

# GUIA DE INTRODUCCIÓN A LA ECONOMETRÍA UTILIZANDO GRETl

**M<sup>a</sup> ISABEL CAL BOUZADA – M<sup>a</sup> VICTORIA VERDUGO MATÉS**

<b>Capítulo 1. INTRODUCCIÓN A GRETL</b>	<b>1</b>
<b>1.1. Presentación de Gretl</b>	<b>1</b>
<b>1.2. Inicio de una sesión de trabajo</b>	<b>1</b>
1.2.1. Ventana Principal o Ventana del Programa	2
<b>1.3. ¿Cómo ejecutar Gretl?</b>	<b>3</b>
<b>1.4. Configuración de Gretl</b>	<b>4</b>
<b>Capítulo 2. MANEJO Y ANÁLISIS DE DATOS</b>	<b>7</b>
<b>2.1. Presentación</b>	<b>7</b>
<b>2.2. Introducción de datos</b>	<b>7</b>
2.2.1. Introducción de datos directamente en el programa	7
2.2.1.1. ¿Cómo introducir datos?	8
2.2.2. Introducción de datos a través de la Consola Gretl y/o de los guiones o ficheros de comandos	10
2.2.3. ¿Cómo crear un fichero de datos?	15
2.2.3.1. Utilizar el menú Archivo	15
2.2.3.2. Utilizar el comando store	16
2.2.4. Recuperación de datos desde un fichero	17
2.2.4.1. Ficheros de datos de Gretl	17
2.2.4.1.1. Utilizar el menú Archivo	18
2.2.4.1.2. Utilizar el comando open	18
2.2.4.2. Ficheros de datos de Excel	18
2.2.4.2.1. Utilizar el menú Archivo	19
2.2.4.2.2. Utilizando el comando open	20
<b>2.3. ¿Cómo ver las variables?</b>	<b>21</b>
2.3.1. Utilizar el comando print	22
<b>2.4. ¿Cómo generar nuevas variables?</b>	<b>22</b>
2.4.1. Utilizar el comando genr	23
2.4.2. Comandos para generar varias variables a la vez	29
<b>2.5. ¿Cómo saber que variables están disponibles en la sesión de trabajo?</b>	<b>31</b>
2.5.1. Utilizar el menú Ver	31
2.5.2. Utilizar el comando varlist	32
<b>2.6. ¿Cómo cambiar el nombre de las variables?</b>	<b>33</b>
2.6.1. Utilizar el menú Variable	33
2.6.2. Utilizar el comando rename	33
<b>2.7. ¿Cómo borrar variables?</b>	<b>34</b>
2.7.1. Utilizar el comando delete	34
<b>2.8. ¿Cómo ordenar variables?</b>	<b>35</b>
<b>2.9. ¿Cómo generar vectores y matrices?</b>	<b>35</b>
2.9.1. Utilizar el menú Añadir	35
2.9.2. Utilizar el comando matrix	36
2.9.2.1. ¿Cómo identificar los elementos de una matriz?	48
2.9.2.2. ¿Cómo identificar las filas de una matriz?	48
2.9.2.3. ¿Cómo identificar las columnas de una matriz?	49
<b>2.10. Relación entre los comandos genr, matrix y scalar</b>	<b>49</b>
<b>2.11. ¿Cómo calcular los estadísticos descriptivos de las variables?</b>	<b>50</b>
2.11.1. Utilizar el menú Ver	50
2.11.2. Utilizar el comando summary	51

<b>2.12. ¿Cómo hacer representaciones gráficas?</b>	<b>55</b>
2.12.1. Utilizar el menú Ver	55
2.12.2. Utilizar el comando gnuplot	57
2.12.3. Utilizar el comando scatters	59
2.12.4. Utilizar el comando textplot	60
<b>2.13. ¿Cómo modificar el rango muestral?</b>	<b>61</b>
2.13.1. Utilizar el menú Muestra	61
2.13.2. Utilizar el comando smpl	62
<b>Capítulo 3. MCO: MODELO DE REGRESIÓN LINEAL CLÁSICO</b>	<b>64</b>
<b>3.1. Presentación e Hipótesis Básicas del Modelo de Regresión Lineal Múltiple</b>	<b>64</b>
<b>3.2. ¿Cómo estimar por MCO en Gretl?</b>	<b>66</b>
3.2.1. Utilizar el menú Modelo	66
3.2.2. Utilizando el comando ols	69
<b>3.3. Análisis de la información proporcionada para una estimación MCO de un modelo formulado con ordenada en el origen</b>	<b>70</b>
<b>3.4. Análisis de los residuos</b>	<b>73</b>
<b>3.5. Análisis de las sumas de cuadrados: tabla ANOVA</b>	<b>74</b>
<b>3.6. Análisis gráfico: menú Gráficos de la Ventana Modelo</b>	<b>75</b>
<b>3.7. Estimación MCO de un modelo formulado sin ordenada en el origen</b>	<b>76</b>
<b>3.8. Interpretación de los coeficientes</b>	<b>78</b>
<b>3.9. Formas funcionales alternativas</b>	<b>80</b>
<b>3.10. Predicción</b>	<b>81</b>
3.10.1. Utilizar el menú Análisis	81
3.10.2. Utilizando el comando fcast	83
<b>3.11. Ejercicio de aplicación</b>	<b>84</b>
3.11.1. Estimación MCO del modelo con ordenada en el origen	84
3.11.1.1. Resultados de la estimación utilizando álgebra matricial	86
3.11.2. Interpretación de coeficientes	97
3.11.3. Análisis gráfico	100
3.11.4. Estimación MCO del modelo sin ordenada en el origen	103
3.11.4.1. Resultados de la estimación utilizando álgebra matricial	104
3.11.5. Formas funcionales alternativas	110
3.11.6. Predicción	114
3.11.6.1. Predicción a través del comando fcast	115
3.11.6.2. Predicción utilizando álgebra matricial	115
<b>Capítulo 4. CONTRASTES DE HIPÓTESIS Y REGIONES DE CONFIANZA</b>	<b>118</b>
<b>4.1. Hipótesis del Modelo de Regresión Lineal Normal Clásico (MRLNC)</b>	<b>118</b>
<b>4.2. Contrastes de hipótesis</b>	<b>118</b>
<b>4.3. ¿Cómo realizar contrastes de hipótesis en Gretl?</b>	<b>120</b>
4.3.1. Utilizar el menú Contrastes	120
Debe de tenerse en cuenta que aunque las versiones t y F son válidas para muestras pequeñas, la versión W sólo es válida si la muestra es lo suficientemente grande.	122
4.3.2. Utilizando el comando restrict	122
<b>4.4. ¿Cómo buscar el p-valor y el valor crítico en una distribución?</b>	<b>122</b>
4.4.1. Valor crítico	122
4.4.2. P-valor	124

<b>4.5. Regiones de Confianza</b>	<b>127</b>
4.5.1. ¿Cómo calcular intervalos y regiones de confianza en Gretl?	127
4.5.2. ¿Cómo utilizar el intervalo de confianza en el contraste de hipótesis?	128
<b>4.6. Ejercicio de aplicación</b>	<b>128</b>
4.6.1. Estimación MCO del modelo	128
4.6.2. Contrastes de nulidad individual	129
4.6.2.1. Contrastes de nulidad individual utilizando del estadístico t	129
4.6.2.2. Contrastes de nulidad individual a través del comando restrict	131
4.6.3. Contraste de nulidad conjunta para todos los parámetros del modelo	132
4.6.4. Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas del modelo	133
4.6.5. Contraste de nulidad conjunta para un subconjunto paramétrico	134
4.6.6. Contraste de nulidad para una combinación lineal	135
4.6.7. Intervalos de confianza	136
4.6.8. Regiones de confianza	139
4.6.9. Contrastes de hipótesis utilizando la matriz de restricciones lineales (R)	140
4.6.10. Contrastes de hipótesis mediante sumas residuales	151
4.6.11. Contrastes de nulidad utilizando del comando omit	155
4.6.12. Contrastes de nulidad utilizando el comando add	156
4.6.13. Contrastes de nulidad utilizando el comando coeffsum	156
4.6.14. Contrastes de no nulidad	157

## ÍNDICE ILUSTRACIONES

Ilustración 1-1. Formas de iniciar Gretl.	1
Ilustración 1-2. <b>Ventana Principal</b> del programa	2
Ilustración 1-3. <b>Ventana Principal</b> con un ejemplo donde Gretl informa del fichero de datos (DATOS1.gdt), del rango muestral (1-20) y de algunas de las características de las variables disponibles (número de identificación, denominación y descripción).	3
Ilustración 1-4. <b>Ventana Consola Gretl.</b>	3
Ilustración 1-5. <b>Ventana archivo de comandos.</b>	4
Ilustración 1-6. Cuadro de diálogo para definir el directorio de trabajo por defecto.	5
Ilustración 1-7. Configuración de preferencias.	6
Ilustración 2-1. Cuadro de diálogo para dimensionar la base de datos.	7
Ilustración 2-2. Cuadro de diálogo para seleccionar la estructura del conjunto de datos.	9
Ilustración 2-3. Secuencia para acceder a la <b>Ventana editar datos</b> utilizando el submenú <b>Añadir</b> del <b>Menú Principal</b> para introducir datos en Gretl.	10
Ilustración 2-4. Añadir variables utilizando la <b>Ventana editar datos.</b>	11
Ilustración 2-5. Guardar datos utilizando el menú <b>Archivo</b> de la <b>Barra de Menú</b> de la <b>Ventana Principal.</b>	16
Ilustración 2-6. Recuperación de datos desde un fichero de datos de Gretl utilizando el menú <b>Archivo</b> de la <b>Barra de Menú</b> de la <b>Ventana Principal.</b>	17
Ilustración 2-7. Recuperación de datos desde un fichero de datos de Excel utilizando el menú <b>Archivo</b> de la <b>Barra de Menú</b> de la <b>Ventana Principal.</b>	19
Ilustración 2-8. Autoetiquetado de variables.	31
Ilustración 2-9. Editor del conjunto de datos de las variables disponibles en una sesión de trabajo.	31
Ilustración 2-10. Editor del conjunto de escalares disponibles en una sesión de trabajo.	31
Ilustración 2-11. Ejemplo de como se pueden cambiar los atributos de una variable.	33
Ilustración 2-12. Ejemplo de ventana de ordenación de variables.	35
Ilustración 2-13. Ejemplos de generación de matrices utilizando el menú <b>Añadir.</b>	35
Ilustración 2-14. Menú que permite ver el contenido y las propiedades de una matriz, editarla, renombrarla y guardarla.	36
Ilustración 2-15. Formas de acceder a la <b>Ventana estadísticos principales.</b>	50
Ilustración 2-16. Formas de acceder a la <b>Ventana distribución de frecuencias y del histograma.</b>	54
Ilustración 2-17. Ejemplo de construcción de un gráfico de dispersión utilizando el menú <b>Ver.</b>	56
Ilustración 2-18. Menú emergente de la <b>Ventana gráfico.</b>	57
Ilustración 2-19. Secuencia para abrir y modificar un gráfico desde un fichero gnuplot.	59
Ilustración 2-20. Diferencia entre el comando <b>scatters</b> y el comando <b>gnuplot.</b>	60
Ilustración 2-21. Opciones del menú <b>Muestra</b> de la <b>Ventana Principal.</b>	61
Ilustración 3-1. Menú <b>Modelo</b> de la <b>Barra de Menú</b> de la <b>Ventana Principal.</b>	66
Ilustración 3-2. <b>Ventana especificar modelo:</b> selección de variables.	67
Ilustración 3-3. Menú <b>Muestra</b> de la <b>Barra de Menú</b> de la <b>Ventana Principal:</b> establecer rango muestral.	67
Ilustración 3-4. Menú <b>Guardar</b> de la <b>Ventana Modelo.</b>	68
Ilustración 3-5. Gráfico valores observados – valores estimados.	75
Ilustración 3-6. Gráfico de residuos.	76
Ilustración 3-7. Opción <b>Predicciones</b> del menú <b>Análisis</b> de la <b>Ventana Modelo.</b>	82
Ilustración 3-9. Gráficos de residuos.	101
Ilustración 3-8. Gráficos de valores observados y estimados del regresando.	101
Ilustración 3-10 Gráfico de valores estimados y observados del regresando respecto a las variables explicativas.	102
Ilustración 3-11. Gráfico valores estimados del consumo y Gráfico residuos equivalentes.	114
Ilustración 4-1. Contrastes de hipótesis lineales utilizando el menú <b>Contrastes.</b>	120
Ilustración 4-2. Búsqueda del valor crítico.	123
Ilustración 4-3. Búsqueda del p-valor.	125
Ilustración 4-4. Intervalos de confianza del 95% utilizando el menú <b>Análisis</b> de la <b>Ventana Modelo.</b>	139
Ilustración 4-5. Elipse de confianza del 95% utilizando el menú <b>Análisis</b> de la <b>Ventana Modelo.</b>	140
Ilustración 4-6. Ejemplos de elipses de confianza.	140

## ÍNDICE TABLAS

<b>Tabla 2-1.</b> Tabla de estadísticos descriptivos. _____	51
<b>Tabla 3-1.</b> Salida asociada al submenú <b>Mínimos Cuadrados Ordinarios</b> del menú <b>Modelo</b> de la <b>Ventana Principal</b> para la estimación MCO de un modelo formulado con ordenada en el origen. _____	70
<b>Tabla 3-2.</b> Salida asociada al submenú <b>Mostrar variable observada, estimada, residuos</b> del menú <b>Análisis</b> de la <b>Ventana Modelo</b> . _____	73
<b>Tabla 3-3.</b> Salida asociada al submenú <b>ANOVA</b> del menú <b>Análisis</b> de la <b>Ventana Modelo</b> . _____	74
<b>Tabla 3-4.</b> Salida asociada al submenú <b>Mínimos Cuadrados Ordinarios</b> del menú <b>Modelo</b> de la <b>Ventana Principal</b> para la estimación MCO de un modelo formulado sin ordenada en el origen. _____	77
<b>Tabla 3-5.</b> Salida asociada al submenú <b>Predicciones</b> del menú <b>Análisis</b> de la <b>Ventana Modelo</b> . _____	83
<b>Tabla 3-6.</b> Tabla <b>ANOVA</b> respecto al origen. _____	93
<b>Tabla 3-7.</b> Tabla <b>ANOVA</b> respecto a la media. _____	108
<b>Tabla 3-8.</b> Tabla <b>ANOVA</b> respecto al origen. _____	108
<b>Tabla 3-9.</b> Comparación bondad de ajuste. _____	114
<b>Tabla 4-1.</b> Salida asociada al submenú <b>Restricciones lineales</b> del menú <b>Contrastes</b> de la <b>Ventana Modelo</b> . _____	121
<b>Tabla 4-2.</b> Resumen estadísticos contraste nulidad individual. _____	132
<b>Tabla 4-3.</b> Resumen intervalos de confianza. _____	138

---

*ÍNDICE CUADROS*

<i>Cuadro 2-1. Algunas opciones del comando <b>store</b>.</i>	<i>16</i>
<i>Cuadro 2-2. Algunos operadores disponibles con el comando <b>genr</b>.</i>	<i>23</i>
<i>Cuadro 2-3. Algunas variables temporales disponibles.</i>	<i>32</i>
<i>Cuadro 2-4. Algunos operadores disponibles con el comando <b>matrix</b>.</i>	<i>39</i>
<i>Cuadro 2-5. Algunas opciones del comando <b>summary</b>.</i>	<i>52</i>
<i>Cuadro 2-6. Algunas opciones del comando <b>gnuplot</b>.</i>	<i>58</i>
<i>Cuadro 3-1. Opciones del comando <b>ols</b>.</i>	<i>69</i>
<i>Cuadro 3-2. Variables temporales asociadas al comando <b>ols</b>.</i>	<i>69</i>
<i>Cuadro 3-3. Algunas opciones del comando <b>fcast</b>.</i>	<i>83</i>

# Capítulo 1. INTRODUCCIÓN A GRETTL

## 1.1. Presentación de Gretl

Gretl, acrónimo de Gnu Regression Econometric y Time Series, es un paquete de software libre desarrollado en la Universidad de Wake Forest por Allin Cottrell. Es relativamente fácil de usar, con una gran flexibilidad y con grandes posibilidades en el manejo de datos, incluyendo operaciones de álgebra matricial. Además de permitir la estimación y el contraste de multitud de modelos econométricos, permite establecer rutinas de programación que facilitarán en gran medida el trabajo del usuario avanzado.

Una de las "ventajas" de Gretl es su presentación bajo el entorno Windows, basada en menús desplegables y ventanas, lo que facilita al usuario una rápida familiarización con el mismo y la utilización de todos los programas de dicho entorno de forma totalmente compatible.

A pesar de que las **ventanas o cuadros de diálogo** pueden resultar insuficientes<sup>1</sup> para realizar determinadas tareas, es el modo aconsejable para iniciarse en el manejo del programa, pues no requiere conocimiento previo alguno, salvo estar familiarizado con programas que funcionen bajo el entorno Windows. No es necesario conocer las instrucciones ni su formato de escritura, es suficiente con que el usuario haga la selección oportuna en cada una de las ventanas o cuadros de diálogo que se van abriendo desde que se selecciona la tarea o procedimiento en los distintos menús y submenús disponibles en la **Ventana Principal** del programa.

## 1.2. Inicio de una sesión de trabajo

Para iniciar una sesión de Gretl se deben realizar las mismas operaciones que para el inicio de cualquier programa que funcione bajo entorno Windows. Se podrá optar por las siguientes alternativas (véase Ilustración 1-1):

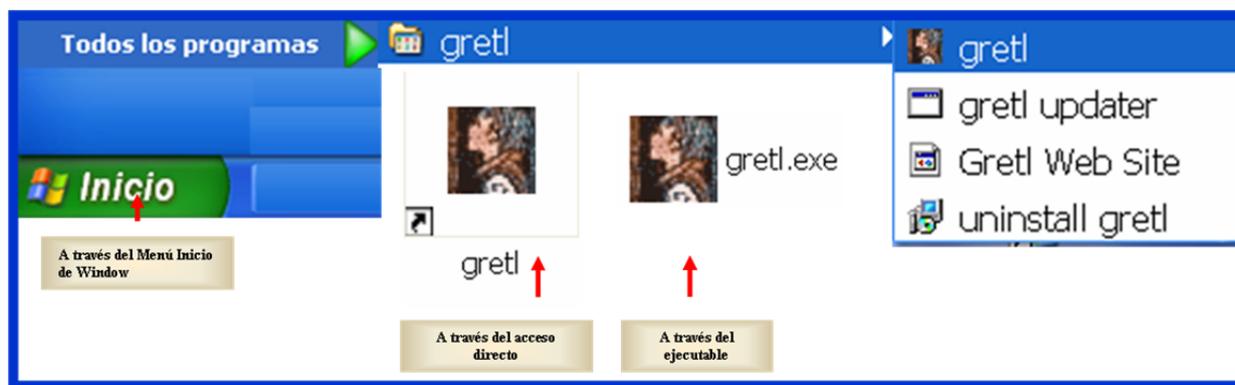


Ilustración 1-1. Formas de iniciar Gretl.

- Utilizar un acceso directo al programa, que previamente se habrá creado en el Escritorio de Windows, siendo esta la alternativa más cómoda para iniciar una sesión de trabajo.
- Desplegar el menú de Inicio de Windows y acceder a Todos los programas → gretl → gretl.
- Desde el Escritorio de Windows acceder al ejecutable *gretl.exe*.

Cualquiera de estas alternativas permite abrir la Ventana Principal de Gretl e iniciar una sesión de trabajo.

<sup>1</sup> Aunque no todos los comandos o instrucciones tienen asociados **ventanas o cuadros de diálogo**, Gretl abarca una amplia gama de procedimientos econométricos.

### 1.2.1. Ventana Principal o Ventana del Programa

La Ventana Principal o Ventana del Programa es el marco en el que se abren y distribuyen las restantes ventanas y/o cuadros de diálogo disponibles en Gretl.

En esta ventana, se pueden distinguir los siguientes elementos:

- La **Barra de Título**, que en su parte izquierda contiene el icono y el nombre del programa y en su parte derecha los iconos para **Minimizar**, **Restaurar** o **Cerrar** dicha ventana.
- La **Barra de Menú**, que contiene los botones que dan acceso al menú principal, los cuales no estarán totalmente disponibles hasta que no se carguen datos (por lo que aparecerán escritos en color claro).

Al situar el cursor encima de uno de los botones operativos de la barra de menú se abre el submenú correspondiente, que puede no estar totalmente disponible, dependiendo del tipo de datos que se hayan cargado.

- La **Barra de información del fichero de datos**, que indicará si no se han cargado datos, si se han cargado pero no se han guardado o el nombre y extensión del fichero si se han guardado.
- La **Zona de información de las variables**, espacio que utiliza Gretl para informar sobre algunas de las características de las variables disponibles en la sesión de trabajo: identificador, nombre y etiqueta.
- La **Barra de información del tipo de datos y del rango muestral**, espacio utilizado para informar del tipo de datos (sin fecha, anuales, mensuales, ...) y del rango (completo y actual).
- La **Barra de Tareas**, que contiene los iconos que permiten acceder a las tareas más habituales de forma más operativa, sin necesidad de desplegar menús y submenús.

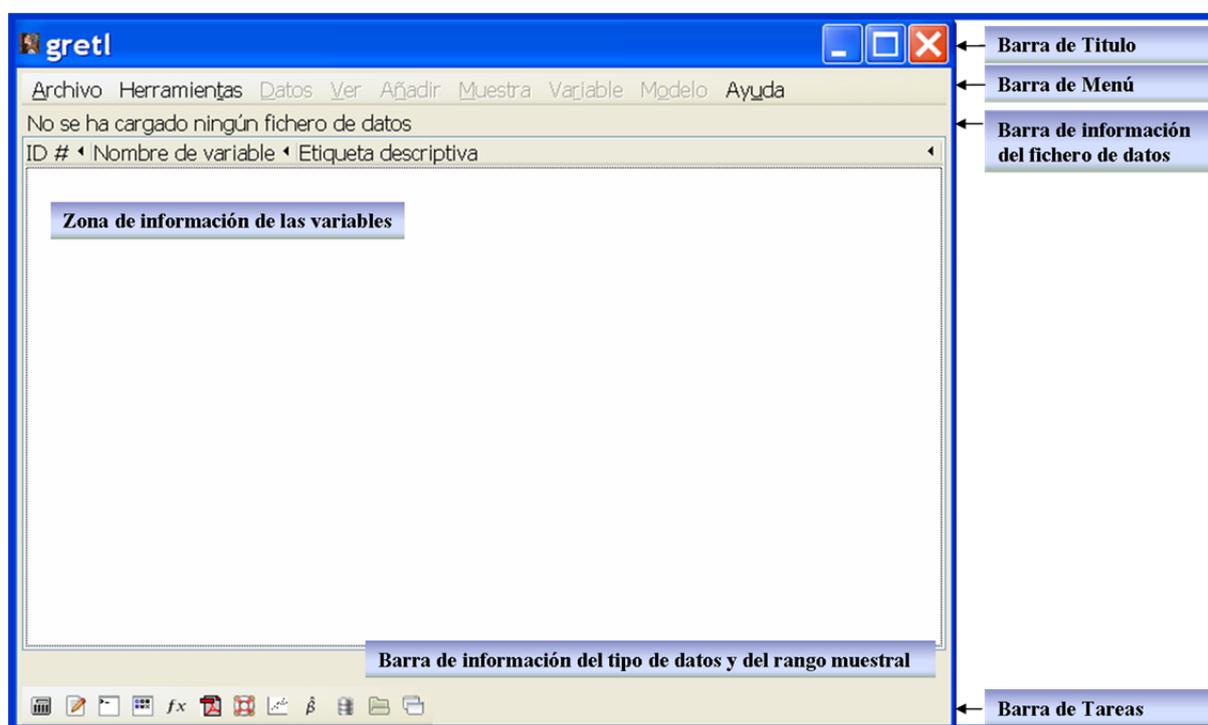


Ilustración 1-2. Ventana Principal del programa

Aunque al inicio de la sesión de trabajo la **Ventana Principal** está vacía y algunos botones desactivados (véase Ilustración 1-2), cuando se cargan datos, todos los botones pasan a estar operativos (véase Ilustración 1-3).



Ilustración 1-3. **Ventana Principal** con un ejemplo donde Gretl informa del fichero de datos (DATOS1.gdt), del rango muestral (1-20) y de algunas de las características de las variables disponibles (número de identificación, denominación y descripción).

El usuario debe ser consciente de que cerrar la **Ventana Principal** significa salir del programa y, por tanto, finalizar la sesión de trabajo.

### 1.3. ¿Cómo ejecutar Gretl?

Existen tres formas de ejecutar Gretl: por **menús**, modo **talk** o consola y modo **batch** o por lotes.

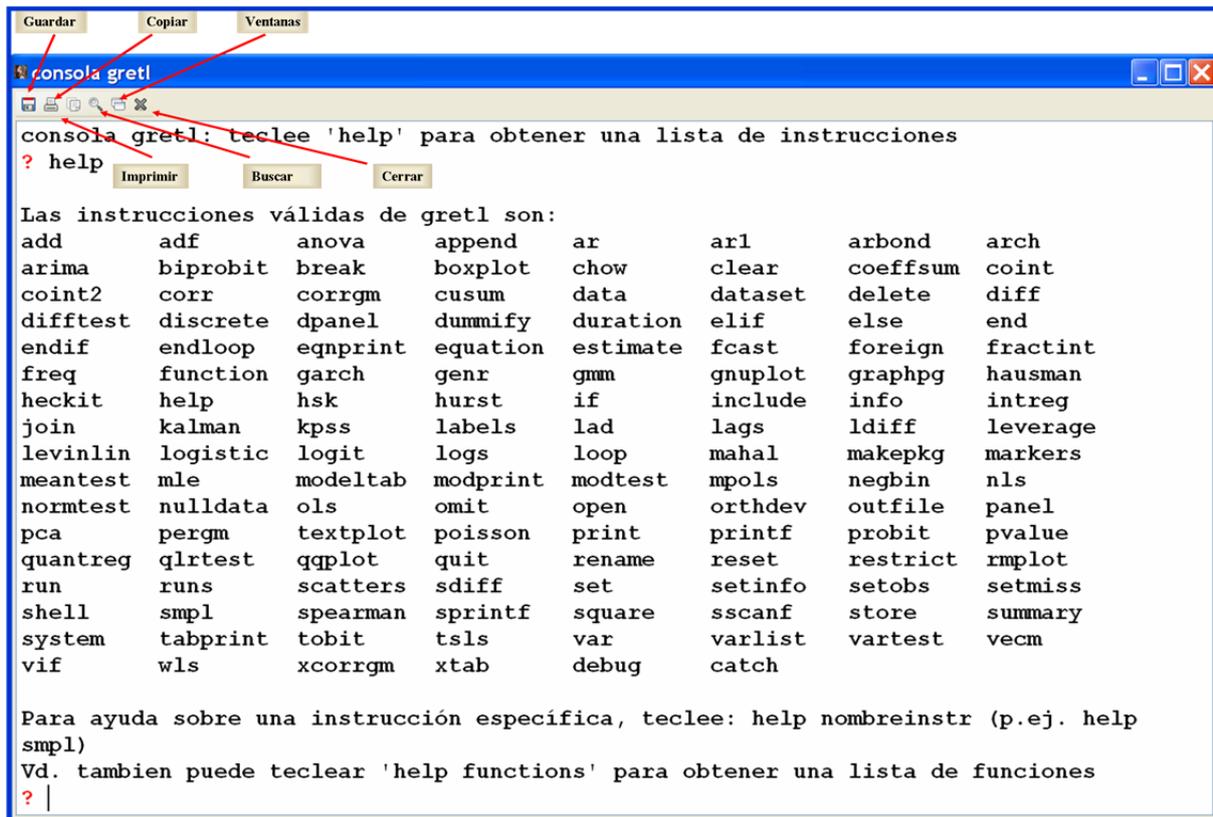


Ilustración 1-4. **Ventana Consola Gretl.**

Utilizando los **menús** se puede acceder a la mayoría de los procedimientos disponibles en Gretl. La ventaja de este modo de trabajo es que el usuario no necesita conocer los comandos ni su formato de escritura, pero su gran desventaja es que los resultados aparecen en ventanas diferentes, por lo que cuando se ha generado un número importante de ellas, su análisis y gestión pueda resultar engorroso.

En el modo **talk** se “habla” con Gretl. Se puede acceder a la **Consola Gretl** a través del menú **Herramientas** o a través del icono correspondiente de la **Barra de Tareas** de la **Ventana Principal**<sup>2</sup>. En este modo de trabajo es necesario teclear un comando de cada vez, es decir, las instrucciones se introducen una a una y Gretl responde presentando en la misma ventana y de forma inmediata los resultados del comando utilizado y un signo de cierre de interrogación (?) que indica que se puede teclear un nuevo comando. Además, en este modo de trabajo, Gretl señala los errores cometidos.

En el modo **batch** o por lotes se utiliza un archivo de guión o fichero de instrucciones que almacena los comandos y opciones que se desean ejecutar. Gretl ejecutará todos los comandos de una sola vez y mostrará la salida en una nueva ventana denominada “**resultados de guión**”.

El fichero de comandos<sup>3</sup> se puede crear directamente en Gretl a través de su editor, al que se puede acceder a través del menú **Archivo** (Archivo → archivo de guión → nuevo guión → guión Gretl) o a través del icono correspondiente de la **Barra de Tareas** de la **Ventana Principal**.

Para ejecutar un fichero de instrucciones existente hay dos opciones:

1. Abrir el fichero y ejecutarlo utilizando el icono “**Ejecutar**” de la ventana archivo de comandos<sup>4</sup>.
2. Abrir la **Consola Gretl** y ejecutar **run path:\nombre del fichero.inp**.

A pesar de que el modo de trabajo por menús resulta ineficiente cuando es necesario ejecutar muchos comandos, es el modo aconsejable para iniciarse en el manejo del programa.

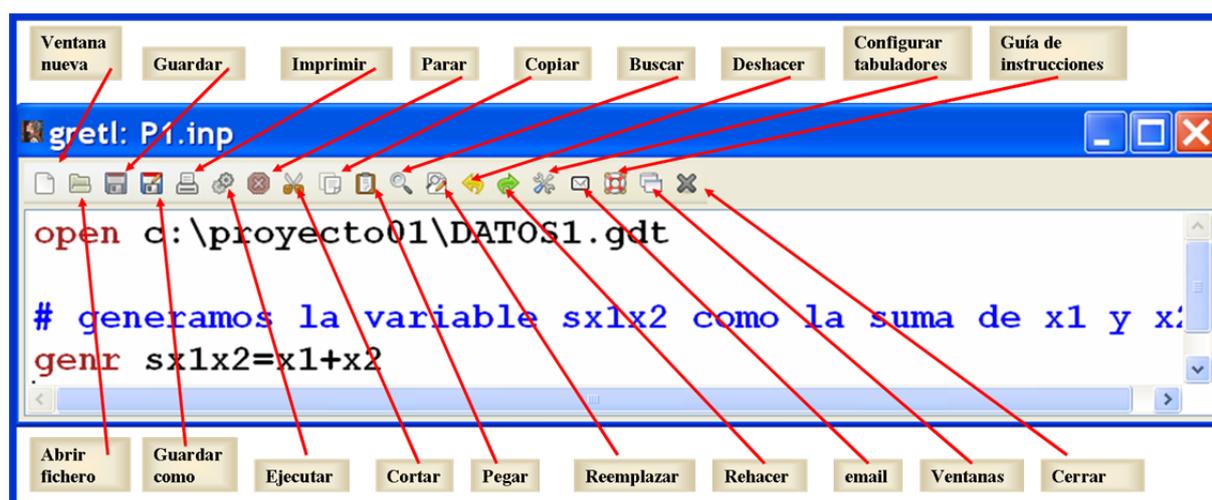


Ilustración 1-5. Ventana archivo de comandos.

## 1.4. Configuración de Gretl

La primera tarea a realizar es la configuración de Gretl de acuerdo con las exigencias de la investigación, de manera que su uso sea lo más cómodo posible<sup>5</sup>.

<sup>2</sup> En la Ilustración 1-4 se puede ver la **Ventana Consola Gretl** y su **Barra de Tareas**.

<sup>3</sup> Los ficheros de instrucciones o comandos tienen la extensión *.inp*.

<sup>4</sup> En la Ilustración 1-5 se puede ver la **Ventana archivo de comandos** y su **Barra de Tareas**.

<sup>5</sup> Antes de iniciar una sesión de trabajo parece conveniente, aunque no es estrictamente necesario, crear una carpeta de trabajo donde se almacenen los ficheros que se van creando en la sesión y configurar Gretl para que asuma dicha carpeta por defecto.

Para definir la ruta de la carpeta de trabajo por defecto, el usuario debe situarse en la **Barra de Menú** de la **Ventana Principal** y seleccionar **Archivo → Directorio de trabajo** y, en el cuadro de diálogo que se abre, utilizando el botón “**Revisar ...**”, seleccionar la carpeta de destino. Véase que en el ejemplo recogido en la Ilustración 1-6 se ha decidido que en la carpeta "proyecto01", se almacenen por defecto todos los ficheros (de datos, de comandos, de gráficos, ...)<sup>6</sup>.

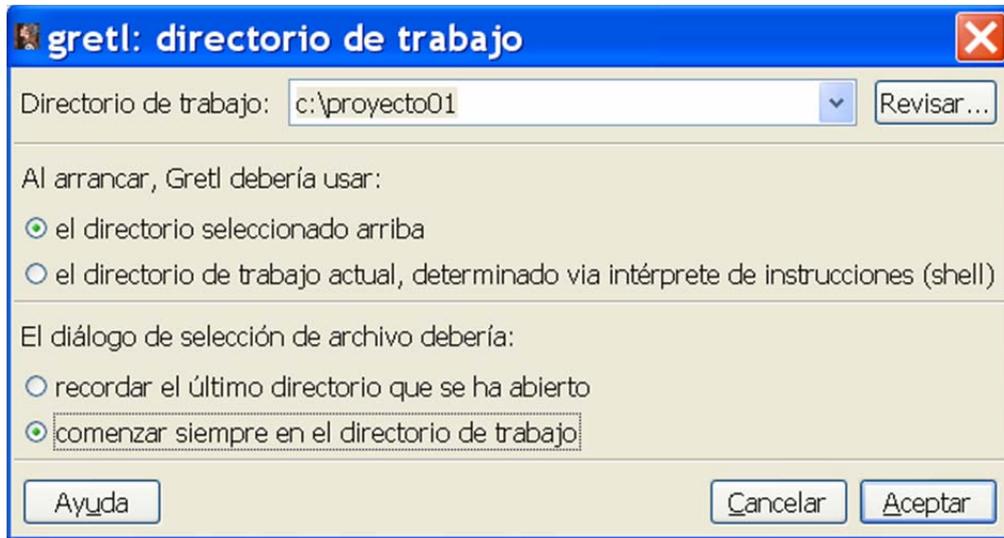


Ilustración 1-6. Cuadro de diálogo para definir el directorio de trabajo por defecto.

Con el botón **Herramientas** de la **Ventana Principal** y seleccionando la opción **Preferencias**, Gretl permite que el usuario establezca sus preferencias en relación a determinados aspectos (véase Ilustración 1-7). Uno de los aspectos que puede modificar es el tipo, estilo y tamaño de la fuente, cuestión importante para la presentación de resultados y para facilitar la accesibilidad a los usuarios con problemas de visión.

Además, permite que el usuario determine el idioma en el que quiere comunicarse con Gretl. En este caso se ha elegido el español, pero Gretl permite trabajar también en gallego, catalán, inglés, portugués, italiano, griego, chino, francés y alemán, entre otros.

Aparte de la elecciones anteriores, los usuarios poco familiarizados con el programa es adecuado que seleccionen:

- Preguntar si se desea guardar la sesión.
- Emular la apariencia de Windows.
- Mostrar automáticamente la vista de iconos.

<sup>6</sup> Para que la selección funcione debe existir la carpeta con el nombre indicado en el lugar especificado, sino es así, el usuario deberá crearla situándose en el directorio correspondiente y utilizando el botón derecho del ratón seleccionar **Nuevo → Carpeta**. Una vez creada la carpeta, se le asignará un nombre, en este caso, "proyecto01".

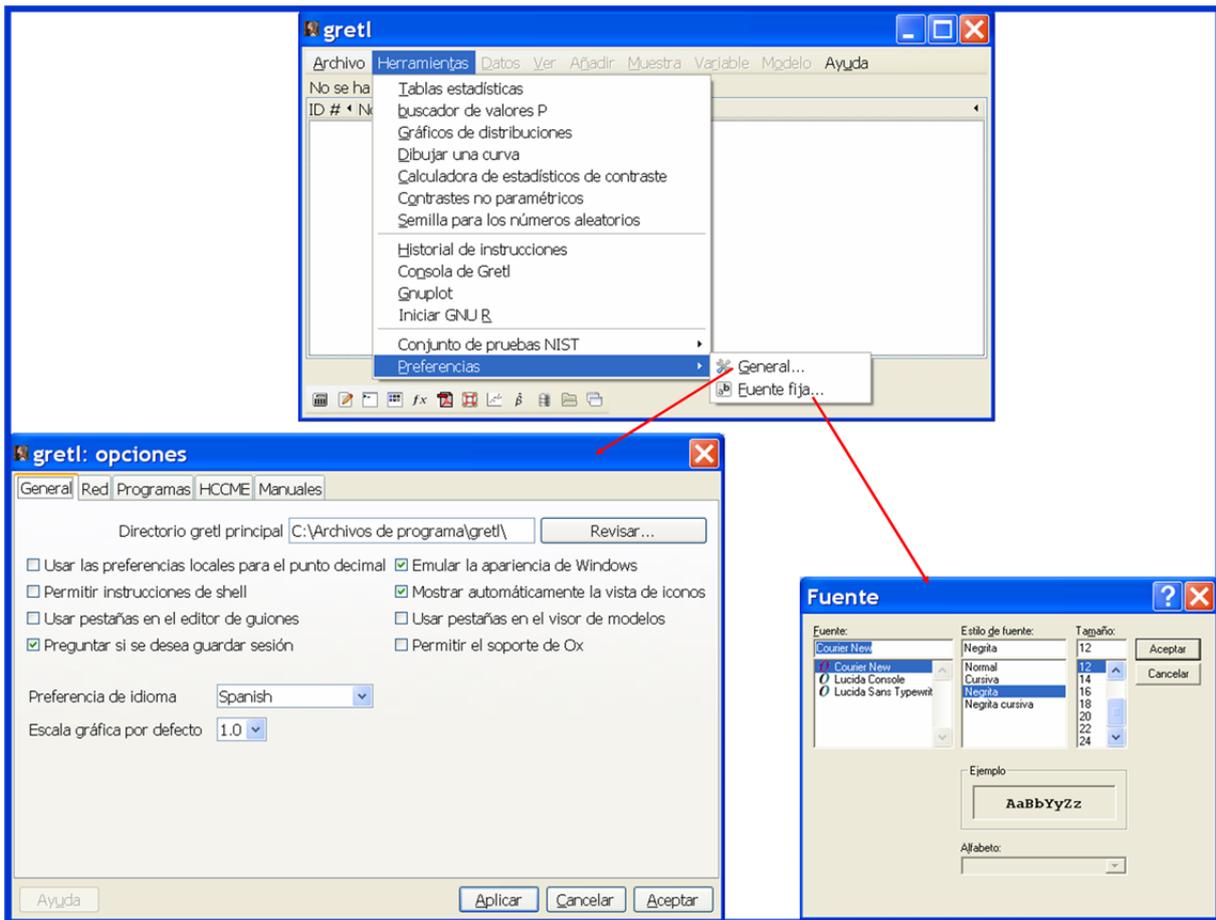


Ilustración 1-7. Configuración de preferencias.

## Capítulo 2. MANEJO Y ANÁLISIS DE DATOS

### 2.1. Presentación

Una tarea fundamental a la hora de abordar cualquier investigación aplicada es la recogida y depuración de datos. Adicionalmente, toda investigación económica implica trabajar con datos que proceden de situaciones reales y, por tanto, no controladas, lo que conlleva determinadas limitaciones a las que el investigador no debe ser insensible. Además, tampoco debe ser ajeno a los métodos de obtención y tratamiento de dichas magnitudes.

Antes de formular un modelo econométrico que intente explicar la realidad objeto de estudio, se debe realizar un análisis preliminar de las variables que se van a utilizar en la investigación, ya que dicho análisis puede arrojar luz sobre el comportamiento de esas variables y de las relaciones que pueda haber entre ellas.

Este capítulo tiene una doble finalidad, por una parte, familiarizar al usuario con el manejo de datos estadísticos mediante el programa Gretl y, por otra, examinar algunos instrumentos útiles para realizar un análisis preliminar de dichos datos.

### 2.2. Introducción de datos

La primera tarea con la que se debe enfrentar el usuario es la introducción de los datos de las variables seleccionadas para la investigación. Estos datos pueden estar disponibles en soporte papel o en soporte informático.

Cuando los datos estén en papel será necesaria su introducción tecleándolos, directamente utilizando el teclado o a través de la Consola Gretl. Si por el contrario, la información está almacenada en ficheros o en "bases de datos", dicha tarea se simplifica notablemente, pues basta con cargarlos.

#### 2.2.1. Introducción de datos directamente en el programa

La introducción directa de datos a través del teclado se puede realizar desde el **Menú Principal**, seleccionando **Archivo → Nuevo conjunto de datos** o utilizando el teclado (**Ctrl+N**).

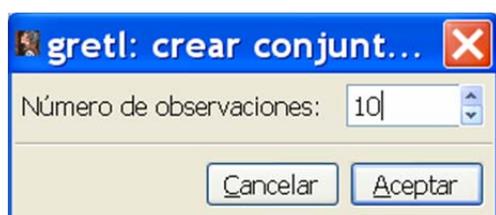


Ilustración 2-1. Cuadro de diálogo para dimensionar la base de datos.

Por ambas vías se abre el cuadro de diálogo “**crear conjunto de datos**”, en el que se debe introducir el número de observaciones de las variables (dimensión de la base de datos). En el ejemplo de la Ilustración 2-1 se ha optado por un tamaño “10”.

Definida la dimensión de la base de datos, se hace clic en “**Aceptar**” y se abre el cuadro de diálogo “**Organizador de estructura de datos**”, donde se debe elegir la

estructura del conjunto de datos. Debe de tenerse en cuenta que dicha elección determina los procedimientos o tareas que se pueden llevar a cabo<sup>7</sup>.

Como puede verse en la Ilustración 2-2, la información requerida por Gretl es diferente dependiendo de la estructura de datos elegida:

- Si se elige una estructura de sección cruzada es suficiente con indicar el número de observaciones.

<sup>7</sup> Por ejemplo, determinados gráficos (correlograma, periodograma, ...) sólo están disponibles si la base de datos es temporal.

- Si se elige una estructura de serie temporal, además de indicar el número de observaciones, será necesario establecer la frecuencia de los datos (anual, trimestral, semanal, ...) y la observación inicial.
- Si se elige una estructura de panel, además de indicar el número de observaciones, será necesario establecer el número de unidades de sección cruzada y el número de períodos temporales a los que se refieren los datos.

Una vez realizadas las elecciones oportunas, en el cuadro de diálogo “**Organizador de la estructura de datos**” se debe confirmar la estructura del conjunto de datos, ya que Gretl informa de la estructura seleccionada y pide confirmación. Si se ha cometido algún error al efectuar dicha elección, existe la posibilidad de ir hacia atrás y corregirlo antes de continuar. No obstante, a través del **Menú Principal**, seleccionando **Datos** → **Estructura del conjunto de datos**, Gretl permite acceder en cualquier momento al cuadro de diálogo “**Organizador de la estructura del conjunto de datos**” para modificar dicha estructura.

Cuando se confirma la estructura de la base de datos Gretl vuelve a la **Ventana Principal** del programa e informa:

1. De las variables iniciales de dicha base de datos:
  - 1.1. Una variable denominada “**const**”, generada por Gretl automáticamente, que toma siempre el valor uno a lo largo de la muestra. En terminología econométrica se le denomina **regresor ficticio**. El identificador que utiliza Gretl para esta variable es “**0**”.
  - 1.2. Una variable denominada “**index**” etiquetada por Gretl como “**variable índice**” que toma valores consecutivos a lo largo de la muestra empezando en 1. El identificador que utiliza Gretl para esta variable es “**1**”.
2. Del tipo de datos y del rango completo de la base de datos, informando de cuales son las observaciones inicial y final.
3. De que los datos no están guardados.

#### **2.2.1.1. ¿Cómo introducir datos?**

Generada la estructura de la base de datos se deben introducir los nombres y los datos de las variables que formarán parte de la misma.

Existen varias formas de acceder a la Ventana “editar datos”:

1. Seleccionando “empezar a introducir los valores de los datos” antes de confirmar la estructura de los datos en el último cuadro de diálogo del “**Organizador de estructura de datos**” (véase Ilustración 2-2).
2. A través del **Menú Principal**, seleccionando “**Añadir**” → “**Definir nueva variable**”.
3. En cualquiera de los casos se abre el cuadro de diálogo que permite dar nombre a la variable e introducir sus datos bajo la denominación “**nombre variable**” o “**añadir variable**”, respectivamente.

A cada una de las variables de la base de datos se le debe asignar un nombre, que será una combinación de letras y números no superior a treinta y un caracteres y que necesariamente comience con una letra<sup>8</sup>.

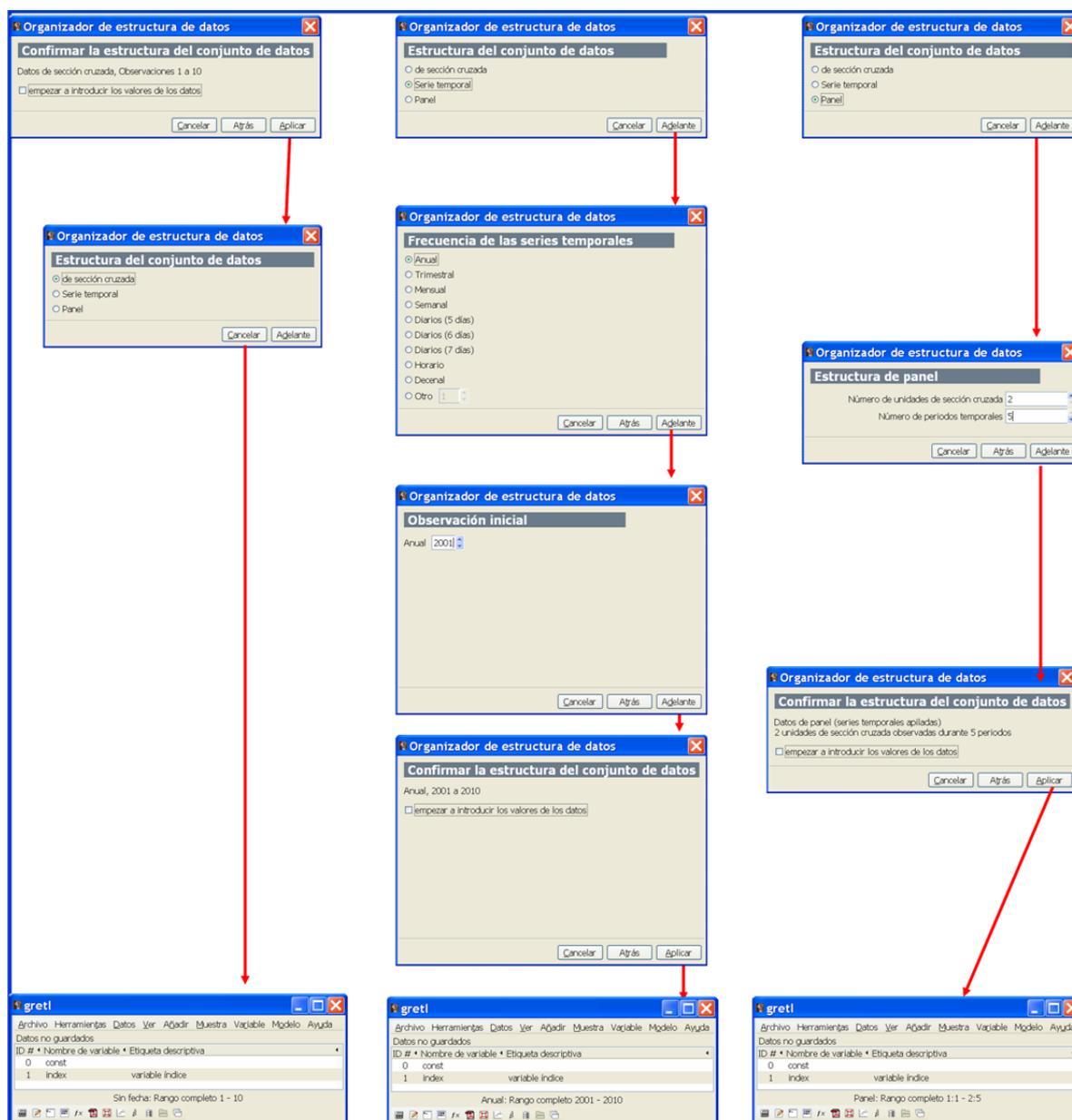


Ilustración 2-2. Cuadro de diálogo para seleccionar la estructura del conjunto de datos.

Una vez que se ha dado nombre a la variable y se hace clic en el botón “**Aceptar**” del cuadro de diálogo “añadir variable”, se accede a la ventana “editar datos” (véase Ilustración 2-3). Para la introducción de datos se debe situar el cursor en la celda correspondiente, teclear el valor de la variable

<sup>8</sup> En Econometría es frecuente nombrar a la variable dependiente con la letra Y y a la variable independiente con la letra X (en el caso, de que haya más de una variable dependiente o independiente, se les asignan subíndices numéricos para distinguirlas). Estas denominaciones aunque son ampliamente utilizadas, no aportan información descriptiva de las variables, por lo que un enfoque más eficiente sería asignar a cada variable un nombre que describa el tipo de datos que contiene.

para esa observación, situar el cursor en la celda siguiente y proceder de la misma forma hasta completar el valor para la última observación de la variable. Para moverse por las celdas se pueden utilizar las flechas “arriba” y “abajo” del teclado.

Una vez introducidos los datos de la variable, es conveniente hacer un chequeo de los mismos para comprobar que se han introducido correctamente. En el caso de que se detecte algún error, este se corregirá situando el cursor en la celda correspondiente, borrando el dato erróneo y tecleando el correcto.

Una vez se han introducido los datos, es necesario utilizar el botón “Aplicar” de la ventana “editar datos”, para que los cambios sean efectivos. No obstante, si en el momento de cerrar la ventana “editar datos”, los cambios no estuviesen guardados, Gretl preguntará al usuario si desea guardarlos.

El usuario puede editar las variables de una en una o hacerlo de forma conjunta (todas las variables). Además, a través del botón **Añadir** de esta ventana se pueden añadir nuevas variables (véase Ilustración 2-4).

### 2.2.2. Introducción de datos a través de la Consola Gretl y/o de los guiones o ficheros de comandos

Otra opción para introducir datos es utilizar la **Consola Gretl** y/o los **ficheros de comandos**, para lo cual el usuario debe conocer los comandos, su formato de escritura y las opciones disponibles.

Antes de introducir los datos es necesario determinar la dimensión de la base de datos, para lo cual se utiliza el comando **nulldata**, cuyo su formato es:

**nulldata** *número de observaciones*

```
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10
```

El siguiente paso es definir la estructura de los datos, es decir, determinar si se trata de datos atemporales, temporales o mixtos:

1. Por defecto Gretl define datos atemporales:

```
# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

? print index --byobs
```

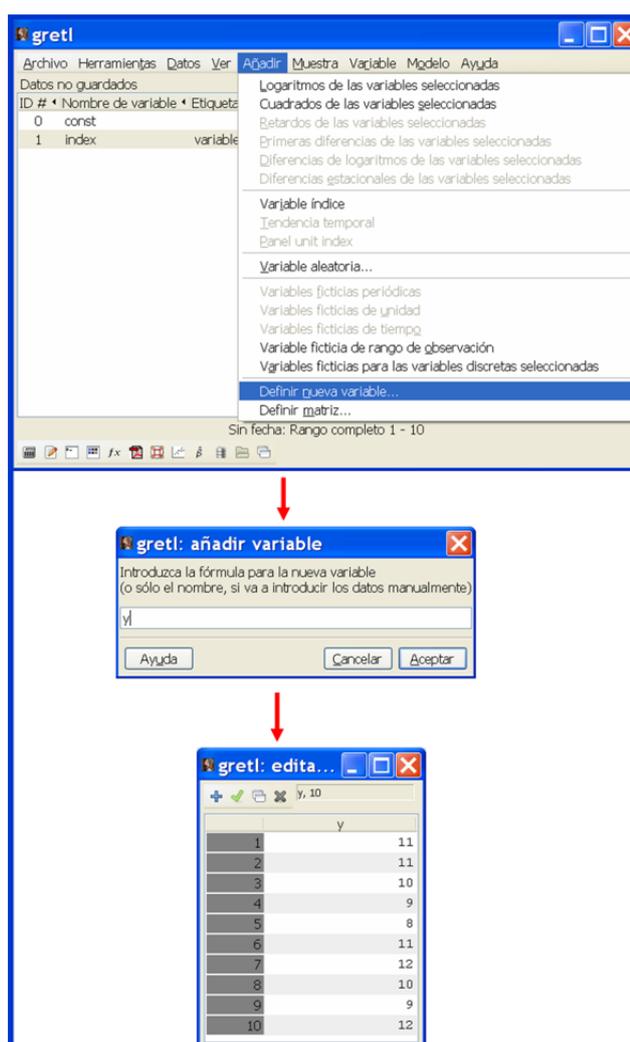


Ilustración 2-3. Secuencia para acceder a la Ventana editar datos utilizando el submenú **Añadir** del **Menú Principal** para introducir datos en Gretl.

```

index
1      1
2      2
3      3
4      4
5      5
6      6
7      7
8      8
9      9
10     10

# Para generar series atemporales es suficiente con definir la dimensión de la base de datos.
    
```

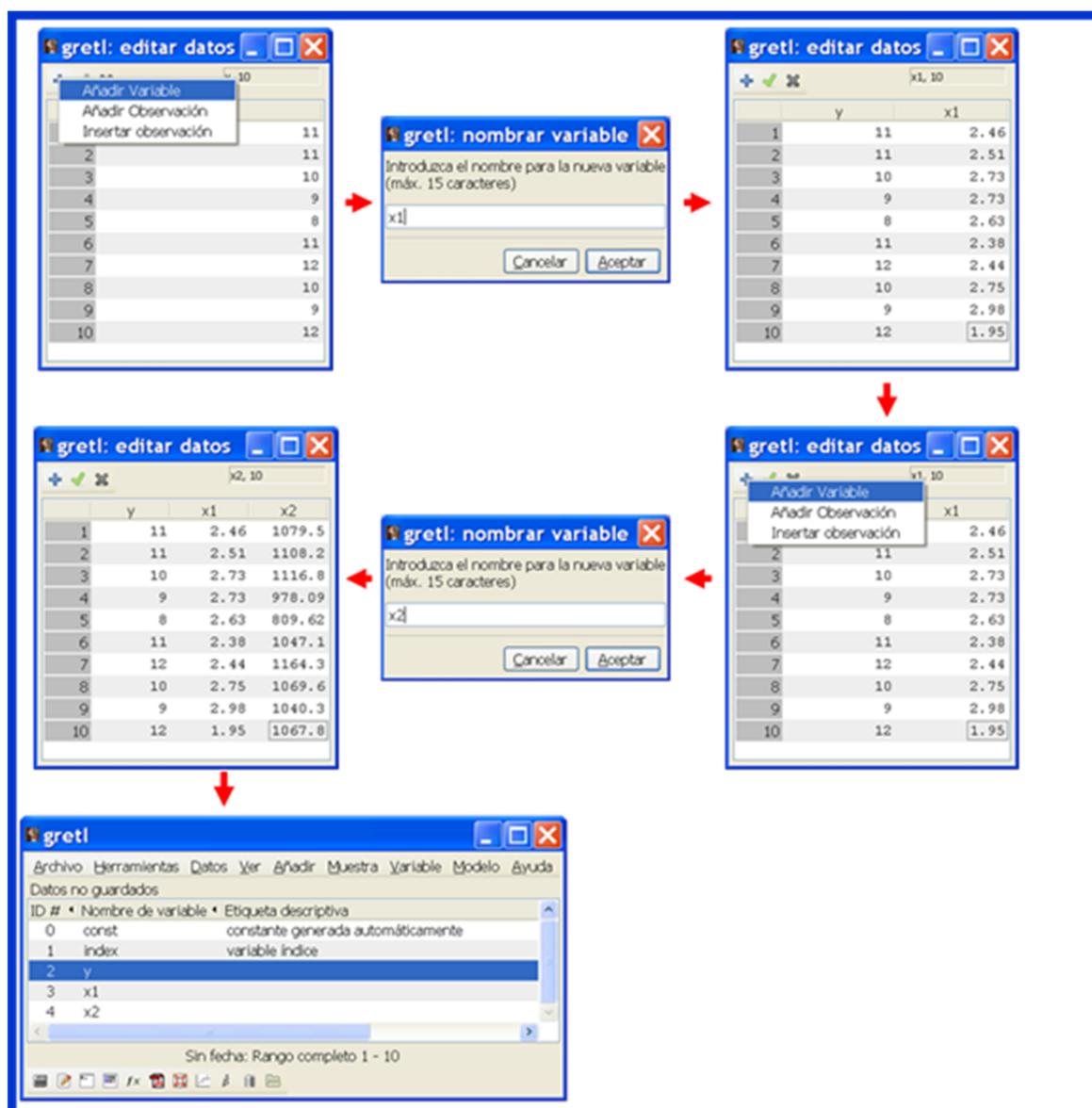


Ilustración 2-4. Añadir variables utilizando la Ventana editar datos.

2. Para datos temporales se utiliza el comando **setobs**, cuyo formato es:

```
setobs frecuencia fecha inicial --time-series
```

En las siguientes tablas se recogen algunos ejemplos de bases de datos temporales generadas utilizando comandos ejecutados desde la **Consola Gretl** y/o desde un **fichero o guión de comandos**:

```
# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir una serie temporal de datos anuales que empiece en el año 2001
? setobs 1 2001 --time-series
Rango de datos completo: 2001 - 2010 (n = 10)

? print index --byobs
      index
2001      1
2002      2
2003      3
2004      4
2005      5
2006      6
2007      7
2008      8
2009      9
2010     10

# Para generar series anuales se debe especificar una frecuencia igual a "1" seguida del primer año
de la serie (en este caso año 2001). Utilizando un comando print se puede observar a que año
corresponde cada observación (por ejemplo, la quinta observación corresponde al año 2005).
```

```
# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir una serie temporal de datos trimestrales que empiece en el primer trimestre del año 2001
? setobs 4 2001:1 --time-series
Rango de datos completo: 2001:1 - 2003:2 (n = 10)

? print index --byobs
      index
2001:1      1
2001:2      2
2001:3      3
2001:4      4
2002:1      5
2002:2      6
2002:3      7
2002:4      8
2003:1      9
2003:2     10

# Para generar series trimestrales se debe especificar una frecuencia igual a "4" seguida del primer
trimestre de la serie año:n°, donde n° varía desde 1 (primer trimestre) hasta 4 (cuarto trimestre).
Utilizando un comando print se puede observar que el primer dato de la serie hace referencia al
primer trimestre de 2001 y el último al segundo trimestre de 2003.
```

```
# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir una serie temporal de datos mensuales que empiece en el mes de enero del año 2001
? setobs 12 2001:01 --time-series
Rango de datos completo: 2001:01 - 2001:10 (n = 10)

? print index --byobs
      index
2001:01      1
2001:02      2
2001:03      3
2001:04      4
2001:05      5
2001:06      6
```

```

2001:07      7
2001:08      8
2001:09      9
2001:10     10

# Para generar series mensuales se debe especificar una frecuencia igual a "12" seguida del primer
mes de la serie año:n°, donde n° varia desde 01 (enero) hasta 12 (diciembre). Utilizando un comando
print se puede observar que el primer dato de la serie hace referencia al mes de enero de 2001 y el
último al mes de octubre de 2001.

```

```

# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir una serie temporal de datos semanales que empiece en la primera semana del año 2011
? setobs 52 2011/01/03 --time-series
Rango de datos completo: 2011/01/03 - 2011/03/07 (n = 10)

? print index --byobs
      index
2011/01/03      1
2011/01/10      2
2011/01/17      3
2011/01/24      4
2011/01/31      5
2011/02/07      6
2011/02/14      7
2011/02/21      8
2011/02/28      9
2011/03/07     10

# Para generar series semanales se debe especificar una frecuencia igual a "52" seguida de la fecha
del lunes de la primera semana de la serie año/n1°/n2°, donde n1° indica el mes y n2° la fecha
correspondiente del lunes de dicha semana. Utilizando un comando print se puede observar que el
primer dato de la serie hace referencia al primer lunes del mes de enero de 2011 (3/1/2011) y el
último al primer lunes del mes de marzo de 2011 (7/3/2011).

```

```

# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir una serie temporal de datos diarios de una "semana de 5 días" (de lunes a viernes) que
empiece en la primera semana del año 2011
? setobs 5 2011/01/03 --time-series
Rango de datos completo: 2011/01/03 - 2011/01/14 (n = 10)

? print index --byobs
      index
2011/01/03      1
2011/01/04      2
2011/01/05      3
2011/01/06      4
2011/01/07      5
2011/01/10      6
2011/01/11      7
2011/01/12      8
2011/01/13      9
2011/01/14     10

# Para generar series diarias de semanas de cinco días se debe especificar una frecuencia igual a
"5" seguida de la fecha del lunes de la primera semana de la serie año/n1°/n2°, donde n1° indica el
mes y n2° la fecha correspondiente a ese día de la semana. Utilizando un comando print se puede
observar que, en este caso, no aparecen en la serie las fechas correspondientes a los sábados y
domingos.

```

3. Para datos de panel se utiliza el comando **setobs**, cuyo formato es:

a. Para series temporales apiladas:

**setobs** *número de unidades temporales*  $n_1^o.n_2^o$  *--stacked-time-series*

```
# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir un panel de datos que contenga 2 unidades de periodos temporales apilados y que empiece en
la primera observación del primer panel.
? setobs 2 1.1 --stacked-time-series
Rango de datos completo: 1:1 - 5:2 (n = 10)

? print index --byobs
      index
1:1      1
1:2      2
2:1      3
2:2      4
3:1      5
3:2      6
4:1      7
4:2      8
5:1      9
5:2     10
```

b. Para series de corte transversal apiladas

**setobs** *número de unidades de sección cruzada*  $n_1^o.n_2^o$  *--stacked cross-section*

```
# Definir una base de datos con 10 observaciones
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir un panel de datos que contenga 2 unidades de sección cruzada apiladas y que empiece en la
primera observación del primer panel.
? setobs 2 1.1 --stacked-cross-section
Rango de datos completo: 1:1 - 2:5 (n = 10)

? print index --byobs
      index
1:1      1
1:2      3
1:3      5
1:4      7
1:5      9
2:1      2
2:2      4
2:3      6
2:4      8
2:5     10
```

Una vez definida la estructura de datos, se puede utilizar el comando **series** para introducir datos a través de la **Consola Gretl** y/o **fichero o guión de comandos**, cuyo formato es:

**series** *nombre de la variable* =  $\{d_1, d_2, \dots, d_T\}$

donde  $d_i$  son los valores numéricos que toma la variable a lo largo del rango muestral.

```
# Definir una serie atemporal
? nulldata 10
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

# Definir series
? series C = {11, 11, 10, 9, 8, 11, 12, 10, 9, 12}
Se ha generado la serie C (ID 2)

? series P = {2.46, 2.51, 2.73, 2.73, 2.63, 2.38, 2.44, 2.75, 2.98, 1.95}
```

```

Se ha generado la serie P (ID 3)

? series RF = {1079.5, 1108.2, 1116.8, 978.09, 809.62, 1047.1, 1164.3, 1069.6, 1040.3, 1067.8}
Se ha generado la serie RF (ID 4)

? print C P RF --byobs

```

	C	P	RF
1	11	2.46	1079.50
2	11	2.51	1108.20
3	10	2.73	1116.80
4	9	2.73	978.09
5	8	2.63	809.62
6	11	2.38	1047.10
7	12	2.44	1164.30
8	10	2.75	1069.60
9	9	2.98	1040.30
10	12	1.95	1067.80

### 2.2.3. ¿Cómo crear un fichero de datos?

Introducidos los datos de las variables, es necesario guardarlos en un fichero para poder trabajar con ellos en sesiones posteriores, puesto que si esto no se hace, el usuario se verá obligado a introducir de nuevo dichos datos. Mientras no se guarden los datos, una de las informaciones que Gretl muestra en su **Ventana Principal** es “*datos no guardados*” (véase parte inferior de la Ilustración 2-4).

Existen varias alternativas para guardar datos en un fichero:

- Utilizar el menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.
- Utilizar el comando **store**.

#### 2.2.3.1. Utilizar el menú Archivo

Existen varias alternativas para guardar datos utilizando el menú **Archivo**:

- Seleccionar la opción “**Guardar datos como**”, con lo que se guardarán en el fichero especificado las variables existentes en la sesión de trabajo (véase secuencia de cuadros que aparecen en la Ilustración 2-5). Por defecto, Gretl guarda el fichero de datos en el **directorio de trabajo**, por lo que si se desea guardarlo en una ubicación distinta sería necesario especificarla de una de las siguientes formas:
  - Teclar **path\nombre del fichero**.
  - Buscar la carpeta donde se quiere guardar utilizando el panel “**Lugares**” que aparece en la parte izquierda de dicho recuadro.

En el ejemplo recogido en la Ilustración 2-5, los datos se guardan en el fichero “datos1.gdt” situado en la carpeta “proyecto01”, que es la que se ha configurado en el capítulo anterior como **directorio de trabajo**. El usuario debe tener en cuenta que aunque no indique la extensión “.gdt”, es la que Gretl atribuye por defecto a los ficheros de datos.

- Seleccionar la opción “**Guardar datos**”. Cuando el fichero de datos se guarda por primera vez, Gretl despliega la misma secuencia de cuadros de diálogo que con la opción **Guardar datos como**, en caso contrario, reemplaza el fichero existente sin más.

La diferencia entre utilizar la opción “**Guardar datos**” y la opción “**Guardar datos como**” para un fichero existente, es que la primera “guarda” el archivo en la misma ubicación y con el mismo nombre, reemplazándolo automáticamente, mientras que la segunda permite almacenarlo con un nombre distinto y/o en una ubicación diferente si se desea.

Las opciones vistas hasta el momento permiten guardar los datos en formatos directamente reconocibles por el programa, pero hay que señalar que Gretl también permite guardarlos en otros

formatos con la opción **Exportar datos** del menú **Archivo** (csv, Gnu Octave, Gnu R, ...), lo que facilita el intercambio con otros paquetes de software.

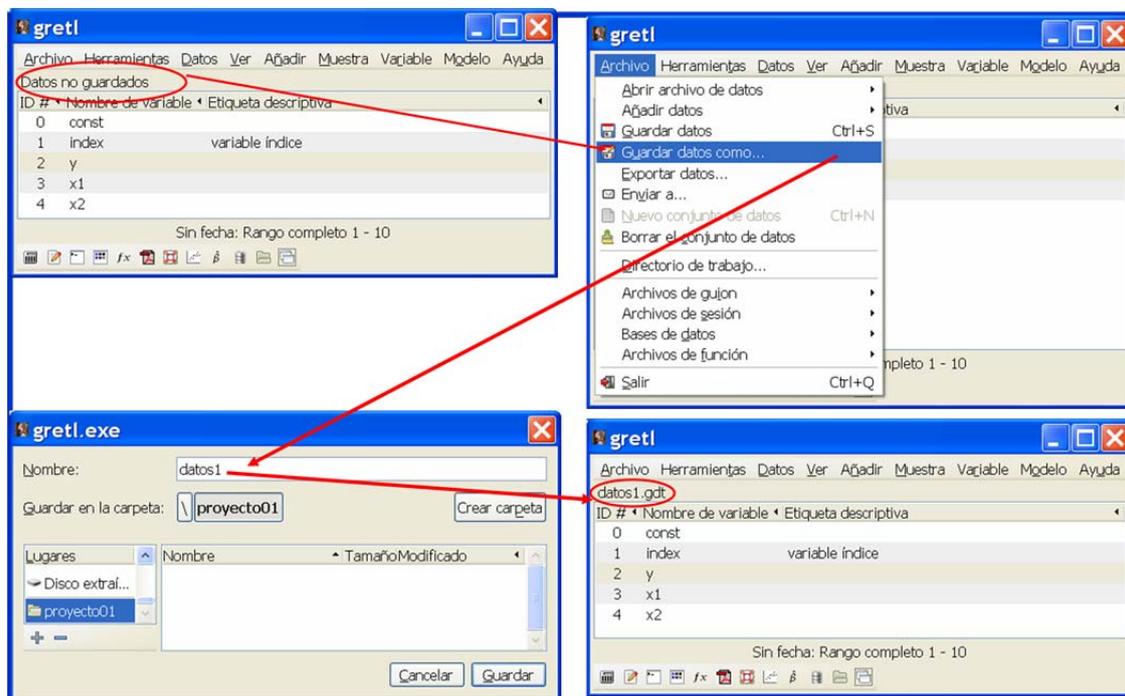


Ilustración 2-5. Guardar datos utilizando el menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.

Además, se puede utilizar la opción **Exportar datos** para seleccionar las variables a guardar, es decir, no se tienen que guardar forzosamente todas las variables disponibles en ese momento en la sesión de trabajo.

### 2.2.3.2. Utilizar el comando **store**

El formato del comando **store** es:

**store nombre del fichero.extensión v<sub>1</sub> v<sub>2</sub> ... v<sub>k</sub> --opciones**

En el Cuadro 2-1 se recogen algunas de las opciones disponibles con el comando **store**.

<b>--csv</b>	→ guarda los datos en formato csv.
<b>--gnu-octave</b>	→ guarda los datos en formato Gnu Octave.
<b>--gnu-R</b>	→ guarda los datos en formato Gnu R.
<b>--dat</b>	→ guarda los datos en formato PcGive ASCII.
<b>--gzipped</b>	→ aplica una comprensión gzip.
<b>--database</b>	→ guarda los datos en formato base de datos de Gretl.
...	

Cuadro 2-1. Algunas opciones del comando **store**.

Por ejemplo, para guardar las variables “index, C, P y RF” en un fichero de datos de Gretl denominado **dconsumo.gdt**, se tendrá que ejecutar el comando:

```
? store dconsumo.gdt index C P RF
wrote c:\proyecto01\dconsumo.gdt
```

Obsérvese que después de ejecutar el comando **store**, Gretl informa de que los datos se han guardado y señala la ruta o path de la carpeta donde se han guardado (la que se ha elegido por defecto como **directorio de trabajo**). Si se desea guardar el fichero en una carpeta diferente será necesario especificar el path o ruta de acceso a dicha carpeta:

```
? store c:\proyecto02\dconsumo.gdt index C
wrote c:\proyecto02\dconsumo.gdt
```

## 2.2.4. Recuperación de datos desde un fichero

La recuperación de datos desde ficheros o bases de datos es una tarea relativamente sencilla. Aunque son muchos los tipos de ficheros<sup>9</sup> de los que Gretl permite leer o importar datos, este capítulo se centrará únicamente en los ficheros de datos del propio programa (\*.gdt) y los ficheros de Excel (\*.xls).

### 2.2.4.1. Ficheros de datos de Gretl

Existen varias alternativas para recuperar datos desde un fichero de datos de Gretl:

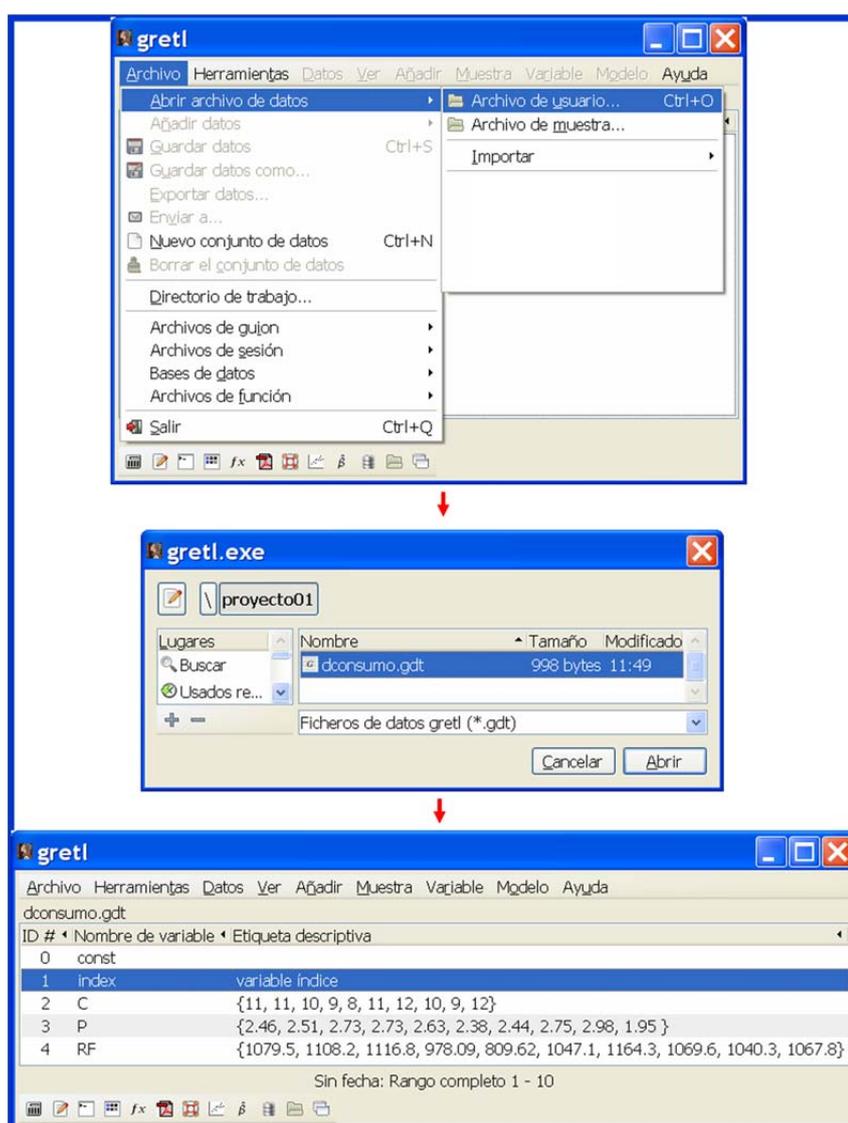


Ilustración 2-6. Recuperación de datos desde un fichero de datos de Gretl utilizando el menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.

- Utilizar el menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.

<sup>9</sup> Gretl permite importar datos de ficheros de datos de Excel, texto (csv), SAS, STATA y SPSS, entre otros.

- Utilizar el comando **open**.

#### 2.2.4.1.1. Utilizar el menú Archivo

Para recuperar datos desde un fichero de datos de Gretl, el usuario puede seguir los siguientes pasos (véase parte superior de la Ilustración 2-6):

- 1) Pinchar el botón **Archivo** de la **Barra de Menú** de la **Ventana Principal** y seleccionar **Abrir datos** → **Archivo de usuario**.
- 2) Seleccionar el fichero de datos y pinchar **Abrir**.

Los datos se cargan de forma automática en la sesión de trabajo y se puede ver su descripción en la Ventana Principal del programa (véase parte inferior de la Ilustración 2-6). Tal como puede observarse en el segundo recuadro, por defecto, Gretl busca el fichero de datos en el **directorio de trabajo** (en este caso **proyecto01**), por lo que si dicho fichero se encontrara en una ubicación distinta, sería necesario especificarla de alguna de las siguientes formas:

- Teclar **path\nombre del fichero.gdt**.
- Buscar la carpeta donde se encuentra utilizando el panel “**Lugares**” que aparece en la parte izquierda de dicho recuadro.

#### 2.2.4.1.2. Utilizar el comando open

El formato del comando **open** es:

**open nombre del fichero.gdt**

Por ejemplo, para recuperar los datos del fichero “dconsumo.gdt” se tendrá que ejecutar el comando:

```
? open dconsumo.gdt
Leer fichero de datos c:\proyecto01\dconsumo.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Listando 5 variables:
0) const    1) index    2) C          3) P          4) RF
```

Obsérvese que después de ejecutar el comando **open**, Gretl informa de la ruta o path de la carpeta desde donde se han recuperado los datos, de su periodicidad, del número máximo de observaciones, del rango muestral y de los nombres de las variables recuperadas (además de las variables “**const**” e “**index**” que se generan automáticamente).

Si se desea recuperar un fichero de una carpeta diferente será necesario indicar el path o ruta de acceso en el comando. Por ejemplo, para recuperar los datos del fichero “dconsumo.gdt”, ubicado en la carpeta “proyecto02”, se tendrá que ejecutar el comando:

```
? open c:\proyecto02\dconsumo.gdt
Leer fichero de datos c:\proyecto02\dconsumo.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Listando 5 variables:
0) const    1) index    2) C          3) P          4) RF
```

#### 2.2.4.2. Ficheros de datos de Excel

Existen varias alternativas para recuperar datos desde un fichero de Excel:

- Utilizar el menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.
- Utilizar el comando **open**.

### 2.2.4.2.1. Utilizar el menú Archivo

Para recuperar datos desde un fichero de Excel, el usuario puede seguir los siguientes pasos (véase Ilustración 2-7):

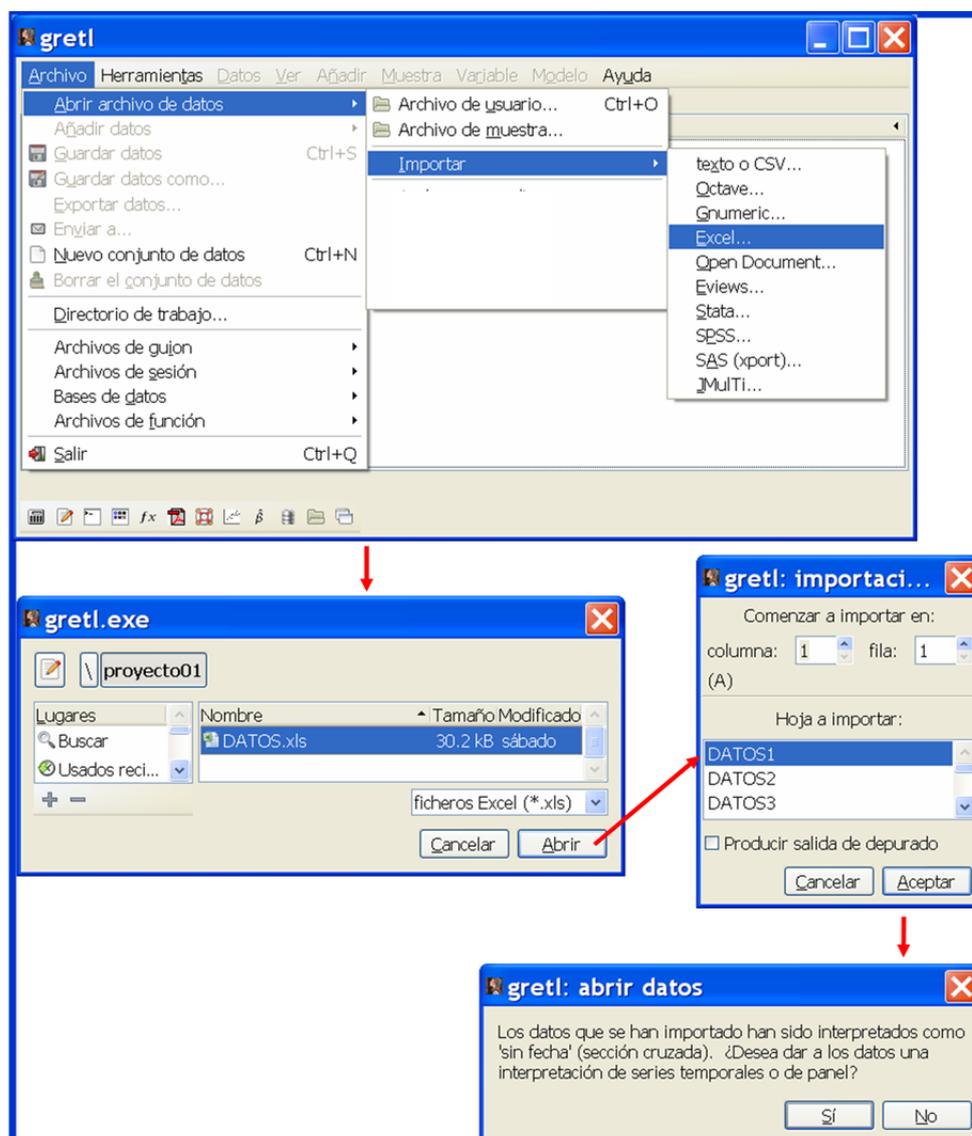


Ilustración 2-7. Recuperación de datos desde un fichero de datos de Excel utilizando el menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.

- 1) Pinchar en el botón **Archivo** de la **Barra de Menú** de la **Ventana Principal** y seleccionar **Abrir datos** → **Importar** → **Excel**.
- 2) Seleccionar el fichero y la hoja donde se encuentran los datos a importar<sup>10</sup>.

Antes de pinchar **Aceptar**, es conveniente indicar la columna y la fila en la que se empezará a importar los datos ya que, por defecto, Gretl empezará en la primera columna y fila.

El usuario debe tener en cuenta que Gretl:

<sup>10</sup> Un fichero de Excel puede tener múltiples hojas, por lo que será necesario especificar la hoja donde se encuentran los datos que se desean cargar.

- Importa los datos columna a columna.
- Sólo importa datos numéricos a excepción de los nombres de las variables que pueden encabezar las columnas de datos y deben empezar por una letra<sup>11</sup>, en caso contrario Gretl emitirá un mensaje de error: *“El primer carácter del nombre de variable ('...') no es válido, el primero debe ser alfabético. Por favor, cambie de nombre a esta variable e inténtelo de nuevo”*.  
Las denominaciones que superen los 31 caracteres serán recortadas por Gretl y los espacios en blanco y los caracteres especiales serán sustituidos por guiones bajos.
- En caso de encontrar información no numérica diferente al nombre de las variables, Gretl emite un mensaje de error *“se espera encontrar un dato numérico, pero se encontro una cadena ...”* y aconseja revisar y corregir el archivo de Excel.
- Las celdas vacías son interpretadas por Gretl como observaciones perdidas.

Por defecto, Gretl interpreta los datos importados como de corte transversal (véase el último cuadro de diálogo de la Ilustración 2-7). Para establecer una estructura diferente (panel de datos o series temporales) basta con seleccionar el botón “**Si**” de este cuadro de diálogo.

Tal como puede observarse en el segundo recuadro que aparece en la Ilustración 2-7, por defecto, Gretl busca el fichero de datos en el *directorio de trabajo* (en este caso *proyecto01*), por lo que si se encontrara en una ubicación distinta, sería necesario especificarla de alguna de las siguientes formas:

- Teclar *path\nombre del fichero.xls*.
- Buscar la carpeta donde se encuentra utilizando el panel “*Lugares*” que aparece en la parte izquierda de dicho recuadro.

#### 2.2.4.2.2. Utilizando el comando open

El formato del comando **open** es:

```
open nombre del fichero.xls --sheet=nº
```

Para una correcta importación de datos, es necesario ejecutar el comando **open** con la opción en la que se especifique la hoja del fichero en la que se encuentran los datos, puesto que en caso contrario se importarán de la primera hoja del fichero.

Por ejemplo, si se quieren recuperar los datos de la segunda hoja del fichero “*Datos.xls*”, se tendrá que ejecutar el comando:

```
? open datos.xls --sheet=2
Listando 5 variables:
  0) const   1) Y           2) X1        3) X2        4) X3

? store datos2.gdt
store: usando nombre de fichero c:\proyecto01\datos2.gdt
Los datos se han escrito correctamente.
```

Obsérvese, que después de ejecutar el comando **open**, Gretl informa del listado de variables que ha importado<sup>12</sup>.

<sup>11</sup> En el caso de que el fichero de Excel contenga sólo datos numéricos, Gretl denominará a las variables v1, v2, ....

<sup>12</sup> Una vez cargados los datos, puede resultar conveniente guardarlos en un fichero de datos de Gretl.

Si se desea recuperar un fichero de una carpeta diferente es necesario especificar su path o ruta de acceso en dicho comando. Por ejemplo, si se quieren recuperar los datos de la segunda hoja del fichero “*Datos.xls*” ubicado en la carpeta “proyecto02”, se tendrá que ejecutar el comando:

```
? open c:\proyecto02\datos.xls --sheet=2
Listando 5 variables:
0) const    1) Y        2) x1        3) x2        4) x3
```

El usuario debe tener en cuenta que cada vez que se abra un fichero de datos, se cerrará automáticamente el fichero de datos con el que se esté trabajando y se perderán los resultados que no se hayan guardado. Por tanto, si el usuario no desea que los nuevos datos sustituyan a los existentes, sino anexarlos a la sesión de trabajo, debe proceder a añadirlos de alguna de las siguientes formas:

- Utilizar el submenú *Añadir datos* del menú **Archivo** de la **Barra de Menú** de la **Ventana Principal**.
- Utilizar el comando **append**.

El formato del comando **append** es:

**append** *path\ nombre del fichero.gdt*

```
? open c:\proyecto01\datos1.gdt
Leer fichero de datos c:\proyecto01\datos1.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Listando 5 variables:
0) const    1) index    2) y        3) x1        4) x2

? open c:\proyecto01\dconsumo.gdt
Leer fichero de datos c:\proyecto01\dconsumo.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Listando 5 variables:
0) const    1) index    2) C        3) P        4) RF
```

En el ejemplo anterior los datos de las variables del fichero *dconsumo.gdt* sustituyen a las del fichero *datos1.gdt*.

```
? open c:\proyecto01\datos1.gdt
Leer fichero de datos c:\proyecto01\datos1.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Listando 5 variables:
0) const    1) index    2) y        3) x1        4) x2

? append c:\proyecto01\dconsumo.gdt
Leer fichero de datos c:\proyecto01\dconsumo.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Los datos se han añadido correctamente
Listando 8 variables:
0) const    1) index    2) y        3) x1        4) x2
5) C        6) P        7) RF
```

En el ejemplo anterior los datos de las variables del fichero *dconsumo.gdt* se añaden a las del fichero *datos1.gdt*.

### 2.3. ¿Cómo ver las variables?

Existen varias alternativas para visualizar las variables disponibles en una sesión de trabajo:

- Utilizar la opción **Mostrar valores** del menú **Variable** de la **Barra de Menú** de la **Ventana Principal**. Esta opción está disponible cuando se tenga seleccionada una única variable en la **Ventana Principal**.
- Utilizar la opción **Mostrar valores** del menú **Datos** de la **Barra de Menú** de la **Ventana Principal**, una vez seleccionada/s la/s variable/s en la **Ventana Principal**.
- Utilizar la opción **Mostrar valores** del menú emergente que aparece al hacer clic con el botón izquierdo del ratón sobre una/s variable/s seleccionada/s previamente en la **Ventana Principal**.
- Utilizar el comando **print**.

### 2.3.1. Utilizar el comando print

Para visualizar alguna/s variable/s disponible/s en una sesión de trabajo a través de la **Consola Gretl** o de un **Fichero de comandos** se puede utilizar el comando **print**, cuyo formato es:

```
print v1 v2 ... vk --opciones
```

Debe de tenerse en cuenta que el comando **print** no proporciona una salida de impresora sino una visualización en pantalla de las variables especificadas en el comando.

```
? print y x1 x2

y:
Rango de datos completo: 1 - 10 (n = 10)
11.0000 11.0000 10.0000 9.00000 8.00000 11.0000 12.0000 10.0000
9.00000 12.0000

x1:
Rango de datos completo: 1 - 10 (n = 10)
2.46000 2.51000 2.73000 2.73000 2.63000 2.38000 2.44000 2.75000
2.98000 1.95000

x2:
Rango de datos completo: 1 - 10 (n = 10)
1079.50 1108.20 1116.80 978.090 809.620 1047.10 1164.30 1069.60
1040.30 1067.80
```

Por defecto, el comando **print** muestra las variables una a una (formato horizontal). Si interesa mostrar las variables en formato vertical (observación a observación), se utilizará la opción **byobs**.

```
? print y x1 x2 --byobs

      y      x1      x2
1      11      2.46      1079.50
2      11      2.51      1108.20
3      10      2.73      1116.80
4       9      2.73      978.09
5       8      2.63      809.62
6      11      2.38      1047.10
7      12      2.44      1164.30
8      10      2.75      1069.60
9       9      2.98      1040.30
10     12      1.95      1067.80
```

## 2.4. ¿Cómo generar nuevas variables?

Existen varias alternativas para generar nuevas variables:

- Utilizar el menú **Añadir** de la **Barra de Menú** de la **Ventana Principal** (véase epígrafe 2.2.1.1).
- Utilizar el comando **genr**.
- Utilizar comandos que permitan generar varias variables a la vez: **logs**, **square**, **lags**, **diff**, **ldiff** y **sdiff**.

### 2.4.1. Utilizar el comando **genr**

Para crear nuevas variables se puede utilizar el comando **genr**, cuyo formato es:

**genr** *nueva variable* = *f(variables existentes)*

Si el resultado de la ecuación de cálculo utilizada en el comando **genr** es un escalar y el usuario desea definir una serie con todos los valores iguales entre sí, deberá utilizar el comando **genr series**<sup>13</sup>.

En el Cuadro 2-2 se recogen algunos de los operadores que pueden utilizarse con el comando **genr**.

#### I. Funciones:

- |  |  |
|--|--|
| (1) Valor absoluto de la variable $v \rightarrow \mathbf{abs}(v)$  | (1) Exponencial de la variable $v (e^v) \rightarrow \mathbf{exp}(v)$         |
| (1) Retardo de orden $p$ de la variable $v \rightarrow \mathbf{v}(-p)$   | (1) Adelanto de orden $p$ de la variable $v \rightarrow \mathbf{v}(+p)$      |
| (1) Logaritmo natural de la variable $v \rightarrow \mathbf{log}(v)$   | (1) Raíz cuadrada de la variable $v \rightarrow \mathbf{sqrt}(v)$            |
| (1) Máximo de la variable $v \rightarrow \mathbf{max}(v)$  | (1) Mínimo de la variable $v \rightarrow \mathbf{min}(v)$                    |
| (1) Media aritmética de la variable $v \rightarrow \mathbf{mean}(v)$   | (1) Mediana de la variable $v \rightarrow \mathbf{median}(v)$                |
| (1) Cuasivarianza de la variable $v \rightarrow \mathbf{var}(v)$   | (1) Cuasidesviación típica de la variable $v \rightarrow \mathbf{sd}(v)$     |
| (1) Cuasicovarianza entre las variables $v$ y $w \rightarrow \mathbf{cov}(v,w)$  | (1) Correlación entre las variables $v$ y $w \rightarrow \mathbf{corr}(v,w)$ |
| (1) Número de observaciones de la variable $v \rightarrow \mathbf{nobs}(v)$  | (1) Primeras diferencias de la variable $v \rightarrow \mathbf{diff}(v)$     |
| (1) Suma los valores de la variable $v$ para el rango muestral $\rightarrow \mathbf{sum}(v)$                                   |  |
| (1) Suma acumulada de los valores de la variable $v$ para el rango muestral $\rightarrow \mathbf{cum}(v)$                      |  |
| (1) Cuantil $p$ de la variable $v \rightarrow \mathbf{quantile}(v, p)$ donde $p$ será del 0.01, 0.02, ... ó 0.99               |  |
| (1) Retarda la variable $v$ $p$ períodos $\rightarrow \mathbf{lags}(v,p)$ proporcionando una lista de $p$ variables retardadas |  |
| (1) Suma de cuadrados de las desviaciones de la variable $v$ respecto a su media $\rightarrow \mathbf{sst}(v)$                 |  |
| (1) Ordena de menor a mayor los valores de la variable $v \rightarrow \mathbf{sort}(v)$  |  |
| (1) Ordena de mayor a menor los valores de la variable $v \rightarrow \mathbf{dsort}(v)$                                       |  |

#### II. Operadores matemáticos:

- |                                    |                              |                                    |
|------------------------------------|------------------------------|------------------------------------|
| (4) Suma $\rightarrow +$           | (4) Resta $\rightarrow -$    |                                    |
| (3) Multiplicación $\rightarrow *$ | (3) División $\rightarrow /$ | (2) Exponenciación $\rightarrow ^$ |

#### III. Relaciones lógicas:

- |                           |                                 |                                      |
|---------------------------|---------------------------------|--------------------------------------|
| (5) Igual $\rightarrow =$ | (5) Distinto $\rightarrow \neq$ | (5) Mayor o igual $\rightarrow \geq$ |
| (5) Mayor $\rightarrow >$ | (5) Menor $\rightarrow <$       | (5) Menor o igual $\rightarrow \leq$ |
- Si la relación propuesta se cumple la variable generada toma el valor uno, tomando el valor cero en caso contrario.

#### IV. Operadores lógicos:

- |                        |                          |                          |
|------------------------|--------------------------|--------------------------|
| (6) No $\rightarrow !$ | (6) Y $\rightarrow \&\&$ | (6) O $\rightarrow \ \ $ |
|------------------------|--------------------------|--------------------------|
- Si la relación propuesta se cumple la variable generada toma el valor uno, tomando el valor cero en caso contrario.

Cuadro 2-2 Algunos operadores disponibles con el comando **genr**.

Debe de tenerse en cuenta que estos operadores se ejecutan de acuerdo a su nivel de prioridad (se señala entre paréntesis con el número que antecede a cada operador) y, que entre operaciones de la misma prioridad, las expresiones se ejecutan de izquierda a derecha, pudiéndose utilizar tantos niveles de paréntesis como sea necesario.

Si se suma, resta, multiplica y divide  $x_1$  y  $x_2$ :

```
# Se genera la variable sx1x2 como suma de x1 y x2
? genr sx1x2 = x1 + x2
Se ha generado la serie sx1x2 (ID 5)
```

<sup>13</sup> En los contextos adecuados el comando **genr** es equivalente no sólo al comando **escalar** sino también al comando **matrix**.

```

# Se genera la variable dx1x2 como diferencia de x1 y x2
? genr dx1x2 = x1 - x2
Se ha generado la serie dx1x2 (ID 6)

# Se genera la variable px1x2 como producto de x1 y x2
? genr px1x2 = x1 * x2
Se ha generado la serie px1x2 (ID 7)

# Se genera la variable cx1x2 como cociente entre x1 y x2
? genr cx1x2 = x1 / x2
Se ha generado la serie cx1x2 (ID 8)

? print x1 x2 sx1x2 dx1x2 px1x2 cx1x2 --byobs

```

	x1	x2	sx1x2	dx1x2	px1x2	cx1x2
1	2.46	1079.50	1081.96	-1077.04	2655.570	0.002279
2	2.51	1108.20	1110.71	-1105.69	2781.582	0.002265
3	2.73	1116.80	1119.53	-1114.07	3048.864	0.002444
4	2.73	978.09	980.82	-975.36	2670.186	0.002791
5	2.63	809.62	812.25	-806.99	2129.301	0.003248
6	2.38	1047.10	1049.48	-1044.72	2492.098	0.002273
7	2.44	1164.30	1166.74	-1161.86	2840.892	0.002096
8	2.75	1069.60	1072.35	-1066.85	2941.400	0.002571
9	2.98	1040.30	1043.28	-1037.32	3100.094	0.002865
10	1.95	1067.80	1069.75	-1065.85	2082.210	0.001826

### Si se calcula el cuadrado, la raíz cuadrada, la exponencial y el logaritmo de $x_1$ :

```

# Se genera la variable cx1 como el cuadrado de x1
? genr cx1 = x1^2
Se ha generado la serie cx1 (ID 9)

# Se genera la variable rx1 como la raíz cuadrada de x1
? genr rx1 = sqrt(x1)
Se ha generado la serie rx1 (ID 10)

# Se genera la variable ex1 como la exponencial de x1
? genr ex1 = exp(x1)
Se ha generado la serie ex1 (ID 11)

# Se genera la variable lx1 como el logaritmo de x1
? genr lx1 = log(x1)
Se ha generado la serie lx1 (ID 12)

? print x1 cx1 rx1 ex1 lx1 --byobs

```

	x1	cx1	rx1	ex1	lx1
1	2.46	6.0516	1.568439	11.70481	0.900161
2	2.51	6.3001	1.584298	12.30493	0.920283
3	2.73	7.4529	1.652271	15.33289	1.004302
4	2.73	7.4529	1.652271	15.33289	1.004302
5	2.63	6.9169	1.621727	13.87377	0.966984
6	2.38	5.6644	1.542725	10.80490	0.867100
7	2.44	5.9536	1.562050	11.47304	0.891998
8	2.75	7.5625	1.658312	15.64263	1.011601
9	2.98	8.8804	1.726268	19.68782	1.091923
10	1.95	3.8025	1.396424	7.02869	0.667829

### Si se retarda y adelanta $x_1$ :

Si se utiliza  $v(-p)$ ,  $p$  indica el  $n^\circ$  de retardos, de manera que se estará retardando “ $p$ ” períodos la variable. Nótese que cuando se retarda una variable  $p$  períodos y las primeras  $p$  observaciones no están definidas, Gretl deja estas observaciones en blanco<sup>14</sup>.

<sup>14</sup> Las observaciones sin datos son tratadas por Gretl como “valores ausentes”, “valores perdidos” o “missing values”.

Si utiliza **v(+p)**, p indica el nº de adelantos, de manera que se estará adelantando “p” períodos la variable y, en este caso, son las p observaciones finales las que pueden no estar definidas, por lo que Gretl las deja en blanco.

```
# Se genera la variable r4x1 como la variable x1 retardada cuatro períodos
? genr r4x1 = x1(-4)
Se ha generado la serie r4x1 (ID 13)

# Se genera la variable r1x1 como el primer retardo de la variable x1
? genr r1x1 = x1(-1)
Se ha generado la serie r1x1 (ID 14)

# Se genera la variable ra4x1 como la variable x1 adelantada cuatro períodos
? genr ra4x1 = x1(+4)
Se ha generado la serie ra4x1 (ID 15)

# Se genera la variable ralx1 como la variable x1 adelantada un período
? genr ralx1 = x1(+1)
Se ha generado la serie ralx1 (ID 16)

? print x1 r4x1 r1x1 ra4x1 ralx1 --byobs
      x1      r4x1      r1x1      ra4x1      ralx1
1      2.46
2      2.51      2.46      2.38      2.73
3      2.73      2.51      2.44      2.73
4      2.73      2.73      2.75      2.63
5      2.63      2.46      2.73      2.98      2.38
6      2.38      2.51      2.63      1.95      2.44
7      2.44      2.73      2.38
8      2.75      2.73      2.44
9      2.98      2.63      2.75      1.95
10     1.95      2.38      2.98
```

**Si se calcula la suma de  $x_1$ :**

La función **sum(v)** genera un escalar igual a la suma de los valores de la variable para el rango muestral corriente.

```
# Se genera sumx1 como la suma de los valores de la variable x1 para el rango muestral
? genr sumx1 = sum(x1)
Se ha generado el escalar sumx1 = 25.56
```

Nótese que cuando el resultado del comando **genr** es un escalar, no será necesario utilizar el comando **print** para visualizarlo, ya que Gretl lo muestra de forma inmediata conjuntamente con el mensaje de que se ha generado el escalar.

**Si se calcula la suma acumulada de  $x_1$ :**

```
# Se genera cumx1 como la suma acumulada de los valores de la variable x1 para el rango muestral
? genr cumx1 = cum(x1)
Se ha generado la serie cumx1 (ID 17)

? print x1 sumx1 cumx1 --byobs
      x1      cumx1
1      2.46      2.46
2      2.51      4.97
3      2.73      7.70
4      2.73      10.43
5      2.63      13.06
6      2.38      15.44
7      2.44      17.88
8      2.75      20.63
9      2.98      23.61
10     1.95      25.56

      sumx1 = 25.560000
```

Obsérvese que el último elemento de la serie suma acumulada coincide con el sumatorio de la variable para el rango muestral.

### Si se genera una variable igual a un valor concreto de otra variable:

```
# Se genera la variable o2x1 igual a la segunda observación de la variable x1
? genr series o2x1 = x1[2]
Se ha generado la serie o2x1 (ID 18)

? print x1 o2x1 --byobs
      x1      o2x1
1      2.46      2.51
2      2.51      2.51
3      2.73      2.51
4      2.73      2.51
5      2.63      2.51
6      2.38      2.51
7      2.44      2.51
8      2.75      2.51
9      2.98      2.51
10     1.95      2.51
```

Obsérvese que para generar una variable que tome un valor concreto de otra variable es necesario incorporar en el comando **genr** “series”, puesto que en caso contrario se generará un escalar. Obsérvese que los valores de una variable se identifican por el nombre de la variable seguido de un número entre paréntesis que hace referencia a la observación correspondiente. En el ejemplo el número es dos, puesto que se trata de definir una variable que sea igual a la segunda observación de la variable  $x_1$  y que se ha denominado o2x1.

Para generar un escalar o una constante se puede utilizar el comando **scalar**. Por ejemplo, si se quiere definir una constante que contenga el valor de la segunda observación de  $x_1$ :

```
? scalar eo2x1=x1[2]
Se ha generado el escalar eo2x1 = 2.51
```

### Si se aplican las relaciones lógicas igual (=) y distinto ( $\neq$ ):

```
? genr v1 =(x1=2.73)
Se ha generado la serie v1 (ID 19)

? genr v2 =(x1!=2.73)
Se ha generado la serie v2 (ID 20)

? print x1 v1 v2 --byobs
      x1      v1      v2
1      2.46      0      1
2      2.51      0      1
3      2.73      1      0
4      2.73      1      0
5      2.63      0      1
6      2.38      0      1
7      2.44      0      1
8      2.75      0      1
9      2.98      0      1
10     1.95      0      1
```

Resulta conveniente que la relación lógica vaya entre paréntesis.

### Si se aplican las relaciones lógicas mayor o igual ( $\geq$ ), mayor ( $>$ ), menor o igual ( $\leq$ ) y menor ( $<$ ):

```
? genr v3 = (x1>=2.73)
Se ha generado la serie v3 (ID 21)

? genr v4 = (x1>2.73)
Se ha generado la serie v4 (ID 22)

? genr v5 = (x1<=2.73)
Se ha generado la serie v5 (ID 23)

? genr v6 = (x1<2.73)
Se ha generado la serie v6 (ID 24)

? print x1 v3 v4 v5 v6 --byobs
```

	x1	v3	v4	v5	v6
1	2.46	0	0	1	1
2	2.51	0	0	1	1
3	2.73	1	0	1	0
4	2.73	1	0	1	0
5	2.63	0	0	1	1
6	2.38	0	0	1	1
7	2.44	0	0	1	1
8	2.75	1	1	0	0
9	2.98	1	1	0	0
10	1.95	0	0	1	1

Si se aplican los operadores lógicos disyunción (“o”), conjunción (“y”) y negación (“no”):

```
? genr v7 = (x1=2.73) || (x1>=2.51)
Se ha generado la serie v7 (ID 25)

? genr v8 = (x1=2.73) && (x1>=2.51)
Se ha generado la serie v8 (ID 26)

? genr v9 = !((x1>2.46) && (x1<2.63)) && ((x2> 978) && (x2<1116))
Se ha generado la serie v9 (ID 27)

? print x1 x2 v7 v8 v9 --byobs
```

	x1	x2	v7	v8	v9
1	2.46	1079.50	0	0	1
2	2.51	1108.20	1	0	0
3	2.73	1116.80	1	1	1
4	2.73	978.09	1	1	1
5	2.63	809.62	1	0	1
6	2.38	1047.10	0	0	1
7	2.44	1164.30	0	0	1
8	2.75	1069.60	1	0	1
9	2.98	1040.30	1	0	1
10	1.95	1067.80	0	0	1

Obsérvese que si se utilizan los operadores lógicos y/o relaciones lógicas, cuando la relación propuesta se cumple, la variable generada tomará el valor uno y tomará el valor cero en caso contrario.

Si se aplica la función retardo de p períodos:

A diferencia de **v(-p)** donde se genera una única serie donde la variable v aparece retardada p períodos, **lags(p,v)**, genera una lista de p series retardadas donde la primera se corresponde con el primer retardo y la última con el retardo p de la variable v.

```
# Se genera x1sr2 como una lista de 2 series de retardos de x1
? genr x1sr2 = lags(2,x1)
Se ha generado la lista x1sr2

? print x1 x1sr2 --byobs
```

	x1	r1x1	x1_2
1	2.46		
2	2.51	2.46	
3	2.73	2.51	2.46
4	2.73	2.73	2.51
5	2.63	2.73	2.73
6	2.38	2.63	2.73
7	2.44	2.38	2.63
8	2.75	2.44	2.38
9	2.98	2.75	2.44
10	1.95	2.98	2.75

En este caso Gretl sólo genera una variable nueva (x1\_2), dado que la serie r1x1 que contiene el primer retardo de la variable ya ha sido generada anteriormente al utilizar la función **v(-1)**.

```
# Se genera x1sr4 como una lista de 4 series de retardos de x1
? genr x1sr4 = lags(4,x1)
Se ha generado la lista x1sr4

? print x1 x1sr4 --byobs
```

	x1	r1x1	x1_2	x1_3	r4x1
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

1	2.46				
2	2.51	2.46			
3	2.73	2.51	2.46		
4	2.73	2.73	2.51	2.46	
5	2.63	2.73	2.73	2.51	2.46
6	2.38	2.63	2.73	2.73	2.51
7	2.44	2.38	2.63	2.73	2.73
8	2.75	2.44	2.38	2.63	2.73
9	2.98	2.75	2.44	2.38	2.63
10	1.95	2.98	2.75	2.44	2.38

En este caso Gretl sólo genera una variable nueva ( $x1\_3$ ), dado que las demás series de la lista ya han sido definidas anteriormente.

### Si se aplica la función primeras diferencias:

```
# Se genera difly como las primeras diferencias de y
? genr difly = diff(y)
Se ha generado la serie difly (ID 30)

? print y difly --byobs
      y      difly
1      11         0
2      11         0
3      10        -1
4       9        -1
5       8        -1
6      11         3
7      12         1
8      10        -2
9       9        -1
10     12         3
```

Además, el usuario puede utilizar el comando **genr** para calcular algunos de los estadísticos más utilizados para el análisis descriptivo de variables: máximo, mínimo, media, cuasivarianza, mediana, cuantiles, ... .

```
# Se genera maxx1 como el máximo de x1
? genr maxx1 = max(x1)
Se ha generado el escalar maxx1 = 2.98

# Se genera minx1 como el mínimo de x1
? genr minx1 = min(x1)
Se ha generado el escalar minx1 = 1.95

# Se genera mediax1 como la media de x1
? genr mediax1 = mean(x1)
Se ha generado el escalar mediax1 = 2.556

# Se genera medianax1 como la mediana de x1
? genr medianax1 = median(x1)
Se ha generado el escalar medianax1 = 2.57

# Se genera cuantil1x1 como el primer cuartil de x1
? genr cuantil1x1 = quantile(x1,0.25)
Se ha generado el escalar cuantil1x1 = 2.425

# Se genera cuantil2x1 como el segundo cuartil de x1
? genr cuantil2x1 = quantile(x1,0.5)
Se ha generado el escalar cuantil2x1 = 2.57

# Se genera cuantil3x1 como el tercer cuartil de x1
? genr cuantil3x1 = quantile(x1,0.75)
Se ha generado el escalar cuantil3x1 = 2.735

# Se genera sctx1 como la suma de cuadrados de las desviaciones de x1 respecto a su media
? genr sctx1 = sst(x1)
Se ha generado el escalar sctx1 = 0.70644

# Se genera nobsx1 como el número de observaciones "no ausentes" de x1
? genr nobsx1 = nobs(x1)
Se ha generado el escalar nobsx1 = 10
```

```
# Se genera varx1 como la cuasivarianza de x1
? genr varx1 = var(x1)
Se ha generado el escalar varx1 = 0.0784933

# Se genera sdx1 como la cuasidesviación típica de x1
? genr sdx1 = sd(x1)
Se ha generado el escalar sdx1 = 0.280167

# Se genera covx1x2 como la cuasicovarianza entre x1 y x2
? genr covx1x2 = cov(x1,x2)
Se ha generado el escalar covx1x2 = -5.3369

# Se genera corx1x2 como la correlación entre x1 y x2
? genr corx1x2 = corr(x1,x2)
Se ha generado el escalar corx1x2 = -0.195431
```

Como se verá más adelante, algunos de estos estadísticos se pueden visualizar, por ejemplo, utilizando el comando **summary**:

```
? summary x1 x2
```

	Media	Mediana	Mínimo	Máximo
x1	2.5560	2.5700	1.9500	2.9800
x2	1048.1	1068.7	809.62	1164.3

	Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
x1	0.28017	0.10961	-0.72470	0.44031
x2	97.472	0.092996	-1.4868	1.7583

	Rango IQ	Observaciones ausentes
x1	0.31000	0
x2	85.602	0

### 2.4.2. Comandos para generar varias variables a la vez

Gretl ofrece la posibilidad de que con un solo comando se generen varias variables simultáneamente. Los formatos de los comandos disponibles para ello son:

- logs**  $v_1 v_2 \dots v_k \rightarrow$  calcula las series de los logaritmos naturales de las variables incluidas en el comando y las guarda bajo la denominación  $l\_v1, l\_v2, \dots, l\_vk$
- square**  $v_1 v_2 \dots v_k \rightarrow$  calcula las series de los cuadrados de las variables incluidas en el comando y las guarda bajo la denominación  $sq\_v1, sq\_v2, \dots, sq\_vk$
- lags p;**  $v_1 v_2 \dots v_k \rightarrow$  calcula **p** series retardadas de cada una de las variables incluidas en el comando y las guarda bajo la denominación  $v1\_1$  (primer retardo de  $v_1$ ),  $v1\_2$  (segundo retardo de  $v_1$ ), ...,  $v1\_p$  (p-ésimo retardo de  $v_1$ ),  $v2\_1$  (primer retardo de  $v_2$ ), ...,  $v2\_p$  (p-ésimo retardo de  $v_2$ ), ...
- diff**  $v_1 v_2 \dots v_k \rightarrow$  calcula las series de las primeras diferencias de las variables incluidas en el comando y las guarda bajo la denominación  $d\_v1, d\_v2, \dots, d\_vk$
- ldiff**  $v_1 v_2 \dots v_k \rightarrow$  calcula las series de las primeras diferencias de los logaritmos naturales de las variables incluidas en el comando y las guarda bajo la denominación  $ld\_v1, ld\_v2, \dots, ld\_vk$
- sdiff**  $v_1 v_2 \dots v_k \rightarrow$  calcula las series de las diferencias estacionales de las variables incluidas en el comando y las guarda bajo la denominación  $sd\_v1, sd\_v2, \dots, sd\_vk$

Si se aplican estos comandos a las variables  $x_1$  y  $x_2$ :

```
? open c:\proyecto01\datos1.gdt
Leer fichero de datos c:\proyecto01\datos1.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10
Listando 5 variables:
  0) const   1) index   2) y           3) x1          4) x2

? logs x1 x2
Listando 7 variables:
  0) const   1) index   2) y           3) x1          4) x2
```

```
5) l_x1    6) l_x2
```

Véase que se han generado dos nuevas variables denominadas `l_x1` y `l_x2`. Otra forma de generar las series de logaritmos naturales es seleccionar las variables en la **Ventana Principal** y pinchar **Logaritmos de las variables seleccionadas** del submenú **Añadir** de dicha ventana.

```
? square x1 x2
Listando 9 variables:
  0) const    1) index    2) y          3) x1         4) x2
  5) l_x1     6) l_x2     7) sq_x1     8) sq_x2
```

Véase que se han generado dos nuevas variables denominadas `sq_x1` y `sq_x2`. Otra forma de generar las series de cuadrados es seleccionar las variables en la **Ventana Principal** y pinchar **Cuadrados de las variables seleccionadas** del submenú **Añadir** de dicha ventana.

```
? lags 2; x1 x2
Listando 13 variables:
  0) const    1) index    2) y          3) x1         4) x2
  5) l_x1     6) l_x2     7) sq_x1     8) sq_x2     9) x1_1
 10) x1_2    11) x2_1    12) x2_2
```

Véase que se han generado cuatro nuevas variables denominadas `x1_1`, `x1_2`, `x2_1` y `x2_2`. Otra forma de generar las series de retardos es seleccionar las variables en la **Ventana Principal** y pinchar **Retardos de las variables seleccionadas** del submenú **Añadir** de dicha ventana.

```
? diff x1 x2
Listando 15 variables:
  0) const    1) index    2) y          3) x1         4) x2
  5) l_x1     6) l_x2     7) sq_x1     8) sq_x2     9) x1_1
 10) x1_2    11) x2_1    12) x2_2    13) d_x1     14) d_x2
```

Véase que se han generado dos nuevas variables denominadas `d_x1` y `d_x2`. Otra forma de generar las series de primeras diferencias es seleccionar las variables en la **Ventana Principal** y pinchar **Primeras diferencias de las variables seleccionadas** del submenú **Añadir** de dicha ventana.

```
? ldiff x1 x2
Listando 17 variables:
  0) const    1) index    2) y          3) x1         4) x2
  5) l_x1     6) l_x2     7) sq_x1     8) sq_x2     9) x1_1
 10) x1_2    11) x2_1    12) x2_2    13) d_x1     14) d_x2
 15) ld_x1   16) ld_x2
```

Véase que se han generado dos nuevas variables denominadas `ld_x1` y `ld_x2`. Otra forma de generar las series de primeras diferencias de los logaritmos naturales es seleccionar las variables en la **Ventana Principal** y pinchar en **Diferencias de logaritmos de las variables seleccionadas** del submenú **Añadir** de dicha ventana.

```
? setobs 4 2010:01
Rango de datos completo: 2010:1 - 2012:2 (n = 10)

? sdiff x1 x2
Listando 19 variables:
  0) const    1) index    2) y          3) x1         4) x2
  5) l_x1     6) l_x2     7) sq_x1     8) sq_x2     9) x1_1
 10) x1_2    11) x2_1    12) x2_2    13) d_x1     14) d_x2
 15) ld_x1   16) ld_x2   17) sd_x1    18) sd_x2
```

Véase que se han generado dos nuevas variables denominadas `sd_x1` y `sd_x2`. Otra forma de generar las series de diferencias estacionales es seleccionar las variables en la **Ventana Principal** y pinchar **Diferencias estacionales de las variables seleccionadas** del submenú **Añadir** de dicha ventana.

El usuario debe de tener en cuenta que no todos los comandos están disponibles con todas las estructuras de datos, por ejemplo, las diferencias estacionales, tan sólo están disponibles cuando se trabaja con series temporales no anuales.

Si no se está de acuerdo con la denominación que Gretl da a las variables creadas con estos comandos, se podrá cambiar el nombre de las mismas (véase epígrafe 2.6. ). Además, tal y como puede verse en la

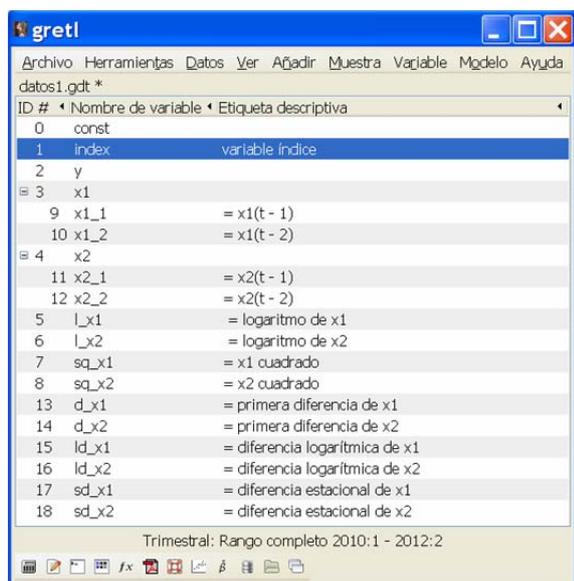


Ilustración 2-8. Autoetiquetado de variables.

Ilustración 2-8, para facilitar la identificación de las variables, Gretl proporciona una etiqueta que permite recordar como han sido generadas dichas variables.

## 2.5. ¿Cómo saber que variables están disponibles en la sesión de trabajo?

Existen varias alternativas para saber que variables están disponibles en la sesión de trabajo:

1. Consultar el espacio que el programa utiliza en la **Ventana Principal** para proporcionar información sobre el número de identificación (ID), la denominación y la descripción<sup>15</sup> de las variables disponibles en la sesión de trabajo.
2. Utilizar el menú **Ver** de la **Barra de Menú** de la **Ventana Principal**.
3. Utilizar el comando **varlist**.

### 2.5.1. Utilizar el menú Ver

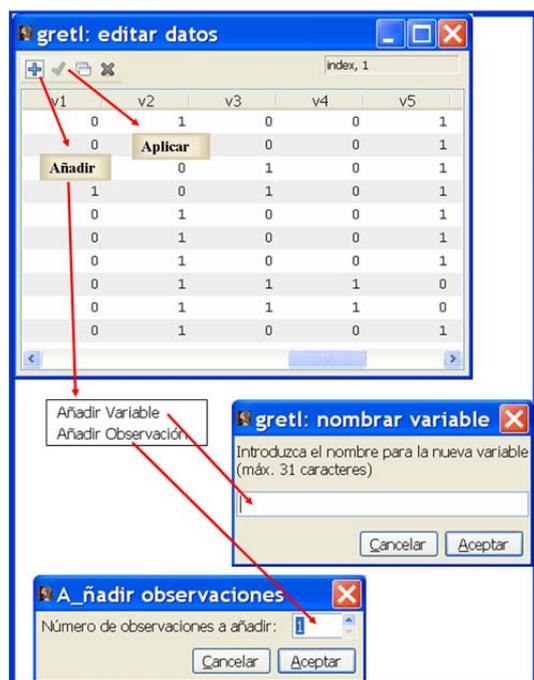


Ilustración 2-9. Editor del conjunto de datos de las variables disponibles en una sesión de trabajo.

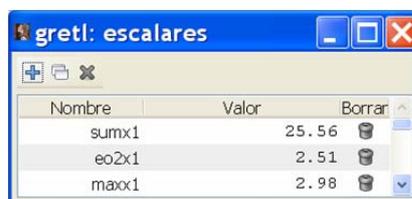


Ilustración 2-10. Editor del conjunto de escalares disponibles en una sesión de trabajo.

Utilizando el menú **Ver** se puede acceder al editor de la tabla de datos de los escalares y de las variables:

- Para acceder al editor de la tabla de datos de las variables, seleccionar **Vista de iconos** y en esta ventana, hacer un doble clic en el icono “**Conjunto de datos**”. Además, este editor permite modificar el conjunto de datos, añadiendo nuevas variables y/o observaciones e insertando o intercalando nuevas observaciones entre las existentes (véase Ilustración 2-9).
- Para acceder a la tabla de escalares, seleccionar **Vista de iconos** → **Escalares**. Además de permitir visualizar los escalares disponibles en la sesión de

<sup>15</sup> Si las variables han sido generadas a través de expresiones algebraicas, por defecto, en su descripción aparecerá la fórmula de cálculo.

trabajo, permite añadir nuevos escalares, borrar los existentes y modificar sus valores (véase Ilustración 2-10).

### 2.5.2. Utilizar el comando varlist

Si se desea información sobre las variables disponibles en una sesión de trabajo se utiliza el comando **varlist**, que muestra el nombre de las variables disponibles en la sesión de trabajo, bien porque hayan sido cargadas de un fichero o base de datos o porque hayan sido generadas en la sesión de Gretl.

En este caso:

```
? varlist
Listando 31 variables:
 0) const    1) index    2) y         3) x1        4) x2
 5) sx1x2    6) dx1x2    7) px1x2    8) cx1x2    9) cx1
10) rx1     11) ex1     12) lx1     13) r4x1    14) r1x1
15) ra4x1   16) ra1x1   17) cumx1   18) o2x1   19) v1
20) v2      21) v3      22) v4      23) v5      24) v6
25) v7      26) v8      27) v9      28) x1_2   29) x1_3
30) difly
```

La opción **--scalars** del comando **varlist** muestra el nombre de los escalares disponibles en la sesión.

```
? varlist --scalars
sumx1 = 25.56
eo2x1 = 2.51
maxx1 = 2.98
minx1 = 1.95
mediax1 = 2.556
medianax1 = 2.57
cuantil1x1 = 2.425
cuantil2x1 = 2.57
cuantil3x1 = 2.735
sctx1 = 0.70644
nobsx1 = 10
varx1 = 0.07849333333333333
sdx1 = 0.280166617092996
covx1x2 = -5.336895555555555
corx1x2 = -0.195431384903646
```

La opción **--accessors** del comando **varlist** muestra el nombre de las “variables” accesorias o temporales disponibles en la sesión de trabajo.

En el Cuadro 2-3 se recogen algunas de las variables temporales que proporciona Gretl, no relacionadas con la estimación de modelos.

<b>\$nobs</b>	→	escalar que indica el número de observaciones de la muestra corriente.
<b>\$nvars</b>	→	escalar que indica el número de variables de la base de datos actual (incluida la constante o regresor ficticio).
<b>\$pd</b>	→	escalar que indica la frecuencia o periodicidad de los datos.
<b>\$t1</b>	→	escalar que indica la primera observación del rango muestral.
<b>\$t2</b>	→	escalar que indica la última observación del rango muestral.
<b>\$datatype</b>	→	escalar que indica el tipo de datos cargados en la sesión de trabajo: 0 → no hay datos cargados. 1 → datos de corte transversal. 2 → datos temporales. 3 → datos de panel.
<b>\$windows</b>	→	escalar que indica si se trabaja en MS Windows (1) o no (0).
<b>\$version</b>	→	escalar que indica el código de la versión del programa (por ejemplo, el número 10914 indica que se está trabajando con la versión 1.9.14 de Gretl).
...		

Cuadro 2-3. Algunas variables temporales disponibles.

Dado que no se han realizado estimaciones, no existen variables temporales ligadas y Gretl emite el mensaje de que no dispone de variables relativas al modelo (se hará referencia a ellas en los capítulos correspondientes).

```
? varlist --accessors

relativo al modelo
ninguno

Otro
$noobs (scalar: 10)
$nvars (scalar: 31)
$spd (scalar: 1)
$t1 (scalar: 1)
$t2 (scalar: 10)
$datatype (scalar: 1)
$windows (scalar: 1)
$version (scalar: 10914)
...
```

En este caso Gretl informa de que se dispone de 10 observaciones, 31 variables, que los datos son de corte transversal, que el rango muestral comienza en la primera observación y termina en la décima, que se trabaja en Windows y con la versión 1.9.14 de Gretl.

## 2.6. ¿Cómo cambiar el nombre de las variables?

Existen varias alternativas para cambiar el nombre de las variables:

1. Utilizar el menú **Variable** de la **Barra de Menú** de la **Ventana Principal**.
2. Utilizar el comando **rename**.

### 2.6.1. Utilizar el menú Variable

Utilizando el menú **Variable** y seleccionando **Editar atributos**<sup>16</sup> se accede al cuadro de diálogo **atributos de variable** (véase Ilustración 2-11), en el que se puede modificar su nombre, número de identificación, descripción y atribuir una denominación distinta para cuando dicha variable aparezca en gráficos.



Ilustración 2-11. Ejemplo de como se pueden cambiar los atributos de una variable.

### 2.6.2. Utilizar el comando rename

Para renombrar variables se puede utilizar el comando **rename**, cuyo formato es:

```
rename nombre anterior nombre nuevo
```

Cuando se renombra una variable, Gretl ofrece un listado con la denominación de todas las variables, en el que se puede comprobar que la variable deja de aparecer con la denominación anterior y aparece con la nueva denominación.

```
? rename y vy
Listando 31 variables:
0) const 1) index 2) vy 3) x1 4) x2
5) sx1x2 6) dx1x2 7) px1x2 8) cx1x2 9) cx1
```

<sup>16</sup> También se puede acceder al editor de atributos de una variable seleccionando dicha variable en la **Ventana Principal** y presionando el botón izquierdo del ratón.

```

10) rx1    11) ex1    12) lx1    13) r4x1    14) r1x1
15) ra4x1 16) ra1x1 17) cumx1 18) o2x1    19) v1
20) v2     21) v3     22) v4     23) v5     24) v6
25) v7     26) v8     27) v9     28) x1_2   29) x1_3
30) difly

? rename vy y
Listando 31 variables:
  0) const  1) index  2) y      3) x1     4) x2
  5) sx1x2  6) dx1x2  7) px1x2  8) cx1x2  9) cx1
 10) rx1    11) ex1    12) lx1    13) r4x1    14) r1x1
 15) ra4x1 16) ra1x1 17) cumx1 18) o2x1    19) v1
 20) v2     21) v3     22) v4     23) v5     24) v6
 25) v7     26) v8     27) v9     28) x1_2   29) x1_3
 30) difly

```

## 2.7. ¿Cómo borrar variables?

Existen varias alternativas para borrar variables:

1. Para eliminar variables utilizando el teclado es suficiente con seleccionarlas en la **Ventana Principal** y presionar la tecla **Supr**. Para evitar borrados erróneos, antes de eliminar las variables definitivamente, Gretl pregunta si se está seguro de realizar dicho borrado.
2. Utilizar el comando **delete**.

### 2.7.1. Utilizar el comando delete

Para borrar variables se puede utilizar el comando **delete**, cuyo formato es:

```
delete v1 v2 ... vk
```

Si se desea borrar la variable index:

```

? delete index
Borrada index
Tome nota: se han reenumerado las variables
Listando 30 variables:
  0) const  1) y      2) x1     3) x2     4) sx1x2
  5) dx1x2  6) px1x2  7) cx1x2  8) cx1     9) rx1
 10) ex1    11) lx1    12) r4x1    13) r1x1    14) ra4x1
 15) ra1x1 16) cumx1 17) o2x1    18) v1     19) v2
 20) v3     21) v4     22) v5     23) v6     24) v7
 25) v8     26) v9     27) x1_2   28) x1_3   29) difly

```

Véase que cuando se borra/n alguna/s variable/s, Gretl ofrece un listado con el resto de las variables. Además, Gretl reenumera las variables, por lo que su número de identificación (ID) puede modificarse, cuestión que debe tenerse en cuenta en las instrucciones en las que figure dicho número de identificación.

En este ejemplo se ha borrado la variable “index” que tenía como ID “1” y eso ha supuesto que se hayan reenumerado los ID del resto de las variables.

Obsérvese que utilizando este comando se pueden borrar variables, escalares, matrices, series, etc.

## 2.8. ¿Cómo ordenar variables?

Utilizando la opción *Ordenar los datos* del menú **Datos** de la **Ventana Principal**, Gretl permite ordenar los datos de las variables en función de los valores crecientes o decrecientes de la variable que se haya elegido para realizar dicha ordenación (véase Ilustración 2-12).

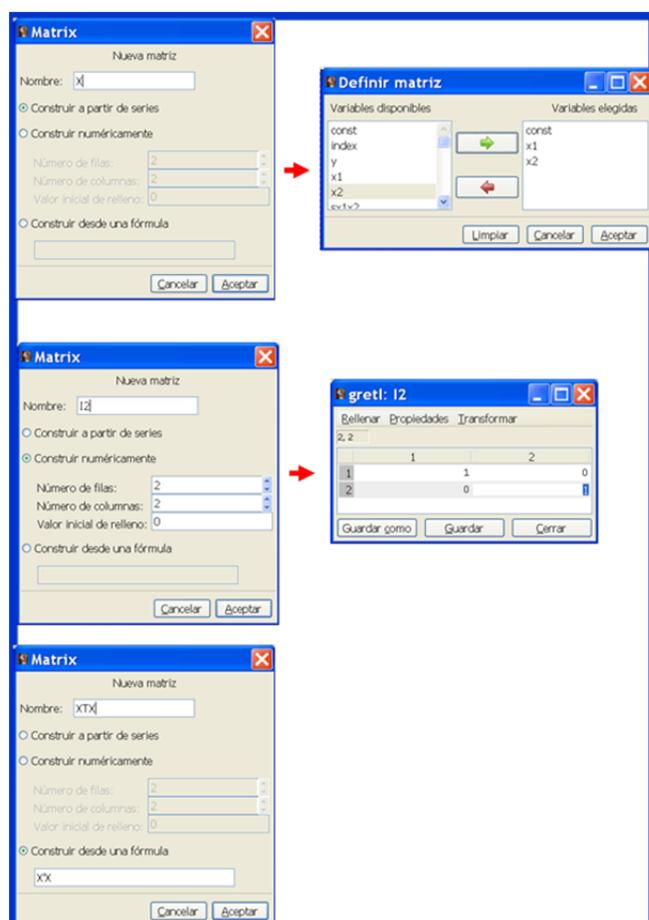


Ilustración 2-13. Ejemplos de generación de matrices utilizando el menú **Añadir**.

inicial de relleno, que podría ser modificado (véase ejemplo de la parte central de la Ilustración 2-13).

3. A partir de expresiones algebraicas, para lo que es necesario conocer los operadores disponibles en Gretl y las reglas de álgebra matricial (véase ejemplo de la parte inferior de la Ilustración 2-13).

Utilizando el menú emergente que aparece cuando en la ventana “*vista de iconos*” se selecciona el icono de la matriz correspondiente y se presiona el botón derecho del ratón, la matriz se puede visualizar, modificar, renombrar, borrar, ... (véase ejemplo de la Ilustración 2-14).

A continuación se hará una breve referencia a cada una de las tareas a las que se puede acceder a través de este menú:

**Ver** → permite visualizar el contenido de la matriz e informa de su dimensión. Además, informa del rango muestral utilizado si la matriz ha sido definida a partir de las variables existentes en la sesión de trabajo.



Ilustración 2-12. Ejemplo de ventana de ordenación de variables.

## 2.9. ¿Cómo generar vectores y matrices?

Existen varias alternativas para generar vectores y matrices:

1. Utilizar el menú **Añadir** de la **Barra de Menú** de la **Ventana Principal**.
2. Utilizar el comando **matrix**.

### 2.9.1. Utilizar el menú Añadir

Utilizando el menú **Añadir** de la **Barra de Menú** de la **Ventana Principal** se pueden definir matrices de tres formas:

1. A partir de variables existentes (véase ejemplo de la parte superior de la Ilustración 2-13).

2. Definiendo su dimensión y el valor

**Editar** → permite acceder al editor de matrices. Utilizando el botón **Rellenar** del editor de matrices, el usuario puede modificar sus valores situando el cursor en la celda correspondiente y tecleando el nuevo valor (para que estos cambios sean permanentes es necesario guardarlos antes de cerrar dicha ventana). El botón **Propiedades** de este editor, es otra forma de poder visualizar las propiedades de la matriz. El botón **Transformar** permite traspasar la matriz, calcular el producto de la traspuesta de la matriz por la matriz, multiplicar y dividir la matriz por un escalar, ....

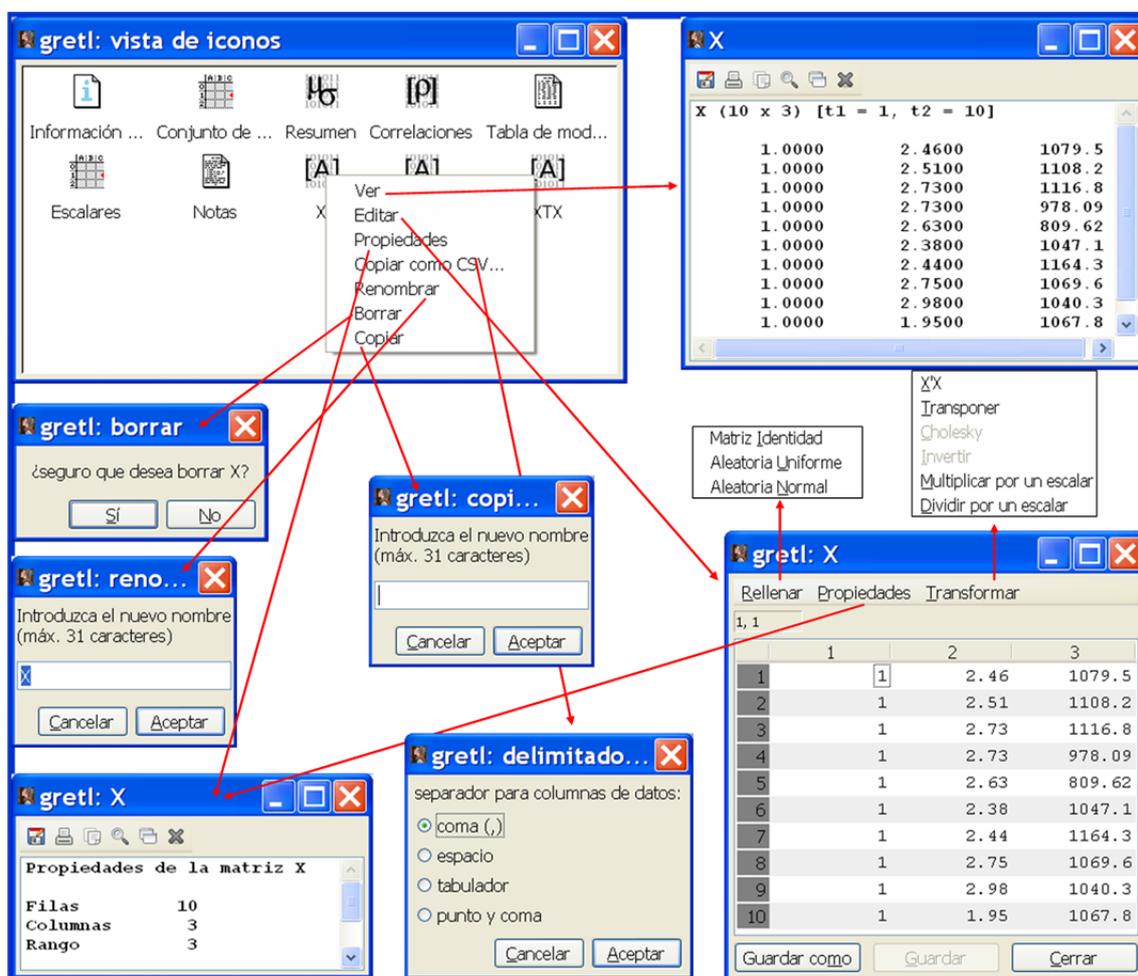


Ilustración 2-14. Menú que permite ver el contenido y las propiedades de una matriz, editarla, renombrarla y guardarla.

**Propiedades** → permite ver algunas “propiedades” o características de la matriz. Las “propiedades” varían en función del tipo de matriz. Así, para matrices cuadradas, además de informar de su dimensión, rango y norma, proporciona la traza, el determinante, los valores propios, si es simétrica, si es definida positiva y si es idempotente.

**Renombrar** → permite cambiar el nombre de la matriz. Debe tenerse en cuenta que cuando se renombra una matriz esta deja de existir con el nombre antiguo y pasa a estar disponible bajo la nueva denominación.

**Borrar** → permite borrar la matriz. Debe tenerse en cuenta que cuando se borra una matriz esta deja de estar disponible.

## 2.9.2. Utilizar el comando matrix

El comando **matrix** permite generar matrices a partir de:

1. Variables o series existentes en la sesión de trabajo:

**matrix** nombre nueva matriz = { $v_1, v_2, \dots, v_k$ }

donde  $v_1, v_2, \dots$  y  $v_k$  son las variables a partir de las cuales se define la matriz.

El número de filas de la matriz coincidirá con el rango muestral corriente y el número de columnas con el número de variables. Los datos de la primera columna de la matriz hacen referencia a la primera variable ( $v_1$ ), los de la segunda columna a la segunda variable ( $v_2$ ) y así sucesivamente.

```
? matrix X = {const, x1, x2}
Se ha generado la matriz X

? print X
X (10 x 3) [t1 = 1, t2 = 10]
  1.0000    2.4600    1079.5
  1.0000    2.5100    1108.2
  1.0000    2.7300    1116.8
  1.0000    2.7300    978.09
  1.0000    2.6300    809.62
  1.0000    2.3800    1047.1
  1.0000    2.4400    1164.3
  1.0000    2.7500    1069.6
  1.0000    2.9800    1040.3
  1.0000    1.9500    1067.8

? smpl 2 4
Rango de datos completo: 1 - 10 (n = 10)
Muestra actual: 2 - 4 (n = 3)

? matrix A = {const, x1, x2}
Se ha generado la matriz A

? print A
A (3 x 3) [t1 = 2, t2 = 4]
  1.0000    2.5100    1108.2
  1.0000    2.7300    1116.8
  1.0000    2.7300    978.09

? smpl 1 10
Rango de datos completo: 1 - 10 (n = 10)
```

Obsérvese que dado que las matrices X y A se han generado a partir de las series const,  $x_1$  y  $x_2$  existentes en la sesión de trabajo, Gretl además de informar de la dimensión de la matriz, informa del rango muestral utilizado para generarla y que en el caso de la matriz X es de la primera observación ( $t1=1$ ) hasta la décima ( $t2=10$ ) y que en el caso de la matriz A es de la observación segunda ( $t1=2$ ) hasta la cuarta ( $t2=4$ ). Se puede comprobar que el número de filas de la matriz A (3) coincide con el tamaño del rango muestral vigente en el momento de su creación y el número de columnas (3) coincide con el número de variables que han intervenido en su generación.

2. A partir de una lista de variables:

**matrix** nombre nueva matriz = {nombre de la lista de variables}

```
? genr list xlist = x1 x2
Se ha generado la lista xlist

? print xlist --byobs
      x1      x2
1      2.46    1079.50
2      2.51    1108.20
3      2.73    1116.80
4      2.73     978.09
5      2.63     809.62
6      2.38    1047.10
7      2.44    1164.30
8      2.75    1069.60
9      2.98    1040.30
10     1.95    1067.80

? matrix X={const, xlist}
Se ha reemplazado la matriz X
```

```
? print X
X (10 x 3) [t1 = 1, t2 = 10]
  1.0000    2.4600    1079.5
  1.0000    2.5100    1108.2
  1.0000    2.7300    1116.8
  1.0000    2.7300    978.09
  1.0000    2.6300    809.62
  1.0000    2.3800    1047.1
  1.0000    2.4400    1164.3
  1.0000    2.7500    1069.6
  1.0000    2.9800    1040.3
  1.0000    1.9500    1067.8
```

En el argumento del comando matrix pueden intervenir simultáneamente variables y series, siempre que tengan la misma longitud.

Obsérvese que las dos últimas columnas de la matriz X se corresponden con las variables que forman la lista “xlist” y la primera se corresponde con la variable **const**.

3. Introduciendo los valores numéricos de los elementos de la matriz:

**matrix** nombre nueva matriz = {  $v_{11}, v_{12}, \dots, v_{1p}; v_{21}, v_{22}, \dots, v_{2p}; \dots; v_{n1}, v_{n2}, \dots, v_{np}$  }

donde  $v_{ij}$  es el valor numérico correspondiente al elemento que ocupa la fila  $i$ -ésima y la columna  $j$ -ésima, el subíndice  $i$  varía desde 1 hasta  $n$  (siendo  $n$  el número de filas de la matriz) y el subíndice  $j$  varía desde 1 hasta  $p$  (siendo  $p$  el número de columnas de la matriz).

Los datos numéricos, correspondientes a los elementos de la matriz, se deben introducir por filas y separados por una “coma”. Para separar unas filas de otras se debe utilizar un “punto y coma”.

```
? matrix C = {1, 2, 3; 4, 5, 6; 7, 8, 9}
Se ha generado la matriz C

? print C
C (3 x 3)
  1  2  3
  4  5  6
  7  8  9

? matrix dt = {1, 2, 3, 4, 5, 6, 7, 8, 9}
Se ha generado la matriz dt

? print dt
dt (1 x 9)
  1  2  3  4  5  6  7  8  9

? matrix d = {1; 2; 3; 4; 5; 6; 7; 8; 9}
Se ha generado la matriz d

? print d
d (9 x 1)
  1
  2
  3
  4
  5
  6
  7
  8
  9
```

4. A partir de matrices<sup>17</sup> existentes en la sesión de trabajo:

**matrix** nombre nueva matriz = f(matrices existentes)

En el Cuadro 2-4 aparecen recogidos algunos de los operadores que pueden utilizarse con el comando **matrix**.

Suma → +	Resta → -	Traspuesta → ' ,
Multiplicación, filas por columnas → *	Multiplicación, elemento a elemento → .*	
Matriz identidad de orden n → I(ndim)	Matriz de unos de orden nxk → ones(ndim,kdim)	
Número de columnas de la matriz A → cols(A)	Número de filas de la matriz A → rows(A)	
Rango de la matriz A → rank(A)	Traza de la matriz A → tr(A)	
Producto kronecker → **	División, elemento a elemento → ./	
Concatenación de matrices por columnas → ~	Concatenación de matrices por filas →	
Matriz de ceros de orden nxk → zeros(ndim,kdim)	Determinante de la matriz A → det(A)	
Inversa de la matriz A → inv(A)	Diagonal principal de la matriz A → diag(A)	
Matriz triangular superior de la matriz A → upper(A)	Matriz triangular inferior de la matriz A → lower(A)	
Vectorializa la matriz A → vec(A)	Matriz cholesky de la matriz A → chol(A)	
Aplica la raíz cuadrada a cada uno de los elementos de la matriz A → sqrt(A)		
Aplica el logaritmo natural a cada uno de los elementos de la matriz A → log(A)		
Aplica la función exponencial a cada uno de los elementos de la matriz A → exp(A)		
Matriz de cuasivarianzas de las columnas de la matriz A → mcov(A)		
Matriz de correlaciones de las columnas de la matriz A → mcorr(A)		
Raíces y vectores característicos de la matriz A → eig(A,null) y para guardar los vectores característicos en la matriz predefinida B → eig(A,B)		
Suma de los elementos de cada columna de la matriz A → sumc(A)		
Media de los elementos de cada columna de la matriz A → meanc(A)		
Máximo de los elementos de cada columna de la matriz A → maxc(A)		
Mínimo de los elementos de cada columna de la matriz A → minc(A)		
Cuasidesviación estándar de los elementos de cada columna de la matriz A → sdc(A)		
Suma de los elementos de cada fila de la matriz A → sumr(A)		
Media de los elementos de cada fila de la matriz A → meanr(A)		
Máximo de los elementos de cada fila de la matriz A → maxr(A)		
Mínimo de los elementos de cada fila de la matriz A → minr(A)		

Cuadro 2-4. Algunos operadores disponibles con el comando **matrix**.

Si se aplican estos operadores a los datos disponibles:

```
# Se genera XT como la traspuesta de la matriz X
? matrix XT = X'
Se ha generado la matriz XT

? print X XT
X (10 x 3) [t1 = 1, t2 = 10]
    1.0000    2.4600    1079.5
    1.0000    2.5100    1108.2
    1.0000    2.7300    1116.8
    1.0000    2.7300    978.09
    1.0000    2.6300    809.62
    1.0000    2.3800    1047.1
    1.0000    2.4400    1164.3
    1.0000    2.7500    1069.6
    1.0000    2.9800    1040.3
    1.0000    1.9500    1067.8

XT (3 x 10)
```

<sup>17</sup> Las variables o series se pueden considerar como matrices de una columna (vectores columna) y los escalares como matrices de una fila y una columna. Por tanto, en las expresiones de cálculo, para definir nuevas matrices pueden intervenir no sólo matrices sino también escalares y series.

1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0000	1.0000						
2.4600	2.5100	2.7300	2.7300	2.6300	2.3800	2.4400	2.7500
2.9800	1.9500						
1079.5	1108.2	1116.8	978.09	809.62	1047.1	1164.3	1069.6
1040.3	1067.8						

Recuerde que la trasposición de matrices consiste en intercambiar filas por columnas y, por tanto, si la dimensión de una matriz es  $m \times n$ , el orden de su traspuesta será  $n \times m$  (en este ejemplo  $10 \times 3$  y  $3 \times 10$ , respectivamente).

Obsérvese que como Gretl no dispone de espacio suficiente para mostrar las 10 columnas que forman la matriz traspuesta, utiliza dos líneas para mostrar cada fila de la matriz.

```
# Se genera XTX como el producto de la traspuesta de X por X
? matrix XTX = XT * X
Se ha generado la matriz XTX

? print XTX
XTX (3 x 3)
  10.000    25.560    10481.
  25.560    66.038    26742.
 10481.    26742.    1.1071e+007
```

Para que el operador **matrix** funcione correctamente es necesario que las matrices que intervienen en la expresión de cálculo tengan la estructura adecuada para realizar la operación.

```
? matrix XX = X * X
Las matrices no tienen la estructura adecuada para poder realizar la operación
```

Obsérvese que la matriz XTX está definida, mientras que la matriz XX no, ello es debido a que para multiplicar matrices es necesario que el número de columnas de la primera matriz coincida con el número de filas de la segunda y, en el caso de la matriz XX esto no se cumple y Gretl emite el mensaje de error: “*las matrices no tienen la estructura adecuada para poder realizar la operación*”.

```
# Se genera XTXD como el determinante de XTX
? matrix XTXD = det(XTX)
Se ha generado la matriz XTXD

? print XTXD
XTXD (1 x 1)
 5.8098e+005
```

Recuerde que el determinante sólo está definido para matrices cuadradas.

```
# Se genera XTXI como la inversa de XTX
? matrix XTXI = inv(XTX)
Se ha generado la matriz XTXI

? print XTXI
XTXI (3 x 3)
 27.503    -4.6284    -0.014858
 -4.6284    1.4718    0.00082674
 -0.014858  0.00082674  1.2159e-005
```

Recuerde que sólo se pueden invertir matrices cuadradas cuyo determinante sea no nulo (XTX tiene inversa porque es una matriz no singular).

```
# Se genera DXTX como la diagonal de XTX
? matrix DXTX = diag(XTX)
Se ha generado la matriz DXTX

? print DXTX
DXTX (3 x 1)
 10.000
 66.038
 1.1071e+007
```

La función **diag** genera un vector columna que contiene los elementos de la diagonal principal de la matriz.

```
# Se genera S como suma de XTX y C
? matrix S = XTX + C
Se ha generado la matriz S

# Se genera D como diferencia de XTX y C
? matrix D = XTX - C
Se ha generado la matriz D

? print XTX C S D
XTX (3 x 3)
  10.000      25.560      10481.
  25.560      66.038      26742.
  10481.      26742.  1.1071e+007

C (3 x 3)
  1  2  3
  4  5  6
  7  8  9

S (3 x 3)
  11.000      27.560      10484.
  29.560      71.038      26748.
  10488.      26750.  1.1071e+007

D (3 x 3)
  9.0000      23.560      10478.
  21.560      61.038      26736.
  10474.      26734.  1.1071e+007
```

Recuerde que para poder sumar o restar matrices es necesario que tengan la misma dimensión.

```
# Se genera CE como la matriz resultante de aplicar la exponencial a cada uno de los elementos de C
? matrix CE = exp(C)
Se ha generado la matriz CE

? print CE
CE (3 x 3)
  2.7183      7.3891      20.086
  54.598      148.41      403.43
  1096.6      2981.0      8103.1

# Se genera CEL como la matriz resultante de aplicar el logaritmo natural a cada uno de los
elementos de CE
? matrix CEL = log(CE)
Se ha generado la matriz CEL

? print CEL
CEL (3 x 3)
  1  2  3
  4  5  6
  7  8  9

# Se genera XR como la matriz resultante de aplicar la raíz cuadrada a cada uno de los elementos
de X
? matrix XR = sqrt(X)
Se ha generado la matriz XR

? print XR
XR (10 x 3)
  1.0000      1.5684      32.856
  1.0000      1.5843      33.290
  1.0000      1.6523      33.419
  1.0000      1.6523      31.274
  1.0000      1.6217      28.454
  1.0000      1.5427      32.359
  1.0000      1.5620      34.122
  1.0000      1.6583      32.705
  1.0000      1.7263      32.254
  1.0000      1.3964      32.677
```

Se debe tener en cuenta que las funciones **exp**, **log** y **sqrt** se aplican a cada uno de los elementos de la matriz.

```
# Se genera RX como el rango de X
? matrix RX = rank(X)
Se ha generado la matriz RX

? print RX
RX (1 x 1)
3
```

Recuerde que el rango de una matriz es el número de columnas linealmente independientes y que para que una matriz cuadrada sea invertible es necesario que su rango coincida con su dimensión (XTX es una matriz cuadrada de orden 3 y con rango es 3, por tanto, tiene inversa).

```
# Se genera TXTX como la traza de XTX
? matrix TXTX = tr(XTX)
Se ha generado la matriz TXTX

? print TXTX
TXTX (1 x 1)
1.1071e+007
```

Recuerde que la traza de una matriz es la suma de los elementos de su diagonal principal (TXTX es la suma de los elementos de la diagonal principal de XTX).

```
# Se genera TRIIXTX como una matriz triangular inferior a partir de XTX
? matrix TRIIXTX = lower(XTX)
Se ha generado la matriz TRIIXTX

# Se genera TRISXTX como una matriz triangular superior a partir de XTX
? matrix TRISXTX = upper(XTX)
Se ha generado la matriz TRISXTX

? print XTX TRIIXTX TRISXTX
XTX (3 x 3)
  10.000      25.560      10481.
  25.560      66.038      26742.
 10481.      26742.    1.1071e+007

TRIIXTX (3 x 3)
  10.000      0.0000      0.0000
  25.560      66.038      0.0000
 10481.      26742.    1.1071e+007

TRISXTX (3 x 3)
  10.000      25.560      10481.
  0.0000      66.038      26742.
  0.0000      0.0000    1.1071e+007
```

Recuerde que sólo se pueden generar matrices triangulares a partir de matrices cuadradas. En las matrices triangulares inferiores los elementos que están por encima de la diagonal principal son nulos y en las matrices triangulares superiores lo son los que están por debajo, los restantes elementos coinciden con los de la matriz de partida.

```
# Se genera I3 como una matriz identidad de orden 3
? matrix I3 = I(3)
Se ha generado la matriz I3

? print I3
I3 (3 x 3)
  1  0  0
  0  1  0
  0  0  1

# Se genera I7 como una matriz identidad de orden 7
? matrix I7 = I(7)
Se ha generado la matriz I7

? print I7
```

```

I7 (7 x 7)

 1  0  0  0  0  0  0
 0  1  0  0  0  0  0
 0  0  1  0  0  0  0
 0  0  0  1  0  0  0
 0  0  0  0  1  0  0
 0  0  0  0  0  1  0
 0  0  0  0  0  0  1

# Se genera una matriz de ceros
? matrix O32 = zeros(3,2)
Se ha generado la matriz O32

? print O32
O32 (3 x 2)
 0  0
 0  0
 0  0

# se genera una matriz de unos
? matrix l32 = ones(3,2)
Se ha generado la matriz l32

? print l32
l32 (3 x 2)
 1  1
 1  1
 1  1

```

Recuerde que una matriz identidad es una matriz cuadrada con elementos diagonales iguales a uno y elementos no diagonales nulos.

La función **zeros** permite generar matrices nulas y la función **ones**, matrices con todos sus elementos iguales a uno.

```

# Se genera la matriz F
? matrix F = {1, 2, 3; 4, 5, 6}
Se ha generado la matriz F

# Producto de F por F elemento a elemento
? matrix FMEF = F .* F
Se ha generado la matriz FMEF

# División de D por D elemento a elemento
? matrix FDEF = F ./ F
Se ha generado la matriz FDEF

? print F FMEF FDEF
F (2 x 3)
 1  2  3
 4  5  6

FMEF (2 x 3)
 1  4  9
16 25 36

FDEF (2 x 3)
 1  1  1
 1  1  1

```

Para poder realizar el producto y el cociente de dos matrices elemento a elemento, es necesario que ambas tengan la misma dimensión. La matriz resultante se obtiene multiplicando o dividiendo los elementos que ocupan el mismo lugar en ambas matrices.

```

# Se genera la matriz G
? matrix G = {7, 8, 9; 10, 11, 12}
Se ha generado la matriz G

# Concatenación por filas
? matrix FGrows = F | G
Se ha generado la matriz FGrows

```

```

# Concatenación por columnas
? matrix FGcols = F ~ G
Se ha generado la matriz FGcols

? print F G FGrows FGcols
F (2 x 3)
  1  2  3
  4  5  6

G (2 x 3)
  7  8  9
 10 11 12

FGrows (4 x 3)
  1  2  3
  4  5  6
  7  8  9
 10 11 12

FGcols (2 x 6)
  1  2  3  7  8  9
  4  5  6 10 11 12

```

Para poder concatenar matrices es necesario que tengan la misma dimensión. El usuario debe tener en cuenta que en la concatenación por filas, las filas de la segunda matriz se añaden a las de la primera por la parte inferior y en la concatenación por columnas, las columnas de la segunda matriz se añaden a las de la primera por la parte derecha.

```

# Se genera vF como un vector a partir de la matriz F
? matrix vF = vec(F)
Se ha generado la matriz vF

? print F vF
F (2 x 3)
  1  2  3
  4  5  6

vF (6 x 1)
  1
  4
  2
  5
  3
  6

```

La función **vec** permite convertir una matriz en un vector columna. Obsérvese que vF es un vector columna de 6 filas, donde en primer lugar aparecen los elementos de la primera columna de F, a continuación los de la segunda y, por último, los de la tercera columna.

```

# Se genera la matriz H
? matrix H = {1, 2; 3, 4}
Se ha generado la matriz H

# Se genera la matriz J
? matrix J = ones(3,2)
Se ha generado la matriz J

# Se genera HPKJ como el producto kronecker de H por J
? matrix HPKJ = H ** J
Se ha generado la matriz HPKJ

? print H J HPKJ
H (2 x 2)
  1  2
  3  4

J (3 x 2)
  1  1
  1  1
  1  1

```

```
HPKJ (6 x 4)
1 1 2 2
1 1 2 2
1 1 2 2
3 3 4 4
3 3 4 4
3 3 4 4
```

Recuerde que el producto kronecker de una matriz  $A_{m \times n}$  por una matriz  $B_{p \times q}$  es una matriz  $C_{(m \times p) \times (n \times q)}$  formada por  $(m \times n)$  bloques de  $(p \times q)$  elementos, donde cada uno de los bloques es el resultado de multiplicar cada uno de los elementos de la primera matriz por cada uno de los elementos de la segunda matriz. En este caso, la matriz resultante del producto kronecker de H por J consta de cuatro bloques de dimensiones  $3 \times 2$  cada uno, donde el primer bloque es el resultado de multiplicar el primer elemento de la primera fila de H por cada uno de los elementos de J; el segundo bloque es el resultado de multiplicar el segundo elemento de la primera fila de H por cada uno de los elementos de J; el tercer bloque es el resultado de multiplicar el primer elemento de la segunda fila de H por cada uno de los elementos de J y el cuarto bloque es el resultado de multiplicar el segundo elemento de la segunda fila de H por cada uno de los elementos de J.

```
# Se genera el número de columnas de G
? matrix colsG = cols(G)
Se ha generado la matriz colsG

# Se genera el número de filas de G
? matrix rowsG = rows(G)
Se ha generado la matriz rowsG

? print G colsG rowsG
G (2 x 3)
 7  8  9
10 11 12

colsG (1 x 1)
 3

rowsG (1 x 1)
 2
```

Las funciones **cols** y **rows** hacen referencia al número de filas y columnas de una matriz respectivamente.

```
# Raíces y vectores característicos
? matrix eigenC = eigengen(C,null)
Se ha generado la matriz eigenC

? matrix vcC = I(2)
Se ha generado la matriz vcC

? matrix eigenC = eigengen(C,&vcC)
Se ha reemplazado la matriz eigenC

? print eigenC vcC
eigenC (3 x 1)
 16.117
 -1.1168
-1.2220e-015

vcC (3 x 3)
-0.23197  -0.78583   0.40825
-0.52532  -0.086751  -0.81650
-0.81867   0.61233   0.40825
```

La función **eigengen** permite calcular y guardar las raíces y vectores característicos de una matriz cuadrada. Para guardar los vectores característicos es necesario hacerlo en una matriz que haya sido definida previamente.

```
# Matriz cholesky
? matrix P = cholesky(XTX)
Se ha generado la matriz P

? matrix XTXc = P * P'
Se ha generado la matriz XTXc

? print P XTX XTXc
P (3 x 3)
  3.1623      0.0000      0.0000
  8.0828      0.84050     0.0000
 3314.5     -57.147      286.78

XTX (3 x 3)
 10.000      25.560      10481.
 25.560      66.038      26742.
10481.      26742.    1.1071e+007

XTXc (3 x 3)
 10.000      25.560      10481.
 25.560      66.038      26742.
10481.      26742.    1.1071e+007
```

Recuerde que la función **cholesky** sólo es aplicable a matrices definidas positivas: *si A es una matriz definida positiva  $\Rightarrow$  existe una matriz P tal que  $P'P=A$*  (dado que XTX es una matriz definida positiva  $P*P'=XTX$ ).

```
# Se genera VXraya como la matriz de cuasivarianzas-cuasicovarianzas de x1 y x2
? matrix Xraya = {x1, x2}
Se ha generado la matriz Xraya

? matrix VXraya = mcov(Xraya)
Se ha generado la matriz VXraya

? print VXraya
VXraya (2 x 2)
 0.078493     -5.3369
 -5.3369      9500.7
```

La función **mcov** permite calcular la matriz de cuasivarianzas-cuasicovarianzas de las variables incluidas en las columnas de la matriz inicial. Dichas matrices son cuadradas y simétricas. En el ejemplo anterior, dado que la primera columna de VXraya se corresponde con la variable x1 y la segunda con x2, el primer elemento de la diagonal principal de VXraya es la cuasivarianza de x1, segundo la cuasivarianza de x2 y el elemento no diagonal es la cuasicovarianza de x1 y x2.

```
# Se genera RXraya como la matriz de correlación de x1 y x2
? matrix RXraya = mcorr(Xraya)
Se ha generado la matriz RXraya

? print RXraya
RXraya (2 x 2)
 1.0000     -0.19543
 -0.19543    1.0000
```

La función **mcorr** permite calcular la matriz de correlaciones de las variables incluidas en las columnas de la matriz inicial. Dicha matriz es cuadrada, simétrica y sus elementos diagonales son iguales a uno. En el ejemplo anterior, dado que la primera columna de VXraya se corresponde con la variable x1 y la segunda con x2, el elemento no diagonal es la correlación entre x1 y x2.

Gretl permite calcular vectores que recogen la suma, media, valores máximos y mínimos tanto por filas como por columnas. Además, por columnas permite calcular el vector que recoge las cuasidesviaciones típicas.

```
# Cálculos por columnas
# Vector fila que contiene la suma de los elementos de cada columna de la matriz C
? matrix Csc = sumc(C)
Se ha generado la matriz Csc
```

```
# Vector fila que contiene la media de los elementos de cada columna de la matriz C
? matrix Cmc = meanc(C)
Se ha generado la matriz Cmc

# Vector fila que contiene los valores máximos de los elementos de cada columna de la matriz C
? matrix Cmaxc = maxc(C)
Se ha generado la matriz Cmaxc

# Vector fila que contiene los valores mínimos de los elementos de cada columna de la matriz C
? matrix Cminc = minc(C)
Se ha generado la matriz Cminc

# Vector fila que contiene las desviaciones estándar de los elementos de cada columna de la matriz C
? matrix Csd = sdc(C)
Se ha generado la matriz Csd

? print C Csc Cmc Cmaxc Cminc Csd
C (3 x 3)
  1  2  3
  4  5  6
  7  8  9

Csc (1 x 3)
 12  15  18

Cmc (1 x 3)
  4  5  6

Cmaxc (1 x 3)
  7  8  9

Cminc (1 x 3)
  1  2  3

Csd (1 x 3)
 2.4495  2.4495  2.4495

# Cálculos por filas
# Vector columna que contiene la suma de los elementos de cada fila de la matriz C
? matrix Csr = sumr(C)
Se ha generado la matriz Csr

# Vector columna que contiene la media de los elementos de cada fila de la matriz C
? matrix Cmr = meanr(C)
Se ha generado la matriz Cmr

# Vector columna que contiene los valores máximos de los elementos de cada fila de la matriz C
? matrix Cmaxr = maxr(C)
Se ha generado la matriz Cmaxr

# Vector columna que contiene los valores mínimos de los elementos de cada fila de la matriz C
? matrix Cminr = minr(C)
Se ha generado la matriz Cminr

? print C Csr Cmr Cmaxr Cminr
C (3 x 3)
  1  2  3
  4  5  6
  7  8  9

Csr (3 x 1)
  6
 15
 24

Cmr (3 x 1)
  2
  5
  8

Cmaxr (3 x 1)
  3
  6
```

```

9
Cminr (3 x 1)
1
4
7

```

Obsérvese que los cálculos por columnas se recogen en vectores fila y los cálculos por filas se recogen en vectores columna.

### 2.9.2.1. ¿Cómo identificar los elementos de una matriz?

Los elementos de una matriz se identifican por el nombre de la matriz seguido de dos números entre corchetes, donde el primero hace referencia a la fila y el segundo hace referencia a la columna en las que se encuentra ese elemento dentro de la matriz:

*Nombre de la matriz*[ $n^o_1, n^o_2$ ]

```

? print C
C (3 x 3)
1 2 3
4 5 6
7 8 9

? matrix c11 = C[1,1]
Se ha generado la matriz c11

? genr c33 = C[3,3]
Se ha generado el escalar c33 = 9

? scalar c12 = C[1,2]
Se ha generado el escalar c12 = 2

? series c21 = C[2,1]
Se ha generado la serie c21 (ID 30)

? print c11 c33 c12 c21 --byobs
      c21
1      4
2      4
3      4
4      4
5      4
6      4
7      4
8      4
9      4
10     4

c11 (1 x 1)
1
      c33 = 9.0000000
      c12 = 2.0000000

```

Dado que los elementos de una matriz se pueden considerar escalares, pueden ser utilizados no sólo con el comando **matrix**, sino también con los comandos **scalar**, **series** y **genr**.

Cuando en un comando **print** intervienen matrices, escalares y series, Gretl visualiza en primer lugar las series, en segundo lugar las matrices y en último lugar los escalares, independientemente del lugar que ocupen los mismos en el comando.

### 2.9.2.2. ¿Cómo identificar las filas de una matriz?

Para identificar las filas de una matriz se indica el primer número que hace referencia a la fila y el segundo, que hace referencia a la columna se deja en blanco:

*Nombre de la matriz*[ $n^o_1,$

```
? matrix c1t = C[1]
Se ha generado la matriz c1t

? matrix c3t = C[3]
Se ha generado la matriz c3t

? print c1t c3t
c1t (1 x 3)
 1  2  3

c3t (1 x 3)
 7  8  9
```

Obsérvese que `c1t` y `c3t` son vectores filas de orden (1x3) que contienen los elementos de la primera y tercera fila de la matriz `C`, respectivamente.

### 2.9.2.3. ¿Cómo identificar las columnas de una matriz?

Para identificar las columnas de una matriz se indica el segundo número que hace referencia a la columna y el primero, que hace referencia a la fila se deja en blanco:

*Nombre de la matriz[,n°2]*

```
? print C
C (3 x 3)
 1  2  3
 4  5  6
 7  8  9

? matrix c1 = C[ , 1]
Se ha generado la matriz c1

? matrix c3 = C[ , 3 ]
Se ha generado la matriz c3

? print c1 c3
c1 (3 x 1)
 1
 4
 7

c3 (3 x 1)
 3
 6
 9
```

Obsérvese que `c1` y `c3` son vectores columna de orden (3x1) que contienen los elementos de la primera y tercera columnas de la matriz `C`, respectivamente.

## 2.10. Relación entre los comandos `genr`, `matrix` y `scalar`

En los contextos adecuados, los comandos **genr**, **matrix** y **scalar** se pueden considerar sinónimos. Además, se debe tener en cuenta que no es estrictamente necesario, aunque si conveniente, que dichos comandos aparezcan de forma explícita en las instrucciones de cálculo (sobretudo para el usuario que se inicie en la utilización del programa).

```
? genr r1 = 3
Se ha generado el escalar r1 = 3

? scalar r2 = 3
Se ha generado el escalar r2 = 3

? matrix r3 = 3
Se ha generado la matriz r3

? r4 = 3
Se ha generado el escalar r4 = 3
```

Obsérvese que no resulta eficiente utilizar el comando **matrix** para definir escalares.

```
? matrix CTC1 = C' * C
Se ha generado la matriz CTC1

? genr CTC2 = C' * C
Se ha generado la matriz CTC2

? CTC3 = C' * C
Se ha generado la matriz CTC3
```

Se debe tener en cuenta que cuando el resultado de la operación matricial es un escalar sólo será considerado como una matriz si para su cálculo se utiliza el comando **matrix**, en cuyo caso, Gretl creará en la **Ventana vista de iconos** el icono correspondiente a la matriz de orden 1x1 que contenga dicho escalar.

```
? genr series h1 =x1
Se ha generado la serie h1 (ID 31)

? series h2 = x1
Se ha generado la serie h2 (ID 32)

? h3 = x1
Se ha generado la serie h3 (ID 33)
```

## 2.11. ¿Cómo calcular los estadísticos descriptivos de las variables?

Existen varias alternativas para calcular los estadísticos descriptivos de las variables:

1. Utilizar el menú **Ver** de la **Barra de Menú** de la **Ventana Principal**.
2. Utilizar el comando **summary**.

### 2.11.1. Utilizar el menú Ver

Antes de acceder al menú **Ver** y seleccionar **Estadísticos principales** será necesario que en la **Ventana Principal** se seleccione/n la/s variable/s de la/s que se quiera obtener los estadísticos descriptivos (véase parte superior izquierda de la Ilustración 2-15).

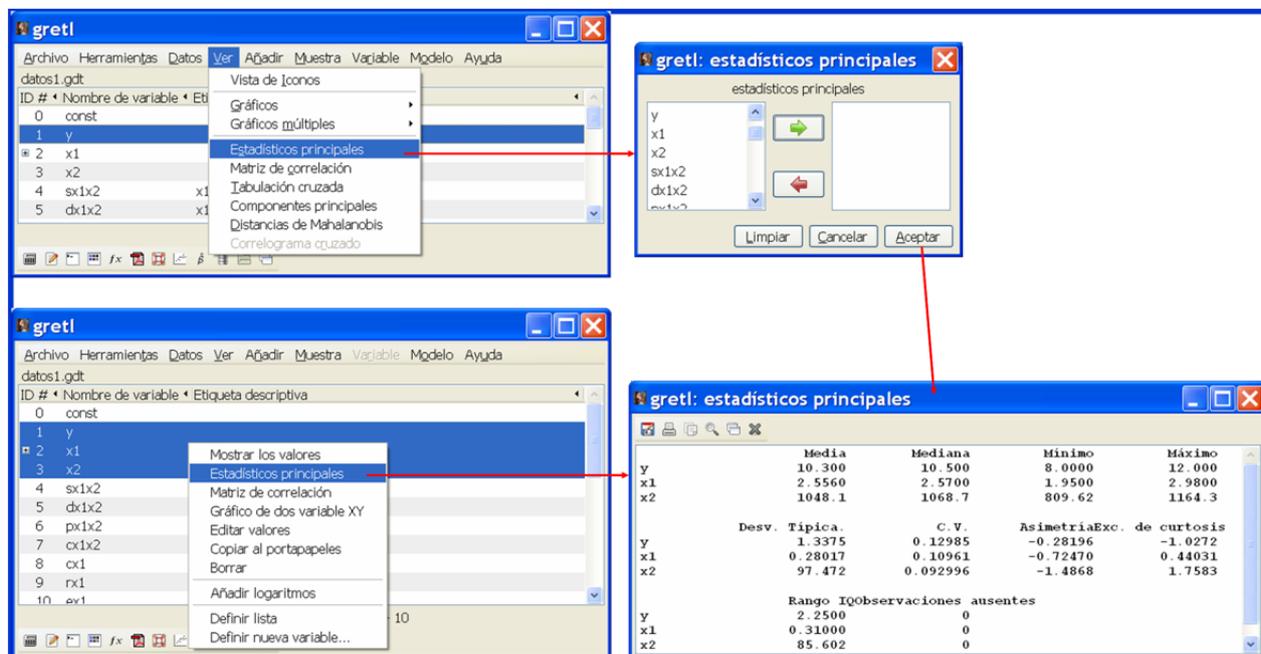


Ilustración 2-15. Formas de acceder a la Ventana estadísticos principales.

Otra forma de acceder a los estadísticos descriptivos es a través del botón derecho del ratón y en el menú emergente seleccionar **Estadísticos descriptivos** (véase parte inferior izquierda de la Ilustración 2-15).

Cuando se esté interesado en ver los estadísticos descriptivos de una sola variable, también se podrá utilizar el menú *Variable* y seleccionar *Estadísticos principales*.

### 2.11.2. Utilizar el comando summary

Para visualizar los estadísticos descriptivos se puede utilizar el comando **summary**, cuyo formato es:

**summary** v<sub>1</sub> v<sub>2</sub> ...

Si en el momento de ejecutar el comando **summary** no se indica un listado de variables, Gretl incluirá todas las variables de la sesión de trabajo.

En la **Tabla 2-1** se recoge la salida estándar de un comando **summary** en el que interviene una única variable genérica X<sub>i</sub> con T observaciones (t=1,2,...,T).

**Tabla 2-1.** Tabla de estadísticos descriptivos.

Estadísticos principales, usando las observaciones **inic–fin** del rango muestral para la variable 'X<sub>i</sub>' (n° observaciones válidas)

<b>Media</b>	Media de la variable X <sub>i</sub> $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it}$
<b>Mediana</b>	Mediana de la variable X <sub>i</sub>
<b>Mínimo</b>	Valor mínimo de la variable X <sub>i</sub>
<b>Máximo</b>	Valor máximo de la variable X <sub>i</sub>
<b>Desviación típica</b>	Desviación típica de la variable X <sub>i</sub> $S_{X_i} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2}$
<b>C.V.</b>	Coficiente de variación de la variable X <sub>i</sub> $CV_{X_i} = \frac{S_{X_i}}{ \bar{X}_i }$
<b>Asimetría</b>	Coficiente de asimetría de la variable X <sub>i</sub> $\gamma_1 = \frac{1}{T} \sum_{t=1}^T \left( \frac{X_{it} - \bar{X}_i}{S_{X_i}} \right)^3$
<b>Exc. de curtosis</b>	Coficiente de exceso de curtosis de la variable X <sub>i</sub> $\gamma_2 = \frac{1}{T} \sum_{t=1}^T \left( \frac{X_{it} - \bar{X}_i}{S_{X_i}} \right)^4 - 3$
<b>Percentil 5%</b>	Percentil del 5% de la variable X <sub>i</sub>
<b>Percentil 95%</b>	Percentil del 95% de la variable X <sub>i</sub>
<b>Rango IQ</b>	Rango intercuartílico de la variable X <sub>i</sub> $RIQ_{X_i} = Q_3 - Q_1$
<b>Observaciones ausentes</b>	Número de observaciones ausentes de la variable X <sub>i</sub>

Como puede observarse en la salida estándar del comando **summary**, se muestran estadísticos descriptivos que reflejan algunas de las características de la distribución de la variable. La media, la mediana, los cuartiles y los percentiles son medidas de posición, la desviación típica, el coeficiente de variación y el rango intercuartílico son medidas de dispersión y los coeficientes de asimetría y exceso de curtosis son medidas de forma.

Las medidas de posición central dan una idea de la situación o centro de los datos, es decir, proporcionan un “valor representativo” del conjunto de datos. La media es el valor promedio y es un estadístico poco robusto frente a la existencia de valores extremos en la muestra (las observaciones anómalas van a tener una gran influencia en el valor que tome dicho estadístico). La mediana es el valor central de la distribución y no varía mucho ante cambios en los valores de unas pocas

observaciones, es decir, que en muestras con observaciones anómalas es más representativa del conjunto de datos la mediana que la media.

Para valorar la representatividad de las medidas de posición se utilizan las medidas de dispersión, que informan de lo poco o muy concentradas que están las observaciones entorno a su valor.

La desviación típica mide el grado de dispersión de los datos alrededor de su media. En general, cuanto más cercana a cero esté la desviación típica, más concentradas estarán las observaciones alrededor de su media y su grado de representatividad será mayor. No obstante, al depender de las unidades de medida, no es fácil comparar la representatividad de varios conjuntos de datos.

El problema anterior se soluciona utilizando el coeficiente de variación, que es adimensional. De manera que si el coeficiente de variación es menor que uno, se considera que la media es bastante representativa del conjunto de datos, puesto que su dispersión es pequeña en relación con el valor de la media.

La asimetría de la distribución da una idea de si los datos se distribuyen simétricamente o no entorno a la media. Será negativa cuando la cola izquierda (asociada a valores por debajo de la media) sea más larga que la cola derecha y positiva en caso contrario. Una asimetría negativa se corresponde con una media inferior a la mediana y positiva al contrario.

El coeficiente de asimetría mide la cantidad de observaciones que se encuentran en las colas en relación con las situadas alrededor de la media, siendo tres su valor de referencia para una distribución normal. Por tanto, un exceso de curtosis negativo indica un menor número de observaciones en las colas y un menor apuntamiento, mientras que si es positivo indica un mayor peso de las colas y un mayor apuntamiento.

En el Cuadro 2-5 se recogen algunas de las opciones disponibles con el comando **summary**.

**--simple** → muestra una versión restringida de los estadísticos descriptivos, es decir, con esta opción sólo se calcula y muestra la media, la desviación típica y los valores máximos y mínimos de las variables que aparezcan en el comando **summary** (sin esta opción también se calculan y muestran los coeficientes de variación, asimetría, curtosis, rango intercuartílico, número de observaciones ausentes y los percentiles del 5% y del 95% cuando el tamaño muestral es igual o superior a 19).

**--by=vi** → calcula y muestra los estadísticos descriptivos de las variables que intervienen en el comando **summary** para cada uno de los valores de la variable (**vi** tiene que ser una variable discreta, por ejemplo, una variable ficticia).

Cuadro 2-5. Algunas opciones del comando **summary**.

A continuación se muestran algunos ejemplos de ejecución del comando **summary** con y sin dichas opciones:

```
# Tabla de estadísticos principales para la variable y
? summary y
Estadísticos principales, usando las observaciones 1 - 10
para la variable 'y' (10 observaciones válidas)
Media                10.300
Mediana              10.500
Mínimo               8.0000
Máximo               12.000
Desviación típica    1.3375
C.V.                 0.12985
Asimetría            -0.28196
Exc. de curtosis     -1.0272
Rango intercuartílico 2.2500
Observaciones ausentes 0

# Tabla reducida de estadísticos principales para la variable y
? summary y --simple
Estadísticos principales, usando las observaciones 1 - 10
para la variable 'y' (10 observaciones válidas)
Media                10.300
Mínimo               8.0000
Máximo               12.000
```

```

Desviación típica          1.3375
Observaciones ausentes      0

# Tabla reducida de estadísticos principales para la variable y para cada uno de los valores de la
variable ficticia v1
? summary y --simple --by=v1
Estadísticos principales de y, por valores de v1
v1 = 0 (n = 8):
  Media          10.500
  Mínimo         8.0000
  Máximo        12.000
  Desviación típica 1.4142
  Observaciones ausentes 0

v1 = 1 (n = 2):
  Media          9.5000
  Mínimo         9.0000
  Máximo        10.000
  Desviación típica 0.70711
  Observaciones ausentes 0

```

Cuando en el comando **summary** interviene más de una variable, el output de dicho comando no hace referencia al rango muestral utilizado:

```

# Tabla de estadísticos principales para las variables x1, x2 e y
? summary x1 x2 y

```

	Media	Mediana	Mínimo	Máximo
x1	2.5560	2.5700	1.9500	2.9800
x2	1048.1	1068.7	809.62	1164.3
y	10.300	10.500	8.0000	12.000

	Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
x1	0.28017	0.10961	-0.72470	0.44031
x2	97.472	0.092996	-1.4868	1.7583
y	1.3375	0.12985	-0.28196	-1.0272

	Rango IQ	Observaciones ausentes
x1	0.31000	0
x2	85.602	0
y	2.2500	0

Los percentiles aparecen en el output del comando **summary** cuando el tamaño muestral es igual o superior a 19:

```

? smpl 1 20
Rango de datos completo: 1 - 60 (n = 60)
Muestra actual: 1 - 20 (n = 20)

? summary x1 x2 y

```

	Media	Mediana	Mínimo	Máximo
x1	2.4630	2.4500	1.9500	2.9800
x2	1053.5	1068.7	809.62	1314.5
y	10.650	11.000	8.0000	13.000

	Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
x1	0.22697	0.092153	0.16015	0.46824
x2	117.19	0.11124	0.058080	-0.021098
y	1.2258	0.11510	-0.18073	-0.32341

	Perc. 5%	Perc. 95%	Rango IQ	Observaciones ausentes
x1	1.9620	2.9685	0.28500	0
x2	814.66	1308.8	158.98	0
y	8.0500	12.950	1.7500	0

Hay que señalar que puede ser interesante, sobre todo en estudios de corte transversal, completar el análisis descriptivo individual de las variables con un estudio de la tabla y del histograma de frecuencias, a los que se puede acceder con el menú emergente que aparece cuando en la **Ventana Principal**, una vez seleccionada la variable, se pincha con el botón derecho del ratón y se selecciona **Distribución de frecuencias** o bien a través de la opción *Distribución de frecuencias* del menú *Variable* (véase Ilustración 2-16).

Como puede verse en la parte central de la Ilustración 2-16, se pueden modificar las características de la tabla y del gráfico de frecuencias realizando las elecciones oportunas en el cuadro de diálogo **distribución de frecuencias**:

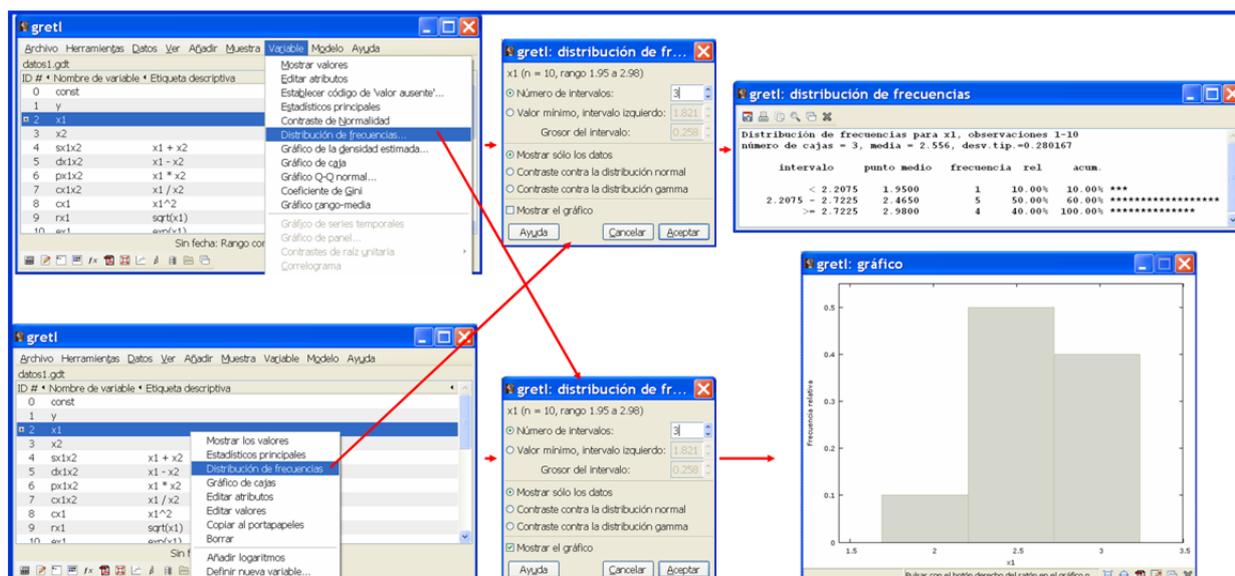


Ilustración 2-16. Formas de acceder a la Ventana distribución de frecuencias y del histograma.

- **Número de intervalos** → Gretl sólo permite seleccionar un número impar de intervalos, siendo tres el número mínimo y el máximo T si T es impar y T-1 si T es par, siendo T el tamaño del rango muestral. No obstante, parece conveniente elegir un número entero próximo a  $\sqrt{T}$  (en el ejemplo de la Ilustración 2-16, como se está trabajando con un rango de 10 observaciones, se ha elegido como número de intervalos 3).
- **Valor mínimo, intervalo izquierdo y grosor del intervalo** → todos los intervalos deben tener la misma amplitud. Por defecto, Gretl los elige de manera que el punto central o marca de clase del primer y último intervalo sean, respectivamente, el valor mínimo y el valor máximo de la variable en el rango muestral utilizado.
- **Mostrar sólo los datos** → se debe elegir esta opción cuando se quiera obtener la tabla de frecuencias.
- **Mostrar el gráfico** → se debe elegir esta opción cuando se quiera obtener el histograma de frecuencias.

Por defecto, Gretl muestra el histograma de frecuencias, debido a que tiene seleccionado simultáneamente **Mostrar sólo los datos** y **Mostrar el gráfico**, por tanto, para visualizar la tabla de frecuencias es necesario no seleccionar **Mostrar el gráfico**.

La opción **Copiar** de la **Ventana distribución de frecuencias** permite importar la tabla de distribución de frecuencias entre otros formatos, como un fichero MS Word. A continuación se

muestra en formato MS Word, la tabla de distribución de frecuencias de la variable  $x_1$  recogida en la parte superior derecha de la Ilustración 2-16:

Distribución de frecuencias para $x_1$ , observaciones 1-10 número de cajas = 3, media = 2.556, desv.típ.=0.280167					
intervalo	punto medio	frecuencia	rel	acum.	
< 2.2075	1.9500	1	10.00%	10.00%	***
2.2075 - 2.7225	2.4650	5	50.00%	60.00%	*****
>= 2.7225	2.9800	4	40.00%	100.00%	*****

En la primera columna, denominada **intervalo**, aparecen los tres intervalos en los que se han dividido los valores que toma la variable  $x_1$  y en la segunda, denominada **punto medio**, aparece la marca de clase de cada intervalo. Bajo la denominación **frecuencia** se muestra la frecuencia absoluta del intervalo (en este caso, hay 5 observaciones comprendidas entre 2.2075 y 2.7225 €). En la columna **rel** se muestra la frecuencia relativa de cada intervalo (porcentaje de observaciones que hay en cada tramo) y en la columna **acum**, aparece la frecuencia relativa acumulada hasta ese intervalo.

Las frecuencias relativas son las que se utilizan para construir el histograma de frecuencias. Por ejemplo, las 5 observaciones que se encuentran en el intervalo [2.2075, 2.7225] constituyen el 50% del total de observaciones y, por tanto, dado que todos los intervalos son de igual amplitud, la altura de la barra central del histograma es su frecuencia relativa asociada en tanto por uno, es decir, 0.5 (véase el gráfico que aparece en la parte inferior derecha de la Ilustración 2-16<sup>18</sup>). Hay que señalar, que en la última columna de la tabla de frecuencias aparecen representadas las frecuencias relativas, no obstante, se trata de una representación de menor calidad puesto que es una plantilla de texto.

Para obtener la tabla de frecuencias con la **Consola Gretl** o con un **fichero de comandos** se puede utilizar el comando **freq**, cuyo formato es:

**freq nombre de la variable**

## 2.12. ¿Cómo hacer representaciones gráficas?

Existen varias alternativas para hacer representaciones gráficas:

1. Utilizar el menú **Ver** de la **Barra de Menú** de la **Ventana Principal**.
2. Utilizar el comando **gnuplot**.
3. Utilizar el comando **scatters**.
4. Utilizar el comando **textplot**.

### 2.12.1. Utilizar el menú **Ver**

En el menú **Ver** existen dos submenús que se pueden utilizar para la construcción de gráficos (véase parte superior de la Ilustración 2-17):

<sup>18</sup> Se podrán modificar las características de este gráfico con la opción **Editar** del menú emergente que aparece al pinchar con el botón derecho del ratón encima del área de dicho gráfico (esta opción será tratada con detalle en el epígrafe relativo a las representaciones gráficas).

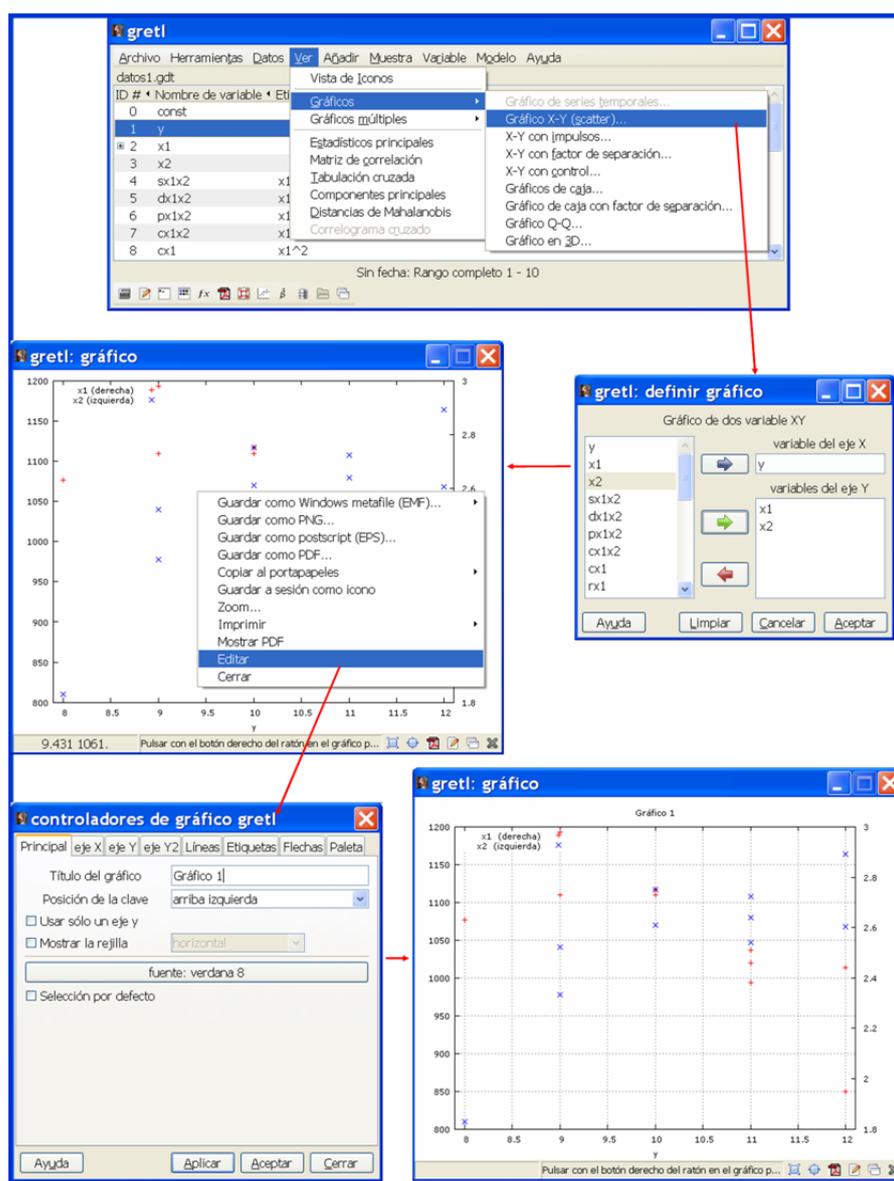


Ilustración 2-17. Ejemplo de construcción de un gráfico de dispersión utilizando el menú Ver.

- **Gráficos:** Con esta opción se pueden construir de una manera sencilla gráficos de dispersión por marcas, por impulsos, con factor de separación (variable ficticia), en tres dimensiones, de control, de caja y QQ.
- **Gráficos múltiples:** A diferencia de la opción anterior en la que se construye un solo gráfico, con esta opción el número de gráficos que se construyen va a depender:
  - En el caso de gráficos de dispersión, del número de variables que se seleccionen para representar en el eje X (eje horizontal) o en el eje Y (eje vertical). Si se selecciona una variable en el eje Y, Gretl proporciona un gráfico; si se seleccionan dos, Gretl proporciona dos gráficos y así sucesivamente.
  - En el caso de gráficos de series temporales, del número de variables que se seleccionen, puesto que a diferencia del caso anterior, en cada uno de estos gráficos, Gretl representa en el eje Y las variables seleccionadas y en el eje X el tiempo.

Para la construcción de gráficos utilizando el menú **Ver**, el usuario tendrá que hacer las selecciones oportunas en las ventanas o cuadros de diálogo que se van abriendo hasta llegar a visualizarlos en la **Ventana gráfico**.

Situándose encima del gráfico y presionando el botón derecho del ratón, aparece un menú emergente que permite, entre otras cosas:

- Guardar como imagen de Windows, PNG, postscript, PDF o copiarlo en el portapapeles.
- Guardar como icono.
- Imprimir.
- Editar. A través de la opción se puede acceder a la *Ventana Controladores de gráfico Gretl*, donde se pueden cambiar los atributos utilizando las pestañas que se van abriendo al seleccionar los botones de esta ventana (véase ejemplo de la Ilustración 2-17).

Hay que señalar que también se puede acceder a algunos de estos gráficos con el menú **Variable** de la **Barra de Menú** de la **Ventana Principal**.

La diferencia entre los comandos **gnuplot - scatters** y el comando **textplot** está en que con este último, los gráficos se muestran como una plantilla de texto dentro del output de Gretl, mientras que si se utilizan los dos primeros, los gráficos se muestran en una pantalla o ventana diferente.

Los gráficos obtenidos con los comandos **gnuplot - scatters** tienen mayor calidad que los obtenidos con el comando **textplot**, además sus características pueden ser fácilmente modificadas con el programa GNUPLOT y pueden incorporarse a un documento de texto y ser tratados como imágenes.

### 2.12.2. Utilizar el comando gnuplot

El formato del comando **gnuplot** es:

```
gnuplot v1 v2 ... vk --opciones
```

Debe de tenerse en cuenta que la última de las variables que interviene en el comando será representada en el eje horizontal.

Por ejemplo, si se ejecuta el comando:

```
? gnuplot y x1
```

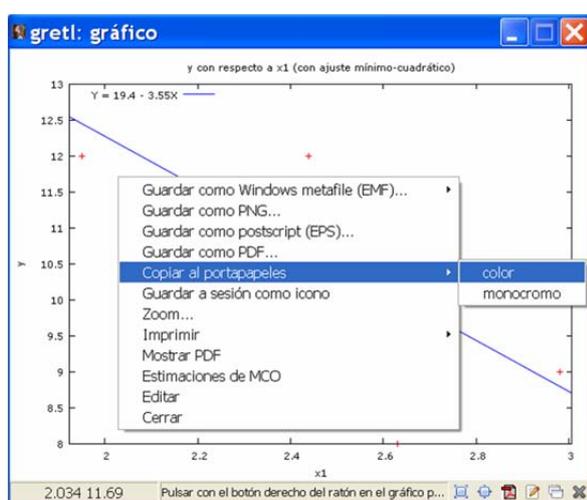


Ilustración 2-18. Menú emergente de la **Ventana gráfico**.

Se abre una nueva ventana "gnuplot graph" en la que aparece la representación gráfica requerida (en este caso, representa a la variable "y" respecto a la variable "x<sub>1</sub>"). Cuando en el comando **gnuplot** intervienen dos variables y se ejecuta sin opciones, Gretl proporciona por defecto la línea de ajuste mínimo-cuadrático.

Si el usuario se sitúa sobre el gráfico y pincha con el botón derecho del ratón, aparecen una serie de opciones que le permitirán modificar su aspecto, guardarlo como imagen, copiarlo en el portapapeles, ... (véase Ilustración 2-18).

Una forma sencilla de incorporar el gráfico a un documento de texto, sería con la opción "copiar al

portapapeles", ir al documento de texto y seleccionar Editar → Pegar y, el gráfico aparecerá en el lugar donde esté situado el cursor.

Hay que señalar que las opciones que permiten incorporar en el gráfico líneas de ajuste sólo están disponibles cuando en el comando intervengan únicamente dos variables y además, dichas opciones son incompatibles entre sí, es decir, Gretl no ejecutará ningún comando en el que intervenga más de una línea de ajuste, emitiendo el correspondiente mensaje de error (opciones incompatibles). Por ejemplo:

```
? gnuplot x1 y --linear-fit --quadratic-fit
Opciones incompatibles
```

Cuando se ejecuta un fichero de instrucciones entre las cuales figuran uno o varios comandos **gnuplot**, Gretl en vez de mostrar dichos gráficos de forma inmediata, abriendo las ventanas correspondientes, los guarda como ficheros con extensión “.plt” en el directorio de trabajo. Por defecto, cada vez que Gretl ejecuta un fichero de instrucciones en el que figuren dichos comandos, reenumera estos ficheros como “gptmpn<sup>o</sup>.plt” donde n<sup>o</sup> empieza en 01, por tanto, una vez terminada la ejecución del programa, es aconsejable cambiar el nombre de dichos ficheros.

En el Cuadro 2-6 se recogen algunas de las opciones disponibles con el comando **gnuplot**.

<b>--with-lines</b>	→	usa líneas.
<b>--with-lp</b>	→	usa líneas y puntos.
<b>--with-impulses</b>	→	usa líneas verticales.
<b>--time-series</b>	→	representa las variables respecto al orden de las observaciones.
<b>--suppress-fitted</b>	→	suprime la línea de ajuste o estimación.
<b>--linear-fit</b>	→	añade línea de ajuste del modelo $\hat{Y}_t = b_0 + b_1 X_t$
<b>--inverse-fit</b>	→	añade línea de ajuste del modelo $\hat{Y}_t = b_0 + b_1 (1/X_t)$
<b>--quadratic-fit</b>	→	añade línea de ajuste del modelo $\hat{Y}_t = b_0 + b_1 X_t + b_2 X_t^2$
<b>--cubic-fit</b>	→	añade línea de ajuste del modelo $\hat{Y}_t = b_0 + b_1 X_t + b_2 X_t^2 + b_3 X_t^3$
<b>--dummy</b>	→	representa a las variables respecto a los valores de una ficticia.
<b>--matrix=A</b>	→	representa las columnas de la matriz A.
...		

Cuadro 2-6. Algunas opciones del comando **gnuplot**.

Por ejemplo, si se ejecutan desde un fichero de comandos las siguientes instrucciones:

```
# Gráfico de dispersión de "y" respecto a "x1"
gnuplot y x1 --with-lines --time-series
gnuplot y x1 --with-impulses
gnuplot y x1 --linear-fit
gnuplot y x1 --quadratic-fit
gnuplot y x1 --inverse-fit

# Gráfico de dispersión de "y" respecto a "x2" de acuerdo con los valores de la variable ficticia "d1"
genr d1 = (x2>1000)
gnuplot y x2 d1 --dummy

# Gráfico de dispersión de la columna 2 de la matriz G respecto de la columna 1
matrix G = {x1, x2}
gnuplot 2 1 --matrix=G
```

En la **Ventana resultados de guión**, Gretl mostrará las siguientes salidas:

```
# Gráfico de dispersión de "y" respecto a "x1"
? gnuplot y x1 --with-lines --time-series
escribió c:\proyecto01\gpttmp01.plt
```

```
? gnuplot y x1 --with-impulses
escribió c:\proyecto01\gpttmp02.plt

? gnuplot y x1 --linear-fit
escribió c:\proyecto01\gpttmp03.plt

? gnuplot y x1 --quadratic-fit
escribió c:\proyecto01\gpttmp04.plt
? gnuplot y x1 --inverse-fit
escribió c:\proyecto01\gpttmp05.plt

# Gráfico de dispersión de "y" respecto a "x2" de acuerdo con los valores de la variable ficticia
"d1"
? genr d1 = (x2>1000)
Se ha generado la serie d1 (ID 5)
? gnuplot y x2 d1 --dummy
escribió c:\proyecto01\gpttmp06.plt

# Gráfico de dispersión de la columna 2 de la matriz G respecto de la columna 1
? matrix G = {x1, x2}
Se ha generado la matriz G
? gnuplot 2 1 --matrix=G
escribió c:\proyecto01\gpttmp07.plt
```

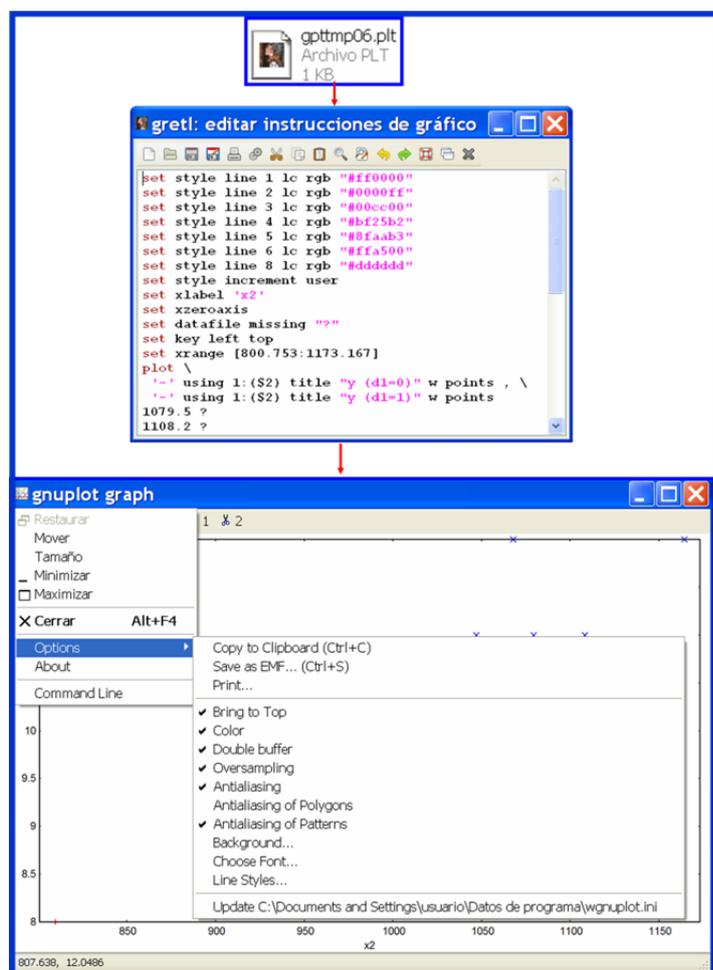


Ilustración 2-19. Secuencia para abrir v modificar un gráfico desde un fichero gnuplot.

Como puede observarse, Gretl crea siete ficheros gráficos en la carpeta de trabajo y los numera desde el 01 hasta el 07. Una alternativa para abrir y modificar dichos ficheros es hacer un doble clic en su nombre, acceder al conjunto de instrucciones que aparecen en el “editor de instrucciones del gráfico” y pinchar en el botón “ejecutar” que aparece en dicha ventana (véase Ilustración 2-19). En este caso, una vez que se ejecuta el fichero de instrucciones de gráfico, se abre el programa GNUPLLOT y para modificar los atributos del gráfico, el usuario debe estar familiarizado con dicho programa.

### 2.12.3. Utilizar el comando **scatters**

El comando **scatters** permite construir múltiples gráficos de dispersión con un solo comando. Su formato es:

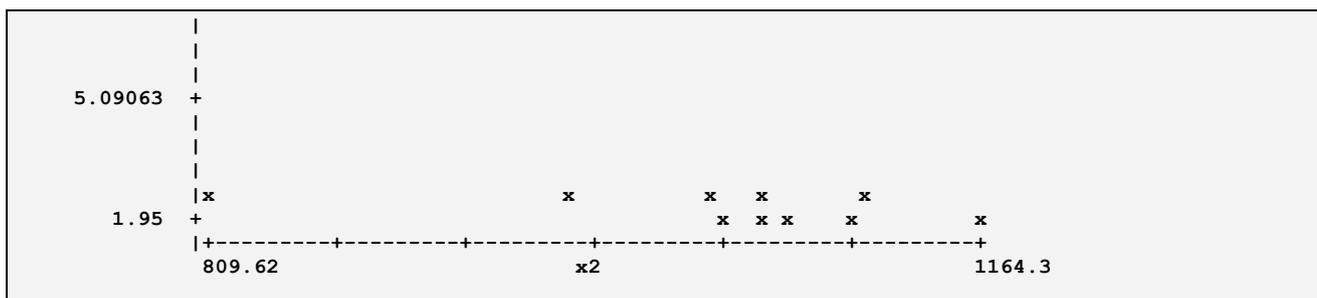
```
scatters v1 v2 ... vk
```

Debe de tenerse en cuenta que, por defecto, el comando **scatters**, representa a las variables respecto al orden de las observaciones y, por tanto, el número de gráficos que proporciona depende del número de variables que intervengan en

el comando. Nótese que dicho comando es equivalente a:

```
gnuplot v1 v2 ... vk --with-lines --time-series
```





Debe de tenerse en cuenta que la última de las variables que intervienen en el comando **textplot** será la representada en el eje horizontal.

Cuando en la representación gráfica se solapen puntos, Gretl lo indicará con un símbolo “+” (valores coincidentes para las variables representadas).

### 2.13. ¿Cómo modificar el rango muestral?

Existen varias alternativas para modificar el rango muestral:

1. Utilizar el menú **Muestra** de la **Barra de Menú** de la **Ventana Principal**.
2. Utilizar el comando **simpl**.

#### 2.13.1. Utilizar el menú Muestra

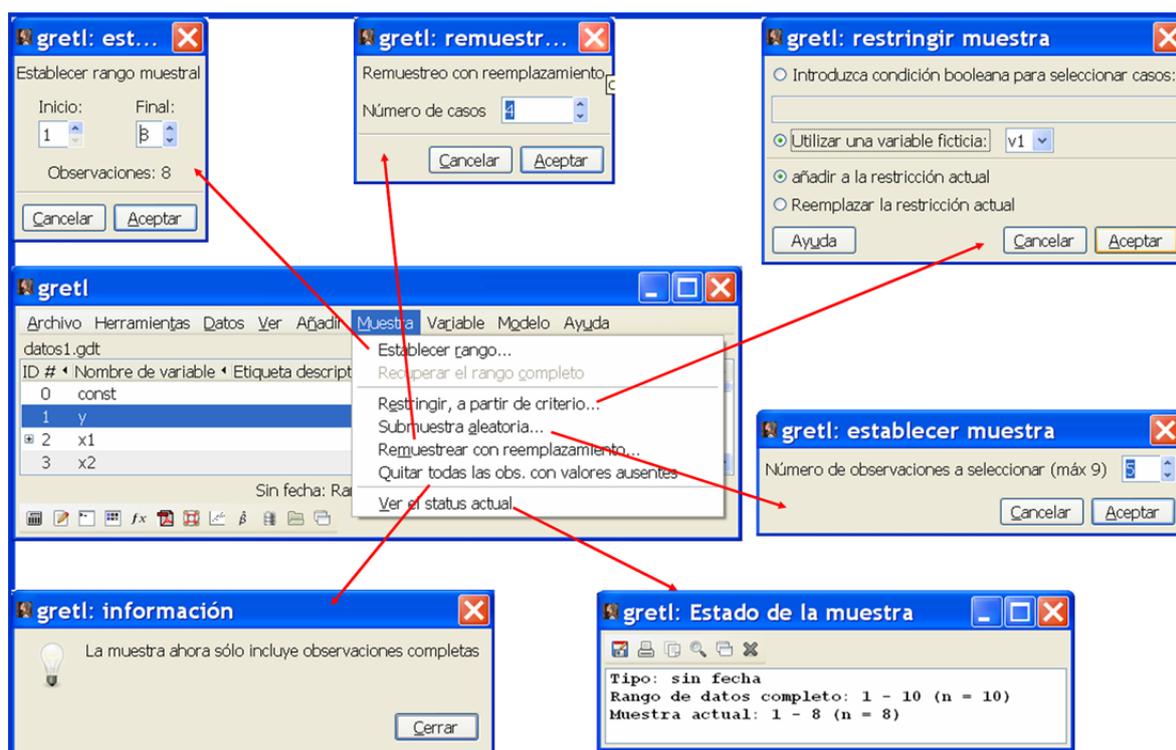


Ilustración 2-21. Opciones del menú **Muestra** de la **Ventana Principal**.

Con el menú **Muestra** se puede (véase Ilustración 2-21):

- Establecer un rango muestral inferior al número de observaciones disponibles.
- Recuperar el rango completo (que el rango a utilizar coincida con el número de observaciones disponibles).

- Restringir el rango muestral a las observaciones que cumplan una determinada condición. Como criterio de selección se puede utilizar una variable ficticia existente en la sesión de trabajo o escribir directamente en el cuadro de diálogo que se abre una nueva condición booleana para la selección de observaciones. Además, Gretl da la posibilidad de que los nuevos criterios de selección se añadan a los existentes o los reemplacen.
- Definir submuestras aleatorias con y sin reemplazamiento.
- Quitar todos los datos con valores ausentes. Cuando se selecciona esta opción, Gretl emite el mensaje “*La muestra ahora sólo incluye observaciones completas*”.
- Ver el status actual del rango muestral.

### 2.13.2. Utilizar el comando `smpl`

Existen varias formas de modificar el rango muestral utilizando el comando **smpl**:

**smpl** *primera observación última observación* → selecciona un rango muestral inferior al número de observaciones disponible.

**smpl --full** → recupera el rango muestral completo (coincidente con el número de observaciones disponibles).

**smpl d --dummy** → genera muestras a través de las observaciones no nulas de una variable ficticia d.

**smpl n --random** → genera muestras aleatorias de tamaño n (n tiene que ser inferior al número de observaciones disponibles).

**smpl --no-missing** → elimina los datos con valores ausentes.

Un ejemplo:

```
? open c:\proyecto01\datos1.gdt

Leer fichero de datos c:\proyecto01\datos1.gdt
periodicidad: 1, máx.obs: 10
rango de observaciones: 1-10

Listando 5 variables:
  0) const   1) index   2) y           3) x1           4) x2

# Se define un rango muestral desde la observación 2 hasta la 8
? smpl 2 8
Rango de datos completo: 1 - 10 (n = 10)
Muestra actual: 2 - 8 (n = 7)

? print y x1 x2 --byobs
      y           x1           x2
2      11          2.51         1108.20
3      10          2.73         1116.80
4       9          2.73          978.09
5       8          2.63          809.62
6      11          2.38         1047.10
7      12          2.44         1164.30
8      10          2.75         1069.60
```

Cada vez que se utilice un comando **smpl**, Gretl informa del rango de datos completo y de la muestra actual (en este caso, comienza en la observación 2 y termina en la 8).

```
? smpl --full
Rango de datos completo: 1 - 10 (n = 10)

# Se define un rango muestral que deja fuera las observaciones sin dato para alguna de las variables
? genr x1r3 = x1(-3)
Se ha generado la serie x1r3 (ID 5)

? smpl --no-missing
Rango de datos completo: 1 - 10 (n = 10)
```

```
Muestra actual: 4 - 10 (n = 7)

? print y x1 x2 x1r3 --byobs
      y      x1      x2      x1r3
4      9      2.73     978.09     2.46
5      8      2.63     809.62     2.51
6     11      2.38    1047.10     2.73
7     12      2.44    1164.30     2.73
8     10      2.75    1069.60     2.63
9      9      2.98    1040.30     2.38
10    12      1.95    1067.80     2.44
```

El nuevo rango muestral contiene 7 observaciones, desde la 4 hasta la 10, es decir, se dejan fuera las tres primeras observaciones, que coinciden con los datos ausentes de la única variable que contiene este tipo de datos (x1r3).

```
? smpl --full
Rango de datos completo: 1 - 10 (n = 10)

# Se define un rango muestral que utiliza como criterio de selección los valores no nulos de la
variable ficticia d1
? genr d1 = x1<2.5
Se ha generado la serie d1 (ID 6)

? smpl d1 --dummy
Conjunto de datos completo: 10 observaciones
Muestra actual: 4 observaciones

? print y x1 x2 d1 --byobs
      y      x1      x2      d1
1     11      2.46    1079.5      1
2     11      2.38    1047.1      1
3     12      2.44    1164.3      1
4     12      1.95    1067.8      1
```

El nuevo rango muestral está formado por las observaciones a las que le corresponde el valor uno de la variable ficticia (en este caso las cuatro primeras).

```
? smpl --full
Rango de datos completo: 1 - 10 (n = 10)

# Se define una muestra aleatoria de tamaño 5
? smpl 5 --random
Conjunto de datos completo: 10 observaciones
Muestra actual: 5 observaciones

? print y x1 x2 --byobs
      y      x1      x2
1     11      2.51    1108.20
2     10      2.73    1116.80
3      9      2.73     978.09
4     12      2.44    1164.30
5     12      1.95    1067.80

? smpl --full
Rango de datos completo: 1 - 10 (n = 10)
```

En la muestra aleatoria de tamaño 5 que se ha generado, el orden de las observaciones no tiene porque ser el mismo que en la muestra completa.

# Capítulo 3. MCO: MODELO DE REGRESIÓN LINEAL CLÁSICO

## 3.1. Presentación e Hipótesis Básicas del Modelo de Regresión Lineal Múltiple

El **Modelo de Regresión Lineal Múltiple (MRLM)** es el caso más simple de modelización econométrica, por lo que será el modelo de partida. Se denomina **lineal** porque la relación entre las variables es de tipo lineal y **múltiple** porque tiene una única ecuación y varias variables explicativas:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \varepsilon_t \quad \forall t=1, 2, \dots, T$$

Donde:

T: Tamaño muestral (número de observaciones disponibles).

$Y_t$ : Variable endógena, variable explicada o regresando.

$X_{0t}, X_{1t}, X_{2t}, \dots, X_{Kt}$ : Variables predeterminadas o regresores ( $X_{0t}$  es el regresor ficticio y  $X_{1t}, X_{2t}, \dots, X_{Kt}$  son las variables explicativas).

Las variables explicativas tienen dos subíndices, el primero (i) da nombre a la variable y el segundo (t) se refiere a la observación muestral, representando el tiempo si la serie es temporal y la unidad económica si la serie es atemporal.

$\varepsilon_t$ : Perturbación Aleatoria, variable no observable que representa el efecto de todos los factores no incluidos de forma explícita en el modelo.

$\beta_0, \beta_1, \beta_2, \dots, \beta_K$ : Parámetros, son los factores desconocidos cuyos valores se suponen constantes a lo largo de toda la muestra.

Aunque no es lo habitual, en algunos modelos puede no aparecer el parámetro que acompaña al regresor ficticio ( $\beta_0$ ), estando en este caso ante modelos formulados sin ordenada en el origen.

El Modelo de Regresión Lineal Múltiple se puede escribir matricialmente:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_t \\ \dots \\ Y_T \end{pmatrix}_{Tx1} = \begin{pmatrix} 1 & X_{11} & X_{21} & \dots & X_{K1} \\ 1 & X_{12} & X_{22} & \dots & X_{K2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1t} & X_{2t} & \dots & X_{Kt} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1T} & X_{2T} & \dots & X_{KT} \end{pmatrix}_{Tx(K+1)} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{pmatrix}_{(K+1)x1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \\ \dots \\ \varepsilon_T \end{pmatrix}_{Tx1} \Rightarrow Y = X\beta + \varepsilon$$

Donde:

Y: Vector columna de orden Tx1, que incluye las T observaciones del regresando.

X: Matriz de orden Tx(K+1), que contiene las observaciones de los regresores (el regresor ficticio y las K variables explicativas). A la matriz X se le denomina **Matriz de Diseño**.

$\beta$ : Vector columna de orden (K+1)x1, que contiene los K+1 parámetros del modelo.

$\varepsilon$ : Vector columna de orden Tx1, que contiene las perturbaciones del modelo.

Las hipótesis básicas de un Modelo de Regresión Lineal Múltiple son las siguientes:

### H<sub>1</sub>. Forma funcional lineal.

El valor esperado del regresando es una combinación lineal de los regresores, sin embargo, la relación que liga al regresando con los regresores es estocástica ya que aparece el término perturbación aleatoria, es decir, se trata de una relación lineal no exacta.

La hipótesis de linealidad se justifica por la facilidad de su tratamiento analítico y como primer paso para la especificación de formas funcionales más complicadas.

## H<sub>2</sub>. No existen errores de observación en las variables.

### H<sub>3</sub>. La perturbación es ruido blanco.

En terminología estadística una variable ruido blanco es una variable aleatoria que se caracteriza por tener **esperanza nula** ( $E\varepsilon_t = 0 \forall t$ ), **varianza constante** ( $E\varepsilon_t^2 = \sigma^2 \forall t$ ) y **covarianza nula** ( $E\varepsilon_t\varepsilon_s = 0 \forall t \neq s$ ), por lo que a este modelo también se le denomina **Modelo de Regresión Lineal Clásico (MRLC)**. En álgebra matricial, las hipótesis de variable ruido blanco se pueden escribir matemáticamente como:

$$\mathbf{H}_{3.1}. E\varepsilon_t = 0 \forall t \Rightarrow E\varepsilon = \mathbf{0}_{T \times 1}$$

Esta hipótesis supone que todos los factores no incluidos de forma explícita en el modelo y, por tanto, incluidos en el término perturbación, no producen efectos sistemáticos, al compensarse en promedio, los efectos positivos con los negativos.

$$\mathbf{H}_{3.2}. E\varepsilon\varepsilon' = \sigma^2 I_T = V$$

Siendo V una matriz simétrica y escalar de orden T x T, denominada matriz de varianzas-covarianzas de las perturbaciones. Sus elementos diagonales son las **varianzas** de los elementos del vector de perturbaciones y sus elementos no diagonales son las **covarianzas** entre dichos elementos.

Esta hipótesis recoge las hipótesis de homocedasticidad e incorrelación de las perturbaciones de un Modelo de Regresión Lineal Múltiple:

$$. E\varepsilon_t^2 = \sigma^2 \forall t \text{ (Hipótesis de Homocedasticidad).}$$

La varianza de la perturbación es constante e independiente de la observación de que se trate, es decir, los factores causales recogidos de forma implícita en la perturbación, actúan de manera análoga en cada observación.

$$. E\varepsilon_t\varepsilon_s = 0 \forall t \neq s \text{ (Hipótesis de Incorrelación entre las perturbaciones).}$$

Las covarianzas entre las distintas observaciones de la perturbación son nulas, lo que significa que no están correlacionadas entre sí, por ello, lo que ocurra en cada observación en los factores integrados en la perturbación no va a estar relacionado con lo que ocurra en las observaciones anteriores o posteriores.

Las hipótesis de esperanza nula y matriz de varianzas-covarianzas escalar, suelen expresarse conjuntamente con la denominación de **perturbación esférica**.

## H<sub>4</sub>. Hipótesis relativas a la matriz X:

$$\mathbf{H}_{4.1}. \text{Rango}(X) = K+1 \text{ (Condición de rango o hipótesis de rango pleno).}$$

Con esta condición se exige que el rango de la matriz X coincida con el número de columnas de dicha matriz (K+1), por lo que todas las columnas de la matriz X deben ser linealmente independientes (hipótesis de no colinealidad de los regresores). La independencia lineal entre los regresores del modelo hace posible aislar el efecto de cada uno de ellos.

Esta hipótesis afecta a la posibilidad de hacer la estimación del modelo, ya que es una condición necesaria para poder calcular la inversa de la matriz X'X.

$$\mathbf{H}_{4.2}. T > K+1.$$

Con esta condición se exige que el número de filas de la matriz  $X$  ( $T$ ) sea mayor que el número de columnas de dicha matriz ( $K+1$ ). Es una condición necesaria, aunque no suficiente, para poder abordar la estimación del modelo, pues garantiza que el número de observaciones de las variables sea mayor que el número de parámetros a estimar, con lo que se asegura el suficiente número de grados de libertad.

Además, es conveniente que el tamaño de la muestra ( $T$ ) sea grande, ya que ello contribuirá a la obtención de mejores estimadores de los parámetros.

#### H<sub>4.3</sub>. $X$ no es estocástica (Hipótesis de Exogeneidad).

Esta hipótesis supone que la matriz de regresores no varía al pasar de una muestra a otra, siendo una hipótesis que simplifica algunas demostraciones, aunque hay que destacar que las buenas propiedades de los estimadores de un modelo clásico se mantienen aunque se sustituya la hipótesis de regresores no estocásticos por la de regresores estocásticos pero independientes de la perturbación.

Las hipótesis que se acaban de indicar son suficientes para obtener estimadores puntuales de los parámetros del modelo. Como criterio de estimación se elegirá el método consistente en minimizar la Suma de Cuadrados de los Errores, denominado **Método Mínimo Cuadrático**, ya que bajo las hipótesis del Modelo de Regresión Lineal Múltiple proporciona los estimadores con las mejores propiedades. Los estimadores obtenidos por este método se denominan **Estimadores Mínimo Cuadráticos Ordinarios** (EMCO) de los parámetros del modelo.

### 3.2. ¿Cómo estimar por MCO en Gretl?

Para estimar por Mínimos Cuadrados Ordinarios se tienen dos opciones:

- Utilizar el menú **Modelo** de la **Barra de Menú** de la **Ventana Principal**.
- Utilizar el comando **ols**.

#### 3.2.1. Utilizar el menú Modelo

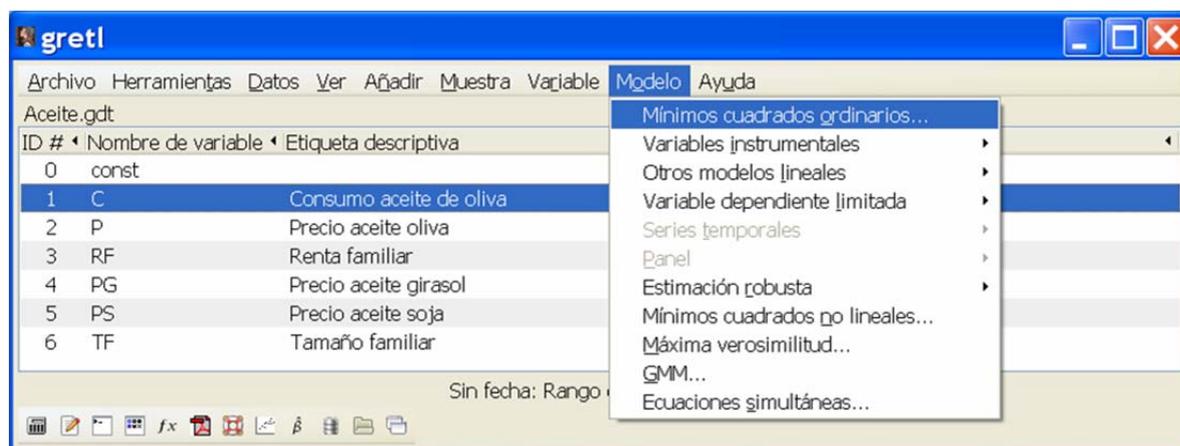


Ilustración 3-1. Menú **Modelo** de la **Barra de Menú** de la **Ventana Principal**.

Al seleccionar el menú **Modelo** de la **Barra de Menú** de la **Ventana Principal**, se abre un cuadro de diálogo que informa de las tareas que se pueden realizar utilizando este asistente (véase Ilustración 3-1), en el que se tiene que seleccionar el procedimiento que se desea utilizar (en este caso, **Mínimos Cuadrados Ordinarios**). Para seleccionarlo, basta con situarse encima del procedimiento deseado (queda sombreado en un color más oscuro) y al hacer clic con el botón izquierdo del ratón se abre el cuadro de diálogo **especificar modelo** (véase Ilustración 3-2), donde en el recuadro situado a la izquierda, Gretl informa al usuario de las variables disponibles para realizar la estimación.

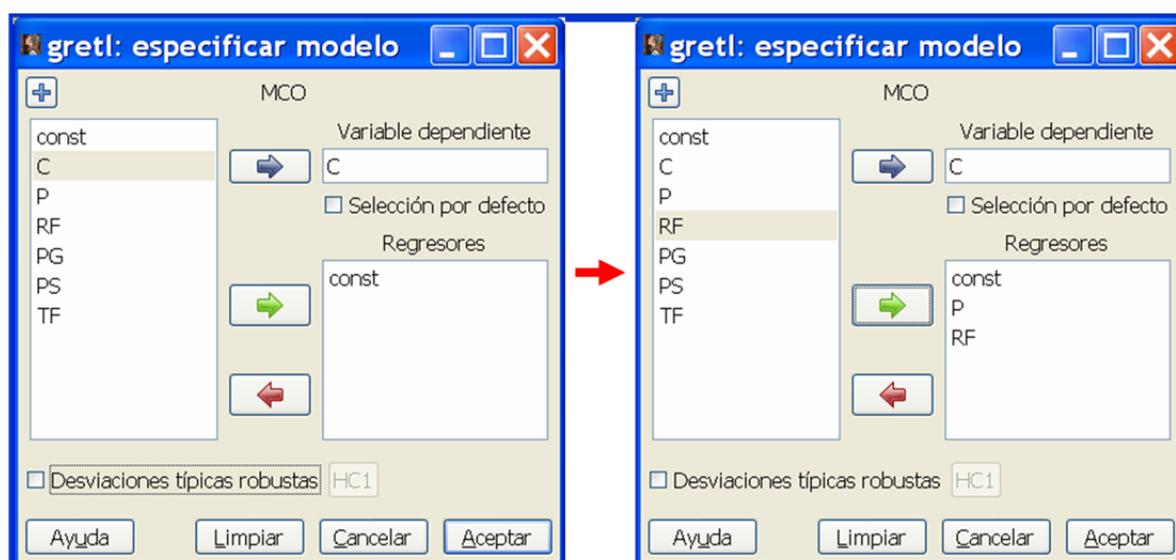


Ilustración 3-2. **Ventana especificar modelo**: selección de variables.

Para seleccionar la variable dependiente o variable a explicar por el modelo, el usuario debe elegir una variable situándose con el cursor encima y haciendo clic en la “flecha azul” situada a la izquierda del recuadro “variable dependiente” y dicha variable aparecerá en dicho recuadro. En caso de equivocación al hacer la selección, se elige la variable correcta, se hace clic en la flecha y dicha variable aparecerá como endógena.

Para seleccionar las variables independientes o variables explicativas del modelo el procedimiento es similar, el usuario debe elegir la/s variable/s situándose con el cursor encima y hacer clic en la “flecha verde” situada a la izquierda del recuadro “variables independientes” y, dicha/s variable/s aparecerá/n en el recuadro “variables independientes. En caso de equivocación al hacer la selección, se deben seleccionar en el recuadro “variables independientes” y hacer clic en la “flecha roja”, con lo que la/s variable/s dejará/n de ser variable/s explicativa/s.

En el recuadro “variables independientes” aparece seleccionado por defecto el **regresor ficticio**, que Gretl etiqueta como “**const**”, por lo que si se quiere realizar una estimación sin ordenada en el origen, será necesario situarse con el cursor encima y hacer clic en la “flecha roja”.



Ilustración 3-3. Menú **Muestra** de la **Barra de Menú** de la **Ventana Principal**: establecer rango muestral.

Antes de seleccionar las variables que intervienen en la estimación, se debe especificar el rango muestral que se desea utilizar para dicha tarea. La estimación del modelo se realizará con toda la muestra, salvo que se especifique lo contrario utilizando el menú **Muestra** de la **Barra de Menú** de la

**Ventana Principal** (véase Ilustración 3-3). Con dicho menú se accede a la ventana **Establecer rango**, en la se pueden modificar las observaciones inicial y final del rango muestral a utilizar en la estimación del modelo.

Seleccionadas las variables en la ventana **especificar modelo**, se hace clic en “Aceptar” y se abre la ventana **Modelo**, donde aparece la información básica para el análisis de los resultados de la estimación MCO. Además, a través del menú de dicha ventana, se pueden guardar algunos resultados de la estimación (botón **Guardar**), realizar contrastes relativos tanto a los parámetros del modelo como a las hipótesis del mismo (botón **Contrastes**), construir gráficos (botón **Gráficos**), realizar predicciones (botón **Análisis**), etc.

Utilizando el botón “**Guardar**”, se pueden guardar con la denominación deseada algunos resultados de la estimación, que podrán ser utilizados para nuevos cálculos. Al realizar los cálculos se debe de tener en cuenta la distinta naturaleza de dichos resultados: series, escalares, ... (véase Ilustración 3-4). Por defecto Gretl proporciona una denominación y descripción para cada una de las series y/o escalares que se guarden, que podrá ser cambiada por el usuario. Dado que en una misma sesión de trabajo se puede estimar más de un modelo econométrico, Gretl enumera dichos modelos y sus estadísticos asociados de forma correlativa empezando en “1”.

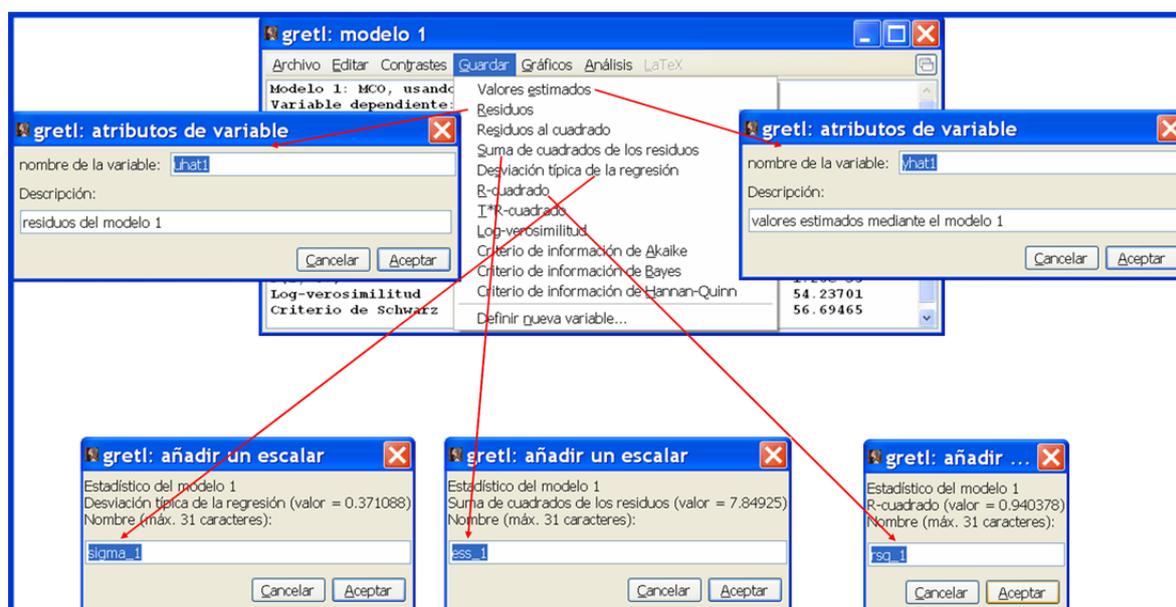


Ilustración 3-4. Menú **Guardar** de la **Ventana Modelo**.

En el ejemplo recogido en la Ilustración 3-4, los comandos que aparecen en el **Historial de instrucciones** y que podrían ser guardados en un fichero de comandos y ejecutados posteriormente son:

```
# modelo 1
ols C const P RF PG
series yhat1 = $yhat
series uhat1 = $uhat
scalar ess_1 = $ess
scalar sigma_1 = $sigma
scalar rsq_1 = $rsq
```

Se le ha indicado a Gretl que estime por MCO un modelo formulado con ordenada en el origen, donde la variable a explicar es “C” y las variables explicativas son “P, RF y PG” y que guarde las series de valores estimados del regresando y de residuos, además de los escalares Suma de Cuadrados de Errores, estimador de la desviación típica de la perturbación y coeficiente de determinación.

Cuando se muestren los resultados guardados con un comando **print**, estos tendrán más decimales que en la salida estándar del comando **ols**, ya que en esta última, por razones de espacio, Gretl ajusta el número de decimales a mostrar.

### 3.2.2. Utilizando el comando **ols**

Otra opción para obtener la estimación MCO es escribir el comando **ols** en la **Consola Gretl** o en un **Fichero de Comandos**, para lo cual el usuario debe conocer su formato de escritura y las opciones disponibles.

El formato del comando **ols** es:

**ols depvar indepvars --opciones**

donde *depvar* es la variable dependiente e *indepvars* es la lista de variables independientes, entre las cuales se incluye el regresor ficticio (Gretl lo etiqueta como **const** y lo sitúa en primer lugar). En el Cuadro 3-1 aparece recogida una breve descripción de algunas de las opciones disponibles con el comando **ols**.

<b>--vcv</b>	→ Imprime la matriz de varianzas-covarianzas estimada de los estimadores.
<b>--simple-print</b>	→ No imprime los estadísticos auxiliares.
<b>--quiet</b>	→ Suprime la impresión de resultados.
<b>--anova</b>	→ Imprime la tabla ANOVA.
<b>--no-df-corr</b>	→ Utiliza el tamaño muestral ( <i>T</i> ) en lugar de los grados de libertad ( <i>DF</i> ) para el cálculo de determinados estadísticos.

Cuadro 3-1. Opciones del comando **ols**.

Cuando se ejecutan algunos comandos, Gretl guarda en memoria una serie de “variables temporales”, que reciben esta denominación porque sólo están disponibles de forma temporal hasta que se ejecute un nuevo comando que lleve asociadas dichas variables. Son variables cuya denominación empieza por \$ y, una forma rápida de acceder a ellas es ejecutar el comando **varlist --accessors** después del comando **ols**.

En el Cuadro 3-2 aparece recogida una breve descripción de algunas de las variables temporales asociadas al comando **ols**, que aparecen en la salida del comando **varlist --accessors** agrupadas bajo la denominación “**relativo al modelo**”. En dicha salida, además de la denominación que Gretl da a cada una de dichas variables, se informa de si se trata de una serie, una matriz, un grupo de variables o un escalar, proporcionando también su valor numérico en este último caso.

<b>\$ess</b> → Suma de Cuadrados de Errores (SCE).	<b>\$T</b> → Tamaño muestral ( <i>T</i> ).
<b>\$rsq</b> → Coeficiente de determinación ( $R^2$ ).	<b>\$df</b> → Grados de libertad ( <i>DF</i> ).
<b>\$ncoeff</b> → Número de parámetros.	<b>\$lnl</b> → Logaritmo de la función de verosimilitud.
<b>\$aic</b> → Criterio de información de Akaike (AIC).	<b>\$bic</b> → Criterio bayesiano de Schwarz (SC).
<b>\$hq</b> → Criterio de Hannan y Quinn (HQC).	<b>\$rsq</b> → $T \cdot R^2$
<b>\$Fstat</b> → Estadístico <i>F</i> ( $F_2$ en modelos formulados con ordenada en el origen y $F_1$ en modelos sin ordenada).	<b>\$yhat</b> → Serie de valores estimados del regresando ( $\hat{Y}$ ).
<b>\$uhat</b> → Serie de residuos o errores de estimación ( <i>e</i> ).	
<b>\$coeff</b> → Vector de estimadores de los parámetros ( <i>b</i> ).	
<b>\$sigma</b> → Estimador de la desviación estándar de la perturbación ( <i>S</i> ).	
<b>\$vcv</b> → Matriz de varianzas-covarianzas estimada de los estimadores $\hat{V}(b)$ .	
<b>\$stderr</b> → Vector de estimadores de las desviaciones típicas de los estimadores de los parámetros.	
<b>\$xlist</b> → Lista de variable/s independiente/s.	<b>\$ylist</b> → Lista variable/s dependiente/s.

Cuadro 3-2. Variables temporales asociadas al comando **ols**.

Si el usuario quiere disponer de estas variables en cualquier momento será necesario que las guarde, lo que puede hacer utilizando un comando **genr**.

### 3.3. Análisis de la información proporcionada para una estimación MCO de un modelo formulado con ordenada en el origen

La información básica para el análisis de los resultados de la estimación MCO de un modelo formulado con ordenada en el origen se puede obtener con la **Ventana Modelo**, ejecutando un comando **ols** en la **Consola Gretl** o ejecutando un **Fichero de comandos**:

**ols** *Y const*  $X_1 X_2 \dots X_k$  *--opciones*

En econometría es frecuente denominar a la variable dependiente con la letra Y y a las variables independientes con la letra X con distintos subíndices numéricos para distinguirlas. Gretl denomina al “regresor ficticio” como “const” y, por defecto, lo selecciona y lo sitúa en el primer lugar de la lista de variables independientes, no obstante, el usuario puede modificar su posición e incluso dejar de seleccionarlo.

Gretl proporciona las salidas de estimación y algunos estadísticos asociados<sup>19</sup> (véase **Tabla 3-1**).

**Tabla 3-1.** Salida asociada al submenú **Mínimos Cuadrados Ordinarios** del menú **Modelo** de la **Ventana Principal** para la estimación MCO de un modelo formulado con ordenada en el origen.

Modelo n°: MCO, usando las observaciones n° observación inicial – n° observación final ( $n = n^\circ$ )  
**(ST)** Variable dependiente: Y (**Sdepvar**)

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
<i>const</i>	$b_0$	$S_{b_0}$	$t_0$	$PV_{t_0}$	
$X_1$	$b_1$	$S_{b_1}$	$t_1$	$PV_{t_1}$	*
$X_2$	$b_2$	$S_{b_2}$	$t_2$	$PV_{t_2}$	**
...	...	...	...	...	
$X_K$	$b_K$	$S_{b_K}$	$t_K$	$PV_{t_K}$	***
Media de la vble. Dep	$\bar{Y}$		D.T. de la vble. dep.	$S_Y$	
Suma de cuad. residuos	$SCE$ ( <b>Sess</b> )		D.T. de la regresión	$S$ ( <b>Ssigma</b> )	
R-cuadrado	$R^2$ ( <b>Srsq</b> )		R-cuadrado corregido	$\bar{R}^2$	
F(K, T-K-1)	$F_2$ ( <b>SFstat</b> )		Valor p (de F)	$PV_{F_1}$	
Log-verosimilitud	$L$ ( <b>Slnl</b> )		Criterio de Akaike	<b>Saic</b>	
Criterio de Schwarz	$SC$ ( <b>Sbic</b> )		Crit. de Hannan-Quinn	<b>Sbqc</b>	

En la **Tabla 3-1** aparecen sombreadas algunas de las variables temporales asociadas a la salida de una estimación MCO. Esta salida proporciona información sobre el método de estimación empleado, el rango muestral, el tamaño muestral y el nombre de la variable dependiente. Como en una sesión de trabajo se puede estimar más de un modelo econométrico, para facilitar la identificación de las salidas, Gretl las etiqueta correlativamente empezando en “1”, como “**Modelo n°**”.

<sup>19</sup> Para facilitar el seguimiento por parte del lector, en cada uno de los capítulos se analizarán sólo los estadísticos directamente relacionados con el tema a tratar.

Coefficiente=Coeficiente estimado ( $b_i$ )

$$b = (X'X)^{-1}X'Y = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_K \end{pmatrix}_{(K+1) \times 1} \Rightarrow b_i$$

Los **coeficientes estimados** recogen el valor de los estimadores de los parámetros que acompañan a los regresores. Cada uno de estos coeficientes indica el cambio que experimenta la variable explicada ante un cambio unitario de la variable explicativa a la que acompaña dicho coeficiente, suponiendo que el resto de las variables se mantienen constantes.

Gretl bajo la denominación “const” proporciona el estimador de la ordenada en el origen (que normalmente no tiene interpretación económica) y demás cálculos asociados.

Desv. Típica=Desviación típica estimada de los estimadores (error estándar) ( $S_{b_i}$ )

$$\widehat{V}(b) = S^2(X'X)^{-1} = \begin{pmatrix} S_{b_0}^2 & \dots & \dots & \dots & \dots \\ S_{b_1b_0} & S_{b_1}^2 & \dots & \dots & \dots \\ S_{b_2b_0} & S_{b_2b_1} & S_{b_2}^2 & \dots & \dots \\ \dots & \dots & \dots & \ddots & \dots \\ S_{b_Kb_0} & S_{b_Kb_1} & S_{b_Kb_2} & \dots & S_{b_K}^2 \end{pmatrix}_{(K+1) \times (K+1)} \Rightarrow S_{b_i} = \sqrt{S_{b_i}^2}$$

Las **desviaciones típicas estimadas** de los estimadores miden la precisión con la que son estimados los parámetros, es decir, indican el “grado de confianza” de las estimaciones (siempre que los estimadores sean insesgados).

La matriz de varianzas-covarianzas estimada de los estimadores es una matriz de orden  $(K+1) \times (K+1)$ , cuyos elementos diagonales son las varianzas estimadas de los estimadores mínimo cuadráticos ordinarios y cuyos elementos no diagonales son las covarianzas estimadas entre dichos estimadores.

Estadístico t=Ratio t

$$t_i = \frac{b_i}{S_{b_i}} \quad \forall i=0,1,2, \dots, K$$

Las **ratios t** o **estadísticos t** se definen como el cociente entre los estimadores y sus desviaciones típicas estimadas. Consideradas en valor absoluto, serán indicadores de la “fiabilidad” de los estimadores, de modo que a mayor cuantía de dicho valor mayor fiabilidad del estimador.

Como se verá en el capítulo dedicado a contrastes de hipótesis, estos estadísticos van a permitir contrastar la hipótesis de nulidad individual de los parámetros, es decir, verificar si la variable a la que acompañan es individualmente significativa para la explicación del regresando. Además, Gretl proporciona los **p-valores** asociados a dichos estadísticos t y, etiqueta con un triple asterisco aquellos estimadores que son estadísticamente significativos al nivel del 1 por ciento, con un doble asterisco aquellos que lo son entre más del 1 y el 5 por ciento y con un asterisco indica la significatividad entre más del 5 y el 10 por ciento.

*Media de la vble dep = Media de la variable dependiente*

$$\bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$$

*D.T. de la vble dep = Desviación típica de la variable dependiente*

$$S_Y = \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y})^2$$

Las características básicas de la variable dependiente son recogidas por su media y su varianza o cuasivarianza<sup>20</sup>.

*Suma de cuad. residuos = Suma de Cuadrados de los Errores*

$$SCE = \sum_{t=1}^T e_t^2 = e'e$$

El método de Mínimos Cuadrados Ordinarios se basa en la minimización de la **Suma de Cuadrados de los Errores**, es decir, en este método los estimadores se obtienen de tal modo que las diferencias entre el valor observado y estimado del regresando son las menores posibles.

D.T. de la regresión=Estimador de la desviación típica de la perturbación o desviación residual

$$S = \sqrt{S^2} = \sqrt{\frac{SCE}{T-K-1}}$$

El **estimador de la varianza de la perturbación** es una medida que sirve para analizar la capacidad explicativa del modelo, ya que es el error cometido en la estimación ponderado por los grados de libertad del modelo.

R-cuadrado=Coeficiente de determinación

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT}$$

El **coeficiente de determinación**<sup>21</sup> es una de las medidas de bondad de ajuste más utilizada, mide la proporción en que la varianza muestral del regresando (SCT/T) es explicada por la varianza del regresando estimado (SCR/T), por tanto, indica en que medida el modelo explica las variaciones del regresando.

Esta medida de bondad de ajuste presenta dos ventajas importantes, es invariante ante cambios de escala o unidades de medida y posee limite inferior y superior ( $0 \leq R^2 \leq 1$ ), tomando el valor uno cuando el ajuste es perfecto y cero cuando es pésimo.

Una de las desventajas del coeficiente de determinación es que se incrementa siempre que se incluyan nuevas variables explicativas en el modelo, aún cuando éstas sean irrelevantes para explicar el comportamiento del regresando. El **coeficiente de determinación corregido** penaliza la inclusión de este tipo de variables, por lo que resulta más adecuado que el coeficiente de determinación para comparar la bondad de ajuste de modelos con distinto número de variables explicativas.

R-cuadrado corregido=Coeficiente de determinación corregido o ajustado

$$\bar{R}^2 = 1 - \frac{\frac{SCE}{T-K-1}}{\frac{SCT}{T-1}} = 1 - \frac{T-1}{T-K-1} (1-R^2)$$

El  $\bar{R}^2$  toma el valor uno cuando el ajuste es perfecto, pero no está acotado inferiormente, pudiendo tomar valores negativos cuando el ajuste realizado es muy malo.

<sup>20</sup> Debe recordarse que en un modelo formulado con ordenada en el origen la media del regresando coincide con la media de su valor estimado  $\bar{Y} = \bar{\hat{Y}}$ .

<sup>21</sup> Debe de tenerse en cuenta que la última expresión del coeficiente de determinación sólo es correcta cuando el modelo está formulado con ordenada en el origen, ya que se obtiene aplicando la descomposición de la Suma de Cuadrados Totales (SCT) en Suma de Cuadrados de los Errores (SCE) y Suma de Cuadrados de la Regresión (SCR), conocida como descomposición de la varianza.

El valor del coeficiente de determinación ajustado será siempre menor que el del coeficiente de determinación, ya que está corregido o ajustado por los grados de libertad.

### 3.4. Análisis de los residuos

Una vez estimado el modelo, es importante realizar un diagnóstico del mismo. El análisis de los residuos MCO juega un papel fundamental en esta tarea (puede ayudar a determinar la existencia de valores atípicos, la presencia de problemas de autocorrelación, heterocedasticidad, etc).

La salida asociada al submenú **Mostrar variable observada, estimada, residuos** del menú **Análisis** de la **Ventana Modelo** proporciona información sobre el rango de estimación, el estimador de la desviación típica de la perturbación y, para cada una de las observaciones muestrales, el valor de la variable dependiente, su valor estimado y el error cometido en su estimación (véase **Tabla 3-2**). Además, tal y como puede verse en la parte inferior de la **Tabla 3-2** proporciona algunos estadísticos para la evaluación de la predicción que no serán analizados en este capítulo.

Tabla 3-2. Salida asociada al submenú <b>Mostrar variable observada, estimada, residuos</b> del menú <b>Análisis</b> de la <b>Ventana Modelo</b> .			
Rango de estimación del modelo: n° observación inicial – n° observación final			
Desviación típica de la regresión = $S$			
Observaciones	Y	estimada	residuo
1	$Y_1$	$\hat{Y}_1$	$e_1$
2	$Y_2$	$\hat{Y}_2$	$e_2$
3	$Y_3$	$\hat{Y}_3$	$e_3$
...	...	...	...
T	$Y_T$	$\hat{Y}_T$	$e_T$
Estadísticos de evaluación de la predicción			
Error medio		EM	
Error cuadrático medio		ECM	
Raíz del Error cuadrático medio		RECM	
Error absoluto medio		EAM	
Porcentaje del error medio		PEM	
Porcentaje del error absoluto medio		PEAM	
U de Theil		U	

Se denomina **residuo** o **error** ( $e_t$ ) a la diferencia entre el valor observado del regresando ( $Y_t$ ) y su valor estimado por el modelo ( $\hat{Y}_t$ ).

*Nº de observación (t)*  
*Y = Valor observado del regresando ( $Y_t$ )*  
*estimada = Valor estimado del regresando ( $\hat{Y}_t$ )*  
 $\hat{Y}_t = X'_t b \Rightarrow \hat{Y} = X' b$   
*residuo = Error de estimación*  
 $e_t = Y_t - \hat{Y}_t \Rightarrow e = Y - \hat{Y}$

Recuérdese que en modelos formulados con ordenada en el origen, la media residual, es decir, el error medio es nulo.

### 3.5. Análisis de las sumas de cuadrados: tabla ANOVA

Las **Tablas ANOVA** o **Tablas de Análisis de la Varianza** proporcionan la descomposición de la variabilidad de la variable dependiente en dos componentes, la información recibida de los regresores y la recibida de los residuos. En la **Tabla 3-3** se recoge la salida asociada al submenú **ANOVA** del menú **Análisis de la Ventana Modelo**.

**Tabla 3-3.** Salida asociada al submenú ANOVA del menú Análisis de la Ventana Modelo.

Análisis de Varianza:

	Suma de cuadrados	gl	Media de cuadrados
Regresión	$SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{\hat{Y}})^2 = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$	K	$\frac{SCR}{K}$
Residuo	$SCE = \sum_{t=1}^T e_t^2 = e'e = \sum_{t=1}^T (e_t - \bar{e})^2$	T-K-1	$\frac{SCE}{T-K-1}$
Total	$SCT = \sum_{t=1}^T (Y_t - \bar{Y})^2$	T-1	$\frac{SCT}{T-1}$

$$R^2 = \frac{SCR}{SCT}$$

$$F_2 = \frac{\frac{SCR}{K}}{\frac{SCE}{T-K-1}} \quad PV_{F_2}$$

En la tabla **ANOVA** aparece recogida la siguiente información:

- La Suma de Cuadrados de la Regresión (SCR), que es la suma de cuadrados de las desviaciones de los valores estimados del regresando ( $\hat{Y}_t$ ) respecto a su media muestral ( $\bar{\hat{Y}}$ ). Por estar el modelo formulado con ordenada en el origen, la media del regresando estimado coincide con la media del regresando ( $\bar{\hat{Y}} = \bar{Y}$ ) y, por tanto, se puede calcular como  $SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$ .
- La Suma de Cuadrados de los Errores (SCE), que es la suma de residuos al cuadrado. En modelos formulados con ordenada en el origen coincide con la suma de cuadrados de las desviaciones de los valores de los residuos ( $e_t$ ) respecto a su media muestral ( $\bar{e}$ ), dado que, en este caso, la media de los residuos es nula.
- La Suma de Cuadrados Totales (SCT), que es la suma de cuadrados de las desviaciones de los valores observados del regresando ( $Y_t$ ) respecto a su media muestral ( $\bar{Y}$ ).

Hay que señalar que la tabla ANOVA sólo está disponible si el modelo está formulado con ordenada en el origen. En la penúltima fila se proporciona el cálculo del coeficiente de determinación<sup>22</sup> y en la última, el cálculo del estadístico  $F_2$  y su p-valor (que serán comentados en el capítulo dedicado a contrastes de hipótesis).

### 3.6. Análisis gráfico: menú Gráficos de la Ventana Modelo

Para tener una idea inicial de la “calidad” de la estimación, puede resultar interesante la representación gráfica de los valores del regresando y del regresando estimado, así como de los residuos. El menú **Gráficos** de la **Ventana Modelo** de la **Estimación mínimo cuadrática** proporciona entre otras representaciones gráficas: una de los residuos y otra de los valores observados y estimados de la variable dependiente del modelo.

Para modificar las propiedades de dichos gráficos es necesaria su edición, para lo que se hace clic en cualquier parte del gráfico con el botón derecho del ratón y, en el menú emergente se selecciona **Editar**, con lo que se abrirá la **Ventana controladores de gráficos de Gretl**, en la cual, de una manera relativamente sencilla, se podrán cambiar algunos de sus atributos: títulos, colores, rango de ejes, añadir líneas, etc.

El gráfico de valores observados y estimados permite analizar, de forma rápida, en que medida los valores estimados por el modelo se ajustan a los valores observados. Por ejemplo, a la vista del gráfico de la Ilustración 3-5, el modelo parece recoger de forma bastante razonable la evolución de la variable endógena, ya que no se observan grandes errores.

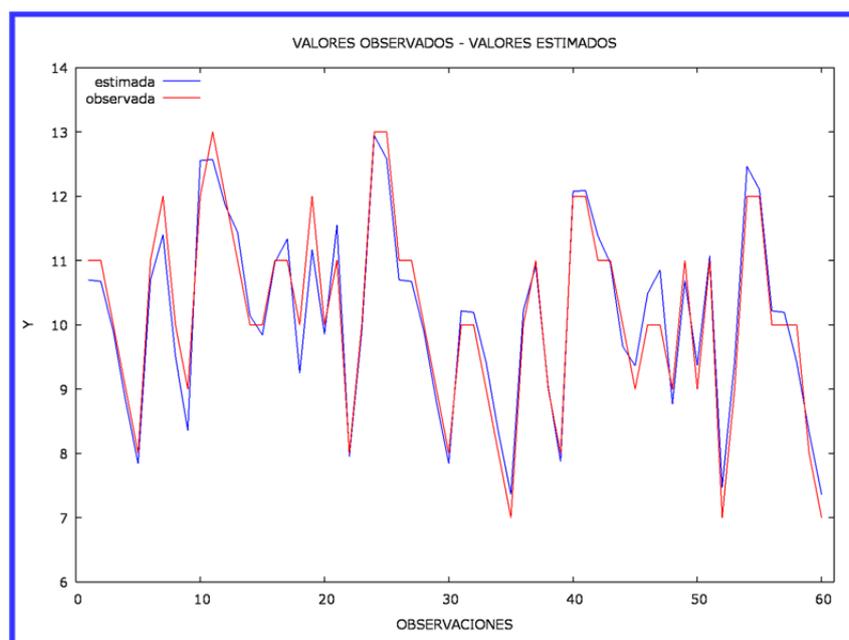


Ilustración 3-5. Gráfico valores observados – valores estimados.

Para facilitar el análisis del gráfico de los residuos resulta conveniente trazar unas líneas adicionales paralelas al eje horizontal:

<sup>22</sup> Se puede comprobar, que en modelos formulados con ordenada en el origen SCT es igual a SCE más SCR, lo que hace que el coeficiente de determinación esté acotado entre cero y uno.

- Una línea que pase por el valor medio de los residuos (cero ya que el modelo está formulado con ordenada en el origen). Esta línea permitirá ver de forma rápida en que observaciones se comete un error de estimación mayor y si el error cometido es positivo o negativo.
- Dos líneas para definir una banda de variación que represente más-menos dos veces la desviación típica. Dicha banda permitirá analizar de forma sencilla si los residuos se distribuyen de acuerdo con una normal, en cuyo caso sus valores estarían con un 95% de probabilidad comprendidos dentro de esta banda.

Además, el gráfico de residuos puede ayudar a determinar la existencia de valores u observaciones anómalas. Por ejemplo, observando el gráfico de la Ilustración 3-6, dos de los errores supera la banda de confianza y quizás sea necesario darle un tratamiento diferenciado del resto. Además, este gráfico, puede ayudar a sospechar problemas de autocorrelación (se analizará en el capítulo correspondiente).

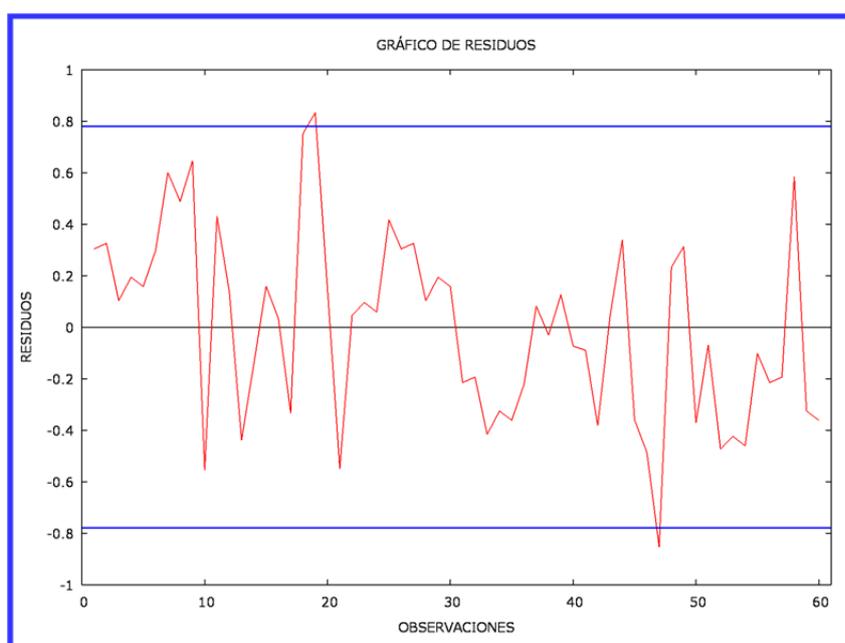


Ilustración 3-6. Gráfico de residuos.

Aunque se trata de análisis muy intuitivos y poco exactos, pueden dar pistas sobre posibles problemas que se pueden plantear por el incumplimiento de las hipótesis básicas, pero difícilmente serán concluyentes.

### 3.7. Estimación MCO de un modelo formulado sin ordenada en el origen

Aunque no es lo habitual, el modelo puede formularse sin ordenada en el origen:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \varepsilon_t \quad \forall t = 1, 2, \dots, T$$

La diferencia con respecto al modelo formulado con ordenada en el origen (epígrafe 3.3. ) es que no aparece el parámetro  $\beta_0$ .

Matricialmente este modelo se puede escribir:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_t \\ \dots \\ Y_T \end{pmatrix}_{Tx1} = \begin{pmatrix} X_{11} & X_{21} & \dots & X_{K1} \\ X_{12} & X_{22} & \dots & X_{K2} \\ \dots & \dots & \dots & \dots \\ X_{1t} & X_{2t} & \dots & X_{Kt} \\ \dots & \dots & \dots & \dots \\ X_{1T} & X_{2T} & \dots & X_{KT} \end{pmatrix}_{TxK} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{pmatrix}_{Kx1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \\ \dots \\ \varepsilon_T \end{pmatrix}_{Tx1} \Rightarrow Y = X\beta + \varepsilon$$

Para estimar un modelo sin ordenada en el origen, en el comando **ols** no se debe incluir como variable independiente el regresor ficticio (**const**).

**ols** Y  $X_1 X_2 \dots X_k$  --opciones

Aparentemente, las salidas del comando **ols** para un modelo con y sin ordenada en el origen son similares (véase **Tabla 3-4**), sin embargo, los cálculos de algunos de los estadísticos difieren de manera importante, cuestión que es necesario tener en cuenta para realizar de forma correcta las interpretaciones.

**Tabla 3-4.** Salida asociada al submenú **Mínimos Cuadrados Ordinarios** del menú **Modelo** de la **Ventana Principal** para la estimación MCO de un modelo formulado sin ordenada en el origen.

Modelo n°: MCO, usando las observaciones n° observación inicial – n° observación final ( $n = n^\circ$ ) (**ST**) Variable dependiente: Y (**Sdepvar**)

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
$X_1$	$b_1$	$S_{b_1}$	$t_1$	$PV_{t_1}$	*
$X_2$	$b_2$	$S_{b_2}$	$t_2$	$PV_{t_2}$	**
...	...	...	...	...	...
$X_K$	$b_K$	$S_{b_k}$	$t_K$	$PV_{t_K}$	***
Media de la vble. Dep	$\bar{Y}$		D.T. de la vble. dep.	$S_Y$	
Suma de cuad. residuos	$SCE$ ( <b>Sess</b> )		D.T. de la regresión	$S$ ( <b>Ssigma</b> )	
R-cuadrado	$R_{RAW}^2$ ( <b>Srsq</b> )		R-cuadrado corregido	$\bar{R}^2$	
F(K, T-K-1)	$F_1$ ( <b>SFstat</b> )		Valor p (de F)	$PV_{F_1}$	
Log-verosimilitud	$L$ ( <b>Stul</b> )		Criterio de Akaike	<b>Saic</b>	
Criterio de Schwarz	$SC$ ( <b>Sbic</b> )		Crit. de Hannan-Quinn	<b>Sbhc</b>	

Por ejemplo, bajo la denominación R-cuadrado, en este caso Gretl no proporciona el coeficiente de determinación sino el coeficiente de determinación bruto ( $R_{RAW}^2$ ).

R-cuadrado=Coeficiente de determinación

$$R_{RAW}^2 = \frac{SCR}{\sum_{t=1}^T Y_t^2}$$

En modelos formulados sin ordenada en el origen, no se cumple la descomposición de SCT en SCE y SCR. Por tanto, el **coeficiente de determinación** deja de estar acotado entre cero y uno y carece de la interpretación habitual, es decir, ya no mide la proporción en que la varianza muestral del regresando (SCT/T) es explicada por la varianza del regresando estimado (SCR/T) y, por tanto, no podrá

interpretarse como el porcentaje de variaciones del regresando explicado por las variaciones de las variables explicativas. Por ello, en estos modelos el **coeficiente de determinación** pierde su atractivo como medida de bondad de ajuste y, Gretl calcula en su lugar el **coeficiente de determinación bruto**, que aunque carece de la interpretación habitual (puesto que no se basa en la descomposición de SCT en SCE y SCR), está acotado entre cero y uno (ya que se basa en la descomposición de  $Y'Y$  en  $\hat{Y}'\hat{Y}$  y  $e'e$ , descomposición que se cumple independientemente de que el modelo esté formulado con o sin regresor ficticio).

Además, debe de tenerse en cuenta que las expresiones de cálculo de los estadísticos se ven afectadas por el hecho de que el modelo esté formulado sin ordenada en el origen.

D.T. de la regresión=Estimador de la desviación típica de la perturbación o desviación residual

$$S = \sqrt{S^2} = \sqrt{\frac{SCE}{T-K}}$$

Al calcular el **estimador de la varianza de la perturbación** se debe tener en cuenta que los grados de libertad del modelo son  $T-K$ , puesto que el número de regresores coincide con el número de variables explicativas.

Debe recordarse que en un modelo formulado sin ordenada en el origen la **media del regresando** no coincide con la media del regresando estimado  $\bar{Y} \neq \bar{\hat{Y}}$ .

Coeficiente=Coeficiente estimado ( $b_i$ )

$$b = (X'X)^{-1}X'Y = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{pmatrix}_{K \times 1} \Rightarrow b_i$$

Los **coeficientes estimados** recogen el valor de los estimadores de los parámetros que acompañan a cada una de las variables explicativas.

Desv. Típica=Desviación típica estimada de los estimadores (error estándar) ( $S_{b_i}$ )

$$\hat{V}(b) = S^2(X'X)^{-1} = \begin{pmatrix} S_{b_1}^2 & \dots & \dots & \dots \\ S_{b_2 b_1} & S_{b_2}^2 & \dots & \dots \\ \dots & \dots & \ddots & \dots \\ S_{b_K b_1} & S_{b_K b_2} & \dots & S_{b_K}^2 \end{pmatrix}_{K \times K} \Rightarrow S_{b_i} = \sqrt{S_{b_i}^2}$$

Las **desviaciones típicas estimadas** de los estimadores miden la precisión con la que son estimados los parámetros, es decir, indican el “grado de confianza” de las estimaciones siempre que los estimadores sean insesgados.

La matriz de varianzas-covarianzas estimada de los estimadores es una matriz de orden  $K \times K$ , cuyos elementos diagonales son las varianzas estimadas de los estimadores mínimo cuadráticos ordinarios y cuyos elementos no diagonales son las covarianzas estimadas de dichos estimadores.

### 3.8. Interpretación de los coeficientes

En un Modelo de Regresión Lineal Múltiple, a los parámetros que acompañan a las variables explicativas se les denomina **coeficientes de regresión parcial**, pues miden el efecto que ocasionan sobre la variable explicada los cambios en la variable explicativa a la que acompañan cuando están presentes otras variables:

$\beta_i = \frac{\partial Y_t}{\partial X_{it}}$   $i = 1, 2, \dots, K$  o bien  $\beta_i = \frac{\partial Y_t}{\partial X_i}$   $i = 1, 2, \dots, K$ , dado que es independiente de la observación.

La estimación de  $\beta_i$  puede considerarse una medida del efecto causal, una vez se han neutralizado, tanto en el regresando como en el regresor las variaciones causadas por las restantes variables explicativas del modelo y se interpreta como el cambio en la variable explicada producido por un cambio unitario en la variable explicativa a la que acompaña, manteniendo constantes las demás variables.

Una de las hipótesis del MRLC es la ausencia de relaciones lineales entre los regresores (condición de rango o hipótesis de rango pleno), siendo la hipótesis que hace posible aislar el efecto de cada uno de los regresores. No obstante, las buenas propiedades del vector de estimadores se siguen cumpliendo siempre que la dependencia lineal entre los regresores no sea exacta (multicolinealidad aproximada), sin embargo, en estas situaciones la interpretación de los coeficientes debe hacerse con cautela, puesto que dichos coeficientes pueden estar recogiendo el efecto no sólo de la variable a la que acompañan, sino también el efecto de las variables relacionadas con ella. En el análisis económico y/o empresarial es difícil encontrar regresores completamente ortogonales, lo habitual es que exista un determinado grado de dependencia lineal entre ellos (que no debe ser demasiado grande para que su efecto no sea demasiado pernicioso).

Debe de tenerse en cuenta que los coeficientes de regresión parcial dependen de las unidades de medida de las variables a las que acompañan, por lo que no son directamente comparables entre sí.

Véase en la **Tabla 3-4** que en la salida del comando **ols**, cada coeficiente de regresión parcial estimado ( $b_i$ ) viene acompañado por su error estándar ( $S_{b_i}$ ), que indica la precisión de la estimación y por su ratio t ( $t_i$ ), que indica su grado de significación.

Para completar la interpretación de los resultados de la estimación se pueden calcular los coeficientes estandarizados y las elasticidades en media, que no dependen de las unidades de medida de las variables a las que acompañan (son adimensionales, por lo que son directamente comparables entre sí).

Los **coeficientes beta** se estiman a través de los coeficientes de regresión parcial estimados ( $b_i$ ) ajustados por el cociente entre la desviación estándar de la variable independiente y la desviación estándar de la variable dependiente ( $\frac{S_{X_i}}{S_Y}$ ), por lo que miden la importancia relativa de las variables independientes<sup>23</sup>:

$$b_i = \frac{S_{X_i}}{S_Y} \quad i = 1, 2, \dots, K$$

Los coeficientes beta o coeficientes estandarizados permiten determinar cual es la variable explicativa que tiene mayor peso para la explicación del regresando.

La interpretación de los coeficientes beta es similar a las de los coeficientes de regresión parcial teniendo en cuenta que tanto la variable explicada como las explicativas están medidas en unidades de desviación estándar: miden el cambio en la variable dependiente (en unidades de desviación estándar)

<sup>23</sup> Los coeficientes beta estimados se pueden obtener a partir de la estimación del siguiente modelo transformado:

$$\frac{Y_t - \bar{Y}}{S_Y} = \beta_1^* \frac{X_{1t} - \bar{X}_1}{S_{X_1}} + \beta_2^* \frac{X_{2t} - \bar{X}_2}{S_{X_2}} + \dots + \beta_K^* \frac{X_{Kt} - \bar{X}_K}{S_K} + \frac{\varepsilon_t}{S_Y}$$

donde las variables han sido estandarizadas (se les ha restado su media y se las ha dividido por su desviación estándar).

La estandarización de las variables facilita la comparación entre los coeficientes aunque, en este caso, la interpretación debe hacerse en términos de desviación estándar.

producido por un cambio unitario en la variable independiente a la que acompaña (en unidades de desviación estándar) manteniendo constantes las demás variables.

Las **elasticidades en media** se estiman a través de los coeficientes de regresión parcial estimados ( $b_i$ ) ajustados por el cociente entre la media de la variable independiente y la media de la variable dependiente ( $\bar{X}_i/\bar{Y}$ ), por lo que miden la sensibilidad de la variable dependiente a los cambios en las variables independientes<sup>24</sup>:

$$\hat{E}_i = b_i \frac{\bar{X}_i}{\bar{Y}} \quad i = 1, 2, \dots, K.$$

Las elasticidades en media miden el cambio porcentual en la variable dependiente producido por un cambio porcentual en la variable independiente a la que acompaña, manteniendo constantes las demás variables.

### 3.9. Formas funcionales alternativas

En algunos casos una forma funcional lineal no caracteriza adecuadamente la relación entre el regresando y los regresores, ya que las variaciones en las variables explicativas no producen siempre el mismo efecto sobre la variable que se pretende explicar, tal como supone el MRLC<sup>25</sup>.

La no linealidad en las variables se puede incorporar en el modelo sin demasiada dificultad, pero no ocurriría lo mismo si la no linealidad afectase a los parámetros que intervienen en la relación, ya que tales modelos requieren tratamientos con un mayor grado de dificultad (por lo que no serán tratados en este manual).

A continuación se analizarán algunas de las formas funcionales no lineales en las variables pero lineales en los parámetros, ya que son fácilmente linealizables a través de transformaciones sencillas. Debe de tenerse en cuenta que la interpretación de los coeficientes es distinta dependiendo del modelo considerado: modelo lineal, modelo lin-log, modelo log-log y modelo log-lin.

- **Modelo Lineal**  $\rightarrow Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \varepsilon_t$

Los coeficientes estimados de las variables explicativas miden el cambio absoluto que se produce en el regresando ante un cambio absoluto producido en la variable explicativa a la que acompañan, bajo la cláusula “ceteris paribus”:

$$b_i = \frac{\partial Y}{\partial X_i} \Rightarrow b_i = \frac{\Delta Y}{\Delta X_i}$$

- **Modelo Lin-Log**  $\rightarrow Y_t = \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \dots + \beta_K \ln X_{Kt} + \varepsilon_t$

Los coeficientes estimados de las variables explicativas miden el cambio absoluto que se produce en el regresando ante un cambio relativo producido en la variable explicativa a la que acompañan, bajo la cláusula “ceteris paribus”. Si este cambio relativo se multiplica por 100, el coeficiente estimado

<sup>24</sup> Las elasticidades en media estimadas se pueden obtener a partir de la estimación del siguiente modelo transformado:

$$\ln Y_t = \beta_0 + E_1 \ln X_{1t} + E_2 \ln X_{2t} + \dots + E_K \ln X_{Kt} + \varepsilon_t$$

donde las variables han sido transformadas aplicándole logaritmos neperianos.

<sup>25</sup> Por ejemplo, determinar el consumo en función de la renta mediante un modelo lineal resulta demasiado restrictivo, ya que se puede pensar que la parte del incremento de la renta que se consume no es constante, sino que disminuye con el nivel de renta, mientras que el modelo lineal supone una derivada del consumo respecto a la renta constante y, por lo tanto, independiente del nivel de renta.

quedaría dividido por 100 y se podría interpretar como el cambio absoluto producido en el regresando ante un cambio porcentual en la variable explicativa correspondiente:

$$b_i = \frac{\partial Y}{\partial \ln X_i} \Rightarrow b_i = \frac{\Delta Y}{\frac{\Delta X_i}{X}}$$

- **Modelo Log-Log**  $\rightarrow \ln Y_t = \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \dots + \beta_K \ln X_{Kt} + \varepsilon_t$

Los coeficientes estimados de las variables explicativas miden el cambio relativo que se produce en el regresando ante un cambio relativo producido en la variable explicativa a la que acompañan, bajo la cláusula “ceteris paribus”. El coeficiente estimado se podría interpretar como el cambio porcentual producido en el regresando ante un cambio porcentual en la variable explicativa correspondiente:

$$b_i = \frac{\partial \ln Y_t}{\partial \ln X_i} \Rightarrow b_i = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_i}{X}}$$

- **Modelo Log-Lin**  $\rightarrow \ln Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \varepsilon_t$

Los coeficientes estimados de las variables explicativas miden el cambio relativo que se produce en el regresando ante un cambio absoluto producido en la variable explicativa a la que acompañan, bajo la cláusula “ceteris paribus”. Si este cambio relativo se multiplica por 100, el coeficiente estimado quedaría multiplicado por 100 y se podría interpretar como el cambio porcentual producido en el regresando ante un cambio absoluto en la variable explicativa correspondiente:

$$b_i = \frac{\partial \ln Y_t}{\partial X_i} \Rightarrow b_i = \frac{\frac{\Delta Y}{Y}}{\Delta X_i}$$

### 3.10. Predicción

Un objetivo fundamental a tener en cuenta en la construcción de un modelo econométrico es la valoración de la capacidad predictiva como opción para comprobar la validez del modelo. Dicha validación debe realizarse con datos diferentes a los que se utilizaron para la estimación y/o diagnóstico del modelo.

Habitualmente, se trabaja con datos de series temporales y es frecuente que dichos datos pertenezcan a períodos posteriores a la muestra, por ello a la estimación de estos valores externos se le denomina **predicción**. No obstante, cuando los datos son transversales, también puede resultar interesante obtener estimaciones con datos externos a la muestra, que por comodidad se seguirá denominando predicción.

En este epígrafe se analizará la predicción óptima en el ámbito de un MRLNC bajo la hipótesis de estabilidad postmuestral.

Existen varias alternativas para realizar las predicciones:

1. Utilizar la opción **Predicciones** del menú **Análisis** de la **Ventana Modelo**.
2. Utilizar el comando **fcast**.

#### 3.10.1. Utilizar el menú Análisis

Utilizando el menú **Análisis** y seleccionando **Predicciones** se accede al cuadro de diálogo **predicción** (véase Ilustración 3-7) donde se debe elegir:

- El dominio de la predicción: hay que tener en cuenta que el rango de predicción no debe coincidir con el rango de estimación del modelo.
- El tipo de predicción: se puede elegir entre predicción estática, dinámica o recursiva. La predicción dinámica, tan sólo, está disponible cuando se trabaja con series temporales y entre las variables explicativas aparece algún retardo de la variable endógena. En cuanto a las predicciones recursivas “n” paso(s) adelante, hay que señalar que la predicción para una observación concreta “t” se obtiene multiplicando los coeficientes estimados del modelo con la muestra desde la observación “1” hasta la observación “t-4” y los regresores evaluados en la observación “t” (la predicción recursiva será realmente una predicción sólo si todos los regresores estocásticos son variables retardadas).
- El número de observaciones a representar anteriores a la predicción: se debe tener en cuenta que para que Gretl elabore un gráfico es necesario seleccionar un mínimo de 5 observaciones.
- Elegir la forma de representar el intervalo de confianza de las predicciones: barra de error, línea inferior y superior o área sombreada.
- Elegir el nivel de confianza: mínimo del 60% y máximo del 99%.
- Elegir entre el intervalo para la predicción puntual (actual Y) o para la predicción media (mean Y).

Una vez hechas las elecciones se pincha en “Aceptar”y se abren dos ventanas de resultados (véase Ilustración 3-7):

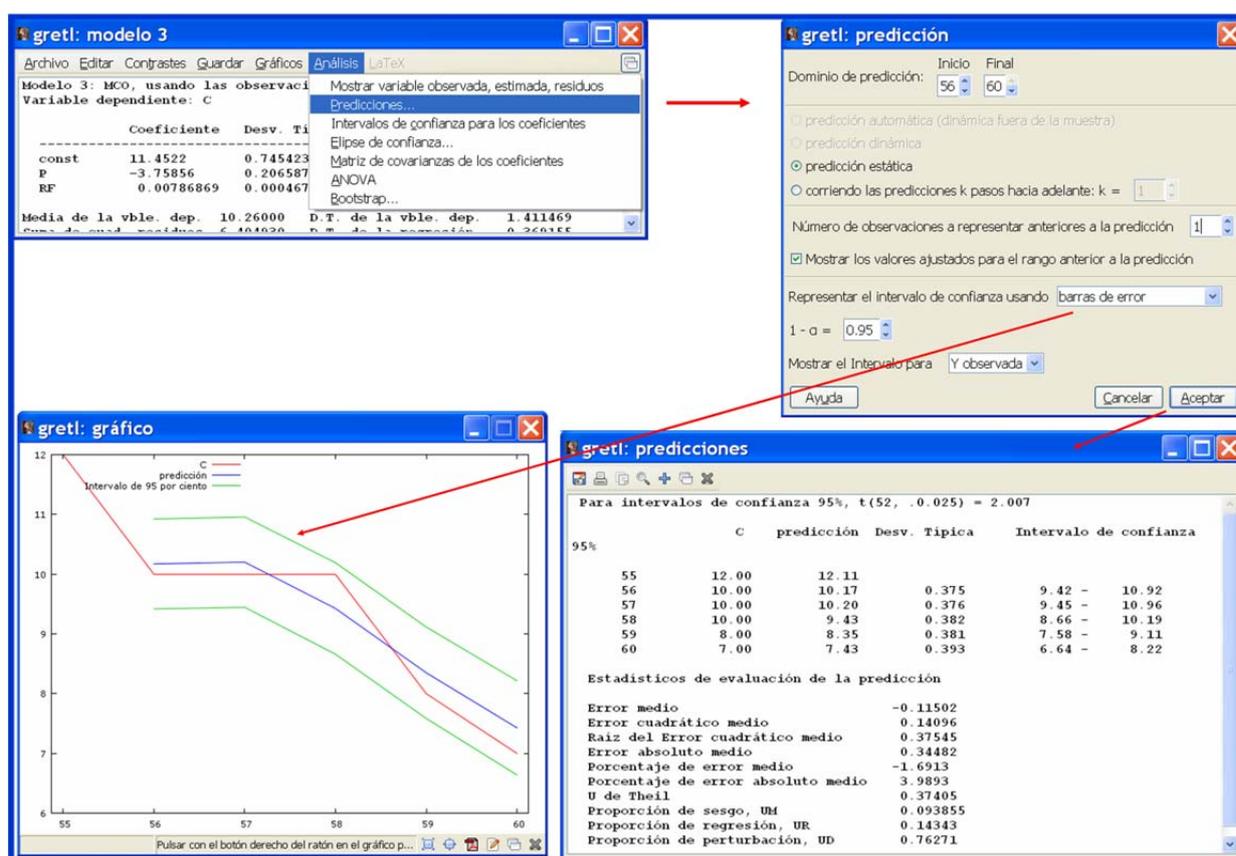


Ilustración 3-7. Opción **Predicciones** del menú **Análisis** de la **Ventana Modelo**.

- La Ventana predicciones, en la que Gretl proporciona entre otros datos, los valores predichos, sus desviaciones típicas, sus intervalos de confianza así como el nivel de confianza y el valor crítico de la t de Student utilizado para su construcción.

– La Ventana gráfico, en la que Gretl proporciona una representación gráfica de los valores observados o valores medios del regresando, los valores predichos y sus respectivos intervalos de confianza.

Véase que en el ejemplo de la Ilustración 3-7, el rango de predicción está formado por las observaciones de la 56 a la 60, se ha optado por una predicción estática, por predicciones puntuales y por intervalos de confianza del 95%.

### 3.10.2. Utilizando el comando **fcast**

Para que el comando **fcast** se ejecute correctamente debe ir precedido de un comando de estimación. El formato del comando **fcast** es:

**fcast** *observación inicial observación final --opciones*

Si se desean guardar los resultados de la predicción es necesario incluir en el comando **fcast** el nombre de la variable en la que se quieren guardar dichos resultados. En este caso, Gretl ejecuta el comando omitiendo la salida, por lo que para visualizar la serie de valores predichos será necesario utilizar un comando **print**.

**fcast** *observación inicial observación final nombre variable --opciones*

La mayor parte de la salida del comando **fcast** se utiliza para realizar un análisis de la capacidad predictiva del modelo, por lo que no será analizada en este capítulo.

En el Cuadro 3-3 se recogen algunas de las opciones disponibles con este comando.

<b>--static</b>	→	proporciona estimaciones estáticas.
<b>--dynamic</b>	→	proporciona estimaciones dinámicas.
<b>--mean-y</b>	→	proporciona la predicción media.
<b>--no-stats</b>	→	suprime la visualización de los estadísticos de predicción.
<b>--quiet</b>	→	ejecuta el comando <b>fcast</b> pero suprime la visualización de la salida.

Cuadro 3-3. Algunas opciones del comando **fcast**.

En la **Tabla 3-5** se recoge la salida estándar del comando **fcast**.

Tabla 3-5. Salida asociada al submenú <b>Predicciones</b> del menú <b>Análisis</b> de la <b>Ventana Modelo</b> .				
Para intervalos de confianza 95%, $t(T-K-1, 0.025)$				
Observaciones	Y	Predicción	Desv. Típica	Intervalo de 95%
1	$Y_1$	$\hat{Y}_1$	$S_{e_1}$	$\hat{Y}_1 \pm t_{\alpha/2}^{T-K-1} S_{e_1}$
2	$Y_2$	$\hat{Y}_2$	$S_{e_2}$	$\hat{Y}_2 \pm t_{\alpha/2}^{T-K-1} S_{e_2}$
...	...	...	...	...
$\tau$	$Y_\tau$	$\hat{Y}_\tau$	$S_{e_\tau}$	$\hat{Y}_\tau \pm t_{\alpha/2}^{T-K-1} S_{e_\tau}$
Estadísticos de evaluación de la predicción				
Error medio			EM	
Error cuadrático medio			ECM	
Raíz del Error cuadrático medio			RECM	
Error absoluto medio			EAM	
Porcentaje de error medio			PEM	
Porcentaje de error absoluto medio			PEAM	

U de Theil	U
Proporción de sesgo	UM
Proporción de regresión	UR
Proporción de perturbación	UD

Tal y como puede verse en la **Tabla 3-5**, Gretl además de proporcionar el valor predicho, la desviación estándar del error predicho y el intervalo de predicción, proporciona estadísticos para evaluar la capacidad predictiva del modelo.

### 3.11. Ejercicio de aplicación

En este ejercicio se analizará la estimación de un modelo con y sin ordenada en el origen, su bondad de ajuste y las propiedades de los residuos. También se analizará la estimación de formas funcionales no lineales. Para ello se plantean los siguientes modelos:

$$C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \varepsilon_t \quad \forall t = 1, 2, \dots, 10$$

$$C_t = \beta_1 P_t + \beta_2 RF_t + \varepsilon_t \quad \forall t = 1, 2, \dots, 10$$

En este caso, se utilizarán los 10 primeros datos del fichero `aceite.gdt`: consumo familiar mensual medio de aceite de oliva en litros (C), precio medio de compra del aceite de oliva en euros/litro (P) e ingresos familiares mensuales medios en euros (RF).

```
? open "C:\Proyecto01\Aceite.gdt"
Leer fichero de datos C:\Proyecto01\Aceite.gdt
periodicidad: 1, máx.obs: 60
rango de observaciones: 1-60

Listando 7 variables:
  0) const   1) C       2) P       3) RF      4) PG
  5) PS     6) TF
```

#### 3.11.1. Estimación MCO del modelo con ordenada en el origen

```
? smpl 1 10
Rango de datos completo: 1 - 60 (n = 60)
Muestra actual: 1 - 10 (n = 10)
```

Obsérvese que aunque se dispone de 60 observaciones para cada variable, sólo se van a utilizar las 10 primeras. Además, el fichero de datos contiene más variables que las que se van a utilizar en la estimación.

```
# Modelo con ordenada en el origen con la opción anova
? ols C const P RF --anova

Modelo 1: MCO, usando las observaciones 1-10
Variable dependiente: C

-----

```

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	8.28657	1.28569	6.445	0.0004	***
P	-2.93379	0.297456	-9.863	2.34e-05	***
RF	0.00907529	0.000854912	10.62	1.44e-05	***

```

Media de la vble. dep. 10.30000 D.T. de la vble. dep. 1.337494
Suma de cuad. residuos 0.420867 D.T. de la regresión 0.245202
R-cuadrado 0.973859 R-cuadrado corregido 0.966390
F(2, 7) 130.3903 Valor p (de F) 2.89e-06
Log-verosimilitud 1.650732 Criterio de Akaike 2.698536
Criterio de Schwarz 3.606292 Crit. de Hannan-Quinn 1.702731

Análisis de Varianza:
Suma de cuadrados gl Media de cuadrados
```

Regresión	15.6791	2	7.83957
Residuo	0.420867	7	0.0601239
Total	16.1	9	1.78889
R <sup>2</sup> = 15.6791 / 16.1 = 0.973859			
F(2, 7) = 7.83957 / 0.0601239 = 130.39 [Valor p 2.89e-006]			
# Se genera el tamaño muestral			
? genr T = \$T			
Se ha generado el escalar T = 10			
# Se genera el número de variables explicativas			
? genr k = \$ncoeff - 1			
Se ha generado el escalar k = 2			

### - Interpretación de resultados

El modelo estimado por MCO es:

$$\hat{C}_t = 8.28657 - 2.93379P_t + 0.00907529RF_t$$

(1.28569)      (0.297456)      (0.000854912)

$$t_0 = 6.445 \quad t_1 = -9.863 \quad t_2 = 10.62$$

$$R^2 = 0.9738$$

Los valores que aparecen entre paréntesis debajo de cada coeficiente estimado ( $b_i$ ) hacen referencia a su desviación típica estimada ( $S_{b_i}$ ) e informan de lo cerca o lejos que está el valor del coeficiente estimado de su valor esperado (informan de su “confiabilidad” o precisión), por ello, cuanto más pequeños sean mejor. El problema que presentan estas desviaciones es que se ven afectadas por las unidades de medida de las variables. Por ello, a veces, entre paréntesis en vez de figurar estas desviaciones figuran las ratios  $t$  ( $t_i$ ), que son adimensionales, de forma que cuanto mayor sea esta ratio en valor absoluto mayor será la precisión del estimador.

De los resultados obtenidos se deduce que:

- El estimador de la ordenada en el origen (8.28657) se podría interpretar como el consumo medio de aceite de oliva de una familia, cuando el nivel de renta mensual y el precio del litro de aceite son nulos. Esta interpretación “mecánica” que se acaba de hacer de  $b_0$  no siempre tiene sentido (este caso es un ejemplo), además cabe señalar que la muestra utilizada no incluye el cero entre los valores observados de dichas variables, por lo que lo más correcto sería interpretar  $b_0$  como el consumo mensual medio de aceite de las familias que no es debido a las variables que aparecen de forma explícita en el modelo.
- Manteniendo constante el nivel de renta, ante un incremento de un euro en el precio del litro de aceite de oliva, la demanda de aceite (medida en términos de consumo mensual por familia), disminuye en 2.93379 litros.
- Manteniendo constante el precio del litro de aceite de oliva, ante un incremento de un euro en el nivel de renta de las familias, la demanda de aceite aumenta en 0.00907529 litros.

Dado que el modelo está formulado con ordenada en el origen, el coeficiente de determinación está acotado entre 0 y 1 y, por tanto, un valor de 0.9738 se puede considerar que representa un elevado poder explicativo (modelo con elevada bondad de ajuste), indicando que aproximadamente el 97% de las variaciones en el consumo mensual de aceite por las familias, son explicadas por las variaciones en el precio del aceite y en el nivel de ingresos familiares mensuales.

### 3.11.1.1. Resultados de la estimación utilizando álgebra matricial

Una de las ventajas del paquete econométrico Gretl es que permite realizar fácilmente operaciones de cálculo matricial<sup>26</sup>. En este epígrafe, con el fin de que el usuario entienda el procedimiento de estimación y no se quede únicamente en la interpretación de resultados numéricos, se calculan de forma detallada los resultados de la estimación del modelo haciendo uso del álgebra matricial.

#### - Vector de estimadores

Para obtener los estimadores mínimo cuadráticos ordinarios del vector de parámetros es necesario conocer los elementos de la matriz  $X'X$  y del vector  $X'Y$ , que se obtendrán a partir de los datos de la muestra:

$$b = (X'X)^{-1}X'Y = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_K \end{pmatrix}_{(K+1) \times 1}$$

donde:

- $X'X$  es la matriz de orden  $(K+1) \times (K+1)$ , que contiene los sumatorios de los productos cruzados de los regresores:

$$X'X = \begin{pmatrix} T & \sum_{t=1}^T X_{1t} & \dots & \sum_{t=1}^T X_{Kt} \\ \sum_{t=1}^T X_{1t} & \sum_{t=1}^T X_{1t}^2 & \dots & \sum_{t=1}^T X_{1t}X_{Kt} \\ \dots & \dots & \ddots & \dots \\ \sum_{t=1}^T X_{Kt} & \sum_{t=1}^T X_{1t}X_{Kt} & \dots & \sum_{t=1}^T X_{Kt}^2 \end{pmatrix}_{(K+1) \times (K+1)}$$

- $X'Y$  es el vector de orden  $(K+1) \times 1$ , que contiene los sumatorios de los productos cruzados del regresando con los regresores:

$$X'Y = \begin{pmatrix} \sum_{t=1}^T Y_t \\ \sum_{t=1}^T Y_t X_{1t} \\ \dots \\ \sum_{t=1}^T Y_t X_{Kt} \end{pmatrix}_{(K+1) \times 1}$$

Recuerde que, siempre que se inicia una sesión de trabajo, Gretl genera de forma automática el regresor ficticio bajo la denominación de **const**.

<sup>26</sup> Las instrucciones para llevar a cabo dichos cálculos ya han sido analizadas en un capítulo anterior.

```
# Se genera la matriz de diseño (X)
? matrix X = {const, P, RF}
Se ha generado la matriz X

# Se muestra la matriz X (que en este caso es de orden 10x3)
? print X
X (10 x 3) [t1 = 1, t2 = 10]
  1.0000    2.4600    1079.5
  1.0000    2.5100    1108.2
  1.0000    2.7300    1116.8
  1.0000    2.7300    978.09
  1.0000    2.6300    809.62
  1.0000    2.3800    1047.1
  1.0000    2.4400    1164.3
  1.0000    2.7500    1069.6
  1.0000    2.9800    1040.3
  1.0000    1.9500    1067.8
```

Cuando se definen matrices a partir de variables, Gretl recuerda el rango muestral con el que se está trabajando, que en este caso comienza en la primera observación ( $t1=1$ ) y termina en la décima ( $t2=10$ ).

```
# Se genera el vector que recoge las observaciones del regresando (Y)
? matrix Y = {C}
Se ha generado la matriz Y

# Se muestra el vector Y (en este caso, es de orden 10x1)
? print Y
Y (10 x 1) [t1 = 1, t2 = 10]
  11
  11
  10
  9
  8
  11
  12
  10
  9
  12
```

Nótese que Gretl no ha creado una serie sino un vector columna denominado Y que contiene las 10 primeras observaciones de la serie C. El usuario debe tener esto en cuenta, puesto que determinadas funciones del comando **genr** se pueden utilizar con series pero no con matrices. En el caso de que se desee definir como serie, basta sustituir el comando **genr** por el comando **series** y Gretl añadiría la serie a la lista de variables iniciales.

```
# Se genera la serie que recoge las observaciones del regresando (y)
? series y = {C}
Se ha generado la serie y (ID 7)

# Se muestra la serie y
? print y
Muestra actual: 1 - 10 (n = 10)
11.0000 11.0000 10.0000 9.00000 8.00000 11.0000 12.0000 10.0000
9.00000 12.0000
```

Como puede verse, Gretl asigna a la variable o serie “y”, el identificador “ID 7” puesto que en la lista inicial ya tiene 6 variables.

Dado que se ha utilizado la denominación Y para un vector, no se podrá utilizar la misma denominación para una variable, si así se hiciese Gretl emitiría el mensaje de error “*la variable Y es de tipo matrix*”:

```
# Se genera la serie que recoge las observaciones del regresando (Y)
? series Y = {C}
La variable Y es de tipo matrix
```

Se debe recordar que se podría utilizar la misma denominación para definir “elementos” de las mismas características, pero que la información nueva sustituiría a la información antigua. Por ejemplo, si se ejecutará el comando:

```
? matrix Y = {P}
Se ha reemplazado la matriz Y

? print Y
Y (10 x 1) [t1 = 1, t2 = 10]
2.46
2.51
2.73
2.73
2.63
2.38
2.44
2.75
2.98
1.95

? matrix Y = {C}
Se ha reemplazado la matriz Y

? print Y
Y (10 x 1) [t1 = 1, t2 = 10]
11
11
10
9
8
11
12
10
9
12
```

Gretl informaría de que el vector Y ha sido reemplazado, dejaría de contener las 10 primeras observaciones de la variable C y pasaría a contener las 10 primeras observaciones de la variable P.

```
# Se calcula y se muestra la matriz de productos cruzados de los regresores
? matrix XTX = X' * X
Se ha generado la matriz XTX

? print XTX
XTX (3 x 3)
    10.000      25.560      10481.
    25.560      66.038      26743.
    10481.      26743.     1.1072e+007

# Se calcula y se muestra la inversa de X'X
? gener XTXI = inv(XTX)
Se ha generado la matriz XTXI

? print XTXI
XTXI (3 x 3)
    27.493      -4.6269      -0.014852
    -4.6269      1.4716      0.00082567
    -0.014852    0.00082567    1.2156e-005

# Se calcula y se muestra el vector de productos cruzados de los regresores y el regresando
? matrix XTY = X' * Y
Se ha generado la matriz XTY

? print XTY
XTY (3 x 1)
    103.00
    260.76
    1.0888e+005
```

Con la inversa de la matriz  $X'X$  y el vector  $X'Y$ , se calcula el vector de estimadores MCO:

```
# Se calcula y se muestra el vector de estimadors (b)
```

```
? matrix b = XTXI * XTY
Se ha generado la matriz b

? print b
b (3 x 1)
   8.2866
  -2.9338
   0.0090753
```

Cuando se hace un cálculo de este tipo no es necesario realizar todas las operaciones por separado (traspuesta, inversa,...), ya que Gretl permite hacer determinados cálculos y operaciones de una sola vez, por ejemplo, para obtener el vector de estimadores bastaría con una única instrucción:

```
? matrix b = inv(X' * X) * (X' * Y)
```

Además, no sería necesario calcular el vector Y, en este caso, se podría utilizar directamente C en las expresiones de cálculo:

```
? matrix b = inv(X' * X) * (X' * C)
```

### - Regresando estimado

$$\hat{Y} = Xb$$

```
# Se calcula y se muestra el regresando estimado
? matrix YE = X * b
Se ha generado la matriz YE

? print YE
YE (10 x 1) [t1 = 1, t2 = 10]
  10.866
  10.980
  10.413
  9.1538
  7.9183
  10.807
  11.695
  9.9256
  8.9854
  12.256
```

Si lo que interesa obtener es la estimación para una única observación del regresando, se puede utilizar la expresión:

$$\hat{Y}_t = X_t' b = b_0 + b_1 X_{1t} + \dots + b_K X_{Kt}$$

```
# Se calcula el valor estimado para la primera observación
? gener YE1 = b[1] + b[2] * X[1,2] + b[3] * X[1,3]
Se ha generado el escalar YE1 = 10.8663
```

Recuerde que cuando se utiliza el comando **gener** y el resultado es un escalar, Gretl informa de su valor y no será necesario utilizar un comando **print** para visualizarlo (sería necesario en el caso de generar series o matrices).

Los valores de los regresores para una observación concreta se pueden obtener a partir de la matriz de diseño (X) o a partir de las series de datos, que son reconocidas por Gretl como vectores columna<sup>27</sup>:

```
? matrix YE1M = b[1] + b[2] * P[1] + b[3] * RF[1]
Se ha generado la matriz YE1M

? print YE1M
YE1M (1 x 1)
```

<sup>27</sup> Recuerde que esta afirmación no se cumple a la inversa, es decir, los vectores columna no son reconocidos por Gretl como series.

10,866

Para generar la serie que contiene los valores estimados del regresando:

```
# Se genera y se muestra la serie que recoge las observaciones del regresando estimado (ye)
? series ye = X * b
Se ha generado la serie ye (ID 8)

? print ye
Muestra actual: 1 - 10 (n = 10)
10.8663 10.9802 10.4130 9.15379 7.91825 10.8067 11.6949 9.92559
8.98537 12.2559
```

#### - Residuos

$$e = Y - \hat{Y}$$

```
# Se calcula y se muestra el vector de residuos
? matrix E = Y - YE
Se ha generado la matriz E

? print E
E (10 x 1) [t1 = 1, t2 = 10]
 0.13367
 0.019811
 -0.41298
 -0.15379
 0.081748
 0.19328
 0.30514
 0.074408
 0.014631
 -0.25592
```

Para obtener un residuo para una observación concreta:

$$e_t = Y_t - \hat{Y}_t$$

```
# Se calcula el valor estimado para la primera observación
? genr e1 = Y[1] - YE[1]
Se ha generado el escalar e1 = 0.133674
```

Para obtener la serie de residuos:

```
# Se genera y se muestra la serie de residuos
? series e = Y - YE
Se ha generado la serie e (ID 9)

? print e
Muestra actual: 1 - 10 (n = 10)
0.133674 0.0198114 -0.412985 -0.153788 0.0817480 0.193283
0.305141 0.0744077 0.0146306 -0.255922
```

#### - Suma de Cuadrados Totales

$$SCT = \sum_{t=1}^T (Y_t - \bar{Y})^2 = Y'Y - T\bar{Y}^2$$

```
# Se calcula la SCT
? genr SCT = sum((y - mean(y))^2)
Se ha generado el escalar SCT = 16.1

# Otra alternativa de cálculo
? genr SCT = Y' * Y - T * (mean(Y))^2
Se ha reemplazado el escalar SCT = 16.1
```

#### - Suma de Cuadrados de Regresión

$$SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 = \hat{Y}'\hat{Y} - T\bar{Y}^2 = b'X'Y - T\bar{Y}^2$$

```
# Se calcula la SCR
? genr SCR = sum((ye - mean(ye))^2)
Se ha generado el escalar SCR = 15.6791
```

```
# Otras alternativas de cálculo
? genr SCR = YE' * YE - T * (mean(YE))^2
Se ha reemplazado el escalar SCR = 15.6791

? genr SCR = b' * X' * Y - T * (mean(YE))^2
Se ha reemplazado el escalar SCR = 15.6791
```

### - Suma de Cuadrados de Errores

$$SCE = e'e = \sum_{t=1}^T e_t^2 = Y'Y - b'X'Y$$

```
# Se calcula la SCE
? genr SCE = sum(e^2)
Se ha generado el escalar SCE = 0.420867

# Otras alternativas de cálculo
? genr SCE = Y' * Y - b' * X' * Y
Se ha reemplazado el escalar SCE = 0.420867

? genr SCE = E' * E
Se ha reemplazado el escalar SCE = 0.420867
```

Al estar el modelo formulado con ordenada en el origen, también se podría calcular SCE como diferencia entre SCT y SCR:

```
? genr SCE = SCT - SCR
```

### - Estimador de la varianza de la perturbación

$$S^2 = \frac{SCE}{T - K - 1}$$

```
# Se calcula el estimador de la varianza de la perturbación
? genr s2 = SCE / (T - k -1)
Se ha generado el escalar s2 = 0.0601239

# Se calcula el estimador de la desviación típica de la perturbación
? genr s = sqrt(s2)
Se ha generado el escalar s = 0.245202
```

### - Matriz de varianzas-covarianzas estimada de los EMCO

$$\hat{V}(b) = S^2(X'X)^{-1} = \begin{pmatrix} S_{b_0}^2 & \dots & \dots & \dots & \dots \\ S_{b_1 b_0} & S_{b_1}^2 & \dots & \dots & \dots \\ S_{b_2 b_0} & S_{b_2 b_1} & S_{b_2}^2 & \dots & \dots \\ \dots & \dots & \dots & \ddots & \dots \\ S_{b_K b_0} & S_{b_K b_1} & S_{b_K b_2} & \dots & S_{b_K}^2 \end{pmatrix}_{(K+1) \times (K+1)}$$

```
# Se calcula y se muestra la matriz de varianzas-covarianzas estimada de los estimadores
? matrix Vb = s2 * inv(X' * X)
Se ha generado la matriz Vb

? print Vb
Vb (3 x 3)
  1.6530      -0.27819  -0.00089295
 -0.27819    0.088480  4.9642e-005
 -0.00089295 4.9642e-005  7.3088e-007

# Se calcula y se muestra el vector que contiene las varianzas estimadas de los EMCO
? matrix s2b = diag(Vb)
Se ha generado la matriz s2b

? print s2b
s2b (3 x 1)
  1.6530
  0.088480
  7.3088e-007
```

```
# Se calcula y se muestra el vector que contiene las desviaciones típicas estimadas de los EMCO
? matrix sb = sqrt(s2b)
Se ha generado la matriz sb

? print sb
sb (3 x 1)
    1.2857
    0.29746
    0.00085491
```

#### - Ratios t de los EMCO

$$t_i = \frac{b_i}{S_{b_i}}$$

```
# Se calcula las ratios t
? genr t0 = b[1] / sb[1]
Se ha generado el escalar t0 = 6.44524

? genr t1 = b[2] / sb[2]
Se ha generado el escalar t1 = -9.86292

? genr t2 = b[3] / sb[3]
Se ha generado el escalar t2 = 10.6155
```

Otra forma de calcular los ratios t sería a través del comando **matrix** y de su operador **división elemento a elemento**:

```
? matrix t = b ./ sb
Se ha generado la matriz t

? print t
t (3 x 1)
    6,4452
   -9,8629
   10,615
```

Nótese que en este caso se generaría un vector columna de orden 3x1.

#### - Coeficiente de determinación y Tabla ANOVA respecto a la media

$$R^2 = \frac{SCR}{SCT}$$

```
# Cálculo del coeficiente de determinación
? genr R2 = SCR/SCT
Se ha generado el escalar R2 = 0.973859
```

Como el modelo está formulado con ordenada en el origen, aplicando la descomposición de SCT en SCE y SCR, el  $R^2$  también se puede obtener como:

$$R^2 = 1 - \frac{SCE}{SCT}$$

```
? genr R2 = 1 - SCE/SCT
```

La información necesaria para el cálculo del coeficiente de determinación en modelos formulados con ordenada en el origen está disponible en la tabla de **Análisis de Varianza**<sup>28</sup> que proporciona Gretl al ejecutar un comando **ols** con la opción **--anova**:

```
Análisis de Varianza:

                Suma de cuadrados          gl  Media de cuadrados
```

<sup>28</sup> Otros paquetes econométricos como Shazam, la denominan ANOVA respecto a la media.

Regresión	15.6791	2	7.83957
Residuo	0.420867	7	0.0601239
Total	16.1	9	1.78889

$R^2 = 15.6791 / 16.1 = 0.973859$   
 $F(2, 7) = 7.83957 / 0.0601239 = 130.39$  [Valor p 2.89e-006]

**- Tabla ANOVA respecto al origen y coeficiente de determinación bruto**

La descomposición de SCT en SCE y SCR, sólo se cumple en modelos formulados con ordenada en el origen. Pero la descomposición de  $\sum_{t=1}^T Y_t^2$  en  $\sum_{t=1}^T \hat{Y}_t^2$  y  $\sum_{t=1}^T e_t^2 = SCE$  se cumple siempre (modelos formulados con o sin ordenada en el origen<sup>29</sup>).

Aunque no aparezca entre las salidas de Gretl, parece conveniente calcular la tabla ANOVA respecto al origen, por ser la base para calcular el coeficiente de determinación en modelos formulados sin ordenada en el origen (coeficiente de determinación bruto):

```

# ANOVA respecto al origen
? genr scy = sum(y^2)
Se ha generado el escalar scy = 1077

? genr scye = sum(ye^2)
Se ha generado el escalar scye = 1076.58

? genr sce = sum(e^2)
Se ha generado el escalar sce = 0.420867
    
```

Nótese que la única diferencia con la tabla ANOVA respecto a la media, es que en la tabla ANOVA respecto al origen, la suma de cuadrados de las desviaciones no se calculan respecto a la media sino respecto al cero. Puede observarse que la suma de cuadrados de los residuos respecto a su media y respecto al cero coinciden puesto que, en modelos formulados con ordenada en el origen, la media de los residuos es cero (véase **Tabla 3-6**).

Gretl sólo proporciona la tabla ANOVA respecto a la media cuando el modelo está formulado con ordenada en el origen, en caso contrario, no muestra ninguna tabla de análisis de la varianza.

Tabla 3-6. Tabla ANOVA respecto al origen.			
Análisis de Varianza:			
		Sumas de cuadrados	gl
Regresión	$\sum_{t=1}^T (\hat{Y}_t - 0)^2 = \sum_{t=1}^T \hat{Y}_t^2$	1076.58	3
Residuo	$\sum_{t=1}^T (e_t - 0)^2 = \sum_{t=1}^T e_t^2 = SCE$	0.420867	7
Total	$\sum_{t=1}^T (Y_t - 0)^2 = \sum_{t=1}^T Y_t^2$	1077	10
$R^2 = 1076.58 / 1077 = 0.9996$ $F(3, 7) = 358.86 / 0.0601239 = 5968.674$ [Valor p 2.74605e-012]			

*Observe que la SCE puede calcularse utilizando ambas tablas.*

```

# Cálculo del coeficiente de determinación bruto
? genr R2RAW = scye/scy
Se ha generado el escalar R2RAW = 0.999609
    
```

<sup>29</sup> Algunos paquetes econométricos como Shazam, por analogía recogen esta descomposición en una tabla denominada, ANOVA respecto al origen.

```
? genr R2RAW = 1 - sce/scy
Se ha reemplazado el escalar R2RAW = 0.999609
```

- **Coefficiente de determinación corregido o ajustado**

$$\bar{R}^2 = 1 - \frac{\frac{SCE}{T - K - 1}}{\frac{SCT}{T - 1}}$$

```
# Se calcula el coeficiente de determinación ajustado
? genr R2A = 1 - ((SCE / (T - k - 1)) / (SCT / (T - 1)))
Se ha generado el escalar R2A = 0.96639
```

- **Matrices H y M**

La matriz H se define como:

$$H = X(X'X)^{-1}X'$$

```
# Se calcula la matriz H
? matrix H = X * inv(X' * X) * X'
Se ha generado la matriz H
? print H
H (10 x 10) [t1 = 1, t2 = 10]
0.12055 0.12345 0.10068 0.058765 0.019414 0.11999 0.14848 0.084099 0.048727 0.17584
0.12345 0.14243 0.14442 0.048353 -0.066478 0.10245 0.18255 0.11135 0.086936 0.12455
0.10068 0.14442 0.22167 0.085860 -0.11031 0.043911 0.17745 0.18168 0.22500 -0.070346
0.058765 0.048353 0.085860 0.18408 0.28351 0.065864 -0.0052496 0.12327 0.18956 -0.034012
0.019414 -0.066478 -0.11031 0.28351 0.77050 0.11851 -0.21961 0.022011 0.084786 0.097665
0.11999 0.10245 0.043911 0.065864 0.11851 0.14591 0.11176 0.046188 -0.0089576 0.25438
0.14848 0.18255 0.17745 -0.0052496 -0.21961 0.11176 0.26168 0.11375 0.058040 0.17114
0.084099 0.11135 0.18168 0.12327 0.022011 0.046188 0.11375 0.16786 0.22528 -0.075487
0.048727 0.086936 0.22500 0.18956 0.084786 -0.0089576 0.058040 0.22528 0.35985 -0.26922
0.17584 0.12455 -0.070346 -0.034012 0.097665 0.25438 0.17114 -0.075487 -0.26922 0.62548
```

H se denomina **matriz de proyección**, pues proyecta la variable dependiente sobre el espacio de las variables independientes, obteniendo el regresando estimado:

$$\hat{Y} = HY$$

```
matrix HY=H*Y
Se ha generado la matriz HY
? print HY YE
HY (10 x 1) [t1 = 1, t2 = 10]
10.866
10.980
10.413
9.1538
7.9183
10.807
11.695
9.9256
8.9854
12.256
YE (10 x 1) [t1 = 1, t2 = 10]
10.866
10.980
10.413
9.1538
7.9183
10.807
11.695
9.9256
8.9854
12.256
```

La matriz M se define como:

$$M = I - H$$

```
# Se calcula la matriz M
? matrix M = I(T) - H
```

```

Se ha generado la matriz M
? print M
M (10 x 10) [t1 = 1, t2 = 10]
0.87945   -0.12345   -0.10068   -0.058765   -0.019414   -0.11999   -0.14848   -0.084099   -0.048727   -0.17584
-0.12345   0.85757   -0.14442   -0.048353   0.066478   -0.10245   -0.18255   -0.11135   -0.086936   -0.12455
-0.10068   -0.14442   0.77833   -0.085860   0.11031   -0.043911   -0.17745   -0.18168   -0.22500   0.070346
-0.058765   -0.048353   -0.085860   0.81592   -0.28351   -0.065864   0.0052496   -0.12327   -0.18956   0.034012
-0.019414   0.066478   0.11031   -0.28351   0.22950   -0.11851   0.21961   -0.022011   -0.084786   -0.097665
-0.11999   -0.10245   -0.043911   -0.065864   -0.11851   0.85409   -0.11176   -0.046188   0.0089576   -0.25438
-0.14848   -0.18255   -0.17745   0.0052496   0.21961   -0.11176   0.73832   -0.11375   -0.058040   -0.17114
-0.084099   -0.11135   -0.18168   -0.12327   -0.022011   -0.046188   -0.11375   0.83214   -0.22528   0.075487
-0.048727   -0.086936   -0.22500   -0.18956   -0.084786   0.0089576   -0.058040   -0.22528   0.64015   0.26922
-0.17584   -0.12455   0.070346   0.034012   -0.097665   -0.25438   -0.17114   0.075487   0.26922   0.37452
    
```

M proyecta la variable dependiente sobre el espacio ortogonal a ella o espacio residual, obteniendo el error de estimación:

$$e = MY$$

```

? matrix MY = M * Y
Se ha generado la matriz MY

? print MY E
MY (10 x 1) [t1 = 1, t2 = 10]
0.13367
0.019811
-0.41298
-0.15379
0.081748
0.19328
0.30514
0.074408
0.014631
-0.25592

E (10 x 1) [t1 = 1, t2 = 10]
0.13367
0.019811
-0.41298
-0.15379
0.081748
0.19328
0.30514
0.074408
0.014631
-0.25592
    
```

Se comprueba fácilmente que son matrices simétricas, idempotentes y ortogonales entre sí:

```

# Simetría de H: H' = H
? matrix HT = H'
Se ha generado la matriz HT

? print HT
HT (10 x 10)
0.12055   0.12345   0.10068   0.058765   0.019414   0.11999   0.14848   0.084099   0.048727   0.17584
0.12345   0.14243   0.14442   0.048353   -0.066478   0.10245   0.18255   0.11135   0.086936   0.12455
0.10068   0.14442   0.22167   0.085860   -0.11031   0.043911   0.17745   0.18168   0.22500   -0.070346
0.058765   0.048353   0.085860   0.18408   0.28351   0.065864   -0.0052496   0.12327   0.18956   -0.034012
0.019414   -0.066478   -0.11031   0.28351   0.77050   0.11851   -0.21961   0.022011   0.084786   0.097665
0.11999   0.10245   0.043911   0.065864   0.11851   0.14591   0.11176   0.046188   -0.0089576   0.25438
0.14848   0.18255   0.17745   -0.0052496   -0.21961   0.11176   0.26168   0.11375   0.058040   0.17114
0.084099   0.11135   0.18168   0.12327   0.022011   0.046188   0.11375   0.16786   0.22528   -0.075487
0.048727   0.086936   0.22500   0.18956   0.084786   -0.0089576   0.058040   0.22528   0.35985   -0.26922
0.17584   0.12455   -0.070346   -0.034012   0.097665   0.25438   0.17114   -0.075487   -0.26922   0.62548

# Idempotencia de H: HH = HHH = ... = H
? matrix HH = H * H
Se ha generado la matriz HH

? print HH
HH (10 x 10) [t1 = 1, t2 = 10]
0.12055   0.12345   0.10068   0.058765   0.019414   0.11999   0.14848   0.084099   0.048727   0.17584
0.12345   0.14243   0.14442   0.048353   -0.066478   0.10245   0.18255   0.11135   0.086936   0.12455
0.10068   0.14442   0.22167   0.085860   -0.11031   0.043911   0.17745   0.18168   0.22500   -0.070346
0.058765   0.048353   0.085860   0.18408   0.28351   0.065864   -0.0052496   0.12327   0.18956   -0.034012
0.019414   -0.066478   -0.11031   0.28351   0.77050   0.11851   -0.21961   0.022011   0.084786   0.097665
0.11999   0.10245   0.043911   0.065864   0.11851   0.14591   0.11176   0.046188   -0.0089576   0.25438
0.14848   0.18255   0.17745   -0.0052496   -0.21961   0.11176   0.26168   0.11375   0.058040   0.17114
0.084099   0.11135   0.18168   0.12327   0.022011   0.046188   0.11375   0.16786   0.22528   -0.075487
    
```

```

0.048727 0.086936 0.22500 0.18956 0.084786 -0.0089576 0.058040 0.22528 0.35985 -0.26922
0.17584 0.12455 -0.070346 -0.034012 0.097665 0.25438 0.17114 -0.075487 -0.26922 0.62548

# Por simetría e idempotencia de H: H'H = HH' = H
? matrix HTH = H' * H
Se ha generado la matriz HTH

? print HTH
HTH (10 x 10) [t1 = 1, t2 = 10]
0.12055 0.12345 0.10068 0.058765 0.019414 0.11999 0.14848 0.084099 0.048727 0.17584
0.12345 0.14243 0.14442 0.048353 -0.066478 0.10245 0.18255 0.11135 0.086936 0.12455
0.10068 0.14442 0.22167 0.085860 -0.11031 0.043911 0.17745 0.18168 0.22500 -0.070346
0.058765 0.048353 0.085860 0.18408 0.28351 0.065864 -0.0052496 0.12327 0.18956 -0.034012
0.019414 -0.066478 -0.11031 0.28351 0.77050 0.11851 -0.21961 0.022011 0.084786 0.097665
0.11999 0.10245 0.043911 0.065864 0.11851 0.14591 0.11176 0.046188 -0.0089576 0.25438
0.14848 0.18255 0.17745 -0.0052496 -0.21961 0.11176 0.26168 0.11375 0.058040 0.17114
0.084099 0.11135 0.18168 0.12327 0.022011 0.046188 0.11375 0.16786 0.22528 -0.075487
0.048727 0.086936 0.22500 0.18956 0.084786 -0.0089576 0.058040 0.22528 0.35985 -0.26922
0.17584 0.12455 -0.070346 -0.034012 0.097665 0.25438 0.17114 -0.075487 -0.26922 0.62548

# Simetría de M: M' = M
? matrix MT = M'
Se ha generado la matriz MT

? print MT
MT (10 x 10)
0.87945 -0.12345 -0.10068 -0.058765 -0.019414 -0.11999 -0.14848 -0.084099 -0.048727 -0.17584
-0.12345 0.85757 -0.14442 -0.048353 0.066478 -0.10245 -0.18255 -0.11135 -0.086936 -0.12455
-0.10068 -0.14442 0.77833 -0.085860 0.11031 -0.043911 -0.17745 -0.18168 -0.22500 0.070346
-0.058765 -0.048353 -0.085860 0.81592 -0.28351 -0.065864 0.0052496 -0.12327 -0.18956 0.034012
-0.019414 0.066478 0.11031 -0.28351 0.22950 -0.11851 0.21961 -0.022011 -0.084786 -0.097665
-0.11999 -0.10245 -0.043911 -0.065864 -0.11851 0.85409 -0.11176 -0.046188 0.0089576 -0.25438
-0.14848 -0.18255 -0.17745 0.0052496 0.21961 -0.11176 0.73832 -0.11375 -0.058040 -0.17114
-0.084099 -0.11135 -0.18168 -0.12327 -0.022011 -0.046188 -0.11375 0.83214 -0.22528 0.075487
-0.048727 -0.086936 -0.22500 -0.18956 -0.084786 0.0089576 -0.058040 -0.22528 0.64015 0.26922
-0.17584 -0.12455 0.070346 0.034012 -0.097665 -0.25438 -0.17114 0.075487 0.26922 0.37452

# Idempotencia de M: MM = MMM = ... = M
? matrix MM = M * M
Se ha generado la matriz MM

? print MM
MM (10 x 10) [t1 = 1, t2 = 10]
0.87945 -0.12345 -0.10068 -0.058765 -0.019414 -0.11999 -0.14848 -0.084099 -0.048727 -0.17584
-0.12345 0.85757 -0.14442 -0.048353 0.066478 -0.10245 -0.18255 -0.11135 -0.086936 -0.12455
-0.10068 -0.14442 0.77833 -0.085860 0.11031 -0.043911 -0.17745 -0.18168 -0.22500 0.070346
-0.058765 -0.048353 -0.085860 0.81592 -0.28351 -0.065864 0.0052496 -0.12327 -0.18956 0.034012
-0.019414 0.066478 0.11031 -0.28351 0.22950 -0.11851 0.21961 -0.022011 -0.084786 -0.097665
-0.11999 -0.10245 -0.043911 -0.065864 -0.11851 0.85409 -0.11176 -0.046188 0.0089576 -0.25438
-0.14848 -0.18255 -0.17745 0.0052496 0.21961 -0.11176 0.73832 -0.11375 -0.058040 -0.17114
-0.084099 -0.11135 -0.18168 -0.12327 -0.022011 -0.046188 -0.11375 0.83214 -0.22528 0.075487
-0.048727 -0.086936 -0.22500 -0.18956 -0.084786 0.0089576 -0.058040 -0.22528 0.64015 0.26922
-0.17584 -0.12455 0.070346 0.034012 -0.097665 -0.25438 -0.17114 0.075487 0.26922 0.37452

# Por simetría e idempotencia de M: M'M = MM' = M
? matrix MTM = M' * M
Se ha generado la matriz MTM

? print MTM
MTM (10 x 10) [t1 = 1, t2 = 10]
0.87945 -0.12345 -0.10068 -0.058765 -0.019414 -0.11999 -0.14848 -0.084099 -0.048727 -0.17584
-0.12345 0.85757 -0.14442 -0.048353 0.066478 -0.10245 -0.18255 -0.11135 -0.086936 -0.12455
-0.10068 -0.14442 0.77833 -0.085860 0.11031 -0.043911 -0.17745 -0.18168 -0.22500 0.070346
-0.058765 -0.048353 -0.085860 0.81592 -0.28351 -0.065864 0.0052496 -0.12327 -0.18956 0.034012
-0.019414 0.066478 0.11031 -0.28351 0.22950 -0.11851 0.21961 -0.022011 -0.084786 -0.097665
-0.11999 -0.10245 -0.043911 -0.065864 -0.11851 0.85409 -0.11176 -0.046188 0.0089576 -0.25438
-0.14848 -0.18255 -0.17745 0.0052496 0.21961 -0.11176 0.73832 -0.11375 -0.058040 -0.17114
-0.084099 -0.11135 -0.18168 -0.12327 -0.022011 -0.046188 -0.11375 0.83214 -0.22528 0.075487
-0.048727 -0.086936 -0.22500 -0.18956 -0.084786 0.0089576 -0.058040 -0.22528 0.64015 0.26922
-0.17584 -0.12455 0.070346 0.034012 -0.097665 -0.25438 -0.17114 0.075487 0.26922 0.37452

# Ortogonalidad ente H y M: HM = 0
? matrix HM=H*M
Se ha generado la matriz HM

? print HM
HM (10 x 10) [t1 = 1, t2 = 10]
-3.3168e-015 -3.4367e-015 -3.7651e-015 -3.5082e-015 -3.0315e-015 -3.1397e-015 -3.4553e-015 -3.7345e-015 -4.0288e-015 -2.5465e-015
-3.3956e-015 -3.6187e-015 -3.6891e-015 -3.1823e-015 -2.4960e-015 -3.2562e-015 -3.7051e-015 -3.5686e-015 -3.5450e-015 -3.1941e-015
-3.8027e-015 -3.7375e-015 -3.0465e-015 -2.7001e-015 -2.5992e-015 -4.0509e-015 -4.1434e-015 -2.9497e-015 -2.0302e-015 -5.3330e-015
-3.6041e-015 -3.2643e-015 -2.7506e-015 -3.8676e-015 -5.4187e-015 -4.1121e-015 -3.0234e-015 -3.0247e-015 -2.8879e-015 -4.7282e-015
-3.2263e-015 -2.6346e-015 -2.7030e-015 -5.4900e-015 -8.7276e-015 -3.7391e-015 -1.4166e-015 -3.6786e-015 -4.3837e-015 -3.0334e-015
-3.1035e-015 -3.3144e-015 -3.9323e-015 -3.9056e-015 -3.6173e-015 -2.9096e-015 -3.1002e-015 -3.9936e-015 -4.7010e-015 -1.5600e-015
-3.4771e-015 -3.6209e-015 -4.1134e-015 -2.9076e-015 -1.2543e-015 -2.9550e-015 -4.0781e-015 -3.6599e-015 -3.6368e-015 -2.7684e-015
-3.7781e-015 -3.6032e-015 -2.8856e-015 -3.0622e-015 -3.5775e-015 -4.0595e-015 -3.7647e-015 -2.8898e-015 -2.2236e-015 -5.3692e-015
-4.1156e-015 -3.6563e-015 -2.1646e-015 -2.8601e-015 -4.4464e-015 -4.7931e-015 -3.8243e-015 -2.2995e-015 -9.1360e-016 -7.5419e-015
-2.4428e-015 -3.1050e-015 -5.2579e-015 -4.5428e-015 -2.6968e-015 -1.5184e-015 -2.7300e-015 -5.1960e-015 -7.3202e-015 2.6132e-015

```

**- Propiedades de los residuos MCO**

Los residuos MCO cumplen una serie de propiedades que se enumeran y se comprueban a continuación:

**P1. La suma de los residuos es nula**

$$\sum_{t=1}^T e_t = 0 \Rightarrow t'e = 0 \text{ donde } t' = (1 \ 1 \ \dots \ 1)_{1 \times T}$$

```
# Suma residual nula
? genr se = sum(E)
Se ha generado el escalar se = -2.63967e-012
```

**Consecuencia:** La suma de los valores observados del regresando es igual a la suma de sus valores estimados  $\Rightarrow$  La media del regresando coincide con la media de su valor estimado.

$$\sum_{t=1}^T Y_t = \sum_{t=1}^T \hat{Y}_t \Rightarrow \bar{Y} = \bar{\hat{Y}}$$

```
# Suma de los valores observados igual a la suma de los valores estimados
? genr sy = sum(Y)
Se ha generado el escalar sy = 103

? genr sye = sum(YE)
Se ha generado el escalar sye = 103

# Media de los valores observados igual a la media de los valores estimados
? genr ym = mean(Y)
Se ha generado el escalar ym = 10.3

? genr yem = mean(YE)
Se ha generado el escalar yem = 10.3
```

**P2. Los residuos y los regresores presentan incorrelación muestral**

$$X'e = 0_{(K+1) \times 1} \Rightarrow \sum_{t=1}^T X_{it}e_t = 0$$

```
# Incorrelación entre residuos y regresores
? matrix XE = X' * E
Se ha generado la matriz XE

? print XE
XE (3 x 1)
-2.6397e-012
-6.7777e-012
-2.7423e-009
```

**P3. No existe correlación muestral entre la variable explicada estimada y los residuos**

$$\hat{Y}'e = 0 \Rightarrow \sum_{t=1}^T \hat{Y}_t e_t = 0$$

```
# Incorrelación entre residuos y valores estimados del regresando
? genr YEE = YE' * E
Se ha generado el escalar YEE = -2.6876e-011
```

**P4. El hiperplano de regresión ajustado por MCO pasa por el punto de valores medios  $(\bar{Y}, \bar{X}_1, \bar{X}_2, \dots, \bar{X}_K)$** 

```
# La recta de regresión pasa por (ym, x1m, x2m x3m x4m)
? genr ymc = b[1] + b[2] * mean(P) + b[3] * mean(RF)
Se ha generado el escalar ymc = 10.3
```

**3.11.2. Interpretación de coeficientes**

Los estimadores de los coeficientes de regresión parcial dependen de las unidades de medida de las variables a las que acompañan (por lo que no son directamente comparables). Por ello, que para enriquecer la interpretación de los resultados de la estimación del modelo resulta conveniente calcular estimadores adimensionales.

**- Estimadores de los coeficientes estandarizados**

$$b_i^* = b_i \frac{S_{X_i}}{S_Y}$$

```
# Coeficientes estandarizados
? genr b0stand = b[1] * sqrt(var(const)) / sqrt(var(C))
Se ha generado el escalar b0stand = 0

? genr b1stand = b[2] * sqrt(var(P)) / sqrt(var(C))
Se ha generado el escalar b1stand = -0.614544

? genr b2stand = b[3] * sqrt(var(RF)) / sqrt(var(C))
Se ha generado el escalar b2stand = 0.661434
```

De los resultados obtenidos se deduce que:

- Manteniendo constante el nivel de renta, ante un incremento de una unidad de desviación estándar en el precio del litro de aceite de oliva, la demanda de aceite disminuye en 0.614544 unidades de desviación estándar.
- Manteniendo constante el precio del litro de aceite de oliva, ante un incremento de una unidad de desviación estándar en el nivel de renta de las familias, la demanda de aceite aumenta en 0.661434 unidades de desviación estándar.

Se puede concluir que la variable que tiene mayor peso en la explicación del consumo mensual de aceite por las familias (C) es el nivel de ingresos mensuales (RF) ya que es la variable con el mayor estimador del coeficiente estandarizado en valor absoluto.

Además, si los regresores son ortogonales, el cuadrado de los coeficientes beta mide la varianza del regresando explicada por esa variable y, por tanto, el coeficiente de determinación vendrá dado por la suma de los cuadrados de dichos coeficientes beta. No obstante, no es habitual trabajar con regresores ortogonales y, por tanto, es imposible aislar del todo el efecto de cada uno de ellos.

```
# Se calculan los cuadrados de los estimadores de los coeficientes beta
? genr b1stand2 = b1stand^2
Se ha generado el escalar b1stand2 = 0.377664

? genr b2stand2 = b2stand^2
Se ha generado el escalar b2stand2 = 0.437495

? genr emulti = 2 * b[2] * b[3] * cov(P,RF) / var(C)
Se ha generado el escalar emulti = 0.1587

? genr R2cemulti = b1stand2 + b2stand2 + emulti
Se ha generado el escalar R2cemulti = 0.973859
```

El  $R^2$  indica que aproximadamente el 97.38% de las variaciones en el consumo mensual de aceite por las familias, son explicadas por las variaciones en el precio del aceite y en el nivel de ingresos mensuales. Además de ese 97% de variaciones en el consumo mensual de aceite por las familias, el 37.77% son explicadas por variaciones en el precio del aceite, el 43.74% son explicadas por variaciones en el nivel de ingresos y el 15.87 % son explicadas por variaciones conjuntas en el nivel de precios y en el nivel de ingresos, dado que ambas variables están correlacionadas:

```
? genr corPRF = corr(P,RF)
Se ha generado el escalar corPRF = -0.195213
```

Otra forma de calcular los estimadores de los coeficientes beta es realizando la estimación del modelo con todas sus variables estandarizadas:

$$\frac{C_t - \bar{C}}{S_C} = \beta_1^* \frac{C_t - \bar{C}}{S_C} + \beta_2^* \frac{RF_t - \bar{RF}}{S_{RF}} + \varepsilon_t^*$$

```
# Variables estandarizadas
? genr Cs = (C - mean(C)) / sqrt(var(C))
Se ha generado la serie Cs (ID 10)
```

```
? genr Ps = (P - mean(P)) / sqrt(var(P))
Se ha generado la serie Ps (ID 11)

? genr RFs = (RF - mean(RF)) / sqrt(var(RF))
Se ha generado la serie RFs (ID 12)

# Estimación del modelo con variables estandarizadas
? ols Cs Ps RFs --simple-print

MCO, usando las observaciones 1-10
Variable dependiente: Cs
```

	Coeficiente	Desv. Típica	Estadístico t	Valor p
Ps	-0.614544	0.0582843	-10.54	5.71e-06 ***
RFs	0.661434	0.0582843	11.35	3.28e-06 ***

```
SCR = 0.235267, R-cuadrado = 0.973859
```

El usuario debe tener en cuenta que Gretl denota la SCE (Suma de Cuadrados de los Errores o Residuos) como SCR. En este manual se utiliza SCR para denominar a la Suma de Cuadrados de Regresión.

Cuando las variables de un modelo están centradas, la ordena en el origen se anula y, dado que las variables estandarizadas no son más que las variables centradas divididas por sus desviaciones estándar, no será necesario introducir ordenada en el origen en el modelo:

```
? ols Cs const Ps RFs --simple-print

MCO, usando las observaciones 1-10
Variable dependiente: Cs
```

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	0.000000	0.0579738	0.0000	1.0000
Ps	-0.614544	0.0623085	-9.863	2.34e-05 ***
RFs	0.661434	0.0623085	10.62	1.44e-05 ***

```
SCR = 0.235267, R-cuadrado = 0.973859
```

Se observa que cuando se incluye el regresor ficticio, el estimador de la ordenada en el origen es cero y la probabilidad de rechazar que ese sea su verdadero valor es uno. Otra evidencia de que realmente la ordenada en el origen es cero, es que el coeficiente de determinación del modelo con ordenada en el origen coincide con el coeficiente de determinación bruto del modelo sin ordenada en el origen.

Además, si se comparan los resultados de la estimación del modelo con y sin ordenada en el origen, se obtienen los mismos resultados para los estimadores de los coeficientes beta de las variables explicativas y lo único que varía son los estimadores de las desviaciones estándar, debido a que los grados de libertad en ambos modelos es diferente (en el primero se incluyen dos regresores mientras que en el segundo se incluyen tres). Lo que sí resulta relevante y pone de manifiesto que los estimadores de los coeficientes de las variables estandarizadas son directamente comparables es que sus desviaciones típicas estimadas coinciden.

**- Estimadores de las elasticidades en media**

$$\hat{E}_i = b_i \frac{\bar{X}_i}{\bar{Y}}$$

```
# Elasticidades en media
? genr EP = b[2] * mean(P) / mean(C)
Se ha generado el escalar EP = -0.728034

? genr ERF = b[3] * mean(RF) / mean(C)
Se ha generado el escalar ERF = 0.923513
```

De los resultados obtenidos se deduce que:

- Manteniendo constante el nivel de renta, ante un incremento de un 1% en el precio del litro de aceite de oliva, la demanda de aceite, disminuye un 0.728034%.
- Manteniendo constante el precio del litro de aceite de oliva, ante un incremento de un 1% en el nivel de renta de las familias, la demanda de aceite aumenta en 0.923513%.

Otra forma de calcular los estimadores de las elasticidades es con la estimación del modelo con todas sus variables tomando logaritmos:

$$\ln C_t = \beta_0 + E_1 \ln P_t + E_2 \ln RF_t + \varepsilon_t^*$$

```
# Variables en términos logarítmicos
? genr LC = log(C)
Se ha generado la serie LC (ID 13)

? genr LP = log(P)
Se ha generado la serie LP (ID 14)

? genr LRF = log(RF)
Se ha generado la serie LRF (ID 15)

# Estimación del modelo con variables en términos logarítmicos
? ols LC const LP LRF --simple-print

MCO, usando las observaciones 1-10
Variable dependiente: LC
```

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	-3.47164	0.701716	-4.947	0.0017	***
LP	-0.665544	0.0854378	-7.790	0.0001	***
LRF	0.923194	0.0981724	9.404	3.20e-05	***

```
SCR = 0.00593285, R-cuadrado = 0.963157
```

Las discrepancias existentes con los resultados anteriores son debidas a que en el modelo logarítmico se obtienen estimadores de las elasticidades mientras que con la primera aproximación se obtienen estimadores de las elasticidades medias.

### 3.11.3. Análisis gráfico

Para hacerse una idea de la “calidad” de la estimación puede resultar interesante la representación gráfica de los valores del regresando y del regresando estimado (véase Ilustración 3-8), así como de los residuos (véase Ilustración 3-9), ya sea respecto al orden de las observaciones o de las variables observables que intervienen en el modelo. La forma más cómoda de acceder a estos gráficos es a través del menú **Gráficos** de la **Ventana Modelo**, en la que se podrán modificar de forma sencilla los atributos de los mismos. Se puede acceder a esta **Ventana Gráfico** ejecutando el correspondiente comando **gnuplot** desde la **Consola Gretl**.

Si el comando **gnuplot** se ejecuta desde un fichero de comandos, no se tiene acceso directamente al gráfico, sino que Gretl en la **salida de guión** informa de la ruta donde se encuentra el fichero que hay que ejecutar para poder acceder a dicho gráfico (por lo que resulta más complicado). En este caso, para modificar sus atributos, el usuario tendrá que estar familiarizado con el **programa gráfico gnuplot**, mientras que si se hace a través de la **Ventana Gráfico**, las modificaciones resultan mucho más sencillas.

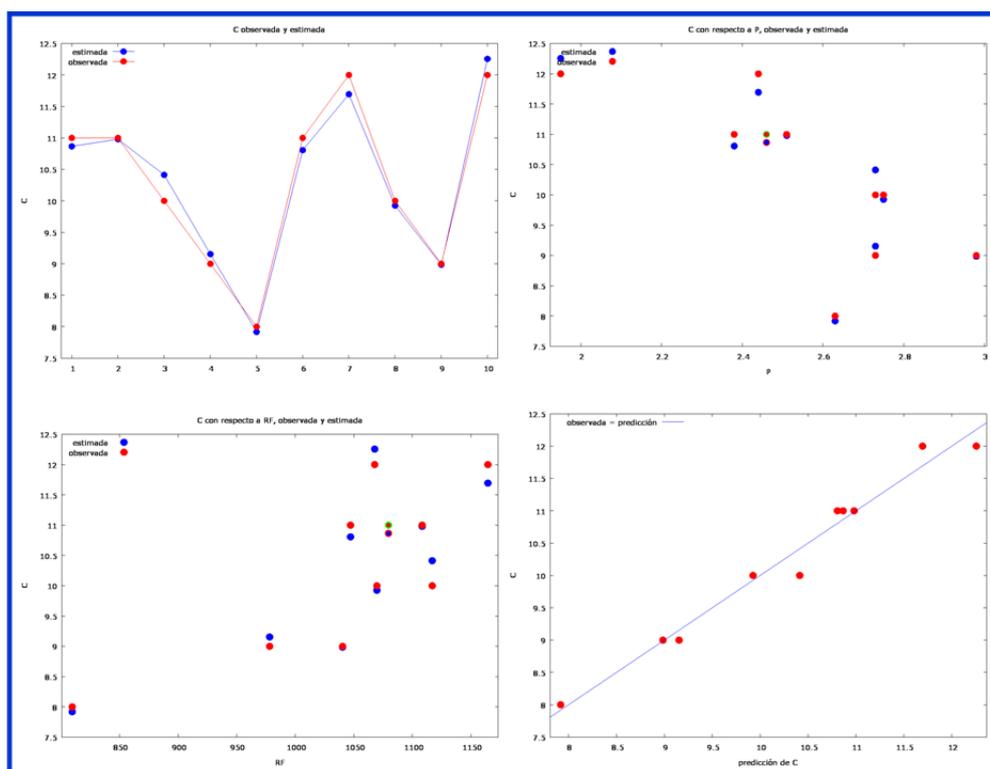


Ilustración 3-8. Gráficos de valores observados y estimados del regresando.

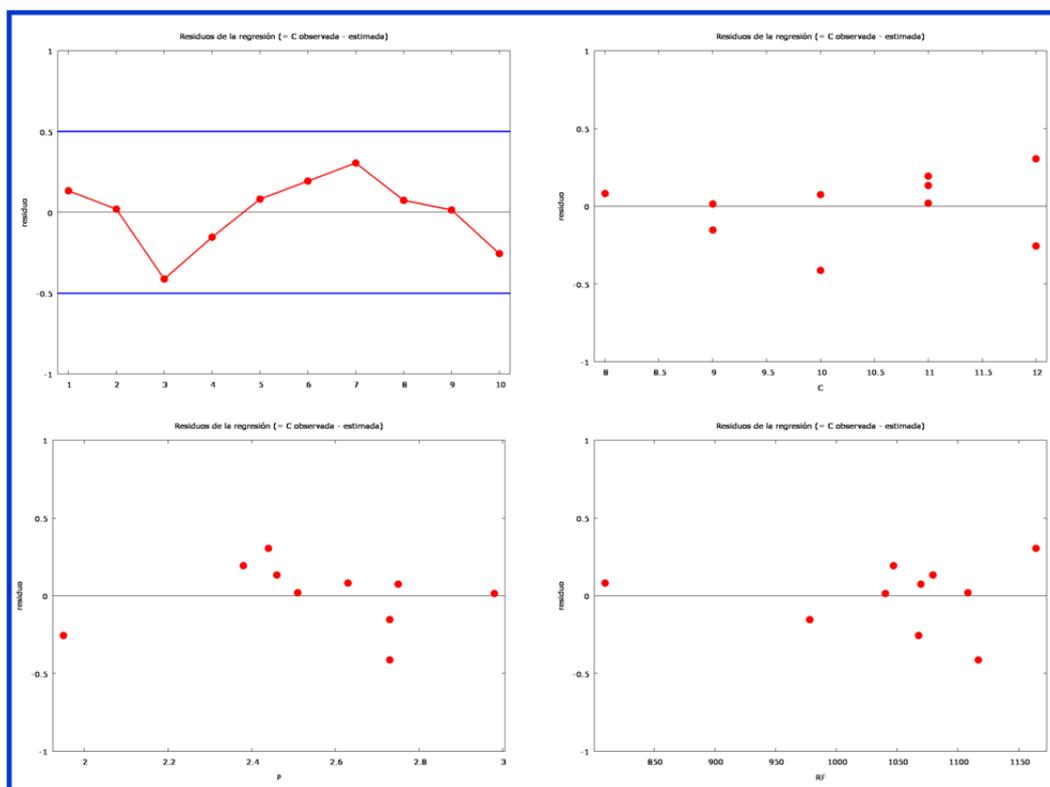


Ilustración 3-9. Gráficos de residuos.

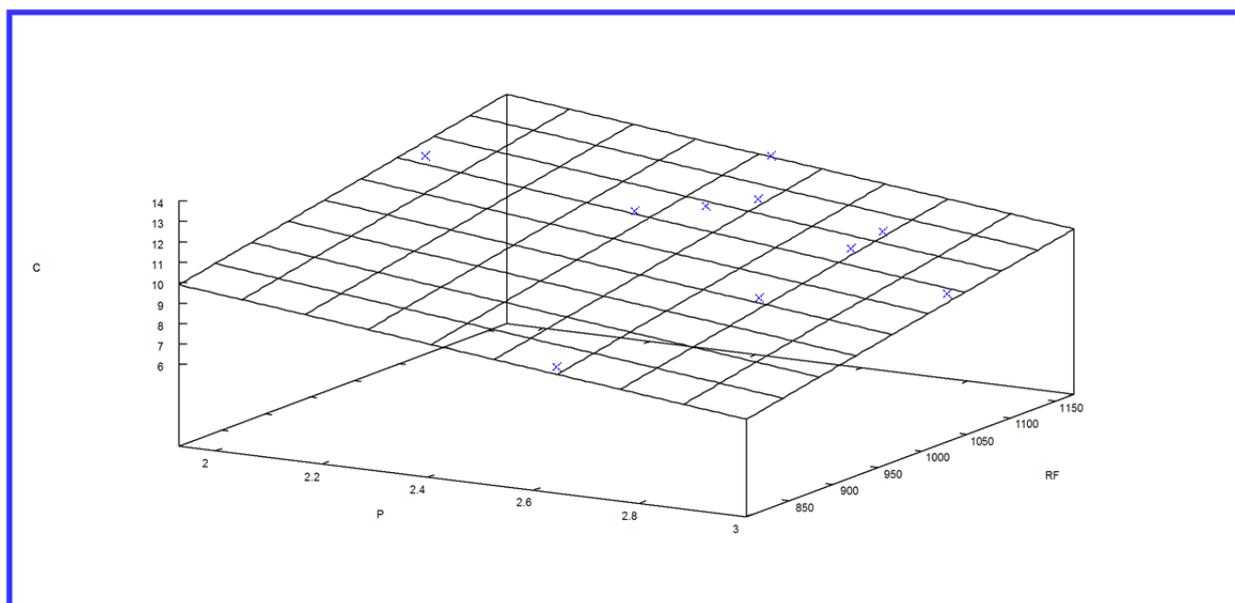


Ilustración 3-10 Gráfico de valores estimados y observados del regresando respecto a las variables explicativas.

Otra forma de construir gráficos es utilizando el comando **graph**. Los gráficos así obtenidos, al ser de texto plano son de peor calidad, pero quizás sean los más apropiados para utilizar en los ficheros de comandos y, recurrir a los de alta calidad, únicamente para los gráficos que deban ir en el informe final del proyecto.

Los gráficos de la Ilustración 3-9 se podrían obtener a partir de la ejecución de los comandos:

```
# Se genera la serie o
? series o = time
Se ha generado la serie o (ID 17)

# Gráfico de dispersión de los residuos
? textplot e o

e
0.305141 |
0.185453 +
| o
0.0 +-----+
-0.213505 +
-0.412985 +
|+-----+-----+-----+-----+
1 10

# Gráfico de dispersión de los residuos
? gnuplot e o
escribió c:\proyecto01\gpttmp01.plt

# Gráfico de dispersión de los residuos respecto a la variable "C"
# textplot e C
```

```
# Gráfico de dispersión de los residuos respecto a la variable "P"
# textplot e P

# Gráfico de dispersión de los residuos respecto a la variable "RF"
# textplot e RF

# Gráfico de dispersión de los residuos respecto a la variable "C"
# gnuplot e C

# Gráfico de dispersión de los residuos respecto a la variable "P"
# gnuplot e P

# Gráfico de dispersión de los residuos respecto a la variable "RF"
# gnuplot e RF
```

Cuando se ejecuta un comando **textplot** aparece un gráfico de texto plano y cuando se ejecuta un comando **gnuplot** aparece la ruta donde se encuentra el fichero correspondiente a dicho gráfico (se ha hecho únicamente para el primer gráfico, pero si se está interesado en obtener los restantes, se debe ejecutar el comando correspondiente y obviar el símbolo # que aparece al principio de la línea del comando).

### 3.11.4. Estimación MCO del modelo sin ordenada en el origen

```
# Modelo sin ordenada en el origen con anova
? ols C P RF --anova

Modelo 2: MCO, usando las observaciones 1-10
Variable dependiente: C
```

	Coefficiente	Desv. Típica	Estadístico t	Valor p
P	-1.53921	0.502789	-3.061	0.0156 **
RF	0.0135517	0.00122794	11.04	4.05e-06 ***

```

Media de la vble. dep. 10.30000 D.T. de la vble. dep. 1.337494
Suma de cuad. residuos 2.918482 D.T. de la regresión 0.603995
R-cuadrado 0.997290 R-cuadrado corregido 0.996951
F(2, 8) 1472.110 Valor p (de F) 5.39e-11
Log-verosimilitud -8.031779 Criterio de Akaike 20.06356
Criterio de Schwarz 20.66873 Crit. de Hannan-Quinn 19.39969
```

Obsérvese que aún cuando el comando **ols** se ha ejecutado con la opción **--anova**, dado que el modelo carece de ordenada en el origen, Gretl no muestra dicha tabla.

#### - Interpretación de resultados

El modelo estimado por MCO es:

$$\hat{C}_t = - \frac{1.53921}{(0.502789)} P_t + \frac{0.0135517}{(0.00122794)} RF_t$$

$$t_1 = -3.061 \quad t_2 = 11.04$$

$$R^2 = 0.99729$$

De los resultados obtenidos se deduce que:

- Manteniendo constante el nivel de renta, ante un incremento de un euro en el precio del litro de aceite de oliva, la demanda de aceite (medida en términos de consumo mensual por familia), disminuye en 1.53921 litros.
- Manteniendo constante el precio del litro de aceite de oliva, ante un incremento de un euro en el nivel de renta de las familias, la demanda de aceite aumenta en 0.0135517 litros.

Dado que el modelo está formulado sin ordenada en el origen, bajo la denominación de coeficiente de determinación se calcula el coeficiente de determinación bruto, que aunque está acotado entre 0 y 1, carece de la interpretación habitual.

### 3.11.4.1. Resultados de la estimación utilizando álgebra matricial

En este epígrafe, con la finalidad de entender las diferencias de cálculo entre el modelo con y sin ordenada en el origen, se calculan de forma detallada los resultados de la estimación del modelo.

#### - Vector de estimadores

Para obtener los estimadores mínimo cuadráticos ordinarios del vector de parámetros es necesario conocer los elementos de la matriz  $X'X$  y del vector  $X'Y$ , que se obtendrán a partir de los datos de la muestra:

$$b = (X'X)^{-1}X'Y = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{pmatrix}_{K \times 1}$$

donde:

- $X'X$  es la matriz de orden  $K \times K$ , que contiene los sumatorios de los productos cruzados de las variables explicativas:

$$X'X = \begin{pmatrix} \sum_{t=1}^T X_{1t}^2 & \sum_{t=1}^T X_{1t}X_{2t} & \dots & \sum_{t=1}^T X_{1t}X_{Kt} \\ \sum_{t=1}^T X_{2t}X_{1t} & \sum_{t=1}^T X_{2t}^2 & \dots & \sum_{t=1}^T X_{2t}X_{Kt} \\ \dots & \dots & \ddots & \dots \\ \sum_{t=1}^T X_{Kt}X_{1t} & \sum_{t=1}^T X_{Kt}X_{2t} & \dots & \sum_{t=1}^T X_{Kt}^2 \end{pmatrix}_{K \times K}$$

- $X'Y$  es el vector de orden  $K \times 1$ , que contiene los sumatorios de los productos cruzados del regresando con las variables explicativas:

$$X'Y = \begin{pmatrix} \sum_{t=1}^T Y_t X_{1t} \\ \sum_{t=1}^T Y_t X_{2t} \\ \dots \\ \sum_{t=1}^T Y_t X_{Kt} \end{pmatrix}_{K \times 1}$$

```
# Se genera la matriz de diseño (Xso)
? matrix Xso = {P, RF}
Se ha generado la matriz Xso

# Se muestra la matriz Xso que, en este caso, es de orden 10x2
? print Xso
Xso (10 x 2) [t1 = 1, t2 = 10]
  2.4600    1079.5
  2.5100    1108.2
  2.7300    1116.8
  2.7300    978.09
  2.6300    809.62
```

2.3800	1047.1
2.4400	1164.3
2.7500	1069.6
2.9800	1040.3
1.9500	1067.8

```
# Se calcula y se muestra la matriz de productos cruzados de las variables explicativas
? matrix XTXso = Xso'* Xso
Se ha generado la matriz XTXso

? print XTXso
XTXso (2 x 2)
    66.038    26743.
   26743.    1.1072e+007

# Se calcula y se muestra la inversa de X'Xso
? gener XTXsoI = inv(XTXso)
Se ha generado la matriz XTXsoI

? print XTXsoI
XTXsoI (2 x 2)
    0.69295   -0.0016738
   -0.0016738  4.1332e-006

# Se calcula y se muestra el vector de productos cruzados de las variables explicativas y el
regresando
? matrix XTsoY = Xso'* Y
Se ha generado la matriz XTsoY

? print XTsoY
XTsoY (2 x 1)
    260.76
   1.0888e+005

# Se calcula bso
? matrix bso = XTXsoI * XTsoY
Se ha generado la matriz bso

? print bso
bso (2 x 1)
   -1.5392
    0.013552
```

### - Regresando estimado

$$\hat{Y} = Xb$$

```
# Se calcula el regresando estimado
? matrix YEso = Xso * bso
Se ha generado la matriz YEso
? print YEso
YEso (10 x 1) [t1 = 1, t2 = 10]

    10.843
    11.155
    10.933
    9.0527
    6.9236
    10.526
    12.023
    10.262
    9.5117
    11.468
```

Si lo que interesa obtener es la estimación para una única observación del regresando, se puede utilizar:

$$\hat{Y}_t = X'_t b = b_1 X_{1t} + \dots + b_K X_{Kt}$$

```
# Se calcula y se muestra el valor estimado para la primera observación
? gener YEso1 = bso[1] * Xso[1,1] + bso[2] * Xso[1,2]
Se ha generado el escalar YEso1 = 10.8427
```

Para generar la serie que contiene los valores estimados del regresando se utiliza :

```
# Se genera y se muestra la serie que recoge las observaciones del regresando estimado (ye)
? series yeso = Xso * bso
Se ha generado la serie yeso (ID 17)

? print yeso
Muestra actual: 1 - 10 (n = 10)
10.8427 11.1548 10.9330 9.05273 6.92359 10.5264 12.0231 10.2621
9.51165 11.4685
```

#### - Residuos

$$e = Y - \hat{Y}$$

```
# Se calcula el vector de residuos
? matrix Eso = Y - YEso
Se ha generado la matriz Eso
? print Eso
Eso (10 x 1) [t1 = 1, t2 = 10]

0.15727
-0.15484
-0.93302
-0.052727
1.0764
0.47362
-0.023102
-0.26206
-0.51165
0.53151
```

Para obtener un residuo para una observación concreta:

$$e_t = Y_t - \hat{Y}_t$$

```
# Se calcula y se muestra el valor estimado para la primera observación
? genr esol = Y[1] - YEso[1]
Se ha generado el escalar esol = 0.157273
```

Para obtener la serie de residuos:

```
# Se genera y se muestra la serie de residuos
? series eso = Y - YEso
Se ha generado la serie eso (ID 18)
? print eso
Muestra actual: 1 - 10 (n = 10)
0.157273 -0.154836 -0.933024 -0.0527265 1.07641 0.473617 -0.0231017
-0.262058 -0.511652 0.531507
```

#### - Suma de Cuadrados de Errores

$$SCE = e'e = \sum_{t=1}^T e_t^2 = Y'Y - b'X'Y$$

```
# Se calcula la SCE
? genr SCEso = Eso' * Eso
Se ha generado el escalar SCEso = 2.91848
```

#### - Estimador de la varianza de la perturbación

$$S^2 = \frac{SCE}{T - K}$$

```
# Se calcula el estimador de la varianza de la perturbación
? genr s2so = SCEso / (T - k)
Se ha generado el escalar s2so = 0.36481

# Se calcula el estimador de la desviación típica de la perturbación
? genr sso = sqrt(s2so)
Se ha generado el escalar sso = 0.603995
```

#### - Matriz de varianzas-covarianzas estimada de los EMCO

$$\widehat{V}(b) = S^2 (X'X)^{-1} = \begin{pmatrix} S_{b_1}^2 & S_{b_1 b_2} & \dots & S_{b_1 b_K} \\ S_{b_2 b_1} & S_{b_2}^2 & \dots & S_{b_2 b_K} \\ \dots & \dots & \ddots & \dots \\ S_{b_K b_1} & S_{b_K b_2} & \dots & S_{b_K}^2 \end{pmatrix}_{K \times K}$$

```
# Se calcula y se muestra la matriz de varianzas-covarianzas estimada de los estimadores
? matrix Vbso = s2so * inv(Xso' * Xso)
Se ha generado la matriz Vbso

? print Vbso
Vbso (2 x 2)
    0.25280  -0.00061061
 -0.00061061  1.5078e-006

# Se calcula y se muestra el vector que contiene las varianzas estimadas de los EMCO
? matrix s2bso = diag(Vbso)
Se ha generado la matriz s2bso

? print s2bso
s2bso (2 x 1)
    0.25280
    1.5078e-006

# Se calcula y se muestra el vector que contiene las desviaciones típicas estimadas de los EMCO
? matrix sbso = sqrt(s2bso)
Se ha generado la matriz sbso

? print sbso
sbso (2 x 1)
    0.50279
    0.0012279
```

#### - Ratios t de los EMCO

$$t_i = \frac{b_i}{S_{b_i}}$$

```
# Se calcula las ratios t
? genr t1so = bso[1] / sbso[1]
Se ha generado el escalar t1so = -3.06135

? genr t2so = bso[2] / sbso[2]
Se ha generado el escalar t2so = 11.0361
```

#### - Tabla ANOVA respecto a la media y Tabla ANOVA respecto al origen

Aunque Gretl para modelos sin ordenada en el origen no proporciona ninguna de las tablas de Análisis de la Varianza, se ha considerado importante construirlas para que se pueda entender como se calcula la bondad de ajuste para estos modelos y para que quede claro que aunque el coeficiente calculado está acotado, no tiene la interpretación habitual, puesto que su cálculo se basa en la Tabla ANOVA respecto al origen (véase **Tabla 3-8**) y no en la Tabla ANOVA respecto a la media (véase **Tabla 3-7**).

```
# ANOVA respecto a la media
# Se calcula la SCT
? genr SCT = sum((y - mean(y))^2)
Se ha reemplazado el escalar SCT = 16.1

# Se calcula la SCR
? genr SCRso = sum((yeso - mean(yeso))^2)
Se ha generado el escalar SCRso = 19.3814

# Se calcula la SCE
? genr SCEso = Eso' * Eso
Se ha reemplazado el escalar SCEso = 2.91848
```

**Tabla 3-7.** Tabla ANOVA respecto a la media.

Análisis de Varianza:

	Sumas de cuadrados	gl	Media de cuadrados
Regresión $SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$	19.3814	2	...
Residuo $SCE = \sum_{t=1}^T e_t^2$	2.91848	8	...
Total $SCT = \sum_{t=1}^T (Y_t - \bar{Y})^2$	16.1	9	...

$R^2 = 19.3814 / 16.1 = 1.2038$

...

```
# ANOVA respecto al origen
? genr scy = sum(y^2)
Se ha reemplazado el escalar scy = 1077

? genr scyeso = sum(yeso^2)
Se ha generado el escalar scyeso = 1074.08

? genr sceso = sum(eso^2)
Se ha generado el escalar sceso = 2.91848
```

**Tabla 3-8.** Tabla ANOVA respecto al origen.

Análisis de Varianza:

	Sumas de cuadrados	gl	Media de cuadrados
Regresión $\sum_{t=1}^T (\hat{Y}_t - 0)^2 = \sum_{t=1}^T \hat{Y}_t^2$	1074.08	2	537.04
Residuo $\sum_{t=1}^T (e_t - 0)^2 = \sum_{t=1}^T e_t^2 = SCE$	2.91848	8	0.36481
Total $\sum_{t=1}^T (Y_t - 0)^2 = \sum_{t=1}^T Y_t^2$	1077	10	107.7

$R^2 = 1074.08 / 1077 = 0.9972$

$F(2, 8) = 537.04 / 0.36481 = 1472.110$  [Valor p 2.74605e-012]

**- Cálculo del coeficiente de determinación**

Como el modelo está formulado sin ordenada en el origen:

$$R^2 = \frac{SCR}{SCT} \neq 1 - \frac{SCE}{SCT}$$

```
# Cálculo del coeficiente de determinación
? genr R2soc1 = SCRso/SCT
Se ha generado el escalar R2soc1 = 1.20381

? genr R2soc2 = 1 - SCEso/SCT
Se ha generado el escalar R2soc2 = 0.818728
```

Se debe tener en cuenta que si en un modelo sin ordenada en el origen se calcula el coeficiente de determinación como  $SCR / SCT$ , aunque la cota inferior seguirá siendo cero, dejará de estar acotado superiormente (como en este caso, podrá ser superior a uno).

Si se calcula dicho coeficiente como  $1 - (SCE / SCT)$ , seguirá teniendo cota superior uno, pero ya no estará acotado inferiormente (pudiendo en este caso, tomar valores negativos).

#### - Coeficiente de determinación bruto

Se puede comprobar que el valor que aparece bajo la denominación **R-cuadrado** no es el coeficiente de determinación propiamente dicho sino el coeficiente de determinación bruto: **R-cuadrado 0.997290**, cálculo que se basa en la Tabla ANOVA respecto al origen

```
# Cálculo del coeficiente de determinación bruto
? genr R2RAWso = scyeso/scy
Se ha generado el escalar R2RAWso = 0.99729

? genr R2RAWso = 1 - sceso/scy
Se ha reemplazado el escalar R2RAWso = 0.99729
```

#### - Coeficiente de determinación corregido o ajustado

$$\bar{R}^2 = 1 - \frac{\frac{SCE}{T-K}}{\frac{\sum_{t=1}^T Y_t^2}{T-1}}$$

```
# Se calcula el coeficiente de determinación ajustado
? genr R2Aso = 1 - ((SCEso / (T - k)) / (scy / (T - 1)))
Se ha generado el escalar R2Aso = 0.996951
```

#### - Propiedades de los residuos MCO

Cuando el modelo está formulado sin ordenada en el origen, los residuos MCO cumplen las siguientes propiedades:

##### P1. La suma de los residuos no es nula

$$\sum_{t=1}^T e_t \neq 0 \Rightarrow t'e \neq 0 \text{ donde } t' = (1 \ 1 \ \dots \ 1)_{1 \times T}$$

```
# Suma residual no es nula
? genr se = sum(eso)
Se ha reemplazado el escalar se = 0.301405
```

**Consecuencia:** La suma de los valores observados del regresando no es igual a la suma de sus valores estimados  $\Rightarrow$  La media del regresando no coincide con la media de su valor estimado.

$$\sum_{t=1}^T Y_t \neq \sum_{t=1}^T \hat{Y}_t \Rightarrow \bar{Y} \neq \bar{\hat{Y}}$$

```
# Suma de los valores observados no es igual a la suma de los valores estimados
? genr sy = sum(y)
Se ha reemplazado el escalar sy = 103

? genr syeso = sum(yeso)
Se ha generado el escalar syeso = 102.699

# Media de los valores observados no es igual a la media de los valores estimados
? genr ym = mean(y)
Se ha reemplazado el escalar ym = 10.3

? genr yesom = mean(yeso)
Se ha generado el escalar yesom = 10.2699
```

##### P2. Los residuos y las variables explicativas presentan incorrelación muestral

$$X'e = 0_{K \times 1} \Rightarrow \sum_{t=1}^T X_{it} e_t = 0$$

```
# Incorrelación entre residuos y variables explicativas
? matrix XsoEso = Xso' * Eso
Se ha generado la matriz XsoEso

? print XsoEso
XsoEso (2 x 1)
7.3632e-013
```

3.0567e-010

**P3. No existe correlación muestral entre la variable explicada estimada y los residuos**

$$\widehat{Y}'e = 0 \Rightarrow \sum_{t=1}^T \widehat{Y}_t e_t = 0$$

```
# Incorrelación entre residuos y valores estimados
? genr YEsosEso = YEsos' * Eso
Se ha generado el escalar YEsosEso = 3.00679e-012
```

**P4. El hiperplano de regresión ajustado por MCO no pasa por el punto de valores medios**

El hiperplano de regresión ajustado no pasa por el punto  $(\bar{Y}, \bar{X}_1, \bar{X}_2, \dots, \bar{X}_K)$  sino que pasa por el punto  $(\widehat{Y}, \bar{X}_1, \bar{X}_2, \dots, \bar{X}_K)$ :

```
# La recta de regresión pasa por (yem, x1m, x2m x3m x4m)
? genr yesomc = bso[1] * mean(P) + bso[2] * mean(RF)
Se ha generado el escalar yesomc = 10.2699

? print yesomc yesom ym

      yesomc = 10.269859
      yesom  = 10.269859
      ym     = 10.300000
```

**3.11.5. Formas funcionales alternativas**

Debe tenerse en cuenta, que la interpretación de los coeficientes estimados es distinta dependiendo de la forma funcional del modelo especificado y que el coeficiente de determinación, tan sólo, se puede utilizar para comparar modelos que tengan la misma variable dependiente, el mismo número de regresores e idéntica forma funcional. En este caso, no se podría utilizar como criterio de selección de modelos, puesto que no serían directamente comparables, por ello, para poder realizar dicha comparación se ha calculado el Coeficiente de determinación equivalente:

$$R_{equi}^2 = 1 - \frac{SCE_{en\ unidades\ de\ Y}}{SCT_{en\ unidades\ de\ Y}}$$

```
# Transformación de las variables
? genr LC = log(C)
Se ha generado la serie LC (ID 7)

? genr LP = log(P)
Se ha generado la serie LP (ID 8)

? genr LRF = log(RF)
Se ha generado la serie LRF (ID 9)

? genr PI = 1 / P
Se ha generado la serie PI (ID 10)

? genr RFI = 1 / RF
Se ha generado la serie RFI (ID 11)
```

**- Modelo lineal o Modelo Lin-Lin**

$$C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \varepsilon_t$$

```
# Modelo 1: Modelo Lin-Lin
? ols C const P RF --simple-print

MCO, usando las observaciones 1-10
Variable dependiente: C

-----
      Coeficiente   Desv. Típica   Estadístico t   Valor p
-----
const      8.28657      1.28569         6.445          0.0004 ***
P          -2.93379      0.297456       -9.863         2.34e-05 ***
RF          0.00907529    0.000854912    10.62          1.44e-05 ***

SCR = 0.420867, R-cuadrado = 0.973859
```

$b_1 = \partial C / \partial P \Rightarrow b_1 = \Delta C / \Delta P \rightarrow$  Manteniendo constante el nivel de renta de las familias, si se incrementa en 1 euro el precio del litro de aceite, la cantidad demandada disminuye en 2.934 litros.

$b_2 = \partial C / \partial RF \Rightarrow b_2 = \Delta C / \Delta RF \rightarrow$  Manteniendo constante el precio del litro de aceite, si se incrementa en 1 euro la renta de las familias, la cantidad demandada aumenta en 0.009075 litros.

Dado que la variable dependiente es lineal, el coeficiente de determinación equivalente coincide con el coeficiente de determinación.

```
# Cálculo del coeficiente de determinación equivalente
? series CEM1 = $yhat
Se ha generado la serie CEM1 (ID 12)

? series EEQUIM1 = C - CEM1
Se ha generado la serie EEQUIM1 (ID 13)

? genr SCEQUIM1 = sum(EEQUIM1^2)
Se ha generado el escalar SCEQUIM1 = 0.420867

? genr R2EQUIM1 = 1 - SCEQUIM1 / (sum((C - mean(C))^2))
Se ha generado el escalar R2EQUIM1 = 0.973859
```

- **Modelo Lin-Log**

$$C_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln RF_t + \varepsilon_t$$

```
# Modelo 2: Modelo Lin-Log
? ols C const LP LRF --simple-print

MCO, usando las observaciones 1-10
Variable dependiente: C
```

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	-44.0633	7.49873	-5.876	0.0006	***
LP	-7.05398	0.913011	-7.726	0.0001	***
LRF	8.76808	1.04910	8.358	6.89e-05	***

```
SCR = 0.677511, R-cuadrado = 0.957919
```

$b_1 = \partial C / \partial \ln P \Rightarrow b_1 = \Delta C / (\Delta P / P) \rightarrow$  Manteniendo constante el nivel de renta de las familias, si se incrementa en un 1% el precio del litro de aceite, la cantidad demandada disminuye en 0.07054 litros.

$b_2 = \partial C / \partial \ln RF \Rightarrow b_2 = \Delta C / (\Delta RF / RF) \rightarrow$  Manteniendo constante el precio del litro de aceite, si se incrementa en un 1% la renta de las familias, la cantidad demandada aumenta en 0.08768 litros.

Dado que la variable dependiente es lineal, el coeficiente de determinación equivalente coincide con el coeficiente de determinación.

```
# Cálculo del coeficiente de determinación equivalente
? series CEM2 = $yhat
Se ha generado la serie CEM2 (ID 14)

? series EEQUIM2 = C - CEM2
Se ha generado la serie EEQUIM2 (ID 15)

? genr SCEQUIM2 = sum(EEQUIM2^2)
Se ha generado el escalar SCEQUIM2 = 0.677511

? genr R2EQUIM2 = 1 - SCEQUIM2 / (sum((C - mean(C))^2))
Se ha generado el escalar R2EQUIM2 = 0.957919
```

- **Modelo Log-Log**

$$\ln C_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln RF_t + \varepsilon_t$$

```
# Modelo 3: Modelo Log-Log
? ols LC const LP LRF --simple-print
```

```
MCO, usando las observaciones 1-10
Variable dependiente: LC
```

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	-3.47164	0.701716	-4.947	0.0017	***
LP	-0.665544	0.0854378	-7.790	0.0001	***
LRF	0.923194	0.0981724	9.404	3.20e-05	***

SCR = 0.00593285, R-cuadrado = 0.963157

$b_1 = \partial \ln C / \partial \ln P \Rightarrow b_1 = (\Delta C / C) / (\Delta P / P) \rightarrow$  Manteniendo constante el nivel de renta de las familias, si se incrementa en un 1% el precio del litro de aceite, la cantidad demandada disminuye en un 0.6655%.

$b_2 = \partial \ln C / \partial \ln RF \Rightarrow b_2 = (\Delta C / C) / (\Delta RF / RF) \rightarrow$  Manteniendo constante el precio del litro de aceite, si se incrementa en un 1% la renta de las familias, la cantidad demandada aumenta en un 0.9232%.

En este caso, los coeficientes estimados miden la elasticidad-precio y la elasticidad-renta de la demanda de aceite de oliva. Como ambas elasticidades son menores que 1 en términos absolutos, se puede concluir que la demanda de aceite de oliva es inelástica tanto respecto al precio como respecto al nivel de renta.

Dado que la variable dependiente no es lineal, el coeficiente de determinación equivalente no coincide con el coeficiente de determinación.

```
# Cálculo del coeficiente de determinación equivalente
? series LCEM3 = $yhat
Se ha generado la serie LCEM3 (ID 16)

? series CEM3 = exp(LCEM3)
Se ha generado la serie CEM3 (ID 17)

? series EEQUIM3 = C - CEM3
Se ha generado la serie EEQUIM3 (ID 18)

? genr SCEQUIM3 = sum(EEQUIM3^2)
Se ha generado el escalar SCEQUIM3 = 0.711406

? genr R2EQUIM3 = 1 - SCEQUIM3 / (sum((C - mean(C))^2))
Se ha generado el escalar R2EQUIM3 = 0.955813
```

#### - Modelo Log-Lin

$$\ln C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \varepsilon_t$$

```
# Modelo4 log-lin
? ols LC const P RF --simple-print
```

MCO, usando las observaciones 1-10  
Variable dependiente: LC

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	2.03348	0.121789	16.70	6.76e-07	***
P	-0.277068	0.0281771	-9.833	2.39e-05	***
RF	0.000953075	8.09834e-05	11.77	7.24e-06	***

SCR = 0.00377653, R-cuadrado = 0.976548

$b_1 = \partial \ln C / \partial P \Rightarrow b_1 = (\Delta C / C) / \Delta P \rightarrow$  Manteniendo constante el nivel de renta de las familias, si se incrementa en 1 euro el precio del litro de aceite, la cantidad demandada disminuye en un 27.71%.

$b_2 = \partial \ln C / \partial RF \Rightarrow b_2 = (\Delta C / C) / \Delta RF \rightarrow$  Manteniendo constante el precio del litro de aceite, si se incrementa en 1 euro la renta de las familias, la cantidad demandada aumenta en un 0.09531%.

Dado que la variable dependiente no es lineal, el coeficiente de determinación equivalente no coincide con el coeficiente de determinación.

```
# Cálculo del coeficiente de determinación equivalente
? series LCEM4 = $yhat
Se ha generado la serie LCEM4 (ID 19)

? series CEM4 = exp(LCEM4)
Se ha generado la serie CEM4 (ID 20)

? series EEQUIM4 = C - CEM4
Se ha generado la serie EEQUIM4 (ID 21)

? genr SCEQUIM4 = sum(EEQUIM4^2)
Se ha generado el escalar SCEQUIM4 = 0.44362

? genr R2EQUIM4 = 1 - SCEQUIM4 / (sum((C - mean(C))^2))
Se ha generado el escalar R2EQUIM4 = 0.972446
```

Cuando en una sesión de trabajo se estiman varios modelos con el menú **Modelo** de la **Ventana Principal** y se guardan los resultados que Gretl muestra en la **Ventana Modelo** como icono, se permite crear lo que Gretl denomina **Tabla de modelos**, donde se muestran los resultados de los principales estadísticos obtenidos en las estimaciones realizadas. Para crear la **Tabla de modelos** se debe acceder a la **Ventana vista de iconos** y situarse con el botón derecho del ratón encima del icono correspondiente al modelo cuyos resultados se deseen añadir a dicha tabla y en el menú emergente, seleccionar **Añadir a la tabla de modelos**.

Estimaciones de MCO		
Variable dependiente: C		
	(Modelo 1)	(Modelo 2)
const	8.287** (1.286)	-44.06** (7.499)
P	-2.934** (0.2975)	
RF	0.009075** (0.0008549)	
LP		-7.054** (0.9130)
LRF		8.768** (1.049)
n	10	10
R-cuadrado	0.9739	0.9579
lnL	1.651	-0.7298

Desviaciones típicas entre paréntesis  
 \* indica significativo al nivel del 10 por ciento  
 \*\* indica significativo al nivel del 5 por ciento

En este caso, dado que la variable dependiente no es la misma en todos los modelos, no se van a poder incluir todos los resultados en una única tabla. Se han tenido que utilizar dos tablas: en la primera se recogen los dos modelos cuya variable dependiente es el consumo y en la segunda los dos modelos restantes, que tienen como variable dependiente el logaritmo neperiano del consumo.

Estimaciones de MCO		
Variable dependiente: LC		
	(Modelo 3)	Modelo 4)
const	-3.472** (0.7017)	2.033** (0.1218)

LP	-0.6655**	
	(0.08544)	
LRF	0.9232**	
	(0.09817)	
P	-0.2771**	
	(0.02818)	
RF	0.0009531**	
	(8.098e-05)	
PI		
RFI		
n	10	10
R-cuadrado	0.9632	0.9765
lnL	22.96	25.22

Desviaciones típicas entre paréntesis  
 \* indica significativo al nivel del 10 por ciento  
 \*\* indica significativo al nivel del 5 por ciento

- ¿Qué forma funcional es la más adecuada?

Tabla 3-9. Comparación bondad de ajuste.		
	$R^2$	$R^2_{equi}$
Modelo Lineal	0.9739	0.9739
Modelo Lin-Log	0.9579	0.9579
Modelo Log-Log	0.9632	0.9558
Modelo Log-Lin	0.9765	0.9725

Para comparar modelos doblemente logarítmicos y semilogarítmicos en la variable dependiente con modelos lineales, no se debe utilizar el coeficiente de determinación, sino el coeficiente de determinación equivalente (véase **Tabla 3-9**).

Al fijarse en el coeficiente de determinación equivalente, se selecciona el modelo lineal, por ser el que presenta un mayor valor para este coeficiente.

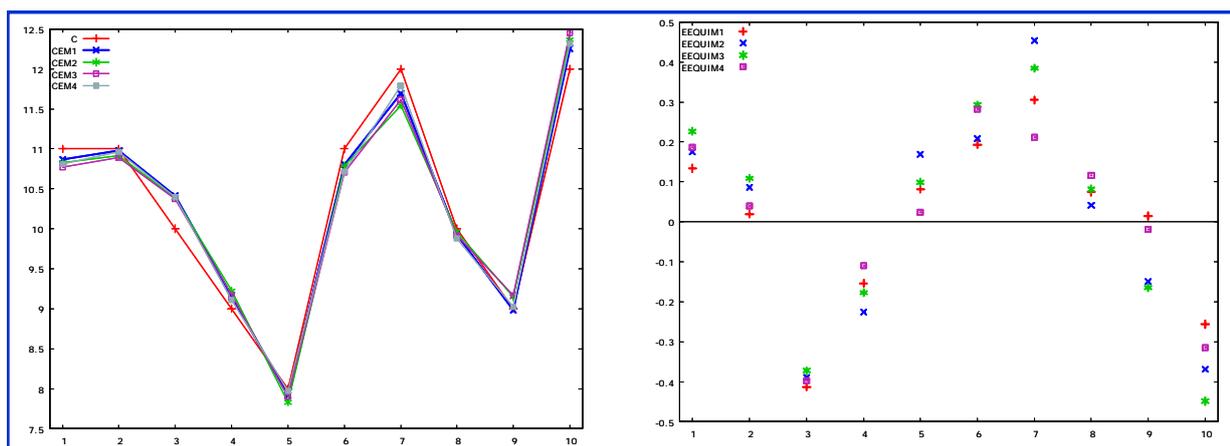


Ilustración 3-11. Gráfico valores estimados del consumo y Gráfico residuos equivalentes.

Los gráficos que aparecen en la Ilustración 3-11 también pueden ayudar a vislumbrar cual es la forma funcional más adecuada, aún cuando, hay que señalar, que en este ejemplo las diferencias son mínimas.

### 3.11.6. Predicción

Para realizar una predicción es necesario que en un mismo fichero se encuentren los datos muestrales (estimación) y los datos extramuestrales (predicción). El fichero aceite.gdt contiene 60 observaciones,

se han utilizado las 10 primeras para la estimación y se utilizarán las observaciones 11 y 12 para efectuar predicciones.

### 3.11.6.1. Predicción a través del comando fcast

El comando **fcast** debe ir precedido necesariamente de un comando **ols**, de lo contrario Gretl indicará mediante un mensaje que se está cometiendo un error.

```
? fcast 11 12 --static
No se puede hacer esto: ningún modelo ha sido aún estimado
Error de datos
```

```
? smpl 1 10
Rango de datos completo: 1 - 60 (n = 60)
Muestra actual: 1 - 10 (n = 10)

# Estimación del modelo
? ols C const P RF --quiet
```

#### - Predicción puntual

```
? smpl 1 12
Rango de datos completo: 1 - 60 (n = 60)
Muestra actual: 1 - 12 (n = 12)

? fcast 11 12 --static --no-stats

Para intervalos de confianza 95%, t(7, .0.025) = 2.365
```

Observaciones	C	predicción	Desv. Típica	Intervalo de confianza 95%
11	13.00	13.00	0.341	12.19 - 13.80
12	12.00	12.10	0.282	11.43 - 12.77

Dado que se ha ejecutado el comando **fcast** con las opciones **--static** y **--no-stats**, se obtienen predicciones estáticas puntuales y se suprime la salida de los estadísticos de análisis de la capacidad predictiva.

#### - Predicción media

```
? fcast 11 12 --mean-y --static --no-stats

Para intervalos de confianza 95%, t(7, .0.025) = 2.365
```

Observaciones	C	predicción	Desv. Típica	Intervalo de confianza 95%
11	13.00	13.00	0.237	12.44 - 13.56
12	12.00	12.10	0.140	11.77 - 12.43

Para obtener predicciones medias puntuales es necesario que el comando **fcast** se ejecute con la opción **--mean-y**.

Obsérvese que la predicción puntual y la predicción media coinciden, lo que varía es el estimador de la desviación típica y, por tanto, los intervalos de confianza, siendo más amplios para la predicción media que para la predicción puntual.

### 3.11.6.2. Predicción utilizando álgebra matricial

#### - Regresando predicho

Bajo las hipótesis del MRLNC y asumiendo estabilidad postmuestral, el predictor óptimo en el sentido de minimizar la varianza del error de predicción, viene dado por:

$$\hat{Y}_\tau = X'_\tau b = b_0 + b_1 X_{1\tau} + b_2 X_{2\tau} + \dots + b_K X_{K\tau} \quad \forall \tau = 1, 2, \dots, N$$

Para realizar una predicción es necesario conocer los valores de los regresores en el período de predicción.

```
# Se define el vector que recoge las observaciones de los regresores para la observación 11
? matrix X11 = {const[11], P[11], RF[11]}
Se ha generado la matriz X11

? print X11
X11 (1 x 3)
      1.0000      2.4600      1314.5

# Se define el vector que recoge las observaciones de los regresores para la observación 12
? matrix X12 = {const[12], P[12], RF[12]}
Se ha generado la matriz X12

? print X12
X12 (1 x 3)
      1.0000      2.3400      1176.9

# Predicción para la observación 11
? genr Cp11 = X11 * b
Se ha generado el escalar Cp11 = 12.9986

# Predicción para la observación 12
? genr Cp12 = X12 * b
Se ha generado el escalar Cp12 = 12.1022
```

**- Error de predicción**

$$e_{\tau} = Y_{\tau} - \hat{Y}_{\tau} \quad \forall \tau = 1, 2, \dots, N$$

```
# Error de predicción para la observación 11
? genr ep11 = C[11] - Cp11
Se ha generado el escalar ep11 = 0.00143375

# Error de predicción para la observación 12
? genr ep12 = C[12] - Cp12
Se ha generado el escalar ep12 = -0.102223
```

**- Varianza estimada del error de la predicción puntual**

$$S_{e_{\tau}}^2 = S^2 [1 + X'_{\tau} (X'X)^{-1} X_{\tau}] \quad \forall \tau = 1, 2, \dots, N$$

```
# Varianza estimada del error de predicción puntual para la observación 11
? genr s2epp11 = s2 * (1 + X11 * inv(X'X) * X11')
Se ha generado el escalar s2epp11 = 0.116251

# Varianza estimada del error de predicción puntual para la observación 12
? genr s2epp12 = s2 * (1 + X12 * inv(X'X) * X12')
Se ha generado el escalar s2epp12 = 0.0796202
```

Al interpretar el resultado, si la varianza del error de predicción es pequeña se puede esperar que los errores de predicción también sean pequeños.

**- Desviación típica estimada del error de la predicción puntual**

$$S_{e_{\tau}} = \sqrt{S_{e_{\tau}}^2} = \sqrt{S^2 [1 + X'_{\tau} (X'X)^{-1} X_{\tau}]} \quad \forall \tau = 1, 2, \dots, N$$

```
# Desviación típica estimada del error de predicción puntual para la observación 11
? genr sepp11 = sqrt(s2epp11)
Se ha generado el escalar sepp11 = 0.340956

# Desviación típica estimada del error de predicción puntual para la observación 12
? genr sepp12 = sqrt(s2epp12)
Se ha generado el escalar sepp12 = 0.28217
```

**- Varianza estimada del error de la predicción media**

$$S_{e_{\tau}}^2 = S^2 [X'_{\tau} (X'X)^{-1} X_{\tau}] \quad \forall \tau = 1, 2, \dots, N$$

```
# Varianza estimada del error de predicción media para la observación 11
? genr s2epm11 = s2 * (X11 * inv(X'X) * X11')
Se ha generado el escalar s2epm11 = 0.0561274
```

```
# Varianza estimada del error de predicción media para la observación 12
? genr s2epm12 = s2 * (X12 * inv(X'X) * X12')
Se ha generado el escalar s2epm12 = 0.0194963
```

**- Desviación típica estimada del error de la predicción media**

$$S_{e_\tau} = \sqrt{S_{e_\tau}^2} = \sqrt{S^2[X'_\tau(X'X)^{-1}X_\tau]} \quad \forall \tau = 1, 2, \dots, N$$

```
# Desviación típica estimada del error de predicción media para la observación 11
? genr sepm11 = sqrt(s2epm11)
Se ha generado el escalar sepm11 = 0.236912
```

```
# Desviación típica estimada del error de predicción media para la observación 12
? genr sepm12 = sqrt(s2epm12)
Se ha generado el escalar sepm12 = 0.139629
```

**- Intervalo de confianza para la predicción puntual**

$$\hat{Y}_\tau \pm t_{\alpha/2}^{T-K-1} S_{e_\tau} \quad \forall \tau = 1, 2, \dots, N$$

```
# Valor crítico para una distribución t de Student de 7 grados de libertad (T-k-1) y un nivel de
significación del 5%
```

```
? genr talpham = critical(t,7,0.025)
Se ha generado el escalar talpham = 2.36462
```

```
# Intervalo de confianza para un nivel de significación del 5% para la observación 11
```

```
? genr LIIPP11 = Cp11 - talpham * sepp11
Se ha generado el escalar LIIPP11 = 12.1923
```

```
? genr LSIPP11 = Cp11 + talpham * sepp11
Se ha generado el escalar LSIPP11 = 13.8048
```

```
# Intervalo de confianza para un nivel de significación del 5% para la observación 12
```

```
? genr LIIPP12 = Cp12 - talpham * sepp12
Se ha generado el escalar LIIPP12 = 11.435
```

```
? genr LSIPP12 = Cp12 + talpham * sepp12
Se ha generado el escalar LSIPP12 = 12.7694
```

**- Intervalo de confianza para la predicción media**

Dado que el predictor del valor esperado coincide con el predictor puntual:

$$\hat{Y}_\tau = E\hat{Y}_\tau \quad \forall \tau = 1, 2, \dots, N$$

El intervalo de confianza para el predictor del valor esperado se puede calcular como:

$$\hat{Y}_\tau \pm t_{\alpha/2}^{T-K-1} S_{e_\tau} \quad \forall \tau = 1, 2, \dots, N$$

```
# Intervalo de confianza para un nivel de significación del 5% para la observación 11
```

```
? genr LIIPM11 = Cp11 - talpham * sepm11
Se ha generado el escalar LIIPM11 = 12.4384
```

```
? genr LSIPM11 = Cp11 + talpham * sepm11
Se ha generado el escalar LSIPM11 = 13.5588
```

```
# Intervalo de confianza para un nivel de significación del 5% para la observación 12
```

```
? genr LIIPM12 = Cp12 - talpham * sepm12
Se ha generado el escalar LIIPM12 = 11.7721
```

```
? genr LSIPM12 = Cp12 + talpham * sepm12
Se ha generado el escalar LSIPM12 = 12.4324
```

# Capítulo 4. CONTRASTES DE HIPÓTESIS Y REGIONES DE CONFIANZA

## 4.1. Hipótesis del Modelo de Regresión Lineal Normal Clásico (MRLNC)

Las hipótesis del MRLC introducidas en el capítulo 3 son suficientes para obtener estimadores puntuales de los parámetros del modelo, pero si se quiere abordar la estimación por intervalo y/o los contrastes de hipótesis, es necesario hacer referencia a la distribución de probabilidad de las perturbaciones.

### H5. Hipótesis de normalidad de las perturbaciones:

$$\varepsilon \xrightarrow{sd} N(0_{Tx1}, \sigma^2 I_T) \Rightarrow \varepsilon_t \xrightarrow{sd} N(0, \sigma^2)$$

Esta hipótesis establece que el vector de perturbaciones sigue una distribución normal multivariante (el valor más probable es el valor esperado y la probabilidad disminuirá de forma sistemática a medida que nos alejemos del valor promedio).

Como los elementos del vector de perturbaciones no están correlacionados y como para variables normales, la incorrelación implica independencia, las perturbaciones aleatorias serán variables independientes entre sí.

El fundamento para la introducción de esta hipótesis se basa en el **Teorema Central del Límite**, según el cual, bajo condiciones muy generales, la suma de variables independientes tiende hacia la ley normal. La perturbación aleatoria se introduce en los modelos econométricos para incorporar el efecto de una serie de factores, que considerados de manera individual tienen escasa incidencia sobre el regresando, pero que considerados conjuntamente ejercen una influencia estadísticamente significativa para la explicación del regresando, es lógico suponer que dichos factores son variables aleatorias independientes, cuya suma sigue una distribución normal.

Por tanto, un **Modelo de Regresión Lineal Normal Clásico (MRLNC)** es un modelo econométrico que además de satisfacer las hipótesis de un MRLC, cumple la hipótesis de normalidad de las perturbaciones.

## 4.2. Contrastes de hipótesis

En este capítulo se analizarán los contrastes de hipótesis y la construcción de regiones de confianza para los parámetros de un modelo estimado por Mínimos Cuadrados Ordinarios.

Para hacer contrastes de hipótesis se seguirán los siguientes pasos:

1. **Formular** claramente la **hipótesis** que se quiere **contrastar**.
2. **Construir** el **estadístico de prueba** que permita decidir, si se acepta o rechaza la hipótesis formulada.
3. **Analizar** en cada caso cual es la **regla de decisión**.

En todo contraste se tendrán dos **hipótesis**, la hipótesis a contrastar, denominada **hipótesis nula** ( $H_0$ ) y otra hipótesis, denominada **hipótesis alternativa** ( $H_1$ ).

Cuando se efectúan contrastes de hipótesis es posible cometer dos tipos de **errores**: rechazar la hipótesis nula siendo cierta (**error tipo I**) o aceptar la hipótesis nula siendo falsa (**error tipo II**). Sería deseable minimizar ambos tipos de errores, pero desafortunadamente, dado un tamaño muestral, no es posible minimizar ambos simultáneamente.

La econometría clásica considera más grave el error tipo I, por lo que fija la probabilidad de cometer un error de este tipo a un nivel relativamente bajo (1%, 5% o como máximo 10%) y selecciona estadísticos de prueba que minimizan la probabilidad de cometer un error tipo II. Al límite que acota la

probabilidad de cometer un error tipo I se le denomina **nivel de significación** y a la probabilidad de no cometer un error tipo II se le denomina **potencia del test**.

El problema relacionado con la selección del valor apropiado del nivel de significación ( $\alpha$ ) se puede evitar utilizando el **p-valor** del estadístico de prueba, que se define como el nivel de significación más bajo al cual puede rechazarse la hipótesis nula. Por tanto, si el contraste es de la cola derecha, el p-valor es la probabilidad real de obtener un valor del estadístico de prueba igual o mayor que el obtenido y si el contraste es de la cola izquierda, el p-valor es la probabilidad real de obtener un valor del estadístico de prueba menor o igual que el obtenido.

Un procedimiento general para efectuar contrastes para cualquier conjunto de hipótesis lineales es plantearlo de la siguiente forma:

$$H_0: R_{q \times (K+1)} \beta_{(K+1) \times 1} = r_{q \times 1}$$

$$H_1: R_{q \times (K+1)} \beta_{(K+1) \times 1} \neq r_{q \times 1}$$

donde R y r son conocidos, R es una matriz cuyo número de filas coincide con el número de restricciones lineales impuestas a los parámetros (q) y su número de columnas coincide con el número de regresores del modelo econométrico (k+1)<sup>30</sup> y r es un vector columna cuyo número de filas coincide con el número de restricciones<sup>31</sup>.

$$\text{Si } H_0 \text{ es cierta } \Rightarrow F = \frac{\frac{(Rb - r)' [R(X'X)^{-1}R']^{-1} (Rb - r)}{q}}{\frac{SCE}{T - K - 1}} \xrightarrow{sd} \frac{\sigma^2 \chi_q^2}{q} \xrightarrow{sd} F_{T-K-1}^q$$

Para aceptar o rechazar la hipótesis nula, existen tres alternativas:

1. Utilizar el p-valor, que indica el menor nivel de significación al cuál se rechaza la hipótesis nula.
2. Comparar el p-valor con un nivel de significación prefijado y utilizar como criterio de decisión:
  - Si p-valor  $\geq \alpha \Rightarrow$  se acepta  $H_0$
  - Si p-valor  $< \alpha \Rightarrow$  se rechaza  $H_0$
3. Comparar el valor del estadístico obtenido con el valor crítico de la distribución correspondiente para el nivel de significación fijado y utilizar la regla de decisión adecuada en cada caso, dependiendo de si se utiliza la cola izquierda, la cola derecha o ambas colas para rechazar  $H_0$ .

Si la muestra es lo suficientemente grande, también, se podría utilizar el estadístico de Wald para realizar este tipo de contrastes. La distribución del estadístico de Wald se basa en la distribución de la forma cuadrática que interviene en el numerador del estadístico F:

$$W = \frac{(Rb - r)' [R(X'X)^{-1}R']^{-1} (Rb - r)}{S^2} \xrightarrow{sd} \chi_q^2$$

Hay que tener en cuenta que la distribución de este estadístico como una chi-cuadrado con q grados de libertad es asintótica (para muestras lo suficientemente grandes), dado que en su calculo no se utiliza la varianza de la perturbación sino su estimador ( $S^2$ ).

<sup>30</sup> Se asume que el modelo está formulado con ordenada en el origen, en caso contrario, el número de regresores sería K.

<sup>31</sup> Debe de tenerse en cuenta que el número de restricciones no puede ser superior al número de regresores del modelo.

La "equivalencia" entre los estadísticos F y W vendría dada por:

$$F = \frac{(Rb - r)'[R(X'X)^{-1}R']^{-1}(Rb - r)}{q S^2} = \frac{W}{q} \Rightarrow W = q F$$

Cuando en el contraste interviene una única restricción, sería posible y aconsejable, utilizar la distribución t de Student, sobre todo porque permite construir intervalos de confianza. Debe recordarse que la equivalencia entre estos estadísticos es la siguiente:

$$F = t^2$$

### 4.3. ¿Cómo realizar contrastes de hipótesis en Gretl?

Para realizar contrastes de hipótesis lineales en Gretl básicamente se pueden utilizar dos opciones<sup>32</sup>:

- Utilizar la opción **Restricciones lineales** del menú **Contrastes** de la **Ventana Modelo**.
- Utilizar el comando **restrict**.

#### 4.3.1. Utilizar el menú Contrastes

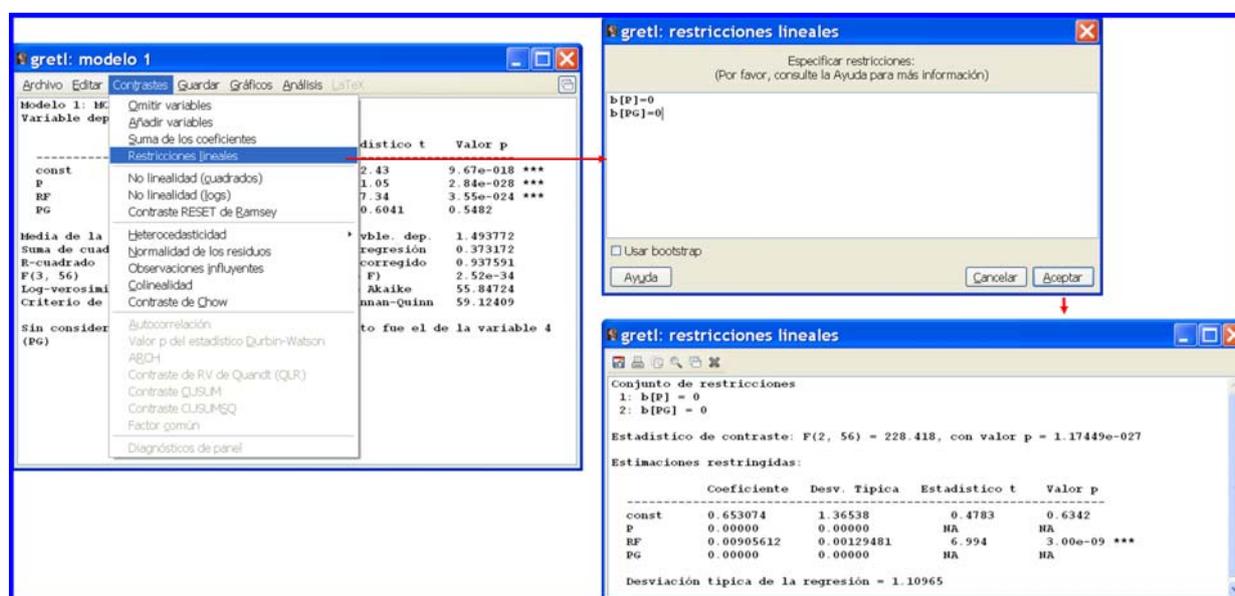


Ilustración 4-1. Contrastes de hipótesis lineales utilizando el menú Contrastes.

Con la opción **Restricciones lineales** del menú **Contrastes** de la **Ventana Modelo** se accede a la ventana **restricciones lineales**, donde se tiene que especificar la ecuación o ecuaciones de la hipótesis que se quiere contrastar. Para especificar de forma correcta las restricciones que intervienen en el contraste, el usuario debe:

- Especificar una ecuación para cada restricción.
- Utilizar una línea para cada ecuación (restricción).

<sup>32</sup> Las otras tres alternativas [**Omitir variables** (**omit**), **Añadir variables** (**add**) y **Suma de coeficientes** (**coeffsum**)] que proporciona Gretl con el menú **Contrastes**, son muy restrictivas y perfectamente abordables a través de la opción **Restricciones lineales**, sin más que identificar el modelo en el que se debe realizar el contraste y escribir adecuadamente las ecuaciones de las restricciones. En el ejercicio de aplicación que aparece al final de este capítulo, se hace una breve referencia a como abordar ciertos contrastes de nulidad de los parámetros a través de dichos comandos.

- Identificar cada parámetro como **b**[*nombre de la variable a la que acompaña*].
- Reformular las restricciones de forma que en lado derecho de las ecuaciones aparezcan sólo valores numericos.

Una vez escritas todas las líneas del contraste, seleccionando el botón "**Aceptar**" de la ventana **restricciones lineales** aparecerá el resultado del contraste y una estimación del modelo restringido (véase parte inferior derecha de la Ilustración 4-1). Se podría acceder a las instrucciones de este contraste con la opción **Historial de instrucciones** del menú **Herramientas** de la **Ventana Principal** y guardarlas en un archivo de comandos. En la ventana de dicho archivo, pinchando el botón "**Ejecutar**" se podrá ver y analizar su salida en la **Ventana resultados guía** correspondiente.

En la **Tabla 4-1** aparece recogida la información que proporciona la ejecución de un contraste de restricciones lineales.

<b>Tabla 4-1. Salida asociada al submenú Restricciones lineales del menú Contrastes de la Ventana Modelo.</b>				
<b>Conjunto de restricciones</b>				
1: ecuación 1				
2: ecuación 2				
...				
q: ecuación q				
<b>Estadístico de contraste: <math>F(q, T-k-1) = \dots</math> con valor <math>p = \dots</math></b>				
<b>Estimaciones restringidas:</b>				
	Coeficiente	Desv. Típica	Estadístico t	Valor p
Const	...	...	...	...
$X_1$	...	...	...	...
$X_2$	...	...	...	...
...	...	...	...	...
$X_k$	...	...	...	...
Desviación típica de la regresión = ...				

Dicha salida se puede estructurar en tres bloques:

1. Conjunto de restricciones del contraste.
2. Estadístico F y p-valor asociado.
3. Estimación del modelo restringido.

En el caso de que en el contraste intervenga una única restricción, si se desea utilizar la distribución t de Student en lugar de la F de Snedecor para realizar el contraste, debe recordarse que la equivalencia entre ambos estadísticos es la siguiente:

$$t = \sqrt{F}$$

Y si se desea utilizar la distribución chi-cuadrado [estadístico de Wald (**W**)], la equivalencia es la siguiente:

$$W = qF$$

Debe de tenerse en cuenta que aunque las versiones **t** y **F** son válidas para muestras pequeñas, la versión **W** sólo es válida si la muestra es lo suficientemente grande.

### 4.3.2. Utilizando el comando **restrict**

Otra opción para realizar contrastes de hipótesis es escribir directamente el comando **restrict** en la **Consola Gretl** o en un **Fichero de comandos**, para lo que es necesario que el usuario conozca el formato de escritura y las opciones disponibles.

El formato del comando **restrict** es:

```
restrict
ecuación 1
ecuación 2
...
ecuación q
end restrict
```

donde “q” es el número de restricciones que intervienen en el contraste. Debe tenerse en cuenta que Gretl identifica a los parámetros como  $b[\text{nombre de la variable a la que acompaña}]^{33}$ .

El formato de escritura de este comando requiere que en la **Consola Gretl** cada restricción ocupe una línea. Además, requiere una primera línea donde aparezca un comando **restrict** y una última línea donde aparezca un comando **end restrict**, para indicarle al programa el principio y final de las restricciones que forman parte del contraste.

Para que Gretl ejecute un comando **restrict** es necesario que dicho comando vaya precedido por un comando **ols**. Aún cuando no es necesario que el comando **restrict** se ejecute inmediatamente después del comando **ols**, resulta conveniente, ya que de esta forma el usuario se asegura que el contraste hace referencia al modelo deseado. Si no se está interesado en ver la salida del comando **ols**, dicho comando se ejecutará con la opción **--quiet** (ejecuta el comando omitiendo su salida). De manera análoga, si no se está interesado en ver la estimación del modelo restringido, se ejecutará el comando **restrict** con la opción **--quiet**.

## 4.4. ¿Cómo buscar el p-valor y el valor crítico en una distribución?

A la hora de efectuar contrastes de hipótesis resulta imprescindible conocer el valor crítico asociado a un nivel de significación prefijado o el p-valor asociado al valor del estadístico de prueba utilizado en dicho contraste. Ambos valores se buscarán en la distribución de probabilidad que siga el estadístico de prueba de dicho contraste bajo la hipótesis nula.

### 4.4.1. Valor crítico

Para buscar el valor crítico se puede utilizar el menú **Herramientas** de la **Ventana Principal**.

<sup>33</sup> Otra alternativa, aunque no aconsejable, es sustituir el nombre de la variable por el lugar que ocupa en el vector de estimadores de los parámetros (**b**) el estimador del parámetro que acompaña a dicha variable.

Al seleccionar la opción **Tablas estadísticas** en dicho menú, se abre el cuadro de diálogo **valores críticos**, donde pinchando en el botón correspondiente (normal, t,  $\chi^2$ , F) se debe elegir el tipo de distribución y proporcionar la información que Gretl necesita en cada caso. Como puede verse en los ejemplos de la Ilustración 4-2, son varias las casillas que se deben ir rellenando, variando su número y contenido de unas distribuciones a otras. En la casilla **probabilidad en la cola derecha** se debe especificar el nivel de significación ( $\alpha$ ) para los contrastes de la cola derecha y el nivel de confianza ( $1 - \alpha$ ) para los contrastes de la cola izquierda. En las distribuciones simétricas y para el contraste de ambas colas, se debe especificar la mitad del nivel de significación ( $\alpha/2$ ). En la distribución normal se debe especificar la media y desviación típica<sup>34</sup> y, en las distribuciones t,  $\chi^2$  y F se deben especificar sus grados de libertad.

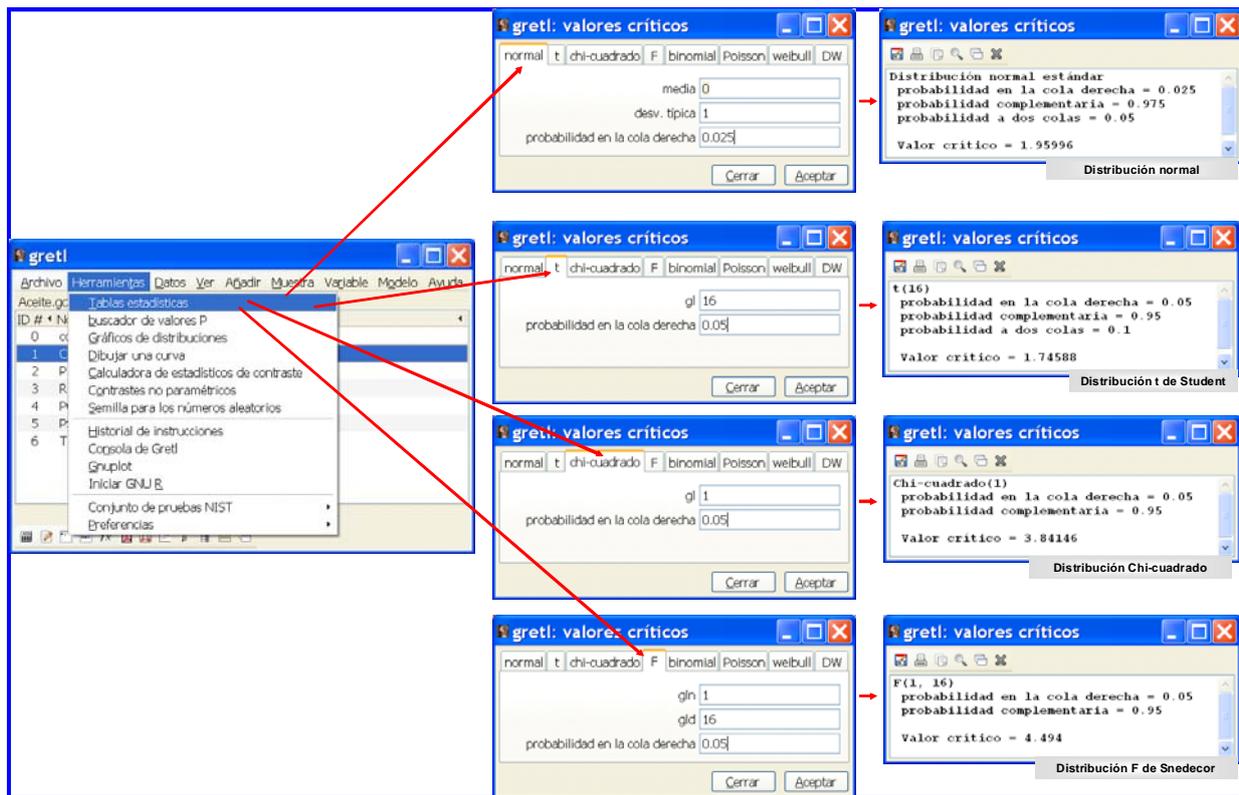


Ilustración 4-2. Búsqueda del valor crítico.

Una vez que se han rellenado todas las casillas del cuadro de diálogo **valores críticos**, se selecciona el botón "Aceptar" y aparece el resultado de la búsqueda en la ventana **valores críticos**, donde se informa (ver Ilustración 4-2):

1. Del tipo y características de la distribución.
2. De la probabilidad en la cola derecha y de la probabilidad complementaria. Además, en las distribuciones simétricas, informa de la distribución en ambas colas.
3. Del valor crítico.

<sup>34</sup> Por defecto, Gretl tiene seleccionadas las características de la distribución normal reducida, es decir, media nula y desviación típica unitaria.

Se puede utilizar la función **critical** del comando **genr** para guardar el valor crítico y tenerlo disponible para cálculos posteriores:

**genr nueva variable = critical(tipo de distribución, parámetros de la distribución, nivel de significación)**

El usuario debe tener en cuenta que dicha función hace referencia al área de la cola derecha.

En distribuciones simétricas:

- Para guardar el valor crítico en ambas colas:

**genr nueva variable = critical(tipo de distribución, parámetros de la distribución, [nivel de significación/2])**

y el valor crítico de la cola izquierda:

**genr nueva variable = - critical(tipo de distribución, parámetros de la distribución, nivel de significación)**

Algunos ejemplos:

```
# Valor critico, para un nivel de significación del 5%, de una distribución chi-cuadrado de 1 grados
de libertad
? genr vcchi = critical(X,1,0.05)
Se ha generado el escalar vcchi = 3.84146

# Valor critico, para un nivel de significación del 5%, de una distribución F de Snedecor de 1
grados de libertad en el numerador y 16 grados de libertad en el denominador
? genr vcF = critical(F,1,16,0.05)
Se ha generado el escalar vcF = 4.494

# Valor critico, para un nivel de significación del 5% en la cola derecha, de una distribución
normal reducida
? genr vcxcd = critical(z,0.05)
Se ha generado el escalar vcxcd = 1.64485

# Valor critico, para un nivel de significación del 5% en la cola izquierda, de una distribución
normal reducida
? genr vcxcd = - critical(z,0.05)
Se ha reemplazado el escalar vcxcd = -1.64485

# Valor critico, para un nivel de significación del 5% en ambas colas, de una distribución normal
reducida
? genr vc2c = critical(z,0.025)
Se ha generado el escalar vc2c = 1.95996

# Valor critico, para un nivel de significación del 5% en la cola derecha, de una distribución t de
Student de 16 grados de libertad
? genr vct = critical(t,16,0.05)
Se ha generado el escalar vct = 1.74588

# Valor critico, para un nivel de significación del 5% en la cola izquierda, de una distribución t
de Student de 16 grados de libertad
? genr vctci = -critical(t,16,0.05)
Se ha generado el escalar vctci = -1.74588

# Valor critico, para un nivel de significación del 5% en ambas colas, de una distribución t de
Student de 16 grados de libertad
? genr vct2c = critical(t,16,0.025)
Se ha generado el escalar vct2c = 2.11991
```

#### 4.4.2. P-valor

Para buscar el p-valor se puede utilizar el menú **Herramientas** de la **Ventana Principal** o el comando **pvalue**.

Al seleccionar la opción **buscador de valores-p** del menú **Herramientas** de la **Ventana Principal** se abre el cuadro de diálogo **buscador de valor p**, donde se debe elegir el tipo de distribución pinchando en el botón correspondiente (normal, t,  $\chi^2$ , F) y proporcionar la información que Gretl necesita en cada caso. Tal y como puede verse en los ejemplos recogidos en la Ilustración 4-3, son varias las casillas que se deben ir rellenando, variando su número y contenido de unas distribuciones a otras. En todos los casos, en la casilla **valor**, se debe especificar el valor del estadístico cuyo p-valor se está buscando. En la distribución normal se debe especificar la media y desviación típica<sup>35</sup>. En las distribuciones t,  $\chi^2$  y F se deben especificar sus grados de libertad.

Una vez que se han rellenado todas las casillas del cuadro de diálogo **buscador de valor p**, se selecciona el botón "**Aceptar**" y aparece el resultado de la búsqueda en la ventana **valor p** donde se informa del tipo y características de la distribución (ver Ilustración 4-3):

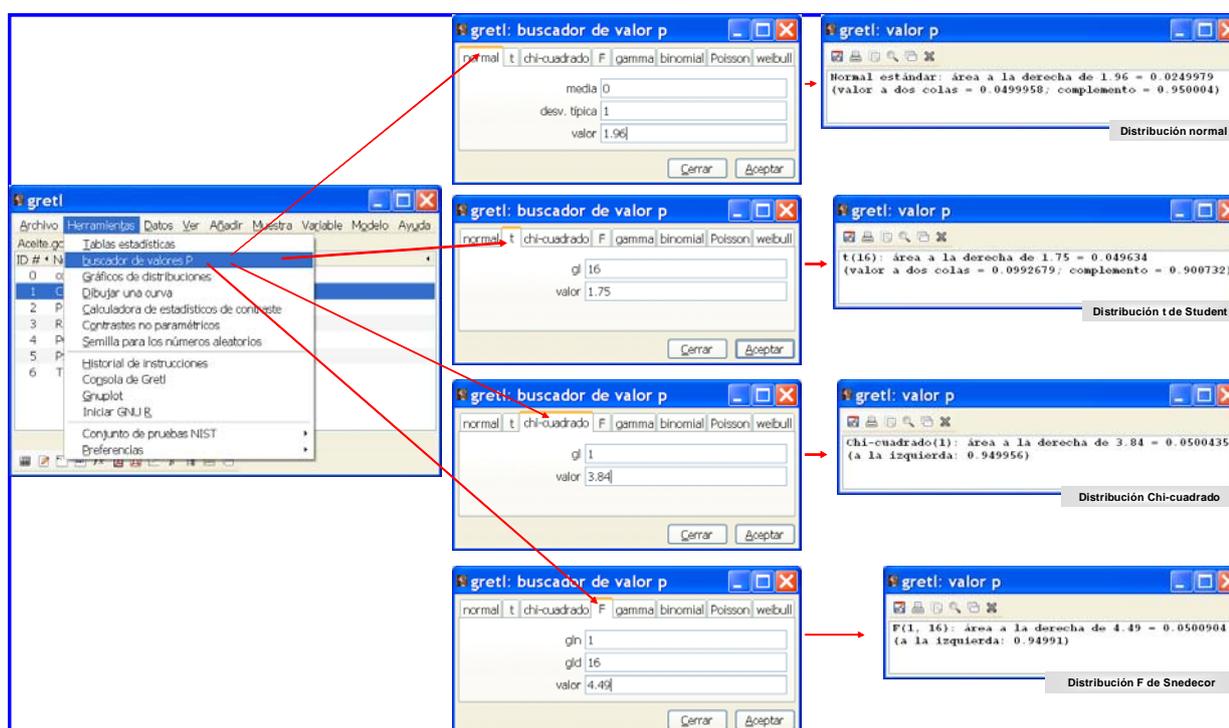


Ilustración 4-3. Búsqueda del p-valor.

1. En las distribuciones  $\chi^2$  y F de Snedecor, del área situada a la derecha del valor del estadístico y del área situada a su izquierda.
2. En las distribuciones normal y t de Student (distribuciones simétricas), del área en la cola derecha, en ambas colas y el complemento (área situada entre ambas colas).

Nótese que la suma de todas estas áreas es siempre la unidad.

La otra opción para obtener los p-valores es escribir directamente el comando **pvalue** en la **Consola Gretl** o en un **Archivo de comandos**, para lo que es necesario que el usuario conozca su formato de escritura y las opciones disponibles.

El formato del comando **pvalue** es:

<sup>35</sup> Por defecto, Gretl tiene seleccionadas las características de la distribución normal reducida, es decir, media nula y desviación típica unitaria.

***pvalue*** tipo de distribución parámetros de la distribución valor del estadístico

En dicho comando se debe especificar:

1. El tipo de distribución, Gretl identifica la distribución t de Student por **t**, la  $\chi^2$  por **X** y la F de Snedecor por **F**.
2. Los parámetros o grados de libertad de la distribución. Se debe tener en cuenta que en el caso de la distribución **F** será necesario especificar en primer lugar los grados de libertad del numerador y en segundo lugar los grados de libertad del denominador.
3. El valor del estadístico.

Gretl identifica la distribución normal reducida por **z** y en el comando **pvalue** no será necesario especificar ningún parámetro para identificar dicha distribución.

Además, se puede utilizar la función **pvalue** del comando **genr** para guardar el p-valor y tenerlo disponible para cálculos posteriores:

**genr** nueva variable = **pvalue**(tipo de distribución, parámetros de la distribución, valor del estadístico)

El usuario debe tener en cuenta que dicha función hace referencia al área de la cola derecha, por tanto, para guardar el p-valor para la cola izquierda se debe ejecutar el comando:

**genr** nueva variable = 1- **pvalue**(tipo de distribución, parámetros de la distribución, valor del estadístico)

En distribuciones simétricas, para guardar el p-valor en ambas colas:

**genr** nueva variable = 2\* **pvalue**(tipo de distribución, parámetros de la distribución, valor del estadístico)

Algunos ejemplos :

```
# P-valor, para un valor de 3.84, de una distribución chi-cuadrado de 1 grado de libertad
? pvalue X 1 3.84
Chi-cuadrado(1): área a la derecha de 3.84 = 0.0500435
(a la izquierda: 0.949956)

? genr pvchi = pvalue(X,1,3.84)
Se ha generado el escalar pvchi = 0.0500435

# P-valor, para un valor de 4.49, de una distribución F de Snedecor de 1 grado de libertad en el
numerador y 16 grados de libertad en el denominador
? pvalue F 1 16 4.49
F(1, 16): área a la derecha de 4.49 = 0.0500904
(a la izquierda: 0.94991)

? genr pvF = pvalue(F,1,16,4.49)
Se ha generado el escalar pvF = 0.0500904

# P-valor, para un valor de 1.64 en la cola derecha, de una distribución normal reducida
? pvalue z 1.64
Normal estándar: área a la derecha de 1.64 = 0.0505026
(valor a dos colas = 0.101005; complemento = 0.898995)

? genr pvzcd = pvalue(z,1.64)
Se ha generado el escalar pvzcd = 0.0505026

# P-valor, para un valor de 1.64 en la cola izquierda, de una distribución normal reducida
? pvalue z -1.64
Normal estándar: área a la derecha de -1.64 = 0.949497
(a la izquierda: 0.0505026)
(valor a dos colas = 0.101005; complemento = 0.898995)

? genr pvzci = 1-pvalue(z,-1.64)
Se ha generado el escalar pvzci = 0.0505026

# P-valor, para un valor de 1.96 en ambas colas, de una distribución normal reducida
? pvalue z 1.96
```

```

Normal estándar: área a la derecha de 1.96 = 0.0249979
(valor a dos colas = 0.0499958; complemento = 0.950004)

? genr pvz2c = 2*pvalue(z,1.96)
Se ha generado el escalar pvz2c = 0.0499958

? genr pvz2c = 2*(1-pvalue(z,-1.96))
Se ha reemplazado el escalar pvz2c = 0.0499958

# P-valor, para un valor de 1.74 en la cola derecha, de una distribución t de Student de 1 grado de
libertad en el numerador y 16 grados de libertad en el denominador
? pvalue t 16 1.74
t(16): área a la derecha de 1.74 = 0.0505273
(valor a dos colas = 0.101055; complemento = 0.898945)

? genr pvt = pvalue(t,16,1.74)
Se ha generado el escalar pvt = 0.0505273

# P-valor, para un valor de 1.74 en la cola izquierda, de una distribución t de Student de 1 grado
de libertad en el numerador y 16 grados de libertad en el denominador
? pvalue t 16 -1.74
t(16): área a la derecha de -1.74 = 0.949473
(a la izquierda: 0.0505273)
(valor a dos colas = 0.101055; complemento = 0.898945)

? genr pvtci = 1-pvalue(t,16,-1.74)
Se ha generado el escalar pvtci = 0.0505273

# P-valor, para un valor de 2.11 en la cola derecha, de una distribución t de Student de 1 grado de
libertad en el numerador y 16 grados de libertad en el denominador
? genr pvt2c = 2*pvalue(t,16,2.11)
Se ha generado el escalar pvt2c = 0.0509572

# P-valor, para un valor de 2.11 en la cola derecha, de una distribución t de Student de 1 grado de
libertad en el numerador y 16 grados de libertad en el denominador
? pvalue t 16 2.11
t(16): área a la derecha de 2.11 = 0.0254786
(valor a dos colas = 0.0509572; complemento = 0.949043)

? genr pvt2c = 2*pvalue(t,16,2.11)
Se ha generado el escalar pvt2c = 0.0509572

```

## 4.5. Regiones de Confianza

Para un subconjunto de  $(K-h)$  parámetros  $(\beta_{sub})$ , la región de confianza viene dada por todos aquellos puntos que verifiquen:

$$P\left((b_{sub} - \beta_{sub})' [(X'X)_{sub}^{-1}]^{-1} (b_{sub} - \beta_{sub}) \leq F_{\alpha}[K - h]S^2\right) = 1 - \alpha$$

En el caso particular de que el subconjunto paramétrico esté formado por un solo parámetro, a la región de confianza se le denomina intervalo de confianza y estará formado por todos aquellos puntos que pertenezcan al intervalo de la recta real:

$$\beta_i \in (b_i \pm t_{\alpha/2}^{T-K-1} S_{b_i})$$

La construcción de regiones de confianza es una forma alternativa de efectuar contrastes de hipótesis. Si el valor establecido en la hipótesis nula para el subconjunto paramétrico pertenece a la región de confianza se aceptará la hipótesis nula y en caso contrario se rechazará.

### 4.5.1. ¿Cómo calcular intervalos y regiones de confianza en Gretl?

A través de la opción **Intervalos de confianza para los coeficientes** del menú **Análisis** de la **Ventana Modelo** se accede a la ventana **intervalos de confianza para los coeficientes** en la que Gretl

proporciona para un nivel de significación del 5%, el valor crítico de una distribución  $t$  de Student cuyos grados de libertad coinciden con los del modelo<sup>36</sup>, los estimadores de los parámetros y los intervalos de confianza para todos los parámetros que intervienen en el modelo.

Con la opción **Elipse de confianza** del menú **Análisis** de la **Ventana Modelo** se accede a la ventana **contrastes de hipótesis** donde, después de elegir dos variables<sup>37</sup> y el nivel de confianza, se accede a la **Ventana gráfico**, donde se muestra la elipse de confianza y los intervalos marginales de confianza para los parámetros seleccionados. Aunque por defecto, Gretl utiliza un nivel de confianza del 95% para construir la región de confianza, existe la posibilidad de cambiar dicho nivel en la construcción de las regiones de confianza conjunta y marginales.

#### 4.5.2. ¿Cómo utilizar el intervalo de confianza en el contraste de hipótesis?

La probabilidad de construir un intervalo de confianza que contenga el verdadero valor del parámetro es igual al nivel de confianza  $(1-\alpha)$ . Por tanto, construir un intervalo de confianza del  $(1-\alpha)\%$  para el parámetro genérico  $\beta_i$ , significa que si se seleccionan 100 muestras de tamaño  $T$  y se construyen 100 intervalos de confianza, se espera que  $(1-\alpha)$  de ellos contengan el verdadero valor del parámetro.

Los intervalos de confianza se pueden utilizar en los contrastes bilaterales. De manera que si el valor que se establece para el parámetro en la hipótesis nula, pertenece al intervalo de confianza se aceptará dicha hipótesis para un nivel de significación  $\alpha$  y, en caso contrario, se rechazará dicha hipótesis.

Aceptar una hipótesis no significa necesariamente que sea cierta, puede ocurrir que se tenga gran incertidumbre sobre el verdadero valor del parámetro y, en dicha situación, la aceptación tan sólo significa que con la información de la que se dispone no hay elementos de juicio suficientes para rechazar dicha hipótesis.

### 4.6. Ejercicio de aplicación

En este ejercicio se analizarán los contrastes de hipótesis y la construcción de intervalos y regiones de confianza para los parámetros de un modelo estimado por Mínimos Cuadrados Ordinarios. Para ello se plantea el siguiente ejemplo:

$$C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t \quad \forall t = 1, 2, \dots, 60$$

Se utilizarán los datos del fichero `aceite.gdt`: consumo familiar mensual medio de aceite de oliva en litros (C), precio medio de compra del aceite de oliva en euros/litro (P), ingresos familiares mensuales medios en euros (RF) y precio medio de compra del aceite de girasol en euros/litro (PG).

```
? open "C:\Proyecto01\Aceite.gdt"
Leer fichero de datos C:\Proyecto01\Aceite.gdt
periodicidad: 1, máx.obs: 60
rango de observaciones: 1-60

Listando 7 variables:
0) const    1) C        2) P        3) RF       4) PG
5) PS      6) TF
```

#### 4.6.1. Estimación MCO del modelo

- Estimación del modelo utilizando el comando **ols**:

```
? ols C const P RF PG --anova
```

```
Modelo 1: MCO, usando las observaciones 1-60
```

<sup>36</sup> (T-K-1) para modelos formulados con ordenada en el origen y (T-K) para modelos formulados sin ordenada en el origen.

<sup>37</sup> Las variables disponibles son las incluidas como variables independientes en el modelo.

Variable dependiente: C

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	11.9116	0.958136	12.43	9.67e-018 ***
P	-3.85611	0.183205	-21.05	2.84e-028 ***
RF	0.00788497	0.000454796	17.34	3.55e-024 ***
PG	-0.305693	0.506005	-0.6041	0.5482

Media de la vble. dep.	10.15000	D.T. de la vble. dep.	1.493772
Suma de cuad. residuos	7.798423	D.T. de la regresión	0.373172
R-cuadrado	0.940764	R-cuadrado corregido	0.937591
F(3, 56)	296.4569	Valor p (de F)	2.52e-34
Log-verosimilitud	-23.92362	Criterio de Akaike	55.84724
Criterio de Schwarz	64.22462	Crit. de Hannan-Quinn	59.12409

Sin considerar la constante, el valor p más alto fue el de la variable 4 (PG)

Análisis de Varianza:

	Suma de cuadrados	gl	Media de cuadrados
Regresión	123.852	3	41.2839
Residuo	7.79842	56	0.139258
Total	131.65	59	2.23136

R<sup>2</sup> = 123.852 / 131.65 = 0.940764  
 F(3, 56) = 41.2839 / 0.139258 = 296.457 [Valor p 2.52e-034]

- Obtención del vector de estimadores y de su matriz de varianzas-covarianzas estimada utilizando álgebra matricial:

```
# Cálculo del vector de estimadores
? matrix X = {const, P, RF, PG}
Se ha generado la matriz X

? matrix Y = {C}
Se ha generado la matriz Y

? matrix b = inv(X'X) * (X'Y)
Se ha generado la matriz b

# Cálculo del estimador de la varianza de la perturbación
? gener s2 = (Y'Y - b'*X'*Y) / (60-3-1)
Se ha generado el escalar s2 = 0.139258

# Cálculo de la matriz de varianzas-covarianzas de los estimadores
? matrix Vb = s2 * inv(X'X)
Se ha generado la matriz Vb

? print b Vb
b (4 x 1)
  11.912
 -3.8561
  0.0078850
 -0.30569

Vb (4 x 4)
  0.91802    -0.11592   -0.00030589   -0.33910
 -0.11592    0.033564   1.3686e-005    0.018695
 -0.00030589 1.3686e-005   2.0684e-007    6.1112e-005
 -0.33910    0.018695   6.1112e-005    0.25604
```

## 4.6.2. Contrastes de nulidad individual

### 4.6.2.1. Contrastes de nulidad individual utilizando del estadístico t

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$\text{Si } H_0 \text{ es cierta } \Rightarrow t_i = \frac{b_i}{S_{b_i}} \xrightarrow{sd} t^{T-K-1}$$

El estadístico  $t_i$  se utiliza para el contraste de nulidad individual de los parámetros y se calcula como el cociente entre el estimador del parámetro cuya nulidad se contrasta y el estimador de su desviación típica:

```
? genr t0=b[1]/sqrt(Vb[1,1])
Se ha generado el escalar t0 = 12.4321

? genr t1=b[2]/sqrt(Vb[2,2])
Se ha generado el escalar t1 = -21.0481

? genr t2=b[3]/sqrt(Vb[3,3])
Se ha generado el escalar t2 = 17.3374

? genr t3=b[4]/sqrt(Vb[4,4])
Se ha generado el escalar t3 = -0.604129
```

*Compruébese que estos estadísticos están disponible en la columna “Estadístico t” de salida estándar del comando **ols** con su p-valor asociado.*

**Interpretación de los contrastes de nulidad individual atendiendo a las tres alternativas señaladas:**

1. El p-valor indica el menor nivel de significación al cuál se rechaza la hipótesis nula.

```
# Pvalores dos colas
? genr pvt0 = 2*pvalue(t,56,abs(t0))
Se ha generado el escalar pvt0 = 9.6678e-018

? genr pvt1 = 2*pvalue(t,56,abs(t1))
Se ha generado el escalar pvt1 = 2.83627e-028

? genr pvt2 = 2*pvalue(t,56,abs(t2))
Se ha generado el escalar pvt2 = 3.55102e-024

? genr pvt3 = 2*pvalue(t,56,abs(t3))
Se ha generado el escalar pvt3 = 0.548197
```

Si se observa el p-valor de los estadísticos  $t$ , este es cero para  $t_0$ ,  $t_1$  y  $t_2$ , es decir, la probabilidad observada de rechazar la hipótesis nula siendo cierta es cero, esto indica que se rechaza la hipótesis nula para cualquier nivel de significación. Para  $t_3$  el p-valor es 0.5482, es decir, la probabilidad observada de rechazar la hipótesis nula siendo cierta es del 54,82%, por tanto, se aceptará la hipótesis nula para cualquier nivel de significación igual o inferior al 54,82% y se rechazará para un nivel de significación superior.

2. Si se fija un nivel de significación  $\alpha$ , se rechaza la hipótesis de nulidad para los parámetros cuyo p-valor sea inferior a dicho nivel de significación y no se rechaza para los parámetros cuyo p-valor sea igual o superior.

Lo primero es fijar el nivel de significación, siendo lo más habitual trabajar con un nivel de significación del 5% ( $\alpha=0.05$ ) o, lo que es equivalente, trabajar con un nivel de confianza del 95% ( $1-\alpha=0.95$ ).

En este caso, para un nivel de significación del 5% se rechaza la hipótesis de nulidad para los parámetros  $\beta_0$ ,  $\beta_1$  y  $\beta_2$  y no se rechaza para el parámetro  $\beta_3$ .

3. Una vez fijado el nivel de significación se puede buscar el valor crítico de la distribución  $t$  que deja a su derecha una cola igual a  $\alpha/2$ , por lo que la regla de decisión sería rechazar  $H_0: \beta_i = 0$  si  $|t_i| > t_{\alpha/2}^{T-K-1}$  y aceptarla en caso contrario.

```
# Valor crítico dos colas
? genr vct= critical(t,56,0.025)
Se ha generado el escalar vct = 2.00324
```

Para un nivel de significación del 5%, el valor crítico de una  $t$  de Student con 56 grados de libertad es 2.00324. Se rechaza la hipótesis de nulidad para los parámetros  $\beta_0$ ,  $\beta_1$  y  $\beta_2$ , dado que sus  $t$ -ratios en valor absoluto son superiores al valor crítico ( $|t_0| = 12.43$ ,  $|t_1| = 21.05$ ,  $|t_2| = 17.34$ ) y, se acepta la nulidad del parámetro  $\beta_3$ , dado que su  $t$ -ratio en valor absoluto es inferior al valor crítico ( $|t_3| = 0.6041$ ). Por lo tanto, se llega a la misma conclusión que utilizando el  $p$ -valor.

De los resultados obtenidos se puede concluir que para un nivel de significación del 5% y dadas las demás variables explicativas incluidas en el modelo:

1. El precio medio de compra del aceite de oliva en euros/litro (P) tiene un efecto individual estadísticamente significativo sobre el consumo familiar mensual medio de aceite de oliva en litros (C).
2. Los ingresos familiares mensuales medios en euros (RF) tienen un efecto individual estadísticamente significativo para la explicación del consumo familiar mensual medio de aceite de oliva en litros (C).
3. El precio medio de compra del aceite de girasol en euros/litro (PG) no tiene un efecto individual estadísticamente significativo sobre el el consumo familiar mensual medio de aceite de oliva en litros (C).

Esta prueba de hipótesis consistente en la contrastación de la nulidad individual para los parámetros del modelo está relacionada con los intervalos de confianza, como se verá en un epígrafe posterior.

#### 4.6.2.2. *Contrastes de nulidad individual a través del comando restrict*

Aún cuando en el comando **restrict** intervenga una única restricción, Gretl proporciona el estadístico F. El usuario debe recordar la equivalencia entre ambos estadísticos: el estadístico  $t$  es la raíz cuadrada del estadístico F con el signo del estimador del parámetro.

##### - Contraste de nulidad individual para $\beta_0$

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[const] = 0
? end restrict

Restricción:
b[const] = 0

Estadístico de contraste: F(1, 56) = 154.556, con valor p = 9.6678e-018

? genr FI0=t0^2
Se ha generado el escalar FI0 = 154.556
```

##### - Contraste de nulidad individual para $\beta_1$

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[P] = 0
? end restrict

Restricción:
b[P] = 0

Estadístico de contraste: F(1, 56) = 443.021, con valor p = 2.83627e-028

? genr FI1=t1^2
Se ha generado el escalar FI1 = 443.021
```

##### - Contraste de nulidad individual para $\beta_2$

```
# Contrastes de nulidad del parámetro que acompaña a la variable RF
? ols C const P RF PG --quiet
? restrict --quiet
? b[RF] = 0
```

```
? end restrict

Restricción:
b[RF] = 0

Estadístico de contraste: F(1, 56) = 300.584, con valor p = 3.55102e-024

? genr FI2=t^2
Se ha generado el escalar FI2 = 300.584
```

**- Contraste de nulidad individual para  $\beta_3$**

```
# Contrastes de nulidad del parámetro que acompaña a la variable PG
? ols C const P RF PG --quiet
? restrict --quiet
? b[PG] = 0
? end restrict
Restricción:
b[PG] = 0

Estadístico de contraste: F(1, 56) = 0.364972, con valor p = 0.548197

? genr FI3=t^3
Se ha generado el escalar FI3 = 0.364972
```

**Tabla 4-2. Resumen estadísticos contraste nulidad individual.**

	t	F	w
const	12.43	99.922	154.556
P	-21.50	-42.231	443.021
RF	17.34	0.0069739	300.584
PG	-0.6041	-13193	0.364972

En la **Tabla 4-2** se recoge el valor de los estadísticos **t**, **F** y **W** para todos los parámetros del modelo. Los cuadrados de las ratios t coinciden con la versión F ( $F_i = t_i^2$ ) y, en este caso, al intervenir una única restricción en el contraste ( $q=1$ ), la versión chi-cuadrado coincide numéricamente con la versión F ( $W_i = F_i$ ).

**4.6.3. Contraste de nulidad conjunta para todos los parámetros del modelo**

$$H_0: \beta = 0_{(K+1) \times 1} \Rightarrow H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_K = 0$$

$$H_1: \beta \neq 0_{(K+1) \times 1} \Rightarrow H_1: \text{alguno ó todos} \neq 0$$

$$\text{Si } H_0 \text{ es cierta} \Rightarrow F_1 = \frac{Q_1}{\frac{SCE}{T-K-1}} = \frac{\frac{b'X'Y}{K+1}}{\frac{SCE}{T-K-1}} \xrightarrow{sd} F_{T-K-1}^{K+1}$$

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[const] = 0
? b[P] = 0
? b[RF] = 0
? b[PG] = 0
? end restrict

Conjunto de restricciones
1: b[const] = 0
2: b[P] = 0
3: b[RF] = 0
4: b[PG] = 0

Estadístico de contraste: F(4, 56) = 11319.3, con valor p = 1.07515e-080
```

Compruébese que este estadístico está disponible en la salida estándar del comando **ols** de un modelo formulado sin ordenada en el origen.

Las tres posibilidades de interpretación de este contraste son:

1. Como el p-valor es nulo, se rechaza la hipótesis nula para cualquier nivel de significación.
2. Para un nivel de significación del 5%, dado que el p-valor es menor que dicho nivel, se rechaza la hipótesis nula.
3. Si se utiliza el valor crítico, la regla de decisión será, si  $F_1 > F_{T-K-1}^{K+1}(\alpha) \Rightarrow$  se rechaza  $H_0: \beta = 0$  y se acepta en caso contrario.

```
# Valor crítico
? genr vcF1 = critical(F,4,56,0.05)
Se ha generado el escalar vcF1 = 2.53658
```

Para un nivel de significación del 5%, el valor crítico de una F de Snedecor con 4 grados de libertad en el numerador y 56 grados de libertad en el denominador es 2.53658. Dado que el valor del estadístico F (11319.3) supera dicho valor crítico, se rechaza la hipótesis nula.

$$F_1 = 11319.3 > F_{T-K-1}^{K+1}(0.05) = F_{56}^4(0.05) = 2.53658$$

En este caso, la versión chi-cuadrado de este estadístico podría calcularse multiplicando por 4 la versión F, puesto que 4 es el número de restricciones<sup>38</sup>.

Por lo tanto, se rechaza la hipótesis de nulidad conjunta de todos los parámetros del modelo, es decir, conjuntamente los regresores son estadísticamente significativos para la explicación del regresando.

#### 4.6.4. Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas del modelo

$$\begin{aligned} H_0: \underline{\beta} = 0 & \Rightarrow H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0 \\ H_1: \underline{\beta} \neq 0 & \Rightarrow H_1: \text{alguno ó todos} \neq 0 \end{aligned}$$

$$\text{Si } H_0 \text{ es cierta} \Rightarrow F_2 = \frac{\frac{Q_2}{K}}{\frac{SCE}{T-K-1}} = \frac{\frac{b' M_{XY}}{K}}{\frac{SCE}{T-K-1}} = \frac{\frac{SCR}{K}}{\frac{SCE}{T-K-1}} = \frac{\frac{R^2}{K}}{\frac{1-R^2}{T-K-1}} \xrightarrow{sd} F_{T-K-1}^K$$

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[P] = 0
? b[RF] = 0
? b[PG] = 0
? end restrict

Conjunto de restricciones
1: b[P] = 0
2: b[RF] = 0
3: b[PG] = 0

Estadístico de contraste: F(3, 56) = 296.457, con valor p = 2.51999e-034

? genr F2 = $Fstat
Se ha generado el escalar F2 = 296.457
```

Compruébese que este estadístico está disponible en la salida estándar del comando **ols** de un modelo formulado con ordenada en el origen y en la tabla ANOVA con su p-valor asociado.

Las tres posibilidades de interpretación de este contraste son:

1. El p-valor es nulo, por lo que se rechaza la hipótesis nula para cualquier nivel de significación.

<sup>38</sup> Debe recordarse que la versión t, sólo se calcula cuando en el contraste interviene una única restricción.

- Para un nivel de significación del 5%, dado que el p-valor es menor que dicho nivel, se rechaza la hipótesis nula.
- Si se utiliza el valor crítico la regla de decisión será, si  $F_2 > F_{T-K-1}^K(\alpha) \Rightarrow$  se rechaza  $H_0: \underline{\beta} = 0$  y se acepta en caso contrario.

Valor crítico  
 ? genr vcF2 = critical(F,3,56,0.05)  
 Se ha generado el escalar vcF2 = 2.76943

Para un nivel de significación del 5%, el valor crítico de una F de Snedecor con 3 grados de libertad en el numerador y 56 grados de libertad en el denominador es 2.76943. Dado que el valor del estadístico F (296.457) supera dicho valor crítico, se rechaza la hipótesis nula.

$$F_2 = 296.457 > F_{T-K-1}^K(0.05) = F_{56}^3(0.05) = 2.765943$$

Por lo tanto, se rechaza la hipótesis de nulidad conjunta de los parámetros que acompañan a las variables explicativas del modelo, es decir, se acepta que conjuntamente las variables explicativas son estadísticamente significativas para la explicación del regresando.

A esta prueba también se le denomina **contraste de significación global** del modelo econométrico. La justificación para dicha denominación se basa en que cuando el coeficiente de determinación es nulo, el estadístico F será nulo y a medida que aumenta el coeficiente de determinación ( $R^2$ ), aumenta el estadístico F, hasta que en el límite, cuando  $R^2$  sea uno F será infinito.

Por tanto, se acepta que el modelo es globalmente válido, lo que equivale a que tiene un elevado  $R^2$  ( $0.940764$ )<sup>39</sup>.

#### 4.6.5. Contraste de nulidad conjunta para un subconjunto paramétrico

$$\begin{aligned} H_0: \beta_{sub} = 0 &\Rightarrow H_0: \beta_{h+1} = \beta_{h+2} = \dots = \beta_K = 0 \\ H_1: \beta_{sub} \neq 0 &\Rightarrow H_1: \text{alguno ó todos} \neq 0 \end{aligned}$$

$$\text{Si } H_0 \text{ es cierta} \Rightarrow F_3 = \frac{\frac{Q_3}{K-h}}{\frac{SCE}{T-K-1}} = \frac{\frac{\Delta SCR}{K-h}}{\frac{SCE}{T-K-1}} = \frac{\frac{SCE_1 - SCE}{K-h}}{\frac{SCE}{T-K-1}} \xrightarrow{sd} F_{T-K-1}^{K-h}$$

donde:

SCE = SCE obtenida de la estimación del *modelo sin restringir*, es decir, con todas las variables explicativas  $Y/X_0, X_1, X_2, \dots, X_h, X_{h+1}, X_{h+2}, \dots, X_K$

SCE<sub>1</sub> = SCE obtenida de la estimación del *modelo restringido*, es decir, con las variables explicativas excluidas del contraste  $Y/X_0, X_1, X_2, \dots, X_h$ .

K = n° de variables explicativas del modelo.

h = n° de variables explicativas excluidas del contraste.

K-h = n° de variables explicativas incluidas en el contraste.

En este caso, la hipótesis nula es  $H_0: \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$  frente a su alternativa  $H_1: \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ .

<sup>39</sup> Debe de tenerse en cuenta que, por razones obvias, este contraste sólo está justificado cuando el modelo está formulado con ordenada en el origen.

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[P] = 0
? b[RF] = 0
? end restrict
```

Conjunto de restricciones  
 1: b[P] = 0  
 2: b[RF] = 0

Estadístico de contraste:  $F(2, 56) = 443.712$ , con valor  $p = 4.54448e-035$

Las tres posibilidades de interpretación de este contraste son:

1. Se rechaza la hipótesis nula para cualquier nivel de significación, dado que el p-valor asociado al estadístico de prueba es cero.
2. Para un nivel de significación del 5%, dado que el p-valor es menor a dicho nivel, se rechaza la hipótesis nula.
3. Si se utiliza el valor crítico, la regla de decisión será, si  $F_3 > F_{T-K-1}^{K-h}(\alpha) \Rightarrow$  se rechaza  $H_0: \beta_{sub} = 0$  y se acepta en caso contrario.

```
# Valor crítico
? genr vcF3 = critical(F,2,56,0.05)
Se ha generado el escalar vcF3 = 3.16186
```

Como  $F_3 = 443.712 > F_{56}^2(0.05) = 3.16186$ , se rechaza la hipótesis nula.

Por lo tanto, se rechaza la hipótesis de nulidad conjunta para los parámetros que acompañan a las variables P y RF, es decir, dada la inclusión en el modelo de PG como variable explicativa, se acepta que conjuntamente las variables P y RF tienen un efecto estadísticamente significativo sobre el regresando.

#### 4.6.6. Contraste de nulidad para una combinación lineal

$$H_0: \pi = 0$$

$$H_1: \pi \neq 0$$

Si  $H_0$  es cierta  $\Rightarrow t = \frac{p}{S_p} \xrightarrow{sd} t^{T-K-1}$  donde  $p = R'b$  y  $S_p^2 = R'\hat{V}(b)R$

En este caso, el contraste propuesto es:

$$H_0: \beta_1 = \beta_3 \Rightarrow H_0: \beta_1 - \beta_3 = 0 \Rightarrow H_0: \pi = 0$$

$$H_1: \beta_1 \neq \beta_3 \Rightarrow H_1: \beta_1 - \beta_3 \neq 0 \Rightarrow H_1: \pi \neq 0$$

```
? matrix p = b[2] - b[4]
Se ha generado la matriz p

? matrix s2p = Vb[2,2] + Vb[4,4] - 2*Vb[2,4]
Se ha generado la matriz s2p

? matrix tp = p / sqrt(s2p)
Se ha generado la matriz tp

? print b Vb p s2p tp
b (4 x 1)
    11.912
   -3.8561
    0.0078850
   -0.30569

Vb (4 x 4)
    0.91802   -0.11592  -0.00030589   -0.33910
   -0.11592    0.033564  1.3686e-005    0.018695
  -0.00030589  1.3686e-005  2.0684e-007   6.1112e-005
   -0.33910    0.018695  6.1112e-005    0.25604
```

```

p (1 x 1)
  -3.5504

s2p (1 x 1)
  0.25222

tp (1 x 1)
  -7.0696

# P-valor
? genr pntp = pvalue(t,56,abs(tp[1,1]))
Se ha generado el escalar pntp = 1.32795e-009

# Valor critico
? genr vct = critical(t,56,0.025)
Se ha generado el escalar vct = 2.00324

```

Se puede comprobar fácilmente que, en este caso, dado que  $R = (0 \ 1 \ 0 \ -1)$ , el estimador de la combinación lineal viene dado por  $p = b_1 - b_3$  y el de su desviación típica estimada, por  $S_p =$

$$\sqrt{S_{b_1}^2 + S_{b_3}^2 - 2S_{b_1b_3}}.$$

Las tres posibilidades de interpretación de este contraste son:

1. Al ser nulo el p-valor, se rechaza esta hipótesis para cualquier nivel de significación
2. Para un nivel de significación del 5%, dado que el p-valor es menor a dicho nivel, se rechaza la hipótesis nula.
3. Si se utiliza el valor crítico, la regla de decisión será, si  $|t| > t_{\alpha/2}^{T-K-1} \Rightarrow$  se rechaza  $H_0: \pi = 0$  y se acepta en caso contrario.

Como el valor absoluto del estadístico t de la combinación lineal (7.0696) supera al valor crítico correspondiente ( $t_{\alpha/2}^{T-K-1} = t_{0.025}^{56} = 2.00324$ ), se rechaza la hipótesis de igualdad para los parámetros que acompañan a las variables P y PG, lo que implica que dada la inclusión en el modelo de la variable RF, las variables P y PG no ejercen la misma influencia para la explicación del regresando.

```

? ols C const P RF PG --quiet
? restrict --quiet
? b[P]- b[PG] = 0
? end restrict

Restricción:
  b[P] - b[PG] = 0

Estadístico de contraste: F(1, 56) = 49.9789, con valor p = 2.6559e-009

? genr F = tp^2
Se ha generado el escalar F = 49.9789

```

#### 4.6.7. Intervalos de confianza

Un intervalo de confianza es un rango de valores construido de tal manera que dentro de sus límites está el verdadero valor del parámetro. La probabilidad de construir un intervalo que contenga el verdadero valor del parámetro es igual al nivel de confianza.

$$\beta_i \in (b_i \pm t_{\alpha/2}^{T-K-1} S_{b_i})$$

Para calcular intervalos de confianza es necesario conocer:

1. Los estimadores de los parámetros ( $b_i$ ).
2. Las desviaciones típicas estimadas de los estimadores de los parámetros ( $S_{b_i}$ ).

3. Los grados de libertad del modelo estimado<sup>40</sup>.
4. El nivel de significación  $\alpha$ .
5. El valor crítico correspondiente de la distribución t de Student  $(t_{\alpha/2}^{T-K-1})$ .

**- Intervalos de confianza para un nivel de significación del 5%**

```
? ols C const P RF PG --quiet

# Vector de las desviaciones estándar de los estimadores de los parámetros
? matrix sb = diag(sqrt(Vb))
Se ha generado la matriz sb

# Valor crítico de una t(56) para un nivel de significación del 5% = 2.00324

# Cálculo del limite inferior
? genr liminf5 = b-sb*2.00324
Se ha generado la matriz liminf5

# Cálculo del limite superior
? genr limsup5 = b+sb*2.00324
Se ha generado la matriz limsup5

# Intervalo de confianza para un nivel de significación del 5%
? print liminf5 limsup5
liminf5 (4 x 1)

      9.9922
     -4.2231
    0.0069739
     -1.3193

limsup5 (4 x 1)

     13.831
     -3.4891
    0.0087960
     0.70796
```

**- Intervalos de confianza para un nivel de significación del 10%**

```
# Valor crítico de una t(56) para un nivel de significación del 10% = 1.67252

# Cálculo del limite inferior
? genr liminf10 = b-sb*1.67252
Se ha generado la matriz liminf10

# Cálculo del limite superior
? genr limsup10 = b+sb*1.67252
Se ha generado la matriz limsup10

# Intervalo de confianza para un nivel de significación del 10%
? print liminf10 limsup10
liminf10 (4 x 1)

     10.309
     -4.1625
    0.0071243
     -1.1520

limsup10 (4 x 1)

     13.514
     -3.5497
    0.0086456
```

<sup>40</sup> Los grados de libertad serán  $(T-K-1)$  si el modelo está formulado con ordenada en el origen y  $(T-K)$  si el modelo está formulado sin ordenada en el origen.

0.54061

**- Intervalos de confianza para un nivel de significación del 1%**

```
# Valor crítico de una t(56) para un nivel de significación del 1% = 2.66651

# Cálculo del limite inferior
? genr liminfl = b-sb*2.66651
Se ha generado la matriz liminfl

# Cálculo del limite superior
? genr limsupl = b+sb*2.66651
Se ha generado la matriz limsupl

# Intervalo de confianza para un nivel de significación del 1%
? print b liminfl limsupl
liminfl (4 x 1)

    9.3567
   -4.3446
    0.0066722
   -1.6550

limsupl (4 x 1)

   14.466
   -3.3676
    0.0090977
    1.0436
```

En la **Tabla 4-3** se recogen los intervalos de confianza para todos los parámetros del modelo para los niveles de significación utilizados.

**Tabla 4-3. Resumen intervalos de confianza.**

	<b>b</b>	<b>Límite inferior</b>	<b>Límite superior</b>
<b>IC 5%</b>	11.912	9.9922	13.831
	-3.8561	-4.2231	-3.4891
	0.0078850	0.0069739	0.0087960
	-0.30569	-1.3193	0.70796
<b>IC 10%</b>	11.912	10.309	13.514
	-3.8561	-4.1625	-3.5497
	0.0078850	0.0071243	0.0086456
	-0.30569	-1.1520	0.54061
<b>IC 1%</b>	11.912	93.567	14.466
	-3.8561	-4.3446	-3.3676
	0.0078850	0.0066722	0.0090977
	-0.30569	-1.6550	1.0436

Como se puede ver en la Ilustración 4-4, el intervalo de confianza del 95% para la ordenada en el origen es (9.9922,13.831), lo que significa que si se seleccionan 100 muestras de tamaño 60 y se construyen 100 intervalos de confianza, se espera que 95 de ellos contengan el verdadero valor del parámetro. Dado que el valor cero, no pertenece a este intervalo, se puede rechazar la hipótesis nula  $\beta_0 = 0$  con una confianza del 95%.

El intervalo de confianza del 95% para el coeficiente de la variable P es (-4.2231, -3.4891). Por tanto, dado que el valor cero no pertenece a este intervalo, se puede rechazar la hipótesis nula  $\beta_1 = 0$  con una confianza del 95%.

Siendo (0.0069739, 0.0087960) el intervalo de confianza del 95% para el coeficiente de la variable RF, como en él no aparece el cero, se concluye rechazando la nulidad del parámetro  $\beta_2$ .

El intervalo de confianza del 95% para el coeficiente de la variable PG es (1.3193, 0.70796), intervalo que contiene el cero y, por tanto, se acepta la nulidad del parámetro  $\beta_3$ .

Aceptar una hipótesis no significa necesariamente que dicha hipótesis sea cierta, en algunos casos, puede ocurrir que se tenga una gran incertidumbre sobre el verdadero valor del parámetro y, en dicha

situación, la aceptación sólo significa, que con la información de la que se dispone no se tienen elementos de juicio suficientes para rechazar dicha hipótesis.

Se puede concluir que, para los niveles de significación del 1%, 5% y 10% se acepta la hipótesis de nulidad individual sólo para el parámetro  $\beta_3$ , puesto que su intervalo de confianza es el único que contiene el cero.

A pesar de que en esta versión de Gretl no está disponible ningún comando para la construcción de intervalos de confianza, se puede acceder a la ventana **intervalos de confianza para los coeficientes** a través del menú **Análisis** de la **Ventana Modelo**, en la que Gretl proporciona los intervalos de confianza para un nivel de significación del 5%. Debe observarse que Gretl recuerda el valor crítico y los coeficientes estimados que se han utilizado para dichos cálculos (ver Ilustración 4-4).

VARIABLE	COEFICIENTE	INTERVALO DE CONFIANZA 95%	
const	11.9116	9.99224	13.8310
P	-3.85611	-4.22311	-3.48910
RF	0.00788497	0.00697390	0.00879603
PG	-0.305693	-1.31934	0.707958

Ilustración 4-4. Intervalos de confianza del 95% utilizando el menú **Análisis** de la **Ventana Modelo**.

#### 4.6.8. Regiones de confianza

Mientras que un intervalo de confianza es un rango de valores para un coeficiente y se construye a partir de un estadístico  $t$ , una región de confianza es un rango de valores para varios coeficientes simultáneamente y se construye a partir del correspondiente estadístico  $F$ .

A pesar de que en esta versión de Gretl no está disponible ningún comando para la construcción de regiones de confianza, se debe recordar que se puede acceder a regiones de confianza bidimensionales (elipses) a través de la opción **elipse de confianza** del menú **Análisis** de la **Ventana Modelo**.

La región de confianza más sencilla es la referida a dos coeficientes, se trata de una elipse cuyo centro es el estimador MCO. Al igual que ocurre en los intervalos de confianza para un solo coeficiente, las dimensiones de la elipse dependen directamente de las varianzas estimadas de los estimadores de los coeficientes: a mayor varianza, mayor será la elipse y viceversa. La inclinación de la elipse depende de la covarianza entre los dos coeficientes estimados: la elipse asciende de izquierda a derecha si tal covarianza es positiva, descendiendo en caso contrario.

De forma análoga al intervalo de confianza, la región de confianza se puede utilizar para efectuar contrastes de hipótesis, sin más que examinar si el punto establecido en dicha hipótesis se encuentra dentro o fuera de la región de confianza. Si dicho punto está dentro, se acepta la hipótesis nula y si está fuera se rechaza. En este caso, el punto (0,0) no pertenece a la elipse y, por tanto, se rechaza la hipótesis nula.

Además, Gretl permite comparar la región de confianza bidimensional con los intervalos de confianza individuales, para lo cual traza dos líneas paralelas a cada eje que pasan por los límites de los intervalos individuales.

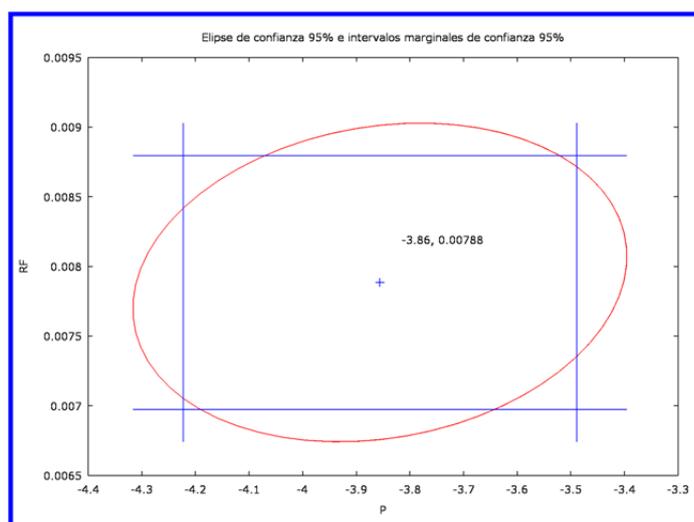


Ilustración 4-5. Elipse de confianza del 95% utilizando el menú **Análisis** de la **Ventana Modelo**.

Con ello queda claro que la región de confianza conjunta no es necesariamente la intersección de los intervalos de confianza individuales y, por tanto, los contrastes individuales y conjuntos son independientes. Es decir, se puede aceptar la nulidad de los parámetros a nivel individual y no a nivel conjunto, a la inversa, etc. En este caso no existen puntos que pertenezcan a la región de confianza sin pertenecer al menos a uno de los intervalos de confianza individuales (véase Ilustración 4-5), siendo ello sintomático de que las variables P y RF son ortogonales o están muy cerca de la ortogonalidad (el grado de correlación entre ambas variables es -0.1172).

```
# Cálculo coeficiente de correlación entre P y RF
? genr corPRF = corr(P,RF)
Se ha generado el escalar corPRF = -0.117228
```

En la parte derecha de la Ilustración 4-6, se puede ver una elipse de confianza para dos parámetros de variables altamente colineales y en la parte izquierda, para dos parámetros de variables ortogonales. En el primer caso, hay puntos de la elipse que no pertenecen a ninguno de los intervalos individuales mientras que en el segundo caso, todos los puntos de la elipse pertenecen al menos a un intervalo individual.

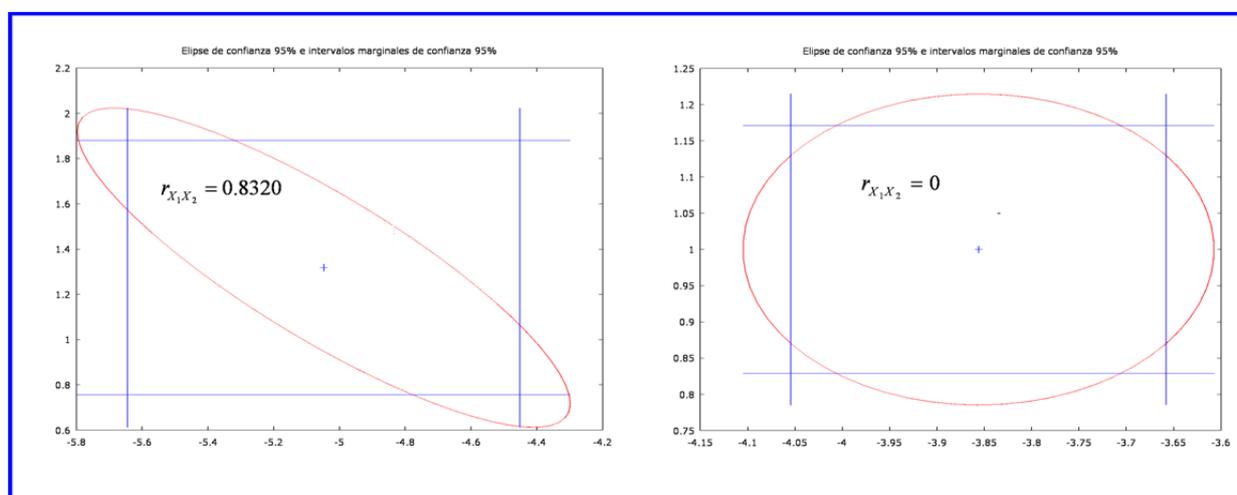


Ilustración 4-6. Ejemplos de elipses de confianza.

### 4.6.9. Contrastes de hipótesis utilizando la matriz de restricciones lineales (R)

Para calcular el estadístico F a partir de la fórmula genérica es necesario conocer:

1. El vector de estimadores.
2. El estimador de la varianza de la perturbación.
3. La inversa de la matriz de productos cruzados de los regresores.

O bien, el vector de estimadores y la matriz de varianzas-covarianzas estimada de los estimadores.

$$F = \frac{\frac{(Rb - r)'[R(X'X)^{-1}R']^{-1}(Rb - r)}{q}}{\frac{SCE}{T - K - 1}} = \frac{(Rb - r)'[R\hat{V}(b)R']^{-1}(Rb - r)}{q}$$

- **Contraste de nulidad individual para  $\beta_0$**

$$\begin{aligned} H_0: \beta_0 &= 0 \\ H_1: \beta_0 &\neq 0 \end{aligned} \Rightarrow F = \frac{b_0' [S_{b_0}^2]^{-1} b_0}{1} = \frac{b_0^2}{S_{b_0}^2} = t_0^2$$

```
? ols C const P RF PG --quiet
? matrix XTXI = inv(X'X)
Se ha generado la matriz XTXI
```

En este caso, la matriz  $R$  es un vector fila de 4 columnas con un uno en el lugar correspondiente al parámetro  $\beta_0$  (parámetro cuya nulidad se contrasta) y el resto de sus elementos nulos, por lo que el producto  $Rb$  selecciona el elemento  $b_0$  del vector  $b$ :

```
? matrix R0 = {1, 0, 0, 0}
Se ha generado la matriz R0

? genr q0 = 1
Se ha generado el escalar q0 = 1

? matrix R0b = R0*b
Se ha generado la matriz R0b

? print R0 b R0b
R0 (1 x 4)
 1  0  0  0

b (4 x 1)
 11.912
 -3.8561
 0.0078850
 -0.30569

R0b (1 x 1)
 11.912
```

$r$  es un escalar y es igual a cero, dado que es el valor que se establece para la restricción en la hipótesis nula:

```
? matrix r0 = 0
Se ha generado la matriz r0
```

El producto  $R(X'X)^{-1}R'$  selecciona el primer elemento de la diagonal principal de la inversa de  $X'X$ :

```
? matrix R0XTXI = R0*XTXI*R0'
Se ha generado la matriz R0XTXI

? print XTXI R0XTXI
XTXI (4 x 4)
 6.5923    -0.83239   -0.0021966   -2.4351
 -0.83239    0.24102    9.8275e-005    0.13425
 -0.0021966  9.8275e-005  1.4853e-006    0.00043884
 -2.4351     0.13425    0.00043884     1.8386

R0XTXI (1 x 1)
 6.5923
```

El producto  $R\hat{V}(b)R'$  selecciona el primer elemento de la diagonal principal de  $\hat{V}(b)$ :

```
? matrix R0Vb = R0*Vb*R0'
Se ha generado la matriz R0Vb

? print Vb R0Vb
Vb (4 x 4)
 0.91802    -0.11592   -0.00030589   -0.33910
```

```

-0.11592    0.033564  1.3686e-005    0.018695
-0.00030589  1.3686e-005  2.0684e-007  6.1112e-005
-0.33910    0.018695  6.1112e-005    0.25604

R0Vb (1 x 1)
0.91802

```

Por lo tanto:

```

? matrix Fb0c1 = (R0b' * inv(R0XTXI) * R0b) / s2
Se ha generado la matriz Fb0c1

? matrix Fb0c2 = R0b' * inv(R0Vb) * R0b
Se ha generado la matriz Fb0c2

? print Fb0c1 Fb0c2
Fb0c1 (1 x 1)
154.56

Fb0c2 (1 x 1)
154.56

```

#### - Contraste de nulidad individual para $\beta_1$

$$H_0: \beta_1 = 0 \quad \Rightarrow F = \frac{b_1' [S_{b_1}^2]^{-1} b_1}{1} = \frac{b_1^2}{S_{b_1}^2} = t_1^2$$

$$H_1: \beta_1 \neq 0$$

En este caso la matriz R es un vector fila de 4 columnas con un uno en el lugar correspondiente al parámetro  $\beta_1$  (parámetro cuya nulidad se contrasta) y el resto de sus elementos nulos, por lo que el producto  $Rb$  selecciona el elemento  $b_1$  del vector  $b$ :

```

? matrix R1 = {0, 1, 0, 0}
Se ha generado la matriz R1

? genr q1 = 1
Se ha generado el escalar q1 = 1

? matrix R1b = R1*b
Se ha generado la matriz R1b

? print R1 b R1b
R1 (1 x 4)
0  1  0  0

b (4 x 1)
11.912
-3.8561
0.0078850
-0.30569

R1b (1 x 1)
-3.8561

```

r es un escalar y es igual a cero dado que es el valor que se establece para la restricción en la hipótesis nula:

```

? matrix r1 = 0
Se ha generado la matriz r1

```

El producto  $R(X'X)^{-1}R'$  selecciona el segundo elemento de la diagonal principal de la inversa de  $X'X$ :

```

? matrix R1XTXI = R1*XTXI*R1'
Se ha generado la matriz R1XTXI

? print XTXI R1XTXI
XTXI (4 x 4)
6.5923    -0.83239   -0.0021966   -2.4351
-0.83239    0.24102   9.8275e-005    0.13425
-0.0021966  9.8275e-005  1.4853e-006  0.00043884
-2.4351    0.13425   0.00043884    1.8386

```

```
R1XTXI (1 x 1)
0.24102
```

El producto  $R\hat{V}(b)R'$  selecciona el segundo elemento de la diagonal principal de  $\hat{V}(b)$ :

```
? matrix R1Vb = R1*Vb*R1'
Se ha generado la matriz R1Vb

? print Vb R1Vb
Vb (4 x 4)
  0.91802    -0.11592   -0.00030589   -0.33910
 -0.11592    0.033564   1.3686e-005    0.018695
 -0.00030589 1.3686e-005   2.0684e-007    6.1112e-005
 -0.33910    0.018695   6.1112e-005    0.25604

R1Vb (1 x 1)
0.033564
```

Por lo tanto:

```
? matrix Fb1c1 = (R1b' * inv(R1XTXI) * R1b) / s2
Se ha generado la matriz Fb1c1

? matrix Fb1c2 = R1b' * inv(R1Vb) * R1b
Se ha generado la matriz Fb1c2

? print Fb1c1 Fb1c2
Fb1c1 (1 x 1)
443.02

Fb1c2 (1 x 1)
443.02
```

#### - Contraste de nulidad individual para $\beta_2$

$$H_0: \beta_2 = 0 \Rightarrow F = \frac{b_2' [S_{b_2}^2]^{-1} b_2}{1} = \frac{b_2^2}{S_{b_2}^2} = t_2^2$$

$$H_1: \beta_2 \neq 0$$

En este caso la matriz R es un vector fila de 4 columnas con un uno en el lugar correspondiente al parámetro  $\beta_2$  (parámetro cuya nulidad se contrasta) y el resto de sus elementos nulos, por lo que el producto  $Rb$  selecciona el elemento  $b_2$  del vector  $b$ :

```
? matrix R2 = {0, 0, 1, 0}
Se ha generado la matriz R2

? genr q2 = 1
Se ha generado el escalar q2 = 1

? matrix R2b = R2*b
Se ha generado la matriz R2b

? print R2 b R2b
R2 (1 x 4)
  0  0  1  0

b (4 x 1)
 11.912
 -3.8561
 0.0078850
 -0.30569

R2b (1 x 1)
0.0078850
```

r es un escalar y es igual a cero dado que es el valor que se establece para la restricción en la hipótesis nula:

```
? matrix r2 = 0
Se ha generado la matriz r2
```

El producto  $R(X'X)^{-1}R'$  selecciona el tercer elemento de la diagonal principal de la inversa de  $X'X$ :

```
? matrix R2XTXI = R2*XTXI*R2'
Se ha generado la matriz R2XTXI

? print XTXI R2XTXI
XTXI (4 x 4)
  6.5923      -0.83239   -0.0021966   -2.4351
 -0.83239     0.24102    9.8275e-005    0.13425
 -0.0021966  9.8275e-005  1.4853e-006    0.00043884
 -2.4351      0.13425    0.00043884    1.8386

R2XTXI (1 x 1)
1.4853e-006
```

El producto  $R\hat{V}(b)R'$  selecciona el tercer elemento de la diagonal principal de  $\hat{V}(b)$ :

```
? matrix R2Vb = R2*Vb*R2'
Se ha generado la matriz R2Vb

? print Vb R2Vb
Vb (4 x 4)
  0.91802     -0.11592   -0.00030589   -0.33910
 -0.11592     0.033564   1.3686e-005    0.018695
 -0.00030589 1.3686e-005   2.0684e-007    6.1112e-005
 -0.33910     0.018695   6.1112e-005    0.25604

R2Vb (1 x 1)
2.0684e-007
```

Por lo tanto:

```
? matrix Fb2c1 = (R2b' * inv(R2XTXI) * R2b) / s2
Se ha generado la matriz Fb2c1

? matrix Fb2c2 = R2b' * inv(R2Vb) * R2b
Se ha generado la matriz Fb2c2

? print Fb2c1 Fb2c2
Fb2c1 (1 x 1)
300.58

Fb2c2 (1 x 1)
300.58
```

#### - Contraste de nulidad individual para $\beta_3$

$$H_0: \beta_3 = 0 \quad \Rightarrow F = \frac{b_3' [S_{b_3}^2]^{-1} b_3}{1} = \frac{b_3^2}{S_{b_3}^2} = t_3^2$$

$$H_1: \beta_3 \neq 0$$

En este caso la matriz  $R$  es un vector fila de 4 columnas con un uno en el lugar correspondiente al parámetro  $\beta_3$  (parámetro cuya nulidad se contrasta) y el resto de sus elementos nulos, por lo que el producto  $Rb$  selecciona el elemento  $b_3$  del vector  $b$ :

```
? matrix R3 = {0, 0, 0, 1}
Se ha generado la matriz R3

? genr q3 = 1
Se ha generado el escalar q3 = 1

? matrix R3b = R3*b
Se ha generado la matriz R3b

? print R3 b R3b
R3 (1 x 4)
  0  0  0  1

b (4 x 1)
 11.912
 -3.8561
 0.0078850
```

```
-0.30569
R3b (1 x 1)
-0.30569
```

$r$  es un escalar y es igual a cero dado que es el valor que se establece para la restricción en la hipótesis nula:

```
? matrix r3 = 0
Se ha generado la matriz r3
```

El producto  $R(X'X)^{-1}R'$  selecciona el cuarto elemento de la diagonal principal de la inversa de  $X'X$ :

```
? matrix R3XTXI = R3*XTXI*R3'
Se ha generado la matriz R3XTXI

? print XTXI R3XTXI
XTXI (4 x 4)
  6.5923      -0.83239   -0.0021966    -2.4351
 -0.83239     0.24102    9.8275e-005    0.13425
-0.0021966   9.8275e-005  1.4853e-006    0.00043884
 -2.4351     0.13425    0.00043884     1.8386

R3XTXI (1 x 1)
  1.8386
```

El producto  $R\hat{V}(b)R'$  selecciona el cuarto elemento de la diagonal principal de  $\hat{V}(b)$ :

```
? matrix R3Vb = R3*Vb*R3'
Se ha generado la matriz R3Vb

? print Vb R3Vb
Vb (4 x 4)
  0.91802     -0.11592   -0.00030589   -0.33910
 -0.11592     0.033564    1.3686e-005    0.018695
-0.00030589  1.3686e-005    2.0684e-007    6.1112e-005
 -0.33910     0.018695    6.1112e-005    0.25604

R3Vb (1 x 1)
  0.25604
```

Por lo tanto:

```
? matrix Fb3c1 = (R3b' * inv(R3XTXI) * R3b) / s2
Se ha generado la matriz Fb3c1

? matrix Fb3c2 = R3b' * inv(R3Vb) * R3b
Se ha generado la matriz Fb3c2

? print Fb3c1 Fb3c2
Fb3c1 (1 x 1)
  0.36497

Fb3c2 (1 x 1)
  0.36497
```

- **Contraste de nulidad conjunta para todos los parámetros del modelo**

$$\begin{aligned}
 H_0: \beta = 0_{(K+1) \times 1} &\Rightarrow H_0: \beta_0 = \beta_1 = \dots = \beta_3 = 0 &\Rightarrow F = F_1 = \frac{\frac{b'(X'X)b}{K+1}}{\frac{SCE}{T-K-1}} = \frac{b'[\hat{V}(b)]^{-1}b}{K+1} \\
 H_1: \beta \neq 0_{(K+1) \times 1} &\Rightarrow H_1: \text{alguno o todos} \neq 0
 \end{aligned}$$

En este caso  $R$  es una matriz de 4 filas y 4 columnas, donde los elementos de su diagonal principal son iguales a uno y los elementos no diagonales son iguales a cero, es decir,  $R$  es una matriz identidad de orden 4, por lo que el producto  $Rb$  selecciona el vector  $b$ :

```
? matrix R0123 = {1, 0, 0, 0; 0, 1, 0, 0; 0, 0, 1, 0; 0, 0, 0, 1}
Se ha generado la matriz R0123

? matrix r0123 = {0; 0; 0; 0}
```

```

Se ha generado la matriz r0123

? genr q0123 = 4
Se ha generado el escalar q0123 = 4

? genr R0123b = R0123*b
Se ha generado la matriz R0123b

? print R0123 b R0123b
R0123 (4 x 4)

  1  0  0  0
  0  1  0  0
  0  0  1  0
  0  0  0  1

b (4 x 1)
  11.912
 -3.8561
  0.0078850
 -0.30569

R0123b (4 x 1)
  11.912
 -3.8561
  0.0078850
 -0.30569

```

$r$  es un vector columna que contiene 4 ceros, dado que es el valor que se fija para las 4 restricciones que se establecen en la hipótesis nula:

```

? print r0123
r0123 (4 x 1)
  0
  0
  0
  0

```

El producto  $R(X'X)^{-1}R'$  selecciona todas las filas y columnas de la inversa de la matriz  $X'X$ :

```

? matrix R0123XTXI = R0123*XTXI*R0123'
Se ha generado la matriz R0123XTXI

? print XTXI R0123XTXI
XTXI (4 x 4)
  6.5923    -0.83239   -0.0021966   -2.4351
 -0.83239    0.24102    9.8275e-005    0.13425
 -0.0021966  9.8275e-005  1.4853e-006    0.00043884
 -2.4351     0.13425    0.00043884     1.8386

R0123XTXI (4 x 4)
  6.5923    -0.83239   -0.0021966   -2.4351
 -0.83239    0.24102    9.8275e-005    0.13425
 -0.0021966  9.8275e-005  1.4853e-006    0.00043884
 -2.4351     0.13425    0.00043884     1.8386

```

El producto  $R\hat{V}(b)R'$  selecciona todas las filas y columnas de la matriz  $\hat{V}(b)$ :

```

? matrix R0123Vb = R0123*Vb*R0123'
Se ha generado la matriz R0123Vb

? print Vb R0123Vb
Vb (4 x 4)
  0.91802    -0.11592   -0.00030589   -0.33910
 -0.11592    0.033564    1.3686e-005    0.018695
 -0.00030589  1.3686e-005    2.0684e-007    6.1112e-005
 -0.33910     0.018695    6.1112e-005     0.25604

R0123Vb (4 x 4)
  0.91802    -0.11592   -0.00030589   -0.33910
 -0.11592    0.033564    1.3686e-005    0.018695
 -0.00030589  1.3686e-005    2.0684e-007    6.1112e-005
 -0.33910     0.018695    6.1112e-005     0.25604

```

Por lo tanto:

```
? matrix Fbc1 = (R0123b' * inv(R0123XTXI) * R0123b) / (q0123*s2)
Se ha generado la matriz Fbc1

? matrix Fbc2 = R0123b' * inv(R0123Vb) * R0123b / q0123
Se ha generado la matriz Fbc2

? print Fbc1 Fbc2
Fbc1 (1 x 1)
    11319.

Fbc2 (1 x 1)
    11319.
```

**- Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas del modelo**

$$\begin{aligned}
 H_0: \underline{\beta} &= 0_{K \times 1} & \Rightarrow & H_0: \beta_1 = \beta_2 = \beta_3 = 0 & \Rightarrow & F = F_2 = \frac{\underline{b}' M_{XX} \underline{b}}{SCE} = \frac{\underline{b}' [\hat{V}(\underline{b})]^{-1} \underline{b}}{K - h} \\
 H_1: \underline{\beta} &\neq 0_{K \times 1} & \Rightarrow & H_1: \text{alguno o todos} \neq 0 & & 
 \end{aligned}$$

En este caso R es una matriz de 3 filas y 4 columnas, donde todos los elementos son cero excepto los 3 últimos de la “diagonal principal” que son iguales a uno, por lo que el producto  $Rb$  selecciona el subvector  $\underline{b}$  del vector  $b$ :

```
? matrix R123 = {0, 1, 0, 0; 0, 0, 1, 0; 0, 0, 0, 1}
Se ha generado la matriz R123

? matrix r123 = {0; 0; 0}
Se ha generado la matriz r123

? genr q123 = 3
Se ha generado el escalar q123 = 3

? matrix R123b = R123*b
Se ha generado la matriz R123b

? print R123 b R123b
R123 (3 x 4)
    0  1  0  0
    0  0  1  0
    0  0  0  1

b (4 x 1)
    11.912
    -3.8561
    0.0078850
    -0.30569

R123b (3 x 1)
    -3.8561
    0.0078850
    -0.30569
```

r es un vector columna que contiene 3 ceros, dado que es el valor que se fija para las 3 restricciones que se establecen en la hipótesis nula.

```
? print r123
r123 (3 x 1)
    0
    0
    0
```

El producto  $R(X'X)^{-1}R'$  selecciona la submatriz formada por las 3 últimas filas y columnas de la inversa de la matriz  $X'X$ <sup>41</sup>.

```
? genr R123XTXI = R123*XTXI*R123'
Se ha generado la matriz R123XTXI

? print XTXI R123XTXI
XTXI (4 x 4)
    6.5923      -0.83239   -0.0021966     -2.4351
   -0.83239     0.24102    9.8275e-005     0.13425
  -0.0021966    9.8275e-005  1.4853e-006     0.00043884
   -2.4351      0.13425    0.00043884     1.8386

R123XTXI (3 x 3)
    0.24102    9.8275e-005     0.13425
   9.8275e-005  1.4853e-006     0.00043884
    0.13425    0.00043884     1.8386
```

El producto  $R\hat{V}(b)R'$  selecciona la submatriz formada por las 3 últimas filas y columnas de la matriz  $\hat{V}(b)$ <sup>42</sup>.

```
? genr R123Vb = R123*Vb*R123'
Se ha generado la matriz R123Vb

? print Vb R123Vb
Vb (4 x 4)
    0.91802     -0.11592   -0.00030589     -0.33910
   -0.11592     0.033564    1.3686e-005     0.018695
  -0.00030589  1.3686e-005    2.0684e-007     6.1112e-005
   -0.33910     0.018695    6.1112e-005     0.25604

R123Vb (3 x 3)
    0.033564    1.3686e-005     0.018695
   1.3686e-005  2.0684e-007     6.1112e-005
    0.018695    6.1112e-005     0.25604
```

Por lo tanto:

```
? matrix Fbrayac1 = (R123b' * inv(R123XTXI) * R123b) / (q123*s2)
Se ha generado la matriz Fbrayac1

? matrix Fbrayac2 = R123b' * inv(R123Vb) * R123b / q123
Se ha generado la matriz Fbrayac2

? print Fbrayac1 Fbrayac2
Fbrayac1 (1 x 1)
    296.46

Fbrayac2 (1 x 1)
    296.46
```

#### - Contrastes de nulidad conjunta para un subconjunto paramétrico

$$\begin{aligned}
 H_0: \beta_{sub} &= 0_{(K-h)x1} \Rightarrow H_0: \beta_1 = \beta_2 = 0 \\
 H_1: \beta_{sub} &\neq 0_{(K-h)x1} \Rightarrow H_1: \text{alguno o ambos} \neq 0 \Rightarrow F = \frac{\frac{b'_{sub} D b_{sub}}{K-h}}{\frac{SCE}{T-K-1}} = \frac{b'_{sub} [\hat{V}(b_{sub})]^{-1} b_{sub}}{K-h}
 \end{aligned}$$

<sup>41</sup> Dicha submatriz coincide con la inversa de la matriz de productos cruzados de las variables explicativas centradas respecto a su media muestral  $[M_{XX}^{-1}]$ .

<sup>42</sup> Dicha submatriz coincide con la matriz de varianzas-covarianzas estimada de los estimadores de los parámetros que acompañan a las variables explicativas del modelo  $[\hat{V}(\underline{b})]$

En este caso  $R$  es una matriz de 2 filas y 4 columnas, donde todos los elementos son cero excepto los 2 centrales de la “diagonal principal” que son iguales a uno, por lo que el producto  $Rb$  selecciona el subvector  $b_{sub} = \begin{pmatrix} b_1 \\ b_1 \end{pmatrix}$  del vector  $b$ :

```
? matrix R12 = {0, 1, 0, 0; 0, 0, 1, 0}
Se ha generado la matriz R12

? genr q12 = 2
Se ha generado el escalar q12 = 2

? matrix R12b = R12*b
Se ha generado la matriz R12b

? print R12 b R12b
R12 (2 x 4)
  0  1  0  0
  0  0  1  0

b (4 x 1)
  11.912
 -3.8561
  0.0078850
 -0.30569

R12b (2 x 1)
 -3.8561
  0.0078850
```

$r$  es un vector columna que contiene 2 ceros, dado que es el valor que se fija para las 2 restricciones que se establecen en la hipótesis nula:

```
? matrix r12 = {0; 0}
Se ha generado la matriz r12

? print r12
r12 (2 x 1)
  0
  0
```

El producto  $R(X'X)^{-1}R'$  selecciona la submatriz  $(X'X)_{sub}^{-1}$  formada por las 2 filas y columnas centrales de la inversa de la matriz  $X'X$ :

```
? genr R12XTXI = R12*XTXI*R12'
Se ha generado la matriz R12XTXI

? print XTXI R12XTXI
XTXI (4 x 4)
  6.5923    -0.83239   -0.0021966   -2.4351
 -0.83239    0.24102    9.8275e-005    0.13425
 -0.0021966  9.8275e-005    1.4853e-006    0.00043884
 -2.4351     0.13425    0.00043884     1.8386

R12XTXI (2 x 2)
  0.24102    9.8275e-005
  9.8275e-005  1.4853e-006
```

El producto  $R\hat{V}(b)R'$  selecciona la submatriz  $\hat{V}(b_{sub})$  formada por las 2 filas y columnas centrales de la matriz  $\hat{V}(b)$ :

```
? genr R12Vb = R12*Vb*R12'
Se ha generado la matriz R12Vb

? print Vb R12Vb
Vb (4 x 4)
  0.91802    -0.11592   -0.00030589   -0.33910
 -0.11592    0.033564    1.3686e-005    0.018695
 -0.00030589  1.3686e-005    2.0684e-007    6.1112e-005
 -0.33910    0.018695    6.1112e-005    0.25604
```

```
R12Vb (2 x 2)
  0.033564  1.3686e-005
  1.3686e-005  2.0684e-007
```

Por lo tanto:

```
? matrix Fb12c1 = (R12b' * inv(R12XTXI) * R12b) / (q12*s2)
Se ha generado la matriz Fb12c1

? matrix Fb12c2 = R12b' * inv(R12Vb) * R12b / q12
Se ha generado la matriz Fb12c2

? print Fb12c1 Fb12c2
Fb12c1 (1 x 1)
  443.71

Fb12c2 (1 x 1)
  443.71
```

#### - Contrastes de nulidad para una combinación lineal

$$H_0: \beta_1 = \beta_3 \Rightarrow \pi = \beta_1 - \beta_3 = 0 \Rightarrow F = \frac{(Rb - r)' [R\hat{V}(b)R']^{-1} (Rb - r)}{q} = \frac{p[S_p^2]^{-1} p}{1} = \frac{p^2}{S_p^2}$$

$$H_1: \pi \neq 0$$

En este caso la matriz R es un vector fila de 4 columnas (0 1 0 -1), por lo que el producto Rb será el estimador de la combinación lineal  $p = b_1 - b_3$ :

```
? matrix Rig13 = {0, 1, 0, -1}
Se ha generado la matriz Rig13

? genr qig13 = 1
Se ha generado el escalar qig13 = 1

? matrix Rig13b = Rig13*b
Se ha generado la matriz Rig13b

? print Rig13 b Rig13b
Rig13 (1 x 4)
  0  1  0  -1

b (4 x 1)
  11.912
 -3.8561
  0.0078850
 -0.30569

Rig13b (1 x 1)
 -3.5504

? matrix p = b[2] - b[4]
Se ha generado la matriz p

? print p
p (1 x 1)
 -3.5504
```

r es un escalar y es igual a cero, dado que es el valor que se establece para la combinación lineal en la hipótesis nula:

```
? matrix rig13 = 0
Se ha generado la matriz rig13
```

El producto  $R(X'X)^{-1}R'$  vendrá dado por la combinación  $x^{22} - 2x^{24} + x^{44}$ , donde  $x^{ij}$  es el elemento de la fila i-ésima y de la columna j-ésima de la inversa de la matriz de productos cruzados de los regresores:

```
? matrix rig13 = 0
Se ha generado la matriz rig13

? matrix Rig13XTXI = Rig13*XTXI*Rig13'
```

```

Se ha generado la matriz Rig13XTXI

? print XTXI Rig13XTXI
XTXI (4 x 4)
  6.5923      -0.83239   -0.0021966    -2.4351
 -0.83239     0.24102    9.8275e-005    0.13425
 -0.0021966  9.8275e-005  1.4853e-006    0.00043884
 -2.4351      0.13425    0.00043884     1.8386

Rig13XTXI (1 x 1)
  1.8111

? matrix xtxi13 = XTXI[2,2] - 2* XTXI[2,4] + XTXI[4,4]
Se ha generado la matriz xtxi13

? print xtxi13
xtxi13 (1 x 1)
  1.8111

```

El producto  $R\hat{V}(b)R'$  será el estimador de la varianza de la combinación lineal  $S_p^2 = S_{b_1}^2 - 2S_{b_1b_3} + S_{b_3}^2$ :

```

? matrix Rig13Vb = Rig13*Vb*Rig13'
Se ha generado la matriz Rig13Vb

? print Vb Rig13Vb
Vb (4 x 4)
  0.91802      -0.11592   -0.00030589   -0.33910
 -0.11592     0.033564    1.3686e-005    0.018695
 -0.00030589  1.3686e-005    2.0684e-007    6.1112e-005
 -0.33910     0.018695    6.1112e-005    0.25604

Rig13Vb (1 x 1)
  0.25222

? matrix s2p = Vb[2,2] - 2* Vb[2,4] + Vb[4,4]
Se ha generado la matriz s2p

? print s2p
s2p (1 x 1)
  0.25222

```

Por lo tanto:

```

? matrix Fbig13c1 = (Rig13b' * (1/Rig13XTXI) * Rig13b) / (qig13*s2)
Se ha generado la matriz Fbig13c1

? matrix Fbig13c2 = Rig13b' * (1/Rig13Vb) * Rig13b / qig13
Se ha generado la matriz Fbig13c2

? print Fbig13c1 Fbig13c2
Fbig13c1 (1 x 1)
  49.979

Fbig13c2 (1 x 1)
  49.979

```

#### 4.6.10. Contrastes de hipótesis mediante sumas residuales

Una forma alternativa de calcular el estadístico F es utilizar las sumas residuales del modelo sin restringir (MSR) y del modelo restringido (MR):

$$F = \frac{\frac{SCE_{MR} - SCE_{MSR}}{q}}{\frac{SCE_{MSR}}{T - K - 1}}$$

```

# Estimación MCO del MSR
? ols C const P RF PG --quiet

# SCE del MSR

```

```
? genr SCEMSR = $ess
Se ha generado el escalar SCEMSR = 7.79842

# Grados de libertad del MSR
? genr GLMSR = 60 - 3 - 1
Se ha generado el escalar GLMSR = 56
```

#### - Contraste de nulidad individual para $\beta_0$

$$H_0: \beta_0 = 0 \Rightarrow \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \beta_0 \neq 0 \Rightarrow \text{MR} \rightarrow C_t = \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

```
# Estimación MCO del MR
? ols C P RF PG --quiet

# SCE del MR
? genr SCEMR = $ess
Se ha generado el escalar SCEMR = 29.3216

# Número de restricciones
? genr q = 1
Se ha generado el escalar q = 1

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha generado el escalar F = 154.556

? print SCEMR SCEMSR GLMSR q F

      SCEMR = 29.321556

      SCEMSR = 7.7984227

      GLMSR = 56.000000

      q = 1.0000000

      F = 154.55631
```

#### - Contraste de nulidad individual para $\beta_1$

$$H_0: \beta_1 = 0 \Rightarrow \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \beta_1 \neq 0 \Rightarrow \text{MR} \rightarrow C_t = \beta_0 + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

```
# Estimación MCO del MR
? ols C const RF PG --quiet

# SCE del MR
? genr SCEMR = $ess
Se ha reemplazado el escalar SCEMR = 69.4925

# Número de restricciones
? genr q = 1
Se ha reemplazado el escalar q = 1

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha reemplazado el escalar F = 443.021

? print SCEMR SCEMSR GLMSR q F

      SCEMR = 69.492458

      SCEMSR = 7.7984227

      GLMSR = 56.000000

      q = 1.0000000

      F = 443.02112
```

#### - Contraste de nulidad individual para $\beta_2$

$$H_0: \beta_2 = 0 \Rightarrow \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \beta_2 \neq 0 \Rightarrow \text{MR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_3 PG_t + \varepsilon_t$$

```
# Estimación MCO del MR
? ols C const P PG --quiet

# SCE del MR
? genr SCEMR = $ess
Se ha reemplazado el escalar SCEMR = 49.657

# Número de restricciones
? genr q = 1
Se ha reemplazado el escalar q = 1

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha reemplazado el escalar F = 300.584

? print SCEMR SCEMSR GLMSR q F
      SCEMR = 49.657019

      SCEMSR = 7.7984227

      GLMSR = 56.000000

      q = 1.0000000

      F = 300.58404
```

**- Contraste de nulidad individual para  $\beta_3$**

$$H_0: \beta_3 = 0 \quad \Rightarrow \quad \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \beta_3 \neq 0 \quad \Rightarrow \quad \text{MR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \varepsilon_t$$

```
# Estimación MCO del MR
? ols C const P RF --quiet

# SCE del MR
? genr SCEMR = $ess
Se ha reemplazado el escalar SCEMR = 7.84925

# Número de restricciones
? genr q = 1
Se ha reemplazado el escalar q = 1

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha reemplazado el escalar F = 0.364972

? print SCEMR SCEMSR GLMSR q F
      SCEMR = 7.8492479

      SCEMSR = 7.7984227

      GLMSR = 56.000000

      q = 1.0000000

      F = 0.36497226
```

**- Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas del modelo**

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \quad \Rightarrow \quad \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \text{alguno ó todos} \neq 0 \quad \Rightarrow \quad \text{MR} \rightarrow C_t = \beta_0$$

```
# Estimación MCO del MR
? ols C const --quiet

# SCE del MR
? genr SCEMR = $ess
Se ha reemplazado el escalar SCEMR = 131.65

# Número de restricciones
? genr q = 3
Se ha reemplazado el escalar q = 3

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha reemplazado el escalar F = 296.457
```

```
? print SCEMR SCEMSR GLMSR q F

      SCEMR = 131.65000
      SCEMSR = 7.7984227

      GLMSR = 56.000000

      q = 3.0000000

      F = 296.45688
```

**- Contraste de nulidad conjunta para un subconjunto paramétrico**

$$H_0: \beta_1 = \beta_3 = 0 \quad \Rightarrow \quad \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \text{alguno ó ambos} \neq 0 \quad \Rightarrow \quad \text{MR} \rightarrow C_t = \beta_0 + \beta_2 RF_t + \varepsilon_t$$

```
# Estimación MCO del MR
? ols C const PG --quiet

# SCE del MR
? genr SCEMR = $ess
Se ha reemplazado el escalar SCEMR = 131.379

# Número de restricciones
? genr q = 2
Se ha reemplazado el escalar q = 2

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha reemplazado el escalar F = 443.712

? print SCEMR SCEMSR GLMSR q F

      SCEMR = 131.37877
      SCEMSR = 7.7984227

      GLMSR = 56.000000

      q = 2.0000000

      F = 443.71150
```

**- Contraste de nulidad para una combinación lineal**

$$H_0: \beta_1 = \beta_3 \quad \Rightarrow \quad \text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

$$H_1: \beta_1 \neq \beta_3 \quad \Rightarrow \quad \text{MR} \rightarrow C_t = \beta_0 + \beta_1 (P_t + PG_t) + \beta_2 RF_t + \varepsilon_t$$

```
? genr PMPG = P + PG
Se ha generado la serie PMPG (ID 7)

# Estimación MCO del MR
? ols C const PMPG RF --quiet
# SCE del MR

? genr SCEMR = $ess
Se ha reemplazado el escalar SCEMR = 14.7584

# Número de restricciones
? genr q = 1
Se ha reemplazado el escalar q = 1

? genr F = ((SCEMR - SCEMSR)/q)/(SCEMSR/GLMSR)
Se ha reemplazado el escalar F = 49.9789

? print SCEMR SCEMSR GLMSR q F

      SCEMR = 14.758364
      SCEMSR = 7.7984227

      GLMSR = 56.000000
```

```
q = 1.0000000
```

```
F = 49.978915
```

#### 4.6.11. Contrastes de nulidad utilizando del comando omit

Una forma alternativa de realizar los contrastes de nulidad para los parámetros, es utilizar la opción **Omitir variables** del menú **Contrastes** de la **Ventana Modelo** o en la **Cónsola de Gretl**, con el comando **omit**. Aunque no es necesario que el comando **omit** se ejecute inmediatamente después de un comando **ols**, resulta conveniente, pues de esta forma el usuario se asegura que el contraste hace referencia al modelo deseado.

```
ols depvar indepvars --quiet
```

```
omit indepvars a omitir --quiet
```

En este caso, el modelo al que se deben referir los contrastes de nulidad es el **modelo sin restringir**, por lo que será el mismo para todos los contrastes:

$$\text{MSR} \rightarrow C_t = \beta_0 + \beta_1 P_t + \beta_2 RF_t + \beta_3 PG_t + \varepsilon_t$$

El modelo con restricciones se obtiene sustituyendo éstas en el modelo sin restricciones.

```
# Contraste de nulidad individual para  $\beta_0$ 
? ols C const P RF PG --quiet
? omit const --quiet

Hipótesis nula: el parámetro de regresión es cero para const
Estadístico de contraste: F(1, 56) = 154.556, con valor p = 9.6678e-018

# Contraste de nulidad individual para  $\beta_1$ 
? ols C const P RF PG --quiet
? omit P --quiet

Hipótesis nula: el parámetro de regresión es cero para P
Estadístico de contraste: F(1, 56) = 443.021, con valor p = 2.83627e-028

# Contraste de nulidad individual para  $\beta_2$ 
? ols C const P RF PG --quiet
? omit RF --quiet

Hipótesis nula: el parámetro de regresión es cero para RF
Estadístico de contraste: F(1, 56) = 300.584, con valor p = 3.55102e-024

# Contraste de nulidad individual para  $\beta_3$ 
? ols C const P RF PG --quiet
? omit PG --quiet

Hipótesis nula: el parámetro de regresión es cero para PG
Estadístico de contraste: F(1, 56) = 0.364972, con valor p = 0.548197

# Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas
? ols C const P RF PG --quiet
? omit P RF PG --quiet

Hipótesis nula: los parámetros de regresión son cero para las variables
P, RF, PG

Estadístico de contraste: F(3, 56) = 296.457, con valor p = 2.51999e-034

# Contrastes de nulidad conjunta para un subconjunto paramétrico
? ols C const P RF PG --quiet
? omit P RF --quiet

Hipótesis nula: los parámetros de regresión son cero para las variables
P, RF

Estadístico de contraste: F(2, 56) = 443.712, con valor p = 4.54448e-035
```

#### 4.6.12. Contrastes de nulidad utilizando el comando add

Una forma alternativa de realizar los contrastes de nulidad de los parámetros es utilizar la opción **Añadir variables** del menú **Contrastes** de la **Ventana Modelo** o en la **Cónsola de Gretl**, con el comando **add**. Aunque no es necesario que el comando **add** se ejecute inmediatamente después de un comando **ols**, resulta conveniente, pues de esta forma el usuario se asegura que el contraste hace referencia al modelo deseado.

```
ols depvar indepvars --quiet
add indepvars a añadir --quiet
```

En este caso, el modelo al que se deben referir los contrastes de nulidad es el **modelo restringido**, por lo que será diferente para cada contraste. Por tanto, se parte del **modelo restringido** y añadiendo las “*nuevas variable independientes*” se debe llegar al **modelo sin restringir**.

```
# Contraste de nulidad individual para  $\beta_0$ 
? ols C P RF PG --quiet
? add const --quiet

Hipótesis nula: el parámetro de regresión es cero para const
Estadístico de contraste: F(1, 56) = 154.556, con valor p = 9.6678e-018

# Contraste de nulidad individual para  $\beta_1$ 
? ols C const RF PG --quiet
? add P --quiet

Hipótesis nula: el parámetro de regresión es cero para P
Estadístico de contraste: F(1, 56) = 443.021, con valor p = 2.83627e-028

# Contraste de nulidad individual para  $\beta_2$ 
? ols C const P PG --quiet
? add RF --quiet

Hipótesis nula: el parámetro de regresión es cero para RF
Estadístico de contraste: F(1, 56) = 300.584, con valor p = 3.55102e-024

# Contraste de nulidad individual para  $\beta_3$ 
? ols C const P RF --quiet
? add PG --quiet

Hipótesis nula: el parámetro de regresión es cero para PG
Estadístico de contraste: F(1, 56) = 0.364972, con valor p = 0.548197

# Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas
? ols C const --quiet
? add P RF PG --quiet

Hipótesis nula: los parámetros de regresión son cero para las variables
P, RF, PG

Estadístico de contraste: F(3, 56) = 296.457, con valor p = 2.51999e-034

# Contrastes de nulidad conjunta para un subconjunto paramétrico
? ols C const PG --quiet
? add P RF --quiet

Hipótesis nula: los parámetros de regresión son cero para las variables
P, RF

Estadístico de contraste: F(2, 56) = 443.712, con valor p = 4.54448e-035
```

#### 4.6.13. Contrastes de nulidad utilizando el comando coeffsum

Una forma alternativa de realizar ciertos contrastes de nulidad para una combinación lineal de parámetros es utilizar la opción **Suma de coeficientes** del menú **Contrastes** de la **Ventana Modelo** o en la **Cónsola de Gretl**, con el comando **coeffsum**. Aunque no es necesario que el comando **coeffsum**

se ejecute inmediatamente después de un comando **ols**, resulta conveniente pues de esta forma, el usuario se asegura que el contraste hace referencia al modelo deseado.

**ols depvar indepvars --quiet**  
**coeffsum indepvars cuyos coeficientes se suman --quiet**

Se debe tener en cuenta que el número de variables independientes que deben aparecer en la instrucción **coeffsum** y, por tanto, el número de parámetros que aparecerán en la restricción, dependerá de la combinación lineal cuya nulidad se quiera contrastar.

La limitación de este comando es que sólo permite contrastar la nulidad de combinaciones lineales del tipo  $\beta_1 + \dots + \beta_j = 0$ .

A continuación se muestra un contraste realizado con el comando **coeffsum** y su equivalente con el comando **restrict**:

```
# Contraste para ver si las variables P y PG tienen una influencia similar de signo contrario sobre C
? ols C const P RF PG --quiet
? coeffsum P PG

Variables: P PG
Suma de los coeficientes = -4.1618
Desviación típica = 0.571835
t(56) = -7.27797 con valor p = 1.20157e-009

# Contraste a través del comando restrict
? ols C const P RF PG --quiet
? restrict --quiet
? b[P] + b[PG] = 0
? end restrict
Restricción:
b[P] + b[PG] = 0

Estadístico de contraste: F(1, 56) = 52.9689, con valor p = 1.20157e-009

? genr tsum=sqrt(52.9689)
Se ha generado el escalar tsum = 7.27797
```

#### 4.6.14. Contrastes de no nulidad

Aún cuando los contrastes más habituales en Econometría son los de nulidad de los parámetros<sup>43</sup>, algunos de los cuales se han analizado en los epígrafes anteriores utilizando distintas alternativas, en algunas ocasiones puede interesar realizar contrastes de no nulidad. Si estos contrastes se realizan con el comando **restrict**, a diferencia de los contrastes de nulidad, el número que aparece a la derecha en las ecuaciones de las restricciones lineales deja de ser cero (al menos, en alguna/s).

Algunos ejemplos:

- **Contraste de no nulidad individual para un parámetro**

$$H_0: \beta_1 = -0.5 \quad \Rightarrow \quad F = \frac{(b_1 + 0.5)' [S_{b_1}^2]^{-1} (b_1 + 0.5)}{1} = \frac{(b_1 + 0.5)^2}{S_{b_1}^2}$$

$$H_1: \beta_1 \neq -0.5$$

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[P] = -0.5
? end restrict
Restricción:
b[P] = -0.5
```

<sup>43</sup> Contrastes de nulidad individual, de nulidad conjunta o de nulidad para una combinación lineal.

Estadístico de contraste:  $F(1, 56) = 335.581$ , con valor  $p = 2.5611e-025$

```
# Cálculo manual
? matrix Fb1 = (b[2]+0.5)^2*(Vb[2,2])^(-1)
Se ha generado la matriz Fb1

? print Fb1
Fb1 (1 x 1)

335.58
```

**- Contraste de no nulidad conjunta para un subconjunto paramétrico**

$$H_0: \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \Rightarrow F = \frac{\left[ \begin{pmatrix} b_1 \\ b_3 \end{pmatrix} - \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \right]' \begin{pmatrix} S_{b_1}^2 & S_{b_1 b_3} \\ S_{b_1 b_3} & S_{b_3}^2 \end{pmatrix}^{-1} \left[ \begin{pmatrix} b_1 \\ b_3 \end{pmatrix} - \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \right]}{2}$$

$$H_1: \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} \neq \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$$

```
? ols C const P RF PG --quiet
? restrict --quiet
? b[P] = -0.5
? b[PG] = 0.5
? end restrict

Conjunto de restricciones
1: b[P] = -0.5
2: b[PG] = 0.5

Estadístico de contraste:  $F(2, 56) = 170.094$ , con valor  $p = 1.61454e-024$ 

# Cálculo manual
? matrix bsub = {b[2]; b[4]}
Se ha generado la matriz bsub

? print bsub
bsub (2 x 1)
-3.8561
-0.30569

? matrix rbsub = {-0.5; 0.5}
Se ha generado la matriz rbsub

? print rbsub
rbsub (2 x 1)
-0.5
0.5

? matrix Vbsub = {Vb[2,2], Vb[2,4]; Vb[4,2], Vb[4,4]}
Se ha generado la matriz Vbsub

? print Vbsub
Vbsub (2 x 2)
0.033564 0.018695
0.018695 0.25604

? matrix Fbsub = (bsub - rbsub)' * inv(Vbsub) * (bsub - rbsub) / 2
Se ha generado la matriz Fbsub

? print Fbsub
Fbsub (1 x 1)
170.09
```

**- Contraste de no nulidad para una combinación lineal**

$$H_0: \beta_1 + 2\beta_3 = 1 \Rightarrow F = \frac{(p-1)^2}{S_p^2} \text{ donde } p = R'b \text{ y } S_p^2 = R'\hat{V}(b)R$$

$$H_1: \beta_1 + 2\beta_3 \neq 1$$

```
? ols C const P RF PG --quiet
? restrict --quiet
```

```
? b[P] + 2*b[PG] = 1
? end restrict

Restricción:
  b[P] + 2*b[PG] = 1

Estadístico de contraste: F(1, 56) = 26.3958, con valor p = 3.65037e-006

# Cálculo manual
? matrix p = b[2] + 2*b[4]
Se ha generado la matriz p

? matrix s2p = Vb[2,2] + 4*Vb[2,4] + 4*Vb[4,4]
Se ha generado la matriz s2p

? genr Fc1 = (p - 1)^2 / s2p
Se ha generado el escalar Fc1 = 26.3958
? print p s2p Fc1
p (1 x 1)
  -4.4675

s2p (1 x 1)
  1.1325

Fc1 = 26.395778
```