

Manual de Introducción a SAS Enterprise Guide[®] v4.1

Llorenç Badiella. Director Técnico del Servei d'Estadística Natàlia Adell. Asesora estadística del Server d'Estadística Anabel Blasco. Asesora estadística del Servei d'Estadística Daniel Cuadras. Asesor estadístico del Servei d'Estadística Anna Espinal. Asesora estadística del Servei d'Estadística Oliver Valero. Asesor estadístico del Servei d'Estadística

Manual de Introducción a SAS Enterprise Guide® v4.1

Servei d'Estadística Universitat Autònoma de Barcelona

Campus UAB - Edifici D 08193 Cerdanyola del Vallès (Barcelona) Tel. 93.581.13.47 s.estadistica@uab.es www.uab.es/s-estadistica/

Manual de Introducción a SAS Enterprise Guide® v4.1			
Dublicado por al Correi d'Estadística de la UAR			
Publicado por el Servei d'Estadística de la UAB			
Diciembre 2007.			
Este documento puede ser copiado y libremente distribuido, siempre y cuando sea preservada su integridad, referenciado su origen y comunicado su uso al Servei d'Estadística de la UAB. No está permitido añadir, borrar o cambiar ninguna de sus partes, o extraer páginas para su uso en otros documentos.			

CONTENIDOS

1	PRESENTACIÓN	6
2	INICIO DE LA SESIÓN EN SAS EG 4.1	9
2.1	Inicio de la sesión	10
2.3	Gestión de servidores y privilegios	
3	LECTURA DE DATOS	
3.1 3.2 3.3 3.4	Introducción de datos	
3.5	Importar datos	
4	FUSIÓN DE ARCHIVOS	26
4.1 4.2 4.3	Ordenar datos	28
5	GESTIÓN DE DATOS	
5.1 5.2 5.3 5.4 5.5	Filtrar casos Filtros dinámicos Agregar datos Contar datos Crear nuevas variables	
6	RESUMEN DE DATOS	
6.1 6.2 6.3 6.4 6.5	Listar Datos Resumir los datos Tablas de resumen personalizadas Tablas de frecuencia Tablas de contingencia	
7	GRÁFICOS	53
7.1 7.2 7.3 7.4 7.5	Gráfico de sectores Gráfico de barras Histograma Diagrama de dispersión Diagrama de cajas	
8	OTRAS OPERACIONES CON DATOS	62
8.1	Transposición de datos	62
9	PRESENTACIÓN	65
9.1	Presentación de resultados	65
10	ASPECTOS AVANZADOS	66
10.1 10.2		
11	EJERCICIOS	68
11.1 11.2 11.3 11.4	Ejercicio 1	
11.4	r Ejeteteto 7	0>

Manual de Introducción a SAS Enterprise Guide® v4.1

12	ANEXO	70
11.5	Ejercicio 5	69
		_

PRESENTACIÓN

Durante los últimos años la estadística aplicada ha ganado mucha importancia en el mundo de la investigación, debido principalmente a la gran potencia de cálculo de los ordenadores y en especial al elevado rendimiento de los diferentes paquetes estadísticos.

Los paquetes estadísticos también se han adaptado a las nuevas necesidades, de manera que ahora ya no se habla de "paquetes estadísticos", término demasiado concreto, sino de sistemas de soporte a la toma de decisiones. Por ejemplo, **SAS** inicialmente significaba "Statistical Analysis System", pero a medida que este programa iba incorporando nuevas capacidades la mayoría de las cuales no eran propiamente estadísticas, las siglas dejaron de tener sentido.

Actualmente, los diferentes módulos de **SAS** hacen que éste sea un software de los que se llaman como "de inicio a fin". Permite crear gráficos, trabajar como una hoja de cálculo, compilar programas en lenguaje C, incluye herramientas para construir interfases para la WWW, herramientas para construir y tratar el Datawarehouse, para explotar datos con la filosofía del Datamining, etc. Todas estas tareas se pueden llevar a cabo empleando un interfaz amigable (**SAS Enterprise Guide**), útil para usuarios finales o bien mediante código de programación (**SAS System**), especialmente útil para los usuarios responsables de producir las mencionadas tareas.

Este manual de introducción a **SAS Enterprise Guide** para Windows pretende ser una primera aproximación al uso de uno de los programas del software SAS para aquellas personas con algunos conocimientos de estadística que quieran iniciarse en el tratamiento de los datos mediante este programa. Este manual sólo trata aquellas instrucciones más básicas y sus opciones más habituales. Es altamente recomendable haber tenido contactos previos con otros programas de análisis y manipulación de datos, hojas de cálculo, bases de datos y lenguajes de programación. Por otro lado, se da por supuesto que el usuario ya tiene experiencia suficiente en el entorno Windows.

Cada sección consta de diferentes apartados: leer datos, tabular los datos, etc. pero en cada uno de ellos se introducen a modo de observaciones diversas características generales con el objetivo de remarcar otros detalles necesarios del funcionamiento del programa. Por este motivo es recomendable seguir el manual de forma lineal y consecutiva.

Para seguir los ejemplos propuestos, es necesario almacenar los ficheros del curso en el directorio de trabajo, siendo "C:\Temp\" el directorio de trabajo utilizado en el presente manual. Los datos se pueden conseguir en la página web www.uab.es/s-estadistica en la sección de materiales.

SAS Institute proporciona distintas herramientas para el manejo y análisis de datos cada una destinada en principio a usuarios con diferentes perfiles

PRODUCTOS SAS INSTITUE

SAS System	Manipulación de datos y análisis estadístico	Producción
SAS Enterprise Guide	Facilidad para el análisis estadístico	Usuario final
SAS Enterprise Miner	DataMining	Analista DM

El programa **SAS System** proporciona un entorno de trabajo para usuarios programadores, ya que todas las instrucciones se ejecutan desde una ventana de sintaxis.

Los procedimientos necesarios para llevar a cabo las distintas tareas con **SAS System** se encuentran agrupados en módulos. Los diferentes módulos del programa **SAS** son los siguientes:

En negrita los módulos más habituales.

MÓDULOS DE SAS

BASE	SAS Básico			
SAS/ACCESS	Interfase a otros formatos de ficheros			
SAS/AF	Lenguaje para el diseño de aplicaciones SAS			
SAS/ASSIST	Asistencia para el uso de SAS			
SAS/CONNECT	Conjunto de herramientas cliente/servidor			
SAS/EIS	Permite construir Bases de datos multidimensionales			
	(MDDB) y relacionales (RDBMS)			
SAS/ETS	Análisis de Series Temporales			
SAS/FSP	Facilidades para la entrada de datos de forma			
	interactiva			
SAS/GIS	GIS (Sistema de Información Geográfica)			
SAS/GRAPH	Módulo gráfico			
SAS/IML	Manipulación de matrices			
SAS/INSIGHT	Herramienta para explorar y analizar datos			
SAS/MDDB	Trabajo con Bases de datos Multidimensionales			
SAS/ODBC	Conexiones ODBC			
SAS/OR	Para resolver problemas de optimización			
SAS/QC	Control de Calidad			
SAS/SPECTRAVIEW	Visualización de datos como representaciones gráficas			
SAS/STAT	Análisis Estadísticos			
SAS/INTRNET	Permite a los usuarios ejecutar programas desde un			
	navegador.			
SAS/WHAREHOSE	Herramientas para el administrador de repositorios de datos.			
ADMINISTRATOR	·			

Por otro lado, **SAS** Enterprise Guide viene a ser una máscara que permite utilizar la potencia de **SAS** sin utilizar la sintaxis.

De hecho, **SAS Enterprise Guide** permite llevar a cabo las tareas de los módulos:

BASE SAS Básico
SAS/GRAPH Módulo gráfico
SAS/STAT Análisis Estadísticos
SAS/QC Control de Calidad

SAS/ETS Análisis de Series Temporales

En definitiva, los módulos accesibles permiten realizar las tareas de manipulación de datos, reporting y análisis estadístico.

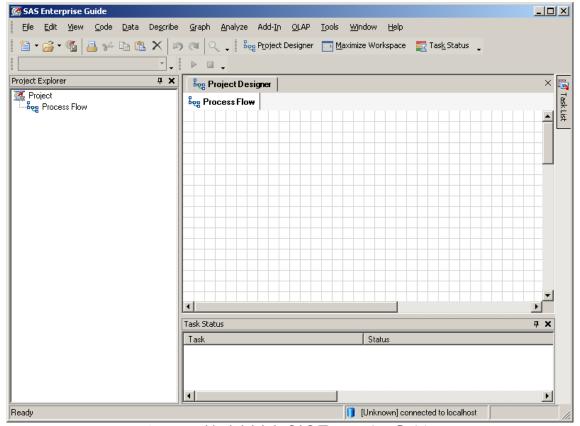
Desde **SAS** Enterprise Guide, el usuario podrá acceder a la mayoría de opciones disponibles por sintaxis en SAS System, aunque en ocasiones necesitará trabajar explícitamente con sintaxis para realizar determinados pasos.

2 INICIO DE LA SESIÓN EN SAS EG 4.1

2.1 Inicio de la sesión

SAS Enterprise Guide es una máscara que permite utilizar la potencia de **SAS** sin utilizar la sintaxis. Este programa proporciona:

- Un interfaz point-and-click a SAS System.
- Acceso transparente a bases de datos tanto de SAS como de otras fuentes.
- Ventanas interactivas que permiten realizar gran cantidad de tares de análisis y reporting.
- Capacidad de exportar resultados a otras aplicaciones Windows y a la Web.
- La posibilidad de programar la ejecución de tareas.



Presentación inicial de SAS Enterprise Guide

SAS Enterprise Guide está configurado de manera que el usuario crea un proyecto y añade diferentes bases de datos o tareas, obteniendo resultados o nuevas bases de datos.

Las tareas de **SAS Enterprise Guide** generan de forma automática código de **SAS System** y proporcionan resultados listos para crear informes.

El usuario de SAS System puede crear tareas de código, escribiendo directamente la sintaxis.

Al ejecutar el programa se abren distintas ventanas:

Process Flow – Aparecen en modo de Flujo de Tareas las operaciones que desea ejecutar y realizar el usuario.

Project – Aparecen en modo de lista, las mismas tareas que aparecen en la ventana Process Flow.

Task List – Lista de tareas disponibles.

Task Status – Estado de la ejecución de las tareas.

Server List – Lista de servidores con datos accesibles.

Workspace - Área de Resultados: Aparecen los resultados a medida que se ejecutan y las tablas de bases de datos.

Para minimizar las ventanas se puede pulsar el botón:

Para desplazar una ventanas se debe pulsar la barra superior y arrastrarla hasta el destino.

Además, en la barra de herramientas aparecen diversos menús desplegables, recogiendo las opciones estándar File, Edit, View, Window, Help, además de menús para grupos de tareas.

2.2 **Proyectos**

Para crear un proyecto nuevo, se puede pulsar el botón: 🍱



Se puede abrir un proyecto antiguo mediante los comandos:

FILE -> Open -> Project.

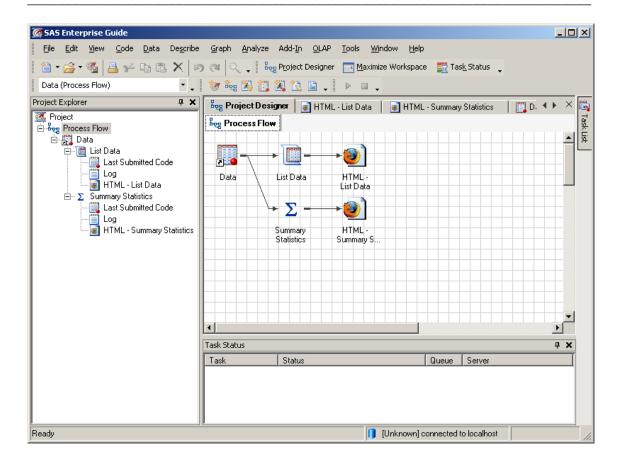
Se pueden añadir notas al proyecto con el botón: 🔼



Observación: Si no se visualizan los botones anteriores, pulsar:

View -> Toolbars -> Project

En la ventana **Process Flow** se muestra un diagrama del proyecto en forma de grafo cuyos nodos son las tareas y las bases de datos.



Ejemplo de proyecto con SAS Enterprise Guide

El usuario selecciona una base de datos desde la ventana de Servidores o una tarea desde la ventana de Tareas y la arrastra hasta la ventana del proyecto. A continuación modifica los parámetros de la tarea seleccionada y la ejecuta.

De esta forma se crean flujos de tareas hasta obtener tablas resumen o resultados gráficos.

Tras cada paso, se obtiene un listado con el código SAS ejecutado, siendo posible modificarlo y personalizarlo.

Se puede personalizar el aspecto de la ventana **Process Flow**, activándola, pulsando el botón derecho y seleccionando las opciones **Zoom** o **Grid**.

Al guardar un proyecto, el programa solicita en qué servidor se desea almacenar el proyecto. Al cerrarlo, solicita qué hacer con los datos temporales.

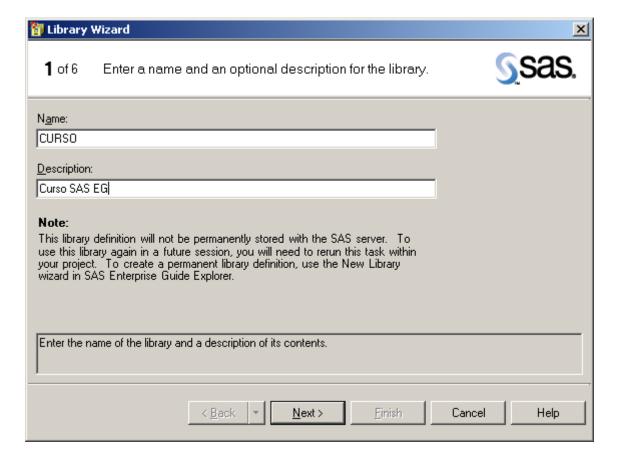
2.3 Gestión de servidores y privilegios

Para empezar a trabajar en un proyecto, es indispensable definir una "LIBRERÍA" personalizada.

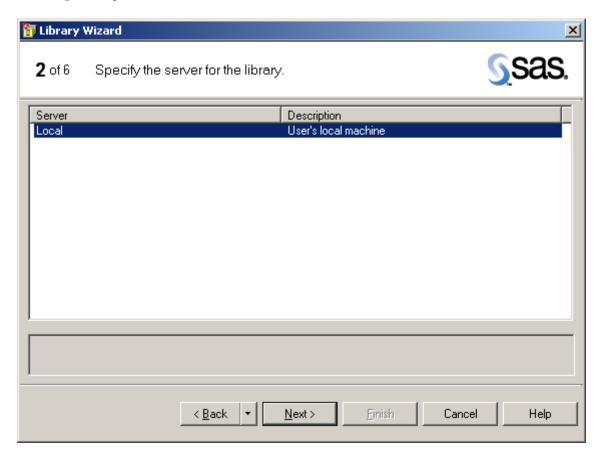
Las librerías son ubicaciones físicas del ordenador donde se almacenan las bases de datos del proyecto. Se pueden interpretar como si se tratara de una definición o enlace.

Tools -> Assign Library

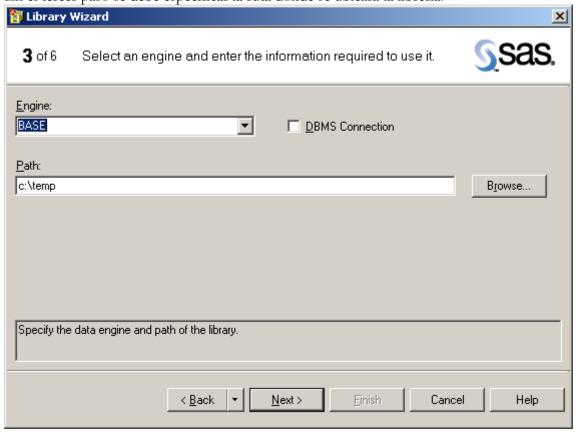
A continuación se escribe el nombre de la nueva librería y se asignan las propiedades.



En el siguiente paso, se selecciona el servidor.



En el tercer paso se debe especificar la ruta donde se ubicará la librería.

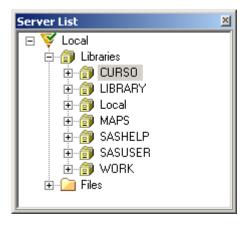


A continuación se aceptan las opciones pulsando Next, hasta Finalizar el proceso.

Ahora, puede comprobarse que la librería se ha creado adecuadamente, desplegando las librerías de la ventana Server List.

Si no se visualiza la nueva librería, pulsar el Botón derecho en la lista de servidores y seleccionar **Refresh**.

Para ver la ventana Server List, pulsar View -> Server List.



Es altamente recomendable crear una nueva librería para cada proyecto.

Con doble-click en la parte superior de la ventana, ésta se incorpora a la derecha de la ventana de trabajo.

3 LECTURA DE DATOS

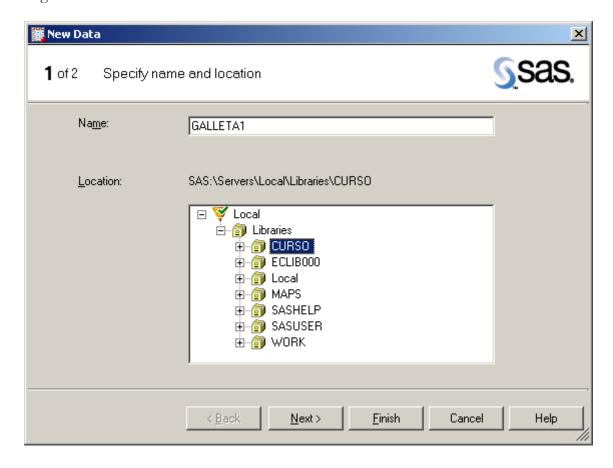
3.1 Introducción de datos

Para introducir datos directamente a **SAS EG**, es necesario abrir un nuevo **Dataset** con las opciones:

File -> New -> Data

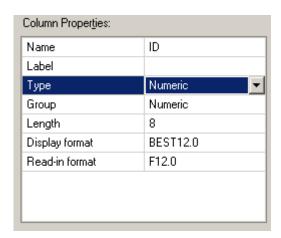
Para crear una nueva base de datos también se puede pulsar el botón:

Ahora, aparece una ventana dónde debemos seleccionar la librería de destino así como asignar un nombre al nuevo fichero de datos.



Si se selecciona la librería **WORK**, los datos serán temporales. Al cerrar el proyecto, el programa solicitará si se desean almacenar los datos temporales. Si se selecciona una librería creada por el usuario, los datos se almacenarán físicamente en la ruta asociada a la librería.

En el segundo paso se debe especificar el número de variables que serán introducidas, el nombre y la definición de tipo de éstas.



Si el usuario decide no editar las propiedades de las variables, al abrir la nueva base de datos, inicialmente, todas las variables son alfanuméricas (cadena), identificadas por el símbolo: . Tras introducir algún número, el programa detecta que se trata de una variable numérica, y cambia la identificación por el símbolo: .

Así, en estos datos de ejemplo, los nombres y etiquetas que corresponden a cada columna son:

A ->	ID	Numérica
B ->	SEXO (M/F)	Alfanumérica
C ->	TIPO (X=Chocolate / B=Blanco)	Alfanumérica
D ->	MARCA (S=marca original/N=marca blanca)	Alfanumérica
E ->	ENMASCARAMIENTO (S=Sí/N=No)	Alfanumérica
F ->	VALORACION (0-100)	Numérica
New ->	DULCE (4=Muy/3=Bastante/2=Poco/1=Nada)	Numérica

New Data 2 of 2 Create columns and specify their properties Columns: Column Properties: DULCE Name Length (in bytes) Name 📵 ID 8 Label 12 🔌 SEXO Туре Numeric 🕭 TIPO 12 Group Numeric 🔌 MARCA 12 Length 🔌 ENMAS... 12 BEST12.0 Display format 📵 VALORA... 8 Read-in format F12.0 📵 DULCE 8 <u>N</u>ew <u>D</u>uplicate Paste. Next> Finish Cancel k <u>B</u>ack Help

Y a continuación pulsar Finish

En la ventana **Process Flow**, aparece un nodo correspondiente a la nueva base de datos. También aparece en la ventana del Proyecto.

Tras hacer doble-click en el icono GALLETA1, se puede empezar a introducir datos.

Todas las bases de datos responden a la siguiente estructura:

Las filas son registros, sujetos, unidades experimentales, pacientes o cuadernos.

Las columnas son variables, ítems o cualidades.

Datos de ejemplo que deben ser introducidos manualmente:

8	eg Project I	Designer [3 GALLETA	1			
	₁ ID	<u></u> SEXO		<u></u> MARCA	ENMASCA RAMIENTO	VALORA CION	DULCE
1	1	F	×	S	S	75	3
2	1	F	×	N	S	70	4
3	1	F	В	S	S	50	2
4	1	F	В	N	S	60	2
5	1	F	Χ	S	N	80	4
6	1	F	Χ	N	N	70	4
7	1	F	В	S	N	60	2
8	1	F	В	N	N	45	4
9							
10							
11							
12							

Se pueden introducir datos procedentes de otras aplicaciones seleccionando, copiando y pegando directamente en la base de datos de SAS.

Cuando hay pocos datos, es recomendable eliminar las filas y columnas que el programa crea por defecto y no serán utilizadas, seleccionando los casos o las columnas sobrantes, pulsando el botón derecho y a continuación **Delete**.

Este ejemplo corresponde al siguiente estudio:

Para comprobar las preferencias de los consumidores de cierto tipo de galletas se realizó un experimento de cata, obteniendo su valoración global (0-100) y su apreciación del sabor dulce (4=Muy dulce/ 3=Bastante dulce / 2=Poco Dulce / 1= Nada dulce). Cada sujeto participante probó 8 galletas "diferentes" según la siguiente combinación:

Enmascarado	Marca	Tipo
No	Original	Chocolate negro
No	Original	Chocolate blanco
No	Blanca	Chocolate negro
No	Blanca	Chocolate blanco
Sí	Original	Chocolate negro
Sí	Original	Chocolate blanco
Sí	Blanca	Chocolate negro
Sí	Blanca	Chocolate blanco

En el estudio se incluyeron 32 sujetos.

3.2 Definición de etiquetas

Para obtener una presentación adecuada de los resultados y simplificar el manejo de los datos, es habitual identificar las variables con un nombre simple y con una etiqueta explicativa, y al mismo tiempo asignar etiquetas de valores a los códigos numéricos.

Pulsando el botón derecho encima del nombre de una variable y seleccionando Properties, se puede cambiar su nombre y su etiqueta.



Así, en este ejemplo, los nombres y etiquetas que corresponden a cada columna son:

ID Identificador del sujeto

SEXO

TIPO Tipo (X=Chocolate / B=Blanco)

MARCA Marca (S=marca original/N=marca blanca)

ENMASCARAMIENTO Enmascaramiento (S=Si/N=No)

VALORACION Valoración (0-100)

DULCE Dulzura

Para asignar etiquetas a los valores, se debe seleccionar desde la opción FORMATS, el formato de etiqueta deseado. Cuando las etiquetas son proporcionadas por el usuario, como paso previo deben ser creadas.

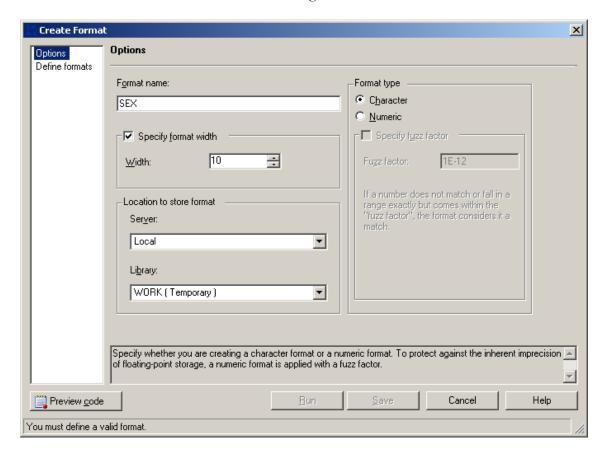
Para crear una nueva definición de etiquetas, se debe seleccionar la tarea **Create Format** del apartado **Data** de la lista de tareas. O bien desde el Menú desplegable **Data**.

Para el ejemplo, se definirán los formatos SEX y DUL con las siguientes etiquetas de valores:

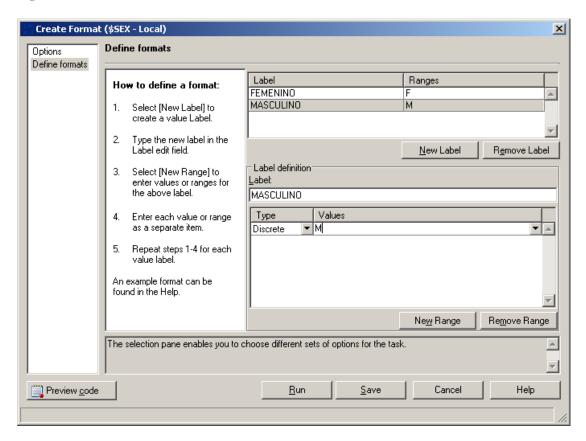
```
SEX (Carácter) DUL (Numérico)
"F"="FEMENINO" 1="NADA"
"M"="MASCULINO" 2="POCO"
3="BASTANTE"
4="MUCHO"
```

Es importante especificar el ancho de las etiquetas correctamente ya que en la presentación de resultados podrían aparecer etiquetas recortadas.

También es importante seleccionar la librería **WORK** como destino de los formatos ya que por defecto **SAS EG** busca las definiciones en esta librería. También sería posible indicar una librería de formatos distinta mediante código de sintaxis extra.

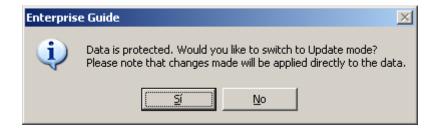


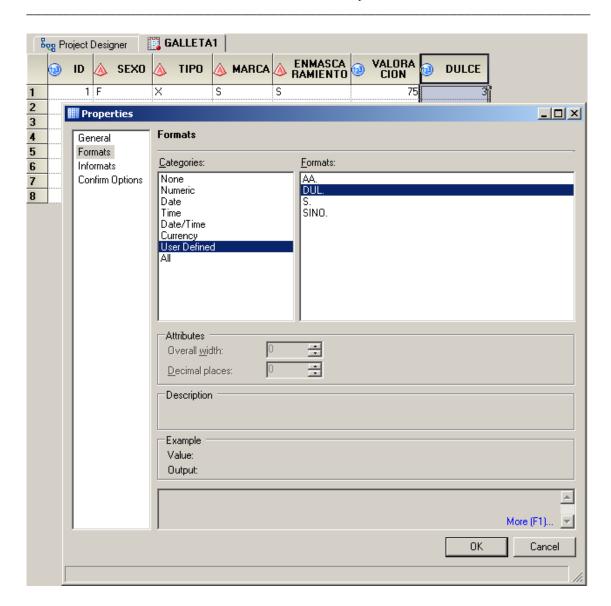
Ahora, seleccionando **New Label** y **New Range** alternativamente podemos incorporar las definiciones de las etiquetas de los valores. Es posible seleccionar rangos de valores en lugar de valores discretos.



Tras crear el formato numérico DUL repitiendo todo el proceso de nuevo, editando las propiedades de las variables de la base de datos **GALLETA1**, se podrán asignar los formatos creados a las variables que lo requieran.

Para poder editar la base de datos es necesario desactivar la protección siguiendo los pasos: **Data -> Read-only**





En las propiedades de la variable, en la opción Formatos y "User Defined" se puede seleccionar el formato creado anteriormente.

Obteniendo como resultado una visualización de las etiquetas de valores.



Seleccionando una o varias columnas o filas, el usuario puede variar el ancho/alto de visualización, esconderlas o fijarlas desde el menú:

Data -> Row

Data -> Column

También se pueden hacer búsquedas o reemplazos en la base de datos creada o existente desde:

Edit -> Search

3.3 Exportar datos

Para exportar una base de datos de **SAS**, se debe seleccionar el objeto base de datos correspondiente, pulsar el botón derecho, y marcar la opción **Export**.

Se selecciona el disco o servidor que corresponda y a continuación el tipo de fichero de la base de datos final. Entre las opciones disponibles hay: **Excel, Acces** o **Texto**.

Otra posibilidad para almacenar bases de datos temporales de **SAS** (que pertenecen a la librería **WORK**) es convertirlas en bases de datos permanentes, esto es, cambiarlas de librería.

3.4 Incorporar datos SAS

Para incorporar datos al proyecto simplemente se deben arrastrar las bases de datos (en formato SAS) de la carpeta o librería donde estén ubicadas hasta la ventana del **Project Flow**.

Siguiendo el ejemplo, se puede incorporar la base de datos GALLETA2, creada anteriormente y almacenada en formato SAS.

3.5 Importar datos

Las bases de datos con formato correcto de base de datos, es decir procedentes de programas que gestionan bases de datos (**Excel**, **Acess**, etc.) y con la estructura adecuada, se pueden trabajar directamente después de localizarlas en la ventana Server List, seleccionarlas y arrastrarlas al proyecto.

SAS EG también permite trabajar con bases de datos OLAP (cubos).

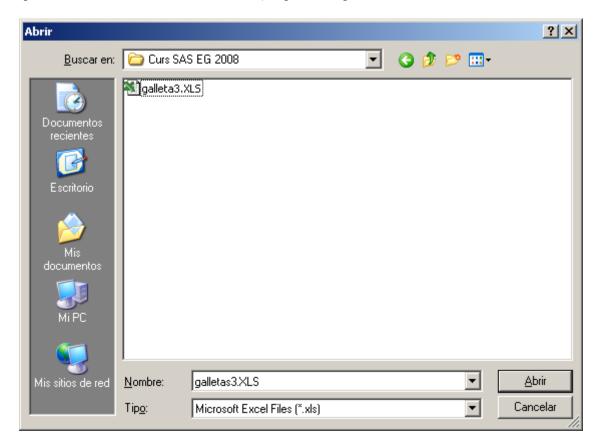
Así mismo, también puede conectarse a fuentes de datos **ODBC** u **OLE DB**.

Las bases de datos procedentes de otras fuentes como texto u otras bases de datos con estructura "extraña", deben ser sometidas a un proceso de importación.

Para importar una base de datos, se debe seleccionar el objeto inicial que se desea transformar a base de datos **SAS**, luego pulsar el botón derecho y seleccionar la opción **Import Data**, apareciendo la siguiente ventana, o en el menú **File – Import data**.



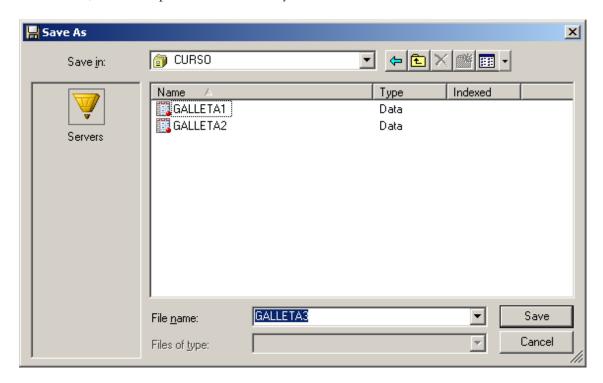
Seguidamente se selecciona el nombre del fichero que se desea importar y el tipo de fichero que es. A continuación se muestra un ejemplo de importación del fichero **GALLETA3.xls**



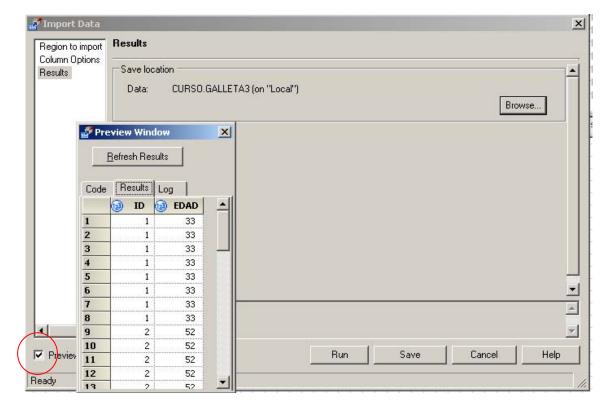
Cuando el fichero es de texto, en las diferentes opciones que aparecen, se debe especificar el tipo de estructura del fichero que se desea leer: delimitado, texto fijo, si en la primera fila están los nombres, la región a importar, etc. Seguidamente, se deben especificar las opciones de las columnas a leer: nombre, etiqueta, tipo, longitud.

A menudo, la automatización del proceso de importación provoca errores. Los más frecuentes son las confusiones provocadas por la lectura de variables numéricas en formato de cadena y las provocadas por los campos de fecha.

Finalmente, se debe especificar el nombre y librería del fichero de datos **SAS** resultante.



Es útil emplear la opción: **Preview Window (Results)**, y comprobar cómo se está importando la base de datos externa.



En este instante, el usuario debe tener las bases de datos Galleta1, Galleta2 y Galleta3 incorporadas al proyecto.

La base de datos **Galleta1** contiene información sobre el sujeto 1 La base de datos **Galleta2** contiene información sobre los sujetos 2 a 32

La base de datos Galleta3 contiene la edad de los sujetos 1 a 32

4 FUSIÓN DE ARCHIVOS

4.1 Ordenar datos

Una de las tareas previas para fusionar datos es la ordenación de las bases de datos.

Para ordenar una base de datos, se arrastra la tarea **SORT** des de el apartado **Data** del menú de tareas.

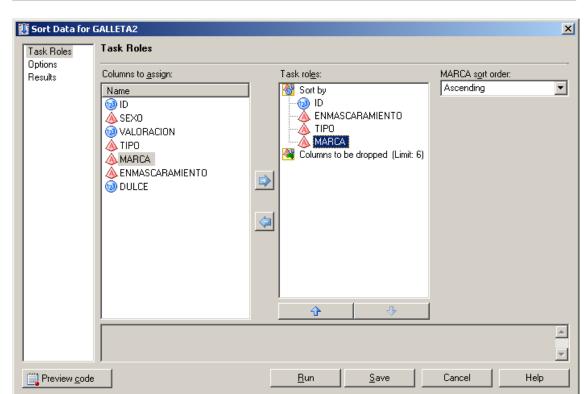
Para realizar cualquier tarea con una base de datos que ha sido modificada por el usuario, ésta debe ser protegida, para evitar manipulaciones posteriores.



Para desproteger la base de datos y poder editarla de nuevo, es necesario desactivar la protección siguiendo los pasos:

Data -> Read-only

Tras aceptar la protección, aparece la ventana de la tarea **SORT**, donde se pueden seleccionar las variables clave para definir el orden.



En esta misma pantalla, se puede indicar si se desean las variables en orden ascendente o descendente.

En la ventana de opciones de la tarea **SORT**, se puede indicar que se eliminen los casos duplicados.

En la ventana de resultados de esta tarea, se selecciona la base de datos resultante.

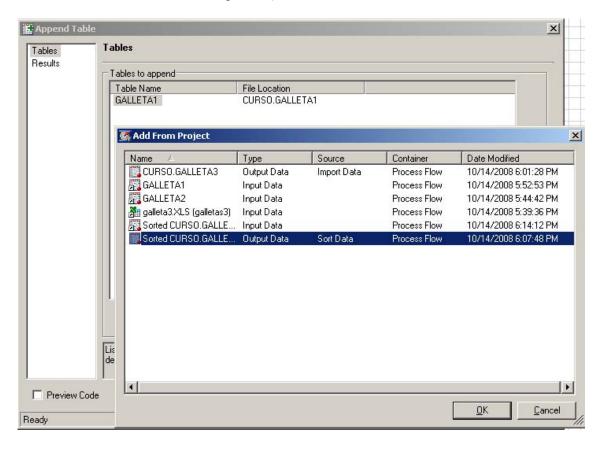
En general, todas las tareas funcionan de forma similar, una ventana de selección de variables, una de opciones específicas y una ventana para indicar cómo deseamos almacenar el resultado. También existe la posibilidad de previsualizar el código **SAS** de la tarea.

4.2 Fundir archivos: Añadir casos

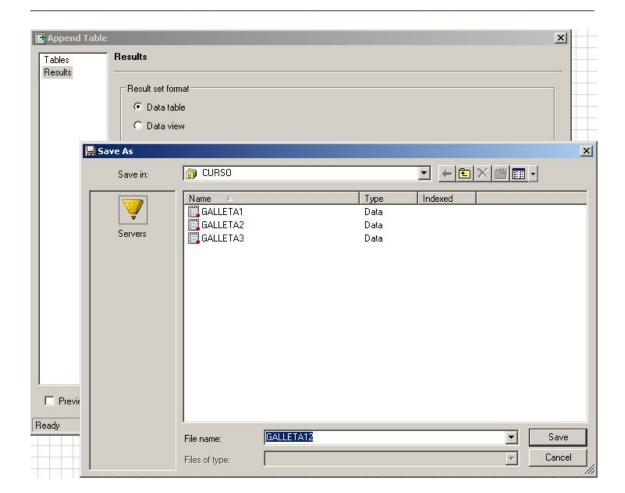
La tarea necesaria para añadir datos es **Append Table**, del grupo de tareas **Data**. Mediante el botón **Add Table**, se especifican las bases de datos a fusionar.

Al añadir casos, no es necesario ordenar previamente.

En el ejemplo, es necesario fundir las bases de datos **Galleta1** y **Galleta2**. Contienen exactamente la misma estructura, pero sujetos diferentes.



En la opción **Results**, se puede especificar la base de datos resultante y la librería, por ejemplo **CURSO.GALLETA12**



Finalmente el resultado es una base con 256 observaciones y 7 columnas.

4.3 Fundir archivos: Añadir variables

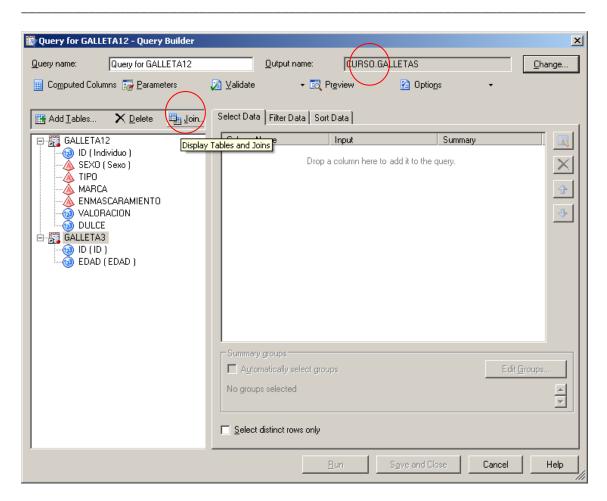
La fusión de dos archivos se puede realizar mediante la opción: Filter and Query, del bloque de tareas Data.

En primer lugar se añaden las bases de datos que se desean fusionar, pulsando el botón **Add Tables** y seleccionando el proyecto como lugar donde se encuentran los datos.

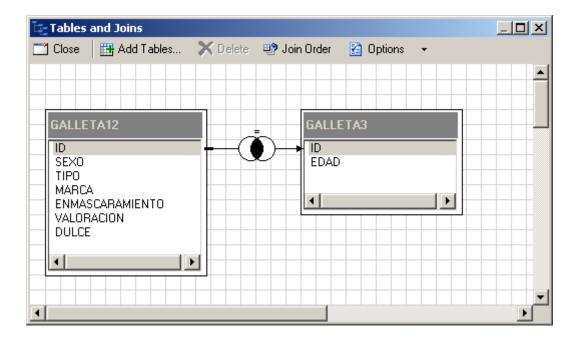
Siguiendo el ejemplo, seleccionar las bases de datos **Galleta12** y **Galleta3**. Ambas variables están identificadas por la variable ID.

La base de datos Galleta3 únicamente contiene un registro por individuo.

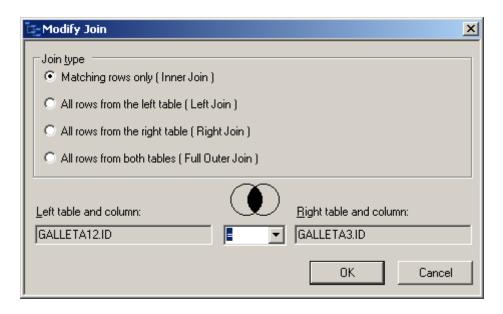
Al activar la ventana, el usuario puede cambiar el nombre de la base de datos resultante.



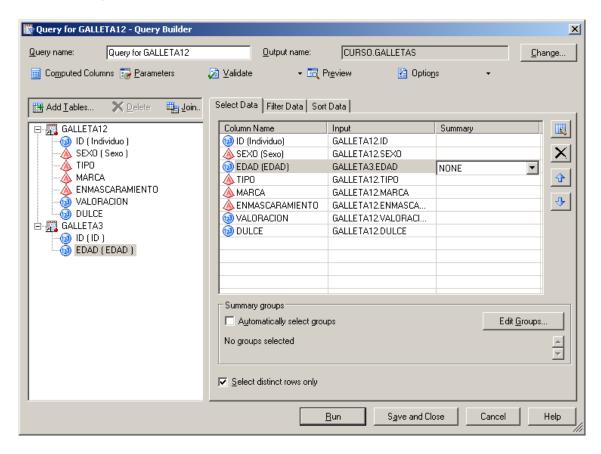
En la pestaña **Join** se abre otra ventana donde aparecerá la vista de la fusión que se realizará.



Es posible modificar la estructura de la fusión de modo que haya una base de datos principal (**Left – Join, Right Join, Full Join**), pulsando con el botón derecho el símbolo = y seleccionando **Modifiy Join**. El programa, propone una clave de fusión. Si no es la deseada, se puede eliminar y arrastrar una de las claves hasta la correspondiente clave en la otra base de datos.



En la pestaña **Select data**, se debe añadir las variables que se desea disponer en la nueva base de datos, arrastrándolas del menú de tablas.

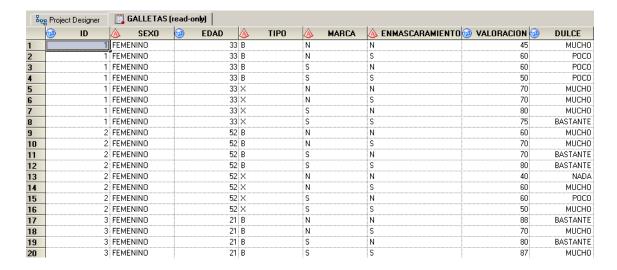


Se puede modificar el orden de las variables seleccionando una variable y pulsando los botones: .

Se puede pre-visualizar (Preview - Results) los resultados y comprobar que la fusión es correcta.

Con el botón Validate, se comprueba que la sintaxis es correcta.

El resultado de la fusión es la siguiente base de datos:

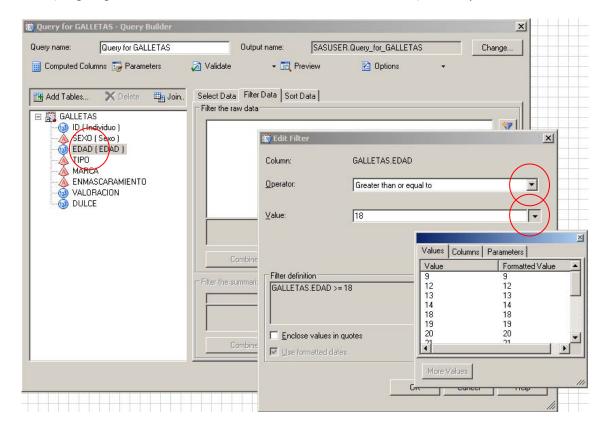


5 GESTIÓN DE DATOS

5.1 Filtrar casos

Tras seleccionar una base de datos, en el menú, **Data**, se debe escoger **Filter and Query**. Para filtrar datos, se arrastra el campo que se desea filtrar a la región **Filter Data**. Se abre una nueva ventana donde podemos editar el filtro.

Por ejemplo, podemos crear una nueva base de datos con los sujetos mayores de edad.



Con la opción Get Values, se obtienen las modalidades de esa variable.

Una vez seleccionado el filtro, se debe indicar qué variables se desean para la nueva base de datos. Para ello, se deben arrastrar las variables de interés en la pestaña **Select Data**.

Pulsando el botón: 🍱 Add a new filter, se pueden diseñar filtros más complejos.

Cuando el filtro no es correcto, aparece un mensaje de error. En caso contrario, se pueden visualizar los resultados para comprobar que el resultado de la tarea es el deseado.

La base de datos resultante contiene únicamente filas en que la condición del filtro se cumple.

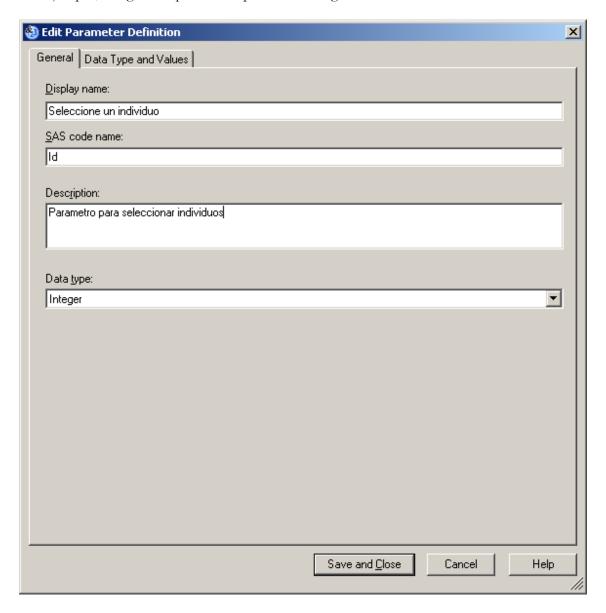
5.2 Filtros dinámicos

Otra posibilidad interesante para usuarios finales es la configuración de filtros dinámicos. Los valores filtrados son introducidos de forma dinámica por el usuario cuando se ejecuta la tarea.

En primer lugar se deben seleccionar todas las variables en el apartado **Select Data**.

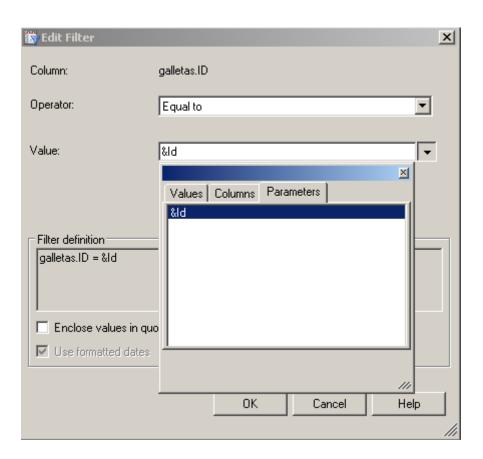
A continuación, en la ventana de **Query**, se debe crear un parámetro, desde la pestaña

Por ejemplo, el siguiente parámetro permitirá escoger uno de los 32 individuos.

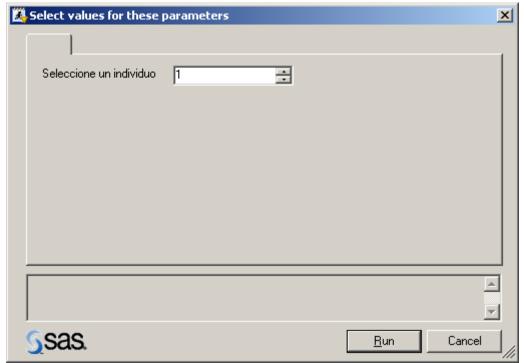


Ahora, creamos un filtro para la variable ID en la pestaña Filter Data.

Seleccionamos como valor en la pestaña Parameters, el parámetro &id creado anteriormente.



Ahora, al ejecutar la Query, el programa nos solicita que introduzcamos un número para la variable ID.



Servei d'Estadística Universitat Autònoma de Barcelona

Así, al ejecutar la tarea, se consigue un filtro dinámico.



Cada vez que se ejecuta esta tarea, el programa pide de nuevo la introducción del parámetro.

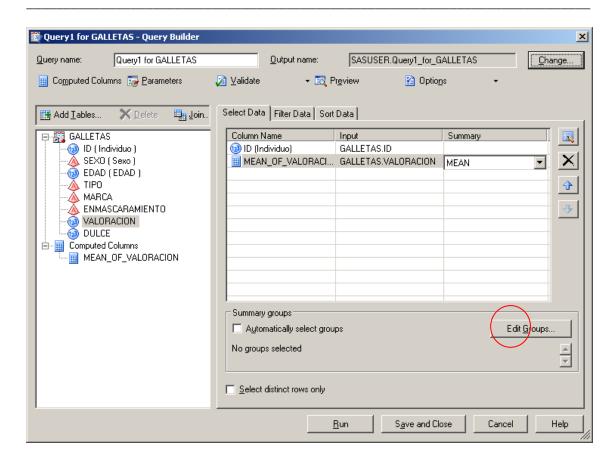
Igualmente, la base de datos resultante contiene únicamente casos en que la condición del filtro se cumple.

5.3 Agregar datos

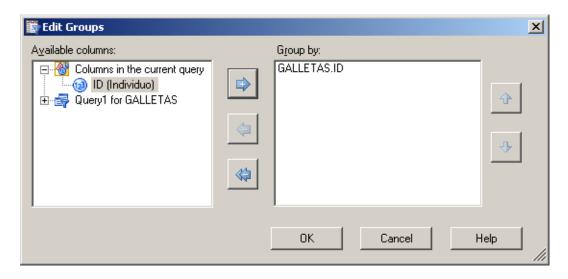
Para agrupar datos, se puede emplear una nueva **Query**. Se selecciona la base de datos de trabajo, y se escoge la tarea **Filter and Query** del grupo de tareas **Data**.

En la ventana **Select Data**, se añaden las variables que nos interesan, se definen las variables de agrupación con el botón **Edit Groups** y se especifica las medidas de resumen de los datos que se van a agrupar (contador, suma, promedio, etc.).

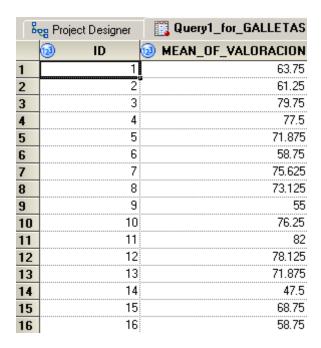
Por ejemplo, para obtener una agregación de las valoraciones por individuo utilizando la base de datos **GALLETAS**:



Pulsando Edit Groups, se identificarán las variables clasificatorias o de agrupación.



La ejecución de esta **Query** da lugar a una nueva base de datos con los valores agrupados (calculando el promedio de la valoración, en este caso) según el individuo (ID):

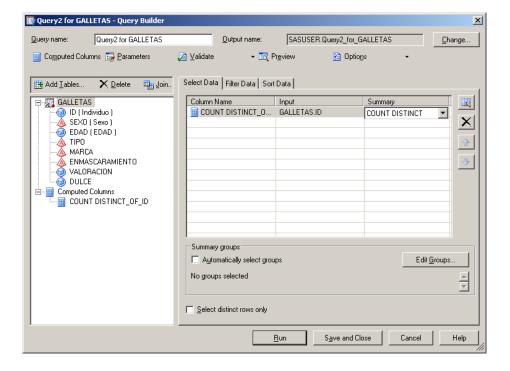


5.4 Contar datos

Otra utilidad para controlar y comprobar si las operaciones con las bases de datos se han realizado correctamente es el recuento de casos.

Para contar datos, se puede emplear una nueva **Query**. Se selecciona la base de datos de trabajo, y se escoge la tarea **Filter and Query** del grupo de tareas **Data**. La función que permite contar es **COUNT**, aunque si hay casos duplicados, es adecuado **COUNT DISTINCT**.

Por ejemplo, para contar cuántos sujetos diferentes hay en la base de datos.

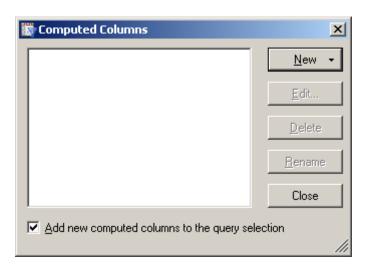


5.5 Crear nuevas variables

Para crear nuevas variables a partir de variables existentes, se puede emplear una nueva **Query**.

Se selecciona la base de datos de trabajo, y se escoge la tarea **Filter and Query** del grupo de tareas **Data**.

En la opción **Computed Columns:** Que aparece en el menú se crea una variable y se editan sus propiedades.



El usuario puede recodificar o construir una variable a partir de una fórmula.

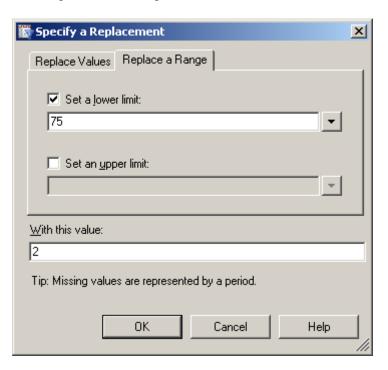


Por ejemplo, para crear una variable que recodifique la Valoración en >75 o <=75, podemos emplear la recodificación siguiente empleando rangos:

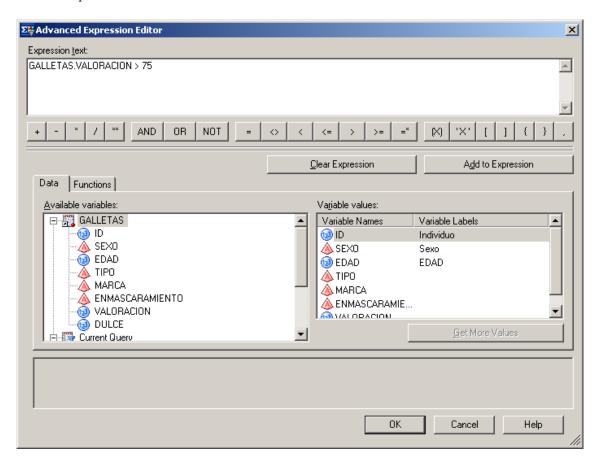
Es necesario comprobar cómo se han recodificado los valores en caso de que los rangos se solapen.

🗑 Recode Column - GALLETAS.VALORACION Recode_VALORACION New column name: <u>A</u>dd... Remove Replacements: With Replace <= 75 >= 75 2 Other values: Replace all other values with: • The current value C A missing value <u>I</u>his value: New column type Character Numeric ΟK Cancel Help

En el apartado inferior podemos especificar cómo se recodificarán los valores no contemplados en las especificaciones del usurio.



o bien empleando una fórmula:



Con doble-click en la variable creada, se pueden modificar sus propiedades.

Es posible introducir expresiones altamente complejas, por lo cual el programa proporciona el listado de funciones disponibles y valores de las variables existentes.

En la recodificación mediante la tarea **Query**, es recomendable añadir en la lista de variables, la variable antigua que ha sido recodificada, de este modo, no se pierde la información inicial.

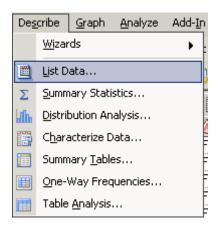
6 RESUMEN DE DATOS

6.1 Listar Datos

Un proceso habitual para comprobar si las operaciones con los ficheros se han realizado correctamente, es listar los datos. Sin embargo, las opciones disponibles permiten que los listados puedan convertirse en auténticos informes.

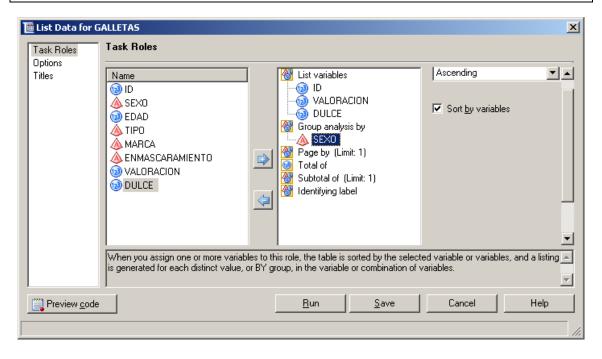
A partir de esta sección, en los ejemplos se emplea la base de datos CURSO.GALLETAS.

La tarea que permite listar los datos se encuentra en el grupo **Describe** y es **List Data**.



Tras seleccionar la base de datos que se desea listar, se seleccionan aquellas variables que aparecerán en el informe.

Las variables asignadas a **Group analysis by** provocarán análisis separados para cada una de sus categorías. Esta propiedad es común en prácticamente todas las tareas.



En la ventana **Opciones** se pueden especificar diferentes opciones de presentación, etiquetas, títulos así como especificar si el listado será sólo de un número pequeño de datos.

El resultado del listado es:

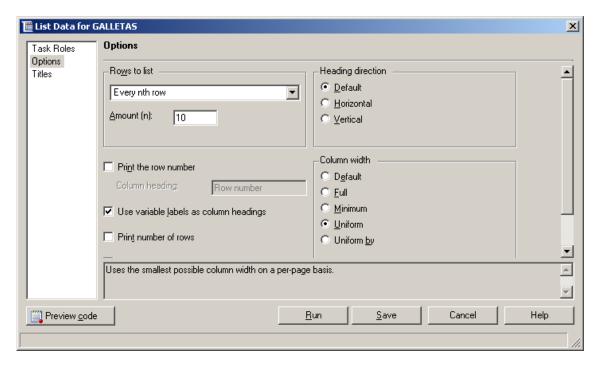


Report Listing

Sexo=FEMENINO

Row number	Individuo	VALORACION	DULCE
1	1	45	мисно
2	1	60	POCO
3	1	60	POCO
4	1	50	POCO
5	1	70	MUCHO

El usuario puede personalizar los listados añadiendo diferentes opciones de presentación. También puede seleccionar listar una muestra de los datos.



6.2 Resumir los datos

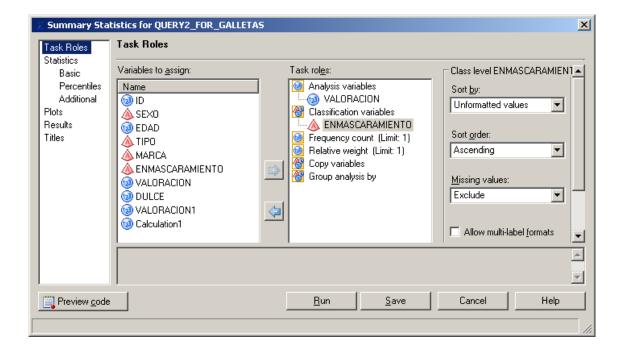
Para resumir los datos y obtener Estadísticos de resumen básicos se debe seleccionar **Summary Statistics** del grupo de tareas **Describe**.

El usuario puede seleccionar una o más variables cuantitativas y si lo desea algunas variables de agrupación.

Las variables asignadas a **Frequency count** o bien **Relative weight** permiten asignar pesos a las observaciones de la base de datos (ponderar los casos). Esta posibilidad es común en todas las tareas de resumen y análisis.

Se puede escoger qué índices de resumen se desea mostrar.

En la opción **Plot**, se pueden solicitar gráficos univariantes: histograma y diagrama de cajas.



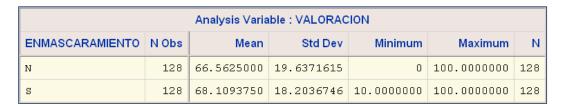
También es posible indicar que los resultados se almacenen directamente en una nueva base de datos, de forma que esta función es equivalente a la función de agrupación de datos vista anteriormente.

El resultado es:

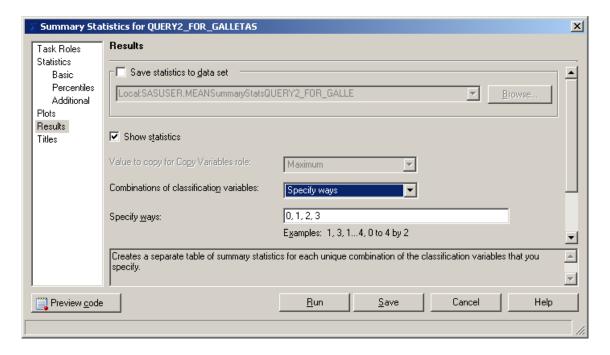


Summary Statistics Results

The MEANS Procedure



Si el usuario selecciona diversas variables de clasificación, puede también especificar qué nivel de combinación de categorías desea: totalmente anidado, sólo variable a variable, etc, en la ventana:

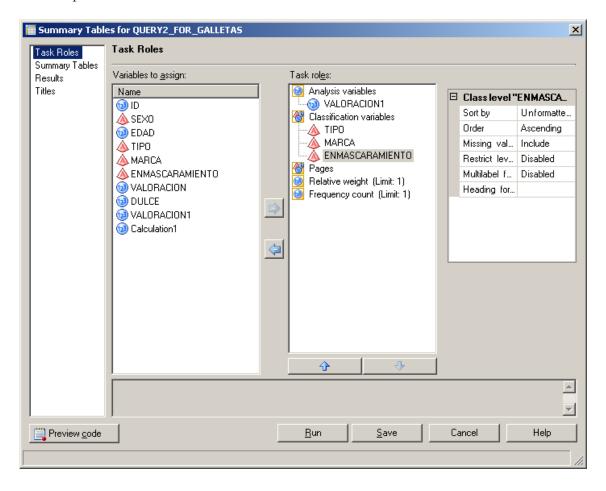


6.3 Tablas de resumen personalizadas

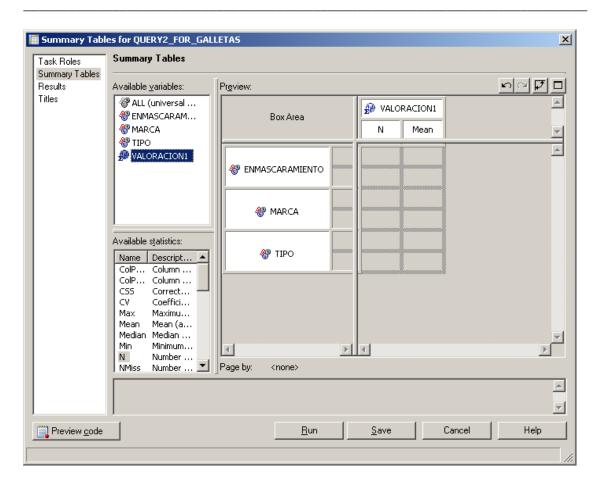
Otra tarea que permite resumir datos es Summary Tables.

Sin embargo, esta tarea permite configurar una tabla resumen personalizada.

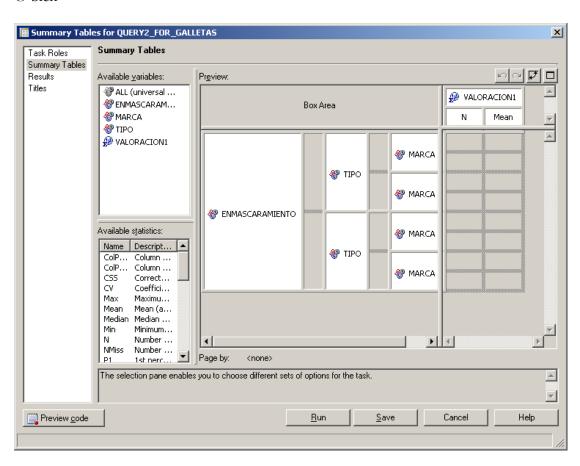
En la ventana de selección de variables se debe indicar el papel de cada variable que formará parte de la tabla.



A continuación, en la ventana de **Summary Tables**, se puede construir la tabla deseada arrastrando cuidadosamente las variables y las medidas de resumen hasta la ubicación deseada.



O bien



Esta tarea, permite obtener resultados con un formato más personalizado, incluyendo si se desean las etiquetas de las variables.

El resultado es:



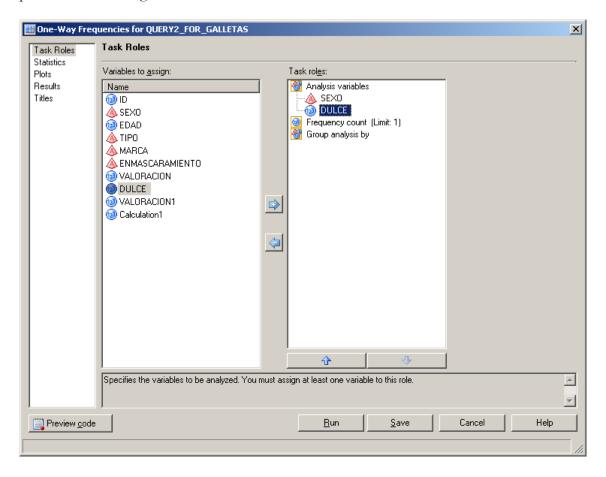
The Power to Know...

Summary Tables

			VALC	RACION1
			N	Mean
ENMASCARAMIENTO	TIPO	MARCA		
N	В	N	32	67.81
		S	32	80.59
	X	N	32	51.03
		S	32	66.81
s	В	N	32	74.84
		S	32	76.50
	X	N	32	53.75
		S	32	67.34

6.4 Tablas de frecuencia

Para generar tablas de frecuencia, existe la tarea: **One-Way frequencies**, siendo también posible solicitar un gráfico de barras.



En la ventana **Statistics**, es posible calcular la prueba Chi-Cuadrado para tablas de frecuencia

El resultado es:



The Power to Know...

One-Way Frequencies Results

The FREQ Procedure

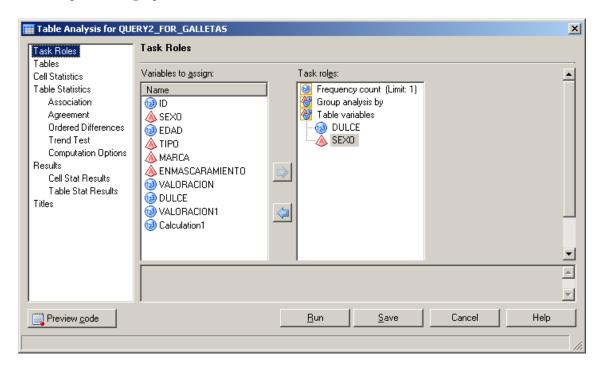
Sexo						
SEXO	Frequency Percent Cumulative Cumulative Per					
F	72	28.13	72	28.13		
М	184	71.88	256	100.00		

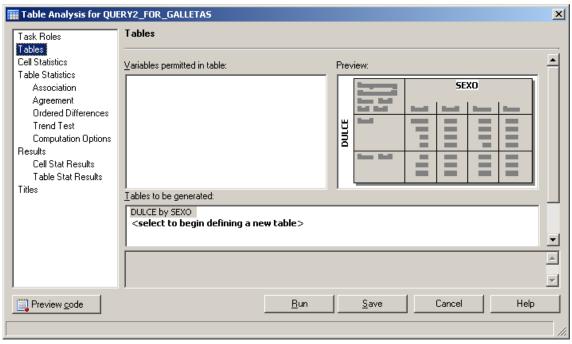
DULCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	14	5.47	14	5.47
2	47	18.36	61	23.83
3	98	38.28	159	62.11
4	97	37.89	256	100.00

6.5 Tablas de contingencia

Para obtener tablas de contingencia, existe la tarea **Table Analysis**. En primer lugar se han de identificar al menos dos variables categóricas, y a continuación, configurar la tabla y los estadísticos que se resumirán en cada casilla.

A continuación se muestra un ejemplo de una tabla de contingencia, donde se quiere obtener para cada grupo de la variable **SEXO**, la distribución de la variable **DULCE**.





En la ventana **Table Statistics**, es posible calcular la prueba Chi-Cuadrado u otras pruebas para tablas de contingencia

El resultado es:



Enterprise Guide»

The Power to Know...

Table Analysis Results

The FREQ Procedure

Frequency Col Pct

Table of DULCE by SEXO							
	SEXO	(Sexo)					
DULCE	F	М	Total				
1	2 2.78	12 6.52	14				
2	16 22.22	31 16.85	47				
3	24 33.33	74 40.22	98				
4	30 41.67	97					
Total	72	184	256				

7 GRÁFICOS

Existen numerosas posibilidades para generar gráficos, entre ellos los más habituales:

- Pasteles / Sectores
- Barras
- Histograma
- Líneas
- Diagramas de caja
- Dispersión

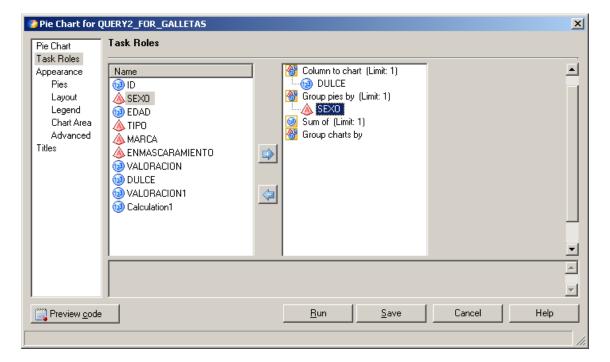
Todas estas tareas se encuentran el grupo Graph.

En cada caso, el usuario deberá seleccionar las opciones que considere apropiadas, pero es recomendable realizar un repaso de todas las opciones que ofrece el programa.

7.1 Gráfico de sectores

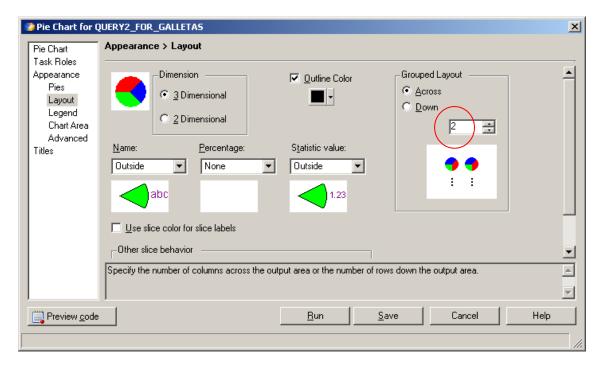
Para obtener un gráfico de sectores se debe seleccionar la tarea Pie Chart.

Se pueden obtener gráficos simples o agrupados.



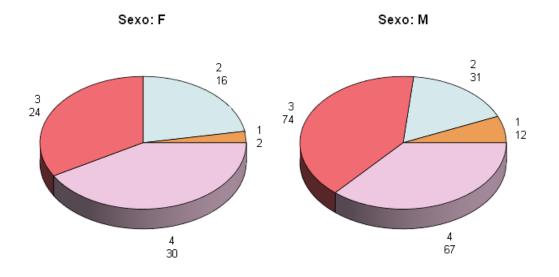
Los sectores representan las frecuencias relativas, en caso de seleccionar una variable en el rol **Sum of**, se obtendrían sectores en función de la suma de la variable de interés.

Es recomendable personalizar la representación.



El resultado del gráfico es:



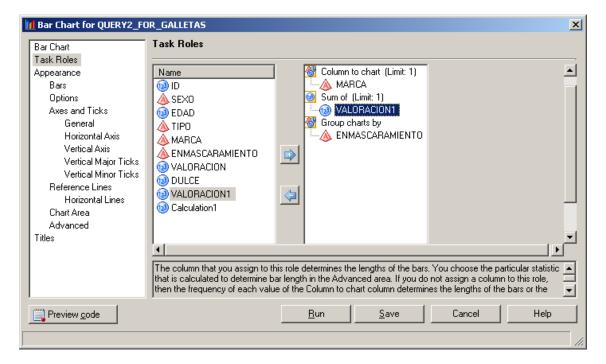


7.2 Gráfico de barras

Para obtener un gráfico de barras se debe seleccionar la tarea Bar Chart.

En caso de asignar una variable a **Sum of**, las barras representarán la suma, el promedio o porcentajes relativos a esta variable.

En la opción **Bars**, será necesario seleccionar **One bar for each unique value**, si la variable que representa las columnas es una variable discreta.



En la opción **Advanced**, si se ha solicitado una variable de resumen, se podrá especificar si se requieren barras de error para los intervalos de confianza.

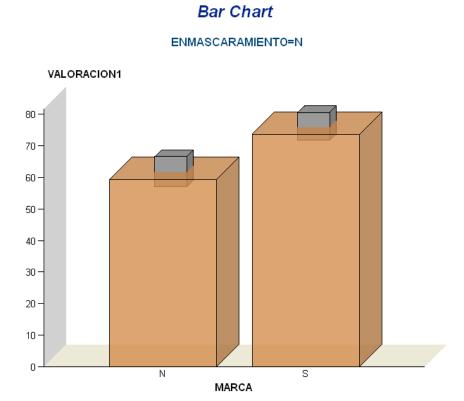
M Bar Chart for QUERY2_FOR_GALLETAS X Appearance > Advanced Bar Chart Task Roles • Appearance Statistic used to calculate bar: Accept missing values Bars Options Average ▾ Skip zero values Axes and Ticks General Error bars are valid only when the chart statistic is Average or Percentage and the chart type is not a 3D grouped bar chart. Horizontal Axis Calculate percentages and cumulative percentages for each group (G100) Vertical Axis Vertical Major Ticks Vertical Minor Ticks Display error bars Reference Lines CLM: Horizontal Lines Chart Area 95 🚉 Additional statistical value to show next to bar: Bars • Advanced Titles Error bar color: Displays confidence intervals on the bars Preview code <u>R</u>un Save Cancel Help

Enterprise Guide»

The Power to Know...

El gráfico de barras resultante es:

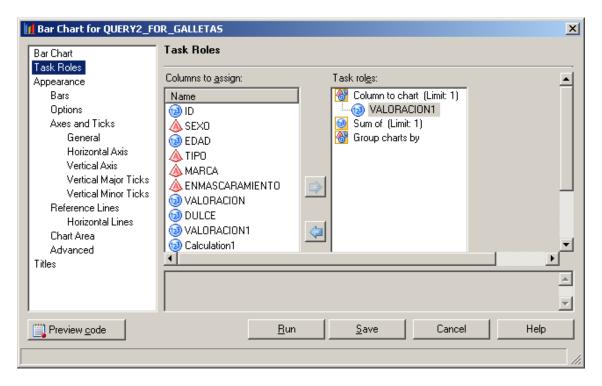
SSAS.



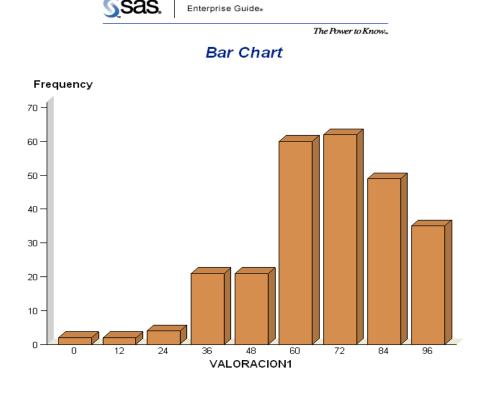
Ahora, el gráfico se puede editar y personalizar activando con el botón derecho la barra de herramientas del gráfico o las propiedades del gráfico. Seleccionando tipo de diagrama, se puede reconvertir el gráfico inicial a otro tipo de gráfico.

7.3 Histograma

La tarea Bar Chart también permite obtener histogramas seleccionando **Simple Vertical Bar**. Para que el programa cree las marcas de clase, no se debe indicar que la variable es discreta.



El histograma también se puede obtener desde la tarea **Summary Statistics**. El resultado es:

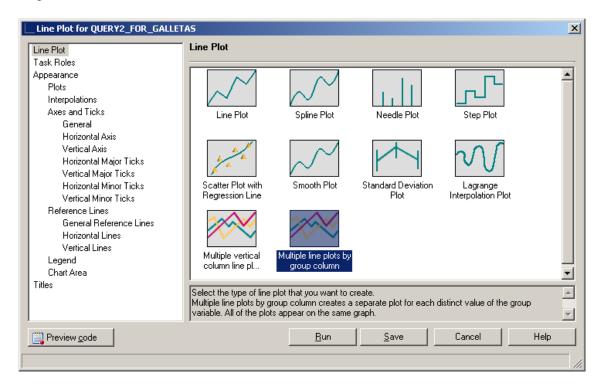


.....

7.4 Diagrama de dispersión

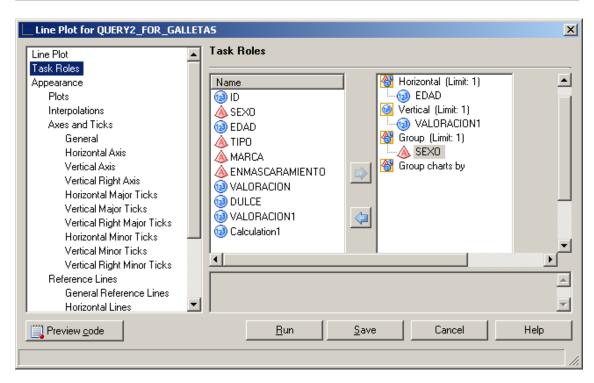
Para obtener un diagrama de dispersión, seleccionar la tarea Line Plot. La tarea Scatter Plot también permite representar diagramas en 3 dimensiones.

En función del tipo particular de diagrama, el usuario debe seleccionar qué gráfico desea emplear.

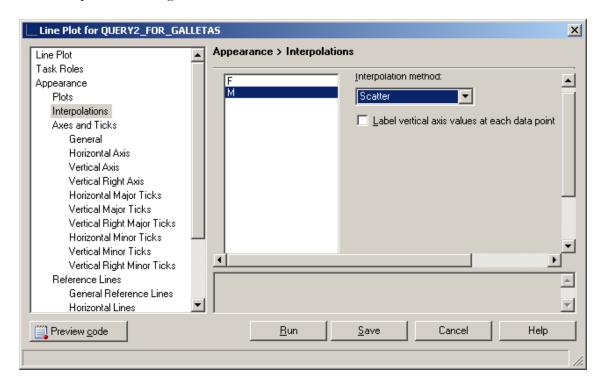


Por ejemplo, para representar la valoración respecto la Edad, identificando los individuos, podríamos emplear el gráfico: **Multiple plots by group column**

Es recomendable situar la variable respuesta en el eje vertical.



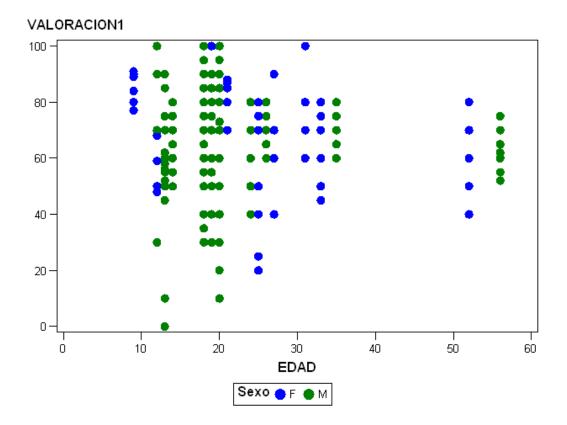
En la opción **Interpolation**, el usuario puede seleccionar qué representación de los puntos desea considerar: ajuste lineal (**Line**), diagrama de puntos (**Scatter**). Esta selección debe realizarse para cada categoría de la variable de clasificación.



El resultado del gráfico es:

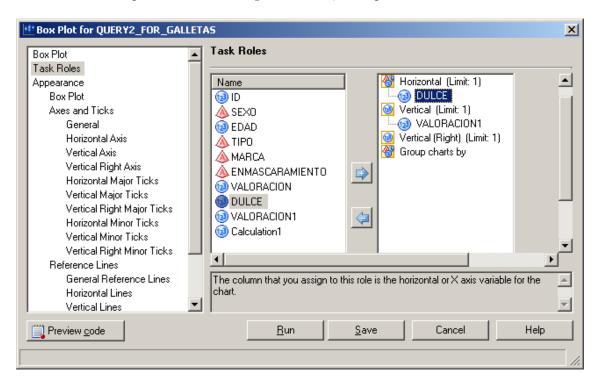


Line Plot



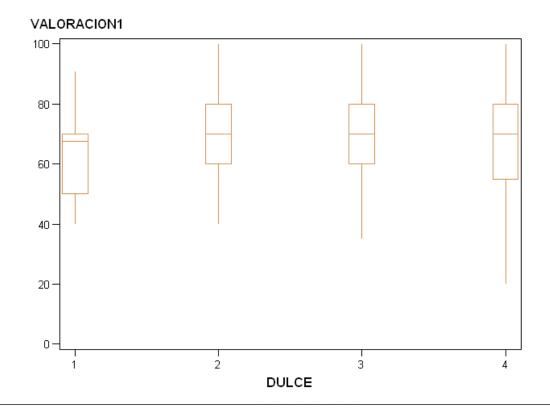
7.5 Diagrama de cajas

La tarea **Box Plot** permite obtener diagramas de caja comparativos.



El diagrama de caja también se puede obtener desde la tarea **Summary Statistics**. El resultado es:

Box Plot



8 OTRAS OPERACIONES CON DATOS

8.1 Transposición de datos

Existen básicamente dos formas de transponer los datos.

A partir de unos datos que recogen medidas repetidas en los mismos sujetos y estas medidas se encuentran en diferentes columnas, una forma de transposición consiste en obtener diferentes filas para cada una de las medidas repetidas.

IND	M1	M2		IND	V	M
1	10	20	\rightarrow	1	1	10
2	11	21		1	2	20
3	12	22		2	1	11
				2	2	21
				3	1	12
				3	2	22

Este tipo de transposición se lleva a cabo mediante la tarea **Stack Columns**, del grupo de tareas **Data**

En la casilla **Group analysis by**, será necesario incluir una o varias variables de forma que cada fila de la base de datos quede identificada de forma única.

El resultado de este proceso de transposición sería:

	1ND		StackedValues
1	1	M1	10
2	1	M2	20
3	2	M1	11
4	2	M2	21
5	3	M1	12
6	3	M2	22

Otra posible manera de transponer los datos consiste en obtener diferentes columnas a partir de las medidas repetidas en diferentes individuos que se encuentran en diferentes filas.

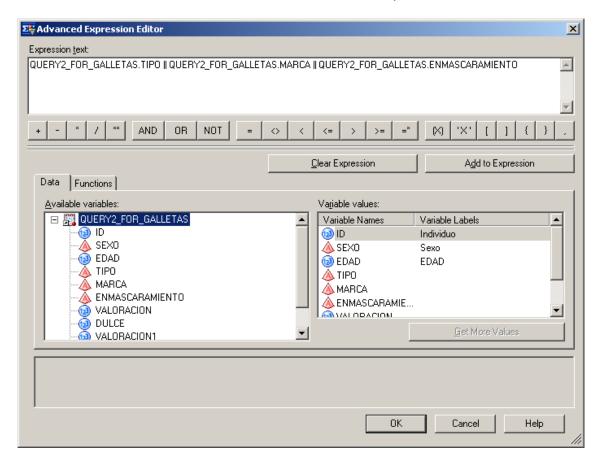
IND	V	M		IND	M1	M2
1	1	10		1	10	20
1	2	20		2	11	21
2	1	11	\rightarrow	3	12	22
2	2	21				
3	1	12				
3	2	22				

Este tipo de transposición se lleva a cabo mediante la tarea **Split Columns**, del grupo de tareas **Data**.

En el proceso de transposición se pierden las variables que no intervienen en ningún rol.

Por ejemplo, si se desea transponer los datos de cada individuo, podemos realizar las operaciones siguientes:

Primero crear una única variable para identificar cada una de las valoraciones. Por ejemplo, recodificando las variables TIPO, ENMASCARAMIENTO y MARCA.



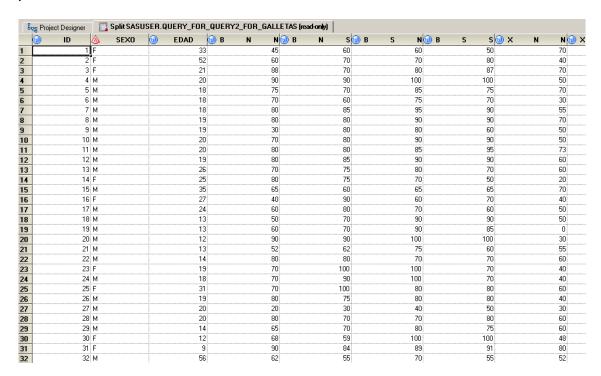
La expresión contempla la concatenación de los valores de las variables de interés.

Esta variable se ha renombrado a **IDVALORACION**.

Ahora, emplear la tarea SPLIT.

Split Columns for QUERY_FOR_QUERY2_FOR_GALLETAS × Task Roles Task Roles Results Ascending ◥▴ 🞳 Column to split (Limit: 1) Name VALORACION 📵 ID 🚳 Value identifier column (Limi 🔌 SEXO ✓ Sort by variables 🔌 IDVALORACION 📵 EDAD Value label column (Limit: 1) 🕭 TIPO 發 Group analysis by 🔌 MARCA 13 D ENMASCARAMIENTO SEXO ∀ALORACION EDAD DULCE WALORACION 1 Calculation1 IDVALORACION | F Runs the task with the options that you have selected. 📜 Preview <u>c</u>ode <u>R</u>un Save Cancel Help

y el resultado es:



9 PRESENTACIÓN

9.1 Presentación de resultados

SAS EG permite almacenar los resultados en distinto formatos: WORD (RTF), HTML, PDF o texto.

El usuario puede indicar qué tipo de resultados prefiere modificando las opciones mediante:

Tools -> Options -> Results

y seleccionando otro tipo de formato de salida.

Mediante:

Tools -> Options -> Tasks -> Tasks general

se pueden cambiar los títulos que se aplican a todos los resultados.

También puede cambiarse y personalizarse el estilo de resultados en (incluso el gráfico):

Tools -> Style Manager

Finalmente, dispone de la opción de construir documentos a partir de los objetos de salida del proyecto mediante el Document Builder

Tools -> Create HTML Document

Tras abrir el **Document builder**, el usuario puede añadir aquellos objetos de resultado obtenidos en formato **HTML** que desee incorporar en un informe final.

10 ASPECTOS AVANZADOS

10.1 Sintaxis de SAS

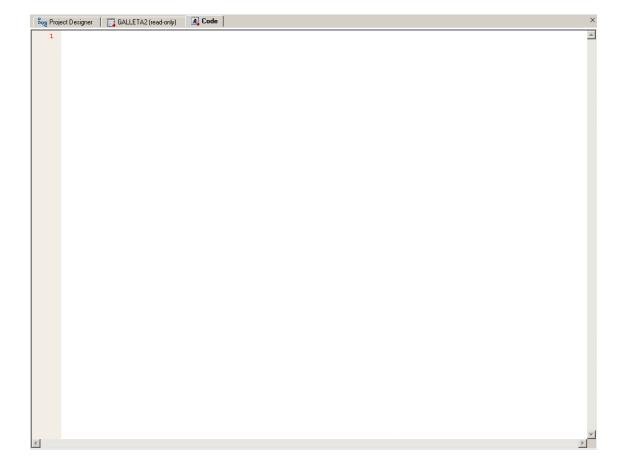
Una forma habitual de trabajar con el programa SAS System es emplear sintaxis para reflejar todos los pasos que realiza el usuario de forma compacta. En SAS Enterprise Guide, se puede escribir sintaxis de SAS.

Al abrir una tarea y seleccionar las opciones pertinentes, también aparece la casilla **Preview Code**, que permite visualizar el correspondiente código de SAS System.

El usuario tiene además la posibilidad de personalizar la sintaxis de la tarea pulsando el botón Insert Code.

Otra posibilidad más práctica para poder incorporar código de SAS System, consiste en añadir una tarea específica de código. Esta tarea recibe el nombre de **Create Code** y se encuentra en el grupo: **Create New Items in Project**. También se puede pulsar el botón

Al abrirlo aparece una hoja editable.



El código se ejecuta después de cerrar la ventana, seleccionando el icono de código y seleccionando desde el menú, **Code** y **Run code on local**. Alternativamente, pulsando **F8** o bien con el botón derecho del ratón, y seleccionando **Run code on local**.

10.2 Otros análisis estadísticos

Aparte de las tareas comentadas, **SAS EG** proporciona un gran número de procedimientos estadísticos, entre ellos:

- Análisis de la distribución de los datos
- Correlaciones
- Calcular rangos o estandarizar variables
- Modelos de regresión
- Análisis de la Varianza
- Control de Calidad
- Análisis Multivariante
- Series temporales
- Análisis de la supervivencia

11 EJERCICIOS

11.1 Ejercicio 1

Crear una base de datos que se llame ESTUDIOFC1 a partir de los siguientes datos:

IB	SEXC) FUN	MADOR	EDA	ADFC1	FC2	FC2	FC1	STATUS FARMACO
4	2	2	22.6	96	155	59	1	1	
2	2	1	20	78	154	76	1	2	
23	2	1	21.6	76	148	72	1	1	
36	1	1	21.4	62	132	70	1	2	
35	2	2	24.5	80	126	46	1	1	

Las etiquetas de las variables son:

FC1 – Frec. Cardiaca Inicial FC2 – Frec. Cardiaca Inicial

FC2FC1 - Incremento Frec. Cardiaca

STATUS - Status Clínico

Los valores que deben codificarse son:

SEXO 1=Hombre 2=Mujer
FUMADOR 1=No 2=Sí
STATUS 1=Normal 2=Alterado
FARMACO 1=Tratamiento A 2=Tratamiento B

11.2 Ejercicio 2

Importación de ficheros. Es recomendable visualizar el contenido de las bases de datos una vez incorporadas al proyecto, para evaluar qué opciones serán necesarias.

- a) Importar la base de datos "FixWidth.txt"
- b) Importar la base de datos "CommaSep.txt" Indicación: Explorar la base de datos después de incorporarla al proyecto.
- c) Importar la base de datos "as_products.xls" Indicación: Es necesario leer el nombre de las variables.
- d) Importar la base de datos "sales.html" Indicación: Es necesario leer el nombre de las variables.

.....

11.3 Ejercicio 3

Fundir el archivo ESTUDIOFC1 con el archivo ESTUDIOFC2, creando la base de datos ESTUDIOFC3

11.4 Ejercicio 4

Evaluar de forma descriptiva si existen diferencias entre tratamientos en la variable FC2. ¿Qué variables del estudio están relacionadas con la variable FC2?

11.5 Ejercicio 5

Contestar a las siguientes preguntas referidas a la base de datos GALLETAS mediante un análisis descriptivo:

- a) ¿Hay diferencias en la valoración de las galletas según SEXO?
- b) ¿Hay relación entre la Edad y la Valoración?
- c) ¿El enmascaramiento influye en la Valoración?
- d) ¿El enmascaramiento influye en la Valoración por igual en las galletas de marca original y en las galletas sin marca?
- e) ¿El enmascaramiento influye en la Valoración por igual en las galletas de chocolate de marca original y en las galletas de chocolate sin marca?

12 ANEXO

Muestra del project flow después de terminar todos los pasos siguiendo el ejemplo.

