTO CABEZAS	SANDRA MILENA GOMEZ
DIANA ESPERANZA OROZCO	SANDRA PAOLA RAMIREZ
DIANA KATHERINE SANCHEZ	SERGIO ORJUELA RUIZ
DIANA PATRICIA TELLEZ	
DIEGO FELIPE CORTÉS	
EDSSON DIRCEU RODRÍGUEZ	
EDWIN MONTAÑO	

Este trabajo es resultado del esfuerzo de todo el equipo perteneciente a la Unidad de Informática.

Esta obra está bajo una licencia Reconocimiento-No comercial 2.5 Colombia de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by/2.5/co/> o envíe una carta a Creative Commons, 171 Second Street, Suite 30 San Francisco, California 94105, USA

**UNIVERSIDAD NACIONAL COLOMBIA  
FACULTAD DE CIENCIAS ECONÓMICAS  
UNIDAD DE INFORMÁTICA Y COMUNICACIONES  
BOGOTÁ D.C.  
MARZO 2006**

**TABLA DE CONTENIDO**

TABLA DE CONTENIDO.....3  
I. INTRODUCCIÓN.....4

# 1. INTRODUCCIÓN

En el desarrollo de los ejercicios econométricos es de vital importancia utilizar un software que brinde las herramientas para realizar un análisis completo de los datos. Stata 9.0 es una herramienta que permite hacer este tipo de análisis brindando mucha más capacidad que otros software. Se pretende así, a través de este manual revisar las herramientas básicas que este posee las cuales fueron exploradas con anterioridad en el primer manual, con el objetivo de generar a través de estas bases la revisión de un manejo un tanto más avanzado aplacándolas al análisis de series de tiempo básico. Se hará la revisión de dos modelos de regresión simple como primera aproximación y luego uno de series de tiempo.

En la unidad de informático ya se había desarrollado un manual básico para manejar Stata en una versión anterior, Stata 8.0, este manual se basa en una versión del software más avanzada, la siguiente, Stata 9.0. Sin embargo entre una versión y otra, las rutas para acceder a las diversas operaciones que el programa permite desarrollar no varían mucho en términos generales.

## 3.1 Marco teórico

Para hacer una primera aproximación de la aplicación de las herramientas básicas del software, se escogió un modelo lineal sencillo el cual requiere regresión lineal simple. El modelo ejemplo fue extraído de un curso de econometría<sup>1</sup>:

“El siguiente conjunto de datos era tomado sobre grupos de trabajadoras de Inglaterra y Galés en el período de 1970-72. Cada grupo está formado por trabajadores de la misma profesión (médicos, decoradores, trabajadores textiles,...etc.), en cada uno de los veinticinco grupos muestrados se han observado dos variables: *el índice de estandarizado de consumo de cigarrillos* (variable regresora,  $x$ ) y *el índice de muertes por cáncer de pulmón* (variable dependiente,  $y$ ). Se desea estudiar la relación entre estas dos variables. ”

$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$
77	84	102	88	133	146	116	155	112	96	91	85
137	116	91	104	115	128	102	101	113	144	100	120
117	123	104	129	105	115	111	118	110	139	76	60
94	128	107	86	87	79	93	113	125	113	66	51
88	104										

<sup>1</sup> Extraído de [www.udc.es](http://www.udc.es)

Se ha realizado un ajuste paramétrico lineal, obteniendo  $R^2 = 0.513$ , aplicando el siguiente modelo:

$$Y = -2.855 + 1.088x$$

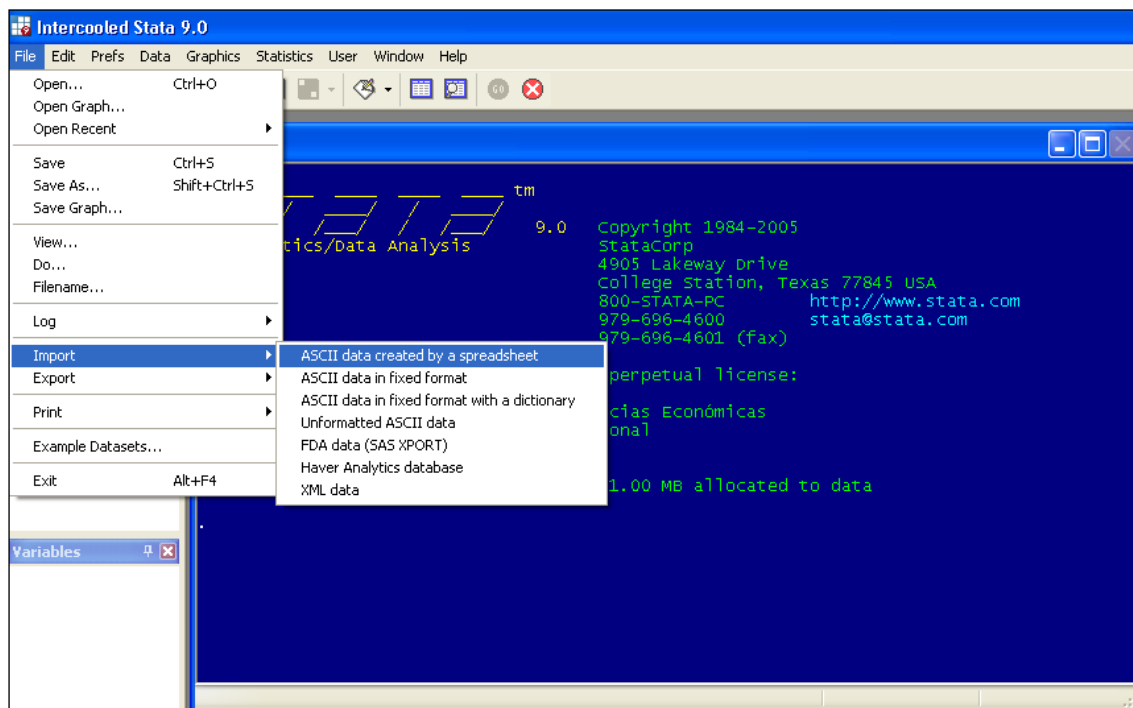
Se pretende revisar el análisis descriptivo tanto del modelo como de las variables y realizar pruebas sobre el mismo.

## 3.2 Revisión herramientas básicas

### 3.2.1 Importación de datos

Stata 9.0 permite importar datos desde diversos tipos de archivos, entre estos se encuentran los de formato .txt, .xls, entre otros. El archivo del ejemplo a desarrollar es de formato .xls, este archivo es reconocido desde Stata 9.0 y desde otro software como un archivo que separa con comas, siendo más explícitos, en formato .csv. Por tanto, para que el archivo sea reconocido por Stata 9.0, antes de importar el archivo de Excel, es necesario que este sea guardado directamente desde el programa en este formato, formato csv.

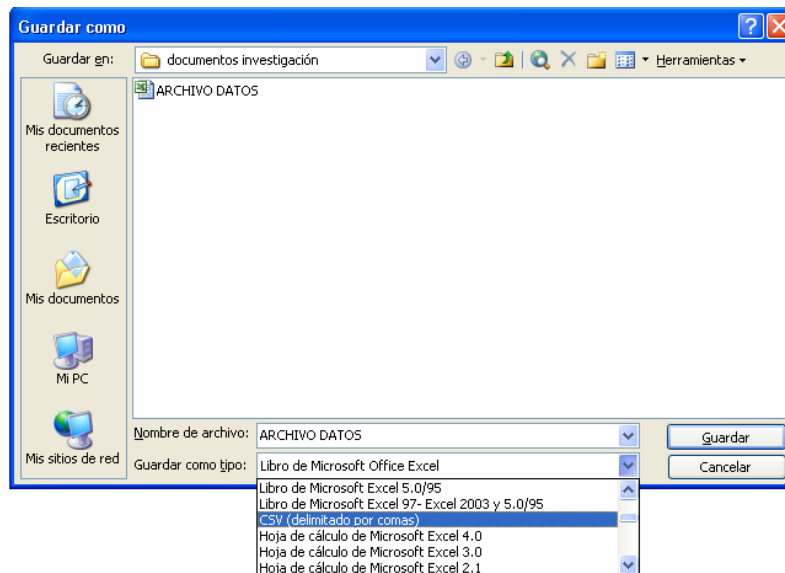
Las opciones a escoger para realizar la importación se ilustran en la Gráfica 1.



GRAFICA 1. Importación de datos

UNIVERSIDAD NACIONAL COLOMBIA  
FACULTAD DE CIENCIAS ECONÓMICAS  
UNIDAD DE INFORMÁTICA Y COMUNICACIONES

Para importar archivos con extensión .xls es necesario seleccionar las opciones ilustradas en la gráfica, sin embargo es necesario que el archivo sea guardado desde Excel con formato .csv, procedimiento ilustrado en la siguiente Gráfica.

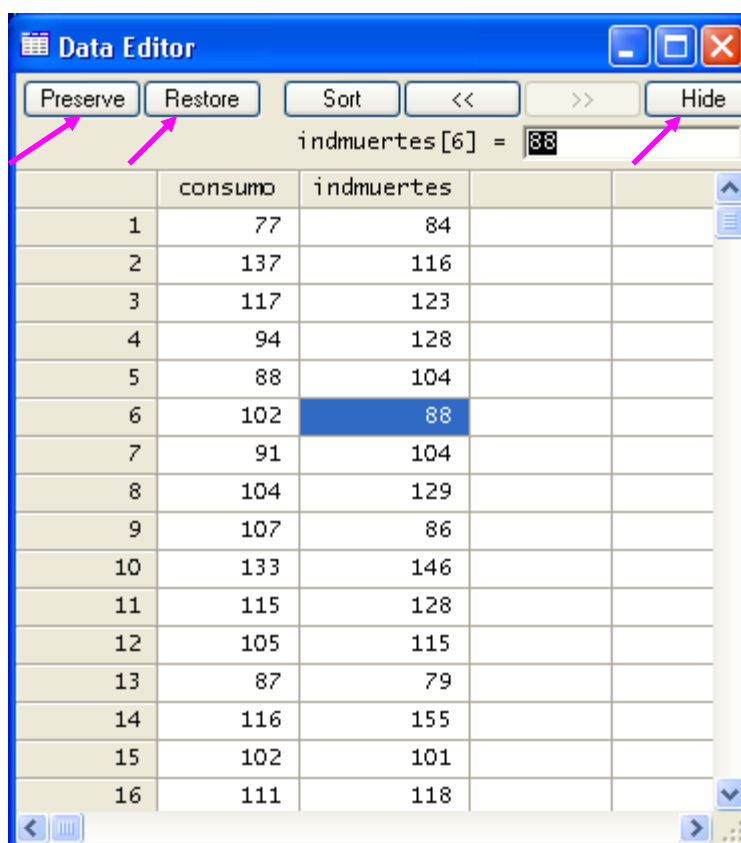


**GRAFICA 2. Procedimiento para guardar los archivos de extensión .xls**

Para mayor información de cómo importar datos desde archivos de otro tipo revisar el manual de introducción.

### 3.2.2 Introducción de datos

Cuando se introducen los datos en el **Data Editor** de manera manual, se visualizan tal como se aprecian en la Gráfica 3.



GRAFICA 3. Data Editor (Introducción de datos)

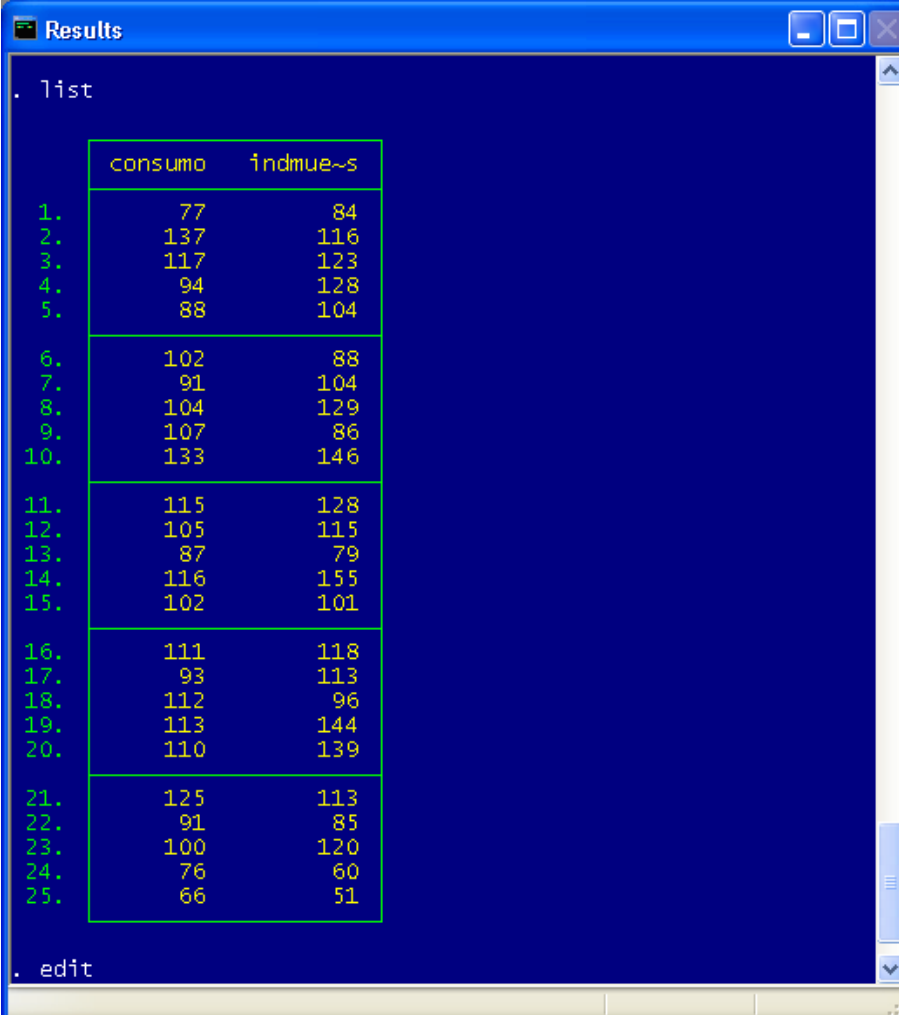
En el editor de datos se encuentran las siguientes opciones las cuales permiten manipular la base de datos, estas se encuentran señaladas en la Gráfica 1:

- **Sort:** organiza los datos de la variable en orden ascendente en función de la variable seleccionada.
- **Preserve:** permite guardar las observaciones de la variable en el archivo, una vez seleccionado las variables quedan modificadas según los cambios hechos en el archivo.
- **Hide:** elimina la serie de datos seleccionada.
- **Restore:** la base de datos queda como se guardo la última vez antes de efectuar cualquier cambio.

Al dar clic derecho en alguna de las celdas del data editor, se puede encontrar la opción **Font**, en donde se puede cambiar la fuente del título de la serie y el formato. Otra opción es **Variable**, donde se puede modificar el Label de la variable y su nombre en el Data Editor.

### 3.2.3 Descripción de las variables

Al escribir en la zona de comandos el comando ***list*** (en minúscula), al ser ejecutado este oprimiendo **Intro**, nos deja apreciar las variables del modelo de la siguiente forma:



```
. list
```

	consumo	índmue~s
1.	77	84
2.	137	116
3.	117	123
4.	94	128
5.	88	104
6.	102	88
7.	91	104
8.	104	129
9.	107	86
10.	133	146
11.	115	128
12.	105	115
13.	87	79
14.	116	155
15.	102	101
16.	111	118
17.	93	113
18.	112	96
19.	113	144
20.	110	139
21.	125	113
22.	91	85
23.	100	120
24.	76	60
25.	66	51

```
. edit
```

GRAFICA 4. Comando "List"

Como se puede apreciar, la ventana de resultados mostrará el comando introducido al ejecutarse este.

El comando ***list*** también permite realizar una especie de filtro, es decir, si por ejemplo se requiere analizar solamente un número determinado de observaciones que componen a una variable, en el cuadro de comandos y aplicándolo al ejemplo se escribe la siguiente secuencia de comandos:

UNIVERSIDAD NACIONAL COLOMBIA  
FACULTAD DE CIENCIAS ECONÓMICAS  
UNIDAD DE INFORMÁTICA Y COMUNICACIONES

*"list consumo in 1/15"*

A lo cual la hoja de resultados mostrará lo que se aprecia en la Gráfica 4:

```

Results
. count
25
. list consumo in 1/15

+-----+
| consumo |
+-----+
1.      77 |
2.     137 |
3.     117 |
4.      94 |
5.      88 |
+-----+
6.     102 |
7.      91 |
8.     104 |
9.     107 |
10.    133 |
+-----+
11.    115 |
12.    105 |
13.     87 |
14.    116 |
15.    102 |
+-----+

```

GRAFICA 5. Comando

Se visualizan las 15 primeras observaciones de la variable seleccionada.

En la Gráfica 4 también se puede observar el resultado de ejecutar el comando **"count"**. Este comando contabiliza los datos que conforman la base de datos, es decir, los que están contenidas en el archivo. Al digitarlo en el cuadro de comandos y ejecutarlo se podrá visualizar el número de datos, tal y como se aprecia en la Gráfica 4 útil para conocer la muestra en su totalidad.

Con el comando **"tabstat"** aparece la media de las variables, pero debe ser escrito junto con los nombres asignados a las variables dentro del software de la siguiente forma:

UNIVERSIDAD NACIONAL COLOMBIA  
 FACULTAD DE CIENCIAS ECONÓMICAS  
 UNIDAD DE INFORMÁTICA Y COMUNICACIONES

*“tabstat (nombre variable1) (nombrevariable2)”*

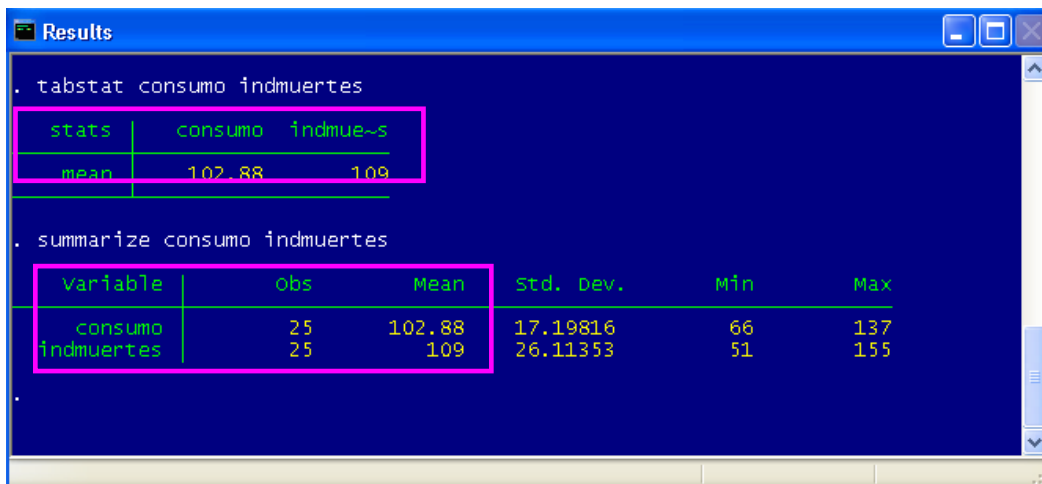
Es importante que se separe tanto el comando como las variables por medio de un espacio. Para el ejemplo escogido los comandos a introducir serían:

*“tabstat consumo indmue-tes”*

Sin embargo con el comando **“summarize”** aparecen muchas más estadísticas que hacen más completo el análisis de las variables. La forma de escribir el comando en la ventana para el ejemplo escogido sería la siguiente:

*“summarize consumo indmue-tes”*

Los resultados se pueden visualizar en la Gráfica 3.



```

. tabstat consumo indmue-tes

+-----+-----+
| stats | consumo | indmue-tes |
+-----+-----+
| mean  | 102.88  | 109        |
+-----+-----+

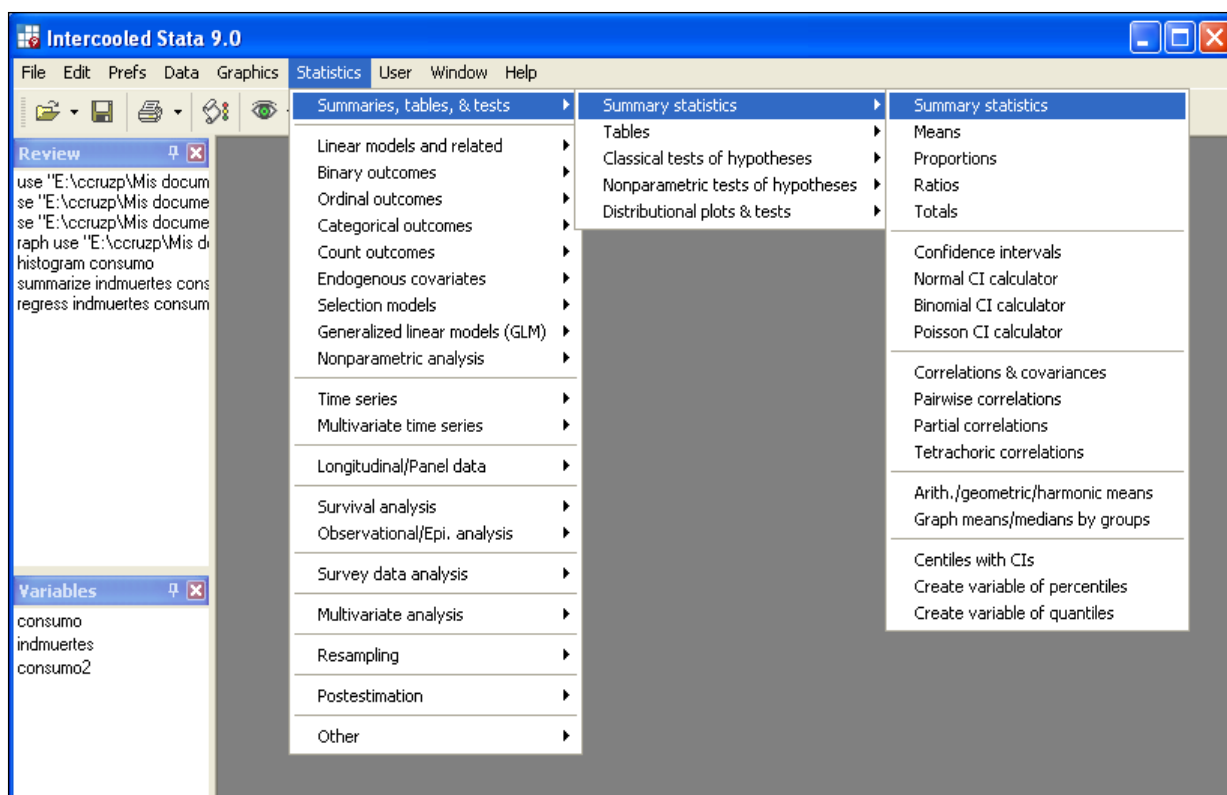
. summarize consumo indmue-tes

+-----+-----+-----+-----+-----+
| Variable | Obs | Mean | Std. Dev. | Min | Max |
+-----+-----+-----+-----+-----+
| consumo  | 25  | 102.88 | 17.19816  | 66  | 137 |
| indmue-tes | 25  | 109    | 26.11353  | 51  | 155 |
+-----+-----+-----+-----+-----+

```

GRAFICA 6. Comandos “tabstat” y “sumaries”

Para observar las estadísticas descriptivas de las variables, también se puede realizar a través del menú **Statistics → Summaries, tables & test → Summary statistics** y de nuevo esta opción. Como se aprecia en la gráfica:



GRAFICA 7. Selección menú para estadísticas descriptivas

Agredando el comando **“detail”** a la secuencia de comandos antes utilizada, se puede observar con mayor detalle las estadísticas de las variables. La secuencia de comandos introducida para generar esta descripción y el resultado que genera se aprecia en la ventana de resultados. Ver Gráfica 8.

```

Results
r(199);
. summarize consumo indmueertes, detail
      indice del consumo de cigarrillo
-----
      Percentiles      Smallest
  1%                66          66
  5%                76          76
 10%                77          77      Obs                25
 25%                91          87      Sum of Wgt.        25

 50%                104
                                Mean                102.88
                                Std. Dev.          17.19816
 75%                113          Largest
 90%                125          117
 95%                133          125      Variance           295.7767
 99%                137          133      Skewness           -.0983426
                                Kurtosis           2.739708

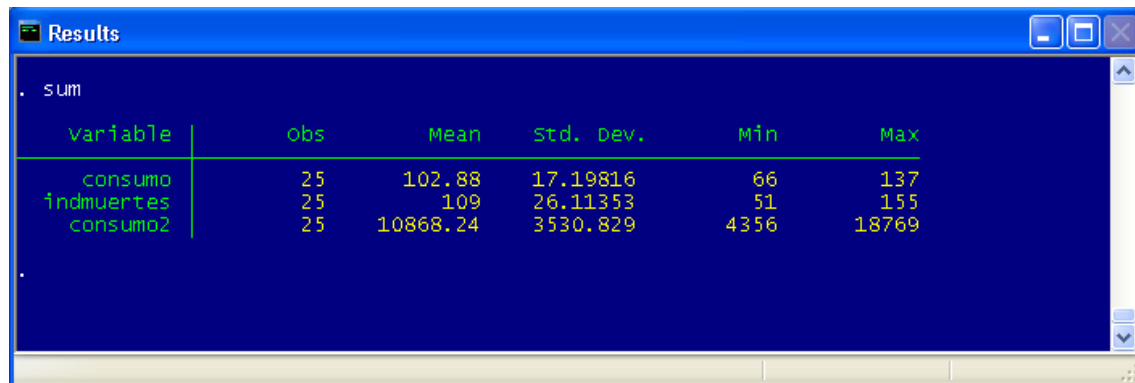
      indice de muertes por cancer de pulmon
-----
      Percentiles      Smallest
  1%                51          51
  5%                60          60
 10%                79          79      Obs                25
 25%                88          84      Sum of Wgt.        25

 50%                113
                                Mean                109
                                Std. Dev.          26.11353
 75%                128          Largest
 90%                144          139
 95%                146          144      Variance           681.9167
 99%                155          146      Skewness           -.3546932
                                Kurtosis           2.638516

```

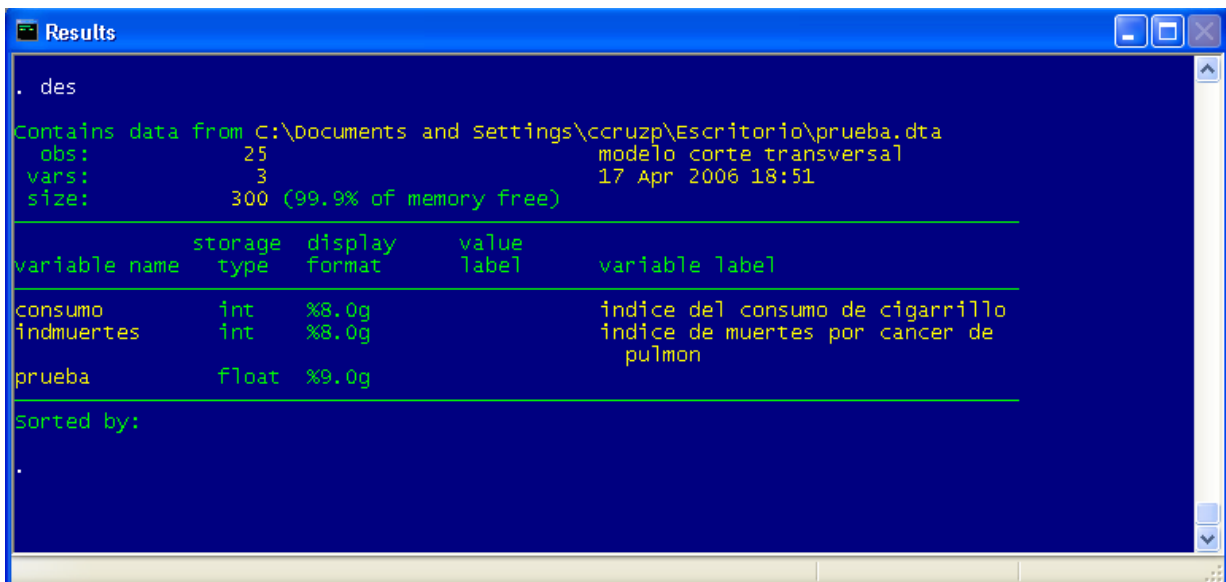
GRAFICA 8. Comando “detail”

El comando “**sum**” es la abreviatura de este comando, es otra forma de acceder a las estadísticas descriptivas.



variable	Obs	Mean	Std. Dev.	Min	Max
consumo	25	102.88	17.19816	66	137
indmuertes	25	109	26.11353	51	155
consumo2	25	10868.24	3530.829	4356	18769

Con el comando “**des**” se pueden apreciar los datos generales descriptivos de las variables y del archivo, como son el número de observaciones, el espacio que ocupan las variables en la memoria, la ubicación del archivo, la etiqueta completa de cada variable entre otros, tal como se aprecia en la Gráfica 8.



```

. des
Contains data from C:\Documents and Settings\ccruzp\Escritorio\prueba.dta
  obs:      25
  vars:      3
  size:     300 (99.9% of memory free)
-----
variable name  storage  display  value  variable label
              type   format   label
consumo        int     %8.0g
indmuertes     int     %8.0g
prueba         float  %9.0g
-----
Sorted by:

```

variable name	storage type	display format	value label	variable label
consumo	int	%8.0g		índice del consumo de cigarrillo
indmuertes	int	%8.0g		índice de muertes por cancer de pulmon
prueba	float	%9.0g		

GRAFICA 9. Comando “des”

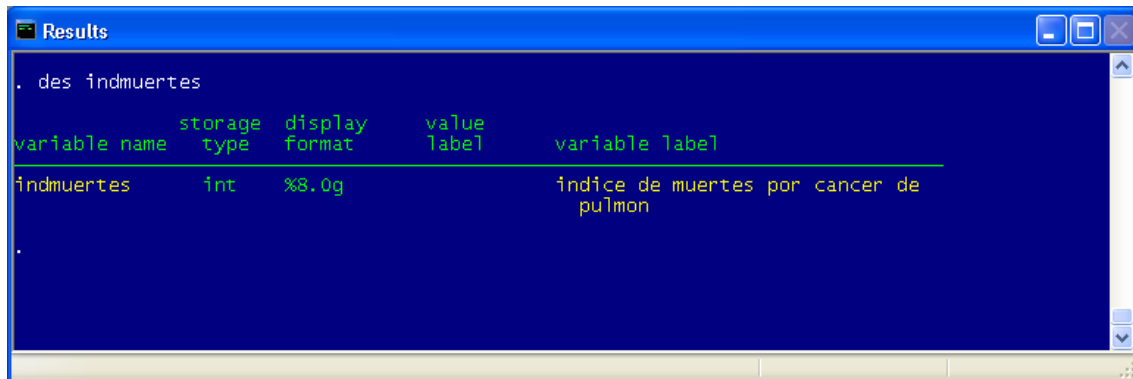
Si solo se requiere la descripción de una de las variables de la base de datos los comandos para realizarla son:



```
Command
des indmuertes
```

GRAFICA 10. Secuencia comando “des”

Con los cuales en la ventana de resultados aparecerá la descripción de la variable seleccionada:



```
Results
. des indmuerter

variable name      storage   display   value   variable label
type              format
-----
indmuerter         int       %8.0g          indice de muertes por cancer de
                    pulmon
```

GRAFICA 11. Comando “des”

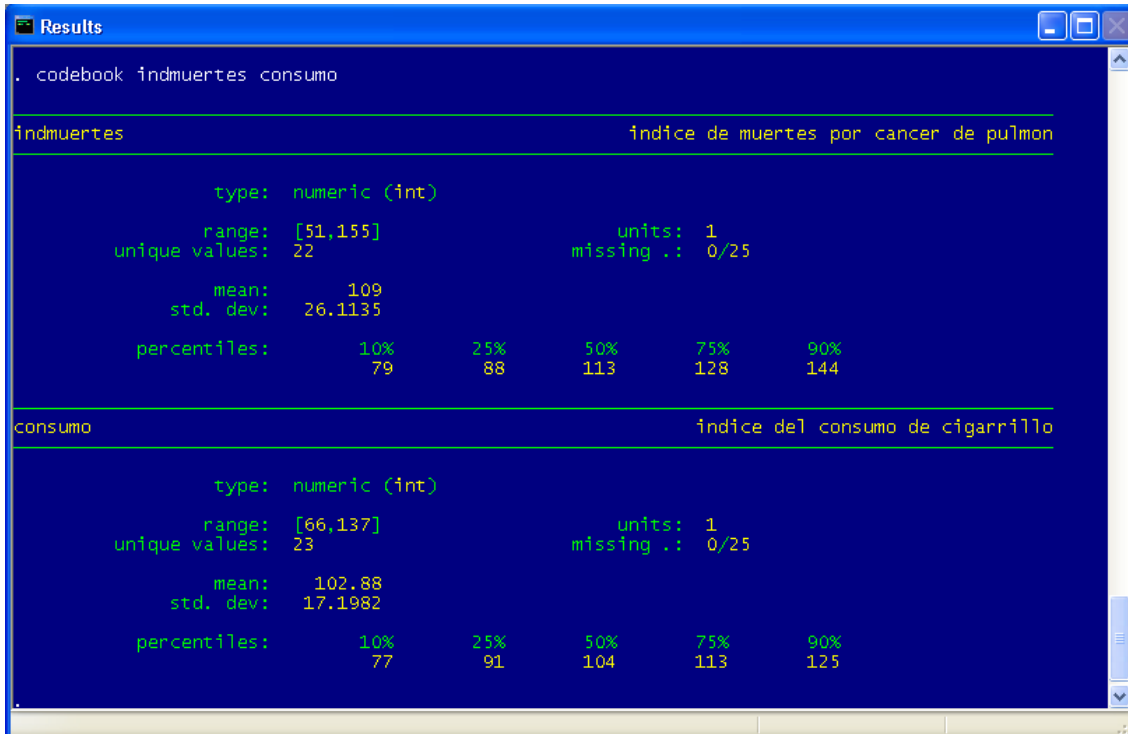
El comando “**codebook**” permite visualizar mayores detalles de las variables:



```
Command
codebook indmuerter consumo
```

GRAFICA 12. Secuencia comando “codebook”

Al ejecutar el comando, la ventana de resultados arrojará las estadísticas que se pueden visualizar en la Gráfica 10.



```

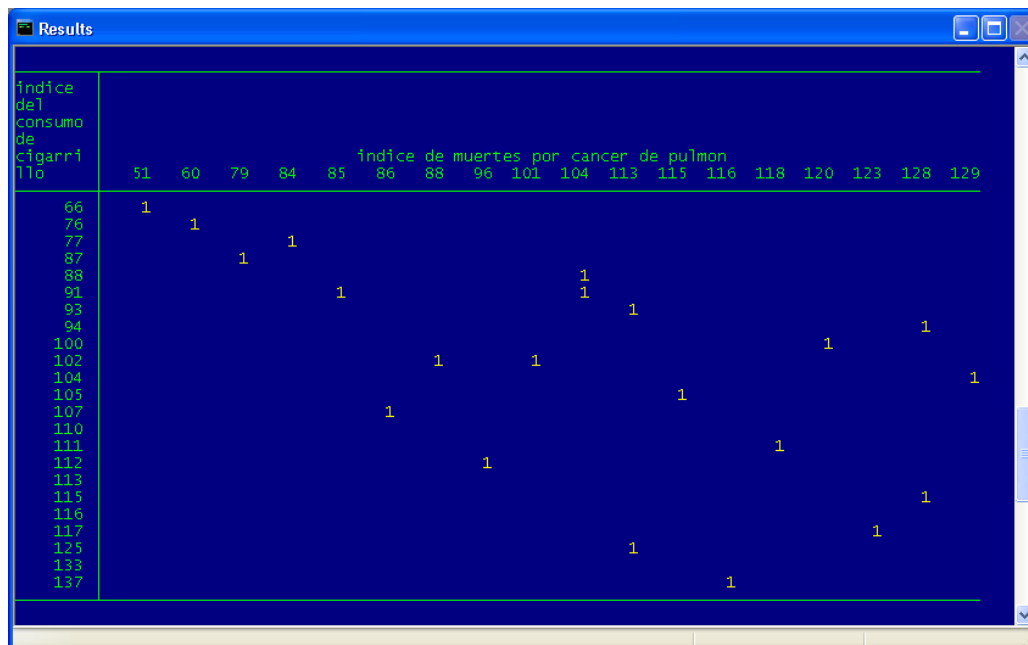
. codebook indmueertes consumo

-----
indmueertes                               índice de muertes por cancer de pulmon
-----
      type: numeric (int)
      range: [51,155]
unique values: 22                          units: 1
                                          missing .: 0/25
      mean:      109
      std. dev:  26.1135
      percentiles: 10%    25%    50%    75%    90%
                   79     88     113   128   144
-----
consumo                                    índice del consumo de cigarrillo
-----
      type: numeric (int)
      range: [66,137]
unique values: 23                          units: 1
                                          missing .: 0/25
      mean:      102.88
      std. dev:  17.1982
      percentiles: 10%    25%    50%    75%    90%
                   77     91    104    113   125

```

GRAFICA 13. Comando “codebook”

El comando **“table”** genera una tabla de frecuencias de las variables de la base de datos, enfrenta la variable endógena con la exógena.



```

-----
indice del consumo de cigarrillo
-----
      51  60  79  84  85  86  88  96 101 104 113 115 116 118 120 123 128 129
-----
66      1
76      1
77      1
87      1
88      1
91      1
93      1
94      1
100     1
102     1
104     1
105     1
107     1
110     1
111     1
112     1
113     1
115     1
116     1
117     1
125     1
133     1
137     1

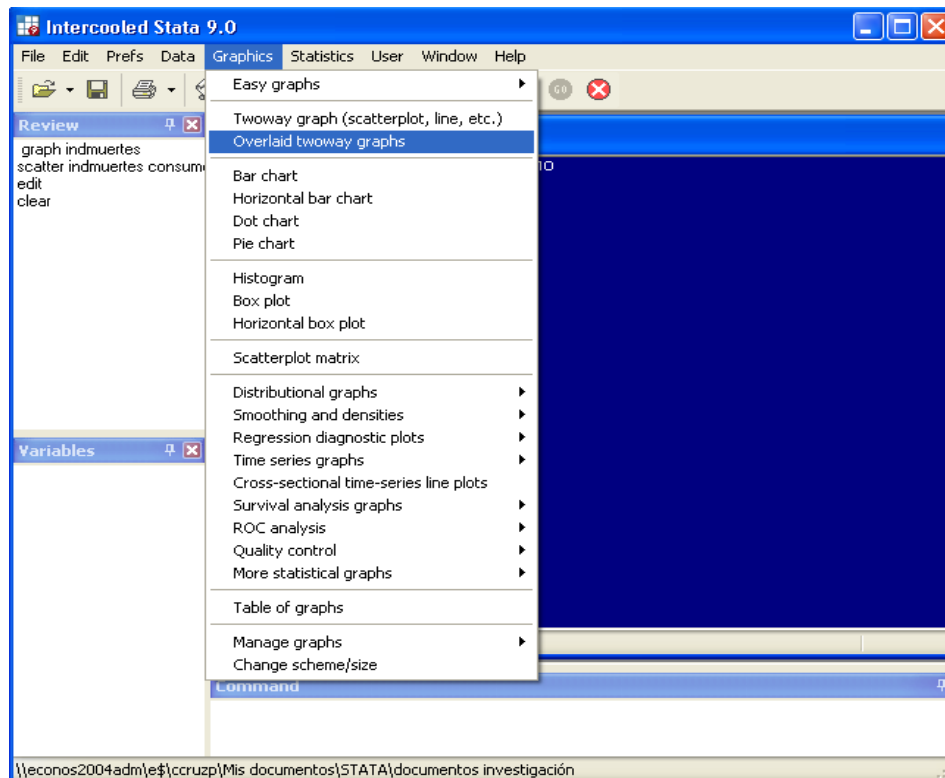
```

GRAFICA 14. Comando “table”

### 3.2.3.1 Gráficos

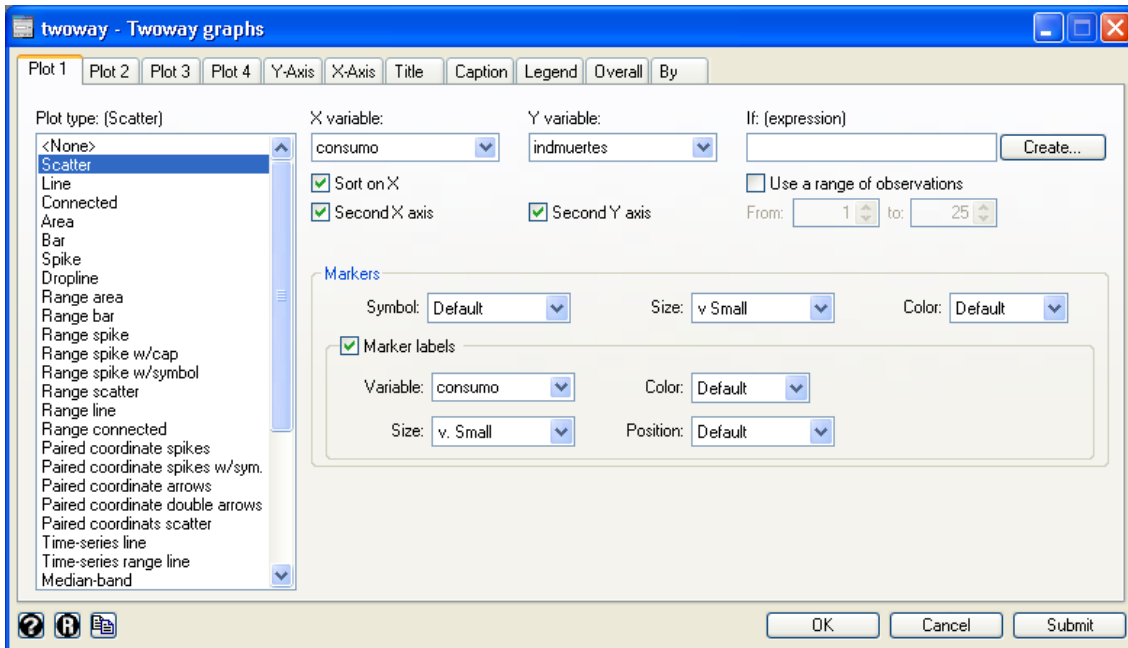
Eligiendo diversas opciones que presenta el menú del software se pueden elaborar distintos tipos de gráficos.

Otro de los gráficos que se pueden elaborar son los que se generan al seleccionar **Graphics** → **Overlaid twoway graph** como se puede apreciar en la siguiente gráfica:



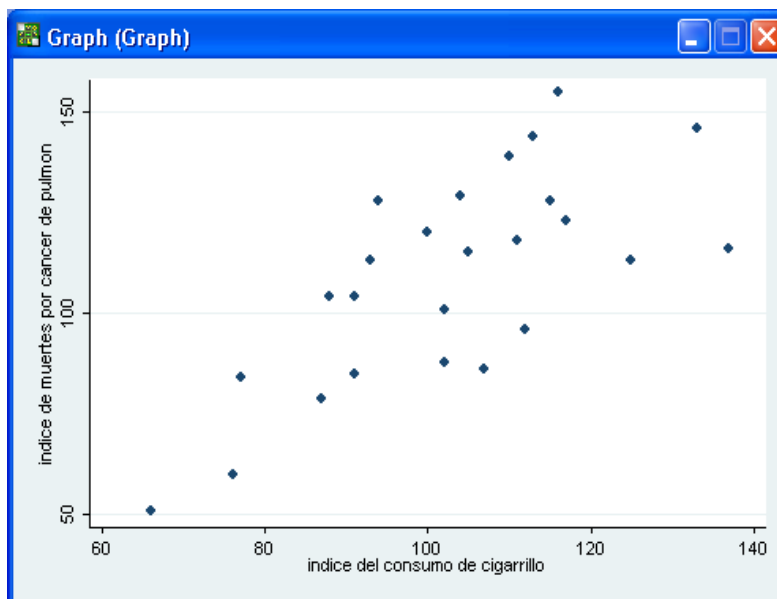
GRAFICA 15. Selección pasos gráfico sencillo

Aparecerán las siguientes opciones:



GRAFICA 16. Gráfico sencillo

Seleccionando la primera y seleccionando las especificaciones que se muestran en la Gráfica 10 se pueden generar desde gráficos sencillos que representen la tendencia de los datos a través de gráficos de puntos en el plano cartesiano, como el que se aprecia a continuación para el modelo escogido:

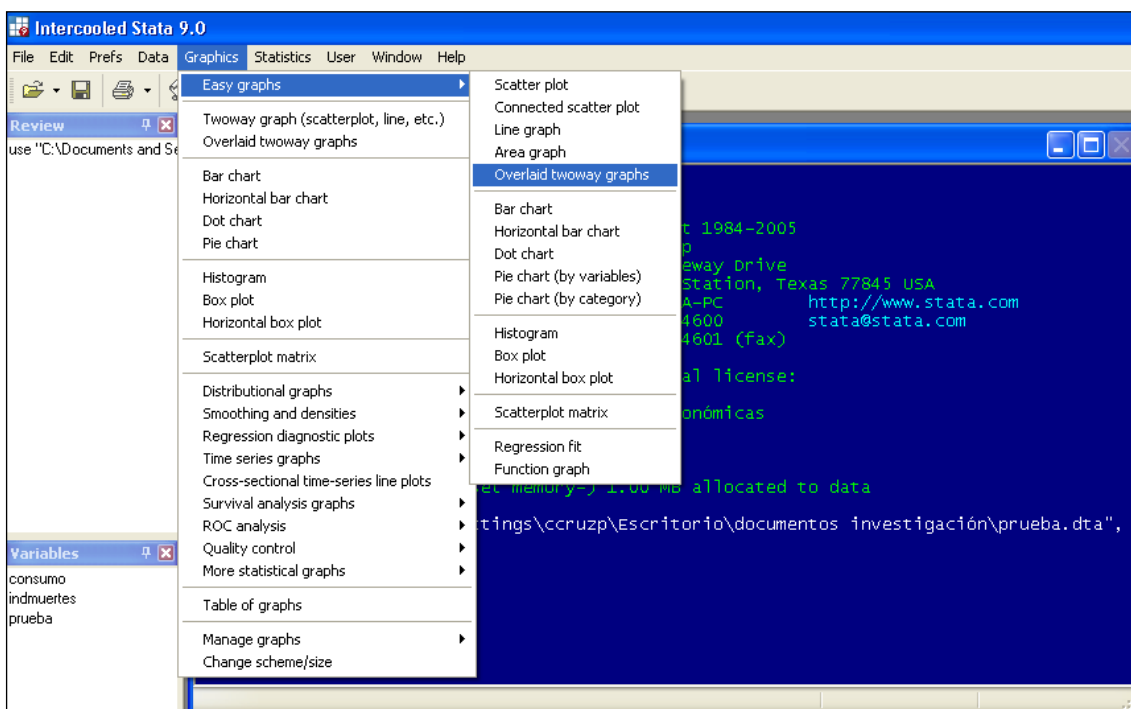


GRAFICA 17. Gráfico sencillo “tendencia de las observaciones”

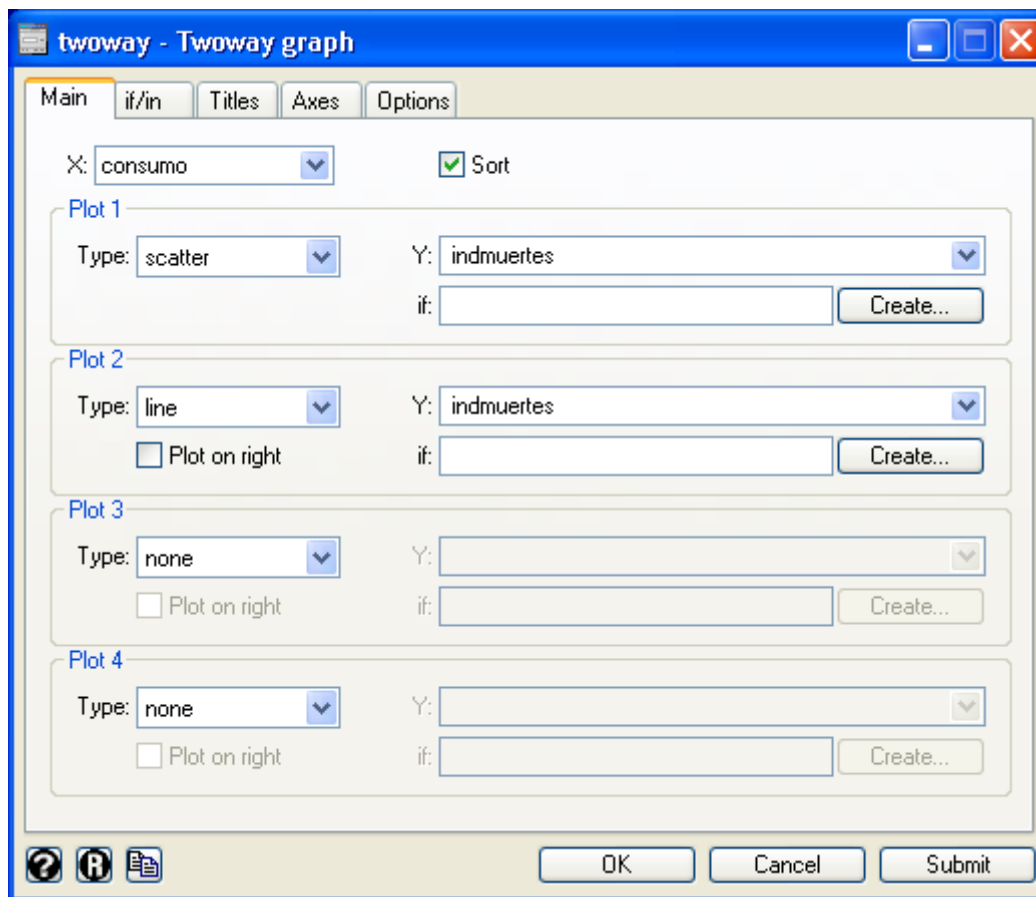
UNIVERSIDAD NACIONAL COLOMBIA  
FACULTAD DE CIENCIAS ECONÓMICAS  
UNIDAD DE INFORMÁTICA Y COMUNICACIONES

Este gráfico sencillo permite hacer una primera aproximación acerca de la relación existente entre las variables y sobre la dispersión de la muestra. Aunque existe dispersión en los datos, se puede afirmar que la relación entre las variables es directamente proporcional, lo cual es bastante lógico desde la interpretación de las mismas. A mayor consumo de cigarrillo, se genera un mayor índice de muertes por cáncer de pulmón.

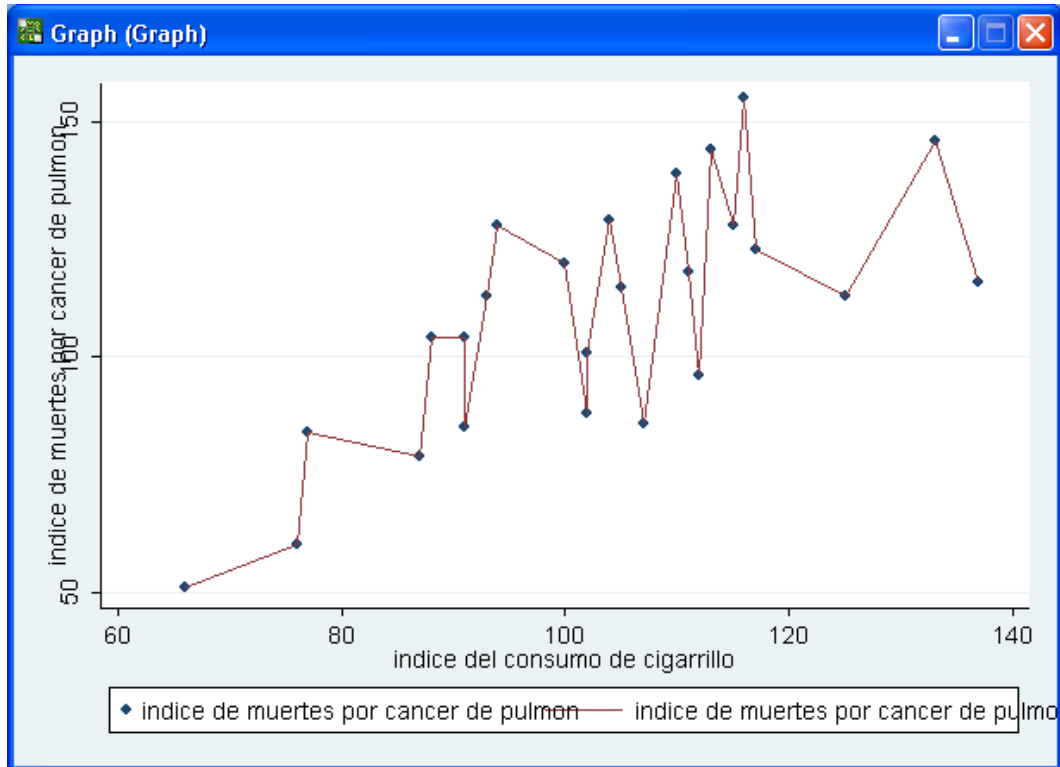
Para generar un gráfico un poco más elaborado que muestre de forma más clara la relación entre las variables del modelo se puede explorar las opciones del software y combinar dos tipos de gráficos que representen las variables. Al seleccionar las opciones ilustradas en las Gráficas 13 y 14, pero seleccionando además las opciones para combinar los tipos de gráfico, se pueden generar gráficos un poco más elaborados como los que se aprecian en la Gráfica 15.



**GRAFICA 18. Opciones gráficos compuestos**



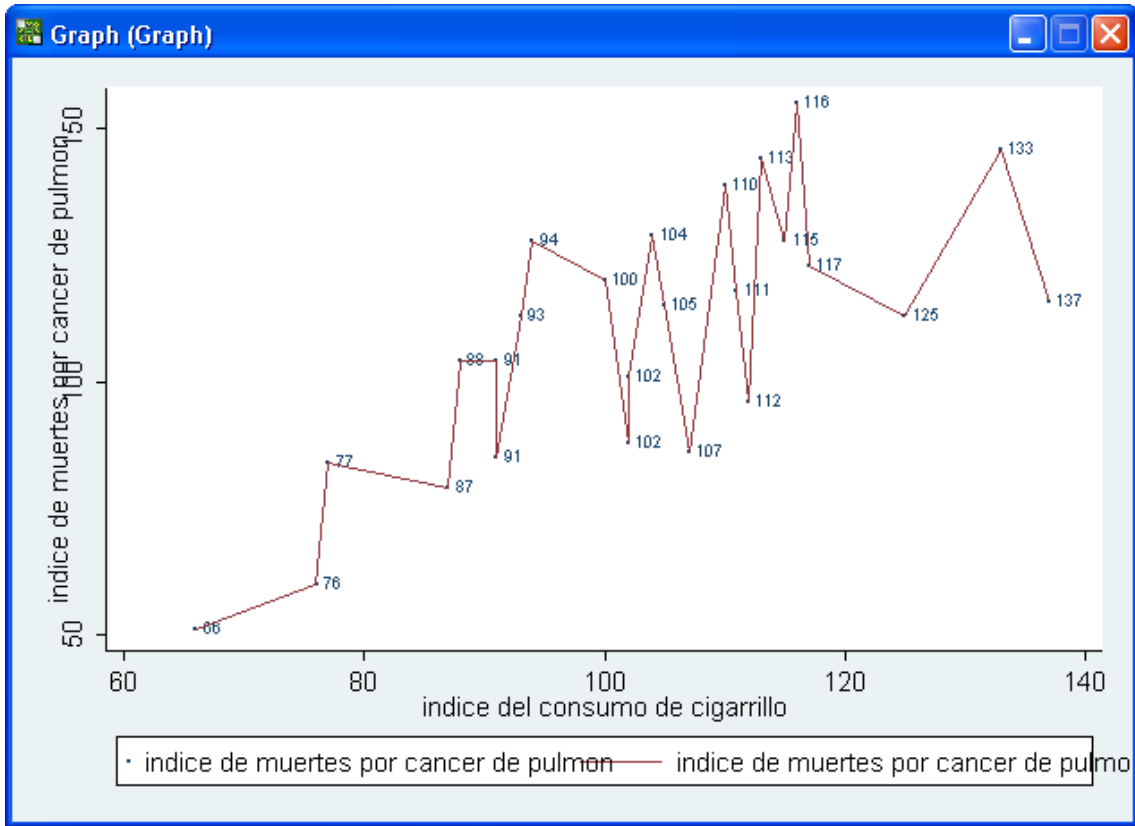
GRAFICA 19. Combinación de gráficos



**GRAFICA 20. Gráfico más elaborado "Tendencia de las observaciones"**

De esta forma desde el menú **Graphics** se puede observar de forma más clara que la relación entre las dos. Al revisar más a fondo la herramienta que ofrece el software se pueden elaborar gráficos con rótulos de valor igual que en Excel.

Y así se genera la gráfica que a su vez es mucho más completa como la que se observa en la Gráfica 15.



GRAFICA 21. Gráfica con rótulos

### 3.2.4 Generación de variables

Stata 9.0 genera variables a través del comando **“generate”**, que también se puede utilizar de forma abreviada, **“gen”**. Para ejecutar este comando es necesario acompañarlo del nombre que llevara la nueva variable a generar, el nombre de la variable generada es con el cual la reconocerá el software si posteriormente se desea realizar operaciones con ella.

### 3.2.5 Operaciones sencillas

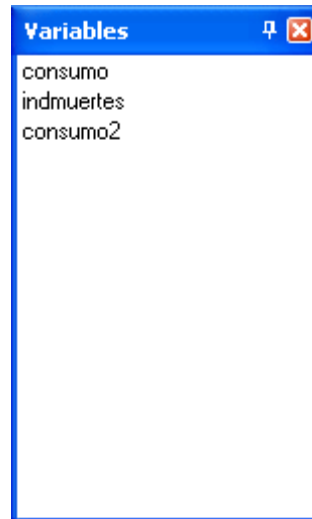
Si se desea realizar operaciones entre variables que estén contenidas en la base de datos y generar otras nuevas a través de estas operaciones, este es un ejemplo elaborado con las variables de la base de datos ejemplo:



```
Command
generate consumo2= consumo * consumo
```

GRAFICA 22. Operaciones básicas

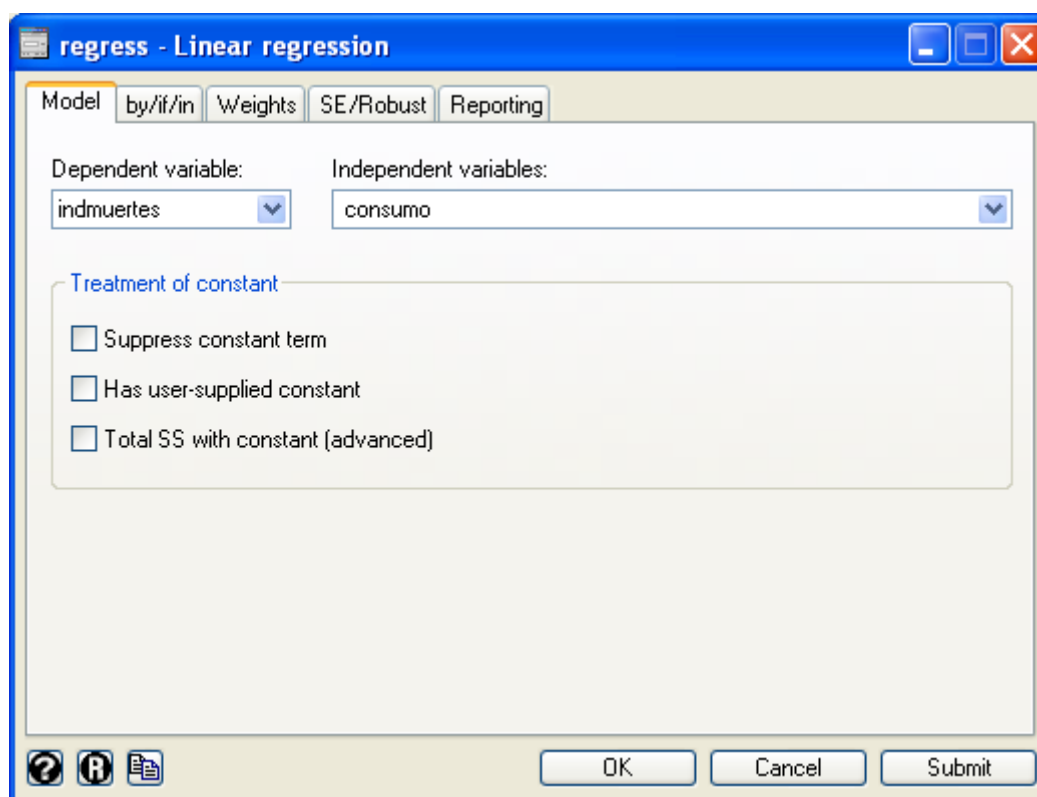
Al pulsar Enter, el comando se ejecuta la nueva variable será generada en el archivo y por tanto aparecerá en la ventana de variables, como se observa:



GRAFICA 23. Ventana de variables

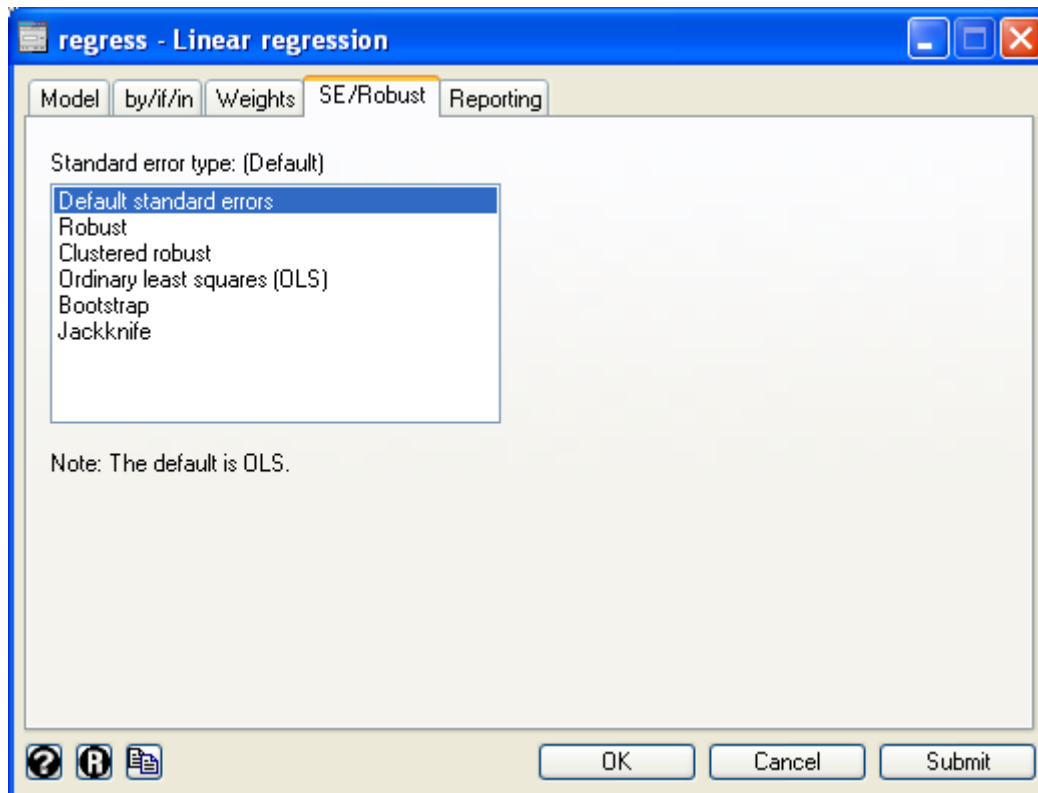
### 3.2.6 Regresión Lineal Simple:

**Stata 9.0** nos permite realizar regresiones fácilmente a través del menú **Statistics** → **Linear models and Related** → **Linear Regresión**. Si se deseara elaborar la regresión por medio de comandos se escribiría en la Ventana Command la secuencia que también se puede apreciar en esta gráfica, la secuencia es automáticamente ejecutada por el software en el momento de realizar las selecciones del menú ya mencionadas y por tanto aparecerá en la ventana de resultados y así mismo en la ventana **Review**.



GRAFICA 24. Cuadro regresión

Este cuadro de diálogo varía frente a la versión anterior, ya que en esta podíamos elegir el nivel de confianza al cual se deseaba realizar la estimación, para esta versión el software realiza la regresión con un nivel de confianza predeterminado, 95%. Sin embargo con parámetros como el p-value de los coeficientes o del F-estadísticos, el estudiante es quien decide con que nivel de confianza va a aplicar las pruebas, es decir a que nivel de significancia de los estadísticos rechaza las hipótesis nulas.



GRAFICA 25. Pestaña SE/ROBUST

En esta pestaña es donde Stata 9.0 permite escoger el método de estimación con el cual se va a elaborar la regresión. Por ejemplo, según las necesidades del estudiante, se puede escoger el método de Mínimos Cuadrados Ordinarios (**Ordinary Least Squares, OLS**) que es el más utilizado.

Así después de ejecutar el comando para elaborar la regresión o haber hecho las selecciones ya mencionadas se obtienen los siguientes resultados:

The screenshot shows the STATA Results window with the following content:

```
. regress indmuertes consumo
```

Source	SS	df	MS			
Model	8395.74904	1	8395.74904	Number of obs =	25	
Residual	7970.25096	23	346.53265	F( 1, 23) =	24.23	
Total	16366	24	681.916667	Prob > F =	0.0001	
				R-squared =	0.5130	
				Adj R-squared =	0.4918	
				Root MSE =	18.615	

indmuertes	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
consumo	1.087532	.2209452	4.92	0.000	.6304724	1.544592
_cons	-2.885319	23.03372	-0.13	0.901	-50.5342	44.76356

GRAFICA 26. Regresión Lineal Simple

Es importante revisar detenidamente la estructura de los comandos si se realiza por este medio:

*“regress indmuertes consumo”*

Como se puede observar, lo que primero se anota es el comando que permite generar la regresión (“regress”), luego separado por un espacio para que el software reconozca que es un comando y así mismo una variable se anota la variable endógena de la regresión (en este caso “indmuertes”, y por último una a una las variables exógenas separadas por espacios para que sean identificadas en el orden en que se desea aparezcan en la regresión.

Esta estructura aplica para la escritura de comandos en general, ya que primero se debe escribir el comando a ejecutar y luego el nombre de la variable a la cual se desea aplicar la ejecución del comando, todo separado por espacios.

Como se puede apreciar en la gráfica de resultados, la ecuación que determina la relación entre las variables que fue arrojada como resultado por el software fue la siguiente:

$$\text{Indmuertes} = -2.8853 + 1.0875 \text{ Consumo}$$

$$Y = -2.8853 + 1.0875$$

La cual efectivamente coincide con la estructura propuesta inicialmente. De lo cual se concluye que por un incremento de 1% en el consumo de cigarrillo, el índice de muertes por cáncer de pulmón se incrementa en 1.08% para la muestra

seleccionada. Y si el índice de consumo de cigarrillo fuese 0, el índice de muertes sería de -2.88% para la muestra seleccionada.

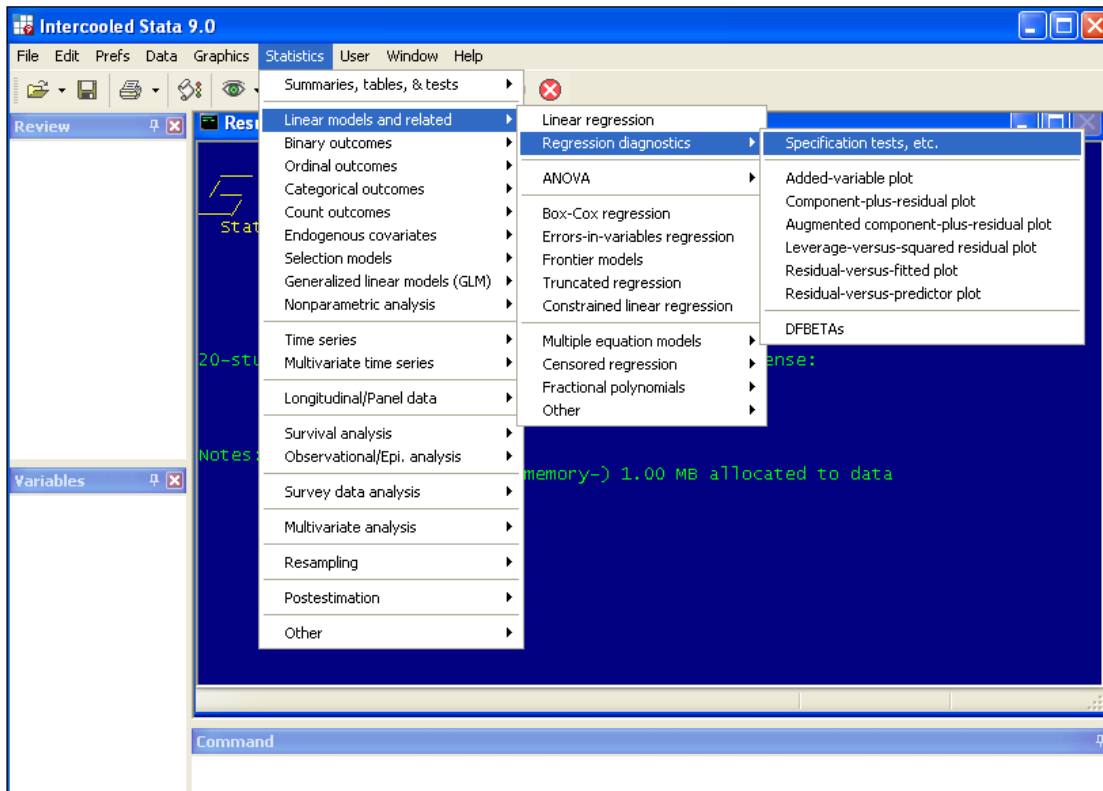
Además en el resultado también se pueden apreciar los límites inferior y superior de un intervalo de confianza del 95% para la variable exógena del modelo generado.

El menú **Graphics** → **Bar Chart**, permite graficar las estadísticas que se requieran de la variable que se elija. Es decir, si deseamos generar la gráfica de la media de la variable consumo, esta se puede elaborar por medio de la selección de estas opciones.

El menú **Graphics** → **Histogram** nos permite graficar la estadística de la variable que sea de nuestro interés analizar. Por ejemplo, es de interés generar un Histograma de los residuos al realizar un modelo econométrico.

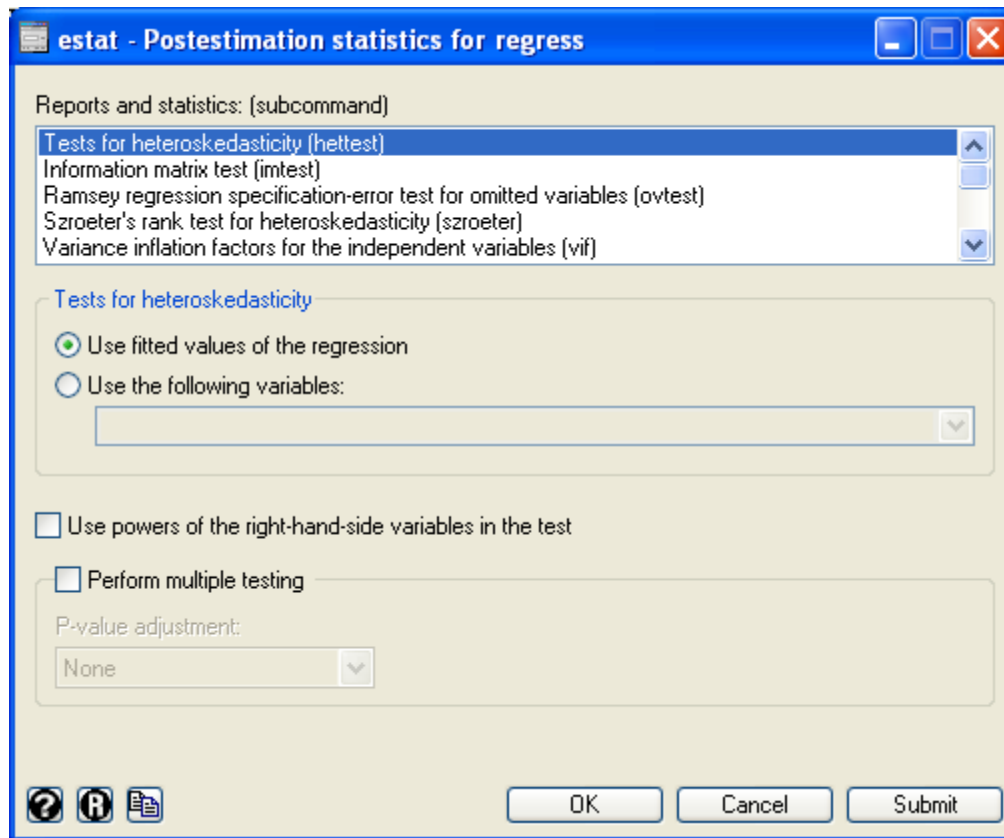
### 3.2.7 Pruebas de hipótesis

Al dirigimos al menú **Statistics** y seleccionar las opciones que se pueden apreciar en la Gráfica 25, se pueden elaborar distintos tipos de test al modelo.



GRAFICA 27

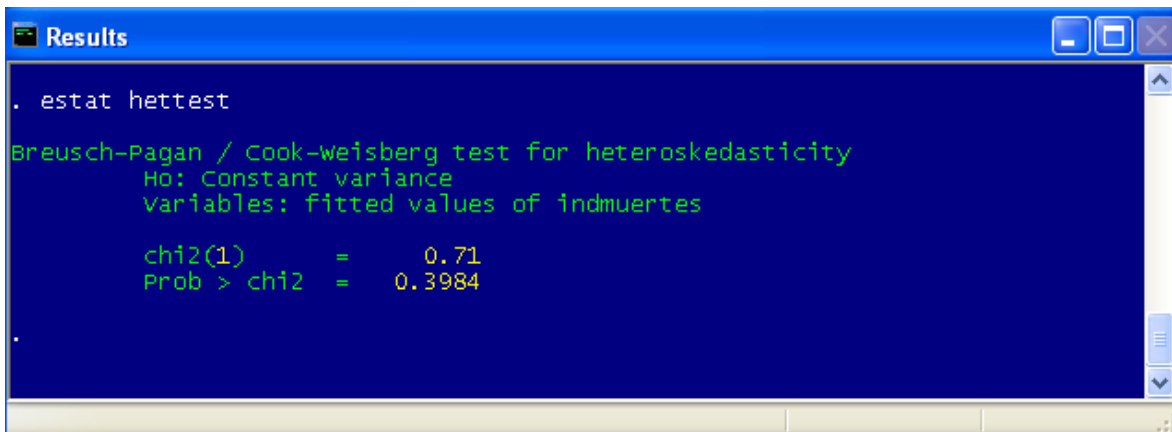
Las opciones de pruebas a realizar se aprecian en la Gráfica 28:



GRAFICA 28

Como se puede apreciar, entre las pruebas que se pueden realizar al modelo están las pruebas de heterocedasticidad y la matriz de varianzas y covarianzas con la cual es posible elaborar las pruebas de hipótesis individual a los coeficientes. Además, como se puede apreciar en la Gráfica 26, se puede realizar la prueba a una variable determinada o a los coeficientes estimados a través de la regresión por medio de la opción **Use fitted values of the regression**.

Es necesario haber generado la regresión para que este comando pueda ser ejecutado, ya que al seleccionar en el menú **File** → **Save As...** el archivo solo almacena las variables más no los resultados de las operaciones o comandos ejecutados a menos que estos sean guardados como un archivo Log.



```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
H0: Constant variance
variables: fitted values of indmuertes

chi2(1)      =      0.71
Prob > chi2   =      0.3984
```

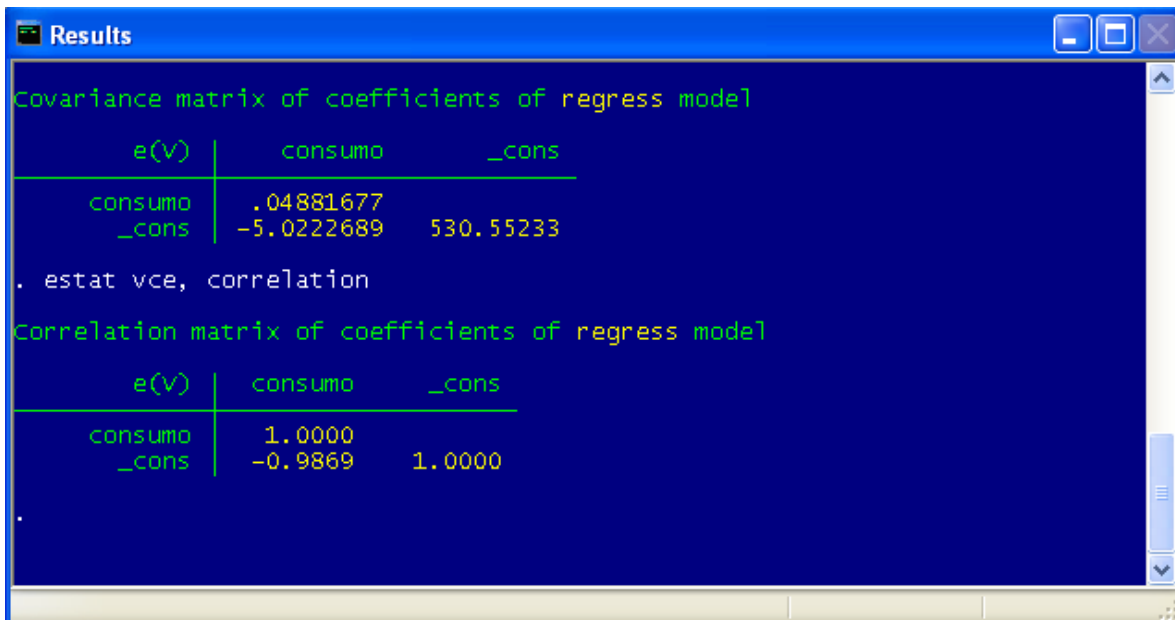
GRAFICA 29

Como se puede ver, los comandos a utilizar si la prueba se realizara por la ventana de comandos serían:

*“estat hettest”*

Los resultados que se aprecian nos permiten concluir que el modelo presenta Heterocedasticidad, ya que el p-value del estadístico Chi-cuadrado es mayor a 0.05 por lo que al 95% de confianza no se rechaza la hipótesis nula de Heterocedasticidad. Es decir, el modelo presenta varianzas que cambian a través de la muestra.

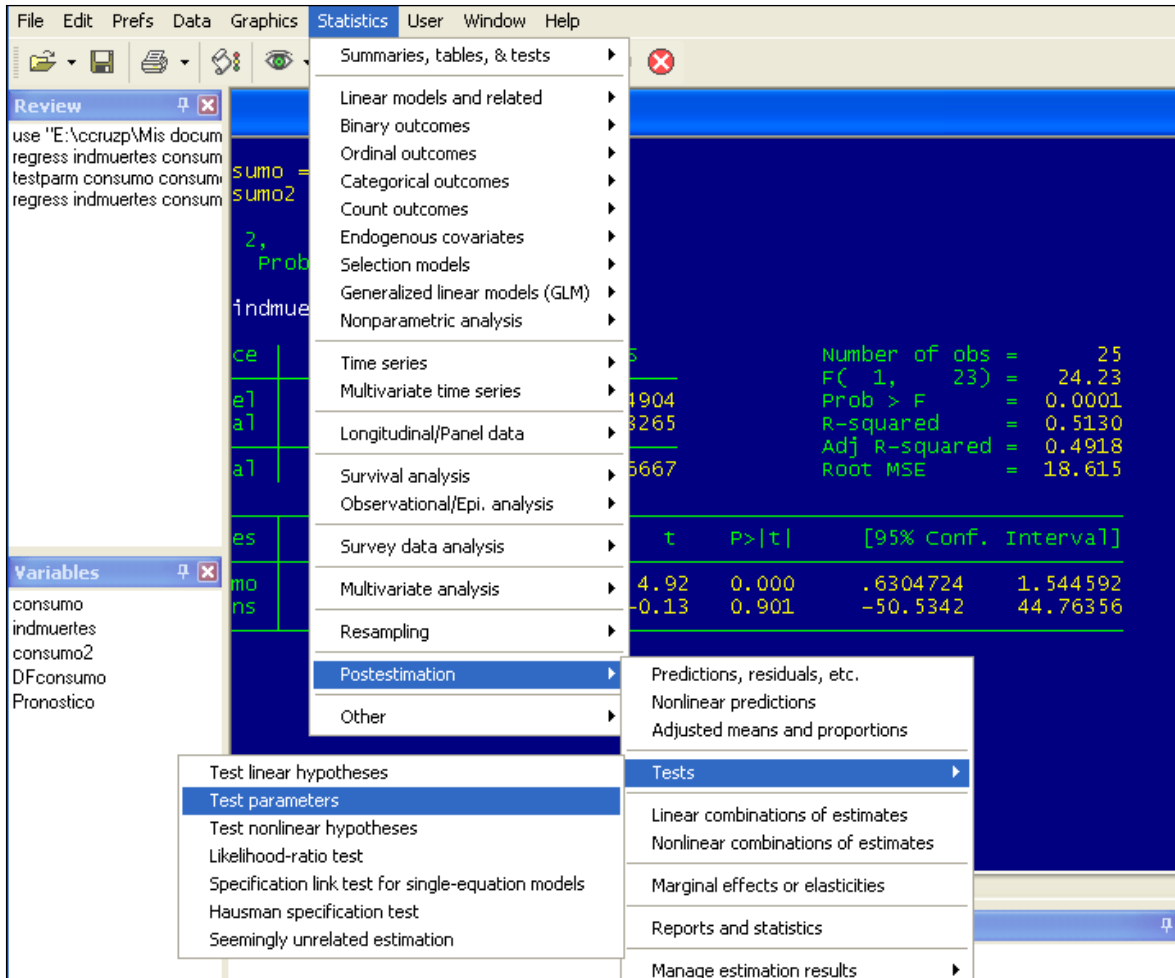
Al elegir la opción **Covariance matrix estimates**, el software nos permite además elaborar la matriz de correlaciones entre las variables, lo cual ayuda a determinar la existencia de multicolinealidad entre las variables del modelo. Para el ejemplo desarrollado no es muy interesante observar las correlaciones dado que solo existe una variable exógena en el modelo. Sin embargo, en la Gráfica 29 se presentan los resultados de ejecutar estas matrices.



GRAFICA 30

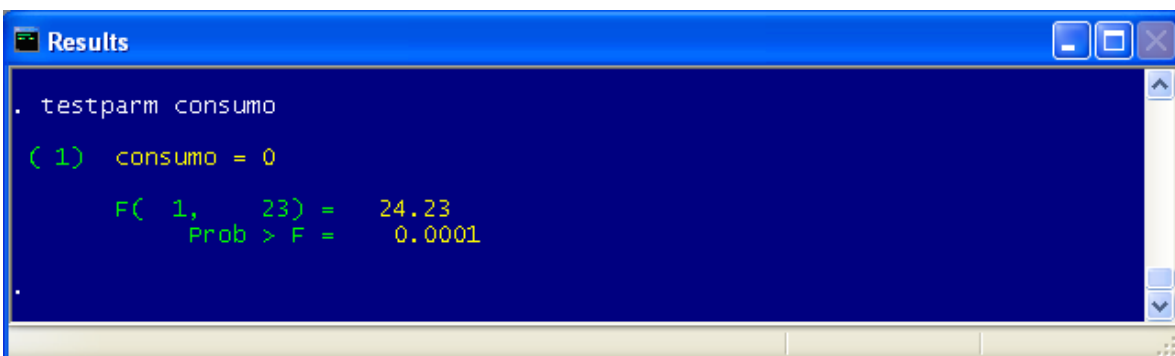
Como se aprecia hay una correlación muy alta entre el término constante y la variable exógena.

En la selección **Postestimation**, también se encuentra una selección que permite observar de forma individual los resultados de la prueba de Significancia Global del modelo que se evalúa con el estadístico F y que también se pueden observar en la salida de la estimación, esta se puede apreciar en la Gráfica:



GRAFICA 31. Prueba F individual

Arrojando así los siguientes resultados:

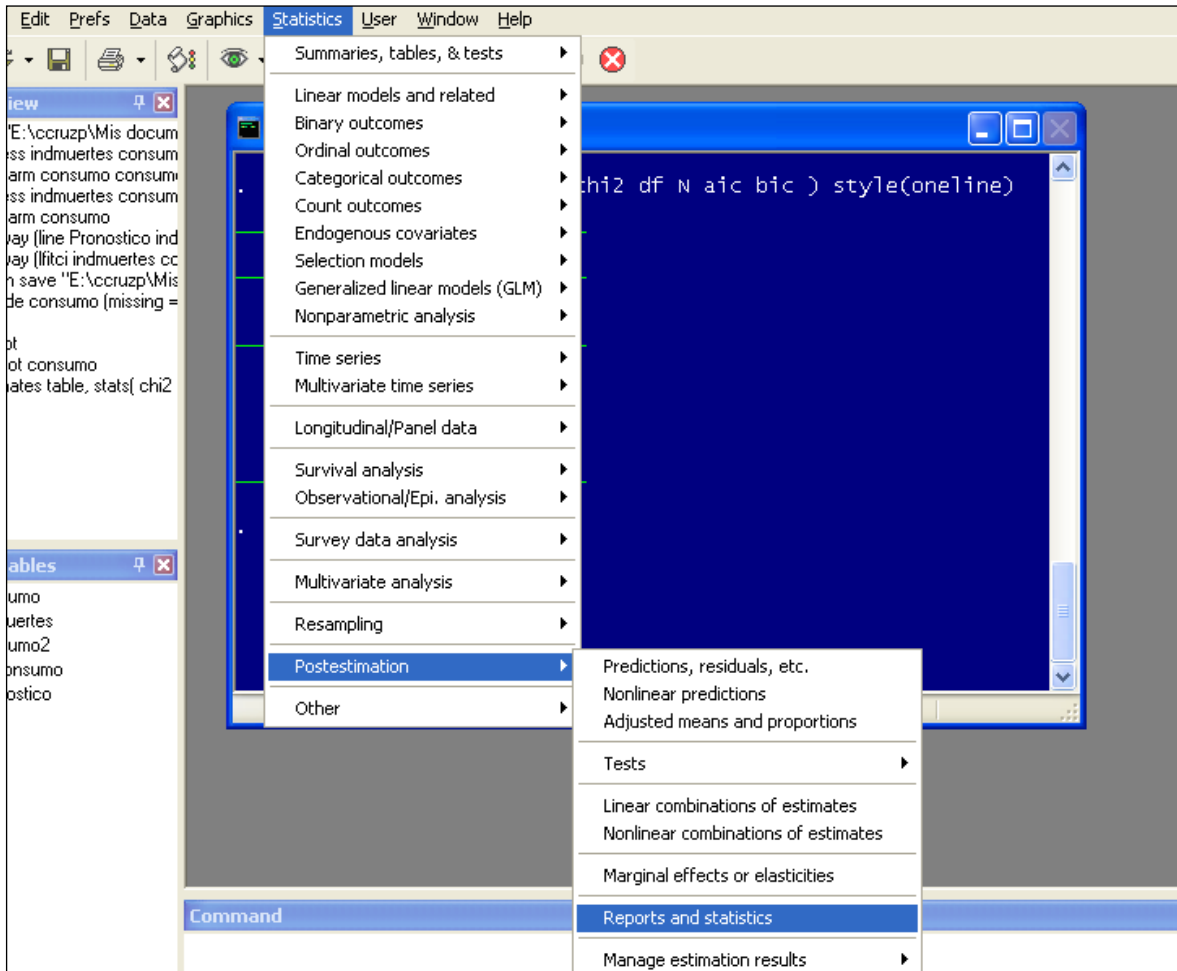


GRAFICA 32. Resultados Test

Como se puede apreciar en la gráfica, si se desea elaborar la prueba a través de

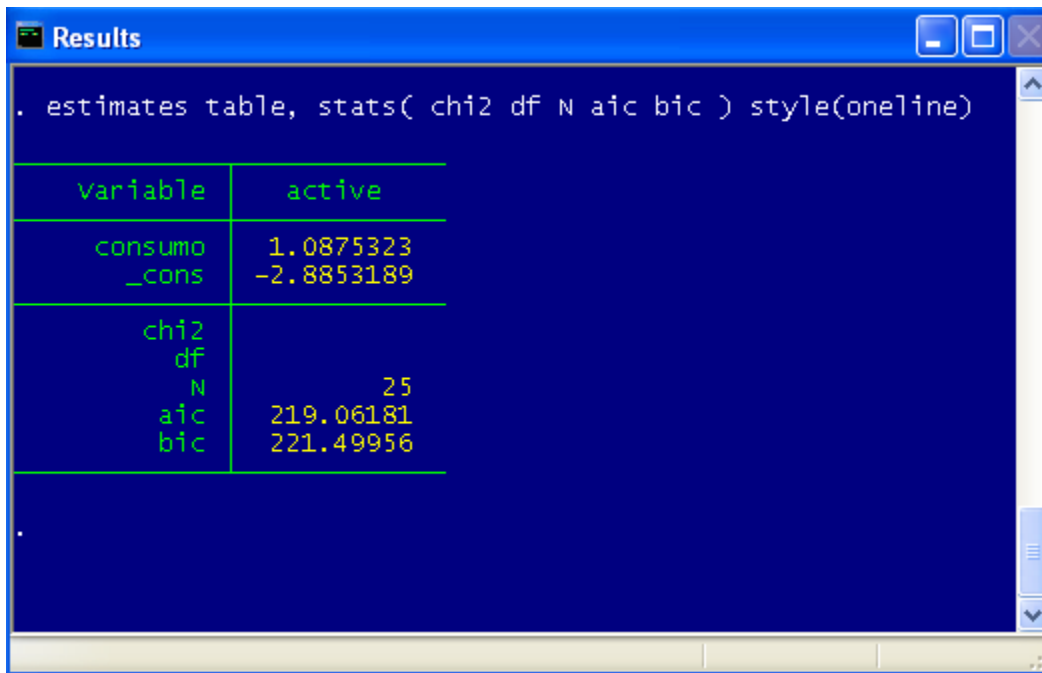
comandos, el comando a usar es el comando **“testparm”**. Los resultados de la prueba arrojan que al 95% de confianza no existe evidencia estadística para aceptar la hipótesis nula, por lo que se podría aceptar la hipótesis alterna la cual busca probar que por lo menos uno de los parámetros es diferente de 0.

Para complementar el análisis descriptivo, el menú **Statistics** ofrece opciones que nos permiten generar cuadros de resultados.



**GRAFICA 33. Pasos estadísticas complementarias**

Al seleccionar la opción que se aprecia en la gráfica, se activa un cuadro de diálogo que permite escoger al usuario, el reporte que desea se despliegue en la ventana de resultados.



Variable	active
consumo	1.0875323
_cons	-2.8853189
chi2	25
df	25
N	219.06181
aic	221.49956
bic	221.49956

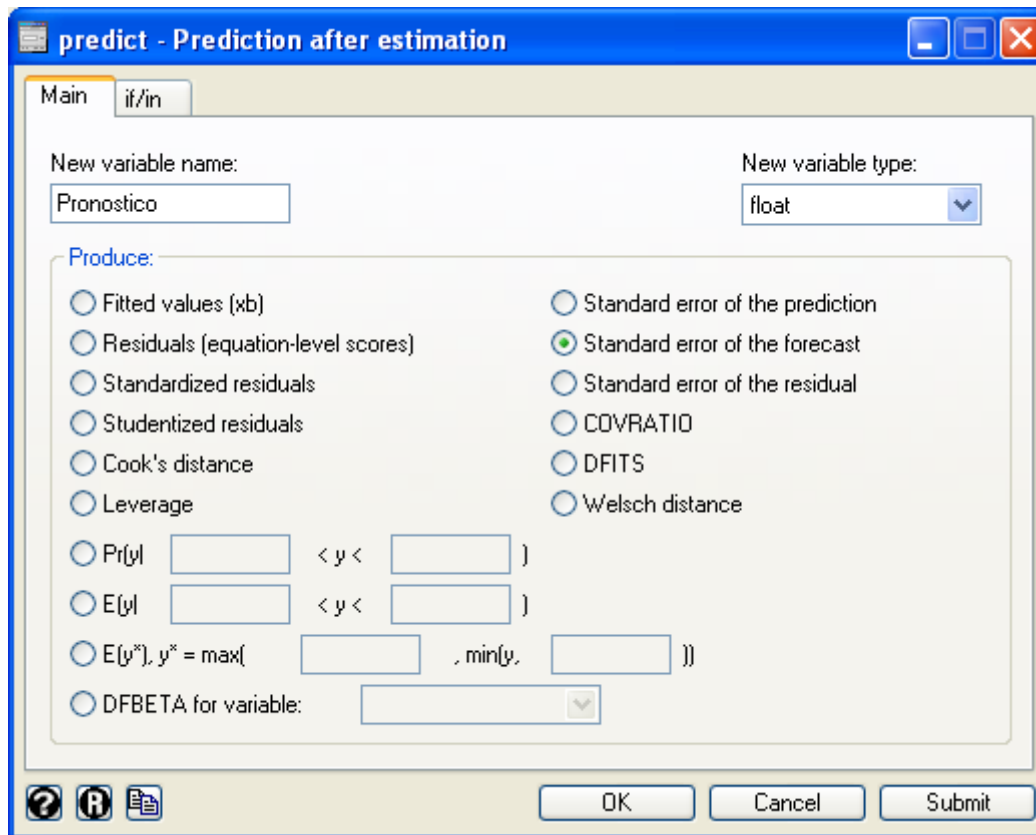
GRAFICA 34. Estadísticas Complementarias

Estas estadísticas complementarias del modelo, permiten observar los criterios Akaike Info Criterion (aic) y Schwartz Info Criterion (bic), los cuales son criterios que permiten escoger entre un modelo y otro alternativo.

### 3.2.8 Estimación y Predicción

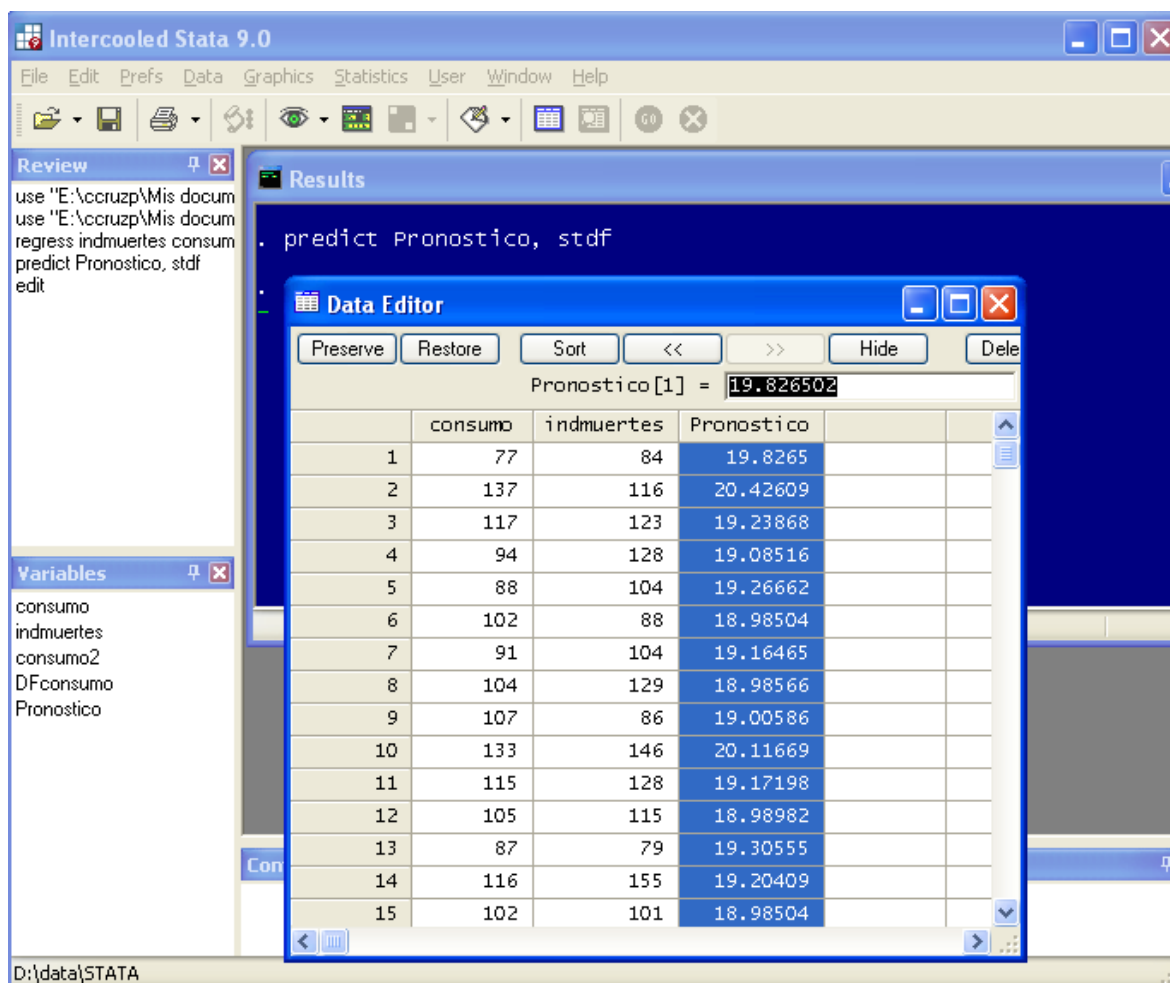
Stata 8.0 presenta la opción **Obtain predictions, residuals, etc., after estimation** en el menú **Statistics**, en la cual se realizan los pronósticos y permite revisar la serie de residuos. Esta opción se presenta de forma mucho más completa en Stata 9.0. Su ubicación se presenta en la Gráfica 35.





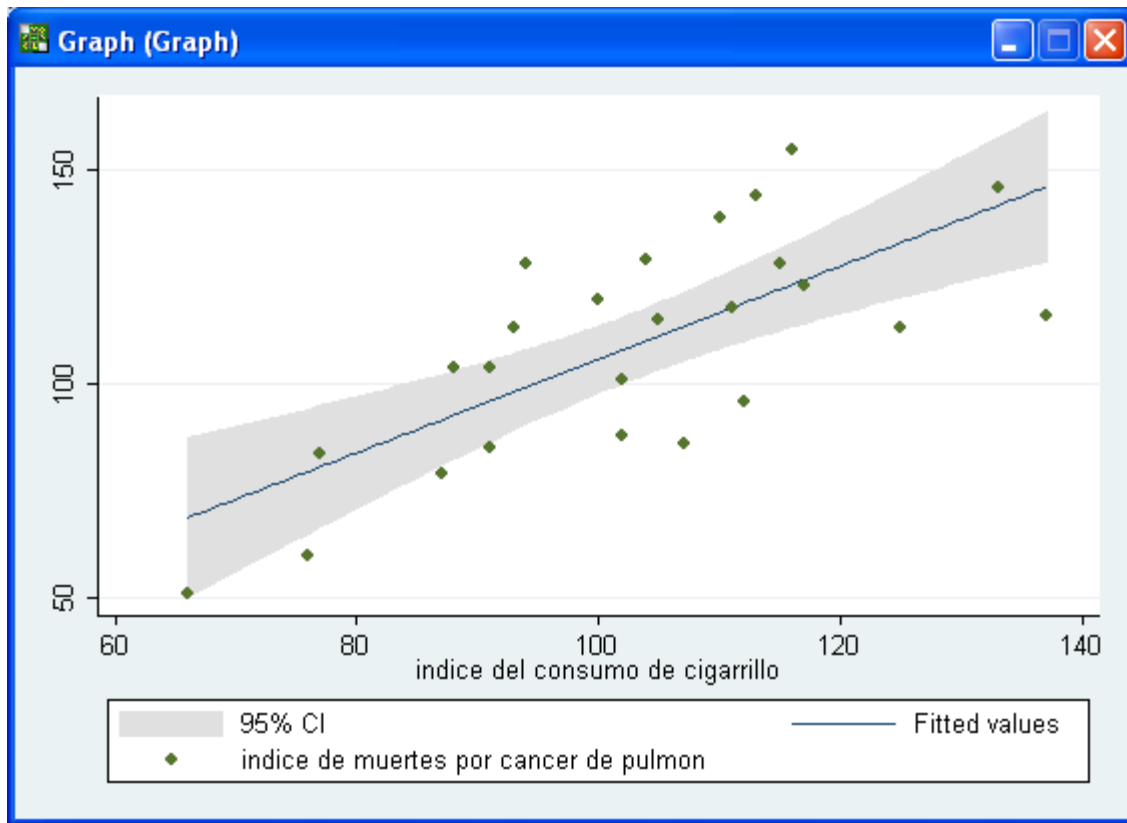
GRAFICA 36. Predict

Al seleccionar esta opción, aparecen las elecciones que se aprecian en la Gráfica 37. Al seleccionar OK, se genera una serie de pronósticos de la variable endógena en el Editor de Datos (**Data Editor**) como se aprecia en la siguiente Gráfica.



**GRAFICA 37. Generación de pronósticos**

Teniendo los pronósticos de la variable endógena, resulta conveniente graficarlos frente al valor observado de la variable y así poder concluir acerca de la calidad de la estimación realizada.



**GRAFICA 38. Datos estimados vs. Datos observados**

A través de este gráfico se puede apreciar el comportamiento de los valores ajustados, con un intervalo de confianza del 95% frente a los valores observados. Aunque los datos presentan dispersión frente a la línea de ajuste, según los resultados obtenidos a través de los estadísticos, el ajuste realizado es adecuado.

### 3.3 Series de Tiempo

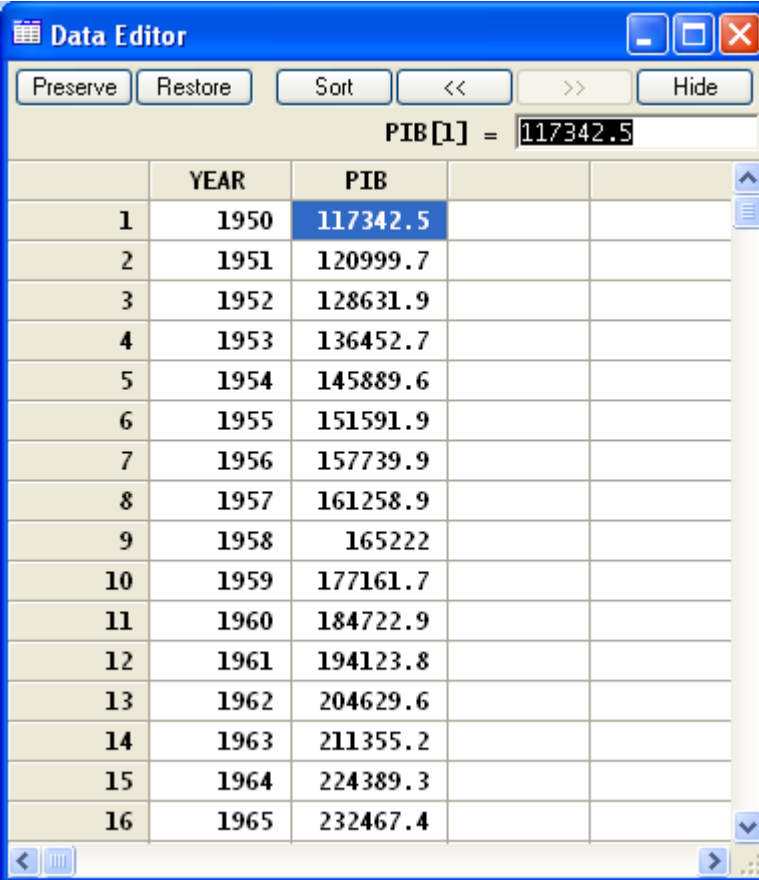
Los análisis se elaborarán a partir de la serie del Producto Interno Bruto (PIB) de Colombia para el período 1950-1996, la muestra se compone de 47 datos a través de los cuales se intentará definir el comportamiento de la serie representándolo a través de un modelo matemático.

#### 3.3.1 Introducción de datos:

La importación de la serie debe hacerse como se señala en secciones anteriores, es decir, luego de haber guardado previamente la serie con el formato csv. Es importante al importar el archivo, que previamente se le haya dado nombre a las series en el archivo de Excel. En los nombres, es recomendable no incluir letras

como ñ ni caracteres como la tilde ya que Stata 9.0 no los reconoce. Así, para la serie ejemplo en vez de AÑOS, se colocó YEAR como nombre para la misma. También es vital configurar en el computador a través de **Panel de Control** → **Configuración Regional** → **Idioma**, el separador de miles, es necesario que se coloque como separador el punto, para que Stata 9.0 no presente problemas al reconocer el archivo .csv o realizar el reemplazo de las comas por puntos desde el menú **Edición** → **Buscar** que ofrece **Excel**.

La siguiente gráfica muestra como se visualiza la serie en el **Data Editor** después de realizar el proceso de importación:



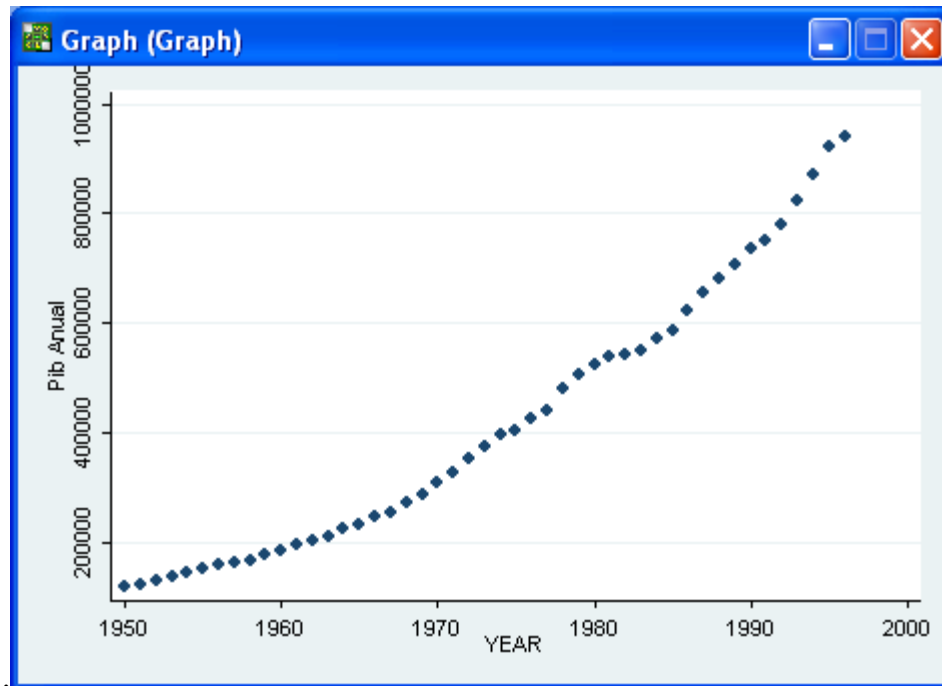
	YEAR	PIB		
1	1950	117342.5		
2	1951	120999.7		
3	1952	128631.9		
4	1953	136452.7		
5	1954	145889.6		
6	1955	151591.9		
7	1956	157739.9		
8	1957	161258.9		
9	1958	165222		
10	1959	177161.7		
11	1960	184722.9		
12	1961	194123.8		
13	1962	204629.6		
14	1963	211355.2		
15	1964	224389.3		
16	1965	232467.4		

GRAFICA 39. Serie en Data Editor

Se señala que también se pueden introducir los datos de manera manual como ya se había señalado antes.

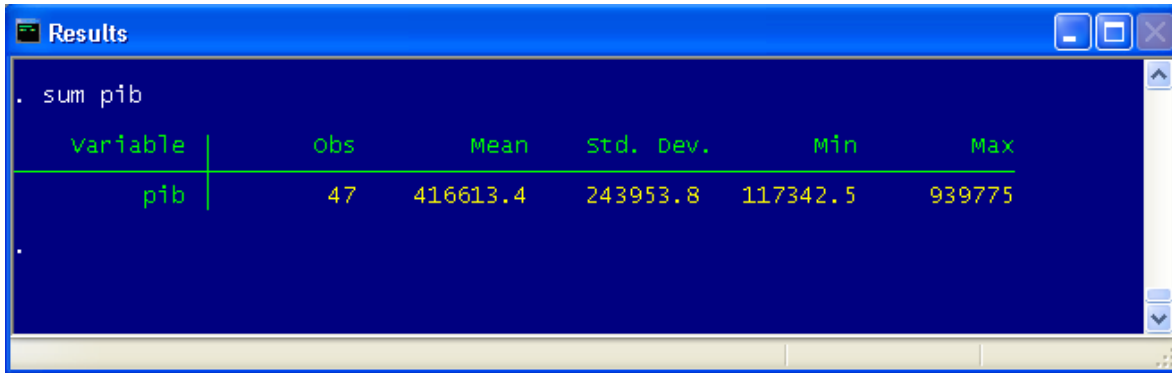
### 3.3.2 Descripción de los datos

Es importante observar el comportamiento gráfico de la serie antes de iniciar el proceso, dado que este puede indicar si el comportamiento de la serie es fluctuante, entre otros también facilita detección de cambio estructural. El comportamiento de la serie a analizar se aprecia en la siguiente gráfica:



**GRAFICA 40. Gráfica Serie PIB**

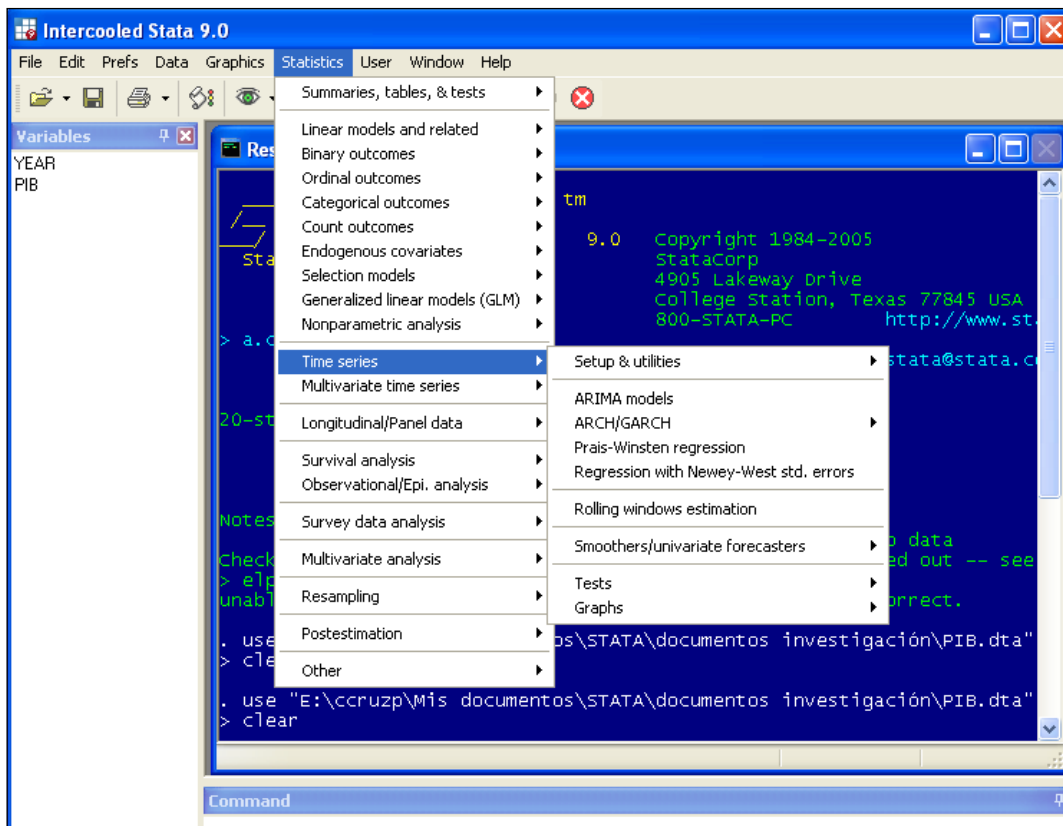
Para observar las estadísticas descriptivas de la serie aplicamos el comando “*sum*” ya visto anteriormente, para lo cual el software arroja los siguientes resultados.



**GRAFICA 41. Estadísticas descriptivas de la serie**

Como se aprecia en las estadísticas descriptivas, la muestra es de 47 datos, los cuales presentan una media de 416613,4 millones de pesos y una desviación de 243953.8.

Stata 9.0 permite elaborar diversos análisis para series de tiempo, su ubicación se puede apreciar en la gráfica que se observa a continuación:



**GRAFICA 42 Ubicación análisis de series de tiempo**

Estos análisis serán explorados en las siguientes secciones.

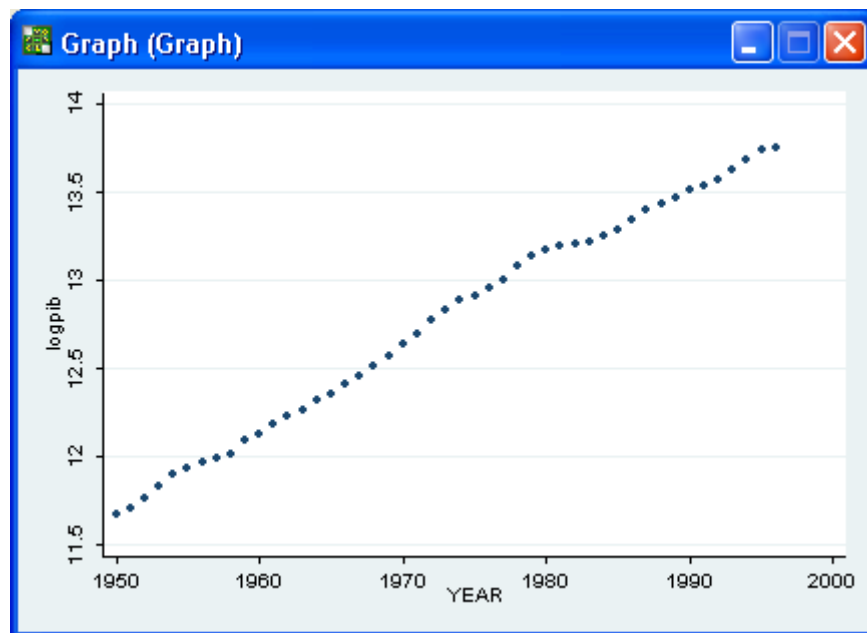
### 3.3.3 Generación del modelo:

El comportamiento de la serie es lineal en términos generales, sin embargo siguiendo la metodología de Box-Cox es necesario suavizar la varianza y la media de la misma. El proceso de suavizamiento de la media requiere la transformación de la serie, para este tipo de serie la más adecuada es la transformación logarítmica. La transformación se realiza por medio de comandos que fueron analizados en manuales anteriores. Para el ejemplo, aplica la siguiente secuencia:

```
Command
gen logpib = log(pib)
```

GRAFICA 43. Comando log

Con lo cual como se observo en una de las secciones anteriores, genera una nueva variable en la ventana de variables, como se aprecia en la Gráfica 43. El comportamiento de esta nueva serie se aprecia en la siguiente gráfica:

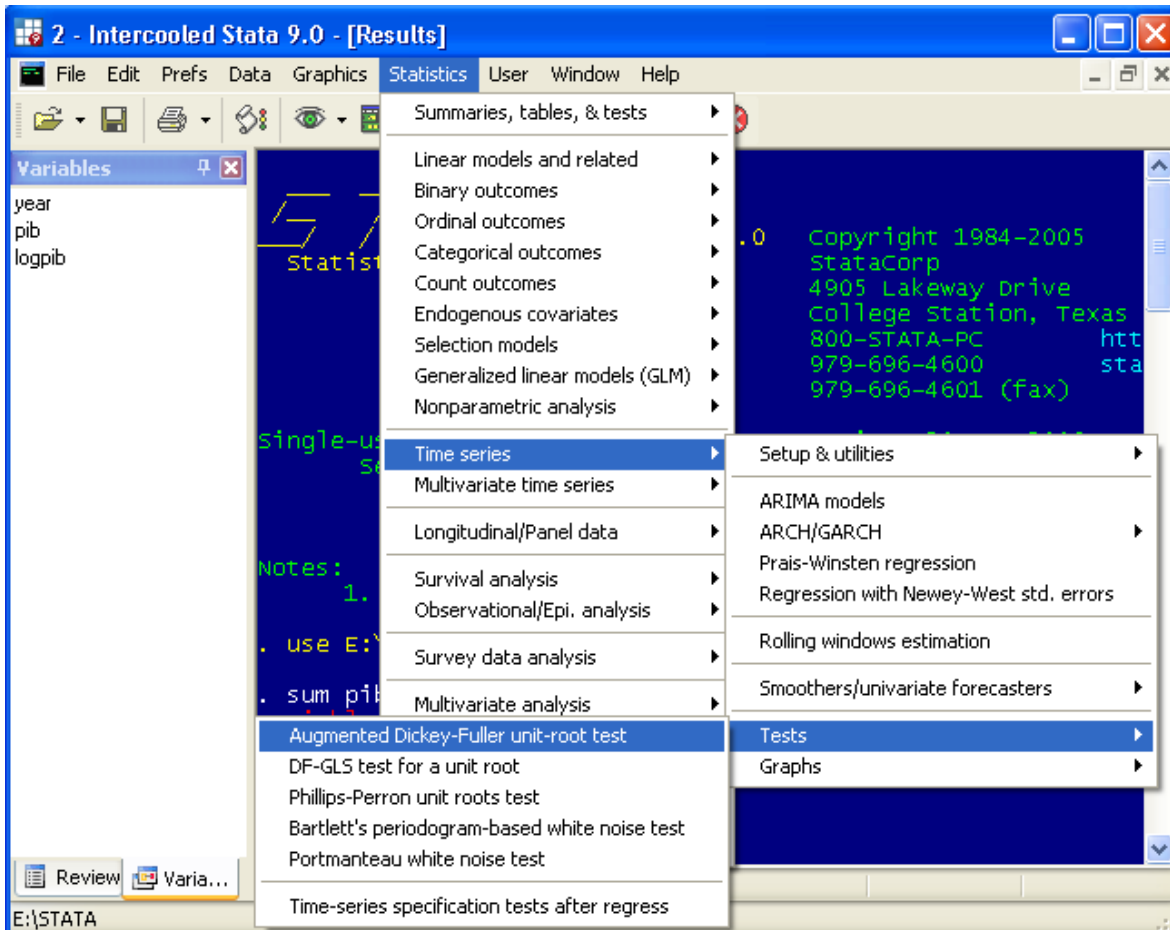


GRAFICA 44. Gráfico Logpib

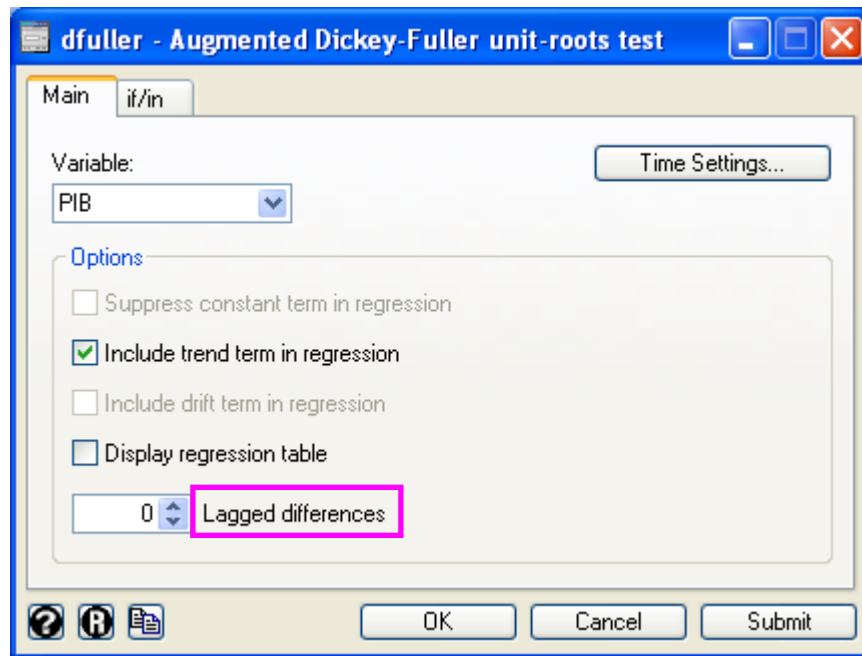
El siguiente paso según la metodología, es suavizar la media de la serie, lo cual se realiza por medio de la diferenciación. Para esto es necesario realizar la prueba de raíz unitaria a la serie, la cual permitirá determinar si la serie requiere

diferenciación y si así es, cuantas requiere través de la significancia del Estadístico Dickey Fuller.

Stata 9.0 permite aplicar esta prueba a través del menú **Statistics**, en donde se encuentra la opción **Time Series**→**Test**, como se ve en la Gráfica:



GRAFICA 45. Ubicación Prueba de raíz Unitaria

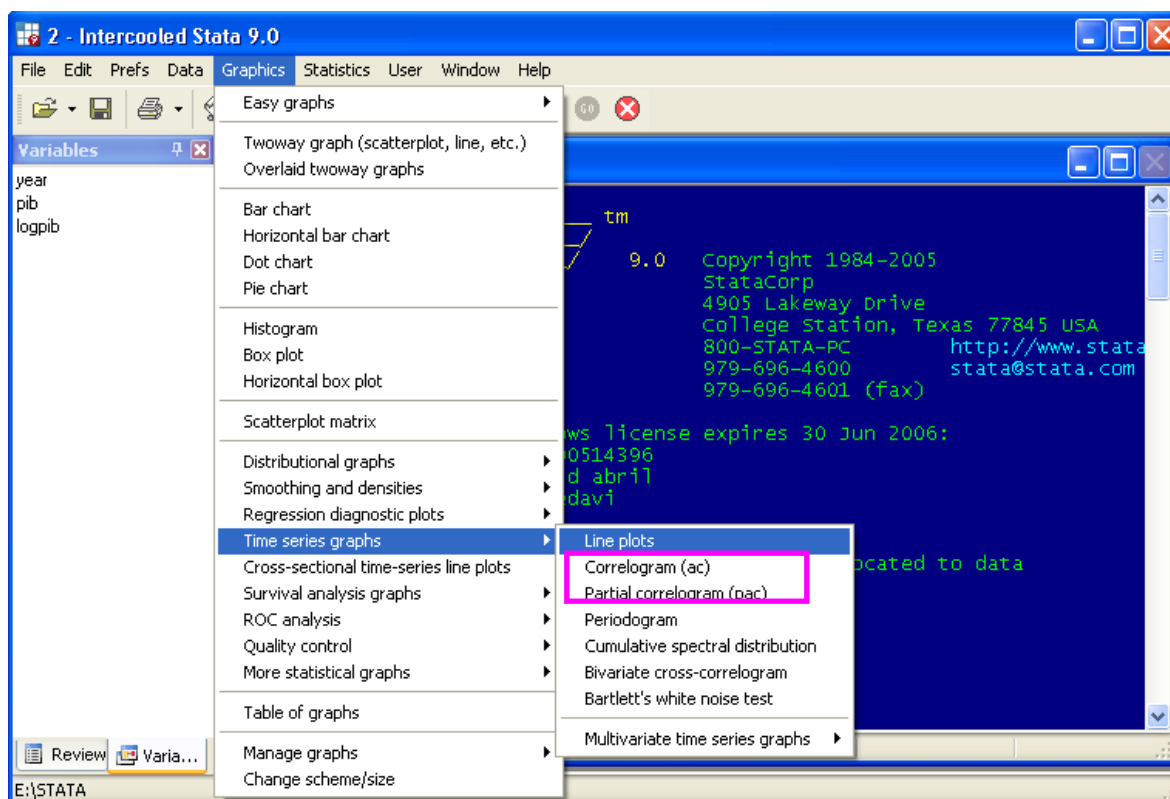


GRAFICA 46. Prueba de raíz unitaria

A través de la significancia del estadístico que arroja la prueba se puede determinar si esta quiere requiere diferenciación o no, es decir, si la significancia del estadístico es menor a 0.05 se rechaza la hipótesis nula de raíz unitaria y la serie no requiere diferenciación. Es decir, si la serie original arroja un estadístico con un nivel de significancia menor a 0.05, esta no requiere diferenciación, si el estadístico es mayor se aplica primera diferencia a la serie o si es necesario, segunda diferencia. Esto se puede determinar en la opción del cuadro denominada **Lagged differences** que se señala en la Gráfica 45.

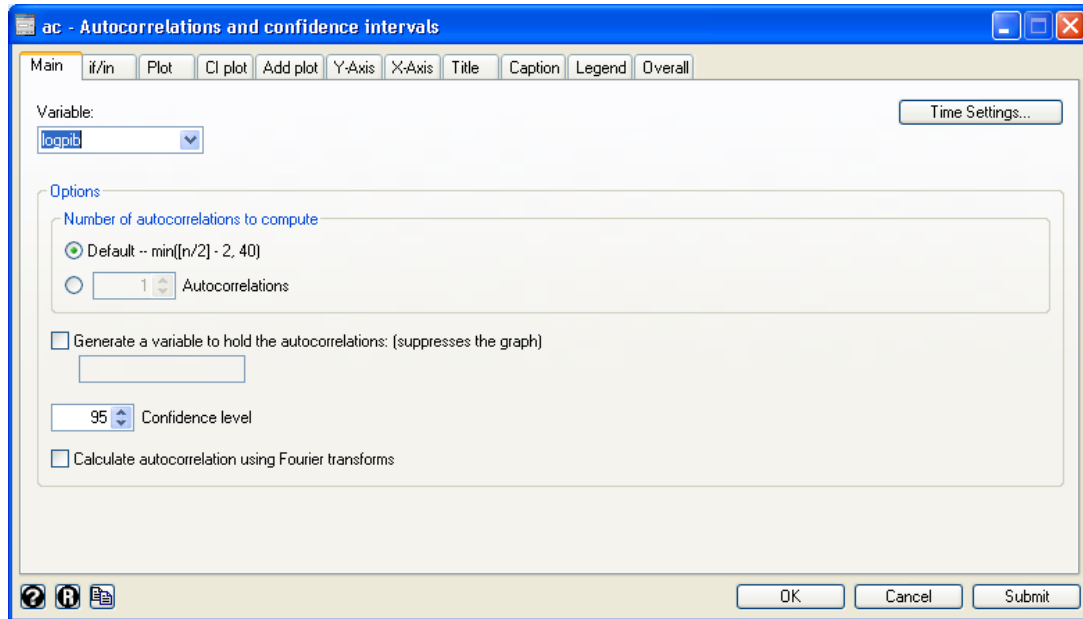
Luego de determinar el modelo de diferenciaciones se elaboran la FAS (Función de Autocorrelación Simple) y la FAP (Función de Autocorrelación Parcial) de la serie las cuales permiten tener una primera aproximación del modelo que mejor representaría la serie a través de la significancia de los rezagos.

Stata 9.0 permite elaborar estas dos funciones a través de la opción **Graphics**→**Time Series Graphs**, su ubicación se aprecia en la siguiente gráfica:



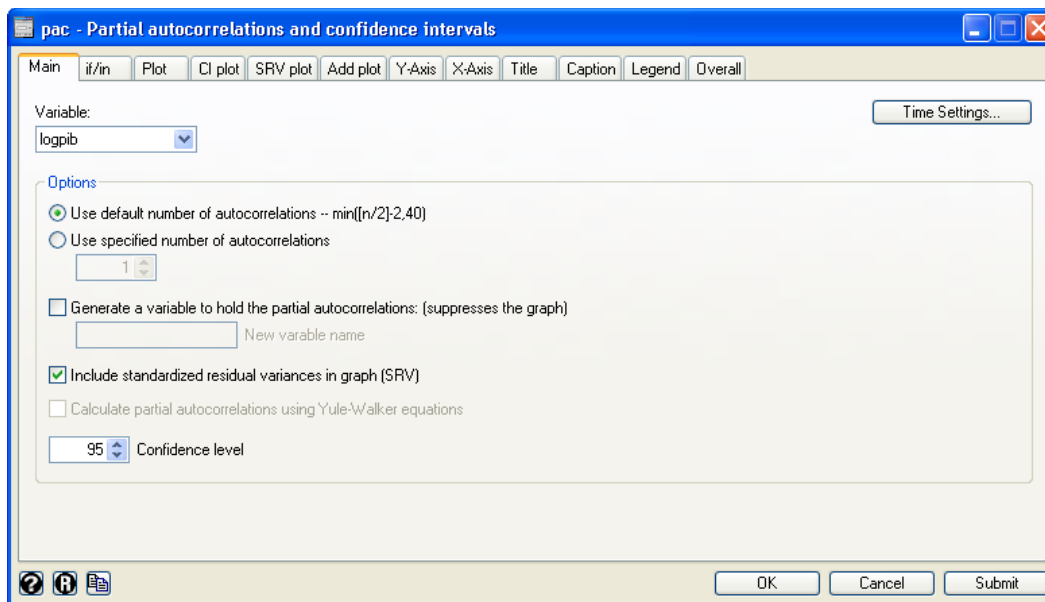
**GRAFICA 47. Ubicación FAS y FAP**

Al seleccionar la primera de estas opciones, **Correlogram (ac)** aparece el siguiente cuadro:



**GRAFICA 48. Función de Autocorrelación Simple**

Al seleccionar la opción **Partial Correlogram (pac)**, se generan las siguientes opciones:



**GRAFICA 49. Función de Autocorrelación Parcial**

Los resultados que arrojan estas pruebas permitirán determinar una aproximación del modelo matemático que mejor representa la serie escogida.