

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

María Pérez Marqués



Minería de datos a través de ejemplos
María Pérez Marqués

ISBN: 978-84-941801-4-9
EAN: 9788494180149
IBIC: UNF

Copyright © 2014 RC Libros
© RC Libros es un sello y marca comercial registrados

Minería de datos a través de ejemplos

Reservados todos los derechos. Ninguna parte de este libro incluida la cubierta puede ser reproducida, su contenido está protegido por la Ley vigente que establece penas de prisión y/o multas a quienes intencionadamente reprodujeren o plagiaran, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución en cualquier tipo de soporte existente o de próxima invención, sin autorización previa y por escrito de los titulares de los derechos de la propiedad intelectual. La infracción de los derechos citados puede constituir delito contra la propiedad intelectual. (Art. 270 y siguientes del Código Penal). Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra a través de la web www.conlicencia.com; o por teléfono a: 91 702 19 70 / 93 272 04 47.

RC Libros, el Autor, y cualquier persona o empresa participante en la redacción, edición o producción de este libro, en ningún caso serán responsables de los resultados del uso de su contenido, ni de cualquier violación de patentes o derechos de terceras partes. El objetivo de la obra es proporcionar al lector conocimientos precisos y acreditados sobre el tema tratado pero su venta no supone ninguna forma de asistencia legal, administrativa ni de ningún otro tipo, si se precisase ayuda adicional o experta deberán buscarse los servicios de profesionales competentes. Productos y marcas citados en su contenido estén o no registrados, pertenecen a sus respectivos propietarios.

RC Libros
Calle Mar Mediterráneo, 2. Nave 6
28830 SAN FERNANDO DE HENARES, Madrid
Teléfono: +34 91 677 57 22
Fax: +34 91 677 57 22
Correo electrónico: info@rclibros.es
Internet: www.rclibros.es

Diseño de colección, cubierta y pre-impresión: Grupo RC
Impresión y encuadernación: Service Point
Depósito Legal: M-3204-2014
Impreso en España

18 17 16 15 14 (1 2 3 4 5 6 7 8 9 10 11 12)

INTRODUCCIÓN

De un modo sencillo podemos definir la minería de datos como un conjunto de técnicas encaminadas al descubrimiento de la información contenida en grandes conjuntos de datos. Se trata de analizar comportamientos, patrones, tendencias, asociaciones y otras características del conocimiento inmerso en los datos. Actualmente se dispone de grandes cantidades de datos y es más necesario que nunca poder analizarlos ordenadamente para extraer de un modo automatizado la inteligencia contenida en ellos utilizando técnicas especializadas apoyadas en herramientas informáticas. Estas técnicas constituyen la minería de datos.

El intenso desarrollo de las herramientas de tratamiento automatizado de la información ha llevado aparejado el uso de las técnicas estadísticas de análisis multivariante de datos de una forma sencilla. Al crecer los medios informáticos se ha facilitado sobremanera la manejabilidad de los algoritmos estadísticos.

Pero por otro lado, podemos decir que las técnicas de minería de datos son tan antiguas como la estadística misma. De hecho, las técnicas estadísticas que utiliza la minería de datos coinciden en su mayoría con las técnicas estadísticas de análisis multivariante de datos. Las herramientas de minería de datos presentan en sus menús, de un modo ordenado, las técnicas de análisis multivariante de datos lógicamente secuenciadas.

Este libro analiza las herramientas más habituales en minería de datos y sus posibilidades de trabajo. Se utilizarán IBM SPSS, IBM SPSS MODELER, SAS, SAS ENTERPRISE GUIDE y SAS ENTERPRISE MINER.

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

A través de ejemplos totalmente resueltos a lo largo del libro se irán presentando las diferentes técnicas de minería de datos. Los archivos se encuentran en la página web www.rclibros.es en la sección *Zona de archivos*.

Se trata de exponer, con sencillez y mediante una metodología interactiva, los conceptos de minería de datos e inteligencia de negocios.

TÉCNICAS DE MINERÍA DE DATOS Y HERRAMIENTAS

CLASIFICACIÓN DE LAS TÉCNICAS DE MINERÍA DE DATOS Y HERRAMIENTAS MÁS COMUNES

Inicialmente las técnicas de minería de datos pueden clasificarse en técnicas de modelado originado por la teoría (en las que las variables pueden clasificarse en dependientes e independientes), técnicas de modelado originado por los datos (en las que todas las variables tienen inicialmente el mismo) y técnicas auxiliares.

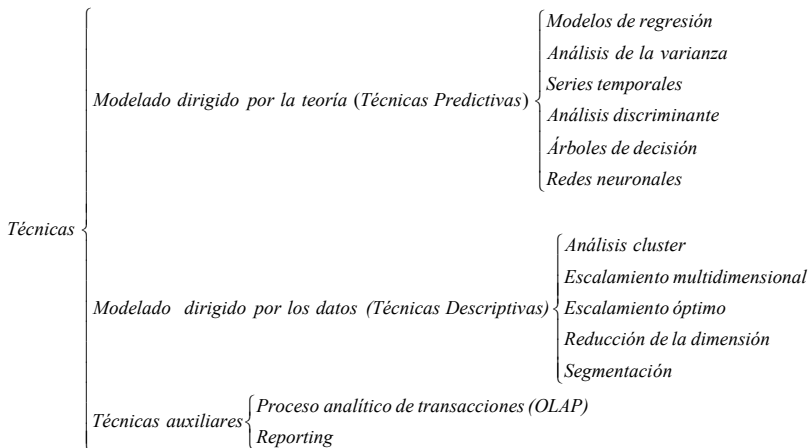
Las *técnicas de modelado originado por la teoría* especifican el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después del proceso de minería de datos antes de aceptarlo como válido. Formalmente, la aplicación de todo modelo debe superar las fases de *identificación objetiva* (a partir de los datos se aplican reglas que permitan identificar el mejor modelo posible que ajuste los datos), *estimación* (proceso de cálculo de los parámetros del modelo elegido para los datos en la fase de identificación), *diagnosis* (proceso de contraste de la validez del modelo estimado) y *predicción* (proceso de utilización del modelo identificado, estimado y validado para predecir valores futuros de las variables dependientes). Podemos incluir entre estas técnicas todos los tipos de regresión y asociación, análisis de la varianza y covarianza, análisis discriminante y series temporales.

En las *técnicas de modelado originado por los datos* no se asigna ningún papel predeterminado a las variables. No se supone la existencia de variables dependientes ni independientes y tampoco se supone la existencia de un modelo previo para los datos. Los modelos se crean automáticamente partiendo del reconocimiento de patrones. El modelo se obtiene como mezcla del conocimiento obtenido antes y después de la minería de datos y también debe contrastarse antes de aceptarse como válido. Por ejemplo, las *redes neuronales* permiten descubrir modelos complejos y afinarlos a medida que progresa la exploración de los datos. Gracias a su capacidad de aprendizaje, permiten descubrir relaciones complejas entre variables sin ninguna intervención externa.

Por su parte, las *técnicas de clasificación* extraen perfiles de comportamiento o clases, siendo su objetivo construir un modelo que permita clasificar cualquier nuevo dato. Asimismo, los *árboles de decisión* permiten dividir datos en grupos basados en los valores de las variables. Esta técnica permite determinar las variables significativas para un elemento dado. El mecanismo de base consiste en elegir un atributo como raíz y desarrollar el árbol según las variables más significativas. Además de las redes neuronales, los árboles de decisión y las técnicas de clasificación (cluster, etc.), podemos incluir en este grupo las *técnicas de reducción de la dimensión* (factorial, componentes principales, correspondencias, etc.), las técnicas de escalamiento óptimo y multidimensional, y el análisis conjunto.

Las *técnicas auxiliares* son herramientas más superficiales y limitadas. Son nuevos métodos basados en técnicas estadísticas descriptivas e informes.

A continuación se muestra una *clasificación inicial de las técnicas de minería de datos*.



Entre las herramientas de minería de datos más habituales tenemos actualmente el software de IBM y el software de SAS. IBM dispone de las herramientas IBM SPSS Statistics e IBM SPSS Modeler. La primera de ellas es un software para el trabajo estadístico en general (que contiene varios procedimientos de minería de datos) y la segunda es una herramienta específica de minería de datos sucesora de SPSS Clementine. Por su parte SAS dispone del software estadístico general, de SAS Enterprise Guide para el trabajo con procedimientos estadísticos y de minería por menús y del software SAS Enterprise Miner, específico de minería de datos.

MODELADO ORIGINADO POR LA TEORÍA (TÉCNICAS EXPLICATIVAS O DE LA DEPENDENCIA)

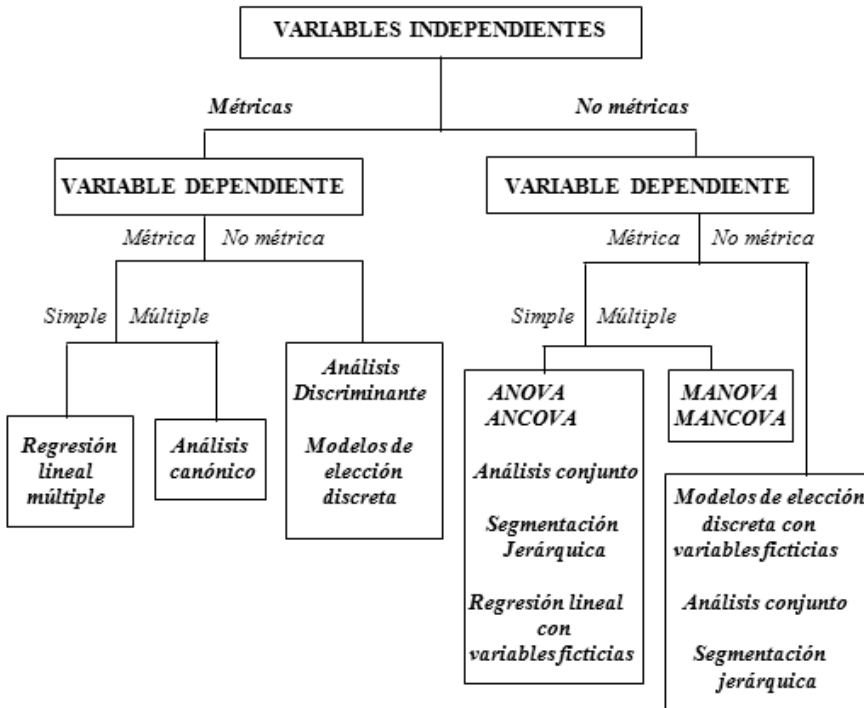
En los análisis que no es aceptable una importancia equivalente en las variables que intervienen, porque *alguna variable se destaca como dependiente principal (MÉTODOS DE DEPENDENCIA)*, habrá que utilizar *técnicas analíticas o inferenciales*, considerando la variable dependiente como explicada por las demás variables independientes explicativas, y tratando de *relacionar todas las variables* por medio de una posible ecuación o modelo que las ligue. El método elegido podría ser entonces la *regresión lineal*, generalmente con todas las variables cuantitativas.

Una vez configurado el modelo matemático se podrá llegar a *predecir el valor de la variable dependiente* conocido el perfil de todas las demás. Si la variable dependiente fuera cualitativa dicotómica (1,0; sí o no) podrá usarse como clasificadora, estudiando su relación con el resto de variables clasificativas a través de la *regresión logística*.

Si la variable dependiente cualitativa observada constatará la asignación de cada individuo a grupos previamente definidos (dos, o más de dos), puede ser utilizada para *clasificar nuevos casos en que se desconozca el grupo a que probablemente pertenecen*, en cuyo caso estamos ante el *análisis discriminante*, que resuelve el problema de asignación en función de un perfil cuantitativo de variables clasificativas.

Si la variable dependiente es cuantitativa y las explicativas son cualitativas estamos ante los *modelos del análisis de la varianza*, que puede extenderse a los *modelos loglineales* para el análisis de tablas de contingencia de dimensión elevada. Si la variable dependiente puede ser cualitativa o cuantitativa y las independientes cualitativas, estamos ante el caso de la *Segmentación*.

Con la intención de clarificar un poco más ese tipo de técnicas de análisis de la dependencia se presenta el cuadro siguiente, que las clasifica en función de la naturaleza métrica o no métrica de las variables independientes y dependientes.



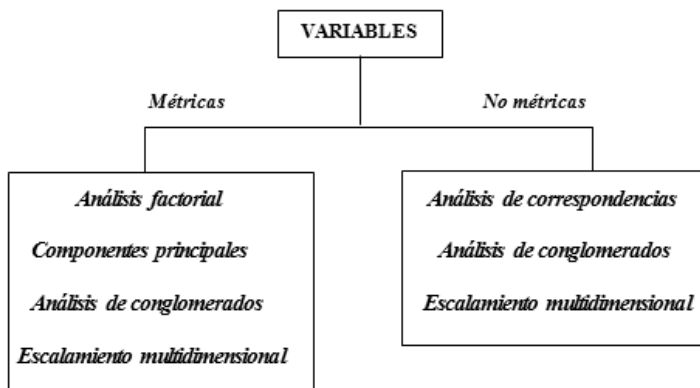
MODELADO ORIGINADO POR LOS DATOS (TÉCNICAS DESCRIPTIVAS O DE LA INTERDEPENDENCIA)

El investigador tendrá que considerar si asigna a todas sus variables una importancia equivalente, es decir, si *ninguna variable destaca como dependiente principal (MÉTODOS DE INTERDEPENDENCIA)* en el objetivo de la investigación. Si es así, puede acudir para su tratamiento en bloque a lo que podría llamarse *técnicas multivariantes descriptivas*.

Y puede hacerlo con dos orientaciones diferentes: por una parte, para *reducir la dimensión de una tabla de datos excesivamente grande* por el elevado número de variables que contiene y quedarse con unas cuantas variables ficticias que, aunque no observadas, sean combinación de las reales y sinteticen la mayor parte de la información contenida en sus datos. En este caso también deberá tener en cuenta el tipo de variables que maneja. Si son *variables cuantitativas*, las técnicas que le permiten este tratamiento pueden ser el *Análisis de componentes principales* y el *Análisis factorial*, si son *variables cualitativas*, puede acudir al *Análisis de correspondencias*, y si son variables cualitativas ordinales se acude al *Escalamiento multidimensional*.

La *Tipología* acepta *variables cualitativas y cuantitativas*. Por otra parte, la otra orientación posible ante una colección de variables sin ninguna destacada en dependencia, sería la de *clasificar sus individuos en grupos más o menos homogéneos con relación al perfil* que en aquellas presenten, en cuyo caso utilizará por ejemplo el *Análisis de clusters*, donde los grupos, no definidos previamente, serán configurados por las propias variables que utiliza.

Con la intención de clarificar un poco más este tipo de técnicas de análisis de la interdependencia se presenta el cuadro siguiente, que las clasifica en función de la naturaleza métrica o no métrica de las variables.



EL ENTORNO DE TRABAJO DE IBM SPSS

Para empezar a trabajar con el programa, basta elegir la opción *Programas* del menú *Inicio* de Windows y seleccionar la subopción *IBM SPSS Statistics 20* de la opción *IBM SPSS Statistics* (Figura 1-1). Se obtiene la pantalla temporal de la Figura 1-2 con información de la versión. A continuación se obtiene la pantalla de la Figura 1-3, que nos permitirá comenzar la tarea con SPSS de varias formas. Se puede comenzar ejecutando el tutorial, introduciendo los datos para el análisis, ejecutando una consulta en lenguaje SPSS creada anteriormente, creando una nueva consulta o abriendo una fuente de datos ya existente u otro tipo de archivo a elegir en la Figura 1-3. No obstante, si se desea que esta pantalla no vuelva a aparecer al ejecutar posteriormente SPSS, basta señalar la casilla *No volver a mostrar este cuadro de diálogo*. Si se pulsa *Cancelar*, se obtiene la pantalla de entrada de SPSS con la hoja de cálculo disponible para introducir los datos de nuestro análisis en sus celdas (Figura 1-4). Cada columna será una variable.

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

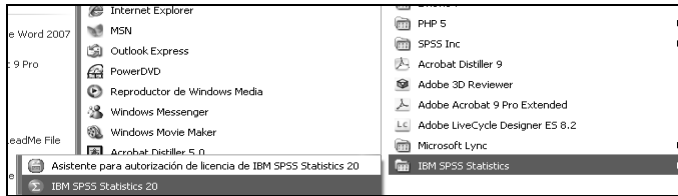


Figura 1-1

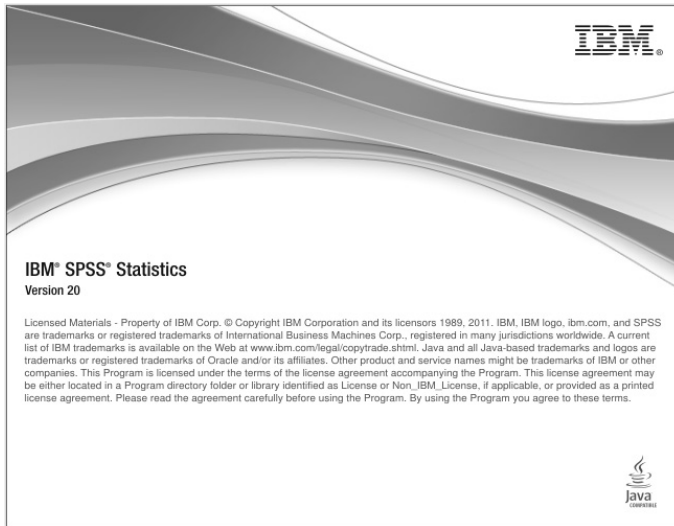


Figura 1-2

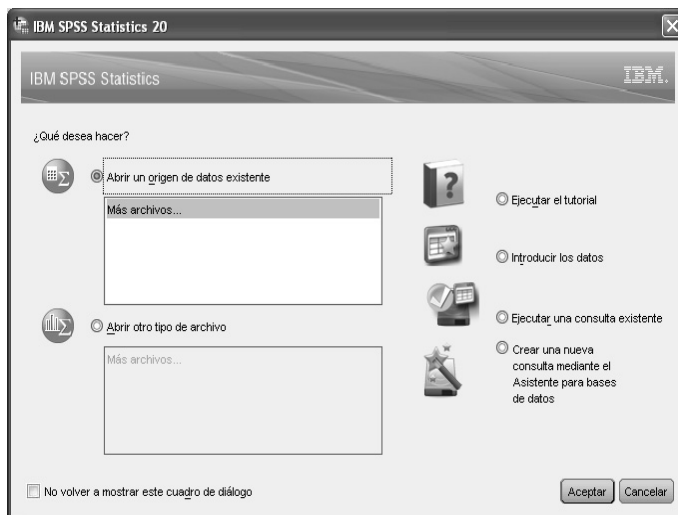


Figura 1-3

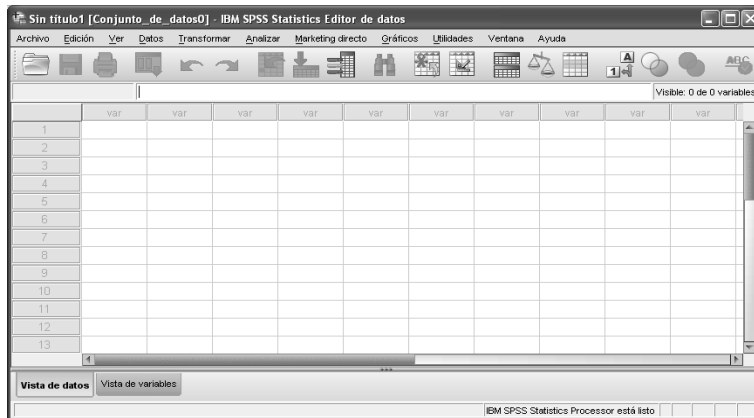


Figura 1-4

La pantalla inicial de la Figura 1-4 (*Editor de datos de SPSS*), que es el marco de trabajo inicial que se utilizará para introducir los datos y elegir el procedimiento adecuado para el análisis. En la línea superior de esta pantalla vemos el icono de SPSS y el nombre del archivo de datos activo. En la línea siguiente se presenta la *barra de menú*, que contiene el menú general de SPSS con todas sus opciones (*Archivo, Edición, Ver, Datos, Transformar, Analizar, Marketing directo, Gráficos, Utilidades, Ventana y Ayuda*). Cada una de estas opciones contiene distintos procedimientos para el trabajo estadístico. La tercera línea presenta la *barra de herramientas*, cuyo contenido son diferentes iconos que permiten acceder rápidamente a los procedimientos más comunes en el trabajo con la aplicación, sin necesidad de acudir al menú general (Figura 1-5). El significado de cada icono puede verse dinámicamente situando el ratón sobre el propio icono.



Figura 1-5

Por orden de colocación de izquierda a derecha, los iconos de la barra de herramientas significan lo siguiente:

- Abrir archivo
- Guardar archivo
- Imprimir
- Recuperar cuadros de diálogo
- Deshacer
- Rehacer
- Ir a caso
- Ir a variable
- Variables
- Buscar

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

- Insertar caso
- Insertar variable
- Segmentar archivo
- Ponderar casos
- Seleccionar casos
- Etiquetas de valor
- Usar conjuntos de variables
- Mostrar todas las variables
- Corregir ortografía

Cuando se ha abierto un archivo en el editor (por ejemplo, abrimos el archivo *accidents.sav* mediante *Archivo* → *Abrir* → *Datos*), en la línea situada debajo de la barra de herramientas se presenta el nombre de la variable relativa a la celda en la que están situados actualmente el cursor y su valor (Figura 1-6). En la parte inferior de la Figura 1-6 se presentan las solapas *Vista de datos* (Figura 1-6) y *Vista de variables* (Figura 1-7) para el fichero cargado actualmente. En la *Vista de variables* se define nombre, tipo, anchura, número de decimales, etiquetas, rango de valores para las categorías de las variables cualitativas, valores perdidos, columnas, tipo de variable y papel de las variables. Para las variables, y en la *Vista de datos* se introducen sus datos sobre las celdas. Cada columna es una variable.

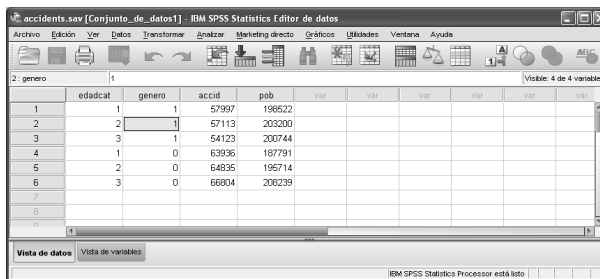


Figura 1-6

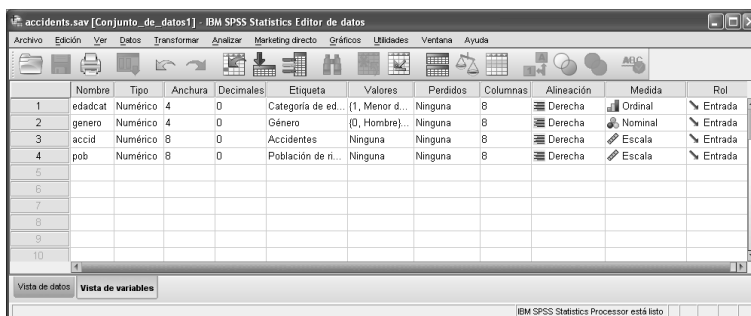


Figura 1-7

IBM SPSS contiene los procedimientos de minería de datos en las opciones del menú general *Datos* (Figura 1-8), *Transformar* (Figura 1-9), y sobre todo en *Analizar* (Figura 1-10). En las opciones *Datos* y *Transformar* se encuentran varios procedimientos relativos a las fases de selección y transformación de minería de datos como iremos viendo posteriormente a lo largo de los capítulos correspondientes. En la opción *Analizar* se encuentran todos los procedimientos relativos a las fases de exploración y limpieza de datos, análisis de datos descriptivo y predictivo y evaluación. Esta opción es la más importante del programa en cuanto a minería de datos.

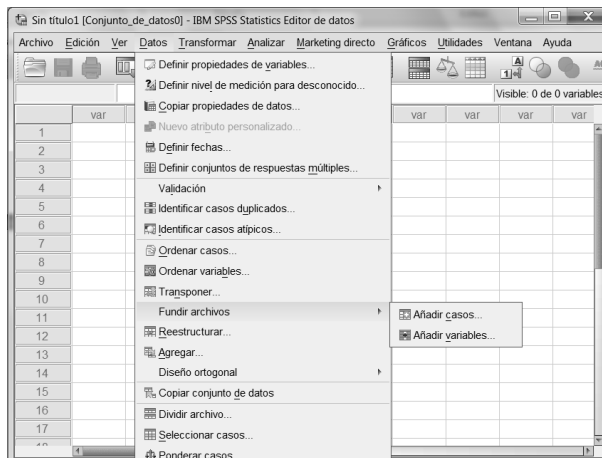


Figura 1-8

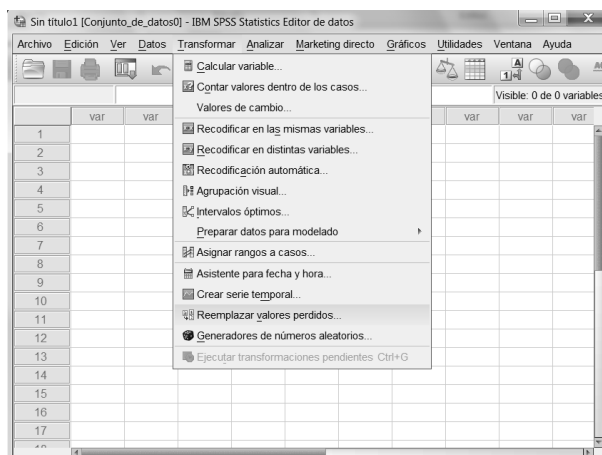


Figura 1-9

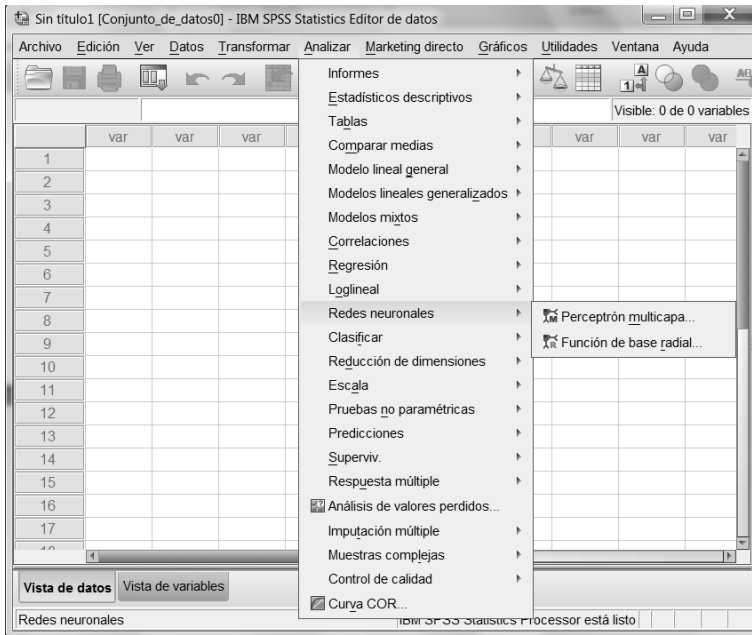


Figura 1-10

ENTORNO DE TRABAJO DE IBM SPSS MODELER

IBM SPSS Modeler es el entorno de trabajo de IBM para minería de datos. Este software ha sufrido varios cambios en su aspecto externo en los últimos años, pero la filosofía y el modo de trabajo no han cambiado desde las primeras versiones. De hecho, el nombre de este software hasta la versión 12 era SPSS Clementine. Pero el aspecto del programa no ha variado. En este libro me referiré indistintamente a las dos nomenclaturas del programa y todo el trabajo es válido para ambas. El entorno de IBM SPSS Modeler (Clementine) está basado en nodos que se van utilizando y conectando para formar un flujo, o *stream*, traducido por Clementine también como “ruta”. Los *streams* pueden alojarse en ficheros separados (.str) o se pueden organizar en proyectos. De hecho, tanto los *streams* como los proyectos de minería de datos se almacenan en ficheros separados que se pueden cargar, guardar, modificar, reejecutar o reorganizar y que son independientes de las fuentes de datos.

En la Figura 1-11 se muestra la estructura típica de un trabajo en Modeler (Clementine) con varios nodos interconectados que van realizando tareas secuenciales adaptadas a la teoría de la minería de datos. En concreto se muestra una ruta que finaliza con un árbol de decisión y una red neuronal para predecir efectos de drogas.

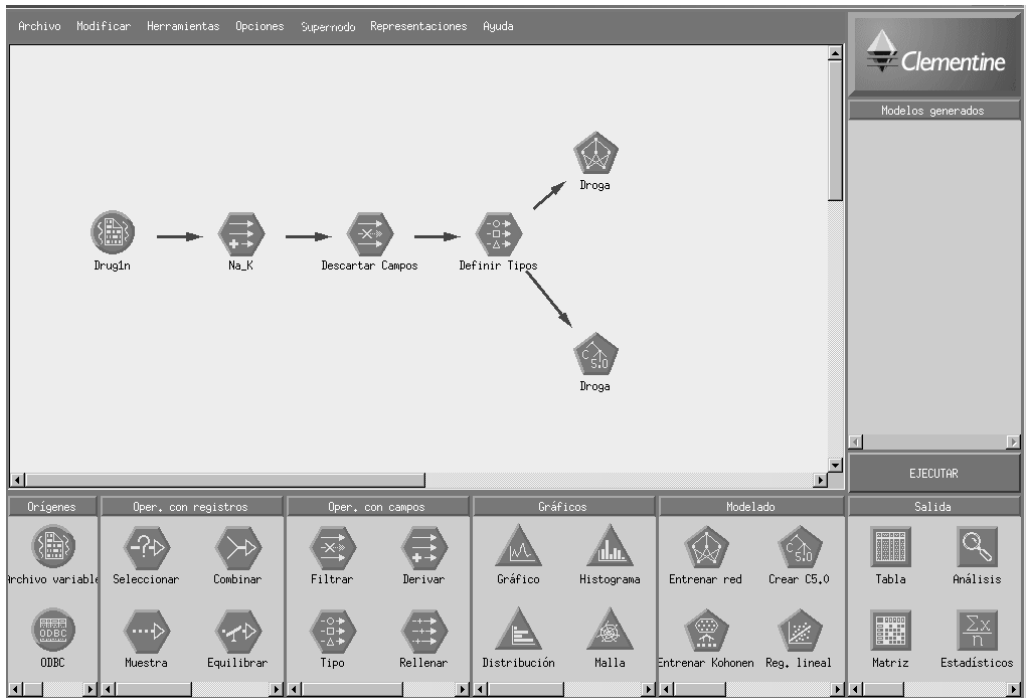


Figura 1-11

Modeler presenta varias paletas en la parte inferior de la Figura 1-11 que clasifican los nodos en seis categorías (Figura 1-12):

- *Orígenes*: nodos para situar las fuentes de datos en el entorno de trabajo.
- *Oper. con registros*: nodos para realizar selecciones y combinaciones con la finalidad de modificar o combinar registros (filas) de distintas fuentes.
- *Oper. con campos*: nodos para modificar o combinar campos (columnas).
- *Gráficos*: nodos para realizar gráficos.
- *Modelado*: nodos para trabajar con modelos predictivos y técnicas descriptivas.
- *Salida*: nodos para tabular datos, presentar estadísticas, exportar datos y analizar modelos.

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

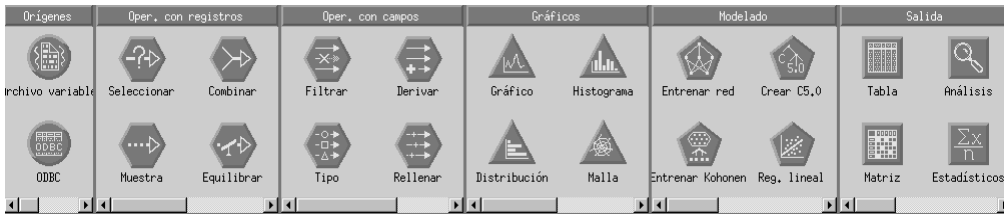


Figura 1-12

En la parte superior derecha de la pantalla se encuentra la paleta *Modelos generados* (Figura 1-13) que muestra los resultados de los modelos generados durante el trabajo. Estos modelos podrán ejecutarse posteriormente desde esta paleta.



Figura 1-13

ENTORNO DE TRABAJO DE SAS ENTERPRISE MINER

Para acceder a SAS Enterprise Miner es necesario ejecutar SAS previamente. A continuación, basta con escribir *miner* en la caja de comandos de SAS Explorer o, en los menús de SAS, ejecutar *Soluciones* → *Análisis* → *Enterprise Miner* (Figura 1-14).

Ambas vías nos llevan a la ventana de la aplicación *Enterprise Miner* de la Figura 1-15, en la que se distinguen las siguientes partes significativas:

Zona de trabajo: área en la que se realizan las tareas de minería de datos.

Barra de herramientas: zona de menú de acceso directo a los procedimientos más importantes de minería de datos (Figura 1-16).

Navegador del proyecto: presenta pestañas para realizar diagramas, presentar herramientas y realizar informes. La pestaña de diagramas (*Diagrams*) del navegador de proyectos permite seleccionar los proyectos y diagramas abiertos. La pestaña herramientas (*Tools*) despliega el conjunto de tareas que pueden ser realizadas por *Enterprise Miner* (Figura 1-17). La pestaña de informes (*Reports*) contiene los informes generados a través de nodos de informes.

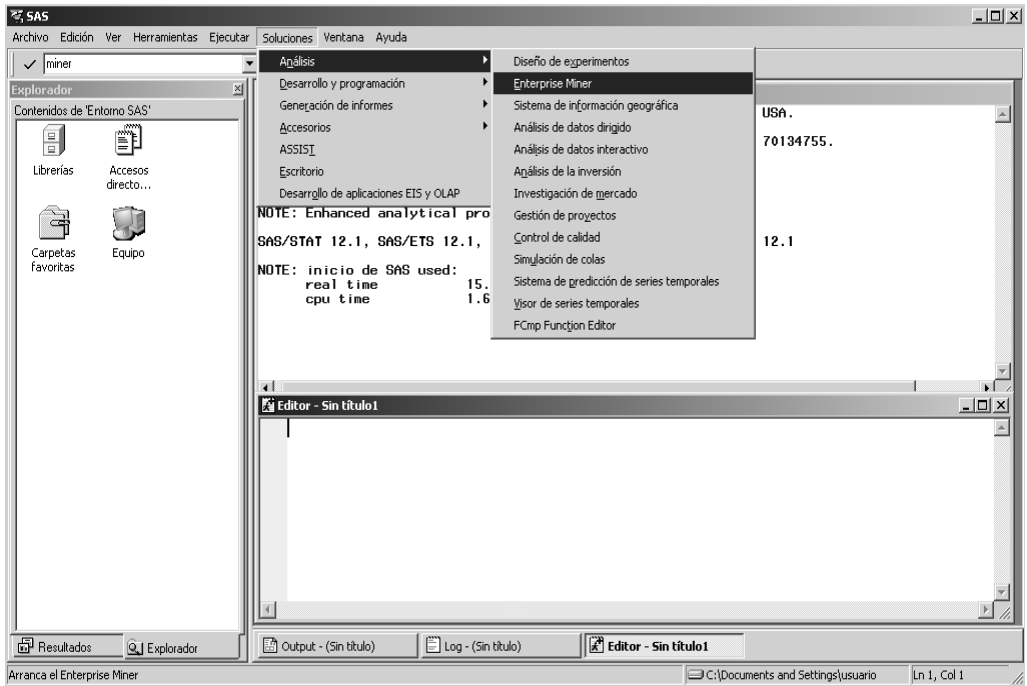


Figura 1-14

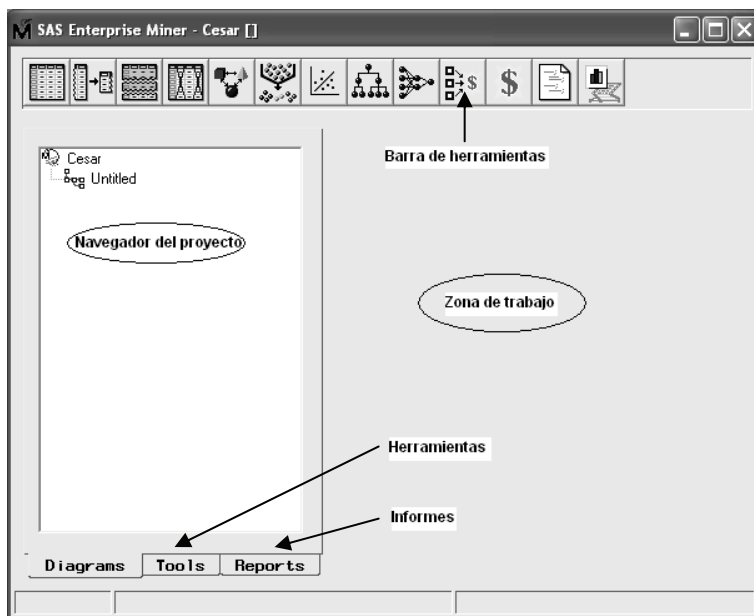


Figura 1-15

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

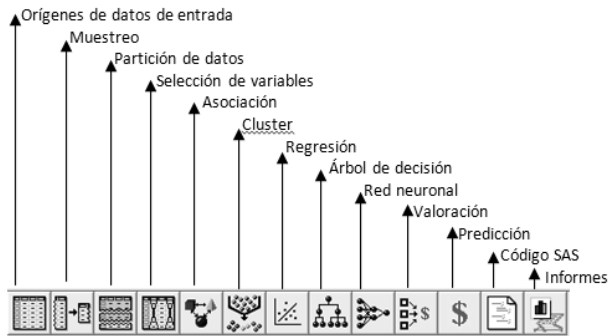


Figura 1-16

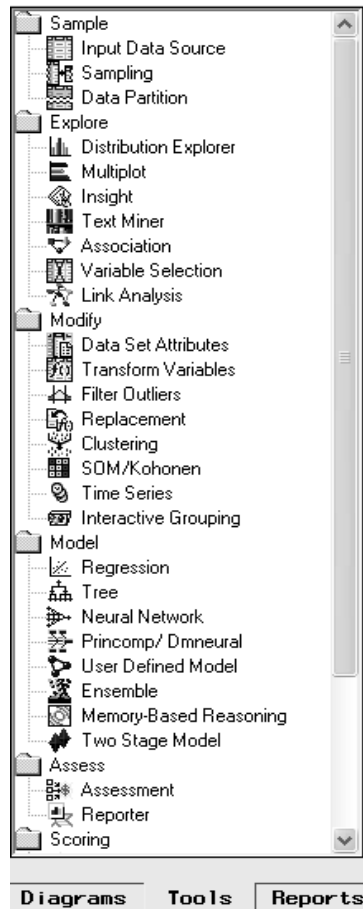


Figura 1-17

Los grupos de herramientas del menú de Enterprise Miner se adaptan a las fases de la metodología SEMMA (*Sample, Explore, Modify, Model* y *Assess*) de SAS para la minería de datos. Esta sostiene que en el proceso de extracción del conocimiento se observa la secuencia de fases siguiente: SELECCIÓN → EXPLORACIÓN → LIMPIEZA → TRANSFORMACIÓN → MINERÍA DE DATOS → EVALUACIÓN → DIFUSIÓN. La fase de *Selección* SAS la denomina *Sample*; la fase de *Exploración* SAS, *Explore*, las fases de *Limpieza* y *Transformación* SAS, *Modify*, la fase de *Análisis de Datos* (o Minería de Datos) SAS, *Model* y la fase de evaluación SAS, *Assess*.



Para empezar a trabajar con SAS Enterprise Miner es necesario comenzar definiendo un proyecto. Para ello se selecciona *File* → *New* → *Project* (Figura 1-18). A continuación, en la pantalla *Create new project* (Figura 1-19) se eligen el nombre y la ubicación del proyecto. Al hacer clic en *create* se crea un proyecto sin título (*untitled*) que será nombrado al guardarlo (Figura 1-20). Una vez creado el proyecto se habilita la zona de trabajo para realizar tareas de minería de datos.

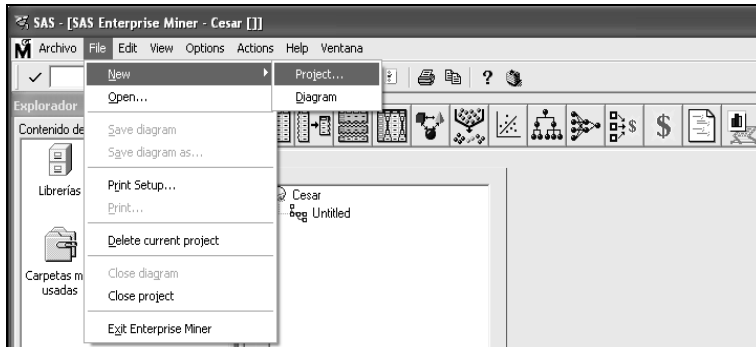


Figura 1-18

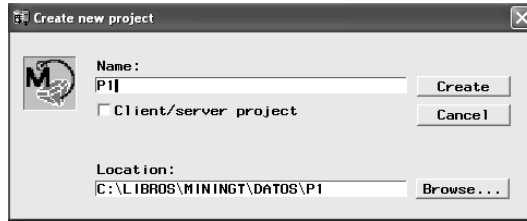


Figura 1-19

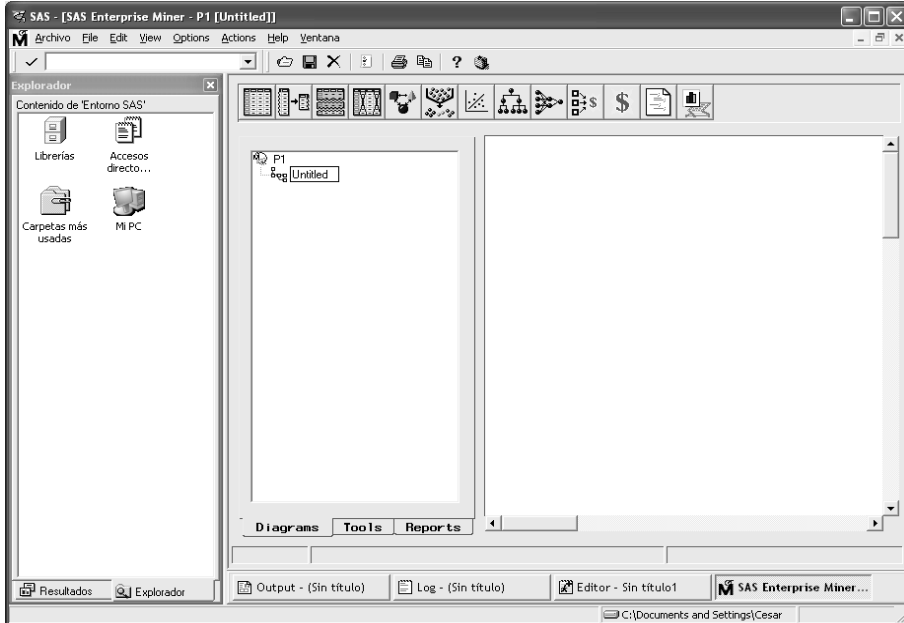


Figura 1-20

LOS ENTORNOS DE TRABAJO DE SAS Y SAS ENTERPRISE GUIDE

El software SAS a través de su módulo SAS Enterprise Guide permite realizar tareas de minería de datos por medio de menús de una forma sencilla. Basta hacer doble clic sobre la opción *SAS Enterprise Guide 5.1* de la Figura 1-21 para obtener la pantalla de entrada de la aplicación (Figura 1-24) después de la pantalla temporal de la Figura 1-22 y de la elección de la opción *Nuevo proyecto* de la Figura 1-23. Las opciones del menú *Tareas* de la Figura 1-24 permiten realizar trabajos de minería de datos. La opción *Tareas* permite realizar los distintos trabajos. Pero antes es necesario asignar una librería al proyecto que contenga los conjuntos de datos a utilizar con el programa mediante la opción *Herramientas* → *Asignar una librería al proyecto* (Figura 1-25). Las Figuras 1-26 a 1-29 muestran los pasos para la asignación.

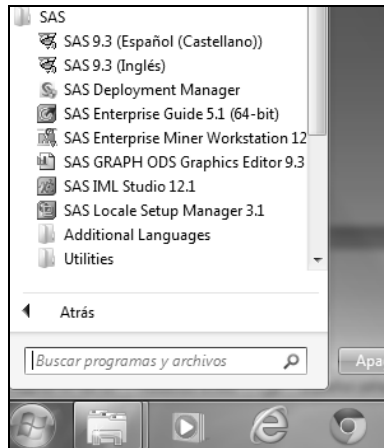


Figura 1-21

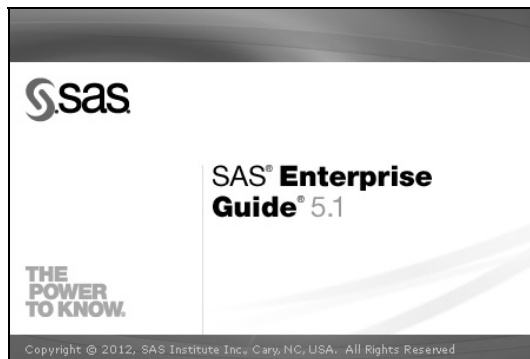


Figura 1-22

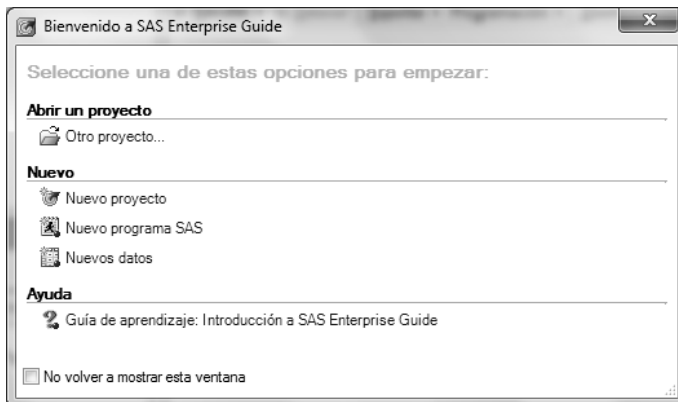


Figura 1-23

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

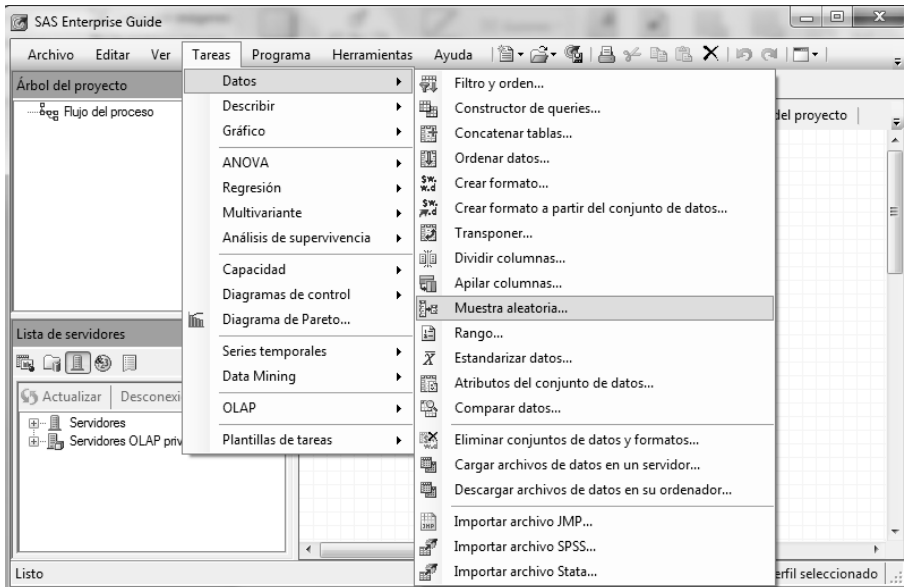


Figura 1-24

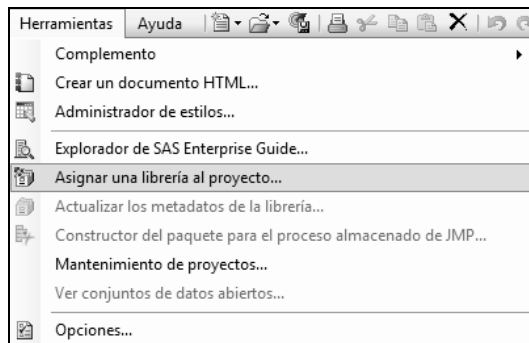


Figura 1-25

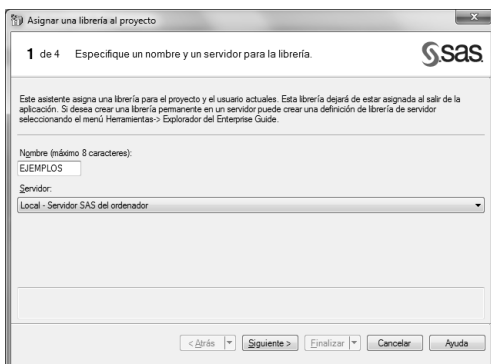


Figura 1-26

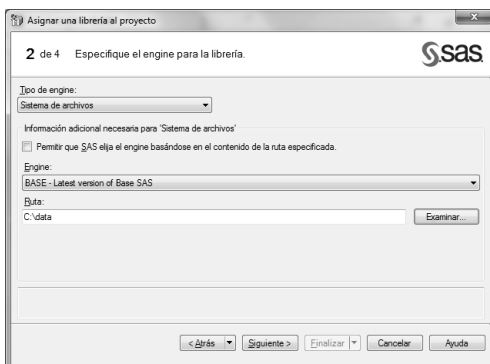


Figura 1-27

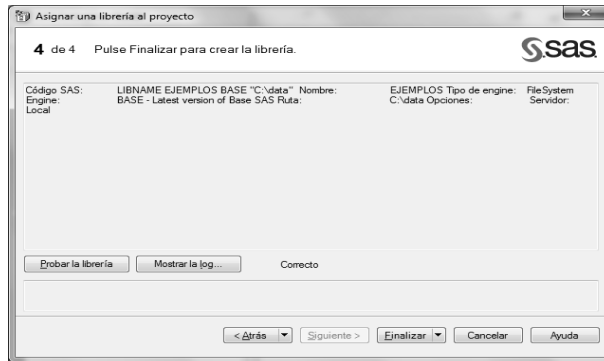


Figura 1-28

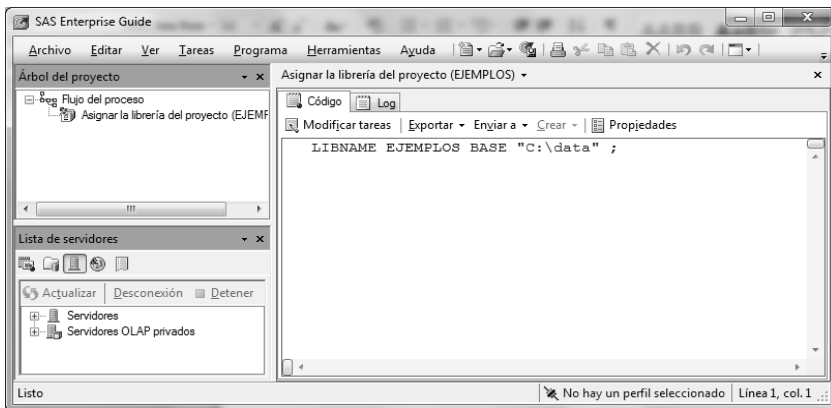


Figura 1-29

La opción SAS 9.3 de la Figura 1-21 nos lleva al marco de trabajo de SAS (Figura 1-31) después de la pantalla temporal de la Figura 1-30.

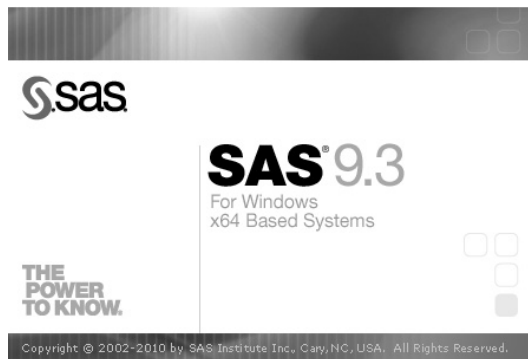


Figura 1-30

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

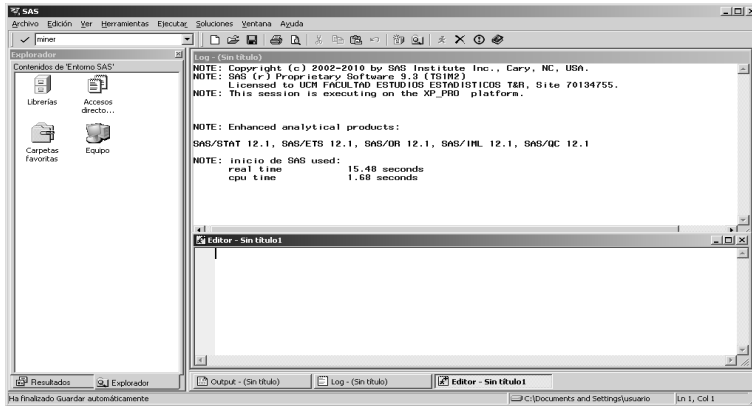


Figura 1-31

Para utilizar de forma óptima la extraordinaria flexibilidad y potencia de SAS, es necesario **trabajar con los procedimientos SAS**, lo que exige como mínimo un conocimiento básico del lenguaje de programación de SAS y en concreto de la estructura de cada uno de los procedimientos. En este texto se analizará la sintaxis de los citados procedimientos y se ilustrará con ejemplos. Como muestra podemos realizar un ajuste de regresión con variables cualitativas y cuantitativas (Figura 1-32).

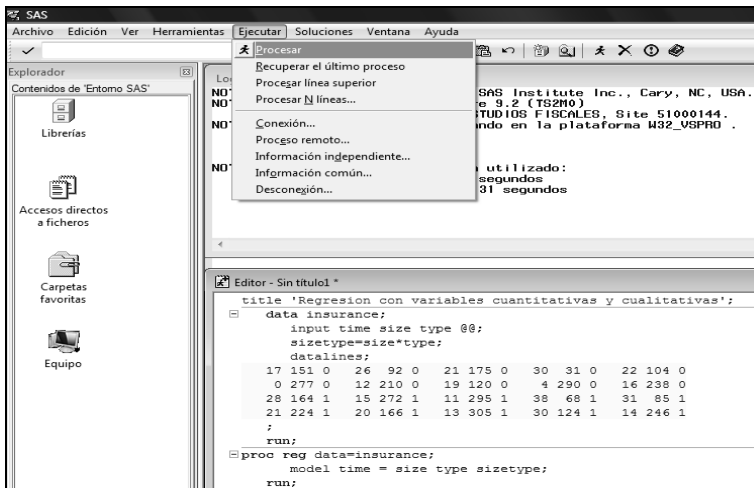



Figura 1-32

Al ejecutar el procedimiento con el icono  o con la opción *Procesar* del menú *Ejecutar*, el entorno de SAS presenta por defecto las tres ventanas de la Figura 1-33. En la parte superior aparece la **ventana OUTPUT**, que presenta la salida completa del programa procesado. Si maximizamos esta ventana, se observará la salida completa (Figura 1-34).

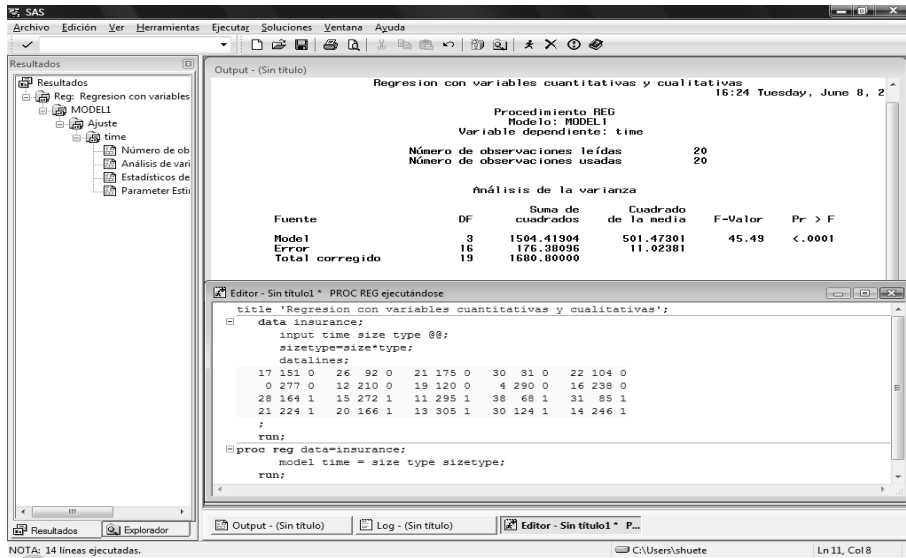


Figura 1-33

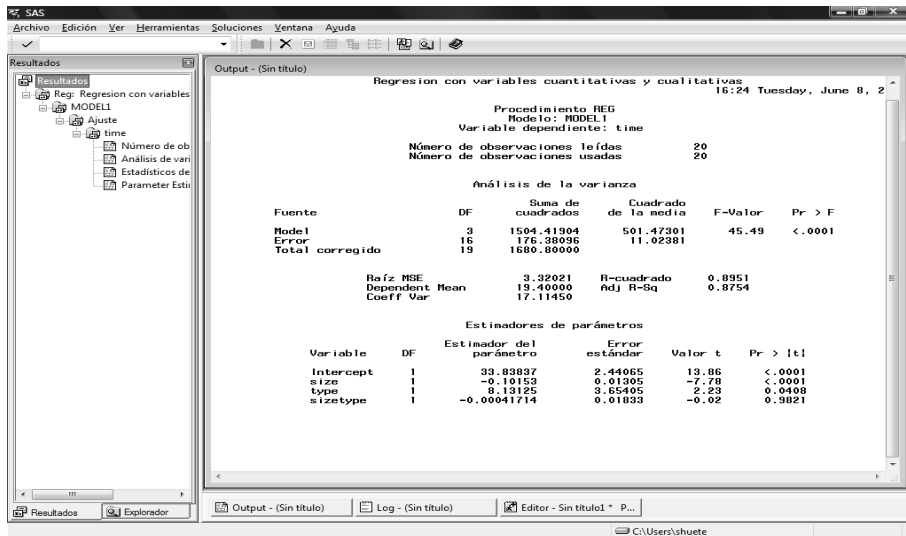


Figura 1-34

La ventana inferior de la Figura 1-33 es la **ventana PROGRAM EDITOR**, que se ha utilizado para escribir el programa SAS. Haciendo clic en el botón **Log** de la barra de iconos de ventanas de la parte inferior de la pantalla, se obtiene la **ventana LOG** que resume todas las incidencias de la ejecución de los programas SAS (Figura 1-35).

MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS

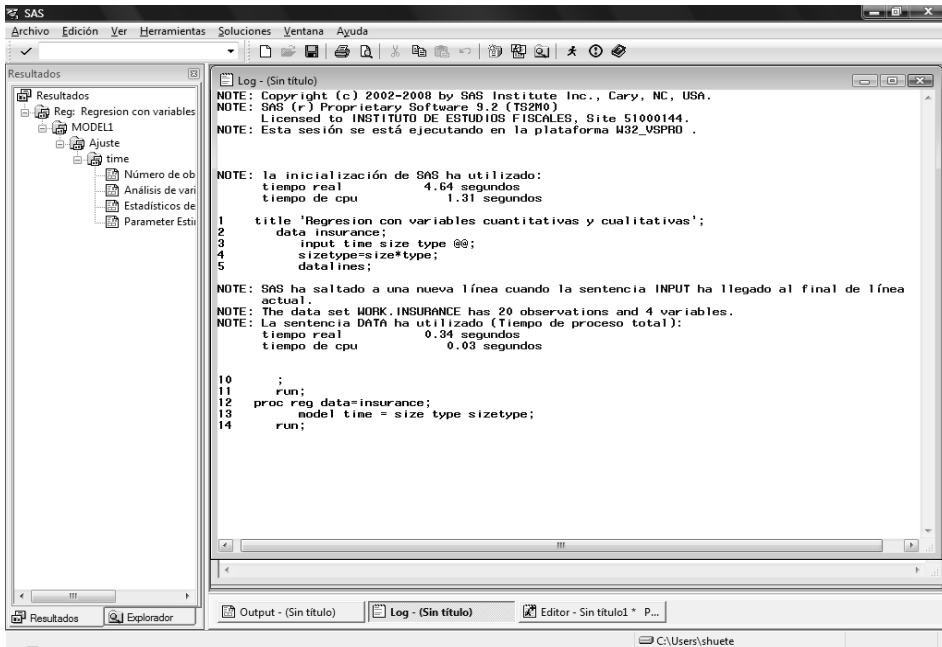



Figura 1-35

En la parte derecha de la pantalla se sitúa la **ventana RESULTADOS**, que divide en zonas la salida de la ventana OUTPUT. Haciendo clic en una entrada de esta ventana, se ve la parte de la salida correspondiente.

Para crear un programa en SAS, se escribe el código de programación directamente sobre la ventana PROGRAM EDITOR. Una vez finalizada la sintaxis, para ejecutar el programa se selecciona la opción *Procesar* del menú *Ejecutar* de SAS. También se puede *ejecutar solo una parte de la sintaxis* seleccionándola en la ventana PROGRAM EDITOR y haciendo clic en *Procesar*. Una vez ejecutado el programa, si este tiene salida gráfica, se mostrará en la **ventana GRAPH**, que es otro tipo de ventana que se maneja en SAS. Para pasar de un tipo de ventana a otro pueden utilizarse los menús *Ver* o *Ventana* o los botones de la barra de iconos de ventanas de la parte inferior de la pantalla.

Es común tener que corregir un programa una vez ejecutado por fallos de sintaxis (que se detectan en la venta LOG). Si nos situamos con el cursor del ratón en la ventana PROGRAM EDITOR y seleccionamos la opción *Recuperar el último proceso* del menú *Ejecutar* de la Figura 1-32, se recupera el texto del programa tal y como se había escrito antes de ejecutarlo. Se realizan las correspondientes correcciones y se vuelve a ejecutar con la opción *Submit* o con el icono .