

# ÍNDICE

<b>Capítulo 1. INTRODUCCIÓN A SHAZAM PROFESSIONAL</b>	<b>1</b>
<b>1.1. Presentación de Shazam Professional</b>	<b>1</b>
<b>1.2. Inicio de una sesión de trabajo</b>	<b>1</b>
<b>1.3. Principales Ventanas</b>	<b>2</b>
1.3.1. Ventana Principal o Ventana del Programa	4
1.3.2. Panel Project-Resources	5
1.3.2.1. Ventana Proyecto o Project	5
1.3.2.2. Ventana Guías de Ayuda o Resources	7
1.3.3. Ventana Input o Editor de Comandos	7
1.3.4. Ventana Output o Ventana de Resultados	10
1.3.5. Panel de Depuración	11
<b>1.4. Configuración de Shazam Professional</b>	<b>13</b>
<b>1.5. Wizards</b>	<b>15</b>
<b>Capítulo 2. MANEJO Y ANÁLISIS DE DATOS</b>	<b>18</b>
<b>2.1. Presentación</b>	<b>18</b>
<b>2.2. Introducción de datos</b>	<b>18</b>
2.2.1. Introducción de datos directamente en el programa	18
2.2.1.1. ¿Cómo dar nombre a las variables?	19
2.2.1.2. ¿Cómo introducir datos?	20
2.2.1.3. ¿Cómo guardar datos?	21
2.2.2. Recuperación de datos desde un fichero	22
2.2.2.1. Ficheros de datos de Shazam	22
2.2.2.2. Ficheros de datos de Excel	22
2.2.3. Recuperación de datos desde una Base de Datos	24
<b>2.3. ¿Cómo generar nuevas variables?</b>	<b>25</b>
2.3.1. A través del Editor de Datos	25
2.3.2. A través del comando GENR	27
<b>2.4. ¿Cómo calcular los estadísticos descriptivos de las variables?</b>	<b>28</b>
2.4.1. A través del WIZARD	28
2.4.2. A través del comando STAT	32
<b>2.5. ¿Cómo hacer representaciones gráficas?</b>	<b>34</b>
<b>Capítulo 3. ÁLGEBRA MATRICIAL</b>	<b>37</b>
<b>3.1. Presentación</b>	<b>37</b>
<b>3.2. Definición de matrices</b>	<b>37</b>
3.2.1. Utilizando el Editor de Matrices	37
3.2.2. Utilizando el comando READ	38
3.2.3. Utilizando el comando COPY	39
<b>3.3. Operaciones con matrices: el comando MATRIX</b>	<b>42</b>
3.3.1. Tipos de matrices	43
3.3.2. Matriz traspuesta	45
3.3.3. Igualdad de matrices	45
3.3.4. Suma y resta de matrices	46
3.3.5. Producto de matrices	47
3.3.5.1. Aplicaciones del producto de matrices	49
3.3.6. Determinante de una matriz	50
3.3.7. Rango de una matriz	52
3.3.8. Matriz inversa	53
3.3.9. Traza de un matriz	54
<b>Capítulo 4. ESTIMACIÓN MCO: MODELO DE REGRESIÓN LINEAL CLÁSICO</b>	<b>56</b>

<b>4.1.</b>	<b>Presentación e hipótesis básicas del Modelo de Regresión Lineal Múltiple</b>	<b>56</b>
<b>4.2.</b>	<b>¿Cómo estimar por MCO en Shazam?</b>	<b>58</b>
4.2.1.	A través del WIZARD	58
4.2.2.	A través del comando OLS	62
<b>4.3.</b>	<b>Análisis de la información básica proporcionada por un comando OLS</b>	<b>63</b>
<b>4.4.</b>	<b>Análisis de los residuos: opción LIST</b>	<b>66</b>
<b>4.5.</b>	<b>Análisis de las sumas de cuadrados: opción ANOVA</b>	<b>67</b>
4.5.1.	Tabla ANOVA respecto a la media	68
4.5.2.	Tabla ANOVA respecto al cero	69
<b>4.6.</b>	<b>Análisis gráfico: opción GRAPH</b>	<b>69</b>
<b>4.7.</b>	<b>Estimación MCO de un modelo formulado sin ordenada en el origen</b>	<b>71</b>
<b>4.8.</b>	<b>Interpretación de los coeficientes</b>	<b>73</b>
<b>4.9.</b>	<b>Formas funcionales alternativas</b>	<b>75</b>
<b>Capítulo 5. CONTRASTES DE HIPÓTESIS Y REGIONES DE CONFIANZA</b>		<b>79</b>
<b>5.1.</b>	<b>Hipótesis del Modelo de Regresión Lineal Normal Clásico (MRLNC)</b>	<b>79</b>
<b>5.2.</b>	<b>Contrastes de hipótesis</b>	<b>79</b>
5.2.1.	Contraste de nulidad individual o contraste de nulidad para un parámetro	80
5.2.2.	Contraste de nulidad conjunta para todos los parámetros del modelo	81
5.2.3.	Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas del modelo	82
5.2.4.	Contraste de nulidad conjunta para un subconjunto paramétrico	83
5.2.5.	Contraste de nulidad para una combinación lineal	83
<b>5.3.</b>	<b>¿Cómo realizar contrastes de hipótesis en Shazam?</b>	<b>84</b>
5.3.1.	A través del WIZARD	84
5.3.2.	A través del comando TEST	86
<b>5.4.</b>	<b>¿Cómo buscar el P-valor y el valor crítico en una distribución prefijada?</b>	<b>88</b>
5.4.1.	A través del WIZARD	88
5.4.2.	A través del comando DISTRIB	91
5.4.3.	¿Cómo utilizar el comando DISTRIB en el contraste de hipótesis?	91
<b>5.5.</b>	<b>Regiones de Confianza</b>	<b>93</b>
5.5.1.	¿Cómo calcular intervalos de confianza en Shazam?	93
5.5.1.1.	A través del WIZARD	93
5.5.1.2.	A través del comando CONFID	95
5.5.2.	¿Cómo utilizar el intervalo de confianza en el contraste de hipótesis?	96

## Ilustraciones

- Ilustración 1-1. Formas de iniciar Shazam.  
Ilustración 1-2. Selección modo inicio programa.  
Ilustración 1-3. Principales Ventanas de Shazam Professional.  
Ilustración 1-4. Menú Principal.  
Ilustración 1-5. Barra de Tareas Principal.  
Ilustración 1-6. Ventanas del Panel Project-Resources.  
Ilustración 1-7. Ventana Input.  
Ilustración 1-8. Ventanas Output.  
Ilustración 1-9. Panel de Depuración.  
Ilustración 1-10. Posibilidades de configuración del programa antes de iniciar una sesión de trabajo.  
Ilustración 1-11. Secuencia de cuadros de diálogo desde que se accede a un Wizard hasta que la instrucción aparece en el Editor de Comandos correspondiente.
- Ilustración 2-1. Dimensión de la base de datos.  
Ilustración 2-2. Ventana Editor de Datos o Data Editor.  
Ilustración 2-3. Cuadro de diálogo para dar nombre o renombrar variables.  
Ilustración 2-4. Introducción de datos.  
Ilustración 2-5. Tipos de ficheros de datos.  
Ilustración 2-6. Procedimiento para añadir un fichero a un Project y para autocargarlo automáticamente.  
Ilustración 2-7. Procedimiento para recuperar datos de un fichero de excel.  
Ilustración 2-8. Procedimiento para la recuperación de datos de una Base de Datos.  
Ilustración 2-9. Procedimiento para generar nuevas variables dentro de una base de datos.  
Ilustración 2-10. Generación de variables para un rango menor al muestral.  
Ilustración 2-11. Secuencia de cuadros de diálogo del wizard “**Write or Print Active Variables**”.  
Ilustración 2-12. Secuencia de cuadros de diálogo del wizard “**Descriptive Statistics**”.  
Ilustración 2-13. Opciones del comando STAT que se pueden insertar a través de su Wizard.  
Ilustración 2-14. Secuencia de cuadros de diálogo del wizard “**Create Graph**”.  
Ilustración 2-15. Ventana Gráfico o Shazam Graph.
- Ilustración 3-1. Dimensión de la matriz.  
Ilustración 3-2. Ventana Editor de Matrices.  
Ilustración 3-3. Extensión de los ficheros de matrices.  
Ilustración 3-4. Pasos a seguir para definir una matriz a través del Editor de Comandos.
- Ilustración 4-1. Secuencia de cuadros del Wizard “**Ordinary Least Square Regression**”.  
Ilustración 4-2. Selección de variables para usar en la estimación.  
Ilustración 4-3. Opciones del comando OLS.  
Ilustración 4-4. Opciones para guardar algunos resultados de la estimación.  
Ilustración 4-5. Gráficos asociados a la opción GRAPH del comando OLS.  
Ilustración 4-7. Gráficos asociados a la opción GRAPH del comando OLS modificados a partir de su Ventana Gráfico.  
Ilustración 4-6. Gráfico de residuos con bandas de confianza del 95%.  
Ilustración 4-8. Gráfico de residuos con Excel.  
Ilustración 4-9. Transformaciones lineales.
- Ilustración 5-1. Secuencia de cuadros de diálogo del wizard “**Hypothesis Tests**”.  
Ilustración 5-2. Secuencia de cuadros de diálogo del wizard “**Probability Distributions**”.  
Ilustración 5-3. Opciones del comando DISTRIB que se pueden insertar a través de su Wizard.  
Ilustración 5-4. Ejemplo de cómo utilizar el comando DISTRIB para buscar los valores críticos en los distintos tipos de contrastes.  
Ilustración 5-5. Ejemplo de cómo utilizar el comando DISTRIB para buscar P-valores en los distintos tipos de contrastes.  
Ilustración 5-6. Secuencia de cuadros de diálogo del wizard “**Confidence Intervals**”.  
Ilustración 5-7. Opciones del comando CONFID que se pueden insertar a través de su Wizard.  
Ilustración 5-8. Ejemplo de la información que proporciona la salida de un comando CONFID en el que intervienen dos parámetros.

## Cuadros de Texto

Cuadro 2-1. Funciones y operadores del comando GENR.

Cuadro 2-2. Descripción opciones del comando STAT.

Cuadro 3-1. Funciones y operadores del comando MATRIX.

Cuadro 4-1. Descripción opciones comando OLS.

Cuadro 4-2. Opciones transformaciones lineales.

Cuadro 5-1. Salida de un comando TEST.

Cuadro 5-2. Descripción opciones comando DISTRIB.

## Tablas

Tabla 4-1. Salida básica de la estimación MCO de un modelo formulado con ordenada en el origen

Tabla 4-2. Salida estándar asociada a la opción LIST

Tabla 4-3. Salida estándar asociada a la opción ANOVA

Tabla 4-4. Salida básica de la estimación MCO de un modelo formulado sin ordenada en el origen

# Capítulo 1. INTRODUCCIÓN A SHAZAM PROFESSIONAL

## 1.1. Presentación de Shazam Professional

Shazam Professional Edition 10.1 es una versión para Windows de un conjunto de herramientas diseñadas originalmente para Shazam Standard.

La "ventaja" de la edición Professional de Shazam 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.

Shazam Professional es un paquete relativamente fácil de usar, con una gran flexibilidad y con grandes posibilidades en el manejo de datos, incluyendo operaciones de álgebra matricial. Permite la estimación y el contraste de multitud de modelos econométricos, siendo posible el establecimiento de rutinas de programación, que facilitarán en gran medida el trabajo del usuario avanzado.

## 1.2. Inicio de una sesión de trabajo

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

- Utilizar un acceso directo al programa, que previamente se habrá creado en el Escritorio de Windows. Si se dispone de dicho acceso, tan sólo, se tendrá que hacer un doble clic en el mismo, siendo esta la alternativa más cómoda para iniciar una sesión de trabajo.
- Desplegar el menú de Inicio de Windows y acceder a Programas → Shazam → Professional Edition → Shazam Professional Edition.
- Desde el Escritorio de Windows acceder al ejecutable *SHAZAMP.exe*.

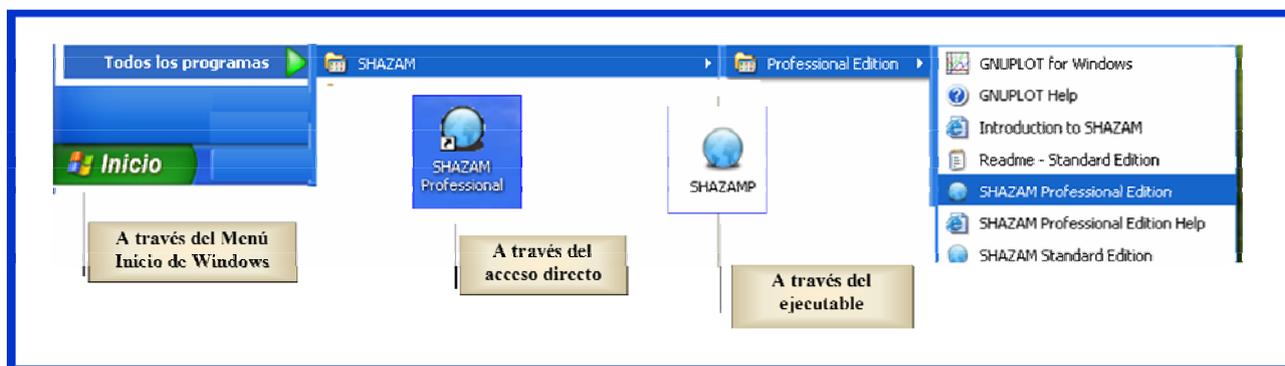


Ilustración 1-1. Formas de iniciar Shazam.

Cualquiera de estas opciones permite entrar en el programa e iniciar una sesión de trabajo. Durante unos segundos, la pantalla inicial de Shazam informa de la versión del programa que se acaba de iniciar.

A continuación se abre un cuadro de diálogo que permite seleccionar el modo en que se inicia la sesión de trabajo (véase Ilustración 1-2):

- Múltiples ventanas:
  - Ventanas en cascada.
  - Ventanas en vertical.
  - Ventanas en horizontal.
- Ventana única.

A pesar de que el programa permite elegir entre trabajar con una o múltiples ventanas, hay que señalar que es en la modalidad de múltiples ventanas donde la versión Professional de Shazam proporciona mayores ventajas. La principal diferencia entre trabajar en el “**Modo Ventana Única**” y “**Modo Ventanas Múltiples**” radica en que en este último, los comandos y el output correspondiente a su ejecución aparecen en ventanas separadas. En este caso, se ha optado por iniciar la sesión de trabajo en el “**Modo Ventanas Múltiples**” en formato vertical. Hay que señalar que, tanto en el modo vertical como horizontal, las pantallas se visualizan simultáneamente, mientras que en la modalidad de cascada, las ventanas aparecen superpuestas.

Este cuadro de diálogo también permite que la selección efectuada sea la que asuma Shazam por defecto cada vez que se inicie una sesión de trabajo.



Ilustración 1-2. Selección modo inicio programa.

### 1.3. Principales Ventanas

Cinco son las ventanas principales de Shazam (véase Ilustración 1-3), todas ellas interrelacionadas en mayor o menor medida y que podrán ser minimizadas o cerradas, según interese en cada momento:

1. **Ventana Principal** o **Ventana del Programa**.
2. **Panel Proyecto-Recursos: Ventana Proyecto o Project** (Project View) / **Ventana Recursos o Guías de Ayuda** (Resources).
3. **Ventana Input** o **Editor de Comandos** (Command Editor).
4. **Ventana Output** o **Ventana de Resultados** (Raw Output).
5. **Panel de Depuración: Ventana Mensajes** (Messages) / **Ventana Comandos Inmediatos** (Immediate) / **Ventana Variables Observadas** (Watch) / **Ventana Variables Temporales** (Temp).

En estas ventanas, se pueden distinguir algunos de los siguientes elementos:

- La **Barra de Título**, que da nombre a la ventana y que se oscurece al situar el cursor en ella, indicando que es la ventana operativa en ese momento. En la mayoría de las ventanas en la parte derecha de la barra de título se sitúan los iconos para **Minimizar**, **Restaurar** o **Cerrar** dicha ventana.
- La **Barra de Menú**, que contiene los botones que dan acceso al menú principal de la ventana.

Al situar el cursor encima de cada uno de los botones de la barra de menú se abre el submenú correspondiente, que puede no estar totalmente disponible, dependiendo de la ventana que se encuentre activa y de la tarea que se esté realizando, por ello, en los distintos menús y submenús, los botones pueden aparecer escritos en oscuro o claro, dependiendo de si están disponibles o no en ese momento. Además, si a algún botón le sigue una punta de flecha, ello implica un nuevo submenú asociado al mismo. Hay que señalar que la **Barra de Menú**, tan sólo, está disponible en la **Ventana Principal** o **Ventana del Programa**.

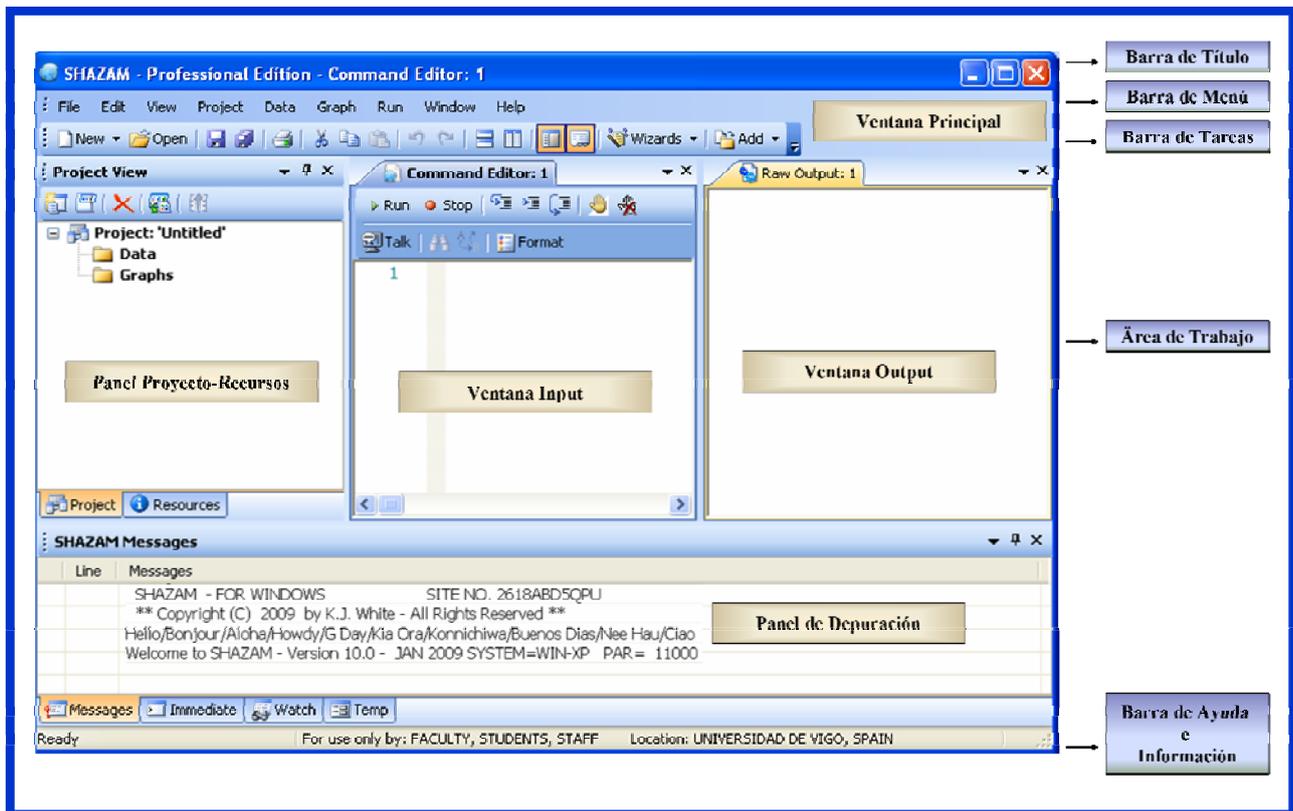


Ilustración 1-3. Principales Ventanas de Shazam Professional.

Como el programa se ejecuta en entorno Windows, la combinación de la tecla Ctrl y una letra es un método abreviado que permite activar determinadas tareas. Cuando estas combinaciones sean posibles, aparecerán indicadas a la derecha del botón correspondiente.

- La **Barra de Tareas**, que contiene los iconos que permiten acceder a las tareas más habituales de una forma más operativa, es decir, sin necesidad de desplegar menús y submenús, siendo su funcionamiento similar al de la barra de menú. Se debe señalar que la barra de tareas de la Ventana Principal puede ser personalizada por el usuario de acuerdo con sus necesidades, para lo cual debe hacer clic en la última de las puntas de flecha de dicha barra (**toolbar options**) y realizar la selección que considere oportuna.
- El **Área de Trabajo**, zona de color blanco que permite la comunicación entre Shazam y el usuario.
- La **Barra de Ayuda e Información**, que informa y proporciona ayuda, de manera que cuando se sitúa el cursor en un botón del submenú del Menú Principal o en un icono de la barra de tareas de alguna de las ventanas, es el espacio utilizado para informar al usuario de la función de dicho botón o icono.

### 1.3.1. 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 disponibles en Shazam Professional.

En la parte izquierda de la **Barra de Título** de esta ventana aparecen el nombre del programa y el de la ventana que está activa en ese momento. Por ejemplo, en la Ilustración 1-4 se aprecia que únicamente está abierta la **Ventana del Programa** puesto que, tan sólo, aparece el nombre del programa en la **Barra de Título**.

En la parte derecha de la **Barra de Título** se sitúan los iconos para **Minimizar**, **Restaurar** o **Cerrar** la **Ventana Principal**. El usuario debe ser consciente de que cerrar esta ventana significa salir del programa y, por tanto, finalizar la sesión de trabajo.

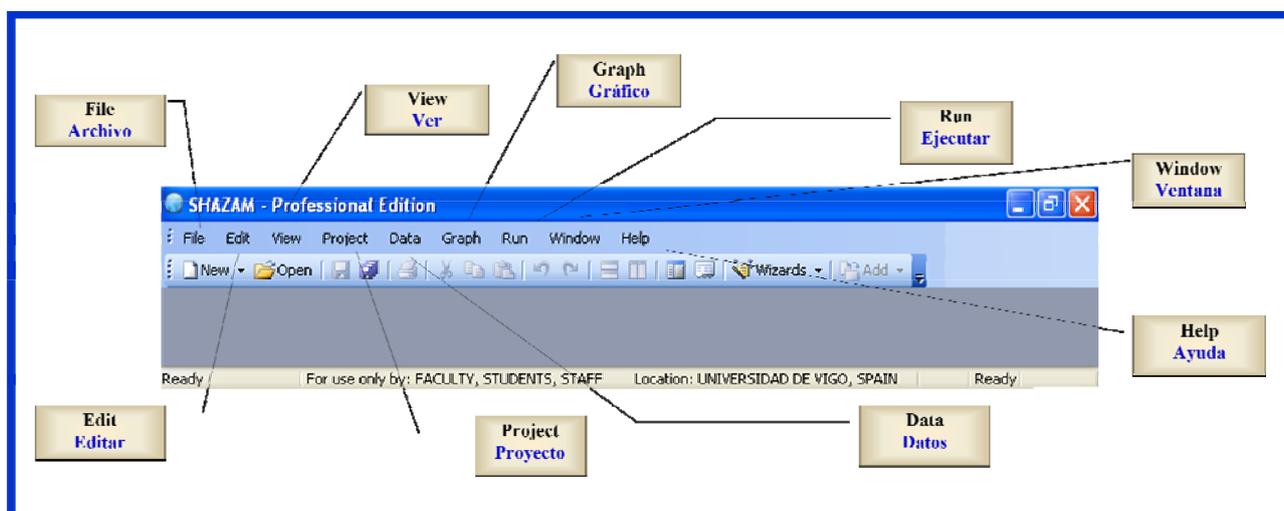


Ilustración 1-4. Menú Principal.

Debajo de la barra de título se encuentra la **Barra de Menú**, que contiene las distintas opciones del **Menú Principal** de Shazam Professional. Situando el cursor encima de cada una de estas opciones de la barra de menú se despliega el submenú correspondiente. A través del **Menú Principal**, el usuario puede acceder a las tareas típicas de cualquier programa que funcione en un entorno **Windows**: abrir, cerrar y guardar archivos; imprimir; cortar, copiar y pegar; buscar y/o reemplazar un texto concreto; etc. Además, permite crear gráficos, ejecutar comandos, crear y modificar bases de datos y, trabajar con proyectos.

Situada debajo de la **Barra de Menú** se encuentra la **Barra de Tareas**, que permite acceder a las tareas más habituales de una forma mucho más operativa, es decir, sin necesidad de desplegar los submenús del **Menú Principal**, basta con hacer clic en el icono correspondiente.

Como se puede ver en la Ilustración 1-5, existen algunos iconos (**New**, **Wizards**) que poseen una punta de flecha situada a la derecha que permite si se hace un clic sobre la misma acceder al submenú correspondiente que será comentado con detalle cuando se aborden tareas específicas. Además, en esta Ilustración 1-5, se puede observar que algunos botones aparecen en oscuro y otros en claro, dependiendo si están disponibles o no en ese momento.

Debajo de la barra de tareas se distribuyen las restantes ventanas disponibles en Shazam (véase Ilustración 1-3).

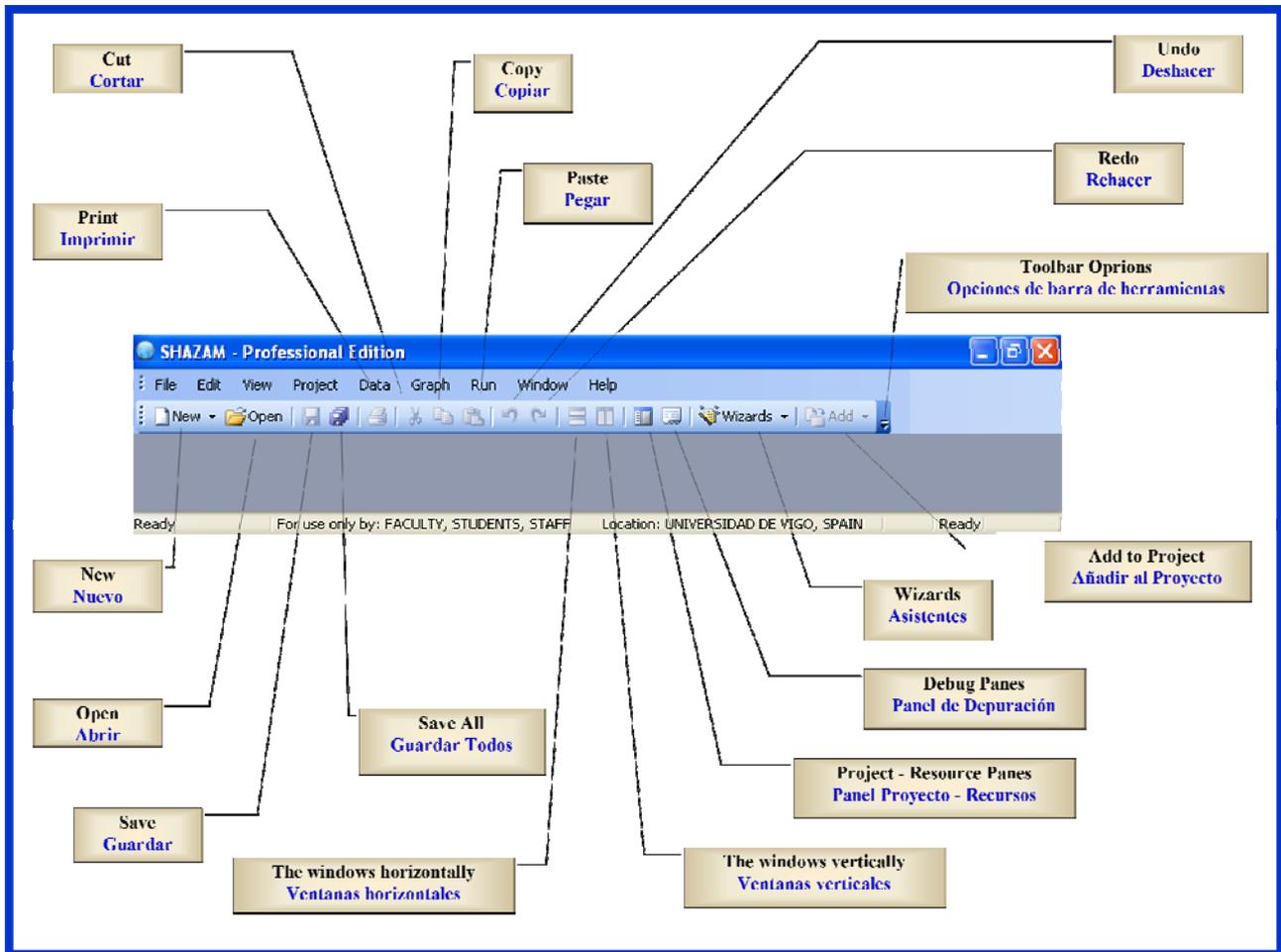


Ilustración 1-5. Barra de Tareas Principal.

Por último, en la parte inferior de la Ventana Principal se encuentra la **Barra de Ayuda e Información**. En la parte central de esta barra se indica los usuarios autorizados para la utilización del programa y su localización. Tanto en la parte izquierda como en la derecha se indica que Shazam está preparado para trabajar (**Ready**), con la particularidad de que cuando se sitúa el cursor en un botón del **Menú Principal** o en un icono de una barra de tareas, Shazam utiliza la parte izquierda para informar al usuario de la función de dicho botón o icono.

### 1.3.2. Panel Project-Resources

El Panel Project-Resources está formado por dos ventanas superpuestas: la **Ventana Project** que es la que abre Shazam de manera automática al iniciar una sesión de trabajo y la **Ventana Resources**, que se encuentra oculta tras la primera.

#### 1.3.2.1. Ventana Proyecto o Project

Como ya se ha comentado, la **Ventana Project** forma parte del denominado por Shazam **Panel Project-Resources** y, en el momento de iniciar una sesión de trabajo, Shazam la abre de manera automática.

A través de los iconos de la barra de tareas de la **Ventana Project** se puede acceder a las tareas más habituales a realizar con un **Project** de forma mucho más rápida y operativa. En la parte izquierda de la Ilustración 1-6 se puede observar que a través de esta barra de tareas se pueden crear y exportar proyectos, renombrar y eliminar items y, autocargar ficheros de datos<sup>1</sup>.

A diferencia de la versión Standard, la versión 10.1 de Shazam Professional permite una nueva manera de utilización del programa denominada "**Project**". Esta nueva forma de trabajo permite el manejo de forma conjunta de ficheros de comandos (\*.sha, \*.shz), ficheros de resultados (\*.out), ficheros de

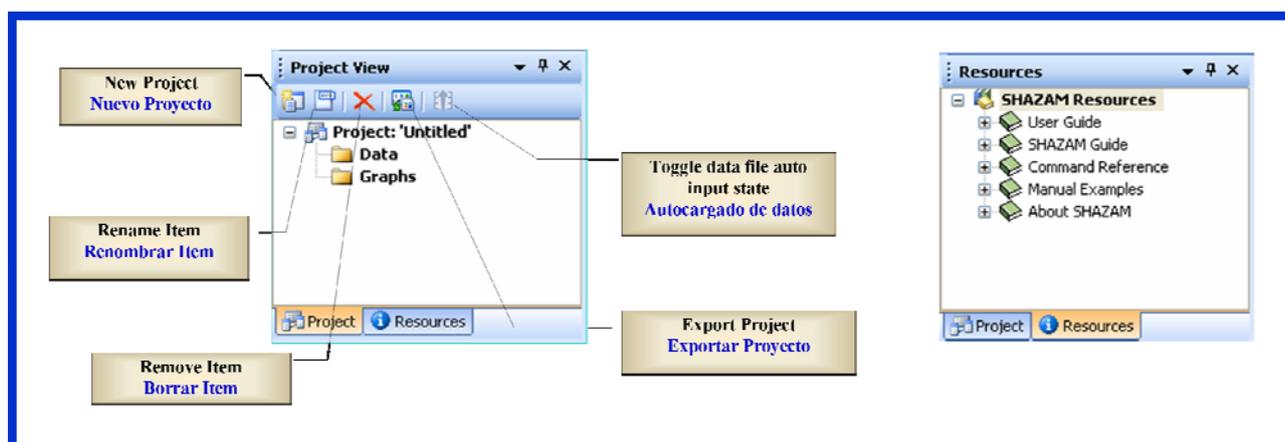


Ilustración 1-6. Ventanas del Panel Project-Resources.

gráficos (\*.gnu), ficheros de datos (\*.dat, \*.txt, \*.shd, \*.dta), ficheros de matrices (\*.mtx); ficheros de procedimiento (\*.pre); etc. A cada uno de los ficheros que forman parte de un Project, Shazam los denomina **item**. En los denominados "**ficheros project**" (\*.shp), Shazam guarda, no sólo, los enlaces con los diferentes items que lo componen, sino también toda la información interna que necesita para procesar dichos items. Eliminar<sup>2</sup> un item de un **Project** no implica eliminar dicho item, sino que dicho item dejará de formar parte del **Project** pero seguirá existiendo como fichero independiente.

Es conveniente que los diferentes items utilizados y/o generados por Shazam en una sesión de trabajo se vayan añadiendo a un **Project**, puesto que de esta forma se podrán visualizar conjuntamente en la **Ventana Project** y, se podrá acceder a su contenido y revisarlo fácilmente, sin más que hacer un doble clic sobre el item seleccionado. Además, cabe destacar que cuando se añaden determinados items a un **Project**, Shazam pone a disposición del usuario una serie de opciones no disponibles en caso contrario:

- Los ficheros de datos, creados o importados, se pueden seleccionar para que se autocarguen automáticamente<sup>3</sup> y, por tanto, no sea necesario utilizar un comando **READ**. Los ficheros de

<sup>1</sup> Otra forma en la que se puede acceder a estas mismas tareas es a través de la opción **Project** del **Menú Principal**. A través de esta opción del **Menú Principal**, el usuario además puede abrir y guardar Projects y, añadir nuevos items al Project corriente.

<sup>2</sup> Para eliminar un item de un **Project** se puede utilizar:

- El **Menú Principal** → **Project** → **Remove item**.
- El botón "**Remove**" de la barra de tareas de la **Ventana Project**
- La tecla "**Supr**".

<sup>3</sup> En una sesión de trabajo, el usuario puede tener seleccionado más de un fichero para que sus datos se autocarguen automáticamente. No obstante, debe ser cauteloso y procurar que en dichos ficheros no haya variables con el mismo nombre, puesto que en ese caso, Shazam establece un rango de preferencias y, trabajará, en lo que se refiere a las variables que posean el mismo nombre, con los datos cargados en último lugar, es decir, los que en la carpeta "**Data**" de la **Ventana Project** aparezcan más abajo.

datos aparecerán en la **Ventana Project** colgando de la carpeta “**Data**” y la denominación de dichos ficheros aparecerá precedida por una rejilla con o sin flecha que será lo que le indique a Shazam si debe cargar (rejilla con flecha) o no (rejilla sin flecha) automáticamente los datos de dichos ficheros en la sesión de trabajo. Para seleccionar o deseleccionar el autocargado de un fichero de datos basta con hacer clic sobre el nombre de dicho fichero y después, hacer clic en el botón de la “rejilla con flecha” (**Toogle data file auto input state**) de la barra de tareas de la **Ventana Project**.

- Los ficheros de datos creados por Shazam utilizando el comando **WRITE** se añaden de forma automática como item al **Project** y, por tanto, se pueden visualizar de forma inmediata en la **Ventana Project**.
- Los ficheros de gráficos creados por Shazam se añaden de forma automática como item al **Project**, una vez que éstos han sido guardados y, a partir de ese momento, se pueden visualizar en la **Ventana Project** colgando de la carpeta “**Graph**”. Tan sólo, si el usuario añade el gráfico al **Project**, se activarán los botones “**Properties**” y “**Edit file**” de la barra de tareas de la **Ventana Gráficos** y, el usuario podrá modificar tanto las propiedades de dichos gráficos como los datos utilizados en su construcción.

### 1.3.2.2. *Ventana Guías de Ayuda o Resources*

Como ya se ha comentado, la **Ventana Resources** forma parte del denominado por Shazam **Panel Project-Resources** y en el momento de iniciar una sesión de trabajo se encuentra oculta tras la **Ventana Project**. Si en la barra de información de este **Panel de Ventanas** se hace clic en el botón **Resources**, la **Ventana Project** da paso a la **Ventana Guías de Ayuda**. Esta ventana proporciona al usuario ayuda sobre el programa y sus comandos (véase parte derecha de la Ilustración 1-6).

### 1.3.3. **Ventana Input o Editor de Comandos**

El usuario utilizará la **Ventana Input** o **Editor de Comandos** para indicar a Shazam las instrucciones que desea ejecutar en la sesión de trabajo. Como se verá más adelante para escribir dichos comandos el usuario debe estar familiarizado con el lenguaje que utiliza Shazam o bien utilizar el asistente (**Wizard**) adecuado para su escritura, el cual mediante una secuencia de cuadros de diálogo, solicita al usuario que vaya haciendo las elecciones necesarias para la tarea que desee realizar. Utilizar los **Wizards** hace que la Versión Professional sea mucho más amigable que la Versión Standard porque permite trabajar en un entorno **Windows** y no es estrictamente necesario conocer las instrucciones y su formato para poder trabajar con el programa. Shazam permite guardar las instrucciones del **Editor de Comandos** en los ficheros de comandos (\*.sha), lo que facilita que dichos ficheros puedan ser abiertos, modificados y ejecutados, en sesiones posteriores.

La barra de tareas de la **Ventana Input** permite acceder a las tareas más habituales a realizar con los comandos, sin necesidad de utilizar el **Menú Principal** o la **Barra de Tareas Principal**. En la Ilustración 1-7 se puede observar que a través de esta barra de tareas, el usuario puede elegir la forma de ejecutar los comandos (modo interactivo, modo Batch, ...); parar el proceso; decidir que el output esté formateado; buscar y/o reemplazar texto; establecer líneas de ruptura; etc.

Existen tres maneras de ejecutar los comandos o instrucciones de una **Ventana Input** o **Editor de Comandos**:

- **Modo Batch**: todos los comandos se ejecutan de una sola vez. La forma más rápida de acceder a esta tarea es a través del botón “**Run**” de la barra de tareas del **Editor de Comandos** activo<sup>4</sup>.

---

<sup>4</sup> Seleccionando la opción “**Run**” del submenú “**Run**” del **Menú Principal** o presionando la tecla “**F5**”, también, se ejecutan todas las instrucciones de la **Ventana Input** activa en ese momento.

- **Modo Talk:** es un modo interactivo en el que Shazam responde de forma inmediata a cada una de las instrucciones que va recibiendo. A diferencia del “**Modo Batch**”, en el “**Modo Talk**”, las instrucciones se ejecutan de una en una.

Por defecto, el icono “**Talk**” de la barra de tareas de la **Ventana Input** no está operativo, de manera que si se presiona pasará a estar operativo (color naranja<sup>5</sup>) y, en esta situación, aparecerá una punta de flecha amarilla en la primera línea de comandos. Si se sitúa el cursor en esta primera línea y se presione la tecla “**Enter**”, Shazam ejecutará la instrucción contenida en la misma y pasará a situar el cursor en la siguiente línea, la cual se ejecutará cuando se presione de nuevo la tecla “**Enter**” y, así sucesivamente, hasta alcanzar la última línea de instrucciones<sup>6</sup>. Una vez alcanzada esta última línea, se pueden escribir nuevas instrucciones que se ejecutarán a medida que se vaya presionando la tecla “**Enter**”.

Este modo de trabajo interactivo tiene la ventaja de permitir al usuario analizar con detalle la salida correspondiente a cada comando y es aconsejable para los usuarios que utilizan por primera vez Shazam.

- **Por etapas o pasos:** en este caso los comandos o instrucciones se ejecutan línea a línea o por grupos:
  - a. **Run to cursor:** ejecuta todas las líneas de comandos, desde la primera hasta la línea donde está situado el cursor, incluyendo dicha instrucción.
  - b. **Step over to cursor:** ejecuta sólo la línea donde está situado el cursor.
  - c. **Step through line o Step into:** ejecuta los comandos línea a línea, empezando por la primera.

La forma más operativa de ejecutar los comandos por etapas es acceder a estas tareas activando el botón oportuno en la **Ventana Input**<sup>7</sup>.

---

<sup>5</sup> También se puede activar el “**Modo Talk**” a través del **Menú Principal: Run → Talk Mode**.

<sup>6</sup> Hay que señalar, que en el “**Modo Talk**” Shazam puede ejecutar cualquier comando independientemente de la línea que ocupe, bastará que se sitúe el cursor en la línea cuyo comando se desee ejecutar y se presione la tecla “**Enter**”. Se puede ir hacia delante o hacia atrás tantas veces como se desee, no obstante, esto puede ser contraproducente, sobre todo, si existen instrucciones interdependientes.

<sup>7</sup> También se puede acceder a dichas tareas a través del submenú de la opción “**Run**” del **Menú Principal** o bien presionando la tecla correspondiente (Step into → **F6**; Run to cursor → **F7**; Step over to cursor → **F8**).

Como ya se ha comentado todas las ventanas de Shazam están interrelacionadas en mayor o menor medida. Cuando se ejecutan comandos, Shazam utiliza la **Ventana Mensajes** para comunicar al usuario las incidencias ocurridas durante la ejecución. Utiliza esta ventana para informar al usuario de los posibles errores cometidos: explicita el número de línea que contiene la instrucción errónea, el tipo de error y, en algunos casos, advierte de lo que puede estar provocando dicho error. Si el usuario hace un clic sobre uno de estos mensajes, Shazam marcará en naranja en la **Ventana Input** correspondiente, la línea de comandos que ha dado lugar a dicho mensaje, lo que facilitará la tarea de rectificación del mismo.

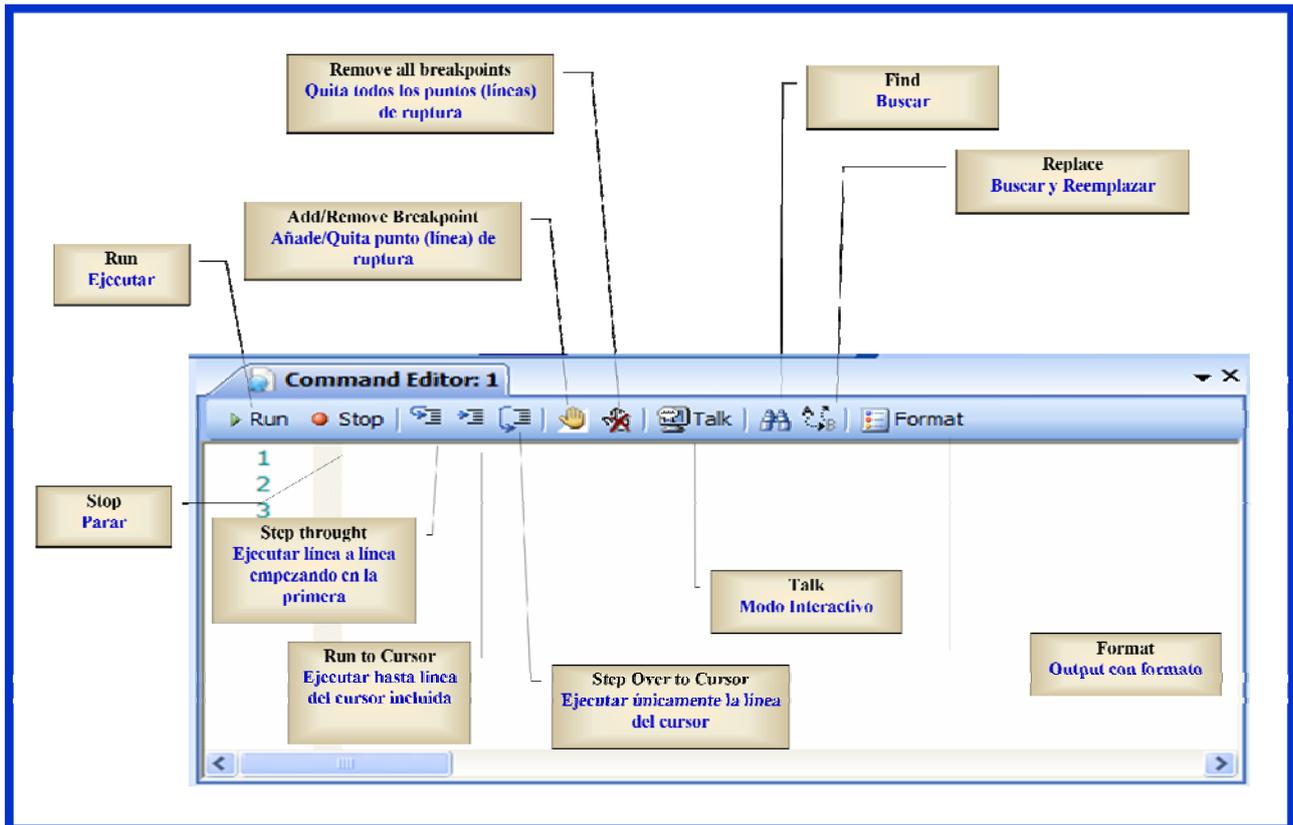


Ilustración 1-7. Ventana Input.

En determinadas ocasiones, el usuario puede estar interesado en establecer “**puntos o líneas de ruptura**” en la ejecución de un fichero de comandos. Esto lo puede hacer directamente a través de la barra de tareas de la **Ventana Input** o a través del **Menú Principal**. El establecimiento de dichos “**puntos o líneas de ruptura**”, permite que el usuario pueda observar el valor de las “**variables temporales**”<sup>8</sup> disponibles en la memoria interna de Shazam, hasta ese momento. Esta posibilidad resulta muy interesante para la depuración del fichero de comandos, sobre todo, si en dicho fichero existen bucles (**loops**).

Debe de tenerse en cuenta que en una sesión de trabajo puede estar disponible más de un **Editor de Comandos**, por lo que las ventanas correspondientes se irán enumerando de manera correlativa. Para abrir un nuevo editor de comandos se puede utilizar el **Menú Principal** seleccionando **File → New → Comand Editor** o se puede utilizar la **Barra de Tareas Principal** seleccionando el icono **New**.

<sup>8</sup> Cuando se ejecutan algunos comandos, Shazam guarda en memoria una serie de variables que reciben el nombre de “variables temporales” porque sólo están disponibles de forma temporal hasta que se ejecute un nuevo comando que lleve asociadas dichas variables. Son variables cuya denominación comienza por el símbolo \$.

### 1.3.4. Ventana Output o Ventana de Resultados

Toda **Ventana Output** está asociada a un **Editor de Comandos** y muestra los resultados relativos a la última ejecución (“**Run**”), es decir, cada vez que se pincha el icono **Run** de la barra de tareas del **Editor de Comandos**, se eliminan de la **Ventana Output** los resultados anteriores y se sustituyen por los nuevos (**Reset**).

Cuando no hay una **Ventana Output** asociada al **Editor de Comandos**, Shazam crea dicha ventana. De manera análoga, si se abre o crea más de un **Editor de Comandos**, la primera vez que se ejecutan, Shazam les asocia una **Ventana de Resultados** a cada uno de ellos, que también irán numerados de manera correlativa, coincidiendo dicha numeración con la de la **Ventana Input** correspondiente.

Hay que señalar que hay dos formas de ver los resultados: “sin formatear” y formateados. Shazam muestra los resultados sin formatear en la **Ventana Raw Output** y los resultados formateados en la **Ventana Formatted Output**. La **Ventana Raw Output** es la que Shazam muestra por defecto. Si el

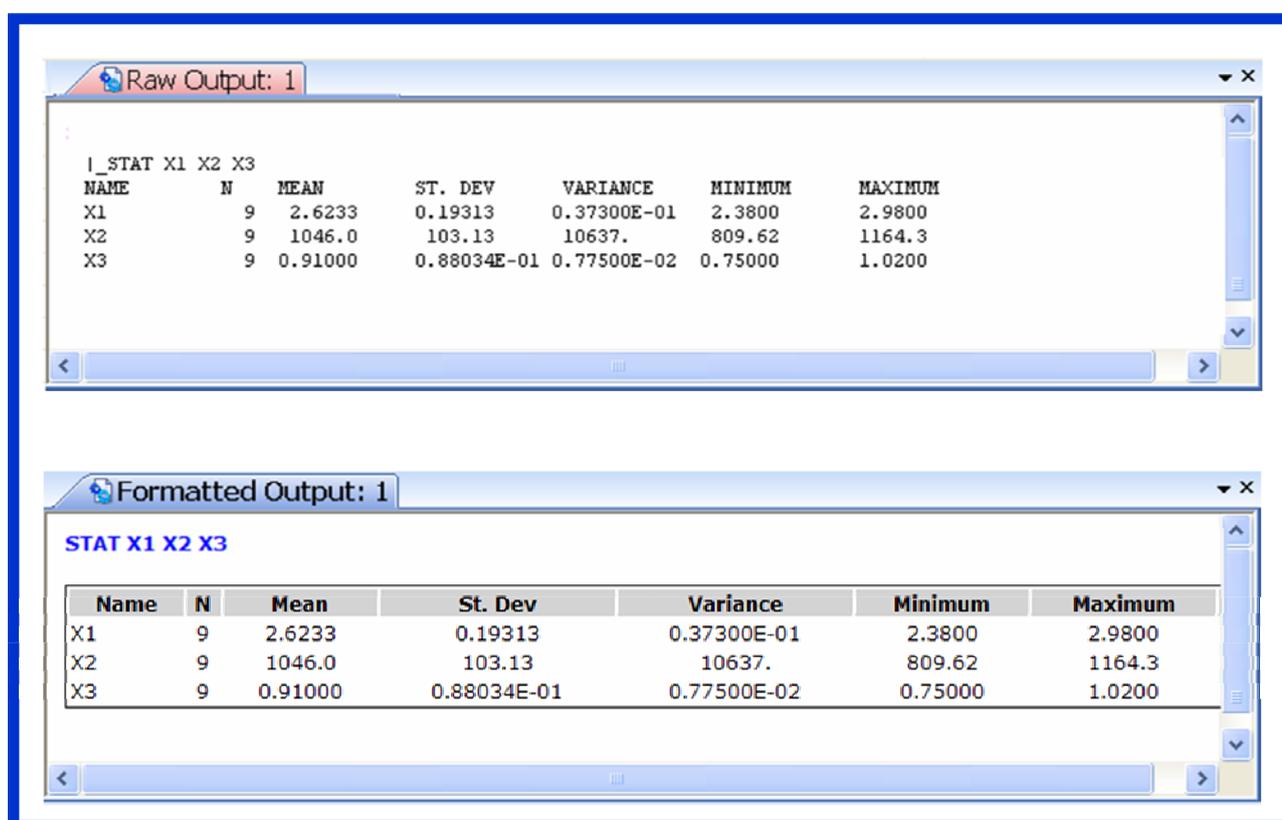


Ilustración 1-8. Ventanas Output.

usuario está interesado en obtener los resultados formateados, es decir, más estructurados, será necesario que en el momento de la ejecución tenga activado el icono **Format** del **Editor de Comandos**. Un ejemplo de la salida formateada y sin formatear puede verse en la Ilustración 1-8, donde se puede observar la diferencia de apariencia de la salida de un comando “**Stat**” estructurada en forma de tabla o sin estructurar.

### 1.3.5. Panel de Depuración

El Panel de Depuración ocupa la parte inferior de la Ventana Principal y consta de cuatro ventanas superpuestas (véase Ilustración 1-9): Ventana Mensajes, Ventana Comandos Inmediatos, Ventana Variables Observadas y Ventana Variables Temporales. La Ventana Mensajes es la que se abrirá por defecto al iniciar la sesión de trabajo.

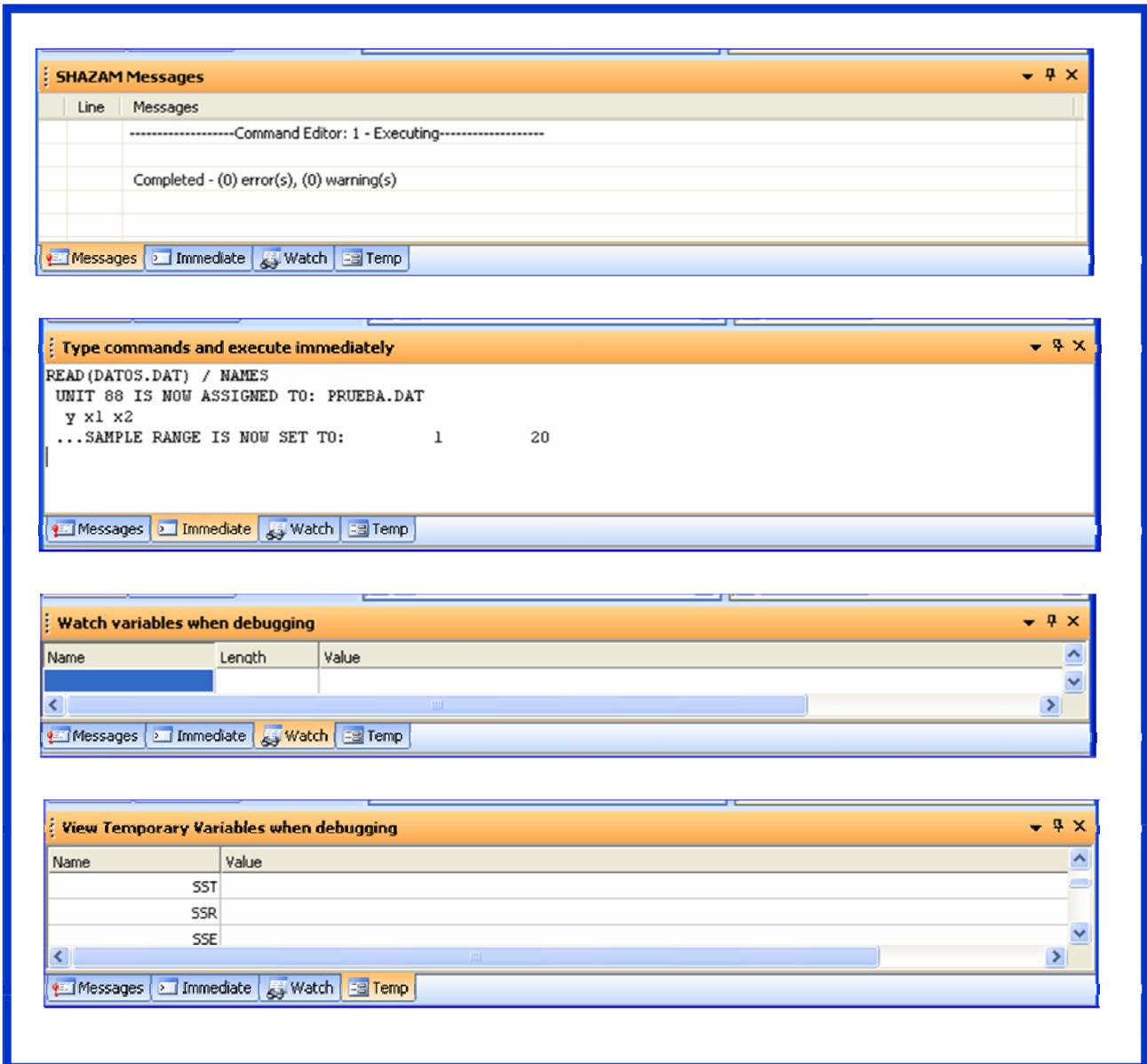


Ilustración 1-9. Panel de Depuración.

La **Ventana Mensajes** es el espacio que utiliza Shazam para informar tanto de los errores cometidos como de sus posibles causas. Los mensajes de error deben ser utilizados para chequear y rectificar las instrucciones correspondientes. Hay que tener en cuenta que Shazam no paraliza la ejecución de un fichero de comandos a pesar de encontrar errores, no obstante, los resultados de la ejecución de dichos comandos pueden no ser correctos por estar viciados por dichos errores. Por tanto, el fichero de comandos debe ser modificado hasta que su ejecución esté libre de estos errores. Para facilitar este chequeo y/o corrección, Shazam explicita el número de línea que contiene la instrucción errónea, el tipo de error y, en algunos casos, advierte de lo que puede estar provocando dicho error. Si el usuario hace clic sobre uno de estos mensajes, Shazam marcará en naranja en la **Ventana Input**

correspondiente, la línea de comandos que ha dado lugar a dicho mensaje, lo que facilitará la tarea de rectificación del mismo.

Los errores más frecuentes con los que el usuario se puede encontrar, pueden ser de cuatro tipos:

- **Formato incorrecto:** estos mensajes aparecen cuando se violan los formatos de escritura de las instrucciones. Todas las instrucciones de Shazam tienen un formato concreto y si éste no se respeta, Shazam será incapaz de interpretar dichas ordenes. Hay que señalar que si se utilizan los **Wizards** para escribir los comandos, este tipo de mensajes se reducirá considerablemente.
- **Operaciones incorrectas:** estos mensajes aparecen, por ejemplo, en cálculos que involucran divisiones por cero o logaritmos de números negativos. Cuando en la ejecución de comandos aparecen “operaciones incorrectas”, Shazam asigna como resultado de las mismas un “código de valor perdido<sup>9</sup>” e imprime el correspondiente mensaje de advertencia, no sólo, en la **Ventana Mensajes** sino también en la **Ventana Output** correspondiente. Se trata de errores que deben ser corregidos porque la utilización de variables con valores perdidos en las instrucciones de Shazam, hará que sus salidas no sean correctas.
- **Problemas irresolubles:** no siempre se pueden conseguir soluciones satisfactorias, va a depender de los datos con los que se esté trabajando. Por ejemplo, no se le puede pedir a Shazam que invierta una matriz singular<sup>10</sup>.
- **Resultados sin sentido:** puede ocurrir que aún cuando un fichero de comandos se ejecute sin ningún mensaje de error, los resultados que se obtengan no tengan mucho sentido. Si el usuario detecta en el “**output**” resultados sin sentido, deberá tener en cuenta otras consideraciones de cara a la rectificación de los comandos y, no sólo, los mensajes de error. El fichero de comandos debe revisarse cuidadosamente. Es una buena idea visualizar y chequear los datos de las variables disponibles en la sesión de trabajo, tanto las cargadas desde ficheros de datos como las generadas, para asegurarse de que dichos datos se han cargado o generado correctamente.

Si en la barra de información del **Panel de Depuración** se selecciona el botón **Immediate**, se activa la **Ventana Comandos Inmediatos**. Para poder trabajar en esta ventana es necesario que la primera tarea sea cargar los datos a través de un comando **READ**. Si el usuario tiene asociado al **Project** abierto la base de datos con la que desea trabajar, no sería necesario especificar la ruta del fichero a cargar, bastaría con especificar el nombre del fichero entre paréntesis y explicitar la opción **Names**. Trabajar en esta ventana requiere que el usuario esté familiarizado no sólo con la denominación de los comandos sino también con sus formatos de escritura, puesto que aquí no podría utilizar la ayuda de los **Wizards** para escribirlos. Utilizar esta ventana sería como trabajar con la versión Standard de Shazam.

Si en la barra de información del **Panel de Depuración** se selecciona el botón **Watch** se activa la **Ventana Variables Observadas**. Cuando se ejecuta el conjunto de comandos por partes, es decir, se establecen líneas o puntos de ruptura en la ejecución del programa, cada vez que la ejecución se detiene por el corte, en la **Ventana Variables Observadas** el usuario puede observar las características de determinadas variables disponibles cuyo nombre debe explicitar previamente a la ejecución. Se trata de una característica muy interesante cuando se ejecutan loops.

Si en la barra de información del **Panel de Depuración** se selecciona el botón **Temp**, se activa la **Ventana Variables Temporales**. Funciona de forma similar a la **Ventana Variables Observadas**.

---

<sup>9</sup> El valor que asigna Shazam a los datos perdidos es de “-99999”.

<sup>10</sup> Este mensaje de error suele aparecer, por ejemplo, cuando en un comando **OLS** exista un problema de multicolinealidad perfecta entre los regresores.

Cuando se ejecuta el conjunto de comandos por partes, es decir, se establecen líneas o puntos de ruptura en la ejecución del programa, cada vez que la ejecución se detiene por el corte, en la **Ventana Variables Temporales** el usuario puede observar los valores de las “variables temporales” disponibles hasta ese momento.

## 1.4. Configuración de Shazam Professional

La primera tarea que se debe hacer al inicio de una sesión de trabajo es la configuración del programa de acuerdo con las exigencias de la investigación, de manera que su uso sea lo más cómodo posible.

Para la configuración de Shazam, el usuario debe situarse en la **Barra de Menú** de la **Ventana del Programa** y seleccionar **Project → Options** y en el cuadro de diálogo que se abre, elegir las opciones que considere oportunas.

Se debe tener en cuenta, que algunas de estas opciones son permanentes, es decir, se mantienen aunque se cierre el programa y otras son temporales, por lo que sólo, afectan a la sesión de trabajo, es decir, al abandonar el programa, dichas especificaciones se pierden.

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 a lo largo de la sesión. Si se hace de este modo, no será necesario especificar el path o ruta de la carpeta cada vez que se quieran guardar, puesto que esta será la que Shazam asuma por defecto.

Para definir la ruta de trabajo por defecto, se debe seleccionar la opción "**General**". En el cuadro de diálogo que se abre se debe especificar el directorio o carpeta para los **Projects** y sus ficheros asociados y el directorio o carpeta donde se quiere almacenar los ficheros de procedimiento y a continuación pinchar en “Aceptar”. Por comodidad, en este caso, se ha decidido que en la carpeta denominada "PROYECTOS", se almacene toda la información (véase parte superior izquierda de la Ilustración 1-10).

Lógicamente, para que la selección anterior 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 PROYECTOS.

Realizadas estas operaciones, se abre un cuadro de diálogo, informando de que todo lo especificado se tendrá en cuenta en las próximas ventanas que se abran o se creen en la sesión de trabajo. Además, por comodidad, se ha optado porque Shazam no vuelva a mostrar este mensaje.

Si se selecciona la opción "**Comand Editor**", Shazam proporciona al usuario varias alternativas de diseño para el **Editor de Comandos** (véase parte superior derecha de la Ilustración 1-10):

- Que el **Editor de Comandos** se autoguarde cada vez que se incorpore o ejecute un nuevo comando.
- Que los comandos se escriban en letras mayúsculas.
- Que se muestren los espacios en blanco.
- Que se suprima en la salida el mensaje “Salvar Cambios”.

Esta opción también permite modificar el diseño de los márgenes de esta ventana:

- Ocultar el número de línea.
- Ocultar la banda vertical del margen.

- Suprimir margen, es decir, suprimir los espacios en blanco iniciales que por defecto tiene establecidos Shazam.

En este caso se ha optado por que los comandos se escriban en letras mayúsculas, ya que ello facilitará una lectura más rápida y eficiente del fichero de comandos.

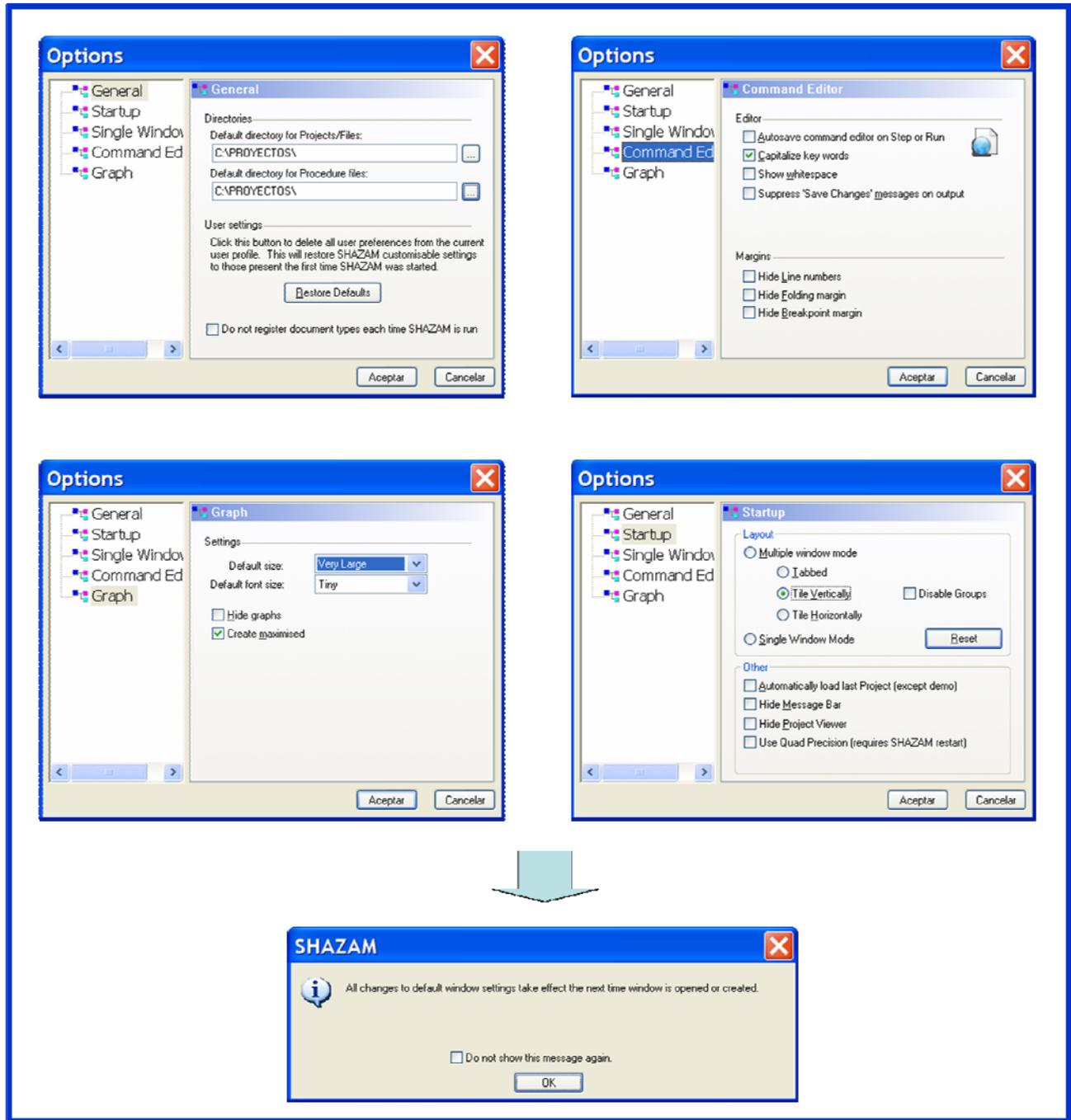


Ilustración 1-10. Posibilidades de configuración del programa antes de iniciar una sesión de trabajo.

Si se elige la opción "**Graph**", el cuadro de diálogo que se abre permite seleccionar el aspecto de los gráficos por defecto (véase parte inferior izquierda de la Ilustración 1-10):

- El tamaño del gráfico.

- El tamaño de la letra.
- La posibilidad de ocultar los gráficos.
- La posibilidad de mostrar los gráficos en la ventana maximizada.

Si la elegida es la opción "**Startup**", el cuadro de diálogo permite seleccionar el modo en que se desea trabajar, es decir, ventana única ó múltiples ventanas y, en este último caso, se puede optar por trabajar con ventanas en cascada, horizontales o verticales. En este cuadro de diálogo también se pueden elegir otras opciones adicionales (véase parte inferior derecha de la Ilustración 1-10):

- Que automáticamente se abra el último **Project** con el que se ha trabajado (excepto la demo).
- Que se oculte la **Ventana Mensajes**.
- Que se oculte la **Ventana Project**.
- Que aumente la precisión en los cálculos<sup>11</sup>.

Shazam Professional también permite trabajar en la modalidad de "**Ventana Única**". Seleccionando la opción "**Single Window**" el cuadro de diálogo permite elegir el aspecto de dicha ventana<sup>12</sup>.

## 1.5. Wizards

Para ejecutar Shazam es necesario conocer su lenguaje. Este conocimiento es imprescindible para explicitar las instrucciones en el **Editor de Comandos**. A diferencia de lo que ocurre en la Versión Standard, donde es imprescindible conocer los comandos y sus opciones, así como su formato de escritura, en la Versión Professional, en la mayoría de los casos, se puede utilizar un "asistente" denominado **Wizard**, que simplifica mucho esta tarea. Una de las principales ventajas y/o atractivos de la Versión Shazam Professional es precisamente la posibilidad de utilizar el apoyo de los **Wizards** a la hora de escribir las instrucciones o comandos.

---

<sup>11</sup> El usuario debe de tener en cuenta que una mayor precisión implica una menor rapidez en la ejecución de los comandos y que para que esta elección surta efecto es necesario reinicializar el programa.

<sup>12</sup> Trabajar en el "**Modo Ventanas Múltiples**" tiene muchas más ventajas que trabajar en el "**Modo Ventana Única**". Quizás la ventaja más importante, sobre todo, para los que utilizan el programa por primera vez, es que en el primer caso, el usuario no está obligado a conocer el lenguaje de Shazam puesto que puede utilizar los **Wizards** para la escritura de comandos y dichos asistentes no están disponibles en el entorno de "**Ventana Única**".

A pesar de que los **Wizards** pueden resultar insuficientes<sup>13</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 que considere oportuna en cada uno de los cuadros de diálogo que se van abriendo desde que se activa un **Wizard** hasta que la instrucción aparece escrita en el **Editor de Comandos** correspondiente y, por ello, será la modalidad que se utilice a lo largo de este manual siempre que sea posible.

Para acceder al asistente se pincha en el botón "**Wizards**" de la barra de tareas de la **Ventana del Programa**<sup>14</sup> y, en este primer cuadro de diálogo (véase Ilustración 1-11), Shazam informa que se pueden utilizar los **Wizards** para ejecutarlos directamente o como asistente para añadir nuevos comandos al **Editor de Comandos**.

Los Wizards permiten mediante la selección de variables y especificando las opciones oportunas con el ratón:

- Estimar modelos
- Construir intervalos de confianza
- Realizar tests de diagnóstico
- Efectuar contrastes de hipótesis
- Construir gráficos
- Realizar predicciones

Presionando en el botón "**Next**", aparece un segundo cuadro de diálogo donde Shazam informa al usuario de los diferentes procedimientos disponibles:

- Modelos Arima
- Modelos con Autocorrelación
- Intervalos de Confianza
- Estadísticos Descriptivos
- Contrastes de Diagnóstico
- Estimación Mínimo Cuadrática Generalizada
- Predicción
- Modelos Heterocedásticos
- Contrastes de Hipótesis
- Estimación por Variables Instrumentales
- Modelo Logit
- Modelos de Regresión no Lineal
- Estimación Mínimo Cuadrática Ordinaria
- Distribuciones de Probabilidad
- Modelo Probit
- Restricciones
- Sistemas de Ecuaciones
- Modelo Tobit
- Estimación Bayesiana con Restricciones de Desigualdad
- Contrastes de Cointegración y Raíz Unitaria

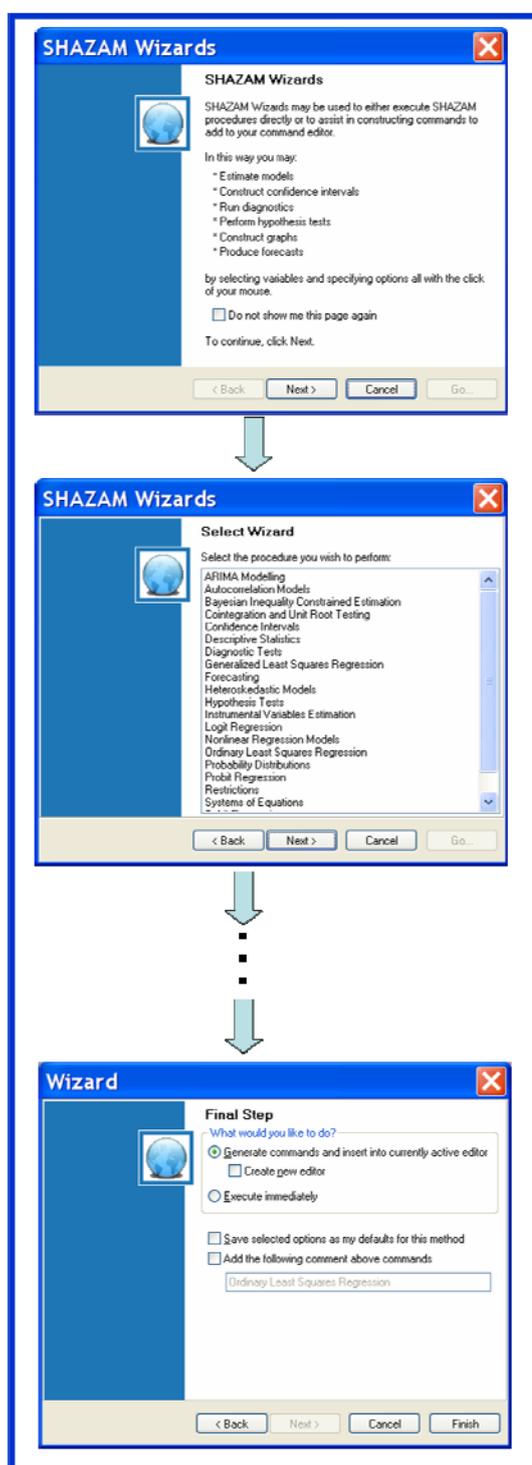


Ilustración 1-11. Secuencia de cuadros de diálogo desde que se accede a un Wizard hasta que la instrucción aparece en el Editor de Comandos correspondiente.

<sup>13</sup> Hay que señalar que todavía no todos los comandos o instrucciones tienen asociados **Wizards**, sin embargo, abarcan una amplia gama de procedimientos econométricos.

<sup>14</sup> A través de este botón se puede acceder a la mayoría de los asistentes disponibles en Shazam, pero no a todos, por ejemplo, no se puede acceder al asistente para construir gráficos o al asistente para visualizar variables.

---

- Estimación de Mínimos Cuadrados en Dos Etapas

Una vez seleccionado el procedimiento, se presiona “**Next**” para pasar al siguiente cuadro de diálogo donde el usuario, una vez más, elegirá las opciones deseadas y así sucesivamente hasta llegar al cuadro de diálogo del final de proceso de escritura del comando. En este último cuadro de selección, Shazam permite varias opciones:

- Generar el comando e insertarlo en el Editor de Comandos activo en ese momento.
- Generar el comando e insertarlo en un nuevo Editor de Comandos.
- Ejecutar el comando de forma inmediata.
- Guardar las opciones seleccionadas para que las utilice este método por defecto.
- Añadir comentarios en líneas anteriores al comando<sup>15</sup>.

Se selecciona una opción y se presiona “**Finish**”. En el caso que el usuario opte por generar el comando e insertarlo en el **Editor de Comandos** activo, podrá comprobar como en la **Ventana Input** aparece escrita dicha instrucción con las opciones que se han seleccionado.

En caso de que sea necesario modificar las opciones de un comando para los que Shazam tenga asociados Wizards, bastará con que se sitúe el cursor en la línea del programa que contiene dicha instrucción y se acceda a la función *Edit Comand* que está disponible en el botón **Wizard** de la **Barra de Tareas** de la **Ventana Principal**.

En la Ilustración 1-11 se recogen los dos cuadros de diálogo iniciales<sup>16</sup> y el final<sup>17</sup> de la secuencia de “ventanas” que aparece desde que se accede a un Wizard de un determinado procedimiento hasta que la instrucción aparece en el Editor de Comandos correspondiente. Estos tres cuadros de diálogo son comunes a todos los wizards y, dado que han sido comentados con detalle en este epígrafe, no parece conveniente que se vuelvan a comentar cuando en capítulos posteriores se aborde con detalle la secuencia de “ventanas” de los diferentes procedimientos para los cuales Shazam tiene disponible un asistente. Dichos cuadros de diálogo también se obviarán de algunas ilustraciones con el fin de no sobrecargarlas.

---

<sup>15</sup> Para que Shazam no tenga en cuenta las notas aclaratorias a la hora de ejecutar el programa, dichas notas van precedidas por el símbolo “\*”.

<sup>16</sup> Shazam denomina a estos cuadros de diálogo “**Shazam Wizards**” y “**Select Wizard**” respectivamente.

<sup>17</sup> Shazam denomina al cuadro de diálogo final “**Final Step**”.

## 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. Además, toda investigación económica lleva implícito trabajar con datos que proceden de situaciones reales y, por tanto, no controladas, lo que conlleva a 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 los datos relativos a 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 en el manejo de datos estadísticos mediante el programa Shazam Professional y, por otra, examinar algunos instrumentos útiles para realizar un análisis preliminar de los mismos.

### 2.2. Introducción de datos

La primera tarea con la que se debe enfrentar el investigador 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 uno a uno. Si por el contrario, la información está almacenada en ficheros o en "bases de datos", dicha tarea se simplifica mucho.

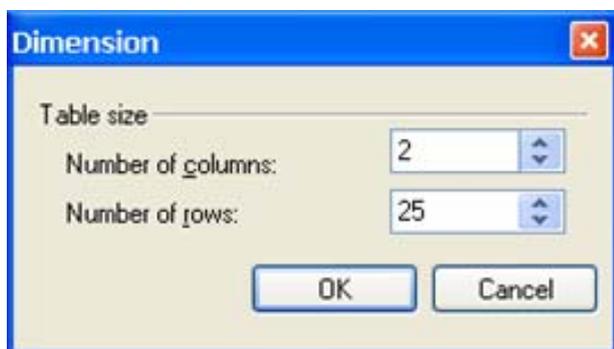


Ilustración 2-1. Dimensión de la base de datos.

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

La introducción directa de datos a través del teclado se puede realizar utilizando el **Menú Principal**, seleccionando **Data → New Dataset** o utilizando la barra de tareas, seleccionando **New → Dataset**.

Se abre un cuadro de diálogo en el que se debe indicar la dimensión de la base de datos, es decir, el número de series o variables (number of columns) y el número de observaciones de esas variables (number of rows). Por ejemplo, si se quieren introducir datos de dos variables con veinticinco observaciones cada una, se indicará 2 columnas y 25 filas y se hará clic en el botón "**OK**" (véase Ilustración 2-1).

Se abre la **Ventana Editor de Datos (Data Editor)** y como en una misma sesión de trabajo se puede utilizar más de un editor de datos, Shazam los enumera empezando en 1.

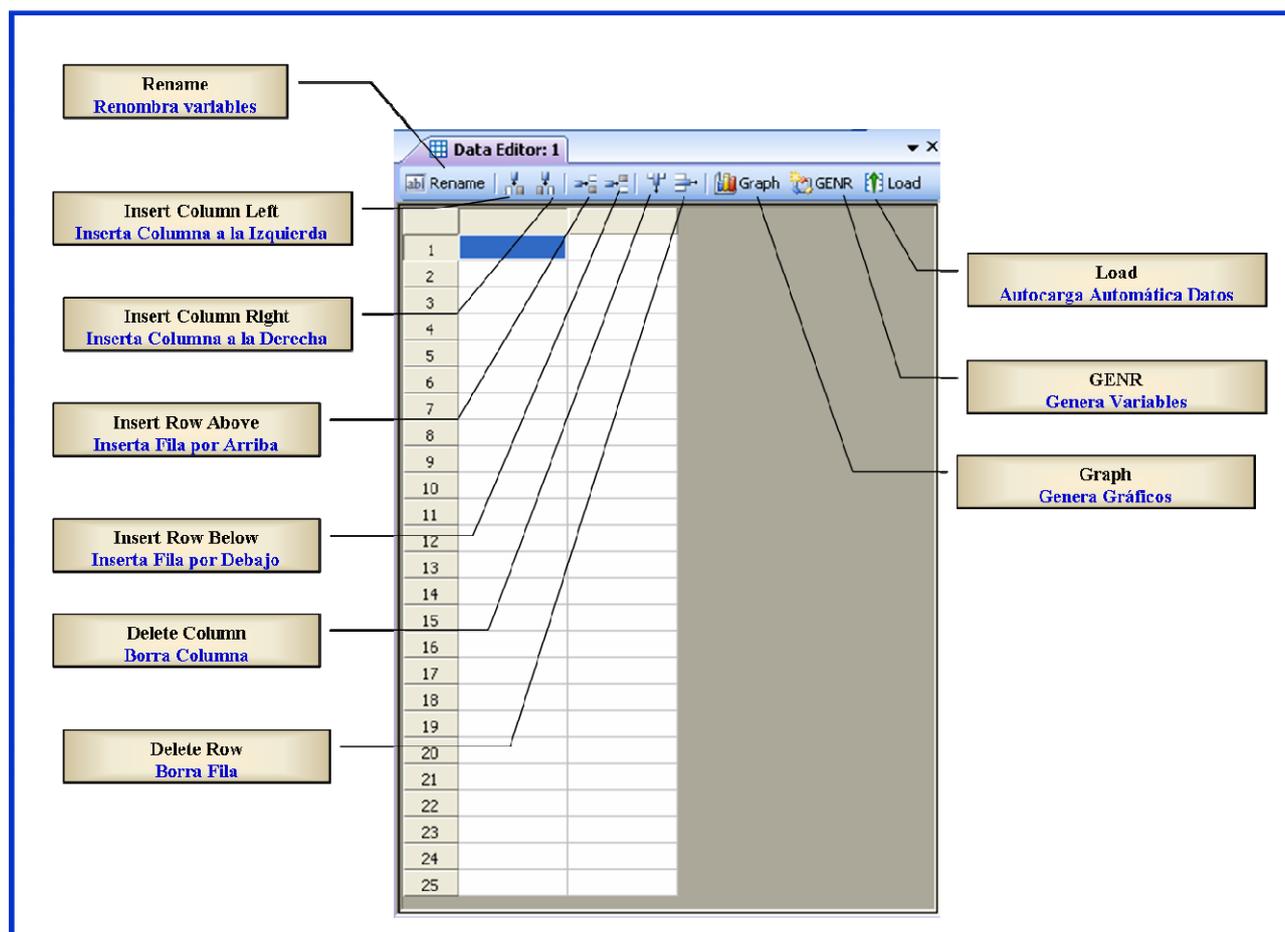


Ilustración 2-2. Ventana Editor de Datos o Data Editor.

Como se puede observar en la Ilustración 2-2, a través de los iconos de la barra de tareas de la **Ventana Editor de Datos** se puede acceder a las tareas más habituales a realizar en una base de datos de una forma más rápida y operativa, sin necesidad de tener que recurrir al submenú **Data** del **Menú Principal**. A través de esta barra de tareas se puede: dar nombre o renombrar las variables; modificar la base de datos insertando o eliminando tanto filas como columnas; crear nuevas variables; hacer representaciones gráficas de las variables incluidas en dicha base de datos y autocargar automáticamente en la sesión de trabajo los datos de las variables contenidos en dicho **Editor de Datos**<sup>18</sup>.

### 2.2.1.1. ¿Cómo dar nombre a las variables?

Para nombrar o cambiar el nombre de una variable, en el **Editor de Datos** se selecciona la columna de dicha variable (que aparecerá marcada en color azul), se selecciona el icono **Rename** de la **Barra Tareas** y en el cuadro de diálogo emergente se especifica el nombre y se hace clic en **OK** (véase Ilustración 2-3).

<sup>18</sup> El submenú **Data** del **Menú Principal** permite además: insertar y eliminar celdas; modificar el aspecto de las columnas (ancho) y filas (alto); seleccionar columnas y filas; reordenar la base de datos y poder utilizar una fila para dar nombre a las variables (lógicamente dicha fila deberá contener texto y no valores numéricos).

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, siempre y cuando comience con una letra y su longitud no supere los ocho caracteres.

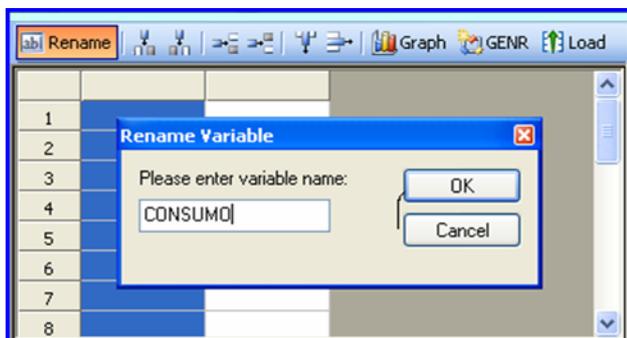


Ilustración 2-3. Cuadro de diálogo para dar nombre o renombrar variables.

En Econometría es frecuente nombrar a la variable dependiente por la letra Y y a la variable independiente por la letra X (en el caso, de que haya más de una variable dependiente o independiente, se les asignan subíndices numéricos distintos para distinguirlas). Estas denominaciones aún cuando son ampliamente utilizadas, no aportan ninguna información descriptiva sobre las variables, por lo que un enfoque más sensato sería asignar a cada variable un nombre que describa el tipo de datos que

contiene.

### 2.2.1.2. ¿Cómo introducir datos?

Una vez que se ha dado nombre a las variables, se deben introducir los datos de las mismas. Debe tenerse en cuenta que no es necesario dar nombre primero a las variables y después introducir sus datos, puede hacerse el procedimiento a la inversa, no obstante, el sentido común, parece aconsejar que el primer camino es el idóneo.

En este caso, la base de datos tiene 25 filas y 2 columnas. En el ejemplo y tal como se recoge en la Ilustración 2-4, la primera columna hace referencia a la variable "CONSUMO" y la segunda a la variable "RENTA". Antes de introducir los datos, las 50 celdas están vacías.

Para la introducción de datos el usuario debe situar el cursor en la celda correspondiente, teclear el valor de la variable para esa observación y, a continuación situar el cursor en la celda siguiente y proceder de la misma forma hasta completar la última observación de la última variable. Para moverse por las distintas celdas puede utilizar las flechas de "arriba", "abajo", "derecha" e "izquierda" del teclado.

	CONSUMO	RENTA
1	360	
2		
3		
4		
5		
6		
7		
8		

Ilustración 2-4. Introducción de datos.

Una vez introducidos todos los datos, es conveniente que se haga un chequeo de los mismos para comprobar que se han introducido correctamente. En el caso de que se detecte algún error, este se corregirá sin más que situar el cursor en la celda correspondiente, borrar el dato erróneo y teclear el correcto. Si el error se detecta en la denominación de una variable, el usuario podrá utilizar el icono **Rename** para cambiarle el nombre.

2.2.1.3. ¿Cómo guardar datos?

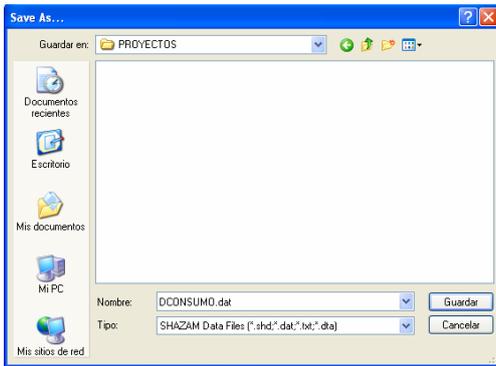


Ilustración 2-5. Tipos de ficheros de datos.

Una vez se han introducido los datos de todas las variables es conveniente guardarlos en un fichero de datos para que estén disponibles en futuras sesiones de trabajo.

Se debe mantener activo el **Editor de Datos** que contiene los valores de las variables que se quieren almacenar y utilizando el **Menú Principal**, seleccionar **File → Save**. A continuación informar a Shazam que se trata de un fichero de datos, es decir, identificarlo con cualquiera de las extensiones que Shazam reconoce por defecto para este tipo de ficheros (.dat; .shd; .txt y .dta) (véase Ilustración 2-5) y, por último, dar nombre al fichero. Como puede observarse en la Ilustración 2-5, en este caso, el fichero se ha denominado

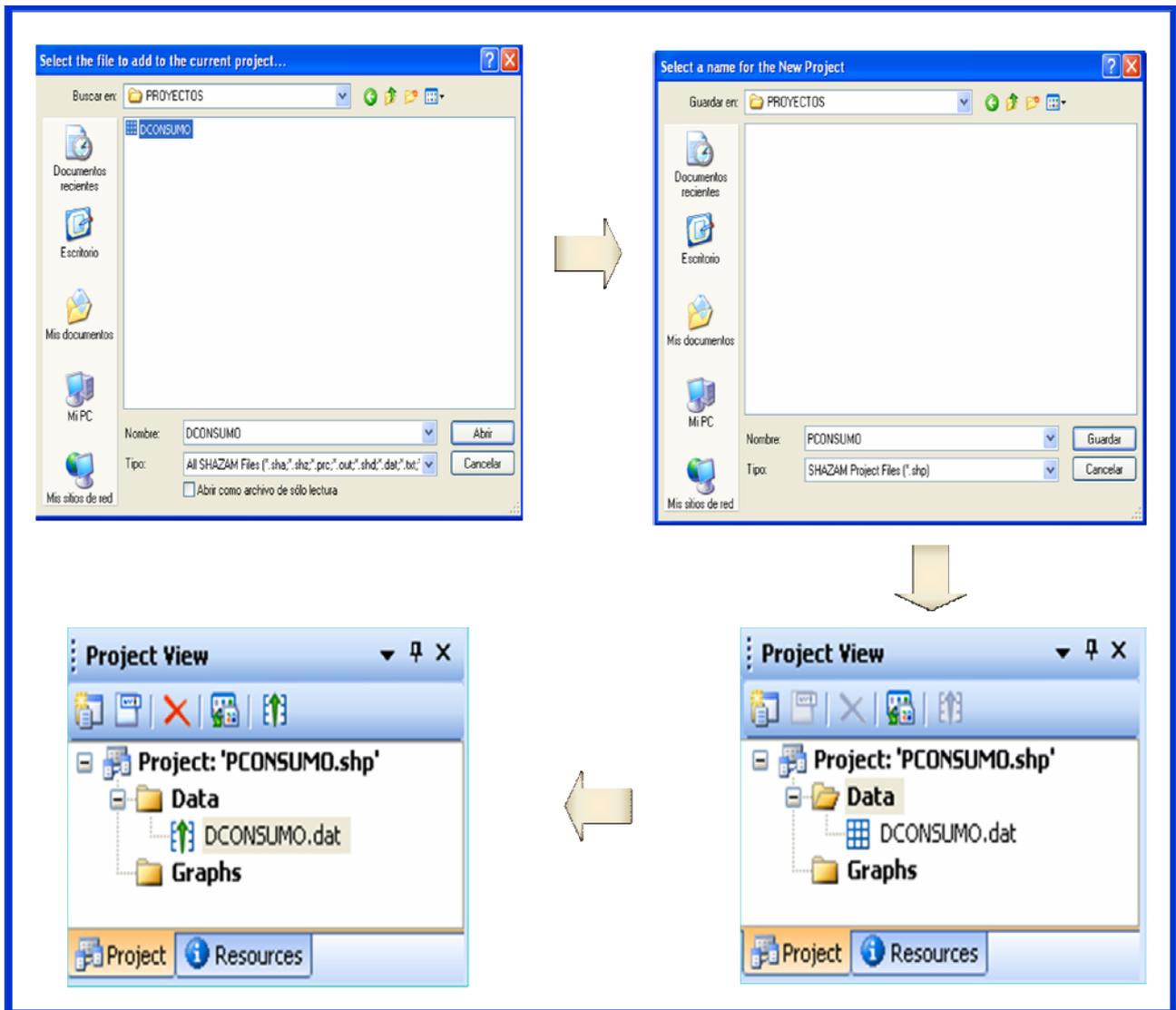


Ilustración 2-6. Procedimiento para añadir un fichero a un Project y para autocargarlo automáticamente.

DCONSUMO.dat. Una vez que los datos están almacenados en el fichero, este puede y debe

adjuntarse a un Project, para lo cuál en el **Menú Principal** se selecciona **Project → Add item to Project → File(s)** y en el cuadro de diálogo se selecciona el fichero a añadir al **Project**.

Se selecciona la ubicación en la que se encuentra el fichero (en este caso, la carpeta "PROYECTOS" que cuelga de "C") y se indica que se trata de un fichero de datos. A continuación, se hace clic en abrir y en la ventana siguiente se le da un nombre al Project (hasta el momento no se había hecho) y se selecciona guardar. De forma inmediata, en la **Ventana Project**, colgando de la carpeta Data, aparece el fichero DCONSUMO.dat (véase Ilustración 2-6).

Para que los datos esten disponibles en todo momento en la sesión de trabajo, es necesario activarlos. Para ello se selecciona el fichero de datos que debe estar situado en la carpeta Data de la **Ventana Project** y, en el **Menú Principal** se selecciona **Project → Toggle Autload** o en la barra de tareas de la **Ventana Project** se selecciona el icono **Toggle Data File Auto Input State**. En ese momento en el rectángulo cuadriculado situado en el lado izquierdo del fichero aparece una flecha hacia arriba, indicando que los datos están disponibles para futuras tareas, sin necesidad de utilizar el comando **READ** para cargarlos.

### 2.2.2. Recuperación de datos desde un fichero

La recuperación de datos desde ficheros o bases de datos es una tarea que se ha simplificado mucho en Shazam Professional. Además, son más los tipos de ficheros de datos que Shazam reconoce por defecto.

#### 2.2.2.1. Ficheros de datos de Shazam

Cuando se trabaja en el entorno de múltiples ventanas, Shazam permite que los ficheros de datos con extensión .dat, .txt, .shz y .dta se puedan cargar automáticamente sin necesidad de tener que utilizar el comando **READ**, lo cual facilita enormemente esta tarea.

#### 2.2.2.2. Ficheros de datos de Excel

Para leer datos desde un fichero de Excel, el usuario debe seguir los pasos recogidos en la Ilustración 2-7.

En primer lugar ir al Menú Principal, seleccionar File → Open.

En el cuadro de diálogo que se abre, seleccionar la carpeta o directorio donde se encuentre el fichero de datos, seleccionar el fichero de datos a recuperar, informar a Shazam de que se trata de un fichero de datos de Excel y hacer clic en “**abrir**”.

A continuación se abre un nuevo cuadro de diálogo donde Shazam informa que para que los datos sean leídos correctamente, el usuario se asegure de que están distribuidos en columnas, que en la primera fila aparecen los nombres de las variables y que en las restantes filas aparecen únicamente valores numéricos. Además, informa de que si se tienen problemas, se utilice "**Advance Data Connector**" de la opción **Data** del **Menú Principal**. Y, a continuación, pregunta si se quiere continuar. Además, proporciona la posibilidad de que Shazam no vuelva a mostrar este mensaje. Es conveniente y aconsejable, que se permitan y se mantengan estos tipos mensajes, sobre todo, mientras el usuario no alcance cierta soltura en la utilización y manejo del programa.

Se hace clic en "Yes" y, en el nuevo cuadro de diálogo, seleccionar la "**Hoja**" del fichero de Excel donde se encuentren almacenados los datos. En la Ilustración 2-7 se puede observar que, en este caso, los datos se encuentran en la hoja denominada "DCONSUMO" y que el fichero tiene dos "Hojas" más. En el momento en que el usuario hace clic en “**OK**”, se abre un **Editor de Datos** cuya denominación coincide con el nombre del fichero a recuperar, en el que aparecen los nombres y valores numéricos de las variables leídas.

El siguiente paso lógico será añadir dicho fichero al **Project** actual y activarlo, para que Shazam pueda utilizar los datos de estas variables en la ejecución de los comandos. La forma más rápida de realizar dicha tarea es utilizar “**Active Document**” disponible tanto en el **Menú Principal (Project → Add Item to Project)** como en el botón “**Add**” de la barra de tareas. Cuando se selecciona “**Active Document**”, Shazam solicita que se guarde dicho fichero con una extensión directamente reconocible por él. La denominación del nuevo fichero puede ser la misma que la del fichero de Excel, no hay ningún problema puesto que se trata de dos ficheros distintos, es decir, los ficheros tendrán el mismo

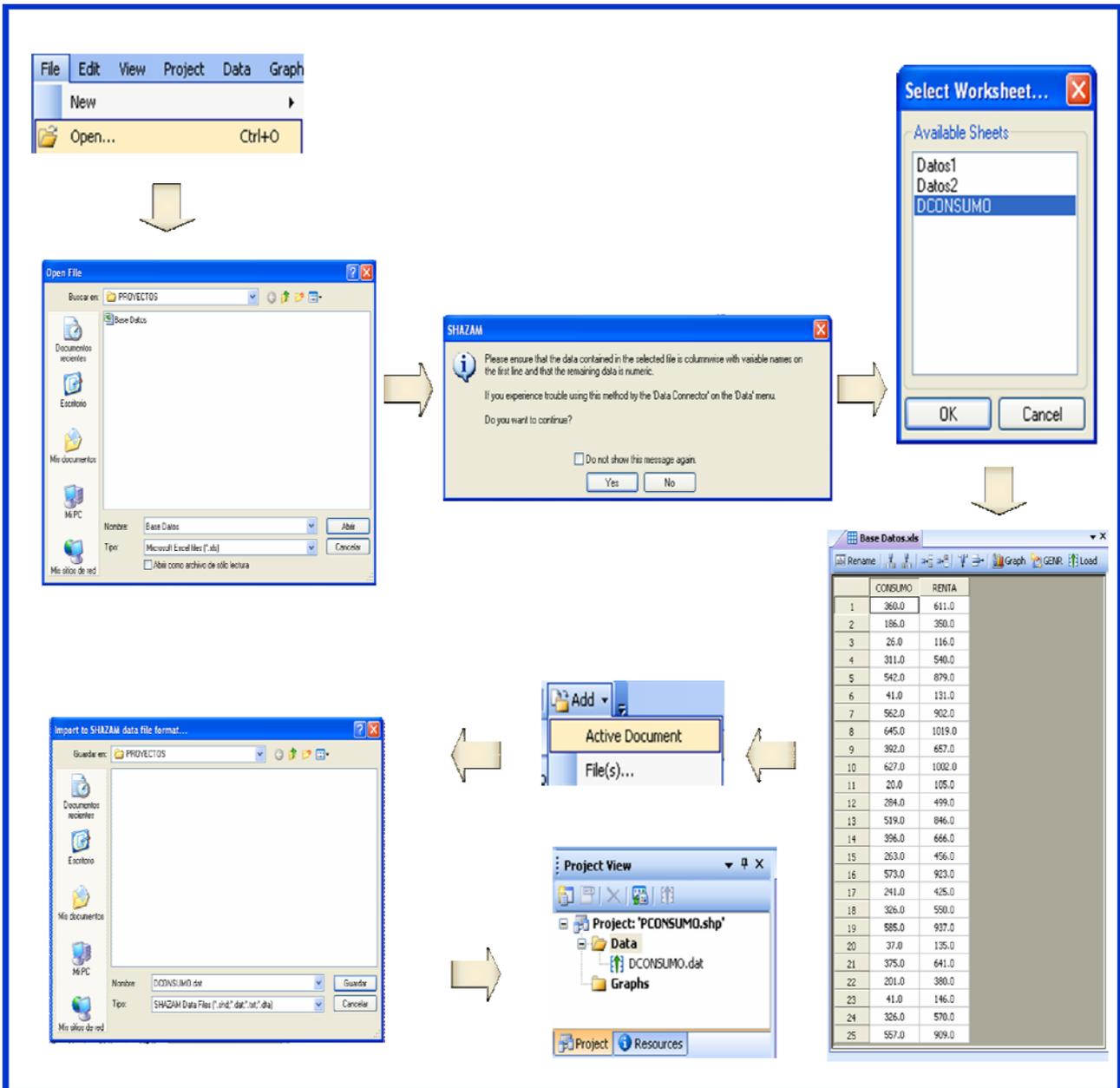


Ilustración 2-7. Procedimiento para recuperar datos de un fichero de excel.

nombre y distinta extensión (.xls el primero y .dat el segundo). Una vez que se hace clic en "**Guardar**", se puede observar que colgando de la carpeta "**Data**" de la **Ventana Project** aparece activado el nuevo fichero de datos.

### 2.2.3. Recuperación de datos desde una Base de Datos

Shazam Professional incorpora un “**Conector de Datos Avanzado**” que permite un acceso universal a base de datos que utilicen mecanismos ADO de Microsoft para conectarse a bancos de datos locales, Networks o en Internet. La versión 10.1 de Shazam Professional permite importar datos de la mayoría de orígenes de datos<sup>19</sup> incluidas las bases de datos OLAP (véase primer recuadro de la parte superior izquierda de la Ilustración 2-8).

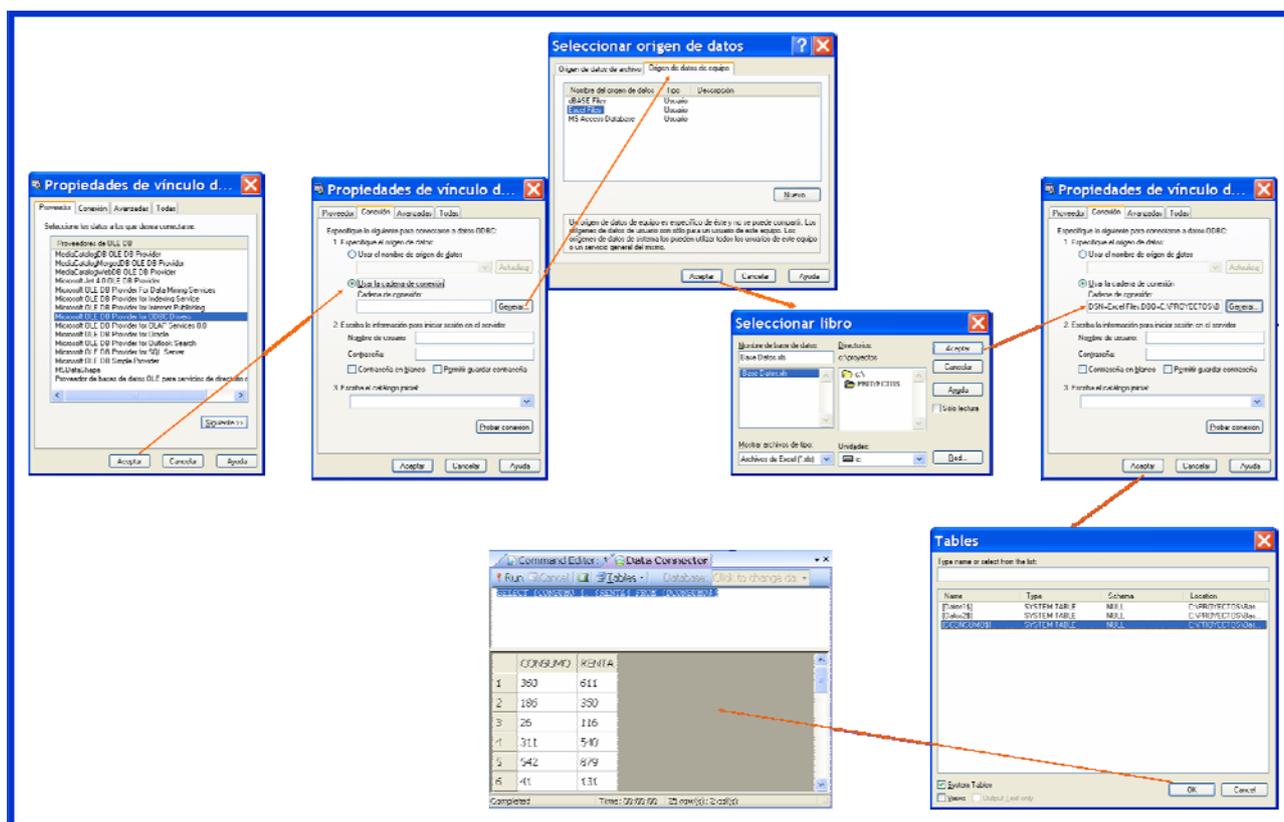


Ilustración 2-8. Procedimiento para la recuperación de datos de una Base de Datos.

El “**Conector de Datos Avanzado**” consta de dos componentes: una primera componente denominada **Microsoft Datos Link** y una segunda, denominada **Ventana Vista Previa**. A través de **Microsoft Datos Link** se establece el tipo de enlace a utilizar en la conexión con la base de datos. Este será el enlace que utilizará SHAZAM para comunicarse con el proveedor de los datos y conectarse al banco de datos. Tal y como puede observarse en la parte inferior derecha de la Ilustración 2-8, una conexión exitosa acaba con el despliegue de una **Ventana Vista Previa**, donde se pueden observar los datos contenidos en la base de datos. Dicha **Ventana Vista Previa** está dividida en dos tablas: una con la información SQL<sup>20</sup> y otra con los datos de las variables. A partir de los datos desplegados en esta

<sup>19</sup> Si no se encuentra el origen de datos que desea entre los especificados por defecto puede recurrir a Windows SharePoint Services. La instalación predeterminada de Windows SharePoint Services incluye un servicio de obtención de datos para conectarse a los bancos de datos en las listas de SharePoint. Un administrador de sitios de SharePoint puede instalar Componentes WebPart de Microsoft Office a fin de agregar más servicios de obtención de datos para Microsoft SQL Server y Microsoft Business Solutions. El programa de instalación de Componentes WebPart de Microsoft Office está disponible en el sitio Web de Descargas en Microsoft Office Online

<sup>20</sup> Structured Query Language (SQL) es un lenguaje estándar y muy potente, usado mundialmente para comunicarse con los proveedores de datos. Sus declaraciones son fáciles de entender y permite a los usuarios establecer criterios específicos para seleccionar la información que desean obtener del banco de datos. No obstante, Shazam a través de su “**Conector de Datos Avanzado**” crea todas las declaraciones SQL necesarias para seleccionar los datos requeridos del banco de datos, sin

ventana y para que puedan ser utilizados por Shazam, debe crearse un fichero de datos (**dataset**) de Shazam y sería aconsejable que dicho fichero se añadiera al **Project** actual puesto que de esta manera dichos datos se podrían cargar de forma automática en la sesión de trabajo. La forma más rápida de conseguir lo anterior es, manteniendo activa la **Ventana Vista Previa**, hacer clic en el botón “**Add**” de la **Barra Principal de Tareas** y dar un nombre a dicho fichero y, de forma inmediata aparecerá colgando de la carpeta “**Data**” de la **Ventana Project**.

Lo primero que se debe hacer para conectarse a una base de datos es acceder al submenú “**Data**” del **Menú Principal** → **Data Connector** → **Connect**. A partir de ese momento se abren una serie de cuadros de diálogo en los que Shazam solicita información acerca de determinados aspectos a tener en cuenta a la hora de hacer la conexión. Un ejemplo sencillo de este procedimiento aparece recogido en la Ilustración 2-8:

- Seleccionar el tipo de proveedor de datos al que desea conectarse.
- Especificar el origen de los datos. En el ejemplo se ha optado por establecer una cadena de conexión dentro del equipo del usuario puesto que el objetivo es acceder al fichero de datos de consumo que figuran en la “base de datos” existente en dicho equipo denominada “Base Datos.xls”.
  - Definir el tipo de datos. En el ejemplo se ha seleccionado Ficheros Excel.
  - Seleccionar el “libro” de excel donde se encuentran los datos. En el ejemplo “C:\PROYECTOS\Base Datos.xls”
- Seleccionar la “hoja” del libro de la base de datos al que se desee acceder.
- Acceder a la **Ventana Vista Previa**.

### 2.3. ¿Cómo generar nuevas variables?

Existen dos alternativas para generar nuevas variables:

- Utilizar el botón **GENR** del **Editor de Datos**.
- Utilizar el comando **GENR** en el **Editor de Comandos**.

#### 2.3.1. A través del Editor de Datos

La ventaja de utilizar el botón **GENR** del **Editor de Datos** es que las nuevas variables se incorporan de forma inmediata a la base de datos que se tenga abierta en ese momento (no obstante, antes de salir de la base de datos se debe guardar la nueva información sino se quiere perder) (véase Ilustración 2-9).

Si las nuevas variables se generan a través del comando **GENR** en el **Editor de Comandos**, estarán disponibles en la sesión de trabajo, pero no aparecerán en la base de datos, a no ser que se anexionen con un comando **WRITE**.

---

necesidad de que el usuario conozca dicho lenguaje. En el ejemplo de la Ilustración 2-8 puede verse el siguiente mensaje en lenguaje SQL: “**SELECT** [CONSUMO], [RENTA] **FROM** [DCONSUMOS]”.

Si se tiene activada una base de datos, es decir, un **Editor de Datos** que contenga datos numéricos de una o más variables, la forma de crear nuevas variables en dicha base de datos es haciendo clic en el botón **GENR** de la barra de tareas de dicha ventana y, en el cuadro de diálogo que se abre, especificar el nombre de la nueva variable y la ecuación de cálculo.

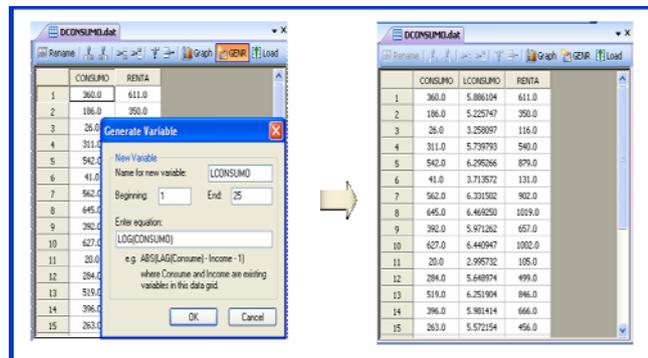


Ilustración 2-9. Procedimiento para generar nuevas variables dentro de una base de datos.

Nótese que Shazam establece la observación inicial y final del rango de la nueva variable, dicho rango por defecto coincide con el tamaño muestral. No obstante, el usuario puede definir la nueva variable para un rango menor.

Algunos de los operadores que pueden utilizarse para definir la ecuación de cálculo de las nuevas variables se recogen en el Cuadro 2-1.

Debe de tenerse en cuenta que estos operadores y funciones se ejecutan de acuerdo a su nivel de

prioridad, por lo que el orden 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. Por ello, para evitar problemas, se pueden utilizar tantos niveles de paréntesis como sea necesario. Obsérvese que las relaciones y operadores lógicos van siempre entre dos puntos, uno antes y otro después de la relación o del operador.

**I. Funciones:**

(1) Exponencial ( $e^V$ ) → <b>EXP(V)</b>	(1) Retardo de orden n → <b>LAG(V,n)</b>
(1) Logaritmo natural → <b>LOG(V)</b>	(1) Raíz cuadrada → <b>SQRT(V)</b>
(1) Suma acumulativa → <b>SUM(V)</b>	(1) Suma de n en n → <b>SUM(V,n)</b>
(1) Valor absoluto → <b>ABS(V)</b>	(1) Variable ficticia → <b>DUM(V)</b>

La función **DUM(V)** creará una variable dummy igual a uno cuando V sea positiva o nula e igual a cero cuando sea negativa.

(1) Valor máximo entre dos variables observación a observación → **MAX(V1,V2)**

(1) Valor mínimo entre dos variables observación a observación → **MIN(V1,V2)**

(1) Suma de una variable más una serie de valores consecutivos que empieza en n+1 → **TIME(V,n)**

**II. Operadores matemáticos:**

(4) Suma → <b>+</b>	(4) Resta → <b>-</b>	(3) Multiplicación → <b>*</b>	(3) División → <b>/</b>	(2) Exponenciación → <b>**</b>
---------------------	----------------------	-------------------------------	-------------------------	--------------------------------

**III. Relaciones lógicas:**

(5) Igual → <b>.EQ.</b>	(5) Distinto → <b>.NE.</b>	(5) Mayor o igual → <b>.GE.</b>
(5) Mayor → <b>.GT.</b>	(5) Menor o igual → <b>.LE.</b>	(5) Menor → <b>.LT.</b>

De manera que si la relación propuesta se cumple la variable que se está generando toma el valor uno, tomando el valor cero en caso contrario.

**IV. Operadores lógicos:**

(6) No → <b>.NOT.</b>	(6) Y → <b>.AND.</b>	(6) O → <b>.OR.</b>
-----------------------	----------------------	---------------------

De manera que si la relación propuesta se cumple la variable que se está generando toma el valor uno, tomando el valor cero en caso contrario.

Cuadro 2-1. Funciones y operadores del comando GENR.

Como ya se ha comentado, en el cuadro de diálogo que se abre tras pinchar en el botón **GENR** del **Editor de Datos**, se debe especificar el nombre de la nueva variable y la ecuación de cálculo. En el ejemplo que aparece recogido en la Ilustración 2-9, se puede observar, que se tiene activada una base de datos denominada “DCONSUMO.dat”, que contiene datos numéricos de las variables

“CONSUMO” y “RENTA” y, que se ha generado una nueva variable denominada "LCONSUMO" como el logaritmo natural de la variable “CONSUMO” para las 25 observaciones muestrales.

Una vez tecleado el nombre de la nueva variable, la ecuación de cálculo y seleccionado el rango muestral y tras pinchar en "OK", se puede observar que esa nueva variable es añadida de forma inmediata a la base de datos.

El usuario debe tener en cuenta que no es conveniente que en una misma base de datos aparezcan variables con el mismo nombre, puesto que en el momento en el que dichos datos se activen para ser utilizados en la sesión de trabajo, tan sólo, aparecerá bajo esa denominación la información incluida en

último lugar y, que en la base de datos será la que figure en las columnas situadas más a la derecha.

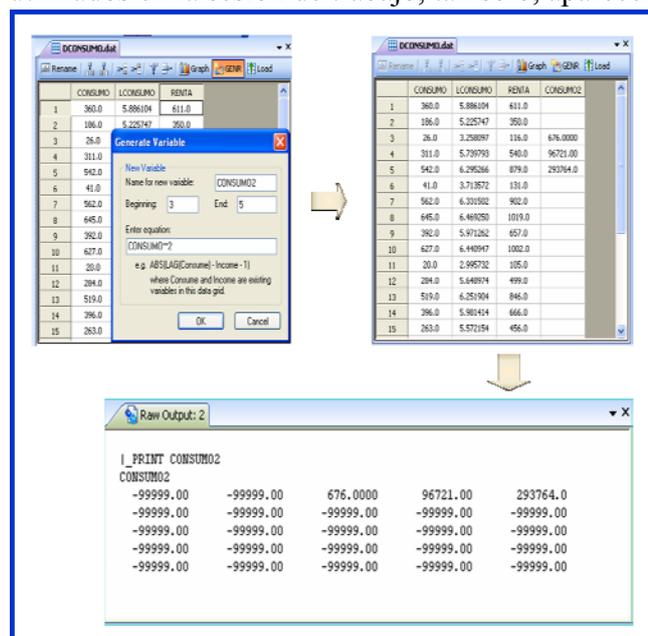


Ilustración 2-10. Generación de variables para un rango menor al muestral.

Ya se ha apuntado que por defecto Shazam genera variables para todo el rango muestral, pero no hay inconveniente en que se generen para un rango menor. Por ejemplo, como se puede observar en la Ilustración 2-10, si se genera “CONSUMO2” como el cuadrado de la variable “CONSUMO” para el rango “3 – 5”, dicha variable, tan sólo, tiene valores para las observaciones 3, 4 y 5, las demás celdas de esa columna están vacías. Si se quiere incorporar la nueva información a la sesión de trabajo, se debe activar de nuevo esa base de datos, lo que se puede hacer pinchando el botón "Load" del **Editor de Datos** (las celdas vacías Shazam la identificará en la sesión de trabajo con el número "-99999"). Hay que señalar que al hacer cálculos, Shazam utiliza dicho valor "-99999"

para las observaciones sin valores numéricos, por tanto, se debe ser cauteloso y comprobar si en la base de datos con la que se está trabajando, existen observaciones sin datos, puesto que ello estaría distorsionando los resultados. No obstante, Shazam permite la posibilidad de eliminar estas observaciones de la muestra y, una manera sencilla de hacerlo, es a través de la modificación del rango muestral, aunque también hay otras opciones.

### 2.3.2. A través del comando GENR

La otra opción para definir nuevas variables es escribir directamente el comando **GENR** en el **Editor de Comandos**, para lo cual el usuario debe conocer, no sólo, la denominación de los operadores que puede utilizar para definir su ecuación de cálculo sino también su formato de escritura.

El formato del comando GENR es:

$$\text{GENR nueva variable} = f(\text{variables existentes})$$

A través del **Editor de Comandos**, Shazam además de permitir definir nuevas variables, también permite generar escalares o constantes, para lo cual se utiliza el comando **GENI**.

La forma más cómoda para acceder a las variables generadas a través del **Editor de Comandos** es utilizando el Wizard “Variables” o la función F12 del teclado. Tal y como se recoge en el ejemplo de la Ilustración 2-11, el usuario puede acceder de forma sencilla a todas las variables activas en ese momento y que se encuentran en la memoria “**system**” y puede seleccionar un subconjunto de ellas para guardarlas en el fichero cuyo nombre especifique o imprimirlas. El usuario debe tener en cuenta que dicho Wizard sólo estará disponible si el fichero de comandos o instrucciones se ejecuta por etapas, es decir, tal y como se ha visto en el capítulo 1, en vez de ejecutar el fichero directamente a través del botón **Run** del **Editor de Comandos**, se deberá ejecutar a través de los botones **Run to cursor**, **Step over to cursor** o **Step through line**.

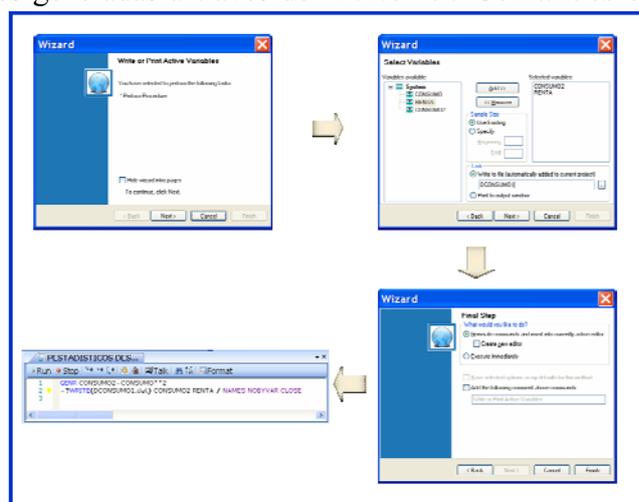


Ilustración 2-11. Secuencia de cuadros de diálogo del wizard “Write or Print Active Variables”.

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

Para calcular los estadísticos descriptivos tenemos dos opciones:

- Utilizar el **Wizard** correspondiente.
- Utilizar directamente el comando **STAT**.

### 2.4.1. A través del WIZARD

Al seleccionar el botón Wizards del Menú Principal → se abre un cuadro de diálogo que informa de las distintas tareas que se pueden implementar a través de este asistente → se hace clic en el botón Next → se abre un nuevo cuadro de diálogo en el que se tiene que seleccionar el procedimiento que se desea implementar y que, en este caso, es “**Descriptive Statistics**”, para seleccionarlo, basta con situarse encima y hacer clic con el botón izquierdo del ratón. Al seleccionarlo queda sombreado en un color más oscuro → se hace clic en “**Go**” → se abre el cuadro de diálogo denominado “**Descriptive Statistic**” donde Shazam informa de que se está en el Wizard de estadísticos descriptivos y de las tareas que va a ejecutar que, en este caso, es única y viene seleccionada por defecto. Se pincha en el botón “Next” y se abre un nuevo cuadro de diálogo denominado “**Select Variables**” donde se tienen que seleccionar las variables y el rango muestral a considerar en los cálculos.

En el cuadro de “**Selección de Variables**”, existen dos recuadros: uno situado a la izquierda donde aparecen todas las variables disponibles y el recuadro de la derecha donde aparecerán las variables seleccionadas por el usuario para el cálculo de los estadísticos descriptivos. Las variables disponibles son las que están activadas en las bases de datos asociadas al actual **Project** y las que están en memoria “**system**”. Para que el **Wizard** funcione correctamente es necesario que la denominación de las variables sea única, es decir, no aparezcan más de una variable con el mismo nombre. El recuadro de variables seleccionadas, en un principio, está vacío. La selección de variables se hará una en una, para ello, en el recuadro de “variables disponibles” se marca una variable y se pincha el botón “Add”, tan sólo, cuando el nombre de dicha variable figure en el recuadro “variables seleccionadas”, estará seleccionada. Si en algún momento, se selecciona una variable equivocada, se marca en el recuadro “variables seleccionadas” y se pincha el botón “Remove” para que deje de estar seleccionada.

Además, también, se puede especificar el "rango muestral" para el que se desea efectuar dicha tarea. Por defecto, está seleccionado el rango existente, pero no hay ningún problema en especificar uno distinto.

Una vez seleccionadas las variables y el rango muestral, se pasa al siguiente cuadro de diálogo haciendo clic en “Next”. En este cuadro de diálogo denominado “**Descriptive Statistics Options**” se pueden seleccionar diferentes opciones en relación con los estadísticos descriptivos:

- Calcular los estadísticos para todas las variables.
- Visualizar la mediana, la moda y los cuantiles.
- Utilizar como divisor N (el tamaño muestral).
- Visualizar la matriz de covarianzas.
- Visualizar la matriz de productos cruzados en desviaciones.
- Utilizar 120 columnas en vez de 80.
- Visualizar ANOVA.
- Visualizar la matriz de correlación.
- Visualizar la matriz de productos cruzados.
- Tratar a las matrices como simples variables.

Además, de visualizar los resultados de estos estadísticos, algunos de ellos se pueden guardar como

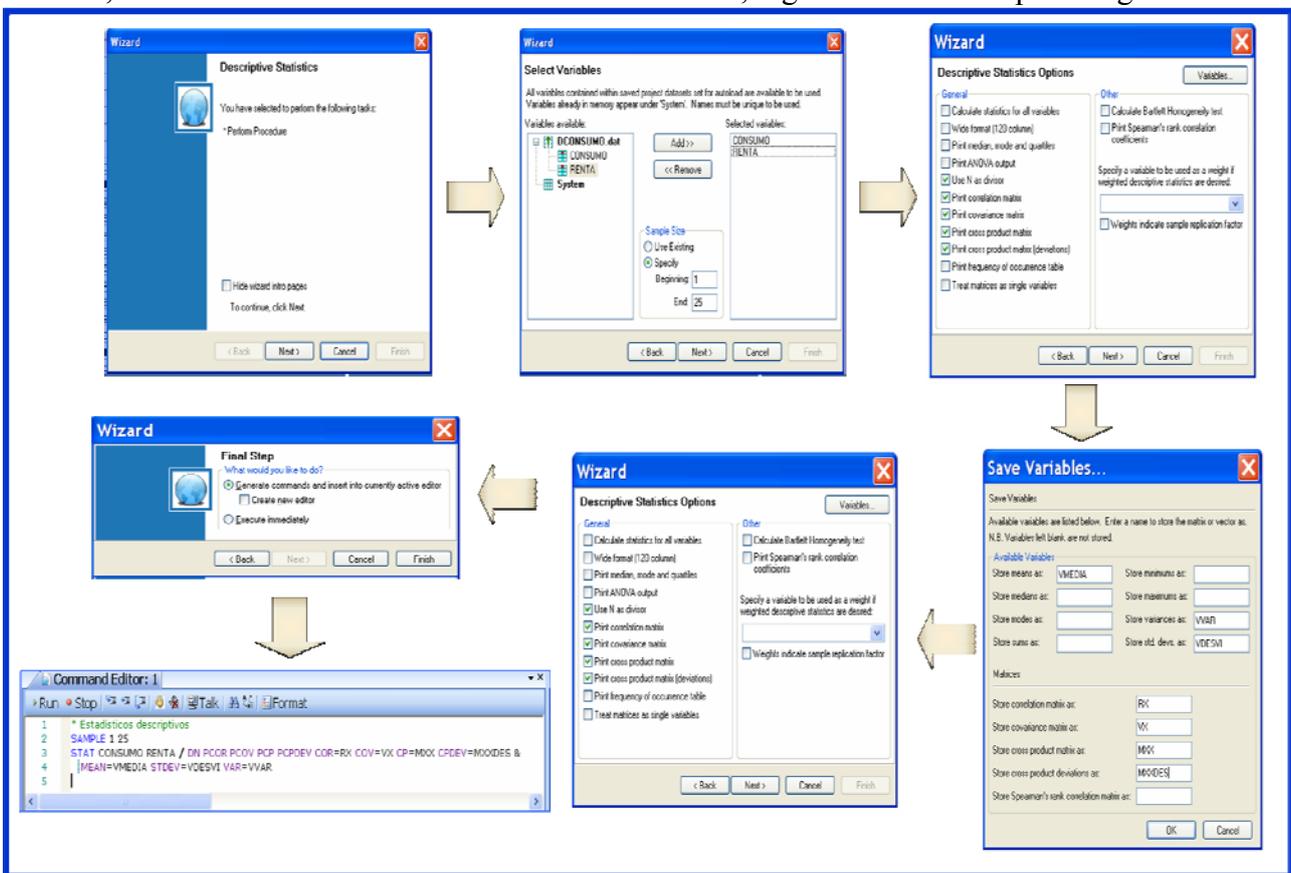


Ilustración 2-12. Secuencia de cuadros de diálogo del wizard “Descriptive Statistics”.

“variables”. Para ello se hace clic en el botón "variables" que aparece en el lado superior derecho de este cuadro de diálogo, con lo que se accede al cuadro de diálogo “Save Variables” donde se puede dar nombre a dichas variables.

Una vez que se le haya dado nombre a las variables que se quieran guardar, se hace clic en el botón "OK" y se vuelve al cuadro de diálogo anterior "Descriptive Statistics Options" y, se selecciona el botón "Next" y con ello se llega al cuadro de diálogo final "Final Step". Al hacer clic en el botón "Finish" se sale del Wizard y, en la Ventana Editor Comandos aparece escrita la instrucción correspondiente.

Una vez escrito el comando en el Editor de Comandos, el siguiente paso lógico es guardarlo en un fichero de comandos. Para ello, con la Ventana Editor de Comandos activada, se selecciona la opción "File" del Menú Principal y, en el submenú que se abre se selecciona "Save as". En el cuadro de diálogo que se abre, se selecciona el directorio o carpeta donde se quiere almacenar dicho fichero, se elige un nombre para el mismo (por ejemplo: PESTADISTICOS DESCRIPTIVOS.SHA) y se hace clic en el botón "Guardar". Una vez que se tiene el archivo de comandos, se puede anexionar al Project con el que esté trabajando. Para ello, se selecciona "Project" del Menú Principal y después, se selecciona Add Item to Project → File y, a continuación, se selecciona el fichero que se quiere añadir y se hace clic en "Abrir". Una vez hecho esto, de forma inmediata, aparece el nuevo fichero en la Ventana Project. Obsérvese que dado que no es un fichero de gráficos ni un fichero de datos, no cuelga ni de la carpeta "Graph" ni de la carpeta "Data" de la Ventana Project. Para abrir este fichero, bastará con seleccionarlo y hacer un doble clic con el botón izquierdo del ratón.

Si se quiere ver la salida de este comando, tan sólo, se tendrá que ejecutar y, para ello, una alternativa es pinchar en el botón "Run" de su ventana. A continuación se va a la Ventana Output correspondiente y se podrá ver y analizar la salida. La salida del comando STAT del ejemplo de la Ilustración 2-12 es la siguiente:

```

|_* Estadísticos descriptivos
|_SAMPLE 1 25
|_STAT CONSUMO RENTA / DN PCOR PCOV PCP PCPDEV COR=RX COV=VX CP=MXX CPDEV=MXXDES &
|   MEAN=VMEDIA STDEV=VDESVI VAR=VVAR
NAME      N      MEAN      ST. DEV      VARIANCE      MINIMUM      MAXIMUM
CONSUMO    25    337.44    200.10    40039.        20.000      645.00
RENDA      25    575.80    295.02    87037.        105.00     1019.0

CORRELATION MATRIX OF VARIABLES -          25 OBSERVATIONS
CONSUMO    1.0000
RENDA      0.99979          1.0000
           CONSUMO      RENDA

COVARIANCE MATRIX OF VARIABLES -          25 OBSERVATIONS
CONSUMO    40039.
RENDA      59020.          87037.
           CONSUMO      RENDA

CROSSPRODUCT MATRIX IN DEVIATIONS -        25 OBSERVATIONS
CONSUMO    0.10010E+07
RENDA      0.14755E+07  0.21759E+07
           CONSUMO      RENDA

CROSSPRODUCT MATRIX OF VARIABLES -          25 OBSERVATIONS
CONSUMO    0.38476E+07
RENDA      0.63330E+07  0.10465E+08
           CONSUMO      RENDA

```

Obsérvese que cuando Shazam muestra las matrices que se han seleccionado, por tratarse de matrices simétricas, sólo permite visualizar los elementos de la diagonal principal y los que están por debajo de dicha diagonal es decir, muestra matrices triangulares inferiores. Además, como referencia para que el usuario pueda identificar sin problemas los datos de dichas matrices, identifica filas y columnas con el

nombre de las variables. El orden en que aparecen las variables en las filas y columnas, es el orden que ocupan dichas variables en el comando **STAT**.

Para visualizar las matrices que se han guardado será necesario ejecutar un comando **PRINT**:

```

|_PRINT RX VX MXX MXXDES VMEDIA VDESVI VVAR
RX
  2 BY      2 MATRIX - LOWER TRIANGLE PRINTED
  1.000000
  0.9997908      1.000000
VX
  2 BY      2 MATRIX - LOWER TRIANGLE PRINTED
  40038.81
  59020.21      87036.64
MXX
  2 BY      2 MATRIX - LOWER TRIANGLE PRINTED
  3847614.
  6332954.      0.1046456E+08
MXXDES
  2 BY      2 MATRIX - LOWER TRIANGLE PRINTED
  1000970.
  1475505.      2175916.
VMEDIA
  337.4400      575.8000
VDESVI
  200.0970      295.0197
VVAR
  40038.81      87036.64

```

Nótese que cuando se visualizan con el comando **PRINT** las matrices que se han guardado, ya no aparecen las referencias de las filas y columnas. Debe tenerse en cuenta que Shazam informa de la dimensión de las matrices y de los vectores fila. Por tanto, los vectores VMEDIA, VDESVI y VVAR a pesar de que se muestra en formato fila, se trata de vectores columna de dos filas cada uno<sup>21</sup>.

<sup>21</sup> Si fuesen vectores filas, Shazam informaría de su dimensión.

## 2.4.2. A través del comando STAT

La otra opción para calcular los estadísticos descriptivos es escribir directamente el comando **STAT** en el **Editor de Comandos**, para lo cual el usuario debe conocer no sólo su formato de escritura sino también la denominación de las opciones disponibles.

El formato del comando STAT es:

**STAT**  $X_1 X_2 \dots X_K$  / **OPCIONES**

En la Ilustración 2-13 aparecen recogidas en color violeta la denominación de las opciones que se pueden insertar en un comando **STAT** a través de su wizard. Como ya se ha comentado el usuario que utilice el wizard del comando **STAT**, tan sólo, debe hacer la elección oportuna en los cuadros de diálogo que se van abriendo y las opciones seleccionadas aparecerán automáticamente en el comando que se inserta en el editor.

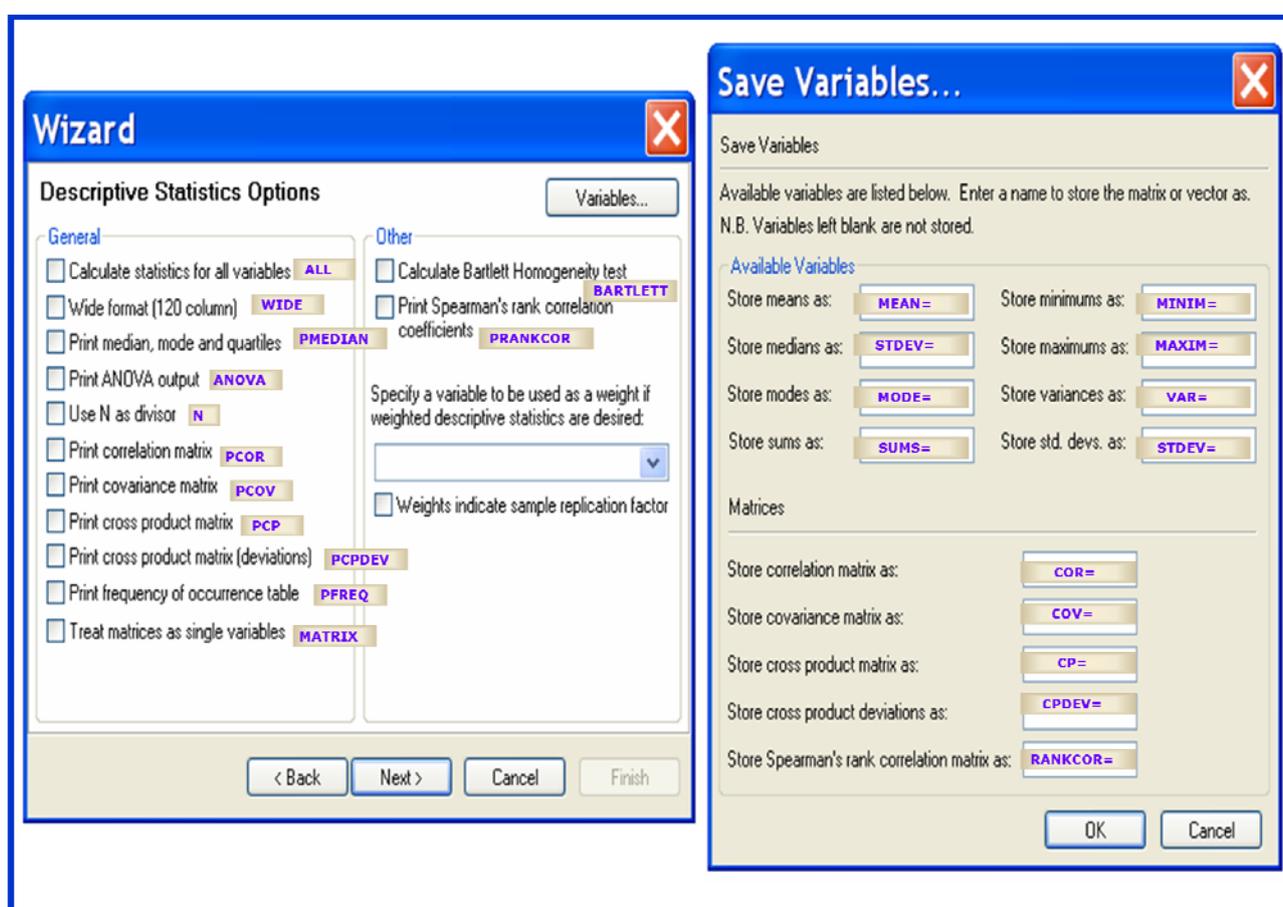


Ilustración 2-13. Opciones del comando STAT que se pueden insertar a través de su Wizard.

En el Cuadro 2-2 aparece recogida una breve descripción de algunas de las opciones disponibles con el comando **STAT**.

- ALL** → Calcula los estadísticos descriptivos básicos para todas las variables. Por lo tanto, cuando se utiliza esta opción, no es necesario especificar el nombre de las variables.
- DN** → Utiliza T en lugar de (T-1) como divisor para el cálculo de determinados estadísticos. Cuando esta opción es activada, el comando STAT proporciona las varianzas, covarianzas y desviaciones estándar de las variables explicitadas en el comando.
- MATRIX** → Permite que cualquier matriz o matrices contenidas en la lista de variables sea tratada como una variable individual (si esta opción no se especifica, Shazam tratará cada una de las columnas de la matriz como una variable individual).
- WIDE** → Usa 120 columnas y NOWIDE usa 80 columnas.
- BEG= END=** → Especifica las observaciones inicial y final para el comando.

**PCOR** → Muestra la matriz de correlaciones de las variables que intervienen en el comando.

$$Cor(X_i, X_j) = r_{X_i X_j} = \frac{Cov(X_i, X_j)}{S_{X_i} S_{X_j}} \quad \forall i, j = 1, 2, \dots, K$$

$$R = \begin{pmatrix} r_{X_1 X_1} & & & & \\ r_{X_2 X_1} & r_{X_2 X_2} & & & \\ r_{X_3 X_1} & r_{X_3 X_2} & r_{X_3 X_3} & & \\ \dots & \dots & \dots & \dots & \\ r_{X_K X_1} & r_{X_K X_2} & r_{X_K X_3} & \dots & r_{X_K X_K} \end{pmatrix}$$

**COR=** → Guarda la matriz de correlaciones de las variables que intervienen en el comando en la matriz indicada.

**PCOV** → Muestra la matriz de varianzas-covarianzas de las variables que intervienen en el comando.

$$Cov(X_i, X_j) = \frac{1}{T-1} \sum_{t=1}^T (X_{it} - \bar{X}_i)(X_{jt} - \bar{X}_j) \quad \forall i, j = 1, 2, \dots, K$$

$$\frac{1}{T-1} \begin{pmatrix} \sum_{t=1}^T (X_{1t} - \bar{X}_1)^2 & & & & \\ \sum_{t=1}^T (X_{2t} - \bar{X}_2)(X_{1t} - \bar{X}_1) & \sum_{t=1}^T (X_{2t} - \bar{X}_2)^2 & & & \\ \sum_{t=1}^T (X_{3t} - \bar{X}_3)(X_{1t} - \bar{X}_1) & \sum_{t=1}^T (X_{3t} - \bar{X}_3)(X_{2t} - \bar{X}_2) & \sum_{t=1}^T (X_{3t} - \bar{X}_3)^2 & & \\ \dots & \dots & \dots & \dots & \\ \sum_{t=1}^T (X_{Kt} - \bar{X}_K)(X_{1t} - \bar{X}_1) & \sum_{t=1}^T (X_{Kt} - \bar{X}_K)(X_{2t} - \bar{X}_2) & \sum_{t=1}^T (X_{Kt} - \bar{X}_K)(X_{3t} - \bar{X}_3) & \dots & \sum_{t=1}^T (X_{Kt} - \bar{X}_K)^2 \end{pmatrix}$$

**COV=** → Guarda la matriz de varianzas-covarianzas de las variables que intervienen en el comando en la matriz indicada.

**PCP** → Muestra la matriz de sumatorios de los productos cruzados de las variables que intervienen en el comando.

$$\begin{pmatrix} \sum_{t=1}^T X_{1t}^2 & & & & \\ \sum_{t=1}^T X_{2t} X_{1t} & \sum_{t=1}^T X_{2t}^2 & & & \\ \sum_{t=1}^T X_{3t} X_{1t} & \sum_{t=1}^T X_{3t} X_{2t} & \sum_{t=1}^T X_{3t}^2 & & \\ \dots & \dots & \dots & \dots & \\ \sum_{t=1}^T X_{Kt} X_{1t} & \sum_{t=1}^T X_{Kt} X_{2t} & \sum_{t=1}^T X_{Kt} X_{3t} & \dots & \sum_{t=1}^T X_{Kt}^2 \end{pmatrix}$$

**CP=** → Guarda la matriz de sumatorios de los productos cruzados de las variables que intervienen en el comando en la matriz indicada.

**PCPDEV** → Muestra la matriz de sumatorios de los productos cruzados en desviaciones respecto a la media de las variables que intervienen en el comando.

$$\begin{pmatrix} \sum_{t=1}^T (X_{1t} - \bar{X}_1)^2 & & & & \\ \sum_{t=1}^T (X_{2t} - \bar{X}_2)(X_{1t} - \bar{X}_1) & \sum_{t=1}^T (X_{2t} - \bar{X}_2)^2 & & & \\ \sum_{t=1}^T (X_{3t} - \bar{X}_3)(X_{1t} - \bar{X}_1) & \sum_{t=1}^T (X_{3t} - \bar{X}_3)(X_{2t} - \bar{X}_2) & \sum_{t=1}^T (X_{3t} - \bar{X}_3)^2 & & \\ \dots & \dots & \dots & \dots & \\ \sum_{t=1}^T (X_{Kt} - \bar{X}_K)(X_{1t} - \bar{X}_1) & \sum_{t=1}^T (X_{Kt} - \bar{X}_K)(X_{2t} - \bar{X}_2) & \sum_{t=1}^T (X_{Kt} - \bar{X}_K)(X_{3t} - \bar{X}_3) & \dots & \sum_{t=1}^T (X_{Kt} - \bar{X}_K)^2 \end{pmatrix}$$

**CPDEV=** → Guarda la matriz de sumatorios de los productos cruzados en desviaciones respecto a la media de las variables que intervienen en el comando en la matriz indicada.

**MAXIM=** → Guarda los valores máximos de las variables que intervienen en el comando como un vector en la variable especificada.

**MINIM=** → Guarda los valores mínimos de las variables que intervienen en el comando como un vector en la variable especificada.

**MEAN=** → Guarda las medias de las variables que intervienen en el comando como un vector en la variable especificada.

**STDEV=** → Guarda las cuasi-desviaciones típicas de las variables que intervienen en el comando como un vector en la variable especificada.

**SUMS=** → Guarda la suma de las variables que intervienen en el comando como un vector en la variable especificada.

**VAR=** → Guarda las cuasi-varianzas de las variables que intervienen en el comando como un vector en la variable especificada.

Cuadro 2-2. Descripción opciones del comando STAT.

## 2.5. ¿Cómo hacer representaciones gráficas?

A través del botón “**Graph**” del **Editor de Datos** o del submenú “**Graph**” del **Menú Principal** se puede acceder al Wizard de creación de gráficos.

En el cuadro de diálogo que se abre denominado “**Create Graph Wizard**” se debe elegir si se quiere construir un gráfico en dos o en tres dimensiones<sup>22</sup>. Una vez hecha esta elección se hace clic en el botón “**Next**” para pasar al siguiente cuadro de diálogo denominado “**Chart Details**”, donde se debe seleccionar los detalles del gráfico: el estilo de la representación (sólo puntos, sólo líneas, líneas y puntos, líneas verticales o barras), las variables que se quieren representar teniendo en cuenta que la seleccionada en primer lugar será la que se represente en el eje horizontal y, también, permite seleccionar el rango para el cual se quiere hacer la representación y que, por defecto, coincide con el rango muestral. En el ejemplo de la Ilustración 2-14 se puede observar que se ha seleccionado una representación únicamente con puntos, que en el eje horizontal se representa la variable “**RENTA**” y en el eje vertical se representa la variable “**CONSUMO**”. Además, se puede observar que dicha representación se hará para las observaciones de la 1 a la 25 que, en ese ejemplo, coincide con el número de observaciones existentes en la base de datos “**DCONSUMO.dat**”

Una vez elegidas todas las características mencionadas, se hace clic en el botón “**Next**” y se abre un cuadro de diálogo, en el que Shazam informa que dispone de toda la información para crear el gráfico. Además, en dicho cuadro de diálogo, ofrece la posibilidad de que dicho gráfico se añada al **Project** actual, recordándole al usuario que si lo hace así, le va a permitir cambiar muchos de sus atributos (títulos, leyendas, estilos, etc). Por tanto, la lógica aconseja añadirlo, puesto que de otro modo no se tendrá la posibilidad de modificarlo. Por otro lado, si se hace así, en todo momento se podrá localizar colgando de la carpeta “**Graph**” de la **Ventana Project**. En dicha carpeta, el usuario podrá comprobar cuales son los gráficos creados hasta ese momento, que denominaciones se les ha dado y, los podrá visualizar de una manera muy sencilla, simplemente haciendo un doble clic sobre ellos.

Para terminar con el proceso de creación de un nuevo gráfico se debe hacer clic en el botón “**Finish**”. Una vez que se hace clic en este botón, Shazam informa una vez más de que, tan sólo, se pueden modificar las características del gráfico si se añade al **Project** corriente. Para añadirlo al **Project** se le tiene que dar un nombre, por ejemplo, “**Grafico1**” y hacer clic en el botón “**Guardar**”. Después de unos instantes, se puede ver como dicho fichero aparece colgado de la carpeta **Graph** de la **Ventana Project** y de manera simultánea, se abre la **Ventana Gráfico** o **Shazam Graph** donde aparece el gráfico que se ha creado. Antes de mostrar el gráfico, Shazam informa de que para modificar sus características se puede utilizar el botón “**Properties**” de la **Ventana Gráfico** (véase Ilustración 2-15).

---

<sup>22</sup> Debe tenerse en cuenta que se requiere un mínimo de dos variables para los gráficos bidimensionales y de tres, para los tridimensionales.

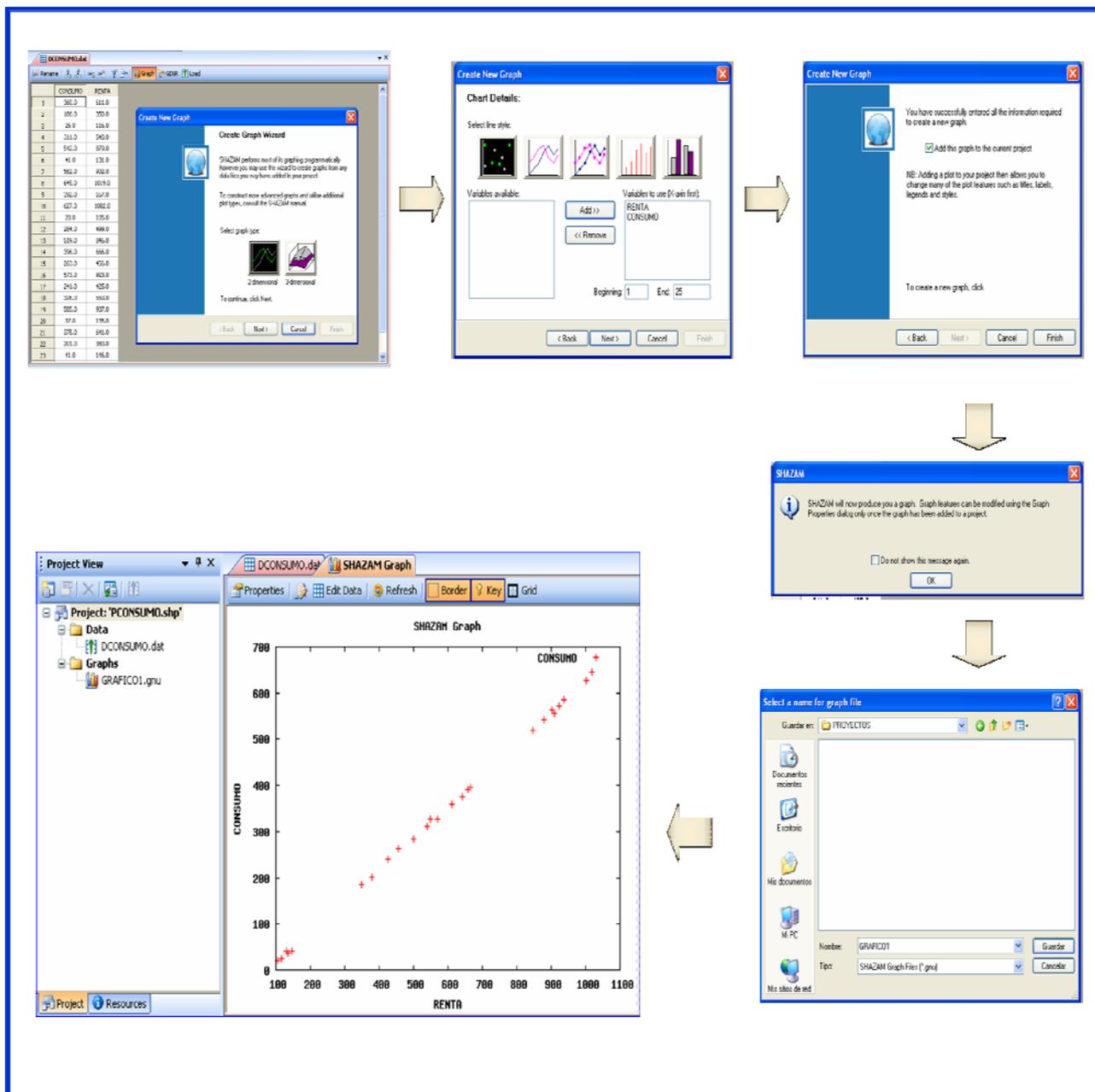


Ilustración 2-14. Secuencia de cuadros de diálogo del wizard “Create Graph”.

The image displays the Shazam Graph interface. The main window, titled 'GRAFICO1.gnu', contains a scatter plot labeled 'Gráfico 1'. The vertical axis is labeled 'CONSUMO' and ranges from 0 to 700. The horizontal axis is labeled 'RENTA' and ranges from 100 to 1100. The plot shows a positive linear relationship between the two variables, with data points marked by red '+' symbols. A grid is overlaid on the plot. The plot area includes a toolbar with buttons for 'Properties', 'Edit Data', 'Refresh', 'Border', 'Key', and 'Grid'. Below the plot, several callout boxes identify these buttons: 'Properties Propiedades', 'Edit Data Editar Datos', 'Refresh Refrescar', 'Grid Líneas de División', 'Key Leyenda', and 'Border Bordes'. To the right of the main window are two 'Graph Properties' dialog boxes. The top dialog shows the 'General' tab with options for 'Show Grid', 'Show Border', 'Hide Legend', 'Box width', 'Font size', 'Line style / Clipping', 'Tics', and 'Legend'. The bottom dialog shows the 'Axes' tab with settings for the X-axis (Label: RENTA), Y-axis (Label: CONSUMO), and Z-axis, including 'Scale' and 'Tic Interval' options.

Ilustración 2-15. Ventana Gráfico o Shazam Graph.

# Capítulo 3. ÁLGEBRA MATRICIAL

## 3.1. Presentación

El uso del álgebra matricial permite presentar de una manera clara y sintética los desarrollos y resultados de los diferentes métodos econométricos. Este capítulo pretende familiarizar al usuario con el manejo de matrices mediante el programa Shazam Professional y, al mismo tiempo, proporcionar un resumen razonablemente conciso sobre el tema.

## 3.2. Definición de matrices

Una **matriz** es una colección de números ordenados rectangularmente:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{T1} & a_{T2} & \dots & a_{Tk} \end{bmatrix}$$

Las matrices se suelen designar con letras mayúsculas (**A**) y sus elementos, con la misma denominación, pero en minúsculas y con dos subíndices ( $a_{ij}$ ), donde el primero hace referencia a la fila y el segundo a la columna, por ejemplo, el elemento genérico  $a_{ij}$  será el que se sitúe en la fila  $i$ -ésima y en la columna  $j$ -ésima.

La **dimensión** de una matriz indica el número de filas y el número de columnas que contiene. **A** es una matriz  $T$  por  $k$  ( $T \times k$ ), es decir, tiene  $T$  filas y  $k$  columnas.

Un **vector** es una colección de números ordenados en una fila (**vector fila**<sup>23</sup>) o en una columna (**vector columna**<sup>24</sup>). Por tanto, una matriz también puede ser interpretada como un conjunto de vectores columna o un conjunto de vectores fila. La interpretación de una matriz como un conjunto de vectores columna sugiere una interpretación natural del conjunto de datos de una muestra, lo cual facilita los desarrollos econométricos<sup>25</sup>.

Shazam permite definir matrices de varias formas:

- Utilizando el editor de matrices
- Utilizando el comando **READ**
- Utilizando el comando **COPY**

### 3.2.1. Utilizando el Editor de Matrices

La introducción directa de los datos de una matriz a través del teclado se puede realizar utilizando el **Menú Principal**, seleccionando **Data** → **New Matrix** o utilizando la barra de tareas, seleccionando **New** → **Matrix**.

---

<sup>23</sup> Un **vector fila** puede ser considerado como una matriz de una única fila.

<sup>24</sup> Un **vector columna** puede ser considerado como una matriz de una única columna.

<sup>25</sup> No obstante, en el próximo capítulo veremos que será necesario cambiar la información que suministran los subíndices: el primer subíndice suele hacer referencia a la columna (nombre de la variable) y el segundo a la fila (observación muestral)

$$X = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{k1} \\ X_{12} & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ X_{1T} & X_{2T} & \dots & X_{kT} \end{bmatrix}$$

En cualquiera de los casos, se abre un cuadro de diálogo en el que se debe indicar la dimensión de la matriz, es decir, el número de columnas (number of columns) y el número de filas (number of rows). Por ejemplo, si se quiere definir una matriz *3 por 3*, se indicará 3 columnas y 3 filas y se hará clic en el botón “OK” (véase Ilustración 3-1).

Se abre la **Ventana Editor de Matrices (Matrix Editor)** y como en una sesión se puede utilizar más de un editor de matrices, Shazam los enumera empezando por 1.

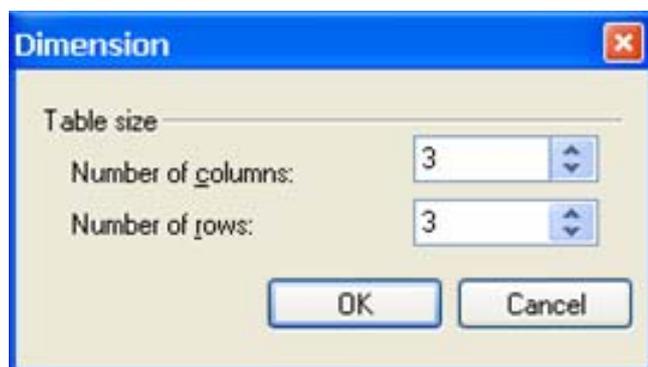


Ilustración 3-1. Dimensión de la matriz.

deseo introducir y teclear dicho valor. Para moverse por las distintas celdas de la matriz puede utilizar las flechas “arriba”, “abajo”, “derecha” e “izquierda” del teclado.

Una vez introducidos todos los datos de la matriz, es conveniente que se haga un chequeo de los mismos para comprobar que estos se han introducido correctamente. En el caso de que se detecte algún error, este se corregirá sin más que situar el cursor en la celda correspondiente, borrar el dato erróneo y teclear el correcto.

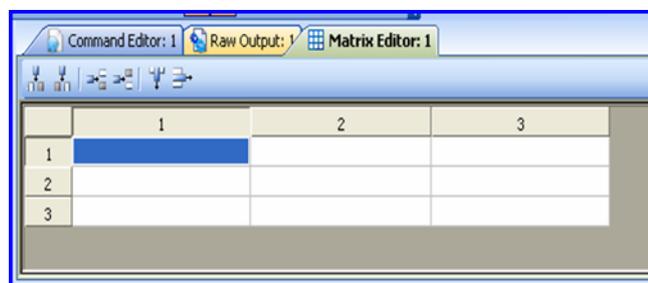


Ilustración 3-2. Ventana Editor de Matrices.

Una vez que se han tecleado todos los datos de la matriz es conveniente guardarlos en un fichero con extensión “.mtx” para que puedan estar disponibles para futuras sesiones de trabajo. Para ello, se debe mantener activo el **Editor de Matrices** que contiene los valores de la matriz que se quiere almacenar y, utilizando el **Menú Principal**, seleccionar **File** → **Save**. A continuación informar a Shazam de que se trata de un fichero de matrices, es decir, identificarlo como un fichero con extensión “.mtx” y, por último, dar un nombre a ese fichero. Como puede observarse en la Ilustración 3-3, en este caso, la matriz se ha denominado A.mtx. Una vez que los datos están almacenados en el fichero, este puede adjuntarse a un Project, para lo cual en el **Menú Principal** se selecciona **Project** → **Add item to Project** → **File(s)** y en el cuadro de diálogo se selecciona el fichero a añadir al **Project**.

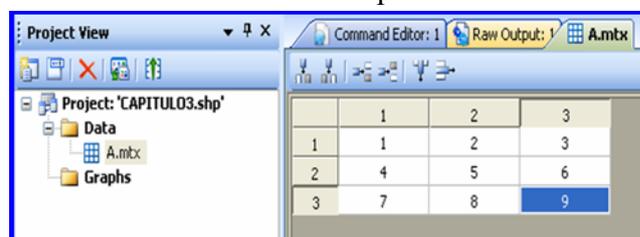


Ilustración 3-3. Extensión de los ficheros de matrices.

### 3.2.2. Utilizando el comando READ

El usuario también puede utilizar el **Editor de Comandos** para introducir matrices de datos y guardarlas en ficheros.

El formato del comando **READ** es:

**READ nombre matriz / ROWS= COLS=**

En la Ilustración 3-4 aparecen recogidos, a través de un ejemplo, los pasos a seguir para definir y visualizar una matriz a través del Editor de Comandos.

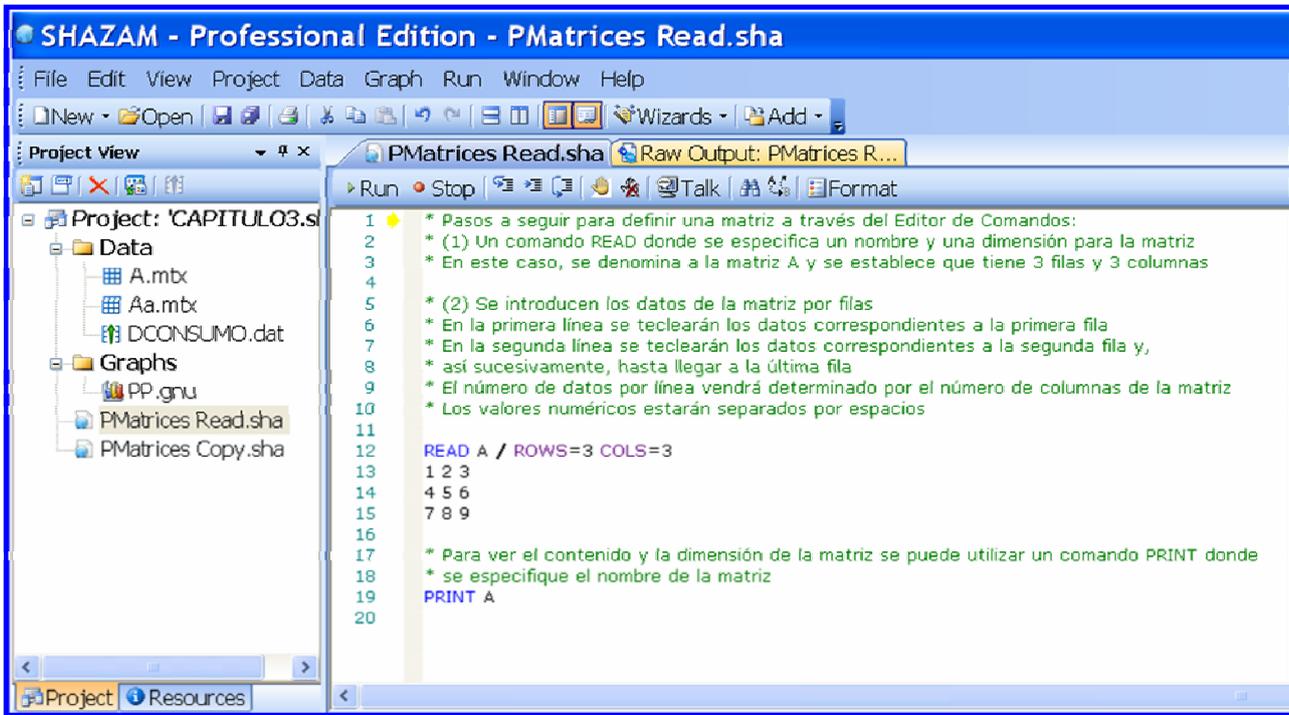


Ilustración 3-4. Pasos a seguir para definir una matriz a través del Editor de Comandos.

A continuación se muestra la salida de la ejecución de los comandos del fichero “PMatrices Read.sha” que aparece en la Ilustración 3-4:

```

|_READ A / ROWS=3 COLS=3
      3 ROWS AND          3 COLUMNS, BEGINNING AT ROW      1

...SAMPLE RANGE IS NOW SET TO:      1      3

|_PRINT A
A
      3 BY      3 MATRIX
1.000000      2.000000      3.000000
4.000000      5.000000      6.000000
7.000000      8.000000      9.000000
    
```

**3.2.3. Utilizando el comando COPY**

A diferencia del comando **READ**, donde se deben teclear los datos de la matriz, la utilización del comando **COPY** permite definir matrices a partir de otras matrices existentes en la sesión de trabajo.

Para llevar a cabo una investigación econométrica es necesario disponer de datos relativos a las variables observables que intervengan en los modelos econométricos formulados para tal fin y, por tanto, una de las primeras tareas de cualquier investigador es la creación de una base de datos que contenga dicha información. A dicha base de datos se le denomina **muestra** y, de forma genérica, se considera que está formada por un conjunto de vectores columna (variables) de tamaño **T**.

El formato del comando **COPY** es:

```
COPY  $M_i$   $M_f$  / FROW= n° primera fila de  $M_i$ ; n° última fila de  $M_i$ 
TROW= n° primera fila de  $M_f$ ; n° última fila de  $M_f$ 
FCOL= n° primera columna de  $M_i$ ; n° última columna de  $M_i$ 
TCOL= n° primera columna de  $M_f$ ; n° última columna de  $M_f$ 
```

Cuando se utiliza un comando **COPY** debe de tenerse en cuenta que las dimensiones especificadas de la matriz inicial [ $M_i$ ] y final [ $M_f$ ] sean compatibles. Por defecto, Shazam considera que el número de columnas de la matriz final coincide con el número de columnas de la matriz inicial y el número de filas con el tamaño del rango muestral con el que se esté trabajando en cada momento, por tanto, si el usuario desea copiar matrices con un número de columnas y filas distinto deberá utilizar las opciones **FCOL=** y **FROW=** con el fin de identificar<sup>26</sup> las columnas y las filas de la matriz inicial cuyos datos se quieren copiar y las opciones **TCOL=** y **TROW=** para identificar las columnas y las filas de la matriz final donde se quieren copiar dichos datos.

Shazam permite utilizar en el comando **COPY** más de una matriz inicial, únicamente, cuando se trata de vectores columna<sup>27</sup>.

A través de la ejecución de los comandos recogidos en el siguiente ejemplo, el usuario puede constatar con unos datos concretos lo expuesto anteriormente:

#### SAMPLE 1 10

\*Generar T como una variable de valores consecutivos que empiece en uno

```
GENR T=TIME(0)
```

\*Generar X0 como una variable que tome siempre el valor 1

```
GENR X0=1
```

\*Mostrar los datos de las variables T, CONSUMO, RENTA y X0

```
PRINT T CONSUMO RENTA X0
```

T	CONSUMO	RENTA	X0
1.000000	360.0000	611.0000	1.000000
2.000000	186.0000	350.0000	1.000000
3.000000	26.00000	116.0000	1.000000
4.000000	311.0000	540.0000	1.000000
5.000000	542.0000	879.0000	1.000000
6.000000	41.00000	131.0000	1.000000
7.000000	562.0000	902.0000	1.000000
8.000000	645.0000	1019.000	1.000000
9.000000	392.0000	657.0000	1.000000
10.00000	627.0000	1002.000	1.000000

\*Observar como todas las variables implicadas en el comando **PRINT** son vectores columna de 10 filas cada uno

\*Generar la matriz X a partir de los vectores columna RENTA y X0

```
COPY RENTA X0 X
```

```
PRINT X
```

<sup>26</sup> Para identificar filas y/o columnas se utilizan dos números separados por “;”: el primer número identifica la primera fila y/o columna y el segundo, identifica la última fila y/o columna seleccionadas en la matriz inicial (**FROW=n°;n°** **FCOL=n°;n°**) y/o en la matriz final (**TROW=n°;n°** **TCOL=n°;n°**). Por tanto, se pueden copiar varias filas y/o columnas de una sola vez, la única restricción es que debe tratarse de filas o columnas contiguas. En el caso, de que se quiera copiar, tan sólo, una fila o una columna, ambos números coincidirán.

<sup>27</sup> Shazam considera a las matrices como un conjunto de vectores columna.

10 BY 2 MATRIX

X

```
611.0000  1.000000
350.0000  1.000000
116.0000  1.000000
540.0000  1.000000
879.0000  1.000000
131.0000  1.000000
902.0000  1.000000
1019.000  1.000000
657.0000  1.000000
1002.000  1.000000
```

\*Observar que dado que la matriz X tiene más de una columna, Shazam informa de su dimensión (10 filas por 2 columnas)

\*Notése que el número de filas coincide con el tamaño muestral y el número de columnas coincide con el número de vectores que intervienen en el comando COPY

\*Generar el vector fila X1T a partir de la primera fila de la matriz X

**COPY X X1T / FROW=1;1 TROW=1;1 FCOL=1;2 TCOL=1;2**

**PRINT X1T**

1 BY 2 MATRIX

X1T

```
611.0000  1.000000
```

\*Observar que dado que X1T es un vector fila, Shazam informa de su dimensión (1 fila por 2 columnas)

\*Generar el vector columna IOTA a partir de la última columna de X

**COPY X IOTA / FROW=1;10 TROW=1;10 FCOL=2;2 TCOL=1;1**

**PRINT IOTA**

IOTA

```
1.000000  1.000000  1.000000  1.000000  1.000000
1.000000  1.000000  1.000000  1.000000  1.000000
```

\*Observar que a pesar de que IOTA es un vector columna, Shazam para ahorrar espacio, lo muestra en formato horizontal y no informa de su dimensión. Además, dado que no dispone de espacio suficiente para mostrar todos esos datos en una única línea, utiliza 2 líneas para mostrarlo. Para determinar el orden que ocupa cada uno de esos datos en el vector, se debe empezar a numerar el orden desde la primera línea a la última y de izquierda a derecha

\*Definir un nuevo rango muestral que contenga, tan sólo, las 5 primeras observaciones

**SAMPLE 1 5**

\*Generar la matriz C a partir de los vectores columna RENTA y X0

**COPY RENTA X0 C**

**PRINT C**

5 BY 2 MATRIX

C

```
611.0000  1.000000
350.0000  1.000000
116.0000  1.000000
540.0000  1.000000
879.0000  1.000000
```

\*Observar que a diferencia de la matriz X, la matriz C tiene 5 filas puesto que se ha definido con un tamaño muestral 5, mientras X se ha definido con un tamaño muestral 10

**SAMPLE 1 10**

**COPY X X CC**

**...ERROR...ONLY ONE MATRIX MAY BE READ AT A TIME**

**COPY IOTA IOTA IOTA II**

**PRINT II**

10 BY 3 MATRIX

II

```
1.000000  1.000000  1.000000
1.000000  1.000000  1.000000
1.000000  1.000000  1.000000
1.000000  1.000000  1.000000
1.000000  1.000000  1.000000
```

```
1.000000 1.000000 1.000000
1.000000 1.000000 1.000000
1.000000 1.000000 1.000000
1.000000 1.000000 1.000000
1.000000 1.000000 1.000000
```

\*Observar que cuando se intenta ejecutar un comando COPY con más de una matriz inicial, Shazam informa del error. No obstante, hay que señalar que cuando se trata de vectores columna de la misma dimensión si lo permite. La matriz II es una matriz de 10 filas por 3 columnas, en la que todos sus elementos son unos.

COPY C CC

...ERROR..C HAS LESS ROWS THAN REQUESTED 5 10

COPY X X3T / FROW=3;3 TROW=1;2

...ERROR..REQUESTED ROWS OR COLUMNS INCOMPATIBLE

\*Observar que cuando se intenta ejecutar un comando COPY donde las dimensiones de las matrices iniciales y finales no son compatibles, shazam informa del error

La inclusión de los comandos **READ** y **COPY** en un fichero de comandos permite al usuario avanzado definir matrices sin necesidad de utilizar el **Editor de Matrices**. Además, permite superar la gran limitación que supone utilizar dicho editor: la necesidad de utilizar un editor y/o un fichero de extensión “.mtx” distinto para cada una de las matrices, lo cual, en ocasiones puede resultar contraproducente, pues podría ocurrir que el número de editores y/o ficheros se disparase y resultase poco operativo, seguramente al usuario no le interese disponer de dichos ficheros sino, tan sólo, de las matrices en la sesión de trabajo.

### 3.3. Operaciones con matrices: el comando MATRIX

Una de las ventajas de Shazam respecto de otros paquetes econométricos es la posibilidad de operar con matrices de una forma relativamente sencilla, lo que permitirá al usuario reproducir los cálculos inherentes a los diferentes métodos econométricos y no centrarse únicamente en la interpretación de los resultados finales.

Operadores matemáticos:					
Multiplicación → *	Suma → +	Resta → -	Exponenciación → **	Traspuesta → '	Concatenación →
Funciones:					
<b>DET</b> (matriz) → Calcula el determinante de la matriz.				<b>DIAG</b> (matriz) → Calcula la diagonal principal de la matriz.	
<b>EIGVAL</b> (matriz) → Calcula los autovalores de la matriz.				<b>EIGVEC</b> (matriz) → Calcula los autovectores de la matriz.	
<b>EXP</b> (matriz) → Aplica el operador exponencial a cada elemento de la matriz.				<b>INV</b> (matriz) → Calcula la inversa de la matriz.	
<b>LOG</b> (matriz) → Calcula el logaritmo natural de cada elemento de la matriz.				<b>RANK</b> (matriz) → Calcula el rango de la matriz.	
<b>SQRT</b> (matriz) → Calcula la raíz cuadrada de cada elemento de la matriz.				<b>TRACE</b> (matriz) → Calcula la traza de la matriz.	
<b>VEC</b> (matriz,arrows) → Convierte un vector en una matriz con n filas.				<b>VEC</b> (matriz) → Convierte una matriz en un vector.	
<b>TRI</b> (matriz) → Crea una matriz triangular inferior a partir de una matriz cuadrada.				<b>LAG</b> (matriz,n) → Cada columna de la matriz se retarda n periodos.	
<b>IDEN</b> (ndim) → Crea una matriz identidad de orden ndim.					
<b>IDEN</b> (ndim,ndiag) → Crea una matriz de n filas y n columnas con una diagonal de unos en la n-ésima paralela inferior a la diagonal principal (n=2,3,..., n). Como caso particular, ndiag=1 proporciona una matriz identidad de orden n.					

Cuadro 3-1. Funciones y operadores del comando MATRIX

El comando **MATRIX** permite hacer operaciones con matrices<sup>28</sup> y su formato es el siguiente:

<sup>28</sup> El usuario debe tener en cuenta que los escalares se pueden considerar como matrices de una fila y una columna; que los vectores son matrices de una fila o una columna dependiendo de si se trata de vectores fila o columna. Además, como ya se ha comentado, las variables existentes en la muestra son consideradas por Shazam como vectores columna.

**MATRIX nombre nueva matriz = f[nombre matrices existentes]**

Algunos de los operadores que pueden utilizarse para definir la ecuación de cálculo de nuevas matrices, a partir de matrices existentes en la sesión de trabajo, se recogen en el Cuadro 3-1. Hay que señalar que, en sentido estricto, algunas de la funciones recogidas en dicho cuadro no se corresponden con operaciones matriciales propiamente dichas, pero Shazam las incorpora porque facilitan cálculos que el usuario puede estar interesado en realizar.

**3.3.1. Tipos de matrices**

Existen denominaciones especiales para las matrices que cumplen determinados requisitos acerca de su dimensión, del valor que toman sus elementos o de como están estos distribuidos dichos elementos dentro de la matriz. En este epígrafe vamos a hacer referencia a algunas de dichas denominaciones:

- Se denomina **matriz cuadrada** a una matriz cuyo número de filas y columna coinciden:

$$\text{si } T = K \Rightarrow A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1T} \\ a_{21} & a_{22} & \dots & a_{2T} \\ \dots & \dots & \dots & \dots \\ a_{T1} & a_{T2} & \dots & a_{TT} \end{bmatrix} \Rightarrow A \text{ es una matriz cuadrada}$$

- Se denomina **matriz diagonal** a una matriz cuadrada en la que todos los elementos son nulos excepto alguno/s de su diagonal principal<sup>29</sup>:

$$\text{si } a_{ij} = 0 \forall i \neq j \text{ y } a_{ij} \neq 0 \text{ para algún } i = j \Rightarrow A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{TT} \end{bmatrix} \Rightarrow A \text{ es una matriz diagonal}$$

La función disponible en Shazam para calcular la diagonal principal de una matriz es **DIAG(matrix)**.

**MATRIX DIAGA=DIAG(A)**

**PRINT A DIAGA**

**3 BY 3 MATRIX**

```
A
1.000000  2.000000  3.000000
4.000000  5.000000  6.000000
7.000000  8.000000  9.000000
DIAGA
1.000000  5.000000  9.000000
```

\*Nótese que, en este caso, dado que la matriz A es de dimensión 3x3, la función DIAG(A) permite guardar en un vector columna de 3 filas, los elementos la diagonal principal de dicha matriz.

<sup>29</sup> La diagonal principal de una matriz cuadrada esta formada por los elementos situados en la diagonal que va desde la esquina superior izquierda hasta la esquina inferior derecha.

- Se denomina **matriz identidad** a una matriz diagonal en la que todos los elementos de su diagonal principal son iguales a la unidad. Se denota por I:

$$\text{si } a_{ij} = 0 \forall i \neq j \text{ y } a_{ij} = 1 \forall i = j \Rightarrow I_{\text{TxT}} = I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \Rightarrow I \text{ es una matriz identidad}$$

La función disponible en Shazam para definir matrices identidades es `IDEN(ndim)` o `IDEN(ndim,ndiag)`.

```
MATRIX I3=IDEN(3)
```

```
PRINT I3
```

```
3 BY 3 MATRIX
```

```
I3
```

```
1.000000  0.000000  0.000000
0.000000  1.000000  0.000000
0.000000  0.000000  1.000000
```

\*Nótese que, en este caso, dado que en la función `IDEN(n°)` aparece entre paréntesis el número 3, se genera una matriz identidad de orden 3x3.

```
MATRIX I5=IDEN(5,1)
```

```
PRINT I5
```

```
5 BY 5 MATRIX
```

```
I5
```

```
1.000000  0.000000  0.000000  0.000000  0.000000
0.000000  1.000000  0.000000  0.000000  0.000000
0.000000  0.000000  1.000000  0.000000  0.000000
0.000000  0.000000  0.000000  1.000000  0.000000
0.000000  0.000000  0.000000  0.000000  1.000000
```

\*Nótese que, en este caso, dado que en la función `IDEN(n°,n°)` aparece entre paréntesis (5,1), se genera una matriz de orden 5x5 que posee unos en su diagonal principal. Esta alternativa sólo permite definir matrices identidades si el segundo número es 1. No obstante, Shazam permite generar matrices donde los unos se sitúan en una diagonal distinta a la diagonal principal. Como ya se ha comentado, Shazam identifica a la diagonal principal con el "1", a la primera paralela inferior a la diagonal principal con el "2" y, así sucesivamente.

```
MATRIX I53=IDEN(5,3)
```

```
PRINT I53
```

```
5 BY 5 MATRIX
```

```
I53
```

```
0.000000  0.000000  0.000000  0.000000  0.000000
0.000000  0.000000  0.000000  0.000000  0.000000
1.000000  0.000000  0.000000  0.000000  0.000000
0.000000  1.000000  0.000000  0.000000  0.000000
0.000000  0.000000  1.000000  0.000000  0.000000
```

\*Nótese que, en este caso, dado que en la función `IDEN(n°,n°)` aparece entre paréntesis (5,3), se genera una matriz de orden 5x5 que posee unos en su segunda diagonal inferior.

- Se denomina **matriz nula** a una matriz en la que todos sus elementos son nulos. A diferencia de una matriz identidad, una matriz nula no tiene porque ser cuadrada:

$$\text{si } a_{ij} = 0 \forall i, j \Rightarrow A_{T \times K} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix} \Rightarrow A \text{ es una matriz nula}$$

### 3.3.2. Matriz traspuesta

La trasposición de matrices consiste en intercambiar filas por columnas. Si A es una matriz  $T \times K$ , su **traspuesta**  $A'$  es una matriz  $K \times T$ .

$$\text{Si } A_{T \times K} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{T1} & a_{T2} & \dots & a_{Tk} \end{bmatrix} \Rightarrow A'_{K \times T} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{T1} \\ a_{12} & a_{22} & \dots & a_{T2} \\ \dots & \dots & \dots & \dots \\ a_{1k} & a_{2k} & \dots & a_{Tk} \end{bmatrix}$$

Si una matriz es igual a su traspuesta, se dice que es una **matriz simétrica**. Para que una matriz sea simétrica es necesario que sea una **matriz cuadrada**, es decir, que tenga el mismo número de filas y columnas:

$$\text{si } T = K \Rightarrow A \text{ simétrica} \Leftrightarrow A = A'$$

En una matriz simétrica los elementos que están por encima de la diagonal principal coinciden simétricamente con los que están por debajo:

$$\text{si } A \text{ simétrica} \Rightarrow a_{ij} = a_{ji} \forall i \neq j$$

La traspuesta de la matriz traspuesta es la matriz original:

$$(A')' = A$$

El operador disponible en Shazam para trasponer matrices es el apóstrofe (') situado en la tecla de cierre de interrogación.

```
*Generar la traspuesta de la matriz C
MATRIX CT=C'
PRINT C CT
```

5 BY 2 MATRIX		2 BY 5 MATRIX				
C		CT				
611.0000	1.000000	611.0000	350.0000	116.0000	540.0000	879.0000
350.0000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
116.0000	1.000000					
540.0000	1.000000					
879.0000	1.000000					

\*Comprobar que la trasposición consiste en intercambiar filas por columnas. Dado que la dimensión de C es 5 filas por 2 columnas, la dimensión de su traspuesta (CT) es de 2 filas por 5 columnas

### 3.3.3. Igualdad de matrices

Para que dos matrices **A** y **B** sean iguales es necesario que cumplan dos requisitos:

- que tengan la misma dimensión y

- que los elementos que ocupan los mismos lugares en ambas matrices coincidan

$$A = B \Leftrightarrow a_{ik} = b_{ik} \forall i \text{ y } \forall k$$

```
*Definir la matriz E
READ E / ROWS=2 COLS=3
1 2 3
4 5 6
*Definir la matriz F
READ F / ROWS=2 COLS=3
1 2 3
4 5 6
*Visualizar las matrices E y F
PRINT E F
E
  2 BY      3 MATRIX
  1.000000  2.000000  3.000000
  4.000000  5.000000  6.000000
F
  2 BY      3 MATRIX
  1.000000  2.000000  3.000000
  4.000000  5.000000  6.000000
*Observar que las matrices E y F son iguales ya que tienen la misma dimensión y los elementos que ocupan los mismos lugares en ambas matrices coinciden
```

### 3.3.4. Suma y resta de matrices

Para que dos o más matrices puedan sumarse o restarse es necesario que tengan la misma dimensión. Los elementos de la matriz suma o resta se obtienen sumando o restando los elementos que ocupan el mismo lugar en las matrices que intervienen en la operación:

$$A \pm B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{T1} & a_{T2} & \dots & a_{Tk} \end{bmatrix} \pm \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \dots & \dots & \dots & \dots \\ b_{T1} & b_{T2} & \dots & b_{Tk} \end{bmatrix} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \dots & a_{1k} \pm b_{1k} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \dots & a_{2k} \pm b_{2k} \\ \dots & \dots & \dots & \dots \\ a_{T1} \pm b_{T1} & a_{T2} \pm b_{T2} & \dots & a_{Tk} \pm b_{Tk} \end{bmatrix}$$

Propiedades:

- Conmutativa<sup>30</sup>:  $A + B = B + A$
- Asociativa:  $[A \pm B] \pm C = A \pm [B \pm C]$
- La traspuesta de una suma o una resta es igual a la suma o resta de las traspuestas:  
 $[A \pm B]' = A' \pm B'$

El operador disponible en Shazam para sumar matrices es el signo “+” y para restar es el signo “-”.

```
*Generar la suma de E+F
MATRIX SUMA=E+F
*Generar la resta de E-F
MATRIX RESTA=E-F
*Visualizar las matrices SUMA y RESTA
PRINT SUMA RESTA
```

<sup>30</sup> La resta sólo cumpliría la propiedad conmutativa cuando ambas matrices fuesen iguales.

2 BY 3 MATRIX

SUMA

2.000000 4.000000 6.000000  
8.000000 10.000000 12.000000

2 BY 3 MATRIX

RESTA

0.000000 0.000000 0.000000  
0.000000 0.000000 0.000000

### 3.3.5. Producto de matrices

Como ya se ha comentado en el apartado 3.2, una matriz puede ser considerada como un conjunto de vectores columna o como un conjunto de vectores fila. Por tanto, dada una matriz A de orden TxK (T filas y k columnas), puede considerarse como un conjunto de k vectores columna de T filas cada uno o como un conjunto de k vectores filas de T columnas cada uno:

$$\text{Sea } A_{T \times K} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ik} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{T1} & a_{T2} & \dots & a_{Tj} & \dots & a_{Tk} \end{bmatrix} = \begin{bmatrix} a_{1T \times 1} & a_{2T \times 1} & \dots & a_{jT \times 1} & \dots & a_{kT \times 1} \end{bmatrix} = \begin{bmatrix} a'_{1 \times k} \\ a'_{2 \times k} \\ \dots \\ a'_{i \times k} \\ \dots \\ a'_{T \times k} \end{bmatrix}$$

donde se denota por  $a_j$  al vector columna formado por los elementos de la columna j-ésima de la

matriz A:  $a_{jT \times 1} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{ij} \\ \dots \\ a_{Tj} \end{bmatrix}$

y se denota por  $a'_i$  al vector fila formado por los elementos de la fila i-ésima de la matriz A:

$$a'_{i \times k} = [a_{i1} \ a_{i2} \ \dots \ a_{ij} \ \dots \ a_{ik}]$$

Para multiplicar dos matrices es necesario que el número de columnas de la primera matriz coincida con el número de filas de la segunda matriz. El número de filas de la matriz producto coincidirá con el número de filas de la primera matriz y el número de columnas, con el número de columnas de la segunda matriz:

$$\text{Si } A_{T \times k} B_{k \times T} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \dots \\ a_{ik} \\ \dots \\ a_{Tk} \end{bmatrix} \begin{bmatrix} b_{1k1} & b_{2k1} & \dots & b_{jk1} & \dots & b_{Tk1} \end{bmatrix} \Rightarrow C_{T \times T} = \begin{bmatrix} a_1' b_1 & a_1' b_2 & \dots & a_1' b_j & \dots & a_1' b_T \\ a_2' b_1 & a_2' b_2 & \dots & a_2' b_j & \dots & a_2' b_T \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_i' b_1 & a_i' b_2 & \dots & a_i' b_j & \dots & a_i' b_T \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_T' b_1 & a_T' b_2 & \dots & a_T' b_j & \dots & a_T' b_T \end{bmatrix}$$

donde  $a_i' b_j$  se define como el producto interno de la fila  $i$ -ésima de la matriz A (vector  $a_i'$ ) por la columna  $j$ -ésima de la matriz B (vector  $b_j$ )

$$a_i' b_j = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ik} \end{bmatrix} \begin{bmatrix} b_{j1} \\ b_{j2} \\ \dots \\ b_{ji} \\ \dots \\ b_{jk} \end{bmatrix} = a_{i1} b_{j1} + a_{i2} b_{j2} + \dots + a_{ij} b_{ji} + \dots + a_{ik} b_{jk} = \sum_{s=1}^k a_{is} b_{js}$$

**Propiedades:**

- El producto de matrices generalmente no es conmutativo. Para que exista el producto AB y BA es necesario que el número de columnas de A sea igual al número de filas de B y que el número de columnas de B sea igual al número de filas de A<sup>31</sup>. No obstante, esto es una condición necesaria pero no suficiente para que AB y BA sean dos matrices iguales, puesto que su dimensión puede ser distinta<sup>32</sup> y aún cuando tengan la misma dimensión, sus elementos pueden diferir<sup>33</sup>.
- Asociativa:  $[AB]C = A[BC]$
- Distributiva respecto a la suma:  $A[B + C] = AB + AC$
- La traspuesta de un producto<sup>34</sup>:  $[AB]' = B' A'$

El operador disponible en Shazam para multiplicar matrices es el signo “\*”.

El **producto de un escalar por una matriz** se obtiene multiplicando el escalar por cada uno de los elementos de la matriz:

<sup>31</sup> Si  $A_{T \times k}$  y  $B_{k \times p} \Rightarrow$  existiría  $AB = C_{T \times p}$  pero no BA puesto que el número de columnas de B no coincide con el número de filas de A. Con esta premisa no sería correcto decir que A y B son matrices multiplicables. Por tanto, lo correcto sería explicitar el orden en que las matrices se multiplican, por ejemplo, en el producto AB, diríamos que la matriz A premultiplica a la matriz B o que la matriz B postmultiplica a la matriz A.

<sup>32</sup> Si  $A_{T \times k}$  y  $B_{k \times T} \Rightarrow C_{T \times T} = AB$  y  $D_{k \times k} = BA$ .

<sup>33</sup> Si A y B son matrices simétricas y multiplicables y,  $A = B' \Rightarrow AB = BA$ .

<sup>34</sup> La traspuesta de un producto es el producto de las traspuestas en orden inverso a su escritura, es decir, de derecha a izquierda.

$$\text{Sea } c_{1 \times 1} A_{T \times K} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1j} & \dots & ca_{1k} \\ ca_{21} & ca_{22} & \dots & ca_{2j} & \dots & ca_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ca_{i1} & ca_{i2} & \dots & ca_{ij} & \dots & ca_{ik} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ca_{T1} & ca_{T2} & \dots & ca_{Tj} & \dots & ca_{Tk} \end{bmatrix}$$

El producto de matrices permite expresar de forma sintética sistemas de ecuaciones, lo cual facilita los desarrollos econométricos. Así, un **sistema de T ecuaciones con (k+1) incógnitas** se puede expresar como el producto de una matriz de orden T x (k+1) y un vector columna de (k+1) filas:

$$\begin{aligned} \hat{y}_1 &= X_{01}b_0 + X_{11}b_1 + X_{21}b_2 + \dots + X_{k1}b_k \\ \hat{y}_2 &= X_{02}b_0 + X_{12}b_1 + X_{22}b_2 + \dots + X_{k2}b_k \\ &\dots \\ \hat{y}_T &= X_{0T}b_0 + X_{1T}b_1 + X_{2T}b_2 + \dots + X_{kT}b_k \end{aligned} \Rightarrow \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \dots \\ \hat{y}_T \end{bmatrix} = \begin{bmatrix} X_{01} & X_{11} & X_{21} & \dots & X_{k1} \\ X_{02} & X_{12} & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ X_{0T} & X_{1T} & X_{2T} & \dots & X_{kT} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_k \end{bmatrix} \Rightarrow \hat{Y}_{T \times 1} = X_{T \times (k+1)} b_{(k+1) \times 1}$$

Cualquier matriz premultiplicada o postmultiplicada por una matriz identidad<sup>35</sup> del orden adecuado da la misma matriz:

$$A_{p \times q} I_{q \times q} = I_{p \times p} A_{p \times q} = A_{p \times q}$$

### 3.3.5.1. Aplicaciones del producto de matrices

El producto de matrices y/o vectores permite que determinadas sumas de elementos se puedan expresar de forma sintética:

- 1) Cualquier vector (fila o columna) premultiplicado o postmultiplicado por un vector (columna o fila) del orden adecuado, en el que todos sus elementos son iguales a uno<sup>36</sup>, proporciona la suma de los

elementos de dicho vector:  $\sum_{i=1}^T x_{it} = x_{i1} + x_{i2} + \dots + x_{iT} = t'x_i = [1 \quad 1 \quad \dots \quad 1] \begin{bmatrix} x_{i1} \\ x_{i2} \\ \dots \\ x_{iT} \end{bmatrix}$

- 2) Si todos los elementos de un vector son la misma constante, dicho vector se puede expresar como el producto de la constante por un vector de unos del mismo orden que el vector original.

si  $x_{it} = a \forall t \Rightarrow x_i = \begin{bmatrix} a \\ a \\ \dots \\ a \end{bmatrix} = at$  y, por tanto, la suma de sus elementos será:  $\sum_{j=1}^T x_{it} = t'x_i = t'(at) = a(t'i) = Ta$

```
| *_Definir la matriz EF=E*F
| _MATRIX EF=E*F
| *_Definir la matriz FE=F*E
```

<sup>35</sup> Una matriz identidad es una matriz cuadrada en la que todos los elementos de su diagonal principal son iguales a uno y los elementos no diagonales son todos iguales a cero. A través del comando **MATRIX nombre matriz = IDEN(número)** se pueden definir matrices identidad de cualquier orden, teniendo en cuenta que dicho orden estará definido por el número que el usuario exprese entre paréntesis. Por ejemplo, si se quiere definir una matriz identidad de orden 2x2 denominada I2, se tendrá que ejecutar el siguiente comando: **MATRIX I2 = IDEN(2)**.

<sup>36</sup> Habitualmente al vector columna de unos se le designa por *iota* (*i*) y al vector fila, por su traspuesto (*i'*).

```

|_ MATRIX FE=F*E
|_ PRINT E F EF FE
E
  2 BY      3 MATRIX
  1.000000      2.000000      3.000000
  4.000000      5.000000      6.000000
F
  3 BY      2 MATRIX
  1.000000      2.000000
  3.000000      4.000000
  5.000000      6.000000
EF
  2 BY      2 MATRIX
  22.000000     28.000000
  49.000000     64.000000
FE
  3 BY      3 MATRIX
  9.000000     12.000000     15.000000
  19.000000    26.000000    33.000000
  29.000000    40.000000    51.000000

*Nótese que, en este caso, existe tanto el producto EF como FE, aunque se trata de matrices de
distintas dimensiones: EF es una matriz 2x2 mientras que FE es de orden 3x3. Esto es posible porque
el número de filas de E coincide con el número de columnas de F y, el número de columnas de E
coincide con el número de filas de F.

|_ MATRIX GE=G*E
|_ PRINT E G GE
E
  2 BY      3 MATRIX
  1.000000      2.000000      3.000000
  4.000000      5.000000      6.000000
G
  2 BY      2 MATRIX
  1.000000      1.000000
  1.000000      1.000000
GE
  2 BY      3 MATRIX
  5.000000      7.000000      9.000000
  5.000000      7.000000      9.000000

|_ MATRIX EG=E*G
...ERROR..INVALID MULT FUNCTION
...MATRIX DIMENSIONS:      2 BY      3      OR      2 BY      2

*Compruebe que para que dos matrices se puedan multiplicar es necesario que el número de columnas de
la primera matriz sea igual al número de filas de la segunda. En este caso existe el producto GE
pero no EG. Obsérvese que Shazam al no poder realizar la multiplicación, muestra un mensaje
indicando que hay un error y que puede estar relacionado con las dimensiones de las matrices que se
pretenden multiplicar.

```

### 3.3.6. Determinante de una matriz

El determinante de una matriz es un escalar que sólo se puede calcular si se trata de una matriz cuadrada, es decir, aquella en que el número de filas y de columnas coincide. Para denotarlo se precede el nombre de la matriz por “**det**” o se incluye dicho nombre entre dos barra verticales “| |”.

Una regla general para calcular el determinante de cualquier matriz sea del orden que sea es a través del uso de sus cofactores.

Se denomina **cofactor** del elemento  $a_{ij}$  y se denota habitualmente por  $A_{ij}$ , al producto del determinante de la matriz que resulta de eliminar la fila y la columna en la que se sitúa dicho elemento por  $(-1)^{i+j}$ .

Sea  $A$  una matriz cuadrada de orden  $M \times M$ , el **cofactor** del elemento  $a_{ij}$  no será más que el determinante de la matriz  $(M-1) \times (M-1)$  que resulta de eliminar la fila  $i$ -ésima y la columna  $j$ -ésima, cambiado de signo si la suma de los subíndices correspondientes a su fila y columna es impar.

$$\text{Sea } A_{M \times M} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{iM} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{M1} & a_{M2} & \dots & a_{Mj} & \dots & a_{MM} \end{bmatrix} \Rightarrow \text{cofactor } A_{ij} = (-1)^{i+j} \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1j-1} & a_{1j+1} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2j-1} & a_{2j+1} & \dots & a_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i-11} & a_{i-12} & \dots & a_{i-1j-1} & a_{i-1j+1} & \dots & a_{i-1M} \\ a_{i+11} & a_{i+12} & \dots & a_{i+1j-1} & a_{i+1j+1} & \dots & a_{i+1M} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{M1} & a_{M2} & \dots & a_{Mj-1} & a_{Mj+1} & \dots & a_{MM} \end{vmatrix}$$

La regla general para obtener el determinante de una matriz consiste en seleccionar una fila o una columna de dicha matriz y multiplicar cada uno de sus elementos por sus cofactores correspondientes y sumar los resultados. Por ejemplo, utilizando como base de los cálculos la fila  $i$ -ésima, el determinante se calcularía como:

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{iM} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{M1} & a_{M2} & \dots & a_{Mj} & \dots & a_{MM} \end{vmatrix} = a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{ij}A_{ij} + \dots + a_{iM}A_{iM}$$

La “fórmula de los cofactores” permite reducir un determinante de cualquier orden a una combinación de determinantes de orden inferior. Por tanto, si se tuviera que calcular el determinante de forma manual, se tendría que desarrollar la expresión hasta llegar a los determinantes de menor orden posible, es decir,  $1 \times 1$ . Para simplificar el proceso sería conveniente elegir la fila o columna más apropiada, es decir, la que permitiera llegar al resultado final con menos cálculos<sup>37</sup>.

**Propiedades:**

- Si se intercambian dos filas/columnas cualesquiera de una matriz, su determinante cambia de signo.
- Si se multiplican todos los elementos de una fila/columna de una matriz por un escalar, su determinante queda multiplicado por ese escalar.
- El valor del determinante queda inalterado si se suma a cualquier fila/columna, un múltiplo de cualquier otra fila/columna.
- Si una matriz tiene dos filas/columnas iguales, su determinante es nulo.
- El determinante de una matriz coincide con el de su traspuesta:  $|A| = |A^t|$
- El determinante del producto de dos matrices es igual al producto de sus determinantes:  $|AB| = |A||B|$

Cuando el determinante de una matriz es nulo, se dice que es una **matriz singular** y cuando su determinante es distinto de cero, se dice que es una **matriz no singular**.

La función disponible en Shazam para calcular el determinante de una matriz es [DET\(matrix\)](#).

```
| _MATRIX DETB=DET (B)
```

<sup>37</sup> Si existen elementos nulos, se elegiría la fila o columna con más ceros. En el caso de que no existiesen elementos nulos, podría resultar conveniente aplicar previamente al cálculo del determinante la técnica denominada “búsqueda de ceros”, puesto que el determinante quedaría inalterado si se suma a cualquier fila/columna un múltiplo de cualquier otra fila/columna.

```

|_PRINT B  DETB
B
  2 BY      2 MATRIX
 14.00000  32.00000
 32.00000  77.00000
DET B
 54.00000

*Dado que el determinante de la matriz B es distinto de cero, B es una matriz singular y, por tanto,
se podrá calcular su inversa.

|_MATRIX DETA=DET(A)
|_PRINT A  DETA
A
  3 BY      3 MATRIX
 17.00000  22.00000  27.00000
 22.00000  29.00000  36.00000
 27.00000  36.00000  45.00000
DETA
 0.000000

*Dado que el determinante de la matriz A es igual a cero, A es una matriz no singular y, por tanto,
no se podrá calcular su inversa.

|_PRINT E
E
  2 BY      3 MATRIX
 1.000000  2.000000  3.000000
 4.000000  5.000000  6.000000
|_MATRIX DETE=DET(E)
...ERROR..INVALID DET  FUNCTION
...MATRIX DIMENSIONS:   1 BY      1      OR      2 BY      3

*Sólo se pueden calcular determinantes de matrices cuadradas. Dado que la matriz E no es cuadrada,
Shazam emite un mensaje de que hay un error y que puede estar relacionado con la dimensión de la
matriz.

```

### 3.3.7. Rango de una matriz

Una matriz puede ser interpretada como un conjunto de vectores columna (variables del modelo econométrico) y, por tanto, su rango puede ser interpretado como el mayor número de columnas (variables) linealmente independientes. Se denomina rango de una matriz al orden del mayor determinante no nulo que se pueda calcular con sus elementos.

si  $A_{T \times k} \Rightarrow \text{rang}(A) = \text{mayor número de columnas linealmente independientes}$

#### Propiedades:

- El rango de una matriz es igual al rango de su traspuesta:  $\text{rang}(A) = \text{rang}(A')$
- Si el rango de una matriz cuadrada es pleno, es decir, coincide con el número de sus columnas, se dice que dicha matriz es no singular: si  $A_{k \times k} \Rightarrow \text{rang}(A) = k \Rightarrow A$  es una matriz no singular. Como se verá en el próximo epígrafe para que una matriz se pueda invertir debe ser cuadrada y tener rango pleno.
- El rango de una matriz siempre será un número menor o igual al mínimo entre el número de columnas y número de filas de la matriz: si  $A_{T \times k} \Rightarrow \text{rang}(A) \leq \min(T, k)$
- El rango de un producto de matrices será menor o igual al mínimo de los rangos de las matrices que se multiplican:  $\text{rang}(AB) \leq \min\{\text{rang}(A); \text{rang}(B)\}$

La función disponible en Shazam para calcular el rango es **RANK(matrix)**.

```

|_MATRIX RANGB=RANK(B)
|_PRINT B  RANGB
B

```

```

2 BY      2 MATRIX
14.00000  32.00000
32.00000  77.00000
RANGB
2.000000

*Nótese que la matriz B es de orden 2x2 y dado que su rango es 2, B tiene rango pleno y, por tanto, se podrá invertir.

|_MATRIX RANGA=RANK(A)
|_PRINT A RANGA
A
3 BY      3 MATRIX
17.00000  22.00000  27.00000
22.00000  29.00000  36.00000
27.00000  36.00000  45.00000
RANGA
2.000000

*Nótese que la matriz A es de orden 3x3 y dado que su rango es 2, A no tiene rango pleno y, por tanto, no se podrá invertir.
    
```

### 3.3.8. Matriz inversa

La inversa de una matriz A, es una matriz denotada  $A^{-1}$  tal que, si se premultiplica o postmultiplica por A, da como resultado la matriz identidad:

$$A A^{-1} = A^{-1} A = I$$

Dos son las condiciones necesarias para que una matriz tenga inversa:

- Que sea una matriz cuadrada, es decir, que el número de filas coincida con el número de columnas.
- Que sea una matriz no singular, es decir, que su determinante sea no nulo.

La inversa de una matriz cuadrada A se obtiene a partir de su determinante y de su matriz adjunta<sup>38</sup>:

$$A^{-1} = \frac{1}{|A|} adjA$$

#### Propiedades:

- La inversa del producto de matrices es igual al producto de las inversas de dichas matrices en orden inverso a su escritura, es decir, de derecha a izquierda:  $(A B \dots F)^{-1} = F^{-1} \dots B^{-1} A^{-1}$
- La traspuesta de la inversa de una matriz es igual a la inversa de su traspuesta:  $(A^{-1})^t = (A^t)^{-1}$
- La inversa de una matriz simétrica, también es simétrica.

La función disponible en Shazam para calcular el rango es **INV(matrix)**.

```

|_MATRIX BI=INV(B)
|_PRINT BI
BI
2 BY      2 MATRIX
1.425926  -0.5925926
    
```

<sup>38</sup> Se denomina matriz adjunta a la traspuesta de su matriz de cofactores:

$$adjA = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1M} \\ A_{21} & A_{22} & \dots & A_{2M} \\ \dots & \dots & \dots & \dots \\ A_{M1} & A_{M2} & \dots & A_{MM} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{M1} \\ A_{12} & A_{22} & \dots & A_{M2} \\ \dots & \dots & \dots & \dots \\ A_{1M} & A_{2M} & \dots & A_{MM} \end{bmatrix}$$

```

-0.5925926      0.2592593

|_MATRIX AI=INV(A)

STEP 3 PIVOT =      0.0000000      INVERSION DISCONTINUED
...ERROR...SINGULAR MATRIX

*Nótese que la matriz A no se puede invertir. Se ha comprobado en epígrafes anteriores que no se
trata de una matriz de rango pleno (es una matriz de orden 3x3 y su rango es 2) y, por tanto, no se
podrá invertir por tratarse de una matriz singular (mensaje de error que emite Shazam).
    
```

### 3.3.9. Traza de un matriz

Se denomina traza de una matriz cuadrada a la suma de los elementos de su diagonal principal:

$$\text{Si } A_{k \times k} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} \Rightarrow tr(A) = \sum_{i=1}^k a_{ii}$$

#### Propiedades:

- La traza de una matriz identidad coincide con su dimensión:

$$\text{Si } I_{k \times k} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \Rightarrow tr(I_k) = \sum_{i=1}^k 1 = k$$

- La traza de una matriz coincide con la traza de se traspuesta:  $tr(A) = tr(A')$
- La traza de la suma de matrices es igual a la suma de las trazas:  $tr(A + B) = tr(A) + tr(B)$
- La traza del producto de un escalar por una matriz es igual al producto del escalar por la traza de la matriz:  $tr(cA) = c tr(A)$
- La traza del producto de matrices es igual al producto de las trazas:  $tr(AB) = tr(A) tr(B)$
- Si  $\exists AB$  y  $BA \Rightarrow tr(AB) = tr(BA)$

La función disponible en Shazam para calcular la traza es **TRACE(matrix)**.

```

|_MATRIX I3=IDEN(3)
|_MATRIX TRI3=TRACE(I3)
|_PRINT I3 TRI3
I3
  3 BY      3 MATRIX
  1.000000      0.000000      0.000000
  0.000000      1.000000      0.000000
  0.000000      0.000000      1.000000
TRI3
  3.000000

*Compruebe que la traza de una matriz identidad coincide con su dimensión (en este caso, como la
matriz identidad es de orden 3x3, su traza es 3).

|_MATRIX DT=D'
|_MATRIX TRD=TRACE(D)
|_MATRIX TRDT=TRACE(DT)
|_PRINT D TRD DT TRDT
D
    
```

```

3 BY      3 MATRIX
1.000000  2.000000  3.000000
4.000000  5.000000  6.000000
7.000000  8.000000  9.000000

```

```

TRD
15.00000

```

```

DT

```

```

3 BY      3 MATRIX
1.000000  4.000000  7.000000
2.000000  5.000000  8.000000
3.000000  6.000000  9.000000

```

```

TRDT
15.00000

```

\*Compruebe que la traza de una matriz coincide con la traza de su traspuesta, ya que los elementos de la diagonal principal de ambas matrices coinciden.

```

|_PRINT E

```

```

E

```

```

2 BY      3 MATRIX
1.000000  2.000000  3.000000
4.000000  5.000000  6.000000

```

```

|_MATRIX TRE=TRACE(E)

```

```

...ERROR..INVALID TRAC FUNCTION

```

```

...MATRIX DIMENSIONS: 1 BY 1 OR 2 BY 3

```

\*La traza sólo está definida para matrices cuadradas. Por ello, dado que la matriz E es de orden 2x3, al intentar calcular la traza Shazam emite un mensaje de error indicando que el orden de la matriz es inadecuado.

# Capítulo 4. ESTIMACIÓN MCO: MODELO DE REGRESIÓN LINEAL CLÁSICO

## 4.1. Presentación e hipótesis básicas del Modelo de Regresión Lineal Múltiple

El Modelo de Regresión Lineal Múltiple es el caso más simple de modelización econométrica, por lo que será nuestro 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:

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

Donde:

T: Tamaño muestral (número de observaciones disponibles).

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

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

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.

$\beta_1, \beta_2, \dots, \beta_K, \beta_0$ : 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  $\beta_0$  que acompaña al regresor ficticio, encontrándonos en este caso ante modelos formulados sin ordenada en el origen.

El Modelo de Regresión Lineal Múltiple se puede escribir matricialmente:

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

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 (las K variables explicativas y el regresor ficticio). A la matriz X también 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:

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

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.

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

Donde V es una matriz simétrica y escalar de orden  $T \times T$ , denominada matriz de varianzas-covarianzas de las perturbaciones de manera que 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:

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

La varianza de las perturbaciones 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.

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

Las covarianzas entre las distintas perturbaciones son nulas, lo que significa que las perturbaciones no están correlacionadas entre sí, por ello, lo que ocurra en cada observación en esos factores integrados en la perturbación no va a estar relacionado con lo que ocurra en la observación anterior o posterior.

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

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

**H<sub>4.1</sub>. Rango (X) = K+1 (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 ( $X'X$ ).

#### **H<sub>4.2</sub>. $T > K+1$ .**

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, 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 elegiremos 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 estimadores con buenas propiedades. Los estimadores obtenidos por este método se denominarán **Estimadores Mínimo Cuadráticos Ordinarios (EMCO)** de los parámetros del modelo.

## **4.2. ¿Cómo estimar por MCO en Shazam?**

Para estimar por Mínimos Cuadrados Ordinarios tenemos dos opciones:

- Utilizar el **Wizard** correspondiente.
- Utilizar directamente el comando **OLS**.

### **4.2.1. A través del WIZARD**

En la Ilustración 4-1 se recoge la secuencia de los cuadros cuadros de diálogo intermedios que Shazam Professional proporciona al usuario para facilitarle la escritura de un comando OLS con sus diferentes opciones.

Al seleccionar el botón Wizards del Menú Principal → se abre un cuadro de diálogo que informa de las distintas tareas que se pueden implementar a través de este asistente → se hace clic en el botón Next → se abre un nuevo cuadro de diálogo en el que se tiene que seleccionar el procedimiento que se desea implementar y que, en este caso, es “**Ordinary Least Square Regression**”, para seleccionarlo, basta

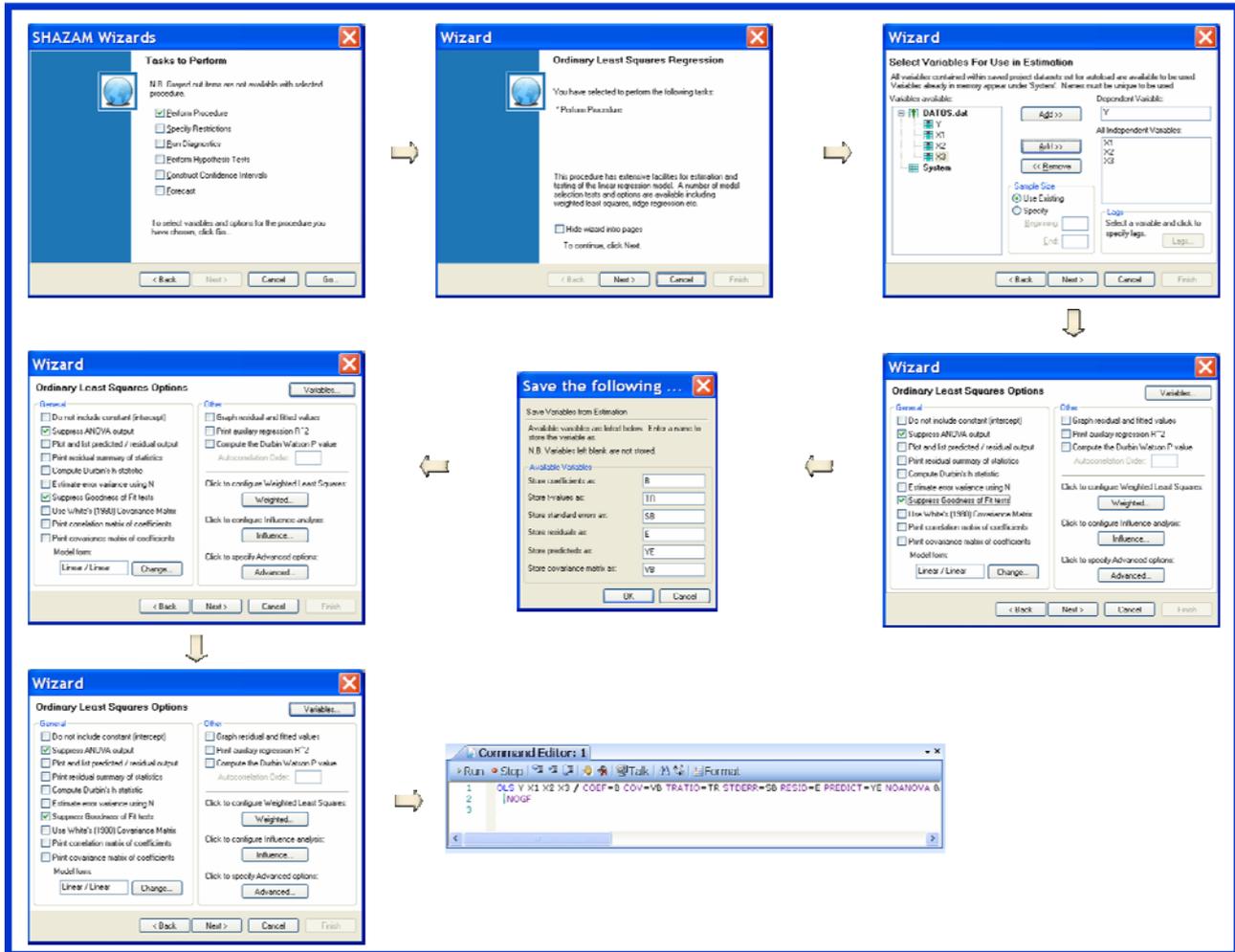


Ilustración 4-1. Secuencia de cuadros del Wizard “Ordinary Least Square Regression”.

con situarse encima y hacer clic con el botón izquierdo del ratón. Al seleccionarlo queda sombreado en un color más oscuro → se hace clic en “Next” → se abre el cuadro de diálogo “Tasks to Perform” donde Shazam, por defecto, tiene seleccionado el objetivo principal de este asistente, que es la implementación de una estimación MCO. Aunque Shazam tiene seleccionada dicha tarea por defecto y la ejecutará siempre que se ejecute una estimación MCO, permite seleccionar tareas adicionales que tienen que ser ejecutadas precedidas por un comando **OLS** y a las que se puede acceder a través de otros “wizards” y, por este motivo, no serán analizadas en este epígrafe:

- a. Especificar restricciones cuyo wizard es **Restrictions**.
- b. Especificar tests de diagnóstico cuyo wizard es **Diagnostic Tests**.
- c. Especificar contrastes de hipótesis cuyo wizard es **Hypothesis Tests**.
- d. Construir intervalos de confianza cuyo wizard es **Confidence Intervals**.
- e. Realizar predicciones cuyo wizard es **Forecasting**.

Una vez seleccionada la tarea → se hace clic en “Go” → se abre el cuadro de diálogo “**Select Variables for use in Estimation**” donde se informa a Shazam de las variables con las que debe realizar dicha tarea, así como del rango muestral para realizarla.

Según se puede ver en la Ilustración 4-2, en este cuadro de diálogo, en el recuadro denominado “variables disponibles” situado a la izquierda, Shazam informa al usuario de todas las variables disponibles para realizar la estimación y que pueden proceder de ficheros de datos y/o de las que Shazam tenga en memoria “system”<sup>39</sup>. Como ya se ha comentado, en una misma sesión de trabajo, pueden estar disponibles variables procedentes de más de un fichero de datos, por ello, Shazam informa de que ficheros están cargados automáticamente (rejilla con una flecha hacia arriba) y que

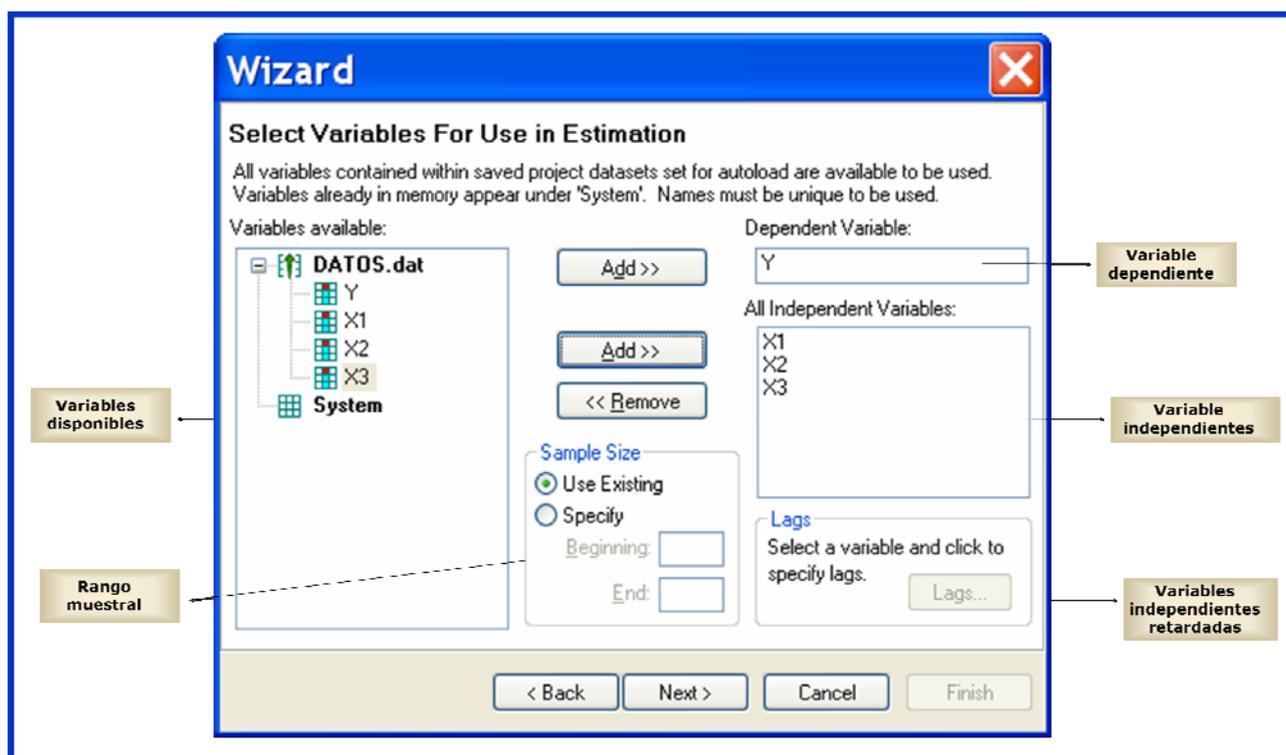


Ilustración 4-2. Selección de variables para usar en la estimación.

variables están disponibles en cada uno de ellos.

El usuario debe tener la precaución de que no exista más de una variable con la misma denominación, pues si este fuera el caso, la información disponible en dicha variable sería, tan sólo, la cargada en último lugar.

<sup>39</sup> Como ya se ha comentado, para poder acceder a las variables que Shazam tiene en memoria, será necesario haber ejecutado previamente instrucciones a través de los botones *Run to cursor*, *Step over to cursor* o *Step through line* que impliquen la generación de nuevas variables a través del Editor de Comandos.

Para seleccionar la variable dependiente o variable a explicar por el modelo, el usuario debe elegir una variable dentro de las variables disponibles, situándose con el cursor encima y hacer clic en el botón “**Add**” y dicha variable será seleccionada y aparecerá en el recuadro denominado “dependent variable”. En caso de que el usuario se equivoque al hacer la selección y la variable que aparece como variable dependiente no sea la deseada, tan sólo, debe seleccionar la correcta y hacer de nuevo clic en el botón “**Añadir**” y la nueva variable aparecerá como variable endógena.

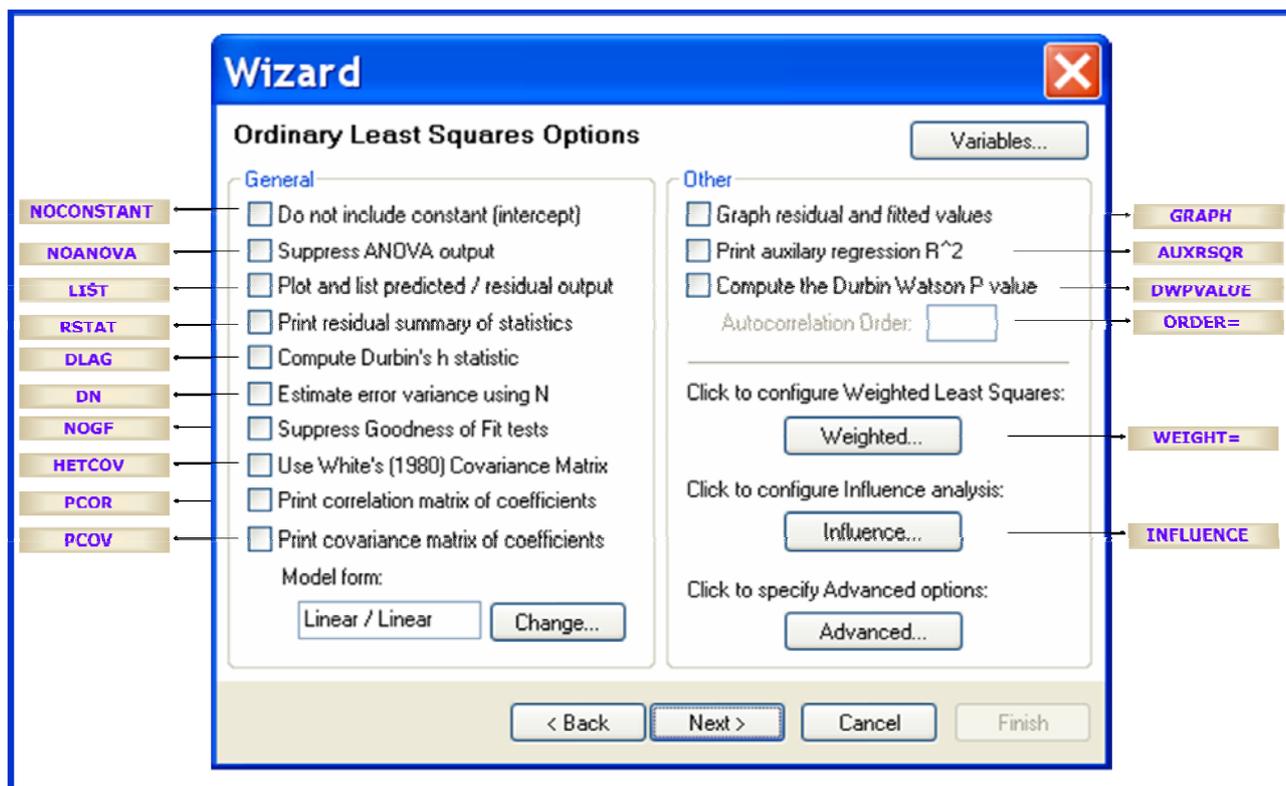


Ilustración 4-3. Opciones del comando OLS.

Para seleccionar las variables independientes o variables explicativas del modelo, el procedimiento es similar, el usuario debe elegir de una en una las variable dentro de las variables disponibles, situándose con el cursor encima y hacer clic en el botón “**Add**” correspondiente a este tipo de variables y que aparece debajo del anterior y, dicha variable será seleccionada y aparecerá en el recuadro denominado “all independent variables”. En caso de que el usuario se equivoque al hacer la selección, tan sólo, debe seleccionarla en el cuadro “all independent variables” y hacer clic en el botón “**Remove**” y la variable dejará de estar seleccionada como variable explicativa.

Además, de seleccionar las variables que intervienen en la estimación, se debe especificar el rango muestral que se desea utilizar para realizar dicha tarea. El usuario puede especificar la observación inicial y final del rango muestral a utilizar en la estimación del modelo, sino lo hace Shazam utiliza toda la muestra (opción seleccionada por defecto).

Una vez seleccionadas las variables → se hace clic en “Next” → se abre el cuadro de diálogo “**Ordinary Least Squares Options**”, donde el usuario debe especificar las opciones con las que desea ejecutar el comando **OLS** (véase Ilustración 4-3).

Haciendo clic en el botón “variables” que aparece en la parte superior derecha de la Ilustración 4-3, el usuario puede guardar, bajo las denominaciones que decida, determinados resultados de la estimación, que podrán ser utilizados para nuevos cálculos. Al realizar los cálculos se debe tener en cuenta la distinta naturaleza de dichos resultados: series, matrices, ... (véase Ilustración 4-4). Una vez dado nombre a los resultados que se deseen guardar → se hace clic en “OK” y se vuelve al cuadro de diálogo anterior “**Ordinary Least Squares Options**” y, se selecciona el botón “Next”, con lo que se llega al cuadro de diálogo final “**Final Step**”. Al hacer clic en el botón “Finish” se sale del asistente y, en la **Ventana Editor Comandos** aparece escrita la instrucción correspondiente.

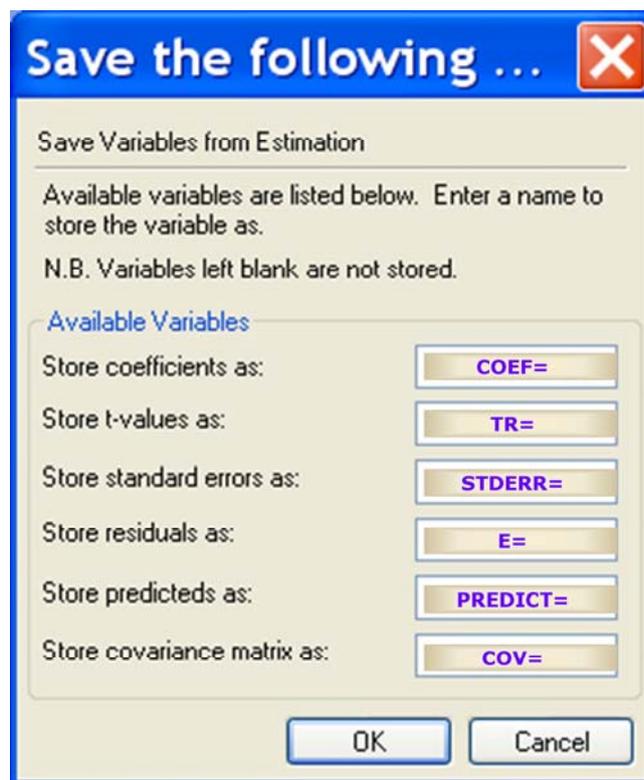


Ilustración 4-4. Opciones para guardar algunos resultados de la estimación.

En el ejemplo recogido en la Ilustración 4-1, la instrucción que aparece en el **Editor de Comandos** es:

```
OLS Y X1 X2 X3 / COEF=B COV=VB TRATIO=TR STDERR=SB RESID=E PREDICT=YE NOANOVA &
NOGF
```

Se le ha indicado a Shazam que estime por MCO un modelo formulado con ordenada en el origen donde la variable a explicar es Y y las variables explicativas son  $X_1$ ,  $X_2$  y  $X_3$ , y que guarde el vector de estimadores bajo la denominación B, la matriz de varianzas-covarianzas estimada de los estimadores como VB, los errores de estimación en E, el regresando estimado en YE, las desviaciones típicas estimadas en SB y las ratios t en TR.

El usuario puede comprobar que cuando muestre los elementos de B, VB, ... con un comando PRINT, dichos elementos tienen más decimales que en la salida estándar del comando OLS, ya que en esta última, por razones de espacio, Shazam ajusta el número de decimales a mostrar.

#### 4.2.2. A través del comando OLS

La otra opción para obtener la estimación MCO es escribir directamente el comando **OLS** en el **Editor de Comandos**, para lo cual el usuario debe conocer su formato de escritura y las opciones disponibles.

El formato del comando OLS es:

```
OLS Y X1 X2 ... XK / OPCIONES
```

En la Ilustración 4-3 y en la Ilustración 4-4 aparecen recogidas en color violeta la denominación de algunas de las opciones que se pueden insertar en un comando **OLS** a través de su wizard. Como ya se ha comentado para utilizar el wizard del comando **OLS** no es necesario conocer su denominación, tan sólo, se debe hacer la elección oportuna en los cuadros de diálogo que se van abriendo y las opciones seleccionadas aparecerán automáticamente en el comando que se inserta en el editor.

<b>NOCONSTANT</b>	→ Estima el modelo de regresión sin incluir ordenada en el origen.
<b>NOANOVA</b>	→ Suprime las tablas de Análisis de la Varianza y los tests de selección de modelos.
<b>LIST</b>	→ Muestra los residuos, los valores estimados de la variable dependiente, algunos estadísticos residuales y un gráfico de los residuos.
<b>RSTAT</b>	→ Muestra algunos estadísticos residuales asociados a la autocorrelación.
<b>DN</b>	→ Utiliza T en lugar de (T-K-1) como divisor para el cálculo de determinados estadísticos.
<b>PCOR</b>	→ Muestra la matriz de correlación de los coeficientes estimados.
<b>PCOV</b>	→ Muestra la matriz de varianzas-covarianzas de los coeficientes estimados.
<b>GRAPH</b>	→ Proporciona dos representaciones gráficas: una de los residuos y otra de los valores observados y estimados de la variable dependiente.
<b>COEF=</b>	→ Guarda los coeficientes estimados (estimadores de los parámetros) en el vector indicado.
<b>COV=</b>	→ Guarda la matriz de varianzas-covarianzas estimada de los estimadores en la matriz indicada.
<b>PREDICT=</b>	→ Guarda los valores estimados del regresando en el vector indicado.
<b>RESID=</b>	→ Guarda los valores de los residuos de la estimación en el vector indicado.
<b>STDERR=</b>	→ Guarda las desviaciones típicas estimadas de los estimadores (errores estándar) en el vector indicado.
<b>TRATIO=</b>	→ Guarda los valores de las ratios t de los coeficientes en el vector indicado.

Cuadro 4-1. Descripción opciones comando OLS.

En el Cuadro 4-1 aparece recogida una breve descripción de algunas de las opciones disponibles con el comando **OLS**.

### 4.3. Análisis de la información básica proporcionada por un comando OLS

La información básica para el análisis de los resultados de una estimación MCO, se puede obtener ejecutando un comando OLS con las opciones **NOANOVA** y **NOGF**:

#### OLS Y $X_1 X_2 \dots X_K$ / **NOANOVA NOGF**

Dichas opciones permiten eliminar de la salida del comando OLS, las tablas ANOVA, los tests de selección de modelos y los tests de bondad de ajuste:

Tabla 4-1. Salida básica de la estimación MCO de un modelo formulado con ordenada en el origen

Tabla 4-1. Salida básica de la estimación MCO de un modelo formulado con ordenada en el origen							
<b>OLS ESTIMATION</b>							
T OBSERVATIONS				Y			
DEPENDENT VARIABLE							
...NOTE...SAMPLE RANGE SET TO:	1,	T	( <b>\$N</b> )				
R-SQUARE				<b>R<sup>2</sup></b>	<b>(<b>\$R2</b>)</b>		
R-SQUARE ADJUSTED				<b><math>\bar{R}^2</math></b>	<b>(<b>\$ADR2</b>)</b>		
VARIANCE OF THE ESTIMATE-SIGMA**2				<b>S<sup>2</sup></b>	<b>(<b>\$SIG2</b>)</b>		
STANDARD ERROR OF THE ESTIMATE-SIGMA				<b>S</b>			
SUM OF SQUARED ERRORS-SSE				<b>SCE</b>	<b>(<b>\$SSE</b>)</b>		
MEAN OF DEPENDENT VARIABLE				<b><math>\bar{Y}</math></b>			
LOG OF THE LIKELIHOOD FUNCTION				<b>L</b>	<b>(<b>\$LLF</b>)</b>		
Variable Name	Estimated Coefficient	Standard Error	T-Ratio	p-Value (DF <b>(<b>\$DF</b>)</b> )	Partial Correlation	Standardised Coefficient	Elasticity at Means
X1	<b>b<sub>1</sub></b>	<b>s<sub>b1</sub></b>	<b>t<sub>1</sub></b>	<b>PVt<sub>1</sub></b>	<b>c<math>\hat{p}</math><sub>1</sub></b>	<b>b<sub>1</sub><sup>*</sup></b>	<b><math>\hat{E}_1</math></b>
X2	<b>b<sub>2</sub></b>	<b>s<sub>b2</sub></b>	<b>t<sub>2</sub></b>	<b>PVt<sub>2</sub></b>	<b>c<math>\hat{p}</math><sub>2</sub></b>	<b>b<sub>2</sub><sup>*</sup></b>	<b><math>\hat{E}_2</math></b>
...	...	...	...	...	...	...	...

Tabla 4-1. Salida básica de la estimación MCO de un modelo formulado con ordenada en el origen

XX	$b_k$	$S_{b_k}$	$t_k$	$PVt_k$	$c\hat{p}_k$	$b_k^*$	$\hat{E}_k$
CONSTANT	$b_0$	$S_{b_0}$	$t_0$	$PVt_0$	$c\hat{p}_0$	$b_0^*$	$\hat{E}_0$

Shazam bajo la denominación CONSTANT proporciona el estimador de la ordenada en el origen y demás cálculos asociados.

Cuando se ejecutan algunos comandos, Shazam guarda en memoria una serie de "variables temporales", que reciben esta denominación precisamente porque sólo están disponibles de forma temporal hasta que se ejecute un nuevo comando que lleve asociadas dichas variables<sup>40</sup>. En la tabla anterior aparecen sombreadas en verde algunas de las variables temporales asociadas a la salida de una estimación MCO. Son variables cuya denominación comienza por \$.

En aras a facilitar el seguimiento por parte del lector, en cada uno de los capítulos se analizarán, tan sólo, los estadísticos directamente relacionados con el tema a tratar.

Como se puede observar, esta salida proporciona, en primer lugar, información sobre el método de estimación empleado, el tamaño muestral, el nombre de la variable dependiente y el rango muestral.

A continuación proporciona las salidas de estimación y algunos estadísticos asociados.

R - SQUARE = Coeficiente de determinación

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

El **coeficiente de determinación** 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.

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). Este es el procedimiento empleado por Shazam para el cálculo de dicho coeficiente.

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 ( $0 \leq R^2 \leq 1$ ). El  $R^2$  toma el valor uno cuando el ajuste es perfecto y cero cuando es pésimo.

R - SQUARE ADJUSTED = Coeficiente de determinación corregido o ajustado

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

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  $R^2$  para comparar la bondad de ajuste de modelos con distinto número de variables explicativas.

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

<sup>40</sup> El valor de las variables temporales se puede visualizar, como ya se ha comentado en el Capítulo 1, en la **Ventana de Variables Temporales** a la que se accede a través del botón **Temp** del **Panel de Depuración**, siempre y cuando los comandos se ejecuten estableciendo líneas de ruptura.

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.

VARIANCE OF THE ESTIMATE - SIGMA \*\*2 = Estimador de la  
varianza de la perturbación o varianza residual

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

STANDARD ERROR OF THE ESTIMATE-SIGMA = Estimador de la  
desviación típica de la perturbación o desviación típica residual

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

El **estimador de la varianza de la perturbación** es otra medida que sirve para analizar la capacidad explicativa del modelo, pues no es más que el error cometido en la estimación, ponderado por los grados de libertad del modelo.

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.

Las características básicas de la variable independiente vienen recogidas por su media y su varianza o cuasivarianza. Recuerde que la varianza y cuasivarianza del regresando coinciden con las de la perturbación.

MEAN OF DEPENDENT VARIABLE = Media de la variable dependiente

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

Además, recuerde que en un modelo formulado con ordenada en el origen la media del regresando coincide con la media de su valor estimado  $\bar{Y} = \bar{\hat{Y}}$ .

SUM OF SQUARD ERRORS - SSE = Suma de Cuadrados de los Errores

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

VARIABLE NAME = Nombre de la variable

ESTIMATED COEFFICIENT = Coeficiente estimado ( $b_i$ )

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

Los **coeficientes estimados** recogen el valor de los estimadores de los parámetros asociados a cada uno de 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, suponiendo que el resto de las variables se mantienen constantes.

Shazam bajo la denominación CONSTANT proporciona el estimador de la ordenada en el origen (que normalmente no tiene interpretación económica) y demás cálculos asociados.

STANDARD ERROR = Error estándar (desviación típica estimada de los estimadores) ( $s_{b_i}$ )

$$\hat{V}(b) = s^2 (X'X)^{-1} = \begin{pmatrix} s_{b_1}^2 & & & & \\ s_{b_1 b_2} & s_{b_2}^2 & & & \\ \dots & \dots & \dots & \dots & \\ s_{b_1 b_k} & s_{b_2 b_k} & \dots & s_{b_k}^2 & \\ s_{b_1 b_0} & s_{b_2 b_0} & \dots & s_{b_k b_0} & s_{b_0}^2 \end{pmatrix}_{(K+1) \times (K+1)} \Rightarrow s_{b_i} = \sqrt{s_{b_i}^2}$$

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

La matriz de varianzas-covarianzas estimada de los estimadores MCO es una matriz de orden (K+1)x(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.

T - RATIO = Ratio t

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

Las **ratios t** se definen como el cociente entre el estimador y su desviación típica estimada. 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 tema de 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 o no individualmente significativa para la explicación del regresando.

#### 4.4. Análisis de los residuos: opción LIST

Una vez estimado el modelo, es importante llevar a cabo un diagnóstico del mismo. El análisis de los residuos MCO juega un papel fundamental en esta tarea. Por ejemplo, nos pueden ayudar a determinar la existencia de valores atípicos, la presencia de problemas de autocorrelación, heterocedasticidad, etc.

La salida estándar asociada a esta opción proporciona, en primer lugar, información sobre el valor de la variable dependiente, su valor estimado y el error cometido en su estimación, para cada una de las observaciones muestrales. Además proporciona un gráfico representando los residuos respecto a su media. Recuérdese que en modelos formulados con ordenada en el origen, la media residual es nula.

**Tabla 4-2. Salida estándar asociada a la opción LIST**

Obs. No.	Observed Value	Predict Value	Calculated Residual		
1	$Y_1$	$\hat{Y}_1$	$e_1$		* I
2	$Y_2$	$\hat{Y}_2$	$e_2$	*	I
...	...	...	...	...	...
T	$Y_T$	$\hat{Y}_T$	$e_T$		I *

DURBIN-WATSON = $dw$	VON NEUMANN RATIO = $v$	RHO = $\hat{\rho}$			
RESIDUAL SUM = $SE$	RESIDUAL VARIANCE = $S^2$	<b>(\$SIG2)</b>			
SUM OF ABSOLUTE ERRORS = $SEA$					
R-SQUARE BETWEEN OBSERVED AND PREDICTED = $r^2$	<b>(\$R2OP)</b>				
RUNS TEST:	$n$ RUNS,	$N_1$ POS,	$N_0$ ZERO,	$N_2$ NEG	NORMAL STATISTIC = $N$

OBS.NO. = N° de observación (t)

OBSERVED VALUE = Valor observado del regresando ( $Y_t$ )

PREDICTED VALUE = Valor estimado del regresando

$$\hat{Y}_t = X_t' b \Rightarrow \hat{Y} = Xb$$

CALCULATED RESIDUAL = Residuo

$$e_t = Y_t - \hat{Y}_t \Rightarrow e = Y - \hat{Y}$$

Se denomina **residuo** o **error** ( $e_t$ ) a la diferencia entre el valor observado del regresando ( $Y_t$ ) y el valor estimado por el modelo ( $\hat{Y}_t$ ).

RESIDUAL SUM = Suma residual

$$SE = \sum_{t=1}^T e_t$$

Además, proporciona la suma residual, la varianza residual, la suma de errores absolutos y el coeficiente de determinación entre observados y estimados. Para modelos formulados con ordenada en el origen la suma residual es nula (propiedad de los residuos MCO).

RESIDUAL VARIANCE = Varianza residual

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

La varianza residual aparece también en la primera parte de la salida del comando OLS, bajo la denominación VARIANCE OF THE ESTIMATE-SIGMA\*\*2.

SUM OF ABSOLUTE ERRORS = Suma de errores absolutos

$$SEA = \sum_{t=1}^T |e_t|$$

En la suma de errores absolutos se evita que los residuos de distinto signo se compensen entre sí, por tanto, esta suma tan sólo será nula si todos los residuos lo son.

R - SQUARE BETWEEN OBSERVED AND PREDICTED = Coeficiente

de determinación entre valores observados y estimados

$$r^2 = \frac{\left[ \sum_{t=1}^T (\hat{Y}_t - \bar{Y})(Y_t - \bar{Y}) \right]^2}{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 \sum_{t=1}^T (Y_t - \bar{Y})^2}$$

El coeficiente de determinación entre observados y estimados ( $r^2$ ) no es más que el coeficiente de determinación resultante de la regresión del regresando sobre el regresando estimado:  $Y_t = \alpha_0 + \alpha_1 \hat{Y}_t + u_t$ .

$$r^2 = \frac{\left[ \sum_{t=1}^T (\hat{Y}_t - \bar{Y})(Y_t - \bar{Y}) \right]^2}{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 \sum_{t=1}^T (Y_t - \bar{Y})^2} = \frac{\left[ \sum_{t=1}^T (\hat{Y}_t - \bar{Y})(Y_t - \bar{Y}) \right]^2}{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 \sum_{t=1}^T (Y_t - \bar{Y})^2} = \frac{SCR^2}{SCR \cdot SCT} = \frac{SCR}{SCT} = R^2$$

Cuando el modelo está formulado con ordenada en el origen,  $\bar{Y}$  coincide con  $\bar{Y}$  y, por tanto,  $r^2$  coincide con  $R^2$ .

Debe de tenerse en cuenta que cuando se ejecuta la opción LIST del comando OLS se activa automáticamente la opción RSTAT, que proporciona resultados asociados a la autocorrelación (dw, v,  $\hat{\rho}$  y test de rachas).

Señalar que \$K es el nº de regresores del modelo (K+1) y \$N es el número de observaciones (tamaño muestral). Si se quiere disponer de estas variables en cualquier momento es necesario guardarlas, lo que se hace con un comando GEN1 por tratarse de escalares.

Cuando en un comando GENR o GEN1 interviene alguna “variable temporal (\$)”, Shazam recuerda el valor numérico de esa variable temporal, porque hace referencia al último de los comandos ejecutados que llevan asociados dichas variables y el usuario podría haber ejecutado otro comando que tuviese disponibles las mismas variables temporales, por lo que la información a utilizar podría no ser la deseada. Si el usuario no está interesado en comprobar dichos valores, puede ejecutar dichos comandos precedidos con un signo “?”.

## 4.5. Análisis de las sumas de cuadrados: opción 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. Dicha variabilidad puede considerarse respecto al origen o respecto a la media, dando lugar a dos tipos de tablas ANOVA. En Shazam Professional la opción ANOVA se ejecuta por defecto, es decir, que si el usuario no desea que en la salida del comando OLS aparezcan

dichas tablas será necesario que lo ejecute con la opción NOANOVA. En la Tabla 4-3 se recoge la salida estándar de la opción ANOVA del comando OLS.

Tabla 4-3. Salida estándar asociada a la opción ANOVA				
<b>MODEL SELECTION TESTS - SEE JUDGE ET AL. (1985,P.242)</b>				
AKAIKE (1969) FINAL PREDICTION ERROR – FPE				PC
(FPE IS ALSO KNOWN AS AMEMIYA PREDICTION CRITERION - PC)				
AKAIKE (1973) INFORMATION CRITERION – LOG AIC				LAIC
SCHWARZ (1978) CRITERION - LOG SC				LSC
<b>MODEL SELECTION TESTS - SEE RAMANATHAN (1998,P.165)</b>				
CRAVEN-WAHBA (1979)				
GENERALIZED CROSS VALIDATION – GCV				GCV
HANNAN AND QUINN (1979) CRITERION				HQ
RICE (1984) CRITERION				RICE
SHIBATA (1981) CRITERION				SHIB
SCHWARZ (1978) CRITERION – SC				SC
AKAIKE (1974) INFORMATION CRITERION – AIC				AIC
<b>ANALYSIS OF VARIANCE – FROM MEAN</b>				
	<b>SS</b>	<b>DF</b>	<b>MS</b>	<b>F</b>
REGRESSION	SCR <b>(\$SSR)</b>	K	SCR/K	$F_2$
ERROR	SCE <b>(\$SSE)</b>	T-K-1	SCE/(T-K-1)	P-VALUE
TOTAL	SCT <b>(\$SST)</b>	T-1	SCT/(T-1)	$PVF_2$
<b>ANALYSIS OF VARIANCE – FROM ZERO</b>				
	<b>SS</b>	<b>DF</b>	<b>MS</b>	<b>F</b>
REGRESSION	$\sum \hat{Y}_t^2$ <b>(\$ZSSR)</b>	K+1	$\sum \hat{Y}_t^2 / K+1$	$F_1$
ERROR	SCE <b>(\$SSE)</b>	T-K-1	SCE/(T-K-1)	P-VALUE
TOTAL	$\sum Y_t^2$ <b>(\$ZSST)</b>	T	$\sum Y_t^2 / T$	$PVF_1$

Teniendo en cuenta que los estadísticos de selección de modelos y los estadísticos F no forman parte del objetivo de este capítulo, se centrará la atención en la primera columna de las tablas de Análisis de la Varianza.

### 4.5.1. Tabla ANOVA respecto a la media

En la tabla ANOVA respecto a la media aparece recogida la siguiente información:

- (a) 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 ( $\bar{Y}$ ).

Por estar el modelo formulado con ordenada en el origen, la

media del regresando coincide con la media del regresando estimado ( $\bar{\hat{Y}} = \bar{Y}$ ) y, por tanto, se

puede calcular como  $SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$ .

REGRESION	$SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$
ERROR	$SCE = \sum_{t=1}^T (e_t - \bar{e})^2 = \sum_{t=1}^T e_t^2 = e'e$
TOTAL	$SCT = \sum_{t=1}^T (Y_t - \bar{Y})^2$

- (b) La Suma de Cuadrados de los Errores (SCE), que es la suma de cuadrados de las desviaciones de los valores de los residuos ( $e_t$ ) respecto a su media muestral ( $\bar{e}$ ), que en modelos formulados con ordenada en el origen es nula, por lo que SCE no es más que la suma de residuos al cuadrado.
- (c) 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 ( $\bar{Y}$ ).

Por estar el modelo formulado 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.

#### 4.5.2. Tabla ANOVA respecto al cero

Nótese que la única diferencia con la tabla ANOVA respecto a la media es que en esta tabla, la suma de cuadrados de las desviaciones no se calculan respecto a las medias sino respecto al cero. Además, puede observarse que la suma de cuadrados de las desviaciones de los errores 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.

REGRESION	$\sum_{t=1}^T (\hat{Y}_t - 0)^2 = \sum_{t=1}^T \hat{Y}_t^2$
ERROR	$SCE = \sum_{t=1}^T (e_t - 0)^2 = \sum_{t=1}^T e_t^2$
TOTAL	$\sum_{t=1}^T (Y_t - 0)^2 = \sum_{t=1}^T Y_t^2$

En la tabla **ANOVA respecto al cero** aparece recogida la siguiente información:

- (b) La suma de los cuadrados de los valores estimados del regresando ( $\sum_{t=1}^T \hat{Y}_t^2$ ).
- (c) La suma de los cuadrados de los errores ( $\sum_{t=1}^T e_t^2$ ).
- (d) La suma de cuadrados de los valores observados del regresando ( $\sum_{t=1}^T Y_t^2$ ).

La descomposición  $\sum_{t=1}^T Y_t^2 = \sum_{t=1}^T \hat{Y}_t^2 + \sum_{t=1}^T e_t^2$  se cumple tanto si el modelo está formulado con o sin ordenada en el origen.

Shazam proporciona la tabla ANOVA respecto al cero para modelos con ó sin ordenada en el origen, mientras que la tabla ANOVA respecto a la media sólo la proporciona si el modelo incluye regresor ficticio. Obsérvese que la SCE se puede obtener en ambas tablas ANOVA.

#### 4.6. Análisis gráfico: opción GRAPH

Para dar una primera idea 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. La opción GRAPH del comando OLS proporciona dos representaciones gráficas: una de los residuos y otra de los valores observados y estimados de la variable dependiente del modelo.

Para poder modificar las propiedades de dichos gráficos es necesario incorporarlos al Project de trabajo, únicamente de este modo se puede acceder a los distintos botones de la **Ventana Gráfico** o **Shazam Graph** y cambiar su apariencia. Utilizando esta posibilidad, a los gráficos recogidos en la Ilustración 4-5, además de trazarles el “enrejado”, se le ha cambiado el título y la denominación de los ejes, tal y como se recoge en la Ilustración 4-7.

El gráfico de valores observados y estimados permite ver 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 valores observados–valores estimados recogido en la parte derecha de la Ilustración 4-7, el modelo parece recoger de forma bastante razonable la evolución de la endógena puesto que no se observan grandes errores.

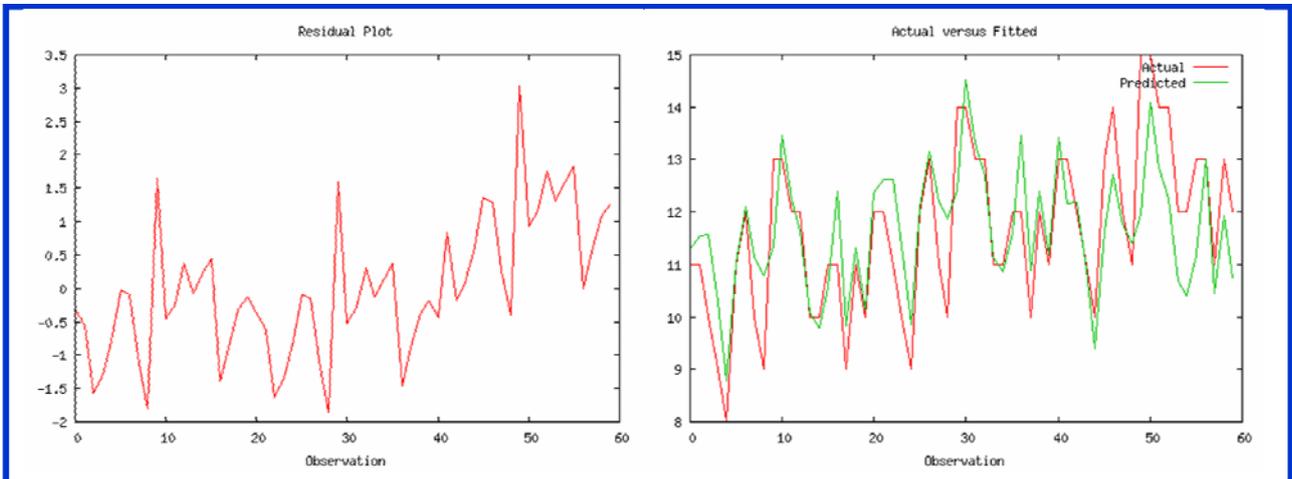


Ilustración 4-5. Gráficos asociados a la opción GRAPH del comando OLS.

Para facilitar el análisis del gráfico de residuos puede resultar conveniente trazar en dicho gráfico líneas adicionales (todas ellas paralelas al eje horizontal):

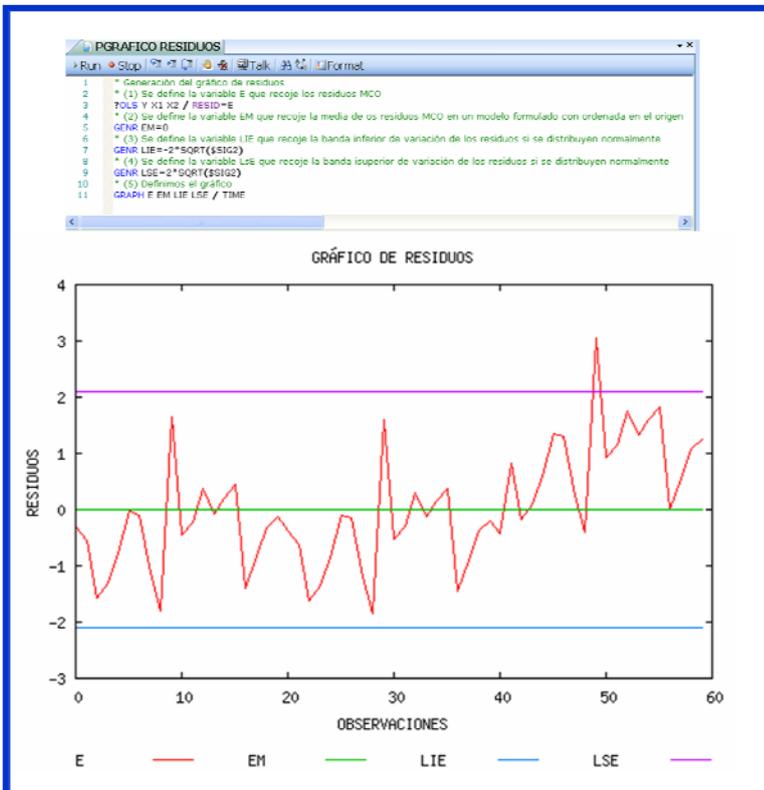


Ilustración 4-6. Gráfico de residuos con bandas de confianza del 95%.

- Una que pase por el valor medio de los residuos (cero dado que se asume que el modelo está formulado con ordenada en el origen). Esto 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 su desviación típica. Dicha banda permitirá analizar de forma sencilla si los residuos se distribuyen o no como una normal. Si los residuos siguiesen una distribución normal, sus valores estarían con un 95% de probabilidad comprendidos dentro de esta banda.

Este gráfico puede ayudar a determinar la existencia de valores u observaciones anómalas. Por ejemplo, observando el gráfico de residuos recogido en la Ilustración 4-6, uno de los errores supera la banda de confianza y quizás sea necesario darle un tratamiento diferenciado del resto. Además, dicho gráfico, puede ayudar a vislumbrar problemas de autocorrelación tal y como se analizará en el capítulo correspondiente. No obstante, 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

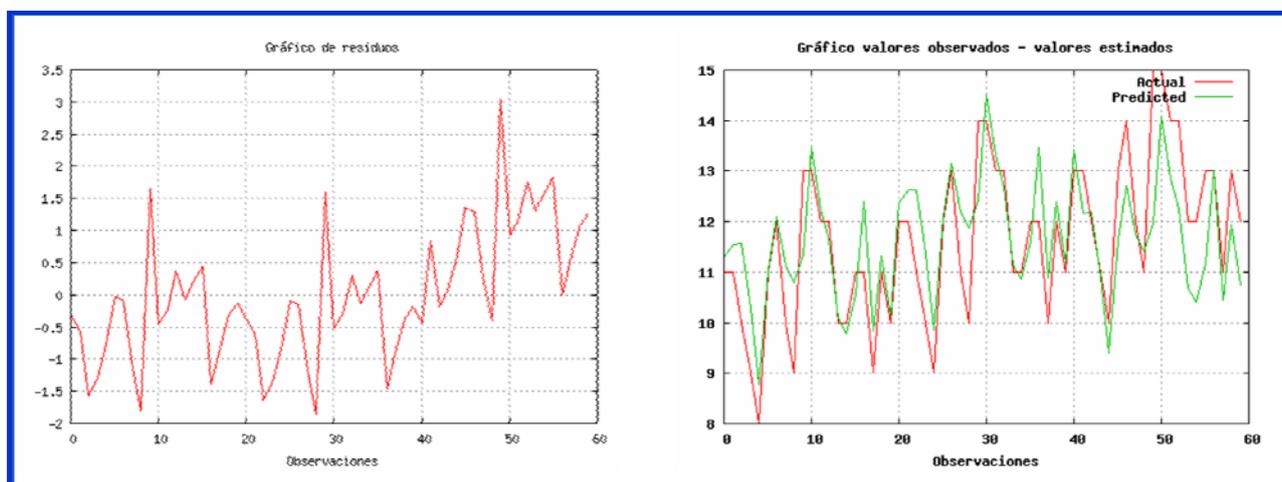


Ilustración 4-7. Gráficos asociados a la opción GRAPH del comando OLS modificados a partir de su Ventana Gráfico.

difícilmente serán concluyentes.

A pesar de que los gráficos en la Versión Professional son mucho mejores que en la Versión Standard, todavía siguen siendo muy restrictivos. No obstante, dado que a cada gráfico Shazam le asocia un fichero de datos y que permite guardar dichos datos en múltiples formatos (.xls, .txt, .csv, ...), el usuario de una manera relativamente sencilla puede utilizar otros programas más potentes en gráficos. En la Ilustración 4-8 aparece el gráfico de residuos realizado con Excel y para cuya representación sólo se necesita la serie de residuos, puesto que el resto de los atributos son opciones que por defecto ofrece el programa.

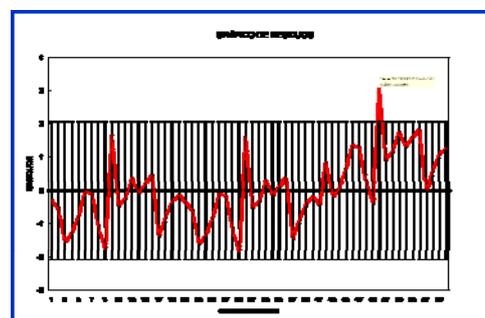


Ilustración 4-8. Gráfico de residuos con Excel.

#### 4.7. Estimación MCO de un modelo formulado sin ordenada en el origen

Aunque no es lo habitual, algunos modelos pueden 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 en el epígrafe 4.1. (modelo formulado con ordenada en el origen) es que no aparece el parámetro  $\beta_0$ .

Matricialmente 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 se debe incluir como opción NOCONSTANT.

**Tabla 4-4. Salida básica de la estimación MCO de un modelo formulado sin ordenada en el origen**

<b>OLS ESTIMATION</b>							
T OBSERVATIONS DEPENDENT VARIABLE	Y						
...NOTE...SAMPLE RANGE SET TO: 1, T	(\$N)						
R-SQUARE	R <sup>2</sup> (\$R2)						
R-SQUARE ADJUSTED	R̄ <sup>2</sup> (\$ADR2)						
VARIANCE OF THE ESTIMATE-SIGMA**2	S <sup>2</sup> (\$SIG2)						
STANDARD ERROR OF THE ESTIMATE-SIGMA	S						
SUM OF SQUARED ERRORS-SSE	SCE (\$SSE)						
MEAN OF DEPENDENT VARIABLE	Ȳ						
LOG OF THE LIKELIHOOD FUNCTION	L (\$LLF)						
RAW MOMENT R-SQUARE	R <sup>2</sup> <sub>RAW</sub> (\$RAW)						
Variable Name	Estimated Coefficient	Standard Error	T-Ratio	p-Value (DF)	Partial Correlation	Standardised Coefficient	Elasticity at Means
X1	b <sub>1</sub>	s <sub>b<sub>1</sub></sub>	t <sub>1</sub>	PVt <sub>1</sub>	c $\hat{p}$ <sub>1</sub>	b <sup>*</sup> <sub>1</sub>	E $\hat{E}$ <sub>1</sub>
X2	b <sub>2</sub>	s <sub>b<sub>2</sub></sub>	t <sub>2</sub>	PVt <sub>2</sub>	c $\hat{p}$ <sub>2</sub>	b <sup>*</sup> <sub>2</sub>	E $\hat{E}$ <sub>2</sub>
...	...	...	...	...	...	...	...
XK	b <sub>k</sub>	s <sub>b<sub>k</sub></sub>	t <sub>k</sub>	PVt <sub>k</sub>	c $\hat{p}$ <sub>k</sub>	b <sup>*</sup> <sub>k</sub>	E $\hat{E}$ <sub>k</sub>

La salida que proporciona el comando OLS con la opción NOCONSTANT es similar a su salida estándar, siendo la única diferencia que incorpora el coeficiente de determinación bruto (RAW MOMENT R-SQUARE).

El **coeficiente de determinación** en modelos formulados sin regresor ficticio ya no mide la proporción en que la varianza muestral del regresando (SCT/T) es explicada por la varianza del regresando estimado (SCR/T), por tanto no puede interpretarse como el porcentaje de variaciones del regresando explicado por las variaciones de las variables explicativas.

Shazam utiliza siempre la primera expresión para el cálculo del coeficiente de determinación, tanto si el modelo está formulado con ordenada en el origen como si no lo está. Por ello, se debe ser cauteloso a la hora de interpretar dicho coeficiente, puesto que en modelos formulados sin ordenada en el origen no se cumple la descomposición de SCT en SCE y SCR, por lo que la interpretación habitual del R<sup>2</sup> no sería correcta.

En este tipo de modelos, esta medida de bondad de ajuste pierde una de sus ventajas, sigue teniendo la unidad como límite superior, pero carece de límite inferior, pudiendo tomar incluso valores negativos.

Además, en este caso, el coeficiente de determinación entre observados y estimados no coincide con  $R^2$  por estar el modelo formulado sin ordenada en el origen.

Por ello, en modelos formulados sin ordenada en el origen Shazam proporciona, además, el coeficiente de determinación bruto, que aunque si está acotado entre cero y uno, también carece de la interpretación habitual. Ello se debe a que su cálculo no se basa en la descomposición de SCT en SCE y SCR (véase tabla ANOVA respecto a la media), sino en la descomposición de  $Y'Y$  en  $\hat{Y}'\hat{Y}$  y  $e'e$  (véase tabla ANOVA respecto al cero), la cuál se cumple independientemente de que el modelo esté formulado con o sin regresor ficticio.

RAW MOMENT R - SQUARE = Coeficiente de determinación bruto (SRAW)

$$R_{RAW}^2 = 1 - \frac{SCE}{\sum_{t=1}^T Y_t^2}$$

El lector debe de ser consciente de 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.

A la hora de calcular el **estimador de la varianza de la perturbación** se debe tener en cuenta que los grados de libertad de este tipo de modelos son  $T-K$ , puesto que ahora el número de regresores del modelo 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  $\bar{Y} \neq \bar{\hat{Y}}$ .

VARIANCE OF THE ESTIMATE - SIGMA \* \*2 = Estimador de la varianza de la perturbación o varianza residual

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

STANDARD ERROR OF THE ESTIMATE - SIGMA = Estimador de la desviación típica de la perturbación o desviación típica residual

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

ARIABLE NAME = Nombre de la variable

ESTIMATED COEFFICIENT = 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 asociados a cada una de las variables explicativas.

STANDARD ERROR = Error estándar (desviación típica estimada de los estimadores) ( $s_{b_i}$ )

$$\hat{V}(b) = S^2 (X'X)^{-1} = \begin{pmatrix} s_{b_1}^2 & & \\ s_{b_1 b_2} & s_{b_2}^2 & \\ \dots & \dots & \\ s_{b_1 b_K} & s_{b_2 b_K} & \dots & s_{b_K}^2 \end{pmatrix}_{K \times K} \Rightarrow s_{b_i} = \sqrt{s_{b_i}^2}$$

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

La matriz de varianzas-covarianzas estimada de los estimadores MCO 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.

### 4.8. Interpretación de los coeficientes

En un Modelo de Regresión Lineal Múltiple a los parámetros que acompañan a las variables explicativas también 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:

$$\beta_i = \frac{\partial Y_t}{\partial X_{it}} \quad i = 1, 2, \dots, K \quad \text{o bien} \quad \beta_i = \frac{\partial Y}{\partial X_i} \quad i = 1, 2, \dots, K, \quad \text{dado que es independiente de la}$$

observación en la que se encuentre.

En este sentido, la estimación de  $\beta_i$  puede considerarse como 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 interpretaría como el cambio en la variable explicada producido por un cambio unitario en la variable explicativa a la que acompañan, 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), y es precisamente esta hipótesis de independencia lineal de los regresores la que hace posible aislar el efecto de cada uno de ellos. No obstante, las buenas propiedades del vector de estimadores se siguen cumpliendo siempre y cuando la dependencia lineal entre los regresores no sea exacta (multicolinealidad aproximada), sin embargo, hay que señalar que en estas situaciones la interpretación de los coeficientes debe hacerse con cautela, puesto que éstos 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 bastante difícil encontrar regresores completamente ortogonales, lo habitual es que exista un determinado grado de dependencia lineal entre ellos, el cual debe procurarse que no sea demasiado grande, para que su efecto no sea demasiado pernicioso.

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

Véase que en la salida del comando OLS cada coeficiente de regresión parcial estimado ( $b_i$ ) [ESTIMATED COEFFICIENT] viene acompañado por su error estándar ( $S_{b_i}$ ) [STANDARD ERROR], que indica la precisión de la estimación y por su ratio t ( $t_i$ ) [T-RATIO], que indica su grado de significación.

Para poder enriquecer la interpretación económica y/o empresarial de los resultados de la estimación de un modelo de regresión, en la salida del comando OLS, Shazam proporciona además de los estimadores de los coeficientes de regresión parcial, los estimadores de los coeficientes estandarizados (STANDARDIZED COEFFICIENT) y de las elasticidades en media (ELASTICITY AT MEANS). Tanto los coeficientes estandarizados como las elasticidades en media no dependen de las unidades de medida de las variables a las que acompañan, es decir, son adimensionales por lo que son directamente comparables entre sí.

Los **coeficientes beta** se estiman a través de los coeficientes de regresión parcial ( $b_i$ ) ajustados por el cociente entre la desviación estándar de la variable independiente y la desviación estándar de la variable dependiente  $\left( \frac{S_{X_i}}{S_Y} \right)$ , por lo que miden la importancia relativa de las variables independientes:

$$b_i^* = b_i \frac{S_{X_i}}{S_Y} \quad i = 0, 1, 2, \dots, K$$

Los coeficientes beta o coeficientes estandarizados permiten determinar cual es la variable explicativa que tiene mayor peso para la explicación del regresando<sup>41</sup>.

La interpretación 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) producido por un cambio unitario en la variable independiente a la que acompaña (en unidades de desviación estándar) manteniendo constantes las demás variables.

La **elasticidad en media** mide 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<sup>42</sup>.

Las elasticidades en media se estiman a través de los coeficientes de regresión parcial ( $b_i$ ) ajustados por el cociente entre la media de la variable independiente y la media de la variable dependiente  $\left(\frac{\bar{X}_i}{\bar{Y}}\right)$ , por lo que miden la sensibilidad de la variable dependiente a los cambios en las variables

independientes:  $\hat{E}_i = b_i \frac{\bar{X}_i}{\bar{Y}}$ .

#### 4.9. Formas funcionales alternativas

En algunos casos una forma funcional lineal no caracterizará 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. 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.

Este tipo de no linealidad en las variables se puede incorporar en el modelo sin demasiada dificultad, 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.

En este apartado se analizarán algunas de las formas funcionales no lineales en las variables pero lineales en los parámetros más habituales, formas 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, modelo log-lin, modelo lin-inv y modelo log-inv.

<sup>41</sup> Los coeficientes beta estimados se pueden obtener a partir de la estimación del siguiente modelo transformado:

$$\frac{Y_i - \bar{Y}}{S_y} = \beta_1 \frac{X_{1i} - \bar{X}_1}{S_{x_1}} + \beta_2 \frac{X_{2i} - \bar{X}_2}{S_{x_2}} + \dots + \beta_k \frac{X_{ki} - \bar{X}_k}{S_{x_k}} + \frac{\varepsilon_i}{S_y}$$

donde las variables han sido estandarizadas, es decir, se les ha restado su media y se ha dividido por su desviación estándar.

La estandarización de las variables facilita la comparación entre los distintos coeficientes, aunque la interpretación debe hacerse ahora en términos de desviación estándar.

<sup>42</sup> Las elasticidades medias estimadas se pueden obtener a partir de la estimación del siguiente modelo transformado:

$$\ln Y_i = \beta_0 + E_1 \ln X_{1i} + E_2 \ln X_{2i} + \dots + E_k \ln X_{ki} + \varepsilon_i$$

donde las variables han sido transformadas aplicándole logaritmos neperianos.

$$\text{Modelo Lineal} \rightarrow Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \beta_0 + \varepsilon_t$$

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”:  $b_i = \frac{\partial Y}{\partial X_i} \Rightarrow b_i = \frac{\Delta Y}{\Delta X_i}$

$$\text{Modelo Lin-Log} \rightarrow Y_t = \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \dots + \beta_k \ln X_{kt} + \beta_0 + \varepsilon_t$$

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 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:  $b_i = \frac{\partial Y}{\partial \ln X_i} \Rightarrow b_i = \frac{\Delta Y}{\frac{\Delta X_i}{X}}$

$$\text{Modelo Log-Log} \rightarrow \ln Y_t = \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \dots + \beta_k \ln X_{kt} + \beta_0 + \varepsilon_t$$

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 también se podría interpretar como el cambio porcentual producido en el regresando ante un cambio porcentual en la variable explicativa

correspondiente:  $b_i = \frac{\partial \ln Y}{\partial \ln X_i} \Rightarrow b_i = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_i}{X}}$

$$\text{Modelo Log-Lin} \rightarrow \ln Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \beta_0 + \varepsilon_t$$

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:

$$b_i = \frac{\partial \ln Y}{\partial X_i} \Rightarrow b_i = \frac{\frac{\Delta Y}{Y}}{\Delta X_i}$$

$$\text{Modelo Lin-Inv} \rightarrow Y_t = \beta_1 \frac{1}{X_{1t}} + \beta_2 \frac{1}{X_{2t}} + \dots + \beta_k \frac{1}{X_{kt}} + \beta_0 + \varepsilon_t$$

Bajo la cláusula “ceteris paribus”, 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 ponderado por su inversa cambiada de signo:

$$b_i = \frac{\partial Y}{\partial \left( \frac{1}{X_i} \right)} \Rightarrow b_i = - \frac{\Delta Y}{\frac{\Delta X_i}{X_i^2}} = \frac{\Delta Y}{X_i} \left( - \frac{1}{X_i} \right)$$

**Modelo Log-Inv**  $\rightarrow \ln Y_t = \beta_1 \left( -\frac{1}{X_{1t}} \right) + \beta_2 \left( -\frac{1}{X_{2t}} \right) + \dots + \beta_k \left( -\frac{1}{X_{kt}} \right) + \beta_0 + \varepsilon_t$

Bajo la cláusula “ceteris paribus”, 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 ponderado por su inversa:  $b_i = \frac{\partial \ln Y}{\partial \left( -\frac{1}{X_i} \right)} \Rightarrow b_i = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_i}{X_i^2}} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_i}{X_i} \frac{1}{X_i}}$

Como se acaba de ver, una gran número de modelos econométricos son fácilmente linealizables mediante una sencilla transformación de los datos y, por tanto, no será necesario recurrir a métodos distintos del MCO para obtener dicha estimación. Para que las elasticidades en media calculadas por Shazam sean correctas, será necesario que el usuario informe a Shazam de cual es la transformación que se ha hecho en las variables. Esto se puede hacer (Ilustración 4-9) a través del cuadro de diálogo “**Least Squares Options**” del wizard “**Ordinary Least Squares Regression**”, pinchando el botón “**Change ...**” y eligiendo la transformación adecuada. Por defecto, Shazam realiza la estimación considerando que tanto la variable dependiente como las independientes son lineales.

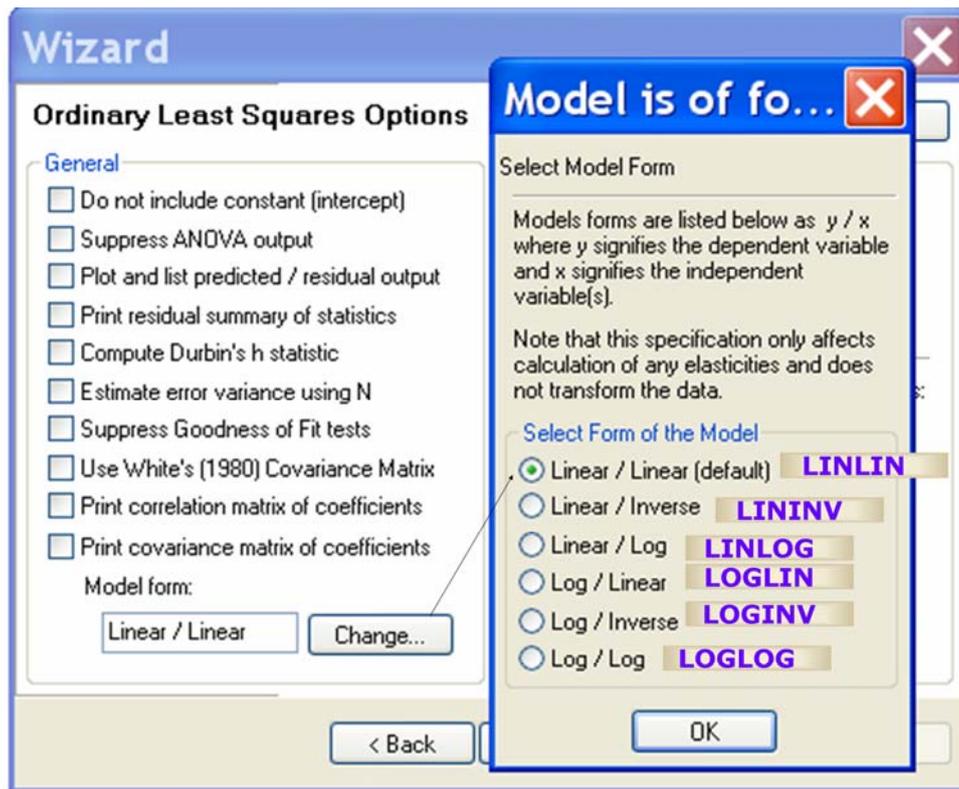


Ilustración 4-9. Transformaciones lineales.

En la Ilustración 4-9, también se recoge la denominación de las diferentes opciones que aparecen en el comando OLS dependiendo de la transformación elegida en las variables y que el usuario podría escribir directamente en un comando OLS sin necesidad de acudir a dicho asistente de escritura. Además, en el Cuadro 4-2 se recoge una breve descripción de dichas opciones y la expresión que se utiliza en cada uno de los casos para el cálculo de las elasticidades.

- LINLIN** → Opción que debe ser utilizada cuando tanto la variable dependiente como las independientes son lineales. En este caso, las elasticidades son estimadas como  $\hat{\epsilon}_i = b_i \frac{X_i}{Y}$ . Es la que selecciona Shazam por defecto.
- LINLOG** → Opción que debe ser utilizada cuando la variable dependiente es lineal y las independientes están expresadas en términos logarítmicos. En este caso, las elasticidades son estimadas como  $\hat{\epsilon}_i = \frac{b_i}{Y}$ .
- LOGLOG** → Opción que debe ser utilizada cuando la variable dependiente y todas las independientes están expresadas en términos logarítmicos. En este caso, las elasticidades son estimadas como  $\hat{\epsilon}_i = b_i$ .
- LOGLIN** → Opción que debe ser utilizada cuando la variable dependiente está expresada en forma logarítmica y las independientes están expresadas en forma lineal. En este caso, las elasticidades son estimadas como  $\hat{\epsilon}_i = b_i X_i$ .
- LOGINV** → Opción que debe ser utilizada cuando la variable dependiente está expresada en forma logarítmica y las independientes están expresadas en forma inversa. En este caso, las elasticidades son estimadas como  $\hat{\epsilon}_i = \frac{b_i}{X_i}$ .
- LININV** → Opción que debe ser utilizada cuando la variable dependiente es lineal y las independientes están expresadas en forma inversa. En este caso, las elasticidades son estimadas como  $\hat{\epsilon}_i = -\frac{b_i}{Y X_i}$ .

Cuadro 4-2. Opciones transformaciones lineales.

# Capítulo 5. CONTRASTES DE HIPÓTESIS Y REGIONES DE CONFIANZA

## 5.1. Hipótesis del Modelo de Regresión Lineal Normal Clásico (MRLNC)

Las hipótesis del MRLC introducidas en el capítulo 4 son suficientes para obtener estimadores puntuales de los parámetros del modelo, pero si se quiere abordar la estimación por intervalo y/o contrastes de hipótesis, es necesario hacer referencia a la distribución de probabilidad de las perturbaciones.

### H5. Hipótesis de normalidad de las perturbaciones:

$$\varepsilon \xrightarrow{sd} N(0_{T \times 1}, \sigma^2 I_T) \Rightarrow \varepsilon_t \xrightarrow{sd} N(0, \sigma^2)$$

Esta hipótesis establece que el vector de perturbaciones tiene 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 en el caso de 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.

## 5.2. 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 denominado **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 no es más que 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 no es más que la probabilidad real de obtener un valor del estadístico de prueba menor o igual que el obtenido.

Un procedimiento general para efectuar contrastes de hipótesis 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 impuestas a los parámetros (q) y su número de columnas con el número de regresores del modelo econométrico  $(k+1)^{43}$  y r es un vector columna cuyo número de filas coincide con el número de restricciones. El número de restricciones no puede ser, en ningún caso, superior al número de regresores del modelo.

$$\text{Si } H_0 \text{ es cierta } \Rightarrow F = \frac{\frac{(Rb - r)' [R(X'X)^{-1}R']^{-1} (Rb - r)}{q}}{\frac{SCE}{T - k - 1}} \xrightarrow{sd} F_{T-k-1}^q$$

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 el siguiente 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  $H_0$ .

### 5.2.1. Contraste de nulidad individual o contraste de nulidad para un parámetro

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

$$\Rightarrow H_0 : R\beta = r \Rightarrow (0 \ 0 \ \dots \ 1 \ \dots \ 0)_{1 \times (k+1)} \begin{pmatrix} \beta_1 \\ \dots \\ \beta_i \\ \dots \\ \beta_k \\ \beta_0 \end{pmatrix}_{(k+1) \times 1} = 0_{1 \times 1}$$

$$H_1 : R\beta \neq r$$

<sup>43</sup> Se asume que el modelo está formulado con ordenada en el origen, en caso contrario, el número de regresores sería igual a k.

Dado que en la hipótesis nula se establece una única restricción relativa a la nulidad del parámetro  $\beta_i$  del modelo:

- La matriz R se convierte en un vector fila de (k+1) columnas con un uno en el lugar correspondiente al parámetro iésimo (parámetro cuya nulidad se desea contrastar) y el resto de sus elementos nulos. Por tanto, en este caso, el producto  $R\beta$  selecciona el elemento  $\beta_i$  del vector  $\beta$ .
- r es un escalar y es igual a cero dado que es el valor que se establece para la restricción en la hipótesis nula.
- El producto  $R(X'X)^{-1}R'$  selecciona el elemento iésimo de la diagonal de la inversa de X'X (elemento  $x^{ii}$ ).

Por ello, el estadístico F se puede calcular como la la ratio entre el cuadrado del estimador del parámetro cuya nulidad se quiere contrastar y su varianza estimada.

Por tanto, si  $H_0$  es cierta, el estadístico F se distribuye como una F de Snedecor de 1 grado de libertad en el numerador y (T-k-1) grados de libertad en el denominador:

$$F = \frac{\frac{b_i'(X^{ii})^{-1}b_i}{SCE}}{T-k-1} = \frac{b_i^2}{S^2 X^{ii}} = \frac{b_i^2}{S_{b_i}^2} \xrightarrow{sd} F_{T-k-1}^1$$

Dada la equivalencia existente entre una distribución F de Snedecor y una distribución t de Student, también se podría utilizar esta distribución para efectuar este contraste:

$$\sqrt{F} = t_i = \frac{b_i}{S_{b_i}} \xrightarrow{sd} t_{T-K-1}$$

### 5.2.2. Contraste de nulidad conjunta para todos los parámetros del modelo

$$\begin{aligned} H_0 : \beta &= 0_{(k+1) \times 1} \\ H_1 : \beta &\neq 0_{(k+1) \times 1} \end{aligned} \Rightarrow \begin{aligned} H_0 : \beta_1 = \dots = \beta_k = \beta_0 = 0 \\ H_1 : \text{alguno o todos} \neq 0 \end{aligned}$$

$$\Rightarrow H_0 : R\beta = r \Rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{(k+1) \times (k+1)} \begin{pmatrix} \beta_1 \\ \dots \\ \beta_i \\ \dots \\ \beta_k \\ \beta_0 \end{pmatrix}_{(k+1) \times 1} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \dots \\ 0 \end{pmatrix}_{(k+1) \times 1}$$

$H_1 : R\beta \neq r$

Dado que en la hipótesis nula se establece (k+1) restricciones relativas a la nulidad conjunta de todos los parámetros del modelo:

- La matriz R se convierte en una matriz de (k+1) filas y (k+1) columnas, donde todos los elementos de su diagonal principal son iguales a uno y los elementos no diagonales son iguales a cero, es decir, R se convierte en una matriz identidad de orden (k+1). Por tanto, en este caso, el producto  $R\beta$  selecciona el vector  $\beta$ .

- r es un vector columna que contiene (k+1) ceros, dado que es el valor que se fija para las (k+1) restricciones que se establecen en la hipótesis nula.
- El producto  $R(X'X)^{-1}R'$  selecciona todas las filas y columnas de la inversa de la matriz X'X (matriz  $(X'X)^{-1}$ ).

Por tanto, si  $H_0$  es cierta, el estadístico F se convierte en:

$$F = F_1 = \frac{\frac{b'(X'X)b}{k+1}}{\frac{SCE}{T-k-1}} = \frac{Q_1}{\frac{SCE}{T-k-1}} = \frac{b'X'Y}{\frac{SCE}{T-k-1}} \xrightarrow{sd} F_{T-k-1}^{k+1}$$

### 5.2.3. Contraste de nulidad conjunta para los parámetros que acompañan a las variables explicativas del modelo

$$\begin{aligned} H_0 : \underline{\beta} = 0_{k \times 1} &\Rightarrow H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 : \underline{\beta} \neq 0_{k \times 1} &\Rightarrow H_1 : \text{alguno o todos} \neq 0 \end{aligned}$$

$$\Rightarrow H_0 : R\underline{\beta} = r \Rightarrow \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{k \times (k+1)} \begin{pmatrix} \beta_1 \\ \dots \\ \beta_i \\ \dots \\ \beta_k \\ \beta_0 \end{pmatrix}_{(k+1) \times 1} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \dots \\ 0 \end{pmatrix}_{k \times 1}$$

$H_1 : R\underline{\beta} \neq r$

Dado que en la hipótesis nula se establecen k restricciones relativas a la nulidad conjunta de los parámetros que acompañan a las variables explicativas del modelo:

- La matriz R se convierte en una matriz de k filas y (k+1) columnas, donde todos los elementos son cero excepto los de la “diagonal principal” de las k primeras columnas que son iguales a uno. Por tanto, en este caso, el producto  $R\underline{\beta}$  selecciona el subvector  $\underline{\beta}$  del vector  $\underline{\beta}$ .
- r es un vector columna que contiene k ceros, dado que es el valor que se fija para las k restricciones que se establecen en la hipótesis nula.
- El producto  $R(X'X)^{-1}R'$  selecciona la submatriz formada por las k primeras filas y columnas de la inversa de la matriz X'X. Dicha submatriz coincide con la inversa de  $M_{xx}$ , que no es más que la matriz de productos cruzados de las variables explicativas centradas respecto a su media muestral.

Por tanto, si  $H_0$  es cierta, el estadístico F se convierte en:

$$F = F_2 = \frac{\frac{\underline{b}' M_{xx} \underline{b}}{k}}{\frac{SCE}{T-k-1}} = \frac{Q_2}{\frac{SCE}{T-k-1}} = \frac{\underline{b}' M_{xy}}{\frac{SCE}{T-k-1}} = \frac{SCR}{\frac{SCE}{T-k-1}} = \frac{R^2}{\frac{1-R^2}{T-k-1}} \xrightarrow{sd} F_{T-k-1}^k$$

### 5.2.4. Contraste de nulidad conjunta para un subconjunto paramétrico

$$\begin{aligned} H_0 : \beta_{sub_{(k-h),x1}} &= 0_{(k-h),x1} \\ H_1 : \beta_{sub_{(k-h),x1}} &\neq 0_{(k-h),x1} \end{aligned} \Rightarrow \begin{aligned} H_0 : \beta_{h+1} = \beta_{h+2} = \dots = \beta_k &= 0 \\ H_1 : \text{alguno o todos} \neq 0 \end{aligned}$$

$$\Rightarrow H_0 : R\beta = r \Rightarrow \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{(k-h)x(k+1)} \begin{pmatrix} \beta_1 \\ \dots \\ \beta_h \\ \beta_{h+1} \\ \beta_{h+2} \\ \dots \\ \beta_k \\ \beta_0 \end{pmatrix}_{(k+1)x1} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \dots \\ 0 \end{pmatrix}_{(k-h)x1}$$

$H_1 : R\beta \neq r$

Dado que en la hipótesis nula se establece (k-h) restricciones relativas a la nulidad conjunta de los parámetros que acompañan a (k-h) variables explicativas del modelo:

- La matriz R se convierte en una matriz de (k-h) filas y (k+1) columnas, donde todos los elementos son cero excepto los de la “diagonal principal” de las (k-h) penúltimas columnas que son iguales a uno. Por tanto, en este caso, el producto  $R\beta$  selecciona el subvector  $\beta_{sub}$  del vector  $\beta$ .
- r es un vector columna que contiene (k-h) ceros, dado que es el valor que se fija para las (k-h) restricciones que se establecen en la hipótesis nula.
- El producto  $R(X'X)^{-1}R'$  selecciona la submatriz  $(X'X)_{sub}^{-1}$  formada por las (k-h) penúltimas filas y columnas de la inversa de la matriz X'X.

Por tanto, si  $H_0$  es cierta, el estadístico F se convierte en:

$$F = F_{sub} = \frac{\frac{b'_{sub}(X'X)_{sub}^{-1}b_{sub}}{k-h}}{\frac{SCE}{T-k-1}} = \frac{\frac{Q_{sub}}{k-h}}{\frac{SCE}{T-k-1}} = \frac{\frac{\Delta SCR}{k-h}}{\frac{SCE}{T-k-1}} = \frac{\frac{SCE_{sub} - SCE}{k-h}}{\frac{SCE}{T-k-1}} \xrightarrow{sd} F_{T-k-1}^{k-h}$$

### 5.2.5. Contraste de nulidad para una combinación lineal

$$\begin{aligned} H_0 : \pi &= 0 \\ H_1 : \pi &\neq 0 \end{aligned} \Rightarrow H_0 : R\beta = r \Rightarrow (R_1 \dots R_i \dots R_k \ R_0)_{1x(k+1)} \begin{pmatrix} \beta_1 \\ \dots \\ \beta_i \\ \dots \\ \beta_k \\ \beta_0 \end{pmatrix}_{(k+1)x1} = 0_{1x1}$$

$H_1 : R\beta \neq r$

Dado que en la hipótesis nula se establece una única restricción relativa a la nulidad de una combinación lineal de parámetros ( $\pi$ ):

- La matriz R se convierte en un vector fila de (k+1) columnas que posee tantos elementos no nulos como parámetros intervienen en la combinación lineal, es decir, si el parámetro  $\beta_i$  interviene en la combinación lineal, el elemento  $R_i$  tomará un valor igual a la ponderación con que dicho parámetro interviene en la combinación lineal, en caso contrario, tomará el valor cero.
- 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.

En este caso, el estadístico F se puede calcular como la la ratio entre el cuadrado del estimador de la combinación lineal ( $p$ ) cuya nulidad se quiere contrastar y su varianza estimada.

Por tanto, si  $H_0$  es cierta, el estadístico F se distribuye como una F de Snedecor de 1 grado de libertad en el numerador y (T-k-1) grados de libertad en el denominador:

$$F = \frac{P_i^2}{S_p^2} \xrightarrow{sd} F_{T-k-1}^1$$

Dada la equivalencia existente entre una distribución F de Snedecor y una distribución t de Student, también se podría utilizar esta distribución para efectuar este contraste:

$$\sqrt{F} = t = \frac{P}{S_p} \xrightarrow{sd} t_{T-k-1}$$

### 5.3. ¿Cómo realizar contrastes de hipótesis en Shazam?

Para realizar contrastes de hipótesis tenemos dos opciones:

- Utilizar el **Wizard** correspondiente.
- Utilizar directamente el comando **TEST**.

#### 5.3.1. A través del WIZARD

Al seleccionar el botón Wizards del Menú Principal → se abre un cuadro de diálogo que informa de las distintas tareas que se pueden implementar a través de este asistente → se hace clic en el botón Next → se abre un nuevo cuadro de diálogo en el que tenemos que seleccionar el procedimiento que se desea implementar y que, en este caso, es “**Hypothesis Tests**”, para seleccionarlo, basta con situarse encima y hacer un clic con el botón izquierdo del ratón. Al seleccionarlo queda sombreado en un color más oscuro → se hace clic en “**Go**” → se abre el cuadro de diálogo “**Hypothesis Tests**” donde Shazam informa de que se está en el Wizard de contrastes de hipótesis y de las tareas que va a ejecutar que, en este caso, es única y viene seleccionada por defecto. Se pincha en el botón "Next" y se abre un nuevo cuadro de diálogo donde se tiene que especificar la ecuación o ecuaciones de la hipótesis que se quiere contrastar.

Tal y como puede verse en la Ilustración 5-1, en este cuadro de diálogo, Shazam da una pequeña ayuda de cómo deben especificarse las ecuaciones a través de las cuales se establecen las restricciones que se fijan en la hipótesis nula o hipótesis a contrastar. Para especificar las restricciones que intervienen en el contraste, hacer clic en el botón “New” que se distingue por un icono donde aparece una bombilla encendida y teclear la restricción que se desee contrastar, teniendo en cuenta que Shazam identifica a los parámetros con el nombre de las variables a las que acompañan en el modelo econométrico. Si en la hipótesis nula interviene más de una restricción, es decir, sea necesario explicitar más de una ecuación,

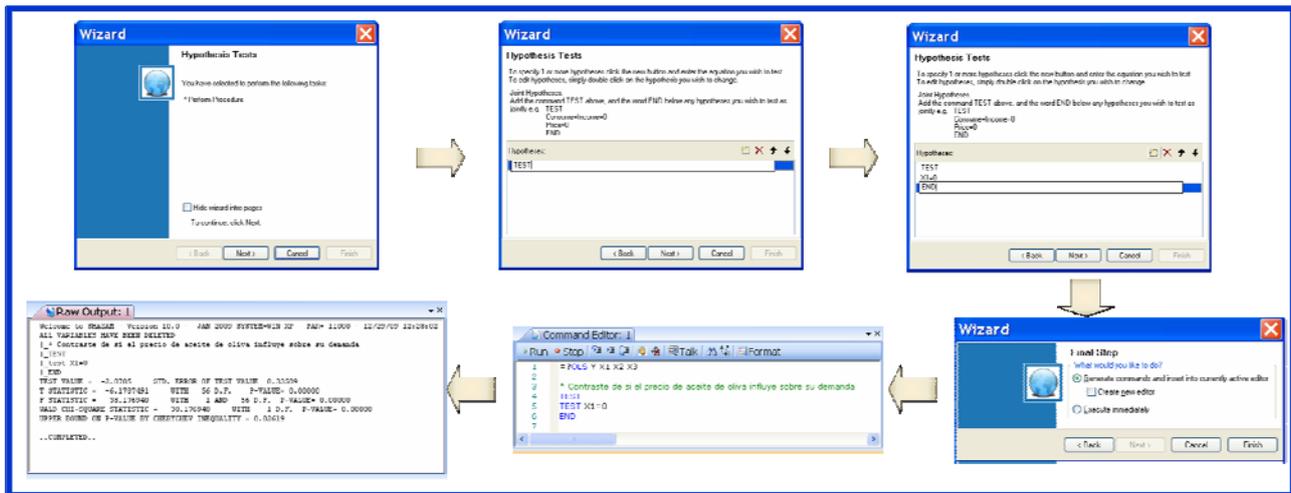


Ilustración 5-1. Secuencia de cuadros de diálogo del wizard “Hypothesis Tests”.

es imprescindible que en la primera línea del contraste aparezca la palabra “TEST” y en la última, aparezca la palabra “END”. En el caso de que la hipótesis nula contenga una única restricción se pueden añadir también estas líneas aunque no sería necesario, es decir, bastaría simplemente con escribir la línea de la ecuación de la restricción.

Este wizard permite de una forma sencilla corregir posibles errores en la especificación de la hipótesis a contrastar. El usuario puede editar cualquier ecuación de cualquier línea del contraste, simplemente, haciendo un doble clic sobre la restricción que desee cambiar o corregir, puede borrar líneas utilizando el botón “Delete” que se distingue por un icono donde aparece un aspa de color rojo y puede mover hacia arriba o hacia abajo las líneas utilizando los botones “Move Up” y “Move Down” que se distinguen por iconos donde aparece una flecha hacia arriba en el primer caso y hacia abajo en el segundo.

Una vez que se hayan escrito todas las líneas del contraste, se selecciona el botón “Next” y con ello llegamos al cuadro de diálogo final “Final Step”. Al hacer clic en el botón “Finish” saldremos del Wizard y, en ese momento, aparecerá escrita la instrucción o instrucciones correspondientes en la Ventana Editor Comandos.

Un ejemplo del procedimiento de construcción de un comando TEST a través de su asistente, aparece recogido en la Ilustración 5-1. Una vez escrito el comando en el Editor de Comandos, si se quiere ver la salida de este comando, tan sólo, se tendrá que ejecutar y, para ello, una alternativa es pinchar en el botón “Run” de dicha ventana. A continuación se va a la Ventana Output correspondiente y se podrá ver y analizar la salida.

En el Cuadro 5-1 aparece recogida la información que proporciona la ejecución de un comando **TEST** en el que sólo interviene una restricción. Se debe tener en cuenta que la información relativa al estadístico **t** no aparece cuando en el contraste interviene más de una restricción. Debe recordarse que la equivalencia entre el estadístico **t**, el estadístico **F** y el estadístico de Wald (**W**) es la siguiente:

$$W = q F = q t^2$$

Además, debe de tenerse en cuenta que aunque las versiones **t** y **F** son válidas para muestras pequeñas, la versión **W**, tan sólo, es válida si la muestra es lo suficientemente grande.

Para que Shazam ejecute un comando **TEST** será necesario que dicho comando vaya precedido de manera inmediata por un comando **OLS**<sup>44</sup>. Si en el output no se está interesado en ver la salida del comando **OLS**, dicho comando se precederá, tal y como puede observarse en la Cuadro 5-1, por los símbolos “=?”, donde el signo de cierre de interrogación permite que se ejecute el comando pero que se omita su salida y el signo de igualdad omite la escritura de dicha línea de comandos en el output.

TEST VALUE = Numerador del estadístico t (diferencia entre la restricción estimada y la restricción bajo la hipótesis nula) (**\$VAL**)  
 STD. ERROR OF TEST VALUE = Denominador del estadístico t (estimador de la desviación estándar de la discrepancia anterior) (**\$STEST**)  
 T STATISTIC = Estadístico t (t) (**\$T**)                      WITH DF = con T-K-1 grados de libertad (**\$DF**)                      P-VALUE = P-valor

F STATISTIC = Estadístico F (F) (**\$F**)  
 WITH DF1 = y DF2 = con q (nº de restricciones que intervienen en el contraste) y T-K-1 grados de libertad en el numerador y denominador respectivamente (**\$DF1,\$DF2**)  
 P-VALUE = P-valor

WALD CHI-SQUARE STATISTIC = Estadístico de Wald (W) (**\$CHI**)                      WITH DF = con q (nº de restricciones que intervienen en el contraste) grados de libertad  
 P-VALUE = P-valor

UPPER BOUND ON P-VALUE BY CHEBYCHEV INEQUALITY = Cota superior de la desigualdad de Chebychev

Cuadro 5-1. Salida de un comando TEST.

### 5.3.2. A través del comando TEST

La otra opción para realizar contrastes de hipótesis es escribir directamente el comando **TEST** en el **Editor de Comandos**, para lo cual el usuario debe conocer no sólo su formato de escritura sino también la denominación de las opciones disponibles.

El formato del comando **TEST** es:

```
TEST
TEST ecuación 1
TEST ecuación 2
...
TEST ecuación q
END
```

<sup>44</sup> Para que Shazam ejecute un comando **TEST** debe precederse por un comando **OLS**, en caso contrario, Shazam emite un mensaje de error donde señala que dicho comando será ignorado (**...ERROR..ABOVE COMMAND WILL BE IGNORED**). Para evitar cometer este tipo de error, Shazam permite acceder al asistente de contrastes desde el asistente de estimación MCO y ligar de forma automática ambos comandos.

donde “q” es el número de restricciones que intervienen en el contraste. Debe tenerse en cuenta que para que Shazam identifique los parámetros que intervienen en cada una de las restricciones (ecuaciones) que intervienen en el contraste, se le debe indicar el nombre de la variable a la que acompañan tal como se hizo en la estimación del modelo.

El formato de escritura de este comando requiere que en el **Editor de Comandos** cada restricción ocupe una línea y que cada ecuación sea precedida por un comando **TEST**. Además, requiere una primera línea donde aparezca únicamente un comando **TEST** y una última línea donde aparezca un comando **END**, en caso contrario, asumirá “q” contrastes con una restricción cada uno y no un único contraste con “q” restricciones. Por tanto, cuando en el contraste intervenga una única restricción, el usuario podrá omitir el comando **TEST** inicial y el comando **END** final:

**TEST ecuación 1**

## 5.4. ¿Cómo buscar el P-valor y el valor crítico en una distribución prefijada?

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<sup>45</sup>. Ambos valores se buscarán en la distribución de probabilidad que sigue el estadístico de prueba de dicho contraste bajo la hipótesis nula.

Para buscar dichos valores tenemos dos opciones:

- Utilizar el **Wizard** correspondiente.
- Utilizar directamente el comando **DISTRIB**.

### 5.4.1. A través del WIZARD

Al seleccionar el botón Wizards del Menú Principal → se abre un cuadro de diálogo que informa de las tareas que se pueden implementar a través de este asistente → se hace clic en el botón Next → se abre un nuevo cuadro de diálogo en el que se tiene que seleccionar el procedimiento que se desea implementar y que, en este caso, es “**Probability Distributions**”, para ello basta con situarse encima y hacer un clic con el botón izquierdo del ratón. Al seleccionarlo queda sombreado en un color más oscuro → se hace clic en “**Go**” → se abre el cuadro de diálogo “**Probability Distributions**” donde Shazam informa que está en el Wizard de distribuciones de probabilidad y de las tareas que va a ejecutar que, en este caso, es única y viene seleccionada por defecto. Se pincha en el botón "Next" y se abre el cuadro de diálogo “Select Variables” donde se seleccionan las “variables” con las que debe realizar dicha tarea, así como del rango muestral para realizarla.

En este cuadro de diálogo, en el recuadro “variables disponibles” situado a la izquierda (Ilustración 5-2) Shazam informa al usuario de las variables disponibles (que pueden proceder de ficheros de datos y/o de las que Shazam tenga disponibles en memoria “system”). El usuario debe tener la precaución de que no exista más de una variable con la misma denominación, pues si este fuera el caso, la información disponible en dicha variable sería, tan sólo, la cargada en último lugar.

Para seleccionar la/s variable/s, el usuario debe elegir una/s variable/s dentro de las disponibles, situándose con el cursor encima y haciendo clic en el botón “**Add**” y, dicha variable será seleccionada y aparecerá en el recuadro “selected variables”. En caso de que el usuario se equivoque al hacer la selección, tan sólo, debe seleccionarla en el cuadro “selected variables” y hacer clic en el botón “**Remove**” y la variable dejará de estar seleccionada.

Además de seleccionar las variables, se puede explicitar el rango muestral que se desea para realizar dicha tarea, en caso contrario, Shazam utilizará toda la muestra (véase que es la opción seleccionada por defecto).

---

<sup>45</sup> La mayoría de los estadísticos asociados a contrastes que Shazam ofrece en sus salidas llevan asociados un P-valor.

Una vez realizada la selección de variables y del rango muestral, se hace clic en **Next** y se pasa al siguiente cuadro de diálogo, donde se debe seleccionar el tipo de distribución y especificar algunos de sus atributos<sup>46</sup>. Si se está interesado en buscar valores críticos en la distribución será necesario que además de especificar los atributos, se seleccione **“Compute Inverse Survival Function”**. Una vez hecha dicha selección y después de presionar el botón **“OK”**, Shazam permite guardar dichos valores bajo la denominación que el usuario desee.

En este cuadro de diálogo, también se permite:

- Guardar la **“Función de Distribución Acumulada”** o **“Cumulative Distribution Function (CDF)”**, que se puede interpretar como el P-valor asociado al estadístico de prueba en contrastes de la cola izquierda, razón por la cual Shazam no la calculará en caso de haber seleccionado **“Compute Inverse Survival Function”**.
- Calcular el **“log likelihood function”** o **“logaritmo de la función de verosimilitud”**.

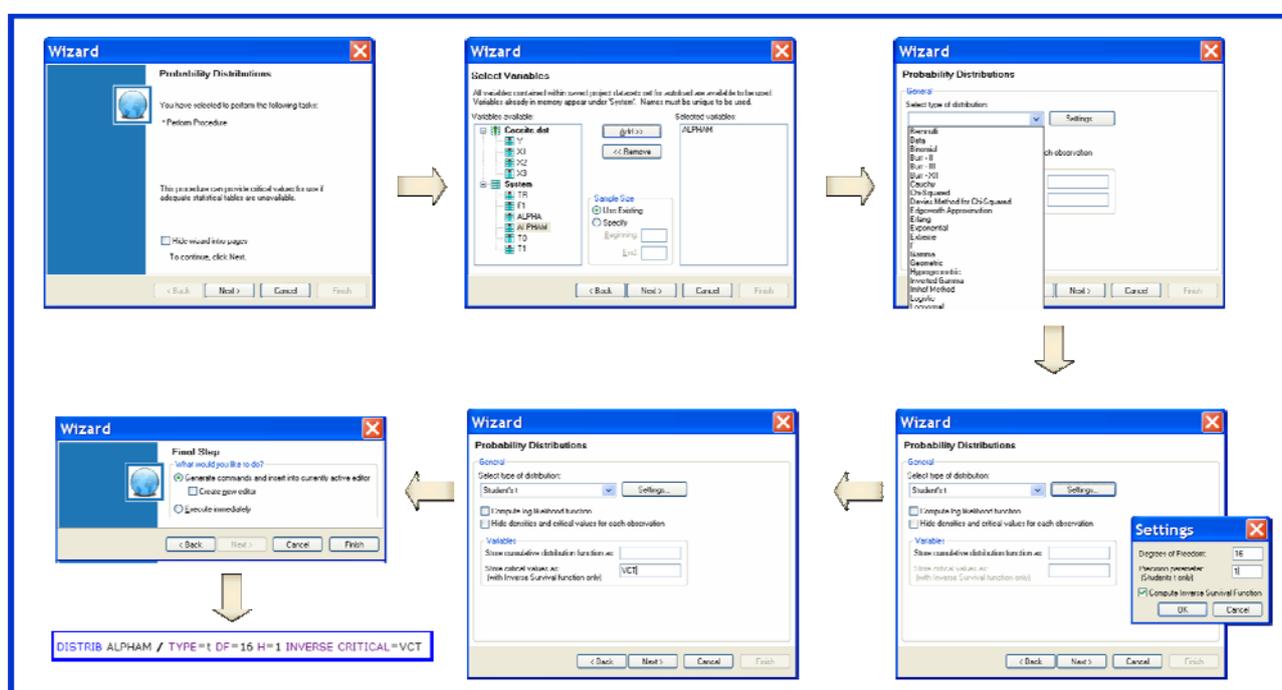


Ilustración 5-2. Secuencia de cuadros de diálogo del wizard **“Probability Distributions”**.

- Ocultar las probabilidades y los valores críticos asociados a cada observación de la muestra (**Hide densities and critical values for each observations**).

<sup>46</sup> Los atributos que se deben especificar van a depender de la distribución seleccionada. En el caso de elegir una distribución t de Student será necesario especificar sus grados de libertad y el parámetro de precisión que por defecto será “1” (Ilustración 5-2). En caso de elegir una distribución normal será necesario especificar su media y su varianza. Si se elige una distribución F de Snedecor será necesario especificar sus grados de libertad en el numerador y en el denominador. En caso de elegir una distribución chi-cuadrado será necesario especificar sus grados de libertad . . . .

Una vez hechas todas las elecciones en este cuadro de diálogo, se selecciona el botón “**Next**” y se llega al último cuadro de diálogo “**Final Step**”. Al hacer clic en el botón “**Finish**” se sale del asistente y, en la **Ventana Editor Comandos** aparece escrita la instrucción correspondiente. En el ejemplo recogido en la Ilustración 5-2, el comando que se inserta es:

**DISTRIB** ALPHAM / **TYPE=t** DF=16 **H=1** **INVERSE** **CRITICAL=VCT**

Este comando le indica a Shazam que para un valor ALPHAM, busque y guarde bajo la denominación VCT, el valor crítico de una distribución t de Student con 16 grados de libertad que deje a su derecha una cola de probabilidad igual a ALPHAM. Dado que en el ejemplo que nos ocupa, ALPHAM es un escalar y se trabaja con toda la muestra, si se ejecutase la instrucción **PRINT VCT**, se podría comprobar que VCT es un vector columna cuyo número de filas coincide con el rango muestral, donde el primer elemento sería el valor crítico y los restantes, serían iguales a -99999, que es como Shazam identifica a los “valores perdidos” o “missing values”. Este problema se podría evitar eligiendo un rango muestral unitario **SAMPLE 1 1**.

Como ya se ha comentado, en la Ilustración 5-2 se recoge un ejemplo de búsqueda del valor crítico. Si el usuario desea buscar un P-valor, la única diferencia es que en el cuadro de diálogo “**Settings**” no se selecciona “**Compute Inverse Survival Function**”.

**DISTRIB** T / **TYPE=t** DF=16 **H=1** **CDF=CDFT**

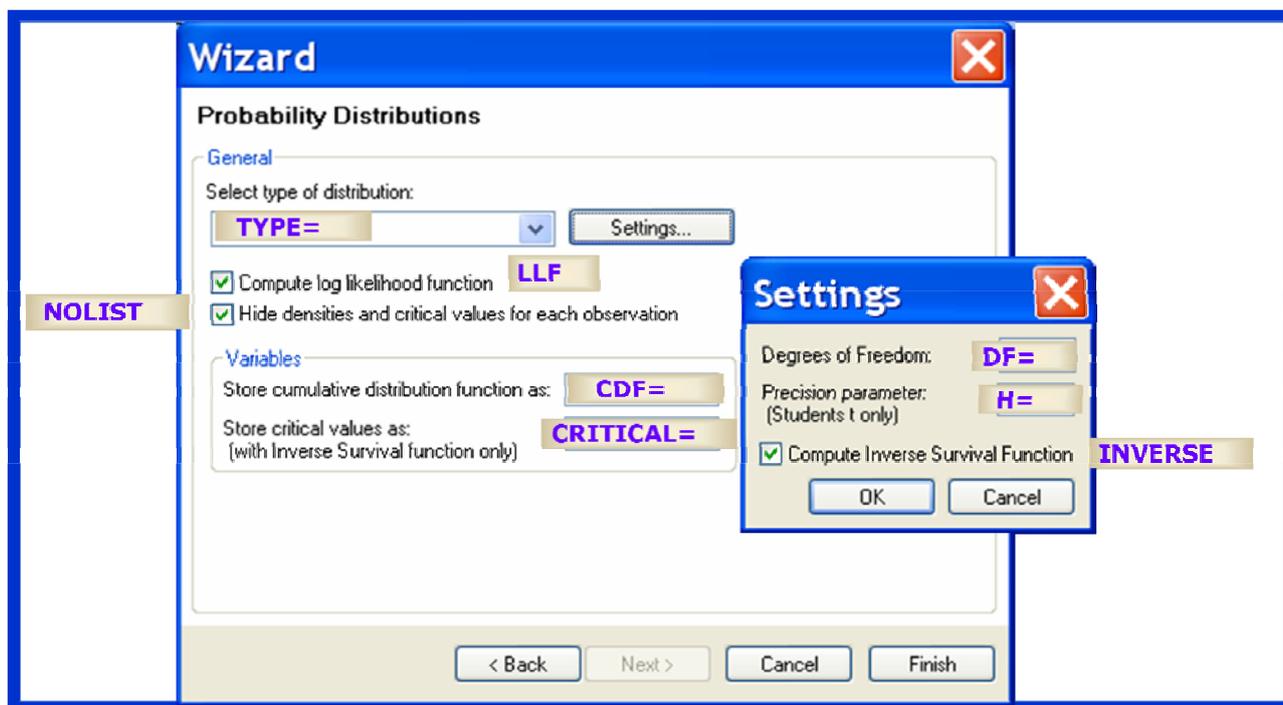


Ilustración 5-3. Opciones del comando DISTRIB que se pueden insertar a través de su Wizard.

Este comando le indica a Shazam que para un valor T, busque y guarde bajo la denominación CDFT, el valor de la función de distribución acumulada de una t de Student con 16 grados de libertad que coincidirá con el P-valor si se utiliza la cola izquierda para el contraste. En contrastes de la cola derecha, el P-valor será 1-CDF.

### 5.4.2. A través del comando DISTRIB

La otra opción para obtener los valores críticos y los P-valores es escribir directamente el comando **DISTRIB** en el **Editor de Comandos**, para lo cual el usuario debe conocer su formato de escritura y la denominación de las opciones disponibles.

El formato del comando DISTRIB es:

#### DISTRIB VARIABLE / OPCIONES

En la Ilustración 5-3 aparecen recogidas en color violeta la denominación de algunas<sup>47</sup> de las opciones que se pueden insertar en un comando **DISTRIB** a través de su wizard.

**INVERSE** → Proporciona el valor crítico de la distribución en vez de la probabilidad (P-valor).

**CRITICAL=** → Guarda los valores críticos cuando utilizamos la opción INVERSE.

**TYPE=** → Especifica el tipo de distribución.

Algunos de los tipos de distribución disponibles son:

- **CHI** → Distribución chi cuadrado (se debe utilizar con **DF=**).
- **F** → Distribución F de Snedecor (se debe utilizar con **DF1=** y **DF2=**).
- **NORMAL** → Distribución Normal (se debe utilizar con **MEAN=** y **VAR=**).
- **T** → Distribución t de Student (se debe utilizar con **DF=**).

**DF=** → Especifica los grados de libertad. Se usa solamente con **TYPE=T** y **TYPE=CHI**.

**DF1= DF2=** → Especifica los grados de libertad para el numerador y el denominador. Se usa sólo con **TYPE=F**.

**CDF=** → Guarda los valores de la función de distribución acumulativa en la variable especificada.

Cuadro 5-2. Descripción opciones comando DISTRIB.

En el Cuadro 5-2 aparece recogida una breve descripción de algunas de las opciones disponibles con el comando **DISTRIB**.

### 5.4.3. ¿Cómo utilizar el comando DISTRIB en el contraste de hipótesis?

Para efectuar contrastes de hipótesis, se debe conocer el P-valor o el valor crítico y ambos pueden ser determinados en Shazam a través del comando DISTRIB.

Para buscar el valor crítico será necesario conocer la distribución del estadístico de prueba y fijar el nivel de significación ( $\alpha$ ) o el nivel de confianza ( $1-\alpha$ ). El usuario debe tener en cuenta que el **valor crítico** que proporciona Shazam se corresponde con el valor de la distribución que deja a su derecha una cola de probabilidad igual al **valor** fijado por el usuario en la “**variable**” seleccionada en el comando DISTRIB. Por tanto, el contenido de dicha “**variable**” va a depender del tipo de contraste, es decir, si la probabilidad de rechazo de la hipótesis nula se sitúa en ambas colas, en la cola derecha o en la cola izquierda:

- En contrastes de la cola derecha, el valor de la variable seleccionada en el comando DISTRIB será igual al nivel de significación ( $\alpha$ ).

<sup>47</sup> Las opciones que se pueden insertar a través del cuadro de diálogo de “**Settings**” van a depender del tipo de distribución elegida.

- En contrastes de la cola izquierda, el valor de la variable seleccionada en el comando DISTRIB será igual al nivel de confianza (1- $\alpha$ ).
- En contrastes de dos colas, será necesario determinar dos valores: para determinar el de la cola derecha se utilizara la mitad del nivel de significación ( $\alpha/2$ ) y para determinar el de la cola izquierda ( $1-\alpha/2$ ).

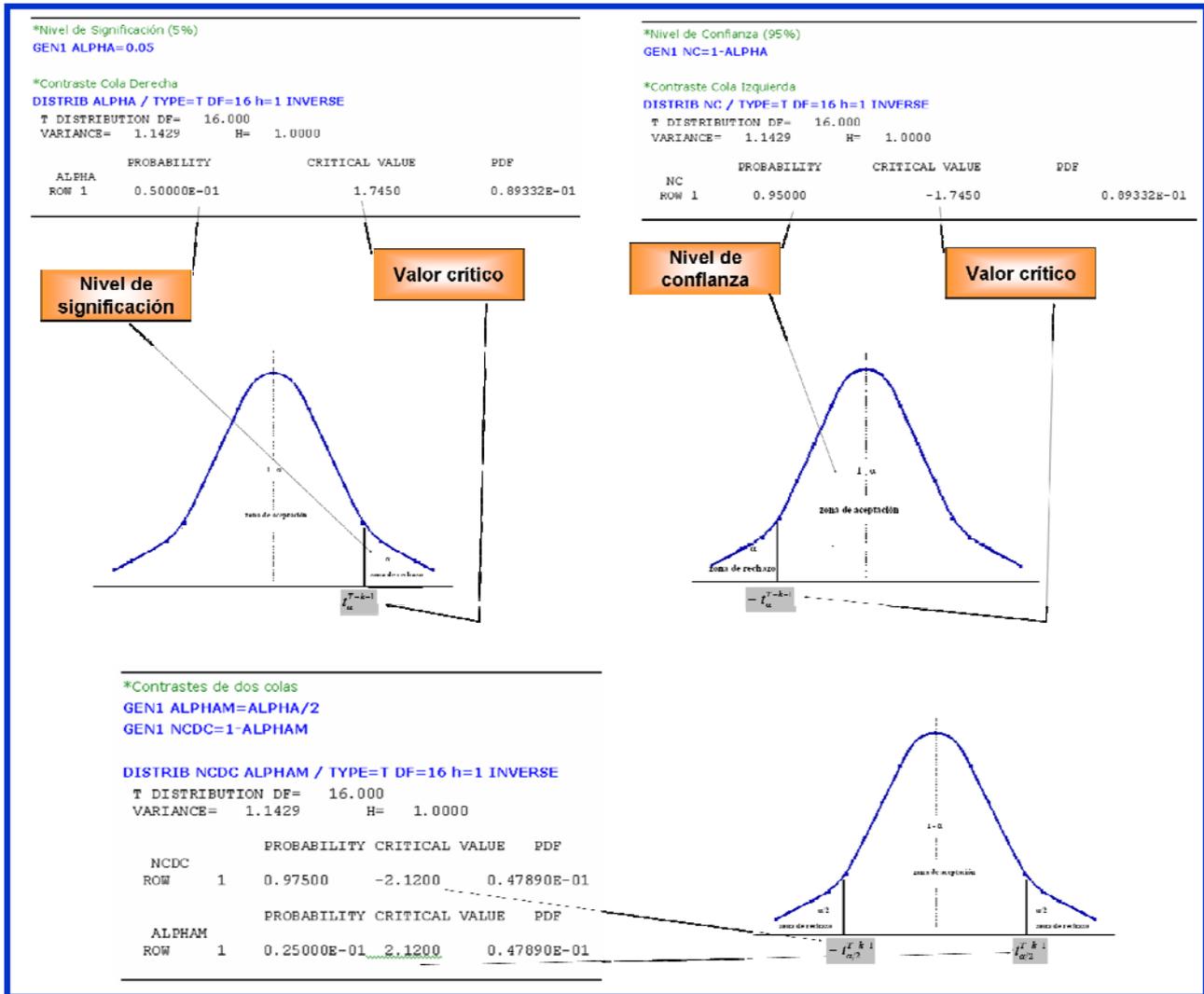


Ilustración 5-4. Ejemplo de cómo utilizar el comando DISTRIB para buscar los valores críticos en los distintos tipos de contrastes.

Un ejemplo de cómo se puede utilizar el comando DISTRIB en la búsqueda de valores críticos para efectuar contrastes de hipótesis aparece recogido en la Ilustración 5-4. Dado que en este ejemplo se ha considerado que el estadístico de prueba sigue una distribución t de Student y dicha distribución es simétrica respecto al cero, no sería necesario buscar el valor crítico en la cola izquierda puesto que para una misma cola de probabilidad coincidiría con el obtenido en la cola derecha cambiado de signo.

Para buscar el P-valor será necesario conocer el valor del estadístico de prueba y su distribución. En la Ilustración 5-5 aparece recogido un ejemplo a partir de una distribución t de Student de 16 grados de libertad. Para facilitar al lector no sólo el proceso de búsqueda sino también la interpretación del P-valor, se ha optado por tomar como valores de los estadísticos los valores críticos obtenidos en el ejemplo de la Ilustración 5-4. El usuario debe tener claro que en contrastes de la cola derecha, Shazam proporciona el P-valor bajo la denominación “1-CDF” y en contrastes de la cola izquierda bajo la denominación “CDF”.

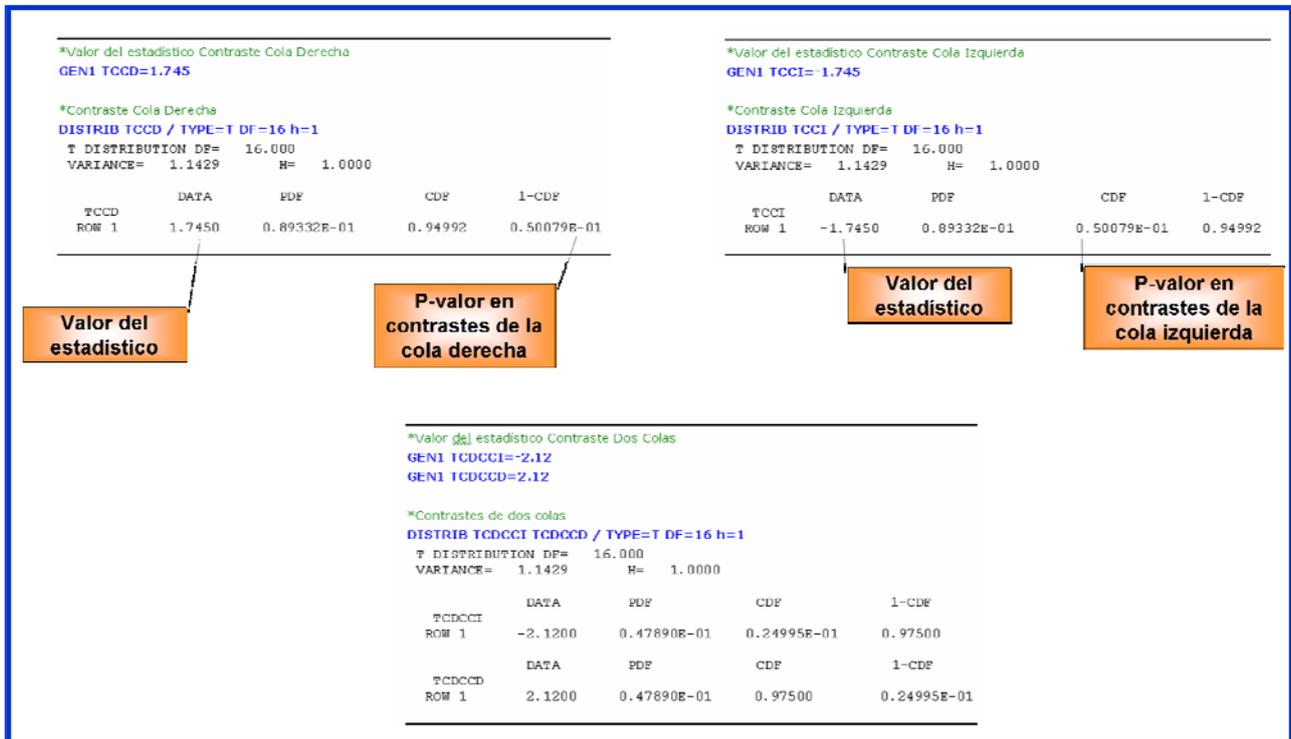


Ilustración 5-5. Ejemplo de cómo utilizar el comando DISTRIB para buscar P-valores en los distintos tipos de contrastes.

### 5.5. Regiones de Confianza

Para un subconjunto de (k-h) parámetros ( $\beta_{sub}$ ), la región de confianza viene dada por todos aquellos puntos que verifiquen:

$$P\left\langle (b_{sub} - \beta_{sub})' [(X'X)_{sub}^{-1}]^{-1} (b_{sub} - \beta_{sub}) \leq F_{\alpha} [k - h] S^2 \right\rangle = 1 - \alpha$$

En el caso 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 permiten 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, en caso contrario, se rechazará.

#### 5.5.1. ¿Cómo calcular intervalos de confianza en Shazam?

Para calcular intervalos de confianza tenemos dos opciones:

- Utilizar el **Wizard** correspondiente.
- Utilizar directamente el comando **CONFID**.

##### 5.5.1.1. A través del WIZARD

Al seleccionar el botón Wizards del Menú Principal → se abre un cuadro de diálogo que informa de las tareas que se pueden implementar a través de este asistente → se hace clic en el botón Next → se abre un nuevo cuadro de diálogo en el que se tiene que seleccionar el procedimiento que se desea implementar y que, en este caso, es “**Confidence Intervals**”. Para seleccionarlo, basta con situarse

encima y hacer un clic con el botón izquierdo del ratón. Al seleccionarlo queda sombreado en un color más oscuro → se hace clic en “Go” → se abre el cuadro de diálogo “**Confidence Intervals**” donde Shazam informa que está en el Wizard de intervalos de confianza y de las tareas que va a ejecutar que, en este caso, es única y viene seleccionada por defecto. Se pincha en el botón "Next" y se abre un cuadro de diálogo donde el usuario debe seleccionar entre las variables disponibles la/s variable/s para construir el intervalo o intervalos de confianza.

Para seleccionar la/s variable/s, el usuario debe elegir una variable dentro de las disponibles que aparecen en el recuadro “available”, situándose con el cursor encima y haciendo clic en el botón “>” y dicha variable será seleccionada y aparecerá en el recuadro “selected”. En caso de que el usuario se equivoque al hacer la selección, debe seleccionarla en el cuadro “selected” y hacer clic en el botón “<” y la variable dejará de estar seleccionada.

Además, como puede observarse en la Ilustración 5-6, se debe explicitar la distribución de probabilidad

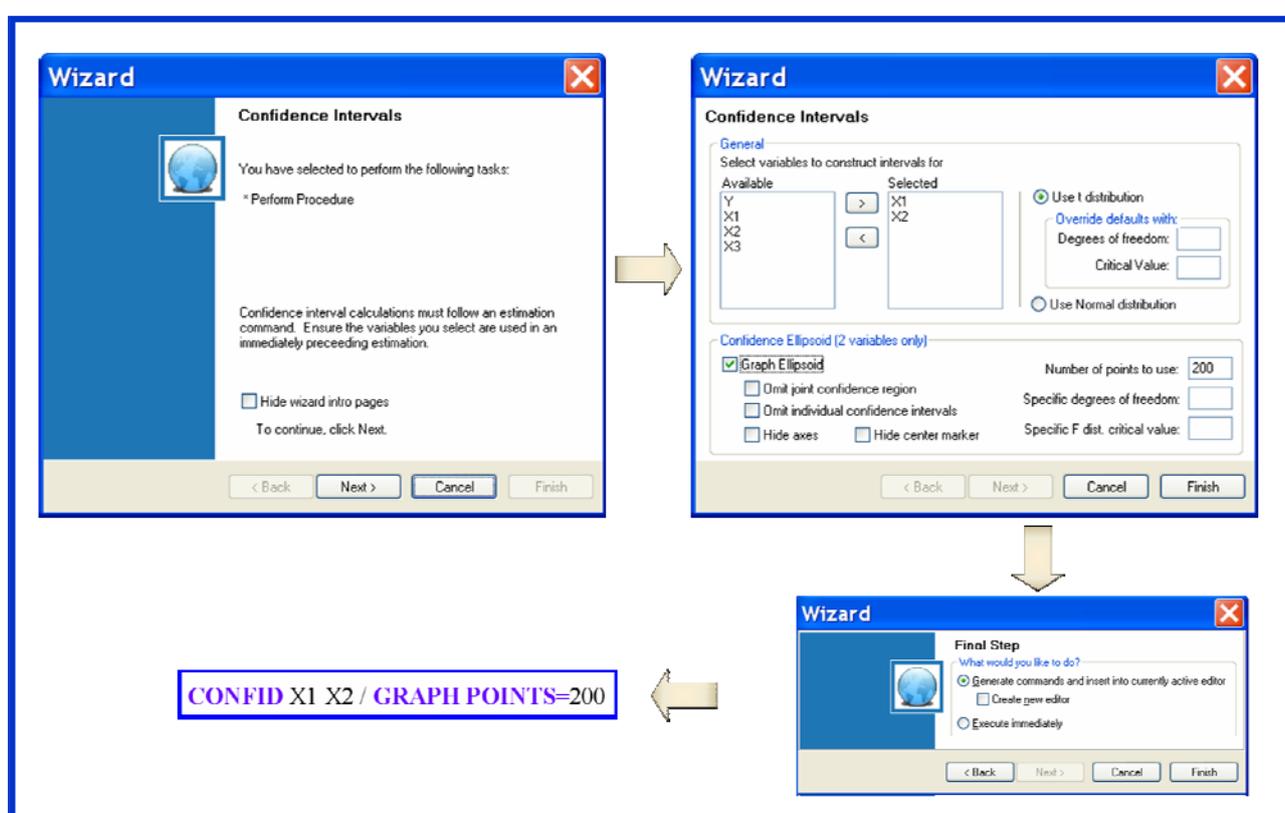


Ilustración 5-6. Secuencia de cuadros de diálogo del wizard “**Confidence Intervals**”.

a utilizar para construir dichos intervalos. El usuario puede optar entre seleccionar una t de Student o una normal (por defecto Shazam selecciona una t de Student)

Para que el comando **CONFID** se ejecute adecuadamente debe ir precedido por un comando **OLS** donde intervengan las variables utilizadas para la construcción de dichos intervalos, en caso contrario, Shazam emitirá mensajes de error del tipo: “**...ERROR CONFID COMMAND MUST FOLLOW ESTIMATION**” o “**...ERROR INVALID COEFFICIENTS OR OPTIONS**”. Este tipo de error se pueden obviar accediendo al wizard de construcción de intervalos de confianza desde el wizard de la estimación MCO, puesto que de esta manera, además de aparecer precedido por el comando OLS, no habrá posibilidad de equivocarse a la hora de elegir los parámetros, que en Shazam se identifican por el nombre de la variable a la que acompañan en la estimación MCO (Shazam en el recuadro de variables disponibles sólo ofrecerá las que intervengan en dicho comando OLS).

Por defecto, a partir de una *t* de Student cuyos grados de libertad coinciden con los de la estimación MCO que le precede, Shazam construye para cada parámetro dos intervalos de confianza (para niveles de significación del 5% y del 10%). No obstante, si lo considera conveniente, el usuario tiene la posibilidad de cambiar los grados de libertad y el valor crítico de la *t* de Student a utilizar en la construcción de dichos intervalos.

En este mismo cuadro de diálogo, en el caso de que el usuario seleccione dos variables, Shazam proporciona la posibilidad de construir un gráfico denominado “elipsoide de confianza” que no es más que la región de confianza conjunta para ambos parámetros. Por defecto y, para facilitar el análisis de dicha región de confianza, en este gráfico se proporciona el punto medio de dicho elipsoide y los puntos extremos de los intervalos individuales, los cuales pueden ayudar a ver la relación entre los contrastes de nulidad individual de los parámetros y el contraste de nulidad conjunta y, en última instancia, si existe un problema serio de multicolinealidad.

Por defecto, para construir dicha región de confianza, Shazam utiliza un nivel de significación del 5% y una *F* de Snedecor de dos grados de libertad en el numerador y, en el denominador, tantos grados de libertad como los de la estimación MCO que le precede. No obstante, existe la posibilidad de cambiar los grados de libertad del denominador y el valor crítico de la *F* de Snedecor a utilizar en la construcción de dicha región de confianza.

Se selecciona el botón "Next" y con ello se llega al cuadro de diálogo final “**Final Step**”. Al hacer clic en el botón "Finish" se sale del **Wizard** y, en la **Ventana Editor Comandos** aparece escrita la instrucción correspondiente.

#### 5.5.1.2. A través del comando **CONFID**

La otra opción para obtener los intervalos de confianza es escribir directamente el comando **CONFID** en el **Editor de Comandos**, para lo cual el usuario debe conocer su formato de escritura y la denominación de las opciones disponibles.

El formato del comando **CONFID** es:

**CONFID X1 X2 ... / OPCIONES**

En la Ilustración 5-7 aparecen recogidas en color violeta la denominación de algunas de las opciones que se pueden insertar en un comando **CONFID** a través de su wizard.

En el ejemplo recogido en la Ilustración 5-6, el comando que se inserta es:

**CONFID X1 X2 / GRAPH POINTS=200**

Si se quiere ver la salida de este comando, tan sólo, se tendrá que ejecutar y, a continuación en la **Ventana Output** se podrá ver y analizar la salida.

**Wizard**

**Confidence Intervals**

**General**

Select variables to construct intervals for

Available		Selected
Y	>	X1
X1		X2
X2	<	
X3		

Use t distribution  
 Override defaults with:  
 Degrees of freedom:   
 Critical Value:

Use Normal distribution

**Confidence Ellipsoid (2 variables only)**

Graph Ellipsoid

Omit joint confidence region  
 Omit individual confidence intervals  
 Hide axes  Hide center marker

Number of points to use:   
 Specific degrees of freedom:   
 Specific F dist. critical value:

Ilustración 5-7. Opciones del comando CONFID que se pueden insertar a través de su Wizard.

La salida del comando **CONFID** del ejemplo de la Ilustración 5-6 se recoge en la Ilustración 5-8.

### 5.5.2. ¿Cómo utilizar el intervalo de confianza en el contraste de hipótesis?

Los intervalos de confianza se pueden utilizar en el contraste de dos colas.

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.

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.

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 tan sólo significa, que con la información de la que se dispone no se tienen elementos de juicio suficientes para rechazar dicha hipótesis.

De acuerdo con los intervalos de confianza que aparecen en la Ilustración 5-8, una conclusión que se podría sacar es que para los niveles de confianza del 90% y del 95% y para los parámetros  $\beta_1$  y  $\beta_2$ , ambas variables son significativas a nivel individual, dado que el valor cero no pertenece a dichos intervalos, es decir, se rechaza la hipótesis nula  $H_0 : \beta_i = 0$  en ambos casos.

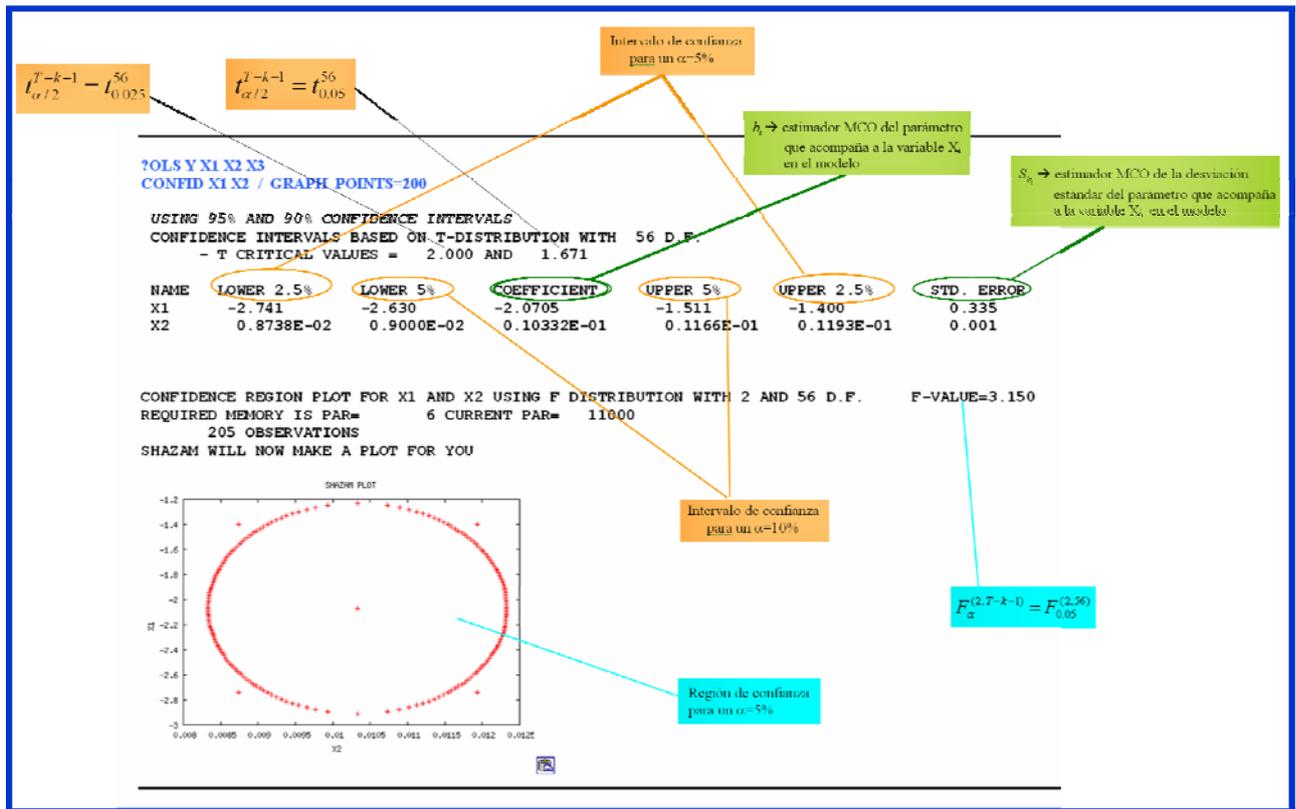


Ilustración 5-8. Ejemplo de la información que proporciona la salida de un comando CONFID en el que intervienen dos parámetros.